

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Αναγνώριση Προτύπων

2016-2017

Γιώργος Δασούλας
03112010
Νικολάος Ζαρίφης
03112178

Προπαρασκευαστική 3ης εργαστηριακής άσκησης
Προθεσμία : 5/12/2016

Σκοπός

Στη συγκεκριμένη άσκηση υλοποιείται ένα σύστημα ταξινόμησης μουσικών αποσπασμάτων με βάση το συναίσθημα (***Music Emotion Classification***) . Στην συγκεκριμένη προπαρασκευαστική χρησιμοποιούμε εργαλεία όπως το **MIR Toolbox** του MATLAB , με τα οποία εξάγουμε χαρακτηριστικά από το ακουστικό σήμα . Επίσης , διαθέτοντας και 3 επισημειωτές , οι οποίοι παρέχουν τις διαστάσεις **valence** και **activation** , αναλύουμε το πόσο συμφωνούν αυτοί οι επισημειωτές , έτσι ώστε να λάβουμε ένα δείγμα ποιότητας των δειγμάτων .

Τεχνικά Θέματα

Για την υλοποίηση του ταξινομητή χρησιμοποιούμε την κλάση **Sound**, της οποίας ο builder αναλαμβάνει , όπως φαίνεται στον κώδικα , όλα τα βήματα της προεπεξεργασίας , της εξαγωγής συντελεστών μέσω των MIR εργαλείων . Επίσης , χρησιμοποιήσαμε την έτοιμη συνάρτηση **kriAlpha** του MATLAB, για τον υπολογισμό του συντελεστή **Krippendorff's alpha**.

Βήμα 1 :

Αρχικά , μέσω της εντολής **audioread** φορτώσαμε τα 412 αρχεία ήχου στο **workspace**.

Σε αυτό το βήμα κάνουμε προεπεξεργασία των δεδομένων/δειγμάτων . Συγκεκριμένα , μέσω της συνάρτησης **preprocess** μετατρέπουμε τα δείγματα (samples) από **stereo** σε **mono** , παίρνοντας τον μέσο όρο των 2 τιμών σε κάθε δείγμα . Επίσης, μέσω συναρτήσεων αλλάζουμε το ρυθμό δειγματοληψίας .

Σημείωση :

- 1) Επειδή στα παρακάτω ερωτήματα χρησιμοποιούμε το MIRTtoolbox , προτιμήσαμε να χρησιμοποιήσουμε κατευθείαν την εντολή **miraudio** , αντί της **audioread** μέσω της οποίας ορίσαμε και απευθείας τις κατάλληλες παραμέτρους δειγματοληψίας και επιλογές stereo/mono . Έτσι , κρατάμε τα δείγματα σε μορφή κατάλληλη για τη διαδικασία εξαγωγής χαρακτηριστικών .
- 2) Λόγω του πολύ μεγάλου μεγέθους δεδομένων , δεν αποθηκεύουμε τελικά τα δείγματα εξ'ολοκλήρου , καθώς με το που υπολογίζουμε τους συντελεστές μας αρκούν αυτοί για την εξαγωγή χαρακτηριστικών . Αυτό το κάνουμε , καθώς αντιμετωπίσαμε μεγάλη φόρτωση της RAM .

Βήμα 2 :

Στο βήμα αυτό αξιοποιήσαμε τις εξόδους που μας παρέχουν οι 3 επισημειωτές των δειγμάτων για τις δύο διαστάσεις του χώρου συναισθήματος **valence** και **activation** . Οι επισημειωτές για την κάθε διάσταση δίνουν μια τιμή που κυμαίνεται στο διάστημα $\{1,...,5\}$. Οι δύο αυτές διαστάσεις μας βοηθάνε στην ταξινόμηση του κάθε αποσπάσματος για το συναίσθημα που εκφράζει . Συγκεκριμένα , αυτή η δισδιάστατη αναπαράσταση του συναισθήματος στους άξονες της Χαράς/Λύπης (Valence) και Ενεργοποίησης/Απενεργοποίησης (Activation) .

Αρχικά , υπολογίσαμε τις μετρήσεις πρώτου βαθμού και συγκεκριμένα τη μέση τιμή και την τυπική απόκλιση για κάθε annotator και για την κάθε διάσταση .

Συγκεκριμένα , προέκυψαν τα παρακάτω αποτελέσματα :

Valence :

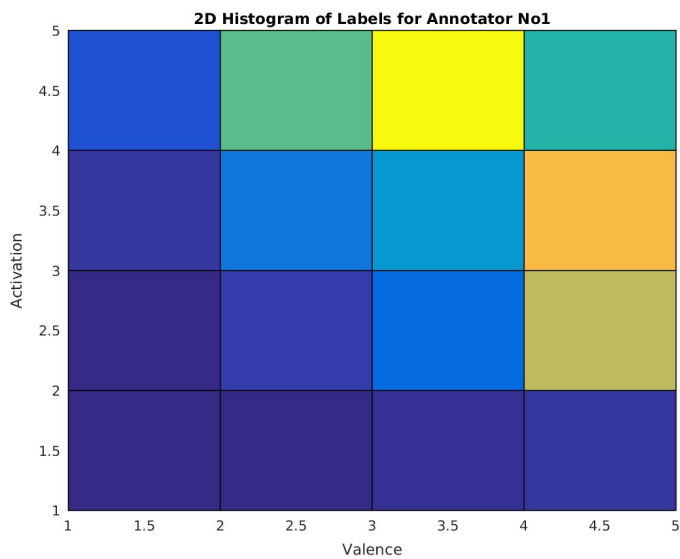
	1	2	3
1	3.3447	3.0316	3.2524
2	0.8921	0.8721	1.0530

Activation :

	1	2	3
1	3.2015	3.2646	2.6529
2	0.9152	0.9283	1.0173

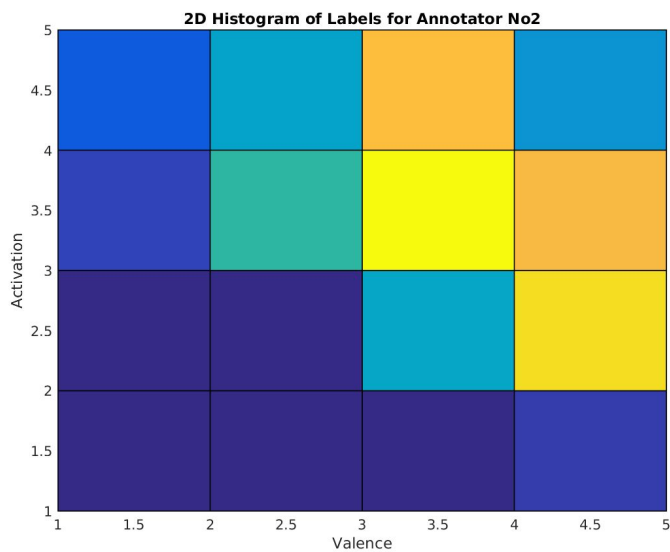
όπου η πρώτη γραμμή αφορά τη μέση τιμή και η δεύτερη την τυπική απόκλιση ενώ οι στήλες τον κάθε επιφωνητή. Από αυτά παρατηρούμε πως στην διάσταση του **Activation** έχουμε μεγαλύτερη διασπορά από ότι στη διάσταση του **Valence** . Επίσης , παρατηρούμε πως οι επιφωνητές έχουν μέση τιμή που κεντράρεται περίπου στο 3 .

Επίσης , παρακάτω ακολουθούν τα 2D-ιστογράμματα **Valence Vs Activation Vs Number of Occurrences** :

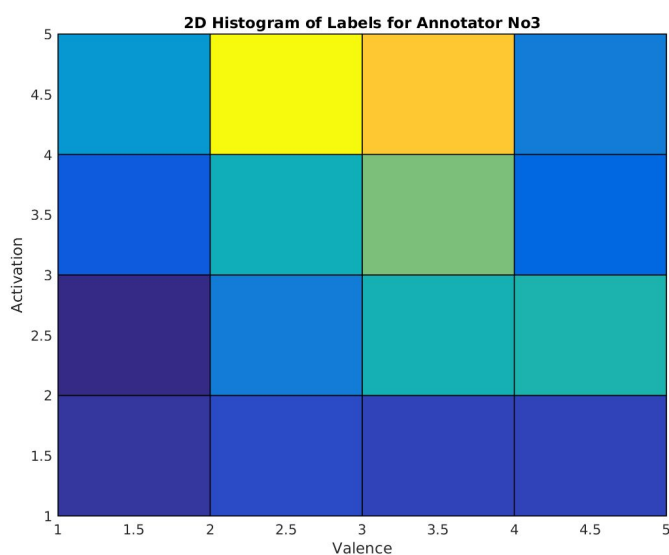


Annotator 1 :

	1	2	3	4	5
1	0	0	2	3	0
2	0	4	13	59	1
3	3	17	27	68	11
4	9	46	83	40	1
5	1	20	4	0	0



	1	2	3	4	5
1	0	0	0	4	6
2	0	0	26	64	17
3	5	34	69	57	3
4	8	25	58	21	2
5	1	8	2	2	0

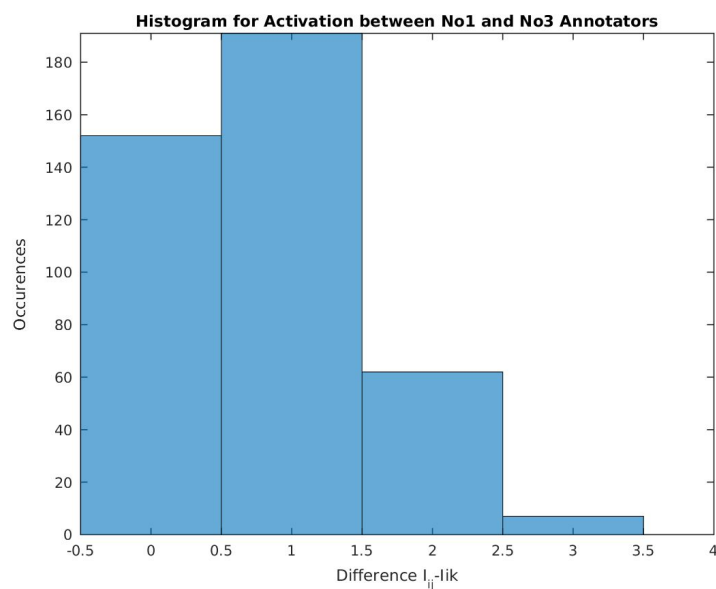
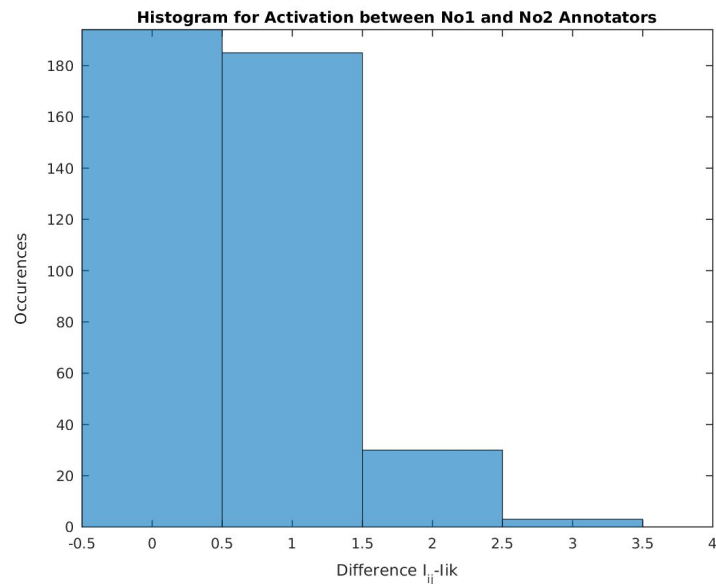


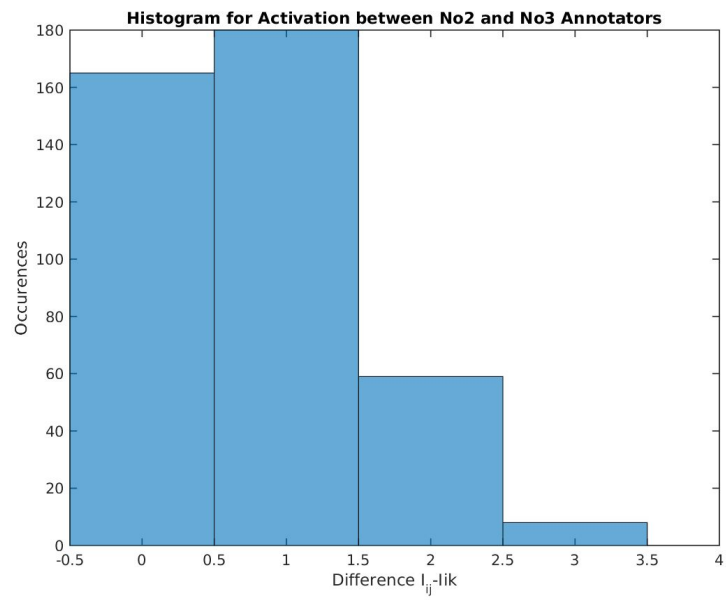
	1	2	3	4	5
1	3	6	5	5	4
2	0	16	32	33	7
3	8	30	43	11	3
4	23	71	62	16	2
5	19	11	2	0	0

Βήμα 3 :

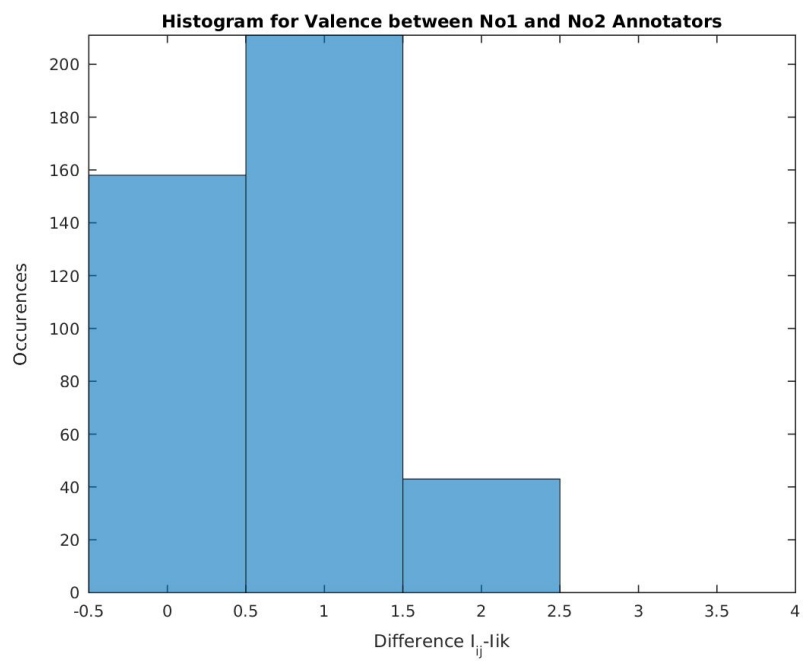
Σε αυτό το βήμα , βλέπουμε την συμφωνία που έχουν οι τρεις επισημειωτές , το οποίο καθορίζει την ποιότητα των δειγμάτων που έχουμε πάρει . Συγκεκριμένα , υπολογίσαμε το Observed Agreement , σύμφωνα με τον τύπο που δίνεται . Επίσης , παρακάτω δίνονται τα ιστογράμματα των διαφορών για κάθε ζεύγος επισημειωτών .

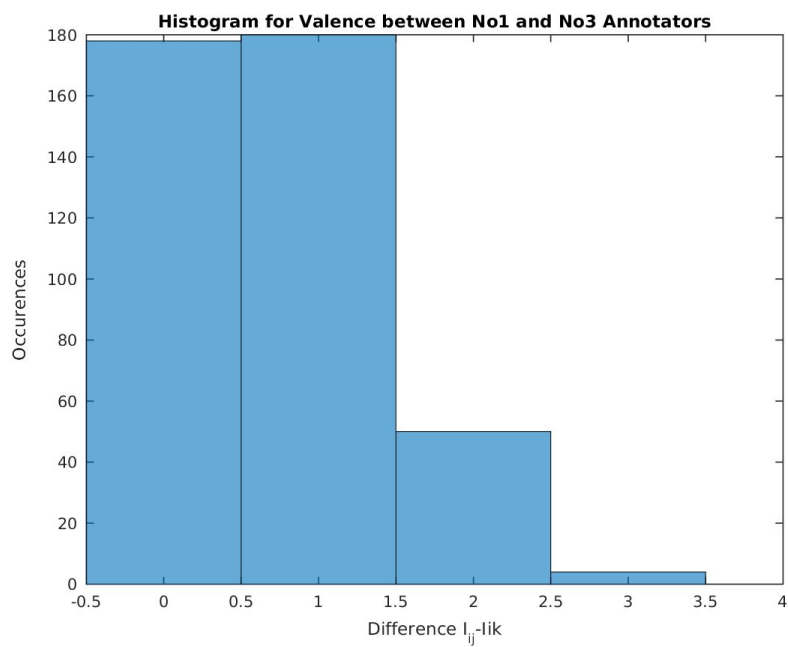
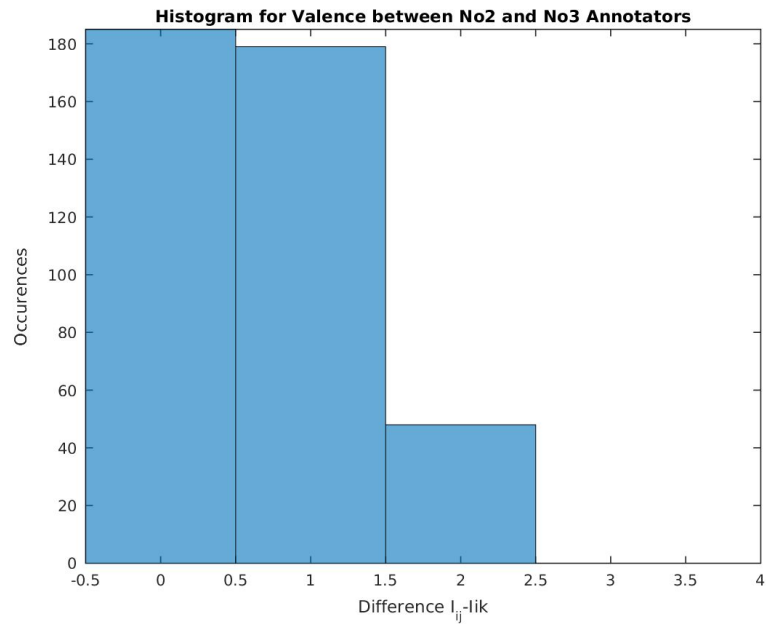
Αρχικά , για **Activation** έχουμε τα εξής :





Αντίστοιχα, για **Valence** έχουμε τα εξής :





Από τα παραπάνω βλέπουμε πως έχουμε γενικά μεγάλη συμφωνία μεταξύ των annotators , καθώς δεν βλέπουμε καθόλου τιμές 4 και πολύ λίγες τιμές 3 . Επίσης , διαφορές τιμής 3 (δηλαδή μεγάλες διαφορές) παρατηρούμε περισσότερο στην περίπτωση του **Activation** . Επίσης , παρατηρούμε πως μάλλον οι επιφωνητές 1 και 2 έχουν τη μεγαλύτερη συμφωνία. Το σημαντικό που παρατηρούμε είναι η καλή ποιότητα των δειγμάτων , το οποίο φαίνεται από την υψηλή συμφωνία των επιφωνητών .

Βήμα 4 :

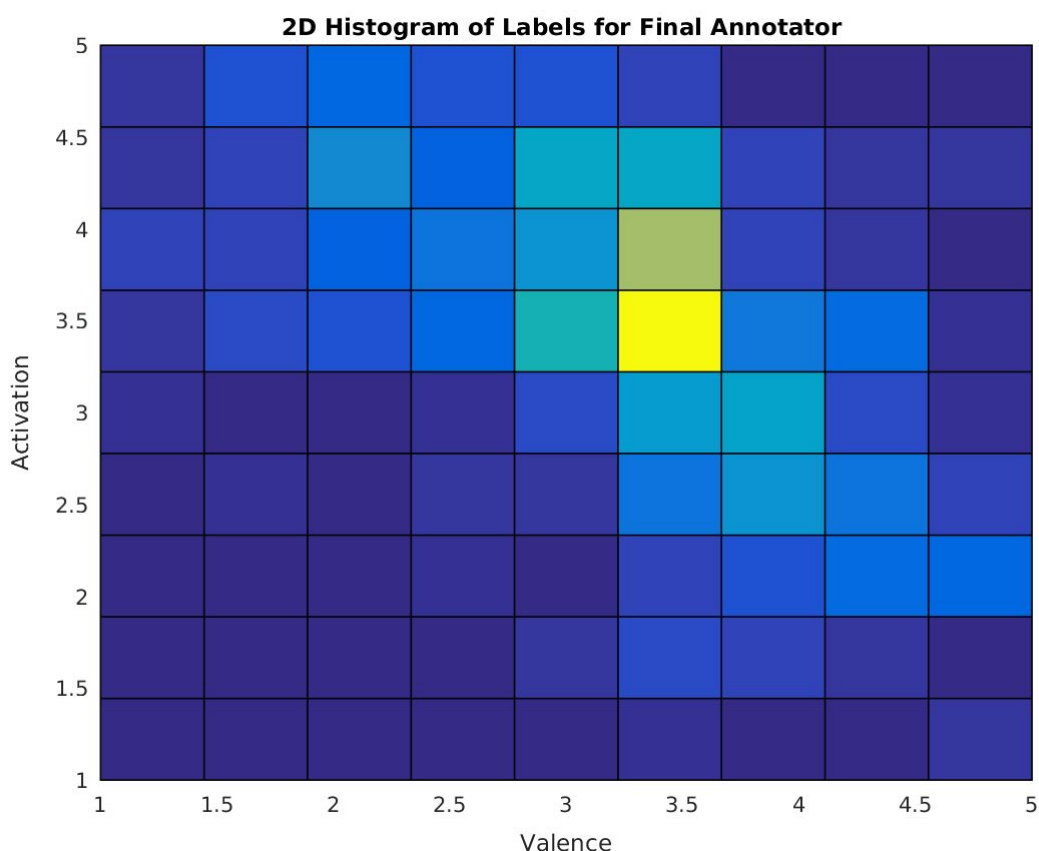
Στο συγκεκριμένο βήμα , υπολογίζουμε τον συντελεστή **Krippendorff's alpha** για ordinal δεδομένα σε κάθε διάσταση . Συγκεκριμένα τα αποτελέσματα είναι :

alpha_act =	alpha_val =
0.4398	0.4747

Το συντελεστή αυτό το χρειαζόμαστε , καθώς λαμβάνει υπόψιν και την αναμενόμενη συμφωνία . Αυτό είναι αναγκαίο , καθώς μπορεί δύο επιφωνητές να συμφωνούν τυχαία και όχι επειδή έχουμε αξιόπιστο label . Όπως βλέπουμε , και πάλι έχουμε ισχυρή συμφωνία μεταξύ των επιφωνητών και , όπως είδαμε και στα διαγράμματα έχουμε μεγαλύτερη συμφωνία στο Valence από ότι στο Activation .

Βήμα 5 :

Για τις τελικές επισημειώσεις διαλέγουμε το μέσο όρο των 3 επισημειώσεων για κάθε δείγμα και κάθε διάσταση . Προκύπτουν έτσι το παρακάτω 2D-ιστόγραμμα :



Βήμα 6-7

Χρησιμοποιώντας το mirtoolbox εξάγαμε τα χαρακτηριστικά τα οποία θα χρησιμοποιήσουμε για classification. Πέρα από τα MFCCs που ασχοληθήκαμε στην προηγούμενη άσκηση, εδώ έχουμε:

Roughness: Roughness ορίζεται σύμφωνα με την φυσιολογία ως η απόσταση από τις κρίσιμες συχνότητες (το φάσμα συχνοτήτων ανάμεσα στο φιλτράρισμα που γίνεται στο κοχλία (σημείο μέσα στο αυτί)). Πρακτικά αυτό που μας δείχνει είναι κατά ποσό είναι αρμονικό το κομμάτι δηλαδή πόσο ταιριάζουν οι νότες, πχ οι νότες που ανήκουν στην ίδια κλίμακα ή είναι πολύ κοντά τονικά σε αντίθεση με νότες που είναι σε διαφορετική κλίμακα πχ ντο, ντο+8. Παίρνουμε μέση τιμή, τυπική απόκλιση, και μέση τιμή κι απόκλιση για μεγαλύτερο κι μικρότερο του median στοιχείου.

Fluctuation: Το οποίο είναι μια τεχνική για ανάλυση του ρυθμού, χρησιμοποιώντας την ενέργεια του φάσματος σε κάθε band. Το αποτέλεσμα είναι η κατανομή της ρυθμικής περιοδικότητας. Παίρνουμε την μεγαλύτερη τιμή κι την μέση τιμή ως χαρακτηριστικά

Key Clarity: Εδώ πρακτικά ψάχνουμε να βρούμε ποια είναι η πιο πιθανή κλίμακα που παίζεται, πρακτικά προσπαθεί να βρει το range όπου είναι οι 7 νότες (8 αν υποθέσουμε ότι η πρώτη είναι κι στο τέλος ως άνω φράγμα), από πίσω ψάχνει να βρει τα local -maximum στο spectrum (peaks) κι σύμφωνα με αυτά κάνει συσχέτιση με διαφορές κλίμακες. Ως χαρακτηριστικό η μέση τιμή σε όλα τα frames.

Modality: Γενικά οι κλίμακες χωρίζονται σε major κι σε minor. Θεωρείτε ότι οι major συγχορδίες/κλίμακες είναι πιο χαρούμενες σε αντίθεση με την minor. Πέρα από τα major/minor οποιοσδήποτε συνδυασμός των 7(ή 8) νότων φτιάχνουν κάτι που μοιάζει με κλίμακα κι αποκαλείται modes. Γενικά γράφοντας μουσική σε διαφορετικά modes αλλάζει τον τρόπο που αισθάνεσαι την μελωδία κι ας παίζεις πάνω στις ίδιες νότες. Εμείς εδώ χρησιμοποιούμε το αποτέλεσμα mirmode που δείχνει πόσο κοντά σε major ή minor είναι το δείγμα μας. Ως χαρακτηριστικό έχουμε την μέση τιμή σε όλα τα frames.

Spectral Novelty: Εδώ βρίσκουμε τα σημεία που υπάρχουν σημαντικές αλλαγές στο φάσμα. Πάλι εδώ παίρνουμε μέση τιμή σε όλα τα frames.

Harmonic Change Detection Function: Ψάχνει να βρει την κυρία ροή της κεντρικής συχνότητας, δηλαδή με πιο άπλα λόγια μας δείχνει τις αλλαγές που γίνονται στις συγχορδίες. Παίρνουμε την μέση τιμή σε όλα τα frames.

MFCCs Deltas: Εδώ σε σχέση με την προηγούμενη άσκηση, έχουμε επίσης κι της “παραγώγους” των MFCC, που δείχνουν το πως μεταβάλλονται οι συντελεστές ανάμεσα στα frames. Ο τρόπος που υπολογίζονται είναι με την μέθοδο linear regression.

Τα Deltas2 είναι σαν δεύτερες παράγωγοι

Μαζί με τα MFCCs, Delta 1, Delta 2 έχουμε 39 χαρακτηριστικά για κάθε frame, από αυτά θα κρατήσουμε για το τελικό μας classify για κάθε ένα από τα 39, την μέση τιμή και τυπική απόκλιση ανάμεσα σε όλα τα frames και την μέση τιμή κι τυπική απόκλιση των 10% μεγαλύτερων

Τα αποτελέσματα του παραπάνω ερωτήματος βρίσκονται σε ένα αρχείο .mat που επισυνάπτουμε τα οποία έχουν δομή Sound.