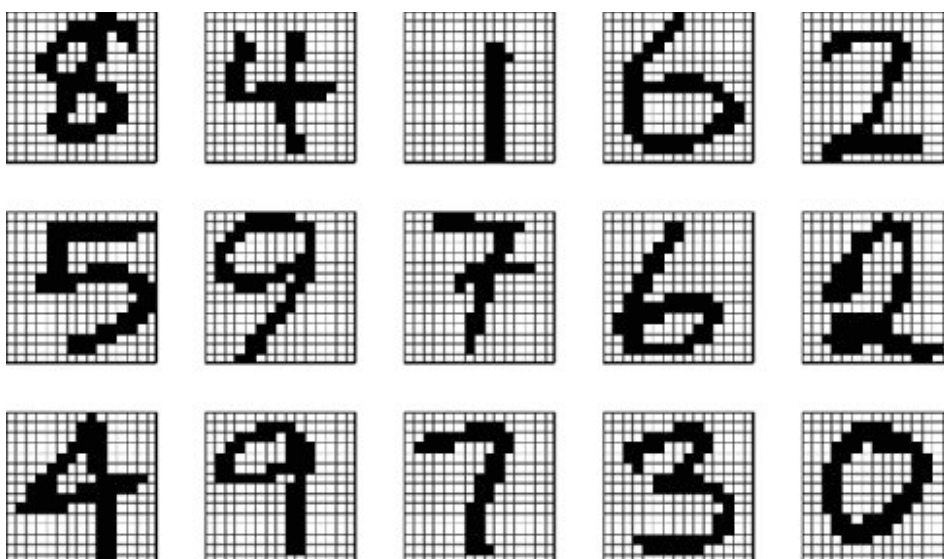


Προπαρασκευή 3ης Εργαστηριακής Άσκησης

Αναγνώριση Συναισθήματος στη Μουσική

Μάθημα : Αναγνώριση Προτύπων



Ροή Σ

Συνεργάτες :

- Βαβουλιώτης Γεώργιος (Α.Μ. : 03112083)
- Σταυρακάκης Δημήτριος (Α.Μ. : 03112017)

Σκοπός: Σκοπός της τρίτης εργαστηριακής άσκησης είναι η αυτόματη ταξινόμηση μουσικών αποσπασμάτων με βάση το συναίσθημα που δημιουργεί στον ακροατή. Στην προπαρασκευή πραγματοποιείται η εξαγωγή χαρακτηριστικών από τα διαθέσιμα δεδομένα, δηλαδή τα μουσικά σήματα. Τα συναισθήματα των μουσικών σημάτων θα περιγραφούν με μια διδιάστατη αναπαράσταση, η οποία έχει στον έναν άξονα το valence (χαρά/λύπη) και στον άλλο το activation (ενεργοποίηση/απενεργοποίηση). Το τελικό πακέτο χαρακτηριστικών θα περιλαμβάνει ένα συνδυασμό χαρακτηριστικών όπως χαρακτηριστικά εμπνευσμένα από τη μουσική και στατιστικά μεγέθη των Mel-Frequency Cepstral Coefficients (MFCCs).

Εκτέλεση Άσκησης

Βήμα 1 : Στο βήμα αυτό καλούμαστε να πάρουμε ένα σύνολο από κομμάτια των Beatles, τα οποία έχουν παρόμοιο συναισθηματικό περιεχόμενο και να εξάγουμε τα ζητούμενα χαρακτηριστικά. Η ζητούμενη τροποποίηση των αποσπασμάτων έγινε με χρήση της συνάρτησης `resample()`, με την βοήθεια της οποίας έγινε αλλαγή στη συχνότητα δειγματοληψίας και επίσης αλλάξαμε το `format` σε `mono` για να έχουμε 8 bits/sample. Μετά από αυτή την επεξεργασία σώζουμε τα επεργασμένα αποσπάσματα σε ένα `mat` file για να μην κάνουμε την ίδια διαδικασία κάθε φορά που τρέχουμε τον κώδικα.

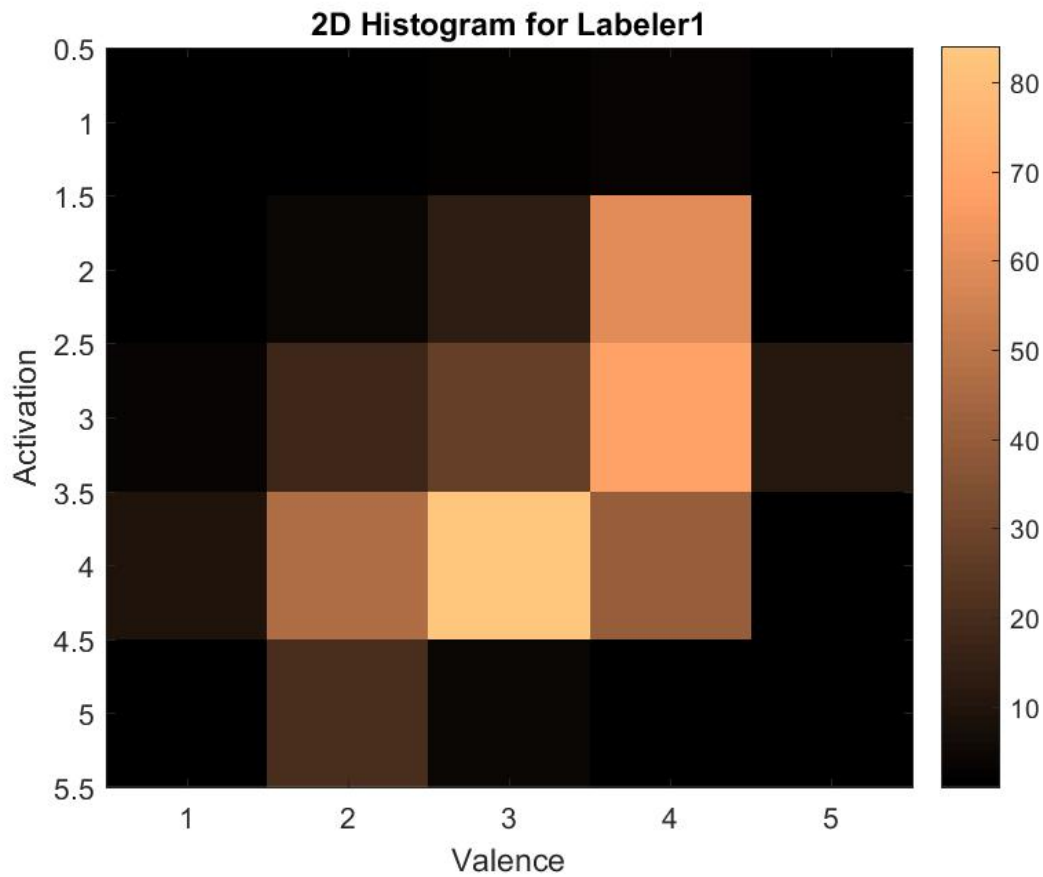
Βήμα 2 : Στο βήμα αυτό θα κάνουμε χρήση του διδιάστατου χώρου valence-activation για τον οποίο κάναμε αναφορά στην αρχή της αναφοράς. Τα valence και activation παίρνουν τιμές από το σύνολο {1,2,3,4,5}. Εμείς αυτό που θέλουμε να κάνουμε είναι να υπολογίσουμε στατιστικά των δεδομένων επισημείωσης. Για τον σκοπό αυτό εισάγουμε στο matlab τα δοσμένα δεδομένα (3 .mat αρχεία) τα οποία περιέχουν τις τιμές των valence και activation για κάθε μουσικό απόσπασμα. Στη συνέχεια σας παρουσιάζουμε την μέση τιμή και η τυπική απόκλιση ανά επισημειωτή για τα valence και activation:

Valence			
	Labeler 1	Labeler 2	Labeler 3
Μέση Τιμή	3.34466019417476	3.03155339805825	3.25242718446602
Τυπική Απόκλιση	0.892055027089517	0.872100417532829	1.05303127827738

Activation			
	Labeler 1	Labeler 2	Labeler 3
Μέση Τιμή	3.20145631067961	3.26456310679612	2.65291262135922
Τυπική Απόκλιση	0.915238825430635	0.928283114690778	1.01732738076082

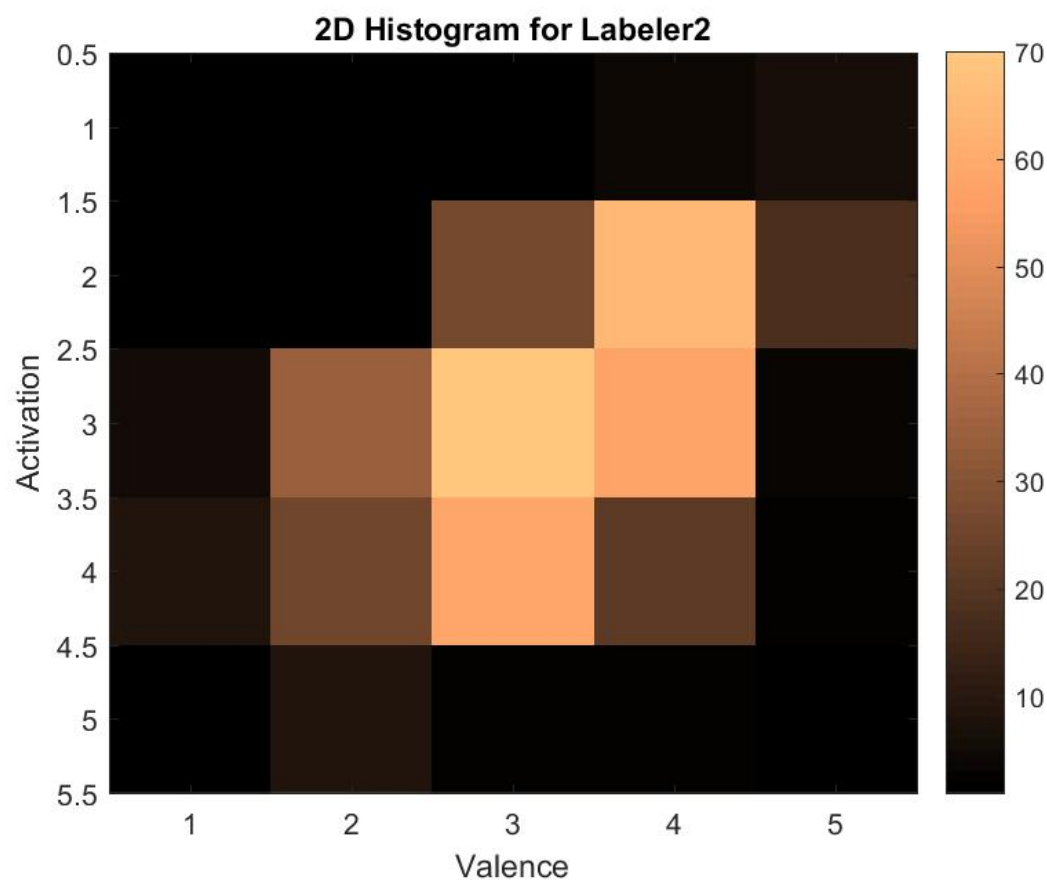
Έπειτα για κάθε επισημειωτή δίνεται ο co-occurrence πίνακας τόσο με τη μορφή διδιάστατου πίνακα όσο και με τη μορφή ιστογράμματος:

Labeler 1 :



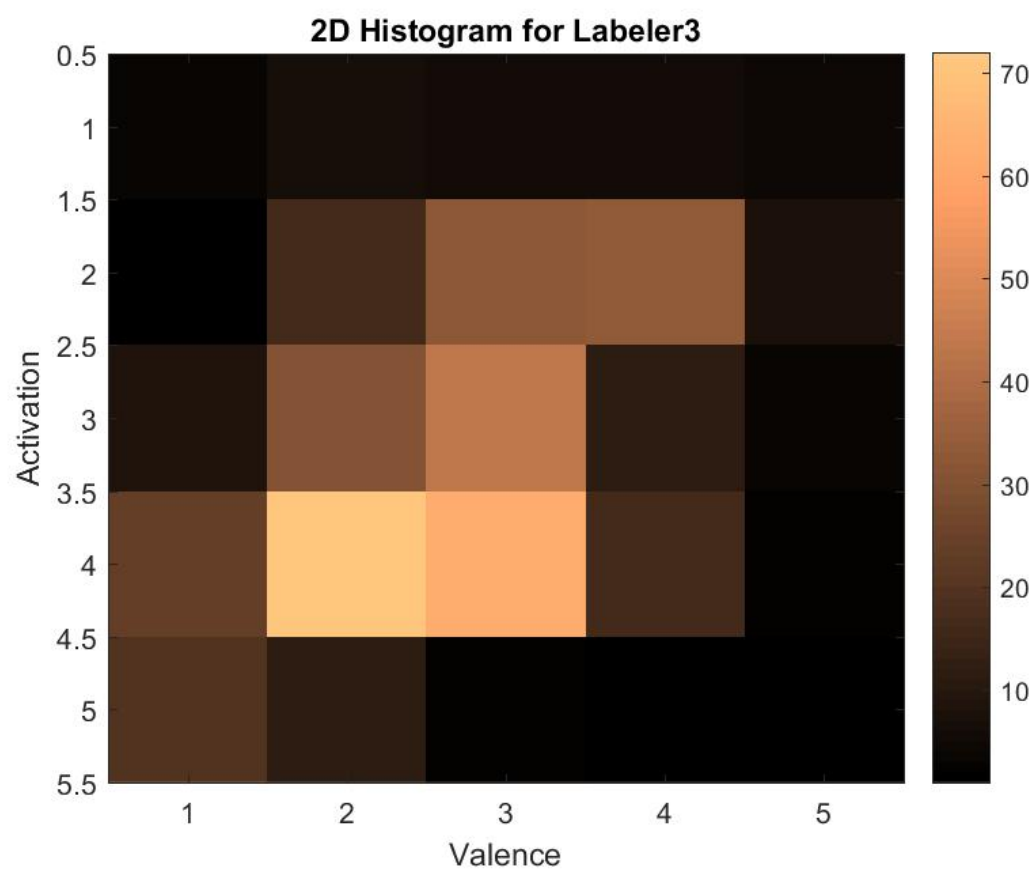
Valence					
Ac tiv ati on	0	0	2	3	0
	0	4	13	59	1
	3	17	27	68	11
	9	46	83	40	1
	1	20	4	0	0

Labeler 2 :



Valence					
Ac tiv ati on	0	0	0	4	6
	0	0	26	64	17
	5	34	69	57	3
	8	25	58	21	2
	1	8	2	2	0

Labeler 3 :



Valence					
Ac tiv ati on	3	6	5	5	4
	0	16	32	33	7
	8	30	43	11	3
	23	71	62	16	2
	19	11	2	0	0

Σχολιασμός Αποτελεσμάτων :

- Στον επισημειωτή 3 παρατηρείται μεγαλύτερη διασπορά διότι έχει κατατάξει αρκετά τραγούδια σε κάθε τιμή variance. Το ίδιο ισχύει και για την διασπορά του activation.
- Απο τους παραπάνω πίνακες έχουμε ότι και για τους τρεις επισημειωτές, η μέση τιμή του valence είναι κοντά στο 3, που είναι η διάμεσος των τιμών που έχουμε. Οσον αφορά το activation, η μέση τιμή είναι πάλι σχετικά κοντά στο 3, αλλά ο επισημειωτής 3 βλέπουμε ότι έχει αρκετά μικρότερη μέση τιμή σε σχέση με τους άλλους δύο διότι δεν έχειπολύ έντονα συναισθήματα προς τα συγκεκριμένα αποσπάσματα τραγουδιών.
- Αν θέλουμε να μιλήσουμε πιο συγκεκριμένα για τα παραπάνω αποτελέσματα, μπορούμε να πούμε ότι παρατηρώντας πιο αναλυτικά τους co-occurrence πίνακες μπορούμε επίσης να συμπερανούμε ότι ο labeler1 έχει κατατάξει σημαντική πλειοψηφία στο ζεύγος valence, activation = 3,4. Επίσης, έχει κατατάξει μεγάλο μέρος στα ζεύγη (4,2), (4,3) και (4,4), δηλαδή ένα σημαντικό υποσύνολο των αποσπασμάτων του έχει προκαλέσει θετικά συναισθήματα. Ο labeler2 έχει κατατάξει τα περισσότερα αποσπάσματα στα ζεύγη (4,2), (4,3) και (3,4). Ο labeler3 κατατάσσει τα συγκεκριμένα αποσπάσματα είτε σε ζεύγος με valence 2 και activation 4 είτε στο (3,4), δηλαδή θεωρεί ότι υπάρχει έντονο συναίσθημα στο κάθε κομμάτι.

Βήμα 3 : Η υψηλή συμφωνία μεταξύ των επισημειωτών αποτελεί συνήθως δείκτη πως τα συγκεκριμένα μουσικά αποσπάσματα αποτελούν «καλά παραδείγματα» για την χρήση τους στην εκπαίδευση ενός μοντέλου ταξινόμησης. Για να επιβεβαιώσουμε το παραπάνω σενάριο, υπολογίζουμε τη τιμή της συμφωνίας σύμφωνα με τη σχέση:

$$1 - E \left\{ \frac{|(l_{ij} - l_{ik})|}{4} \right\}$$

όπου l_{ij} είναι η τιμή του επισημειωτή j στο δείγμα i και l_{ik} η τιμή του επισημειωτή k στο δείγμα i . Επίσης υπολογίζουμε και τη συνολική συμφωνία, η οποία αποτελεί το μέσο όρο των ποσοστών από τις δύο διαστάσεις για κάθε ζεύγος επισημειωτών. Τα αποτελέσματα φαίνονται στη συνέχεια:

Labeler1 VS Labeler2

Valence Observed Agreement : 0.81978210
 Activation Observed Agreement : 0.84587410
 Total Mean Agreement : 0.83282810

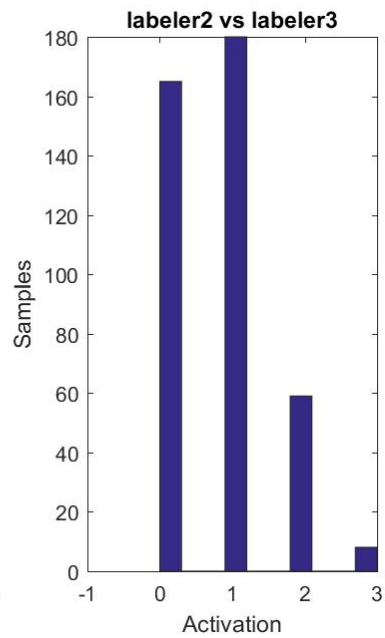
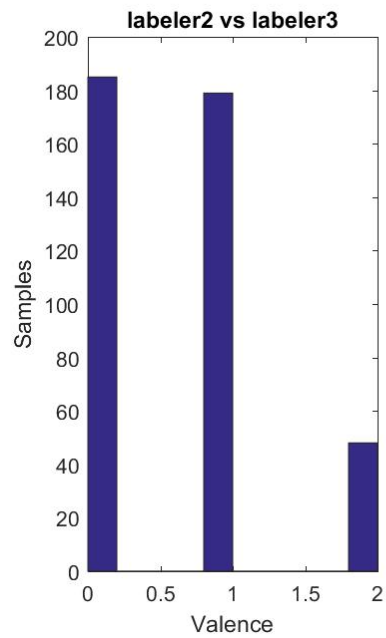
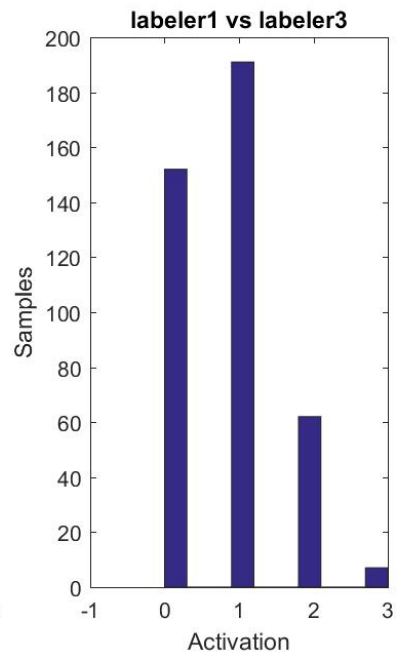
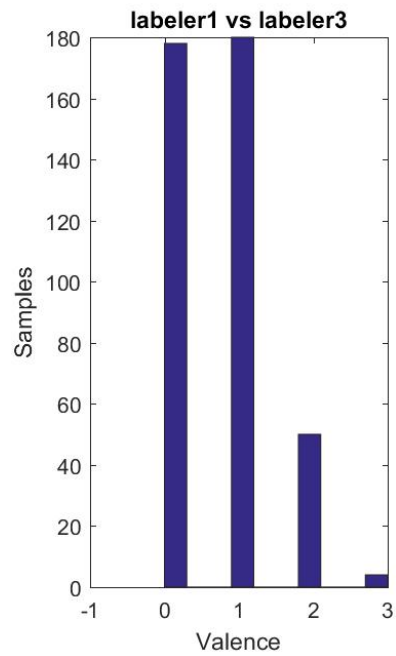
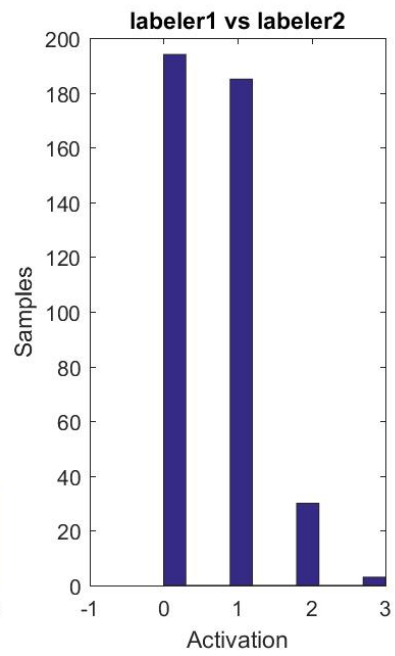
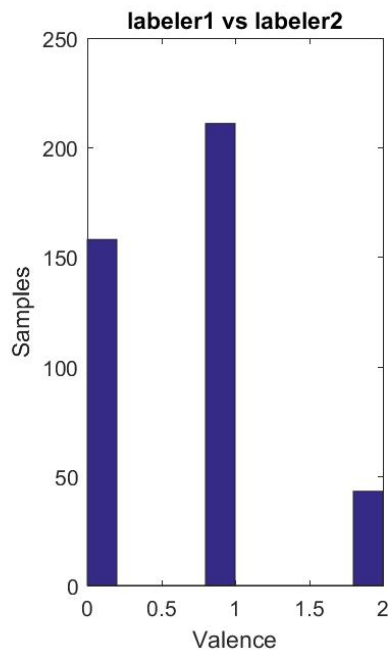
Labeler1 VS Labeler3

Valence Observed Agreement:0.82281610
 Activation Observed Agreement:0.79611710
 Total Mean Agreement:0.80946610

Labeler2 VS Labeler3

Valence Observed Agreement:0.83313110
 Activation Observed Agreement:0.80461210
 Total Mean Agreement:0.81887110

Στη συνέχεια παρατίθενται τα ιστογράμματα διαφοράς ανά ζεύγος παρατηρητών για κάθε μία διάσταση:



Σχολιασμός Αποτελεσμάτων : Παρατηρούμε αρκετά μεγάλα ποσοστά συμφωνίας τόσο σε καθε μία διάσταση ξεχωριστά όσο και συνολικά, άρα το δείγμα μας είναι αρκετά καλό για την κατασκευή του ταξινομητή. Απο τα ιστογράμματα προκύπτει οτι δεν έχουμε σχεδόν καθόλου διαφορές μεγαλύτερες του 2 για το valence. Το συγκεκριμένο δεν ισχύει για το activation διότι σε κάθε ζεύγος παρατηρούμε δείγματα με διαφορές ίσες με 3, καθώς σε κάποιον επισημειωτή είναι πιθανό ενα τραγούδι να βγάζει εντονότερο συναίσθημα.

Βήμα 4 : Στο βήμα αυτό υπολογίζουμε τον συντελεστή Krippendorff's alpha(για ordinal δεδομένα) σε κάθε διάσταση. Το alpha μετράει τη συμφωνία μεταξύ των επισημειωτών και δίνετε ως:

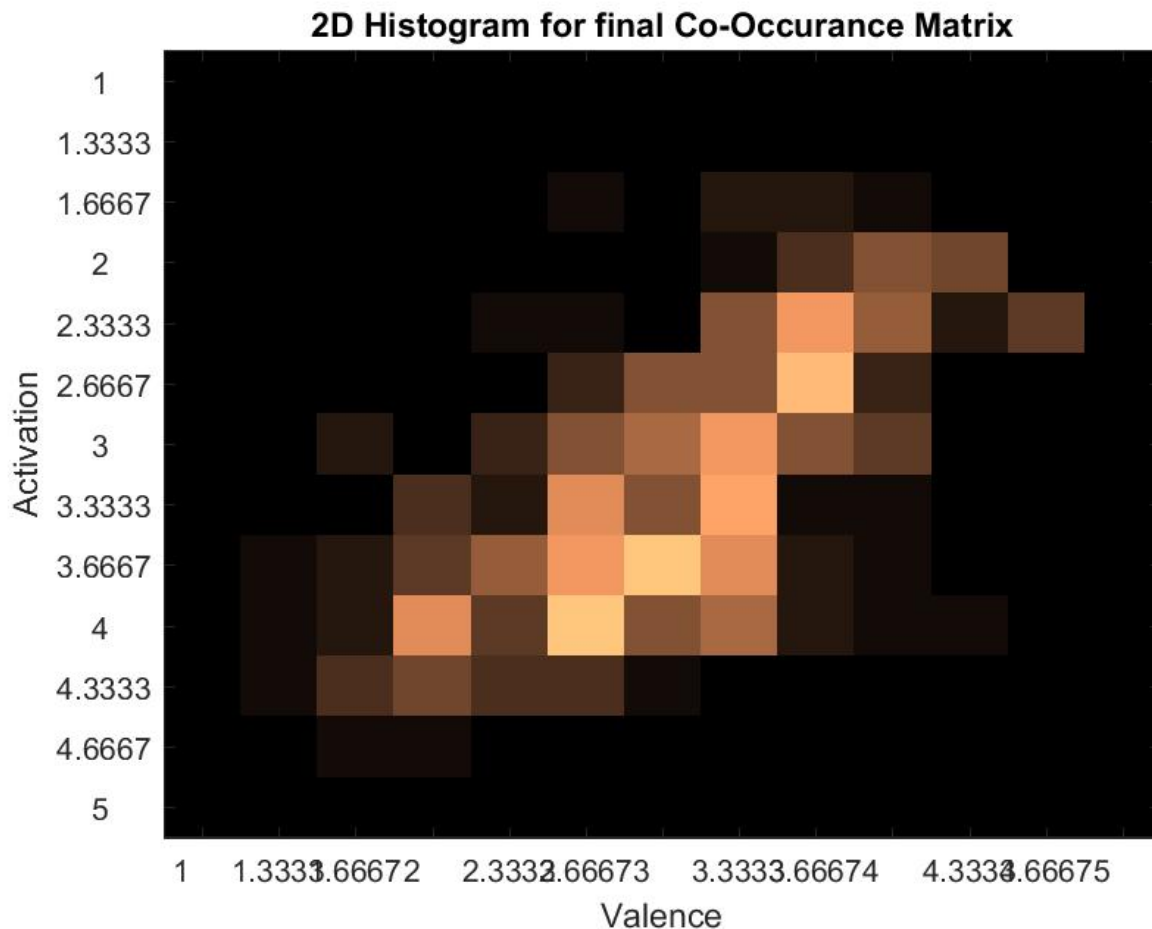
$$\alpha = 1 - \frac{\text{Observed Disagreement}}{\text{Expected Disagreement}}$$

όπου η παρατηρούμενη διαφωνία (Observed Disagreement) είναι η μέση τιμή όλων των σημείων διαφωνίας των δεδομένων (μετρημένη ως απόλυτη διαφορά) και η αναμενόμενη διαφωνία (Expected Disagreement) είναι η μέση τιμή της διαφοράς μεταξύ δύο τιμών i και k σε όλα τα πιθανά ζευγάρια $n(n-1)$ τιμών. Τα αποτελέσματα τα οποία πήραμε είναι τα ακόλουθα :

Krippendorff's alpha coefficient for Valence : 0.4746924
Krippendorff's alpha coefficient for Activation : 0.4398084

Σχολιασμός Αποτελεσμάτων : Προφανώς οι παραπάνω τιμές είναι μικρότερες από τις αντίστοιχες για την παρατηρούμενη συμφωνία διότι κάθε συντελεστής λαμβάνεται υπόψιν και η αναμενόμενη διαφωνία για τον κάθε υπολογισμό.

Βήμα 5 : Στο βήμα αυτό για τον υπολογισμό των τιμών των επισημειώσεων, οι οποίες και θα χρησιμοποιηθούν για την ταξινόμηση των δεδομένων, χρησιμοποιούμε τη μέση τιμή (mean) των τριών επισημειωτών σε κάθε μουσικό απόσπασμα αλλά και για τη κάθε διάσταση ξεχωριστά. Οι τελικές επισημειώσεις θα έχουν τιμές {1, 1.3333, 1.6667, 2, 2.3333, 2.6667, 3, 3.3333, 3.6667, 4, 4.3333, 4.6667, 5}. Κατασκευάζουμε και πάλι τον co-occurrence πίνακα και το αντίστοιχο ιστόγραμμα, τα οποία φαίνονται παρακάτω :



Βήμα 6 : Για την εξαγωγή των χαρακτηριστικών αρχικά χρησιμοποιούμε την βραχέος χρόνου ανάλυση, και στη συνέχεια θα υπολογίζουμε στατιστικά που αφορούν την κατανομή των εν λόγω χαρακτηριστικών στο σύνολο των παραθύρων ανάλυσης. Για την εξαγωγή των χαρακτηριστικών χρησιμοποιούμε τις προεπιλεγμένες από το MIRtoolbox τιμές, δηλαδή διάρκεια πλαισίου ανάλυσης $T=50\text{msec}$ και $\text{Toverlap} = 25\text{msec}$. Από αυτήν την διαδικασία εξάγονται 10 χαρακτηριστικά καθώς τόσο ορίζει η εκφώνηση αλλά θα μπορούσαν να εξαχθούν και άλλα χαρακτηριστικά, μέσω άλλων συναρτήσεων του toolbox, όπως τις *miremotion()* και *mirchromagram()*.

Βήμα 7 : Στο βήμα αυτό γίνεται εξαγωγή των Mel Frequency Cepstral Coefficients (MFCCs) για διάρκεια πλαισίου $T=25\text{msec}$, επικάλυψη $\text{Toverlap}=10\text{msec}$, αριθμός συστοιχίας φίλτρων $Q=26$, αριθμός ζητούμενων χαρακτηριστικών $N_c = 13$. Συγκεκριμένα, λαμβάνονται 39 χαρακτηριστικά, που είναι τα 13 προαναφερθέντα και οι πρώτες και δεύτερες παράγωγοί τους. Από αυτά πραγματοποιούνται στατιστικές μετρήσεις που αφορούν την μέση τιμή, την τυπική απόκλιση και τη μέση τιμή του 10% των μεγαλύτερων και μικρότερων τιμών αντίστοιχα.

Βήμα 8 : Έγινε προετοιμασία των αλγορίθμων ταξινόμησης που υλοποιήσαμε κατά τη διάρκεια της 1^{ης} Εργαστηριακής Άσκησης.

Βήμα 9 : Έγινε εγκατάσταση του Weka και εξοικειωθήκαμε με αυτό.