ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΜΕΘΟΔΟΛΟΓΙΑΣ, ΙΣΤΟΡΙΑΣ ΚΑΙ ΘΕΩΡΙΑΣ
ΤΗΣ ΕΠΙΣΤΗΜΗΣ

΄Ενας Ταξινομητής ΄Ηχων

Σωτήρης Λαμπρινίδης

Πτυχιακή εργασία

Επιβλέπουσα:
Ελπίδα Τζαφέστα

14 Σεπτεμβρίου 2015

Περιεχόμενα

Περίληψη

Τα τελευταία χρόνια, η αύξηση της χρήσης του διαδικτύου και των ηλεκτρονικών υπολογιστών έχει οδηγήσει σε μία έκρηξη δεδομένων. Είτε πρόκειται για εμπορικές εφαρμογές είτε για έρευνα, το πεδίο της εξαγωγής πληροφοριών από ακουστικά σήματα παρουσιάζει μια αυξημένη κινητικότητα, προσπαθώντας να αποσπάσει και να χρησιμοποιήσει τις πληροφορίες που περιέχονται στα δεδομένα αυτά. Η παρούσα εργασία απασχολείται με το πρόβλημα της αυτόματης Ταξινόμησης ήχων, χωρίς να προϋποτίθεται κάτι για την φύση τους. Χρησιμοποιώντας καλώς τεκμηριωμένες μεθόδους πάνω στην εξαγωγή χαρακτηριστικών από ήχους και την Μηχανική Μάθηση, υλοποιήσαμε έναν Ταξινομητή ήχων. Η υλοποίησή μας βασίζεται στην σύνθεση πολλών χαρακτηριστικών και την εκπαίδευση ενός ταξινομητή Λογιστικής Παλινδρόμησης.

Τα αποτελέσματά μας συμβαδίζουν με την βιβλιογραφία και δείχνουν πως μία γενική προσέγγιση στην Ταξινόμηση Ήχων μπορεί να αποδώσει καρπούς και οφείλει να ερευνηθεί περαιτέρω.

1. ΕΙΣΑΓΩΓΗ

Ο ανθρωπος, μαζί με τις πιο πολλές μορφές έμβιας ύλης, έχει εξελιχθεί ώστε να μπορεί να αναγνωρίζει και να χειρίζεται πρότυπα προερχόμενα από κάθε αισθητήριό του. Ίσως φαίνεται αστεία η ερώτηση αν αυτό που ακούγεται είναι μουσική ή ομιλία, αν αυτό το κομμάτι που παίζει είναι ροκ ή κλασσική μουσική, αν ο ήχος που ακούγεται παράγεται από κοντραμπάσσο ή από βιολί. Όταν μάλιστα κάποιος περάσει αρκετό χρόνο με ένα ερέθισμα, μπορεί να εμβαθύνει ακόμα πιο πολύ: ο οδηγός γνωρίζει ότι το αυτοκίνητό του έχει κάποιο πρόβλημα από τον ήχο που κάνει η μηχανή, έναν ήχο σχεδόν κενό πληροφοριών για τον επιβάτη.

Υπάρχουν πράγματα που ο άνθρωπος φαίνεται να φέρνει εις πέρας με μεγάλη ευκολία, και που είναι πολύ δύσκολο να κάνει μία μηχανή. Αυτό το τεράστιο χάσμα πάντα έφερνε σε αμηχανία τους ερευνητές της Τεχνητής Νοημοσύνης. Η παρούσα μελέτη ασχολείται με την υλοποίηση ενός λογισμικού ταξινόμησης ηχητικών δειγμάτων.

1.1. Ο Στόχος της παρούσας εργασίας

Ας φέρουμε στο μυαλό μας μία μουσική βιβλιοθήκη που μπορεί να είναι η δική μας, ενός γνωστού μας ή ενός ραδιοφωνικού παραγωγού - το μόνο σίγουρο είναι πως τα αρχεία θα είναι με κάποιο τρόπο *ταξινομημένα*. Εμείς ενδέχεται να έχουμε χωρίσει τα κομμάτια σε φακέλους, δηλαδή σε *κατηγορίες*, ανάλογα με το μουσικό είδος, κάποιος άλλος ανάλογα με την διάθεση που του εμπνέουν, ενώ ο ραδιοφωνικός παραγωγός, που ας υποθέσουμε ότι εργάζεται σε δύο διαφορετικούς σταθμούς, ίσως τα έχει χωρισμένα με βάση τον σταθμό στον οποίο επιλέγει το κάθε κομμάτι.

Η δουλειά του παραγωγού, όμως, περιλαμβάνει και το να είναι ενημερωμένος πάνω στις τελευταίες κυκλοφορίες. Έτσι, ακούει καινούργια κομμάτια και τα κατατάσσει είτε σε κατάλληλα για κάποιον από τους δύο σταθμούς, είτε σε κατάλληλα για το καλάθι των αχρήστων. Σίγουρα μπορεί να επιχειρηματολογήσει για το ποιό κομμάτι ταιριάζει σε κάθε περίπτωση, είναι όμως εύλογη η υπόθεση πως με μία προσεκτική ακρόαση ο καθένας θα παρατηρήσει κάποια κοινά μεταξύ των κομματιών που ανήκουν σε μία κατηγορία και θα είναι σε θέση να αποφανθεί και αυτός για το πού θα ταξινομηθεί μία καινούργια κυκλοφορία.

Αυτή η διαδικασία που περιγράψαμε, η ταξινόμηση ενός νέου δείγματος σε μία από κάποιες ορισμένες κατηγορίες, είναι και ο στόχος του προγράμματος το οποίο θα υλοποιήσουμε.

1.2. Ήχος

Ένας ορισμός του ήχου θα μπορούσε να είναι η αίσθηση που προκαλείται λόγω της διέγερσης των αισθητηρίων οργάνων της ακοής από μεταβολές της πίεσης του αέρα. Φυσικά, ένας ήχος μπορεί να έχει πολλά χαρακτηριστικά. Μία απλή ανάλυση θα μπορούσε να αναφέρει:

Τονικό ύψος
: Το πόσο ¨ψηλά¨ ή ¨χαμηλά¨ αντιλαμβανόμαστε έναν ήχο, ή πιο απόλυτα η *συχνότητά* του.

Χροιά
: Το τονικό χρώμα ενός ήχου, το πώς αντιλαμβανόμαστε έναν ήχο σαν σύνολο των επιμέρους τονικών και αρμονικών συχνοτήτων και οργάνων/φορέων ήχου.

Δυναμικές
: Η απόλυτη και σχετική ένταση ενός ήχου

Χρόνος
: Το πώς μεταβάλλεται ένας ήχος σε ένα χρονικό πεδίο ορισμού.

(Owen, 2000)

Βλέπουμε ότι μερικά χαρακτηριστικά είναι ξεκάθαρα, π.χ. μπορούμε πολύ εύκολα να μετρήσουμε το πιο δυνατό σε ένταση σημείο ενός κομματιού με μεγάλη ακρίβεια, ενώ άλλα, όπως η χροιά, δείχνουν ότι θα είναι πιο δύσκολο να μετρηθούν.

1.3. Ήχος και Επιστήμη Υπολογιστών

Σήμερα οι λέξεις ήχος και υπολογιστές ίσως έχουν μεγάλο βαθμό συσχετισμού για τον μέσο δυτικό άνθρωπο, αυτή η κατάκτηση όμως ήρθε μετά από την απαιτούμενη έρευνα. Όταν περίπου στα μισά του 20ου αιώνα κατέστη δυνατό να μεταγράψουμε ένα φυσικό σήμα σε ψηφιακό και αντίστροφα, άρχισαν να διαφαίνονται οι ατελείωτες δυνατότητες χειρισμού των ακουστικών πληροφοριών. Ήδη το 1959, η παρατήρηση ότι υψηλές συχνότητες δεν γίνονται αντιληπτές όταν συνυπάρχουν με χαμηλές συχνότητες (Ehmer, 1959) κατέστησε δυνατό να μπορούμε να συμπιέσουμε ηχητικά σήματα χωρίς να χάνουμε σημαντική, για τον ακροατή, πληροφορία, μικραίνοντας τον χώρο που καταλαμβάνει ένα αρχείο ήχου και οδηγώντας σταδιακά στην ηγεμονία του ψηφιακού μέσου.

Το πώς θα αναπαρασταθεί η ακουστική πληροφορία μέσα από ηχητικά σήματα είναι και το θέμα του άρθρου των Yang, Wang και Shamma (Yang et al., 1992) με τίτλο *Auditory Representations of Acoustic Signals*, όπου ερευνούν αλγόριθμους αναπαράστασης χαρακτηριστικών ήχου σε

συμφωνία με το ακουστικό σύστημα του ανθρώπου. Η υπόθεση εργασίας της έρευνας είναι ακριβώς αυτή: ότι κατανοώντας την λειτουργία του ανθρώπινου ακουστικού συστήματος θα ωφεληθούν προβλήματα αναγνώρισης προτύπων.

Πιο σύγχρονες προσπάθειες έχουν αναδείξει πληθώρα επιλογών για την ανάκτηση πληροφοριών με νόημα από ηχητικά σήματα. Στο άρθρο του Müller (Müller et al., 2011) παρουσιάζεται μια πληθώρα μεθόδων ακόμη και για περίπλοκα προβλήματα όπως αναγνώριση μελωδίας και αρμονίας. Στο project CUIDADO (Peeters, 2004), ένα ενοποιημένο περιβάλλον για περιγραφή μουσικών αρχείων που περιλαμβάνει και ακουστικούς περιγραφείς, καταγράφονται με λεπτομέρεια οι μέθοδοι που χρησιμοποιήθηκαν για εξαγωγή πάνω από 50 διαφορετικών περιγραφών. Τέλος, παρόμοιες τεχνικές χρησιμοποιούνται σε σύγχρονους αλγόριθμους συμπίεσης ήχου (Painter and Spanias, 2000). Οι πληροφορίες που εξάγουμε με την βοήθεια αλγορίθμων από ηχητικά δείγματα θα αναφέρονται από εδώ και στο εξής ως Χαρακτηριστικά (features).

1.4. Μηχανική Μάθηση

Ακόμα και αν καταφέρουμε να εξάγουμε χαρακτηριστικά με νόημα από ένα ηχητικό σήμα δεν θα μπορέσουμε να πάμε πολύ μακριά χωρίς να έχουμε σχεδιάσει και ένα έργο μηχανικής μάθησης. Ένα πρόγραμμα μηχανικής μάθησης διαφοροποιείται από οποιοδήποτε άλλο λογισμικό ως προς το ότι η απόδοσή του βελτιώνεται από την εμπειρία (Mitchell, 1997). Αυτό γίνεται χρησιμοποιώντας διάφορους αλγόριθμους ταξινόμησης από τα μαθηματικά και την στατιστική.

Ίσως το πιο συνηθισμένο έργο μηχανικής μάθησης πάνω σε ηχητικά σήματα είναι η ταξινόμηση μουσικού είδους. Το πρόγραμμα εκπαιδεύεται σε κάποια κομμάτια από διάφορα μουσικά είδη (πχ. ποπ, ροκ, κλασσική κλπ.) και έπειτα μπορεί να αποφανθεί για το τί είδους είναι ένα κομμάτι που δεν έχει 'ακούσει' ποτέ. Οι Tzanetakis και Cook (Tzanetakis and Cook, 2002) χρησιμοποίησαν ένα συνδυασμό χαρακτηριστικών σε τρία σετ (χροιά, ρυθμός και τονικό ύψος) τα οποία τροφοδότησαν σε ταξινομητή αναγνώρισης προτύπων πετυχαίνοντας ακρίβεια 61%, η οποία οι συγγραφείς υποστηρίζουν πως είναι κοντά στην ακρίβεια ανθρώπινων υποκειμένων στο ίδιο έργο. Για το συγκεκριμένο έργο έχουν χρησιμοποιηθεί σύνολα ταξινομητών (Silla et al., 2007), όπου χρησιμοποιώντας (Support Vector Machines / (SVC)) πέτυχαν ακρίβεια περίπου 65%, και νευρωνικά δίκτυα (Scluter and Osendorfer, 2008). Η απόδοση των παραπάνω τεχνικών έχει μελετηθεί και σε δείγματα από μη-δυτική μουσική (Norowi et al., 2005).

Στο ίδιο πλαίσιο αξίζει να αναφέρουμε την ταξινόμηση είδους μέσω μουσικής σημειογραφίας (McKay, 2004), καθώς και την ταξινόμηση με ετικέτες αντί για τάξεις, όπου μπορεί ένα δείγμα να έχει πάνω από

μία ετικέτα (Mandel and Ellis, 2008). Η ταξινόμηση μουσικού είδους δείχνει να έχει ένα σημαντικό εκτόπισμα στην κοινότητα της επιστήμης των υπολογιστών και φαίνεται να υπάρχουν εξελίξεις στο μέλλον (McKay and Fujinaga, 2006).

Ενδιαφέρον υπάρχει επίσης και για έργα αναγνώρισης μουσικών οργάνων (Essid et al., 2004), (Mazarakis et al., 2006), ταξινόμησης με βάση συναισθηματικές ποιότητες της μουσικής (Kim et al., 2010), και διάκρισης μουσικής/ομιλίας (Sheirer and Slaney, 1997). Τέλος, αξίζει να αναφερθούμε στην επέκταση των τεχνικών σε έργα αναγνώρισης φωνημάτων μπαμπουίνων (Janvier et al., 2013) καθώς και τον παγκόσμιο διαγωνισμό του xeno-canto [1] για αυτόματη ταξινόμηση τραγουδιών πτηνών.

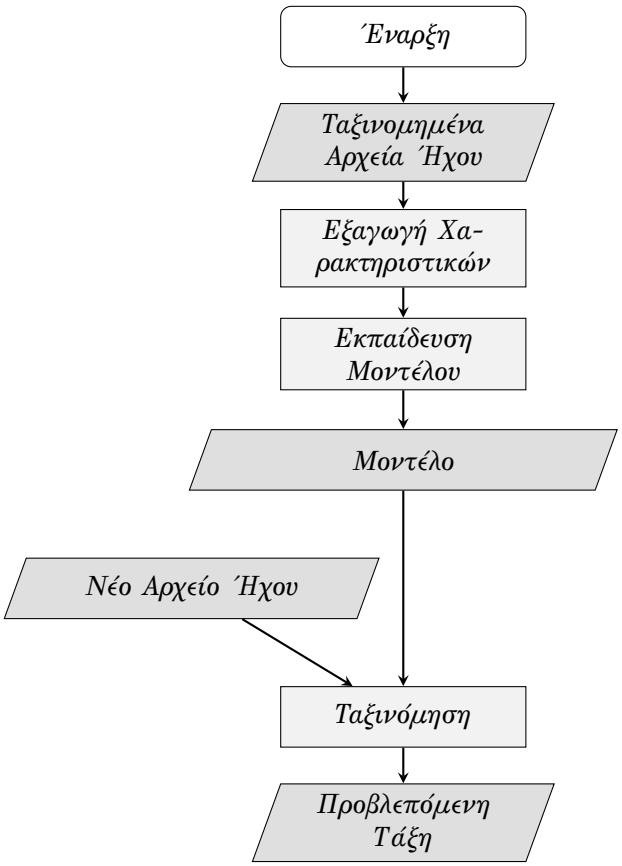
1.5. Το πρόβλημα

Συγκεκριμένες τεχνικές λοιπόν έχουν χρησιμοποιηθεί σε ένα σημαντικό εύρος προβλημάτων, καταδεικνύοντας μία ομοιογένεια ως προς το αντικείμενο, τις τεχνικές και τις μεθόδους που μπορούν να χρησιμοποιηθούν. Με βάση την γνώση που ήδη υπάρχει, θα συνθέσουμε ένα λογισμικό εποπτευόμενης μάθησης το οποίο θα μπορεί να λύσει οποιοδήποτε παρόμοιο με τα παραπάνω πρόβλημα ταξινόμησης ήχων.

Υπάρχουν κάποιες αποφάσεις όμως που πρέπει να πάρουμε πρίν σχεδιαστεί το σύστημα. Το πρόβλημά μας είναι η ταξινόμηση ήχου με βάση το πώς *ακούγεται*, δηλαδή μας ενδιαφέρουν χαρακτηριστικά για ένα χρονικό παράθυρο της τάξης των δεκάδων χιλιοστών του δευτερολέπτου από τα οποία μπορούμε να εξάγουμε πληροφορίες σε σχέση με την χροιά.

Δεδομένου ότι τα χρονικά χαρακτηριστικά δεν φαίνεται να επηρεάζουν την ταξινόμηση ήχων (Li et al., 2001), ο ταξινομητής δεν θα λαμβάνει υπόψιν χρονικά ή τονικά χαρακτηριστικά του σήματος. Το σχεδιάγραμμα του προβλήματός μας φαίνεται στο σχήμα (1):

[1] πρόκειται για μία βάση δεδομένων με πλήθος ήχων πτηνών http://www.xeno-canto.org



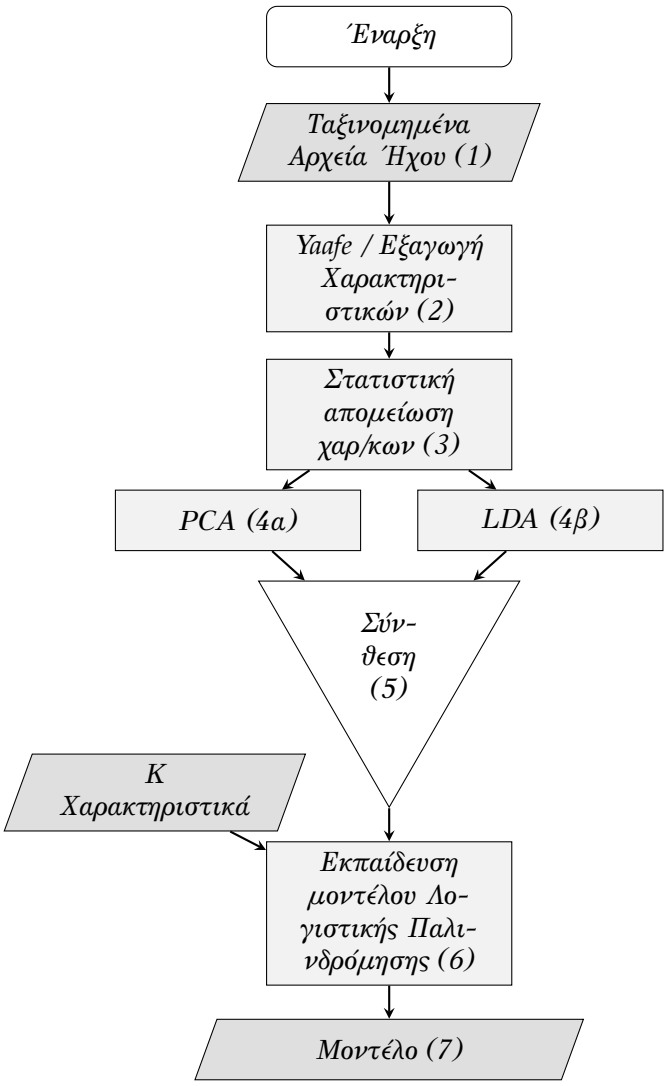
Σχήμα 1: *Σχεδιάγραμμα του προβλήματος*

2. ΣΧΕΔΙΑΣΜΟΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ

Όπως αναφέραμε, ξεκινάμε την έρευνα με στόχο ένα γενικό μοντέλο ταξινόμησης ήχου. Παρόμοιες προσπάθειες με την παρούσα στοχεύουν σε ταξινόμηση συγκεκριμένων τύπων ήχου, όπως αναφέρονται στην παράγραφο 1.4. Αυτό δεν σημαίνει πως η έρευνα αφορά έναν Γενικό Ταξινομητή Ήχων, αλλά πως κατ΄ αρχήν δεν κάνουμε καμία υπόθεση για την φύση του ήχου που θα τροφοδοτήσουμε στον ταξινομητή μας, εκτός από το ότι θα είναι ένα ηχητικό κύμα.

Το πρόγραμμά μας γράφτηκε σε γλώσσα προγραμματισμού Python και χρησιμοποιεί την βιβλιοθήκη Yaafe[2] (Mathieu et al., 2010) για την εξαγωγή των χαρακτηριστικών από τα αρχεία ήχου και την βιβλιοθήκη Scikit-learn[3] (Pedregosa et al., 2011) για το έργο της μηχανικής μάθησης.

[2]http://yaafe.sourceforge.net/
[3]http://scikit-learn.org/stable/



Σχήμα 2: *Σχεδιάγραμμα της Λύσης*

2.1. Ταξινομητής

Βήμα 1

Αρχικά χρειαζόμαστε κάποια αρχεία ήχου. Ένα αρχείο ήχου φέρει 4 χαρακτηριστικά: τον τύπο κωδικοποίησης (η ΄γλώσσα΄ που είναι γραμμένο το αρχείο), τον ρυθμό δειγματοληψίας (συνήθως της τάξης των δεκάδων Khz), το βάθος Bit, δηλαδή με πόσα bits αναπαρίσταται κάθε δείγμα, και την διάρκεια, δηλαδή το πλήθος των δειγμάτων. Επεξεργαζόμαστε

τα αρχεία μέσω του Yaafe, οπότε όσον αφορά τον τύπο αρχείου, υποστηρίζονται διάφοροι τύποι αρχείων όπως WAV, OGG, MP3 και άλλα[4]. Ο ρυθμός δειγματοληψίας πρέπει να εισαχθεί ανάλογα με τα αρχεία που διατίθενται, και μία συνηθισμένη τιμή είναι τα 22050 Hz, ενώ όσον αφορά το βάθος bit, η τιμή 16 θεωρείται δεδομένη. Η διάρκεια για κάθε κομμάτι μπορεί να είναι αυθαίρετη, προτείνεται όμως μιά τιμή μεταξύ 30 και 60 δευτερολέπων. Για τους μετασχηματισμούς των αρχείων χρησιμοποιήθηκε το πρόγραμμα SoX[5].

Βήμα 2

Χρησιμοποιείται η βιβλιοθήκη Yaafe για να εξάγουμε σχεδόν όλα τα διαθέσιμα χαρακτηριστικά που είναι δυνατόν, καθώς χρειαζόμαστε πληθώρα επιλογών για να καλύψουμε ένα δυνητικά ποίκιλο φάσμα εισροών[6]. Όλα τα χαρακτηριστικά υπολογίζονται αρχικά για πλαίσιο 512 δειγμάτων με βήμα 128 δείγματα (παραλείποντας 128 δείγματα για κάθε πλαίσιο). Αυτό είναι και το αρχικό παράθυρο ανάλυσης, και είναι κατάλληλο για να εξάγουμε πληροφορίες σε σχέση με την χροιά. Στην συνέχεια υπολογίζεται ο χρονικός μέσος και η τυπική απόκλιση για κάθε 40 πλαίσια, με βήμα 8, ώστε να πάρουμε μία πιο γενική εικόνα για το δείγμα, μειώνοντας το πλήθος των δεδομένων.

Βήμα 3

Στο σημείο αυτό, και μεμονωμένα για κάθε χαρακτηριστικό, εξάγουμε κάποιους στατιστικούς περιγραφείς για κάθε διάσταση. Αυτοί είναι το ελάχιστο, ο μέσος όρος, η τυπική απόκλιση και το μέγιστο. Για παράδειγμα, το χαρακτηριστικό MFCC αποτελείται από 26 διαστάσεις (13 για τη μέση τιμή και 13 για την τυπική απόκλιση), αντιπροσωπεύοντας διαφορετικά συχνοτικά εύρη, και έχει μήκος ανάλογο της διάρκειας του δείγματος. Μετά τον μετασχηματισμό, έχουμε ένα μονοδιάστατο διάνυσμα μήκους $4 \times a$, όπου a το πλήθος των διαστάσεων, και για το παράδειγμά μας θα έχουμε ένα διάνυσμα μήκους 104[7]. Με αυτόν τον τρόπο όχι μόνο μειώνεται το πλήθος των δεδομένων προς επεξεργασία αλλά και εξισώνονται τα τελικά διανύσματα ανεξάρτητα από την διάρκεια του κάθε δείγματος.

[4] Δυστυχώς στην παρούσα έκδοση του προγράμματος μπορέσαμε να δοκιμάσουμε επιτυχώς μόνο αρχεία WAV.

[5] http://sox.sourceforge.net/

[6] βλ. Παράρτημα Α΄

[7] βλ. Παράρτημα Β΄

Ψευδοκώδικας:

ΠΑΡΑΤΑΞΗ-ΑΡΧΕΙΩΝ-Χ ⇐ **κάθε** ΑΡΧΕΙΟ-ΗΧΟΥ (Χ)
ΠΡΟΣΩΡΙΝΗ-ΠΑΡΑΤΑΞΗ-Χ ⇐ ΚΕΝΗ-ΠΑΡΑΤΑΞΗ
για κάθε Χ στο ΠΑΡΑΤΑΞΗ-ΑΡΧΕΙΩΝ-Χ **κάνε**
  ΕΛ ← ΕΛΑΧΙΣΤΟ(Χ, άξονας=0)
  ΤΔ ← ΤΥΠΙΚΗ-ΔΙΑΚΥΜΑΝΣΗ(Χ, άξονας=0)
  ΜΟ ← ΜΕΣΟΣ-ΟΡΟΣ(Χ, άξονας=0)
  ΜΓ ← ΜΕΓΙΣΤΟ(Χ, άξονας=0)
  ΠΡΟΣΩΡΙΝΗ-ΠΑΡΑΤΑΞΗ-Χ += ΠΑΡΑΤΑΞΗ(ΕΛ, ΤΔ, ΜΟ, ΜΓ)
τέλος για

Βήμα 4, 5

Στα βήματα 4 και 5, μειώνουμε το μήκος των χαρακτηριστικών χρησιμοποιώντας δύο τεχνικές ανάλυσης: τις Principal Component Analysis (PCA) και Linear Discriminant Analysis(LDA). Οι συγκεκριμένες τεχνικές επιλέχθηκαν μετά από εμπειρική δοκιμή των μεθόδων απομείωσης που προσφέρει η βιβλιοθήκη Scikit-Learn. Με τις τεχνικές αυτές αναλύεται το αρχικό δείγμα και επιχειρούμε να κρατήσουμε μόνο τα σημεία που διακρίνουν αποτελεσματικά τις τάξεις. Αρχικά κανονικοποιούμε τις τιμές των χαρακτηριστικών στο εύρος [0, 1], και στη συνέχεια ανάλογα με το μήκος του διανύσματος επιλέγεται το πλήθος των συνιστωσών[8]. Τα δύο παραγόμενα διανύσματα στην συνέχεια συντίθενται σε ένα. Για να πραγματοποιηθούν αυτοί οι μετασχηματισμοί χρησιμοποιούμε τις εξής μεθόδους της βιβλιοθήκης Scikit-Learn:

FeatureUnion: Η μέθοδος αυτή μας επιτρέπει να χρησιμοποιήσουμε εύκολα δύο ή παραπάνω μεθόδους απομείωσης λειτουργώντας παράλληλα: το δείγμα περνάει αυτούσιο από κάθε επιμέρους μετασχηματισμό και στην συνέχεια τα αποτελέσματα συντίθενται σε ένα μονοδιάστατο διάνυσμα.

Pipeline: Η Μέθοδος αυτή είναι παρόμοια με την παραπάνω, επεξεργάζεται όμως τα δείγματα *σειριακά*. Στο παράδειγμά μας, αφού το δείγμα περάσει από επεξεργασία από την MinMaxScaler για να κανονικοποιηθεί στο εύρος [0, 1], στην συνέχεια τροφοδοτείται στην διάταξη PCA-LDA.

Ψευδοκώδικας:

σημ.: όπου SCL: Βιβλιοθήκη Scikit-Learn

ΠΑΡΑΤΑΞΗ-ΑΡΧΕΙΩΝ-Χ ⇐ **κάθε** ΑΡΧΕΙΟ-ΗΧΟΥ (Χ)
ΠΑΡΑΤΑΞΗ-ΤΑΞΕΩΝ-Υ ⇐ **κάθε** ΤΑΞΗ-ΑΡΧΕΙΟΥ (Υ)
ΜΗΚΟΣ-Χ = ΜΗΚΟΣ(ΠΑΡΑΤΑΞΗ-ΑΡΧΕΙΩΝ-ΗΧΟΥ[0])
εάν ΜΗΚΟΣ-Χ > 30 **τότε**
  ΤΟΜΗ = 13
αλλιώς εάν ΜΗΚΟΣ-Χ > 15 **τότε**

[8] βλ. Παράρτημα Γ΄

```
  ΤΟΜΗ = 7
αλλιώς
  ΤΟΜΗ = 3
τέλος εάν
ΣΥΝΘΕΣΗ = SCL.FeatureUnion[SCL.LDA(συνιστώσες=ΤΟΜΗ-1), SCL.PCA(συνιστώσες=ΤΟΜΗ)]

ΑΠΟΜΕΙΩΤΗΣ = SCL.Pipeline[MinMaxScaler(0,1),  ΣΥΝΘΕΣΗ]
ΤΕΛΙΚΗ-ΠΑΡΑΤΑΞΗ-Χ =
  ΑΠΟΜΕΙΩΤΗΣ.ΜΕΤΑΣΧΗΜΑΤΙΣΕ(
    ΠΑΡΑΤΑΞΗ-ΑΡΧΕΙΩΝ-Χ, ΠΑΡΑΤΑΞΗ-ΤΑΞΕΩΝ-Υ)
```

Βήμα 6

Αξιολογούνται τα διαφορετικά χαρακτηριστικά με κριτήριο την ακρίβεια του ταξινομητή και παρουσιάζεται στον χρήστη ένας πίνακας με τα σκόρ για κάθε αύξοντα συνδυασμό καλύτερων χαρακτηριστικών. Ο χρήστης εισάγει τον επιθυμητό αριθμό Κ καλύτερων χαρακτηριστικών και το τελικό μοντέλο εκπαιδεύεται με βάση αυτά. Για την ταξινόμηση χρησιμοποιούμε έναν Ταξινομητή Λογιστικής Παλινδρόμησης. Δοκιμάστηκαν και άλλες επιλογές για την ταξινόμηση, όπως Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machine - SVC), και ο συγκεκριμένος ταξινομητής επιλέχθηκε μετά από εμπειρική δοκιμή με κριτήριο την ακρίβεια για διαφορετικά μήκη διανυσμάτων.

Βήμα 7

Συγκεντρώνονται όλα τα απαιτούμενα μέρη για μελλοντικές ταξινομήσεις και αποθηκεύεται το μοντέλο. Το αρχείο, στις δοκιμές μας, δεν ξεπέρασε το 1MB. Επίσης δίνεται η δυνατότητα για αποθήκευση έκθεσης ταξινόμησης και πίνακα αλληλοεπικάλυψης.

Εδώ το έργο της ταξινόμησης έχει τελειώσει. Σειρά έχει η δοκιμή.

2.2. Δοκιμή

Το πρόγραμμα δοκιμής παίρνει ως εισροή ένα αποθηκευμένο μοντέλο, ένα αρχείο ήχου, και τον ρυθμό δειγματοληψίας του και παρουσιάζει την τάξη στην οποία ανήκει καθώς και την πιθανότητα να ανήκει στην τάξη αυτή.[9].

[9] βλ. Παράρτημα Δ΄

3. ΑΠΟΤΕΛΕΣΜΑΤΑ

Για την αξιολόγηση του Ταξινομητή χρησιμοποιήσαμε έξι υπάρχουσες βάσεις δεδομένων και μία νέα. Αυτές είναι:

i. Βάση δεδομένων μουσικής / ομιλίας GTZAN [10]

ii. Συλλογή μουσικών ειδών GTZAN [11]

iii. Βάση δεδομένων διαγωνισμού ταξινόμησης μουσικού είδους ISMIR2004 [12]

iv. Βιβλιοθήκη δειγμάτων μουσικών οργάνων του Πανεπιστημίου της IOWA [13]

v. Βάση δεδομένων ¨μουσική και συναίσθημα¨ [14]

vi. Βάση δεδομένων ¨απογραφή¨ / an4 του Πανεπιστημίου Carnegie Mellon [15]

vii. Συνθέτες του Ελληνικού τραγουδιού

Αξιολογήθηκε η ακρίβεια, το σκόρ F1 και η ανάκληση[16] του ταξινομητή για κάθε έργο χωρίζοντας τα δείγματα σε σετ εκπαίδευσης και δοκιμής, με το σετ δοκιμής να είναι το 20% του συνόλου. Η αξιολόγηση επικυρώθηκε με τεχνική (Stratified Tenfold Cross-Validation), όπως αυτή υλοποιείται στην βιβλιοθήκη Scikit-learn. Εξαίρεση στο σχήμα αποτελεί η βιβλιοθήκη ISMIR2004, όπου ορίζεται ξεχωριστό σετ εκπαίδευσης και δοκιμής.

[10] Διαθέσιμη στον σύνδεσμο http://marsyasweb.appspot.com/download/data_sets/

[11] Βλ. υπόσημείωση 10.

[12] Διαθέσιμη στον σύνδεσμο http://ismir2004.ismir.net/genre_contest/index.html

[13] Διαθέσιμη στον σύνδεσμο http://theremin.music.uiowa.edu/MIS.html

[14] Διαθέσιμη στον σύνδεσμο http://cvml.unige.ch/databases/emoMusic/

[15] Διαθέσιμη στον σύνδεσμο http://www.speech.cs.cmu.edu/databases/an4/

[16] Ως ακρίβεια ορίζεται το κλάσμα των σωστών προβλέψεων προς όλες τις προβλέψεις, ενώ ο όρος ανάκληση αναφέρεται στο κλάσμα των σωστών προβλέψεων προς το πλήθος των δειγμάτων της κατηγορίας. Το σκόρ F1 είναι ένας δείκτης που λαμβάνει υπ΄όψιν και την ακρίβεια αλλά και την ανάκληση.

3.1. Βάση δεδομένων μουσικής / ομιλίας GTZAN

Έργο: διαχωρισμός μουσικής / ομιλίας
Τάξεις: 2
Δείγματα: 120
Ρυθμός δειγματοληψίας: 22050 Hz
Διάρκεια δειγμάτων: 30 s

Η βάση δεδομένων μουσικής / ομιλίας GTZAN είναι μία από τις δύο συλλογές για έργα ανάκτησης μουσικών πληροφοριών του ερευνητή Γιώργου Τζανετάκη. Περιλαμβάνει δύο τάξεις, δείγματα μουσικής και δείγματα ομιλίας. Στην ακρόαση που πραγματοποιήθηκε, η βάση φάνηκε να έχει ποικιλία δειγμάτων, αρκετά μουσικά είδη και διάφορες γλώσσες στην κατηγορία της ομιλίας.

Η ταξινόμηση σε αυτή την περίπτωση είναι ένα εύκολο έργο, δεδομένων των μεγάλων διαφορών μεταξύ των τάξεων. Με ένα μόνο χαρακτηριστικό ο ταξινομητής πέτυχε ακρίβεια 97.4% και κοντα στο 100% για τρία χαρακτηριστικά. Επιλέξαμε 3 Χαρακτηριστικά, με τελικό μήκος διανύσματος 46, τα οποία είναι:

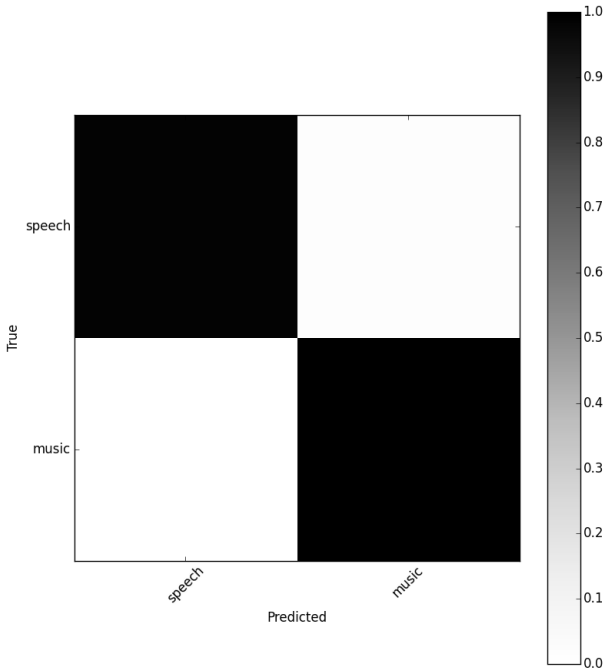
- Loudness
- Temporal Shape Statistics
- Spectral Flatness per Band

Η επίδοση του ταξινομητή φαίνεται στον Πίνακα (1).

Τάξη	Ακρίβεια	Σκόρ F1	Ανάκληση	Υποστήριξη
Ομιλία	0.99	1.00	1.00	130
Μουσική	1.00	0.99	1.00	130
Μέσος Όρος	1.00	1.00	1.00	260

Πίνακας 1: *Έκθεση ταξινόμησης για την βάση δεδομένων μουσικής / ομιλίας GTZAN*

Ο πίνακας αλληλοεπικάλυψης είναι μια καλή οπτική αναπαράσταση για την ακρίβεια ενός ταξινομητή πολλών τάξεων. Στον άξονα Υ βλέπουμε την αληθινή τάξη των αρχείων που εισρέουν και στον άξονα Χ την τάξη που προβλέπει ο Ταξινομητής. Οι τιμές φαίνονται από το χρώμα του τετραγώνου, με 1.0 να σημαίνει πρόβλεψη του 100% των δειγμάτων και 0.0 πρόβλεψη του 0% των δειγμάτων. Είναι προφανές ότι ο πίνακας αλληλοεπικάλυψης ενός τέλειου ταξινομητή σχηματίζει μια μάυρη διαγώνιο απο άνω αριστερά προς κάτω δεξιά.



Σχήμα 3: *Πίνακας αλληλοεπικάλυψης για την βάση δεδομένων μουσικής / ομιλίας GTZAN*

3.2. Συλλογή μουσικών ειδών GTZAN

Έργο: ταξινόμηση μουσικού είδους
Τάξεις: 10 (Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, Rock)
Δείγματα: 1000
Ρυθμός δειγματοληψίας: 22050 Hz
Διάρκεια δειγμάτων: 30 s

Η συλλογή μουσικών ειδών GTZAN (Tzanetakis and Cook, 2002) είναι ίσως η πιο γνωστή βιβλιοθήκη για έργα ταξινόμησης μουσικού είδους. Περιλαμβάνει δείγματα από δέκα διαφορετικά μουσικά είδη, όπως φαίνονται παραπάνω.

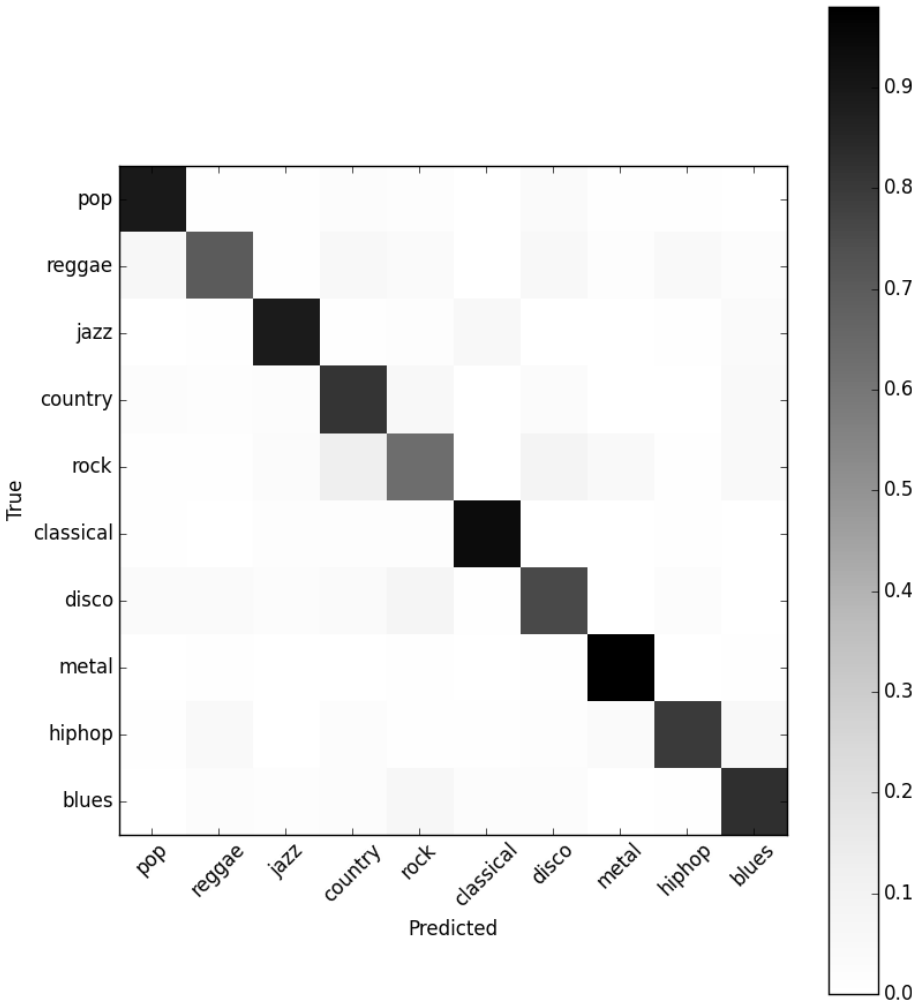
Ο ταξινομητής μας για ένα χαρακτηριστικό πέτυχε ακρίβεια 70.8%. Επιλέχθηκαν 7 χαρακτηριστικά, με μήκος διανύσματος 154, τα οποία είναι τα εξής:

- Amplitude Modulation
- Loudness
- MFCC
- OBSI
- OBSIR
- Spectral Crest Factor per Band
- Spectral Flatness per Band

Η επίδοση του ταξινομητή φαίνεται στον Πίνακα (2).

Τάξη	Ακρίβεια	Σκόρ F1	Ανάκληση	Υποστήριξη
Pop	0.86	0.90	0.88	200
Reggae	0.82	0.70	0.76	200
Jazz	0.88	0.89	0.88	200
Country	0.72	0.81	0.77	200
Rock	0.69	0.64	0.66	200
Classical	0.92	0.94	0.93	200
Disco	0.75	0.76	0.75	200
Metal	0.90	0.98	0.94	200
Hiphop	0.89	0.80	0.84	200
Blues	0.80	0.83	0.82	200
Μέσος Όρος	0.82	0.82	0.82	2000

Πίνακας 2: *Έκθεση ταξινόμησης για την συλλογή μουσικών ειδών GTZAN*



Σχήμα 4: *Πίνακας αλληλοεπικάλυψης για την συλλογή μουσικών ειδών GTZAN*

Από τον παραπάνω πίνακα μπορούμε να εξάγουμε αρκετά συμπεράσματα: Η ροκ και η ντίσκο φαίνεται να έχουν μια αλληλοεπικάλυψη, κομμάτια ροκ δείχνουν περίπου 10-15% σαν κάντρι και κομμάτια ρέγκε δείχουν περίπου 5-10% σαν ποπ, κάντρι, ντίσκο ή χιπ-χοπ.

3.3. Βάση δεδομένων διαγωνισμού ταξινόμησης μουσικού είδους ISMIR2004

Έργο: ταξινόμηση μουσικού είδους
Τάξεις: 6 (Rock / Pop, Jazz / Blues, Classical, Electronic, Metal / Punk, World)
Δείγματα: 722
Ρυθμός δειγματοληψίας: 22050 Hz
Διάρκεια δειγμάτων: 4 μέχρι 59 s

Η κοινότητα International Society for Music Information Retrieval - ISMIR διοργανώνει κάθε χρόνο το ομόνυμο συνέδριο όπου παρουσιάζονται άρθρα πάνω στο συγκεκριμένο πεδίο και διοργανώνονται διαγωνισμοί σε έργα ανάκτησης πληροφορίας από μουσική. Στα πλαίσια του ISMIR το 2004 υπήρξε και διαγνωσιμός στην ταξινόμηση μουσικού γένους. Οι διαγωνιζόμενοι είχαν κάποια δείγματα σε ένα σετ εκπαίδευσης, έπρεπε να χτίσουν ένα ταξινομητή και στη συνέχεια να μετρήσουν την ευστοχία του ταξινομητή σε ένα καινούργιο σετ δειγμάτων δοκιμής. Με τον ίδιο τρόπο αξιολογείται και η δική μας υλοποίηση.

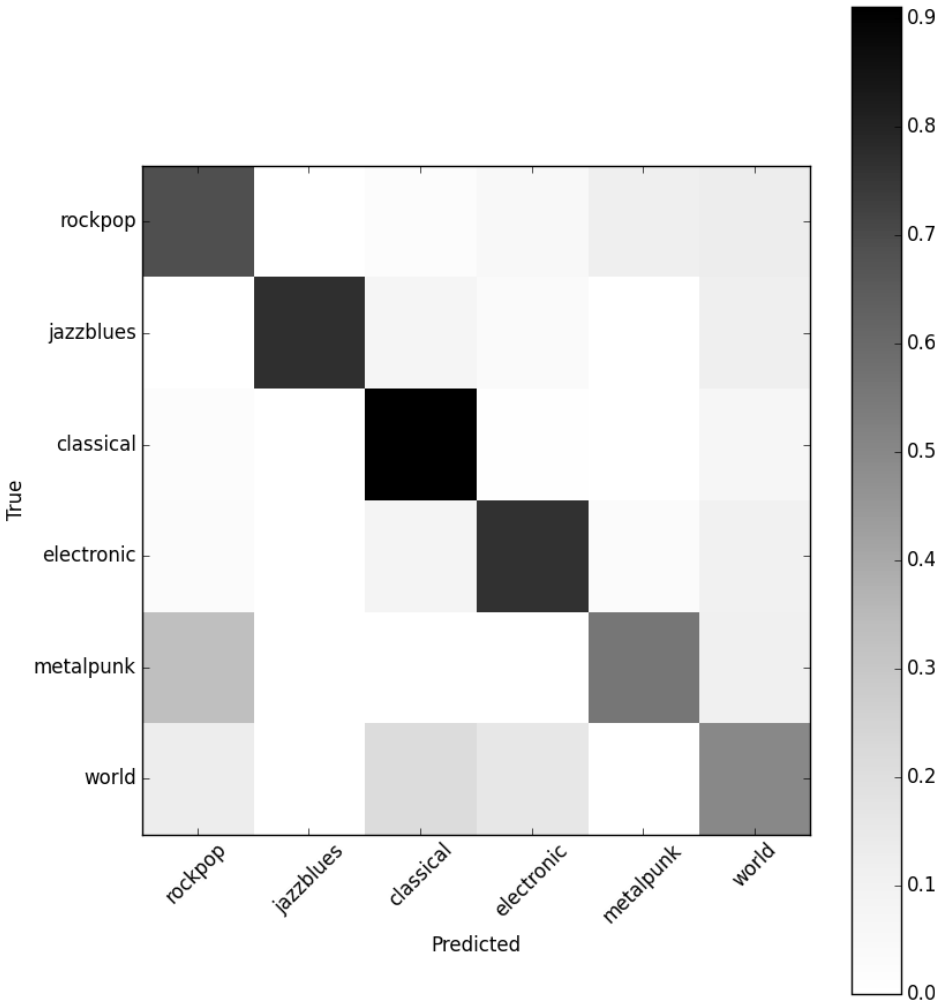
Εφαρμόστηκε επαναδειγματοληψία στα 22050 Hz και περικοπή ούτως ώστε τα δείγματα μεγαλύτερα του ενός λεπτού να περιοριστούν σε 40 δευτερόλεπτα, παραλείποντας τα 20 πρώτα δευτερόλεπτα του κομματιού. Ο ταξινομητής μας για ένα χαρακτηριστικό πέτυχε ακρίβεια 72.6%. Αξίζει να επισημάνουμε πως με την προσθήκη περισσότερων χαρακτηριστικών δεν επήλθε η βελτίωση που υπήρχε σε άλλες βάσεις δεδομένων. Επιλέχθηκαν 4 χαρακτηριστικα, με μήκος διανύσματος 72, τα οποία είναι τα εξής:

- MFCC
- OBSI
- OBSIR
- Spectral Shape Statistics

Η επίδοση του ταξινομητή φαίνεται στον Πίνακα (3).

Τάξη	Ακρίβεια	Σκόρ F1	Ανάκληση	Υποστήριξη
Rock / Pop	0.56	0.57	0.57	200
Jazz / blues	0.76	0.56	0.64	50
Classical	0.87	0.95	0.90	630
Electronic	0.72	0.73	0.72	230
Metal / Punk	0.52	0.44	0.48	90
World	0.62	0.54	0.58	250
Μέσος Όρος	0.74	0.74	0.74	1450

Πίνακας 3: *Έκθεση ταξινόμησης για την βάση δεδομένων ISMIR2004*



Σχήμα 5: *Πίνακας αλληλοεπικάλυψης για την βάση δεδομένων ISMIR2004*

Οι Πίνακες αλληλοεπικάλυψης συνεχίζουν να μας φανερώνουν πολύτιμες πληροφορίες. Στον παραπάνω πίνακα, βλέπουμε ότι όλα τα είδη μερικές φορές δείχνουν σαν μουσικές του κόσμου με φανερή εξαίρεση την κλασσική μουσική. Αυτό είναι αρκετά εύλογο, καθώς η κατηγορία μουσικές του κόσμου αναφέρεται σε ένα ευρύ φάσμα μουσικών ειδών τα οποία αδυνατούμε να κατατάξουμε σε κάποια άλλη κατηγορία δυτικής μουσικής.

3.4. Βάση δειγμάτων μουσικών οργάνων του Πανεπιστημίου της IOWA - MIS (Musical Instrument Samples)

Έργο: ταξινόμηση μουσικών οργάνων
Τάξεις: 19 (Clarinet, Oboe, Trumpet, Tambourine, Xylophone, Cymbals, Basoon, Flute, Violin, Horn, Bass, Tuba, Trombone, Cello, Vibraphone, Viola, Sax, Bells, Marimba)
Δείγματα: 1623
Ρυθμός δειγματοληψίας: 44100 Hz
Διάρκεια δειγμάτων: 4 μέχρι 59 s

Το Πανεπιστήμιο της IOWA αξιοποίησε τα στούντιο μουσικής που διαθέτει για να συγκροτήσει μία εκτενέστατη βάση δεδομένων με δείγματα από μουσικά όργανα. Μπορούμε να βρούμε δείγματα για πολλά όργανα τα οποία χωρίζονται σε ένα αρχείο ήχου για κάθε νότα στο εύρος του οργάνου. Επιλέξαμε μερικά από αυτά για να δοκιμάσουμε τον ταξινομητή.

Ο Ταξινομητής μας για ένα χαρακτηριστικό πέτυχε ακρίβεια 96.9%. Λόγω της απομόνωσης των δειγμάτων είναι εύκολο να πετύχουμε μεγάλη ακρίβεια. Επιλέχθηκαν 3 χαρακτηριστικα, με μήκος διανύσματος 75, τα οποία είναι τα εξής:

- Loudness
- OBSI
- OBSIR

Η επίδοση του ταξινομητή φαίνεται στον Πίνακα (4).

Τάξη	Ακρίβεια	Σκόρ F1	Ανάκληση	Υποστήριξη
Κλαρινέτο	0.99	1.00	0.99	260
Όμποε	0.96	0.99	0.97	70
Τρομπέτα	1.00	1.00	1.00	70
Ντέφι	1.00	1.00	1.00	70
Ξυλόφωνο	0.99	1.00	1.00	30
Κύμβαλα	1.00	0.99	1.00	150
Φαγκότο	0.96	1.00	0.98	80
Φλάουτο	0.99	1.00	0.99	220
Βιολί	0.98	0.96	0.97	180
Κόρνο	1.00	0.98	0.99	90
Κοντραμπάσσο	1.00	0.99	0.99	210
Τούμπα	0.95	1.00	0.97	70
Τρομπόνι	1.00	0.97	0.99	120
Τσέλο	0.98	0.99	0.99	190
Βιμπράφωνο	0.98	0.97	0.98	250
Βιόλα	0.98	0.99	0.99	200
Σαξόφωνο	0.98	0.98	0.98	130
Κουδνούνια	1.00	0.95	0.97	80
Μαρίμπα	0.99	0.99	0.99	490
Μέσος Όρος	0.99	0.99	0.99	3240

Πίνακας 4: *Έκθεση ταξινόμησης για την βάση δεδομένων μουσικών οργάνων του Πανεπιστημίου της IOWA*

Λόγω των μικρών διαφορών, ο πίνακας αλληλοεπικάλυψης δεν εμφανίζει χρήσιμη πληροφορία για αυτό και παραλείπεται.

3.5. Βάση δεδομένων ¨μουσική και συναίσθημα¨

Έργο: ταξινόμηση μουσικής με βάση το συναίσθημα
Τάξεις: 4 (Sad, Happy, Relax, Angry)
Δείγματα: 2906
Ρυθμός δειγματοληψίας: 22050 Hz
Διάρκεια δειγμάτων: 30 ή 60 s

Σε άρθρο που παρουσιάστηκε στο 12ο συνέδριο ISMIR, εισήχθη μία βάση δεδομένων για ταξινόμηση μουσικής με βάση το συναίσθημα (Song et al., 2012). Οι ετικέτες ανακτήθηκαν από το γνωστό ιστότοπο υπηρεσιών μουσικού περιεχομένου last.fm και έχουν διαμορφωθεί από χιλιάδες χρήστες ανά τον κόσμο.

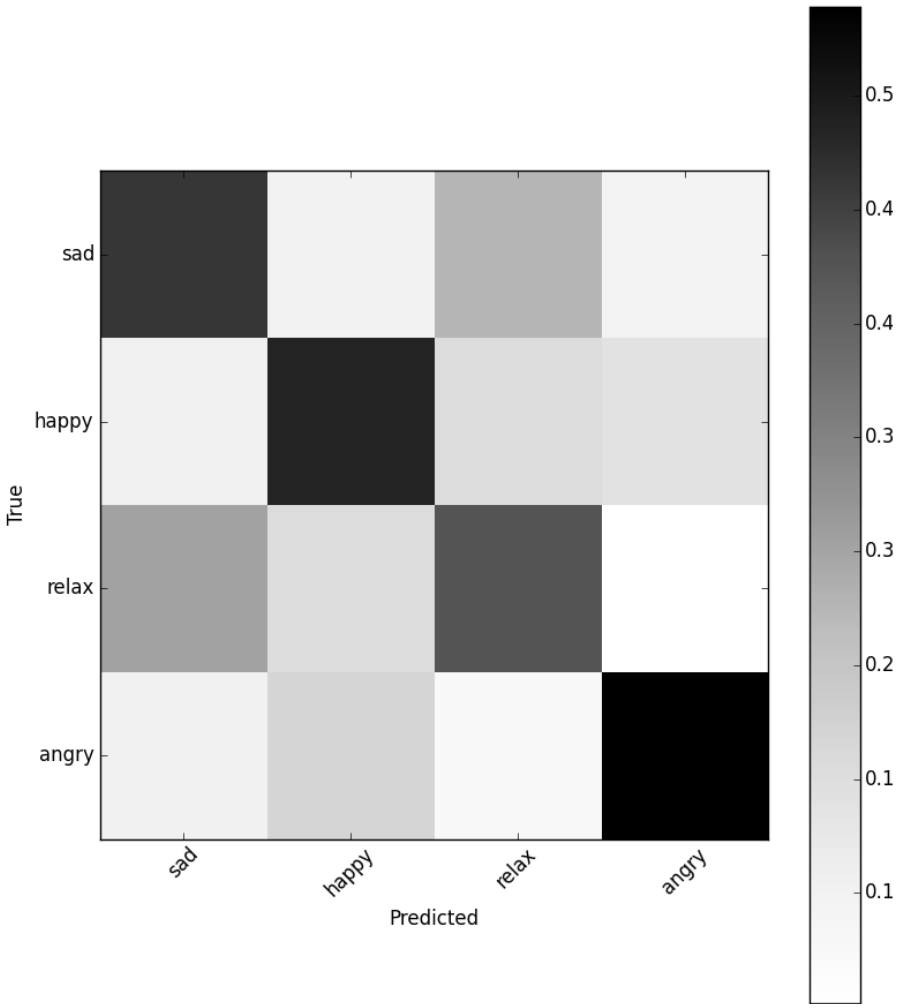
Ο ταξινομητής μας για ένα χαρακτηριστικό πέτυχε ακρίβεια 48.2%. Ο αριθμός αυτός μπορεί να φαίνεται χαμηλός, πρόκειται όμως για μια βάση δεδομένων με αρκετά υποκειμενικά χαρακτηριστικά. Επιλέχθηκαν 4 χαρακτηριστικά, με μήκος διανύσματος 64, τα οποία είναι τα εξής:

- Loudness
- MFCC
- OBSI
- OBSIR

Η επίδοση του Ταξινομητή φαίνεται στον Πίνακα (5).

Τάξη	Ακρίβεια	Σκόρ F1	Ανάκληση	Υποστήριξη
Λυπημένο	0.49	0.50	0.49	1530
Χαρούμενο	0.52	0.52	0.52	1510
Χαλαρό	0.46	0.45	0.45	1500
Θυμωμένο	0.59	0.59	0.59	1280
Μέσος Όρος	0.51	0.51	0.51	5820

Πίνακας 5: *Έκθεση ταξινόμησης για την βάση δεδομένων μουσική και συναίσθημα*



Σχήμα 6: *Πίνακας αλληλοεπικάλυψης για την βάση δεδομένων μουσική και συναίσθημα*

Έχει ενδιαφέρον να παρατηρήσουμε οτι ενώ οι κατηγορίες ¨θυμωμένο¨ και ¨χαρούμενο¨ είναι πιο αναγνωρίσιμες από τον ταξινομητή, οι κατηγορίες ¨λυπημένο¨ και ¨χαλαρό¨ επικαλύπτονται σχετικα μεταξύ τους, πράγμα που είναι και εύλογο. Επίσης είναι προφανές ότι δείγματα που ανήκουν στην τάξη ¨χαλαρό¨ είναι απίθανο να ταξινομηθούν ώς ¨θυμωμένο¨.

3.6. Βάση δεδομένων ¨απογραφή¨ / an4 του Πανεπιστημίου Carnegie Mellon

Έργο: αναγνώριση ομιλητή
Τάξεις: 84 (84 διαφορετικοί ομιλητές)
Δείγματα: 1087
Ρυθμός δειγματοληψίας: 16000 Hz
Διάρκεια δειγμάτων: <5 s

Η Βάση Δεδομένων ¨απογραφή¨ του τμήματος αναγνώρισης ομιλίας του Πανεπιστημίου Carnegie Mellon ηχογραφήθηκε εσωτερικά στο Πανεπιστήμιο με σκοπό να χρησιμοποιηθεί σε έργα αναγνώρισης λόγου. Οι Συμμετέχοντες κλήθηκαν να πουν κάποιες τυχαίες πληροφορίες, και το αποτέλεσμα μοιάζει με απογραφή, εξ αυτού και το όνομα της βάσης. Στην έρευνά μας θα χρησιμοποιηθεί για κάτι λίγο διαφορετικό: για Αναγνώριση Ομιλητή.

Ο Ταξινομητής μας για ένα χαρακτηριστικό πέτυχε ακρίβεια 80%. Έχει ενδιαφέρον το γεγονός οτι με επιπλέον χαρακτηριστικά υπήρξε μεγάλη βελτίωση στην ακρίβεια του ταξινομητή. Επιλέχθηκαν 4 χαρακτηριστικά, με μήκος διανύσματος 100, τα οποία είναι τα εξής:

- Loudness
- MFCC
- OBSI
- Spectral Flatness per Band

Λόγω του μεγάλου αριθμούυ των τάξεων δέν έχει νόημα να αναφέρουμε τα αποτελέσματα για όλες τις επιμέρους τάξεις. Αντί αυτού, θα αναφέρουμε μερικά χαρακτηριστικά.

Η μικρότερη ακρίβεια που είχαμε για τάξη είναι 0.73, και η μέγιστη 1.00. Η μέση ακρίβεια και ανάκληση είναι 0.94 και κάθε τάξη υποστηρίζεται από 20 ή 30 δείγματα ανάλογα με το πλήθος των δειγμάτων, 2390 σύνολο για όλες τις τάξεις.

3.7. Συνθέτες του Ελληνικού τραγουδιού

Έργο: ταξινόμηση συνθέτη
Τάξεις: 12 (Ασίκης, Διαμαντίδης, Γκόγκος, Χατζηχρήστος, Καλδάρας, Μητσάκης, Μπάτης, Παπαϊωάννου, Περιστέρης, Σκαρβέλης, Τούντας, Βαμβακάρης)
Δείγματα: 685
Ρυθμός δειγματοληψίας: 22050 Hz
Διάρκεια δειγμάτων: 45 s

Δημιουργήσαμε μία βάση δεδομένων με Έλληνες λαϊκούς συνθέτες της περιόδου 1920-1950. Τα κομμάτια συλλέχθηκαν από την συλλογή "Έλληνες Συνθέτες" της MINOS - EMI. Ο ταξινομητής μας για ένα χαρακτηριστικό πέτυχε ακρίβεια 61%. Τα κομμάτια περικόπηκαν από το 20ό δευτερόλεπτο ώστε η διάρκεια τους να είναι 45 δευτερόλεπτα. Επιλέχθηκαν 8 χαρακτηριστικά, με μήκος διανύσματος 181, τα οποία είναι τα εξής:

- Loudness
- LPC
- MFCC
- OBSI
- OBSIR
- Spectral Crest Factor per Band
- Spectral Flatness Per Band
- Spectral Shape Statistics

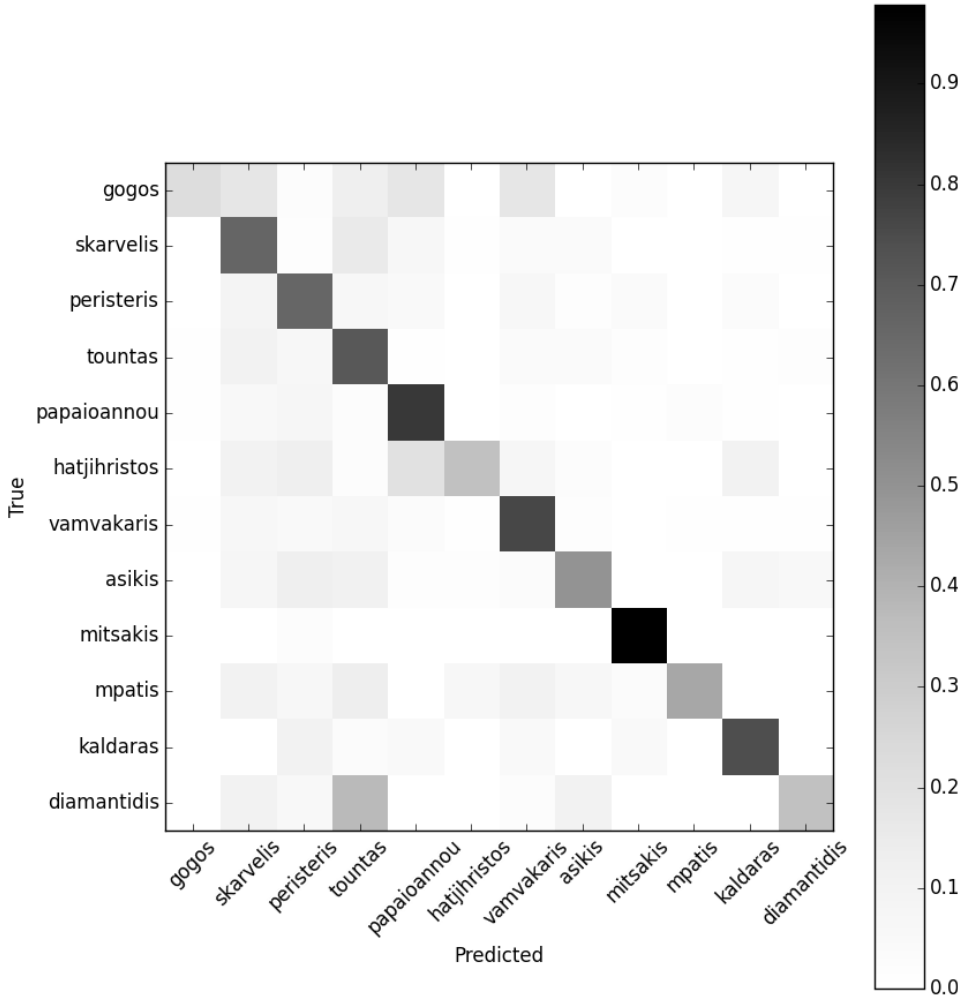
Η επίδοση του ταξινομητή φαίνεται στον Πίνακα (6).

Τάξη	Ακρίβεια	Σκόρ F1	Ανάκληση	Υποστήριξη
Γκόγκος	0.75	0.23	0.35	40
Σκαρβέλης	0.59	0.66	0.62	180
Περιστέρης	0.64	0.66	0.65	180
Τούντας	0.63	0.71	0.67	220
Παπαϊωάννου	0.71	0.80	0.75	140
Χατζηχρήστος	0.67	0.35	0.46	40
Βαμβακάρης	0.78	0.76	0.77	220
Ασίκης	0.55	0.50	0.52	70
Μητσάκης	0.90	0.98	0.94	140
Μπάτης	0.76	0.43	0.55	30
Καλδάρας	0.71	0.74	0.73	70
Διαμαντίδης	0.56	0.35	0.43	40
Μέσος Όρος	0.69	0.69	0.69	1370

Πίνακας 6: *Έκθεση ταξινόμησης για την βάση δεδομένων Ελλήνων συνθετών*

Θεωρούμε ότι η επίδοση είναι άκρως ικανοποιητική, ιδίως αν συνυ-

πολογίσουμε οτι για έναν συνθέτη έχουμε διαφορετικούς ερμηνευτές και διαφορετικές ενορχηστρώσεις.



Σχήμα 7: *Πίνακας αλληλοεπικάλυψης για την βάση δεδομένων Ελλήνων συνθετών*

Ενδεχομένως κάποιος που γνωρίζει μπορεί να επιχειρηματολογήσει, ότι ο Διαμαντίδης μοιάζει πολύ με τον Τούντα καθώς και οι δύο χρησιμοποιούσαν κατά κόρον βιολί, ή ότι ο Μητσάκης είναι κορυφαίος συνθέτης διότι δεν μοιάζει με κανέναν.

4. ΣΥΖΗΤΗΣΗ

χρησιμοποιώντας την υπάρχουσα γνώση πάνω στην εξαγωγή πληροφοριών από ηχητικά δείγματα και την μηχανική μάθηση συνθέσαμε έναν ταξινομητή ικανό να δουλέψει με ένα φάσμα ήχων. Τα αποτελέσματά μας φανερώνουν μία πάρα πολύ καλή ακρίβεια. Για το έργο (2), μία από τις πιο επιτυχημένες υλοποιήσεις είχε ακρίβεια 83% (Bergestra et al., 2006), σε σχέση με το δικό μας 82%. Στο έργο (3) ο ταξινομητής μας ήταν ακριβής στο 74% των περιπτώσεων, ενώ ίσως το κορυφαίο σκόρ που έχει επιτευχθεί είναι 83.5% (Holzapfel and Stylianou, 2008). Στο έργο (5), οι δημιουργοί της βάσης δεδομένων είχαν ακρίβεια 54% (Song et al., 2012), μόλις 3% πάνω από την δική μας ταξινόμηση.

Όσον αφορά τα υπόλοιπα έργα, δεν έχουν ξαναχρησιμοποιηθεί οι βάσεις δεδομένων για παρόμοια έργα και κατ΄ επέκταση δεν έχουμε κάτι με το οποίο να πραγματοποιήσουμε σύγκριση. Μπορούμε να πούμε με ασφάλεια πως υπερ-καλύψαμε απλά έργα όπως ο διαχωρισμός μουσικής / ομιλίας, η ταξινόμηση μουσικών οργάνων και η αναγνώριση ομιλητή. Όπως είδαμε στο τελευταίο έργο, οι πληροφορίες που εξάγουμε μπορούν ενδεχομένως να χρησιμοποιηθούν και σε μουσικολογικές έρευνες.

Θεωρούμε πως είναι εύλογο να ερευνήσουμε για ένα γενικό μοντέλο ταξινόμησης ήχου σε αντιδιαστολή με εξειδικευμένα μοντέλα. Ενδεχομένως πολλά συστήματα π.χ. για ταξινόμηση μουσικού είδους, να μπορούν να ταξινομήσουν και άλλου τύπου δείγματα, προκύπτει λοιπόν το ερώτημα για το αν είναι σωστή η κατεύθυνση της εξειδίκευσης. Το πρόγραμμά μας σε όλα τα έργα έφερε αποτελέσματα πολύ κοντά σε αποτελέσματα εξειδικευμένων σε ένα έργο διατάξεων. Μπορούμε να υποθέσουμε πως κάθε ήχος ανεξαρτήτως του αν είναι μουσική, ομιλία ή θόρυβος φέρει κάποια καθολικά ηχητικά χαρακτηριστικά τα οποία μπορεί ένα πρόγραμμα να επεξεργαστεί σε ένα αγνωστικιστικό πλαίσιο ως προς την προέλευση του, οδηγώντας σε μία γενική λύση του προβλήματος της ταξινόμησης του.

Σε καμία περίπτωση, βέβαια, δεν θεωρούμε ολοκληρωμένη την υλοποίησή μας. Κατάρχάς, χρειάζεται να δοκιμαστούν και άλλες βάσεις δεδομένων με ήχους πέρα απο μουσική και ομιλία. Όσον αφορά τις τεχνικές που χρησιμοποιήθηκαν, Θα χρειαστεί να ερευνηθούν τρόποι για ακόμη πιο επιλεκτική μείωση των χαρακτηριστικών, ενδεχομένως και με νευρωνικά δίκτυα, έτσι ώστε να μεγιστοποιήσουμε την σχετική πληροφορία για κάθε έργο, μειώνοντας τον όγκο των δεδομένων. Με αυτό τον τρόπο εικάζουμε πως θα βελτιωθεί και το εύρος της γενικότητας του ταξινομητή, καθώς θα χρησιμοποιούνται μόνο οι σχετικές πληροφορίες για τον κάθε τύπο δείγματος. Υπάρχει πλούσια βιβλιογραφία πάνω σε νέες μεθόδους που μπορούν να δοκιμαστούν, και σε μεθόδους που ήδη λειτουργούν και χρησιμοποιούνται. Μπορούμε να διακινδυνέυσουμε την πρόβλεψη ότι τα επόμενα χρόνια η αυτόματη ταξινόμηση ήχων θα χρησιμοποιείται καθημερινά απο όλους τους δυτικούς ανθρώπους.

Βιβλιογραφία

J. Bergestra, N. Casagrande, D. Erhan, D. Eck, B. Kegl, *Aggregate Features and ADABOOST for Music Classification*, Machine Learning, Vol. 65, Issue 2-3, 2006.

R, Ehmer, *Masking by Tones vs Noise Bands*, The Journal of the Acoustical Society of America, Vol. 31, Issue: 9, Acoustical Society of America, New York, USA, 1959.

S. Essid, G. Richard, B. David, *Efficient Musical Instrument Recognition on Solo Performance Music Using Basic Features*, Audio Engineering Society 25th International Conference, London, UK, 2004.

M. Janvier, R. Horaud, L. Girin, F. Berthommier, L. Boe et al., *Supervised Classification of Baboon Vocalizations*, NIPS4B - Neural Information Processing Scaled for Bioacousics, Nevada, USA, 2013

A. Holzapfel, Y. Stylianou, *Musical Genre Classification Using Nonnegative Matrix Factorization-Based Features*, IEEE Transactions on Audio, Speech, and Language Processing, Vol.15, No. 2, 2008.

Y. Kim, Erik, Schimdt, R. Migneco, O, Morton, P. Richardson, J. Scorr, J. Speck, D. Turnbull, *Emotion Recognition: A State of the Art Review (2010)*, Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR), Utrecht, Netherlands, 2010.

S. Li, I. Sethi, N. Dimitrova, T. McGee, *Classification of General Audio Data for Content-Based Retrieval*, Pattern Recognition Letters, Vol. 22, Issue 5, 2001.

M. Mandel, D. Ellis, *Multiple-Instance Learning for Music Information Retrieval*, Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR), Philadelphia, USA, 2008.

B. Mathieu, S. Essid, T. Fillon, J. Prado, G. Richard, *YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software*, Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR), Utrecht, Netherlands, 2010.

G. Mazarakis, P. Tzevelekos, G. Kouroupetroglou, *Musical Instrument Recognition and Classification Using Time Encoded Signal Processing and Fast Artificial Neural Networks*, 4th Hellenic Conderence on AI, SETN, Heraklion, Greece, 2006.

C. McKay, *Automatic Genre Classification of MIDI Recordings*, Master Thesis, McGill University, Montreal, Canada, 2004.

C. McKay, I. Fujinaga, *Musical Genre Classification: Is it Worth Pursuing and how can it be Improved?*, Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR), Victoria, Canada, 2006.

T. Mitchell, *Machine Learning*, McGraw Hill, Ohio, USA, 1997.

M. Müller, D. P. W. Ellis, A. Klapuri, G. Richard, *Signal Processing for Music Analysis*, IEEE Jounal of Selected Topics in Signal Procesing, Vol. 5, Issue: 6, IEEE, USA, 2011.

N. Norowi, S. Doraisamy, R. Wirza, *Factors Affectig Automatic Genre Classification: An Investigation Incorporatin Non-Western Musical Forms*, Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR), London, UK, 2005.

H. Owen, *Music Theory Resource Book*, Oxford University Press, Oxford, UK, 2000.

T. Painter, A. Spanias, *Perceptual Coding of Digital Audio*, Proceedings of the IEEE, Vol. 88, Issue: 4, IEEE, USA, 2000.

F. Predregosa, G. Varoquax, A. Gramfort, V. Michel, B. Thirion, B. Dubourg, J. Vanderpls, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research, Vol. 12, 2011.

G. Peeters, *A Large set of Audio Features for Sound Description (Similarity and Classification) in the CUIDAO project*, Technical report, 2004.

E. Scheirer, M. Slaney, *Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator*, IEEE International Conference on Acoustics, Speech and Signal Processing, Munich, Germany, 1997.

J. Schlüter, C. Osendorfer, *Music Similarity Estimation with the Mean-Covariance Restricted Boltzmann Machine*, 11th International Conference on Machine Learning, Vol. 2, IEEE, USA, 2011.

C. Silla Jr., C. Kaestner, A. Koerich, *Automatic Music Genre Classification Using Ensemble Classifiers*, IEEE International Conference on Systems, Man and Cybernetics, IEEE, USA, 2007.

Y. Song, S. Dixom, M. Pearce, *Evaluation of Musical Features for Emotion Classification*, Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR), Porto, Portugal, 2012

G. Tzanetakis, P. Cook, *Musical Genre Classification of Audio Signals*, IEEE Transactions on Speech and Audio Processing, Vol. 10, Issue: 5, IEEE, 2002.

X. Yang, K. Wang and S. A. Shamma, *Auditory Representations of Acoustic Signals*, IEEE Transactions on Information Theory, Vol. 38, Issue: 2, IEEE USA, 1992.

Παράρτημα

Α΄. Τα Ακουστικά Χαρακτηριστικά

Amplitude Modulation
Περιγράφει το τρέμολο και τον κόκκο, δηλαδή ταχείς περιοδικές διακυμάνσεις στον ήχο.

Auto Correlation
Δίνει πληροφορίες σχετικά με επαναλαμβανόμενα μοτίβα στον ήχο, το πόσο ένα σήμα συσχετίζεται με τον εαυτό του.

Complex Domain Onset Detection
Περιγράφει την εμφάνιση ήχων σε ένα δείγμα.

Envelope Shape Statistics
Εκτιμά το κέντρο βάρους, την εξάπλωση, την ασυμετρία και την κύρτωση για ένα σήμα με βάση τον Φάκελο (Envelope). Πρόκεται για βασικά περιγραφικά χαρακτηριστικά για έναν ήχο.

Loudness
Εκτιμά την υποκειμενική εντύπωση για το πόσο δυνατός είναι ένας ήχος.

LPC
Κωδικοποιεί το συχνοτικό φάσμα ενός σήματος χρησιμοποιώντας γραμμική πρόβλεψη.

Mel Spectrum
Χωρίζει το σήμα σε τμήματα σύμφωνα με τις μπάντες Mel, οι οποίες βασίζεται στην ανθρώπινη ακοή.

MFCC
Υπολογίζει μια συχνοτική αναπαράσταση χρησιμοποιώντας αντιληπτικά αξιώματα.

OBSI
Περιγράφει την σχετική κατανομή του σήματος χωρίζοντας το σήμα σε οκτάβες.

OBSIR
Υπολογίζει τον λογάριθμο του χαρακτηριστικού OBSI.

Perceptual Sharpness
Υπολογίζει την Οξύτητα ενός ήχου.

Perceptual Spread
: Υπολογίζει το εύρος της χροιάς, πόσο εκτείνονται σε συχνότητες τα χαρακτηριστικά της χροιάς ενός ήχου.

Spectral Crest Factor per Band
: Υπολογίζει πόσο δυνατά είναι τα δυνατά σημεία ενός ήχου (Peaks), σε σχέση με τον μέσο όρο για κάθε συχνοτικό εύρος.

Spectral Decrease
: Υπολογίζει το πόσο απότομα μειώνεται το σύνολο των συχνοτήτων του ήχου.

Spectral Flatness
: Υπολογίζει το πόσο επίπεδο συχνοτικά είναι ένα σήμα.

Spectral Flatness per Band
: Υπολογίζει το πόσο επίπεδο συχνοτικά είναι ένα σήμα, αφού χωριστεί σε συχνοτικά τμήματα.

Spectral Rolloff
: Υπολογίζει τον θόρυβο σε ένα σήμα υπολογίζοντας το υψηλό όριο της συχνότητας για το έυρος στο οποίο περιέχεται το 95% της ενέργειας του σήματος.

Spectral Shape Statistics
: Εκτιμά το κέντρο Βάρους, την εξάπλωση, την ασυμετρία και την κύρτωση για το συχνοτικό φάσμα ενός σήματος.

Spectral Slope
: Υπολογίζει πόσο μειώνεται το συχνοτικό φάσμα ενός σήματος στις υψηλές συχνότητες.

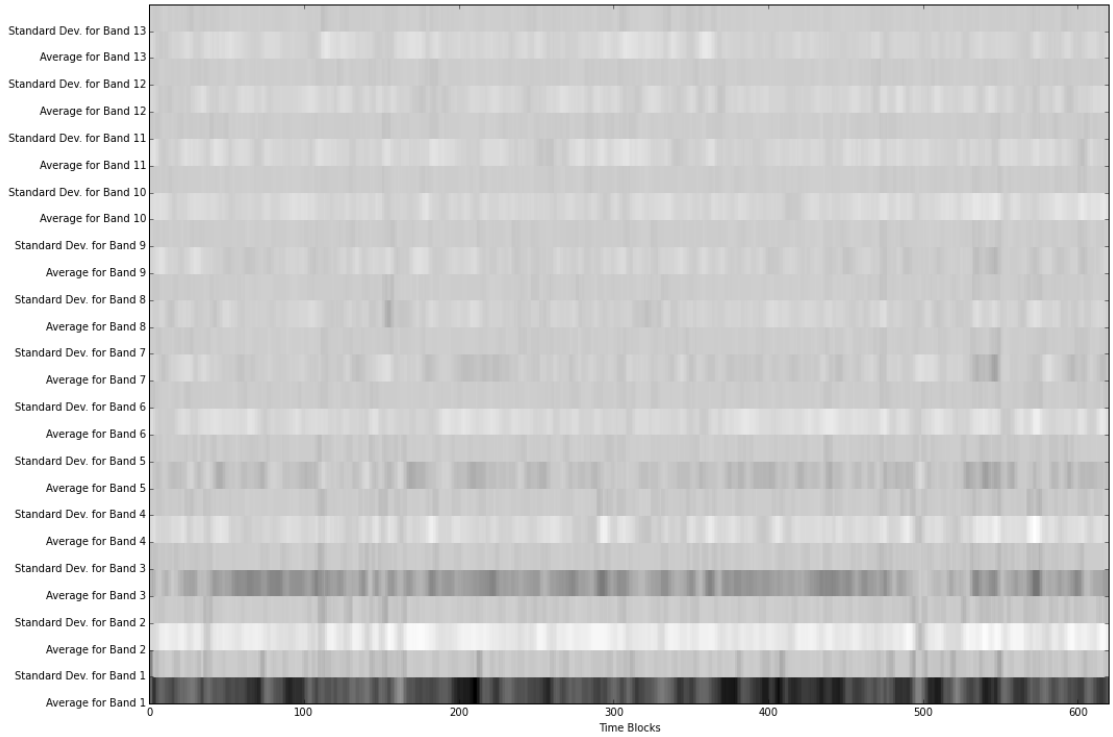
Spectral Variation
: Υπολογίζει τις συχνοτικές διακυμάνσεις ενός σήματος.

ZCR
: Υπολογίζει τον ρυθμό με τον οποίο σήμα αλλάζει από θετικό σε αρνητικό και αντίστροφα. Περιγράφει ένα σήμα σε όρους θορύβου και άλλων βασικών χαρακτηριστικών.

Β΄. Μετασχηματισμοί σήματος

Στο βήμα (2), τα χαρακτηριστικά αρχικά δείχνουν κάπως έτσι:

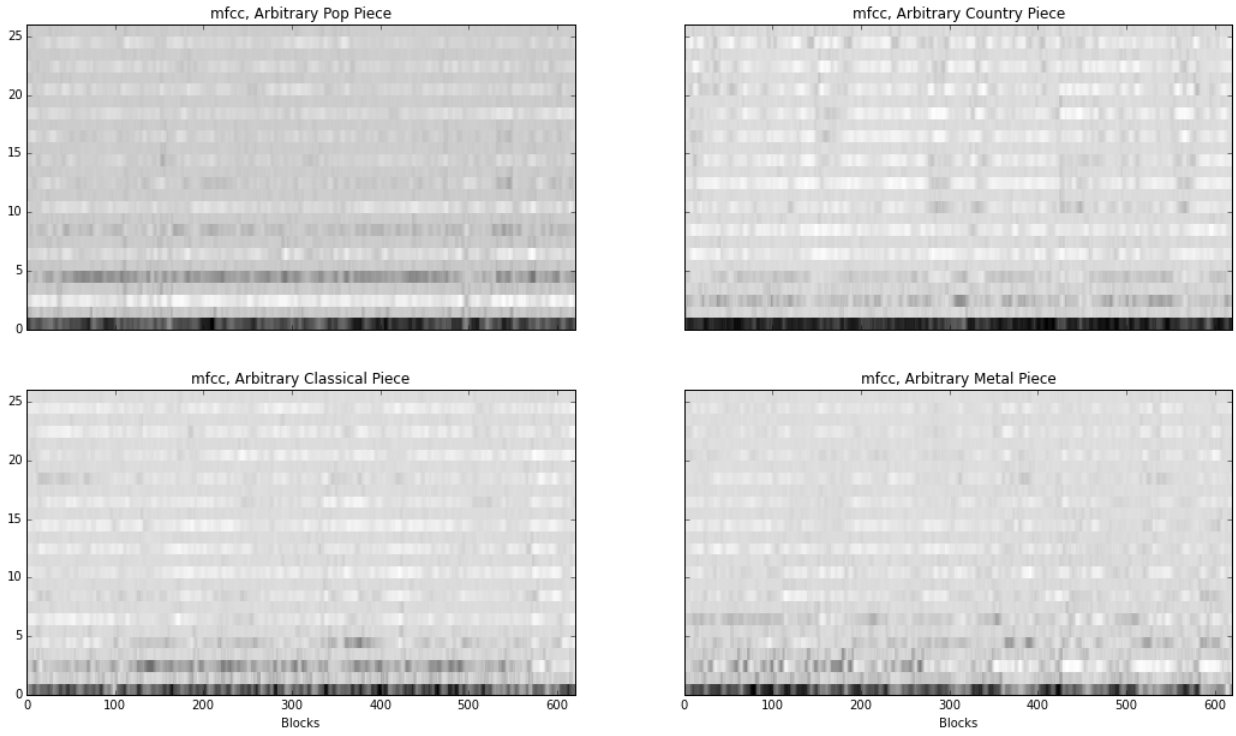
Σχήμα 8: *Αναπαράσταση του χαρακτηριστικού MFCC*



Στον άξονα x έχουμε τα πλαίσια τα οποία έχουμε χωρίσει το σήμα, και στον άξονα y βλέπουμε να εναλλάσσεται ο μέσος όρος και η τυπική απόκλιση για τις διαστάσεις του κάθε χαρακτηριστικού.

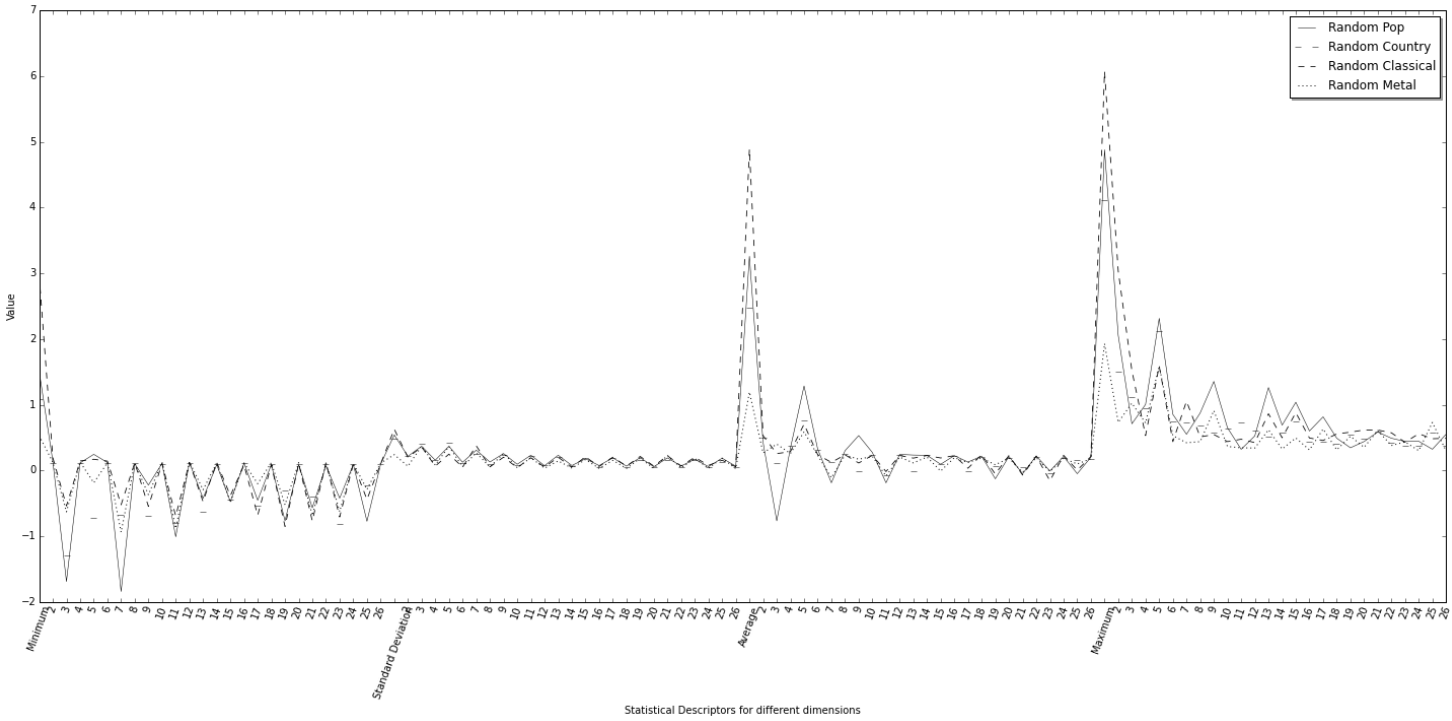
Στην παρακάτω εικόνα έχουμε μερικά δείγματα από διαφορετικά είδη:

Σχήμα 9: *Διάφορα είδη για το χαρακτηριστικό MFCC*



Παρατηρείται πως το πλήθος των δεδομένων για ένα κομμάτι είναι 600×26, δηλαδή πάρα πολλά. Γι΄ αυτό μετά το βήμα 3 έχουμε το εξής σχήμα:

Σχήμα 10: *Διάφορα είδη για το χαρακτηριστικό MFCC - Στατιστικά απομειωμένα*



Τώρα πια τα δεδομένα για ένα κομμάτι είναι 4×26

Γ΄. Επιλογή Συνιστωσών

Η Επιλογή των συνιστωσών για το βήμα (4) της μεθόδου έγινε ανάλογα με τις διαστάσεις κάθε χαρακτηριστικού σύμφωνα με τον παρακάτω πίνακα:

Πίνακας 7: *Επιλογή Συνιστωσών*

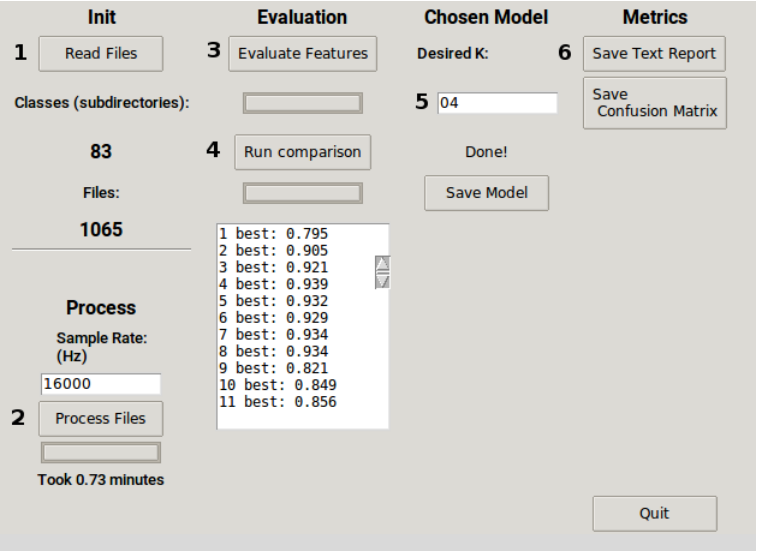
Διαστάσεις	Συνιστώσες PCA	Συνιστώσες LDA
N>30	13	12
30>N>15	7	6
N<15	3	2

Οι παραπάνω τιμές προέκυψαν μετά από δοκιμές με κριτήριο την ακρίβεια της ταξινόμησης και με προσοχή για να μην υπάρξει υπερπροσαρμογή (overfitting) του ταξινομητή.

Δ΄. Γραφικό Περιβάλλον

Βλέπουμε το κεντρικό περιβάλλον του ταξινομητή, με τα βήματα όπως περιγράφονται στις μεθόδους.

Σχήμα 11: *Κεντρικό περιβάλλον ταξινόμησης*



Η Τυπική χρήση του προγράμματος συνίσταται στα εξής βήματα:

1

Αφού έχουμε τα αρχεία χωρισμένα σε φακέλους ανάλογα με τις τάξεις, πατώντας το κουμπί 'Read Files' το πρόγραμμα μας ενημερώνει για το πλήθος των αρχείων και τάξεων.

2

Εισάγουμε τον ρυθμό δειγματοληψίας και πατάμε το κουμπί 'Process Files'. Το πρόγραμμα εξάγει χαρακτηριστικά από τα αρχεία ήχου, μας ενημερώνει μέσω της αντίστοιχης μπάρας για την πρόοδο της διαδικασίας και στο τέλος αναγράφεται ο χρόνος που παρήλθε.

3

Πατώντας το κουμπί 'Evaluate Features' το πρόγραμμα αξιολογεί το κάθε χαρακτηριστικό. Η πρόοδος φαίνεται στην αντίστοιχη μπάρα.

4

Πατώντας το κουμπί 'Run comparison' το πρόγραμμα δημιουργεί αύξοντες συνδυασμούς των καλύτερων χαρακτηριστικών, όπως προέκυψαν στο προηγούμενο βήμα, και μας πληροφορεί στον αντίστοιχο πίνακα για την ακρίβειά τους.

5

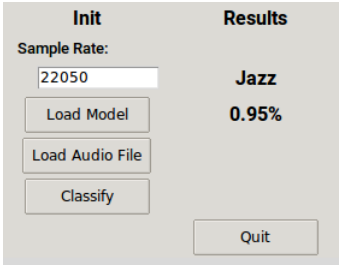
Σε αυτό το σημείο αφού συμβουλευτούμε τον πίνακα που προέκυψε στο προηγούμενο βήμα, εισάγουμε το πλήθος των χαρακτηριστικών που επιθυμούμε να κρατήσουμε για το μοντέλο, και στη συνέχεια αποθηκεύουμε το μοντέλο στον δίσκο.

6

Εδώ δίνεται η δυνατότητα αποθήκευσης της έκθεσης ταξινόμησης, η οποία αναφέρει την ακρίβεια καθώς και άλλους δείκτες για την απόδοση του ταξινομητή, καθώς και ποιά επιμέρους χαρακτηριστικά επιλέχθηκαν. Επίσης μπορεί να αποθηκευθεί και ο πίνακας σύγχησης, σε μορφή PNG.

Στην συνέχεια, ανοίγουμε ένα δεύτερο πρόγραμμα, τον Δοκιμαστή:

Σχήμα 12: *Δοκιμαστής*



Ο Δοκιμαστής, προβλέπει την τάξη για ένα νέο αρχείο ήχου. Χρειάζεται να εισάγουμε τον ρυθμό δειγματοληψίας του αρχείου, να φορτώσουμε ένα μοντέλο που προέκυψε απο το βήμα 5 του ταξινομητή, και το αρχείο που θέλουμε να ταξινομήσουμε, και πατώντας το κουμπί 'Classify' μας εμφανίζεται η προβλεπόμενη τάξη, καθώς και η σιγουριά του ταξινομητή ότι το αρχείο ανήκει όντως στην τάξη αυτή.