«ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ»

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Αναγνώριση Προτύπων με Έμφαση στην Αναγνώριση Φωνής

Προπαρασκευή 3ης Εργαστηριακής Άσκησης

9ο Εξάμηνο

Αθανασίου Νικόλαος

03112074

Σκοπός της προπαρασκευής της 3ης εργαστηριακής άσκησης είναι αρχικά η προεπεξεργασία των δεδομένων για την ευκολότερη επεξεργασία τους, η σύγκριση των 3 δοθέντων επισημειωτών ως προς τις επιμέρους διαστάσεις τους, δηλαδή valence και activation, αλλά και συνολικά, καθώς και η εξαγωγή χρήσιμων χαρακτηριστικών όσον αφορά το συναίσθημα για κάθε μουσικό κομμάτι, με τη βοήθεια του MIR Toolbox

Βήμα 1

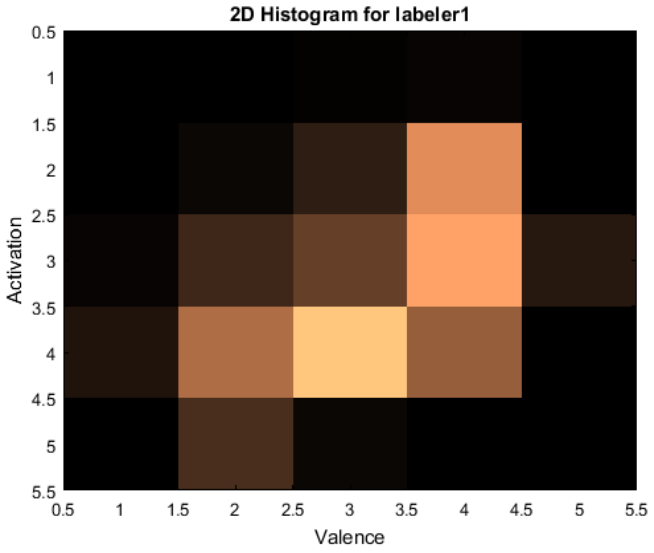
Τα δεδομένα διαβάζονται με τη βοήθεια της *audioread(),* από την έξοδο της οποίας διαπιστώνεται η συχνότητα δειγματοληψίας τους, καθώς και ότι πρόκειται για stereo, αφού ο πίνακας έχει 2 στήλες μια για κάθε κανάλι.Συνεπώς, λαμβάνεται το ημιάθροισμά τους, για να βρούμε τις mono τιμές του ακουστικού σήματος και διαιρώντας με το 22050Hz, βρίσκεται το ratio επαναδειγματοληψίας τους, η οποία υλοποιείται με τη βήθεια της *resample()*. Τέλος, αποθηκεύουμε τα επεξεργασμένα δεδομένα για μελλοντική χρήση χωρίς κάποια επξεργασία.

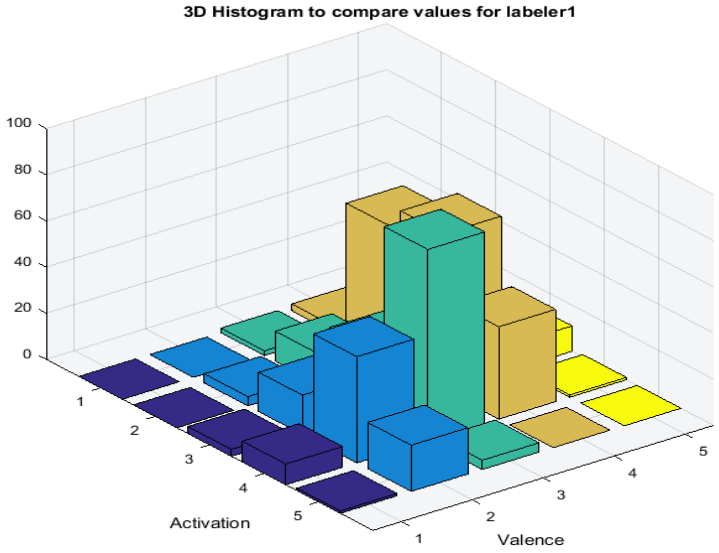
Βήμα 2

Αρχικά, υπολογίζονται οι μέσες τιμές και τυπικές αποκλίσεις για κάθε επισημειωτή και στις δύο διαστάσεις valence (θετικότητα συναισθήματος) και activation (ένταση συναισθήματος). Παρατηρούμε ότι υπάρχουν και θετικά, αλλά και αρνητικά συναισθήματα (valence) είτε μικρής είτε μεγάλης έντασης (activation), χωρίς όμως την κυριαρχία ακραίων συναισθημάτων μεγάλης έντασης (τυπική απόκλιση), όπως ήταν αναμενόμενο από τα τραγούδια των pop/rock Beatles, με τον επισημειωτή 3 να έχει τις μεγαλύτερες διασπορές σε κάθε περίπτωση, γεγονός που υποδεικνύει την ποικιλία του στην κατηγοριοποίηση.

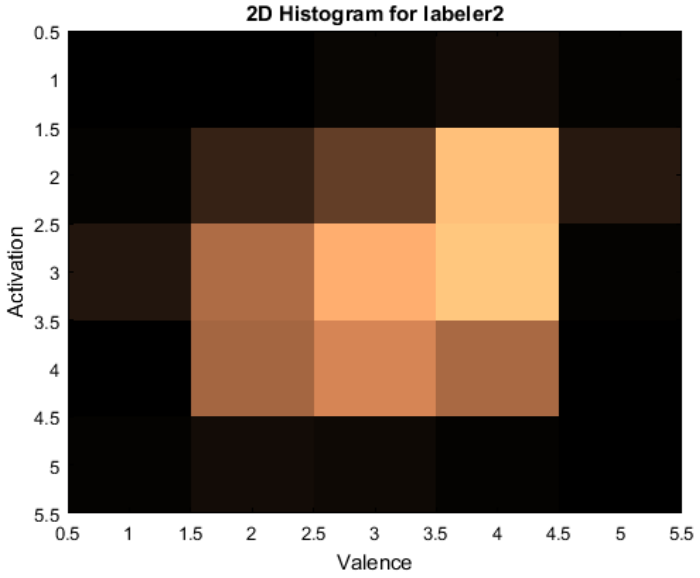
		Επισημειωτής 1	*Επισημειωτής 2*	*Επισημειωτής 3*
Valence	***Μέση τιμή***	3,34466019417476	3,03155339805825	3,25242718446602
	Τυπική Απόκλιση	0,892055027089517	0,872100417532829	1,05303127827738
Activation	***Μέση τιμή***	3,20145631067961	3,26456310679612	2,65291262135922
	Τυπική Απόκλιση	0,915238825430635	0,928283114690778	1,01732738076082

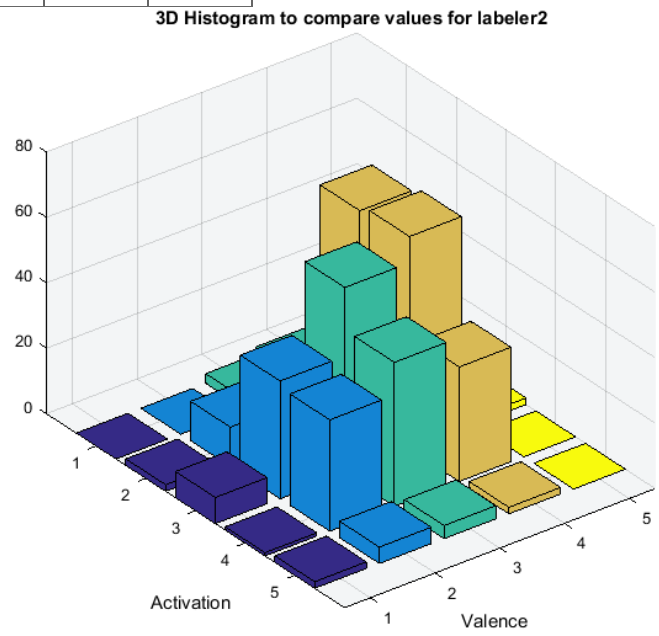
Στη συνέχεια παρουσιάζονται οι co-occurrence πίνακες σε 3 μορφές, ώστε να είναι προφανείς οι μεταξύ τους διαφορές.



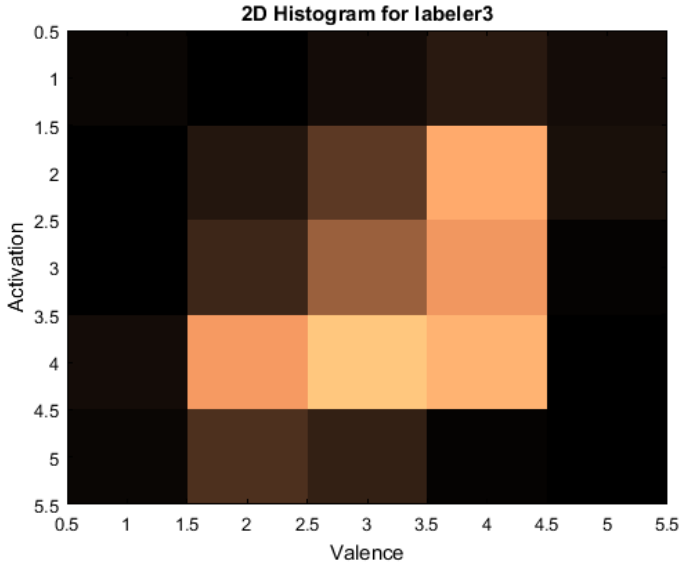


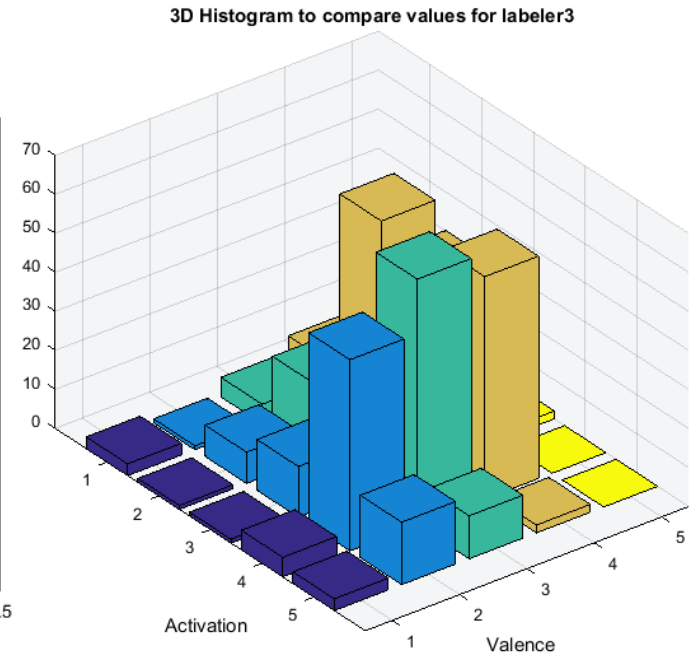
0	0	2	3	0
0	4	13	59	1
3	17	27	68	11
9	46	83	40	1
1	20	4	0	0





0	0	3	5	2
2	12	21	63	9
8	36	57	65	2
1	34	44	35	0
2	5	4	2	0





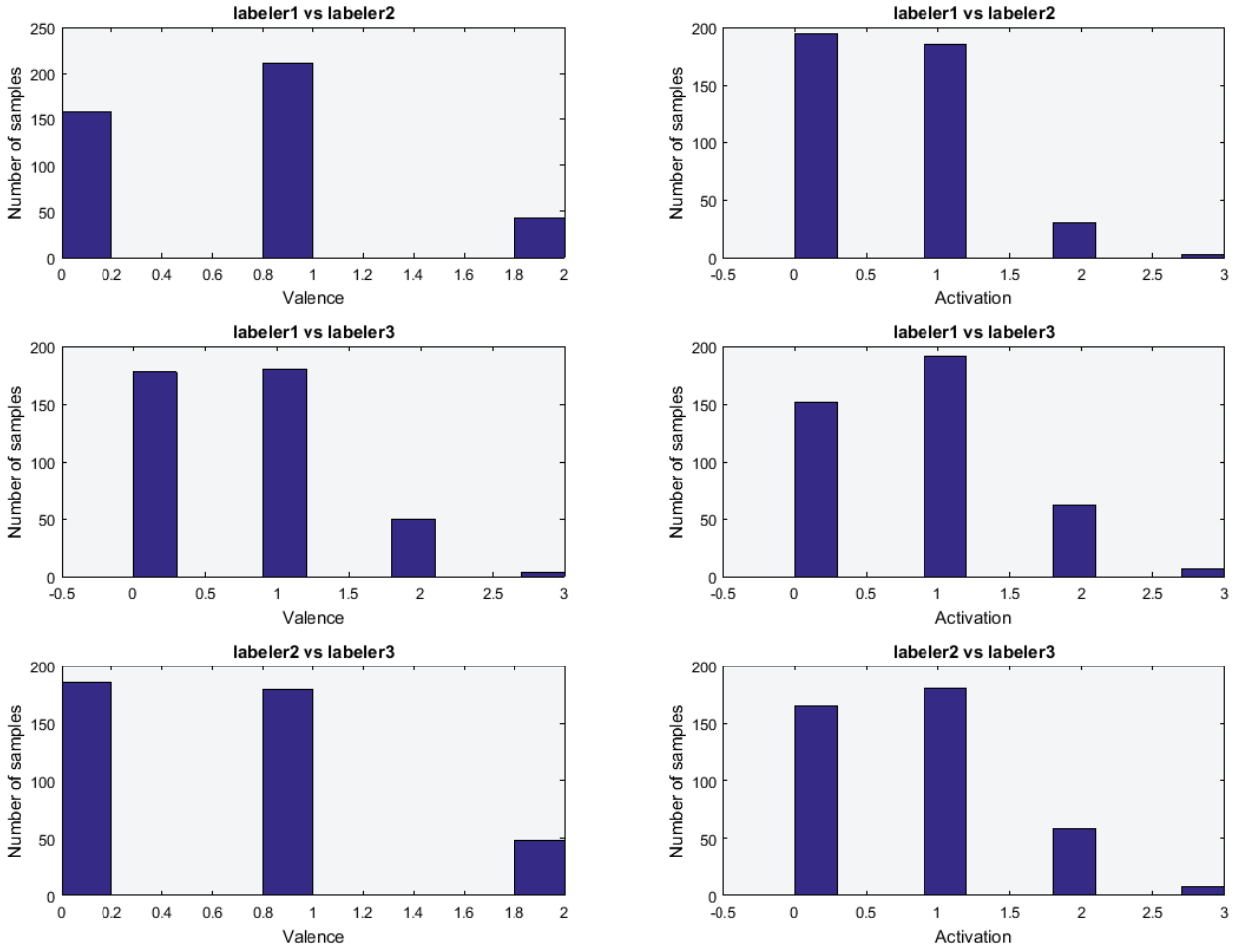
3	**1**	**5**	**9**	**5**
1	**8**	**19**	**54**	**6**
1	**13**	**31**	**48**	**2**
5	**49**	**63**	**57**	**0**
3	**16**	**11**	**2**	**0**

Παρατηρούμε ότι οι επισημειωτές 1 και 2 έχουν παρόμοια χαρακτηριστικά και ο 3 μεγαλύτερη διασπορά και στη διάσταση valence, αλλά και στην activation, πράγμα που σημαίνει ότι κατατάσσει περισσότερα τραγούδια σε κάθε κατηγορία, ακόμα και σε αυτές που οι άλλοι επισημειωτές, σύμφωνα με τον co-occurrence matrix, δεν κατατάσσουν, τόσο όσον αφορά το συναίσθημα όσο και την ένταση. Επιπλέον, η μέση τιμή του activation του επισημειωτή 3 δείχνει να είναι σημαντικά μικρότερη σε σχέση με των άλλων 2, άρα δεν αντιλαμβάνεται την ένταση των συναισθημάτων που αντιλαβάνονται οι άλλοι 2 επισημειωτές. Οι άλλοι 2 επισημειωτές, σαν γενική εικόνα, μπορεί να διαπιστωθεί ότι έχουν περιεχόμενο κυρίως στο μέσο κάθε διάστασης. Ειδικότερα, σύμφωνα με τους co-occurrence matrices, ο επισημειωτής 1 κατατάσσει τα περισσότερα δείγματα στην περιοχή [valence,activation]=[{3,4},{3,4}] –ιδιαίτερα στα 3,4 και 4,3 κατατάσσει τα περισσότερα κομμάτια- άρα τα χαρακτηρίζει είτε μεσαίας έντασης προς το χαρούμενο είτε μεγάλης έντασης προς το ουδέτερο, με κυρίαρχο το πρώτο από τα 2 συναισθήματα. Στην περίπτωση του επισημειωτή 2, παρατηρούμε από κυρίως συναισθήματα χαράς, μέτριας έως και μικρής έντασης κατά κύριο λόγο (κελιά (2,4) (3,4)). Τέλος, αναφορικά με τον τελευταίο επισημειωτή, τη μεγαλύτερη τιμή λαμβάνουμε στην απόλυτη ουδετερότητα συναισθήματος και έντασης. Ωστόσο, εμφανίζεται έντονη παρουσία

δειγμάτων στη τιμή 4 για valence, δηλαδή θετικά συναισθήματα σε κάθε τιμή έντασης, εκτός από τις 2 ακραίες.

Βήμα 3

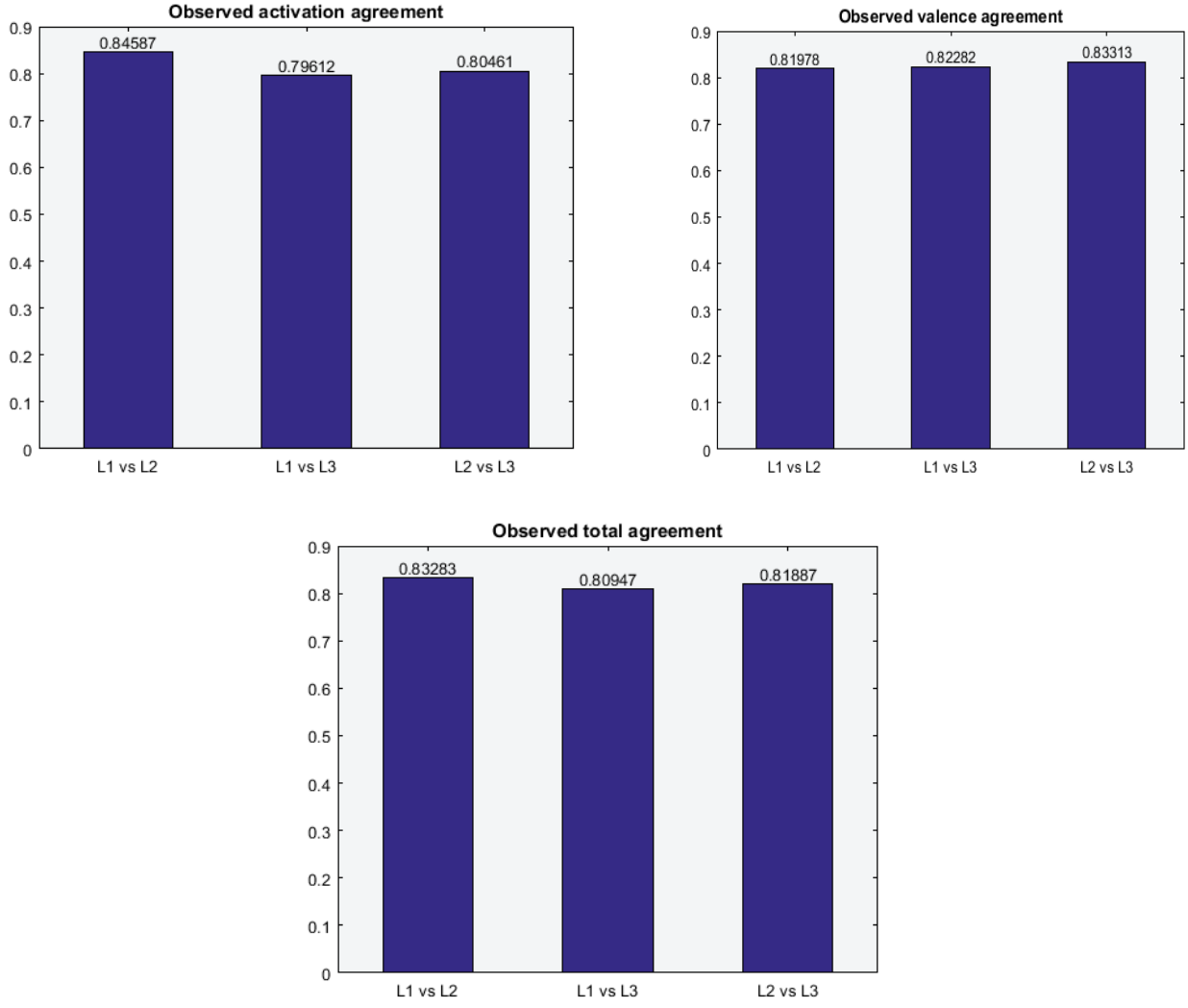
Στη συνέχεια, υπολογίζονται και παρουσιάζονται τα διαγράμματα διαφοράς τιμών για activation και valence και οι τιμές συμφωνίας για τους 3 συνδυασμούς ζυγών επισημειωτών, τα οποία υπολογίστηκαν σύμφωνα με τη σχέση της εκφώνησης:



Παρουσιάζεται μεγάλη συμφωνία και των τριών αναφορικά με το συναίσθημα, πράγμα λογικό καθώς είναι ευκολότερα διαχωρίσιμο χαρακτηριστικό. Παρατηρούνται διαφορές μεγέθους 2 για το πολύ το 12% των δειγμάτων (~50), ενώ στις περισσότερες περιπτώσεις είτε συμφωνούν είτε διαφέρουν κατά 1.

Στην περίπτωση του activation, παρατηρούνται διαφορές μεγέθους 2, αυτή τη φορά σχεδόν πάντα για το ~12% των δειγμάτων και διαφορές που φτάνουν μέχρι και 3, γεγονός που καταδεικνύει διαφορά στην αντίληψη της έντασης μεταξύ των επισημειωτών και κύριο χαρακτηρστικό που μειώνει τη «συμφωνία» τους. Στη συνέχεια, αναπαρίσταται η παρατηρούμενη συμφωνία, η οποία σε γενικές γραμμές είναι άνω του 80%, κάτι που φαινόταν εξ αρχής, λόγω των συγκλίνουσων μέσων τιμών και διασπορών. Ωστόσο, δεν ξεπερνάει το 83-84%, κάτι το οποίο φάνηκε από τη διαφορά τους στην

αντίληψη του συναισθήματος, ιδιαίτερα στην περίπτωση του επισημειωτή 3, που παρουσιάζει ελαφρώς διαφορετική συμπεριφορά από τους άλλους 2 σε αυτή τη διάσταση.



Βήμα 4

Στο βήμα αυτό υπολογίζεται ο συντελεστής Krippendorff's alpha, ο οποίος είναι ένα χρήσιμο μέτρο για εξετάσουμε κατά πόσο είναι έμπιστοι οι επισημειωτές μας :

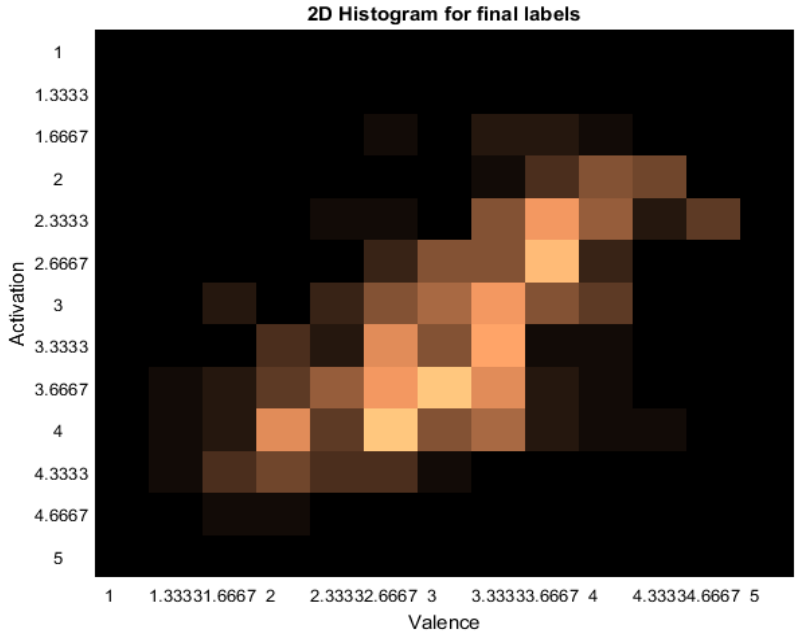
Activation : 0.4398

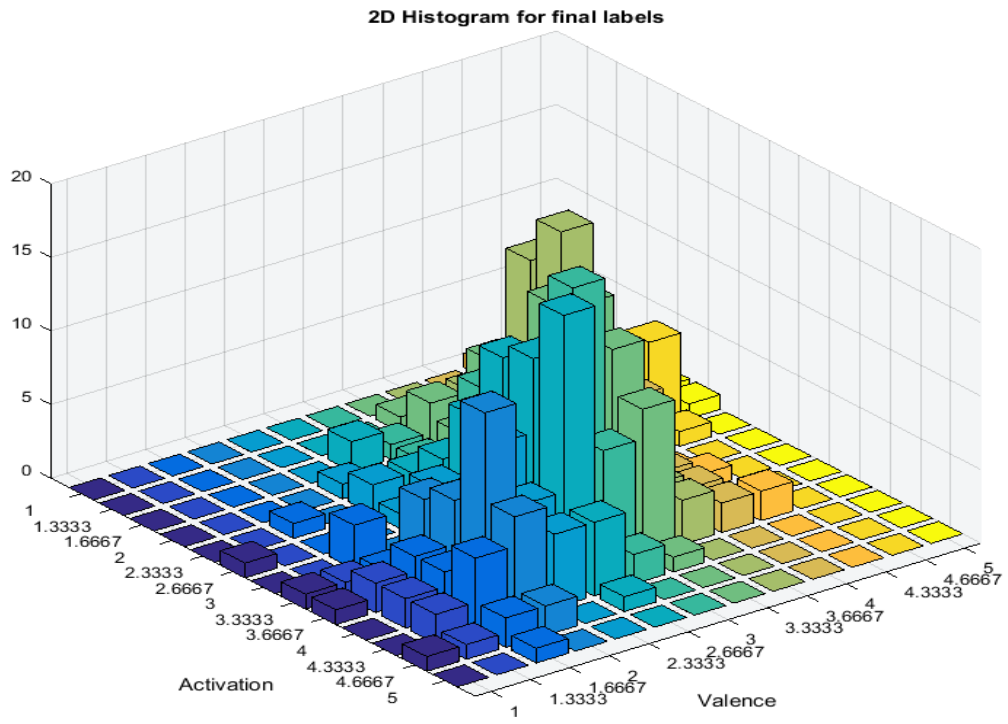
Valence : 0.4747

Τα αποτελέσματα αυτά δείχνουν ότι βρισκόμαστε αρκετά πάνω από την τυχαιότητα και ότι σαν σύνολο, οι επισημειωτές μας μπορούν να θεωρηθούν στην παρούσα φάση αξιόπιστοι. Ο κύριος παράγοντας μείωσης της τιμής του alpha αποτελεί το ότι λαμβάνεται υπόψη η αναμενόμενη διαφωνία, η οποία επηρεάζει αρνητικά το αποτέλεσμα.

Βήμα 5

Στο βήμα αυτό, λαμβάνεται ο μέσος όρος των τριών επισημειωτών για κάθε διάσταση και ουσιαστικά «κβαντίζεται» στις δοθείσες τιμές της εκφώνησης. Λαμβάνει, δηλαδή, την τιμή στην οποία είναι πιο κοντά της, που υπολογίζεται σύμφωνα με το ελάχιστο της απόλυτης τιμής της διαφοράς μεταξύ της παρούσας τιμής του –που προέκυψε από το μέσο όρο- και των δοθέντων τιμών. Ακολούθως, παρουσιάζεται ο co-occurrence πίνακας σε μορφές κατάλληλες για εξαγωγή ασφαλών συμπερασμάτων.





Όπως αναμενόταν, παρατηρούμε συγκέντρωση των τιμών στο κέντρο του πίνακα με τις μεγαλύτερες τιμές να εντοπίζονται για τιμή valence 3 και 3,67, δηλαδή θετικά συναισθήματα, όπως παρατηρείται και από τις μέσες τιμές, οι οποίες είναι ελαφρώς μεγαλύτερες του 3 και στις τιμές activation 4 και 4,67 που δείχνουν μεγάλη ένταση. Αν και τα μέγιστα δεν απέχουν τόσο από τις υπόλοιπες τιμές του

co-occurrence matrix, οι οποίες υποδεικνύουν μια ελαφρά συναισθηματική θετικότητα ουδέτερης προς έντονης έντασης.

Βήμα 6

Στο παρόν βήμα, εξάγονται ορισμένα χαρακτηριστικά για κάθε κομμάτι, όσον αφορά το συναισθηματικό του περιεχόμενο, τα οποία αναλύονται παρακάτω:

Ακουστική «Τραχύτητα»(roughness)

Χρησιμοποιώντας τη συνάρτηση *mirrougness(),* υπολογίζουμε το πόσο «σύμφωνο» (αρμονικό, πρίμο) ή «παράφωνο» (πιο μπάσος απότομος ήχος) είναι το κάθε ακουστικό δείγμα. Ουσιαστικά, υπολογίζεται το φάσμα του ακουστικού σήματος και πιο συγκεκριμένα, οι κορυφές του. Οι σύμφωνοι ήχοι έχουν πιο ομοιόμορφα κατανεμημένες κορυφές, σε αντίθεση με τους πιο παράφωνους ήχους. Ο διαχωρισμός των ήχων, σύμφωνα με αυτή με τη μετρική, έχει να κάνει με τις αντίστοιχες νότες, οκτάβες που περιέχει κάθε τραγούδι και είναι πιο πολύπλοκος από ό,τι φαίνεται.

Fluctuation (διακύμανση/ρυθμική περιοδικότητα ματαξύ ακουστικών καναλιών)

Χρησιμοποιώντας τη *mirfluctuation(),* υπολογίζουμε τις ρυθμικές διακυμάνσεις του κομματιού, δηλαδή τις εναλλαγές τέμπο. Πάλι, βασίζεται σε φασματική ανάλυση του σήματος και ειδικότερα, με την παράμετρο Summary που του δίνεται, μας παρουσιάζει μια καθολική αναπαράσταση της μετρικής, αθροίζοντας το φασματικό αποτέλεσμα κάθε band.

Key clarity

Η συνάρτηση *mirkey()* επιλέγει τις κορυφές από την καμπύλη που της επιστρέφει η *mirkeystrength(),* η οποία, χρησιμοποιώντας το χρωμόγραμμα της *mirchromogramm(),* μέσω ετεροσυσχέτισης, ουσιαστικά κατατάσσει τονικά (σύμφωνα δηλαδή με τις νότες) το αντίστοιχο πλαίσιο του μουσικού κομματιού. Να σημειωθεί ότι λόγω του προκαθορισμένου παράθυρου της *mirkey()* (1 sec), της δώσαμε εκτός από παράμετρο παραθύρου('Frame') και το μήκος του παραθύρου (το overlap της είναι 50%, οπότε καλυπτόμαστε). Τέλος, αυτή η μετρική ουσιαστικά μας δίνει τα τονικά χαρακτηριστικά του τραγουδιού, δηλαδή τις νότες στις οποίες «παίζεται».

Modality

Ουσιαστικά, το modality στη μουσική είναι η κλίμακα μαζί με χαρακτηριστικά μελοδικής συμπεριφοράς για ένα κομμάτι. Χρησιμοποιώντας και πάλι την αντίστοιχη συνάρτηση, επιστρέφεται ως τιμή στο διάστημα [-1,1] και όσο πιο κοντά στο 1 είναι, τόσο πιο ματζόρε είναι ο ήχος, ενώ αντίστοιχα στο -1, τόσο πιο μινόρε είναι.

Novelty

Αυτό το χαρακτηριστικό σχετίζεται κυρίως με τις διαδοχές τέμπο στο κομμάτι. Χρησιμοποιείται (προκαθορισμένο) η MULTI-GRANUL, η οποία ουσιαστικά για κάθε «στιγμή» μέσα στο κομμάτι,

υπολογίζει το τέμπο (temporal scale) του προηγούμενου μέρους και το βαθμό αντίθεσης του προηγούμενου με το επόμενο μέρος. Για την επιλογή της άλλης μεθόδου, η οποία είναι πιο κοστοβόρα υπολογιστικά, πρέπει να τεθεί στη συνάρτηση η παράμετρος 'Kernel'.

Harmonic Changes in Detection Function

Αυτό το χαρακτηριστικό μας δείχνει τις εναλλαγές αρμονίας μέσα σε ένα μουσικό κομμάτι και οι υπολογισμοί που γίνονται, περιλαμβάνουν και πάλι φασματική ανάλυση, χρωμόγραμμα. Τα αποτελέσματα που μας δίνει είναι η ροή του «τονικού» κέντρου βάρους κάθε κομματιού.

Σε όλες τις περιπτώσεις χρησιμοποιήθηκαν οι δοσμένες συναρτήσεις της εκφώνησης, οι αντίστοιχες παράμετροι και υπολογίστηκαν οι αντίστοιχες ζητούμενες μετρικές με τις συναρτήσεις *max() mean() std()*. Επίσης, να τονιστεί ότι η συνάρτηση *mirgetdata()* χρησιμοποιείται σε κάθε βήμα, για να εξάγουμε τα αποτελέσματα κάθε κληθείσας συνάρτηση στην απαιτούμενη μορφή για να επεξεργαστούμε. Περιέχονται πολλές ενδιαφέρουσες συναρτήσεις στο toolbox, όπως το chromogram που εξάγει σημαντικά χαρακτηριστικά για ένα μουσικό κομμάτι και το χρησιμοποιούν πολλές από τις παραπάνω συναρτήσεις για τους υπολογισμούς τους και το emotion, που από μόνο του αποτελεί έναν εκτιμητή συναισθήματος του κομματιού σε 3 διαστάσεις (Valence, Activation, Tension προαιρετικά είναι 3) και υπολογίζει και το concept του τραγουδιού ανάμεσα σε 5 κατηγορίες -θυμό, χαρά, λύπη, φόβο, τρυφερότητα-.

Βήμα 7

Στο βήμα αυτό, υπολογίζουμε τους mfcc συντελεστές, με τη χρήση της συνάρτησης *mirmfcc()* ως ακολούθως:

- `26 κανάλια,13 συντελεστές και συγκεκριμένες πραμέτρους παραθύρων`
  ```
  MIRMFCCs =
  mirgetdata(mirmfcc(sample,'Frame',0.025,'s',0.01,'s','Bands',26,'Rank',1:13));
  ```
- `Όμοια με πριν όμως τώρα θέλουμε τις πρώτες παραγώγους άρα Delta,1`
  ```
  MIRMFCCs_d =
  mirgetdata(mirmfcc(sample,'Frame',0.025,'s',0.01,'s','Bands',26,'Rank',1:13,'Delt
  a',1))
  ```
- `Όμοια με πριν όμως τώρα θέλουμε τις δεύτερες παραγώγους άρα Delta,2`
  ```
  MIRMFCCs_d =
  mirgetdata(mirmfcc(sample,'Frame',0.025,'s',0.01,'s','Bands',26,'Rank',1:13,'Delt
  a',1))
  ```

Σε κάθε περίπτωση εξάγονται οι ζητούμενες μετρικές και όσον αφορά τις μέσες τιμές των 10% μεγαλύτερων και μικρότερων στοιχειών, χρησιμοποιείται η συνάρτηση sort του Matlab, στη μια περίπτωση ως έχει (10% μικρότερων) και στην άλλη, με παράμετρο 'descend' για φθίνουσα ταξινόμηση, ώστε να παρθούν τα 1 ως 10%* στοιχεία. Αν και θα μπορούσε να μην χρησιμοποιηθεί η παράμετρος και να λαμβανόταν από τον ήδη αύξων ταξινομημένο πίνακα το [90%*#elements, end] αυτού του πίνακα, που αντιστοιχεί στα ίδια κελιά με τα αντίσοιχα του φθίνοντα ταξινομημένου.

Βήμα 8

Προετοιμάστηκαν οι ταξινομητές (NNR-K,NNR-1, Naïve Bayes, SVM), οι οποίοι χρησιμοποιήθηκαν στην $1^{η}$ άσκηση για ταξινόμηση ψηφίων.

Βήμα 9

Εγκαταστήθηκε το εργαλείο Weka και πειραματίστηκα με τα έτοιμα δεδομένα σε μορφή *.artff που έχει, τα οποία είναι σε μορφή προς χρήση από το ίδιο το πρόγραμμα για αλγόριθμους machine learning (πχ svm).