



ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΚΡΗΤΗΣ
ΠΑΡΑΡΤΗΜΑ ΡΕΘΥΜΝΟΥ
ΤΜΗΜΑ ΜΟΥΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΑΚΟΥΣΤΙΚΗΣ

ΣΥΣΤΗΜΑ ΑΥΤΟΜΑΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ
ΜΟΥΣΙΚΩΝ ΟΡΓΑΝΩΝ

ΝΙΚΟΛΑΟΣ ΣΤ. ΚΟΡΡΕΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων Καθηγητής: Δρ. Ηλίας Ποταμίτης

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της παρούσης πτυχιακής εργασίας, κύριο Ηλία Ποταμίτη, που οι απαντήσεις στα ερωτήματα μου και οι συμβουλές του ήταν πολύ χρήσιμες και ανεκτίμητες. Στην συνέχεια, θα ήθελα να ευχαριστήσω τους γονείς μου, Στυλιανό και Ιακωβίνα Κορρέ, για την συμπαράσταση τους όλα αυτά τα χρόνια τόσο στα υλικά όσο και στα πνευματικά και ηθικά αγαθά, καθώς και στην διαρκή και ανιδιοτελή υποστήριξη, αγάπη, διαπαιδαγώγηση και εμπύχωση που μου παρείχαν. Επίσης, θα ήθελα να ευχαριστήσω όλους όσους μου συμπαραστάθηκαν στη διάρκεια εκπόνησης της παρούσης πτυχιακής εργασίας.

Περίληψη

Ο κύριος σκοπός της συγκεκριμένης πτυχιακής εργασίας είναι η υλοποίηση και η αξιολόγηση ενός αυτόματου συστήματος αναγνώρισης μουσικών οργάνων. Σκοπός του συστήματος είναι να αναγνωρίσει το μουσικό όργανο από το σήμα μιας νότας του, το οποίο σήμα εισάγεται μέσω ενός μικροφώνου στο σύστημα. Το πρόγραμμα υλοποιήθηκε με τεχνικές που χρησιμοποιούνται στην Αναγνώριση Ομιλητή και Ομιλίας.

Τα όργανα που καλείται να αναγνωρίσει είναι το πιάνο, το ξυλόφωνο, η φλογέρα, η κιθάρα και η φουσαρμόνικα. Ο ταξινομητής που χρησιμοποιήθηκε ήταν η Μίξη Γκαουσιανών Μοντέλων ή Gaussian Mixture Models(GMM) και τα χαρακτηριστικά ως προς τα οποία εξετάστηκαν τα δεδομένα της εκπαίδευσης και της εξέτασης είναι τα χαρακτηριστικά MFCC, delta και delta-delta. Η χρήση αυτών των χαρακτηριστικών, δίνει αρκετά ικανοποιητικά αποτελέσματα και συγκεκριμένα η επιτυχής αναγνώριση(επί τοις %) των μουσικών οργάνων από το σήμα της νότας, που εισάγεται στο σύστημα μέσω του μικροφώνου, είναι 92.65%. Χρήση λιγότερων χαρακτηριστικών μειώνει την απόδοση του συστήματος. Επίσης, αλλαγές σε κάποιες παραμέτρους(όπως το μήκος του frame), μειώνουν την απόδοση. Οι παράμετροι που μεταβλήθηκαν καθώς και τα αποτελέσματα της επιτυχούς αναγνώρισης του συστήματος παραθέτονται στο Κεφάλαιο 6°.

Λέξεις Κλειδιά:

Μουσικά όργανα, Αναγνώριση μουσικών οργάνων, Ψηφιακή Επεξεργασία Σήματος, Μίξη Γκαουσιανών Μοντέλων, Gaussian Mixture Models, MFCC, delta, delta-delta, εξαγωγή χαρακτηριστικών, ξυλόφωνο, φλογέρα, κιθάρα, πιάνο, φουσαρμόνικα, αναγνώριση προτύπων, φυσιολογία αυτιού.

Abstract

The main goal of this thesis is the development and evaluation of an automatic musical instrument recognition system. Aim of the system is to recognize a musical instrument from a note's signal, which signal is imported via a system's microphone. The system is developed with techniques that are used in Speaker and Speech Recognition.

The instruments, which are called to recognize are piano, xylophone, recorder, guitar and harmonica. The classifier that has been used is the Gaussian Mixture Models (GMM) and the features as for the training and testing data are MFCC, delta and delta-delta. Use of these features gives very good results and correct recognition of musical instruments (in percent) from the note's signal, which imported to system via microphone, is 92.65%. Use of less features decrease the efficiency of the system.

Also, changes in some parameters (such as framesize) decrease the efficiency. The parameters that were changed as well as the results of correct recognition of the system are reported in Chapter 6.

Keywords:

Musical instruments, Musical instruments Recognition, Digital Signal Processing, Gaussian Mixture Models, MFCC, delta, delta-delta, feature extraction, xylophone, recorder, guitar, piano, harmonica, pattern recognition, ear physiology.

ΠΕΡΙΕΧΟΜΕΝΑ

<i>ΕΙΣΑΓΩΓΗ</i>	8
ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ	
<i>ΑΚΟΥΣΤΙΚΗ ΤΩΝ ΜΟΥΣΙΚΩΝ ΟΡΓΑΝΩΝ ΚΑΙ ΦΥΣΙΟΛΟΓΙΑ ΤΟΥ ΑΥΤΙΟΥ</i>	13
1.1 Παραγωγή του ήχου και μουσικά όργανα	13
1.2 Βασικές κατηγορίες μουσικών οργάνων	15
1.2.1 Έγχορδα μουσικά όργανα	15
1.2.2 Αερόφωνα μουσικά όργανα	20
1.2.3 Ιδιόφωνα μουσικά όργανα	23
1.3 Επιλογή πέντε μουσικών οργάνων των παραπάνω κατηγοριών	27
1.3.1 Κιθάρα	27
1.3.2 Πιάνο	30
1.3.3 Φλογέρα(Recorder)	32
1.3.4 Ξυλόφωνο	34
1.3.5 Φυσαρμόνικα	35
1.4 Η φυσιολογία του αυτιού	37
ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ	
<i>ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ</i>	47
2.1 Βασικά της ψηφιακής επεξεργασίας σήματος	47
2.2 Φίλτρα	50
2.3 Παράθυρα και Πλαίσια	54
2.4 Μετασχηματισμός Fourier και Διακριτός Μετασχηματισμός Fourier	58
2.4.1 Μετασχηματισμός Fourier	60
2.4.2 Διακριτός Μετασχηματισμός Fourier	61
2.5 Εξαγωγή Χαρακτηριστικών	63
ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ	
<i>ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ</i>	71
3.1 Εισαγωγή στην Αναγνώριση Προτύπων	71
3.2 Ταξινόμηση προτύπων	72
3.2.1 Ντετερμινιστική ταξινόμηση προτύπων	69
3.2.1.1 Γεωμετρικές αποστάσεις	69
3.2.1.2 Γραμμικός διαχωρισμός και γραμμικές διακρίνουσες συναρτήσεις	75
3.2.1.3 Τεχνητά Νευρωνικά Δίκτυα	76
3.2.1.4 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)	81
3.2.2 Στατιστική ταξινόμηση προτύπων	77
3.2.2.1 Gaussian Mixture Models	861
ΚΕΦΑΛΑΙΟ ΤΕΤΑΡΤΟ	
<i>ΣΥΣΤΗΜΑΤΑ ΑΝΑΓΝΩΡΙΣΗΣ</i>	86
4.1 Συστήματα αναγνώρισης ηχητικής πηγής	86
4.1.1 Αναγνώριση περιβαλλοντικών ήχων	89
4.1.2 Αναγνώριση ομιλητή	89
4.1.3 Αναγνώριση μουσικών οργάνων	90
ΚΕΦΑΛΑΙΟ ΠΕΜΠΤΟ	
<i>ΑΥΤΟΜΑΤΟ ΣΥΣΤΗΜΑ ΑΝΑΓΝΩΡΙΣΗΣ ΜΟΥΣΙΚΩΝ ΟΡΓΑΝΩΝ</i>	93
5.1 Γενική λειτουργία του συστήματος	93
5.1.1 Στάδιο Εκπαίδευσης(Training Stage)	94
5.1.2 Στάδιο της Εξέτασης(Testing Stage)	96
5.2 Αναλυτικά τα στάδια της εκπαίδευσης και εξέτασης	97

5.2.1 End-Point Detection εκπαίδευσης	98
5.2.2 Εξαγωγή MFCC, delta, delta-delta χαρακτηριστικών και εκπαίδευση	102
5.2.3 Αυτόματη ηχογράφηση, αποθήκευση και εκπαίδευση των αρχείων ήχου...	105
5.2.4 End-Point Detection εξέτασης.....	110
5.3 Γραφικό Περιβάλλον Διεπαφής (ΓΠΔ) ή <i>Graphical User Interface (GUI)</i>	115
ΚΕΦΑΛΑΙΟ ΕΚΤΟ	
<i>ΠΕΙΡΑΜΑΤΙΚΑ ΚΑΙ ΣΥΓΚΡΙΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ</i>	117
6.1 Πειραματικά αποτελέσματα	117
6.2 Συγκριτικά αποτελέσματα	120
6.2.1 Χρονική διάρκεια frame.....	120
6.2.2 Αριθμός Συνιστωσών της Μίξης Γκαουσιανών Μοντέλων(ΓΜΜ)	124
6.2.3 Επιλογή χαρακτηριστικών.....	129
6.3 Συμπεράσματα	132
ΒΙΒΛΙΟΓΡΑΦΙΑ	133
ΠΑΡΑΡΤΗΜΑ Α.....	136

ΕΙΣΑΓΩΓΗ

Η φύση του ήχου είχε απασχολήσει τον άνθρωπο από την αρχαιότητα. Αναφέρεται ότι είχε γίνει μελέτη της φύσης του ήχου από τον Πυθαγόρα(570-497 π.Χ.), διατυπώνοντας ότι ο ήχος δημιουργείται από σώματα τα οποία ταλαντώνονται. Παρατήρησε ότι η μεταβολή του μήκους της χορδής ενός μονόχορδου μουσικού οργάνου αλλάζει το τονικό ύψος του ήχου και όρισε την οκτάβα. Στην συνέχεια, ο Αριστοτέλης(524-480 π.Χ.) διατύπωσε ένα ορισμό για την παραγωγή και την διάδοση του ήχου. Κατά το πέρασμα των χρόνων, μελετήθηκε η φύση του ήχου και από άλλους ερευνητές, όπως είναι ο Γάλλος φιλόσοφος Marin Mersene(1588-1648), Robert Boyle(1627-1691) , Galileo Galilei(1564-1642) και Isaac Newton(1643-1727).

Επίσης, στο πέρασμα των αιώνων, η μουσική αναπτύχθηκε και μαζί με αυτή και τα μουσικά όργανα, για να εξυπηρετούνται οι εκτελεστές αλλά και να μεταβάλλεται το ηχόχρωμα των έργων των συνθετών. Αν και ο ήχος, η μουσική και τα μουσικά όργανα είχαν μελετηθεί από την αρχαιότητα, η καταγραφή και η αναπαραγωγή του ήχου από κάποια συσκευή άργησε να εμφανιστεί. Πιο συγκεκριμένα στον 18^ο και 19^ο αιώνα γίνεται πιο συστηματική μελέτη της φύσης του ήχου και της ομιλίας και γίνεται έρευνα για την δημιουργία συσκευών για την καταγραφή του ήχου και την αναπαραγωγή του. Πρώτοι οι De Kempelien(1791) και Leon Scott(1817-1879), έκαναν το βήμα της καταγραφής του ήχου, και οδήγησαν τον Thomas Alva Edison(1847-1931) στην εφεύρεση του πρώτου φωνογράφου το 1877, το οποίο κατέγραψε τον ήχο και την ομιλία σε φύλλο αλουμινίου ή κερί για την μετέπειτα αναπαραγωγή του. Στην συνέχεια η ανακάλυψη του τηλεφώνου, έδωσε έναυσμα νέων ερευνών για την καταγραφή και αναπαραγωγή του ήχου με χρήση του ηλεκτρικού ρεύματος. Στην συνέχεια η ανάπτυξη συσκευών καταγραφής, αναπαραγωγής, μετάδοσης , λήψης ήχου αλλά και εικόνας ήταν ραγδαία. Δεν μπορούμε να μην αναφέρουμε και το Nicola Tesla(1856-1943) έναν από τους μεγαλύτερους εφευρέτες, που οι μελέτες και οι ανακαλύψεις του στην επιστήμη του ηλεκτρισμού άλλαξαν τον αιώνα στον οποίο ζούμε. Μέσα στις εφευρέσεις του είναι η γεννήτρια υψηλών συχνοτήτων(1890) και το πηνίο Tesla(1891), το οποίο είναι ένας μετασχηματιστής με σημαντικές εφαρμογές στο πεδίο των ραδιοεπικοινωνιών.

Στον 20^ο αιώνα έγινε η μεγάλη εξέλιξη της επιστήμης του ηλεκτρισμού, της ηλεκτρονικής, των Η/Υ και της Ψηφιακής Επεξεργασίας Σήματος, όπου κατέστησαν δυνατή την ψηφιακή καταγραφή, αναπαραγωγή, επεξεργασία , συμπίεση, αποθήκευση και κωδικοποίηση του ήχου.

Η μελέτη της αυτόματης αναγνώρισης της φωνής έγινε στα εργαστήρια AT&T's Bell το 1936. Όμως , η έρευνα χρηματοδοτήθηκε και πραγματοποιείτο σε Πανεπιστήμια και για λογαριασμό της Αμερικανικής Κυβέρνησης. Μόλις στην δεκαετία του 80', η τεχνολογία αυτή απασχόλησε την αγορά. Κατά την διάρκεια του 20^{ου} αιώνα έγινε χρήση κάποιων τεχνικών και μεθόδων στην βελτίωση της αναγνώρισης της φωνής, όπως διάφορα χαρακτηριστικά(LPC, cepstrum, spectrum, cochlear κ.α.) χρησιμοποιήθηκαν στην αναγνώριση την δεκαετία του 60'. Την δεκαετία του 90' έγινε η χρήση της Μίξης Γκαουσιανών Μοντέλων(GMM), των Νευρωνικών δικτύων, αλλά και των Κρυφών Μαρκοβιανών Μοντέλων(HMM). Από την δεκαετία του 80' και έπειτα η ανάπτυξη και η βελτίωση της τεχνολογίας της αυτόματης αναγνώρισης φωνής και ήχου, έχει γνωρίσει ραγδαία ανάπτυξη. Σε αυτό το χρονικό

διάστημα έγινε και η προσπάθεια αυτόματης αναγνώρισης μουσικών οργάνων, η οποία χρησιμοποιεί τεχνικές και μεθόδους που είναι παρόμοιες με τις τεχνικές και μεθόδους που χρησιμοποιούνται στις εφαρμογές της αναγνώρισης ομιλίας και ηχητικής πηγής.

Η Αναγνώριση μουσικών οργάνων (Musical Instruments Recognition) είναι ένας επιστημονικός κλάδος, ο οποίος είναι υπό-πεδίο του επιστημονικού κλάδου της Αναγνώρισης Ηχητικής Πηγής (Sound-Source Recognition) και προσπαθεί να λύσει διάφορα προβλήματα, όπως της αναγνώρισης του είδους της μουσικής, της καταγραφής παρτιτούρας από ένα ηχητικό σήμα, της δομημένης κωδικοποίησης και άλλα εργαλεία για τους μουσικούς αλλά και εξαγωγή χρήσιμων συμπερασμάτων για άλλες εφαρμογές της αναγνώρισης ηχητικής πηγής, όπως είναι η αναγνώριση του ομιλητή (Speaker Recognition).

Σκοπός της παρούσας πτυχιακής εργασίας είναι η υλοποίηση και η αξιολόγηση ενός αυτόματου συστήματος αναγνώρισης μουσικών οργάνων, το οποίο είναι ικανό να αναγνωρίσει το μουσικό όργανο από το οποίο προέρχεται από μια 'απομονωμένη' νότα, ανάμεσα στα μουσικά όργανα της κιθάρας, πιάνου, ξυλοφώνου, φλογέρας και φουσαρμόνικας. Στο σύστημα χρησιμοποιείται η Μίξη Γκαουσιανών Μοντέλων ή Gaussian Mixture Models τόσο στην εκπαίδευση όσο και στην εξέταση των δεδομένων και ως χαρακτηριστικά ως προς τα οποία εξετάζονται τα δεδομένα της εκπαίδευσης και της εξέτασης είναι τα MFCC, delta και delta-delta χαρακτηριστικά. Οι τεχνικές, οι μέθοδοι, ο τρόπος εξαγωγής χαρακτηριστικών καθώς και η υλοποίηση του συστήματος θα εξηγηθούν με πιο λεπτομέρεια στα κεφάλαια που ακολουθούν.

ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ

ΑΚΟΥΣΤΙΚΗ ΤΩΝ ΜΟΥΣΙΚΩΝ ΟΡΓΑΝΩΝ ΚΑΙ ΦΥΣΙΟΛΟΓΙΑ ΤΟΥ ΑΥΤΙΟΥ

1.1 Παραγωγή του ήχου και μουσικά όργανα

Ο ήχος είναι ένα διάμηκες μηχανικό κύμα και δημιουργείται από την περιοδική μεταβολή της πίεσης ενός υλικού μέσου, στο οποίο διαδίδεται (π.χ. ένα μέσο που διαδίδεται ο ήχος είναι ο αέρας). Οι ήχοι, στους οποίους είναι ευαίσθητο το ανθρώπινο αυτί κυμαίνονται στις συχνότητες $20\text{Hz} \leq f \leq 20\text{KHz}$. Η περιοχή αυτή ονομάζεται ακουστική περιοχή και οι ήχοι σε αυτή την περιοχή ονομάζονται ακουστικοί ήχοι. Κάτω από αυτή την περιοχή, δηλαδή $f < 20\text{Hz}$ ονομάζονται υπήχοι (infrasounds) και πάνω από την ακουστική περιοχή, δηλαδή $f > 20\text{KHz}$ ονομάζονται υπέρηχοι (ultrasound).

Θεωρώντας ότι τα κύματα είναι επίπεδα (δηλαδή διαδίδονται προς μια κατεύθυνση π.χ. προς τον θετικό άξονα ενός ορθοκανονικού συστήματος συντεταγμένων), και τα οποία διαδίδονται σε ένα υλικό μέσο χωρίς απώλειες (π.χ. απορρόφηση) η πίεση του ήχου ικανοποιεί την σχέση:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \frac{\partial^2 p}{\partial x^2} \quad (1.1.1)$$

Η παραπάνω σχέση είναι η μονοδιάστατη κυματική εξίσωση, όπου αποδεικνύει την κυματική φύση του ήχου.

Ας θεωρήσουμε ότι ένα επίπεδο κύμα διαδίδεται προς μία κατεύθυνση στο επίπεδο και συγκεκριμένα στην διεύθυνση +x. Η στιγμιαία ηχητική πίεση του κύματος (δηλαδή η υπέρ-πίεση που δημιουργήθηκε στο μέσο διάδοσης) δίνεται από την σχέση

$$p(x,t) = p_0 e^{i(\omega t - kx)} \quad (1.1.2)$$

Η πραγματική ποσότητα της παραπάνω σχέσης είναι ίση με:

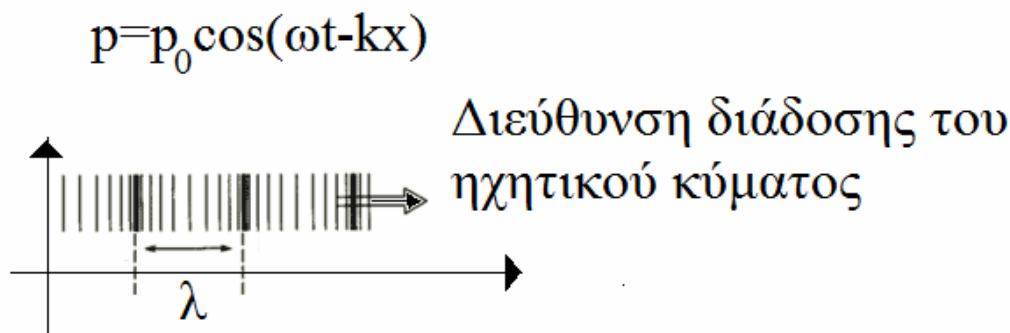
$$p(x,t) = p_0 \cos(\omega t - kx) \quad (1.1.3)$$

, όπου ω είναι η κυκλική γωνιακή συχνότητα, k ο κυματάριθμος ($k=2\pi/\lambda$), p_0 το μέγιστο πλάτος της ακουστικής πίεσης. Η σχέση 1.1.3 είναι η λύση της μονοδιάστατης κυματικής εξίσωσης για ένα επίπεδο κύμα το οποίο διαδίδεται προς τον άξονα $+x$.

Για σφαιρικό κύμα ισχύει:

$$p(r, t) = \frac{p_0}{r} \cos(\omega t - kr) \quad (1.1.4)$$

, όπου r η ακτινική απόσταση.



Εικόνα 1.1.1. Διάδοση επίπεδων ηχητικών κυμάτων.

Στο παραπάνω σχήμα φαίνεται η διάδοση ενός επίπεδου ηχητικού κύματος. Το μήκος κύματος λ είναι η απόσταση μεταξύ δυο διαδοχικών χαμηλών ή υψηλών μετώπων πίεσης και ισούται με $\lambda=c/f$ και μετριέται σε m, όπου c η ταχύτητα του ήχου (m/s) και f η συχνότητα του (Hz). Η περίοδος ενός ήχου δίνεται από την σχέση $T=1/f$ και μετριέται σε sec.

Η ταχύτητα διάδοσης του ήχου σε ένα υλικό μέσο είναι ίση με

$$c = \sqrt{\frac{B}{\rho}} \quad (1.1.5)$$

, όπου B η ελαστικότητα του μέσου και ρ η πυκνότητα του. Η παραπάνω σχέση δίνει ταχύτητα του ηχητικού κύματος ίση με: $c=331$ m/s, όταν το ηχητικό κύμα διαδίδεται στον αέρα και η θερμοκρασία του περιβάλλοντος είναι ίση με 0°C .

Τα όργανα που χρησιμοποιήθηκαν στην παρούσα εργασία είναι κλασική κιθάρα, φουσαρμόνικα, πιάνο, recorder (φλογέρα) και ξυλόφωνο. Πριν εξετάσουμε τα όργανα αυτά κάθε αυτά, θα ήταν εύλογο να εξετάσουμε κάποια βασικές ιδιότητες των κατηγοριών που ανήκουν.

1.2 Βασικές κατηγορίες μουσικών οργάνων

Η φυσική ακουστική ή μουσική ακουστική μελετά την ακουστική των μουσικών οργάνων και περιγράφει με έννοιες της Φυσικής την παραγωγή, την διάδοση και την αντίληψη των ήχων που προέρχονται από όργανα κάθε κατηγορίας μουσικών οργάνων. Οι βασικές κατηγορίες στις οποίες ταξινομούνται τα μουσικά όργανα είναι:

- Χορδόφωνα (Έγχορδα)
- Αερόφωνα (Πνευστά)
- Μεμβρανόφωνα (Κρουστά)
- Ιδιόφωνα (Αυτόφωνα, Αυτοηχούμενα)
- Ηλεκτρόφωνα

1.2.1 Έγχορδα μουσικά όργανα

Τα Χορδόφωνα ή Έγχορδα μουσικά όργανα έχουν βασικό στοιχείο τους μια ή περισσότερες χορδές. Ο ήχος, σε αυτά τα όργανα παράγεται από τις παλμικές δονήσεις μιας ή περισσότερων χορδών. Συνήθως, αυτές οι χορδές είναι τεντωμένες και προσαρμοσμένες σε ένα αντηχείο, οι ταλαντώσεις της ή των χορδών μεταδίδονται στο αντηχείο και αυτό με τη σειρά του πάλλεται και ενισχύει και 'εμπλουτίζει' τον ήχο που παράγεται από το όργανο. Οι χορδές είναι συνήθως κατασκευασμένες από μέταλλο, πλαστικές ίνες ή έντερα ζώων. Για την κατασκευή του αντηχείου συνήθως χρησιμοποιείται κάποιο ελαστικό και ανθεκτικό ξύλο.

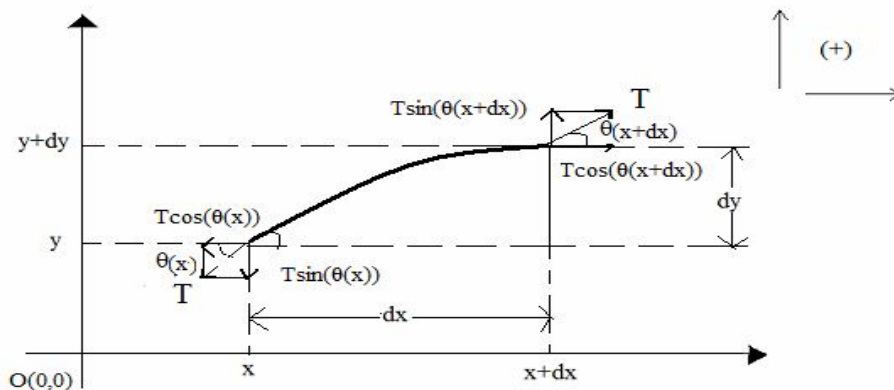
Ανάλογα με το πώς διεγείρονται οι χορδές για να παραχθεί ο ήχος, τα έγχορδα μουσικά όργανα διακρίνονται σε τρεις βασικές κατηγορίες:

- *Έγχορδα με τόξο*: Η διέγερση των χορδών προκαλείται από την τριβή τους με ένα δοξάρι ή τόξο, π.χ. βιολί, βιόλα κ.τ.λ.
- *Νοκτά ή νυσσόμενα έγχορδα*: Εδώ οι χορδές νύσσονται, 'τσιμπιούνται' απ' ευθείας με τα δάκτυλα του οργανοπαίκτη, π.χ. άρπα, κιθάρα κ.τ.λ. και
- *Κρουόμενα έγχορδα*: Οι χορδές σε αυτή την κατηγορία διεγείρονται από την κρούση ειδικών μπαγκετών ή μικρών σφυριών πάνω στις χορδές, π.χ. σαντούρι ή πιάνο, αντίστοιχα.

Οι χορδές είναι υλικά σώματα, όπου οι διάσταση που λαμβάνεται υπόψη περισσότερο είναι η διάσταση του μήκους. Οι ταλαντώσεις της χορδής είναι εγκάρσιες. Οι ιδιοσυχνότητες της χορδής, δίνονται από την σχέση:

$$f = \frac{n}{2L} \sqrt{\frac{T}{\mu}} \quad (1.2.1)$$

με $n=1,2,3,\dots$, T είναι η τάση της χορδής, L το μήκος της χορδής και μ η γραμμική πυκνότητα της χορδής, και $\sqrt{T/\mu}$ είναι η ταχύτητα του ήχου στην χορδή. Συνήθως οι χορδές είναι πακτωμένες σε δυο άκρα και αποτελούν την πιο απλή περίπτωση πηγής, όπου η πηγή ταλαντώνεται σε ένα άπειρο αριθμό συχνοτήτων.



Εικόνα 1.2.1 Μικρό τμήμα της χορδής.

Η κυματική εξίσωση ενός κύματος σε μια ιδανική χορδή μπορεί να υπολογιστεί από τον 2^ο νόμο του Νεύτωνα για ένα απειροελάχιστο τμήμα της χορδής. Το τμήμα της χορδής που φαίνεται στην παραπάνω εικόνα είναι πολύ μικρότερο σε σχέση με το συνολικό μήκος της χορδής και θεωρούμε ότι αυτό το μήκος του απειροελάχιστου τμήματος της χορδής είναι dx . Επίσης, θεωρούμε ότι η μετατόπιση y της χορδής είναι πολύ μικρή και η αλλαγή της τάσης λόγω της μετατόπισης αμελητέα.

Ως προς τον άξονα $x'x$: Η συνισταμένη των δυνάμεων ως προς τον άξονα $x'x$ είναι ίση με:

$$T\cos(\theta(x+dx))-T\cos(\theta(x)) \quad (1.2.2)$$

Θεωρώντας την γωνία θ πολύ μικρή, τόσο το $\cos(\theta(x+dx))$, όσο $\cos(\theta(x))$ είναι ίσες με την μονάδα. Έτσι, η διαφορά τους είναι περίπου ίση με το μηδέν και προσεγγιστικά θεωρούμε ότι δεν υπάρχει κίνηση του απειροελάχιστου τμήματος της χορδής στην διεύθυνση $x'x$ και κινείται μόνο ως προς τον άξονα $y'y$. Επομένως, θέλουμε να υπολογίσουμε την συνισταμένη των δυνάμεων στην διεύθυνση $y'y$.

Ως προς τον άξονα $y'y$: Η συνισταμένη των δυνάμεων ως προς τον άξονα $y'y$ είναι ίση με: $\Sigma F_y = dma$, δηλαδή $T\sin(\theta(x+dx))-T\sin(\theta(x)) = ma$. Και συμβολίζουμε το $T\sin(\theta(x+dx))-T\sin(\theta(x))$ με dF_y . Έτσι, η συνολική δύναμη που τείνει να επαναφέρει στην θέση ισορροπίας το στοιχειώδες τμήμα της χορδής είναι ίση με:

$dF_y = F_y(x+dx) - F_y(x)$. Η διαφορά των δυο τάσεων μπορεί να υπολογιστεί από το θεώρημα του Taylor, δηλαδή εφαρμόζοντας την σχέση:

$$F_y(x + dx) = F_y(x) + \frac{\partial F_y(x)}{\partial x} dx - F_y(x) \quad (1.2.3)$$

στην εξίσωση (1.2.4) έχουμε ότι:

$$dFy \approx Fy(x + dx) - Fy(x) = Fy(x) + \frac{\partial Fy(x)}{\partial x} dx - Fy(x) \quad (1.2.4)$$

$$dFy \approx \frac{\partial Fy(x)}{\partial x} dx \quad (1.2.5)$$

Η δύναμη $Fy(x)$ στο σημείο x , όπως φαίνεται και από το σχήμα είναι ίση με: $Fy(x) = T \sin\theta(x)$ και επειδή η μετατόπιση είναι μικρή, μπορούμε να γράψουμε ότι:

$$\sin\theta(x) \approx \tan\theta(x) \approx \frac{\partial y}{\partial x} \quad (1.2.6)$$

Άρα,

$$Fy(x) \approx T \frac{\partial y}{\partial x} \quad (1.2.7)$$

Επομένως η σχέση (1.2.7) μπορεί να γίνει:

$$dFy \approx \frac{\partial(T \frac{\partial y}{\partial x})}{\partial x} dx \quad (1.2.8)$$

Η οποία γίνεται ίση με:

$$\frac{\partial y}{\partial x} \quad dFy \approx \frac{T \partial^2 y}{\partial x^2} dx \quad (1.2.9)$$

Έτσι, ο 2^{ος} Νόμος του Νεύτωνα είναι ίσος με:

$$\Sigma Fy = dm \cdot a \Leftrightarrow$$

$$dFy = dm \cdot a \quad (1.2.10)$$

Λόγω ότι έχουμε θεωρήσει ότι κυρίαρχη διάσταση της χορδής είναι το μήκος της, μπορούμε να πούμε ότι η μάζα της χορδής είναι ίση με: $dm = \rho \cdot dV = \rho \cdot dx$.

Η επιτάχυνση είναι ίση με: $a = \frac{\partial^2 y}{\partial t^2}$

Επομένως,

$$\frac{T \partial^2 y}{\partial x^2} dx = \rho dx \frac{\partial^2 y}{\partial t^2} \quad (1.2.11)$$

Απαλείφοντας τους όρους dx , προκύπτει η κυματική εξίσωση της κίνησης της χορδής για εγκάρσια κύματα, η οποία είναι ίση με:

$$\frac{\partial^2 y}{\partial t^2} = \frac{T}{\rho} \frac{\partial^2 y}{\partial x^2} \quad (1.2.12)$$

, όπου T/ρ είναι ίσο με το τετράγωνο της ταχύτητας του ήχου στην χορδή, δηλαδή $c^2=T/\rho$. Έτσι, η παραπάνω σχέση γίνεται ίση με:

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}$$

, η οποία έχει την ίδια μορφή της εξίσωσης (1.1.1).

Η γενική λύση της κυματικής εξίσωσης της κίνησης της χορδής έχει την μορφή δυο ηχητικών κυμάτων τα οποία διαδίδονται προς αντίθετες κατευθύνσεις. Δηλαδή,

$$y(x,t)=f(ct-x)+g(ct+x) \quad (1.2.13)$$

Μπορούμε να θεωρήσουμε ότι η συναρτήσεις $f(ct-x)$, $g(ct+x)$ είναι εκθετικής μορφής και συγκεκριμένα:

$$f(ct-x)=Ae^{ik(ct-x)} \Leftrightarrow f(ct-x)=Ae^{i(\omega t-kx)} \text{ και } g(ct+x)=Be^{ik(ct+x)} \Leftrightarrow \\ g(ct+x)=Be^{i(\omega t+kx)} \quad (1.2.14)$$

ή ισοδύναμα

$$y(x,t)=(A\cos(\omega t)+B\cos(\omega t))(C\cos(kx)+D\sin(kx)) \quad (1.2.15)$$

, όπου A και B σταθερές που προσδιορίζονται από τις αρχικές συνθήκες και C και D σταθερές που προσδιορίζονται από τις συνοριακές συνθήκες.

Η μορφή της γενικής λύσης της κυματικής εξίσωσης εξαρτάται από τον τρόπο πάκτωσης της χορδής, από την θέση και τον τρόπο εφαρμογής της εφαρμοζόμενης δύναμης. Αυτός είναι και ο λόγος που η χορδή ενός πιάνου έχει διαφορετική φασματική ανάλυση(ηχόχρωμα) από την χορδή μιας κιθάρας, έστω και αν έχουν τα ίδια χαρακτηριστικά και εφαρμόζεται η ίδια τάση.

Έστω για παράδειγμα ότι έχουμε μια χορδή με πυκνότητα ρ και τα δυο άκρα της είναι πακτωμένα. Η χορδή έχει τάση T και η απόσταση ανάμεσα στα δυο άκρα που είναι στερεωμένα η χορδή είναι L .

Η γενική λύση της κυματικής εξίσωσης της κίνησης της χορδής, όπως είδαμε είναι ίση με την σχέση (1.2.14):

$$y(x,t)=(A\cos(\omega t)+B\cos(\omega t))(C\cos(kx)+D\sin(kx))$$

Για τις συνοριακές συνθήκες μπορούμε να γράψουμε ότι: Αφού είναι τα δύο άκρα της χορδής πακτωμένα, για το αριστερό και το δεξιό άκρο της χορδής για κάθε χρονική στιγμή t ισχύει ότι:

$$\text{Αριστερό άκρο}(x=0): y(0,t)=0 \Leftrightarrow (A\cos(\omega t)+B\cos(\omega t))(C\cos(kx)+D\sin(kx))=0 \Leftrightarrow \\ (A\cos(\omega t)+B\cos(\omega t))C=0 \Leftrightarrow C=0.$$

Άρα, $y(x,t)=(A\cos(\omega t)+B\sin(\omega t))D\sin(kx)$ ή θέτοντας $A'=AD$ και $B'=BD$,

$$y(x,t)=(A'\cos(\omega t)+B'\sin(\omega t))\sin(kx) \quad (1.2.16)$$

Δεξί άκρο ($x = L$): $y(L,t)=0 \Leftrightarrow (A\cos(\omega t)+B\sin(\omega t))\sin(kL)=0 \Leftrightarrow$

$$\sin(kL)=0 \Leftrightarrow kL=n\pi, n=1,2,3\dots \quad (1.2.17)$$

, όπου k είναι ο κυματάριθμος, που λαμβάνει διακριτές τιμές k_n που σχηματίζουν μια ακολουθία:

$$k_n=2\pi/\lambda_n \quad (1.2.18)$$

Έτσι, μπορούμε να γράψουμε ότι:

$$(2\pi/\lambda_n) L=n\pi \Leftrightarrow \lambda_n=2 L/n \Leftrightarrow c/f_n=2 L/n \Leftrightarrow$$

$$f_n = \frac{nc}{2L} \quad (1.2.19)$$

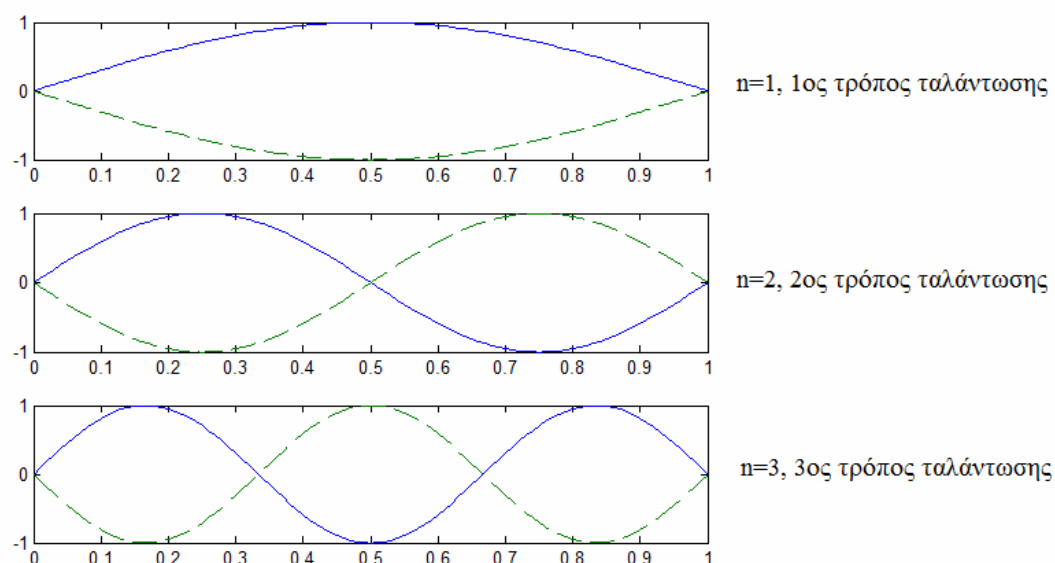
, όπου $c=\sqrt{T/\rho}$ και έτσι καταλήγουμε στην σχέση (1.2.1), που είναι ίση με :

$$f = \frac{n}{2L} \sqrt{\frac{T}{\mu}}$$

Για $n=1$ παίρνουμε την θεμελιώδη συχνότητα f_1 και για τις υπόλοιπες συχνότητες έχουμε ότι: $f_n= nf_1$. Έτσι, οι κανονικοί τρόποι ταλάντωσης(συχνότητες συντονισμού) δίνονται από την σχέση (1.2.15) αν αντικαταστήσουμε το k_n με ω_n/c , όπου $\omega_n=2\pi f_n$, δηλαδή:

$$y(x,t)=(A'\cos(\omega t)+B'\sin(\omega t))\sin(\omega_n x/c) \quad (1.2.20)$$

, όπου $k_n=\omega_n/c$. Οι κανονικοί τρόποι ταλάντωσης είναι αρμονικοί διότι: $f_n= nf_1$.



Εικόνα 1.2.2 Τρεις πρώτοι τρόποι ταλάντωσης.

1.2.2 Αερόφωνα μουσικά όργανα

Όπως είδαμε, η ακουστική βάση, των εγχόρδων οργάνων είναι οι ταλαντευόμενες χορδές. Η ακουστική βάση των αερόφωνων οργάνων(πνευστών) είναι οι ταλαντευόμενες αέριες στήλες.

Στις κατηγορίες των αερόφωνων μουσικών οργάνων συγκαταλέγονται οι ακόλουθες κατηγορίες:

- 1) *Τα ξύλινα πνευστά:* Σε αυτή την κατηγορία ανήκουν π.χ. τα φλάουτα, κλαρινέτο, φαγκότο, όμποε και χωρίζεται σε δύο επιμέρους κατηγορίες:
 - i) Τα φλαουτοειδή, όπου ο ήχος παράγεται από την διέγερση και ταλάντωση ενός αέριου πίδακα, ο οποίος με την σειρά του ταλαντώνει την αέρια στήλη που βρίσκεται εσωτερικά στον σωλήνα(φλάουτο, φλογέρα).
 - ii) Τα πνευστά που έχουν μονή ή διπλή γλωττίδα, όπου ο ήχος παράγεται από τις ταλαντώσεις της μίας ή των δύο γλωττίδων, αντίστοιχα(κλαρινέτο, όμποε).
- 2) *Τα χάλκινα:* όπου ο ήχος παράγεται από τις παλμικές δονήσεις(ταλαντώσεις) των χειλιών του οργανοπαίχτη.
- 3) *Τα πολύαυλα αερόφωνα και τα αερόφωνα με ελεύθερες γλωττίδες:* Στην κατηγορία αυτή τα πολύαυλα μουσικά όργανα ταλαντώνουν την αέρια στήλη χρησιμοποιώντας κάποιο μηχανισμό πλήκτρων(πληκτροφόρα αερόφωνα) είτε όχι(άσκαυλοι). Σε αυτή την κατηγορία ανήκουν το εκκλησιαστικό όργανο, το ακορντεόν(πολύαυλα αερόφωνα), οι παιδικές σβούρες και το γιο-γιο(αερόφωνα με ελεύθερες γλωττίδες).
- 4) *Η ανθρώπινη φωνή:* Η ανθρώπινη φωνή θεωρείται και αυτή ως αερόφωνο, όπου ο ήχος παράγεται από την ταλάντωση δυο φωνητικών χορδών(μυών) με την χρήση αέρα που εκπνέεται από τα πνευμόνια και οι χορδές πάλλονται σαν γλωττίδες και οι δονήσεις των, παράγουν ήχο. Ύστερα η αέρια στήλη που περιέχεται στο λαιμό,

, ταλαντώνεται και μαζί με τον λαιμό και άλλες κοιλότητες π.χ. στόμα, μύτη, κρανίο, επηρεάζουν, ενισχύοντας και ‘εμπλουτίζοντας’ τον παραγόμενο ήχο από τις φωνητικές χορδές.

Σημαντικό μέρος ενός πνευστού οργάνου είναι η γεωμετρική κατασκευή του. Συνήθως τα όργανα αυτά μπορούμε να πούμε ότι είναι σωλήνες που οι μορφές τους είναι παρεκκλίσεις κυλινδρικών ή κωνικών σωλήνων. Αυτές οι παρεκκλίσεις, όμως έχουν σημαντική μουσική σπουδαιότητα. Μέσα στους σωλήνες των μουσικών οργάνων δημιουργούνται τα στάσιμα κύματα. Τα πνευστά επίσης έχουν οπές, με τις οποίες ο οργανοπαίχτης μπορεί να ‘μεταβάλλει το μήκος του σωλήνα’, ώστε να μεταβάλλει την συχνότητα συντονισμού που καθορίζει το τονικό ύψος του μουσικού ήχου.

Στην χορδή, είδαμε ότι διαδίδονται εγκάρσια ηχητικά κύματα κατά μήκος της. Στον σωλήνα ενός πνευστού οργάνου, εσωτερικά υπάρχει μια αέρια στήλη, όπου μέσω αυτής διαδίδονται διαμήκη ηχητικά κύματα.

Η εξίσωση που περιγράφει την κίνηση των ηχητικών κυμάτων στην αέρια στήλη κατά μήκος του σωλήνα(στον οποίο περιέχεται η αέρια στήλη) είναι ίδια με την σχέση (1.1.1), η οποία είναι η μονοδιάστατη κυματική εξίσωση:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \frac{\partial^2 p}{\partial x^2}$$

, όπου $c^2 = \sqrt{B/\rho}$, όπου B είναι το μέτρο συμπίεστικότητας της αέριας στήλης και ρ είναι η πυκνότητα της αέριας στήλης.

Θεωρητικά, αν δημιουργηθούν δυο κύματα με ίσο πλάτος και συχνότητα, τα οποία διαδίδονται προς αντίθετες κατευθύνσεις μέσα στον σωλήνα, τότε θα δημιουργηθεί ένα διάμηκες στάσιμο κύμα.

Έτσι, στην περίπτωση δημιουργίας στάσιμων κυμάτων, η παραπάνω κυματική εξίσωση γίνεται ίση με την μονοδιάστατη εξίσωση του Helmholtz, η οποία δίνεται από την σχέση:

$$\frac{\partial^2 p(x)}{\partial x^2} + k^2 p(x) = 0 \tag{1.2.21}$$

Η γενική λύση της παραπάνω εξίσωσης είναι ίση με:

$$p(x) = A \cos(kx) + B \sin(kx) \tag{1.2.22}$$

, όπου A και B είναι αυθαίρετες σταθερές που προσδιορίζονται από τις συνοριακές συνθήκες. Οι σωλήνες μπορούν να χωριστούν σε Ανοιχτό-Ανοιχτό σωλήνα, Κλειστό-Ανοιχτό σωλήνα και Κλειστό-Κλειστό σωλήνα.

Ανοιχτός-Ανοιχτός σωλήνας.

Στον σωλήνα αυτού του τύπου τα δυο άκρα είναι κόμβοι πίεσης(κοιλίες μετατόπισης). Όπως, αναφέρεται και παραπάνω η σχέση (1.2.13) μπορεί να καταλήξει στην εξίσωση (1.2.14), η οποία μας δίνει τις συχνότητες στις οποίες δημιουργούνται στάσιμα κύματα:

$$f_n = \frac{nc}{2L}, \quad n=1,2,3\dots$$

,όπου $c = \sqrt{\frac{B}{\rho}}$ η ταχύτητα του ήχου στον αέρα και L το μήκος του σωλήνα. Ο

τύπος αυτός ισχύει και για Κλειστό-Κλειστό σωλήνα.

Για παράδειγμα, έστω ότι έχουμε ένα σωλήνα που και τα δυο άκρα του είναι ανοιχτά και έχει μήκος L . Στα άκρα του όπως υπάρχουν κόμβοι πίεσης και αναγκάζουν ένα επίπεδο ηχητικό κύμα να ανακλαστεί στο ένα άκρο του, να αλλάξει η φάση 180° και να συμβάλλει με το προσπίπτον ηχητικό κύμα, δημιουργώντας στάσιμα κύματα. Επειδή, στα δύο άκρα υπάρχουν κόμβοι πίεσης, αυτό σημαίνει ότι η πίεση στα άκρα αυτά θα είναι μηδενική, δηλαδή:

$$p(0)=0 \text{ και } p(L)=0$$

Χρησιμοποιώντας της λύση της μονοδιάστατης εξίσωση του Helmholtz, προκύπτει ότι:

$$p(x)=A\cos(kx)+B\sin(kx)$$

Αριστερό άκρο: $p(0)=0 \Leftrightarrow A=0$

Δεξί άκρο: $p(L)=0 \Leftrightarrow B\sin kL=0 \Leftrightarrow k_n L=n\pi \Leftrightarrow k_n=n\pi/L \Leftrightarrow$

$$f_n=nc/2L, \quad n=1,2,3\dots$$

, αφού $k_n=2\pi f_n/c$. Σε αυτές τις συχνότητες δημιουργούνται στάσιμα κύματα και οι συχνότητες συντονισμού είναι αρμονικές αφού: $f_n=nf_1$.

Κλειστός-Ανοιχτός σωλήνας

Σε αυτού του είδους τον σωλήνα στο άκρο που είναι ανοιχτό υπάρχει κόμβος πίεσης(μηδενική πίεση), ενώ στο κλειστό άκρο υπάρχει κοιλία πίεσης(μέγιστη πίεση) και οι συχνότητες στις οποίες δημιουργούνται τα στάσιμα κύματα, δίνονται από την σχέση:

$$f_n = \frac{nc}{4L}, \quad n=1,3,5\dots \quad (1.2.23)$$

,όπου $c = \sqrt{\frac{B}{\rho}}$ η ταχύτητα του ήχου στον αέρα και L το μήκος του σωλήνα.

Παρατηρούμε ότι μόνο οι περιττές αρμονικές 'επιζούν'.

Για παράδειγμα, έστω ότι έχουμε ένα σωλήνα που και ένα από τα δυο άκρα του είναι ανοιχτό και το άλλο κλειστό και έχει μήκος L . Στο ένα άκρο υπάρχει κόμβος πίεσης και αναγκάζει ένα επίπεδο ηχητικό κύμα να ανακλαστεί στο ένα άκρο του και να αλλάξει η φάση του κατά 180° και να συμβάλλει με το προσπίπτον ηχητικό κύμα, δημιουργώντας στάσιμα κύματα. Επειδή, στο κλειστό άκρο υπάρχει κοιλία πίεσης και στο αριστερό άκρο υπάρχει κόμβος πίεσης, οι συνοριακές συνθήκες γίνονται ίσες με:

$$dp(0)/dx=0 \text{ και } p(L)=0$$

Χρησιμοποιώντας της λύση της μονοδιάστατης εξίσωση του Helmholtz, προκύπτει ότι:

$$p(x)=A\cos(kx)+B\sin(kx)$$

$$\text{Αριστερό άκρο: } dp(0)/dx=0 \Leftrightarrow kB=0 \Leftrightarrow B=0$$

$$\text{Δεξί άκρο: } p(L)=0 \Leftrightarrow A\cos(kL)=0 \Leftrightarrow k_n L=(2n+1)\pi/2 \Leftrightarrow k_n=(2n+1)\pi/2L \Leftrightarrow$$

$$f_n=(2n+1)c/4L, \quad n=1,2,3,\dots$$

ή

$$f_n=nc/4L, \quad n=1,3,5,\dots$$

, όπου $k_n=2\pi f_n/c$. Σε αυτές τις συχνότητες δημιουργούνται στάσιμα κύματα και οι συχνότητες συντονισμού είναι περιττές αρμονικές αφού: $f_n=nf_1, \quad n=1,3,5,\dots$

1.2.3 Ιδιόφωνα μουσικά όργανα

Σε αυτή την κατηγορία μουσικών οργάνων η ακουστική βάση είναι οι μπάρες(ή ράβδοι). Ο τρόπος παραγωγής ήχου στα ιδιόφωνα μουσικά όργανα ποικίλει και μπορεί να είναι από κρούση, τριβή, ξύσιμο, νύξη, δόνηση κ.τ.λ. του οργάνου, που θα προκαλέσει την παλμική κίνηση του, ώστε να παραχθεί ο τελικός ήχος.

Στην κατηγορία των ιδιόφωνων μουσικών οργάνων ανήκουν το ξυλόφωνο, η Marimba, το μεταλλόφωνο, το τρίγωνο, φουσαρμόνικα, κ.τ.λ. Συνήθως, ο ήχος παράγεται κρούοντας τις μπάρες, από τις οποίες αποτελούνται αυτά τα μουσικά όργανα, με κάποιους επικρουστήρες, όπως μπαγκέτες ή σφυράκια ή στην περίπτωση της φουσαρμόνικας φυσώντας ή 'ρουφώντας' ρεύμα αέρα. Αποτέλεσμα είναι η δόνηση των μπαρών, μέσω των οποίων παράγεται ήχος.

Στην μπάρα υπάρχουν τόσο διαμήκεις δονήσεις όσο και εγκάρσιες. Έτσι, σε μια μπάρα παράγονται στάσιμα διαμήκη και εγκάρσια ηχητικά κύματα. Οι συχνότητες συντονισμού της μπάρας για τις εγκάρσιες ταλαντώσεις της δεν είναι αρμονικές μεταξύ τους, όπως οι συχνότητες συντονισμού που παράγονται από τις διαμήκεις ταλαντώσεις της μπάρας.

Η εξίσωση που περιγράφει την κίνηση των διαμηκών ηχητικών κυμάτων στην μπάρα κατά μήκος της μπάρας είναι η μονοδιάστατη κυματική εξίσωση:

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}$$

, όπου $c^2=E/\rho$, όπου E είναι το μέτρο ελαστικότητας Young της μπάρας και ρ είναι η πυκνότητα της.

Η γενική λύση της παραπάνω εξίσωσης είναι ίδιας με αυτή των χορδών.

Έστω, ότι δημιουργούνται στάσιμα κύματα κατά μήκος της μπάρας. Τότε, η παραπάνω εξίσωση, γίνεται ίση με την μονοδιάστατη εξίσωση του Helmholtz, η οποία είναι ίση με:

$$\frac{\partial^2 y(x)}{\partial x^2} + k^2 y(x) = 0$$

Η γενική λύση της παραπάνω εξίσωσης είναι ίση με:

$$y(x) = A \cos(kx) + B \sin(kx)$$

, όπου A και B είναι αυθαίρετες σταθερές που προσδιορίζονται από τις συνοριακές συνθήκες. Ως προς τις συνοριακές συνθήκες μπορούμε να διακρίνουμε τρεις περιπτώσεις α) Μπάρα με τα δύο άκρα ελεύθερα β) Μπάρα με το ένα άκρο ελεύθερο και το άλλο πακτωμένο και γ) Μπάρα με τα δύο άκρα πακτωμένα.

Μπάρα με τα δύο άκρα ελεύθερα

Σε αυτή την περίπτωση οι συχνότητες συντονισμού της μπάρας που οφείλονται από την δημιουργία στάσιμων κυμάτων από την συμβολή διαμηκών ηχητικών κυμάτων, δίνονται από την σχέση:

$$f_n = \frac{nc}{2L}, \quad n=1,2,3\dots$$

, όπου $c = \sqrt{E/\rho}$ η ταχύτητα του ήχου στην μπάρα και L το μήκος της μπάρας. Το στάσιμο κύμα στη θεμελιώδη συχνότητα έχει κοιλία μετατόπισης στις άκρες της μπάρας και δεσμό στην μέση της.

Ο τύπος αυτός ισχύει και για την περίπτωση που η μπάρα έχει και τα δύο άκρα της πακτωμένα. Αφού τα δυο άκρα είναι πακτωμένα, το στάσιμο κύμα στη θεμελιώδη συχνότητα θα έχει δεσμούς στις άκρες της μπάρας και κοιλία στην μέση της μπάρας.

Μπάρα με ένα άκρο ελεύθερο

Σε αυτή την περίπτωση οι συχνότητες συντονισμού της μπάρας δίνονται από την σχέση:

$$f_n = \frac{nc}{4L}, \quad n=1,3,5\dots$$

, όπου $c = \sqrt{E/\rho}$ η ταχύτητα του ήχου στην μπάρα και L το μήκος της μπάρας. Στην περίπτωση, της μπάρας με το ένα άκρο ελεύθερο, στη θεμελιώδη συχνότητα, το στάσιμο κύμα έχει κοιλία μετατόπισης στο ελεύθερο άκρο και δεσμό στο πακτωμένο άκρο.

Θα δώσουμε ένα παράδειγμα για την πρώτη περίπτωση. Έστω ότι έχουμε μια μπάρα που και τα δυο άκρα της είναι ελεύθερα και έχει μήκος L. Λόγω ότι η μπάρα ταλαντώνεται ελεύθερα στα άκρα της, δεν υπάρχουν ελαστικές δυνάμεις σε αυτά και έτσι ισχύει ότι: $F=0$, που ισοδυναμεί με $dy/dx=0$. Έτσι, μπορούμε να γράψουμε για τις συνοριακές συνθήκες ότι:

$$dy(x)/dx=0 \text{ και } dy(x)/dx=0$$

Χρησιμοποιώντας της λύση της μονοδιάστατης εξίσωση του Helmholtz, προκύπτει ότι:

$$y(x)=A\cos(kx)+B\sin(kx) \Leftrightarrow \\ dy(x)/dx=-kA\sin(kx)+ kB \cos(kx)$$

Αριστερό άκρο: $dy(0)/dx =0 \Leftrightarrow B=0$

Δεξί άκρο: $dy(L)/dx =0 \Leftrightarrow -kA\sin kL=0 \Leftrightarrow k_n L=n\pi \Leftrightarrow k_n=n\pi/L \Leftrightarrow$

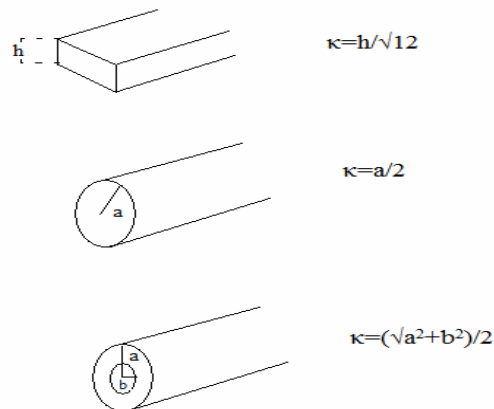
$$f_n=nc/2 L, n=1,2,3 \dots$$

, αφού $k_n=2\pi f_n/c$. Σε αυτές τις συχνότητες δημιουργούνται στάσιμα κύματα και οι συχνότητες συντονισμού είναι αρμονικές αφού: $f_n=nf_1$.

Όπως, έχει αναφερθεί παραπάνω, εκτός από διαμήκη κύματα στην μπάρα, διαδίδονται και εγκάρσια ηχητικά κύματα. Η εγκάρσια δόνηση της μπάρας είναι αρκετά πιο πολύπλοκη από τις προηγούμενες περιπτώσεις που έχουμε εξετάσει. Η εξίσωση που ισχύει για τα εγκάρσια κύματα στην περίπτωση της μπάρας είναι ίση με:

$$\frac{\partial^2 y}{\partial t^2} = -\kappa^2 c^2 \frac{\partial^4 y}{\partial x^4} \tag{1.2.24}$$

, όπου κ είναι η ακτίνα περιστροφής μιας διατομής και εξαρτάται από το σχήμα της μπάρας(βλέπε στην παρακάτω εικόνα), $\sqrt{c=E/\rho}$ η ταχύτητα του ήχου στην μπάρα.



Εικόνα 1.2.3. Το κ για διαφορετικά σχήματα .

Η παραπάνω εξίσωση διαφέρει από την κυματική εξίσωση για τα διαμήκη ηχητικά κύματα. Οι λύσεις της μορφής: $y(x,t)=f(ct-x)+g(ct+x)$ δεν επαληθεύουν την εξίσωση. Αυτό σημαίνει ότι τα εγκάρσια ηχητικά κύματα στην μπάρα δεν ταξιδεύουν στον άξονα $x'x$ με σταθερή ταχύτητα και με αμετάβλητο σχήμα. Η παραπάνω εξίσωση μπορεί να λυθεί με την μέθοδο της διαχώρισης των μεταβλητών χρησιμοποιώντας την μιγαδική εγκάρσια μετατόπιση από την θέση ισορροπίας $y=\Psi(x)e^{i\omega t}$ και την αντικαθιστούμε στην εξίσωση της κίνησης των εγκαρσίων κυμάτων στην μπάρα.

Η γενική λύση που προκύπτει είναι ίση με:

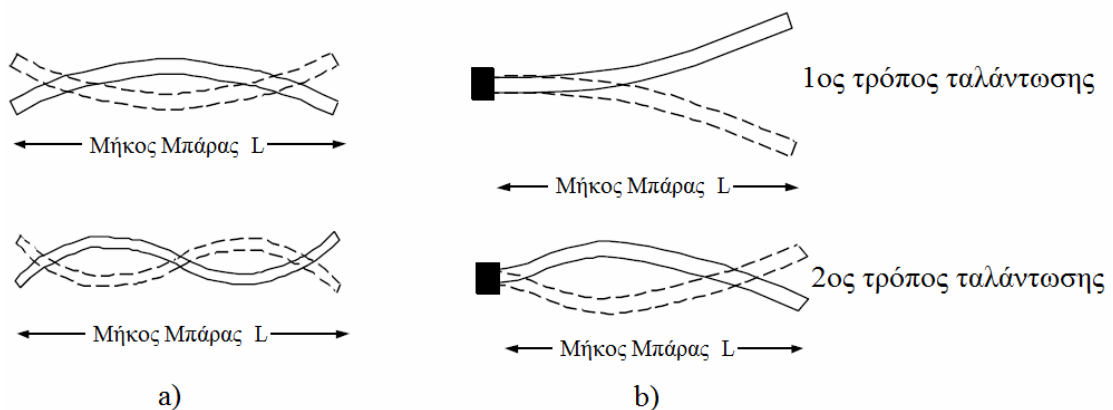
$$Y = e^{i\omega t} (Ae^{\omega x/v} + B e^{-\omega x/v} + Ce^{i(\omega x/v)} + De^{-i(\omega x/v)}) \quad (1.2.25)$$

, όπου **A**, **B**, **C**, **D** μιγαδικές σταθερές (μιγαδικά πλάτη) και $v = \sqrt{κωσ}$.

Για τα άκρα της μπάρας, όπως και την περίπτωση των διαμηκών κυμάτων υπάρχουν οι περιπτώσεις η μπάρα να είναι ελεύθερη και στα δύο άκρα, πακτωμένη και στα δύο άκρα και πακτωμένη στο ένα άκρο και το άλλο να είναι ελεύθερο. Θα πάρουμε μόνο την περίπτωση της μπάρας που έχει και τα δύο άκρα ελεύθερα. Οι συχνότητες συντονισμού της μπάρας με ελεύθερα και τα δύο άκρα δίνονται από την σχέση:

$$f_m = \frac{\pi c k m^2}{8L^2}, m = 3, 5, 7, (2n+1) \dots \quad (1.2.26)$$

, όπου c είναι η ακτίνα περιστροφής μιας διατομής, $\sqrt{c} = E/\rho$ η ταχύτητα του ήχου στην μπάρα και L το μήκος της μπάρας. Όπως, φαίνεται και στην παραπάνω εξίσωση οι συχνότητες συντονισμού της μπάρας δεν είναι αρμονικές γιατί δεν ισχύει $f_m = m f_1$, $m = 1, 2, 3 \dots$, αλλά $f_m = m^2 f_1$, $m = 3, 5, 7 \dots$



Εικόνα 1.2.4. Οι εγκάρσιες ταλαντώσεις της μπάρας, όταν έχουν δημιουργηθεί στην μπάρα στάσιμα κύματα α) για μπάρα και με ελεύθερα τα δυο άκρα και β) για μπάρα με το ένα άκρο πακτωμένο.

Στις περισσότερες μπάρες οι συχνότητες συντονισμού για τις διάμηκες ταλαντώσεις της μπάρας είναι σε πολύ υψηλότερες συχνότητες από τις συχνότητες συντονισμού για τις εγκάρσιες ταλαντώσεις.

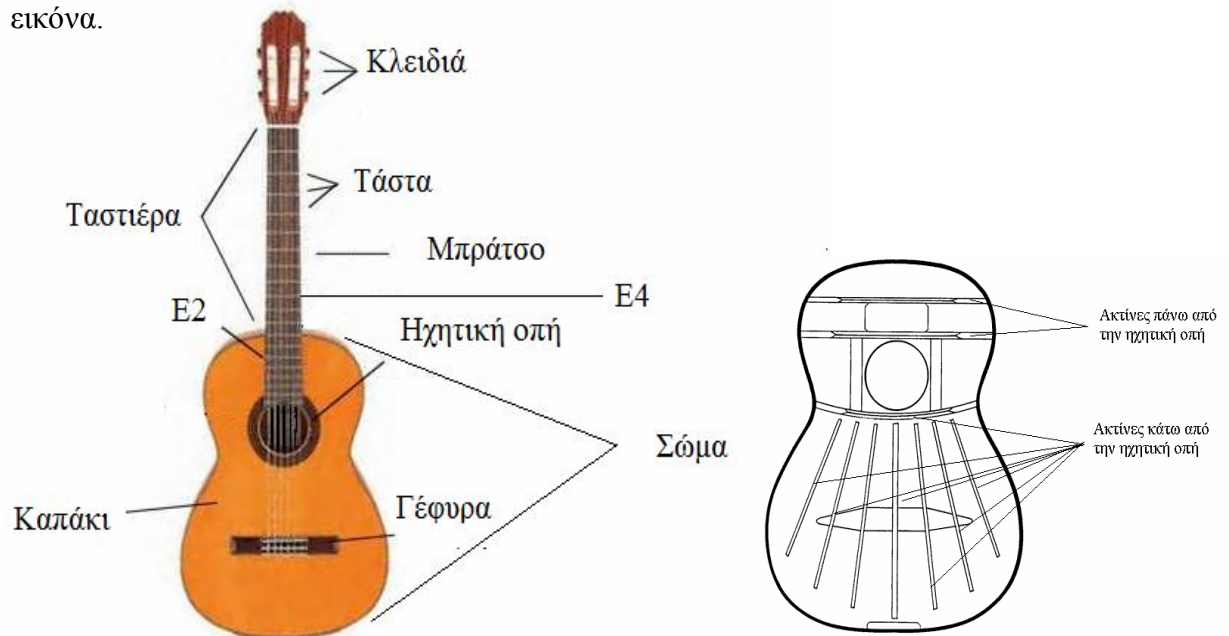
1.3 Επιλογή πέντε μουσικών οργάνων των παραπάνω κατηγοριών

1.3.1 Κιθάρα

Η κιθάρα εμφανίστηκε, όπως και το λαούτο, την εποχή του Μεσαίωνα στη Ισπανία, προερχόμενη από τους Άραβες. Η κιθάρα, καθώς και το λαούτο είχαν διαδοθεί στην Ευρώπη από τα τέλη του 14^{ου} αιώνα και συνυπήρχαν για αρκετά χρόνια. Η κιθάρα, λόγω του πιο εύκολου παιξίματος χρησιμοποιείτο στην λαϊκή μουσική, ενώ το λαούτο και η βιουέλα (σχήμα κιθάρας με πλήθος χορδών και με τρόπο παιξίματος του λαούτου) στην έντεχνη μουσική. Με το πέρασμα του χρόνου η δημοτικότητα του λαούτου και της βιουέλας εξασθένιζε, ενώ η δημοτικότητα της κιθάρας είναι αρκετά μεγάλη μέχρι και στις μέρες μας.

Υπάρχουν πολύ τύποι κιθάρας, από τους οποίους η ισπανική κιθάρα (με πέντε χορδές) προσεγγίζει την σύγχρονη που έχει έξι χορδές. Η έκτη χορδή προστέθηκε κατά την διάρκεια του 18^{ου} αιώνα.

Η κλασική κιθάρα ανήκει στην κατηγορία των έγχορδων μουσικών οργάνων και παίζεται τραβώντας τις χορδές με τον αντίχειρα ή τα δάκτυλα του χεριού ή με μια πένα. Ενώ το ένα χέρι διεγείρει την χορδή, το άλλο πιέζει την χορδή σε διάφορα σημεία της, αυξομειώνοντας το μήκος των χορδών αλλάζοντας τονικό ύψος. Τα βασικά μέρη από τα οποία αποτελείται η κλασική κιθάρα είναι το σώμα, οι χορδές και η ταστιέρα. Το πλήθος των χορδών της είναι έξι και οι θεμελιώδεις συχνότητες κάθε μίας, είναι a) 82Hz (E2), b) 110Hz (A2), c) 146Hz (D3), d) 196Hz (G3), e) 246Hz (B3), f) 329Hz (E4), ξεκινώντας από την χορδή που έχει την μεγαλύτερη πυκνότητα μέχρι την χορδή που έχει την μικρότερη πυκνότητα, δηλαδή από την χορδή που βρίσκεται πιο αριστερά μέχρι την χορδή που βρίσκεται πιο δεξιά όπως φαίνεται στη παρακάτω εικόνα.



Εικόνα 1.3.1. Η κιθάρα και τα μέρη της.

Το σώμα της κιθάρας είναι ένα ξύλινο και έχει ένα κοίλο σχήμα που αποτελείται από την μπροστινή επιφάνεια(καπάκι) και την πίσω επιφάνεια(πλάτη). Η ταστιέρα προσαρμόζεται στο καπάκι και στα πλαϊνά μέρη της κιθάρας. Στην πίσω επιφάνεια της κιθάρας(πλάτη) τοποθετούνται λεπτές τραβέρσες, έτσι ώστε να την κάνουν άκαμπτη. Στο καπάκι της κιθάρας υπάρχει ένα κυκλικό άνοιγμα, το οποίο ονομάζεται ηχητική οπή. Στο καπάκι, στο πίσω τμήμα του τοποθετούνται ακτίνες πάνω και κάτω από την ηχητική οπή, όπου πάνω από την ηχητική οπή οι ακτίνες υποβαστών το μπράτσο και την ταστιέρα. Στο καπάκι της κιθάρας, το τμήμα, πάνω από την ηχητική οπή, είναι πολύ σταθερό(υψηλή ακαμψία), λόγω των ακτινών και της ταστιέρας. Η περιοχή του καπακιού που κυρίως ταλαντώνεται είναι κάτω από την ηχητική οπή και καλείται ελεύθερο τμήμα του καπακιού της κιθάρας. Η γέφυρα(καβαλάρης) τοποθετείται λίγο πιο πάνω από την μέση του ελεύθερου τμήματος και ενισχύει την ακαμψία του. Οι χορδές προσαρμόζονται στη γέφυρα, η οποία είναι κολλημένη στο καπάκι, και εκτείνονται μέχρι τα κλειδιά, περνώντας από την ταστιέρα.

Βασική λειτουργία κιθάρας.

Στο παίξιμο της κιθάρας, ο οργανοπαίχτης τραβάει προς μια διεύθυνση την χορδή και μετά την αφήνει ελεύθερη. Έτσι, η χορδή αρχίζει να ταλαντώνεται. Το πλάτος των ταλαντώσεων της χορδής είναι αρκετά μεγάλο, αλλά η χορδή έχει πολύ μικρή επιφάνεια και έτσι, εκπέμπεται πολύ μικρή ένταση ήχου. Όμως, ένα μεγάλο μέρος της ενέργειας ταλάντωσης της χορδής, μεταφέρεται στο καπάκι και έτσι μετατρέπεται σε ενέργεια ταλάντωσης του καπακιού. Βέβαια, περισσότερο ταλαντώνεται το ελεύθερο τμήμα του καπακιού της κιθάρας απ' ότι το πάνω τμήμα της ηχητικής οπής, λόγω μικρότερης ακαμψίας. Επίσης, μέρος της ενέργειας μεταφέρεται στα πλαϊνά μέρη της κιθάρας, στην πλάτη και στην αέρια κοιλότητα που περικλείεται στο εσωτερικό του σώματος της κιθάρας, η ταλάντωση των οποίων είναι μικρότερη σε σχέση με την ταλάντωση του ελεύθερου τμήματος του καπακιού. Η μεγάλη επιφάνεια του καπακιού το καθιστά ικανοποιητική πηγή εκπομπής ήχου. Έτσι, το σώμα της κιθάρας λειτουργεί σαν ενισχυτής του ασθενούς ήχου των χορδών. Όταν, μεταφέρεται μέρος της ενέργειας από τις χορδές στο σώμα, το σώμα δεν ενισχύει ομοιόμορφα όλες τις συχνότητες με αποτέλεσμα το φάσμα του εκπεμπόμενου ήχου του καπακιού να είναι διαφορετικό από το φάσμα της ταλαντευόμενης χορδής. Επιρροή στο εκπεμπόμενο ήχο της κιθάρας, έχει και ο τρόπος που διεγείρεται η χορδή. Όσο πιο μαλακά διεγείρεται η χορδή τόσο μικρότερο είναι το πλάτος των υψηλών αρμονικών του συνολικού φάσματος, ενώ όσο πιο δυνατά διεγείρεται τόσο μεγαλύτερο είναι το πλάτος των υψηλών αρμονικών. Επομένως, ανάλογα της αρχικής διέγερσης των χορδών από τον οργανοπαίχτη και την μηχανική απόκριση του σώματος, καθορίζεται η φασματική απόκριση του εκπεμπόμενου ήχου από την κιθάρα, δηλαδή το ηχόχρωμα και η ποιότητα της.

Χορδές

Οι χορδές είναι υλικά σώματα, όπου οι διάσταση που λαμβάνεται υπόψη περισσότερο είναι η διάσταση του μήκους. Οι ταλαντώσεις της χορδής είναι εγκάρσιες. Οι ιδιοσυχνότητες της χορδής, δίνονται από την σχέση:

$$f = \frac{1}{2L} \sqrt{\frac{T}{\mu}} n$$

με $n=1,2,3\dots$, T είναι η τάση της χορδής, L το μήκος της χορδής και μ η γραμμική πυκνότητα της χορδής, και $\sqrt{T/\mu}$ είναι η ταχύτητα του ήχου στην χορδή.

Αέρια κοιλότητα και σώμα της κιθάρας

Όπως, αναφέρθηκε σε προηγούμενη ενότητα, μέρος της ενέργειας από την ταλάντωση των χορδών μεταφέρεται στο καπάκι, το οποίο στην συνέχεια ταλαντώνεται. Επίσης, μέρος της ενέργειας μεταφέρεται και στην αέρια κοιλότητα της κιθάρας.

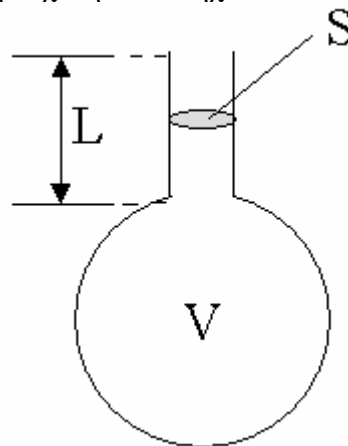
Για τους χαμηλότερους τρόπους ταλάντωσης η αέρια κοιλότητα της κιθάρας είναι ένα αντηχείο Helmholtz. Η συχνότητα συντονισμού της αέριας κοιλότητας αυτού του αντηχείου εξαρτάται από τον όγκο της κοιλότητας και από την διάμετρο της οπής και δίνεται από την σχέση:

$$f = \frac{c}{2\pi} \sqrt{\frac{S}{V \cdot l}} \tag{1.3.1}$$

ή

$$f = \frac{c}{2\pi} \sqrt{\frac{\pi r^2}{1.7 \cdot V \cdot r}} \approx \frac{c}{2\pi} \sqrt{\frac{1.8 \cdot r}{V}} \tag{1.3.2}$$

, όπου r είναι η ακτίνα της ηχητικής οπής, $S=\pi r^2$ είναι το εμβαδό της ηχητικής οπής, V είναι ο όγκος της αέριας κοιλότητας, $L(L=1.7r$ ισχύει για κιθάρα) είναι το μήκος του λαιμού και c είναι η ταχύτητα του ήχου στον αέρα.

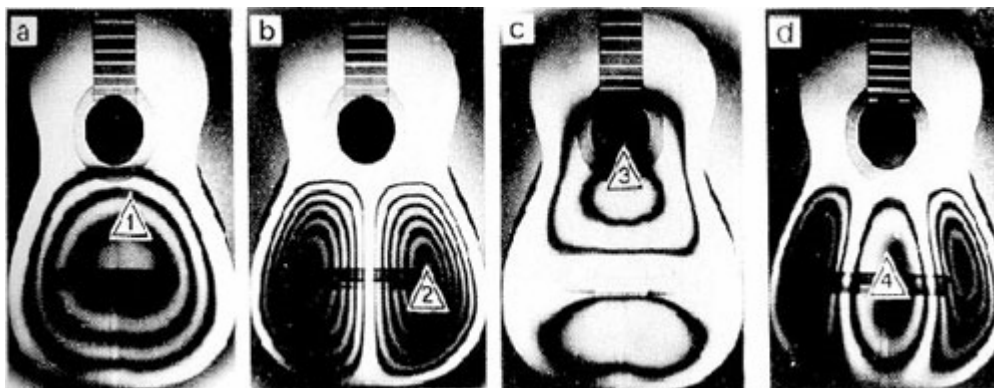


Εικόνα 1.3.2. Συντονιστής(Αντηχείο) Helmholtz.

Οι υψηλότεροι τρόποι δόνησης μοιάζουν με τα στάσιμα κύματα σε ένα παραλληλεπίπεδο δωμάτιο. Οι συντονισμοί αυτοί δεν μεταδίδονται μέσω του ανοίγματος(δεν εκπέμπουν ήχο), αλλά γίνεται σύζευξη αυτών των συντονισμών με

τους υψηλότερους τρόπους δόνησης(υψηλότερους συντονισμούς)του καπακιού της κιθάρας.

Όσον αφορά τους συντονισμούς του σώματος της κιθάρας, οι περισσότεροι από της χαμηλής συχνότητας συντονισμούς αποδίδονται σε συζευγμένες κινήσεις του καπακιού, της πλάτης και του περικλειόμενου αέρα στην κοιλότητα της κιθάρας. Επίσης, οι ταλαντώσεις είναι μεγαλύτερες στο ελεύθερο τμήμα του καπακιού της κιθάρας και είναι μικρού πλάτους κοντά στην γέφυρα εκτός από τον πρώτο συντονισμό.



Εικόνα 1.3.3. Τέσσερις πρώτοι συντονισμοί του καπακιού της κιθάρας και συγκεκριμένα a)185Hz b)287Hz c)460Hz και d)508Hz.Οι μαύρες καμπύλες δείχνουν σημεία της κιθάρας ίσου πλάτους ταλάντωσης και τα τρίγωνα δείχνουν τα σημεία διέγερσης του καπακιού της κιθάρας στην εκάστοτε περίπτωση.

Το συχνοτικό εύρος της κιθάρας κυμαίνεται προσεγγιστικά από 82Hz έως 4KHz, που είναι οι πολύ υψηλές αρμονικές που μπορούν να παραχθούν στην κιθάρα.

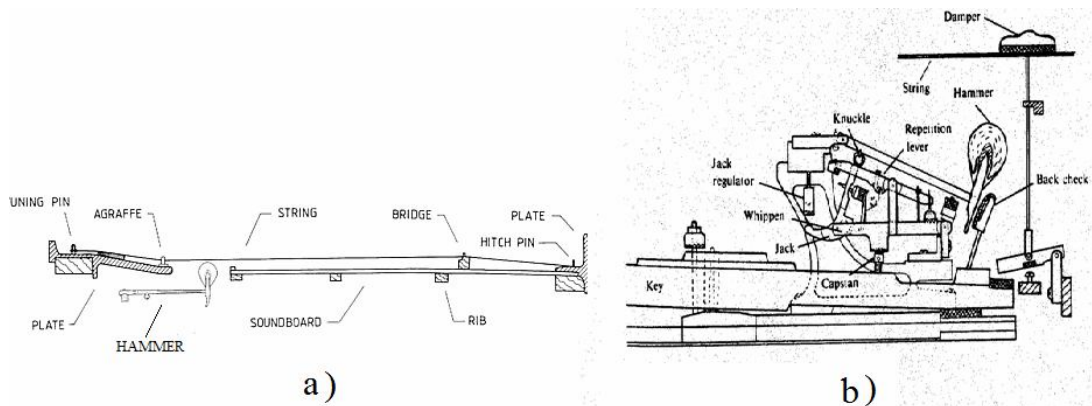
1.3.2 Πιάνο

Παρόλο που το πιάνο επινοήθηκε από έναν Ιταλό, οι κατασκευαστές που το βελτίωσαν αλλά και το εξέλιξαν στο πρώτο μισό του 18^{ου} αιώνα ήταν Γερμανοί. Πιο συγκεκριμένα ήταν ο Γκότφριντ Ζίλμπερμαν(Gottfried Silbermann), ο μαθητής του Γιοχάνες Τσούμπε(Johannes Zumpe) και ο Γιοχάνες Αντρέας Στάιν.(Johannes Andreas Stein). Οι συνεχείς βελτιώσεις στην κατασκευή του πιάνου συνεχίστηκαν τα επόμενα 100 χρόνια στην Ευρώπη και την Αμερική. Αυτές οι βελτιώσεις αφορούσαν κυρίως την μουσική έκταση του μουσικού οργάνου και την ενίσχυση του σκελετού(σώματος) του. Στις μέρες μας το πιάνο είναι ένα από τα πιο δημοφιλή μουσικά όργανα και χρησιμοποιείται είτε σαν σολιστικό είτε σαν συνοδευτικό μουσικό όργανο.

Στην συνέχεια θα αναφερθούμε συνοπτικά στο βασικό μηχανισμό παραγωγής του ήχου στο πιάνο.

Στο πιάνο η παραγωγή του ήχου γίνεται μέσω μια μεταλλικής χορδής, στην οποία εφαρμόζεται μια δύναμη από ένα σφυράκι(hammer). Η μεταλλική χορδή εκτείνεται μεταξύ δύο υποστηρίγμάτων υπό υψηλή τάση. Τα δύο υποστηρίγματα, τα οποία συγκρατούν την χορδή στα δύο άκρα της, ονομάζονται Agrafee(αριστερό άκρο) και

Hitch pin(δεξί άκρο). Αυτά στερεώνονται σε ένα μεταλλικό πλαίσιο που ονομάζεται πλάτη(plate). Κοντά στο δεξιό άκρο της, η χορδή περνάει από μια ξύλινη πλάκα, την γέφυρα(bridge), η οποία είναι κολλημένη σε μια μεγάλη και λεπτή ξύλινη επιφάνεια, που ονομάζεται καπάκι ή soundboard. Η τοποθεσία της γέφυρας είναι τέτοια ώστε η χορδή να ασκεί μια κατακόρυφη δύναμη στην γέφυρα και στο καπάκι. Ρόλος του καπακιού είναι η ενίσχυση του ασθενούς ήχου που παράγεται από την χορδή. Η χορδή διεγείρεται από ένα σφυράκι το οποίο αρχίζει να κινείται από το πάτημα ενός πλήκτρου του πιάνου, μέσω ενός πολύπλοκου συστήματος από μοχλούς, όπως φαίνεται στην παρακάτω εικόνα.



Εικόνα 1.3.4. α)Εσωτερικά μέρη πιάνου που είναι στηριγμένη η χορδή και επιδρούν στην τελική παραγωγή του ήχου του οργάνου και β) βασικός μηχανισμός για την κίνηση του σφυριού προς την χορδή και κατ' επέκταση την παραγωγή του ήχου στο πιάνο.

Η χορδή του πιάνου έχει πολύ μικρή σύζευξη με τον αέρα. Έτσι, τα ηχητικά κύματα που θα παραχθούν θα έχουν πολύ μικρή ένταση. Για την ενίσχυση τους χρειάζεται, αυτά να διαδοθούν στο καπάκι(soundboard) μέσω της γέφυρας(bridge), ώστε να ταλαντωθεί μεγαλύτερη ποσότητα αέρα και κατ' επέκταση ο ήχος να έχει υψηλότερη ένταση. Το περίεργο σχήμα της γέφυρας και η τοποθεσία της, εξασφαλίζουν ότι το καπάκι θα ταλαντωθεί ισχυρά σε όλες τις συχνότητες. Επίσης, ένα μοναδικό χαρακτηριστικό του πιάνου, συγκρινόμενο με το βιολί και το harpsichord, είναι ότι οι χορδές είναι συγκρατημένες στα άκρα τους, έχοντας πολύ υψηλή τάση. Αυτό έχει ως συνέπεια την αύξηση της αποθηκευμένης ενέργειας ταλάντωσης, αλλά και την αύξηση της δυναμικής ισχύος και εύρους του πιάνου.

Όπως, έχει αναφερθεί στην ενότητα που παρουσιάστηκαν τα έγχορδα όργανα, μια ιδανική χορδή(θεωρητικά, μηδενικής ακτίνας) θα ταλαντωθεί περισσότερο στις συχνότητες που είναι ακέραια πολλαπλάσια της θεμελιώδους, δηλαδή: $f_n = n f_1$, όπου $f_1 = c/2L$ με $c = \sqrt{T/\mu}$.

Μια πραγματική χορδή, δηλαδή πεπερασμένης ακτίνας, θα ταλαντωθεί στις συχνότητες που δίνονται από την σχέση:

$$f_n = n \cdot f_1 \left[\frac{1 + (n^2 - 1) \tau^4 \cdot \kappa}{T \cdot L^2} \right], \quad n = 1, 2, 3, \dots \quad (1.3.3)$$

, όπου τ είναι η ακτίνα της χορδής, T είναι η τάση της, κ είναι ένα μέτρο που δείχνει πόσο ισχυρή είναι η χορδή και L είναι το μήκος της χορδής.

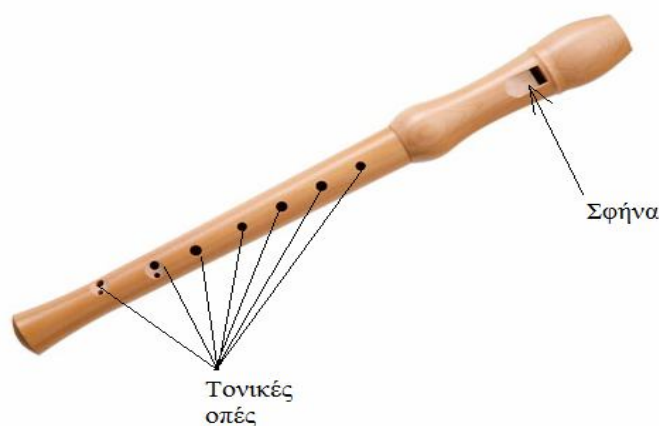
Οι συχνότητες που θα παραχθούν, δεν θα είναι τέλεια αρμονικές, αλλά αυτό το αποτέλεσμα είναι και το επιθυμητό στο πιάνο. Ο λόγος που χρειάζεται λίγη μη-

αρμονικότητα είναι ότι θα κάνει πιο πλούσιο ένα τόνο στην χροιά του και επίσης θα 'επισκιάσει' (mask) κάποιες νότες που είναι λίγο ξεκούρδιστες σε μια συγχορδία που θα παίζει ένας οργανοπαίχτης. Επίσης, ένα άλλο χαρακτηριστικό στο πιάνο είναι ότι οι νότες κάθε οκτάβας είναι κουρδισμένες στην δεύτερη αρμονική, σε σχέση με τις νότες της προηγούμενης(χαμηλότερης) οκτάβας. Το συχνοτικό εύρος του πιάνου είναι πολύ μεγάλο και κυμαίνεται προσεγγιστικά από 28Hz έως 4186Hz.

1.3.3 Φλογέρα(Recorder)

Τα ξύλινα πνευστά θεωρούνται από τα παλιότερα μουσικά όργανα, όπου τα πρώτα ευρήματα, που σχετίζονται με αυτά χρονολογούνται από την παλαιολιθική εποχή (περίπου 20000π.Χ.). Ορισμένα από αυτά μπορούμε να τα χαρακτηρίσουμε ως προγόνους των σημερινών φλάουτων, σε μια πρωτόγονη μορφή που χρησιμοποιούσαν κούφια κόκκαλα ζώων ανοίγοντας τους μια τρύπα και έτσι παρήγαν ένα και μοναδικό ήχο. Στην κατηγορία των ξύλινων πνευστών ανήκει το μουσικό όργανο recorder ή φλογέρα, όπως λέγεται στην ελληνική γλώσσα. Πολλές φορές συναντιέται και με το όνομα φλάουτο με ράμφος. Η φλογέρα χρησιμοποιείται σαν όργανο πάνω από 5000 χρόνια. Ευρήματα του τύπου της φλογέρας έχουν βρεθεί σχεδόν σε κάθε χώρα του κόσμου. Κατά την διάρκεια του Μεσαίωνα, η φλογέρα παίζονταν από περιπλανώμενους διασκεδαστές ή τροβαδούρους καθώς ταξίδευαν στην επαρχία. Στις φλογέρες συνηθιζόταν να παίζονται απλές μελωδίες για να συνοδεύουν χορό, επαναστάσεις ή εορταστικές εκδηλώσεις της εποχής. Η φλογέρα έφτασε στο ζενίθ της δημοτικότητας της κατά την διάρκεια της περιόδου της Αναγέννησης, όπου πολλά σπίτια είχαν πάνω από 20 φλογέρες το καθένα. Στα μέσα του 18^{ου} αιώνα αναπτύχθηκαν όργανα, όπως το φλάουτο ή πλαγιάυλος(transverse flute), όπου λόγω πλουσιότερου ηχοχρώματος και μεγαλύτερου συχνοτικού εύρους, χρησιμοποιείτο ως όργανο της συμφωνικής ορχήστρας. Έτσι, η φλογέρα έγινε σιγά σιγά λιγότερο δημοφιλές όργανο. Παρ' όλα αυτά η φλογέρα συνεχίζει να χρησιμοποιείται στην σύγχρονη folk μουσική και ενίοτε χρησιμοποιείται και σε ποπ-ροκ τραγούδια.

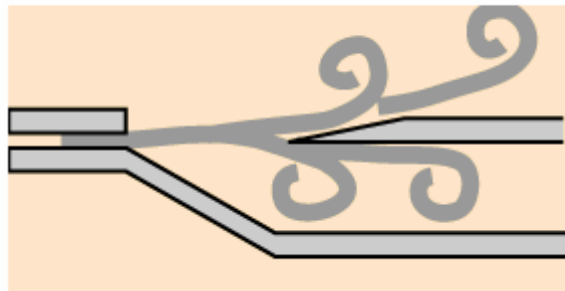
Στο μουσικό όργανο Recorder ή φλογέρα, η παραγωγή του ήχου, δηλαδή η δημιουργία στάσιμου κύματος, γίνεται διαμέσου ενός ταλαντευόμενου αέριου πίδακα. Η μορφή της φλογέρας φαίνεται στο παρακάτω σχήμα.



Εικόνα 1.3.5. Φλογέρα(Recorder).

Βασικά μέρη της φλογέρας είναι οι τονικές οπές, το διάφραγμα, η σφήνα, το στόμιο(χώρος ανάμεσα στο διάφραγμα και την σφήνα).

Αέρας ρέει από το στόμα του οργανοπαίχτη στο διάφραγμα της φλογέρας. Τότε δημιουργείται ένας λεπτός πίδακας αέρα από την άλλη πλευρά του διαφράγματος. Η κίνηση αυτού του αέριου πίδακα είναι ασταθής και περιγράφεται από μια ημιτονοειδή κίνηση με διαρκώς αυξανόμενες διακυμάνσεις(τυρβώδη ροή). Σε αυτή τη διαταραχή του αέριου πίδακα, οφείλεται η ηχητική ταλάντωση μέσα στο σωλήνα της φλογέρας που προκαλεί τον αέρα να ρέει διαδοχικά από την μία ή την άλλη πλευρά της σφήνας(μέσα και έξω από τον σωλήνα της φλογέρας) κάθε μισή περίοδο ταλάντωσης(Βλέπε παρακάτω σχήμα).



Εικόνα 1.3.6. Ταλάντωση μιας αέριας στήλης.

Αν η ταχύτητα του πίδακα είναι τέτοια ώστε ο πίδακας να εισέρχεται στην κάτω πλευρά της σφήνας(προς τα μέσα του σωλήνα) και ταυτόχρονα εισέρχεται ο περιβάλλον αέρας(που φυσάει ο οργανοπαίχτης διαμέσου του διαφράγματος), τότε γίνεται ενίσχυση του ήχου και έχουμε αύξηση του πλάτους των ταλαντώσεων, στην βασική συχνότητα συντονισμού του σωλήνα. Το πλάτος θα γίνει μέγιστο στην κατάσταση ισορροπίας, όταν η ενέργεια που χάνεται λόγω τριβών της αέριας στήλης με τα τοιχώματα του σωλήνα ισοσταθμιστεί με την ενέργεια που παρέχεται από τον πίδακα.

Ο περιβάλλον αέρας που εξέρχεται από το διάφραγμα(που φυσάει ο οργανοπαίχτης διαμέσου του διαφράγματος), έχει μεγάλο φάσμα συχνοτήτων(θόρυβος) και παράγει μια μικρή ταλάντωση στην στήλη αέρα μέσα στο σωλήνα, η οποία ταλάντωση ενισχύεται όταν ισχύει η συνθήκη που περιγράφηκε παραπάνω, δηλαδή όταν ο πίδακας αέρα εισέρχεται στο κάτω μέρος της σφήνας και ταυτόχρονα εισέρχεται ο περιβάλλον αέρας.

Οι συχνότητες συντονισμού του σωλήνα της φλογέρας, δίνονται από τις σχέση:

$$f_n = nc/2L, n=1,2,3\dots$$

, αφού η φλογέρα είναι Ανοιχτός-Ανοιχτός σωλήνας. Αυτές οι συχνότητες συντονισμού είναι αρμονικές αφού: $f_n = nf_1$.

Οι τονικές οπές σε μια φλογέρα είναι οι οπές που ανοίγονται κατά μήκος του σωλήνα της φλογέρας, ώστε όταν καλύπτονται αυτές με τα δάκτυλα, αλλάζουν οι συχνότητες συντονισμού και έτσι παράγονται διαφορετικοί τόνοι. Οι συχνότητες συντονισμού αλλάζουν, διότι ένα άνοιγμα μια οπής, ισοδυναμεί με το να κάνουμε τον σωλήνα μικρότερο. Δηλαδή, μια οπή είναι σαν να μεταβάλλει το μήκος του σωλήνα. Το μήκος στο οποίο θα δημιουργηθεί το στάσιμο κύμα λέγεται

αποτελεσματικό μήκος (efficiency length) Όσο μεγαλύτερη είναι η οπή τόσο το αποτελεσματικό μήκος μειώνεται μέχρι το σημείο, όπου το μέγεθος της οπής γίνει συγκρίσιμο με το πάχος του σωλήνα. Τότε, η οπή αυτή είναι το πέρας του μήκους του σωλήνα. Από την σχέση, που δίνει τις συχνότητες συντονισμού του σωλήνα παρατηρούμε ότι μικρότερο μήκος σημαίνει υψηλότερη συχνότητα, άρα ο ήχος θα γίνει οξύτερος. Έτσι, όταν τα δάκτυλα καλύπτουν όλες τις οπές της φλογέρας, τότε παράγεται η χαμηλότερη συχνότητα συντονισμού, ενώ με ανοιχτές όλες τις οπές παράγεται πολύ υψηλή συχνότητα, που αντιστοιχεί στην έβδομη νότα μιας μουσικής κλίμακας. Αν όλες οι οπές είναι κλειστές και χρησιμοποιηθεί η τεχνική overblowing (υπερφύσημα), τότε η συχνότητα διπλασιάζεται σε σχέση με την συχνότητα που παράγεται χωρίς να κάνουμε overblowing, άρα αυτό σημαίνει ότι μουσικά το τονικό ύψος ανεβαίνει μια οκτάβα.

Υπάρχουν τέσσερις τύποι φλογέρας που μπορούν να χρησιμοποιηθούν και να παράγουν διαφορετικά τονικά ύψη: α) Σοπράνο (Soprano) β) Αλτο (Alto) γ) Τενόρο (Tenor) και δ) Μπάσο (Bass).

Συνήθως οι φλογέρες είναι κουρδισμένες σε Ντο (C) ή Φα (F) νότα. Και έχει έκταση περίπου δυο οκτάβων και ενός τόνου.

1.3.4 Ξυλόφωνο

Το ξυλόφωνο είναι και αυτό πολύ παλιό όργανο και δημιουργήθηκε ανεξάρτητα, στην Αφρική και την Ασία. Το 17^ο αιώνα Αφρικανοί οργανοπαίχτες πήραν μαζί τους το ξυλόφωνο στην Κεντρική Αμερική, όπου εκεί τροποποιήθηκε και έγινε γνωστό ως μαρίμπα (marimba). Η μαρίμπα εξακολουθεί να είναι δημοφιλής στο Μεξικό και στην Κεντρική Αμερική και θεωρείται εθνικό όργανο της Γουατεμάλα. Οι Αφρικανοί που είναι υπεύθυνοι για την διάδοση του οργάνου αυτού, ανέπτυξαν και αποτελεσματικές μεθόδους για το σωστό κούρδισμα του. Δημιουργούσαν μια εγκοπή σε σχήμα τόξου από την κάτω μεριά της ξύλινης μπάρας από την οποία αποτελείται και έτσι η μπάρα κουρδίζονταν σωστά. Αυτή η μέθοδος είναι το κλειδί του σωστού κούρδισματος του ξυλοφώνου, της μαρίμπα και άλλων μελών της οικογένειας ξυλοφώνου.

Το ξυλόφωνο έκανε την εμφάνισή του στην Ευρώπη, πλέον με την Ευρωπαϊκή μορφή του, τον 15^ο αιώνα και αναπτύχθηκε από την Κεντρική και Ανατολική Ευρώπη.

Οι συνθέτες της Δύσης ή δεν είχαν γνωρίσει ακόμα αυτό το όργανο ή άρχισαν να γράφουν μουσική για αυτό στα μέσα του 18^{ου} αιώνα.

Τον 19^ο αιώνα, το ξυλόφωνο τροποποιήθηκε προσθέτοντας του επιπλέον σειρές μπαρών και οι τέσσερις σειρές έγιναν πρότυπο ξυλοφώνου. Σύγχρονοι μουσικοί 'αναβίωσαν' την χρήση του ξυλοφώνου το 1960 δίνοντας μεγάλο ενδιαφέρον στην σύνθεση μουσικής γι' αυτό το όργανο.

Το ξυλόφωνο είναι ένα μουσικό όργανο το οποίο ανήκει στην κατηγορία των ιδιόφωνων μουσικών οργάνων και αποτελείται από κάποιες μπάρες ίδιου υλικού (π.χ. ξύλου), ίδιου πάχους αλλά διαφορετικού μήκους. Τα άκρα είναι ελαφρώς στηριζόμενα και μπορεί να θεωρηθούν σαν ελεύθερα άκρα και να παράγονται οι συχνότητες όπως παρουσιάστηκαν στην ενότητα με τα ιδιόφωνα όργανα. Η παραγωγή του ήχου γίνεται όταν στο κτυπήσουμε με ειδικούς κρουστήρες (μικρές μπαγκέτες) τις μπάρες από τις οποίες αποτελείται. Η μορφή του ξυλοφώνου φαίνεται στην παρακάτω εικόνα.



Εικόνα 1.3.7. Ξυλόφωνο .

Όπως, έχει αναφερθεί, οι μπάρες δεν παράγουν αρμονικές συχνότητες. Μεγάλη επίδραση στην αρμονικότητα του τόνου παίζει η δεύτερη συχνότητα συντονισμού (δεύτερη αρμονική), διότι η τρίτη αρμονική και οι υπόλοιπες συχνότητες αποσβένουν πάρα πολύ γρήγορα (δηλαδή, οι συχνότητες αυτές ηχούν για πολύ μικρό χρονικό διάστημα). Όμως, η δεύτερη αρμονική είναι 2.76 φορές μεγαλύτερη από την πρώτη. Για να διορθωθεί αυτό το πρόβλημα οι μπάρες του ξυλοφώνου κόβονται σε μορφή τόξου από την κάτω πλευρά, και έτσι το κέντρο της μπάρας είναι λεπτότερο απ' ότι στα άκρα. Αυτό έχει ως αποτέλεσμα η δεύτερη συχνότητα συντονισμού να είναι 3 φορές μεγαλύτερη από την πρώτη, και έτσι να είναι αρμονική σε σχέση με την πρώτη και καλυτερεύοντας τον τόνο που παράγεται. Το συχνοτικό εύρος ενός ξυλοφώνου προσεγγιστικά κυμαίνεται από 392Hz έως 2500 Hz.

1.3.5 Φυσαρμόνικα

Όπως και τα άλλα μουσικά όργανα, έτσι και η φυσαρμόνικα έχουν μεγάλη προϊστορία. Η ιστορία της φυσαρμόνικας αρχίζει από το 3000 π.Χ. όταν ο Κινέζος αυτοκράτορας Nyn-Kwya εφεύρε ένα μουσικό όργανο που είχε ελεύθερη γλωττίδα (free reed), το οποίο ονομάζονταν 'Sheng' (θαυμάσια φωνή). Το όργανο αυτό θεωρείται πρόδρομος της σύγχρονης φυσαρμόνικας. Το 'Sheng' παρουσιάστηκε στην Ευρώπη τον 18^ο αιώνα και η ιδέα των ελεύθερων γλωττίδων έφερε την ανάπτυξη και άλλων οργάνων μεταξύ των οποίων το reed organ, το ακορντεόν, η κονσερτίνα, το σαξόφωνο και η φυσαρμόνικα. Η σύγχρονη φυσαρμόνικα εφευρέθηκε το 1821 από έναν γερμανικό ωρολογοποιό, εν ονόματι Christian Buschman που έβαλε δεκαπέντε σωλήνες διαφορετικών τονικών υψών μαζί για να δημιουργήσει ένα περίεργο μικρό μουσικό όργανο. Στην αρχή η φυσαρμόνικες παράγονταν από ωρολογοποιούς και ήταν στο περιθώριο ως μουσικά όργανα, έως ότου το 1857 ο Matthias Hohner αποφάσισε να κατασκευάσει και να κάνει μαζική παραγωγή της φυσαρμόνικας πηγαίνοντας στο Trossingen της Γερμανίας.

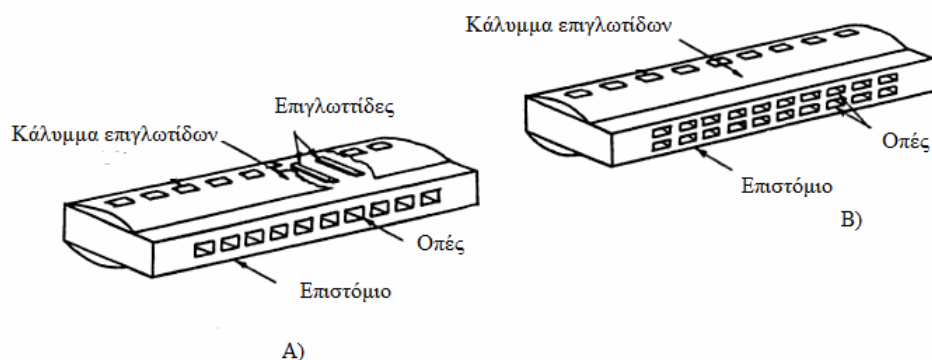
Έτσι, η φυσαρμόνικα διαδόθηκε σε ολόκληρη την Γερμανία και αυτή η μαζική μετανάστευση πολλών Γερμανών πολιτών στα τέλη του 19^{ου} αιώνα, κατέστησε δυνατή την διάδοση της σε ολόκληρο τον κόσμο. Κατά την διάρκεια του

Αμερικανικού εμφυλίου πολέμου, η φουσαρμόνικα είχε φτάσει τις Ηνωμένες Πολιτείες και πολλοί στρατιώτες και από τις δύο πλευρές έπαιζαν φουσαρμόνικα.

Η πρώτη χρήση της ήταν για λαϊκά τραγούδια, επαναστάσεις ύμνους και άλλα παρόμοια και σιγά σιγά χρησιμοποιήθηκε από τους έγχρωμους μετανάστες της Αμερικής και η δυνατότητα της ως όργανο της μπλούζ μουσικής ήρθε στο φως. Στις ημέρες μας η φουσαρμόνικα παίζεται σε πολλά είδη μουσικής με καθιέρωση της στην μπλουζ και την ροκ μουσική.

Η φουσαρμόνικα αποτελείται από μια σειρά ελεύθερες ‘γλωττίδες’ που είναι κουρδισμένες στις νότες μιας μουσικής κλίμακας. Αυτές οι ελεύθερες ‘γλωττίδες’ είναι τοποθετημένες πάνω σε ένα ξύλινο ή πλαστικό ή μεταλλικό κουτί. Ο αέρας που τις διεγείρει παρέχεται από το στόμα του οργανοπαίκτη από τις οπές που βρίσκονται στο επιστόμιο της φουσαρμόνικας. Ο αέρας παρέχεται είτε έχοντας κάποια πίεση (φυσώντας) είτε κάποια υποπίεση (εισπνέοντας), ή κενό (vacuum) ανάλογα αν ο οργανοπαίκτης εκπνέει ή εισπνέει, αντίστοιχα. Κάθε οπή έχει ένα κανάλι (channel), που συνδέεται με δύο ‘γλωττίδες’. Έτσι, κάθε κανάλι της φουσαρμόνικας είναι ικανό να παράγει δυο τόνους διαφορετικών συχνοτήτων, οι οποίοι παράγονται όταν κάποιος εισπνέει αέρα ή όταν εκπνέει αέρα από την οπή ή οπές της φουσαρμόνικας.

Υπάρχουν φουσαρμόνικες τριών κατηγοριών: α) Η απλή φουσαρμόνικα (simple harmonica) β) Η φουσαρμόνικα τύπου concert (concert harmonica) και γ) Η χρωματική φουσαρμόνικα (chromatic harmonica). Στην παρακάτω εικόνα φαίνονται δύο φουσαρμόνικες διαφορετικού τύπου.



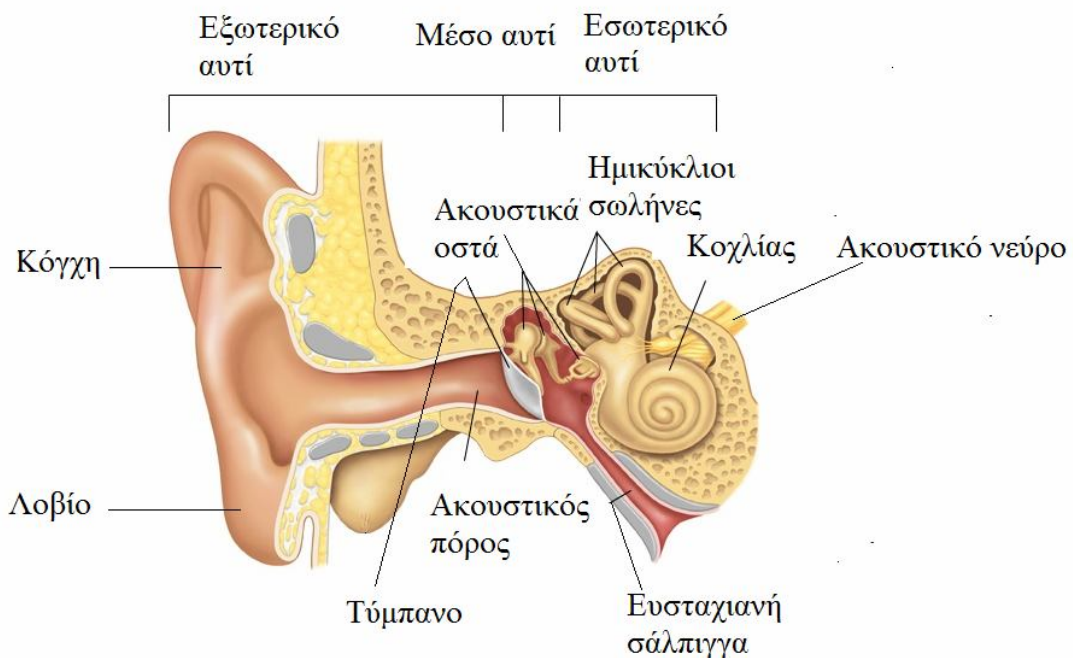
Εικόνα 1.3.8. Α) Απλή φουσαρμόνικα και Β) Φουσαρμόνικα τύπου Concert .

Η απλή φουσαρμόνικα αποτελείται από 10 οπές και μπορεί να παράγει δυο τόνους διαφορετικής συχνότητας σε κάθε οπή ανάλογα αν ο αέρας εισπνέεται ή εκπνέεται από το όργανο. Η φουσαρμόνικα τύπου concert έχει δυο σειρές από οπές. Στην πάνω σειρά οι ‘γλωττίδες’ είναι κουρδισμένες μια οκτάβα υψηλότερα από ότι οι ‘γλωττίδες’ της κάτω σειράς. Οι χρωματικές φουσαρμόνικες, έχουν παρόμοιο σχήμα με τις φουσαρμόνικες τύπου concert, και η διαφορά τους έγκειται στο ότι στις χρωματικές κλίμακες η πάνω σειρά είναι κουρδισμένη ένα ημιτόνιο υψηλότερα απ’ ότι η κάτω σειρά. Στην χρωματική φουσαρμόνικα υπάρχει μια λαβή στα δεξιά, όπου είναι συνδεδεμένη με το εσωτερικό του οργάνου. Αν είναι έξω αυτή η λαβή, τότε οι πάνω οπές είναι ανοιχτές και οι κάτω κλειστές και αν η λαβή είναι προς τα μέσα, τότε οι πάνω οπές είναι κλειστές και οι κάτω ανοιχτές. Γενικά, οι φουσαρμόνικες είναι

κατασκευασμένες σε διάφορα κλειδιά, δηλαδή άλλες έχουν κάποια κλίμακα της C minor, ενώ άλλες στην E major.

1.4 Η φυσιολογία του αυτιού

Όταν παράγεται ένας ήχος από κάποια πηγή π.χ. κιθάρα, φωνή, μηχανή αυτοκινήτου κ.τ.λ. προκαλούνται μεταβολές πίεσης στον αέρα, και αυτές οι μεταβολές διαδίδονται ως ηχητικά κύματα, τα οποία διεγείρουν το αισθητήριο της ακοής και προκαλούν ένα ακουστικό αίσθημα, που λέγεται ήχος. Το αισθητήριο της ακοής που διεγείρεται για να προκαλέσει το αίσθημα του ήχου ονομάζεται αυτί.



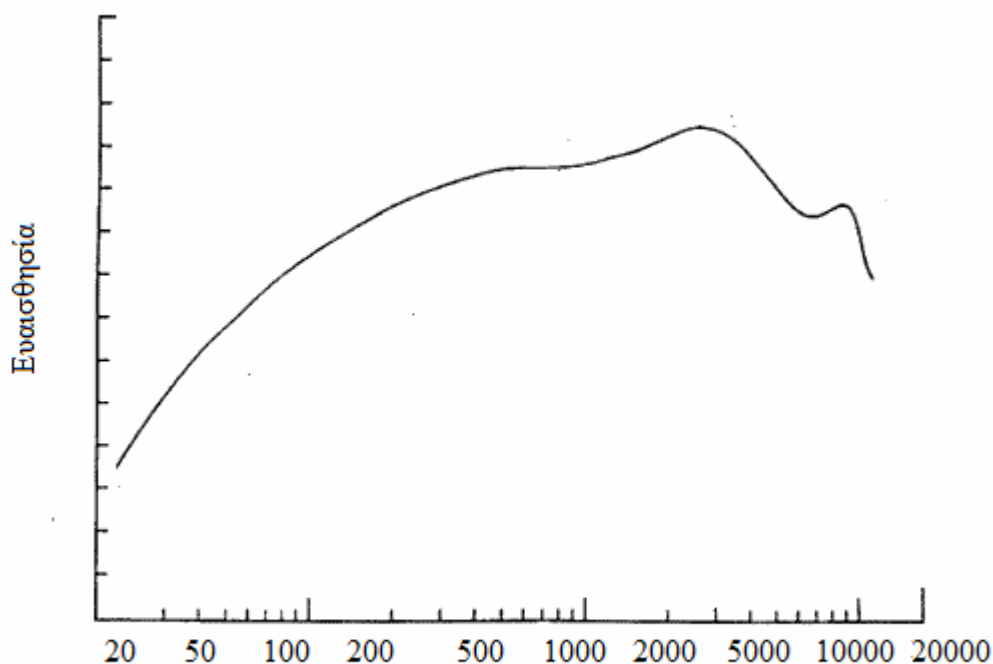
Εικόνα 1.4.1. Το αυτί, το αισθητήριο της ακοής.

Το αυτί, όπως φαίνεται και στην παραπάνω εικόνα χωρίζεται σε τρία μέρη και συγκεκριμένα στο εξωτερικό αυτί, στο μέσο αυτί και το εσωτερικό αυτί. Το εξωτερικό αυτί περιέχει την κόγχη και τον ακουστικό πόρο. Όταν ένα ηχητικό κύμα διαδίδεται μέσω του αέρα στο αυτί, προσπίπτει στην κόγχη, η οποία συγκεντρώνει το ηχητικό κύμα και το μεταφέρει στον ακουστικό πόρο. Ο ακουστικός πόρος λειτουργεί σαν ένα σωλήνα Ανοιχτό-Κλειστό, που έχουμε δει σε προηγούμενη ενότητα και, διεγείρει περισσότερο κάποιες συγκεκριμένες συχνότητες που ονομάζονται συχνότητες συντονισμού και δίνονται από την σχέση:

$$f_n = nc/4L, n=1,2,3\dots$$

, όπου c είναι η ταχύτητα ήχου στον αέρα και L το μήκος του ακουστικού πόρου. Το μήκος του ακουστικού πόρου είναι περίπου ίσο με $L \approx 28\text{mm}$, που η θεμελιώδη

συχνότητα συντονισμού είναι ίση με: $f_1=c/4 L=343/4*0.028=343/0.112\approx 3062\text{Hz}$. Αυτό μας επιβεβαιώνει γιατί και η υψηλότερη ευαισθησία του ανθρώπινου αυτιού είναι στις μεσαίες συχνότητες από 3000-4000Hz, σύμφωνα με την καμπύλη ακουστότητας Fletcher-Munson. Η καμπύλη συντονισμού είναι ευρεία(βλέπε παρακάτω εικόνα) και σε κάποιες συχνότητες έχουμε αύξηση της ηχητικής πίεσης κατά 10dB στο τύμπανο.



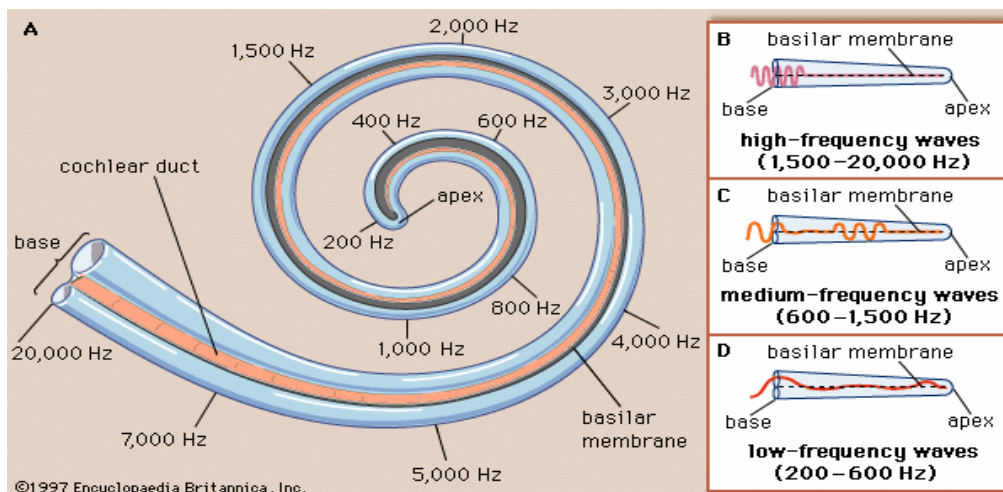
Εικόνα 1.4.2. Καμπύλη συντονισμού του εξωτερικού αυτιού.
Συχνότητα(Hz)

Το εξωτερικό αυτί διαχωρίζεται από το μέσο αυτί, μέσω του τυμπάνου. Σε αυτό το σημείο του αυτιού προσαρμόζεται η μηχανική σύνθετη αντίσταση του τυμπάνου με την ειδική σύνθετη αντίσταση του αέρα. Σε συχνότητες 800Hz και άνω, αυτή η προσαρμογή(ταίριασμα) των εμποδίσεων είναι αρκετά καλή, αλλά είναι μικρή η προσαρμογή των εμποδίσεων κάτω από τα 800Hz. Αν δεν υπήρχε το μέσο αυτί, δεν θα υπήρχε καλή προσαρμογή των εμποδίσεων με αποτέλεσμα την μείωση της στάθμης της ηχητικής πίεσης σε πολλές συχνότητες.

Το τύμπανο ή τυμπανική μεμβράνη, όπως ονομάζεται είναι ο μηχανισμός που θα μετατρέψει τις μεταβολές της ακουστικής πίεσης του αέρα σε μηχανικές δονήσεις, οι οποίες θα διαδοθούν στο εσωτερικό αυτί, και συγκεκριμένα στον κοχλία, μέσω των τριών ακουστικών οστών, της σφύρας(malleus), του άκμονα(incus) και του αναβολέα(stapes). Τα οστά αυτά συνδέονται με μυϊκούς ιστούς. Οι μυϊκοί ιστοί λειτουργούν σαν προστατευτικός μηχανισμός του αυτιού σε ισχυρά ερεθίσματα(ήχοι υψηλής έντασης), περιορίζοντας το πλάτος της δόνησης. Στον κοχλία, που βρίσκεται στο εσωτερικό του αυτιού γίνεται η ανάλυση του ήχου, μετατρέποντας τα ηχητικά κύματα σε ηλεκτροχημικά σήματα, τα οποία μέσω του ακουστικού νεύρου μεταφέρονται στον εγκέφαλο.

Τον κοχλία μπορούμε να τον παρομοιάσουμε σαν ένα Αναλυτή Fourier(Fourier Analyzer), ο οποίος διαχωρίζει την συχνοτική πληροφορία του ηχητικού κύματος για

στον εγκέφαλο. Κατά μήκος του, υπάρχει μια μεμβράνη, η οποία αποκαλείται βασική μεμβράνη ή basilar membrane. Η βασική μεμβράνη αποτελείται από 23000 έως 24000 νευρικές απολήξεις με την μορφή τριχιδίων. Το κάθε τριχίδιο συντονίζεται σε κάποια συγκεκριμένη συχνότητα καθώς όλη η μεμβράνη ταλαντώνεται. Έχουν προταθεί διάφορες θεωρίες για τον τρόπο που επιτυγχάνεται η ανάλυση συχνότητας του ήχου από τον κοχλία. Μια από αυτές, που θεωρείται επικρατέστερη, είναι η γνωστή ως χωρική θεωρία της αντίληψης του τονικού ύψους που είχε προταθεί από τον von Bekesy. Ο von Bekesy παρατήρησε σχηματισμούς πάνω σε πραγματικές βασικές μεμβράνες όταν αυτές διεγείρονταν από ηχητικά κύματα διαφορετικών συχνοτήτων. Έτσι, παρατηρήθηκε ότι οι υψηλές συχνότητες διεγείρουν μέγιστα(ταλαντώνουν με μέγιστο πλάτος) την βασική μεμβράνη στην αρχή της, στην βάση(base). Αντίθετα, οι χαμηλότερες συχνότητες διεγείρουν μέγιστα την βασική μεμβράνη στο τέλος της, στην κορυφή(apex). Οι παρατηρήσεις του von Bekesy συσχετίζουν τις αποστάσεις με το αντιλαμβανόμενο τονικό ύψος του ήχου(βλέπε παρακάτω εικόνα). Αξιοσημείωτο είναι ότι η απόσταση του σημείου από την κορυφή του κοχλία που ταλαντώνεται με το μέγιστο πλάτος είναι ανάλογη με τον λογάριθμο της συχνότητας. Η αναγνώριση της έντασης του ήχου, δεχόμαστε ότι σχετίζεται με το πλάτος ταλάντωσης της μεμβράνης.



Εικόνα 1.4.3. Βασική μεμβράνη του κοχλία, και συσχέτιση του αντιλαμβανόμενου τονικού ύψους με την απόσταση από την κορυφή(Apex) του κοχλία.

Όπως αναφέραμε, σύμφωνα με τη Φυσική ο ήχος είναι ένα ακουστικό κύμα. Δηλαδή, είναι η πίεση και η κίνηση των μορίων του αέρα ή κάποιου άλλου μέσου, που δημιουργείται όταν ένα υλικό δονείται π.χ. όταν ένα διαπασών δονείται(ταλαντώνεται) περιοδικά στον αέρα. Οι μεταβολές της πίεσης(τα ηχητικά κύματα) μεταδίδονται και φθάνουν στο ανθρώπινο ακουστικό νεύρο, το οποίο τις μεταφέρει στον εγκέφαλο, όπου και μεταφράζονται σε ήχους.

Όμως ο ήχος έχει διαφορετική επίδραση σε κάθε άνθρωπο και εξαρτάται από την ιδιοσυγκρασία του αλλά και από άλλους παράγοντες και υπάρχουν κάποια μεγέθη που σχετίζονται με το πώς ένας άνθρωπος αντιλαμβάνεται ένα ήχο, όπου αυτά τα μεγέθη έχουν προκύψει από στατιστική έρευνα σε ένα μεγάλο πλήθος ατόμων, το οποίο αυτό δείγμα είναι αντιπροσωπευτικό για διάφορες κατηγορίες ατόμων.

Γενικά, ένας ήχος έχει ένα σύνολο χαρακτηριστικών που τον περιγράφουν. Αυτά τα χαρακτηριστικά χωρίζονται σε δυο βασικές κατηγορίες, οι οποίες είναι:

α) αντικειμενικά χαρακτηριστικά και β) υποκειμενικά χαρακτηριστικά. Στην πρώτη περίπτωση αυτά τα χαρακτηριστικά είναι ανεξάρτητα από την υποκειμενική αντίληψη ενός ακροατή, ενώ στην δεύτερη περίπτωση εξαρτώνται από την προσωπική αντίληψη του. Στον παρακάτω πίνακα φαίνονται ταξινομημένα τα υποκειμενικά και αντικειμενικά χαρακτηριστικά του ήχου.

Αντικειμενικά χαρακτηριστικά	Υποκειμενικά χαρακτηριστικά
Συχνότητα f(Hz)	Ύψος
Ένταση (W/m^2)	Ακουστότητα
Αρμονικές	Χροιά

Πίνακας 1. Χαρακτηριστικά του ήχου.

Αντικειμενικά χαρακτηριστικά

Ένας ήχος χαρακτηρίζεται από την συχνότητα του και την ένταση του.

- Η συχνότητα είναι ο αριθμός των ταλαντώσεων των σωματιδίων του υλικού μέσου ανά δευτερόλεπτο λόγω της διάδοσης του ηχητικού κύματος, όπου το ηχητικό κύμα δημιουργείται από το υλικό σώμα όταν αυτό δονείται. Η συχνότητα έχει μονάδα μέτρησης τα Hertz.

- Η ηχητική ένταση σε ένα σημείο του ηχητικού πεδίου και προς μια καθορισμένη διεύθυνση είναι το πηλίκο της μέσης ηχητικής ισχύος που διαπερνά κάθετα μια στοιχειώδη επιφάνεια δια του εμβαδού της επιφάνειας. Μονάδα μέτρησης της έντασης του ήχου είναι το $Watt/m^2$.

- Ένας ήχος αν περιέχει μόνο μια συχνότητα ονομάζεται απλός τόνος. Συνήθως στην φύση συναντιούνται ήχοι, οι οποίοι περιέχουν παραπάνω από μια συχνότητα, (π.χ. η φωνή) και μπορούν να αναλυθούν ως ένα πεπερασμένο πλήθος συγκεκριμένων ημιτονοειδών κυμάτων. Μια γραφική παράσταση, που αναπαριστά το πλάτος των απλών ημιτονοειδών κυμάτων συναρτήσει της συχνότητας, λέγεται φάσμα. Το φάσμα δείχνει την ενέργεια που μεταφέρουν οι συνιστώσες συχνότητες από τις οποίες αποτελείται ο ήχος. Εάν το φάσμα του σύνθετου ήχου, περιέχει μια συνιστώσα συχνότητα που είναι ίση με την συχνότητα του σύνθετου περιοδικού ήχου, αυτή η συχνότητα λέγεται θεμελιώδης (fundamental). Αν οι υπόλοιπες συχνότητες από τις οποίες αποτελείται είναι ακέραια πολλαπλάσια της θεμελιώδους συχνότητας λέγονται αρμονικές (harmonics). Συνήθως, οι αρμονικές έχουν διαφορετικά πλάτη σε σχέση με την θεμελιώδη συχνότητα αλλά και μεταξύ τους. Η πιο συνηθισμένη περίπτωση είναι οι σύνθετοι τόνοι να μην είναι περιοδικοί και το φάσμα των συχνοτήτων αποτελείται από ένα μεγάλο πλήθος συχνοτήτων περιλαμβανομένων αρμονικών και μη. Στην πράξη, και οι περιοδικοί σύνθετοι περιέχουν περισσότερες συχνότητες στο φάσμα τους, με την θεμελιώδη και τις αρμονικές να υπερισχύουν έναντι των άλλων συχνοτήτων.

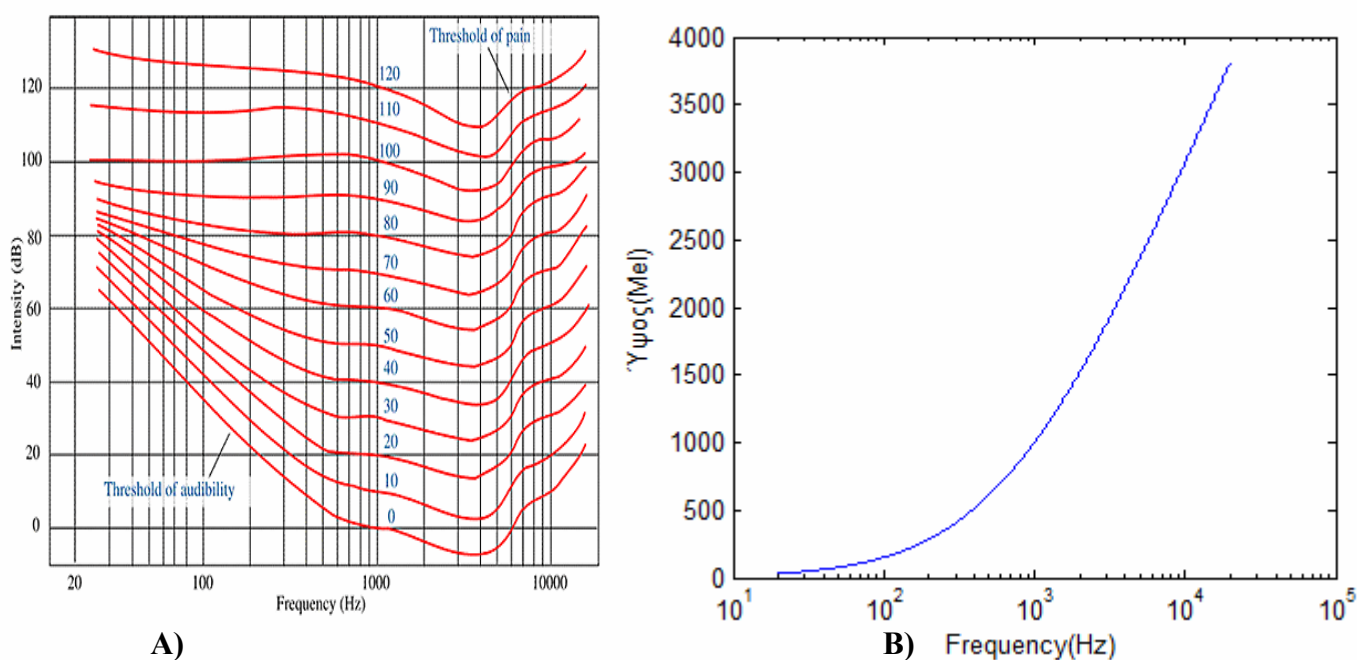
Υποκειμενικά χαρακτηριστικά

Τα υποκειμενικά χαρακτηριστικά ενός ήχου είναι η ακουστικότητα ή ακουστότητα (loudness), το τονικό ύψος (pitch) και η χροιά (timbre).

- Η ακουστότητα είναι η υποκειμενική ένταση του ήχου, που εξαρτάται από την ακουστική ικανότητα του ακροατή. Η ακουστότητα είχε μελετηθεί από δυο ερευνητές, τους Fletcher and Munson, οι οποίοι μέσα από κάποια στατιστικό πείραμα κατάληξαν στις επονομαζόμενες καμπύλες ίσης ακουστότητας ή όπως είναι γνωστές καμπύλες Fletcher and Munson. Οι καμπύλες αυτές μας πληροφορούν ότι ένας τόνος 100Hz έχει στάθμη ηχητικής έντασης 60dB είναι το ίδιο ακουστός με τον τόνο των 1000Hz που έχει στάθμη ηχητικής έντασης 40dB, δηλαδή, η ένταση που αντιλαμβανόμαστε είναι ίδια και για τους δύο τόνους η οποία μετριέται σε Phon. Έτσι, σύμφωνα με τις καμπύλες η στάθμη ακουστότητας των τόνων των 100Hz και 1000Hz είναι ίση με 40Phon. Γενικά, οι ήχοι διακρίνονται σε ασθενείς και ισχυρούς. Ένας ασθενής ήχος έχει περίπου ακουστότητα 1 Phon, ενώ ένας ήχος, ο οποίος είναι ισχυρός και μπορεί να προκαλέσει βλάβη ή πόνο στο αυτί έχει ακουστότητα ίση με 120 Phon.

- Το τονικό ύψος σχετίζεται με την συχνότητα ενός ήχου και είναι η υποκειμενική απόκριση του αυτιού στην συχνότητα. Το τονικό ύψος είναι το γνώρισμα του ήχου με το οποίο μπορούμε να διακρίνουμε τους ήχους σε οξείς (πρίμους, υψηλής συχνότητας) και σε βαρείς (μπάσους, χαμηλής συχνότητας). Για να μετρήσουμε το ύψος σαν ένα υποκειμενικό αίσθημα, χρησιμοποιούμε την μονάδα Mel σε αντίθεση με την μονάδα μέτρησης της συχνότητας του ήχου που είναι τα Hertz (ή κύκλοι ανά δευτερόλεπτο). Γενικά, η μονάδα μέτρησης Mel ορίζεται ως εξής: Ένας απλός τόνος συχνότητας 1000Hz που η στάθμη της έντασης του είναι 60dB έχει ύψος 1000 Mel. Το πόσο καλά, ο άνθρωπος, μπορεί να αναγνωρίσει και να διακρίνει ήχους διαφορετικής συχνότητας εξαρτάται από την συχνότητα και την ένταση των ήχων.

- Η χροιά ενός ήχου είναι ένα χαρακτηριστικό το οποίο μας επιτρέπει να ξεχωρίζουμε ήχους, οι οποίοι μπορεί να έχουν το ίδιο τονικό ύψος και την ίδια ένταση. Αυτό το χαρακτηριστικό ονομάζεται και ηχόχρωμα και έχει σχέση με τους σύνθετους ήχους, οι οποίοι έχουν την ίδια θεμελιώδη συχνότητα αλλά διαφορετικές αρμονικές και τα σχετικά πλάτη των αρμονικών είναι διαφορετικά.



Εικόνα 1.4.4. Α) Καμπύλη ίσης ακουστότητας (ή καμπύλη Fletcher-Munson) και Β) Σχέση μεταξύ τονικού ύψους (σε Mel) με την συχνότητα (λογαριθμική κλίμακα).

Το σύστημα ακοής ενός ανθρώπου είναι αρκετά εξελιγμένο, όπου μπορεί να αναγνωρίσει πολλά γεγονότα και αντικείμενα μέσω του ήχου που αυτά παράγουν. Επίσης, έχει την ικανότητα να αναγνωρίσει μεγάλο πλήθος ήχων σε ένα περιβάλλοντα χώρο που τα χαρακτηριστικά αυτού του χώρου αλλάζουν. Π.χ. μπορεί ένα άνθρωπος να είναι σε μια καφετέρια και να συνομιλεί με το κινητό του τηλέφωνο με ένα άλλο πρόσωπο. Παρ' όλο που υπάρχει μουσική στην καφετέρια, κάποιος αριθμός ανθρώπων που συνομιλούν μεταξύ τους εκείνη την χρονική στιγμή, ο άνθρωπος που μιλάει στο τηλέφωνο έχει την ικανότητα να αναγνωρίσει το πρόσωπο με το οποίο συνομιλεί, να αναγνωρίσει το μουσικό κομμάτι που παίζεται από τα μεγάφωνα της καφετέριας κ.τ.λ.

Επομένως, σε έναν περιβάλλοντα χώρο που θα βρεθεί ένας άνθρωπος, συνήθως δεν ακούγονται οι πηγές ήχου ξεχωριστά, αλλά παράγονται ήχοι σχεδόν ταυτόχρονα. Ο ακροατής με κάποιο τρόπο πρέπει να οργανώσει αυτή την περίπλοκη ακουστική σκηνή, με ένα τέτοιο τρόπο ώστε να μπορεί να καταλάβει την συνεισφορά της κάθε πηγής στο συνολικό ηχητικό αποτέλεσμα που έχει δημιουργηθεί.

Η ακουστική σκηνή ανάλυσης είναι ένα πεδίο που σχετίζεται με την έρευνα ψυχοφυσικών φαινομένων, προσπαθώντας να εξηγήσει το πώς ένας ακροατής καταλαβαίνει μια συνεχή μίξη ήχων καθώς αυτοί προέρχονται από ένα σύνολο ανεξάρτητων πηγών.

Το εγχείρημα της ανάλυσης της ακουστικής σκηνής είναι αρκετά δύσκολο λόγω της φύσης του ήχου, ο οποίος φθάνει στο αισθητήριο της ακοής, το αυτί. Κάθε ηχητική πηγή δημιουργεί μικρές μεταβολές στη ατμοσφαιρική πίεση, δηλαδή δημιουργεί ηχητικά κύματα, τα οποία διαδίδονται μακριά από την πηγή. Η δυσκολία έγκειται στο ότι τα ηχητικά κύματα που προέρχονται από ανεξάρτητες πηγές, φθάνουν στο αυτί ως άθροισμα ανεξάρτητων ηχητικών κυμάτων με συνέπεια ο

ακροατής να μην μπορεί να έχει 'πρόσβαση' σε κάθε ηχητικό κύμα ξεχωριστά παρά μόνο στην μίξη των ηχητικών κυμάτων.

Στην πραγματικότητα, όπως αναφέρθηκε οι πηγές που μπορεί να βρίσκονται στον χώρο, δρουν σχεδόν ταυτόχρονα και σπάνια συμβαίνει να είναι συγχρονισμένα τα ηχητικά γεγονότα. Έτσι, υπάρχουν κάποιοι περιορισμοί στον περιβάλλοντα χώρο. Πεδίο μελέτης της ανάλυσης της ακουστικής σκηνής είναι να προσδιορίσει αυτούς τους περιορισμούς, την επίδραση τους στην μίξη των ήχων και πιθανές στρατηγικές για την 'εξερεύνηση' τους για την κατανόηση τους. Στο βιβλίο του Bregman (Auditory Scene Analysis, 1990), που έχει θέσει τις βάσεις του πεδίου της ανάλυσης ακουστικής σκηνής, παρουσιάζει ένα σύνολο περιορισμών και στρατηγικών, καθώς και στοιχεία για την χρησιμοποίησή τους από τους ακροατές.

Στην περίπτωση που μια συγκεκριμένη ηχητική πηγή παράγαγε το ίδιο ηχητικό κύμα κάθε φορά η αναγνώριση θα ήταν εύκολη. Σε γενικές γραμμές θα μπορούσαμε να απομνημονεύσουμε κάθε ήχο και όταν αναπαρήγαγε η ηχητική πηγή τον ίδιο ήχο θα τον ταιριάζαμε με το αποθηκευμένο πρότυπο στην μνήμη, αναγνωρίζοντας τον. Στην πραγματικότητα, κάτι τέτοιο είναι σχεδόν αδύνατο, διότι υπάρχει μεγάλη μεταβλητότητα στα ηχητικά κύματα που παράγονται από μια συγκεκριμένη πηγή σε διαφορετικούς χρόνους. Αυτές οι αποκλίσεις (η μεταβλητότητα) είναι μέρος της πολυπλοκότητας του περιβάλλοντος.

Για παράδειγμα, φυσικοί ήχοι(που δεν παράγονται από αντικείμενα που είναι κατασκευασμένα από ανθρώπους), διαφέρουν σε διαφορετικούς χρόνους παραγωγής των, διότι μια φυσική διαδικασία της παραγωγής ήχου δεν είναι ποτέ ίδια δυο φορές.

Η φωνή ανήκει στην κατηγορία των φυσικών ήχων, αφού η παραγωγή της είναι μια φυσική διαδικασία. Παράγοντας μέσω της φωνής το φωνήεν 'α' της ελληνικής αλφαβήτου σε διαφορετικές χρονικές στιγμές, το ηχητικό κύμα που θα δημιουργηθεί θα είναι διαφορετικό κάθε φορά. Αυτό συμβαίνει, διότι κάθε φορά η τάση που εφαρμόζεται στις φωνητικές χορδές αλλάζει, οι φωνητικές χορδές ταλαντώνονται με διαφορετικό τρόπο, μεταβάλλεται η απόκριση της στοματικής κοιλότητας λόγω αλλαγής της γλώσσας, της σιαγόνας κ.τ.λ.

Ωστόσο, ο άνθρωπος λειτουργεί με έναν πιο αφαιρετικό τρόπο ώστε να μπορέσει να αναγνωρίσει ένα ήχο. Στο προηγούμενο, παράδειγμα, καθώς προφέρεται το γράμμα 'α' υπάρχουν κάποιοι παράγοντες που μεταβάλλουν το ηχητικό κύμα σε μεγάλο βαθμό. Όμως υπάρχουν και άλλοι παράγοντες που παραμένουν σταθεροί και δεν μεταβάλλονται κατά την διαδικασία παραγωγής του ήχου. Αυτοί οι παράγοντες είναι κάποια ιδιαίτερα χαρακτηριστικά που έχει ένας ήχος και ονομάζονται ακουστικές σταθερές(acoustic invariants). Έτσι, αυτά τα χαρακτηριστικά ενός ήχου που σχετίζονται με τις ακουστικές σταθερές, είναι χρήσιμο να προσδιοριστούν ώστε να κατανοήσουμε τον μηχανισμό αναγνώρισης ηχητικών γεγονότων από τον άνθρωπο.

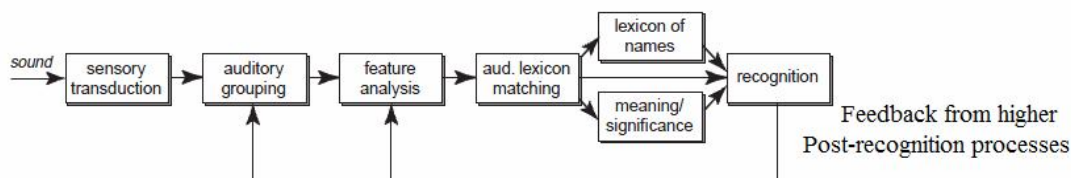
Ο Handel και ο McAdams προτείνουν ότι όποιο πόρισμα βασίζεται στην ανίχνευση των σταθερών παραγόντων του ήχου είναι η πιο πιθανή βάση της αναγνώρισης της ηχητικής πηγής από άνθρωπο.

Ο τρόπος διέγερσης και οι συντονισμοί μιας ηχητικής πηγής επιδρούν ταυτόχρονα στις ιδιότητες ενός ηχητικού κύματος και υπάρχουν πολλά δυναμικά ακουστικά χαρακτηριστικά για την χρησιμοποίησή τους για αναγνώριση. Δεν υπάρχει όμως, κάποιο χαρακτηριστικό ή ομάδα χαρακτηριστικών που μπορούν σε κάθε περίπτωση και υπό όλες τις συνθήκες να είναι καθοριστικά για την αναγνώριση. Τα χαρακτηριστικά αυτά μεταβάλλονται και δεν είναι τελείως ανεξάρτητα μεταξύ τους. Το ένα επηρεάζει το άλλο. Ειδικά στην περίπτωση που υπάρχει πλήθος ήχων που επικαλύπτονται και είναι δύσκολο να είναι γνωστό σε μεγάλο βαθμό ποια

χαρακτηριστικά θα είναι διαθέσιμα και χρήσιμα για την διαδικασία της αναγνώρισης. Η στρατηγική που θα χρησιμοποιήσει ο ακροατής για την αναγνώριση πρέπει να είναι ευέλικτη.

Ο Mc Adams προτείνει ότι η διαδικασία της αναγνώρισης με αφαιρετικό τρόπο, σε ένα μεγάλο μέρος είναι διαδοχική, δηλαδή τα ηχητικά κύματα μετατρέπονται, κατά την αισθητήρια μεταγωγή τους (κατά την διάδοση τους από ‘αισθητήρες’), σε μια αναπαράσταση όπου μια ακουστική ομαδοποίηση (auditory grouping) μπορεί να πραγματοποιηθεί.

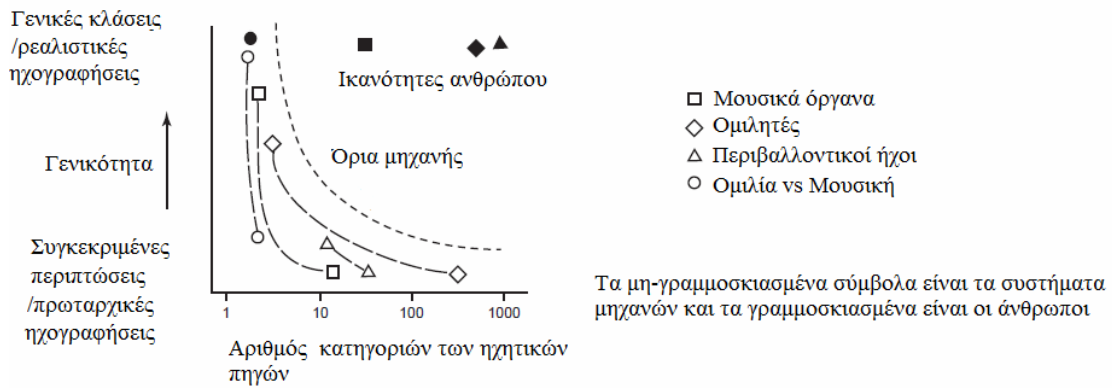
Τα στοιχεία της ομαδοποίησης αναλύονται σε ένα σύνολο χαρακτηριστικών, τα οποία χρησιμοποιούνται ως η βάση της διαδικασίας της αναγνώρισης. Ο Mc Adams πιστεύει ότι η αναγνώριση λαμβάνει χώρα μετά την διαδικασία ομαδοποίησης της ανάλυσης της ακουστικής σκηνής. Παρ’ όλα αυτά αφήνει την πιθανότητα της ανατροφοδότησης από διαδικασίες μετά- αναγνώρισης, υψηλότερου επιπέδου. Αυτή η κυκλική επανάληψη της ανατροφοδότησης είναι απαραίτητη και πρέπει να ληφθεί υπόψη σε φαινόμενα, όπως η αποκατάσταση των φωνημάτων.



Εικόνα 1.4.5. Στάδια ακουστικής επεξεργασίας για αναγνώριση ηχητικής πηγής σύμφωνα με τον Mc Adams.

Στην παραπάνω εικόνα φαίνονται τα στάδια της επεξεργασίας για την αναγνώριση ηχητικής πηγής από τον άνθρωπο, όπου σύμφωνα με τον Mc Adams είναι διαδοχική. Γενικά, χρησιμοποιούμε ό,τι γνωρίζουμε για μια συγκεκριμένη ηχητική πηγή για να καλύψουμε τα κενά στα διαθέσιμα αισθητήρια στοιχεία της ακοής, δηλαδή, τα στοιχεία που δεν ακούσαμε. Σύμφωνα με τα φαινόμενα ακουστικής αποκατάστασης (Warren), συμπληρώνουμε λεπτομερώς τα κενά με την προϋπάρχουσα γνώση, βάσει των προσδοκιών μας και των αισθήσεων μας. Η συγκεκριμένη διαδικασία είναι εξολοκλήρου απρόσιτη στη συνείδηση μας. Δεν γνωρίζουμε ότι το κάνουμε και πιστεύουμε ότι ακούμε περισσότερο και με περισσότερη λεπτομέρεια από ότι στην πραγματικότητα ακούμε. Πράγματι, η κυκλική επανάληψη της ανατροφοδότησης στην αρχιτεκτονική του Mc Adams (τα στάδια ακουστικής επεξεργασίας) είναι ουσιαστική. Ως άνθρωποι είμαστε ικανοί να αναγνωρίσουμε περιπτώσεις από ένα μεγάλο πλήθος γενικών κατηγοριών, σε πραγματικές ακουστικές συνθήκες και κάτω από ένα ευρύ φάσμα πολυπλοκότητας που προκύπτει από μια μίξη σχεδόν-ταυτόχρονων ήχων. Η ανθρώπινη αναγνώριση μειώνεται, όταν οι συνθήκες που λαμβάνει χώρα η αναγνώριση χειροτερεύουν. Η εκμάθηση μας είναι εξαιρετικά ευέλικτη. Μπορούμε να αναγνωρίσουμε την δομή ενός χώρου χωρίς να έχει δοθεί ‘ετικέτα’ (label) σε κάθε αντικείμενο και μαθαίνουμε συνεχώς, προσθέτοντας καινούριες κατηγορίες αντικειμένων καθόλη την διάρκεια της ζωής μας. Εκτός από μια τέτοια μη-επιβλέπουσα μάθηση (unsupervised learning), μπορούμε να μάθουμε καινούριες κατηγορίες από κάποια διδασκαλία. Σε πολλές περιπτώσεις χρειαζόμαστε μόνο μερικά παραδείγματα, μερικές φορές ακόμα και ένα, για να μάθουμε μια καινούργια κατηγορία (Sayre, 1965). Συμπληρώνοντας τα παραπάνω, ο εγκέφαλος μας λειτουργεί σε πραγματικό χρόνο και όχι μόνο σε γενικές γραμμές. Οι άνθρωποι

ακροατές ξεπερνούν τα συστήματα μηχανών σε κάθε κριτήριο, όπως φαίνεται στο παρακάτω σχήμα.



Εικόνα 1.4.6. Ικανότητες ανθρώπου σε σχέση με τα συστήματα μηχανών για την αναγνώριση περιπτώσεων, όπως μουσικά όργανα, ομιλητές, περιβαλλοντικοί ήχοι, ομιλία μαζί με μουσική. Το γράφημα είναι για αναγνώριση είτε κάποια συγκριμένη κατηγορία ή σε γενική κατηγορία.

ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ

ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΣΗΜΑΤΟΣ

2.1 Βασικά της ψηφιακής επεξεργασίας σήματος

Η ψηφιακή επεξεργασία σήματος(Digital Signal Processing) ασχολείται με την ψηφιακή αναπαράσταση των σημάτων, την ανάλυση τους και την επεξεργασία αυτών και των πληροφοριών που μεταφέρουν. Σε αυτό το κεφάλαιο θα γίνει αναφορά σε κάποιες βασικές έννοιες της ψηφιακής επεξεργασίας σήματος.

Τα περισσότερα σήματα τα οποία συναντώνται στο φυσικό κόσμο, αλλά και σε διάφορες εφαρμογές είναι κυρίως αναλογικά σήματα. Π.χ. τα σήματα ήχου, τα βιολογικά σήματα, τα ηλεκτρικά σήματα κ.τ.λ. είναι κυρίως αναλογικά σήματα.

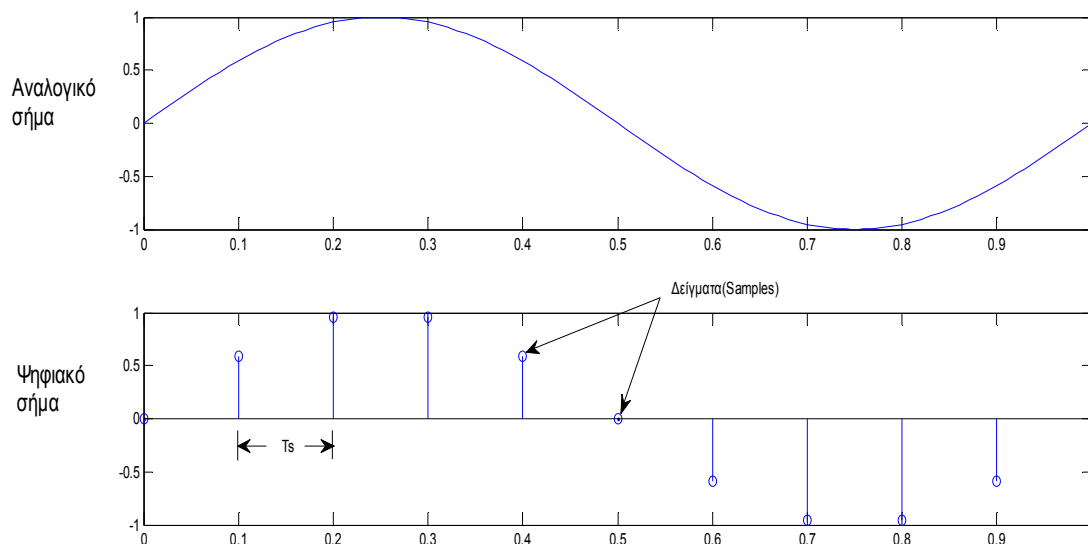
Γενικά, μπορούμε να πούμε ότι:

- Ένα σήμα είναι μια συνάρτηση μιας ή πολλών ανεξάρτητων μεταβλητών που περιέχει κάποια πληροφορία.
- Ένα Αναλογικό σήμα είναι μια συνεχής συνάρτηση, όπου τόσο η εξαρτημένη, όσο και η ανεξάρτητη μεταβλητή παίρνουν συνεχείς τιμές. Σε πολλές περιπτώσεις η ανεξάρτητη μεταβλητή είναι ο χρόνος που εξελίσσεται το σήμα και η εξαρτημένη μεταβλητή το πλάτος ενός σήματος. Οι τιμές του χρόνου παίρνουν οποιαδήποτε τιμή σε ένα διάστημα τιμών (α, β) που είναι το πεδίο ορισμού της συνάρτησης και οι τιμές του πλάτους της παίρνουν οποιαδήποτε τιμή σε ένα διάστημα (α', β'), που είναι το πεδίο τιμών της συνάρτησης.
- Ένα ψηφιακό σήμα είναι μια ακολουθία πραγματικών ή μιγαδικών αριθμών ή είναι το σήμα για το οποίο τόσο η ανεξάρτητη μεταβλητή του χρόνου, όσο και οι τιμές του σήματος(εξαρτημένη μεταβλητή) παίρνουν διακριτές(κβαντισμένες) τιμές.

Σε πολλές περιπτώσεις θέλουμε να μετατρέψουμε τα αναλογικά σήματα σε ψηφιακά, ώστε να εισαχθούν σε ένα υπολογιστικό σύστημα, όπως είναι ο ηλεκτρονικός υπολογιστής(που δέχεται τις πληροφορίες ως διακριτές τιμές) για ανάλυση, επεξεργασία κ.τ.λ.

Έτσι, για να εισαχθεί και να επεξεργασθεί ένα αναλογικό σήμα σε έναν ηλεκτρονικό υπολογιστή θα πρέπει να μετατραπεί αυτό το σήμα σε ψηφιακό, δηλαδή οι συνεχείς τιμές του σήματος να μετατραπούν σε διακριτές(ψηφιακές τιμές). Για να επιτευχθεί κάτι τέτοιο γίνεται μετατροπή από έναν αναλογικό-ψηφιακό μετατροπέα(ADC-Analog to Digital Converter). Συνήθως το σύστημα αυτό σε έναν Η/Υ περιέχεται στην κάρτα γραφικών. Το αναλογικό σήμα μετατρέπεται σε ψηφιακό, κατά την εκτέλεση δύο σταδίων:

- α)της δειγματοληψίας(Sampling) και
- β)της κβαντοποίησης ή της ανάλυσης(Quantization ή Resolution).



Εικόνα 2.1.1. Αναπαράσταση ενός αναλογικού και ενός ψηφιακού σήματος.

Όπως, είπαμε ένα αναλογικό σήμα είναι μια συνεχής μεταβολή, συναρτήσει του χρόνου και μπορεί να πάρει άπειρες τιμές τόσο η μεταβλητή του χρόνου, όσο και η μεταβλητή του πλάτους του σήματος. Έτσι, στο πρώτο στάδιο μετατροπής του αναλογικού σήματος σε ψηφιακό, δηλαδή κατά την διαδικασία της δειγματοληψίας λαμβάνονται δείγματα του σήματος ανά τακτά χρονικά διαστήματα. Το πόσα δείγματα λαμβάνονται στην μονάδα του χρόνου εξαρτάται από την συχνότητα δειγματοληψίας f_s . Όσο πιο μεγάλη είναι η συχνότητα δειγματοληψίας τόσο πιο πιστά θα αναπαρασταθεί το αρχικό σήμα.

Η επιλογή της συχνότητας δειγματοληψίας γίνεται με βάση το θεώρημα των Nyquist και Shannon που αναφέρει ότι:

Πρέπει να υπάρχουν τουλάχιστον δύο δείγματα ανά κύκλο ενός σήματος για να υπάρχει επαρκής πληροφορία.

Έτσι, εάν έχουμε ένα σήμα το οποίο είναι 20000 κύκλοι το δευτερόλεπτο, δηλαδή 20KHz, τότε σύμφωνα με το παραπάνω θεώρημα πρέπει να ληφθούν 2 δείγματα ανά κύκλο, άρα η συχνότητα δειγματοληψίας πρέπει να είναι ίση ή μεγαλύτερη των $2 \cdot 20000 = 40000$ κύκλους το δευτερόλεπτο, δηλαδή 40KHz.

Τα παραπάνω μπορούν να διατυπωθούν διαφορετικά ως εξής:

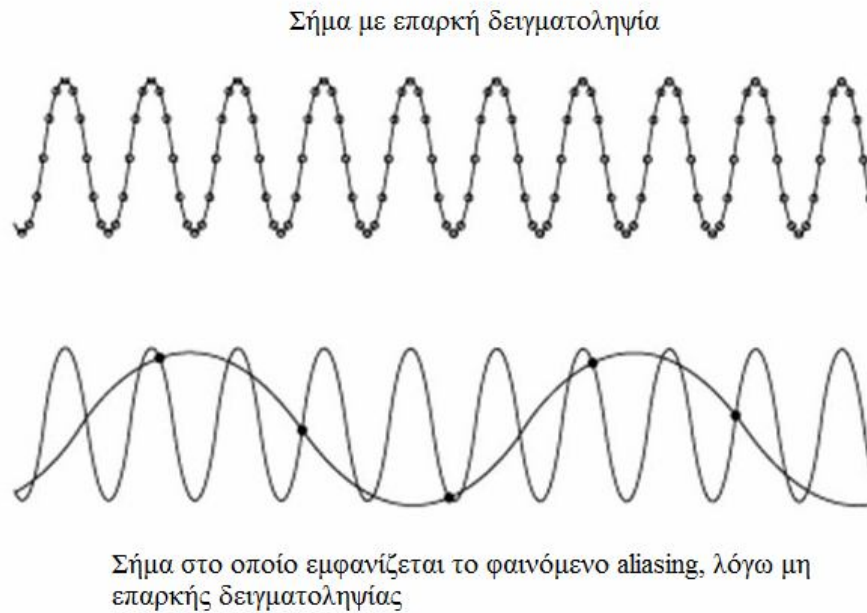
Για να ανακατασκευαστεί ένα σήμα έχοντας επαρκή πληροφορία, πρέπει η συχνότητα δειγματοληψίας να είναι τουλάχιστον διπλάσια της υψηλότερης συχνότητας που περιέχει το σήμα στο οποίο θα γίνει η δειγματοληψία.

Δηλαδή,

$$f_s \geq 2f_{\max} \quad (2.1.1)$$

Η συχνότητα f_{\max} καλείται συχνότητα αποκοπής ή συχνότητα Nyquist.

Σε περίπτωση, που η συχνότητα δειγματοληψίας είναι μικρότερη της $2f_{\max}$, τότε θα παρουσιαστούν κάποια σφάλματα στον ψηφιακό σήμα και πιο συγκεκριμένα το φασματικό περιεχόμενο του σήματος θα χαθεί μετά την f_{\max} κατά την διακριτικοποίηση (δειγματοληψία). Το φαινόμενο αυτό ονομάζεται αναδίπλωση, ή επικάλυψη (aliasing) του φάσματος.



Εικόνα 2.1.2. Σήμα με επαρκή πληροφορία και σήμα στο οποίο εμφανίζεται το φαινόμενο aliasing.

Στο δεύτερο στάδιο, δηλαδή κατά της διαδικασία της κβαντοποίησης, οι τιμές του πλάτους του σήματος πρέπει να πάρουν μεμονωμένες(διακριτές) τιμές. Δηλαδή, έχουν ληφθεί κάποια δείγματα(samples), που έχουν κάποιες τιμές πλάτους(στάθμες). Οι τιμές του πλάτους των δειγμάτων που ‘επιλέχθηκαν’ κατά την διαδικασία της δειγματοληψίας πρέπει να αντιστοιχηθούν στην πλησιέστερη διακριτή στάθμη του συστήματος.

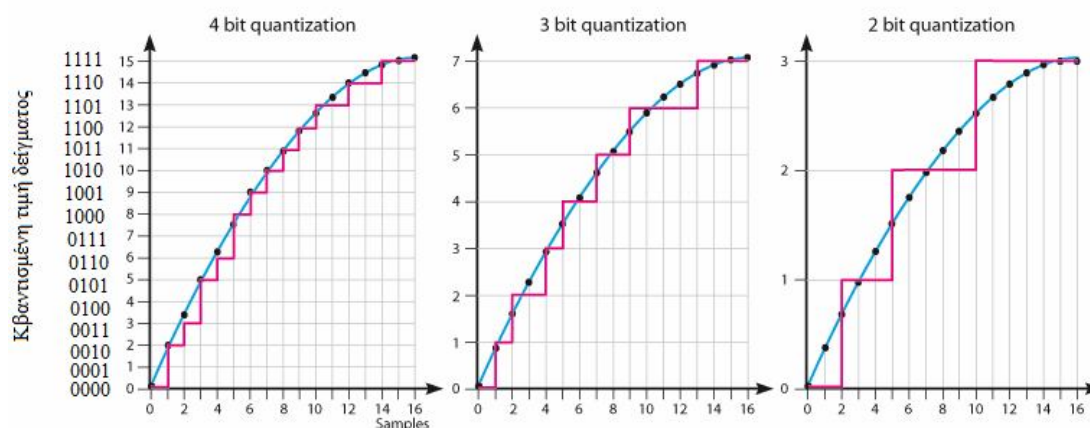
Οι διακριτές στάθμες του συστήματος, στις οποίες θα αντιστοιχηθούν οι τιμές του πλάτους του δείγματος, που έχει επιλεχθεί κατά την δειγματοληψία, σχετίζεται με την ανάλυση του συστήματος ή όπως λέγεται το bit-rate του συστήματος. Αν το σύστημα έχει bit-rate στα 4-bit, τότε το πλήθος των διακριτών σταθμών θα είναι ίσο με 2^4 bit, δηλαδή θα είναι ίσο με 16 στάθμες(επίπεδα), ένα παράδειγμα φαίνεται στην παρακάτω εικόνα. Το bit-rate σε ένα CD, είναι ίσο με 16 bit, άρα υπάρχει ανάλυση $2^{16}=65536$ διακριτών σταθμών (επίπεδων). Η σχέση των τιμών bit-rate με τα επίπεδα που δημιουργούνται στο σύστημα είναι ίση με:

$$n\text{-bit-rate} = 2^n \text{ διακριτές στάθμες(επίπεδα)} \quad (2.1.2)$$

Όμως, λόγω του περιορισμένου αριθμού επιπέδων κατά την διαδικασία της κβαντοποίησης, εμφανίζονται κάποια σφάλματα στο σήμα και τα οποία ονομάζονται σφάλματα κβαντισμού (quantization errors). Η διαφορά μεταξύ δυο συνεχόμενων διακριτών τιμών των στάθμων, ονομάζεται κβάντο(quantum) ή μέγεθος βήματος(step-size). Η μέση τετραγωνική τιμή του σφάλματος είναι ίση με:

$$\bar{q}^2 = d/12 \quad (2.1.3)$$

, όπου d είναι η τιμή του κβάντου ή του μεγέθους του βήματος.



Εικόνα 2.1.3. Κβαντοποίηση μιας συνάρτησης με ανάλυση συστήματος στα 4-bit, 3-bit και 2-bit, αντίστοιχα.

Στην συνέχεια τα δείγματα θα αποθηκευτούν σαν δυαδικά ψηφία(κώδικα λέξης) και αυτά θα χρησιμοποιηθούν για την ανακατασκευή του σήματος.

2.2 Φίλτρα

Θεωρούμε ότι έχουμε ένα γραμμικό και χρονικά αμετάβλητο σύστημα με κρουστική απόκριση ίση με $h(n)$. Έτσι, η έξοδος του συστήματος για οποιαδήποτε είσοδο $x(n)$ είναι ίση με:

$$y(n) = x(n) * h(n) \quad (2.2.1)$$

Δηλαδή, η έξοδος είναι ίση με το άθροισμα της συνέλιξης του σήματος εισόδου με την κρουστική απόκριση του συστήματος.

Η συνέλιξη ισχύει ότι είναι ίση με:

$$x(n) * h(n) \Leftrightarrow X(e^{j\omega}) H(e^{j\omega})$$

, όπου $X(e^{j\omega}) \rightarrow \text{DFT}\{x(n)\}$ και $H(e^{j\omega}) \rightarrow \text{DFT}\{h(n)\}$.

Η $X(e^{j\omega})$ είναι ίση με την συχνοτική απόκριση του σήματος εισόδου και η $H(e^{j\omega})$ είναι συχνοτική απόκριση του συστήματος και καθορίζει την επίδραση του συστήματος στο συχνοτικό περιεχόμενο του σήματος εισόδου. Ο DFT, είναι ο Διακριτός Μετασχηματισμός Fourier, όπου η έννοια του θα εξηγηθεί στην ενότητα 2.4.2.

Η $H(e^{j\omega})$ είναι μια ποσότητα που παίρνει και μιγαδικές τιμές, άρα μπορεί να γραφεί με την μορφή:

$$H(e^{j\omega}) = |H(e^{j\omega})| e^{j\phi(\omega)} \quad (2.2.2)$$

, όπου $|H(e^{j\omega})|$ είναι το πλάτος και $\varphi(\omega)$ είναι η φάση της.

Έτσι, η μελέτη της συχνοτικής απόκρισης αφορά τον υπολογισμό και την γραφική απεικόνιση του πλάτους και της φάσης της. Αν η φάση είναι γραμμική συνάρτηση του ω , τότε το σύστημα καλείται σύστημα γραμμικής φάσης.

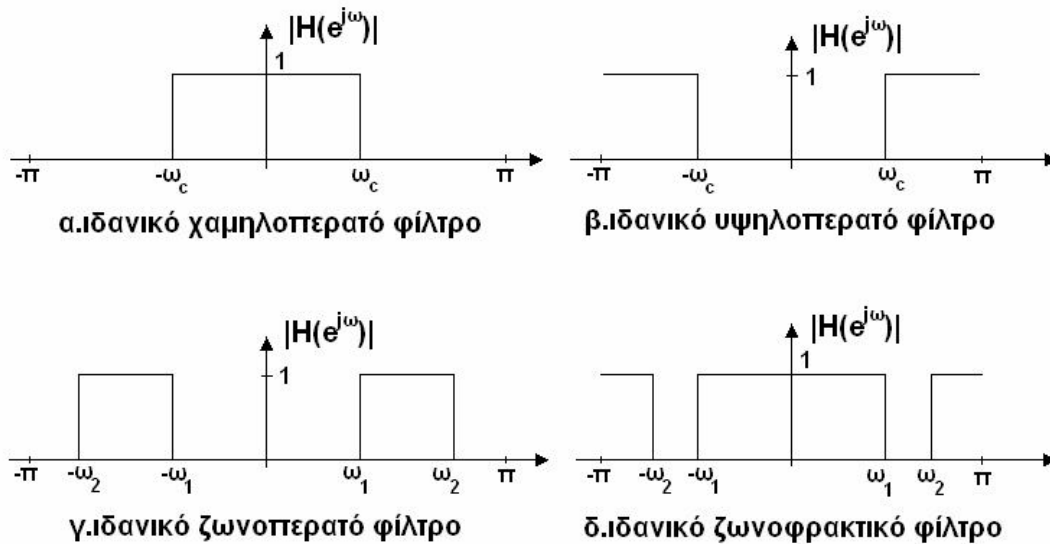
Η καθυστέρηση ομάδας(group delay) $\tau(\omega)$ είναι ίση με:

$$\tau\omega = -\frac{d\varphi(\omega)}{d\omega} \quad (2.2.3)$$

, όπου δείχνει την καθυστέρηση σε πλήθος δειγμάτων που υφίσταται το σήμα εισόδου από το σύστημα συναρτήσει της μεταβλητής ω .

Υπάρχουν διάφοροι τύποι φίλτρων μερικοί από τους οποίους φαίνονται παρακάτω.

Ολοπέρατο φίλτρο, καλείται το φίλτρο στο οποίο ισχύει ότι το πλάτος της συχνοτικής τους απόκρισης είναι ίσο με μια σταθερά, δηλαδή ίσο με $|H(e^{j\omega})|=c$.



Εικόνα 2.2.1. Ιδανικά φίλτρα.

Σημαντικά είναι τα φίλτρα που παρουσιάζουν σταθερό κατά τμήματα πλάτος απόκρισης συχνότητας. Τέτοια φίλτρα φαίνονται στην παραπάνω εικόνα και τα οποία είναι τα χαμηλοπερατά (low-pass) φίλτρα, τα υψηλοπερατά (high-pass) φίλτρα, τα ζωνοπερατά (band-pass) φίλτρα και τα ζωνοφρακτικά (band-stop) φίλτρα. Τα τμήματα στα οποία το πλάτος της απόκρισης συχνότητας είναι ίσο 1 καλούνται ζώνη διέλευσης, ενώ στα τμήματα που το πλάτος που ισούται με 0 καλούνται ζώνη αποκοπής. Οι ζώνες διέλευσης και αποκοπής χαρακτηρίζονται από κάποιες οριακές συχνότητες, που αντιστοιχούν στα όρια των ζωνών και οι οποίες ονομάζονται συχνότητες αποκοπής(cut-off frequencies). Έτσι:

Α) Στην πρώτη περίπτωση ισχύει ότι: $H(e^{j\omega})=1$, όταν $|\omega| \leq \omega_c$ (δηλαδή, όταν $-\omega_c \leq \omega \leq \omega_c$) και $H(e^{j\omega})=0$, όταν $\omega_c < |\omega| \leq \pi$ (δηλαδή, όταν $-\pi \leq \omega < -\omega_c$ και $\omega_c < \omega \leq \pi$), όπου οι τιμές ω_c και $-\omega_c$ είναι οι οριακές συχνότητες της ζώνης διέλευσης και λέγονται συχνότητες αποκοπής.

Β) Στην δεύτερη περίπτωση ισχύει ότι: $H(e^{j\omega})=1$, όταν $|\omega| \geq \omega_c$ (δηλαδή, $\omega \geq \omega_c$ και $\omega \leq -\omega_c$) και $H(e^{j\omega})=0$, όταν $|\omega| \leq \omega_c$ (δηλαδή, όταν $-\omega_c < \omega < \omega_c$), όπου οι τιμές ω_c

και $-\omega_c$ είναι οι οριακές συχνότητες της ζώνης διέλευσης και λέγονται συχνότητες αποκοπής.

Γ) Στην τρίτη περίπτωση ισχύει ότι: $H(e^{j\omega})=1$, όταν $\omega_1 \leq |\omega| \leq \omega_2$ (δηλαδή, όταν $-\omega_2 \leq \omega \leq -\omega_1$ και $\omega_1 \leq \omega \leq \omega_2$) και $H(e^{j\omega})=0$, όταν $-\omega_1 < \omega < \omega_1$ και $-\pi \leq \omega < -\omega_2$ και $\omega_2 < \omega \leq \pi$, όπου οι τιμές και $-\omega_1$, $-\omega_2$, ω_1 και ω_2 είναι οι οριακές συχνότητες της ζώνης διέλευσης και λέγονται συχνότητες αποκοπής.

Δ) Στην τέταρτη περίπτωση ισχύει ότι: $H(e^{j\omega})=1$, όταν $|\omega| \leq \omega_1$ και $|\omega| \geq \omega_2$ (δηλαδή, όταν $-\omega_1 \leq \omega \leq \omega_1$ και $\omega \geq \omega_2$ και $\omega \leq -\omega_2$, αντίστοιχα) και $H(e^{j\omega})=0$, όταν $-\omega_2 \leq \omega \leq -\omega_1$ και $\omega_1 \leq \omega \leq \omega_2$, όπου οι τιμές και $-\omega_1$, $-\omega_2$, ω_1 και ω_2 είναι οι οριακές συχνότητες της ζώνης διέλευσης και λέγονται συχνότητες αποκοπής.

Φίλτρο Butterworth

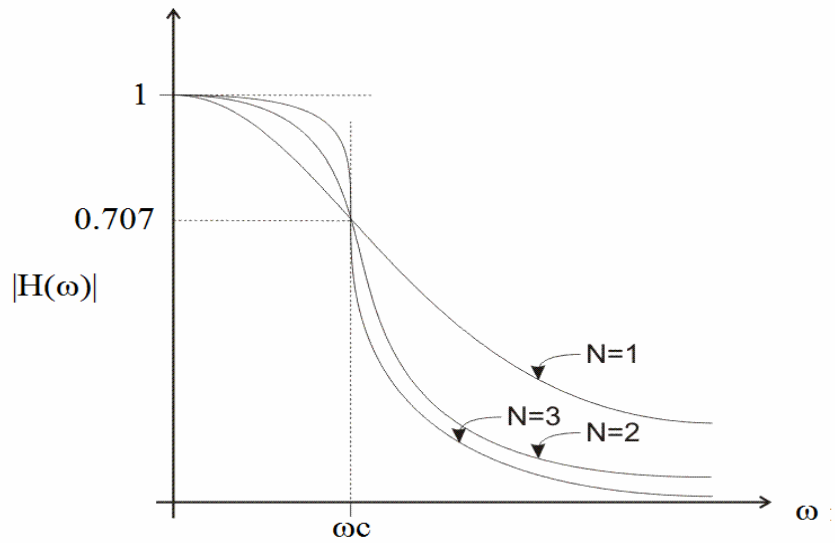
Ένα δημοφιλές φίλτρο είναι το λεγόμενο φίλτρο Butterworth, του οποίου η συνάρτηση της συχνοτικής απόκρισης είναι ίση με:

$$|H(j\omega)|^2 = \frac{1}{1 + \left(\frac{j\omega}{j\omega_c}\right)^{2N}} \quad (2.2.4)$$

ή

$$|H(\omega)|^2 = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}} \quad (2.2.5)$$

, όπου ω είναι η κυκλική συχνότητα, ω_c είναι η συχνότητα αποκοπής και η τιμή του N εκφράζει την τάξη του φίλτρου. Εάν, η συχνότητα ω πάρει την τιμή μηδέν, τότε το μέτρο της συχνοτικής απόκρισης θα γίνει ίσο με $|H(0)|=1$ και έτσι αυτή η συνιστώσα της συχνότητας θα περάσει, χωρίς καμιά επίδραση από το φίλτρο. Αν, η συχνότητα ω πάρει την τιμή που προσεγγίζει το άπειρο (∞), τότε το μέτρο της συχνοτικής απόκρισης θα γίνει ίσο με $|H(\infty)|=0$ και έτσι αυτή η συνιστώσα της συχνότητας θα αποκοπεί. Στην περίπτωση που η συχνότητα ω , είναι ίδια με την συχνότητα αποκοπής ω_c , τότε το μέτρο της συχνοτικής απόκρισης θα γίνει ίσο με $|H(\omega_c)|^2 = 1/2 \Leftrightarrow |H(\omega_c)| = \sqrt{2}/2 = 0.707$, που αντιστοιχεί σε μείωση 3dB από το μέγιστο πλάτος, αν εκφράσουμε την συχνοτική απόκριση σε dB (δηλαδή, $20 \log(H(\omega_c)) = 20 \log(0.707) \Leftrightarrow \text{dB} = -3$). Έτσι, οι τιμές του μέτρου της συχνοτικής απόκρισης του φίλτρου κυμαίνονται από $0 < |H(\omega)|^2 < 1$. Η παρακάτω εικόνα δίνει την γραφική απεικόνιση διαφορετικών τάξεων Butterworth φίλτρων.



Εικόνα 2.2.2. Butterworth φίλτρα πρώτης τάξης, δεύτερης τάξης και τρίτης τάξης, δηλαδή για N=1, N=2 και N=3, αντίστοιχα.

Όταν έχουμε ένα σύστημα το οποίο είναι γραμμικό και χρονικό αμετάβλητο(ΓΧΑ σύστημα), όπου η έξοδος του ανατροφοδοτεί την είσοδο του τότε, η είσοδος και η έξοδος του συστήματος μπορούν να γραφτούν με την βοήθεια εξισώσεων διαφορών με σταθερούς συντελεστές.

Δηλαδή:

$$y(n) = -\sum_{k=1}^p a(k)y(n-k) + \sum_{k=0}^q b(k)x(n-k) \quad (2.2.6)$$

Χρησιμοποιώντας τις ιδιότητες της γραμμικότητας και της μετατόπισης, του Μετασχηματισμό Fourier Διακριτού χρόνου (DTFT), η εξίσωση διαφορών στο πεδίο της συχνότητας εκφράζεται ως:

$$Y(e^{j\omega}) = -\sum_{k=1}^p a(k) \cdot e^{-jk\omega} \cdot Y(e^{j\omega}) + \sum_{k=0}^q b(k) \cdot e^{-jk\omega} \cdot X(e^{j\omega}) \Leftrightarrow$$

$$Y(e^{j\omega}) \left(1 + \sum_{k=1}^p a(k) \cdot e^{-jk\omega}\right) = X(e^{j\omega}) \sum_{k=0}^q b(k) \cdot e^{-jk\omega} \Leftrightarrow$$

$$\frac{Y(e^{j\omega})}{X(e^{j\omega})} = \frac{\sum_{k=0}^q b(k) \cdot e^{-jk\omega}}{1 + \sum_{k=1}^p a(k) \cdot e^{-jk\omega}} \quad (2.2.7)$$

, όπου $\frac{Y(e^{j\omega})}{X(e^{j\omega})}$ ισχύει από το θεώρημα της συνέλιξης ότι είναι ίσο με H(e^{jω}) που είναι η συχνοτική απόκριση τους συστήματος.

Η παραπάνω σχέση μας δίνει την γενική μορφή της εξίσωσης που περιγράφει την συχνοτική απόκριση ενός ΓΧΑ συστήματος. Αν η ποσότητα στον παρανομαστή είναι ίση με μηδέν ή το $p=0$, τότε η παραπάνω σχέση περιγράφει την συχνοτική απόκριση ενός φίλτρου Πεπερασμένης Κρουστικής Απόκρισης(Finite Impulse Response,FIR). Αν $p \neq 0$, τότε η παραπάνω σχέση περιγράφει την συχνοτική απόκριση ενός φίλτρου Άπειρης Κρουστικής Απόκρισης(Infinite Impulse Response,IIR).

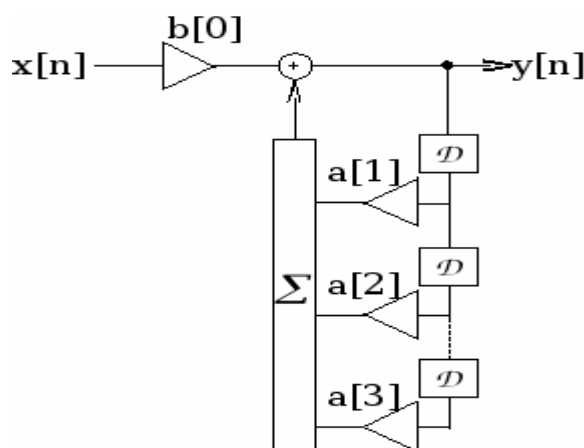
Στην εικόνα που ακολουθεί φαίνεται ένα σύστημα που εξίσωση διαφοράς που συνδέει την είσοδο με την έξοδο του συστήματος είναι ίση με:

$y(n)=b(0)x(n)+a(1)y(n-1)+a(2)y(n-2)+a(3)y(n-3)$ και η συχνοτική απόκριση είναι ίση με:

$$Y(e^{j\omega})=b(0)X(e^{j\omega})+a(1)Y(e^{j\omega})e^{-j\omega}+a(2)Y(e^{j\omega})e^{-j2\omega}+a(3)Y(e^{j\omega})e^{-j3\omega} \Leftrightarrow$$

$$Y(e^{j\omega})[1-a(1)e^{-j\omega}-a(2)e^{-j2\omega}-a(3)e^{-j3\omega}]=b(0)X(e^{j\omega}) \Leftrightarrow$$

$$H(e^{j\omega})=Y(e^{j\omega})/X(e^{j\omega})=\frac{b(0)}{1-a(1)e^{-j\omega}-a(2)e^{-j2\omega}-a(3)e^{-j3\omega}}$$



Εικόνα 2.2.3. Φίλτρο Άπειρης Κρουστικής Απόκρισης(IIR).

2.3 Παράθυρα και Πλαίσια

Στην επεξεργασία σημάτων, συνήθως το σήμα χωρίζεται σε μικρότερα τμήματα μικρού χρονικού εύρους, τα οποία ονομάζονται πλαίσια ή χρονικά παράθυρα(frames ή time windows). Σε κάθε ένα από αυτά τα πλαίσια μικρού χρονικού εύρους, εφαρμόζεται διαδοχικά μια συνάρτηση παραθύρου(window function) και η διαδικασία αυτή ονομάζεται παραθυροποίηση ή windowing. Έτσι, η επεξεργασία ενός σήματος γίνεται μέσω της επεξεργασίας κάθε ξεχωριστού πλαισίου ή χρονικού παραθύρου, που έχουμε χωρίσει το σήμα.

Για παράδειγμα, θεωρούμε ότι έχουμε ένα σήμα $x(n)$, μήκους(πλήθος δειγμάτων) $L1$. Από αυτό το σήμα παίρνουμε μικρότερα τμήματα(πλαίσια), τα οποία έχουν μήκος $L2$, όπου $L2 < L1$. Στην συνέχεια ολισθαίνουμε το πλαίσιο, κατά ένα πλήθος δειγμάτων μικρότερο του συνολικού μήκους του πλαισίου. Ύστερα, επαναλαμβάνουμε την διαδικασία της ολίσθησης μέχρι να φτάσουμε στο τέλος του σήματος. Στην συνέχεια πολλαπλασιάζουμε κάθε πλαίσιο με μια συνάρτηση παραθύρου. Αποθηκεύουμε το πλήθος των πλαισίων(στα οποία έχει εφαρμοστεί η συνάρτηση παραθύρου), που έχουμε χωρίσει το σήμα, και, έχοντας επιλέξει ένα κατάλληλο παράθυρο και ένα κατάλληλο

πλήθος δειγμάτων ολίσθησης, καμιά περιοχή του σήματος δεν θα ‘χαθεί’ στην αποθήκευση. Η επιλογή του πλήθους δειγμάτων ολίσθησης είναι περίπου 1/2 ή 2/3 του μήκους του πλαισίου(δηλαδή, 1/2 ή 2/3 του πλήθους των δειγμάτων που περιέχονται σε ένα πλαίσιο).

Παρακάτω θα διατυπωθεί η ορολογία που χρησιμοποιείται συνήθως για κάποιες έννοιες.

Για να περιγράψουμε το μήκος του πλαισίου χρησιμοποιούμε τον όρο Frame size, που σχετίζεται με το πλήθος των δειγμάτων που περιέχει ένα πλαίσιο. Με τον όρο Frame overlap, δηλώνεται το πλήθος των δειγμάτων που επικαλύπτουν τα διαδοχικά πλαίσια.

Ο όρος Frame step ή Hop size, είναι η διαφορά του πλήθους των δειγμάτων που επικαλύπτουν διαδοχικά πλαίσια από το πλήθος των δειγμάτων του πλαισίου, δηλαδή, Frame step=Frame size-Frame overlap, όπου μας δηλώνει το πλήθος δειγμάτων ολίσθησης του πλαισίου και ο όρος Frame rate, μας δηλώνει τον αριθμό των πλαισίων ανά δευτερόλεπτο, που ισούται με την συχνότητα δειγματοληψίας διαιρεμένη από την ποσότητα Frame step, δηλαδή,

Frame rate=Sampling Frequency/Frame step, και μονάδα μέτρησης της είναι frames per second(frames/sec).

Για παράδειγμα, θεωρούμε ότι έχουμε χρησιμοποιήσει συχνότητα δειγματοληψίας fs=8000samples/sec, χρονική διάρκεια πλαισίου t=25ms και επικάλυψη(overlap) διαδοχικών πλαισίων, σε χρονική διάρκεια t_{overlap}=15ms. Έτσι,

$$\text{Frame size} = fs \cdot t = 200 \text{ δείγματα(samples)}$$

$$\text{Frame overlap} = fs \cdot t_{\text{overlap}} = 120 \text{ δείγματα(samples)}$$

$$\text{Frame step} = \text{Frame size} - \text{Frame overlap} = 80 \text{ δείγματα(samples)}$$

$$\text{Frame rate} = fs / \text{Frame step} = 100 \text{ παράθυρα/δευτερόλεπτο(frames/sec)}$$

Η διαδικασία της παραθυροποίησης είναι σαν να έχουμε ένα σήμα x_v(n) του ν-οστού πλαισίου, το οποίο πολλαπλασιάζεται με μια συνάρτηση παραθύρου w(n), δηλαδή,

$$x_v'(n) = x_v(n)w(n) \quad (2.3.1)$$

Η συνάρτηση παραθύρου w(n) είναι ακολουθία πεπερασμένου μήκους N και η οποία μηδενίζεται έξω από το διάστημα [0,N] και είναι συμμετρική ως προς N/2, δηλαδή ισχύει ότι:

$$w(n) = w(N-n) \quad (2.3.2)$$

Κάποιες από τις συνήθεις συνάρτησεις παραθύρου, που υπάρχουν είναι:

- Το ορθογώνιο παράθυρο

$$w(n) = \begin{cases} 1, & n \in [0, N] \\ 0, & n \notin [0, N] \end{cases} \quad (2.3.3)$$

Το ορθογώνιο παράθυρο λόγω της απότομης αλλαγής που προκαλεί στο σήμα δεν χρησιμοποιείται. Το παράθυρο, που εφαρμόζεται στο σήμα, μηδενίζει απότομα την αρχή και το τέλος του σήματος με συνέπεια, να υπάρχουν σοβαρά προβλήματα στο πεδίο της συχνότητας.

- Το παράθυρο Hanning

$$w(n) = \begin{cases} 0.5 - 0.5 \cos\left(\frac{2\pi n}{N}\right), & n \in [0, N] \\ 0, & n \notin [0, N] \end{cases} \quad (2.3.4)$$

- Το παράθυρο Hamming

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N}\right), & n \in [0, N] \\ 0, & n \notin [0, N] \end{cases} \quad (2.3.5)$$

Στην επεξεργασία των μουσικών σημάτων, αλλά και της φωνής χρησιμοποιείται συνήθως το παράθυρο Hamming, το οποίο δίνει αρκετά ικανοποιητικά αποτελέσματα.

- Το παράθυρο Blackman

$$w(n) = \begin{cases} 0.42 - 0.5 \cos\left(\frac{2\pi n}{N}\right) + 0.08 \cos\left(\frac{4\pi n}{N}\right), & n \in [0, N] \\ 0, & n \notin [0, N] \end{cases} \quad (2.3.6)$$

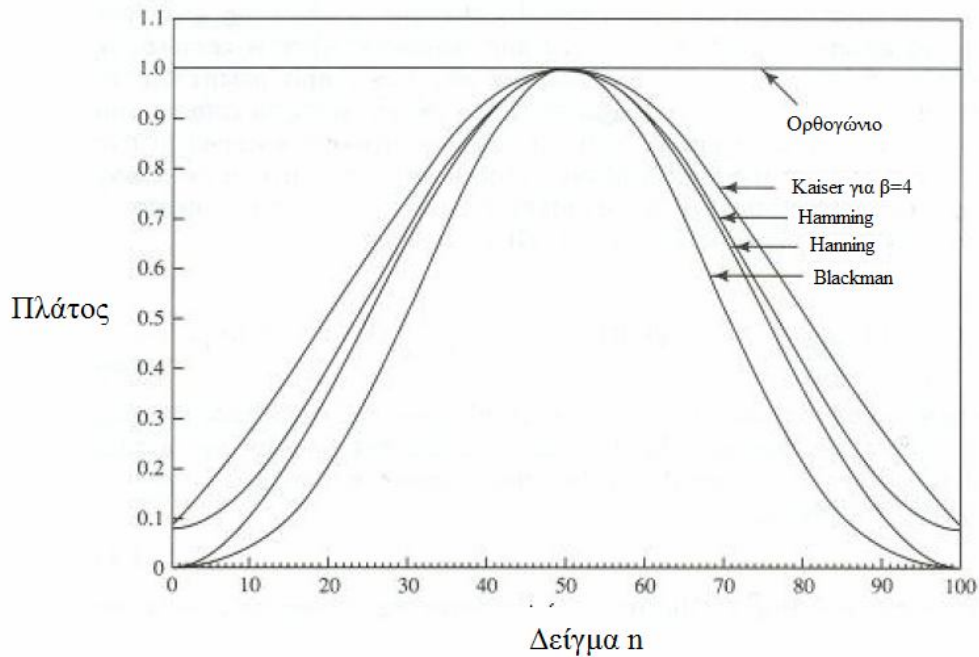
- Το παράθυρο Kaiser

$$w(n) = \begin{cases} \frac{I_0 \left[\beta \left[1 - \left(\frac{n - \alpha}{\alpha} \right)^2 \right]^{1/2} \right]}{I_0 \beta}, & n \in [0, N] \\ 0, & n \notin [0, N] \end{cases} \quad (2.3.7)$$

, όπου $\alpha = N/2$ και $I_0(\cdot)$ είναι μια μηδενικού βαθμού συνάρτηση Bessel πρώτου είδους που παράγεται από το ανάπτυγμα της δυναμοσειράς που φαίνεται παρακάτω:

$$I_0 = 1 + \sum_{k=1}^{\infty} \left[\left(\frac{(x/2)^k}{k!} \right)^2 \right] \quad (2.3.8)$$

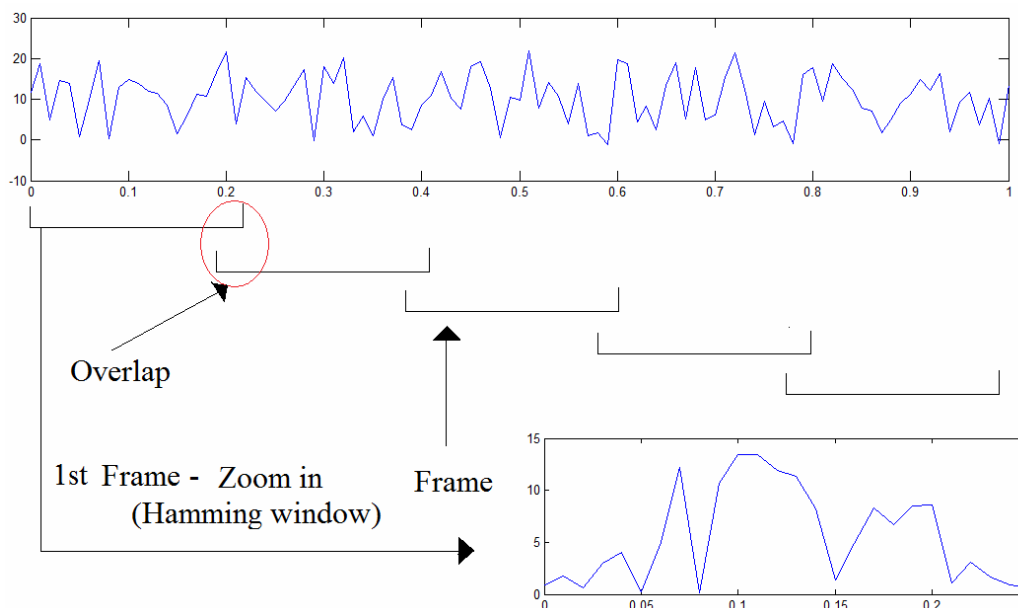
Στην εικόνα 2.3.1, φαίνονται οι διάφοροι τύποι παραθύρων που περιγράφηκαν παραπάνω.



Εικόνα 2.3.1. Διάφορες συναρτήσεις παραθύρων με μήκος N=101.

Εφαρμόζοντας μια από τις παραπάνω συναρτήσεις παραθύρου, μπορούμε να επεξεργαστούμε τόσο το φασματικό περιεχόμενο του σήματος όσο και τη χρονική του μεταβολή. Αν επεξεργαστούμε την χρονική μεταβολή τους σήματος, μπορούμε να εξάγουμε κάποια χαρακτηριστικά, τα οποία ονομάζονται χρονικά χαρακτηριστικά (temporal features).

Ένα παράδειγμα της εφαρμογής του παραθύρου Hamming σε ένα τυχαίο σήμα φαίνεται σχηματικά στην εικόνα 2.3.2, που παρατίθεται παρακάτω.



Εικόνα 2.3.2. Διαίρεση του σήματος σε μικρότερα τμήματα(παράθυρα), με χρήση σε κάθε ένα από αυτά του παραθύρου Hamming και Frame rate=5frames/sec.

2.4 Μετασχηματισμός Fourier και Διακριτός Μετασχηματισμός Fourier

Ο μετασχηματισμός Fourier οφείλεται στον γάλλο μαθηματικό και φυσικό Jean Baptiste Joseph Fourier (1768-1830). Ο Fourier ενδιαφέρονταν για την μεταφορά της θερμότητας σε στερεά σώματα και έτσι παρουσίασε μια εργασία του στο Ινστιτούτο της Γαλλίας για την χρήση των ημιτόνων στην αντιπροσώπευση των κατανομών της θερμοκρασίας. Σε αυτή την εργασία του ανέφερε ότι οποιοδήποτε συνεχές περιοδικό σήμα μπορεί να αντιπροσωπευτεί ως το άθροισμα κατάλληλων επιλεγμένων ημιτονοειδών κυμάτων. Ο μετασχηματισμός και οι σειρές Fourier παρέχουν την δυνατότητα του μετασχηματισμού μιας συνάρτησης $x(t)$ που ανήκει στο πεδίο του χρόνου σε μια άλλη συνάρτηση $X(f)$, η οποία παριστάνεται στο πεδίο της συχνότητας και ο αντίστροφος μετασχηματισμός Fourier παρέχει την δυνατότητα του μετασχηματισμού της συνάρτησης $X(f)$ από το πεδίο της συχνότητας στην αρχική συνάρτηση $x(t)$, που παριστάνεται στο πεδίο του χρόνου.

Με αυτή την διαδικασία είναι εφικτό να υπολογιστεί το φασματικό περιεχόμενο ενός σήματος $x(t)$.

Ένα σήμα μπορεί να είναι είτε διακριτό είτε συνεχές. Επίσης, μπορεί να είναι είτε περιοδικό είτε απεριοδικό. Έτσι, τα σήματα μπορούν να χωριστούν σε τέσσερις βασικές κατηγορίες:

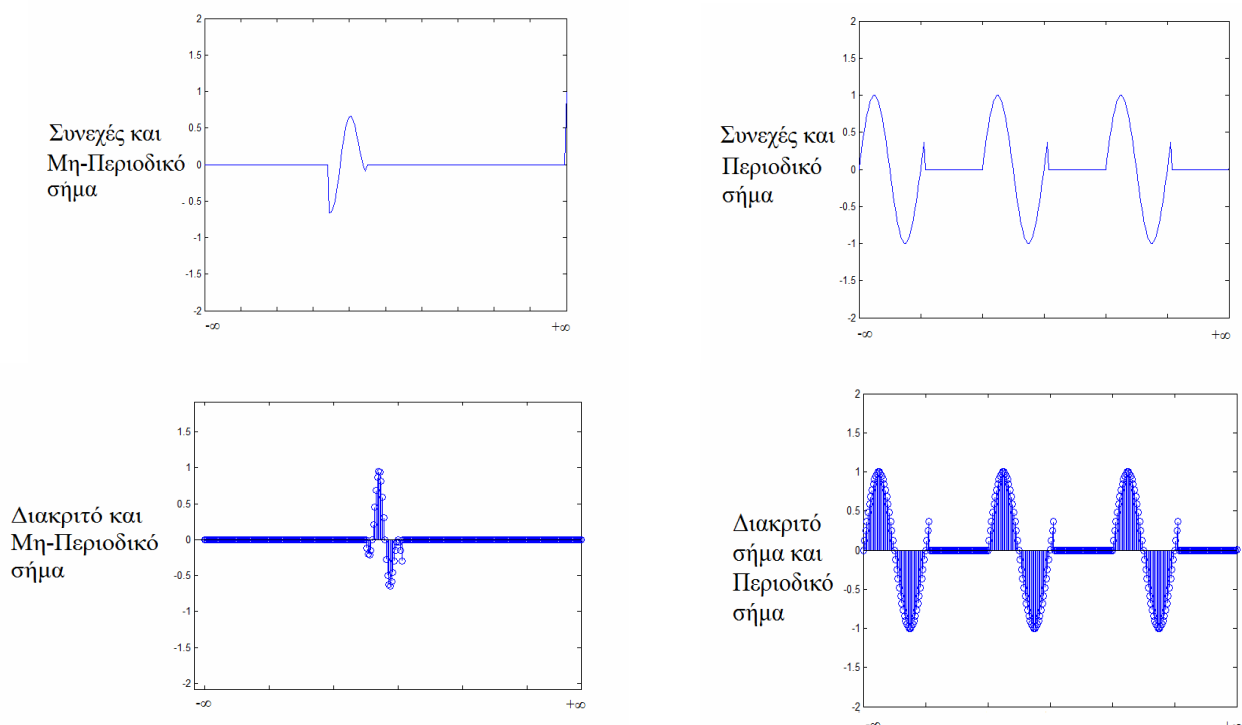
➤ **Συνεχή και μη-περιοδικά:** Σε αυτή την κατηγορία ανήκουν τα σήματα τα οποία εκτείνονται από $-\infty$ έως το $+\infty$, χωρίς όμως να επαναλαμβάνονται σε τακτά χρονικά διαστήματα σαν ένα περιοδικό σήμα. Σε αυτή την περίπτωση ο μετασχηματισμός καλείται Μετασχηματισμός Fourier(Fourier Transform).

➤ **Συνεχή και περιοδικά:** Σε αυτή την κατηγορία ανήκουν τα σήματα τα οποία εκτείνονται από το $-\infty$ έως το $+\infty$, και επαναλαμβάνονται κατά τακτά χρονικά διαστήματα, όπως ένα ημίτονο ή ένα συνημίτονο. Ο μετασχηματισμός για αυτή την κατηγορία καλείται Σειρές Fourier(Fourier Series).

➤ **Διακριτά και μη-περιοδικά:** Τα σήματα σε αυτή την περίπτωση, ορίζονται για διακριτές τιμές(δείγματα) που εκτείνονται από το $-\infty$ έως το $+\infty$, αλλά δεν επαναλαμβάνονται σε τακτά χρονικά διαστήματα. Εδώ, ο μετασχηματισμός καλείται Μετασχηματισμός Fourier Διακριτού Χρόνου(Discrete Time Fourier Transform, (DTFT)).

➤ **Διακριτά και περιοδικά:** Τα σήματα σε αυτή την περίπτωση, ορίζονται για διακριτές τιμές, τα οποία έχουν έναν επαναλαμβανόμενο ‘χαρακτήρα’ σε τακτά χρονικά διαστήματα και εκτείνονται από το $-\infty$ έως το $+\infty$. Αυτός ο μετασχηματισμός ονομάζεται Διακριτός Μετασχηματισμός Fourier(Discrete Fourier Transform, DFT).

Στην εικόνα 2.4.1, απεικονίζονται σχηματικά οι παραπάνω κατηγορίες σημάτων.



Εικόνα 2.4.1.Οι τέσσερις βασικές κατηγορίες σημάτων: Συνεχή-Απεριοδικά, Συνεχή-Περιοδικά, Διακριτά-Απεριοδικά και Διακριτά-Περιοδικά.

Όπως, είδαμε παραπάνω τα σήματα ορίζονται εκτεινόμενα από το $-\infty$ έως το $+\infty$. Δεν μπορείς όμως να χρησιμοποιήσεις μια ομάδα άπειρου μήκους σημάτων και με αυτά να συνθέσεις ένα σήμα που έχει πεπερασμένο μήκος. Λύση, γι' αυτό είναι το πεπερασμένο πλήθος δειγμάτων που χρησιμοποιείται για την περιγραφή ενός σήματος να μοιάζει σαν

ένα σήμα με άπειρο μήκος (άπειρο πλήθος δειγμάτων). Ο τρόπος για να γίνει αυτό είναι να θεωρηθεί, ότι αριστερά και δεξιά από το σήμα που σε ενδιαφέρει, υπάρχει άπειρο πλήθος δειγμάτων και ότι αυτά τα δείγματα, που βρίσκονται αριστερά και δεξιά του σήματος αναφοράς (βοηθητικά δείγματα), είναι μηδενικά. Σε αυτή την περίπτωση το σήμα θα μοιάζει περισσότερο με διακριτό-απεριοδικό και θα χρησιμοποιηθεί ο Μετασχηματισμός Fourier Διακριτού Χρόνου (DTFT).

Αντί να θεωρήσουμε ότι τα βοηθητικά δείγματα είναι μηδενικά, θεωρούμε ότι οι τιμές τους αντιστοιχούν στο σήμα αναφοράς και τότε το σήμα αναφοράς, φαίνεται να επαναλαμβάνεται και να μοιάζει με διακριτό-περιοδικό. Σε αυτή την περίπτωση εφαρμόζεται ο Διακριτός Μετασχηματισμός Fourier. Επειδή, ένας Η/Υ, αναγνωρίζει μόνο διακριτές τιμές, και με πεπερασμένο πλήθος αυτών των τιμών, ο μετασχηματισμός που χρησιμοποιείται στους Η/Υ είναι ο Διακριτός Μετασχηματισμός Fourier (DFT).

Σε αυτή την ενότητα θα ασχοληθούμε με τον Μετασχηματισμό Fourier, τον Διακριτό Μετασχηματισμό Fourier και κάποιες βασικές τους ιδιότητες.

2.4.1 Μετασχηματισμός Fourier

Θεωρούμε ένα μη-περιοδικό σήμα, σαν ένα σήμα που έχει προκύψει από ένα περιοδικό με περίοδο που εκτείνεται από το $-\infty$ έως το $+\infty$. Τότε, για ένα σήμα που είναι μια συνάρτηση του χρόνου με περίοδο από το $-\infty$ έως το $+\infty$, μπορούμε να διατυπώσουμε την εξής σχέση:

$$F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t} dt \quad (2.4.1)$$

και θεωρούμε ότι η συνάρτηση $F(\omega)$ υπάρχει για κάθε τιμή της κυκλικής συχνότητας ω , και αυτή την συνάρτηση την ονομάζουμε Μετασχηματισμό Fourier ή ολοκλήρωμα Fourier. Με τον Μετασχηματισμό Fourier, γίνεται αναπαράσταση του σήματος από το πεδίο του χρόνου στο πεδίο της συχνότητας.

Γενικά, ο Μετασχηματισμός Fourier, είναι μια μιγαδική συνάρτηση και μπορούμε να την εκφράσουμε ως ένα άθροισμα των πραγματικών και των φανταστικών μερών της, ή με ως εκθετική μορφή. Δηλαδή,

$$F(\omega) = \text{Real}\{F(\omega)\} + j\text{Im}\{F(\omega)\} \quad (2.4.2)$$

ή

$$F(\omega) = |F(\omega)| \cdot e^{-j\varphi(\omega)} \quad (2.4.3)$$

,αντίστοιχα.

Ο αντίστροφος Μετασχηματισμός Fourier (Inverse Fourier Transform, IFT), δίνεται από την σχέση:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega)e^{j\omega t} d\omega \quad (2.4.4)$$

, όπου με τον αντίστροφο Μετασχηματισμό Fourier, γίνεται αναπαράσταση του σήματος από το πεδίο της συχνότητας στο πεδίο του χρόνου.

Για να αναπαραστήσουμε τον Μετασχηματισμό Fourier και τον αντίστροφο του, χρησιμοποιούμε τους παρακάτω συμβολισμούς.

$F\{f(t)\}=F(\omega)$ για τον Μετασχηματισμό Fourier και

$F^{-1}\{F(\omega)\}=f(t)$ για τον Αντίστροφο Μετασχηματισμό Fourier.

Στον παρακάτω πίνακα φαίνονται κάποιες βασικές ιδιότητες του Μετασχηματισμού Fourier.

<i>Ιδιότητα</i>	<i>f(t)</i>	<i>F(ω)</i>
Γραμμικότητα	$\alpha_1 f_1(t) + \alpha_2 f_2(t) + \dots + \alpha_n f_n(t)$	$\alpha_1 F_1(\omega) + \alpha_2 F_2(\omega) + \dots + \alpha_n F_n(\omega)$
Συμμετρία	$F(t)$	$2\pi f(-\omega)$
Χρονική μετατόπιση	$f(t - t_0)$	$F(\omega) e^{-j\omega t_0}$
Χρονική Κλιμάκωση	$f(\alpha t)$	$\frac{1}{ \alpha } F\left(\frac{\omega}{\alpha}\right)$
Συχνотική μετατόπιση	$f(t) e^{j\omega_0 t}$	$F(\omega - \omega_0)$
Παραγωγή ως προς τον χρόνο	$\frac{d^n f(t)}{dt^n}$	$(j\omega)^n F(\omega)$
Παραγωγή ως προς την συχνότητα	$(jt)^n f(t)$	$\frac{d^n F(\omega)}{d\omega^n}$
Χρονική συνέλιξη	$f_1(t) * f_2(t)$	$F_1(\omega) F_2(\omega)$
Συχνотική συνέλιξη	$f_1(t) f_2(t)$	$F_1(\omega) * F_2(\omega)$
Συζυγείς συναρτήσεις	$f^*(t)$	$F^*(-\omega)$
Περιοχή κάτω από την f(t)	$F(0) = \int_{-\infty}^{+\infty} f(t) dt$	
Περιοχή κάτω από την F(ω)	$f(0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) d\omega$	

Πίνακας 2. Βασικές ιδιότητες του Μετασχηματισμού Fourier.

2.4.2 Διακριτός Μετασχηματισμός Fourier

Ο διακριτός μετασχηματισμός Fourier, δίνεται από την σχέση:

$$X[m] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi \frac{mn}{N}} \quad (2.4.5)$$

, όπου $\omega = \left(\frac{2\pi}{NT}\right)m$ για $m = 0, 1, 2, \dots, N-1$ και N είναι ο αριθμός των δειγμάτων, τα οποία ισαπέχουν σε ένα διάστημα 0 έως 2π ενός μοναδιαίου κύκλου. Η παραπάνω σχέση

αναφέρεται ως N-δειγμάτων Διακριτός Μετασχηματισμός Fourier της $X[m]$ ή για συντομία N –δειγμάτων DFT.

Ο Αντίστροφος Διακριτός Μετασχηματισμός Fourier, δίνεται από την σχέση:

$$x[n] = \frac{1}{N} \sum_{m=0}^{N-1} X[m] \cdot e^{j2\pi \frac{mn}{N}}, \text{ όπου } n=0,1,2,\dots,N-1 \quad (2.4.6)$$

Όπως, και στην περίπτωση του Μετασχηματισμού Fourier, έτσι και στον Διακριτό Μετασχηματισμό Fourier, η συνάρτηση $X[m]$ είναι μιγαδική και έτσι μπορούμε να την γράψουμε με την μορφή:

$$X[m] = \text{Real}\{X[m]\} + j\text{Im}\{X[m]\}$$

, όπου $m=0,1,2,\dots,N-1$.

Για να αναπαραστήσουμε τον Διακριτό Μετασχηματισμό Fourier και τον αντίστροφο του, χρησιμοποιούμε τους παρακάτω συμβολισμούς.

$\text{DFT}\{x[n]\} = X[m]$ για τον Διακριτό Μετασχηματισμό Fourier και

$\text{DFT}^{-1}\{X[m]\} = x[n]$ για τον Αντίστροφο Διακριτό Μετασχηματισμό Fourier. Ο όρος $e^{-j2\pi n/N}$, είναι στην ουσία ένα στρεφόμενο διάνυσμα, και στρέφεται από γωνία 0 έως 2π , δηλαδή $0 \leq \theta \leq 2\pi$ και διαιρείται σε $360/N$ ίσα τμήματα. Γι' αυτό το λόγο, συχνά αυτός ο όρος αντιπροσωπεύεται με W_N , δηλαδή,

$$W_N = e^{-j\frac{2\pi}{N}} \quad (2.4.7)$$

και

$$W_N^{-1} = e^{j\frac{2\pi}{N}} \quad (2.4.8)$$

Έτσι, έχουμε ότι:

$X[m] = \sum_{n=0}^{N-1} x[n] W_N^{mn}$ για τον Διακριτό Μετασχηματισμό Fourier και

$x[n] = \frac{1}{N} \sum_{m=0}^{N-1} X[m] W_N^{-mn}$ για τον Αντίστροφο Διακριτό Μετασχηματισμό Fourier.

Στον πίνακα, φαίνονται κάποιες βασικές ιδιότητες του Διακριτού Μετασχηματισμού Fourier.

<u>Ιδιότητα</u>	<u>$f(t)$</u>	<u>$F(\omega)$</u>
Γραμμικότητα	$\alpha_1 x_1[t] + \alpha_2 x_2[t] + \dots + \alpha_v x_v[t]$	$\alpha_1 X_1[m] + \alpha_2 X_2[m] + \dots + \alpha_v X_v[m]$
Χρονική μετατόπιση	$x[n-k]$	$X[m] W_N^{km}$
Συχνотική μετατόπιση	$x[n] W_N^{-km}$	$X[m-k]$
Χρονική συνέλιξη	$x_1[n] * x_2[n]$	$X_1[m] X_2[m]$
Συχνотική συνέλιξη	$x_1[n] x_2[n]$	$1/N (X_1[m] * X_2[m])$

Πίνακας 3. Βασικές ιδιότητες του Διακριτού Μετασχηματισμού Fourier.

Χαρακτηριστικό είναι, ότι το σήμα στο πεδίο του χρόνου, όπως έχουμε αναφέρει στην προηγούμενη ενότητα, είναι διακριτό και περιοδικό και εφαρμόζοντας τον Διακριτό Μετασχηματισμό Fourier, το μετασχηματισμένο σήμα που παριστάνεται στο πεδίο των συχνοτήτων είναι επίσης διακριτό και περιοδικό.

Επίσης, στο Μετασχηματισμό Fourier, το σήμα στο πεδίο του χρόνου είναι συνεχές και απεριοδικό, εφαρμόζοντας τον Μετασχηματισμό Fourier, το μετασχηματισμένο σήμα που παριστάνεται στο πεδίο των συχνοτήτων είναι επίσης συνεχές και απεριοδικό.

Αν θεωρήσουμε ότι η ακολουθία $x[n]$ είναι το σήμα εισόδου και ο Μετασχηματισμός Fourier Διακριτού Χρόνου είναι $X(e^{j\omega})$, τότε το Θεώρημα Parseval, διατυπώνεται ως εξής:

$$\sum_{n=-\infty}^{\infty} |x(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(e^{j\omega})|^2 d\omega \quad (2.4.9)$$

Το θεώρημα Parseval αναφέρεται και ως θεώρημα διατήρησης της ενέργειας, διότι καθώς εφαρμόζεται ο Διακριτός Μετασχηματισμός Fourier σε μια ακολουθία $x[n]$, τότε διατηρείται η συνολική ενέργεια κατά την μετάβαση από το πεδίο του χρόνου στο πεδίο της συχνότητας.

2.5 Εξαγωγή Χαρακτηριστικών

Για την αναγνώριση μιας πηγής, όπως είναι ένας ομιλητής, ένα μουσικό όργανο κ.τ.λ. βασικό ρόλο παίζει η χρησιμοποίηση κατάλληλων χαρακτηριστικών για την επιτυχή αναγνώριση της. Πολλά από τα ακουστικά χαρακτηριστικά που θέλουμε να εξάγουμε από ένα ακουστικό σήμα, έχουν σχέση με την ανθρώπινη αντίληψη και έτσι δεν μπορούν να αντιπροσωπευθούν ακριβώς με μαθηματικές σχέσεις και ποσότητες.

Όμως, έχουν γίνει προσπάθειες ώστε να μπορούν αυτά τα χαρακτηριστικά, να εκφραστούν με μαθηματικές σχέσεις, ώστε να μπορούν να χρησιμοποιηθούν από έναν ηλεκτρονικό υπολογιστή για τους απαραίτητους υπολογισμούς και περαιτέρω ανάλυση τους. Αυτές οι σχέσεις που χρησιμοποιούνται, προσπαθούν να μιμηθούν προσεγγιστικά την ανθρώπινη αντίληψη. Έτσι, μπορεί να χρησιμοποιηθούν χαρακτηριστικά που είναι τα φυσικά ή αντικειμενικά χαρακτηριστικά, τα χαρακτηριστικά που βασίζονται στην ανθρώπινη αντίληψη και τα χαρακτηριστικά που είναι συνδυασμοί αυτών. Και στις τρεις περιπτώσεις, ο στόχος είναι, να εξαχθούν χαρακτηριστικά από κατάλληλα δείγματα(πρότυπα) και με βάση αυτά να δημιουργηθεί ένα διάλυμα χαρακτηριστικών. Έχοντας εξετάσει τα κατάλληλα πρότυπα ως προς ένα ή περισσότερα χαρακτηριστικά, δημιουργείται ένα σύνολο που ονομάζεται σύνολο χαρακτηριστικών εκπαίδευσης ή σύνολο εκπαίδευσης (training set). Αυτά τα χαρακτηριστικά εισάγονται σε ένα σύστημα για την εκπαίδευση του ταξινομητή και αυτά θα συνδράμουν μαζί με την επιλογή του ταξινομητή στην απόδοση του συστήματος, στην ταξινόμηση και κατ' επέκταση στην αναγνώριση των προτύπων, όταν αυτά θα εισάγονται στο στάδιο της ταξινόμησης.

Η βασική προσέγγιση στο πρόβλημα της εξαγωγής χαρακτηριστικών ενός ακουστικού σήματος, μπορεί να αναφερθεί περιληπτικά στα ακόλουθα βήματα:

- *Βήμα 1^ο*: Πρέπει να χρησιμοποιήσουμε την τεχνική της εφαρμογής πλαισίων στο σήμα που μας ενδιαφέρει, τεμαχίζοντας το σε ένα σύνολο από χρονικά παράθυρα (πλαίσια). Η χρονική διάρκεια του κάθε πλαισίου πρέπει να είναι περίπου 20-30ms (Αν $f_s=8000$, 160-300δείγματα(samples)). Συνήθως η ποσότητα είναι ίση με ένα πλήθος δειγμάτων που είναι δύναμη του 2(μπορεί να διαιρεθεί ακριβώς με τον αριθμό 2), όπως 256, 512, 1024 κ.τ.λ. δείγματα(samples). Αν η χρονική διάρκεια του πλαισίου, είναι πάρα πολύ μεγάλη, τότε δεν θα μπορούν να υπολογιστούν τα χρονικά-μεταβαλλόμενα χαρακτηριστικά του σήματος. Από την άλλη πλευρά, αν η χρονική διάρκεια είναι πολύ μικρή, τότε δεν θα μπορούν να εξαχθούν ποικίλα ακουστικά χαρακτηριστικά. Γενικά, ένα πλαίσιο πρέπει να περιέχει πάνω από μία θεμελιώδη περίοδο του σήματος(ανάλογα με την θεμελιώδη συχνότητα του σήματος).

- *Βήμα 2^ο*: Στην συνέχεια πρέπει να ορίσουμε την ποσότητα Frame overlap. Όπως, έχουμε αναφέρει και σε προηγούμενη ενότητα, η ποσότητα αυτή πρέπει να είναι περίπου το 1/2 ή 2/3 του μήκους του πλαισίου(Frame size). Όσο, πιο μεγάλη είναι η επικάλυψη, τόσο περισσότερο είναι το υπολογιστικό κόστος.

- *Βήμα 3^ο*: Γενικά, θεωρούμε ότι όταν έχουμε ένα σήμα και το έχουμε τεμαχίσει σε επιμέρους τμήματα(πλαίσια), τότε σε κάθε ένα πλαίσιο το σήμα είναι σταθερό, και με βάση αυτή την υπόθεση μπορούμε να εξάγουμε ακουστικά χαρακτηριστικά, όπως zero crossing, ένταση, ύψος, MFCC, LPC, συντελεστές delta και συντελεστές delta-delta κ.τ.λ.

- *Βήμα 4^ο*: Χρησιμοποιώντας τα χαρακτηριστικά zero crossing και ένταση, μπορούμε να εφαρμόσουμε μια ανίχνευση του σημείου τέλους(endpoint detection) του σήματος, παραλείποντας τα πλαίσια που περιέχουν σήματα που είναι σε σιγή και λαμβάνοντας υπόψιν μόνο τα πλαίσια που περιέχουν σήμα με τιμή έντασης και τιμή zero crossing πάνω από μια τιμή κατωφλίου, για περαιτέρω ανάλυση.

Επιπλέον χαρακτηριστικά και τα χαρακτηριστικά, που αναφέρονται παραπάνω θα αναλυθούν στην συνέχεια.

Κεντρική συχνότητα φάσματος

Η κεντρική συχνότητα φάσματος ή Spectral Centroid, είναι ένα πολύ απλό χαρακτηριστικό. Έχει δειχθεί, ότι αποτελεί το σημείο ισορροπίας του φάσματος και ότι σχετίζεται με το πόσο ένα σήμα είναι πλούσιο σε αρμονικές υψηλών συχνοτήτων και κατ' επέκταση με την λαμπρότητα(brightness) του σήματος. Έτσι, μπορούμε να πούμε ότι είναι ένα μέτρο λαμπρότητας του σήματος, και εκφράζεται από την σχέση:

$$SC = \frac{\sum_{n=1}^N f_n X[n]}{\sum_{n=1}^N X[n]} \quad (2.5.1)$$

, όπου f_n δηλώνει την κεντρική συχνότητα του n φίλτρου της τράπεζας φίλτρων(filterbank), N είναι το πλήθος των φίλτρων της τράπεζας φίλτρων(filterbank), και $X[n]$ είναι το πλάτος (ή η ενέργεια) του σήματος του n φίλτρου. Η μονάδα μέτρησης της κεντρικής συχνότητας φάσματος είναι κύκλοι ανά δευτερόλεπτο, δηλαδή μετριέται σε Hertz(Hz).

Spectral Flux

Η διακύμανση φάσματος ή Spectral flux ή Spectral Variation, όπως λέγεται, είναι μια ποσότητα, η οποία μας δείχνει πόσο αργά ή γρήγορα μεταβάλλεται η ενέργεια του φάσματος ενός σήματος. Πιο συγκεκριμένα ορίζεται ως η διακύμανση των τιμών του φάσματος δυο διαδοχικών παραθύρων(πλαισίων). Η διακύμανση φάσματος δίνεται από την σχέση:

$$SF = \sum_{n=1}^N (X_f[n] - X_{f-1}[n])^2 \quad (2.5.2)$$

, όπου $X[n]$ είναι το κανονικοποιημένο(normalized) πλάτος του n δείγματος (sample) του Διακριτού Μετασχηματισμού Fourier, N είναι το πλήθος των δειγμάτων και f , $f-1$ το τρέχον και το προηγούμενο παράθυρο(frame), αντίστοιχα.

Zero crossing

Για συντομία αυτό το χαρακτηριστικό το συμβολίζουμε με ZCR. Το ZCR χαρακτηριστικό είναι ίσο με το πλήθος των επαναλήψεων, που η κυματομορφή του σήματος τέμνει τον οριζόντιο άξονα(x'x) σε κάθε δοσμένο πλαίσιο.

Από το χαρακτηριστικό ZCR, προκύπτουν κάποια πολύ χρήσιμα συμπεράσματα, όπως:

Η τιμή του ZCR χαρακτηριστικού του θορύβου περιβάλλοντος είναι μεγαλύτερη όταν εισάγουμε ένα σήμα, το οποίο έχει κάποια αισθητή θεμελιώδη περίοδο. Επίσης, αν το σήμα είναι πολύ απλό μπορούμε να χρησιμοποιήσουμε τις τιμές του ZCR, για να βρούμε τη θεμελιώδη περίοδο. Π.χ. αν είναι γνωστή η τιμή του ZCR ανά δευτερόλεπτο και γνωστή η τιμή του ZCR ανά κύκλο(περίοδο), τότε η θεμελιώδης συχνότητα του

σήματος είναι $f = \frac{\text{zeros / sec}}{\text{zeros / cycle}}$.

Αν όμως το σήμα είναι αρκετά σύνθετο τότε η εκτίμηση της θεμελιώδους συχνότητας του είναι αναξιόπιστη.

Η σχέση που μας δίνει την τιμή του χαρακτηριστικού ZCR, είναι η παρακάτω:

$$ZCR = \frac{1}{2} \sum_{n=1}^N | \text{sign}(x[n]) - \text{sign}(x[n-1]) | \quad (2.5.3)$$

, όπου $x[n]$ είναι το πλάτος του σήματος, n είναι ο δείκτης των δειγμάτων(samples) στο πεδίο του χρόνου, N είναι το πλήθος των δειγμάτων και sign δηλώνει την συνάρτηση signum, που έχει οριστεί σε προηγούμενο κεφάλαιο και είναι η συνάρτηση πρόσημου. Αν το $x[n]$ έχει θετική τιμή δίνει 1, αν είναι ίσο με το μηδέν δίνει 0 και αν έχει αρνητική τιμή δίνει -1.

Ενέργεια Βραχέως Χρόνου

Η Ενέργεια Βραχέως Χρόνου ή Short Time Energy(STE) αποτελεί ένα μέτρο της στιγμιαίας ενέργειας του σήματος. Το STE χαρακτηριστικό δίνεται από την σχέση:

$$\text{STE} = \frac{1}{N} \sum_{n=1}^N (x[m]w[n-m])^2 \quad (2.5.4)$$

, όπου $w[n]$ είναι μια συνάρτηση παραθύρου.

Συντελεστές Cepstral

Οι συντελεστές Cepstral ή Cepstral coefficients είναι ένας πολύ κατάλληλος τρόπος για την μοντελοποίηση της κατανομής της φασματικής ενέργειας. Οι συντελεστές Cepstral δίνονται από την σχέση:

$$c(k) = \text{IDFT} \{ \log | \text{DFT}(x[n]) | \} \quad (2.5.5)$$

, όπου IDFT δηλώνει τον Αντίστροφο Διακριτό Μετασχηματισμό Fourier και DFT τον Διακριτό Μετασχηματισμό Fourier.

Η παραπάνω σχέση μπορεί να γραφτεί και ως:

$$c(k) = \frac{1}{N} \sum_{m=0}^{N-1} \log(|X[m]|) e^{j2\pi \frac{mk}{N}} \quad (2.5.6)$$

Οι συντελεστές Cepstral υπολογίζονται για παράθυρα μικρού μήκους με την πάροδο του χρόνου.

Συνήθως επιλέγονται οι πρώτοι συντελεστές που έχουν υπολογιστεί, όπου το πλήθος τους είναι M (όπου $M < N$). Αυτοί οι M συντελεστές χρησιμοποιούνται ως χαρακτηριστικά. Όλοι οι συντελεστές μοντελοποιούν το ακριβές φάσμα, αλλά οι πρώτοι M συντελεστές δίνουν μια γενική εικόνα του φάσματος. Όσο μεγαλύτερο είναι το πλήθος των συντελεστών που θα επιλεγθεί, τόσο μεγαλύτερη είναι η ακρίβεια στην μοντελοποίηση του φάσματος. Επίσης, ο πρώτος συντελεστής Cepstral, που σχετίζεται με την ενέργεια, συνήθως δεν χρησιμοποιείται. Ένα εμπειρικός υπολογισμός του πλήθους των πρώτων συντελεστών που θα επιλεγθούν είναι $M = fs/2000\text{Hz}$.

Όμως, οι συντελεστές Cepstral, κάνουν χρήση γραμμικής κλίμακας στον άξονα συχνοτήτων, αλλά θα ήταν πιο αποδοτικό να χρησιμοποιείται λογαριθμική κλίμακα στον άξονα των συχνοτήτων. Σημαντικό, καθώς και επιθυμητό είναι ότι αν υπάρχει μια μικρή μεταβολή στο διάλυμα χαρακτηριστικών, να προκληθεί μικρή αλλαγή στην αντίληψη αυτής της μεταβολής, και το αντίστροφο.

Ένα χαρακτηριστικό που πληροί τις παραπάνω συνθήκες, είναι ο υπολογισμός των MFCC συντελεστών, όπου η ανάλυση τους θα γίνει παρακάτω.

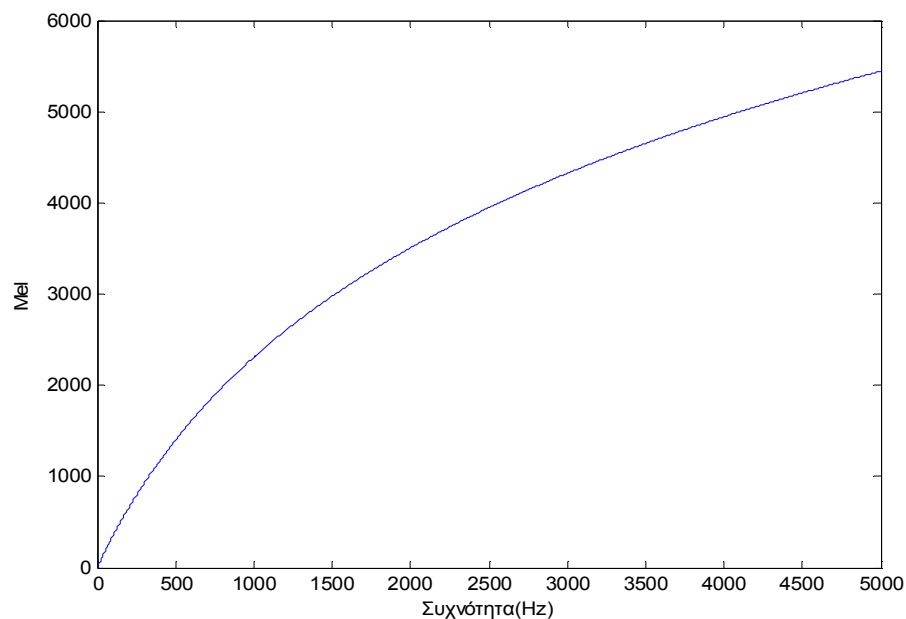
Συντελεστές MFCC, delta και delta-delta

Οι συντελεστές MFCC μπορούμε να πούμε ότι είναι μια βελτίωση των απλών Cepstral συντελεστών και όπως αναφέραμε στην προηγούμενη ενότητα η φασματική ανάλυση γίνεται σε λογαριθμική κλίμακα. Η λογαριθμική κλίμακα που χρησιμοποιείται είναι η κλίμακα Mel, διότι αυτή η κλίμακα βασίζεται στο ανθρώπινο σύστημα ακοής.

Για τον λόγο αυτό οι συντελεστές MFCC, χρησιμοποιούνται ως εξαγωγείς χαρακτηριστικών σε εφαρμογές της αναγνώρισης ομιλίας και μουσικών οργάνων. Η μονάδα Mel είναι μονάδα μέτρησης του αντιληπτού τονικού ύψους ενός σήματος ή η συχνότητα ενός απλού τόνου. Η κλίμακα Mel συνδέεται με την συχνότητα που απεικονίζεται σε γραμμικό άξονα, από την σχέση:

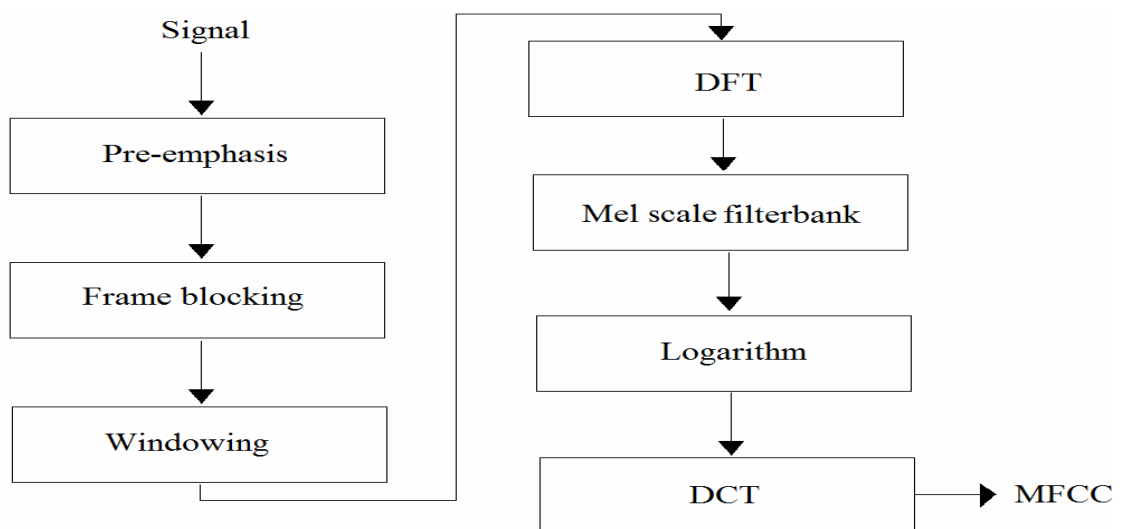
$$f_{\text{Mel}} = 2595 \log_{10}(1 + f/700) \quad (2.5.7)$$

, όπου f_{Mel} είναι η αντιληπτή συχνότητα μονάδα μέτρησης τα Mels και f είναι η πραγματική συχνότητα που μετριέται σε Hz. Η σχέση της κλίμακας Mel με την κλίμακα της συχνότητας (γραμμική κλίμακα), φαίνεται στο παρακάτω σχήμα.



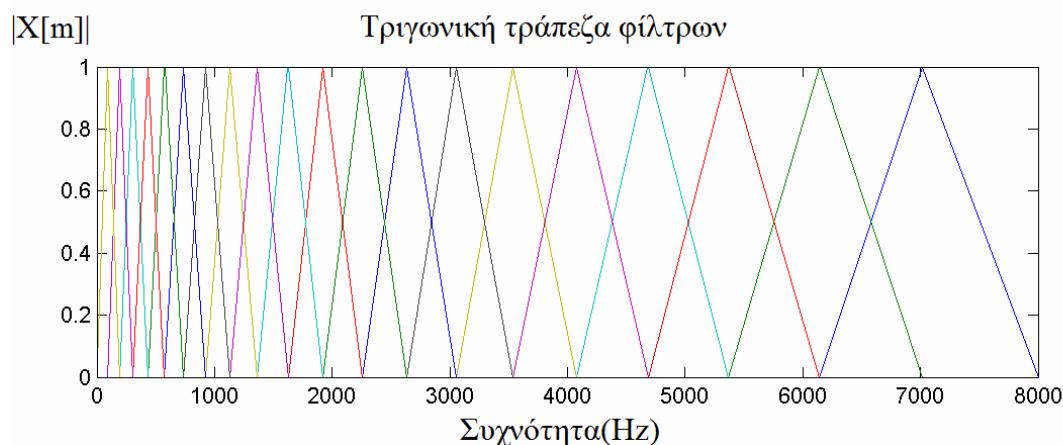
Εικόνα 2.5.1. Κλίμακα Mel σε σχέση με την συχνότητα f .

Η βασική διαδικασία που ακολουθείται για τον υπολογισμό των MFCC είναι η παρακάτω.



Εικόνα 2.5.2. Διαδικασία υπολογισμού του χαρακτηριστικού MFCC.

Αρχικά, το σήμα στέλνεται σε ένα φίλτρο, όπου μπορούν να ενισχυθούν κάποιοι σημαντικοί συντονισμοί υψηλών συχνοτήτων. Στην συνέχεια, το φιλτραρισμένο σήμα χωρίζεται σε πλαίσια(frames) των 20-30ms, με επικάλυψη περίπου το 1/2 -2/3 του μήκους του πλαισίου(frame size). Στο αμέσως, επόμενο στάδιο σε κάθε πλαίσιο εφαρμόζεται ένας τύπος παραθύρου(πολλαπλασιάζεται το σήμα κάθε πλαισίου με μια συνάρτηση παραθύρου). Συνήθως, χρησιμοποιείται παράθυρο Hamming. Ο λόγος, που χρησιμοποιούμε το παράθυρο είναι για να διατηρηθεί η συνέχεια των πρώτων και των τελευταίων δειγμάτων σε κάθε πλαίσιο, δηλαδή να υπάρχει εξομάλυνση της αρχής και του τέλους ενός σήματος για να χρησιμοποιηθεί στον Διακριτό Μετασχηματισμό Fourier. Μετά, για κάθε ένα πλαίσιο εφαρμόζουμε τον Διακριτό Μετασχηματισμό Fourier, και παίρνουμε το μέτρο των τιμών που θα προκύψουν. Ύστερα, χρησιμοποιούμε κάποια τράπεζα φίλτρων(filterbanks). Η τράπεζα φίλτρων αποτελείται από τριγωνικά φίλτρα, τα οποία φαίνονται στην παρακάτω εικόνα. Τέτοια φίλτρα υπολογίζουν το φάσμα γύρω από κάθε κεντρική συχνότητα, όπου τα εύρη ζώνης που την περιέχουν συνεχώς αυξάνονται.



Εικόνα 2.5.3. Τριγωνική τράπεζα φίλτρων.

Έχοντας, τα μέτρα των τιμών, που έχουν προκύψει από τον Διακριτό Μετασχηματισμό Fourier, τα πολλαπλασιάζουμε με ένα σύνολο N ζωνοπερατών τριγωνικών φίλτρων(triangular band-pass filters), για να υπολογίσουμε την ενέργεια κάθε φίλτρου. Μετά, λογαριθμίζουμε τις προκύπτουσες ενέργειες για να υπολογίσουμε την λογαριθμική ενέργεια(log energy) κάθε ζωνοπερατού τριγωνικού φίλτρου. Οι κεντρικές συχνότητες κάθε ζωνοπερατού φίλτρου ισαπέχουν στον άξονα των μελ-συχνοτήτων. Σε αυτό το στάδιο, χρησιμοποιώντας τις λογαριθμικές ενέργειες κάθε φίλτρου, εφαρμόζουμε τον Διακριτό Μετασχηματισμό Συνημίτονου(Discrete Cosine Transform, DCT). Τα αποτελέσματα αυτού του υπολογισμού είναι το διάνυσμα χαρακτηριστικών MFCC. Η σχέση που δίνει τον Διακριτό Μετασχηματισμό Συνημίτονου είναι:

$$C_m = \sum_{k=1}^N S[k] \cos\left[\frac{\pi m}{N}(k - 0.5)\right] \quad (2.5.8)$$

, όπου $m=1,2,\dots,L$ είναι ένα δείκτης που μας δείχνει τον αριθμό του συντελεστή C_m που υπολογίζεται από την παραπάνω σχέση, N είναι ο αριθμός των ζωνοπερατών

τριγωνικών φίλτρων που χρησιμοποιήθηκαν στο προηγούμενο στάδιο, $S[k]$ είναι η λογαριθμική ενέργεια του k ζωνοπερατού τριγωνικού φίλτρου. Λόγω, του ότι έχουμε εφαρμόσει δύο μετασχηματισμούς στο σήμα εισόδου(DFT και DCT), το πεδίο συχνότητας μετασχηματίζεται σε ένα άλλο πεδίο, το οποίο αποκαλείται πεδίο quefrency. Για καλύτερη απόδοση μιας εφαρμογής αναγνώρισης ομιλίας, μουσικού οργάνου, ομιλητή κ.τ.λ. μπορούμε να χρησιμοποιήσουμε τους συντελεστές delta και delta-delta, που προκύπτουν με την βοήθεια των MFCC. Αυτοί οι συντελεστές θα αναλυθούν παρακάτω.

Αν πάρουμε τις χρονικές παραγώγους των MFCC, τότε προκύπτουν νέα χαρακτηριστικά, τα οποία μας δίνουν την ‘ταχύτητα’ και την ‘επιτάχυνση’ των MFCC. Αυτά τα νέα χαρακτηριστικά ονομάζονται συντελεστές delta(‘ταχύτητα’ των MFCC) και συντελεστές delta-delta(‘επιτάχυνση’ των MFCC). Για να υπολογίσουμε τους συντελεστές delta, χρησιμοποιούμε την παρακάτω σχέση:

$$\Delta C[m] = \frac{\sum_{i=-M}^M (c[m+i]i)}{\sum_{i=-M}^M i^2} \quad (2.5.9)$$

, όπου $c[m]$ είναι m^{th} συντελεστής MFCC και M είναι το μέγεθος του πλαισίου. Οι συντελεστές delta-delta, δίνονται από την σχέση:

$$\Delta^2 C[m] = \frac{\sum_{i=-M}^M c[m+i](b_0 i^2 + b_1)}{b_2} \quad (2.5.10)$$

, όπου οι τιμές b_0 , b_1 και b_2 , είναι κάποιες σταθερές τιμές οι οποίες υπολογίζονται από το μέγεθος παραθύρου για τους συντελεστές delta-delta. Οι συντελεστές MFCC είναι ένα διάνυσμα χαρακτηριστικών 13-διαστάσεων (περιέχει 13 συντελεστές, εκ των οποίων, ο πρώτος σχετίζεται με την συνολική μετρούμενη λογαριθμική ενέργεια όλων των φίλτρων), οι συντελεστές delta είναι επίσης ένα διάνυσμα χαρακτηριστικών 13-διαστάσεων, όπως και οι συντελεστές delta-delta είναι ένα διάνυσμα χαρακτηριστικών 13-διαστάσεων. Το συνολικό πλήθος των συντελεστών αυτών είναι 39. Άρα, έχουμε ένα διάνυσμα χαρακτηριστικών 39-διαστάσεων, που το οποίο χρησιμοποιείται και πιο συχνά σε εφαρμογές αναγνώρισης ομιλίας.

ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ

3.1 Εισαγωγή στην Αναγνώριση Προτύπων

Η αναγνώριση προτύπων θεωρείται ένα βασικό χαρακτηριστικό των έμβιων όντων, όπως είναι και ο άνθρωπος. Έτσι, στην καθημερινή ζωή οι άνθρωποι έχουν την εμπειρία της αναγνώρισης της μητέρα τους από την φωνή της, του πρόσωπου ενός φίλου από μια φωτογραφία, της ελαστικότητας μιας επιφάνειας από την αφή κ.τ.λ. Στα ερεθίσματα αυτά: αντικείμενα, ιδέες, ήχους, εικόνες, έμβια όντα κ.τ.λ., τα οποία καλούμαστε να αναγνωρίσουμε, τους δίνουμε την ονομασία πρότυπα(patterns). Τα πρότυπα είναι οποιοδήποτε ερέθισμα που μπορεί να περιγραφεί, και τα οποία θέλουμε να ταξινομήσουμε σε έναν αριθμό κατηγοριών. Επομένως, μπορούμε να δώσουμε τον ακόλουθο ορισμό της αναγνώρισης προτύπων:

Η Αναγνώριση Προτύπων(Pattern Recognition) είναι ο επιστημονικός κλάδος, που αντικείμενο της είναι η περιγραφή και η κατάταξη προτύπων σε έναν αριθμό κατηγοριών(κλάσεων).

Ένα σύστημα(έναν αλγόριθμος), ο οποίος έχει σχεδιαστεί και υλοποιηθεί ώστε να αναγνωρίζει πρότυπα ονομάζεται ταξινομητής(classifier). Η λειτουργία του ταξινομητή χωρίζεται σε δύο φάσεις : α) την εκπαίδευση(training) και β) την ταξινόμηση(classify). Στην φάση της εκπαίδευσης, βασικό είναι να υπάρχει ένα πλήθος προτύπων για την εκπαίδευση του ταξινομητή, το οποίο καλείται σύνολο εκπαίδευσης(training set). Στην συνέχεια, προσδιορίζονται οι κλάσεις ή οι συγκεντρώσεις(clusters) των προτύπων και όλοι οι παράμετροι του συστήματος για τους οποίους θα καταστήσουν το σύστημα ικανό για την ταξινόμηση των προτύπων στις σωστές κλάσεις. Στη φάση της ταξινόμησης, το σύστημα αποδίδει ένα πρότυπο σε μια κλάση ή σε μια συγκέντρωση.

Ο τρόπος που θα γίνει η εκπαίδευση του ταξινομητή χωρίζεται σε τρεις βασικές κατηγορίες:

i)Την επιβλεπόμενη εκπαίδευση ή μάθηση(supervised learning): σε αυτή την κατηγορία είναι γνωστή η κλάση, στην οποία ανήκει το κάθε πρότυπο του συνόλου εκπαίδευσης και βασική επίτευξη είναι η απόδοση ενός προτύπου που δεν είναι στο σύνολο εκπαίδευσης, σε μια από τις προκαθορισμένες κλάσεις.

ii)Την μη- επιβλεπόμενη εκπαίδευση ή μάθηση(unsupervised learning): σε αυτή την κατηγορία είναι άγνωστη η κλάση, στην οποία ανήκει το κάθε πρότυπο του συνόλου εκπαίδευσης και βασική επίδωξη είναι η εύρεση των συγκεντρώσεων των προτύπων(clustering).

iii) Ενισχυτική εκπαίδευση ή μάθηση(Reinforcement learning): σε αυτή την κατηγορία εισάγουμε ένα πρότυπο στον ταξινομητή στην φάση της εκπαίδευσης.

Υπολογίζουμε δοκιμαστικά την κλάση στην οποία ανήκει και στην συνέχεια χρησιμοποιούμε την γνωστή κλάση, που ανήκει το πρότυπο βελτιώνοντας έτσι την απόδοση του ταξινομητή. Δηλαδή, στην αρχή δεν δίνεται η κλάση στην οποία ανήκει το πρότυπο και το σύστημα το ταξινομεί σε μια κλάση, η οποία λέγεται δοκιμαστική. Στην συνέχεια στο σύστημα χρησιμοποιείται η διαδικασία της ανατροφοδότησης, που είναι επιβλεπόμενη και αποφασίζεται εάν η δοκιμαστική κλάση είναι σωστή ή όχι.

3.2 Ταξινόμηση προτύπων

Για την ταξινόμηση των προτύπων σε διάφορες κλάσεις, συνήθως χρησιμοποιούνται κάποιες διακρίνουσες συναρτήσεις ή συναρτήσεις απόφασης. Γι' αυτές τις συναρτήσεις, συνήθως δεν υπάρχουν πολλές πληροφορίες, με αποτέλεσμα να γίνεται χρήση κάποιων 'αυθαίρετων' συναρτήσεων και με την χρήση κάποιου αλγόριθμου βελτιστοποίησης αυτές αποκτούν ικανοποιητική μορφή για την έκφραση της κάθε κλάσης.

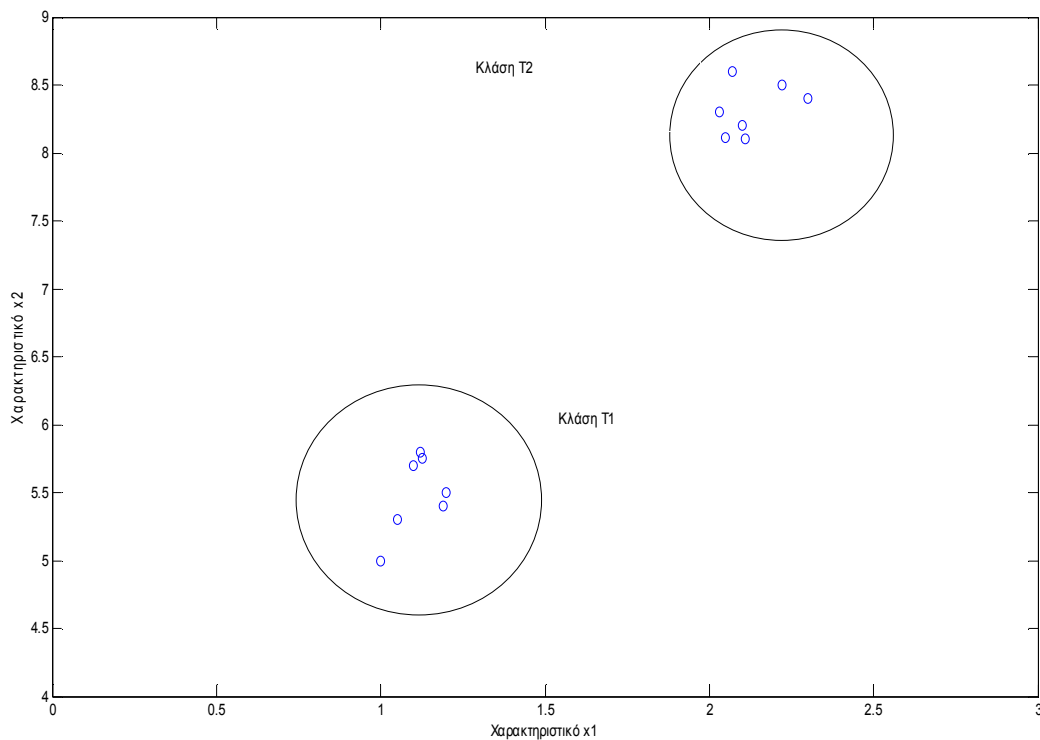
Έστω, ότι έχουμε ένα σύνολο Ω , που περιέχει K πρότυπα και τα εξετάζουμε ως προς N χαρακτηριστικά και έστω, ότι θέλουμε να τα ταξινομήσουμε σε T κλάσεις ή τάξεις.

Τότε, οποιοδήποτε πρότυπο το θεωρούμε ως ένα N -διάστατο διάνυσμα το οποίο παριστάνεται σε ένα N -διάστατο Ευκλείδιο χώρο(E^N)ή χώρο χαρακτηριστικών (feature space). Το πρότυπο αυτό είναι ένα διάνυσμα στήλης ή ένας πίνακας στήλης τύπου $N \times 1$ και συμβολίζεται με \mathbf{x} :

$$\mathbf{x} = [x_1, x_2, \dots, x_N]^T$$

,όπου $[.]^T$ εκφράζει τον ανάστροφο πίνακα, (x_1, x_2, \dots, x_N) είναι τα χαρακτηριστικά ως προς τα οποία εξετάζουμε ένα πρότυπο.

Επομένως, τα πρότυπα παριστάνονται ως σημεία(που είναι τα περατά των διανυσμάτων, λαμβάνοντας ως αρχή την αρχή των αξόνων) και οι κλάσεις είναι σύνολο από σημεία, τα οποία απέχουν λίγο ή πολύ μεταξύ τους. Ένα παράδειγμα φαίνεται στην εικόνα 3.3.1.



Εικόνα 3.1.1. Παράδειγμα απεικόνισης κλάσεων, που περιέχουν πρότυπα που έχουν εξεταστεί ως προς δύο χαρακτηριστικά (x_1, x_2) στον διδιάστατο Ευκλείδιο χώρο.

Βασική επιδίωξη του συστήματος είναι η ταξινόμηση των προτύπων σε μία από τις T κλάσεις, που έχουμε καθορίσει ότι πρέπει να ανήκουν τα πρότυπα.

Ας υποθέσουμε ότι οι διακρίνουσες συναρτήσεις, που καθορίζουν κατά κάποιο τρόπο τα όρια της κάθε κλάσης είναι της μορφής: $g_1(x), g_2(x), \dots, g_T(x)$ και είναι μονοσήμαντες ως προς το διάνυσμα ενός προτύπου x .

Εάν $g_i(x) > g_j(x)$, για κάθε $i \neq j$, τότε αποφάσισε ότι το πρότυπο x ανήκει στην κατηγορία ω_i , στην περίπτωση που η κλάση ω_i σχετίζεται με την συνάρτηση $g_i(x)$ και $i, j = 1, 2, \dots, T$. Δηλαδή, το πρότυπο ανήκει στην κλάση ω_i που σχετίζεται με την διακρίνουσα συνάρτηση που έχει την μέγιστη τιμή.

Οι διακρίνουσες συναρτήσεις μπορεί να είναι ντετερμινιστικές (αιτιοκρατικές) ή στοχαστικές. Επίσης, μπορεί να είναι ορισμένοι αλγόριθμοι που παίρνουν σαν είσοδο πρότυπα με την μορφή διανυσμάτων, και καθιστούν ένα σύστημα ταξινόμησης ικανό να ταξινομήσει ένα πρότυπο στη σωστή κλάση.

Στις ντετερμινιστικές μεθόδους ανήκουν αλγόριθμοι που κάνουν χρήση γεωμετρικών αποστάσεων όπως ο K-means, Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks) κ.τ.λ., ενώ στις στοχαστικές μεθόδους ανήκουν αλγόριθμοι που κάνουν χρήση πιθανοτήτων και σε αυτή την κατηγορία ανήκουν τα κρυφά μοντέλα Markov (Hidden Markov Models, HMM), τα μοντέλα Γκαουσιανών μίξεων (Gaussian mixture models, GMM) και αλγόριθμοι που χρησιμοποιούν την απόφαση Bayes.

Στη συνέχεια, παρουσιάζονται οι πιο γνωστές ντετερμινιστικές και στοχαστικές μέθοδοι ταξινόμησης προτύπων.

3.2.1 Ντετερμινιστική ταξινόμηση προτύπων

3.2.1.1 Γεωμετρικές αποστάσεις

Όπως αναφέραμε και στην παραπάνω ενότητα, στη ντετερμινιστική ταξινόμηση προτύπων κατατάσσονται οι ταξινομητές, οι οποίοι χρησιμοποιούν αλγόριθμους που κάνουν χρήση των γεωμετρικών αποστάσεων. Έτσι, γίνεται χρήση κάποιων συναρτήσεων απόστασης που βασίζονται στον κανόνα του κοντινότερου γείτονα και βασική επιδίωξη τόσο στην εκπαίδευση όσο και στην αναγνώριση είναι η ταξινόμηση των εισαχθέντων προτύπων στο σύστημα σε μια από τις προκαθορισμένες κλάσεις. Οι πιο γνωστές συναρτήσεις απόστασης είναι:

Η Ευκλείδεια απόσταση (Euclidean distance): έστω ότι έχουμε δυο πρότυπα x και y , τα οποία παριστάνονται ως διανύσματα N διαστάσεων στον N διαστάσεων Ευκλείδειο χώρο. Τότε η Ευκλείδεια απόσταση δίνεται από την σχέση:

$$d_E(\bar{x}, \bar{y}) = \sqrt{\sum_{v=1}^N (x(v) - y(v))^2} \quad (3.2.1)$$

ή

$$d_E(\bar{x}, \bar{y}) = \sqrt{(\mathbf{x} - \mathbf{y})(\mathbf{x} - \mathbf{y})^T} \quad (3.2.2)$$

, όπου N είναι το πλήθος των χαρακτηριστικών ως προς τα οποία εξετάζουμε τα πρότυπα x και y . Η παραπάνω σχέση εκφράζει την Ευκλείδεια απόσταση του διανύσματος x από το διάνυσμα y και χρησιμοποιείται ευρέως δίνοντας αρκετά καλά αποτελέσματα.

Επομένως, αν $d_E(\bar{x}, \bar{y})=0$, τότε λέμε ότι τα πρότυπα x και y είναι ίσα μεταξύ τους

και έτσι η Ευκλείδεια απόσταση μπορεί να χρησιμοποιηθεί και ως κριτήριο ισότητας δυο προτύπων (διανυσμάτων).

Η απόσταση Mahalanobis (Mahalanobis distance): η απόσταση Mahalanobis εκφράζει όπως και η Ευκλείδεια απόσταση, την απόσταση ενός διανύσματος x από ένα διάνυσμα y σε έναν N διαστάσεων χώρο και δίνεται από την σχέση:

$$D_R(\bar{x}, \bar{y}) = (\mathbf{x} - \mathbf{y})\Sigma^{-1}(\mathbf{x} - \mathbf{y})^T \quad (3.2.3)$$

, όπου Σ είναι ο πίνακας συνδιασποράς του διανύσματος x και y (την διαφορά των τιμών των χαρακτηριστικών του πίνακα y από τις αντίστοιχες τιμές των χαρακτηριστικών του πίνακα x), Σ^{-1} είναι ο αντίστροφος του πίνακα συνδιασποράς και $(.)^T$ εκφράζει τον ανάστροφο πίνακα.

3.2.1.2 Γραμμικός διαχωρισμός και γραμμικές διακρίνουσες συναρτήσεις

Τα σημεία στον N-διάστατο χώρο χαρακτηριστικών (δηλαδή, τα πρότυπα που θέλουμε να ταξινομήσουμε), διαχωρίζονται μεταξύ τους με τρόπο που καθορίζεται από τις συναρτήσεις απόφασης και στην γραμμική τους μορφή δίνονται από την παρακάτω σχέση:

$$\theta = \sum_{v=1}^N w_v x_v \quad (3.2.4)$$

$$\theta = \mathbf{w}^T \mathbf{x} \quad (3.2.5)$$

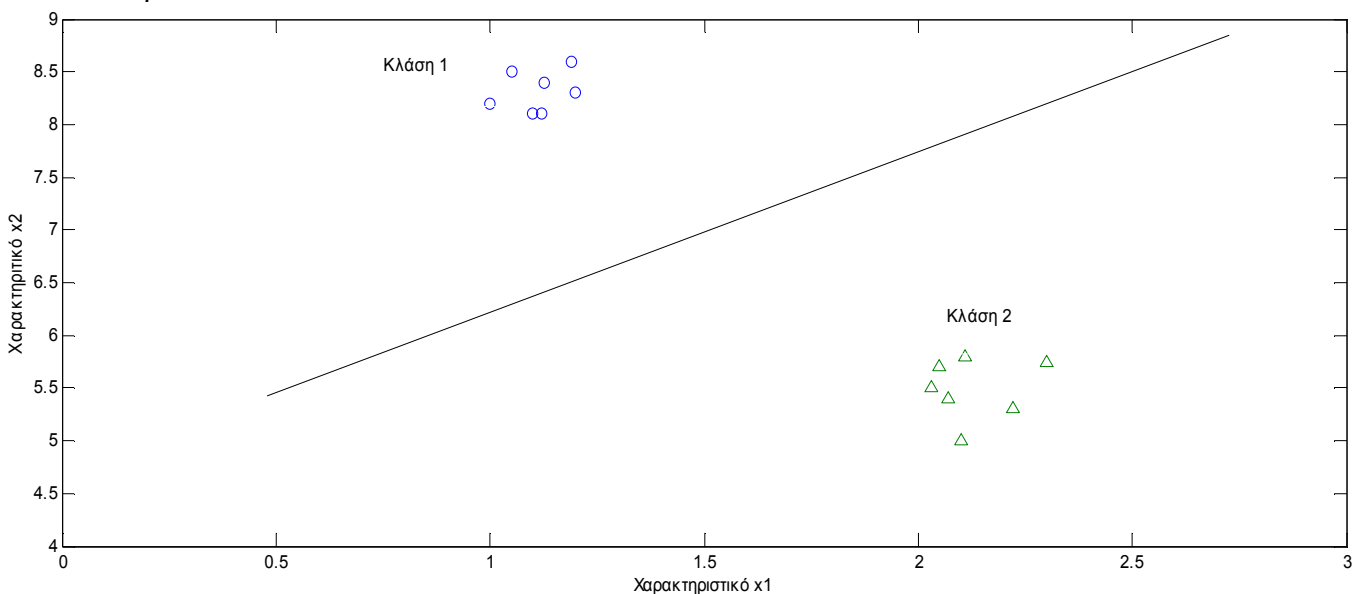
, όπου \mathbf{w} είναι το διάνυσμα διαχωρισμού και είναι ίσο με: $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_N]^T$ και εκφράζει τους συντελεστές των αγνώστων μιας γραμμικής εξίσωσης της μορφής $w_1 x_1 + w_2 x_2 + \dots + w_N x_N = \theta$ και \mathbf{x} είναι διάνυσμα εισόδου, δηλαδή το σύνολο εκπαίδευσης (training set).

Υπολογίζοντας το διάνυσμα $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_N]^T$, διαμορφώνεται κατάλληλα το υπερεπιπέδο που διαχωρίζει τις κλάσεις των προτύπων.

Στην περίπτωση, που θέλουμε να διαχωρίσουμε δύο κλάσεις μεταξύ τους και εξετάζουμε τα πρότυπα ως προς 2 χαρακτηριστικά, τότε η γραμμική εξίσωση είναι της μορφής:

$$w_1 x_1 + w_2 x_2 = \theta \quad (3.2.6)$$

Η παραπάνω σχέση είναι η εξίσωση μιας ευθείας και αυτή η ευθεία θα διαχωρίσει τις δυο κλάσεις μεταξύ τους. Σε αυτή την περίπτωση λέμε ότι τα σημεία της κάθε κλάσης είναι γραμμικά διαχωρίσιμα μεταξύ τους. Ένα παράδειγμα φαίνεται στην παρακάτω εικόνα.



Εικόνα 3.2.1. Παράδειγμα γραμμικού διαχωρισμού δύο κλάσεων από μια ευθεία.

Στο παραπάνω παράδειγμα προϋπόθεση είναι οι τιμές των χαρακτηριστικών να είναι τέτοιες ώστε να διαχωρίζονται γραμμικά τα σημεία των δύο κατηγοριών(κλάσεων) και έτσι να μπορούν να διαχωριστούν από μια ευθεία της μορφής: $w_1x_1+w_2x_2 = \theta$. Στην περίπτωση που εξετάζουμε τα πρότυπα ως προς περισσότερα των δύο χαρακτηριστικών, τότε ο διαχωρισμός των κλάσεων γίνεται μέσω ενός υπερεπιπέδου, που δίνεται από την σχέση 3.2.4.

3.2.1.3 Τεχνητά Νευρωνικά Δίκτυα

Τα Τεχνητά Νευρωνικά Δίκτυα(Artificial Neural Networks) είναι ένα πλήθος από νευρώνες, οι οποίοι συνδέονται μεταξύ τους. Οι νευρώνες είναι κάποιες μονάδες επεξεργασίας(Processing Units). Κάθε νευρώνας έχει ένα πεπερασμένο πλήθος εισόδων και μία μόνο έξοδο, η οποία είναι είσοδος σε κάποιους άλλους νευρώνες. Κάθε είσοδος έχει και ένα συντελεστή στάθμισης(ή βάρους) και προσδιορίζει την σημαντικότητα της κάθε σύνδεσης, δηλαδή το πόσο συμβάλει κάθε είσοδος στο δίκτυο. Η επεξεργασία που γίνεται στον κάθε νευρώνα καθορίζεται από τη συνάρτηση μεταφοράς, η οποία καθορίζει την κάθε έξοδο συναρτήσει των εισόδων και των συντελεστών βάρους. Για συντομία, τα Τεχνητά Νευρωνικά Δίκτυα θα γράφονται με χρήση των ακρωνύμιων τους, ως ΤΝΔ.

Όπως αναφέρθηκε και πάνω ένας νευρώνας είναι μια μονάδα επεξεργασίας πληροφορίας. Ένα μοντέλο ενός νευρώνα χαρακτηρίζεται από τρία βασικά στοιχεία, τα οποία είναι τα εξής:

- Ένα σύνολο από συνάψεις ή συνδέσεις.
- Έναν αθροιστή και
- Μια συνάρτηση ενεργοποίησης ή μια συνάρτηση μεταφοράς.

Κάθε σύνδεση έχει και ένα συντελεστή βάρους, έτσι ένα σήμα x_n στην είσοδο της σύναψης v που συνδέεται στον νευρώνα T , πολλαπλασιάζεται με το συντελεστή βάρους w_v . Το βάρος w_v είναι θετικό αν η σύνδεση είναι διεγερτική και αρνητικό αν σύνδεση είναι απαγορευτική. Ο αθροιστής χρησιμοποιείται για την πρόσθεση των σημάτων εισόδου και τέλος η συνάρτηση ενεργοποίησης(ή μεταφοράς) χρησιμοποιείται για τη μείωση του εύρους της εξόδου του νευρώνα.

Νευρωνικό Δίκτυο Perceptron

Μια απλή μορφή ενός Νευρωνικού Δικτύου είναι το Perceptron . Το Νευρωνικό Δίκτυο Perceptron προτάθηκε το 1958 από τον Rosenblatt. Ένα τέτοιο σύστημα ήταν μια από τις πρώτες προσπάθειες ώστε να μπορεί να υπάρχει η δυνατότητα να ταξινομηθούν πρότυπα, τα οποία ανήκουν σε δύο κλάσεις, έχοντας σαν προϋπόθεση ότι τα πρότυπα είναι γραμμικά διαχωρίσιμα. Αυτό το απλό δίκτυο είναι ένας απλός νευρώνας. Στην περίπτωση που έχουμε παραπάνω από μία κλάσεις ή τάξεις, για παράδειγμα T , που μπορούμε να ταξινομήσουμε κάποια πρότυπα, τότε χρησιμοποιούμε T Perceptrons, ένα για κάθε κλάση, θεωρώντας θετικά πρότυπα για

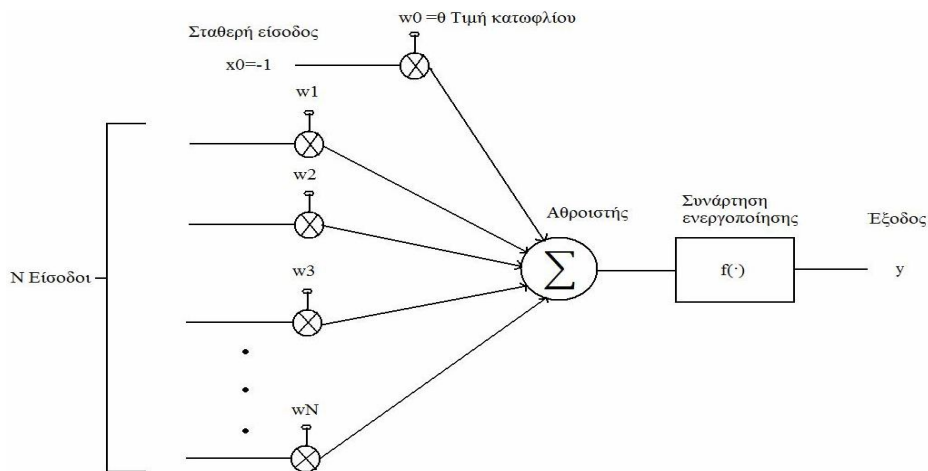
τη κλάση K τα πρότυπα που ανήκουν στην K , και αρνητικά τα πρότυπα των υπόλοιπων κατηγοριών (K').

Η έξοδος του Perceptron δίνεται από τη σχέση

$$y = f\left(\sum_{v=0}^N w_v x_v\right) = f(\mathbf{w}^T \mathbf{x}) \quad (3.2.7)$$

,όπου $f\left(\sum_{v=0}^N w_v x_v\right)$ ή $f(\mathbf{w}^T \mathbf{x})$ είναι η συνάρτηση σκληρού περιοριστή, η οποία δίνει -1 όταν $\left(\sum_{v=0}^N w_v x_v\right) < 0$ και $+1$ όταν $\left(\sum_{v=0}^N w_v x_v\right) > 0$. Η συνάρτηση 3.2.7 περιγράφει πλήρως ένα νευρώνα.

Τα διανύσματα \mathbf{w} και \mathbf{x} είναι N διαστάσεων και θεωρούμε ότι $x_0 = -1$ και $w_0 = \theta$, όπου το θ ονομάζεται κατώφλι (threshold) ή πόλωση (bias).



Εικόνα 3.2.2. Μοντέλο ενός Νευρωνικού Δικτύου Perceptron .

Εάν το συνολικό άθροισμα από τις υπόλοιπες εισόδους του νευρώνα είναι μεγαλύτερο από την τιμή θ , τότε ο νευρώνας ενεργοποιείται, διαφορετικά παραμένει ανενεργός. Στο παράδειγμα που περιγράφηκε παραπάνω η βασική επιδίωξη του Perceptron είναι να ταξινομήσει τα πρότυπα εισόδου \mathbf{x} στις δύο κλάσεις K και K' .

Η συνάρτηση ενεργοποίησης, που είδαμε παραπάνω, ορίζει την έξοδο ενός νευρώνα σε σχέση με το βαθμό ενεργοποίησης της εισόδου. Οι συνήθεις συναρτήσεις ενεργοποίησης είναι οι εξής:

- Συνάρτηση Κατώφλιου ή βηματική συνάρτηση
- Τμηματικά Γραμμική Συνάρτηση
- Σιγμοειδής Συνάρτηση

Θεωρούμε ότι $x = \left(\sum_{v=1}^N w_v x_v - \theta\right)$. Η βηματική συνάρτηση δίνεται από την σχέση:

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (3.2.8)$$

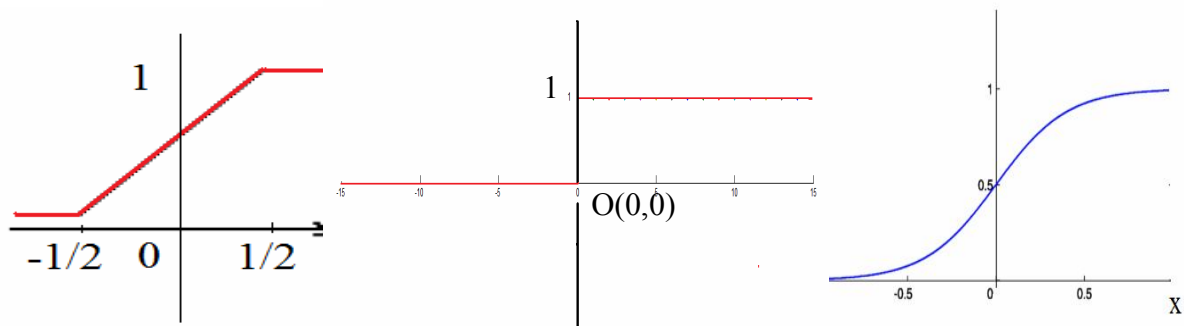
Η τμηματική γραμμική συνάρτηση δίνεται από την σχέση:

$$f(x) = \begin{cases} 1, & x \geq 1/2 \\ x, & -1/2 < x < 1/2 \\ 0, & x \leq -1/2 \end{cases} \quad (3.2.9)$$

Ενώ, η σιγμοειδής συνάρτηση, η οποία είναι η πιο συνηθισμένη μορφή συνάρτησης ενεργοποίησης που χρησιμοποιείται στα ΤΝΔ, δίνεται από την σχέση:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.2.10)$$

Οι γραφικές παραστάσεις, της βηματικής συνάρτησης, της τμηματικής γραμμικής συνάρτησης και της σιγμοειδής συνάρτησης φαίνονται στην παρακάτω εικόνα.



Εικόνα 3.2.3. Τμηματική γραμμική συνάρτηση(Αριστερά), Βηματική συνάρτηση(Μέση) και σιγμοειδής συνάρτηση(Δεξιά).

Αφού παρουσιάσαμε τις συναρτήσεις ενεργοποίησης είναι εύλογο να παρουσιάσουμε και τη λειτουργία του Νευρωνικού Δικτύου Perceptron. Ο αλγόριθμος μάθησης φαίνεται παρακάτω:

Στο Perceptron εισάγουμε ένα ένα τα πρότυπα του συνόλου εκπαίδευσης(training set) και παίρνουμε την έξοδό του. Η έξοδος του πρέπει να είναι +1 όταν του παρουσιάζεται πρότυπο που ανήκει στην κατηγορία Κ και -1 όταν ανήκει στην κατηγορία Κ'. Όταν η έξοδος είναι σωστή τότε δεν χρειάζεται να επιτελεστεί κάποια ενέργεια και επομένως εισάγουμε το επόμενο πρότυπο. Στην περίπτωση που το αποτέλεσμα της εξόδου είναι λάθος, τότε πρέπει να τροποποιηθούν οι συντελεστές βάρους w_i με τέτοιο τρόπο που παρουσιάζεται παρακάτω:

Αν το x ανήκει στην κλάση Κ και η έξοδος του Perceptron είναι -1 τότε:

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \mathbf{ax}(t) \quad (3.2.11)$$

Αν το x ανήκει στην κλάση Κ' και η έξοδος είναι +1 τότε

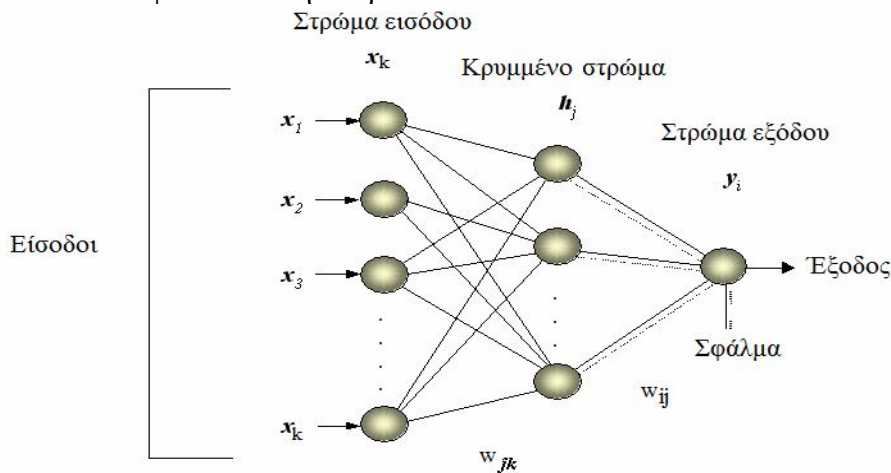
$$\mathbf{w}(t+1) = \mathbf{w}(t) - \mathbf{ax}(t) \quad (3.2.12)$$

Η ποσότητα a λέγεται παράμετρος εκμάθησης και παίρνει τιμές από το διάστημα (0,1). Το t είναι ένας μετρητής επανάληψης της διαδικασίας της εκμάθησης. Έτσι,

σκοπός ενός Νευρωνικού Δικτύου Perceptron είναι να διορθώσει το σφάλμα όταν ένα πρότυπο ταξινομείται σε λάθος κλάση, με τελικό αποτέλεσμα την ορθή ταξινόμηση των προτύπων στις κλάσεις στις οποίες ανήκουν. Συνήθως, τα βάρη w_i αρχικοποιούνται σε μηδενικές τιμές ή σε πολύ μικρές τυχαίες τιμές. Η διαδικασία τερματίζεται όταν το Perceptron ταξινομήσει σωστά όλα τα πρότυπα ή όταν ο αριθμός των λαθών ταξινόμησης είναι ο ελάχιστος δυνατός.

Πολύστρωματικό Νευρωνικό δίκτυο Perceptron και Αλγόριθμος
όπισθεν διάδοσης σφάλματος (Back-propagation error algorithm)

Το Νευρωνικό δίκτυο Perceptron που περιγράφηκε στην προηγούμενη ενότητα είναι ενός επιπέδου (στρώματος) ΤΝΔ (Single Layer Artificial Neural Network), ενώ τα πολυεπίπεδα Perceptron αποτελούν γενίκευση του απλού Perceptron και εκτός από το στρώμα εισόδου και εξόδου περιλαμβάνουν και ένα ή περισσότερα κρυμμένα στρώματα (Hidden Layers). Έτσι, η πληροφορία από το επίπεδο εισόδου μεταφέρεται στους νευρώνες του ενδιάμεσου επιπέδου ή των ενδιάμεσων επιπέδων και από εκεί στο επίπεδο εξόδου. Συνήθως κάθε νευρώνας του επιπέδου εισόδου ή ενός ενδιάμεσου επιπέδου συνδέεται με όλους τους νευρώνες του επόμενου επιπέδου. Ένα τέτοιο δίκτυο φαίνεται στην παρακάτω εικόνα.



Εικόνα 3.2.4. Πολυεπίπεδο πρόσθιας τροφοδότησης ΤΝΔ με τρία στρώματα Εισόδου-Κρυμμένου-Εξόδου.

Ο πιο συνηθισμένος τρόπος εκπαίδευσης πολυεπίπεδων ΤΝΔ πρόσθιας ανατροφοδότησης είναι ο αλγόριθμος της όπισθεν διάδοσης σφάλματος, ο οποίος επινοήθηκε πρώτα από τον Werbos το 1974 και βελτιώθηκε αργότερα από τον Parker και από τους Rumelhart, Hinton and Williams.

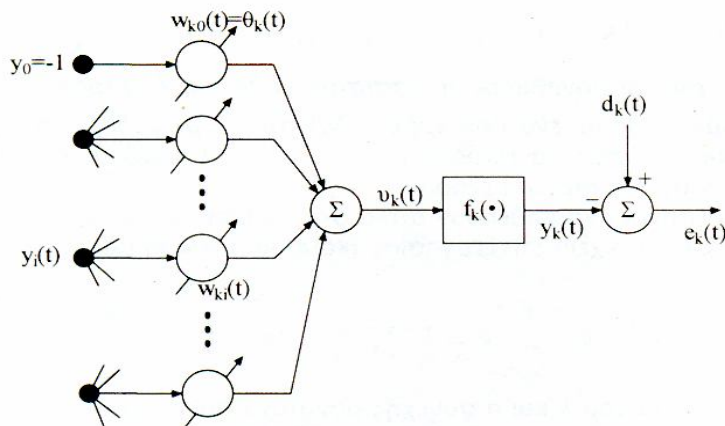
Βασικός σκοπός κατά την εκτέλεση του αλγόριθμου όπισθεν διάδοσης σφάλματος είναι ότι με βάση την ελαχιστοποίηση του τετραγωνικού σφάλματος της εξόδου του Νευρωνικού Δικτύου να γίνεται η αλλαγή των τιμών των συντελεστών βάρους των νευρώνων. Δηλαδή, οι τιμές των συντελεστών βάρους αλλάζουν ώστε το τετραγωνικό σφάλμα :

$$E(t) = \frac{1}{2} \sum_{i=1}^M e_i^2(t) \quad (3.2.13)$$

να ελαχιστοποιείται. Στην παραπάνω σχέση η ποσότητα $e_i^2(t)$ είναι ίση με:

$$e_i^2(t) = d_i(t) - y_i(t) \quad (3.2.14)$$

, όπου $d_i(t)$ είναι η αναμενόμενη ή επιθυμητή έξοδος του δικτύου και $y_i(t)$ είναι η πραγματική έξοδος του Νευρωνικού δικτύου και $i = 1, 2, \dots, M$, που είναι ο δείκτης του προσδιορίζει τον νευρώνα εξόδου και M είναι το πλήθος των νευρώνων εξόδου.



Εικόνα 3.2.5. Δομή ενός νευρώνα εξόδου.

Από το παραπάνω σχήμα παρατηρούμε ότι η έξοδος του k νευρώνα, που βρίσκεται στο επίπεδο εξόδου είναι ίση με:

$$v_k(t) = \sum_{i=0}^n w_{ki}(t) y_i(t) \quad (3.2.15)$$

, όπου $y_i(t)$ είναι η έξοδος του προηγούμενου επιπέδου και w_{ki} είναι ο συντελεστής βάρους που συνδέει τον k νευρώνα με τον i νευρώνα του προηγούμενου επιπέδου. Όπως αναφέραμε παραπάνω, στόχος της εκπαίδευσης είναι η ελαχιστοποίηση του τετραγωνικού σφάλματος της εξόδου, προσαρμόζοντας τις τιμές των συντελεστών βάρους του δικτύου.

Η λειτουργία του αλγόριθμου χωρίζεται σε δύο στάδια. Το πρώτο αναφέρεται ως πέρασμα εμπρόσθιας φοράς (forward pass) και το δεύτερο ως πέρασμα ανάστροφης φοράς (reverse pass). Όταν εκτελείται το πέρασμα εμπρόσθιας φοράς, τότε τα βάρη των συνδέσεων ή συνάψεων παραμένουν αναλλοίωτα σε όλο το δίκτυο σε κάθε νευρώνα και υπολογίζονται οι αντίστοιχες έξοδοι. Αρχικά εισάγονται τα χαρακτηριστικά ως προς τα οποία εξετάσαμε το πρότυπο, σαν είσοδοι στο δίκτυο και στην συνέχεια εισάγονται στο πρώτο κρυμμένο επίπεδο. Αν υπάρχει και άλλο κρυμμένο επίπεδο εισάγονται σε αυτό οι έξοδοι του προηγούμενου κρυμμένου επιπέδου και τέλος οι έξοδοι του τελευταίου κρυμμένου επιπέδου εισάγονται στο επίπεδο εξόδου με τον υπολογισμό του σφάλματος για κάθε νευρώνα αυτού του επιπέδου. Στο δεύτερο στάδιο, της ανάστροφης φοράς, το πέρασμα αρχίζει από το επίπεδο εξόδου με σκοπό να διαδοθεί μέχρι την είσοδο του δικτύου. Σε αυτό το στάδιο μέσω κάποιων υπολογισμών καθίσταται εφικτή η μεταβολή των τιμών των συντελεστών βάρους. Και οι υπολογισμοί γίνονται μέχρι να διαδοθεί το τετραγωνικό σφάλμα εξόδου στο επίπεδο εισόδου του δικτύου.

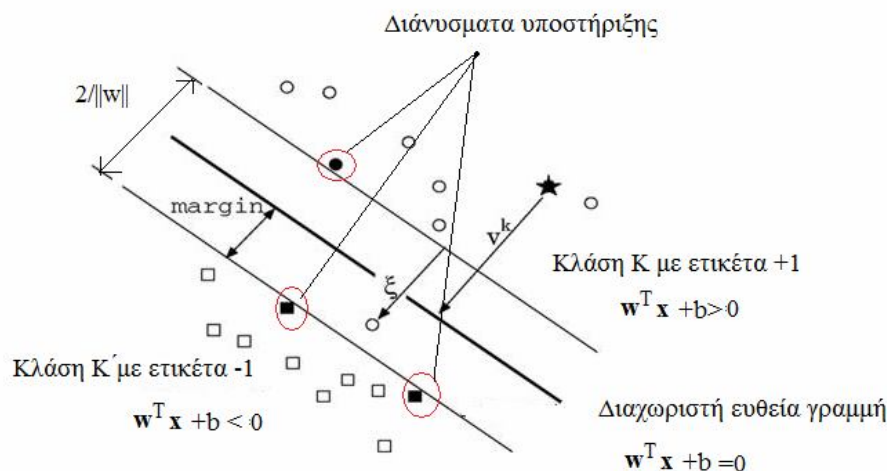
Η διαδικασία των δύο προαναφερθέντων περασμάτων εκτελείται για κάθε πρότυπο του συνόλου εκπαίδευσης με μία κυκλική επανάληψη μέχρι να εισαχθούν στο δίκτυο όλα τα πρότυπα του συνόλου εκπαίδευσης. Έτσι ταξινομούνται τα πρότυπα του συνόλου εκπαίδευσης με το ελάχιστο σφάλμα.

3.2.1.4 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)

Το πολυεπίπεδο Perceptron είναι ένα Νευρωνικό Δίκτυο εμπρόσθιας τροφοδότησης. Οι Μηχανές Διανυσμάτων Υποστήριξης που θα παρουσιαστούν σε αυτή την ενότητα ανήκουν στα Νευρωνικά Δίκτυα Ανατροφοδότησης (Neural Networks feedback) και επινοήθηκαν από τον Vapnik.

Έστω ότι θέλουμε να χωρίσουμε τα πρότυπα σε δύο κλάσεις και συγκεκριμένα στην κλάση K και στην κλάση K' . Τα πρότυπα που ανήκουν στην κλάση K , χαρακτηρίζονται από την ετικέτα τους που είναι ίση με $+1$ και τα πρότυπα που ανήκουν στην κλάση K' έχουν ετικέτα -1 .

Εισάγοντας ενός νέο πρότυπο στο σύστημα θέλουμε να προβλέψουμε σε ποια κλάση ανήκει. Για να γίνει η αναγνώριση των προτύπων, χρειάζεται ο προσδιορισμός ενός ορίου ανάμεσα στις δύο κλάσεις και ο έλεγχος της θέσης, του υπό εξέταση προτύπου. Το όριο αυτό θα είναι μια ευθεία γραμμή αν τα πρότυπα παριστάνονται σε έναν Ευκλείδειο χώρο δύο διαστάσεων (E^N , με $N=2$) ή ένα υπερεπίπεδο, αν οι διαστάσεις του Ευκλείδειου χώρου ξεπερνάνε τις δύο (E^N , με $N>2$). Για απλοποίηση της διαδικασίας θεωρούμε ότι έχουμε εξετάσει τα πρότυπα του συνόλου εκπαίδευσης προς δύο χαρακτηριστικά, άρα ο Ευκλείδειος χώρος είναι δύο διαστάσεων και το όριο ανάμεσα στις δύο κλάσεις K και K' είναι μια ευθεία γραμμή. Αυτή η περίπτωση φαίνεται στην εικόνα που ακολουθεί.



Εικόνα 3.2.6. Παράδειγμα γραμμικού διαχωρισμού κλάσεων.

Η ιδανική ευθεία γραμμή (όριο) που μπορεί να διαχωρίσει τις δύο κλάσεις είναι η μεσοπαράλληλος των παραλλήλων ευθειών (με αχνό μαύρο χρώμα) που φαίνονται στην παραπάνω εικόνα. Η ιδανική αυτή ευθεία γραμμή (ή ιδανικό υπερεπίπεδο) ονομάζεται Διαχωριστή ευθεία γραμμή (ή Διαχωριστό υπερεπίπεδο εάν E^N , έχει τιμή $N>2$). Η απόσταση ανάμεσα στις δύο παράλληλες ευθείες είναι μέγιστη. Τα πρότυπα που βρίσκονται πάνω σε αυτές τις παράλληλες ευθείες γραμμές ονομάζονται διανύσματα υποστήριξης (support vectors).

Όλα αυτά που περιγράφηκαν παραπάνω μπορούν να συνοψιστούν σε μια σχέση, η οποία είναι η εξής:

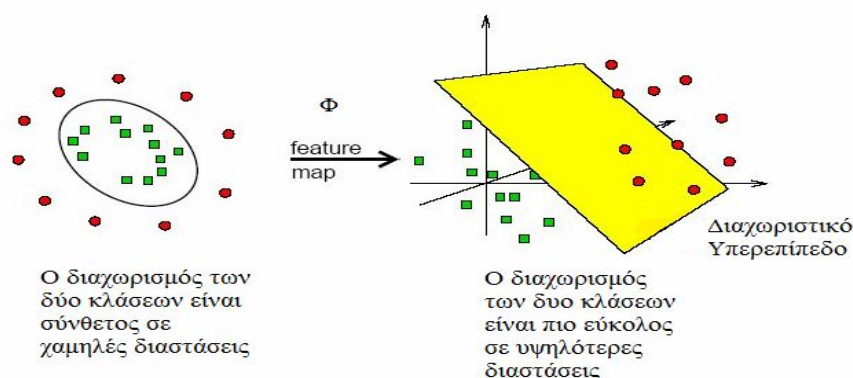
$$y_i ((w^T x_i) + b) > 0, \text{ για κάθε } i \quad (3.2.16)$$

, όπου $y_i \in \{\pm 1\}$, $i=1,2,\dots, m$, η απόσταση (margin) μεταξύ των δύο επιπέδων είναι $2/\|\mathbf{w}\|$, με $\|\mathbf{w}\|$ την Ευκλείδεια νόρμα του διανύσματος \mathbf{w} και το διάνυσμα \mathbf{w} είναι το διάνυσμα των συντελεστών των αγνώστων.

Η εξίσωση της διαχωριστικής ευθείας γραμμής (ή υπερεπίπεδο στην γενική μορφή) είναι ίση με:

$$((\mathbf{w}^T \mathbf{x}) + b) = 0 \quad (3.2.17)$$

Πρακτικά είναι αδύνατη η εύρεση ενός υπερεπίπεδου (ορίου) με γραμμικά χαρακτηριστικά. Τη λύση δίνουν οι μηχανές διανυσμάτων υποστήριξης. Σκοπός τους είναι η απεικόνιση των προτύπων σε ένα χώρο μεγαλύτερων διαστάσεων από τον αρχικό με μια απεικόνιση Φ και η δημιουργία ενός υπερεπίπεδου με μέγιστο εύρος (margin) ανάμεσα στις κλάσεις. Ένα παράδειγμα αυτής της απεικόνισης φαίνεται στην παρακάτω εικόνα.



Εικόνα 3.2.7. Απεικόνιση του διαχωρισμού δύο κλάσεων σε χαμηλές διαστάσεις και υψηλές διαστάσεις.

3.2.2 Στατιστική ταξινόμηση προτύπων

Σε αντίθεση με την Ντετερμινιστική ταξινόμηση προτύπων, που περιγράφηκε παραπάνω, η Στατιστική ή Στοχαστική ταξινόμηση προτύπων χρησιμοποιεί μοντέλα που στηρίζονται στις πιθανότητες, για να ταξινομήσει ένα πρότυπο σε μια κλάση. Μια από τις πιο δημοφιλείς και σημαντικές στατιστικές προσεγγίσεις του προβλήματος της ταξινόμησης προτύπων είναι η θεωρία απόφασης του Bayes.

Βάση της είναι η σύγκριση μεταξύ διαφόρων αποφάσεων ταξινόμησης με βάση τις πιθανότητες και τα κόστη που σχετίζονται με τις αποφάσεις αυτές. Γίνεται θεώρηση ότι το πρόβλημα της απόφασης για την ταξινόμηση των προτύπων ορίζεται με πιθανοθεωρητικούς όρους και είναι γνωστές όλες οι πιθανότητες που θα χρησιμοποιηθούν. Βασικός στόχος είναι μεγιστοποίηση της πιθανότητας της σωστής απόφασης και ελαχιστοποίηση του κόστους που θα έχει αυτή η απόφαση, διαλέγοντας ανάμεσα από τις δυνατές περιπτώσεις, αυτή που ενδέχεται να είναι η σωστή.

Θα περιγράψουμε συνοπτικά κάποιους συμβολισμούς που θα χρησιμοποιηθούν στην περιγραφή της θεωρίας απόφασης του Bayes.

- Τα ενδεχόμενα ή οι κλάσεις(κατηγορίες) θα συμβολίζονται με ω . Έτσι, με ω_j συμβολίζουμε το πρότυπο που η πραγματική του κλάση είναι ω_j .
- Η εκ των προτέρων (priori) πιθανότητα συμβολίζεται με $P(\omega_j)$ και περιγράφει την γνώση που έχουμε από πριν για τις κλάσεις στις οποίες ανήκουν τα πρότυπα. Έτσι, με $P(\omega_j)$ συμβολίζεται η πιθανότητα η πραγματική κλάση ενός προτύπου να είναι η ω_j .
- Με \mathbf{x} συμβολίζουμε μια τυχαία διανυσματική μεταβλητή, που είναι ένα διάνυσμα N διαστάσεων και ανήκει σε ένα N διαστάσεων Ευκλείδειο χώρο(χώρος χαρακτηριστικών). Το διάνυσμα \mathbf{x} είναι ένα πίνακας στήλη που περιέχει τις τιμές των N χαρακτηριστικών ενός προτύπου, που θέλουμε να ταξινομήσουμε.
- Η υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας του διανύσματος \mathbf{x} με δεδομένο ότι η πραγματική κλάση που ανήκει το πρότυπο είναι ω_j συμβολίζεται με $p(\mathbf{x}|\omega_j)$ και εκφράζει την πιθανότητα μια κλάση ω_j να περιλαμβάνει ένα διάνυσμα \mathbf{x} .
- Η εκ των υστέρων (a posteriori) πιθανότητα $P(\omega_j|\mathbf{x})$ εκφράζει την πιθανότητα πραγματοποίησης του ενδεχομένου ω_j με δεδομένο ότι έχει πραγματοποιηθεί(έχει παρατηρηθεί) ένα διάνυσμα \mathbf{x} . Δηλαδή, είναι η πιθανότητα ένα διάνυσμα \mathbf{x} να ανήκει σε μια κατηγορία ω_j .

Στην γενική περίπτωση που θέλουμε να ταξινομήσουμε τα πρότυπα σε c κλάσεις, τότε ο τύπος του Bayes είναι ίσος με:

$$P(\omega_j | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_j) \cdot P(\omega_j)}{p(\mathbf{x})} \quad (3.2.18)$$

, όπου $p(\mathbf{x}) = \sum_{j=1}^c p(\mathbf{x} | \omega_j) P(\omega_j)$

Με λόγια ο παραπάνω τύπος μπορεί να εκφραστεί ως:

Εκ των υστέρων πιθανότητα = $\frac{\text{πιθανοφάνεια } X \text{ εκ των προτέρων πιθανότητα}}{\text{γεγονός}}$

Η $p(\mathbf{x}|\omega_j)$ καλείται συνάρτηση πιθανοφάνειας της ω_j σε σχέση με το \mathbf{x} και εκφράζει ότι, εάν όλες οι υπόλοιπες παράμετροι είναι ίσες, η κλάση ω_j για την οποία η $p(\mathbf{x}|\omega_j)$ έχει μεγάλη τιμή έχει μεγαλύτερη πιθανότητα να είναι η σωστή κλάση. Στην παραπάνω σχέση το γινόμενο της συνάρτησης πιθανοφάνειας με την εκ των προτέρων πιθανότητα συμβάλει στον υπολογισμό της εκ των υστέρων πιθανότητας.

Ο παράγοντας $p(\mathbf{x})$, θεωρείται ως ένας παράγοντας κανονικοποίησης(normalize factor), τέτοιος ώστε το άθροισμα των εκ των υστέρων πιθανοτήτων να ισούται με τη μονάδα, δηλαδή, $\sum_{j=1}^c P(\omega_j|\mathbf{x})=1$.

Επίσης, μπορούμε να ορίσουμε και μια συνάρτηση κόστους $\lambda(a_i|\omega_j)$ που εκφράζει το κόστος, της ενέργειας a_i που θα εκτελέσουμε, όταν ω_j είναι η πραγματική κλάση στην οποία ανήκει το υπό ταξινόμηση πρότυπο.

Έχοντας την $P(\omega_j|\mathbf{x})$, ως η πιθανότητα η πραγματική κλάση ενός προτύπου να είναι η ω_j για δεδομένο \mathbf{x} και την συνάρτηση κόστους, τότε το υπό συνθήκη ρίσκο που σχετίζεται με την ενέργεια a_i είναι ίσο με:

$$R(a_i|x) = \sum_{j=1}^c \lambda(a_i|\omega_j) P(\omega_j|x) \quad (3.2.19)$$

Για κάθε μετρούμενο διάνυσμα \mathbf{x} , το αναμενόμενο κόστος ή ρίσκο είναι ελάχιστο αν επιλεγεί μια ενέργεια a_i , η οποία θα ελαχιστοποιήσει το υπό συνθήκη ρίσκο.

Στην συνέχεια θα θεωρήσουμε ότι κάθε ενέργεια a_i θα εκφράζει την απόφαση ότι η πραγματική κλάση του υπό ταξινόμηση προτύπου είναι η ω_i . Έτσι, εάν εκτελέσουμε την ενέργεια a_i (δηλαδή εάν παρθεί η απόφαση ότι το πρότυπο ανήκει στην κατηγορία ω_i) και η πραγματική κλάση είναι ω_i , τότε η απόφαση είναι σωστή, διαφορετικά εάν η πραγματική κλάση είναι ω_j , τότε η απόφαση είναι λάθος.

Έτσι, αν πάρουμε την απόφαση ότι το πρότυπο ανήκει στην κλάση ω_j , τότε η απόφαση είναι σωστή εάν $i=j$ και λάθος όταν $i \neq j$. Έτσι ορίζεται μια συνάρτηση κόστους, η οποία ονομάζεται συμμετρική ή μηδέν-ένα συνάρτηση κόστους (zero-one loss function) και είναι ίση με:

$$\lambda(a_i | \omega_j) = \begin{cases} 0, & i=j \\ 1, & i \neq j \end{cases} \quad (3.2.20)$$

, όπου $i, j=1, 2, \dots, c$.

Επομένως, όταν η απόφαση είναι σωστή η συνάρτηση δεν αντιστοιχεί κανένα κόστος, ενώ αν είναι λάθος η απόφαση αντιστοιχεί μοναδιαίο κόστος. Σαν συμπέρασμα από τα παραπάνω βγάζουμε ότι όλα τα λάθη είναι ισοδύναμα από πλευράς κόστους.

Το ρίσκο που εξάγεται, χρησιμοποιώντας αυτή την συνάρτηση κόστους ισούται με:

$$\begin{aligned} R(a_i|x) &= \sum_{j=1}^c \lambda(a_i|\omega_j) P(\omega_j|x) \Leftrightarrow \\ R(a_i|x) &= \sum_{i \neq j} P(\omega_j|x) \Leftrightarrow \\ R(a_i|x) &= 1 - P(\omega_i|x) \end{aligned} \quad (3.2.21)$$

, όπου $P(\omega_i|x)$ είναι η υπό συνθήκη (δεσμευμένη) πιθανότητα ότι η ενέργεια a_i είναι σωστή. Δηλαδή, το ρίσκο ισούται $\sum_{i \neq j} P(\omega_j|x)$ που δείχνει την πιθανότητα του ολικού λάθους και μπορεί να γραφτεί ως $1 - P(\omega_i|x)$ σύμφωνα με τους κανόνες λογισμού των πιθανοτήτων, αφού η πιθανότητα $P(\omega_i|x)$ είναι η πιθανότητα της σωστής απόφασης. Το συνολικό ρίσκο είναι ίσο με την μέση πιθανότητα λάθους και θέλουμε ένα κανόνα απόφασης να ελαχιστοποιεί την πιθανότητα λάθους, δηλαδή το ρυθμό λάθους. Όπως, αναφέρθηκε παραπάνω, για να γίνει ελάχιστο το ρίσκο, πρέπει να αποφασιστεί η ενέργεια a_i που θα ελαχιστοποιήσει το υπό συνθήκη ρίσκο. Η ελαχιστοποίηση του υπό συνθήκη ρίσκου γίνεται εάν η πιθανότητα $P(\omega_i|x)$ γίνει μέγιστη. Έτσι, πρέπει να γίνει η επιλογή του i , το οποίο θα μεγιστοποιεί την εκ των υστέρων πιθανότητα $P(\omega_i|x)$.

Έτσι, χρησιμοποιείται ο κανόνας απόφασης του Bayes, ο οποίος ελαχιστοποιεί το ρίσκο. Έτσι ο κανόνας απόφασης ελαχίστου ρυθμού λάθους, όπως ονομάζεται, διατυπώνεται ως εξής:

Αποφάσισε ω_i εάν $P(\omega_i|x) > P(\omega_j|x)$ για κάθε $i \neq j$

Με αυτό τον κανόνα απόφασης αν η εκ των υστέρων πιθανότητα $P(\omega_i|\mathbf{x})$ είναι μέγιστη τότε ο κανόνας απόφασης οδηγεί στην ταξινόμηση του προτύπου με διάνυσμα \mathbf{x} στην κατηγορία ω_i .

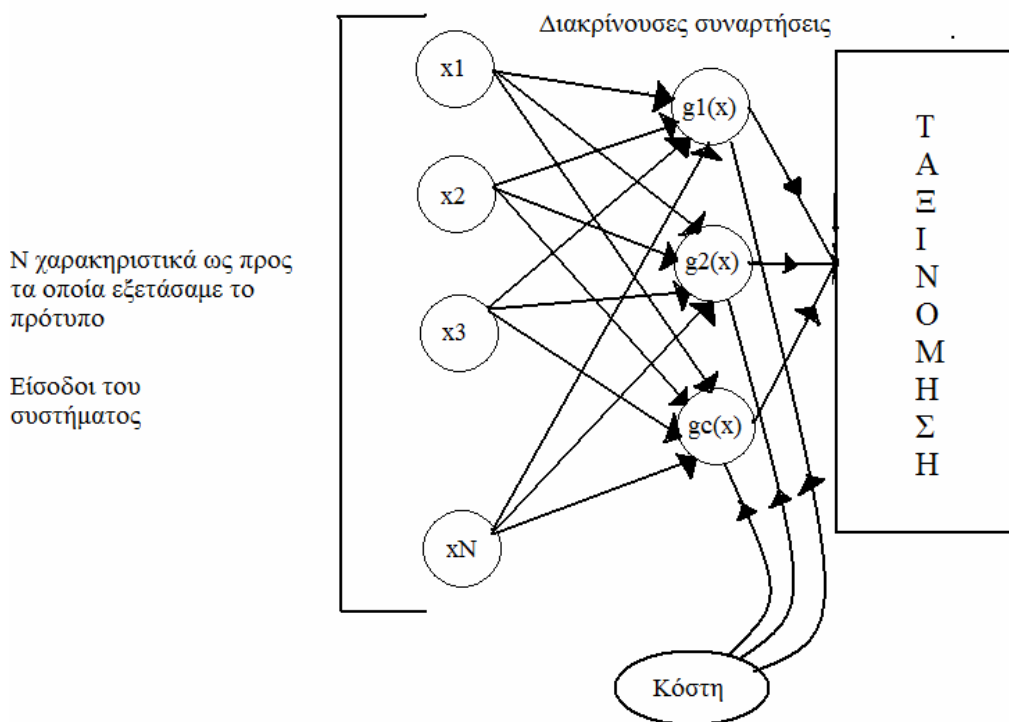
Γενικά, υπάρχουν πολλοί τρόποι για να αναπαραστήσεις ταξινομητές προτύπων. Ένας από τους πιο χρήσιμους τρόπους είναι η χρήση ενός συνόλου από διακρίνουσες συναρτήσεις $g_i(\mathbf{x}), i = 1, \dots, c$.

Και ο κανόνας απόφασης διατυπώνεται ως εξής:

$$\text{Αποφάσισε } \omega_i \text{ εάν } g_i(\mathbf{x}) > g_j(\mathbf{x}) \text{ για κάθε } i \neq j$$

Επομένως, ο ταξινομητής προτύπων αντιμετωπίζεται ως ένα δίκτυο το οποίο υπολογίζει τις τιμές για c διαφορετικές διακρίνουσες συναρτήσεις και επιλέγει την κλάση που αντιστοιχεί στη διακρίνουσα συνάρτηση που έχει τη μεγαλύτερη τιμή. Έτσι λοιπόν, ένας ταξινομητής Bayes μπορεί να αναπαρασταθεί με αυτόν τον τρόπο. Στην περίπτωση που χρησιμοποιούμε την μηδέν-ένα συνάρτηση κόστους, πρέπει να θέσουμε για μια διακρίνουσα συνάρτηση ότι: $g_i(\mathbf{x}) = P(\omega_i|\mathbf{x})$.

Έτσι, αν η τιμή της εκ των υστέρων πιθανότητα $P(\omega_i|\mathbf{x})$ είναι μέγιστη, τότε και η διακρίνουσα συνάρτηση $g_i(\mathbf{x})$ γίνεται μέγιστη. Και χρησιμοποιώντας τον παραπάνω κανόνα απόφασης, ταξινομείται το πρότυπο με διάνυσμα \mathbf{x} στην κλάση ω_i . Η δομή ενός γενικού στατιστικού ταξινομητή φαίνεται στην παρακάτω εικόνα.



Εικόνα 3.2.8. Η δομή ενός γενικού στατιστικού ταξινομητή προτύπων που περιλαμβάνει N εισόδους (χαρακτηριστικά ως προς τα οποία εξετάσαμε το πρότυπο) και c διακρίνουσες συναρτήσεις $g_i(\mathbf{x})$.

3.2.2.1 Gaussian Mixture Models

Στο πρόβλημα ταξινόμησης προτύπων και στο πρόβλημα αναγνώρισης ηχητικής πηγής, οι μίξεις των Γκαουσιανών μοντέλων είναι από τις πιο διαδεδομένες τεχνικές. Πρώτη φορά εισήχθηκε σαν τεχνική από τον Reynolds για να αναπτύξει ένα σύστημα αναγνώρισης ομιλητή, το 1995.

Από το 1995 και έπειτα πολλοί ερευνητές που ασχολήθηκαν με το πρόβλημα της ταξινόμησης προτύπων, χρησιμοποίησαν τις μίξεις Γκαουσιανών μοντέλων, όπου τα αποτελέσματα ήταν ιδιαίτερα ικανοποιητικά. Τις μίξεις Γκαουσιανών μοντέλων, για απλοποίηση, θα τις γράφουμε με τα ακρωνύμια της λέξης Gaussian Mixture Models, ως GMM. Πριν αναλύσουμε τα GMM, είναι εύλογο να γίνει μια εισαγωγή για την λεγόμενη κανονική ή Γκαουσιανή συνάρτηση πυκνότητας πιθανότητας μιας μεταβλητής και πολλών μεταβλητών.

Η κανονική ή Γκαουσιανή συνάρτηση πυκνότητας μιας μεταβλητής δίνεται από την σχέση:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (3.2.22)$$

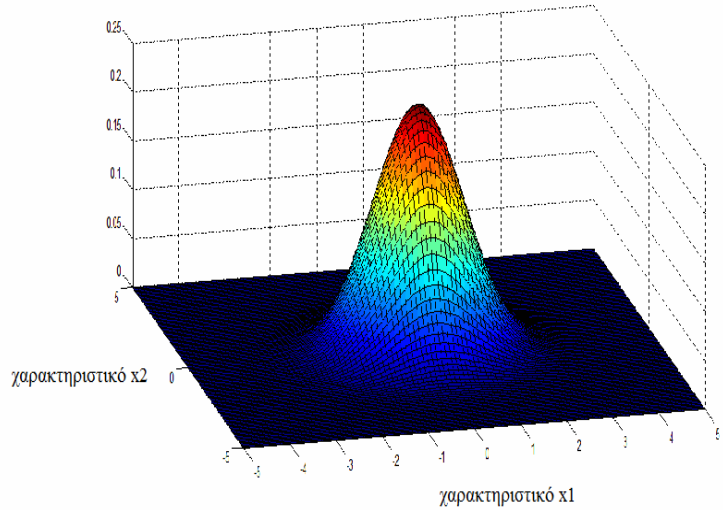
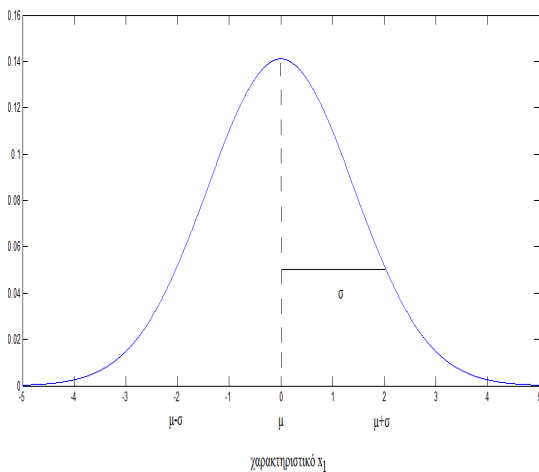
, όπου μ είναι η μέση τιμή των τιμών του x χαρακτηριστικού και σ^2 είναι το μέτρο διακύμανσης ή διασποράς των τιμών του x χαρακτηριστικού σε σχέση με την μέση τιμή μ . Για συντομία μια κανονική ή Γκαουσιανή συνάρτηση πυκνότητας πιθανότητας την συμβολίζουμε με $p(x) = N(\mu, \sigma^2)$, αφού μπορεί να καθοριστεί μια Γκαουσιανή κατανομή πλήρως από δυο παραμέτρους που είναι η μέση τιμή της μ και το μέτρο διασποράς σ^2 .

Η γενική μορφή της Γκαουσιανής κατανομής είναι η κανονική ή Γκαουσιανή συνάρτηση πυκνότητας πιθανότητας πολλών μεταβλητών που δίνεται από την σχέση:

$$p(x) = \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma|}} e^{-\frac{1}{2}[(x-\mu)^T \Sigma^{-1} (x-\mu)]} \quad (3.2.23)$$

, όπου x είναι ένα N διαστάσεων διάνυσμα στήλης, το μ είναι ένα N διαστάσεων διάνυσμα μέσης τιμής, το Σ είναι ένας $N \times N$ πίνακας συνδιασποράς, $|\Sigma|$ είναι η ορίζουσα του πίνακα συνδιασποράς Σ , Σ^{-1} είναι ο αντίστροφος πίνακας του πίνακα Σ και $(x-\mu)^T$ συμβολίζεται ο ανάστροφος πίνακας του $(x-\mu)$. Η ποσότητα $(x-\mu)^T \Sigma^{-1} (x-\mu)$ είναι η απόσταση Mahalanobis, του διανύσματος x από το διάνυσμα μέσης τιμής μ .

Στην παρακάτω εικόνα φαίνεται η κανονική ή Γκαουσιανή κατανομή μιας μεταβλητής και η κανονική ή Γκαουσιανή κατανομή πολλών μεταβλητών.



A)

B)

Εικόνα 3.2.9. A) Γκαουσιανή κατανομή μιας μεταβλητής B) Γκαουσιανή κατανομή δύο μεταβλητών.

Τα GMM αποτελούνται από έναν σταθμισμένο γραμμικό συνδυασμό(μίξη) Γκαουσιανών κατανομών.

Από την Θεωρία Απόφασης Bayes , ισχύει ότι:

$$P(\omega_j | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_j) \cdot P(\omega_j)}{p(\mathbf{x})}$$

, όπου $p(\mathbf{x}) = \sum_{j=1}^c p(\mathbf{x} | \omega_j) P(\omega_j)$

Οι όροι $p(\mathbf{x})$ και $P(\omega_j)$ μπορούν να μην ληφθούν υπόψη, διότι ο παράγοντας $p(\mathbf{x})$ δεν παίζει κάποιο ρόλο στην λήψη απόφασης και ο παράγοντας $P(\omega_j) = 1/c$, για κάθε $j=1,2,\dots,c$, που σημαίνει ότι οι πιθανότητες η πραγματική κλάση ενός προτύπου να είναι $\omega_1, \omega_2, \dots, \omega_c$ είναι ίδιες (ισοπίθανες υποθέσεις). Στην παραπάνω σχέση, η ποσότητα $p(\mathbf{x} | \omega_j)$ είναι η κανονική ή Γκαουσιανή συνάρτηση πυκνότητας πιθανότητας που εκφράζει την πυκνότητα πιθανότητας του \mathbf{x} με δεδομένο ότι η πραγματική κλάση είναι ω_j και $P(\omega_j | \mathbf{x})$ είναι οι εκ των υστέρων πιθανότητα η κλάση να είναι ω_j με δεδομένο ότι έχει παρατηρηθεί ένα διάνυσμα \mathbf{x} .

Μια Γκαουσιανή κατανομή της μορφής:

$$b_i(\mathbf{x}) = p(\mathbf{x} | \omega_j) = \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma_i|}} e^{\left(-\frac{1}{2}[(\mathbf{x}-\mu_i)^T \Sigma_i^{-1}(\mathbf{x}-\mu_i)]\right)} \quad (3.2.24)$$

θα την συμβολίζουμε με $b_i(\mathbf{x})$.

Η συνάρτηση πυκνότητας πιθανότητας Γκαουσιανής μίξης είναι ένα σταθμισμένο άθροισμα από M συνιστώσες Γκαουσιανών κατανομών και δίνεται από την σχέση:

$$\sum_{i=1}^M p_i b_i(\mathbf{x})$$

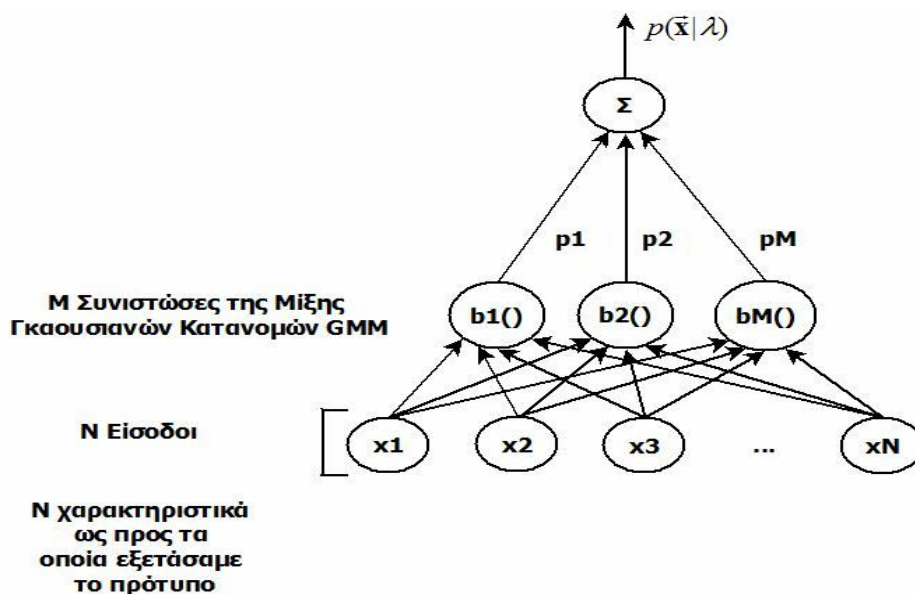
$$p(\mathbf{x}|\lambda)= \quad (3.2.25)$$

,όπου p_i είναι τα βάρη που πολλαπλασιάζονται με τις συναρτήσεις πυκνότητας πιθανότητας και πρέπει να ισχύει γι' αυτά ότι: $\sum_{i=1}^M p_i=1$ και το λ αντιπροσωπεύει την συνολική συνάρτηση πυκνότητας πιθανότητας Γκαουσιανής Μίξης, που καθορίζεται πλήρως από τα διανύσματα μέσης τιμής, από τους πίνακες συνδιασποράς και τα βάρη από όλες τις συνιστώσες Γκαουσιανών κατανομών.

Έτσι, το λ είναι μια συντομογραφία για την οποία ισχύει ότι:

$$\lambda=\{p_i, \mu_i, \Sigma_i\}, \text{ όπου } i=1,2,\dots,M$$

Στην αναγνώριση ομιλητή, κάθε ομιλητής αντιπροσωπεύεται από ένα GMM και αντιστοιχίζεται σε αυτόν ή αυτήν ένα μοντέλο λ . Έτσι, και στην αναγνώριση μουσικών οργάνων, κάθε μουσικό όργανο αντιπροσωπεύεται από ένα GMM και αντιστοιχίζεται σε ένα μουσικό όργανο ένα μοντέλο λ . Στην παρακάτω εικόνα φαίνεται πως σχηματίζεται μια συνάρτηση πυκνότητας πιθανότητας Γκαουσιανής μίξης με M συνιστώσες, με δοσμένο ένα διάνυσμα \mathbf{x} .



Εικόνα 3.2.10. Πυκνότητα πιθανότητας Γκαουσιανής μίξης με M συνιστώσες.

Με δοσμένα κάποια δεδομένα εκπαίδευσης, ο στόχος της εκπαίδευσης είναι να γίνει εκτίμηση των παραμέτρων του GMM, δηλαδή του μοντέλου λ , το οποίο θα ταιριάζει καλύτερα στην κατανομή των διανυσμάτων των χαρακτηριστικών, που έχουν χρησιμοποιηθεί στην εκπαίδευση. Υπάρχει πληθώρα τεχνικών για την εκτίμηση των παραμέτρων του GMM και μια από τις πιο δημοφιλείς είναι η μέθοδος της 'Εκτίμησης της Μέγιστης Πιθανοφάνειας' (Maximum Likelihood Estimation) ή συντομογραφικά ML εκτίμηση. Ο στόχος της εκτίμησης είναι να βρει τις παραμέτρους του μοντέλου, οι οποίες θα μεγιστοποιούν την πιθανοφάνεια του GMM, για δοσμένα δεδομένα εκπαίδευσης. Δηλαδή:

Έστω ότι δίνονται K πρότυπα (διανύσματα χαρακτηριστικών) για την εκπαίδευση, $X=\{x_1, x_2, \dots, x_K\}$. Η πιθανοφάνεια του GMM δίνεται από την σχέση:

$$P(\mathbf{X}|\lambda) = \prod_{k=1}^K p(\mathbf{x}_k|\lambda) \quad (3.2.26)$$

, έχοντας υποθέσει ότι τα δεδομένα εκπαίδευσης είναι ανεξάρτητα μεταξύ τους. Αυτή η συνάρτηση είναι μια μη γραμμική συνάρτηση ως προς τις παραμέτρους λ , με συνέπεια να μην είναι πιθανό να γίνει η μεγιστοποίηση της άμεσα και άρα και ο υπολογισμός των παραμέτρων. Παρ' όλα αυτά η Εκτίμηση της Μέγιστης Πιθανοφάνειας για τον υπολογισμό των παραμέτρων μπορεί να γίνει χρησιμοποιώντας τον αλγόριθμο της 'Μεγιστοποίησης Προσδοκίας' (Expectation-Maximization, EM algorithm). Η βασική ιδέα του EM αλγόριθμου είναι να ξεκινήσει με ένα αρχικό μοντέλο λ για να υπολογίσει (εκτιμήσει) ένα καινούριο μοντέλο λ' , τέτοιο ώστε να ισχύει ότι:

$$P(\mathbf{X}|\lambda') \geq P(\mathbf{X}|\lambda)$$

Αυτό το μοντέλο λ' γίνεται το αρχικό μοντέλο στην επανάληψη της διαδικασίας και αυτή η διαδικασία επαναλαμβάνεται μέχρι κάποιο κατώτατο όριο σύγκλισης επιτευχθεί. Παρουσιάζονται, παρακάτω οι τύποι ανανέωσης των παραμέτρων σε κάθε επανάληψη του EM αλγορίθμου, οι οποίοι εγγυώνται μια μονότονη αύξηση της τιμής της πιθανοφάνειας του μοντέλου. Έτσι, τα βάρη της μίξης ανανεώνονται, σύμφωνα με την σχέση:

$$p_i' = \frac{1}{K} \sum_{k=1}^K p(i | \mathbf{x}_k, \lambda) \quad (3.2.27)$$

Έτσι, το διάνυσμα μέσης τιμής ανανεώνεται, σύμφωνα με την σχέση:

$$\mu_i' = \frac{\sum_{k=1}^K p(i | \mathbf{x}_k, \lambda) \cdot \mathbf{x}_k}{\sum_{k=1}^K p(i | \mathbf{x}_k, \lambda)} \quad (3.2.28)$$

Έτσι, το μέτρα διασποράς ανανεώνονται, σύμφωνα με την σχέση:

$$\sigma_i^{2'} = \frac{\sum_{k=1}^K p(i | \mathbf{x}_k, \lambda) \cdot \mathbf{x}_k^2}{\sum_{k=1}^K p(i | \mathbf{x}_k, \lambda)} - \mu_i'^2 \quad (3.2.29)$$

, όπου σ_i^2 , \mathbf{x}_k , μ_i είναι κάποια αυθαίρετα στοιχεία των διανυσμάτων σ_i^2 , \mathbf{x}_k , μ_i , αντίστοιχα. Η «εκ των υστέρων» (a posteriori) πιθανότητα της i κατανομής (ακουστικής κλάσεως) δίνεται από την σχέση:

$$p(i | \mathbf{x}, \lambda) = \frac{p_i b_i(\mathbf{x})}{\sum_{k=1}^M p_k b_k(\mathbf{x})} \quad (3.2.30)$$

Δύο βασικοί παράγοντες παίζουν ρόλο στην εκπαίδευση του GMM, οι οποίες είναι α) η τάξη ή ο αριθμός των συνιστωσών της μίξης και β) οι παράμετροι του μοντέλου που θα εκτιμηθούν(υπολογιστούν) από την εκτέλεση του EM αλγορίθμου.

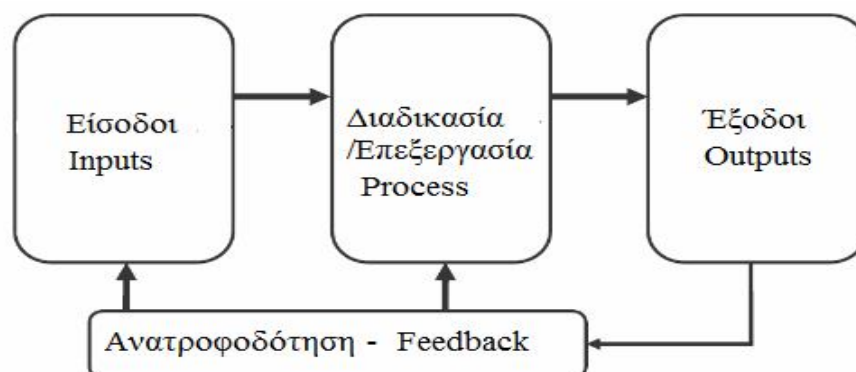
ΚΕΦΑΛΑΙΟ ΤΕΤΑΡΤΟ

ΣΥΣΤΗΜΑΤΑ ΑΝΑΓΝΩΡΙΣΗΣ

4.1 Συστήματα αναγνώρισης ηχητικής πηγής

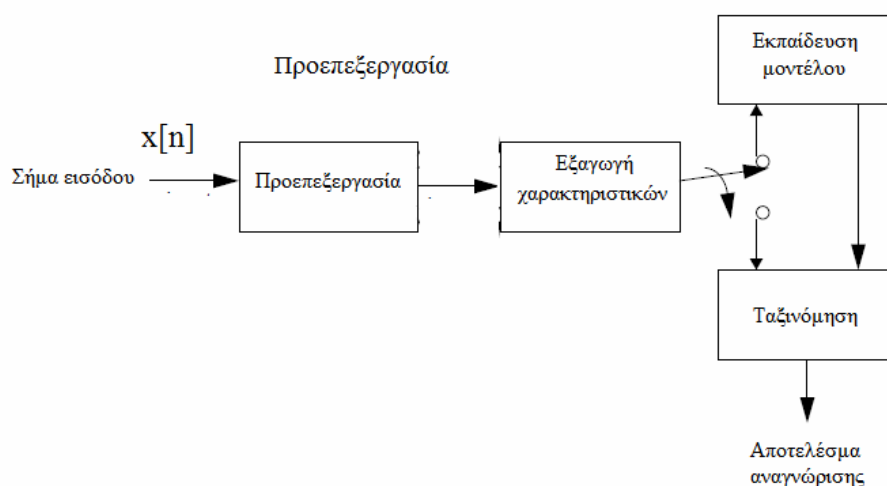
Ένας τρόπος μελέτης ενός τεχνολογικού συστήματος είναι η εξέταση των εισόδων του(inputs), των διαδικασιών(processes) και των εξόδων του(outputs). Έτσι, διαπιστώνεται ότι κάθε σύστημα περιλαμβάνει εισόδους, διαδικασίες και εξόδους. Το μοντέλο που περιγράφει τα συστήματα με αυτό τον τρόπο ονομάζεται καθολικό υπόδειγμα συστήματος (universal system model). Έτσι, το σύστημα αυτό χρειάζεται κάποιο ‘ερέθισμα’, δηλαδή κάποια είσοδο π.χ. πληροφορίες, υλικά, ενέργεια, οικονομικοί πόροι, ανθρώπινη προσπάθεια κ.τ.λ. ώστε να γίνει ροή της εισόδου στο στάδιο της διαδικασίας ή επεξεργασίας. Σε αυτό το στάδιο, ανάλογα με την είσοδο ή τις εισόδους του συστήματος, γίνεται η επεξεργασία τους και κατ’ επέκταση εκτελούνται κάποιες εντολές. Το τρίτο στάδιο είναι η έξοδος ή οι εξοδοί του συστήματος, όπου εξάγονται ή εκτελούνται τα αποτελέσματα των διαδικασιών. Συνήθως σε ένα καθολικό υπόδειγμα συστήματος περιλαμβάνεται και ένα άλλο ακόμα στάδιο, το οποίο είναι αρκετά σημαντικό και το οποίο είναι το στάδιο της ανάδρασης ή ανατροφοδότησης(feedback). Σε αυτό το στάδιο γίνεται ανατροφοδότηση των εισόδων ή των διαδικασιών από την έξοδο(ή εξόδους) του συστήματος. Έτσι, επιτυγχάνεται κάποια κυκλική ροή. Βέβαια, αυτό έχει μια επίπτωση στο συνολικό σύστημα.

Ένα καθολικό υπόδειγμα συστήματος μαζί με το στάδιο της ανατροφοδότησης φαίνεται στην εικόνα 4.1.1.



Εικόνα 4.1.1. Καθολικό υπόδειγμα συστήματος(Universal System Model) μαζί με το στάδιο της ανατροφοδότησης.

Με τον παραπάνω τρόπο μπορούμε να περιγράψουμε με απλό τρόπο ένα σύστημα αναγνώρισης ηχητικής πηγής. Αυτό το σύστημα περιλαμβάνει μια είσοδο(π.χ. μια νότα από ένα μουσικό όργανο), επιτελεί μια διαδικασία(π.χ. μια σύγκριση των χαρακτηριστικών του σήματος εισόδου με τα αποθηκευμένα αρχεία που περιέχονται στο σύστημα) και η έξοδος, όπου επιτελείται η ταξινόμηση του μουσικού οργάνου στην κατηγορία που έχει εκτιμηθεί ότι μπορεί να ανήκει.



Εικόνα 4.1.2. Μπλοκ διάγραμμα ενός συστήματος αναγνώρισης ηχητικής πηγής.

Στην παραπάνω εικόνα φαίνεται ένα απλοποιημένο μπλοκ διάγραμμα ενός συστήματος αναγνώρισης ηχητικής πηγής. Όπως, φαίνεται και από την παραπάνω εικόνα ένα σύστημα αναγνώρισης ηχητικής πηγής αποτελείται από την βαθμίδα προεπεξεργασίας του σήματος(pre-processing), από την βαθμίδα της εξαγωγής χαρακτηριστικών(feature extraction), την βαθμίδα του καθορισμού του μοντέλου που περιγράφει αυτά τα χαρακτηριστικά ή την βαθμίδα δημιουργίας της μνήμης του συστήματος και τέλος την βαθμίδα της ταξινόμησης(classifier), όπου σε αυτήν προσδιορίζεται η ταυτότητα των ακολουθιών(σημάτων εισόδου) χρησιμοποιώντας την μνήμη του συστήματος. Ανάλογα με τις τεχνικές που χρησιμοποιούνται στις επιμέρους βαθμίδες, τα συστήματα αναγνώρισης ηχητικής πηγής ταξινομούνται ανάλογα:

- την τεχνική που χρησιμοποιείται στην βαθμίδα εξαγωγής χαρακτηριστικών (φασματική ανάλυση, cepstral μετασχηματισμός, μετασχηματισμός Fourier κ.λ.π.)
- τη βασική μονάδα αναγνώρισης / κωδικοποίησης που χρησιμοποιείται από το σύστημα
 - το τρόπο ταξινόμησης από την βαθμίδα ταξινόμησης των εισαγόμενων προτύπων (για παράδειγμα, συστήματα σύγκρισης προτύπων (Template Matching), πιθανοτικά μοντέλα(π.χ. Κρυμμένα Μοντέλα Markov– HMM, Μίξη Γκαουσιανών Μοντέλων-GMM κ.τ.λ), συντακτικά συστήματα και συστήματα δικτύων(π.χ. Νευρωνικά Δίκτυα-Neural Networks).
 - το τρόπο δημιουργίας της μνήμης του συστήματος ή την τεχνική του προσδιορισμού του μοντέλου που περιγράφει τα χαρακτηριστικά του εισερχόμενου προτύπου, από την βαθμίδα της εκπαίδευσης μοντέλου.

Πολλές φορές κατηγοριοποιούμε ένα σύστημα ως:

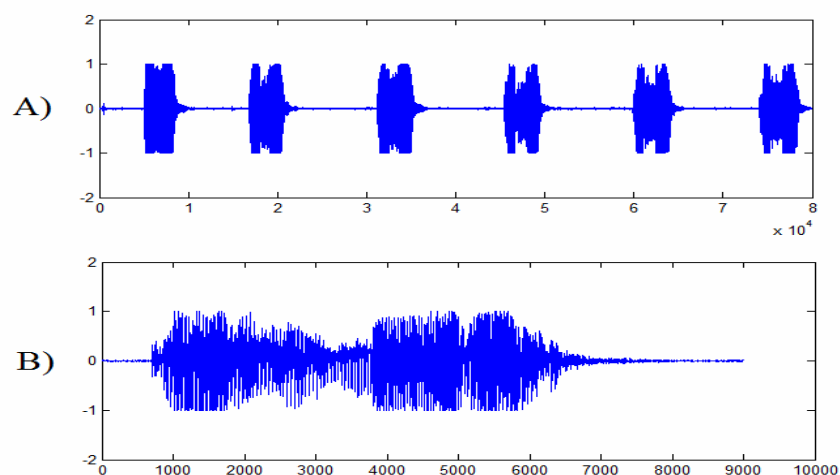
- **Ανεξάρτητο από την ηχητική πηγή:** το σύστημα έχει την ικανότητα να αναγνωρίσει πρότυπα ηχητικής πηγής από ένα μεγάλο αριθμό ηχητικών πηγών.
- **Εξαρτώμενο από την ηχητική πηγή:** μπορεί να αναγνωρίσει ηχητικές πηγές, οι οποίες χρησιμοποιήθηκαν και ως δεδομένα στην εκπαίδευση.
- **Προσαρμοζόμενο στην ηχητική πηγή:** μίγμα των δύο παραπάνω και έχει υλοποιηθεί έτσι ώστε να προσαρμόζει την λειτουργία του σε κάθε νέα ηχητική πηγή που καλείται να αναγνωρίσει, βάση γενικών μοντέλων περιγραφής των χαρακτηριστικών της ηχητικής πηγής.

Επίσης, ένα σύστημα διαφέρει από ένα άλλο ανάλογα με τον τύπο του σήματος που καλείται να αναγνωρίσει. Έτσι, διακρίνουμε και ένα σύστημα ως:

- **Μη συνεχές σύστημα αναγνώρισης:** όταν καλείται να αναγνωρίσει μεμονωμένα σήματα από την ηχητική πηγή, δηλαδή απαιτεί το σύστημα να αναγνωρίσει σήματα με μικρές παύσεις ανάμεσα τους. Για παράδειγμα όταν έχει εκπαιδευτεί το σύστημα να αναγνωρίζει μεμονωμένες λέξεις που προφέρει ο χρήστης. Σε αυτή την περίπτωση το σύστημα καλείται σύστημα αναγνώρισης διακριτής ή μη συνεχής ομιλίας.

- **Συνεχές σύστημα αναγνώρισης:** όταν το σύστημα έχει υλοποιηθεί, ώστε να αναγνωρίζει ένα συνεχές σήμα που παράγεται από την πηγή, δηλαδή απαιτεί το σύστημα να αναγνωρίσει σήματα που το ένα διαδέχεται το άλλο χωρίς παύσεις ανάμεσα τους. Για παράδειγμα όταν το σύστημα έχει υλοποιηθεί να αναγνωρίζει μια ή περισσότερες φράσεις καθώς ομιλεί ένας άνθρωπος. Σε αυτή την περίπτωση το σύστημα ονομάζεται σύστημα αναγνώρισης συνεχούς ομιλίας.

Για να μην προκληθεί σύγχυση, ένα σύστημα χαρακτηρίζεται συνεχές ή μη συνεχές, χωρίς να σχετίζεται αν το σήμα που παράγει η πηγή είναι αναλογικό ή ψηφιακό, αντίστοιχα. Σχετίζεται αν το σύστημα καλείται να αναγνωρίσει μεμονωμένους ή όχι ήχους. Αυτό φαίνεται καλύτερα στην παρακάτω εικόνα.



Εικόνα 4.1.3. Α) Σήμα διακριτής (ή μη συνεχής) ομιλίας, της λέξης 'Ένα', το οποίο επαναλαμβάνεται 6 φορές και Β) Σήμα συνεχούς ομιλίας, της λέξης 'Καλημέρα'.

Τα συστήματα αναγνώρισης ηχητική πηγής μπορούν να διακριθούν σε πολλές κατηγορίες, όμως θα αναφέρουμε μόνο τρεις από αυτές:

- Αναγνώριση περιβαλλοντικών ήχων

- Αναγνώριση ομιλητή
- Αναγνώριση μουσικών οργάνων

4.1.1 Αναγνώριση περιβαλλοντικών ήχων

Στην αναγνώριση περιβαλλοντικών ήχων περιλαμβάνονται και τα ηχητικά εφέ. Είναι ένα πολύ ευρύ πεδίο και για λόγους συντομίας θα αναφερθούν δύο παραδείγματα.

Ένα σύστημα αναγνώρισης ήταν του Klassner και συγκεκριμένα το Sound Understanding Testbed (SUT), που σκοπό είχε να αναγνωρίζει συγκεκριμένους ήχους από το σπίτι και το περιβάλλον[Klassner96].

Το σύστημα SUT είχε μια βιβλιοθήκη με 40 ήχους, από τους οποίους τα μοντέλα παράγονταν με το χέρι. Το υλικό της εξέτασης(test) κατασκευάστηκε τοποθετώντας 4 ανεξάρτητα ηχητικά παραδείγματα από την βιβλιοθήκη(ηχογράφηση πέντε δευτερολέπτων) και το σύστημα έπρεπε να αναγνωρίσει ποιο γεγονός πραγματοποιήθηκε και πότε.

Ανάλογα με το αν τα μοντέλα που είχαν παραχθεί, χρειάζονταν ως αναφορές ή αυτά ήταν που χρησιμοποιούντουσαν κατευθείαν στην ηχογράφηση, οι αποδόσεις του συστήματος ήταν 59% και 61% , αντίστοιχα.

Το άλλο παράδειγμα έχει να κάνει με τον Zhang and Kuo. Οι Zhang και Kuo δοκίμασαν ένα σχέδιο Query by Example (QBE)(γλώσσα προγραμματισμού, ερωτοαπαντήσεων για σχεσιακές βάσεις δεδομένων(relational database)) για την αναγνώριση ηχητικών εφέ[Zhang00]. Χρησιμοποιήθηκαν κατηγορίες ηχητικών εφέ, όπως χειροκρότημα, βήματα, έκρηξη, βροχή κ.τ.λ.. Με 18 κατηγορίες ηχητικών εφέ, αναφέρθηκε μια απόδοση 86%.

4.1.2 Αναγνώριση ομιλητή

Η αναγνώριση ομιλητή είναι το πρόβλημα αναγνώρισης ηχητικής πηγής που έχει μελετηθεί περισσότερο[Martin99]. Ο στόχος της αναγνώρισης ομιλητή διαφέρει από την αναγνώριση μουσικών οργάνων σε ποικίλα σημεία. Ένα από αυτά τα σημεία διαφοράς είναι, ότι ο αριθμός των μουσικών οργάνων είναι αρκετά περιορισμένος σε αντίθεση με τις δεσεκατομμύρια υπάρχουσες ανθρώπινες φωνές. Από την άλλη πλευρά, το εύρος της θεμελιώδης συχνότητας της ανθρώπινης φωνής είναι σχετικά περιορισμένο σε σχέση με το παραγόμενο εύρος της θεμελιώδης συχνότητας από τα περισσότερα μουσικά όργανα. Γενικά, μια φωνή μπορεί να παράγει μεγαλύτερη ποικιλία ήχων από ένα μουσικό όργανο[Handel95].

Η ιδέα της αναγνώρισης, στην συγκεκριμένη κατηγορία είναι ο προσδιορισμός των έμφυτων διαφορών στα αρθρωτικά όργανα και στον τρόπο της ομιλίας [Mammone96]. Τα αρθρωτικά όργανα που αναφέρονται παραπάνω είναι η δομή της φωνητικής οδούς, το μέγεθος της ρινικής κοιλότητας και τα χαρακτηριστικά των φωνητικών χορδών ή της επιγλωττίδας. Στην πράξη στα υλοποιημένα συστήματα χρησιμοποιούνται μόνο χαρακτηριστικά που σχετίζονται με τα χαρακτηριστικά και την δομή της φωνητικής οδούς.

Πρόσφατες μέθοδοι, στόχο έχουν την σύγκριση των χαρακτηριστικών μεταξύ όμοιων φωνητικών ήχων(phonetic sounds) μέσα στα σύνολα της

εκπαίδευσης(training) και της εξέτασης(test). Όμως, η τελευταία μέθοδος, η οποία χρησιμοποιείται περισσότερο σήμερα, περιλαμβάνει την αναζήτηση των συγκεντρώσεων των ακουστικών χαρακτηριστικών στο στάδιο της εκπαίδευσης και ταξινόμησης ή αναγνώρισης, χωρίς επόπτη(unsupervised clustering of acoustic features). Το στατιστικό μοντέλο που χρησιμοποιείται, επίσης πάρα πολύ στις εφαρμογές της αναγνώρισης ομιλητή είναι η Μίξη Γκαουσιανών Μοντέλων ή Gaussian Mixture Models(GMM). Δυναμικά πεδία για περαιτέρω έρευνα στην αναγνώριση ομιλητή είναι η χρήση της θεμελιώδους συχνότητας(fundamental frequency) και του ρυθμού της ομιλίας(speech rhythm).

Σύμφωνα με τους Murthy και Alonso [Murthy99, Alonso00], ένα από τα κυριότερα προβλήματα της αναγνώρισης ομιλητή είναι ότι η απόδοση του συστήματος μειώνεται όταν οι ακουστικές συνθήκες ποικίλουν από αυτές που επικρατούν στην διάρκεια της εξέτασης(δοκιμής, testing). Όπως επίσης, η απόδοση μειώνεται όταν άλλοι ήχοι συμβάλλουν με το σήμα της ομιλίας ή όταν το μέγεθος του πληθυσμού αυξάνει[Reynolds95]. Για παράδειγμα αναφέρεται ότι με 10 ομιλητές η απόδοση του συστήματος ήταν 94%, ενώ με 113 ομιλητές, η απόδοση μειώθηκε στα 83%. Παρ' όλα αυτά όσο αναφορά το πεδίο των συστημάτων αναγνώρισης ηχητικής πηγής, τα συστήματα αναγνώρισης ομιλητή είναι πιο κοντά στην πρακτική εφαρμογή και οι μέθοδοι που χρησιμοποιούνται είναι οι πιο αναπτυγμένες.

4.1.3 Αναγνώριση μουσικών οργάνων

Σε αυτή την κατηγορία, η αναγνώριση μουσικών οργάνων μπορεί να διαιρεθεί σε τέσσερις επιμέρους κατηγορίες, ως εξής:

- Παραγωγή χώρων χροιάς(Generation of timbre spaces)
- Αναγνώριση μεμονωμένων τόνων (Recognition of single tones)
- Αναγνώριση μονοφωνικών φράσεων (Recognition of monophonic phrases)
- Πολυφωνική αναγνώριση(Polyphonic Recognition)

Για αυτές τις επιμέρους κατηγορίες θα αναφερθούμε συνοπτικά στην συνέχεια.

Παραγωγή των χώρων χροιάς(Generation of timbre spaces)

Σε αυτή την κατηγορία, έχουν γίνει προσπάθειες για παραγωγή των χώρων χροιάς χρησιμοποιώντας τεχνικές που προσπαθούν να μοντελοποιήσουν το ανθρώπινο σύστημα ακοής. Στόχος ήταν η συγκέντρωση σε ομάδες, των χαρακτηριστικών των μουσικών ήχων που αντιληπτικά ήταν όμοιοι, δηλαδή ο σκοπός ήταν να ομαδοποιήσουν ήχους που είχαν ορισμένα κοινά χαρακτηριστικά μεταξύ τους. Αυτές οι προσπάθειες μπορούν να συσχετιστούν με την αναγνώριση μουσικών οργάνων και παρ' όλο τη δυσκολία να δώσουν κάποια καλή απόδοση στην ταξινόμηση των οργάνων, δίδουν τι μπορεί να επιτευχθεί με αυτή την προσέγγιση.

Πολλές από αυτές τις μελέτες συνηθίζουν να χρησιμοποιούν ενός είδος ακουστικού μοντέλου ως είσοδο σε ένα αυτό-οργανωμένο χάρτη του Kohonen (Kohonen Self-Organizing Map, SOM).

Αναγνώριση μεμονωμένων τόνων (Recognition of single tones)

Στα συστήματα αυτής της κατηγορίας χρησιμοποιούνται μεμονωμένες νότες ενός μουσικού οργάνου ως υλικό εξέτασης, για ποικίλο αριθμό μουσικών οργάνων και τονικό υψών.

Ένα παράδειγμα αυτής της κατηγορίας είναι των Kaminskyj και Materka, που χρησιμοποίησαν χαρακτηριστικά που προέρχονταν από την περιβάλλουσα της RMS(Root-Mean-Square) ενέργειας και κάποια από αυτά επιλέχθηκαν μέσω της Κύριας Ανάλυσης Συνιστωσών(Principal Component Analysis, PCA) και ο τύπος ταξινομητή που χρησιμοποιήθηκε ήταν τύπος Νευρωνικού Δικτύου(Neural Network classifier) και ένας ταξινομητής k-κοντινότερου γείτονα(k-nearest neighbor), που σκοπό είχε την ταξινόμηση τόνων(πέρα μιας ζώνης οκτάβας) της κιθάρας, του πιάνου, της μαρίμπα και του ακορντεόν [Kaminskyj95]. Και οι δύο ταξινομητές που χρησιμοποιήθηκαν, πραγματοποίησαν καλή απόδοση, γύρω στο 98%. Το σύστημα είχε κάποιους περιορισμούς, όπως τα όργανα που χρησιμοποιήθηκαν ήταν τελείως διαφορετικά μεταξύ τους, υπήρχε μόνο ένα παράδειγμα για κάθε όργανο, το εύρος των νοτών ήταν μικρό και τόσο τα δεδομένα της εκπαίδευσης όσο και τα δεδομένα της εξέτασης ήταν από την ίδια session ηχογράφησης. Άρα, δεν μπορούν να εξαχθούν συμπεράσματα για ποια θα ήταν η απόδοση του συστήματος, αν γινόταν χρήση μουσικών οργάνων της ίδια οικογένειας, μεγαλύτερου εύρους νοτών κ.τ.λ.

Ένα άλλο παράδειγμα συστήματος αναγνώρισης μεμονωμένων νοτών, είναι του Kostek. Ο Kostek υπολόγισε διάφορα χαρακτηριστικά που σχετίζονται με το φασματικό περιεχόμενο και τα χαρακτηριστικά της ατάκας. Τα χαρακτηριστικά αυτά τα εξήγαγε από χρωματικές κλίμακες, οι οποίες είναι παιγμένες με διαφορετικούς τρόπους[Kostek99]. Σαν ταξινομητής χρησιμοποιήθηκε ένα Εμπρόσθιας-τροφοδότησης νευρωνικό δίκτυο δύο επιπέδων(Two Layer feed-forward Neural Network classifier). Ο Kostek αναφέρει εξαιρετικά ποσοστά αναγνώρισης για τέσσερα μουσικά όργανα και συγκεκριμένα για το μπάσο τρομπόνι, τρομπόνι, Αγγλικό κόρνο και κόντρα-φαγγότο.

Τα παραπάνω παραδείγματα ήταν συστήματα που χρησιμοποιούσαν ένα παράδειγμα για κάθε όργανο. Ένα παράδειγμα χρησιμοποίησης πολλών παραδειγμάτων για κάθε όργανο είναι του Martin. Ο Martin χρησιμοποίησε ένα ευρύ σύνολο χαρακτηριστικών για να περιγράψει τις ακουστικές ιδιότητες[Martin99]. Ένας ενισχυμένος Μπεϋζιανός ταξινομητής με ταξινομική ιεραρχία χρησιμοποιήθηκε στην βαθμίδα ταξινόμησης. Για την ταξινόμηση 137 νοτών από 14 μουσικά όργανα, από την συλλογή McGill και με στόχο 27 κλάσεις, η καλύτερη απόδοση που αναφέρθηκε ήταν 39% από ανεξάρτητα όργανα και 76% για όργανα της ίδιας οικογένειας. Κατά συνέπεια, όσο πιο απαιτητική είναι αξιολόγηση του υλικού που χρησιμοποιείται, τα ποσοστά αναγνώρισης είναι σημαντικά χαμηλά, σε σχέση με συστήματα που περιγράφηκαν παραπάνω.

Αναγνώριση μονοφωνικών φράσεων (Recognition of monophonic phrases)

Όπως και στην περίπτωση της αναγνώρισης μεμονωμένων τόνων έτσι και σε αυτή την περίπτωση, υπάρχουν συστήματα που χρησιμοποιούν ένα παράδειγμα για κάθε

όργανο και άλλα συστήματα, που χρησιμοποιούν περισσότερα του ενός παραδείγματα. Ένα σύστημα που κάνει χρήση ενός παραδείγματος για κάθε όργανο είναι των Dubnov και Rodet.

Οι Dubnov και Rodet χρησιμοποίησαν cepstral and delta cepstral συντελεστές, οι οποίοι υπολογίστηκαν από 18 μουσικά κομμάτια από ένα μεγάλο πλήθος οργάνων[Dubnov98]. Τα διανύσματα των χαρακτηριστικών στην αρχή κβαντοποιήθηκαν (vector quantized) και ύστερα τροφοδοτήθηκαν σε ένα στατιστικό ταξινομητή αναζήτησης συγκεντρώσεων(statistical clustering algorithm). Τα αποτελέσματα της συγκέντρωσης(clustering) ήταν αρκετά υποσχόμενα, λόγω ότι τα χαρακτηριστικά από διαφορετικά όργανα, συγκεντρώθηκαν σε διαφορετικές ομάδες συγκεντρώσεων(clusters), αλλά οι Dubnov και Rodet δεν αναφέρουν κάποια ποσοστά αναγνώρισης.

Ένα άλλο σύστημα, όμως αυτή τη φορά με χρήση πολλαπλών παραδειγμάτων για κάθε όργανο είναι της Brown, που χρησιμοποίησε τεχνικές από την αναγνώριση ομιλητή για την ταξινόμηση μεταξύ όμποε, σαξοφώνου, φλάουτου και κλαρινέτου[Brown01].

Χρησιμοποίησε ανεξάρτητα δεδομένα και για την διαδικασία της εκπαίδευσης αλλά και για την διαδικασία της εξέτασης. Τα δεδομένα αυτά είναι ποικίλης ποιότητας και έχουν παρθεί από ηχογραφήσεις διαφημίσεων. Χρησιμοποίησε μια queffrency παράγωγο των συντελεστών σταθερού-Q(constant-Q coefficients). Παρατήρησε μια ακρίβεια του 84%, η οποία είναι συγκρίσιμη με την ακρίβεια των ανθρώπων σε μια εξέταση ακρόασης που πραγματοποίησε από ένα υποσύνολο των δειγμάτων των δεδομένων. Άλλα, επιτυχημένα χαρακτηριστικά ήταν οι συντελεστές cepstral και συντελεστές αυτοσυσχέτισης(autocorrelation coefficients). Σε παλαιότερη έρευνα, το σύστημα της ταξινόμησης δειγμάτων όμποε και σαξοφώνου έχει ακρίβεια που άγγιζε το 94%[Brown99].

Πολυφωνική αναγνώριση(Polyphonic Recognition)

Η εφαρμογή της υλοποίησης συστήματος αναγνώρισης πολυφωνικού ήχου είναι ένα δύσκολο αλλά και ενδιαφέρον πρόβλημα στο πεδίο της αναγνώρισης μουσικών οργάνων. Στόχος του συστήματος είναι η αναγνώριση των ξεχωριστών οργάνων που είναι ηχογραφημένα σε ένα κομμάτι. Ένα παράδειγμα τέτοιου συστήματος είναι από την δουλειά του Kashino, που στόχος είναι η αντιστοίχιση μουσικού σήματος σε νότες(music transcription). Στην διαδικασία της αντιστοίχισης μουσικού σήματος σε νότες περιέχεται και η αναγνώριση μουσικών οργάνων. Σε αυτό το σύστημα του Kashino, ζητούμενο ήταν να αντιστοιχίσει τυχαία σήματα από κλαρινέτο, φλάουτο, πιάνο, τρομπέτα και βιολί σε νότες και αναφέρθηκε πολύ καλή απόδοση[Kashino95]. Μετά από αυτό το εγχείρημα του Kashino, ο ίδιος και ο Murase υλοποίησαν ένα σύστημα στόχος του οποίου ήταν η αντιστοίχιση τριών μελωδιών από διάφορα όργανα σε νότες[Kashino98, Kashino99]. Τα όργανα που χρησιμοποιήθηκαν ήταν βιολί, φλάουτο και πιάνο με απόδοση που φτάνει τα 88.5%.

ΚΕΦΑΛΑΙΟ ΠΕΜΠΤΟ

ΑΥΤΟΜΑΤΟ ΣΥΣΤΗΜΑ ΑΝΑΓΝΩΡΙΣΗΣ ΜΟΥΣΙΚΩΝ ΟΡΓΑΝΩΝ

5.1 Γενική λειτουργία του συστήματος

Στα προηγούμενα κεφάλαια έγινε αναφορά στις τεχνικές της ψηφιακής επεξεργασίας σήματος, στις τεχνικές εξαγωγής χαρακτηριστικών αλλά και στις μεθόδους που χρησιμοποιούνται για την εκπαίδευση αλλά και την τελική αναγνώριση, που είναι και ο σκοπός του συστήματος.

Στην παρούσα πτυχιακή εργασία, σκοπός είναι η υλοποίηση ενός αυτόματου συστήματος αναγνώρισης μουσικών οργάνων. Δηλαδή, βασική επιδίωξη του συστήματος είναι η αναγνώριση διάφορων μουσικών οργάνων, όταν ο ήχος καταγράφεται από ένα μικρόφωνο που είναι συνδεδεμένο στον ηλεκτρονικό υπολογιστή(H/Y). Η διαδικασία θα αναφερθεί συνοπτικά στην συγκεκριμένη ενότητα, αλλά στην συνέχεια θα γίνει μια πιο αναλυτική παρουσίαση των τεχνικών και μεθόδων που χρησιμοποιήθηκαν για το σύστημα.

Καταρχήν το σύστημα είναι εξαρτώμενο από το μουσικό όργανο, που σημαίνει ότι αναγνωρίζει τον ήχο από τα συγκριμένα μουσικά όργανα που χρησιμοποιήθηκαν στο στάδιο της εκπαίδευσης. Δηλαδή, πρέπει να χρησιμοποιηθεί το μουσικό όργανο π.χ. η κιθάρα που χρησιμοποιήθηκε για την εκπαίδευση του συστήματος. Διαφορετική κιθάρα θα έχει αλλά αποτελέσματα στην απόδοση της σωστής αναγνώρισης. Σημαντικό ρόλο διαδραματίζουν, επίσης, και οι υπάρχουσες ακουστικές ιδιότητες του χώρου. Αν έχει ηχογραφηθεί το μουσικό όργανο σε δωμάτιο που η στάθμη θορύβου περιβάλλοντος είναι μικρή, τότε πρέπει να ισχύουν ίδιες, πιο σωστά όμοιες συνθήκες και κατά την διάρκεια της αναγνώρισης. Πράγμα που σημαίνει ότι αν η στάθμη του περιβάλλοντος θορύβου είναι αρκετά υψηλή στο στάδιο της αναγνώρισης από την στάθμη του σταδίου της εκπαίδευσης, τότε η απόδοση θα μειωθεί κατακόρυφα.

Κατά δεύτερον το σύστημα αναγνωρίζει μόνο μεμονωμένες νότες και όχι διάφορες μουσικές φράσεις. Δηλαδή, το σύστημα έχει εκπαιδευθεί με την χρήση μια νότας από κάθε όργανο και αναγνωρίζει το όργανο όταν παίζουμε την ίδια νότα από το εκάστοτε όργανο. Διαφορετικά, αν παίζουμε π.χ. μια μουσική φράση ή κάποια συγχορδία, η απόδοση του συστήματος μειώνεται.

Υπάρχουν δύο στάδια για την υλοποίηση του συστήματος και αυτά είναι:

- α) Το στάδιο της εκπαίδευσης(training stage) και
- β) το στάδιο της εξέτασης ή αναγνώρισης(testing or recognition stage).

Στο στάδιο της εκπαίδευσης περιλαμβάνεται η 'εκμάθηση' του συστήματος. Αυτό γίνεται δημιουργώντας, θα λέγαμε, ένα 'λεξικό', το οποίο είναι ένα ακουστικό

μοντέλο για κάθε μουσικό όργανο. Το σύστημα χρειάζεται αυτό το 'λεξικό' (ακουστικά μοντέλα) για το στάδιο της αναγνώρισης. Στο σύστημα αυτό, το 'λεξικό' αυτό περιλαμβάνει ακουστικά μοντέλα από 5 μουσικά όργανα, τα οποία είναι η κιθάρα, το πιάνο, η φλογέρα, το ξυλόφωνο και η φουσαρμόνικα. Το ακουστικό μοντέλο για κάθε όργανο έχει δημιουργηθεί προσαρμόζοντας τις Γκαουσιανές Μίξεις Μοντέλων (GMM ή GMM) στις τιμές των χαρακτηριστικών που έχουν εξαχθεί. Συγκεκριμένα τα χαρακτηριστικά αυτά είναι οι συντελεστές MFCC, delta και delta-delta. Το επονομαζόμενο 'λεξικό' είναι θέσεις μνήμης που περιέχουν τις Γκαουσιανές Μίξεις Μοντέλων(GMM) για κάθε ένα μουσικό όργανο.

Στο στάδιο της εξέτασης ή αναγνώρισης γίνεται χρήση των ακουστικών μοντέλων για κάθε μουσικό όργανο για την αναγνώριση μεμονωμένων νοτών. Η αναγνώριση γίνεται χρησιμοποιώντας ένα αλγόριθμο ταξινόμησης(classification algorithm) και συγκεκριμένα τον ταξινομητή GMM(GMM classifier) που χρησιμοποιεί μια μίξη από Γκαουσιανά μοντέλα, μιας και οι Γκαουσιανές κατανομές προσεγγίζουν αρκετά καλά τις ακουστικές ιδιότητες ενός ηχητικού σήματος. Επίσης, η μίξη Γκαουσιανών μοντέλων χρησιμοποιείται για την δημιουργία του ακουστικού μοντέλου κάθε οργάνου στην φάση της εκπαίδευσης.

Για το συγκεκριμένο πρόγραμμα χρησιμοποιήθηκε η γλώσσα προγραμματισμού 'MATLAB Version 7' και τα εργαλεία που χρειάζονται είναι τα 'Data Acquisition Toolbox', 'Signal Processing Toolbox' και 'Statistics Toolbox'. Επίσης, χρησιμοποιήθηκαν τα mfcc.m από το Auditory toolbox του Matlab, deltacoeff.m του Speechcore που υπάρχει στην ιστοσελίδα του Matlab και το epdcode.m από το internert, με διάφορες τροποποιήσεις σε κάποια από αυτά τα αρχεία για την προσαρμογή τους στο σύστημα.

5.1.1 Στάδιο Εκπαίδευσης(Training Stage)

Τόσο στο στάδιο της εκπαίδευσης όσο και στο στάδιο της αναγνώρισης γίνεται η προ-επεξεργασία και η εξαγωγή χαρακτηριστικών από τα σήματα που εισάγονται στην είσοδο του συστήματος.

Στο στάδιο της εκπαίδευσης, το ηχητικό σήμα από την νότα που παράγει ένα όργανο καταγράφεται μέσω ενός μικροφώνου SHARP RMICD0231 και το σήμα το αποθηκεύουμε σαν αρχείο .wav στην μνήμη του Η/Υ. Τη νότα μπορούμε να την διεγείρουμε με διάφορους τρόπους και να την αφήσουμε να παίζει όσο χρονικό διάστημα επιθυμούμε, βέβαια στα όρια των 5 ή 10 δευτερολέπτων που έχουμε επιλέξει για την ηχογράφηση. Για παράδειγμα, παίζουμε μια νότα από την κιθάρα με νύξη του δακτύλου μας σε μια από τις χορδές της κιθάρας. Η νότα σιγά-σιγά αποσβένει και μετά το διάστημα των 5 ή 10 δευτερολέπτων, σταμάτα η ηχογράφηση. Το σήμα μας έχει πλέον καταγραφεί και το αποθηκεύουμε στην συνέχεια στην μνήμη του Η/Υ. Για το σύστημα στην παρούσα πτυχιακή εργασία, έχουμε χρησιμοποιήσει μνήμη 20 θέσεων για αυτά τα 5 όργανα, που σημαίνει ότι ηχογραφούμε 4 διαφορετικές εκδοχές της νότας για κάθε όργανο και έτσι αποθηκεύουμε 20 αρχεία ήχου. Αυτές τις τέσσερις διαφορετικές εκδοχές, τις ονομάζουμε παραδείγματα για λόγους ευκολίας. Μια εκδοχή ή ένα παράδειγμα μπορεί να περιέχει μια νότα που έχουμε παίζει ή την ίδια νότα επαναλαμβανομένη 2,3 ή παραπάνω φορές(Προτείνεται η ηχογράφηση κάθε παραδείγματος να περιέχει την ίδια νότα παιγμένη 2 ή 3 φορές, για αύξηση της απόδοσης του συστήματος). Βασική παράμετρος που πρέπει να ληφθεί υπόψη στην καταγραφή του ήχου από κάθε όργανο είναι η συχνότητα

δειγματοληψίας, η τιμή της οποίας είναι ίση με 16000 samples/sec. Με αυτό τον τρόπο έχουμε στην διάθεση μας 20 αρχεία από 5 διαφορετικά μουσικά όργανα που έχουν συχνότητα δειγματοληψίας ίση με $fs=16000$ samples/sec.

Στην συνέχεια, διαλέγουμε ένα-ένα τα αποθηκευμένα αρχεία και τα εισάγουμε στην διαδικασία της εκπαίδευσης. Στην εκπαίδευση γίνεται μια προ-επεξεργασία του σήματος και χωρίζουμε κάθε σήμα σε frames. Το χρονικό διάστημα κάθε frame έχει δειχθεί ότι είναι καλό να είναι περίπου στα 20ms. Στην παρούσα εργασία έχουμε χρησιμοποιήσει αρχικά frame χρονικής διάρκειας 10ms. Έτσι, λόγω της συχνότητας δειγματοληψίας που χρησιμοποιούμε το μήκος του πλαισίου(frame) είναι ίσο με $16000 \cdot 0.01 = 160$ samples ή δείγματα. Επομένως, θα χρησιμοποιηθεί ένα μήκος επικάλυψης(overlapsize) ίσο με (μήκος του πλαισίου/2)= $160/2=80$ samples. Το μήκος μεταπήδησης ή hopsize ή βήμα θα είναι ίσο με (μήκος του πλαισίου-μήκος επικάλυψης)= $160-80=80$ samples. Πράγμα που σημαίνει ότι το frame rate είναι ίσο με $frame\ rate = fs/80 = 16000/80 = 200\text{Hz}$ (200frames/sec).

Αναφέρεται στην βιβλιογραφία της αναγνώρισης μεμονωμένων λέξεων, ότι για να εντοπίσεις μια λέξη είναι χρήσιμο να υπολογίσεις την ενέργεια του σήματος και το χαρακτηριστικό zero-crossing για κάθε ένα frame. Μέσω αριθμητικών και λογικών πράξεων από το περιβάλλον της γλώσσας προγραμματισμού MATLAB, εξάγουμε και αποθηκεύουμε τα samples(δείγματα) των frames που υπάρχει κάποια σήμα και όχι θόρυβος (χρησιμοποιείται ένα σύστημα End-Point Detection ή EPD που σκοπός του είναι σε ένα σύστημα αναγνώρισης ομιλίας, ο εντοπισμός ενός ήχου και η διάκριση του από τον θόρυβο περιβάλλοντος αλλά και ο εντοπισμός αν είναι voiced ομιλία(π.χ. φωνήεντα) ή unvoiced(π.χ. σύμφωνα)). Τα δεδομένα που το σύστημα υπολογίζει και θεωρεί ότι περιέχουν σήμα(την νότα) χρησιμοποιούνται ως είσοδος σε ένα άλλο κώδικα για την εξαγωγή των MFCC, delta και delta-delta χαρακτηριστικών.

Στην συνέχεια αφού επιστραφεί ένα (39αριθμός frames) διάνυσμα χαρακτηριστικών που περιέχει τα προαναφερθέντα χαρακτηριστικά τότε για τις τιμές αυτές των χαρακτηριστικών αντιστοιχούμε ένα άθροισμα Γκαουσιανών κατανομών. Αυτή η Γκαουσιανή κατανομή σχηματίζει το μοντέλο της νότας για το εκάστοτε όργανο που χρησιμοποιήσαμε. Η Γκαουσιανή συνάρτηση πυκνότητας πιθανότητας παραμετροποιείται από τα βάρη των απλών Γκαουσιανών κατανομών, το διάνυσμα μέση τιμή των τιμών κάθε χαρακτηριστικού και τον πίνακα συνδιασποράς από όλες τις συνιστώσες(μίξεις) των απλών Γκαουσιανών κατανομών. Η μεμονωμένη νότα (που έχει επαναληφθεί μια ή παραπάνω φορές στο αποθηκευμένο αρχείο ήχου) κάθε οργάνου μπορεί πλέον να αντιπροσωπευθεί από τις παραμέτρους της μίξης Γκαουσιανών μοντέλων(GMM). Για να γίνει εκτίμηση των παραμέτρων ενός GMM για ένα σύνολο από MFCC, delta και delta-delta διανύσματα χαρακτηριστικών που έχουν εξαχθεί κατά την φάση της εκπαίδευσης πρέπει να χρησιμοποιηθεί ο EM αλγόριθμος(Expectation-Maximization) ώστε να γίνει εκτίμηση της μέγιστης πιθανοφάνειας(Maximum Likelihood) .

Στο περιβάλλον της γλώσσας προγραμματισμού MATLAB αυτό μπορεί να γίνει εφικτό, χρησιμοποιώντας από το εργαλείο Statistics Toolbox την συνάρτηση `gmdistribution`. Δίνοντας ένα σύνολο από MFCC, delta και delta-delta χαρακτηριστικά και τον αριθμό των συνιστωσών των απλών Γκαουσιανών κατανομών που θα χρησιμοποιηθούν, τότε υπολογίζονται οι παράμετροι του GMM για κάθε μοντέλο. Με αυτό τον τρόπο η διαδικασία της εκπαίδευσης ολοκληρώνεται έχοντας στην διάθεση πλέον τις παραμέτρους των GMM για κάθε μοντέλο της νότας από κάθε όργανο που χρησιμοποιήσαμε.

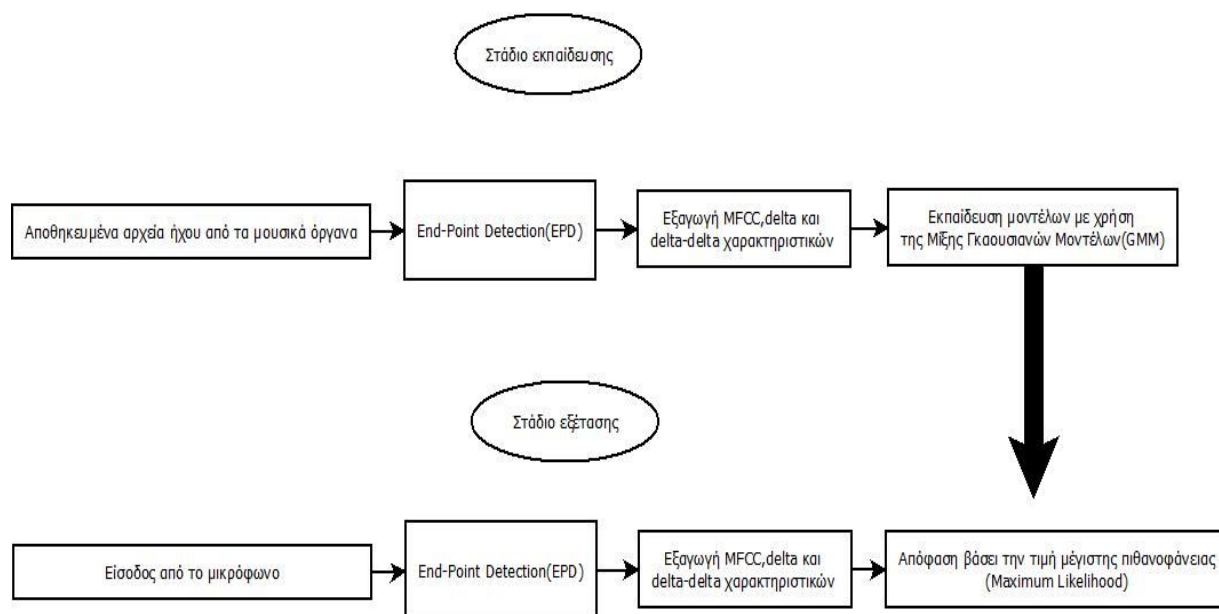
5.1.2 Στάδιο της εξέτασης(Testing Stage)

Στην διαδικασία της εξέτασης ή αναγνώρισης πρέπει να γίνει η επιλογή ενός κατάλληλου ταξινομητή για την σωστή ταξινόμηση κάθε προτύπου(που το σήμα του εισάγεται στο υπολογιστικό σύστημα μέσω του μικροφώνου) στην κλάση που ανήκει.

Και στην διαδικασία της εξέτασης γίνεται η χρήση ενός EPD συστήματος για τον εντοπισμό μιας νότας. Στην συνέχεια εξάγουμε τα χαρακτηριστικά των MFCC, delta και delta-delta από κάθε frame. Σκοπός είναι η εύρεση ενός μοντέλου της νότας(που έχουμε στην διάθεση μας από το στάδιο της εκπαίδευσης)με την μέγιστη εκ των υστέρων πιθανότητα για το σύνολο των διανυσματικών χαρακτηριστικών που έχουν προκύψει από τα πρότυπα στο στάδιο της εξέτασης. Χρησιμοποιώντας την posterior συνάρτηση που περιέχεται στο Statistics Toolbox μπορούμε να υπολογίσουμε την αρνητική λογαριθμική πιθανοφάνεια:

$$[\text{Posterior, Negative log-likelihood}] = \text{posterior}(\text{model}(\text{note}), \text{testdata}) \quad (5.1.1)$$

Τα ορίσματα που χρειάζεται η συνάρτηση για τον υπολογισμό της αρνητικής λογαριθμικής πιθανοφάνειας(negative log-likelihood) είναι το GMM μοντέλο μιας νότας(που έχουμε στην διάθεση μας από το στάδιο της εκπαίδευσης) και το διανύσμα χαρακτηριστικών του προτύπου που εισάγουμε στο σύστημα(δεδομένα εξέτασης από ένα σύνολο frames διάρκειας άνω των 255ms). Επαναλαμβάνοντας αυτό τον υπολογισμό για κάθε μοντέλο νότας, τότε στο στάδιο της εξέτασης το μουσικό όργανο θα ταξινομηθεί ως το μουσικό όργανο που η GMM της νότας του παράγει την μέγιστη τιμή πιθανοφάνειας. Αυτό είναι εφικτό όταν υπολογίσει κανείς την ελαχίστη τιμή της αρνητικής λογαριθμικής πιθανοφάνειας. Έτσι, το σύστημα θα είναι ικανό να αναγνωρίσει το μουσικό όργανο, από μια συγκεκριμένη νότα που χρησιμοποιείται ως είσοδος από το μικρόφωνο στο σύστημα. Ένα γενικό block διάγραμμα φαίνεται στην παρακάτω εικόνα.



Εικόνα. 5.1.1. Διάγραμμα ροής ενός συστήματος αναγνώρισης μουσικών οργάνων.

5.2 Αναλυτικά τα στάδια της εκπαίδευσης και εξέτασης

Σε αυτό το σημείο θα αναφέρουμε πιο αναλυτικά την λειτουργία κάθε σταδίου του συστήματος, ώστε να προβληθούν οι βήμα προς βήμα διαδικασίες για την επιτέλεση της αναγνώρισης μουσικών οργάνων, που είναι και ο τελικός στόχος του συστήματος και της παρούσας εργασίας. Στην παρουσίαση των σταδίων θα προβληθούν διαγράμματα ροής, όπου θα παρουσιάζεται η γενική φιλοσοφία των υπολογισμών και επεξεργασιών για την καλύτερη κατανόηση της λειτουργίας του αλγορίθμου κάθε σταδίου. Συνήθως, οι περισσότερες εντολές που θα αναφέρονται στα διαγράμματα ροής απευθύνονται στις εντολές που υποστηρίζει η γλώσσα προγραμματισμού MATLAB και οι υπόλοιπες, απλώς φανερώνουν την λειτουργία που επιτελούν, χωρίς η ονομασία τους να υποστηρίζεται από το περιβάλλον του MATLAB. Αν και συνήθως δεν ενδείκνυται η χρήση εντολών μιας γλώσσας προγραμματισμού σε ένα διάγραμμα ροής, στην συγκεκριμένη περίπτωση, τις χρησιμοποιούμε για λόγους που βοηθούν στην καλύτερη κατανόηση των αλγορίθμων. Οι κώδικες από τους οποίους αποτελείται το πρόγραμμα θα παρατεθούν στις τελευταίες σελίδες.

Θα ξεκινήσουμε από το στάδιο της εκπαίδευσης και συγκεκριμένα από τον κώδικα End-Point Detection εκπαίδευσης, που είναι αρκετά σημαντικός στο στάδιο της εκπαίδευσης των μοντέλων των ‘μεμονωμένων’ νοτών από κάθε μουσικό όργανο.

5.2.1 End-Point Detection εκπαίδευσης

Στο End-Point Detection εκπαίδευσης, εισάγεται το σήμα που έχουμε ηχογραφήσει από το παίξιμο της νότας από ένα μουσικό όργανο και έχουμε αποθηκεύσει στην μνήμη του Η/Υ ως ένα αρχείο .wav (π.χ. μουσικό _όργανο_No1.wav). Στην συνέχεια, το σήμα αυτό το χωρίζουμε σε πλαίσια(frames), το πλήθος των οποίων είναι ίσο με (αριθμός samples του σήματος νότας)/hopsize-1. Το hopsize που χρησιμοποιείται είναι 80 samples(δείγματα), λόγω του ότι το μέγεθος του πλαισίου(frame) είναι ίσο με 160 δείγματα και το μέγεθος της επικάλυψης(overlapsize)είναι ίσο με (μέγεθος πλαισίου)/2=160/2=80δείγματα. Άρα, το hopsize είναι ίσο με (μέγεθος πλαισίου)- (μέγεθος της επικάλυψης)=160-80=80 δείγματα. Έτσι, η χρονική διάρκεια κάθε frame είναι ίση με (μέγεθος πλαισίου)/(συχνότητα δειγματοληψίας)=160/16000=0.01sec ή 10 msec και η χρονική διάρκεια της επικάλυψης και του βήματος(hopsize) είναι ίση με 80/16000=0.005sec ή 5 msec. Από κάθε frame αφαιρούμε την μέση τιμή του frame. Στην συνέχεια για κάθε ένα από τα frames που έχουν δημιουργηθεί, υπολογίζουμε την λογαριθμική ενέργεια του σήματος που περιέχεται σε αυτά. Η σχέση του υπολογισμού της λογαριθμικής ενέργειας είναι ίση με:

$$Frame_energy = \log \left(\sum_{i=1}^{αριθμός\ frames} x_i^2 \right) \quad (5.2.1)$$

, όπου x είναι η ενέργεια του σήματος που περιέχεται σε κάθε frame. Επίσης, υπολογίζεται και ο αριθμός των zero crossing σε κάθε ένα frame. Στην συνέχεια θεωρούμε ότι στα πρώτα 50 frames, υπάρχει ο θόρυβος περιβάλλοντος (όχι frames που αφορούν το σήμα της νότας) και για αυτά τα frames υπολογίζεται η χαμηλότερη ενέργεια threshold (low energy threshold). Στην συνέχεια, υπολογίζουμε την τιμή κατωφλίου (threshold) της ενέργειας για τα πρώτα 50 επικαλυπτόμενα frames που θεωρούμε ότι περιέχουν τον θόρυβο περιβάλλοντος. Η σχέση που υπολογίζει την τιμή κατωφλίου της ενέργειας είναι ίση με:

$$\text{Energy threshold} = \frac{1}{N} \sum_{i=1}^N Z(i) + a \cdot \frac{1}{N} \sum_{i=1}^N (Z(i) - \bar{x})^2 \quad (5.2.2)$$

ή με άλλα λόγια

$$\text{Energy threshold} = \text{μέση τιμή}(Z) + a \cdot \text{τυπική απόκλιση}(Z) \quad (5.2.3)$$

, όπου Z είναι ο πίνακας που έχουμε αποθηκεύσει τις τιμές λογαριθμικής ενέργειας για τα πρώτα 50 frames, a είναι μια μεταβλητή που η τιμή της δίνεται ανάλογα με τον θόρυβο περιβάλλοντος που επικρατεί στο δωμάτιο ηχογράφησης, N είναι ίση με 50 (το πλήθος των frames που υπολογίζουμε την τιμή κατωφλίου) και \bar{x} είναι η μέση τιμή των τιμών του πίνακα Z , που είναι ίση με:

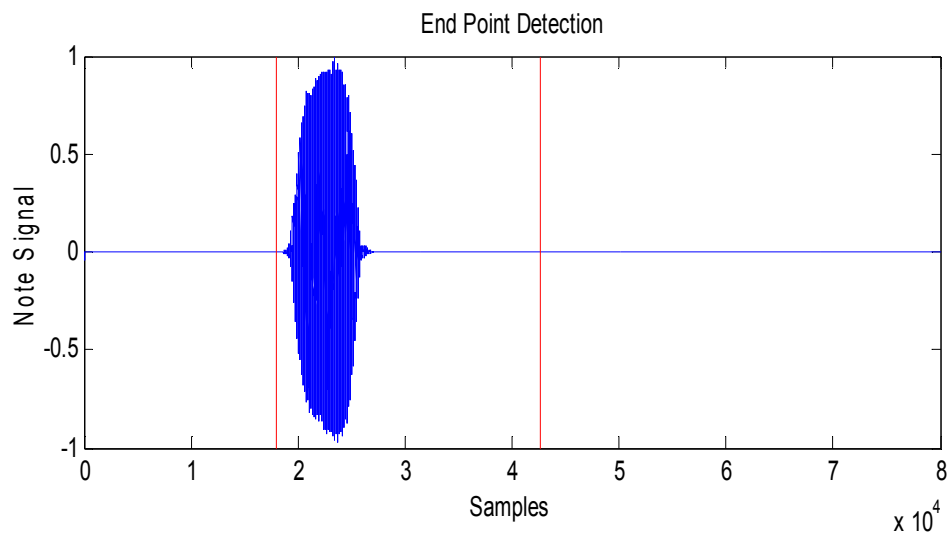
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N Z(i) \quad (5.2.4)$$

Έτσι, έχουμε στην διάθεση μας μια τιμή κατωφλίου. Με την τιμή κατωφλίου της ενέργειας (energy threshold) μπορούμε να υπολογίσουμε μια άλλη τιμή κατωφλίου (Amplitude_threshold) την οποία χρειαζόμαστε για την λειτουργία του κώδικα. Ο τύπος που υπολογίζει αυτή την τιμή είναι ίσος με:

$$\text{Amplitude_threshold} = \max(\text{ενέργεια frame})^2 - |\text{energy threshold}| \quad (5.2.5)$$

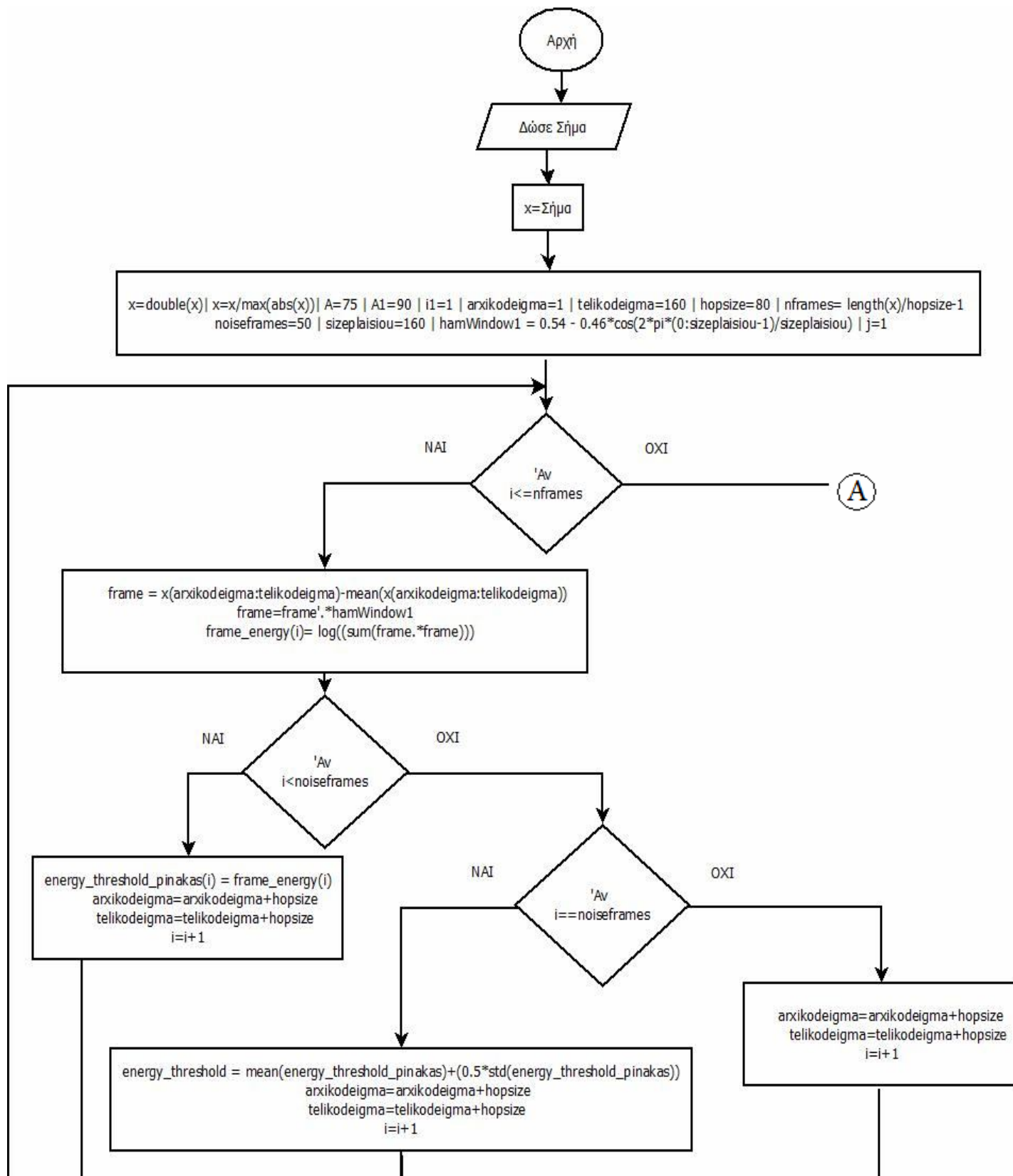
Μετά τα 50 frames, στα υπόλοιπα frames συγκρίνουμε την ενέργεια τους με την δεύτερη τιμή κατωφλίου (Amplitude_threshold). Αν ισχύει η συνθήκη, η τιμή της ενέργειας τους να είναι μεγαλύτερη ή ίση από αυτή την τιμή κατωφλίου, τότε δημιουργείται ένας πίνακας με την πρώτη του θέση να περιέχει την τιμή του δείγματος που ισχύει αυτή η συνθήκη και η τελευταία θέση περιέχει την τελευταία τιμή του δείγματος που ισχύει αυτή η συνθήκη. Έτσι, παίρνοντας από τον πίνακα την αρχική και τελική του τιμή, έχουμε τα άκρα που ανάμεσα σε αυτά περιέχεται το σήμα της νότας του μουσικού οργάνου που έχουμε ηχογραφήσει το σήμα. Αυτή είναι αρχή λειτουργίας του End-Point Detection. Οι τιμές του σήματος που περιέχονται ανάμεσα σε αυτά τα άκρα είναι και η έξοδος του End-Point Detection εκπαίδευσης που θα εισαχθεί σαν είσοδος στον κώδικα της εξαγωγής των MFCC, delta και delta-delta χαρακτηριστικών.

Μεγάλο μέρος του σήματος (κατά προτίμηση όλο το σήμα της νότας) πρέπει να περιέχεται ανάμεσα στις κόκκινες κάθετες γραμμές, όπως φαίνεται και στην παρακάτω εικόνα.

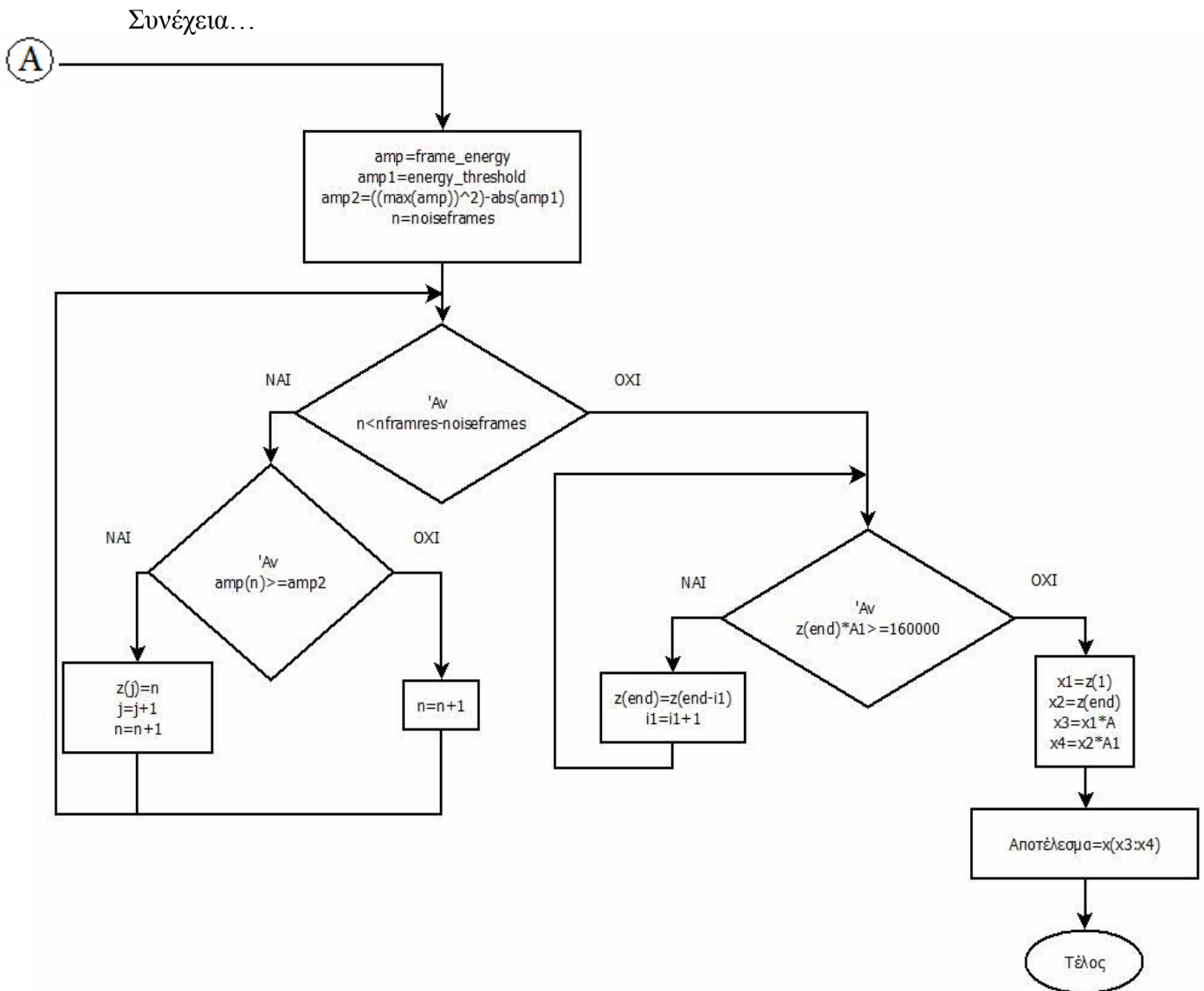


Εικόνα 5.2.1. Σήμα μιας νότας E από μια φουσαρμόνικα. Οι κάθετες κόκκινες γραμμές οριοθετούν το σήμα το οποίο εξάγεται από τον κώδικα του End-Point Detection. Έτσι, οι τιμές των δειγμάτων που θα εξαχθούν θα είναι οι τιμές που είναι ανάμεσα στις κόκκινες γραμμές.

Δηλαδή, οι τιμές του σήματος που θα εξαχθούν (το εξαγόμενο σήμα) περιέχονται μέσα στις κάθετες κόκκινες γραμμές που φαίνονται στο παραπάνω σχήμα. Για τον τρόπο που λειτουργεί το End-Point Detection εκπαίδευσης θα παραθέσουμε το διάγραμμα ροής του.



Συνεχίζεται στην επόμενη σελίδα...



Εικόνα 5.2.2. Διάγραμμα ροής του End-Point Detection εκπαίδευσης.

5.2.2 Εξαγωγή MFCC, delta, delta-delta χαρακτηριστικών και εκπαίδευση

Σε αυτό το στάδιο, σκοπός είναι η εξαγωγή των MFCC, delta και delta-delta χαρακτηριστικών από τα σήματα των νοτών, με σκοπό την χρησιμοποίησή τους για την εύρεση των μίξεων των Γκαουσιανών μοντέλων για κάθε νότα.

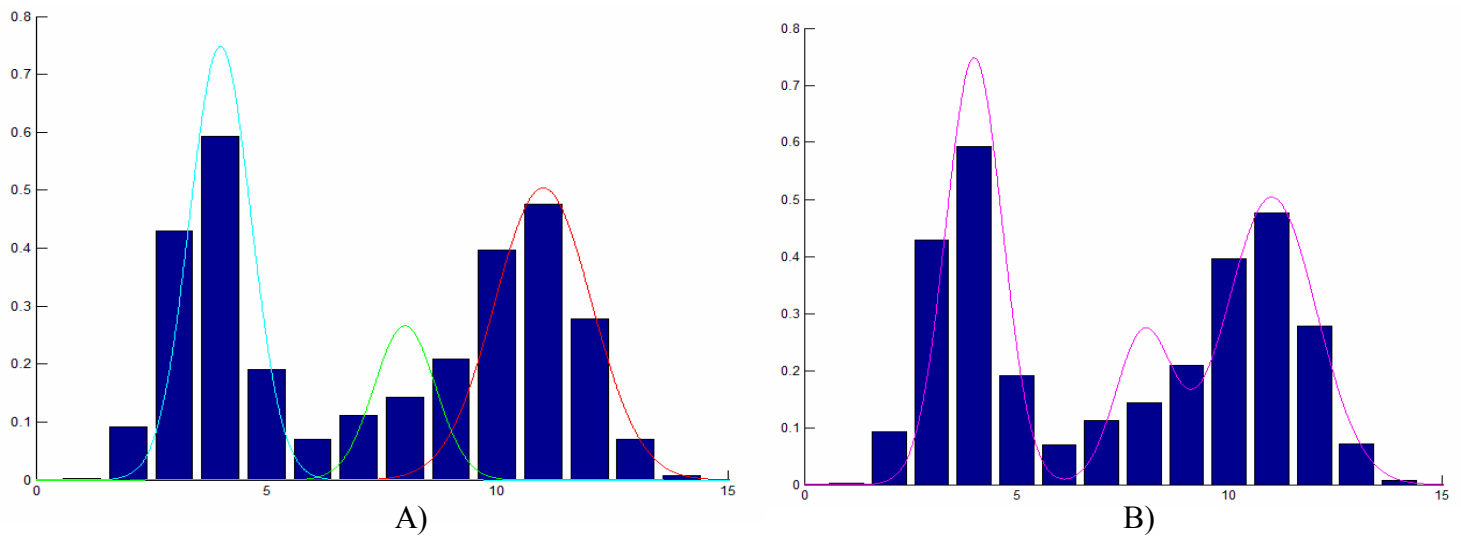
Στο στάδιο αυτό αρχικά γίνεται η χρήση του End-Point Detection για κάθε μία νότα και εξάγεται το σήμα που περιέχεται ανάμεσα στις κόκκινες κάθετες γραμμές, όπως φαίνεται στην εικόνα 5.2.1. Το σήμα κάθε μίας νότας χωρίζεται σε frames, το πλήθος των οποίων είναι ίσο με $\text{length}(\text{σήμα νότας})/\text{hopsizel}-1$. Το hopsize που χρησιμοποιείται είναι 80 samples(δείγματα), λόγω του ότι το μέγεθος του πλαισίου(frame) είναι ίσο με 160 δείγματα και το μέγεθος της επικάλυψης(overlapsize)είναι ίσο με $(\text{μέγεθος πλαισίου})/2=160/2=80$ δείγματα. Άρα, το hopsize είναι ίσο με $(\text{μέγεθος πλαισίου})- (\text{μέγεθος της επικάλυψης})=160-80=80$ δείγματα. Έτσι, κάθε frame εισάγεται σε ένα πίνακα τύπου (μέγεθος πλαισίου) \times (αριθμός πλαισίων,frames). Τα δεδομένα που περιέχει αυτός ο πίνακας γίνονται είσοδοι στο κώδικα του υπολογισμού των MFCC, delta και delta-delta χαρακτηριστικών. Ο θεωρητικός υπολογισμός των MFCC, delta και delta-delta χαρακτηριστικών έχει αναφερθεί σε προηγούμενη ενότητα. Στην συνέχεια, χρησιμοποιείται η εντολή

gmdistribution.fit(mfccδεδομένα',8,'CovType', 'diagonal') (5.2.6)

, όπου για κάθε μοντέλο υπολογίζει την μίξη των Γκαουσιανών κατανομών, δοσμένων των χαρακτηριστικών MFCC, delta και delta-delta(έναν πίνακα 39 διαστάσεων, δηλαδή 39 \times (αριθμός frames)) και του αριθμού των συνιστωσών των Γκαουσιανών κατανομών που θα χρησιμοποιηθούν(στην περίπτωση αυτή θα χρησιμοποιηθούν 8 Γκαουσιανές κατανομές για τα MFCC, delta και delta-delta χαρακτηριστικά, η μίξη των οποίων θα αντιστοιχηθεί σε κάθε παράδειγμα νότας που έχουμε ηχογραφήσει).

Στην παρούσα εργασία χρησιμοποιούμε 20 θέσεις μνήμης για τα δεδομένα της εκπαίδευσης και λόγω ότι χρησιμοποιούμε 5 όργανα, αυτό σημαίνει ότι, χρησιμοποιούμε 4 παραδείγματα για κάθε νότα ενός οργάνου. Όπως, έχει αναφερθεί παραπάνω τα επωνομαζόμενα ως παραδείγματα είναι τέσσερις διαφορετικές εκδοχές των νοτών που παίζονται από κάθε όργανο. Επίσης, σε κάθε παράδειγμα μπορεί να περιέχεται μια νότα, η οποία επαναλαμβάνεται μία ή περισσότερες φορές.

Άρα, στην περίπτωση που εισάγεται το δεύτερο παράδειγμα της νότας από το μουσικό όργανο No 5(π.χ. που είναι η φλογέρα), τότε γι' αυτό θα αντιστοιχηθεί μια μίξη Γκαουσιανών κατανομών. Ένα παράδειγμα μίξης Γκαουσιανών κατανομών για το μιας διάστασης διάνυσμα MFCC χαρακτηριστικού φαίνεται στην παρακάτω εικόνα.



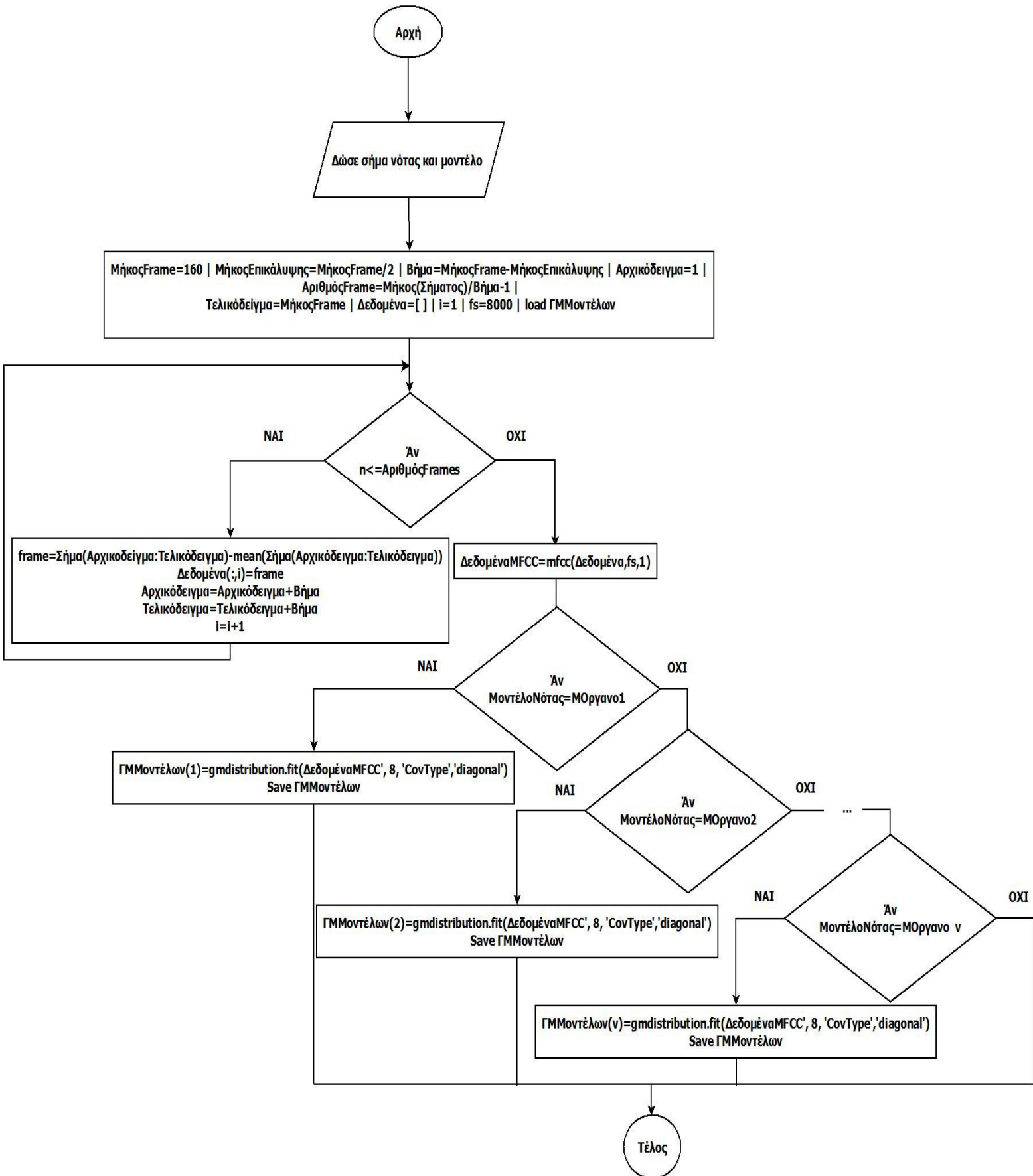
Εικόνα 5.2.3. Τυχαία κατανομή της πρώτης διάστασης για το διάνυσμα του MFCC χαρακτηριστικού και η αντιστοίχιση μιας μίξης 3 απλών Γκαουσιανών κατανομών
 Α) Οι 3 απλές Γκαουσιανές (ή κανονικές) κατανομές Β) Η μίξη των 3 απλών Γκαουσιανών κατανομών και η αντιστοίχιση της στα δεδομένα MFCC.

Σαν είσοδο, στον αλγόριθμο που φαίνεται στο παρακάτω διάγραμμα ροής δίνουμε το σήμα που έχει εξαχθεί από το End-Point Detection εκπαίδευσης (το σήμα που περιέχεται ανάμεσα στις δύο κάθετες κόκκινες γραμμές, όπως αναφέρεται στην ενότητα του End-Point Detection εκπαίδευσης) και το όνομα του μοντέλου που θέλουμε να εκπαιδύσουμε (π.χ. ΜΌργανο1 για το όργανο Νο1, που μπορεί να είναι το σήμα της νότας της κιθάρας).

Έτσι, έχουμε αποθηκευμένες στην μνήμη του Η/Υ τις μίξεις των Γκαουσιανών μοντέλων για κάθε παράδειγμα νότας για τα 5 όργανα που χρησιμοποιούνται.

Η φάση της εκπαίδευσης έχει ολοκληρωθεί και για την επιβεβαίωση της σωστής λειτουργίας του συστήματος πρέπει να χρησιμοποιηθεί το στάδιο της εξέτασης, για να ολοκληρωθεί ο σκοπός του συστήματος, που είναι η αυτόματη αναγνώριση των 5 αυτών μουσικών οργάνων, όταν χρησιμοποιούμε ως είσοδο στο σύστημα το σήμα μιας νότας από κάθε ένα από αυτά τα 5 μουσικά όργανα.

Στο παρακάτω διάγραμμα ροής της εκπαίδευσης χρησιμοποιείται συχνότητα δειγματοληψίας 8000 samples/sec και λόγω ότι μπορεί να αποθηκεύσει n σήματα από μουσικά όργανα, παρουσιάζεται στην γενική του μορφή. Στην περίπτωση της παρούσας εργασίας το $n=20$, διότι γίνεται ηχογράφηση 20 παραδειγμάτων νοτών για όλα τα μουσικά όργανα με 4 παραδείγματα (4 διαφορετικές εκδοχές) για κάθε ένα μουσικό όργανο.



Εικόνα 5.2.4. Διάγραμμα ροής για την εκπαίδευση και αποθήκευση των μοντέλων των νοτών για κάθε όργανο .

5.2.3 Αυτόματη ηχογράφηση, αποθήκευση και εκπαίδευση των αρχείων ήχου

Αν και έχουμε περιγράψει την διαδικασία της εκπαίδευσης παραπάνω, στην ενότητα αυτή θα παρουσιαστεί ένας αλγόριθμος, ο οποίος θα δίνει την διευκόλυνση στο χρήστη να εκπαιδεύσει από την αρχή τα μοντέλα των νοτών, χωρίς να χρειάζεται να γράφει εντολές στο περιβάλλον του MATLAB για αυτή την διαδικασία, αλλά μόνο πατώντας ένα από δύο πλήκτρα του πληκτρολογίου για να επιβεβαιώσει αν θέλει να ηχογραφήσει ή όχι κάποιο από τα παραδείγματα νότας για τα όργανα που θα χρησιμοποιηθούν. Τα πλήκτρα αυτά είναι το Y το οποίο αντιστοιχεί σε κατάφαση(στην λέξη ΝΑΙ) και το N που αντιστοιχεί σε άρνηση(στην λέξη ΟΧΙ). Οι λεπτομέρειες παρουσιάζονται παρακάτω.

Για διευκόλυνση του χρήστη προσφέρεται στο πρόγραμμα, η επιλογή της εκπαίδευσης των μοντέλων, πατώντας απλώς το πλήκτρο της 'Εκπαίδευσης Μοντέλων' στο GUI(Graphical User Interface). Επίσης, μπορεί να επιλέξει την διάρκεια της ηχογράφησης του μοντέλου πατώντας το πλήκτρο 5 ή 10sec, δίνοντας την δυνατότητα ηχογράφησης ενός αρχείου ήχου διάρκειας 5 ή 10sec, αντίστοιχα(Default ή στάνταρ τιμή διάρκειας ηχογράφησης είναι 10sec). Όταν ο χρήστης πατήσει το πλήκτρο της 'Εκπαίδευσης Μοντέλων', τότε θα εμφανιστεί στο παράθυρο εντολών(Command Window) του MATLAB, η δυνατότητα ότι μπορεί να επιλέξει την διάρκεια της ηχογράφησης και η ερώτηση που αναφέρει αν θέλει να ηχογραφήσει το πρώτο παράδειγμα για το πρώτο μουσικό όργανο. Στην περίπτωση που ο χρήστης επιλέξει ότι θέλει να ηχογραφήσει το πρώτο παράδειγμα του μουσικού οργάνου No1, τότε γίνεται η ηχογράφηση διάρκειας 5 ή 10 δευτερολέπτων ανάλογα με την επιλογή που έχει κάνει ο χρήστης. Το ηχογραφημένο σήμα γίνεται είσοδος στον κώδικα του End-Point Detection και εξάγονται τα άκρα του σήματος(νότας) που έχει υπολογίσει, χωρίς να λαμβάνονται τα δείγματα αριστερά και δεξιά της πρώτης και της δεύτερης κόκκινης κάθετης γραμμής(εικόνα 5.2.1). Επίσης, ένας διαπεραστικός ήχος ακούγεται ότι αποθηκεύτηκε η νότα του οργάνου και εμφανίζεται το γράφημα του End-Point-Detection της νότας. Μεγάλο μέρος του σήματος(κατά προτίμηση όλο το σήμα της νότας) πρέπει να περιέχεται ανάμεσα στις κόκκινες κάθετες γραμμές. Σε περίπτωση που δεν περιέχεται, δίνεται η δυνατότητα να ξανά-ηχογραφήσουμε το σήμα της νότας, όπως αναφέρεται και παρακάτω. Σε περίπτωση όμως, που δεν το ξανά-ηχογραφήσουμε και μεγάλο μέρος του σήματος της νότας δεν περιέχεται στις κόκκινες κάθετες γραμμές, τότε κατά την διάρκεια της εκπαίδευσης θα εμφανιστεί κάποιο error στο παράθυρο εντολών του MATLAB.

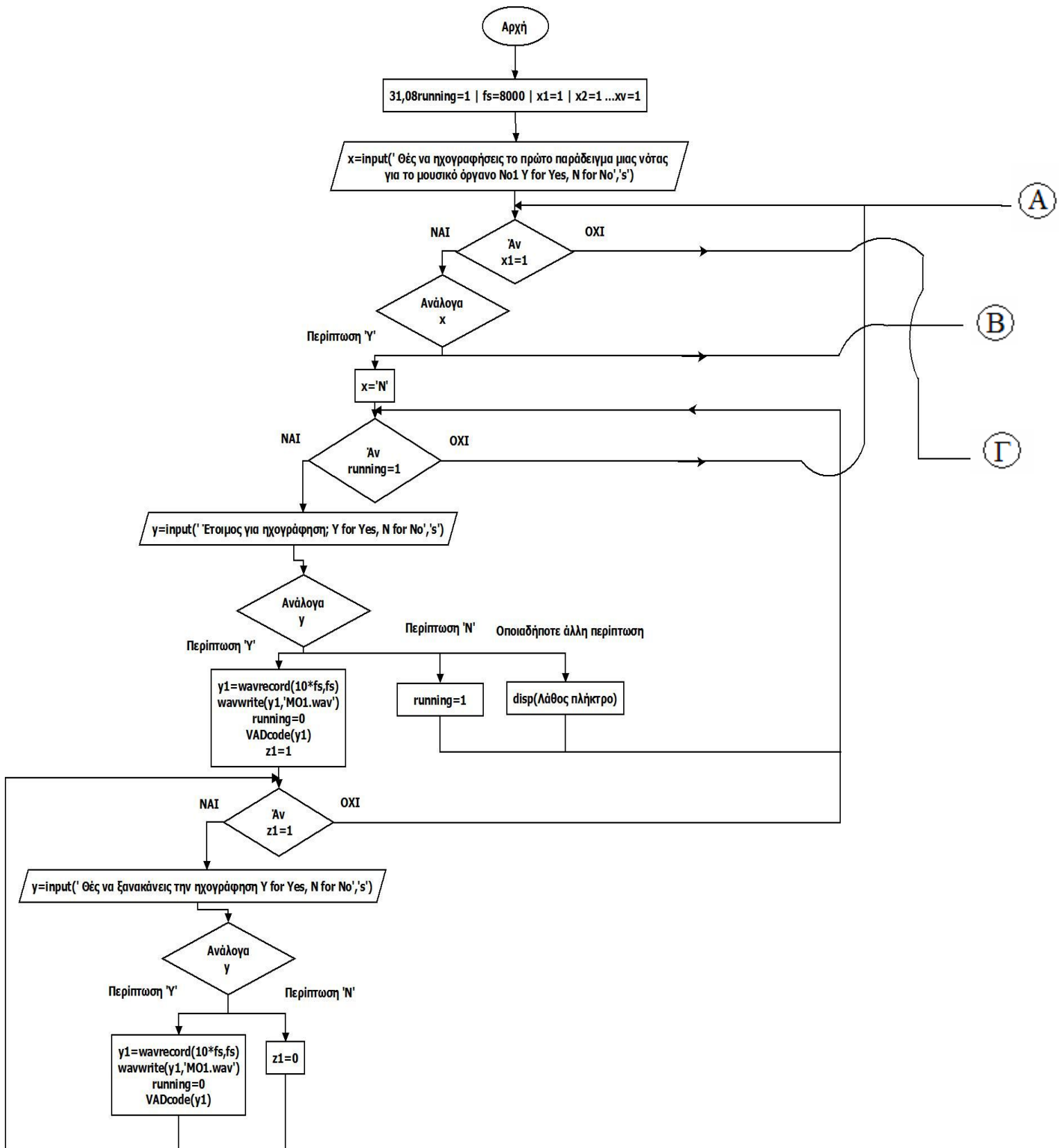
Στην συνέχεια της διαδικασίας, γίνεται ερώτηση αν θέλει ο χρήστης να ξανά-ηχογραφήσει την νότα του μουσικού οργάνου(σε περίπτωση που το σήμα της νότας δεν περιέχεται ανάμεσα στις κάθετες κόκκινες γραμμές, όπως θα φανεί από την γράφημα που θα παρουσιαστεί όταν γίνει η ηχογράφηση του σήματος).

Στην περίπτωση που δέχεται ο χρήστης να κάνει ξανά την ηχογράφηση, η παραπάνω διαδικασία επαναλαμβάνεται, μέχρι ο χρήστης να είναι ικανοποιημένος από το σήμα που θα εξαχθεί από το End-Point Detection. Αν επιθυμεί να μην ηχογραφήσει ξανά την νότα, τότε εμφανίζεται η ερώτηση αν θέλει να συνεχίσει τις ηχογραφήσεις. Αν δεχτεί τότε εμφανίζεται η ερώτηση αν θέλει να ηχογραφήσει το

δεύτερο παράδειγμα της νότας για το πρώτο μουσικό όργανο. Και ο αλγόριθμος ακολουθεί την ίδια πορεία που περιγράφηκε παραπάνω για όλα τα όργανα, ώστε να ληφθούν υπόψη όλα τα παραδείγματα των νοτών των χρησιμοποιούμενων οργάνων. Αν δεν δεχθεί τότε καμία άλλη ηχογράφηση δεν λαμβάνει χώρα και γίνεται η ερώτηση αν θέλει να εκπαιδεύσει τα μοντέλα. Σε περίπτωση που δεχτεί, τότε απλώς τα ηχογραφημένα αρχεία γίνονται είσοδοι στον κώδικα της εκπαίδευσης, διαφορετικά τερματίζει η διαδικασία. Έτσι, ο χρήστης μπορεί να επιλέξει να ηχογραφήσει όλα τα παραδείγματα των νοτών για όλα τα όργανα ή να ηχογραφήσει παραδείγματα νοτών δικής του επιλογής πλήθους οργάνων(μέχρι 5). Επίσης, αν στην αρχή δεν δεχθεί να ηχογραφήσει το πρώτο παράδειγμα της νότας του πρώτου οργάνου, τότε εμφανίζεται η ερώτηση, που φαίνεται και παραπάνω, η οποία ζητάει από το χρήστη να επιλέξει αν θέλει να συνεχίσει τις ηχογραφήσεις.

Στην περίπτωση που έχουν γίνει όλες οι ηχογραφήσεις και έχουν αποθηκευτεί τα σήματα των νοτών των χρησιμοποιούμενων μουσικών οργάνων ως αρχεία .wav στην μνήμη του Η/Υ, τότε αυτά τα αποθηκευμένα αρχεία θα χρησιμοποιηθούν σαν είσοδος σε ένα άλλο κώδικα, τον κώδικα της εκπαίδευσης. Ο κώδικας της εκπαίδευσης έχει αναλυθεί σε παραπάνω ενότητα. Απλώς θα αναφέρουμε περιληπτικά, ότι σε κάθε ένα σήμα νότας θα αντιστοιχηθεί μια μίξη Γκαουσιανών μοντέλων και τα μοντέλα αυτά αποθηκεύονται σε ένα .mat αρχείο, που θα είναι το σύνολο της εκπαίδευσης(training set). Με βάση αυτά τα μοντέλα θα γίνει η σύγκριση με τα πρότυπα της εξέτασης στο στάδιο της αναγνώρισης με σκοπό την αναγνώριση των εκάστοτε μουσικών οργάνων.

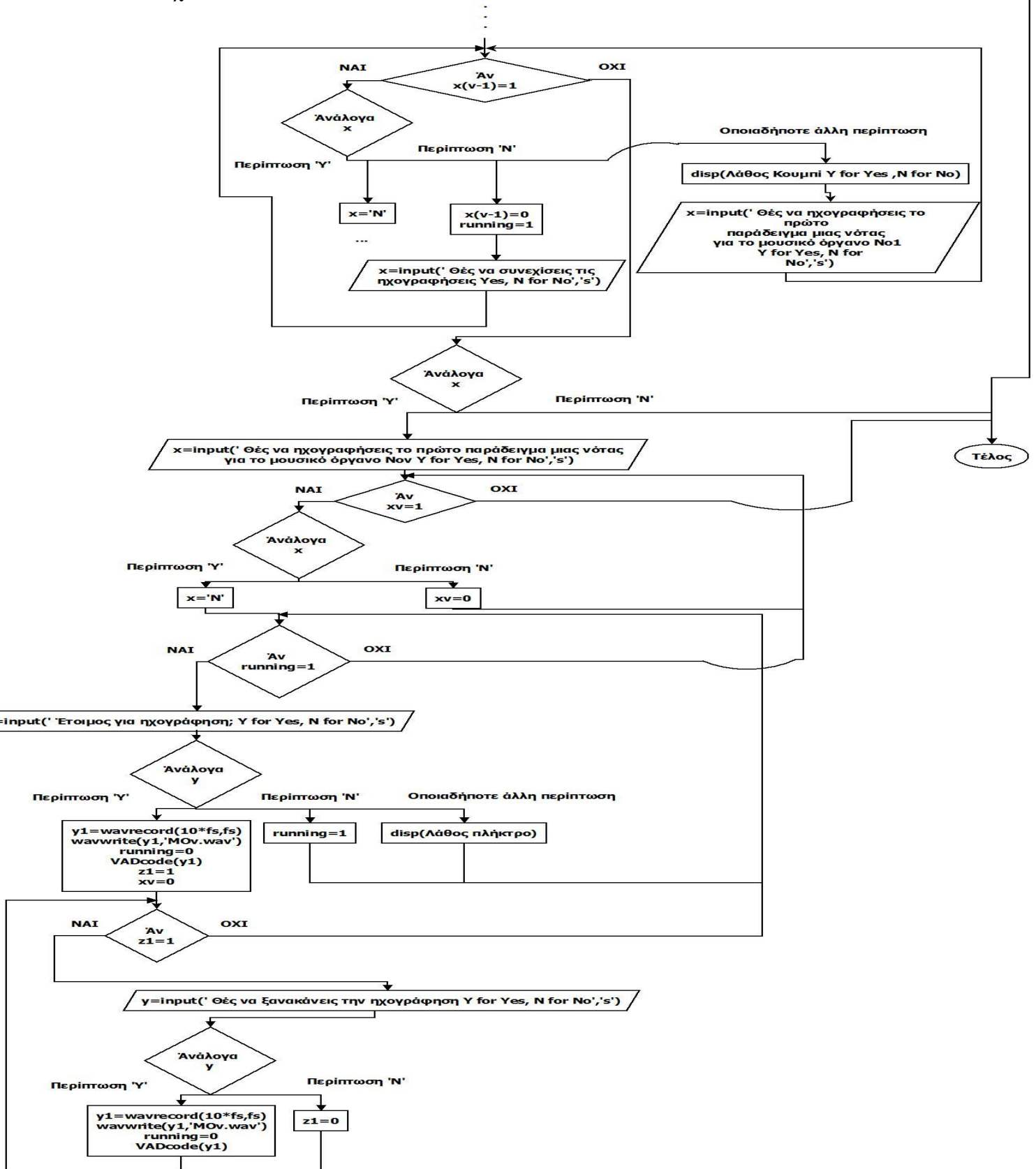
Στην παρακάτω εικόνα, φαίνεται το διάγραμμα ροής της αυτόματης ηχογράφησης και αποθήκευσης των αρχείων ήχου. Τα αρχεία αυτά θα εισαχθούν στον αλγόριθμο της εκπαίδευσης, που φαίνεται στην παραπάνω ενότητα. Η συχνότητα δειγματοληψίας που χρησιμοποιείται στο παρακάτω διάγραμμα ροής είναι 8000 samples/sec. Για λόγους εξοικονόμησης χώρου, το διάγραμμα ροής φαίνεται για την ηχογράφηση και αποθήκευση του πρώτου παραδείγματος νότας για το μουσικό όργανο No1 και του πρώτου παραδείγματος νότας για το μουσικό όργανο No2 και τέλος για το πρώτο παράδειγμα νότας για το μουσικό όργανο No n.



Συνεχίζεται στην επόμενη σελίδα...



Συνέχεια...



Εικόνα 5.2.5. Διάγραμμα ροής αυτόματης ηχογράφησης και αποθήκευσης των αρχείων ήχου.

5.2.4 End-Point Detection εξέταση

Στο End-Point Detection εξέταση, η καταγραφή των δειγμάτων που περιέχονται στα frame γίνονται σε πραγματικό χρόνο(real time) και δεν είναι αποθηκευμένα αρχεία στην μνήμη του Η/Υ, όπως στο End-Point Detection εκπαίδευσης. Στον κώδικα του End-Point Detection εξέταση υπολογίζονται οι τιμές κατωφλίου της ενέργειας και των zero-crossings. Ο υπολογισμός του κατωφλίου ενέργειας είναι πανομοιότυπος με τον τρόπο που υπολογίζεται και στο End-Point Detection εκπαίδευσης(σχέση 5.2.2). Ο υπολογισμός της τιμής κατωφλίου των zero-crossings γίνεται με τον τύπο που φαίνεται παρακάτω:

$$\text{Zero-crossings threshold} = \frac{1}{N} \sum_{i=1}^N \text{ZRC}(i) + a \cdot \frac{1}{N} \sum_{i=1}^N (\text{ZRC}(i) - \bar{x})^2 \quad (5.2.7)$$

ή

$$\text{Zero-crossings threshold} = \text{μέση τιμή(ZRC)} + a \cdot \text{τυπική απόκλιση(ZRC)} \quad (5.2.8)$$

, όπου ZRC είναι ο πίνακας που έχουμε αποθηκεύσει τις τιμές κατωφλίου για τα πρώτα 50 frames, a είναι μια μεταβλητή που η τιμή της δίνεται ανάλογα με τον θόρυβο περιβάλλοντος που επικρατεί στο δωμάτιο ηχογράφησης και x είναι η μέση τιμή των τιμών του πίνακα ZRC, που είναι ίση με:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N \text{ZRC}(i) \quad (5.2.9)$$

Δηλαδή, σε αυτή την περίπτωση υπολογίζουμε και την τιμή κατωφλίου της ενέργειας, αλλά και την τιμή κατωφλίου του πλήθους των zero-crossing για τα 50 πρώτα δείγματα(samples).

Στην συνέχεια, μετά τα 50 frames, στα υπόλοιπα frames συγκρίνουμε την ενέργεια και τα zero-crossings που περιέχουν με τις τιμές threshold της ενέργειας και των zero-crossings, αντίστοιχα. Σε περίπτωση που ικανοποιηθούν οι συνθήκες που εμπεριέχονται στον κώδικα, τότε θα έχει αποθηκευτεί ένας πίνακας, που περιέχει τις τιμές του σήματος που θα γίνουν είσοδο στο κώδικα MFCC, για τον υπολογισμό των συντελεστών MFCC, delta και delta-delta, οι οποίες θα χρησιμοποιηθούν για την διαδικασία της αναγνώρισης. Αν οι τιμές της ενέργειας ή οι τιμές των zero-crossings είναι μεγαλύτερες από τις αντίστοιχες τιμές των κατωφλίων τους, τότε έχει εντοπιστεί κάποιο σήμα, διαφορετικά δεν έχει εντοπιστεί. Στην συνέχεια, πρέπει να ελέγξουμε κατά πόσο αυτό το σήμα που εντοπίστηκε είναι σήμα νότας. Αν έχει εντοπιστεί σήμα νότας, τότε χρησιμοποιούμε τα δεδομένα ως είσοδο στον κώδικα MFCC και εξάγονται τα δεδομένα των MFCC, delta και delta-delta χαρακτηριστικών που θα τα χρησιμοποιήσουμε για τον υπολογισμό της αρνητικής λογαριθμικής πιθανοφάνειας (negative log-likelihood).

Πιο συγκεκριμένα, στο στάδιο της εξέτασης, όπως έχει ήδη αναφερθεί χρησιμοποιούμε κάποια εργαλεία που περιέχονται στο MATLAB. Το πρώτο εργαλείο

που χρησιμοποιείται είναι το Acquisition Toolbox ή εργαλείο καταγραφής. Με αυτό ορίζουμε ότι τα δεδομένα που θα καταγραφούν θα είναι από την κάρτα ήχου του H/Y και ότι θα χρησιμοποιήσουμε συνολικά ένα κανάλι. Επίσης, μπορεί να οριστούν διάφοροι παράμετροι, όπως η συχνότητα δειγματοληψίας, το μέγεθος του πλαισίου(frame) κ.τ.λ. Στην συνέχεια ορίζουμε ότι αρχίζει η καταγραφή. Μέσω μιας εντολής *while* (συγκεκριμένα *while running==1*) αρχίζουμε ένα βρόγχο επανάληψης που τερματίζει όταν η μεταβλητή *running* γίνει ίση με μια τιμή διάφορη του 1($\neq 1$). Μέσα στο βρόγχο επανάληψης χρησιμοποιούμε την εντολή *getdata(AI,(μέγεθος πλαισίου)/2)*.

Στην εντολή αυτή γίνεται η λήψη δεδομένων από την μεταβλητή AI, που έχουμε ορίσει ότι είναι η είσοδος της κάρτας ήχου του υπολογιστή(αναλογική είσοδος). Η δεύτερη παράμετρος (μέγεθος πλαισίου)/2 είναι ότι κάθε φορά που ενεργοποιείται η εντολή *getdata* θα λαμβάνει τόσα δείγματα, όσο θα είναι το αποτέλεσμα (μέγεθος πλαισίου)/2. Στην συγκεκριμένη περίπτωση που το μέγεθος πλαισίου είναι 160 δείγματα, το αποτέλεσμα της παραπάνω παράστασης θα είναι ίσο με 80 δείγματα.

Στην συνέχεια εφαρμόζουμε πλαίσια(frames) στα εισερχόμενα σήματα. Μετά γίνεται χρήση ενός κώδικα End-Point Detection διαφορετικό από το End-Point Detection του σταδίου της εκπαίδευσης. Το EPD εξέτασης καθώς και η όλη διαδικασία της αναγνώρισης φαίνονται στο διάγραμμα ροής της εικόνας 5.2.6, που ακολουθεί.

Αυτό που αναφέρουμε είναι ότι στο EPD εξέτασης μόλις γίνει ο εντοπισμός της νότας, τα δεδομένα που έχουν συλλεχθεί, γίνονται είσοδος στον κώδικα MFCC και εξάγονται τα MFCC, delta και delta-delta χαρακτηριστικά των δεδομένων της εξέτασης. Όπως, έχει αναφερθεί παραπάνω, χρησιμοποιούμε από το εργαλείο Statistics Toolbox ή εργαλείο Στατιστικής του MATLAB την συνάρτηση:

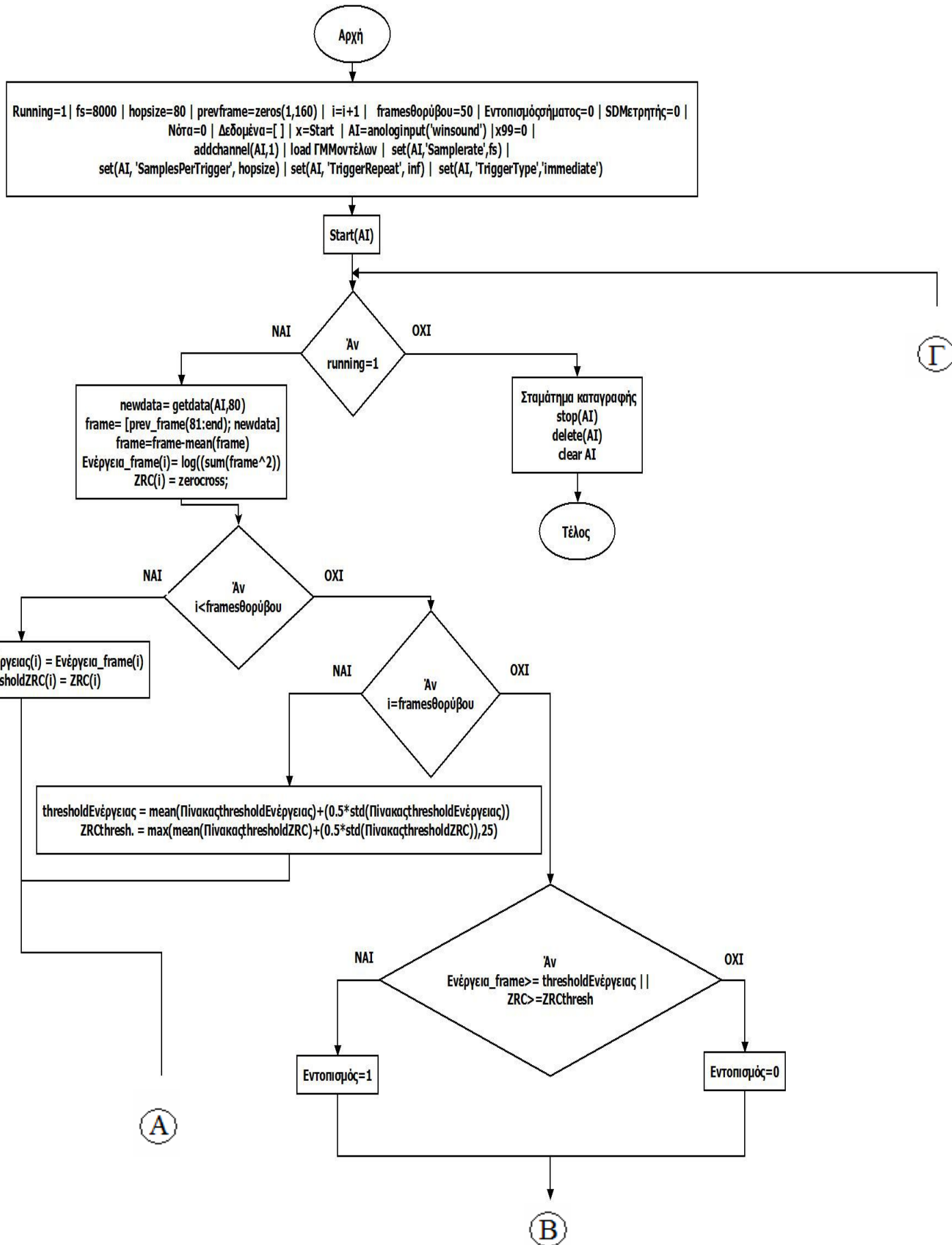
$$[\text{Posterior, log-likelihood}] = \text{posterior}(\text{gmmmodel}(\text{note}), \text{testdata}) \quad (5.2.10)$$

, από την οποία εξάγουμε την εκ των υστέρων πιθανότητα και την αρνητική λογαριθμική πιθανοφάνεια. Τις τιμές της λογαριθμικής πιθανοφάνειας της αποθηκεύουμε σε ένα πίνακα 1X20, όπως είναι και οι θέσεις μνήμης που χρησιμοποιείται από το σύστημα στο στάδιο της εκπαίδευσης για να αποθηκεύσει την μίξη των Γκαουσιανών κατανομών των παραδειγμάτων των νοτών για κάθε μουσικό όργανο. Στην συνέχεια χρησιμοποιούμε μια εντολή, η οποία είναι η εξής:

$$[\text{Negative log-likelihood_Value, Negative log-likelihood_Position}] = \text{min}(\text{Matrix with negative log-likelihood}) \quad (5.2.11)$$

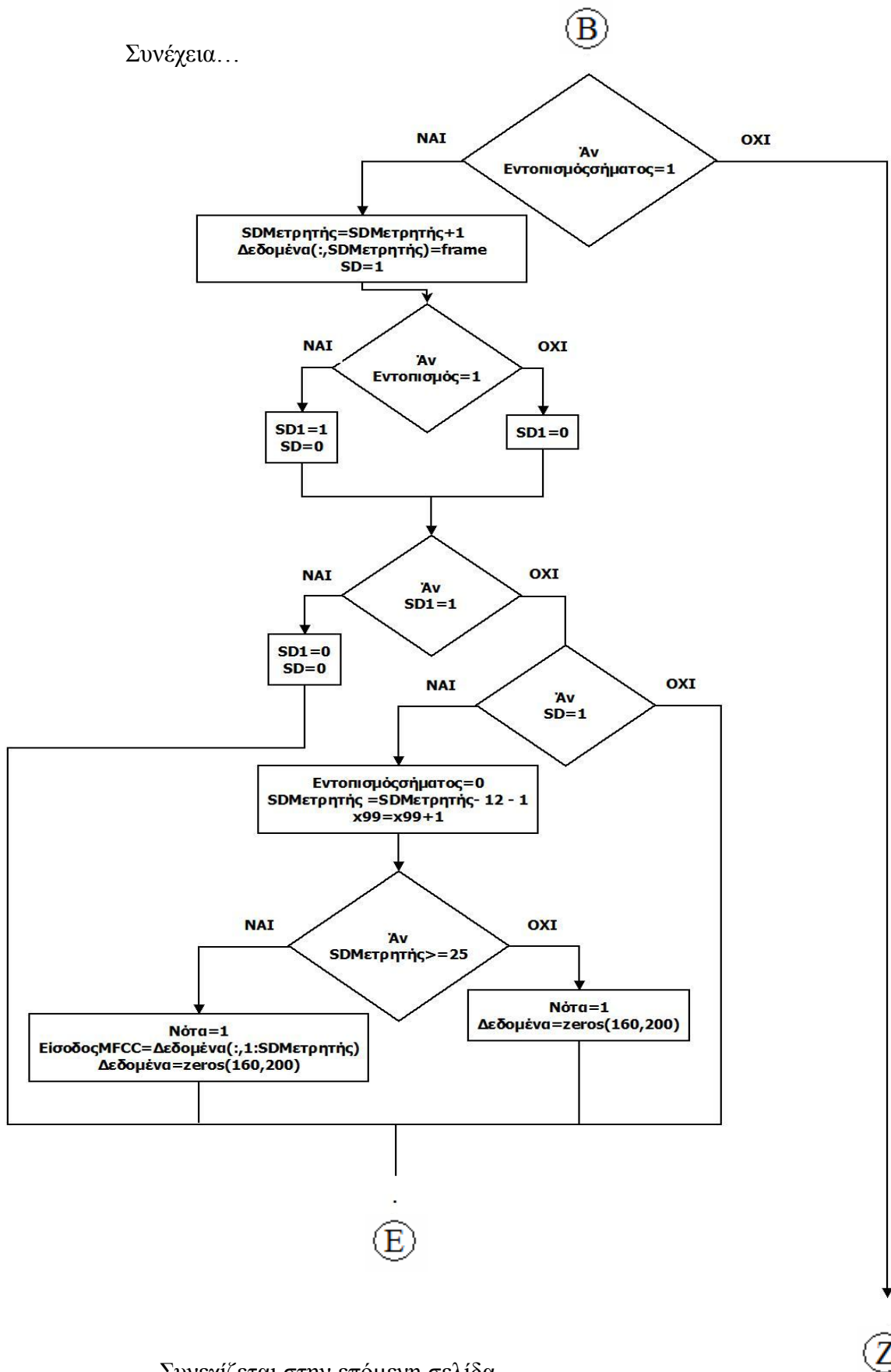
, όπου υπολογίζει την μικρότερη τιμή αρνητικής λογαριθμικής πιθανοφάνειας που αντιστοιχεί στην μέγιστη πιθανοφάνεια(Maximum Likelihood) και την θέση που καταλαμβάνει αυτή στον πίνακα 1X20 που έχουμε αποθηκεύσει τις τιμές των αρνητικών λογαριθμικών πιθανοφανειών. Έτσι, γνωρίζοντας την θέση που κατέχει η μικρότερη τιμή αρνητικής λογαριθμικής πιθανοφάνειας και κατά συνέπεια η μέγιστη πιθανοφάνεια, μπορούμε να αποδώσουμε σε αυτή την θέση κάποια άλλη εντολή, αναλόγως ποιος είναι ο στόχος του συστήματος. Ο στόχος στο συγκεκριμένο σύστημα είναι η αναγνώριση μουσικών οργάνων, άρα χρησιμοποιώντας την τιμή της θέσης που κατέχει η μέγιστη πιθανοφάνεια, χρησιμοποιούμε μια ακόμα συνάρτηση που περιέχει 20 περιπτώσεις και ανάλογα την περίπτωση φορτώνει στο Workspace

του MATLAB μια εικόνα. Με την εντολή **image(imread(Μουσικό όργανο))** την προβάλλει στην οθόνη του υπολογιστή ως γράφημα. Οι πρώτες τέσσερις θέσεις(κατά συνέπεια και περιπτώσεις)του πίνακα τύπου 1X20, αντιστοιχούν στο Πιάνο(1-4), οι επόμενες τέσσερις στην Φυσαρμόνικα(5-8), οι άλλες τέσσερις στο Ξυλόφωνο(10-13), οι επόμενες τέσσερις στην Κιθάρα(13-16) και οι τελευταίες θέσεις στην Φλογέρα(17-20). Έτσι, για παράδειγμα, αν η μικρότερη τιμή αρνητικής λογαριθμικής πιθανοφάνειας βρίσκεται στην θέση 5 του 1X20 πίνακα, τότε θα φορτωθεί στο Workspace του MATLAB η εικόνα της περίπτωσης 5 και μέσω της εντολής **image(imread(Μουσικό όργανο))** θα προβληθεί στην οθόνη του υπολογιστή ως γράφημα η εικόνα της περίπτωσης 5 που θα είναι η εικόνα που περιέχει την λέξη Φυσαρμόνικα. Επίσης, στο σύστημα αυτό υπολογίζεται και εμφανίζεται στην οθόνη του Η/Υ και η κυματομορφή του ηχητικού σήματος. Η λειτουργία του End-Point Detection εξέτασης γίνεται πιο κατανοητή παρατηρώντας το διάγραμμα ροής που απεικονίζεται στην εικόνα που ακολουθεί.



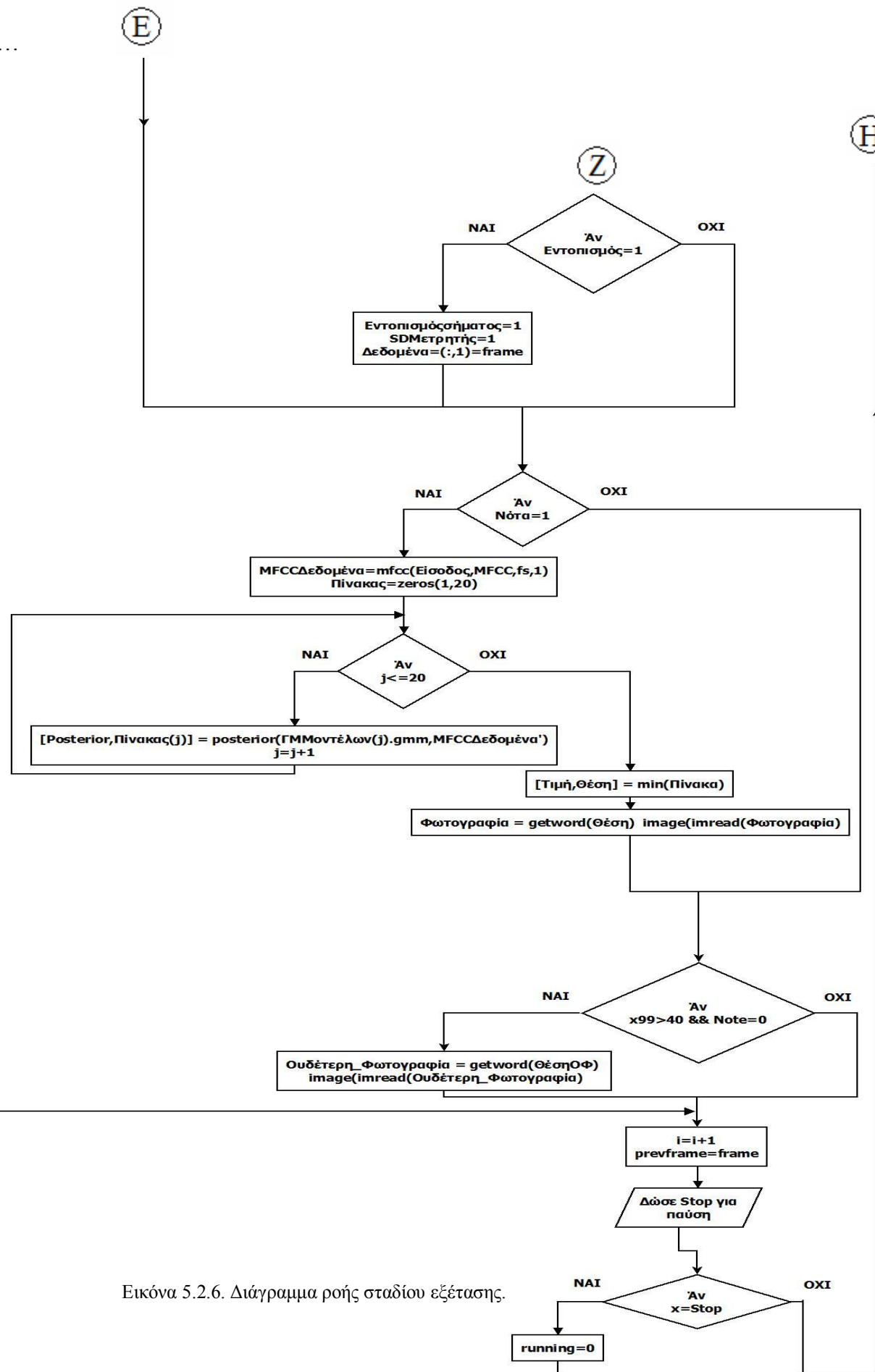
Συνεχίζεται στην επόμενη σελίδα:...

Συνέχεια...



Συνεχίζεται στην επόμενη σελίδα...

Συνέχεια...

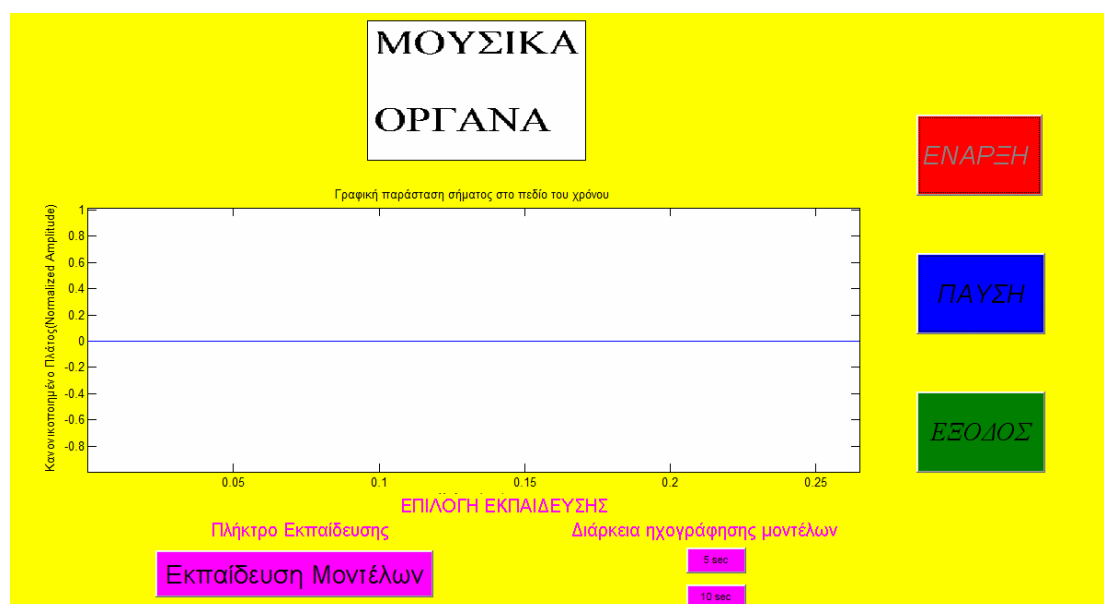


Εικόνα 5.2.6. Διάγραμμα ροής σταδίου εξέτασης.

5.3 Γραφικό Περιβάλλον Διεπαφής (ΓΠΔ) ή Graphical User Interface (GUI)

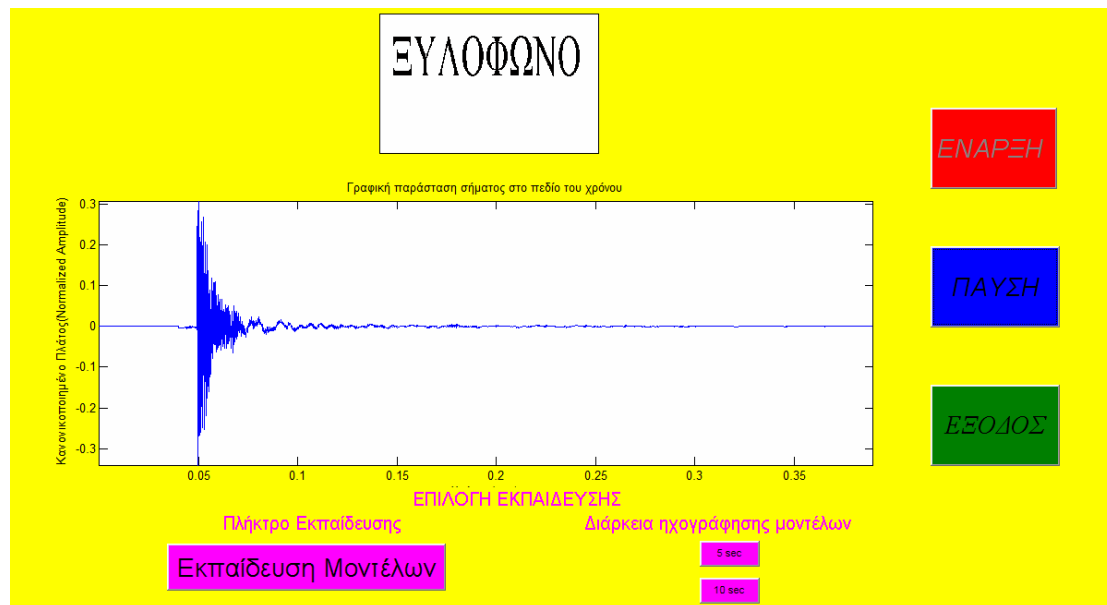
Στις παραπάνω ενότητες, αναφέρθηκαν ξεχωριστά τα στάδια με βάση τα οποία το σύστημα καταλήγει στο πέρας του σκοπού του, που είναι η αναγνώριση μουσικών οργάνων από μεμονωμένες νότες. Για την ευκολία του χρήστη, χρήσιμη είναι η ανάπτυξη ενός γραφικού περιβάλλοντος διεπαφής, όπου ο χρήστης μπορεί να ελέγχει κάποιες παραμέτρους του συστήματος, όπως π.χ. την έναρξη της λειτουργίας του. Για συντομία το γραφικό περιβάλλον διεπαφής θα καταγράφεται ως GUI που είναι τα ακρωνύμια της ξενόγλωσσης ονομασία του, δηλαδή Graphical User Interface.

Η μορφή του GUI φαίνεται στην παρακάτω εικόνα.



Εικόνα 5.3.1. Γραφικό περιβάλλον διεπαφής ή GUI του αυτόματου συστήματος αναγνώρισης μουσικών οργάνων.

Στην παραπάνω εικόνα φαίνεται μια φωτογραφία που αναγράφει την φράση 'Μουσικά Όργανα'. Σε αυτή την θέση θα εμφανίζεται η φωτογραφία που θα αναγράφει το μουσικό όργανο, το οποίο μουσικό όργανο έχει αναγνωρίσει το σύστημα. Στο GUI, εμφανίζεται και μια γραφική παράσταση του σήματος συναρτήσεως του χρόνου. Σε αυτή την γραφική παράσταση φαίνεται η κυματομορφή της νότας του μουσικού οργάνου, το ηχητικό σήμα της οποίας εισάγεται στο σύστημα μέσω ενός μικροφώνου. Στην παρακάτω, εικόνα φαίνεται το σύστημα σε λειτουργία και στην αναγνώριση ενός μουσικού οργάνου.



Εικόνα 5.3.2. Γραφικό περιβάλλον διεπαφής ή GUI του αυτόματου συστήματος αναγνώρισης μουσικών οργάνων σε λειτουργία και η αναγνώριση του ξυλοφώνου από το σύστημα.

Επίσης, στην εικόνα φαίνονται τα πλήκτρα ‘Εναρξη’, ‘Παύση’ και ‘Εξοδος’, που χρησιμεύουν όπως δηλώνει και το όνομα τους στην έναρξη, στην παύση της λειτουργίας του συστήματος και την έξοδο από το σύστημα, αντίστοιχα.

Στο κάτω μέρος του GUI, φαίνονται τα πλήκτρα που αναγράφουν ‘5 sec’ και ‘10 sec’, τα οποία χρησιμεύουν, στην επιλογή της χρονικής διάρκειας της ηχογράφησης των νοτών. Το άλλο πλήκτρο αναγράφει ‘Εκπαίδευση Μοντέλων’ που σκοπό έχει την αυτόματη ηχογράφηση και εκπαίδευση των ηχητικών σημάτων των νοτών από τα όργανα που θέλουμε να αναγνωρίσει το σύστημα. Στην περίπτωση που δεν πατήσουμε πρώτα κάποιο από τα πλήκτρα ‘5 sec’ και ‘10 sec’, τότε η default τιμή της χρονικής διάρκειας της ηχογράφησης των νοτών είναι τα 10 sec. Το πλήκτρο ‘Εκπαίδευση Μοντέλων’, δεν είναι απαραίτητο να χρησιμοποιηθεί, διότι τα ηχογραφημένα αρχεία ήχου των νοτών από κάθε μουσικό όργανο που χρησιμεύουν στην αναγνώριση, είναι αποθηκευμένα στην μνήμη του Η/Υ. Το πλήκτρο αυτό, απλώς δίνει την δυνατότητα στον χρήστη να ξανά-ηχογραφήσει και να ξανά-εκπαιδεύσει τα αρχεία ήχου των νοτών. Αυτό, μπορεί να φανεί χρήσιμο, διότι το σύστημα αυτό είναι εξαρτώμενο από την ηχητική πηγή, που σημαίνει ότι διαφορετικές συνθήκες εξέτασης π.χ. απόκριση χώρου, διαφορετικά όργανα, διαφορετικός Η/Υ, διαφορετικός θόρυβος περιβάλλοντος, από τις συνθήκες που επικρατούσαν την στιγμή της ηχογράφησης και της εκπαίδευσης, μπορεί να μειώσουν αρκετά την απόδοση του συστήματος. Έτσι, αν το σύστημα θέλουμε να το εξετάσουμε σε άλλες συνθήκες, μπορούμε να ηχογραφήσουμε και να εκπαιδεύσουμε τα μουσικά όργανα σε αυτές τις συνθήκες εξέτασης και με αυτό τον τρόπο, να επιτευχθεί υψηλή απόδοση στην αναγνώριση των μουσικών οργάνων από το σύστημα.

ΚΕΦΑΛΑΙΟ ΕΚΤΟ

ΠΕΙΡΑΜΑΤΙΚΑ ΚΑΙ ΣΥΓΚΡΙΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

6.1 Πειραματικά αποτελέσματα

Στην παρούσα πτυχιακή εργασία σκοπός είναι η υλοποίηση ενός συστήματος αυτόματης αναγνώρισης μουσικών οργάνων. Τα μουσικά όργανα που το σύστημα καλείται να αναγνωρίσει είναι η κιθάρα, το ξυλόφωνο, η φλογέρα, το πιάνο και η φουσαρμόνικα. Για κάθε ένα μουσικό όργανο ηχογραφείται και αποθηκεύεται μια νότα(μεμονωμένη που επαναλαμβάνεται μία ή παραπάνω φορές) σε αρχεία ήχου .wav, τέσσερις διαφορετικές φορές(τέσσερα παραδείγματα). Δηλαδή, για κάθε ένα από αυτό τα πέντε όργανα αποθηκεύτηκαν 4 αρχεία ήχου που περιέχουν την ίδια νότα αλλά σε διαφορετική εκδοχή(διαφορετικό 'παίξιμο'). Αυτό σημαίνει, ότι θα ηχογραφηθούν και θα αποθηκευτούν 20 συνολικά παραδείγματα(4 παραδείγματα νοτών για κάθε ένα όργανο) ή αλλιώς 20 αρχεία ήχου.

Υστερα για κάθε μια νότα θα γίνουν τα εξής: Με την χρήση του End-Point Detection εκπαίδευσης, θα γίνει η αναγνώριση της αρχής και του τέλους του σήματος της νότας και το σήμα αυτό θα εξαχθεί(βλέπε 5.2.1). Από εξαγόμενο σήμα από το End-Point Detection εκπαίδευσης, θα γίνει η εξαγωγή των χαρακτηριστικών(feature extraction). Τα χαρακτηριστικά τα οποία θα εξαχθούν από το σήμα της νότας θα είναι τα χαρακτηριστικά MFCC, delta και delta-delta. Στην συνέχεια με βάση αυτά τα χαρακτηριστικά θα υπολογιστεί η Μίξη Γκαουσιανών Μοντέλων(Gaussian Mixture Models) και θα αποδοθεί στα χαρακτηριστικά που εξήχθησαν από το σήμα της νότας.

Επομένως, η νότα, η οποία έχει παραχθεί από ένα από τα πέντε μουσικά όργανα(π.χ. την κιθάρα) θα αντιπροσωπευθεί με αυτή την Μίξη Γκαουσιανών Μοντέλων. Αυτό θα γίνει για όλα τα παραδείγματα των νοτών, δηλαδή σύνολο, θα έχουν αποδοθεί 20 Μίξεις Γκαουσιανών Μοντέλων στα 20 παραδείγματα των νοτών. Όπως, έχει αναφερθεί και στο κεφάλαιο 5^ο, χρησιμοποιήθηκε μια μνήμη 20 θέσεων για την αποθήκευση των Μίξεων Γκαουσιανών Μοντέλων για κάθε ένα από τα 20 παραδείγματα νότας. Έτσι, στην μνήμη των 20 θέσεων είναι αποθηκευμένες 20 Μίξεις Γκαουσιανών Μοντέλων.

Στην εξέταση, μέσω του μικροφώνου καταγράφεται το σήμα από το μουσικό όργανο που θέλουμε να εξετάσουμε(π.χ. κιθάρα), και το σύστημα καλείται να αναγνωρίσει από ποιο όργανο προέρχεται το σήμα. Η αναγνώριση, γίνεται με την βοήθεια των παραπάνω δεδομένων εκπαίδευσης(των Μίξεων Γκαουσιανών Μοντέλων που είναι αποθηκευμένες στην μνήμη των 20 θέσεων).

Σε αυτό το κεφάλαιο θέλουμε να αξιολογήσουμε το σύστημα στην αναγνώριση που επιτυγχάνει για αυτά τα πέντε μουσικά όργανα, όταν το σήμα από κάθε ένα από αυτά εισάγεται στο σύστημα μέσω του μικροφώνου.

Παραθέτουμε τις παραμέτρους που χρησιμοποιήθηκαν τόσο στην εκπαίδευση όσο και στην εξέταση του συστήματος:

Για την εκπαίδευση ισχύει:

1) Ηχογράφιση:

Τα αρχεία ήχου είναι διάρκειας 10 δευτερολέπτων.

Η συχνότητα δειγματοληψίας είναι $fs=16000\text{Hz}$.

2) End-Point Detection εκπαίδευσης:

Μήκος $frame=160$ δείγματα.

Μήκος επικάλυψης= 80 δείγματα.

Πλήθος δειγμάτων ολίσθησης πλαισίου ή $hopsize=80$ δείγματα.

Πλήθος συνιστωσών για την Μίξη Γκαουσιανών Μοντέλων= 8

Μνήμη 20 θέσεων που περιέχουν τις Μίξεις Γκαουσιανών Μοντέλων (39 διαστάσεων) των παραδειγμάτων των νοτών των πέντε μουσικών οργάνων(κιθάρα, ξυλόφωνο, φλογέρα, πιάνο, φουσαρμόνικα).

Οι παράμετροι A και A1 στον κώδικα End-Point Detection είναι ίσες με $A=75$ και $A1=90$.

3) Για να είναι πιο αντικειμενική η αξιολόγηση του συστήματος αναγνώρισης μουσικών οργάνων, ηχογραφήθηκε από όλα τα όργανα(κιθάρα, ξυλόφωνο, φλογέρα, φουσαρμόνικα, πιάνο) η ίδια νότα και συγκεκριμένα η νότα B5(περίπου 987Hz). Στη ίδια νότα θα γίνει και η διαδικασία της εξέτασης.

4) Σε κάθε παράδειγμα νότας ηχογραφήθηκε η ίδια νότα μια ή παραπάνω φορές, για την αύξηση της απόδοσης του συστήματος. Δηλαδή, για το πρώτο παράδειγμα της νότας από ξυλόφωνο ηχογραφήθηκε η ίδια νότα 2 φορές, για το δεύτερο παράδειγμα 3 φορές, για το τρίτο παράδειγμα 2 φορές και για το τέταρτο παράδειγμα ηχογραφήθηκε 3 φορές, ενώ στα τέσσερα παραδείγματα της νότας του πιάνου, ηχογραφήθηκε η νότα από μια φορά.

Για την εξέταση ισχύει:

1) Μήκος $frame=160$ δείγματα.

2) Μήκος επικάλυψης= 80 δείγματα.

3) Πλήθος δειγμάτων ολίσθησης πλαισίου ή $hopsize=80$ δείγματα.

4) Η συχνότητα δειγματοληψίας είναι $fs=16000\text{Hz}$.

5) Θεωρούμε ότι στα πρώτα 50 frames περιλαμβάνεται ο θόρυβος περιβάλλοντος και με βάση τις τιμές που λαμβάνονται δημιουργούνται οι τιμές κατωφλίου τόσο της ενέργειας όσο και των zero-crossing.

6) Απορρίπτονται τα συνεχόμενα frames των οποίων η χρονική διάρκεια είναι μικρότερη των 0.255sec(δηλαδή λαμβάνουμε υπόψη μόνο frames, που το πλήθος τους είναι μεγαλύτερο ή ίσο των 50, $SDMetric_{thts} \geq 50$).

Αυτές είναι οι παράμετροι που θα χρησιμοποιηθούν στο σύστημα αναγνώρισης και τα αποτελέσματα της απόδοσης του συστήματος παρατίθενται στον παρακάτω πίνακα.

Μουσικά όργανα	Πιάνο	Ξυλόφωνο	Φλογέρα	Φυσαρμόνικα	Κιθάρα
Επιτυχής αναγνώριση (επί τοις εκατό %)	86.6%	100%	100%	100%	86.6%
Εσφαλμένη αναγνώριση(όργανα τα οποία αναγνώριζε αντί για τα σωστά)	Κανένα	Κανένα	Κανένα	Κανένα	Φυσαρμόνικα
Συνολική απόδοση συστήματος	94.64%				

Πίνακας 4. Αξιολόγηση του συστήματος.

Τα όργανα εξετάστηκαν από 30 φορές το καθένα και βάσει αυτής της εξέτασης προέκυψε η απόδοση του κάθε οργάνου ξεχωριστά αλλά και η συνολική απόδοση του συστήματος. Η απόδοση του κάθε οργάνου(ξεχωριστά) εξήχθει με βάση την επί τοις εκατό % σχετική συχνότητα επιτυχίας στην αναγνώριση τους από το σύστημα.

Πιο συγκεκριμένα, η σχετική συχνότητα είναι το πλήθος των περιπτώσεων που αναγνωρίστηκε σωστά ένα μουσικό όργανο από το σύστημα προς το σύνολο των περιπτώσεων που αυτό εξετάστηκε. Δηλαδή,

$$fi\% = \frac{\text{Πλήθος των περιπτώσεων της σωστής αναγνώρισης του οργάνου}}{\text{Πλήθος εξετάσεων του οργάνου}}$$

Για παράδειγμα, στην περίπτωση της φλογέρας:

Ο συνολικός αριθμός που εξετάστηκε η κιθάρα είναι ίσος με 30 και το πλήθος σωστής αναγνώρισης της κιθάρας από το σύστημα είναι ίσο με τον αριθμό 28. Κατά συνέπεια η επί τοις εκατό σχετική συχνότητα επιτυχίας στην αναγνώριση της κιθάρας είναι ίση με:

$$fi\% = (28/30) * 100\% = 0.933 * 100\% = 93.3\%$$

Από τον πίνακα 4, προκύπτει ότι η συνολική απόδοση του συστήματος είναι 94.64% που θεωρείται ιδιαίτερα ικανοποιητική.

6.2 Συγκριτικά αποτελέσματα

Στην ενότητα 6.1 αξιολογήσαμε το σύστημα με βάση κάποιες αρχικές συνθήκες στο στάδιο της εκπαίδευσης και στο στάδιο της εξέτασης. Σε αυτή την ενότητα θα αλλάξουμε κάποιες παραμέτρους του συστήματος για να αξιολογήσουμε κατά πόσο μειώνεται ή αυξάνεται η απόδοση του συστήματος αναγνώρισης σε σχέση με το σύστημα που οι παράμετροι του αναφέρονται στην ενότητα 6.1.

6.2.1 Χρονική διάρκεια frame

Η πρώτη παράμετρος που θα μεταβάλλουμε είναι η χρονική διάρκεια του frame, κρατώντας τις άλλες παραμέτρους σταθερές. Πιο συγκεκριμένα θα χρησιμοποιήσουμε frames χρονικής διάρκειας 20ms και 25ms. Οι παράμετροι που ισχύουν και στις δύο περιπτώσεις είναι οι εξής:

Για την εκπαίδευση ισχύει:

1) Ηχογράφηση:

- i) Τα αρχεία ήχου είναι διάρκειας 10 δευτερολέπτων.
- ii) Η συχνότητα δειγματοληψίας είναι $f_s=16000\text{Hz}$.

2) End-Point Detection εκπαίδευσης:

- i) Πλήθος συνιστωσών για την Μίξη Γκαουσιανών Μοντέλων=8
Μνήμη 20 θέσεων που περιέχουν τις Μίξεις Γκαουσιανών Μοντέλων (39 διαστάσεων) των παραδειγμάτων των νοτών των πέντε μουσικών οργάνων(κιθάρα, ξυλόφωνο, φλογέρα, πιάνο, φουσαρμόνικα).

3) Νότα τόσο για την διαδικασία της εκπαίδευσης όσο και κατά την διαδικασία της εξέτασης είναι η νότα B5(περίπου 987Hz).

Για την εξέταση ισχύει:

1) Η συχνότητα δειγματοληψίας είναι $f_s=16000\text{Hz}$.

2) Θεωρούμε ότι στα πρώτα 50 frames περιλαμβάνεται ο θόρυβος περιβάλλοντος και με βάση τις τιμές που λαμβάνονται δημιουργούνται οι τιμές κατωφλίου τόσο της ενέργειας όσο και των zero-crossing.

Οι παράμετροι που θα μεταβληθούν θα είναι το μήκος του frame(framesize), του hopsize και της επικάλυψης(overlapsize).

Τα όργανα εξετάστηκαν από 30 φορές το καθένα και με βάση αυτή την εξέταση προέκυψε η απόδοση του κάθε οργάνου ξεχωριστά αλλά και η συνολική απόδοση του συστήματος.

Τα αποτελέσματα της απόδοσης του συστήματος στην εκάστοτε περίπτωση φαίνονται στους παρακάτω πίνακες.

Frames χρονικής διάρκειας 20ms

Ισχύει και στο στάδιο της εκπαίδευσης και στο στάδιο της εξέτασης:

- i) Μήκος frame=320 δείγματα.
- ii) Μήκος επικάλυψης=160 δείγματα.
- iii) Πλήθος δειγμάτων ολίσθησης πλαισίου ή hopsize=160δείγματα.

Για το στάδιο της εκπαίδευσης ισχύει ότι:

- 1) End-Point Detection εκπαίδευσης:
 - i) $A=90$ και $A1=250$. Οι παράμετροι A και $A1$ είναι κάποιες σταθερές που σχετίζονται με τον αλγόριθμο του End-Point Detection (EPD), οι οποίες καθορίζουν τα άκρα του σήματος, το οποίο και τελικά θα εξαχθεί από αυτόν τον αλγόριθμο.

Για το στάδιο της εξέτασης ισχύει ότι:

- 1) Απορρίπτονται τα συνεχόμενα frames των οποίων η χρονική διάρκεια είναι μικρότερη των 0.26 sec (δηλαδή λαμβάνουμε υπόψιν μόνο frames, που το πλήθος τους είναι μεγαλύτερο ή ίσο των 25, $SDMethrs \geq 25$).

Τα αποτελέσματα για frames χρονικής διάρκειας 20ms, παραθέτονται στον πίνακα 5.

Μουσικά όργανα	Πιάνο	Ξυλόφωνο	Φλογέρα	Φυσαρμόνικα	Κιθάρα
Επιτυχής αναγνώριση (επί τοις εκατό %)	80%	100%	100%	100%	63.3%
Εσφαλμένη αναγνώριση (όργανα τα οποία αναγνώριζε αντί για τα σωστά)	Κιθάρα	Κανένα	Κανένα	Κανένα	Φυσαρμόνικα
Συνολική απόδοση συστήματος	88.66%				

Πίνακας 5. Αξιολόγηση του συστήματος με χρονική διάρκεια frame 20ms.

Όπως, παρατηρούμε και στον παραπάνω πίνακα η απόδοση του συστήματος μειώθηκε από 94.64% στα 88.66%. Η επιτυχία στην αναγνώριση του κάθε οργάνου είναι πολύ μικρότερη από την επιτυχία που είχε το σύστημα που περιγράφεται στην ενότητα 6.1 (frames χρονικής διάρκειας 10ms), εκτός από τις αποδόσεις του ξυλοφώνου, της φλογέρας και της φυσαρμόνικας που παρέμειναν σταθερές (100%). Η μείωση της απόδοσης του συστήματος θεωρείται φυσιολογική, διότι χρησιμοποιεί

μικρότερο πλήθος προς επεξεργασία σε σχέση με το σύστημα στην ενότητα 6.1, και συνεπώς θα μειωνόταν και η απόδοση του.

Frames χρονικής διάρκειας 25ms

Για την εκπαίδευση ισχύει:

1) End-Point Detection εκπαίδευσης:

i) $A=90$ και $A1=310$.

Για την εκπαίδευση και την εξέταση ισχύει:

i) Μήκος frame=400 δείγματα.

ii) Μήκος επικάλυψης=200 δείγματα.

iv) Πλήθος δειγμάτων ολίσθησης πλαισίου ή hopsize=200 δείγματα.

Για το στάδιο της εξέτασης ισχύει ότι:

1) Απορρίπτονται τα συνεχόμενα frames των οποίων η χρονική διάρκεια είναι μικρότερη των 0.25 sec(δηλαδή λαμβάνουμε υπόψη μόνο frames, που το πλήθος τους είναι μεγαλύτερο ή ίσο των 20, $SDMetrhts \geq 20$).

Τα αποτελέσματα της απόδοσης του συστήματος παραθέτονται στον παρακάτω πίνακα.

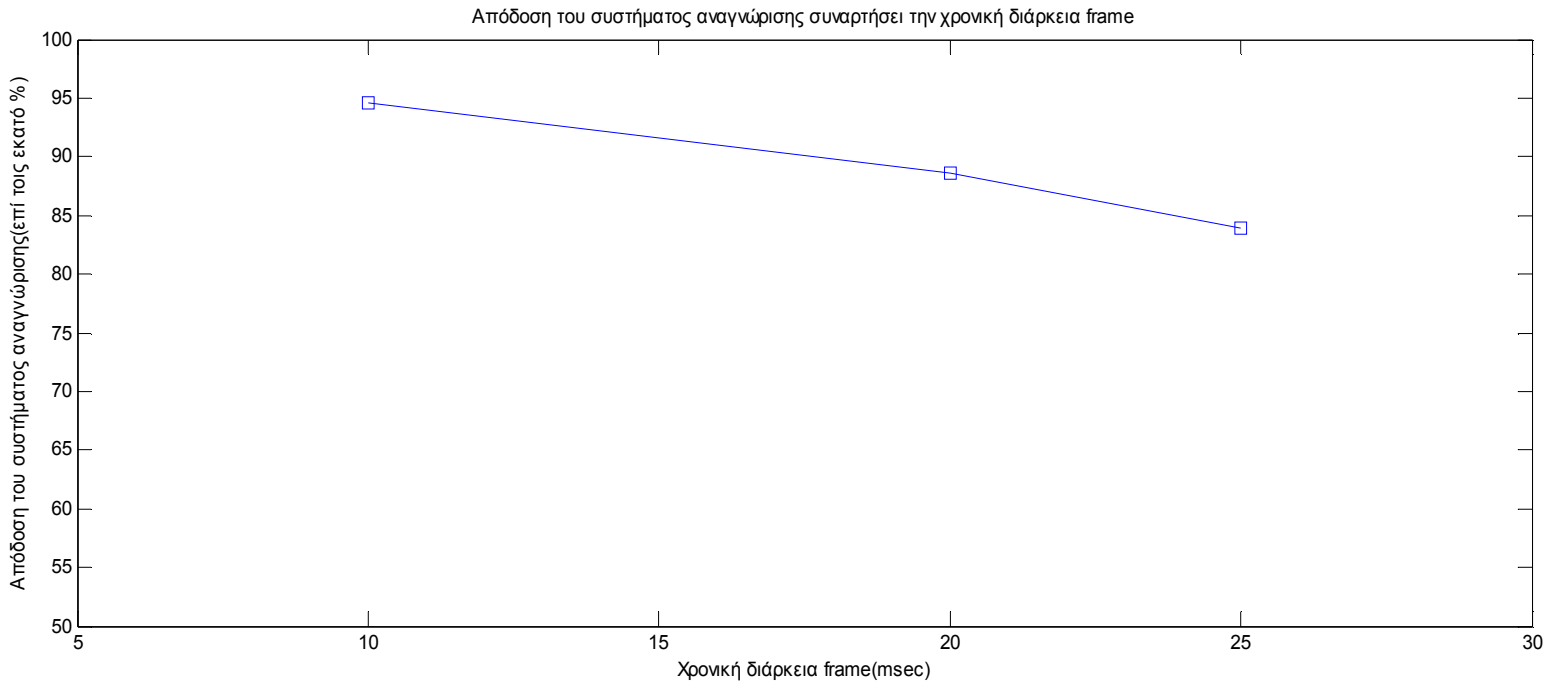
Μουσικά όργανα	Πιάνο	Ξυλόφωνο	Φλογέρα	Φυσαρμόνικ α	Κιθάρα
Επιτυχής αναγνώριση (επί τοις εκατό %)	50%	93.3%	93.3%	100%	83.3%
Εσφαλμένη αναγνώριση(όργανα τα οποία αναγνώριζε αντί για τα σωστά)	Φυσαρμόνικα και Κιθάρα	Φυσαρμόνικα	Φυσαρμόνικα	Κανένα	Φυσαρμόνικα
Συνολική απόδοση συστήματος	83.98%				

Πίνακας 6. Αξιολόγηση του συστήματος με χρονική διάρκεια frame 25ms.

Η απόδοση του συστήματος με frames χρονικής διάρκειας 25ms είναι λίγο χαμηλότερη από την απόδοση του συστήματος με χρονικής διάρκειας 20ms, αλλά και χαμηλότερη σε σχέση με το σύστημα που χρησιμοποιούνται frames χρονικής διάρκειας 10 ms(94.64%). Σε σχέση με το σύστημα με frames 20ms, παρατηρούμε ότι η επιτυχής αναγνώριση μόνο της φυσαρμόνικας παραμένει σταθερή, ενώ στα άλλα μουσικά όργανα μειώνεται αρκετά

με μεγαλύτερη μείωση στο πιάνο(από 80% σε 50%), εκτός από μια αύξηση που παρατηρείται στην κιθάρα(από 63.3% σε 83.3%). Σε σχέση με το σύστημα της ενότητας 6.1, παρατηρούμε ότι η επιτυχής αναγνώριση μόνο της φουσαρμόνικας παραμένει 100%, ενώ στα άλλα μουσικά όργανα μειώνεται αρκετά.

Μια πιο εποπτική εικόνα της απόδοσης του συστήματος σε σχέση με την χρονική διάρκεια των frames φαίνεται στην εικόνα 6.2.2.



Εικόνα 6.2.1. Γραφική παράσταση της απόδοσης του συστήματος αναγνώρισης συναρτήσει της χρονικής διάρκειας frame.

6.2.2 Αριθμός Συνιστωσών της Μίξης Γκαουσιανών Μοντέλων(GMM)

Η δεύτερη παράμετρος που θα μεταβάλλουμε είναι η επιλογή των αριθμών των συνιστωσών της Μίξης Γκαουσιανών Μοντέλων. Στο σύστημα που περιγράφεται στην ενότητα 6.1, ο αριθμός συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M είναι ίσος με $M=8$. Πιο συγκεκριμένα θα μεταβάλλουμε την παράμετρο αυτή με τις τιμές $M=4$, $M=12$, $M=16$, $M=20$ και $M=24$.

Γενικά οι παράμετροι που χρησιμοποιήθηκαν στο σύστημα τόσο κατά την διάρκεια της εκπαίδευσης όσο και κατά την διάρκεια της εξέτασης, παρατίθενται στην συνέχεια.

Για την εκπαίδευση ισχύει:

1) Ηχογράφηση:

Τα αρχεία ήχου είναι διάρκειας 10 δευτερολέπτων.

Η συχνότητα δειγματοληψίας είναι $fs=16000\text{Hz}$.

2) End-Point Detection εκπαίδευσης:

i) $A=75$ και $A1=90$.

ii) Μήκος $frame=160$ δείγματα.

iii) Μήκος επικάλυψης= 80 δείγματα.

iv) Πλήθος δειγμάτων ολίσθησης πλαισίου ή $hopsize=80$ δείγματα.

v) Μνήμη 20 θέσεων που περιέχουν τις Μίξεις Γκαουσιανών Μοντέλων (39 διαστάσεων) των παραδειγμάτων των νοτών των πέντε μουσικών οργάνων (κιθάρα, ξυλόφωνο, φλογέρα, πιάνο, φουσαρμόνικα).

3) Η νότα που χρησιμοποιήθηκε είναι η B5. Στη ίδια νότα θα γίνει και η διαδικασία της εξέτασης.

Για την εξέταση ισχύει:

1) Απορρίπτονται τα συνεχόμενα frames των οποίων η χρονική διάρκεια είναι μικρότερη των 0.255sec (δηλαδή λαμβάνουμε υπόψιν μόνο frames, που το πλήθος τους είναι μεγαλύτερο ή ίσο των 50, $SDMetric \geq 50$).

2) Μήκος $frame=160$ δείγματα.

3) Μήκος επικάλυψης= 80 δείγματα.

4) Πλήθος δειγμάτων ολίσθησης πλαισίου ή $hopsize=80$ δείγματα.

5) Η συχνότητα δειγματοληψίας είναι $fs=16000\text{Hz}$.

6) Θεωρούμε ότι στα πρώτα 50 frames περιλαμβάνεται ο θόρυβος περιβάλλοντος και με βάση τις τιμές που λαμβάνονται δημιουργούνται οι τιμές κατωφλίου τόσο της ενέργειας όσο και των zero-crossing.

Οι παραπάνω παράμετροι παρέμειναν σταθερές καθώς το σύστημα εξετάστηκε για $M=4$, $M=12$, $M=16$, $M=20$ και $M=24$. Επίσης, οι παραπάνω συνθήκες είναι ίδιες για το σύστημα που περιγράφεται στην ενότητα 6.1, όπου $M=8$.

Τα αποτελέσματα της απόδοσης του συστήματος για $M=4$, $M=12$, $M=16$, $M=20$ και $M=24$, παρατίθενται στους πίνακες 6,7,8,9 και 10, αντίστοιχα.

Αναφέρουμε, και ότι, τα όργανα εξετάστηκαν από 30 φορές το καθένα και με βάση αυτή την εξέταση προέκυψε η απόδοση του κάθε οργάνου ξεχωριστά αλλά και η συνολική απόδοση του συστήματος.

Αριθμός Συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=4

Μουσικά όργανα	Πιάνο	Ξυλόφωνο	Φλογέρα	Φυσαρμόνικα	Κιθάρα
Επιτυχής αναγνώριση (επί τοις εκατό %)	93.3%	100%	100%	100%	10%
Εσφαλμένη αναγνώριση(όργανα τα οποία αναγνώριζε αντί για τα σωστά)	Κιθάρα	Κανένα	Κανένα	Κανένα	Πιάνο
Συνολική απόδοση συστήματος	80.66%				

Πίνακας 7. Αξιολόγηση του συστήματος χρησιμοποιώντας αριθμό συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=4.

Γενικά η απόδοση του συστήματος χρησιμοποιώντας αριθμό συνιστωσών M=4, είναι αρκετά μικρή σε σύγκριση με το σύστημα που περιγράφεται στην ενότητα 6.1, το οποίο έχει αριθμό συνιστωσών της Μίξης Γκαουσιανών Μοντέλων ίσο με 8(M=8). Όπως, παρατηρούμε και από τον Πίνακα 7, η επιτυχία της αναγνώρισης της κιθάρας είναι αρκετά χαμηλή, σε σύγκριση με την επιτυχία στην αναγνώριση του συστήματος 6.1.

Αριθμός Συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=12

Μουσικά όργανα	Πιάνο	Ξυλόφωνο	Φλογέρα	Φυσαρμόνικα	Κιθάρα
Επιτυχής αναγνώριση (επί τοις εκατό %)	73.3%	100%	100%	100%	46.6%
Εσφαλμένη αναγνώριση(όργανα τα οποία αναγνώριζε αντί για τα σωστά)	Κιθάρα	Κανένα	Κανένα	Κανένα	Πιάνο
Συνολική απόδοση συστήματος	83.32%				

Πίνακας 8. Αξιολόγηση του συστήματος χρησιμοποιώντας αριθμό συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=12.

Σε αυτή την περίπτωση, η απόδοση του συστήματος χρησιμοποιώντας αριθμό συνιστωσών M=12, είναι αρκετά μικρή σε σύγκριση με το σύστημα που περιγράφεται στην ενότητα 6.1, το οποίο έχει αριθμό συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=8, αλλά και με το σύστημα που περιγράφηκε παραπάνω με αριθμό συνιστωσών ίσο με M=4. Όπως, παρατηρούμε και από τον Πίνακα 8, η επιτυχία της αναγνώρισης της κιθάρας είναι αρκετά χαμηλή.

Αριθμός Συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=16

Μουσικά όργανα	Πιάνο	Ξυλόφωνο	Φλογέρα	Φυσαρμόνικα	Κιθάρα
Επιτυχής αναγνώριση (επί τοις εκατό %)	20%	100%	100%	100%	80%
Εσφαλμένη αναγνώριση(όργανα τα οποία αναγνώριζε αντί για τα σωστά)	Κιθάρα	Κανένα	Κανένα	Κανένα	Φυσαρμόνικα
Συνολική απόδοση συστήματος	80%				

Πίνακας 9. Αξιολόγηση του συστήματος χρησιμοποιώντας αριθμό συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=16.

Από τον πίνακα 9, παρατηρούμε ότι υπάρχει αισθητή μείωση στην απόδοση του συστήματος στην αναγνώριση του πιάνου, όπως και στην συνολική απόδοση του συστήματος.

Αριθμός Συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=20

Μουσικά όργανα	Πιάνο	Ξυλόφωνο	Φλογέρα	Φυσαρμόνικα	Κιθάρα
Επιτυχής αναγνώριση (επί τοις εκατό %)	56.6%	100%	100%	100%	43.3%
Εσφαλμένη αναγνώριση(όργανα τα οποία αναγνώριζε αντί για τα σωστά)	Κιθάρα	Κανένα	Κανένα	Κανένα	Πιάνο και Φυσαρμόνικα
Συνολική απόδοση συστήματος	79.98%				

Πίνακας 10. Αξιολόγηση του συστήματος χρησιμοποιώντας αριθμό συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=20.

Όταν ο αριθμός των συνιστωσών γίνεται ίσος με M=20, η απόδοση του συστήματος είναι αρκετά χαμηλή, και περίπου ίση με την απόδοση του συστήματος για M=16(80%).

Επίσης, στον πίνακα 9, φαίνεται ότι η κιθάρα, έχει επιτυχία αναγνώρισης αρκετά χαμηλή, όπως και το πιάνο.

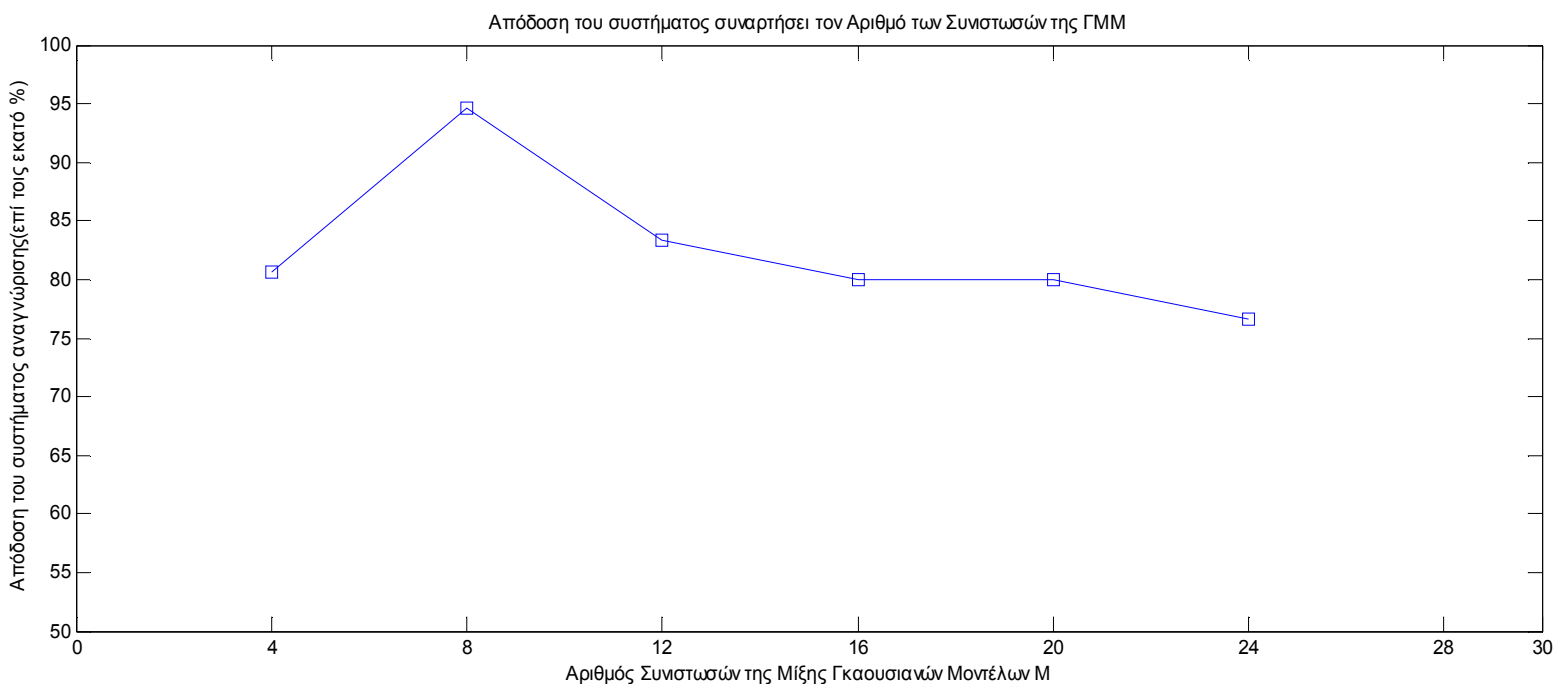
Αριθμός Συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=24

Μουσικά όργανα	Πιάνο	Ξυλόφωνο	Φλογέρα	Φυσαρμόνικα	Κιθάρα
Επιτυχής αναγνώριση (επί τοις εκατό %)	66.6%	100%	100%	100%	16.6%
Εσφαλμένη αναγνώριση(όργανα τα οποία αναγνώριζε αντί για τα σωστά)	Κιθάρα	Κανένα	Κανένα	Κανένα	Πιάνο
Συνολική απόδοση συστήματος	76.64%				

Πίνακας 11. Αξιολόγηση του συστήματος χρησιμοποιώντας αριθμό συνιστωσών της Μίξης Γκαουσιανών Μοντέλων M=24.

Στην περίπτωση που $M=24$, τότε η απόδοση του συστήματος είναι αρκετά ικανοποιητική. Στον πίνακα 10, φαίνεται ότι υπάρχει μείωση της απόδοσης(46.6%) του συστήματος στην αναγνώριση τόσο της κιθάρας όσο και του πιάνου.

Τα αποτελέσματα που φαίνονται στους παραπάνω πίνακες, αναπαρίστανται γραφικά στο παρακάτω διάγραμμα, ώστε να παρουσιαστεί με πιο ευκολία η σχέση μεταξύ του αριθμού των συνιστωσών της Μίξης Γκαουσιανών Μοντέλων και της συνολικής απόδοσης του συστήματος αναγνώρισης.



Εικόνα 6.2.2. Γραφική παράσταση της απόδοσης του συστήματος αναγνώρισης συναρτήσει των αριθμό των Συνιστωσών που χρησιμοποιήθηκαν στην Μίξη Γκαουσιανών Μοντέλων(ΓΜΜ ή GMM).

Στην Εικόνα 6.2.2, παρατηρούμε ότι αύξηση του αριθμού των Μίξεων Γκαουσιανών Μοντέλων, μειώνει τη συνολική απόδοση του συστήματος με το $M=24$ να φτάνει το 76.64%. Εξαιρείται η περίπτωση $M=4$ (80.66%), όπου η απόδοση είναι χαμηλότερη από την περίπτωση $M=8$ (94.64%). Η υψηλότερη επιτυχία στην αναγνώριση πραγματοποιείται στην περίπτωση που ο αριθμός των Μίξεων Γκαουσιανών Μοντέλων είναι ίσος με 8, όπου αγγίζει το 94.64%.

6.2.3 Επιλογή χαρακτηριστικών

Η τρίτη παράμετρος που θα μεταβάλλουμε είναι η επιλογή των χαρακτηριστικών. Πιο συγκεκριμένα μια περίπτωση θα είναι η επιλογή μόνο των MFCC χαρακτηριστικών και στην συνέχεια η επιλογή των MFCC και delta χαρακτηριστικών.

Και στις δύο περιπτώσεις οι παράμετροι έχουν ως εξής:

Για την εκπαίδευση ισχύει:

1) Ηχογράφηση:

- i) Τα αρχεία ήχου είναι διάρκειας 10 δευτερολέπτων.
- ii) Η συχνότητα δειγματοληψίας είναι $fs=16000\text{Hz}$.

2) End-Point Detection εκπαίδευσης και εκπαίδευση:

- i) $A=75$ και $A1=90$
 - ii) Μήκος frame=160 δείγματα.
 - iii) Μήκος επικάλυψης=80 δείγματα.
 - iv) Πλήθος δειγμάτων ολίσθησης πλαισίου ή hopsize=80 δείγματα.
 - v) Πλήθος συνιστωσών για την Μίξη Γκαουσιανών Μοντέλων=8
- 3) Η νότα που χρησιμοποιήθηκε τόσο στην εκπαίδευση όσο και στην εξέταση είναι η B5(περίπου 987Hz).

Για την εξέταση ισχύει:

1) Απορρίπτονται τα συνεχόμενα frames των οποίων η χρονική διάρκεια είναι μικρότερη των 0.255sec(δηλαδή πλήθος frames μικρότερο ή ίσο των 50, $SDMetric \geq 50$).

- 2) Μήκος frame=160 δείγματα.
- 3) Μήκος επικάλυψης=80 δείγματα.
- 4) Πλήθος δειγμάτων ολίσθησης πλαισίου ή hopsize=80 δείγματα.
- 5) Η συχνότητα δειγματοληψίας είναι $fs=16000\text{Hz}$.

Θεωρούμε ότι στα πρώτα 50 frames περιλαμβάνεται ο θόρυβος περιβάλλοντος και με βάση τις τιμές που λαμβάνονται δημιουργείται οι τιμές κατωφλίου τόσο της ενέργειας όσο και των zero-crossing.

Αυτές είναι οι παράμετροι που θα χρησιμοποιηθούν στο σύστημα αναγνώρισης και τα αποτελέσματα της απόδοσης του συστήματος φαίνονται στους παρακάτω πίνακες.

Τα όργανα εξετάστηκαν από 30 φορές το καθένα και βάσει αυτής της εξέτασης προέκυψε η απόδοση του κάθε οργάνου ξεχωριστά αλλά και η συνολική απόδοση του συστήματος.

Χρήση μόνο MFCC χαρακτηριστικών

Στην περίπτωση που γίνεται χρήση μόνο των MFCC χαρακτηριστικών, τότε για την διαδικασία της εκπαίδευσης ισχύει ότι θα χρησιμοποιηθεί μια μνήμη 20 θέσεων που περιέχουν τις Μίξεις Γκαουσιανών Μοντέλων (13 διαστάσεων) των παραδειγμάτων των νοτών των πέντε μουσικών οργάνων(κιθάρα, ξυλόφωνο, φλογέρα, πιάνο, φουσαρμόνικα).

Μουσικά όργανα	Πιάνο	Ξυλόφωνο	Φλογέρα	Φυσαρμόνικα	Κιθάρα
Επιτυχής αναγνώριση (επί τοις εκατό %)	70%	100%	100%	100%	86.6%
Εσφαλμένη αναγνώριση(όργανα τα οποία αναγνώριζε αντί για τα σωστά)	Κιθάρα	Κανένα	Κανένα	Κανένα	Φυσαρμόνικα
Συνολική απόδοση συστήματος	91.32%				

Πίνακας 12. Αξιολόγηση του συστήματος χρησιμοποιώντας μόνο τα MFCC χαρακτηριστικά.

Μόνο η χρήση των MFCC χαρακτηριστικών, δεν είναι ικανή να δώσει ικανοποιητική απόδοση στο σύστημα αναγνώρισης, όπως φαίνεται και στον πίνακα 12.

Χρήση MFCC και delta χαρακτηριστικών

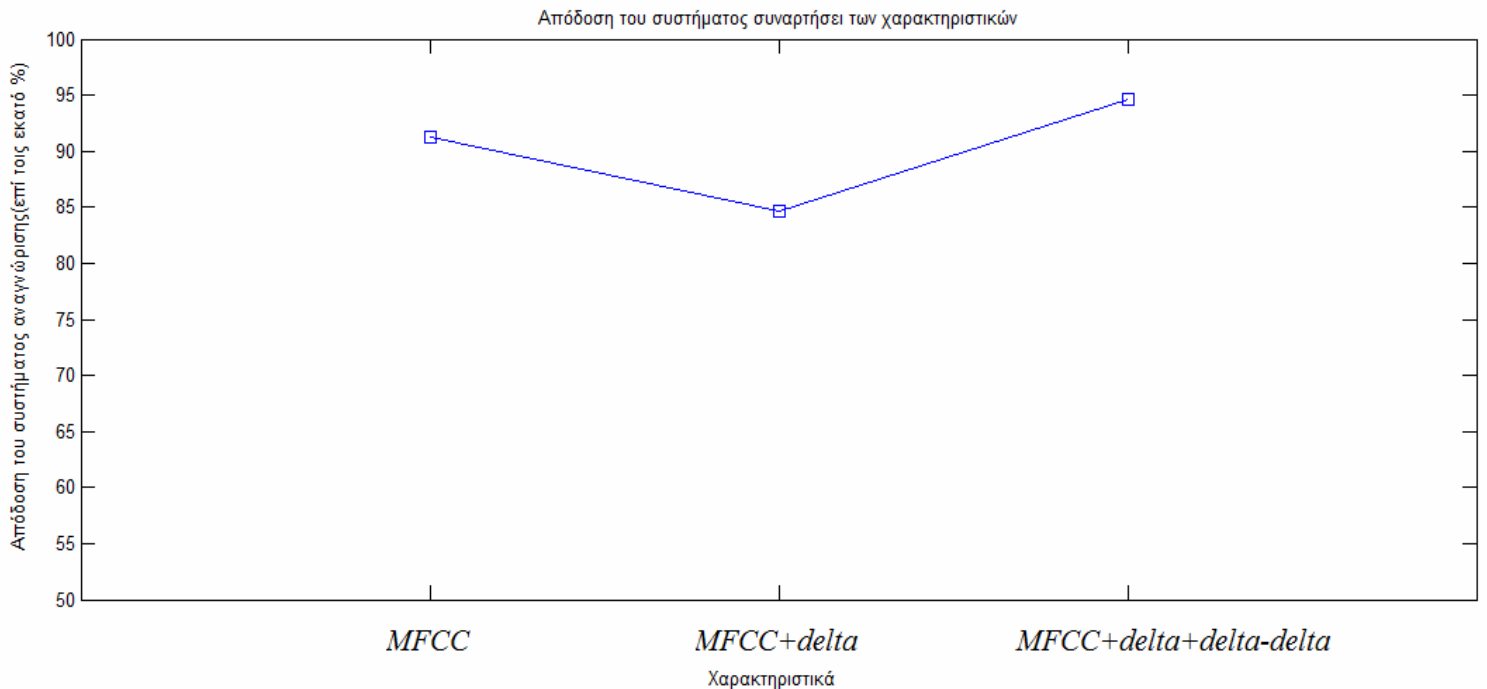
Στην περίπτωση που γίνεται χρήση των MFCC και delta χαρακτηριστικών, τότε για την διαδικασία της εκπαίδευσης ισχύει ότι θα χρησιμοποιηθεί μια μνήμη 20 θέσεων που περιέχουν τις Μίξεις Γκαουσιανών Μοντέλων (26 διαστάσεων) των παραδειγμάτων των νοτών των πέντε μουσικών οργάνων(κιθάρα, ξυλόφωνο, φλογέρα, πιάνο, φυσαρμόνικα). Τα αποτελέσματα φαίνονται στον πίνακα 12.

Μουσικά όργανα	Πιάνο	Ξυλόφωνο	Φλογέρα	Φυσαρμόνικα	Κιθάρα
Επιτυχής αναγνώριση (επί τοις εκατό %)	76.6%	100%	100%	100%	46.6%
Εσφαλμένη αναγνώριση(όργανα τα οποία αναγνώριζε αντί για τα σωστά)	Κιθάρα	Κανένα	Κανένα	Κανένα	Φυσαρμόνικα
Συνολική απόδοση συστήματος	84.64%				

Πίνακας 13. Αξιολόγηση του συστήματος χρησιμοποιώντας τα MFCC και delta χαρακτηριστικά.

Χρησιμοποιώντας μαζί με τα MFCC και τα delta χαρακτηριστικά, τότε η απόδοση του συστήματος αυξάνεται, από το σύστημα που χρησιμοποιεί μόνο τα MFCC χαρακτηριστικά.

Στην εικόνα 6.2.3, παριστάνεται γραφικά η απόδοση του συστήματος αναγνώρισης μουσικών οργάνων συναρτήσει της επιλογής χαρακτηριστικών. Η εικόνα 6.2.3 αναπαριστά το σύστημα που περιγράφεται στην ενότητα 6.1, που χρησιμοποιεί MFCC, delta και delta-delta χαρακτηριστικά.



Εικόνα 6.2.3. Γραφική παράσταση της απόδοσης του συστήματος αναγνώρισης συναρτήσει των χαρακτηριστικών(features).

Από την εικόνα 6.2.3 παρατηρούμε ότι χρησιμοποιώντας τον μεγαλύτερο συνδυασμό χαρακτηριστικών τότε, η απόδοση του συστήματος αυξάνεται. Η χαμηλότερη απόδοση, παρατηρείται χρησιμοποιώντας τα χαρακτηριστικά MFCC και delta(84.64%), αν και το αποτέλεσμα δεν είναι ανάλογο του αναμενόμενου(δηλαδή, με περισσότερα χαρακτηριστικά αναμέναμε επίτευξη υψηλότερης απόδοσης).

6.3 Συμπεράσματα

Από τα συγκριτικά αποτελέσματα που παρουσιάστηκαν παραπάνω, μπορούμε να εξάγουμε το συμπέρασμα ότι η απόδοση του συστήματος δεν παραμένει σταθερή αν μεταβάλλουμε κάποιες παραμέτρους. Το ίδιο συμβαίνει και με την απόδοση του συστήματος στην αναγνώριση κάθε ενός οργάνου ξεχωριστά. Επίσης, μπορούμε να επισημάνουμε ότι εάν χρησιμοποιηθεί παράθυρο μικρής χρονικής διάρκειας περίπου 10ms, αριθμός συνιστωσών της Μίξης Γκαουσιανών Μοντέλων(ΓΜΜ) $M=8$ και ο συνδυασμός των MFCC, delta και delta-delta χαρακτηριστικών, τότε η απόδοση του συστήματος είναι εξαιρετικά υψηλή και ικανοποιητική, που η τιμή της σε ποσοστό επί τοις εκατό % είναι ίση με 94.64% και είναι η απόδοση του συστήματος, το οποίο περιγράφηκε στην ενότητα 6.1. Τέλος, και κάποιοι άλλοι παράγοντες παίζουν καθοριστικό ρόλο στην σωστή αναγνώριση, οι οποίοι συνοπτικά είναι οι εξής:

Ο πρώτος παράγοντας είναι το ηχητικό περιβάλλον κατά την διάρκεια της εκπαίδευσης. Για να υπάρχουν καλύτερα αποτελέσματα στην απόδοση του συστήματος, ορθό είναι τα δεδομένα της εκπαίδευσης να ηχογραφηθούν στον ίδιο χώρο, στον οποίο θα γίνει και η εξέταση. Αν οι ακουστικές συνθήκες είναι παραπλήσιες τόσο κατά την διάρκεια της εκπαίδευσης όσο και κατά την διάρκεια της εξέτασης, τότε θα εξαχθούν καλύτερα αποτελέσματα.

Ένας άλλος παράγοντας είναι το μέγεθος των δεδομένων εκπαίδευσης. Μεγαλύτερος όγκος δεδομένων σημαίνει καλύτερα αποτελέσματα για το σύστημα αναγνώρισης. Στο συγκεκριμένο σύστημα ηχογραφήσαμε 4 διαφορετικές εκδοχές(παραδείγματα) της νότας για κάθε όργανο. Σε κάθε ένα από αυτά τα τέσσερα διαφορετικά παραδείγματα περιέχονται αρκετές επαναλήψεις της ίδιας νότας, λόγω του ότι κατά αυτόν τον τρόπο αυξάνεται η απόδοση του συστήματος.

Ο τρίτος παράγοντας είναι ότι το σύστημα είναι εξαρτώμενο από την ηχητική πηγή, πράγμα που σημαίνει ότι αν τα όργανα που θα εξεταστούν είναι διαφορετικά από τα όργανα που ηχογραφούνται και εκπαιδεύονται, τότε η απόδοση του συστήματος γίνεται αρκετά χαμηλή. Επίσης, αν παιχτεί μια νότα με διαφορετικό τρόπο στο στάδιο της εξέτασης, από το στάδιο της ηχογράφησης και εκπαίδευσης, τότε θα επηρεάσει αυτή η συνθήκη τα αποτελέσματα του συστήματος αναγνώρισης. Γι' αυτό είναι προτιμότερο να ηχογραφούνται και να εκπαιδεύονται περισσότερες επαναλήψεις της ίδιας νότας, αλλά τα παιχνίματα της να ποικίλουν, έτσι ώστε και να παιχθεί η νότα στο στάδιο της εξέτασης με ένα διαφορετικό τρόπο, να έχουν αυξηθεί οι πιθανότητες να ταξινομηθεί και κατά συνέπεια να αναγνωριστεί σωστά το μουσικό όργανο, από το οποίο προέρχεται το σήμα της νότας. Επίσης, αν χρησιμοποιηθεί μικρόφωνο με άλλη συχνοτική απόκριση στην εξέταση από το στάδιο της εκπαίδευσης, τότε και αυτή η συνθήκη θα επηρεάσει το τελικό αποτέλεσμα της αναγνώρισης. Επομένως, μπορούμε να συμπεράνουμε ότι, πρέπει οι ακουστικές συνθήκες(του χώρου, του μικροφώνου, του H/Y, του μουσικού οργάνου και ότι άλλη συνθήκη μπορεί να επηρεάσει το αποτέλεσμα της αναγνώρισης) να είναι όσο το δυνατόν παραπλήσιες τόσο κατά την διάρκεια της εκπαίδευσης όσο και κατά την διάρκεια της εξέτασης.

Τέλος, ο τέταρτος παράγοντας που μπορεί να επηρεάσει το αποτέλεσμα της αναγνώρισης και κατά συνέπεια την απόδοση του συστήματος είναι ο περιβάλλοντας χώρος της εξέτασης. Στην περίπτωση που εξωτερικές ηχητικές πηγές υπάρχουν στον χώρο και η στάθμη του θορύβου περιβάλλοντος είναι αρκετά υψηλή, τότε αυτή η συνθήκη επηρεάζει αρνητικά την απόδοση του συστήματος, και το πρόβλημα αυτό λύνεται με την χρήση κάποιων τεχνικών όπως η ηχοακύρωση θορύβου(noise cancellation), ώστε να επιτευχθεί υψηλή τιμή Signal to Noise Ratio(SNR) ή σήματος προς θόρυβο και να βελτιώσει την απόδοση του συστήματος αναγνώρισης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Σκαρλάτος Δημήτρης, 'Εφαρμοσμένη Ακουστική', Εκδόσεις Φιλομάθεια, 2005, ISBN: 960-87710-1-3.
- [2] Αβέρωφ Έφη, 'Εισαγωγή στην οργανογνωσία', Εκδόσεις Μουσικός Οίκος Φίλιππος Νάκας, 1992, ISBN: 960-290-642-1.
- [3] Κουζούπης Σπύρος, Σημειώσεις Θεωρίας 'Εφαρμοσμένη Ακουστική 3', ΤΕΙ Κρήτης-Παράρτημα Ρεθύμνου- Τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Ρέθυμνο.
- [4] Δρ. Νεκτάριος Παπαδογιάννης, Σημειώσεις Θεωρίας 'Μουσική Ακουστική και Οργανολογία' εν ονόματι 'Φυσική Ακουστική Μουσικών Οργάνων, ΤΕΙ Κρήτης-Παράρτημα Ρεθύμνου- Τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Ρέθυμνο.
- [5] Ταταράκης Μιχάλης, Σημειώσεις Θεωρίας 'Μουσική Ακουστική και Οργανολογία' εν ονόματι 'Φυσική Ακουστική Μουσικών Οργάνων, ΤΕΙ Κρήτης-Παράρτημα Ρεθύμνου- Τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Ρέθυμνο.
- [6] Ελευθερίου Μαρία, Σημειώσεις Εργαστηρίου 'Μουσικής Ακουστικής και Οργανολογίας', ΤΕΙ Κρήτης-Παράρτημα Ρεθύμνου- Τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Ρέθυμνο.
- [7] Lapp R. David, 'The physics of music and musical instruments', Wright Center for innovative science education Tufts university, Medford Massachusetts
- [8] Martin Dana Keith, 'Sound-Source Recognition: A Theory and Computational Model'. PhD Thesis, MIT, June 1999.
- [9] Δρ. Στρουθόπουλος Π. Χαράλαμπος, E-book 'Αναγνώριση Προτύπων-Νευρωνικά Δίκτυα', ΤΕΙ Σερρών,-Σχολή Τεχνολογικών Εφαρμογών-Τμήμα Πληροφορικής και Επικοινωνιών, Σέρρες, Ιούλιος 2008.
- [10] Δρ. Ποταμίτης Ηλίας και Παπαδάκης Νίκος, 'Σημειώσεις Εργαστηρίου Ψηφιακής Επεξεργασίας Σήματος', ΤΕΙ Κρήτης-Παράρτημα Ρεθύμνου-Τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Ρέθυμνο.
- [11] Hayes H. Monson, 'Ψηφιακή Επεξεργασία Σήματος', Εκδόσεις Τζιόλα, Θεσσαλονίκη, 2000, ISBN:960-8050-11-1.
- [12] Φιτσανάκης Μίνως, σημειώσεις Θεωρίας 'Mastering', ΤΕΙ Κρήτης-Παράρτημα Ρεθύμνου- Τμήμα Μουσικής Τεχνολογίας και Ακουστικής, Ρέθυμνο.
- [13] Sanders Mark, 'Τεχνολογία Επικοινωνιών', Εκδόσεις Ίδρυμα Ευγενίδου, Δεύτερη Έκδοση, Αθήνα, 2005, ISBN:960-337-026-6.
- [14] Αδαμόπουλος Λ., Δαμιανού Χ. και Σβέρκος Α., 'Μαθηματικά και Στοιχεία Στατιστικής', ΟΕΔΒ, 8^η Έκδοση, Μάρτιος 1999.
- [15] Διαμαντόπουλος Ταξιάρχης, 'Προγραμματισμός και Σύνθεση Ήχου', Εκδόσεις Ελλην, 2004, ISBN: 960-286-851-1.
- [16] Reynolds, Rose. (1995). "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models". IEEE Transactions on Speech and Audio Processing, Vol. 3, No. 1, January 1995.
- [17] Στρίντζης Μιχάλη, 'Αναγνώριση Προτύπων' Εκδοτικός οίκος Αδελφών Κυριακίδη Α.Ε., 1999, ISBN:978-960-343-290-6.
- [18] Olson F. Harry, 'Music, Physics and Engineering', Dover Publications Inc., Second Edition, 1967, ISBN:486-21769-8.
- [19] Fletcher N. and Rossing T., 'The Physics of Musical Instruments', Springer, Second Edition, 1998, ISBN:0-387-98374-0.

- [20] Rossing T., 'Science of Percussion Instruments', World Scientific, 2000, Series in popular science- Volume 3, ISBN:978-9810241582.
- [21] Smith W. Steven, 'The Scientist and Engineer's Guide to Digital Signal Processing', California Technical Publishing, 1997, ISBN:0-9660176-3-3.
- [22] Eronen, "Automatic Musical Instrument Recognition" Master of Science Thesis, Tampere University of Technology-Department of Information Technology, Tampere, October 2001.
- [23] http://www.physics.ntua.gr/~mmakro/index_files/Kef10_Hxos_Omilia_red.pdf
- [24] <http://neural.cs.nthu.edu.tw/jang/books/audioSignalProcessing/>
- [25] http://www.icsd.aegean.gr/lecturers/kavallieratou/PattRec_files/pr_2.pdf
- [26] Φακωτάκης Νίκος, Σημειώσεις 'Τεχνολογία Ομιλίας', Πανεπιστήμιο Πατρών-Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών, Πάτρα 2005.
- [27] Γκιόκας Α. Άγγελος, 'Αυτόματη Αναγνώριση Μουσικών Οργάνων', Διπλωματική Εργασία, ΕΜΠ-Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών-Τομέας Επικοινωνιών Ηλεκτρονικής και Συστημάτων Πληροφορικής, Αθήνα, Νοέμβριος 2006.
- [28] Βασίλειος-Φοίβος Α. Μανιατάκος, 'Αναγνώριση ηχητικής πηγής-Εφαρμογή αναγνώρισης παραδοσιακών μουσικών οργάνων', ΕΜΠ-Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών-Τομέας Επικοινωνιών Ηλεκτρονικής και Συστημάτων Πληροφορικής, Αθήνα, Σεπτέμβριος 2006.
- [29] Karris T. Steven, 'Signals and Systems' Orchard Publications, Second Edition, 2003, ISBN:0-9709511-8-3.
- [30] Τσακαλίδης Π., 'Αναγνώριση Προτύπων(Pattern Recognition)-Μπεϋζιανή Θεωρία Αποφάσεων(Bayesian Decision Theory), Σημειώσεις από διάλεξη, Πανεπιστήμιο Κρήτης-Τμήμα Επιστήμης Υπολογιστών.
- [31] <http://hyperphysics.phy-astr.gsu.edu/hbase/hframe.html>
- [32] <http://www.bluzkat.com/harphist.htm>
- [33] <http://www.suite101.com/content/teaching-recorders-historical-background-a327801>
- [34] <http://www.cs.tut.fi/sgn/arg/klap/classification.pdf>
- [35] http://www.speech.kth.se/music/5_lectures/introd/introd.html
- [36] <http://www.phys.unsw.edu.au/jw/fluteacoustics.html>
- [37] Whittaker Jed, 'The physics of ear', Σημειώσεις(Lectures Notes), December 2005, pptfile:
http://www.colorado.edu/physics/phys1240/phys1240_fa07/images/JedWhittaker%20on%20the%20ear.ppt
- [38] http://scienceblogs.com/startswithabang/2010/05/dont_you_hear_that.php
- [39] Dr. Matteson E. Samuel, 'The Science and Technology of Musical Sound', Σημειώσεις(Lecture Notes), pptfile:
<http://www.math.unt.edu/~matteson/1251-001/Schd1251.html>
- [40] Ning Daryl, 'Developing an Isolated Word Recognition System in MATLAB', The Mathworks Inc., 2009.
- [41] <http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/>
- [42] <http://www.freewebs.com/lvtaoran/>
- [43] http://caspar.bgsu.edu/~courses/stats/Lectures/Lect_LogLikelihood.shtml
- [44] Peil U., 'Wind and Music', Technische Universität Braunschweig
- [45] <http://www.acousticmasters.com/fanbrace.gif>
- [46] <http://www.freewebs.com/lvtaoran/>
- [47] <http://lear.inrialpes.fr/RecogWorkshop08/documents/panel-lecun-ponce-hebert.pdf>

- [48] <http://library.thinkquest.org/C005705/English/sound/history.htm>
- [49] <http://magnmat.physics.auth.gr/documents/pasxou-resta.pdf>
- [50] <http://www.ideafinder.com/history/inventors/tesla.htm>
- [51] http://www.dragon-medical-transcription.com/history_speech_recognition.html

ΠΑΡΑΡΤΗΜΑ Α

Σε αυτό το παράρτημα θα παρουσιαστούν οι κώδικες που χρησιμοποιήθηκαν για την υλοποίηση του συστήματος αυτόματης αναγνώρισης μουσικών οργάνων. Οι κώδικες, όπως έχει αναφερθεί και στο κεφάλαιο 5^ο είναι υλοποιημένοι σε πρόγραμμα 'MATLAB', έτσι η λειτουργία του συστήματος γίνεται μέσω της γλώσσας προγραμματισμού 'MATLAB'.

Στην αρχή παρουσιάζονται οι κώδικες που αποτελούν το στάδιο της εκπαίδευσης, ενώ στο τέλος παρουσιάζονται οι κώδικες που αποτελούν το στάδιο της εξέτασης ή αναγνώρισης.

Κώδικες εκπαίδευσης

Κώδικας αυτόματης ηχογράφησης, αποθήκευσης και εκπαίδευσης(Doktryprogramma)

Ο κώδικας που φαίνεται παρακάτω, χρησιμοποιείται για την αυτόματη ηχογράφηση, αποθήκευση και εκπαίδευση των μοντέλων των νοτών για κάθε ένα όργανο που χρησιμοποιούμε. Ηχογραφούμε 4 διαφορετικές εκδοχές της νότας για κάθε ένα όργανο, έτσι δημιουργούμε 20 θέσεις μνήμης(λόγω του ότι θέλουμε να εξετάσουμε το σύστημα για 5 μουσικά όργανα, δηλαδή $4 \times 5 = 20$ θέσεις μνήμης).

```
% Χρησιμοποιείται για να μπορεί να επιλέξει ο χρήστης ποια μοντέλα  
θέλει να  
% ηχογραφήσει, να αποθηκεύσει και να εκπαιδεύσει  
running=1;  
fs=16000;  
x1=1;  
x2=1;  
x3=1;  
x4=1;  
x5=1;  
x6=1;  
x7=1;  
x8=1;  
x9=1;  
x10=1;  
x11=1;  
x12=1;  
x13=1;  
x14=1;  
x15=1;  
x16=1;  
x17=1;  
x18=1;  
x19=1;  
x20=1;  
handles.diarkeiarec1=10;
```

```

x=input('Θες να ηχογραφήσεις το πρώτο παράδειγμα για το μουσικό
όργανο No 1; Y(για ΝΑΙ)/N(για ΟΧΙ)', 's')
while x1
switch x
case {'Y','y'}
    x='N'
    while running
        y=input('Έτοιμος/η για ηχογράφηση; Y/N', 's')
        switch y
        case {'Y','y'}
            disp('Ηχογράφηση')
            y1=wavrecord(handles.diarkeiarecl*fs,fs);
            wavwrite(y1, 'M011.wav');
            disp('Αποθηκεύτηκε ως M011.wav')
            x123456=wavread('Hxosgiaprogramma.wav');
            sound(0.5*x123456);
            running=0;
            epdcode(y1);
            x='N' ;
            z1=1;
            while z1
                y=input('Θες να ξανακάνεις την ηχογράφηση; Y(για ΝΑΙ) ή
N(για ΟΧΙ)', 's')
                switch y
                case {'Y','y'}
                    disp('Ηχογράφηση')
                    y1=wavrecord(handles.diarkeiarecl*fs,fs);
                    wavwrite(y1, 'one.wav');
                    disp('Αποθηκεύτηκε ως M011.wav')
                    x123456=wavread('Hxosgiaprogramma.wav');
                    sound(0.5*x123456);
                    running=0;
                    epdcode(y1);
                    case {'N','n'}
                        z1=0;
                end
            end
        case {'N','n'}
            running=1;
        otherwise
            disp('Λάθος κουμπί, πάτα Y για ΝΑΙ ή N για ΟΧΙ')
        end
    end
end

case {'N','n'}
    x1=0;
    running=1;
    x=input('Θες να συνεχίσεις τις ηχογραφήσεις? Y(για
ΝΑΙ)/N(για ΟΧΙ)', 's')
    switch x
    case {'N','n'}

    case {'Y','y'}

```

```

        x=input('Θες να ηχογραφήσεις το δεύτερο παράδειγμα για το
μουσικό όργανο No 1; Y(για ΝΑΙ)/N(για ΟΧΙ) ','s')
        while x2
            switch x
                case {'Y','y'}
                    while running
                        y=input('Ετοιμος/η για ηχογράφηση; Y/N','s')
                        switch y
                            case {'Y','y'}
                                disp('Ηχογράφηση')
                                y1=wavrecord(handles.diarkeiarecl*fs,fs);
                                wavwrite(y1,'M012.wav');
                                disp('Αποθηκεύτηκε ως M012.wav')
                                x123456=wavread('Hxosgiaprogramma.wav');
                                sound(0.5*x123456);
                                running=0;
                                epdcode(y1);
                                x='N';

                                z1=1;
                                while z1
                                    y=input('Θες να ξανακάνεις την ηχογράφηση? Y(for Yes)
or N(for No) ','s')
                                    switch y
                                        case {'Y','y'}
                                            disp('Ηχογράφηση')
                                            y1=wavrecord(handles.diarkeiarecl*fs,fs);
                                            wavwrite(y1,'M012.wav');
                                            disp('Αποθηκεύτηκε ως M012.wav')
                                            x123456=wavread('Hxosgiaprogramma.wav');
                                            sound(0.5*x123456);
                                            running=0;
                                            epdcode(y1);
                                        case {'N','n'}
                                            z1=0;
                                    end
                                end

                                case {'N','n'}
                                    running=1;
                                otherwise
                                    disp('Λάθος κουμπί, πάτα Y για ΝΑΙ ή N
για ΟΧΙ')
                                end
                            end
                        end

                    case {'N','n'}
                        running=1;
                    otherwise
                        disp('Λάθος κουμπί, πάτα Y για ΝΑΙ ή N
για ΟΧΙ')
                    end
                end
            end

            case {'N','n'}
                x2=0;
                running=1;
                x=input('Θες να συνεχίσεις τις ηχογραφήσεις? Y(για ΝΑΙ)/N(για
OXI) ','s')
                switch x
                    case {'N','n'}

                    case{'Y','y'}

                        x=input('Θες να ηχογραφήσεις το τρίτο παράδειγμα για το
μουσικό όργανο No 1; Y(για ΝΑΙ)/N(για ΟΧΙ) ','s')
                        while x3

```

```

switch x
    case {'Y','y'}
        while running
            y=input('Έτοιμος/η για ηχογράφηση; Y/N','s')
            switch y
                case {'Y','y'}
                    disp('Ηχογράφηση')
                    y1=wavrecord(handles.diarkeiarecl*fs,fs);
                    wavwrite(y1,'MO13.wav');
                    disp('Αποθηκεύτηκε ως MO13.wav')
                    x123456=wavread('Hxosgiaprogramma.wav');
                    sound(0.5*x123456);
                    running=0;
                    epdcode(y1);
                    x='N';

                z1=1;
                while z1
                    y=input('Θες να ξανακάνεις την ηχογράφηση; Y(για ΝΑΙ) ή
N(για ΟΧΙ)','s')
                    switch y
                        case {'Y','y'}
                            disp('Ηχογράφηση')
                            y1=wavrecord(handles.diarkeiarecl*fs,fs);
                            wavwrite(y1,'MO13.wav');
                            disp('Αποθηκεύτηκε ως MO13.wav')
                            x123456=wavread('Hxosgiaprogramma.wav');
                            sound(0.5*x123456);
                            running=0;
                            epdcode(y1);
                        case {'N','n'}
                            z1=0;
                    end
                end

                case {'N','n'}
                    running=1;
                    otherwise
                        disp('Λάθος κουμπί, πάτα Y για ΝΑΙ ή N
για ΟΧΙ')
                end
            end
        case {'N','n'}
            x3=0;
            running=1;
            x=input('Θες να συνεχίσεις τις ηχογραφήσεις; Y(για ΝΑΙ)/N(για
ΟΧΙ)','s')
            switch x
                case {'N','n'}

                case{'Y','y'}

                    x=input('Θες να ηχογραφήσεις το τέταρτο παράδειγμα για το
μουσικό όργανο No 1; Y(για ΝΑΙ)/N(για ΟΧΙ)','s')
                    while x4
                        switch x
                            case {'Y','y'}
                                while running
                                    y=input('Έτοιμος/η για ηχογράφηση; Y/N','s')
                                    switch y

```

```

        case {'Y','y'}
        disp('Ηχογράφηση')
        y1=wavrecord(handles.diarkeiarecl*fs,fs);
        wavwrite(y1,'MO14.wav');
        disp('Αποθηκεύτηκε ως MO14.wav')
        x123456=wavread('Hxosgiaprogramma.wav');
        sound(0.5*x123456);
        running=0;
        epdcode(y1);
        x='N'
        z1=1;
        while z1
        y=input('Θες να ξανακάνεις την ηχογράφηση;
Y(για NAI) ή N(για OXI)', 's')
        switch y
        case {'Y','y'}
        disp('Ηχογράφηση')
        y1=wavrecord(handles.diarkeiarecl*fs,fs);
        wavwrite(y1,'MO14.wav');
        disp('Αποθηκεύτηκε ως MO14.wav')
        x123456=wavread('Hxosgiaprogramma.wav');
        sound(0.5*x123456);
        running=0;
        epdcode(y1);
        case {'N','n'}
        z1=0;
        end
        end
        case {'N','n'}
        running=1;
        otherwise
        disp('Λάθος κουμπί, πάτα Y για NAI ή N
για OXI')
        end
        end
        case {'N','n'}
        x4=0;
        running=1;
        x=input('Θες να συνεχίσεις τις ηχογραφήσεις; Y(για NAI)/ N(for
No)', 's')
        switch x
        case {'N','n'}

        case{'Y','y'}

        x=input('Θές να ηχογραφήσεις το πρώτο παράδειγμα για το
μουσικό όργανο No 2; Y(για NAI)/N(για OXI)', 's')
        while x5
        switch x
        case {'Y','y'}
        while running
        y=input('Έτοιμος/η για ηχογράφηση; Y/N', 's')
        switch y
        case {'Y','y'}
        disp('Ηχογράφηση')
        y1=wavrecord(handles.diarkeiarecl*fs,fs);
        wavwrite(y1,'MO21.wav');
        disp('Αποθηκεύτηκε ως MO21.wav')
        x123456=wavread('Hxosgiaprogramma.wav');
        sound(0.5*x123456);

```

```

        running=0;
        epdcode(y1);
        x='N'
        z1=1;
        while z1
Y(για NAI) ή N(για OXI) ', 's')
        y=input('Θες να ξανακάνεις την ηχογράφηση;
        switch y
        case {'Y','y'}
            disp('Ηχογράφηση')
            y1=wavrecord(handles.diarkeiarecl*fs,fs);
            wavwrite(y1,'MO21.wav');
            disp('Αποθηκεύτηκε ως MO21.wav')
            x123456=wavread('Hxosgiaprogramma.wav');
            sound(0.5*x123456);
            running=0;
            epdcode(y1);
            case {'N','n'}
                z1=0;
            end
        end
        case {'N','n'}
            running=1;
            otherwise
                disp('Λάθος κουμπί, πάτα Y για NAI ή N
για OXI')
        end
        end
        case {'N','n'}
            x5=0;
            running=1;
            x=input('Θες να συνεχίσεις τις ηχογραφήσεις;
Y(για NAI)/N(για OXI) ', 's')
            switch x
            case {'N','n'}

            case{'Y','y'}

                x=input('Θές να ηχογραφήσεις το δεύτερο παράδειγμα για το
μουσικό όργανο No 2; Y(για NAI)/N(για OXI) ', 's')
                while x6
                switch x
                case {'Y','y'}
                    while running
                        y=input('Έτοιμος/η για ηχογράφηση; Y/N', 's')
                        switch y
                        case {'Y','y'}
                            disp('Ηχογράφηση')
                            y1=wavrecord(handles.diarkeiarecl*fs,fs);
                            wavwrite(y1,'MO22.wav');
                            disp('Αποθηκεύτηκε ως MO22.wav')
                            x123456=wavread('Hxosgiaprogramma.wav');
                            sound(0.5*x123456);
                            running=0;
                            epdcode(y1);
                            x='N'
                            z1=1;
                            while z1

```



```

        end
        case {'N','n'}
            x10=0;
            running=1;
            x=input('Θες να συνεχίσεις τις ηχογραφήσεις; Y(για
NAI)/N(για OXI) ','s')
            switch x
                case {'N','n'}

                    case{'Y','y'}

                        x=input('Θες να ηχογραφήσεις το τρίτο παράδειγμα για το
μουσικό όργανο No 3; Y(για NAI)/N(για OXI) ','s')
                        while x11
                            switch x
                                case {'Y','y'}
                                    while running
                                        y=input('Έτοιμος/η για ηχογράφηση; Y/N','s')
                                        switch y
                                            case {'Y','y'}
                                                disp('Ηχογράφηση')
                                                y1=wavrecord(handles.diarkeiarecl*fs,fs);
                                                wavwrite(y1,'MO33.wav');
                                                disp('Αποθηκεύτηκε ως MO33.wav')
                                                x123456=wavread('Hxosgiaprogramma.wav');
                                                sound(0.5*x123456);
                                                running=0;
                                                epdcode(y1);
                                                x='N'
                                                z1=1;
                                                while z1
                                                    y=input('Θες να ξανακάνεις την ηχογράφηση;
Y(για NAI) or N(OXI) ','s')
                                                    switch y
                                                        case {'Y','y'}
                                                            disp('Ηχογράφηση')
                                                            y1=wavrecord(handles.diarkeiarecl*fs,fs);
                                                            wavwrite(y1,'MO33.wav');
                                                            disp('Αποθηκεύτηκε ως MO33.wav')
                                                            x123456=wavread('Hxosgiaprogramma.wav');
                                                            sound(0.5*x123456);
                                                            running=0;
                                                            epdcode(y1);
                                                        case {'N','n'}
                                                            z1=0;
                                                    end
                                                end
                                            case {'N','n'}
                                                running=1;
                                                otherwise
                                                    disp('Λάθος κουμπί, πάτα Y για NAI or N
για OXI')
                                                end
                                        end
                                    end
                                end
                            end
                        end
                    end
                end
            end
        end
        case {'N','n'}
            x11=0;
            running=1;

```

```

        x=input('Θες να συνεχίσεις τις ηχογραφήσεις; Y(για
NAI)/N(για OXI) ','s')
        switch x
        case {'N','n'}

        case{'Y','y'}

        x=input('Θες να ηχογραφήσεις το τέταρτο παράδειγμα για το
μουσικό όργανο No 3; Y(για NAI)/N(για OXI) ','s')
        while x12
        switch x
        case {'Y','y'}
        while running
        y=input('Έτοιμος/η για ηχογράφηση; Y/N','s')
        switch y
        case {'Y','y'}
        disp('Ηχογράφηση')
        y1=wavrecord(handles.diarkeiarecl*fs,fs);
        wavwrite(y1,'MO34.wav');
        disp('Αποθηκεύτηκε ως MO34.wav')
        x123456=wavread('Hxosgiaprogramma.wav');
        sound(0.5*x123456);
        running=0;
        epdcode(y1);
        x='N'
        z1=1;
        while z1
        Y(για NAI) ή N(για OXI) ','s')
        y=input('Θες να ξανακάνεις την ηχογράφηση
        switch y
        case {'Y','y'}
        disp('Ηχογράφηση')
        y1=wavrecord(handles.diarkeiarecl*fs,fs);
        wavwrite(y1,'MO34.wav');
        disp('Αποθηκεύτηκε ως MO34.wav')
        x123456=wavread('Hxosgiaprogramma.wav');
        sound(0.5*x123456);
        running=0;
        epdcode(y1);
        case {'N','n'}
        z1=0;
        end
        end
        case {'N','n'}
        running=1;
        otherwise
        disp('Λάθος κουμπί, πάτα Y για NAI ή N
για OXI')

        end

        end
        case {'N','n'}
        x12=0;
        running=1;
        x=input('Θες να συνεχίσεις τις ηχογραφήσεις; Y(για
NAI)/N(για OXI) ','s')
        switch x
        case {'N','n'}

```

```

case{'Y','y'}

    x=input('Θες να ηχογραφήσεις το πρώτο παράδειγμα για το
μουσικό όργανο No 4; Y(για NAI)/N(για OXI) ','s')
    while x13
        switch x
            case {'Y','y'}
                while running
                    y=input('Έτοιμος/η για ηχογράφηση; Y/N','s')
                    switch y
                        case {'Y','y'}
                            disp('Ηχογράφηση')
                            y1=wavrecord(handles.diarkeiarecl*fs,fs);
                            wavwrite(y1,'MO41.wav');
                            disp('Αποθηκεύτηκε ως MO41.wav')
                            x123456=wavread('Hxosgiaprogramma.wav');
                            sound(0.5*x123456);
                            running=0;
                            epdcode(y1);
                            x='N'
                            z1=1;
                            while z1
                                Y(για NAI) or N(για OXI) ','s')
                                switch y
                                    case {'Y','y'}
                                        disp('Ηχογράφηση')
                                        y1=wavrecord(handles.diarkeiarecl*fs,fs);
                                        wavwrite(y1,'MO41.wav');
                                        disp('Αποθηκεύτηκε ως MO41.wav')
                                        x123456=wavread('Hxosgiaprogramma.wav');
                                        sound(0.5*x123456);
                                        running=0;
                                        epdcode(y1);
                                    case {'N','n'}
                                        z1=0;
                                        end
                                    end
                                    case {'N','n'}
                                        running=1;
                                        otherwise
                                            disp('Λάθος κουμπί, πάτα Y για NAI ή N
για OXI')
                                        end
                                end
                            end
                        case {'N','n'}
                            x13=0;
                            running=1;
                            x=input('Θες να συνεχίσεις τις ηχογραφήσεις; Y(για
NAI)/N(για OXI) ','s')
                            switch x
                                case {'N','n'}

                                case{'Y','y'}

                                    x=input('Θες να ηχογραφήσεις το δεύτερο παράδειγμα για το
μουσικό όργανο No 4; Y(για NAI)/N(για OXI) ','s')
                                    while x14

```



```

case {'Y','y'}
disp('Ηχογράφηση')
y1=wavrecord(handles.diarkeiarecl*fs,fs);
wavwrite(y1,'MO43.wav');
disp('Αποθηκεύτηκε ως MO43.wav')
x123456=wavread('Hxosgiaprogramma.wav');
sound(0.5*x123456);
running=0;
epdcode(y1);
x='N'
z1=1;
while z1
Y(για NAI) or N(για OXI)', 's')
y=input('Θες να ξανακάνεις την ηχογράφηση;
switch y
case {'Y','y'}
disp('Ηχογράφηση')
y1=wavrecord(handles.diarkeiarecl*fs,fs);
wavwrite(y1,'MO43.wav');
disp('Αποθηκεύτηκε ως MO43.wav')
x123456=wavread('Hxosgiaprogramma.wav');
sound(0.5*x123456);
running=0;
epdcode(y1);
case {'N','n'}
z1=0;
end
end
case {'N','n'}
running=1;
otherwise
disp('Λάθος κουμπί, πάτα Y για NAI ή N
για OXI')
end

end
case {'N','n'}
x15=0;
running=1;
x=input('Θες να συνεχίσεις τις ηχογραφήσεις; Y(για
NAI)/N(για OXI)', 's')
switch x
case {'N','n'}

case{'Y','y'}

x=input('Θες να ηχογραφήσεις το τέταρτο παράδειγμα για το
μουσικό όργανο No 4; Y(για NAI)/N(για OXI)', 's')
while x16
switch x
case {'Y','y'}
while running
y=input('Έτοιμος/η για ηχογράφηση; Y/N', 's')
switch y
case {'Y','y'}
disp('Ηχογράφηση')
y1=wavrecord(handles.diarkeiarecl*fs,fs);
wavwrite(y1,'MO44.wav');
disp('Αποθηκεύτηκε ως MO44.wav')

```



```

        z1=1;
        while z1
        y=input('Θες να ξανακάνεις την ηχογράφηση;
Y(για NAI) ή N(για OXI)', 's')
        switch y
        case {'Y','y'}
        disp('Ηχογράφηση')
        y1=wavrecord(handles.diarkeiarecl*fs,fs);
        wavwrite(y1,'MO51.wav');
        disp('Αποθηκεύτηκε MO51.wav')
        x123456=wavread('Hxosgiaprogramma.wav');
        sound(0.5*x123456);
        running=0;
        epdcode(y1);
        case {'N','n'}
        z1=0;
        end
        end
        case {'N','n'}
        running=1;
        otherwise
        disp('Λάθος κουμπί, πάτα Y για NAI ή N
για OXI')
        end
        end
        case {'N','n'}
        x17=0;
        running=1;
        x=input('Θες να συνεχίσεις τις ηχογραφήσεις; Y(για
NAI)/N(για OXI)', 's')
        switch x
        case {'N','n'}
        case{'Y','y'}

        x=input('Θες να ηχογραφήσεις το δεύτερο παράδειγμα για το
μουσικό όργανο No 5; Y(για NAI)/N(για OXI)', 's')
        while x18
        switch x
        case {'Y','y'}
        while running
        y=input('Έτοιμος/η για ηχογράφηση; Y/N', 's')
        switch y
        case {'Y','y'}
        disp('Ηχογράφηση')
        y1=wavrecord(handles.diarkeiarecl*fs,fs);
        wavwrite(y1,'MO52.wav');
        disp('Αποθηκεύτηκε ως MO52.wav')
        x123456=wavread('Hxosgiaprogramma.wav');
        sound(0.5*x123456);
        running=0;
        epdcode(y1);
        x='N'
        z1=1;
        while z1
        y=input('Θες να ξανακάνεις την ηχογράφηση;
Y(για NAI) ή N(για OXI)', 's')
        switch y
        case {'Y','y'}
        disp('Ηχογράφηση')

```



```

        running=0;
        epdcode(y1);
        case {'N','n'}
            z1=0;
        end
    end
    case {'N','n'}
        running=1;
        otherwise
            disp('Λάθος κουμπί, πάτα Y για NAI ή N
για OXI')
    end

end

end
case {'N','n'}
    x19=0;
    running=1;
    x=input('Θες να συνεχίσεις τις ηχογραφήσεις; Y(για
NAI)/N(για OXI)', 's')
    switch x
        case {'N','n'}

        case {'Y','y'}

            x=input('Θες να ηχογραφήσεις το τέταρτο παράδειγμα για το
μουσικό όργανο No 5? Y(για NAI)/N(για OXI)', 's')
            while x20
                switch x
                    case {'Y','y'}
                        while running
                            y=input('Έτοιμος/η για ηχογράφιση; Y/N', 's')
                            switch y
                                case {'Y','y'}
                                    disp('Ηχογράφιση')
                                    y1=wavrecord(handles.diarkeiarecl*fs,fs);
                                    wavwrite(y1, 'M054.wav');
                                    disp('Αποθηκεύτηκε ως M054.wav')
                                    x123456=wavread('Hxosgiaprogramma.wav');
                                    sound(0.5*x123456);
                                    running=0;
                                    epdcode(y1);
                                    x='N'
                                    z1=1;
                                    while z1
                                        Y(για NAI) ή N(για OXI)', 's')
                                        switch y
                                            case {'Y','y'}
                                                disp('Ηχογράφιση')
                                                y1=wavrecord(handles.diarkeiarecl*fs,fs);
                                                wavwrite(y1, 'M054.wav');
                                                disp('Αποθηκεύτηκε ως M054.wav')
                                                x123456=wavread('Hxosgiaprogramma.wav');
                                                sound(0.5*x123456);
                                                running=0;
                                                epdcode(y1);
                                                case {'N','n'}
                                                    z1=0;
                                                end
                                            end
                                        end
                                    end
                                end
                            end
                        end
                    end
                end
            end
        end
    end
end

```

```

end
x20=0;
case {'N','n'}
running=1;
otherwise
disp('Λάθος κουμπί, πάτα Y για NAI ή N
για OXI')
end

end
case {'N','n'}
x20=0;
otherwise
disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
x=input('Θες να ηχογραφήσεις το τέταρτο παράδειγμα για το
μουσικό όργανο No 5? Y(για NAI)/N(για OXI) ','s')

end
end

end
otherwise
disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
x=input('Θες να ηχογραφήσεις το τρίτο παράδειγμα για το
μουσικό όργανο No 5? Y(για NAI)/N(για OXI) ','s')

end
end

end
otherwise
disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
x=input('Θες να ηχογραφήσεις το δεύτερο παράδειγμα για το
μουσικό όργανο No 5? Y(για NAI)/N(για OXI) ','s')

end
end

end
otherwise
disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
x=input('Θες να ηχογραφήσεις το πρώτο παράδειγμα για το
μουσικό όργανο No 5? Y(για NAI)/N(για OXI) ','s')

end
end

end
otherwise
disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
x=input('Θες να ηχογραφήσεις το τέταρτο παράδειγμα για το
μουσικό όργανο No 4? Y(για NAI)/N(για OXI) ','s')

end
end

end
otherwise
disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
x=input('Θες να ηχογραφήσεις το τρίτο παράδειγμα για το
μουσικό όργανο No 4? Y(για NAI)/N(για OXI) ','s')

end

```

```
end
    end
    otherwise
    disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
    x=input('Θες να ηχογραφήσεις το δεύτερο παράδειγμα για το
μουσικό όργανο No 4? Y(για NAI)/N(για OXI) ','s')

end
end

    end
    otherwise
    disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
    x=input('Θες να ηχογραφήσεις το πρώτο παράδειγμα για το
μουσικό όργανο No 4? Y(για NAI)/N(για OXI) ','s')

end
end

    end
    otherwise
    disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
    x=input('Θες να ηχογραφήσεις το τέταρτο παράδειγμα για το
μουσικό όργανο No 3? Y(για NAI)/N(για OXI) ','s')

end
end

    end
    otherwise
    disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
    x=input('Θες να ηχογραφήσεις το τρίτο παράδειγμα για το
μουσικό όργανο No 3? Y(για NAI)/N(για OXI) ','s')

end
end

    end
    otherwise
    disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
    x=input('Θες να ηχογραφήσεις το δεύτερο παράδειγμα για το
μουσικό όργανο No 3? Y(για NAI)/N(για OXI) ','s')

end
end

    end
    otherwise
    disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
    x=input('Θες να ηχογραφήσεις το πρώτο παράδειγμα για το
μουσικό όργανο No 3? Y(για NAI)/N(για OXI) ','s')

end
end

    end
    otherwise
    disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
    x=input('Θες να ηχογραφήσεις το τέταρτο παράδειγμα για το
μουσικό όργανο No 2? Y(για NAI)/N(για OXI) ','s')

end
end

    end
    otherwise
    disp('Λάθος κουμπί, πάτα Y για NAI ή N για OXI')
```

```
x=input('Θες να ηχογραφήσεις το τρίτο παράδειγμα για το
μουσικό όργανο No 2? Y(για ΝΑΙ)/N(για ΟΧΙ) ','s')

    end
    end
        end
            otherwise
                disp('Λάθος κουμπί, πάτα Y για ΝΑΙ ή N για ΟΧΙ')
                x=input('Θες να ηχογραφήσεις το δεύτερο παράδειγμα για το
μουσικό όργανο No 2? Y(για ΝΑΙ)/N(για ΟΧΙ) ','s')

                    end
                    end
            end
        otherwise
            disp('Λάθος κουμπί, πάτα Y για ΝΑΙ ή N για ΟΧΙ')
            x=input('Θες να ηχογραφήσεις το πρώτο παράδειγμα για το
μουσικό όργανο No 2? Y(για ΝΑΙ)/N(για ΟΧΙ) ','s')

                end
                end
        end
    otherwise
        disp('Λάθος κουμπί, πάτα Y για ΝΑΙ ή N για ΟΧΙ')
        x=input('Θες να ηχογραφήσεις το τέταρτο παράδειγμα για το
μουσικό όργανο No 1? Y(για ΝΑΙ)/N(για ΟΧΙ) ','s')

            end
            end
        end
    otherwise
        disp('Λάθος κουμπί, πάτα Y για ΝΑΙ ή N για ΟΧΙ')
        x=input('Θες να ηχογραφήσεις το τρίτο παράδειγμα για το
μουσικό όργανο No 1? Y(για ΝΑΙ)/N(για ΟΧΙ) ','s')

            end
            end
        end
    otherwise
        disp('Λάθος κουμπί, πάτα Y για ΝΑΙ ή N για ΟΧΙ')
        x=input('Θες να ηχογραφήσεις το δεύτερο παράδειγμα για το
μουσικό όργανο No 1? Y(για ΝΑΙ)/N(για ΟΧΙ) ','s')

            end
            end
        end
    otherwise
        disp('Λάθος κουμπί, πάτα Y για ΝΑΙ ή N για ΟΧΙ')
        x=input('Θες να ηχογραφήσεις το πρώτο παράδειγμα για το
μουσικό όργανο No 1? Y(για ΝΑΙ)/N(για ΟΧΙ) ','s')

            end
            end
        end
    end
end

x=input('Θές να εκπαιδεύσεις τα μοντέλα? Y(για ΝΑΙ)/N(για
ΟΧΙ) ','s')
switch x
```

```

case {'y','Y'}

%Ekpaideush montelwn
y = wavread('M011.wav');
Ekpaideushcode(y,'M011');
y = wavread('M012.wav');
Ekpaideushcode(y,'M012');
y = wavread('M013.wav');
Ekpaideushcode(y,'M013');
y = wavread('M014.wav');
Ekpaideushcode(y,'M014');
y = wavread('M021.wav');
Ekpaideushcode(y,'M021');
y = wavread('M022.wav');
Ekpaideushcode(y,'M022');
y = wavread('M023.wav');
Ekpaideushcode(y,'M023');
y = wavread('M024.wav');
Ekpaideushcode(y,'M024');
y = wavread('M031.wav');
Ekpaideushcode(y,'M031');
y = wavread('M032.wav');
Ekpaideushcode(y,'M032');
y = wavread('M033.wav');
Ekpaideushcode(y,'M033');
y = wavread('M034.wav');
Ekpaideushcode(y,'M034');
y = wavread('M041.wav');
Ekpaideushcode(y,'M041');
y = wavread('M042.wav');
Ekpaideushcode(y,'M042');
y = wavread('M043.wav');
Ekpaideushcode(y,'M043');
y = wavread('M044.wav');
Ekpaideushcode(y,'M044');
y = wavread('M051.wav');
Ekpaideushcode(y,'M051');
y = wavread('M052.wav');
Ekpaideushcode(y,'M052');
y = wavread('M053.wav');
Ekpaideushcode(y,'M053');
y = wavread('M054.wav');
Ekpaideushcode(y,'M054');
load Pinakasmontelwn
handles.models=MontelaMousikwnOrganwn;
    case {'n','N'}
        load Pinakasmontelwn
handles.models=MontelaMousikwnOrganwn;
    otherwise
        disp('Λάθος κουμπί, παρακαλώ πάτα Y για ΝΑΙ ή Y για ΟΧΙ')
end

```

Κώδικας εκπαίδευσης(Ekpaideushcode)

Ο παρακάτω κώδικας χρησιμοποιείται για την εκπαίδευση των μοντέλων των οργάνων που θα χρησιμοποιήσουμε για την εκτίμηση των Μίξεων Γκαουσιανών Κατανομών(ΓΜΜ). Δέχεται ως όρισμα η συνάρτηση το σήμα της νότας του εκάστοτε οργάνου και την ονομασία του αρχείου με το οποίο είναι αποθηκευμένο στο

H/Y. Στην συνέχεια χρησιμοποιείται ο κώδικας epdcode για τον εντοπισμό των άκρων του σήματος. Έπειτα το σήμα χωρίζεται σε frames για την δημιουργία ενός πίνακα $(\text{sizeplaisiou}/2) \times (\text{αριθμός frames})$ για την χρησιμοποιησή του ως είσοδο στον κώδικα MFCC για τον υπολογισμό των συντελεστών MFCC, delta και delta-delta. Τέλος για κάθε ένα όργανο γίνεται η εκτίμηση των 8 Μίξεων Γκαουσιανών Κατανομών για τον 39-διάστατο πίνακα που ήταν το αποτέλεσμα του κώδικα MFCC και το μοντέλο κάθε οργάνου αποθηκεύεται σε ένα mat αρχείο, τα δεδομένα του οποίου θα χρησιμοποιηθούν στο στάδιο της εξέτασης.

```
function Ekpaideushcode (mousikoshma, model)

EPDshma=epdcode (mousikoshma) ;

sizeplaisiou= 160; % Μήκος frame
overlapsize = sizeplaisiou/2; % Μήκος overlap
hopsizе = sizeplaisiou - overlapsize; % Βήμα
nframes = length (EPDshma)/hopsizе-1; % Αριθμός frames
arithmosM=8; %Αριθμός Γκαουσιανών κατανομών που η μίξη τους θα δώσει
την Μίξη Γκαουσιανών Κατανομών
arxikodeigma = 1;
telikodeigma =sizeplaisiou;
notedata=[];
load Pinakasmontelwn

for i = 1:nframes

    frame = EPDshma (arxikodeigma:telikodeigma) ;

    notedata (:, i)=frame;

    EisodosMFCC = notedata;

    % Ολίσθηση στο επόμενο frame του EPDshma
    arxikodeigma = arxikodeigma + hopsizе;
    telikodeigma = telikodeigma + hopsizе;
end

mfccdata = mfcc (EisodosMFCC,16000,1); % Υπολογισμός των MFCC, delta
και delta-delta συντελεστών
```



```

ArMontelou = arithmosmontelou(model); % Η συνάρτηση arithmosmontelou
επιστρέφει την τιμή του αριθμού του μοντέλου ανάλογα με το μοντέλο
που έχουμε εισάγει
MontelaMousikwnOrganwn(ArMontelou).word = model; %Αποθηκεύει το όνομα
του μοντέλου στις θέσεις μνήμης που θα αποθηκεύσουμε τις GMM των
μοντέλων των νοτιών
options = statset('MaxIter',500,'Display','final');
disp(['GMM εκπαίδευση για το: ' model]);
MontelaMousikwnOrganwn(ArMontelou).gmm =
gmdistribution.fit(mfccdata', arithmosM, 'CovType', ...
    'diagonal', 'Options', options); % Αντιστοίχιση, με χρήση της
παραπάνω συνάρτησης, της Μίξης Γκαουσιανών Κατανομών (GMM) στις τιμές
των εξαγόμενων
... χαρακτηριστικών (από τον κώδικα mfcc) των δεδομένων εκπαίδευσης. Η
GMM της νότιας κάθε οργάνου αποθηκεύεται σε μια από τις 20 θέσεις
μνήμης, π.χ.
...το πρώτο παράδειγμα για το μουσικό όργανο No2 αποθηκεύεται
στην θέση 5 του πίνακα MODELS.

save Pinakasmontelwn MontelaMousikwnOrganwn

end

```

Κώδικας End-Point Detection εκπαίδευσης(epdcode)

Αυτή η συνάρτηση χρησιμοποιείται για τον υπολογισμό των άκρων του σήματος, που δίνουμε ως όρισμα όταν την καλούμε

```

function EPDSignal=epdcode(x);

x=double(x);
x=x/max(abs(x)); %Κανονικοποίηση του σήματος

%Αρχικοποίηση μεταβλητών

A=75;
A1=90;
i1=1;
arxikodeigma=1;
telikodeigma=160;
hopsize=80;
nframes= length(x)/hopsize-1;
noiseframes=50;
sizeplaisiou=160;
hamWindow1 = 0.54 - 0.46*cos(2*pi*(0:sizeplaisiou-1)/sizeplaisiou);

for i = 1:nframes

    frame = x(arxikodeigma:telikodeigma)-
mean(x(arxikodeigma:telikodeigma)); % Αφαίρεση της μέσης τιμής
    frame=frame'.*hamWindow1;

```

```

        frame_energy(i)= log((sum(frame.*frame))); %Υπολογισμός της
        λογαριθμικής ενέργειας κάθε frame

        if i < noiseframes
            energy_threshold_pinakas(i) = frame_energy(i);%
        Αποθήκευση τιμών ενέργειας στον πίνακα για τα πρώτα 50 frames
        elseif i == noiseframes
            energy_threshold =
            mean(energy_threshold_pinakas)+(0.5*std(energy_threshold_pinakas));%
        Υπολογισμός του κατωφλίου ενέργειας

        end

        arxikodeigma=arxikodeigma+hopsiz;
        telikodeigma=telikodeigma+hopsiz;
    end

    amp=frame_energy;
    amp=amp';

    amp1=energy_threshold; %Πρώτη τιμή κατωφλίου για την ενέργεια που θα
    χρησιμεύσει για τον υπολογισμό της δεύτερης τιμής κατωφλίου

    amp2=( (max(amp)^2) )-abs(amp1);% Δεύτερη τιμή κατωφλίου για την
    ενέργεια

    j=1;
    for n=noiseframes:nframes-noiseframes
        if amp(n)>=amp2
            z(j)=n; %Δημιουργία πίνακα που περιέχει τις τιμές των frames
            για τις οποίες ισχύει ότι η ενέργεια
            %τους είναι μεγαλύτερη ή ίση από την τιμή κατωφλίου
            amp2
            j=j+1;

        end
    end

    if z(end)*A1>=160000
        while z(end)*A1>=160000 %(samples=fs*t)
            z(end)=z(end-i1);
            i1=i1+1;
        end
    end

    x1=z(1);%Η πρώτη τιμή για την οποία ισχύει ότι είναι η μεγαλύτερη
    ή ίση με την τιμή κατωφλίου amp2
    x2=z(end);%Η τελευταία τιμή για την οποία ισχύει ότι είναι η
    μεγαλύτερη ή ίση με την τιμή κατωφλίου amp2

```

```

figure(2);

plot(x);
axis([0 length(x) -1 1]);
title('End Point Detection');
xlabel('Samples');
ylabel('Note Signal');
line([x1*A,x1*A],[-1,1],'Color','red');
line([x2*A1,x2*A1],[-1,1],'Color','red');
x3=x1*A;% Αριστερό άκρο του σήματος που θα εξαχθεί από αυτόν τον
κώδικα
x4=x2*A1;%Δεξιό άκρο του σήματος που θα εξαχθεί από αυτόν τον κώδικα
%figure(3);
%plot(1:nframes,
frame_energy,1:nframes,energy_threshold,1:nframes,amp2);
%axis([0 nframes min(frame_energy) max(frame_energy)]);
%title('Log-Energy');
%xlabel('Number of Frames');
%ylabel('Log-Energy of Musical Signal');
%line([x1,x1],[min(frame_energy),max(frame_energy)],'Color','red');
%line([x2,x2],[min(frame_energy),max(frame_energy)],'Color','red');

EPDsignal=x(x3:x4);% Το σήμα που θα εξαχθεί από αυτόν τον κώδικα

end

```

Κώδικας επιστροφής αριθμού μοντέλου(arithmosmontelou)

Συνάρτηση που δέχεται ως όρισμα το όνομα του μοντέλου και μας επιστρέφει ως τιμή τον αριθμό του μοντέλου με τον οποίο θα προσδιορίσουμε την θέση μνήμης που θα αποθηκεύσουμε τα μοντέλα των νοτών των οργάνων στο κώδικα εκπαίδευσης.

```

function AMontelou =arithmosmontelou(model)

switch model

case 'M011'
    AMontelou = 1;
case 'M012'
    AMontelou = 2;
case 'M013'
    AMontelou = 3;
case 'M014'
    AMontelou = 4;
case 'M021'
    AMontelou = 5;
case 'M022'
    AMontelou = 6;
case 'M023'
    AMontelou = 7;
case 'M024'

```

```

        AMontelou = 8;
    case 'M031'
        AMontelou = 9;
    case 'M032'
        AMontelou = 10;
    case 'M033'
        AMontelou = 11;
    case 'M034'
        AMontelou = 12;
    case 'M041'
        AMontelou = 13;
    case 'M042'
        AMontelou = 14;
    case 'M043'
        AMontelou = 15;
    case 'M044'
        AMontelou = 16;
    case 'M051'
        AMontelou = 17;
    case 'M052'
        AMontelou = 18;
    case 'M053'
        AMontelou = 19;
    case 'M054'
        AMontelou = 20;

    end
end

```

Κώδικας ακρόασης των ηχογραφημένων αρχείων(dokimastikoprogramma1)

Ο παρακάτω κώδικας χρησιμοποιείται για την ακρόαση των πρώτων πέντε ηχογραφημένων αρχείων που χρησιμοποιεί ο χρήστης για την εκπαίδευση

```

function varargout = dokimastikoprogramma1(varargin)

% Begin initialization code - DO NOT EDIT

gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                  'gui_Singleton',  gui_Singleton, ...
                  'gui_OpeningFcn', @dokimastikoprogramma1_OpeningFcn, ...
                  'gui_OutputFcn',  @dokimastikoprogramma1_OutputFcn, ...
                  'gui_LayoutFcn',  [], ...
                  'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargin
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});

```

```
end
% End initialization code - DO NOT EDIT

% --- Executes just before dokimastikoprogrammaal is made visible.
function dokimastikoprogrammaal_OpeningFcn(hObject, eventdata,
handles, varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% varargin   command line arguments to dokimastikoprogrammaal (see
VARARGIN)

% Choose default command line output for dokimastikoprogrammaal
handles.output = hObject;
handles.x1=0;handles.x2=1;

plot(zeros(1,160000));
% Update handles structure
guidata(hObject, handles);

% UIWAIT makes dokimastikoprogrammaal wait for user response (see
UIRESUME)
% uiwait(handles.figure1);

% --- Outputs from this function are returned to the command line.
function varargout = dokimastikoprogrammaal_OutputFcn(hObject,
eventdata, handles)
% varargout  cell array for returning output args (see VARARGOUT);
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;

function edit1_Callback(hObject, eventdata, handles)
% hObject    handle to edit1 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit1 as text
%        str2double(get(hObject,'String')) returns contents of edit1
as a double

% --- Executes on button press in pushbutton2.
function pushbutton2_Callback(hObject, eventdata, handles)
handles.x=1;
handles.x2=1;
```

```
guidata(hObject, handles);
% hObject    handle to pushbutton2 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton3.
function pushbutton3_Callback(hObject, eventdata, handles)
if handles.x==1;
    handles.x1=1;
for i=5:0.01:20;

line([100,200],[-1,1],'Color','red');hold on;
line([i*5000,i*5000],[-1,1],'Color','red');
end
plot(handles.y)
axis([0 160000 min(handles.y) max(handles.y)])
end
guidata(hObject, handles);
% hObject    handle to pushbutton3 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton4.
function pushbutton4_Callback(hObject, eventdata, handles)
if handles.x1==1;
    sound(handles.y(5*5000:20*5000),16000);
else
    sound(handles.y,16000);
end
guidata(hObject, handles);

% hObject    handle to pushbutton4 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton5.
function pushbutton5_Callback(hObject, eventdata, handles)
handles.x1=0;

hold off;
plot(handles.y)
axis([0 160000 min(handles.y) max(handles.y)])
guidata(hObject, handles);
% hObject    handle to pushbutton5 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton6.
function pushbutton6_Callback(hObject, eventdata, handles)
handles.y=wavread('M011.wav');
hold off;
plot(handles.y);
axis([0 160000 min(handles.y) max(handles.y)])
```

```
guidata(hObject, handles);
% hObject    handle to pushbutton6 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton11.
function pushbutton11_Callback(hObject, eventdata, handles)
handles.y=wavread('MO21.wav');
hold off;
plot(handles.y);
axis([0 160000 min(handles.y) max(handles.y)])
guidata(hObject, handles);
% hObject    handle to pushbutton11 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton12.
function pushbutton12_Callback(hObject, eventdata, handles)
handles.y=wavread('MO31.wav');
hold off;
plot(handles.y);
axis([0 160000 min(handles.y) max(handles.y)])
guidata(hObject, handles);
% hObject    handle to pushbutton12 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton13.
function pushbutton13_Callback(hObject, eventdata, handles)
handles.y=wavread('MO41.wav');
hold off;
plot(handles.y);
axis([0 160000 min(handles.y) max(handles.y)])
guidata(hObject, handles);
% hObject    handle to pushbutton13 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton14.
function pushbutton14_Callback(hObject, eventdata, handles)
handles.y=wavread('MO51.wav');
hold off;
plot(handles.y);
axis([0 160000 min(handles.y) max(handles.y)])
guidata(hObject, handles);
% hObject    handle to pushbutton14 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton16.
function pushbutton16_Callback(hObject, eventdata, handles)

    close dokimastikoprogramma1
```

```
% hObject    handle to pushbutton16 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
```

Κώδικας επιλογής διάρκειας ηχογράφησης(diarkeiarec)

Η συνάρτηση που φαίνεται παρακάτω χρησιμοποιείται για της αλλαγή της διάρκειας ηχογράφησης ανάλογα με την επιλογή που θα κάνει ο χρήστης. Η επιλογή του δίνεται στο γραφικό περιβάλλον διεπαφής του προγράμματος.

```
function x=diarkeiarec(y);
if y==10;
    x=10;
elseif y==5;
    x=5;
end
end
```

Κώδικες εξέτασης

Κώδικας εξέτασης(ASMIR)

Ο παρακάτω κώδικας μόλις τον ‘τρέξουμε’ στην γλώσσα προγραμματισμού MATLAB, μας εμφανίζει το γραφικό περιβάλλον διεπαφής(GUI), με τις διάφορες επιλογές του, όπως έχουμε αναφέρει στην ενότητα 5.3. Ο κώδικας αυτός περιέχει τον κώδικα επιλογής διάρκεια ηχογράφησης(diarkeiarec), ακρόασης ηχογραφημένων αρχείων(dokimastikoprogramma1), καθώς και τον κώδικα της αυτόματης ηχογράφησης, αποθήκευσης και εκπαίδευσης(Doktryprogramma) (Απλώς χρησιμοποιείται και στο στάδιο εξέτασης σαν επιλογή για το αν θα θέλαμε να ξαναηχογραφήσουμε και να ξαναεκπαιδεύσουμε τα μοντέλα των νοτών). Αφού ‘τρέξουμε’ τον κώδικα ASMIR, μπορούμε να δοκιμάσουμε και να εξετάσουμε το σύστημα αυτόματης αναγνώρισης μουσικών οργάνων, με τα πέντε όργανα που χρησιμοποιήθηκαν και πιο συγκεκριμένα για τα μουσικά όργανα κιθάρα, πιάνο, ξυλόφωνο, φλογέρα και φουσαρμόνικα.

```
function varargout = ASMIR(varargin)
% ASMIR M-file for ASMIR.fig
%     ASMIR, by itself, creates a new ASMIR or raises the existing
%     singleton*.
%
%     H = ASMIR returns the handle to a new ASMIR or the handle to
%     the existing singleton*.
%
%     ASMIR('CALLBACK',hObject,eventData,handles,...) calls the
local
%     function named CALLBACK in ASMIR.M with the given input
arguments.
%
%     ASMIR('Property','Value',...) creates a new ASMIR or raises
the
```



```

%     existing singleton*. Starting from the left, property value
pairs are
%     applied to the GUI before ASMR_OpeningFcn gets called. An
%     unrecognized property name or invalid value makes property
application
%     stop. All inputs are passed to ASMR_OpeningFcn via varargin.
%
%     *See GUI Options on GUIDE's Tools menu. Choose "GUI allows
only one
%     instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help ASMR

% Last Modified by GUIDE v2.5 08-Jul-2011 16:01:22

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                  'gui_Singleton',  gui_Singleton, ...
                  'gui_OpeningFcn', @ASMR_OpeningFcn, ...
                  'gui_OutputFcn',  @ASMR_OutputFcn, ...
                  'gui_LayoutFcn',  [], ...
                  'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if narginout
    [varargout{1:narginout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT

% --- Executes just before ASMR is made visible.
function ASMR_OpeningFcn(hObject, eventdata, handles, varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% varargin   command line arguments to ASMR (see VARARGIN)

% Choose default command line output for ASMR
handles.output = hObject;

handles.sizeplasiou = 160;    %Μήκος frame
handles.noiseframes = 50;    % Αριθμός frames που χρησιμοποιούνται για
την εκτίμηση του θορύβου περιβάλλοντος

% Θέτουμε την ως είσοδο των δεδομένων μας την κάρτα ήχου του
υπολογιστή

handles.AI = analoginput('winsound');
addchannel(handles.AI, 1);

% Ρυθμίσεις της αναλογικής εισόδου

```

```

set(handles.AI, 'SampleRate', 16000);
set(handles.AI, 'SamplesPerTrigger', handles.sizeplaisiou/2);
set(handles.AI, 'TriggerRepeat', inf);
set(handles.AI, 'TriggerType', 'immediate');
set(handles.AI, 'BufferingConfig', [2048,20]);

% Αρχικές συνθήκες
handles.running = 0;
handles.close = 0;
handles.x90=1;% Μεταβλητή για διάρκεια ηχογράφησης όταν επιλεχτεί η
εκπαίδευση μοντέλων
handles.y2=0;%
% Disable "Stop" Button on startup
set(handles.pausebutton, 'Enable', 'off');

% Αρχικοποίηση Γραφήματος
handles.plot = plot(handles.Grafhma, (0:handles.AI.SampleRate-
1)/handles.AI.SampleRate, zeros(handles.AI.SampleRate,1));
title(handles.Grafhma, 'Γραφική παράσταση σήματος στο πεδίο του
χρόνου');
xlabel(handles.Grafhma, 'Χρόνος (sec)');
ylabel('Κανονικοποιημένο Πλάτος (Normalized Amplitude)');
% Αρχικοποίηση musicalinstrument
Eikona = Eikonamontelou(21);
image(imread(Eikona), 'parent', handles.musicalinstrument);%Εμφάνιση
της αρχικής εικόνας που αναγράφει 'ΜΟΥΣΙΚΑ ΟΡΓΑΝΑ' στο GUI
axis(handles.musicalinstrument, 'image');
colormap(handles.musicalinstrument, gray(256));
set(handles.musicalinstrument, 'YTick', [], 'XTick', []);
load Pinakasmontelwn MontelaMousikwnOrganwn% Φόρτωση μοντέλων
handles.models = MontelaMousikwnOrganwn;

% Update handles structure
guidata(hObject, handles);

% UIWT makes ASMR wait for user response (see UIRESUME)
% uiwait(handles.figure1);

% --- Outputs from this function are returned to the command line.
function varargout = ASMR_OutputFcn(hObject, eventdata, handles)
% varargout cell array for returning output args (see VARARGOUT);
% hObject handle to figure
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;

% --- Executes on button press in startbutton.
function startbutton_Callback(hObject, eventdata, handles)

fs = 16000;% Συχνότητα δειγματοληψίας
handles.running=1;

```

```

handles.close=0;
notedata=[];
prev_frame=zeros(handles.sizeplaisiou,1);
handles.noiseframes=50;
i=1;
SDetection =0;
SDetectx1=0;
SDetectx2=0;
Note=0;
x99=0;
start(handles.AI)
set(handles.pausebutton,'Enable','on');
set(handles.startbutton,'Enable','off');
guidata(hObject,handles);

while handles.running

    newdata= getdata(handles.AI,handles.AI.SamplesPerTrigger);

    frame= [prev_frame(81:end); newdata];

    frame=frame-mean(frame);
    frame_energy = log(sum(frame.*frame)+eps);
    frame_zcrossings = zerocross(frame);

    if i < handles.noiseframes
        energy_threshold_buf(i) = frame_energy;
        zcrossings_threshold_buf(i) = frame_zcrossings;
    elseif i == handles.noiseframes
        energy_threshold = mean(energy_threshold_buf) +
0.5*std(energy_threshold_buf);
        zcrossing_threshold = max(mean(zcrossings_threshold_buf) +
0.5*std(zcrossings_threshold_buf),25);
    else

        % Αρχικοποίηση του δείκτη για να γίνει έλεγχος εντοπισμού
ηχητικού σήματος
        if frame_energy >= energy_threshold || frame_zcrossings >=
zcrossing_threshold

            DETECT=1;

        else
            DETECT = 0;
        end

        if SDetection ==1
            SDmetrhths=SDmetrhths+1;
            notedata(:,SDmetrhths)=frame;
            SDetectx2=1;

            if DETECT==1
                SDetectx1=1;
                SDetectx2=0;
            else
                SDetectx1=0;
            end
        end
    end
end

```

```

    if SDetectx1==1
        %Μηδενισμός των μεταβλητών SDetectx1,SDetectx2
        SDetectx1=0;
        SDetectx2=0;

    elseif SDetectx2==1

        SDetection = 0;
        SDmetrhths = SDmetrhths -20 - 1;% Θεωρώ ότι στα για
21 frame είναι σιγή και δεν υπάρχει κάποιος εντοπισμός νότας η
χρονική
% που αντιστοιχεί σε
αυτά τα 21 frames είναι ίση με 0.105sec
% Έτσι, υπολογίζεται
ο αριθμός των διαδοχικών επικαλυπτόμεων frames για τα οποία έχει
% εντοπιστεί κάποιο
σήμα που αντιστοιχεί στην νότα του μουσικού όργάνου
        x99=x99+1; % Μετρητής που θα χρησιμοποιηθεί για την
γραφική παράσταση μηδενικών στο plot, όταν δεν εισάγεται στο σύστημα
σήμα

        if SDmetrhths>=50% Απορρίπτονται τα συνεχόμενα
επικαλυπτόμενα frames των οποίων η χρονική
% διάρκεια είναι μικρότερη ή ίση των 0.255sec και
έχουν ελάχιστη
% ενέργεια (λόγω του ότι 80 δείγματα είναι το μήκος
επικάλυψης, άρα η χρονική διάρκεια της επικάλυψης είναι ίση με
% 80/16000=0.005 και για τον λόγο αυτό η χρονική
διάρκεια
% των 50frames μαζί με το 50ο frame είναι ίση με
% 0.005*51=0.255sec

        Note=1;
        EisodosMFCC=notedata(:,1:SDmetrhths);
        notedata=zeros(handles.sizeplaisiou,200);

    else
        Note=0;
        notedata=zeros(handles.sizeplaisiou,200);
    end
end
else

    if DETECT==1
        SDetection =1;
        SDmetrhths=1;
        notedata(:,1)=frame;
    end
end

if Note==1

        mfccdata = mfcc(EisodosMFCC,fs,1); % Υπολογισμός των
MFCC,delta και delta-delta συντελεστών από το σήμα εισόδου

```

```

% Υπολογισμός της αρνητικής λογαριθμικής πιθανοφάνειας που
η ελάχιστη τιμή της μας δίνει την μέγιστη
%πιθανοφάνεια
Pinakaspithan = zeros(1,20);
for j = 1:20
    [Postpithan,Pinakaspithan(j)] =
posterior(handles.models(j).gmm,mfccdata');
end

[timh,nthesh] = min(Pinakaspithan);% Δίνει την τιμή της
μικρότερης αρνητικής πιθανοφάνειας(timh) και τον αριθμό της θέσης
%που καταλαμβάνει η τιμή αυτή στον πίνακα
'Pinakaspithan'(nthesh)'

Eikona = Eikonamontelou(nthesh);% Η τιμή της θέσης
χρησιμοποιείται στην συνάρτηση 'Eikonamontelou'
%για την φόρτωση της
εικόνας του εκάστοτε μοντέλου που εντοπίστηκε
image(imread(Eikona),'parent',handles.musicalinstrument);
axis(handles.musicalinstrument,'image');
colormap(handles.musicalinstrument,gray(256));
set(handles.musicalinstrument,'YTick',[],'XTick',[]);

Grafikhparastash =
reshape(EisodosMFCC(1:handles.sizeplaisiou/2,:),1,numel(EisodosMFCC(1
:handles.sizeplaisiou/2,:)));%Δημιουργείται ένας πίνακας
γραμμή(1Χαριθμός frames) που τα στοιχεία από τα οποία αποτελείται
είναι από τα στοιχεία του πίνακα EisodosMFCC, δηλαδή είναι σαν να
ενώνουμε τα συνεχόμενα frame (που εντοπίστηκε νότια)μήκους
sizeplaisiou/2 σε ένα κοινό πίνακα και με αυτό δημιουργούμε τα
δεδομένα μας για τον άξονα y της γραφικής μας παράστασης με
αποτέλεσμα την χρήση του set παρακάτω για την εμφάνιση του σήματος
της νότιας που εντοπίστηκε στο GUI

set(handles.plot,'ydata',Grafikhparastash,'xdata',(1:length(Grafikhpa
rastash))/fs);
axis(handles.Grafhma,'tight');
SDmetrhths=1;
Note=0;
x99=0;

end

if x99>30 && Note==0 %Εμφανίζει μια γραφική παράσταση με
μηδενικά στο plot, όταν στο σύστημα δεν εισάγετε κάποιο σήμα και
γίνεται αρχικοποίηση
% του musical instrument με την εικόνα
που αναγράφει 'ΜΟΥΣΙΚΑ ΟΡΓΑΝΑ'
Eikona = Eikonamontelou(21);
image(imread(Eikona),'parent',handles.musicalinstrument);
axis(handles.musicalinstrument,'image');
colormap(handles.musicalinstrument,gray(256));
set(handles.musicalinstrument,'YTick',[],'XTick',[]);
set(handles.plot,'ydata',zeros(1,fs),'xdata',(1:fs)/fs);
title(handles.Grafhma,'Γραφική παράσταση σήματος στο
πεδίο του χρόνου');
xlabel(handles.Grafhma,'Χρόνος (sec)');
ylabel('Κανονικοποιημένο Πλάτος (Normalized Amplitude)');
x99=0;

```

```
        end

    end

    i=i+1;
    prev_frame=frame;
    handles = guidata(hObject);

end

stop(handles.AI);
set(handles.pausebutton,'Enable','off');
set(handles.startbutton,'Enable','on');

% If "Exit" button is pushed, also delete ai object and close GUI
if handles.close
    delete(handles.AI);
    closereq;
end

% hObject    handle to startbutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pausebutton.
function pausebutton_Callback(hObject, eventdata, handles)
handles.running = 0;
handles.close = 0;
handles.Mousikhmeorgana=0;
% Update handles structure
guidata(hObject, handles);
% hObject    handle to pausebutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in extibutton.
function extibutton_Callback(hObject, eventdata, handles)
if handles.running
    handles.running = 0;
    handles.close = 1;
    % Update handles structure
    guidata(hObject, handles);
else
    delete(handles.AI);
    closereq;
end

% hObject    handle to extibutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
```

```
% --- Executes on button press in TrainModels.
function TrainModels_Callback(hObject, eventdata, handles)

disp('Επέλεξε από το πρόγραμμα την διάρκεια ηχογράφησης που
επιθυμείς')

if handles.x90==1;
    handles.diarkeiarec1=10;
elseif handles.x90==0;
    handles.diarkeiarec1=handles.diarkeiarec2;
end

run Doktryprogramma

%guidata(hObject, handles);
% hObject    handle to TrainModels (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton9.
function pushbutton9_Callback(hObject, eventdata, handles)
set(handles.pushbutton9, 'Enable', 'off');
set(handles.pushbutton10, 'Enable', 'on');
handles.x90=0;
handles.diarkeiarec2=diarkeiarec(5); %Diarkeia Hxografhshs 5 sec
% Update handles structure
    guidata(hObject, handles);
% hObject    handle to pushbutton9 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton10.
function pushbutton10_Callback(hObject, eventdata, handles)
set(handles.pushbutton10, 'Enable', 'off');
set(handles.pushbutton9, 'Enable', 'on');
handles.x90=0;
handles.diarkeiarec2=diarkeiarec(10); %Diarkeia Hxografhshs 10 sec
% Update handles structure
    guidata(hObject, handles);
% hObject    handle to pushbutton10 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% --- Executes on button press in pushbutton13.
function pushbutton13_Callback(hObject, eventdata, handles)

run dokimastikoprogramma1
% hObject    handle to pushbutton13 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
```

```
% handles      structure with handles and user data (see GUIDATA)
```

Κώδικας zerocrossing(zerocross)

Συνάρτηση Zerocross για υπολογισμό των zerocrossing για κάθε frame.

```
function ZRCMetrhths = zerocross(frame)

sizeplaisiou=160;
ZRCMetrhths = 0;
  prohgoumenoproshmo = 0;

frame=frame;
  for z1=1:sizeplaisiou
    trexonproshmo=sign(frame(z1));

    if (trexonproshmo*prohgoumenoproshmo)==-1
      ZRCMetrhths = ZRCMetrhths +1;
    end
    if trexonproshmo>0 || trexonproshmo<0
      prohgoumenoproshmo= trexonproshmo;
    end

  end

end
```

Κώδικας επιστροφής εικόνας μοντέλου(Eikonamontelou)

Χρησιμοποιείται για την φόρτωση της εικόνας του οργάνου ανάλογα με την το όρισμα που θα κληθεί η συνάρτηση.Στις 4 πρώτες θέσεις είναι η εικόνα που αναγράφει 'ΠΙΑΝΟ', στις επόμενες 4 θέσεις είναι η εικόνα που αναγράφει 'ΦΥΣΑΡΜΟΝΙΚΑ', στις επόμενες 4 θέσεις είναι η εικόνα που αναγράφει 'ΞΥΛΟΦΩΝΟ', στις άλλες 4 θέσεις είναι η εικόνα που αναγράφει 'ΚΙΘΑΡΑ' και στις τελευταίες 4 θέσεις είναι η εικόνα που αναγράφει 'ΦΛΟΓΕΡΑ'.

```
function x126 = Eikonamontelou(n11_IDX)

switch n11_IDX
  case 1
    x126 = [pwd '\gifs\one.gif'];
  case 2
    x126 = [pwd '\gifs\one.gif'];
  case 3
    x126 = [pwd '\gifs\one.gif'];
  case 4
    x126 = [pwd '\gifs\one.gif'];
  case 5
    x126 = [pwd '\gifs\two.gif'];
  case 6
    x126 = [pwd '\gifs\two.gif'];
  case 7
```



```

        x126 = [pwd '\gifs\two.gif'];
    case 8
        x126 = [pwd '\gifs\two.gif'];
    case 9
        x126 = [pwd '\gifs\three.gif'];
    case 10
        x126 = [pwd '\gifs\three.gif'];
    case 11
        x126 = [pwd '\gifs\three.gif'];
    case 12
        x126 = [pwd '\gifs\three.gif'];
    case 13
        x126 = [pwd '\gifs\four.gif'];
    case 14
        x126 = [pwd '\gifs\four.gif'];
    case 15
        x126 = [pwd '\gifs\four.gif'];
    case 16
        x126 = [pwd '\gifs\four.gif'];
    case 17
        x126 = [pwd '\gifs\five.gif'];
    case 18
        x126 = [pwd '\gifs\five.gif'];
    case 19
        x126 = [pwd '\gifs\five.gif'];
    case 20
        x126 = [pwd '\gifs\five.gif'];
    case 21
        x126 = [pwd '\gifs\eleven.gif'];% Εικόνα για την αρχικοποίηση
του musicalinstrument στο GUI
end

```

Κώδικας εκπαίδευσης και εξέτασης

Κώδικας εξαγωγής MFCC,delta και delta-delta χαρακτηριστικών(mfcc)

Με τον παρακάτω κώδικα γίνεται εφικτή η εξαγωγή των MFCC,delta και delta-delta χαρακτηριστικών. Ο κώδικας αυτός επιστρέφει ένα 39-διαστάσεων πίνακα με τα προαναφερθέντα χαρακτηριστικά, τόσο κατά την διαδικασία της εκπαίδευσης(γίνεται η αντιστοίχιση των Μίξεων Γκαουσιανών Κατανομών στα σήματα των νοτών) όσο και κατά την διαδικασία της εξέτασης(υπολογίζεται η αρνητική λογαριθμική πιθανοφάνεια,negative log-likelihood ή nll).

```

function ceps = mfcc(input, samplingRate, PREPROCESS)
% Χρησιμοποιείται για την εξαγωγή των MFCC, delta και delta-delta
% χαρακτηριστικών
% ceps = mfcc(input, samplingRate)
%
% Find the Mel frequency cepstral coefficients (ceps) corresponding
to an
% input EPDshma signal. Also calculates the first and second
derivatives.
%
% INPUTS:
%
```

```
% input: Matrix of EPDshma frames. Each column represents a frame of
EPDshma.
% samplingrate: Sampling Frequency in Hz.
% PREPROCESS: Flag to apply pre-emphasis and windowing on EPDshma
frames.
%           1 = YES, 0 = NO.
%
% OUTPUTS:
%
% ceps: Matrix of 13 MFCC, 13 delta MFCC, and 13 delta-delta MFCC
%       coefficients. Each feature vector in a column.
%
% Derived from the original function 'mfcc.m' in the Auditory Toolbox
% written by:
%
% Malcolm Slaney
% Interval Research Corporation
% malcolm@interval.com
% http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/
%
% Also uses the 'deltacoeff.m' function written by:
%
% Olutope Foluso Omogbenigun
% London Metropolitan University
% http://www.mathworks.com/matlabcentral/fileexchange/19298

% Get framesize and number of frames
>windowSize,frames] = size(input);

% Filter bank parameters
lowestFrequency = 133.3333;
linearFilters = 13;
linearSpacing = 66.66666666;
logFilters = 27;
logSpacing = 1.0711703;
fftSize = 512;
cepstralCoefficients = 13;

% Keep this around for later....
totalFilters = linearFilters + logFilters;

% Now figure the band edges. Interesting frequencies are spaced
% by linearSpacing for a while, then go logarithmic. First figure
% all the interesting frequencies. Lower, center, and upper band
% edges are all consecutive interesting frequencies.

freqs = lowestFrequency + (0:linearFilters-1)*linearSpacing;
freqs(linearFilters+1:totalFilters+2) = ...
    freqs(linearFilters) * logSpacing.^(1:logFilters+2);

lower = freqs(1:totalFilters);
center = freqs(2:totalFilters+1);
upper = freqs(3:totalFilters+2);

% We now want to combine FFT bins so that each filter has unit
% weight, assuming a triangular weighting function. First figure
% out the height of the triangle, then we can figure out each
% frequencies contribution
mfccFilterWeights = zeros(totalFilters,fftSize);
```

```

triangleHeight = 2./(upper-lower);
fftFreqs = (0:fftSize-1)/fftSize*samplingRate;

for chan=1:totalFilters
    mfccFilterWeights(chan,:) = ...
        (fftFreqs > lower(chan) & fftFreqs <= center(chan)).* ...
        triangleHeight(chan).*(fftFreqs-lower(chan))/(center(chan)-
lower(chan)) + ...
        (fftFreqs > center(chan) & fftFreqs < upper(chan)).* ...
        triangleHeight(chan).*(upper(chan)-fftFreqs)/(upper(chan)-
center(chan));
end

hamWindow = 0.54 - 0.46*cos(2*pi*(0>windowSize-1)/windowSize);

% Figure out Discrete Cosine Transform. We want a matrix
% dct(i,j) which is totalFilters x cepstralCoefficients in size.
% The i,j component is given by
%           cos( i * (j+0.5)/totalFilters pi )
% where we have assumed that i and j start at 0.
mfccDCTMatrix = 1/sqrt(totalFilters/2)*cos((0:(cepstralCoefficients-
1))' * ...
        (2*(0:(totalFilters-1))+1) * pi/2/totalFilters);
mfccDCTMatrix(1,:) = mfccDCTMatrix(1,:) * sqrt(2)/2;

% Filter the input with the preemphasis filter and window.
if PREPROCESS
    preEmphasized = filter([1 -.97], 1, input);
else
    preEmphasized = input;
end

if PREPROCESS==1;
    HamWindow=HamWindow';
    for i=1:frames
        hamWindow(:,i)=HamWindow;
    end
    preEmphasized=preEmphasized.*hamWindow;% Πολλαπλασιασμός
preEmphasized με κάθε μία από τις στήλες του πίνακα hamWindow
end

fftMag = abs(fft(preEmphasized,fftSize));
earMag = log10(mfccFilterWeights * fftMag);
ceps = mfccDCTMatrix * earMag;
stdceps1=std(ceps,1,2);

for i=1:frames
    stdceps(:,i)=stdceps1;
end

ceps = ceps - stdceps;% Αφαίρεση της τυπικής απόκλισης από τις τιμές
των συντελεστών

%Υπολογισμός των συντελεστών delta και delta-delta
d =(deltacoeff(ceps')).*0.6;           % delta & mfcc
d1 = (deltacoeff(d)).*0.4;           % delta-delta & mfcc
ceps = [ceps; d' ;d1'];           % mfcc & delta & delta-delta

```

Οι τιμές των συντελεστών τοποθετούνται σε ένα ενιαίο πίνακα για την εξαγωγή τους

```
function diff = deltacoeff(x)
%Author:      Olutope Foluso Omogbenigun
%Email:      olutopeomogbenigun at hotmail.com
%University:  London Metropolitan University
%Date:       12/07/07
%Syntax:     diff = deltacoeff(Matrix);
%Calculates the time derivative of the MFCC
%coefficients matrix x and returns the result as a new matrix.

[nr,nc] = size(x);

K = 3;          %Number of frame span(backward and forward span
equal)
b = K:-1:-K;    %Vector of filter coefficients

%pads cepstral coefficients matrix by repeating first and last rows
K times
px = [ repmat(x(1,:),K,1);x; repmat(x(end,:),K,1) ];

diff = filter(b, 1, px, [], 1); % filter data vector along each
column
diff = diff/sum(b.^2);          %Divide by sum of square of all span
values
% Trim off upper and lower K rows to make input and output matrix
equal
diff = diff(K + [1:nr],:);
```