



**ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΜΜΕ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ:
ΑΥΤΟΜΑΤΗ ΑΝΑΓΝΩΡΙΣΗ ΗΧΩΝ ΣΕ ΡΑΔΙΟΦΩΝΙΚΕΣ
ΕΚΠΟΜΠΕΣ**




Μανιδάκη Αγάπη

Επιβλέπων Καθηγητής: Κούτρας Αθανάσιος

Πύργος, 2017

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 ΣΠΟΥΔΑΣΤΗ 1	ΑΜ	ΥΠΟΓΡΑΦΗ
ΜΑΝΙΔΑΚΗ ΑΓΑΠΗ	1328	
ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΣΠΟΥΔΑΣΤΗ 2 <i>(σε περίπτωση που είναι απαραίτητο)</i>	ΑΜ	ΥΠΟΓΡΑΦΗ
.....
ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΣΠΟΥΔΑΣΤΗ 3 <i>(σε περίπτωση που είναι απαραίτητο)</i>	ΑΜ	ΥΠΟΓΡΑΦΗ
.....

(Αυτή η σελίδα είναι σκοπίμως κενή)

**ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΜΜΕ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ:
ΑΥΤΟΜΑΤΗ ΑΝΑΓΝΩΡΙΣΗ ΗΧΩΝ ΣΕ ΡΑΔΙΟΦΩΝΙΚΕΣ
ΕΚΠΟΜΠΕΣ**

Μανιδάκη Αγάπη

Επιβλέπων Καθηγητής: Κούτρας Αθανάσιος

Πύργος, 2017

(Αυτή η σελίδα είναι σκοπίμως κενή)

ΠΕΡΙΛΗΨΗ

Στην εργασία αυτή θα ασχοληθούμε με την κατάτμηση και ταξινόμηση ηχητικών σημάτων, ομιλίας και μουσικής, σε ραδιοφωνικές εκπομπές. Χρησιμοποιώντας το πρόγραμμα Marsyas θα εξάγουμε βασικά χαρακτηριστικά από τα ηχητικά μας δείγματα και έπειτα θα τα εισάγουμε στο πρόγραμμα Weka, όπου με χρήση ταξινομητών και των αποτελεσμάτων του προηγούμενου βήματος επεξεργασίας, θα εκτιμήσουμε τα αποτελέσματα ταξινόμησης. Τέλος, σε θεωρητικό επίπεδο θα παρουσιάσουμε βασικά θέματα που σχετίζονται με τον ήχο αλλά και με έννοιες που αφορούν τα πειράματα της εργασίας μας.

ABSTRACT

In this thesis we will deal with the segmentation and classification of speech and music audio signals problems, in radio broadcast shows. Using Marsyas we will extract key features of our audio samples which will be imported into pattern recognition program Weka, where using classifiers and the results we got from the previous step, we will estimate our results. Finally, in theoretical level we will examine basic theoretical issues related to sound and concepts used in our experiments.

Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1	10
1.1 Εισαγωγή.....	10
1.2 Βιβλιογραφική Ανασκόπηση	11
ΚΕΦΑΛΑΙΟ 2	13
2.1 Θεωρία του ήχου.....	13
2.2 Ανθρώπινο σύστημα ακοής.....	13
2.2.1 Εξωτερικό αυτί.....	13
2.2.2 Μέσο αυτί	14
2.2.3 Εσωτερικό αυτί	14
2.3 Είδη ήχων κ θορύβων	14
ΚΕΦΑΛΑΙΟ 3	16
3.1 Χαρακτηριστικά του ήχου	16
3.1.1 Αντικειμενικά χαρακτηριστικά.....	16
3.1.2 Υποκειμενικά χαρακτηριστικά.....	16
3.1.3 Επιπλέον χαρακτηριστικά	18
ΚΕΦΑΛΑΙΟ 4	21
4.1 Αλγόριθμοι κατηγοριοποίησης (Classification algorithms).....	21
4.1.1 Αλγόριθμος υποστήριξης διανυσματικών μηχανών (SVM).....	21
4.1.2 Αλγόριθμος Μπείσιανής κατηγοριοποίησης (Naïve Bayes).....	22
4.1.3 Αλγόριθμοι τεχνητών νευρωνικών δικτύων (Artificial Neural Networks- ANN).....	22
4.1.4 Κατηγοριοποίηση με κ χωρίς επιτήρηση	23
4.1.5 Multilayer Perceptron	24
4.1.6 Logistic model tree (LMT)	25
ΚΕΦΑΛΑΙΟ 5 : Data Analysis.....	26
5.1 Πίνακας Σύγχυσης (Confusion Matrix)	26
5.2 Stratified Cross – Validation.....	26
5.2.1 Kappa – statistics	26
5.2.2 Μέσο απόλυτο σφάλμα (MAE)	27
5.2.3 Ρίζα του μέσου τετραγωνικού σφάλματος (RMSE)	27
5.2.4 Απόλυτο Σχετικό Σφάλμα (Relative absolute error (Rae)).....	27
5.2.5 Root Relative Squared Error (Rrse).....	27
5.3 Ακρίβεια ταξινομητή ανά κατηγορία.....	28

5.3.1 TruePositiverate	28
5.3.2FalsePositiverate	28
5.3.3 Ακρίβεια ταξινόμητη (Precision).....	28
5.3.4 Ανάκληση ταξινόμητη (Recall)	28
5.3.5 F-measure.....	29
5.3.6 ΣυντελεστήςσυσχέτισηςMatthews (Matthews correlation coefficient – MCC).....	29
5.3.7 AUC- AreaUnderCurves	30
ΚΕΦΑΛΑΙΟ 6 : ΠΕΙΡΑΜΑΤΙΚΗ ΔΙΑΔΙΚΑΣΙΑ.....	32
6.1 Προτεινόμενη Μεθοδολογία.....	32
6.1.1 Εξαγόμενα Χαρακτηριστικά	32
6.1.2 Βάση Δεδομένων	33
6.2 Αποτελέσματα.....	33
6.2.1 Ταξινόμητης LibSVM.....	34
6.2.2 Ταξινόμητης LMT	41
6.2.3 Ταξινόμητης Multilayer Perceptron.....	48
6.2.4 Ταξινόμητης Naïve Bayes.....	55
6.3 Συμπεράσματα – Προοπτικές εξέλιξης.....	62
Βιβλιογραφικές Αναφορές.....	66
Διαδικτυακές Πηγές.....	67

ΚΕΦΑΛΑΙΟ 1

1.1 Εισαγωγή

Με την ανάπτυξη της τεχνολογίας όλο και περισσότερες εφαρμογές μπαίνουν στην ζωή μας είτε συμβάλλοντας στην εκπαίδευση, είτε στην ψυχαγωγία, ακόμη και εξυπηρετώντας μας σε καθημερινά ζητήματα. Από τα τεχνολογικά προϊόντα δεν θα μπορούσαν να λείπουν εφαρμογές που έχουν να κάνουν με την ομιλία και την μουσική. Η αναγνώριση και η ταξινόμηση της ομιλίας και της μουσικής είναι κ το θέμα που θα μας απασχολήσει στην παρούσα εργασία.

Αρκετές είναι οι εφαρμογές που συναντάμε γύρω μας και έχουν κοινή βάση με το πρόβλημα μας. Αρχικά, από τα πιο διαδεδομένα προγράμματα σήμερα είναι αυτά που αναγνωρίζουν μουσική. Όλοι σχεδόν που έχουν κινητά τελευταίας τεχνολογίας και γενικά ασχολούνται με gadgets, έχουν χρησιμοποιήσει ένα τέτοιο πρόγραμμα. Η αναγνώριση ενός ήχου γενικά χρησιμοποιείται και σε εκπαιδευτικά multimedia προγράμματα, αλλά και σε αυτόματα συστήματα εξυπηρέτησης μέσω τηλεφώνου, όπου έχουν αντικαταστήσει τους τηλεφωνητές στα διάφορα κέντρα και υπηρεσίες. Από τα πιο σημαντικά επιτεύγματα που έχει σχέση με την αναγνώριση, δεν είναι άλλο από την μετατροπή φυσικής ομιλίας σε κείμενο κ το αντίθετο (μετατροπή κειμένου σε συνθετική ομιλία).

Για να επιτευχθεί η *αναγνώριση* ενός ήχου, χρησιμοποιούνται αρχικά κάποιες άλλες τεχνικές. Αυτές είναι:

1. Ανίχνευση ήχου (Detection)
2. Κατάτμηση ήχου (Segmentation)
3. Ταξινόμηση ήχου (Clustering)

Οι κύριες κατηγορίες που διαχωρίζεται ένας ήχος είναι:

- Μουσική
- Καθαρή Ομιλία
- Περιβαλλοντικοί Θόρυβοι

Μπορούμε όμως από αυτές τις κατηγορίες να διεξάγουμε και άλλες, όπως:

- Ομιλία πάνω από μουσική
- Ομιλία πάνω από περιβαλλοντικό θόρυβο
- Σιωπή

Η ανίχνευση, η κατάτμηση και η ταξινόμηση αφορούν τμήματα ομιλίας, μουσικής, του ρόλου του ομιλητή, του θέματος μιας εκπομπής κ.α.

Για να εκτελεστούν οι τεχνικές αυτές χρησιμοποιούνται διάφορα μοντέλα, αλγόριθμοι, ταξινομητές, καθώς και τα ίδια τα χαρακτηριστικά του ήχου. Κάποια από αυτά, μπορούν να

χρησιμοποιηθούν μόνα τους, αλλά και σε συνδυασμό με κάποια άλλα. Ανάλογα με τα τμήματα που αναφερόμαστε χρησιμοποιούμε και άλλες προσεγγίσεις.

Κάποιοι από τους πιο γνωστούς αλγόριθμους είναι:

- *Αλγόριθμος υποστήριξης διανυσματικών μηχανών (SVM)*
- *Αλγόριθμος Gaussian mixture model (GMM)*
- *Αλγόριθμος Μπεϊσιανής κατηγοριοποίησης (Bayes)*

Κάποια από τα πιο γνωστά μοντέλα είναι:

- *Model-based (HMM, GMM, SVM, BIC)*
- *Metric-based*
- *Όπως και μοντέλα που βασίζονται στα χαρακτηριστικά του ηχητικού σήματος, ή στα χαρακτηριστικά του MPEG-7 κ.α.*

Με όλα τα παραπάνω μπορούμε να πραγματοποιήσουμε αναγνώριση ηχητικών δεδομένων σε ραδιοηλεκτρονικές εκπομπές οποιουδήποτε περιεχομένου (αθλητικές, ενημερωτικές, ειδησεογραφικές κ.α.), σε sportgames, ακόμα και σε εκπομπές άλλων χωρών (ξενόγλωσσες).

1.2 Βιβλιογραφική Ανασκόπηση

Πολλά συστήματα για την κατάτμηση του ήχου σε ομιλία και μουσική έχουν παρουσιαστεί στο παρελθόν. Διαφορετικές τεχνικές, διαφορετικοί αλγόριθμοι εκπαίδευσης, πειραματισμοί με διαφορετικά σύνολα εξαγόμενων χαρακτηριστικών.

Οι Α. Πικράκης, Θ. Γιαννακόπουλος και Σ. Θεοδωρίδης (2006), παρουσίασαν έναν υπολογιστικά αποδοτικό διαχωριστή ομιλίας/μουσικής. Πρόκειται για ένα offline σύστημα που χρησιμοποιεί μία τεχνική με ένα μόνο εξαγόμενο χαρακτηριστικό, την χρωματική ενέργεια. Το σύστημα εξετάστηκε σε ηχογραφημένες ραδιοφωνικές εκπομπές του διαδικτύου και πέτυχε μέση ακρίβεια 93,38%.¹ Μία μέθοδος παρακολούθησης ραδιοφωνικών εκπομπών σε πραγματικό χρόνο προτάθηκε από τον J. Saunders (1996). Το σύστημα βασίστηκε στην χρήση ενέργειας και zerocrossingrate, ως εξαγόμενα χαρακτηριστικά κ πέτυχε ποσοστό ακρίβειας 98%.² Οι R. Gael, R. Mathieu και E. Slim (2007), συνδύασαν επιβλεπόμενη και μη επιβλεπόμενη κατηγοριοποίηση και είχαν ποσοστό επιτυχίας 96,5%, όπου μόνο με επιβλεπόμενη έφταναν στο 95,9%.³ Πολλοί βασίστηκαν στους MFCC για τα πειράματά τους, ωστόσο οι M. Kos, M. Grašič, D. Vlaj και Z. Kačič (2009), χρησιμοποίησαν τους MFCCV (MelFrequencyCepstralCoefficientsVariance) για online κατάτμηση, πετυχαίνοντας 98,06% ακρίβεια κ περνώντας κατά 1,3% την κατάτμηση όπου

¹A.Pikrakis., Th. Giannakopoulos , and S. Theodoridis. "A computationally efficient speech/music discrimination for radio recordings", Department of Informatics and Telecommunications, University of Athens.

²J. Saunders. "Real-Time Discrimination of Broadcast Speech/Music," 1996, Sanders, A Lockheed Martin Co., Nashua.

χρησιμοποιήθηκαν μόνο οι MFCC.⁴ Οι E. Scheirer και M. Slaney (1997), εξήγαγαν 13 χαρακτηριστικά ήχου για να εκπαιδεύσουν τρία πολυδιάστατα συστήματα ταξινόμησης. Ο καλύτερος από τους τρεις έδωσε ποσοστό ακρίβειας πάνω από 94%.⁵

³R. Gael, R. Mathieu, and E. Slim. “Combined supervised and unsupervised approaches for automatic segmentation of radiophonic audio streams,” 2007, GET-ENST, Paris, France.

⁴R. Gael, R. Mathieu, and E. Slim. “Combined supervised and unsupervised approaches for automatic segmentation of radiophonic audio streams,” 2007, GET-ENST, Paris, France.

⁵ Scheiner Eric, Slaney Malcolm. “Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator,” 1997, Interval Research Corp., Palo Alto.

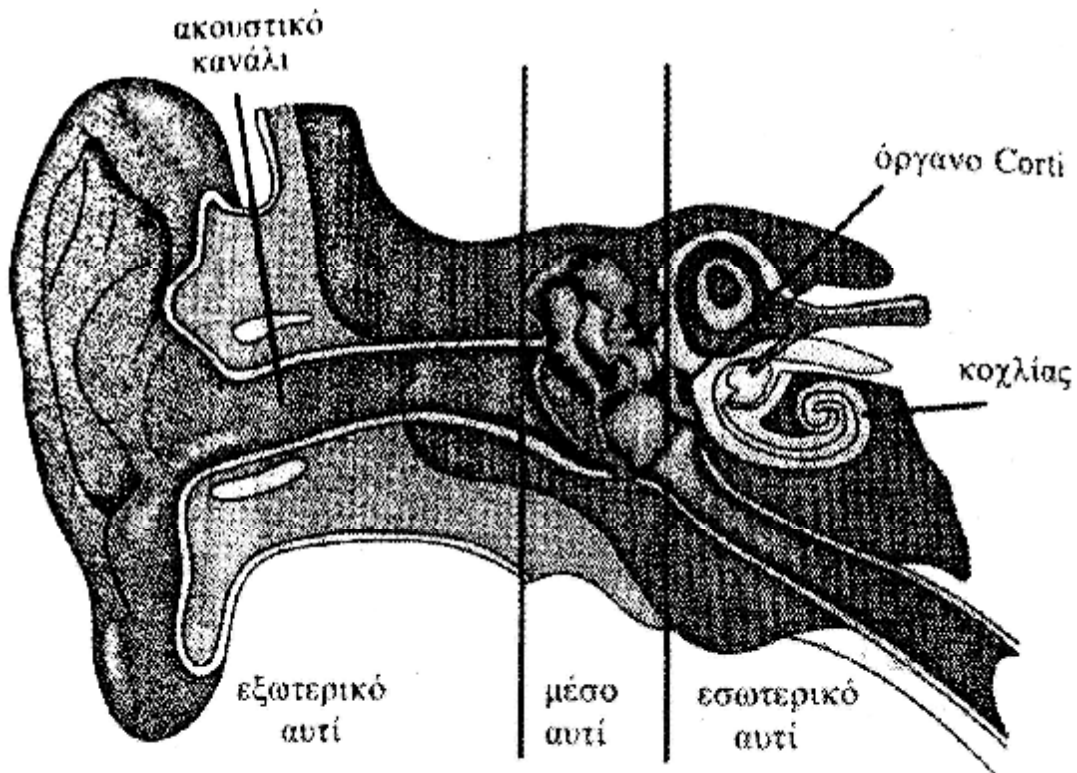
ΚΕΦΑΛΑΙΟ 2

2.1 Θεωρία του ήχου

Ο ήχος είναι ένα είδος ενέργειας, που ονομάζεται ακουστική ενέργεια. Η ακουστική ενέργεια είναι μεταβολές της πίεσης του ατμοσφαιρικού αέρα ή αλλιώς μεταβαλλόμενα κύματα πίεσης, τα οποία διαδίδονται σ' ένα ελαστικό μέσο, που είναι συνήθως ο αέρας.

2.2 Ανθρώπινο σύστημα ακοής

Το ανθρώπινο αυτί αποτελείται από τρία μέρη: το εξωτερικό, το μέσο κ το εσωτερικό αυτί.



Εικόνα 1: Ανθρώπινο σύστημα ακοής

2.2.1 Εξωτερικό αυτί

Το εξωτερικό αυτί περιλαμβάνει το πτερύγιο, το ακουστικό κανάλι κ το τύμπανο. Το πτερύγιο κατευθύνει τα ηχητικά κύματα προς το ακουστικό κανάλι κ ταυτόχρονα εμποδίζει ήχους που προέρχονται από το πίσω μέρος της κεφαλής, βοηθώντας έτσι στον κατ' αρχήν εντοπισμό της πηγής του ήχου. Στο τέλος του ακουστικού καναλιού υπάρχει το τύμπανο ή τυμπανική μεμβράνη, που χωρίζει το εξωτερικό μέρος από το μέσο.

2.2.2 Μέσο αυτί

Πίσω από το τύμπανο, βρίσκεται το μέσο αυτί, δηλαδή μία μικρή κοιλότητα μέσα στην οποία υπάρχουν τρία μικρά οστά: η σφύρα, ο άκμων κ ο αναβολέας. Η σφύρα είναι προσκολλημένη στο τύμπανο, κ μεταδίδει, μέσω του άκμωνος κ του αναβολέα, τις δονήσεις του τυμπάνου στην ωοειδή θυρίδα στην οποία βρίσκεται προσκολλημένος, ο αναβολέας. Τα τρία οστά σχηματίζουν μεταξύ τους ένα σύστημα μοχλών, το ποίο ανιχνεύει τη δόνηση του τύμπανου κ τη μεταφέρει ενισχυμένη στην μεμβράνη της ωοειδούς θυρίδας. Παράλληλα περιορίζουν το πλάτος δόνησης του τύμπανου.

2.2.3 Εσωτερικό αυτί

Αποτελεί το πιο ενδιαφέρον κ περισσότερο πολύπλοκο κομμάτι του ανθρώπινου ακουστικού συστήματος, όπου μετατρέπει την μηχανική ενέργεια των δονήσεων που φτάνουν από το τύμπανο στο μέσο αυτό, σε ηλεκτροχημικό σήμα κ αποτελείται από τον κοχλία κ τον λαβύρινθο. Ο κοχλίας έχει μήκος περίπου 32 mm, αποτελείται από 2.5 στροφές κ περιέχει δύο διαφορετικά οργανικά υγρά, τα οποία διαχωρίζονται από την βασική μεμβράνη.

Ο ρόλος της βασικής μεμβράνης είναι θεμελιώδης για τον μηχανισμό λειτουργίας της ακοής αφού επιτελεί την συχνοτική ανάλυση του ηχητικού σήματος. Συντονίζεται σε διαφορετικά σημεία για διαφορετικές συχνότητες κ λόγω αυτής της ιδιότητας χαρακτηρίζεται «αναλυτής συχνοτήτων». Για πολύ χαμηλές συχνότητες η ταλάντωση της βασικής μεμβράνης μηδενίζεται στο τέρμα του κοχλίου, ενώ για υψηλότερες νωρίτερα. Για την περίπτωση της ύπαρξης περισσότερων από μία συχνοτήτων, παρατηρούνται μέγιστα πλάτους σε περισσότερες θέσεις.

Επάνω στην μεμβράνη βρίσκεται το όργανο του *corti*, που συγκροτείται από τριχοειδή κύτταρα κ μετατρέπει τις μηχανικές ταλαντώσεις της βασικής μεμβράνης σε ηλεκτρικό σήμα που οδηγείται στο ακουστικό νεύρο. Οι κάθετες απομακρύνσεις της βασικής μεμβράνης μετατρέπονται σε ακτινικές κινήσεις τριχιδίων, μεταβάλλοντας με αυτόν τον τρόπο την ηλεκτρική αντίσταση του κυττάρου, το οποίο με ηλεκτροχημικές διαδικασίες, παράγει παλμούς ηλεκτρικού ρεύματος που μεταφέρονται στον εγκέφαλο από ένα μεγάλο αριθμό νευρικών ινών.⁶

2.3 Είδη ήχων κ θορύβων

- Ø Απλός ήχος, είναι ο ήχος που παράγεται από αρμονική διαταραχή.
- Ø Θόρυβος, ονομάζεται κάθε δυσάρεστος ή ανεπιθύμητος ήχος.
 - **Λευκός θόρυβος**, είναι ο θόρυβος που έχει συνεχές ηχητικό φάσμα σε μία ευρεία περιοχή συχνοτήτων που ενδιαφέρει και σταθερή φασματική πυκνότητα.
 - **Ροζ θόρυβος**, είναι ο θόρυβος που έχει συνεχές ηχητικό φάσμα με μία σταθερή ηχητική ισχύ σε κάθε φασματική ζώνη μεταβλητού ζωνικού εύρους.
 - **Θόρυβος Βάθους**, είναι ο θόρυβος που προέρχεται από όλες τις πηγές που βρίσκονται στο περιβάλλον κ δεν μας ενδιαφέρουν.

⁶Λ. Χαδέλλης, *Τεχνολογία Ήχου*, Μάρτιος 2010, Πανεπιστημιακές Εκδόσεις «ΑΡΑΚΥΝΘΟΣ».

- **Θόρυβος Περιβάλλοντος**, είναι ο θόρυβος που εκπέμπεται από συγκεκριμένες πηγές που δεν μας ενδιαφέρουν κατά την στιγμή της μέτρησης. Είναι μέρος του θορύβου βάρους.
- **Σταθερός θόρυβος**, είναι ο θόρυβος που παρουσιάζει αμελητέα μικρές διακυμάνσεις στάθμης κατά την διάρκεια της παρατήρησης.
- **Μεταβλητός θόρυβος**, είναι ο θόρυβος που δεν είναι σταθερός.
- **Κομινόμενος θόρυβος**, είναι ο μεταβλητός θόρυβος που η στάθμη του αλλάζει συνεχώς κ σε σημαντικό βαθμό κατά την διάρκεια της παρατήρησης.⁷

⁷Δ. Σκαρλάτος, *Εφαρμοσμένη Ακουστική*, Γ' Έκδοση, Οκτώβριος 2008, Πάτρα, GOTSIS.

ΚΕΦΑΛΑΙΟ 3

3.1 Χαρακτηριστικά του ήχου

Η ιδιαιτερότητα του κάθε ήχου οφείλεται σε ένα σύνολο από ηχητικά μεγέθη, άλλα από τα οποία είναι ανεξάρτητα από την προσωπική αντίληψη του ακροατή κ αποκαλούνται αντικειμενικά, ενώ άλλα είναι συνδεδεμένα με τον ακροατή κ λέγονται υποκειμενικά χαρακτηριστικά του ήχου.

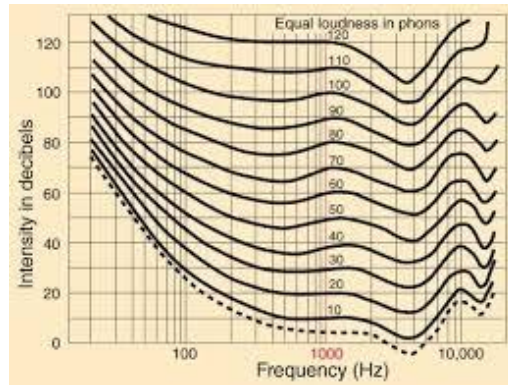
3.1.1 Αντικειμενικά χαρακτηριστικά

- a) ΣΥΧΝΟΤΗΤΑ: Η συχνότητα έχει να κάνει με το πόσο γρήγορα ή αργά πάλλεται το σώμα που δημιουργεί τον ήχο. Ορίζει τον αριθμό των κύκλων που εκτελεί η συνάρτηση ανά δευτερόλεπτο κ μετριέται σε κύκλους ανά δευτερόλεπτο ή Hertz (Hz). Κάθε ήχος για να γίνει αντιληπτός από το ανθρώπινο αυτί, θα πρέπει να έχει συχνότητα μεταξύ 20 και 20000 Hz. Οι πιο χρήσιμες συχνότητες βρίσκονται κάτω από 10 KHz, όπως η μουσική, η ομιλία κ οι διάφοροι θόρυβοι.

- b) ΕΝΤΑΣΗ: Το δεύτερο βασικό χαρακτηριστικό του ήχου είναι η ένταση, η οποία είναι στενά συνδεδεμένη με την ισχύ του ηχητικού σήματος που διεγείρει το αυτί μας. Η κύρια αιτία που οι ήχοι έχουν διαφορετικές εντάσεις, είναι ότι πιέζουν με διαφορετική δύναμη το τύμπανο του αυτιού μας. Στην ουσία μιλάμε για το πόσο δυνατές είναι οι αναταράξεις που προκαλεί το σώμα που παράγει τον ήχο. Όσο πιο μεγάλη ισχύ διαμορφώνουν τα ηχητικά κύματα τόσο περισσότερη δύναμη εξασκούν στον μηχανισμό του αυτιού μας. Η ένταση ορίζεται από το πλάτος της δόνησης με αποτέλεσμα όσο μεγαλύτερο είναι το πλάτος, τόσο ισχυρότερος είναι ο ήχος. Η ένταση μετριέται σε Decibels (dB).

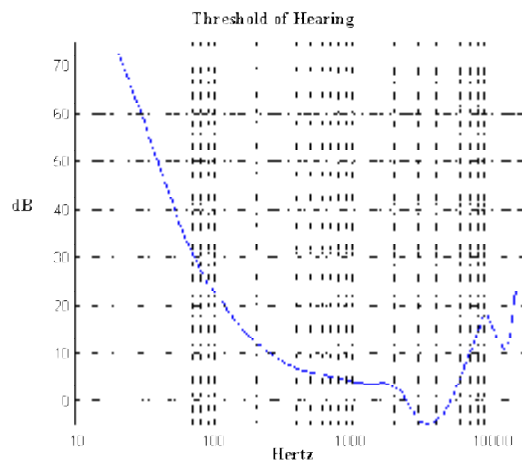
3.1.2 Υποκειμενικά χαρακτηριστικά

- a) ΑΚΟΥΣΤΟΤΗΤΑ: Αποτελεί ένα από τα πιο βασικά χαρακτηριστικά ενός ήχου, αφού ο κάθε ακροατής τα αντιλαμβάνεται διαφορετικά. Για παράδειγμα ένας ήχος με ίδια ένταση θα έχει μικρότερη ακουστότητα για ένα άτομο με πρόβλημα στην ακοή. Η ακουστότητα έχει σχέση με την ένταση του ήχου. Σε σχέση με την ακουστότητα οι ήχοι διακρίνονται σε **ισχυρούς κ ασθενείς**.



Εικόνα 2: Καμπύλες ακουστότητας

Το απόλυτο κατώφλι ακουστότητας υποδηλώνει την στάθμη ηχητικής πίεσης ενός θορύβου ή ενός τονικού σήματος, η οποία είναι μόλις ακουστή. Επομένως, μπορεί να θεωρηθεί ότι είναι το κατώτερο όριο της δυναμικής περιοχής του ανθρώπινου ακουστικού συστήματος.



Εικόνα 3: Απόλυτο κατώφλι ακουστότητας

- b) ΥΨΟΣ: Υποκειμενικό χαρακτηριστικό του ήχου, αντίστοιχο τη συχνότητας, είναι το ύψος του. Η σχέση αυτή χωρίζει τους ήχους σε οξείς κ βαρείς. Όσο μεγαλύτερη είναι η συχνότητα, τόσο οξύτερος είναι κ ο ήχος. Οι χαμηλότερες συχνότητες γίνονται αντιληπτές σαν βαθύς ήχος.
- c) ΧΡΟΙΑ (ή ΗΧΟΧΡΩΜΑ): Λόγω της χροιάς, ένας ήχος ξεχωρίζει ακόμα κ αν τα υπόλοιπα χαρακτηριστικά είναι τα ίδια. Η χροιά αναφέρεται στους σύνθετους ήχους κ εξαρτάται από τους απλούς που αποτελούν το σύνθετο.

3.1.3 Επιπλέον χαρακτηριστικά

a) Zero crossing rate (πλήθος μηδενισμού συνάρτησης)

Στο πλαίσιο των σημάτων διακριτού χρόνου, το πλήθος μηδενισμού συνάρτησης προκύπτει αν τα διαδοχικά δείγματα έχουν διαφορετικά αλγεβρικά σημεία. Ο ρυθμός με τον οποίο εμφανίζεται το zero crossing rate είναι ένα απλό μέτρο του συχνοτικού περιεχομένου ενός σήματος. Μετρά το πόσες φορές περνά το πλάτος των σημάτων ομιλίας από μία τιμή μηδέν, σε ένα δεδομένο χρονικό διάστημα.⁸

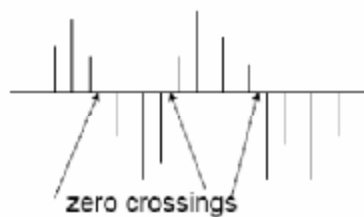
$$Z_n = \sum_{m=-\infty}^{\infty} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| w(n-m)$$

Όπου:

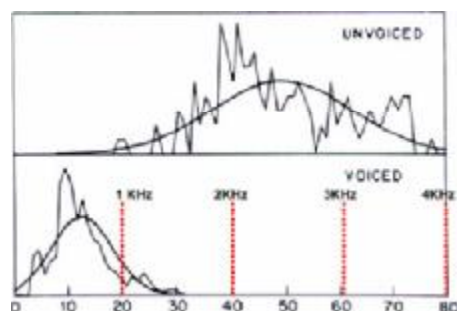
$$\text{sgn}[x(m)] = \begin{cases} 1, & x(m) \geq 0 \\ -1, & x(m) < 0 \end{cases}$$

$$\text{και } w(n) = \frac{1}{2N}, \text{ για } 0 \leq n \leq N-1$$

$$w(n) = 0, \text{ για οτιδήποτε άλλο.}$$



Εικόνα 4: Ορισμός zero crossing rate



Εικόνα 5: Κατανομή zero crossings για ομιλία κ αφωνία

⁸R. G.Bachu, S. Kopparthi S, B. Adapa, B. D. Barkana "Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal," 2008, Electrical Engineering Department, School of Engineering, University of Bridgeport.

b) Spectral features (χαρακτηριστικά φάσματος)

i. *Spectral centroid (φασματικό κέντρο βάρους)*

Το φασματικό κέντρο βάρους ορίζεται ως το κέντρο βάρους του φασματικού μεγέθους του μετασχηματισμού Fourier. Για την ανάλυση της μουσικής, διαιρείται το αρχείο του ήχου σε επικαλυπτόμενα πλαίσια από μικρότερα μήκη k υπολογίζεται ο μετασχηματισμός Fourier του κάθε πλαισίου.

$$C_t = \frac{\sum_{n=1}^N n M_t[n]}{\sum_{n=1}^N M_t[n]}$$

Όπου: $M_t[n]$ είναι το μέγεθος του μετασχηματισμού Fourier στο πλαίσιο t k στο σύνολο συχνοτήτων n .

Συνήθως, η μέση τιμή k η διακύμανση του κέντρου βάρους μεταξύ των διαφορετικών χρονικών πλαισίων στο παράθυρο υφής, χρησιμοποιούνται ως χαρακτηριστικά γνωρίσματα.⁹

ii. *Spectral Roll Off (φασματικό roll off)*

Το μέτρο αυτό είναι χρήσιμο για την διάκριση της ομιλίας από την αφωνία. Η αφωνία έχει υψηλό ποσοστό της ενέργειας που περιέχεται στην περιοχή υψηλών συχνοτήτων στο φάσμα, όπου το μεγαλύτερο μέτρο της ενέργειας για την ομιλία k την μουσική περιέχεται σε χαμηλότερες ζώνες.

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \sum_{n=1}^N M_t[n]$$

Το φασματικό rolloff ορίζεται ως η συχνότητα R_t κάτω από την οποία συγκεντρώνεται το 85% της κατανομής του μεγέθους.¹⁰

iii. *Spectral Flux (φασματική ροή)*

Η φασματική ροή είναι ένα μέτρο που δείχνει το πόσο γρήγορα αλλάζει το φάσμα ισχύος ενός σήματος. Υπολογίζεται με την σύγκριση του φάσματος ισχύος για ένα πλαίσιο έναντι το φάσμα

⁹U. Nam, "Special Area Exam Part II", (April 2008), <https://crrma.stanford.edu/~unjung/AIR/areaExam.pdf>

¹⁰<http://www.egr.msu.edu/~aviyente/ece366pd09.pdf>

ισχύος του προηγούμενου πλαισίου. Η φασματική ροή μπορεί να χρησιμοποιηθεί για τον προσδιορισμό της χροιάς ενός ακουστικού σήματος, ή στην ανίχνευση έναρξης, μεταξύ άλλων.

$$F_t = \sum_{n=1}^N \left(N_t[n] - N_{t-1}[n] \right)^2$$

Όπου, $N_t[n]$ είναι το κανονικοποιημένο μέγεθος του μετασχηματισμού Fourier στο τρέχον πλαίσιο t .¹¹

c) Συντελεστές Mel Frequency Cepstral (MFCCs)

Οι συντελεστές MFC είναι κινητήρια χαρακτηριστικά βασισμένα στον μετασχηματισμό Fourier. Μετά την λήψη του μετασχηματισμού ενός παραθύρου ανάλυσης, το φάσμα μεγέθους διέρχεται μέσω μιας τράπεζας φίλτρου MEL, με μεταβαλλόμενο εύρος ζώνης που μιμείται το ανθρώπινο αυτί, δηλ. μικρό εύρος ζώνης σε χαμηλές συχνότητες κ μεγάλο εύρος ζώνης σε ψηλές συχνότητες.

Πρώτα η συχνότητα κλιμακώνεται χρησιμοποιώντας τράπεζα φίλτρου MELH(k,m) κ στην συνέχεια, λαμβάνεται ο λογάριθμος χρησιμοποιώντας :

$$X'(m) = \ln \sum_{k=0}^{N-1} \left| X(k) \right| H(k, m) \quad \text{για } m=1, 2, \dots$$

με M να είναι ο αριθμός των τραπεζών φίλτρου κ $M \ll N$ ο αριθμός των σημείων συχνοτήτων.

d) Short-time energy

Το πλάτος του σήματος ομιλίας μεταβάλλεται με τον χρόνο. Γενικά, το πλάτος των τμημάτων μη φωνητικού λόγου, είναι πολύ χαμηλότερο από το πλάτος των τμημάτων ομιλίας. Η ενέργεια του σήματος ομιλίας παρέχει μία παράσταση που αντικατοπτρίζει αυτές τις παραλλαγές πλάτους.

$$E_n = \sum_{m=-\infty}^{\infty} \left[x(m)w(n - m) \right]^2$$

¹¹D.Giannoulis, M. Massberg, and J. D. Reiss. "Parameter Automation in a Dynamic Range Compressor," 2013, Queen Mary University of London UK.

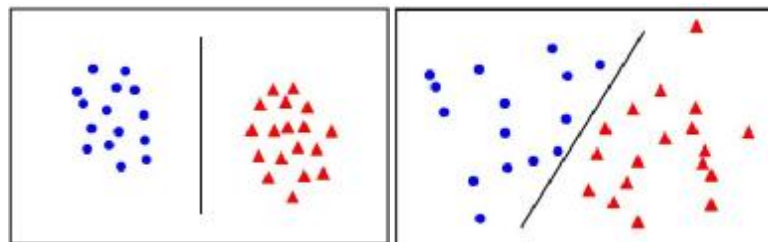
ΚΕΦΑΛΑΙΟ 4

4.1 Αλγόριθμοι κατηγοριοποίησης (Classification algorithms)

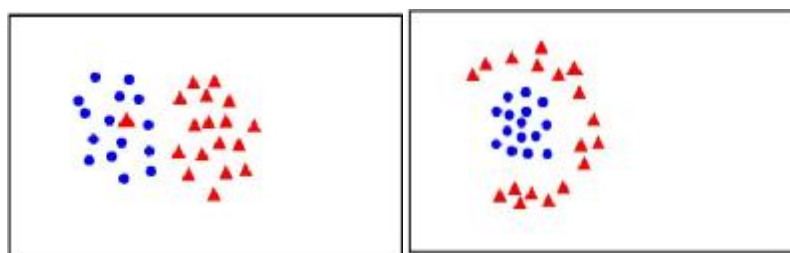
4.1.1 Αλγόριθμος υποστήριξης διανυσματικών μηχανών (SVM)

«Ο αλγόριθμος υποστήριξης διανυσματικών μηχανών χρησιμοποιεί την ανάμειξη γραμμικών μοντέλων βασισμένων σε στιγμιότυπα. Είναι ένας απλός ταξινομητής ο οποίος «εκπαιδεύει» το όριο απόφασης ανάμεσα σε δύο κλάσεις. Για να βρεθεί αυτό το όριο, μεγαλώνει την απόσταση ανάμεσα σε αυτές τις δύο κλάσεις, έχοντας επιλέξει γραμμικούς διαχωριστές.

Οι μηχανές αυτές προβάλλουν δεδομένα από τον χώρο εισόδου στον χώρο παραμέτρων, γραμμικά ή μη γραμμικά, χρησιμοποιώντας μία συνάρτηση πυρήνα τύπου Kernel. Έχοντας λοιπόν την δυνατότητα να παράγουν κ μη γραμμικές επιφάνειες απόφασης, τις καθιστούν πολύ σημαντικές στο να επιλύουν μεγάλο αριθμό προβλημάτων ταξινόμησης. Τέλος, χειρίζονται μεγάλου όγκου δεδομένα, οπότε είναι η κατάλληλη επιλογή για την επίλυση προβλημάτων που αφορούν κατηγοριοποίηση μουσικής πληροφορίας». ¹²



Εικόνα 6: Γραμμικός Διαχωρισμός



Εικόνα 7: Μη Γραμμικός Διαχωρισμός

¹²Ζαρδαβάς, Ιωάννης. «Αυτόματη κατηγοριοποίηση ειδών κρητικής μουσικής με χρήση μεθόδων μηχανικής μάθησης», Ανώτατο Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης, Τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Ρέθυμνο, 2013.

Σε όλα τα πειράματα, χρησιμοποιήθηκε η βιβλιοθήκη **LibSVM** του Weka, όπου έχει 2 στάδια: την εκπαίδευση των δεδομένων για την μάθηση των παραμέτρων του μοντέλου κ την πρόβλεψη πληροφοριών των υπό δοκιμή δεδομένων, με την χρήση του μοντέλου αυτού.

4.1.2 Αλγόριθμος Μπεϊσιανής κατηγοριοποίησης (NaïveBayes)

Από το θεώρημα του Bayes προκύπτει:

$$P(W_j | x) = \frac{P(x | W_j) \times P(W_j)}{P(x)}$$

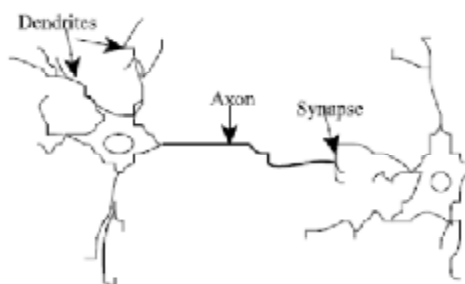
όπου W_j η πιθανότητα της κλάσης j , και x το διάνυσμα των χαρακτηριστικών.

Με βάση το παραπάνω μοντέλο μπορεί να εκτιμηθεί η πιθανότητα κάθε κατηγορίας της κατηγοριοποίησης. Δημιουργείται με την χρήση δεδομένων εκπαίδευσης, αξιοποιώντας τα διανύσματα που αναπαριστούν τα δεδομένα επεξεργασίας. Πειράματα που έχουν πραγματοποιηθεί με την χρήση των διαφόρων δεδομένων κειμένου, απέδειξαν ότι ο παραπάνω αλγόριθμος είναι ιδιαίτερα αποτελεσματικός για πολλές διαδικασίες κατηγοριοποίησης, κυρίως αναγνώρισης ομιλίας.

4.1.3 Αλγόριθμοι τεχνητών νευρωνικών δικτύων (ArtificialNeuralNetworks- ANN)

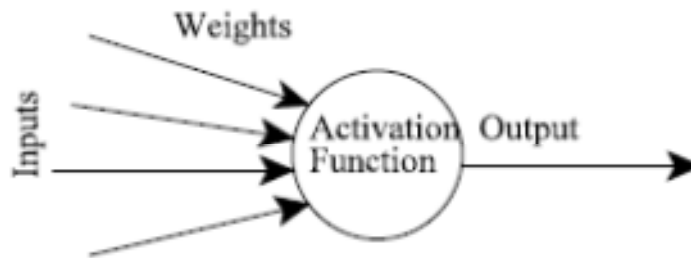
Τεχνητό νευρωνικό δίκτυο, είναι το δίκτυο που βλέπει τους κόμβους ως «τεχνητούς νευρώνες». Ένας τεχνητός νευρώνας είναι ένα υπολογιστικό μοντέλο εμπνευσμένο από τους φυσικούς νευρώνες

Η πολυπλοκότητα των πραγματικών νευρώνων αντλείται σε υψηλό βαθμό όταν μοντελοποιούνται οι τεχνητοί νευρώνες. Αυτές αποτελούνται βασικά από εισόδους, οι οποίες πολλαπλασιάζονται με βάρη, κ στην συνέχεια υπολογίζεται μία μαθηματική συνάρτηση, η οποία καθορίζει την ενεργοποίηση του νευρώνα. Μία άλλη λειτουργία υπολογίζει την έξοδο του τεχνητού νευρώνα. Τα ANN συνδυάζουν τεχνητούς νευρώνες, προκειμένου να επεξεργάζονται πληροφορίες.



Εικόνα 8:Φυσικοί νευρώνες

Όσο μεγαλύτερο είναι το βάρος ενός τεχνητού νευρώνα, τόσο ισχυρότερη είναι η είσοδος με την οποία θα πολλαπλασιαστεί. Με την προσαρμογή των βαρών ενός ANN, μπορούμε να πετύχουμε την έξοδο που θέλουμε για συγκεκριμένες εισόδους. Επειδή, όμως, είναι περίπλοκο να βρει κανείς όλα τα απαραίτητα βάρη όταν έχουμε ένα ANN με εκατοντάδες ή χιλιάδες νευρώνες, μπορούμε να βρούμε αλγόριθμους που μπορούν να προσαρμόσουν τα βάρη, προκειμένου να αποκτήσουν την επιθυμητή έξοδο από το δίκτυο. Αυτή η διαδικασία της προσαρμογής των σταθμίσεων ονομάζεται μάθηση ή εκπαίδευση.¹³



Εικόνα 9: Τεχνητός νευρώνας

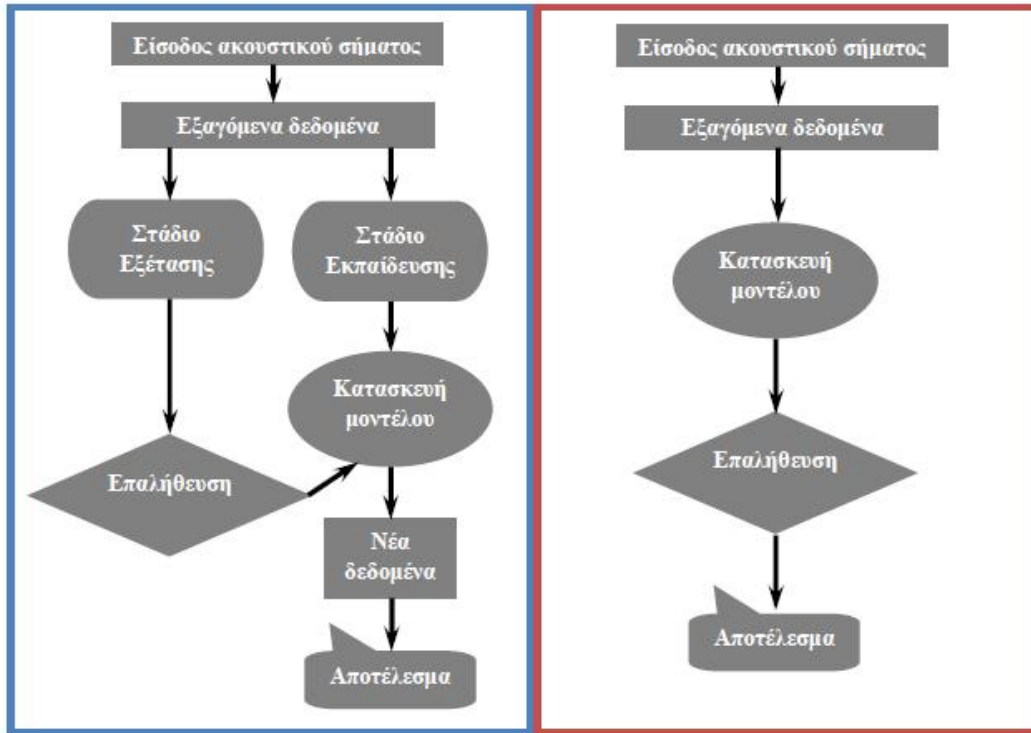
4.1.4 Κατηγοριοποίηση με κ χωρίς επιτήρηση

«Στην κατηγοριοποίηση με επίβλεψη, κάθε παράδειγμα εκπαίδευσης αντιστοιχεί σε ένα διάνυσμα εισόδου $\{x_1, x_2, \dots, x_n\} = y$, όπου x είναι το σύνολο χαρακτηριστικών και y είναι η τιμή της κλάσης εξόδου, η οποία περιγράφει ένα συγκεκριμένο γεγονός για την έννοια στόχο. Για την εκπαίδευση συσχετισμένων εισόδων και εξόδων, πρέπει να παραχθεί μία συνάρτηση, που υποδηλώνει την αντικειμενική σχέση ανάμεσα στις τιμές εισόδου και εξόδου, και αυτή η συνάρτηση να μπορεί να προβλέψει την έξοδο για άγνωστες εισόδους. Στην κατηγοριοποίηση με επίβλεψη ανήκουν τα νευρωνικά δίκτυα και η υποστήριξη διανυσματικών μηχανών.

Στην κατηγοριοποίηση χωρίς επίβλεψη δεν υπάρχει προκαθορισμένο σύνολο τιμών. Τα στοιχεία που βρίσκονται για εκπαίδευση χωρίζονται σε άγνωστες ομάδες βάσει των χαρακτηριστικών τους. Έπειτα, εξετάζονται οι σχέσεις ανάμεσα σε ένα σύνολο χαρακτηριστικών. Σε αυτή την κατηγορία χρησιμοποιείται ο αλγόριθμος που βασίζεται στα Μπείσιανά μοντέλα».¹⁴

¹³Gershenson, Carlos. “Artificial Neural Networks for Beginners,” 2003.

¹⁴Ι. Ζαρδαβάς, ό.π.



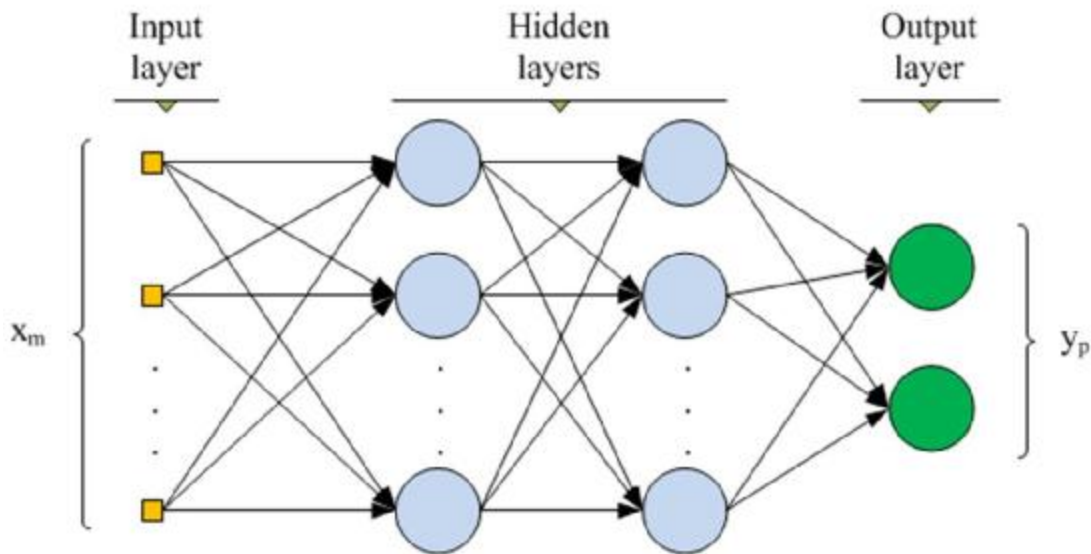
Εικόνα 10: Κατηγοριοποίηση με κ χωρίς επίβλεψη

4.1.5 Multilayer Perceptron

«Τα MLP έχουν χρησιμοποιηθεί σε πάρα πολλά και μάλιστα δύσκολα προβλήματα μάθησης με επίβλεψη, σε ταξινομήσεις και προσεγγίσεις συναρτήσεων, με μεγάλη επιτυχία. Εκπαιδεύονται συνήθως χρησιμοποιώντας τον αλγόριθμο οπισθοδιάδοσης του σφάλματος (error backpropagation) που ουσιαστικά εφαρμόζει τη μέθοδο (βελτιστοποίησης) της καθόδου με βάση την κλίση (gradient descent) για την ελαχιστοποίηση του τετραγωνικού σφάλματος εκπαίδευσης. Η διαδικασία της εκπαίδευσης του αλγορίθμου οπισθοδιάδοσης του σφάλματος αποτελείται από υπολογισμούς που γίνονται σε δύο περάσματα από τα επίπεδα του MLP.

Το πρώτο πέραςμα γίνεται κατά την ευθεία φορά, δηλαδή από την είσοδο προς την έξοδο, και το δεύτερο ανάποδα, από την έξοδο προς την είσοδο. Έτσι το MLP μπορεί να επιλύσει προβλήματα που δεν είναι γραμμικά διαχωρίσιμα αντίθετα με το perceptron που επιλύει μόνο γραμμικά διαχωρίσιμα προβλήματα. Ουσιαστικά το MLP στις περιπτώσεις ταξινόμησης ορίζει υπερεπίπεδα στο χώρο των δεδομένων, δημιουργώντας περιοχές απόφασης από τις τομές αυτών των υπερεπιπέδων».¹⁵

¹⁵ Κ. Κυρίτσης, «Νευρωνικά Δίκτυα και Μηχανές Διανυσματικής Υποστήριξης», Σχολή Θετικών Επιστημών – Τμήμα Μαθηματικών, Πανεπιστήμιο Πατρών (Μάιος 2014).



Εικόνα 11: Επίπεδα MLP

4.1.6 Logistic model tree (LMT)

Στην επιστήμη των υπολογιστών, ένα λογιστικό μοντέλο δέντρου (LMT) είναι ένα μοντέλο κατάταξης, με ένα σχετικό αλγόριθμο εκπαίδευσης με επίβλεψη, που συνδυάζει την λογιστική παλινδρόμηση (LR) και την εκμάθηση δέντρων απόφασης.

Ο βασικός αλγόριθμος επαγωγής LMT χρησιμοποιεί διασταυρωμένη επικύρωση για να βρεθεί μια σειρά από LogitBoost επαναλήψεις που δεν συλλαμβάνει τον θόρυβο στα δεδομένα εκπαίδευσης.

ΚΕΦΑΛΑΙΟ 5 : Data Analysis

Στο κεφάλαιο αυτό, θα εξηγήσουμε τους όρους που συναντάμε στα αποτελέσματα του πειράματος μας.

5.1 Πίνακας Σύγχυσης (Confusion Matrix)

Στον τομέα της μηχανικής μάθησης και συγκεκριμένα το πρόβλημα της στατιστικής ταξινόμησης, το confusionmatrix, επίσης γνωστό ως errormatrix, είναι ένα ειδικό σχεδιάγραμμα που επιτρέπει την απεικόνιση της απόδοσης ενός αλγορίθμου, συνήθως ενός με επιβλεπόμενη εκμάθηση.

Κάθε στήλη του πίνακα αντιπροσωπεύει τις περιπτώσεις σε μια προβλεπόμενη κατηγορία, ενώ κάθε σειρά αντιπροσωπεύει τις περιπτώσεις σε μια πραγματική κατηγορία (ή το αντίστροφο). Το όνομα προέρχεται από το γεγονός ότι το καθιστά εύκολο να δούμε αν το σύστημα συγχέει δύο.

5.2 StratifiedCross – Validation

Σε ένα πρόβλημα πρόβλεψης δίνεται συνήθως σε ένα μοντέλο, ένα γνωστό σύνολο δεδομένων, για να γίνει η εκπαίδευση των δεδομένων. Έπειτα, στο ίδιο μοντέλο δίνεται ένα άγνωστο σύνολο δεδομένων για να γίνει η δοκιμή των δεδομένων.

Ο στόχος του crossvalidation είναι (1) να καθορίσει ένα σύνολο δεδομένων για την δοκιμή του μοντέλου στην φάση της εκπαίδευσης, προκειμένου να περιοριστούν τα προβλήματα, όπως η υπερπροσαρμογή, κ (2) να δώσει μία ιδέα για το πώς το μοντέλο θα γενικευθεί σε ένα ανεξάρτητο σύνολο δεδομένων.

Ένας γύρος του crossvalidation περιλαμβάνει τον καταμερισμό των δειγμάτων των δεδομένων σε συμπληρωματικά υποσύνολα, εκτελώντας την ανάλυση σε ένα υποσύνολο κ επικυρώνοντας την ανάλυση στο άλλο υποσύνολο. Για να μειωθούν οι διακυμάνσεις, πολλοί γύροι εκτελούνται χρησιμοποιώντας διαφορετικά διαμερίσματα, κ βγαίνει ο μέσος όρος ανά γύρο για τα αποτελέσματα επικύρωσης.

Συνοπτικά, το crossvalidation συνδυάζει μέτρα προσαρμογής, για να αποκομίσει μία πιο ακριβή εκτίμηση, της απόδοσης του μοντέλου πρόβλεψης.

5.2.1 Kappa – statistics

Ο συντελεστής kappa είναι ένα στατιστικό μέτρο για ποιοτικά στοιχεία, σχετικά με την συμφωνία μεταξύ των εκτιμητών ή των σχολιαστών. Είναι ένα ισχυρό μέτρο σε σύγκριση με ένα απλό υπολογισμό ποσοστού συμφωνίας, αφού λαμβάνει υπόψη κ συμφωνίες που προκύπτουν τυχαία.

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

Όπου, $P(A)$ είναι το ποσοστό συμφωνίας (για παράδειγμα μεταξύ του ταξινομητή k της απόλυτης αλήθειας), $P(E)$ είναι η τυχαία συμφωνία, το $K=1$ υποδηλώνει την τέλεια συμφωνία k το $K=0$, υποδηλώνει την τυχαία συμφωνία.¹⁶

5.2.2 Μέσο απόλυτο σφάλμα (MAE)

Το MAE μετρά το μέσο εύρος των σφαλμάτων σε μια σειρά από προβλέψεις, χωρίς να λαμβάνεται υπόψη η διεύθυνσή τους. Μετρά την ακρίβεια για τις συνεχείς μεταβλητές.

Εκφρασμένο με λόγια, το MAE είναι ο μέσος όρος του δείγματος επαλήθευσης των απόλυτων τιμών, των διαφορών μεταξύ των προβλέψεων και της αντίστοιχης παρατήρησης. Είναι ένα γραμμικό αποτέλεσμα, το οποίο σημαίνει ότι όλες οι ατομικές διαφορές έχουν την ίδια βαρύτητα στη μέση τιμή.

5.2.3 Ρίζα του μέσου τετραγωνικού σφάλματος (RMSE)

Το RMSE είναι ένας τετραγωνικός κανόνας βαθμολόγησης που μετρά το μέσο εύρος του σφάλματος. Εκφρασμένο σε λέξεις, η διαφορά μεταξύ των προβλέψεων και των αντίστοιχων παρατηρούμενων τιμών τετραγωνίζονται και, στη συνέχεια, βρίσκεται ο μέσος όρος του δείγματος.

Τέλος, λαμβάνεται η τετραγωνική ρίζα του μέσου όρου. Δεδομένου ότι τα σφάλματα τετραγωνίστηκαν για να βρεθεί ο μέσος όρος τους, το RMSE δίνει ένα σχετικά υψηλό βάρος σε μεγάλα σφάλματα. Αυτό σημαίνει ότι το RMSE είναι πιο χρήσιμο όταν μεγάλα σφάλματα είναι ιδιαίτερα ανεπιθύμητα.

Το MAE και το RMSE μπορούν να χρησιμοποιηθούν μαζί για να διαγνωστεί η διακύμανση των σφαλμάτων σε μια σειρά προβλέψεων. Το RMSE θα είναι πάντα μεγαλύτερο ή ίσο με το MAE, όσο μεγαλύτερη η διαφορά μεταξύ τους, τόσο μεγαλύτερη η διακύμανση στα μεμονωμένα σφάλματα στο δείγμα. Εάν το $RMSE = MAE$, τότε όλα τα σφάλματα είναι του ίδιου εύρους.

5.2.4 Απόλυτο Σχετικό Σφάλμα (Relative absolute error (Rae))

Παίρνει το συνολικό απόλυτο σφάλμα k το ομαλοποιεί, διαιρώντας το με το συνολικό απόλυτο σφάλμα της απλής πρόβλεψης.

5.2.5 Root Relative Squared Error (Rrse)

Με τη λήψη της τετραγωνικής ρίζας του σχετικού τετραγωνισμένου σφάλματος, μειώνει το σφάλμα στις ίδιες διαστάσεις, όπως η ποσότητα που είχε προβλεφθεί.

¹⁶B.Eugenio Di, M. Glass, "The Kappa statistic: a second look," 2004.

5.3 Ακρίβεια ταξινομητή ανά κατηγορία

5.3.1 TruePositiverate

Είναι το ποσοστό των παραδειγμάτων που ταξινομήθηκε ως κατηγορία x, μεταξύ όλων των παραδειγμάτων που είναι πραγματικά στην κατηγορία x, δηλ. πόσο μέρος της κατηγορίας κατακτήθηκε. Είναι ισοδύναμο με την ανάκληση. Στον confusionmatrix είναι το διαγώνιο στοιχείο διαιρούμενο από το άθροισμα της σχετικής σειράς.

5.3.2 FalsePositiverate

Είναι το ποσοστό των παραδειγμάτων που είχαν ταξινομηθεί ως κατηγορία x, αλλά ανήκουν σε διαφορετική κατηγορία, ανάμεσα σε όλα τα παραδείγματα που δεν είναι της κατηγορίας x. Στο confusionmatrix, είναι το άθροισμα της στήλης της κατηγορίας x, μείον το διαγώνιο στοιχείο, διαιρούμενο από το άθροισμα των σειρών όλων των άλλων κλάσεων.

5.3.3 Ακρίβεια ταξινομητή (Precision)

Η ακρίβεια είναι ο αριθμός των αληθώς θετικών διαιρούμενος από τον αριθμό των πραγματικά θετικών και των εσφαλμένων θετικών. Με άλλα λόγια, είναι ο αριθμός των θετικών προβλέψεων διαιρούμενος από τον συνολικό αριθμό των θετικών τιμών της προβλεπόμενης κατηγορίας.

Η ακρίβεια μπορεί να θεωρηθεί ως το μέτρο της ακρίβειας των ταξινομητών. Μία χαμηλή ακρίβεια μπορεί να δείξει μεγάλο αριθμό των ψευδώς θετικών.

5.3.4 Ανάκληση ταξινομητή (Recall)

Ανάκληση είναι ο αριθμός των αληθώς θετικών διαιρούμενος από τον αριθμό των αληθώς θετικών και τον αριθμό των ψευδώς αρνητικών. Με άλλα λόγια, είναι ο αριθμός των θετικών προβλέψεων διαιρούμενος από τον αριθμό των θετικών τιμών της κατηγορίας στα δεδομένα δοκιμής. Σε δυαδική ταξινόμηση, η ανάκληση ονομάζεται κ ευαισθησία..

Η ανάκληση μπορεί να θεωρηθεί ως μέτρο της πληρότητας των ταξινομητών. Μια χαμηλή ανάκληση δείχνει πολλά ψευδώς αρνητικά αποτελέσματα.

Ανάκληση στην ανάκτηση πληροφοριών, είναι το κλάσμα των εγγράφων που σχετίζονται με το ερώτημα, που ανακτάται με επιτυχία.

$$recall = \frac{|{\text{relevant_documents}} \cap {\text{retrieved_documents}}|}{|{\text{retrieved_documents}}|}$$

Για παράδειγμα, για την αναζήτηση κειμένου σε ένα σύνολο εγγράφων, ανάκληση είναι ο αριθμός των σωστών αποτελεσμάτων διαιρούμενος από τον αριθμό των αποτελεσμάτων που θα έπρεπε να έχουν επιστραφεί.

Είναι τετριμμένο να επιτευχθεί η ανάκληση 100% επιστρέφοντας όλα τα έγγραφα ως απάντηση σε κάθε ερώτημα. Ως εκ τούτου, η ανάκληση από μόνη της δεν είναι αρκετή, αλλά κάποιος πρέπει να μετρήσει τον αριθμό των μη-σχετικών εγγράφων επίσης, για παράδειγμα με υπολογισμό της ακρίβειας.

Στην αναγνώριση προτύπων και ανάκτησης πληροφοριών με δυαδική ταξινόμηση, η *ακρίβεια* είναι το κλάσμα των ανακτημένων περιπτώσεων που είναι σχετικές, ενώ η *ανάκληση* είναι το κλάσμα των σχετικών περιπτώσεων που ανακτώνται. Τόσο η ακρίβεια και η ανάκληση βασίζονται στην κατανόηση και το μέτρο της συνάφειας.

5.3.5 F-measure

Ένα μέτρο που συνδυάζει την ακρίβεια και την ανάκληση. Είναι ο αρμονικός μέσος όρος της ακρίβειας και της ανάκλησης.

$$F = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Το μέτρο αυτό είναι κατά προσέγγιση ο μέσος όρος των δύο όταν είναι κοντά, και είναι γενικότερα ο αρμονικός μέσος όρος, ο οποίος, για την περίπτωση δύο αριθμών, συμπίπτει με το τετράγωνο του γεωμετρικού μέσου όρου διαιρούμενος από τον αριθμητικό μέσο όρο.

5.3.6 Συντελεστής συσχέτισης Matthews (Matthews correlation coefficient – MCC)

Χρησιμοποιείται στην μηχανική μάθηση ως μέτρο της ποιότητας των δυαδικών ταξινομήσεων (δύο κατηγοριών). Λαμβάνει υπόψη τα αληθινά κ τα ψευδή θετικά κ αρνητικά κ θεωρείται γενικά ως ένα ισορροπημένο μέτρο κ μπορεί να χρησιμοποιηθεί ακόμα κ αν οι δύο κατηγορίες έχουν πολύ διαφορετικό μέγεθος.

$$|MCC| = \sqrt{\frac{x^2}{n}}$$

όπου n , ο συνολικός αριθμός παρατηρήσεων.

Ο MCC είναι στην ουσία ένας συντελεστής συσχέτισης μεταξύ των παρατηρούμενων και των προβλεπόμενων δυαδικών ταξινομητών κ επιστρέφει μια τιμή μεταξύ -1 και +1. Η τιμή 1 αποτελεί μια τέλεια πρόβλεψη, η τιμή 0 δείχνει τίποτα καλύτερο από ό, τι μία τυχαία πρόβλεψη και η τιμή -1 δείχνει συνολική διαφωνία μεταξύ της πρόβλεψης και παρατήρησης.

Αν και δεν υπάρχει τέλειος τρόπος περιγραφής του confusionmatrix των αληθών κ των ψευδών θετικών κ αρνητικών από έναν αριθμό, ο MCC θεωρείται γενικά ως ένα από τα καλύτερα τέτοια μέτρα.

Ο MCC μπορεί να υπολογιστεί απευθείας από τον confusionmatrix χρησιμοποιώντας τον εξής τύπο:

$$MCC = \frac{TP' TN - FP' FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

TP, Ο αριθμός των αληθώς θετικών

FP, Ο αριθμός των ψευδώς θετικών

TN, Ο αριθμός των αληθώς αρνητικών

FN, Ο αριθμός των ψευδώς αρνητικών

Εάν κάποιο από τα τέσσερα ποσά στον παρονομαστή είναι μηδέν, ο παρονομαστής μπορεί να ρυθμιστεί αυθαίρετα σε ένα, αυτό έχει ως αποτέλεσμα ένα MCC μηδέν, ο οποίος μπορεί να αποδειχθεί ότι είναι η σωστή οριακή τιμή.

5.3.7 AUC- AreaUnderCurves

Είναι η περιοχή που συνοψίζει την καμπύλη ROC . Όσο υψηλότερη είναι τόσο καλύτερη είναι η ιεραρχική απόδοση του ταξινομητή. Μπορεί να ερμηνευτεί ως η πιθανότητα ότι ένα τυχαία επιλεγμένο θετικό παράδειγμα θα καταταχθεί υψηλότερα, από ένα τυχαία επιλεγμένο αρνητικό παράδειγμα.

- Receiver operating characteristic - ROC curves

Στη στατιστική, είναι ένα γραφικό διάγραμμα που απεικονίζει την απόδοση ενός ταξινομητή δυαδικού συστήματος ως ποικίλο κατώφλι διάκρισης. Η καμπύλη δημιουργείται με την γραφική αναπαράσταση του TPrate έναντι του FPrate σε διάφορες ρυθμίσεις του κατωφλιού. Γενικά, αν οι κατανομές των πιθανοτήτων είναι γνωστές τόσο για την ανίχνευση όσο κ για τον ψευδή συναγερμό, η καμπύλη αυτή μπορεί να δημιουργηθεί με γραφική αναπαράσταση της αθροιστικής συνάρτησης κατανομής, της πιθανότητας ανίχνευσης στον άξονα Y, σε σχέση με την αθροιστική συνάρτηση κατανομής της πιθανότητας ψευδών συναγερμών στον άξονα X.

· Precision-Recall curves - PR curves

Η ανάκληση είναι ισοδύναμη με το πραγματικό θετικό ποσοστό ή την ευαισθησία (ο y-άξονας στις καμπύλες ROC), ενώ η ακρίβεια είναι το ίδιο με θετική προγνωστική αξία.

Αν το παράδειγμα με τη μεγαλύτερη τιμή εξόδου είναι θετικό ή αρνητικό, μεταβάλλει σε μεγάλο βαθμό την καμπύλη PR (πλησιάζοντας (0, 1) εάν είναι θετική κ (0, 0) εάν είναι αρνητική). Ομοίως, οι εκτιμήσεις της ακρίβειας για ανάκληση κοντά στο 0 τείνουν να έχουν υψηλή διακύμανση, και αυτό είναι μια σημαντική δυσκολία στην κατασκευή των καμπυλών PR.

Μια καμπύλη PR μπορεί να οριστεί ως το σύνολο των σημείων:

$$PR(.) = \{ (Recall(c), Prec(c)), -\infty < c < \infty \}$$

Όπου, $Recall(c) = P(Y > c)$ και $Prec(c) = P(D = 1 | Z > c)$

ROC vs PR curves

Υπάρχει μία πολύ σημαντική διαφορά στο τι αντιπροσωπεύει η καμπύλη ROC έναντι της καμπύλης PR.

Μία καμπύλη ROC αντιπροσωπεύει μια σχέση μεταξύ της **ευαισθησίας** και **εξειδίκευσης**. Η ευαισθησία είναι το άλλο όνομα για την **ανάκληση**, αλλά η εξειδίκευση δεν είναι το ίδιο με την ακρίβεια.

Ανάκληση / ευαισθησία είναι ένα μέτρο του πόσο πολλά από τα θετικά δείγματα έχουν ταυτοποιηθεί ως θετικά.

Εξειδίκευση είναι το μέτρο του πόσο πολλά από τα αρνητικά δείγματα έχουν ταυτοποιηθεί ως αρνητικά.

Η **ακρίβεια** από την άλλη πλευρά είναι διαφορετική. Είναι ένα μέτρο του πόσο πολλά από τα δείγματα που προβλέπεται από το ταξινομητή ως θετικά είναι πράγματι θετικά.

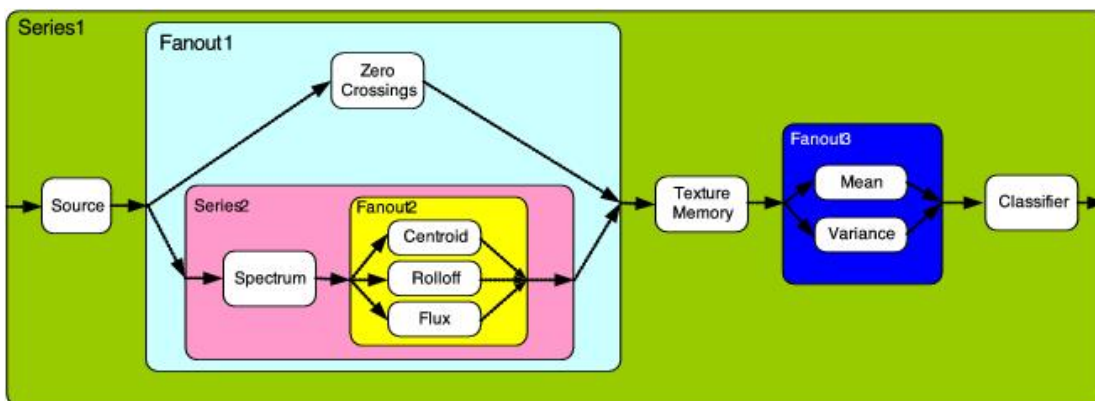
Έτσι, εάν το πρόβλημά μας αφορά το είδος της αναζήτησης, για παράδειγμα: τα θετικά δείγματα κατηγορίας είναι πολύ σπάνια σε σχέση με τα αρνητικά δείγματα κατηγορίας, χρησιμοποιούμε μια καμπύλη PR. Διαφορετικά, χρησιμοποιούμε μια καμπύλη ROC.

ΚΕΦΑΛΑΙΟ 6 : ΠΕΙΡΑΜΑΤΙΚΗ ΔΙΑΔΙΚΑΣΙΑ

6.1 Προτεινόμενη Μεθοδολογία

Όπως προαναφέρθηκε στην εισαγωγή, για την αναγνώριση ενός ήχου χρησιμοποιούνται τρεις μέθοδοι πρώτα: *Ανίχνευση* (Detection), *Κατάτμηση* (Segmentation) και τέλος *Τμηματοποίηση* (Clustering) του ήχου.

Η μεθοδολογία που χρησιμοποιήθηκε βασίζεται στο μοντέλο εξαγωγής ηχητικών χαρακτηριστικών της εφαρμογής ανοιχτού κώδικα MARSYAS. Με χρήση διαφόρων μετασχηματισμών μπορούμε να κάνουμε εξαγωγή χαρακτηριστικών ηχητικών δεδομένων, μέσα από την γραμμή εντολών. Αυτά τα χαρακτηριστικά αποθηκεύονται σε μορφή κειμένου. Γίνεται εισαγωγή στο περιβάλλον του WEKA, όπου υπάρχει η δυνατότητα ταξινόμησης του ήχου με χρήση αλγορίθμων εκπαίδευσης. Αναλυτικά η διαδικασία γίνεται με βάση την αναφορά του *Tzenetakis – MarsyasSubmissionstoMirex2009*.



Εικόνα 22: Διάγραμμα Συστήματος Εξαγωγής Χαρακτηριστικών Marsyas

6.1.1 Εξαγόμενα Χαρακτηριστικά

Τα εξαγόμενα χαρακτηριστικά που επιλέχθηκαν για το πείραμα είναι τα εξής:

STFTMFCC: Spectral Centroid, Spectral Roll-Off, Spectral Flux, Zero Crossing Rate, Mel Frequency Cepstral Coefficients.

Όλα τα παραπάνω χαρακτηριστικά έχουν περιγραφεί αναλυτικά στο **κεφάλαιο 2**.

- Bextract -svmusic.mfspeech.mf -w AllFeatures
- Bextract -sv -mfccmusic.mfspeech.mf -w mfcc
- Bextract -sv -zrcrmusic.mfspeech.mf -w ZeroCrossings
- Bextract -sv -spfemusic.mfspeech.mf -w SpectralFeatures
- Bextract -sv -flxmusic.mfspeech.mf -w Flux
- Bextract -sv -ctdmusic.mfspeech.mf -w SpectralCentroid
- Bextract -sv -rlfmusic.mfspeech.mf -w RollOff

Οι παραπάνω εντολές, έδωσαν τα αρχεία κειμένου, που εισήχθησαν αργότερα στο πρόγραμμα Weka, όπου έγινε κατηγοριοποίηση των κομματιών, με χρήση των εξής ταξινομητών:

- LMT
- LibSVM
- Naïve Bayes
- Multilayer Perceptron

Οι ταξινομητές περιγράφηκαν αναλυτικά στο **κεφάλαιο 3**.

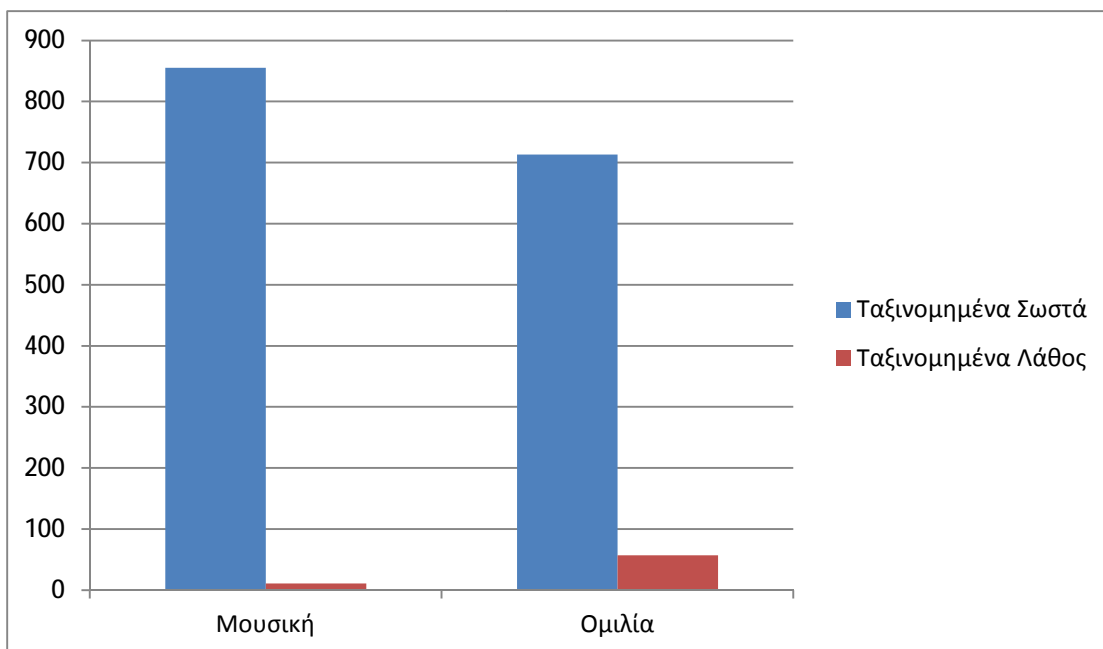
6.1.2 Βάση Δεδομένων

Τα ηχητικά δεδομένα που χρησιμοποιήσα προέρχονται από ραδιοφωνικές εκπομπές του BBC Radio, στις οποίες έγινε χρονικός καταμερισμός και διαχωρισμός σε τμήματα μουσικής και ομιλίας. Το format των αρχείων είναι τύπου .wav, το οποίο υποστηρίζεται κατάλληλα από τα προγράμματα εξαγωγής παραμέτρων MARSYAS και αναγνώρισης προτύπων WEKA. Συγκεκριμένα, για την κατηγορία της μουσικής χρησιμοποιήθηκαν 866 αρχεία τύπου .wav, διάρκειας 2.5 λεπτών το κάθε ένα κατά μέσο όρο. Για την ομιλία τα .wav αρχεία ήταν συνολικά 770, διάρκειας 2 λεπτών κατά μέσο όρο. Συνολικά, η βάση δεδομένων που χρησιμοποιήθηκε περιλαμβάνει 1650 ηχητικά αρχεία.

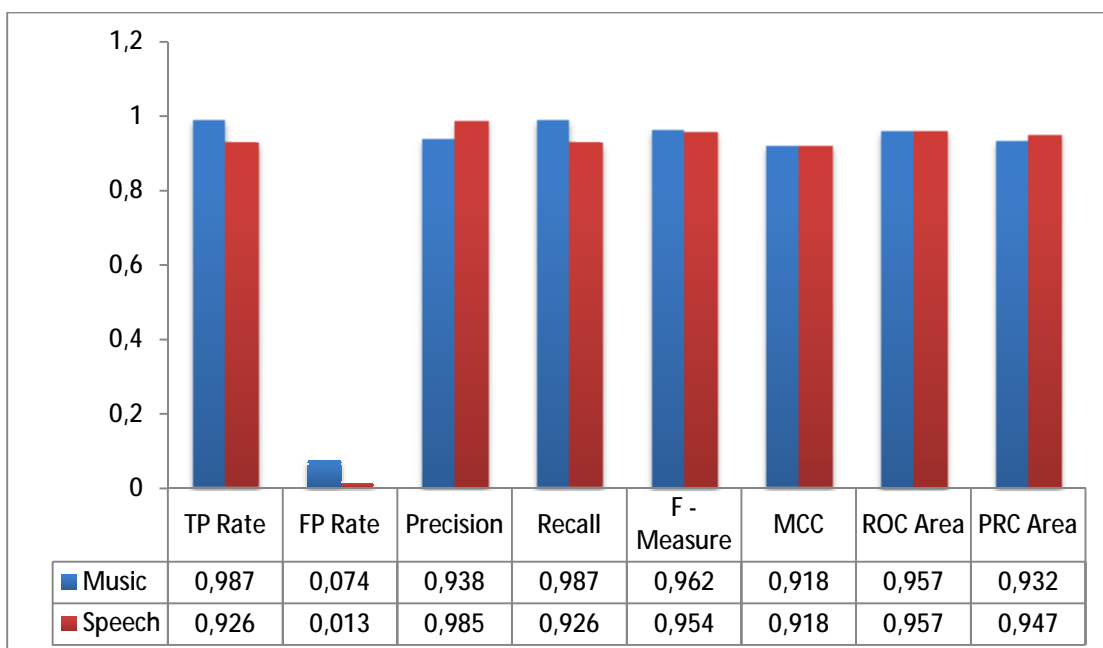
6.2 Αποτελέσματα

Στις ενότητες που ακολουθούν, παρουσιάζονται τα πειραματικά αποτελέσματα αναγνώρισης των ηχητικών κομματιών μουσικής και ομιλίας με χρήση διαφορετικών τύπων ταξινομητών όπως αναφέρθηκαν σε προηγούμενη ενότητα και διαφορετικά σετ ακουστικών παραμέτρων.

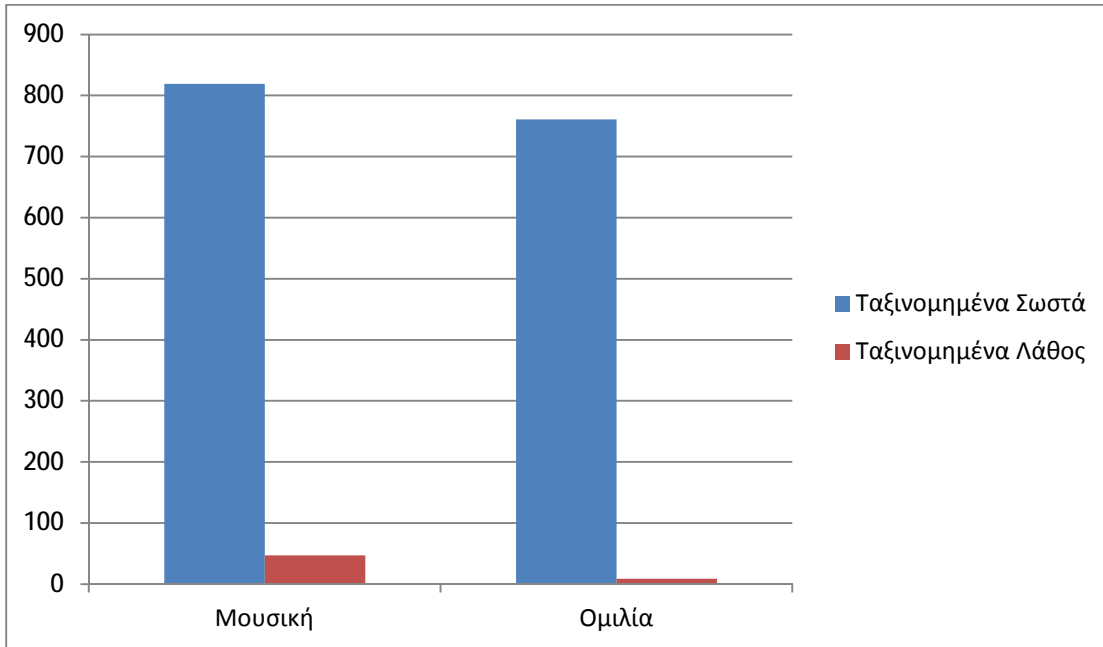
6.2.1 Ταξινομητής LibSVM



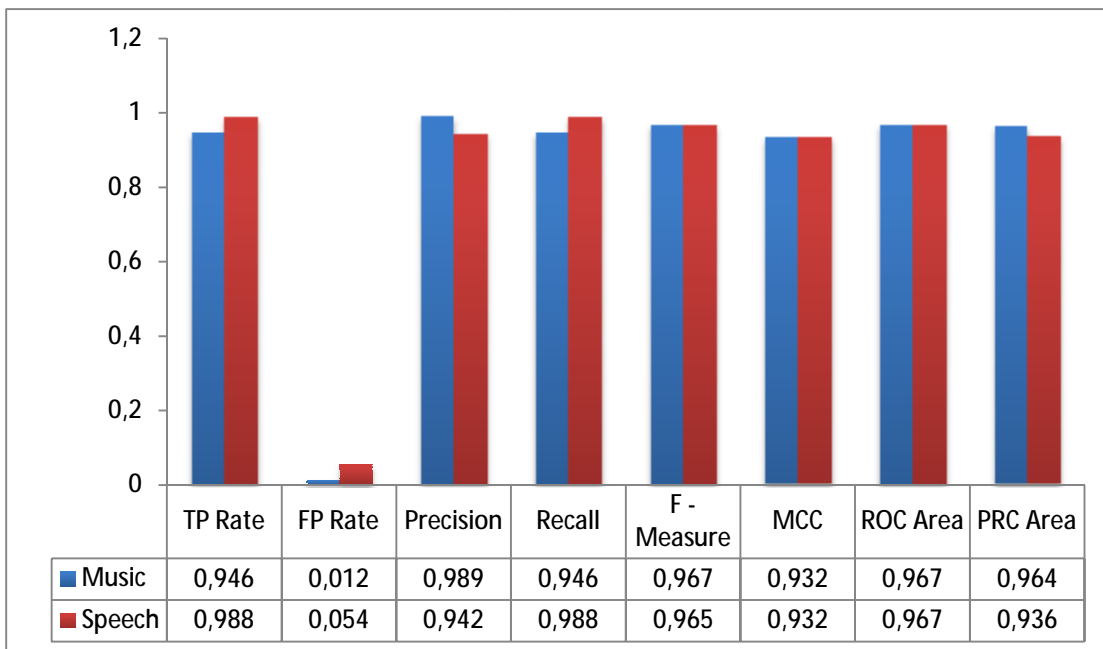
Πίνακας 1: Confusion Matrix All Features



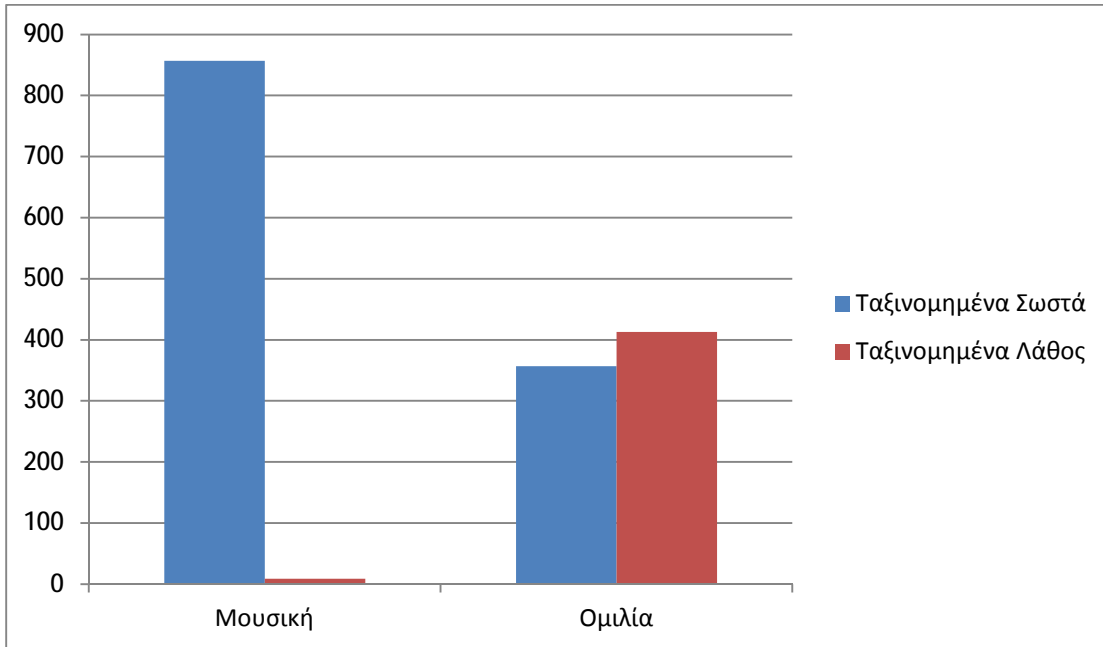
Πίνακας 2: Data All Features



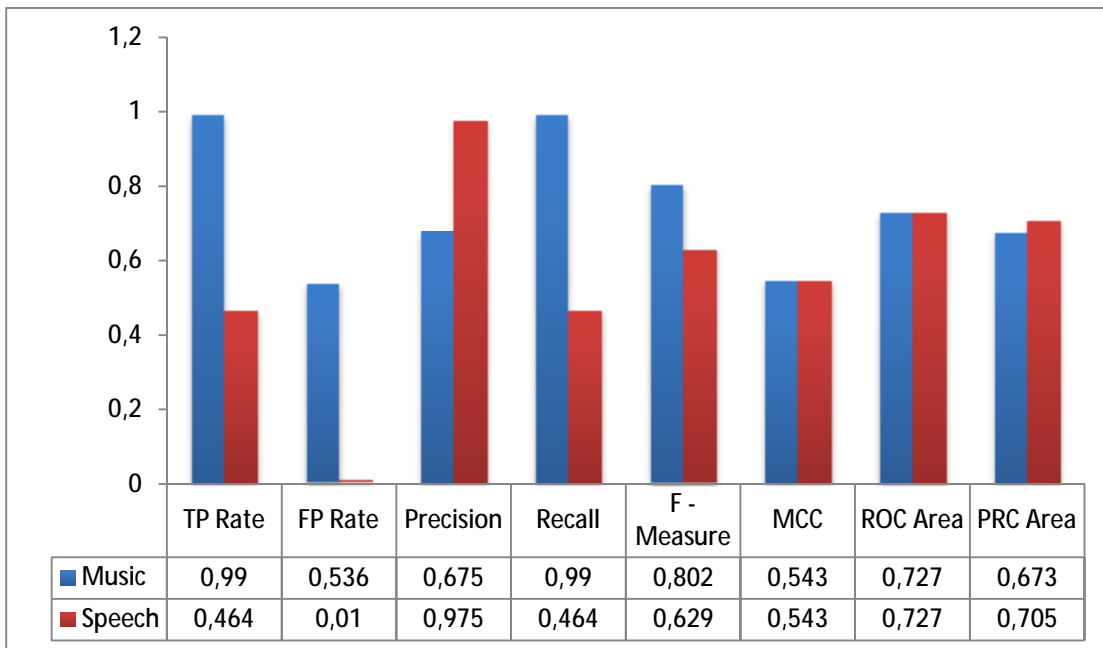
Πίνακας 3: Confusion Matrix Spectral Features



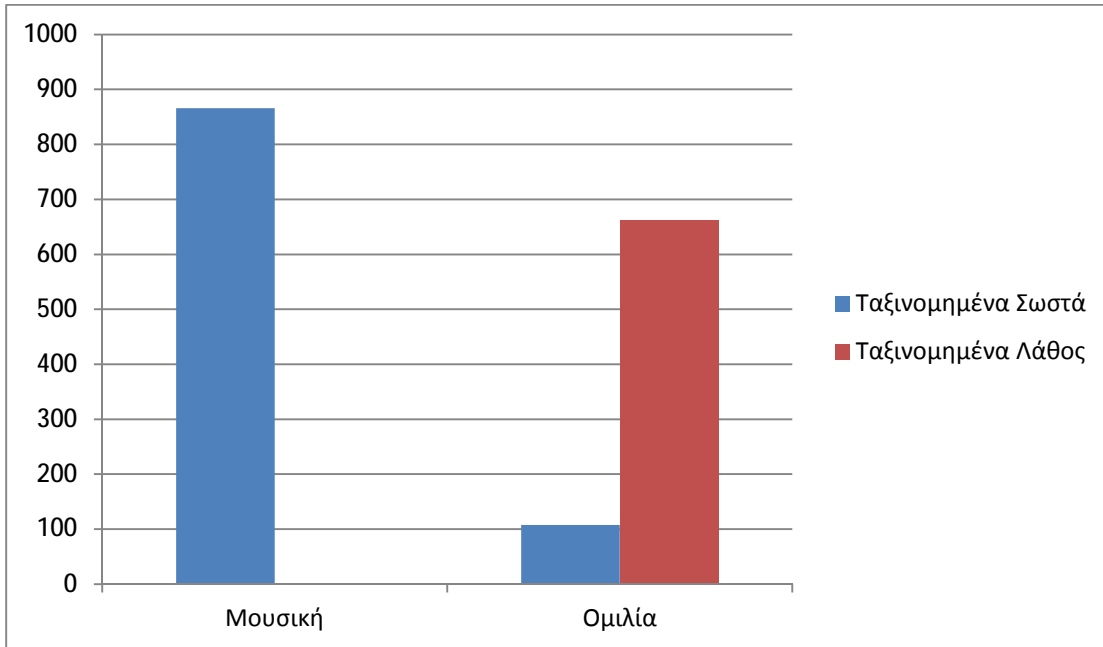
Πίνακας 4: Data Spectral Features



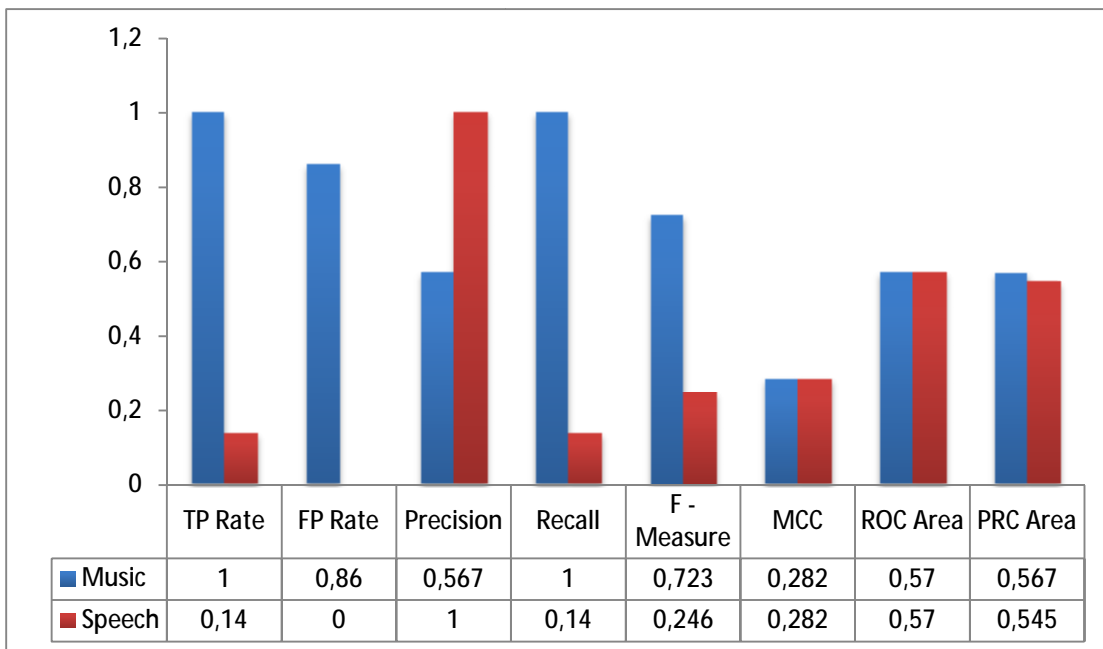
Πίνακας 5:Confusion Matrix Spectral Centroid



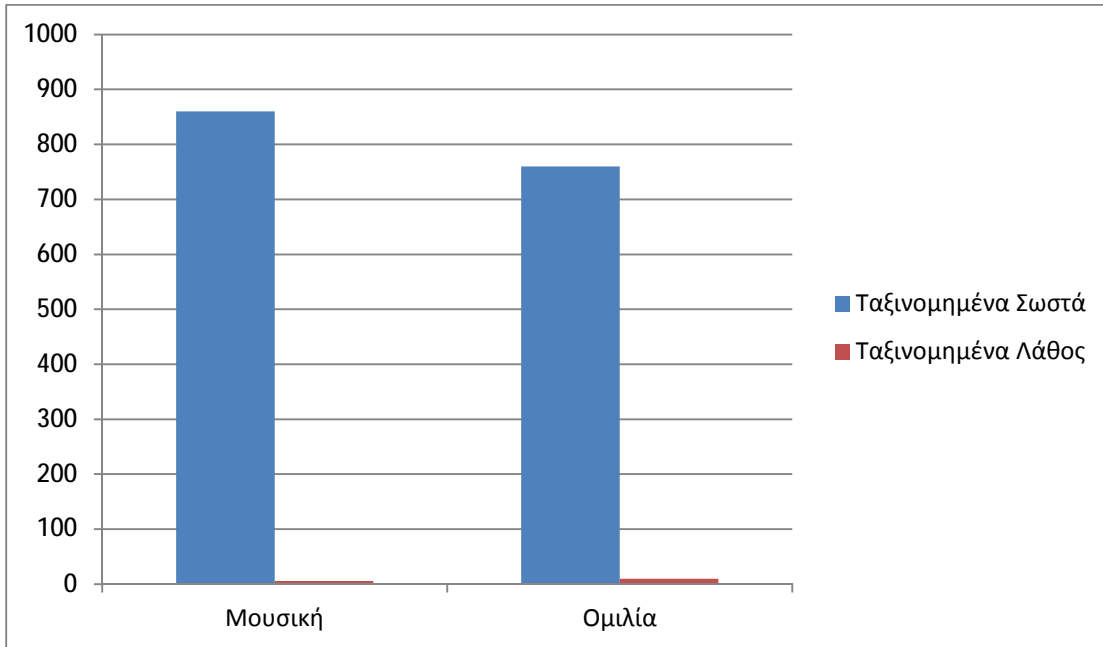
Πίνακας 6:Data Spectral Centroid



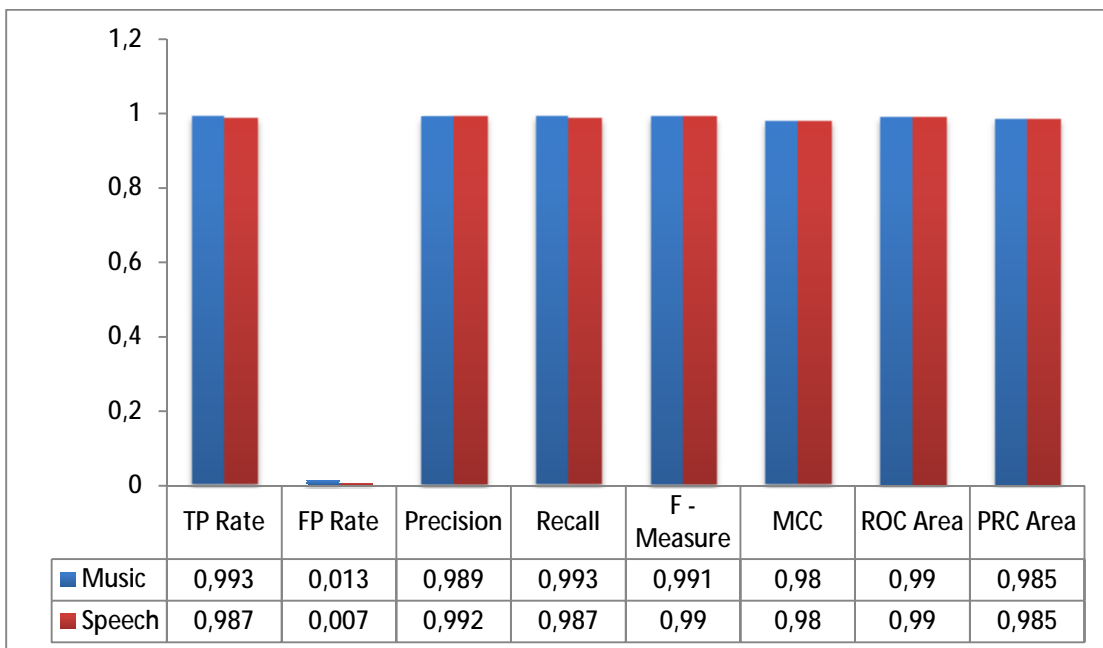
Πίνακας 7:Confusion Matrix Zero Crossing Rate



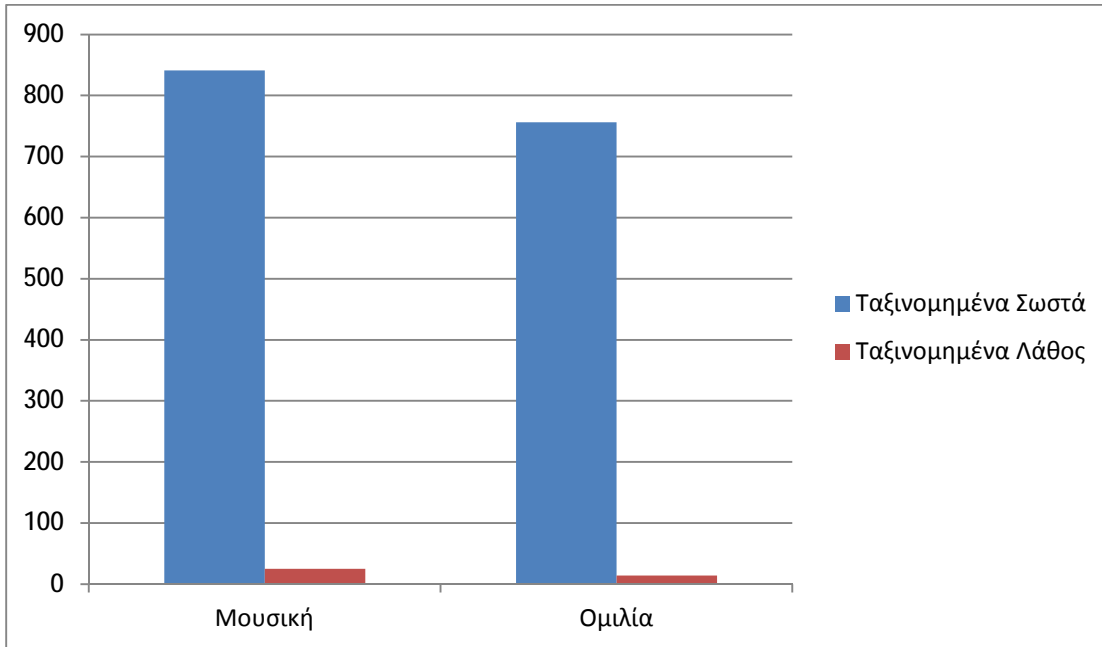
Πίνακας 8:Data Zero Crossing Rate



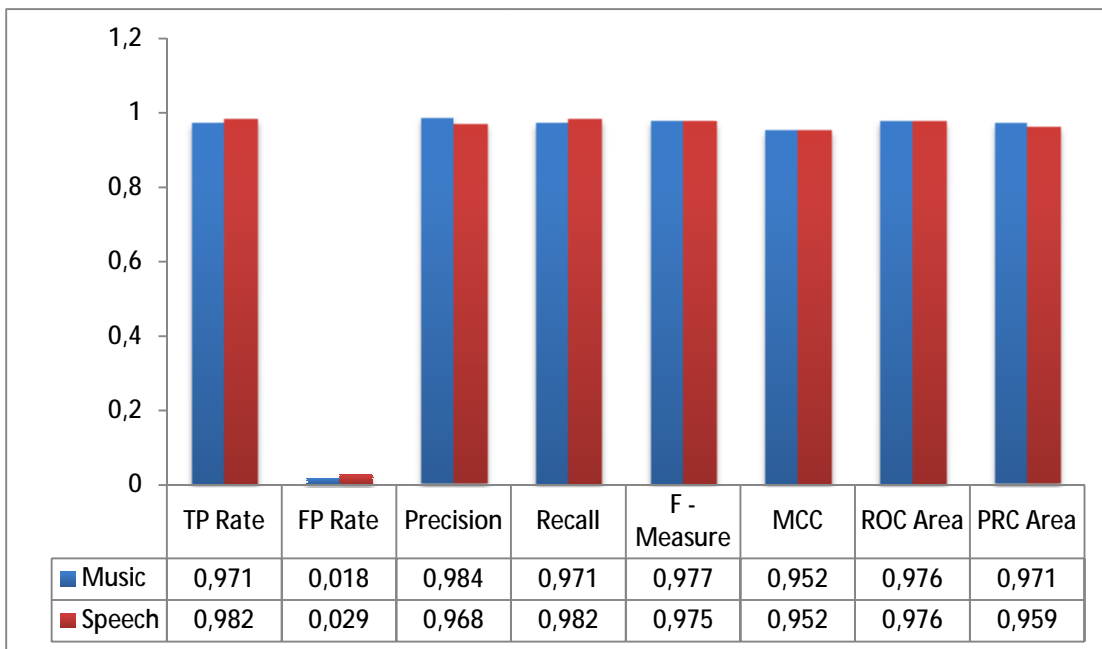
Πίνακας 9:Confusion Matrix MFCC



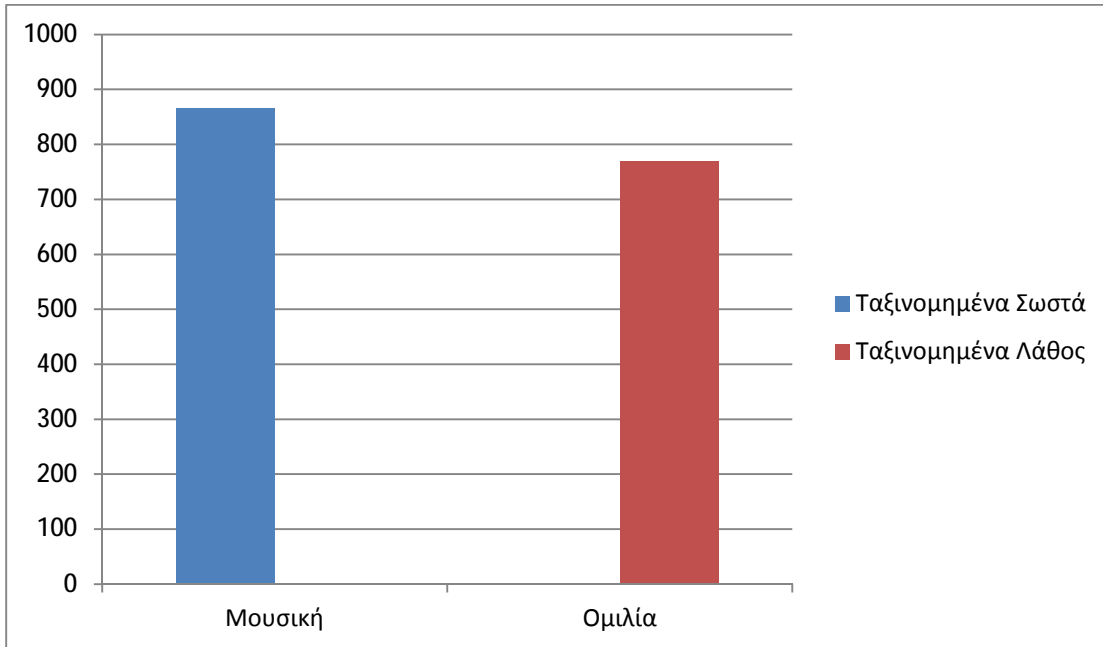
Πίνακας 10:Data MFCC



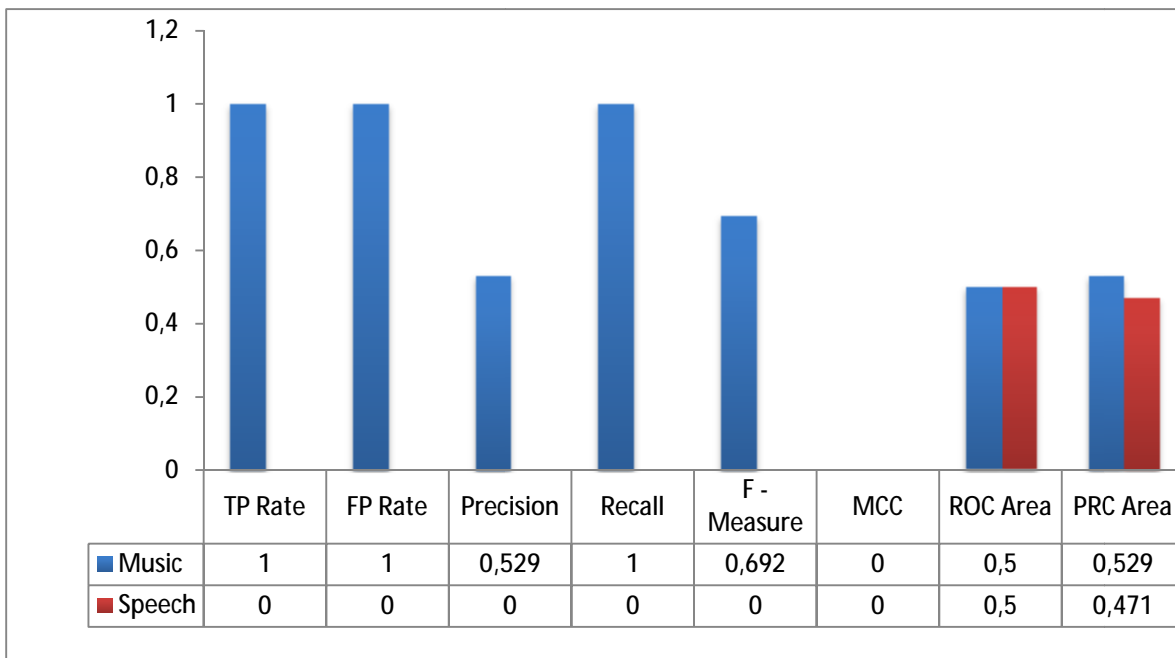
Πίνακας 11:Confusion Matrix Roll Off



Πίνακας 12:Data Roll Off

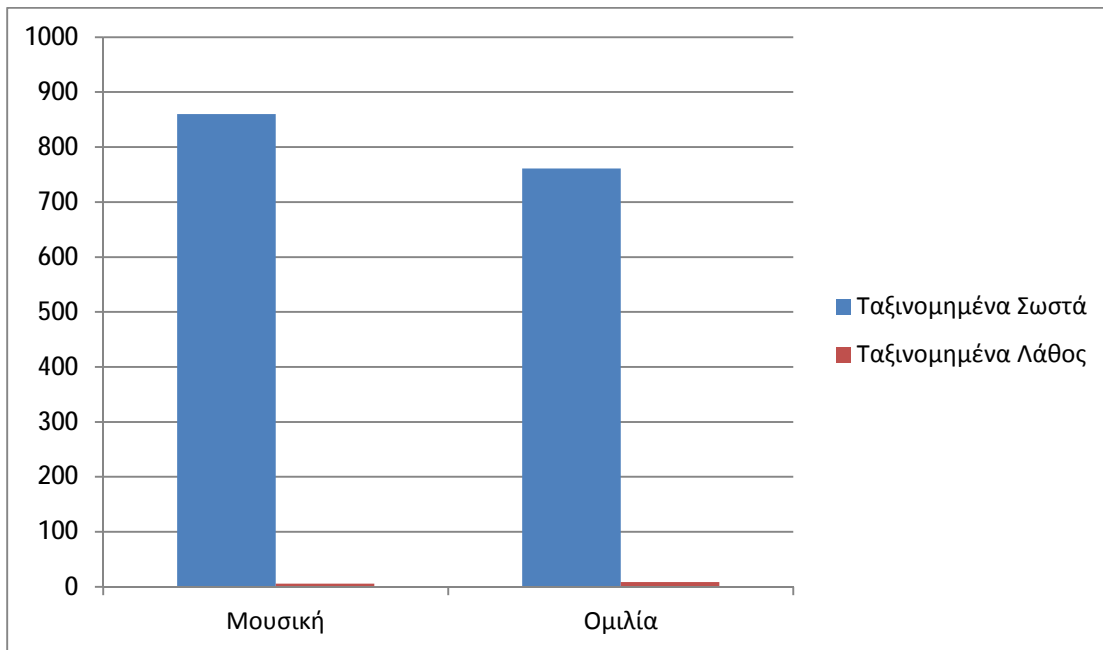


Πίνακας 13:Confusion Matrix Flux

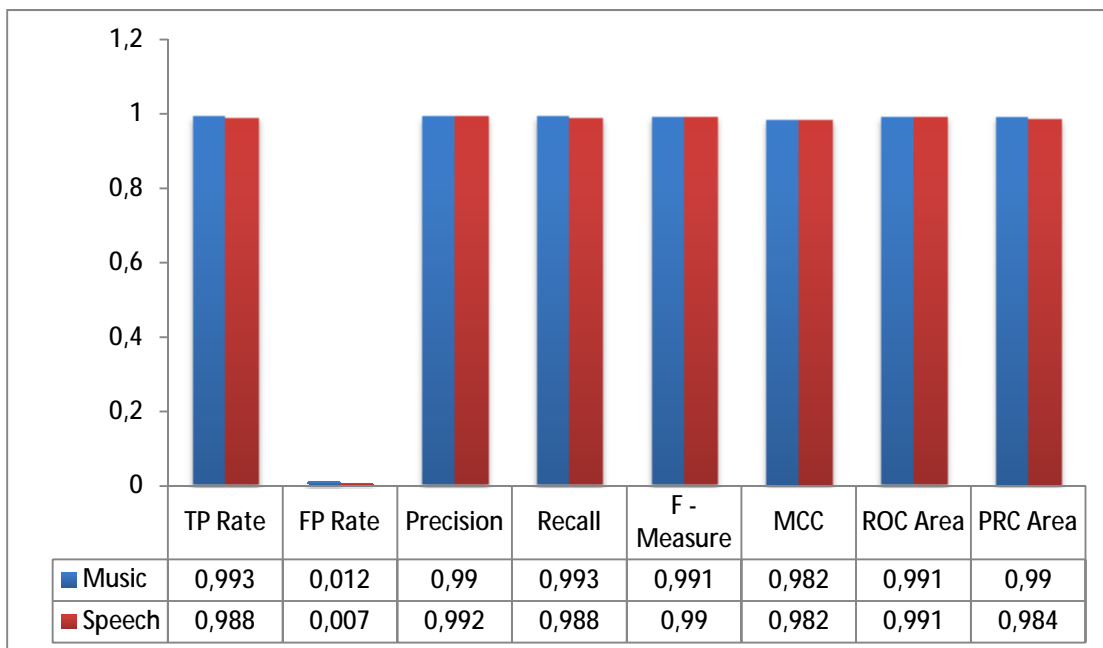


Πίνακας 14:Data Flux

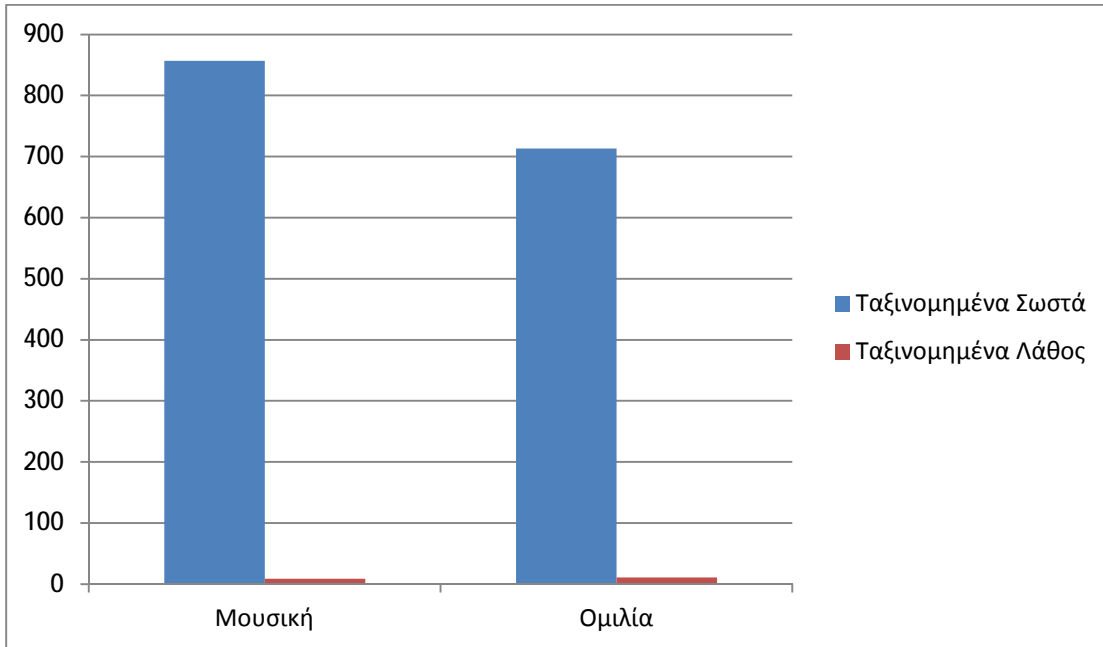
6.2.2 Ταξινομητής LMT



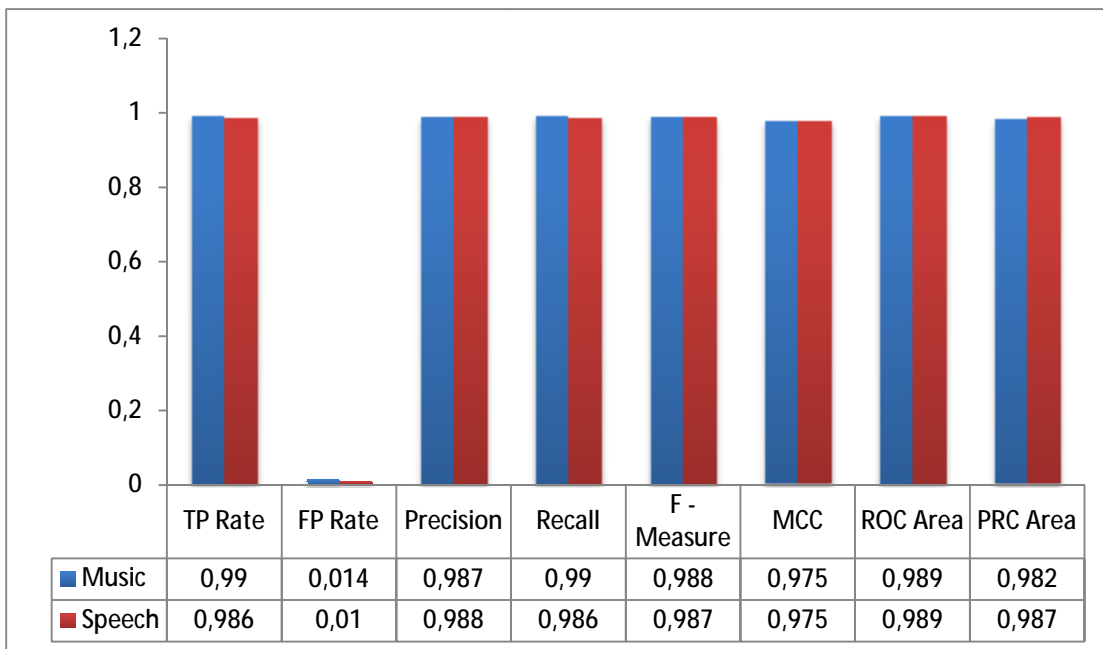
Πίνακας 15: Confusion Matrix All Features



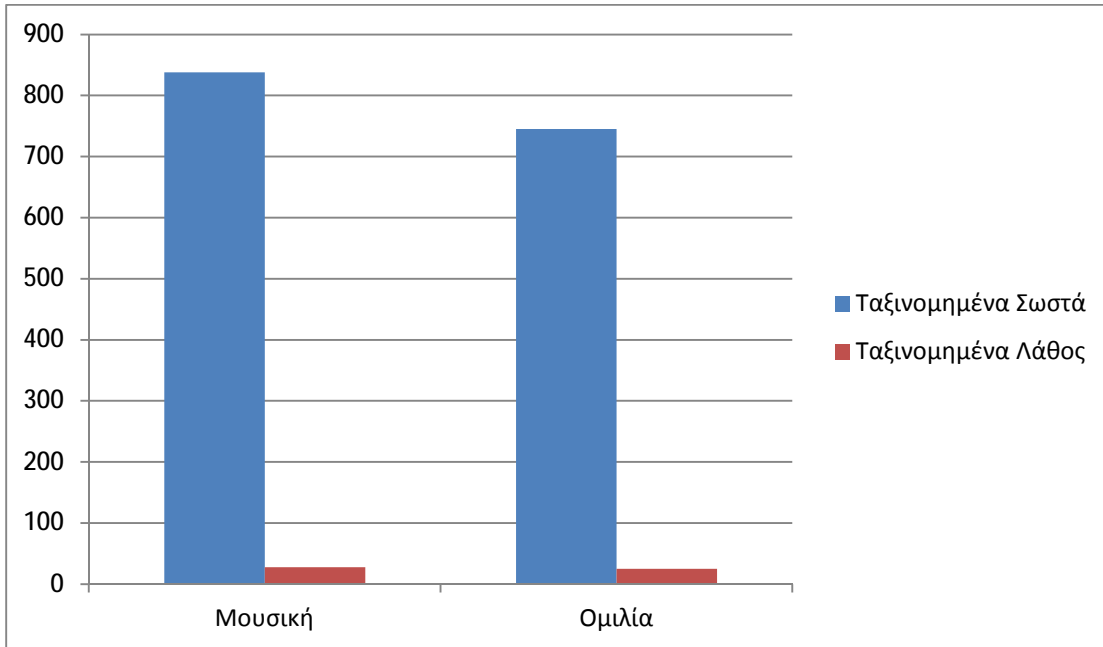
Πίνακας 16: Data All Features



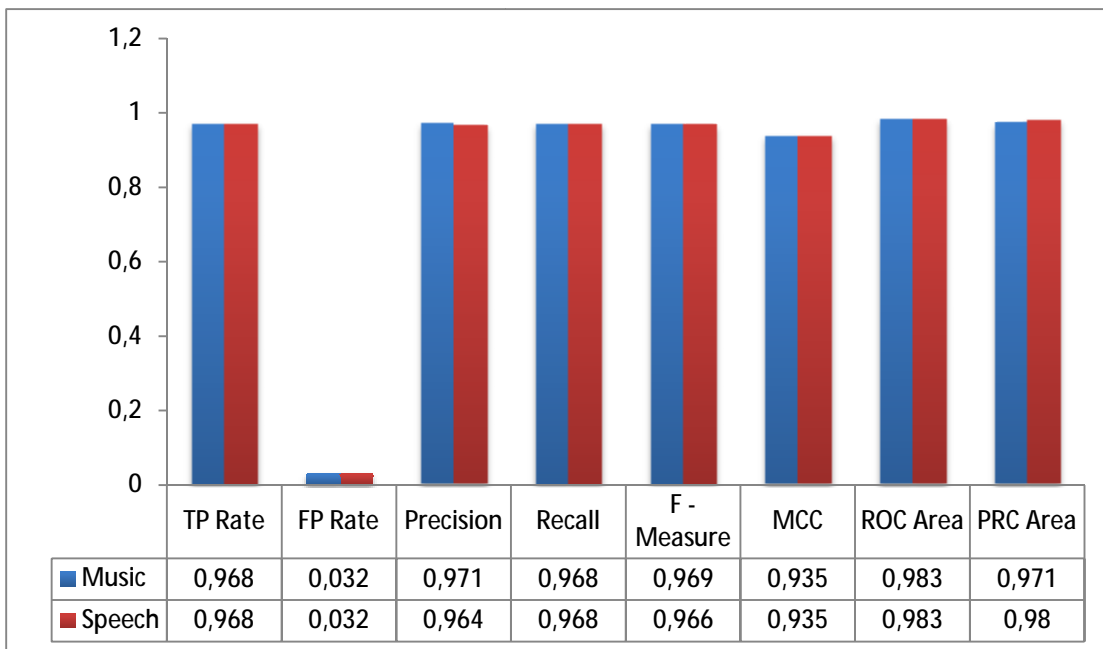
Πίνακας 17:Confusion Matrix Spectral Features



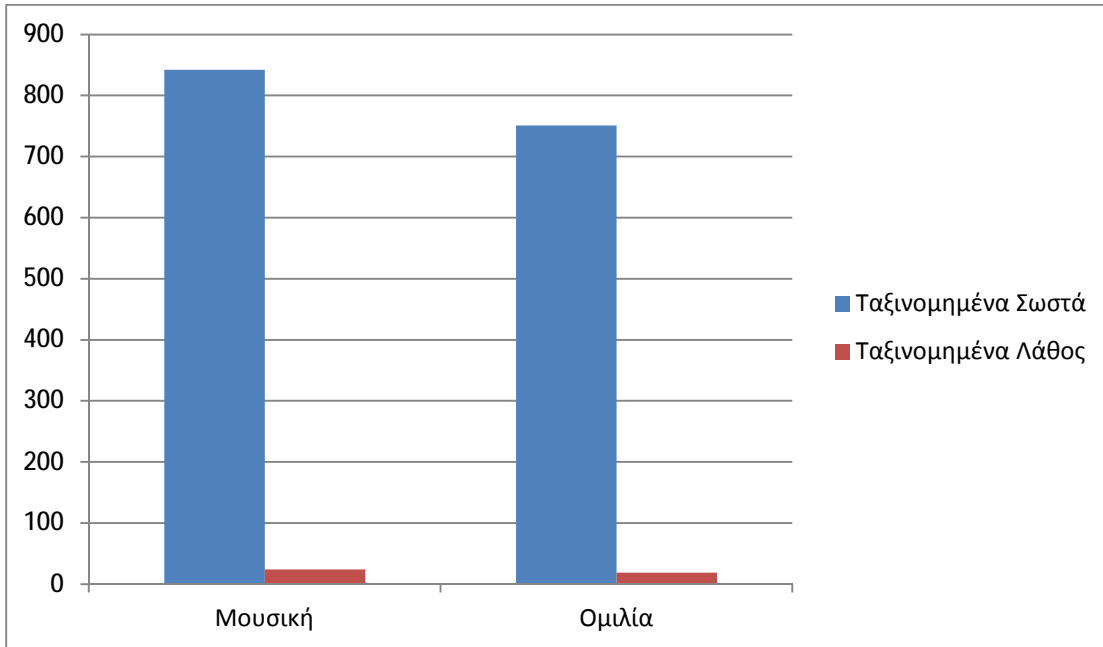
Πίνακας 18:Data Spectral Features



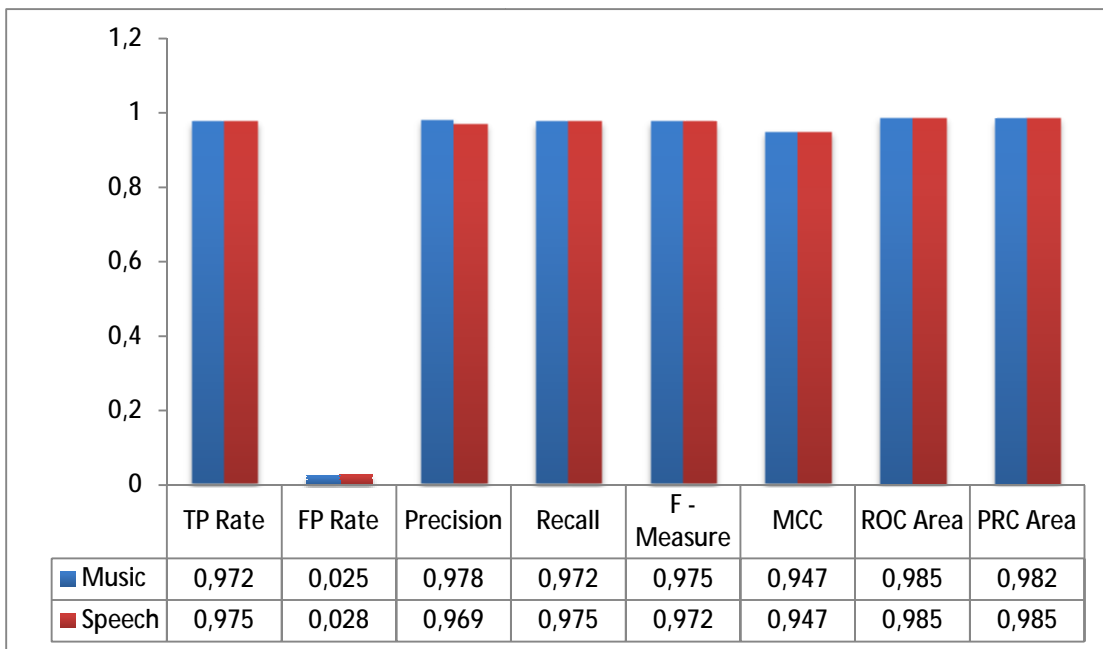
Πίνακας 19:Confusion Matrix Spectral Centroid



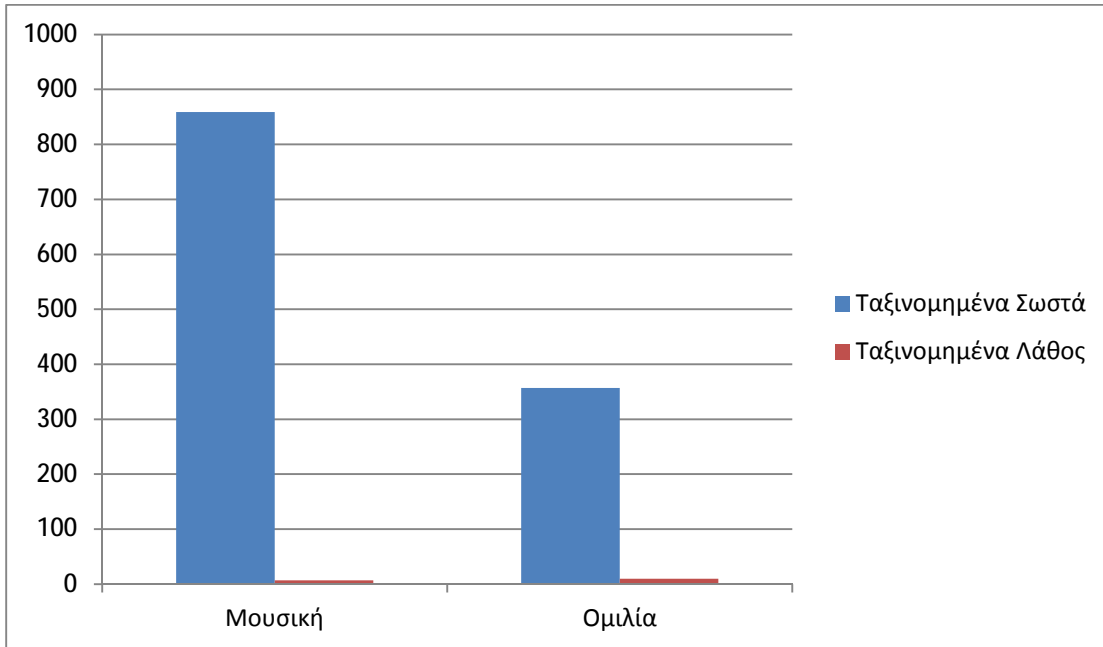
Πίνακας 20:Data Spectral Centroid



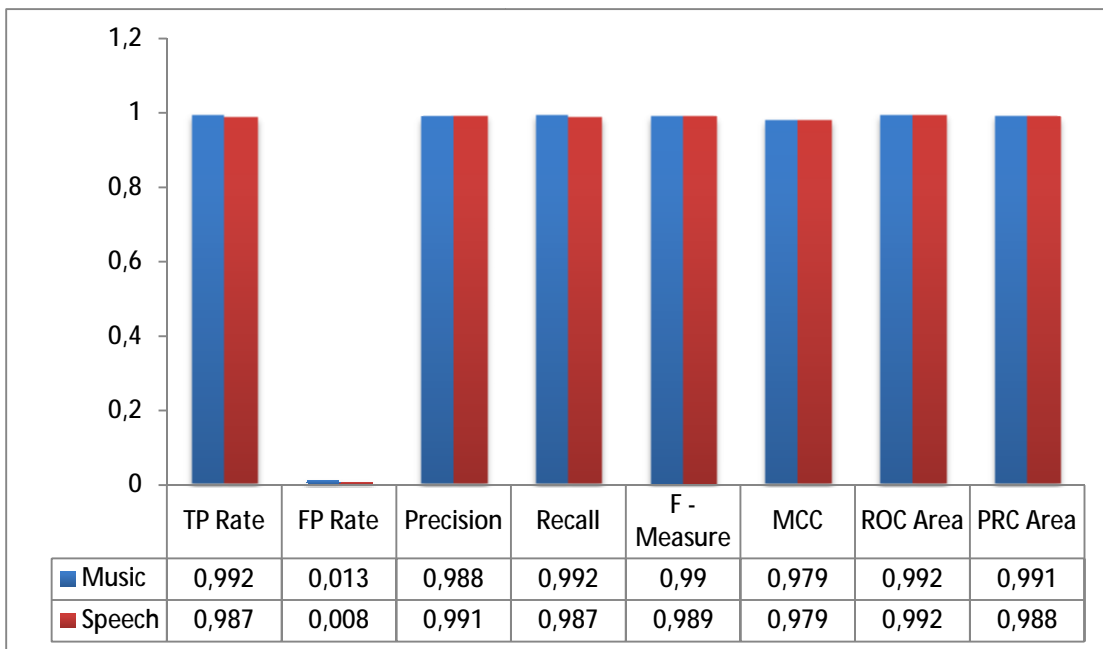
Πίνακας 21:Confusion Matrix Zero Crossing Rate



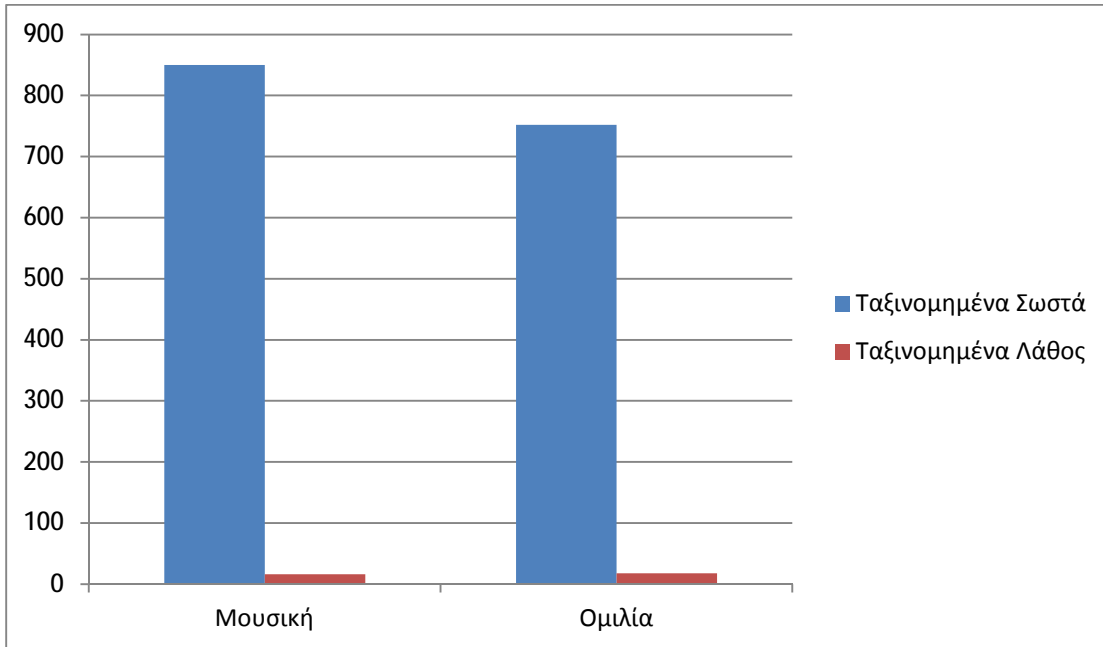
Πίνακας 22:Data Zero Crossing Rate



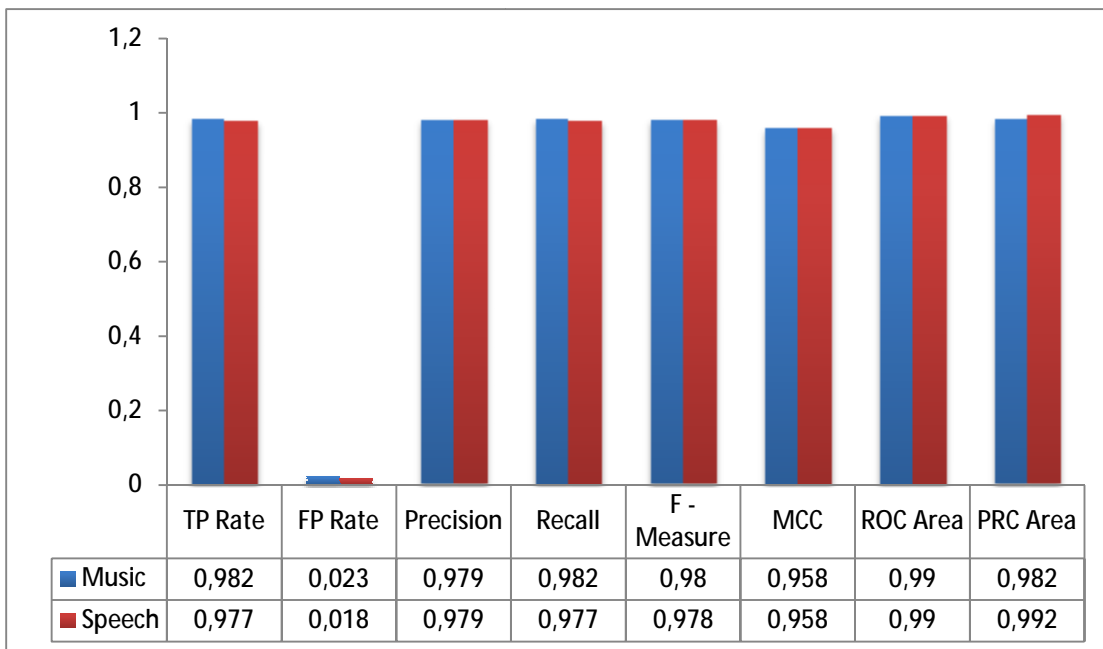
Πίνακας 23:Confusion Matrix MFCC



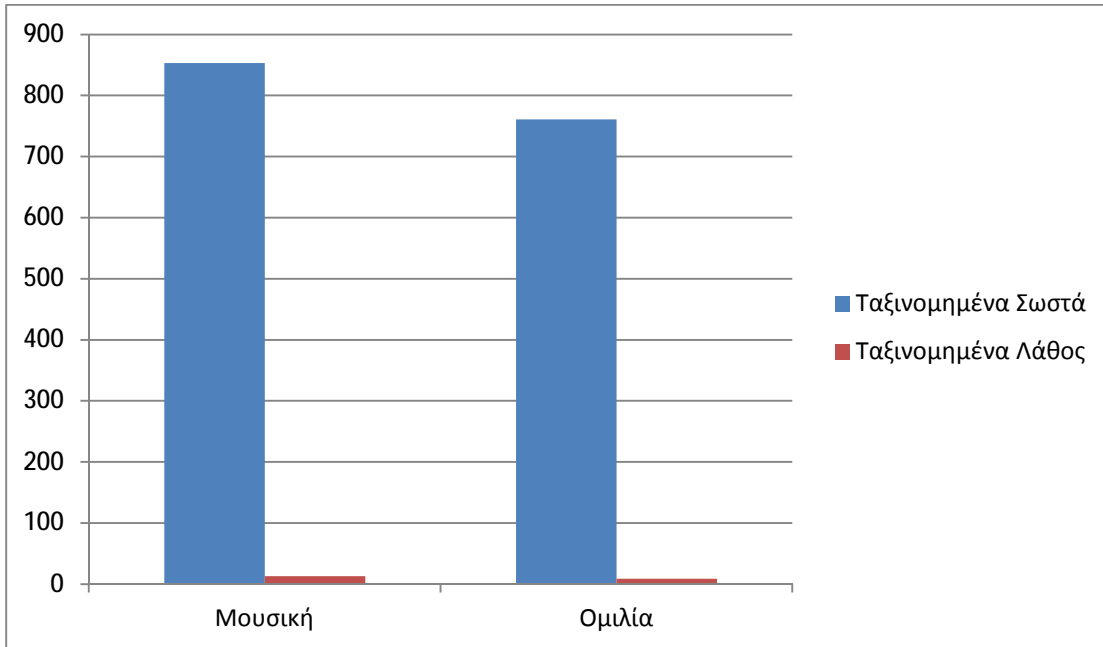
Πίνακας 24:Data MFCC



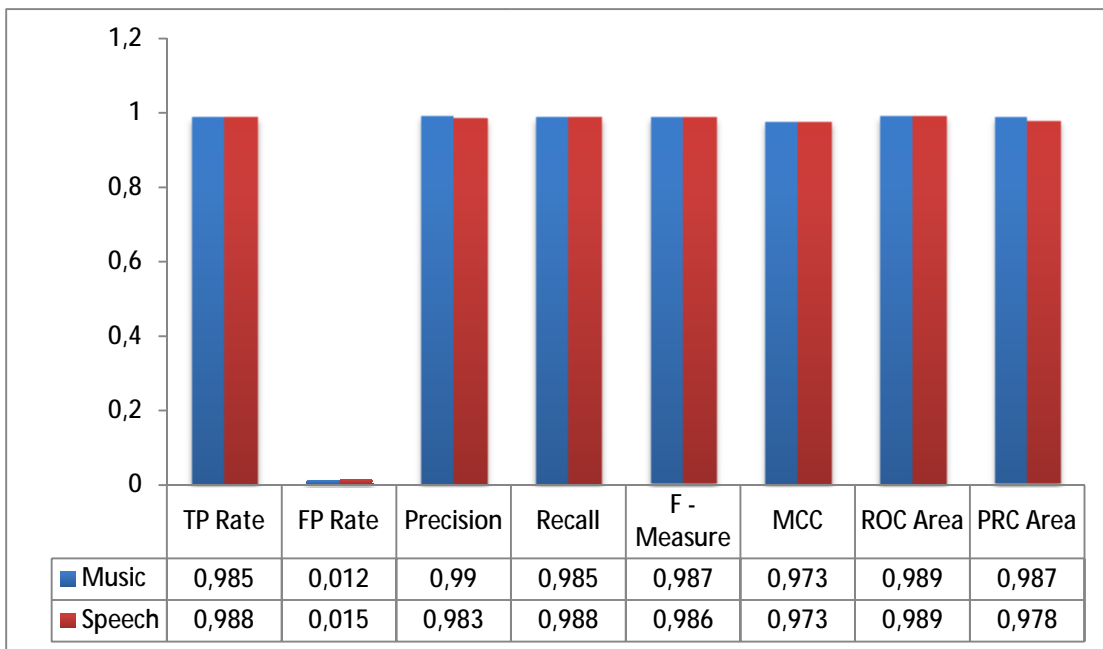
Πίνακας 25:Confusion Matrix Roll Off



Πίνακας 26:Data Roll Off

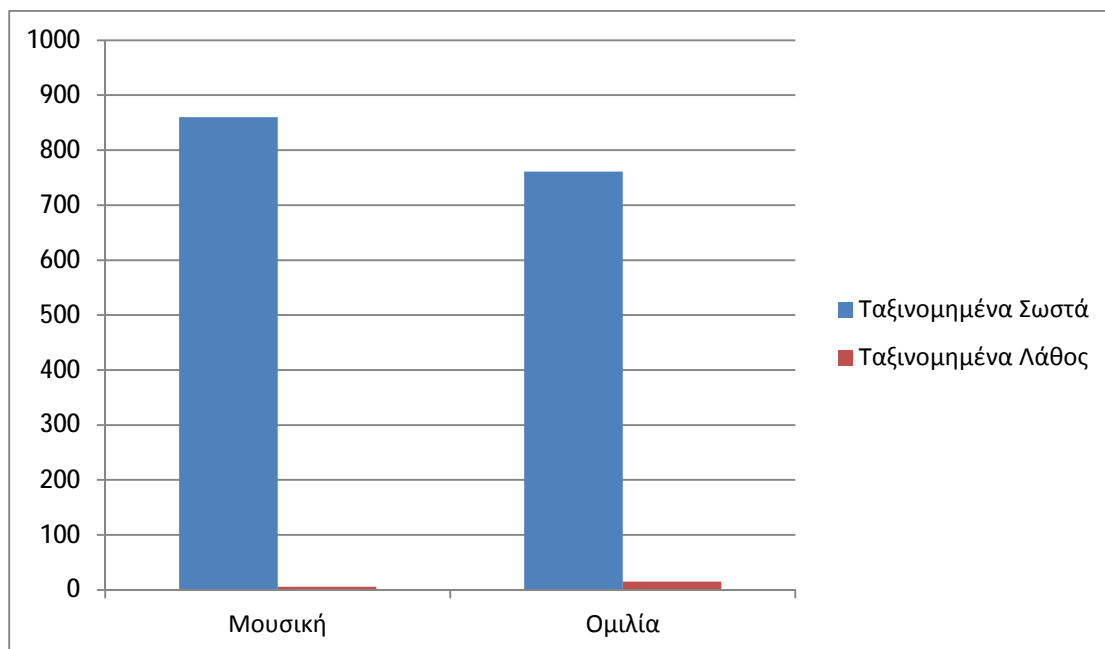


Πίνακας 27:Confusion Matrix Flux

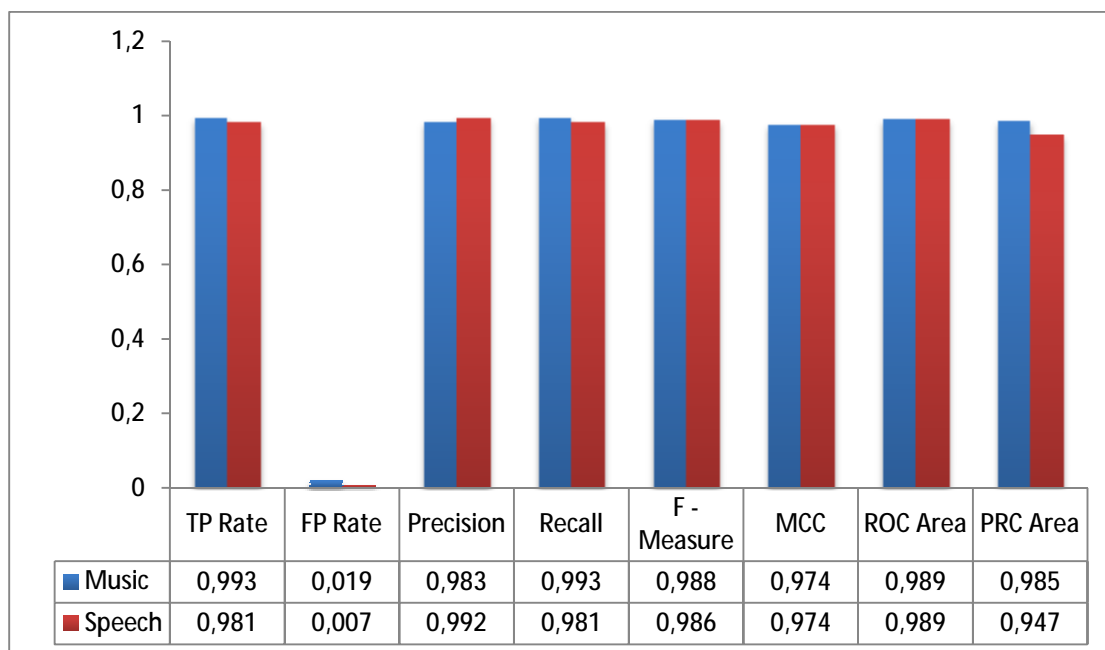


Πίνακας 28:Data Flux

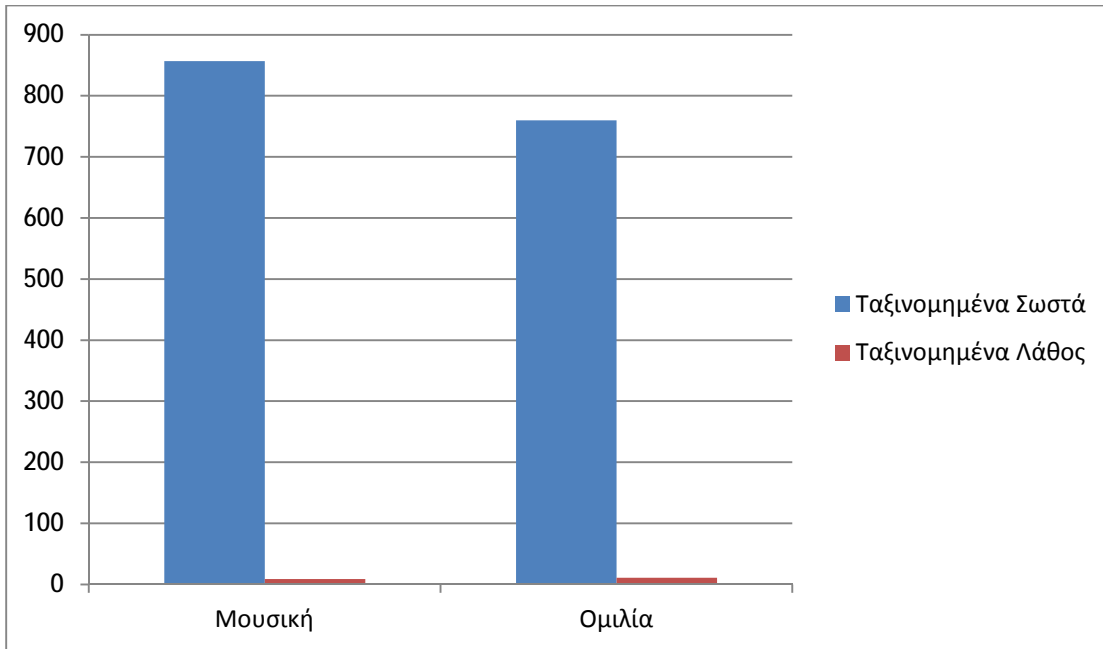
6.2.3 Ταξινομητής Multilayer Perceptron



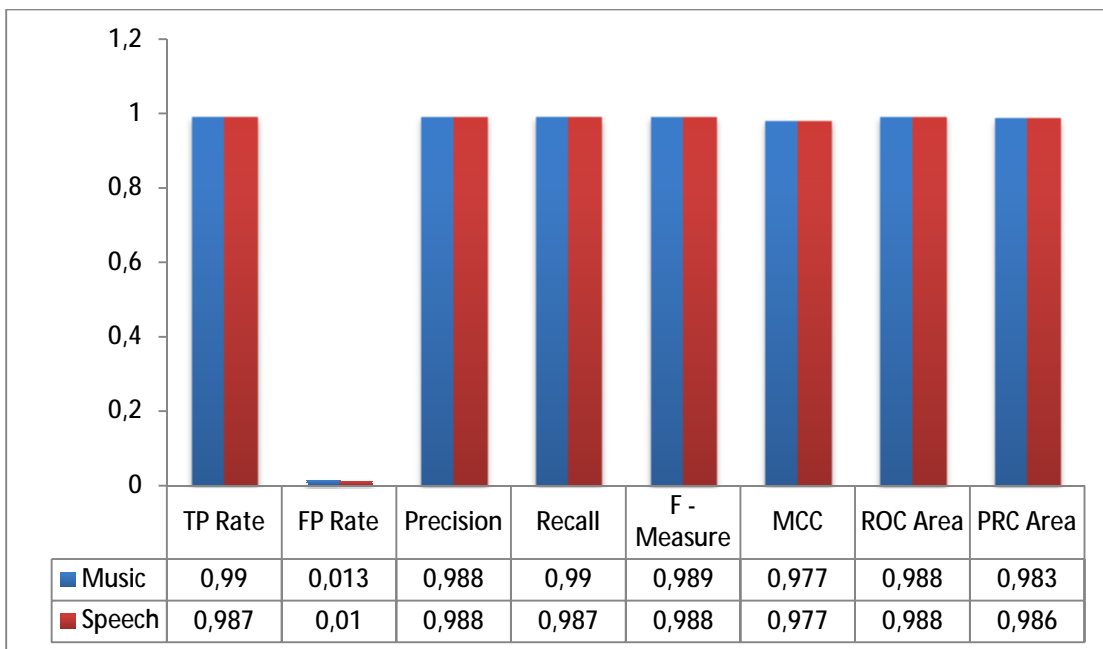
Πίνακας 29:Confusion Matrix All Features



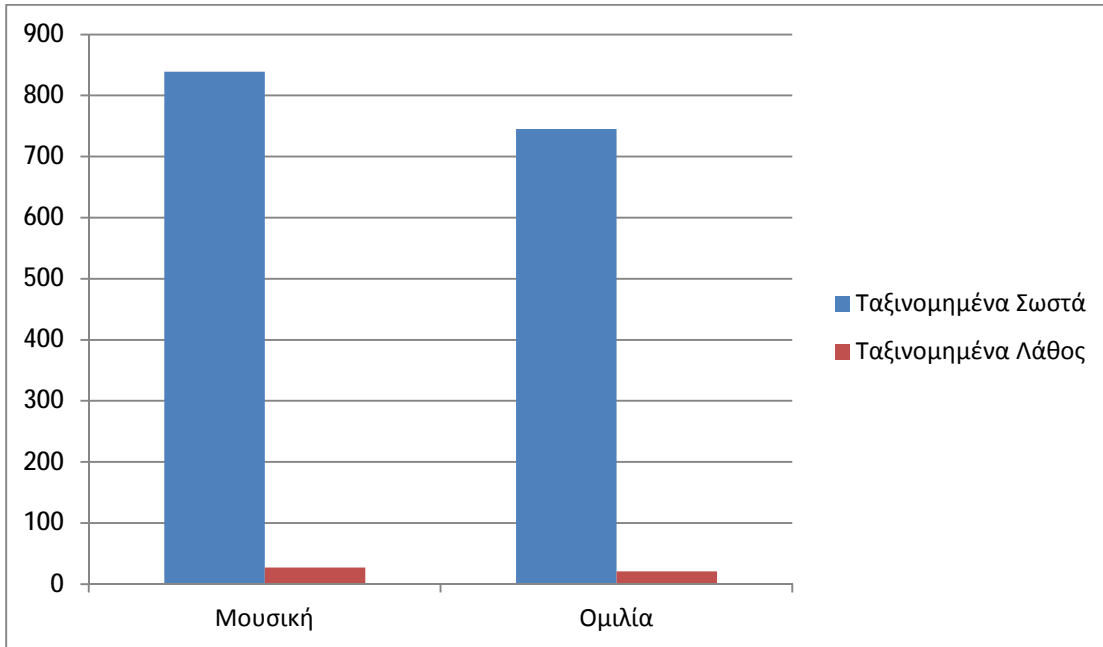
Πίνακας 30:Data All Features



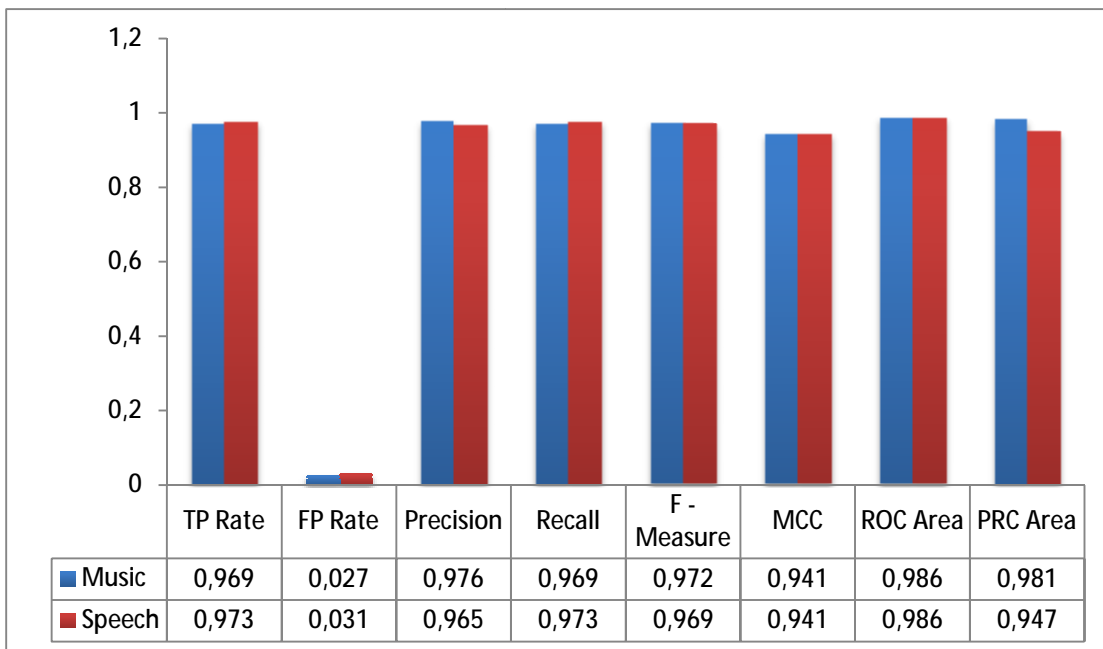
Πίνακας 31:Confusion Matrix Spectral Features



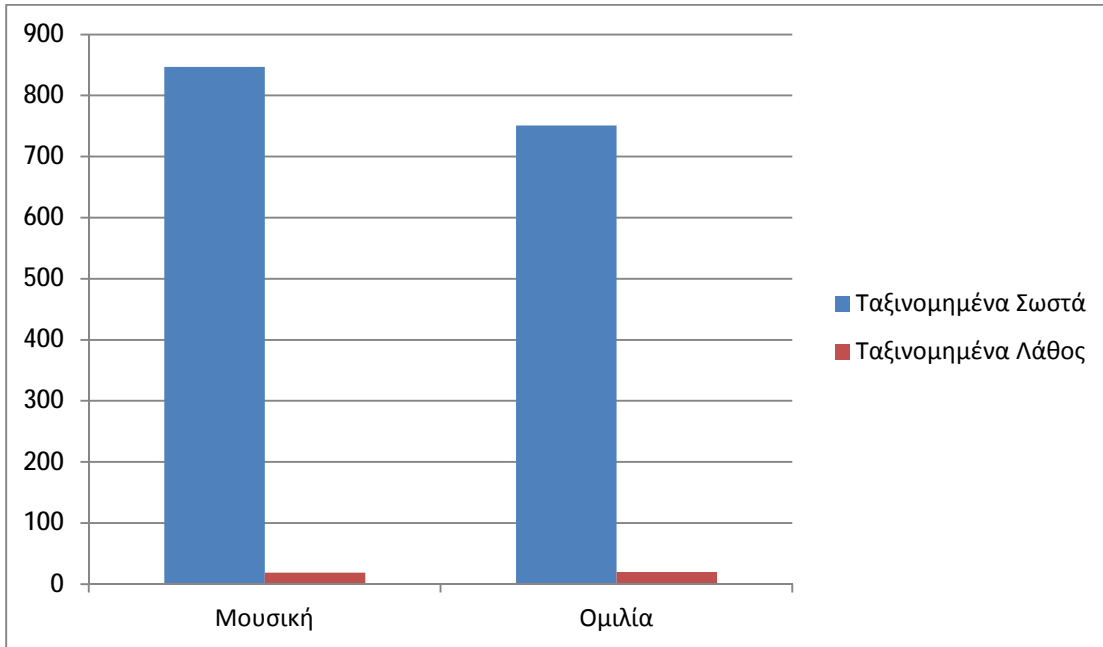
Πίνακας 32:Data Spectral Features



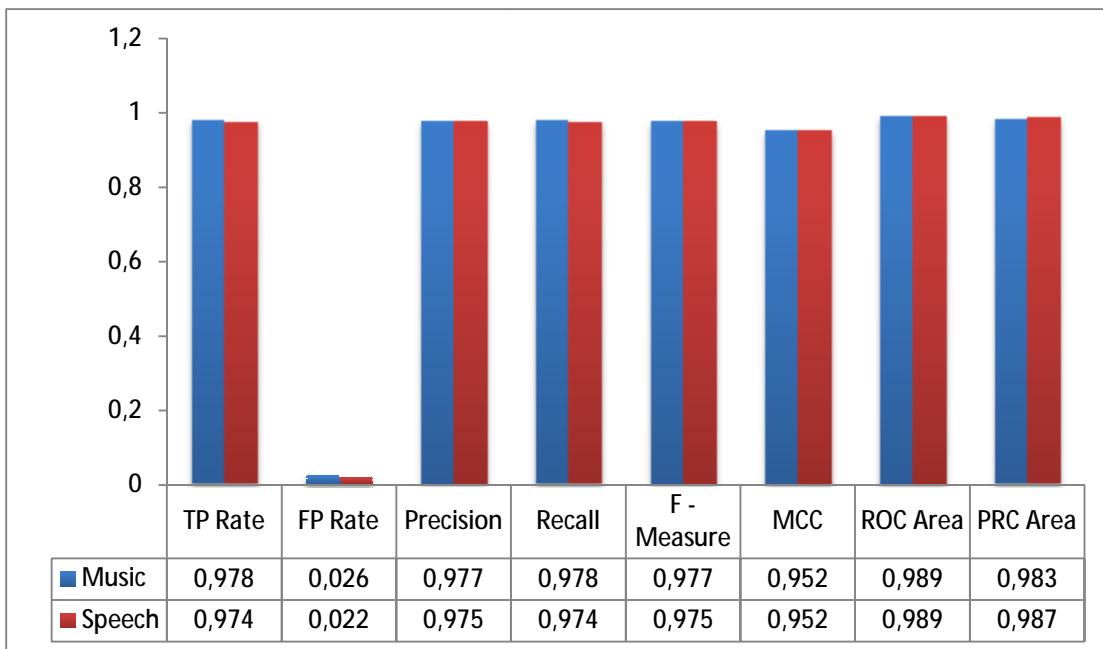
Πίνακας 33:Confusion Matrix Spectral Centroid



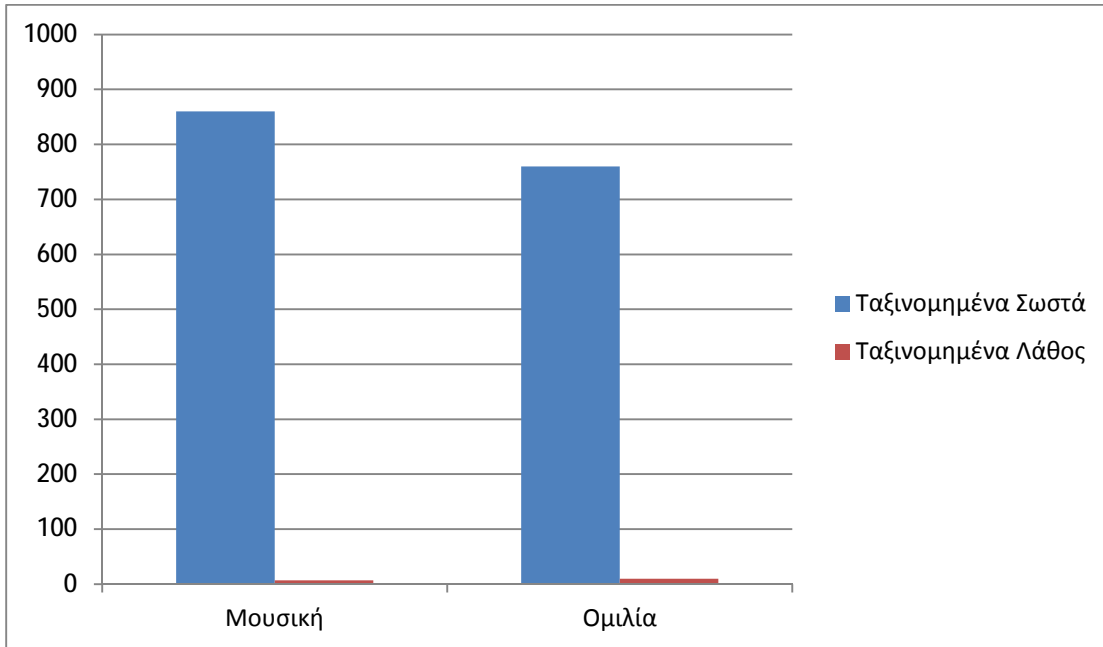
Πίνακας 34:Data Spectral Centroid



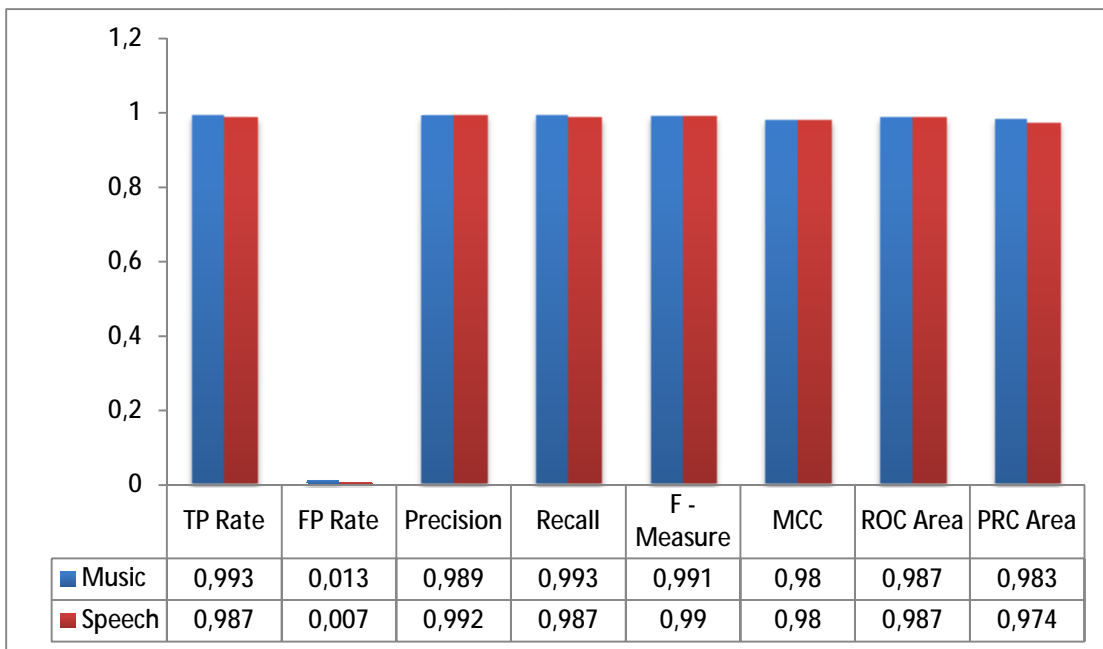
Πίνακας 35:Confusion Matrix Zero Crossing Rate



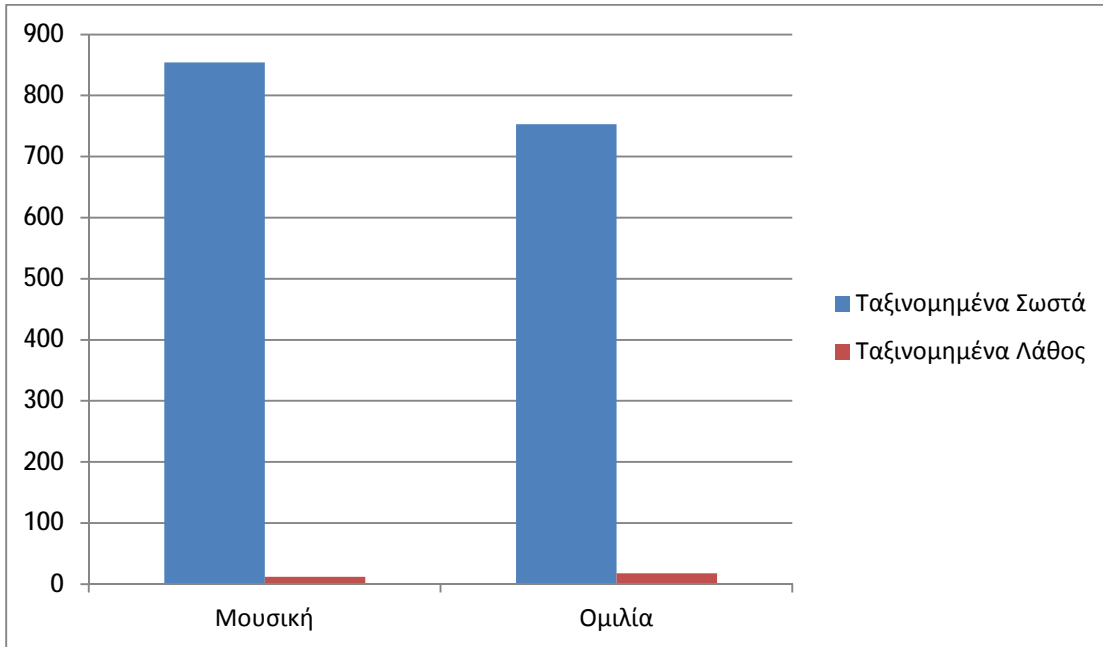
Πίνακας 36:Data Zero Crossing Rate



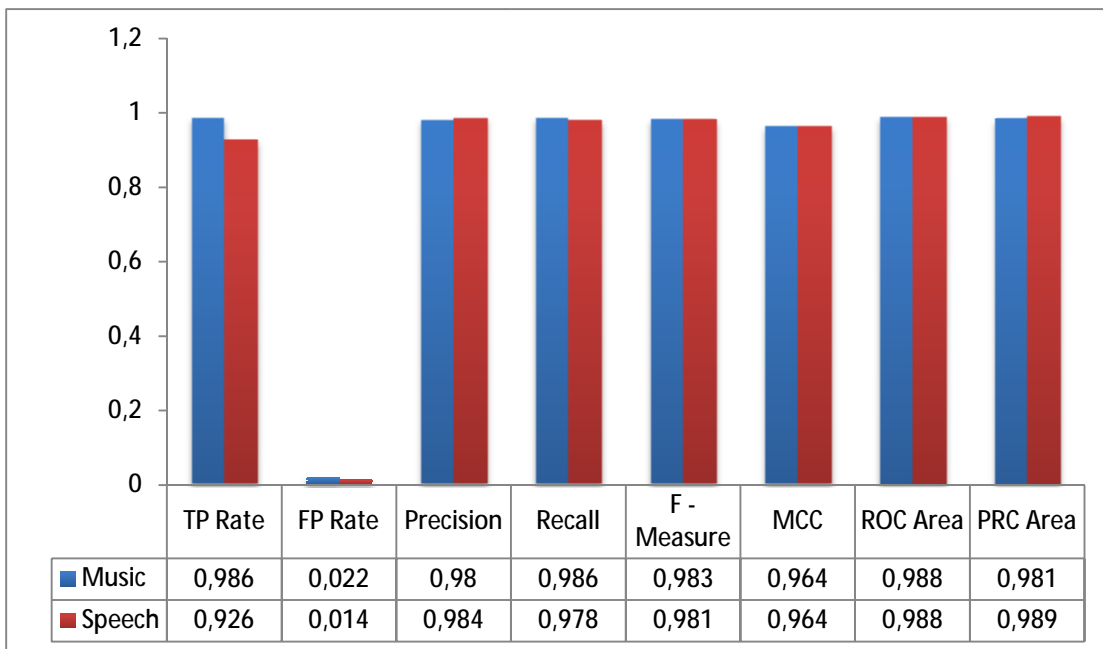
Πίνακας 37:Confusion Matrix MFCC



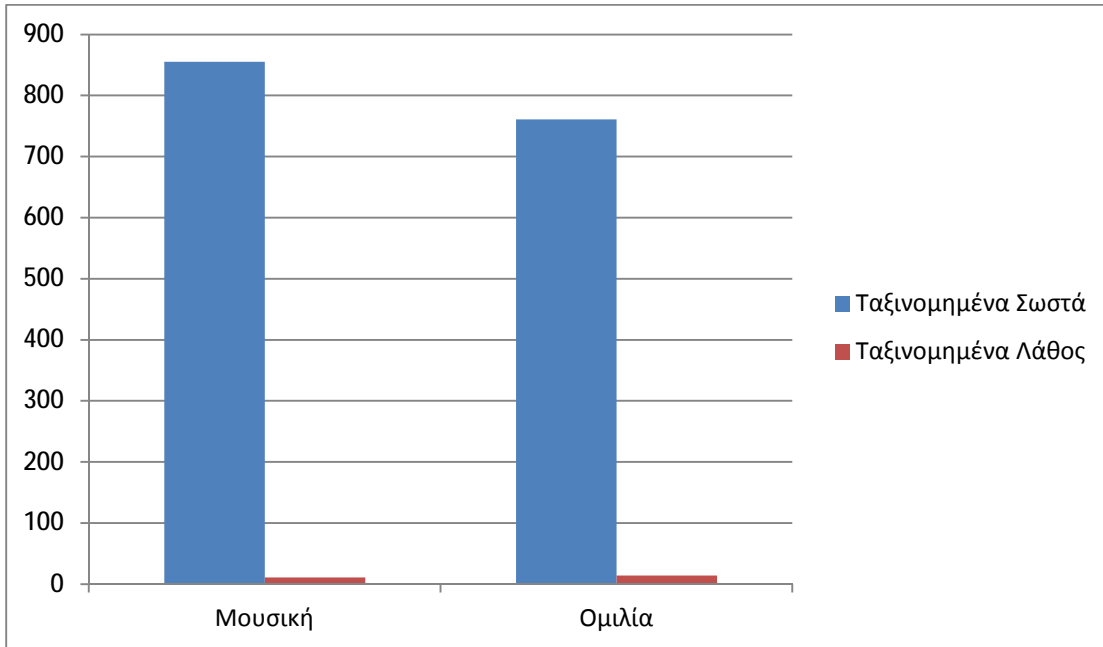
Πίνακας 38:Data MFCC



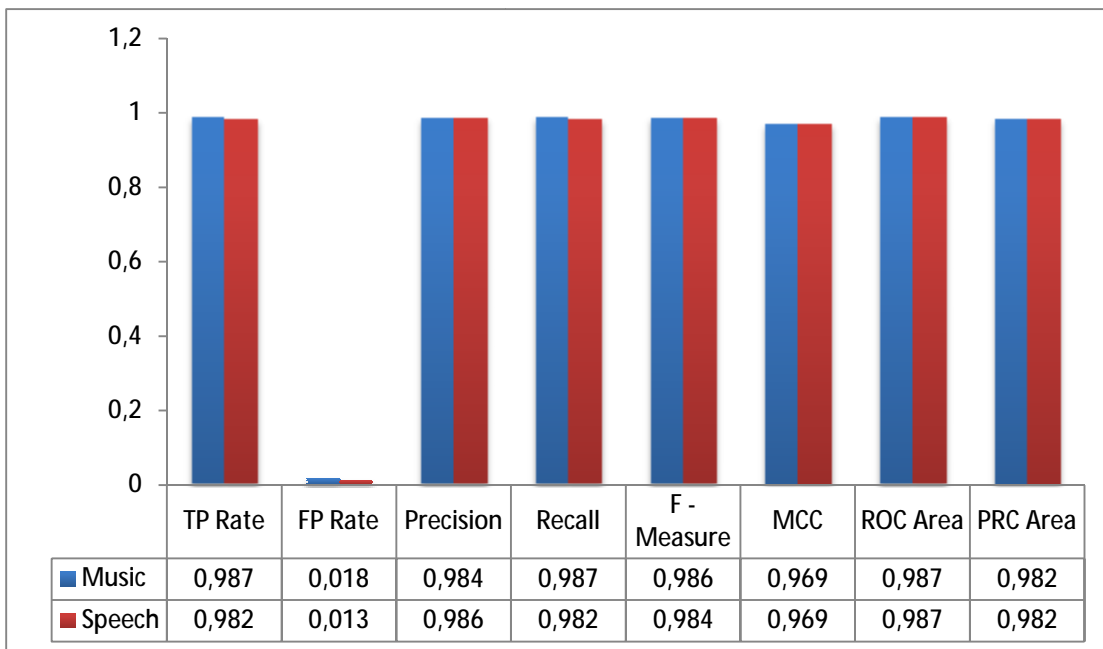
Πίνακας 39: Confusion Matrix Roll Off



Πίνακας 40: Data Roll Off

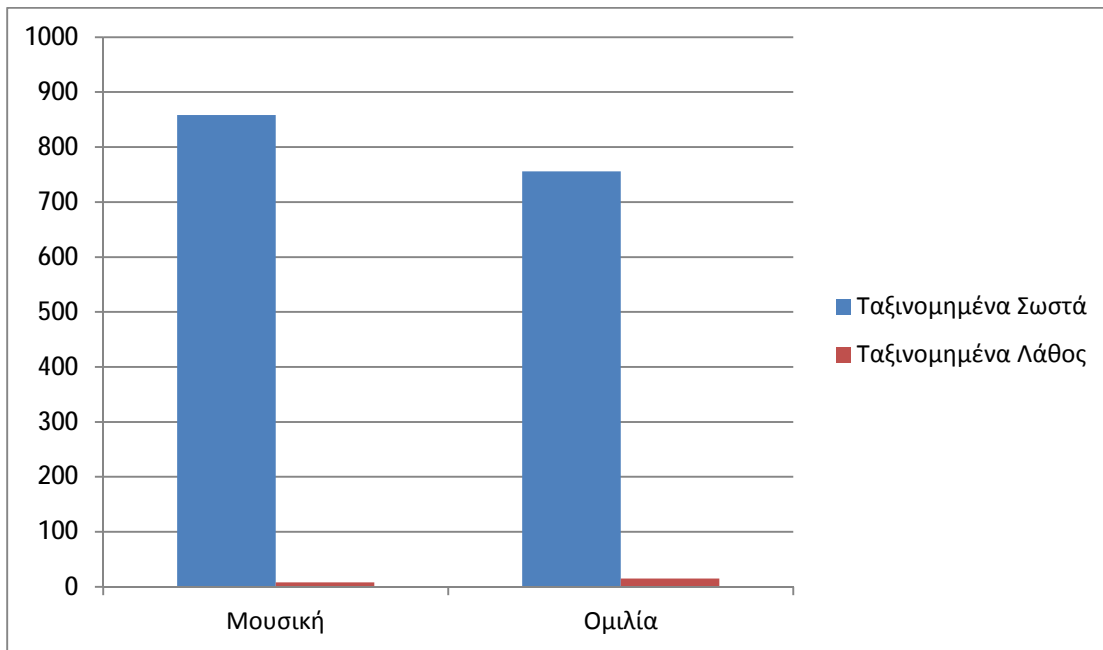


Πίνακας 41:Confusion Matrix Flux

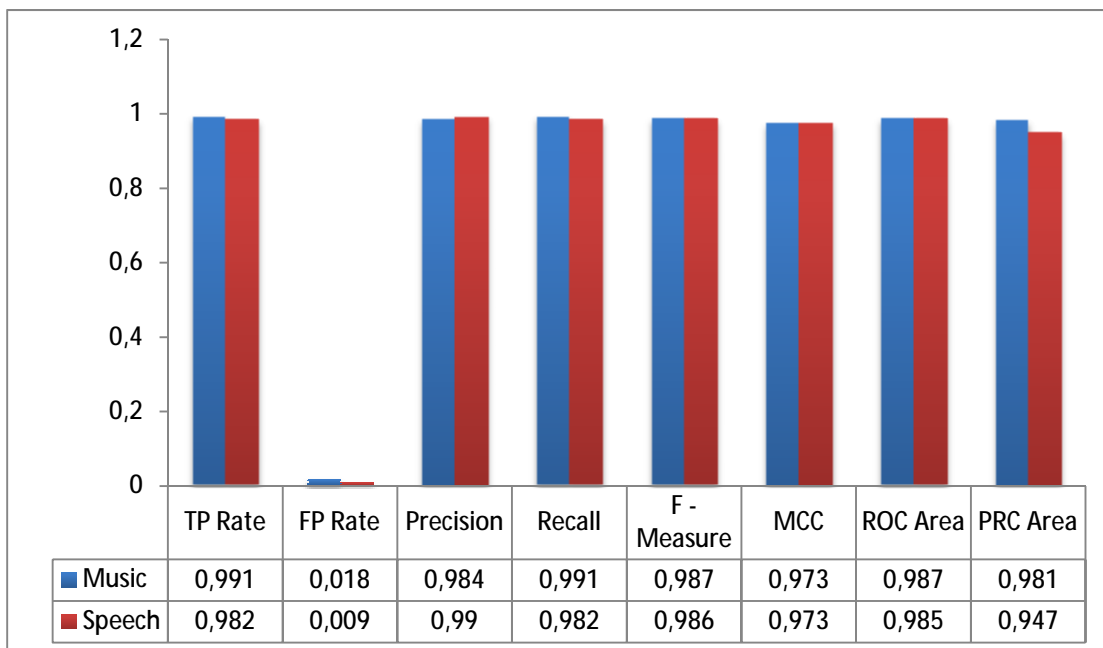


Πίνακας 42:Data Flux

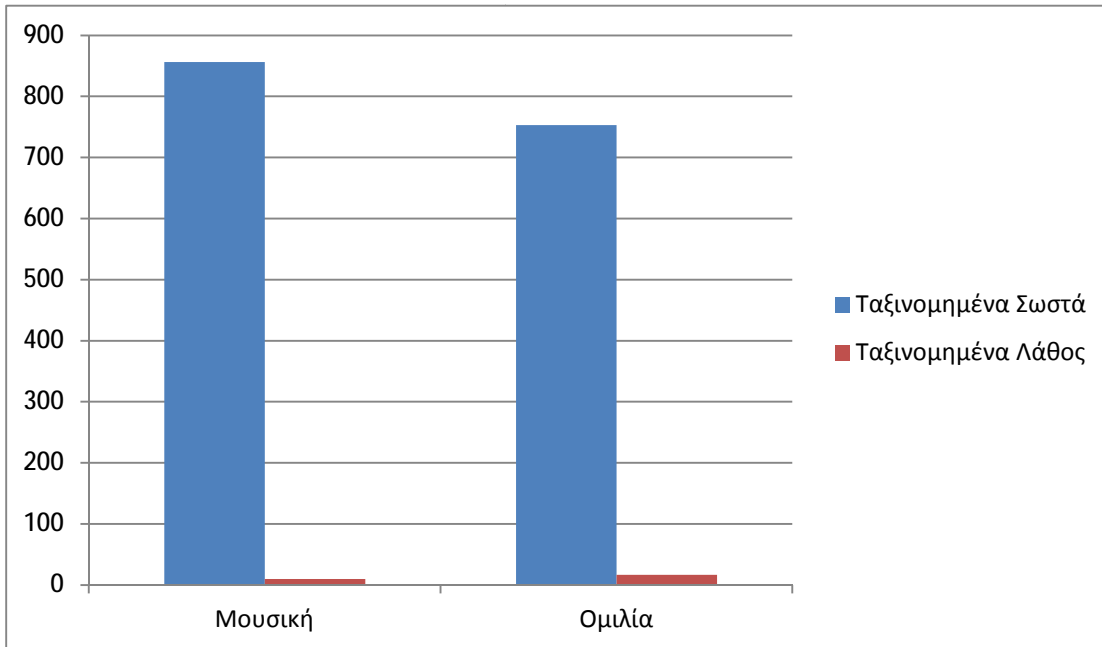
6.2.4 Ταξινομητής Naïve Bayes



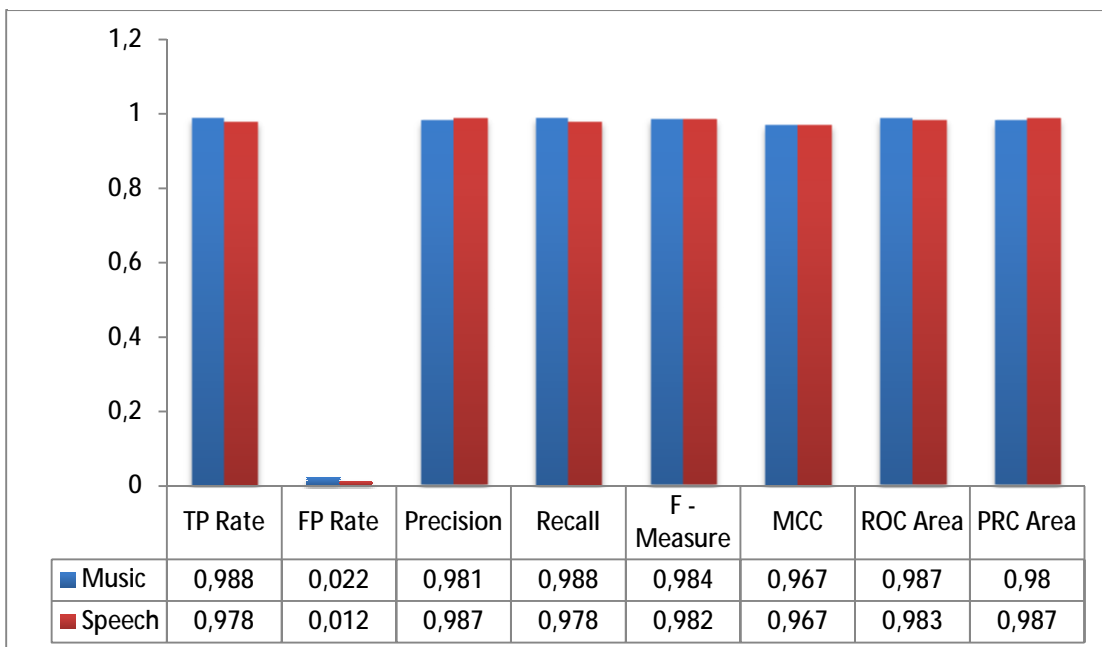
Πίνακας 43:Confusion Matrix All Features



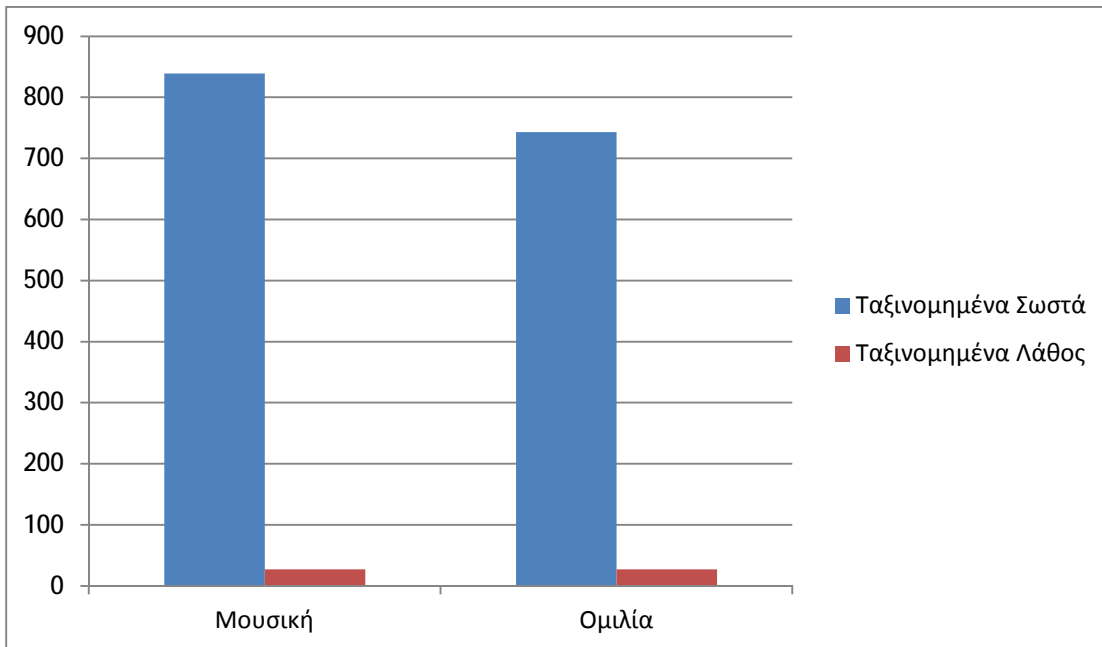
Πίνακας 44:Data All Features



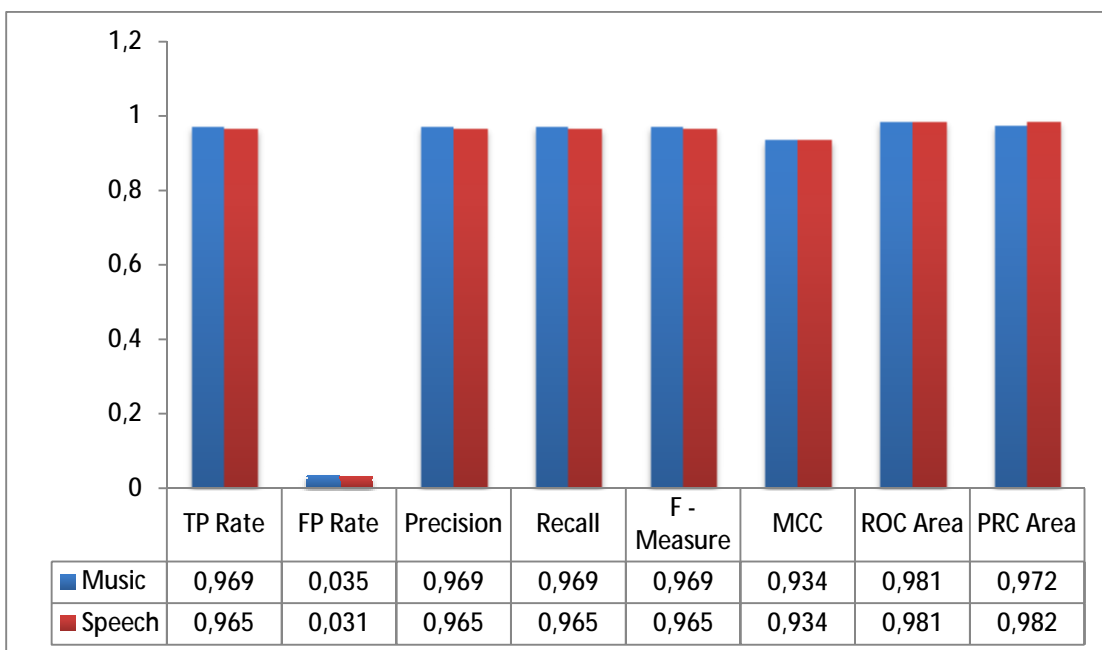
Πίνακας 46:Confusion Matrix Spectral Features



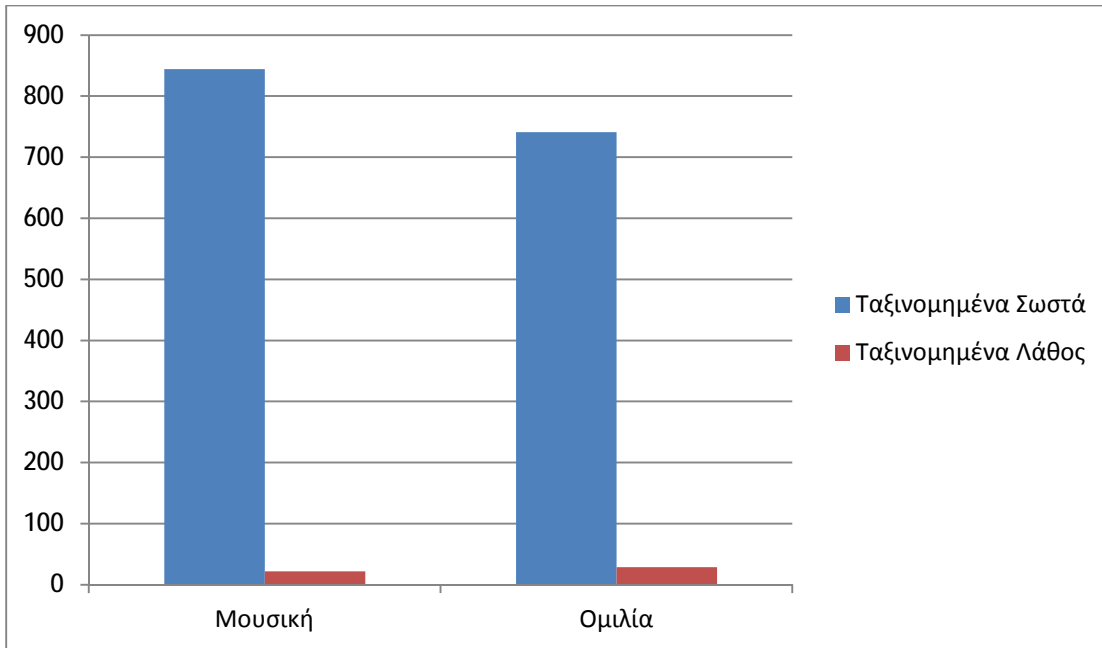
Πίνακας 47:Data Spectral Features



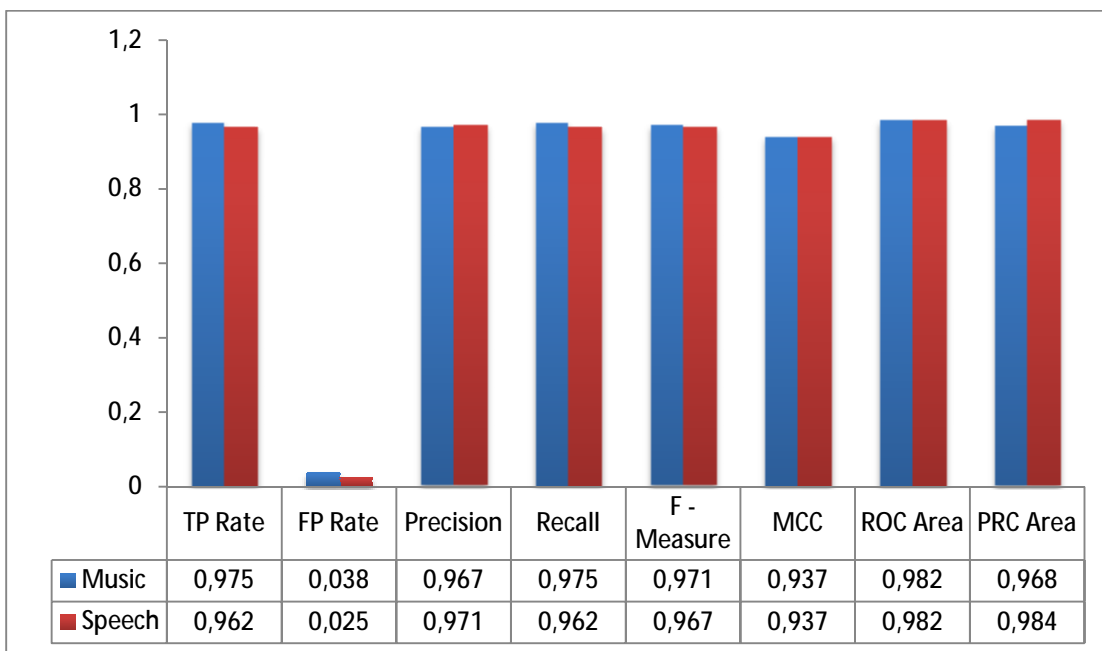
Πίνακας 48:Confusion Matrix Spectral Centroid



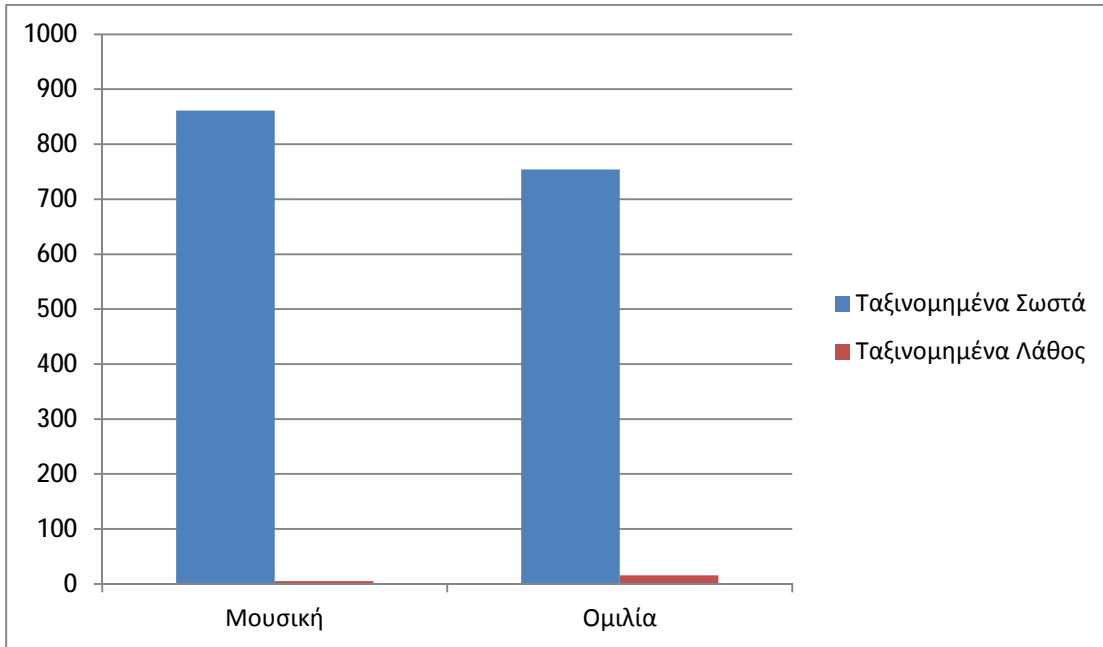
Πίνακας 49:Data Spectral Centroid



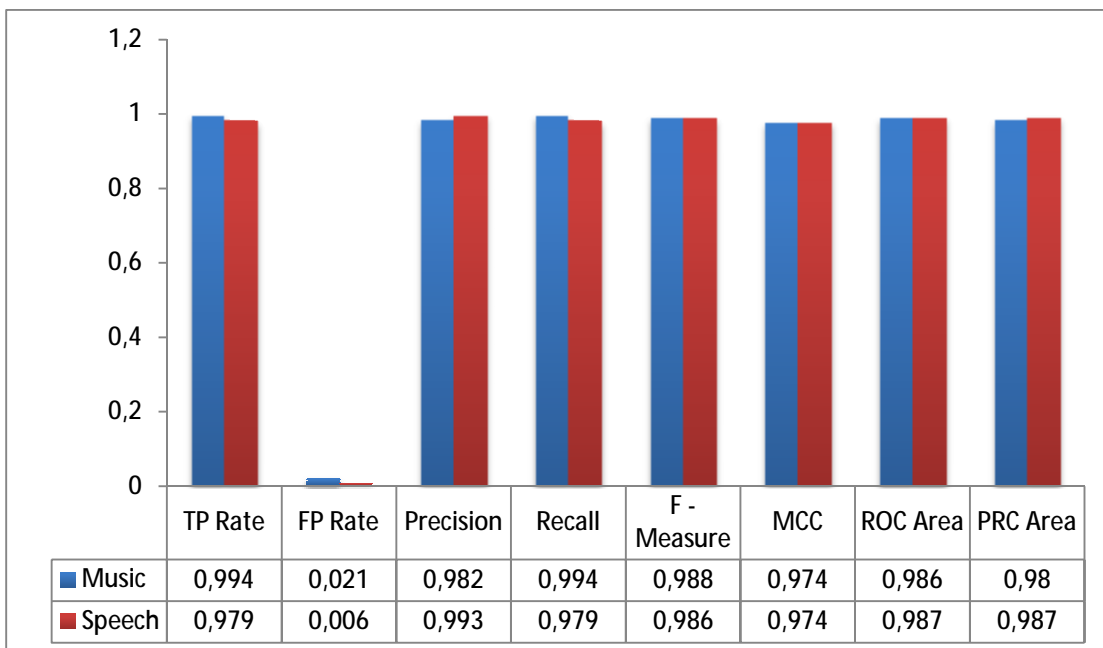
Πίνακας 50:Confusion Matrix Zero Crossing Rate



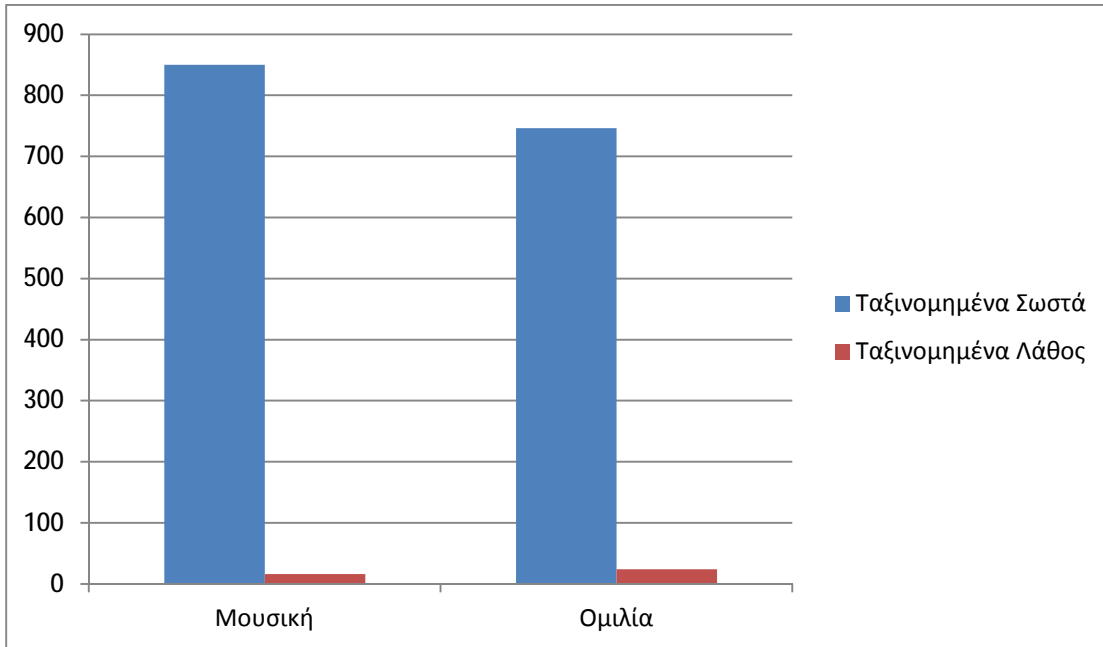
Πίνακας 51:Data Zero Crossing Rate



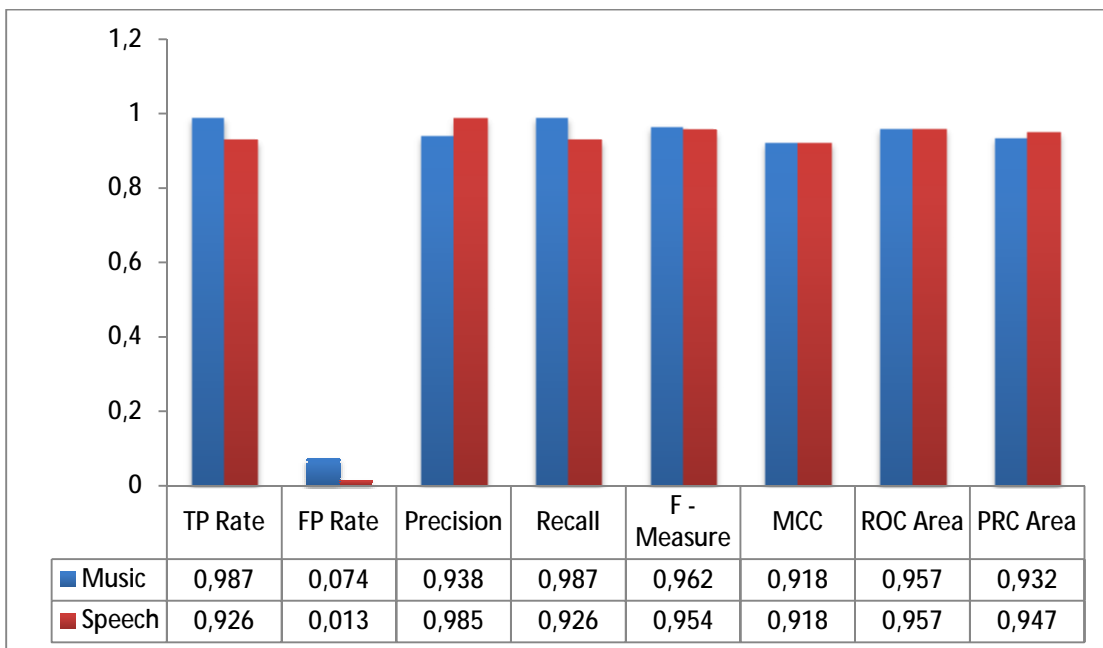
Πίνακας 52:Confusion Matrix MFCC



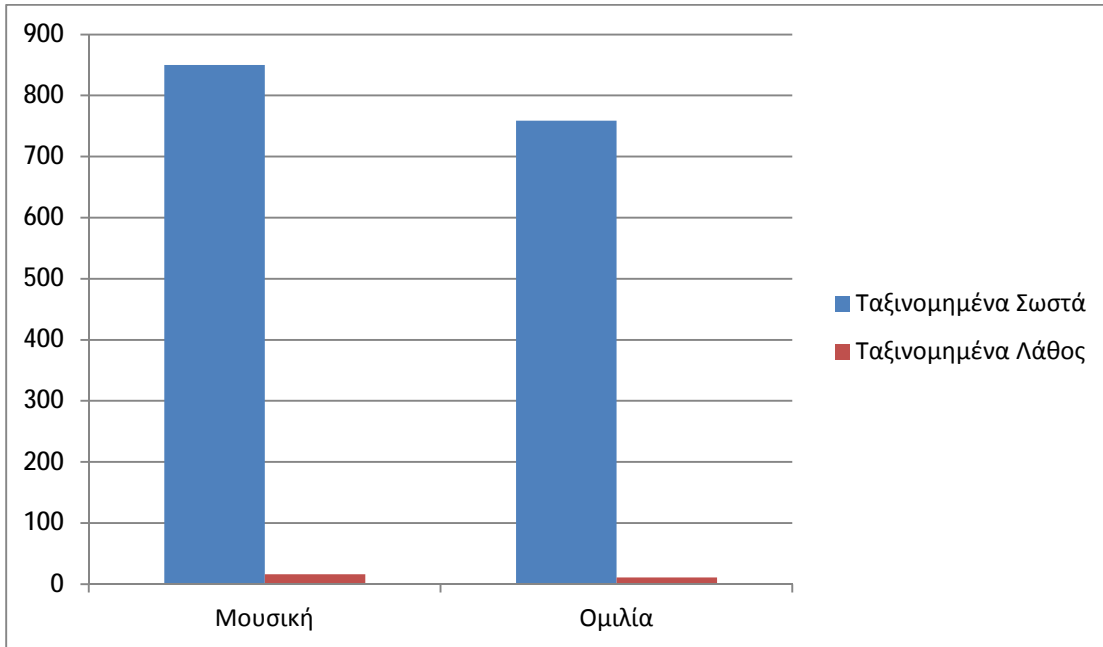
Πίνακας 53:Data MFCC



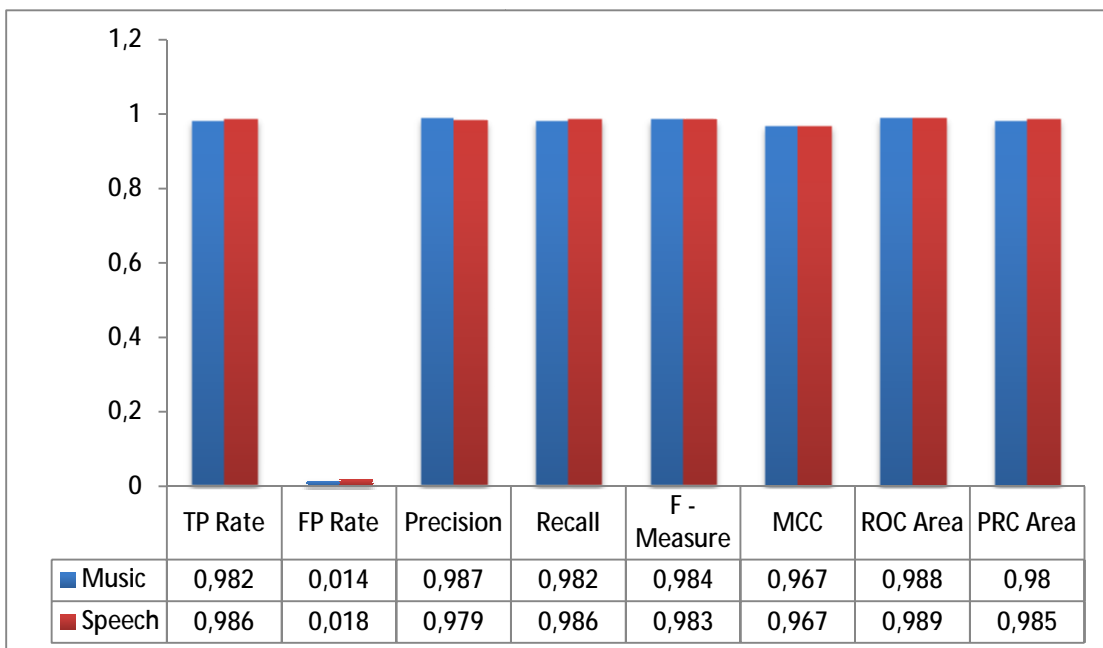
Πίνακας 54: Confusion Matrix Roll Off



Πίνακας 55: Data Roll Off



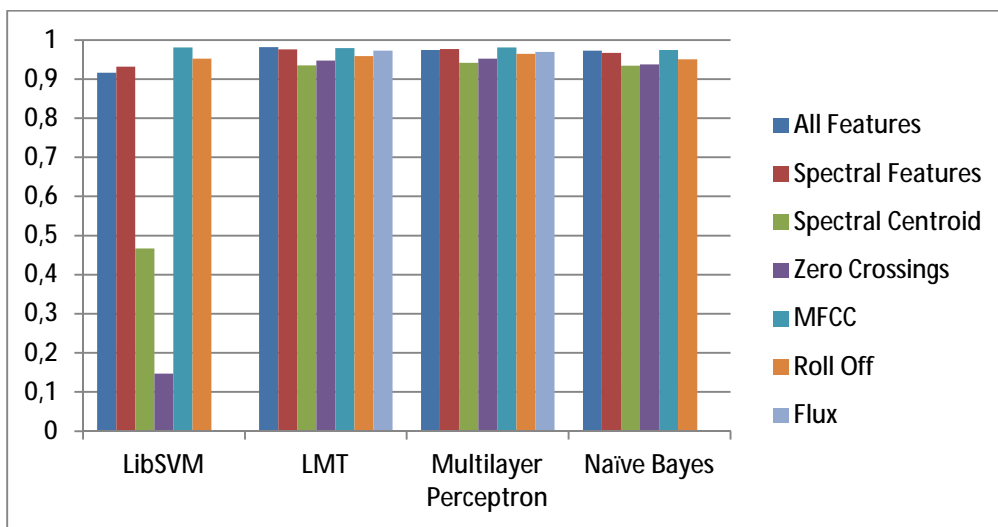
Πίνακας 56:Confusion Matrix Flux



Πίνακας 57:Data Flux

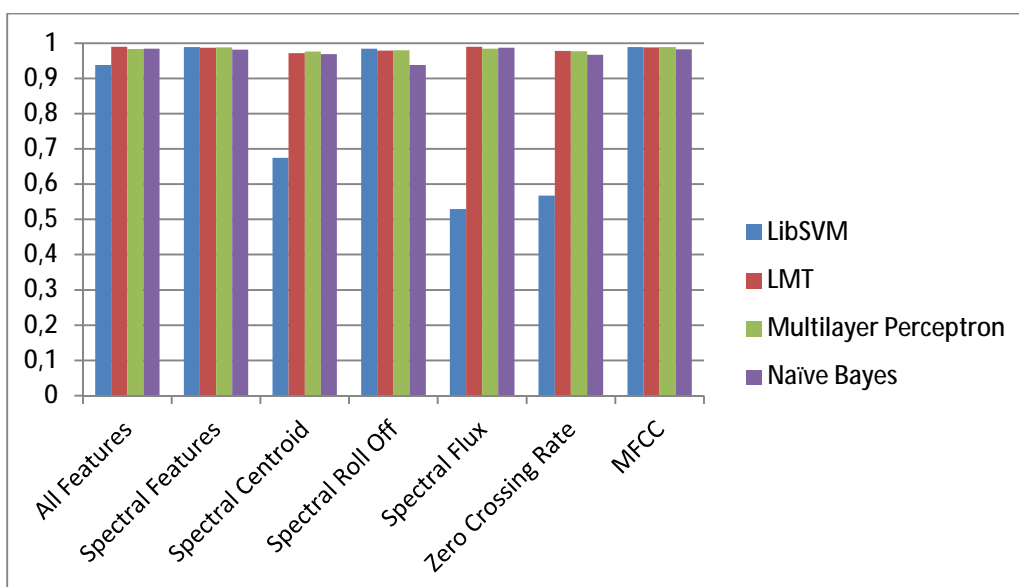
6.3 Συμπεράσματα – Προοπτικές εξέλιξης

Πίνακας 58: Γραφική Παράσταση Karra Statistics



Στον παραπάνω πίνακα βλέπουμε τις τιμές του συντελεστή Karra, όπου όπως αναφέρθηκε σε προηγούμενη ενότητα, όπου $k=1$ έχουμε τέλεια συμφωνία και όπου $k=0$ έχουμε ασυμφωνία. Οι τιμές είναι σχετικά καλές με εξαίρεση τον ταξινομητή SVM, στον οποίο υπάρχουν αρκετά χαμηλές τιμές. Πιο συγκεκριμένα, η φασματική ροή εδώ είναι 0, με δεύτερο χαμηλότερο το πλήθος μηδενισμού συνάρτησης και λίγο κάτω του μετρίου το φασματικό κέντρο βάρους.

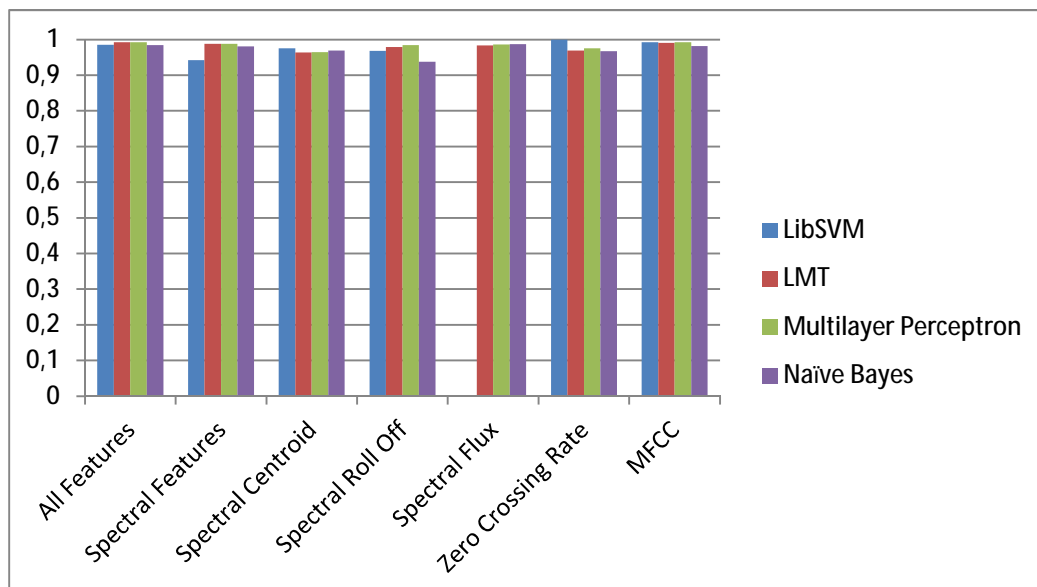
Πίνακας 59: Γραφική Παράσταση Precision-Μουσικής



Στην κατηγορία της μουσικής, συναντάμε το ίδιο πρόβλημα με πριν. Ενώ οι τιμές είναι αρκετά καλές, βλέπουμε πάλι πως ο αλγόριθμος SVM δεν έχει λειτουργήσει σωστά, δίνοντας χαμηλά

αποτελέσματα πάλι σε spectralcentroid, spectralflux και zerocrossingrate. Οι συντελεστές MelFrequency, έχουν δώσει την μεγαλύτερη ακρίβεια.

Πίνακας 60:Γραφική Παράσταση Precision-Ομιλίας



Στην ομιλία σε αντίθεση με την μουσική, τα αποτελέσματα μας είναι σχεδόν άριστα. Το χειρότερο αποτέλεσμα για ακόμη μία φορά έχει έρθει από τον SVM κ την φασματική ροή που είναι 0, αλλά έχουμε και την απόλυτη ακρίβεια στον ίδιο ταξινομητή, με το zerocrossin grate που αντιστοιχεί σε 1.

LibSVM

	All features	Spectral features	MFCC	Spectral centroid	Zero crossings	Flux	Roll off
Ποσοστό επιτυχίας ταξινομητή	95.8435 %	96.577 %	99.022 %	74.2054 %	59.5355 %	52.934 %	97.6161 %
Kappa Statistics	0.9163	0.9315	0.9804	0.4668	0.1473	0	0.9522
RAE	8.3416 %	6.8696 %	1.9627 %	51.7672 %	81.2083 %	94.4567 %	4.7842 %

LMT

	All features	Spectral features	MFCC	Spectral centroid	Zero crossings	Flux	Roll off
Ποσοστό επιτυχίας ταξινομητή	99.0831 %	98.7775 %	98.9609 %	96.7604 %	97.3716 %	98.6553 %	97.9218 %
Kappa Statistics	0.9816	0.9755	0.9791	0.935	0.9473	0.973	0.9583
RAE	3.9375 %	6.277 %	4.6393 %	11.4162 %	10.3996 %	6.914 %	7.9724 %

Multilayer Perceptron

	All features	Spectral features	MFCC	Spectral centroid	Zero crossings	Flux	Roll off
Ποσοστό επιτυχίας ταξινομητή	98.7164 %	98.8386 %	99.022 %	97.066 %	97.6161 %	98.4719 %	98.2274 %
Kappa Statistics	0.9742	0.9767	0.9804	0.9411	0.9522	0.9693	0.9644
RAE	2.7148 %	3.1749 %	2.5068 %	6.5627 %	5.998 %	4.5071 %	4.7349 %

Naïve Bayes

	All features	Spectral features	MFCC	Spectral centroid	Zero crossings	Flux	Roll off
Ποσοστό επιτυχίας ταξινομητή	98.6553 %	98.3496 %	98.7164 %	96.6993 %	96.8826 %	98.3496 %	97.555 %
Kappa Statistics	0.973	0.9669	0.9742	0.9338	0.9374	0.9669	0.9509
RAE	2.6988 %	3.3751 %	2.5761 %	8.9733 %	7.3179 %	3.6426 %	5.5657 %

Σύμφωνα με τους τελικούς πίνακες τα αποτελέσματα των ταξινομητών ήταν αρκετά ικανοποιητικά. Συγκεκριμένα, ο λιγότερο αποτελεσματικός ταξινομητής ήταν ο αλγόριθμος υποστήριξης διανυσματικών μηχανών (SVM) με μέσο όρο, το ποσοστό 68,3024 %. Αρκετά καλύτερα αποτελέσματα έδωσε ο αλγόριθμος Μπείσιανής κατηγοριοποίησης (NaïveBayes), με μέσο όρο 97,8868 %. Τέλος, οι αλγόριθμοι LogisticModelTree (LMT) και MultilayerPerceptron (MLP), έκαναν εξίσου καλή κατηγοριοποίηση με ποσοστά 98,2189% και 98,2797 % αντίστοιχα. Με μικρή διαφορά στο παρόν πείραμα ο ταξινομητής MLP ήταν η καλύτερη επιλογή.

Το συγκεκριμένο σύστημα έχει προοπτικές εξέλιξης και ιδιαίτερα σε ό,τι αφορά τον αλγόριθμο υποστήριξης διανυσματικών μηχανών όπου σε παλαιότερες μελέτες και πειράματα τρίτων, ήταν ο βασικός ταξινομητής και είχε υψηλό ποσοστό επιτυχίας. Μία μελλοντική εργασία θα

μπορούσε να είναι η ίδια μεθοδολογία, όμως με εφαρμογή σε real time ραδιοφωνικές εκπομπές (online). Είναι πιθανόν ο συνδυασμός ταξινομητών να επέφερε σημάδια βελτίωσης στο πείραμα, για το στάδιο της κατηγοριοποίησης. Τέλος, τα χαρακτηριστικά του ήχου ποικίλουν, οπότε υπάρχει πληθώρα συνδυασμών όπου θα μπορούσαν να χρησιμοποιηθούν ως εξαγόμενα χαρακτηριστικά και θα είχε ενδιαφέρον να δούμε αν θα υπήρχε διαφοροποίηση στα ως τώρα αποτελέσματα.

Βιβλιογραφικές Αναφορές

1. Bachu R. G., Kopparthi S., Adapa B., Barkana B. D. “Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal,” 2008, Electrical Engineering Department, School of Engineering, University of Bridgeport.
2. Benetos, E., Kotropoulos, C. “A tensor-based approach for automatic music genre classification,” Paper presented at the EUSIPCO 2008: 16th European Signal Processing Conference, 25-29 Aug 2008, Lausanne, Switzerland.
3. Bouckaert Remco R., Frank Eibe, Hall Mark, Kirkby, Reutemann, Seewald Alex, and Scuse David. “WEKA Manual For Version 3-7-8,” The University of Waikato, (January 2013).
4. Eugenio Di, Barbara, Glass Michael. “The Kappa statistic: a second look,” 2004.
5. Fawcett, Tom. “An introduction to ROC Analysis,” Institute for the Study of Learning and Expertise, Palo Alto, (December 2005).
6. Gael Richard, Mathieu Ramona and Slim Essid. “Combined supervised and unsupervised approaches for automatic segmentation of radiophonic audio streams,” 2007, GET-ENST, Paris, France.
7. Gershenson, Carlos. “Artificial Neural Networks for Beginners,” 2003.
8. Giannoulis Dimitrios, Massberg Michael, and Reiss Joshua D. “Parameter Automation in a Dynamic Range Compressor,” 2013, Queen Mary University of London UK.
9. Gouyon Fabien, Pachet Francois, and Delerue Olivier. “On the use of zero-crossing rate for an application of classification of percussive sounds,” Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, (December 2000).
10. Kendrick Boyd, Eng Kevin H., and Page David C. “Area Under the Precision-Recall Curve: Point Estimates and Confidence Intervals,” 2013.
11. Kos Marco, Grasic Matej, Vlaj Damian and Kacic Zdravko. “On-line Speech/Music Segmentation for Broadcast News Domain,” 2009, Faculty of Electrical Engineering and Computer Science, University of Maribor.
12. Murphy, Kevin P. “Naïve Bayes classifiers,” (October 2006).
13. Nam, Unjung. “Special Area Exam Part II”, (April 2008), <https://ccrma.stanford.edu/~unjung/AIR/areaExam.pdf>
14. Panagiotakis C., Tziritas G. “A Speech/Music Discrimination Based on RMS and Zero-Crossings,” 2002, Department of Computer Science, University of Crete, Heraklion.
15. Pikrakis Aggelos, Giannakopoulos Theodoros and Theodoridis Sergios. “A computationally efficient speech/music discrimination for radio recordings,” 2006, Department of Informatics and Telecommunications, University of Athens.
16. Powers M. W., David. “Evaluation: From Precision, Recall and F-Factor to Roc, Informedness, Markedness & Correlation,” (December 2007), School of Informatics and Engineering Flinders University, Adelaide, Australia.
17. Rybach David, Gollan Christian, Schluter Ralf and Ney Hermann. “Audio segmentation for speech recognition using segment features,” 2009, Human Language Technology and Pattern Recognition Computer Science Department, RWTH Aachen University, Germany.
18. Saunders, John. “Real-Time Discrimination of Broadcast Speech/Music,” 1996, Sanders, A Lockheed Martin Co., Nashua.

19. Scheiner Eric, Slaney Malcolm. “Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator,” 1997, Interval Research Corp., Palo Alto.
20. Sharma Neha, Om Hari. “Comparing the Performance of Data Mining Tools: WEKA and DTREG,” *International Journal of Scientific & Engineering Research*, Volume 5, Issue 4, (April 2014).
21. Theofilis George-Nektarios, “Weka Classifiers Summary,” Athens University of Economics and Business, (November 2013)
22. Ζαρδαβάς, Ιωάννης. «Αυτόματη κατηγοριοποίηση ειδών κρητικής μουσικής με χρήση μεθόδων μηχανικής μάθησης», Ανώτατο Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης, Τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Ρέθυμνο, 2013.
23. Κυρίτσης, Κωνσταντίνος. «Νευρωνικά Δίκτυα και Μηχανές Διανυσματικής Υποστήριξης», Σχολή Θετικών Επιστημών – Τμήμα Μαθηματικών, Πανεπιστήμιο Πατρών (Μάιος 2014).
24. Σκαρλάτος, Δημήτρης. *Εφαρμοσμένη Ακουστική*, Γ΄ Έκδοση, Οκτώβριος 2008, Πάτρα, GOTSIS.
25. Χαδέλλης, Λουκάς. *Τεχνολογία Ήχου*, Μάρτιος 2010, Πανεπιστημιακές Εκδόσεις «ΑΡΑΚΥΝΘΟΣ».

Διαδικτυακές Πηγές

1. [https://en.wikipedia.org/wiki/Cross-validation_\(statistics\)#cite_note-1](https://en.wikipedia.org/wiki/Cross-validation_(statistics)#cite_note-1)
2. <http://www.cs.cmu.edu/~schneide/tut5/node42.html>
3. http://spu.fem.uniag.sk/cvicenia/ksov/fuskova-ulicna/Data%20mining/cvicenie8_classification/Cv8_classification_some%20interpretations.pdf
4. http://sovarr.c4dm.eecs.qmul.ac.uk/wiki/Spectral_Rolloff
5. https://en.wikipedia.org/wiki/Spectral_flux
6. https://en.wikipedia.org/wiki/Receiver_operating_characteristic
7. <http://www.egr.msu.edu/~aviyente/ece366pd09.pdf>