



Τμήμα Μηχανικών Αυτοματισμού
Αλεξάνδρειο Τεχνολογικό Εκπαιδευτικό Ίδρυμα Θεσ/νίκης



Αυτόματη ταξινόμηση βιοακουστικών σημάτων

Automated classification of bioacoustic signals

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

του Μιχαηλίδη Μανώλη

Επιβλέπουσα Καθηγήτρια: Δρ. Παπαδοπούλου Φωτεινή

Σίνδος, Ιούλιος 2016

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά την επιβλέπουσα καθηγήτρια Δρ. Παπαδοπούλου Φωτεινή, για την εμπιστοσύνη και υπομονή καθώς και για τη βοήθειά και καθοδήγησή της σε διάφορα ζητήματα κατά τη διάρκεια της υλοποίησης της εργασίας. Επίσης τους γονείς μου που με στήριξαν κατά τη διάρκεια της μόρφωσής μου.

Copyright © Μιχαηλίδης Μανώλης 2016.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό εκπαιδευτικό ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή της προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό χαρακτήρα πρέπει να απευθύνονται στον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Αλεξάνδρειου Τεχνολογικού Εκπαιδευτικού Ιδρύματος Θεσσαλονίκης. Για οποιαδήποτε χρήση του εμπεριεχομένου υλικού επικοινωνήστε με τον συγγραφέα στη διεύθυνση: manolis.michailidis@yahoo.com

Abstract

In modern era, many bird species face the danger of extinction. The need of study and maintaining the biodiversity, contributed in development of automated observation systems. In current work the effort is done in order to classify 3 different bird species of Greece.

The felicitous extraction of features require the correct processing of the data base. In this work, have been collected only specific parts of the recordings (trills) and with them the classification and recognition system is made. So for each recording firstly the noise level is reduced with the pre-emphasis filter where necessary and then using the Hilbert follower trills are obtained. Then, the fundamental frequency from each trill is computed via the autocorrelation. The features that are used are the average value of the fundamental frequency, the standard deviation, the bandwidth and the rapidness of changing the fundamental for each trill.

Several tests were made with different neural networks (specifically, with different number of nodes) and the results are discussed. Also two different approaches in experiments were done (using different number of inputs and outputs of the neural network). The results are very encouraging for future use and improvement of the system.

Περίληψη

Στη σύγχρονη εποχή, πολλά είδη πτηνών διατρέχουν κίνδυνο εξαφάνισης. Η ανάγκη της μελέτης και της διατήρησης της βιοποικιλότητας συνέβαλε στην ανάπτυξη αυτόματων συστημάτων εποπτείας. Στην παρούσα εργασία, γίνεται προσπάθεια αυτόματης αναγνώρισης τριών πτηνών της Ελλάδας μέσα από ηχητικές καταγραφές τους.

Η εύστοχη εξαγωγή χαρακτηριστικών από τα ηχητικά σήματα προϋποθέτει σωστή επεξεργασία της βάσης δεδομένων. Στην εργασία αυτή, έχουν εξαχθεί μόνο τα τμήματα της τρίλιας και με αυτά γίνεται η προσπάθεια δημιουργίας συστήματος ταξινόμησης και αναγνώρισης. Έτσι, για κάθε ηχογράφιση, μέσω του ακολουθητή Hilbert, συλλέχθηκαν οι τρίλιες, αφού πρώτα ακυρώθηκε ο θόρυβος μέσω φίλτρου προέμφασης, όπου κρίθηκε αναγκαίο. Ύστερα, υπολογίστηκε η θεμελιώδης συχνότητα από κάθε τρίλια με την χρήση της αυτοσυσχέτισης. Τα χαρακτηριστικά που χρησιμοποιήθηκαν για την ταξινόμηση είναι η μέση τιμή της θεμελιώδους συχνότητας, η τυπική απόκλιση, το εύρος τιμών της καθώς και η νευρικότητα (ταχύτητα εναλλαγής θεμελιώδους) στις τρίλιες.

Πραγματοποιήθηκαν δοκιμές σε νευρωνικά δίκτυα ποικίλης αρχιτεκτονικής (νευρωνικά δίκτυα με διαφορετικό αριθμό κόμβων) και παρουσιάζονται τα αποτελέσματα. Επίσης, εφαρμόστηκαν δύο ξεχωριστές προσεγγίσεις, με διαφορετική δομή του εκάστοτε δικτύου (δηλαδή αλλάζοντας τον αριθμό εισόδων και εξόδων του νευρωνικού δικτύου). Τα αποτελέσματα είναι αρκετά ενθαρρυντικά για τη μελλοντική χρήση και βελτίωση του συστήματος.

Πίνακας περιεχομένων

Abstract.....	4
Περίληψη	5
Πίνακας περιεχομένων	6
ΚΕΦΑΛΑΙΟ 1 : ΕΙΣΑΓΩΓΗ	9
1.1 Βιοακουστική	9
1.2 Ήχοι των πτηνών	10
1.3 Κίνδυνοι εξαφάνισης ειδών και λύσεις για την αντιμετώπιση τους	13
1.3.1 Προβλήματα και κίνδυνοι.....	13
1.3.2 Μέτρα αντιμετώπισης προβλήματος	14
1.4 Συμβολή των πτηνών στην ανθρωπότητα.....	14
1.5 Σκοπός της εργασίας.....	15
1.6 Δομή της εργασίας.....	16
ΚΕΦΑΛΑΙΟ 2 : ΤΡΟΠΟΙ ΑΝΙΧΝΕΥΣΗΣ ΚΑΙ ΤΑΞΙΝΟΜΗΣΗΣ ΒΙΟΑΚΟΥΣΤΙΚΟΥ ΣΗΜΑΤΟΣ. 18	
2.1 Εξαγωγή συλλαβών με την ενέργεια και κατηγοριοποίηση μέσω ιστογράμματος των DTW αποκλίσεων.....	18
2.2 Εξαγωγή συλλαβών με ακολουθητή HILBERT και ταξινόμηση μέσω Κρυμμένων Μαρκοβιανών Μοντέλων (HMMs)	20
2.3 Άλλοι τρόποι κατηγοριοποίησης ήχων πτηνών	22

ΚΕΦΑΛΑΙΟ 3 : ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΗΝ ΕΠΕΞΕΡΓΑΣΙΑ ΤΟΥ ΑΚΟΥΣΤΙΚΟΥ ΣΗΜΑΤΟΣ	23
3.1 Συντελεστές MFCCs (Mel Frequency Cepstral Coefficients)	23
3.2 Delta – Παράγωγος 1 ^{ης} Τάξης (Time Derivative) / Delta-Deltas – Παράγωγος 2 ^{ης} Τάξης (Acceleration Derivative)	28
3.3 Cepstrum/Liftering.....	29
3.4 Προέμφαση πρώτου βαθμού (First order pre-emphasis)	33
3.5 Συσχέτιση - Αυτοσυσχέτιση (Correlation - Autocorrelation).....	34
3.6 Μετασχηματισμός Fourier Διακριτού Χρόνου-DTFT (Discrete Time Fourier Transform)/Σύντομος Μετασχηματισμός Fourier- STFT (Short Time Fourier Transform)	36
3.7 Παράθυρα/παραθύρωση (windowing)	38
3.8 Συνέλιξη (Convolution)	38
3.9 Κινούμενος Μέσος Όρος (Moving Average)	39
3.10 Μετασχηματισμός Hilbert/Hilbert Φάκελος(Envelope) – Αναλυτικό σήμα....	40
3.11 Νευρωνικά δίκτυα – Back-propagation Perceptron	41
ΚΕΦΑΛΑΙΟ 4 : ΤΑΞΙΝΟΜΗΣΗ ΚΕΛΑΪΔΙΣΜΑΤΩΝ ΜΕ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	46
4.1 Βάση δεδομένων.....	46
4.2 Προεργασία και τμηματοποίηση	47
4.3 Τμηματοποίηση Hilbert	49
4.4 Βασική/θεμελιώδης συχνότητα (Fundamental frequency).....	49
4.5 Εξαγωγή χαρακτηριστικών	52
4.6 Εκπαίδευση νευρωνικού δικτύου.....	56

4.7 Πειραματική διαδικασία.....	57
4.7.1 Πρώτη πειραματική προσέγγιση	57
4.7.2 Δεύτερη πειραματική προσέγγιση	58
4.8 Συμπεράσματα – Προτάσεις μελλοντικής βελτίωσης	63
ΠΑΡΑΡΤΗΜΑ Α΄	65
ΠΑΡΑΡΤΗΜΑ Β΄	75
Βιβλιογραφία	80

ΚΕΦΑΛΑΙΟ 1 : ΕΙΣΑΓΩΓΗ

1.1 Βιοακουστική

Η βιοακουστική είναι ένας διεπιστημονικός κλάδος που αποτελείται από αυτούς της βιολογίας και της ακουστικής. Άλλωστε, αυτό δηλώνει κι η ετυμολογία της λέξης (βίος + ακουστική). Ασχολείται με τον εντοπισμό, την καταγραφή, ταξινόμηση και τη συλλογή των ήχων που παράγουν και αντιλαμβάνονται τα είδη [1]. Διαθέτει εξειδικευμένες υποκατηγορίες όπου γίνεται η μελέτη μόνο των θαλάσσιων ή μόνο των αμφίβιων ειδών. Βέβαια, υφίστανται και άλλες κατηγορίες, όμως δεν θα γίνει περαιτέρω επέκταση, καθώς η παρούσα εργασία ασχολείται αποκλειστικά με πτηνά.

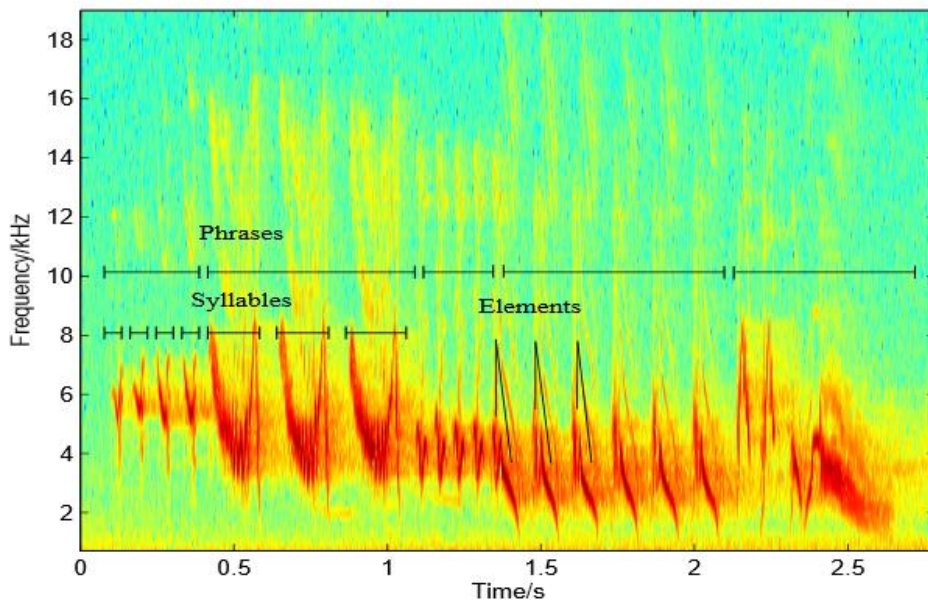
Σε γενικές γραμμές, η βιοακουστική μελετά από νευροβιολογικής και ανατομικής πλευράς, την παραγωγή και τη μετάδοση των ήχων στο περιβάλλον καθώς και την επίδραση τους στα υπόλοιπα είδη (επιστήμη της νευροφυσιολογίας)[2]. Έχει διαπιστωθεί ότι υπήρξε εξέλιξη στους μηχανισμούς παραγωγής ήχων των πτηνών, δηλαδή προσαρμογή της έντασης και της συχνότητας τους. Με αυτούς τους ήχους πραγματοποιείται η επικοινωνία μεταξύ των πτηνών καθώς εκφράζουν τη συναισθηματική τους κατάσταση είτε προειδοποιούν για κινδύνους. Παρόλο που έχει αναπτυχθεί ιδιαίτερα ο κλάδος και οι μελετητές έχουν μάθει αρκετά για την βιοποικιλότητα, ωστόσο ελάχιστα είναι γνωστά από νευροβιολογικής διάστασης που επηρεάζουν την παραγωγή και τη μετάδοση των ήχων τους.

Πρακτικά, πρώτα συλλέγονται οι ήχοι με ειδικές συσκευές καταγραφής (sound recorders), κι έπειτα, με την αξιοποίηση εξειδικευμένων προγραμμάτων, γίνεται ο διαχωρισμός και η μελέτη των καταγραφών. Γενικότερα, μέσω της βιοακουστικής, με τη μελέτη των λειτουργιών και της συμπεριφοράς των ειδών, επιτυγχάνεται η καλύτερη κατανόηση τους.

Η κλασική μέθοδος εξακολουθεί να επιτελεί τον εντοπισμό των ειδών μέσω της ακοής κι όπως γίνεται αντιληπτό, είναι αρκετά δύσκολη, επειδή απαιτεί ειδικούς εμπειρογνώμονες του αντικείμενου, με αρκετά «έμπειρη» ακοή. Μάλιστα πολλές φορές δημιουργείται διχογνωμία μεταξύ τους καθώς οι ηχογραφήσεις ενδέχεται να συνοδεύονται κι από θόρυβο κι έτσι ο εντοπισμός μόνο μέσω της ακουστικής κρίνεται ανεπαρκής [3]. Η ανάπτυξη του κλάδου των Η/Υ και η χρήση των νευρωνικών δικτύων όπως παρουσιάζεται και στα κεφάλαια που ακολουθούν, οδήγησε σε μεθόδους ταξινόμησης και κατηγοριοποίησης με απουσία του ανθρώπινου παράγοντα.

1.2 Ήχοι των πτηνών

Τα πτηνά παράγουν ήχους ποικίλων συχνοτήτων και εντάσεων. Τα τραγούδια τους διαφοροποιούνται ανάλογα με την ηλικία, το φύλο και τον βαθμό μίμησης από άλλα ενήλικα κυρίως υποκείμενα. Ακόμη έχουν παρατηρηθεί διαφορές σε τραγούδια μεταξύ του ίδιου είδους στα αστικά κέντρα και στα δάση, λόγω της επίδρασης της ηχορρύπανσης. Αυτό συμβαίνει γιατί προσαρμόζονται στις συνθήκες του περιβάλλοντος, μεταβάλλοντας την ένταση και τη συχνότητα των τραγουδιών τους [4]. Σε μια ηχογράφιση μερικών δευτερολέπτων δύναται να υπάρχουν συλλαβές, φράσεις/μοτίβα, στοιχεία, τρίλιες. Συλλαβές είναι τα πιο βασικά κομμάτια του τραγουδιού, στοιχεία είναι κάποια τυχαία συμβάντα που δεν συμβαδίζουν με όλα τα προηγούμενα και φράσεις είναι συστάδες συλλαβών που αποτελούν ένα μοτίβο [5]. Στο παρακάτω σχήμα 1.1, απεικονίζεται το φασματικό περιεχόμενο στο πεδίο χρόνου-συχνότητας (STFT) μιας ηχογράφησης και παρουσιάζονται όλα τα προαναφερθέντα ενώ το σχήμα 1.2 επεξηγεί τον χρωματικό κώδικα. Όσο πιο έντονο είναι το κόκκινο χρώμα, τόσο πιο μεγάλη η συγκέντρωση της ενέργειας στο τμήμα αυτό.

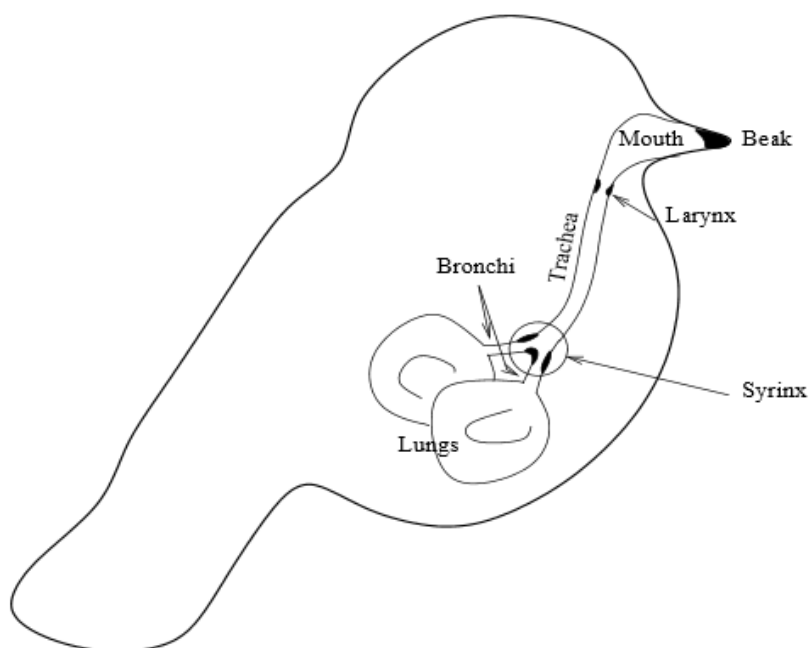


Σχήμα 1.1 STFT ηχογράφησης όπου ο οριζόντιος άξονας είναι χρόνος σε δευτερόλεπτα και ο κατακόρυφος η συχνότητα σε kHz. Παρουσιάζονται συλλαβές, στοιχεία και φράσεις [6].

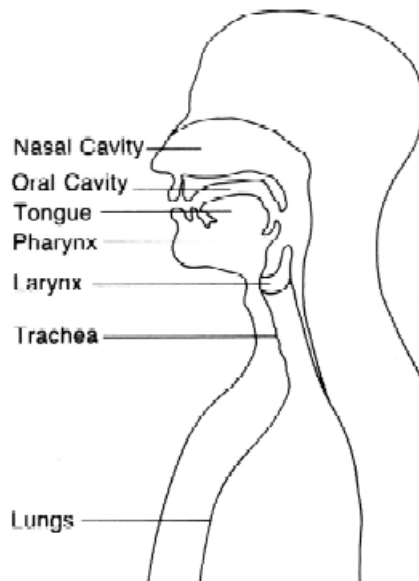


Σχήμα 1.2 Ενδεικτική μπάρα απεικόνισης χρωματικής έντασης της ενέργειας σε λογαριθμική κλίμακα.

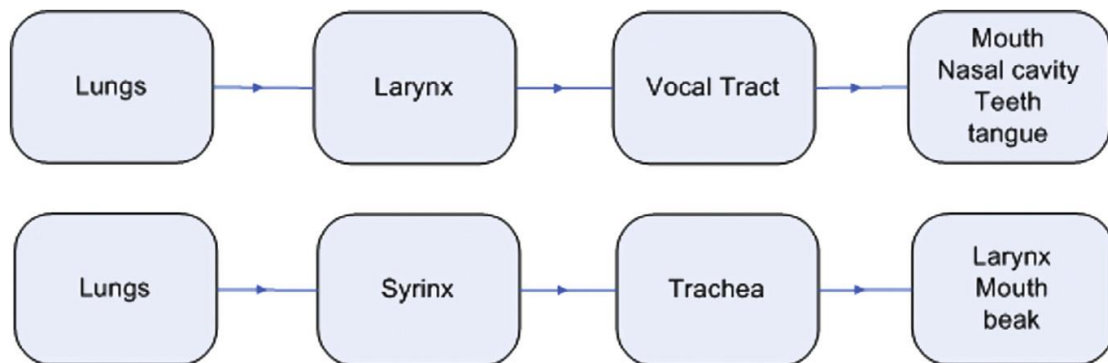
Στο παρακάτω σχήμα (σχ. 1.3), απεικονίζεται το σύστημα παραγωγής ήχου στα πτηνά [6] σε αντιπαραβολή με το σχήμα 1.4, όπου παρουσιάζεται το σύστημα με τα όργανα παραγωγής ήχου στον άνθρωπο [7]. Στο σχήμα που ακολουθεί (σχ. 1.5), δίνεται το συγκριτικό διάγραμμα παραγωγής ήχου σε άνθρωπο και πτηνό [8].



Σχήμα 1.3 Όργανα που συμβάλλουν στην διαδικασία παραγωγής ήχου στα πτηνά. Όπου lungs οι πνεύμονες, bronchi, οι βρόγχοι, syrinx, το σύριγξ, trachea, η τραχεία οδός, larynx, ο λάρυγγας, mouth, το στόμα και beak, το ράμφος [6].



Σχήμα 1.4 Όργανα που συμβάλλουν στην παραγωγή ήχου στον άνθρωπο. Lungs είναι οι πνεύμονες, trachea η τραχεία οδός, larynx ο λάρυγγας, pharynx ο φάρυγγας, tongue η γλώσσα, oral cavity η στοματική κοιλότητα και nasal cavity η ρινική κοιλότητα [7].



Σχήμα 1.5 Διάγραμμα ροής παραγωγής ήχου σε άνθρωπο και πτηνό. Στον άνθρωπο ξεκινάει από τους πνεύμονες μέσω του λάρυγγα συνεχίζει στην ηχητική οδό και καταλήγει στο στόμα (γλώσσα, δόντια, ρινική οδός). Στο πτηνό ξεκινάει από τους πνεύμονες συνεχίζει στην τραχεία μέσω του σύριγξ και καταλήγει στο ράμφος (λάρυγγας, στόμα) [8].

Στον άνθρωπο, η διαδικασία ξεκινάει από τους πνεύμονες, συνεχίζει στον λάρυγγα κι έπειτα, μέσω της φωνητικής οδού, καταλήγει στο στόμα. Εν τέλει, εκεί, μέσω της ρινικής οδού, των δοντιών και της γλώσσας παράγεται ο ήχος. Στα πτηνά, η διαδικασία παρουσιάζει κάποιες διαφορές. Ξεκινάει από τους πνεύμονες συνεχίζει στην τραχεία, αφού πρώτα περάσει μέσω του οργάνου που ονομάζεται σύριγξ (syrinx). Στο τέλος, με την χρήση του λάρυγγα και του ράμφους παράγεται ο ήχος. Οι διαφορές που διαπιστώνονται είναι ότι ενώ ο άνθρωπος διαθέτει λάρυγγα, το πτηνό ως αντίστοιχο του έχει το syrinx, το όργανο που επιτρέπει στα πτηνά την παραγωγή ήχων χωρίς τη χρήση φωνητικών χορδών όπως συμβαίνει στον άνθρωπο. Το όργανο

αυτό επιτρέπει στα πτηνά την παραγωγή πολλαπλών συχνοτήτων με ραγδαίες μεταβολές σε χιλιοστά του δευτερολέπτου [9], [6].

1.3 Κίνδυνοι εξαφάνισης ειδών και λύσεις για την αντιμετώπισή τους

Στην παράγραφο αυτή σημειώνονται οι σύγχρονοι κίνδυνοι εξαφάνισης των πτηνών λόγω της ανθρώπινης παρέμβασης και παρατίθενται μερικές προτάσεις για την προστασία της βιοποικιλότητας.

1.3.1 Προβλήματα και κίνδυνοι

Η παρέμβαση του ανθρώπου στη φύση είχε άμεσες επιπτώσεις στη βιοποικιλότητα. Οι συνέπειες άρχισαν να φαίνονται ξεκάθαρα από τη δεκαετία του '70, όπου και ξεκίνησαν και οι πρώτες προσπάθειες περί ευαισθητοποίησης πολιτών μέσω των ΜΜΕ. Άλλοτε με το κυνήγι, που σε αρκετές περιπτώσεις είχε λαθροθηρικό χαρακτήρα και άλλες φορές με την απερίσκεπτη αποδάσωση περιοχών, ο άνθρωπος δημιούργησε σοβαρά προβλήματα στα πτηνά. Η δημιουργία μεγαλουπόλεων εκεί που κάποτε είχαν το καταφύγιο τους, σε συνδυασμό με την κατασκευή αυτοκινητοδρόμων, που στην ουσία διαχώριζαν τα φυσικά τους εδάφη, επηρέασαν τις ζωές τους σημαντικά. Όπως επίσης η βιομηχανοποίηση, όπου με τα εργοστάσια και τα αυτοκίνητα που με τη σειρά τους, ευθύνονται όχι μόνο για την μόλυνση του περιβάλλοντος με τα χημικά αέρια, αλλά και για την ηχορρύπανση. Όλα τα παραπάνω έχουν δυσκολέψει σε μεγάλο βαθμό τα είδη στην επικοινωνία μεταξύ τους και γενικότερα στην καθημερινότητά τους.

Όμως ίσως τον βασικότερο κίνδυνο για τον αφανισμό κάποιων ειδών αποτελεί, η κλιματική αλλαγή που λαμβάνει χώρα τις τελευταίες δεκαετίες λόγω της ανθρώπινης παρέμβασης [3], [10]. Η μείωση του πληθυσμού πολλών ειδών συνδέεται άμεσα με τον άνθρωπο και συνεπάγεται την υποβάθμιση του φυσικού περιβάλλοντος και της ποιότητας ζωής μας. Στην ιστοσελίδα της ελληνικής ορνιθολογικής εταιρείας διατυπώνεται χαρακτηριστικά: **«Η προσπάθεια για τη διατήρηση των πουλιών είναι και αγώνας για τη διατήρηση του περιβάλλοντος συνολικότερα. Γι' αυτό, ένας καλύτερος κόσμος για τα πουλιά, σημαίνει έναν καλύτερο κόσμο για όλους μας».**

Όπως αναφέρεται στο [11], οι ανθρωπογενείς ήχοι επηρεάζουν τα πουλιά των αστικών κέντρων με την αυξημένη χρήση των αυτοκινήτων, αλλά και με τα εργοστάσια που εκπέμπουν θόρυβους. Αυτοί οι θόρυβοι επιδρούν στα είδη δημιουργώντας δυσκολίες στην επικοινωνία μεταξύ τους. Έτσι τα πουλιά των αστικών κέντρων είναι ικανά να προσαρμόζονται τα κελαϊδίσματα τους σε ένταση και συχνότητα. Όπως έχει παρατηρηθεί τα πουλιά των πόλεων τραγουδούν με

υψηλότερες εντάσεις και συχνότητες από τα αντίστοιχα της υπαίθρου [4]. Εκτός όμως από τον θόρυβο, τα πουλιά διατρέχουν επίσης και τον κίνδυνο των τοξικών αερίων καθώς επηρεάζεται η εγκεφαλική ανάπτυξή τους με αποτέλεσμα να δημιουργούνται προβλήματα υγείας και επικοινωνίας [10]. Τέλος, υφίστανται δυσκολίες αποπροσανατολισμού που αντιμετωπίζουν με τη φωτορύπανση κυρίως των αστικών κέντρων. Γενικό συμπέρασμα είναι ότι με τις μεταβολές που γίνονται στο περιβάλλον, τα πουλιά αλλάζουν τις συνήθειες και τα τραγούδια τους κάθε φορά, προσαρμόζοντάς τα, αντίστοιχα, με βάση τις αλλαγές και τις ανάγκες τους.

1.3.2 Μέτρα αντιμετώπισης προβλήματος

Η ταξινόμηση και η εποπτεία της βιοποικιλότητας φαίνεται να μπορεί να συνεισφέρει σημαντικά στον κλάδο της βιολογίας (ορνιθολογίας). Αυτή θα οδηγήσει στην ευκολότερη αναγνώριση και καλύτερη κατανόηση των τρόπων επικοινωνίας και συμπεριφοράς μεταξύ των πτηνών. Σε αρκετές περιπτώσεις η σοβαρή έλλειψη δεδομένων και γνώσεων συμβάλλει στην εξαφάνιση ειδών σε συνδυασμό με την αδιαφορία, όπως έχει αποδειχθεί. Έτσι απαιτείται η συλλογή δεδομένων από ποικίλα τοπία, κατά διαφορετικά χρονικά διαστήματα, ώστε να είναι διαθέσιμο επαρκές υλικό σε διάφορες χρονικές περιόδους και να συμβάλλει στις προσπάθειες που πραγματοποιούνται υπέρ της διατήρησης της βιοποικιλότητας [12].

Με την ενημέρωση των κινδύνων τόσο για τον άνθρωπο όσο και για την βιοποικιλότητα μέσω των ΜΜΕ, επιτυγχάνεται η ευαισθητοποίηση των πολιτών σε θέματα διατήρησης και προστασίας της βιοποικιλότητας. Μη κερδοσκοπικοί οργανισμοί και ομάδες συμβάλλουν στο έργο σε σημαντικό βαθμό ασκώντας πίεση στις κυβερνήσεις ώστε να λάβουν μέτρα επί του ζητήματος. Ως αποτέλεσμα, ήδη δημιουργήθηκαν εθνικά πάρκα και θεσπίστηκαν νόμοι περί της προστασίας τους.

Εκτός αυτού, ο αριθμός των εθελοντικών ομάδων και διάφορων μη κερδοσκοπικών οργανώσεων έχει αυξηθεί σημαντικά και αποτελείται κυρίως από λάτρεις και παρατηρητές των πουλιών που είτε πραγματοποιούν δωρεές είτε καταγράφουν και παρατηρούν τα είδη. Παράλληλα, έχει αυξηθεί και ο αριθμός των άρθρων στα τελευταία χρόνια που ασχολούνται με την διατήρηση της βιοποικιλότητας [10].

1.4 Συμβολή των πτηνών στην ανθρωπότητα

Τα πουλιά με τους ήχους τους είχαν και συνεχίζουν ακόμα να έχουν, σημαντική θέση στην ανθρωπότητα και στον πολιτισμό μας. Εκτός από την καλαισθησία και την

ομορφιά τους, με τα τραγούδια τους έχουν αποτελέσει σημαντική πηγή έμπνευσης σε διάφορους καλλιτέχνες όπως μουσικούς, συνθέτες και ποιητές [9]. Οι όμορφες μελωδίες τους έχουν ηρεμιστικές ιδιότητες στον άνθρωπο πολλές φορές, επηρεάζοντας την νευροψυχολογική διάθεση του ατόμου.

Σε αρκετές χώρες στο πέρασμα των αιώνων δημιουργήθηκαν μύθοι και θρύλοι για τα πουλιά. Σε μια πιο ακραία περίπτωση η θεοποίηση και η αντίληψη ότι κάποια ανώτερη πνευματικά οντότητα φανερώνεται στους ανθρώπους με την μορφή τους. Τα πιο χαρακτηριστικά παραδείγματα είναι οι θεότητες της αρχαίας Αιγύπτου, όπως επίσης και διάφορες θεότητες στους ιθαγενείς λαούς της Αμερικής. Δεν θα μπορούσαμε να μην αναφέρουμε ότι το περιστέρι αποτελεί το σύμβολο της ειρήνης σε όλη την οικουμένη.

1.5 Σκοπός της εργασίας

Στην εποχή μας οι έρευνες για την αυτόματη αναγνώριση των ειδών μέσω της βιοακουστικής επεκτείνονται συνεχώς. Όλο και περισσότεροι μελετητές και επιστήμονες ασχολούνται με το αντικείμενο αυτό, κυρίως, για ερευνητικούς λόγους, γι' αυτό κρίνεται αναγκαία η δημιουργία ενός ενιαίου εγχειριδίου σχετικού με την αυτόματη αναγνώριση και ταξινόμηση των ποικίλων ειδών [11], ακόμη ίσως και για τον εντοπισμό νέων ειδών που πιθανόν να είναι, μέχρι στιγμής, άγνωστα. Η προσπάθεια γίνεται για τη δημιουργία ενός αυτόματου συστήματος που θα κατηγοριοποιεί και θα εποπτεύει, καθώς η προστασία όσων ειδών απειλούνται κρίνεται αναγκαία για τους λόγους που αναφέρθηκαν παραπάνω.

Ωστόσο, οι μέθοδοι αναγνώρισης που αναπτύχθηκαν είναι σημαντικά χρονοβόρες και βασίζονται αποκλειστικά στις γνώσεις των ειδικών. Έτσι δυο από τα βασικότερα μειονεκτήματα τους είναι α) η αναγκαία παρουσία των εμπειρογνομόνων και β) η διάσταση στις γνώμες μεταξύ των ειδικών, καθώς η αντίληψη είναι σχετική, δηλαδή, διαφέρει από άνθρωπο σε άνθρωπο. Επίσης ο παράγοντας της μίμησης που αναφέρθηκε παραπάνω, στην παράγραφο 1.2, θέτει περισσότερους κινδύνους λάθους ταξινόμησης μέσω της ακουστικής. Τέλος, ο θόρυβος στις καταγραφές, είτε προέρχεται από άλλα είδη ή στοιχεία της φύσης είτε είναι ανθρωπογενής, δημιουργεί επιπλέον δυσκολίες.

Η ανάπτυξη του κλάδου των Η/Υ, συνέβαλε στην ανάγκη για τη δημιουργία συστημάτων αυτόματης αναγνώρισης και ταξινόμησης της βιοποικιλότητας, χωρίς την παρουσία και την επέμβαση του ανθρώπινου παράγοντα, ακόμη και σε περιβάλλοντα δύσκολης πρόσβασης για τον άνθρωπο [11]. Η παρούσα εργασία κινείται στην παραπάνω κατεύθυνση προτείνοντας ένα σύστημα αυτόματης

ταξινόμησης των βιοακουστικών σημάτων που προέρχονται από τρία είδη πουλιών, χαρακτηριστικών στην ελληνική πανίδα.

1.6 Δομή της εργασίας

Η εργασία αυτή διαρθρώνεται σε τέσσερα κεφάλαια.

Το πρώτο κεφάλαιο αποτελεί την εισαγωγή, όπου παρουσιάζονται οι βασικές έννοιες, διατυπώνεται ο σκοπός της εργασίας και περιγράφεται η δομή της. Ιδιαίτερα, αναφέρεται στο πεδίο της βιοακουστικής όπου περιγράφεται, ειδικά, το σύστημα παραγωγής ήχων των πτηνών (σε σύγκριση με αυτό του ανθρώπου) ενώ παράλληλα, παρουσιάζονται οι άμεσοι ή έμμεσοι κίνδυνοι που αντιμετωπίζει η βιοποικιλότητα καθώς και βασικοί λόγοι διατήρησής της.

Στο δεύτερο κεφάλαιο, παρουσιάζονται, επιλεκτικά, κάποιες πρόσφατες εργασίες, σχετικές με τους τρόπους ανίχνευσης και ταξινόμησης ακουστικού σήματος προερχόμενου από πουλιά και αναλύονται οι μέθοδοι στις οποίες αναφέρονται. Ειδικότερα, παρουσιάζεται η εξαγωγή συλλαβών με ενεργειακό κατώφλι και η κατηγοριοποίηση ιστογράμματος των DTW αποκλίσεων. Επιπλέον, περιγράφεται η εξαγωγή συλλαβών με ακολουθητή Hilbert και η ταξινόμηση μέσω Κρυμμένων Μαρκοβιανών Μοντέλων (HMMs). Τέλος, αναφέρονται μερικά πρόσφατα άρθρα με ποικίλους τρόπους εξαγωγής χαρακτηριστικών και κατηγοριοποίησης των πτηνών (οι πηγές δίνονται στην βιβλιογραφία).

Το τρίτο κεφάλαιο αφιερώνεται στα ιδιαίτερα εργαλεία που αναφέρονται ή προτείνονται για την υλοποίηση της παρούσας εργασίας, περιέχει, δηλαδή, το μαθηματικό υπόβαθρο τόσο για την κατανόηση των μεθόδων που περιγράφονται στο προηγούμενο κεφάλαιο, όσο και για την πειραματική διαδικασία που ακολουθεί. Συγκεκριμένα, παρουσιάζονται οι MFCCs, Delta και οι Delta-Deltas συντελεστές, όπως επίσης και η διαδικασία του cepstrum liftering, λόγω της ευρέως χρήσης τόσο στην αναγνώριση ομιλίας όσο και των πτηνών. Περιγράφεται το φίλτρο της προέμφασης, η συσχέτιση και η αυτοσυσχέτιση, ο DTFT όπως επίσης και ο STFT. Επίσης παρουσιάζεται η έννοια της παραθύρωσης, επεξηγείται η συνέλιξη και ο κινούμενος μέσος όρος και ο μετασχηματισμός Hilbert. Τέλος, γίνεται εισαγωγή στα νευρωνικά δίκτυα και εστιάζεται στα back-propagation δίκτυα.

Στο τέταρτο κεφάλαιο, περιγράφεται η διαδικασία που ακολουθήθηκε στο πειραματικό μέρος της εργασίας και παρουσιάζονται τα αποτελέσματά της. Συγκεκριμένα, αναφέρονται οι προσεγγίσεις που ακολουθήθηκαν, ξεκινώντας από την εξεύρεση της βάσης δεδομένων (τράπεζα βιοακουστικών σημάτων), την προετοιμασία των σημάτων, την εξαγωγή χαρακτηριστικών από αυτά και την

εκπαίδευση σχετικού ταξινομητή (νευρωνικό δίκτυο) έως τη δοκιμή σε «άγνωστα» σήματα και την εκτίμηση των αποτελεσμάτων. Τέλος, διατυπώνονται τα συμπεράσματα της εργασίας και προτείνονται πιθανές μελλοντικές επεκτάσεις της.

Η παρούσα εργασία έχει πραγματοποιηθεί σε περιβάλλον Matlab, καθώς αυτή η γλώσσα προγραμματισμού έχει βελτιωθεί αρκετά και διαθέτει ποικίλα εργαλεία με έτοιμες συναρτήσεις που διευκολύνουν σε μεγάλο βαθμό τον προγραμματισμό. Μειώνεται η πολυπλοκότητα των αλγορίθμων και του χρόνου υπολογισμών, έτσι δεν απαιτείται υπολογιστής μεγάλων επιδόσεων και υψηλού κόστους. Επιπλέον, διαθέτει έτοιμα εργαλεία και συναρτήσεις νευρωνικών δικτύων για το στάδιο της κατηγοριοποίησης.

ΚΕΦΑΛΑΙΟ 2 : ΤΡΟΠΟΙ ΑΝΙΧΝΕΥΣΗΣ ΚΑΙ ΤΑΞΙΝΟΜΗΣΗΣ ΒΙΟΑΚΟΥΣΤΙΚΟΥ ΣΗΜΑΤΟΣ

Την τελευταία δεκαετία, παρατηρείται πρόοδος στον τομέα, καθώς όλο και περισσότεροι επιστήμονες ασχολούνται με την αυτόματη κατηγοριοποίηση των ήχων των πτηνών αλλά και άλλων ειδών. Έχουν αναπτυχθεί ποικίλες μέθοδοι προτείνοντας διάφορους αλγόριθμους ενώ παράλληλα υπήρξε σημαντική εξέλιξη σε προγραμματιστικό επίπεδο.

2.1 Εξαγωγή συλλαβών με την ενέργεια και κατηγοριοποίηση μέσω ιστογράμματος των DTW αποκλίσεων

Με τη μέθοδο αυτή, πρώτα εξάγονται οι συλλαβές με βάση την ενέργεια. Ως συλλαβή θεωρείται το κομμάτι που πληροί ένα ορισμένο ενεργειακό κατώφλι. Έτσι κατ' αρχήν, γίνεται ανάλυση βραχέως χρόνου (πεδίο χρόνου-συχνότητας) και υπολογίζεται ο σύντομος μετασχηματισμός Fourier STFT των σημάτων. Συγκεκριμένα, στην εργασία [9] χρησιμοποιήθηκε παράθυρο Kaiser ($\alpha = 8$) με μήκος 256 δείγματα. Το μέγεθος του FFT επιλέχθηκε στα 1024 δείγματα και με 75% επικάλυψη για συχνότητα δειγματοληψίας 44.1KHz. Μια αξιόπιστη τιμή στο κατώφλι θεωρήθηκαν τα 30dB. Σε ένα άλλο άρθρο, επιλέχθηκε το κατώφλι στα $3/2E$, όπου E , ο μέσος όρος της ενέργειας του σήματος [13]. Στο παράρτημα Β' δίνεται ο αλγόριθμος, ο οποίος εξάγει N συλλαβές από μια ηχογράφηση, ξεκινώντας από αυτήν με το μεγαλύτερο πλάτος. Στη συνέχεια, υπολογίζονται οι DTW (Dynamic Time Warping) διαφορές μεταξύ των τροχιών στις συλλαβές (Gaussian syllable prototypes). Επιλέγονται τυχαία k συλλαβές και αυτές θεωρούνται ως αρχικά

πρότυπα για κάθε είδος. Κατόπιν, για κάθε συστάδα συλλαβών που έχουν εξαχθεί από κάθε τμήμα (chunk), υπολογίζονται οι διαφορές από τα αρχικά πρότυπα και ταξινομούνται ανάλογα με την πιο ελάχιστη απόσταση που έχουν στο αντίστοιχα πρότυπα. Ο όρος 'chunk' χαρακτηρίζει ένα μικρό απόσπασμα του αρχικού σήματος. Έτσι αντικαθίστανται τα πρότυπα από αυτά που έχουν μικρότερες διαφορές (κριτήριο για να ανανεωθεί το πρότυπο θεωρείται το ελάχιστο άθροισμα των DTW αποστάσεων). Η διαδικασία επαναλαμβάνεται έως ότου υπάρξει σύγκλιση και συνήθως, χρειάζονται 5 επαναλήψεις.

Αν και απαιτούνται κάποιες ενέργειες (όπως το φιλτράρισμα του σήματος) ώστε να μην ληφθούν υπ' όψιν λανθάνουσες συλλαβές, η μέθοδος αυτή επηρεάζεται ελάχιστα από τον παράγοντα του θορύβου. Η θέση του μεγίστου της τροχιάς του πλάτους σε κάθε συλλαβή είναι ίδια, γεγονός που καθιστά ευκολότερη την σύγκριση διάφορων συλλαβών και επομένως, απλούστερο τον διαχωρισμό των ειδών [12].

Στη συνέχεια, υπολογίζεται η προηγούμενη πιθανότητα προτύπων $i = 1 \dots k$, για κάθε συλλαβή x από την παρακάτω εξίσωση:

$$p(i|x) = \frac{1/(\sqrt{2\pi}\sigma_i)e^{(-d_{ix}^2/2\sigma_i^2)}}{\sum_{j=1}^k 1/(\sqrt{2\pi}\sigma_j)e^{(-d_{jx}^2/2\sigma_j^2)}} \quad (2.1)$$

όπου d_{ix} είναι η DTW απόσταση μεταξύ της συλλαβής x και του προτύπου i και σ_i^2 είναι η διασπορά. Η διασπορά αυτή υπολογίζεται αθροίζοντας τις υψωμένες στο τετράγωνο DTW αποστάσεις μεταξύ του προτύπου i και της συλλαβής η οποία είναι η πιο κοντινή στο πρότυπο αυτό.

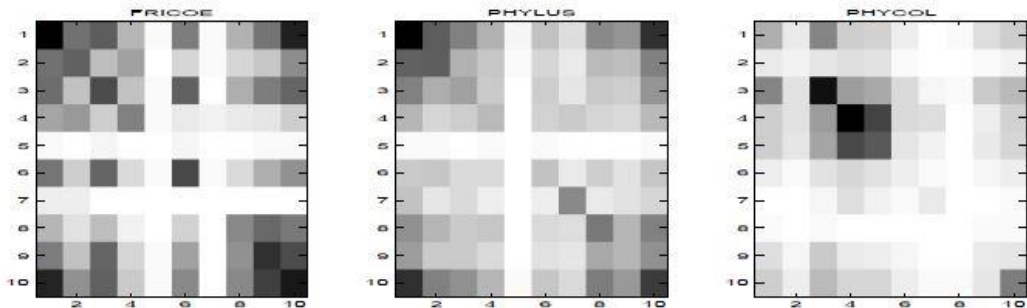
Με δεδομένα τα πρότυπα, η κατηγοριοποίηση πραγματοποιείται μέσω ιστογράμματος. Για κάθε είδος, δεν αρκεί μόνο μια συλλαβή για να χρησιμοποιηθεί ως πρότυπο, γι' αυτό μέσω του γκαουσιανού n -γράμματος (n -gram) εφαρμόζονται όλα τα πρότυπα και όχι μόνο αυτά που έχουν μεγάλη ομοιότητα με το αρχικό πρότυπο. Υπολογίζεται η πιθανότητα ομοιότητας μιας συλλαβής με μια επόμενη και έτσι δημιουργείται για το κάθε είδος το ιστόγραμμα του. Θεωρώντας 2 συνεχόμενες συλλαβές x_t, x_{t-1} και τις γκαουσιανές πιθανότητες p_t, p_{t-1} (bigrams), υπολογίζονται οι τιμές του bigram για 2 συνεχόμενα πρότυπα i, j από τον τύπο:

$$h_{i,j}(t) = \frac{p_{t-1,i}p_{t,j}}{\sum_{i',j'} p_{t-1,i'}p_{t,j'}} \quad (2.2)$$

όπου i' , συμβολίζει το διάνυσμα i που έχει αντιμετατεθεί. Στη συνέχεια, ομαλοποιείται το bigram:

$$H = \sum_{t=2}^L p_{t-1} p_t^T / |p_{t-1} p_t^T| \quad (2.3)$$

όπου p_i^T , το διάνυσμα που έχει αντιμετωπιστεί. Στο σχήμα 2.1 που ακολουθεί, παρουσιάζονται για 3 είδη τα ιστογράμματα συλλαβών όπου χρησιμοποιήθηκαν 10 γκαουσιανές ως πρότυπα [12].



Σχήμα 2.1 Γκαουσιανά ιστογράμματα πρότυπα για 3 είδη, το σκούρο χρώμα δηλώνει μεγάλη ομοιότητα μεταξύ των 10 συλλαβών που χρησιμοποιήθηκαν ως χαρακτηριστικά πρότυπα για κάθε είδος [12].

Το τελικό στάδιο αποτελείται από την σύγκριση των ιστογραμμάτων, υπολογίζοντας τη συσχέτιση μεταξύ τους. Το ομαλοποιημένο bigram H μετατρέπεται σε διάνυσμα-στήλη μιας διάστασης h , όπου για k πρότυπα, ο συντελεστής συσχέτισης για δυο ιστογράμματα h_1, h_2 υπολογίζεται από τη σχέση:

$$c(h_1, h_2) = \frac{h_1^T h_2}{\sqrt{h_1^T h_1} \sqrt{h_2^T h_2}} \quad (2.4)$$

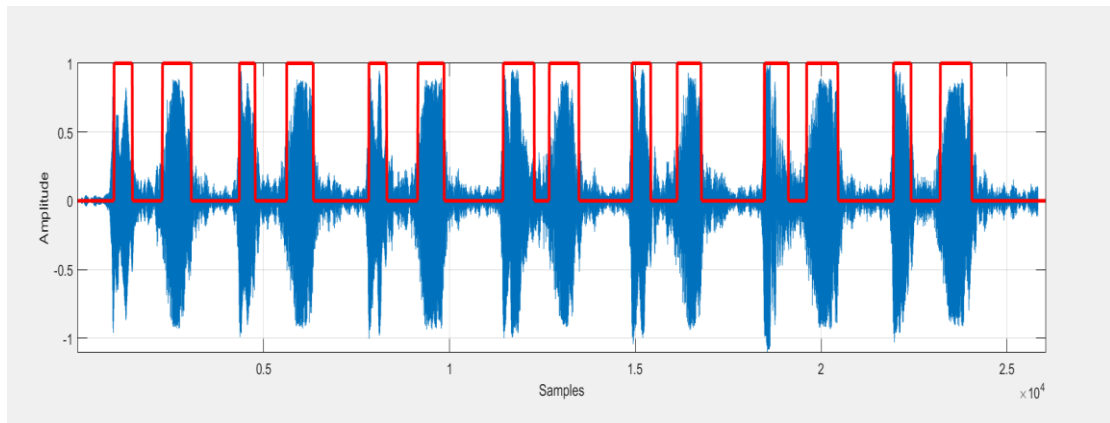
Ο συντελεστής κυμαίνεται από το 0 μέχρι 1 και όσο πιο κοντά είναι στο μηδέν, δηλώνει ότι τα ιστογράμματα δεν παρουσιάζουν αρκετά κοινά στοιχεία ενώ αντίστοιχα, στο 1 δηλώνεται το αντίθετο.

2.2 Εξαγωγή συλλαβών με ακολουθητή HILBERT και ταξινόμηση μέσω Κρυμμένων Μαρκοβιανών Μοντέλων (HMMs)

Σύμφωνα με την προτεινόμενη μέθοδο [8], η προεργασία συνίσταται στον διαχωρισμό του θορύβου από το σήμα. Επειδή πολλά είδη πουλιών κελαηδούν σε διαφορετικές περιοχές συχνότητας, κρίνεται αναγκαία η διαφορετική αντιμετώπιση

για κάθε είδος. Συγκεκριμένα, γίνεται υποδειγματοληψία όσο πιο χαμηλά το επιτρέπει η μέγιστη συχνότητα της ηχογράφησης, δηλαδή περίπου 3 φορές επάνω από τη συχνότητα όπου κυμαίνονται οι ηχογραφήσεις. Στη συνέχεια, εφαρμόζεται ζωνοπερατό φίλτρο, με συχνοτικό εύρος περίπου όπου κυμαίνονται τα τραγούδια, δηλαδή για κάθε είδος, επιλέγεται διαφορετικό εύρος. Το στάδιο αυτό θεωρείται πολύ κρίσιμο καθώς τα τραγούδια των πτηνών επιβαρύνονται από θόρυβο (παρεμβολές από τον αέρα, την βροχή ή και άλλα είδη).

Στη συνέχεια, ο ακολουθητής Hilbert επιτρέπει τον εντοπισμό και την απομόνωση μόνο των συλλαβών σε μια ηχογράφηση. Το πραγματικό μέρος του μετασχηματισμού Hilbert συγκρίνεται με ένα κατώφλι θ , στην προσπάθεια της εξαγωγής συλλαβών. Όταν το σήμα είναι μεγαλύτερο από το κατώφλι τότε αποθηκεύεται, τα υπόλοιπα θεωρούνται θόρυβος. Στο σχήμα 2.2 που ακολουθεί, εφαρμόζεται ο ακολουθητής σε ένα σήμα και μόνο τα διαστήματα όπου ο ακολουθητής έχει πλάτος 1 επιλέγεται να αποθηκευτούν.



Σχήμα 2.2 Ο ακολουθητής Hilbert (κόκκινο χρώμα) εφαρμοσμένος σε ένα σήμα. Αφού πολλαπλασιαστεί με το σήμα (εκεί όπου έχει πλάτος 1), θα αποθηκευτούν τα αντίστοιχα μόνο διαστήματα του σήματος. Ο οριζόντιος άξονας είναι ο αριθμός των δειγμάτων και ο κατακόρυφος το πλάτος της ηχογράφησης [14].

Σε κάθε συλλαβή που έχει εξαχθεί, υπολογίζονται οι αντιληπτικοί γραμμικοί cepstral συντελεστές πρόβλεψης (Perceptual Linear Prediction Cepstral Coefficients, PLP-CC). Έτσι, κάθε ηχογράφηση χωρίζεται σε τμήματα (chunks) των 10ms ενώ κατόπιν εφαρμόζεται παράθυρο hamming με 50% επικάλυψη. Κάθε πλαισιωμένο (framed) παράθυρο περνάει από φίλτρο προ-έμφασης, με συντελεστή 0.97. Έπειτα, υπολογίζεται η ενέργεια και πολλαπλασιάζεται με 41 mel τριγωνικά φίλτρα. Αντί της λογαρίθμησης, εξάγεται η κυβική ρίζα της ενέργειας (16 PLP συντελεστές). Η ενέργεια του σήματος και οι PLP-CC συντελεστές συνδέονται μέσω της σχέσης:

$$P(\omega) = \frac{G^2}{|1 + \sum_{k=1}^p a_k e^{-jk\omega}|^2} \quad (2.5)$$

όπου a_k , οι συντελεστές. Οι συντελεστές αυτοί μετασχηματίζονται (cepstral liftering)

ως εξής:

$$w_c = 1 + L/2 * \sin([1:(ncer) - 1] * \pi/L) \quad (2.6)$$

όπου $L=22$ και $ncer=16$ το πλήθος των cepstral συντελεστών. Τέλος, υπολογίζονται οι δέλτα και δέλτα-δέλτα συντελεστές. Η εκπαίδευση και η ταξινόμηση γίνεται με την χρήση Κρυμμένων Μαρκοβιανών Μοντέλων (HMMs) [8].

2.3 Άλλοι τρόποι κατηγοριοποίησης ήχων πτηνών

Υπάρχουν αρκετές ακόμα μέθοδοι βασιζόμενες σ' αυτές που έχουν ήδη αναφερθεί, με κάποιες διαφοροποιήσεις. Άλλες χρησιμοποιούν επιπλέον χαρακτηριστικά ή μερικά από αυτά, όμως δεν θα γίνει περαιτέρω αναφορά καθώς δεν είναι σκοπός της παρούσας εργασίας η εκτενής ανάλυση των μεθόδων. Ωστόσο, δίνονται στην βιβλιογραφία [15], [16], [17], [18], [13], [19].

ΚΕΦΑΛΑΙΟ 3 : ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΗΝ ΕΠΕΞΕΡΓΑΣΙΑ ΤΟΥ ΑΚΟΥΣΤΙΚΟΥ ΣΗΜΑΤΟΣ

3.1 Συντελεστές MFCCs (Mel Frequency Cepstral Coefficients)

Οι MFCCs συντελεστές παρουσιάστηκαν από τους Steven S. Davis και Paul Mermelstein το 1980 και αποτελούν μέχρι σήμερα την τεχνολογία αιχμής στην αυτόματη ταξινόμηση των ακουστικών σημάτων. Η τεχνική αυτή χρησιμοποιείται σε ποικίλες εφαρμογές, όπως την αναγνώριση μουσικής και της ομιλίας, κάθε φορά όμως προσαρμόζεται ανάλογα με τις απαιτήσεις της εφαρμογής.

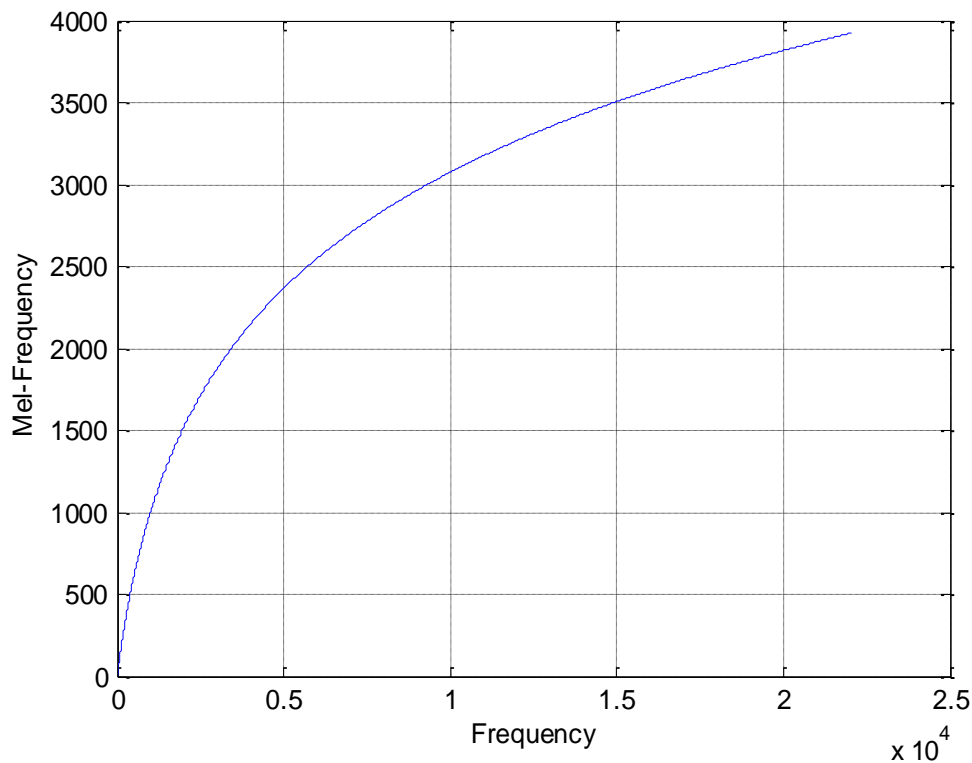
Στην ουσία γίνεται μια προσομοίωση της ανθρώπινης ακοής, η οποία είναι γραμμική μέχρι τα 1000 Hz κι έπειτα γίνεται λογαριθμική. Έτσι η μετατροπή από την κλίμακα των hertz σε mel γίνεται ως εξής:

$$M(f) = 1125 \ln(1 + f/700) \quad (3.1)$$

ενώ από mel σε συχνότητα, η μετατροπή γίνεται με τον παρακάτω τύπο:

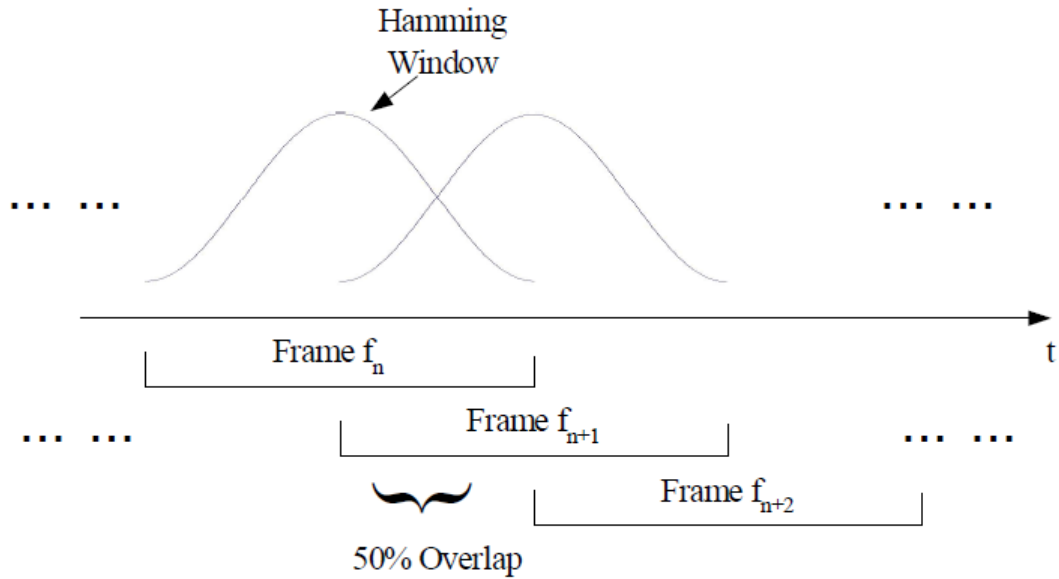
$$M^{-1}(m) = 700(e^{m/1125} - 1) \quad (3.2)$$

Στο παρακάτω σχήμα (σχ. 3.1) απεικονίζεται η μεταβολή της κλίμακας mel σε σχέση με τα hertz.



Σχήμα 3.1 Το διάγραμμα απόκρισης της συχνότητας σε σχέση με την κλίμακα mel, όπου μέχρι τα 1000 είναι σχεδόν γραμμική και από το σημείο αυτό γίνεται λογαριθμική. Ο οριζόντιος άξονας παριστά τη συχνότητα σε hertz και ο κατακόρυφος σε mel κλίμακα.

Πρώτα απ' όλα, το ηχογραφημένο σήμα φιλτράρεται μέσω ενός φίλτρου προέμφασης με συντελεστή 0.97. Έπειτα, αφού έχει περάσει από το φίλτρο, χωρίζεται σε μερικώς επικαλυπτόμενα παράθυρα (frames) διάρκειας μερικών ms ώστε να γίνει υπέρβαση του προβλήματος της μη στασιμότητας ενός σήματος. Συγκεκριμένα, κάποιες ενδεικτικές τιμές για την τμηματοποίηση της ενιαίας ηχογράφησης σε παράθυρα είναι τα 20-40ms [20], [21]. Η στασιμότητα σε ένα παράθυρο εξασφαλίζεται λόγω της μικρής διάρκειας του, όπως και της επικάλυψης που συνήθως είναι 50% [20]. Σχετικά με το πιθανό πρόβλημα δημιουργίας θορύβου υψηλής συχνότητας στην αρχή και στο τέλος κάθε παραθύρου, λόγω της ακαριαίας μεταβολής των τιμών στο παράθυρο σε μηδέν και από το μηδέν σε διάφορες τιμές, εφαρμόζεται σε καθένα από αυτά παραθύρωση Hamming, μήκους όσο και το κάθε frame, προκειμένου να εξασφαλισθεί η συνέχεια [20]. Στο παρακάτω σχήμα 3.2 παρουσιάζεται η προαναφερθείσα διαδικασία.



Σχήμα 3.2 Διάγραμμα σχηματικής αναπαράστασης της τμηματοποίησης της ηχογράφησης σε frames με 50% επικάλυψη με την χρήση παραθύρου Hamming [20].

Στη συνέχεια, πραγματοποιείται ο υπολογισμός της φασματικής ισχύος για κάθε frame. Ο υπολογισμός γίνεται με κάποιο συγκεκριμένο επιλεγμένο αριθμό $nfft$, ο οποίος είναι δύναμη του 2. Ο αριθμός αυτός αποτελεί την διακριτική ικανότητα της ισχύος (συνήθως επιλέγεται 512 ή 1024).

Το επόμενο βήμα περιλαμβάνει τη δημιουργία των τριγωνικών mel φίλτρων. Έτσι, για μια συχνότητα δειγματοληψίας F_s επιλέγεται μια ελάχιστη τιμή της συχνότητας (f_{min}) και μια μέγιστη (f_{max}). Στην αναγνώριση ομιλίας συνιστάται η f_{min} να ισούται με 100 Hz, επιπλέον όταν f_{min} 50 με 60 Hz αποφεύγονται τα AC βουητά της ομιλίας [22]. Έπειτα, γίνεται η μετατροπή των f_{min} και f_{max} σε κλίμακα mel (εξίσωση 3.1), αντίστοιχα, σε mel_{min} και mel_{max} . Γίνεται η ταξινόμηση σε πίνακα γραμμικά των τιμών πλήθους M (όσο και ο αριθμός των φίλτρων), με πρώτη τιμή την τιμή mel_{min} και τελευταία mel_{max} . Κατόπιν, μετασχηματίζεται ο πίνακας αυτός στη συχνότητα μέσω της σχέσης (3.2) και δημιουργείται νέος πίνακας F με γραμμικά διαιρούμενα διαστήματα πλήθους M . Τέλος, υπολογίζονται τα bin διαστήματα μέσω του τύπου:

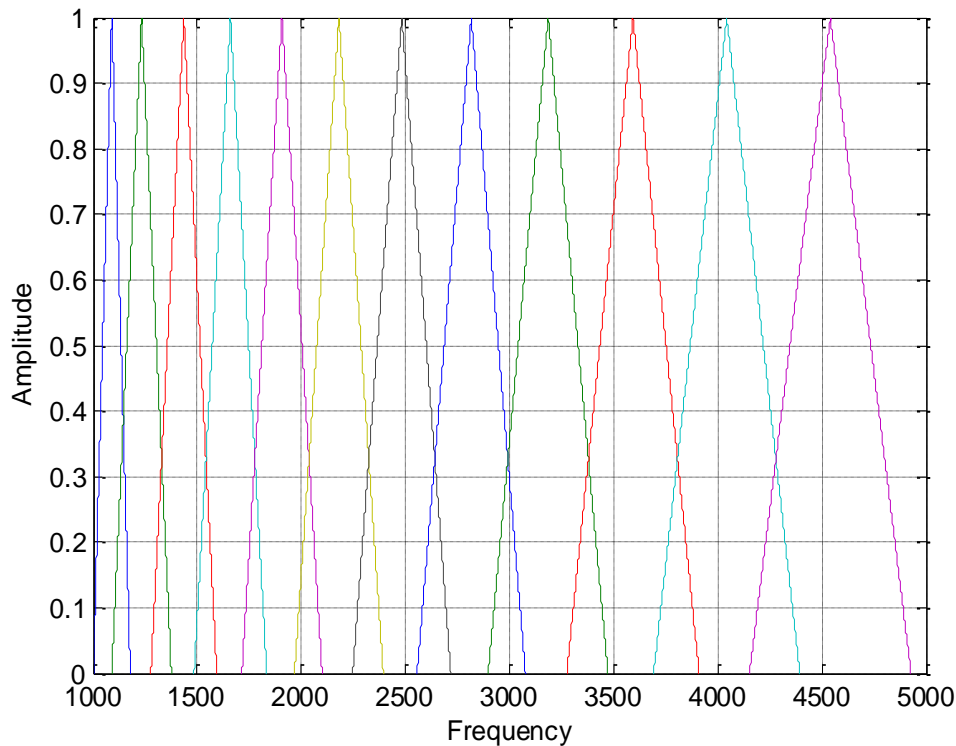
$$b = \frac{(nfft+1)F}{F_s} \quad (3.3)$$

Το πρώτο φίλτρο θα ξεκινάει από το πρώτο σημείο, θα γίνεται μονάδα στο δεύτερο και θα σταματάει στο τρίτο. Το δεύτερο φίλτρο θα ξεκινάει από το δεύτερο σημείο, θα φτάνει στο μέγιστο του στο τρίτο σημείο και έπειτα θα τελειώνει στο τέταρτο κ.ο.κ. [21]. Παρακάτω παρατίθεται ο τύπος για τον υπολογισμό των μοναδιαίων τριγωνικών μελ φίλτρων.

$$H(m, k) = \begin{cases} 0, & k < b(m-1) \text{ ή } k \geq b(m+1) \\ \frac{k-b(m-1)}{b(m)-b(m-1)}, & b(m-1) \leq k < b(m) \\ \frac{b(m+1)-k}{b(m+1)-b(m)}, & b(m) \leq k < b(m+1) \end{cases} \quad (3.4)$$

όπου $1 \leq k \leq K$ και $1 \leq m \leq M$.

Τα φίλτρα αυτά εστιάζουν περισσότερο στις χαμηλές συχνότητες, γιατί εκεί βρίσκεται η πληροφορία που μας ενδιαφέρει συνήθως. Στην αρχή, τα εύρη συχνοτήτων (bandwidth) είναι στενά και τα φίλτρα είναι κοντά μεταξύ τους, όσο όμως αυξάνεται η συχνότητα, αυξάνεται και το bandwidth κάθε φίλτρου και αραιώνουν μεταξύ τους. Στο επόμενο σχήμα (σχ. 3.3), δίνεται ένα τέτοιο παράδειγμα 12 φίλτρων, προσαρμοσμένων να ξεκινάνε από τα 1000Hz και φτάνουν μέχρι τα 5 KHz.



Σχήμα 3.3 Διάγραμμα 12 mel φίλτρων τα οποία ξεκινάνε από τα 1000 Hz και φτάνουν έως τα 5000 Hz. Ο οριζόντιος άξονας είναι η συχνότητα σε Hz και ο κατακόρυφος το πλάτος.

Στο επόμενο στάδιο περιλαμβάνεται ο πολλαπλασιασμός των φίλτρων αυτών με τη φασματική ισχύ καθενός από τα frames. Αφού λογαριθμιστεί το αποτέλεσμα του πολλαπλασιασμού και έτσι δημιουργείται ένας νέος πίνακας L_p , σύμφωνα με τον παρακάτω τύπο:

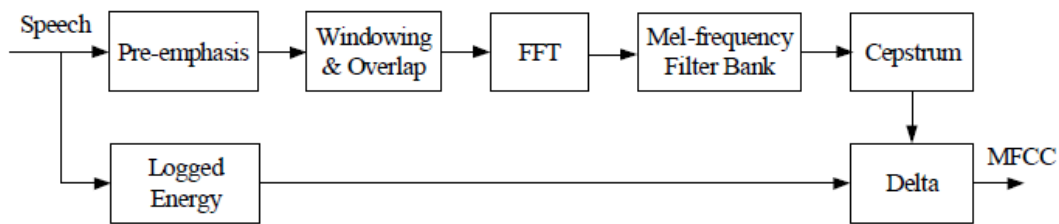
$$L_p(m, k) = \ln\left\{\sum_{k=0}^{N-1} H(m, k) * X_p(k)\right\} \quad (3.5)$$

όπου $1 \leq p \leq P$ είναι το πλήθος των frames, $1 \leq m \leq M$ ο αριθμός των φίλτρων και X είναι ο STFT του frame p .

Το τελευταίο στάδιο περιλαμβάνει τον υπολογισμό του DCT (Discrete Cosine Transform) για κάθε frame του προηγούμενου πίνακα L_p κι έτσι πλέον μετατρέπονται οι τιμές από τη mel συχνότητα, στον χρόνο. Σε κάθε frame αντιστοιχεί κι ένας αντιπροσωπευτικός συντελεστής φάσματος, δηλαδή, δημιουργείται ο πίνακας $\Phi[p]$,

όπου $1 \leq p \leq P$ αποτελεί τον αριθμό των frames. Ο πίνακας αυτός περιλαμβάνει τους ζητούμενους συντελεστές MFCCs ανά frame.

Συνήθως από τους υπολογισμένους mel συντελεστές υπολογίζεται το cepstrum liftering με συντελεστή 15 ή 20. Όπως επίσης υπολογίζονται οι delta (time derivatives) και οι delta-deltas (acceleration derivatives) συντελεστές. Στο επόμενο μπλοκ διάγραμμα παρουσιάζεται ολόκληρη η διαδικασία (σχήμα 3.4) [20].



Σχήμα 3.4 Διάγραμμα απεικόνισης υπολογισμού των MFCCs συντελεστών [20].

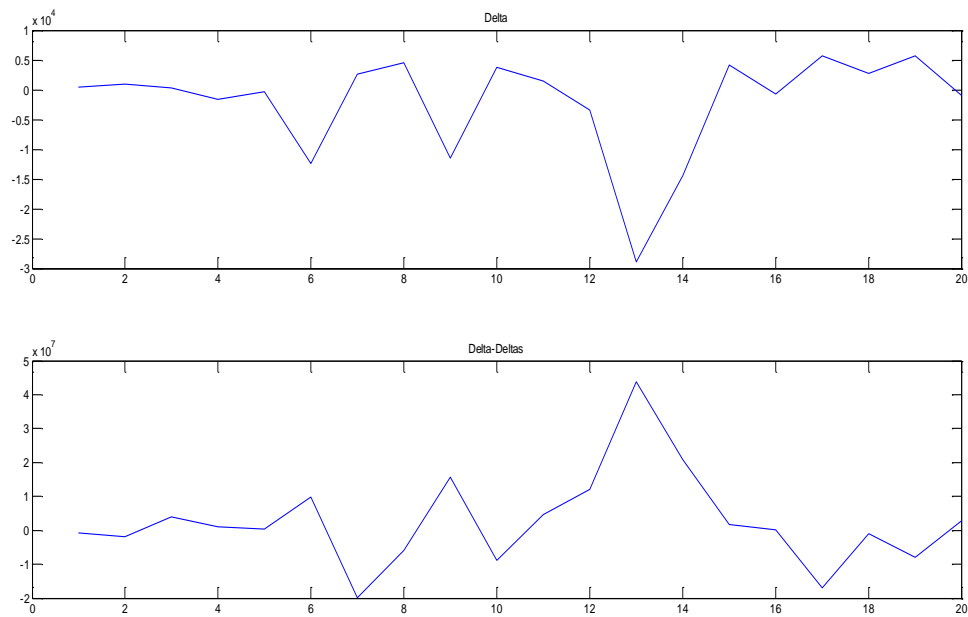
3.2 Delta - Παράγωγος 1ης Τάξης (Time Derivative) / Delta-Deltas - Παράγωγος 2ης Τάξης (Acceleration Derivative)

Στην ομιλία, η ροπή (trend) χάνεται λόγω της παραθύρωσης, γι' αυτό κρίνεται απαραίτητος ο υπολογισμός των συντελεστών delta ώστε να επιτυγχάνονται πιο ακριβή αποτελέσματα στο στάδιο της αναγνώρισης. Η επιπλέον πληροφορία για τα δυναμικά χαρακτηριστικά της ομιλίας δίνονται μέσω αυτών των χαρακτηριστικών [21]. Έστω ότι έχουν υπολογιστεί οι συντελεστές MFCCs, $c(n)$, για κάθε frame.

Τότε, οι delta συντελεστές υπολογίζονται ως εξής :

$$d(n) = \frac{\sum_{i=1}^D i * (c(n+i) - c(n-i))}{\sum_{i=1}^D i^2} \quad (3.6)$$

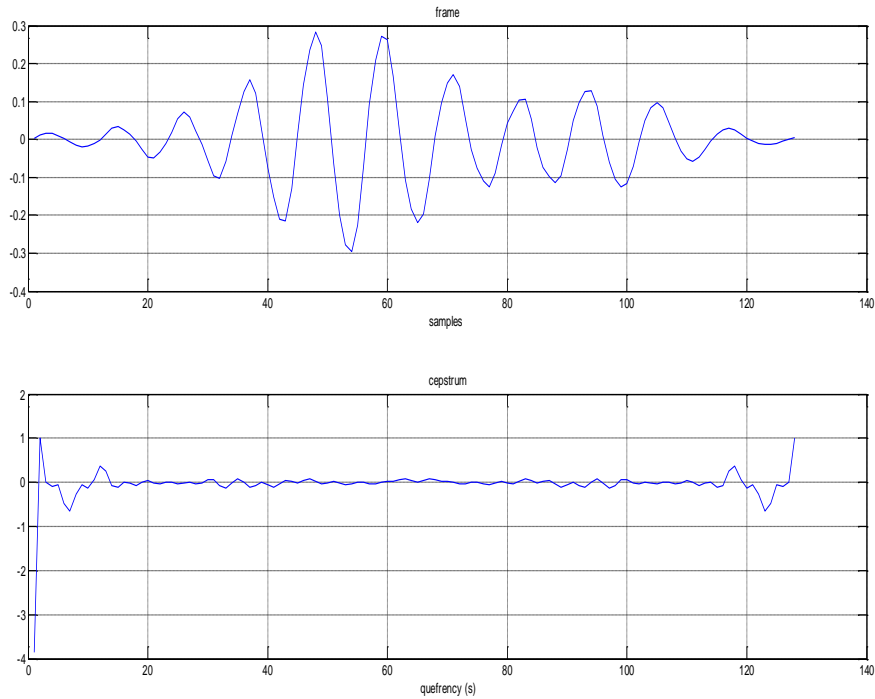
όπου συνήθως $D=2$. Αντίστοιχα, οι delta-deltas υπολογίζονται με την ίδια σχέση μόνο που αντί τους συντελεστές mel αυτή την φορά χρησιμοποιούνται οι ήδη υπολογισμένοι συντελεστές delta [22]. Στο παρακάτω σχήμα (σχ. 3.5) δίδεται η αναπαράσταση των συντελεστών delta και delta-deltas, αντίστοιχα, ενός σήματος.



Σχήμα 3.5 Διάγραμμα delta και delta-deltas 20 συντελεστών αντίστοιχα. Στον οριζόντιο άξονα είναι οι συντελεστές και στον κατακόρυφο οι τιμές τους.

3.3 Cepstrum/Liftering

Ο όρος cepstrum προκύπτει από τον αναγραμματισμό του αγγλικού όρου για το φάσμα (spectrum) και ο υπολογισμός του μοιάζει με τον υπολογισμό του φάσματος αλλά με κάποιες βασικές διαφορές. Μετά από τον διαχωρισμό του σήματος σε frames, υπολογίζεται ο ταχύς μετασχηματισμός Fourier (FFT) για καθένα από αυτά. Κατόπιν, γίνεται λογαρίθμηση και υπολογισμός των αντίστροφων μετασχηματισμών Fourier (IFFT). Το αποτέλεσμα ονομάζεται cepstrum, δηλαδή είναι ο IFFT του λογαριθμημένου φάσματος. Ο οριζόντιος άξονας έχει τις μονάδες της χρονικής περιοδικότητας quefrensy, όρος που κατά αντιστοιχία, προκύπτει από τη συχνότητα (frequency), ενώ ο κατακόρυφος, δείχνει το πλάτος της έντασης. Ακόμη, οι συντελεστές του cepstrum που ισαπέχουν μεταξύ τους χρονικά κατά κάποιο πολλαπλάσιο ονομάζονται rahmonics (αναγραμματισμός του αγγλικού όρου για τις αρμονικές, harmonics) [23]. Ένα παράδειγμα δίνεται στο παρακάτω σχήμα (σχ. 3.6), όπου έχουμε το frame ενός σήματος και το αντίστοιχο cepstrum του.



Σχήμα 3.6 Τμήμα ενός σήματος όπου ο οριζόντιος άξονας είναι τα δείγματα και ο κατακόρυφος η ένταση του. Παρακάτω είναι το cepstrum του frame αυτού, όπου ο οριζόντιος άξονας είναι τα δείγματα και ο κατακόρυφος η ένταση αντίστοιχα.

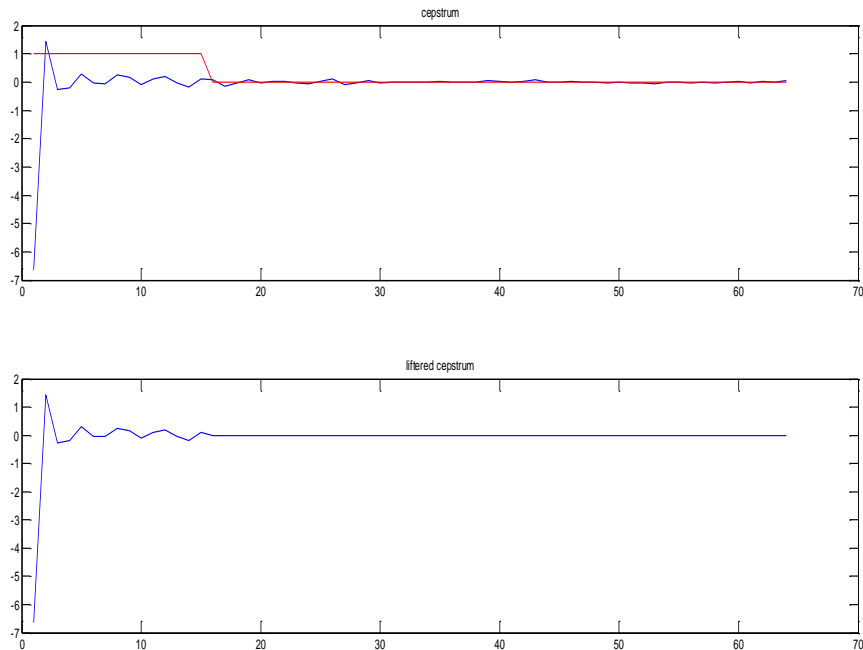
Η διαδικασία του liftering είναι όμοια με αυτή των φίλτρων στον χρόνο (filtering – liftering). Για το φίλτρο, επιλέγεται κατ' αρχήν κατάλληλο παράθυρο με μοναδιαίο πλάτος και κατάλληλο μήκος, το οποίο, στη συνέχεια, πολλαπλασιάζεται με το σήμα. Μεσολαβεί η πράξη της συνέλιξης του φίλτρου με το σήμα (στο χρόνο), ή αλλιώς πολλαπλασιασμός στο πεδίο της συχνότητας. Υπάρχουν δυο κατηγορίες liftering η μια ονομάζεται low-time liftering και η άλλη high-time liftering (από το low-pass filtering και high-pass filtering αντίστοιχα) [24].

Το low-time liftering εφαρμόζεται για τη εξαγωγή των χαρακτηριστικών της ηχητικής οδού (vocal tract) στο πεδίο quefrency. Χρησιμοποιείται στην ανίχνευση των αργών μεταβολών της φωνητικής οδού, υπολογίζοντας τις μέγιστες κορυφές από το cepstrum. Το παράθυρο που χρησιμοποιείται περιγράφεται από τη σχέση:

$$w(n) = \begin{cases} 1, & 0 \leq n \leq L \\ 0, & L \leq n \leq \frac{N}{2} \end{cases} \quad (3.7)$$

όπου L είναι το μήκος αποκοπής του παραθύρου και N το μήκος του cepstrum. Συνήθως χρησιμοποιείται συντελεστής 15 ή 20. Παρακάτω (σχ. 3.7), παρουσιάζεται

το cepstrum ενός frame και το low-time lifter (με συντελεστή 15) με κόκκινο και στη συνέχεια το αποτέλεσμα του low-time liftering στο cepstrum αυτό. Επειδή το cepstrum είναι συμμετρικό, το liftering εφαρμόζεται μόνο για το μισό cepstrum.

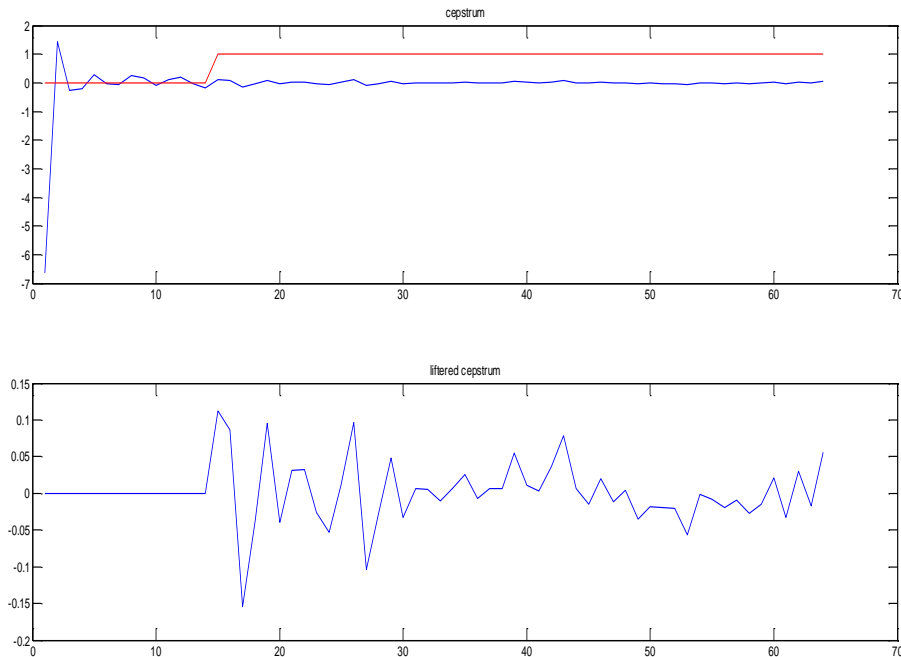


Σχήμα 3.7 Το cestrum πριν το low-time liftering, όπου η κόκκινη γραμμή είναι το φίλτρο με συντελεστή 15. Στον οριζόντιο άξονα είναι τα δείγματα του frame και στον κατακόρυφο η ένταση του cepstrum και του φίλτρου. Στο κάτω διάγραμμα είναι το cepstrum μετά το low-time liftering [24].

Στη επόμενη κατηγορία περιλαμβάνεται το high-time liftering, που χρησιμοποιείται για τα χαρακτηριστικά της διέγερσης ενός σήματος. Το παράθυρο που χρησιμοποιείται περιγράφεται από τη σχέση:

$$w(n) = \begin{cases} 1, & L \leq n \leq \frac{N}{2} \\ 0, & \text{αλλιού} \end{cases} \quad (3.8)$$

όπου L είναι το μήκος αποκοπής του παραθύρου και N το μήκος του cepstrum. Παρακάτω, δίνεται το cepstrum ενός frame (με μπλε) και το high-time lifter συντελεστή 15 (με κόκκινο), ενώ πιο κάτω το αποτέλεσμα του high-time liftering στο cepstrum αυτό (σχήμα 3.8).



Σχήμα 3.8 Το cepstrum πριν το high-time liftering, όπου η κόκκινη γραμμή είναι το φίλτρο με συντελεστή 15. Στον οριζόντιο άξονα είναι τα δείγματα του frame και στον κατακόρυφο η ένταση του cepstrum και του φίλτρο. Παρακάτω είναι το cepstrum που έχει περάσει από το high-time liftering [24].

Καθώς οι μεγαλύτερου βαθμού cepstral συντελεστές έχουν μικρές τιμές σε σχέση με αυτούς που είναι χαμηλού βαθμού, δημιουργείται πρόβλημα επειδή οι συντελεστές δεν συγκλίνουν σε μια περιοχή, αλλά είναι διασκορπισμένοι [22]. Γι' αυτό ακριβώς χρησιμοποιείται το cepstral liftering ώστε δηλαδή, να συγκεντρωθούν προς μια συγκλίνουσα περιοχή, σύμφωνα με τη σχέση:

$$c_l = \left(1 + \frac{L}{2} \sin\left(\frac{\pi n}{L}\right)\right) c(n) \quad (3.9)$$

όπου $c(n)$ είναι το n -οστός συντελεστής και L είναι η προεπιλεγμένη παράμετρος, στην περίπτωσή μας $L = 22$.

3.4 Προέμφαση πρώτου βαθμού (First order pre-emphasis)

Στην επεξεργασία σήματος ομιλίας, το αρχικό σήμα έχει συνήθως αρκετή ενέργεια στις χαμηλές συχνότητες. Ως προεργασία, το προετοιμάζουμε ενισχύοντας τις υψηλότερες συχνότητες, με ένα φίλτρο προέμφασης (pre-emphasis). Η σχέση εισόδου-εξόδου της προέμφασης με συντελεστή k (όπου $0.9 < k < 1$) είναι:

$$y(n) = x(n) - k * x(n - 1) \quad (3.10)$$

και επομένως, για συντελεστή $k = 0.97$ γίνεται:

$$y(n) = x(n) - 0.97 * x(n - 1) \quad (3.11)$$

Προκύπτει, δηλαδή, η εξίσωση της πρώτης προς τα πίσω διαφοράς με συντελεστή 0.97. Η κρουστική της απόκριση θα είναι:

$$h(n) = \delta(n) - 0.97 * \delta(n - 1) \quad (3.12)$$

όπου δ , είναι η συνάρτηση του Dirac ενώ στο πεδίο της συχνότητας, η απόκριση είναι:

$$H(\omega) = 1 - 0.97 * e^{-j\omega} \quad (3.13)$$

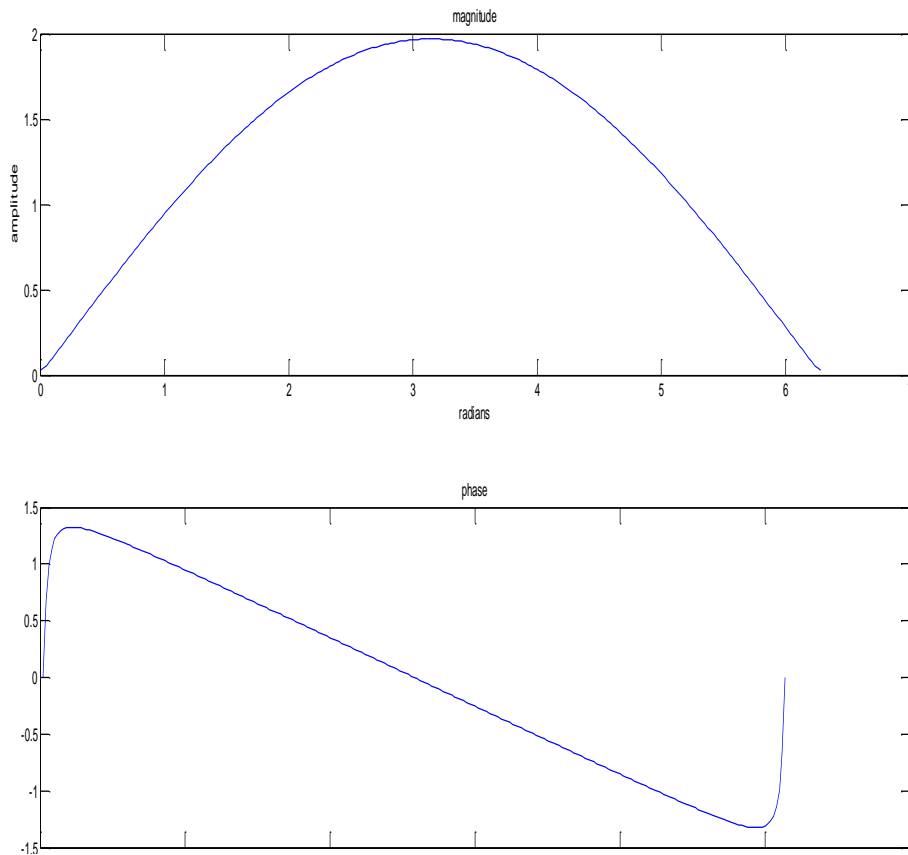
η οποία σύμφωνα με τον τύπο του Euler

$$e^{-j\varphi t} = \cos\varphi t - j\sin\varphi t \quad (3.14)$$

αναλύεται ως εξής:

$$H(\omega) = 1 - 0.97\cos\omega + j0.97\sin\omega \quad (3.15)$$

Εύκολα προκύπτει ότι πρόκειται για ένα FIR (Finite Impulse Response) υψηλοπερατό φίλτρο με γραμμική φάση. Στο παρακάτω σχήμα (σχ. 3.9) φαίνεται το πλάτος και τη φάση της απόκρισης συχνότητας του συγκεκριμένου φίλτρου.



Σχήμα 3.9 Φίλτρο προ-έμφασης: ο οριζόντιος άξονας στα διαγράμματα είναι σε ακτίνια (rad).

3.5 Συσχέτιση - Αυτοσυσχέτιση (Correlation - Autocorrelation)

Η συσχέτιση μεταξύ δυο ακολουθιών χρησιμοποιείται συχνά στην επεξεργασία σήματος. Αποτελεί ένα μέτρο σύγκρισης χρονοσειρών και εκφράζει τον βαθμό ομοιότητας που παρουσιάζουν οι ακολουθίες μεταξύ τους. Έστω οι ακολουθίες $x(n)$,

$y(n)$, τότε η συσχέτισή τους δίνεται από τη σχέση:

$$r_{x,y}(l) = \sum_{n=-\infty}^{\infty} x(n)y(n-l) \quad (3.16)$$

όπου το l είναι η μετατόπιση ή lag παράμετρος. Επιπλέον, η αυτοσυσχέτιση υπολογίζεται από τη σχέση:

$$r_{x,x}(l) = \sum_{n=-\infty}^{\infty} x(n)x(n-l) \quad (3.17)$$

Η αυτοσυσχέτιση σε ένα σήμα πεπερασμένης διάρκειας δίνεται από τη σχέση

$$R_{x,x}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n)x(n-k) \quad (3.18)$$

όπου k η μετατόπιση ή lag και N ο αριθμός των δειγμάτων του σήματος [25].

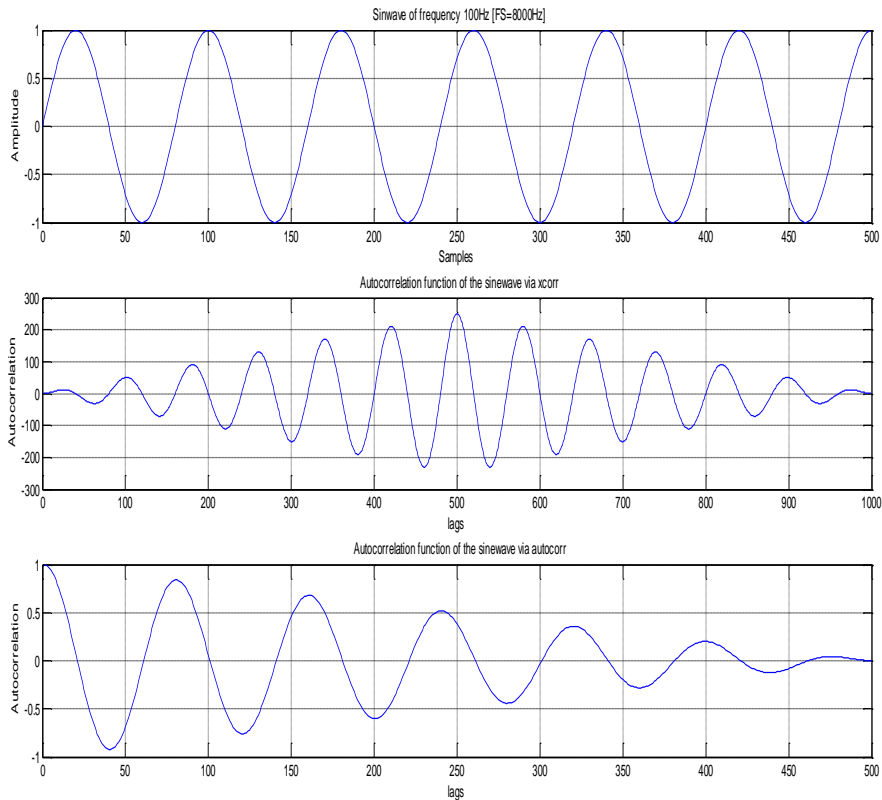
Αποτελεί ένα μέτρο σύγκρισης του σήματος με τον εαυτό του έχοντας καθυστέρηση κάθε φορά κατά k δείγματα. Στην επεξεργασία σήματος χρησιμοποιείται στην αναγνώριση μουσικών οργάνων ή και μιας ομιλίας, στην παρούσα όμως εργασία εφαρμόζεται στον εντοπισμό της θεμελιώδους συχνότητας στις συλλαβές. Τέλος, υπολογίζεται η κανονικοποιημένη συσχέτιση (ως προς τη μέγιστη τιμή της $R_{xx}(0)$),

δηλαδή

$$\rho_{xx}(k) = \frac{R_{xx}(k)}{R_{xx}(0)} \quad (3.19)$$

Έστω ένας ημιτονοειδής παλμός με πλήθος δειγμάτων $N = 500$ με συχνότητα $F = 100 \text{ Hz}$ και συχνότητα δειγματοληψίας $F_s = 8000 \text{ Hz}$. Στο παρακάτω σχήμα

3.10 παρουσιάζονται το ημιτονοειδές σήμα, η αυτοσυσχέτιση μέσω της συνάρτησης `xcorr` του Matlab η οποία έχει συμμετρική απεικόνιση και τέλος η αυτοσυσχέτιση μέσω της `autocorr` που δεν παρουσιάζει την συμμετρία καθώς χρησιμοποιεί μόνο τα μισά lags.



Σχήμα 3.10 Σήμα ημιτονοειδές με συχνότητα δειγματοληψίας 8kHz και συχνότητα του σήματος 100 Hz, στον οριζόντιο άξονα είναι ο αριθμός των δειγμάτων και στον κατακόρυφο το πλάτος του σήματος. Στο σχήμα που ακολουθεί παρουσιάζεται η αυτοσυσχέτιση υπολογισμένη από την `xcorr` στον οριζόντιο είναι οι μετατοπίσεις `lags` και στον κατακόρυφο ο βαθμός ομοιότητας σε κάθε `lag` αντίστοιχα. Τέλος είναι η αυτοσυσχέτιση μέσω της `autocorr`, ο οριζόντιος είναι τα `lags` και ο κατακόρυφος ο βαθμός ομοιότητας κανονικοποιημένος στην μονάδα, δηλαδή όσο πιο κοντά στο 1 τότε έχουμε μεγαλύτερη ομοιότητα.

3.6 Μετασχηματισμός Fourier Διακριτού Χρόνου-DTFT (Discrete Time Fourier Transform)/Σύντομος Μετασχηματισμός Fourier- STFT (Short Time Fourier Transform

Οι ηχογραφήσεις έχουν πραγματοποιηθεί σε πραγματικό χρόνο οπότε στον κατακόρυφο άξονα θα είναι η ένταση του σήματος και στον οριζόντιο ο χρόνος (ή τα δείγματα) όπου για κάθε δείγμα ισχύει:

$$n = T * Fs \tag{3.20}$$

T είναι ο χρόνος και F_s η συχνότητα δειγματοληψίας. Με τον DTFT αναλύεται

περεταίρω το σήμα και συλλέγονται σημαντικές πληροφορίες, όπως το συχνοτικό εύρος όπου κυμαίνονται τα κελαηδήματα κάποιου είδους. Πολλές φορές οι ηχογραφήσεις έχουν πραγματοποιηθεί σε θορυβώδες περιβάλλον με τον ήχο από αυτοκίνητα και τον αέρα να παρεμβάλλονται. Με την φασματική ανάλυση και με την εφαρμογή κατάλληλα επιλεγμένου φίλτρου, απομονώνεται η πληροφορία από τον θόρυβο. Με την έννοια θόρυβος εννοούνται οι οποιεσδήποτε παρεμβολές από διάφορες πηγές που δεν μας ενδιαφέρουν, έτσι μπορεί σε μια ηχογράφιση το τραγούδι από κάποιο άλλο πουλί που δεν μελετάται να θεωρηθεί ως θόρυβος.

Για κάθε ακολουθία $x(n)$ που ισχύει $\sum_{-\infty}^{\infty} |x(n)| < \infty$, υπολογίζεται ο DTFT από τον τύπο:

$$X(e^{j\omega}) \triangleq F[x(n)] = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n} \quad (3.21)$$

Αντίστοιχα, ο αντίστροφος μετασχηματισμός Fourier διακριτού χρόνου (IDTFT) μιας συνεχούς ακολουθίας $X(e^{j\omega})$ υπολογίζεται από τη σχέση:

$$x(n) \triangleq F^{-1}[X(e^{j\omega})] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega n} d\omega \quad (3.22)$$

Στην επεξεργασία σήματος ο DTFT μετασχηματίζει μια διακριτή ακολουθία σε μια συνεχή μιγαδική συνάρτηση $X(e^{j\omega})$, όπου το ω είναι πραγματικός αριθμός και συμβολίζει τη συχνότητα σε radians.

Κατά αντιστοιχία, ο STFT (Short Time Fourier Transform) είναι μετασχηματισμός Fourier για μικρά χρονικά παραθυρωμένα διαστήματα του αρχικού σήματος (frames). Ο STFT διακριτού χρόνου υπολογίζεται ως εξής:

$$X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n - m]e^{-j\omega n} \quad (3.23)$$

όπου w είναι η συνάρτηση του επιλεγμένου παραθύρου (Hamming, Blackman κτλ) [26].

3.7 Παράθυρα/παραθύρωση (windowing)

Στην επεξεργασία σήματος με τον όρο παράθυρα εννοούνται συναρτήσεις που μηδενίζουν το σήμα έξω από κάποια συγκεκριμένα όρια. Τα όρια αυτά επιλέγονται με βάση κάποια κριτήρια, ανάλογα την εφαρμογή. Δηλαδή, πολλαπλασιάζοντας το επιλεγμένο μας παράθυρο με το σήμα, επιτρέπεται η εστίαση σε κάποιο επιλεγμένο κομμάτι του σήματος και η διαδικασία αυτή ονομάζεται παραθύρωση. Η μετατόπιση του παραθύρου κάθε φορά με ένα σταθερό βήμα μας επιτρέπει τη μελέτη ολόκληρου του σήματος, παρόλο που δεν είναι στάσιμο, θεωρώντας ότι στα μικρά αυτά διαστήματα αυτό παραμένει στάσιμο.

Υπάρχουν πολλές συναρτήσεις παραθύρων και όλες τους είναι χρήσιμες. Συγκρίνοντας τις συναρτήσεις αυτές μεταξύ τους παρατηρούνται μερικά μειονεκτήματα σε σχέση με άλλες, όπως για παράδειγμα το τετραγωνικό παράθυρο σε σχέση με το τριγωνικό, όπου η απόκριση του τετραγωνικού δημιουργεί παραμορφώσεις (ripples) στο σήμα. Γι' αυτόν ακριβώς τον λόγο αναφέρθηκε παραπάνω πως ανάλογα με την εφαρμογή επιλέγεται το αντίστοιχο παράθυρο. Το παράρτημα Α' περιέχει τα πιο συνήθη παράθυρα. Στο Matlab υπάρχουν έτοιμες οι συναρτήσεις για τον σχεδιασμό των παραθύρων.

3.8 Συνέλιξη (Convolution)

Στην παρούσα εργασία, με τον όρο συνέλιξη εννοείται η συνέλιξη μιας διάστασης (1D Convolution), καθώς υφίστανται και περισσότερων διαστάσεων όμως δεν θα αναφερθούν άλλες, καθώς στην επεξεργασία σήματος με 1 κανάλι (mono) αυτή χρησιμοποιείται. Έτσι, για δυο σήματα h και x το αποτέλεσμα της συνέλιξης είναι:

$$y_k = \sum_{n=0}^{N-1} h_n * x_{k-n} \quad (3.24)$$

όπου n είναι το πλήθος των στοιχείων στα σήματα εισόδου. Παρακάτω (σχήμα 3.11) παρουσιάζεται σχηματικά ο υπολογισμός της για δυο αιτιατά σήματα [28].

$$\begin{aligned}
 y_0 &= h_0 \cdot x_0 \\
 y_1 &= h_1 \cdot x_0 + h_0 \cdot x_1 \\
 y_2 &= h_2 \cdot x_0 + h_1 \cdot x_1 + h_0 \cdot x_2 \\
 y_3 &= h_2 \cdot x_1 + h_1 \cdot x_2 + h_0 \cdot x_3 \\
 &\vdots
 \end{aligned}$$

Σχήμα 3.11 Απεικόνιση υπολογισμού συνέλιξης για 2 σήματα x και h [28].

3.9 Κινούμενος Μέσος Όρος (Moving Average)

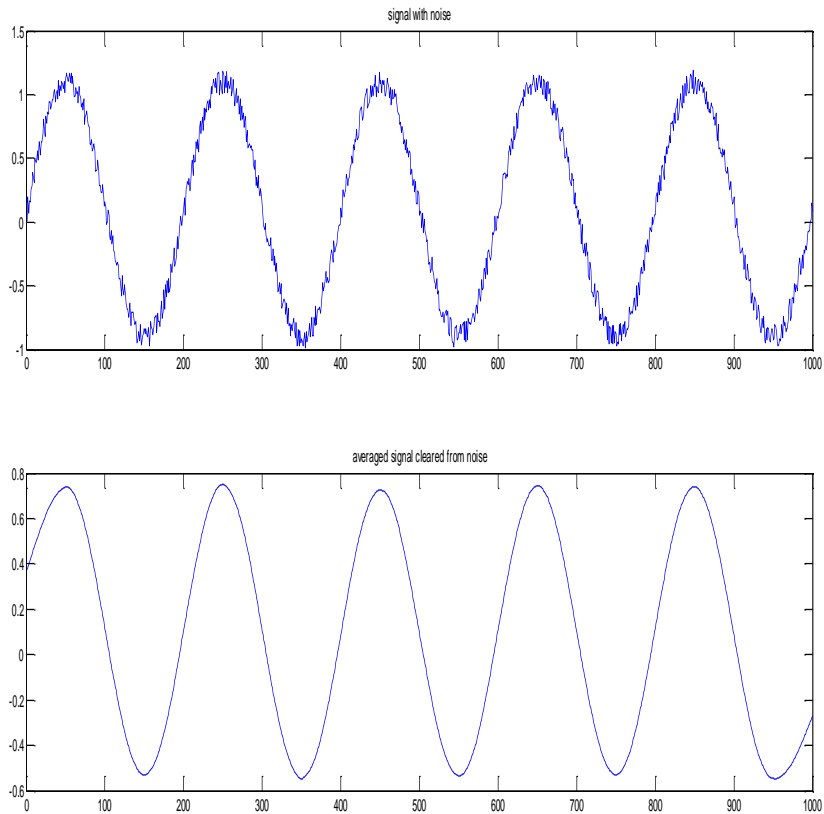
Στην εργασία αυτή έγινε χρήση του κινούμενου μέσου όρου, με στόχο την ελαχιστοποίηση των ανεπιθύμητων και πολύ ακραίων μεταβολών στο σήμα. Κάθε τιμή του υπολογίζεται με βάση τον μέσο όρο των γειτονικών του σημείων. Δηλαδή για κάθε σημείο k , υπολογίζεται ως εξής:

$$y_k = \frac{x_{k-u} + x_{k-u+1} + \dots + x_k + \dots + x_{k+u-1} + x_{k+u}}{w} \quad (3.25)$$

όπου $u = \frac{w-1}{2}$, x είναι το σήμα μας και y είναι ο υπολογισμένος κινητός μέσος όρος

του σήματος. Τέλος w είναι το επιλεγμένο μας παράθυρο και το μέγεθος του οποίου είναι περιττός αριθμός [29]. Έτσι πρώτα επιλέγεται το κατάλληλο μέγεθος του παραθύρου και ύστερα δημιουργείται ένα μοναδιαίο διάνυσμα k μήκους όσο και το μήκος του παραθύρου. Έπειτα πραγματοποιείται η συνέλιξη μεταξύ του σήματος και του διανύσματος k και διαιρείται κάθε φορά με το μήκος του επιλεγμένου

παραθύρου. Παρακάτω (σχήμα 3.12), δίνεται ένα ημιτονοειδές σήμα με προσθετικό θόρυβο και η απαλλαγή του από αυτόν, με τη χρήση του κινούμενου μέσου όρου, με τετραγωνικό παράθυρο μήκους 100 σημείων.



Σχήμα 3.12 Ημιτονοειδές σήμα με θόρυβο, οριζόντιος άξονας είναι τα δείγματα του σήματος και κατακόρυφος το πλάτος του. Παρακάτω έχουμε το ίδιο σήμα απαλλαγμένο από το θόρυβο πλέον με τη χρήση του κινητού μέσου όρου και με τετραγωνικό παράθυρο 100 δειγμάτων, οι άξονες είναι όμοιοι με τους προηγούμενους.

3.10 Μετασχηματισμός Hilbert/Hilbert Φάκελος (Envelope) – Αναλυτικό σήμα

Για κάθε σήμα $g(t)$, ο μετασχηματισμός Hilbert υπολογίζεται ως εξής:

$$H\{g(t)\} = g(t) * \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(\tau)}{t-\tau} d\tau = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(t-\tau)}{\tau} d\tau \quad (3.26)$$

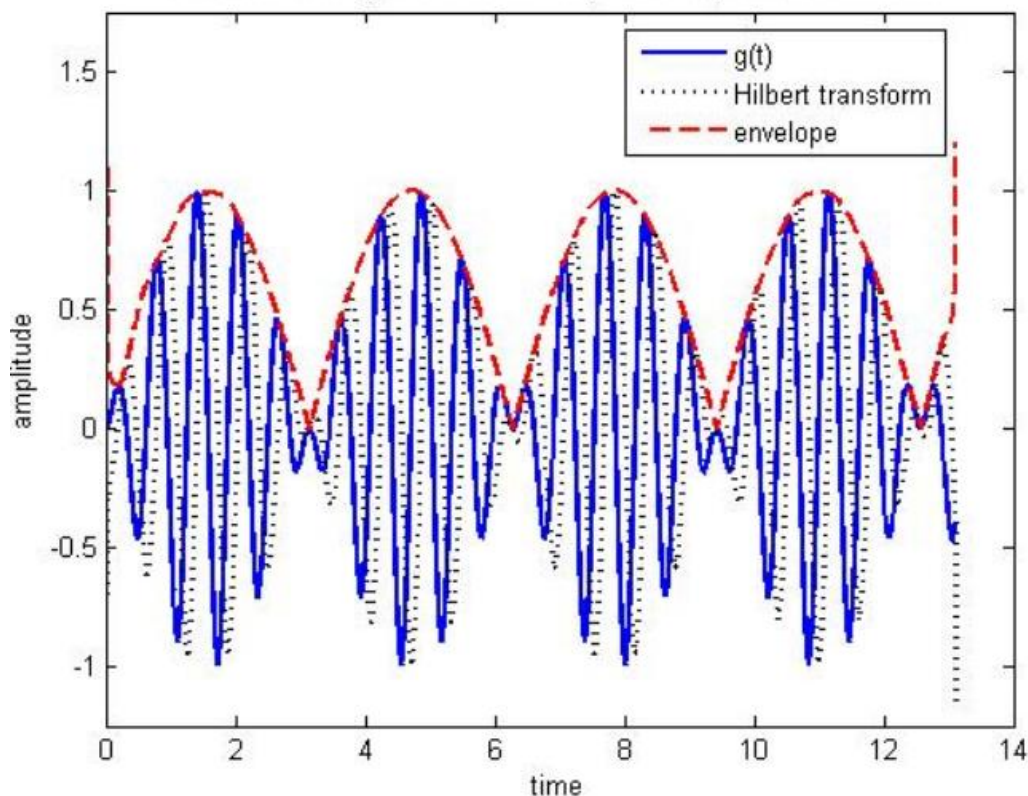
είναι δηλαδή το αποτέλεσμα της συνέλιξης της $g(t)$ με $1/\pi t$ και συνήθως συμβολίζεται με $\hat{g}(t)$ [30]. Φυσικά, για πραγματικά σήματα δεν είναι δυνατόν να

υπάρχουν άπειρα όρια, οπότε τα όρια είναι πεπερασμένα. Σύμφωνα λοιπόν με την αρχή του Cauchy [31] η παραπάνω σχέση γράφεται:

$$\lim_{\varepsilon \rightarrow 0^+} \left(\int_{t-1/\varepsilon}^{t-\varepsilon} \frac{g(\tau)}{t-\tau} d\tau + \int_{t+\varepsilon}^{t+1/\varepsilon} \frac{g(\tau)}{t-\tau} d\tau \right) \quad (3.27)$$

Ο φάκελος Hilbert (Hilbert Envelope) ή το αναλυτικό σήμα, είναι το πραγματικό μέρος του μετασχηματισμού Hilbert, δηλαδή $E(t) = |H\{g(t)\}|$ [30]. Στο παρακάτω

σχήμα 3.13 παρουσιάζεται ένα πραγματικό σήμα $g(t)$ και με κόκκινες διακεκομμένες γραμμές ο φάκελος [31].



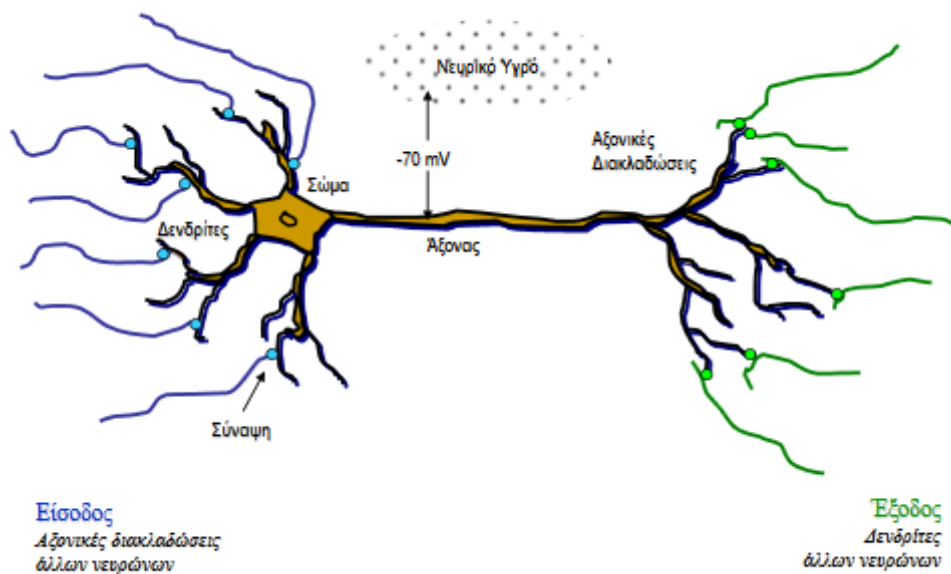
Σχήμα 3.13 Χρώμα μπλε είναι ένα σήμα, με μαύρη διακεκομμένη είναι ο Hilbert μετασχηματισμός του και με κόκκινες διακεκομμένες είναι ο φάκελος του σήματος. Στον οριζόντιο άξονα έχουμε χρόνο σε δευτερόλεπτα και στον κατακόρυφο το πλάτος του σήματος [31].

3.11 Νευρωνικά δίκτυα – Back-propagation Perceptron

Η δομή του ανθρώπινου εγκεφάλου αποτέλεσε τη βάση των νευρωνικών δικτύων. Η ικανότητά του στην παράλληλη επεξεργασία και την εξαγωγή συμπερασμάτων, όσο και στη λήψη αποφάσεων αποτελεί φαινόμενο μοναδικό. Το

κύριο πλεονέκτημά του είναι η δυνατότητα της παράλληλης επεξεργασίας δεδομένων και πληροφοριών σε αντίθεση με την τυπική λειτουργία του υπολογιστή να πραγματοποιεί σειριακή επεξεργασία. Εκτιμάται ότι αποτελείται παραπάνω από 10δισ διασυνδεδεμένων νευρώνων και ο μέσος άνθρωπος που δεν καταναλώνει αλκοόλ και δεν καπνίζει χάνει ημερησίως 1000 λιγότερους σε σχέση με τους άλλους. Μάλιστα πρόσφατα ανακαλύφθηκε πως οι νευρώνες αναγεννώνται. Τα τεχνητά νευρωνικά δίκτυα είναι η γενίκευση των μαθηματικών μοντέλων νευρικών βιολογικών συστημάτων και παρουσιάσθηκαν από τους McCulloch-Pitts το 1943 [32].

Στο παρακάτω σχήμα (σχ. 3.14), απεικονίζεται ο νευρώνας, ο οποίος αποτελείται από τους δενδρίτες, το σώμα, τις συνάψεις και τον άξονα. Δενδρίτες είναι ο συνδετικός κρίκος ενός νευρώνα με τον επόμενο και ονομάσθηκαν έτσι επειδή μοιάζουν με κλαδιά δέντρων. Οι δενδρίτες συνδέονται με τις αξονικές διακλαδώσεις, οι οποίες είναι προεκτάσεις του άξονα. Τα σημεία όπου ενώνονται οι δενδρίτες με τις αξονικές διακλαδώσεις ονομάζονται συνάψεις. Κάθε αξονική διακλάδωση καταλήγει στο σώμα, όπου κάθε σώμα έχει τον πυρήνα του [32].



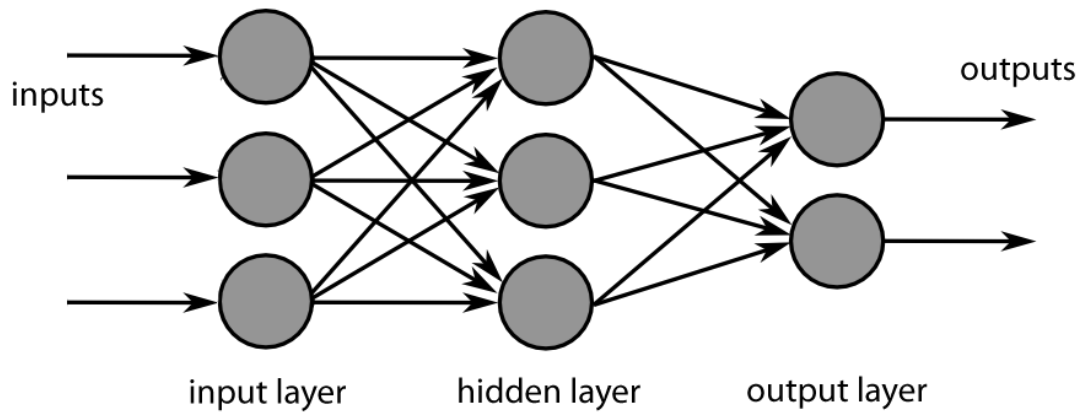
Σχήμα 3.14 Αναπαράσταση βιολογικού νευρώνα [32].

Η μετάδοση σημάτων από έναν νευρώνα σε έναν άλλον μέσω των συνάψεων, είναι μια πολύπλοκη χημική διαδικασία κατά την οποία στέλνονται χημικά στοιχεία. Το αποτέλεσμα είναι η αύξηση ή μείωση της τάσης του στατικού ηλεκτρισμού στον άξονα.

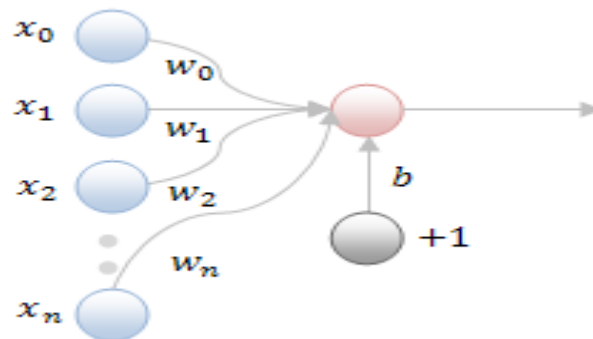
Τα τεχνητά νευρωνικά δίκτυα προέκυψαν ως απόπειρα προσομοίωσης της παράλληλης επεξεργασίας των νευρώνων του εγκεφάλου είτε με παραγωγή

λογισμικού είτε με κατασκευή hardware. Χρησιμοποιούνται ως ταξινομητές και για την ομαδοποίηση δεδομένων εισόδου.

Ανάλογα με τις εφαρμογές, έχουν αναπτυχθεί πολλά είδη νευρωνικών δικτύων. Στην εργασία αυτή, θα αναφερθούμε σε δίκτυα μιας κατεύθυνσης (forward) που αποτελούνται από 3 τμήματα/στρώματα: την είσοδο, το κρυφό στρώμα (N επιπέδων) και την έξοδο. Στο παρακάτω σχήμα (σχ. 3.15), παρουσιάζεται νευρωνικό δίκτυο με 3 εισόδους 1 κρυφό επίπεδο 3 κόμβων και 2 εξόδους [32].



Σχήμα 3.15 Η δομή του νευρωνικού δικτύου perceptron: το δίκτυο αποτελείται από το στρώμα εισόδου (input layer), το κρυφό στρώμα (hidden layer) και το στρώμα εξόδου (output layer) [32].



Σχήμα 3.16 Αναπαράσταση υπολογισμού σε κόμβο νευρωνικού δικτύου, όπου κάθε είσοδος πολλαπλασιάζεται με ένα τυχαία ορισμένο βάρος w και στη συνέχεια αθροίζεται με τον συντελεστή bias [33].

Τα νευρωνικά δίκτυα τύπου perceptron είναι καθοδηγούμενα δίκτυα που εκπαιδεύονται. Περιγράφονται από τις εισόδους, τα βάρη τα οποία μεταβάλλονται κατά τη διάρκεια της εκπαίδευσης, τον παράγοντα bias και τις εξόδους [33]. Κάθε είσοδος πολλαπλασιάζεται με το αντίστοιχο βάρος και αθροίζεται με το bias και τέλος, μέσω μιας συνάρτησης ενεργοποίησης, οδηγείται στην έξοδο. Για n εισόδους, περιγράφεται από την σχέση:

$$Net = \sum_{i=0}^n x_i * w_i + b \quad (3.28)$$

όπου x είναι το διάνυσμα εισόδων, w τα βάρη και b το bias (βλ. σχ. 3.16). Η έξοδος είναι πάντα 0 ή 1 και ανατροφοδοτεί την είσοδο με σκοπό την σύγκριση της με την επιθυμητή έξοδο. Τα βάρη στην αρχή ορίζονται τυχαία και ύστερα ανανεώνονται με βάση τον κανόνα δέλτα. Ο κανόνας αυτός δηλώνει ότι τα βάρη πρέπει να ανανεώνονται έως ότου υπάρξει σύγκλιση της εξόδου με την επιθυμητή, δηλαδή

$$\Delta w_i = c * x_i * \delta \quad (3.29)$$

όπου με δ ισούται η διαφορά της επιθυμητής εξόδου από την πραγματική και c είναι ο ρυθμός μάθησης. Πάντα ορίζουμε και ένα κατώφλι ανοχής σφάλματος (σφάλμα = έξοδος - επιθυμητή τιμή). Η συνάρτηση ενεργοποίησης βρίσκεται στο τμήμα εξόδου του στρώματος. Υπάρχει η γραμμική, η βηματική, όμως στην πειραματική διαδικασία όμως χρησιμοποιήθηκε η σιγμοειδής συνάρτηση ενεργοποίησης. Η γραμμική περιγράφεται από μια ευθεία γραμμή $y = a * x$, η βηματική ως εξής:

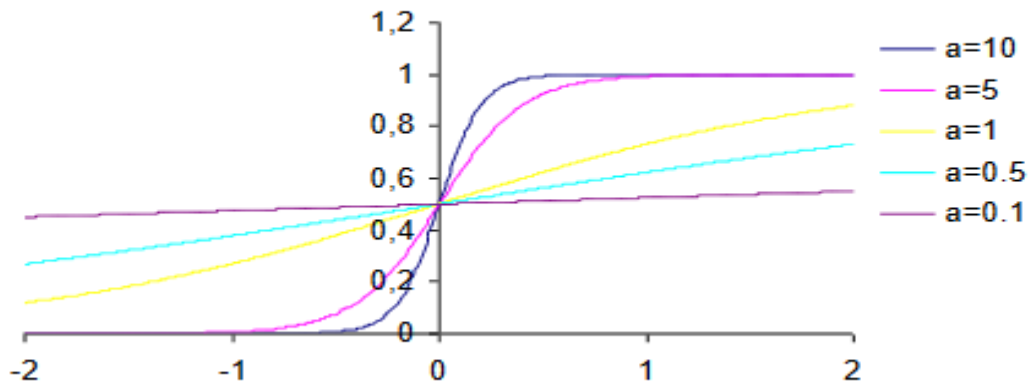
$$y = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (3.30)$$

και τέλος η σιγμοειδής περιγράφεται από την παρακάτω σχέση:

$$y = \frac{1}{1+e^{-a*x}} \quad (3.31)$$

Παρακάτω (σχήμα 3.17), δίνεται η σιγμοειδής συνάρτηση για διάφορες τιμές της παραμέτρου a , όπου η παράμετρος αυτή ρυθμίζει την κλίση της.

Σιγμοειδής συνάρτηση



Σχήμα 3.17 Παρουσίαση της σιγμοειδής συνάρτησης για διάφορες τιμές της παραμέτρου a .

Γενικά για τον νευρώνα q με p κρυφά επίπεδα και είσοδο x για την σιγμοειδή συνάρτηση ισχύει:

$$\delta_p(q) = x_p(q) * [1 - x_p(q)] * \sum w_{p+1}(q, i) * \delta_{p+1}(i) \quad (3.32)$$

Στο παράρτημα Β' δίνεται ο σχετικός αλγόριθμος (σχήμα Β.4) [34].

ΚΕΦΑΛΑΙΟ 4 : ΤΑΞΙΝΟΜΗΣΗ ΚΕΛΑΪΔΙΣΜΑΤΩΝ ΜΕ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

4.1 Βάση δεδομένων

Στην παρούσα εργασία επιλέχθηκαν καταγραφές από πουλιά της Ελλάδας, ώστε να μπορούν να μελετηθούν οι μέθοδοι σε πραγματικά δεδομένα της χώρας. Ωστόσο, πραγματοποιήθηκαν δοκιμές επίσης και από ηχογραφήσεις από γειτονικές μεσογειακές χώρες (Βουλγαρία, Ιταλία). Επιλέχθηκαν ωδικά πουλιά, συγκεκριμένα, η κοινή ευρωπαϊκή καρδερίνα (*Carduelis Carduelis*), το αηδόني (*Luscinia Megarhynchos*) και το τουρλί (*Lullula Arborea*). Η βάση δεδομένων αποτελείται από εκπαιδευτικά cd που λήφθηκαν μέσω του διαδικτύου και οι σύνδεσμοι δίνονται στην βιβλιογραφία ([35],..., [88]). Στην αρχή αποθηκεύτηκαν στην μορφή που ήταν διαθέσιμα και έπειτα μετατράπηκαν σε μορφή wav ενός καναλιού (mono). Επιπλέον, χρησιμοποιήθηκαν ηχογραφήσεις από κλουβιά σε διάφορες περιοχές της χώρας διαθέσιμα στο διαδίκτυο. Παρακάτω πίνακας Π4.1, δίνεται ένας ενδεικτικός πίνακας με ηχογραφήσεις από την ιστοσελίδα του xeno-canto για το αηδόνη. Έτσι για κάθε ηχογράφιση, παρουσιάζεται η χώρα και η τοποθεσία, η ημερομηνία, η ώρα και τέλος, το υψόμετρο όπου λήφθηκε η καθεμιά.

Πίνακας Π4.1: Ενδεικτικός πίνακας βάσης δεδομένων από την ιστοσελίδα του xeno-canto για το αηδόνη, όπου παρουσιάζονται η χώρα, η τοποθεσία, η ημερομηνία, η ώρα και το υψόμετρο.

α/α	Ελλάδα		Ιταλία		Βουλγαρία	
1	Λέσβος 14:30	14/4/15 800m	Tuscany 22:30	30/5/14 460m	Strelcha 19:43	25/5/15 480m
2	Λέσβος 8:00	14/5/14 10m	Turin 20:30	1/5/13 240m	Strelcha 19:58	25/5/15 480m
3	Λέσβος 13:00	8/5/14 10m	Verona 8:00	14/4/13 320m	Strelcha 17:00	25/5/15 480m
4	Ρέθυμνο 7:21	15/4/15 380m	Verona 8:00	14/4/13 550m	Strelcha 17:33	25/5/15 480m
5			Bocchigliero 9:54	2/6/12 1051m	Strelcha 18:50	26/5/15 480m
6			Tuscany 7:00	6/6/12 320m	Strelcha 15:52	1/6/15 480m
7			Verona 9:00	26/4/14 380m	Strelcha 19:19	1/6/15 480m

4.2 Προεργασία και τμηματοποίηση

Οι ηχογραφήσεις ποικίλουν στη διάρκεια τους, κάποιες είναι μόλις μερικά λεπτά ενώ κάποιες άλλες δευτερόλεπτα. Γι' αυτό, τον λόγο, εάν η ηχογράφιση έχει μεγάλη διάρκεια, συνιστάται η περικοπή της μέχρι το σημείο όπου τελειώνει το τραγούδι. Ύστερα, κάθε κομμάτι διαιρείται περαιτέρω σε μικρότερης διάρκειας τμήματα (chunks), όπου υπάρχει μια μικρή προσωρινή παύση. Κάθε τμήμα (chunk) έχει μελετηθεί και εξάγεται από αυτό μόνο το διάστημα της τρίλιας μέσω της ακοής, εφόσον υπάρχει, ειδάλλως η διαδικασία συνεχίζεται στο επόμενο τμήμα (chunk). Στη βάση δεδομένων της παρούσας εργασίας, δεν χρειάστηκε η παραπάνω διαδικασία και έτσι από κάθε ηχογράφιση απευθείας έχουν εξαχθεί μόνον οι τρίλιες. Με τον όρο τρίλιες εννοούνται οι γρήγορες εναλλαγές μεταξύ 2 συχνοτήτων ή η επανάληψη της ίδιας συχνότητας. Αποτελεί ορολογία της μουσικής [89], [90].

Στο στάδιο της ανάλυσης, χρησιμοποιήθηκε παράθυρο μήκους 128 δειγμάτων. Το παράθυρο αυτό κρίνεται κατάλληλο για την επεξεργασία του συγκεκριμένου τύπου σήματος, καθώς, για συχνότητα δειγματοληψίας 44.1 kHz, ο χρόνος που αντιστοιχεί στη διάρκεια του παραθύρου αυτού είναι $T = \frac{128}{44100}$, δηλαδή $T=2.9$ ms.

Εάν το διάστημα αυτό αντιστοιχισθεί στην περίοδο μιας ελάχιστης ανιχνεύσιμης συχνότητας, αυτή υπολογίζεται: $f_{min} = \frac{1}{2.9 \cdot 10^{-3}} = 344 \text{ Hz}$. Το συχνοτικό εύρος που

τραγουδούν τα πουλιά εκτείνεται από τα 518 Hz μέχρι τα 11.050 kHz, εκτός του οποίου ενδέχεται να υπάρχουν λίγες συλλαβές ή καθόλου, με κύριο χαρακτηριστικό τον θόρυβο [13]. Άρα πληρούνται οι ελάχιστες προδιαγραφές και επομένως το παράθυρο αυτό κρίνεται κατάλληλο για τις ηχογραφήσεις της παρούσας εργασίας.

Παρακάτω δίνονται οι πίνακες με τις διάρκειες από κάθε ηχογράφιση και το πλήθος από τις τρίλιες που αποθηκεύτηκαν από την κάθε μια, πρώτα από το YouTube (Π4.2, Π4.4, Π4.6) και ύστερα από το xeno-canto (Π4.3, Π4.5, Π4.7), για τα τρία εξεταζόμενα είδη πουλιών.

Αηδόνι

Πίνακας Π4.2: Διάρκεια και αριθμός τρίλιων από την ιστοσελίδα του YouTube για το αηδόνι.

a/a	Διάρκεια	Τρίλιες
1	5:21	21
2	8:09	44
3	4:57	27
4	6	48

Πίνακας Π4.3: Διάρκεια και αριθμός τρίλιων από την ιστοσελίδα του xeno-canto για το αηδόνι.

α/α	Ελλάδα		Ιταλία		Βουλγαρία	
	Διάρκεια	Τρίλιες	Διάρκεια	Τρίλιες	Διάρκεια	Τρίλιες
1	0:50	6	2:04	11	15:4	171
2	6:41	67	1:24	11	5:06	24
3	4	67	4:35	26	4:24	14
4	0:55	8	2:05	24	3:01	12
5			0:59	16	2:31	4
6			1:15	21	7:03	46
7			2:18	25	15:22	143

Καρδερίνα

Πίνακας Π4.4: Διάρκεια και αριθμός τρίλιων από την ιστοσελίδα του YouTube για την καρδερίνα.

α/α	Διάρκεια	Τρίλιες
1	10:16	41
2	3:10	15
3	8:35	28
4	10:11	57
5	16:51	69
6	8	61
7	4:10	27
8	6:02	20
9	8:11	30
10	5:14	36

Πίνακας Π4.5: Διάρκεια και αριθμός τρίλιων από την ιστοσελίδα του xeno-canto για την καρδερίνα.

α/α	Ελλάδα		Ιταλία		Βουλγαρία	
	Διάρκεια	Τρίλιες	Διάρκεια	Τρίλιες	Διάρκεια	Τρίλιες
1	0:19	2	0:34:1	4	1:18:2	3
2			1:09:2	2	2:51:1	7
3			0:17	4		
4			0:9:4	2		

Τουρλί

Πίνακας Π4.6: Διάρκεια και αριθμός τρίλιων από την ιστοσελίδα του YouTube για το τουρλί.

α/α	Διάρκεια	Τρίλιες
1	0:37:6	17
2	1:03	8
3	7:36	24
4	3:56	12
5	7	36
6	5:06	23

Πίνακας Π4.7: Διάρκεια και αριθμός τρίλιων από την ιστοσελίδα του xeno-canto για το τουρλί.

α/α	Ελλάδα		Ιταλία		Βουλγαρία	
	Διάρκεια	Τρίλιες	Διάρκεια	Τρίλιες	Διάρκεια	Τρίλιες
1	3:16:5	31	1	29	7	43
2			0:35:8	5	3:38	24
3			0:57:1	8	3:58:7	29
4			1:7	8	8:39:2	32
5			1:12:6	9		

4.3 Τμηματοποίηση Hilbert

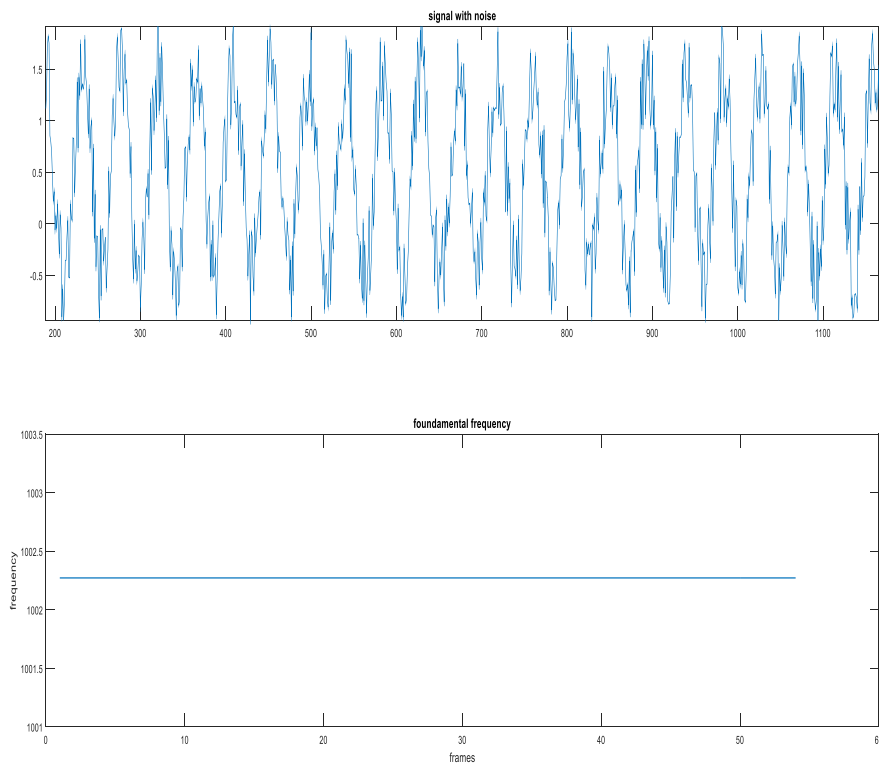
Στο επόμενο στάδιο, εφαρμόστηκε φιλτράρισμα με ζωνοπερατό φίλτρο με ζώνη διέλευσης 1kHz έως 12kHz, στην περιοχή δηλαδή όπου κυμαίνονται τα κελαηδήσμά τους. Κάθε φιλτραρισμένο τμήμα (chunk), τμηματοποιείται περαιτέρω με τη χρήση φακέλου Hilbert. Για ένα ορισμένο παράθυρο εστίασης μήκους D δειγμάτων, υπολογίζονται: α) το προσαρμοσμένο κατώφλι, β) το κατώφλι θορύβου και γ) το κατώφλι σήματος. Πρώτα τίθενται κάποιες αρχικές τιμές για κάθε κατώφλι, ύστερα σε κάθε επανάληψη, τα όρια αυτά ανανεώνονται και με βάση τις τελευταίες τιμές τους, αποθηκεύεται το κομμάτι της ηχογράφησης όπου το αναλυτικό σήμα είναι μεγαλύτερο από το προσωρινό κατώφλι. Εάν στο παράθυρο ανάλυσης δεν ισχύει το παραπάνω, εξετάζεται το εξής ενδεχόμενο: εφόσον ο φάκελος ξεπερνάει το προσωρινό κατώφλι, τότε ανανεώνεται το κατώφλι θορύβου. Δηλαδή, τα τμήματα του σήματος που βρίσκονται κάτω από το επιλεγμένο όριο θορύβου, αντιμετωπίζονται ως θόρυβος. Το κατώφλι σήματος προσδιορίζει την ελάχιστη τιμή στο παράθυρο εστίασης που μπορεί να χαρακτηριστεί ως σήμα/πληροφορία. Στην εργασία αυτή χρησιμοποιήθηκε παράθυρο εστίασης και παράθυρο εξομάλυνσης του φακέλου μήκους 220 δειγμάτων, πολλαπλάσιο της συχνότητας δειγματοληψίας ($0.005 \cdot 44.100$). Στο παράρτημα Β' (σχήμα Β.1) δίνεται το διάγραμμα ροής για την υλοποίηση του κώδικα [14].

4.4 Βασική/θεμελιώδης συχνότητα (Fundamental frequency)

Βασική ή θεμελιώδης θεωρείται η χαμηλότερη συχνότητα ενός περιοδικού σήματος ενώ όλες οι υπόλοιπες είναι πολλαπλάσιες της. Να αναφερθεί εδώ ότι υφίστανται ποικίλοι τρόποι για τον εντοπισμό/εκτίμησή της, μέσω της αυτοσυσχέτισης, των cepstrum, του HPS (Harmonic Product Spectrum) ή και μέσω ειδικά κατασκευασμένων φίλτρων [91]. Η τεχνική που χρησιμοποιήθηκε στην παρούσα εργασία είναι αυτή της αυτοσυσχέτισης.

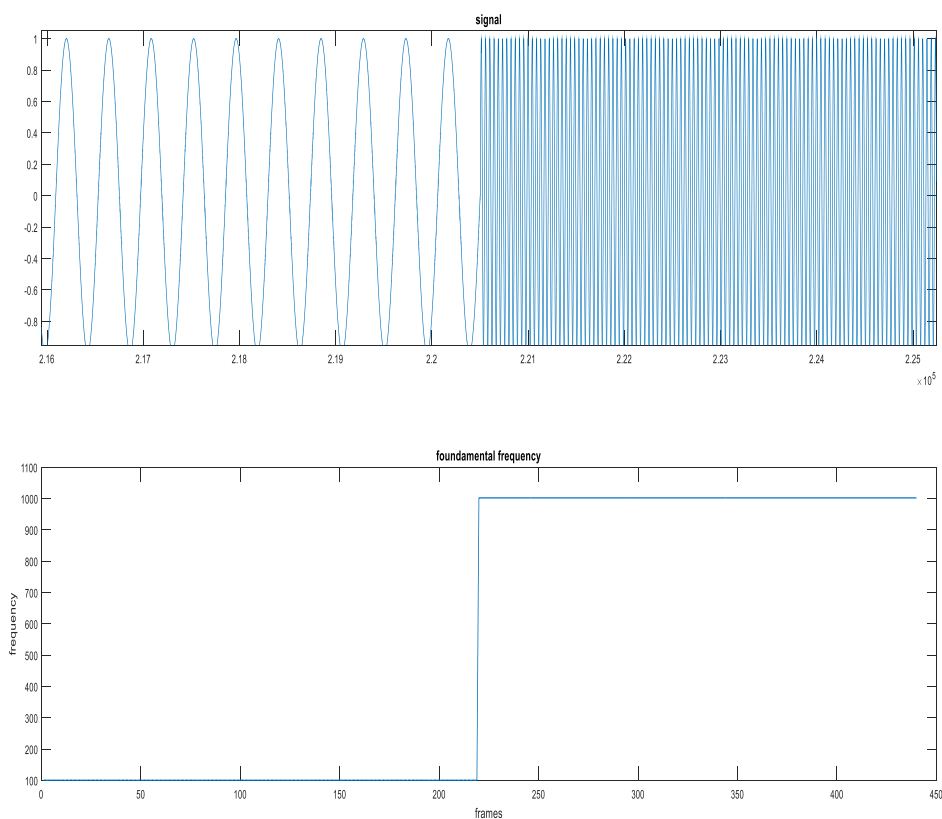
Πρώτα εφαρμόζεται φίλτρο προέμφασης με συντελεστή 0.97 και κατόπιν χωρίζεται το σήμα σε επικαλυπτόμενα κατά 50% πλαίσια (frames) μέσω hamming παραθύρου μήκους 128 δειγμάτων, δημιουργώντας τα αντίστοιχα διανύσματα. Σε κάθε frame υπολογίζεται η αυτοσυσχέτιση και σημειώνονται τα τοπικά μέγιστα με τις ακριβείς τοποθεσίες αυτών ενώ διαγράφεται το πρώτο στοιχείο του διανύσματος της αυτοσυσχέτισης, καθώς η αυτοσυσχέτιση πάντοτε παρουσιάζει το πρώτο μέγιστο της στο lag 0 (για μηδενική μετατόπιση). Το επόμενο βήμα περιλαμβάνει τον υπολογισμό των διαφορών των θέσεων των μεγίστων. Τέλος, υπολογίζεται ο μέσος όρος των διαφορών αυτών, η εκτιμώμενη θεμελιώδης συχνότητα είναι το αποτέλεσμα της διαίρεσης της συχνότητας δειγματοληψίας με τον μέσο όρο των διαφορών. Η διαδικασία αυτή επαναλαμβάνεται για κάθε frame και έτσι κάθε frame θα έχει μια αντιπροσωπευτική θεμελιώδη συχνότητα.

Ο αλγόριθμος αυτός έχει προσαρμοσθεί για την αποφυγή των NAN (Not A Number) τιμών που ενδέχεται να παρουσιάζονται, έτσι αποφεύγεται το πρόβλημα της πιθανής ασυνέχειας της θεμελιώδους συχνότητας. Επίσης είναι σχετικά απλός και δίνει ακριβή αποτελέσματα για σήματα με θόρυβο. Οι υπόλοιπες μέθοδοι που προαναφέρθηκαν παρουσιάζουν μεγαλύτερη ευαισθησία στον παράγοντα του θορύβου και περιλαμβάνουν περισσότερη πολυπλοκότητα. Παρακάτω (σχήμα 4.1), παρουσιάζεται ένα σήμα με συχνότητα 1000Hz με την επίδραση του θορύβου και η υπολογισμένη βασική συχνότητα του σήματος αυτού.



Σχήμα 4.1 Ημιτονοειδές ενθόρυβο σήμα συχνότητας 1kHz: στον οριζόντιο άξονα είναι τα δείγματα και στον κατακόρυφο το πλάτος του. Από κάτω, είναι η εκτιμώμενη θεμελιώδης συχνότητα όπου στον οριζόντιο άξονα έχουμε τα frames και στον κατακόρυφο η συχνότητα σε Hz.

Στο επόμενο σχήμα (σχ. 4.2) δίνεται παράδειγμα σήματος με εναλλαγή 2 συχνοτήτων, από τα 100Hz στα 1000 Hz. Όπως φαίνεται, τα αποτελέσματα είναι αρκετά ακριβή με ελάχιστες αποκλίσεις από τις πραγματικές τιμές.



Σχήμα 4.2 Ημιτονοειδές σήμα εναλλαγής 2 συχνοτήτων: στον οριζόντιο άξονα είναι ο αριθμός των δειγμάτων και στον κατακόρυφο το πλάτος. Ακολουθεί η θεμελιώδης συχνότητα με πολύ καλή ακρίβεια τιμών (στον οριζόντιο είναι τα frames και στον κατακόρυφο η συχνότητα σε Hz).

4.5 Εξαγωγή χαρακτηριστικών

Το στάδιο αυτό κρίνεται πολύ σημαντικό καθώς απαιτείται η ορθή εξαγωγή χαρακτηριστικών (features), τα οποία στη συνέχεια θα τροφοδοτήσουν τα κατάλληλα εργαλεία αναγνώρισης και ταξινόμησης. Έτσι, σε κάθε τρίλια θα εξετασθούν τα εξής χαρακτηριστικά: ο μέσος όρος της θεμελιώδους συχνότητας, η τυπική απόκλιση της θεμελιώδους, το εύρος της τρίλιας και η νευρικότητα της. Με τον όρο νευρικότητα εννοείται το πλήθος των εναλλαγών των δύο συχνοτήτων στη συνολική διάρκεια της τρίλιας. Κάθε τρίλια έχει ιδιαίτερη αντιμετώπιση, σχετικά με 3 διαφορετικές παραμέτρους: το παράθυρο της γκαουσιανής ομαλοποίησης (Nwin), (μέθοδος παρόμοια με τον κινητό μέσο όρο η διαφορά είναι ότι χρησιμοποιείται γκαουσιανό παράθυρο), το κατώφλι ελάχιστης απόστασης των μεγίστων (MPD - Min Peak Distance) και το όριο της ελάχιστης διαφοράς μεταξύ 2 γειτονικών μεγίστων (MPP - Min Peak Prominence) [92]. Οι δυο τελευταίες είναι παράμετροι της συνάρτησης findpeaks του Matlab, που χρησιμοποιούνται στο στάδιο της εύρεσης των μεγίστων και των θέσεων τους στην αυτοσυσχέτιση. Σε αρκετές περιπτώσεις, οι τιμές των

παραμέτρων αυτών μεταβάλλονται, παρόλο που οι τρίλιες προέρχονται από τις ίδιες ηχογραφήσεις, οπότε για κάθε ηχογράφιση χρησιμοποιείται ο μέσος όρος των παραμέτρων αυτών. Παρακάτω, παρουσιάζονται οι παράμετροι αυτές με την εξής σειρά: παράθυρο γκαουσιανής ομαλοποίησης (Nwin), min peak distance (MPD), min peak prominence (MPP), πρώτα για τις τρίλιες από τις ηχογραφήσεις που λήφθηκαν από το YouTube και στην συνέχεια από το xeno-canto (Π4.8, Π4.10, Π4.12 και Π4.9, Π4.11, Π4.13, αντίστοιχα).

Αηδόνι

Πίνακας Π4.8: Παράμετροι για την εξαγωγή χαρακτηριστικών στις ηχογραφήσεις από YouTube για το αηδόνι.

α/α	Παράμετροι		
	Nwin	MPD	MPP
1	28	10	60
2	28	10	50
3	19	8	54
4	14	4	50

Πίνακας Π4.9: Παράμετροι για την εξαγωγή χαρακτηριστικών στις ηχογραφήσεις από xeno-canto για το αηδόνι.

α/α	Ελλάδα			Ιταλία			Βουλγαρία		
	Nwin	MPD	MPP	Nwin	MPD	MPP	Nwin	MPD	MPP
1	30	10	50	16	10	25	20	10	50
2	14	12	25	20	10	25	10	5	25
3	14	8	25	15	8	80	20	5	25
4	14	10	25	15	10	25	5	5	25
5				14	6	25	5	5	25
6				14	8	50	12	5	50
7				14	8	25	20	5	25

Καρδερίνα

Πίνακας Π4.10: Παράμετροι για την εξαγωγή χαρακτηριστικών στις ηχογραφήσεις από YouTube για την καρδερίνα.

α/α	Παράμετροι		
	Nwin	MPD	MPP
1	14	6	53
2	11	10	50
3	12	10	50
4	12	9	50
5	12	18	50
6	18	20	50
7	9	7.5	50
8	20	5	50
9	15	5	50
10	10	10	50

Πίνακας Π4.11: Παράμετροι για την εξαγωγή χαρακτηριστικών στις ηχογραφήσεις από xeno-canto για την καρδερίνα.

α/α	Ελλάδα			Ιταλία			Βουλγαρία		
	Nwin	MPD	MPP	Nwin	MPD	MPP	Nwin	MPD	MPP
1	18	10	50	14	8	25	18	10	50
2				16	10	25	18	10	50
3				16	14	50			
4				14	8	25			

Τουρλί

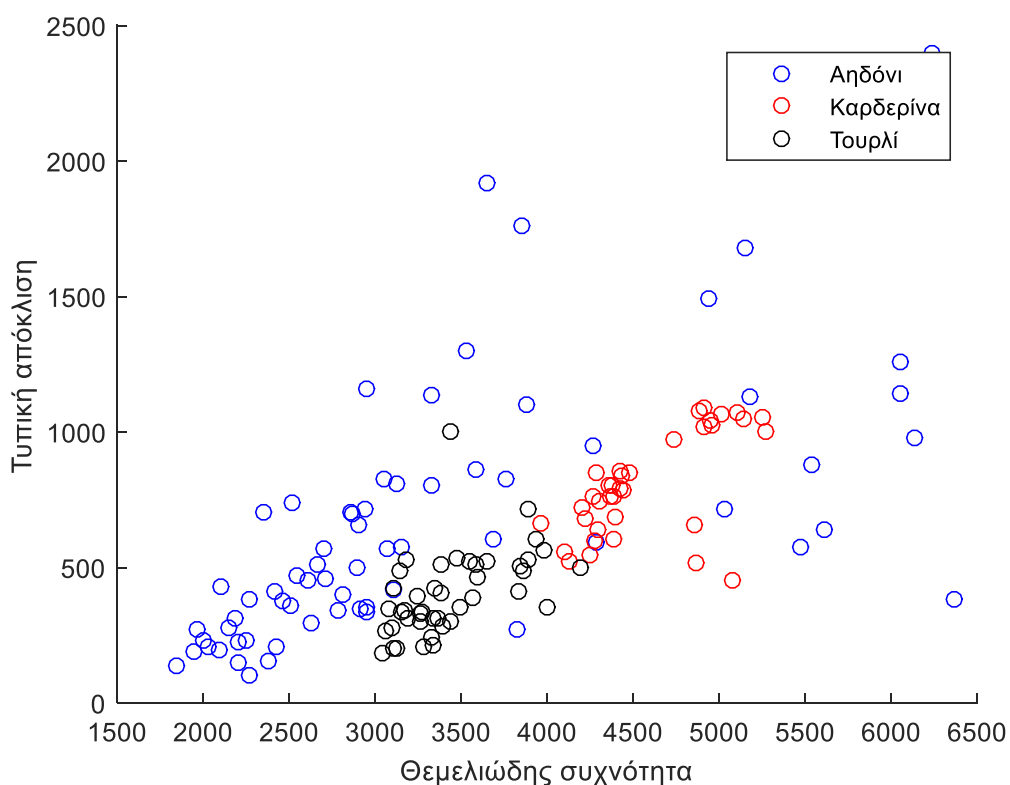
Πίνακας Π4.12: Παράμετροι για την εξαγωγή χαρακτηριστικών στις ηχογραφήσεις από YouTube για το τουρλί.

α/α	Παράμετροι		
	Nwin	MPD	MPP
1	20	10	25
2	9	10	50
3	30	17	50
4	5	15	50
5	14	5	50
6	16	10	50

Πίνακας Π4.13: Παράμετροι για την εξαγωγή χαρακτηριστικών στις ηχογραφήσεις από xeno-canto για το τουρλί.

α/α	Ελλάδα			Ιταλία			Βουλγαρία		
	Nwin	MPD	MPP	Nwin	MPD	MPP	Nwin	MPD	MPP
1	14	10	50	18	10	50	10	10	25
2				18	10	50	10	10	25
3				18	10	50	10	10	25
4				18	10	50	10	10	25
5				18	10	50			

Στο σχήμα 4.3 παρουσιάζονται 2 χαρακτηριστικά σε συνάρτηση μεταξύ τους από 3 ηχογραφήσεις για το κάθε είδος. Επιλέχθηκαν ως χαρακτηριστικά η θεμελιώδης συχνότητα με την τυπική της απόκλιση σε Hz: με μπλε είναι από το αηδόνη με μαύρο από το τουρλί και με κόκκινο από την καρδερίνα.



Σχήμα 4.3 Διάγραμμα διασποράς της θεμελιώδους συχνότητας με την τυπική απόκλισή της (σε Hz) από 3 ηχογραφήσεις. Με μπλε είναι τα δεδομένα από αηδόνη, με μαύρο για το τουρλί και με κόκκινο για την καρδερίνα.

4.6 Εκπαίδευση νευρωνικού δικτύου

Ο συνολικός αριθμός των ηχογραφήσεων που χρησιμοποιήθηκαν στα πλαίσια της παρούσας εργασίας είναι 22, 17 και 16, για αηδόνη, καρδερίνα και τουρλί, αντίστοιχα, από τις οποίες οι μισές χρησιμοποιήθηκαν για την εκπαίδευση του νευρωνικού δικτύου και οι υπόλοιπες, στην αναγνώριση. Δηλαδή, χρησιμοποιήθηκαν 11 ηχογραφήσεις από το αηδόνη, 8 από την καρδερίνα και 8 από το τουρλί στην εκπαίδευση, ενώ στην αναγνώριση 11, 9 και 8, αντίστοιχα. Μελετήθηκαν οι εξής 2 πειραματικές προσεγγίσεις:

1^η προσέγγιση: Από κάθε ηχογράφιση δημιουργήθηκαν συστάδες 3 γειτονικών στοιχείων τριλίων με τέσσερα χαρακτηριστικά (3 μέσοι όροι, 3 τυπικές αποκλίσεις, 3 νευρικότητες και 3 εύρη συχνότητας), οπότε κάθε είσοδος του νευρωνικού δικτύου τροφοδοτήθηκε από διανύσματα 12 γραμμών, για κάθε ηχογράφιση. Δηλαδή από κάθε ηχογράφιση επιλέχθηκαν τρεις διαδοχικές τρίλιες για να τροφοδοτήσουν με τα χαρακτηριστικά τους την είσοδο του νευρωνικού. Ο πίνακας εισόδου αποτελείται από 12 γραμμές με 223 στήλες για την εκπαίδευση του δικτύου. Όπου περισσεύαν στοιχεία, επειδή δεν χωρίζονταν ακριβώς με το 3, δεν χρησιμοποιήθηκαν. Η έξοδος είναι διάνυσμα 3 γραμμών με 223 στήλες. Έτσι για κάθε ηχογράφιση στο στάδιο της αναγνώρισης εξετάστηκαν τα αθροίσματα των εξόδων του δικτύου, κανονικοποιημένα προς το πλήθος των τριλίων. Ο χαρακτηρισμός για το αηδόνη είναι [1 0 0], για την καρδερίνα είναι [0 1 0] και για το τουρλί [0 0 1].

2^η προσέγγιση: Στο πείραμα αυτό, πρώτα έχουν εξαχθεί τα διανύσματα 4 γραμμών όπου κάθε γραμμή αντιστοιχεί σε καθένα χαρακτηριστικό. Η διαδικασία επαναλαμβάνεται για όλες τις ηχογραφήσεις και έτσι δημιουργείται ο πίνακας εισόδου 4 γραμμών με 708 στήλες. Στην συνέχεια τα διανύσματα αυτά τροφοδοτούν την είσοδο χωρίς κάποια περεταίρω επεξεργασία. Η έξοδος είναι διάνυσμα 3 γραμμών και 708 στηλών, όπου ο χαρακτηρισμός για το αηδόνη είναι [1 0 0], για την καρδερίνα είναι [0 1 0] και για το τουρλί [0 0 1]. Στο στάδιο της αναγνώρισης για κάθε ηχογράφιση αθροίζονται οι έξοδοι όλων των τριλίων και διαιρούνται με το πλήθος τους (κανονικοποίηση των εξόδων προς των πλήθος των τριλίων).

Όσον αφορά στην επιλογή των κρυφών επιπέδων, ανάλογα με την πολυπλοκότητα του προβλήματος (ως προς τις περιοχές απόφασης), επιλέγονται ένα ή περισσότερα. Στην παρούσα εργασία, η χρήση ενός κρυφού επιπέδου θεωρείται επαρκής. Όσο για τον αριθμό των κόμβων στο κρυφό επίπεδο χρησιμοποιήθηκαν τα εξής εμπειρικά κριτήρια [93]:

- Ο αριθμός των κόμβων πρέπει να είναι μεταξύ του αριθμού εισόδων και του αριθμού εξόδων.
- Να είναι τα 2/3 του αριθμού εισόδων αθροίζοντας με το μέγεθος των εξόδων.
- Ο αριθμός τους πρέπει να είναι μικρότερος του διπλάσιου μεγέθους εισόδων.

Τα κριτήρια αυτά δεν αποτελούν κανόνα, όμως βάσει αυτών πραγματοποιώντας δοκιμές υπολογίζεται ο ελάχιστος αριθμός κόμβων που δίνουν τα βέλτιστα αποτελέσματα. Στόχος, το νευρωνικό δίκτυο να είναι όσο πιο απλό γίνεται και αυτό συνεπάγεται λιγότερα κρυφά επίπεδα και κόμβους. Η διαδικασία αυτή ονομάζεται γενίκευση (generalization) [94].

4.7 Πειραματική διαδικασία

4.7.1 Πρώτη πειραματική προσέγγιση

Στο πείραμα αυτό, ο πίνακας εισόδου είναι 12 γραμμών με 223 στήλες, η έξοδος είναι 3 γραμμές με 223 στήλες. Το 70% των διανυσμάτων εισόδων χρησιμοποιήθηκε στην εκπαίδευση (train), το 15% για επαλήθευση (validation) και 15% για έλεγχο (test). Η εκπαίδευση περιλαμβάνει την εκπαίδευση του δικτύου με την ανανέωση των βαρών σε κάθε επανάληψη με στόχο το επιθυμητό σφάλμα, η επαλήθευση χρησιμοποιείται στην εύρεση βέλτιστων παραμέτρων και βαρών καθώς και για την αποφυγή της υπερεκπαίδευσης του νευρωνικού. Τέλος ο έλεγχος χρησιμοποιείται για τον επανέλεγχο ώστε η εκπαίδευση του δικτύου να γίνει χωρίς σφάλματα. Οι τιμές για την επαλήθευση και τον έλεγχο επιλέγονται τυχαία από τον πίνακα εισόδου μέσω του αλγορίθμου του νευρωνικού δικτύου. Έτσι το τελικό διάνυσμα εισόδου αποτελείται από 156 στήλες και 12 γραμμές, ενώ αντίστοιχα 33 στήλες αποτελούν το διάνυσμα της επαλήθευσης και ελέγχου. Λαμβάνοντας υπόψη τα εμπειρικά κριτήρια επιλογής του αριθμού των κόμβων που αναφέρθηκαν παραπάνω, πρώτη δοκιμή έγινε με 10 κόμβους στο κρυφό επίπεδο και έπειτα με 23 και με 3. Σε όλες τις περιπτώσεις είχαμε 0% επιτυχία για το αηδόνι και το τουρλί και 100% στην καρδερίνα. Επίσης, η είσοδος του δικτύου τροφοδοτήθηκε από 10 ηχογραφήσεις για αηδόνι, 6 καρδερίνα και τουρλί. Τότε στην περίπτωση αυτή παρατηρείται 100% επιτυχία στο αηδόνι και στην καρδερίνα ενώ στο τουρλί 0%. Οι

ηχογραφήσεις σε κάθε περίπτωση επιλέχθηκαν τυχαία για την εκπαίδευση και την αναγνώριση.

4.7.2 Δεύτερη πειραματική προσέγγιση

Στο πείραμα αυτό, η είσοδος είναι πίνακας 4 γραμμών και 708 στηλών και η έξοδος 3 γραμμές με 708 στήλες. Και πάλι, εκπαιδεύεται το 70% της εισόδου και το 30% χρησιμοποιείται στην επαλήθευση και στον έλεγχο. Ως συνάρτηση για τη σύγκλιση χρησιμοποιήθηκε η Levenberg-Marquardt backpropagation του Matlab και ο μέγιστος αριθμός των συνεχόμενων αποτυχιών ορίστηκε 20. Τέλος, ως παράμετρος επίδοσης δικτύου ορίστηκε το επιτρεπτό σφάλμα MSE κανονικοποιημένο στον μέσο όρο του πίνακα εκπαίδευσης.

Πάλι, σε συμφωνία με τα προαναφερθέντα κριτήρια επιλογής κόμβων, οι δοκιμές πραγματοποιήθηκαν με 4, 5 και 7 κόμβους. Παρακάτω, δίνονται οι πίνακες με τα αποτελέσματα. Τα αποτελέσματα αναγνώρισης είναι για αηδόνι καρδερίνα και τουρλί ανά ηχογράφηση δίνονται στο παρακάτω πίνακα Π4.14.

Πίνακας Π4.14: Αποτελέσματα αναγνώρισης με νευρωνικό 4 κόμβων.

α/α	Αηδόνι	Καρδερίνα	Τουρλί	Αηδόνι	Καρδερίνα	Τουρλί	Αηδόνι	Καρδερίνα	Τουρλί
1	0,57	0,12	0,3	0,22	0,53	0,24	0,12	0,18	0,69
2	0,35	0,16	0,48	0,77	0,2	0,01	0,14	0,07	0,78
3	0,71	0,2	0	0,62	0,38	0	0,1	0,15	0,74
4	0,93	0,04	0,02	0,08	0,91	0	0,17	0,06	0,76
5	0,62	0,2	0,18	0,09	0,9	0	0,1	0,02	0,87
6	0,42	0,45	0,12	0,57	0,42	0	0,09	0,07	0,83
7	0,86	0,01	0,04	0,16	0,76	0,06	0,15	0,1	0,73
8	0,8	0,16	0,03	0,13	0,54	0,32	0,13	0,06	0,8
9	0,7	0,22	0,07	0,18	0,8	0,02			
10	0,78	0,12	0,01						
11	0,6	0,26	0,12						
Επιτυχία	9/11			6/9			8/8		

Πίνακας Π4.15: Αποτελέσματα αναγνώρισης με νευρωνικό 5 κόμβων.

α/α	Αηδόνι	Καρδερίνα	Τουρλί	Αηδόνι	Καρδερίνα	Τουρλί	Αηδόνι	Καρδερίνα	Τουρλί
1	0,54	0,12	0,33	0,12	0,61	0,26	0,13	0,2	0,66
2	0,34	0,17	0,48	0,72	0,26	0	0,2	0,07	0,72
3	0,74	0,17	0	0,35	0,64	0	0,13	0,22	0,64
4	0,99	0	0	0,2	0,8	0	0,21	0,01	0,76
5	0,54	0,26	0,2	0,05	0,94	0	0,14	0	0,86
6	0,41	0,4	0,18	0,42	0,57	0	0,14	0,06	0,8
7	0,78	0,16	0,03	0,08	0,83	0,08	0,18	0,1	0,71
8	0,73	0,24	0,02	0,12	0,6	0,27	0,26	0,03	0,7
9	0,8	0,16	0,03	0,05	0,9	0,04			
10	0,76	0,16	0,08						
11	0,51	0,35	0,13						
Επιτυχία	10/11			8/9			8/8		

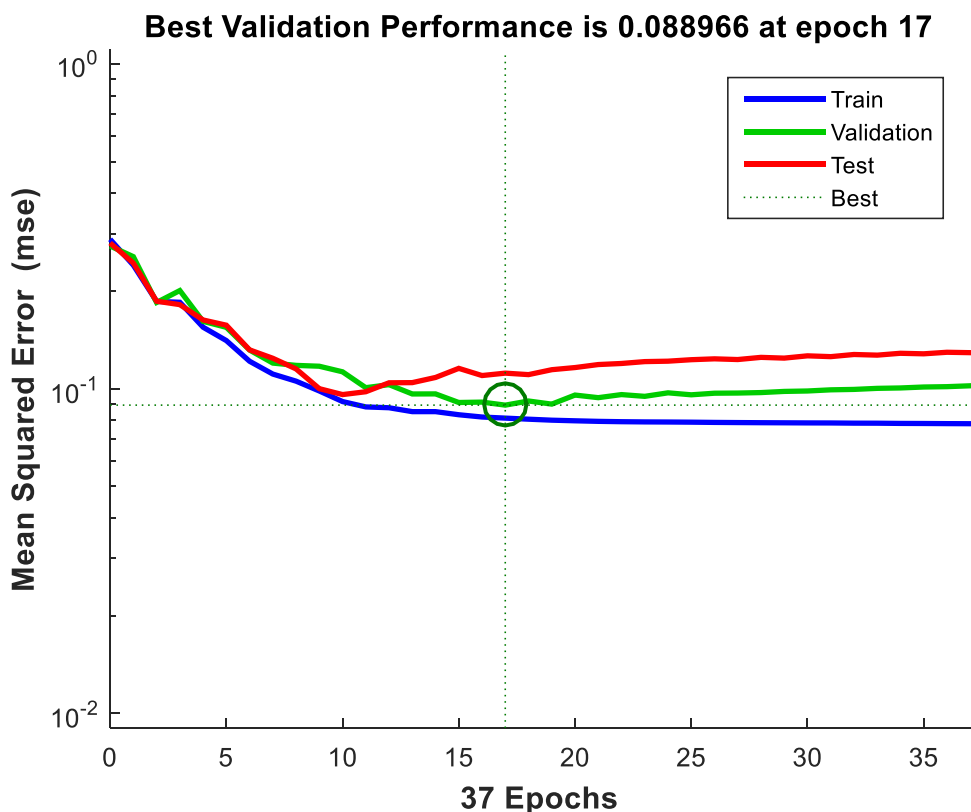
Πίνακας Π4.16: Αποτελέσματα αναγνώρισης με νευρωνικό 7 κόμβων.

α/α	Αηδόνι	Καρδερίνα	Τουρλί	Αηδόνι	Καρδερίνα	Τουρλί	Αηδόνι	Καρδερίνα	Τουρλί
1	0,58	0,11	0,3	0,21	0,6	0,2	0,18	0,13	0,68
2	0,38	0,12	0,5	0,74	0,24	0,02	0,24	0,05	0,71
3	0,75	0,15	0,01	0,31	0,6	0,09	0,19	0,11	0,69
4	0,99	0	0	0,21	0,78	0	0,24	0,01	0,73
5	0,6	0,2	0,2	0,03	0,96	0	0,16	0	0,84
6	0,5	0,34	0,15	0,4	0,58	0,01	0,2	0,04	0,75
7	0,83	0,13	0,04	0,09	0,81	0,1	0,28	0,07	0,65
8	0,74	0,2	0,05	0,3	0,44	0,26	0,31	0,02	0,66
9	0,8	0,16	0,04	0,06	0,9	0,04			
10	0,79	0,1	0,11						
11	0,67	0,2	0,12						
Επιτυχία	10/11			8/9			8/8		

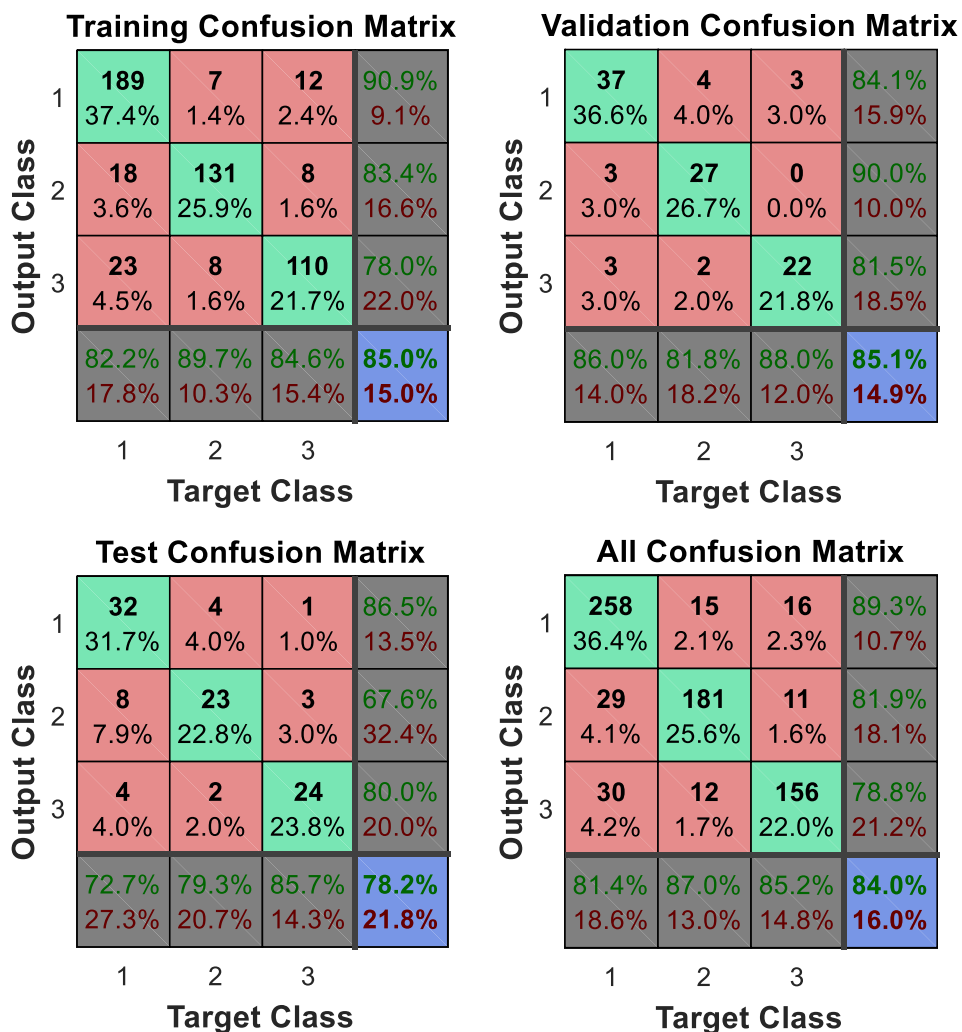
Συγκρίνοντας τους πίνακες διακρίνεται ότι η χρήση 4 κόμβων δεν βελτιστοποιεί την επιτυχία. Αντιθέτως με 5 ή 7 κόμβους λαμβάνεται η βέλτιστη, ωστόσο στους 5 κόμβους η αναγνώριση για το αηδόνι στην 6^η ηχογράφιση έχει την τιμή 0.41 με την αντίστοιχη της καρδερίνας να είναι 0.4. Το ποσοστό αυτό εκτός του ότι βρίσκεται οριακά το σωστό είδος, είναι και σχετικά μικρό (μόλις 0,41). Ποσοστό επιτυχές στην εργασία θεωρείται τουλάχιστον το 0,5. Στις υπόλοιπες ηχογραφήσεις υφίστανται ελάχιστες μεταβολές στα ποσοστά τα οποία δεν επηρεάζουν το τελικό αποτέλεσμα. Άρα ως βέλτιστη λύση επιλέγεται η χρήση 7 κόμβων. Για το δίκτυο αυτό δίνονται παρακάτω η επίδοση του δικτύου (σχήμα 4.4), ο πίνακας με τις επιτυχίες στην εκπαίδευση, στον έλεγχο, στη δοκιμή και ο συνολικός πίνακας (σχήμα 4.5). Επιπλέον, παρουσιάζεται η χαρακτηριστική απόκριση λειτουργίας του δικτύου σχήμα 4.6, όπου η πρώτη κατηγορία είναι το αηδόνι (με μπλε χρώμα), η δεύτερη είναι η καρδερίνα (με πράσινο) και η τρίτη το τουρλί (με κόκκινο).

Εδώ, αναφέρονται κατά φθίνουσα σειρά τα κριτήρια όταν σταματάει την σύγκλιση ο Levenberg-Marquardt backpropagation αλγόριθμος του Matlab [95]:

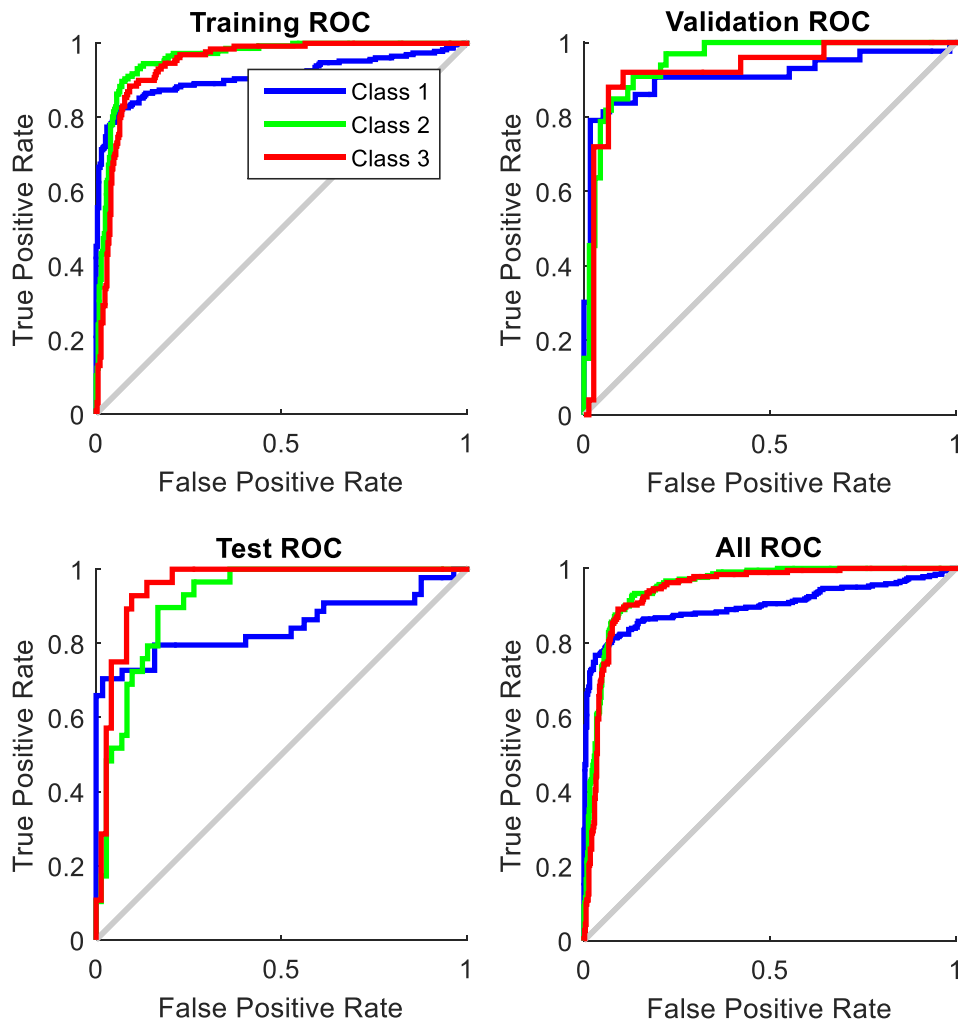
- Ο μέγιστος αριθμός των επαναλήψεων που έχουν ορισθεί, έχει ξεπερασθεί.
- Ο χρόνος εκτέλεσης της ρουτίνας του Matlab έχει περάσει το μέγιστο χρονικό περιθώριο.
- Η επίδοση του δικτύου ελαχιστοποιεί την επιθυμητή έξοδο.
- Η παράμετρος επίδοσης του δικτύου γίνει μικρότερη μιας προκαθορισμένης τιμής.
- Η παράμετρος σύγκλισης (μ) περάσει την προκαθορισμένη τιμή.
- Η επαλήθευση του δικτύου αποτύχει περισσότερες φορές από την προκαθορισμένη τιμή, όταν χρησιμοποιείται η επαλήθευση στο δίκτυο.



Σχήμα 4.4 Διάγραμμα επίδοσης δικτύου 7 κόμβων με βάση το MSE (μέσος όρος τετραγωνικού σφάλματος). Η μπλε γραμμή είναι για την εκπαίδευση, η πράσινη για τον έλεγχο και η κόκκινη για την δοκιμή. Στον κατακόρυφο είναι το mse σφάλμα και στον οριζόντιο οι εποχές σύγκλισης. Με διακεκομμένες είναι το ελάχιστο σφάλμα και η εποχή όπου συμβαίνει.



Σχήμα 4.5 Πίνακες με τα ποσοστά επιτυχίας στο στάδιο της εκπαίδευσης, του ελέγχου, της δοκιμής και του συνόλου τους. Με πράσινο χρώμα είναι τα στοιχεία που ταξινομήθηκαν σωστά ενώ στο κόκκινο τα λάθος. Σε κάθε πίνακα κατακόρυφα είναι οι έξοδοι και οριζόντια οι επιθυμητές έξοδοι, οι συστάδες 1,2 και 3 αντιστοιχούν σε κάθε κατηγορία (αηδόνι, καρδερίνα, τουρλί). Τα γκρι τετράγωνα είναι το άθροισμα ανά στήλη και με μπλε, τα συνολικά ποσοστά του κάθε πίνακα.



Σχήμα 4.6 Διαγράμματα απόκρισης χαρακτηριστικής λειτουργίας του δικτύου, μπλε είναι για το αηδόνι, πράσινο η καρδερίνα και κόκκινο το τουρλί , για την εκπαίδευση τον έλεγχο και τη δοκιμή. Στον οριζόντιο άξονα είναι ο ρυθμός λάθους και στον κατακόρυφο ο ρυθμός σωστής ταξινόμησης.

Τέλος, έχουμε τα βάρη της εισόδου, του κρυφού επιπέδου και τους συντελεστές bias για το δίκτυο αυτό, πίνακες Π.4.17, Π.4.18 και Π.4.19.

Πίνακας Π4.17: Συντελεστές βαρών εισόδου νευρωνικού 7 κόμβων.

Κόμβοι /Είσοδοι	1	2	3	4
1	5.8617	0.1989	0.8139	1.6244
2	-0.3993	-0.4703	1.5258	-0.7322
3	-5.1457	1.4662	0.2440	0.7898
4	7.6218	-0.7815	-3.9726	-0.4151
5	-2.6499	-1.2504	0.9480	-6.4920
6	-3.1714	-4.2778	-0.8022	5.0181
7	-0.2514	-2.0816	2.9531	-5.7917

Πίνακας Π4.18: Συντελεστές βαρών εξόδου νευρωνικού 7 κόμβων.

Έξοδοι/Κόμβοι	1	2	3	4	5	6	7
1	4.5107	-5.6116	7.4720	4.2199	0.0043	-2.5835	1.5088
2	-2.7966	1.8026	-3.5289	-1.1796	-3.9435	-1.0749	2.0299
3	-0.8955	3.8544	-4.8156	-2.7287	5.0435	4.2520	-4.4904

Πίνακας Π4.19: Συντελεστές βαρών bias νευρωνικού 7 κόμβων.

Κόμβοι/Κρυφά επίπεδα	1	2	3	4	5	6	7
1	0.3772	-2.4738	-1.6964	0.3507	-5.2104	2.1195	-2.7353

4.8 Συμπεράσματα – Προτάσεις μελλοντικής βελτίωσης

Στην εργασία μας χρησιμοποιήθηκαν ηχογραφήσεις από πτηνά σε κλουβιά όσο και από δάση. Απομονώθηκε όσο καλύτερα γινόταν η τρίλια από την υπόλοιπη ηχογράφηση μειώνοντας παράλληλα τα επίπεδα θορύβου. Επιλέχθηκε το πιο απλό μοντέλο νευρωνικού για την ταξινόμηση, ύστερα από δοκιμές. Τα αποτελέσματα με την χρήση backpropagation νευρωνικού δικτύου με 1 κρυφό επίπεδο 7 κόμβων προσφέρουν αρκετά αξιόπιστα αποτελέσματα. Στο αηδόνη η επιτυχία είναι 10/11, στην καρδερίνα 8/9 και στο τουρλί 8/8. Αν ληφθεί υπ' όψιν ότι τα τραγούδια τους διαφοροποιούνται ανά περιοχή, φύλο, ηλικία και ότι τα πειράματα πραγματοποιήθηκαν από ηχογραφήσεις γειτονικών χωρών, τότε τα αποτελέσματα είναι αρκετά ενθαρρυντικά για μελλοντική βελτίωση της μεθόδου αυτής. Όσο για την πρώτη πειραματική προσέγγιση, τα αποτελέσματα δεν μπορούν να θεωρηθούν ικανοποιητικά.

Ωστόσο, στη διαδικασία που ακολουθήθηκε, επιλέχθηκαν μόνο τρίλιες από τα τραγούδια τους και η διαδικασία εξαγωγής τους πραγματοποιήθηκε χειροκίνητα.

Όπως γίνεται αντιληπτό η διεργασία εξαγωγής απαιτεί μελέτη καθεμίας ηχογράφησης χωριστά, γεγονός που την καθιστά χρονοβόρα. Η μελέτη των μοτίβων των τρίλιων και η δημιουργία αυτόματου συστήματος εξαγωγής τους θα ήταν προτιμότερη. Στο στάδιο της προεργασίας και τμηματοποίησης, μελετήθηκαν για κάθε ηχογράφηση τα κατάλληλα μεγέθη παραθύρων για την ομαλοποίηση των θεμελιωδών συχνοτήτων. Και πάλι, αυτή η εργασία είναι χρονοβόρα, οπότε καλύτερη λύση θα αποτελούσε η μοντελοποίηση τους με σκοπό την αυτόματη επιλογή των μεγεθών αυτών. Επιπλέον, η μελέτη και κατηγοριοποίηση των ειδών μόνο μέσω της τρίλιας μπορεί να μην θεωρηθεί επαρκής, καθώς σε κάποιες ηχογραφήσεις ενδέχεται να μην έχουμε τρίλιες ή άλλα είδη ενδέχεται να μην παρουσιάζουν συχνά τρίλιες. Στην περίπτωση αυτή η εξαγωγή μοτίβων και προτύπων είναι αναγκαία κάνοντας πιθανόν την ταξινόμηση με Κρυμμένα Μαρκοβιανά Μοντέλα (Hidden Markov Models).

ΠΑΡΑΡΤΗΜΑ Α΄

Στο παράρτημα αυτό αναφέρονται τα πιο σημαντικά παράθυρα. Για την ευχέρεια της κατανόησής τους, θεωρείται ένας συγκεκριμένος αριθμός δειγμάτων N του αρχικού σήματος ενώ $w[n]$, είναι η συνάρτηση του εκάστοτε παραθύρου (βιβλιογραφικές πηγές: [96], [97], [98]).

Τετραγωνικό (Rectangular)

Το πιο απλό είναι το τετραγωνικό παράθυρο που σε μια συγκεκριμένη περιοχή το πλάτος του αντιστοιχεί με 1. Από μαθηματικής πλευράς, εκφράζεται ως

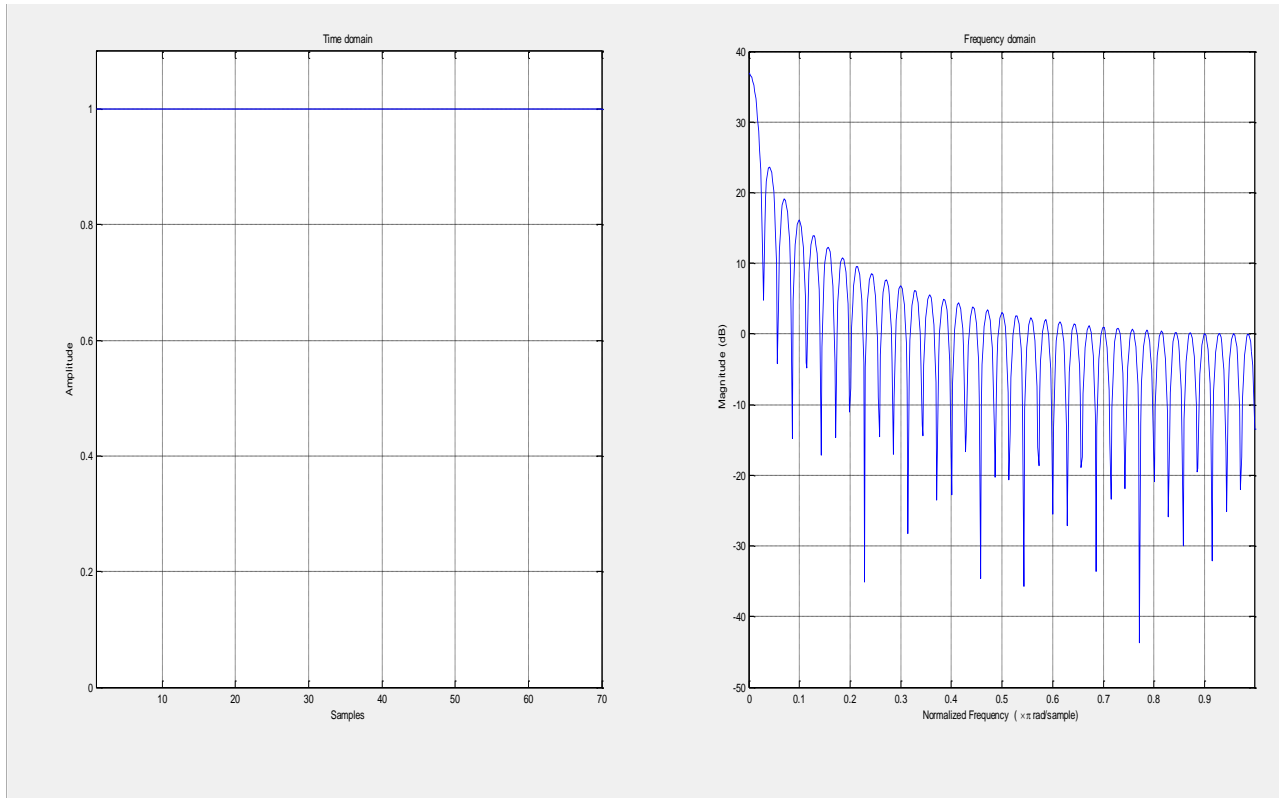
$$w[n] = \begin{cases} 1, & 1 \leq n \leq N \\ 0, & \text{αλλού} \end{cases} \quad (\text{A.1})$$

Στο σχήμα Α.1 παρουσιάζεται το παράθυρο στον χρόνο και στη συχνότητα για 70 δείγματα. Επισημαίνεται ότι η κανονικοποιημένη συχνότητα (Normalized Frequency) υπολογίζεται από τη σχέση

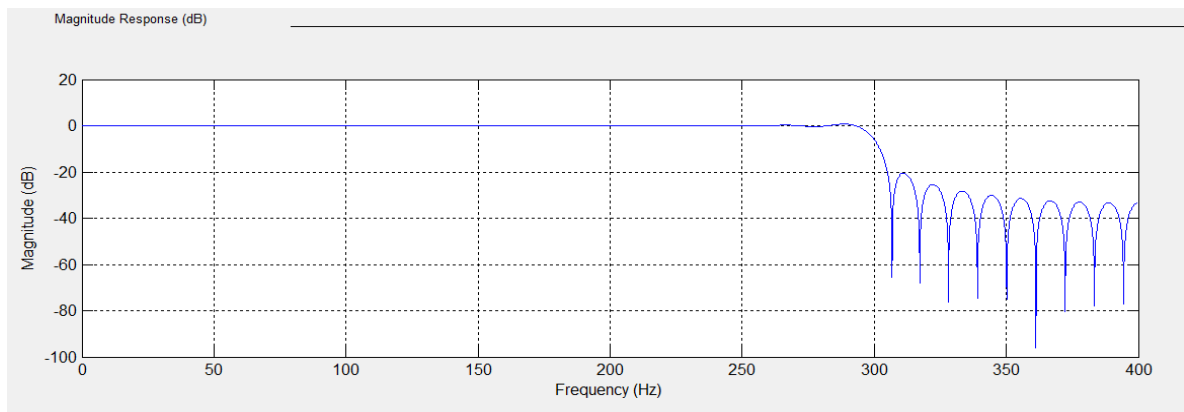
$$F = \frac{f_{\text{norm}} * Fs}{2\pi} \quad (\text{A.2})$$

όπου f_{norm} είναι η κανονικοποιημένη συχνότητα και Fs η συχνότητα δειγματοληψίας.

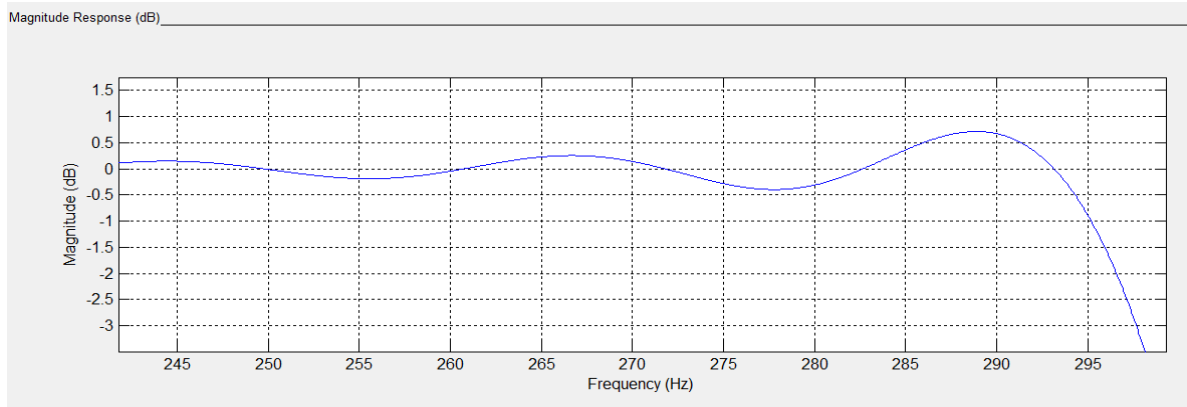
ΠΑΡΑΡΤΗΜΑ Α'



Σχήμα Α.1 Διάγραμμα τετραγωνικού παραθύρου σε δείγματα με πλάτος 1 και της ομαλοποιημένης συχνοτικής του απόκρισης.



Σχήμα Α.2.1 Απόκριση του τετραγωνικού στην συχνότητα. Στον κατακόρυφο άξονα, το πλάτος σε dB και στον οριζόντιο, η συχνότητα σε Hz.



Σχήμα Α.2.2 Απόκριση του τετραγωνικού στην συχνότητα. Στον κατακόρυφο άξονα, το πλάτος σε dB και στον οριζόντιο η συχνότητα σε Hz. Με την οπτική εστίαση φαίνονται οι κυματώσεις.

Στο σχήμα Α.2.1 φαίνεται η απόκριση του παραθύρου σε κλίμακα dB ενώ στο Α.2.2 έχουμε τις κυματώσεις (ripples) και μάλιστα σε έντονο βαθμό αφού η ιδανική απόκριση πρέπει να έχει πλάτος 0 dB, που όπως διακρίνεται ξεπερνάει τα 0.5 dB.

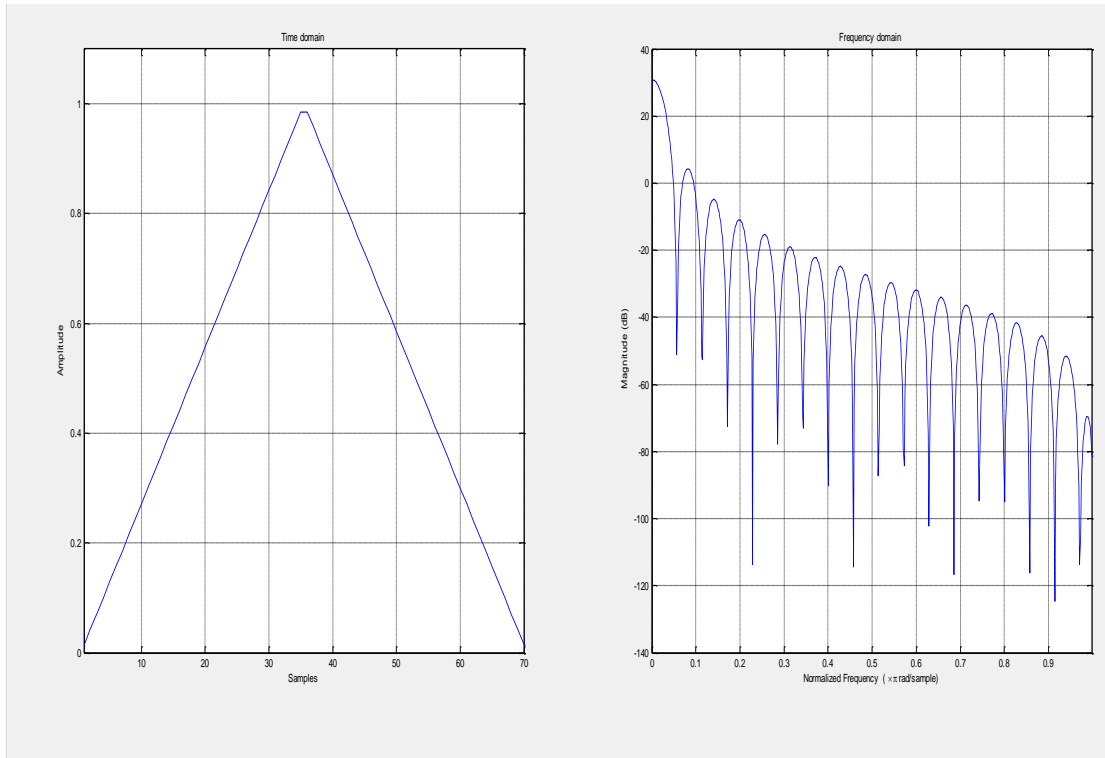
Τριγωνικό (Triangular)

Το τριγωνικό παράθυρο είναι το επόμενο που θα μελετηθεί και όπως θα διαπιστωθεί είναι σαφώς καλύτερο από το τετραγωνικό. Εκφράζεται ως

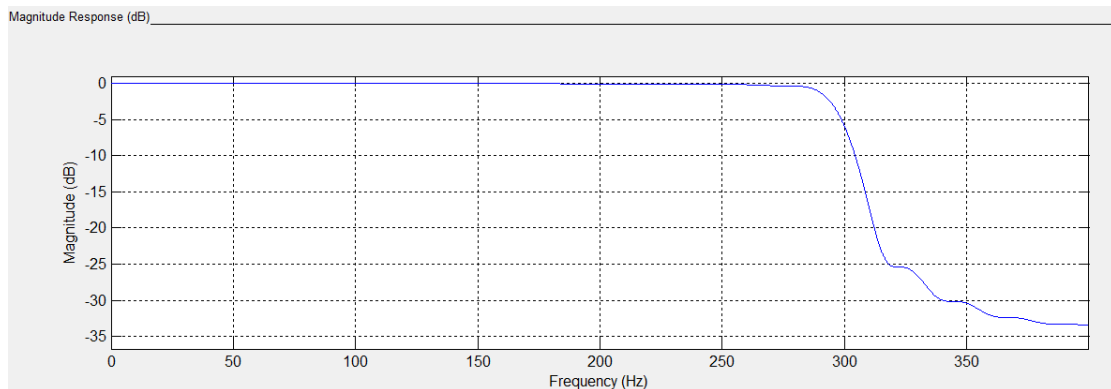
$$w[n] = 1 - \left| \frac{n - \frac{N-1}{2}}{N} \right| \quad (\text{A.3})$$

Στο σχήμα Α.3.1 παρουσιάζεται η τριγωνική συνάρτηση στο πεδίο του χρόνου και της συχνότητας. Τέλος, στο Α.3.2 δίνεται η απόκρισή του και έχει μειωμένες κυματώσεις και πιο ομαλούς δευτερεύοντες λοβούς σε σχέση με το τετραγωνικό παράθυρο.

ΠΑΡΑΡΤΗΜΑ Α΄



Σχήμα Α.3.1 Διάγραμμα τριγωνικού παραθύρου σε δείγματα με πλάτος 1 και της ομαλοποιημένης συχνοτικής του απόκρισης.



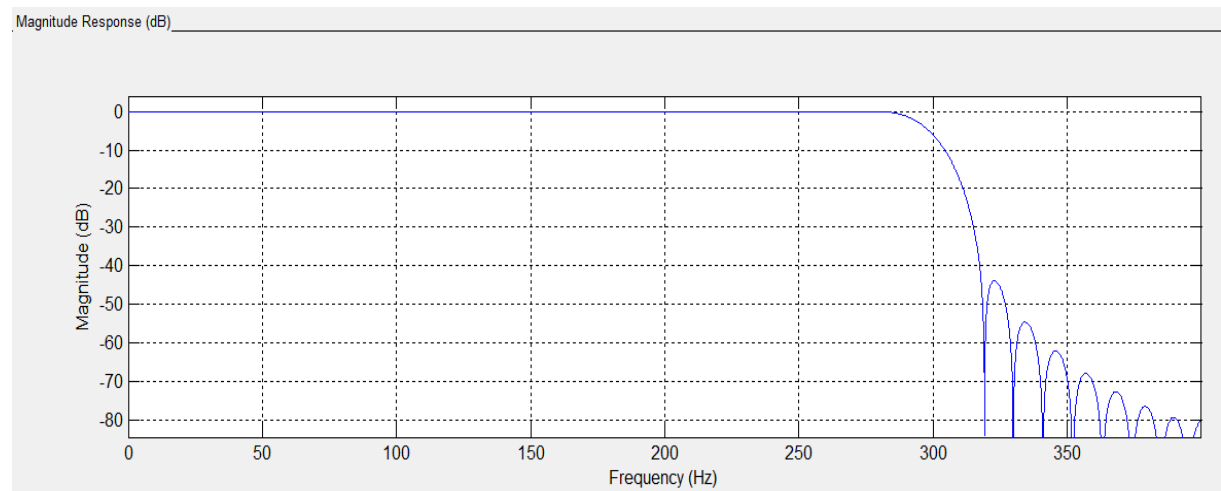
Σχήμα Α.3.2 Απόκριση του τριγωνικού στην συχνότητα. Στον κατακόρυφο άξονα, το πλάτος σε dB και στον οριζόντιο, η συχνότητα σε Hz.

Hanning (Von Hann)

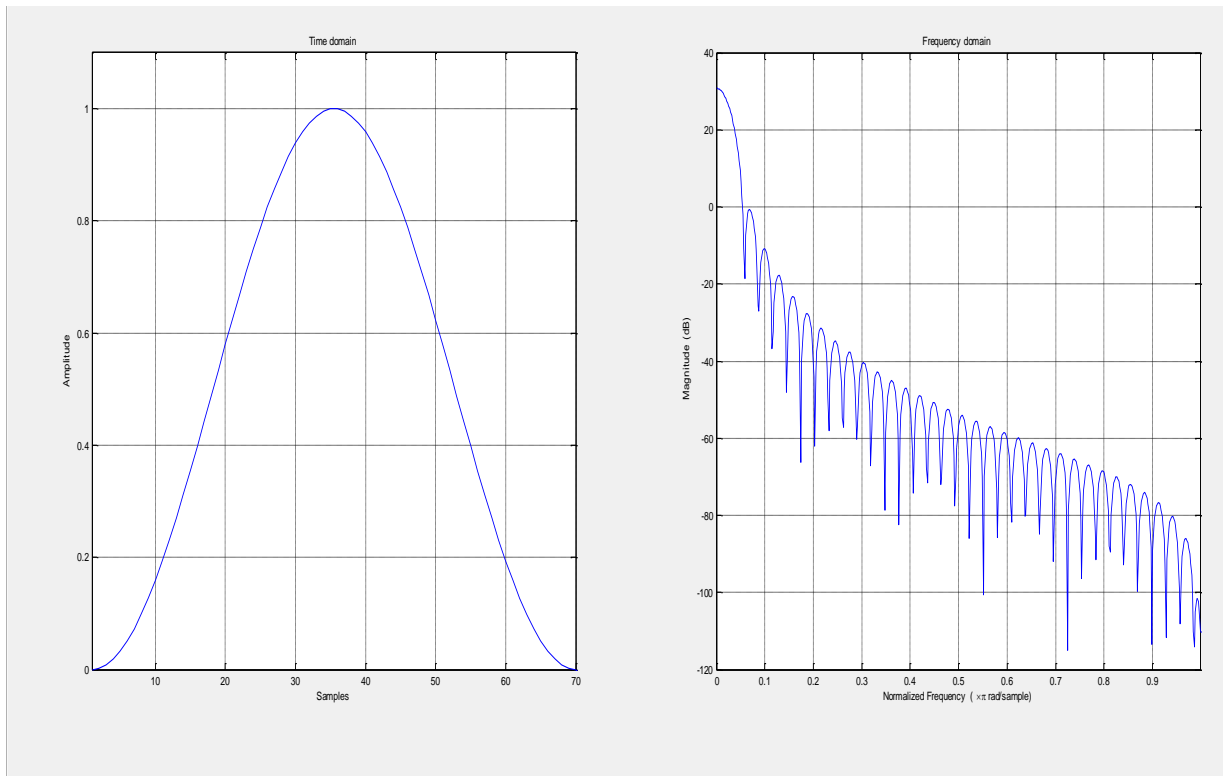
Έχει ημιτονοειδή μορφή και περιγράφεται από τον τύπο:

$$w[n] = 0.5 \left(1 - \cos \left(\frac{2\pi n}{N-1} \right) \right) \quad (\text{A.4})$$

Από τα σχήματα Α.4.1 και Α.4.2 παρατηρείται ότι δεν παρουσιάζεται προβληματικό με τις κυματώσεις και τους λοβούς όπως επίσης, σε σχέση με το τριγωνικό, είναι καλύτερο.



Σχήμα Α.4.1 Απόκριση του hanning στην συχνότητα. Στον κατακόρυφο άξονα, το πλάτος σε dB και στον οριζόντιο, η συχνότητα σε Hz.



Σχήμα Α.4.2 Διάγραμμα hanning παραθύρου σε δείγματα με πλάτος 1 και της ομαλοποιημένης συχνοτικής του απόκρισης.

Hamming

Έχει κατασκευασθεί ώστε να επιτευχθεί η ελαχιστοποίηση του κύριου λοβού. Η σχέση στο χρόνο είναι παρόμοια με την hanning καθώς κι αυτή έχει ημιτονοειδή μορφή και δίνεται από τον τύπο

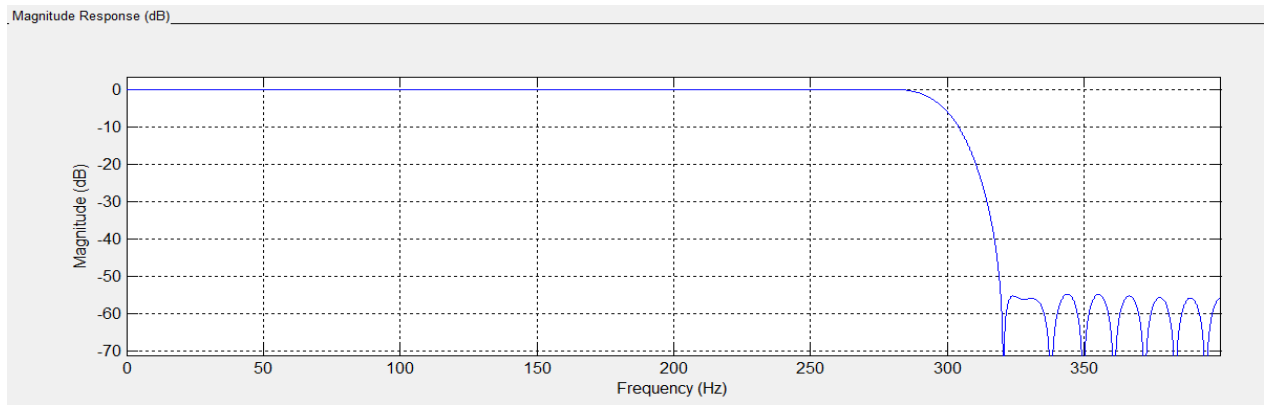
$$w[n] = \alpha - \beta \cos\left(\frac{2\pi n}{N-1}\right) \quad (\text{A.5.1}),$$

όπου η παράμετρος α παίρνει διάφορες τιμές και η β υπολογίζεται ως εξής:

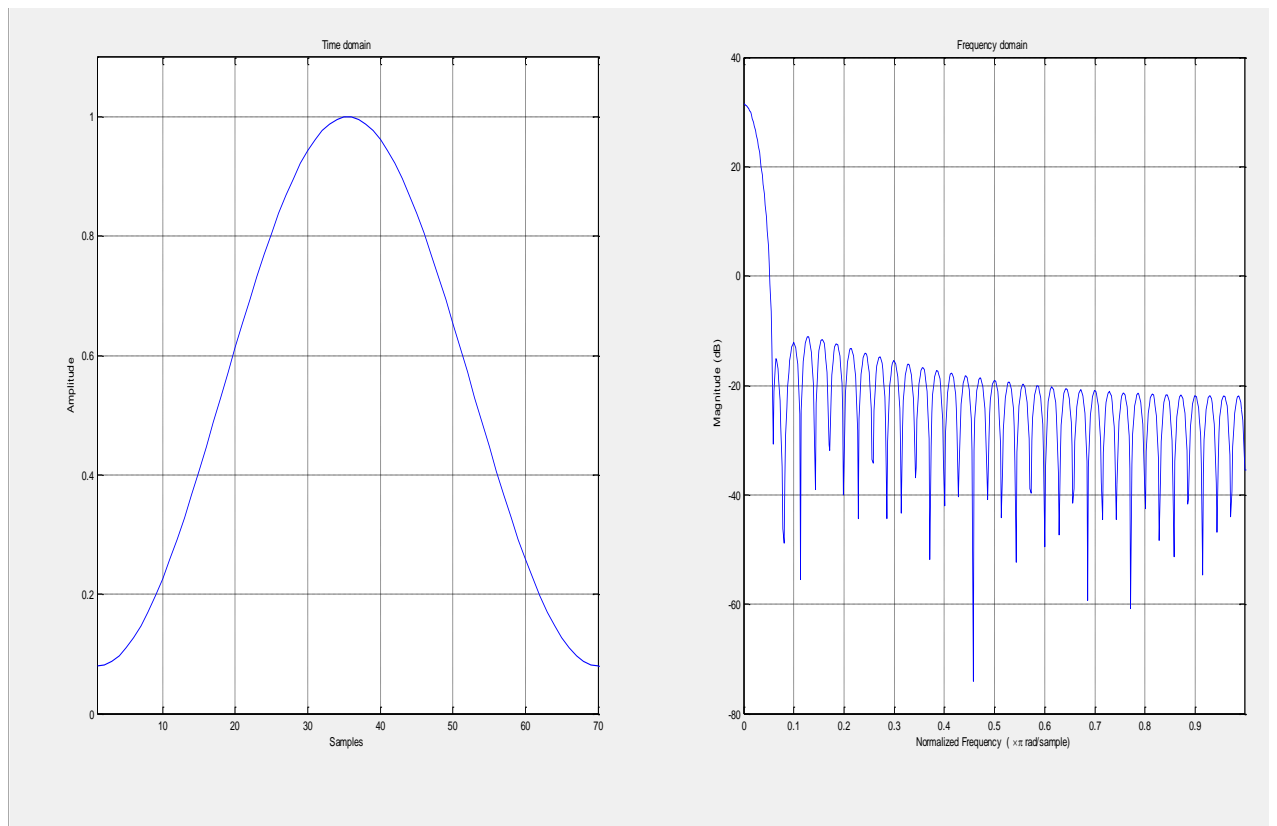
$$\beta = 1 - \alpha \quad (\text{A.5.2}).$$

Στα παρακάτω σχήματα, παρουσιάζονται οι αποκρίσεις συχνότητας για τις παραμέτρους $\alpha = 0.54$ και $\beta = 0.46$. Και αυτό το παράθυρο έχει πολύ καλή

απόκριση όσον αφορά στις κυματώσεις, καθώς διακρίνεται το μειωμένο πλάτος του πρώτου λοβού.



Σχήμα Α.5.1 Απόκριση του hamming στην συχνότητα. Στον κατακόρυφο άξονα, το πλάτος σε dB και στον οριζόντιο, η συχνότητα σε Hz.



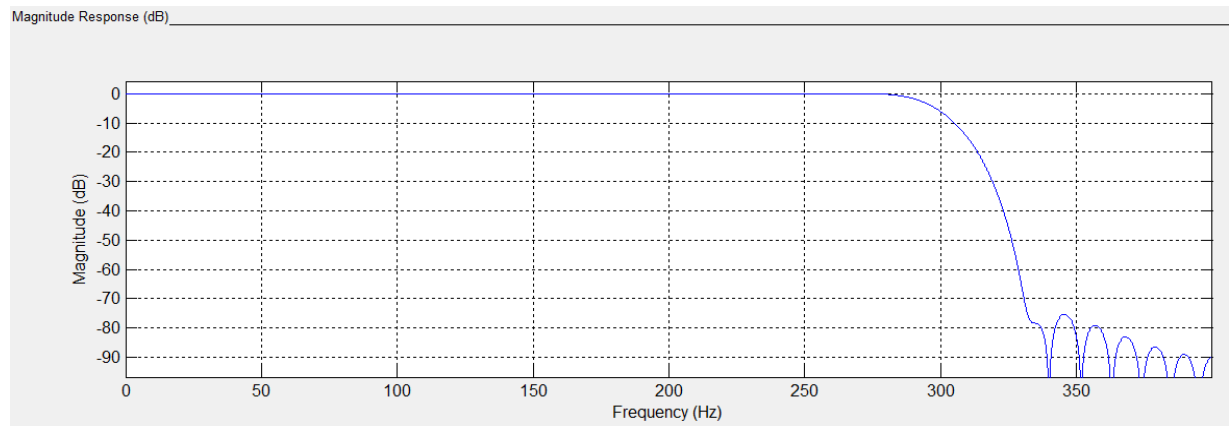
Σχήμα Α.5.2 Διάγραμμα hamming παραθύρου σε δείγματα με πλάτος 1 και της ομαλοποιημένης συχνοτικής του απόκρισης.

Blackman

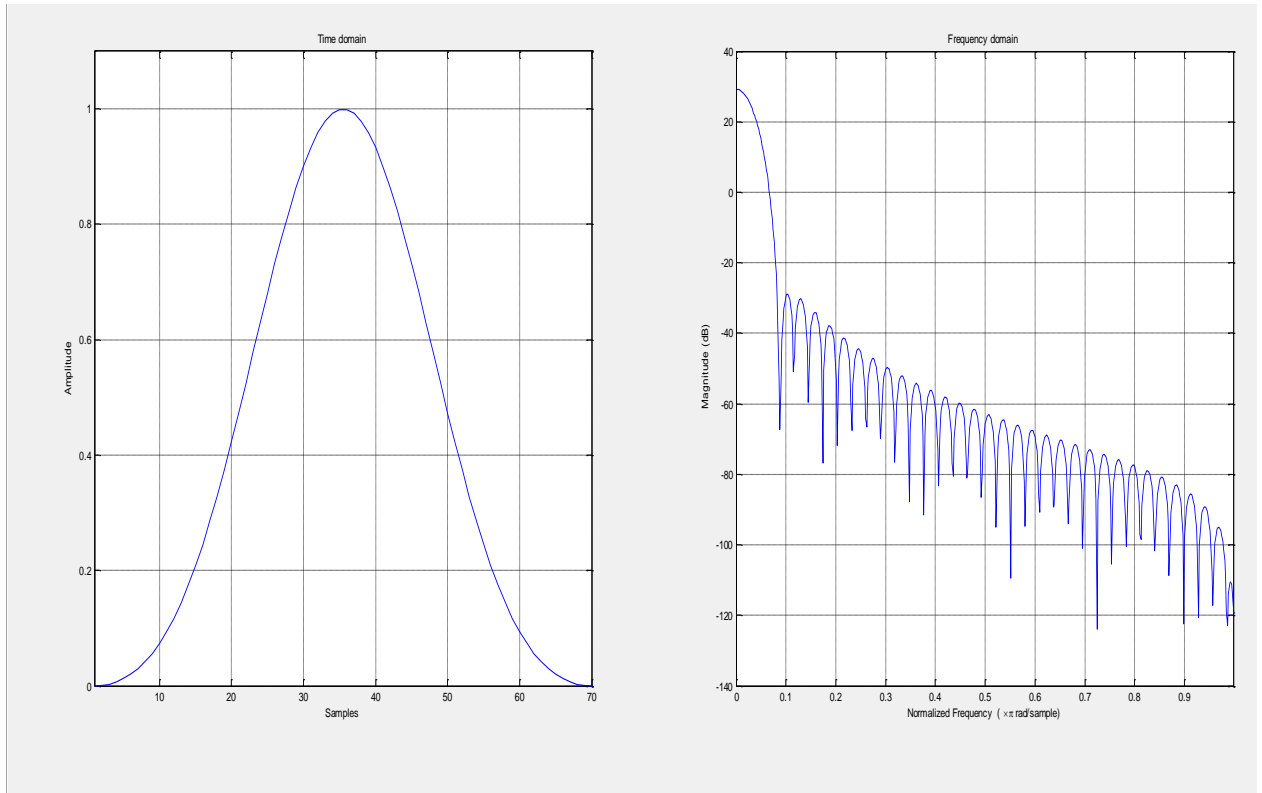
Το παράθυρο αυτό δίνεται από τον τύπο

$$w[n] = \alpha_0 - \alpha_1 \cos\left(\frac{2\pi n}{N-1}\right) + \alpha_2 \cos\left(\frac{4\pi n}{N-1}\right) \quad (\text{A.6})$$

όπου $\alpha_0 = \frac{1-\alpha}{2}$, $\alpha_1 = \frac{1}{2}$, $\alpha_2 = \frac{\alpha}{2}$, για $\alpha = 0.16$. Όπως και οι δυο προηγούμενες συναρτήσεις παραθύρων έτσι κι αυτή έχει ημιτονοειδή μορφή.



Σχήμα Α.6.1 Απόκριση του παραθύρου blackman στη συχνότητα. Στον κατακόρυφο άξονα, το πλάτος σε dB και στον οριζόντιο, η συχνότητα σε Hz.



Σχήμα Α.6.2 Διάγραμμα blackman παραθύρου σε δείγματα με πλάτος 1 και της ομαλοποιημένης συχνοτικής του απόκρισης.

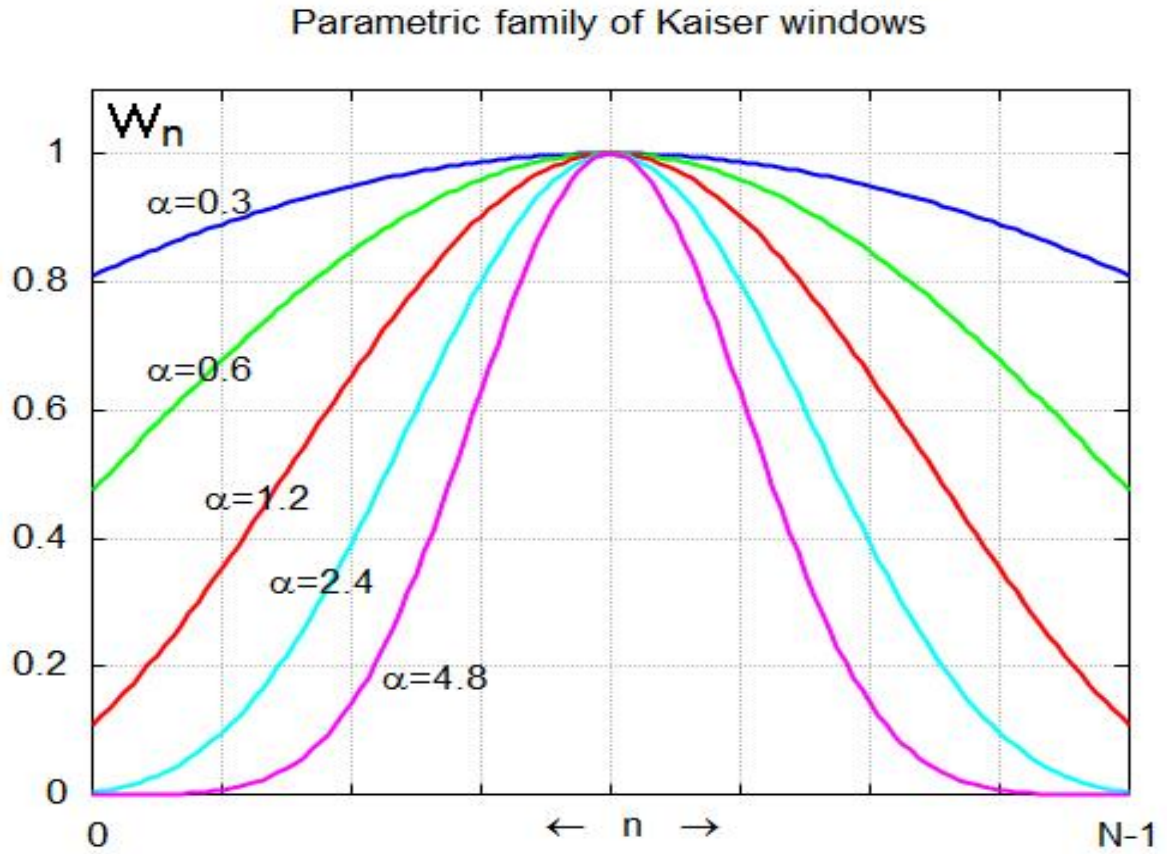
Kaiser

Τέλος, δίνονται τα παράθυρα Kaiser τα οποία είναι πολύ σημαντικά, από την άποψη ότι προσδιορίζονται οι επιτρεπόμενες κυματώσεις (ripples) όταν κατασκευάζονται φίλτρα με τη μέθοδο της παραθύρωσης. Η δυνατότητα αυτή είναι πολύ χρήσιμη και εξίσου σημαντική, επειδή στις πραγματικές εφαρμογές απαιτούνται προσαρμοσμένα παράθυρα. Τα παράθυρα αυτά λόγω της παραμέτρου α , που

αναλύεται παρακάτω, δημιουργεί μια οικογένεια παραμετρικών παραθύρων της οικογένειας Kaiser και έτσι δίνεται η δυνατότητα της επιλογής κάθε φορά του καταλληλότερου. Εκφράζονται από την σχέση

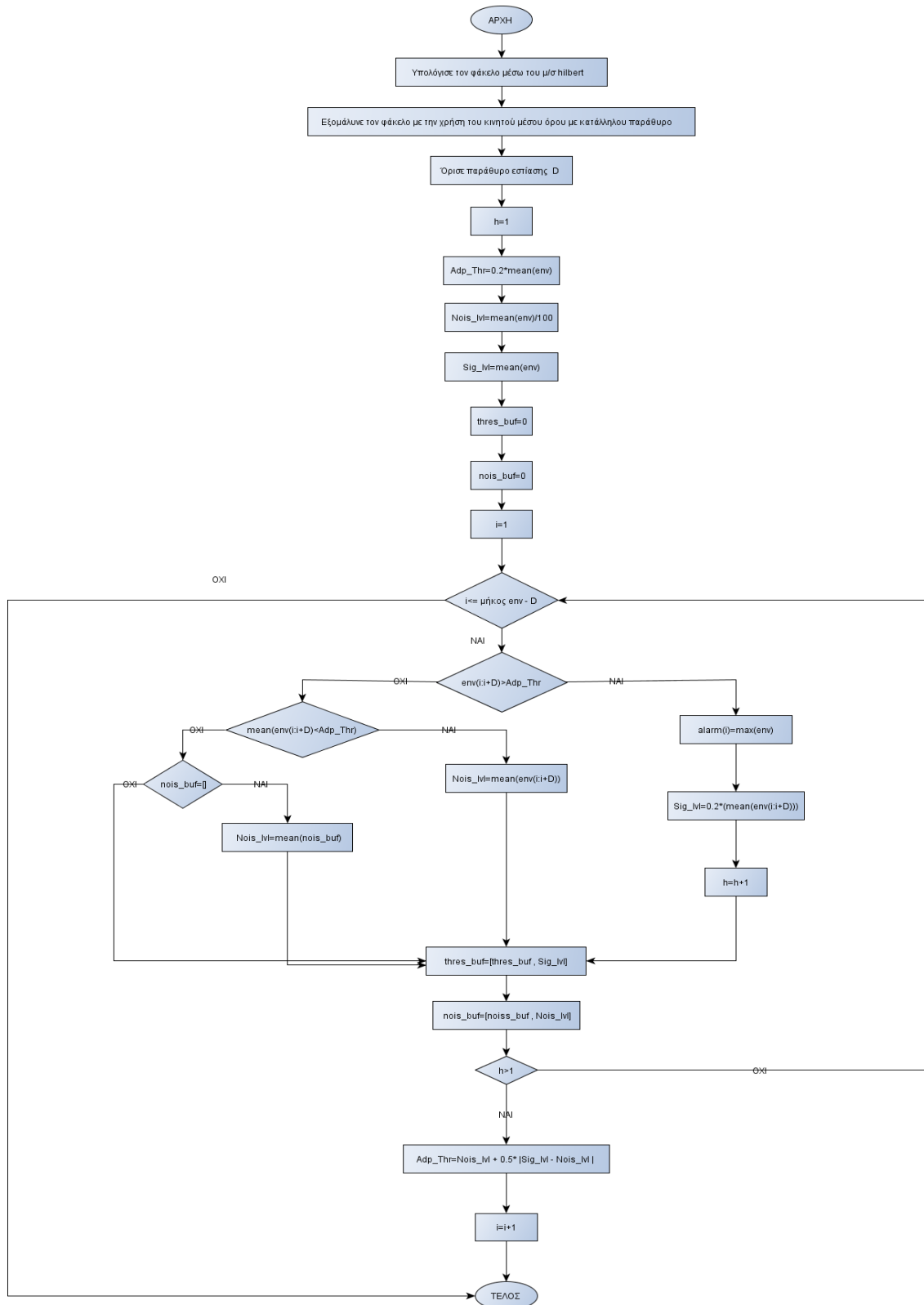
$$w[n] = \begin{cases} \frac{I_0\left(\pi\alpha\sqrt{1-\left(\frac{2n}{N-1}-1\right)^2}\right)}{I_0(\pi\alpha)}, & 0 \leq n \leq N-1 \\ 0, & \text{αλλού} \end{cases} \quad (\text{A.7})$$

όπου I_0 είναι η συνάρτηση Bessel μηδενικού βαθμού και α είναι η πραγματική μη μηδενική παράμετρος. Στο σχήμα 3.16 παρουσιάζονται τα παράθυρα Kaiser για διάφορες τιμές της παραμέτρου α .



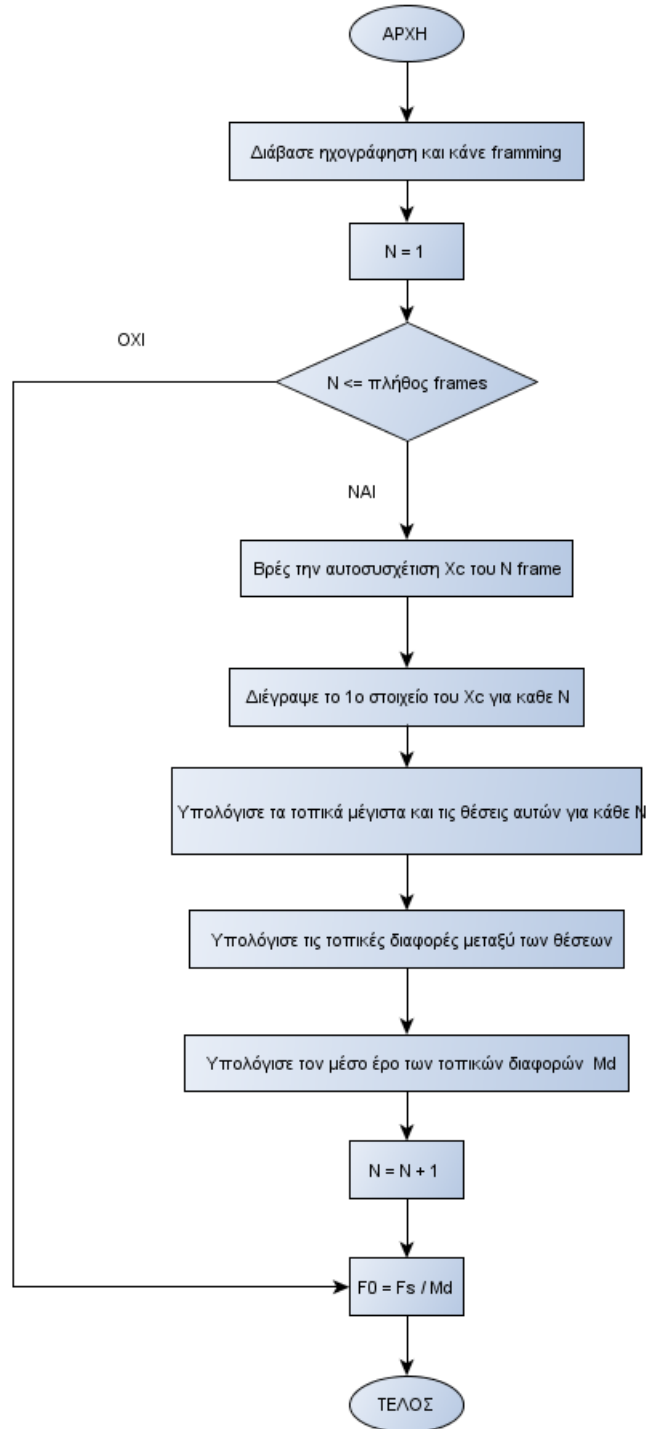
Σχήμα Α.7 Οικογένεια παραμετρικών παραθύρων Kaiser για διάφορες τιμές της παραμέτρου α .

ΠΑΡΑΡΤΗΜΑ Β'



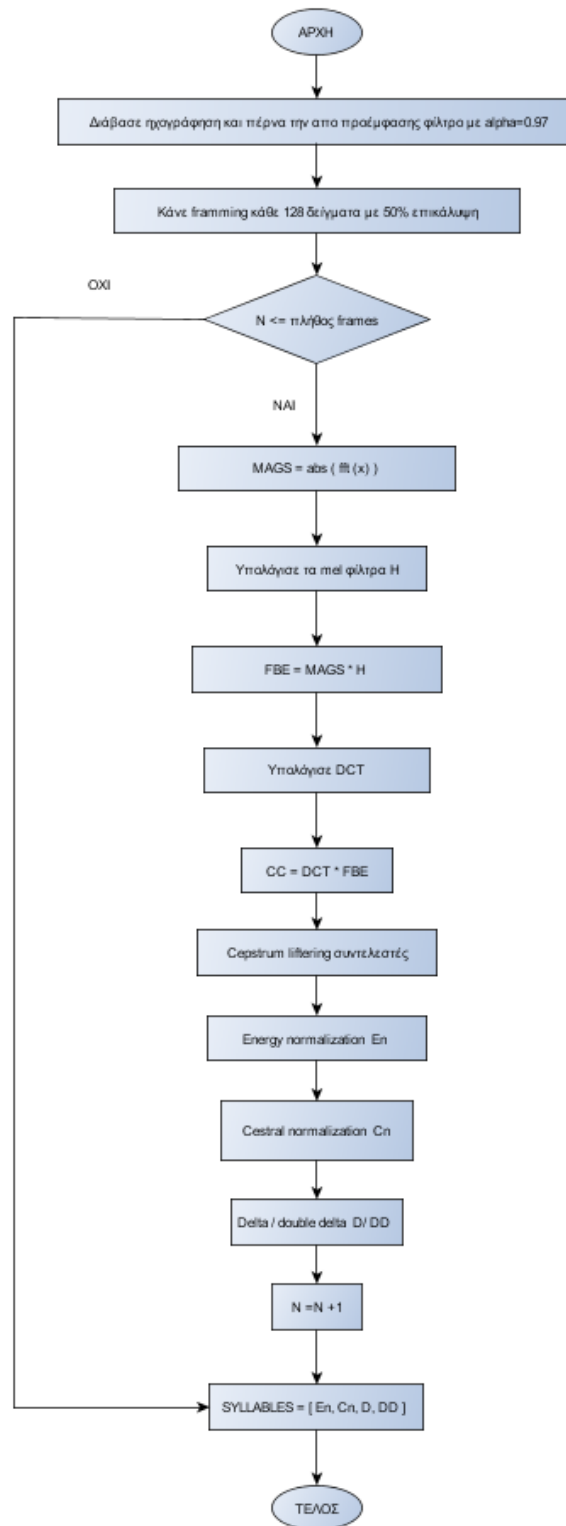
Σχήμα Β.1 Διάγραμμα ροής αναλυτικού σήματος [14].

ΠΑΡΑΡΤΗΜΑ Β΄



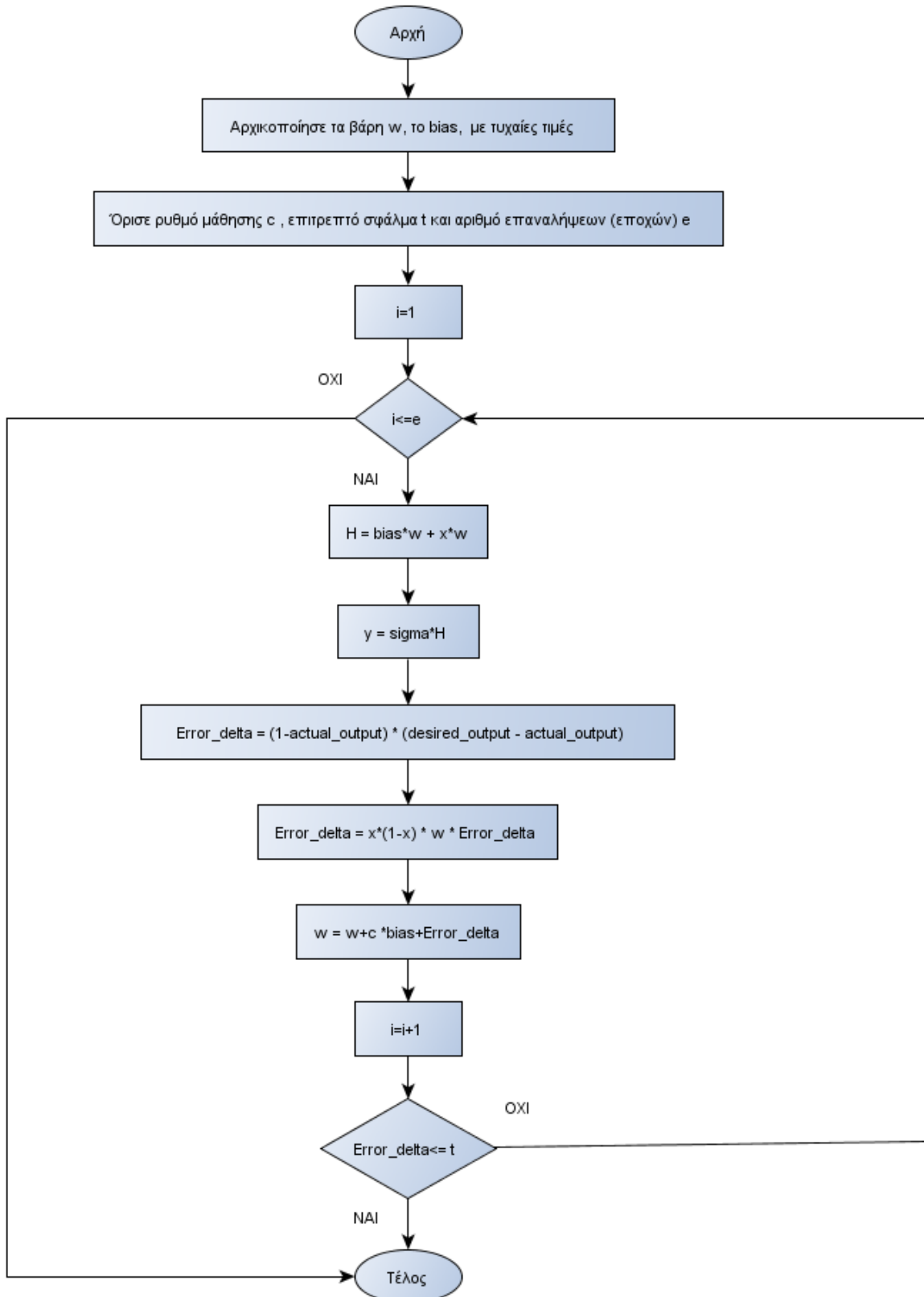
Σχήμα Β.2 Διάγραμμα ροής υπολογισμού θεμελιώδους συχνότητας.

ΠΑΡΑΡΤΗΜΑ Β'



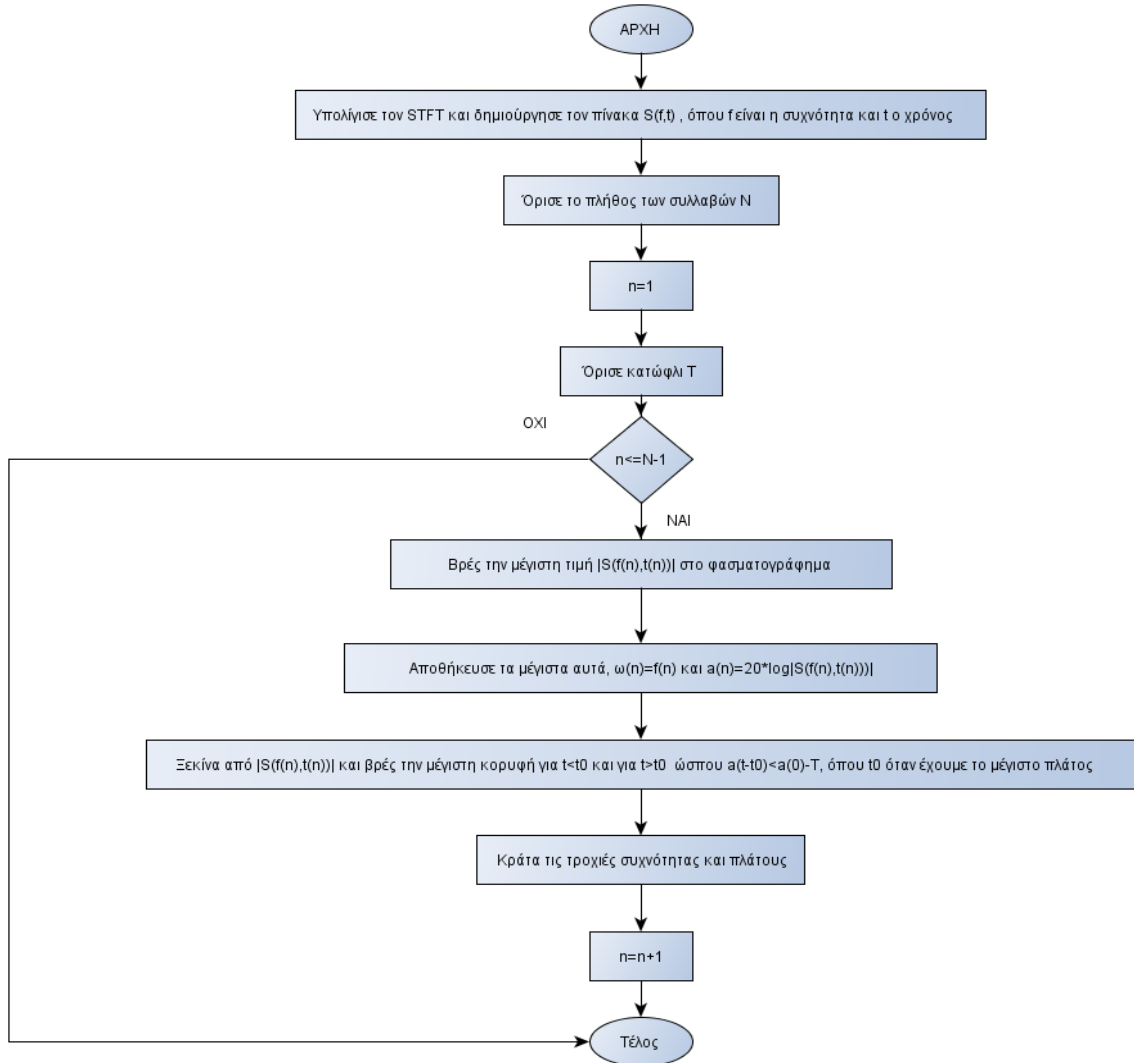
Σχήμα Β.3 Διάγραμμα ροής υπολογισμού συλλαβών με βάση τους συντελεστές MFCCs.

ΠΑΡΑΡΤΗΜΑ Β'



Σχήμα Β.4 Διάγραμμα ροής back-propagation νευρωνικού δικτύου [34].

ΠΑΡΑΡΤΗΜΑ Β΄



Σχήμα Β.5 Διάγραμμα ροής εξαγωγής συλλαβών με κατώφλι T [9].

Βιβλιογραφία

- [1] J. Vielliard. (2003). [Online]. Available: <http://www2.ib.unicamp.br/profs/jacques/ibac2003/history.html>.
- [2] [Online]. Available: <https://en.wikipedia.org/wiki/Bioacoustics>.
- [3] T. M. Aide, C. C. Bravo, M. C. Cerqueira, C. Milan, G. Vega and R. Alvarez, "Real-time bioacoustics monitoring and automated species identification," *PeerJ*, July 13, 2013.
- [4] E. Nemeth, N. Pieretti, S. A. Zollinger, N. Geberzahn, J. Partecke, A. C. Miranda and H. Brumm, "Bird song and anthropogenic noise: vocal constraints may explain why birds sing higher-frequency songs in cities," *Proceeding of the Royal Society*, December 13, 2012.
- [5] D. Wolff, "Detecting Bird Sounds via Periodic Structures: A Robust Pattern Recognition Approach to Unsupervised Animal Monitoring," M.S. thesis, Institut fur Informatik, Rheinische Friedrich-Wilhelms-Universität Bonn, May 28, 2008.
- [6] S. Fagerlund, "Automatic Recognition of Bird Species by Their Sounds," M.S. thesis, Department of Electrical and Communications Engineering Laboratory of Acoustics and Audio Signal Processing, Helsinki University Of Technology, Espoo, Nov.8, 2004.
- [7] X. Huang, A. Acero and H. W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. New Jersey : Prentice Hall, 2001, ch.2.
- [8] I. Potamitis, S. Ntalampiras, O. Jahn and K. Riede, "Automatic bird sound detection in long real-field recordings: Applications and tools," *Elsevier*, Feb. 2014.
- [9] A. Härmä, "Automatic identification of bird species based on sinusoidal modeling of syllables," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 545-548, 2003.
- [10] P. Laiolo, "The emerging significance of bioacoustics in animal species conservation" *Elsevier*, pp. 1636-1642, March 19, 2010.
- [11] D. Chesmore, "Automated bioacoustic identification of species," *Annals of the Brazilian Academy of Science*, pp. 435-440, February 5, 2004.
- [12] P. Somervuo and A. Härmä, "Bird song recognition based on syllable pair histograms," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.5, pp 825–828, 2004.
- [13] F. Briggs, X. Fern and R. Raich, "Technical Report (Not Peer Reviewed): Acoustic Classification of Bird Species from Syllables: an Empirical Study".
- [14] H. Sedghamiz. (2013 April). [Online]. Available: <http://www.mathworks.com/matlabcentral/fileexchange/46139-automatic-enveloping--variance-change-and-activity-detection-with-hilbert-transform>.
- [15] C. H. Lee, Y. K. Lee and R. Z. Huang, "Automatic Recognition of Bird Songs Using Cepstral Coefficients," *Journal of Information Technology and Applications*, vol.1, No.1, pp. 17-23, May 2006.

- [16] F. Briggs, R. Raich, and X. Z. Fern, "Audio classification of bird species: a Statistical Manifold Approach," *IEEE*.
- [17] W. Chu and D. T. Blumstein, "Noise robust bird song detection using syllable pattern-based Hidden Markov Models".
- [18] L. Neal, F. Briggs, R. Raich and X. Z. Fern, "Time-frequency segmentation of bird song in noisy acoustic environments," *IEEE ICASSP*, pp. 2012-2015, 2011.
- [19] L. Ranjard and H. A. Ross, "Unsupervised bird song syllable classification using evolving neural networks," *J. Acoust. Soc. Am.* vol.6, pp. 4358-4368, March 6, 2008.
- [20] W. Han, C. F. Chan, C. S. Choy and K. P. Pun, "An efficient MFCC extraction method in speech recognition," *IEEE ISCAS*, pp. 145-148, 2006.
- [21] [Online]. Available: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>.
- [22] L. M. Narayana and S. K. Kopparapu. "Choice of Mel Filter Bank in Computing MFCC," October 28, 2014.
- [23] [Online]. Available: <http://azimadli.com/vibman/cepstrumterminology.htm>.
- [24] [Online]. Available: <http://iitg.vlab.co.in/?sub=59&brch=164&sim=615&cnt=1>.
- [25] J. W. Leis, *DSP Using Matlab For Students and Researchers*. New Jersey : John Wiley & Sons, 2011, ch.5, pp. 156-159.
- [26] [Online]. Available: https://en.wikipedia.org/wiki/Short-time_Fourier_transform.
- [27] F. R. Kschischang. (2006, October 22). The Hilbert Transform [Online]. Available: <http://www.comm.utoronto.ca/frank/papers/hilbert.pdf>.
- [28] [Online]. Available: http://www.analog.com/media/en/technical-documentation/dsp-book/dsp_book_Ch6.pdf.
- [29] [Online]. Available: <http://matlabtricks.com/post-11/moving-average-by-convolution>.
- [30] Y. W. Liu. (2012, April 25). [Online]. Available: <http://cdn.intechopen.com/pdfs-wm/36434.pdf>.
- [31] J.T. Ulrich. (2006, March). [Online]. Available: https://www.researchgate.net/publication/257547765_Envelope_Calculation_from_the_Hilbert_Transform.
- [32] Γ. Πεφάνιδης. (2011, October). [Online]. Available: <http://ai.uom.gr/Courses/AdvancedNeuralNetworks/Material/NeuralNetworksAll.pdf>.
- [33] [Online]. Available: <http://technobium.com/stock-market-prediction-using-neuroph-neural-networks/>.
- [34] M. Mazur. (2015, March 17). [Online]. Available: <https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/>.

- [35] [Online]. Available: <http://matlabgeeks.com/tips-tutorials/neural-networks-a-perceptron-in-matlab/>.
- [36] (2016, March 28). [Online]. Available: <https://www.youtube.com/watch?v=StKImfDrLf4>.
- [37] (2015, May 30). [Online]. Available: <https://www.youtube.com/watch?v=oDrLeoUWRxY>.
- [38] (2014, October 19). [Online]. Available: <https://www.youtube.com/watch?v=iGyLd5sQcde>.
- [39] (2012, April 23). [Online]. Available: <https://www.youtube.com/watch?v=yntsA5apjHY>.
- [40] (2012, November 25). [Online]. Available: <https://www.youtube.com/watch?v=rCG79aOH-Ag>.
- [41] (2011, March 29). [Online]. Available: <https://www.youtube.com/watch?v=sPLVE6DVXh4>.
- [42] (2014, January 2). [Online]. Available: <https://www.youtube.com/watch?v=NW9m0RxxgLY>.
- [43] (2013, June 30). [Online]. Available: <https://www.youtube.com/watch?v=UKjztDZRsw8>.
- [44] (2013, March 8). [Online]. Available: <https://www.youtube.com/watch?v=7qF0wP5wV64>.
- [45] (2015, March 1). [Online]. Available: https://www.youtube.com/watch?v=WfD_YxMxm54.
- [46] (2013, September 30). [Online]. Available: <https://www.youtube.com/watch?v=h5UG9e84v7Y>.
- [47] (2014, March 8). [Online]. Available: https://www.youtube.com/watch?v=XiLeq_wCQVY.
- [48] (2015, February 1). [Online]. Available: https://www.youtube.com/watch?v=p5wmKZzjC_M.
- [49] (2014, April 23). [Online]. Available: <https://www.youtube.com/watch?v=czqHTfvjNxM>.
- [50] (2014, May 17). [Online]. Available: <https://www.youtube.com/watch?v=QroHX3xHqVE>.
- [51] (2015, January 27). [Online]. Available: <https://www.youtube.com/watch?v=C3NZD2kUEv0>.
- [52] (2015, August 20). [Online]. Available: https://www.youtube.com/watch?v=X8FOWd_gS1A.
- [53] (2010, April 15). [Online]. Available: <https://www.youtube.com/watch?v=Ga864L7p9CQ>.
- [54] (2015 October 12). [Online]. Available: <https://www.youtube.com/watch?v=gLSMAtjK5T8>.
- [55] (2013, June 15). [Online]. Available: <https://www.youtube.com/watch?v=wHbQGeUH0b8>.
- [56] (2015, April 14). [Online]. Available: <http://www.xeno-canto.org/235818>.
- [57] (2014, May 14). [Online]. Available: <http://www.xeno-canto.org/189562>.
- [58] (2014, May 8). [Online]. Available: <http://www.xeno-canto.org/178467>.
- [59] (2015, April 15). [Online]. Available: <http://www.xeno-canto.org/283317>.
- [60] (2014, May 30). [Online]. Available: <http://www.xeno-canto.org/180178>.
- [61] (2013, May 1). [Online]. Available: <http://www.xeno-canto.org/132028>.

- [62] (2014, April 14). [Online]. Available: <http://www.xeno-canto.org/130585>.
- [63] (2013, April 14). [Online]. Available: <http://www.xeno-canto.org/130584>.
- [64] (2012, June 27). [Online]. Available: <http://www.xeno-canto.org/105120>.
- [65] (2012, June 6). [Online]. Available: <http://www.xeno-canto.org/103144>.
- [66] (2014, April 5). [Online]. Available: <http://www.xeno-canto.org/173934>.
- [67] (2015, May 25). [Online]. Available: <http://www.xeno-canto.org/304600>.
- [68] (2015, May 25). [Online]. Available: <http://www.xeno-canto.org/304599>.
- [69] (2015, June 1). [Online]. Available: <http://www.xeno-canto.org/304589>.
- [70] (2015, June 2). [Online]. Available: <http://www.xeno-canto.org/301941>.
- [71] (2015, June 7). [Online]. Available: <http://www.xeno-canto.org/301931>.
- [72] (2015, May 23). [Online]. Available: <http://www.xeno-canto.org/246773>.
- [73] (2015, May 26). [Online]. Available: <http://www.xeno-canto.org/304597>.
- [74] (2015, June 16). [Online]. Available: <http://www.xeno-canto.org/299908>.
- [75] (2015, May 14). [Online]. Available: <http://www.xeno-canto.org/317748>.
- [76] (2013, March 4). [Online]. Available: <http://www.xeno-canto.org/144943>.
- [77] (2011, February 12). [Online]. Available: <http://www.xeno-canto.org/132597>.
- [78] (2010, July 20). [Online]. Available: <http://www.xeno-canto.org/118344>.
- [79] (2015, April 11). [Online]. Available: <http://www.xeno-canto.org/244636>.
- [80] (2015, May 8). [Online]. Available: <http://www.xeno-canto.org/244629>.
- [81] (2014, May 13). [Online]. Available: <http://www.xeno-canto.org/178522>.
- [82] (2015, May 29). [Online]. Available: <http://www.xeno-canto.org/254526>.
- [83] (2013, June 14). [Online]. Available: <http://www.xeno-canto.org/144976>.
- [84] (2012, May 6). [Online]. Available: <http://www.xeno-canto.org/130698>.
- [85] (2013, April 14). [Online]. Available: <http://www.xeno-canto.org/129614>.
- [86] (2016, February 2). [Online]. Available: <http://www.xeno-canto.org/304434>.
- [87] (2016, February 22). [Online]. Available: <http://www.xeno-canto.org/304432>.
- [88] (2015, September 18). [Online]. Available: <http://www.xeno-canto.org/282086>.
- [89] (2015, September 19). [Online]. Available: <http://www.xeno-canto.org/282085>.

Βιβλιογραφία

- [90] [Online]. Available: [https://en.wikipedia.org/wiki/Trill_\(music\)](https://en.wikipedia.org/wiki/Trill_(music)).
- [91] [Online]. Available:
<http://www.musicheaven.gr/html/modules.php?name=Encyclopedia&op=content&tid=58>.
- [92] [Online]. Available: <https://ccrma.stanford.edu/~pdelac/154/m154paper.htm>.
- [93] [Online]. Available:
<http://www.mathworks.com/help/signal/ref/findpeaks.html?requestedDomain=www.mathworks.com>.
- [94] [Online]. Available: <http://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw>.
- [95] M. T. Hagan, H. B. Demuth, M. H. Beale and O. De Jesús. Neural Network Design (2nd ed.) [Online]. Available: <http://hagan.okstate.edu/NNDesign.pdf#page=469>.
- [96] [Online]. Available: <http://www.mathworks.com/help/nnet/ref/trainlm.html>.
- [97] [Online]. Available: http://en.wikipedia.org/wiki/Window_function.
- [98] [Online]. Available: http://en.wikipedia.org/wiki/Kaiser_window.