ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΝΑΓΝΩΡΙΣΗ ΗΧΗΤΙΚΗΣ ΠΗΓΗΣ ΕΦΑΡΜΟΓΗ ΑΝΑΓΝΩΡΙΣΗΣ ΠΑΡΑΔΟΣΙΑΚΩΝ ΜΟΥΣΙΚΩΝ ΟΡΓΑΝΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μανιατάκος Α. Βασίλειος-Φοίβος

Επιβλέπων : Καμπουράκης Γεώργιος
Καθηγητής Ε.Μ.Π

Αθήνα, Σεπτέμβριος 2006

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ
ΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΝΑΓΝΩΡΙΣΗ ΗΧΗΤΙΚΗΣ ΠΗΓΗΣ ΕΦΑΡΜΟΓΗ ΑΝΑΓΝΩΡΙΣΗΣ ΠΑΡΑΔΟΣΙΑΚΩΝ ΜΟΥΣΙΚΩΝ ΟΡΓΑΝΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μανιατάκος Α. Βασίλειος-Φοίβος

Επιβλέπων : Καμπουράκης Γεώργιος
Καθηγητής Ε.Μ.Π

Αθήνα, Σεπτέμβριος 2006

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 7η Σεπτεμβρίου 2006.

.....	.....	.....
Καμπουράκης Γεώργιος	Καγιάφας Ελευθέριος	Λούμος Βασίλειος
Καθηγητής Ε.Μ.Π	Καθηγητής Ε.Μ.Π	Καθηγητής Ε.Μ.Π

Αθήνα, Σεπτέμβριος 2006

……

ΜΑΝΙΑΤΑΚΟΣ Α. ΒΑΣΙΛΕΙΟΣ-ΦΟΙΒΟΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Ηλ. Υπολογιστών Ε.Μ.Π

.....

Μανιατάκος Α.Βασίλειος-Φοίβος

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Η ικανότητα του ανθρώπου να αναγνωρίζει, αποκλειστικά μέσω του ήχου, αντικείμενα στο περιβάλλον είναι εξαιρετικά ανεπτυγμένη. Αντίθετα, η μηχανή δεν είναι ακόμα σε θέση να αναγνωρίζει με την ίδια επιτυχία μια ηχητική πηγή, και μάλιστα σε δύσκολες συνθήκες θορύβου ή συνηχήσεων από πολλές πηγές η απόδοση της μηχανής φθίνει θεαματικά. Με κίνητρο κυρίως την διερεύνηση του τρόπου που αντιλαμβάνεται ο άνθρωπος, υπάρχει τα τελευταία 50 χρόνια έντονη ερευνητική δραστηριοποίηση στον τομέα της αναγνώρισης ηχητικής πηγής.
Στα πλαίσια της διπλωματικής εργασίας αναπτύχθηκε και υλοποιήθηκε ένα Σύστημα Αναγνώρισης Μουσικών Παραδοσιακών Οργάνων (ΣΑΜΠΟ). Το σύστημα αυτό προσπαθεί κάθε φορά να αναγνωρίσει το όργανο από το οποίο προέρχεται ένα προς εξέταση ηχητικό σήμα. Η μελέτη έγινε σχετικά με 4 έγχορδα παραδοσιακά όργανα, την κρητική λύρα, το μπουζούκι, το ούτι και το λαούτο. Για τους σκοπούς της εργασίας ηχογραφήθηκαν τα εν λόγω όργανα και μετά από επεξεργασία των ηχητικών δειγμάτων εξήγθη ένα σύνολο από χαρακτηριστικές ιδιότητες για καθένα δείγμα. Χρησιμοποιήθηκαν συνολικά περί τα 1000 δείγματα του ενός δευτερολέπου, το 45% των οποίων για την εκπαίδευση ενός νευρωνικού δικτύου- ταξινομητή και το 55% για την αξιολόγηση του συστήματος. Εξήγθησαν αποτελέσματα για 3 διαφορετικές παραλλαγές στο σχεδιασμό του συστήματος.

Αναγνώριση, ταξινόμηση, σετ χαρακτηριστικών ,χαρακτηριστικά, ακουστικός, λύρα ,μπουζούκι, ούτι, λαούτο, φάσμα συχνοτήτων, τονικό ύψος, κοχλίας

Abstract

The human ability to recognize objects in the environment from only the sounds they produce is extraordinary robust. On the contrary, machine does not successively confront a source recognition problem, especially in noisy, reverberant and sound competing environments. Due to the motive of explaining human unconscious perceptual processes such as hearing, there is strong research effort taking place the last 50 years.
This Diploma thesis proposes a musical instrument recognition system, specialized on traditional greek musical instruments, which tries to retrieve information concerning the source of a sound signal. The experimental research took place over 4 string traditional instruments, the cretian “lyra”, mpouzouki, hute and lute. Due to research purposes, these four instruments were recorded during solo performance and isolated tones in a home studio-like environment and a sum of statistic features and perceptual attributes were extracted and lead to a neural network classifier. Through a sum of 1000 one-second-each samples, a 45% was utilized for network education, and the rest 55% for testing. There were constructed three versions of the system, each of them leading to concrete results further analyzed at the pre-last chapter of the thesis.

Attributes, recognition, identification, transcription, constraints, acoustics, neural network, musical instruments, mpouzouki, hute, lute, lyra, features extraction, classification.

Στη μητέρα μου,
Τζούλια

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερα τον καθηγητή μου κ. Γιώργο Καμπουράκη για την πολύτιμη καθοδήγηση του και τις ωραίες κουβέντες που κάναμε κατά τη διάρκεια εκπόνησης της διπλωματικής εργασίας.
Θα ήθελα να ευχαριστήσω θερμά τον Σπύρο Ράπτη, γιατί με έκανε αυτά τα δύο πράγματα που αγαπώ εξίσου, την επιστήμη και τη μουσική, να τα αγαπήσω ακόμα περισσότερο.
Θα ήθελα να ευχαριστήσω τον Γιώργο Γιαννακάκη για την ανεκτίμητη βοήθεια του τα τελευταία χρόνια, αλλά κυρίως επειδή είμαστε φίλοι.
Επίσης τον Γιάννη Δημητρακόπουλο και τη Λήδα Μανιατάκου για την μουσική τους συνεισφορά και την υπομονή που επέδειξαν κατά την ηχογράφηση και τις αντιμουσικές υποδείξεις μου
Το Στέφανο Μεσσήνη που έδρασε αστραπή
Ευχαριστώ τέλος θερμά τους: Τάσο Μ.,τον Χρήστο ,την Ελισσάβετ, τον Νίκο, την Ειρήνη και όλους όσους βοήθησαν στην πραγματοποίηση αυτής της εργασίας

Κεφάλαιο 1: Εισαγωγή

Η ακοή είναι ένα από τα βασικότερα μέσα της ανθρώπινης διαδραστικότητας. Ωστόσο, οι γνώσεις μας για το πώς λειτουργεί είναι ακόμα πολύ περιορισμένες, κυρίως λόγω της δυσκολίας ενσυνείδητης πρόσβασης στις διαδικασίες αντίληψης. Επιπλέον, η ίδια η γλώσσα, σε ένα βαθμό, μας περιορίζει στον προσδιορισμό των ήχων.

Πολλές πάντως είναι οι φορές που οι ικανότητες αναγνώρισης ήχων από τον άνθρωπο είναι κάτι παραπάνω από εντυπωσιακές: Κάθεται ,για παράδειγμα, κάποιος, στην πολυθρόνα του σαλονιού του, όταν ακούει τον ήχο των κλειδιών έξω από την πόρτα. Από τον ήχο αυτό καταλαβαίνει ότι κάποιος από την οικογένειά του μόλις έφτασε, και πιθανότατα, ακόμα και το ποιος είναι, από τον τρόπο που τοποθετεί το κλειδί στην κλειδαρία και το θόρυβο του μπρελόκ. Την ίδια στιγμή ακούει το σκύλο στην αυλή να γαυγίζει, σε ένα μίγμα ήχων και από γαυγίσματα και άλλων σκυλιών, όμως μπορεί και διαχωρίζει το γαύγισμα το γαύγισμα του δικού του σκύλου και να το αναγνωρίζει με ακρίβεια στις χρονικές στιγμές που αυτό συμβαίνει. Πόσο μάλλον όταν κάποιος ακούει τα 3 πρώτα μέτρα από την ηχογράφηση στο κοντσέρτο αριθμός 3 του Schostakovitch, και μετά γράφει τις νότες που αντιστοιχούν σε μουσική παρτιτούρα.

Η αναγνώριση ηχητικής πηγής από τον άνθρωπο αποτελεί ένα πολύ σημαντικό κομμάτι στη διερεύνηση του ανθρώπινου ακουστικού συστήματος. Το συγκεκριμένο πεδίο απασχολεί ερευνητές από διαφορετικούς κλάδους, ακουστικούς μηχανικούς, ψυχολόγους, φυσικούς, μηχανικούς Η/Υ . Ωστόσο, η αναγνώριση ηχητικής πηγής από τη μηχανή, ξεκινώντας από τις πιο βασικές και κοινώς αποδεκτές αρχές της ακουστικής επιστήμης και προσπαθώντας να δημιουργήσει ακουστική νοημοσύνη στη μηχανή, αποτελεί το συγκερασμό των παραπάνω ερευνητικών πεδίων και πρόκληση για κάθε ερευνητή. Και αυτό γιατί όταν θα καταφέρουμε να αναλύουμε με επιτυχία μια ακουστική σκηνή (σκηνή που εκτυλίσσεται ακουστικά), σίγουρα θα ξέρουμε πολύ περισσότερα από αυτά που ξέρουμε τώρα για τις διεργασίες που πραγματοποιούνται στον ανθρώπινο εγκέφαλο. Το κίνητρο αυτό από μόνο του αρκεί για να ωθήσει κάποιον να δραστηριοποιηθεί σε τέτοιους τομείς έρευνας, στα πλαίσια δηλαδή και μόνο της εγγενής αναζήτησης του ανθρώπου για το τι πραγματικά είναι.

Όμως, σε αυτό το σημείο κρίνεται σκόπιμη μια σύντομη αναφορά και στις άμεσες και πρακτικές εφαρμογές που έχει η αναγνώριση ηχητικής πηγής. Κάποιες από αυτές είναι:

- Ταξινόμηση multimedia αρχείων σε υπολογιστή

Σε μια εποχή που ο αποθηκευτικός χώρος ανά μέσο χρήστη έχει εκτοξευθεί στα ύψη, και ο καθένας έχει στην κατοχή του σε ψηφιακή μορφή χιλιάδες δευτερόλεπτα μουσικής, είναι εμφανής η δυσκολία οργάνωσης του οπτικοακουστικού υλικού. Η αναγνώριση σίγουρα θα μπορούσε να αποτελέσει τη λύση για την ταξινόμηση και βελτιστοποίηση της διαχείρισης της ακουστικής πληροφορίας.

- Καταγραφή μουσικού αρχείου σε παρτιτούρα.

Αν και είναι νωρίς ακόμα για κάτι τέτοιο και ιδιαίτερα όταν μιλάμε για σήμα από περισσότερα από ένα όργανα, Το συγκεκριμένο θα ήταν ιδιαίτερα χρήσιμο εργαλείο για μουσικούς εκτελεστές, συνθέτες και ψυχολόγους.

- Παρακολούθηση περιβάλλοντος

Μία από τις πιο χρήσιμες εφαρμογές είναι προφανώς η παρακολούθηση των ήχων του περιβάλλοντος, π.χ στην κατασκευή έξυπνων σπιτιών με εφαρμογές home-monitoring και στην βοήθεια ατόμων με ειδικές ανάγκες.

- Συνθετικούς ομιλητές και ακροατές

Κατασκευάζοντας υπολογιστές με τη δυνατότητα να καταλαβαίνουν ήχους και να καταλαβαίνουν περιεχόμενο θα αποτελούσε βάση για φανταστικές εφαρμογές. Θα μπορούσαμε να κατασκευάσουμε εικονικούς δασκάλους μουσικής με ατελείωτη υπομονή και εικονικούς εκτελεστές για να παίζουμε μαζί τους.

Η παρούσα διατριβή έχει θέμα την αναγνώριση ηχητικής πηγής, και συγκεκριμένα την αναγνώριση μουσικών οργάνων. Στα πλαίσια αυτής σχεδιάστηκε ένα θεωρητικό μοντέλο, που στη συνέχεια υλοποιήθηκε σε υπολογιστικό περιβάλλον. Το σύστημα αναγνώρισης που αναπτύχθηκε ονομάζεται ΣΑΜΠΟ – Σύστημα Αναγνώρισης Μουσικών Παραδοσιακών Οργάνων.

Στη συνέχεια θα γίνει εκτενής αναφορά στις έρευνες γύρω από την αναγνώριση ηχητικής πηγής (κεφ.2). Στα κεφάλαια 3 και 4 αποσαφηνίζεται ο τρόπος προσέγγισης ενός τέτοιου προβλήματος και μελετάται το θεωρητικό (μαθηματικό ,φυσικό, ψυχοφυσικό και υπολογιστικό) υπόβαθρο για την επίλυσή του. Στο $5^{ο}$ κεφάλαιο παρουσιάζεται το ΣΑΜΠΟ- Σύστημα Αναγνώρισης Μουσικών Παραδοσιακών Οργάνων: ο σχεδιασμός, υλοποίηση και αποτελέσματα. Τέλος στο $6^{ο}$ κεφάλαιο γίνονται τα γενικά συμπεράσματα και αναφέρονται οι προσδοκίες για μελλοντική έρευνα.

Κεφάλαιο 2 :Αναγνωρίζοντας μουσικές πηγές.

Η αίσθηση της ακοής λειτουργεί σαν μέσο αντίληψης του περιβάλλοντος και των γεγονότων που λαμβάνουν χώρα σε αυτό. Αυτό αυτομάτως προϋποθέτει ο ακροατής να μπορεί να εξάγει συμπεράσματα για την πηγή του ήχου που εκλαμβάνει από το ακουστικό του σύστημα.

Ο άνθρωπος αναγνωρίζει πολλά γεγονότα και αντικείμενα αποκλειστικά μέσω του ήχου. Οι δυνατότητες αυτές είναι είτε έμφυτες είτε αποτελούν αποτέλεσμα μάθησης σε πολύ μικρή ηλικία. Στη βάση της εξέλιξης, η δυνατότητα αναγνώρισης αντικειμένων μέσω του παραγόμενου ήχου έχει παίξει καθοριστικό ρόλο για σχεδόν όλα τα σπονδυλωτά ζώα , από τα πρώτα χρόνια της ύπαρξής τους , ως ένα από τα κύρια όπλα επιβίωσης. Ωστόσο, οι διαδικασίες που κρύβονται πίσω από την αναγνώριση ηχητικής πηγής παραμένουν ακόμα ανεξερεύνητες και τα μόνα σχεδόν στοιχεία σχετικά με αυτές έχουν προκύψει από πειράματα ψυχοφυσικού χαρακτήρα .

Η μελέτη γύρω από αυτά τα πειράματα οδήγησε στην ανάπτυξη θεωριών γύρω από την αντίληψη μέσω της ακοής στους ζωντανούς οργανισμούς, οι οποίες χρησιμοποιήθηκαν σαν βάση για τη δημιουργία μοντέλων αντίληψης του ήχου από τη μηχανή.

Στο κεφάλαιο αυτό γίνεται λόγος για την «ακουστική σκηνή », που σήμερα αποτελεί βασικό πεδίο έρευνας της ψυχοφυσικής, μέρος της οποίας αποτελεί και η αναγνώριση ηχητικής πηγής. Μετά την ακουστική σκηνή και τα βασικά της χαρακτηριστικά κρίνεται σκόπιμη μία αναφορά στις θεωρίες της αντίληψης μέσω της ακοής ως πηγές έμπνευσης για τα αντίστοιχα μηχανικά μοντέλα. Στη συνέχεια θα γίνει αξιολόγηση των κριτηρίων για την αναγνώριση ηχητικής πηγής από τη μηχανή .Τέλος ,θα παρουσιαστούν οι κατηγορίες της αναγνώρισης σε σχέση με το είδος της ηχητικής πηγής.

2.1 Κατανόηση των ακουστικών σκηνών (Auditory Scene Analysis)

Ο κόσμος των ήχων είναι σύνθετος. Σε ένα καθημερινό περιβάλλον, πολλά αντικείμενα παράγουν ήχους ταυτόχρονα και ο ακροατής πρέπει κάπως να οργανώσει την περίπλοκη ακουστική σκηνή κατά τέτοιο τρόπο ώστε να καταλαβαίνει τη συνεισφορά της κάθε ακουστικής πηγής ξεχωριστά. Η ανάλυση ακουστικής σκηνής, ένας τομέας της ψυχοφυσικής έρευνας, προσπαθεί να εξηγήσει πώς ένας ακροατής καταλαβαίνει ένα συνεχές ακουστικό μίγμα ανεξάρτητων ηχητικών πηγών.

Η διαδικασία της ανάλυσης της ακουστικής σκηνής δυσχεραίνεται εν μέρει λόγω της διαφανούς φύση του ήχου. Κάθε ακουστική πηγή δημιουργεί τις μικρές διαφοροποιήσεις στην περιβαλλοντική πίεση αέρα – ηχητικά κύματα - που ταξιδεύουν μακριά από την πηγή. Η δυσκολία προκύπτει επειδή το ηχητικό κύμα που φτάνει στο ανθρώπινο αυτί αποτελεί υπέρθεση των κυμάτων κάθε ανεξάρτητης πηγής , και έτσι ο ακροατής έχει πρόσβαση μόνο στο ακουστικό μίγμα. Ο Helmholtz περισσότερο από αιώνα πριν, είχε παρατηρήσει:

"Το αυτί είναι επομένως σχεδόν στην ίδια κατάσταση που θα ήταν και το μάτι εάν εξέταζε ένα σημείο στην επιφάνεια του νερού μέσω ενός μακριού στενού σωλήνα που θα του επέτρεπε την παρατήρηση της ανόδου του και της καθόδου, και έπειτα του ζητούνταν (του ματιού) μια ανάλυση των σύνθετων κυμάτων της θάλασσας." [1]

Ακόμη και χωρίς αυτήν την πρόσθετη πολυπλοκότητα, η ανάλυση ακουστικής σκηνής έχει πολλά κοινά με την ανάλυση οπτικής σκηνής, η οποία είναι δεν είναι σε καμία περίπτωση ένα πρόβλημα με εύκολη λύση.

2.1.1 Η επίδραση από παράγοντες του περιβάλλοντος

Ο ίδιος ο φυσικός κόσμος που μας περιβάλλει θέτει περιορισμούς στην παραγωγή των ήχων. Σαν ερευνητικός τομέας, η ανάλυση ακουστικής σκηνής ενδιαφέρεται για τον προσδιορισμό αυτών των περιορισμών, της επίδρασής τους στα εκάστοτε ακουστικά μίγματα ,και τη χάραξη στρατηγικών για την κατανόηση τους. Στο βιβλίο που θέτει της βάσης του ερευνητικού αυτού πεδίου, ο Bregman (1990) παρουσιάζει ένα σύνολο τέτοιων περιορισμών και στρατηγικών, μαζί με τα στοιχεία δοκιμής τους από τους ανθρώπους ακροατές.

Για παράδειγμα, ανεξάρτητα ηχητικά γεγονότα σπάνια συμβαίνει να είναι συγχρονισμένα. Σαν συμπέρασμα ηχητικά ερεθίσματα που αρχίζουν, τελειώνουν, ή μεταβάλλονται συγχρονισμένα πιθανότατα έχουν προκύψει από την ίδια ηχητική πηγή. Το ανθρώπινο ακουστικό σύστημα είναι ιδιαίτερα ευαίσθητο στα ταυτόχρονα onsets σε διαφορετικές περιοχές συχνότητας ,και στη αντίστοιχη διαμόρφωση συχνότητας και πλάτους. Τα αντικείμενα στον κόσμο αλλάζουν αργά σε σχέση με τις γρήγορες ταλαντώσεις των ακουστικών κυμάτων, έτσι ήχοι με ταυτόχρονη μεταβολή στο χρόνο και με χαρακτηριστικά που συσχετίζονται, (pitch, ηχηρότητα, φασματικό περιεχόμενο, κλπ) είναι πιθανό έχουν προέλθει από την ίδια πηγή. Μέσα από αυτόν τον μηχανισμό, μια ακολουθία φωνημάτων μπορεί και ακούγεται σαν πρόταση, ή μια ακολουθία από νότες που παράγονται από ένα μουσικό όργανο μπορεί να ακουστεί ως μελωδική φράση.

Η παρακάτω παλιά πλην χρήσιμη και άκρως σημαντική ευριστική αποδίδεται στον Bregman:

"Εάν ένα μέρος από μια ομάδα ακουστικών χαρακτηριστικών μπορεί κάποιος να το αποδώσει στη συνέχεια ενός ήχου που ακούστηκε μόλις πριν, να το κάνει" [2]. Έτσι, αφού μέρη της ακουστικής σκηνής αποδοθούν σε «παλιούς» ήχους που συνεχίζουν, ότι απομένει μπορεί να ερμηνευθεί ως νέοι ήχοι. Αυτό συμβαίνει σε δύο επίπεδα, αφενός στον επίπεδο της βραχέως χρόνου πρόβλεψης που βασίζεται σε στιγμιαία χαρακτηριστικά του ήχου, αφετέρου στο επίπεδο της ευρέως χρόνου πρόβλεψης, η οποία και βασίζεται σε «ηχητικά ρεύματα» μέσα στο χρόνο.

2.1.2 Η επίδραση της μάθησης

Τα προηγούμενα χαρακτηριστικά ήταν ανεξάρτητα από το περιεχόμενο της ηχητικής σκηνής και το επίπεδο γνώσης του ακροατή. Όμως, το

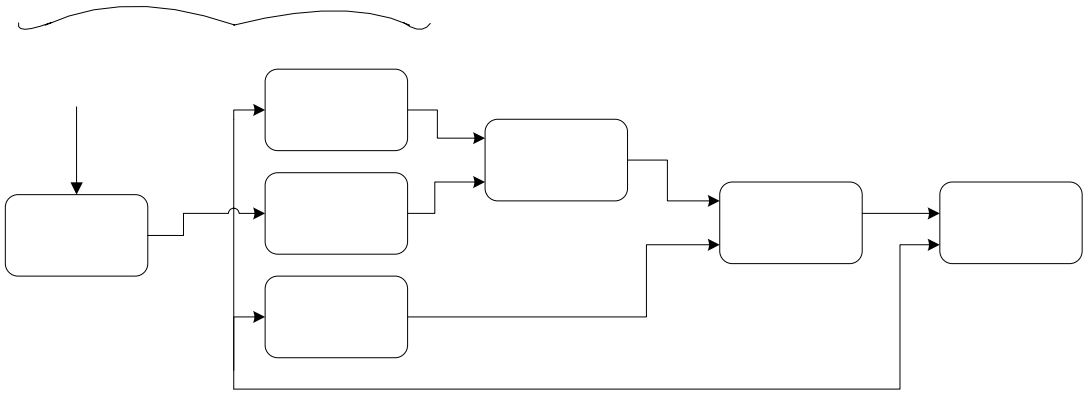
περιεχόμενο των ήχων που φτάνουν σαν μίγμα στο ανθρώπινο αυτί ,όσο και η εμπειρία του ακροατή επηρεάζουν καθοριστικά την αντίληψη. Πάνω σε αυτό το ζήτημα ο Bergman εισάγει την ιδέα των «σχημάτων» ως γνωστικά πρότυπα, τα οποία αλληλεπιδρώντας με τις πιο γενικές στρατηγικές εξηγούν τη θεωρία της ακουστικής σκηνής.

2.1.3 Ανάλυση της ακουστικής σκηνής από τη μηχανή (CASA:Computational Auditory Scene Analysis)

Τα τελευταία χρόνια πολλοί ερευνητές δοκίμασαν να σχεδιάσουν και να υλοποιήσουν υπολογιστικά πλαίσια τα οποία πραγματοποιούν ανάλυση ακουστικής σκηνής. Τα πλαίσια αυτά σε μεγάλο βαθμό υλοποιούσαν στρατηγικές του Bregman.

Ο Ellis το 1996 περιγράφει μια ποικιλία από αυτά τα συστήματα ,με αναφορές στις αυθεντικές παρουσιάσεις τους στις διατριβές των Cooke (1993), Brown (1992) και Mellinger (1991) και σε μία δημοσίευση από τον ίδιο (1994). Ο Ellis καταλήγει στο συμπέρασμα ότι η όλη η διαδικασία της CASA μπορεί να χωριστεί σε 4 κύρια μέρη που το ένα ακολουθεί το άλλο:

1. **Τμήμα Εισόδου**: Όλα τα συστήματα χρησιμοποιούν μία τράπεζα φίλτρων προκειμένου να τεμαχίσουν το ηχητικό σήμα σε διαφορετικές μπάντες συχνοτήτων. Στον ανθρώπινο οργανισμό αυτή η διαδικασία πραγματοποιείται στον κοχλία και από εκεί η πληροφορία ανάλογα με την μπάντα της συχνότητας στην οποία αντιστοιχεί μεταφέρεται σε ανώτερα στρώματα του ανθρώπινου συστήματος ακουστικής αντίληψης. Στο τμήμα αυτό μετά από επεξεργασία κωδικοποιούνται σε πίνακες συγκεκριμένα χαρακτηριστικά (cue maps).
2. **Βασική αναπαράσταση:** Σε δεύτερο επίπεδο η έξοδος του τμήματος εισόδου οργανώνεται σε συγκεκριμένα στοιχεία (atoms) τα οποία και κατασκευάζουν τα ηχητικά αντικείμενα (auditory objects). Τυπικά atoms περιλαμβάνουν τα tracks, τα οποία αναπαριστούν σταθερά ημίτονα που πιθανότητα αντιστοιχούν στις αρμονικές, και τα onsets ,τα οποία και αναπαριστούν απότομη αύξηση της ενέργειας η οποία πιθανότητα σηματοδοτεί ένα καινούργιο ήχο.
3. **Αλγόριθμος ομαδοποίησης:** Σε τρίτο επίπεδο ,ένα υποσύνολο των στρατηγικών του Bregman χρησιμοποιείται προκειμένου να ομαδοποιήσει τα στοιχεία που προέρχονται από τη βασική αναπαράσταση και αυτά να σχηματίσουν ολοκληρωμένα ηχητικά αντικείμενα(auditory objects). Για παράδειγμα τα tracks, συσχετισμένα μεταξύ τους με μια απλή συχνοτική σχέση, σχηματίζουν σε συνδυασμό έναν αρμονικό ήχο.
4. **Έξοδος ,ανασύνθεση δεδομένων:** Στο τελικό στάδιο ,οι ομαδικές αναπαραστάσεις από το τρίτο επίπεδο μετατρέπονται σε μια μορφή επιθυμητή από το εκάστοτε σύστημα. Σε μερικές περιπτώσεις η έξοδος έχει τη μορφή ακουστικών κυματομορφών οι οποίες αντιστοιχούν στα διαφορετικά ηχητικά αντικείμενα (auditory objects).



Σχήμα 2.1 Διάγραμμα ροής της επεξεργασίας στην CASA αρχιτεκτονική Ellis (1996)

Τα πρώτα CASA συστήματα είχαν σχεδιαστικά πολλούς περιορισμούς, όπως: μη ακριβή μέτρηση χαρακτηριστικών, εσφαλμένα και μη αντιστρεπτά συμπεράσματα γενίκευσης καθώς και στο τέλος ,πρόβλημα ανασύνθεση, έλλειψη δυνατότητας διαχείρισης των ομαδοποιημένων δεδομένων.

sound

Ο Ellis προσπάθησε να διορθώσει αυτούς τους περιορισμούς εισάγοντας την πρόβλεψη βραχέως χρόνου, βασισμένη σε στατιστικές ιδιότητες ηχητικών αντικειμένων χαμηλού επιπέδου (νέφη θορύβου, συντελεστές φάσματος). Η προσέγγισή του αυτή (PDCASA: Prediction driven computational auditory scene analysis) σημείωσε αξιόλογη επιτυχία σε αντικείμενα defacto αντιληπτικής έννοιας (κόρνες αυτοκινήτων, πόρτες κλπ),απέτυχε όμως να λύσει προβλήματα όπως για παράδειγμα η αποκατάσταση φωνημάτων.

2.2 Αναγνώριση ηχητική πηγής από τον άνθρωπο

Cochlea model

Εάν μια ηχητική πηγή παρήγαγε το ίδιο ηχητικό ερέθισμα κάθε φορά, η υπόθεση της αναγνώρισής της θα ήταν εύκολη - θα μπορούσαμε απλά (τουλάχιστον σε γενικές γραμμές) να απομνημονεύσουμε κάθε ήχο και να τον ταιριάξουμε με ένα από τα αποθηκευμένα δείγματα στη μνήμη. Στην πραγματικότητα, υπάρχει τεράστια μεταβλητότητα στα ακουστικά κύματα που παράγονται από οποιαδήποτε δεδομένη ηχητική πηγή σε διαφορετικό χρόνο. Αυτή η μεταβλητότητα οφείλεται εν μέρει στην πολυπλοκότητα του περιβάλλοντος - παραδείγματος χάριν, μια λεπτομερής μοντελοποίηση της ακουστικής απόκρισης ενός δωματίου μεταβάλλεται με τη διαφορετική τοποθέτηση ενός αντικειμένου, με τις μεταβολές στην κυκλοφορία του αέρα, και ακόμη και με τις μεταβολές στην υγρασία! Φυσικοί ήχοι - δηλαδή ήχοι που δεν παράγονται από αντικείμενα που έχουν κατασκευαστεί από ανθρώπους - διαφέρουν ακόμη περισσότερο από περίπτωση σε περίπτωση επειδή η φυσική διαδικασία της παραγωγής ήχου δεν είναι ποτέ η ίδια δύο φορές.

Ο ακροατής οφείλει να αντιμετωπίσει αφαιρετικά το «ακατέργαστο» ακουστικό σήμα προκειμένου να ανακαλύψει την ταυτότητα ενός ηχητικού γεγονότος. Αν και υπάρχει πολλή μεταβλητότητα στο ακουστικό σήμα, υπάρχουν συχνά σταθερές – χαρακτηριστικά που δεν αλλάζουν ανάλογα με την περίπτωση -στη διαδικασία παραγωγής του ήχου. Για παράδειγμα, ο τύπος της διέγερσης -ο τρόπος που η ενέργεια εγχέεται στο φυσικό σύστημα, παραδείγματος χάριν με το κτύπημα το φύσημα, ή το ξύσιμο –επηρεάζει το ακουστικό σήμα με πολλούς τρόπους, έμμεσους όσο και άμεσους.

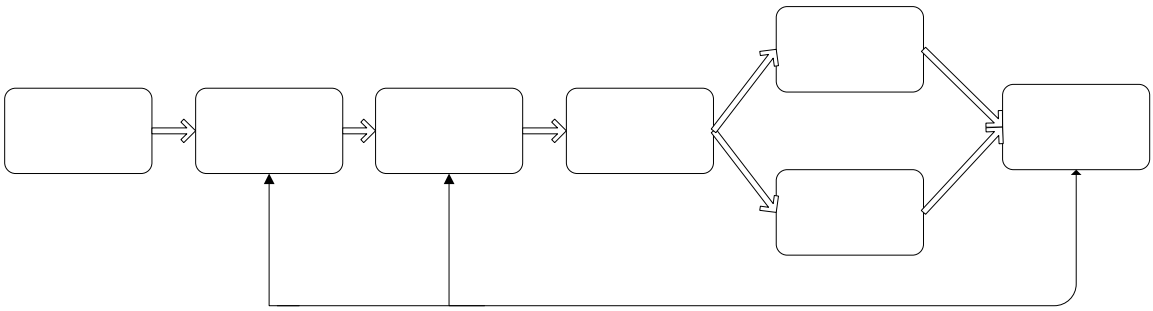
Οι υλικές ιδιότητες και η γεωμετρία του σώματος που προκαλεί τη διέγερση διαμορφώνουν κάποια χαρακτηριστικά που παραμένουν αμετάβλητα κατά τη μεταβολή των εξωτερικών συνθηκών με παρόμοιο αλλά και συμπληρωματικό τρόπο: έχουν επιπτώσεις στο φάσμα συχνότητας, στην ατάκα (Onset) και σβήσιμο (Offset) ενός ήχου ,καθώς και επιπτώσεις κατά τις, από τον ένα ήχο στον άλλο, αλληλοκαλυπτόμενες μεταβάσεις. Με τη μέτρηση χαρακτηριστικών όπως τα παραπάνω, φαντάζει δυνατή η προς τα πίσω διερεύνηση αυτών των σταθερών παραγόντων που διαμορφώνουν αυτά τα χαρακτηριστικά, και από εκεί η ταυτοποίηση ενός ηχητικού συμβάντος με μια ηχητική πηγή. Τόσο ο Handel όσο και ο McAdams πιστεύουν ότι όποιο συμπέρασμα βασισμένο στην ανίχνευση των σταθερών παραγόντων του ήχου είναι η πλέον πιθανή βάση για την αναγνώριση από τον άνθρωπο μιας ηχητικής πηγής. Είναι σημαντικό, εντούτοις, να ερευνηθούν και λιγότερο τετριμμένες σταθερές, από την ταυτοποίηση ενός ηχητικού συμβάντος με μια ηχητική πηγή, οι οποίες λαμβάνουν ως μόνο δεδομένο την ερώτηση.

Επειδή οι ιδιότητες διέγερσης και συντονισμού επηρεάζουν ταυτόχρονα τις ιδιότητες του ηχητικού κύματος, υπάρχουν πολλά μετρήσιμα ακουστικά χαρακτηριστικά που δύνανται να χρησιμοποιηθούν κατά τη διαδικασία της αναγνώρισης. Ωστόσο, δεν υπάρχει κάποιο ή μια ομάδα από κυρίαρχα χαρακτηριστικά που να μπορούν υπό όλες τις συνθήκες να είναι καθοριστικά για μια απόφαση στα πλαίσια της αναγνώρισης. Επιπλέον κανένα από αυτά τα χαρακτηριστικά ποτέ δεν αποτελεί ανεξάρτητη μεταβλητή, αλλά αντιθέτως εξαρτάται από τα υπόλοιπα. Έτσι ένας ακροατής προκειμένου να αποφανθεί για την προέλευση του ήχου θα κάνει χρήση κάθε φορά και ανάλογα με την περίπτωση μιας άλλης ομάδας χαρακτηριστικών. Ιδιαίτερα στην περίπτωση πολλών ήχων που επικαλύπτονται, είναι δύσκολο για τον ακροατή να υποθέσει εκ των προτέρων ποια χαρακτηριστικά θα του φανούν αποτελεσματικά για την κάθε περίπτωση αναγνώρισης .Επομένως, η στρατηγική αναγνώρισης του ακροατή είναι εύκαμπτη.

Ο McAdams περιγράφει την αναγνώριση ως σειρά φαινομένων: "Η αναγνώριση σημαίνει ότι κάτι που ακούγεται αυτήν την χρονική στιγμή αντιστοιχεί με κάποιο τρόπο σε κάτι που έχει ακουστεί στο παρελθόν… Η αναγνώριση μπορεί να συνοδευτεί από μια λίγο-πολύ ισχυρή αίσθηση της οικειότητας, με την αντιστοίχιση του ηχητικού ερεθίσματος με την πηγή (π.χ. ένα κόρνα αυτοκινήτων), και συχνά με μία δήλωση-γνωστοποίηση αυτού που ακούγεται προς τον ακροατή στην τρέχουσα κατάστασή του τον οδηγεί σε κάποια αντίδραση" [3].

Η θεώρηση του McAdams για τη διαδικασία της αναγνώρισης, με αφηρημένη μορφή, παρουσιάζεται στο σχήμα 2.2 (σημειώστε τις ομοιότητες

με το σχήμα 2.1). Ο McAdams πιστεύει ότι η διαδικασία είναι κατά ένα μεγάλο μέρος διαδοχική: το ακουστικό κύμα μετατρέπεται ,μέσω αισθητήρων - μεταγωγέων ,σε μια αναπαράσταση όπου είναι δυνατή η ηχητική ομαδοποίηση . Τα ομαδοποιημένα στοιχεία αναλύονται στη βάση κάποιου συνόλου χαρακτηριστικών , τα οποία χρησιμοποιούνται έπειτα ως κριτήρια στη διαδικασία της αναγνώρισης. Αν και ο McAdams υποστηρίζει ότι η αναγνώριση ακολουθεί των διαδικασιών ομαδοποίησης της ανάλυσης της ακουστικής σκηνής ,αφήνει ανοικτή την πιθανότητα για τη δυνατότητα της ανατροφοδότησης από διαδικασίες υψηλότερου επιπέδου(διαδικασίες μετά-αναγνώρισης). Ο βρόγχος της ανάδρασης ,που φαίνεται και στο σχήμα, κρίνεται απαραίτητος σε διαδικασίες όπως για παράδειγμα η αποκατάσταση φωνημάτων κατά την αναγνώριση φωνής.



Σχήμα 1.2 Στάδια της ηχητικής επεξεργασίας για την αναγνώριση ηχητικής πηγής σύμφωνα με τον McAdams(1993)

Μελέτες στην Ψυχολογία έδειξαν ότι η αναγνώριση αντικειμένου από τον άνθρωπο -σε όλες της μορφές της αντίληψης μέσω των αισθήσεων - παρουσιάζει ενδιάμεσα επίπεδα αφαιρετικότητας. Ο Minsky ορίζει αυτά σαν ζώνες επιπέδων (level bands) (Minsky, 1986). Πιστεύει ότι η λεπτομέρεια στην αναπαράσταση στη μνήμη πάνω από ένα συγκεκριμένο βαθμό δυσχεραίνει το ταίριασμα των καινούργιων καταστάσεων με τις αποθηκευμένες . Απ' την άλλη μεριά ,πάνω από έναν ορισμένο βαθμό αφαίρεσης, οι περιγραφές δεν είναι αρκετά λεπτομερείς για να είναι χρήσιμες —δεν παρέχουν πληροφορίες με διακριτική ικανότητα.

Η ιδέα αυτή μοιάζει με αυτήν της Rosch ((Rosch, 1978; Rosch et al., 1976).

Σύμφωνα με την έρευνά της, τα είδη των αντικειμένων στον κόσμο σχηματίζουν ιεραρχίες στο μυαλό. Υπάρχει ,στη μνήμη, ένα επίπεδο προνομιακό ,το «βασικό επίπεδο», όπου λαμβάνει χώρα ,κατά το μεγαλύτερο μέρος, η διαδικασία αναγνώρισης των αντικειμένων. Στο βασικό επίπεδο, απορροφάται το μεγαλύτερο μέρος της χρήσιμης πληροφορίας και εξάγονται τα πιο χρήσιμα συμπεράσματα, με τη λιγότερη μάλιστα προσπάθεια. Η αντίληψη των «βασικών αντικειμένων», των αντικειμένων που

κατηγοριοποιούνται στο βασικό επίπεδο, είναι η πρώτη διαδικασία αντίληψης που πραγματοποιείται .Τα βασικά αντικείμενα είναι τα πρώτα αντικείμενα που κατηγοριοποιούνται και ορίζονται σαν έννοιες από τα παιδιά, τα πιο κωδικοποιημένα και απαραίτητα για τη γλώσσα. Για παράδειγμα ,όσον αφορά κωδικοποίηση ηχητικών βασικών αντικειμένων, ένας ήχος που παράγεται από ένα αυτοκίνητο που κινείται κωδικοποιείται σε βασικό επίπεδο σαν ένας «δυσάρεστος ήχος μηχανής», πριν αποκωδικοποιηθεί σαν «ήχος από έκρηξη σε μηχανή εσωτερικής καύσης».

Ο Minsky με τη σειρά του, πιστεύει ότι τα αντικείμενα μπορούν να οργανωθούν σε διαφορετικές ιεραρχικές δομές με ισάριθμες διαφορετικές κατατάξεις (classifications). Ανάλογα με το περιεχόμενο του ερεθίσματος χρησιμοποιείται και η αρμόζουσα ιεραρχία κατάταξης. Το επίπεδο και οι ιεραρχίες είναι δυο έννοιες ανεξάρτητες .Για κάθε επίπεδο –όπως ορίζεται από την Rosch-κατά τον Minsky λαμβάνουν χώρα διάφορες μορφές ιεράρχησης, και αυτά τα δύο στοιχεία οδηγούν κάθε φορά στη λήψη απόφασης. Μάλιστα, η πλοήγηση του εγκεφάλου στα επίπεδα και στις ιεραρχήσεις είναι τις περισσότερες φορές διαδικασία αστραπιαία και ως εκ τούτου μη προσβάσιμη κατά την ενδοσκόπηση.

Σ’ αυτό το σημείο θα πρέπει να σημειωθεί η σημασία της ανάδρασης που επισημαίνεται στην έρευνα του Mc Adams σχετικά με την ανάλυση ακουστικής σκηνής (σχήμα 2.2). Κάποιες, εξάλλου, επιδράσεις από συμπεράσματα παραγόμενα στο υψηλό επίπεδο ,έτσι όπως φαίνονται στο διάγραμμα ροής ,είναι κάτι παραπάνω από προφανείς. Κάθε ακροατής είναι πολύ ευαίσθητος στην ακοή του ονόματος του ,ακόμα και σε σύνθετα και θορυβώδη περιβάλλοντα. Επίσης, οι πολύγλωσσοι ομιλητές μπορούν να καταλάβουν ομιλία στην μητρική τους γλώσσα, στην οποία και έχουν υποστεί ένα είδος υπερεκπαίδευσης, σε πολύ δύσκολα περιβάλλοντα .Αντίθετα, για τις γλώσσες που δεν είναι η μητρική ,χρειάζονται συνθήκες μεγαλύτερου σηματοθορυβικού λόγου.

Συμπερασματικά ,χρησιμοποιούμε οτιδήποτε ξέρουμε για μία ηχητική πηγή προκειμένου να συμπληρώσουμε κενά που προκύπτουν από ελλιπή συλλογή στοιχείων κατά την ακοή. Σύμφωνα με τον Warren ,ανάλογα με τις προσδοκίες μας ,συμπληρώνουμε άγνωστα στοιχεία με υποθέσεις βγαλμένες από την ήδη υπάρχουσα γνώση. Μάλιστα, αυτή η διαδικασία είναι απολύτως μη προσβάσιμη στο συνειδητό. Έτσι ,στην πραγματικότητα, νομίζουμε ότι ακούμε περισσότερα από όσα ακούμε στην πραγματικότητα. Η αντίληψή μας λοιπόν ,σύμφωνα πάντα με τον Warren, η αντίληψη είναι ένα αόριστο μίγμα από αισθήσεις και προσδοκίες (“sensations and expectations”).

Συμπερασματικά, ο άνθρωπος μπορεί να αναγνωρίζει διακριτές καταστάσεις μέσα από έναν πολύ μεγάλο αριθμό γενικών κλάσεων. Καταφέρνει να διαχειρίζεται ένα μεγάλο αριθμό από αποθηκευμένα δεδομένα γνώσης έτσι ώστε να μπορεί να αναγνωρίσει ένα νέο ερέθισμα εντάσσοντας το σε μια κλάση η οποία μπορεί να διαφέρει κάθε φορά ,ανάλογα με την ανάγκη της εκάστοτε ταξινόμησης, και αυτό το καταφέρνει μέσω της δυνατότητας επεξεργασίας κάθε φορά και σε διαφορετικό «επίπεδο». Με τη δυνατότητα της ανάδρασης από γνώση υψηλού επιπέδου καταφέρνει να συμπληρώνει κενά στη γνώση που παρέχεται από ένα ερέθισμα, και έτσι καταφέρνει να αναγνωρίζει με επιτυχία στις πραγματικές συνθήκες και σε δύσκολα και θορυβώδη περιβάλλοντα. Η διαδικασία της μάθησης είναι πολύ ευέλικτη: δομούμε τον κόσμο χωρίς απαραίτητα να προϋπάρχει μια ταμπέλα

για το κάθε νοηματικό αντικείμενο και επίσης μαθαίνουμε συνεχώς, προσθέτοντας ανά πάσα στιγμή καινούργιες κλάσεις. Αυτού του είδους η μη επιβλεπόμενη μάθηση συμπληρώνεται από τη μάθηση με επίβλεψη και από μεγάλη ικανότητα για γενίκευση: για παράδειγμα ,μπορεί να χρειαζόμαστε ένα πολύ μικρό αριθμό ακουσμάτων για ένα καινούργιο όργανο , προκειμένου να κατασκευάσουμε μια καινούργια κλάση και να μπορούμε να το αναγνωρίσουμε σε μία μίξη με όργανα διαφορετική από αυτή που χρησιμοποιήθηκε για να το μάθουμε. Και επιπλέον ,όλα αυτά στον άνθρωπο γίνονται σε πραγματικό χρόνο.

2.3 Αναγνώριση ηχητική πηγής από τη μηχανή

Μέχρι τώρα έχουν κατασκευαστεί πολλά συστήματα που είναι ικανά να πραγματοποιούν αναγνώριση ήχων με διαφορετικά κάθε φορά κριτήρια κατάταξης και ταξινόμησης. Τέτοια είναι, για παράδειγμα, συστήματα ικανά να αναγνωρίζουν συγκεκριμένες διαφημίσεις που παίζονται κατά τη διάρκεια προγραμμάτων στο ραδιόφωνο και την τηλεόραση, συστήματα με ικανότητα διαχωρισμού ομιλίας από μουσική, συστήματα αναγνώρισης ομιλητών από το τηλέφωνο, και συστήματα που αναγνωρίζουν όργανα σε μια ηχογράφηση. Η προσπάθεια της κατασκευής τέτοιων συστημάτων και τα προβλήματα που αυτά αντιμετώπιζαν ομαδοποίησε τις δυσκολίες κατασκευής και έθεσε τα κριτήρια αριστοποίησης τους. Παρακάτω θα γίνει μια σύντομη αναφορά σε αυτά τα κριτήρια. Στη συνέχεια θα γίνει μια σύντομη παρουσίαση των συστημάτων μηχανικής αναγνώρισης ήχων, ομαδοποιημένα σύμφωνα με το κριτήριο της κατάταξής τους, το είδος δηλαδή της αναγνώρισης. Στην συγκεκριμένη εργασία ,ακριβώς με τον ίδιο τρόπο που δε μας ενδιαφέρει το περιεχόμενο ενός μηνύματος ομιλίας αλλά η ταυτότητα του ομιλητή, έτσι και στην αναγνώριση της ηχητικής πηγής δεν μας ενδιαφέρει η μουσική φράση που εξήχθη από ένα όργανο αλλά η ταυτότητα του ίδιου του οργάνου. Έτσι τα συστήματα που θα παρουσιαστούν θα έχουν να κάνουν με αναγνώριση ταυτότητας.

2.3.1 Κριτήρια αριστοποίησης των συστημάτων αναγνώρισης ηχητικής πηγής.

1. **Δυνατότητα γενίκευσης**. Διαφορετικά στιγμιότυπα και περιπτώσεις από το ίδιο είδος ήχου θα πρέπει να εκλαμβάνονται στα ίδια. Στην αναγνώριση μουσικών οργάνων για παράδειγμα, το σύστημα θα πρέπει να είναι ικανό να αναγνωρίζει ένα μουσικό όργανο ανεξάρτητα από τον εκτελεστή ή την διαφορετικότητα στο ηχόχρωμα του οργάνου ,ή ακόμα και τη διαφοροποίηση στην ακουστότητα του χώρου στον οποίο γίνεται μια ηχογράφηση. Παρόλη δηλαδή την διαφοροποίηση στην ποιότητα δύο εκτελέσεων π.χ. από ένα μαθητή πιάνου στην $1^{η}$ τάξη που κάνει τις πρώτες του ασκήσεις τεχνικής με την εκτέλεση του Brendel σε μία σονάτα Beethoven ,το σύστημα θα πρέπει να δύναται να αναγνωρίσει και στις δύο περιπτώσεις το ίδιο πράγμα, δηλαδή το πιάνο. Αυτό αποτελεί την ικανότητα γενίκευσης του συστήματος, σε σχέση με κάτι για το οποίο είναι ήδη

εκπαιδευμένο ,καθώς δεν είναι δυνατό ένα σύστημα να εκπαιδευτεί για την εκτέλεση π.χ ενός οργάνου σε όλα τα κομμάτια ,από όλους τους πιθανούς εκτελεστές και υπό όλες τις πιθανές συνθήκες.

2. **Ικανότητα διαχείρισης της πολυπλοκότητας του πραγματικού κόσμου**. Πολύ συχνά, ψυχοακουστικά πειράματα χρησιμοποιούν πολύ απλοποιημένα ερεθίσματα σαν εισόδους, τα οποία όμως έχουν πολύ μικρή σχέση με τους ήχους που συναντούμε στον πραγματικό κόσμο. Σαν αποτέλεσμα, συστήματα τα οποία εκπαιδεύονται με ήχους όπως ημιτονικές κυματομορφές, ήχους προερχόμενους από υπολογιστική σύνθεση και λευκό θόρυβο, ήχους από ηχογραφήσεις σε ανηχοϊκό θάλαμο ,τελικά στην πράξη δεν είναι αποτελεσματικά. Το τελευταίο είναι λογικό να συμβαίνει καθώς τέτοια δείγματα ήχων δεν είναι δυνατό να συσχετίζονται με ήχους της πολυπλοκότητας του καθημερινού κόσμου, των θορύβων τυχαίων κατανομών ,ήχων με πρόσθετα στοχαστικά χαρακτηριστικά. Επιπλέον ,οι ήχοι στον πραγματικό κόσμο σπάνιο συμβαίνει να συναντώνται απομονωμένοι και ακουστικές ανακλάσεις και αντηχήσεις σχεδόν πάντα επηρεάζουν τον ήχο που φτάνει σε ένα μικρόφωνο.
3. **Επεκτασιμότητα.** Ένας άνθρωπος μπορεί πολύ εύκολα να προσθέτει σε ημερήσια διάταξη καινούργιες κλάσεις ήχων. Αντίθετα, ένα υπολογιστικό σύστημα αναγνώρισης ηχητικής πηγής μπορεί να εκπαιδευτεί για συγκεκριμένο και όχι πολύ μεγάλο αριθμό κατηγοριών ηχητικών πηγών. Είναι λοιπόν ένα πολύ σημαντικό κριτήριο το κατά πόσο ένα σύστημα είναι επεκτάσιμο ,δηλαδή μπορούν σε αυτό εκ των υστέρων να προστεθούν νέες κλάσεις, και το κατά πόσο σε αυτήν την περίπτωση το σύστημα παραμένει αποδοτικό.
4. **Ευέλικτη στρατηγική μάθησης.** Τα μηχανικά συστήματα αναγνώρισης χρησιμοποιούν είτε επιβλεπόμενη είτε μη επιβλεπόμενη μάθηση. Στην πρώτη περίπτωση προκαθορίζονται οι κατηγορίες και το σύστημα εκπαιδεύεται καθώς τα δεδομένα αντιστοιχούνται στις κατηγορίες αυτές. Στη μη επιβλεπόμενη μάθηση, οι κατηγορίες κατασκευάζονται μόνες τους και σταδιακά κατά τη διαδικασία της μάθησης. Ωστόσο ο διαχωρισμός αυτός είναι καθαρά τεχνικό θέμα. Και αυτό γιατί στην πραγματικότητα ο άνθρωπος χρησιμοποιεί για την εκπαίδευση τόσο ονοματοποιημένα δεδομένα (labeled data) όσο και μη (unlabeled). Έτσι, ενώ η μηχανή διαχωρίζει τις διαδικασίες της εκπαίδευσης με τη λειτουργία, ο άνθρωπος αντιθέτως μαθαίνει συνεχώς ,τροποποιώντας κάθε τόσο τις κατηγορίες ,προσθέτοντας καινούργιες και βελτιστοποιώντας τα κριτήρια για τις αποφάσεις για τις εκάστοτε κατηγοριοποιήσεις με βάση τα καινούργια δεδομένα. Ωστόσο δεν είναι λίγα πλέον τα συστήματα που χρησιμοποιούν υβριδική μέθοδο μάθησης ,δηλαδή μέθοδο που συνδυάζει και τους δύο προηγούμενους τρόπους εκπαίδευσης. Ωστόσο ,η αποτελεσματικότητα και η βελτιστοποίηση του χρόνου και της ποιότητας της μάθησης έγκειται στον αλγόριθμο που χρησιμοποιεί το κάθε σύστημα ,και γι ‘αυτό το λόγο η μάθηση αποτελεί ένα σημαντικότατο κριτήριο για την αξιολόγηση των συστημάτων αναγνώρισης.
5. **Λειτουργία σε πραγματικό χρόνο.** Ένα από τα βασικά χαρακτηριστικά του ανθρώπου ακροατή είναι ότι αλληλεπιδρούν με το περιβάλλον τους την ίδια στιγμή ακριβώς που διαδραματίζεται το

ηχητικό φαινόμενο. Αντίθετα, τα μηχανικά συστήματα αναγνώρισης βασίζονται σε προεπιλεγμένα ηχητικά αποσπάσματα προορισμένα εκ των προτέρων για ανάλυση κάτι που προφανώς έρχεται σε αντίθεση με την ίδια τη φύση της μουσικής. Άρα λοιπόν προκειμένου μέσα από τη μηχανική αναγνώριση να προσεγγιστεί η φύση της βιολογικής ακουστικής αντίληψης, είναι επιτακτική η ανάγκη της κατασκευής συστημάτων τα οποία εξ' ορισμού θα διενεργούν ακουστική ανάλυση των ερεθισμάτων και θα αποφαίνονται για την κατηγορία τους σε πραγματικό χρόνο.

Εκτός από τα 5 βασικά κριτήρια που προαναφέρθηκαν υπάρχουν και άλλα κριτήρια τα οποία ποικίλουν ανάλογα με την κατεύθυνση που επιλέγει ο κατασκευαστής για το σύστημα. Για παράδειγμα ,δύο συστήματα αναγνώρισης με την ίδια αποτελεσματικότητα μπορούν να εκτιμηθούν διαφορετικά ανάλογα με την απλότητα της κατασκευής την οποία επιδεικνύουν ή ,στην περίπτωση που το ζητούμενο είναι η προσέγγιση της βιολογικής λειτουργίας, το πόσο ρεαλιστικό βιολογικά είναι το μοντέλο που χρησιμοποιήθηκε. Ωστόσο εδώ θα πρέπει να αναφερθεί ότι δεν είναι απαραίτητο κριτήριο η δυνατότητα ανασύνθεσης προκειμένου για την αναπαραγωγή των δεδομένων. Αυτό βέβαια δεν ισχύει στην περίπτωση που η αναγνώριση αποτελεί μέρος της διαδικασίας ανάλυσης μουσικής σκηνής (CASA), καθώς εκεί δεν αρκούν τα λιγοστά δεδομένα που χρειάζεται η αναγνώριση ανεξάρτητα προκειμένου να μπορεί να αναπαρασταθεί η γνώση.

Τέλος, θα πρέπει να αναφερθεί η σημασιολογική διαφορά ανάμεσα σε κάποιους σημαντικούς όρους για το συγκεκριμένο πεδίο έρευνας ,οι οποίοι εκ πρώτης όψεως δείχνουν να είναι πολύ κοντά. Οι όροι αυτοί είναι :*ταξινόμηση (classification)* ,*διερεύνηση (identification)* και *αναγνώριση (recognition)*. *Αναγνώριση* είναι η διαδικασία συλλογής πληροφορίας και εξαγωγής συμπερασμάτων. Η *ταξινόμηση* σαν όρος εμπεριέχει την έννοια, σε τελικό στάδιο, εξαγωγή συμπερασμάτων με μορφή αντιστοίχησης σε κατηγορίες. Η *διερεύνηση*, χρησιμοποιείται για να περιγράψει διαδικασίες αναγνώρισης στις οποίες η επιλογή των κατηγοριών δεν είναι προκαθορισμένη από πριν.

2.3.2 Συστήματα μηχανικής αναγνώρισης ήχων

2.3.2.1 Η αναγνώριση σε πολύ εξειδικευμένες κατηγορίες ήχων

Πολλά συστήματα κατασκευάστηκαν με στόχο να μπορούν να αναγνωρίζουν δείγματα σε πολύ εξειδικευμένες κατηγορίες ήχων. Ένα τυπικό παράδειγμα από ένα τέτοιο σύστημα είναι αυτό του Nooralahiyan και των συνεργατών του το 1998 , το οποίο ήταν σε θέση να αναγνωρίζει από τον ήχο διαφορετικούς τύπους οχημάτων. Στο σύστημα αυτό ένα δείγμα του ήχου, επιλεγμένο από ανθρώπινο χέρι, κωδικοποιούνταν αρχικά μέσω του αλγόριθμου γραμμικής πρόβλεψης LPC (Linear Prediction Algorithm). Στη συνέχεια οι συντελεστές του LPC εισέρχονταν σε ένα νευρωνικό δίκτυο TDNN το οποίο και κατέτασσε τον ήχο σε μία από τις 4 κατηγορίες :νταλίκες, Ι.Χ, μοτοσικλέτες, και μικρά φορτηγά. Πραγματοποιήθηκαν δύο μελέτες:Η μία με όλους τους ήχους ηχογραφημένους προσεκτικά σε συνθήκες εργαστηρίου ,και

η δεύτερη σε πιο ρεαλιστικές συνθήκες. Και στις δύο χρησιμοποιήθηκε για την εκπαίδευση επιβλεπόμενη μάθηση. Το σύστημα παρουσίασε εξαιρετικά αξιόλογη επίδοση, με ποσοστό 96% για τους ήχους της εκπαίδευσης και 81% για τους ήχους της επαλήθευσης. Προφανώς το TDNN βρήκε κάποιο είδος κανονικότητας στα χαρακτηριστικά τα οποία επέτρεψαν την κατάταξη σε κατηγορίες ,αλλά όπως είναι τυπικό για τέτοιου είδους μελέτες, δεν έγινε προσπάθεια να βρεθούν τα πιο σημαντικά χαρακτηριστικά.

Συστήματα με ίδιο προσανατολισμό συναντώνται σε μη συσχετιστικές προσεγγίσεις της διαδικασίας αναγνώρισης για τραγούδια σε μουσικές επενδύσεις ταινιών (Hawley,1993;Pfeiffer et al. ,1996).

Υπάρχουν στη βιβλιογραφία κάποια παραδείγματα συστημάτων τα οποία σκοπό είχαν να συγκριθούν με το ανθρώπινο σύστημα αναγνώρισης στην επιτυχία εξαγωγής συμπερασμάτων σε μη προφανή ζητήματα, όπως για παράδειγμα το φύλο ενός ανθρώπου από τον ήχο τον παπουτσιών του όταν περπατά (Li et al. 1991). Μάλιστα σε αυτή την περίπτωση εξιχνιάστηκε ένα σύνολο ακουστικών χαρακτηριστικών τα οποία συσχετίζονταν με τις ανθρώπινες κρίσεις.

Τα συστήματα αναγνώρισης εξειδικευμένων κατηγοριών διαφέρουν πολύ στην ικανότητα να γενικεύουν σε σχέση με τα δείγματα εκπαίδευσης. Αυτή η διαφοροποίηση έγκειται στο γεγονός ότι η επιλογή των χαρακτηριστικών για την αναγνώριση δεν είναι πολλές φορές η πλέον κατάλληλη, ή από το μικρό αριθμό ή την έλλειψη ανομοιογένειας των δειγμάτων για εκπαίδευση. Κάποια παρουσιάζουν αποδοτικότητα μόνο για ηχογραφήσεις εργαστηρίου και άλλα προσαρμόζονται και σε περιπτώσεις ήχων σε συνθήκες περιβάλλοντος. Κανένα από αυτά δεν μπορεί να αντεπεξέλθει σε ήχους που βρίσκονται σε μίξη με άλλους ήχους ή με θόρυβο, δεν δουλεύει real-time, ενώ δεν έχει κατηγορία «άγνωστο» για δείγματα τα οποία δυσκολεύεται να κατηγοριοποιήσει. Επίσης ,είναι δύσκολο να πει κάποιος αν έχουν επεκτασιμότητα

2.3.2.2 Η αναγνώριση σε πιο γενικές κατηγορίες ήχων

Ένα τυπικό παράδειγμα τέτοιας κατηγορίας είναι η διάκριση μεταξύ ομιλίας και μουσικής, η οποία και έχει εφαρμογή στην αυτοματοποιημένη αναγνώριση φωνής και την κατάτμηση soundtrack. Τέτοια συστήματα έχουμε από τους Spina and Zue,1996, Scheirer and Schlaney,1997. Το 2^{ο} παρουσιάζει τη μεγαλύτερη πληρότητα . Συγκεκριμένα ,για τη μελέτη έγινε εκτεταμένη συλλογή δειγμάτων από το ραδιόφωνο, και δοκιμάστηκαν πολλά διαφορετικά συστήματα αναγνώρισης ,το καθένα με διαφορετικό συνδυασμό εξαγόμενων χαρακτηριστικών (από 13 συνολικά). Το 90% των δειγμάτων χρησιμοποιήθηκε για την εκπαίδευση(επιβλεπόμενη) και το 10% για την αξιολόγηση. Ο καλύτερος classifier χρησιμοποιούσε μόνο 3 από τα 13 χαρακτηριστικά και είχε ποσοστό λάθους μόλις 5,8%, το οποίο έπεφτε στο 1,4% με αύξηση του χρόνου στα δείγματα που προορίζονταν για αξιολόγηση (2.4 sec). Για το σύστημα έγινε και software υλοποίηση πραγματικού χρόνου.

Ένα επίσης σημαντικό σύστημα ήταν αυτό του Wold (1996). Η σημαντικότητα του έγκειται στην επεκτασιμότητα που επιδείκνυε. Μάλιστα, το συγκεκριμένο σύστημα επέτρεπε στο χρήστη να ορίσει μία νέα κατηγορία

μέσω ενός νέου αριθμού παραδειγμάτων. Το σύστημα χρησιμοποιούσε κατά κόρον αντιληπτικά αντί για στατιστικά χαρακτηριστικά όπως ένταση, pitch, λαμπερότητα, εύρος και αρμονικότητα καθώς και την διακύμανση των χαρακτηριστικών αυτών μέσα στο χρόνο. Για την ταξινόμηση χρησιμοποιήθηκαν γκαουσιανά μοντέλα και απόσταση Mahalanobis. Όσον αφορά στην αξιολόγηση του εν λόγω συστήματος, κατά των διαχωρισμό δειγμάτων όπως γυναικεία ομιλία η τονικό τηλέφωνο ήταν αρκετά αποδοτικό, όμως σε πιο δύσκολα προβλήματα όπως είναι η ταξινόμηση μουσικών οργάνων και αναγνώριση ταυτότητας ομιλητών δεν μπόρεσε να αντεπεξέλθει με ιδιαίτερη επιτυχία. Αυτό πιθανότατα να γινόταν δυνατό μέσω της διεύρυνσης των χρησιμοποιούμενων χαρακτηριστικών.

2.3.2.3 Η αναγνώριση ομιλητών

Σύστημα αυτής της κατηγορίας συναντούμε από τους Mammone et al. το 1996. Αυτό όπως και άλλα παρόμοια συστήματα χρησιμοποιούν κατά κύριο λόγο στατιστικές μεθόδους από τη θεωρία αναγνώρισης προτύπων με επιβλεπόμενη μάθηση. Το σύστημα του Reynolds, χρησιμοποιεί MFCC (Mel Frequency Cepstral Coefficients) σαν χαρακτηριστικά για την είσοδο. Τα cepstrum υπολογίζονται σε παράθυρα χρονικού πλάτους 20ms, και σαν τυπικό σύστημα αναγνώρισης φωνής σκοπός είναι να βρεθούν τα formants της φωνής που αντιπροσωπεύουν τους συντονισμούς στο εσωτερικό του φωνητικού σωλήνα. Η επίδοση του μοντέλου κρίνεται σχεδόν άπταιστη για ηχογραφήσεις σε ιδανικές συνθήκες και πλήθος ομιλητών μέχρι τους 630 (ηχογραφήσεις από την βάση δεδομένων TIMIT). Για δείγματα ομιλίας με πιο δύσκολες συνθήκες όπως ομιλία από το τηλέφωνο η σε συνδυασμό με άλλους ήχους η απόδοση πέφτει στο 83% (για πλήθος ομιλητών 113).

Τα συστήματα που κατασκευάστηκαν μέχρι το 2000 έχουν στηριχθεί μόνο σε ένα μέρος των ακουστικών ιδιοτήτων που ακουσία αναλύει το ανθρώπινο μυαλό προκειμένου να αποταθεί για την ταυτότητα ενός ομιλητή. Οι προσεγγίσεις με MFCC δεν επαρκούν γιατί ,για παράδειγμα, δεν λαμβάνουν υπ' όψη τους τη βασική συχνότητα (τον τόνο) της ανθρώπινης ομιλίας. Επίσης ο ρυθμός ομιλίας, μέρος της ανάλυσης στο πεδίο του χρόνου, ο οποίος και αποτελεί σημαντικότατο στοιχείο για την αναγνώριση από τον άνθρωπο ,σπάνια συναντάται με κάποια μορφή σε κάποιο από τα μηχανικά συστήματα.

2.3.2.4 Αναγνώριση ήχων του περιβάλλοντος

Δύο συστήματα αναγνώρισης αυτής της κατηγορίας είναι άξια για αναφορά: Το SUT(Sound Understanding Testbed) από τον Klassner το 1996, και το σύστημα του Saint-Arnaud (1995).

Το SUT είναι ένα πολύ έξυπνο σύστημα με πολλαπλά επίπεδα αφαιρετικότητας στην χρήση των χαρακτηριστικών το οποίο είναι ικανό να αναγνωρίζει 40 είδη ήχων περιβάλλοντος. Στα πλαίσια του SUT έγιναν 2 μελέτες: και οι δύο με τελικό τεστ 4 διαφορετικών ήχων οι οποίοι ακούγονταν σε διάστημα 5 δευτερολέπτων και το σύστημα έπρεπε να αποφανθεί για την ακριβή χρονική στιγμή και για την ταυτότητα των ήχων που ακούστηκαν. Κατά την πρώτη το σύστημα ήταν «ενήμερο» για τους ήχους που θα ακούγονταν , ενώ στη δεύτερη έπρεπε να αποφασίσει ανάμεσα στις 40 επιλογές για τους 4 ήχους. Η πρώτη είχε επιτυχία 61% και η δεύτερη 59%. Σε κάθε περίπτωση ,αν και το σύστημα παρουσίαζε αμφιλεγόμενη επεκτασιμότητα και

ικανότητα γενίκευσης, ωστόσο λόγω της εκτεταμένης γνώσης που συμπεριλάμβανε ήταν ένα βήμα προς τη σωστή κατεύθυνση για την ακουστική ανάλυση σκηνής (Auditory Scene Analysis).

Ο Saint-Arnaud επιχείρησε να εξερευνήσει ένα εύρος πολύ απλών ήχων που τα ονόμασε textures. Textures είναι ήχοι όπως αυτοί ενός φωτοτυπικού μηχανήματος, ο ήχος του νερού που κοχλάζει, φωνές που ψιθυρίζουν. Το χαρακτηριστικό των textures είναι ότι ενώ τοπικά παρουσιάζουν μία μοναδικότητα στη δομή (ίδιοι ήχοι σε διαφορετικές ηχογραφήσεις) ,εντούτοις μακροσκοπικά η τυχαιότητα τους είναι περιορισμένη καθώς παρουσιάζουν μεταξύ τους κοινά χαρακτηριστικά. Το πρώτο που επιχείρησε ήταν να κάνει ένα ψυχοφυσικό πείραμα προκειμένου να διαπιστώσει ποια είναι τα κριτήρια με βάση τα οποία οι άνθρωποι ταξινομούν αυτούς τους απλούς ήχους, δηλαδή παρότρυνε τους ακροατές να κατασκευάζουν αυτοί τις κατηγορίες. Σαν συμπέρασμα έλαβε ότι οι άνθρωποι έχουν την τάση να κατηγοριοποιούν ανάλογα με την ακουστική πηγή ,όπως νερό, φωνές, μηχανές όπως και με βάση τα ακουστικά χαρακτηριστικά ,όπως την περιοδικότητα και το ποσοστό θορύβου. Στη συνέχεια επιχείρησε να κατασκευάσει ένα υπολογιστικό ταξινομητή (classifier) που θα μπορούσε να προσομοιώσει τον ανθρώπινο τρόπο ταξινόμησης. Χρησιμοποίησε μία τεχνική βασισμένη σε cluster πιθανότητες προκειμένου να κατατάξει σήματα που είχαν υποστεί φιλτράρισμα τύπου Q σε 21 μπάντες συχνοτήτων. Σαν μετρό για την διαχωρισιμότητα χρησιμοποίησε κάποιο κατασκευασμένο από τον ίδιο. Καθώς χρησιμοποίησε πολύ μικρό αριθμό δειγμάτων απέφυγε να αξιολογήσει το σύστημα του, αναφέρει όμως ότι τα αποτελέσματά του ήταν ενθαρρυντικά.

2.3.2.5 Αναγνώριση μουσικών οργάνων.

Τα τελευταία 30 χρόνια κατασκευάστηκαν αρκετά συστήματα αναγνώρισης μουσικών οργάνων, με διαφορετική κάθε φορά οπτική ,σκοπό και επίπεδο απόδοσης. Τα περισσότερα από αυτά χρησιμοποιούσαν για δείγματα εκπαίδευσης και αξιολόγησης απομονωμένους τόνους(isolated tones), δηλαδή μεμονωμένες νότες, είτε αυτές ήταν φυσικές είτε προέρχονταν από κάποιο μηχανικό συνθέτη (synthesizer). Τελευταία βρίσκουμε συστήματα τα οποία χρησιμοποιούν μουσικές φράσεις από ηχογραφήσεις για τη μουσική βιομηχανία.

Ο De Poli το 1994 κατασκεύασε μια σειρά από αυτοοργανούμενους χάρτες Kohonen ,γνωστοί σαν νευρωνικά δίκτυα SOM. Οι είσοδοι ήταν μεμονωμένες νότες. Σε κάθε περίπτωση χρησιμοποιήθηκε μία νότα για κάθε όργανο (για 40 διαφορετικά όργανα) .Από τις νότες εισάγονταν πολλά χαρακτηριστικά στο νευρωνικό SOM , όπως για παράδειγμα οι συντελεστές MFCC ,και σε διάφορες εκδοχές της μελέτης ο χώρος των χαρακτηριστικών (feature space) μειωνόταν με χρήση PCA (Principal Component Analysis).

Οι Feiten και Gunzel (1994), σε μια μελέτη που κινήθηκε περίπου στα ίδια πλαίσια, εκπαίδευσαν ένα Kohonen SOM με φασματικά χαρακτηριστικά από 98 νότες που παρήχθησαν από το συνθέτη Roland Soundcanvas. Οι συγγραφείς υποστηρίζουν ότι το νευρωνικό μπορεί κάλλιστα να

χρησιμοποιηθεί για εφαρμογές αναζήτησης, ωστόσο δεν παρουσίασαν επίσημα αποτελέσματα.
Οι Kaminskyj και Materga το 1995 συνέκριναν τις ικανότητες από δύο classifiers ,οι οποίοι σαν είσοδο είχαν ένα κοινό set χαρακτηριστικών φασματικού περιεχομένου από μεμονωμένες νότες: ένα νευρωνικό εμπρόσθιας τροφοδότησης και έναν k-nearest neighbour classifier. Και οι δύο classifiers είχαν απόδοση κοντά στο 98% για νότες που παρήχθησαν συνολικά από 4 όργανα: κιθάρα, πιάνο, μαρίμπα και ακορντεόν ,για τόνος που ανήκαν σε ένα εύρος μιας οκτάβας. Παρόλο που αυτή η απόδοση είναι κάτι παραπάνω από ικανοποιητική ,ωστόσο πρέπει να συνυπολογίσουμε το γεγονός ότι τόσο τα δεδομένα εκπαίδευσης όσο και αξιολόγησης προήλθαν από ηχογράφηση των ίδιων εκτελεστών στις ίδιες συνθήκες. Επίσης, τα όργανα αυτά έχουν πολύ ξεχωριστές ακουστικές ιδιότητες μεταξύ τους ,και ως εκ τούτου a priori διαχωρισιμότητα.

Ο Langmead (1995) εκπαίδευσε ένα νευρωνικό δίκτυο χρησιμοποιώντας μεμονομένους μουσικούς φθόγγους εκτελεσμένους σε διάφορα όργανα. Τα δείγματά του τα ανέλυσε χρησιμοποιώντας κάποιο είδος ημιτονικής ανάλυσης. Αναφέρει ότι «Το εκπαιδευμένο δίκτυο έδειξε επιτυχία στην αναγνώριση μουσικής χροιάς (timbre recognition)» ,χωρίς όμως να παρέχει συγκεκριμένα αποτελέσματα.

Το πρόβλημα της αναγνώρισης μεμονωμένων νοτών (Isolated tones) προσεγγίστηκε με παραδοσιακές τεχνικές αναγνώρισης προτύπων από τουλάχιστον δύο ερευνητές. Ο Bourne (1972) εκπαίδευσε ένα μπεϊζιανό ταξινομητή (Bayesian classifier) με χαρακτηριστικά οπτικής από την σκοπιά του ανθρώπινου τρόπου αντίληψης. Αυτό συμπεριλάμβανε συνολικά φάσματα και ατάκες (onsets) από διαφορετικές «αρμονίες», τα οποία εξήγθησαν από 60 νότες από κλαρινέτα, γαλλικά κόρνα και τρομπέτες. 15 νότες χρησιμοποιήθηκαν για την αξιολόγηση του συστήματος, εκ των οποίων μόλις 8 δεν είχαν χρησιμοποιηθεί και στην εκπαίδευση. Το συγκεκριμένο σύστημα ταξινόμησε σωστά τις 14 από τις 15. Το 1998 ο Fujinaga εκπαίδευσε ένα k-nearest neighbor ταξινομητή με χαρακτηριστικά από 1338 φασματικά στιγμιότυπα, εξαγόμενα από 23 όργανα σε ένα μεγάλο εύρος μουσικών φθόγγων. Χρησιμοποιώντας ένα κριτήριο καταλληλότητας βασισμένο σε γενετικούς αλγόριθμους τύπου leave-one-out, προσπάθησε να διερευνήσει τους συνδυασμούς χαρακτηριστικών που πετύχαιναν μέγιστη διαχωρισιμότητα και έφτασε τελικά το σύστημα του σε μια απόδοση 50%.

Ο Kashino και η ομάδα του κατασκεύασαν μία σειρά από συστήματα τα οποία επιχειρούσαν πολυφωνική ακολουθία φθόγγων (pitch tracking). Το τελευταίο τους σύστημα ,το οποίο χρησιμοποιούσε την διαφοροποίηση στην αρμονικότητα και τον συγχρονισμό στην ατάκα, αναγνώριζε με επιτυχία την πηγή 42 νοτών από τσέμπαλο και φλάουτο παιγμένα διαδοχικά από ένα συνθέτη δειγμάτων (sample synthesizer) (Kashino & Tanaka,1992).

Τα επόμενα χρόνια κατασκευάστηκαν συστήματα που χρησιμοποιούν περισσότερα εξαγόμενα χαρακτηριστικά και μπορούν να αναγνωρίζουν διάφορα όργανα. Ο Kashino και η ομάδα του το 1995 κατασκεύασαν ένα recognizer για κλαρινέτο, φλάουτο, πιάνο, τρομπέτα και βιολί σε τυχαίες συγχορδίες χρησιμοποιώντας για την ταξινόμηση ένα ασυνήθιστο μετρικό με αξιοσημείωτη επιτυχία. Το πιο πρόσφατο σύστημα της ομάδας, μέσω προσαρμοστικών templates και πληροφορίας περιεχομένου καταφέρνει να κάνει αναγνώριση σε τρίο για βιολί, φλάουτο και πιάνο. (Kashino & Murase,

1997-1998). Δεδομένων των μουσικών φθόγγων το σύστημα αναγνώριζε το όργανο με επιτυχία 88.5%. Σε μια άλλη αναφορά προτείνεται η χρήση μιας ιεραρχικής οντολογίας για ήχους, σαν μέσο αναγνώρισης περισσότερων ηχητικών πηγών (Nakatani ry sl., 1997).
Μέχρι πολύ πρόσφατα δεν υπήρξαν δημοσιευμένες αναφορές για συστήματα αναγνώρισης μουσικών οργάνων τα οποία ήταν ικανά να βγάζουν απόφαση για πραγματικές μουσικές ηχογραφήσεις. Ωστόσο κάποια συστήματα φαίνεται ότι κατασκευάστηκαν τελευταία για αυτό το σκοπό, βασισμένα σε τεχνικές από την αναγνώριση φωνής και αναγνώριση ομιλητή.

Ο Brown (1997-1999) έχει περιγράψει έναν ταξινομητή δύο δρόμων ο οποίος αναγνωρίζει όμποε από σαξόφωνο. Ένα GMM (Gaussian Mixture Model) που δέχεται σαν είσοδο συντελεστές Cepstrum τύπου Q εκπαιδεύτηκε πάνω σε δείγματα συνολικού μεγέθους 1 λεπτού για κάθε όργανο. Σε ανεξάρτητα θορυβώδη δείγματα από εμπορικές ηχογραφήσεις το σύστημα ταξινόμησε σωστά το 94% . Το σύστημα επεκτάθηκε σε έναν ταξινομητή 4 δρόμων προσθέτοντας φλάουτο και κλαρινέτο και είχε ποσοστό επιτυχίας 84%.

Ο Dubnov με τον Rodet χρησιμοποίησε έναν διανυσματικό κβαντιστή βασισμένο σε συντελεστές cepstrum σαν την είσοδο σε έναν αλγόριθμό στατιστικής ταξινόμησης (statistical clustering). Το σύστημα εκπαιδεύτηκε με 18 μικρά δείγματα από ισάριθμα όργανα. Δεν αναφέρθηκαν αποτελέσματα ταξινόμησης, ωστόσο ο Vector quantizer φαίνεται πως κατά κάποιον τρόπο «αιχμαλώτησε» κάτι από το χώρο των ήχων τον οργάνων. Παρόλο που δεν υπήρχε αρκετή πληροφορία στη συγκεκριμένη αναφορά για να αξιολογήσουμε το σύστημα ,η μέθοδος έδειχνε να δίνει υποσχέσεις για το μέλλον. Το 2004 ο Rodet με τον Vincent έχουν ονομάσει τη μέθοδό τους ISA (Independent Subspace Analysis) . Στη μέθοδο αυτή αναπαριστούν το φάσμα πολυφωνικής μουσικής σαν σταθμισμένο μη γραμμικό άθροισμα βαρών από πιθανούς συνδυασμούς φασματικών ιδιοχώρων από κάθε όργανο. Η αξιολόγηση γίνεται από μη εμπορικά CDs σε σόλο όμως ηχογραφήσεις όπου πετυχαίνεται ποσοστό αναγνώρισης 90% για τα όργανα και 97% για οικογένειες οργάνων. Το ποσοστό αυτό πέφτει στο 85 % κατά την προσθήκη θορύβου στις ηχογραφήσεις. Τέλος το σύστημα δοκιμάζεται σε πολυφωνική ηχογράφηση όπου με δεδομένο τον συνολικό αριθμό των οργάνων του ορχηστικού συνόλου το σύστημα καταφέρνει να αποφανθεί για τα όργανα και να βρει τα pitch του καθενός. Όμως το συγκεκριμένο ελέγχεται μόνο σε ένα μέρος ενός κομματιού ,με αποτέλεσμα να μην μπορούμε να εξάγουμε γενικά συμπεράσματα για την απόδοση του συστήματος σε πολυφωνικές ηχογραφήσεις.
Ο Marques το 1999 κατασκεύασε ένα σετ από ταξινομητές 9-δρόμων (κλαρινέτο, φλάουτο, τσέμπαλο, εκκλησιαστικό όργανο, πιάνο, τρομπόνι, βιολί κ.α.) .

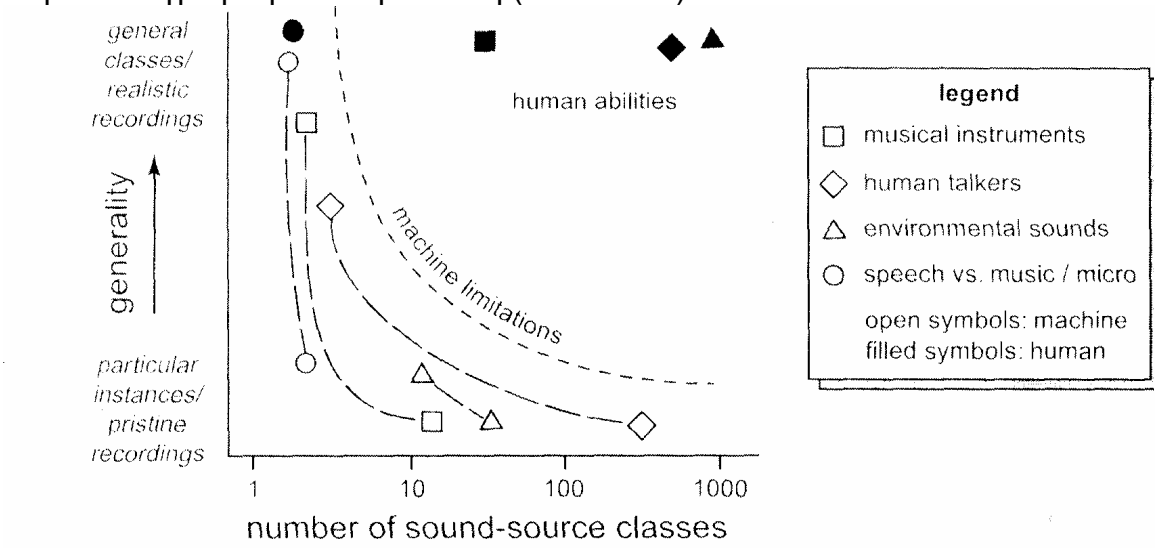
Χρησιμοποιώντας διάφορα σετ από χαρακτηριστικά και ταξινομητές και δείγματα από εμπορικές και ερασιτεχνικές ηχογραφήσεις κατέληξε στο ότι οι καλύτεροι ταξινομητές χρησιμοποιούσαν MFCC συντελεστές. Το ποσοστό επιτυχίας του συστήματος ήταν 72% για τις επαγγελματικές ηχογραφήσεις και 45% για τις μη επαγγελματικές. Ήταν προφανές ότι το σύστημα δεν είχε ικανότητα γενίκευσης που να συγκρίνεται με την ανθρώπινη.

Τέλος οι Essid, Richard και David το 2005 κατασκεύασαν ένα σύστημα για αναγνώριση σε σόλο εμπορικές ηχογραφήσεις. Το σύστημα

χρησιμοποιούσε μια πληθώρα χαρακτηριστικών, τα οποία όμως επιλέγοντας προς χρήση σύμφωνα με την Class Pairwise Feature Selection. Βασισμένοι στην τεχνική IRMFSP (Inertia Radio Maximization using Feature Space Projection) που είχε προταθεί αρχικά από τον Peeters οι συγγραφείς της δημοσίευσης δοκίμασαν να βρίσκουν κάθε φορά το υποσέτ των χαρακτηριστικών που είναι το ιδανικότερο για το διαχωρισμό των οργάνων ανά δύο. Ο ταξινομητής ήταν τύπου GMM. Με δείγματα από σόλο ηχογραφήσεις 10 οργάνων και 45 ταξινομητές (συνδυασμός των 10 ανά δύο) το σύστημα είχε επιτυχία μέχρι και 79%.

2.3.2.6 Συμπεράσματα

Οι άνθρωποι ακροατές έχουν τη δυνατότητα να προσαρμόζονται στις εκάστοτε –πολλές φορές δυσμενείς- συνθήκες του περιβάλλοντος με εξαιρετική ευκολία ,σε αντίθεση με τα μηχανικά συστήματα αναγνώρισης που είμαστε ακόμα σε θέση να κατασκευάσουμε. Προς το παρόν μπορούμε να κατασκευάσουμε τεχνητά συστήματα που μπορούν να αναγνωρίζουν πολλές πηγές ήχου, κυρίως όμως σε συνθήκες εργαστηρίου, και λιγότερες πηγές ήχου σε πιο δύσκολες συνθήκες. Η σύγκριση των ανθρώπων ομιλητών με τα μηχανικά συστήματα αναγνώρισης ως προς την απόδοση, φαίνεται στην παρακάτω γραφική αναπαράσταση (Εικόνα 2-1).



Εικόνα 2-1 Σύγκριση των ανθρώπων ομιλητών με τα μηχανικά συστήματα αναγνώρισης ως προς την απόδοση

Η πρόκληση για το μέλλον είναι να κατασκευάσουμε συστήματα τα οποία μπορούν να αναγνωρίσουν περισσότερες κλάσεις από ακουστικές πηγές με αυξημένη ικανότητα γενίκευσης και κάτω από συνθήκες της πολυπλοκότητας του σύγχρονου κόσμου. Στην αναγνώριση μουσικών οργάνων το απώτερο ζητούμενο είναι να κατασκευαστούν συστήματα τα οποία θα δέχονται σαν είσοδο ένα μουσικό κομμάτι παιγμένο από ένα

ορχηστικό σύνολο και θα έχουν σαν έξοδο το μέρος του κάθε οργάνου ξεχωριστά με τη μορφή μουσικού κειμένου.

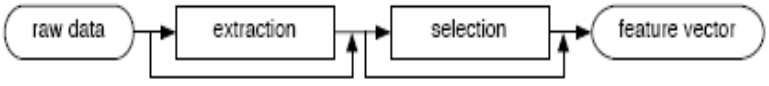
Κεφάλαιο 3 : Εξαγωγή χαρακτηριστικών

Στο παρόν κεφάλαιο παρέχεται μια εισαγωγή στις βασικές έννοιες της εξαγωγής χαρακτηριστικών, δίνοντας το απαραίτητο υπόβαθρο για την επεξεργασία σήματος. Παρατίθενται χαρακτηριστικά που χρησιμοποιούνται συχνά στην ανάλυση μουσικής.

3.1 Το διάνυσμα χαρακτηριστικών

3.1.1 Εισαγωγή

Η εξαγωγή χαρακτηριστικών (feature extraction) αποτελεί μια από τις δύο πλέον συνήθεις χρησιμοποιούμενες τεχνικές προεπεξεργασίας στην ταξινόμηση. Τα νέα χαρακτηριστικά παράγονται από τα ακατέργαστα δείγματα με την εφαρμογή ενός ή περισσότερων μετασχηματισμών. Η άλλη συνήθης τεχνική είναι η επιλογή χαρακτηριστικών (feature selection). Η επιλογή χαρακτηριστικών συναντάται σε δύο μορφές. Στην πρώτη εφαρμόζεται στο αρχικό σύνολο στοιχείων (στα ακατέργαστα δεδομένα). Στη δεύτερη εφαρμόζεται ακριβώς μετά την εξαγωγή χαρακτηριστικών. Ένα σύστημα ταξινόμησης δύναται να χρησιμοποιεί μία ή και τις δύο από αυτές τις τεχνικές. Θεωρητικά, είναι δυνατό να χρησιμοποιηθούν τα ακατέργαστα στοιχεία ως χαρακτηριστικά, εάν αυτά βέβαια είναι ήδη σε μορφή κατάλληλη για την ταξινόμηση. Στην πράξη βέβαια, αυτό είναι σχεδόν αδύνατο, είτε για λόγους χώρου και χρόνου επεξεργασίας, είτε κυρίως λόγω έλλειψης διαχωρισιμότητας (ικανότητας διαχωρισμού) σε δεδομένα τέτοιας μορφής. Αυτό συναντάται κατεξοχήν στα audio αρχεία καθώς το μέγεθος των ακατέργαστων δεδομένων είναι τεράστιο. Έτσι από ένα τέτοιο αρχείο, προσπαθούμε να εξάγουμε χαρακτηριστικά τέτοια ώστε να μειώσουμε το μήκος του διανύσματος των χαρακτηριστικών και να αποφύγουμε την εισαγωγή στοιχείων επαναληψιμότητας και επικάλυψης, ώστε να πετύχουμε τα δεδομένα εισαγωγής του ταξινομητή να είναι διαχωρίσιμα, ανάλογα με την ταξινόμηση που επιθυμούμε να πετύχουμε.



Σχήμα 3. 1 Πως κατασκευάζεται ένα διάνυσμα χαρακτηριστικών

Ένα *διάνυσμα χαρακτηριστικών* (feature vector) ***x*** είναι ένα απλό στοιχείο αναπαράστασης δεδομένων, το οποίο χρησιμοποιείται από τον αλγόριθμο ταξινόμησης και αποτελείται από ***d*** στοιχεία: $\boldsymbol{x} = (x_1, \ldots, x_d)$. Το x_i ονομάζεται *χαρακτηριστικό* , και *η διάσταση του χώρου* των *χαρακτηριστικών* καθορίζεται από το ***d***. Ένα διάνυσμα χαρακτηριστικών αποτελεί ένα σημείο στον χώρο των χαρακτηριστικών. Ένα *σετ προτύπων* το οποίο περιέχει ***n*** στοιχεία είναι:

$$X = \{x1, \ldots, xn\}$$

και το i - οστό διάνυσμα χαρακτηριστικών του $\boldsymbol{X}$ γράφεται:

$$x_i = (x_{i1}, \ldots, x_{id})$$

Στις περισσότερες περιπτώσεις ένα σετ προτύπων έχει τη μορφή $d \times n$ και ονομάζεται *πίνακας προτύπων.*

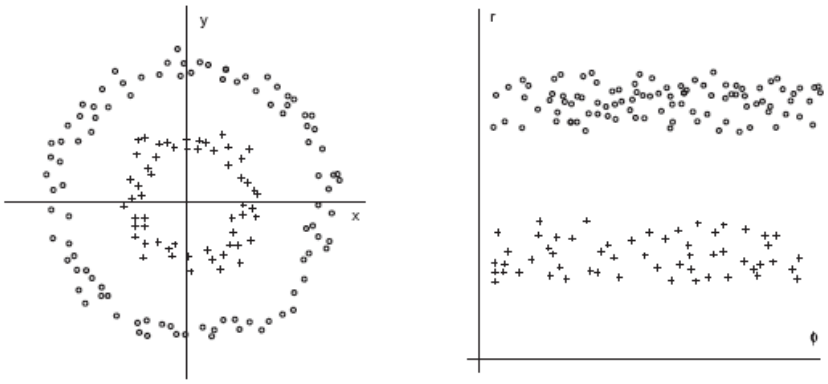
3.1.2 Χρησιμοποιώντας τα σωστά χαρακτηριστικά

Η χρήση των σωστών χαρακτηριστικών είναι μείζονος σημασίας για τη διαδικασία της ταξινόμησης. Ένας αλγόριθμός ταξινόμησης θα έχει πάντα ένα αποτέλεσμα, αλλά μία ελλιπής αναπαράσταση σε επίπεδο χαρακτηριστικών θα οδηγήσει σε ένα αποτέλεσμα το οποίο δεν αντανακλά την φύση των δεδομένων. Η επιλογή των χαρακτηριστικών χρειάζεται να πληρεί τα εξής δύο κριτήρια:

1. Αντικείμενα που θεωρούνται ίδια σύμφωνα με τα κριτήρια της ταξινόμησης θα πρέπει να αναπαρίστανται με σημεία πολύ κοντά στον χώρο των χαρακτηριστικών (feature space). Επίσης, η απόσταση μεταξύ των περιοχών όπου αλλάζει η απόφαση για τον ταξινομητή θα πρέπει να είναι όσο μεγαλύτερη γίνεται.
2. Τα χαρακτηριστικά θα πρέπει να εμπεριέχουν όλη τη σημαντική πληροφορία που περιέχεται στα δεδομένα. Τα δεδομένα θα πρέπει να απλοποιούνται χωρίς να χάνεται πληροφορία. Επιδιώκουμε έτσι κατά κύριο λόγο αντιστρεπτούς μετασχηματισμούς, έτσι ώστε να μπορούμε να επανακτήσουμε πλήρως τα δεδομένα μέσω των χαρακτηριστικών τους. Αυτή είναι μια ιδιότητα, για παράδειγμα, του μετασχηματισμού Fourier. Αυτό το δεύτερο κριτήριο κρίνεται απαραίτητο για την ανάλυση CASA , όχι όμως και για την περίπτωση που απαιτείται μόνο αναγνώριση. Τότε δεν είναι απαραίτητο να κατασκευάσουμε και μηχανισμό ανασύνθεσης των δεδομένων.

Ένα απλό παράδειγμα για τη σημαντικότητα της επιτυχημένης επιλογής χαρακτηριστικών μπορεί να φανεί στο σχήμα 3.2. Η πληροφορία αποτελείται από σημεία στον δισδιάστατο χώρο, τα οποία έχουν περίπου την ίδια απόσταση από την πηγή. Αν οι καρτεσιανές μεταβλητές (x,y) χρησιμοποιούνταν σαν χαρακτηριστικά , οι περισσότεροι αλγόριθμοι ταξινόμησης θα είχαν προβλήματα στην εύρεση των συνόρων των ομαδοποιήσεων προκειμένου για μια σωστή ταξινόμηση. Αντίθετα , αν χρησιμοποιούνταν για την αναπαράσταση των χαρακτηριστικών πολικές συντεταγμένες (r,φ), όπου r η ακτίνα και φ η γωνία, θα ήταν πολύ ευκολότερη η ταξινόμηση, και αυτό λόγω της συγκεκριμένης κατανομής των δειγμάτων στο χώρο των χαρακτηριστικών (feature space).
Η εύρεση των σωστών χαρακτηριστικών είναι μια πολύ δύσκολη διαδικασία η οποία πολλές φορές πραγματοποιείται μέσω δοκιμών. Συνήθως μάλιστα είναι σκόπιμη η μείωση της διαστασιμότητας του χώρου των χαρακτηριστικών προκειμένου να αυξηθεί η απόδοση του συστήματος.

(α) καρτεσιανές συντεταγμένες (β) πολικές συντεταγμένες



Σχήμα 3 2 Παράδειγμα που επιδεικνύει πως η επιλογή των χαρακτηριστικών επηρεάζει την ταξινόμηση. Για το ίδιο σετ χαρακτηριστικών των α, β, οι πολικές συντεταγμένες είναι πιθανό να έχουν μεγαλύτερη αποτελεσματικότητα από τις καρτεσιανές για την ταξινόμηση καθώς στο β αρκεί μια απλή γραμμή για το διαχωρισμό των περιοχών.

3.2. Το μουσικό σήμα ως άθροισμα ημιτόνων

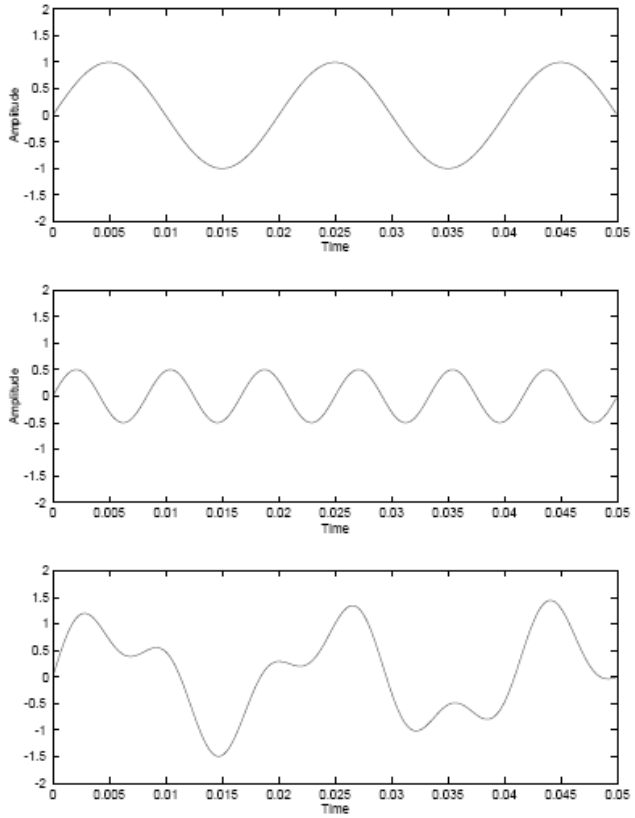
Παρακάτω γίνεται μια σύντομη αναφορά στις αρχές και το μαθηματικό υπόβαθρο των τεχνικών εξαγωγής χαρακτηριστικών.

3.2.1 Ημιτονοειδής υπέρθεση

Ο Jean Baptiste Joseph Fourier (1768-1830) ,γάλλος μαθηματικός και φυσικός, ήταν ο πρώτος που ανακάλυψε ότι ένα συνεχόμενο περιοδικό σήμα μπορεί να αναπαρασταθεί σαν το άθροισμα από ειδικά επιλεγμένα στοιχειώδη ημιτονοειδή σήματα. Η εξίσωση του ημιτόνου είναι ,ως γνωστό:

$$y(t) = A \sin(2\pi f t + \varphi) \qquad (3.1)$$

Με *A* συμβολίζεται το πλάτος του κύματος, *f* είναι η συχνότητα του σήματος και *φ* η γωνία φάσης του κύματος. Όλα τα σήματα, ανεξάρτητα με το πόσο σύνθετα είναι μπορούν να σχηματιστούν αθροίζοντας τέτοιες ημιτονικές συναρτήσεις με διαφορετική συχνότητα, πλάτος και φάση. Καθώς ο ήχος διαδίδεται σαν κύμα, αυτή η αρχή μπορεί να αποτελέσει τη βάση για την ανάλυση των ηχητικών κυμάτων (σχήμα 3.3). Η σύνθεση του σήματος από υπέρθεση ημιτόνων ονομάζεται *Σύνθεση Fourier.*



Σχήμα 3. 3 Υπέρθεση δύο ημιτόνων ,το ένα με συχνότητα 50 Hz και το άλλο με συχνότητα 120 Hz. Καθώς a = sin(2_50t) και b =0.5 sin(2_120t) στο τρίτο διάγραμμα προκύπτει μέσω πρόσθεσης το σήμα c = a + b

3.2.2 Μετασχηματισμός Fourier

Η αντίστροφη διαδικασία από τη σύνθεση Fourier είναι ο διαχωρισμός ενός σήματος σε στοιχειώδη ημιτονικά σήματα και ονομάζεται *Ανάλυση Fourier*. Η *Ανάλυση Fourier* μας επιτρέπει να διαπιστώσουμε ποιες συχνότητες είναι παρούσες στο σήμα και πόσο δυνατή είναι η επίδραση από κάθε μία από αυτές. Η μαθηματική έκφραση της Ανάλυσης Fourier ονομάζεται μετασχηματισμός Fourier και δίνεται σαν :

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{-j2\pi f\, t}\, dt, \qquad (3.2)$$

και ο αντίστροφος μετασχηματισμός σαν:

$$x(t) = \int_{-\infty}^{+\infty} \boldsymbol{X}(f)e^{j2\pi f\, t}\, df, \qquad (3.3)$$

Σε αυτή την εξίσωση το x(t) είναι συνάρτηση χρόνου και το X(f) είναι η αντίστοιχη συνάρτηση συχνότητας.

Με εφαρμογή του μετασχηματισμού Fourier σε ένα σήμα , το μεταφέρουμε από το πεδίο του χρόνου στο πεδίο της συχνότητας. Ο μετασχηματισμός Fourier είναι αντιστρεπτός, έτσι μπορούμε να επιλέξουμε οποιονδήποτε από τους δύο τρόπους για να αναπαραστήσουμε το ίδιο σήμα. Τροποποιώντας το σήμα με ένα τρόπο στο ένα πεδίο ,αυτόματα τροποποιείται και στο άλλο. Για παράδειγμα , μία συνέλιξη στο πεδίο του χρόνου ισοδυναμεί με πολλαπλασιασμό στο πεδίο της συχνότητας και αντιστρόφως.

Ο μετασχηματισμός Fourier μπορεί να χρησιμοποιηθεί θεωρητικά σε ένα σήμα απείρου μήκους. Παρέχει πληροφορία για το φασματικό περιεχόμενο ολόκληρου του σήματος ενώ παράλληλα δεν παρέχει καμία πληροφορία για την εξέλιξη στο πεδίο του χρόνου. Αυτό δεν αποτελεί πρόβλημα για περιοδικά σήματα όπως αυτά του παραδείγματος στο (σχήμα 3.3), αλλά στην περίπτωση μη περιοδικών σημάτων, όπως είναι για παράδειγμα η μουσική , απαιτείται μια διαφορετική αντιμετώπιση, *ο μετασχηματισμός Fourier βραχέως χρόνου (Short-Time Fourier Transform STFT).*

3.2.3 Μετασχηματισμός Fourier βραχέως χρόνου (STFT)

Ο Μετασχηματισμός Fourier βραχέως χρόνου (STFT) κατακερματίζει το σήμα σε μικρά συνεχόμενα κομμάτια (frames) και πραγματοποιεί μετασχηματισμό Fourier σε καθένα από αυτά , αναλύοντας έτσι τη μεταβολή του σήματος και ως προς το χρόνο. Αυτό όμως δημιουργεί καινούργιες δυσκολίες: Αφού ο μετασχηματισμός Fourier απαιτεί το σήμα να είναι άπειρο , κάθε κομμάτι του σήματος θα πρέπει να μετατραπεί πρώτα σε ένα σήμα με άπειρο μήκος. Το απότομο κόψιμο στο τέλος και η απότομη έναρξη στην αρχή ενός frame έχει σαν αποτέλεσμα κατά τον μετασχηματισμό την εμφάνιση ημιτόνων με πολύ μεγάλες συχνότητες που εκφράζουν τις γρήγορες μεταβολές τα οποία δεν υπάρχουν στο αρχικό σήμα.

Η λύση στο παραπάνω πρόβλημα έρχεται από την διαδικασία της παραθύρωσης (windowing). Κάθε frame πολλαπλασιάζεται με μια συνάρτηση παραθύρωσης (windowing function) η οποία σταδιακά μειώνει το σήμα σε κάθε frame ως την τιμή 0. Παράλληλα επιλέγεται ένα ποσοστό επικάλυψης μεταξύ των frames έτσι ώστε μαζί με την κατάλληλη συνάρτηση παραθύρωσης ο μετασχηματισμός να γίνεται με τη μέγιστη ακρίβεια. Μία πολύ γνωστή συνάρτηση παραθύρωσης είναι το *παράθυρο Hamming* το οποίο διατηρεί μια πολύ καλή ισορροπία μεταξύ της ταχύτητας υπολογιστικής επεξεργασίας και του μικρού σφάλματος. Άλλα παράθυρα είναι το gammatone ,το τριγωνικό ,το τετραγωνικό κ.α.

Η μαθηματική σχέση για τον μεταχηματισμό Fourier βραχέως χρόνου είναι:

$$X(f,t) = \int_{-\infty}^{\infty} h(t'-t)x(t)e^{-j2\pi ft}dt \qquad (3.4)$$

3.2.4 Διακριτός Μετασχηματισμός Fourier

Οι τεχνικές που αναφέρθησαν μέχρι τώρα χρησιμεύουν για το μετασχηματισμό συνεχούς σήματος. Τα υπολογιστικά συστήματα όμως διαχειρίζονται τη μουσική σαν ένα διακριτό σήμα. Έτσι απαιτείται ένα εργαλείο για την πραγματοποίηση του μετασχηματισμού στο πεδίο της συχνότητας για σήματα διακριτού χρόνου. Τέτοιο εργαλείο είναι ο Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Transform ,DFT).

Στην περίπτωση μιας συνάρτησης $f(t) \rightarrow f(t_k)$,για συντομία γράφουμε : $f_k \equiv f(t_k)$, όπου $t_k \equiv k\Delta$, με $k = 0, \ldots, N-1$. Επιλέγοντας ρυθμό δειγματοληψίας

$$\upsilon_n = \frac{n}{N\Delta} \qquad (3.5)$$

με $n = -N/2, \ldots, 0, \ldots, N/2$. Ο μετασχηματισμός DFT δίνεται από

$$F_n = \frac{1}{N}\sum_{k=0}^{N-1} f_k e^{-j2\pi nk/N} \qquad (3.6)$$

και ο αντίστροφος μετασχηματισμός από :

$$f_k = \sum_{n=0}^{N-1} F_n e^{j2\pi nk/N}. \qquad (3.7)$$

Εδώ θα πρέπει να σημειωθεί ότι η απόκριση συχνότητας που προκύπτει από τον DFT είναι συχνά περίπλοκη, παρόλο που το αυθεντικό σήμα είναι απόλυτα αληθινό.

Υπάρχουν πολλοί τρόποι για τον υπολογισμό του DFT , όπως για παράδειγμα η λύση γραμμικών εξισώσεων . Ο πιο ευρείας χρήσης τρόπος και ταυτόχρονα ο πιο αποδοτικός είναι ο FFT, ο οποίος θα περιγραφεί σε επόμενο κεφάλαιο.

3.3 Το ανθρώπινο ακουστικό σύστημα

3.3.1 Φυσικά χαρακτηριστικά και χαρακτηριστικά αντίληψης

Συνήθως, τα χαρακτηριστικά που χρησιμοποιούνται για οποιαδήποτε ταξινόμηση που αφορά στον ήχο διακρίνονται σε δύο κατηγορίες : τα φυσικά χαρακτηριστικά και τα χαρακτηριστικά αντίληψης. Τα φυσικά χαρακτηριστικά βασίζονται σε μαθηματική και στατιστική ανάλυση των ιδιοτήτων του σήματος ήχου. Παραδείγματα για φυσικά χαρακτηριστικά είναι η βασική συχνότητα , η στιγμιαία και συνολική ενέργεια και ο ρυθμός διασταυρώσεων από το 0 (zero crossing rate). Τα χαρακτηριστικά αντίληψης βασίζονται στον τρόπο με τον οποίο ο άνθρωπος αντιλαμβάνεται τον ήχο , όπως για παράδειγμα το pitch , η χροιά (timbre) και ο ρυθμός.

Είναι φανερό πως όλα τα χαρακτηριστικά αντίληψης σχετίζονται άμεσα ή έμμεσα με τα φυσικά χαρακτηριστικά. Κάποια μπορούν να αντιστοιχισθούν κατ' ευθείαν: Το πλάτος του σήματος του ήχου σχετίζεται με την ηχηρότητα, η βασική συχνότητα σχετίζεται με το pitch. Ωστόσο , εξετάζοντας με περισσότερη λεπτομέρεια σε αυτή την αντιστοίχηση συμπεραίνουμε ότι πρόκειται για επιπόλαιη προσέγγιση. Για παράδειγμα ,η υποκειμενική ηχηρότητα που εκλαμβάνει ο ακροατής εξαρτάται από το φασματικό περιεχόμενο του σήματος, ενώ η έννοια του pitch εξαρτάται επίσης τόσο από το περιεχόμενο του σήματος σε αρμονικές συχνότητες όσο και από τον ακουστικό σωλήνα και τη διαμόρφωση που γίνεται στο ίδιο το αυτί. Πολλά από τα χαρακτηριστικά αντίληψης είναι δύσκολο ως αδύνατο να περιγραφούν μαθηματικά , καθώς εξαρτώνται παράλληλα από αλληλοεξαρτούμενες φυσικές παραμέτρους. Γεγονός πάντως είναι πως ο άνθρωπος έχει αξιοθαύμαστες ικανότητες για ταξινόμηση που αφορά στη μουσική. Δεν είναι τυχαίο άλλωστε που τα τελευταία χρόνια γίνεται μεγάλη προσπάθεια , ο τρόπος που ο άνθρωπος αντιλαμβάνεται τον ήχο να ενσωματωθεί στα μηχανικά συστήματα αναγνώρισης ,ακόμα κι αν αυτό συνεπάγεται μεγάλο υπολογιστικό κόστος και απογοητευτικά αποτελέσματα.

Έτσι, ο σύνθετος τρόπος της ανθρώπινης αντίληψης και η αποκρυπτογράφηση αυτής τροφοδοτεί το πεδίο έρευνας της αναγνώρισης της ηχητικής πηγής και θεωρείται ως η βάση για τη δημιουργία συστημάτων στο μέλλον που θα επιλύουν πλήρως το πρόβλημα.

Στο παρακάτω κεφάλαιο γίνεται μια περιγραφή του ανθρώπινου ακουστικού συστήματος , και παρατίθενται έρευνες της ψυχοακουστικής σχετικά με την ανθρώπινη αντίληψη του ήχου. Επιπλέον, παρουσιάζονται με συντομία τα μουσικά όργανα κατά οικογένειες σύμφωνα με τις ακουστικές τους ιδιότητες. Τέλος, παρουσιάζονται οι μαθηματικές αναπαραστάσεις για τα φυσικά χαρακτηριστικά και τα χαρακτηριστικά της ανθρώπινης αντίληψης που έχουν χρησιμοποιηθεί κατά την προσπάθεια δημιουργίας αποδοτικών ταξινομητών για μουσικά όργανα.

3.3.2 Το ανθρώπινο αυτί

Η μελέτη της κατασκευής του αυτιού είναι μια μελέτη φυσιολογίας. Η μελέτη της ανθρώπινης αντίληψης του ήχου αποτελεί όμως κεφάλαιο της

ψυχολογίας. Η ψυχοακουστική είναι ένας συνολικός όρος που περιλαμβάνει την φυσική κατασκευή του αυτιού, τις διαδρομές κίνησης του ήχου, την αντίληψη του ήχου, και τις σχέσεις μεταξύ τους.

Το ηχητικό κύμα-ερέθισμα που φτάνει στο αυτί προκαλεί μηχανικές κινήσεις που έχουν σαν αποτέλεσμα λειτουργίες των νεύρων που καταλήγουν στον εγκέφαλο και δημιουργούν μια αίσθηση. Το επόμενο ερώτημα είναι: «Με ποιο τρόπο αναγνωρίζονται και ερμηνεύονται αυτοί οι ήχοι;». Παρά την έντονη ερευνητική δραστηριότητα σε όλες τις απόψεις της ανθρώπινης ακοής, οι γνώσεις μας εξακολουθούν να είναι απελπιστικά ατελείς.

3.3.2.1 Ευαισθησία του αυτιού

Η λεπτότητα και η ευαισθησία της ακοής μας μπορεί να φαίνεται έντονα με ένα μικρό πείραμα:

«Ανοίγεται αργά η ογκώδης πόρτα ενός ανηχοϊκού θαλάμου, και φαίνονται οι εξαιρετικά παχείς τοίχοι, και οι σφήνες από υαλόνημα μήκους ενός μέτρου, με τις κορυφές τους προς τα μέσα, που καλύπτουν όλους τους τοίχους, την οροφή, και αυτό που θα μπορούσε να ονομαστεί πάτωμα, εκτός από το ότι περπατούμε πάνω σε ένα ανοικτό ατσάλινο πλέγμα.

Φέρνουμε μια καρέκλα και καθόμαστε. Το πείραμα χρειάζεται χρόνο, και επειδή έχουμε ενημερωθεί προηγούμενα, περιμένουμε, μετρώντας τις σφήνες από υαλόνημα για να περάσει η ώρα. Ο χώρος είναι αλλόκοτος. Η θάλασσα του ήχου και των θορύβων της ζωής και των δραστηριοτήτων όπου συνήθως βρισκόμαστε, και την οποία κανονικά σπάνια συνειδητοποιούμε, είναι τώρα εμφανής με την απουσία της.

Η σιωπή μας πιέζει(σχεδόν απόλυτα), για 10 λεπτά, και περνά μισή ώρα. Ανακαλύπτονται νέοι ήχοι, ήχοι που προέρχονται μέσα από το ίδιο μας το σώμα. Στην αρχή, το δυνατό χτύπημα της καρδιάς μας, που μόλις συνέρχεται από την νέα κατάσταση. Περνά μια ώρα. Ακούγεται το αίμα μέσα στις φλέβες μας. Τέλος, αν τα αυτιά μας είναι τεντωμένα, η υπομονή μας ανταμείβεται από έναν παράξενο σφυριχτό ήχο μεταξύ των χτυπημάτων της καρδιάς και την ροή του αίματος. Τι είναι αυτό; Είναι ο ήχος των σωματιδίων του αέρα που χτυπούν στα τύμπανα μας. Η κίνηση των τύμπανων που είναι αποτέλεσμα αυτού του σφυριχτού ήχου είναι απίστευτα μικρή, μόνο το 1/100 του εκατομμυριοστού του εκατοστού του μέτρου, ή το 1/10 της διαμέτρου του μορίου του υδρογόνου!»

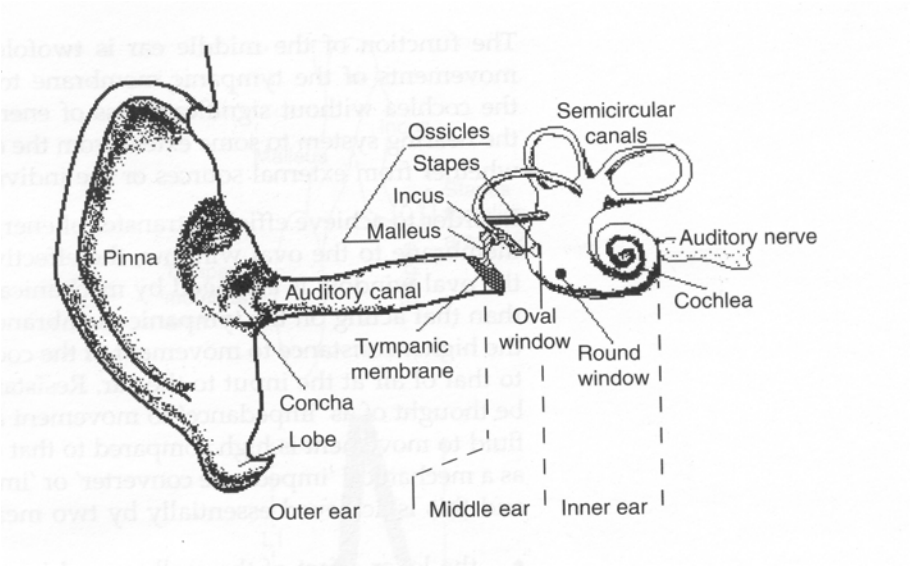
Το ανθρώπινο αυτί δεν μπορεί να ανιχνεύσει ήχους απαλότερους από την ροή των μορίων του αέρα στο τύμπανο. Αυτό είναι και το κατώφλι της ακοής. Δεν υπάρχει λόγος να έχουμε περισσότερο ευαίσθητα αυτιά, επειδή οποιοσδήποτε ήχος μικρότερης στάθμης θα πνιγόταν από τον θόρυβο των σωματιδίων του αέρα. Αυτό σημαίνει ότι η μεγαλύτερη ευαισθησία της ακοής μας μόλις και ταιριάζει με τους απαλότερους ήχους που είναι δυνατοί σε ένα αέριο μέσο. Είναι τυχαίο; Είναι θέμα προσαρμογής; Είναι θέμα σχεδίασης;

Στο άλλο άκρο, τα αυτιά μας μπορούν να αντιδράσουν στον βρυχηθμό του κανονιού, στον θόρυβο της εκτόξευσης πυραύλου, ή σε ένα αεριωθούμενο που απογειώνεται. Ειδικά προστατευτικά χαρακτηριστικά του αυτιού προστατεύουν τον ευαίσθητο μηχανισμό από ζημιά από όλους τους θορύβους εκτός από τους εξαιρετικά έντονους.

3.3.2.2 Η ανατομία του αυτιού

Τα τρία κύρια τμήματα του ανθρώπινου ακουστικού συστήματος, που φαίνεται στην εικόνα 3 -1, είναι το έξω αυτί (outer ear), το μέσο αυτί (middle ear), και το έσω αυτί (inner ear). Το έξω αυτί αποτελείται από το πτερύγιο (pinna) και από το ακουστικό κανάλι (Auditory meatus). Το ακουστικό κανάλι

τελειώνει στην ακουστική μεμβράνη ή τύμπανο (tympanic membrane). Το μέσο αυτί είναι ένας χώρος γεμάτος αέρα που διασχίζεται από τα τρία μικρά οστά (ossicles) που ονομάζονται σφύρα (malleus), άκμονας (incus), και αναβολέας (stapes). Η σφύρα είναι κολλημένη στο τύμπανο και ο αναβολέας κολλημένος στο ελλειψοειδές παράθυρο (oval window) του έσω αυτιού. Όλα μαζί αυτά τα τρία οστά σχηματίζουν μια μηχανική σύνδεση μοχλού μεταξύ του τύμπανου το οποίο ενεργοποιείται από τον αέρα και του κοχλία του έσω αυτιού ο οποίος είναι γεμάτος υγρό. Το έσω αυτί τελειώνει στο ακουστικό νεύρο(auditory nerve), το οποίο στέλνει ερεθίσματα στον εγκέφαλο.



Εικόνα 3-1 Η ανατομία του ανθρώπινου αυτιού

3.3.2.2.1 Το πτερύγιο: Κατευθυντικός κωδικοποιητής ήχου

Στην αρχαιότητα, το πτερύγιο θεωρούνταν είτε σαν υποτυπώδες όργανο είτε σαν απλή συσκευή συλλογής ήχων. Και πράγματι, είναι συσκευή συλλογής ήχων. Το πτερύγιο προσφέρει διαφοροποίηση, των ήχων που προέρχονται από εμπρός σε σχέση με τους ήχους που προέρχονται από πίσω. Αν βάλουμε την παλάμη μας πίσω από το αυτί αυξάνεται το ενεργό μέγεθος του πτερυγίου και με τον τρόπο αυτό αυξάνεται και η φαινομενική ηχηρότητα κατά ποσό που εξαρτάται από τη συχνότητα. Για τις σημαντικές συχνότητες ομιλίας (2.000 μέχρι 3.000 Hz), η πίεση ήχου στο τύμπανο αυξάνεται κατά περίπου 5 dB. Αυτή η διαφοροποίηση εμπρός/πίσω είναι η ηπιότερη συνεισφορά του πτερυγίου.

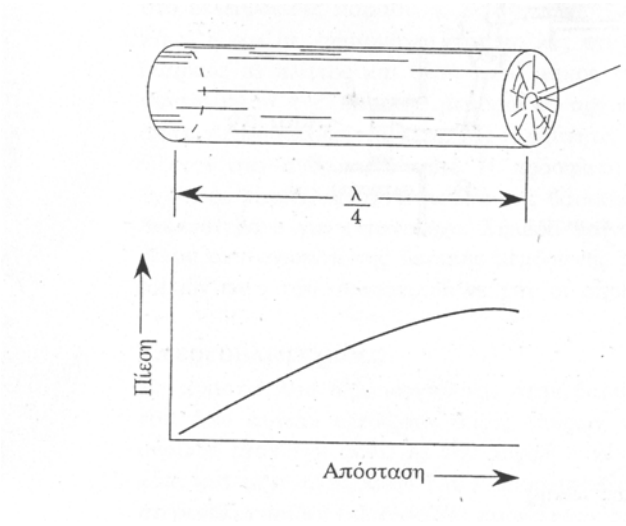
Οι πρόσφατες έρευνες έδειξαν ότι το πτερύγιο εκτελεί μια πολύ αποφασιστική λειτουργία χαράσσοντας κατευθυντικές πληροφορίες σε όλους τους ήχους που συλλαμβάνονται από το αυτί. Αυτό σημαίνει ότι πληροφορίες σχετικές με την πηγή του ήχου προστίθενται στο περιεχόμενο του ήχου έτσι ώστε η τελική ακουστική πίεση στο τύμπανο να δίνει την δυνατότητα στον

εγκέφαλο να ερμηνεύει και το περιεχόμενο του ήχου και την κατεύθυνση από την οποία αυτός προέρχεται.

Ένα απλό ψυχοακουστικό πείραμα μπορεί να δείξει τον τρόπο με τον οποίο από τις απλές μεταβολές του ήχου που πέφτει στο αυτί έχουμε σαν αποτέλεσμα υποκειμενικές εντυπώσεις κατεύθυνσης. Με ένα ακουστικό στο ένα αυτί, ακούμε με ένα φίλτρο ρυθμιζόμενου ανοίγματος έναν τυχαίο θόρυβο με πλάτος ζώνης μιας οκτάβας και με κεντρική συχνότητα τα 8 kHz. Αν ρυθμίσουμε το φίλτρο στα 7.2 kHz ο ήχος θα φαίνεται ότι προέρχεται από πηγή στο ύψος του παρατηρητή. Όταν το άνοιγμα ρυθμιστεί στα 8 kHz ο ήχος φαίνεται ότι έρχεται από επάνω. Όταν το άνοιγμα είναι στα 6.3 kHz ο ήχος φαίνεται να έρχεται από κάτω. Το πείραμα αυτό δείχνει ότι το ανθρώπινο ακουστικό σύστημα εξάγει πληροφορίες κατεύθυνσης από το σχήμα του φάσματος του ήχου στο τύμπανο.

3.3.2.2.2Το κανάλι του αυτιού

Το κανάλι του αυτιού αυξάνει και αυτό την ηχηρότητα των ήχων που το διαπερνούν. Στο Σχ. 3.4, το κανάλι του αυτιού, που έχει μέση διάμετρο περίπου 0.7 cm και μήκος περίπου 3 cm, φαίνεται σε σχηματική μορφή σαν ευθύ και με ομογενή διάμετρο σε όλο το μήκος του. Από ακουστικής πλευράς, η προσέγγιση είναι λογική. Έχουμε ένα αγωγό σε σχήμα σωλήνα που στο μέσα άκρο του κλείνει με το τύμπανο.

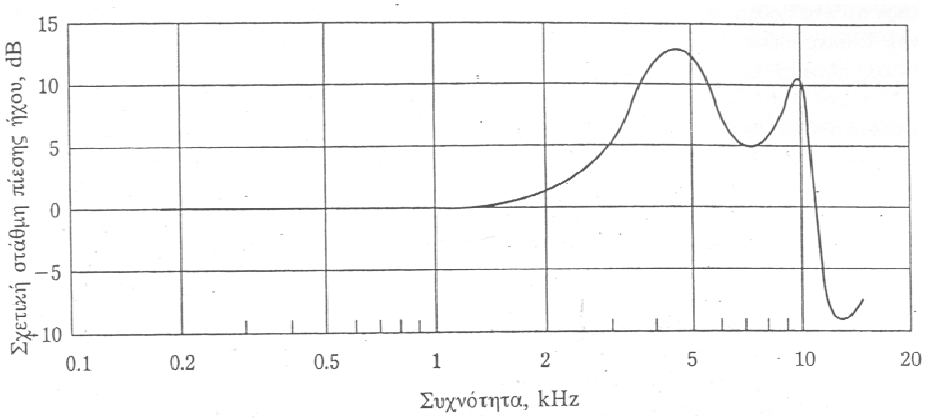


Σχήμα 3. 4 Το ακουστικό κανάλι, κλειστό στο ένα άκρο από το τύμπανο, λειτουργεί σαν "σωλήνας μουσικού οργάνου" ενός τετάρτου μήκους κύματος. Ο συντονισμός δίνει ακουστική ενίσχυση νια τις σημαντικές συχνότητες φωνής

Όταν η επιστήμη της ακουστικής βρισκόταν στα πρώτα της βήματα, οι πρώτοι ερευνητές μελέτησαν έντονα τους σωλήνες των μουσικών οργάνων. Η ακουστική ομοιότητα του καναλιού του αυτιού με σωλήνα μουσικού οργάνου δεν διέφυγε από τους πρώτους ερευνητές στον τομέα αυτό. Το φαινόμενο του συντονισμού στο κανάλι του αυτιού αυξάνει την πίεση ήχου στο τύμπανο σε ορισμένες συχνότητες. Το μέγιστο βρίσκεται κοντά στην συχνότητα στην οποία ο σωλήνας των 3 cm έχει ένα τέταρτο μήκους κύματος, περίπου 3.000 Hz.

Το ακουστικό κανάλι, κλειστό στο ένα άκρο από το τύμπανο(σχήμα 3.4), λειτουργεί σαν "σωλήνας μουσικού οργάνου" ενός τετάρτου μήκους κύματος. Ο συντονισμός δίνει ακουστική ενίσχυση νια τις σημαντικές συχνότητες φωνής Στο Σχ. 3.5 φαίνεται η αύξηση της πίεσης ήχου στο τύμπανο σε σχέση με την πίεση στο άνοιγμα του καναλιού του αυτιού. Παρατηρούμε μια πρωτεύουσα κορυφή περίπου στα 3.000 Hz που προκαλείται από το φαινόμενο συντονισμού του σωλήνα στο τέταρτο του μήκους κύματος. Ο πρωτεύων

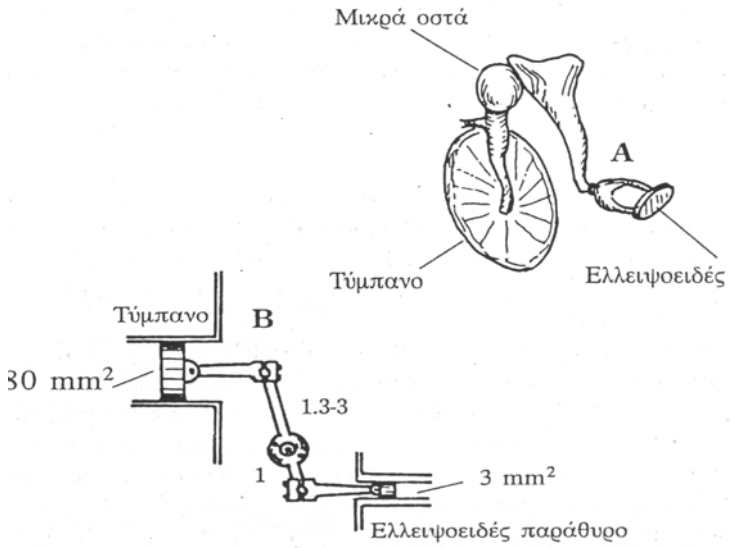
συντονισμός του σωλήνα αυξάνει την πίεση ήχου στο τύμπανο περίπου κατά 12 dB στον κύριο συντονισμό περίπου στα 3.000 Hz. Υπάρχει ένας δευτερεύων συντονισμός κοντύτερα στα 9.000 Hz με μικρότερη μέγιστη πίεση.



Σχημα 3.5 Η συνάρτηση μεταφορά (απόκριση συχνότητας) του καναλιού του αυτιού. Πρόκειται για μια σταθερή συνιστώσα που προστίθεται με κάθε κατευθυντικά κωδικοποιημένο ήχο που φτάνει στο τύμπανο.

3.3.2.2.3 Το μέσο αυτί

`Η μετάδοση της ενέργειας του ήχου από ένα αραιό μέσο όπως είναι ο αέρας σε πυκνό μέσο όπως το νερό αποτελεί σοβαρό πρόβλημα. Αν δεν υπάρχει πολύ εξειδικευμένος εξοπλισμός, ο ήχος από τον αέρα ανακλάται στο νερό όπως το φως σε καθρέπτη. Το ζήτημα καταλήγει σε προσαρμογή αντιστάσεων, και στην περίπτωση μας ο λόγος αντιστάσεων είναι περίπου 4.000 : 1. Ας σκεφτούμε πόσο ικανοποιητικό θα ήταν να οδηγούμε το πηνίο φωνής ενός μεγαφώνου 1 Ohm με έναν ενισχυτή που έχει αντίσταση εξόδου 4.000 Ohm. Είναι φανερό ότι δεν πρόκειται να μεταδοθεί αρκετή ισχύς.



Σχημα 3.6 (A) Τα μικρά οστά (σφύρα, άκμονας, αναβολέας) του μέσου αυτιού, που μεταδίδουν μηχανικές ταλαντώσεις του τυμπάνου στο ελλειψοειδές παράθυρο του κοχλία. (B) Μηχανικό ανάλογο της λειτουργίας προσαρμογής αντιστάσεων στο μέσο

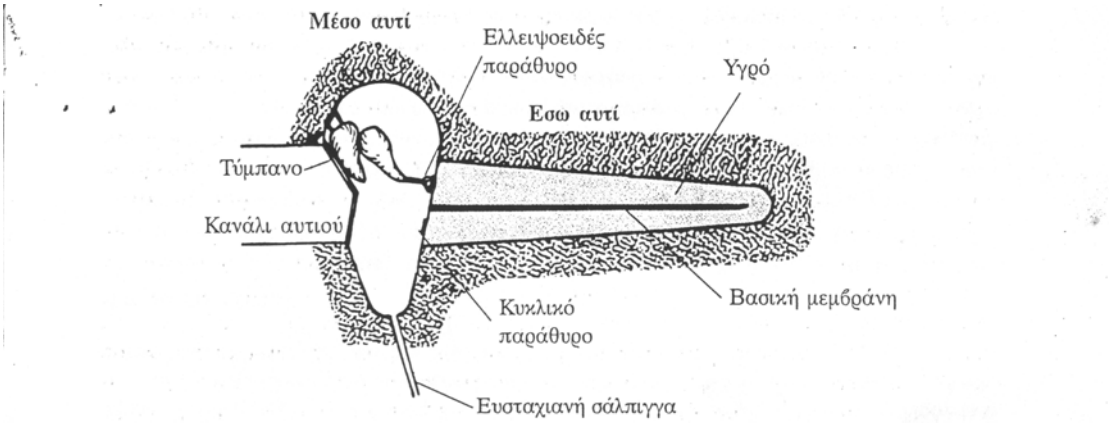
αυτί. Η διαφορά επιφάνειας μεταξύ τύμπανου και ελλειψοειδούς παραθύρου, μαζί με την μηχανική σύνδεση υποβιβασμού, προσαρμόζουν την κίνηση του τυμπάνου που ενεργοποιείται από τον αέρα στο ελλειψοειδές παράθυρο με υγρό

Το θέμα είναι να μεταφερθεί με μέγιστη απόδοση η ασθενική ενέργεια που παριστάνεται από την κίνηση ταλάντωσης ενός μάλλον εύθραυστου διαφράγματος, στο υγρό του έσω αυτιού. Στο Σχ. 3.6 προτείνεται η διπλή λύση. Τα τρία μικρά οστά σχηματίζουν μια μηχανική σύνδεση μεταξύ του τυμπάνου και του ελλειψοειδούς παραθύρου, το οποίο βρίσκεται σε απαλή επαφή με το υγρό του έσω αυτιού. Το πρώτο από τα τρία οστά, η σφύρα, είναι συνδεδεμένο με το τύμπανο. Το τρίτο, ο αναβολέας, στην πραγματικότητα αποτελεί τμήμα του ελλειψοειδούς παραθύρου. Στη σύνδεση αυτή υπάρχει κίνηση μοχλού με αναλογία από 1.3 : 1 μέχρι 3 : 1. Δηλαδή, η κίνηση του τυμπάνου ελαττώνεται κατά το μέγεθος αυτό στο ελλειψοειδές παράθυρο του έσω αυτιού.

Το πρόβλημα της προσαρμογής ήχου στον αέρα σε ήχο στο υγρό του έσω αυτιού λύνεται όμορφα με την μηχανική του μέσου αυτιού. Απόδειξη ότι η προσαρμογή αντιστάσεων μαζί με την ενίσχυση συντονισμού στο Σχ. 3.5 λειτουργούν αποτελεσματικά, αποτελεί το γεγονός ότι μια κίνηση διαφράγματος μικρή όσο οι μοριακές διαστάσεις δίνει την αντίληψη κατωφλίου.

Στο Σχ. 3.7 δίνεται ένα σχηματικό διάγραμμα του αυτιού. Το κωνικό τύμπανο στο εσωτερικό άκρο του ακουστικού καναλιού σχηματίζει την μια πλευρά του μέσου αυτιού που είναι γεμάτο αέρα. Το μέσο αυτί συνδέεται με το άνω μέρος του λαιμού πίσω από τη ρινική κοιλότητα με την ευσταχιανή σάλπιγγα. Το τύμπανο λειτουργεί σαν σύστημα «ακουστικής ανάρτησης», λειτουργώντας ενάντια στην υποχωρητικότητα του αέρα που είναι παγιδευμένος στο μέσο αυτί. Η ευσταχιανή σάλπιγγα είναι κατάλληλα μικρή ώστε να μη καταστρέφει αυτή την υποχωρητικότητα. Το κυκλικό παράθυρο διαχωρίζει το γεμάτο αέρα μέσο αυτί από το πρακτικά ασυμπίεστο υγρό του έσω αυτιού.

Η ευσταχιανή σάλπιγγα εξυπηρετεί μια δεύτερη λειτουργία εξισώνοντας την στατική πίεση του αέρα στο μέσο αυτί με την εξωτερική ατμοσφαιρική πίεση έτσι ώστε να λειτουργούν σωστά το τύμπανο και οι ευαίσθητες μεμβράνες του έσω αυτιού. Κάθε φορά που καταπίνουμε, οι ευσταχιανές σάλπιγγες ανοίγουν, ισορροπώντας την πίεση στο μέσο αυτί. Όταν ένα αεροπλάνο (τουλάχιστο από αυτά που δεν έχουν καμπίνα υπό πίεση) αλλάζει απότομα ύψος, υπάρχει περίπτωση οι επιβάτες να εμφανίσουν προσωρινή κώφωση ή πόνο μέχρις ότου η πίεση στο μέσο αυτί να ισορροπήσει με κατάποση. Στην πραγματικότητα η ευσταχιανή σάλπιγγα έχει και μια τρίτη λειτουργία ανάγκης, την απαγωγή των υγρών σε περίπτωση μόλυνσης του μέσου αυτιού.



Σχήμα 3.7 Εξιδανικευμένο σχέδιο του ανθρώπινου αυτιού όπου φαίνεται ξετυλιγμένος ο γεμάτος υγρό κοχλίας. Ο ήχος που μπαίνει στο κανάλι του αυτιού προκαλεί την ταλάντωση του τύμπανου. Αυτή η ταλάντωση μεταδίδεται στον κοχλία μέσω της μηχανικής σύνδεσης του μέσου αυτιού. Ο ήχος αναλύεται μέσω στάσιμων κυμάτων που δημιουργούνται στην βασιλική (βασική) μεμβράνη.

3.3.2.2.4 Το έσω αυτί

Μέχρι το σημείο αυτό έχουμε εξετάσει μόνο τους ακουστικούς ενισχυτές και τα χαρακτηριστικά μηχανικής προσαρμογής της αντίστασης του μέσου αυτιού. Όλα αυτά είναι σχεδόν καλά κατανοητά. Η πολύπλοκη λειτουργία του κοχλία εξακολουθεί ακόμη να βρίσκεται μέσα σε μυστήριο, αλλά η εκτενής έρευνα μας δίνει σταθερά νέες γνώσεις.

Στην εικόνα 3-1 φαίνεται πόσο κοντά βρίσκονται τα τρία, κάθετα μεταξύ τους, ημικυκλικά κανάλια του μηχανισμού του προθαλάμου, του οργάνου ισορρόπησης, και του κοχλία, δηλαδή του οργάνου που αναλύει τον ήχο. Και τα τρία περιέχουν το ίδιο υγρό, αλλά οι λειτουργίες τους είναι ανεξάντλητες. Ο κοχλίας, που έχει μέγεθος όσο ένα μπιζέλι, βρίσκεται μέσα σε στερεό κόκαλο.

Είναι στριμμένος σαν σαλιγκάρι απ' όπου παίρνει και το όνομα του. Στο Σχ. 3.7, για περιγραφικούς λόγους, αυτό το στρίψιμο των 2 στροφών τεντώνεται στο πλήρες μήκος του, που είναι περίπου 2.5 εκατοστά. Το γεμάτο υγρό έσω αυτί χωρίζεται κατά μήκος με δύο μεμβράνες, την μεμβράνη τον Reissner και την βασιλική (ή βασική) μεμβράνη. Άμεσα μας ενδιαφέρει η βασιλική μεμβράνη και η απόκριση της στις ταλαντώσεις ήχου μέσα στο υγρό.

Η ταλάντωση του τύμπανου ενεργοποιεί τα μικρά οστά. Η κίνηση του αναβολέα, που είναι κολλημένος στο ελλειψοειδές παράθυρο, προκαλεί την ταλάντωση του υγρού του έσω αυτιού. Μια κίνηση του ελλειψοειδούς παραθύρου προς τα μέσα έχει σαν αποτέλεσμα την ροή του υγρού γύρω από το πέρα άκρο της βασικής μεμβράνης, με αποτέλεσμα κίνηση προς τα έξω του κυκλικού παραθύρου. Ο ήχος που ενεργοποιεί το ελλειψοειδές παράθυρο έχει σαν αποτέλεσμα την δημιουργία στάσιμων κυμάτων στην βασιλική μεμβράνη. Η θέση των μέγιστων πλατών των στάσιμων κυμάτων στην βασική μεμβράνη μεταβάλλεται καθώς αλλάζει η συχνότητα του ήχου που προκαλεί την διέγερση.

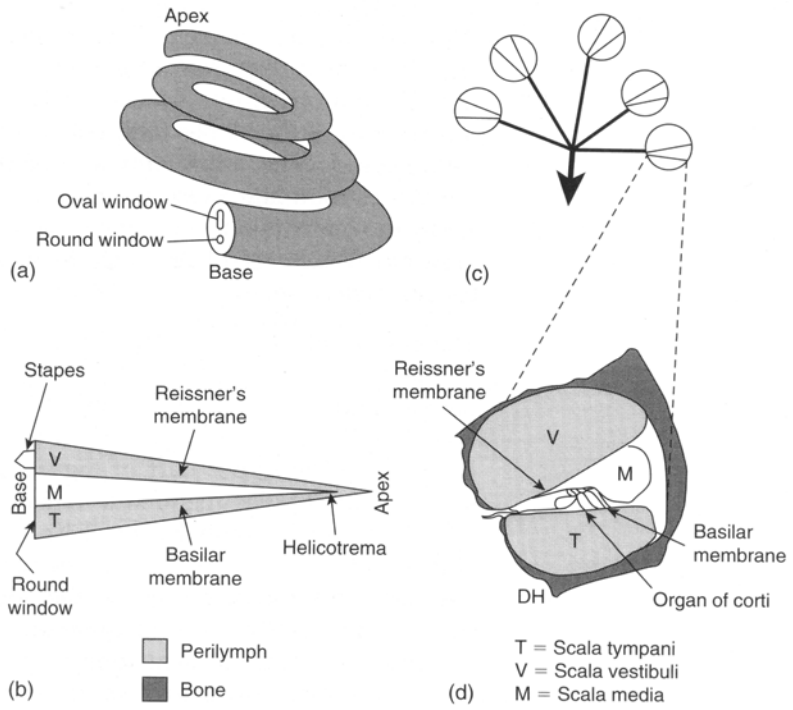
Ο ήχος χαμηλής συχνότητας έχει σαν αποτέλεσμα μέγιστο πλάτος κοντά στο πέρα άκρο της βασικής μεμβράνης. Ο ήχος υψηλής συχνότητας δημιουργεί κορυφές κοντά στο ελλειψοειδές παράθυρο. Σε περίπτωση πολύπλοκου

σήματος όπως είναι η μουσική ή η ομιλία, δημιουργούνται πολλές στιγμιαίες κορυφές, οι οποίες μετακινούνται διαρκώς σε πλάτος και θέση κατά μήκος της βασικής μεμβράνης. Αυτές οι κορυφές συντονισμού της βασικής μεμβράνης αρχικά θεωρούνταν ότι ήταν πολύ πλατειές ώστε να μη μπορεί να εξηγηθεί η οξύτητα του διαχωρισμού συχνοτήτων που εμφανίζεται στο ανθρώπινο αυτί. Η πρόσφατη έρευνα έδειξε ότι σε χαμηλές εντάσεις ήχου, οι καμπύλες συντονισμού της βασικής μεμβράνης είναι πολύ οξείες, και φαρδαίνουν μόνο για έντονο ήχο. Σήμερα φαίνεται ότι η οξύτητα των καμπύλων μηχανικού συντονισμού της βασικής μεμβράνης μπορεί να συγκριθεί με την οξύτητα των απλών ινών του ακουστικού νεύρου οι οποίες την γεμίζουν.

3.3.2.2.5 Στερεοβλεφαρίδες

Τα κύματα που δημιουργούνται στην βασική μεμβράνη του γεμάτου υγρό σωλήνα του έσω αυτιού ερεθίζουν άκρες νεύρων ψιλές σαν τρίχες οι οποίες μεταφέρουν σήματα στον εγκέφαλο με την μορφή εκκένωσης νευρώνων. Περίπου 15.000 εξωτερικά κύτταρα με περίπου 140 μικροσκοπικές τρίχες το καθένα οι οποίες ονομάζονται στερεοβλεφαρίδες (sterocilia) προεξέχουν από το καθένα. Επιπλέον, υπάρχουν περίπου 3.500 εσωτερικά κύτταρα με περίπου 40 στερεοβλεφαρίδες το καθένα. Αυτές οι στερεοβλεφαρίδες είναι οι πραγματικοί μετατροπείς ηχητικής ενέργειας σε ηλεκτρική. Υπάρχουν δύο είδη κυττάρων με τρίχες, τα εσωτερικά και τα εξωτερικά, που ονομάζονται έτσι ανάλογα με την θέση και την διάταξη τους. Καθώς ο ήχος αναγκάζει το υγρό του κοχλία και την βασική μεμβράνη να κινούνται, οι στερεοβλεφαρίδες στα κύτταρα με τις τρίχες λυγίζουν, οπότε ηλεκτρικά κύματα οδηγούνται προς τον ακουστικό φλοιό του εγκεφάλου.

Όταν ο ήχος διεγείρει το υγρό στο έσω αυτί, ερεθίζονται η μεμβράνη και τα κύτταρα με τις τρίχες, και στέλνουν ένα ηλεκτρικό κύμα μέσα από τον ιστό που τα περιβάλλει. Αυτά τα επονομαζόμενα ακουστικά δυναμικά (acoustic potentials), που είναι αναλογικά, λαμβάνονται και ενισχύονται, αναπαράγοντας τον ήχο που πέφτει στο αυτί, το οποίο λειτουργεί σαν βιολογικό μικρόφωνο. Τα δυναμικά αυτά είναι ανάλογα με την πίεση ήχου, και έχουν γραμμική απόκριση σε μια περιοχή μέχρι 80 dB. Αν και είναι ενδιαφέρον, αυτό το μικροφωνικό δυναμικό δεν πρέπει να μπερδεύεται με τα δυναμικά δράσης (action potentials) του ακουστικού νεύρου, τα οποία μεταφέρουν πληροφορίες στον εγκέφαλο.



Σχήμα 3 .8 (a) Σπειροειδής φύση του κοχλία (b) Ο ξετυλιγμένος κοχλίας (c) Κοχλίας σε τομή (d) Λεπτομερής άποψη του σωληνωτού κοχλία

Η κάμψη των στερεοβλεφαρίδων σκανδαλίζει τις νευρικές διεγέρσεις που μεταφέρονται με το ακουστικό νεύρο στον εγκέφαλο. Ενώ τα μικροφωνικά σήματα είναι αναλογικά, οι διεγέρσεις που στέλνονται στον ακουστικό φλοιό είναι διεγέρσεις που παράγονται από εκκενώσεις νευρώνων. Μία απλή ίνα νεύρου είτε διεγείρεται είτε δεν διεγείρεται (δυαδική!). Όταν διεγείρεται, προκαλεί την διέγερση της διπλανής της, κ.ο.κ. Οι φυσιολόγοι παρομοιάζουν την διαδικασία με το άναμμα του φυτιλιού της πυρίτιδας. Η ταχύτητα μετακίνησης δεν έχει καμία σχέση με τον τρόπο που ανάβει το φυτίλι. Υποθέτουμε ότι η ηχηρότητα του ήχου έχει σχέση με το πλήθος των ινών νεύρων που έχουν διεγερθεί και τις ταχύτητες επανάληψης μιας τέτοιας διέγερσης. Όταν διεγείρονται όλες οι ίνες νεύρων (περίπου 15.000), έχουμε την μέγιστη ηχηρότητα που μπορούμε να αντιληφθούμε. Η ευαισθησία κατωφλίου μπορεί να παρασταθεί με την διέγερση μιας μόνο ίνας. Ακόμη και σήμερα δεν έχει προταθεί μια συνολική αποδεκτή θεωρία του τρόπου με τον οποίο πραγματικά λειτουργούν το έσω αυτί και ο εγκέφαλος.
Τα παραπάνω αποτελούν μια εξαιρετικά απλουστευμένη παρουσίαση ενός πολύ πολύπλοκου μηχανισμού στον οποίο έχει αφιερωθεί μεγάλη ποσότητα πρόσφατης έρευνας. Μερικοί από τους αριθμούς που έχουν χρησιμοποιηθεί και μερικές από τις θεωρίες που έχουν εξεταστεί δεν είναι γενικά-εκ των υστέρων- αποδεκτές.

3.3.3 Δυνατότητες αναγνώρισης του ανθρώπινου ακουστικού συστήματος.

3.3.3.1 Ακουστική και αντίληψη μουσικών οργάνων-Έρευνες

Δυστυχώς οι δυνατότητες αντίληψης του ήχου από τον άνθρωπο δεν έχουν ακόμα αποσαφηνιστεί. Πολύ λίγοι ερευνητές έχουν ασχοληθεί με το κατά πόσο ο άνθρωπος μπορεί να αναγνωρίσει με εγκυρότητα τα μουσικά όργανα. Μάλιστα, τα όποια πειράματα έχουν γίνει αφορούν σε μεμονωμένους μουσικούς τόνους και όχι μουσικές φράσεις με πληροφορία που να σχετίζεται με κάποιο μουσικό περιεχόμενο. Αυτό βέβαια σε καμία περίπτωση δεν ανταποκρίνεται στις συνθήκες του πραγματικού κόσμου.

Ωστόσο, από τα πειράματα και τις μελέτες που έχουν γίνει, έχουν προκύψει συμπεράσματα, αρκετά από τα οποία συγκεντρώνουν ομοφωνία απόψεων. Τα σημαντικότερα από αυτά παρατίθενται παρακάτω.

Όσον αφορά σε μεμονωμένους μουσικούς τόνους (isolated tones) , **είναι πιο εύκολη η αναγνώριση ενός τόνου αν είναι παρούσα και η ατάκα του συγκεκριμένου τόνου (onset)**, δηλαδή αν ο τόνος ακούγεται από την αρχική στιγμή της διέγερσης του οργάνου από το οποίο και προέρχεται. Αυτό ειπώθηκε αρχικά από τον Stumpf το (1911) ,και στη συνέχεια από τους Eagleson & Eagleson (1947), Berger (1964), Saldanha & Corso (1964),Thayer (1972), Volodin (1972), Elliott (1975) ,Dillon (1981). Όμως δεν συμφώνησε μαζί τους και ο Kendal (1986) ο οποίος από τις δικές του έρευνες δεν διαπίστωσε κάτι τέτοιο.

Κάποια όργανα είναι ευκολότερο να αναγνωριστούν σε σχέση με άλλα. Σε μια μελέτη σχετική με την αναγνώριση 9 οργάνων (βιολί, άλτο κόρνο, τρομπέτα, πίκολο, φλάουτο, κλαρίνο, σαξόφωνο, καμπάνες και κύμβαλα) τα οποία εκτελούσαν την νότα ντο C4 (261 Hz) ,διαπιστώθηκε ότι το βιολί ,η τρομπέτα και οι καμπάνες αναγνωρίζονταν πιο εύκολα ,σε αντίθεση με το κόρνο, το φλάουτο και το πίκολο φλάουτο που αναγνωρίζονταν πιο δύσκολα (Eagleson & Eagleson,1947) . Οι Saldanha και Corso το 1964 και ο Berger (1964) πραγματοποίησαν παρόμοια πειράματα με διαφορετικό σετ οργάνων προς αναγνώριση κάθε φορά. Τα αποτελέσματα έδειχναν ότι σε κάθε πείραμα υπήρχαν όργανα εύκολα και όργανα δύσκολα προς αναγνώριση, όμως αυτά ήταν διαφορετικά κάθε φόρα και έτσι δεν θα μπορούσε ακόμα να εξαχθεί κάποιο γενικό συμπέρασμα.

Προσπαθώντας να κατηγοριοποιήσουν τα λάθη, οι ερευνητές έκαναν κάποιες σημαντικές παρατηρήσεις. Οι Saldanha και Korso (1964) **βρήκαν ομάδες οργάνων που είναι εύκολο να μπερδέψει κάποιος ακροατής**: το φαγκότο με το σαξόφωνο, το όμποε με το αγγλικό κόρνο, την τρομπέτα με την κορνέτα, το σαξόφωνο, και το αγγλικό κόρνο. Ο Berger παρατήρησε παρομοίως δυσκολία στο διαχωρισμό του γαλλικού κόρνου με το σαξόφωνο και την τρομπέτα, και του άλτο με το τενόρο σαξόφωνο. Στο εργαστήριο του Melville Clark στο MIT, μετά από μια σειρά πειραμάτων οι ερευνητές κατέληξαν στο συμπέρασμα **ότι η μεγαλύτερη δυσκολία διαχωρισμού παρατηρούνταν μεταξύ οργάνων που ανήκαν στην ίδια «οικογένεια».** Το ίδιο συμπέρασμα επιβεβαιώθηκε από τον Robertson το 1961 (τα χάλκινα, και τα ζευγάρια βιολί-βιόλα και τσέλο-κοντραμπάσο) και τον Schlossberg το 1960 (τρομπόνι-τρομπέτα).

Περισσότερες έρευνες οδήγησαν στο συμπέρασμα ότι **κάποιοι άνθρωποι είναι πιο ικανοί στην αναγνώριση οργάνων σε σχέση με άλλους**. Ο

Milner το 1963 ανακάλυψε ότι οι μουσικοί κάνουν λιγότερα λάθη στην αναγνώριση οργάνων σε σχέση με τους μη μουσικούς. Ο Kendal το 1986, αντίστοιχα, παρατήρησε ότι οι απόφοιτοι ενός πανεπιστημίου μουσικής κάνουν λιγότερα λάθη από τους σπουδαστές. Όλα αυτά τα πειράματα, αλλά και αυτό που μας λέει και η κοινή λογική και η μουσική παιδεία στη δυτική μουσική εδώ και 400 χρόνια, συντείνουν στο συμπέρασμα ότι **η αναγνώριση μουσικών οργάνων από τον άνθρωπο είναι μία ικανότητα η οποία καλλιεργείται.**

Οι Sandanha και Corso το 1964 κατέληξαν επίσης και σε ένα άλλο πολύ σημαντικό συμπέρασμα: η διαδικασία της αναγνώρισης ,για μεμονωμένους τόνους, εξαρτάται σε μεγάλο βαθμό από το μουσικό ύψος ,το pitch, του τόνου που εξετάζεται. Δηλαδή, **ανάλογα με τον τόνο, είναι δυνατόν ένα όργανο να αναγνωρίζεται λιγότερο ή περισσότερο εύκολα.** Επίσης, **η παρουσία vibrato** (μουσικός όρος που ισοδυναμεί φυσικά με τη διαμόρφωση συχνότητας ενός σήματος με συχνότητα κοντά στα 6 Hz-ηχητικά ερμηνεύεται σαν μια παλινδρόμηση γύρω από τον κεντρικό τόνο και αποτελεί πολύ συχνό φαινόμενο για τη φωνή και τα έγχορδα) **διευκολύνει την αναγνώριση**. Οι Campell και Heller το 1979 υποστήριξαν ότι **ο τρόπος που εκτελείται ένα legato** (νότες που παίζονται κολλημένες η μία μετά την άλλη συνεχόμενα και δημιουργούν μια φράση με απολυτη συνοχή), λόγω της μοναδικότητας του σε κάθε όργανο, **λειτουργεί αποφασιστικά σε μια απόφαση κατά την αναγνώριση μουσικών οργάνων.**

Μετά από μια σωρεία τέτοιων πειραμάτων ήταν **ο** , ο οποίος για πρώτη φορά **αμφισβήτησε την εγκυρότητα των πειραμάτων που στηρίζονταν σε τεστ μεμονωμένων μουσικών τόνων**. Υποστήριξε ότι οι απομονωμένοι τόνοι είναι μη συνηθισμένοι και μη φυσικοί και άρα τα πειράματα που στηρίζονταν πάνω σε αυτούς δύσκολά θα μπορούσαν να οδηγήσουν σε κάποια ασφαλή συμπεράσματα. Προκειμένου να επιβεβαιώσει τις επιφυλάξεις του, πραγματοποίησε πειράματα με εκτελέσεις, από όργανα, απομονωμένων τόνων χωρίς την αρχή τους (το onset),και μουσικών φράσεων από διάφορα είδη μουσικής (κλασσικά έργα, παραδοσιακά κομμάτια κ.α). Τα πειράματα έδειξαν ότι αυτοί οι παράγοντες που εξαρτούνται από την ατάκα της νότας δεν είναι ούτε ικανοί ούτε αναγκαίοι για την αναγνώριση των μουσικών οργάνων. Αντίθετα, σύμφωνα πάντα με την άποψή του, πιο σημαντικό ήταν το σταθερό μέρος (steady state) ενός απομονωμένου μουσικού τόνου.

Δυο μελέτες των τελευταίων 10 χρόνων είναι άξιες αναφοράς. Ο Crummer το 1994 πραγματοποίησε μετρήσεις της εγκεφαλικής δραστηριότητας κατά την αναγνώριση μουσικών οργάνων. Οι μετρήσεις αυτές έδειξαν ότι **ένας μουσικός χρειάζεται λιγότερη προσπάθεια από ένα μη μουσικό προκειμένου να αναγνωρίσει σωστά ένα συγκεκριμένο ακουστικό δείγμα**. Τέλος , οι Sandel και Χρονόπουλος, το 1996, δημοσίευσαν ότι **οι ακροατές καταφέρνουν να ξεπεράσουν τις δυσκολίες της αναγνώρισης οργάνων που μοιάζουν στο άκουσμα (π.χ αγγλικού κόρνου και όμποε) ,όταν έχουν πρωτύτερα εκπαιδευτεί με βάση το άκουσμα από πολλούς μουσικούς τόνους παρά μόνο από έναν. Επίσης ,με νότες από ένα περιορισμένο εύρος, οι ακροατές που εκπαιδεύτηκαν με μουσικές φράσεις έχουν καλύτερη επίδοση αναγνώρισης σε σχέση με αυτούς που εκπαιδεύτηκαν με μεμονωμένες νότες.**

3.3.3.2 Ακουστική και αντίληψη μουσικών οργάνων-οι διαστάσεις του ήχου

Ηχηρότητα (loudness)

Η ηχηρότητα είναι ένα υποκειμενικό μέτρο για την ένταση του ήχου που αντιλαμβάνεται ο άνθρωπος. Παρόλο που η ληφθείσα (από το αυτί) ηχηρότητα ενός ήχου σχετίζεται με το πλάτος της ακουστικής πίεσης, ωστόσο δεν υπάρχει κάποια εμφανής σχέση που να συνδέει αποκλειστικά αυτά τα δύο.

Ψυχοακουστικά πειράματα έχουν δείξει πως είναι δυνατόν ένα ακουστικό κύμα με μεγαλύτερο πλάτος ακουστικής πίεσης να ακούγεται λιγότερο έντονα από ένα κύμα με μικρότερο πλάτος. Αυτό συμβαίνει γιατί η ευαισθησία του ανθρώπινου αυτιού διαφέρει από συχνότητα σε συχνότητα. Ο Fletcher και ο Manson το 1933, μετά από πειραματικές μετρήσεις κατέληξαν στη γνωστή καμπύλη της στάθμης της ηχητικής πίεσης και της ληφθείσας-από τον άνθρωπο- ηχηρότητας, για τις διάφορες συχνότητες. Σύμφωνα με αυτές, ήχοι χαμηλών συχνοτήτων γίνονται λιγότερο αντιληπτοί από ήχους ψηλών συχνοτήτων (πάνω από 1 kHZ) της ίδιας στάθμης ακουστικής πίεσης.

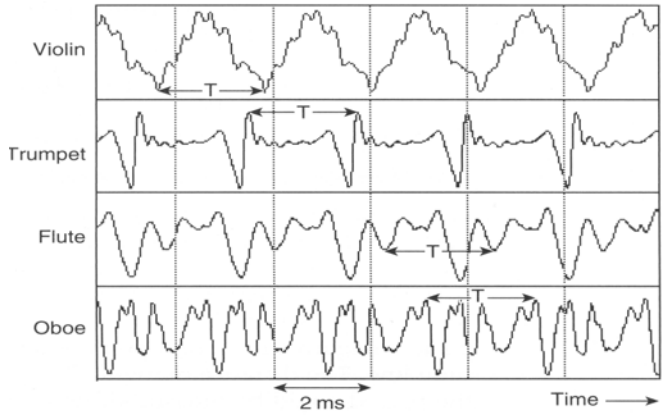
Τονικό ύψος(Pitch)

Σαν τονικό ύψος (pitch) ονομάζουμε, στη δυτική μουσική, τη νότα –ή πιο σωστά το φθόγγο που αντιπροσωπεύεται από τη νότα- στο συγκερασμένο σύστημα χορδίσματος, που παράγεται από κάποια ηχητική πηγή. Το pitch καθορίζεται, στη μουσική βιβλιογραφία , είτε με τις ονομασίες Ντο, ρε , μι κτλ και τις διέσεις και υφέσεις αυτών και την οκτάβα στην οποία ανήκουν, είτε ,συμφωνά με την αγγλόφωνη μουσική βιβλιογραφία, με τις ονομασία X_y ,όπου X={C,D,E,F,G,A,B} τα αντίστοιχα ντο, ρε, μι κτλ και y={1,2,3,4,5,6,7} οι οκτάβες στις οποίες βρίσκεται η παραπάνω νότα.

Όταν ακούμε μία νότα από ένα μουσικό όργανο μας δημιουργείται η αίσθηση ενός συγκεκριμένου τονικού ύψους (pitch). Αυτό συμβαίνει επειδή το μουσικό όργανο παράγει ένα κύμα ακουστικής πίεσης το οποίο έχει κάποιο είδος περιοδικότητας (quasi-periodic) (σχήμα 3.9) . Το τονικό ύψος στην περίπτωση αυτή αντιστοιχεί σε μια συχνότητα f_0 η οποία βρίσκεται από τον τύπο :

$$f_o = \frac{1}{T}, \qquad (3.8)$$

όπου T η περίοδος του ήχου σε sec.



Σχήμα 3. 9 Κυματομορφή της ακουστικής πίεσης στη νότα λα Α4 440 Hz

Στο σχήμα 3.9 αναπαρίστανται οι κυματομορφές της νότας λα (Α4 ,440Hz) εκτελεσμένης από το βιολί, την τρομπέτα, το φλάουτο και το όμποε. Παρατηρούμε ότι η περίοδος και στις τέσσερις αυτές κυματομορφές είναι η ίδια και ίση με Τ. Σύμφωνα με τον παραπάνω τύπο μπορούμε να υπολογίσουμε και τη συχνότητα του pitch f_o.

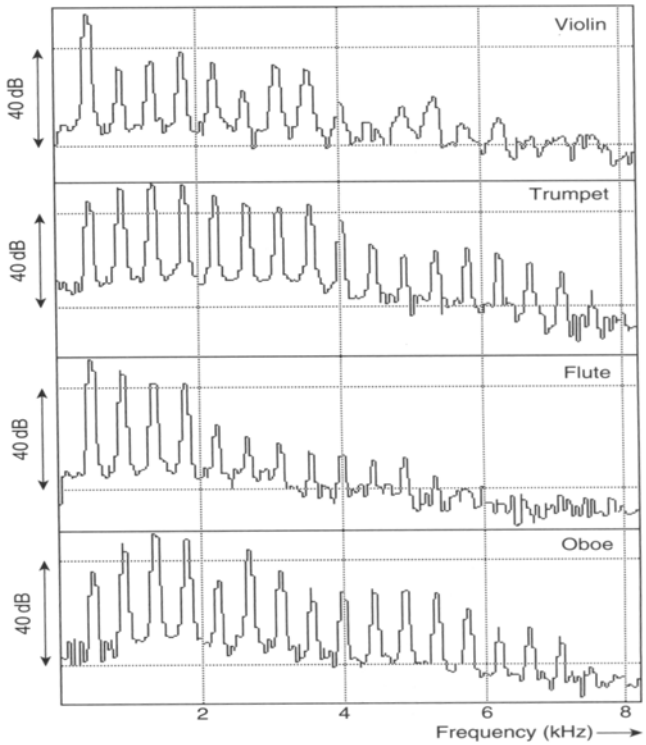
Χροιά (Timbre)

Πολύ μεγάλο μέρος της έρευνας στην ψυχοφυσική έχει αφιερωθεί σχετικά με τη λέξη χροιά(Timbre), όσον αφορά στη μουσική.

Σαν χροιά ονομάζουμε το χαρακτηριστικό της αίσθησης της ακοής μέσου του οποίου ένας ακροατής μπορεί να αποφανθεί ότι δύο ήχοι με το ίδιο τονικό ύψος (pitch), την ίδια ηχηρότητα και την ίδια διάρκεια είναι διαφορετικοί. Η χροιά εξαρτάται κατά το μεγαλύτερο ποσοστό από τη μορφή του φάσματος του ηχητικού ερεθίσματος, αλλά εξαρτάται επίσης από την κυματομορφή, την ακουστική πίεση, την περιοχή συχνοτήτων του φάσματος και τις περιβαλλοντικές συνθήκες.

Κατά πολλούς η λέξη χροιά δεν έχει καμία επιστημονική σημασία. Παρομοιάζεται μάλιστα με τη λέξη εμφάνιση (appearance) στην όραση: Σημαίνει διαφορετικά πράγματα για διαφορετικούς ανθρώπους, ενώ συμπεριλαμβάνει μέσα της πολλά χαρακτηριστικά και ποσότητες.

Ωστόσο η λέξη χροιά χρησιμοποιείται από πολλούς ερευνητές σαν συνώνυμο με την ηχητική πηγή (sound source), σαν ένα σύνολο χαρακτηριστικών που καθορίζουν μια ηχητική πηγή.



Σχήμα 3. 10 Τα φάσματα των κυματομορφων του σχήματος 3.9

Ενώ οι περίοδοι στις κυματομορφές της νότας Α4 εκτελεσμένης από τα 4 όργανα (βιολί, τρομπέτα, φλάουτο, όμποε) είναι ίσες, ωστόσο τα σχήματα που έχουν οι κυματομορφές είναι τελείως διαφορετικά. Το pitch που θα αντιληφθεί ο ακροατής είναι το ίδιο και στις τέσσερις περιπτώσεις και ο διαφορετικός ήχος του λα του κάθε οργάνου, δηλαδή η χροιά της ηχητικής πηγής (timbre) (βλ. κεφ. 3.3.3.2.2) θα μπορούσαμε να πούμε ότι αντικατοπτρίζεται στη διαφορετικότητα των κυματομορφών. Οι διαφοροποιήσεις αυτές στις κυματομορφές δεν είναι παρά διαφοροποιήσεις στην ακουστική πίεση, η οποία λαμβάνεται από το τύμπανο και εν συνεχεία μεταφέρεται μηχανικά στην βασική (βασιλική) μεμβράνη του αυτιού του ανθρώπου, έχοντας αναλυθεί σε συντελεστές συχνοτήτων . Έτσι δύο διαφορετικά όργανα, ακόμα και στην περίπτωση μιας ίδιας νότας, έχοντας διαφορετικές κυματομορφές, δημιουργούν διαφορετικές διεγέρσεις στην βασική μεμβράνη και άρα διαφορετικές διεγέρσεις στο ακουστικό νεύρο. Μέσω αυτής της διαδικασίας ο άνθρωπος αντιλαμβάνεται τη διαφορετικότητα στη χροιά.

Στο σχήμα 3.10 παριστάνεται το φάσμα πλάτους της συχνότητας για την νότα λα από τα ίδια όργανα. Στο διάγραμμα αυτό παρατηρούμε τοπικά μέγιστα σε συχνότητες πολλαπλάσιες του f_o=440Hz. Αυτό εξηγείται από τη θεωρία ακουστικής και έχει σχέση με τη φυσιολογία των μουσικών οργάνων. Τα πλάτη στις συχνότητες αυτές είναι διαφορετικά. Έτσι αυτό που διαμορφώνει τη χροιά ,σε μία γενική προσέγγιση, είναι οι διαφορετικές τιμές του πλάτους στα πολλαπλάσια της βασικής συχνότητας f_o που αντιπροσωπεύει το τονικό ύψος. Οι συχνότητες πολλαπλάσιες του f_o ονομάζονται αρμονικές (harmonic partials, overtones).

Ωστόσο η παραπάνω προσέγγιση για το pitch και για τη χροιά είναι αρκετά γενικευμένη και παρουσιάζει πολλά προβλήματα. Και αυτό γιατί σε ήχους από τους οποίους απουσιάζει η θεμελιώδης συχνότητα f_o , ή σε ήχους οι οποίοι δεν έχουν περιοδικότητα ,όπως είναι κάποια φυσικά φαινόμενα π.χ. ο ήχος των κυμάτων ή της βροχής, συμβαίνει να είναι αντιληπτή πολλές φορές η αίσθηση τονικού ύψους. Σε εξήγηση αυτού αναπτύχθηκαν οι συμπληρωματικά η τοπική και εν συνεχεία η χρονική θεωρία για την αντίληψη του pitch, θεωρίες που όμως ξεφεύγουν από το θέμα της παρούσας διατριβής.

3.4 Οικογένειες μουσικών οργάνων

Τα μουσικά όργανα, όπως τα ξέρουμε από τη δυτική μουσική παράδοση και αν εξαιρέσουμε τα κρουστά, χωρίζονται σε 3 μεγάλες κατηγορίες: τα χάλκινα, τα πνευστά και τα έγχορδα.

Προηγούμενες έρευνες (κεφ. 3.3) έδειξαν πως οι περισσότερες δυσκολίες στην αναγνώριση παρουσιάζονταν σε όργανα της ίδιας οικογένειας. Σε μια προσπάθεια να ταξινομήσουμε το πλήθος των οργάνων ανάλογα με το σχήμα τους και τις γεωμετρικές τους ιδιότητες, παρατηρούμε ότι τα όργανα αυτά λαμβάνουν γειτονική θέση. Αξίζει να κάνουμε μια σύντομη επισκόπηση στη μελέτη των ακουστικών ιδιοτήτων των οργάνων των τριών μεγάλων κατηγοριών.

Τα χάλκινα

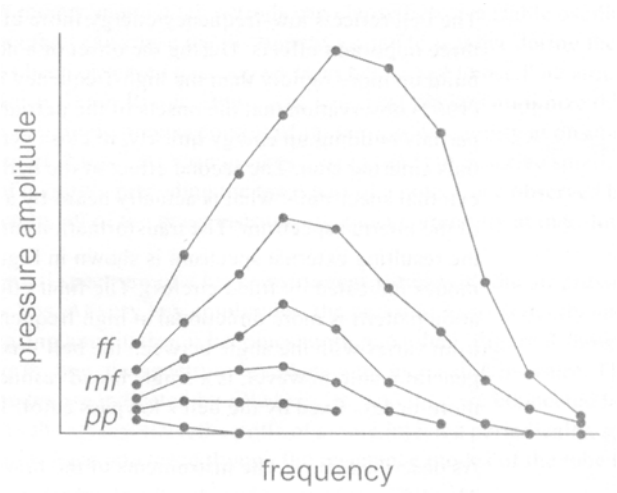
Τα χάλκινα όργανα έχουν την πιο απλή δομή σε σχέση με τις άλλες κατηγορίες. Η οικογένεια των χάλκινων περιλαμβάνει όργανα ορχήστρας όπως : κορνέτα, τρομπέτα, τρομπόνι, γαλλικό κόρνο, επιφώνιο και τούμπα. Κάθε όργανο από αυτά αποτελείται από ένα μακρύ και συνήθως χάλκινο σωλήνα ,ο οποίος καταλήγει στο τέλος του σε ένα άνοιγμα-καμπάνα. Ο εκτελεστής παίζει φυσώντας στο μπροστινό, στενό μέρος του σωλήνα μέσω ενός εξαρτήματος που ονομάζεται επιστόμιο.

Η διαδικασία της εξαγωγής ήχου από ένα τέτοιο όργανο έχει ως εξής: Ο εκτελεστής φυσώντας επιτρέπει μικρές δόσεις αέρα να ταξιδεύουν στο σωλήνα. Ο αέρας όταν φτάνει στην καμπάνα συναντά διαφορετική εμπέδηση, η οποία έχει σαν αποτέλεσμα τη δημιουργία στάσιμων κυμάτων κατά μήκος του σωλήνα. Τα στάσιμα αυτά κύματα έχουν συχνότητες- πολλαπλάσια μιας συχνότητας που είναι ίση με την ταχύτητα του ήχου δια το διπλάσιο μήκος του σωλήνα .

Τα στάσιμα κύματα κατά μήκος του σωλήνα χρειάζονται κάποιο χρονικό διάστημα προκειμένου να δημιουργηθούν. Κατά τη διάρκεια αυτή ,ο ήχος του οργάνου παραμένει εξίσου ασταθής, και στην περίπτωση υψηλού pitch μπορεί να χρειαστούν πολλές περίοδοι pitch προκειμένου να δημιουργηθεί μια σταθερή ταλάντωση. Κατά τη διάρκεια της ασταθούς ατάκας(onset) παρατηρείται και μικρές εκρήξεις ενέργειας μη αρμονικού ήχου, ιδιαίτερα σε σχετικά άπειρους εκτελεστές.

Το φάσμα συχνοτήτων σε ένα χάλκινο όργανο χαρακτηρίζεται ,για μικρά πλάτη ακουστικής πίεσης, από περιοδικότητα (το κύμα είναι σχεδόν ημιτονοειδές), όμως με αύξηση της ακουστικής πίεσης (μέσω αύξησης του

φυσήματος του εκτελεστή μέσα στο σωλήνα) το σήμα που εξάγεται προσεγγίζει την κρουστική απόκριση. Μια σχηματική αναπαράσταση του φάσματος σε σχέση με τη στάθμη της ακουστικής πίεσης σε «μουσική» μονάδα (pp=πολύ σίγα έως ff= πολύ δυνατά) δίνεται στο σχήμα 3.11.



Σχήμα 3. 11 Το φάσμα ενός χάλκινου οργάνου, για 6 στάθμες ακουστικής πίεσης,σύμφωνα με τον Benade

Τα πνευστά

Τα πνευστά (ή ξύλινα) αποτελούν τη λιγότερο ομογενής κατηγορία από τις 3, τουλάχιστον όσον αφορά στα όργανα μιας ορχήστρας. Η ανομοιογένεια αυτή είναι τόσο ακουστική όσο και «αντιληπτική» και έχει σαν συνέπεια τον διαχωρισμό των ξύλινων σε υποκατηγορίες :ξύλινα με διπλό καλάμι, τα κλαρινέτα, τα φλάουτα, τα υπόλοιπα με μονό καλάμι και τα σαξόφωνα.

Τα ξύλινα εξάγουν ήχο με τη δημιουργία στάσιμων κυμάτων σε ένα σωλήνα, του οποίου το ωφέλιμο μήκος είναι μεταβλητό και εξαρτάται από το κλείσιμο ή όχι ,κατ' επιλογή, ενός αριθμού από τρύπες. Σε αντίθεση με τα χάλκινα, τα ξύλινα έχουν και έναν αριθμό από ειδικά κλειδιά (registers) τα οποία με το πάτημά τους απελευθερώνουν ένα αριθμό από άλλες μικρές τρύπες.

Βασικό χαρακτηριστικό των ξύλινων είναι ο μικρός χρόνος που απαιτείται για να φτάσουν σε σταθερή κατάσταση. Ωστόσο, ανάλογα με την υποκατηγορία στην οποία ανήκουν, τα ξύλινα έχουν διαφορετικές ακουστικές ιδιότητες, όπως συχνότητες συντονισμού, φασματικό περιεχόμενο κ.α. Περισσότερα για τις ακουστικές ιδιότητες των πνευστών μπορούν να αναζητηθούν σε μελέτες των Rossing, Moore, Luce, Benade .

Τα έγχορδα

Τα έγχορδα (strings) χωρίζονται επίσης σε δυο μεγάλες υποκατηγορίες: τα *έγχορδα με δοξάρι* (bowed strings) και τα *νυκτά έγχορδα* (plucked strings).

Συνηθισμένα έγχορδα με δοξάρι της δυτικής μουσικής παράδοσης είναι το βιολί ,η βιόλα, το τσέλο και το κοντραμπάσο, ενώ γνωστά νυκτά είναι η κιθάρα και η άρπα.

Ένα τυπικό έγχορδο όργανο αποτελείται από ένα ξύλινο σώμα και έναν προεξέχων λαιμό. Οι χορδές είναι τοποθετημένες σε τεντωμένη θέση κατά μήκος του λαιμού και όντας πιασμένες η κάθε μία από ένα κλειδί στην άκρη του, μέχρι ένα σημείο στο κάτω μέρος του κυρίως σώματος του οργάνου, τη *γέφυρα*. Όταν πάλλονται οι χορδές, η παλμική κίνηση μεταφέρεται μέσω της γέφυρας στο σώμα του οργάνου και δημιουργεί ταλάντωση στα μόρια του αέρα που βρίσκονται μέσα του. Ο εκτελεστής θέτει μια χορδή σε κίνηση είτε σέρνοντας ένα *δοξάρι*, που αποτελείται από τρίχες ζώου, κάθετα στη χορδή, είτε τσιμπώντας τη χορδή με το νύχι του η ένα πλαστικό ειδικά διαμορφωμένο εξάρτημα (πένα).

Στα έγχορδα με δοξάρι, το δοξάρι «κολλάει» κατά μικρά χρονικά διαστήματα πάνω στη χορδή, με αποτέλεσμα η χορδή να ακολουθεί σε αυτά τα διαστήματα την κίνηση του δοξαριού και ξαφνικά να επανέρχεται.

Ο ήχος στη σταθερή κατάσταση (steady state) είναι περιοδικός, και η μορφή του κύματος του ήχου που εξάγεται εξαρτάται από την πίεση που ασκείται μέσω του δοξαριού, αλλά και από το πόσο κοντά στη γέφυρα βρίσκεται το δοξάρι. Ωστόσο είναι λίγο παραπλανητικό να γίνεται λόγος για σταθερή κατάσταση σε ένα έγχορδο με δοξάρι, καθώς αυτή εξαρτάται από κινήσεις οι οποίες έχουν ακρίβεια χιλιοστού. Επίσης ,τόσο η ατάκα όσο και το σβήσιμο (onset και offset) είναι αρκετά πολύπλοκες μη ντετερμινιστικές διαδικασίες. Γενικά το φάσμα τους δεν είναι αρμονικό, ενώ λάθη στο onset και στο offset μπορεί να οδηγήσουν σε τρίξιμο της χορδής.

Το φάσμα ενός νυκτού έγχορδου δεν είναι ποτέ αρμονικό. Αυτό οφείλεται στη διαφορετική ταχύτητα των αρμονικών των διάφορων συχνοτήτων που κινούνται κατά μήκος της χορδής γεγονός που έχει σαν αποτέλεσμα το τονικό ύψος των υψηλότερων αρμονικών να είναι ελαφρώς χαμηλότερο.

Η παρουσία της γέφυρας καθορίζει τις ακουστικές ιδιότητες ενός έγχορδου μουσικού οργάνου. Δημιουργεί συντονισμούς στο φάσμα συχνοτήτων (π.χ. για το βιολί οι συντονισμοί που οφείλονται στη γέφυρα δημιουργούνται κοντά στα 3 και στα 6 kHz. Με τη βοήθεια ενός ειδικού εξαρτήματος, *της σουρντίνας*, που τοποθετείται στη γέφυρα, αυξάνεται η μάζα της γέφυρας και οι συχνότητες συντονισμού γίνονται μικρότερες, με αποτέλεσμα το όργανο να εξάγει ένα πιο σκοτεινό, θα λέγαμε, ήχο.

Το σώμα ενός έγχορδου έχει πολλές συχνότητες συντονισμού, τόσο λόγω του αέρα που βρίσκεται μέσα όσο και λόγω των ξύλινων επιστρώσεων. Οι συχνότητες αυτές έχουν στενό πλάτος (υψηλό Q) . Οι πρώτοι χαμηλοί -σαν συχνότητα- συντονισμοί (αέρα και ξύλου) είναι απόλυτα ελεγχόμενες κατά την κατασκευή ποιοτικών, τουλάχιστον, οργάνων. Δεν συμβαίνει όμως το ίδιο και με τους συντονισμούς υψηλής συχνότητας, οι οποίοι διαφέρουν από όργανο σε όργανο, και μπορούν μάλιστα να μεταβάλλονται με το πέρασμα των χρόνων.

Μία ενδιαφέρουσα τεχνική που χρησιμοποιείται κατά την εκτέλεση ενός έγχορδου είναι το vibrato, που ισοδυναμεί ηχητικά με μια παλινδρόμηση γύρω από ένα κεντρικό pitch. Αυτό πετυχαίνεται με την παλινδρόμηση μπρος-πίσω του δαχτύλου που πατά την χορδή στο λαιμό του οργάνου. Αυτή η «διαμόρφωση» στο μήκος της ενεργού χορδής ισοδυναμεί, μιλώντας με

τεχνικούς όρους, με διαμόρφωση του pitch, δηλαδή της συχνότητας. Η διαμόρφωση συχνότητας προκαλεί διαμόρφωση πλάτους για το φάσμα και άρα και για κάθε αρμονική συχνότητα χωριστά. Ανάλογα με τη θέση της κάθε αρμονικής σε σχέση με τις συχνότητες συντονισμού , η διαμόρφωση πλάτους μπορεί να έχει διαφορετική κατεύθυνση.

Οι συντονισμοί επηρεάζουν και την ατάκα αλλά και την απόσβεση κάθε νότας. Σαν συνέπεια, κάθε όργανο χρειάζεται διαφορετικό χρονικό διάστημα για την ατάκα και την επίτευξη της σταθερής κατάστασης (πίνακας 1).

Όργανο	Χρόνος για την επίτευξη σταθερής κατάστασης(ms)	Χρόνος για την επίτευξη μέγιστου πλάτους(ms)
Βιολί	100	200
Βιόλα	40	100
Τσέλο	120	350
Κοντραμπάσο	80	100

Πίνακας 1. Ατάκα και μέγιστο πλάτος για τα 4 βασικά έγχορδα της ορχήστρας

3.5 Αναπαράσταση των χαρακτηριστικών

Στο κεφάλαιο 3.1 δόθηκε ο ορισμός του διανύσματος των χαρακτηριστικών. Στα κεφάλαια 3.2-3.4 έγινε περιγραφή του τρόπου προσέγγισης του προβλήματος της αναγνώρισης από μαθηματική σκοπιά, καθώς και την οπτική της ανθρώπινης αντίληψης αλλά και των ακουστικών ιδιοτήτων των οργάνων. Τελικά ,τίθεται το ζήτημα, ποια είναι τα κατάλληλα χαρακτηριστικά προκειμένου να μπορέσουμε να αναγνωρίσουμε μία ηχητική πηγή όπως η αναγνώριση οργάνων; Οι διάφορες προσεγγίσεις που έγιναν από ερευνητές στο παρελθόν, σε όποια κατηγορία χαρακτηριστικών(φυσικά ή αντιληπτικά) και αν βασίστηκαν, τελικό στόχο είχαν τη δημιουργία ενός διανύσματος χαρακτηριστικών όπως αυτό της παραγράφου 3.1. με βάση τα χαρακτηριστικά από κατάλληλα δείγματα κατασκευάζονται τα σετ χαρακτηριστικών εκπαίδευσης και αξιολόγησης, τα οποία εισάγονται διαδοχικά σε κάποιον τυπικό ταξινομητή προκειμένου να κατασκευαστεί και εν συνεχεία να αξιολογηθεί το σύστημα της αναγνώρισης.

Στην παράγραφο αυτή θα γίνει μια αναφορά στις μεθόδους αναπαράστασης χαρακτηριστικών, είτε πρόκειται για φυσικά είτε για χαρακτηριστικά αντίληψης, τα οποία χαίρουν συχνής χρήσης από τους ερευνητές. Τα χαρακτηριστικά αυτά, μεμονωμένα ή και σε συνδυασμούς χρησιμοποιούνται για την κατασκευή του διανύσματος χαρακτηριστικών και καθορίζουν σε μεγάλο βαθμό τη φύση του ταξινομητή που θα χρησιμοποιηθεί σαν τελευταία βαθμίδα επεξεργασίας στο σύστημα.

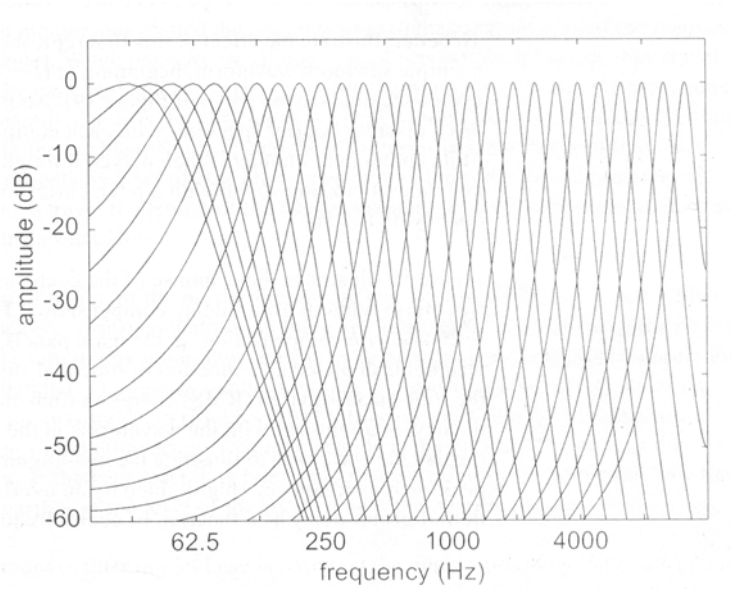
3.5.1 Χαρακτηριστικά σε έκφραση του φάσματος που προκύπτουν από ακουστικά μοντέλα

Ένα σύνηθες μουσικό αρχείο έχει υποστεί δειγματοληψία 44.100 Hz το δευτερόλεπτο. Η χρήση λοιπόν, αυτούσιου του ακατέργαστου μουσικού αρχείου από την ηχογράφηση ενός οργάνου, σαν χαρακτηριστικό που εισάγεται σε κάποιον ταξινομητή κρίνεται αυτομάτως από δύσκολη ως αδύνατη. Επιπλέον, οι μελέτες από την ψυχοακουστική δείχνουν ότι μέσα στα χαρακτηριστικά θα πρέπει να εισάγουμε και στοιχεία που αφορούν στον ανθρώπινο τρόπο αντίληψης και αναγνώρισης. Τα δύο αυτά στοιχεία έχουν συντελέσει στην εισαγωγή χαρακτηριστικών που προκύπτουν από ακουστικά μοντέλα για τον κοχλία του αυτιού. Σύμφωνα με αυτά , στον κοχλία το σήμα σε γενικές γραμμές υφίσταται ανάλυση συχνότητας ,μια μορφή φιλτραρίσματος από μια τράπεζα ζωνοπερατών φίλτρων. Σε μια προσπάθεια μοντελοποίησης του αυτιού, κατά τον ορισμό των συχνοτήτων των φίλτρων πρέπει να ληφθούν υπόψη και οι μη γραμμικότητες που διακρίνουμε στην ανθρώπινη ακοή. Ωστόσο, ο τρόπος που λειτουργεί ο κοχλίας δεν είναι ακόμα απόλυτα γνωστός , με αποτέλεσμα την διαφορετική προσέγγιση από κάθε μοντέλο στο συγκεκριμένο ζήτημα. Παρακάτω αναφέρονται τα σημαντικότερα από τα μοντέλα που χρησιμοποιούνται για την εξαγωγή των χαρακτηριστικών κατά την διαδικασία αναγνώρισης οργάνων.

- ERB (equivalent rectangular bandwidth)

Το μοντέλο αυτό προτάθηκε από τον Patterson το 1990 (Patterson & Moore ,1990 ; Patterson & Holdsworth,1993) και σε κάποιες περιπτώσεις το συναντούμε σε συνδυασμό με νευροφυσιολογικές και ψυχοφυσικές μελέτες του Meddis (1986).

Ο Patterson προσομοιώνει τον κοχλία με μια τράπεζα φίλτρων. Καθένα από αυτά τα φίλτρα προκύπτει από 4 διπολικά φίλτρα με ίδιους πόλους αλλά διαφορετικά μηδενικά, τα οποία σχηματίζουν ένα οχταπολικό φίλτρο που προσεγγίζει τη απόκριση της γαμματονικής συνάρτησης. Το εύρος ζώνης των φίλτρων, στις χαμηλές συχνότητες, είναι κάπως μεγαλύτερο, προσομοιώνοντας τα “critical bands” του κοχλία και όσο αυξάνεται η συχνότητα σταθεροποιείται. Ο αριθμός των φίλτρων ανά το φάσμα συχνοτήτων αλλάζει από εφαρμογή σε εφαρμογή, με μέσο όρο γύρω στα 6 φίλτρα (6 κεντρικές συχνότητες) ανά οκτάβα (σχήμα 3.12)

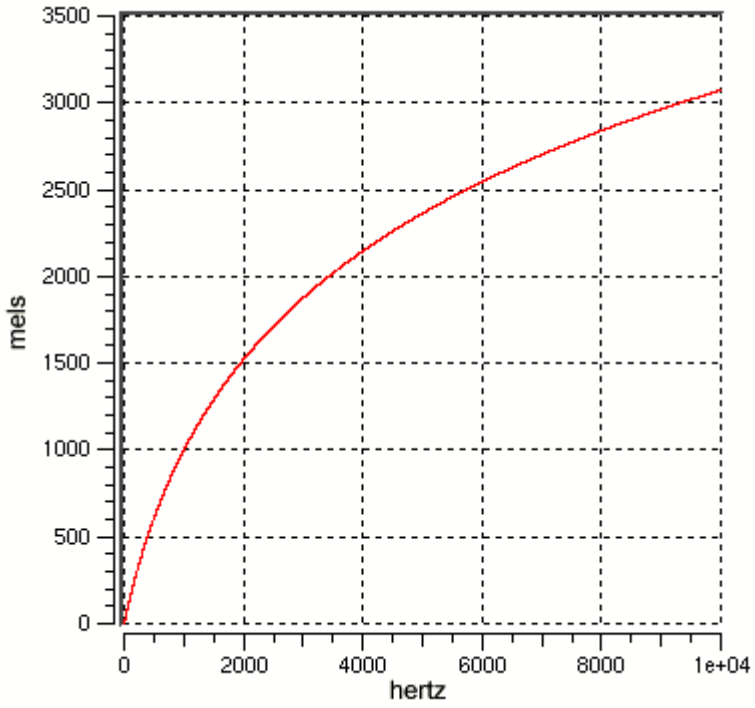


Σχήμα 3. 12 Γενική απόκριση συνάρτησης του κοχλία σε λογαριθμική κλίμακα

- MFCC (Mel Frequency Cepstral Coefficients)

Η κλίμακα mel, προτεινόμενη από τους Stevens, Volkmann και Newman το (Journal of Acoustic Society of America),1937,p.185 - 190) είναι μια κλίμακα τονικών υψών τα οποία θεωρούνται από τους ακροατές ίσα στην απόσταση μεταξύ τους. Το σημείο αναφοράς μεταξύ αυτής της κλίμακας και της κανονικής μέτρησης της συχνότητας σε Hertz καθορίζεται με την εξίσωση ενός τόνου συχνότητας 1000 Hz , 40 db πάνω από το κατώτατο όριο ηχηρότητας για τον ακροατή, με ένα τόνο (pitch) 1000 mel. Πάνω από 500 Hz, ολοένα και μεγαλύτερα διαστήματα θεωρούνται από τους ακροατές ότι παραγάγουν ίσες αυξήσεις τονικών υψών. Κατά συνέπεια, τέσσερις οκτάβες στην κλίμακα Hertz (πάνω από 500 Hz) ισοδυναμούν περίπου με δύο οκτάβες στην κλίμακα mel. Το όνομα mel προέρχεται από τη λέξη μελωδία και δείχνει ότι η κλίμακα είναι βασισμένη στις συγκρίσεις τονικών υψών. Η ακριβής αντιστοιχία των mel με τα Hertz φαίνεται στη γραφική παράσταση (σχήμα 13).

Τα MFCC είναι συντελεστές οι οποίοι βασίζονται στην κλίμακα Mel και προκύπτουν από μία σειρά μετασχηματισμών . Συγκεκριμένα πρώτα το σήμα πολλαπλασιάζεται με ένα παράθυρο, π.χ. παράθυρο Hamming και στη συνέχεια λαμβάνεται το FFT του σήματος. Βάση της κλίμακας mel κατασκευάζεται μία τράπεζα επικαλυπτόμενων τριγωνικών φίλτρων με κεντρικές συχνότητες που κατανέμονται αρχικά γραμμικά (οι πρώτες 13 έχουν απόσταση 133,3 Hz) και στη συνέχεια λογαριθμικά (οι υπόλοιπες 27 χωρίζονται με ένα παράγοντα 1,071). Το FFT του σήματος φιλτράρεται μέσα από τα φίλτρα αυτά. Το πλάτος της απόκρισης συχνότητας της τράπεζας φίλτρων εξαρτάται από την κλίμακα mel, και φθίνει εκθετικά στις μεγαλύτερες συχνότητες. Στη συνέχεια λαμβάνεται ο λογάριθμός με βάση 10 των εξόδων των φίλτρων. Τέλος για την μείωση της διαστασιμότητας χρησιμοποιείται ο μετασχηματισμός DCT , και τελικά λαμβάνεται ένα διάνυσμα με τους 12 πιο σημαντικούς συντελεστές συν ένας συντελεστής που αποτελεί στην ενέργεια στο frame.



Σχήμα 3.13 Τα mels σε συνάρτηση με τα Hz

Ο τύπος που δίνει τα mel σε σχέση με τα Hz είναι :

$$m = 1127.01048 \log_e(1 + f/700) \tag{3.9}$$

Λύνοντας ως προς f έχουμε:

$$f = 700(e^{m/1127.01048} - 1) \tag{3.10}$$

Αν με P_j συμβολίσουμε το λογάριθμο του πλάτους του FFT από ενός φίλτρου j της τράπεζας ο τύπος που δίνει τον i συντελεστή MFCC είναι:

$$c_i = \Sigma_{j=1}^{N} P_j \cos(i\pi/N\,(j\text{-}0.5)). \tag{3.11}$$

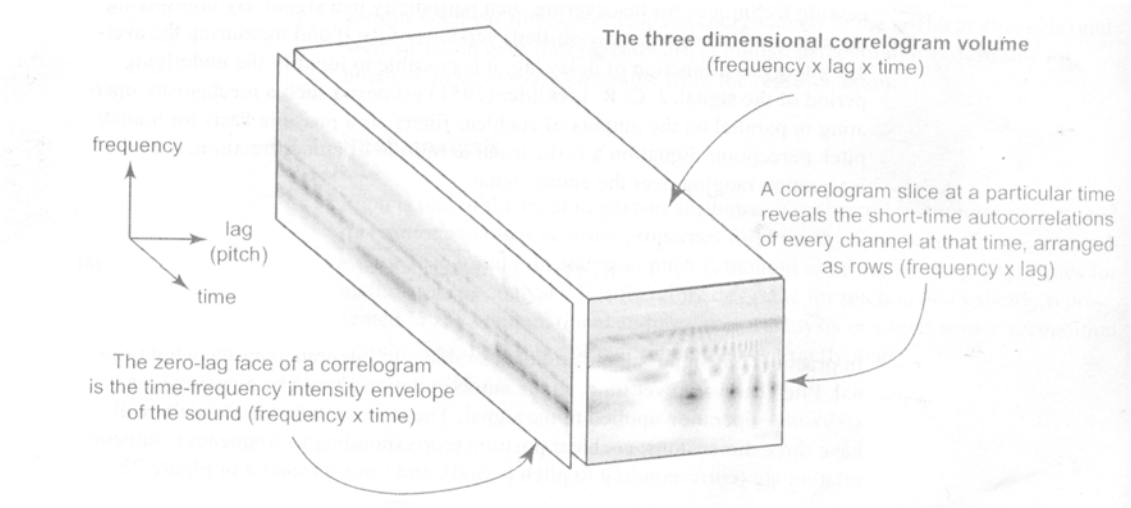
3.5.2 Χαρακτηριστικά που εκφράζουν ιδιότητες στη νότα

- Τονικό ύψος

Το τονικό ύψος (pitch) είναι ίσως το πιο σημαντικό χαρακτηριστικό στα μουσικά όργανα. Το ίδιο σημαντικό είναι και για την ανάλυση ακουστικής σκηνής.
Ένα περιοδικό σήμα, σε κάθε έξοδο φίλτρου ενός μοντέλου που προσομοιώνει τον κοχλία, θα έχει σαν αποτέλεσμα ένα επίσης περιοδικό σήμα. Η περιοδικότητα στα διάφορα κανάλια είναι η βάση για την κατανόηση του pitch από τον άνθρωπο. Η αυτοσυσχέτιση είναι ένα από τα πιο συνηθισμένα εργαλεία της επεξεργασίας σήματος για την εύρεση της περιοδικότητας. Πολλαπλασιάζοντας ένα σήμα με αντίγραφα του εαυτού του που έχουν υποστεί καθυστέρηση στο πεδίο του χρόνου, και υπολογίζοντας τη μέση ενέργεια σε συνάρτηση της καθυστέρησης, καθίσταται δυνατή η εξεύρεση της περιοδικότητας του σήματος. Ο μαθηματικός τύπος για την αυτοσυσχέτιση ενός συνεχούς σήματος είναι:

$$Rxx(\tau) = \int_{-\infty}^{\infty} x(t)x(t-\tau)dt \qquad (3.12)$$

Στην πράξη ,το ζητούμενο είναι η αυτοσυσχέτιση να υπολογίζεται όχι σε ολόκληρο αλλά σε frames του σήματος. Άλλωστε το pitch, σε ένα μουσικό αρχείο που προέκυψε από εκτέλεση μουσικού οργάνου, μεταβάλλεται συνήθως με πολύ γρήγορο ρυθμό. Ως εκ τούτου, για την αναπαράσταση του pitch απαιτούνται τρεις διαστάσεις: Το κανάλι του κοχλία (αντιστοιχεί στη συχνότητα), η αυτοσυσχέτιση σε σχέση με την καθυστέρηση(αντιστοιχεί στην περίοδο του pitch), και ο χρόνος. Σε ένα σήμα διακριτού χρόνου , η τρίτη διάσταση στην αναπαράσταση θα αντιστοιχεί σε χρονικές στιγμές (frames) όπου θα εξετάζεται το σήμα σε ένα χρονικό διάστημα (παράθυρο). Η τρισδιάστατη αυτή μορφή αναπαράστασης ονομάζεται κορελόγραμμα (σχήμα 3.14). Σε κάθε frame αντιστοιχεί μια φέτα κορελογράμματος (correlogram slice).



Σχήμα 3. 14 Το κορελόγραμμα

Οι πράξη της παραθύρωσης γίνεται συνήθως πριν την αυτοσυσχέτιση. Συγκεκριμένα:

$$X_w(t-t_0) = x(t)w(t-t_0)$$

(3.13)

$$Rx_w x_w = \int_{-\infty}^{\infty} x_w(t-t_0)x_w(t-\tau,t_0)dt$$

Από τα χρονικά στιγμιότυπα του κορελογράμματος ,μέσω ενός αλγορίθμου που αναζητεί τα τοπικά μέγιστα στον κοχλία και τα ομαδοποιεί σαν αρμονικά πολλαπλάσια ,υπολογίζεται το pitch.

Το pitch χρησιμοποιείται αυτούσια σαν χαρακτηριστικό στον μηχανισμό της ταξινόμησης. Μπορεί όμως να είναι πιο χρήσιμο να ελέγχεται απλά το κατά πόσο το pitch του ήχου προς αναγνώριση αποτελεί ένα από τα δυνατά pitch σύμφωνα με το εύρος του κάθε οργάνου. Στις επόμενες παραγράφους φαίνεται και η σημαντικότητα του pitch όταν συνδυάζεται και με άλλα χαρακτηριστικά.

- Vibrato, Tremolo

Το Vibrato χρησιμοποιείται σε συνδυασμό με το pitch. Εξετάζεται το κατά πόσο το η συχνότητα του pitch υφίσταται κάποιο είδος περιοδικής διαμόρφωσης συχνότητας. Λόγω της διαφορετικότητας της διαμόρφωσης αυτής από όργανο σε όργανο, ή και της μη ύπαρξης vibrato σε ορισμένα όργανα, η τεχνική αυτή βοηθά στην αναγνώριση.

Κατά τον ίδιο τρόπο χρησιμοποιείται και το tremolo σαν περιοδική διαμόρφωση της ενέργειας του σήματος.

- Ατάκα (attack transients)

Από τις πολυάριθμες μελέτες που έχουν γίνει με δείγματα μεμονωμένες νότες από όργανα σχετικά με την χρησιμότητα των χαρακτηριστικών της ατάκας (το μικρό χρονικό διάστημα από την στιγμή που ενεργοποιείται μια νέα νότα μέχρι να σταθεροποιηθεί η ενέργεια του ήχου). Οι μελέτες έχουν δείξει ότι τα στοιχεία αυτά μπορούν να προκαλέσουν αξιόλογη διαχωρισιμότητα ανάμεσα στα διάφορα μουσικά όργανα. Τέτοια στοιχεία είναι για παράδειγμα ο ρυθμός αύξησης του πλάτους των αρμονικών (σε dB/ms), που υπολογίζεται με μεθόδους παρόμοιες με την ανίχνευση ακμών στην εικόνα .Ωστόσο είναι ένα ζήτημα το κατά πόσο μπορεί να υπολογιστεί με ακρίβεια το χρονικό διάστημα που διαρκεί μια ατάκα (πολλοί ερευνητές θέτουν σαν τέλος της ατάκας το χρονικό σημείο όπου το σήμα, ξεκινώντας από το 0, φτάνει στο 75% του συνολικού πλάτους). Κατά τον ίδιο τρόπο, χαρακτηριστικά αποτελούν το steady state και το decay (ουρά του ήχου), που όμως η έλλειψη αξιοπιστίας στην εξεύρεσή τους τα καθιστά αρκετά αμφιλεγόμενα χαρακτηριστικά. Η διερεύνησή τους σε ένα σήμα που προέρχεται από κανονική ηχογράφηση με συνεχόμενες νότες είναι ακόμη δυσκολότερη, γεγονός που καθιστά ακόμα πιο αναξιόπιστη τη χρησιμοποίηση τους σαν

χαρακτηριστικά για την αναγνώριση. Πάντως θεωρείται σχεδόν βέβαιο ότι μελλοντικά η χρησιμότητα τέτοιου είδους χαρακτηριστικών θα διερευνηθεί περαιτέρω.

3.5.3 Χαρακτηριστικά ενέργειας σήματος

- Κεντρική συχνότητα φάσματος

Η κεντρική συχνότητα του φάσματος αποτελεί ένα από τα απλούστερα χαρακτηριστικά της κατηγορίας. Αντιπροσωπεύει το σημείο ισορροπίας του φάσματος και είναι μέτρο για την λαμπρότητα (brightness), δηλαδή το κατά πόσο ένα σήμα είναι πλούσιο σε αρμονικές υψηλών συχνοτήτων. Ο μαθηματικός τύπος είναι :

$$C = \frac{\sum_{n=1}^{N} M_t[n] \cdot n}{\sum_{n=1}^{N} M_t[n]}, \qquad (3.14)$$

όπου M το πλάτος (ή η ενέργεια) του σήματος και n οι μπάντες συχνοτήτων του μοντέλου για τον κοχλία (ή τα σημεία του FFT).

- Κατώφλι ενέργειας

Είναι η συχνότητα που αντιστοιχεί στο r % της ενέργειας του φάσματος (ξεκινώντας από τις χαμηλές συχνότητες). Ισχύει:

$$\sum_{n=1}^{R} M_t[n] = r \cdot \sum_{n=1}^{N} M_t[n]. \qquad (3.15)$$

Προφανώς πρόκειται για μια γενίκευση της κεντρικής συχνότητας φάσματος (συχνότητα που αντιστοιχεί στο r=50).

- Παράγωγος (Flux)

Αποτελεί μέτρο της τοπικής μεταβλητότητας φάσματος. Είναι:

$$F = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2 \qquad (3.16)$$

όπου $N_t[n]$ το κανονικοποιημένο πλάτος του Fourier στο παράθυρο t

- Ρυθμός Zero-crossing

Είναι το πλήθος των διασταυρώσεων του σήματος με τον άξονα των x (τιμή πλάτους ίση με το 0). Το zero-crossing αποτελεί ένα μέτρο για το βαθμό θορύβου (υψηλές συχνότητες) του σήματος. Δίνεται από τον τύπο:

$$Z = \sum_{n=1}^{N} |s(x[n]) - s(x[n-1])| \qquad (3.17)$$

όπου $x[n]$ το πλάτος σήματος στο διακριτό χρόνο n και $s()$ η συνάρτηση signum={1,0}.

- Ενέργεια βραχέως χρόνου (Short Time Energy)

Η ενέργεια βραχέως χρόνου αποτελεί μέτρο της στιγμιαίας ενέργειας του σήματος. Δίνεται από τον τύπο:

$$E_n = \frac{1}{N} \sum_{m} [x(m)w(n-m)]^2, \qquad (3.18)$$

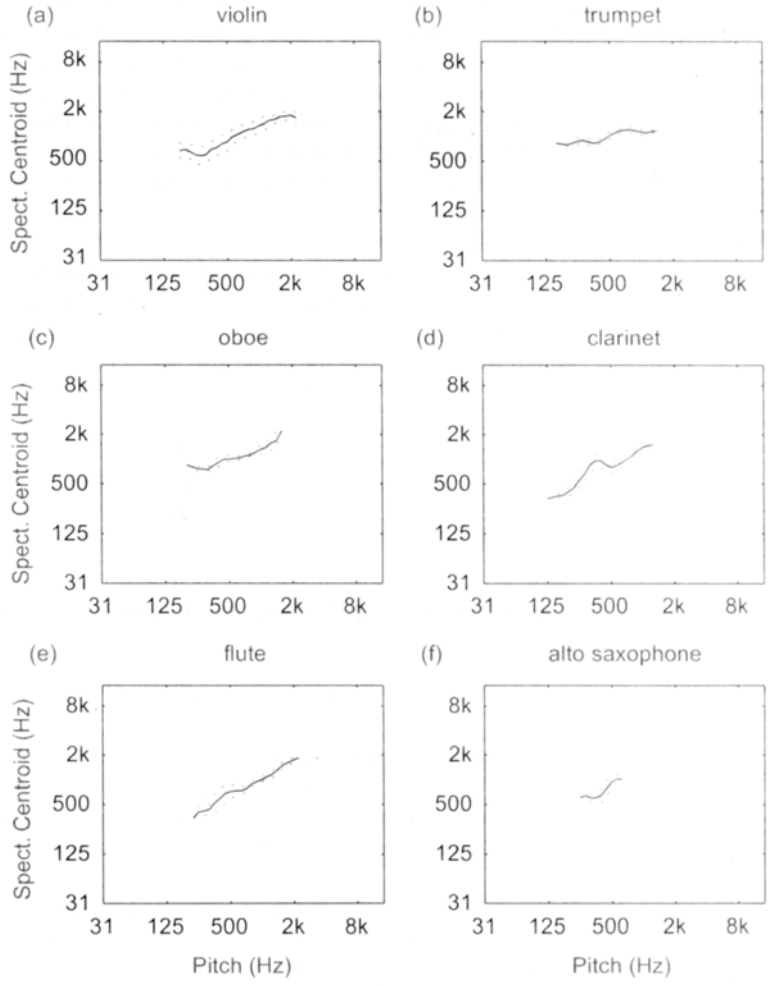
όπου :

$$w(n) = \begin{cases} 1/2 & 0 \le n \le N-1 \\ 0 & \text{otherwise.} \end{cases} \qquad (3.19)$$

Χρησιμοποιείται κυρίως στην επεξεργασία φωνής και συγκεκριμένα στην διαδικασία της κατάτμησης (segmentation) κατά το διαχωρισμό του σήματος από τη σιωπή (silence) και των έμφωνων από τους άφωνους ήχους.

- Έλλειψη αρμονικότητας (inharmonicity)

Πρόκειται για ένα πολύ σημαντικό χαρακτηριστικό που μετρά την έλλειψη αρμονικότητας στο σήμα, δηλαδή το κατά πόσο το σήμα αναλύεται ακριβώς σε ακέραια πολλαπλάσια της θεμελιώδους συχνότητας ή τα spikes (βουνά) βρίσκονται σε μη αρμονικές συχνότητες.



Σχήμα 3. 15 Κεντροειδής συχνότητα φάσματος σαν συνάρτηση του pitch για όργανα της ορχήστρας.

Κεφάλαιο 4 : Ταξινόμηση

Η ταξινόμηση αποτελεί ένα υποπεδίο της θεωρίας αποφάσεων. Βασίζεται στην απλή θεώρηση ότι υπάρχουν κάποιες βασικές ιδιότητες οι οποίες μπορούν να κατηγοριοποιήσουν φαινομενικά διαφορετικά πρότυπα κάτω από την ταμπέλα κάποιων κοινών στοιχείων.

Στην περίπτωση της αναγνώρισης οργάνων, μια σονάτα Beethoven για βιολί μπορεί να διαφέρει ριζικά στο περιεχόμενο από την καντέντζα στο concerto για βιολί του Schoenberg, ωστόσο ο άνθρωπος μπορεί να αναγνωρίσει με ευκολία ότι και στις δύο περιπτώσεις το όργανο που ακούγεται είναι το ίδιο.

Στο κεφάλαιο 3 έγινε αναφορά στις μεθόδους και την διαδικασία της εξαγωγής του σετ χαρακτηριστικών. Αυτά τα διανύσματα χαρακτηριστικών χρησιμοποιούνται για να αντιστοιχίσουν κάθε αντικείμενο σε μια συγκεκριμένη κατηγορία. Αυτό είναι το μέρος ταξινόμησης, το οποίο αποτελεί την τελευταία βασική δομική μονάδα ενός συστήματος αναγνώρισης μουσικών οργάνων.

Υπάρχει μια μεγάλη ποικιλία των τεχνικών ταξινόμησης οι οποίες κατηγοριοποιούνται σε σχέση με τον τρόπο προσέγγισης . Στη συνέχεια θα γίνει παρουσίαση των κατηγοριών αυτών και των σημαντικότερων μεθόδων κάθε μίας.

4.1 Γενικά

Εκπαίδευση και μάθηση

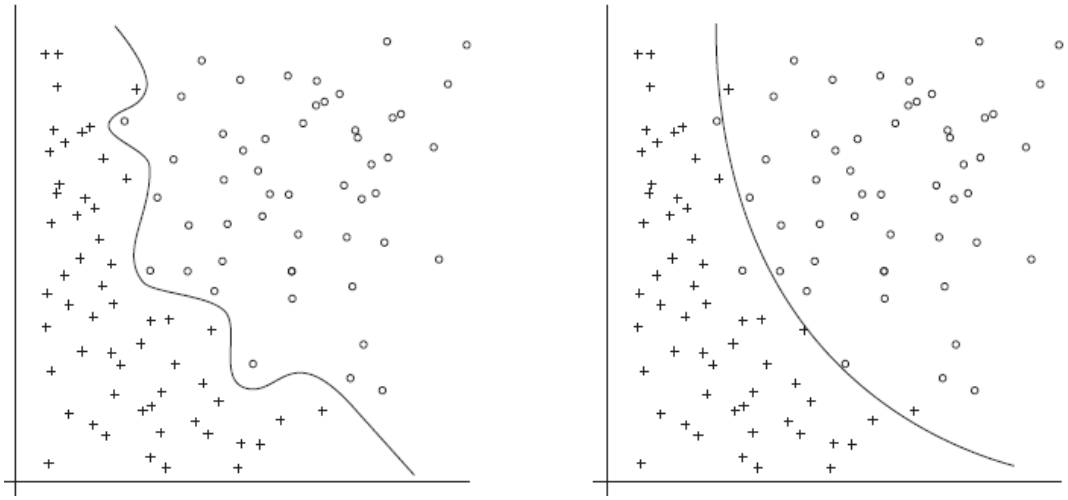
Η δημιουργία ενός ταξινομητή συνήθως ισοδυναμεί με την διαδικασία της παραμετροποίησης ενός μοντέλου μέσα από την εκπαίδευση. Η εκπαίδευση λοιπόν είναι η διαδικασία κατά την οποία χρησιμοποιείται πληροφορία από ένα αριθμό δειγμάτων προκειμένου να οριστούν οι παράμετροι ενός ταξινομητή.

Η ταξινόμηση πολλές φορές ονομάζεται μάθηση υπό επίβλεψη. Η μάθηση υπό επίβλεψη χρησιμοποιεί διανύσματα χαρακτηριστικών και τα αντιστοιχίζει σε προκαθορισμένες κατηγορίες προκειμένου να ορίσει τις παραμέτρους του ταξινομητή. Ταξινόμηση όμως μπορεί να γίνει και χωρίς επίβλεψη, χωρίς δηλαδή να προηγείται κάποιου είδους εκπαίδευση. Έτσι, με τον όρο μάθηση χωρίς επίβλεψη ονομάζουμε τον ταξινομητή ο οποίος πραγματοποιεί από μόνος του φυσικές ομαδοποιήσεις κατά τη διάρκεια της ίδιας της ταξινόμησης.

Ικανότητα γενίκευσης

Ένα σημαντικό ζήτημα είναι το κατά πόσο το σύστημα ταξινόμησης έχει προσαρμοστικότητα στα καινούργια δεδομένα (ικανότητα γενίκευσης) ή είναι αυστηρά προσαρμοσμένο στα δεδομένα της μάθησης (overfitting). Στην

δεύτερη περίπτωση, παρόλο που η προσαρμογή στα δεδομένα μάθησης μπορεί να είναι φαινομενικά άριστη, η ικανότητα του συστήματος να ταξινομεί καινούργια δεδομένα να είναι μειωμένη (σχήμα 4.1).



Σχήμα 4. 1 Παράδειγμα για το φαινόμενο του overfitting. Στην αριστερή εικόνα το σύστημα έχει ταξινομήσει τέλεια όλα τα δεδομένα από τη μάθηση, ωστόσο έχει μειωμένη πιθανότητα να ανταποκριθεί με επιτυχία σε καινούργια δεδομένα. Αντίθετα, στην εικόνα στα δεξιά, το σύνορο της απόφασης έχει απλούστερη μορφή και πιθανότατα αντανακλά την πραγματική φύση των δεδομένων. Παρόλο που κάποια δεδομένα έχουν ταξινομηθεί λάθος, η συνολική απόδοση σε άγνωστες εισόδους θα είναι καλύτερη**.**

Όπως αναφέρθηκε και στην αρχή του κεφαλαίου, οι ταξινομητές διακρίνονται σε κατηγορίες ανάλογα με την προσέγγισή τους στο εκάστοτε πρόβλημα. Η προσέγγιση αυτή καθορίζεται σε μεγάλο βαθμό και από την εκ των προτέρων γνώση που έχουμε για τη μορφή που έχει η κατανομή των δεδομένων στο διανυσματικό χώρο των χαρακτηριστικών. Έτσι, πολλές φορές είναι δόκιμο να προσεγγίσουμε ένα πρόβλημα ταξινόμησης μέσω μιας κατανομής γνωστής μορφής και πιθανοτικών σχέσεων, ή στην περίπτωση που δεν ξέρουμε τίποτα για την κατανομή των δεδομένων του προβλήματος να προσπαθήσουμε να αναζητήσουμε την κατανομή μέσα από τη διαδικασία της μάθησης. Τέλος μπορούμε να προσεγγίσουμε το πρόβλημα και με τεχνικές προσομοίωσης του ανθρώπινου τρόπου αντίληψης (νευρωνικά δίκτυα). Τέλος είναι δυνατός και ο συνδυασμός μεθόδων διαφορετικών προσεγγίσεων.

Στη συνέχεια του κεφαλαίου θα γίνει αναφορά στις σημαντικότερες μεθόδους ταξινόμησης, αναφορικά με τη φύση του προβλήματος που επιζητούν να επιλύσουν.

4.2 Μπεϊζιανή θεωρία απόφασης (Bayesian Decision Theory)

Σύμφωνα με τα προλεγόμενα σχετικά με την ταξινόμηση, το ζητούμενο είναι η απόφαση του συστήματος για την κατηγορία στην οποία ανήκει ένα αντικείμενο. Μια τέτοια απόφαση ενδέχεται να είναι και λανθασμένη. Η

επιτυχία της απόφασης λοιπόν έγκειται στην μεγιστοποίηση της πιθανότητας της σωστής απόφασης. Μία τέτοια προσέγγιση του προβλήματος είναι η μπεϊζιανή θεωρία απόφασης, η οποία κάνει χρήση της θεωρίας πιθανοτήτων για να διαλέξει, ανάμεσα στις δυνατές περιπτώσεις για απόφαση, αυτή που ενδέχεται να είναι η σωστότερη.

Σε αυτό το σημείο κρίνεται χρήσιμο να οριστούν μαθηματικά τα δομικά στοιχεία της θεωρίας.

- *Τα ενδεχόμενα* συμβολίζονται με ω *(*κατηγορίες).
- Η *εκ των προτέρων πιθανότητα (prior propability)* $P(\omega_j)$ περιγράφει την γνώση που έχουμε από πριν για τις κατηγορίες στις οποίες ανηκουν τα δεδομένα. Για παράδειγμα, στην περίπτωση αναγνώρισης οργάνων, η πιθανότητα αυτή ισοδυναμεί με την κατανομή που ξέρουμε για τα δεδομένα (π.χ ότι από τα 200 δείγματα μουσικής εκτέλεσης τα 50 αντιστοιχούν σε βιολί (ενδεχόμενο ω_1) και τα 150 σε κοντραμπάσο (ενδεχόμενο ω_2), με αντίστοιχες a priori πιθανότητες $P(\omega_1)$= 25% και $P(\omega_2)$= 75%.
- *Η συνάρτηση πυκνότητας δεσμευμένης πιθανότητας κλάσης* $p(\mathbf{x}|\omega_j)$, εκφράζει την πιθανότητα μια κατηγορία ω_j να περιλαμβάνει ένα διάνυσμα **x .**
- Η *εκ των υστέρων (a posteriori) πιθανότητα* $P(\omega_j|\mathbf{x})$ δηλώνει την πιθανότητα ένα διάνυσμα **x** να ανήκει σε μια κατηγορία ω_j . Αυτή είναι και η μορφή της πιθανότητας που επιθυμούμε να υπολογίσουμε.

Σύμφωνα με τη φόρμουλα Bayes, ισχύει:

$$P(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{p(\mathbf{x})}, \tag{4.1}$$

όπου:

$$p(\mathbf{x}) = \sum_{j=1}^{c} p(\mathbf{x}|\omega_j)P(\omega_j). \tag{4.2}$$

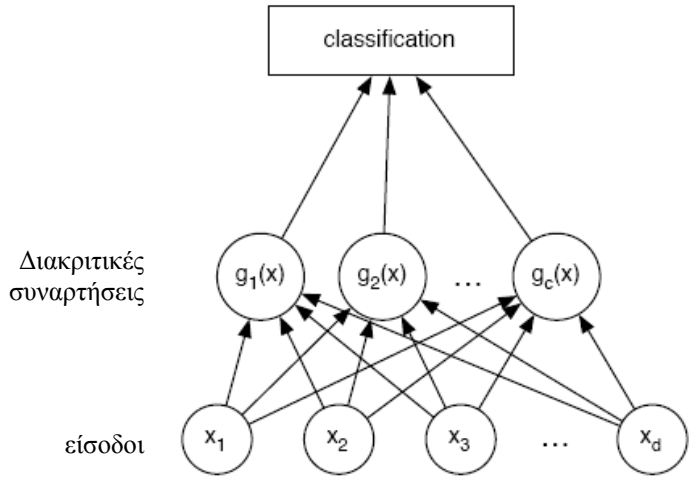
Προκειμένου να ελαχιστοποιηθεί η πιθανότητα λάθους, το σύστημα θα πρέπει να αποφασίσει για το ω_j με δεδομένο ένα διάνυσμα **x** έτσι ώστε η πιθανότητα $P(\omega_j|\mathbf{x})$ να είναι η μέγιστη . Αυτός λέγεται ο *μπεϊζιανός κανόνας απόφασης* και μπορεί να εκφραστεί ως:

$$\omega_i \text{ if } P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x}) \text{ for all } i \neq j. \tag{4.3}$$

Έτσι, ένας ταξινομητής Bayes είναι ένα σύστημα το οποίο υπολογίζει έναν αριθμό *συναρτήσεων ταξινόμησης* *c* της μορφής $g_i(\mathbf{x})$, όπου *i = 1,…,c ,* και επιλέγει την κατηγορία η οποία παρουσιάζει τη μεγαλύτερη διακριτική ικανότητα. Ο ταξινομητής αντιστοιχίζει ένα διάνυσμα **x** σε μια κατηγορία ω_i *αν* $g_i(\mathbf{x}) < g_j(\mathbf{x})$ για όλα τα $i \neq j$. Η συνάρτηση ταξινόμησης με τη μεγαλύτερη τιμή αντιστοιχεί στην μέγιστη εκ των υστέρων πιθανότητα.

Οι συναρτήσεις ταξινόμησης δύνονται από τον τύπο:

$$g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i) \tag{4.4}$$



Σχήμα 4. 2 Δομή ταξινομητή με d εισόδους x_i και c διακριτικές συναρτήσεις $g_i(x)$. Για κάθε είσοδο υπολογίζεται η τιμή των διακριτικών συναρτήσεων. Κάθε τέτοια συνάρτηση αντιστοιχεί και σε μια κατηγορία ω_i . Το σύστημα διαλέγει την κατηγορία για την οποία το $g_i(x)$ γίνεται μέγιστο.

Σύμφωνα με τα παραπάνω, αρκεί να ξέρουμε τις πυκνότητες δεσμευμένης πιθανότητας κλάσης, έτσι ώστε να βρούμε τις τιμές των συναρτήσεων ταξινόμησης και στη συνέχεια, με βάση τον κανόνα απόφασης Bayes, την μέγιστη εκ των υστέρων πιθανότητα ένα διάνυσμα εισόδου να ανήκει σε μια συγκεκριμένη κατηγορία .Αυτή αποτελεί την κατηγορία στην οποία θα εισαχθεί ένα δείγμα.

Στα περισσότερα προβλήματα ταξινόμησης οι πυκνότητες δεσμευμένης πιθανότητας κλάσης δεν είναι γνωστές. Σε περίπτωση που οι a priori πιθανότητες είναι γνωστές, αρκεί ο προσδιορισμός τους ώστε να οριστεί πλήρως ένα σύστημα ταξινόμησης.

Για τον υπολογισμό των πυκνοτήτων δεσμευμένης πιθανότητας υπάρχουν δύο προσεγγίσεις: η παραμετρική και η μη παραμετρική.

4.2.1 Παραμετρική προσέγγιση

Η παραμετρική προσέγγιση χρησιμοποιείται στις περιπτώσεις όπου μπορεί να γίνει μια υπόθεση για την μορφή των πυκνοτήτων δεσμευμένης πιθανότητας κλάσης, δηλαδή τα διανύσματα των χαρακτηριστικών **x** ακολουθούν μια συγκεκριμένη κατανομή. Ξέροντας τη συνάρτηση κατανομής ,αρκεί να προσδιορίσουμε τις παραμέτρους στις γενικές εξισώσεις και έχουμε ορίσει πλήρως τον ταξινομητή.

Έστω η περίπτωση που η πυκνότητα δεσμευμένης πιθανότητας κλάσης είναι η πυκνότητα κανονικής κατανομής. Η γενική της μορφή είναι:

$$p(x) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \tag{4.5}$$

όπου **x** είναι ένα διάνυσμα διάστασης d , Σ είναι ο $d*d$ πίνακας συμμεταβλητότητας και Σ^{-1}ο αντίστροφός του.
Αυτή είναι μια συχνή περίπτωση, όπου όλα τα διανύσματα **x** δεδομένης μιας κλάσης ω_j κατανέμονται γύρω από ένα κεντρικό διάνυσμα μ_j με παρουσία θορύβου κανονικής κατανομής. Πολλά φυσικά συστήματα μοντελοποιούνται με αυτόν τον τρόπο.

Στον παραπάνω τύπο αρκεί να υπολογίσουμε το μ και τον πίνακα συμμεταβλητότητας. Μία από τις συνήθεις τεχνικές για αυτό είναι ο υ*πολογισμός των παραμέτρων μέγιστης πιθανοφάνειας ,* δηλαδή η προσέγγιση του μ που μεγιστοποιεί την πιθανότητα να λάβουμε πίσω τα δείγματα εκπαίδευσης. Για παράδειγμα, θα μπορούσαμε να έχουμε:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x_k} \tag{4.6}$$

όπου $\hat{\mu}$ μια προσέγγιση του μ σαν στατιστικός μέσος όρος.

4.2.2 Μη παραμετρική προσέγγιση

Στην περίπτωση όπου δεν ξέρουμε την μορφή των πυκνοτήτων δεσμευμένης πιθανότητας κλάσης , ή όπου το σύστημα, για να προσομοιώσει μια κατανομή δειγμάτων, χρειάζεται παραπάνω από ένας στατιστικός μέσος (εξίσωση 4.6), οι παραμετρικές μέθοδοι κρίνονται αποτελεσματικοί. Κρίνεται σκόπιμο στις περιπτώσεις αυτές, ο ταξινομητής να εκπαιδεύεται αποκλειστικά από τα διανύσματα των χαρακτηριστικών. Μέσω αυτών θα βρίσκονται οι πυκνότητες δεσμευμένης πιθανότητας $p(\mathbf{x}|\omega_j)$, ή ακόμα και απευθείας οι εκ των υστέρων πιθανότητες $P(\omega_j|\mathbf{x})$. Παρακάτω αναφέρονται οι σημαντικότερες μέθοδοι μη παραμετρικής ταξινόμησης.

- Ο κανόνας του κοντινότερου γείτονα

Μία από τις πιο γνωστές μη παραμετρικές μεθόδους είναι ο κανόνας του κοντινότερου γείτονα. Για κάθε κατηγορία υπάρχει και ένα διάνυσμα-κέντρο. Συνολικά τα κέντρα αυτά από όλες τις τάξεις συνιστούν ένα διάνυσμα: **D**={$\mathbf{x_1},\ldots,\mathbf{x_n}$}. Προκειμένου να κατατάξουμε ένα διάνυσμα **x** , σύμφωνα με τον κανόνα του κοντινότερου γείτονα αυτό θα καταταχθεί στην ίδια κατηγορία με το διάνυσμα **x'** $\in$ **D**, όπου **x'** το διάνυσμα το κοντινότερο , με βάση κάποιο μέτρο απόστασης, ως προς το **x**.

- Ο κανόνας του k- κοντινότερου γείτονα

Πρόκειται για μια γενίκευση του κανόνα κοντινότερου γείτονα, με τη διαφορά ότι αντί για την απόσταση από τα διανύσματα-κέντρα εξετάζεται η απόσταση από τα k γειτονικά διανύσματα σε μια απόσταση d από το διάνυσμα προς ταξινόμηση x. Από το συνολικό αριθμό των κοντινών διανυσμάτων υπολογίζεται ο αριθμός αυτών που ανήκουν σε κάθε κατηγορία, και οι συχνότητες της κάθε κατηγορίας αντικαθιστούν στον κανόνα Bayes τις $P(\omega_j|\mathbf{x})$ εκ των υστέρων πιθανότητες. Η μεγαλύτερη κρίνει την ταξινόμηση και του εν λόγω δείγματος σε μία από τις κατηγορίες.

Ανάλογα με το εκάστοτε πρόβλημα ταξινόμησης, είναι δόκιμο να χρησιμοποιείται διαφορετικός τρόπος για την μέτρηση της απόστασης των δειγμάτων στον διανυσματικό χώρο. Η πιο συνηθισμένη απόσταση είναι η απόσταση Minkowski, η οποία δίνεται από το μαθηματικό τύπο:

$$d_p(\mathbf{x_i},\mathbf{x_j}) = \sqrt[p]{\sum_{k=1}^{d}(|x_{i,k}-x_{j,k}|)^p} = \|\mathbf{x_i}-\mathbf{x_j}\|_p. \quad (4.7)$$

Η εφαρμογή της για p=2 είναι η γνωστή ευκλείδεια απόσταση, η οποία και χρησιμοποιείται ευρέως στα δισδιάστατα και τρισδιάστατα προβλήματα.

4.3 Τεχνητά νευρωνικά δίκτυα

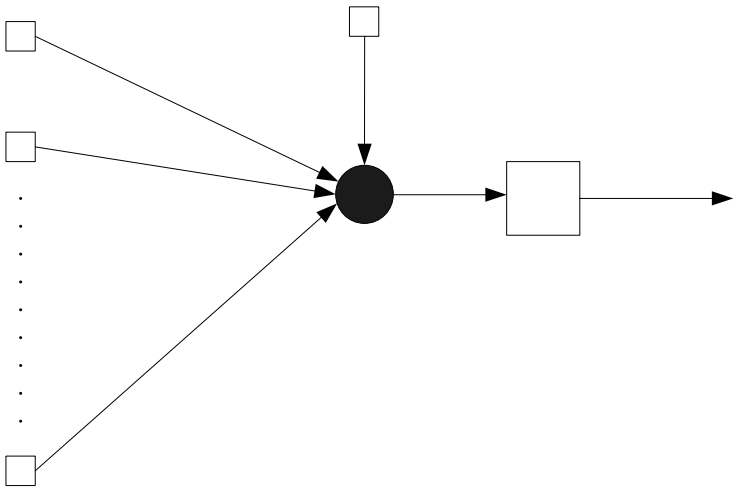
4.3.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζονται οι βασικές αρχές λειτουργίας των Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ). Αρχικά, περιγράφονται οι βασικές κατηγορίες νευρωνικών δικτύων. Ακολουθούν τα χαρακτηριστικά και οι ιδιότητες τους, καθώς επίσης και εφαρμογές τους σε τομείς της τεχνολογίας. Βαρύτητα δίνεται στα πολυεπίπεδα νευρωνικά δίκτυα και στην εκπαίδευση με

τον αλγόριθμο της όπισθεν διάδοσης σφάλματος. Τέλος, αναλύεται ο αλγόριθμος αυτός καθώς και οι τροποποιημένες μορφές του.

4.3.2 Περιγραφή Τεχνητών Νευρωνικών Δικτύων

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) προέρχονται από την προσπάθεια προσομοίωσης του τρόπου λειτουργίας του ανθρώπινου εγκεφάλου και γενικότερα του νευρικού συστήματος. Είναι παράλληλοι κατανεμημένοι επεξεργαστές αποτελούμενοι από απλές μονάδες επεξεργασίας που λέγονται νευρώνες και οι οποίοι έχουν τη δυνατότητα αποθήκευσης «γνώσης» μέσω της εμπειρίας που αποκτούν κατά την διαδικασία εκπαίδευσης [4]. Η στοιχειώδης μονάδα επεξεργασίας ενός ΤΝΔ λέγεται *τεχνητός νευρώνας (neuron)*. Κάθε νευρώνας έχει έναν συγκεκριμένο αριθμό εισόδων και εξόδων αλλά και συνάψεων οι οποίες τον συνδέουν με τους άλλους νευρώνες και χαρακτηρίζονται από μία τιμή βάρους. Οι τιμές των βαρών των συνδέσεων αποτελούν την γνώση που είναι αποθηκευμένη στο δίκτυο και καθορίζουν την λειτουργικότητά του. Ένας τυπικός τεχνητός νευρώνας φαίνεται στο σχήμα 4.3.



Σχήμα 4.3: Τεχνητός Νευρώνας

Όπως φαίνεται ο νευρώνας είναι μία σύναψη των εισόδων του x_i $i=1,2,\ldots,p$ δηλαδή άθροισμα με συντελεστές βάρους w_i $i=1,2,\ldots,p$. Ο νευρώνας ενδεχομένως μπορεί να έχει μία επιπλέον είσοδο θ , που είναι γνωστή ως *πόλωση (bias),* και χρησιμοποιείται προκειμένου να αυξήσει ή να μειώσει το αποτέλεσμα της σύναψης ενός νευρώνα ανάλογα με το αν αυτό είναι θετικό ή αρνητικό. Έτσι το ολικό άθροισμα του νευρώνα δίνεται από τον τύπο

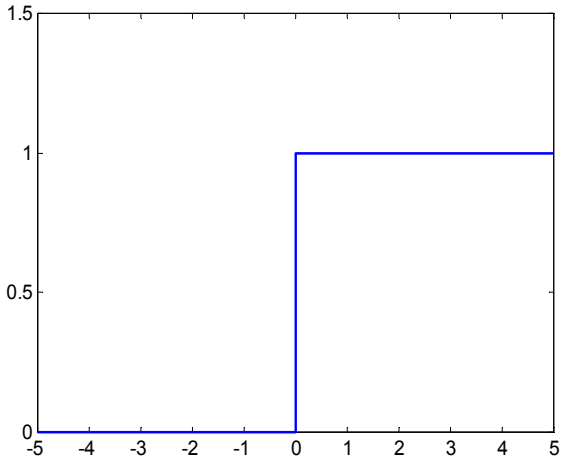
$$u = \sum_{i=1}^{p} w_i x_i - \theta \tag{4.8}$$

Το άθροισμα αυτό αποτελεί το όρισμα μιας γραμμικής ή μη γραμμικής συνάρτησης μετασχηματισμού *f*, η οποία ονομάζεται συνάρτηση ενεργοποίησης οπότε η έξοδος y του ΤΝΔ δίνεται από τη σχέση

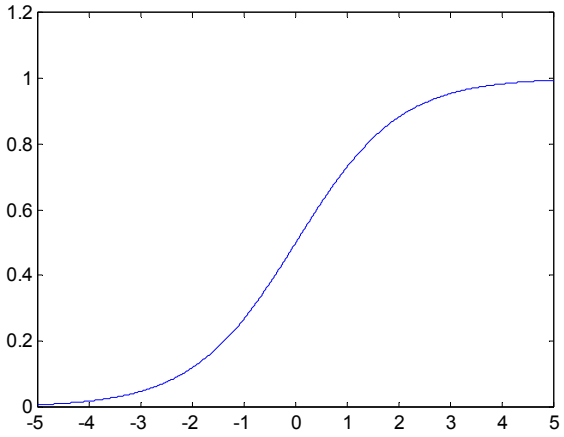
$$y = f(u) \qquad (4.9)$$

Υπάρχουν διάφορες συναρτήσεις ενεργοποίησης οι κυριότερες των οποίων είναι

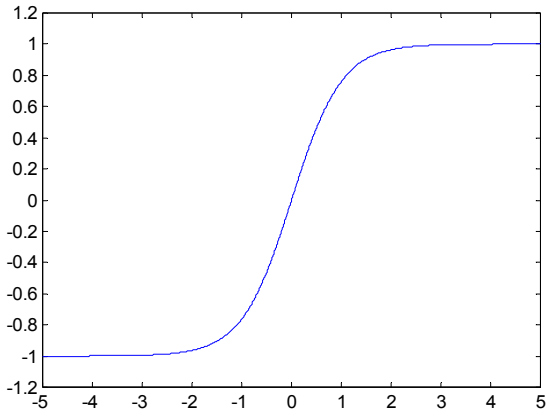
- Συνάρτηση κατωφλίου (σχ. 4α)
- Σιγμοειδής εφαπτομενική συνάρτηση (σχ. 4β)
- Σιγμοειδής υπερβολική εφαπτομενική συνάρτηση (σχ. 4γ)



(α)



(β)



(γ)

Σχήμα 4.4: Συναρτήσεις ενεργοποίησης

Οι νευρώνες σε ένα ΝΔ είναι οργανωμένοι σε επίπεδα (layers). Τα εξωτερικά σήματα εφαρμόζονται στους νευρώνες του επιπέδου εισόδου (input layer). Οι έξοδοι των νευρώνων του επιπέδου εισόδου μεταφέρουν τις πληροφορίες τους στους νευρώνες των ενδιάμεσων ή κρυμμένων επιπέδων (hidden layers), οι οποίοι δεν έχουν άμεση σχέση με το περιβάλλον. Τέλος, οι νευρώνες του επιπέδου εξόδου (output layer) ενημερώνουν το χρήστη για την έξοδο του ΝΔ. Όταν καθένας από τους νευρώνες ενός επιπέδου συνδέεται με όλους τους νευρώνες του επόμενου επιπέδου, τότε το ΝΔ είναι πλήρως συνδεδεμένο (fully connected), αλλιώς είναι μερικώς συνδεδεμένο (partially connected).

4.3.2.1 Ιδιότητες Τεχνητών Νευρωνικών Δικτύων

Τα ΤΝΔ έχουν ιδιαίτερη υπολογιστική ισχύ λόγω της ιδιότητάς τους να εκπαιδεύονται και να επεξεργάζονται παράλληλα πληροφορίες. Άμεση συνέπεια της ικανότητας εκπαίδευσης είναι η *ικανότητα γενίκευσης,* δηλαδή η ικανότητα τους να ανταποκρίνονται σε δεδομένα που δεν έχουν αντιμετωπίσει άλλη φορά αλλά για τα οποία μπορούν να εξάγουν συμπεράσματα από την εκπαίδευση την οποία έχουν υποστεί. Οι κυριότερες εφαρμογές τους είναι σε μη-γραμμικά προβλήματα που δεν μπορούν εύκολα να περιγραφούν με κανόνες ή μαθηματικούς τύπους. Η εκτεταμένη χρήση σε τέτοιου είδους προβλήματα οφείλεται στη μη απαίτηση *a priori* υποθέσεων για τη στατιστική κατανομή των χρησιμοποιούμενων δεδομένων, αλλά και στη δυνατότητα εξαγωγής κρυμμένης πληροφορίας από αυτά, κάτι που δεν μπορεί να γίνει εύκολα με τις συνήθεις στατιστικές μεθόδους.

Ένα άλλο χαρακτηριστικό των ΝΔ είναι η ανοχή τους σε σφάλματα, γεγονός που οφείλεται στο μεγάλο αριθμό νευρώνων από τον οποίο αποτελούνται. Έτσι, βλάβη σε έναν ή περισσότερους νευρώνες δεν επηρεάζει αισθητά τη συνολική απόδοση του ΤΝΔ[5]. Επιπλέον, η δυνατότητα που υπάρχει για Very Large Scale Integration (VLSI) [6] υλοποίηση των ΤΝΔ τα καθιστά ένα ισχυρό εργαλείο, αφού ακόμη και σύνθετες μορφές ΤΝΔ μπορούν

με τη μορφή ολοκληρωμένου κυκλώματος να χρησιμοποιηθούν σε πραγματικού χρόνου εφαρμογές με σχετικά χαμηλό κόστος.

Πρέπει πάντως να τονιστεί ότι τα ΤΝΔ δεν δίνουν πάντα βέλτιστες λύσεις σε όλα προβλήματα. Η χρήση τους δεν ενδείκνυται σε προβλήματα για τα οποία έχουν ήδη αναπτυχθεί μαθηματικά ή αλγοριθμικά μοντέλα επίλυσης και τα οποία τις περισσότερες φορές είναι πιο αξιόπιστα από τα ΤΝΔ. Πέρα από αυτό, η υλοποίηση μέσω ΤΝΔ είναι ατελέσφορη σε τέτοιες περιπτώσεις αφού απαιτείται μελέτη διαδικασίας επιλογής του κατάλληλου ΤΝΔ, επιλογής και επεξεργασίας των διανυσμάτων εκπαίδευσης, καθώς και της εκπαίδευσης του ΤΝΔ. Επίσης τα ΤΝΔ είναι ένα «μαύρο κουτί» (black box) για τους επιστήμονες στο οποίο είναι δύσκολη η εξήγηση του τρόπου με τον οποίο φτάνει στην επίλυση του προβλήματος, με αποτέλεσμα να μην εξάγεται επαρκής πληροφορία για παραπέρα έρευνα ή κατανόηση των βαθύτερων αιτιών που διέπουν το εκάστοτε πρόβλημα.

4.3.2.2 Τύποι Νευρωνικών Δικτύων

Τα ΝΔ διακρίνονται σε διάφορες κατηγορίες ανάλογα
(α) με την αρχιτεκτονική
(β) το χρησιμοποιούμενο αλγόριθμο εκπαίδευσης.

Η επιλογή του κατάλληλου ΝΔ εξαρτάται από τον τύπο του μελετούμενου προβλήματος

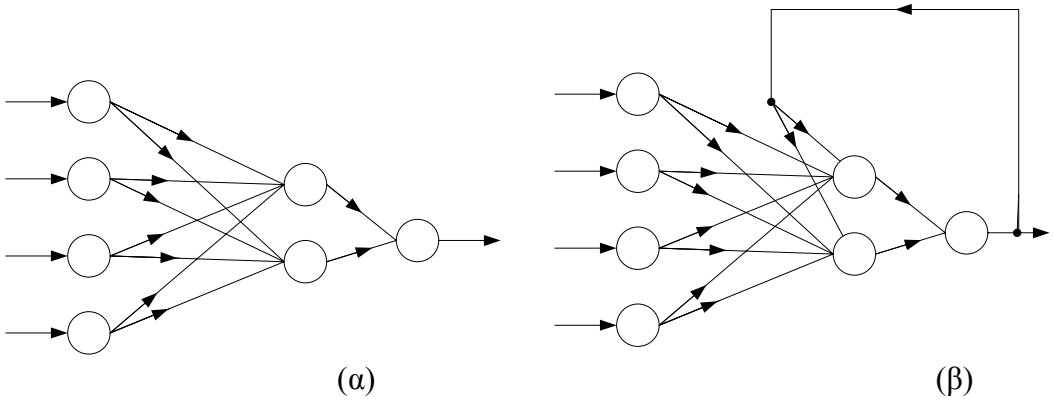
4.3.2.3 Αρχιτεκτονική Νευρωνικών Δικτύων

Οι δύο βασικές ιδιότητες που καθορίζουν την αρχιτεκτονική ενός ΤΝΔ είναι το πλήθος των στρωμάτων και ο τρόπος σύνδεσης των νευρώνων μεταξύ τους.
Βάσει αυτής της κατηγοριοποίησης διακρίνονται σε δύο μεγάλες κατηγορίες:

- Δίκτυα πρόσθιας τροφοδότησης (feed-forward)
- Δίκτυα ανατροφοδότησης (recurrent)

Στα δίκτυα πρόσθιας τροφοδότησης τα σήματα μεταφέρονται προς μία μόνο κατεύθυνση, από επίπεδα εισόδου προς επίπεδα εξόδου. Αν δεν υπάρχει κάποιο ενδιάμεσο επίπεδο αναφερόμαστε σε ΝΔ πρόσθιας τροφοδότησης ενός επιπέδου (single layer feed-forward), ενώ στην περίπτωση ενός ή περισσότερων ενδιάμεσων επιπέδων αναφερόμαστε σε πολυεπίπεδα ΝΔ πρόσθιας τροφοδότησης (multilayer feed-forward). Στα ΝΔ πρόσθιας τροφοδότησης δεν επιτρέπονται αναδράσεις, δηλαδή η έξοδος ενός νευρώνα να αποτελεί την είσοδο σε νευρώνα του ίδιου ή προηγούμενου επιπέδου.

Αντίθετα, στην περίπτωση των ΝΔ ανατροφοδότησης επιτρέπονται αναδράσεις και συνεπώς τα σήματα μπορούν να μεταφέρονται και προς τις δύο κατευθύνσεις. Και σε αυτή την περίπτωση μπορούμε να έχουμε ένα ή περισσότερα ενδιάμεσα επίπεδα. Τα παραπάνω φαίνονται στο σχήμα 4.5



Σχήμα 4.5: (α)ΤΝΔ πρόσθιας τροφοδότησης και (β) ΤΝΔ ανατροφοδότησης

Υπάρχουν βέβαια και κατηγοριοποιήσεις σε σχέση με τον αριθμό των επιπέδων (στο σχήμα 4.5 υπάρχουν 3 επίπεδα). Επίσης τα ΤΝΔ διαχωρίζονται ανάλογα με τον αριθμό των νευρώνων που έχουν σε κάθε επίπεδο αλλά και του τρόπου που αυτά συνδέονται μεταξύ τους.

4.3.2.4 Αλγόριθμοι Εκπαίδευσης

Ένα ΤΝΔ για να μπορεί να εξάγει συμπεράσματα από τα δεδομένα που εισάγονται σε αυτό πρέπει πρώτα να περάσει από μία διαδικασία εκπαίδευσης. Στην διαδικασία αυτή, μεταβάλλονται οι συντελεστές βάρους του προκειμένου να βρεθεί η βέλτιστη τιμή τους, δηλαδή η τιμή που θα ανταποκρίνεται καλύτερα στις ανάγκες του προβλήματος. Για τον σκοπό αυτό έχουν προταθεί διάφοροι αλγόριθμοι εκπαίδευσης που χωρίζονται σε 3 κατηγορίες

- εκπαίδευση με επίβλεψη (supervised training)
- εκπαίδευση χωρίς επίβλεψη (unsupervised training)
- υβριδική εκπαίδευση (hybrid training).

Στην εκπαίδευση με επίβλεψη οι είσοδοι εφαρμόζονται στο ΤΝΔ ταυτόχρονα με την επιθυμητή έξοδο του δικτύου και τα βάρη μεταβάλλονται με τέτοιον τρόπο ώστε να ελαχιστοποιηθεί η διαφορά μεταξύ της εξόδου του ΤΝΔ και της επιθυμητής εξόδου. Ο πιο δημοφιλής αλγόριθμος αυτής της κατηγορίας είναι ο αλγόριθμος όπισθεν διάδοσης σφάλματος (back-propagation). Η ενισχυμένη εκπαίδευση (reinforcement training) είναι ειδική περίπτωση της εκπαίδευσης με επίβλεψη. Στην ενισχυμένη εκπαίδευση ή εκπαίδευση με ημιεπίβλεψη, αντί για τη σωστή έξοδο εξάγεται ένας χαρακτηρισμός σχετικός με την απόδοση του ΤΝΔ και τα βάρη μεταβάλλονται έτσι ώστε να ελαχιστοποιηθεί η πιθανότητα «κακού» χαρακτηρισμού του ΤΝΔ.

Στην εκπαίδευση χωρίς επίβλεψη η επιθυμητή έξοδος δεν είναι γνωστή. Εφαρμόζεται στο ΤΝΔ ένα σύνολο από εισόδους-χαρακτηριστικά και το ΤΝΔ αναζητά από μόνο του την «κρυμμένη πληροφορία» που υπάρχει στα δεδομένα για να πραγματοποιήσει την ταξινόμηση.

Στην υβριδική εκπαίδευση, ένα μέρος των βαρών ανανεώνεται με εκπαίδευση με επίβλεψη, ενώ ένα άλλο με εκπαίδευση χωρίς επίβλεψη. Συνήθως η εκπαίδευση χωρίς επίβλεψη εφαρμόζεται στα πρώτα επίπεδα

νευρώνων, έτσι ώστε τα δεδομένα να ομαδοποιηθούν ανάλογα με τη σχετική ομοιότητα τους, ενώ σε επόμενα επίπεδα νευρώνων εφαρμόζεται ο αλγόριθμος της Όπισθεν Διάδοσης Σφάλματος, έτσι ώστε οι ομάδες που προέκυψαν να συσχετισθούν με την επιθυμητή έξοδο.

4.3.2.4.1 Αλγόριθμος Όπισθεν Διάδοσης Σφάλματος

Όπως αναφέρθηκε παραπάνω, ο πιο δημοφιλής τρόπος εκπαίδευσης πολυεπίπεδων ΤΝΔ πρόσθιας ανατροφοδότησης είναι ο αλγόριθμος της όπισθεν διάδοσης σφάλματος, ο οποίος επινοήθηκε πρώτα από τον Werbos [7] το 1974 και βελτιώθηκε αργότερα από τον Parker [8] και από τους Rumelhart, Hinton and Williams [6]. Ο αλγόριθμος αυτός θα χρησιμοποιηθεί και στην παρούσα εργασία για την εκπαίδευση των ΤΝΔ που βρίσκονται στην μονάδα ταξινόμησης του συστήματος αναγνώρισης οργάνων που θα αναπτυχθεί στο κεφάλαιο 5. Σε ένα τέτοιο πολυεπίπεδο ΤΝΔ πρόσθιας τροφοδότησης εισάγονται τα p διανύσματα του συνόλου εκπαίδευσης μαζί με τις επιθυμητές τους εξόδους d_i , $i=1,2,\ldots,p$. 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. Η τιμή εξόδου του y_j του j νευρώνα για την εμφάνιση του προτύπου n, $n=1,2,\ldots,p$ δίνεται από τη σχέση

$$y_j(n) = f(u_j(n)) \qquad (4.10)$$

όπου f η συνάρτηση ενεργοποίησης του συγκεκριμένου νευρώνα και u_j το συναπτικό άθροισμα, δηλαδή

$$u_j(n) = \sum_{i=1}^{p} w_{ji}(n) x_i(n) + b_j \qquad (4.11)$$

αν ο νευρώνας ανήκει στο επίπεδο εισόδου και

$$u_j(n) = \sum_{i \in C} w_{ji}(n) y_i(n) + b_j \qquad (4.12)$$

όπου w_{ji} συνδέει τον j νευρώνα με τον i νευρώνα του προηγούμενου επιπέδου και y_i οι έξοδοι των νευρώνων του προηγούμενου επιπέδου.
Αφού το σήμα φτάσει στο επίπεδο εξόδου, μπορεί να υπολογιστεί τα σφάλμα στην έξοδο του νευρώνα j που ανήκει σε επίπεδο εξόδου και το οποίο δίνεται από τον τύπο:

$$e_i(n)=d_i(n) - y_i(n) \quad (4.13)$$

όπου $d_i(n)$ η επιθυμητή και $y_i(n)$ η πραγματική έξοδος του νευρώνα j.
Ορίζουμε την στιγμιαία τιμή του τετραγωνικού σφάλματος για το νευρώνα εξόδου j ως $\frac{1}{2} e_j^{\,2}(n)$. Αντίστοιχα, ορίζουμε το άθροισμα των τετραγωνικών σφαλμάτων όλων των νευρώνων εξόδου ως

$$J(n) = \frac{1}{2}\sum_{j=1}^{M} e_j^{\,2}(n) \qquad (4.14)$$

όπου M είναι ο αριθμός των νευρώνων εξόδου.
Επίσης ορίζουμε την μέση τιμή των σφαλμάτων ως

$$J_{av} = \frac{1}{p}\sum_{n=1}^{p} J(n) \qquad (4.15)$$

Τόσο το J όσο και το J_{av} είναι συναρτήσεις των ελεύθερων παραμέτρων του δικτύου (βάρη των συνδέσεων καν πολώσεις). Ο αντικειμενικός στόχος της διαδικασίας εκπαίδευσης είναι η ελαχιστοποίηση του J_{av} προσαρμόζοντας τις ελεύθερες παραμέτρους του δικτύου. Η ελαχιστοποίηση της συνάρτησης σφάλματος γίνεται με τη μέθοδο *απότομης καθόδου (gradient steepest descent)* [9]. Η εφαρμογή της μεθόδου έγκειται στη μεταβολή των βαρών με τέτοιο τρόπο, ώστε η μεταβολή των βαρών γίνεται ανάλογα με την κλίση που έχει η συνάρτηση σφάλματος.
Η ενημέρωση των συντελεστών w_{ji} και b_j γίνεται, σύμφωνα με τον αλγόριθμο, από τους παρακάτω τύπους

$$\Delta w_{ji}(n) = -\eta \frac{\partial J(n)}{\partial w_{ji}(n)} = \eta \delta_j(n) y_i(n) \qquad \Delta b_j(n) = \eta \delta_j(n)$$

(4.16)

όπου το η είναι ο ρυθμός εκπαίδευσης και $\delta_j(n)$ ονομάζεται τοπική κλίση για την οποία ισχύει ότι

$$\delta_j(n) = \begin{cases} e_j(n) f_j'(u_j(n)) & , \ \alpha \nu j \ \alpha \nu τιστοιχεί σε νευρώνα εξόδου \\ f_j'(u_j(n)) \sum_k \delta_k(n) w_{kj}(n) & , \ \alpha \nu j \ \alpha ντιστοιχεί σε νευρώνα ενδιάμεσου επιπέδου \end{cases}$$

(4.17)

Όπως διαφαίνεται η λειτουργία του αλγόριθμου έχει δύο στάδια. Το πρώτο αναφέρεται ως πέρασμα ορθής φοράς (forward pass), και το δεύτερο ως πέρασμα ανάστροφης φοράς (reverse pass). Στο ευθύ πέρασμα, τα βάρη των συνδέσεων παραμένουν αναλλοίωτα σε όλο το δίκτυο και τα σήματα υπολογίζονται σε κάθε νευρώνα υπολογίζοντας τις αντίστοιχες εξόδους. Το πέρασμα αυτό ξεκινά στο πρώτο κρυμμένο επίπεδο, με την παρουσίαση του διανύσματος εισόδου, και τελειώνει στο επίπεδο εξόδου με τον υπολογισμό του σφάλματος για κάθε νευρώνα αυτού του επιπέδου.

Το ανάστροφο πέρασμα, αντίθετα, αρχίζει από το επίπεδο εξόδου, περνώντας τα σήματα σφάλματος προς τα πίσω και υπολογίζοντας αναδρομικά το *δ* για κάθε νευρώνα. Η διαδικασία αυτή επιτρέπει στα βάρη των συνδέσεων να υποστούν αλλαγές σύμφωνα με τον κανόνα δέλτα. Για τους νευρώνες του επιπέδου εξόδου, το *δ* ισούται με το γινόμενο του σφάλματος επί την πρώτη παράγωγο της μη-γραμμικής συνάρτησης. Στη συνέχεια, χρησιμοποιούμε αυτό το *δ* για να υπολογίσουμε τα *δ* του προτελευταίου επιπέδου και ούτω καθεξής για τα υπόλοιπα. Δηλαδή το

σφάλμα εξόδου διαδίδεται από το επίπεδο εξόδου μέσω των ενδιάμεσων επιπέδων προς την είσοδο (όπισθεν διάδοση σφάλματος).

Σημειώνεται ότι η διαδικασία των δύο περασμάτων εκτελείται για κάθε πρότυπο του συνόλου εκπαίδευσης με μία κυκλική επανάληψη μέχρι να εισαχθούν όλα τα πρότυπα που ανήκουν στο σύνολο εκπαίδευσης.

Ο αλγόριθμος της όπισθεν διάδοσης σφάλματος μπορεί να εφαρμοστεί με δύο τρόπους σε ένα σύνολο εκπαίδευσης. Ο ένας τρόπος, γνωστός και ως on-line εκπαίδευση, ανανεώνει τα βάρη κάθε φορά που εφαρμόζεται στο ΝΔ ένα νέο διάνυσμα του συνόλου εκπαίδευσης. Ο άλλος τρόπος εκπαίδευσης, γνωστός και ως batch εκπαίδευση, ανανεώνει τα βάρη μετά την εφαρμογή όλων των διανυσμάτων του συνόλου εκπαίδευσης. Η περίοδος με την οποία ανανεώνονται τα βάρη ονομάζεται επανάληψη ή εποχή (epoch). Κάθε ένας από τους προαναφερθέντες τρόπους εκπαίδευσης έχει πλεονεκτήματα και μειονεκτήματα. Συγκεκριμένα, στην on-line εκπαίδευση η ανανέωση των βαρών ανά εφαρμοζόμενο διάνυσμα καθιστά την αναζήτηση στο χώρο των βαρών στοχαστική, που σημαίνει ότι ο αλγόριθμος είναι δύσκολο να εγκλωβιστεί σε κάποιο τοπικό ελάχιστο (local minimum). Η on-line εκπαίδευση προτιμάται στις hardware υλοποιήσεις ΝΔ, όπου υπάρχουν περιορισμοί στη διαθεσιμότητα πόρων αποθήκευσης. Στην batch εκπαίδευση το διάνυσμα της κλίσης (gradient) υπολογίζεται με μεγαλύτερη ακρίβεια και ο αλγόριθμος μπορεί να συγκλίνει σε κάποιο ελάχιστο, αλλά ύστερα από μεγαλύτερο χρόνο εκπαίδευσης.

4.3.2.4.2 Προσθήκη του όρου ορμής στον Αλγόριθμο Όπισθεν Διάδοσης Σφάλματος

Ο αλγόριθμος back-propagation δίνει μία προσέγγιση της τροχιάς των βαρών που υπολογίζεται με τη μέθοδο της απότομης καθόδου. Ο ρυθμός εκμάθησης καθορίζει το μέγεθος της μεταβολής των συντελεστών βάρους σε κάθε εποχή. Όσο μικρότερος είναι ο ρυθμός μάθησης, τόσο μικρότερη είναι η μεταβολή των βαρών σε κάθε επανάληψη και ομαλότερη η τροχιά σύγκλισης της καμπύλης των βαρών. Αυτή η «βελτίωση» όμως έχει ως κόστος τον αργό ρυθμό εκπαίδευσης. Από την άλλη μεριά αν χρησιμοποιήσουμε υψηλό ρυθμό μάθησης μπορεί μεν να επιτευχθεί επιτάχυνση της εκπαίδευσης αλλά οι μεγάλες μεταβολές σε κάθε επανάληψη προκαλούν κίνδυνο αστάθειας του αλγορίθμου εκπαίδευσης (ταλαντώσεις). Μία απλή μέθοδος αύξησης του ρυθμού, με αποφυγή των κινδύνων αστάθειας, είναι η τροποποίηση του κανόνα ενημέρωσης των βαρών με εισαγωγή ενός *όρου ορμής α* (momentum term), όπως φαίνεται στην παρακάτω εξίσωση

$$\Delta w_{ji}(n)=\alpha\,\Delta w_{ji}(n-1)+\eta\delta_j(n)y_i(n) \qquad (4.18)$$

όπου $0<\alpha<1$.

Η τιμή της παραμέτρου α δεν είναι καθορισμένη και προκύπτει μετά από εμπειρική μελέτη ανάλογα με το είδος του προβλήματος. Προκειμένου να δούμε την επίπτωση της σταθεράς α στην παρουσίαση των προτύπων, ανα-διατυπώνουμε την προηγούμενη εξίσωση ως μία χρονική ακολουθία με δείκτη t. Ό δείκτης t έχει αρχική τιμή 0 και φθάνει μέχρι την τρέχουσα χρονική στιγμή

n. Αυτό που προκύπτει είναι μία διαφορική εξίσωση πρώτης τάξης. Λύνοντας ως προς $\Delta w_{ji}(n)$ προκύπτει

$$\Delta w_{ji}(n) = \eta \sum_{t=0}^{n} a^{n-t} \delta_j(t) y_i(t)$$

ή

$$\Delta w_{ji}(n) = \eta \sum_{t=0}^{n} a^{n-t} \frac{\partial J(n)}{\partial w_{ji}(n)} \qquad (4.19)$$

όπου δείχνει την επίπτωση της ορμής α στην μεταβολή των βαρών.
Βασισμένοι σε αυτή τη σχέση μπορούμε να κάνουμε τις εξής παρατηρήσεις:
Όταν η μερική παράγωγος $\frac{\partial J(n)}{\partial w_{ji}(n)}$ έχει το ίδιο πρόσημο σε συνεχόμενες επαναλήψεις, τότε το $\Delta w_{ji}(n)$ μεγαλώνει σημαντικά, και το βάρος $w_{ji}(n)$ μεταβάλλεται σε μεγάλο βαθμό. Τότε η εισαγωγή του παράγοντα ορμής στον αλγόριθμο back-propagation τείνει να επιταχύνει σε σταθερή κατεύθυνση την κάθοδο. Αντίθετα όταν η μερική παράγωγος $\frac{\partial J(n)}{\partial w_{ji}(n)}$ έχει αντίθετο πρόσημο σε συνεχόμενες επαναλήψεις, τότε το $\Delta w_{ji}(n)$ μειώνεται σημαντικά, και το βάρος $w_{ji}(n)$ μεταβάλλεται σε μικρό βαθμό. Επομένως η εισαγωγή του παράγοντα momentum στον αλγόριθμο back-propagation έχει σταθεροποιητικό αποτέλεσμα.

Έτσι, προκειμένου να αυξηθεί η ταχύτητα σύγκλισης έχει προταθεί η χρήση μεταβλητού ρυθμού εκμάθησης [10][11]. Πιο συγκεκριμένα, η εκπαίδευση ξεκινά με μια «ασφαλή» τιμή ρυθμού εκμάθησης (π.χ. μικρή) και μεταβάλλεται ανάλογα με τη συμπεριφορά του ΝΔ. Η γενική ιδέα [10] μιας από τις πιο δημοφιλείς μεθόδους μεταβολής του ρυθμού εκμάθησης είναι ότι αν κατά την εκπαίδευση η μήτρα βαρών οδηγεί σε μικρότερο σφάλμα, τότε ο ρυθμός εκμάθησης αυξάνεται. Στην περίπτωση που οδηγεί σε μεγαλύτερο σφάλμα, ο ρυθμός εκμάθησης μειώνεται. Τέλος, στην περίπτωση που η μήτρα βαρών δεν οδηγεί σε μεταβολή της τιμής σφάλματος, οι συντελεστές βαρύτητας και πόλωσης διατηρούνται, η ορμή γίνεται 0 και το βήμα επαναλαμβάνεται.

4.3.3 Παράμετροι βελτιστοποίησης ΤΝΔ

4.3.3.1 Αρχικοποίηση Βαρών

Έχει αποδειχτεί ότι η ταχύτητα σύγκλισης ενός ΤΝΔ εξαρτάται σε μεγάλο βαθμό από τα αρχικά βάρη. Για το λόγο αυτό πραγματοποιείται αρχικοποίηση της μήτρας βαρών είτε με τυχαία αρχικά βάρη, ή χρησιμοποιώντας ως αρχικά βάρη τις προσεγγιστικές λύσεις που προκύπτουν από άλλες τεχνικές μοντελοποίησης όπως είναι η μέθοδος πρωτευουσών συνιστωσών [12], ο ταξινομητής κοντινότερου γείτονα [13] κ.λ.π. με σκοπό τόσο τη μείωση του απαιτούμενου χρόνου εκπαίδευσης, όσο και τη μείωση της πιθανότητας σύγκλισης σε κάποιο τοπικό ελάχιστο (μη τυχαία αρχικοποίηση).

Στην τυχαία αρχικοποίηση τα αρχικά βάρη παίρνουν μικρές τυχαίες τιμές. Οι τιμές αυτές χρειάζεται να είναι τυχαίες προκειμένου να μην υπάρχουν συμμετρίες, δηλαδή κάθε νευρώνας να επεξεργάζεται διαφορετικές συναρτήσεις. Στην αντίθετη περίπτωση η απόκριση όλων των νευρώνων του ίδιου επίπεδου θα ήταν παρόμοια, το συγκεκριμένο επίπεδο θα συμπεριφερόταν σαν να αποτελείται ουσιαστικά από ένα νευρώνα, με αποτέλεσμα η πληροφορία για το σφάλμα να είναι παρόμοια και συνεπώς η μεταβολή των βαρών κατά τη διάρκεια της εκπαίδευσης θα ήταν δύσκολη. Η επιλογή «μικρών» τιμών αρχικοποίησης των βαρών είναι απαραίτητη προκειμένου να αποφεύγεται η μετάβαση των σιγμοειδών συναρτήσεων σε κόρο. Μεγάλα βάρη μπορούν να ενισχύσουν μια μεσαίου μεγέθους είσοδο παράγοντας πολύ μεγάλες τιμές για τα αθροίσματα της σχέσης (4.8) που σημαίνει πολύ μεγάλη είσοδο για τους νευρώνες του επόμενου επιπέδου. Αυτό θα έχει ως συνέπεια οι νευρώνες να κινηθούν σε επίπεδες επιφάνειες κοντά στο σημείο εκκίνησης ή να εγκλωβιστούν σε κάποιο τοπικό ελάχιστο. Από την άλλη, οι τιμές των βαρών δεν επιτρέπεται να είναι πάρα πολύ μικρές, γιατί αυτό σημαίνει πολύ μικρή τιμή για το σήμα σφάλματος επιδρώντας στην ταχύτητα εκπαίδευσης. Για την τυχαία αρχικοποίηση των βαρών έχουν αναπτυχθεί διάφορες μεθοδολογίες οι περισσότερες από τις οποίες αναφέρονται στο εύρος των αρχικών τιμών [14].

Στη μη τυχαία αρχικοποίηση, το ΝΔ ξεκινά από μια σχετικά καλή λύση και ο αλγόριθμος της όπισθεν διάδοσης σφάλματος χρησιμοποιείται στη συνέχεια για την εύρεση της βέλτιστης λύσης. Με τον τρόπο αυτό μειώνεται ο απαιτούμενος χρόνος εκπαίδευσης προκειμένου το σύστημα να συγκλίνει στο πραγματικό ολικό ελάχιστο και όχι σε κάποιο τοπικό.

4.3.3.2 Τερματισμός Εκπαίδευσης

Ο καθορισμός της χρονικής στιγμής τερματισμού της εκπαίδευσης ενός ΤΝΔ είναι πολύ σημαντικό ζήτημα. Αν η εκπαίδευση σταματήσει πολύ νωρίς τότε ενδεχομένως το ΤΝΔ να μην έχει εκπαιδευτεί επαρκώς (δεν έχει μάθει σωστά) και δεν μπορεί να ανταποκριθεί στις ανάγκες του προβλήματος κάνοντας λάθος εκτιμήσεις για τα δεδομένα. Αντίθετα αν το ΤΝΔ αφεθεί να εκπαιδευτεί παραπάνω από όσο χρειάζεται τότε κινδυνεύει να υπερεκπαιδευτεί και να χάσει την ικανότητα γενίκευσης. Σε αυτή την περίπτωση το ΤΝΔ «αποστηθίζει» τα δεδομένα εκπαίδευσης σε τέτοιο βαθμό που δεν μπορεί να εκτιμήσει σωστά δεδομένα που απέχουν λίγο από αυτά. Είναι ζητούμενο λοιπόν η εύρεση του βέλτιστου αριθμού εποχών εκπαίδευσης ώστε να μεγιστοποιηθεί η απόδοση του δικτύου.

Συνήθη κριτήρια τερματισμού της εκπαίδευσης είναι τα παρακάτω [4]:

- Η τιμή της συνάρτησης σφάλματος να είναι μικρότερη από μία προκαθορισμένη τιμή.
- Να έχει εκτελεστεί προκαθορισμένος αριθμός επαναλήψεων
- Η κλίση της συνάρτησης σφάλματος να είναι μικρότερη από μια προκαθορισμένη τιμή.

Τα παραπάνω κριτήρια έχουν το μειονέκτημα ότι από τη μία εξαρτώνται αποκλειστικά από τις παραμέτρους του ΝΔ και από την άλλη δε λαμβάνουν υπόψη τον παράγοντα γενίκευσης του ΝΔ, δηλαδή την ικανότητα του να

ανταποκρίνεται αξιόπιστα σε νέα δεδομένα που δεν έχουν χρησιμοποιηθεί κατά τη διάρκεια της εκπαίδευσης.
Για βελτίωση της γενίκευσης έχουν προταθεί διάφοροι μέθοδοι τερματισμού της εκπαίδευσης.

Μία από αυτές είναι η μέθοδος του *έγκαιρου τερματισμού (early stopping)* [15]. Για την υλοποίηση της απαιτούνται δύο φάσεις: η *φάση εκπαίδευσης (training phase)* και η *φάση γενίκευσης (testing phase),* Στη φάση εκπαίδευσης χρησιμοποιούνται δύο σύνολα δεδομένων, το *σύνολο εκπαίδευσης (training set)* και το *σύνολο επαλήθευσης (validation set),* ενώ στη φάση γενίκευσης χρησιμοποιείται το *σύνολο γενίκευσης (testing set).* Τα ΤΝΔ εκπαιδεύεται με τα δεδομένα του συνόλου εκπαίδευσης και η διαδικασία της εκπαίδευσης ολοκληρώνεται όταν μεγιστοποιηθεί η απόδοση του ΤΝΔ στο σύνολο επαλήθευσης, οπότε οι βέλτιστες τιμές για τις μήτρες των βαρών και των πολώσεων αποθηκεύονται για να χρησιμοποιηθούν στη φάση γενίκευσης. Στη φάση γενίκευσης οι είσοδοι στο ΝΔ είναι τα δεδομένα του συνόλου γενίκευσης, ενώ η προκύπτουσα έξοδος δίνει μια εκτίμηση για την ικανότητα του ΝΔ να ανταποκρίνεται σε νέα άγνωστα δεδομένα.

Μια άλλη μέθοδος τερματισμού είναι η μέθοδος της *διεπικύρωσης (cross-validation)* [16], γενική περίπτωση της οποίας είναι η *k* διεπικύρωση. Η μέθοδος της *k* διεπικύρωσης, όπως και η μέθοδος του έγκαιρου τερματισμού, πραγματοποιείται σε δύο φάσεις. Αρχικά τα δεδομένα της φάσης εκπαίδευσης διαιρούνται σε k υποσύνολα ίδιας διάστασης. Το ΝΔ εκπαιδεύεται *k* φορές χρησιμοποιώντας ως σύνολο εκπαίδευσης τα k-1 υποσύνολα δεδομένων, ενώ το k υποσύνολο δεδομένων, διαφορετικό κάθε φορά, χρησιμοποιείται ως σύνολο επαλήθευσης. Η ικανότητα γενίκευσης του ΝΔ ελέγχεται στα δεδομένα του συνόλου επαλήθευσης και προκύπτει ως ο μέσος όρος της απόδοσης του ΤΝΔ μετά το πέρας των k εκπαιδεύσεων. Το κύριο μειονέκτημα αυτής της μεθόδου είναι το υψηλό υπολογιστικό κόστος, μιας και κάθε ΝΔ πρέπει να εκπαιδευτεί k φορές.

4.3.3.3 Επιλογή Συνόλου Εκπαίδευσης, Επαλήθευσης και Γενίκευσης

Η απόδοση και η ικανότητα γενίκευσης ενός ΝΔ εξαρτώνται άμεσα από την ποσότητα και την ποιότητα των δεδομένων που θα χρησιμοποιηθούν για την εκπαίδευση του. Το σύνολο εκπαίδευσης θα πρέπει να είναι αντιπροσωπευτικό των διαφορετικών προτύπων που χαρακτηρίζουν το προς επίλυση πρόβλημα. Η απόκριση ενός εκπαιδευμένου ΝΔ είναι πάρα πολύ καλή σε δεδομένα παραπλήσια αυτών που χρησιμοποιήθηκαν κατά τη διάρκεια εκπαίδευσης του *(παρεμβολή-interpolation),* ενώ δεν ισχύει το ίδιο σε περιπτώσεις ακραίες σε σχέση με τα δεδομένα εκπαίδευσης *(υπερβολή-extrapolation).* Αυτό καθιστά προφανή την απαίτηση για δεδομένα εκπαίδευσης που καλύπτουν όσο το δυνατό μεγαλύτερο εύρος πιθανών διαφορετικών τιμών, έτσι ώστε να μειώνεται η πιθανότητα παρουσίασης ακραίων περιπτώσεων. Επίσης, τα διάφορα πρότυπα θα πρέπει να εκπροσωπούνται με παρόμοια ποσοστά στα σύνολα εκπαίδευσης, επαλήθευσης και γενίκευσης.

4.3.3.4 Σχεδίαση Νευρωνικού Δικτύου

Ο αριθμός των επιπέδων και το πλήθος νευρώνων από τους οποίους αποτελείται το καθένα καθορίζουν την αρχιτεκτονική και είναι πολύ σημαντικοί παράμετροι για την λειτουργία ενός ΤΝΔ.

Το πλήθος των επιπέδων, και δη των ενδιάμεσων επιπέδων αφού το επίπεδο εισόδου και εξόδου είναι υποχρεωτικά, καθορίζει την πολυπλοκότητα του προβλήματος που μπορεί να λύσει το ΤΝΔ. Ένα ενδιάμεσο επίπεδο δημιουργεί μια υπερεπιφάνεια, ενώ δύο ενδιάμεσα επίπεδα συνδυάζουν τις υπερεπιφάνειες για τη δημιουργία κυρτών περιοχών απόφασης. Στην πράξη δε χρησιμοποιούνται σχεδόν ποτέ περισσότερα από δύο ενδιάμεσα επίπεδα, αφού αυτά είναι αρκετά για τη δημιουργία περιοχών ταξινόμησης οποιασδήποτε μορφής. Ωστόσο, έχει δειχτεί ότι ακόμα και με ένα ενδιάμεσο επίπεδο το ΤΝΔ είναι ικανό να αναπαραστήσει οποιαδήποτε συνεχή συνάρτηση πολλών μεταβλητών [17].

Μετά το πλήθος των επιπέδων χρειάζεται να καθοριστεί το πλήθος των νευρώνων που θα έχει κάθε επίπεδο. Για το επίπεδο εισόδου ο αριθμός των νευρώνων εξαρτάται από το πλήθος των χαρακτηριστικών που χρησιμοποιούνται για το προς επίλυση πρόβλημα. Για το επίπεδο εξόδου οι νευρώνες καθορίζονται πάλι από την φύση του προβλήματος ανάλογα με τι εξόδους χρειάζεται να δώσει. Αυτό το οποίο χρειάζεται ιδιαίτερη προσοχή στη σχεδίαση του ΤΝΔ είναι ο αριθμός των νευρώνων των ενδιάμεσων επιπέδων αφού αυτός συνδέεται με την απόδοση του ΤΝΔ. Συγκεκριμένα, μικρός αριθμός νευρώνων στα ενδιάμεσα επίπεδα μπορεί να αποτύχει να λύσει το πρόβλημα, ενώ μεγάλος αριθμός έχει χρονοβόρα εκπαίδευση και μπορεί να υποπέσει σε υπερεκπαίδευση.

Γενικά δεν υπάρχει μέθοδος που να βρίσκει τον βέλτιστο αριθμό νευρώνων στο ενδιάμεσο επίπεδο αλλά μόνο εμπειρικές μελέτες που οι περισσότερες σχετίζονται άμεσα με την φύση του προβλήματος, το μέγεθος του συνόλου εκπαίδευσης, την ποιότητα των δεδομένων κλπ. Το εύρος τιμών που πρέπει να έχουν οι νευρώνες στο ενδιάμεσο επίπεδο έχει τεθεί [17] να έχει κατώτατο όριο το 2 και ως ανώτατο 2N+1*,* όπου *N* ο αριθμός δεδομένων του διανύσματος εισόδου. Η εμπειρική μελέτη δείχνει ότι ο πιο πιθανός αριθμός νευρώνων είναι $\sqrt{MN}$, όπου M ο αριθμός των νευρώνων του επιπέδου εξόδου [15], ή το 75% του αριθμού των νευρώνων του επιπέδου εισόδου [15]. Πάντως για την ολοκληρωμένη μελέτη ενός ΤΝΔ εκτελούνται συνεχείς εκπαιδεύσεις του ΝΔ για διαφορετικό αριθμό ενδιάμεσων νευρώνων και επιλέγεται εκείνος που ικανοποιεί με καλύτερο τρόπο το κριτήριο τερματισμού της εκπαίδευσης, ανεξάρτητα *από* τους παραπάνω κανόνες οι οποίοι απλά δίνουν μια τάξη μεγέθους για το ζήτημα. Έτσι, υπάρχουν μέθοδοι που ξεκινούν την εκπαίδευση με μικρό αριθμό νευρώνων και συνεχίζουν προσθέτοντας επιπλέον νευρώνες (μέθοδοι ανάπτυξης - constructive methods), ανάλογα με τη συμπεριφορά του κριτηρίου τερματισμού, ή αντίστροφα ξεκινούν από μεγάλο αριθμό νευρώνων και χρησιμοποιώντας *μεθοδολογίες περιορισμού νευρώνων (pruning methods)* βρίσκουν το βέλτιστο αριθμό τους.

4.3.3.5 Προεπεξεργασία Δεδομένων

Τα διανύσματα εισόδου πριν εφαρμοστούν στους νευρώνες του επιπέδου εισόδου πρέπει να υποστούν κάποιο είδος προεπεξεργασίας. Η προεπεξεργασία έχει ως στόχο την αναγωγή των δεδομένων σε τέτοιο εύρος τιμών έτσι ώστε αυτά να μην βρίσκονται σε περιοχή των σιγμοειδών συναρτήσεων που παρουσιάζουν κόρο. Παράλληλα τα δεδομένα εισόδου κανονικοποιούνται, έτσι ώστε η μέση τιμή τους να είναι ίση με μηδέν και η τυπική απόκλιση ίση με ένα. Με αυτόν τον τρόπο όλοι οι συντελεστές βάρους «μαθαίνουν» με την ίδια περίπου ταχύτητα. Εναλλακτικά, τα δεδομένα μπορούν να κανονικοποιηθούν από 0 έως 1 ή -1 έως 1 ανάλογα με το αν η χρησιμοποιούμενη συνάρτηση μετασχηματισμού είναι η σιγμοειδής εφαπτομενική ή η υπερβολική. Τέλος, για την επιτάχυνση του αλγόριθμου της όπισθεν διάδοσης σφάλματος θα πρέπει οι μεταβλητές του διανύσματος εισόδου να είναι ασυσχέτιστες, κάτι που μπορεί να επιτευχθεί με διάφορες μεθοδολογίες μείωσης της διάστασης του διανύσματος εισόδου σε ένα ΝΔ. Ιδανικά, θα έπρεπε να εξεταστούν όλοι οι 2^N διαφορετικοί συνδυασμοί των N μεταβλητών του διανύσματος εισόδου και να επιλεγούν εκείνες οι μεταβλητές που ικανοποιούν με βέλτιστο τρόπο κάποια συνάρτηση ποιότητας που μπορεί να είναι είτε το κριτήριο τερματισμού του ΝΔ ή κάποιο μέτρο απόστασης, πληροφορίας, εξάρτησης κ.λ.π. Επειδή, η συγκεκριμένη μέθοδος έχει μεγάλο υπολογιστικό κόστος ακόμη και για σχετικά μικρό πλήθος μεταβλητών συνήθως χρησιμοποιούνται τεχνικές, οι οποίες βασίζονται σε ευρετικές μεθόδους επιλογής, οι οποίες επιχειρούν να εξισορροπήσουν την υπολογιστική πολυπλοκότητα με την ικανοποίηση της συνάρτησης ποιότητας.

4.3.4 Εφαρμογές Τεχνητών Νευρωνικών Δικτύων

Τα ΤΝΔ έχουν ευρεία εφαρμογή σε ένα μεγάλο φάσμα περιοχών και τα αποτελέσματα που δίνουν τα καθιστούν ως ένα αρκετά υποσχόμενο εργαλείο στον τομέα ταξινόμησης προτύπων αλλά και σε πολύπλοκα προβλήματα στατιστικής υφής. Τα πιο δημοφιλή ΤΝΔ, με βάση το πλήθος των εφαρμογών τους, είναι τα πολυεπίπεδα ΤΝΔ πρόσθιας τροφοδότησης εκπαιδευμένα με τον αλγόριθμο της όπισθεν διάδοσης σφάλματος.

Εφαρμογές τους έχουν ήδη αναπτυχθεί στην αναγνώριση γραφής, αναγνώριση εικόνων, αναγνώριση φωνής, μετατροπή κειμένου σε φωνή. Επίσης έχουν εφαρμοστεί επιτυχώς για τη βελτιστοποίηση διαδικασιών ελέγχου, που αφορούν είτε στην παραγωγή, για συμπίεση εικόνων και για αναζήτηση σε μεγάλες βάσεις δεδομένων με σκοπό την εξόρυξη χρήσιμης πληροφορίας. Τα τελευταία χρόνια τα ΝΔ χρησιμοποιούνται και ως εργαλεία βραχυπρόθεσμων και μακροπρόθεσμων προβλέψεων σε χρηματιστηριακές εφαρμογές μετεωρολογικά φαινόμενα κ.λ.π.[18]. Επίσης έχουν χρησιμοποιηθεί για την επίλυση κλασσικών αλγορίθμων όπως αυτό του περιπλανώμενου πωλητή (travelling salesman's problem).

Κεφάλαιο 5 εφαρμογή- Σύστημα Αναγνώρισης Μουσικών Παραδοσιακών Οργάνων «ΣΑΜΠΟ»

5.1.1 Εισαγωγή-σκοπός

Στα προηγούμενα κεφάλαια εκτενής περιγραφή συστημάτων αναγνώρισης ήχου. Αυτά συνίστανται σε δύο βασικά κομμάτια, το κομμάτι της εξαγωγής των χαρακτηριστικών και αυτό της ταξινόμησης.

Έγινε σαφές από τα προηγούμενα ότι υπάρχουν παραπάνω από ένας τρόπος προσέγγισης για την επίλυση ενός προβλήματος αναγνώρισης μουσικής πηγής. Ανάλογα με τη φύση του προβλήματος κρίνεται σκόπιμη η εξαγωγή συγκεκριμένων κάθε φορά χαρακτηριστικών, και αντίστοιχα η χρήση ενός εξίσου συγκεκριμένου ταξινομητή στην ουρά του συστήματος. Στο τέλος, η αξιολόγηση του συστήματος κρίνεται πάντα με βάση το βαθμό δυσκολίας που παρουσιάζει το εκάστοτε εγχείρημα αναγνώρισης, σε συνδυασμό με τις παραδοχές που έγιναν κατά τη διάρκειά του.

Κατά την παρούσα διατριβή έγινε σχεδίαση και υλοποίηση ενός ταξινομητή προκειμένου για την **αναγνώριση μουσικών παραδοσιακών οργάνων.** Το Σύστημα Αναγνώρισης Μουσικών Παραδοσιακών Οργάνων που υλοποιήθηκε στα πλαίσια της παρούσας διατριβής θα ονομάζεται στη συνέχεια του κεφαλαίου **ΣΑΜΠΟ** για λόγους συντομίας.

Η σχεδίαση του συστήματος αναγνώρισης απαιτούσε τη λήψη αποφάσεων σε αρκετά επίπεδα. Συγκεκριμένα:

- Επιλογή οργάνων προς αναγνώριση

Τα όργανα που επιλέχθηκαν προς αναγνώριση είναι ελληνικά παραδοσιακά όργανα. Οι γνώσεις που έχουμε για τα τεχνικά και ακουστικά χαρακτηριστικά για τα συγκεκριμένα όργανα είναι πολύ περιορισμένες. Επιπλέον δεν έχει προηγηθεί κάποια απόπειρα αναγνώρισης τους.

- Δείγματα προς εκπαίδευση και αναγνώριση

Οι μελέτες που έχουν πραγματοποιηθεί μέχρι τώρα δεν χρησιμοποιούν ηχητικά δείγματα από εκτέλεση μουσικών οργάνων από μια κοινή βάση δεδομένων, καθώς μια τέτοια βάση σε παγκόσμιο επίπεδο δεν υφίσταται (όπως για παράδειγμα η βάση TIMIT για την αναγνώριση φωνής). Έτσι έγκειται στην ευχέρεια του σχεδιαστή να χρησιμοποιήσει για την εκπαίδευση και την αξιολόγηση του συστήματος του είτε ηχογραφήσεις μέσα σε ανηχοϊκούς θαλάμους, είτε ηχογραφήσεις από τη μουσική βιομηχανία, είτε ακόμα και ερασιτεχνικές ηχογραφήσεις. Πέρα από την ποιότητα της ηχογράφησης, επιπλέον, από μελέτη σε μελέτη, παρατηρούμε διαφοροποιήσεις και ως προς το περιεχόμενο των δειγμάτων. Έτσι, παρατηρούμε στο χώρο μελέτες με δείγματα από μεμονωμένους ήχους (πολλές φορές από μία μόνο νότα), από ακολουθίες νοτών, ή από ολόκληρες μουσικές φράσεις.

1. Στην παρούσα εφαρμογή κρίθηκε σκόπιμο το σύνολο των δειγμάτων να αντικατοπτρίζει όσο γίνεται περισσότερο τις πραγματικές συνθήκες αναγνώρισης. Στην καθημερινή ζωή καλούμαστε να μπορούμε να αναγνωρίζουμε μια μουσική πηγή ανεξαρτήτως της

ακουστότητας του χώρου που διαδραματίζεται η ακουστική σκηνή, ή της παρουσίας θορύβου πάσης φύσης. Για το λόγο αυτό τα ηχητικά δείγματα του ΣΑΜΠΟ έχουν αποσπασθεί από ηχογραφήσεις που κυκλοφορούν στο εμπόριο σε μορφή CD,καθώς και από ερασιτεχνικές ηχογραφήσεις και ηχογραφήσεις που έλαβαν χώρα ειδικά για την κατασκευή του συγκεκριμένου συστήματος .

2. Το περιεχόμενο των δειγμάτων μεμονωμένων νοτών αποτελεί βάση για εξαγωγή χαρακτηριστικών χρονικού (temporal) χαρακτήρα, όπως είναι η ατάκα (onset) ,η σταθερή κατάσταση (steady state) κ.α (κεφάλαιο 3.5). Η πληροφορία αυτή είναι εξαιρετικά χρήσιμη κατά την μηχανική αναγνώριση και επιπλέον μπορεί να την επιταχύνει χρονικά , καθώς απαιτεί ελάχιστο χρόνο επεξεργασίας δείγματος. Όμως , η χρήση χρονικών χαρακτηριστικών που αφορούν σε νότες, στην περίπτωση μουσικών φράσεων απαιτεί προεπεξεργασία κατάτμησης (segmentation) της φράσης αυτής στις νότες (pitch) που τη συνιστούν, διαδικασία απαιτητική σε χρόνο και αρκετά αναξιόπιστη. Έτσι, η εκτενής μελέτη τέτοιων χαρακτηριστικών κρίνεται, κατά την δική μου άποψη, δευτερεύουσας σημασίας, καθώς μπορεί να οδηγήσει σε εσφαλμένα συμπεράσματα. Επιπλέον, σε μια προσπάθεια αναζήτησης χαρακτηριστικών με «μουσικό» χαρακτήρα και φυσική έννοια, κατά τον ανθρώπινο τρόπο αντίληψης, τα δείγματα από μεμονωμένες νότες θεωρήθηκαν εσφαλμένος τρόπος προσέγγισης κατά την σχεδίαση του συστήματος. Και αυτό γιατί ο άνθρωπος στηρίζεται περισσότερο σε πληροφορία που στηρίζεται στο περιεχόμενο και σε μια μακροσκοπική αίσθηση σχετική με το φάσμα παρά σε ηχητικά δρώμενα που εκτυλίσσονται σε λίγα msec.

Οι λόγοι αυτοί οδήγησαν στην επιλογή δειγμάτων από κανονικές μουσικές φράσεις. Το γεγονός αυτό ανέβασε κατά πολύ το δείκτη δυσκολίας της διαδικασίας της αναγνώρισης και περιόρισε τα ποσοστά επιτυχίας του ΣΑΜΠΟ, αλλά κρίθηκε αναγκαίο, στα πλαίσια κατασκευής ενός συστήματος που θα αποτελέσει γόνιμη βάση και για μελλοντική έρευνα.

- <u>Επιλογή χαρακτηριστικών</u>

Τα κριτήρια για την επιλογή των χαρακτηριστικών προς ταξινόμηση στηρίχθηκαν στο ίδιο πλαίσιο αντιλήψεων που έκριναν και την επιλογή της φύσης των ηχητικών δειγμάτων. Προτιμήθηκαν χαρακτηριστικά με «μουσικό» χαρακτήρα που συνάδουν με την διαδικασία ανθρώπινης αντίληψης. Από ένα πλήθος διαθέσιμων χαρακτηριστικών που υλοποιήθηκαν, τελικά επιλέχθηκαν μόνο αυτά που είχαν φυσικό νόημα. Βαρύτητα δόθηκε στην επιλογή ενός ακουστικού μοντέλου για την ανθρώπινη ακοή .

- <u>Επιλογή τάξεων</u>

Κατά τη σχεδίαση του ΣΑΜΠΟ δεν δόθηκε βαρύτητα στην αριστοποίηση, σε χρόνο και σε μνήμη, ενός εκτελέσιμου προγράμματος. Επιπλέον δεν υπήρξε επιδίωξη για real-time λειτουργία. Αυτό έδωσε περιθώριο για περισσότερες δοκιμές, στα πλαίσια ενός framework υπολογιστικής προσομοίωσης πειραματικού χαρακτήρα, με τελικό σκοπό την αριστοποίηση ενός συστήματος απόφασης για τις πλέον δύσκολες συνθήκες ταξινόμησης. Και αυτό γιατί στο ΣΑΜΠΟ, διενεργείται ταξινόμηση δειγμάτων από εκτέλεση τεσσάρων μουσικών οργάνων που ανήκουν στην ίδια κατηγορία, των οποίων

η αναγνώριση από τον άνθρωπο παρουσιάζει αξιοσημείωτες δυσκολίες. Τα όργανα αυτά είναι έγχορδα παραδοσιακά ,τα τρία νυκτά και το ένα με δοξάρι και συγκεκριμένα : κρητική λύρα, μπουζούκι, ούτι και λαούτο.

- Επιλογή ταξινομητή

Κατά τον ίδιο τρόπο και ο ταξινομητής θα έπρεπε να εναρμονίζεται με το σκεπτικό της εφαρμογής, δηλαδή την ανθρώπινη αντίληψη, την ικανότητα γενίκευσης, την μη γραμμικότητα αλλά και την ανοχή σε λάθη επιμέρους στοιχείων του ταξινομητή.

5.1.2 Ελληνικά Λαϊκά Παραδοσιακά Όργανα

5.1.2.1 Λαούτο

Το λαούτο πανελλήνια είναι γνωστό ως λαούτο, λαβούτο ή λαγούτο (απ' το αραβικό al oud = ξύλο ή κατά τον Sachs , ευλύγιστο ραβδί). Παλιότερα λεγόταν και τυφλοσούρτης, γιατί, με το να κρατάει το ρυθμό , βοηθούσε το μελωδικό όργανο (βιολί ή κλαρίνο ή λύρα) να παίζει σωστά ρυθμικά. Το λαούτο έχει μεγάλο αχλαδόσχημο (επιπεδόκυρτο) ηχείο και μακρύ χέρι «σπασμένο» προς τα πίσω. Στο επάνω μέρος έχει κλειδιά απ' τα πλάγια, τέσσερις διπλές χορδές στερεωμένες στον καβαλάρη, πάνω στο καπάκι, και παίζεται με πένα.

Στα τέλη του 19^{ου} αι. το λαούτο κατασκευαζόταν σε τρία μεγέθη. Σήμερα έχει επικρατήσει το μεσαίο μέγεθος. Κι αυτό όμως διαφέρει στις διαστάσεις του από κατασκευαστή σε κατασκευαστή, αν και οι διαφορές είναι μικρές και χωρίς σημασία για την λειτουργία του οργάνου.

Για την κατασκευή του λαούτου ακολουθείται πάντα η ίδια σειρά ως προς τα τμήματα που το απαρτίζουν (*σκάφη, χέρι, καπάκι*), από την οποία εξαρτάται και ουσιαστικά η σταθερότητα, η καλή λειτουργία και η ηχητική απόδοση του οργάνου. Για την κατασκευή του χρησιμοποιούνται σκληρά ξύλα για την σκάφη: έβενο, παλισάνδρη, σφεντάμι, μαόνι, καρυδιά κ.α., φλαμούρι ή άλλο μαλακό ξύλο για το σκελετό του χεριού, και λευκή ξυλεία – συνήθως πεύκο – για το καπάκι. Οι χορδές του λαούτου ήταν παλιότερα από έντερο. Σταδιακά, μαζί με τις εντερικές χορδές άρχισαν να χρησιμοποιούνται και μεταλλικές – τα λεγόμενα *τσέλια* - , στην αρχή μόνο για τις υψηλές χορδές. Σήμερα χρησιμοποιούνται αποκλειστικά μεταλλικές χορδές, τέλια και *χρυσές* (σύρμα με μεταλλική περιτύλιξη).

Η πένα του λαούτου κατασκευάζεται από φτερό αρπακτικού πουλιού (συνήθως γύπα ή αετού), αλλά σε ορισμένες περιπτώσεις και από πλαστικό.

Ο ήχος του λαούτου παράγεται από το τσίμπημα των χορδών από την πένα. Το λαούτο κουρδίζεται πάντα κατά πέμπτες. Από τα τέσσερα ζεύγη των χορδών του, το πρώτο κουρδίζεται ουνίσονο και τα άλλα σε διαστήματα οκτάβας.

Η ένταση του ήχου εξαρτάται από το παίξιμο του λαουτιέρη. Το μαλακό παίξιμο δίνει ήχο γλυκό, όχι όμως πολύ μεγάλο, σε δύο κυρίως χρωματισμούς (*πιάνο* – σιγά – και *φόρτε* – δυνατά -). Με σκληρό παίξιμο έχουμε δυνατότερο ήχο, μόνο *φόρτε*. Οι παραπάνω χρωματισμοί δεν πρέπει να νοηθούν με την έννοια της δυναμικής της κλασσικής μουσικής, της προοδευτικής δηλαδή έντασης του ήχου από το πιάνο στο φόρτε ή το αντίθετο. Ο λαουτιέρης παίζει απλώς σιγά ή δυνατά, μαλακά ή σκληρά.

Η μελωδική έκταση του λαούτου είναι δύο οκτάβες και μια έκτη. Η έκταση όμως αυτή έχει περισσότερο θεωρητική αξία , καθότι το λαούτο σήμερα είναι ένα όργανο κυρίως συνοδευτικό. Συνοδεύει ρυθμικά και αρμονικά ορισμένα από τα κατεξοχήν μελωδικά όργανα: βιολί, κλαρίνο, λύρα;, περιορισμένο σε λίγες σύντομες μελωδικές φράσεις.

Το λαούτο ,όταν συνοδεύει, είναι εκείνο που «κρατάει το ρυθμό», το μέτρο με την περιοδική κυρίως επανάληψη ενός ρυθμικού σχήματος. Η αρμονική συνοδεία του λαούτου περιοριζόταν παλιότερα σε ένα ίσο (ισοκράτη, pedal).

5.1.2.2 Ούτι

Το ούτι (απ' το αραβικό al oud = ξύλο) έχει μεγάλο αχλαδόσχημο (επιπεδόκυρτο) ηχείο, κοντό και φαρδύ χέρι χωρίς μπερντέδες, κεφαλή που σχηματίζει σχεδόν ορθή γωνία με το χέρι και κλειδιά απ' τα πλάγια. Παίζεται με πλήκτρο (πένα, από φλούδα κορμού ή κλαδιού κερασιάς ή από κέρατο βοδιού και σήμερα από πλαστική ύλη). Έχει συνήθως πέντε διπλές εντέρινες χορδές (σε ουνίσονο), στερεωμένες στον καβαλάρη, πάνω στο καπάκι, κουρδισμένες κατά τέταρτες (σε ουνίσονο), εκτός από τη βαρύτερη χορδή που κουρδίζεται σε διάστημα τόνου από την επόμενη. Στην κατασκευή ακολουθεί την ίδια διαδικασία με το λαούτο.

Το ούτι, αν και το συναντάμε στην Ελλάδα, παίζεται σε περιορισμένη κλίμακα. Ούτι έπαιζαν αποκλειστικά οι Έλληνες της Μικράς Ασίας και της Κωνσταντινούπολης, οι οποίοι και αγνοούσαν το ελληνικό λαούτο με το μακρύ χέρι. Μετά την καταστροφή του 1922 και την ανταλλαγή των πληθυσμών , το ούτι χρησιμοποιήθηκε και στον ελλαδικό χώρο κάπως περισσότερο σε μικρασιατικά συνήθως γλέντια και σε συγκροτήματα που έπαιζαν πολύ και τούρκικη μουσική. Αυτό όμως δεν επηρέασε ποτέ την σταθερή θέση του λαούτου με το μακρύ χέρι που εξακολουθεί να χρησιμοποιείται πανελλήνια ως το κατεξοχήν ρυθμικό και αρμονικό όργανο, τόσο στη νησιώτικη ζυγιά όσο και στην κομπανία.

5.1.2.3 Λύρα

Η λύρα έχει ένα αχλαδόσχημο (επιπεδόκυρτο) ηχείο και κοντό χέρι, χωρίς μπερντέδες, που συνεχίζει το ηχείο, κλειδιά από πίσω προς τα εμπρός, καβαλάρη, τρεις μονές χορδές, στερεωμένες στο άκρο του ηχείου ή στο χτένι, και παίζεται με δοξάρι. Φτιάχνεται συνήθως από τον ίδιο τον εκτελεστή σε διάφορα μεγέθη, ανάλογα με τις διαστάσεις του ξύλου που θα βρει, ανάλογα με τη σωματική του διάπλαση και τη «φωνή» που θέλει να έχει η λύρα του (ψιλή και διαπεραστική, χοντρή και βαθιά).

Τα κύρια μέρη της λύρας είναι η *σκάφη*, η *κεφαλή* και το *χέρι*. Για την κατασκευή τους χρησιμοποιείται συνήθως μονοκόμματο ξύλο (μουριά, κισσός, πικροδάφνη, αγριαχλαδιά, καρυδιά, καστανιά, σφεντάμι, οξυά, δαμασκηνιά, κυπαρίσσι κ.α.). Η λύρα διαθέτει τρία κλειδιά, τα στριφτάλια, τα οποία φτιάχνονται σε διάφορα μεγέθη ανάλογα με τις ανάγκες του λυράρη. Στα κλειδιά αυτά τυλίγονται τρείς χορδές, που ακουμπούν στον καβαλάρη και δένονται στο άκρο της λύρας.

Η αχλαδόσχημη λύρα κουρδίζεται κατά πέμπτες καθαρές. Η μελωδία παίζεται στην πρώτη, υψηλότερη χορδή. Η δεύτερη χορδή (μια πέμπτη χαμηλότερη από την πρώτη) χρησιμοποιείται πολύ λίγο για την μελωδία και η

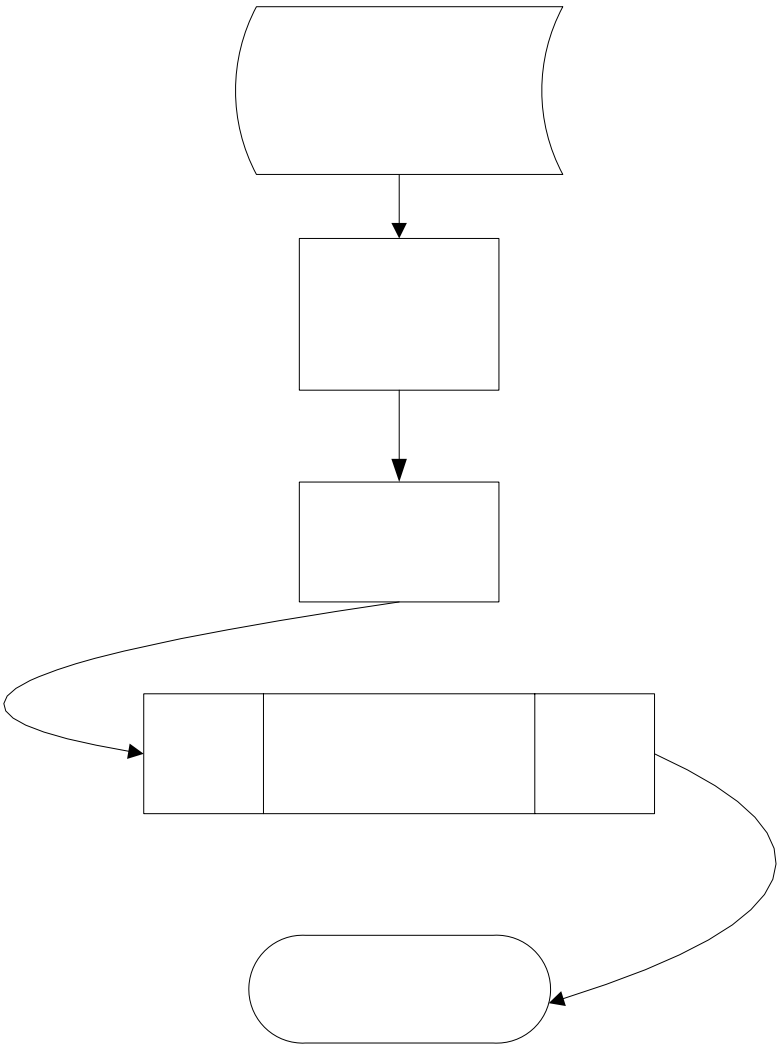
τρίτη χορδή (σε διάστημα πέμπτης ή τέταρτης από την δεύτερη) χρησιμοποιείται σπάνια.

Η αχλαδόσχημη λύρα είναι ένα όργανο κυρίως για γρήγορους σκοπούς, για χορευτικές μελωδίες, που παίζονται κατά κανόνα στην πούντα του δοξαριού, με γρήγορε χωριστές δοξαριές.

Με το πέρασμα των χρόνων η αχλαδόσημη λύρα παρουσιάζει ορισμένες μορφολογικές αλλαγές που συντελούνται προοδευτικά με κέντρο την Κρήτη. Η αρχική μορφή της αχλαδόσχημης λύρας είναι το *λυράκι*. Οι ανάγκες ωστόσο της μουσικής . ‘άλλες στο χορό και άλλες στο τραγούδι, περιορίζουν σιγά-σιγά το λυράκι για τις χορευτικές μελωδίες, καθότι αυτό παράγει οξύ και διαπεραστικό ήχο. Αντίθετα η συνοδεία του τραγουδιού δημιουργεί με τον καιρό ένα μεγαλύτερο τύπο οργάνου, που σιγά-σιγά γίνεται η βροντόλυρα.

5.2 Σύστημα ταξινόμησης παραδοσιακών ελληνικών οργάνων

Το γενικό διάγραμμα ροής για το εν λόγω σύστημα έχεί ως εξής:



σχήμα 5. 1 Γενικό διάγραμμα ροής του ΣΑΜΠΟ

5.2.1 Δεδομένα

Για την εκπαίδευση και ταξινόμηση του συστήματος ταξινόμησης, χρησιμοποιήθηκαν τόσο δείγματα από εμπορικές κυκλοφορίες, ηχογραφήσεις που έγιναν σε studio ηχογράφησης αλλά και ηχογραφήσεις σε home studio ειδικά για τους σκοπούς του ΣΑΜΠΟ.

Τα εμπορικά CD's από όπου εξήγχθησαν τα δείγματα ανήκουν στη σειρά "The Greek Volk Instruments" από την εταιρία F.M Records S.A. Συγκεκριμένα χρησιμοποιήθηκαν τα CD με αριθμό Volume 7(ούτι) και Volume 10 (λαούτο). Ηχογραφήσεις που έγιναν σε στούντιο αποτέλεσαν την πηγή για δείγματα που χρησιμοποιήθηκαν για την κρητική λύρα. Η ηχογράφηση πραγματοποιήθηκε στα πλαίσια ηχογράφησης του demo album ενός μουσικού συγκροτήματος , για την ηχογράφηση και μίξη του οποίου και ήμουν υπεύθυνος. Τα δείγματα εξήγθησαν από tracks τα οποία ήταν δοκιμαστικά και δεν χρησιμοποιήθηκαν περαιτέρω στη μίξη του album, και ως εκ τούτου δεν αποτελούν πνευματική ιδιοκτησία του συγκροτήματος.

Στα πλαίσια της παρούσας διατριβής, πραγματοποίησα ηχογραφήσεις στο προσωπικό μου home studio για τα όργανα: μπουζούκι, κρητική λύρα, ούτι, τουμπελέκι. Κατά την ηχογράφηση ελήφθησαν δείγματα από μεμονωμένες νότες σε όλη την γκάμα τονικών υψών και των τεσσάρων οργάνων, καθώς επίσης και από τραγούδια και αυτοσχεδιασμούς επίσης και για τα τέσσερα όργανα. Κατά την πορεία του σχεδιασμού του ΣΑΜΠΟ, όπως αναφέρθηκε και προηγούμενα, αποφασίστηκε να χρησιμοποιηθούν αποκλειστικά μουσικές φράσεις και όχι μεμονωμένες νότες, έτσι τα δείγματα ελήφθησαν από τις εκτελέσεις τραγουδιών και τους αυτοσχεδιασμούς.

Οι γενικές προδιαγραφές των ηχογραφήσεων, για αποφυγή ανομοιογένειας από τεχνητούς παράγοντες που δεν έχουν σχέση με τη φύση των οργάνων ,ήταν παντού οι ίδιες: Μονοφωνική, δειγματοληψία 44100 Hz και κβαντισμός 16 bits. Αυτές είναι άλλωστε και οι προδιαγραφές για τα εμπορικά CD. Τα κομμάτια που ελήφθησαν από τα CD, λόγω του ότι είχαν μορφή stereo, μετατράπηκαν σε μονοφωνικά χωρίς ιδιαίτερη επεξεργασία, απλά και μόνο με την κράτηση του ενός ηχητικού καναλιού από τα δύο. Αυτό δεν θα δημιουργούσε προβλήματα αφού επρόκειτο για ηχογραφήσεις μονοφωνικού σήματος από ένα όργανο. Οι ηχογραφήσεις που πραγματοποιήθηκαν στο home studio χαίρουν διαφοροποιήσεων όσον αφορά τη θέση και γωνία λήψης του μικροφώνου, τη ρύθμιση του κέρδους (gain) του προενισχυτή, καθώς και τη θέση του εκτελεστή στο χώρο ηχογράφησης. Για τις ηχογραφήσεις αυτές χρησιμοποιήθηκε υπερκαρδιοειδές μικρόφωνο AKG 535, προενισχυτές RME και κάρτα ήχου RME Fireface, και το λογισμικό Steinberg Cubase 3.1 για Windows.

Στη συνέχεια, μέσω του περιβάλλοντος Matlab 7 έγινε κατάτμηση και τυχαία ανάμιξη του μουσικού υλικού ανά όργανο. Τα τελικά δείγματα αποτελούνται από ένα δευτερόλεπτο μουσικής το καθένα και είναι ταξινομημένα ανά μουσικό όργανο. Στην πρώτη έκδοση του ταξινομητή χρησιμοποιούνται συνολικά 894 τέτοια δείγματα, από τα οποία τα 395 για την εκπαίδευση του συστήματος και τα 499 για την αξιολόγηση, δηλαδή περίπου 44% και 56% αντίστοιχα. Η κατανομή αυτή καθιστά σαφές ότι η εξέταση των

δειγμάτων απαιτούσε a priori τη δυνατότητα του συστήματος για γενίκευση. Τα παραπάνω αποτελούν την κύρια έκδοση του συστήματος για τα τέσσερα έγχορδα (λύρα, ούτι, μπουζούκι, λαούτο). Αναλυτικά τα δείγματα για τις διαδικασίες της εκπαίδευσης και της αξιολόγησης φαίνονται στον πίνακα 5.1.

	εκπαίδευση	Αξιολόγηση	**Σύνολο**
Λύρα	141	165	**306**
Ούτι	80	94	**174**
Μπουζούκι	50	63	**113**
Λαούτο	124	177	**301**
Σύνολο	**395**	**499**	**894**

Πίνακας 5. 1 Αριθμητική κατανομή των δειγμάτων της εφαρμογής

5.2.2 Εξαγωγή χαρακτηριστικών

Μετά από τη διαδικασία κατάτμησης του μουσικού υλικού, τυχαίας ανάμιξής του και ταξινόμησης του κατά όργανο, σειρά έχει η εξαγωγή των χαρακτηριστικών.

- Η εξαγωγή των χαρακτηριστικών στηρίχθηκε κατά μεγάλο μέρος στην τεχνική των MFCC (Mel Frequency Cepstral Coefficients) που περιγράφεται αναλυτικά στο 4^{ο} κεφάλαιο, παρ. 5. Γνωστή από τη χρήση της στην τεχνολογία φωνής, η εφαρμογή της τεχνικής αυτής για αρχεία ήχου, αν και λιγότερο διαδεδομένη, κρίνεται ιδιαίτερα χρήσιμη. Η χρησιμότητά της για το ΣΑΜΠΟ έγκειται, κατά τη γνώμη μου, σε δύο λόγους:

1. Κάνει χρήση της μοντελοποίησης της ανθρώπινης ακοής βάσει της κλίμακας MEL, η οποία φαίνεται στην πράξη να εκφράζει τον τρόπο αντίληψης των τονικών υψών (pitch) από τον άνθρωπο
2. Μειώνει τον αριθμό συντελεστών σε 13, συμπεριλαμβανομένης και της ενέργειας του σήματος ανά frame. Αυτό συνεπάγεται αισθητή μείωση του της διαστασιμότητας του διανυσματικού χώρου, σε σχέση με άλλα μοντέλα που χρησιμοποιούν μέχρι και 60 συντελεστές, γεγονός που έχει φανεί στην πράξη ότι πετυχαίνει διαχωρισιμότητα με μικρότερο αριθμό δειγμάτων προς εκπαίδευση, ενώ παράλληλα μειώνει την πολυπλοκότητα σε μνήμη και σε χρόνο.

Στην προσπάθεια προσομοίωσης του ανθρωπίνου τρόπου αντίληψης του ήχου και αναγνώρισης κάποιου μουσικού οργάνου, πραγματοποιήθηκε μια μάλλον ασυνήθιστη παραμετροποίηση κατά την εξαγωγή των cepstral συντελεστών. Συγκεκριμένα, σαν χρονικό παράθυρο για την εφαρμογή του FFT 256 σημείων χρησιμοποιήθηκε ολόκληρο το δείγμα διάρκειας ενός δευτερολέπτου. Αυτό έρχεται σε αντίθεση με τις εφαρμογές φωνής, στα οποία το παράθυρο κυμαίνεται μεταξύ 256-2024 samples σε μια δειγματοληψία

τάξεως 16kHz, και τις εφαρμογές μουσικής στις οποίες συνήθως δεν υπερβαίνει τα 0.5 δευτερόλεπτα. Προτιμήθηκε σε σχέση με την κατάτμηση του χρονικού παραθύρου και την εφαρμογή των MFCC σε επί μέρους frames προκειμένου να εισάγει μη γραμμικότητες στο μηχανικό σύστημα αναγνώρισης και επιπλέον να προσδώσει κάτι από την αργή επεξεργασία των δεδομένων και την εξαγωγή γενικευμένων συμπερασμάτων σε σχέση με τονικά ύψη και όχι με συχνότητες που λαμβάνει χώρα στον ανθρώπινο εγκέφαλο.

Από κάθε δείγμα ενός δευτερολέπτου έγινε εξαγωγή 13 συντελεστών cepstrum, ο ένας εκ των οποίων είναι η ενέργεια του φάσματος για το συγκεκριμένο δείγμα .

- Ένα δεύτερο χαρακτηριστικό που χρησιμοποιήθηκε αυτούσιο είναι η κεντροειδής του φάσματος (κεφ. 3.5.3). Αυτό λαμβάνει κάθε φορά την τιμή της συχνότητας στην οποία η ενέργεια του φάσματος είναι 50%, σαρώνοντας από τις χαμηλές συχνότητες.
- Χρησιμοποιήθηκε ακόμα το κατώφλι ενέργειας, όπως περιγράφεται στο 3.5.3. Σαν κρίσιμες τιμές έχουμε το 15%, 30% και 80%. Για λόγους αποτελεσματικότητας, μετά από σειρά δοκιμών κρατήθηκε το 30%.
- Τα τονικά ύψη (pitch) για το κάθε δείγμα. Το χαρακτηριστικό αυτό εξήγθη μέσω της τεχνικής του κορελογράμματος, όπως αυτό περιγράφεται στο κεφάλαιο 4.5.3. Από το κορελόγραμμα βρίσκεται το pitch για κάθε frame, και προκύπτει μια γραφική παράσταση του pitch σε συνάρτηση του χρόνου, γνωστή στη βιβλιογραφία σαν pitch weft. Στη συνέχεια λαμβάνεται ο χρονικός μέσος όρος και αποθηκεύεται σαν ένα μέσο pitch ανά δείγμα. Για την ανάλυση χρησιμοποιήθηκαν τα MFCC πριν την εφαρμογή του DCT και χρονικό παράθυρο 0.1 msec.

- Τα επόμενα χαρακτηριστικά είναι συνδυασμός των παραπάνω. Δεν κρίθηκε σκόπιμο να προστεθούν στα προηγούμενα χαρακτηριστικά για την αποφυγή πλεονάζουσας πληροφορίας, και γι’ αυτό δεν χρησιμοποιήθηκαν στην πρώτη εκδοχή του ΣΑΜΠΟ αλλά στις παραλλαγές του συστήματος αντικαθιστώντας εν μέρει τα προηγούμενα χαρακτηριστικά. Κατά την εξαγωγή των χαρακτηριστικών αυτών έγινε προσπάθεια συσχετισμού, με κάποιο τρόπο, του pitch με το ενεργειακό περιεχόμενο ανά συχνότητα. Προκειμένου να διατηρηθεί μικρή διαστασιμότητα στο διάνυσμα των χαρακτηριστικών, το pitch συνδυάστηκε με το κατώφλι ενέργειας στις διάφορες εκδοχές του και όχι με το ενεργειακό περιεχόμενο ανά μπάντα συχνοτήτων. Περισσότερα τις εν λόγω παραλλαγές στα χαρακτηριστικά αναφέρονται στο κεφάλαιο 5.4: «Παραλλαγές». Δεν έχω υπ’ όψη μου αναφορά για χρήση αυτού ή παρόμοιου χαρακτηριστικού στη σημερινή βιβλιογραφία, εντούτοις τόσο θεωρητικά όσο και στην πράξη είναι άξιο προσοχής.

Ανακεφαλαιώνοντας, η πρώτη εκδοχή του προγράμματος έχει σαν χαρακτηριστικά

1. Τους 13 συντελεστές MFCC
2. την κεντροειδή του φάσματος
3. Τη συχνότητα κατωφλίου ενέργειας 30%

Το 1^{o} και το τονικό ύψος για την εκδοχή v3, υλοποιήθηκαν σε περιβάλλον Matlab 7 με τη βοήθεια της βιβλιοθήκης συναρτήσεων “Auditory toolbox” για Matlab κατασκευασμένη από τον Slaney, το 1993, η διανομή της οποίας είναι ελεύθερη κάτω από το καθεστώς GNU. Όλα τα υπόλοιπα χαρακτηριστικά που εξήγθησαν στην πρώτη εκδοχή του ΣΑΜΠΟ και στις παραλλαγές, καθώς και ο κώδικας για την ταξινόμηση και για ότι άλλο χρειάστηκε υλοποιήθηκε χωρίς τη χρήση κάποιας άλλης εξωτερικής βιβλιοθήκης συναρτήσεων.

5.2.3 Ταξινόμηση

Η διαδικασία της ταξινόμησης στο σύστημα ΣΑΜΠΟ βασίστηκε στη θεωρία των νευρωνικών δικτύων. Συγκεκριμένα, έγινε μια υλοποίηση ενός πολυεπίπεδου Perceptron (multilayer Neural Network) μάθησης υπό επίβλεψη που ακολουθεί τον αλγόριθμό της όπισθεν διάδοσης σφάλματος (Back Propagation), όπως αυτός περιγράφεται αναλυτικά στο κεφάλαιο 4.3.2.4. Το νευρωνικό δίκτυο που κατασκευάστηκε περιέχει ένα κρυμμένο επίπεδο με νευρώνες ανάμεσα σε 8 και 15, ο αριθμός των οποίων ρυθμίζεται αυτόματα και κάθε φορά, βάση κριτηρίων βελτιστοποίησης. Τα κριτήρια αυτά είναι :

1. η επιτυχία ταξινόμησης στα δείγματα εκπαίδευσης να ξεπερνά το 83,29%.
2. Το κανονικοποιημένο μέσο τετραγωνικό σφάλμα να μην πέφτει κάτω από το κατώφλι του 0.067 για να αποφευχθεί το φαινόμενο της υπερεκπαίδευσης.
3. Οι εποχές εκπαίδευσης να μην ξεπερνούν τις 3000.

Όσον αφορά στις υπόλοιπες παραμέτρους του νευρωνικού, ο αλγόριθμος back propagation χρησιμοποιήθηκε στην batch εκδοχή του. Δεν χρειάστηκε να γίνει κάποια αρχικοποίηση στις μεταβλητές, καθώς ο αλγόριθμος άρχιζε πάντα να συγκλίνει από πολύ μικρό αριθμό εποχών χωρίς να παρατηρούνται σχεδόν καθόλου ταλαντώσεις καθ’ όλη τη διάρκεια της εκπαίδευσης. Σαν συνάρτηση-παράμετρος για την εκπαίδευση ,καταλληλότερη κρίθηκε η traingdx, μετά από δοκιμές, με κριτήρια την ταχύτητα σύγκλισης και τον αριθμό ταλαντώσεων.

5.2.4 Εξαγωγή αποτελεσμάτων

Για την εξαγωγή των τελικών αποτελεσμάτων, δημιουργήθηκαν παράλληλα 10 νευρωνικά δίκτυα του ίδιου τύπου τα οποία μεταχειρίζονταν επίσης τα ίδια δεδομένα. Αυτό έγινε με στόχο την αποφυγή εξαγωγής συμπερασμάτων που οφείλονται σε τυχαίους παράγοντες(τυχαία αρχικοποίηση στις τιμές των βαρών του νευρωνικού). Έτσι τα αποτελέσματα παρουσιάζονται φραγμένα σε τιμές ελάχιστης και μέγιστης ποσοστιαίας επιτυχίας. Στην πρώτη εκδοχή του ΣΑΜΠΟ η αξιολόγηση γίνεται για δείγματα

διάρκειας ενός δευτερολέπτου, ίδιας δηλαδή χρονικής διάρκειας με τα δείγματα της εκπαίδευσης. Τα αποτελέσματα θα αφορούν σε συνολικά ποσοστά επιτυχίας αναγνώρισης καθώς και ποσοστά επιτυχίας ανά όργανο, επί αγνώστων δειγμάτων.

5.3 Αποτελέσματα

Στην πρώτη εκδοχή του ΣΑΜΠΟ, χρησιμοποιήθηκαν τα χαρακτηριστικά που αναφέρθηκαν παραπάνω. Το διάγραμμα ροής του συστήματος φαίνεται στο σχήμα 5.2.

Το νευρωνικό πραγματοποίησε δέκα διαφορετικές εκπαιδεύσεις με τα κριτήρια που προαναφέρθηκαν. Η συνάρτηση εκπαίδευσης traingdx" επιλέχθηκε με βάση την ικανότητα γενίκευσης και την ταχύτητα εκπαίδευσης. Για κάθε εκπαίδευση έγινε ταξινόμηση για τα test data.

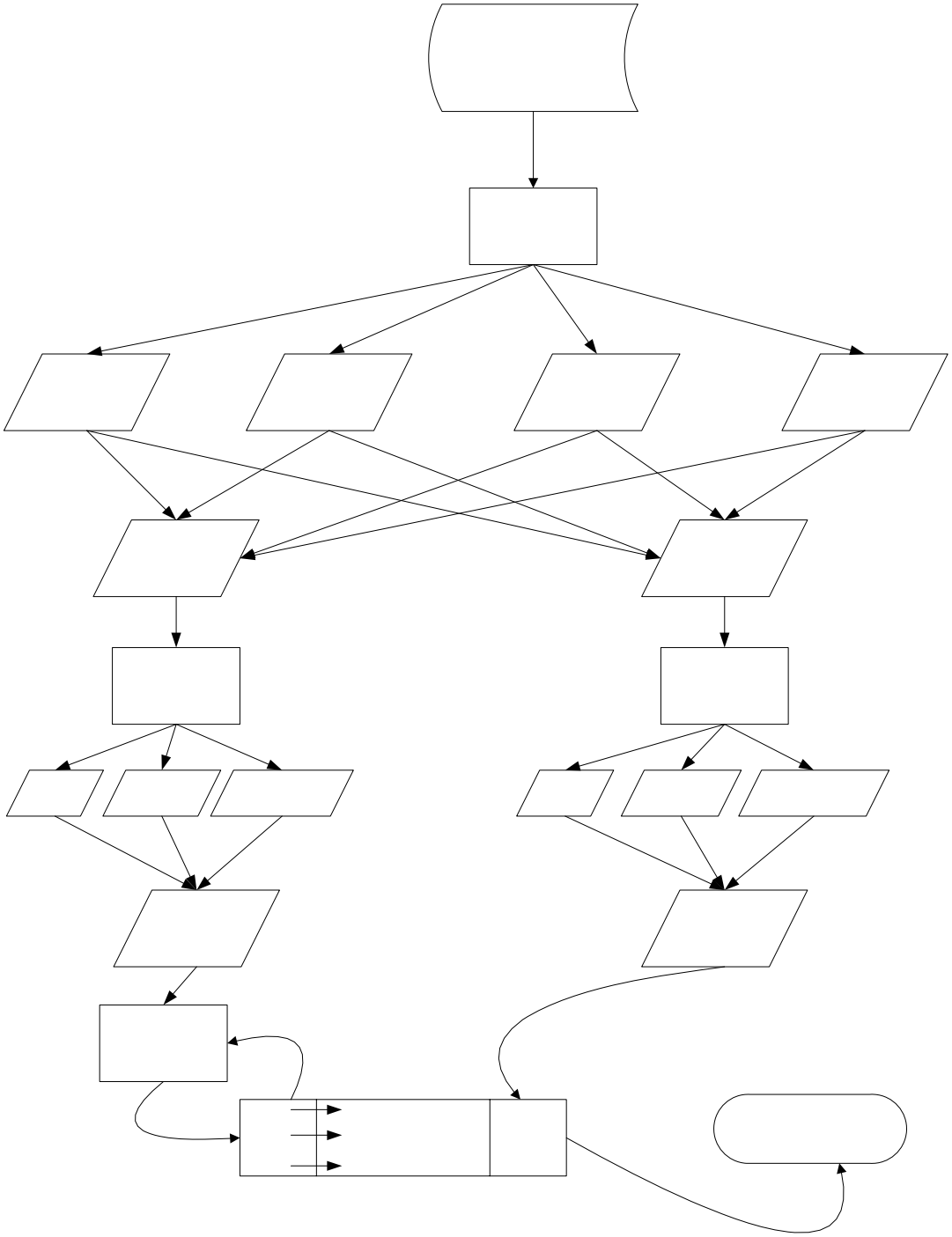
Η μέση επιτυχία του συστήματος στις 10 εκπαιδεύσεις φαίνεται στους πίνακες 5.2, 5.3.

	Ποσόστιαία μέση επιτυχία (10 επαναλήψεις)	Μέσος Αριθμός νευρώνων
Επιτυχία εκπαίδευσης	83,80%	16,1
Επιτυχία αγνώστων δειγμάτων	80,56%	

Πίνακας 5. 2 Μέση επιτυχία της πρώτης εκδοχής του ΣΑΜΠΟ

	Ποσοστιαία κυμαινόμενη επιτυχία (10 επαναλήψεις)	Κυμαινόμενος Αριθμός νευρώνων
Επιτυχία εκπαίδευσης	83,29 %– 84,81%	14-18
Επιτυχία αγνώστων δειγμάτων	78,96% - 81,16%	

Πίνακας 5. 3 Κυμαινόμενη επιτυχία πρώτης εκδοχής ΣΑΜΠΟ



σχήμα 5. 2 Το διάγραμμα ροής της πρώτης εκδοχής του ΣΑΜΠΟ.

Τα ποσοστά επιτυχίας αναγνώρισης επί του συνόλου των αγνώστων δειγμάτων όλων των οργάνων κυμάνθηκαν μεταξύ 78,96% και 81,16%, με μέσο ποσοστό επιτυχίας 80,56%. Τα αντίστοιχα ποσοστά για τα δείγματα εκπαίδευσης κυμάνθηκαν μεταξύ 83,29 % και 84,81%, με μέσο ποσοστό επιτυχημένης αναγνώρισης 83,8%. Σύμφωνα με τα κριτήρια σύγκλισης που είχαν τεθεί ,η απόδοση του νευρωνικού δικτύου στην εκπαίδευση γινόταν βέλτιστη για 16,1 νευρώνες, κατά μέσο όρο, στο κρυμμένο επίπεδο.

Η επιτυχία στην αναγνώριση και η λάθος ταξινόμηση ανά όργανο για την πρώτη εκδοχή, για την εκπαίδευση και την αξιολόγηση, φαίνεται στους πίνακες 5.4, 5.5. Η διαγώνιος των πινάκων εκφράζει τις επιτυχημένες απόπειρες αναγνώρισης, ενώ τα υπόλοιπα στοιχεία των σειρών τις λανθασμένες αποφάσεις.

Στους πίνακες 5.6, 5.7 φαίνονται τα αντίστοιχα ποσοστά επί τοις εκατό.

ΕΚΠΑΙΔΕΥΣΗ		Ταξινόμηση που πραγματοποίησε το σύστημα				
		Λύρα	Μπουζούκι	Ούτι	Λαούτο	**Σύνολο**
Προέλευση των δειγμάτων	Λύρα	**131,5**	1,9	2,7	4,9	**141**
	Μπουζούκι	4,8	**39,3**	1,9	4,0	**50**
	Ούτι	14	1,7	**48,4**	15,9	**80**
	Λαούτο	2,8	3,7	5,7	**111,8**	**124**

Πίνακας 5. 4 Αναγνώριση στα δείγματα της εκπαίδευσης, αναλυτικά ανά όργανο.

ΑΞΙΟΛΟΓΗΣΗ (άγνωστα δείγματα)		Ταξινόμηση που πραγματοποίησε το σύστημα				
		Λύρα	Μπουζούκι	Ούτι	Λαούτο	**Σύνολο**
Προέλευση των δειγμάτων	Λύρα	**156**	3,2	2,6	3,2	**165**
	Μπουζούκι	10	**37,5**	5,7	9,8	**63**
	Ούτι	17,2	3,9	**44,5**	28,4	**94**
	Λαούτο	8	2,7	6,8	**159,5**	**177**

Πίνακας 5. 5 Αναγνώριση στα άγνωστα δείγματα, αναλυτικά ανά όργανο.

ΕΚΠΑΙΔΕΥΣΗ		Ταξινόμηση που πραγματοποίησε το σύστημα %			
		Λύρα	Μπουζούκι	Ούτι	Λαούτο
Προέλευση των δειγμάτων %	Λύρα	**93,2624**	1,3475	1,9149	3,4752
	Μπουζούκι	9,6	**78,6**	3,8	8,0
	Ούτι	17,5	2,125	**60,5**	19,875
	Λαούτο	2,2581	2,9839	4,5968	**90,1613**

Πίνακας 5. 6 Ποσοστό επιτυχίας αναγνώρισης στα δείγματα της εκπαίδευσης αναλυτικά για κάθε όργανο

ΑΞΙΟΛΟΓΗΣΗ (άγνωστα δείγματα)		Ταξινόμηση που πραγματοποίησε το σύστημα %			
		Λύρα	Μπουζούκι	Ούτι	Λαούτο
Προέλευση των δειγμάτων %	Λύρα	**94,5455**	1,9394	1,5758	1,9394
	Μπουζούκι	15,8730	**59,5238**	9,0476	15,5556
	Ούτι	18,2979	4,1489	**47,3404**	30,2128
	Λαούτο	4,5198	1,5254	3,8418	**90,1130**

Πίνακας 5. 7 Ποσοστό επιτυχίας αναγνώρισης στα άγνωστα δείγματα αναλυτικά για κάθε όργανο

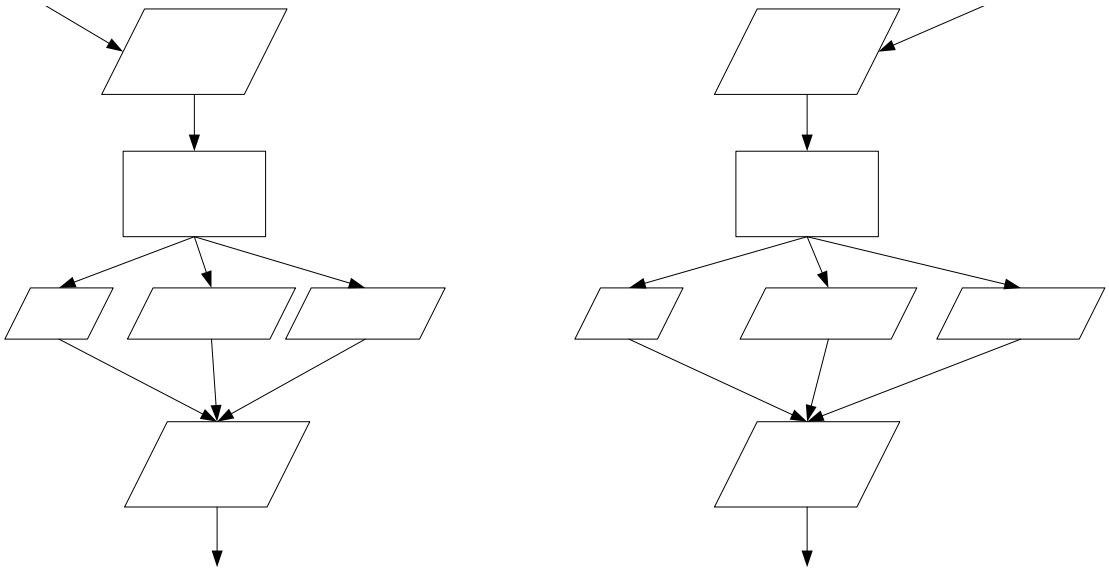
5.4 Παραλλαγές

5.4.1 Διαφοροποίηση στα φασματικά χαρακτηριστικά (ΣΑΜΠΟ v3).

Όπως ειπώθηκε και στο κεφάλαιο 5.2.2, έγινε προσπάθεια να συνδυαστούν με κάποιο τρόπο τα φασματικά χαρακτηριστικά με το pitch. Και αυτό γιατί η ενέργεια δεν είναι ανεξάρτητο χαρακτηριστικό, αλλά εκτός από τη φυσιολογία του οργάνου, εξαρτάται σε μεγάλο βαθμό και από το τονικό ύψος pitch, της νότας που παίζει το όργανο. Για παράδειγμα, όσο μειώνεται το pitch, το φάσμα γεμίζει με χαμηλές αρμονικές ,και η κεντροειδής συχνότητα του φάσματος μετατοπίζεται προς τα αριστέρα (μειώνεται). Κρίθηκε λοιπόν σκόπιμο να εισαχθεί ένα νέο χαρακτηριστικό: **Η σχετική συχνότητα κατωφλίου ενέργειας**. Αυτη δίνεται από τον τύπο:

$$REnRolloff = \log \frac{EnRolloff}{Pitch} \qquad (5.1)$$

όπου . Οι τιμές που δόθηκαν για το r ήταν 30% και 50% στην τιμή του $Enrolloff$ (κεφ. 3.5), και τα χαρακτηριστικό του pitch και της συχνότητας κατωφλιού ενέργειας αντικαταστάθηκαν από τα νέα χαρακτηριστικά.



σχήμα 5. 3 Η αλλαγή στη μονάδα εξαγωγής χαρακτηριστικών για την εκδοχή ΣΑΜΠΟ v3

Το νέο αυτό χαρακτηριστικό έχει την σχεδιάστηκε με την προσδοκία να παρουσιάσει αυξημένη συσχέτιση μεταξύ των δειγμάτων του ίδιου οργάνου. Ακολουθήθηκε η ίδια ακριβώς μέθοδος με την πρώτη εκδοχή του ΣΑΜΠΟ (10 εκπαιδεύσεις με τα ίδια κριτήρια ταξινόμησης). Τα αποτελέσματα είναι αντιστοίχως τα ακόλουθα.

	Ποσόστιαία μέση επιτυχία (10 επαναλήψεις)	Μέσος Αριθμός νευρώνων
Επιτυχία εκπαίδευσης	83,95%	17,4
Επιτυχία αγνώστων δειγμάτων	81.04%	

Πίνακας 5. 8 Μέση επιτυχία της εκδοχής v3 ΣΑΜΠΟ

	Κυμαινόμενη επιτυχία (10 επαναλήψεις)	Κυμαινόμενος Αριθμός νευρώνων
Επιτυχία εκπαίδευσης	83,54 %– 84,30%	14-20
Επιτυχία αγνώστων δειγμάτων	79,76% - 82,16%	

Πίνακας 5. 9 Κυμαινόμενη επιτυχία της εκδοχής v3 ΣΑΜΠΟ

ΕΚΠΑΙΔΕΥΣΗ		Ταξινόμηση που πραγματοποίησε το σύστημα				
		Λύρα	Μπουζούκι	Ούτι	Λαούτο	**Σύνολο**
Προέλευση των δειγμάτων	Λύρα	**132,4**	1,8	2,6	4,2	**141**
	Μπουζούκι	7,4	**37,6**	2,4	2,6	**50**
	Ούτι	11,0	2,6	**49,6**	16,8	**80**
	Λαούτο	3,2	3,4	5,4	**112,0**	**124**

Πίνακας 5. 10 Αναγνώριση στα δείγματα της εκπαίδευσης, αναλυτικά ανά όργανο

ΑΞΙΟΛΟΓΗΣΗ (άγνωστα δείγματα)		Ταξινόμηση που πραγματοποίησε το σύστημα				
		Λύρα	Μπουζούκι	Ούτι	Λαούτο	**Σύνολο**
Προέλευση των δειγμάτων	Λύρα	**159.8**	2.0	0.6	2.6	**165**
	Μπουζούκι	11.4	**40.2**	2.6	8.8	**63**
	Ούτι	20.0	3.4	**39.4**	31.2	**94**
	Λαούτο	5.8	2.4	3.8	b	**177**

Πίνακας 5. 11 Αναγνώριση στα άγνωστα δείγματα, αναλυτικά ανά όργανο

ΕΚΠΑΙΔΕΥΣΗ		Ταξινόμηση που πραγματοποίησε το σύστημα %			
		Λύρα	Μπουζούκι	Ούτι	Λαούτο
Προέλευση των δειγμάτων %	Λύρα	**93.9007**	1.2766	1.8440	2.9787
	Μπουζούκι	14.8	**75.2**	4.8	5.2
	Ούτι	13.7500	3.2500	**62.0**	21. 0
	Λαούτο	2.5806	2.7419	4.3548	**90.3226**

Πίνακας 5. 12 Ποσοστό επιτυχίας αναγνώρισης στα δείγματα της εκπαίδευσης αναλυτικά για κάθε όργανο

ΑΞΙΟΛΟΓΗΣΗ (άγνωστα δείγματα)		Ταξινόμηση που πραγματοποίησε το σύστημα %			
		Λύρα	Μπουζούκι	Ούτι	Λαούτο
Προέλευση των δειγμάτων %	Λύρα	**96.8485**	1.2121	0.3636	1.5758
	Μπουζούκι	18.0952	**63.8095**	4.1270	13.9683
	Ούτι	21.2766	3.6170	**41.9149**	33.1915
	Λαούτο	3.2768	1.3559	2.1469	**93.2203**

Πίνακας 5. 13 Ποσοστό επιτυχίας αναγνώρισης στα άγνωστα δείγματα αναλυτικά για κάθε όργανο

Τα ποσοστά παρουσιάζονται ελαφρώς βελτιωμένα, γεγονός που ενισχύει τη σημασία του νέου χαρακτηριστικού.

5.4.2 Διαφοροποίηση ως προς το χρονικό μέγεθος του δείγματος (ΣΑΜΠΟ v2).

Στην παραλλαγή αυτή δοκίμασα να αυξήσω το μέγεθος του χρονικού παραθύρου στα δείγματα. Αυτό έγινε στα πλαίσια του γεγονότος ότι, κατά την προσωπική μου άποψη, ο άνθρωπος συμβαίνει πολλές φορές να συνδυάζει παραμέτρους ξανά και ξανά ώσπου να αποφανθεί για την ταυτότητα ενός οργάνου. Έτσι τυχαίνει να επαναλάβει πάνω από μια φορά μια ευριστική, είτε για εξακρίβωση είτε απλά για επαλήθευση.

Για τη συγκεκριμένη εκδοχή του ΣΑΜΠΟ εξήγθησαν από τα ίδια ακριβώς αρχεία δείγματα χρονικής διάρκειας 2 sec. Η επεξεργασία , η εξαγωγή χαρακτηριστικών και η εκπαίδευση έγινε κατά χρονικά παράθυρα 1 sec. Πρόκειται λοιπόν για την περίπτωση όπου το σύστημα έχει 2 sec άγνωστου δείγματος στη διάθεσή του προκειμένου να αποφανθεί για την πηγή από την οποία αυτό προέρχεται. Το σύστημα κάνει την ανάλυση ανά sec και περιμένει να βγάλει ίδιο αποτέλεσμα και στα δύο υποδείγματα.
Η σκέψη αυτή υλοποιήθηκε με προσθήκη, μετά τη χρήση του νευρωνικού και στο τελευταίο μέρος του συστήματος μιας πρόσθετης μονάδας επεξεργασίας (σχήμα 4).

Η αλλαγή αυτή επέφερε, όπως ήταν αναμενόμενο, αξιόλογη αύξηση της τάξεως του 5%. Αναλυτικά:

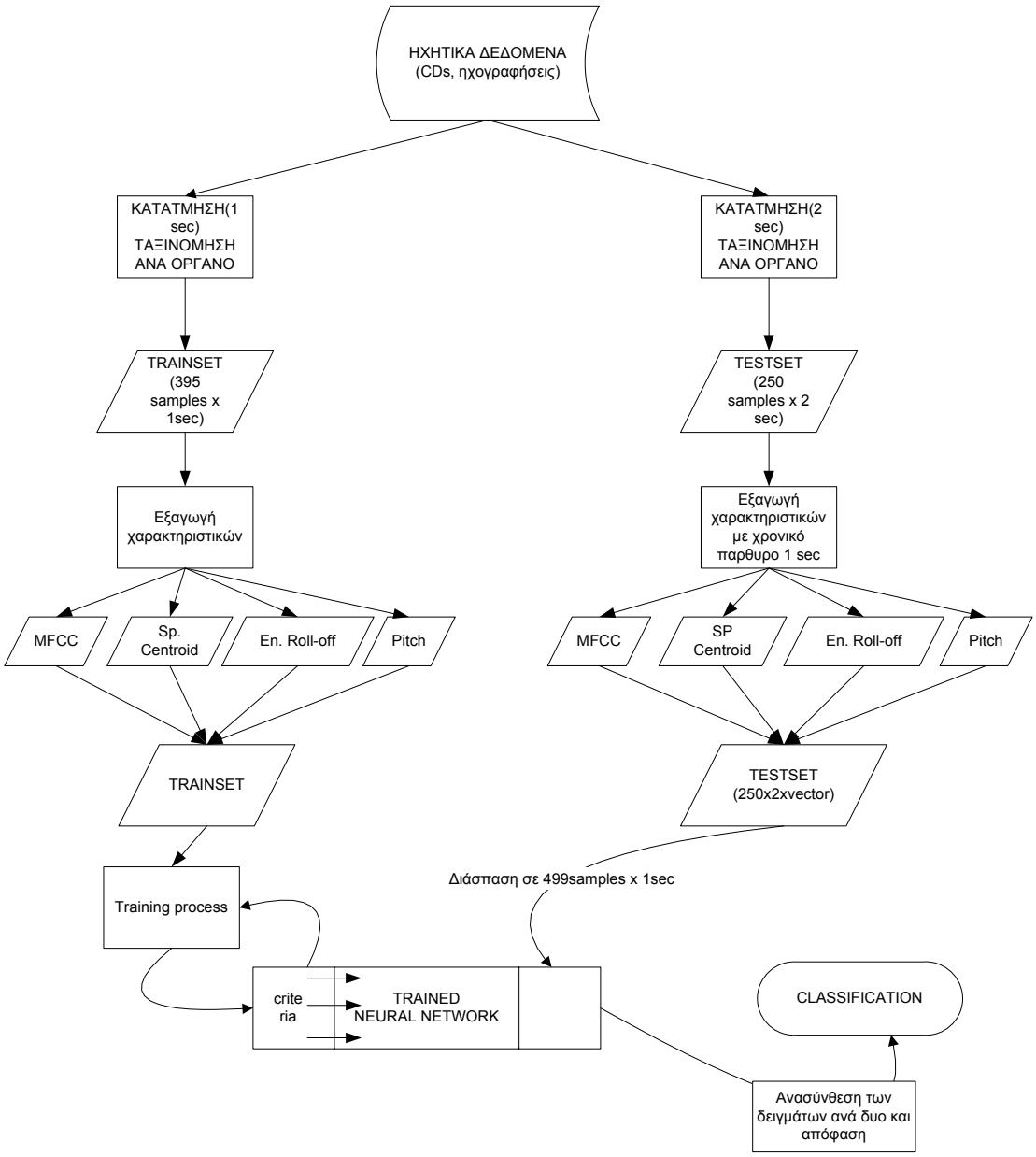
0.8375 15.6000
0.8557 15.6000

	Ποσόστιαία μέση επιτυχία (10 επαναλήψεις)	Μέσος Αριθμός νευρώνων
Επιτυχία εκπαίδευσης	83,75%	15,6
Επιτυχία αγνώστων δειγμάτων	85.57%	

Πίνακας 5. 13 Μέση επιτυχία της εκδοχής v2 ΣΑΜΠΟ

	Κυμαινόμενη επιτυχία (10 επαναλήψεις)	Κυμαινόμενος Αριθμός νευρώνων
Επιτυχία εκπαίδευσης	83,54 %– 84,30%	14-18
Επιτυχία αγνώστων δειγμάτων	84,37% - 86,77%	

Πίνακας 5. 14 Κυμαινόμενη επιτυχία της εκδοχής v2 ΣΑΜΠΟ



σχήμα 5. 4 Το διάγραμμα ροής στην εκδοχή ΣΑΜΠΟ v2

Τα αποτελέσματα της διαδικασίας αυτής παρουσιάζονται αναλυτικά στο Παράρτημα 1.

5.5 Συμπεράσματα

Από τα αποτελέσματα για την πρώτη έκδοση αλλά και τις δύο παραλλαγές v2 και v3 του Συστήματος Αναγνώρισης Μουσικών Παραδοσιακών Οργάνων ΣΑΜΠΟ εξήγησαν πολλά και σημαντικά συμπεράσματα.

- Επαληθεύτηκε η άποψη ότι τα MFCC παρουσιάζουν πολύ μεγάλη ευελιξία και στην αναγνώριση μουσικής πηγής. Αυτό γιατί βασίζονται στην κλίμακα Mel, η διατύπωση της οποίας αν και προηγείται χρονικά (1937) από άλλα πιο σύνθετα και ολοκληρωμένα μοντέλα για τον κοχλία ,ωστόσο φαίνεται πως ανταποκρίνεται στη φύση της ανθρώπινης ακοής.
- Για μια αναγνώριση 4 οργάνων, ο αριθμός των χαρακτηριστικών δεν πρέπει να υπερβαίνει την διαστασιμότητα 15-16d, καθώς τότε μειώνεται η διαχωριστική ικανότητα του ταξινομητή. Για περισσότερα χαρακτηριστικά κρίνεται απαραίτητη η χρήση κάποιας μεθόδου μείωσης της διαστασιμότητας του χώρου χαρακτηριστικών.
- Τα στιγμιαία (temporal) χαρακτηριστικά, αν εξαιρέσουμε το pitch το οποίο μπορεί να θεωρηθεί ένα από αυτά, δεν είναι απαραίτητα για την αναγνώριση. Αντιθέτως, τα φασματικά χαρακτηριστικά όπως αυτά που χρησιμοποιούνται στο ΣΑΜΠΟ πετυχαίνουν ικανοποιητική απόδοση αναγνώρισης.
- Τα χαρακτηριστικά της ενέργειας όπως η κεντροειδής συχνότητα φάσματος και η συχνότητα κατωφλίου ενέργειας είναι πολύ σημαντικά για την αναγνώριση. Για τη συχνότητα κατωφλίου ενέργειας, όταν προκύπτει από επεξεργασία σήματος που έχει ληφθεί υπό συνθήκες θορύβου του πραγματικού κόσμου, πιο αποδοτικές για την ταξινόμηση είναι οι τιμές r<50% (συχνότητα μικρότερη της κεντροειδούς). Αυτό είναι λογικό καθώς στις υψηλές τιμές r οι συχνότητες του μουσικού σήματος είναι έντονα αναμεμειγμένες με θόρυβο.
- Η σχετική συχνότητα κατωφλίου ενέργειας (εξ.5.1), το νέο χαρακτηριστικό που παρουσιάστηκε στα πλαίσια της παρούσας διατριβής, κρίνεται αποδοτικό, καθώς συνδιάζει δύο μεταβλητές-χαρακτηριστικά που αλληλοεξαρτώνται, και αποτελεί πιο ανεξάρτητο κριτήριο .
- Η αύξηση της απόδοσης ενός συστήματος αναγνώρισης μπορεί να επιτευχθεί και με μεθόδους που δεν στηρίζονται άμεσα στις τεχνικές εξαγωγής χαρακτηριστικών και ταξινόμησης, αλλά εκμεταλλεύονται πλεονεκτήματα του εκάστοτε συστήματος προς αναγνώριση, όπως τεχνικές ανθρώπινης και τεχνητής νοημοσύνης (κεφ. 5.4.2 ΣΑΜΠΟ v2).
- Η μεγάλη δυσκολία συνίσταται στον διαχωρισμό ανάμεσα στο ούτι και στο λαούτο. Τα δύο αυτά όργανα είναι «αδέλφια», καθώς το ούτι δεν είναι παρά η «ανατολική» εκδοχή του λαούτου και έχουν πολύ κοντινές ακουστικές ιδιότητες. Ως εκ τούτου είναι πολύ λογικό να υπάρχει πρόβλημα στο διαχωρισμό τους. Δυστυχώς, προς απόδειξη αυτού δεν υπήρξε δυνατότητα για πείραμα με ανθρώπους ακροατές.
- Τέλος τα δυνατά ποσοστά στην αναγνώριση της λύρας πιστοποιούν ότι το σύστημα αποκρίθηκε με επιτυχία στην αναγνώριση του μοναδικού

οργάνου με δοξάρι, ανάμεσα σε 3 άλλα νυκτά όργανα, γεγονός που αποδυκνύει ότι στα εξαγόμενα χαρακτηριστικά ενσωματώθηκαν άριστα οι ακουστικές ιδιότητες των οργάνων.

Κεφάλαιο 6 Συμπεράσματα

6.1 Σύνοψη και συμπεράσματα

Στα πλαίσια της παρούσας διατριβής εξετάστηκε το πρόβλημα της αναγνώρισης μουσικών οργάνων.

Στο κεφάλαιο 1, έγινε μια εισαγωγή για το πρόβλημα της αναγνώρισης μουσικής πηγής, την ταυτότητα του ερευνητικού πεδίου αυτού και της σημασίας του.

Στο κεφάλαιο 2 έγινε ιστορική αναφορά στα πεπραγμένα και τη βιβλιογραφία του ερευνητικού πεδίου της αναγνώρισης μουσικής πηγής και τις εκφάνσεις που αυτό έχει σήμερα. Σαν τελευταία αναφερθείσα κατηγορία, η αναγνώριση μουσικών οργάνων παρουσιάστηκε σαν πρόβλημα και τέθηκαν τα κριτήρια για την αξιολόγηση ενός τέτοιου συστήματος.

Στο κεφάλαιο 3 αφού δόθηκε το μαθηματικό υπόβαθρο που σχετίζεται με την κατασκευή διανύσματος χαρακτηριστικών, έγινε αναφορά των θεωριών για την ανθρώπινη ακοή και τη φυσιολογία του ανθρώπινου αυτιού καθώς και τις ακουστικές ιδιότητες γνωστών οργάνων. Παρουσιάστηκε η ερευνητική δραστηριότητα στην αναγνώριση μουσικών οργάνων. Στη συνέχεια έγινε λόγος για τον διαχωρισμό της φύσης των χαρακτηριστικών που αποτελούν τη βάση για την ταξινόμηση και τα κριτήρια επιλογής τους. Παρουσιάστηκαν τα πιο σημαντικά από αυτά.

Στο κεφάλαιο 4 δόθηκαν οι αρχές τις ταξινόμησης. Στο τέλος του κεφαλαίου γίνεται εκτενής αναφορά στα νευρωνικά δίκτυα, στον αλγόριθμο όπισθεν διάδοσης στον οποίο στηρίζεται η εκπαίδευσή του και τις εφαρμογές τους.

Στο κεφάλαιο 5 παρουσιάστηκε η εφαρμογή που υλοποιήθηκε στα πλαίσια της παρούσας διπλωματικής διατριβής, το «Σύστημα Αναγνώρισης Μουσικών Παραδοσιακών Οργάνων – Σ.Α.Μ.Π.Ο», ο σχεδιασμός και τα αποτελέσματα.

Σαν συμπέρασμα από όλη τη μελέτη και την έρευνα που πραγματοποιήθηκε στα πλαίσια της παρούσας διατριβής , η αναγνώριση οργάνων είναι αυτή τη στιγμή ένα πρόβλημα που παραμένει μεν ανοικτό, ωστόσο μπορεί να αντιμετωπιστεί επιμεριστικά με επιτυχία.

Όπως και σε πολλά άλλα ερευνητικά πεδία, έτσι και στο πεδίο της μηχανικής αναγνώρισης, η ανάδραση από την ανθρώπινη αντίληψη και φυσικά ανάλογα θα κρίνει σε μεγάλο βαθμό ,κατά τη γνώμη μου, τις εξελίξεις.

Η σταθεροποίηση των συστημάτων αναγνώρισης σε ένα μεγάλο ποσοστό επιτυχίας διευκολύνεται με τη χρήση υβριδικών μεθόδων από νευρωνικά δίκτυα , τεχνητή νοημοσύνη και θεωρία επεξεργασίας σήματος και στατιστικές μεθόδους. Με τη χρήση των κατάλληλων μεθόδων τόσο στην εξαγωγή χαρακτηριστικών όσο και στην ταξινόμηση μπορούμε να επιτύχουμε ένα υψηλό βαθμό απόδοσης του συστήματος, τουλάχιστον όσον αφορά την αναγνώριση μονοφωνικού σήματος. Για την αναγνώριση και μουσική καταγραφή πολυφωνικού σήματος, απαιτούνται ακόμα πολλά βήματα τόσο σε ερευνητικό ψυχοφυσικό επίπεδο όσο και σε μαθηματικό και υπολογιστικό, προκειμένου να φτάσουμε να έχουμε αξιόλογα αποτελέσματα. Το συγκεκριμένο αποτελεί και πρόβλημα πολλές φορές και για τον άνθρωπο-

ακροατή, γεγονός που για πολλούς ίσως βάζει ένα ανώτατο όριο στην επιτυχία προσέγγισης του προβλήματος.

6.2 Μελλοντική εργασία

Στα πλαίσια της παρούσας διατριβής, αποκόμισα πολλά θετικά στοιχεία – εναύσματα για μελλοντική εργασία πάνω στον τομέα της αναγνώρισης μουσικών οργάνων. Το ΣΑΜΠΟ απέδειξε ότι είναι δυνατή η αναγνώριση οργάνων της ίδιας κατηγορίας με πολύ ικανοποιητικά , σε σχέση με το βαθμό δυσκολίας του εγχειρήματος, ποσοστά επιτυχίας.

Το γεγονός ότι δόθηκε βάρος στα αντιληπτικά χαρακτηριστικά και το ότι εξήγθησαν τα συγκεκριμένα αποτελέσματα βάσει «πραγματικών» δειγμάτων αποτέλεσε ένα επίσης άκρως θετικό στοιχείο.
Η επιτυχία του νέου χαρακτηριστικού στην εκδοχή ΣΑΜΠΟ v3, με οδηγεί στην διερεύνηση και άλλων δυνατών συνδυασμών αλληλοεξαρτώμενων χαρακτηριστικών.

Έτσι, στο άμεσο μέλλον θα διερευνηθεί πιθανή μερική αποτυχία του συστήματος οφειλόμενη στις αναλογίες και την έλλειψη πλουραλισμού ιδιοτήτων στα δείγματα. Η νέα μελέτη θα συνδυαστεί και με πείραμα με ανθρώπους –ακροατές πάνω στα ίδια δείγματα, και θα εξαγθεί και ποσοστό επιτυχίας επί των ανθρώπινων επιδόσεων.

Το ποσοστό επιτυχίας που αγγίζει, στην έκδοση ΣΑΜΠΟ v2 το 86%, δίνει το ερέθισμα για υλοποίηση του συστήματος αναγνώρισης μονοφωνικής πηγής-οργάνου με τη μορφή εφαρμογής σε παραθυρικό περιβάλλον. Μάλιστα, η σκέψη είναι να χρησιμοποιηθεί σαν back-end σε επεκτάσιμη εφαρμογή ,προκειμένου να αναγνωρίζει όργανα μεταξύ της ίδιας οικογένειας, συνδυαζόμενη με πιο στιβαρές και αποτελεσματικές μεθόδους για την αναγνώριση οικογενειών των οργάνων, σε προηγούμενο στάδιο επεξεργασίας.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Helmholz H. (1954), *On the sensations of Tone as a Physiological Basis for the Theory of Music.* (A.J. Ellis, translator, original work published in 1885, Second English Ed.) New York: Dover , p. 29

[2] Bregman A. (1990), *Auditory Scene Analysis*. Cambridge: MIT Press , p.222

[3] McAdams S. (1993), Recognition of sound sources and events. In *Thinking in Sound: the Cognitive of Human Audition*. Oxford University Press, p.147

[4] S. Haykin(1999), Neural Networks: A Comprehensive Foundation, Prentice-Hall

[5] Σ. Τζαφέστα (2002), Υπολογιστική Νοημοσύνη, Τόμος Α: Μεθοδολογίες, Ε.Μ.Π.

[6] V. Beiu, J.G. Taylor (1995), *Optimal VLSI Implementations of Neural Networks: VLSI-Friendly Learning Algorithms*. The Applied Decision Technologies ADT'95

[7] P.J. Werbos (1974), Beyond Regression: New tools for prediction and analysis in the Behavioral Sciences. Ph.D. thesis, Harvard University,

[8] D.B. Parker (1985), Learning-Logic, Technical Report TR-47, MIT

[9] D.E. Rumelhart, G.E. Hinton, R.J. Williams (1986), *Learning Internal Representations by Error Propagation, Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*, MIT Press, Cambridge, MA

[10] T.P. Vogl, J.K. Mangis, A.K. Rigler, W.T. Zink, D.L. Alkon (1988), *Accelerating the Convergence of the Back-propagation Method, Biological Cybernetics*, vol. 59, p. 257-263

[11] R.A. Jacobs (1988), *Increased Rates of Convergence through Learning Rate Adaptation*, Neural Networks, vol. 1, no. 4, pp. 295-307

[12] G.M. Georgiou, C. Koutsougeras (1992), Embedding Domain Information in Backpropagation, In *Proceedings of SPIE Conference on Adaptive and Learning Systems*, Orlando, Fla. Society of Photo-Optical Instrumentation Engineers, Bellingham, WA

[13] S.G. Smyth (1992), Designing Multi Layer Perceptrons from Nearest Neighbor Systems, *IEEE Transactions on Neural Networks*, vol. 3, no. 2, p. 329-333

[14] R. Rojas (1994), Optimal Weight Initialization for Neural Networks, In *Proceedings of the International Conference on Artificial Neural Networks* (ICANN'94), pp. 577-580, Springer Verlag, London

[15] S.M. Weiss, C.A. Kulikowski (1991), *Computer Systems that Learn*, Morgan Kaufmann

[16] M. Stone (1974), Cross-validation Choice and Assessment of Statistical Predictions, *Journal of Royal Statistical Society*, vol. B32, p. 111-133

[17] A.N. Kolmogorov (1963), *On the Representation of Continuous Functions of Several Variables by Superpositions of Continuous Functions of one Variable and Addition*, American Mathematical Society Translations, vol. 28, p. 55-59

[18] Σ. Μουγιακάκου (2003), *Ανάπτυξη Συστημάτων Υποστήριξης Κλινικών Αποφάσεων με Χρήση Μεθόδων Τεχνητής Νοημοσύνης*, Διδακτορική Διατριβή, Ε.Μ.Π.

Adler Samuel (2002), *The Study of Orchestration*, 3rd Ed.: W.W. Norton & Company, Inc.

Alan V. Oppenheim, Ronald W. Schafer (1989), *Discrete – Time Signal Processing*, International Ed.: Prentice-Hall International, Inc.

Berger K.W. (1964), Some factors in the recognition of timbre, *J.Acoust.Soc.Am.*36, 1888-1891

Bourne, J.B. (1972), *Musical Timbre Recognition Based on a Model of the Auditory System*, Master Thesis. Massachusetts Institute of Technology, Cambridge ,MA

Brown G.J. (1992), Computational Auditory Scene analysis: A Representational Approach. Ph.D. thesis, University of Sheffield

Brown G.J. & Cooke M. (1994), Perceptual grouping of musical sounds: A computational model. *Journal of New Music Research* 23, p.107-132

Brown J. C. (1997), Computer identification of musical instruments using pattern recognition. *Presented at the 1997 Conference of the Society of Music Perception And Cognition*, Cambridge, MA

Brown J. C. (1997), Cluster – based probability model for musical instrument identification . *J. Acoust.Soc.Am*. 101, 3167 (abstract only)

Brown J. C. (1998), Personal communication

Brown J. C. (1998), Computer identification of wind instruments using cepstral coefficients. *J. Acoust.Soc.Am*. 103, 2967 (abstract only)

Brown J. C. (1998), Computer identification of wind instruments using cepstral coefficients. *In Proceedings of the 16th International Concress on Acoustics and 135th Meeting of the Acoustical Society of America (pp. 1889-1890).* Seattle

Brown J. C. (1999), Computer identification of wind instruments using cepstral coefficients. *J. Acoust.Soc.Am*. 105(3) 1933-1941

Brown R. (1981). An experimental study of the relative importance of acoustic parameters for auditory speaker recognition. *Language and Speech* 24, 295-310

Campell W.C. & Heller J.J. (1979), Convergence procedurerces for investingating music listening tasks. *Bulletin of the Council for Research in Music Education* 59, 18-23

Cooke M. (1993), Modeling Auditory Processing and Organization. Cambridge: Cambridge University Press

De Poli G. & Prandoni P. (1997), Sonological models for timbre characterization, *Journal of New Music Research* 26, 170-197

De Poli G. & Tonella P. (1993), Self – organizing neural networks and Grey's timbre space. In *Proceedings of the 1993 International Computer Music Conference* (pp. 441- 444)

Dillon H. (1981), The Perception of Musical Instruments (Doctoral dissertation, University of South Wales Australia, 1979) *Dissertation Abstracts Internationa*l 41, 2703B-2704B

Dubnov S. & Rodet X. (1998), Timbre recognition with combined stationary and temporal features. In *Proceedings of the 1998 International Computer Music Conference* (pp102-108)

Eagleson H.V. & Eagleson O.W. (1947), Identification of musical instruments when heard directly and over a public-adress system. J. Acoust. Soc. Am. 19(2), 338-342

Elliott C. (1975), Attacks and releases as factors in instrument identification, *Journal of Research in Music Education*23, 35-40 (As cited by Kendall, 1986)

Ellis D. & Rosental D. (1995), Mid-level representations for computational auditory scene analysis. In *Proceedings of the International Joint Conference on Artificial Intelligence workshop on Computational Auditory Scene Analysis.*

Ellis D.P.W. (1994), A computer implementation of psychoacoustic grouping rules. *In Proceedings of the 12th Intl. Conf. on Pattern Recognition*. Jerusalem.

Ellis D.P.W. (1996), *Prediction-driven computational auditory scene analysis*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA

Everest F. Alton. *Εγχειριδιο Ακουστικής*, 3η Έκδοση, Εκδόσεις Α. Τζιόλα Ε., Θεσσαλονίκη

Feiten B. & Gunzel S. (1994), Automatic indexing of a sound database using self-organizing neural nets. *Computer Music Journal* 18(3), 53-65

Fujinaga I. (1998), Machine recognition of timbre using steady-state tone of acoustic musical instruments. In *Proceedings of the 1998 International Computer Music Conference* (pp. 207-210)

Handel S. (1989), *Listening*, Cambridge :MIT Press

Handel S. (1995), Timbre perception and auditory object identification. In B.C.J. Moore (ed.) *Hearing,* New York :Academic Press

Haykin Simon (1995), *Συστήματα Επικοινωνίας*, επιμέλεια Ε.Δ. Συκάς, Μ.Ε. Θεολόγου, εκδ. Παπασωτηρίου

Howard M. David, Angus Jamie (2005), *Acoustics and Psychoacoustics*, 3rd Ed.:Focal Press

Hawley M.J. (1993), Structure out of Sound, Ph.D. thesis, Massachusetts Institute of Technology, Program in Media Arts and Sciences, Cambridge MA

Kaminskyj I. & Materka A. (1995), Automatic source identification of monophonic musical instrument sounds. In Proceedings *of the 1995 IEEE International Conference on Neural Networks* (pp. 189-194)

Kashino K. & Murase H. (1997), Sound source identification for ensemble music based on the music stream extraction. In *Proceedings of the 1997 International Conference on Acoustics, Speech and Signal Processing*, Seattle

Kashino K. & Murase H. (1998), Music recognition using note transition context. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle

Kashino K., Nakadai K., Kinoshita T. & Tanaka H. (1995), Application of Bayesian probability network to music scene analysis. In *Proceedings of the 1995 International Joint Conference on Artificial Intelligence*

Kashino K. & Tanaka H. (1993), A sound source separation system with the ability of automatic tone modeling. In *Proceedings of the 1993 International Computer Music Conference*

Kendall R.A. (1986), The role of acoustic signal partitions in listener categorization of musical phrase. *Music Perception* 4(2), 185-214

Klasser F.I. (1996), Data *Reprocessing in Signal Understanding Systems*, Ph.D. thesis, University of Massachusetts Amherst

Langmead C.J. (1995), Sound analysis, comparison and modelification based on a perceptual model of timbre. In *Proceedings of the 1995 International Computer Music Conference*

Langmead C.J. (1995), *A Theoretical Model of Timbre Perception Based on Morphological Representations of Time-Varying Spectra.* Master thesis. Dartmouth College

Li X., Logan R.J. & Pastore R.E. (1991), Perception of acoustic source characteristics: Walking sounds, *J.Acoust.Soc.*A.90, 3036-3049

Mammone R., Zhang X. & Ramachandran R.P. (1996), Robus speacker recognition: A feature-based approach. *IEEE Signal Processing Magazine*13(5), 58-71

Marques J. (1999), An *Automatic Annonation System for Audio Data Containing Music*, Master's thesis Massachusetts Institute of Technology, Cambridge MA

McAdams S. & Cunible J.C. (1992). Perception of timbral analogies. *Phil. Trans.R.Soc.Lond .* B336, 383-389

McAdams S., Winsberg S., Donnadieu S., De Soete G. & Krimphoff J. (1995), Perceptual scaling of synthesized musical timbres: Coomon dimensions, specificities, and latent subject classes. *Phychol. Res.58*, 177-192

Melliger D.K. (1991). *Event Formation and Separation in Musical Sound*, Ph.D. thesis, Stanford University

Misnky M. (1974), *A Framework for representing Knowledge*, Massachusetts Institute of Technology, Cambridge MA

Misnky M. (1986), *The Society of Mind*, New York: Simon & Schuster

Nakatani T., Kashino K. & Okuno H.G. (1997), Integration of speech stream and music stream segregations based on a sound ontology. In *Proceedings of the 1997 International Joint Conference on Artificial Intelligence*

Nooralahiyan A. Y., Kirby H.R. & McKeown D. (1998), Vehicle classification by acoustic signature, *Mathl. Comput. Modeling* 27(9-11), 205-214

Pfeifer S., Fischer S. & Effelsberg W. (1996), Automatic audio content analysis. Universität Mannheim Technical Report, Mannheim, Germany

Reynolds D.A. (1995), Speaker identification and verification using Gaussian mixture speaker models, *Speech Communication*17, 97-108

Rosch E. (1978), Principles of categorization. In E. Rosch & B.B. Lloyd (eds.), *Cognition and Categorization*, Hillsdate, NJ: Lawrence Erlbaum

Rosch E., Mervis C.B., Gray W.D., Johnson D.M. & Boyes – Braem P. (1976), Basic objects in natural categories*. Cognitive Phycology*8, 382-439

Rossing D. Thomas, Moore F. Richard, Wheeler Paul A. (2001), *The Science of Sound*, 3rd Ed.: Addition Wesley

Saint-Arnaud N. (1995), *Classification of Sound Textures*. Master's thesis, Massachusetts Institute of Technology, Cambridge MA

Saldanha E.L. & Corso J.F. (1964), Timbre cues and the identification of musical instruments, *J.Acoust. Soc.Am*. 36, 2021-2026

Scheirer E.D. & Slaney M. (1997), Construction and evalution of a robust multifeature speech/music discriminator. In *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech and Signal Processing*, Munich

Spina M. & Zue V. (1996), Automatic transcription of general audio data: Preliminary analyses. In *Proceedings of the International Conference on Spoken Language Processing* (pp. 594-597)

Stumpf C. (1926), *Die Sprachlaute*, Berlin: Springer – Verlag (As cited by Kendall, 1986)

Thayer R. (1972), The effect of the attack transient on aural recognition of instrumental timbre. In J. Heller & W. Campbell (eds.), *Computer Analysis of the auditory characteristics of musical performance* (pp. 80-101). Final Report (proj. No 9-0564A), U.S. Department of Health, |Education and Welfare Bureau of Research (As citied by Kendall, 1986)

Volodin A. (1972). [The perception of transient processes in musical sounds]. Voprosy Psikholgii18(4), 51-60 (As citied by Kendall, 1986)

Warren H. & Verbrugge R.R. (1984), Auditory perception of breaking and bouncing events: A case study in ecological acoustics, *J. Exp. Psycol: Hum. Percept. Perform.* 10, 704-712

Waren R.M. (1970), *Auditory Perception: A New Analysis and Sunthesis*. Cambridge: Cambridge University Press

Waren R.M., Obusek C.J. & Ackroff J.M. (1972), Auditory induction: Perceptual synthesis of absent sounds. *Science*179, 1149-1151

Wold E., Blum T., Keislar D. & Wheaton J. (1996), Content-based classification, search, and retrieval of audio, *IEEE Multimedia* (Fall), 27-36

Martin Dana Keith (1999), Soynd – Source Recognition: A Theory and Computational Model, Ph.D. thesis, Massachusetts Institute of Technology

Καραγιάννης Γ., Σταϊνχάουερ Γ. (2001) *Μάθηση Μηχανών και Αναγνώριση Προτύπων*, Ε.Μ.Π.

Παραρτημα 1

Αποτελέσματα

Version 1

average_char(:,:,1) =

```
    0.8380   16.1000
```

average_char(:,:,2) =

```
    0.7966   16.1000
```

average_orig(:,:,1) =

```
  131.5000    1.9000    2.7000    4.9000  141.0000
    4.8000   39.3000    1.9000    4.0000   50.0000
   14.0000    1.7000   48.4000   15.9000   80.0000
    2.8000    3.7000    5.7000  111.8000  124.0000
```

average_orig(:,:,2) =

```
  156.0000    3.2000    2.6000    3.2000  165.0000
   10.0000   37.5000    5.7000    9.8000   63.0000
   17.2000    3.9000   44.5000   28.4000   94.0000
    8.0000    2.7000    6.8000  159.5000  177.0000
```

average_perc(:,:,1) =

```
   93.2624    1.3475    1.9149    3.4752
    9.6000   78.6000    3.8000    8.0000
   17.5000    2.1250   60.5000   19.8750
    2.2581    2.9839    4.5968   90.1613
```

average_perc(:,:,2) =

```
   94.5455    1.9394    1.5758    1.9394
   15.8730   59.5238    9.0476   15.5556
   18.2979    4.1489   47.3404   30.2128
    4.5198    1.5254    3.8418   90.1130
```

sampo_test_char(:,:,1) =

0.7896 16.0000

sampo_test_char(:,:,2) =

0.8116 14.0000

sampo_test_char(:,:,3) =

0.7936 16.0000

sampo_test_char(:,:,4) =

0.7896 14.0000

sampo_test_char(:,:,5) =

0.8036 18.0000

sampo_test_char(:,:,6) =

0.8036 16.0000

sampo_test_char(:,:,7) =

0.7916 16.0000

sampo_test_char(:,:,8) =

0.7916 16.0000

sampo_test_char(:,:,9) =

0.7856 18.0000

sampo_test_char(:,:,10) =

0.8056 17.0000

sampo_test_orig(:,:,1) =

158	2	2	3	165
9	36	9	9	63
19	3	44	28	94
9	4	8	156	177

sampo_test_orig(:,:,2) =

157	2	2	4	165
8	37	4	14	63
16	4	47	27	94
8	4	1	164	177

sampo_test_orig(:,:,3) =

158	3	3	1	165
10	38	6	9	63
18	4	46	26	94
9	3	11	154	177

sampo_test_orig(:,:,4) =

157	3	2	3	165
9	38	4	12	63
16	4	40	34	94
10	3	5	159	177

sampo_test_orig(:,:,5) =

156	3	3	3	165
9	37	6	11	63
15	4	46	29	94
7	1	7	162	177

sampo_test_orig(:,:,6) =

156	2	1	6	165
11	39	5	8	63
19	4	42	29	94
8	2	3	164	177

sampo_test_orig(:,:,7) =

153	4	4	4	165
10	38	5	10	63
18	4	42	30	94
6	3	6	162	177

sampo_test_orig(:,:,8) =

154	6	3	2	165
13	35	7	8	63
14	4	52	24	94
9	3	11	154	177

sampo_test_orig(:,:,9) =

155	4	3	3	165
12	37	7	7	63
18	4	42	30	94
7	2	10	158	177

sampo_test_orig(:,:,10) =

156	3	3	3	165
9	40	4	10	63
19	4	44	27	94
7	2	6	162	177

sampo_test_perc(:,:,1) =

95.7576	1.2121	1.2121	1.8182
14.2857	57.1429	14.2857	14.2857
20.2128	3.1915	46.8085	29.7872
5.0847	2.2599	4.5198	88.1356

sampo_test_perc(:,:,2) =

95.1515	1.2121	1.2121	2.4242
12.6984	58.7302	6.3492	22.2222
17.0213	4.2553	50.0000	28.7234
4.5198	2.2599	0.5650	92.6554

sampo_test_perc(:,:,3) =

95.7576	1.8182	1.8182	0.6061
15.8730	60.3175	9.5238	14.2857
19.1489	4.2553	48.9362	27.6596
5.0847	1.6949	6.2147	87.0056

sampo_test_perc(:,:,4) =

95.1515	1.8182	1.2121	1.8182
14.2857	60.3175	6.3492	19.0476
17.0213	4.2553	42.5532	36.1702
5.6497	1.6949	2.8249	89.8305

sampo_test_perc(:,:,5) =

94.5455	1.8182	1.8182	1.8182
14.2857	58.7302	9.5238	17.4603
15.9574	4.2553	48.9362	30.8511
3.9548	0.5650	3.9548	91.5254

sampo_test_perc(:,:,6) =

94.5455	1.2121	0.6061	3.6364
17.4603	61.9048	7.9365	12.6984
20.2128	4.2553	44.6809	30.8511
4.5198	1.1299	1.6949	92.6554

sampo_test_perc(:,:,7) =

92.7273	2.4242	2.4242	2.4242
15.8730	60.3175	7.9365	15.8730
19.1489	4.2553	44.6809	31.9149
3.3898	1.6949	3.3898	91.5254

sampo_test_perc(:,:,8) =

93.3333	3.6364	1.8182	1.2121
20.6349	55.5556	11.1111	12.6984
14.8936	4.2553	55.3191	25.5319
5.0847	1.6949	6.2147	87.0056

sampo_test_perc(:,:,9) =

93.9394	2.4242	1.8182	1.8182
19.0476	58.7302	11.1111	11.1111
19.1489	4.2553	44.6809	31.9149
3.9548	1.1299	5.6497	89.2655

sampo_test_perc(:,:,10) =

94.5455	1.8182	1.8182	1.8182
14.2857	63.4921	6.3492	15.8730
20.2128	4.2553	46.8085	28.7234
3.9548	1.1299	3.3898	91.5254

version 2

average_char(:,:,1) =

0.8375	15.6000

average_char(:,:,2) =

0.8557	15.6000

average_perc(:,:,1) =

93.5461	1.4184	2.0567	2.9787
10.4000	78.6000	4.4000	6.6000
17.3750	3.1250	61.2500	18.2500
2.1774	3.0645	5.5645	89.1935

average_perc(:,:,2) =

97.9394	0.8485	0.8485	0.3636
12.3810	74.6032	6.0317	6.9841
19.1489	0.6383	49.1489	31.0638
1.2429	0.4520	1.0169	97.2881

sampo_test_char(:,:,1) =

0.8437	14.0000

sampo_test_char(:,:,2) =

0.8557	15.0000

sampo_test_char(:,:,3) =

0.8517	16.0000

sampo_test_char(:,:,4) =

0.8437	18.0000

sampo_test_char(:,:,5) =

0.8597	15.0000

sampo_test_char(:,:,6) =

0.8397	16.0000

sampo_test_char(:,:,7) =

0.8677	15.0000

sampo_test_char(:,:,8) =

0.8597	17.0000

sampo_test_char(:,:,9) =

```
    0.8878   16.0000
```

sampo_test_char(:,:,10) =

```
    0.8477   14.0000
```

sampo_test_perc(:,:,1) =

```
   97.5758    1.2121    1.2121         0
    9.5238   77.7778    6.3492    6.3492
   27.6596         0   42.5532   29.7872
    1.1299    1.1299    1.1299   96.6102
```

sampo_test_perc(:,:,2) =

```
   98.7879         0    1.2121         0
    9.5238   77.7778    6.3492    6.3492
   17.0213    4.2553   44.6809   34.0426
    1.1299         0    1.1299   97.7401
```

sampo_test_perc(:,:,3) =

```
   98.7879    1.2121         0         0
   12.6984   68.2540    9.5238    9.5238
   21.2766    2.1277   48.9362   27.6596
    1.1299         0    1.1299   97.7401
```

sampo_test_perc(:,:,4) =

```
   95.1515    1.2121         0    3.6364
   19.0476   68.2540    6.3492    6.3492
   17.0213         0   51.0638   31.9149
    1.1299         0    1.1299   97.7401
```

sampo_test_perc(:,:,5) =

```
   98.7879         0    1.2121         0
   12.6984   74.6032    3.1746    9.5238
   17.0213         0   48.9362   34.0426
    1.1299         0    1.1299   97.7401
```

sampo_test_perc(:,:,6) =

```
   97.5758         0    2.4242         0
   15.8730   65.0794    6.3492   12.6984
   19.1489         0   46.8085   34.0426
    1.1299         0    1.1299   97.7401
```

sampo_test_perc(:,:,7) =

98.7879	1.2121	0	0
12.6984	77.7778	6.3492	3.1746
19.1489	0	53.1915	27.6596
1.1299	1.1299	1.1299	96.6102

sampo_test_perc(:,:,8) =

97.5758	1.2121	1.2121	0
12.6984	74.6032	9.5238	3.1746
17.0213	0	51.0638	31.9149
1.1299	0	1.1299	97.7401

sampo_test_perc(:,:,9) =

100.0000	0	0	0
9.5238	84.1270	3.1746	3.1746
17.0213	0	57.4468	25.5319
1.1299	1.1299	1.1299	96.6102

sampo_test_perc(:,:,10) =

96.3636	2.4242	1.2121	0
9.5238	77.7778	3.1746	9.5238
19.1489	0	46.8085	34.0426
2.2599	1.1299	0	96.6102

Version 3

average_char(:,:,1) =

0.8395	17.4000

average_char(:,:,2) =

0.8104	17.4000

average_perc(:,:,1) =

93.9007	1.2766	1.8440	2.9787
14.8000	75.2000	4.8000	5.2000
13.7500	3.2500	62.0000	21.0000

```
    2.5806    2.7419    4.3548   90.3226


average_perc(:,:,2) =

   96.8485    1.2121    0.3636    1.5758
   18.0952   63.8095    4.1270   13.9683
   21.2766    3.6170   41.9149   33.1915
    3.2768    1.3559    2.1469   93.2203


sampo_test_char(:,:,1) =

    0.7976   16.0000


sampo_test_char(:,:,2) =

    0.8136   20.0000


sampo_test_char(:,:,3) =

    0.8096   14.0000


sampo_test_char(:,:,4) =

    0.8096   20.0000


sampo_test_char(:,:,5) =

    0.8216   17.0000


sampo_test_perc(:,:,1) =

   95.1515    2.4242    0.6061    1.8182
   19.0476   61.9048    1.5873   17.4603
   21.2766    2.1277   39.3617   37.2340
    3.3898    1.1299    2.2599   93.2203


sampo_test_perc(:,:,2) =

   97.5758         0    0.6061    1.8182
   17.4603   58.7302    7.9365   15.8730
   19.1489    3.1915   45.7447   31.9149
    3.3898    1.1299    2.2599   93.2203


sampo_test_perc(:,:,3) =

   96.3636    1.8182         0    1.8182
```

17.4603	65.0794	6.3492	11.1111
22.3404	4.2553	39.3617	34.0426
2.8249	1.6949	1.1299	94.3503

sampo_test_perc(:,:,4) =

97.5758	0.6061	0	1.8182
19.0476	65.0794	3.1746	12.6984
23.4043	4.2553	39.3617	32.9787
3.9548	1.1299	1.6949	93.2203

sampo_test_perc(:,:,5) =

97.5758	1.2121	0.6061	0.6061
17.4603	68.2540	1.5873	12.6984
20.2128	4.2553	45.7447	29.7872
2.8249	1.6949	3.3898	92.0904

Κώδικας MATLAB

SAMPO V1

```
clear all;

load lyra_Training.mat;
load mpouzouki_Train.mat;
load laouto_Train.mat;
load oyti_Training.mat;
laouto_training=laouto_training(:,2:125);

load lyra_Test.mat;
load mpouzouki_Test.mat;
load laouto_Test.mat;
load oyti_Test.mat;
laouto_testing=laouto_testing(:,1:177);

load pitch.mat


            %construction of TRAIN MFCC INSTRUMENT matrix

%construction of lyra mfcc train matrix
lyra_mfcc=zeros(13,141);
lyra_freq=zeros(256,141);
   for(i=1:141)
      [lyra_mfcc(:,i),lyra_freq(:,i)]=mfcc(lyra_training(:,i),44100,2);
   end
C1(1,:)=centroid(lyra_freq,256,4);
C1(2,:)=energy_perc(lyra_freq,256,4,0.3);
C1(3,:)=energy_perc(lyra_freq,256,4,0.8);
%P1=pitchh(lyra_training);
lyra_mfcc=[lyra_mfcc ; C1 ;P1];

%construction of mpouzouki MFCC train matrix
mpouzouki_mfcc=zeros(13,50);
   for(i=1:50)
      [mpouzouki_mfcc(:,i),mpouzouki_freq(:,i)]=mfcc(mpouzouki_training(:,i),44100,2);
   end
C2(1,:)=centroid(mpouzouki_freq,256,4);
C2(2,:)=energy_perc(mpouzouki_freq,256,4,0.3);
C2(3,:)=energy_perc(mpouzouki_freq,256,4,0.8);
%P2=pitchh(mpouzouki_training);
mpouzouki_mfcc=[mpouzouki_mfcc ; C2 ;P2];

%construction of laouto MFCC train matrix
%laouto_mfcc=zeros(13,10);
   for(i=1:124)
      [laouto_mfcc(:,i),laouto_freq(:,i)]=mfcc(laouto_training(:,i),44100,2);
   end
C3(1,:)=centroid(laouto_freq,256,4);
C3(2,:)=energy_perc(laouto_freq,256,4,0.3);
C3(3,:)=energy_perc(laouto_freq,256,4,0.8);
%P3=pitchh(laouto_training);
laouto_mfcc=[laouto_mfcc ; C3 ;P3];

%construction of oyti mfcc train matrix
%oyti_mfcc=zeros(13,141);
```

```
%oyti_freq=zeros(256,141);
   for(i=1:80)
      [oyti_mfcc(:,i),oyti_freq(:,i)]=mfcc(oyti_training(:,i),44100,2);
   end
C4(1,:)=centroid(oyti_freq,256,4);
C4(2,:)=energy_perc(oyti_freq,256,4,0.3);
C4(3,:)=energy_perc(oyti_freq,256,4,0.8);
%P4=pitchh(oyti_training);
oyti_mfcc=[oyti_mfcc ; C4 ; P4];



        %construction of TEST MFCC INSTRUMENT MATRIX


%construction of lyra mfcc TEST matrix
   for(i=1:165)
      [lyra_mfcc_test(:,i),lyra_freq_test(:,i)]=mfcc(lyra_testing(:,i),44100,2);
   end
Ct1(1,:)=centroid(lyra_freq_test,256,4);
Ct1(2,:)=energy_perc(lyra_freq_test,256,4,0.3);
Ct1(3,:)=energy_perc(lyra_freq_test,256,4,0.8);
%Pt1=pitchh(lyra_testing);
lyra_mfcc_test=[lyra_mfcc_test ; Ct1 ;Pt1];

%construction of mpouzouki mfcc TEST matrix
mpouzouki_mfcc_test=zeros(13,63);
   for(i=1:63)

[mpouzouki_mfcc_test(:,i),mpouzouki_freq_test(:,i)]=mfcc(mpouzouki_testing(:,i),4410
0,2);
   end
Ct2(1,:)=centroid(mpouzouki_freq_test,256,4);
Ct2(2,:)=energy_perc(mpouzouki_freq_test,256,4,0.3);
Ct2(3,:)=energy_perc(mpouzouki_freq_test,256,4,0.8);
%Pt2=pitchh(mpouzouki_testing);
mpouzouki_mfcc_test=[mpouzouki_mfcc_test ; Ct2; Pt2];

%construction of laouto mfcc TEST matrix
%laouto_mfcc_test=zeros(13,14);
   for(i=1:177)
      [laouto_mfcc_test(:,i),laouto_freq_test(:,i)]=mfcc(laouto_testing(:,i),44100,2);
   end
Ct3(1,:)=centroid(laouto_freq_test,256,4);
Ct3(2,:)=energy_perc(laouto_freq_test,256,4,0.3);
Ct3(3,:)=energy_perc(laouto_freq_test,256,4,0.8);
%Pt3=pitchh(laouto_testing);
laouto_mfcc_test=[laouto_mfcc_test ; Ct3 ;Pt3];

%construction of oyti mfcc TEST matrix
   for(i=1:94)
      [oyti_mfcc_test(:,i),oyti_freq_test(:,i)]=mfcc(oyti_testing(:,i),44100,2);
   end
Ct4(1,:)=centroid(oyti_freq_test,256,4);
Ct4(2,:)=energy_perc(oyti_freq_test,256,4,0.3);
Ct4(3,:)=energy_perc(oyti_freq_test,256,4,0.8);
%Pt4=pitchh(oyti_testing);
oyti_mfcc_test=[oyti_mfcc_test ; Ct4 ;Pt4];
```

```
   %CONSTRUCTION OF TRAINSET

TrainSet(:,1:141)=lyra_mfcc(:,1:141);
TrainSet(:,142:191)=mpouzouki_mfcc(:,1:50);
TrainSet(:,192:271)=oyti_mfcc(:,1:80);
TrainSet(:,272:395)=laouto_mfcc(:,1:124);


   %CONSTRUCTION OF TRAINING (TARGETSET01)
for(i=1:141)
   TargetSet01(1,i)=1;
   TargetSet01(2,i)=0;
   TargetSet01(3,i)=0;
   TargetSet01(4,i)=0;
end
for(i=142:191)
   TargetSet01(1,i)=0;
   TargetSet01(2,i)=1;
   TargetSet01(3,i)=0;
   TargetSet01(4,i)=0;
end
for(i=192:271)
   TargetSet01(1,i)=0;
   TargetSet01(2,i)=0;
   TargetSet01(3,i)=1;
   TargetSet01(4,i)=0;
end
for(i=272:395)
   TargetSet01(1,i)=0;
   TargetSet01(2,i)=0;
   TargetSet01(3,i)=0;
   TargetSet01(4,i)=1;
end


   %CONSTRUCTION OF TESTSET

TestSet(:,1:165)=lyra_mfcc_test(:,1:165);
TestSet(:,166:228)=mpouzouki_mfcc_test(:,1:63);
TestSet(:,229:322)=oyti_mfcc_test(:,1:94);
TestSet(:,323:499)=laouto_mfcc_test(:,1:177);


   %CONSTRUCTION OF TESTING (TARGETSET02)
for(i=1:165)
   TargetSet02(1,i)=1;
   TargetSet02(2,i)=0;
   TargetSet02(3,i)=0;
   TargetSet02(4,i)=0;
end
for(i=166:228)
   TargetSet02(1,i)=0;
   TargetSet02(2,i)=1;
   TargetSet02(3,i)=0;
   TargetSet02(4,i)=0;
end
for(i=229:322)
   TargetSet02(1,i)=0;
   TargetSet02(2,i)=0;
   TargetSet02(3,i)=1;
```

```
   TargetSet02(4,i)=0;
end
for(i=323:499)
   TargetSet02(1,i)=0;
   TargetSet02(2,i)=0;
   TargetSet02(3,i)=0;
   TargetSet02(4,i)=1;
end


    %   C H A N G I N G   T R A I N S E T  WITH  T E S T S E T !!!!


%temp=TrainSet;
%TrainSet=TestSet;
%TestSet=temp;

%temp=TargetSet01;
%TargetSet01=TargetSet02;
%TargetSet02=temp;

TrainSet=TrainSet(1:15,:);
TestSet=TestSet(1:15,:);

save Sampo_Data.mat TrainSet TestSet TargetSet01 TargetSet02;
clear all;
```

```
SAMPO V2
clear all;

load lyra_Training.mat;
load mpouzouki_Train.mat;
load laouto_Train.mat;
load oyti_Training.mat;
laouto_training=laouto_training(:,2:125);

load lyra_Test.mat;
load mpouzouki_Test.mat;
load laouto_Test.mat;
load oyti_Test.mat;
laouto_testing=laouto_testing(:,1:177);

%load pitch.mat


            %construction of TRAIN MFCC INSTRUMENT matrix

%construction of lyra mfcc train matrix
lyra_mfcc=zeros(13,141);
lyra_freq=zeros(256,141);
   for(i=1:141)
      [lyra_mfcc(:,i),lyra_freq(:,i)]=mfcc(lyra_training(:,i),44100,2);
   end
C1(1,:)=centroid(lyra_freq,256,4);
C1(2,:)=energy_perc(lyra_freq,256,4,0.3);
```

```
C1(3,:)=energy_perc(lyra_freq,256,4,0.8);
P1=pitchh(lyra_training);
C1(1,:)=log(C1(1,:)./P1);
C1(2,:)=log(C1(2,:)./P1);
lyra_mfcc=[lyra_mfcc ; C1 ;P1];

%construction of mpouzouki MFCC train matrix
mpouzouki_mfcc=zeros(13,50);
   for(i=1:50)
      [mpouzouki_mfcc(:,i),mpouzouki_freq(:,i)]=mfcc(mpouzouki_training(:,i),44100,2);
   end
C2(1,:)=centroid(mpouzouki_freq,256,4);
C2(2,:)=energy_perc(mpouzouki_freq,256,4,0.3);
C2(3,:)=energy_perc(mpouzouki_freq,256,4,0.8);
P2=pitchh(mpouzouki_training);
C2(1,:)=log(C2(1,:)./P2);
C2(2,:)=log(C2(2,:)./P2);
mpouzouki_mfcc=[mpouzouki_mfcc ; C2 ;P2];

%construction of laouto MFCC train matrix
%laouto_mfcc=zeros(13,10);
   for(i=1:124)
      [laouto_mfcc(:,i),laouto_freq(:,i)]=mfcc(laouto_training(:,i),44100,2);
   end
C3(1,:)=centroid(laouto_freq,256,4);
C3(2,:)=energy_perc(laouto_freq,256,4,0.3);
C3(3,:)=energy_perc(laouto_freq,256,4,0.8);
P3=pitchh(laouto_training);
C3(1,:)=log(C3(1,:)./P3);
C3(2,:)=log(C3(2,:)./P3);
laouto_mfcc=[laouto_mfcc ; C3 ;P3];

%construction of oyti mfcc train matrix
%oyti_mfcc=zeros(13,141);
%oyti_freq=zeros(256,141);
   for(i=1:80)
      [oyti_mfcc(:,i),oyti_freq(:,i)]=mfcc(oyti_training(:,i),44100,2);
   end
C4(1,:)=centroid(oyti_freq,256,4);
C4(2,:)=energy_perc(oyti_freq,256,4,0.3);
C4(3,:)=energy_perc(oyti_freq,256,4,0.8);
P4=pitchh(oyti_training);
C4(1,:)=log(C4(1,:)./P4);
C4(2,:)=log(C4(2,:)./P4);
oyti_mfcc=[oyti_mfcc ; C4 ; P4];



        %construction of TEST MFCC INSTRUMENT MATRIX


%construction of lyra mfcc TEST matrix
   for(i=1:165)
      [lyra_mfcc_test(:,i),lyra_freq_test(:,i)]=mfcc(lyra_testing(:,i),44100,2);
   end
Ct1(1,:)=centroid(lyra_freq_test,256,4);
Ct1(2,:)=energy_perc(lyra_freq_test,256,4,0.3);
Ct1(3,:)=energy_perc(lyra_freq_test,256,4,0.8);
Pt1=pitchh(lyra_testing);
Ct1(1,:)=log(Ct1(1,:)./Pt1);
```

```
Ct1(2,:)=log(Ct1(2,:)./Pt1);
lyra_mfcc_test=[lyra_mfcc_test ; Ct1 ;Pt1];

%construction of mpouzouki mfcc TEST matrix
mpouzouki_mfcc_test=zeros(13,63);
   for(i=1:63)

[mpouzouki_mfcc_test(:,i),mpouzouki_freq_test(:,i)]=mfcc(mpouzouki_testing(:,i),4410
0,2);
   end
Ct2(1,:)=centroid(mpouzouki_freq_test,256,4);
Ct2(2,:)=energy_perc(mpouzouki_freq_test,256,4,0.3);
Ct2(3,:)=energy_perc(mpouzouki_freq_test,256,4,0.8);
Pt2=pitchh(mpouzouki_testing);
Ct2(1,:)=log(Ct2(1,:)./Pt2);
Ct2(2,:)=log(Ct2(2,:)./Pt2);
mpouzouki_mfcc_test=[mpouzouki_mfcc_test ; Ct2; Pt2];

%construction of laouto mfcc TEST matrix
%laouto_mfcc_test=zeros(13,14);
   for(i=1:177)
      [laouto_mfcc_test(:,i),laouto_freq_test(:,i)]=mfcc(laouto_testing(:,i),44100,2);
   end
Ct3(1,:)=centroid(laouto_freq_test,256,4);
Ct3(2,:)=energy_perc(laouto_freq_test,256,4,0.3);
Ct3(3,:)=energy_perc(laouto_freq_test,256,4,0.8);
Pt3=pitchh(laouto_testing);
Ct3(1,:)=log(Ct3(1,:)./Pt3);
Ct3(2,:)=log(Ct3(2,:)./Pt3);
laouto_mfcc_test=[laouto_mfcc_test ; Ct3 ;Pt3];

%construction of oyti mfcc TEST matrix
   for(i=1:94)
      [oyti_mfcc_test(:,i),oyti_freq_test(:,i)]=mfcc(oyti_testing(:,i),44100,2);
   end
Ct4(1,:)=centroid(oyti_freq_test,256,4);
Ct4(2,:)=energy_perc(oyti_freq_test,256,4,0.3);
Ct4(3,:)=energy_perc(oyti_freq_test,256,4,0.8);
Pt4=pitchh(oyti_testing);
Ct4(1,:)=log(Ct4(1,:)./Pt4);
Ct4(2,:)=log(Ct4(2,:)./Pt4);
oyti_mfcc_test=[oyti_mfcc_test ; Ct4 ;Pt4];


   %CONSTRUCTION OF TRAINSET

TrainSet(:,1:141)=lyra_mfcc(:,1:141);
TrainSet(:,142:191)=mpouzouki_mfcc(:,1:50);
TrainSet(:,192:271)=oyti_mfcc(:,1:80);
TrainSet(:,272:395)=laouto_mfcc(:,1:124);


   %CONSTRUCTION OF TRAINING (TARGETSET01)
for(i=1:141)
   TargetSet01(1,i)=1;
   TargetSet01(2,i)=0;
   TargetSet01(3,i)=0;
   TargetSet01(4,i)=0;
end
for(i=142:191)
```

```
   TargetSet01(1,i)=0;
   TargetSet01(2,i)=1;
   TargetSet01(3,i)=0;
   TargetSet01(4,i)=0;
end
for(i=192:271)
   TargetSet01(1,i)=0;
   TargetSet01(2,i)=0;
   TargetSet01(3,i)=1;
   TargetSet01(4,i)=0;
end
for(i=272:395)
   TargetSet01(1,i)=0;
   TargetSet01(2,i)=0;
   TargetSet01(3,i)=0;
   TargetSet01(4,i)=1;
end


   %CONSTRUCTION OF TESTSET

TestSet(:,1:165)=lyra_mfcc_test(:,1:165);
TestSet(:,166:228)=mpouzouki_mfcc_test(:,1:63);
TestSet(:,229:322)=oyti_mfcc_test(:,1:94);
TestSet(:,323:499)=laouto_mfcc_test(:,1:177);


   %CONSTRUCTION OF TESTING (TARGETSET02)
for(i=1:165)
   TargetSet02(1,i)=1;
   TargetSet02(2,i)=0;
   TargetSet02(3,i)=0;
   TargetSet02(4,i)=0;
end
for(i=166:228)
   TargetSet02(1,i)=0;
   TargetSet02(2,i)=1;
   TargetSet02(3,i)=0;
   TargetSet02(4,i)=0;
end
for(i=229:322)
   TargetSet02(1,i)=0;
   TargetSet02(2,i)=0;
   TargetSet02(3,i)=1;
   TargetSet02(4,i)=0;
end
for(i=323:499)
   TargetSet02(1,i)=0;
   TargetSet02(2,i)=0;
   TargetSet02(3,i)=0;
   TargetSet02(4,i)=1;
end


    %  C H A N G I N G   T R A I N S E T  WITH  T E S T S E T !!!!


%temp=TrainSet;
%TrainSet=TestSet;
%TestSet=temp;
```

```
%temp=TargetSet01;
%TargetSet01=TargetSet02;
%TargetSet02=temp;

TrainSet=TrainSet(1:15,:);
TestSet=TestSet(1:15,:);

save Sampo_Data3.mat TrainSet TestSet TargetSet01 TargetSet02;
clear all;
```

Παράρτημα 2

Αρχεία wav

CD: The greek volk instruments, FM records, vol 7 track 11, vol 10 track 15

Demo CD:Aherusia "Eros Aenaos"

Η λίστα της ηχογράφησης στα FebMan Home Studios υπάρχει στο CD.

Έπαιξαν οι μουσικοι:

Γιάννης Δημητρακόπουλος : Κρητική Λύρα, Μπουζούκι
Λήδα Μανιατάκου : Ούτι , Τουμπελέκι.