Τεχνολογικό Εκπαιδευτικό Ίδρυμα Δυτικής Ελλάδας

Σχολή Διοίκησης και Οικονομίας

Τμήμα Πληροφορικής και ΜΜΕ

(ΠΑΡΑΡΤΗΜΑ ΠΥΡΓΟΥ)

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Τίτλος: Δημιουργία συστήματος αυτόματης αναγνώρισης συναισθήματος σε τραγούδια

Χατζησταμάτη Στέλλα

Επιβλέπων Καθηγητής: Κούτρας Αθανάσιος

2014

‘Music is the shorthand of emotion’

Leo Tolstoy

Abstract

The emotion recognition from music is a fundamental problem in the field of computers. In recent years there have been many studies using systems that serve this purpose. In this thesis, we focus on creating a system that has as main objective to automatically detect and classify emotions on music. This study concerns the recognition of emotions in music files (mp3).

Περίληψη

Η αναγνώριση συναισθημάτων από μουσικά κομμάτια αποτελεί ένα θεμελιώδες πρόβλημα στον τομέα των ηλεκτρονικών υπολογιστών και της αναγνώρισης προτύπων. Τα τελευταία χρόνια έχουν πραγματοποιηθεί πολλές μελέτες χρησιμοποιώντας συστήματα που εξυπηρετούν αυτόν τον σκοπό. Σε αυτήν την πτυχιακή εργασία, θα επικεντρωθούμε στην δημιουργία ενός συστήματος που έχει ως κύριο στόχο να αναγνωρίζει αυτόματα και να ταξινομεί τα συναισθήματα μουσικών κομματιών. Εδώ η μελέτη αφορά την αναγνώριση συναισθημάτων σε μουσικά αρχεία (mp3).

Περιεχόμενα

Κεφάλαιο 1 - Μουσική και συναίσθημα

1.1. Εισαγωγή

Μουσική: Ως μουσική ορίζεται η τέχνη που βασίζεται στην οργάνωση ήχων με σκοπό την σύνθεση, εκτέλεση και ακρόαση/λήψη ενός μουσικού έργου. Με τον όρο εννοείται επίσης, και το σύνολο ήχων από το οποίο απαρτίζεται ένα μουσικό κομμάτι.

Διάθεση(mood): Η διάθεση είναι μια συναισθηματική κατάσταση μακράς διαρκείας. Οι διαθέσεις διαφέρουν από τα απλά συναισθήματα, γι' αυτό και είναι λιγότερο σαφές, λιγότερο έντονο και λιγότερο πιθανό να ενεργοποιηθούν από ένα συγκεκριμένο ερέθισμα ή γεγονός (Thayer 1989)[Panda, 2010].

Συναίσθημα(emotion): Συναίσθημα είναι ένα σύμπλεγμα από επιδράσεις ανάμεσα σε υποκειμενικούς και αντικειμενικούς παράγοντες, ενδιάμεσα από νευρικά, ορμονικά συστήματα, τα οποία μπορούν α) να προσφέρουν επιδραστικές εμπειρίες όπως τα συναισθήματα της διέγερσης, ευχαρίστησης, δυσανασχέτησης, β) να παράγουν γνωστικές διαδικασίες όπως αντιληπτικά συναφή αποτελέσματα, εκτιμήσεις, διαδικασίες προσθήκης ετικετών, γ) να ενεργοποιήσουν ευρέως διαδεδομένες φυσιολογικές προσαρμογές κατά τις συνθήκες διέγερσης και δ) να οδηγήσουν σε συμπεριφορά η οποία είναι συχνά, αλλά όχι πάντα εντυπωσιακή, στοχο-προσανατολιστική και προσαρμοστική (Meyers 2007)[Panda, 2010].

1.2 Πως επηρεάζει η μουσική το συναίσθημα;

1.2.1 Εισαγωγή

Η μουσική είναι ένα σημαντικό κομμάτι στη ζωή του ανθρώπου. Ιδιαίτερα την σημερινή εποχή που η διάδοση της μουσικής είναι ευκολότερη και οι χρήστες έχουν άμεση και πιο γρήγορη πρόσβαση σε αυτήν. Η παρούσα πτυχιακή εργασία εστιάζει στην μουσική ταξινόμηση και ανάκτηση με βάση το συναίσθημα. Αυτό αποτελεί μία πρόκληση: Πρώτον, γιατί η συναισθηματική αντίληψη είναι υποκειμενική και οι άνθρωποι μπορούν να αντιληφθούν διαφορετικά συναισθήματα για το ίδιο κομμάτι. Αυτό, λοιπόν το θέμα της υποκειμενικότητας είναι η βασική δυσκολία ενός τέτοιου προβλήματος αφού δεν υπάρχει συμφωνία στο αποτέλεσμα της ταξινόμησης. Δεύτερον, είναι δύσκολη η περιγραφή ενός συναισθήματος γιατί τα επίθετα που το

περιγράφουν είναι διφορούμενα και η χρήση των επιθέτων για το ίδιο συναίσθημα ποικίλει από άτομο σε άτομο. Τρίτον, είναι ακόμα ανεξήγητο πώς η μουσική προκαλεί συναισθήματα. Είναι δύσκολο να κατανοηθεί, ποιο είναι αυτό το στοιχείο της μουσικής που δημιουργεί ένα ιδιαίτερο συναίσθημα στον ακροατή.

Η σχέση ανάμεσα σε μουσική και συναίσθημα αποτελεί ένα θέμα που έχει απασχολήσει πολλούς ερευνητές από διάφορους τομείς όπως ανθρωπολογία, φιλοσοφία, ψυχολογία, μουσικολογία, κοινωνιολογία, βιολογία[Juslin & Sloboda, 2001]. Ένας μεγάλος αριθμός από μελέτες, έχει γίνει στην ψυχολογία. Σε αυτές τις μελέτες, το συναίσθημα κατηγοριοποιείται σε λίστες: εκφραζόμενο, λαμβανόμενο και προκαλούμενο[Juslin & Sloboda, 2001; Gabrielsson, 2002; Hallam, Cross & Thau, 2008; Huron, 2006]. Το πρώτο αναφέρεται στο συναίσθημα με το οποίο ο ερμηνευτής προσπαθεί να επικοινωνήσει με τους ακροατές, ενώ τα τελευταία δυο αναφέρονται στην επιδραστική ανταπόκριση των ακροατών. Τα δυο τελευταία, ιδιαίτερα το προκαλούμενο εξαρτάται από την αλληλεπίδραση παραγόντων όπως μουσικοί, προσωπικοί και καταστατικοί. Το λαμβανόμενο συναίσθημα επηρεάζεται λιγότερο από καταστατικούς παράγοντες.

Μια τυπική προσέγγιση ενός συστήματος αναγνώρισης συναισθήματος από μουσική κατηγοριοποιεί τα συναισθήματα σε ομάδες-κατηγορίες (όπως *χαρούμενος, θυμωμένος, λυπημένος και χαλαρωμένος*) και εφαρμόζει τεχνικές αναγνώρισης προτύπων για να εκπαιδεύσει έναν ταξινομητή [Hu, Downie, Laurier, Bay & Ehmann, 2008; Katayose, Imai & Inokuchi, 1998; Laurier, Grivolla & Herrera, 2008; Li & Ogihara, 2003; Liu, Yang, Wu & Chen, 2006; Livingstone & Brown, 2005; Lu, Liu & Zhang, 2006; Schuller, Hage, Schuller & Rigoll, 2010; Skowronek, McKinney & Van de Par, 2006; Trohidis, Tsoumakas, Kalliris & Vlahavas, 2008; Wang, Zhang & Zhu, 2004; Wu & Jeng, 2008; Yang, Liu & Chen, 2006]. Άλλες πάλι, λόγω της ασάφειας των χαρακτηρισμών που ορίζουν τις συναισθηματικές τάξεις με ένα διάγραμμα valence-arousal. Το valence ορίζει πόσο συναρπαστικό ή ήρεμο είναι ένα τραγούδι ενώ το arousal πόσο θετικό ή αρνητικό είναι το συναίσθημα που προκαλεί. Ένα αντιπροσωπευτικό παράδειγμα είναι το διάγραμμα VA του Thayer που χωρίζει τις τάξεις των συναισθημάτων σε τεταρτημόρια[Chang, Lo, Wang & Chung, 2010], όπως φαίνεται παρακάτω στην ενότητα 1.3.2 το σχήμα 5. Παρ' όλα αυτά υπάρχει μια ασάφεια στην ταξινόμηση των συναισθημάτων σε κατηγορίες. Για παράδειγμα στο πρώτο τεταρτημόριο, βρίσκονται τα συναισθήματα ***exciting***, ***happy*** και ***pleasure*** τα οποία είναι διαφορετικά εκ φύσεως. Η ασάφεια αυτή, μπερδεύει τόσο τους ακροατές στα υποκειμενικά τεστ όσο και τους χρήστες όταν ανακτούν ένα μουσικό κομμάτι σύμφωνα με την συναισθηματική τους κατάσταση.

Συνήθως τα χαρακτηριστικά τόνου, ρυθμού και αρμονίας της μουσικής εξάγονται για να αναπαραστήσουν την ακουστική ιδιότητα ενός μουσικού κομματιού. Παράλληλα, αρκετοί αλγόριθμοι ταξινόμησης έχουν εφαρμοστεί για να βρουν τη σχέση ανάμεσα σε μουσικά χαρακτηριστικά και ετικέτες συναισθήματος όπως είναι τα support vector machines [Hu, Downie, Laurier, Bay & Ehmann, 2008; Li & Ogihara, 2003; Wang, Zhang & Zhu, 2004; Bischoff, Firan, Paiu, Nejdl, Laurier & Sordo, 2009], τα Gaussian mixture models[Fernandes & Paiva ,2010; Liu, Zhang & Lu, 2003], τα neural networks[Feng, Zhuang & Pan 2003] και τα k-nearest neighbor[Yang, Liu & Chen, 2006; Wieczorkowska, 2004].

1.2.2 Προβλήματα στην αναγνώριση συναισθημάτων

Προκύπτουν, τέσσερα πολύ σημαντικά θέματα που πρέπει να λυθούν[Yang & Chen, 2011]. Η ασάφεια και διακριτότητα της περιγραφής του συναισθήματος, το βαρύ γνωστικό φορτίο του σχολιασμού συναισθήματος, η υποκειμενικότητα της συναισθηματικής αντίληψης και τέλος το σημασιολογικό χάσμα ανάμεσα στα χαμηλού επιπέδου ηχητικά σήματα και στην υψηλού επιπέδου ανθρώπινη αντίληψη.

1.2.2.1 Ασάφεια ή σύγχυση και διακριτότητα

Η ασάφεια ή σύγχυση είναι ένα χαρακτηριστικό της φυσικής γλώσσας των κατηγοριών[Hofmann, 1999]. Τα συναισθήματα είναι πολύ συγκεχυμένες έννοιες. Για να αποφευχθεί αυτή η ασάφεια των επιδραστικών όρων για να μειωθεί η προσπάθεια του αναπτυγμένου συστήματος, πολλοί ερευνητές χρησιμοποίησαν τα βασικά συναισθήματα όπως ***happy, sad, angry*** και ***relaxing*** ή ομάδες συναισθημάτων, ταξινομώντας τα σε κατηγορίες[Lu, Liu & Zhang, 2006; Wang, Zhang & Zhu, 2004; Feng, Zhuang & Pan 2003; Hu & Downie, 2007; Yang & Lee, 2004]. Η διακριτότητα αναφέρεται στον αριθμό των τάξεων των συναισθημάτων. Δηλαδή η ταξινόμηση του συναισθήματος είναι πολύ μικρότερη σε σύγκριση με την πληθώρα συναισθημάτων που αισθάνεται ο ακροατής. Αυτό είναι ανεπιθύμητο επειδή ένα σύστημα σύστασης με περιορισμένο λεξιλόγιο συναισθημάτων, ίσως δεν ικανοποιήσει τις απαιτήσεις του χρήστη στον πραγματικό κόσμο των εφαρμογών μουσικής ανάκτησης. Παρ' όλα αυτά, χρησιμοποιώντας μια λεπτομερής περιγραφή των συναισθημάτων, δεν είναι απαραίτητο να διευθετηθεί η διακριτότητα, επειδή μειώνεται η ασάφεια μεταξύ των συναισθηματικών όρων[Bartoszewski,

Kwasnicka, Markowska-Kaczmar & Myszkowski, 2008] και επειδή η ανάπτυξη ενός αυτόματου συστήματος που ταξινομεί με μικρό σφάλμα την μουσική σε ένα μεγάλο αριθμό τάξεων είναι πολύ δύσκολο.

1.2.2.2 Απαιτούμενες γνώσεις

Για να συλλέξουμε το ground truth (ακρίβεια του τρόπου εκπαίδευσης του συστήματος) εκπαιδεύοντας ένα αυτόματο μοντέλο, μια υποκειμενική δοκιμή τυπικά γίνεται προσκαλώντας εθελοντές να χαρακτηρίσουν συναισθηματικά ένα μουσικό κομμάτι. Για να μειωθεί ο χρόνος εξαγωγής των αποτελεσμάτων, κάθε μουσικό κομμάτι χαρακτηρίζεται το λιγότερο από τρεις εθελοντές[Li & Ogihara, 2003; Lu, Liu & Zhang, 2006; Skowronek, McKinney & Van de Par, 2006; Trohidis, Tsoumakas, Kalliris & Vlahavas, 2008]. Αυτή η πρακτική είναι προβληματική γιατί τα καθημερινά σχόλια αυτών που έχουν γνώσεις πάνω στο αντικείμενο είναι πολύ διαφορετικά από αυτά των εθελοντών που δεν έχουν καμία γνώση και απαιτείται ξεχωριστή μεταχείριση[Sloboda, O'Neill & Ivaldi, 2001]. Το γνωστικό φορτίο συλλογής ετικετών για τα συναισθήματα ενός κατηγορηματικού συστήματος έχει λυθεί πρόσφατα με την αύξηση ιστοσελίδων που καταχωρούν οι χρήστες ετικέτες στα συναισθήματα όπως το All Music Guide[http://www.allmusic.com/] και ο Last.fm[http://www.last.fm/]. Αντίθετα στο διαστατικό σύστημα απαιτούνται αξιολογήσεις συναισθήματος στο επίπεδο που ορίζεται από τους δύο άξονες valence και arousal. Αυτές οι αξιολογήσεις δεν αποκτούνται διαδικτυακά αλλά βρίσκονται χρησιμοποιώντας είτε σταθερή κλίμακα αξιολόγησης είτε γραφική κλίμακα αξιολόγησης, δίνοντας στους εθελοντές τις απαραίτητες γνώσεις. Είναι δύσκολο να βεβαιώσουμε μια σταθερή κλίμακα αξιολόγησης ανάμεσα σε διαφορετικούς εθελοντές αλλά και στον ίδιο εθελοντή[Ovadia, 2004]. Ως αποτέλεσμα, η ποιότητα των τιμών του ground truth, μπορεί να υποβαθμίσει την ακρίβεια του συστήματος.

1.2.2.3 Υποκειμενικότητα συναισθηματικής αντίληψης

Η μουσική αντίληψη είναι υποκειμενική και εξαρτάται από πολλούς παράγοντες όπως η ηλικία, το γένος, η προσωπικότητα και άλλα[Abeles & Chung, 1996]. Οι επιδράσεις ανάμεσα στην μουσική και τον ακροατή εξαρτώνται από τις προτιμήσεις του πάνω στη μουσική[Holbrook & Schindler, 1989; Jargreaves & North, 1997] και από την οικειότητά του με αυτήν[Jargreaves North, 1997]. Εξαιτίας της υποκειμενικότητας είναι δύσκολο να υπάρξει ομοφωνία σχετικά με το ποιος συναισθηματικός όρος χαρακτηρίζει

καλύτερα ένα μουσικό κομμάτι. Στην κατηγορηματική προσέγγιση που κάθε μουσικό κομμάτι χαρακτηρίζεται από μια τάξη συναισθήματος, το θέμα της υποκειμενικότητας δεν αντιμετωπίζεται σωστά. Αντίθετα, στην διαστατική προσέγγιση το πρόβλημα λύνεται επειδή κάθε χρήστης απαντά διαφορετικά στο ίδιο κομμάτι. Παρά το γεγονός ότι η υποκειμενική φύση της συναισθηματικής αντίληψης είναι αναγνωρισμένη, δεν έχουν γίνει αρκετές προσπάθειες για την επίλυση του θέματος.

1.2.2.4 Σημασιολογικό χάσμα ανάμεσα στα χαμηλού επιπέδου ηχητικά σήματα και στην υψηλού επιπέδου ανθρώπινη αντίληψη

Εξαιτίας αυτού του χάσματος είναι δύσκολο να υπολογιστούν με ακρίβεια οι τιμές του συναισθήματος, ιδιαίτερα του valence[Lu, Liu & Zhang, 2006; Schubert, 1999; Tolos, Tato & Kemp, 2005; Yang, Lin, Su & Chen 2008]. Το στοιχείο αυτό της μουσικής που προκαλεί ένα συναίσθημα στον ακροατή, δεν έχει κατανοηθεί ακόμα καλά. Κατά συνέπεια, η απόδοση των συμβατικών μεθόδων που αξιοποιούν μόνο τα χαρακτηριστικά ήχου χαμηλού επιπέδου φαίνεται να έχουν φτάσει σε ένα όριο. Στην Audio Mood Classification(AMC), έχει πραγματοποιηθεί από το 2007 το MIREX (music information retrieval evaluation exchange), με σκοπό την προώθηση της έρευνας του συστήματος Music Emotion Recognition και την παροχή συγκρίσεων αναφοράς[Panda, Malheiro, Rocha, Oliveira & Paiva, 2008; Panda & Paiva DAF 2012; Panda & Paiva MML 2012; Panda & Paiva CISUC 2011; Hu, Downie, Laurier, Bay & Ehmann, 2008]. Χρησιμοποιούνται πέντε ομάδες συναισθήματος: *passionate, rollicking, literate, humorous και aggressive*[Hu & Downie, 2007]. Τα ποσοστά ακρίβειας της ταξινόμησης από το 2007 έως το 2010 δεν ξεπέρασαν ποτέ το 70%.

1.3 Προσεγγίσεις

Η σχέση ανάμεσα σε μουσική και συναίσθημα έχει μελετηθεί από ψυχολόγους για δεκαετίες. Τα θέματα που αντιμετώπισαν στην έρευνα τους αφορούν α) το αν τα καθημερινά συναισθήματα είναι τα ίδια με εκείνα που λαμβάνονται από την μουσική, β) με εκείνα τα συναισθήματα που λαμβάνονται από τον ακροατή ή με εκείνα που αισθάνεται ο ακροατής, γ) κατά πόσο οι παράγοντες μουσικοί, προσωπικοί και καταστατικοί επηρεάζουν την συναισθηματική αντίληψη και τέλος, δ) πως πρέπει να αντιληφθούμε το

μουσικό συναίσθημα[Juslin & Sloboda, 2001]. Έτσι κατέληξαν σε δυο προσεγγίσεις για την μουσική αντίληψη: την κατηγορηματική και την διαστατική.

1.3.1 Κατηγορηματική προσέγγιση

Κατηγορηματική προσέγγιση. Εδώ ο ακροατής αντιλαμβάνεται τα συναισθήματα ως κατηγορίες διαφορετικές μεταξύ τους. Σημαντικό σε αυτήν την προσέγγιση είναι η έννοια των βασικών συναισθημάτων, δηλαδή η ιδέα του να υπάρχουν έμφυτες και παγκόσμιες κατηγορίες συναισθήματος όπως *anger, fear, sad, happiness*, από τις οποίες έχουν δημιουργηθεί όλες οι άλλες τάξεις συναισθήματος[Schuller, Hage, Schuller & Rigoll, 2010; Anderson & McOwan, 2006; Ekman, 1992; Jonghwa & Ande, 2008; Keltner & Ekman, 2000; Laurier, Meyers, Serra, Blech, Herrera & Serra, 2009; Lee & Narayanan, 2005; Picard, Vyzas & Healey, 2001]. Ο ερευνητής Paul Eckman[Ekman, 1992; Ekman, 1999], για παράδειγμα χρησιμοποίησε την εξής κατηγοριοποίηση, *anger, fear, sadness, happiness*, και *disgust*. Μια τυπική κατηγορηματική προσέγγιση μπορεί να αποτελείται από επτά διακριτές τάξεις. Η αντίληψη των βασικών συναισθημάτων διαφοροποιείται. Διαφορετικοί ερευνητές βρήκαν διαφορετικά σύνολα βασικών συναισθημάτων[Sloboda & Juslin, 2001]. Ένα μεγάλο μειονέκτημά της προσέγγισης αυτής είναι ότι ο άνθρωπος αντιλαμβάνεται πολύ περισσότερα συναισθήματα από τον αριθμό των κύριων τάξεων συναισθημάτων που ορίζει αυτή. Επιπλέον η χρήση μεγάλου αριθμού τάξεων συναισθημάτων μπερδεύει τους ακροατές και είναι μη πρακτική για ψυχολογικές μελέτες[Sloboda & Juslin, 2001].

Ο πρώτος ερευνητής που βασίστηκε στην κατηγορηματική προσέγγιση και δημιούργησε ένα κυκλικό μοντέλο ήταν η Kate Hevner (1936), χωρίζοντας τα συναισθήματα σε 8 κατηγορίες: *dignified, sad, dreamy, serene, graceful, happy, exciting, vigorous*[Fernandes & Paiva, 2010; Panda, 2010; Hevner, 1935; Hevner, 1936]. Το ίδιο μοντέλο χρησιμοποιήθηκε και από τον Owen Craigie Meyers το 2007.

Κυκλικό διάγραμμα με 8 κατηγορίες συναισθήματος(Hevner 1936)

Η Emery Schubert το 2003 χρησιμοποίησε επίσης, την κατηγορηματική προσέγγιση και δημιούργησε ένα νέο μοντέλο με εννιά κατηγορίες[Schubert, 2003; Schubert, 1999].

Cluster	Emotions in each cluster
1	Bright, cheerful, happy, joyous
2	Humorous, light, lyrical, merry, playful
3	Calm, delicate, graceful, quiet, relaxed, serene, soothing, tender
4	Dreamy, sentimental
5	Dark, depressing, gloomy, melancholy, mournful, sad, solemn
6	Heavy, majestic, sacred, serious, spiritual, vigorous
7	Tragic, yearning
8	Agitated, angry, restless, tense
9	Dramatic, exciting, exhilarated, passionate, sensational, soaring, triumphant

9 κατηγορίες συναισθήματος (Schubert 2003)

Η ταξινόμηση συναισθημάτων που ακολουθεί είναι από το MIREX (’07-’10) (music information retrieval evaluation exchange)[Panda, Malheiro, Rocha, Oliveira & Paiva, 2008; Panda & Paiva DAF 2012; Panda & Paiva MML 2012; Panda & Paiva, CISUC 2011; Panda & Paiva, AES 2011; Cardoso, Panda & Paiva, INForum 2011; Panda & Paiva, CISUC 2012; Hu, Downie, Laurier, Bay & Ehmann, 2008; Panda, Rocha & Paiva, CISUC].

Cluster	Description
1	passionate, rousing, confident, boisterous, rowdy
2	rollicking, cheerful, fun, sweet, amiable/good natured
3	literate, poignant, wistful, bittersweet, autumnal, brooding
4	humorous, silly, campy, quirky, whimsical, witty, wry
5	aggressive, fiery, tense/anxious, intense, volatile, visceral

*Έρευνα **MIREX** με 5 κατηγορίες συναισθήματος**(‘07-‘10)***

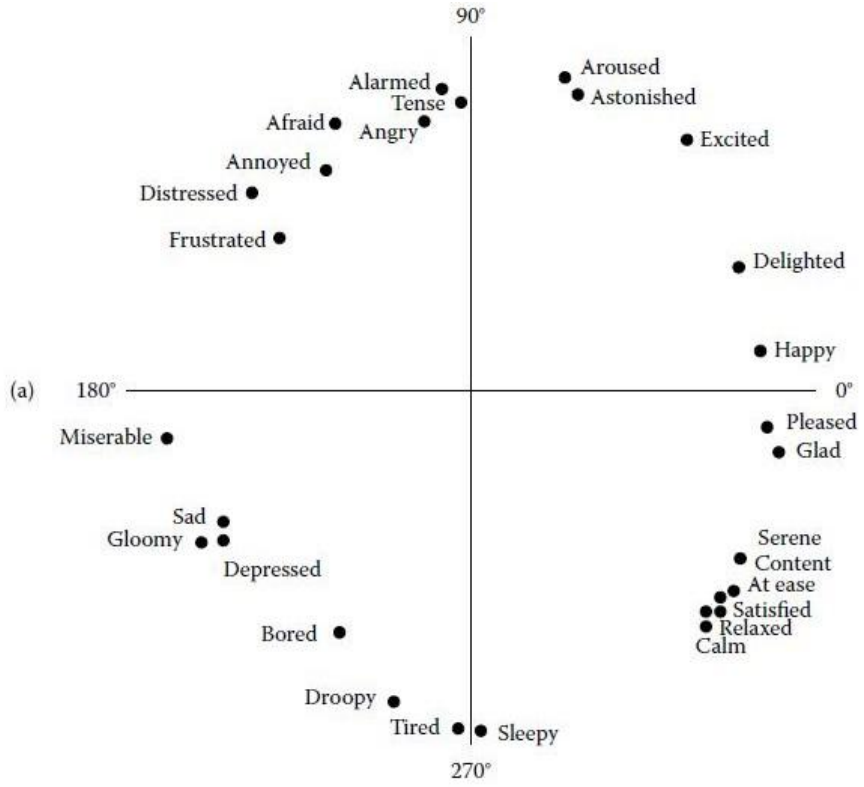
1.3.2 Διαστατική προσέγγιση

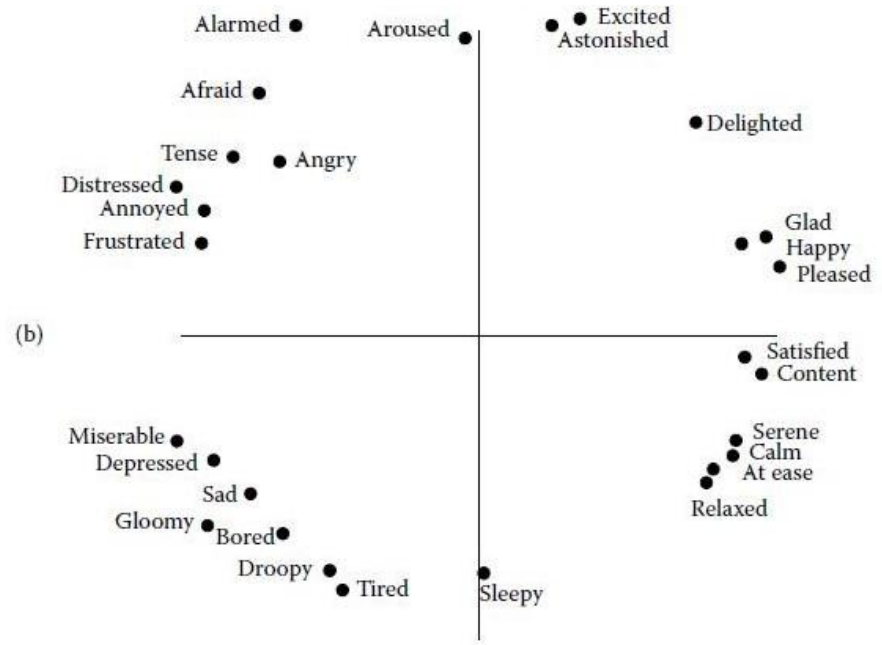
Η προσέγγιση αυτή εστιάζει σε ταυτοποιημένα συναισθήματα, βασισμένα στις θέσεις τους σε ένα επίπεδο συναισθημάτων με άξονες, οι οποίοι αντιπροσωπεύουν τις εσωτερικές ανθρώπινες εκφράσεις του συναισθήματος. Αυτές οι εσωτερικές διαστάσεις συναισθήματος βρίσκονται αναλύοντας τον συσχετισμό ανάμεσα σε συναισθηματικούς όρους. Αυτό πραγματοποιείται ζητώντας από ανθρώπους να σχολιάσουν συναισθηματικά ένα μουσικό κομμάτι με την χρήση μιας κλίμακας αξιολόγησης από ποικίλα συναισθήματα. Έπειτα με τεχνικές ανάλυσης παραγόντων, συλλέγονται τα αποτελέσματα και αποκτάται ο συσχετισμός ανάμεσα στους συναισθηματικούς όρους. Στους δυο άξονες χρησιμοποιούνται τα ***valence*** και ***arousal.*** Στους τρεις άξονες χρησιμοποιούνται τα εξής: ***valence, arousal, dominance***. Το valence ορίζει πόσο συναρπαστικό ή ήρεμο είναι ένα τραγούδι, το arousal πόσο θετικό ή αρνητικό είναι το συναίσθημα που προκαλεί και dominance που δείχνει την δραστικότητα ή έλεγχο του συναισθήματος[Scherer, 2004]. Αυτή η προσέγγιση ωστόσο έχει επικριθεί. Υπάρχει μια διαφωνία από ερευνητές στο ότι αυτή η προσέγγιση μπερδεύει σημαντικές ψυχολογικές διαφορές και κατ’ επέκταση καλύπτει σημαντικές πλευρές στη συναισθηματική διαδικασία[Ali, 2006]. Για παράδειγμα, τα συναισθήματα angry και fear που ανήκουν στο δεύτερο

τεταρτημόριο του διαγράμματος valence-arousal όπως δείχνει το σχήμα 4 στην ενότητα 1.3.2 έχουν πολύ διαφορετικές επιπτώσεις στον οργανισμό. Το ίδιο ισχύει και για την boredom και melancholy(Σχήμα 4). Χρησιμοποιώντας το δισδιάστατο μοντέλο δεν διακρίνονται τα μουσικά κομμάτια που παράγουν ένα συναίσθημα από ένα άλλο και δεν επιτρέπουν την θεωρητική εξέταση της προέλευσης και των μηχανισμών τέτοιων επιδραστικών επιπτώσεων[Sloboda & Juslin, 2001]. Επιπλέον, υπάρχει και η διαφωνία ότι χρησιμοποιώντας λίγες διαστάσεις συναισθήματος δεν μπορούν να περιγραφούν όλα τα συναισθήματα χωρίς σημαντικά σφάλματα[Collier, 2007].

Ο James Russell (1980) χρησιμοποίησε πρώτος την διαστατική προσέγγιση σε κυκλική μορφή[Russell, 1980; Fernandes & Paiva, 2010; Panda, 2010;].

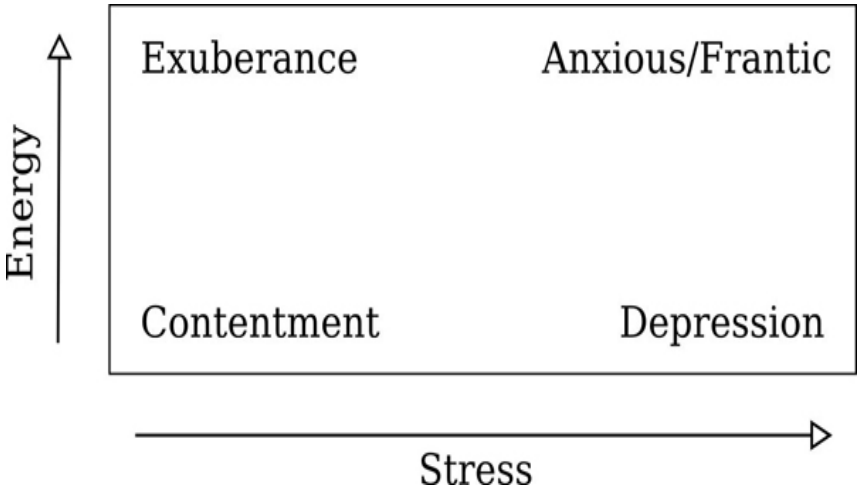
Σχήμα 4





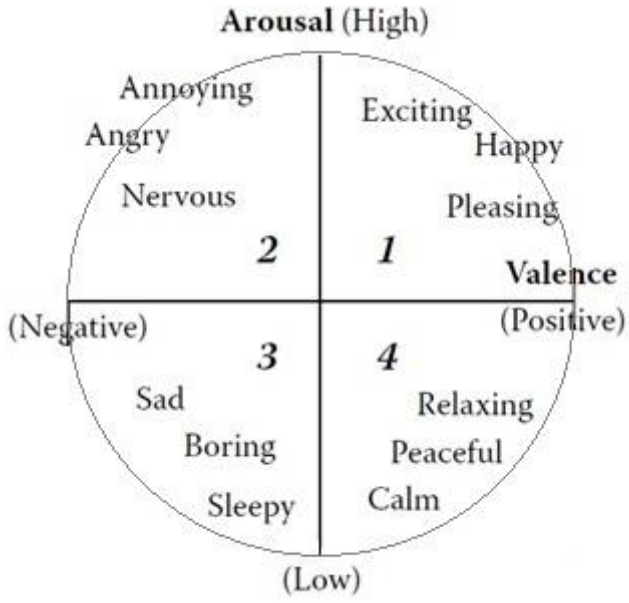
Κυκλικό μοντέλο με 28 συναισθήματα(Russell 1980)

Ο ερευνητής Robert Thayer (1989) αργότερα χρησιμοποίησε την ίδια προσέγγιση, δημιουργώντας ένα διάγραμμα συντεταγμένων με μεταβλητές valence και arousal[Thayer, 1989; Fernandes & Paiva, 2010; Panda, 2010;].



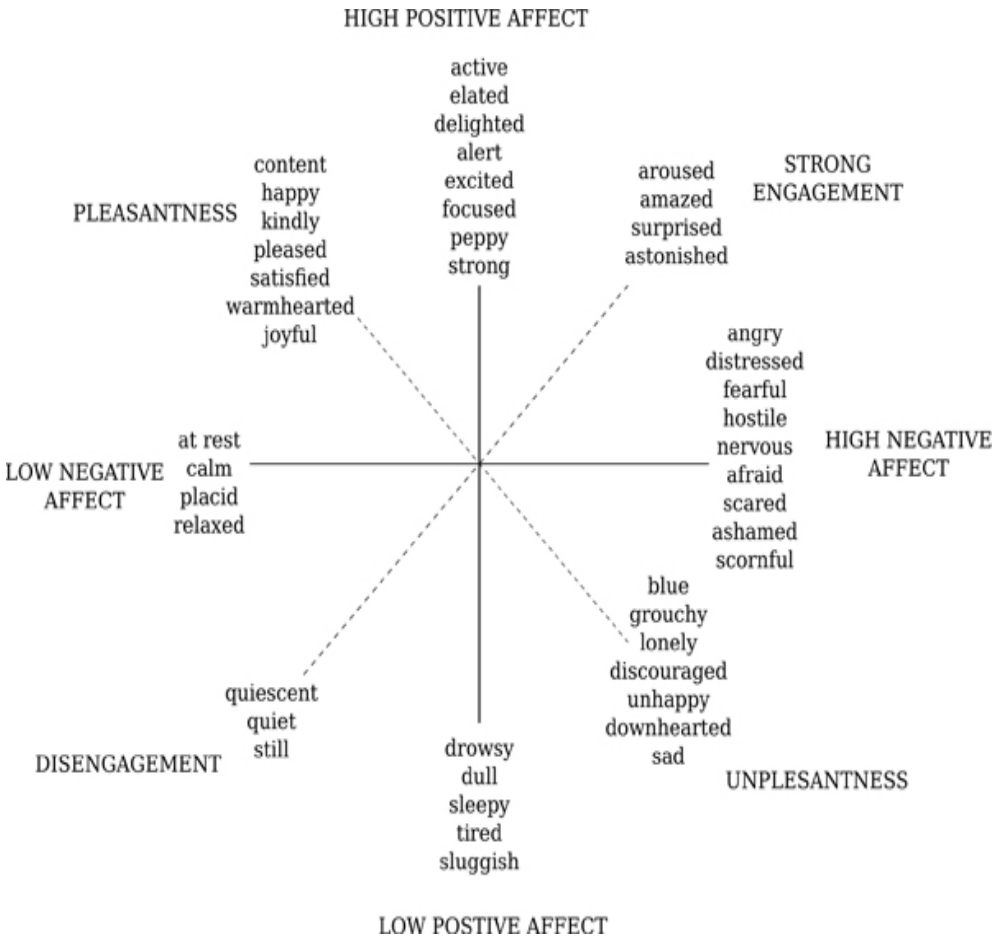
Μοντέλο Thayer (1989)

Ο Yang το 2007, 2008, 2009 χρησιμοποίησε τη διαστατική προσέγγιση βασισμένος στο μοντέλο του Thayer [Thayer, 1989; Yang, Liu & Chen, 2006; Yang & Chen, IEEE; Yang, Lin & Chen, 2009; Yang, Su, Lin & Chen, 2007].



*Μοντέλο **Yang** βασισμένο στο μοντέλο του **Thayer(2008)***

Οι ερευνητές Tellegen-Watson-Clarc (1999), χρησιμοποίησαν την διαστατική προσέγγιση με έναν διαφορετικό τρόπο, δημιουργώντας ένα διάγραμμα με μεταβλητές ***high** και **low positive affect** και **high** και **low negative affect***[Tellegen, Watson & Clark, 1999]. Στο διάγραμμα αυτό βασίστηκε και ο Laar το 2005[Tellegen, Watson & Clark, 1999; Laar, 2006; Fernandes & Paiva, 2010; Panda, 2010;].



Κυκλικό μοντέλο με 8 κατηγορίες συναισθήματος(Tellegen-Watson-Clarc 1999)

Η παρούσα πτυχιακή εργασία, σκοπό έχει τη δημιουργία ενός συστήματος που θα αναγνωρίζει αυτόματα το συναίσθημα σε τραγούδια. Θα χρησιμοποιηθούν 5 κατηγορίες συναισθημάτων όπως του MIREX (music information retrieval evaluation exchange) [Panda, Malheiro, Rocha, Oliveira & Paiva, 2008; Panda & Paiva DAF 2012; Panda & Paiva MML 2012; Panda & Paiva, CISUC 2011; Panda & Paiva, AES 2011; Cardoso, Panda & Paiva, INForum 2011; Panda & Paiva, CISUC 2012; Hu, Downie, Laurier, Bay & Ehmann, 2008; Panda, Rocha & Paiva, CISUC], το λογισμικό Marsyas[Tzanetakis & Cook, 2002] για την εξαγωγή παραμέτρων που θα περιγράφουν το συναισθηματικό περιεχόμενο των τραγουδιών και το πακέτο Weka[Panda, Malheiro, Rocha, Oliveira & Paiva, 2008;] για την αναγνώριση και εξαγωγή των αποτελεσμάτων.

Η δομή της εργασίας οργανώνεται ως εξής: Στο κεφάλαιο 2, περιγράφονται οι σημαντικότερες μελέτες που έχουν γίνει πάνω στο αντικείμενο. Στο κεφάλαιο 3, περιγράφονται τα χαρακτηριστικά της μουσικής. Στο κεφάλαιο 4, περιγράφεται η μεθοδολογία της δικής μας έρευνας. Στο

κεφάλαιο 5 περιγράφονται τα πειράματα που διεξήγαμε. Στο κεφάλαιο 6, γίνεται η αξιολόγηση των πειραμάτων και στο κεφάλαιο 7 δίνονται τα συμπεράσματα και η μελλοντική δουλειά.

Κεφάλαιο 2 - Ανασκόπηση Βιβλιογραφίας

2.1 Σημαντικότερες μελέτες

2.1.1 Εισαγωγή

Τα τελευταία χρόνια έχουν διεξαχθεί πολλές μελέτες με αντικείμενο την αναγνώριση συναισθημάτων από τραγούδια. Οι ερευνητές χρησιμοποίησαν διάφορα μοντέλα και συστήματα είτε με την χρήση διαδικτύου είτε χωρίς που εξυπηρετούσαν αυτόν τον σκοπό. Ο ενδιαφερόμενος μπορεί να αναζητήσει πληροφορίες στην βιβλιογραφία[1-213] για μελέτες πάνω στο αντικείμενο. Οι σημαντικότερες μελέτες παρουσιάζονται στην επόμενη ενότητα.

2.1.2 Έρευνες ανά χρονολογία

2001: Οι Conor Hayes και Padraig Cunningham ανέπτυξαν μια εφαρμογή διαδικτυακής μουσικής στο Trinity College του Δουβλίνου[Hayes & Cunningham, 2001]. Το smart radio όπως ονομάστηκε δίνει την δυνατότητα στους χρήστες να δημιουργήσουν, να διευθύνουν και να μοιραστούν μουσικά προγράμματα. Αποτελεί ένα πρωτότυπο σύστημα για ένα υψηλό εύρος ζώνης σε πάντα online σύνδεση στο διαδίκτυο.

Αρχικά, το Smart Radio Project αφομοιώνει δυο παραδείγματα: την ρύθμιση διανομής μουσικής στο διαδίκτυο και την υπηρεσία προγραμματισμού ενός προσωπικού ραδιοφώνου. Το Smart Radio χρησιμοποιεί *streaming audio technology* που αναπτύχθηκε από το Real Networks. Η αρχιτεκτονική του Smart Radio δεν χρησιμοποιεί αυτήν την τεχνολογία, όμως στην προκειμένη περίπτωση η ανάπτυξη του Real Networks παρέχει το πιο δημοφιλές Streaming Client Player.

Η επιλογή του Streaming technology είναι κατάλληλη για πολλούς λόγους. Πρώτον, ο χρήστης συνδέεται, επιλέγει το προτεινόμενο πρόγραμμα ή ένα αγαπημένο πρόγραμμα και αυτόματα και αμέσως, αυτό εμφανίζεται στην επιφάνεια εργασίας του υπολογιστή του. Δεύτερον, είναι η λύση στην πειρατεία. Τα Streaming Audio Files δεν αποθηκεύονται στο σκληρό δίσκο του υπολογιστή του χρήστη ούτε κοινοποιούνται στο διαδίκτυο για παράνομη διανομή. Τρίτον, αυτή η υπηρεσία δεν δίνει τη δυνατότητα στο χρήστη να κατεβάσει μουσικά αρχεία αλλά του είναι διαθέσιμα μόνο όταν είναι συνδεδεμένος στο διαδίκτυο.

Στη συνέχεια, οι ερευνητές παρουσιάζουν τεχνικές για να παρέχουν ένα προσωπικό ραδιόφωνο στους χρήστες. Υποστηρίζουν ότι ο χρήστης πρέπει να

έχει την δυνατότητα να δημιουργεί το δικό του μουσικό πρόγραμμα. Στην παρούσα έρευνα η ανταπόκριση του χρήστη γίνεται σε δυο επίπεδα: επίπεδο κομματιού και επίπεδο προγράμματος. Περαιτέρω σε αυτό, συζητούν πως το σύστημα σύστασης Smart radio ενθαρρύνει τη συμμετοχή της κοινότητας, επιτρέποντας στους χρήστες να γνωρίζουν ποιοι είναι οι πιο συνεπής τους 'φίλοι'[Hill, Rosenstein & Furnas,1995].

Έπειτα, παρουσιάζεται η αρχιτεκτονική του Smart radio. Το Smart Radio αποτελεί μια διαδικτυακή εφαρμογή. Ο χρήστης συνδέεται στο διαδίκτυο, επιλέγει να δημιουργήσει νέα προγράμματα, να ακούσει παλαιότερα, ή να επιλέξει προγράμματα που του προτείνουν άλλοι χρήστες. Ο προγραμματισμός του server έχει γίνει με Java servlets. Κάθε χρήστης που συνδέεται βλέπει έναν εικονικό Smart Radio Server.

Τέλος, η λήψη ενός προγράμματος μουσικής γίνεται σε δυο σκηνές: δημιουργία νέου προγράμματος και προτεινόμενο πρόγραμμα ή παλαιότερο αγαπημένο πρόγραμμα. Όταν ο χρήστης δείξει ότι το πρόγραμμα είναι έτοιμο να παίξει ο browser περνάει την session id στο plug-in, το οποίο ζητάει την αντίστοιχη session id του προγράμματος. Το τελευταίο κομμάτι της έρευνας κλείνει με προτάσεις για συστήματα επί πληρωμή και επισημαίνει την σημαντικότητα του συστήματος του Smart Radio σε σύνδεση με τεχνολογίες υψηλού εύρους ζώνης όπως το ADSL.

2006: Οι Peter Knees, Tim Pohle, Markus Schedl και Gerhard Widmer, παρουσίασαν μια τεχνική δημιουργίας αυτόματης playlist που συνδυάζει την ομοιότητα μουσικής βασισμένη σε ηχητικά σήματα με αυτήν του μουσικού καλλιτέχνη από το διαδίκτυο[Widmer, Pohle, Schedl & Knees, 2006]. Παρουσίασαν έναν νέο και αποτελεσματικό τρόπο για να συνδυάζονται οι πηγές πληροφορίας για την παραγωγή μιας αποτελεσματικής μουσικής playlist.

Οι προηγούμενες προσεγγίσεις χρησιμοποιούσαν μόνο τα ακουστικά χαρακτηριστικά. Εδώ, ενσωματώνονται και τα χαρακτηριστικά από το διαδίκτυο για να μειωθεί ο υπολογισμός της ακουστικής ομοιότητας μιας και αυτό βελτιώνει την ποιότητα των παραγόμενων playlists. Δημιούργησαν έτσι, μια επαφή για χρήστες που ακούν μουσική μέσω κινητού, τον οποίο ονόμασαν 'wheel'.

Αρχικά, στην παρούσα έρευνα εξετάζονται τρεις προσεγγίσεις που συνδυάζουν ακουστικά και διαδικτυακά δεδομένα. Έπειτα, δίνεται μια επισκόπηση των συστημάτων παραγωγής αυτόματης playlist μεταξύ αυτών και του δικού τους *'wheel'*, ο οποίος θα χρησιμοποιηθεί για να δείξει την αποτελεσματικότητα αυτής της τεχνικής.

Πρώτη προσέγγιση. Η ακουστική και η διαδικτυακή ταξινόμηση του είδους μουσικής χρησιμοποιείται για να ανιχνεύσει το είδος ενός συνόλου από

πέντε είδη μουσικής, πέντε καλλιτεχνών το καθένα. Συνδυάζοντας τις προβλέψεις που έγιναν και από τις δυο μεθόδους, ακουστικά και διαδικτυακά δεδομένα, γραμμικά αποδίδει καλύτερη συνολική πρόβλεψη για όλα τα τεστ[Whitman & Smaragdis, 2002].

Δεύτερη προσέγγιση. Η ακουστική ομοιότητα ενός τραγουδιού είναι συνδυασμένη με την ομοιότητα του καλλιτέχνη από το διαδίκτυο για να αποκτήσει μια καινούρια μέτρηση ομοιότητας[Baumann, 2005].

Τρίτη προσέγγιση. Η αύξηση μιας διεπαφής σε μουσικές συλλογές με χαρακτηριστικά που πήραν από το διαδίκτυο. Η διεπαφή αποτελείται από ένα τοπίο τρισδιάστατου νησιού που τοποθετεί τα μουσικά κομμάτια σύμφωνα με την δική τους ομοιότητα στον ήχο. Ο χρήστης πλοηγείται ελεύθερα σε ένα εικονικό περιβάλλον. Η εξερεύνηση υποστηρίζεται από τους όρους στο τοπίο που είναι σχετικές με το ακουστικό περιεχόμενο σε αυτήν την περιοχή και τους αντίστοιχους καλλιτέχνες. Έτσι, παρέχει μια σημασιολογική αντιστοίχιση βασισμένη στη μουσική[Knees, Pohle, Schedl & Widmer, ACM Multimedia, 2006].

Στη συνέχεια, επεκτείνεται η διεπαφή ‘wheel’. Αν και η αυθεντική προσέγγιση περιλαμβάνει τον υπολογισμό ακουστικών ομοιοτήτων ανάμεσα σε κάθε κομμάτι τραγουδιού σε μια συλλογή, ενσωματώνονται τα δεδομένα από το διαδίκτυο για να μειωθούν οι απαραίτητοι υπολογισμοί ομοιότητας. Συγκεντρώνουν πληροφορία από το διαδίκτυο για τους καλλιτέχνες και την χρησιμοποιούν για να αξιολογήσουν τις ομοιότητες μεταξύ τους. Με αυτόν τον τρόπο πετυχαίνεται η βελτίωση της αρχικής προσέγγισης που βασίζεται μόνο στον ήχο.

Μεθοδολογία. Στόχος της παραγωγής playlist είναι να μεγιστοποιήσουν το μέσο όρο ομοιότητας ανάμεσα σε συνεχόμενα τραγούδια. Στην συγκεκριμένη έρευνα η playlist που δημιουργούν είναι μια προβολή μιας ολόκληρης μουσικής συλλογής σε μια διάσταση. Αυτή η συλλογή διατάσσεται πάνω σε ένα κυκλικό τροχό και κάθε διαφορετικό μουσικό είδος τοποθετείται γύρω του. Κάθε φορά που θέλει ο χρήστης να επιλέξει κάποιο είδος γυρνάει τον τροχό.

Αξιολόγηση. Η αξιολόγηση αυτής της προσέγγισης γίνεται σε δυο φάσεις. Την ποσοτική αξιολόγηση και την υποκειμενική αξιολόγηση. Στην ποσοτική αξιολόγηση διεξάγονται κάποια πειράματα χρησιμοποιώντας δυο συλλογές.

Η πρώτη συλλογή αποτελείται από 3.456 τραγούδια 339 καλλιτεχνών από 7 ομοιόμορφα είδη μουσικής: classical 14,7%, dance 15,0%, hip-hop 14,5%, jazz 13,6%, metal 14,9%, pop 11,6% και punk 15,6%. Ο μικρότερος αριθμός τραγουδιών ανά καλλιτέχνη είναι 1 και ο μεγαλύτερος 317.

Η δεύτερη συλλογή αποτελείται από κάποια πολύ ιδιαίτερα είδη μουσικής και αυτό για να δείξουν την εφαρμοστικότητα της προσέγγισης σε μη σταθερή μουσική. Επιλέχθηκαν 2.545 τραγούδια από 103 καλλιτέχνες από 13 είδη μουσικής: a cappella 4,4%, acid jazz 2,7%, blues 2,5%, bossa nova 2,8%, celtic 5,2%, electronica 21,1%, folk rock 9,4%, Italian 5,6%, jazz 5,3%, metal 16,1%, punk rock 10,2%, rap 12,9% και reggae 1,8% . Ο μικρότερος αριθμός τραγουδιών ανά καλλιτέχνη είναι 8 και ο μεγαλύτερος 61. Η αξιολόγηση γίνεται συγκρίνοντας την δική τους προσέγγιση με την αρχική. Εστιάζουν στον περιορισμό των απαραίτητων υπολογισμών της ομοιότητας και προσπαθούν να βρουν τρόπο να αξιολογήσουν την ποιότητα της playlist. Ο χρήστης υπολογίζουν ότι ακούει περίπου το ένα όγδοο των τραγουδιών, 432 για την πρώτη συλλογή και 318 για την δεύτερη.

Στην υποκειμενική αξιολόγηση αποδεικνύουν την πρακτική χρησιμότητα της προσέγγισης. Διεξάγουν έτσι, μια μελέτη αποτελούμενη από 10 άτομα. Χρησιμοποιούν την πρώτη συλλογή δημιουργώντας μια playlist και εξάγουν 10 ακολουθίες με 10 κομμάτια την καθεμία. Τέλος, τυχαία ορίζουν μια αφετηρία στον τροχό και διαδοχικά εξάγουν τις playlists των 10 τραγουδιών. Κάθε άτομο βαθμολογεί τις playlists στην κλίμακα από 1 έως 5. Από την μελέτη αυτή είδαν ότι 7 στις 10 playlists έχουν βαθμολογηθεί με πάνω από 3, 33 τραγούδια βαθμολογήθηκαν με 5 και 29 τραγούδια βαθμολογήθηκαν με 4. Αυτό δείχνει ότι οι playlists που δημιουργήθηκαν είναι χρήσιμες και συνοχικές για τον χρήστη.

2008: Οι Kunsu Kim, Donghoon Lee, Tae-Bok Yoon και Jee-Hyong Lee πρότειναν ένα σύστημα μουσικής σύστασης βασισμένο στις προτιμήσεις των χρηστών[Kim, Lee, Yoon & Lee, 2008]. Το παρόν σύστημα δημιουργεί μουσικά μοντέλα χρησιμοποιώντας τα Hidden Markov models με Mel Frequency Cepstral Coefficients, τα οποία είναι τα χαρακτηριστικά των ηχητικών κυμάτων. Τα τραγούδια που έχει ακούσει ο χρήστης στο παρελθόν ομαδοποιούνται και αναλύονται. Με βάση λοιπόν, αυτήν την ανάλυση το σύστημα συστήνει κομμάτια μουσικής στους χρήστες.

Ξεκινώντας, η ραγδαία ανάπτυξη του διαδικτύου καθώς και ο αριθμός των ιστοσελίδων online μουσικής έχει αλλάξει τον τρόπο με τον οποίο ο χρήστης ακούει μουσική. Παλιότερα οι χρήστες άκουγαν μουσική μέσω τηλεόρασης ή ραδιοφώνου ή αγοράζοντας ένα CD σε αντίθεση με την σημερινή εποχή που η online μουσική είναι περισσότερο προσβάσιμη. Παρ' όλα αυτά οι παροχείς μουσικής στο διαδίκτυο δίνουν κομμάτια στους χρήστες χωρίς να υπολογίζουν τις προτιμήσεις τους.

Σε αυτήν την έρευνα, προτείνουν ένα σύστημα το οποίο θα βασίζεται στην προτίμηση του χρήστη μέσω της ανάλυσης μουσικού περιεχομένου.

Αντιλαμβάνονται την προτίμηση του χρήστη αναλύοντας κάποια χαρακτηριστικά της μουσικής όπως τόνος, τέμπο, ρυθμός και αρμονία[Scaringella, Zoia & Mlynek, 2007].

Οι περιγραφείς που χρησιμοποιούνται είναι οι MFCC.

MFCC. Περιλαμβάνουν τις πληροφορίες των χαρακτηριστικών όπως η τιμή των συντελεστών και τους χρησιμοποιούν για να δημιουργήσουν το μουσικό μοντέλο που αναγνωρίζει τις προτιμήσεις των χρηστών. Οι συγκεκριμένοι συντελεστές χρησιμοποιούνται στην αναγνώριση ομιλίας, παρ' όλα αυτά χρησιμεύουν στην ανάλυση της μουσικής, γιατί περιέχουν πληροφορία για το ηχητικό κύμα[Logan, 2000]. Πολλοί ερευνητές ανάλυσαν την μουσική χρησιμοποιώντας MFCCs. Ο G. Tzanetakis εργάστηκε πάνω στην ταξινόμηση του είδους μουσικής των ηχητικών σημάτων, χρησιμοποιώντας αρκετά σύνολα χαρακτηριστικών που χρησιμοποιούν MFCCs[Tzanetakis & Cook, 2002], ενώ ο J. J. Aucouturier εργάστηκε πάνω στην μέτρηση μουσικής ομοιότητας που βασίζεται σε MFCCs[Aucouturier & Pachet, 2002].

HMM. Το προτεινόμενο σύστημα αποτελείται από το στάδιο της μοντελοποίηση μουσικής, της ανάλυσης προτίμησης και της διαδικασίας σύστασης. Κατά την μουσική μοντελοποίηση, ένα κομμάτι μουσικής μετατρέπεται σε ένα HMM μοντέλο[Rabiner, 1993]. Το μοντέλο χρησιμοποιείται στις άλλες διαδικασίες της ανάλυσης και της σύστασης. Η προτίμηση του χρήστη περιλαμβάνει και την ομαδοποίηση των κομματιών που έχει ακούσει. Στην διαδικασία της σύστασης, το σύστημα υπολογίζει τα αποτελέσματα των υποψήφιων τραγουδιών για σύσταση, βασισμένη στις προτιμήσεις των χρηστών και δημιουργεί μια προτεινόμενη λίστα χρησιμοποιώντας τα αποτελέσματα.

Πείραμα και αξιολόγηση. Στο πείραμά τους χρησιμοποίησαν 160 τραγούδια. Επέλεξαν τραγούδια από 4 διαφορετικά είδη μουσικής, Jazz, R&B, Punk, Hip Hop και 40 τραγούδια ανά είδος. Εξήγαν τους MFCC από τα δείγματα και έφτιαξαν 160 HMM. Τα αποτελέσματα του πειράματος χωρίζονται σε δυο πίνακες. Στον πρώτο πίνακα η ακρίβεια για τα R&B είναι χαμηλότερη αλλά λογική. Σύγκριναν την δική τους προσέγγιση με μια προηγούμενη και στον δεύτερο πίνακα έδωσαν τα αποτελέσματα της προηγούμενης μεθόδου με την ίδια πειραματική υπόθεση. Εδώ η Jazz έχει χαμηλότερη ακρίβεια αλλά τα άλλα είδη κατά μέσο όρο βελτιώνονται.

Δεύτερο πείραμα και αξιολόγηση. Εξετάζουν 12 τύπους χρηστών. Ο χρήστης κάθε τύπου ακούει 10 τραγούδια από δυο είδη: 7 από το ένα είδος και 3 από το άλλο. Έχουμε έτσι 4 είδη μουσικής και 12 δυνατούς συνδυασμούς. Παρουσιάζουν τα αποτελέσματα σε ένα τρίτο πίνακα. Στην πρώτη περίπτωση το σύστημα συστήνει 64% Jazz, 25% R&B με ποσοστά παρόμοια στην προτίμηση του χρήστη. Σε άλλες περιπτώσεις η ακρίβεια σύστασης του

συστήματος είναι μικρότερη, διότι εκεί που ο χρήστης προτιμάει περισσότερο R&B από Hip Hop το σύστημα προτείνει το αντίθετο. Αυτό οφείλεται στην ομοιότητα των δυο ειδών μουσικής. Για αυτό το λόγο οι χρήστες μπερδεύουν τα δυο είδη και οι περισσότεροι σχεδιαστές τέτοιων συστημάτων ταξινομούν στην ίδια κατηγορία τα δυο είδη. Στο δικό τους σύστημα τα Hip Hop τραγούδια έχουν παρόμοια ομοιότητα διανυσμάτων και τα R&B τραγούδια είναι ταξινομημένα σε αρκετές ομάδες. Στην τελευταία περίπτωση είναι αυτός ο λόγος που το σύστημα προτείνει το Hip Hop.

Συνοψίζοντας, η προτεινόμενη προσέγγισή, δίνει έμφαση στην προτίμηση του χρήστη. Χρησιμοποίησαν έτσι, ένα σύστημα που προτείνει μουσική και συγκρίνοντας το με τις προτιμήσεις των χρηστών συμπέραναν από τα πειράματα ότι είναι παρόμοιες μεταξύ τους.

2009:Οι Cheng-Che Lu και Vincent S.Tseng στην μελέτη τους παρουσίασαν μια πρωτότυπη μέθοδο που την ονόμασαν ***personalized hybrid music recommendation*** και κατασκεύασαν ένα σύστημα το οποίο θα συστήνει μουσική στους χρήστες[Lu & Tseng, 2009]. Κάποια διαδικτυακά μουσικά καταστήματα, προτείνουν μουσική, η οποία έχει ρυθμιστεί από πολλούς ανθρώπους. Παρ' όλα αυτά, η παρούσα έρευνα προσπαθεί να αντιμετωπίσει τα εξής προβλήματα: 1. Πως μπορούμε να συστήνουμε αγαπημένη μουσική χωρίς αυτή να ρυθμίζεται από κανέναν 2. Πως να αποφεύγουμε επανειλημμένη πρόταση μη αγαπημένης μουσικής για χρήστες 3. Πως να προτείνεται περισσότερο ενδιαφέρουσα μουσική και σε χρήστες οι οποίοι δεν ακούν μουσική.

Οι χρήστες έχουν πρόσβαση στη μουσική μέσω του διαδικτύου, αλλά είναι δύσκολο να αναζητήσουν αγαπημένη μουσική από μουσικές βάσεις δεδομένων. Στο παρελθόν έχουν πραγματοποιηθεί μελέτες με μεθόδους σύστασης μουσικής, οι οποίες αντιμετωπίζουν δυο σημαντικά προβλήματα: δεν συστήνουν μουσική η οποία δεν έχει αξιολογηθεί και ίσως οι χρήστες δεν ενδιαφέρονται για ψηλά αξιολογημένη μουσική. Επιπλέον, πολυάριθμες μελέτες έχουν αφιερωθεί σε ανάλυση πολυμέσων και προσωπικές συστάσεις όπως μουσική ταξινόμηση και συστήματα σύστασης[Lee & Lu, 2003; Lu & Tseng, 2006; Tseng, Su & Huang, 2006; Tseng, Su, Wang & Lin, 2007; Lee, Lu & Liu, 2002]. Σε αυτήν την έρευνα δημιουργούμε μια πρωτότυπη μέθοδο, η οποία έχει ως στόχο να αντιμετωπίσει τα τρία προαναφερθέντα προβλήματα. Για να υπολογίσουν την ακρίβεια σύστασης, κατασκευάζουν ένα σύστημα που συστήνει μουσική στους χρήστες αφότου συγκεντρώσει τα ενδιαφέροντα των χρηστών.

Πείραμα και αξιολόγηση. Οι ερευνητές χρησιμοποίησαν 680 MIDI αρχεία από δημόσιες πηγές στο διαδίκτυο[http://www.downwiththeloads.com/, http://www.ingeb.org/, http://www.kididdles.com/, http://www.mididb.com/].

Το σύστημα που κατασκεύασαν συστήνει μουσική στους χρήστες, αφότου ερευνήσει τα μουσικά τους ενδιαφέροντα. Σε αυτό το σύστημα ο χρήστης καλείται να αξιολογήσει κάποια τραγούδια. Αρχικά, επιλέγει αν το τραγούδι του άρεσε ή όχι ή την επιλογή του κενού. Έπειτα, ανάλογα με την απάντησή του δηλώνει τους λόγους της επιλογής του κάνοντας κλικ στα χαρακτηριστικά που του παρατίθενται από το σύστημα, τόνος, ρυθμός κλπ. Με αυτόν τον τρόπο τα στοιχεία επιλέγονται ως αγαπημένα. Στα πειράματα χρησιμοποίησαν 27 εθελοντές. Έλεγξαν έτσι, αν η μέθοδός τους είναι ακριβής. Το σύστημα διαθέτει 10 στοιχεία σε κάθε χρήστη. Τα αποτελέσματα των πειραμάτων σε τρεις διαφορετικές χρονικές στιγμές είχαν ακρίβεια 84,5%, 90 % και 86,5% αντίστοιχα.

Κλείνοντας, στην παρούσα έρευνα μελέτησαν μια μέθοδο που συστήνει αγαπημένη μουσική στους χρήστες, παρακάμπτοντας τα ενδεχόμενα προβλήματα που προέκυψαν σε παλαιότερες έρευνες. Τα αποτελέσματα έδειξαν ότι η ακρίβεια του συστήματος είναι περίπου 90%.

2009: Οι Bo Shao, Dingding Wang, Tao Li και Mitsunori Ogihara παρουσίασαν μια στρατηγική για μουσική σύσταση με δύο απαιτήσεις[Shao, Wang, Li & Ogihara, 2009]. Πρώτη απαίτηση: ***High Recommendation Accuracy***. Εδώ, στόχος είναι η δημιουργία ενός συστήματος το οποίο θα παράγει μικρές λίστες από τραγούδια και στη συνέχεια ο χρήστης θα έχει την δυνατότητα να τα χαρακτηρίσει ως αγαπημένες ή όχι. Δεύτερη απαίτηση: ***High Recommendation Novelty***. Εδώ, αναφέρονται στην ποικιλία των καλλιτεχνών και στην ποικιλία του μουσικού περιεχομένου. Πιο συγκεκριμένα, το μουσικό περιεχόμενο θα πρέπει να δίνει πληροφορία για τον τόνο, το τέμπο και το ρυθμό και κατ' επέκταση αυτό να μην αποκλίνει από τις προτιμήσεις των χρηστών[Tzanetakis & Cook, 2002].

Παλαιότερα, ποικίλες έρευνες μουσικής σύστασης, που είχαν αναπτυχθεί, χρησιμοποίησαν πληροφορία τόσο για τον χρήστη και τα μουσικά περιεχόμενα, όσο και για το ιστορικό της μουσικής που άκουσε ο χρήστης και για την δισκογραφία[Cai, Zhang, Zhang & Ma, 2007; Chen & Chen, 2001; Logan, 2004; Oliver & Kreger-Stickles, 2006; Pachet, Cazaly & Roy, 2000; Pauws, Verhaegh & Vossen, 2006; Platt, Burges, Swenson, Weare & Zheng, 2002; Uitdenbogerd & Schyndel, 2002]. Τέτοιες προσεγγίσεις διαχωρίστηκαν σε δυο ομάδες: ***Collaborative-filtering methods*** και ***Content-based methods***.

Collaborative-filtering methods. Αυτές οι μέθοδοι συνιστούν τραγούδια σε χρήστες τα οποία είναι βασισμένα σε αξιολογήσεις άλλων χρηστών[Chen & Chen, 2001; Breese, Heckerman & Kadie, 1998; Cohen & Fan, 2000]. Σε περίπτωση που η βαθμολογία ενός στοιχείου από έναν χρήστη δεν είναι διαθέσιμη, τότε αυτές η μέθοδοι εκτιμούν το στοιχείο υπολογίζοντας

τον μέσο όρο όλων των αξιολογήσεων των στοιχείων από άλλους χρήστες. Αξίζει να σημειωθεί εδώ ότι για να είναι αποτελεσματική μια τέτοια μέθοδος χρειάζεται μεγάλος αριθμός αξιολογήσεων από τους χρήστες. Αυτός είναι και ο μεγαλύτερος περιορισμός[Schafer, Konstan & Riedi, 1999; Sarwar, Karypis, Konstan & Riedi, 2000].

Content-based methods. Αντίθετα, αυτές οι μέθοδοι συστήνουν μουσική χρησιμοποιώντας meta-data όπως τα είδη, τα στυλ, τους καλλιτέχνες και τους στίχους των τραγουδιών[Pauws, Verhaegh & Vossen, 2006; Ragno, Burges & Herley, 2005; Yoshii, Goto, Komatani, Ogata & Okuno, 2006] ή/και εξαγόμενα ακουστικά χαρακτηριστικά από δείγματα μουσικής[Wang, Li & Ogihara, 2009; Li & Ogihara, 2004; Huang & Jenor, 2004; Li, Kim, Guan & Oh, 2004]. Στη μουσική σύσταση, τα αντανακλαστικά και σταθερά ακουστικά χαρακτηριστικά εκπροσωπούν τραγούδι συγκεκριμένων χαρακτηριστικών όπως τα παραπάνω. Συγκρίνοντάς τα με τα ακουστικά χαρακτηριστικά συμπεραίνουμε ότι ένας μεγάλος αριθμός meta-data είναι περιγραφές περιεχομένου από μουσικούς. Η απόκτηση των meta-data παρόλα αυτά απαιτεί πολύ χρόνο και παράλληλα δεν παρέχει επαρκής πληροφορίες για τις προτιμήσεις των χρηστών[Li, Kim, Guan & Oh, 2004].

Στην παρούσα έρευνα, προτείνουν μια προσέγγιση την οποία ονόμασαν *DWA(Dynamic Weighting Approach)*. Η προσέγγιση αυτή δοκιμάζεται μέσα από μια σειρά πειραμάτων σε ένα πραγματικό σύνολο δεδομένων κατασκευασμένο από ανώνυμους χρήστες στο διαδίκτυο [http://www.newwisdom.net]. Χρησιμοποιεί το dynamic weighting scheme, βασισμένο σε σχέδια πρόσβασης χρήστη. Αυτό το σχέδιο συγκρίνει με ακουστικά χαρακτηριστικά και σχέδια πρόσβασης χρήστη με βάση κάθε αναπαράσταση. Με αυτό καταφέρνει να κερδίσει περισσότερη ακρίβεια για την ανθρώπινη αντίληψη στη μουσική.

Εξαγωγή χαρακτηριστικών. Εδώ, χρησιμοποιούν τονικά χαρακτηριστικά και συντελεστές ιστογραμμάτων κύματος για να εξάγουν χαρακτηριστικά. Το σύνολο των εξαγόμενων χαρακτηριστικών αποτελείται από 80 χαρακτηριστικά και τρία συστατικά, τους *Mel-Frequency Cepstral Coefficients (MFCCs)*, τα *Short-Term Fourier Transform Features (STFT)* και τα *Daubechies Wavelet Coefficient Histograms (DWCH)*.

Mel-Frequency Cepstral Coefficients. Είναι ένα σύνολο χαρακτηριστικών το οποίο χρησιμοποιείται κυρίως σε έρευνες για την αναγνώριση ομιλίας. Σχεδιάστηκε για να συλλαμβάνει βραχυπρόθεσμα spectral-based χαρακτηριστικά. Δηλαδή, πρώτα υπολογίζουν τον λογάριθμο πλάτους φάσματος βασισμένο στον μετασχηματισμό Fourier, όπου οι συχνότητες διαιρούνται σε δεκατρία bins χρησιμοποιώντας την κλιμάκωση Mel-Frequency και έπειτα αυτό ο διάνυσμα αποσυνδέεται χρησιμοποιώντας

διακριτό μετασχηματισμό συνημίτονου. Αυτό είναι το διάνυσμα MFCC (Mel-Frequency Cepstral Coefficient). Τα πρώτα 5 bins επιλέγονται και υπολογίζεται ο μέσος όρος και η διακύμανση του καθενός πάνω από τα καρέ.

Short-Term Fourier Transform Features. Αυτό είναι ένα σύνολο χαρακτηριστικών[Tzanetakis & Cook, 2002] σχετικό με τις τονικές υφές και δεν συλλαμβάνεται χρησιμοποιώντας τον MFCC. Αποτελείται από τους εξής τύπους χαρακτηριστικών: *spectral centroid, spectral rolloff, spectral flux, zero crossing* και *low energy*[Yang & Chen, 2011]. Το spectral centroid είναι το κέντρο βάρους του φασματικού πλάτους του μετασχηματισμού STFT. Το κέντρο βάρους είναι η μέτρηση της φασματικής μορφής. Το spectral rolloff είναι η ποσότητα της υψηλής συχνότητας στο σήμα και είναι και αυτό μέτρηση της φασματικής μορφής. Το spectral flux εκτιμάει την ποσότητα της γενικής φασματικής αλλαγής. Το zero-crossing μετρά το θόρυβο του σήματος που υπολογίζεται παίρνοντας την μέση και σταθερή απόκλιση του αριθμού των τιμών του σήματος που περνούν από το άξονα μηδέν σε κάθε χρόνο. Το low energy είναι το ποσοστό των καρέ που έχει ενέργεια μικρότερη από τον μέσο όρο ενέργειας όλου του σήματος[Yang & Chen, 2011].

Daubechies Wavelet Coefficient Histograms. Τα Daubechies Wavelet Filters είναι φίλτρα, τα οποία χρησιμοποιούνται στην ανάκτηση εικόνας[Daubechies, 1992; Li, Ogihara & Li, 2003]. Τα Daubechies Wavelet βασίζονται στο Ingrid Daubechies που είναι μια οικογένεια ορθογώνιων κυματιδίων η οποία καθορίζει ένα διακριτό μετασχηματισμό κυματιδίων και χαρακτηρίζεται από έναν μέγιστο αριθμό εκλιπόντων στιγμών για ένα δεδομένο χρονικό περιθώριο.

Πείραμα και αξιολόγηση. Εδώ, παρουσιάζουν την αξιολόγηση των επιδόσεων αυτής της προσέγγισης. Πολλές μελέτες χαρακτηρίζουν το σύστημά τους ποιοτικό.

Αρχικά, χρησιμοποίησαν ένα site το οποίο επισκέπτονται 6000 εγγεγραμμένοι χρήστες[http://www.newwisdom.net]. Οι χρήστες έχουν την δυνατότητα τόσο να ακούσουν μουσική όσο και να δημιουργήσουν τις δικές τους λίστες αναπαραγωγής. Τώρα το site διαθέτει 10.000 τραγούδια και εκατοντάδες λίστες αναπαραγωγής. Πάνω από το 80% των τραγουδιών είναι κινέζοι καλλιτέχνες αλλά και αμερικάνοι, ευρωπαίοι, γιαπωνέζοι και κορεάτες. Τα τραγούδια καλύπτουν ένα φάσμα από όλα τα είδη μουσικής. Στα πειράματά χρησιμοποιήθηκαν 2829 τραγούδια. Οι ερευνητές χρησιμοποίησαν λίστες αναπαραγωγής στις οποίες οι χρήστες είχαν συμπεριλάβει περισσότερα από 10 τραγούδια και λιγότερα από 20. Κατέληξαν στις 274 λίστες αναπαραγωγής.

Στη συνέχεια, επεξεργάστηκαν τα τραγούδια και τις λίστες αναπαραγωγής για να συλλέξουν τα χαρακτηριστικά περιεχομένου και τα σχέδια πρόσβασης χρήστη. Εδώ, το dynamic weighting scheme και ο

αλγόριθμος μουσικής αναβάθμισης, εφαρμόζεται για να παράγει τις ταυτοποιήσεις σύστασης των μουσικών κομματιών.

Έπειτα, χρησιμοποίησαν 50 τραγούδια από τρεις διαφορετικές τάξεις τις οποίες επέλεξαν οι χρήστες. Αυτές στα διαγράμματα έχουν τρία σχήματα αντίστοιχα, διαμάντι, κύκλος, αστέρι. Αξιολόγησαν τις μεθόδους που χρησιμοποίησαν και συμπέραναν ότι η δική τους προσέγγιση ξεπερνά τις άλλες μεθόδους επιλογής χαρακτηριστικών.

Τέλος, σύγκριναν την δική τους προσέγγιση με άλλες πέντε προσεγγίσεις. Η πρώτη προσέγγιση είναι η ***Content-Based Approach*** η οποία βασίζεται σε εξαγόμενα ακουστικά χαρακτηριστικά περιεχομένου από τα κομμάτια των τραγουδιών. Η δεύτερη προσέγγιση είναι η ***Artist-Based Approach*** η οποία είναι η μόνη που βασίζεται στον καλλιτέχνη ονομαστικά και έτσι συστήνει μουσική. Η τρίτη προσέγγιση είναι η ***Access-Pattern-Based Approach***. Εδώ, αυτή η προσέγγιση βασίζεται σε σχέδια πρόσβασης χρηστών. Δηλαδή, επιλέγει τα κομμάτια τα οποία εμφανίζονται πιο συχνά στις λίστες αναπαραγωγής. Η τέταρτη προσέγγιση αναφέρθηκε και προηγουμένως είναι η ***Hybrid Approach***. Αυτή η προσέγγιση προσπαθεί να ενσωματώσει την collaborative filtering method και την content-based method βασισμένη σε αλγόριθμους. Η τελευταία προσέγγιση είναι η DWA. Βασίζεται στην δική τους προσέγγιση και χρησιμοποιεί τα σχέδια πρόσβασης χρήστη για να βρει τα βάρη για το περιεχόμενο κάθε χαρακτηριστικού και έπειτα εκτελεί διάδοση ετικέτας και διαβάθμιση για την μουσική σύσταση.

Διεξήγαν μια σειρά από πειράματα για να συγκρίνουν την επίδοση των παραπάνω προσεγγίσεων. Η Content-Based Approach και η παρούσα έρευνα προσέγγιση έχουν μεγαλύτερη ποικιλία καλλιτεχνών από την Hybrid Approach και την Access-Pattern-Based Approach. Στην ποικιλία περιεχομένου η Content-Based Approach έχει την υψηλότερη ομοιότητα περιεχομένου και χαμηλή ποικιλία, ενώ η Access-Pattern-Based Approach έχει αρκετή ποικιλία έλλειψη ομοιότητας περιεχομένου. Η Hybrid Approach και η παρούσα έρευνα έχουν συγκρίσιμες επιδόσεις στην καλά ισορροπημένη ποικιλία περιεχομένου. Στην παραγωγή playlist αποτελέσματα δείχνουν ότι η παρούσα έρευνα έχει καλύτερη επίδοση ενώ η Content-Based Approach και η Hybrid Approach βρίσκονται στο ίδιο επίπεδο με χαμηλότερη επίδοση. Ελέγχοντας τα στατιστικά αποτελέσματα είδαν καθαρά ότι η δική τους υπερτερεί όλων των υπολοίπων.

Συνοψίζοντας, η δική τους προσέγγιση αποδείχτηκε αποτελεσματική και είναι η αφορμή για περαιτέρω έρευνα. Επιπλέον, μπορούν να ερευνηθούν περισσότερα περιεκτικά χαρακτηριστικά μουσικού περιεχομένου για παρόμοιες μετρήσεις.

2009:Οι Αλέξανδρος Νανόπουλος, Δημήτριος Ραφαιλίδης, Παναγιώτης Συμεωνίδης και Γιάννης Μανολόπουλος στην μελέτη τους εξέτασαν τη προσωπική μουσική σύσταση βασισμένη σε social tags[Νανόπουλος, Ραφαιλίδης, Συμεωνίδης & Μανολόπουλος, 2009]. Χρησιμοποιώντας μεθόδους βρήκαν συσχετισμούς ανάμεσα σε χρήστες-ετικέτες-μουσικά στοιχεία. Τέτοιοι μέθοδοι είναι οι three-order tensors. Πειραματικά χρησιμοποιούν πραγματικά δεδομένα από τον last.fm.

Πολλά web sites δίνουν την δυνατότητα σε χρήστες να βάλουν ετικέτες σε τραγούδια, άλμπουμ ή καλλιτέχνες. Οι social tags γίνονται όλο και πιο δημοφιλή στην ανάκτηση μουσικής πληροφορίας(MIR), τέτοιες όπως είναι το είδος, το στυλ, η διάθεση, η άποψη των χρηστών και η ενορχήστρωση. Έτσι αντίστροφα, σε ένα μοναδικό κομμάτι πληροφορίας όπως το είδος αναθέτεται από μια ταξινόμηση, τα social tags παρέχουν μια πολύπλευρη πηγή πληροφορίας για το μουσικό περιεχόμενο[Lamere, 2008].

Ωστόσο, υπάρχουν κάποιες προκλήσεις που θέτει η ελεύθερη φύση των social tags. Η πρώτη πρόκληση είναι ότι οι ετικέτες μπορεί να έχουν παραπάνω από μια σημασία και αυτό αποτελεί πρόβλημα (***polysemy***). Δηλαδή, μπορεί οι χρήστες να χαρακτηρίσουν ως ‘κλασσική’ για παράδειγμα την δεκαετία του ’80 και παράλληλα να έχουν χαρακτηριστεί το ίδιο και η ροκ μουσική της δεκαετίας του ’60. Έτσι ο χρήστης μπορεί να ανακτήσει μια ανάμειξη από μουσικά κομμάτια και από τις δυο κατηγορίες. Η δεύτερη πρόκληση είναι η ύπαρξη διαφορετικών ετικετών που έχουν παρόμοια σημασία το λεγόμενο πρόβλημα της συνωνυμίας (***synonymy***). Για παράδειγμα, κάποια κομμάτια της δεκαετίας του ’80 χαρακτηρίζονται ‘ορχηστρικά’ και δεν ανακτούνται μαζί με άλλα τα οποία χαρακτηρίζονται ‘κλασσικά’. Η τρίτη προσέγγιση δεν σχετίζεται με τις ερμηνείες των ετικετών. Αφορά μουσικά στοιχεία που έχουν χαρακτηριστεί ‘φτωχά’ (***sparsity*** ή ***cold-start***). Δηλαδή, αυτό αφορά ένα νέο κομμάτι το οποίο δεν έχει χαρακτηριστεί αλλά παρ’ όλα αυτά είναι στις πρώτες θέσεις των charts.

LSA. Αποτελεί μια τεχνική που ονομάζεται ***Latent Semantic Analysis***[Furnas, Deerwester & Dumais, 1988]. Είναι μια μέθοδος ανάκτησης μουσικής πληροφορίας (MIR), που διευθετεί τις παραπάνω προκλήσεις[Levy & Sandler, 2008]. Η συγκεκριμένη ανάλυση αποκαλύπτει κρυφές δομές σε δεδομένα χρησιμοποιώντας τεχνικές όπως το ***Singular Value Decomposition (SVD)***. Προσεγγίσεις που βασίζονται στο LSA[Levy & Sandler, 2008] υπολογίζουν το SVD μιας δισδιάστατης κλίμακας που αναπαριστά σχέσεις δύο τρόπων ανάμεσα σε μουσικά στοιχεία και ετικέτες.

HOSVD. Αποτελεί μια επέκταση της τεχνικής SVD, ονομάζεται ***Higher-Order SVD***, εφαρμόζεται σε πολλά επιστημονικά πεδία και βάσει αυτής προτείνονται οι καλύτερες συστάσεις[Kolda & Bader, 2009]. Η μέθοδος

αυτή έχει χρησιμοποιηθεί σε παλαιότερή τους μελέτη χωρίς όμως να έχει διευθετήσει την τρίτη προσέγγιση που αναφέρθηκε παραπάνω. Σε άλλες πάλι έρευνες το πρόβλημα αντιμετωπίστηκε με χαρακτηριστικά εξαγόμενα από ήχο[Eck, Lamere, Bertin-Mahieux & Green, 2007; Sordo, Laurier & Celma, 2007]. Ωστόσο αυτές οι έρευνες εστιάζουν στην αυτόματη πρόβλεψη ετικέτας και όχι σε προσωπική σύσταση μουσικών στοιχείων.

Εδώ, προτείνεται μια μέθοδος που βασίζεται στο HOSVD για να επεκταθεί παλαιότερή τους έρευνα πάνω στην σύσταση μουσικής, συνδυάζοντας εξαγόμενες ομοιότητες από ακουστικά χαρακτηριστικά με social tags.

Σχετικές έρευνες. Η μουσική σύσταση έχει απασχολήσει πολλούς ερευνητές. Κάποιες από αυτές τις έρευνες χρησιμοποίησαν μεθόδους βασισμένες στην μέτρηση ακουστικής ομοιότητας[Logan, 2004]. Άλλες πάλι, έγιναν στην προσπάθεια να γεφυρώσουν το σημασιολογικό χάσμα που υπάρχει και να αναπτύξουν υβριδικές μεθόδους μουσική σύστασης. Επίσης, κάποιες άλλες συνδύασαν collaborative filtering δεδομένα με δεδομένα ακουστικού περιεχομένου και άλλες κάνανε εκτίμηση των προτιμήσεων του χρήστη[Yoshii, Goto, Komatan, Ogata & Okuno, 2006]. Η παρούσα μέθοδος είναι διαφορετική από τις προαναφερθείσες μεθόδους στο γεγονός ότι δεν εκμεταλλεύεται τα social tags, των οποίων η μεγάλη προοπτική για το MIR είναι μόνο πρόσφατα αναγνωρισμένη[Lamere, 2008; Levy & Sandler, 2008].

Πειράματα και αξιολόγηση. Πειραματικά, σύγκριναν την δική τους προτεινόμενη μέθοδο, που την ονόμασαν ***MusicBox***, με δυο άλλες μεθόδους. Την σύσταση βασισμένη στο HOSVD που δεν μελετάει τα ακουστικά χαρακτηριστικά[Symeonidis, Ruxanda, Nanopoulos & Manolopoulos, 2008] και την σύσταση βασισμένη στο LSA που εφαρμόζεται σε items-tags σχέσεις, με παραγόμενες συστάσεις που βασίζονται στον Item-based (IB) αλγόριθμο[Sarwar, Karypis, Konstan & Riedi, 2001]. Τα πειράματά τους δείχνουν την ποιότητα της δικής τους σύστασης έναντι των άλλων μεθόδων, αφού είναι βελτιωμένη σε σχέση με τις άλλες μεθόδους και τις καταστέλλει. Αυτό οφείλεται στο γεγονός ότι, όλες οι μέθοδοι tensor-based (MusicBox και HOSVD) ξεπερνούν την μέθοδο LSA. Επιπλέον, εκμεταλλευόμενοι τις ακουστικές ομοιότητες, η μέθοδος είναι αποτελεσματική στην μείωση του sparsity. Τέλος, ταιριάζει καλύτερα με τις προσωπικές προοπτικές του κάθε χρήστη.

Συνοψίζοντας, εξέτασαν το πρόβλημα της προσωπικής μουσικής σύστασης βασισμένη σε social tags. Βρήκαν δηλαδή, συσχετισμούς ανάμεσα σε χρήστες ετικέτες και μουσικά στοιχεία, με την χρήση three-order tensors. Η σύστασή αυτή βασίζεται στην ανακάλυψη μιας κρυφής δομής σε αυτό το μοντέλο χρησιμοποιώντας το HOSVD το οποίο επεκτείνει το SVD σε high-

dimensional matrixes (tensors). Επιπλέον, βελτίωσαν την ποιότητας της σύστασης αντιμετωπίζοντας την τρίτη πρόκληση που προαναφέρθηκε, εκμεταλλευόμενοι τις ομοιότητες ανάμεσα στα μουσικά στοιχεία με βάση τα ακουστικά χαρακτηριστικά. Χρησιμοποίησαν πραγματικά δεδομένα από τον last.fm και τα πειράματα έδειξαν την υπεροχή της δικής τους μεθόδου όσον αφορά την ποιότητα της σύστασης. Τέλος, προτείνουν ως μελλοντική δουλειά, την επέκταση της δικής τους μεθόδου.

2009: Οι Byeong-jun Han, Seungmin Rho, Roger B. Dannenberg και Eenjun Hwan παρουσίασαν το *SMERS (SVR-based Music Emotion Recognition System)* με την χρήση ενός support vector regression (SVR) βασισμένο στο σύστημα μουσικής συναισθηματικής αναγνώρισης[Han, Rho, Dannenberg & Hwang, 2009]. Το μουσικό συναίσθημα παίζει σημαντικό ρόλο στην μουσική ανάκτηση, στον εντοπισμό της διάθεσης και στις εφαρμογές σχετικές με την μουσική. Πολλά θέματα έχουν διευθετηθεί σε διάφορους κλάδους όπως φυσιολογία, ψυχολογία, γνωστική επιστήμη και μουσικολογία. Η διαδικασία αναγνώρισης γίνεται σε τρία βήματα. Πρώτον, 7 διαφορετικά χαρακτηριστικά εξάγονται από την μουσική. Δεύτερον, αυτά τα χαρακτηριστικά χαρτογραφούνται σε 11 κατηγορίες συναισθημάτων χρησιμοποιώντας το δισδιάστατο μοντέλο συναισθήματος του Thayer[Thayer, 1989]. Τρίτον, δυο συναρτήσεις παλινδρόμησης εκπαιδεύονται χρησιμοποιώντας το SVR και προβλέπονται οι τιμές valence και arousal.

Με πρόσφατες έρευνες στον τομέα της ανάκτησης μουσικής πληροφορίας, υπάρχει ένα εμφανές ενδιαφέρον για την ανάλυση και κατανόηση του συναισθηματικού περιεχομένου της μουσικής. Εξαιτίας της ποικιλίας και της ευρύτητας του μουσικού περιεχομένου πολλοί ερευνητές έχουν αναζητήσει ένα πλήθος από ερευνημένα θέματα σε αυτόν τον τομέα στην μουσικολογία και στην ψυχολογία. Εδώ, αναπτύσσεται ένα σύστημα μουσικής αναγνώρισης συναισθήματος για την πρόβλεψη του valence και arousal ενός τραγουδιού βασισμένο σε ακουστικό περιεχόμενο.

Περιγραφή συστήματος. Το SMERS αποτελείται από τρία βήματα. Πρώτον, εξαγωγή χαρακτηριστικών. Εδώ, εξάγονται και αναλύονται 7 διαφορετικά χαρακτηριστικά. Δεύτερον, χαρτογράφηση. Εδώ, τα εξαγόμενα χαρακτηριστικά χαρτογραφούνται σε 11 κατηγορίες συναισθήματος στο δισδιάστατο μοντέλο συναισθήματος του Thayer[Thayer, 1989]. Τρίτον, εκπαίδευση. Το σύστημα χρησιμοποιεί τα εξαγόμενα χαρακτηριστικά ως εισαγόμενα διανύσματα για να εκπαιδεύσει το SVR.

Στην παρούσα έρευνα χρησιμοποιούνται 165 δυτικά pop τραγούδια. Συλλέχτηκαν 15 τραγούδια από καθεμία από τις 11 κατηγορίες συναισθήματος σε μια μεγάλη βάση δεδομένων μουσικής, το All Music

Guide[http://www.allmusic.com/], το οποίο παρέχει 180 κατηγορίες συναισθήματος για ταξινόμηση εισαγόμενων τραγουδιών. Τα μουσικά χαρακτηριστικά είναι ποικίλα όπως ***scale, intensity, rhythm,*** και ***harmonics***. Η κλίμακα είναι ένας συνολικός κανόνας τονικής διαμόρφωσης της μουσικής. Ο μέσος όρος ενέργειας του συνολικού κύματος συχνότητας χρησιμοποιείται ευρέως για μετρήσει την ένταση της μουσικής. Ο ρυθμός αφορά τα ρυθμικά χαρακτηριστικά όπως ***tempo*** και ***beat***, σημαντικά στοιχεία για την μουσική. Ο χτύπος είναι βασικό στοιχείο της μουσικής ενώ το τέμπο είναι οι χτύποι ανά λεπτό οι οποίοι αντιπροσωπεύουν ολόκληρο το ρυθμικό χαρακτηριστικό της μουσικής. Οι αρμονίες παρατηρούνται σε μουσικούς τόνους. Σε μονοφωνική μουσική παρατηρούνται εύκολα αρμονίες στο διάγραμμα φάσματος, αντίθετα στην πολυφωνική μουσική είναι δύσκολο να βρούμε αρμονίες γιατί πολλά μουσικά όργανα και φωνές εκτελούνται αμέσως.

SVR (Support Vector Regression) εκπαίδευση[Yang, Lin, Su & Chen, 2008; Smola, 2004]. Είναι μια εφαρμογή της SVM για να βρεθεί η συνάρτηση χαρτογράφησης ανάμεσα στην είσοδο και έξοδο. Βασίζεται στη βιβλιοθήκη LIBSVM[Chih-Chung & Chih-Jen, 2001].

SVM (Support Vector Machine) εκπαίδευση. Για την ταξινόμηση του συναισθήματος χρησιμοποιείται μια SVM με πολλές τάξεις. Η SVM ταξινομεί μόνο μια τάξη την στιγμή. Έτσι, χρησιμοποιούνται 11 SVMs για να ταξινομήσουμε κάθε συναίσθημα ξεχωριστά.

GMM (Gaussian Mixture Model) εκπαίδευση. Είναι μοντέλο το οποίο μοντελοποιεί τα μουσικά χαρακτηριστικά. Εδώ, χρησιμοποιούνται 7 τέτοια μοντέλα για τα σύνολα valence και arousal. Κάθε GMM εκπαιδεύεται χρησιμοποιώντας τον αλγόριθμο EM(Expectation Maximization).

Πειράματα και αξιολόγηση. Στα πειράματα χρησιμοποιούνται δυο συστήματα το Καρτεσιανό και το Πολικό. Τα αποτελέσματα των SVMs στο Καρτεσιανό σύστημα είναι καλά σε ιδιαίτερα μουσικά συναισθήματα όπως ***angry, boring*** και ***relaxing***. Ωστόσο κάποια άλλα διαγώνια στοιχεία είχαν άσχημα αποτελέσματα. Αλλάζοντας την SVM σε SVR αυξήθηκε η ερμηνεία. Κατά μέσο όρο, 9,5 τραγούδια ταξινομήθηκαν σωστά ενώ 21 τραγούδια δεν ταξινομήθηκαν στο relaxing. Τα αποτελέσματα του GMM στο Καρτεσιανό σύστημα 12,8 τραγούδια ταξινομήθηκαν σωστά. Παρ' όλα αυτά κάποια συναισθήματα δεν ταξινομήθηκαν σωστά όπως angry (4 τραγούδια), sad (5 τραγούδια) και boring (2 τραγούδια). Τα αποτελέσματα του SVR στο Πολικό σύστημα δείχνουν ότι οι μη ισορροπημένες ταξινομήσεις ήταν σημαντικά μειωμένες. Δηλαδή, 14,2 τραγούδια ταξινομήθηκαν σωστά και τα μη σωστά ταξινομημένα ήταν μόνο relaxing (2 τραγούδια) και boring (3 τραγούδια). Όλα τα αποτελέσματα στο Καρτεσιανό σύστημα έχουν μέγιστη ακρίβεια 91,52% .

Στο Πολικό σύστημα αντίθετα η ακρίβεια αυξάνεται κατά 94,55% χρησιμοποιώντας τον SVR και 92,73% χρησιμοποιώντας τον GMM.

Κλείνοντας, αυτή η έρευνα ανέπτυξε ένα σύστημα χρησιμοποιώντας τους εξής αλγόριθμους: SVR, SVM και GMM. Τα πειράματά τους έδειξαν ότι η ακρίβεια του SVR στο Πολικό σύστημα αυξήθηκε από 63.03% σε 94.55% ενώ ο GMM στο Πολικό σύστημα έχει αυξημένη ακρίβεια από 91.52% σε 92.73%.

2010:Οι Chuan-Yu Chang, Chun-Yen Lo, Chi-Jane Wang και Pau-Choo Chung στην μελέτη τους εφαρμόζουν έναν συσχετισμό ο οποίος καθορίζει τα χαρακτηριστικά της μουσικής που προκαλούν ένα συναίσθημα[Chang, Lo, Wang & Chung, 2010]. Στην συνέχεια χρησιμοποιούν δυο Support Vector Machines που ταξινομούν την μουσική που προκαλεί συναισθήματα όπως ***happiness, anger, sadness*** και ***peacefulness***.

Το να ακούς μουσική είναι χαλαρωτικό. Παρ' όλα αυτά η μουσική δεν είναι πάντα αποτελεσματική σε μεταβαλλόμενα συναισθήματα. Υπάρχουν συστήματα που αναγνωρίζουν συναισθήματα αυτόματα από μουσική. Ένα σύστημα μουσικής σύστασης αναλύει μουσική και συναισθήματα. Η ανάλυση των συναισθημάτων από μουσική και η ταξινόμησή τους είναι δυο διαφορετικά κομμάτια[Yang, Lin, Su & Chen, 2008]. Το τι συναίσθημα θα προκληθεί στον κάθε χρήστη ακούγοντας μουσική είναι υποκειμενικό γι' αυτό και είναι δύσκολο σε ένα σύστημα να ταξινομήσει τα συναισθήματα. Έτσι, στην παρούσα έρευνα προτείνεται μια προσέγγιση που μειώνει αυτήν την υποκειμενικότητα.

Σχετικές έρευνες. Με σκοπό να ανακαλύψουν ερευνητές την σχέση ανάμεσα σε μουσική και συναίσθημα το οποίο ίσως προκαλείται, τα συναισθήματα έχουν κατηγοριοποιηθεί σε πολλές τάξεις και έχουν παραχθεί πολλά σχέδια αναγνώρισης τα οποία ταξινομούν τη μουσική[Lu, Liu & Zhang, 2006; Yang, Lin, Su & Chen, 2008]. Συναισθήματα όπως ***happiness, anger, sadness***, έχουν ταξινομηθεί χρησιμοποιώντας ποικίλα μοντέλα συναισθήματος όπως το μοντέλο του Thayer[Thayer, 1989], το μοντέλο valence-arousal[Osgood, Suci & Tannenbaum, 1957], και το μοντέλο του Russell[Russell, 1980]. Ο Yang πρότεινε ένα συνδυαστικό μοντέλο, το οποίο χρησιμοποιεί το μοντέλο valence-arousal του Thayer, που ενσωματώνει το μοντέλο valence-arousal[Osgood, Suci & Tannenbaum, 1957] και το μοντέλο του Russell[Russell, 1980]. Με βάση το μοντέλο του Thayer μπορούν να δημιουργηθούν 4 τύποι συναισθήματος: ***happiness, nervousness, sadness*** και ***peacefulness***. Το arousal είναι η ένταση του συναισθήματος, ενώ το valence ο βαθμός του θετικού η αρνητικού συναισθήματος. Εδώ, χρησιμοποιείται το μοντέλο του Thayer για την ταξινόμηση συναισθήματος.

Εξαγωγή χαρακτηριστικών. Έχουν προταθεί πολλά χαρακτηριστικά για να περιγράψουν την μουσική. Τέτοια είναι τα: *linear prediction coefficients (LPC)*, οι *linear prediction cepstrum coefficients (LPCC)*[Dhanalakshmi, Palanivel & Ramalingam, 2009; Changsheng, Maddage & Xi, 2005], οι *Mel-frequency cepstral coefficients (MFCC)*[Shao, Wang, Li & Ogihara, 2009; Dhanalakshmi, Palanivel & Ramalingam, 2009; Changsheng, Maddage & Xi, 2005; Lee, Lin, Yu & Shih, 2009], *entropy και dynamism*[Ajmera, McCowan & Bourlard, 2003], *timbre*[Shao, Wang, Li & Ogihara, 2009; Lu, Liu & Zhang, 2006; Zhu, Shi, Kim & Eom, 2006; Dhanalakshmi, Palanivel & Ramalingam, 2009; Ajmera, McCowan & Bourlard, 2003], *intensity*[Lu, Liu & Zhang, 2006], *rhythm*[Lu, Liu & Zhang, 2006; Tzanetakis & Cook, 2002], *pitch*[Tzanetakis & Cook, 2002], *amplitude envelope*[Changsheng, Maddage & Xi, 2005] και *Daubechies wavelet coefficients histograms*[Shao, Wang, Li & Ogihara, 2009]. Αυτά τα χαρακτηριστικά εφαρμόζονται αμέσως, χαρακτηρίζουν την μουσική βραχυπρόθεσμα αλλά ίσως παραμελήσουν σημαντικές ιδιότητες σε μακροπρόθεσμη πληροφορία που προκαλεί συναισθήματα σε χρήστες.

Στην παρούσα έρευνα χρησιμοποίησαν μια μακροπρόθεσμη προσέγγιση που οργανώνει τα χαρακτηριστικά της μουσικής σε μια σειρά από ακολουθίες χαρακτηριστικών. Εξάγονται 21 χαρακτηριστικά για κάθε καρέ: LPCC, MFCC, intensity (average, variance, maximum, intensity), timbre (centroid, bandwidth, rolloff, flux, peak, valley, contrast, zero crossing rate) στον τομέα της συχνότητας και intensity (average, variance, intensity), timbre (centroid, bandwidth, rolloff, flux) στον τομέα του χρόνου. 7 χαρακτηριστικά επιλέγονται για το arousal στον τομέα του χρόνου τα variance, centroid και στον τομέα της συχνότητας average, variance, maximum, rolloff, flux. Για το valence εξάγονται 9 στον τομέα του χρόνου τα average, variance, bandwidth, rolloff, flux και στον τομέα της συχνότητας τα average, variance, centroid, rolloff.

Ταξινόμηση. Για την ταξινόμηση των συναισθημάτων απαιτείται ένας ταξινομητής. Σε παλαιότερες μελέτες, έχουν χρησιμοποιηθεί ταξινομητές όπως *support vector machine (SVM)*[Dhanalakshmi, Palanivel & Ramalingam, 2009], *support vector regression(SVR)*[Yang, Lin, Su & Chen, 2008; Changsheng, Maddage & Xi, 2005], *gaussian mixture models (GMM)*[Lu, Liu & Zhang, 2006], *hidden markov model (HMM)*[Ajmera, McCowan & Bourlard, 2003], K-*nearest neighbor (KNN)*[Tzanetakis & Cook, 2002; Li & Ogihara, IEEE 2006]. Εδώ, χρησιμοποιήθηκε ο ταξινομητής support vector machine (SVM) γιατί έχει την υψηλότερη ακρίβεια στην ταξινόμηση.

Πειράματα και αξιολόγηση. Στα πειράματά τους χρησιμοποιήθηκαν 293 αρχεία μουσικής συμπεριλαμβάνοντας ερμηνείες πιάνου, συμφωνική

μουσική, φλάουτο, ερμηνείες άρπας και hip hop μουσική. Τα κομμάτια είναι μόνο μουσικά όργανα και όχι φωνητικά. Το κάθε κομμάτι παίζει για δυο λεπτά και όταν σταματήσει ο χρήστης σημειώνει ποιό συναίσθημα του προκάλεσε το κομμάτι. Το πείραμα συνεχίζεται μέχρι τουλάχιστον έξι κομμάτια για κάθε συναίσθημα που ταυτοποιείται. Στο σύνολο ο κάθε χρήστης ακούει 24 κομμάτια. Όσον αφορά στην αξιολόγηση, σύγκριναν την δική τους μέθοδο με την μέθοδο του Yang[Yang, Lin, Su & Chen, 2008] που χρησιμοποιεί το SVR για ταξινόμηση. Τα ποσοστά του Yang είναι 33,85% μέσο όρο αναγνώρισης συναισθήματος, 48,46% arousal και 60,77% valence ενώ της δικής τους είναι 73,08%, 81,54% και 85,38% αντίστοιχα. Με βάση τα ποσοστά η δική τους υπερτερεί, έχει υψηλή ακρίβεια ταξινόμησης και βρίσκεται πιο κοντά στην αντίληψη του χρήστη.

Κλείνοντας, πρότειναν μια προσέγγιση για να βρουν ακολουθίες συναισθηματικής μουσικής η οποία ίσως προκαλεί ένα ιδιαίτερο συναίσθημα στους χρήστες. Χρησιμοποίησαν το SVM για να ταξινομήσουν τα συναισθήματα για κάθε χρήστη. Κατάφεραν να δημιουργήσουν μια προσωπική μουσική βάση δεδομένων που αναγνωρίζει συναισθήματα από μουσική για τους χρήστες. Τα πειράματα έδειξαν την αποτελεσματικότητα του συστήματος.

2011:Οι Seungjae Lee, Jung Hyun Kim, Sung Min Kim και Won Young Yoo παρουσιάζουν ένα πρόγραμμα σύστασης μουσικής βασισμένο στη διάθεση, το SMOODI[Lee, Kim, Kim & Yoo, 2011]. Είναι μια εφαρμογή που χωρίζεται σε τρεις ενότητες: ***Mood Square, Cover Flow, Mood Cloud***. Στο Mood Square, οι χρήστες ελέγχουν την διανομή συναισθήματος των τοπικών clips και παράγουν λίστες αγγίζοντας τα κελιά διάθεσης. Στο Cover Flow, οι χρήστες μπορούν να βρουν μουσική πληροφορία και να παράγουν λίστες παρόμοιων τραγουδιών ‘σέρνοντας’ ένα τραγούδι. Στο Tag Cloud, ο χρήστης μπορεί να δημιουργήσει λίστες αναπαραγωγής επιλέγοντας ετικέτες διάθεσης. Για αυτήν την εφαρμογή, ανέπτυξαν ένα νέο μοντέλο διάθεσης από συλλεγμένες ετικέτες και τιμές θέσης valence-arousal και σχεδίασαν μια συνάρτηση για να εκτιμήσουν πιθανότητες διάθεσης.

Η μουσική είναι μια τέχνη η οποία περιγράφει την ανθρώπινη σκέψη και συναίσθημα από ήχους που βρίσκονται τριγύρω. Η ανάπτυξη της τεχνολογίας της πληροφορίας βοήθησε στην απόκτηση περισσότερης μουσικής και στην πρόσβαση σε αυτήν. Ωστόσο, υπάρχουν λίγοι τρόποι για να βρει κάποιος την μουσική που θέλει. Το κλειδί είναι η έρευνα, τρόπος στον οποίο κάποιος μπορεί να βρει μουσική από το όνομα του καλλιτέχνη ή του τραγουδιού, αλλά πολλές φορές είναι αδύνατον να θυμάται όλα τα ονόματα τραγουδιών και καλλιτεχνών. Επιπλέον αν ο χρήστης δεν χρησιμοποιήσει την σωστή λέξη-κλειδί στην αναζήτηση κάποιου τραγουδιού τότε τα

αποτελέσματα θα είναι λάθος. Αντίθετα η αναζήτηση ενός συναισθήματος είναι πιο εύκολη γιατί ο καθένας μπορεί να επιλέξει μουσική από αυτό. Η ταξινόμηση συναισθήματος χωρίζεται σε τρεις κατηγορίες: *κατηγοριοποίηση συναισθήματος ή διάθεσης, εξαγωγή χαρακτηριστικού και ταξινόμηση*. Παλαιότερες έρευνες προτείνουν τέτοιες μεθόδους που είναι αρκετά καλές[Lu, Liu & Zhang, 2006; Feng, Zhuang & Pan, 2003; Yang, Lin, Su & Chen, 2008; Korhonen, Clausi & Jernigan, 2006; Hu, 2010], παρ' όλα αυτά όχι τόσο πετυχημένες λόγω της ασάφειας της μουσικής διάθεσης και της υποκειμενικότητας της απόφασης της μουσικής διάθεσης. Εδώ, παρουσιάζουν το SMOODI το οποίο προέρχεται από τις λέξεις smart και mood.

Μοντέλο διάθεσης. Το συναισθηματικό μοντέλο διαιρείται σε δυο προσεγγίσεις, την κατηγορηματική και την διαστατική. Στην πρώτη, τα συναισθήματα είναι τα εξής: ***angry, fear, sad, happy*** και ***disgusting***[Ekman, 1992]. Στην δεύτερη, τα συναισθήματα παρουσιάζονται σε άξονες δύο ή τριών διαστάσεων[Thayer, 1989; Russell, 1980]. Το μοντέλο βασίζεται σε τιμές valence-arousal σε αντίθεση με προηγούμενα μοντέλα.

Μοντέλο σύστασης. Η διανομή των valence-arousal δείχνει την χρησιμότητα των δύο αυτών χαρακτηριστικών στην σύσταση μουσικής. Στη σύσταση μουσικής υπάρχουν δυο δυνατές προσεγγίσεις: σύσταση με βάση το διάγραμμα valence-arousal και σύσταση με βάση την διαβάθμιση. Στην πρώτη κάθε κομμάτι έχει μια θέση στο διάγραμμα valence-arousal και έτσι συστήνεται η μουσική. Στην δεύτερη κάθε κατηγορία διάθεσης εκτιμάται από τις τιμές των valence-arousal και με αυτόν τον τρόπο παράγονται λίστες αναπαραγωγής.

Πειράματα και αξιολόγηση. Στα πειράματα χρησιμοποιήθηκαν 8 κατηγορίες διάθεσης και 1000 clips από ποικίλα είδη μουσικής. Το προτεινόμενο μοντέλο διάθεσης επέλεξε τυχαία 10 ταξινομημένα μουσικά clips για κάθε κατηγορία. 80 clips χρησιμοποιήθηκαν για το τεστ ταξινόμησης. Απαιτείται ο χρήστης να επιλέξει ναι ή όχι όταν παίζει κάθε κομμάτι και παρουσιάζεται η περισσότερο πιθανή κατηγορία διάθεσης. Το πείραμα διεξάχθηκε σε 15 χρήστες. Ο μέσος όρος ταξινόμησης είναι 67,5%. Κάθε μουσικό clip έχει ποικίλα συναισθήματα και μόνο σε μια κατηγορία μπορεί να περιγραφεί η διάθεση του τραγουδιού. Αν υποθέσουμε ότι 2 από τις 8 κατηγορίες περιγράφουν την διάθεση του τραγουδιού, η πιθανότητα τουλάχιστον μιας σωστής κατηγορίας φτάνει στο 80,33%.

Συνοψίζοντας, σε αυτή την μελέτη παρουσίασαν μια εφαρμογή μουσικής σύστασης, η οποία παρέχει εύκολες μεθόδους για να συστήσει μουσική και να δημιουργήσει λίστες αναπαραγωγής από ένα 'άγγιξε' και 'σύρε'. Για αυτήν την εφαρμογή, προτείνουν το μοντέλο διάθεσης και το μοντέλο σύστασης. Τα πειράματα έδειξαν την χρησιμότητα αυτής της

εφαρμογής στην παραγωγή λιστών αναπαραγωγής για τους ακροατές. Τέλος, το μοντέλο διάθεσης μπορεί να χρησιμοποιηθεί σε μελλοντικές εφαρμογές.

Κλείνοντας το αυτό το κεφάλαιο συνειδητοποιούμε την σημαντικότητα που έχει η μουσική στη ζωή του ανθρώπου. Είναι εμφανές πως η μουσική επηρεάζει το συναίσθημα. Ανακεφαλαιώνοντας, κάποια από τα συστήματα που αναπτύχθηκαν ήταν τα εξής: smart radio, DWA, SMERS, SMOODI.

Κεφάλαιο 3 – Εξαγωγή Παραμέτρων

3.1 Τα χαρακτηριστικά της μουσικής

Η εμπειρία του να ακούς μουσική είναι πολυδιάστατη. Για παράδειγμα το arousal έχει σχέση με το ρυθμό, τον τόνο, την ένταση και το ηχόχρωμα, ενώ το valence έχει σχέση με την λειτουργία και την αρμονία[Gabrielsson & Lindstrom, 2001]. Το valence ορίζει πόσο συναρπαστικό και ήρεμο είναι ένα τραγούδι ενώ το arousal πόσο θετικό ή αρνητικό είναι το συναίσθημα που προκαλείται. Η συναισθηματική αντίληψη σχετίζεται με τον συνδυασμό πολλών παραγόντων της μουσικής[Hevner, 1935; Rigg, 1964]. Δηλαδή, μουσικά κομμάτια με υψηλού τόνου συγχορδίες έχουν πιο θετικό valence σε αντίθεση με ήρεμου τόνου συγχορδίες. Τα χαρακτηριστικά γνωρίσματα που εξάγονται όταν ακούμε μουσική χωρίζονται σε πέντε κατηγορίες: *ενεργειακά, ρυθμικά, διαχρονικά, φασματικά και αρμονικά*.

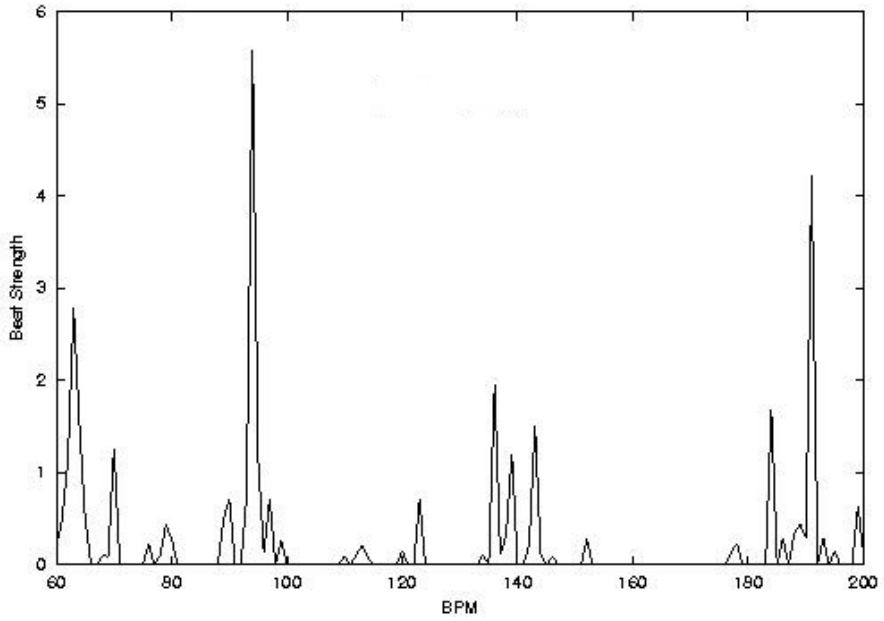
3.1.1 Ενεργειακά χαρακτηριστικά

Η ενέργεια σχετίζεται σε μεγάλο βαθμό με το arousal[Gabrielsson & Lindstrom, 2001]. Η μέτρηση της λαμβανόμενης έντασης γίνεται με την χρήση του μοντέλου δυναμικής έντασης των Chalupper και Fastl[Chalupper & Fastl, 2002] και εφαρμόζεται στο PsySound[Cabrera, 1999; Ricard, 2004]. Το PsySound είναι ένα πρόγραμμα υπολογιστή που οι παράμετροι των μοντέλων της ακουστικής αίσθησης βασίζονται σε ψυχο-ακουστικά μοντέλα όπως η Bark critical band[Zwicker, 1961] για μοντελοποίηση ακουστικών φίλτρων στα αυτιά μας, το ακουστικό μοντέλο χρονικής ολοκλήρωσης, και το μοντέλο για την μοντελοποίηση της ευκρίνειας των Zwicker και Fastl, μια υποκειμενική μέτρηση του ήχου σε μια επεκτατική κλίμακα από το πολύ βαρετό στο οξύ[Zwicker & Fastl, 1999]. Η sound description toolbox (SDT) εξάγει έναν αριθμό MPEG-7 standard descriptors και άλλους από ηχητικά σήματα[Benetos, Kotti & Kotropoulos, 2007]. Εξάγονται 40 ενεργειακά χαρακτηριστικά όπως *audio power (AP), total loudness (TL),* και *specific loudness sensation coefficients (SONE)*. Το AP είναι η δύναμη του ηχητικού σήματος. Η εξαγωγή των total loudness (TL) και specific loudness sensation coefficients (SONE) βασίζεται στα αντιληπτικά μοντέλα που εφαρμόζονται στην εργαλειοθήκη μουσικής ανάλυσης (Music Analysis toolbox)[Pampalk, 2004], συμπεριλαμβάνοντας το outer ear model, την Bark critical-band rate scale και

την φασματική συγκάλυψη. Το φάσμα ισχύος που προκύπτει αντικατοπτρίζει την ένταση της ανθρώπινης αίσθησης και ονομάζεται sonogram[Pampalk, Rauber & Merkl, 2002]. Οι specific loudness sensation coefficients (SONE) είναι οι συντελεστές που υπολογίζονται από το sonogram, το οποίο αποτελείται από έως και 24 Bark critical bands (ο ακριβής αριθμός εξαρτάται από την συχνότητα δειγματοληψίας του ηχητικού σήματος). Το total loudness (TL) υπολογίζεται ως μια συνάθροιση βασισμένη στη μέθοδο του Stevens[Hartmann, 1998], η οποία παίρνει το σύνολο του μεγαλύτερου SONE συντελεστή και μια αναλογία 0,15 του συνόλου των υπολειπόμενων συντελεστών.

3.1.2 Ρυθμικά χαρακτηριστικά

Ρυθμός είναι το μοτίβο των παλμών ποικίλης δύναμης. Συνήθως περιγράφεται με τους όρους τέμπο, μέτρο ή διατύπωση. Ένα κομμάτι με γρήγορο ρυθμό έχει υψηλό arousal ενώ με συνεχή ρυθμό έχει θετικό valence και ο σταθερός ρυθμός έχει αρνητικό valence[Gabrielsson & Lindstrom, 2001]. Για την εξαγωγή των ρυθμικών χαρακτηριστικών χρησιμοποιούνται πολλές εργαλειοθήκες όπως είναι το Marsyas[Tzanetakis & Cook, 2002]. Το Marsyas είναι ένα δωρεάν λογισμικό το οποίο χρησιμοποιείται για να αναπτύξει γρήγορα και να αξιολογήσει εφαρμογές ήχου στον υπολογιστή. Το Marsyas χρησιμοποιείται για να υπολογίσει το ***beat histogram*** της μουσικής και να εξάγει 6 χαρακτηριστικά από αυτό. Η εικόνα 1 είναι ένα παράδειγμα.



beat histogram

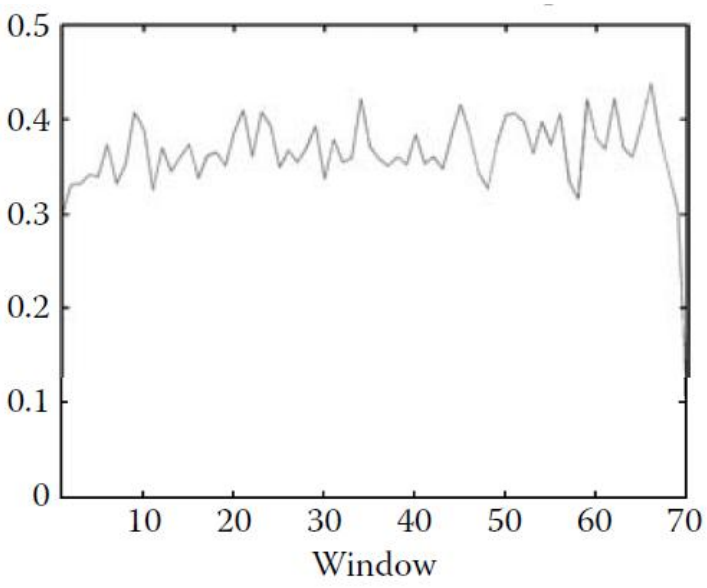
Τα βασικά από αυτά τα χαρακτηριστικά είναι: *beat strength, amplitude* και *period* της πρώτης και της δεύτερης κορυφής του beat histogram, και *ratio* της δύναμης των δυο κορυφών σε χτύπους ανά λεπτό (bpm). Το beat histogram κατασκευάζεται υπολογίζοντας την αυτοσυσχέτιση του σήματος σε κάθε ζώνη συχνότητας οκτάβας. Οι κυρίαρχες κορυφές της συνάρτησης αντιστοιχούν σε διάφορες περιοδικότητες του σήματος. Η εργαλειοθήκη μουσικής ανάλυσης (MA toolbox)[Pampalk, 2004] χρησιμοποιείται για να εξάγει το *rhythm pattern* που περιέχει πληροφορία για το πόσο δυνατοί και γρήγοροι είναι οι χτύποι που παίζονται μέσα στις αντίστοιχες ζώνες συχνοτήτων[Pampalk, Rauber & Merkl, 2002]. Αυτή η εργαλειοθήκη χρησιμοποιεί τον μετασχηματισμό STFT (short-time Fourier transform) για να αποκτηθεί η διαμόρφωση πλάτους του SONE για κάθε τμήμα 6 δευτερολέπτων του μουσικού κομματιού. Το rhythm pattern ενσωματώνεται σε ένα ιστόγραμμα ρυθμού των 60-bin αθροίζοντας τους συντελεστές διαμόρφωσης πλάτους κατά μήκος των ζωνών. Η μέση τιμή του ιστογράμματος ρυθμού είναι και ο μέσος όρος του τέμπο. Αυτά τα χαρακτηριστικά εξάγονται από τον rhythm pattern (RP) extractor[Lidy & Rauber, 2005]. Παράλληλα, με την εργαλειοθήκη ανάκτησης μουσικής πληροφορίας (MIR toolbox)[Lartillot & Toiviainen, 2007], όπως προτείνεται σε έρευνες, εξάγονται πέντε ρυθμικά χαρακτηριστικά: *rhythm strength, rhythm regularity, rhythm clarity, average onset frequency,* και *average tempo*[Lu, Liu & Zhang, 2006]. Το rhythm strength υπολογίζεται από τον μέσο όρο έναρξης της δύναμης της καμπύλης εντοπισμού έναρξης, που υπολογίζεται με βάση τον αλγόριθμο του Klapuri[Klapuri, 1999]. Ο όρος έναρξη αναφέρεται στον χρόνο έναρξης του τραγουδιού. Τα rhythm regularity και rhythm clarity υπολογίζονται κάνοντας αυτοσυσχέτιση στην καμπύλη εντοπισμού έναρξης. Το average onset frequency υπολογίζεται ως ο αριθμός των ενάρξεων ανά δευτερόλεπτο, ενώ το average tempo υπολογίζεται εντοπίζοντας την περιοδικότητα από την καμπύλη εντοπισμού έναρξης.

3.1.3 Διαχρονικά χαρακτηριστικά

Στα διαχρονικά χαρακτηριστικά χρησιμοποιείται η SDToolbox (Sound Description)[Benetos, Kotti & Kotropoulos, 2007] για να εξάγει τα *zero-crossing rate, temporal centroid,* και *log attack time* για να βρεθεί η διαχρονική ποιότητα της μουσικής. Το zero-crossing rate μετρά το θόρυβο του σήματος που υπολογίζεται παίρνοντας την μέση και σταθερή απόκλιση του αριθμού των τιμών του σήματος που περνούν από το άξονα μηδέν σε κάθε χρόνο. Το *zero-crossing rate* υπολογίζεται από την παρακάτω σχέση:

$$\text{Zero crossing rate} = \frac{1}{T}\sum_{t=m-T+1}^{m}\frac{\left|\text{sgn}(s_t) - \text{sgn}(s_{t-1})\right|}{2}\times w(m-t),$$

Όπου το *T* είναι το μήκος του παράθυρου του χρόνου s_t είναι το πλάτος των *t* δειγμάτων χρόνου και *w(.)* είναι ένα ορθογώνιο παράθυρο. Το αρχικό μήκος του παραθύρου του χρόνου είναι 1,6% του συνολικού μήκους του εισαγόμενου σήματος με 10% επικάλυψη. Η εικόνα 2 παρουσιάζεται ένα παράδειγμα.



Zero crossing rate

Το temporal centroid και log attack time είναι δυο MPEG-7 harmonic instrument timbre περιγραφείς που περιγράφουν την ενέργεια του σήματος[Allamanche, Herre, Helmuth, Froba, Kasten & Cremer, 2001]. Και τα δυο χρησιμοποιούνται στην ταξινόμηση μουσικού οργάνου. Τέλος, η σταθερή διακύμανση του zero-crossing rate ίσως είναι χρήσιμη στην πρόβλεψη του valence.

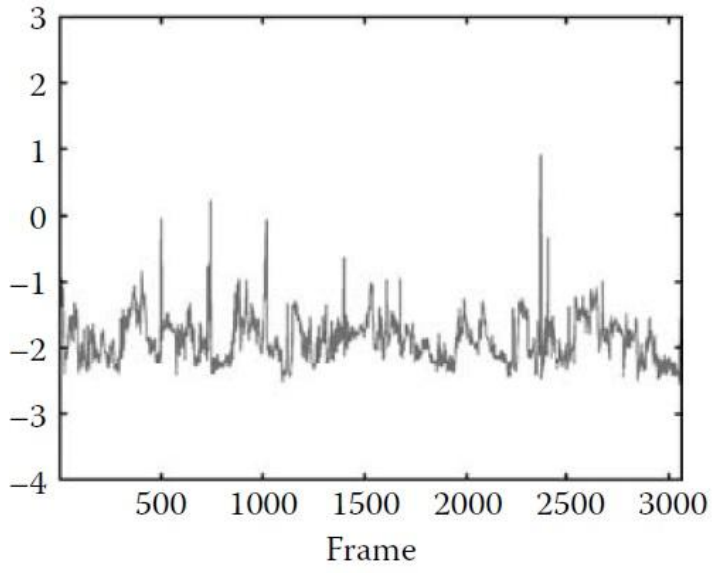
3.1.4 Φασματικά χαρακτηριστικά

Αυτά τα χαρακτηριστικά υπολογίζονται από το μετασχηματισμό STFT του ηχητικού σήματος[Peeters, 2004]. Για αυτά, χρησιμοποιείται το λογισμικό Marsyas και εξάγει τα εξής timbral texture χαρακτηριστικά: *spectral centroid, spectral rolloff, spectral flux, spectral flatness measures (SFM),* και *spectral*

crest factors (SCF)[Tzanetakis & Cook, 2002]. Αυτά τα χαρακτηριστικά εξάγονται σε κάθε καρέ και υπολογίζοντας την μέση τιμή και τυπική απόκλιση για κάθε δευτερόλεπτο. Το spectral centroid είναι το κέντρο βάρους του φασματικού πλάτους του μετασχηματισμού STFT. Το κέντρο βάρους είναι η μέτρηση της φασματικής μορφής. Το spectral centroid υπολογίζεται από την παρακάτω σχέση:

$$\text{spectral centroid} = \frac{\sum_{n=1}^{N} nA_t^n}{\sum_{n=1}^{N} A_t^n},$$

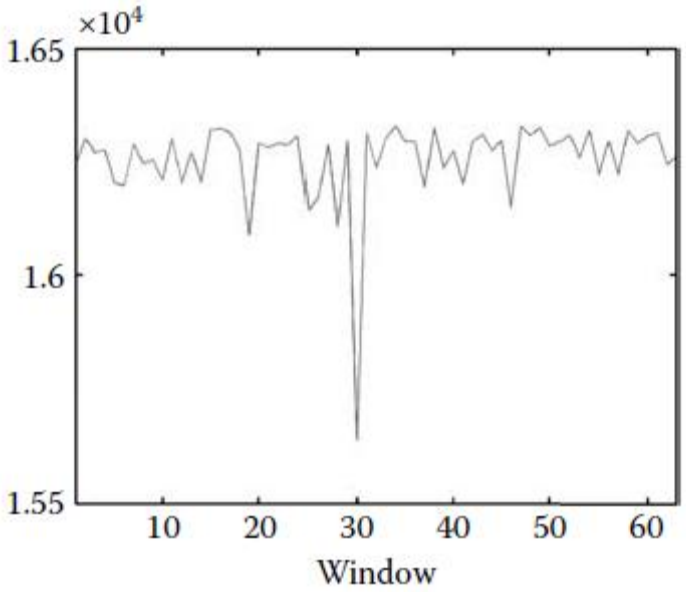
όπου A_t^n είναι πλάτος του φάσματος των t καρέ και n η συχνότητα των bin, και N ο συνολικός αριθμός των bins. Η εικόνα 3 είναι ένα παράδειγμα.



spectral centroid

Το spectral rolloff είναι η ποσότητα της υψηλής συχνότητας στο σήμα και είναι και αυτό μέτρηση της φασματικής μορφής. Το spectral rolloff ορίζεται ως η συχνότητα k_t παρακάτω η οποία περιέχει ένα ορισμένο κλάσμα της συνολικής ενέργειας. Η αναλογία είναι 0,85. Η εικόνα 4 είναι ένα παράδειγμα.

$$\sum_{n=1}^{N} A_t^n = 0.85 * \sum_{n=1}^{N} A_t^n$$



spectral rolloff

Το spectral flux εκτιμάει την ποσότητα της γενικής φασματικής αλλαγής. Το spectral flux ορίζεται ως ο κύκλος της διαφοράς ανάμεσα στα ισορροπημένα πλάτη των επιτυχημένων καρέ και δίνεται από την παρακάτω σχέση:

$$\text{spectral flux} = \sum_{n=1}^{N} \left(a_t^n - a_{t-1}^n \right),$$

όπου a δηλώνει το ισορροπημένο πλάτος του φάσματος (ισορροπημένο για κάθε καρέ).

Οι spectral flatness measures και οι spectral crest factors σχετίζονται με την τονικότητα του ηχητικού σήματος[Allamanche, Herre, Helmuth, Froba, Kasten & Cremer, 2001]. Η τονικότητα έχει να κάνει με το valence. Για παράδειγμα χαρούμενες και γαλήνιες μελωδίες είναι τονικές και οι μελωδίες που δείχνουν θυμό είναι άτονες[Thompson & Robitaille, 1992]. Οι spectral flatness measures (SFM) είναι η αναλογία ανάμεσα στο γεωμετρικό μέσο της φασματικής δύναμης και στο αριθμητικό μέσο. Τα spectral flatness measures υπολογίζονται από την σχέση:

$$\text{spectral flatness measures}=\frac{\left(\prod_{n\in B_k} A_t^n\right)^{1/N_k}}{\frac{1}{N_k}\sum_{n\in B_k} A_t^n},$$

όπου A_t^n είναι πλάτος του φάσματος των *t* καρέ, το B^k δηλώνει τις *k* υποζώνες συχνότητας, και N_k είναι ο αριθμός των bins στο B^k.

Αντίθετα, οι spectral crest factors (SCF) είναι η αναλογία ανάμεσα στην κορυφή πλάτους και την ρίζα του μέσου τετραγωνικού πλάτους.

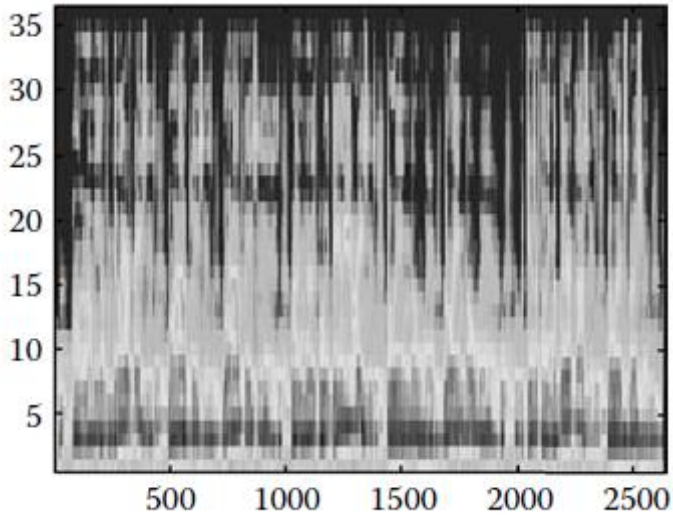
$$\text{spectral crest factors}=\frac{\max_{n\in B^k} A_t^n}{\frac{1}{N_k}\sum_{n=1}^{N} A_t^n},$$

όπου το B^k δηλώνει τις *k* υποζώνες συχνότητας, και N_k είναι ο αριθμός των bins στο B^k.

Στο λογισμικό Marsyas χρησιμοποιούνται 24 υποζώνες. Μια από αυτές χρησιμοποιεί την εργαλειοθήκη μουσικής ανάλυσης (MA toolbox)[Pampalk, 2002] για να εξάγει τους *Mel-frequency cepstral coefficients (MFCCs)*. Οι MFCC είναι οι συντελεστές του μετασχηματισμού διακριτού συνημίτονου (discrete cosine transform) κάθε βραχυπρόθεσμου λογάριθμου του φάσματος ισχύος που εκφράζεται σε μια κλίμακα Mel-frequency μη γραμμικής αντίληψης για να αναπαραστήσει μορφότυπες κορυφές του φάσματος[Davis & Mermelstein, 1980]. Η κανονική συχνότητα f hertz μπορεί να μετατραπεί στο εύρος mel και δίνεται από τον παρακάτω τύπο:

$$m = 1127.01048 \log (1 + f / 700)$$

Οι MFCC φαίνονται από την παρακάτω εικόνα:



MFCC

Λαμβάνεται έτσι, η μέση και τυπική απόκλιση των πρώτων 13 ή 20 MFCC του κάθε καρέ. Ο MFCC είναι ευρέως διαδεδομένος στην επεξεργασία σήματος ομιλίας και στην έρευνα Music Information Retrieval[Casey, Veltkamp, Goto, Leman, Rhodes & Slaney, 2008]. Παρόλα αυτά ο συντελεστής MFCC χάνει την σχετική φασματική πληροφορία[Lu, Liu & Zhang, 2006]. Για αυτό το λόγο προτείνεται να βρεθεί η σχετική διανομή ενέργειας στα αρμονικά συστατικά του φάσματος. Έτσι η φασματική κορυφή (spectral peak), η φασματική κοιλάδα (spectral valley) και οι δυναμικές τους σε κάθε υποζώνη αντανακλούν την σχετική κατανομή των αρμονικών και μη αρμονικών συστατικών στο φάσμα. Αυτό δείχνει την υπεροχή των φασματικών χαρακτηριστικών αντίθεσης (spectral contrast features)[Jiang, Lu, Zhang, Tao & Cai, 2002] έναντι του MFCC για την μουσική ταξινόμηση. Άλλο ένα φασματικό χαρακτηριστικό που μπορεί να χρησιμοποιηθεί στο παρόν σύστημα είναι το ***Daubechies wavelets coefficient histogram (DWCH)***[Li & Ogihara, 2003; Li & Ogihara, 2004; Li & Ogihara, 2006]. Αυτό, υπολογίζεται σε διαφορετικές υποζώνες συχνότητας με διαφορετικές αναλύσεις. Τόσο τα χαρακτηριστικά αντίθεσης όσο και το DWCH εφαρμόζονται στο MATLAB[http://www.mathworks.com/products/wavelet/]. Μπορεί επιπλέον να χρησιμοποιηθεί η εργαλειοθήκη μουσικής ανάκτησης πληροφορίας (MIRtoolbox)[Lartillot & Toiviainen, 2007] για να παραχθούν τρία αισθητήρια χαρακτηριστικά, *roughness, irregularity, inharmonicity*. Το πρώτο μετράει τον θόρυβο του φάσματος[Sethares,1998]. Το δεύτερο μετράει το βαθμό παραλλαγής των διαδοχικών κορυφών[Fujinaga & McMillan, 2000; Jensen, 1999]. Το τρίτο αναπαριστά την απόκλιση των φασματικών συστατικών του σήματος από ένα καθαρά αρμονικό σήμα[Osgood, Suci & Tannenbaum, 1957]. Οι ψυχολογικές μελέτες δείχνουν ότι το valence

σχετίζεται με το αρμονικό περιεχόμενο των μουσικών σημάτων. Τέλος, άλλα τρία χαρακτηριστικά που χρησιμοποιούνται είναι τα *tristimulus, even-harm,* και *odd-harm*[Zhu, Shi, Kim & Eom, 2006; Wieczorkowska, Synak, Lewis & Ras, 2005; Wieczorkowska, Synak & Ras2006]. Το πρώτο είναι μια παράμετρος ενώ τα άλλα δυο αναπαριστούν τις άρτιες και περιττές αρμονικές του φάσματος.

3.1.5 Αρμονικά χαρακτηριστικά

Τα αρμονικά χαρακτηριστικά είναι χαρακτηριστικά που υπολογίζονται από την ημιτονοειδή αρμονική μοντελοποίηση του σήματος[Peeters,2004]. Για αυτά, χρησιμοποιείται η εργαλειοθήκη μουσικής ανάκτησης πληροφορίας (MIR toolbox)[Lartillot & Toiviainen, 2007] και εξάγονται δυο χαρακτηριστικά τόνου, *salient pitch, chromagram center* και τρία χαρακτηριστικά τονικότητας, *key clarity, mode, harmonic change*. Η MIRtoolbox εκτιμά το *pitch*, ή την *perceived fundamental frequency* του κάθε βραχυπρόθεσμου καρέ (50 ms, 1/2 επικάλυψη) βασισμένη στον αλγόριθμο multi-pitch detection των Tolonen και Karjalainen[Tolonen & Karjalainen, 2000]. Επιπλέον, υπολογίζει το *wrapped chromagram*, ή το *pitch class profile*, για κάθε καρέ και χρησιμοποιεί το κέντρο βάρους του χρωμογράμματος (chromagram centroid) ως άλλη μια εκτίμηση της θεμελιώδους συχνότητας. Το wrapped chromagram σχεδιάζει το φάσμα συχνότητας σε 12 bins εκπροσωπώντας ο καθένας 12 διαφορετικά ημιτόνια (χρώματα) της μουσικής οκτάβας. Κάθε bin αντιστοιχεί σε μια από τις 12 τάξεις ημιτόνου σε μια κλίμακα δυτικού δωδεκάτονου ίσης ιδιοσυγκρασίας. Επόμενο βήμα, είναι η σύγκριση ενός χρωμογράμματος σε 24 μικρότερα και μέγιστα προφίλ κλειδιών[Gomez, 2006] για να εκτελεστεί ο εντοπισμός κλειδιού και να εκτιμηθεί η δύναμη του καρέ σε συνδυασμό με κάθε κλειδί. Η δύναμη που σχετίζεται με το καλύτερο κλειδί, που είναι αυτό με την υψηλότερη δύναμη, επιστρέφεται ως *key clarity*. Η διαφορά ανάμεσα στο μέγιστο και στο μικρότερο κλειδί είναι η δύναμη του κλειδιού που επιστρέφεται ως μια εκτίμηση της μουσικής λειτουργίας, η οποία περιγράφει μια καθορισμένη διάταξη των διατονικών τόνων μιας οκτάβας[Oliveira & Cardoso, 2008]. Η λειτουργία σχετίζεται με το valence[Gabrielsson & Lindstrom, 2001; Oliveira & Cardoso, 2009]. Ωστόσο δεν είναι σίγουρο το αν οι τιμές της μουσικής λειτουργίας συσχετίζονται με το valence. Η εργαλειοθήκη MIR χρησιμοποιεί έναν αλγόριθμο[Harte, Sandler & Gasser, 2006] για να υπολογίσει ένα διάνυσμα χαρακτηριστικού έξι διαστάσεων που ονομάζεται *tonal centroid* από το χρωμόγραμμα και εντοπίζει αρμονικές αλλαγές σε μουσικό ήχο. Η υψηλή αρμονική αλλαγή δείχνει μεγάλη διαφορά στο αρμονικό περιεχόμενο ανάμεσα

σε συνεχή καρέ. Τα βραχυπρόθεσμα χαρακτηριστικά συγκεντρώνονται λαμβάνοντας την μέση και τυπική απόκλιση[Meng, Ahrendt, Larsen & Hansen, 2007]. Η εργαλειοθήκη Marsyas υπολογίζει το *pitch histogram* και εξάγει τα εξής χαρακτηριστικά: *tonic, main pitch class, octave range* του *dominant pitch, main tonal interval relation,* και *overall pitch strength*[Tzanetakis & Cook, 2002]. Τέλος, το PsySound συγκρίνει 16 χαρακτηριστικά τόνου, όπως *mean, standard deviation, skewness*, *kurtosis* του τόνου και *pitch strength time series* που εκτιμούνται από το *SWIPE (sawtooth waveform inspired pitch estimator)* και το *SWIPE'*. Το τελευταίο περιορίζει την ανάλυση των πρώτων και κύριων αρμονικών χαρακτηριστικών[Cabrera, 1999; Camacho, 2007].

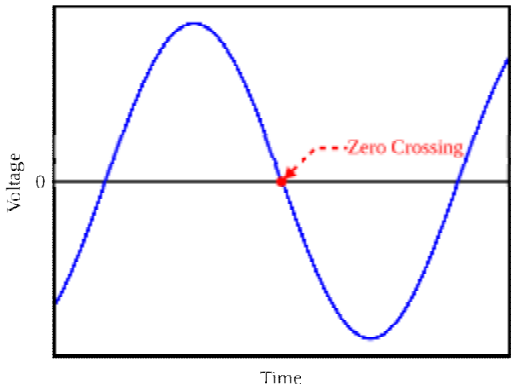
Στην παρούσα εργασία χρησιμοποιείται μόνο το λογισμικό Marsyas[Tzanetakis & Cook, 2002] για να εξάγει τα εξής χαρακτηριστικά: Beat Histogram, Linear Prediction Cepstral Coefficients(LPCC), LSP(Line Spectral Pair), Mel-Frequency Cepstral Coefficients(MFCC), Spectral Flatness Measures(SFM), Spectral Crest Factors(SCF), Zero Crossing, Spectral Centroid, Spectral Rolloff, Spectral Flux. Τα LPCC, LSP και zero crossings δεν συμπεριλαμβάνονται στις παραπάνω κατηγορίες. Οι linear prediction cepstrum coefficients (LPCC) είναι και αυτοί χαρακτηριστικά που περιγράφουν την μουσική. Χρησιμοποιούνται στην αναγνώριση ομιλίας ως μια εκτίμηση του speech vocal tract filter. Συνήθως χρησιμοποιούνται στα μουσικά σήματα. Ο τύπος είναι:

$$c_q = \begin{cases} ln(Q) & : q = 0 \\ -b_q + \frac{1}{m}\sum_{q=1}^{Q} -(m-q)b_q c_{m-q} & : q > 0 \end{cases}$$

Οι line spectral pair (LSP), χρησιμοποιούνται για να αναπαραστήσουν τους linear prediction coefficients για την μετάδοση πέρα του ενός καναλιού. Είναι πολύ χρήσιμοι στην κωδικοποίηση ομιλίας. Ο τύπος είναι:

$$A(z) = 0.5[P(z) + Q(z)], \text{ όπου}$$
$$P(z) = A(z) + z^{-(p+1)}A(z^{-1})$$
$$Q(z) = A(z) - z^{-(p+1)}A(z^{-1})$$

Τα zero crossings μετρούν το θόρυβο του σήματος στο σημείο 0. Ένα παράδειγμα είναι η εικόνα 6.



Zero crossing

Κεφάλαιο 4 – Πειραματική Διαδικασία

4.1 Μεθοδολογία

4.1.1 Εισαγωγή

Στην έρευνα μας χρησιμοποιούμε το Marsyas[Tzanetakis & Cook, 2002]. Το Marsyas είναι ένα δωρεάν λογισμικό δημιουργημένο από τον Γ. Τζανετάκη. Χρησιμοποιείται κυρίως για την γρήγορη ανάπτυξη και στην συνέχεια αξιολόγηση εφαρμογών ήχου στον υπολογιστή.

4.1.2 Απαραίτητες ενέργειες

Από το εγχειρίδιο του Marsyas που βρίσκουμε στο blog εγκαθιστούμε μια σειρά από προγράμματα απαραίτητα για να λειτουργήσει το λογισμικό. Αυτά είναι: Cmake, TortoiseSVN, Visual Studio, DirectX SDK και το συμπιεσμένο αρχείο Libmad.

Cmake: είναι μια δωρεάν πλατφόρμα προγράμματος λογισμικού. Από αυτό κατασκευάζεται ένα λογισμικό με την μέθοδο ανεξαρτήτου μεταγλωττιστή.

TortoiseSVN: είναι ένα δωρεάν λογισμικό για προγραμματιστές λογισμικού.

DirectX SDK: είναι μια συλλογή από προγραμματιστικά περιβάλλοντα σχετικά με πολυμέσα, κυρίως παιχνίδια και βίντεο.

Visual Studio: είναι ένα ολοκληρωμένο περιβάλλον ανάπτυξης για την ανάπτυξη προγραμμάτων στον υπολογιστή για Microsoft Windows. Η έκδοση που χρησιμοποιούμε είναι η Visual Studio 12.

Libmad: τα αρχικά mad σημαίνουν MPEG Audio Decoder. Το Libmad ουσιαστικά είναι μια βιβλιοθήκη που αποκωδικοποιεί αρχεία που έχουν κωδικοποιηθεί με ένα MPEG Audio codec. Αποτελείται από το libmad, το software library, και το madplay ένα πρόγραμμα γραμμής εντολών για την αναπαραγωγή mp3. Η έκδοση που χρησιμοποιούμε είναι η libmad-1.15.1b.

4.1.3 Εγκατάσταση απαραίτητων προγραμμάτων

Αρχικό μας βήμα είναι η δημιουργία ενός φακέλου στον τοπικό δίσκο του υπολογιστή μας που θα αποθηκεύονται τα αρχεία μας. Στη δεδομένη περίπτωση ο φάκελός μας ονομάζεται Research και περιέχει μέσα έναν υποφάκελο με όνομα Marsyas, εκεί όπου αποθηκεύονται τα δεδομένα μας.

Αφού εγκαταστήσουμε τα παραπάνω προγράμματα ξεκινάμε

ανοίγοντας το Cmake. Εδώ, επιλέγουμε την έκδοση Visual Studio που έχουμε ‘κατεβάσει’, στην προκειμένη περίπτωση του 12. Στις πιθανές επιλογές επιλέγουμε Visual Studio 11 όπως μας έχει υποδείξει το εγχειρίδιο του Marsyas. Τα δεδομένα στο παράθυρο του προγράμματος εμφανίζονται κοκκινισμένα. Πατάμε το κουμπί configure και το ‘κοκκίνισμα’ φεύγει. Στη συνέχεια αφήνουμε ως έχουν τα δεδομένα που είναι επιλεγμένα και επιλέγουμε το WITH MAD. Εδώ μας χρησιμεύει το συμπιεσμένο αρχείο libmad-1.15.1b. Μετά την επιλογή WITH MAD εμφανίζονται δυο δεδομένα με όνομα mad include_DIR και mad library αντίστοιχα. Στο πρώτο ορίζουμε το path του φακέλου libmad-1.15.1b που βρίσκεται στο φάκελο του Marsyas. Στο δεύτερο ορίζουμε το path του φακέλου libmad-1.15.1b που βρίσκεται στις λήψεις, τον υποφάκελο msvc++, τον υποφάκελο debug και τέλος βρίσκουμε το αρχείο libmad.lib. Στη συνέχεια πατάμε το κουμπί generate. Αφού έχει ‘χτιστεί’ το λογισμικό μας μέσα στο φάκελο Marsyas έχει δημιουργηθεί ένας φάκελος με όνομα build.

Στη συνέχεια, μέσα στο φάκελο βρίσκουμε το αρχείο ALLBUILD.vcxproj, το οποίο αναγνωρίζει το πρόγραμμα Visual Studio 12. Το αρχείο αυτό περιέχει το project όλου του κώδικα του προγράμματος Marsyas, το οποίο θα πρέπει να γίνει compile. Η διαδικασία του compile θα είναι έγκυρη όταν δεν εμφανιστεί κανένα σφάλμα, παρά μόνο κάποιες (πιθανόν) προειδοποιήσεις.

Τα αποτελέσματα της εξαγωγής παραμέτρων εμφανίζονται στο πρόγραμμα που ονομάζεται WEKA[Panda, Malheiro, Rocha, Oliveira & Paiva, 2008; Percival & Τζανετακης Marsyas user manual for version 0.3]. Το WEKA(Waikato Environment Knowledge Analysis) είναι μια δημοφιλής σουίτα λογισμικού μάθησης μηχανής γραμμένο σε γλώσσα προγραμματισμού Java που αναπτύχθηκε στο πανεπιστήμιο του Waikato της Νέας Ζηλανδίας. Είναι ένα δωρεάν λογισμικό υπό την άδεια GNU General Public License. Βασική προϋπόθεση χρήσης του WEKA είναι να υπάρχει ένα σύνολο από αρχεία μουσικής πάνω στο οποίο θα γίνονται οι δοκιμές. Όπως προαναφέρθηκε στη συλλογή δεδομένων έχουμε δημιουργήσει έναν φάκελο με όνομα music.

Αρχικά, ανοίγουμε μια γραμμή εντολών. Στη συνέχεια, ορίζουμε το path που βρίσκεται ο φάκελος music με τοποθετημένα τα μουσικά αρχεία και δημιουργούμε ένα αρχείο music.mf το οποίο περιέχει τη λίστα με τα μουσικά κομμάτια και ανοίγει σε Notepad. Στη συνέχεια χρησιμοποιούμε το πρόγραμμα bextract[Percival & Τζανετακης Marsyas user manual for version 0.3]. Εδώ, ουσιαστικά ζητάμε από το πρόγραμμα Marsyas να εξάγει όλες τις παραμέτρους που περιγράφηκαν παραπάνω για κάθε ένα μουσικό κομμάτι. Σε περίπτωση που θέλουμε να εξάγουμε κάποιο συγκεκριμένο χαρακτηριστικό μετά την παράμετρο bextract εισάγουμε το όνομα του χαρακτηριστικού. Για παράδειγμα

αν θέλουμε να εξάγουμε τα τονικά χαρακτηριστικά πληκτρολογούμε bextract –timbral all.mf –w ms1.arff ή για τα φασματικά bextract –spfe all.mf –w ms2.arff. Με το ίδιο τρόπο εξάγεται και κάθε χαρακτηριστικό ξεχωριστά. Αν για παράδειγμα θέλουμε να εξάγουμε μόνο τους MFCC πληκτρολογούμε bextract –mfcc all.mf –w mfcc.arff. Ομοίως και για τα υπόλοιπα. Σε περίπτωση που θέλουμε τα χαρακτηριστικά να εμφανίζονται σε ένα μόνο συγκεντρωτικό διάνυσμα ανεξαρτήτου του μήκους του μουσικού κομματιού, χρησιμοποιούμε την παράμετρο –sv δηλαδή πληκτρολογούμε bextract –sv all.mf –w ms.arff. στην παραπάνω περίπτωση εξάγαμε όλες τις παραμέτρους. Με τον ίδιο τρόπο εξάγουμε και κάθε μία ξεχωριστά. Ουσιαστικά, πληκτρολογούμε τα χαρακτηριστικά που θέλουμε να εξάγουμε, το αρχείο music.mf και διπλά το όνομα αρχείου που θα αποθηκευτούν. Αυτό το αρχείο έχει format .arff, αρχείο το οποίο διαβάζει το WEKA.

Για το κομμάτι της αναγνώρισης, ανοίγουμε το αρχείο με όνομα ms.arff με το WEKA. Εδώ, είναι απαραίτητες κάποιες ρυθμίσεις. Από την καρτέλα classify, επιλέγουμε τον φάκελο functions και στη συνέχεια την LibSVM. Η LibSVM είναι μια βιβλιοθήκη η οποία είναι απαραίτητη για την εκτέλεση της ταξινόμησης, χρησιμοποιώντας Support Vector Machines. Εδώ, για την εκπαίδευση μοντέλων ταξινόμησης χρησιμοποιείται ο ταξινομητής SVM που είναι αρκετά αξιόπιστος. Στη συνέχεια, κάνουμε ‘κλικ’ στην επιλογή cross-validation και βάζουμε 10 επαναλήψεις (default). Έπειτα, πατάμε start. Έχουμε την επιλογή stop σε περίπτωση που θελήσουμε να σταματήσουμε την διαδικασία. Όταν τελειώσει η διαδικασία τα αποτελέσματα της ταξινόμησης εμφανίζονται δεξιά του παραθύρου.

4.2 Συλλογή δεδομένων

Δημιουργούμε έναν υποφάκελο με όνομα test μέσα στο φάκελο Research που έχουμε ήδη δημιουργήσει. Στο φάκελο test δημιουργούμε έναν υποφάκελο με όνομα music όπου αποθηκεύουμε τα αρχεία μουσικής για δοκιμή. Τα αρχεία μουσικής που έχουμε συλλέξει είναι 903, από διάφορα είδη μουσικής, επιλέγοντας 30 δευτερόλεπτα αντιπροσωπευτικά για κάθε κομμάτι [Panda, Malheiro, Rocha, Oliveira & Paiva, 2008; Panda & Paiva DAF 2012; Panda & Paiva MML 2012; Panda & Paiva, CISUC 2011; Panda & Paiva, AES 2011; Cardoso, Panda & Paiva, INForum 2011; Panda & Paiva, CISUC 2012; Hu, Downie, Laurier, Bay & Ehmann, 2008; Panda, Rocha & Paiva, CISUC]. Τα κομμάτια αυτά έχουν επιλεχθεί από το MIREX, το οποίο τα έχει χωρίσει σε κατηγορίες με βάση τους συναισθηματικούς σχολιασμούς εθελοντών. Κάθε τραγούδι μπορεί να έχει παραπάνω από μία ετικέτα. Οι ετικέτες κάθε τραγουδιού ομαδοποιούνται από μια κατηγορία και ο

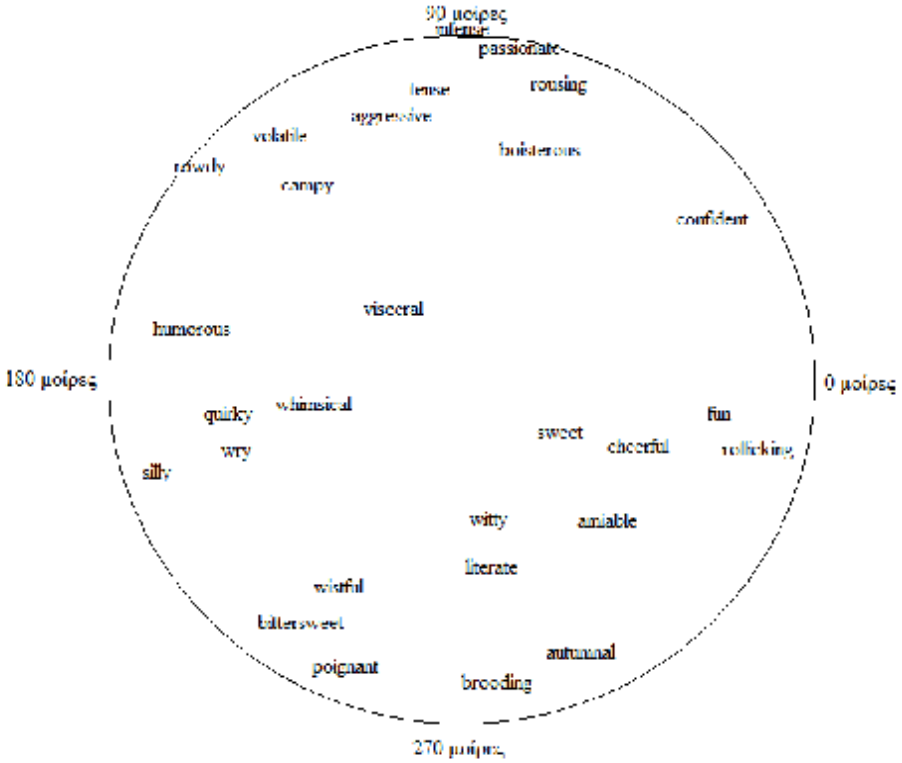
συναισθηματικός σχολιασμός του κάθε τραγουδιού βασίζεται στην πιο σημαντική κατηγορία, του τραγουδιού με τις περισσότερες ετικέτες (για παράδειγμα ένα τραγούδι με μια ετικέτα στην κατηγορία 1 και τρεις στην κατηγορία 5, ταξινομείται στην κατηγορία 5. Όλα τα τραγούδια έχουν μοιραστεί ισορροπημένα στις 5 κατηγορίες. 18.8% στην κατηγορία 1, 18.2% στην κατηγορία 2, 23.8% στην κατηγορία 3, 21.2% στην κατηγορία 4 and 18.1% στην κατηγορία 5. Η βάση δεδομένων που χρησιμοποιήθηκε στα πειράματά μας βρίσκεται στο διαδίκτυο (www.mir.dei.uc.pt).

4.3 Ταξινόμηση συναισθημάτων

Η προσέγγιση μας είναι στην αυτόματη αναγνώριση συναισθήματος από μουσικά τραγούδια είναι κατηγορηματική και χρησιμοποιήσαμε τις εξής 5 κατηγορίες[Panda, Malheiro, Rocha, Oliveira & Paiva, 2008; Panda & Paiva DAF 2012; Panda & Paiva MML 2012; Panda & Paiva, CISUC 2011; Panda & Paiva, AES 2011; Cardoso, Panda & Paiva, INForum 2011; Panda & Paiva, CISUC 2012; Hu, Downie, Laurier, Bay & Ehmann, 2008; Panda, Rocha & Paiva, CISUC]:

1 passionate, rousing, confident, boisterous, rowdy
2 rollicking, cheerful, fun, sweet, amiable/good natured
3 literate, poignant, wistful, bittersweet, autumnal, brooding
4 humorous, silly, campy, quirky, whimsical, witty, wry
5 aggressive, fiery, tense/anxious, intense, volatile, visceral

Η εικόνα που ακολουθεί είναι μια απεικόνιση των παραπάνω κατηγοριών σε συντεταγμένες (κατά προσέγγιση).



Ταξινόμηση των 5 κατηγοριών σε συντεταγμένες

4.4 Εξαγωγή χαρακτηριστικών

Στην παρούσα εργασία χρησιμοποιήθηκε το λογισμικό Marsyas[90] για να εξάγει τα εξής χαρακτηριστικά (όπως αναφέρθηκαν και στο κεφάλαιο 2): Beat Histogram, Linear Prediction Cepstral Coefficients(LPCC), LSP(Line Spectral Pair), Mel-Frequency Cepstral Coefficients(MFCC), Spectral Flatness Measures(SFM), Spectral Crest Factors(SCF), Zero Crossing, Spectral Centroid, Spectral Rolloff, Spectral Flux καθώς και οι συνδυασμοί τους..

Κεφάλαιο 5 - Πειράματα

5.1 Πειράματα

Στα πειράματα μας, χρησιμοποιούμε μια έτοιμη βάση δεδομένων με 903 τραγούδια από διάφορα είδη μουσικής, με επιλεγμένα 30 δευτερόλεπτα αντιπροσωπευτικά για κάθε τραγούδι, για την εκπαίδευση μοντέλου. Τα κομμάτια αυτά έχουν ταξινομηθεί σε πέντε κατηγορίες όπως προαναφέρθηκαν στο κεφάλαιο 4. Η βάση δεδομένων με τα τραγούδια έχει ληφθεί από το διαδίκτυο (www.mir.dei.uc.pt) και είναι η ίδια που χρησιμοποιήθηκε στο MIREX (music information retrieval evaluation exchange) [Panda, Malheiro, Rocha, Oliveira & Paiva, 2008; Panda & Paiva DAF 2012; Panda & Paiva MML 2012; Panda & Paiva, CISUC 2011; Panda & Paiva, AES 2011; Cardoso, Panda & Paiva, INForum 2011; Panda & Paiva, CISUC 2012; Hu, Downie, Laurier, Bay & Ehmann, 2008; Panda, Rocha & Paiva, CISUC]. Τα χαρακτηριστικά που εξάγονται από το Marsyas[Tzanetakis & Cook, 2002] όπως προαναφέρθηκε είναι τα εξής: beat histogram, LPCC, LSP, MFCC, SFM, SCF, Zero Crossing, Spectral Centroid, Spectral Rolloff, Spectral Flux. Για την χρήση των παραπάνω χαρακτηριστικών συνδυάζουμε τις εξής παραμέτρους: -n (normalize), -ws (window size), -hp (hop size). Το normalize χρησιμοποιείται για να ισορροπεί το πλάτος του ηχητικού σήματος. Το window size είναι το μέγεθος σε δείγματα της ανάλυσης παραθύρου. Η αρχική του τιμή είναι 512 δείγματα. Το hop size είναι το μέγεθος σε δείγματα του μεγέθους ανάλυσης μετάβασης παραθύρου. Η αρχική του τιμή είναι 512, δηλαδή χωρίς επικάλυψη (χρησιμοποιώντας παράθυρο ανάλυσης επίσης 512 σημείων). Τα window size και hop size συνδυάζονται μεταξύ τους. Το hop size πρέπει να είναι το μισό του window size, για καλύτερο αποτέλεσμα.

Δοκιμάσαμε διαφορετικές τιμές των παραπάνω παραμέτρων για κάθε σετ χαρακτηριστικών ξεχωριστά, για να βρούμε ποιος συνδυασμός τιμών αποδίδει καλύτερο ποσοστό στην αναγνώριση συναισθημάτων στα μουσικά κομμάτια. Τα πειράματα διεξάγονται σε δύο φάσεις. Στην πρώτη φάση, εξετάζουμε τις 5 κατηγορίες συναισθημάτων. Στην δεύτερη φάση, συγχωνεύουμε την κατηγορία 1 με την 5 και την 2 με την 4. Αυτό γίνεται γιατί οι κατηγορίες 1 με 5 και 2 με 4 έχουν ομοιότητες στα συναισθήματα και επαναλαμβάνονται όλα τα σετ πειραμάτων.

5.1.1 Πρώτη φάση

Σε αυτό το πείραμα χρησιμοποιούμε πέντε κατηγορίες και αφήνοντας τις τιμές των window size και hop size ως default(512). Στην άλλη περίπτωση κρατάμε το window size στην αρχική του τιμή και το hop size το μειώνουμε στο μισό(256). Τα αποτελέσματα για κάθε χαρακτηριστικό της μουσικής που εξάγεται από το Marsyas με τις παραπάνω παραμέτρους είναι τα εξής:

Beat Histogram

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 23.8649% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω (confusion matrix):

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	340	0	0
Cl2	0	**0**	328	0	0
Cl3	0	0	**431**	0	0
Cl4	0	0	382	**0**	0
Cl5	0	0	325	0	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 38.7597% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**48**	13	33	32	44
Cl2	26	**21**	47	46	24
Cl3	15	8	**152**	23	17
Cl4	27	13	63	**49**	39
Cl5	23	4	29	27	**80**

Linear Prediction Cepstral Coefficients(LPCC)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 19.7674% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**34**	34	0	170	102
Cl2	33	**33**	0	164	98
Cl3	43	43	**0**	216	129
Cl4	38	38	0	**191**	115
Cl5	33	32	0	161	**99**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 23.8095% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	170	0	0
Cl2	0	**0**	164	0	0
Cl3	0	0	**215**	0	0
Cl4	0	0	191	**0**	0
Cl5	0	0	163	0	**0**

Line Spectral Pair(LSP)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 23.8649% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	340	0	0
Cl2	0	**0**	328	0	0
Cl3	0	0	**431**	0	0
Cl4	0	0	382	**0**	0
Cl5	0	0	325	0	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 23.8095% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	170	0	0
Cl2	0	**0**	164	0	0
Cl3	0	0	**215**	0	0
Cl4	0	0	191	**0**	0
Cl5	0	0	163	0	**0**

Mel-Frequency Cepstral Coefficients(MFCC)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 19.7674% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**34**	34	0	170	102
Cl2	33	**33**	0	164	98
Cl3	43	43	**0**	216	129

Cl4	38	38	0	**191**	115
Cl5	33	32	0	161	**99**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 37.5415% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**43**	3	39	40	45
Cl2	34	**6**	45	60	19
Cl3	14	0	**161**	28	12
Cl4	34	3	67	**51**	36
Cl5	25	2	29	29	**78**

Spectral Flatness Measures(SFM) - Spectral Crest Factors(SCF)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 23.8649% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	340	0	0
Cl2	0	**0**	328	0	0
Cl3	0	0	**431**	0	0
Cl4	0	0	382	**0**	0
Cl5	0	0	325	0	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 29.9003% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	5	56	80	29
Cl2	2	**9**	67	67	19
Cl3	6	3	**127**	66	13
Cl4	5	4	72	**94**	16
Cl5	6	3	40	74	**40**

Spectral Features(STFT)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 23.8649% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	340	0	0
Cl2	0	**0**	328	0	0
Cl3	0	0	**431**	0	0
Cl4	0	0	382	**0**	0
Cl5	0	0	325	0	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 23.8095% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	170	0	0
Cl2	0	**0**	164	0	0

Cl3	0	0	**215**	0	0
Cl4	0	0	191	**0**	0
Cl5	0	0	163	0	**0**

Timbral features

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 23.8649% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	340	0	0
Cl2	0	**0**	328	0	0
Cl3	0	0	**431**	0	0
Cl4	0	0	382	**0**	0
Cl5	0	0	325	0	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 38.7597% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**48**	13	33	32	44
Cl2	26	**21**	47	46	24
Cl3	15	8	**152**	23	17
Cl4	27	13	63	**49**	39
Cl5	23	4	29	27	**80**

Spectral Features(STFT) - Mel-Frequency Cepstral Coefficients(MFCC)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 23.8649% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	340	0	0
Cl2	0	**0**	328	0	0
Cl3	0	0	**431**	0	0
Cl4	0	0	382	**0**	0
Cl5	0	0	325	0	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 37.2093% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**46**	2	42	37	43
Cl2	38	**3**	47	57	19
Cl3	17	0	**159**	26	13
Cl4	39	3	66	**51**	32
Cl5	27	0	30	29	**77**

Όλα τα παραπάνω

Στο τελευταίο κομμάτι της πρώτης φάσης εξάγουμε όλα τα παραπάνω χαρακτηριστικά χρησιμοποιώντας τις ίδιες παραμέτρους. Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και

hop size είναι 23.8649% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	340	0	0
Cl2	0	**0**	328	0	0
Cl3	0	0	**431**	0	0
Cl4	0	0	382	**0**	0
Cl5	0	0	325	0	**0**

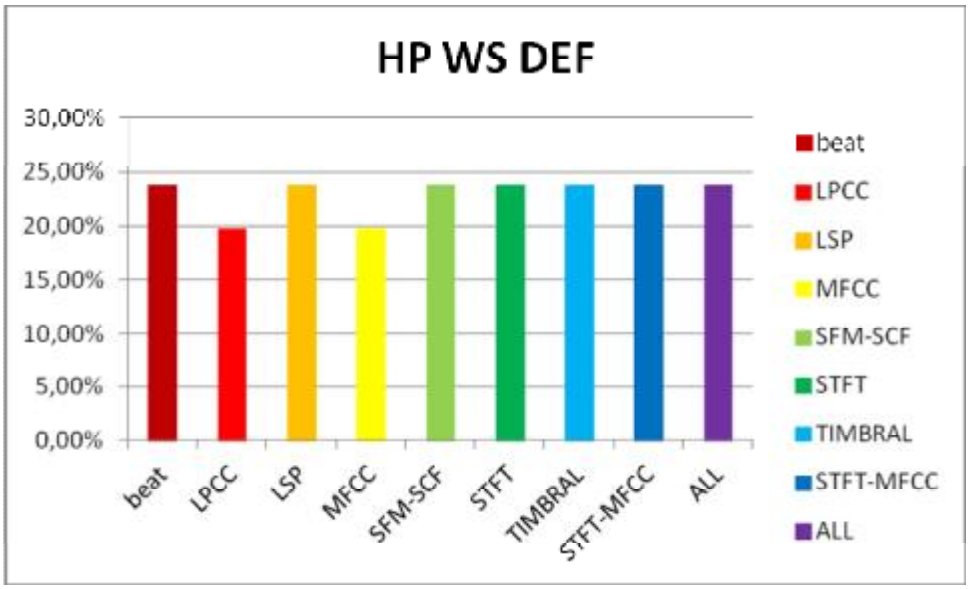
Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 23.8095% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3	Cl4	Cl5
Cl1	**0**	0	170	0	0
Cl2	0	**0**	164	0	0
Cl3	0	0	**215**	0	0
Cl4	0	0	191	**0**	0
Cl5	0	0	163	0	**0**

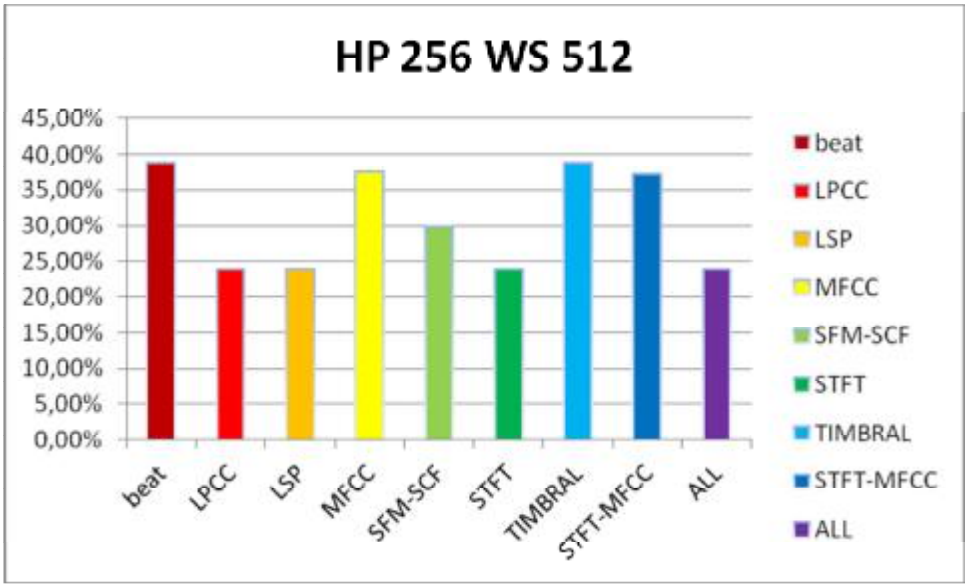
5.1.1.1 Συγκεντρωτικοί πίνακες αποτελεσμάτων

WINDOW SIZE, HOP SIZE DEFAULT(512,512)	
BEAT	23,86%
LPCC	19,76%
LSP	23,86%
MFCC	19,76%
SFM-SCF	23,86%

STFT	23,86%
TIMBRAL	23,86%
STFT-MFCC	23,86%
ALL	23,86%



WINDOW SIZE, HOP SIZE(512,256)	
BEAT	38,75%
LPCC	23,80%
LSP	23,80%
MFCC	37,54%
SFM-SCF	29,90%
STFT	23,80%
TIMBRAL	38,75%
STFT-MFCC	37,20%
ALL	23,80%



5.1.2 Δεύτερη φάση

Σε αυτό το πείραμα χρησιμοποιούμε τρεις κατηγορίες, συγχωνεύοντας την πρώτη κατηγορία με την πέντε και την δυο με την τέσσερα, η τρία μένει ως έχει. Αφήνουμε τις τιμές των window size και hop size ως default(512). Στην άλλη περίπτωση κρατάμε το window size στην αρχική του τιμή και το hop size το μειώνουμε στο μισό(256). Τα αποτελέσματα για κάθε χαρακτηριστικό της μουσικής που εξάγεται από το Marsyas με τις παραπάνω παραμέτρους είναι τα εξής:

Beat Histogram

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 39.3688% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**0**	666	0
Cl2	0	**711**	0
Cl3	0	429	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 53.9313% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**181**	130	22
Cl2	101	**207**	47
Cl3	27	89	**99**

Linear Prediction Cepstral Coefficients(LPCC)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 28.2392% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**67**	134	465
Cl2	71	**142**	498
Cl3	42	86	**301**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 39.3134% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**0**	333	0
Cl2	0	**325**	0
Cl3	0	215	**0**

Line Spectral Pair(LSP)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 39.3688% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**0**	666	0
Cl2	0	**711**	0
Cl3	0	429	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 39.3134% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**0**	333	0
Cl2	0	**355**	0
Cl3	0	215	**0**

Mel-Frequency Cepstral Coefficients(MFCC)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 28.2392% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**67**	134	465
Cl2	71	**142**	498
Cl3	42	86	**301**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 54.2636% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**169**	133	31
Cl2	93	**216**	46
Cl3	21	89	**105**

Spectral Flatness Measures(SFM) - Spectral Crest Factors(SCF)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 39.3688% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**0**	666	0
Cl2	0	**711**	0
Cl3	0	429	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 46.1794% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**190**	136	7
Cl2	131	**220**	4
Cl3	57	151	**7**

Spectral Features(STFT)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 39.3688% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**0**	666	0
Cl2	0	**711**	0
Cl3	0	429	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 39.3134% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**0**	333	0
Cl2	0	**355**	0
Cl3	0	215	**0**

Timbral features

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 39.3688% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**0**	666	0
Cl2	0	**711**	0
Cl3	0	429	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 53.9313% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**181**	130	22
Cl2	101	**207**	47
Cl3	27	89	**99**

Spectral Features(STFT) - Mel-Frequency Cepstral Coefficients(MFCC)

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 39.3688% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**0**	666	0
Cl2	0	**711**	0
Cl3	0	429	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 54.1528% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

	Cl1	Cl2	Cl3
Cl1	**169**	134	30
Cl2	90	**218**	47
Cl3	22	91	**102**

Όλα τα παραπάνω

Στο τελευταίο κομμάτι της δεύτερης φάσης εξάγουμε όλα τα παραπάνω χαρακτηριστικά χρησιμοποιώντας τις ίδιες παραμέτρους. Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις default τιμές των window size και hop size είναι 39.3688% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

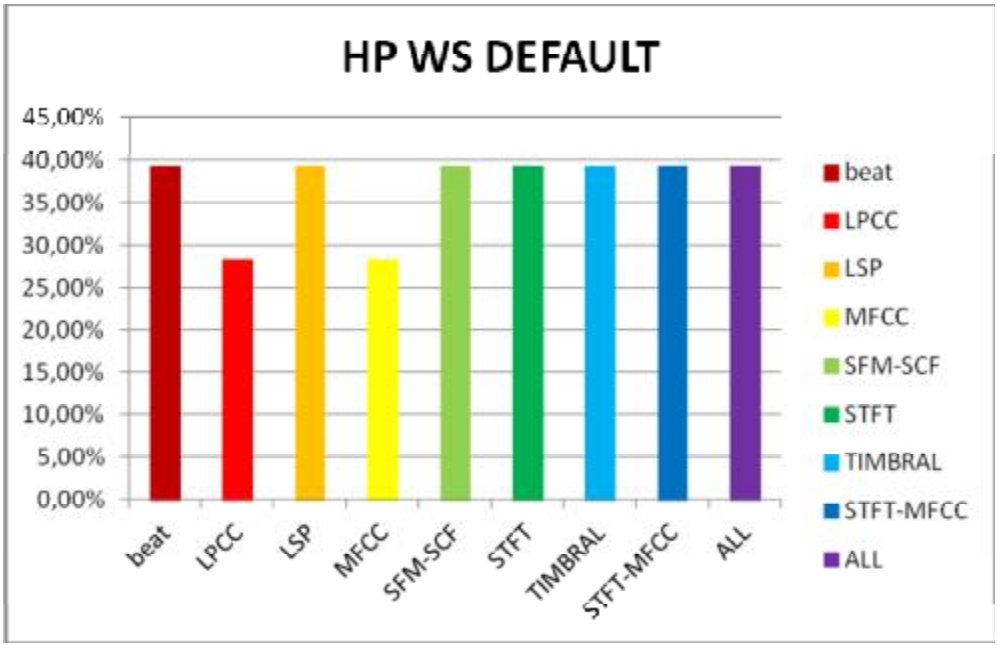
	Cl1	Cl2	Cl3
Cl1	**0**	666	0
Cl2	0	**711**	0
Cl3	0	429	**0**

Το ποσοστό ακρίβειας ταξινόμησης χρησιμοποιώντας τις τιμές 512 για window size και 256 για hop size είναι 39.3134% και ο πίνακας που δείχνει που μπερδεύεται το σύστημα είναι ο παρακάτω:

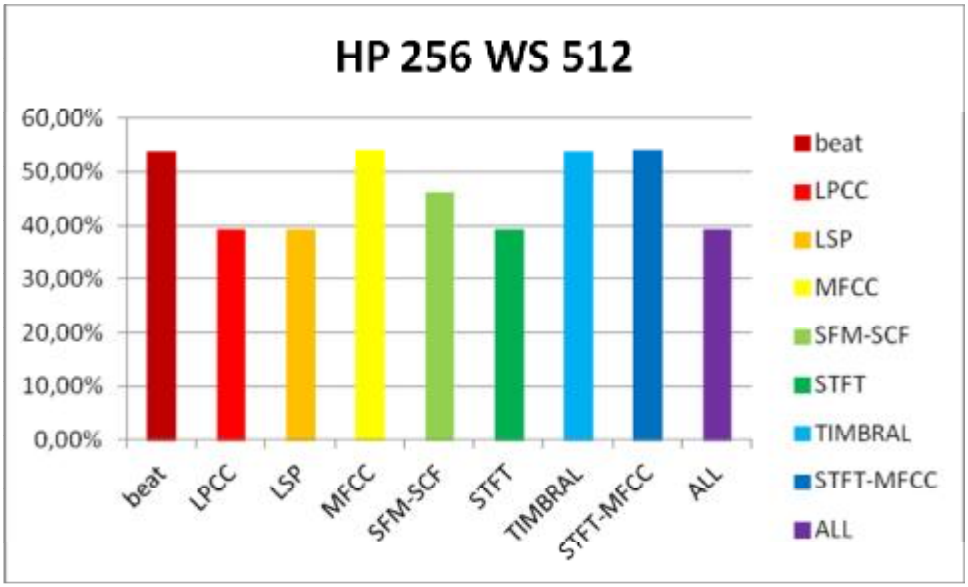
	Cl1	Cl2	Cl3
Cl1	**0**	333	0
Cl2	0	**355**	0
Cl3	0	215	**0**

5.1.2.1Συγκεντρωτικοί πίνακες αποτελεσμάτων

WINDOW SIZE, HOP SIZE DEFAULT(512,512)	
BEAT	39,36%
LPCC	28,23%
LSP	39,36%
MFCC	28,23%
SFM-SCF	39,36%
STFT	39,36%
TIMBRAL	39,36%
STFT-MFCC	39,36%
ALL	39,36%



WINDOW SIZE, HOP SIZE(512,256)	
BEAT	53,93%
LPCC	39,31%
LSP	39,31%
MFCC	54,26%
SFM-SCF	46,17%
STFT	39,31%
TIMBRAL	53,93%
STFT-MFCC	54,25%
ALL	39,31%


HP 256 WS 512
60,00%
50,00%
40,00%
30,00%
20,00%
10,00%
0,00%
beat
LPCC
LSP
MFCC
SFM-SCF
STFT
TIMBRAL
STFT-MFCC
ALL
beat
LPCC
LSP
MFCC
SFM-SCF
STFT
TIMBRAL
STFT-MFCC
ALL

Κεφάλαιο 6 - Αποτελέσματα

6.1 Αξιολόγηση

Η ακρίβεια του συστήματος όπως δείχνουν και τα ποσοστά είναι χαμηλή, ιδιαίτερα τα αποτελέσματα από τις πέντε κατηγορίες συναισθημάτων. Τα αποτελέσματα μπερδεύονται στις κατηγορίες 1 με 5και 2 με 4 γιατί αυτές οι κατηγορίες έχουν ομοιότητες στα συναισθήματα.

6.1.1 Αξιολόγηση πρώτης φάσης

Όπως καταλαβαίνουμε από τα πειράματα, το ποσοστό ακρίβειας ταξινόμησης είναι πολύ χαμηλό και το σύστημα μπερδεύει την κατηγορία ένα με την πέντε και την κατηγορία δύο με την τέσσερα, όπως προαναφέρθηκε. Τα ποσοστά ακρίβειας κυμαίνονται από 19 % έως 38%. Τα ποσοστά καλυτερεύουν όταν το window size είναι 512 και το hop size 256.

Πιο συγκεκριμένα, για το beat histogram τα ποσοστά είναι 23% για τις τιμές default των window size και hop size, και 38% όταν τις αλλάξουμε. Το ίδιο ισχύει και για τα timbral features. Στα ίδια επίπεδα κυμαίνονται και τα άλλα χαρακτηριστικά με τους LPCC και MFCC να έχουν τις χειρότερες τιμές(19%) όταν εφαρμόζονται οι αρχικές τιμές των window size και hop size. Οι LPCC και MFCC με window size 512 και hop size 256 έχουν τιμές 23 και 37% αντίστοιχα. Το τέταρτο και πέμπτο καλύτερο ποσοστό μετά τα beat histogram, timbral features και MFCC, είναι ο συνδυασμός των spectral features – MFCC, και οι SFM-SCF με τιμές 23% με default window size και hop size και 37% με window size 512 και hop size 256, και για 23% με default window size και hop size και 29% με window size 512 και hop size 256.

Υπάρχουν και τρεις περιπτώσεις στις οποίες το ποσοστό χειροτερεύει αλλά όχι σε μεγάλο βαθμό (κατά 0,05%). Δηλαδή, έχουν καλύτερο αποτέλεσμα χρησιμοποιώντας τις τιμές default για το window size και hop size. Αυτό παρατηρείται για τα χαρακτηριστικά LSP και spectral features καθώς και όταν εξάγουμε όλα τα χαρακτηριστικά.

6.1.2 Αξιολόγηση δεύτερης φάσης

Όπως παρατηρήθηκε, στην πρώτη φάση των πειραμάτων, τα ποσοστά ακρίβειας του συστήματος είναι χαμηλά. Στη δεύτερη φάση των πειραμάτων

μειώσαμε τις κατηγορίες για να αυξήσουμε το ποσοστό ακρίβειας του συστήματος. Η αύξηση του ποσοστού είναι της τάξης του 14% περίπου.

Πιο συγκεκριμένα, τα timbral features και beat histogram εμφάνισαν ποσοστό 39% για τις τιμές default των window size και hop size και 53% με window size 512 και hop size 256. Οι MFCC και LPCC με τιμές default των window size και hop size φτάνουν το 28% και αλλάζοντας τες σε window size 512 και hop size 256 φτάνουν σε 54 και 39% αντίστοιχα. Οι SFM-SCF βγάζουν ποσοστό 39% με τιμές default των window size και hop size και 46% με window size 512 και hop size 256. Το καλύτερο αποτέλεσμα το έχει ο συνδυασμός των spectral features με τους MFCC με ποσοστό 39% για τις default τιμές των window size και hop size και 54% με window size 512 και hop size 256.

Και εδώ εμφανίζονται χειρότερα αποτελέσματα με window size 512 και hop size 256 για τα ίδια χαρακτηριστικά όπως στην πρώτη φάση. Το ποσοστό βρίσκεται στο 39% για τα LSP, spectral features και εξάγοντας όλα τα χαρακτηριστικά μαζί. Η μείωση συνεχίζει να είναι 0,05% όπως στην πρώτη φάση, για το window size 512 και hop size 256.

Κλείνοντας, διαπιστώνουμε ότι, η αύξηση ακρίβειας ταξινόμησης του συστήματος μειώνοντας τις κατηγορίες κυμαίνεται από 9 μέχρι 16%. Το ποσοστό αυτό δεν είναι και τόσο σημαντικό. Αυτό ίσως οφείλεται στο γεγονός ότι δεν είναι απόλυτα σωστός ο τρόπος που έχουν ταξινομηθεί τα κομμάτια σε κατηγορίες. Οι ερευνητές Panda και Fernandes[Panda, Malheiro, Rocha, Oliveira & Paiva, 2008; Panda & Paiva DAF 2012; Panda & Paiva MML 2012; Panda & Paiva, CISUC 2011; Panda & Paiva, AES 2011; Cardoso, Panda & Paiva, INForum 2011; Panda & Paiva, CISUC 2012; Hu, Downie, Laurier, Bay & Ehmann, 2008; Panda, Rocha & Paiva, CISUC], χρησιμοποιώντας τα ίδια δεδομένα, εμφανίζουν ποσοστά που αγγίζουν το 60%. Αυτό οφείλεται στο γεγονός ότι, αναλύουν μαζί σταθερά χαρακτηριστικά ήχου, μελωδικά χαρακτηριστικά, χαρακτηριστικά MIDI και χαρακτηριστικά στίχων. Επιπλέον χρησιμοποιούν πέρα του ενός λογισμικού(Marsyas, Psysound, MIR Toolbox).

Κεφάλαιο 7 - Συμπεράσματα

7.1 Συμπεράσματα και μελλοντική δουλειά

Παρουσιάσαμε ένα σύστημα το οποίο εκπαιδεύεται για να αναγνωρίζει αυτόματα συναισθήματα από μουσική. Τα δεδομένα που χρησιμοποιήσαμε τα αντλήσαμε από το διαδίκτυο. Το λογισμικό που χρησιμοποιήσαμε είναι το Marsyas[Tzanetakis & Cook, 2002] από το οποίο εξάγαμε τα μουσικά χαρακτηριστικά. Τα πειράματά μας πραγματοποιήθηκαν σε δύο φάσεις. Στην πρώτη φάση χρησιμοποιήσαμε 5 κατηγορίες συναισθήματος και στην δεύτερη 3 κατηγορίες. Οι 5 κατηγορίες εμφάνισαν χειρότερα αποτελέσματα σε σχέση με τις 3. Η διαφορά τους κυμαίνεται από 9 με 16% αύξηση της δεύτερης φάσης. Το καλύτερο αποτέλεσμα για τις πέντε κατηγορίες είναι 38% ενώ για τις 3 κατηγορίες 54%.

Σε μια μελλοντική δουλειά θα μπορούσε να γίνει καλύτερη ταξινόμηση στα κομμάτια ή και επαναπροσδιορισμός τους, ώστε να μην μπερδεύεται το σύστημα και να εμφανίζει καλύτερα ποσοστά ακρίβειας. Επιπλέον, θα μπορούσαν να πραγματοποιηθούν και ground truth πειράματα, για να δούμε πως ταξινομούν τα συναισθήματα οι εθελοντές. Έτσι, να γίνει σύγκριση των αποτελεσμάτων από τα πειράματα και από αυτά του συστήματος και να βελτιωθεί η ποιότητα του. Τέλος, θα μπορούσε να γίνει ένας συνδυασμός από παραμέτρους: μουσικής, αρχείων MIDI και στίχων για καλύτερα αποτελέσματα.

Βιβλιογραφία

Πηγές:

www.sciencedirect.com

www.ieeexplore.com

www.mir.dei.uc.pt

www.repository.cmu.edu

www.marsyasweb.appspot.com

www.marsyas.info

www.marsology.blogspot.com

Άρθρα- Βιβλία- Εγχειρίδια:

Αλέξανδρος Νανόπουλος, Δημήτριος Ραφαΐλίδης, Παναγιώτης Συμεωνίδης, Γιάννης Μανολόπουλος, 'MusicBox: Personalized music recommendation system based on cubic analysis of social tags', IEEE 2009.

A. P. Oliveira and A. Cardoso. 'Automatic manipulation of music to express desired emotions'. In Proc. Sound and Music Computing Conf., pages 265–270, 2009.

A. P. Oliveira and A. Cardoso. 'Controlling music affective content: A symbolic approach'. In Proc. Conf. Interdisciplinary Musicology, 2008.

A. Camacho. 'SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music'. PhD thesis, Univ. Florida, 2007.

A.Gabrielsson. 'Emotion perceived and emotion felt: Same or different?', Musicae Scientiae, pages 123–147, 2002. special issue.

A. Gabrielsson and E. Lindstrom. 'The influence of musical structure on emotional expression'. In P. N. Juslin and J. A. Sloboda, editors, Music and Emotion: Theory and Research. Oxford University Press, New York, 2001.
A. Klapuri. 'Sound onset detection by applying psychoacoustic knowledge'. In Proc. Int. Conf. Acoustics, Speech, and Signal Processing, pages 3089–3092, 1999.

A. Meng, P. Ahrendt, J. Larsen, and L. K. Hansen. 'Temporal feature integration for music genre classification'. IEEE Trans. Audio, Speech & Language Processing, 15(5):1654–1663, 2007.

A. Tellegen, D. Watson and L.Clark. 'On the dimensional and hierarchical structure of affect'. 1999 Psychological Science
A. Uitdenbogerd and R. van Schyndel. 'A review of factors affecting music recommender success'. In Proc. ISMIR, 2002

A. Wieczorkowska. ‘Towards extracting emotions from music’. In Proc. Int. Workshop on Intelligent Media Technology for Communicative Intelligence, pages 228–238, 2004.

A. Wieczorkowska, P. Synak, R. A. Lewis, and Z. W. Ras. ‘Extracting emotions from music data’. In Proc. Int. Symp. Intelligent Systems, pages 456–465, 2005.

A. Wieczorkowska, P. Synak, and Z.W. Ras. ‘Multi-label classification of emotions in music’. In Proc. Intelligent Information Processing and Web Mining, pages 307–315, 2006.

All music guide. [Online] http://www.allmusic.com/.

B. Logan, ‘Mel Frequency Cepstral Coefficients for music modeling’. The First International Symposium on Music Information Retrieval, 2000.

B. Logan. ‘Music recommendation from song sets’, In Proc. ISMIR, Oct. 2004, pp. 425-428.

B. Sarwar, G. Karypis, J. Konstan, J.Riedi. ‘Application of dimensionality reduction in recommender systems-a case study’. In Proc. ACM WebKDD workshop, 2000
B. Sarwar, G. Karypis, J. Konstan, J.Riedi. ‘Item-based collaborative filtering recommendation algorithms’, In Proc. 10th WWW Conf., 2001, pp.285-295.

B. Schuller, C. Hage, D. Schuller, and G. Rigoll. ‘MisterD.J., CheerMeUp!: Musical and textual features for automatic mood classification’. J. New Music Research, 39(1):13–34,2010.

B. Whitman and P. Smaragdis. ‘Combining musical and cultural features for intelligent style detection’. In proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR ’02), Paris, France, October 2002

Bo Shao, Dingding Wang, Tao Li, Mitsunori Ogihara, ‘Music recommendation based on acoustic features and user access patterns’, IEEE 2009.

Bram van de Laar. ‘Emotion detection in music, a survey’. 2006

Byeong-Jun Han, Seungmin Rho, Roger B. Dannenberg, Eenjun Hwang, ‘Smers: music emotion recognition using Support Vector Regression’, ISMIR 2009.

C. C. Liu, Y.-H. Yang, P.-H. Wu, and H. H. Chen. ‘Detecting and classifying emotion in popular music’. In Proc. Joint Int. Conf. Information Sciences, pages 996–999, 2006.

C. E. Osgood, G. J. Suci, P. H. Tannenbaum, ‘The measurement of meaning’. University of Illinois Press, 1957

C. H. Lee, J. L. Shih, K. M. Yu, H. S. Lin. 'Automatic music genre classification based on modulation spectral, analysis of spectral and cepstral features'. IEEE Transitions on Multimedia, vol. 11, pp. 670-682, 2009.

C. M. Lee and S. S. Narayanan. 'Toward detecting emotions in spoken dialogs'. IEEE Trans. Speech and Audio Processing, 13(2):293–303, 2005.

C. Harte, M. Sandler, and M. Gasser. 'Detecting harmonic change in musical audio'. In Proc. ACM Workshop on Audio and Music Computing Multimedia, pages 21–26, 2006.

C. Laurier, J. Grivolla, and P. Herrera. 'Multimodal music mood classification using audio and lyrics'. In Proc. Int. Conf. Machine Learning and Applications, pages 105–111, 2008.

C. Laurier, O. Meyers, J. Serra, M. Blech, P. Herrera, and X. Serra. 'Indexing music by mood: Design and integration of an automatic content-based annotator'. Multimedia Tools and Applications, 2009.

Cheng-Che Lu, Vincent S. Tseng, 'A novel method for personalized music recommendation', Science Direct 2009.

Chih-Chung. Chang, and Lin, Chih-Jen: 'LIBSVM: a library for support vector machines', 2001. Available: http://www.csie.ntu.edu.tw/ ~cjlin/libsvm.

Chuan-Yu Chang, Chun-Yen Lo, Chi-Jane Wang, Pau-Choo Chung, 'A music recommendation system with consideration of personal emotion' IEEE 2010.

Conor Hayes, Padraig Cunningham, 'Smart radio – community based music radio', Science Direct 2001.

D. J. Jargreaves and A. C. North. 'The Social Psychology of Music'. Oxford University Press, Oxford, UK, 1997.

D. N. Jiang, L. Lu, H. J. Zhang, J. H. Tao, and L. H. Cai. 'Music type classification by spectral contrast features'. In Proc. IEEE Int. Conf. Multimedia Expo., pages 113–116, 2002.

D. Cabrera. 'Psysound: A computer program for psycho-acoustical analysis'. In Proc. Australian Acoustic Society Conf., pages 47–54, 1999. [Online] http://psysound. wikidot.com/.

D. Eck, P. Lamere, T. Bertin-Mahieux, S. Green. 'Automatic generation of social tags for music recommendation'. In Proc. 21st NIPS Conf., 2007, pp. 385-392.

D. Huron. 'Sweet Anticipation: Music and the Psychology of Expectation', MIT Press, Cambridge, Massachusetts, 2006.

D. Keltner and P. Ekman. 'The psychophysiology of emotion'. Handbook of Emotions, M. Lewis and J.M. Haviland-Jones, eds., pages 236–249, 2000.

D. Liu, L. Lu, and H.-J. Zhang. ‘Automatic music mood detection from acoustic music data’. In Proc. Int. Conf. Music Information Retrieval, pages 81–87, 2003.

D. Yang and W.-S. Lee. ‘Disambiguating music emotion using software agents’. In Proc. Int. Conf. Music Information Retrieval, 2004.

DownWithTheLoads. http://www.downwiththeloads.com/.

E. Allamanche, J. Herre, O. Helmuth, B. Froba, T. Kasten, and M. Cremer. ‘Content-based identification of audio material using MPEG-7 low level description’. In Proc. Int. Conf. Music Information Retrieval, pages 197–204, 2001.

E. Benetos, M. Kotti, and C. Kotropoulos. ‘Large scale musical instrument identification’. In Proc. Int. Conf. Music Information Retrieval, 2007. [Online] http://www.ifs.tuwien.ac.at/mir/muscle/del/audiotools.html#SoundDescrToolbox.

E. Gomez. ‘Tonal Description of Music Audio Signal’. PhD thesis, Universitat Pompeu Fabra, Barcelona, 2006.

E. Pampalk. ‘A MATLAB toolbox to compute music similarity from audio’. In Proc. Int. Conf. Music Information Retrieval, 2004. [Online] http://www.ofai.at/ elias.pampalk/ma/.

E. Pampalk, A. Rauber, and D. Merkl. ‘Content-based organization and visualization of music archives’. In Proc. ACM Int. Conf. Multimedia, pages 570–579, 2002.

E. Schubert. ‘Update of the Hevner adjective check list’. Perceptual and Motor Skills, 96:1117– 1122, 2003.

E. Schubert. ‘Measurement and Time Series Analysis of Emotion in Music’. PhD thesis, School of Music Education, Univ. New South Wales, Sydney, Australia, 1999

E. Zwicker. ‘Subdivision of the audible frequency range into critical bands’. J. Acoustical Society of America, 33, 1961.

E. Zwicker and H. Fastl. ‘Psychoacoustics: Facts and Models’. Springer, New York, 1999.

F. Pachet, D. Cazaly and P. Roy. ‘A combinatorial approach to content-based music selection’, IEEE Multimedia vol. 7, no. 1, pp. 457-462, Jul. 2000.

Folkongs. http://www.ingeb.org/.

G. Collier. ‘Beyond valence and activity in the emotional connotations of music’. Psychology of Music, 35(1):110–131, 2007.

G. Furnas, S. Deerwester, S. Dumais. ‘Information retrieval using a singular value decomposition model of latent semantic structure’. In Proc. 11th ACM SIGIR Conf., 1988, pp. 465-480.

G. Peeters. ‘A large set of audio features for sound description (similarity and classification) in the CUIDADO project’. Technical report, IRCAM, 2004.

G. Tzanetakis and P. Cook. ‘Musical genre classification of audio signals’. IEEE Trans. Speech & Audio Processing, 10(5):293–302, 2002. [Online] http://marsyas.sness. net/.

Graham Percival, Γιώργος Τζανετακης, ‘Marsyas user manual for version 0.3’.

H. C Chen and A. L. P. Chen. ‘A music recommendation system based on music data grouping and user interests’. In Proc. CIKM ’01: Proc. 10th Int. Conf. Inf. Knowledge Manag., New York, 2001, pp. 231-238.

H. F. Abeles and J. W. Chung. ‘Responses to Music’, pages 285–342. IMR Press, San Antonio, TX, 1996.

H. Fastl. ‘Fluctuation strength and temporal masking patterns of amplitude-modulated broad-band noise’. 8(1):59–69, 1982.

H. Katayose, M. Imai, and S. Inokuchi. ‘Sentiment extraction in music’. In Proc. Int. Conf. Pattern Recognition, pages 1083–1087, 1998.

I. Daubechies, Ten Lectures on Wavelets. Philadelphia, PA: SIAM, 1992.

I. Fujinaga and K. McMillan. ‘Real-time recognition of orchestral instruments’. In Proc. Int. Computer Music Conf., pages 141–143, 2000.

J. A. Russell. ‘A circumplex model of affect’. J. Personality & Social Psychology, 39(6):1161– 1178, 1980.

J. A. Sloboda, S. A. O’Neill, and A. Ivaldi. ‘Functions of music in every day life: An exploratory study using the experience sampling methodology’. Musicae Scientiae, 5(1):9–32, 2001

J. A. Sloboda and P. N. Juslin. ‘Psychological perspectives on music and emotion’. In P. N. Juslin and J. A. Sloboda, editors, Music and Emotion: Theory and Research. Oxford University Press, New York, 2001.

J. B. Schafer, J. Konstan, J.Riedi. ‘Recommender Systems in e-commerce’. In Proc. EC ’99: Proc. 1st ACM Conf. Electronic Commerce, 1999, pp. 158-166.

J. C. Platt. C. J. C. Burges, S. Swenson, C. Weare, A. Zheng. ‘Learning a Gaussian process prior for automatically generating music playlists’. In Advances in Neural Information Processing Systems 14, 2002, pp. 1425-1432.

J. J. Aucouturier and Francois Pachet, ‘Music Similarity Measures: What’s the Use?’, International Symposium on Music Information Retrieval, pp. 157-163, 2002.

J. S. Breese, D. Heckerman, C. Kadie. ‘Empirical analysis of predictive algorithms for collaborative filtering’, In Proc. 14th Annu. Conf. Uncertainty Artif. Intell., 1998, pp. 43-52

J. Ajmera, I. McCowan, H. Bourlard. ‘Speech/music segmentation using entropy and dynamism features in a HMM classification framework’. Speech Communication, vol. 40, pp. 351- 363, 2003.

J. Chalupper and H. Fastl. ‘Dynamic loudness model for normal and hearing-impaired listeners’. 88:378–386, 2002.

J. Ricard. ‘Towards Computational Morphological Description of Sound’. PhD thesis, Univ. Pompeu Fabra, Barcelona, 2004.

J. Skowronek, M. F. McKinney, and S. van de Par. ‘Ground truth for automatic music mood classification’. In Proc. Int. Conf. Music Information Retrieval, pages 395–396, 2006.

Jensen. ‘Timbre models of musical sounds’. Technical report, University of Copenhagen, 1999.

João André Ferro Fernandes, Rui Pedro Paiva, ‘Automatic Playlist Generation via Music Mood Analysis’, University of Coimbra 2010.

K. R. Scherer. ‘Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them’. J. New Music Research, 33(5):239–251, 2004.

K. Anderson and P. W. McOwan. ‘A real-time automated system for the recognition of human facial expressions’. IEEE Trans. System, Man & Cybernetics, 36(1):96–105, 2006.

K. Bischoff, C. S. Firan, R. Paiu, W. Nejdl, C. Laurier, and M. Sordo. ‘Music mood and theme classification-a hybrid approach’. In Proc. Int. Conf. Music Information Retrieval, pages 657–662, 2009.

K. Hevner. ‘Expression in music: A discussion of experimental studies and theories’. Psycho- logical Review, 48(2):186–204, 1935.

K. Hevner. ‘Experimental studies of the elements of expression in music’. American J. Psychology, 48:246–268, 1936.

K. Jonghwa and E. Ande. ‘Emotion recognition based on physiological changes in music listening’. IEEE Trans. Pattern Analysis & Machine Intelligence, 30(12):2067–2083, 2008.

K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas. ‘Multi-label classification of music into emotions’. In Proc. Int. Conf. Music Information Retrieval, pages 325–330, 2008.

K. Yoshii, M. Goto, K. Komatani, T. Ogata, H. G. Okuno. 'Hybrid collaborative and content-based music recommendation using probabilistic model with latent user preferences', In Proc. ISMIR, 2006

KIDiddles. http://www.kididdles.com/.

Kunsu Kim, Donghoon Lee, Tae-Bok Yoon, Jee-Hyong Lee, 'A music recommendation system based on personal preference analysis', IEEE 2008.
L. R. Rabiner, 'Fundamentals of speech recognition', Prectice-Hall, 1993.

L. Lu, D. Liu, and H. Zhang. 'Automatic mood detection and tracking of music audio signals'. IEEE Trans. Audio, Speech & Language Processing, 14(1):5–18, 2006.

Last.fm. [Online] http://www.last.fm/.

Lee, W. P., Lu, C. C. (2003). 'Customing WAP-based information services on mobile networks'. Personal and Ubiquitous Computing 7(6) 321-330
Lee, W. P., Lu, C. C., Liu, C. H. (2002). 'Intelligent agent-based systems for personalized recommendations in internet commerce'. Expert Systems with Applications, 22(4) 275-284

Lu, C. C., Tseng, V. S. (2006). 'Music Classification by parsing main melody'. In Proceedings of the International Computer Symposium (pp. 281-285)

Luís Cardoso, Renato Panda, Rui Pedro Paiva, 'MOODetector: A Prototype Software Tool for Mood-based Playlist Generation', INForum 2011.

M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. 'Content-based music information retrieval: Current directions and future challenges'. Proceedings of the IEEE, 96(4):668–696, 2008.

M. B. Holbrook and R. M. Schindler. 'Some exploratory findings on the development of musical tastes'. J. Consumer Research, 16:119 –124, 1989.

M. D Korhonen, D. A. Clausi, and M. E. Jernigan, 'Modeling emotional content of music using system identification', IEEE Trans. On Sys. Man. and Cyber., vol. 36, no. 3, pp.588-599, 2006.

M. G. Rigg. 'The mood effects of music: A comparison of data from four investigators'. J. Psychology, 58:427–438, 1964.

M. Y. Wang, N.-Y. Zhang, and H.-C. Zhu. 'User-adaptive music emotion recognition'. In Proc. IEEE Int. Conf. Signal Processing, pages 1352–1355, 2004.

M. Bartoszewski, H. Kwasnicka, U. Markowska-Kaczmar, and P. B. Myszkowski. 'Extraction of emotional content from music data'. In Proc. Computer Information Systems and Industrial Management Applications, pages 293–299, 2008.

M. Levy and M. Sandler. 'Learning latent semantic models for music from social tags'. J. New Music Res., vol. 37, no. 2, pp.137-150, 2008.

M. Sordo, C. Laurier, O. Celma. 'Annotating music collections: How content-based similarity helps to propagate labels', In Proc. 8th ISMIR Conf., 2007, pp. 531-534.

M. Tolos, R. Tato, and T. Kemp. 'Mood-based navigation through large collections of musical data'. In Proc. IEEE Consumer Communications & Network Conf., pages 71–75, 2005.

MATLAB wavelet toolbox. [Online] http://www.mathworks.com/products/ wavelet/.

MIDI Database. http://www.mididb.com/.

N. Oliver and L. Kreger-Stickles. 'Papa: Physiology and purpose-aware automatic playlist generation'. In Proc. 7th Int. Conf. Music Inf. Retrieval, Oct. 2006, pp. 250-253.

N. Scaringella, G. Zoia, and D. Mlynek, 'Automatic Genre Classification of Music Content', IEEE Signal Process Magazine, Vol. 23, No. 2, pp. 133-141, 2007.

New wisdom http://www.newwisdom.net.

O. Lartillot and P. Toiviainen. 'MIR in MATLAB (II): A toolbox for musical feature extraction from audio'. In Proc. Int. Conf. Music Information Retrieval, pages 127–130, 2007. [Online] http://users.jyu.fi/ lartillo/mirtoolbox/.

Owen Craigie Meyers, 'A Mood-Based Music Classification and Exploration System', Program in Media Arts and Sciences, Massachusetts Institute of Technology 2007

P. N. Juslin and J. A. Sloboda. 'Music and Emotion: Theory and Research', Oxford University Press, New York, 2001.

P. Dhanalakshmi, S. Palanivel, V. Ramalingam. 'Classification of audio signals using SVM and RBFNN'. Expert Systems of Applications, vol. 36, pp. 6069-6075, 2009.

P. Ekman. 'An argument for basic emotions'. Cognition and Emotion, 6(3):169–200, 1992.

P. Ekman, 'Basic Emotions', Handbook of Cognition and emotion, 1999, pp. 45-60

P. Lamere. Social tagging and music information retrieval. J. New Music Research, 37(2):101–114, 2008.

P. Symeonidis, M. Ruxanda, A. Nanopoulos, Y. Manolopoulos. 'Ternary semantic analysis of social tags for personalized music recommendation'. In Proc. 9th ISMIR Conf., 2008, pp. 219-224.

Peter Knees, Tim Pohle, Markus Schedl, Gerhard Widmer, 'Combining audio-based similarity with web-based data to accelerate automatic music playlist generation', Science direct 2006.

Peter Knees, Tim Pohle, Markus Schedl, Gerhard Widmer. 'An Innovative Three-Dimensional User Interface for Exploring Music Collections Enriched with Meta-Information from the Web'. In Proceedings of the ACM Multimedia 2006, Santa Barbara, California, USA, October 2006.

Q. Li, B.-M. Kim, D.-H. Guan, D.-W. Oh. 'A music recommender based audio features'. In SIGIR '04: Proc. 27th Annual Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, New York, 2004, pp. 532-533.

R. E Thayer, 'The Biopsychology of Mood and Arousal', New York: Oxford University Press, 1989.

R. W. Picard, E. Vyzas, and J. Healey. 'Toward machine emotional intelligence: Analysis of affective physiological state'. IEEE Trans. Pattern Analysis & Machine Intelligence, 23(10):1175–1191, 2001.

R. Cai, C. Zhang, L. Zhang, and W.-Y. Ma. 'Scalable music recommendation by search', In Proc. MULTIMEDIA '07: Proc. 15th Int. Conf. Multimedia , 2007, pp. 1065-1074.

R. Panda, B. Rocha, R. P. Paiva, 'Dimensional Music Emotion Recognition: Combining Standard and Melodic Audio Features', CISUC.

R. Panda, R. Malheiro, B. Rocha, A. Oliveira, R. P. Paiva, 'Multi-Modal Music Emotion Recognition: A New Dataset, Methodology and Comparative Analysis', CISUC 2008.

R. Ragno, C. J. C. Burges, C. Herley. 'Inferring similarity between music objects with application to playlist generation'. In Proc. 7th ACM SIGMM Int. Workshop Multimedia Inf. Retrieval, 2005, pp. 73-80.

Renato Panda, Rui Pedro Paiva, 'Automatic creation of mood playlists in the Thayer plane: a methodology and a comparative study', CISUC 2011.
Renato Eduardo Silva Panda, 'Automatic mood tracking in audio music', University of Coimbra 2010.

Renato Panda, Rui Pedro Paiva 'Mirex 2012: mood classification tasks submission', CISUC 2012.

Renato Panda, Rui Pedro Paiva, 'Music Emotion Classification: Analysis of a Classifier Ensemble Approach', MML 2012.

Renato Panda, Rui Pedro Paiva, 'Music emotion classification: dataset acquisition and comparative analysis', DAF 2012.

Renato Panda, Rui Pedro Paiva, 'Using Support Vector Machines for automatic mood tracking in audio music', AES 2011.

S. O. Ali. 'Songs and emotions: are lyrics and melodies equal partners'. Psychology of Music, 34(4):511–534, 2006.

S. R. Livingstone and A. R. Brown. ‘Dynamic response: A real-time adaptation for music emotion’. In Proc. Australian Conf. Interactive Entertainment, pages 105–111, 2005.

S. Baumann. ‘Artificial Listening Systems: Modellierung and Approximation der subjektiven Perzeption von Musikahnlichkeit’. PhD thesis, Technical University of Kaiserslautern 2005.

S. Davis and P. Mermelstein. ‘Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences’. IEEE Trans. Acoustics, Speech & Signal Processing, 28(4):357–366, 1980.

S. Hallam, I. Cross, and M. Thaut. ‘The Oxford Handbook of Music Psychology’, Oxford University Press, New York, 2008.

S. Ovadia. ‘Ratings and rankings: Reconsidering the structure of values and their measurement’. Int. J. Social Research Methodology, 7(5):403–414, 2004.

S. Pauws, W. Verhaegh and M. Vossen. ‘Fast generation of optimal music playlists using local search’, In Proc. 7th Int. Conf. Music Inf. Retrieval, Oct. 2006, pp. 138-143.

Seungjae Lee Jyng, Hyun Kim, Sung Min Kim, Won Young Yoo, ‘Smoodi: mood-based music recommendation player’, IEEE 2011.

Smola, Alex J., et al.: ‘A tutorial on support vector regression’, Statistics and Computing, Vol.14, pp.199-222, 2004.

T. L. Wu and S.-K. Jeng. ‘Probabilistic estimation of a novel music emotion model’. In Proc. Int. Multimedia Modeling Conf., pages 487–497, 2008.

T. Hofmann. ‘Probabilistic latent semantic indexing’. In Proc. ACM Int. Conf. Information Retrieval, pages 50–57, 1999.

T. Kolda, T. Bader, ‘Tensor decompositions and applications’, SIAM Rev., vol. 51, no. 3, Sep. 2009.

T. Li, M. Ogihara, Q. Li. ‘A comparative study on content-based music genre classification’. In Proc. SIGIR, 2003, pp. 282-289.

T. Li and M. Ogihara. ‘Content-based music similarity search and emotion detection’. In Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, pages 17–21, 2004.

T. Li and M. Ogihara. ‘Content-based music similarity search and emotion detection’. In Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Vol.5 pages 705-708, 2004.

T. Li and M. Ogihara. ‘Detecting emotion in music’. In Proc. Int. Conf. Music Information Retrieval, pages 239–240, 2003.

T. Li and M. Ogihara. ‘Toward intelligent music information retrieval’. IEEE Trans. Multi- media, 8(3):564–574, 2006.

T. Lidy and A. Rauber. ‘Evaluation of feature extractors and psycho-acoustic transformations for music genre classification’. In Proc. Int. Conf. Music Information Retrieval, pages 34–41, 2005. [Online] http://www.ifs.tuwien.ac. at/mir/audiofeatureextraction.html.

T. Tolonen and M. Karjalainen. ‘A computationally efficient multipitch analysis model’. IEEE Trans. Speech Audio Processing, 8(6):708–716, 2000.

Tseng, V. S., Su, J. H., Huang, J. H. (2006). ‘A novel video annotation method by integrating visual features and frequent patterns’. In ACM KDD workshop on multimedia data mining , Philadelphia, USA.

Tseng, V. S., Su, J. H., Wang, B. W., Lin, Y. M. (2007). ‘Web image annotation by fusing visual features and textual information’. In ACM Symposium on applied computing(SAC), Korea.

W. A. Sethares. ‘Tuning, Timbre, Spectrum, Scale’. Springer-Verlag, 1998.

W. F. Thompson and B. Robitaille. ‘Can composers express emotions through music?’, Empirical Studies of the Arts, 10:79–89, 1992.

W. L. Hill, M. Rosenstein, G. Furnas, ‘Recommending and Evaluating Choices in a Virtual Community of Use’, Proceedings on the Conference on Human Factors in Computing Systems (CHI95)ACM Press, Denver, CO, 1995, pp.194-201.

W. M. Hartmann. ‘Signals, Sound, and Sensation’. Springer, New York, 1998.

W. W. Cohen, W. Fan. ‘Web-collaborative filtering: Recommending music by crawling the web’, Comput. Netw., vol. 33, no. 1-6, pp. 685-698, 2000.

X. Changsheng, N. C. Maddage, S. Xi. ‘Automatic music classification and summarization’, IEEE Transitions on Speech and Audio Processing, vol. 13, pp. 441-450, 2005.

X. Hu and J. S. Downie. ‘Exploring mood metadata: Relationships with genre, artist and usage metadata’. In Proc. Int. Conf. Music Information Retrieval, 2007.

X. Hu, J. S. Downie, C. Laurier, M. Bay, and A. F. Ehmann. ‘The 2007 MIREX audio mood classification task: Lessons learned’. In Proc. Int. Conf. Music Information Retrieval, pages 462–467, 2008.

X. Zhu, Y. Y. Shi, H. G. Kim, K. W. Eom. ‘An integrated music recommendation system’. IEEE Transitions on Consumer Electronics, vol. 52, pp. 917-925, 2006

Xiao Hu, ‘Improving music mood classification using lyrics, audio, and social tag’, Doctoral Dissertation, University of Illinois at Urbana-Champaign, 2010.

Y. C. Huang, S.-K Jenor. ‘An audio recommendation system based on audio signature description scheme in mpeg-7 audio’. In 2004 IEEE Int. Conf. Multimedia Expo, 2004, vol. 1, pp. 639-642.

Y. H. Yang, C. C. Liu, and H. H. Chen. ‘Music emotion classification: A fuzzy approach’. In Proc. ACM Int. Conf. Multimedia, pages 81–84, 2006.

Y. H. Yang and H. H. Chen. ‘Predicting the distribution of perceived emotions of a music signal for content retrieval’. IEEE Trans. Audio, Speech & Language Processing. Submitted.

Y. H. Yang, Y.-C. Lin, and H. H. Chen. ‘Personalized music emotion recognition’. In Proc. ACM Int. Conf. Information Retrieval, pages 748–749, 2009.

Y. H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen. ‘A regression approach to music emotion recognition’. IEEE Trans. Audio, Speech & Language Processing, 16(2):448–457, 2008.

Y. H. Yang, Y.-F. Su, Y.-C. Lin, and H. H. Chen. ‘Music emotion recognition: The role of individuality’. In Proc. ACM Int. Workshop on Human-Centered Multimedia, pages 13–21, 2007. [Online] http://mpac.ee.ntu.edu.tw/ yihsuan/MER/hcm07/.

Y. Feng, Y. Zhuang, and Y. Pan. ‘Popular music retrieval by detecting mood’. In Proc. ACM Int. Conf. Information Retrieval, pages 375–376, 2003.

Yi-Hsuan Yang, Homer H. Chen, ‘Music Emotion Recognition’, Multimedia Computing, Communication and Intelligence 2011.