ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΜΕΛΕΤΗ ΣΥΣΤΗΜΑΤΟΣ ΑΥΤΟΜΑΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ ΑΝΤΙΓΡΑΦΩΝ ΜΟΥΣΙΚΩΝ ΤΡΑΓΟΥΔΙΩΝ

ΝΤΑΛΟΥΚΑ ΝΙΚΗ (ΑΜ 2197)
(πρώην Τμήματος ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΜΜΕ, ΤΕΙ ΔΥΤ. ΕΛΛΑΔΑΣ)

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ: ΚΟΥΤΡΑΣ ΑΘΑΝΑΣΙΟΣ

ΠΑΤΡΑ, 2021

[Αυτή η σελίδα είναι κενή]

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΜΗ ΛΟΓΟΚΛΟΠΗΣ

Βεβαιώνω ότι είμαι συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Ακόμα δηλώνω ότι αυτή η γραπτή εργασία προετοιμάστηκε από εμένα προσωπικά και αποκλειστικά και ειδικά για την συγκεκριμένη πτυχιακή εργασία και ότι θα αναλάβω πλήρως τις συνέπειες εάν η εργασία αυτή αποδειχθεί ότι δεν μου ανήκει.

ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΦΟΙΤΗΤΗ 1	**ΑΜ**	**ΥΠΟΓΡΑΦΗ**
ΝΤΑΛΟΥΚΑ ΝΙΚΗ	**2197**	
.....	.....	.....

[Αυτή η σελίδα είναι κενή]

ΕΥΧΑΡΙΣΤΙΕΣ

Η εκπόνηση της Πτυχιακής εργασίας έγινε με την υποστήριξη ενός συνόλου ανθρώπων τους οποίους θα ήθελα να ευχαριστήσω. Πρώτα απ' όλα ευχαριστώ τον επιβλέποντα καθηγητή μου, κ. Κούτρα Αθανάσιο. Θα ήθελα επίσης να ευχαριστήσω τη μητέρα μου για τη συμπαράσταση στην ολοκλήρωση των σπουδών μου, καθώς και την υπόλοιπη οικογένειά μου.

[Αυτή η σελίδα είναι κενή]

ΠΡΟΛΟΓΟΣ

Γνωρίζοντας τις μουσικές όψεις και πως αυτές μπορεί να χρησιμοποιηθούν για την απόσπαση σημαντικής πληροφορίας, μπορεί να κατασκευαστεί ένα σύστημα αναγνώρισης τραγουδιών "cover". Ο στόχος αυτής της κατασκευής είναι η χρήση ενός ήδη υπάρχοντος συστήματος, που αναλύει τα αποτελέσματά του και αναπτύσσει έναν τρόπο βελτίωσής τους.

Σε αυτήν την περίπτωση οι βελτιώσεις κατευθύνονται προς την αναγνώριση των τραγουδιών "cover", που είναι όσο το δυνατόν πλησιέστερα στα πλαίσια των στίχων και της ενορχήστρωσης του αυθεντικού τραγουδιού.

Η έρευνα αυτή θα ξεκινήσει δίνοντας μια γενική πληροφορία για τις μουσικές όψεις και πως επηρεάζουν τη διαδικασία της αναγνώρισης των διασκευών. Θα εξετάσει τις πιο κοινές προσεγγίσεις που χρησιμοποιούν τα συστήματα αναγνώρισης τραγουδιών "cover", προκειμένου να παράγουν ποιοτικά αποτελέσματα και να διευθύνεται η πρόσφατη εργασία στην περιοχή.

[Αυτή η σελίδα είναι κενή]

ΠΕΡΙΛΗΨΗ

Η παρούσα πτυχιακή εργασία ασχολείται με τη μελέτη συστημάτων αναγνώρισης τραγουδιών "cover", καθώς και με την υλοποίηση μιας τέτοιας εφαρμογής. Ο όρος του τραγουδιού "cover", είναι η εναλλακτική απόδοση από ένα προηγούμενο ηχογραφημένο τραγούδι. Ένα τραγούδι "cover" μπορεί να διαφέρει από το αυθεντικό κομμάτι ως προς την ποιότητα του τόνου, τον ρυθμό, τη δομή, το μουσικό κλειδί, τον κανονισμό ή τη γλώσσα των φωνητικών. Έχουν προταθεί αρκετές μέθοδοι, ώστε να υλοποιηθεί η αυτόματη αναγνώριση ενός τραγουδιού "cover" από την κοινότητα ανάκτησης μουσικής πληροφορίας (MIR), η οποία έχει δώσει ιδιαίτερη σημασία στο συγκεκριμένο έργο τα τελευταία χρόνια. Αυτές είναι :

· Η εξαγωγή μουσικών δακτυλικών αποτυπωμάτων για την αναγνώριση τραγουδιών "cover" κλασσικής μουσικής. Στόχος εδώ είναι να προσδιοριστούν οι διαφορετικές εκδόσεις της ίδιας μουσικής μέσω συγκρίσεων ομοιότητας των μουσικών δακτυλικών αποτυπωμάτων.

· Τα δυναμικά διανύσματα χαρακτηρισμού χρώματος με εφαρμογές αναγνώρισης τραγουδιού "cover", τα οποία περιγράφουν τις αλλαγές γειτονικών διαστημάτων βαθμού εντάσεως.

· Η αναγνώριση τραγουδιού "cover" σε μεγάλη κλίμακα με τη χρήση διακριτών σημείων χαρακτηριστικού ίχνους χρώματος. Με αυτή τη μέθοδο γίνεται η εύρεση "cover" από μία βάση δεδομένων εκατομμυρίων τραγουδιών με τη χρήση ανάλογων αλγόριθμων.

· Η αναγνώριση τραγουδιού "cover" με την άμεση εξαγωγή χαρακτηριστικού χρώματος από αρχεία AAC. Το σύστημα αυτό σχεδιάζει κατευθείαν τους τροποποιημένους συντελεστές διακεκριμένου μετασχηματισμού συνημίτονων σε χαρακτηριστικό χρώμα 12 διαστάσεων χωρίς να το αποκωδικοποιεί πλήρως.

· Η αναγνώριση ενός "live" τραγουδιού γνωστού καλλιτέχνη με τη χρήση "audio hashprints" (σχεδίαση μιας σειράς σήματος συνεχούς χρόνου σε μια ακολουθία διακεκριμένων συμβόλων που είναι κατάλληλα για αντίστροφη ευρετηρίαση και αποτελεσματική σύγκριση ζευγών).

· Η αναγνώριση τραγουδιού "cover" χρησιμοποιώντας μία μήτρα διασταύρωσης ομοιότητας από τραγούδι σε τραγούδι με συνελικτικό νευρωνικό δίκτυο (CNN).

ABSTRACT

The present thesis deals with the study of "cover" song recognition systems, as well as implementing such an application. The term "cover" is an alternative rendition of a previously recorded song. A "cover" song may differ from the original track in terms of tone quality, rhythm, structure, musical key, arrangement or phonetic language. Several methods have been proposed to enable automatic recognition of a "cover" song

by the Music Recovery Community (MIR), which has given particular importance to this project in recent years. These are :

· The export of musical fingerprints to identify classical music "cover" songs. The purpose here is to determine the different versions of the same music through comparisons of musical fingerprints.

· Dynamic chroma feature vectors with applications to "cover" song identification, which describe changes in adjacent degrees of intensity.

· Large-scale "cover" song recognition using hashed chroma landmarks. With this method we a "cover" from a database of millions of songs is found by the use of similar algorithms.

· Cover song indentification with direct chroma feature extraction from AAC files. This system directly designs modified coefficients of discrete cosine transforms in a 12-dimensional chroma feature without fully decoding it.

· Known-artist live song identification using audio hashprints (drawing a series of continuous time signals in a sequence of distinct symbols suitable for reverse indexing and efficient pair comparison).

· Cover song identification using song-to-song cross-similarity matrix with convolutional neural network (CNN).

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

Αναγνώριση τραγουδιού “cover”, Μουσική ομοιότητα, Ανάκτηση μουσικής πληροφορίας, Δομή μεταβλητότητας, Εξαγωγή χαρακτηριστικών, Κρίσιμη απόσταση, Χαρακτηριστικά χρώματος.

[Αυτή η σελίδα είναι κενή]

ΠΕΡΙΕΧΟΜΕΝΑ

[Αυτή η σελίδα είναι κενή]

[Αυτή η σελίδα είναι κενή]

ΕΙΣΑΓΩΓΗ

Η αναγνώριση τραγουδιών "cover", μελετάται αρκετά τα τελευταία χρόνια από την κοινότητα MIR και η αρμοδιότητά της έχει πολλές οπτικές γωνίες. Παίζει σημαντικό ρόλο στην επεξαργασία περιεχομένου του ήχου, αφού εξετάζει τις ομοιότητες της μουσικής, δίνοντας τρόπους να μετρηθούν και να μοντελοποιηθούν. Στον ορισμό της έννοιας της ομοιότητα μιας μουσικής, εμπλέκονται πολλοί παράγοντες, μπορεί να καθοριστεί και από πολιτιστικές και προσωπικές απόψεις[1].

Για να ανιχνευτεί ένα τραγούδι "cover", πρέπει να αντληθεί κάποια αμετάβλητη αναπαράσταση από το τραγούδι ολόκληρο ή ίσως από κάποια κρίσημα τμήματά του. Δεν είναι γνωστό ποια είναι η ουσιώδης πληροφορία που πρέπει να αποκωδικοποιηθεί, ώστε να λυθεί το πρόβλημα αυτό από τους ακροατές. Είναι σίγουρα σχετική η πληροφορία που λαμβάνεται από την ευαισθησία ή από την έλλειψη αυτής κατά τις μελωδικές παραλλαγές[2],[3]. Όταν ένα τραγούδι "cover" είναι αρκετά παρόμοιο σε χροιά με το αυθεντικό, ακόμα και μικρα αποσπάσματά του αρκούν για να γίνει η αναγνώριση[4]. Ένα πρόβλημα αποτελεί η μνήμη αναπαράστασης στους ανθρώπους. Στη μία περίπτωση το "cover" και το αυθεντικό κομμάτι θα μοιάζουν πολύ, οπότε θα γίνεται η αναγνώριση κατά την κωδικοποίηση, διαφορετικά θα γίνεται αποθήκευση της κωδικοποιημένης ηχογράφησης στη μνήμη, και έπειτα θα πραγματοποιούνται οι υπολογισμοί ομοιότητας.

Από την εμπορική άποψη, η ανίχνευση τραγουδιών "cover", έχει επίπτωση σε ότι αφορά τα μουσικά δικαιώματα. Ωστόσο, η ανάλυση της μουσικής ομοιότητας, αποτελεί "κλειδί" στην αναζήτηση, την ανάκτηση και την οργάνωση μουσικών συλλογών. Αυτό μπορεί να θεωρηθεί ως μειονέκτημα για τη μελλοντική μουσική βιομηχανία, αλλά επωφελεί τις προσωπικές μουσικές συλλογές. Υπάρχουν διάφορες βάσεις δεδομένων ή ιστοσελίδες στο διαδίκτυο, όπου μπορεί κάποιος να βρει πολλές παραλλαγές ενός τραγουδιού όπως τα : Second Hand Songs, Coverinfo, Coverville, Midomi, Fancovers, Youtube.

[Αυτή η σελίδα είναι κενή]

1 ΕΠΙΣΤΗΜΟΝΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

Καθώς γίνεται έρευνα σε προσεγγίσεις που αναφέρονται στην ομοιότητα και την ανάκτηση των τραγουδιών, όσον αφορά συμβολικούς και ηχητικούς τομείς, όπως σε συστήματα αναζήτησης μέσω query-by-humming, ανάκτησης μουσικής βάση περιεχομένου, ταξινόμησης φύλου, ή ηχητικών δακτυλικών αποτυπομάτων, λαμβάνουμε πληροφορίες για για τις ομοιότητες στα τραγούδια "cover".

Τα συστήματα αναζήτησης μέσω query-by-humming ανήκουν στο συμβολικό τομέα[5], από τον οποίο πηγάζουν αρκετές ιδέες για συστήματα αναγνώρισης τραγουδιών "cover". Στα συστήματα αυτά, ο χρήστης τραγουδά ή βουίζει κάποια μελωδία και το σύστημα ψάχνει για αντιστοίχηση σε μια μουσική βάση δεδομένων. Αυτά τα συστήματα διαχειρίζονται κάποιο συμβολικό είδος μουσικής πληροφορίας (συνήθως αρχεία MIDI), έτσι πρέπει να γίνει μεταγραφή του ερωτήμτος και του μουσικού υλικού σε συμβολικό τομέα. Το μειονέκτημα των συστημάτων αυτών είναι πως δεν επιτυγχάνουν ακόμα υψηλή ακρίβεια στα ηχητικά μουσικά σήματα του πραγματικού κόσμου. Οι αλγόριθμοι της σύγχρονης μουσικής τεχνολογίας αποδίδουν συνολικές ακρίβειες γύρω στο 75%, ακόμα και για την εκτίμηση της μελωδίας.

Η οργάνωση της ανάκτησης μουσικής βάση περιεχομένου, καθορίζεται από περιπτώσεις χρήσης, όπως έναν τύπο ερωτήματος, την αίσθηση αντιστοίχησης και τη μορφή απόδοσης[6],[7]. Η αίσθηση της αντιστοίχησης υποδηλώνει διαφορετικούς βαθμούς εξειδίκευσης. Μπορεί να είναι ακριβής (ανάκτηση μουσικής με συγκεκριμένο περιεχόμενο) ή κατά προσέγγιση (ανάκτηση κοντινών γειτόνων σε μουσικό χώρο, όπου η εγγύτητα αποκωδικοποιεί διαφορετικές αισθήσεις της μουσικής ομοιότητας)[6]. Μία πρωτότυπη περίπτωση χρήσης είναι η ταξινόμηση φύλου[8], όπου επιτυγχάνεται η ομαδοποίηση των τραγουδιών σύμφωνα με το φύλο. Αυτή η μέθοδος έχει χαμηλή εξειδίκευση αντιστοίχησης[6]. Τα μουσικά δακτυλικά αποτυπώματα είναι παράδειγμα υψηλής εξειδίκευσης αντιστοίχησης[9]. Συνίσταται για ιδιαίτερη ερμηνεία συγκεκριμένου τραγουδιού (διπλή ανίχνευση).

Πολλές έρευνες χρησιμοποιούν την ενδιάμεση εξειδίκευση αντιστοίχησης, στην οποία οι αλγόριθμοι πολλών ηχητικών δακτυλικών αποτυπωμάτων, χρησιμοποιούν περιγραφείς βάση της τονοθέτησης[10],[11],[12].[13]. Αυτές οι προσεγγίσεις μπορούν να χαρακτηριστούν με όρους όπως, αναγνώριση ήχου, αντιστοίχηση ήχου, ή ανάκτηση πολυφωνικού ήχου.

1.1 Συνοπτική παρουσίαση κεφαλαίων

Εφόσον υπάρχει η γνώση των μουσικών όψεων και πως αυτές μπορεί να χρησιμοποιηθούν για να αποσπαστεί σημαντική πληροφορία, μπορεί να κατασκευαστεί ένα σύστημα αναγνώρισης τραγουδιών "cover". Ο στόχος αυτής της κατασκευής είναι η χρήση ενός ήδη υπάρχοντος συστήματος, που αναλύει τα αποτελέσματά του και αναπτύσσει έναν τρόπο βελτίωσής τους. Σε αυτήν την περίπτωση οι βελτιώσεις κατευθύνονται προς την αναγνώριση των τραγουδιών

"cover", που είναι όσο το δυνατόν πλησιέστερα στα πλαίσια των στίχων και της ενορχήστρωσης του αυθεντικού τραγουδιού.

Η έρευνα αυτή θα ξεκινήσει δίνοντας μια γενική πληροφορία για τις μουσικές όψεις και πως επηρεάζουν τη διαδικασία της αναγνώρισης των διασκευών. Θα εξετάσει τις πιο κοινές προσεγγίσεις που χρησιμοποιούν τα συστήματα αναγνώρισης τραγουδιών "cover", προκειμένου να παράγουν ποιοτικά αποτελέσματα και να διευθύνεται η πρόσφατη εργασία στην περιοχή.

2 ΜΟΥΣΙΚΕΣ ΟΨΕΙΣ ΚΑΙ ΠΡΟΣΕΓΓΙΣΕΙΣ ΓΙΑ ΤΗΝ ΑΝΙΧΝΕΥΣΗ ΤΟΥ “COVER”

Με τον όρο "cover", χαρακτηρίζεται η εναλλακτική απόδοση ενός προηγούμενου ηχογραφημένου τραγουδιού. Αυτό πραγματοποιείται σε περίπτωση που θέλει ο καλλιτέχνης να αποδώσει το αυθεντικό τραγούδι μεταγλωτισμένο σε διαφορετική γλώσσα, είτε να το προσαρμόσει στα γούστα κάποιας συγκεκριμένης περιοχής. Επίσης, μπορεί να πραγματοποιηθεί για τον εκσυγχρονισμό κάποιου παλιού τραγουδιού, είτε για την παρουσίαση νέων καλλιτεχνών ή τιμή σε κάποιον άλλο καλλιτέχνη, είτε απλά για την ευχαρίστηση της ερμηνείας κάποιου οικείου τραγουδιού.

2.1 Τύποι “cover”

Μεταξύ των τραγουδιών "cover", μπορούν να γίνουν πολλές διακρίσεις. Οι διακρίσεις αυτές βασίζονται κυρίως σε μουσικά χαρακτηριστικά. Βάση αυτών των χαρακτηριστικών αναφέρονται κάποια παραδείγματα "cover"[14] :

· Remaster (Ανανεωμένη έκδοση) : Η δημιουργία μιας νέας αυθεντίας, που σημαίνει κάποιο είδος βελτιστοποίησης ήχου σε ένα προηγούμενο, υπάρχων γινόμενο.

· Instumental (Ενορχηστρωμένο) : Έκδοση χωρίς στίχους.

· Acapella: Έκδοση κάποιου τραγουδιού μόνο με φωνητικά.

· Live performance (Ζωντανή εκτέλεση) : Κάποιο καταγεγραμμένο κομμάτι από ζωντανή εκτέλεση.

· Acoustic (Ακουστικό) : Καταγεγραμμένο κομμάτι με διαφορετική σειρά από ακουστικά όργανα σε πιο οικεία μορφή.

· Demo (Έκδοση Επίδειξης) : Ένα παράδειγμα της ιδέας του καλλιτέχνη.

· Duet (Ντουέτο) : Κομμάτι ερμηνευμένο με επέκταση του αριθμού των βασικών ερμηνευτών.

· Medley (Ποτ Πουρί) : Ερμηνεία σειράς κομματιών χωρίς παύση.

· Remix (Αναμειγμένο κομμάτι με άλλο ή άλλη μουσική) : Μία εναλλακτική αυθεντία τραγουδιού, με την πρόσθεση ή την αφαίρεση στοιχείων, ή απλά αλλάζοντας την εξίσωση, τη δυναμική, το βαθμό εντάσεως, το ρυθμό, το χρόνο που παίζεται, ή οτιδήποτε άλλο από τα διάφορα μουσικά συστατικά. Με τις αλλαγές αυτές, κάποια

κομμάτια αλλοιώνονται αρκετά, με αποτέλεσμα να μη θυμίζουν τόσο το αρχικό κομμάτι.

· Quotation (Αναφορά) : Συνήθως μελωδική αναφορά. Η ενσωμάτωση από ένα σχετικά σύντομο τμήμα μιας ήδη υπάρχουσας μουσικής. Δεν είναι μέλος της κεντρικής ουσίας της δουλειάς.

2.2 Οι μουσικές όψεις

Με την έννοια του "cover", μπορούν να θεωρηθούν οι μουσικές διαστάσεις στις οποίες, ένα τέτοιο κομμάτι μπορεί να διαφέρει από το αυθεντικό. Στην κλασσική μουσική διαφορετικές ερμηνείες του ίδιου τραγουδιού μπορεί να εμφανίσουν διακριτικές παραλλαγές που περιλαμβάνουν διαφορετική δυναμική, ρυθμό, ποιότητα τόνου ή προφορά κ.α.. Από την άλλη, στη λαϊκή μουσική, ο σκοπός είναι η ριζικά διαφορετική ερμηνεία του αυθεντικού, όταν αυτό πραγματοποιηθεί. Κάποια βασικά χαρακτηριστικά τα οποία αλλάζουν σε ένα κομμάτι "cover" είναι τα ακόλουθα :

· Τέμπο/Ποιότητα τόνου : Οι επικρατέστερες ομάδες παραλλαγών που αλλάζουν το γενικό χρώμα ή την πλοκή του ήχου είναι :

i. οι τεχνικές παραγωγής : η διαφορετική καταγραφή ήχου και οι τεχνικές επεξεργασίας (δυναμική συμπίεση, μικρόφωνα, εξισορρόπηση, κλπ).

ii. ενορχήστρωση : οι καινούριοι ερμηνευτές μπορεί να χρησιμοποιούν διαφορετικά όργανα, συνθέσεις ή διαδικασίες καταγραφής.

· Ρυθμός : ο ρυθμός μπορεί να αλλάζει στην καταγραφή ζωντανής μετάδοσης, αφού δεν είναι τόσο εύκολος ο έλεγχός του σε συναυλία. Στην κλασσική μουσική, μικρές διακυμάνσεις παρουσιάζονται για διαφορετικές αποδόσεις του ίδιου κομματιού.

· Συγχρονισμός : η ρυθμική δομή μπορεί να αλλάξει ανάλογα με την πρόθεση ή το συναίσθημα του ερμηνευτή.

· Δομή : αυτή η τροποποίηση μπορεί να γίνει τόσο απλά, όπως παρακάμπτοντας μια μικρή εισαγωγή, επαναλαμβάνοντας το ρεφραίν, παρουσιάζοντας ένα ενόργανο τμήμα, είτε μικραίνοντας ένα.

· Κλειδί : το κομμάτι μπορεί να μεταφερθεί σε ένα διαφορετικό κλειδί ή τονικότητα. Αυτό συμβαίνει για να προσαρμοστεί το εύρος των βημάτων σε διαφορετικό τραγουδιστή ή όργανο, ώστε να προκλυθούν κάποιες αλλαγές διάθεσης στον ακροατή.

· Εναρμόνιση : Καθώς συντηρείται το κλειδί, η πρόοδος της χορδής μπορεί να αλλάξει (προσθέτοντας ή αφαιρώντας χορδές, αντικαθιστώντας αυτές από γειτονικές, τροποποιώντας τους τύπους των χορδών, προσθέτοντας εντάσεις, κλπ).

· Στίχοι και γλώσσα : Ένας από τους λόγους που παρουσιάζεται ένα κομμάτι "cover", είναι για να μεταφραστεί σε άλλη γλώσσα, ώστε να γνωστοποιηθεί σε άλλες κοινότητες.

· Θόρυβος : Άλλες εκδηλώσεις ήχου που μπορεί να υπάρχουν σε καταγραφή τραγουδιού, κυρίως από κοινό, όπως χειροκρότημα, φωνές, ή σφυρίγματα, και πιο σπάνια σε συμπίεση ήχου και κωδικοποίηση ομιλίας.

2.3 Προσεγγίσεις

Η βασική προσέγγιση για να μετρηθεί η ομοιότητα μεταξύ τραγουδιών "cover" είναι να εκμεταλλευτούν οι μουσικές όψεις μοιρασμένες μεταξύ τους. Από τη στιγμή που υποβάλλονται για μεταβολή κάποια σημαντικά χαρακτηριστικά, όπως η ποιότητα του τόνου, το κλειδί, η εναρμόνιση, ο ρυθμός, ο συγχρονισμός, η δομή και ούτω καθεξής, τα συστήματα αναγνώρισης τραγουδιών "cover", πρέπει να είναι ισχυρά ενάντια σε αυτές τις μεταβολές.

Υπεύθυνα για την υπερνίκηση της πλειοψηφίας των μουσικών αλλαγών μεταξύ των "covers", είναι συνήθως τα εξαγόμενα περιγραφικά στοιχεία, αλλά καθώς αυτά δεν μπορούν να διαχειριστούν κάποιες πολύ συχνές αλλαγές, όπως ο ρυθμός, το κλειδί ή τη δομή της μεταβλητότητας, δίνεται ιδιαίτερη έμφαση στην επίτευξή τους. Επομένως, τα βασικά λειτουργικά διαγράμματα που αποτελούν τα στοιχεία των συστημάτων αναγνώρισης τραγουδιών "cover" που υπάρχουν, είναι τέσσερα: η εξαγωγή χαρακτηριστικών, η βασική μεταβλητή, η μεταβλητή ρυθμού και η μεταβλητή δομής. Ένα επιπλέον διάγραμμα μπορεί να θεωρηθεί στο τέλος της αλυσίδας για το τελικό μέτρημα των ομοιοτήτων που χρησιμοποιήθηκε.

2.3.1 Εξαγωγή χαρακτηριστικών

Οι διαφορετικές εκδοχές του ίδιου κομματιού, διατηρούν την κεντρική μελωδία και/ή την αρμονική κίνηση, ανεξάρτητα από το κεντρικό κλειδί του. Λόγω αυτής της ιδιότητας, το τονικό ή το αρμονικό περιεχόμενο είναι ένα χαρατηριστικό μεσαίου επιπέδου που θα μπορούσε να ληφθεί υπ' όψην για τη δυναμική αναγνώριση των "covers".

Ο όρος τονικότητα, υποδηλώνει ένα σύστημα σχέσεων μεταξύ μιας σειράς βαθμών εντάσεως, που μπορούν να σχηματίσουν μελωδίες και αρμονίες, έχοντας κάποιο τόνο ως το πιο σημαντικό στοιχείο του[14].

Μια τονική αλληλουχία μπορεί να γίνει αντιληπτή ως μία σειρά διαφορετικών συνδιασμών νότας παιγμένων διαδοχικά. Αυτές οι νότες μπορεί να είναι είτε μοναδικές (μελωδία), είτε να παιχτούν από κοινού με άλλες (συγχορδία ή αρμονική κίνηση). Όλοι οι αλγόριθμοι αναγνώρισης "cover", εκμεταλλεύονται αναπαραστάσεις τονικής αλληλουχίας, που εξάγονται από ακατέργαστα σήματα ήχου. Εξαίρεση αποτελούν τα πρώιμα συστήματα.

Η μελωδία αποτελεί έναν προεξέχων μουσικό περιγραφέα από ένα κομμάτι μουσικής[15] και αρκετά συστήματα αναγνώρισης "cover", χρησιμοποιούν μελωδικές αναπαραστάσεις ως κεντρικό περιγραφέα[16],[17],[18],[19],[20]. Αρχικά αυτά τα συστήματα εξάγουν από το ακατέργαστο σήμα την επικρατέστερη μελωδία[21]. Η εξαγωγή αυτή σχετίζεται με την ανίχνευση βαθμού εντάσεως, η οποία γίνεται περίπλοκη, γιατί αν και πολλαπλοί βαθμοί εντάσεως παρουσιάζονται την ίδια στιγμή,

στα περισσότερα μόνο ένα από αυτά θα είναι η μελωδία. Για να εκκαθαριστεί η επικρατέστερη αναπαράσταση, τα συστήματα ανίχνευσης του "cover" συνδυάζουν έναν αποσυμπιεστή μελωδίας με έναν ανιχνευτή φωνής και άλλες μονάδες μέτρησης επεξεργασίας για να επιτευχθεί μια αξιόπιστη αναπαράσταση[18],[19],[20]. Μία άλλη πιθανότητα είναι να παράγουν τη λεγόμενη αναπαράσταση μεσαίου επιπέδου για αυτές τις μελωδίες, που η έμφαση τοποθετείται εκτός από την εξαγωγή μελωδίας, στη σκοπιμότητα της περιγραφής ήχου με τρόπο που διευκολύνει την ανάκτηση.

Εναλλακτικά, η ομοιότητα των "cover", μπορεί να εκτιμηθεί με αρμονικές αλληλουχίες, χρησιμοποιώντας τα λεγόμενα χαρακτηριστικά χρώματος PCP[22],[23],[24],[25]. Τα χαρακτηριστικά PCP προέρχονται από την ενέργεια που βρέθηκε στα πλαίσια ενός δοσμένου εύρους συχνοτήτων (συνήθως από 50 εώς 5000 Hz) σε σύντομο χρονικό διάστημα φασματικών αναπαραστάσεων (100 msec) ηχητικών σημάτων εξαγώμενων σε μία βάση καρέ καρέ. Η ενέργεια συνήθως καταπίπτει σε μια οκτάβα 12-bin ανεξάρτητου ιστογράμματος που αναπαριστά τη σχετική ένταση για κάθε ένα από τα 12 ημίτονα από ίση χρωματική κλίμακα. Τα αξιόπιστα χαρακτηριστικά PCP πρέπει να: α) αναπαριστούν την κατανομή τάξης βαθμού εντάσεως για μονοφωνικά και πολυφωνικά σήματα, β) εξετάζουν την παρουσία της αρμονικής συχνότητας, γ) είναι εύρωστα στο θόρυβο και στους μη τονικούς ήχους, δ) είναι ανεξάρτητα της ποιότητας του τόνου και του οργάνου που παίχτηκε, ε) είναι ανεξάτητα της έντασης και της δυναμικής, και ζ) είναι ανεξάρτητα του κουρδίσματος, έτσι ώστε η αναφερόμενη συχνότητα να μπορεί να είναι διαφορετική από την πρότυπη A 440 Hz[23]. Η πλειοψηφία των συστημάτων χρησιμοποιεί ένα χαρακτηριστικό βασισμένο σε PCP σαν πρωταρχική πηγή πληροφορίας[26],[27],[28],[29],[30].

Μία παραλλαγή χρήσης PCP, είναι η μέθοδος που χρησιμοποιεί διανυσματική κβαντοποίηση, όπου οι αλληλουχίες PCP καταπίπτουν σε αλληλουχίες αλυσίδων, δηλαδή συνοψίζοντας αρκετά χαρακτηριστικά διανυσμάτων από κοντινό αντιπρόσωπο, που γίνεται μέσω των K αλγόριθμων[31]. Η διανυσματική κβαντοποίηση εκτελείται με δυαδικό υπολογισμό συνιστώσας διανύσματος χαρακτηριστικού PCP.

Για τον υπολογισμό ομοιότητας μπορεί να χρησιμοποιηθεί η συγχορδία ή πρότυπες αλληλουχίες κλειδιού[32],[33],[34],[35]. Η διαδικασία για την εκτίμηση χορδών αποτελείται από την προεπεξεργασία του ήχου σε αναπαράσταση χαρακτηριστικού διανύσματος και την προσέγγιση της πιθανότερης ακολουθίας χορδών από αυτούς τους φορείς (μέσω templatematching ή expecting-maximization ή μέσω μοντέλων Hidden Markov Models)[36].

2.3.2 Μεταβλητή κλειδιού

Τα "covers" μπορεί να μεταφερθούν σε διαφορετικά κλειδιά. Καθώς οι βαθμοί εντάσεως γίνονται αντιληπτοί ως σχετικοί μεταξύ τους, παρά σε απόλυτες κατηγορίες, οι μεταφερόμενες εκδοχές είναι ισοδύναμες στους ακροατές[37]. Παρά την ύπαρξη μιας κοινής αλλαγής μεταξύ των εκδόσεων, ορισμένα συστήματα δεν εξετάζουν ρητά τις μεταφορές. Αυτή είναι η περίπτωση των συστημάτων που δεν επικεντρώνονται ειδικά στα τραγούδια "cover" ή δεν χρησιμοποιούν τονική αναπαράσταση[34],[38].

Η μετατόπιση αντανακλάται ως μετατόπιση του δακτυλίου σε σχέση με το "pitch axis" της αναπαράστασης χαρακτηριστικών. Υπάρχουν στρατηγικές για την αντιμετώπιση της μεταφοράς και η καταλληλότητά τους εξαρτάται από την εκλεκτική αναπαραγωγή χαρακτηριστικών. Γενικά, η μεταβλητότητα της μεταθέσεως μπορεί να επιτευχθεί με τη σχετική κωδικοποίηση χαρακτηριστικών, με εκτίμηση κλειδιού, με μετασχηματισμούς αμετάβλητης μετατόπισης ή με εφαρμογή διαφορετικών μεταθέσεων.

Για την επίτευξη αμετάβλητης κλειδιού γίνεται έλεγχος όλων των πιθανών μεταφορών στοιχείων[27],[28],[17]. Στην περίπτωση μιας ανεξάρτητης αναπαράστασης οκτάβας, αυτό υποδηλώνει τον υπολογισμό της μέτρησης της ομοιότητας για όλες τις κυκλικές ή δακτύλιες μετατοπίσεις στον άξονα βήματος για κάθε τραγούδι. Αυτή η τεχνική εγκυάται τη μέγιστη ανάκτηση ακρίβειας[39].

Προσεγγίσεις επιτάχυνσης αυτής της μεθόδου έχουν παρουσιαστεί[39],[40]. Δίνοντας μια τονική αναπαράσταση δύο τραγουδιών, αυτοί οι αλγόριθμοι υπολογίζουν τις πιο πιθανές σχετικές μεταθέσεις, δεδομένου ενός συνόλου αναπαράστασης της τονικής περιεκτικότητας για κάθε τραγούδι[26],[40]. Έτσι, αυτό το σύνολο αναπαράστασης μπορεί να είναι ένα απλό σύνολο PCP χαρακτηριστικών από όλη την αλληλουχία και να υπολογίζεται "off-line". Εν τέλη, οι πιο πιθανές μετατοπίσεις που έχουν επιλεχθεί ειναι οι K. Η περαιτέρω αξιολόγηση προτείνει ότι για αναπαραστάσεις βασισμένες σε 12 bin PCP, μια σχεδόν βέλτιστη ακρίβεια μπορεί να επιτευχθεί με 2 μετατοπίσεις[39], μειώνοντας έτσι 6 φορές το υπολογιστικό φορτίο. Ορισμένα συστήμα προκαθορίζουν ένα συγκεκριμένο αριθμό μετατοπίσεων προς υπολογισμό. Αυτές μπορεί να επιλεχθούν αυθαιρέτως[19],[2], ή να βασίζονται σε μουσική γνώση[33].

Μια προσέγγιση είναι να εκτιμηθεί το κεντρικό κλειδί off-line και έπειτα να εφαρμοστεί η μετατόπιση αναλόγως[29],[30]. Έτσι, τα λάθη μεταδίδονται ταχύτερα και μπορούν να χειροτερεύσουν την ανάκτηση ακρίβειας ανεπανόρθωτα[39],[40].

Σε περίπτωση που χρησιμοποιείται μια συμβολική αναπαράσταση όπως οι χορδές, θα μπορούσε αυτή να τροποποιηθεί παρεταίρω, ώστε να περιγράφονται γειτονικές αλλαγές χορδών. Με αυτόν τον τρόπο αποκτάται μια ακολουθία ανεξάρτητη κλειδιού[18],[32],[35]. Αυτή η ιδέα έχει επεκταθεί στις PCP ακολουθίες, χρησιμοποιόντας την έννοια της άριστης ένδειξης της μετατόπισης[40].

Για την επίτευξη της μεταβατικής μεταβλητότητας, γίνεται να χρησιμοποιηθεί ένα φάσμα ισχύος 2D[17], ή μια 2D λειτουργία αυτοσυσχέτισης[41]. Η αυτοσυσχέτιση είναι χειρισμός για τη μετατροπή σημάτων σε μια αναπαράσταση επιβράδυνσης ή σε μια αμετάβλητη μετατόπιση[42]. Το φάσμα ισχύος, που ορίζεται ως η μετατόπιση της αυτοσυσχέτισης του Fourier, είναι επίσης αμετάβλητη μετατόπιση. Άλλες 2D μετατοπίσεις μπορούν επίσης να χρησιμοποιηθούν.

2.3.3 Χρονική μεταβλητή

Σε διαφορετικές αποδόσεις ενός ίδιου κομματιού, ο ρυθμός μπορεί να τροποποιηθεί ώστε οι εξαγόμενες ακολουθίες να μην μπορούν να συγκριθούν άμεσα. Αν το "cover"

έχει ρυθμό 2 φορές γρηγορότερο, τότε ένα πλαίσιο ίσως αντιστοιχεί σε δύο πλαίσια στο αυθεντικό κομμάτι και αυτό δημιουργεί την ανάγκη για την έρευση ενός τρόπου που να είναι σε θέση να αντιστοιχεί αποτελεσματικά αυτά τα πλαίσια.

Ένας τρόπος είναι η χρήση της εξαγόμενης γραμμής μελωδίας για τον προσδιορισμό της αναλογίας της διάρκειας μεταξύ δύο διαδοχικών νοτών[18] (συστήματα query-by-humming)[5], ή αλλιώς η εκτίμηση του ρυθμού με την ανίχνευση του παλμού και τη συγκέντρωση των περιγραφικών πληροφοριών που αντιστοιχούν στον ίδιο παλμό (συστήματα βασισμένα σε PCP ή μελωδική αναπαράσταση). Μια εναλλακτική λύση στο τελευταίο είναι να γίνει χρονική συμπίεση και επέκταση που συνίσταται στην επανάληψη δειγματοληψίας της γραμμής μελωδίας σε συμπιεσμένες και εκτεταμένες μουσικές εκδόσεις που θα συγκριθούν έτσι ώστε να είναι η σωστή επαναδειγματοληψία που προσδιορίζεται. Ο μετασχηματισμός Fourier 2D, μπορεί επίσης να χρησιμοποιηθεί για να επιτευχθεί χρονική ανανέωση.

Τέλος, μπορούν να χρησιμοποιηθούν δυναμικές τεχνικές προγραμματισμού για την αυτόματη ανεύρεση τοπικών αντιστοιχιών. Λαμβάνοντας υπόψη τους γειτονικούς περιορισμούς και τα πρότυπα, μπορούν να προσδιοριστούν οι τοπικές αποκλίσεις του ρυθμού που μπορεί να επαναληφθούν. Οι αλγόριθμοι Dynamic Time Warping (DTW)[28] είναι η τυπική επιλογή επειδή ο κύριος στόχος τους είναι να ευθυγραμμίσουν δύο αλληλουχίες σε χρόνο που να επιτυγχάνουν μια βέλτιστη αντιστοίχιση.

2.3.4 Μεταβλητή δομής

Η βασική προσέγγιση για να γίνει μια δομή συστήματος αμετάβλητη είναι να συνοψιστεί ένα τραγούδι στα πιο αντιπροσωπευτικά ή επαναλαμβανόμενα μέρη του[16],[17],[29]. Προκειμένου να γίνει αυτό, το σύστημα πρέπει να είναι ικανό να κάνει κατάτμηση της δομής και να προσδιορίσει τα σημαντικότερα τμήματα.

Ο κατασκευαστικός τομέας, για τον εντοπισμό των βασικών κατασκευαστικών τμημάτων, είναι μια άλλη ενεργή έρευνα εντός της κοινότητας MIR, και κάνει τον αγώνα της κάθε χρόνο στη MIREX, αλλά όπως συμβαίνει στην αναγνώριση του "cover", οι λύσεις δεν είναι τέλειες. Σε αυτό πρέπει να ληφθεί υπ' όψιν πως το πιο αναγνωρίσιμο τμήμα ενός μουσικού κομματιού είναι ένα μικρό τμήμα, όπως μια εισαγωγή ή γέφυρα και όχι πάντα το πιο επαναλαμβανόμενο, όπως μια χορωδία.

Αλγόριθμοι δυναμικού προγραμματισμού, ειδικότερα, αλγόριθμοι τοπικής ευθυγράμμισης, όπως ο αλγόριθμος Smith-Waterman, μπορεί επίσης να χρησιμοποιηθούν για να αντιμετωπιστούν οι τεχνητές αλλαγές μεταξύ δύο τραγουδιών.

2.3.5 Υπολογισμός ομοιοτήτων

Το τελικό βήμα ενός συστήματος αναγνώρισης είναι να ανακτήσει μια λίστα από "covers" από μία μουσική συλλογή, η οποία κατατάσεται βάση μέτρησης ομοιοτήτων, έτσι ώστε τα κομμάτια να ξεκινάνε από τα πιο όμοια. Τα συστήματα αυτά κάνουν αυτές τις μετρήσεις ανάμεσα σε ζευγάρια κομματιών και λειτουργούν στην αναπαράσταση που έχει ληφθεί μετά την εξαγωγή χαρακτηριστικών, την μεταβλητή κλειδιού, ρυθμού, και των μονάδων μέτρησης μεταβλητής δομής.

Οι τεχνικές δυναμικού προγραμματισμού, που επιτυγχάνουν μεταβλητή ρυθμού παρέχουν ήδη μέτρηση ομοιότητας ως έξοδο[43],[44],[45]. Τα συστήματα αναγνώρισης τέτοιων τεχνικών χρησιμοποιούν τη μέτρηση ομοιότητας που αυτές παρέχουν. Αυτό ισχυεί για συστήματα που χρησιμοποιούν επεξεργασία αποστάσεων[33],[18] ή αλγόριθμους στρέβλωσης δυναμικού χρόνου[19],[20],[29],[30],[34],[35],[38]Βάση των μηκών των αναπαραστάσεων, δημιουργείται μια εξομάλυνση, που ίσως παράγει στις εκδόσεις με διαφορά στη διάρκεια, αντιθέσεις. Αν χρησιμοποιούνται τεχνικές δυναμικού προγραμματισμού τοπικής ευθυγράμμισης, η μέτρηση ομοιότητας αντιστοιχεί στο μήκος της ακολουθίας αντιστοίχησης που βρέθηκε[26],[40].

Τέλος, μπορεί να γίνει χρήση συμβατικών μετρήσεων ομοιότητας, όπως η διασταυρούμενη συσχέτιση[16],[27],[28], η Fobenius norm[46], η Euclidean distance[17],[41], ή το dot product[47],[48],[49]. Για την αντιμετόπιση των αλλαγών δομής, μπορεί να χρησιμοποιηθεί η ακολουθία στρατηγικής παραθύρου, όπου οι μετρήσεις συνδιάζονται με βήματα πολλαπλής επεξεργασίας όπως ο ορισμός κατωφλίου , τα βάθη TFIDF[17] ή οι αναλογίες αντιστοιχίας[49]. Λίγες από αυτές τις μετρήσεις συμπεριλαμβάνουν την εξομαλυσμένη συμπίεση απόστασης[32] και το μοντέλο του Markov.

3 ΕΚΤΙΜΗΣΗ ΤΩΝ ΚΡΙΤΗΡΙΩΝ

Για την εκτίμηση συστημάτων αναγνώρισης "cover" είναι δύσκολο να βρεθεί κοινή μεθολογία. Στη MIREX[7] βρέθηκε μια απόπειρα να γίνει σύγκριση μεταξύ εκδοχών συστημάτων αναγνώρισης, η οποία για κάθε σύστημα παρέχει μια συνολική ακρίβεια. Η υλοποίηση ανεξάρτητων εκτιμήσεων για την εξαγωγή χαρακτηριστικών ή τον υπολογισμό ομοιοτήτων, αποτελεί σημαντική βελτίωση, αφού αναλύει τις συνεισφορές που παρέχουν σε όλο το σύστημα.

Η διαδικασία της εκτίμησης, γίνεται με βάση την υποβολή από το χρήστη ενός τραγουδιού προς αναζήτηση[50]. Το σύνολο που ανακτάται από κάποια έτοιμη συλογή, αξιολογείται για την ακρίβειά του, λαμβάνοντας υπ' όψιν τις μετρήσεις εκτίμησης και το μουσικό υλικό που χρησιμοποιήθηκε.

3.1 Μετρήσεις εκτίμησης

Μία μέτρηση εκτίμησης που χρησιμοποιείται σε πειθαρχίες ανάκτησης πληροφορίας είναι το μέσο ακρίβειας της μέσης τιμής (MAP). Η MIREX χρησιμοποιεί τη MAP στην αξιολόγηση του ζητήματος της αναγνώρισης των "cover"[50].

Κάποιες άλλες μετρήσεις εκτίμησης είναι η R-Ακρίβεια[33],[34], παραλλαγές της ακρίβειας ή της ανάκλησης σε διαφορετικά επίπεδα τάξης[27],[28],[38],[46],[41],[47],[48],[49],[19],[20], η μέση τιμή ακρίβειας[51] και ανάκλησης και η F-μέτρηση[29],[30],[40].

3.2 Μουσικό υλικό

Το μουσικό υλικό που εξετάζεται σε κάθε κομμάτι είναι πολύ σημαντικό ζήτημα όσον αφορά την εκτίμηση. Το ζήτημα αυτό, ανάλογα την περίπτωση, εξαρτάται από τη μουσική συλλογή που μελετάται και τον τύπο της εκδοχής που θέλουμε να αναγνωρίσουμε, που μπορεί να είναι από επεξεργασμένα κομμάτια, έως τελείως διαφορετικά τραγούδια.

3.2.1 Είδος

Η σύγκριση δύο συστημάτων που εκτιμώνται σε διαφορετικές συνθήκες και έχουν σχεδιαστεί για να λύσουν διαφορετικά προβλήματα είναι πολύπλοκη. Οι δουλειές που έχουν υψηλές ακρίβειες αφορούν την κλασική μουσική , αν και δεν έχουν καλή ποιότητα τόνου, κατασκευή και ρυθμικές μεταβολές. Άλλες δουλειές χρησιμοποιούν ένα πιο παραλλαγμένο είδος κατανομής στις μουσικές τους συλλογές, κάποιες φορές είναι ακόμα ασαφές ποιοι τύποι εκδοχών χρησιμοποιούνται. Συνήθως είναι μιξαρισμένα και μπορεί να περιέχουν επεξεργασμένα κομμάτια (ευκολότερη ανίχνευση), ποτ-πουρί (η κεντρική άποψη μπορεί να είναι αλλαγές μεταβλητών στη δομή του τραγουδιού), ντέμο (έχουν σημαντικές παραλλαγές και σεβασμό στο

τραγούδι που θα κυκλοφορήσει), ρεμίξ και παραπομπές (χαμηλή διάρκεια και παραμορφωμένη αρμονία). Οι συλλογές στη MIREX περιέχουν μεγάλη ποικιλία ειδών, στυλ και ενορχηστρώσεων. Είναι όμως άγνωστος ο τύπος των "cover" που παρουσιάζονται. Η εφαρμοσιμότητα της μεθόδου που αναπτύσσεται εξασφαλίζεται μόνο με μεγάλες ποικιλίες.

3.2.2 Μεταβλητότητα

Οι ποσοτικές πτυχές του μουσικού υλικού πρέπει επίσης να ληφθούν υπόψιν. Τα τελικά ποσοστά ακριβείας μπορεί να επηρεαστούν από το πλήθος των τραγουδιών όπως και τη διανομή τους. Για την εξέταση αυτού θα χρειαστεί μια μουσική συλλογή να αποσυνδεθεί σε σειρές "cover", οι οποίες δίνουν τον πληθάριθμό τους. Ένα τεστ που έχει παρουσιαστεί, είναι βασισμένο σε μια μουσική συλλογή 2135 τραγουδιών "cover". Βάση των 2 παραμέτρων εκτελέσθηκαν 30 τυχαία τραγούδια. Στη συνέχεια υπολογίστηκε και σχεδιάστηκε για όλες τις θέσεις η μέση τιμή MAP. 'Ετσι προκείπτει πως με λιγότερες από 50 μουσικές σειρές ή μόνο με τον πληθάριθμο 2 αποδόσεων τα αποτελέσματα είναι ψηλά, κάτι το μη ρεαλιστικό, καθώς μεγαλύτερες τιμές την ίδια στιγμή , λήγουν σε μία σταθερή περιοχή ακρίβειας. Επίσης με λιγότερες από 50 εισαγωγές "cover" μια μεγάλη μεταβλητότητα παρατηρείται στην ακρίβεια εκτίμησης, που ίσως οφείλεται στο υποσύνολο που επιλέχθηκε. Αν αυξηθεί ο αριθμός των σειρών "cover" και ο πληθάριθμός τους γίνεται μικρότερη η μεταβλητότητα.

3.2.3 Ζητήματα μεγέθους

Πολλές μελέτες χρησιμοποιούν λιγότερες από 50 σειρές "cover" , που σημαίνει πως δεν μπορούμε να είμαστε βέβαιοι για τις ακρίβειες. Το ίδιο μπορεί να συμβεί και σε ένα σύνολο δεδομένων που χρησιμοποιούν ερευνητές , για τον έλεγχο των συστημάτων, το covers80, με 80 σειρές "cover" με έναν πληθάριθμο από 2.

Αν δεν υπάρχει μεγάλη εκτίμηση συνόλου δεδομένων, μπορεί να γίνει απόπειρα αποζημίωσης των τεχνουργημάτων, ενώ προστίθενται θόρυβος ή τραγούδια ελέγχου. Η συμπερίληψη των τραγουδιών αυτών ενδεχομένως να καθιστά το θέμα δύσκολο, καθώς η πιθανότητα να ληφθούν σχετικά αντικείμενα στα πλαίσια των αντικειμένων που κατατάχθηκαν πρώτα, είναι χαμηλή. Αυτό γίνεται και στη δομή της MIREX. Έτσι λοιπόν τα δεδομένα που δοκιμάζονται απαρίζονται από 30 σειρές "cover" αποτελούμενες από 11 εκδοχές, καταλήγοντας σε μια συλλογή από 330 τραγούδια. Το θέμα της ανίχνευσης γίνεται δυσκολότερο με την πρόσθεση 670 ατομικών τραγουδιών, σειρές "cover" με πληθάριθμο από 1.

Έτσι, αν η μουσική συλλογή είναι μεγάλη και με μεγάλη ποικιλία, τα αποτελέσματα εκτός δείγματος είναι πιο παρόμοια. Τα αντίθετα δεδομένα έχουν ασυνήθιστες ακρίβειες και ανακριβείς υπολογισμούς παραμέτρων.

4 ΔΑΚΤΥΛΙΚΑ ΑΠΟΤΥΠΩΜΑΤΑ ΗΧΟΥ ΚΑΙ ΣΥΓΚΡΙΣΗ ΑΛΓΟΡΙΘΜΩΝ

Μια αποτύπωση αποτυπωμάτων ήχου είναι ένα μικρό σύνολο χαρακτηριστικών που προσδιορίζει με μοναδικό τρόπο ένα τραγούδι. Είναι μια τεχνική που μπορεί να χρησιμοποιηθεί για την ταυτοποίηση του ήχου από το σήμα απευθείας. Μπορεί επίσης να χρησιμοποιηθεί για την επισήμανση μη επισημασμένου ήχου και για πιο σοβαρές εφαρμογές όπως παρακολούθηση εκπομπής. Το γενικό πλαίσιο για την οικοδόμηση ενός αποτυπώματος περιλαμβάνει μια αφετηρία και ένα μπλοκ μοντελοποίησης αποτυπώματος. Δύο βασικοί αλγόριθμοι που χρησιμοποιούνται σε αυτό είναι οι PRH και MLH.

Δύο φάσεις εμπλέκονται στην αναγνώριση ενός ηχητικού κλιπ από το αποτύπωμα του:

· **Φάση Εγγραφής :** Μια βάση δεδομένων ή ένα αποθετήριο είναι γεμάτο με αποτυπώματα και συναφή μετα-δεδομένα μεγάλου αριθμού τραγουδιών.

· **Φάση αναγνώρισης :** Τα αποτυπώματα άγνωστων τραγουδιών εξάγονται και συγκρίνονται με τα στοιχεία της βάσης δεδομένων. Εάν το αποτύπωμα του ηχητικού κλιπ βρει μια αντιστοιχία στη βάση δεδομένων, το τραγούδι θα αναγνωριστεί.

Το αποτύπωμα ενός ήχου είναι μια συμπαγής υπογραφή βασισμένη στο περιεχόμενο, που συνοψίζει αποτελεσματικά ολόκληρη την ηχογράφηση ή μέρος αυτής.

4.1 Εξαγωγή μουσικού αποτυπώματος για αναγνώριση τραγουδιών "cover" κλασικής μουσικής

Στην περίπτωση των τραγουδιών "cover" κλασικής μουσικής προτείνεται ένας αλγόριθμος για την εξαγωγή μουσικών δακτυλικών αποτυπωμάτων απευθείας από ένα ηχητικό σήμα, ο οποίος επιδιώκει να εγκλωβίσει διάφορες πτυχές των μουσικών πληροφοριών, όπως η συνολική κατανομή των σημείων, η δομή της αρμονίας και οι χρονικές μεταβολές τους, σε μια συμπαγή αναπαράσταση.

Δεδομένου ότι ο ήχος μουσικής αντιπροσωπεύει τυπικά ένα μείγμα πολλών μουσικών οργάνων ή φωνών, είναι απαραίτητο να χειρίζονται πολλαπλές θέσεις. Επίσης, η μαθηματική μοντελοποίηση της σύνθετης δυναμικής και η αλληλεπίδραση μεταξύ διάφορων πτυχών της μουσικής, όπως ο ρυθμός, η δομή της αρμονίας και η εξέλιξη της χορδής, δημιουργούν αρκετά ανοιχτά προβλήματα. Για όλα αυτά οι ερευνητές οδηγήθηκαν στην εστίαση του χειρισμού συγκεκριμένων πτυχών που περιγράφουν την υποκείμενη μουσική ανάλογα με την εφαρμογή.

Ειδικά στην αναγνώριση "cover", τα χαρακτηριστικά μουσικών πληροφοριών, όπως η κατανομή σημείων, οι δομές αρμονίας και η τάση αλλαγής σημείων, συλλέγουν βασικά χαρακτηριστικά που περιγράφουν τη μουσική. Έτσι προτείνεται μια εκπροσώπηση μουσικής, η οποία θα ενσωματώνει τις σχετικές μουσικές πληροφορίες, παρόμοιας της Unal et al [1] και της Hatisma et al [2].

Η ικανότητα της εκτέλεσης μετρήσεων ομοιότητας που έχουν νόημα για την εφαρμογή που στοχεύεται, είναι πολύ βασική για το σχεδιασμό μουσικών

αποτυπωμάτων. Προσεγγίσεις αυτού είναι η Mandel et al[3], η Lee[4], η Unal[5], η Ellis et al[6].

Η εφαρμογή κλασικής μουσικής που χρειάζεται για να αναγνωριστεί ένα κομμάτι, θα πρέπει να λαμβάνει υπ' όψιν την παρουσία πολλών διαφορετικών εκδόσεων ενός τραγουδιού, τα οποία πιθανόν να διαφέρουν στο κλειδί, στο ρυθμό, στους αγωγούς ή στους "players" και στην ενορχήστρωση. Το μουσικό αυτό αποτύπωμα επιδιώκει να συλλάβει και να μοντελοποιήσει τα βασικά μουσικά χαρακτηριστικά που είναι ανθεκτικά σε αυτές τις παραλλαγές.

4.1.1 Χαρακτηριστικό βασισμένο στο χρώμα του ήχου

Τα χαρακτηριστικά χρώματος που χρησιμοποιούνται σε αυτήν την περίπτωση, βασίζονται στο μοντέλο του Shepard helix[7], που παραγοντοποιεί την αντίληψη της συχνότητας σε ύψος τόνου και χρώματος :

$f = 2^{h+c}$ h ∈ Z, c ∈ [0,1)

Τα h,c,f, δηλώνουν το ύψος τόνου, χρώμα και συχνότητα. Το χρωματογράφημα μπορεί να υπολογιστεί κάνοντας πρώτα μια σύντομη ανάλυση φάσματος ισχύος :

$$x_c(t) = \sum_k s\ (t, 2^{c+k})$$

όπου το s (t,2c+k) είναι ένα σύντομο φάσμα ισχύος σε χρόνο t. Γίνεται κβαντοποίηση χρώματος σε 12 επίπεδα αποδόσεων παραγωγής, ένα διάνυσμα 12 διαστάσεων x(t) πολύ παρόμοιο με τις κατηγορίες χρωματικού βήματος δυτικού τύπου (A έως G #). Αυτές οι ποσότητες είναι το χαρακτηριστικό διανύσματος του χρώματος και χρησιμεύουν στην επεξεργασία ήχου μουσικής[1,4,6]. Τα στοιχεία του διανύσματος αντιπροσωπεύουν την ενέργεια για την αντίστοιχη τάξη βήματος στην χρονική στιγμή t. Υπάρχουν φορές που χρησιμοποιείται ο δείκτης n σαν διακριτός και όχι συνεχόμενος χρόνος (t). Οι Ellis και Poliner έχουν προτείνει ένα χαρακτηριστικό χρώματος χρησιμοποιώντας ένα παράθυρο ανάλυσης συγχρονισμού παλμών, ανθεκτικό στις παραλλαγές ρυθμού[6]. Βάση αυτού δημιουργείται το μουσικό αρχείο της παρούσας εργασίας.

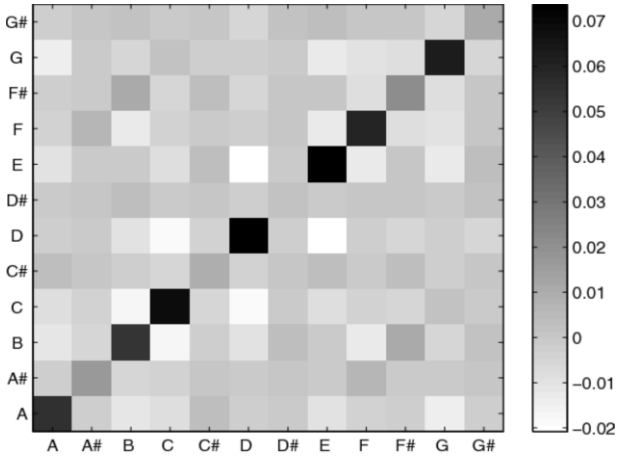
4.1.2 Μουσικό αποτύπωμα

Είναι προτιμότερο να υπάρχει λιγότερη μνήμη και πολυπλοκότητα, όπως και μεγαλύτερη ακρίβεια στην καταγραφή των μοναδικών χαρακτηριστικών της μουσικής. Μια απλή μήτρα συνδιακύμανσης, από τα διανύσματα χαρακτηριστικών χρώματος συγχρονισμένου βήματος προτείνεται σαν μουσικό αποτύπωμα :

$$\Phi = E[(x - E[x])(x - E[x])^T$$

όπου το Τ αντιπροσωπεύει τη μήτρα που μεταφέρεται.

Το **Σχήμα 1** δείχνει ένα παράδειγμα του αποτυπώματος. Τα στοιχεία της μήτρας είναι μία ποσότητα που συνδέεται με την ενέργεια. Τα διαγώνια στοιχεία του πίνακα αντιπροσωπεύουν το βαθμό παρουσίας κάθε τάξης βήματος ενέργειας. Οι στήλες δείχνουν το βαθμό συσχέτισης κάθε τάξης βήματος με δεδομένη κατηγορία αυτής. Έτσι φαίνεται η δομή της αρμονίας.

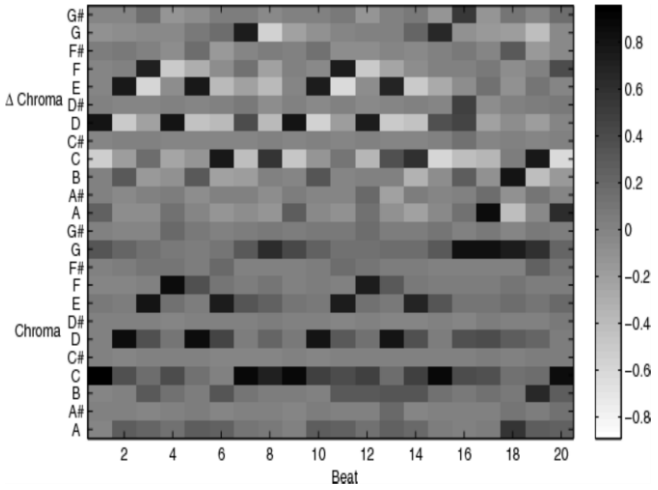


Σχήμα 1 Παράδειγμα αποτυπώματος με χρήση χαρακτηριστικών διανυσμάτων χρώματος.

Για να μοντελοποιηθούν οι δυναμικές χρονικές πληροφορίες, χρησιμοποιείται ο υπολογισμός των χαρακτηριστικών δέλτα. Για να γίνει πιο απλό γίνεται χρήση ενός μόνο διανύσματος χαρακτηριστικών από το γειτονικό βήμα για να μοντελοποιηθεί η χρονική πληροφορία.

$$\Delta x(n) = x(n+1) - x(n)$$

Η εξίσωση είναι η δυναμική μεταξύ δύο διαδοχικών βημάτων. Το **Σχήμα 2** δείχνει παραδείγματα από διανύσματα χαρακτηριστικών χρώματος και χρώματος δέλτα. Το κάτω μισό είναι τα διανύσματα χρώματος και το πάνω χρώματος δέλτα. Οι θετικές και οι αρνητικές τιμές δηλώνουν την επιλεγμένη ένταση και την ένταση απελευθέρωσης, αντίστοιχα.



Σχήμα 2 Παράδειγμα χαρακτηριστικών διανυσμάτων δέλτα χρώματος.

Για την κατασκευή του μουσικού αρχείου γίνεται δημιουργία ενός υπερ διάνυσμα από τη μίξη των δύο χρωμάτων, όπου ο υπολογισμός της μήτρας είναι :

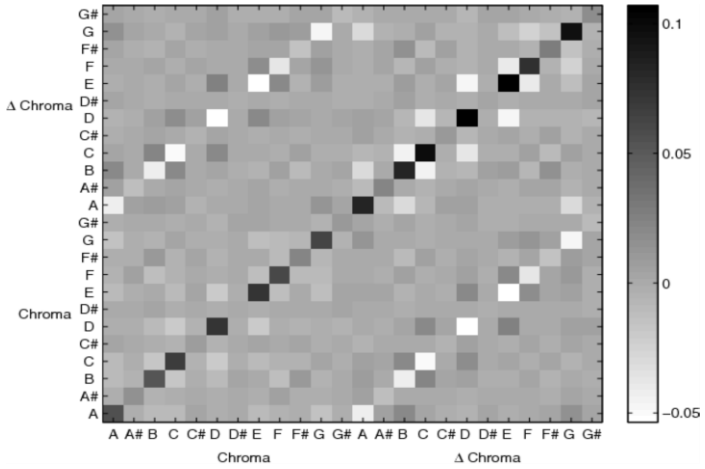
$$\Phi_{\Delta} = E[(x_{\Delta} - E[x_{\Delta}])(x_{\Delta} - E[x_{\Delta}])^{T}]$$

όπου

$$x_{\Delta} = \begin{matrix} x \\ \Delta x \end{matrix}$$

Το Σχήμα 3 είναι ένα παράδειγμα. Αφού το μήκος του διανύσματος διπλασιάζεται, το μέγεθος του αποτυπώματος είναι 4 φορές μεγαλύτερο. Το πρώτο τεταρτημόριο αντιπροσωπεύει τη μήτρα των διανυσμάτων χαρακτηριστικών δέλτα χρωμάτων. Οι ποσότητες σχετίζονται με τις χρονικές μεταβολές της έντασης, Τα διαγώνια στοιχεία δηλώνουν το βαθμό των χρονικών μεταβολών στην ένταση των τάξεων βήματος. Τα στοιχεία των στηλών δείχνουν το βαθμό των χρονικών αλλαγών που συμβαίνουν με μια δεδομένη τάξη βήματος ταυτόχρονα. Οι θετικές τιμές αντιπροσωπεύουν συμβάντα που έχουν οριστεί για την αντίστοιχη τάξη βήματος, ενώ οι αρνητικές τιμές αντιπροσωπεύουν γεγονότα απελευθέρωσης.

Το δεύτερο ή το τέταρτο τεταρτημόριο, δείχνει τη μήτρα μετασχηματισμού μεταξύ των διανυσμάτων χρωμάτων και των δέλτα χρωμάτων. Κάθε διάνυσμα περιγράφει τις κινήσεις των σημείων που ακολουθούν. Αν η τιμή είναι θετική, υπάρχει τάση να εμφανίζεται το σημείο μετά το δεδομένο σημείο, ενώ αν είναι αρνητική υπάρχει τάση απελευθέρωσης μετά από αυτό.



Σχήμα 3 Παράδειγμα υπερ διανυσμάτων.

4.1.3 Πειράματα

α) Διάταξη του πειράματος

Στην βάση δεδομένων του πειράματος χρησιμοποιήθηκαν 107 ξεχωριστά κομμάτια κλασικής μουσικής. Καταγράφηκαν στη μορφή MIDI, και το ηχητικό σήμα για κάθε ένα δημιουργήθηκε χρησιμοποιώντας την εργαλειοθήκη Timidity ++[8], με ρυθμό δειγματοληψίας 16kHz. Κάθε κομμάτι της μουσικής έχει δύο διαφορετικές εκδόσεις. Χρησιμοποιείται η μία ως ερώτημα και η άλλη ως αναφορά. Για το βασικό σύστημα, γίνεται χρήση του συστήματος Elis et al που πρότεινε μια αλληλοσυσχέτιση.

β) Μέτρηση ομοιότητας

Υπολογισμός ομοιότητας μουσικής i και j :

$$S_{ij} = \sum_{k} \sum_{l} \Phi_{kl}^{(i)} \Phi_{kl}^{(j)}$$

όπου φkl αντιπροσωπεύει το $k-th$ και το $l-th$ στοιχείο (γραμμή) του μουσικού αποτυπώματος Φ. Για να αντισταθμιστεί η πιθανή μεταφορά του κλειδιού, μετατοπίζουμε κυκλικά ένα αποτύπωμα στη διαγώνια κατεύθυνση με ένα ημιτονικό βήμα για να πάρουμε τη μέγιστη τιμή ομοιότητας :

$$S_{ij} = \max_{m} \sum_{k} \sum_{l} \Phi_{kl}^{(i)} \Phi_{kl}^{m(j)} \qquad ; 0 \le m \le 11,$$

όπου**,**

$$\boldsymbol{\Phi_{kl}^{m} = \Phi_{mod\left(\frac{k+m}{12}\right) mod\left(\frac{l+m}{12}\right)}}$$

και mod (·) αντιπροσωπεύει το μέτρο της διαίρεσης. Στην περίπτωση δέλτα χρωμάτων, η διαδικασία μετατόπισης γίνεται ξεχωριστά σε κάθε τεταρτημόριο.

Για την κανονικοποίηση της μήτρας συνδιακύμανσης, ώστε να ληφθεί η κατάληλη μέτρηση ομοιότητας, πρέπει να γίνει σωστή επιλογή του σχεδίου κανονικοποίησης, που εξαρτάται από το είδος των πληροφοριών που τονίζει η εφαρμογή :

$$S_{ij} = \max_{m} \sum_{k} \sum_{l} N(\Phi_{kl}^{(i)}) N(\Phi_{kl}^{m(j)})$$

όπου N (·) αντιπροσωπεύει τον επιλεγμένο αλγόριθμο κανονικοποίησης. Έτσι δύο τύποι κανονικοποίησης έχουν χρησιμοποιηθεί : μια ολική κανονικοποίηση (ON), που εξετάζει συνολική κατανομή ενέργειας κάθε σημείου και της συνύπαρξής τους διαιρώντας το τετραγωνικό τους άθροισμα στο αποτύπωμα και μια κανονικοποίηση σε στήλη (CN), που δίδει έμφαση στη δομή της αρμονίας της μουσικής διαιρώντας το τετραγωνικό άθροισμα στη στήλη του αποτυπώματος:

$$\text{(ON)} \ : \quad N(\Phi_{kl}) = \frac{\Phi_{kl}}{\sqrt{\sum_m \sum_n (\Phi_{mn})^2}}$$

$$\text{(CN)} \ : \quad N(\Phi_{kl}) = \frac{\Phi_{kl}}{\sqrt{\sum_m (\Phi_{ml})^2}}$$

γ) Αποτελέσματα

	[7]	Fingerprint w/ ON (11)	Fingerprint w/ CN (12)
Accuracy (%)	59.6	68.6	**80.7**
Approx. Searching Time (sec)	386	6	6

(1)

	[7]	[7] w/ $\mathbf{x}_\Delta$	FP-ON w/ $\mathbf{x}_\Delta$	FP-CN w/ $\mathbf{x}_\Delta$
Accuracy (%)	59.6	65.1	77.1	**85.3**
Approx. Searching Time (sec)	386	845	23	23

(2)

Ο **πίνακας 1** δείχνει την ακρίβεια και την ταχύτητα αναζήτησης του συστήματος σε σύγκριση με το βασικό σύστημα. Η προσέγγιση αυτή επιταχύνει την ταχύτητα αναζήτησης κατά 60 φορές περίπου με βελτίωση σχετικής ακρίβειας 30%.

Ο **πίνακας 2** δείχνει την απόδοση αποτυπώματος με υπερ-διανύσματα χαρακτηριστικών χρώματος και δέλτα χρώματος. Έχουμε 40% βελτίωση της ακρίβειας, και έναν παράγοντα 20 επιταχύνσεων σε σύγκριση με το συμβατικό σύστημα. Υπάρχουν

επίσης βελτιώσεις στην απαίτηση μνήμης αποθήκευσης, το σύστημα αποθηκεύει μόνο το μουσικό αποτύπωμα που είναι 576 Byte για κάθε τραγούδι και στην περίπτωση δέλτα χρωμάτων, απαιτείται 1728 Byte.

4.2 Διανύσματα χαρακτηριστικών δυναμικού χρώματος με εφαρμογές στην αναγνώριση "cover"

Μία νέα εκδοχή δυναμικού διανύσματος χαρακτηριστικών χρώματος προτείνεται εμπνευσμένο από ψυχοφυσικές παρατηρήσεις. Η Warren et al[1], χρησιμοποίησε λειτουργικό σύστημα απεικόνισης μαγνητικού συντονισμού, για να δείξει τις ψυχοφυσικές επιδράσεις των αλλαγών του βήματος στον ανθρώπινο εγκέφαλο χειρίζοντας το βήμα ενός δεδομένου σήματος. Η απεικόνιση της αλλαγής βήματος του χρώματος έγινε στον τον πρόσθιο έως τον πρωταρχικό ακουστικό φλοιό, ενώ η αλλαγή ύψους βήματος απεικονίστηκε στον οπίσθιο στον πρωταρχικό ακουστικό φλοιό.

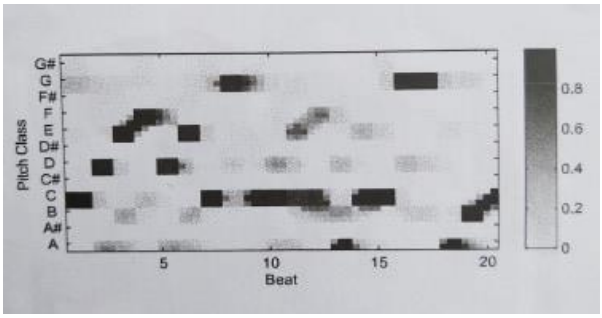
Μετά από έρευνες που έγιναν με εφαρμογή αναγνώρισης τραγουδιών "cover" κλασικής μουσικής, προτείνεται ένας αλγόριθμος που επιχειρεί να μοντελοποιήσει σχετικές μεταβολές χρωμάτων σε συνδυασμό με τα συμβατικά διανύσματα χαρακτηριστικών χρωμάτων.

4.2.1 Ο αλγόριθμος

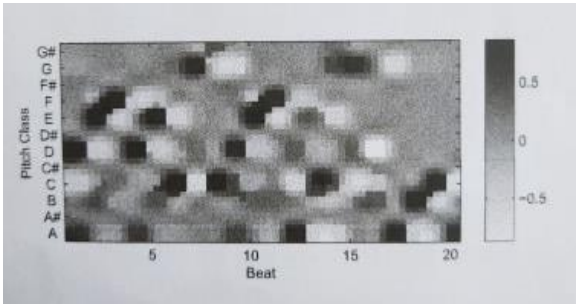
α) Το προτεινόμενο χαρακτηριστικό δέλτα χρώματος

Η χρωματική αλλαγή ερμηνεύεται ως ένα σχετικό διάστημα μεταξύ των τάξεων βημάτων που παίζονται σε όρους ημιτονίου. Αν η τάξη βήματος "D" αναπαράγεται μετά του "C", το διάστημα αυτό είναι +2 ημιτόνια. Αν η μελωδία είναι μονοφωνική, η αλλαγή είναι μια κλιμακωτή τιμή. Όταν έχουμε πολυφωνικό σήμα, οι αλλαγές είναι πολλές και γίνονται ταυτόχρονα. Αν οι τάξεις "C" και "G" παίζονται ταυτόχρονα μετά του "D", το διάστημα αλλαγής μπορεί να είναι -2, αλλά και +5 ημιτονίων. Για να αντιμετωπιστεί αυτό απαιτείται αναπαράσταση διανύσματος.

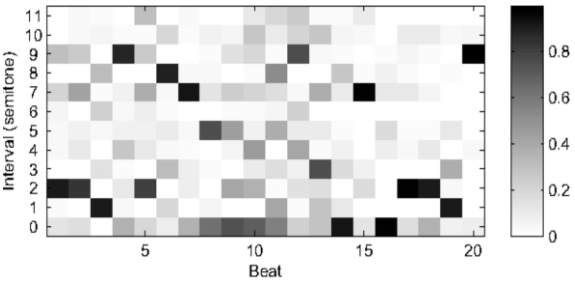
Στην εξίσωση $\Delta c[n] = c[n+1] - c[n]$, $\Delta c[n]$ είναι η ευκλείδεια απόσταση μεταξύ 2 γειτονικών διανυσμάτων. Μπορεί επίσης να ερμηνευτεί ως η αλλαγή χρώματος με μηδενικό διάστημα (καμία αλλαγή).



α) Διανύσματα χαρακτηριστικών χρώματος



β) Διανύσματα χαρακτηριστικών χρώματος δέλτα



γ) Προτεινόμενα διανύσματα χαρακτηριστικών χρώματος δέλτα

Μπορούν να ληφθούν παρόμοιες ποσότητες λαμβάνοντας υπόψη οποιοδήποτε διάστημα αλλαγής i με κυκλική περιστροφή του τελευταίου διανύσματος :

$$\left\|\Delta c^i[n]\right\| = \left\|c^i[n+1] - c[n]\right\| \quad ; 0 \leq i \leq 11$$

όπου c^i αντιπροσωπεύει το περιστρεφόμενο διάνυσμα c του οποίου τα στοιχεία μετακινούνται κυκλικά από i ημιτόνια. Η τιμή αντιπροσωπεύει την απίθανη μετακίνηση προς το i. Για απλότητα, ορίζεται το εύρος του i όπως στην παραπάνω εξίσωση. Το i είναι συντελεστής του 12, έτσι ένα -2 διάστημα μπορεί να ερμηνευτεί ως + 10.

Έτσι ορίζεται μια αναπαράσταση διανύσματος, που δείχνει την πιθανότητα μετακίνησης προς μεμονωμένα διαστήματα αλλαγής. Θα χρειαστεί μια συνάρτηση που μετατρέπει το απίθανο σε πιθανό. Θέτεται ένα πλην (-) και γίνεται πρόσθεση της μέγιστης τιμής μεταξύ των στοιχείων για να δημιουργηθεί ένα διάνυσμα θετικών στοιχείων :

$$\nabla \mathbf{c}[\boldsymbol{n}] = \{\nabla \mathbf{c}_0[\boldsymbol{n}], \nabla \boldsymbol{c}_1[\boldsymbol{n}], \ldots, \nabla \boldsymbol{c}_{11}[\boldsymbol{n}]\}^{\boldsymbol{T}}$$

όπου,

$$\nabla \boldsymbol{c_i}[\boldsymbol{n}] = -\left\|\Delta \boldsymbol{c^i}[\boldsymbol{n}]\right\| + \boldsymbol{C_{max}}$$

και

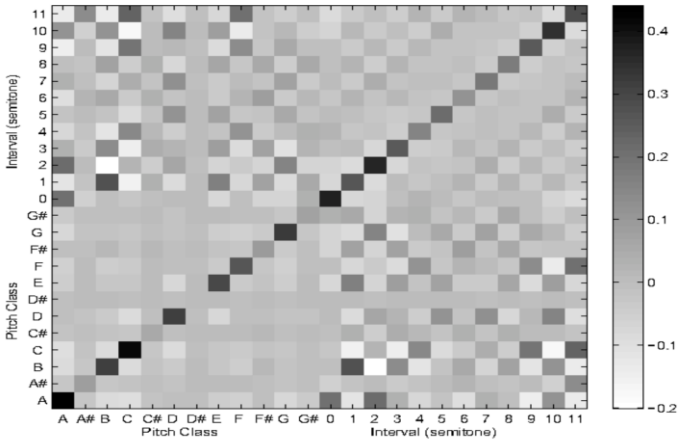
$$C_{max} = \max_j \left\|\Delta c^j[n]\right\|$$

β) Μουσικό αποτύπωμα

Υπάρχει ο αλγόριθμος από προηγούμενη εργασία[2] :

$$\Phi = E[(x - E[x])(x - E[x])^T],$$

Όπου το Τ δηλώνει τη μετάθεση της μήτρας. Εάν το x είναι υπερ διάνυσμα του διανύσματος χρώματος και του διανύσματος δέλτα χρώματος, δηλαδή,$x = [c^T \Delta c^T]^T$, το αρχείο συμπιέζει επιπλέον γνωρίσματα, όπως πληροφορίες χρονικής αλλαγής μετά την αναπαραγωγή μίας τάξης βήματος.



Σχήμα 1 Μουσικό αποτύπωμα με χρήση του προτεινόμενου διανύσματος χαρακτηριστικού χρώματος δέλτα

Αν γίνει χρήση$x = [c^T \nabla c^T]^T$, αυτό που προκύπτει αρχικά είναι πως οι άξονες του αποτυπώματος δεν αποτελούνται μόνο από τις τάξεις βήματος αλλά και από τα σχετικά διαστήματα. Στο πρώτο τεταρτημόριο της εικόνας, τα διαγώνια στοιχεία αντιπροσωπεύουν τις εντάσεις των αλλαγών χρώματος, κάθε διάνυσμα περιγράφει πώς συμβαίνουν οι αλλαγές χρώματος ταυτόχρονα με την αντίστοιχη αλλαγή χρώματος.

Στο δεύτερο τεταρτημόριο , κάθε διάνυσμα απεικονίζει σε ποια κατεύθυνση συμβαίνει η αλλαγή χρώματος μετά την αναπαραγωγή της αντίστοιχης τάξης βήματος. Όσο μεγαλύτερη είναι η τιμή, τόσο ισχυρότερη είναι η τάση για τις τάξεις βήματος που παίζονται ταυτόχρονα με την αντίστοιχη να κινηθούν προς το αντίστοιχο διάστημα. Μετά την αναπαραγωγή της τάξης "A", υπάρχει η τάση τα σημεία που παίζονται με αυτή να διατηρηθούν ή να μετακινηθούν 2 ημιτόνια πάνω.

γ) Μέτρηση ομοιότητας

Γίνεται χρήση μιας προσέγγισης αντιστοίχισης για να μετρηθεί η ομοιότητα δύο αποτυπωμάτων, π.χ. i και j :

$$S_{ij} = \sum_{k=1}^{24} \sum_{l=1}^{24} \Phi_{kl}^{(i)} \Phi_{kl}^{(j)}$$

Όπου $\Phi_{kl}^{(i)}$ αντιπροσωπεύει το στοιχείο της σειράς $k-th$ και στήλης $l-th$ του αποτυπώματος Φ του κομματιού i.

Όσο μεγαλύτερη η τιμή είναι μεγαλύτερη η ομοιότητα. Ακόμη και αν οι μουσικές είναι ίδιες το κλειδί μπορεί να μεταφερθεί. Λαμβάνεται η μέγιστη τιμή ομοιότητας μεταξύ των πιθανών μετατοπίσεων για να αντισταθμιστεί. Στα διανύσματα δέλτα χρώματος, γίνεται κυκλική μετακίνηση κάθε τεταρτημορίου σε διαγώνια κατεύθυνση. Δεδομένου ότι το διάστημα αλλαγής είναι μια σχετική τιμή και ανεξάρτητη από το κλειδί, δεν πρέπει να μετακινηθεί κατά τη διαδικασία αντιστάθμισης.

δ) Αποτελέσματα πειράματος

Στη βάση δεδομένων υπάρχουν 2000 κομμάτια κλασσικής μουσικής. Καταγράφηκαν στη μορφή MIDI[3] και το κάθε ένα δημιουργήθηκε χρησιμοποιώντας Timidity++[4] με ρυθμό δειγματοληψίας 16kHz. Όσα τραγούδια ξεπερνούν τα 5 λεπτά, κόπηκαν στα 5 ακριβώς. Η μία από τις δύο εκδόσεις χρησιμοποιείται ως ερώτημα και η άλλη ως αναφορά. Η ταξινόμηση, γίνεται με βάση τη μέγιστη βαθμολογία ομοιότητας μεταξύ του συνόλου δεδομένων :

$$arg \max_j S_{ij}$$

	Chroma	Delta	Proposed Delta
Accuracy (%)	74.7	76.7	**79.0**
Realtive Improvement (%)	-	2.5	**5.8**

Πίνακας 1

	Delta	Proposed Delta
Φ_{1st}	**62.2**	51.7
Φ_{2nd} or Φ_{4th}	69.5	**75.2**
Φ_{3rd}	74.7	74.7
Φ_{all}	76.6	**79.0**

Πίνακας 2

Ο **Πίνακας 1** δείχνει την ακρίβεια αναγνώρισης σύμφωνα με τους τύπους χαρακτηριστικών. Εξηγεί τη βελτίωση της απόδοσης χρησιμοποιώντας τις χρονικές δυναμικές πληροφορίες στα διανύσματα χρωμάτων. Υποδεικνύει ότι οι δυναμικές πληροφορίες μπορούν να προσφέρουν συμπληρωματικά χαρακτηριστικά για να συλληφθούν τα μοναδικά γνωρίσματα του κομματιού.

Ο **Πίνακας 2** δείχνει μια βαθιά ανάλυση, βάση των περιπτώσεων χρήσης του καθενός τεταρτημόριου. Δεδομένου ότι αυτά περιλαμβάνουν διαφορετικές πτυχές των μουσικών χαρακτηριστικών, εκτελούμε με αυτά ζητήματα αναγνώρισης. Το Φ υποδεικνύει ότι χρησιμοποιούμε το αντίστοιχο τεταρτημόριο, ώστε να ερευνηθούν οι επιδράσεις των μουσικών ιδιοτήτων που είναι ενσωματωμένες σε αυτό.

4.2 Αναγνώριση τραγουδιών "cover" μεγάλης κλίμακας με χρήση κατακερματισμένων σημείων χρώματος

Οι τεχνικές αναγνώρισης μπορεί επίσης να χρησιμοποιηθούν για να βρεθούν πρότυπα και δομή σε μεγάλα σύνολα μουσικών δεδομένων. Κυκλοφόρησε πρόσφατα το Million Song Dataset (MSD)[1] και μια συμπληρωματική λίστα τραγουδιών, το SecondHandSongs dataset (SHSD). Το SHSD απαριθμεί 12.960 "cover" και 5.236 δοκιμαστικά "cover", μέρη του MSD. Το MSD περιέχει χαρακτηριστικά ήχου και άλλα, για 1.000.000 κομμάτια.

Σε αυτήν τη εργασία γίνεται αναζήτηση μιας σειράς χαρακτηριστικών εμπνευσμένης από αποτυπώματα, που μπορούν να χρησιμοποιηθούν ως κωδικοί κατακερματισμού. Στο αποτύπωμα Shazam, ο Wang[2] προσδιορίζει τα σημεία στο σήμα, που είναι οι αναγνωρίσιμες κορυφές και κάνει μέτρηση της μεταξύ τους απόστασης. Αυτό αποτελεί ένα πολύ ακριβές αναγνωριστικό.

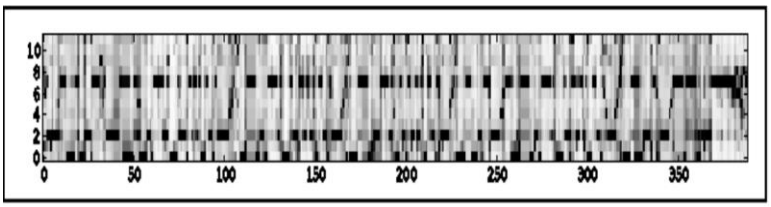
Ένα σύστημα κατακερματισμού περιέχει δύο μέρη : Την εξαγωγή, που υπολογίζει τους κωδικούς κατακερματισμού, μια αργή διαδικασία και τη σύγκριση κωδικών κατακερματισμού.

4.3.1 Σύστημα “hashing”

Σε αυτήν την εργασία αναλύεται πώς να γίνει μετάβαση από μια μήτρα χρώματος σε μερικές δωδεκάδες ακέραιους για κάθε τραγούδι.

α) Αναπαράσταση δεδομένων

Η ανάλυση χαρακτηριστικών που χρησιμοποιείται βασίζεται στο API της Echo Nest[3]. Ένα χρωματογράφημα (Σχήμα 1) είναι παρόμοιο με φασματογράφημα σταθερού Q εκτός αν το περιεχόμενο βήματος διπλώνεται σε μία οκτάβα 12 ξεχωριστών δοχείων, που αντιστοιχούν σε ένα συγκεκριμένο ημιτόνιο. Για κάθε τραγούδι στο MSD, ο Echo Nest δίνει ένα διάνυσμα χρώματος (μήκος 12) για κάθε μουσικό γεγονός, και μια κατάτμηση του τραγουδιού σε παλμούς, οι οποίοι μπορεί να καλύπτουν ή να υποδιαιρούν τμήματα. Υπολογίζεται η μέση τιμή ανά τμήμα χρόνου παλμού χρώματος δίνοντας μια αναπαράσταση συγχρονισμένου παλμού. Τα διανύσματα χρώματος Echo Nest ρυθμίζονται για να έχουν σε κάθε στήλη μεγαλύτερη τιμή 1.



Σχήμα 1 Ευθυγραμμισμένοι παλμοί χρωμάτων του τραγουδιού "Wild Rover" από τον Dropkick Murphys.

Όταν ευθυγραμμίζονται τα διανύσματα χρώματος στους παλμούς γίνεται να τα επαναρυθμιστούν με πολλούς τρόπους. Εδώ ρυθμίζεται με το μέγιστο.

β) Ο κωδικός "hash"

Βάση του Σχεδίου 1 εντοπίζονται τα σημεία αναφοράς, χρησιμοποιώντας ένα προσαρμοστικό κατώφλι :

Αρχικοποίηση T, το διάνυσμα κατωφλίου μεγέθους 12 ως μέγιστο ανά ημιτόνιο κατά τα 10 πρώτα χρονικά πλαίσια.

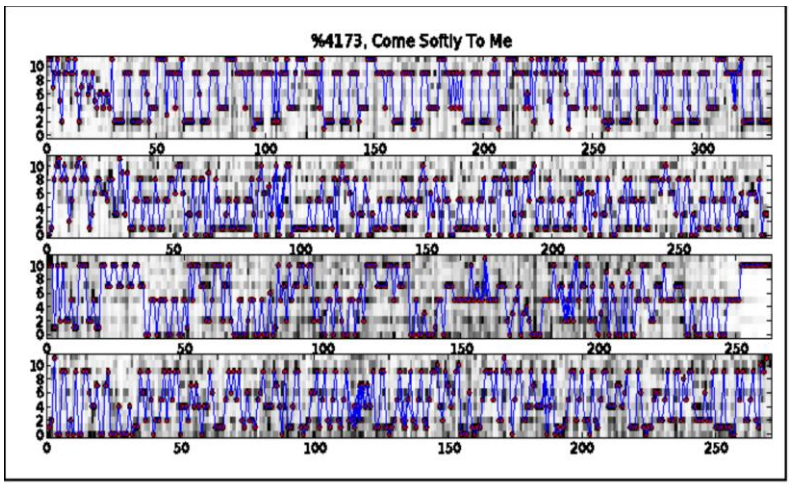
Τον χρόνο t, θεωρείται ένα δοχείο ως σημείο (u), αν η τιμή του ξεπερνά το κατώφλι.

$T_u = u$, και $T_i = max(T_u * \psi T_i), \psi$ ο παράγοντας αποσύνθεσης. Περιορίζεται ο αριθμός των σημείων ανά χρονικό πλαίσιο σε 2.

Μετακινώντας από t σε$t + 1$, χρησιμοποιείται ο παράγοντας αποσύνθεσης ψ, έχοντας $T_{t+1} = T_t * \psi$.

Εντοπίζονται σημεία δια μέσου εμπρός και πίσω φάσης και εξετάζονται όλοι οι δυνατοί συνδυασμοί, δηλαδή σύνολο αλμάτων σε παράθυρο W μεγέθους. Το Σχήμα 2 είναι μια

απεικόνιση αυτών, μια λίστα διαφορών σε χρόνο και ημιτονίου μεταξύ σημείων αναφοράς με ένα αρχικό ημιτόνιο επιπλέον.



Σχήμα 2 Σημεία και άλματα για 4 τραγούδια της εργασίας %4173 του SHSD. Γίνεται χρήση "jumpcode 2", πίνακας 1.

γ) Κωδικοποίηση και ανάκτηση

Χρησιμοποιώντας τη βάση δεδομένων SQLite των αλμάτων, γίνονται συγκρίσεις. Για να υπολογιστεί ένα περιστρεφόμενο άλμα, γίνεται χρήση του παρακάτω σχεδίου.

Σειρά αλμάτων k, μέγεθος παραθύρου χρόνου W. Άλματα (δοχείο χρώματος, χρονικό πλαίσιο): $(b_1, 0), (b_2, t_2), (b_k, t_k)$. Γίνεται υπολογισμός :

$$(b_1 - b_0) + (t_1 - 0 + 1) * 12$$

$$+ [(b_2 - b_1) + (t_2 - t_1 + 1) * 12 + 1] * 12 * (W + 1)$$

$$+ \ldots$$

Γίνεται να ανακτηθεί η τιμή αρχικού ημιτονίου με modulo 12. Κάνοντας αυτό και προσθέτοντας τιμή περιστροφής 0-11, μπορεί να γίνει περιστροφή άλματος.

Σε κάποια τραγούδια υπάρχουν περισσότερα άλματα, σε αυτά εκχωρείται ένα βάρος για να αντιμετοπιστεί, το οποίο είναι ο αριθμός των εμφανίσεων ενός άλματος διαιρούμενος με το λογάριθμο του συνολικού αριθμού των αλμάτων.

Στην ανάκτηση προκύπτουν όλα τα άλαματα για όλα τα τραγούδια. Εάν δωθεί ένα τραγούδι σαν ερώτημα λανβάνονται τα ζευγάρια που 1) το άλμα ανήκει επίσης στο τραγούδι αυτό και 2) το βάρος αυτού εμπίπτει σ' ένα περιθώριο γύρω από αυτό το βάρος. Γίνεται υπολογισμός με ποσοστό $\propto$ ως $βάρος_{ερώτημα} * (1-\propto) \leq βάρος \leq$

$\beta\acute{\alpha}\rho o s_{\epsilon\rho\acute{\omega}\tau\eta\mu\alpha} * (1+\propto)$. Έπειτα υπολογίζεται η συχνότητα, που όσο μεγαλύτερη είναι τόσο πιθανό να είναι το "cover" του τραγουδιού.

4.3.2 Πειράματα

Δημιουργήθηκε ένα σύνολο 500 δυαδικών ζητημάτων από τα 12.960 τραγούδια από το σετ εκπαίδευσης. Ο αλγόριθμος επιλέγει από 2 το σωστό "cover" μετά από ερώτημα.

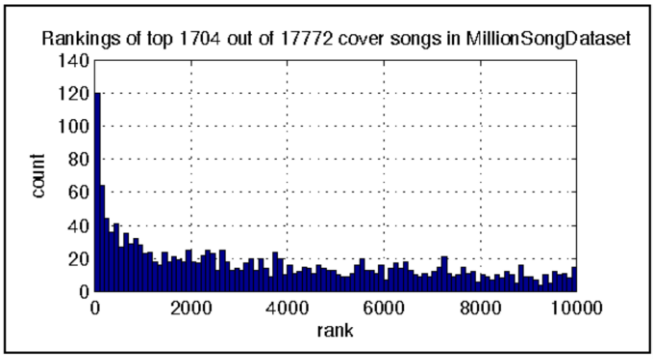
Παρουσιάζονται 2 ρυθμίσεις. Η πρώτη (jumpcodes 1) έχει: $\psi = 0.96, W = 6, \alpha = 0.5$, σημεία ανά άλμα: 1 και 4, ευθυγράμμιση χρώματος σε κάθε παλμό. Δίνει πολλά άλματα και είναι πολύπλοκη. Η δεύτερη : $\psi = 0.995, W = 3, \alpha = 0.6$, σημεία ανά άλμα: 3, ευθυγράμμιση χρώματος ανά 2 παλμούς. Πολύ λιγότερα άλματα.

	accuracy	#hashes
random	50.0%	-
jumpcodes 1	79.8%	11,794
jumpcodes 2	**77.4%**	176
correlation	76.6%	-

Πινακας 1 Αποτελέσματα συνόλου 500 ερωτημάτων.

Υπολογίζεται η συσχέτιση μεταξύ των αναπαραστάσεων χρώματος ευθυγραμμισμένων με τους παλμούς, συμπεριλαμβανομένων περιστροφών και χρονικών αντισταθμίσεων. Η μέθοδος είναι ενστικτώδης, εκτελεί ομοίως.

Στο Σχήμα 3, φαίνεται η κατάταξη των 1,704 από τα 17, 771 ζεύγη ερωτημάτων (9,6%)(τραγούδι ερώτημα, τραγούδι στόχος), στο κορυφαίο 1% (10.000 κομμάτια) των εκατομμυρίων τραγουδιών.



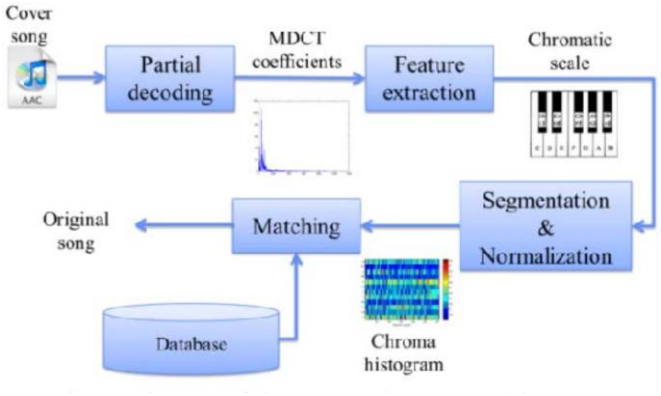
Σχήμα 3 Ιστόγραμμα της κατάταξης των 1,704 ζευγών ερωτημάτων των οποίων η τάξη είναι κάτω από 10.000

4.4 Αναγνώριση "cover" με άμεση εξαγωγή χαρακτηριστικών χρώματος από αρχεία AAC

Στην εργασία αυτή προτείνεται μια μέθοδος εξαγωγής χαρακτηριστικών που αυξάνει την ταχύτητα επεξεργασίας, προέρχεται από τα αρχεία AAC. Περιλαμβάνει χαρτογράφηση παραμέτρου σε χρώμα. Ο συντελεστής τροποποιημένου διακριτού μετασχηματισμού συνημίτονου (MDCT) χρησιμοποιείται άμεσα για την εξαγωγή χαρακτηριστικών.

4.4.1 Το προτεινόμενο σύστημα

Στο Σχήμα 1 το αρχείο AAC αποκωδικοποιείται όχι όμως ολοκληρωτικά για τον εντοπισμό των συντελεστών MDCT, που χαρτογραφούνται σε ένα 12-διάστατο χαρακτηριστικό χρώματος. Για να μειωθεί η χρονική διάσταση στο χώρο, πολλά καρέ συγχωνεύονται σε ένα τμήμα. Ο δυναμικός προγραμματισμός τοπικού αλγόριθμου ευθυγράμμισης (DPLA), υπολογίζει την ομοιότητα αρχικού και "cover" τραγουδιού στην αντιστοίχιση.



Εικόνα 1 Διάγραμμα της αρχιτεκτονικής του προτεινόμενου συστήματος

α) Εξαγωγή χαρακτηριστικών

Το εργαλείο ανάλυσης χρόνου συχνότητας είναι το MDCT[1], ορίζεται ως :

$$X(k) = 2\sum_{n=0}^{N-1} x(n)\cos\left(\frac{2\pi}{N}\left(n+\frac{N}{4}+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\right)$$

x είναι η αλληλουχία εσαγωγής παραθύρου, n ο δείκτης δείγματος, k ο δείκτης συντελεστή φάσματος και N το μήκος του παραθύρου μετασχηματισμού.

Οι συντελεστές $X(k)$ δηλώνουν το μέγεθος ενέργειας μιας ζώνης συχνοτήτων και μπορεούν να κατανεμηθούν σε δοχείο χρώματος b[2] :

$$b = mod\left(round\left(12 \times log_2\left(\frac{f}{f_0}\right), 12\right)\right)$$

f η αντίστοιχη συχνότητα του $X(k)$ και f_0 είναι $16.352\,Hz$, που αντιπροσωπεύεται από το $C0$, η χαμηλότερη συχνότητα βήματος. Ένα χαρακτηριστικό χρώματος καταγράφει την ένταση ενέργειας που συσχετίζεται με κάθε ένα από τα 12 ημιτόνια $(C, C\,\#, \text{B}, A\,\#, B)$.

β) Τμηματοποίηση και Κανονικοποίηση

Στο πρότυπο MPEG-2 AAC, το σήμα εισόδου κωδικοποιείται ανά καρέ[1]. Ένα τραγούδι των 3,5 λεπτών έχει περίπου 3000 καρέ με ρυθμό δειγματοληψίας $16k\,Hz$. Στην προσπάθεια να συγχωνευθούν πολλά καρέ σε ένα τμήμα, φαίνεται ότι αυτό που έχει οριστεί σε 1 δευτερόλεπτο επιτυγχάνει τα βέλτιστα αποτελέσματα.

γ) Αντιστοίχιση

Το πρώτο βήμα είναι να δημιουργηθεί ένας δυαδικός πίνακας ομοιότητας (BSM) μεταξύ των 2 τραγουδιών. Το δεύτερο βήμα περιλαμβάνει τον αλγόριθμο

τοπικής ευθυγράμμισης δυναμικού προγραμματισμού για υπολογισμό βαθμού ομοιότητας ακόμα και σε αρκετούς ρυθμούς[3]. Στην Εικόνα 2 η μέγιστη τιμή της μήτρας DPLA, χρησιμοποιείται ως βαθμός ομοιότητας :

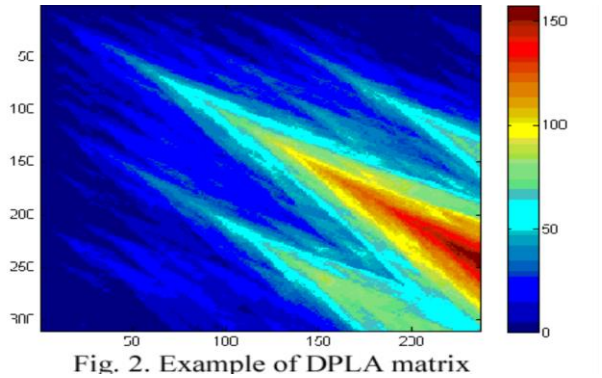


Fig. 2. Example of DPLA matrix

Εικόνα 2 Παράδειγμα της μήτρας DPLA

4.4.2 Αποτελέσματα πειραμάτων

Το πρότυπο MPEG-2 AAC[1] παρέχει πολλά εργαλεία ανάλυσης και υψηλής συμπίεσης στη διαδικασία κωδικοποίησης. Στο προτεινόμενο σύστημα, το χρώμα εξάγεται από τους συντελεστές MDCT χωρίς πλήρη αποκωδικοποίηση, γεγονός που μπορεί να εξοικονομήσει πάνω από 70% πολυπλοκότητα[4].

Ο **Πίνακας 1** δείχνει τις επιδόσεις του προτεινόμενου συστήματος και του συστήματος Ellis με Top-N τον αριθμό των σωστών τραγουδιών που βρέθηκαν στο N-υψηλότερο σκορ αντιστοίχισης που προέκυψαν. Το προτεινόμενο σύστημα έχει ακρίβεια 62%.

	Top-1	Top-3	Top-5	Top-10	Matching time(s)	Time saving
Ellis	59	65	65	67	1355	0%
Proposed	74	77	79	81	884	35%

4.5 "Live" Αναγνώριση τραγουδιών γνωστών καλλιτεχνών με χρήση ηχητικών "hashprints"

Στη ζωντανή αναγνώριση τραγουδιών κάποιος ηχογραφεί σ'ένα app ένα τραγούδι και παίρνει μια απάντηση, όπως το Shazam και το SoundHound.

Δύο προβλήματα απαρτίζουν αυτή τη διαδικασία. Η αποτύπωση ήχου που επιχειρεί να αναγνωρίσει ένα τμήμα ήχου σε μια βάση δεδομένων με καθαρές εγγραφές, αλλά αποτελεί πρόκληση το ότι τα ερωτήματα είναι σύντομα και θορυβώδη και το σύστημα πρέπει να λειτουργεί σε πραγματικό χρόνο, και η ανίχνευση τραγουδιού "cover" που προσπαθεί να εντοπίσει εκδόσεις του ίδιου τραγουδιού, η αντιστοίχιση όμως είναι κάπως ασαφής.

Τα αποτυπώματα ήχου αποτελούν ενδιαφέρον για τη βιομηχανία. Εταιρείες που παράγουν τέτοιες δουλειές είναι η Philips[1],[2], η Google[3],[4], η Telefonica[5],[6] και η Gracenote[7]. Επίσης η εργασία TRECVID που είναι βασισμένη στην ανίχνευση αντιγράφων περιέχει τέτοια αποτυπώματα[8]. Η εκτίμηση “cover” της MIREX[9], ανέπτυξε την ανίχνευση τους[10-13]. Η συλλογή 1000 τραγουδιών[14] βοήθησε στις συλλογές μεγάλης κλίμακας να ερευνηθούν[15-19]. Υπάρχουν ωστόσο και πολλές εργασίες στη αντιστοίχηση ήχου[20-23].

Η αναγνώριση "live" τραγουδιών βασισμένη σε ερωτήματα κινητών είναι ανεξερεύνητη. Ένας λόγος είναι η δυσκολία συλλογής ενός κατάλληλου συνόλου δεδομένων. Οι Rafii et al[24] προτείνουν μέθοδο βασισμένη σε μια διμερή αναπαράσταση του μετασχηματισμού σταθερού Q.

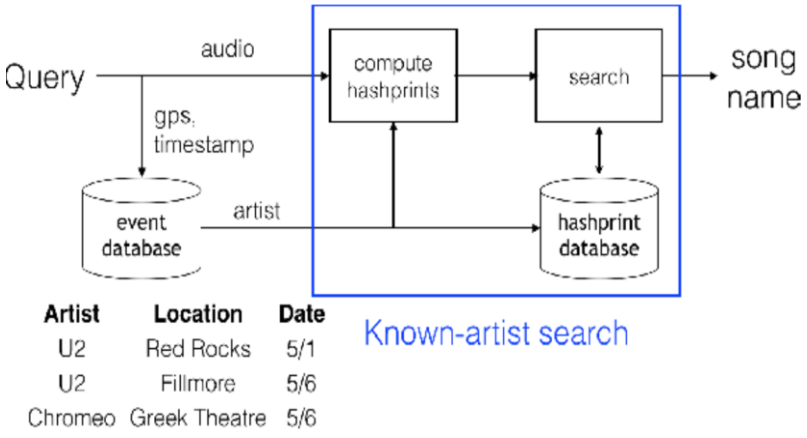
Σε αυτήν την εργασία προτείνεται μια λύση 2 στοιχείων: μιας δυαδικής αναπαράστασης ήχου που ονομάζεται hashprints, βασισμένη στην εκμάθηση ενός συνόλου φασματοχρονικών φίλτρων σε έναν ανεξάρτητο καλλιτεχνικό τρόπο και ενός απλού αλγόριθμου αντιστοίχισης που επιτρέπει σε κάποιον να ανταλλάξει την ακρίβεια για την αποτελεσματικότητα προκειμένου να προσαρμοστεί το μέγεθος της βάσης δεδομένων αναζήτησης του κάθε καλλιτέχνη.

4.5.1 Περιγραφή συστήματος

α) Αρχιτεκτονική

Όταν υποβάλλεται ένα ερώτημα, οι συντεταγμένες GPS και οι πληροφορίες χρονικής σήμανσης κάνουν τη συσχέτιση του με μια συναυλία για την εύρεση του καλλιτέχνη. Μετά γίνεται αναζήτηση του κομματιού.

Το σύστημα υποθέτει ότι το πρόγραμμα των συναυλιών είναι διαθέσιμο στο διαδίκτυο και πως ο καλλιτέχνης κάνει εκτέλεση από ένα ηχογραφημένο άλμπουμ σε στούντιο.



Σχήμα 1 Αρχιτεκτονική συστήματος αναγνώρισης τραγουδιού "live".

β) Αναπαράσταση “hashprint”

Κίνητρο: Στην εφαρμογή μας κάθε hashprint είναι ένας ενιαίος ακέραιος 64-bit που περιέχει έως και 64 bits πληροφοριών. Μπορεί να υπολογιστεί η απόσταση Hamming ανάμεσα σε δύο hashprints πολύ αποτελεσματικά, με την επιβολή μίας μοναδικής λογικής λειτουργίας xor μεταξύ 2 ακέραιων 64 bit και να μετρηθεί ο αριθμός των 1 bits στο αποτέλεσμα.

Η αναπαράσταση hashprint χαρακτηρίζεται από το να είναι συμπαγής (αποτελεσματική) και εύρωστη (ανθεκτική στο θόρυβο). Τα hashprints είναι αμετάβλητα στην ένταση, δηλαδή ένα σήμα θα αποδώσει την ίδια αναπαράσταση hashprint ακόμη και αν πολλαπλασιαστεί με έναν σταθερό παράγοντα.

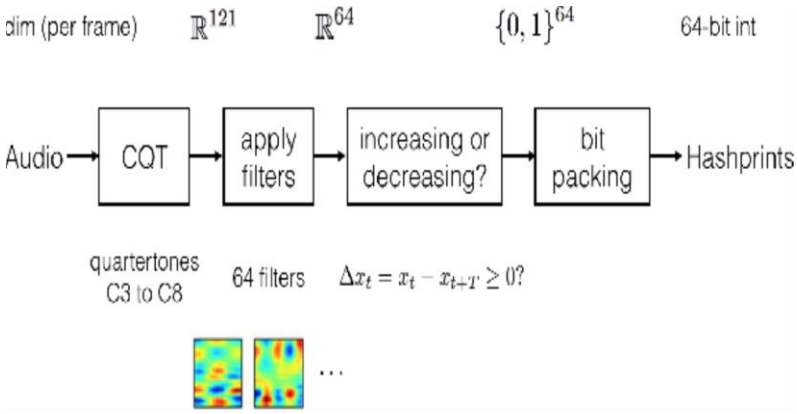
Μηχανική: Το πρώτο βήμα είναι να υπολογιστεί ένας σταθερός μετασχηματισμός Q (CQT), μια χρονική αναπαράσταση του ήχου από ένα σύνολο φίλτρων λογαριθμικά διαχωρισμένων με σταθερό συντελεστή Q που επιτρέπει να υπολογιστούν οι μετατοπίσεις βήματος. Στα πειράματά χρησιμοποιήθηκε η περίπτωση των Schorkhuber και Klapuri[25]. Θεωρούνται 24 υποζώνες ανά οκτάβα μεταξύ $C3$ $(130.81Hz)$ και $C8$ $(4186.01Hz)$. Με ανάλυση τετραγωνικού τόνου για τις αλλαγές και λαμβάνοντας τον λογάριθμο των ενεργειακών τιμών της υποζώνης, προκύπτουν 121 τιμές υπο-ζώνης λογαριθμικής ενέργειας κάθε 12,4 ms.

Στο επόμενο βήμα εφαρμόζονται φίλτρα. Σε κάθε καρέ εφαρμόζονται $N = 64$ φίλτρα και παράγονται N χαρακτηριστικά. Τα φασματοχρονικά χαρακτηριστικά είναι συνδιασμός των τιμών CQT log-ενέργειας του τρέχοντος και γειτονικών πλαισίων.

Το τρίτο βήμα είναι να προσδιοριστεί αν κάθε φασματο-χρονικό χαρακτηριστικό αυξάνεται ή μειώνεται με το χρόνο, με υπολογισμό των χαρακτηριστικών στον διαχωρισμό των T δευτερολέπτων και οριοθέτηση των χαρακτηριστικών στο 0 για να

γίνει επιβεβαίωση ότι τα ψηφία είναι ισορροπημένα. Χρησιμοποιούμε τιμή $T = .992\ sec$. Στο τέλος έχουμε N δυαδικές τιμές ανά καρέ.

Το τέταρτο βήμα είναι να "πακεταριστούν" οι $N = 64$ τιμές ανά πλαίσιο σε έναν ακέραιο $64 - bit$. Αυτό επιτρέπει την αποθήκευση hashprints και τον υπολογισμό των αποστάσεων Hamming με χρήση logical operators bit.



Σχήμα 2 Δομικό διάγραμμα υπολογισμού hashprint.

Γνώση του φίλτρου: Η επιλογή των φίλτρων γίνεται για να μεγιστοποιηθεί η διακύμανση των φασματοχρονικών χαρακτηριστικών που προκύπτουν, ενώ βεβαιώνουν πως δεν συνδέονται.

Θεωρείται ενα διάνυσμα $\in R^{121w}$ που περιέχει τις τιμές log-ενέργειας CQT, όπου w ο αριθμός πλαισίων. Μια δέσμη αυτών των διανυσμάτων μπορεί να θεωρηθεί σε μία μήτρα $A \in R^{M*121w}$, που M ο αριθμός των καρέ. $S \in R^{121w*121w}$είναι η μήτρα συνδιακύμανσης του A και $x_i \in R^{121w}$ οι συντελεστές του i φίλτρου. Για $i = 1, \ldots, N$, :

maximize $\boldsymbol{x_i^T S x_i}$

subject to $\|\boldsymbol{x_i}\|_2^2 = 1$

$\boldsymbol{x_i^T x_j} = 0$ $\boldsymbol{, j} =$ 1 , $\ldots, \boldsymbol{i} - 1.$

γ) Αναζήτηση αλγόριθμου

Στη διάρκεια της διαδικτυακής αναζήτησης, η ακολουθία ερωτημάτων hashprint συγκρίνεται με τη βάση δεδομένων για να κάνει αντιστοίχιση.

Σταυρωτή συσχέτιση: Σε κάθε ακολουθία hashprint, προσδιορίζεται η μετατόπιση μεγιστοποίησης ρυθμού συμφωνίας bit μεταξύ της ακολουθίας ερωτήματος και του αντίστοιχου τμήματος ακολουθίας αναφοράς.

Υποδειγματοληψία: Για επιτάχυνση γίνεται υποδειγματοληψία στις ακολουθίες ερωτήματος και αναφοράς από έναν παράγοντα B.

Αποκατάσταση: Η αποκατάσταση γίνεται με χρήση των ακολουθιών hashprint χωρίς υποδειγματοληψία στις κορυφαίες L ακολουθίες.

4.5.2 Εκτίμηση

α) Δεδομένα

Η εκτίμηση γίνεται στο δείκτη αναφοράς Gracenote. Τα στοιχεία προέρχονται από 10 καλλιτέχνες διαφορετικών ειδών και φαίνονται σον πίνακα 1.

Artist Name	ID	Genre	Songs	Dur (hrs)
Big K.R.I.T.	Big	hip hop	71	4.2
Chromeo	Chr	electro-funk	44	3.0
Death Cab for Cutie	Dea	indie rock	87	6.0
Foo Fighters	Foo	hard rock	86	6.1
Kanye West	Kan	hip hop	92	6.6
Maroon 5	Mar	pop rock	66	4.0
One Direction	One	pop boy band	60	3.4
Taylor Swift	Tay	country, pop	71	4.9
T.I.	TI	hip hop	154	10.8
Tom Petty	Tom	rock, blues rock	193	12.1

TABLE I

Πίνακας 1 Σφαιρική εικόνα της φιοριτούρας των δεδομένων ταυτοποίησης του τραγουδιού “live”. Η βάση δεδομένων περιέχει ολόκληρα κομμάτια από τα “studio album” των καλλιτεχνών. Τα ερωτήματα αποτελούνται από 1000 6 δευτερολέπτων ηχογραφήσεων κινητών τηλεφώνων από “live” εκτελέσεις (1000 ερωτήματα ανά καλλιτέχνη).

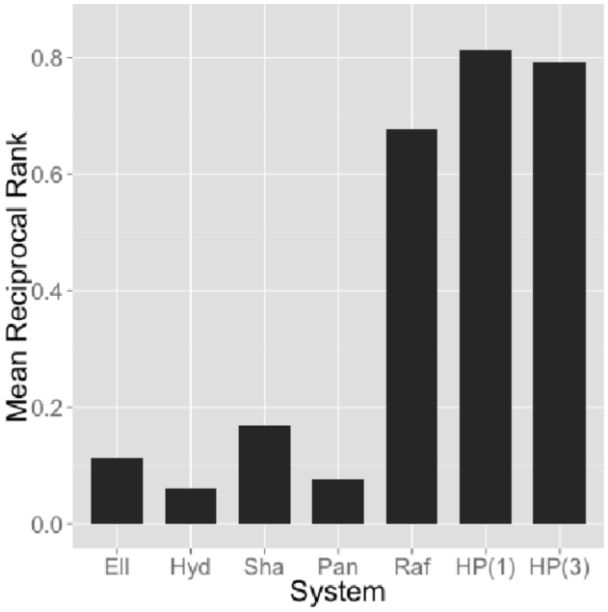
β) Μέτρηση εκτίμησης

Η μέτρηση που χρησιμοποιείται είναι η μέση αμοιβαία σειρά (MRR)[26] και δίνεται από τον τύπο:

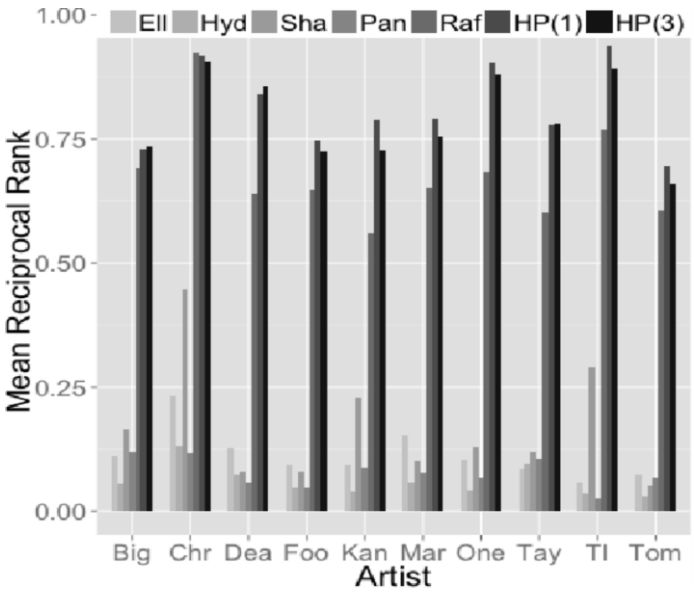
$$MRR = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{R_i}$$

N τα ερωτήματα αναφοράς και R_iη τάξη του σωστού στοιχείου για το i.

γ) Αποτελέσματα



Σχήμα 3 Απόδοση του προτεινόμενου συστήματος hashprint («HP») σε σύγκριση με άλλα πέντε βασικά συστήματα. Ο αριθμός της παρένθεσης δείχνει τον συντελεστή δειγματοληψίας της αναζήτησης αλληλοσυσχέτισης.



Σχήμα 4 Ανάλυση των αποτελεσμάτων από καλλιτέχνες. Τα πρώτα τρία γράμματα του ονόματος του καλλιτέχνη εμφανίζονται στο κάτω μέρος.

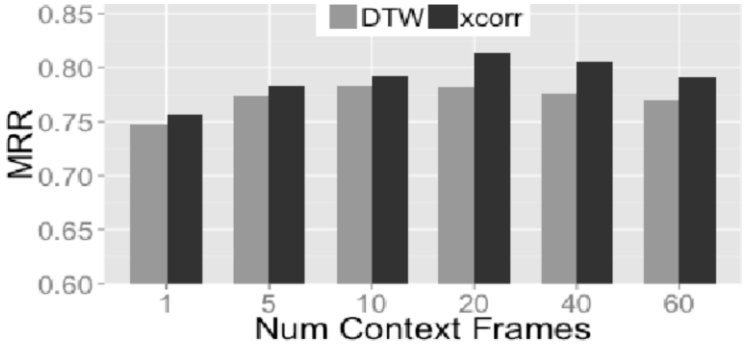
ID	Ref	System	Description
Ell	[12]	cover song	max cross-correlation, beat-level chroma
Hyd	[11]	cover song	system combination of [12], [57], and [58]
Sha	[33]	fingerprint	number of matching spectral peak pairs
Pan	[59]	fingerprint	number of matching spectral peak triples
Raf	[25]	live song	Hamming similarity, binarized CQT

Πίνακας 2 Περίληψη των 5 συστημάτων που χρησιμοποιούντα ως συγκρίσεις.

4.5.3 Επιδράσεις

α) Μη αντιστοιχία ρυθμού

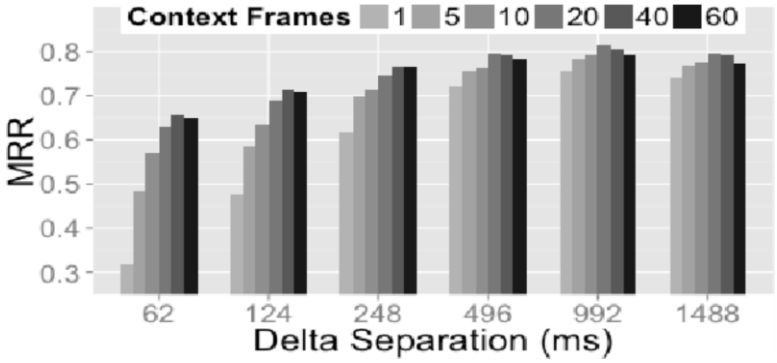
Σύγκριση αλγόριθμου αναζήτησης διασταυρώσεων με ακολουθία DTW, που είναι ένας τρόπος ευθυγράμμισης 2 ακολουθιών χαρακτηριστικών με τοπικές διαφορές στο ρυθμό και η ακολουθία DTW είναι παραλαγή που επιτρέπει σε μία ακολουθία να αρχίσει σε οποιαδήποτε μετατόπιση στην άλλη ακολουθία[27].



Σχήμα 5 Σύγκριση των προσεγγίσεων DTW και συσχέτισης. Ο οριζόντιος άξονας αναφέρεται στον αριθμό πλαισίων που καλύπτονται από κάθε hashprint.

β) Πλαίσιο και Δέλτα

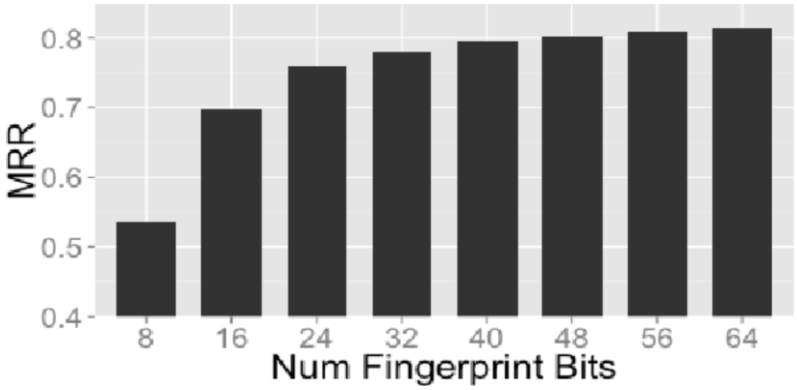
Γίνεται εκτ'ελεση πειραμάτων σε ένα εύρος τιμών w και διαχωρισμού δέλτα T, με τη χρήση αντιστοιχιών σταυρωτής συσχέτισης με βαθμό υποδειγματοληψίας 1.



Σχήμα 6 Επίδραση του μήκους παραθύρου περιβάλλοντος hashprint και τιμής διαχωρισμού δέλτα στην απόδοση του συστήματος.

γ) Αριθμός Bits

Γίνεται χρήση αντιστοίχισης σταυρωτής συσχέτισης με συντελεστή δειγματοληψίας 1.



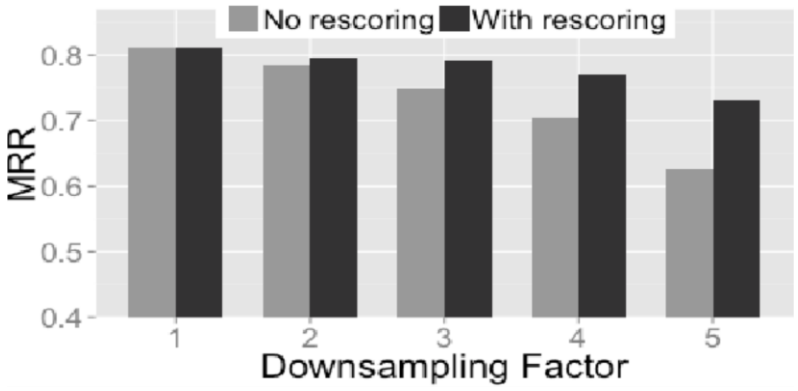
Σχήμα 7 Επίδραση του αριθμού των bits hashprint στην απόδοση του συστήματος.

δ) Γνώση

Επαναλαμβάνεται το πείραμα 'HP' (Σχήμα 3) με χρήση φίλτρων με τυχαίους συντελεστές. Η χρήση τυχαίων συντελεστών αντιστοιχεί σε προσέγγιση ευαίσθητης περιοχής (LSH).

ε) Υποδειγματοληψία και αποκατάσταση

Εξετάζεται η ακρίβεια και η αποτελεσματικότητα του συστήματος.



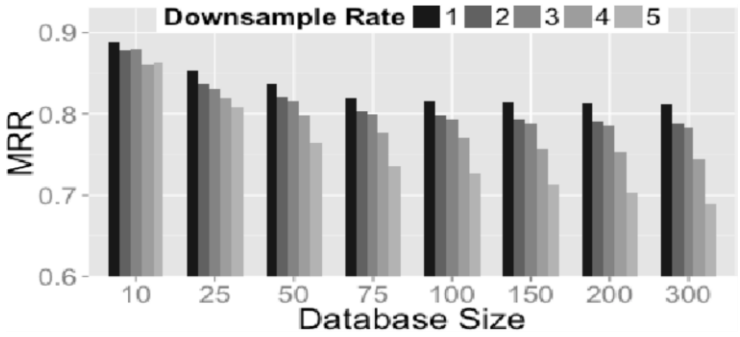
Σχήμα 8 Επίδραση της υποδειγματοληψίας και αποκατάστασης στην απόδοση του συστήματος.

Downsample	Rescoring	CQT	Search	Total
1	no	.51	2.87	3.43
2	no	.50	.68	1.23
3	no	.49	.31	.85
4	no	.52	.18	.74
5	no	.49	.12	.66
1	yes	.53	2.90	3.48
2	yes	.50	.72	1.26
3	yes	.51	.35	.90
4	yes	.50	.21	.76
5	yes	.49	.15	.69

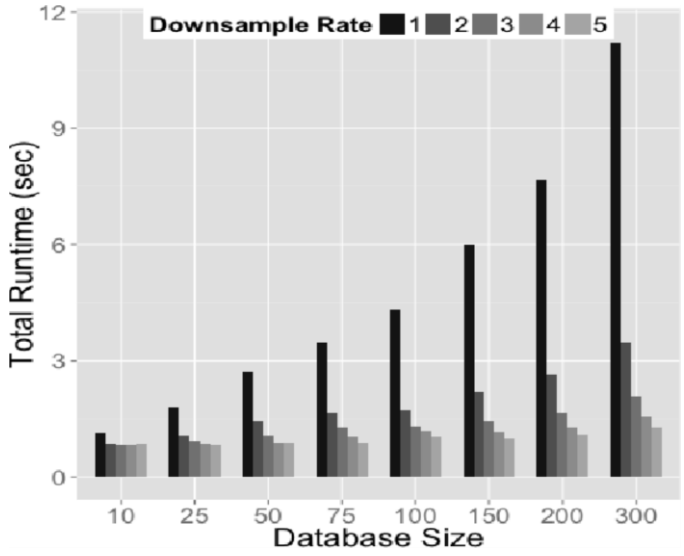
Πίνακας 3 Επίδραση της υποδειγματοληψίας και αποκατάστασης στο μέσο χρόνο επεξεργασίας ανά ερώτημα. Οι στήλες 3 και 4 δείχνουν το μέσο χρόνο σε δευτερόλεπτα που απαιτείται για τον υπολογισμό CQT και πραγματοποίηση της αναζήτησης αντίστοιχα, Η στήλη 5 δείχνει το μέσο συνολικό χρόνο επεξεργασίας ανά ερώτημα. Τα πειράματα εκτελέθηκαν σε ένα μονό πυρήνα 2.2 GHz Intel Xeon prossecor.

ζ) Μέγεθος βάσης δεδομένων

Γίνεται εκτέλεση ελεγχόμενων πειραμάτων στα οποία έχει διορθωθεί το μέγεθος της βάσης δεδομένων. Όταν αυτό είναι μικρότερο του πραγματικού, αφαιρούνται τραγούδια για να επιτευχθεί το επιθυμητό. Στην αντίθετη περίπτωση, γίνεται γέμισμα με τυχαία τραγούδια από τους άλλους καλλιτέχνες.



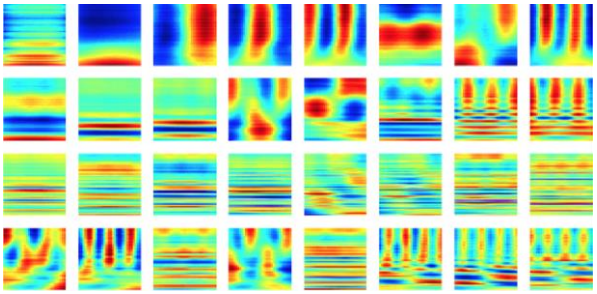
Σχήμα 9 Επίδραση του μεγέθους της βάσης δεδομένων στην απόδοση του συστήματος.



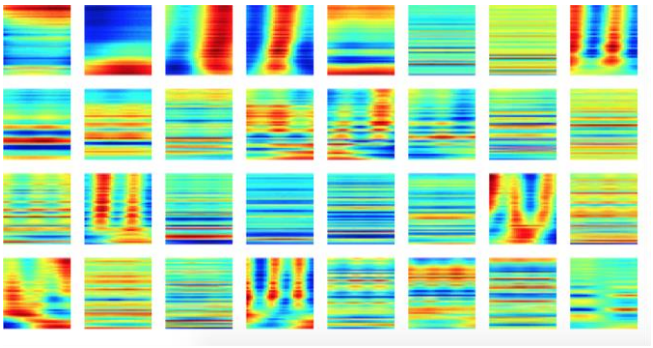
Σχήμα 10 Επίδραση του μεγέθους της βάσης δεδομένων στο μέσο χρόνο επεξεργασίας ανά ερώτημα.

η) Φίλτρα

Στα σχήματα φαίνεται το είδος των πληροφοριών που συλλαμβάνουν τα hashprints.



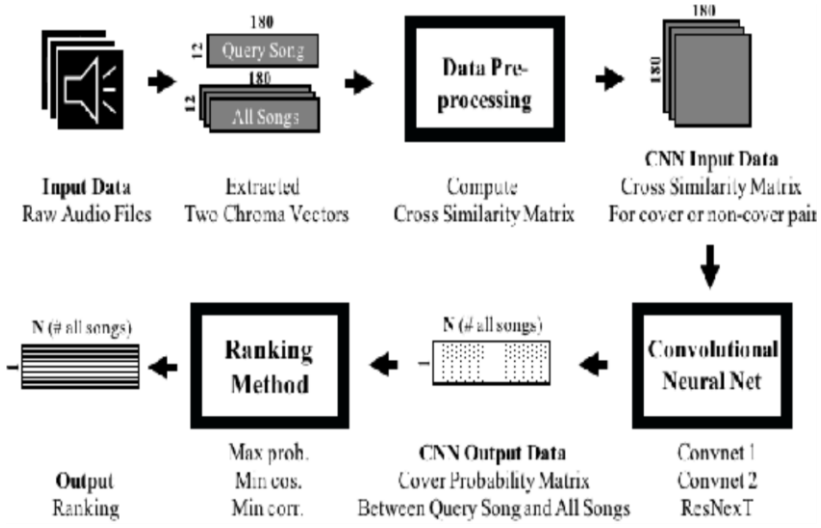
Σχήμα 11 Τα 32 καλύτερα γνωστά φίλτρα για το Big KRIT. Είναι διατεταγμένα πρώτα από αριστερά προς τα δεξιά και στη συνέχεια από πάνω προς τα κάτω. Κάθε φίλτρο εκτείνεται σε 0,372 δευτερόλεπτα και καλύπτει μια περιοχή συχνοτήτων από C3 έως C8.



Σχήμα 12 Τα 32 καλύτερα γνωστά φίλτρα για τον Taylor Swift.

4.6 Αναγνώριση τραγουδιών “cover” με χρήση μήτρας σταυρωτής ομοιότητας ανά τραγούδι με συμβατικό νευρωνικό δίκτυο

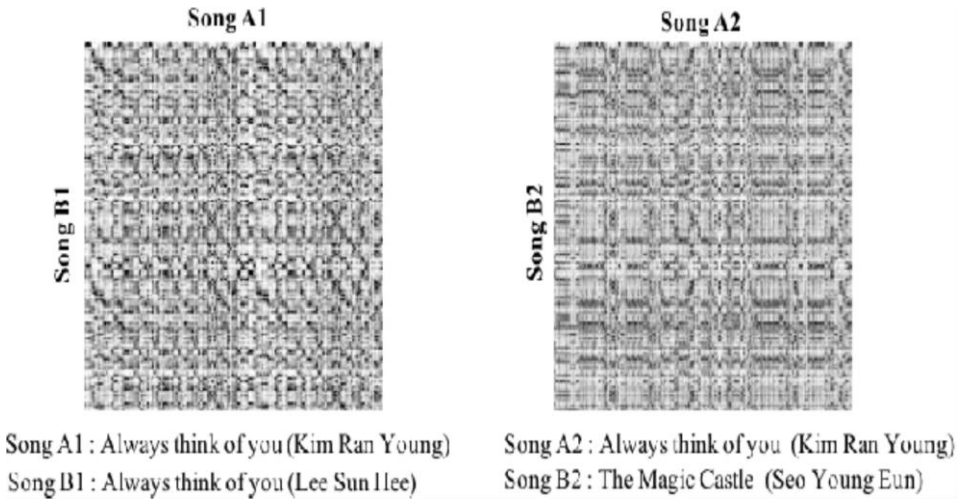
Σε αυτήν την εργασία γίνεται μέτρηση ομοιότητας με μήτρα σταυρωτής ομοιότητας σε 2 διανύσματα χρώματος. Μετά με σχεδίαση CNN[1],[2], υπολογίζονται οι τιμές των πιθανοτήτων, με τις οποίες γίνεται αναπαράσταση των διανυσμάτων όπως και η απόσταση και η συσχέτιση μεταξύ τους[3].



Σχήμα 1 Επισκόπηση συστήματος: Από τον ήχο κάθε τραγουδιού εξάγεται ένας διάνυσμα χρώματος. Χρησιμοποιώντας τη μήτρα σταυρωτής ομοιότητας από δύο χρώματα, το CNN εξάγει την πιθανότητα "cover". Η κατάταξη δίνεται με βάση τον πίνακα πιθανότητας.

4.6.1 Μήτρα σταυρωτής ομοιότητας

Κάνοντας σύγκριση ακολουθίας μελωδίας εντοπίζεται η σχέση 2 τραγουδιών. Με τη χρήση διανυσμάτων χρώματος 12-d που εξάγονται από τον ακατέργαστο ήχο και αναπαραστούν την ενέργεια για 12 ημιτόνια ανά μονάδα χρόνου, υπολογίζεται η μήτρα σταυρωτής ομοιότητας.



Σχήμα 2 Παράδειγμα μητρών σταυρωτής ομοιότητας που δημιουργούνται από ζεύγη "cover" (αριστερά) και μη "cover" (δεξιά): ένα διαγώνιο στοιχείο βρέθηκε στη μήτρα των "cover".

4.6.2 CNN

Το ConvNet-1 είναι τύπος CNN με 10 συνελικτικά στρώματα 0.58×10^6 παραμέτρων. Αρχικά υποδεικνύει την είσοδο με φίλτρο 5×5. Το Convnet-2 περιέχει 25.28×10^6 παράμετρούς και το φίλτρο είναι 3×3. Το ResNeXt[2] είναι ένα εκτεταμένο CNN.

Block #	Input layer	Block 1	Block 2	Block 3	Block 4	Block 5	Final layers	
Compo-nents	-	$\left\{\begin{array}{l} Conv\,(32 \times 5 \times 5), \text{ReLU} \\ Conv\,(32 \times 5 \times 5), \text{ReLU} \\ Maxpool\,(2 \times 2) \\ BN \end{array}\right\} \times 1$	$\left\{\begin{array}{l} Conv\,(32 \times 3 \times 3), \text{ReLU} \\ Conv\,(16 \times 3 \times 3), \text{ReLU} \\ Maxpool\,(2 \times 2) \\ BN \end{array}\right\} \times 4$				$DropOut_1(0.5)$ $FC(256)$, ReLU $DropOut_2(0.25)$	$FC(2)$ softmax
Output	(1, 180, 180)	(32,90,90)	(16,45,45)	(16,22,22)	(16,11,11)	(16,5,5)	(,,256)	(,,2)

Πίνακας 1 Αρχιτεκτονική ConvNet-1: Μέσα στις αγκύλες υπάρχουν συνελικτικά blocks και έξω ο αριθμός των στοιβαγμένων blocks. Το Conv υποδηλώνει ένα ίδιο στρώμα συνέλιξης με βήμα = 1, είναι (κανάλι × πλάτος × ύψος). Το Maxpool δηλώνει ένα στρώμα μέγιστης συγκέντρωσης με βήμα = 1 και οι εσωτερικές του παρενθέσεις είναι (μέγεθος που συγκεντρώθηκε). ReLU, BN, FC (ο αριθμός των βαρών), και DropOut (ποσοστό) υποδηλώνει την επαναλαμβανόμενη λειτουργία γραμμικής ενεργοποίησης μονάδας, την κανονικοποίηση της παρτίδας, το πλήρως συνδεδεμένο στρώμα και την τακτοποίηση λόγω αποχώρησης, αντίστοιχα.

Block #	Input layer	Block 1	Block 2	Block 3	Final layers	
Compo-nents	.	$\left\{ \begin{matrix} Conv\,(16\times3\times3),\mathrm{ReLU} \\ BN \end{matrix} \right\}\times2$ $Maxpool(2\times2)$	$\left\{ \begin{matrix} Conv\,(32\times3\times3),\mathrm{ReLU} \\ BN \end{matrix} \right\}\times3$ $Maxpool(2\times2)$	$\left\{ \begin{matrix} Conv\,(48\times3\times3),\mathrm{ReLU} \\ Conv\,(64\times3\times3),\mathrm{ReLU} \\ Conv\,(80\times3\times3),\mathrm{ReLU} \\ Conv\,(96\times3\times3),\mathrm{ReLU} \end{matrix} \right\}\times1$ $Maxpool(2\times2)$	$FC(1024),\mathrm{ReLU}$ $DropOut_3(0.5)$ $FC(200),\mathrm{ReLU}$ $DropOut_4(0.8)$	$FC(2)$ softmax
Output	(1,180,180)	(16,176,176)	(32,41,41)	(96,16,16)	(,,200)	(,,2)

Πίνακας 2 Αρχιτεκτονική ConvNet-2: Μέσα στις αγκύλες υπάρχουν συνελικτικά blocks και έξω ο αριθμός των στοιβαγμένων blocks. Το Conv υποδηλώνει ένα έγκυρο στρώμα συνέλιξης με βήμα = 1, και επαναχρησιμοποιούμε τις σημειώσεις του Πίνακα 1.

4.6.3 Μέθοδος κατάταξης

Με την υπόθεση N τραγουδιών, προκύπτουν $N \times N$ ζεύγη από μήτρες ως είσοδο στα CNN. Η έξοδος μετά παράγει μήτρα πιθανότητας ομοιότητας $P \in R^{N*N}$ όπου $P_{i,j}$ με $i,j \in \{1,2,\dots,N\}$ είναι δείκτης τραγουδιού. Με P υπολογίζεται :

$$R_i^{MaxProb} = sort_{des}\left(P_{i,j}\ για\ όλα\ τα\ j\right)$$

$R_i^{MaxProb} \in R^{1*N}$ είναι η κατάταξη της πιθανότητας "cover" του τραγουδιού i.

Παρουσιάζονται και άλλες μέθοδοι με διανύσματα του P. Πρώτα παρουσιάζεται το i τραγούδι με $P_{i,j}$ για όλα τα j και μετά υπολογίζεται η κατάταξη βασισμένη στην απόσταση των διανυσμάτων αναπαράστασης. Το διάνυσμα αυτό για το i $P_{i,}: \in R^{1*N}$ αποτελείται από τις τιμές πιθανότητας "cover" $P_{i,j}$ για όλα τα j και έχει τη μορφή $\left[P_{i,1}, P_{i,2}, \dots, P_{i,N}\right]$. Καθορίζεται R_i^{MinCos} :

$$R_i^{MinCos} = sort_{asc}\left(dist_{cos}\left(P_{i,:}, P_{j,:}\right)για\ όλα\ τα\ j\right)$$

$dist_{cos}$ επιστρέφει την απόσταση συνημιτόνου[4] μεταξύ 2 διανυσμάτων και $sort_{asc}$ επιστρέφει τον δείκτη των στοιχείων ταξινομημένα σε αύξουσα σειρά. Μπορούμε να αντικαταστήσουμε τη λειτουργία $dist$:

$$R_i^{MinCorr} = sort_{asc}\left(dist_{corr}\left(P_{i,:}, P_{j,:}\right)για\ όλα\ τα\ j\right)$$

$dist_{corr}$ η συσχέτιση 2 διανυσμάτων[4].

4.6.4 Εκτίμηση

Στον πίνακα 3 το σύνολο δεδομένων εξάσκησης αποτελείται από 1.175 τραγούδια και το σύνολο δεδομένων δοκιμής από 1.000 που συλλέχθηκαν από τους Heo et al[5]. Τα σύνολα διαχωρίζονται.

Dataset	# cover	# non-cover
Train 2 K	2,113	2,113
Train 30 K	2,113	30,000
Train 100 K	2,113	100,000
Validation	322	322
Test	3,300	496,200

Πίνακας 3 Πληροφορίες συνόλου δεδομένων

Χρησιμοποιούνται 3 μετρήσεις:

- MNIT10: ο μέσος αριθμός των αληθινών "cover" των 10 ανώτατων κομματιών που θεωρούνται "cover" για κάθε τραγούδι.
- MAP: η μέση ακρίβεια.
- MR1: η μέση κατάταξη 1.

Στο πείραμα αυξάνεται το μέγεθος των μη "cover" τραγουδιών για την κατανόηση των επιπτώσεων του CNN στην αναγνώριση. Εξασκήθηκαν 9 CNN μοντέλα: Χρησιμοποιήθηκαν ο Adam optimizer[6] με τη λειτουργία απώλειας σταυρωτής εντροπίας[7] και η ακρίβεια επικύρωσης κάθε μοντέλου κυμαίνεται από 0,83-0,88. Εφαρμόστηκαν 3 μέθοδοι κατάταξης στην έξοδο των CNN. Τέλος, αξιολογήθηκαν τα αποτελέσματα 2 βάσεων και 27 προτεινόμενων αλγορίθμων, όπως παρουσιάζονται στον Πίνακα 4. Το καλύτερο αποτέλεσμα μας έφθασε 12,8% μεγαλύτερο σημείο MNIT10 και 12% μεγαλύτερο MAP.

Model	Train set	ranking method	# correct answer	MNIT10	MAP	MR1
DTW+ML	-	MinEuclid	2406	7.29	0.75	26.55
SimPLe+ML	-		2602	7.88	0.81	15.05
ConvNet-1	2 K	MaxProb	2273	6.89	0.72	2.62
		MinCos	1657	5.02	0.52	16.5
		MinCorr	2090	6.33	0.67	8.67
	30 K	MaxProb	2655	8.05	0.83	2.50
		MinCos	3007	9.11	**0.93**	7.59
		MinCorr	3022	**9.16**	**0.93**	4.80
	100 K	MaxProb	2521	7.64	0.78	4.06
		MinCos	2888	8.75	0.89	8.32
		MinCorr	2911	8.82	0.90	10.7
ConvNet-2	2 K	MaxProb	1899	5.75	0.57	3.46
		MinCos	1621	4.91	0.52	10.5
		MinCorr	1852	5.61	0.60	8.01
	30 K	MaxProb	2647	7.99	0.82	2.92
		MinCos	2913	8.83	0.90	7.86
		MinCorr	2941	8.91	0.91	5.93
	100 K	MaxProb	2649	8.03	0.82	3.03
		MinCos	3008	9.12	0.92	10.3
		MinCorr	**3023**	**9.16**	**0.93**	7.01
ResNeXt	2 K	MaxProb	2525	7.65	0.77	2.90
		MinCos	1561	4.73	0.50	20.9
		MinCorr	1970	5.97	0.63	10.2
	30 K	MaxProb	2607	7.90	0.81	3.03
		MinCos	2955	8.95	0.91	3.26
		MinCorr	2954	8.95	0.91	2.87
	100 K	MaxProb	2705	8.20	0.84	**1.96**
		MinCos	3013	9.13	**0.93**	5.41
		MinCorr	3016	9.14	**0.93**	4.84

Πίνακας 4 Απόδοση αναγνώρισης τραγουδιού "cover" για τους αλγόριθμους βάσης (DTW ML, SimPLe ML) και προτεινόμενων αλγορίθμων. Οι DTW + ML και SimPLe + ML είναι αλγόριθμοι που εφαρμόζουν μέτρηση μάθησης στις τιμές εξόδου DTW και SimPLe, αντίστοιχα[5].

5 ΣΥΜΠΕΡΑΣΜΑΤΑ

Βάση των παραπάνω μεθόδων και των πειραμάτων που έχουν εκτελεστεί προκύπτουν κάποια συμπεράσματα και διαπιστώσεις που οδηγούν σε ιδέες για εξέλιξη των ζητημάτων σε μελλοντικές εργασίες.

5.1 Συμπεράσματα για την εξαγωγή μουσικού αποτυπώματος για αναγνώριση τραγουδιών "cover" κλασικής μουσικής

Η πρώτη εργασία αφορούσε μια προσέγγιση μουσικού αποτυπώματος για αναγνώριση τραγουδιών κλασικής μουσικής. Το αποτύπωμα αυτό είναι αποτελεσματικό γιατί μπορεί να εφαρμοστεί σε φορητές συσκευές και προσωπικούς υπολογιστές με την ανάλωση πολύ μικρής υπολογιστικής δύναμης και μνήμης, μειώνοντας σημαντικά το χρόνο αναζήτησης και αυξάνοντας την ακρίβεια.

Προκύπτει από τα αποτελέσματα πως η δομή της αρμονίας μιας ατομικής παρατήρησης περιλαμβάνει περισσότερη πληροφορία από τη συνολική διανομή παρατηρήσεων. Επιπλέον, φαίνεται πως η χρονική δυναμική πληροφορία είναι πολύ σημαντικη στην αναγνώριση “cover”. Στο μέλλον θα μελετηθεί η σημασία αυτών σε άλλες σχετικές εφαρμογές, όπως αυτοματοποιημένος συνθέτης, είδος και ταξινόμηση διάθεσης.

5.2 Συμπεράσματα για τα διανύσματα χαρακτηριστικών δυναμικού χρώματος με εφαρμογές στην αναγνώριση "cover"

Το προτεινόμενο διάνυσμα χαρακτηριστικού χρώματος μετά από πειράματα, υποδεικνύεται πως μπορεί πολύ αποτελεσματικά να υιοθετηθεί ως συμπληρωματικό χαρακτηριστικό. Σε μελλοντική εργασία θα ήταν χρήσιμο να μελετηθούν οι επιδράσεις των μουσικών χαρακτηριστικών που είναι ενσωματωμένα στη μουσική πληροφορία, ώστε να να μπορέσουν με το βέλτιστο τρόπο να συγχωνευθούν αυτά τα χαρακτηριστικά βάση ενός αλγόριθμου. Επίσης, θα πρέπει να μελετηθούν εφαρμογές ανάκτησης μουσικής πληροφορίας, όπως συνθέτης αναγνώρισης και κατάτμηση μουσικής με το προτεινόμενα διανύσματα χαρακτηριστικού δέλτα χρώματος.

5.3 Συμπεράσματα για την αναγνώριση τραγουδιών "cover" μεγάλης κλίμακας με χρήση κατακερματισμένων σημείων χρώματος

Σε αυτήν την εργασία χορηγήθηκε για πρώτη φορά στη βάση δεδομένων SHSD ένα σημείο αναφοράς. Σε σύγκριση με πειράματα μικρότερων συλλογών, τα αποτελέσματα ήταν πιο αρνητικά.

Όσον αφορά την αναγνώριση “cover”, αυτή η βάση δεδομένων παρέχει τα δεδομένα που οδηγούν στη γνώση μιας μέτρησης ομοιότητας. Για άλλα ζητήματα ανάκτησης μουσικής πληροφορίας, πολλά από αυτά μπορούν να να γίνουν στο MSD, θα ήταν χρήσιμο να βρεθούν καλύτεροι αλγόριθμοι για σύσταση και ταξινόμηση.

5.4 Συμπεράσματα για την αναγνώριση "cover" με άμεση εξαγωγή χαρακτηριστικών χρώματος από αρχεία AAC

Στην εργασία αυτή αναπτύχθηκε μια γρήγορη μέθοδος εξαγωγής χαρακτηριστικών. Σχεδιάστηκαν οι φασματικοί συντελεστές σε χαρακτηριαστικά χρώματος 12-bin χωρίς να αποκωδικοποιούνται πλήρως. Η κατάτμηση των χαρακτηριστικών χρησιμοποιήθηκε ώστε να μειωθεί ο χρόνος της διάστασης με σκοπό να επιταχυνθεί η διαδικασία αντιστοίχησης. Το σύστημα κατάφερε να παρέχει μια ακρίβεια 62% και ο χρόνος αντιστίχοισης μειώθηκε κατά 35%, έτσι αποτελεί λύση για βάσεις δεδομένων μεγάλης κλίμακας.

5.5 Συμπεράσματα για τη "live" αναγνώριση τραγουδιών γνωστών καλλιτεχνών με χρήση ηχητικών "hashprints"

Δύο συστατικά απαρτίζουν το σύστημα που προτάθηκε σε αυτήν την εργασία : 1) μια δυαδική αναπαράσταση ήχου (“hashprints”) , 2) ένας αλγόριθμος αντιστοίχισης βασισμένος στην αλληλοσυσχέτιση που μπορεί να συντονιστεί για την επίτευξη της επιθυμητής καθυστέρησης του χρόνου εκτέλεσης. Το σύστημα αυτό ταυτόχρονα βελτιώνει τη μέση αμοιβαία κατάταξη από το .68 στο .79 και μειώνει την καθυστέρηση χρόνου εκτέλεσης από τα 10 στα 0.9 δευτερόλεπτα. Παρατηρείται πως και η αναπαράσταση του ήχου και ο αλγόριθμος αντιστοίχησης είναι συγκεκριμένα για τον καλλιτέχνη, προσαρμοσμένα στα ειδικά χαρακτηρικά της μουσικής του και στο μέγεθος της βάσης δεδομένων.

Το σύστημα αυτό υποθέτει πως είναι διαθέσιμη μια βάση δεδομένων με πληροφορίες της εκδήλωσης του κονσέρτου, καθώς και ότι το ερώτημα είναι μια ζωντανή παρουσίαση τραγουδιού πάνω σε ένα άλμουμ ηχογραφημένο σε στούντιο. Στο μέλλον θα πρέπει να μπορούν να διευθετηθούν οι περιπτώσεις που παραβιάζουν τις υποθέσεις αυτές , για παράδειγμα αν το ερώτημα δεν ανήκει στη βάση δεδομένων. Ωστόσο, θα πρέπει να επεκταθεί η βάση ώστε να περιέχει ποικιλία καλλιτεχνών, στυλ και ειδών. Με αυτό θα μπορεί να μελετηθεί η σχέση της ειδικότητας της γνώσης και της επίδοσης όταν τα φίλτρα μαθαίνονται ανά καλλιτέχνη, είδος και σε μία βάση δεδομένων με ποικιλία.

5.6 Συμπεράσματα για την αναγνώριση τραγουδιών “cover” με χρήση μήτρας σταυρωτής ομοιότητας ανά τραγούδι με συμβατικό νευρωνικό δίκτυο

Στην προσέγγιση αυτής της εργασίας παρατηρήθηκε πως η σχέση τραγουδιού “cover” μπορεί να παρουσιαστεί σαν ένα διαγώνιο συστατικό σε μήτρα σταυρωτής ομοιότητας. Με βάση αυτό έγινε διεύθυνση των CNN, ώστε να καταταχθούν σε “cover” και “non-cover” με χρήση της μήτρας αυτής. Στη συνέχεια προτάθηκε νέο διάνυσμα αναπαράστασης για τη μείωση της απόστασης μεταξύ των κομματιών με τη χρήση πιθανότητας εξόδου των CNN. Η απόδοση των CNN που προέκυψε από τα πειράματα με τη μέθοδο απευθείας κατάταξης στην πιθανότητα εξόδου ήταν καλύτερη από άλλες. Μετά την εφαρμογή της μεθόδου που προτάθηκε επιτεύχθηκε βελτίωση σημείου της MAP 12% από την πρώτη. Επίσης έγινε ανάλυση της επίδρασης των αυξημένων παραδειγμάτων τραγουδιών “non-cover” και η επίδραση της χρήσης αναπαράστασης διανύσματος για τη μέθοδο κατάταξης.

Τα αποτελέσματα φαίνονται ελπιδοφόρα όμως δεν παρέχουν ένα ολόκληρο πλαίσιο για την εύρεση “cover” σε μεγάλη κλίμακα. Στο μέλλον θα γίνει ανάπτυξη αυτού του θέματος.

ΑΝΑΦΟΡΕΣ

2-3

[1] D. P. W. Ellis, B. Whitman, A. Berenzweig, and S. Lawrence. The quest for ground truth in musical artist similarity. Int. Symp. on Music Information Retrieval (ISMIR), pages 518–529, October 2002.

[2] W. J. Dowling and J. L. Harwood. Music cognition. Academic Press, 1985.

[3] B. W. White. Recognition of distorted melodies. American Journal of Psychology, 73:100– 107, 1960.

[4] E. G. Schellenberg, P. Iverson, and M. C. McKinnon. Name that tune: identifying familiar recordings from brief excerpts. Psychonomic Bulletin and Review, 6(4):641–646, 1999.

[5] R. B. Dannenberg, W. P. Birmingham, B. Pardo, N. Hu, C. Meek, and G. Tzanetakis. A comparative evaluation of search techniques for query-by-humming using the musart testbed. Journal of the American Society for Information Science and Technology, 58(5):687–701, 2007.

[6] M. Casey, R. C. 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, April 2008.

[7] J. S. Downie. The music information retrieval evaluation exchange (2005–2007): a window in to music information retrieval research. AcousticalScienceandTechnology,29(4):247–255, 2008.

[8] N. Scaringella, G. Zoia, and D. Mlynek. Automatic genre classification of music content: a survey. IEEE Signal Processing Magazine, 23(2):133–141, 2006.

[9] P. Cano, E. Batlle, T. Kalker, and J. Haitsma. A review of audio fingerprinting. Journal of VLSI Signal Processing, 41:271–284, 2005.

[10] M. Casey, C. Rhodes, and M. Slaney. Analysis of minimum distances in high-dimensional musical spaces. IEEE Trans. on Audio, Speech, and Language Processing, 16(5):1015–1028, July 2008.

[11] R.Miotto and N.Orio. A music identification system based on chroma indexing and statistical modeling. Int. Symp. on Music Information Retrieval (ISMIR), pages 301–306, September 2008.

[12] M. Riley, E. Heinen, and J. Ghosh. A text retrieval approach to content-based audio retrieval. Int. Symp. on Music Information Retrieval (ISMIR), pages 295–300, September 2008.

[13] E.Unal and E.Chew. Statistical modeling and retrieval of polyphonic music. IEEE Workshop on Multimedia Signal Processing (MMSP), pages 405–409, 2007.

[14] C. Larkin, editor. The Encyclopedia of Popular Music. 3rd edition, November 1998.

[15] E. Selfridge-Field. Conceptual and representational issues in melodic comparison. MIT Press, Cambridge, USA, 1998.

[16] M. Marolt. A mid-level melody-based representation for calculating audio similarity. Int. Symp. on Music Information Retrieval (ISMIR), pages 280–285, October 2006.

[17] M. Marolt. A mid-level representation for melody-based retrieval in audio collections. IEEE Trans. on Multimedia, 10(8):1617–1625, December 2008.

[18] C. Sailer and K. Dressler. Finding cover songs by melodic similarity. MIREX extended abstract, September 2006.

[19] W. H. Tsai, H. M. Yu, and H. M. Wang. A query-by-example technique for retrieving cover versions of popular songs with similar melodies. Int. Symp. on Music Information Retrieval (ISMIR), pages 183–190, 2005.

[20] W. H. Tsai, H. M. Yu, and H. M. Wang. Using the similarity of main melodies to identify cover versions of popular songs for music document retrieval. Journal of Information Science and Engineering, 24(6):1669–1687, November 2008.

[21] G. E. Poliner, D. P. W. Ellis, A. Ehmann, E. G´omez, S. Streich, and B. S. Ong. Melody transcription from music audio: approaches and evaluation. IEEE Trans. on Audio, Speech, and Language Processing, 15:1247–1256, 2007.

[22] T. Fujishima. Realtime chord recognition of musical sound: a system using common lisp music. Int. Computer Music Conference (ICMC), pages 464–467, 1999.

[23] E. G´omez. Tonal description of music audio signals. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2006. Available online: http://mtg.upf.edu/node/472.

[24] H.Purwins. Proles of pitch classes. Circularity of relative pitch and key: experiments, models, computational music analysis, and perspectives. PhD thesis, Berlin University of Technology, Germany, 2005.

[25] G.Tzanetakis. Pitch histograms in audio and symbolic music information retrieval. Int. Symp. on Music Information Retrieval (ISMIR), pages 31–38, 2002.

[26] A. Egorov and G. Linetsky. Cover song identification with if-f0 pitch class profiles. MIREX extended abstract, September 2008.

[27] D.P.W. Ellis and C.Cotton. The 2007 labrosa cover song detection system. MIREX extended abstract, September 2007.

[28] D. P. W. Ellis and G. E. Poliner. Identifying cover songs with chroma features and dynamic programming beat tracking. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 4:1429–1432, April 2007.

[29] E. G´omez and P. Herrera. The song remains the same: identifying versions of the same song using tonal descriptors. Int. Symp. on Music Information Retrieval (ISMIR), pages 180–185, October 2006.

[30] E. G´omez, B. S. Ong, and P. Herrera. Automatic tonal analysis from music summaries for version identification. Conv. of the Audio Engineering Society (AES), October 2006. CDROM, paper no. 6902.

[31] R. Xu and D. C. Wunsch. Clustering. IEEE Press, 2009.

[32] T. E. Ahonen and K. Lemstrom. Identifying cover songs using normalized compression distance. Int. Workshop on Machine Learning and Music (MML), July 2008.

[33] J. P. Bello. Audio-based cover song retrieval using approximate chord sequences: testing shifts, gaps, swaps and beats. Int. Symp. On Music Information Retrieval (ISMIR), pages239– 244, September 2007.

[34] ¨O. Izmirli. Tonal similarity from audio using a template based attractor model. Int. Symp. on Music Information Retrieval (ISMIR), pages 540–545, 2005.

[35] K. Lee. Identifying cover songs from audio using harmonic representation. MIREX extended abstract, September 2006.

[36] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. Proc. of the IEEE, 1989.

[37] W. J. Dowling. Scale and contour: two components of a theory of memory for melodies. Psychological Review, 85(4):341–354, 1978.

[38] J. Foote. Arthur: Retrieving orchestral music by long-term structure. Int. Symp. on Music Information Retrieval (ISMIR), October 2000.

[39] J. Serr`a, E. G´omez, and P. Herrera. Transposing chroma representations to a common key. IEEE CS Conference on The Use of Symbols to Represent Music and Multimedia Objects, pages 45–48, October 2008.

[40] J. Serr`a, E. G´omez, P. Herrera, and X. Serra. Chroma binary similarity and local alignment applied to cover song identification. IEEE Trans. On Audio, Speech, and Language Processing, 16(6):1138–1152, August 2008.

[41] J. H. Jensen, M. G. Christensen, and S. H. Jensen. A chroma-based tempo-in sensitive distance measure for cover song identification using the 2d autocorrelation. MIREX extended abstract, September 2008.

[42] A. V. Oppenheim, R. W. Schafer, and J. B. Buck. Discrete-Time Signal Processing. Prentice Hall, 2 edition, February 1999.

[43] D. Gusfield. Algorithms on strings, trees and sequences: computer sciences and computational biology. Cambridge University Press, 1997.

[44] L. R. Rabiner and B. H. Juang. Fundamentals of speech recognition. Prentice, 1993.

[45] D. Sankoff and J. Kruskal. Time warps, string edits, and macromolecules. Addison-Wesley, 1983.

[46] J. H. Jensen, M. G. Christensen, D. P. W. Ellis, and S. H. Jensen. A tempo-insensitive distance measure for cover song identification based on chroma features. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pages 2209–2212, April 2008.

[47] S. Kim and S. Narayanan. Dynamic chroma feature vectors with applications to cover song identification. IEEE Workshop on Multimedia Signal Processing (MMSP), pages 984–987, October 2008.

[48] S. Kim, E. Unal, and S. Narayanan. Fingerprint extraction for classical music cover song identification. IEEE Int. Conf. on Multimedia and Expo (ICME), pages 1261–1264, June 2008.

[49] F. Kurth and M. M¨uller. Efficient index-based audio matching. IEEE Trans. On Audio, Speech, and Language Processing, 16(2):382–395, 2008.

[50] R. Baeza–Yates and B. Ribeiro-Neto. Modern Information Retrieval. ACMPressBooks,1999.

[51] H. Nagano, K. Kashino, and H. Murase. Fast music retrieval using polyphonic binary feature vectors. IEEE Int. Conf. on Multimedia and Expo (ICME), 1:101–104, 2002.

4.1

[1] E.Unal, E. Chew, P. Georgiou, S. Narayanan, “Challenging Uncertainly in Query-by-Humming Systems: A Fingerprinting Approach,” Special Issue of the IEEE transaction on Audio, Speech and Language Processing on Music Information Retrieval (MIR), Vol.16, No.2, 2008.

[2] J. Haitsma, T. Kalker, “A highly robust audio fingerprinting system, ”International Symposium on Music Information Retrieval (ISMIR), 2002.

[3] M.I. Mandel, D.P.W. Ellis, “Song-level features and SVM for music classification,” International Symposium on Music Information Retrieval (ISMIR), 2006.

[4] K. Lee, “Identifying cover songs from audio using harmonic representation, ”International Symposium on Music Information Retrieval (ISMIR), 2006.

[5] E. Unal, S. Narayanan, “Statistical modeling and retrieval of polyphonic music,” International workshop on Multimedia Signal Processing (MMSP), 2007.

[6] D.P.W. Ellis, G.E. Poliner “Identifying ‘cover songs’ with chroma features and dynamic programming beat tracking,” Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2007.

[7] R.N. Shepard, “Circularity in judgments of relative pitch,” Journal of the Acoustic Society of America, Vol. 36, No. 12, 1964.

[8] http://timidity.sourceforge.net/

4.2

[1] D. Warren, S. Uppenkamp, R. D. Patterson, and T. D. Griffiths, “Separating pitch chroma and pitch height in the human brain,” Proc. Natl. Acad. Sci. U.S.A., Vol. 100, No. 17, 2003.

[2] S. Kim, E. Unal, and S. Narayanan, “Music fingerprint extraction for classical music cover song identification,” International Conference of Multimedia and Expo (ICME), in press, Jun. 2008.

[3] http://www.classicalarchives.com

[4] http://timidity.sourceforge.net/

4.3

[1] T. Bertin-Mahieux, D. Ellis, B. Whitman, and P. Lamere, "The million song dataset," in Proceedings of the 11th 1nternational Society for Music 1nformation Retrieval Conference (ISM1R 2011),2011.

[2] A. Wang, "An industrial strength audio search algorithm," in Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003), 2003.

[3] The Echo Nest Analyze, API, http://developer.echonest.com.

4.4

[1] International Organization for Standardization, "Information Technology - Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding (AAC)," ISO/IEC 138187, 1997.

[2] E. Ravelli, G. Richard, and L. Daudet, “Audio Signal Representations for Indexing in the Transform Domain,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 3, pp. 434-446, March. 2010.

[3] J. Serra, E. Gomez, P. Herrera, and X. Serra, “Chroma Binary Similarity and Local Alignment Applied to Cover Song Identification,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 6, pp. 1138-1151, Aug. 2008.

[4] T. H. Tsai and C. Liu, “A Configurable Common Filterbank Processor for Multi-Standard Audio Decoder,” IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. 90, no. 9, pp. 1913-1923, Sep. 2007.

4.5

[1] J. Haitsma and T. Kalker, “A highly robust audio fingerprinting system,” in Proc. Int. Soc. Music Inf. Retrieval, 2002, pp. 107–115.

[2] J.Haitsma, T.Kalker, and J.Oostveen,“Robust audio hashing for content identification,” in Proc. Int. Workshop Content-Based Multimedia Indexing, 2001, vol. 4, pp. 117–124.

[3] S. Baluja and M. Covell, “Audio fingerprinting: Combining computer vision & data stream processing, ”in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Apr. 2007, vol. 2, pp. 213–216.

[4] S. Baluja and M. Covell, “Waveprint: Efficient wavelet-based audio fingerprinting,” Pattern Recog., vol. 41, no. 11, pp. 3467–3480, May 2008.

[5] X. Anguera, A. Garzon, and T. Adamek, “MASK: Robust local features for audio fingerprinting,” in Proc. IEEE Int. Conf. Multimedia Expo, Jul. 2012, pp. 455–460.

[6] E. Younessian, X. Anguera, T. Adamek, N. Oliver, and D. Marimon, “Telefonica research at TRECVID 2010 content-based copy detection,” in Proc. TRECVID, 2010.

[7] B. Coover and J. Han, “A power mask based audio fingerprint,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., May 2014, pp. 1394– 1398.

[8] P. Over et al., “TRECVID 2011—An overview of the goals, tasks, data, evaluation mechanisms and metrics,” in Proc. TRECVID, 2011.

[9] J.Downie, M.Bay, A.Ehmann, and M.Jones, “Audio cover song identification: MIREX 2006-2007 results and analyses,” in Proc. Int. Soc. Music Inf. Retrieval, 2008, pp. 468–474.

[10] S.Ravuriand D.Ellis, “Cover song detection: classification,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Mar. 2010, pp. 65–68.

[11] D. Ellis and G. Poliner, “Identifying ‘cover songs’ with chroma features and dynamic programming beat tracking,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Apr. 2007, pp. 1429–1432.

[12] J. Serra, X. Serra, and R. Andrzejak, “Cross recurrence quantification for cover song identification,” New J. Phys., vol. 11, no. 9, 2009, Art. no. 093017.

[13] J. Serra, E. G´omez, and P. Herrera, “Audio cover song identification and similarity: Background, approaches, evaluation, and beyond,” in Proc. Adv. Music Inf. Retrieval Conf., 2010, pp. 307–332.

[14] T. Bertin-Mahieux, D. Ellis, B. Whitman, and P. Lamere, “The million song dataset,” in Proc. Int. Soc. Music Inf. Retrieval, 2011, pp. 591–596.

[15] T. Bertin-Mahieux and D. Ellis, “Large-scale cover song recognition using the 2D fourier transform magnitude,” in Proc. Int. Soc. Music Inf. Retrieval, 2012, pp. 241–246.

[16] E.J.Humphrey, O.Nieto, and J.P.Bello, “Data driven and discriminative projections for large-scale cover song identification,” in Proc. Int. Soc. Music Inf. Retrieval, 2013, pp. 149–154.

[17] M. Khadkevich and M. Omologo, “Large-scale cover song identification using chord profiles,” in Proc. Int. Soc. Music Inf. Retrieval, 2013, pp. 233–238.

[18] J.Osmalsky, J.-J.Embrechts, P.Foster, and S.Dixon, “Combining features for cover song identification,” in Proc. Int. Soc. Music Inf. Retrieval, 2015, pp. 462–468.

[19] T. Bertin-Mahieux and D. P. Ellis, “Large-scale cover song recognition using hashed chroma landmarks,” in Proc. IEEE Workshop Appl. Signal Process. Audio Acoust., Oct. 2011, pp. 117–120.

[20] M. Casey, C. Rhodes, and M. Slaney, “Analysis of minimum distances in high-dimensional musical spaces,” IEEE Trans. Audio, Speech, Lang. Process., vol. 16, no. 5, pp. 1015–1028, Jul. 2008.

[21] P. Grosche and M. M¨uller, “Toward characteristic audio shingles for efficient cross-version music retrieval,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Mar. 2012, pp. 473–476.

[22] F. Kurth and M. M¨uller, “Efficient index-based audio matching,” IEEE Trans. Audio, Speech, Lang. Process., vol. 16, no. 2, pp. 382–395, Feb. 2008.

[23] M. M¨uller, F. Kurth, and M. Clausen, “Audio matching via chroma based statistical features,” in Proc. Int. Soc. Music Inf. Retrieval, 2005, pp. 288–295.

[24] Z. Rafii, B. Coover, and J. Han, “An audio fingerprinting system for live version identification using image processing techniques,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., May 2014, pp. 644–648.

[25] C. Sch¨orkhuber and A. Klapuri, “Constant-q transform toolbox for music processing,” in Proc. Sound Music Comput. Conf., 2010, pp. 3–64.

[26] E. Voorhees, “The TREC-8 question answering track report,” in Proc. 8th Text Retrieval Conf., 1999, pp. 77–82.

[27] M. M¨uller, Fundamentals of Music Processing. Berlin, Germany: Springer, 2015.

4.6

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.

[2] Saining Xie, Ross Girshick, Piotr Doll´ar, Zhuowen Tu, and Kaiming He, “Aggregated residual transformations for deep neural networks,” arXiv preprint arXiv:1611.05431, 2016.

[3] Sungkyun Chang, Juheon Lee, Sang Keun Choe, and Kyogu Lee, “Audio cover song identification using convolutional neural network,” NIPS Machine Learning for Audio (ML4Audio) Workshop, 2017.

[4] Sung-Hyuk Cha, “Comprehensive survey on distance/similarity measures between probability density functions,” City, vol. 1, no. 2, pp. 1, 2007.

[5] Hoon Heo, Hyunwoo J Kim, Wan Soo Kim, and Kyogu Lee, “Cover song identification with metric learning using distance as a feature,” ISMIR, 2017.

[6] Diederik Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[7] Pieter-Tjerk De Boer, Dirk P Kroese, Shie Mannor, and Reuven Y Rubinstein, “A tutorial on the cross-entropy method,” Annals of operations research, vol.134, no.1, pp. 19–67, 2005.