Εθνικο Μετσοβιο Πολυτεχνειο

Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων

Τομεας Σηματων, Ελεγχου και Ρομποτικης

Αναγνώριση βιοσήματος οδοντοκητών με συνελικτικά αναδρομικά νευρωνικά δίκτυα

Διπλωματικη Εργασια

του

ΔΗΜΗΤΡΙΟΥ Ν. ΜΑΚΡΟΠΟΥΛΟΥ

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Συνεπιβλέποντες:

Δρ. Αντιγόνη Τσιάμη
Μεταδιδακτορική Ερευνήτρια Ε.Μ.Π.

Δρ. Αριστείδης Προσπαθόπουλος
Ειδικός Λειτουργικός Επιστήμονας Α΄
(Ι.Ω. ΕΛ.ΚΕ.Θ.Ε)

ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ

Αθήνα, Οκτώβριος 2022

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής
ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ
ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ

Αναγνώριση βιοσήματος οδοντοκητών με συνελικτικά αναδρομικά νευρωνικά δίκτυα

Διπλωματικη Εργασια

του

ΔΗΜΗΤΡΙΟΥ Ν. ΜΑΚΡΟΠΟΥΛΟΥ

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 31η Οκτωβρίου 2022.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Πέτρος Μαραγκός	Αθανάσιος Ροντογιάννης	Γεράσιμος Ποταμιάνος
Καθηγητής Ε.Μ.Π.	Αν. Καθηγητής ΕΜΠ	Αν. Καθ. Τμ.ΗΜΜΥ
		Παν/μιο Θεσσαλιας

Αθήνα, Οκτώβριος 2022

:.....

Δημητριος Ν. Μακροπουλος
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Στην εργασία που ακολουθεί εστιάζουμε στην ανάπτυξη υβριδικών συνελικτικών αναδρομικών νευρωνικών δικτύων (Convolutional Recurrent Neural Networks) για την κατηγοριοποίηση βιοσημάτων που έχουν συλλεχθεί στην Ελληνική Τάφρο και προέρχονται από δύο είδη κητωδών, φυσητήρες (*Physeter macrocephalus*) και ζωνοδέλφινα (*Stenella coeruleoalba*). Μετατρέπουμε τα ηχητικά σήματα σε φασματογραφήματα κλίμακας mel (mel-spectrograms) πριν τα εισάγουμε ως εισόδους σε βαθύ συνελικτικό δίκτυο ResNet που έχει σχεδιαστεί για να εξάγει χρονοσυχνοτικά πρότυπα. Στην κορυφή του δικτύου τοποθετείται ένα στρώμα κατανεμημένο στο χρόνο (time-distributed layer) που ανασχηματίζει τη διάσταση του διανύσματος χαρακτηριστικών για να το εισάγει σε κάποια εκδοχή αναδρομικού νευρωνικού δίκτυου, Long Short-Term Memory (LSTMs) ή Gated Recurrent Units (GRUs) και να αναγνωρίσει σε αυτό μακροπρόθεσμες χρονικές εξαρτήσεις. Αποδεικνύεται ότι το υβριδικό δίκτυο κατηγοριοποιεί με ακρίβεια ηχητικά σήματα σε ένα πρόβλημα ταξινόμησης βιοσημάτων σε κητώδη έναντι περιβάλλοντος θορύβου ενώ επιδεικνύει ισχυρή ικανότητα μάθησης σε πρόβλημα αναγνώρισης που περιλαμβάνει αλληλεπικαλυπτόμενες ηχητικές αναπαραστάσεις. Η προτεινόμενη αρχιτεκτονική επιτυγχάνει να διαχωρίσει το χώρο εισόδου αποδοτικότερα τόσο σε σχέση με παραδοσιακές μεθόδους μηχανικής μάθησης όσο και ως προς αρχιτεκτονικές βάσης που περιέχουν είτε μόνο συνελικτικά δίκτυα είτε μόνο αναδρομικά ή δομές που συνδυάζουν παράλληλα συνελικτικά και αναδρομικά δίκτυα.

Λέξεις Κλειδιά

Μηχανική μάθηση, αναγνώριση προτύπων, βιοακουστική, αναδρομικά δίκτυα, υδρόφωνα

Abstract

In this paper we focus on the development of a convolutional recurrent neural network (CRNN) to categorize biosignals collected in the Hellenic Trench, generated by two cetacean species, sperm whales (*Physeter macrocephalus*) and striped dolphins (*Stenella coeruleoalba*).We convert audio signals into mel-spectrograms applying dynamic compression techniques based on automatic gain control and forward the input into a deep residual network, designed to capture spectral patterns. Next, ResNet's output is reshaped into a time-distributed layer and fed into recurrent network variants, Long Short-Term Memory (LSTMs) or Gated Recurrent Units (GRUs), able to recognize long-term time dependencies on extracted features. The hybrid network is able to perfectly classify audio signals into three categories (dolphins, sperm whales, ambient noise) while it also exhibits high learning ability on recognising intraclass representations of overlapping acoustic patterns (clicks vs whistles and clicks, both emitted by dolphins). The proposed scheme outperforms traditional ML techniques and baseline architectures comprising either ResNet or LSTM structures or their deep parallel combinations.

Keywords

Machine learning, pattern recognition, bioacoustic patterns, recurrent networks, passive acoustic listeners

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή Πέτρο Μαραγκό που δέχτηκε την επίβλεψη της διπλωματικής μου εργασίας. Οι εβδομαδιαίες διαλέξεις του στην Ψηφιακή Επεξεργασία Σήματος, στην Όραση Υπολογιστών και Αναγνώριση Προτύπων, αποτέλεσαν μαζί πηγή έμπνευσης και τον χάρτη πάνω στον οποίο σχεδιάστηκε η εργασία. Κάθε γραμμή από την εργασία μου φέρει επίσης τη συμβολή και την παρουσία της συνεπιβλέπουσάς μου στη διπλωματική, Δρ. Αντιγόνης Τσιάμη. Την ευχαριστώ απεριόριστα για το χρόνο που διέθεσε μαζί με τις συμβουλές, τις επισημάνσεις, την προοπτική που έδωσε στην εργασία και το αίσθημα ασφάλειας που μου μετέφερε σε κάθε γραμμή έρευνας που επιχειρήσαμε. Η ιδέα για μια ερευνητική εργασία που να συνδυάζει μηχανική μάθηση και υποθαλάσσιους ήχους προήλθε από τον Δρ. Αριστείδη Προσπαθόπουλο υπεύθυνο της ερευνητικής ομάδας υποθαλάσσιου ήχου στο Ινστιτούτο Ωκεανογραφίας του ΕΛ.ΚΕ.Θ.Ε. Τον ευχαριστώ πολύ που ανακάλυψε το πρόβλημα για εμένα και μου το εμπιστεύτηκε, για την εμπειρία και τη γνώση που μου μετέφερε ειδικά και γενικά στα προβλήματα βιοακουστικής και διάδοσης του υποθαλάσσιου ήχου. Τέλος, θέλω να ευχαριστήσω την φοιτήτρια Αγγελική-Ελένη Δουκίδη, για την ευφυΐα και καλοσύνη που έφερε στη συνεργασία μας τα τελευταία χρόνια στη σχολή με αφορμή την ανάπτυξη σε μια πολυπληθή σειρά από εργασίες του κοινού προγραμματιστικού μας κώδικα.

Περιεχόμενα

Κατάλογος Σχημάτων

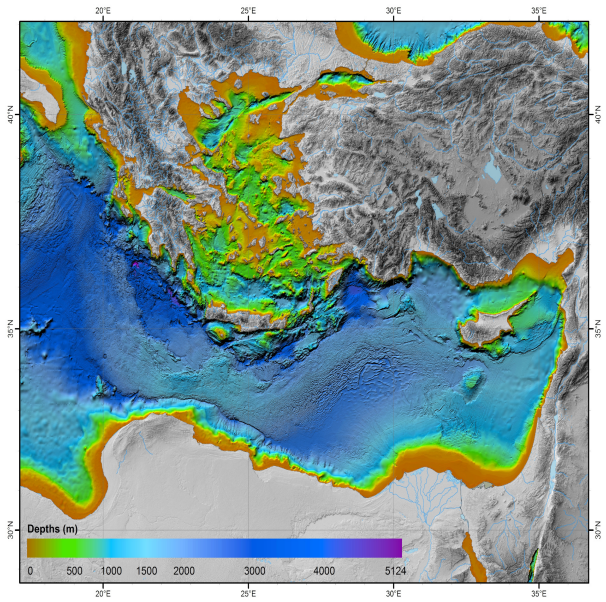
Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Η μελέτη που ακολουθεί εστιάζει σε οδοντοκήτη της Ελληνικής Τάφρου (Hellenic Trench), μιας θαλάσσιας γεωγραφικής περιοχής που εκτείνεται από τα δυτικά των Ιονίων Νήσων και της Πελοποννήσου έως τα νότια της Κρήτης και τη νότια Ρόδο εμβαδού 57.000 περίπου τετραγωνικών χιλιομέτρων. Η γειτονιά της Τάφρου επικαθορίζεται από υποθαλάσσιους γκρεμούς και χαράδρες σε κοντινές αποστάσεις από την ακτή (3-10 χλμ) που οδηγούν σε ραγδαία αύξηση του βάθους του θαλάσσιου πυθμένα. Στον χάρτη της Ανατολικής Μεσογείου (εικόνα 1.1) αποτυπώνονται οι απότομες κλίσεις στο ρήγμα της Τάφρου που κάνουν τον πυθμένα να υπερβαίνει συστηματικά τα 1000 μέτρα και να σχηματίζει 45 μίλια Ν.Δ της Πύλου, στο Φρέαρ των Οινουσσών, το βαθύτερο σημείο της Μεσογείου στα 5,267 χλμ από την επιφάνεια της θάλασσας.

Σχήμα 1.1: Βαθυμετρία Ανατολικής Μεσογείου



Πηγή: Ινστιτούτο Ωκεανογραφίας ΕΛ.ΚΕ.Θ.Ε

Στα βαθιά νερά της Ελληνικής Τάφρου έχει καταγραφεί ίσως ο σημαντικότερος πληθυσμός

φυσητήρων στην Ανατολική Μεσόγειο που αριθμεί περίπου 200-250 άτομα [3]. Η εμφάνιση των φυσητήρων στην επιφάνεια της θάλασσας είναι σπάνια καθώς καταδύονται για χρονικά διαστήματα που υπερβαίνουν συχνά τα 30-40 λεπτά σε βάθη που μπορεί να ξεπερνούν τα 2000 μέτρα [4]. Σε τέτοια βάθη, απουσία φωτός, η αλληλεπίδρασή τους με το περιβάλλον γίνεται μέσω παραγωγής και εκπομπής ηχητικών και υπερηχητικών σημάτων. Τα θηλαστικά αυτά είναι προικισμένα -μέσα από την ιδιαίτερη ανατομία τους- με το ισχυρότερο βιολογικό sonar που μέχρι τώρα έχει καταγραφεί στο ζωικό βασίλειο [5] και εκπέμπουν χαρακτηριστικούς παλμούς ηχοεντοπισμού για να προσανατολιστούν και να εντοπίσουν τη λεία τους ενώ με διακριτό τρόπο παράγουν ακολουθίες παλμών (codas) για σκοπούς επικοινωνίας [6]. Η θαλάσσια περιοχή της Τάφρου αποτελεί ενδιαίτημα και για άλλα οδοντοκήτη: Ζιφιοί *(Ziphious cavuristrus)*, ζωνοδέλφινα *(Stenella coeruleoalba)*, ρινοδέλφινα *(Tursiops truncatus)*, σταχτοδέλφινα *(Grampus griseus)*, κοινά δελφίνια *(Delphinus delphis)*. Έχουν καταγραφεί καταδύσεις ζιφιών σε βάθη μεγαλύτερα από εκείνα στα οποία φτάνουν οι φυσητήρες ενώ τα ζωνοδέλφινα που είναι το είδος δελφινιού που απαντιέται περισσότερο στη Μεσόγειο μπορούν με τη σειρά τους να καταδυθούν σε βάθη που φτάνουν τα 700 μέτρα εκπέμποντας ηχητικούς παλμούς (clicks) ή συνδυασμό παλμών και σφυριγμάτων (whistles).

Σε ένα τέτοιο απαγορευτικό για το φως περιβάλλον η μελέτη της συμπεριφοράς των θαλάσσιων θηλαστικών γίνεται μέσω υδρόφωνων που αποτυπώνουν σε ένα συνεχές από θορύβους περιβάλλοντος περιστασιακές εκφωνήσεις κητωδών στο ηχητικό και υπερηχητικό φάσμα. Το τελικό αποτέλεσμα της καταγραφής αποτυπώνει μια υπέρθεση ήχων που μαζί με τη βιοακουστική υπογραφή των θηλαστικών ενσωματώνει χαρακτηριστικά της σχετικής και απόλυτης θέσης τους ως προς τους αισθητήρες ή αποτυπώματα από τα χαρακτηριστικά της αλληλεπίδρασης του βιοσήματος με το μέσο διάδοσης όπως ανακλάσεις από αντικείμενα στην ακτίνα καταγραφής ή αποσβέσεις. Η μελέτη των ηχητικών σημάτων από θαλάσσια θηλαστικά είναι ένα αντικείμενο στο οποίο συγκλίνουν διαφορετικά επιστημονικά πεδία: Η βιοακουστική μελετά τα ιδιαίτερα χαρακτηριστικά του παραγόμενου από τα κητώδη ήχου, η φυσική αναζητά λύσεις της κυματικής εξίσωσης για τη διάδοση του υποθαλάσσιου ήχου στον ωκεανό, η ψηφιακή επεξεργασία σήματος στοχεύει στην αποθορυβοποίηση και βέλτιστη διαχείριση του προβλήματος διάστασης-συμπίεσης δεδομένων ενώ πρόσφατα η μηχανική μάθηση συνεισφέρει στην αναγνώριση των ήχων και την ταξινόμησή τους. Σκοπός της εργασίας είναι να αναδείξει πτυχές της εικόνας του προβλήματος σε μικρές ενότητες εστιάζοντας σε επιμέρους οπτικές του πριν αναπτυχθεί πιο αναλυτικά ένα βαθύ δίκτυο που να μαθαίνει βιοσήματα οδοντοκητών και να τα κατηγοριοποιεί.

Κεφάλαιο 2

Φυσική του Ήχου

'If I can get a mechanism which will make a current of electricity vary in its intensity, as the air varies in density when a sound is passing through it, I can telegraph any sound, even the sound of speech.'
Alexander Graham Bell

2.1 Εισαγωγή

Το ηχητικό κύμα χαρακτηρίζεται ως διαταραχή του μηχανικού μέσου στο οποίο αυτό διαδίδεται και εμφανίζεται με τη μορφή μιας ακολουθίας διαδοχικά εναλλασσόμενων περιοχών υψηλής και χαμηλής πίεσης (L.D.Landau, E.M.Lifschitz) [7]. Είναι διαμήκες μηχανικό κύμα και μπορεί να ταξιδέψει σε οποιοδήποτε μέσο, στερεό, υγρό ή αέριο όχι όμως στο κενό που εξ ορισμού χαρακτηρίζεται από απουσία μέσου διάδοσης. Καθώς το κύμα διαδίδεται, τα σωματίδια του μέσου (των μορίων του αέρα αν ο ήχος διαδίδεται στην ατμόσφαιρα, των μορίων νερού αν αυτός διαδίδεται στο νερό, των μορίων του υλικού ενός στερεού αν διαδίδεται σε τέτοιο μέσο) πάλλονται παράγοντας μορφές πολύπλοκων μοτίβων μεταβολών πίεσης και πυκνότητας (L.R.Rabiner, R.W.Schafer) [8] κατά μήκος της διεύθυνσης διάδοσης του κύματος .

Η ακολουθία γένεσης διαδοχικών πυκνωμάτων και αραιωμάτων στο μέσο γίνεται με μεγάλη ταχύτητα χωρίς να συνοδεύεται από μεταφορά ενέργειας από το πυκνό μέρος του μέσου στο αραιό με τη μορφή θερμότητας, είναι δηλαδή μια αδιαβατική διαδικασία. Η ιδέα αυτή, ότι η πίεση και η θερμοκρασία μεταβάλλονται αδιαβατικά σε ένα ηχητικό κύμα πιστώνεται στον μαθηματικό P.S Laplace που διόρθωσε την επικρατούσα Νευτώνεια εκτίμηση ότι η μεταβολή της πίεσης με την πυκνότητα στην ηχητική διάδοση είναι ισόθερμη διαδικασία (R.Feynman) [9]. Αν συνέβαινε η θερμοκρασία να διατηρείται παντού σταθερή στο μέσο, θα έπρεπε η ταχύτητα ροής της θερμότητας (που οφείλεται στη μεταφορά μεταφορικής κινητικής ενέργειας από το ένα μόριο στο άλλο κατά τη κρούση τους) να είναι πολύ μεγαλύτερη από την ταχύτητα του ήχου που εκφράζεται από το λόγο της απόστασης μισού μήκους κύματος (από πύκνωμα σε αραίωμα) ανά μισή περίοδο ταλάντωσης (οπότε πυκνώματα και αραιώματα εναλλάσσονται) (F.S Crawford) [10]. Έτσι θα έπρεπε : $v(heatflow) >> \frac{\lambda}{T}$. Αποδεικνύεται ωστόσο ότι κάτι

τέτοιο δεν συμβαίνει.

Η αγωγή της θερμότητας μεταξύ διαφορετικών τμημάτων του ρευστού και συγκεκριμένα από την υψηλής θερμοκρασίας πυκνή περιοχή στη χαμηλής θερμοκρασίας αραιή περιοχή είναι αμελητέα γεγονός που συνεπάγεται αμεταβλητότητα στην εντροπία των σωματιδίων που κινούνται στο χώρο (L.D.Landau, E.M.Lifschitz)[7]. Αυτό οφείλεται στο γεγονός ότι το μήκος κύματος του ήχου είναι μεγάλο σε σχέση με τη μέση ελεύθερη διαδρομή των σωματιδίων του μέσου, αν κάτι τέτοιο δεν συνέβαινε τα σωματίδια του μέσου θα κινούνταν ελεύθερα στις περιοχές υψηλής και χαμηλής πίεσης αντισταθμίζοντας τη διαφορά πίεσης που προϋποθέτει τον ήχο (R.Feynman) [9]. Στη σύγχρονη φυσική η παραπάνω παραδοχή παραμένει θεμελιώδης στην εξήγηση των ηχητικών φαινομένων ενώ κύματα στα οποία παρατηρείται μεταβλητότητα στην εντροπία και είναι γνωστά με την ονομασία 'δεύτερος ήχος' (second sound) συνδέονται με φαινόμενα υπερρευστότητας υλικών όπως το υγρό ήλιο-II (Helium II), είναι κύματα εντροπίας στα οποία μεταβάλλεται η θερμοκρασία και η εντροπία χωρίς το φαινόμενο να συνοδεύεται από αξιόλογες διακυμάνσεις της πίεσης και της πυκνότητας του μέσου. Η παρατήρηση της ασυνήθιστης θερμικής αγωγιμότητας του ήλιου-II είχε ήδη από το 1935 επισημανθεί από τον φυσικό W.Keesom ενώ η εξήγηση του φαινομένου διατυπώθηκε από τους φυσικούς P.L.Kapitza (Nobel prize 1978) το 1938 [11] και L.D.Landau (Nobel prize 1961) το 1941 [12].

2.2 Θεμελιώδεις εξισώσεις ρευστοδυναμικής και κυματική εξίσωση διάδοσης ηχου.

Η διάδοση του ήχου οφείλεται στην παρουσία ηχητικών πηγών στο χώρο και τα χαρακτηριστικά της διάδοσής του εξαρτώνται ισχυρά κατ' αρχάς από τα γεωμετρικά και φυσικά τους χαρακτηριστικά. Στη μαθηματική προτυποποίηση της ηχητικής διάδοσης, οι πηγές προσομοιώνονται συχνά ως σημειακές που αποδίδουν σφαιρική συμμετρία όταν η διάστασή τους είναι πολύ μικρή σε σχέση με το μήκος κύματος του ήχου ή γραμμικές με κυλινδρική συμμετρία οπότε και εκπέμπουν κυλινδρικά κύματα ενώ μεγάλες επίπεδες πηγές παράγουν αντίστοιχα επίπεδα κύματα. Μια ηχητική πηγή μπορεί να παράγει κύματα ασθενώς ή ισχυρά κατευθυντικά ή κύματα σφαιρικά που διαδίδονται προς όλες τις κατευθύνσεις ομοιότροπα. Παράλληλα με την ισχυρή συνεισφορά της πηγής στα χαρακτηριστικά του ήχου, το διαδιδόμενο ηχητικό κύμα εξαρτάται από παραμέτρους και συνοριακές συνθήκες χαρακτηριστικές του ωκεάνιου μέσου [13]. Αποδεικνύεται ότι η περιγραφή της ηχητικής διάδοσης μπορεί να αποτυπωθεί πλήρως σε μια κυματική εξίσωση που παράγεται από την αρχή διατήρησης της μάζας (εξίσωση συνέχειας) και το θεμελιώδη νόμο της μηχανικής του Νεύτωνα (εξίσωση Euler).

2.2.1 Εξίσωση διατήρησης της μάζας - εξίσωση συνέχειας

Έστω ρευστό όγκου dV στον τρισδιάστατο χώρο. Η εξίσωση διατήρησης της μάζας ενός ρευστού εξισώνει τη μάζα του ρευστού που εξέρχεται από μια επιφάνεια που περικλείει το ρευστό στη μονάδα του χρόνου, $\oint_V \rho\mathbf{v}\, d\mathbf{S}$, με το ρυθμό μεταβολής της μάζας του ρευστού μέσα στον στοιχειώδη όγκο dV:

$$\oint_V \rho\mathbf{v}\, d\mathbf{S} = -\frac{\partial}{\partial t}\int \rho\mathbf{v}\, dV$$

όπου το αρνητικό πρόσημο υποδηλώνει την εκροή ρευστού λόγω απώλειας μάζας και η κατεύθυνση της στοιχειώδους επιφάνειας $d\mathbf{S}$ είναι κατά μήκος της καθέτου.

Από το θεώρημα της απόκλισης (Green) το επιφανειακό ολοκλήρωμα που ορίζει την τιμή μιας συνάρτησης στη συνοριακή επιφάνεια που περιβάλλει τον όγκο ισούται με το ολοκλήρωμα της απόκλισης της συναρτησης πάνω στην περιοχή αυτή :

$$\int \nabla\cdot(\rho\mathbf{v})\, dV = \oint_V \rho\mathbf{v}\, d\mathbf{S}$$

Συνεπώς η εξίσωση διατήρησης της μάζας εκφράζεται στη διαφορική της μορφή ως:

$$\nabla\cdot(\rho\mathbf{v}) + \frac{\partial\rho}{\partial t} = 0 \tag{2.1}$$

Η παραπάνω εξίσωση διατυπώνει μαθηματικά τη φυσική συνθήκη ότι η εκροή μάζας ρευστού από μια επιφάνεια ανά μονάδα όγκου σε κάθε σημείο του χώρου στη μονάδα του χρόνου ισούται με τον αρνητικό ρυθμό μεταβολής της πυκνότητας του ρευστού στον όγκο.

2.2.2 Εξίσωση Euler

Έστω ρευστό όγκου dV στον τρισδιάστατο χώρο εντός ενός ρευστού. Η συνολική δύναμη που ασκείται στον όγκο αυτό που περιβάλλεται από ρευστό ισούται με το ολοκλήρωμα της πίεσης στην επιφάνεια που περιορίζει τον όγκο αυτό, $-\oint_V p\, dS$, ενώ μετασχηματίζεται με το θεώρημα της απόκλισης:

$$-\oint_V p\, dS = -\oint_S (\nabla p)\, dV$$

Συνεπώς η δύναμη που ασκείται στον μοναδιαίο όγκο του ρευστού ισούται με ∇p και μέσω αυτής προκύπτει η εξίσωση κίνησης από την Νευτώνεια μηχανική:

$$-\nabla p = \rho d\mathbf{v}/dt$$

που εξισώνει τη δύναμη με τη μάζα ανά μονάδα όγκου (ρ) και το ρυθμό μεταβολής της ταχύτητας δεδομένου σωματιδίου του ρευστού καθώς κινείται στο χώρο. Η στοιχειώδης μεταβολή στην ταχύτητα δεδομένου σωματιδίου ρευστού σε χρόνο dt συνίσταται αφενός μεν στη μεταβολή της ταχύτητας σε σταθερό σημείο του χώρου στο διάστημα αυτό $(\frac{\partial\mathbf{v}}{\partial t})dt$ αφετέρου δε στη διαφορά ταχυτήτων την ίδια χρονική στιγμή δύο σημείων που βρίσκονται σε απόσταση dr μεταξύ τους, όπου $d\mathbf{r}$ η απόσταση που μετακινήθηκε το υπο μελέτη σωματίδιο στο χρόνο dt: $(d\mathbf{r}\cdot\nabla)\mathbf{v}$ (L.D.Landau, E.M.Lifschitz) [7]. Παραγωγίζοντας το παραπάνω άθροισμα ως προς

το χρόνο προκύπτει:

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{\rho}\nabla p \tag{2.2}$$

Η παραπάνω εξίσωση του Euler (1755) είναι θεμελιώδης εξίσωση της ρευστοδυναμικής και συνιστά την εξίσωση κίνησης του ρευστού. Η εμβέλεια της εξίσωσης αυτής αφορά στα ιδανικά ρευστά για τα οποία η απώλεια ενέργειας λόγω τριβής και η θερμική αγωγιμότητα θεωρούνται αμελητέες[7].

2.2.3 Κυματική Εξίσωση στο πεδίο του χρόνου και της συχνότητας

Από τις εξισώσεις διατήρησης μάζας και την εξίσωση του Euler είναι εφικτή η διατύπωση της κυματικής εξίσωσης για ιδανικά ρευστά. Υποθέτοντας πως οι ταλαντώσεις που προκαλεί η ηχητική πηγή είναι μικρού πλάτους μπορούμε να θεωρήσουμε ότι η σωματιδιακή ταχύτητα $\mathbf{v}$ είναι επίσης πολύ μικρή και ο δεύτερος όρος στην εξίσωση του Euler να αγνοηθεί. Για τον ίδιο λόγο, οι μεταβολές πίεσης και πυκνότητας στο ρευστό επίσης είναι μικρές και θα μπορούσαν να εκφραστούν ως $p = p_0 + p'$, $\rho = \rho_0 + \rho'$ όπου οι παράμετροι ισορροπίας πυκνότητας και πίεσης εκφράζονται μέσω των p_0, ρ_0 αντίστοιχα ενώ οι δεύτεροι όροι εκφράζουν τις μεταβολές τους στο ηχητικό κύμα ($p' << p_0, \rho' << \rho_0$). Η σωματιδιακή ταχύτητα $\mathbf{v}$ που προκαλείται από τη διαταραχή στην πίεση θεωρείται συνεπώς μικρή σε σχέση με την ταχύτητα του ήχου (για τους λόγους που αναλύθηκαν και στην εισαγωγή). Η εξίσωση διατήρησης της μάζας μπορεί να αναπροσαρμοστεί στην παρακάτω μορφή:

$$\rho_0 \nabla \cdot \mathbf{v} + \frac{\partial \rho'}{\partial t} = 0 \tag{2.3}$$

ενώ η εξίσωση του Euler γίνεται:

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{1}{\rho_0}\nabla p' = 0 \tag{2.4}$$

Μια μικρή μεταβολή στην πίεση σχετίζεται με μια μικρή μεταβολή στην πυκνότητα μέσω της σχέσης:

$$p' = \left(\frac{\partial p}{\partial \rho_0}\right)_S \rho' \tag{2.5}$$

ενώ

$$\left(\frac{\partial p}{\partial \rho_0}\right)_S \equiv c^2$$

είναι η σχέση που συνδέει την ταχύτητα του ήχου με το ρυθμό μεταβολής της πίεσης με την πυκνότητα υπό σταθερή εντροπία [9]. Παραγωγίζοντας τη σχέση (2.5) και αντικαθιστώντας το ρυθμό μεταβολής της πυκνότητας από την εξίσωση διατήρησης μάζας (2.1) προκύπτει:

$$\frac{\partial p'}{\partial t} + \rho_0 c^2 \nabla \cdot \mathbf{v} = 0 \tag{2.6}$$

Υπολογίζοντας τη μερική παράγωγο ως προς το χρόνο της παραπάνω εξίσωσης και αντικαθιστώντας την παράγωγο της ταχύτητας από την εξίσωση Euler προκύπτει:

$$\frac{\partial^2 p'}{\partial t^2} - \rho_0 c^2 \nabla \cdot \left(\frac{1}{\rho_0}\nabla p'\right) = 0 \tag{2.7}$$

από όπου θεωρώντας την πυκνότητα σταθερή προκύπτει η κυματική εξίσωση για την πίεση :

$$\nabla^2 p' - \frac{1}{c^2}\frac{\partial^2 p'}{\partial t^2} = 0 \tag{2.8}$$

Η κυματική εξισώση περιγράφει πλήρως ένα ηχητικό κύμα και η εμβέλειά της επεκτείνεται και σε άλλα φυσικά μεγέθη, τη σωματιδιακή ταχύτητα, το δυναμικό ταχύτητας ($\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = 0$), τη μετατόπιση ($\nabla^2\psi - (1/c)^2\psi = 0$) ενώ παρουσία πηγών οι παραπάνω εξισώσεις αναδιαμορφώνονται με την εισαγωγή στο δεξιό μέλος του όρου $f(r,t)$ που εκφράζει την πλεονάζουσα έγχυση όγκου συναρτήσει του χώρου και του χρόνου. [13]. Μεταφέροντας την επίλυση της κυματικής εξίσωσης από το πεδίο του χρόνου στο πεδίο της συχνότητας με μετασχηματισμό Fourier προκύπτει η εξίσωση Helmholtz:

$$[\nabla^2 + k^2(r)]\psi(r,\omega) = f(r,\omega)$$

όπου $k(r)$ είναι ο κυματικός αριθμός που ορίζεται από τη σχέση: $k(r) = \omega/c(r)$.

Παρά τη φαινομενική απλότητα της μορφής της εξίσωσης Helmholtz η επίλυσή της χαρακτηρίζεται από πολυπλοκότητα και πολλαπλότητα στην προσέγγιση και τις τεχνικές επίλυσης. Ενδεικτικά, η λύση της εξαρτάται από το προφίλ της ταχύτητας του ήχου και συνεπώς τον κυματαριθμό, τις συνοριακές συνθήκες, τη γεωμετρία πομπού-δέκτη, τη σύσταση του πυθμένα, τη συχνότητα και το εύρος συχνοτήτων [13] ενώ τα μοντέλα που την επιλύουν αναπτύσσουν αριθμητικές ή αναλυτικές λύσεις ή συνδυασμό τους. Η επίλυση της εξίσωσης Helmholtz οδηγεί σε χαρτογραφήσεις του ακουστικού πεδίου εντός του ωκεανού και η έξοδος των μοντέλων διάδοσης του ήχου μετρά επίπεδο ηχητικής έντασης (Sound Pressure Level) σε κάθε σημείο του τρισδιάστατου χώρου (μονάδες μέτρησης: db re $1\mu Pa/3d$ coordinates). Τα μοντέλα που χρησιμοποιούνται ενσωματώνουν στις εξισώσεις τους την εξάρτηση της λύσης αφενός μεν από το βάθος αφετέρου δε από την ιδιαιτερότητα που χαρακτηρίζει την οριζόντια κλίμακα (κεκλιμένος πυθμένας, υποθαλάσσιες οροσειρές κ.ο.κ.).

2.3 Συμπεράσματα - Συζήτηση

Η παραπάνω ανάλυση μας έφερε σε μια αναγκαστική γνωριμία με τα περισσότερα φυσικά μεγέθη που σχετίζονται με τα ηχητικά φαινόμενα (πίεση, θερμοκρασία, συχνότητα, ταχύτητα, ενέργεια κ.ο.κ.). Το ηχητικό φαινόμενο είναι μετρήσιμο και προσδιορίζεται σε συγκεκριμένες μονάδες μέτρησης οδηγώντας στο σημαντικό συμπέρασμα ότι πειραματικές συσκευές μπορούν αποθηκεύοντας ηχητικά σήματα σε κυματομορφές τάσης -όπως έλεγε ο A.G.Bell- να χρησιμοποιηθούν αργότερα για επαναμετρατροπή τους σε db re 1μ Pa και να ακολουθήσει περαιτέρω ανάλυση και επεξεργασία. Η εργασία που ακολουθεί έχει στηριχτεί σε δεδομένα που έχουν συλλεχθεί από ποντίσεις υδρόφωνων σε διάφορες περιοχές της Ελληνικής Τάφρου, τη χρονική περίοδο 2000-2021, κατά τη διάρκεια ερευνητικών αποστολών του Ελληνικού Κέντρου Θαλασσίων Ερευνών (ΕΛ.ΚΕ.Θ.Ε), του Ινστιτούτου Τεχνολογίας Έρευνας (Ι.Τ.Ε.) και του Ινστιτούτου Κητολογικών Ερευνών Πέλαγος΄.

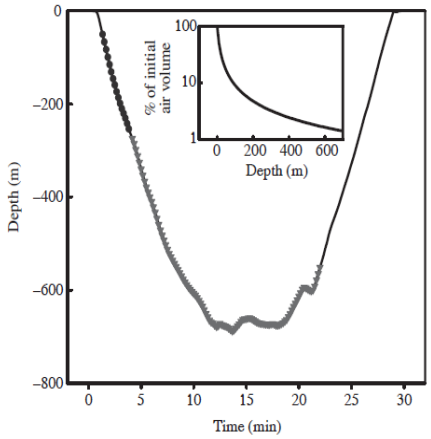
Κεφάλαιο 3

Βιοακουστική

3.1 Εισαγωγή

Σύγχρονες μελέτες παραλληλίζουν την υπερτροφική ρινική δομή των φυσητήρων με βιοακουστική μηχανή, ικανή να παράγει εξαιρετικά θορυβώδεις ήχους –τους πιο θορυβώδεις στο ζωικό βασίλειο (234dB re: 1 µPa at 1m) [14] - μέσω ενός μηχανισμού απότομης συμπίεσης αέρα στα φωνητικά χείλη [1]. Ο πιο κοινός ήχος φυσητήρων είναι τα clicks, που συντίθενται από μια ακολουθία τακτικά χωροθετημένων φθίνοντων παλμών [15]. Συγκεκριμένα εκτιμάται ότι οι φυσητήρες παράγουν αφενός μεν ισχυρά κατευθυντικούς ήχους που χρησιμεύουν στον ηχοεντοπισμό (echolocation) όπως επίσης και λιγότερο κατευθυντικά σήματα που εμφανίζονται ως ακολουθία από στερεότυπα patterns τριών ή περισσοτέρων clicks που εξυπηρετούν στη μεταξύ τους επικοινωνία και φέρουν την ορολογία codas [16].

Σχήμα 3.1: Καμπύλη βάθους κατάδυσης με το χρόνο καί φύση ηχητικών παλμών. Τα codas σημειώνονται στην καμπύλη βάθους κατάδυσης με σφαιρικά σημεία ενώ τα clicks με τρίγωνα



Η εικόνα μεταφέρεται αυτούσια από άρθρο των Madsen et al, (2002) [5].

Η παραπάνω διάκριση που γίνεται ανάμεσα στον τύπο ηχητικού σήματος και τη λειτουργία που επιτελεί ενισχύεται και από το γεγονός ότι η κατευθυντικότητα αντικειμενικά μειώνει το

φυσικό χώρο στον οποίο μπορεί να επιτευχθεί αλληλεπίδραση μεταξύ των θηλαστικών είναι όμως και προϊόν παρατήρησης καθώς τα codas εκπέμπονται σε περιόδους που οι φυσητήρες αναπτύσσουν έντονη κοινωνική συμπεριφορά κοντά στην επιφάνεια στην αρχή και στο τέλος των καταβυθίσεων [17] (σχήμα 3.1) αλλά όχι κατά τη διάρκεια της αναζήτησης θηράματος στο βυθό [5].

Η διαφοροποίηση των ήχων των codas οριοθετεί διακριτούς φωνητικούς πληθυσμούς φυσητήρων (vocal clans). Στον Ατλαντικό ωκεανό η διακριτοποίηση αυτή φαίνεται ότι προσδιορίζεται κάθετα από τη γεωγραφικότητα των πληθυσμών καθώς σε κάθε θαλάσσια γειτονιά μόνο ένα ρεπερτόριο σημάτων codas επικρατεί. Δεν ισχύει όμως αυτό στον Ειρηνικό ωκεανό, όπου σε φωνητικές φυλές η κοινωνικοποίηση των φυσητήρων γίνεται οριζόντια ανάμεσα σε οντότητες που χρησιμοποιούν την ίδια διάλεκτο και δεν σχετίζεται απαραίτητα με τη γεωγραφική εγγύτητα [16]. Στη Μεσόγειο το πιο χαρακτηριστικό πρότυπο σημάτων coda φυσητήρων είναι το '3+1' αποτελείται δηλαδή από 3 κλικ με απόσταση ενός δευτερολέπτου μεταξύ τους και ένα τέταρτο κλικ που ακολουθεί δύο δευτερόλεπτα μετά. Ήδη στο ανεξάντλητο αυτό προς έρευνα θέμα προσαρμόζονται μοντέλα μηχανικής μάθησης που διερευνούν τη σχέση πληθυσμών και διαλέκτων σημάτων ςcodas.

3.2 Μηχανισμός παραγωγής βιοηχητικού σήματος οδοντοκητών

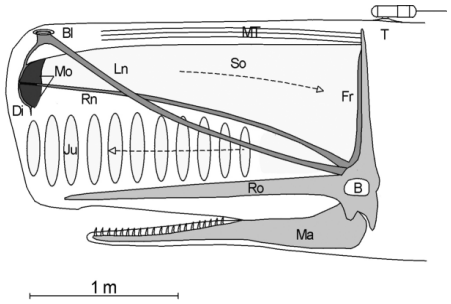
Ο βασικός ρόλος της ρινικής δομής στα οδοντοκήτη αφορά στην εξασφάλιση της αναπνευστικής λειτουργίας ενώ παράλληλα στο ρινικό σύμπλεγμα τα θαλάσσια κήτη παράγουν ακουστικούς ήχους οι οποίοι μεταδίδονται στο θαλάσσιο περιβάλλον[1]. Για την παραγωγή ήχου, τα οδοντοκήτη διαθέτουν φωνητικά χείλη (phonic lips) που συνιστούν ειδικές βιολογικές βαλβίδες ικανές να πάλλονται σε ρεύματα αέρα παράγοντας και κατευθύνοντας ηχητικά κύματα προς γειτονικούς ιστούς που εστιάζουν την ηχητική δέσμη πριν την εκπομπή της στο θαλάσσιο περιβάλλον [14].

Τρεις βασικές ερωτήσεις που καλείται να απαντήσει η βιοακουστική για την παραγωγή βιοηχητικών σημάτων αφορούν στα δομικά βιολογικά στοιχεία που εμπλέκονται στην παραγωγή ήχων, τη σχετική τους θέση τους ως προς τα άλλα βιολογικά όργανα, τέλος τον τρόπο με τον οποίο τελικά παράγονται τα βιοσήματα [18]. Για να γίνει εμφανής ο τρόπος που παράγεται ο ήχος στη ρινική δομή του κήτους παραθέτουμε μια βασική εικόνα της ανατομίας του κεφαλιού ενός φυσητήρα. Εντός του ρινικού συμπλέγματος των φυσητήρων συγκροτούνται 2 μεγάλοι κηρώδεις θύλακες λίπους. Ο άνω θύλακας περιβάλλεται από μια επιμήκη δομή που έχει σχήμα στρογγυλής κωνικής διατομής και ορίζεται από την ονομασία spermaceti organ (σπερμακετικό όργανο), εκτείνεται δε από την κοίλη επιφάνεια πάνω από το κρανίο μέχρι το πρόσθιο άκρο του μετώπου [18] κατά μήκος του ραχιαίου τμήματος του κεφαλιού του κήτους.

[1] Η όσφρηση θεωρείται εξαιρετικά περιορισμένη στα οδοντοκήτη.

Αποτελούσε αντικείμενο εκτεταμένης κητοθηρίας εξαιτίας της υψηλής ποιότητας σπερμακετικού ελαίου που περιέχεται σε αυτό ενώ ο χαμηλότερος θύλακας που ονομάστηκε από τους κητοθήρες ΄κάδος΄ (junk) και καθιερώθηκε ως τέτοιος στην ορολογία και των βιολόγων [2] αποτελείται από πυκνότερο λίπος που οργανώνεται σε τομείς εγκάρσιων διαιρέσεων συνδετικού ιστού που μοιάζουν με φακούς. Επισημαίνεται ότι η κατευθυντικότητα της ακουστικής δέσμης σχετίζεται κατά τους βιοακουστικούς επιστήμονες με τη λειτουργία των τομών αυτών ως ηχητικές εστίες. Οι φυσητήρες φέρουν ένα μόνο εξωτερικό ρινικό άνοιγμα (ρουθούνι) εκπνοής που συνιστά απόληξη δύο διακριτών ρινικών περασμάτων (nasal pasages) των οποίων οι προεκτάσεις ενώνονται τελικά πριν την εξωτερική οπή στην αριστερή πλευρά της άκρης της μύτης του κεφαλιού του κήτους.

Σχήμα 3.2: Ανατομία του κεφαλιού ενός φυσητήρα Mo: Spermaceti organ, Ju: Junk, Di: Distal sac, Fr: Frontal sac



Η εικόνα αναπαράγεται αυτούσια από το άρθρο των (Mohl et al, 2001)) [16]

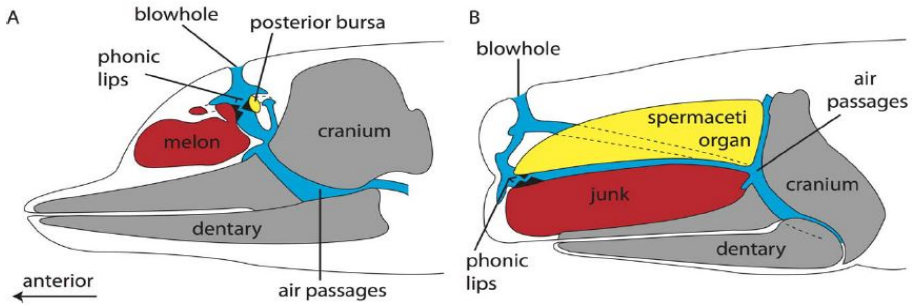
Το βασικό μοντέλο των Mohl et Al εξηγεί την παραγωγή του ηχητικού σήματος εκτιμώντας ότι το πρωταρχικό γεγονός συνίσταται σε εξαναγκασμένη ροή συμπιεσμένου αέρα από το δεξιό ρουθούνι του κήτους Rn ανάμεσα στα φωνητικά χείλη καθώς αυτά ανοίγουν απειροστά για να σχηματίσουν σχισμή. Η προώθηση του αέρα στα φωνητικά χείλη οδηγεί σε διαρροή μικρής ποσότητας ενέργειας εμπρόσθια κατευθείαν προς το θαλάσσιο νερό (τα βέλη στην εικόνα δείχνουν την κατεύθυνση του ήχου) προκαλώντας έναν παλμό χαμηλού πλάτους (P0) ενώ το ηχητικό σήμα οπισθοδιαδίδεται μέσα στο σπερμακετικό όργανο (spermaceti organ). Παρά το γεγονός ότι η ενέργεια του παλμού (P0) είναι μικρότερη από το 10% της παραγόμενης ηχητικής ενέργειας αυτός διαδίδεται προς όλες τις κατευθύνσεις και ανιχνεύεται σε μεγάλες αποστάσεις (και από τα υδρόφωνα) ενώ παράγει ηχώ από ανάκλαση στο βυθό και στην επιφάνεια της θάλασσας [19].

Στο μοντέλο που αναπτύχθηκε από τον Mohl (Bent horn model) οι δύο κοιλότητες αέρα (μετωπιαίος θύλακας Fr και περιφερικός θύλακας E) λειτουργούν ως ανακλαστήρες του σήματος που παράγεται στα φωνητικά χείλη του κήτους. Το ηχητικό κύμα που οπισθοδιαδίδεται αντηχεί αρκετές φορές εντός του σπερμακετικού οργάνου (spermaceti organ) ενώ κατά τη

[2] Ονομάστηκε έτσι από τους κητοθήρες υποδηλώνοντας τη μη χρησιμότητά του για τους ίδιους και υιοθετήθηκε η ορολογία και από τους θαλάσσιους βιολόγους, παρ΄όλα αυτά επιτελεί ουσιαστικές λειτουργίες για τους φυσητήρες.

διάρκεια κάθε κύκλου, μέρος της ηχητικής ενέργειας εξέρχεται στον κάδο (junk). Συγκεκριμένα το ηχητικό σήμα ανακλάται στον μετωπιαίο θύλακα αέρα (frontal air sac) -μπροστά από τα σαγόνια- με κατεύθυνση προς τα κάτω και μπροστά εντός του κάδου (junk) από όπου διαδίδεται στο θαλάσσιο νερό ως ο παλμός (p1) που είναι ισχυρά κατευθυντικός. Από το σημείο αυτό ο ήχος μπορεί να οδηγηθεί μέσω των φακών και του ενδιάμεσου συνδετικού ιστού στο μπροστινό μέρος του κεφαλιού και να διαδοθεί στο νερό. Η εναπομείνασα ενέργεια ανακλάται από τον μετωπιαίο θύλακα αέρα πίσω στο σπερμακετικό όργανο (spermaceti organ) όπου επιστρέφει στον περιφερικό θύλακα επαναλαμβάνοντας το ίδιο pattern παραγωγής ηχητικών παλμών (p2, p3) από διαδοχικές ηχητικές ανακλάσεις ανάμεσα στους περιφερικούς και μετωπιαίους θύλακες αέρα.

Σχήμα 3.3: Απλουστευμένη σχηματική αναπαράσταση της δεξιάς πλευράς του δελφινιού (που είναι συμμετρική με την αριστερή) ως προς το δεξιό τμήμα του κεφαλιού ενός φυσητήρα (που δεν εμφανίζει συμμετρία).



Η εικόνα μεταφέρεται αυτούσια από το άρθρο των Berta et al. [20]

Τα οδοντοκήτη που δεν ανήκουν στην οικογένεια των φυσητηροειδών διαθέτουν δύο ζεύγη φωνητικών χειλιών ενώ η πρόσληψη αέρα από τους πνεύμονες του κήτους προϋποθέτει το πέρασμα του μέσα από τις βιολογικές δομές που παράγουν ήχους, γεγονός που οφείλεται στη συμμετρία που χαρακτηρίζει το αριστερό και δεξιό ημισφαίριο του κεφαλιού τους. Αντιθέτως στους φυσητήρες ένα μόνο ζεύγος φωνητικών χειλιών είναι σε σύνδεση με το δεξιό ρινικό πέρασμα (nasal passage) για την παραγωγή ήχων ενώ το αριστερό ρινικό πέρασμα παρακάμπτει το μηχανισμό παραγωγής ήχου και εξειδικεύεται στην αναπνευστική λειτουργία συνδέοντας το εξωτερικό ρινικό άνοιγμα (ρουθούνι) με τους πνεύμονες [1]. Διαθέτουν συνεπώς οι φυσητήρες ενιαίο σύμπλεγμα παραγωγής ήχων αφενός μεν εξαιτίας της υπερμεγέθους επέκτασης της ρινικής δομής στο δεξιό μέρος του κεφαλιού αφετέρου δε στην εκτιμούμενη αναδιοργάνωση και μείωση των λειτουργικών στοιχείων στο αριστερό τμήμα του κεφαλιού κατά τη διάρκεια της εξέλιξης [20]. Στο παρακάτω σχήμα φαίνονται σε αντιπαραβολή οι αντιστοιχίες ως προς τη λειτουργικότητα ανάμεσα στην ανατομία των δελφινιών και των φυσητήρων.

Στην παραπάνω εικόνα επιχειρείται μια υπεραπλουστευμένη αντιστοιχία των ομόλογων βιολογικών οργάνων δελφινιών και φυσητήρων που εμπλέκονται στην παραγωγή βιολογικών

ήχων. Η λιπώδης μάζα ιστών που ορίζεται με την ονομασία (melon)[3] περιλαμβάνει το μεγαλύτερο μέρος του μετώπου μικρών οδοντοκητών όπως τα δελφίνια, είναι το συμμετρικό αντίστοιχο του κάδου junk του φυσητήρα και επιτελεί παρόμοιο λειτουργικό ρόλο στα δύο οδοντοκήτη, δηλαδή στη μετάδοση και επεξεργασία ηχητικού σήματος [18]. Επίσης ο ραχιαίος θύλακας αρθρικού υγρού (dorsal bursa) θεωρείται ομόλογο όργανο του σπερμακετικού οργάνου (spermacetic organ) του φυσητήρα καθώς και τα δύο όργανα εμφανίζονται στην ίδια πλευρά της κύριας διόδου αέρα που οδηγεί στο ρουθούνι. Ασφαλώς αυτή η αντιστοίχιση λειτουργιών και βιολογικών οργάνων πρέπει να γίνει με εξαιρετική προσοχή αφού οι φυσητήρες έχουν προικιστεί με το μεγαλύτερο ρύγχος στο ζωικό βασίλειο, το μέγεθός τους είναι πολύ μεγαλύτερο από αυτό των δελφινιών και η ανατομία τους φανερά διακριτή.

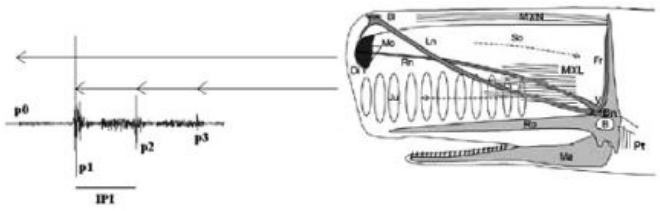
[3]Ονομασία που επίσης δόθηκε από τους κητοθήρες

3.3 Ακολουθία ηχητικών σημάτων οδοντοκητών

3.3.1 Σήματα φυσητήρων

Τα clicks που παράγουν οι φυσητήρες συντίθενται από τρεις χαρακτηριστικούς παλμούς (P0, P1, P2) που είναι προϊόντα του ίδιου ακουστικού γεγονότος, της δημιουργίας ενός βραχέος παλμού στα φωνητικά χείλη[19]. Συγκεκριμένα ένα μικρό μέρος της ηχητικής ενέργειας διαδίδεται εμπρόσθια (παλμός P0) ενώ το μεγαλύτερο μέρος της ηχητικής ενέργειας διοχετεύεται προς τα πίσω μέσω της ιδιαίτερης γεωμετρίας των φωνητικών χειλιών (phonic lips) και ανακλάται μέσω του εμπρόσθιου θύλακα αέρα (Madsen et Al, 2002) (παλμός P1) οπότε και εξέρχεται από το όργανο junk του φυσητήρα [1]. Τέλος ένα υπολοιπόμενο μέρος ενέργειας ανακλάται στον περιφερικό θύλακα αέρα και δημιουργεί ένα νέο ανακλώμενο ηχητικό παλμό που ανιχνεύεται ως παλμός P2. Είναι χρήσιμο να σημειωθεί ότι αν το κήτος εκπέμπει ήχους εκτός άξονα δηλαδή βρίσκεται πίσω από το όργανο καταγραφής η ενέργεια του πανκατευθυντικού παλμού μπορεί να αποτυπώνεται υψηλότερη από την ενέργεια του παλμού P1, οι κυματομορφές συνεπώς που καταγράφονται διαφέρουν ανάλογα με τη σχετική θέση του θηλαστικού ως προς το όργανο μέτρησης [21].

Σχήμα 3.4: Παλμοσειρά φυσητήρα, αναπαράγεται από άρθρο των (Wahlberg et al.) [1]



3.3.2 Σήματα ζωνοδέλφινων

Οι ήχοι που εκπέμπονται από τα δελφινοειδή κατηγοριοποιούνται σε τρεις γενικές κλάσεις: Στα μεμονωμένα clicks ηχοεντοπισμού, σε ακολουθία διαδοχικών παλμικών ήχων (burst pulse signals) και σε σφυρίγματα (whistles). Τα απομονωμένα clicks ηχοεντοπισμού είναι σήματα βραχέος χρόνου και ευρείας ζώνης (broadband) συχνοτήτων (μπορούν να φτάσουν σε συχνότητες μεγαλύτερες των 100 kHz). Οι ακολουθίες διαδοχικών clicks είναι επίσης σήματα υψηλών συχνοτήτων που διαχωρίζονται μεταξύ τους από στοιχειώδη χρονικά διαστήματα (inter-click intervals). Σε αντιδιαστολή με τα διακριτά σήματα, τα σφυρίγματα των δελφινιών είναι συνεχείς ήχοι, στενής συχνοτικής ζώνης (narrow band) που διαρκούν από μερικά δέκατα του δευτερολέπτου έως μερικά δευτερόλεπτα και η θεμελιώδης τους συχνότητά έχει εύρος από 2-30 kHz [22].

3.3.3 Συμπεράσματα - συζήτηση για τον χώρο εισόδου

Οι ακολουθίες παλμών φυσητήρων και ζωνοδέλφινων ή τα συνεχή σήματα που εκπέμπουν τα δελφίνια μας οδηγούν σε κάποια πρώτα συμπεράσματα σχετικά με το χώρο εισόδου του προβλήματος μηχανικής μάθησης που ακολουθεί. Τα clicks των ζωνοδέλφινων διαφοροποιούνται από τα clicks των φυσητήρων ως προς το πλάτος και την ενέργεια που μεταφέρουν, διακρίνονται όμως και ως προς τα συχνοτικά χαρακτηριστικά, καθώς τα δελφίνια μπορούν να εκπέμπουν σε αρκετά υψηλότερες συχνότητες όπως επίσης διαφοροποιούνται και στα χρονικά διαστήματα (inter-click interval) μεταξύ των clicks δεδομένου ότι τα δελφίνια παράγουν πυκνότερες ακολουθίες παλμών. Συνεπώς οι δυνατές επιφάνειες διαχωρισμού των διαφορετικών σημάτων των οδοντοκητών θα μπορούσαν να κατασκευαστούν σε ένα χώρο χαρακτηριστικών που παραμετροποιείται ως προς τις συχνοτική και τη χρονική δομή. Για τους παραπάνω λόγους τα μοντέλα που προσαρμόζουμε θα έπρεπε να ενσωματώσουν στη μηχανική τους τις διαστάσεις της ενέργειας, φάσματος, χρόνου στην εξαγωγή διανυσμάτων χαρακτηριστικών.

Κεφάλαιο 4

Ψηφιακή Επεξεργασία Σήματος και Εικόνας

4.1 Εισαγωγή

Τα σήματα αναπαρίστανται μαθηματικά με συναρτήσεις μιας ή περισσοτέρων ανεξάρτητων μεταβλητών. Τα ηχητικά σήματα αναπαρίστανται με χρονοσειρές ενώ μια εικόνα ορίζεται ως μια δισδιάστατη συνάρτηση f(x,y), χωρικών συντεταγμένων (x,y), της οποίας το πλάτος σε κάθε εικονοστοιχείο (pixel) ορίζει την ένταση της εικόνας στο σημείο αυτό. Το εύρος των τιμών της έντασης για δυαδικές εικόνες (ασπρόμαυρες) είναι [0,1] ενώ για εικόνες αποχρώσεων του γκρι είναι [0,255][23]. Η γραφική αναπαράσταση ενός σήματος στο πεδίο του χρόνου δεν εμπεριέχει με προφανή τρόπο τις περισσότερες φορές φασματική πληροφορία για αυτό [24] και στην παράγραφο που ακολουθεί εστιάζουμε στην εξαγωγή συχνοτικών χαρακτηριστικών από σήματα συνεχή ή διακριτά στο χρόνο και τη συχνότητα κάνοντας μια σύντομη αναφορά στο μετασχηματισμό Fourier και μετεξελίξεις του, τον μετασχηματισμό Gabor ή Short Time Fourier Transform (S.T.F.T.) και τον μετασχηματισμό Wavelet. Στο δεύτερο μέρος του κεφαλαίου 4 το ενδιαφέρον εστιάζεται σε σήματα εικόνας και σε τεχνικές εξαγωγής σημείων ενδιαφέροντος που συνιστά βασική επιδιώξη στην όραση υπολογιστών.

4.2 Ψηφιακή Επεξεργασία Σήματος

4.2.1 Μετασχηματισμός Fourier(Fourier transform)

Ο μετασχηματισμός Fourier εφαρμόζεται σε στάσιμα σήματα, δηλαδή, σήματα των οποίων οι ιδιότητες δεν αλλάζουν με το χρόνο. Για συνεχή σήματα x(t) που χαρακτηρίζονται από την ιδιότητα αυτή, το ζεύγος μετασχηματισμού Fourier ορίζεται από την εξίσωση ανάλυσης:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) exp(-j\omega t)\, dt \tag{4.1}$$

ή

$$X(f) = \int_{-\infty}^{\infty} x(t) exp(-2j\pi f t)\, dt \tag{4.2}$$

όπου ω η κυκλική συχνότητα με μονάδα μέτρησης rad/sec, ενώ f η γραμμική συχνότητα συνεχούς χρόνου σε Hz και από την εξίσωση σύνθεσης:

$$x(t) = \frac{1}{2\pi} \int X(\omega) e^{j\omega t} d\omega \tag{4.3}$$

Οι συντελεστές X(f) εκφράζουν την έννοια της συνολικής συχνότητας (global frequency, f) σε ένα σήμα και υπολογίζονται ως εσωτερικά γινόμενα του σήματος με ημιτονοειδείς συναρτήσεις βάσης άπειρης χρονικής διάρκειας [2]. Για το λόγο αυτό, η ανάλυση Fourier απεικονίζει σωστά το συχνοτικό περιεχόμενο στάσιμων σημάτων αφού απότομες μεταβολές στις ιδιότητες των σημάτων που αλλοιώνουν τη στασιμότητα (όπως μεταβατικά γεγονότα) διαχέονται σε όλο τον άξονα των συχνοτήτων X(f).

Ο μετασχηματισμός ενός απεριοδικού σήματος ονομάζεται ΄φάσμα΄ (spectrum) καθώς εμπεριέχει πληροφοριακό περιεχόμενο για το πως το σήμα $x(t)$ συντίθεται από ημιτονοειδή σήματα διαφορετικής συχνότητας. Ο μηχανισμός που καθιστά τον FT χρήσιμο εργαλείο στη φασματική αναπαράσταση ενός σήματος έγκειται στο γεγονός ότι για κάθε συχνότητα το σήμα συγκρίνεται με το μιγαδικό εκθετικό αποδίδοντας υψηλούς συντελεστές Fourier όταν το φάσμα τους περιέχει τη συχνότητα του εκθετικού ενώ αν ένα σήμα δεν περιέχει φασματική συνιστώσα στη συχνότητα αυτή ο συντελεστής Fourier μηδενίζεται[24]. Παρόλα αυτά, ο μετασχηματισμός Fourier δεν φέρει πληροφορία για το χρονικό εντοπισμό των φασματικών συνιστωσών.

Η τυπική παραδοχή για την περίπτωση μη στάσιμων σημάτων συνήθως γίνεται εισάγοντας μια ΄τοπική συχνότητα΄ (local frequency in time) σε ένα χρονικό παράθυρο στο οποίο προσεγγιστικά το σήμα μπορεί να θεωρηθεί στάσιμο. Ένας δεύτερος ισοδύναμος τρόπος για την επέκταση σε περιπτώσεις μη στάσιμων σημάτων θα μπορούσε να υλοποιηθεί με την τροποποίηση των ημιτονοειδών συναρτήσεων βάσης σε συναρτήσεις περισσότερο εντοπισμένες στο χρόνο αλλά λιγότερο εντοπισμένες στη συχνότητα.

4.2.2 Μετασχηματισμός Fourier διακριτού χρόνου (D.T.F.T.)

Ο μετασχηματισμός Fourier διακριτού χρόνου (D.T.F.T.) ορίζεται από το ζεύγος των εξισώσεων:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega \dot{n}}$$

$$x[n] = \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega n} d\omega$$

και είναι περιοδικός ως προς τη συχνότητα ω με περίοδο 2π συνεπώς η συμπεριφορά του καθορίζεται πλήρως σε αυτό το συχνοτικό διάστημα σε αντίθεση με το συνεχή στο χρόνο μετασχηματισμό που εκτείνεται σε όλο το φάσμα των συχνοτήτων. Ο D.T.F.T προκύπτει περιορίζοντας τον μετασχηματισμό Z στο μοναδιαίο κύκλο θέτοντας $z = e^{j\omega}$ και είναι καλά

ορισμένος αν το πεδίο σύγκλισης του $X(z)$ τον περιέχει. Πρέπει συνεπώς να ισχύει η συνθήκη αθροισιμότητας της απόλυτης τιμής του διακριτού σήματος στο χρόνο: $\sum_{n=-\infty}^{\infty} |x[n]| < \infty$. Ο μετασχηματισμός $X(e^{j\omega})$ λέγεται και φάσμα (spectrum) του σήματος $x[n]$ διότι όπως και ο συνεχής μετασχηματισμός Fourier, περιέχει πληροφορία για τον τρόπο με τον οποίο το σήμα $x[n]$ συντίθεται από μιγαδικά εκθετικά διαφορετικών συχνοτήτων[25].

4.2.3 Διακριτός Μετασχηματισμός Fourier (D.F.T.)

Ο D.F.T μπορεί να θεωρηθεί ως μια δειγματοληπτημένη εκδοχή πεπερασμένου μήκους του μετασχηματισμού D.T.F.T, δηλαδή

$$X[k] = X(e^{j\omega})|_{\omega=(2\pi k/n)}, k = 0, 1, .., N-1$$

Πράγματι, δεδομένης της δυσκολίας υλοποίησης του DTFT αφενός μεν λόγω του άπειρου αθροίσματος αφετέρου δε λόγω της συνέχειας της μεταβλητής της συχνότητας ορίζουμε τον μετασχηματισμό D.F.T στον οποίο η συχνότητα είναι διακριτή μεταβλητή και αντιστοιχεί σε δείγματα του DTFT. Συγκεκριμένα για ένα διακριτό σήμα που αναπαρίσταται στο πεδίο του χρόνου από μια ακολουθία N ακεραίων $x_0, x_1, ..., x_{N-1}$ ο διακριτός μετασχηματισμός Fourier, (D.F.T.) εκφράζεται από το άθροισμα:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{(-j2\pi k/N)n}, k = 0, 1, ..., N-1$$

όπου έχει προκύψει από δειγματοληψία του DTFT $X(e^{j\omega})$ σε N ισαπέχουσες συχνότητες $\omega = (2 \cdot \pi \cdot k/N)$ δηλαδή τα δείγματα του DTFT στις συχνότητες αυτές ταυτίζονται με τις τιμές του DFT που υπολογίζονται από την $X[k]$.

4.2.4 Μετασχηματισμός Gabor ή ανάλυση Fourier βραχέος χρόνου (S.T.F.T.)

Ο μετασχηματισμός Fourier προσαρμόστηκε από τον Gabor προκειμένου να εφαρμοστεί σε σήματα που δεν χαρακτηρίζονται από στασιμότητα. Συγκεκριμένα, ο μετασχηματισμός Gabor κάνει την παραδοχή ότι το σήμα x(t) είναι κατά προσέγγιση στάσιμο για ένα πολύ μικρό χρονικό παράθυρο. Ο κύριος σκοπός της παραθύρωσης του σήματος είναι αφενός ο προσδιορισμός της έκτασης της ακολουθίας στην οποία θα εφαρμοστεί ο μετασχηματισμός, έτσι ώστε τα φασματικά χαρακτηριστικά να είναι σχετικά σταθερά κατά μήκος του παραθύρου αφετέρου η λείανση (smoothing) που προκύπτει από τη συνέλιξη του φάσματος της ακολουθίας με το φάσμα του παραθύρου. Ο μετασχηματισμός Fourier των παραθυρωμένων σημάτων $x(t)g(t-t)$ οδηγεί στον μετασχηματισμό Gabor ή Short Time Fourier Transform (S.T.F.T.) ο οποίος αντιστοιχεί το σήμα σε μια δισδιάσταση συνάρτηση στο χρονοσυχνοτικό επίπεδο (τ, f).

$$STFT(\tau, f) = \int_{-\infty}^{\infty} x(t) g^*(t-\tau) exp(-2j\pi f t)\, dt$$

αν η συνάρτηση αναφοράς είναι συνεχής, διαφορετικά για διακριτή συνάρτηση ο μετασχηματισμός γράφεται:

$$X_n[k] = \sum_{n=0}^{N-1} x[n]w[n-k]e^{(-j2\pi k/N)n}, k = 0, 1, ..., N-1$$

Η παραπάνω μαθηματική διατύπωση του Fourier μετασχηματισμού μπορεί να ερμηνευτεί ποιοτικά με δύο τρόπους ανάλογα με τον παράγοντα k ή n που θεωρούμε σταθερό καθώς ο άλλος μεταβάλλεται.

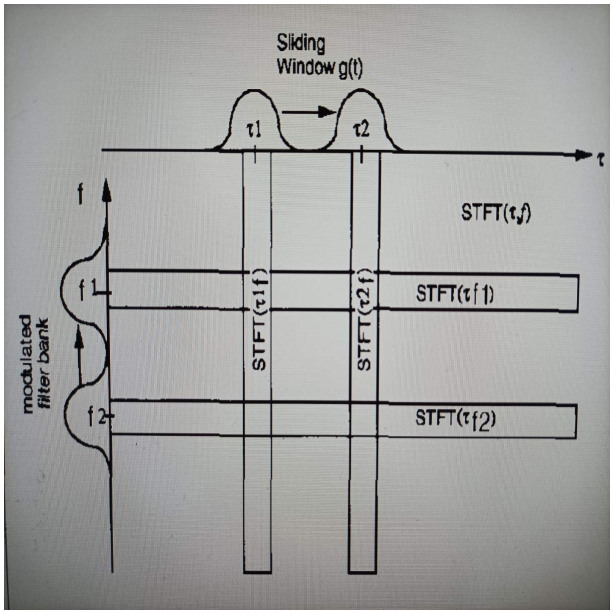
1. Ο πρώτος τρόπος εκφράζει τη θεώρηση ότι για κάθε συγκεκριμένο στοιχείο του παραθυρωμένου χρονικά σήματος n ο μετασχηματισμός υπολογίζει όλες τις STFT συχνότητες σε αυτό.

2. Η δεύτερη θεώρηση ερμηνεύει τον μετασχηματισμό ως μεταβαλλόμενη συνάρτηση του n για μια δεδομένη συχνότητα $\omega_k = 2\pi k/N$. Υπολογίζει συγκεκριμένα το μετασχηματισμό ως τη γραμμική συνέλιξη ενός βαθυπερατού φίλτρου $w[n]$ στο σήμα $x[n]e^{-j(2\pi k/N)n}$ δηλαδή:

$$X_n[k] = w[n] * [x[n]e^{-j(2\pi k/N)n}]$$

Μπορεί συνεπώς να ερμηνευτεί ο μετασχηματισμός ως εφαρμογή βαθυπερατού φίλτρου στο μετατοπισμένο φάσμα του $x[n]$ γύρω από μια περιοχή της συχνότητας ω_k. Οι δύο παραπάνω εκφράσεις μπορούν να περιγραφούν με το παρακάτω σχήμα που εκτίθεται αυτούσιο από άρθρο των Rioul, Vetterli.

Σχήμα 4.1: Η εικόνα μεταφέρεται αυτούσια από το άρθρο των Rioul, Vetterli [2]



Εναλλακτικές ερμηνείες του STFT στο χρονοσυχνοτικό επίπεδο. Στο κάθετο παραλληλόγραμμο ο μετασχηματισμός εφαρμόζεται σε κυλιόμενο χρονικά παράθυρο, ενώ στο οριζόντιο παραλληλόγραμμο ο μετασχηματισμός ερμηνεύεται ως εφαρμογή συστοιχίας ζωνοπερατών φίλτρων σε όλο το σήμα χρονικά με διαμόρφωση της συνάρτησης παραθύρου.

Στην πρώτη περίπτωση ο μετασχηματισμός εστιάζει σε μια κάθετη τομή του STFT και αντιστοιχεί στον μετασχηματισμό Fourier ενός μόνο χρονικού παραθύρου του σήματος ενώ η δεύτερη εκδοχή του μετασχηματισμού αντιστοιχεί στην οριζόντια τομή του σχήματος όπου φαίνεται η μεταβολή του συχνοτικού περιεχομένου σε μια στενή ζώνη στη συχνότητα $f_1, f_2, ...$ γύρω από την οποία εστιάζουμε. Στο σχήμα 4.1 στα κάθετα παραλληλόγραμμα ο STFT ερμηνεύεται σαν ακολουθία μετασχηματισμών ενός παραθυρωμένου τμήματος του σήματος ενώ στα οριζόντια παραλληλόγραμμα, ο μετασχηματισμός ερμηνεύεται ως εφαρμογή σε όλο το σήμα στο χρόνο για δεδομένη συχνότητα, συστοιχίας ζωνοπερατών φίλτρων με κρουστική απόκριση τη συνάρτηση παραθύρου διαμορφωμένη σε αυτή τη συχνότητα.

4.2.5 Αρχή της αβεβαιότητας

Ο μετασχηματισμός Gabor έχει σταθερή χρονοσυχνοτική ανάλυση το οποίο συνεπάγεται ότι μεταχειρίζεται υψηλές και χαμηλές συχνότητες με τον ίδιο τρόπο και από τη στιγμή που το μήκος του παραθύρου επιλεγεί η χρονοσυχνοτική ανάλυση (resolution) είναι σταθερή για όλο το χρονοσυχνοτικό επίπεδο [2]. Όσο το χρονικό παράθυρο γίνεται μικρότερο τόσο η φασματική ανάλυση/ευκρίνεια ελαττώνεται και αντίστροφα. Η επιλογή στενού χρονικού παραθύρου συμβάλλει στην επιλεξιμότητα και ικανότητα διάκρισης γεγονότων (events) που εκτυλίσσονται σε μικρά χρονικά διαστήματα με κόστος όμως χειρότερη συχνοτική ανάλυση δηλαδή οι συχνότητες στο σήμα που βρίσκονται κοντά μεταξύ τους δεν μπορούν να διακριθούν εύκολα. Συγκεκριμένα, αποδεικνύεται μαθηματικά ότι το γινόμενο της ανάλυσης (resolution) στη συχνότητα και στο χρόνο είναι κάτω φραγμένο και ισχύει ότι: $\delta t \cdot \delta f \geq 1/4\pi$ όπου $\delta t^2 = \frac{\int t^2 g(t)^2}{\int G(t)^2}$ και $\delta f^2 = \frac{\int f^2 g(f)^2}{\int G(f)^2}$. Η αρχή της αβεβαιότητας του Gabor επισημαίνει τελικά μια εγγενή αδυναμία του μετασχηματισμού να επιτύχει ευκρινή ανάλυση στο χρόνο και στη συχνότητα ταυτόχρονα.

4.2.6 Φασματογραφήματα (Spectrograms)

Το φασματογράφημα (Spectrogram) υπολογίζεται από το άθροισμα τετραγώνων πραγματικών και μιγαδικών συντελεστών (squared modulus) του STFT και αναπαριστά την κατανομή της ενέργειας ενός σήματος στο χρονοσυχνοτικό επίπεδο. Το φασματογράφημα σε αντίθεση με τον STFT δεν είναι αντιστρέψιμο αφού η φάση που είναι απαραίτητη για την ανακατασκευή του σήματος δεν μπορεί να ανακτηθεί μετά τον τετραγωνισμό πραγματικών και φανταστικών συνιστωσών του σήματος. Σε συμφωνία με την αρχή της αβεβαιότητας που προηγήθηκε, φασματογραφήματα στενής συχνοτικής ζώνης (narrowband spectrograms), χαρακτηρίζονται από καλή ανάλυση στη συχνότητα και αντίστοιχα αβεβαιότητα στο χρόνο ενώ ευρυζωνικά φασματογραφήματα (wideband spectrograms) χαρακτηρίζονται από καλή ανάλυση στο χρόνο (περιορισμένο χρονικό παράθυρο) και αβεβαιότητα στο πεδίο της συχνότητας.

4.2.7 Μετασχηματισμός Κυματιδίων (Wavelets transform)

Το βασικό μειονέκτημα του μετασχηματισμού Fourier είναι το σταθερό χρονικό παράθυρο ανάλυσης σε όλη τη διάρκεια του σήματος. Η ιδέα πίσω από το μετασχηματισμό Wavelet

είναι η χρησιμοποίηση παραθύρου μεταβλητού μήκους, συγκεκριμένα η εφαρμογή στενού παράθυρου στο χρόνο στις υψηλές συχνότητες και παράθυρου μεγάλης διάρκειας στις χαμηλές συχνότητες. Τα κυματίδια είναι συναρτήσεις βάσης μεταβαλλόμενης συχνότητας και πεπερασμένης διάρκειας σε αντίθεση με τα ημίτονα-συνημίτονα του μετασχηματισμού Fourier που εκτείνονται σε άπειρη διάρκεια και έχουν σταθερή συχνότητα. Οι συναρτήσεις βάσης μπορούν να κατασκευαστούν από μια συνάρτηση που λέγεται μητρικό σωματίδιο που κλιμακώνουμε κατά s και μετατοπίζουμε κατά τ: $\psi = \frac{1}{|s|^{0.5}}\psi(\frac{t-\tau}{s})$. Μεγάλα $s > 1$ αντιστοιχούν σε κυματίδια βάσης με ευρύ μήκος και χαμηλή συχνότητα ενώ για μικρά $s < 1$ οι συναρτήσεις βάσεις θα χαρακτηρίζονται από στενότητα και υψηλή συχνότητα. Η κλίμακα συνιστά συνεπώς μέτρο της διάρκειας των συναρτήσεων βάσης. Ο μετασχηματισμός Wavelet προκύπτει τελικά για κάθε ζευγάρι κλίμακας και χρόνου από το ολοκλήρωμα του γινομένου του σήματος με το μητρικό σωματίδιο μετατοπισμένο κατά τ και κλιμακωμένο κατά s:
$WT(s,\tau) = \frac{1}{|s|^{0.5}} \int x(t)\psi(\frac{t-\tau}{s})dt$

4.3 Ψηφιακή Επεξεργασία Εικόνας

4.3.1 Αναγνώριση ακμών

Μελέτες στη σύγχρονη βιολογία συγκλίνουν στην εκτίμηση ότι ο μηχανισμός της ανθρώπινης όρασης στο πρώιμο στάδιο εντοπίζει καταρχάς τις ακμές των αντικειμένων πριν επικεντρωθεί στις λεπτομέρειές τους. Οι ακμές μιας εικόνας –με τη σειρά τους- σχετίζονται με τα ασυνεχή όρια ή σύνορα ενός φυσικού αντικειμένου ή μιας επιφάνειας σε μια σκηνή. Κατανοούμε συνεπώς τις ακμές ως μεταβολές ή ασυνέχειες φυσικών ιδιοτήτων των εικονιζόμενων 3-Δ αντικειμένων που μπορεί για παράδειγμα να αφορούν στο βάθος, την υφή ή την ανταναακλαστικότητα της επιφάνειάς τους. Οι ακμές συνεπώς οριοθετούν ένα αντικείμενο διακρίνοντας συγκεκριμένα χαρακτηριστικά (αποχρώσεις της ασπρόμαυρης κλίμακας, χρώμα, ή άλλη ιδιότητα) σε αντιπαραβολή με το background. Στο παραπάνω πλαίσιο η παρουσία των ακμών θεωρείται υπεύθυνη για μεταβολές στην ένταση (intensity) μιας εικόνας και για τον λόγο αυτό, ο εντοπισμός τους σε μια εικόνα ισοδυναμεί με τον εντοπισμό, την ανίχνευση και μέτρηση μεταβολών της έντασης σε γειτονιές της κατά τη σάρωσή της.

Σε μια πρώτη ανάγνωση της παραπάνω διαπίστωσης η εύρεση των μεταβολών της έντασης σε μια εικόνα θα οδηγούσε σε ανίχνευση των ακμών της. Συγκεκριμένα το διάνυσμα κλίσης της έντασης μιας εικόνας δείχνει στην κατεύθυνση της πιο απότομης αύξησης (steepest ascent) των επιπέδων έντασης τών αποχρώσεων του γκρι στην εικόνα ενώ η κατεύθυνσή του είναι κάθετη στα τοπικά περιγράμματα (contours) [26]. Συνεπώς η κλίση στην ακμή ενός φωτεινού αντικειμένου σε ένα σκοτεινό φόντο, έχει κατεύθυνση προς το αντικείμενο. Αναζητώντας συνεπώς τα σημεία μηδενισμού της κλίσης μπορούμε να εντοπίζουμε πιθανά σύνορα μεταξύ ακμών και φόντου. Η παραγώγιση όμως εικόνων ενισχύει το θόρυβο στις υψηλές συχνότητες όπου η αναλογία θορύβου/σήματος είναι εξορισμού υψηλότερη κάνοντας την εξομάλυνση του σήματος μέσω εφαρμογής βαθυπερατού φίλτρου απαραίτητη πριν την παραγώγισή του.

Το πρόβλημα της ανίχνευσης ακμών μπορεί να χωριστεί σε τρία διαφορετικά υποπροβλήματα:

1. Εξομάλυνση (Smoothing): Οι εντάσεις της εικόνας εξομαλύνονται μέσω εφαρμογής φίλτρου ή προσεγγίζονται τοπικά από κατάλληλες συναρτήσεις. Με την εξομάλυνση επιτυγχάνεται η απομάκρυνση του θορύβου και των ψευδών ακμών δηλαδή ακμών που μπορεί να έχουν σχηματιστεί λόγω κβάντισης. Το βέλτιστο φίλτρο εξομάλυνσης που ικανοποιεί τις απαιτήσεις της βιολογικής όρασης (ομαλό και οριοθετημένο στο πεδίο του χώρου, ομαλό και ζωνοπερατό στο πεδίο της συχνότητας) είναι η συνάρτηση Gauss.

2. Ενίσχυση (Enhancement): Οι εξομαλυμένες εντάσεις μιας εικόνας παραγωγίζονται μια φορά για τον εντοπισμό των τοπικών ακρότατων ή δύο φορές για την ανίχνευση των σημείων μηδενισμού (zero crossing), αντιστοιχία που έχει χαρακτήρα ισοδυναμίας. Πράγματι στις κορυφές και τις κοιλάδες τις πρώτης παραγώγου του σήματος εισόδου αντιστοιχούν σημεία μηδενισμού της δεύτερης παραγώγου του σήματος εισόδου. Η διαφόριση κάνει πιο έντονες τις

ακμές και διευκολύνει την αναγνωρισιμότητα σε πρότυπα της εικόνας στις θέσεις των ακμών (ακρότατα ή σημεία μηδενισμού (zero-crossings)).

3. Κριτήριο Απόφασης: Εάν I(x,y) η αρχική ένταση της εικόνας, οι δυαδικές της ακμές ορίζονται ως το πεδίο ορισμού των εικονοστοιχείων (pixels) που το προηγούμενο στάδιο της ενίσχυσης διακρίνει ως ακρότατα ή σημεία μηδενισμού της πρώτης ή δεύτερης παραγώγου της έντασης της εικόνας αντίστοιχα.

Η μαθηματική διατύπωση των παραπάνω προτάσεων είναι πως για δεδομένη δισδιάστατη εικόνα, $f(x, y)$, υπολογίζεται η Λαπλασιανή που συνιστά ισοτροπικό -συμμετρικό ως προς τη διεύθυνση- διαφορικό τελεστή ως το μέτρο της δεύτερης χωρικής παραγώγου μιας εικόνας:
$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1) - 4f(x, y)$
ενώ η ανίχνευση ακμών μπορεί να γίνει στα σημεία μηδενισμού της (zero-crossings). Στην πραγματικότητα, η Λαπλασιανή τονίζει τις ασυνέχειες της έντασης σε μια εικόνα ενώ θολώνει τις περιοχές της εικόνας στις οποίες τα επίπεδα της έντασης μεταβάλλονται αργά [23].

Στη γραμμική ανίχνευση ακμών τα δύο πρώτα βήματα εξομάλυνσης και της παραγώγισης μπορούν να αντικατασταθούν από την πράξη της γραμμικής συνέλιξης μεταξύ της εικόνας f και μιας συνάρτησης πυρήνα Kernel h τύπου Gauss. Εναλλακτικά συνεπώς, η ανίχνευση των μεταβολών της έντασης μπορεί να γίνει υπολογίζοντας τα σημεία μηδενισμού μιας συνάρτησης $\nabla^2(G_\sigma) * I(x, y)$ για την υπό μελέτη εικόνα $I(x, y)$, όπου $G(x, y)$ δισδιάστατη γκαουσιανή κατανομή $G(x, y) = (2\pi\sigma^2)^{-1} e^{-(x^2+y^2)/2\sigma^2}$ και ∇^2 η λαπλασιανή και ονομάζεται Λαπλασιανή της Γκαουσιανής (LoG): $\nabla^2 G(x, y) = (2\pi\sigma^4)^{-1} e^{-(x^2+y^2)/2\sigma^2} \frac{(x^2+y^2)}{\sigma^2-2}$. Συνεπώς εφαρμόζουμε τον Λαπλασιανό τελεστή σε μια εξομαλυμένη -μέσω εν φίλτρου εξομάλυνσης Gauss- έκδοση της αρχικής εικόνας και σε περιοχές σταθερής έντασης το gradient της συνέλιξης είναι 0, ενώ στη γειτονιά της εικόνας που η ένταση μεταβάλλεται, η LoG θα είναι θετική στη σκοτεινή πλευρά και αρνητική στην φωτεινή. Επιπλέον φίλτρα LoG με μεγάλο σ αναμένεται να ανιχνεύουν πιο ευρείες ακμές ενώ φίλτρα με μικρό σ αναμένεται να εστιάσουν σε μεγαλύτερες λεπτομέρειες. Έτσι αν I(x,y) είναι η αρχική εικόνα έντασης, οι δυαδικές ακμές ορίζονται ως τα zero-crossings της συνέλιξης της με τη συνάρτηση Kernel, δηλαδή τα σημεία (x,y) τ.ω. $\nabla^2(G_\sigma) * I(x, y) = 0$.

4.3.2 Αναγνώριση γωνιών

Μια γωνία είναι το σημείο τομής δύο ή περισσότερων ακμών. Για τον λόγο αυτό οι μεταβολές της έντασης μιας εικόνας συνιστούν το βασικό πεδίο έρευνας όχι μόνο στο πρόβλημα ανίχνευσης ακμών αλλά και στην ανίχνευση γωνιών. Οι γωνίες είναι περιοχές μιας εικόνας που χαρακτηρίζονται από ισχυρές κλίσεις σε περισσότερες από μια κατευθύνσεις. Η ανίχνευση γωνιών διαισθητικά μπορεί να επιτευχθεί με τη βοήθεια ενός κυλιόμενου παραθύρου που σαρώνει την εικόνα συγκρίνοντας τη φωτεινότητα γειτονικών εικονοστοιχείων. Η ανίχνευση έντονων ασυνεχειών στην ένταση της εικόνας καθώς το παράθυρο κυλίεται προς οποιαδήποτε κατεύθυνση συνιστά ένδειξη γωνιότητας και κριτήριο για την απόφαση του αλγορίθμου να χαρακτηρίσει το pixel ως γωνία ενός αντικειμένου. Διατυπωμένο διαφορετικά, αν ένα pixel βρίσκεται εντός ενός αντικειμένου, η γειτονιά του συμπίπτει με τη γειτονιά των γειτονικών pixel και η ένταση είναι σταθερή. Εάν ένα pixel βρίσκεται πάνω σε μια ακμή ενός αντικειμένου, η γειτονιά του διαφέρει από αυτήν των γειτονικών του pixel στην κάθετη διεύθυνση. Τέλος ένα pixel που βρίσκεται πάνω σε μια γωνία έχει γειτονιά διαφορετικής έντασης από τη γειτονιά όλων των γειτόνων του κατά μήκος όλων των κατευθύνσεων. Ο αλγόριθμος ανίχνευσης γωνιών των Harris-Stephens διατυπώνει μαθηματικά την παραπάνω διάκριση στις τρεις κατηγορίες σημείων ενδιαφέροντος εντός μιας εικόνας.

Συγκεκριμένα, ο αλγόριθμος ορίζει ένα σταθμισμένο άθροισμα των υψωμένων στο τετράγωνο διαφορών έντασης της εικόνας που μαθηματικά παίρνει τη μορφή:

$$C(x,y) = \begin{bmatrix} x & y \end{bmatrix} M \begin{bmatrix} x \\ y \end{bmatrix}$$

όπου

$$M = \sum_n \sum_t w(s,t) A$$

και

$$A = \begin{bmatrix} f_x^2 & f_y^2 \end{bmatrix} M \begin{bmatrix} f_x^2 \\ f_y^2 \end{bmatrix}$$

Τα ιδιοδιανύσματα του πίνακα M επειδή είναι συμμετρικός και πραγματικός δείχνουν στην κατεύθυνση μέγιστης διασποράς στα δεδομένα και οι αντίστοιχες ιδιοτιμές είναι ανάλογες του μέτρου της διασποράς στην κατεύθυνση των ιδιοδιανυσμάτων [23]. Το κριτήριο απόφασης του αλγορίθμου βασίζεται στις ιδιοτιμές του πίνακα. Αν και οι δύο ιδιοτιμές παίρνουν μικρές τιμές τότε βρισκόμαστε σε γειτονιά σταθερής έντασης ενώ μια μικρή και μια μεγάλη ιδιοτιμή συνεπάγεται την παρουσία οριζόντιας ή κατακόρυφης ακμής. Τέλος δυο μεγάλες ιδιοτιμές υποδηλώνουν την πιθανή παρουσία κορυφής (ή απομονωμένων φωτεινών σημείων)[23].

4.3.3 Συμπεράσματα - Συζήτηση

Στη θεματική ενότητα της ψηφιακής επεξεργασίας διαπιστώσαμε πως παραδοσιακές τεχνικές ανάλυσης σήματος οδηγούν στην ανάκτηση μιας πληροφορίας που είναι κρυμμένη σε αυτά. Ο Gabor μετασχηματισμός μπορεί να μεταφέρει ένα σήμα από το πεδίο του χρόνου (time-domain) στο πεδίο της συχνότητας (frequency-domain) αναδεικνύοντας τη συχνοτική διάσταση που εμπεριέχεται στις κυματομορφές των δεδομένων. Από την άλλη πλευρά οι τεχνικές αναγνώρισης ακμών και γωνιών σε 2Δ σήματα περιγράφουν κλασικές μεθόδους εξαγωγής χαρακτηριστικών από εικόνες. Στη θεωρητική ενότητα της βαθιάς μάθησης θα προσαρμόσουμε τις παραπάνω τεχνικές εξαγωγής χαρακτηριστικών σε πολυστρωματικές αρχιτεκτονικές εκμεταλλευόμενοι την υπολογιστική ισχύ των σύγχρονων μηχανών για να εξάγουμε για κάθε στρώμα πολλαπλά πρότυπα -μέσω συνελίξεων εικόνων και kernels- για κάθε γειτονιά εικονοστοιχείων εικόνων που έχουν προκύψει από Gabor μετασχηματισμούς του ηχητικού σήματος.

Κεφάλαιο 5

Μηχανική μάθηση

5.1 Εισαγωγή

Η μηχανική μάθηση είναι μια οικογένεια τεχνικών που επιτελούν μια διαδικασία σταδιακής εξαγωγής ενδιάμεσων αφαιρετικών αναπαραστάσεων ενός συνόλου δεδομένων. Η διαδικασία αυτή οδηγεί στη γνώση μιας κατανομής από την οποία έχουν προέλθει τα δεδομένα του χώρου εισόδου και μπορεί να λύσει μια σειρά από προβλήματα: Προβλήματα παλινδρόμησης (regression estimation), αναγνώρισης προτύπων (pattern recognition), εκτίμηση πυκνότητας (density estimation). Η μάθηση μπορεί να είναι επιβλεπόμενη οπότε υφίσταται ένας εξωτερικός ΄δάσκαλος΄ που επιβλέπει την ορθότητα των αποτελεσμάτων της ώστε επαγωγικά να βελτιώσει την ικανότητά της να γενικεύει ή μη επιβλεπόμενη οπότε η μηχανή καλείται να ανακαλύψει ΄κρυμμένες΄ συσχετίσεις, ομαδοποιήσεις, πρότυπα, σε επαγόμενο από μη επισημειωμένα δεδομένα χώρο χαρακτηριστικών των οποίων η γνώση δεν μπορεί να ελεγχθεί πριν ή μετά τη διαδικασία.

5.2 Στατιστική θεωρία μάθησης

Η στατιστική θεωρία μάθησης μελετά τις μαθηματικές ιδιότητες των μηχανών μάθησης οι οποίες ανάγονται σε μεγάλο βαθμό στις ιδιότητες των κλάσεων συναρτήσεων που μια μηχανή μπορεί να μάθει.

Ας θεωρήσουμε τις συναρτήσεις δεικτών (indicator functions) που ορίζονται ως οι συναρτήσεις που διαχωρίζουν ένα σύνολο διανυσμάτων σε δύο υποσύνολα: Το υποσύνολο διανυσμάτων που ταξινομείται στην κλάση 0 και το υποσύνολο που διαχωρίζεται στην κλάση 1. Οριοθετώντας το πρόβλημα της θεωρίας μάθησης, το ζητούμενο είναι κάτω υπό ποιες συνθήκες, μπορεί να κατασκευαστεί από ένα διακριτό αριθμό ανεξάρτητων και προερχόμενων από την ίδια άγνωστη από κοινού συνάρτηση κατανομή $P(y, x) = P(y/x)P(x)$ παρατηρήσεων $(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$, μια μηχανή μάθησης με ικανότητα γενίκευσης. Συγκεκριμένα αναζητούμε μια συνάρτηση $f_l(x)$ προερχόμενη από ένα σύνολο συναρτήσεων δεικτών $f(x) : X- > 0, 1$, ικανή να προσεγγίζει βέλτιστα ένα συναρτησιακό κόστους, τον αναμενόμενο κίνδυνο (expected risk):

$$R(a) = \int L(y - f(x))dP(x, y) \tag{5.1}$$

όπου $L(y - f(x))$ μη αρνητική συνάρτηση σφάλματος.

Η προσέγγιση του παραπάνω συναρτησιακού από ένα συναρτησιακό που κατασκευάζεται στη βάση εμπειρικών δεδομένων του συνόλου εκπαίδευσης οδήγησε στην αρχή ελαχιστοποίησης του εμπειρικού κινδύνου (empirical risk minimization). Η αρχή αυτή προσδιορίζει μια επαγωγική διαδικασία μάθησης της $R(a)$ από την εμπειρική -γιατί βασίζεται σε δεδομένα εκπαίδευσης- συνάρτηση $R^{emp}(a)$ που ελαχιστοποιεί μια συνάρτηση κόστους στο σύνολο των πειραματικών δεδομένων. Η αρχή ελαχιστοποίησης του εμπειρικού κινδύνου είναι επαρκής όταν ο δειγματόχωρος είναι μεγάλος με συνεπαγόμενο κόστος στο χρόνο εκπαίδευσης.

Το σφάλμα στη διαδικασία μάθησης δεν εξαρτάται όμως μόνο από το σφάλμα εκτίμησης αλλά και από το λάθος προσέγγισης που αναφέρεται στο σφάλμα γενίκευσης σε άγνωστο σύνολο δεδομένων. Ο V.Vapnik ανέπτυξε τη θεωρία ελαχιστοποίησης του δομικού κινδύνου (structural risk minimization) που μπορεί να προσαρμοστεί στη μάθηση προτύπων σε προβλήματα μικρής κλίμακας. Θα δούμε στη συνέχεια ότι η προσέγγιση των σφαλμάτων εκτίμησης γίνεται σε όρους μιας διάστασης που εισήγαγαν οι Vapnik και Chervovenkis και λέγεται VC dimension που εκφράζει το μέτρο της ικανότητας ή της έκφρασης ισχύος μιας οικογένειας δυαδικών συναρτήσεων κατηγοριοποίησης που υλοποιεί μια μηχανή μάθησης.

5.3 Το Perceptron του Rosenblatt

Το απλούστερο νευρωνικό δίκτυο που σχεδιάζεται για ταξινόμηση γραμμικά διαχωρίσιμων προτύπων αποτελείται από ένα μόνο νευρώνα και λέγεται Perceptron. Η τομή που έφερε το perceptron βρίσκεται στην ικανότητά του να εκπαιδεύει με βάση ένα διάνυσμα εισόδου και εξόδου τα βάρη ενός δικτύου και να μπορεί να τα κατηγοριοποιεί ως προς το υπερεπίπεδο που αυτά προσδιορίζουν. Ο Rosenblatt που εισήγαγε το Perceptron απέδειξε ότι ένα τέτοιο μοντέλο συγκλίνει σε λύση και οριοθετεί υπερεπίπεδο απόφασης όταν τα πρότυπα που συμμετέχουν στη διαδικασία μάθησης του δικτύου εξάγονται από δυο γραμμικά διαχωρίσιμες κλάσεις. Ο αλγόριθμος που υλοποιείται για τη μάθηση και προσαρμογή των βαρών είναι ο ακόλουθος [27]:

1. Αρχικοποίηση του μετρητή εποχών ν= 1 και του διανύσματος βαρών $(b, w_1[n], w_2[n], ...w_m[n])$
2. Ενεργοποίηση νευρώνα για κάθε χρονικό βήμα του perceptron.
3. Υπολογισμός της τρέχουσας απόκρισης του perceptron από τη σχέση:
$y[n] = sgn[w^T[n]x[n]]$, όπου εφαρμόζεται η συνάρτηση προσήμου στο προσαρμοσμένο από τα βάρη διάνυσμα εισόδου.
4. Προσαρμογή του διανύσματος βάρους από τη σχέση:
$w[n + 1] = w[n] + n(d[n] - y[n])x[n]$, όπου n, παράμετρος σύγκλισης που ορίζεται από τον προγραμματιστή.

5. Αύξηση βήματος n. Έλεγχος αν το βήμα είναι μικρότερο από τον συνολικό αριθμό εποχών N. Αν ναι, ολοκλήρωση της διαδικασίας, διαφορετικά επιστροφή στο βήμα 2.

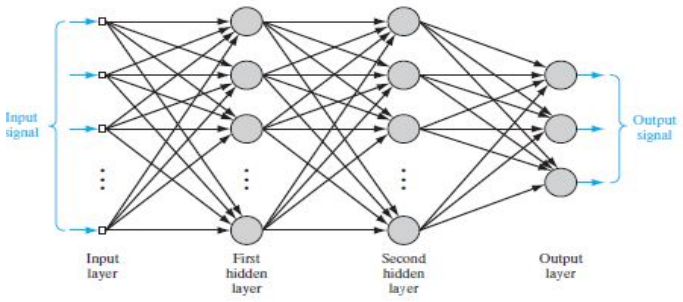
5.4 Πολυστρωματικά Perceptrons (M.L.P.)

Με βάση το γενικό θεώρημα προσέγγισης (universal approximation theory) ένα εμπρόσθιο δίκτυο με γραμμικό στρώμα εξόδου και τουλάχιστον ένα στρώμα με επαρκή αριθμό κρυμμένων νευρώνων ακολουθούμενο από συνάρτηση ενεργοποίησης μπορεί να προσεγγίσει οποιαδήποτε Borel μετρήσιμη συνάρτηση -δηλαδή κάθε συνάρτηση συνεχή σε φραγμένο και κλειστό υπόχωρο του R^N– που ορίζεται και απεικονίζει τιμές σε χώρο πεπερασμένης διάστασης χωρίς να υπάρχει κάτω όριο στο σφάλμα με το οποίο μπορεί να επιτευχθεί η προσέγγιση αυτή. Το θεώρημα αυτό δίνει τη δυνατότητα να αναπαρασταθεί μέσα από ένα μονοστρωματικό νευρωνικό δίκτυο οποιαδήποτε συνάρτηση, δεν εγγυάται όμως: α) ότι το δίκτυο θα αποκτήσει τη δυνατότητα να ΄μάθει΄ τη συνάρτηση αυτή β) ότι θα ΄γενικεύει΄ σε νέα παραδείγματά της[28].

Αυτό συμβάινει για δύο λόγους: Αφενός μεν ο αλγόριθμος μπορεί να μάθει να ταυτίζει τα αποτελέσματα προσεγγιζόμενης και προσεγγιστικής συνάρτησης υπερπροσαρμόζοντας τα βάρη της τελευταίας γενικεύοντας όχι αντιπροσωπευτικά patterns αλλά ψευδή πρότυπα του συνόλου εκπαίδευσης (training set) γενικεύοντας τελικά αλλοιωμένα (aliased) πρότυπα. Δεύτερον ο αλγόριθμος βελτιστοποίησης δεν εγγυάται ότι θα βρεθεί μια βέλτιστη τιμή των παραμέτρων για την επιθυμητή συνάρτηση ολοκληρώνοντας την εκπαίδευση σε τοπικά και όχι ολικά ακρότατα και γενικεύοντας τελικά ανεπαρκώς. Τέλος το θεώρημα δεν επισημαίνει ποιο είναι το μέγεθος του δικτύου που τελικά θα προσεγγίσει βέλτιστα τη συνάρτηση.

Στο παρακάτω σχήμα απεικονίζεται ένα fully connected πολυστρωματικό perceptron το οποίο αποτελείται από δύο κρυμμένα στρώματα και το στρώμα εξόδου.

Σχήμα 5.1: Πολυστρωματικό Νευρωνικό Δίκτυο



Η εικόνα μεταφέρεται αυτούσια από βιβλίο του S.Haykin [27]

Στο πολυστρωματικό perceptron ένα σήμα συνάρτησης διαδίδεται εμπρόσθια ενεργοποιώντας κάθε νευρώνα κάθε στρώματος που συναντάει με κατεύθυνση την έξοδο του δικτύου. Η έξοδος που προβλέπει το δίκτυο συγκρίνεται στη συνέχεια με την επισημείωση που έχει

υποδειχτεί από κάποιον supervisor και υπολογίζεται μια συνάρτηση σφάλματος. Ακολουθεί οπισθοδιάδοση του σφάλματος και αναπροσαρμογή των βαρών κάθε νευρώνα στην κατεύθυνση που προσδιορίζει η μέγιστη κατάβαση δυναμικού (steepest gradient descent) στο χώρο βαρών (weight space). Μετά την ανανέωση του διανύσματος βαρών επαναλαμβάνεται η ίδια διαδικασία μέχρι να επιτευχθεί κάποια σύγκλιση.

5.5 Μηχανές Διανυσμάτων Υποστήριξης (SVM)

Η εισαγωγή των μηχανών διανυσμάτων υποστήριξης έγινε στα πλαίσια της Στατιστικής Θεωρίας Μάθησης (Statistical Learning Theory) που ανέπτυξε ο V.Vapnik. Η στατιστική θεωρία μάθησης αναπτύσσει μεθόδους για την κατασκευή προσεγγίσεων που συγκλίνουν στην επιθυμητή συνάρτηση καθώς ο αριθμός των παρατηρήσεων αυξάνεται [29]. Η βασική ιδέα που υλοποιούν τα SVM είναι η ακόλουθη: Δεδομένου ενός συνόλου δεδομένων εκπαίδευσης η μηχανή διανυσμάτων υποστήριξης στοχεύει στην απεικόνισή τους σε ένα χώρο χαρακτηριστικών υψηλής διάστασης Z μέσω μιας μη γραμμικής απεικόνισης που επιλέγεται a priori. Στον παραγόμενο χώρο χαρακτηριστικών υψηλής διάστασης μπορεί να κατασκευαστεί ένα βέλτιστο υπερεπίπεδο διαχωρισμού των δεδομένων που συνιστά την επιφάνεια απόφασης διαχωρίζοντας τις κλάσεις στις οποίες ανήκουν τα δεδομένα εκπαίδευσης. Το κριτήριο για την κατασκευή του υπερεπιπέδου συνίσταται στη μεγιστοποίηση του περιθωρίου διαχωρισμού (separation margin) μεταξύ θετικών και αρνητικών προτύπων[30] και η επιφάνεια σχεδιάζεται με τρόπο που να απέχει όσο το δυνατόν περισσότερο από τα κοντινότερα θετικά και αρνητικά πρότυπα.

Ο αριθμός των χαρακτηριστικών που συγκροτουν τον κρυμμένο χώρο χαρακτηριστικών (Feature hidden space) καθορίζεται από τον αριθμό των διανυσμάτων υποστήριξης. Συνεπώς η θεωρία των (SVM) παρέχει μια αναλυτική προσέγγιση στον καθορισμό του βέλτιστου μεγέθους του χώρου χαρακτηριστικών, εξασφαλίζοντας βέλτιστη απόκριση σε προβλήματα ταξινόμησης [27]. Τα SVM από τη στιγμή που θα οριστούν οι υπερπαράμετροι του προβλήματος τετραγωνικού προγραμματισμού δίνουν μια ντετερμινιστική λύση συνεπώς υπερβαίνουν προβλήματα που χαρακτηρίζουν τα perceptrons όπως τα τοπικά ελάχιστα και τη διασπορά των λύσεων στο χώρο αναζήτησης ενώ ταυτόχρονα υπερτερούν στο γεγονός ότι μπορούν να παράγουν πιο σύνθετες επιφάνειες διαχωρισμού.

5.5.1 Γραμμικά Διαχωρίσιμα Δεδομένα

Έστω ότι το διάνυσμα εισόδου που αποτελείται από τα δεδομένα

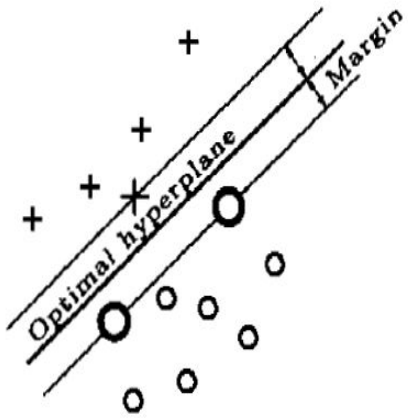
$$(x_1, y_1), (x_2, y_2), ..., (x_N, y_N), x \in R_n, y \in (-1, -1) \tag{5.2}$$

μπορεί να διαχωριστεί από υπερεπίπεδο της μορφής:

$$(wx) - b = 0 \tag{5.3}$$

Λέγεται ότι το παραπάνω σύνολο διανυσμάτων διαχωρίζεται από βέλτιστο υπερεπίπεδο αν διαχωρίζεται χωρίς σφάλμα και η απόσταση ανάμεσα στο κοντινότερο διάνυσμα και το υπερεπίπεδο είναι μέγιστη.

Σχήμα 5.2: Βέλτιστο επίπεδο διαχωρισμού και διανύσματα υποστήριξης



Εικόνα από το βιβλίο του V.Vapnik, The Nature of Statistical Learning Theory [30]

Η μαθηματική περιγραφή του υπερεπιπέδου απόφασης δίνεται από τη σχέση:

$$wx_i - b \geq 1, if y_i = 1 \tag{5.4}$$

$$wx_i - b \leq -1, if y_i = -1 \tag{5.5}$$

Τα ιδιαίτερα εκείνα σημεία (x_i, y_i) για τα οποία στις παραπάνω σχέσεις ικανοποιείται η σχέση της ισότητας ονομάζονται διανύσματα υποστήριξης (support vectors) και γεωμετρικά αντιστοιχούν στα εγγύτερα δυνατά σημεία του συνόλου εκπαίδευσης ως προς το βέλτιστο υπερεπίπεδο ενώ χαρακτηρίζονται από μέγιστο βαθμό δυσκολίας διαχωρισμού [27].

Αποδεικνύεται ότι η μεγιστοποίηση του περιθωρίου διαχωρισμού σε ένα πρόβλημα διαχωρισμού δυαδικών κλάσεων ισοδυναμεί με την ελαχιστοποίηση της Ευκλείδιας νόρμας του διανύσματος ω, $\Phi(w) = \frac{1}{2}||w||^2$ ενώ παράλληλα ικανοποιεί τις παραπάνω σχέσεις που σε συμπαγή μορφή γράφονται: $y_i[(wx_i) - b] \leq 1, i = 1, ..., l$. Η λύση του προβλήματος βελτιστοποίησης προκύπτει από την επίλυση της Lagrangian συνάρτησης:
$L(w, b, a) = \frac{1}{2}||w||^2 - \sum_{i=1}^{l} \lambda_i[(wx_i) - b]y_i - 1$
Υπολογίζοντας τις μερικές παραγώγους της Lagrangian συνάρτησης ως προς (b, w) καταλήγουμε στις παρακάτω ιδιότητες του βέλτιστου υπερεπιπέδου:
1. Οι συντελεστές για το βέλτιστο υπερεπίπεδο θα έπρεπε να ικανοποιούν τους περιορισμούς:
$\sum_{i=1}^{l} \lambda_i^0 y_i = 0, , \lambda_i^0 \geq 0, i = 1, ..., l$
2. Οι παράμετροι του βέλτιστου υπερεπιπέδου είναι γραμμικοί συνδυασμοί των διανυσμάτων του συνόλου εκπαίδευσης:
$w_0 = \sum_{i=1}^{l} y_i \lambda_i^0 x_i, \lambda_i^0 \geq 0, i = 1, ..., l$
3. Η λύση πρέπει να ικανοποιεί την ισότητα για τις σχέσεις (5.4, 5.5)

Προκύπτει από τις παραπάνω ιδιότητες ότι μόνο τα διανύσματα υποστήριξης μπορεί έχουν μη μηδενικούς συντελεστές στο ανάπτυγμα του w_0.

5.5.2 Η μηχανή διανυσμάτων υποστήριξης ως μηχανή Kernel

Στην παράγραφο αυτή θα δείξουμε πως για την κατασκευή του βέλτιστου υπερεπιπέδου διαχωρισμού στον χώρο χαρακτηριστικών δεν είναι απαραίτητη η ανάπτυξη του χώρου χαρακτηριστικών σε αναλυτική μορφή. Χρειάζεται μόνο ο υπολογισμός των εσωτερικών γινομένων μεταξύ των διανυσμάτων υποστήριξης και των διανυσμάτων στο χώρο χαρακτηριστικών. Συγκεκριμένα στο εισαγωγικό σημείωμα στα SVM αναφέραμε τις δύο μαθηματικές λειτουργίες που επιτελούνται από μια μηχανή διανυσμάτων υποστήριξης.
1. Μη γραμμική απεικόνιση ενός διανύσματος εισόδου σε έναν υψηλής διάστασης χώρο χαρακτηριστικών που είναι κρυμμένος τόσο από το χώρο εισόδου όσο και το χώρο εξόδου.
2. Κατασκευή ενός βέλτιστου υπερεπιπέδου για το διαχωρισμό των χαρακτηριστικών που έχουν εξαχθεί από το διάνυσμα εισόδου στο βήμα 1.

Έστω ότι το σύνολο των μη γραμμικών συναρτήσεων που επιτελούν το βήμα 1 είναι οι $\Phi_j(x), j = 1, .., \infty$ ¨ή εναλλακτικά σε διανυσματική μορφή $[\phi_1(x), \phi_2(x), ...]^T$, όπου x ο χώρος εισόδου. Τότε η κατασκευή ενός υπερεπιπέδου διαχωρισμού μπορεί να αναπαρασταθεί από τη σχέση: $w^T\phi(x)$. Το διάνυσμα βάρους μπορεί να αναπαρασταθεί από τη σχέση: $w = \sum_{i=1}^{N_s} a_i d_i \phi(x_i)$, όπου N_s ο αριθμός των διανυσμάτων υποστήριξης. Αντικαθιστώντας στη σχέση που αναπαριστά το υπερεπίπεδο προκύπτει:

$$\sum_{i=1}^{N_s} a_i d_i \phi(x_i)\phi({x_i}^T) = 0 \tag{5.6}$$

η οποία γράφεται

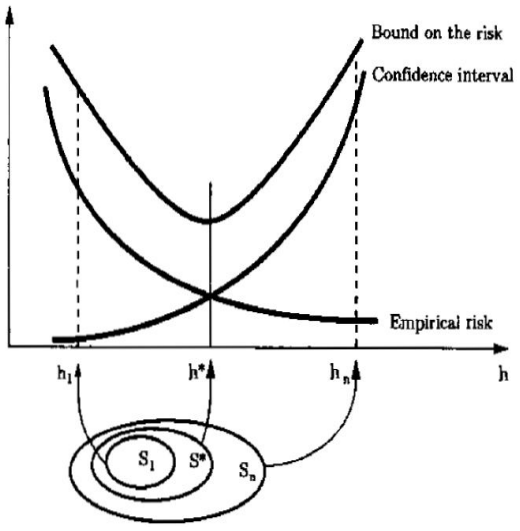
$$\sum_{i=1}^{N_s} a_i d_i k(\mathbf{x}, \mathbf{x}_i) = 0 \tag{5.7}$$

Η συνάρτηση k ονομάζεται εσωτερικό γινόμενο kernel και υπολογίζει το εσωτερικό γινόμενο των απεικονίσεων που παράγονται στο χώρο χαρακτηριστικών υπό την συνάρτηση φ δύο σημείων στο χώρο χαρακτηριστικών. Οι μη γραμμικές μηχανές μπορεί να είναι γραμμικές, πολυωνυμικές, ακτινικές συναρτήσεις βάσης RBF. Σε μια πολυωνυμική μηχανή μάθησης ο πυρήνας Kernel είναι η συνάρτηση $K(x, x_i) = (xx_i + 1)^d$ όπου d επιλέγεται apriori από τον χρήστη ενώ μια μηχανή με ακτινική συνάρτηση βάσης (R.B.F) αναπαρίσταται από την $K(x, x_i) = exp(-\frac{|x-x_i|^2}{2\sigma^2})$ Αποδεικνύεται ότι η γνώση του εσωτερικού γινομένου kernel που αποτελείται από τα διανύσματα του χώρου εισόδου και τα διανύσματα υποστήριξης εξασφαλίζει την κατασκευή του βέλτιστου υπερεπιπέδου διαχωρισμού.

5.6 Vapnik-Chervovenkis (VC) διάσταση και ικανότητα γενίκευσης μηχανών μάθησης

Οι Vapnik, Chervovenkis όρισαν τη VC διάσταση ενός συνόλου συναρτήσεων δεικτών h ως τον μέγιστο αριθμό h διανυσμάτων που μπορεί να διαχωριστούν με 2^h δυνατούς τρόπους χρησιμοποιώντας συναρτήσεις αυτού του συνόλου ενώ δεν υπάρχουν $h+1$ διανύσματα διαχωρίσιμα με 2^{h+1} τρόπους. Συνεπώς η διάσταση VC είναι άπειρη όταν δεν υπάρχει μέγιστο h τέτοιο ώστε να υπάρχει σύνολο h σημείων που η κλάση συναρτήσεων να μπορεί να διαχωρίσει [31]. Η ικανότητα γενίκευσης τόσο των νευρωνικών δικτύων όσο και των SVM εξαρτάται από τη VC διάσταση[32]. Συγκεκριμένα προκειμένου να αντιμετωπιστεί το πρόβλημα της υπερπροσαρμογής (overfitting) στα νευρωνικά δίκτυα πρέπει η VC διάσταση να είναι μικρή που συνεπάγεται όμως δυσκολίες στην προσαρμογή των δεδομένων εκπαίδευσης αφού ο πρώτος όρος του συναρτησοειδούς αναπαριστά τον εμπειρικό κίνδυνο. Η στατιστική θεωρία μάθησης έδειξε τη σημαντικότητα να περιοριστεί το σύνολο συναρτήσεων από το οποίο η συνάρτηση f μπορεί να επιλεχθεί σε αυτήν που είναι κατάλληλη για την ποσότητα των διαθέσιμων δεδομένων εκπαίδευσης. Η ελαχιστοποίηση του ορίου σφάλματος στο test set εξαρτάται από την ελαχιστοποίηση τόσο του εμπειρικού κινδύνου όσο και της ικανότητας της κατηγορίας της συνάρτησης, αρχή που οδήγησε στην ελαχιστοποίηση δομικού κινδύνου (structural risk minimization).

Σχήμα 5.3: Εμπειρικός κίνδυνος και διάστημα εμπιστοσύνης



Η εικόνα αναπαράγεται από βιβλίο του V.Vapnik [30]

5.7 Συμπέρασματα - συζήτηση για την ικανότητα γενίκευσης μηχανών μάθησης

Τα Τεχνητά Νευρωνικά Δίκτυα αποτελούν απλουστευμένη προσομοίωση της λειτουργίας του ανθρώπινου εγκεφάλου ενώ οι μηχανές διανυσμάτων υποστήριξης αντλούν την έμπνευσή τους από τη Στατιστική Θεωρία Μάθησης η οποία θεμελιώθηκε θεωρητικά τη δεκαετία του 1960. Η καινοτομία του Vapnik έγκειται στην εισαγωγή της έννοιας του δομικού κινδύνου που αναδεικνύει κάποιου είδους ανταγωνιστική σχέση ανάμεσα στην ποιότητα προσέγγισης ενός συνόλου δεδομένων και τη συνθετότητα της συνάρτησης προσέγγισής τους. Προσδιορίζει ένα άνω όριο που φράσσει την ικανότητα γενίκευσης ενός δικτύου και το ορίζει ως τη διαφορά αναμενόμενης τιμής σφάλματος και εμπειρικού ρίσκου: $R[f] - R_{emp} \leq \Phi(\frac{l}{2\pi})$ όπου l ο αριθμός των δεδομένων εκπαίδευσης και h η διάσταση VC. Το όριο αυτό ονομάζεται διάστημα εμπιστοσύνης και ο σχεδιασμός μιας πολύπλοκης μηχανής το καθιστά μεγάλο, συνεπώς η μείωση του εμπειρικού σφάλματος δεν θα αποτρέψει την εμφάνιση υπερπροσαρμογής (overfitting). Αντίστροφα η κατασκευή μιας μηχανής με μικρό διάστημα εμπιστοσύνης αποτρέπει τη δυνατότητα μείωσης του εμπειρικού κινδύνου. Πάνω στην έννοια του διαστήματος εμπιστοσύνης μπορεί να διαφανεί η βασική διαφοροποίηση μεταξύ Νευρωνικών Δικτύων και Μηχανών Διανυσμάτων Υποστήριξης που έγκειται στο ότι στα μεν πρώτα επιχειρείται ελαχιστοποίηση μιας συνάρτησης εμπειρικού κινδύνου με σταθερό το διάστημα εμπιστοσύνης ενώ στα SVM ελαχιστοποιείται το διάστημα εμπιστοσύνης διατηρώντας σταθερό τον εμπειρικό κίνδυνο [30].

5.8 Εξαγωγείς χαρακτηριστικών

Κατά την επίλυση ενός προβλήματος μηχανικής μάθησης εξάγεται σε κάποιο ενδιάμεσο στάδιο κάποιο διάνυσμα χαρακτηριστικών (feature extraction) από το διάνυσμα των δεδομένων εισόδου. Η εξαγωγή χαρακτηριστικών συνίσταται σε μια διαδικασία εξαγωγής διανυσμάτων από πρότυπα που εμπεριέχουν σημαντικές πληροφορίες για αυτά. Από το δυναμοσύνολο των χώρων χαρακτηριστικών (feature space) που μπορεί να εξάγονται από ένα χώρο εισόδου (input space) θα επιθυμούσαμε ιδανικά να υλοποιήσουμε μια τεχνική συμπίεσης που να μας οδηγεί σε εκείνο το χώρο του οποίου οι διαστάσεις συμπυκνώνουν τη μέγιστη δυνατή ποσότητα πληροφορίας του χώρου εισόδου.

5.8.1 Ανάλυση Κύριων Συνιστωσών (P.C.A.)

Η τεχνική P.C.A. είναι από τις πιο συνηθισμένες τεχνικές εξαγωγής χαρακτηριστικών και ανάγεται σε έναν ορθογώνιο μετασχηματισμό του συστήματος συντεταγμένων στον οποίο προβάλλουμε τα δεδομένα. Οι νέοι άξονες ονομάζονται κύριοι άξονες των δεδομένων και αποδεικνύεται ότι ένα υποσύνολο τους επαρκεί για να περιγράψει το μεγαλύτερο ποσοστό μεταβλητότητας στα δεδομένα.

Η μέθοδος PCA είναι μια μη επιβλεπόμενη τεχνική, δηλαδή δεν γίνεται με γνώση των κατηγοριών ή τάξεων στις οποίες ανήκει το κάθε δεδομένο, και πετυχαίνει ένα γραμμικό μετασχηματισμό ενός χώρου εισόδου υψηλής διάστασης σε ένα χώρο χαμηλότερης διάστασης του οποίου τα στοιχεία είναι ασυσχέτιστα. Η πρώτη κύρια συνιστώσα PCA-1 αντιπροσωπεύει τη μέγιστη διεύθυνση στη διασπορά των δεδομένων και κάθε παρατήρηση μπορεί να προβληθεί πάνω στη γραμμή αυτή παίρνοντας μια τιμή συντεταγμένων κατά μήκος της PC1. Ο δεύτερος άξονας PC2 προσανατολίζεται με τρόπο που να αντιπροσωπεύει την αμέσως επόμενη σημαντικότερη πηγή διασποράς στα δεδομένα με τον περιορισμό να είναι ορθογώνιος στον PC1 ενώ η ίδια διαδικασία ακολουθείται και για τους επόμενους άξονες PC3, PC4,..., PCn. Με τον τρόπο αυτό προβάλλεται γεωμετρικά ένα σύνολο δεδομένων σε ένα χώρο μικρότερης διάστασης από το χώρο δεδομένων, όπου οι νέες μεταβλητές ονομάζονται κύριες συνιστώσες και συμπυκνώνουν με φθίνοντα τρόπο τη μεταβλητότητα στα δεδομένα που αντιπροσωπεύει το μέτρο του πληροφοριακού τους περιεχομένου.

Για την υλοποίηση του PCA υπολογίζεται ο πίνακας συνδιασποράς R των δεδομένων εισόδου x που ανήκουν σε ένα χώρο διάστασης N. Η υπόθεση για το διάνυσμα εισόδου είναι ότι η μέση τιμή είναι μηδέν, αν κάτι τέτοιο δεν ισχύει αφαιρούμε τη μέση τιμή από το διάνυσμα. Από το φασματικό θεώρημα επειδή ο πίνακας είναι συμμετρικός μπορεί να αποσυντεθεί στη μορφή: $R = ULU^T$, όπου U είναι ορθοκανονικός πίνακας. Προκύπτει συνεπώς $RU = UL$ ή ισοδύναμα $Ru_i = \lambda_i u_i$. Από τη σχέση αυτή υπολογίζονται οι ιδιοτιμές και τα ιδιοδιανύσματά του πίνακα συνδιασποράς και διατάσσονται σε φθίνουσα σειρά ισοδύναμα $\lambda_1, \lambda_2, ..., \lambda_m$. Οι ιδιοτιμές λέγονται κύριες συνιστώσες. Μια εναλλακτική ισοδύναμη θεώρηση της $R = ULU^T$

είναι η $L = U^T RU$. Παρατηρούμε επίσης ότι η προβολή του διανύσματος εισόδου x σε μοναδιαία διανύσματα u ίδιας διάστασης είναι $A = X^T u = u^T X$ και η διασπορά του υπολογίζεται ως: $\sigma^2 = E[A^2] = E[(u^T X)(X^T u)] = u^T E[XX^T]u = u^T Ru$. Από τις παραπάνω σχέσεις προκύπτει ότι η διασπορά του διανύσματος προβολής του διανύσματος εισόδου στα ιδιοδιανύσματα του πίνακα αυτοσυσχέτισης ισούται με την αντίστοιχη ιδιοτιμή λ_i και η συνολική διασπορά $\sum_{i=1}^{N} E[x^2[i]]$ ισούται -υπό τη συνθήκη μηδενικής μέσης τιμής- με το άθροισμα των ιδιοτιμών $\sum_{i=1}^{N} \lambda_i$. Άρα η επιλογή συνιστωσών που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές μεγιστοποιεί τη διασπορά.

Παράλληλα αναπτύσσοντας το διάνυσμα x ως γινόμενο των διανυσμάτων βάσης του e_i και του διανύσματος προβολής σε αυτά προκύπτει $x = \sum_{i=0}^{N-1} a_i e_i$ όπου $a_i = \sum_{i=0}^{N-1} e_i x_i$. Προσεγγίζουμε την έκφραση του διανύσματος εισόδου διατηρώντας μόνο τις μεγαλύτερες m ιδιοτιμές $\hat{x} = \sum_{i=0}^{m-1} a(i)e(i)$ και ελαχιστοποιούμε την ενέργεια προβολής των νέων προσεγγιστικών διανυσμάτων χαμηλής διάστασης ως προς τα διανύσματα εισόδου. Δεδομένου ότι η τετραγωνική ενέργεια του σφάλματος προβολής είναι όπως δείξαμε ίση με το άθροισμα των ιδιοτιμών $E[|x - \hat{x}|^2] = \sum_{i=m}^{N-1} \lambda_i$ όπου λ_i οι ιδιοτιμές του πίνακα συνδιασποράς R_x προκύπτει ισοδυναμία στην ελαχιστοποίηση του τετραγωνικού σφάλματος ανακατασκευής και τη μεγιστοποίηση της διασποράς της προβολής του διανύσματος στον υπόχωρο χαρακτηριστικών.

Συγκεφαλαιώνοντας για τον υπολογισμό των κύριων συνιστωσών υπολογίζουμε τις ιδιοτιμές και τα ιδιοδιανύσματα του πίνακα αυτοσυσχέτισης του διανύσματος εισόδου και προβάλλουμε τα δεδομένα στον υπόχωρο που προσδιορίζεται από τα ιδιοδιανύσματα που ανήκουν στις σημαντικότερες ιδιοτιμές [27]. Οι κύριες συνιστώσες είναι μεταξύ τους ασυσχέτιστες ενώ οι διασπορές τους διατάσσονται σε φθίνουσα σειρά καθώς είναι ίσες με τις ιδιοτιμές του πίνακα αυτοσυσχέτισης. Αν οι ιδιοτιμές του πίνακα αυτοσυσχέτισης φθίνουν με υψηλό ρυθμό τότε η προσέγγιση του χώρου εισόδου μπορεί να γίνει με λίγες κύριες συνιστώσες και αντίστροφα.

5.8.2 Οπτικοποίηση διανύσματος χαρακτηριστικών στο Ευκλείδιο επίπεδο: t-distributed Stochastic Neighborhood Embedding, (t-SNE)

Η τεχνική t-SNE είναι μια τεχνική μη γραμμικής μείωσης διάστασης που επιτρέπει οπτικοποίηση δεδομένων υψηλής διάστασης στον $\mathbb{R}^2$ ή $\mathbb{R}^3$. Συγκεκριμένα, η τεχνική αυτή προτυποποιεί αντικείμενα υψηλής διάστασης σε δισδιάστατα σημεία τέτοια ώστε παρόμοια αντικείμενα στο χώρο εισόδου να αναπαρίστανται σε κοντινές αποστάσεις στον ενσωματωμένο χώρο (embedded space) ενώ ανόμοια αντικείμενα να αναπαρίστανται με μεγάλη πιθανότητα σε απομακρυσμένα 2Δ-σημεία. Η τεχνική αυτή μας επιτρέπει να προσομοιώσουμε τα χαρακτηριστικά με τη μορφή συστάδων (clusters) ή κλάσεων (classes).

Δεδομένου ενός συνόλου δεδομένων που ανήκει σε d-διάστατο χώρο $X = (X_1, X_2, ..., X_n) \subseteq R^d$ η τεχνική t-SNE υπολογίζει έναν ενσωματωμένο σε αυτόν s-διάστατο χώρο σημείων χαμηλότερης διάστασης $Y = (Y_1, Y_2, ..., Y_n) \subseteq R^s$ κατασκευάζοντας μέτρα ομοιότητας στο χώρο εισόδου και στον ενσωματωμένο χώρο ελαχιστοποιώντας κάποια μετρική της μεταξύ τους απόκλισης.
Συγκεκριμένα ένα απο κοινού μέτρο πιθανότητας $p_{i,j}$ που αναπαριστά ομοιότητες [32] στο χώρο εισόδου υπολογίζεται από τη σχέση: $p_{ij} = \frac{p_{i/j}+p_{j/i}}{2n}$ ενώ $p_{i/j}$ υπολογίζεται από την εξίσωση:

$$p_{i/j} = \frac{e^{-\frac{|xi-xj|^2}{2\sigma i^2}}}{\Sigma_{k\neq i} e^{-\frac{|xi-xk|^2}{2{\sigma_i}^2}}} \tag{5.8}$$

Επίσης το μέτρο της ομοιότητας δεδομένων y_i, y_j που ανήκουν στον ενσωματωμένο χώρο χαμηλής διάστασης δίνεται από τη σχέση:

$$q_{i,j} = \frac{(1+|yi-yj|^2)^{-1}}{\Sigma_{k\neq l}(1+|yk-yl|^2)^{-1}} \tag{5.9}$$

Η τεχνική t-SNE βασίζεται στην ελαχιστοποίηση της μετρικής Kullback-Leibler που μετρά την απόκλιση μεταξύ της από κοινού κατανομής P του χώρου εισόδου και της από κοινού κατανομής Q των σημείων στον ενσωματωμένο χώρο:

$$D_{K,L}(P,Q) = \Sigma_{i\neq j} p_{i,j} log(\frac{p_{j/i}}{q_{j/i}}) \tag{5.10}$$

Τα διανύσματα $(Y = Y_1, Y_2, ..., Y_d)$ αρχικοποιούνται τυχαία ενώ η συνάρτηση κόστους βελτιστοποιείται χρησιμοποιώντας τον αλγόριθμο gradient descent.

5.9 Βαθιά μάθηση

Η βαθιά μάθηση είναι υποπεδίο της μηχανικής μάθησης και γνώρισε τα τελευταία χρόνια ραγδαία ανάπτυξη τόσο λόγω της προόδου που έχει επιτευχθεί στο θεωρητικό υπόβαθρο πάνω στο οποίο αναπτύχθηκαν οι αλγόριθμοι αλλά και ιδιαιτέρως λόγω της καινοτομίας σε υλικό και λογισμικό που αναβάθμισε τη δυνατότητα αποθήκευσης πληροφορίας και ενίσχυσε την υπολογιστική ικανότητα των μηχανημάτων. Τομή στη ραγδαία εξέλιξη της βαθιάς μάθησης τα τελευταία χρόνια αποτέλεσε η μεγάλη πρόοδος στους επιταχυντές για την επεξεργασία γραφικών GPUs και η ανάπτυξη αλγορίθμων που μπορούν να 'τρέξουν' σε τέτοια περιβάλλοντα. Οι επεξεργαστές γραφικών αξιοποιούν τον παράλληλο προγραμματισμό, χαρακτηρίζονται από ικανότητα ανάπτυξης μεγάλης υπολογιστικής ισχύος και μπορούν να διαχειρίζονται μεγάλους όγκους δεδομένων. Η ευρεία διάδοσή της βαθιάς μάθησης οφείλεται σε open source βιβλιοθήκες μηχανικής μάθησης όπως οι PyTorch και Tensorflow που προσφέρουν δυνατότητες αξιοποίησης έτοιμων υπολογιστικών components σε ένα μεγάλο εύρος προβλημάτων.

Οι τεχνικές βαθιάς μάθησης είναι μέθοδοι που επιτρέπουν σε μια μηχανή να ανακαλύπτει αυτόματα εσωτερικές αναπαραστάσεις ενός διανύσματος εισόδου και οι αρχιτεκτονικές της αποτελούνται από πολυστρωματικές στοίβες που επιτελούν συνήθως κάποια μη γραμμική απεικόνιση από ένα χώρο εισόδου σε έναν χώρο εξόδου. Η βαθιά μάθηση χρησιμοποιείται σήμερα σε όλο το φάσμα των επιστημών με μεγάλη αποτελεσματικότητα σε προβλήματα όρασης υπολογιστών όπως η αναγνώριση αντικειμένων και η κατάτμηση εικόνας αλλά και πλήθος άλλες εφαρμογές: Αναγνώριση ομιλίας, μεταγραφή λόγου σε κείμενο, αναγνώριση εξωτικών σωματιδίων στη φυσική υψηλής ενέργειας, βιοπληροφορική, εφαρμογές στην ιατρική, βιοακουστική, κ.α.

5.9.1 Συνελικτικά Νευρωνικά Δίκτυα (C.N.N.)

Τα συνελικτικά νευρωνικά δίκτυα (CNN) προσομοιώνουν στη διαδικασία αναγνώρισης τα διαδοχικά στάδια μέσα από τα οποία ο ανθρώπινος εγκέφαλος επεξεργάζεται τις εικόνες στον οπτικό φλοιό. Η ανίχνευση ακμών, γραμμών ή περιγραμμάτων θεωρείται ότι αποτελεί ένα πρώτο στάδιο της επεξεργασίας μιας εικόνας μαζί με άλλες χαμηλού επιπέδου αναπαραστάσεις που αφορούν στην ανίχνευση φωτός ή χρώματος. Σε ένα επόμενο στάδιο ο οπτικός φλοιός αναγνωρίζει περισσότερο σύνθετες φόρμες όπως μοτίβα ή την υφή ενός αντικειμένου. Τέλος διαφορετικά συστατικά της εικόνας συνδέονται προκειμένου να σχηματίσουν μια συνεκτική αντίληψη των αντικειμένων και των σκηνών εντός του περιβάλλοντος στο οποίο βρίσκονται ενσωματώνοντας ταυτόχρονα μηχανισμούς προσοχής. Η παραπάνω ακολουθιακή ιεραρχική δομή αντανακλά τον τρόπο με τον οποίο η ανθρώπινη όραση καταλαβαίνει σταδιακά και ταξινομεί τα αντικείμενα και συνιστά ένα μοντέλο που προσομοιώνουν οι βασικές βαθιές αρχιτεκτονικές των συνελικτικών νευρωνικών δικτύων.

Τα συνελικτικά δίκτυα εξειδικεύονται σε επεξεργασία δεδομένων που έχουν τοπολογία

μονοδιάστατων πλεγμάτων (1d grid) όπως για παράδειγμα χρονοσειρές, δισδιάστατων πλεγμάτων (2d grid pixels) όπως για παράδειγμα εικόνες ή τρισδιάστατων πλεγμάτων (3d grid pixels) όπως για παράδειγμα βίντεο. Η αρχιτεκτονική τους συνίσταται σε συνελικτικά στρώματα (convolutional layers) που ακολουθούνται από στρώματα ομαδοποίησης (pooling layers), και σειριακά από πλήρως συνδεδεμένα στρώματα (fully connected layers). Στα συνελικτικά στρώματα, τα διανύσματα εισόδου συνελλίσσονται με φίλτρα-πυρήνες και εξάγονται χάρτες χαρακτηριστικών (feature maps). Σε μια δισδιάστατη εικόνα I(x,y) η λειτουργία της συνέλιξης με μια συνάρτηση πυρήνα K(x,y) περιγράφεται μαθηματικά ως εξής:

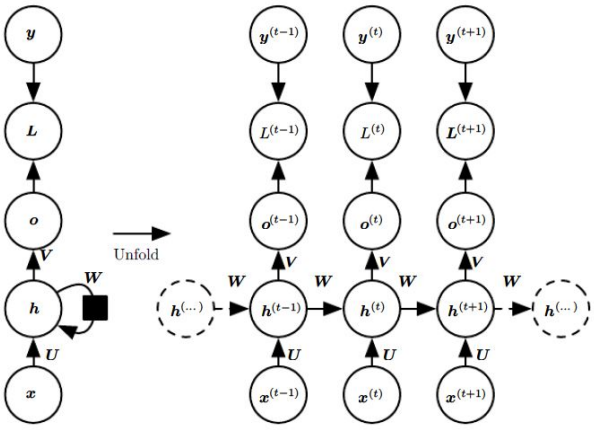
$$S_{i,j} = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) \tag{5.11}$$

Η παραπάνω πράξη συνέλιξης υποδηλώνει πως κάθε σημείο (x,y) του grid αφομοιώνει στο περιεχόμενό του το άθροισμα της έντασης των εικονοστοιχείων της γειτονιάς της εικόνας στην οποία ανήκει με το εφαρμοζόμενο σε αυτή φίλτρο. Τα συνελικτικά δίκτυα χαρακτηρίζονται από διαμοιρασμό όμοιων παραμέτρων (parameter sharing) κατά μήκος πολλαπλών τοποθεσιών της εικόνας, επιτρέποντας σε κάθε στοιχείο του πυρήνα (kernel) να συνελλίσσεται αυτούσιο με τις γειτονιές της εικόνας, περιορίζοντας τον αριθμό των υπό προσαρμογή παραμέτρων και μειώνοντας την πολυπλοκότητα του μοντέλου. Σε ένα επόμενο στάδιο οι χάρτες χαρακτηριστικών που παράγονται από τις συνελίξεις εισέρχονται σε στρώματα ομαδοποίησης (pooling layers) που επιτελούν κάποιου τύπου στατιστική περίληψη μεγιστοποιώντας ή βρίσκοντας το μέσο όρο μιας γειτονιάς στοιχείων του grid. Στόχος των στρωμάτων αυτών είναι να κάνουν την αναπαράσταση σχετικά αμετάβλητη σε μικρές μετατοπίσεις εντός του διανύσματος εισόδου και να ενισχύσουν την ικανότητα γενίκευσης που έχει το δίκτυο εστιάζοντας στην παρουσία χαρακτηριστικών στο διάνυσμα εισόδου παρά στον εντοπισμό της θέσης τους [28].

5.9.2 Αναδρομικά Νευρωνικά Δίκτυα (R.N.N.)

Τα Recurrent Neural Networks, (R.N.N) είναι μια οικογένεια νευρωνικών δικτύων που ειδικεύεται στην ανάλυση και επεξεργασία ακολουθιακών δεδομένων με τον ίδιο τρόπο που τα συνελικτικά δίκτυα αναλύουν συνήθως δισδιάστατα διανύσματα όπως εικόνες (Goodfelow et al.). Στην πραγματικότητα τα RNN συνιστούν μια επέκταση των εμπρόσθιων νευρωνικών δικτύων (feedforward neural networks) ικανών να διαχειριστούν ακολουθίες εισόδου μεταβλητού μήκους με τη βοήθεια αναδρομικών κρυμμένων καταστάσεων (hidden state) των οποίων η ενεργοποίηση κάθε στιγμή εξαρτάται από αυτήν προηγούμενων χρονικών στιγμών. Τα αναδρομικά νευρωνικά δίκτυα μπορούν να μάθουν τη δυναμική των χρονικών προτύπων ενός σήματος μέσω των αναδρομικών συνδέσεων που διαθέτουν μεταξύ των στρωμάτων (layers) επιτρέποντας προγενέστερη ή και μεταγενέστερη πληροφορία εισόδου να διαμορφώσει την εκτίμηση της παρούσας εξόδου. Λόγω της ικανότητάς τους να επεξεργάζονται ακολουθίες μεταβλητού μήκους χρησιμοποιούνται ευρέως στη μοντελοποίηση αναγνώρισης φωνής, στην πρόβλεψη χρονοσειρών, μουσικής σύνθεσης κ.α. Παρ΄ όλα αυτά, το φαινόμενο των φθινουσών ή εκθετικά αυξανομένων κλίσεων (vanishing, exploding gradients) κατά τη διάρκεια εκτέλεσης του αλγορίθμου οπισθοδιάδοσης (backpropagation algorithm) εμποδίζουν τα αναδρομικά δίκτυα να μαθαίνουν μακροπρόθεσμες εξαρτήσεις μεταξύ απομακρυσμένων χρονικά εισόδων πρόβλημα που οδήγησε στα μοντέλα LSTM, GRU τα οποία επιλύουν το πρόβλημα αυτό με τη χρήση θυρών.

Σχήμα 5.4: Αναδρομικό Νευρωνικό Δίκτυο RNN



Η φωτογραφία αναπαράγεται από το βιβλίο Deep Learning, GoodFellow et al.[28]

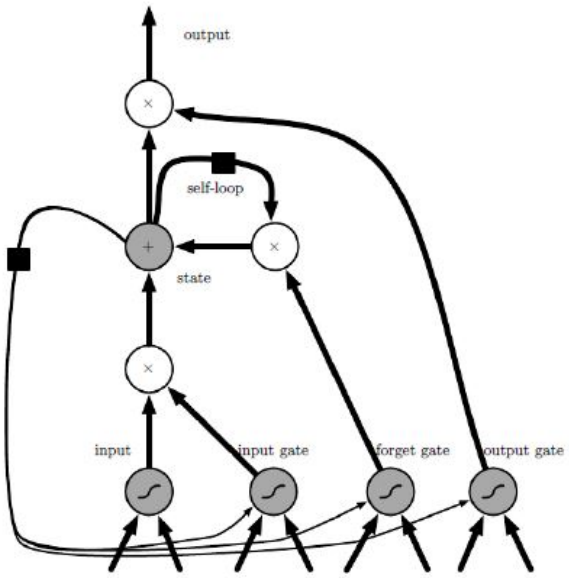
Στην παραπάνω φωτογραφία απεικονίζεται η γενική μορφή ενός αναδρομικού νευρωνικού δικτύου που επεξεργάζεται μια ακολουθία τιμών εισόδου $x_1, x_2, ..., x_T$ σε μια ακολουθία τιμών εξόδου $y_1, y_2, ..., y_T$ μέσω κρυμμένων καταστάσεων. Ο βασικός ρόλος μιας κρυμμένης κατάστασης h_t είναι να συμπυκνώσει περίληψη των προηγούμενων τιμών της ακολουθίας εισόδου λειτουργώντας ως ΄μνήμη΄ του δικτύου. Η λειτουργία της αποτυπώνεται μαθηματικά μέσω μιας δυναμικής μαθηματικής εξίσωσης της μορφής: $h_t = f(h_{t-1}, x_t; \theta)$. Η πληροφορία που αποκρυσταλλώνεται ως μνήμη του δικτύου σε μια κρυμμένη κατάσταση είναι περιληπτική

εκείνης του χώρου εισόδου. Η υποβάθμιση πληροφοριακού περιεχομένου διαπιστώνεται και από το γεγονός ότι ακολουθίες εισόδου αυθαίρετου μήκους απεικονίζονται σε μια κρυμμένη κατάσταση σταθερού μήκους h_t. Ιδανικά συνεπώς η βέλτιστη κατάσταση είναι εκείνη για την οποία το μέγιστο πληροφοριακό περιεχόμενο αποτυπώνεται σε μια κρυμμένη κατάσταση.

5.9.3 Δίκτυο Μακράς Βραχείας Μνήμης (L.S.T.M.)

Τα δίκτυα LSTM αποτελούν ειδική κατηγορία των RNN που επιχειρεί να επιλύσει το πρόβλημα των χρονικά μακροπρόθεσμων εξαρτήσεων επιτρέποντας τη ροή πληροφορίας στο χρόνο, χρησιμοποιώντας ως θεμελιώδεις μονάδες αναδρομικά συνδεδεμένα blocks μνήμης. Τα LSTM διαδίδουν την πληροφορία μέσα στο χρόνο μέσω άθροισης [28]. Το δίκτυο LSTM διαφοροποιείται από τα RNN αρχιτεκτονικά στην εισαγωγή θυρών (Gates) που ελέγχουν την πληροφορία που εισέρχεται σε αυτό (Input gate), διατηρώντας την πληροφορία που πρέπει να συντηρηθεί (Forget gate) λειτουργώντας ως συσσωρευτές και επιλέγοντας την πληροφορία που θα πρέπει να οδηγηθεί στην πύλη εξόδου (Output gate). Τα LSTM δίκτυα έχουν συνθετότερες επαναλαμβανόμενες δομές, συγκεκριμένα 4 στρώματα που αλληλεπιδρούν ώστε να ελέγξουν τη ροή πληροφορίας. Οι πύλες επιτρέπουν τη ροή πληροφορίας μέσω για παράδειγμα ενός σιγμοειδούς στρώματος νευρωνικού δικτύου και έναν πολλαπλασιασμό. Η σιγμοειδής δέχεται το input και το συμπιέζει φράζοντας το στο πεδίο τιμών [0,1] αξιολογώντας στην κλίμακα αυτή τη σημαντικότητα της πληροφορίας που δέχεται ως είσοδο.

Σχήμα 5.5: Αρχιτεκτονική μιας μονάδας LSTM



Η φωτογραφία αναπαράγεται από το βιβλίο Deep Learning, GoodFellow et al.[28]

Στην παρακάνω εικόνα από τους Goodfellow et al. [28] φαίνεται ο μηχανισμός λειτουργίας ενός κυττάρου μνήμης. Όπως περιγράφει ο συγγραφέας, το κύτταρο αποτελείται από ένα συνήθη νευρώνα που τροφοδοτείται από το διάνυσμα εισόδου, το αποτέλεσμα του οποίου ανατροφοδοτεί τις πύλες εισόδου, εξόδου και μνήμης. Η πύλη εισόδου μπορεί να λειτουργήσει ως πύλη επίτρεψης μιας σιγμοειδούς εκδοχής της εισόδου στην κατάσταση 'state'. Ο νευρώνας κατάστασης έχει αναδρομικό βρόχο που ελέγχεται από την πύλη forget ενώ το συνολικό αποτέλεσμα του κυττάρου μνήμης μπορεί να διαβιβαστεί ή να αποκοπεί από τη θύρα εξόδου μέσω της θύρας output. Όλες οι θύρες διαθέτουν έναν μη γραμμικό σιγμοειδή μηχανισμό

ενώ όπως φαίνεται στο σχήμα η μονάδα κατάστασης forget unit μπορεί να χρησιμοποιηθεί ως είσοδος στις μονάδες των θυρών. Οι εξισώσεις κατάστασης του LSTM δικτύου περιγράφονται μαθηματικά από τις παρακάτω σχέσεις:

$$i_t = \sigma(W_i\dot{[}h_{t-1}, x_t] + b_i) \tag{5.12}$$

$$\tilde{C}_t = tanh(W_c\dot{[}h_{t-1}, x_t] + b_c) \tag{5.13}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{5.14}$$

$$f_t = \sigma(W_f\dot{[}h_{t-1}, x_t] + b_f) \tag{5.15}$$

$$o_t = \sigma(W_0\dot{[}h_{t-1}, x_t] + b_0) \tag{5.16}$$

$$h_t = o_t * tanh(C_t) \tag{5.17}$$

Ο παραπάνω μηχανισμός λειτουργίας επιτρέπει να συντηρηθεί η ευρωστία (robustness) στη ροή παραγώγισης κατά τη διαδικασία του backpropagation αλγόριθμου συμβάλλοντας στην ικανότητα μάθησης του LSTM να διαπιστώνει εξαρτήσεις που εκτείνονται σε βάθος χρόνου.

5.9.4 Συμπεράσματα και συζήτηση

Η ανάπτυξη των συνελικτικών νευρωνικών δικτύων και η επιτυχία τους στην επίλυση μιας σειράς προβλημάτων αναγνώρισης προτύπων έγκειται σε μεγάλο βαθμό στην ικανότητα της λειτουργίας της χωρικής συνέλιξης εικόνων με kernels να εξάγει χαρακτηριστικά του χώρου εισόδου, όπως ακμές, κορυφές και περιοχές ενδιαφέροντος. Η παραπάνω λειτουργία παραπέμπει στις τεχνικές εξομάλυνσης και όξυνσης ακμών που αναπτύξαμε στην ενότητα ψηφιακής επεξεργασίας εικόνας (4.3). Παρ΄όλα αυτά στα πλαίσια της βαθιάς μάθησης αυξάνονται οι υπολογιστικές δυνατότητες με παράλληλες εφαρμογές της συνέλιξης σε κάθε στρώμα του δικτύου και παραγωγή πολλαπλών ειδών χαρακτηριστικών για κάθε διαφορετική τοποθεσία στην εικόνα. Εξάγοντας για κάθε περιοχή από εικονοστοιχεία της εικόνας πολλαπλά πρότυπα η βαθιά μάθηση επιτυγχάνει να φτιάξει αναπαραστάσεις χαρακτηριστικών που ενσωματώνουν σημαντικό βαθμό και διάσταση πληροφορίας από τη δομή του χώρου εισόδου.

Κεφάλαιο 6

Πειραματικό μέρος: Αναγνώριση βιοσήματος οδοντοκητών με τεχνικές βαθιάς μάθησης

6.1 Εισαγωγή

[1] Στην εργασία που ακολουθεί εστιάζουμε στην ανάπτυξη ενός συνελικτικού αναδρομικού νευρωνικού δικτύου (Convolutional Recurrent Neural Network) για την κατηγοριοποίηση βιοσημάτων που έχουν συλλεχθεί στην Ελληνική Τάφρο, από δύο είδη κητωδών, τους φυσητήρες (*Physeter macrocephalus*) και τα ζωνοδέλφινα (*Stenella coeruleoalba*). Μετατρέπουμε τα ηχητικά σήματα σε mel-spectrograms εφαρμόζοντας τεχνικές κανονικοποίησης ενέργειας ανά κανάλι (P.C.E.N) και τροφοδοτούμε την είσοδο σε βαθύ ResNet που έχει σχεδιαστεί ώστε να εξάγει φασματικά πρότυπα. Στη συνέχεια ανασχηματίζουμε τις διαστάσεις της εξόδου του ResNet και τροφοδοτούμε το διάνυσμα χαρακτηριστικών σε ένα αναδρομικό δίκτυο (LSTM ή GRU) ικανό να αναγνωρίζει μακροπρόθεσμες χρονικές εξαρτήσεις. Το υβριδικό δίκτυο αναγνωρίζει ηχητικά σήματα που ταξινομούνται σε τρεις διαφορετικές κατηγορίες (βιοσήματα φυσητήρων και δελφινιών έναντι θορύβου περιβάλλοντος) ενώ επιδεικνύει ισχυρή ικανότητα μάθησης στην αναγνώριση ηχητικών εκφωνήσεων που αλληλεπικαλύπτονται (clicks έναντι clicks and whistles δελφινιών). Η προτεινόμενη υβριδική αρχιτεκτονική υπερτερεί σε απόδοση έναντι παραδοσιακών τεχνικών μηχανικής μάθησης αλλά και βασικών αρχιτεκτονικών βαθιάς μάθησης ResNet ή LSTM.

[1] Προκαταρκτικά αποτελέσματα της μελέτης δημοσιεύτηκαν στο [33] ενώ μια πιο αναλυτική εργασία έχει υποβληθεί για δημοσίευση στο [34]

6.2 Σχετική επιστημονική έρευνα

Η βιοακουστική και ιδιαιτέρως το πρόβλημα αναγνώρισης υποθαλάσσιου ήχου συγκεντρώνει με αυξανόμενο ρυθμό το ενδιαφέρον ερευνητών από την περιοχή της μηχανικής μάθησης. Διάφορες προσεγγίσεις έχουν ακολουθηθεί στο σχεδιασμό αλγορίθμων που στοχεύουν να ταξινομήσουν βιοσήματα από κητώδη, περιλαμβάνοντας παραδοσιακές τεχνικές μηχανικής μάθησης που εκπαιδεύονται σε διανύσματα χαρακτηριστικών και μοντέρνες τεχνικές βαθιάς μάθησης. Συγκεκριμένα, αλγόριθμοι GMM και SVM που επιχειρούν να οριοθετήσουν τα σύνορα μεταξύ κατανομών έχουν χρησιμοποιηθεί για να κατασκευάσουν ταξινομητές cepstral χαρακτηριστικών που εξάγονται από φάσματα για τα παρακάτω είδη: Blainville's beaked whales *(Mesoplodon densirostris)*, short-finned pilot whales *(Globicephala macrorhynchus)*, Risso's dolphins *(Grampus griseus)*[35]. Επίσης δίκτυα Hidden Markov Models, HMM που λαμβάνουν υπόψιν τους τόσο τη χρονική όσο και τη φασματική δομή καθώς διαχειρίζονται τα δεδομένα ως ακολουθίες καταστάσεων, μπορούν να αναγνωρίσουν από ένα πληθυσμό γνωστών κητωδών killer whales *(Orcinus orca)* την ταυτότητα συγκεκριμένων ατόμων με διακριτή διάλεκτο [36]. Δίκτυα μη επιβλεπόμενης μάθησης έχουν επίσης χρησιμοποιηθεί στη βάση χρονο-συχνοτικών κατανομών για την εκπαίδευση αυτο-οργανούμενων χαρτών χαρακτηριστικών (self-organizing feature maps) οδηγώντας σε ομαδοποιήσεις διαφορετικών τύπων Μεγάπτερων φαλαινών *(Megaptera novaeangliae)* [37].

Σε πρόσφατες μελέτες, αποδείχτηκε ότι βαθιά CNN εκπαιδευμένα σε φασματογραφήματα από ήχους κητωδών είναι ικανά να ανιχνεύσουν και να ταξινομήσουν clicks φυσητήρων ή ήχους μεγάπτερων φαλαινών (Humpback Whales - *Megaptera novaeangliae*) [38]. Στην ίδια γραμμή έρευνας οι Bermant et al [8] χρησιμοποιούν ένα μοντέλο βαθιάς μάθησης CNN με είσοδο 650 φάσματα φυσητήρων και ηχητικών αρχείων που δεν περιέχουν clicks στοχεύοντας στη διατύπωση ενός boolean κριτηρίου απόφασης (ανίχνευση/μη ανίχνευση) σημάτων ήχο-εντοπισμού (echolocation clicks) επιτυγχάνοντας ακρίβεια αναγνώρισης 99.5%. Επιπλέον δίκτυα αυτο-επιβλεπόμενης μάθησης (self-supervised) που υλοποιούν αναδρομικά δίκτυα (LSTM ή GRU) επιτυγχάνουν ταξινόμηση codas φυσητήρων από δύο διαφορετικά γεωγραφικά datasets σε διακριτές κλάσεις-διαλέκτους. Παρόμοιες αρχιτεκτονικές αναγνωρίζουν με επιτυχία την ταυτότητα ατόμων από το ίδιο γεωγραφικό dataset με απόδοση 99.4%.

6.3 Δεδομένα: Ανάλυση και Προεπεξεργασία

Τα δεδομένα που έχουν χρησιμοποιηθεί για την πειραματική μελέτη προέρχονται από τρεις διαφορετικές πηγές:

1) Υδρόφωνο (Passive Acoustic Listener) με συχνότητα δειγματοληψίας στα 100kHz ποντίστηκε σε παρατηρητήριο στην Πύλο τη χρονική περίοδο 11/11/2008 - 17/09/2009, 10 χιλιόμετρα από τη Δυτική Ακτή της Πελοποννήσου [39].

2) Ακουστικά δεδομένα συλλέχθηκαν μέσω ρυμουλκούμενης συστοιχίας υδροφώνων κατά τη διάρκεια ερευνών από το Ινστιτούτο Κητολογικών Ερευνών Πέλαγος΄ κατά μήκος της Ελληνικής Τάφρου κατά την περίοδο 2001-2020 με συχνότητα δειγματοληψίας 48 kHz [40].

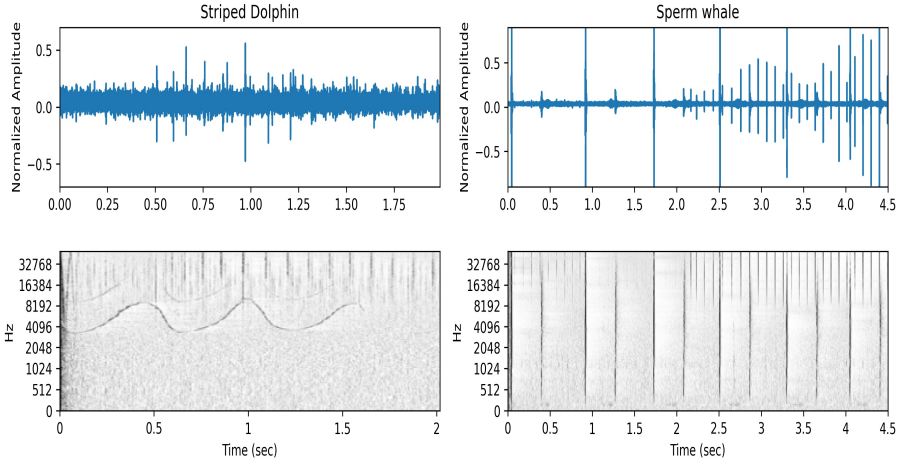
3) Ακουστικές καταγραφές πραγματοποιήθηκαν με συχνότητες έως 100 kHz το 2020-2021 από το παρατηρητήριο SAvE Whales Observatory (Σύστημα για την αποφυγή προσκρούσεων με κήτη σε κίνδυνο) που αποτελείτο από 3 ακουστικούς σταθμούς καθένας από τους οποίους διαθέτει υδρόφωνο σε βάθος 100m [41].

Η βάση είναι επισημειωμένη από εξειδικευμένο θαλάσσιο βιολόγο που έχει αντιστοιχίσει σε κάθε ηχητικό αρχείο, επιγραφή (label) σχετική με το είδος του θηλαστικού/ων που αναγνωρίζει.

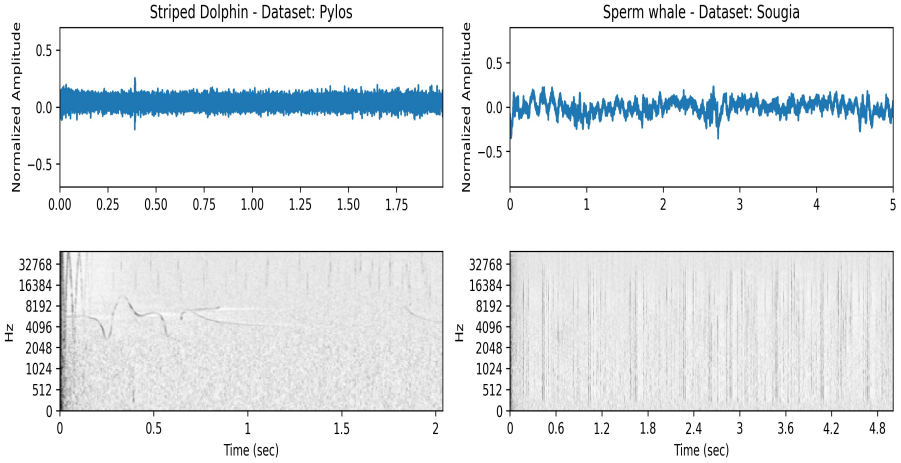
Στο σχήμα 6.1 απεικονίζονται ηχητικές κυματομορφές και φασματογραφήματα για τα δύο είδη κητωδών από το σύνολο δεδομένων της Πύλου. Στην αριστερή πλευρά της εικόνας σχεδιάζεται η ηχητική κυματομορφή ζωνοδέλφινου που αποτελείται από clicks και σφυρίγματα (whistles) μαζί με μια μετατοπισμένη προς τα πάνω αρμονική συνιστώσα του συνεχούς σήματος. Δεξιά απεικονίζεται ηχητική καταγραφή φυσητήρα στην Πύλο μαζί με το φασματογράφημά του. Παρατηρούνται σε αυτό χαρακτηριστικές ακολουθίες από clicks ταυτόχρονα με ήχους χαμηλότερης ενέργειας που προέρχονται από ανακλάσεις στον πυθμένα ή εντός της ρινικής δομής του κήτους. Στο σχήμα 6.2 επίσης απεικονίζεται κυματόμορφή ζωνοδέλφινου και δεξιά ηχητικό σήμα φυσητήρα από το dataset της Σούγιας ενώ στο σχήμα 6.3 αποτυπώνονται clicks ζωνοδέλφινων στην αριστερή πλευρά και ξανά ηχητικά φυσητήρα στο σχήμα που βρίσκεται δεξιά.

Το σύνολο δεδομένων αποτελείται από ακουστικά σήματα που ανήκουν σε 3 κατηγορίες: 291 καταγραφές φυσητήρων, 90 σήματα από περιβάλλοντα θόρυβο που χαρακτηρίζονται από απουσία βιοσημάτων, και 284 καταγραφές ζωνοδέλφινων για ένα πείραμα ταξινόμησης ήχων μεταξύ διαφορετικών κλάσεων. Η επιλογή να χρησιμοποιηθεί σχετικά περιορισμένος αριθμός σημάτων στα οποία να απουσιάζουν βιοσήματα και να αντιπροσωπεύουν τη γενική κλάση υποθαλάσσιου θορύβου περιβάλλοντος αποφασίστηκε ώστε να αυξηθεί και ο βαθμός δυσκολίας του προβλήματος αφού στόχος της εργασίας είναι αναζητήσει τα όρια επιτυχίας του νευρωνι-

Σχήμα 6.1: Κυματομορφές και φάσματογραφήματα κητωδών - dataset από το παρατηρητήριο της Πύλου



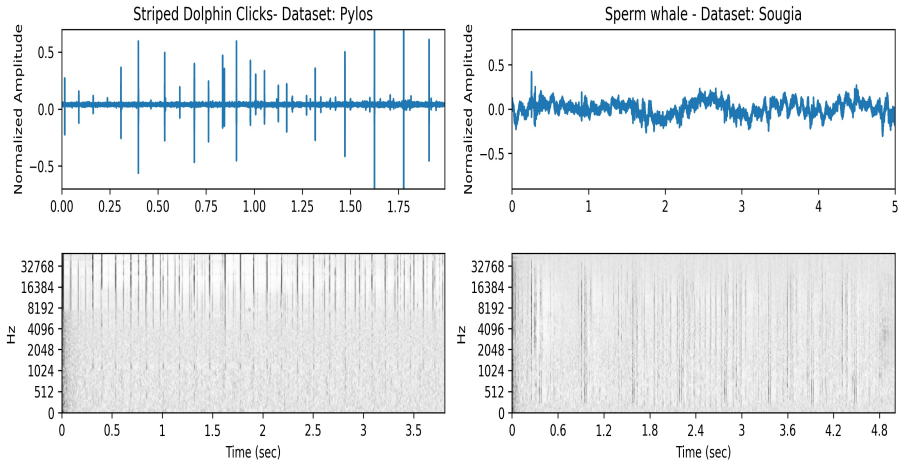
Σχήμα 6.2: Κυματομορφές και φάσματογραφήματα κητωδών - dataset από το παρατηρητήριο της Πύλου και της Σούγιας



κού δικτύου. Στη συνέχεια ακολούθησε η σχεδίαση ενός δεύτερου πειράματος που προέκυψε από τη διαίρεση της κλάσης του ζωνοδέλφινου σε 2 διακριτές -μερικώς αλληλεπικαλυπτόμενες- υποκλάσεις δελφινιών (135 ακουστικά αρχεία αφορούν σε clicks ενώ 149 καταγραφές συντίθενται από whistles and clicks).

Παρατηρήθηκε ότι στα πρώτα δευτερόλεπτα καταγραφής πολλά από τα σήματα χαρακτηρίζονται από ένα άηχο εισαγωγικό τμήμα μεταβλητού χρόνου που οφείλεται σε σφάλμα του οργάνου μέτρησης. Για την αποκοπή του άηχου τμήματος που δεν εμπεριέχει χρήσιμη πληροφορία υπολογίζουμε την ενέργεια βραχέος χρόνου του σήματος $E = \sum_{-\infty}^{\infty}(x[m]w[n-m])^2$ και επιλέγουμε ένα κατώφλι στο οποίο η πρώτη παράγωγός της μεταβάλλεται σημαντικά ώστε να διαχωρίσουμε την ηχητική συνιστώσα από το σήμα υποβάθρου.

Σχήμα 6.3: Ζωνοδέλφινα Clicks από το παρατηρητήριο της Πύλου και της Σούγιας



6.4 Φασματογραφήματα βιοσημάτων

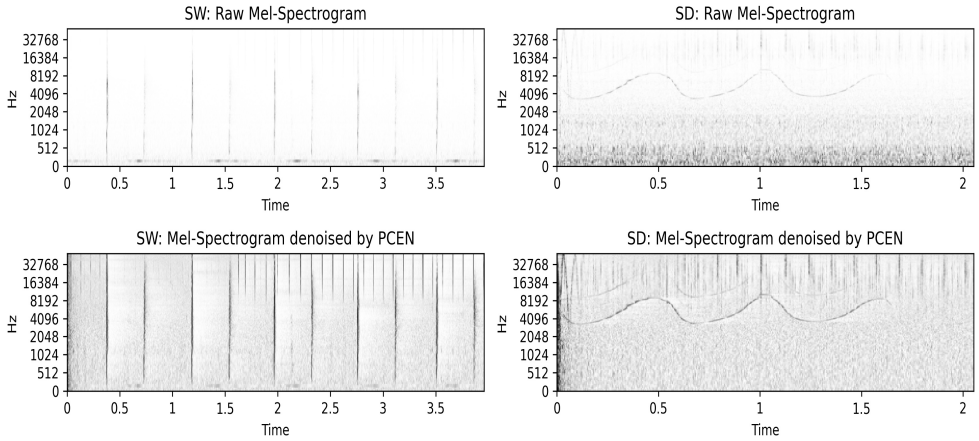
Προκειμένου να αποθορυβοποιηθεί το σήμα από ηλεκτρονικό θόρυβο στενής συχνοτικής ζώνης που προκαλείται από το όργανο μέτρησης και εμφανίζεται στο φασματογράφημα με τη μορφή οριζόντιων γραμμών εφαρμόζουμε την τεχνική κανονικοποίησης ενέργειας ανα συχνοτικό κανάλι (Per Channel Energy Normalization , P.C.E.N). Η τεχνική αυτή εφαρμόζεται σε προβλήματα ανίχνευσης ακουστικών γεγονότων και υλοποιείται σε δύο βήματα. Σε ένα πρώτο βήμα εφαρμόζεται αυτόματος έλεγχος κέρδους (automatic gain control) ώστε να διατηρήθεί η ενέργεια της ακουστικής κυματομορφής κοντά σε ένα προκαθορισμένο επίπεδο-στόχο. Αυτό γίνεται διαιρώντας τον μετασχηματισμό S.T.F.T. με μια χρονικά εξομαλυμένη εκδοχή του ίδιου μετασχηματισμού S.T.F.T.. Η εξομάλυνση πραγματοποιείται ανεξάρτητα για κάθε συχνοτικό bin. Στη συνέχεια εφαρμόζεται δυναμική συμπίεση κλίμακας dynamic root compression.

$$PCEN = (\frac{E(t,f)}{(\epsilon + M(t,f))^a} + \delta)^r - \delta^r \tag{6.1}$$

όπου το M(t,f) είναι η εξομαλυμένη εκδοχή του φάσματος και υπολογίζεται με τη βοήθεια ενός IIR φίλτρου:

$$M(t,f) = (1-s)M(t-1,f) + sE(t,f) \tag{6.2}$$

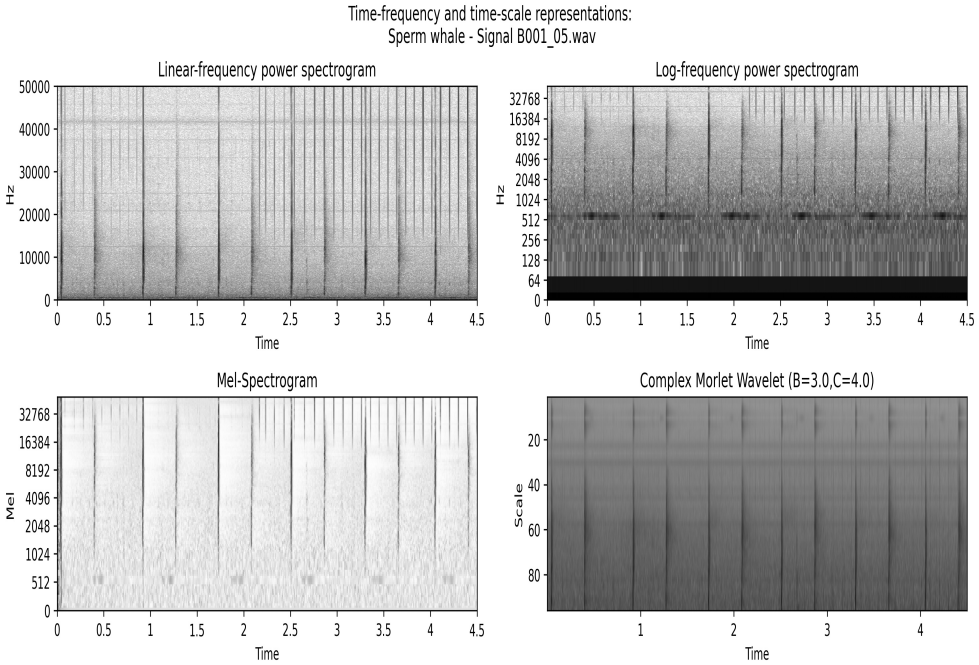
Ο πρώτος όρος της σχέσης $\frac{E(t,f)}{(\epsilon+M(t,f))^a}$ υλοποιεί έναν αυτόματο έλεγχο κέρδους που σκοπεύει στη μείωση του στάσιμου background θορύβου και ελέγχεται από την παράμετρο κανονικοποίησης κέρδους a που ορίζεται στο [0,1]. Η δυναμική συμπίεση κλίμακας (Dynamic Range compression) μειώνει τη διασπορά της έντασης του ήχου (foreground loudness) [42]. Η χρησιμοποίηση του φίλτρου P.C.E.N. καταφέρνει να ενισχύσει την αντίθεση ανάμεσα στο θόρυβο background και στα μεταβατικά (foreground) γεγονότα που οφείλονται στα βιοσήματα που εκπέμπουν τα κητώδη.



Σχήμα 6.4: Στο αριστερό μέρος της εικόνας εμφανίζεται Mel-spectrogram φυσητήρα της Πύλου ενώ στο κάτω μέρος έχει εφαρμοστεί αποθορυβοποίηση με φίλτρο PCEN. Στο δεξιό μέρος της εικόνας το συμμετρικό αντίστοιχο για ζωνοδέλφινο. Διαπιστώνεται ενίσχυση του contrast μεταξύ background θορύβου και foreground μεταβατικών σημάτων.

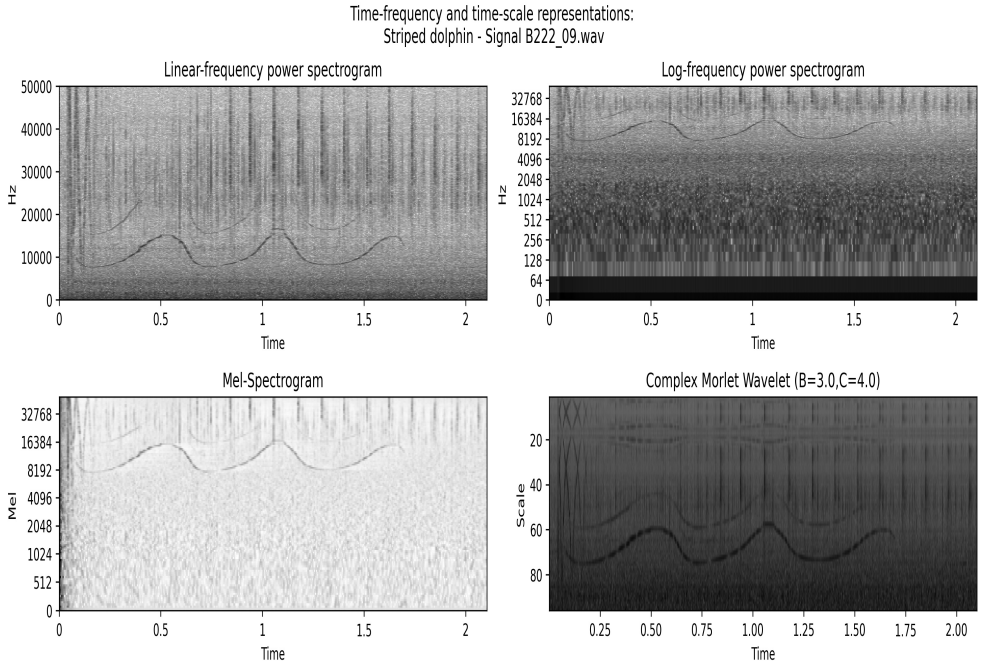
Στην παρακάτω εικόνα από ηχητικό φυσητήρα έχουν σχεδιαστεί φασματογραφήματα προσαρμοσμένα σε διαφορετική κλίμακα (γραμμικά, log, mel) όπως επίσης και ένα scalogram που έχει σχεδιαστεί με μητρικό κυματίδιο Complex Morlet.

Σχήμα 6.5: Απεικόνιση φασματογραφημάτων φυσητήρα σε διαφορετικές κλίμακες και Scalogram



Παρατηρούμε ότι σε ένα log-φασματογράφημα όπως και στην κλίμακα mel αναδεικνύεται σε αντίθεση με το γραμμικό ανάλογό τους η απουσία βιοσήματος από 0-1 kHz που μπορεί

Σχήμα 6.6: Απεικόνιση φασματογραφημάτων ζωνοδέλφινου σε διαφορετικές κλίμακες και Scalogram



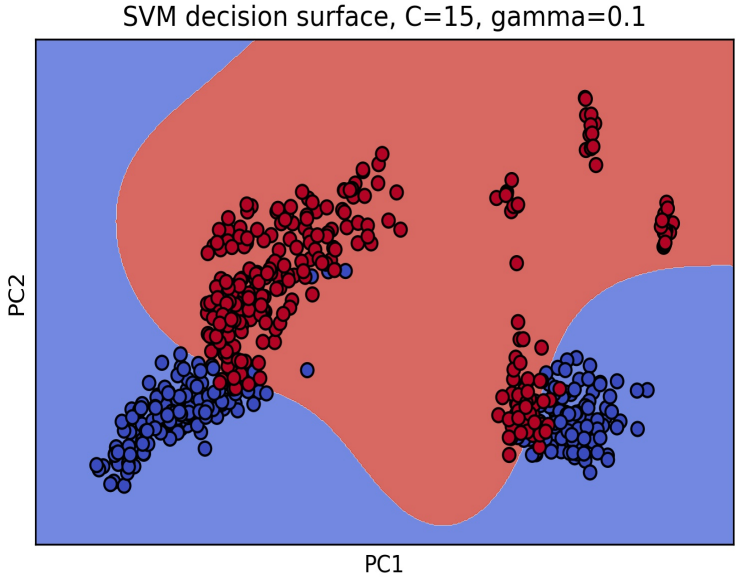
να δικαιολογήσει και την απόφασή μας για αποθορυβοποίηση σε αυτό το εύρος συχνοτήτων. Παρατηρούμε επίσης ότι στο mel-spectrogram στο οποίο έχει χρησιμοποιηθεί τεχνική αποθορυβοποίησης PCEN απουσιάζει ο ιδιοθόρυβος του υδρόφωνου που εμφανιζόταν τόσο στο γραμμικό όσο και στο λογαριθμικό φάσμα με τη μορφή μιας οριζόντιας γραμμής.

6.5 Κατηγοριοποίηση βιοσημάτων με τεχνικές παραδοσιακής μηχανικής μάθησης σε δύο ηχητικές κατηγορίες

Για σκοπούς σύγκρισης αναπτύξαμε ένα παραδοσιακό μοντέλο μηχανικής μάθησης βασισμένο σε Mel Frequency Cepstral Coefficients, (MFCCs) συντελεστές, ένα διάνυσμα χαρακτηριστικών που χρησιμοποιείται ευρέως σε παραδοσιακά μοντέλα μηχανικής μάθησης ακολουθούμενο από SVM ή kNN. Η εξαγωγή των MFCCs χαρακτηριστικών πραγματοποιήθηκε εφαρμόζοντας σε κάθε πλαίσιο (frame) τον Διακριτό Μετασχηματισμό Συνημιτόνου (Discrete Cosine Transform) σε log-Mel φασματογραφήματα κάθε ηχητικού αρχείου. Επιλέξαμε να διατηρήσουμε τους πρώτους 13 cepstral συντελεστές ανά frame που εμπεριέχουν το μεγαλύτερο κομμάτι πληροφορίας, στη συνέχεια απορρίψαμε τον μηδενικό συντελεστή που αντιπροσωπεύει τη μέση log energy του σήματος $log(\sum_{i=0}^{i=N}(x[i]^2))$ και πραγματοποιήσαμε επαύξηση του διανύσματος χαρακτηριστικών υπολογίζοντας τις πρώτες παραγώγους Δ-MFCCs. Τέλος, υπολογίσαμε τη μέση τιμή των MFCC χαρακτηριστικών κατά μήκος όλων των frames για να κατασκευάσουμε ένα διάνυσμα (MFCC, Δ-MFCCs) χαρακτηριστικών διάστασης 24x1 για κάθε ηχητικό αρχείο.

Σε αυτό το πείραμα κατηγοριοποιούμε βιοσήματα σε δύο ακουστικές κλάσεις: Σήματα που προέρχονται από ζωνοδέλφινα και σήματα που προέρχονται από φυσητήρες. Εξάγουμε το διάνυσμα χαρακτηριστικών από τα ηχητικά σήματα και το προβάλλουμε με τη βοήθεια της τεχνικής PCA στο Ευκλείδιο επίπεδο. Στη συνέχεια εφαρμόζουμε έναν ταξινομητή SVM με συνάρτηση διαχωρισμού την RBF για να διαχωρίσει το διάνυσμα χαρακτηριστικών τις κλάσεις. Προκύπτει το επόμενο σχήμα:

Σχήμα 6.7: Προβολή στο δισδιάστατο επίπεδο με PCA



Πίνακας 6.1: Απόδοση διαφορετικών αρχιτεκτονικών

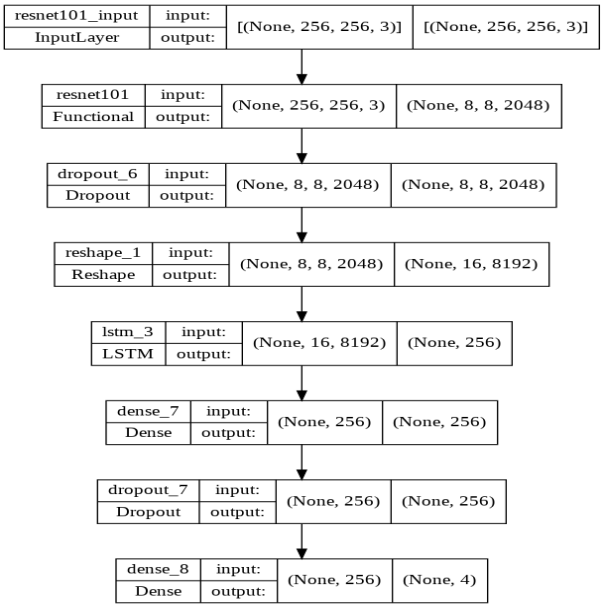
Μοντέλο	**Αποτελέσματα σε άγνωστο set**	
	Accuracy	***Precision***
MFCC - SVM (RBF kernel)	99.8%	100.0%
MFCC - Linear SVM	95.0%	97.0%
MFCC - Poly SVM	98.0%	98.0%
MFCC - Sigmoid SVM	94.0%	96.0%
MFCC-kNN	98.2%	99.0%
MFCC-Gaussian NB	90.5%	98.7%

Τέλος έχοντας εξάγει το διάνυσμα χαρακτηριστικών αξιολογήσαμε ταξινομητές παραδοσιακής μηχανικής μάθησης: linear, polynomial, RBF - SVM, SVM, kNN αξιοποιώντας εξαντλητική αναζήτηση (exhaustive search) σε μια σειρά παραμέτρων και υπολογίζουμε το accuracy. Στον παραπάνω πίνακα φαίνονται τα αποτελέσματα όπου παρατηρούμε ότι ένα SVM με RBF συνάρτηση διαχωρισμού μπορεί να διαχωρίσει και να ταξινομήσει σωστά ένα MFCC διάνυσμα χαρακτηριστικών. Στη συνέχεια δεδομένης της επιτυχίας των παραδοσιακών μεθόδων να λύσουν το βιοακουστικό πρόβλημα χωρίς να υπάρχει ανάγκη προσφυγής σε βαθιά μάθηση θα πραγματοποιήσουμε έναν πιο λεπτομερή πειραματικό σχεδιασμό προσθέτοντας μια ηχητική κλάση που αποτελείται από υποθαλάσσιους θορύβους αλλά όχι βιοσήματα ενώ στη συνέχεια θα υποδιαιρέσουμε το σύνολο των ηχητικών δεδομένων ζωνοδέλφινων σε δύο υποκλάσεις ηχητικών που αλληλεπικαλύπτονται ώστε να φανεί η ικανότητα μάθησης των δικτύων σε ένα πολυπλοκότερο πρόβλημα διαχωρισμού.

6.6 Συνελικτικά Αναδρομικά Νευρωνικά Δίκτυα

Στη μελέτη που ακολουθεί η προτεινόμενη αρχιτεκτονική συνίσταται σε μια ακολουθιακή δομή ενός Residual Network ακολουθούμενο από αναδρομικό δίκτυο LSTM ή GRU. Χρησιμοποιείται συγκεκριμένα προεκπαιδευμένο στο ImageNet δίκτυο ResNet101 ως μοντέλο βάσης. Τα δίκτυα ResNet χρησιμοποιούν πολλαπλά blocks που συνδέονται μεταξύ τους σειριακά, επιτρέποντας στην πληροφορία του σήματος να ρέει χωρίς απώλειες διαμέσου των στρωμάτων επιτελώντας όταν χρειάζεται ταυτοτικές απεικονίσεις (identity mapping). Χρησιμοποιήσαμε σε μια σειρά από πειράματα διαφορετικές εκδοχές αναδρομικών νευρωνικών δικτύων όπως LSTM, GRU ή την δικατευθυντική εκδοχή τους BiLSTMs, BiGRUs που είναι ικανή να αναγνωρίσει χρονικές εξαρτήσεις τόσο σε προηγούμενα όσο και σε επόμενα χρονικά βήματα.

Σχήμα 6.8: Ακολουθιακή αρχιτεκτονική ResNet-LSTM



Στο προηγούμενο σχήμα αποτυπώνεται το προτεινόμενο υβριδικό δίκτυο. Το πρώτο block του δικτύου χρησιμοποιεί αρχιτεκτονική ResNet101 επεξεργαζόμενο mel-spectrograms με εισόδους της μορφής 256x256x3 παράγοντας φασματικούς χάρτες χαρακτηριστικών η έξοδος των οποίων είναι 8x8x2048. Οι χάρτες χαρακτηριστικών ανασχηματίζονται σε ένα χρονοκατανεμημένο στρώμα διαστάσεων 16x8192, και προωθούνται σε ένα δεύτερο block που απότελείται από αναδρομικό δίκτυο με 256 units από το οποίο εξάγονται χρονικά χαρακτηριστικά που προωθούνται είτε σε ένα στρώμα μηχανισμού προσοχής (attention layer) είτε σε ένα πλήρως συνδεδεμένο στρώμα ακολουθούμενο από ένα στρώμα softmax στην κορυφή του για σκοπούς ταξινόμησης.

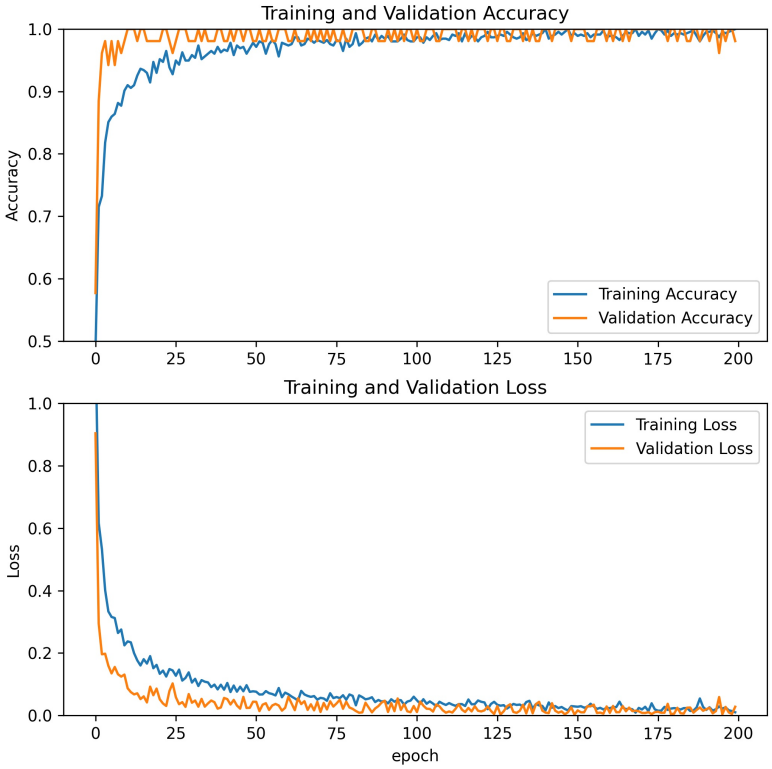
6.7 Βελτιστοποίηση παραμέτρων

Στη πειραματική εφαρμογή, το σύνολο δεδομένων που χρησιμοποιούμε διαιρείται σε 2 υποσύνολα εκπαίδευσης (training dataset) και επικύρωσης (validation dataset) με αναλογίες 80% και 20% αντίστοιχα. Χρησιμοποιήθηκε η K-Fold cross validation τεχνική η οποία συνίσταται στην εκπαίδευση K-1 υποσυνόλων με επικύρωση του K-ου υποσυνόλου. Το τελικό αποτέλεσμα υπολογίζεται ως η μέση απόδοση (mean accuracy) και η μέση ακρίβεια (mean precision) των άγνωστων συνόλων (test sets) των K πειραμάτων. Για τη μελέτη που πραγματοποιήσαμε, θέσαμε K=5 και το μοντέλο εκπαιδεύεται για 150 εποχές με batch size ίσο με 64. Χρησιμοποιήθηκε ο Adam αλγόριθμος βελτιστοποίησης με ρυθμό μάθησης (learning rate) ίσο με 10^{-3} ενώ η συνάρτηση σφάλματος που επιλέχθηκε είναι η categorical cross entropy. Τέλος χρησιμοποιούμε κανονικοποίηση στο τελευταίο στρώμα πριν τον ταξινομητή (dropout regularization) 40% απορρίπτοντας με τυχαίο τρόπο συνάψεις και να αποφευχθεί πιθανή υπερπροσαρμογή του μοντέλου στα δεδομένα.

6.8 Σχεδιασμός πειράματος κατηγοριοποίησης δύο ειδών βιοσημάτων και θορύβου

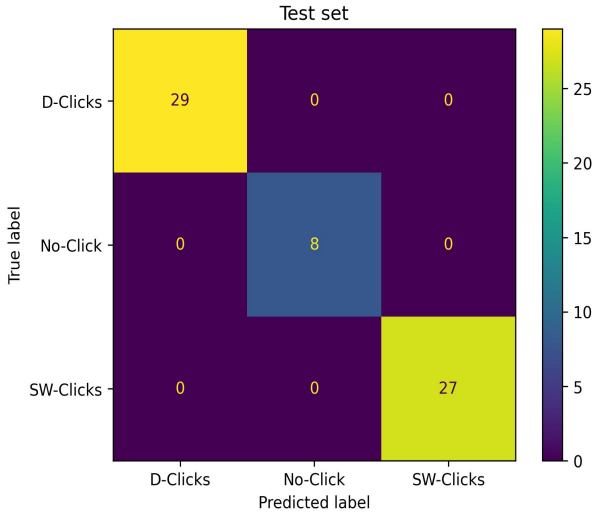
Ο πρώτος πειραματικός σχεδιασμός (ενότητα 6.5) αφορούσε σε ένα πρόβλημα ταξινόμησης δύο κλάσεων βιοσημάτων. Παρατηρήσαμε ότι κλασικά δίκτυα μηχανικής μάθησης που συνδυάζουν εξαγωγή cepstral χαρακτηριστικών και SVM επιλύουν το πρόβλημα ταξινόμησης. Επιχειρώντας να αναβαθμίσουμε τη δυσκολία του προβλήματος ενσωματώνουμε μια τρίτη κλάση που περιέχει θορύβους από το υποθαλάσσιο περιβάλλον -που συνδυάζουν ήχους από μηχανή πλοίου ή και πιο ήσυχα περιβάλλοντα - αλλά απουσιάζουν βιοσήματα. Προσαρμόσαμε αυτή τη φορά ένα βαθύ νευρωνικό δίκτυο ResNet-LSTM και παρατηρήσαμε ότι μετά από λίγες εποχές εκπαίδευσης - περίπου 40- συγκλίνει σε λύση του προβλήματος. Στο παρακάτω σχήμα αποτυπώνεται η εξέλιξη της απόδοσης και του σφάλματος ανά εποχή τόσο του training οσο και του validation set. Στο παρακάτω γράφημα φαίνεται η εξέλιξη ανά εποχή της απόδοσης και του σφάλματος τόσο του training οσο και του validation set.

Σχήμα 6.9: Απόδοση και σφάλμα ενός δικτύου ResNet-LSTM



Ακολουθεί ο πίνακας σύγχυσης confusion matrix του δικτύου ResNet-LSTM για ένα άγνωστο σύνολο όπου φαίνεται πως όλα τα άγνωστα ηχητικά αρχεία έχουν ταξινομηθεί σε σωστές κλάσεις.

Σχήμα 6.10: Confusion Matrix σε ένα πείραμα 3 ηχητικών κλάσεων



Σε μια σειρά από βαθιά δίκτυα που αναπτύξαμε παρατηρήσαμε επίσης υψηλή απόδοση και ικανότητα γενίκευσης σε άγνωστο σύνολο δεδομένων (test set). Τα παραδοσιακά δίκτυα μηχανικής μάθησης επίσης διατηρούν πολύ υψηλή απόδοση, η απουσία βιοσήματος στο φάσμα και η παρουσία ανθρωπογενών ήχων ή θορύβων περιβάλλοντος δεν επιβαρύνει τελικά την απόδοση των δικτύων παραδοσιακών ή βαθιών. Συμπεραίνουμε ότι και σε αυτό το πρόβλημα η επιλογή βέλτιστου δικτύου μπορεί να γίνει με βάση την ελάχιστη υπολογιστική πολυπλοκότητα η οποία καθορίζεται από τον αριθμό των παραμέτρων που πρέπει να εκπαιδευτούν.

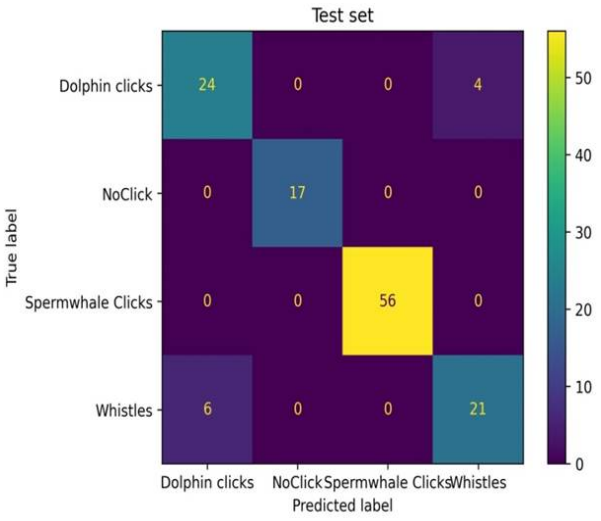
6.8.1 Σχεδιασμός πειράματος με αλληλεπικαλυπτόμενα βιοσήματα

Η επιτυχία της πλειονότητας των δικτύων στο πείραμα με τρεις κλάσεις οδηγεί στο σχεδιασμό ενός τελευταίου πειράματος που συνοδεύεται από αύξηση της πολυπλοκότητας στο χώρο εισόδου. Συγκεκριμένα διαιρούμε την ηχητική κατηγορία των βιοσημάτων δελφινιών σε δύο μερικώς αλληλεπικαλυπτόμενες υποκλάσεις που αναπαριστούν η πρώτη clicks δελφινιών και η δεύτερη ταυτόχρονα whistles/buzzes και clicks δελφινιών. Από το σύνολο των 284 βιοσημάτων ζωνοδέλφινων συνολικά 135 αρχεία χαρακτηρίστηκαν ως ζωνοδέλφινα clicks και 149 whistles ή clicks. Στον πίνακα αποτελεσμάτων παρουσιάζονται οι μετρικές της απόδοσης των δικτύων που υλοποιήθηκαν: accuracy και precision.

Πίνακας 6.2: Απόδοση διαφορετικών αρχιτεκτονικών

Μοντέλο	**Αποτελέσματα σε άγνωστο set**		
	Parameters	***Accuracy***	***Precision***
MFCC-SVM (RBF kernel)	-	83.0%	73.4%
MFCC-kNN	-	75.45%	73.4%
ResNet	1.0M	87.0%	84.7%
ResNet-LSTM	9.77M	91.3%	89.9%
ResNet-BiLSTM	18.5M	90.1%	89.1%
ResNet-GRU	7.6M	90.9%	89.8%
ResNet-BiGRU	14.2M	88.7%	88.0%
ResNet-LSTM-Attention	9.8M	90.4%	89.9%
Parallel ResNet-LSTM	8.2M	89.2%	88.6%

Σχήμα 6.11: Αποτελέσματα ενός από τα πειράματα δικτύου ResNet-LSTM σε άγνωστο set



Από τα παραπάνω αποτελέσματα συμπεραίνουμε ότι:

α) Τα βασικά μοντέλα βαθιάς μάθησης (ResNet) υπεραποδίδουν σε σχέση με τα παραδοσιακά δίκτυα μηχανικής μάθησης. Πραγματικά, διαπιστώνεται καταρχάς μια υποβάθμιση της απόδοσης των παραδοσιακών δικτύων σε σχέση με εκείνες που είχαν εμφανίσει σε πειράματα 2 και 3 κλάσεων. Επίσης τα βασικά δίκτυα βαθιάς μάθησης ResNet παρουσιάζουν υψηλότερο accuracy κατά 4% από τον SVM ταξινομητή.

β) Τα υβριδικά νευρωνικά δίκτυα επιτυγχάνουν υψηλότερες αποδόσεις από ότι τα βασικά βαθιά μοντέλα ResNet ή LSTM. Πραγματικά τόσο η απόδοση όσο και το precision παρουσιάζονται βελτιωμένα στα υβριδικά δίκτυα με κόστος την αύξηση του αριθμού των παραμέτρων και της πολυπλοκότητας του δικτύου.

γ) Τα δίκτυα που είναι δικατευθυντικά (bidirectional) δεν βελτιώνουν την απόδοση. Το αποτέλεσμα αυτό μας οδηγεί στο συμπέρασμα πως η έξοδος του δικτύου δεν εξαρτάται από μελλοντικά inputs. Θα ήταν λάθος να γενικεύσουμε όμως ένα τέτοιο συμπέρασμα σε αρχεία τόσο βραχείας διάρκειας και σωστό θα ήταν να ελεγχθεί σε νέα δεδομένα ή σε αντίστοιχα προβλήματα βιοακουστικής.

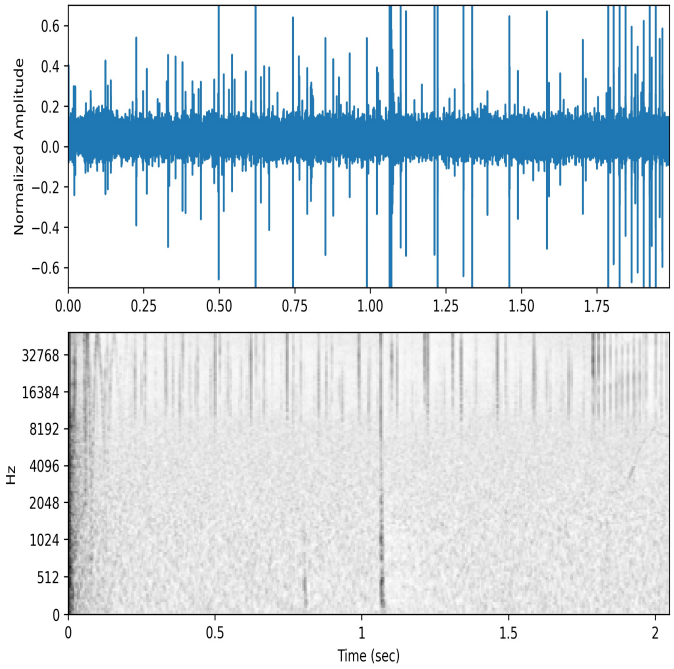
δ) Όλες οι αρχιτεκτονικές κατάφεραν να επιλύσουν ένα -μεταξύ των ειδών- πρόβλημα ταξινόμησης ηχητικών αρχείων με 2 ή 3 κλάσεις ενώ υβριδικές αρχιτεκτονικές απέδειξαν συγκριτικά πλεονεκτήματα στη διάκριση μερικώς αλληλεπικαλυπτόμενων προτύπων. Σε απλά προβλήματα βιοακουστικής όπου ο χώρος εισόδου δεν χαρακτηρίζεται από εμφανή τουλάχιστον αλληλεπικάλυψη των βιοσημάτων απλοί εξαγωγείς χαρακτηριστικών με ταξινομητές μηχανικής μάθησης συνιστούν βέλτιστες επιλογές λόγω της χαμηλής τους πολυπλοκότητας.

ε) Επεκτείνοντας το τέταρτο συμπέρασμα, η επιτυχία των υβριδικών δικτύων έναντι των δικτύων βάσης εκφράζει ότι η επεξεργασία που επιτελούν τα LSTM είναι συμπληρωματική ως προς τα χρονοσυχνοτικά χαρακτηριστικά που εξάγουν τα ResNet παρά το γεγονός ότι τα φασματογραφήματα εμπεριέχουν ήδη χρονοσυχνοτική πληροφορία. Το χρονικό dependency το οποίο εξάγει το αναδρομικό δίκτυο αναβαθμίζει την πληροφορία του διανύσματος χαρακτηριστικών του συνελικτικού δικτύου συμβάλλοντας στην αύξηση της απόδοσης.

Στον πίνακα 6.11 που παραθέτει αποτελέσματα από ένα πείραμα της 5-fold διαδικασίας, διαπιστώνεται πως το δίκτυο ταξινομεί ήχους στα σωστά είδη ενώ υπάρχουν 6 whistles and clicks που καταγράφηκαν σαν clicks και 4 clicks που αναγνωρίστηκαν σαν clicks and whistles. Αποτιμώντας την αιτία των λανθασμένων αποτελεσμάτων του δικτύου παρατηρούμε ότι στα αρχεία που ανήκουν στην κατηγορία clicks and whistles αλλά αναγνωρίζονται σαν clicks από το νευρωνικό δίκτυο η συνιστώσα του συνεχούς σήματος στο φάσμα είναι γενικά ασθενούς έντασης με αποτέλεσμα να κυριαρχεί ακουστικά και γραφικά το αποτύπωμα των clicks. Συγκεκριμένα, η παρουσία των συνεχών σημάτων (whistles) σε αυτά είναι αμυδρή, καταλαμβάνει πολύ μικρή

έκταση του ηχητικού αρχείου ή και τα δύο. Σε μερικές περιπτώσεις ο χαρακτηρισμός ενός αρχείου ως συνδυασμός ηχητικών clicks και whistles είναι οριακός. Για το λόγο αυτό στη συγκεκριμένη κατηγορία η πιθανότητα να χαρακτηριστεί ένα αρχείο σαν click ενώ σε αυτό βρίσκεται και whistle αποκλίνει ισχυρά από τις πιθανότητες αναγνώρισης αρχείων που χαρακτηρίζονται με σωστό label.

Σχήμα 6.12: Επισημειωμένο ως click and whistle αναγνωριστέο σαν click



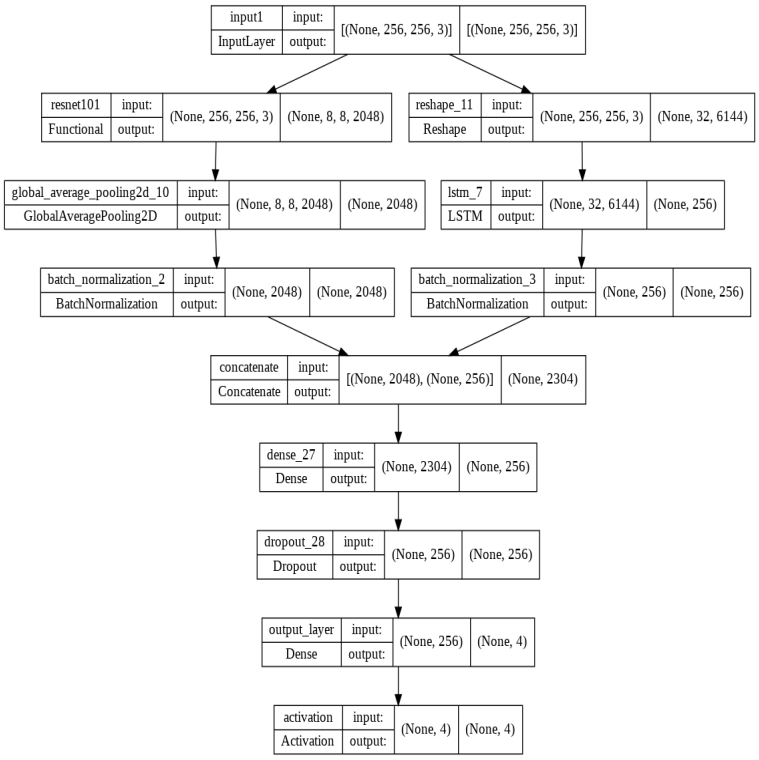
Στο σχήμα 6.12 φαίνεται ένα αρχείο το οποίο έχει επισημειωθεί ως click and whistle ενώ έχει αναγνωριστεί από το νευρωνικό δίκτυο ως click. Παρατηρούμε ότι μόνο στα τελευταία 0.25 sec του αρχείου που διαρκεί μόλις 2 sec υπάρχει το αποτύπωμα ενός αμυδρού στο φάσμα whistle. Το αποτύπωμα των παλμικών ήχων καλύπτει τα πρώτα 1.75 sec και είναι κυρίαρχο στο τελευταίο κομμάτι του ήχου. Συνεπώς τα ποσοστά απόδοσης που παρουσιάζονται εδώ θα έπρεπε να σχετικοποιηθούν αφού οι κυματομορφές και τα παραγόμενα φάσματογραφήματα εξαρτώνται από τη σχετική θέση του αισθητήρα ως προς το κήτος, το σχετικό χρονικό διάστημα εκπομπής και τη σχετική ένταση των παλμικών ή των ταυτόχρονων παλμικών και συνεχών ήχων.

6.9 Σχεδιασμός αρχιτεκτονικών και οπτικοποίηση features με τεχνικές PCA και t-SNE

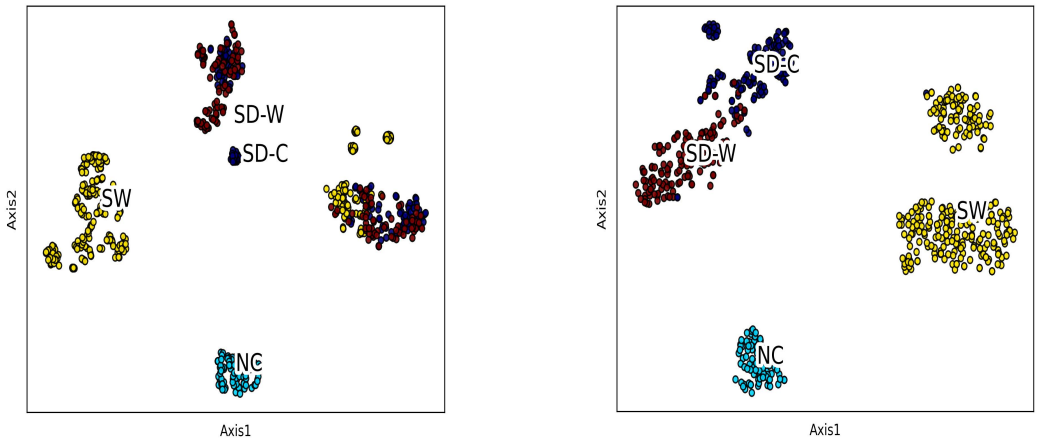
6.9.1 Οπτικοποίηση διανυσμάτων χαρακτηριστικών με τεχνικές PCA και t-SNE

Στη μελέτη αναπτύξαμε μια σειρά από διαφορετικές αρχιτεκτονικές για να δοκιμάσουμε πλεονεκτήματα και μειονεκτήματα κάθε μοντέλου. Στο παρακάτω σχήμα (6.13) παραθέτουμε για παράδειγμα μια παράλληλη αρχιτεκτονική που επίσης χρησιμοποιήθηκε της οποίας όμως η απόδοση ήταν μικρότερη από αυτή της σειριακής αρχιτεκτονικής. Η είσοδος με τα φασματογραφήματα στην περίπτωση αυτή τροφοδοτεί παράλληλα ένα δίκτυο ResNet και ένα LSTM δίκτυο ενώ τα features που εξάγονται από τα δύο δίκτυα συνενώνονται σε ένα ενιαίο διάνυσμα χαρακτηριστικών το οποίο στη συνέχεια τροφοδοτείται σε ένα ταξινομητή.

Σχήμα 6.13: Παράλληλη αρχιτεκτονική ResNet-LSTM



Ο αρχιτεκτονικός σχεδιασμός χαρακτηρίζεται από αναρίθμητο πρακτικά πλήθος διαφορετικών επιλογών και είναι δύσκολο να επιχειρηματολογήσει κανείς για την καταλληλότητα ενός μοντέλου σε όρους διαφορετικούς από αυτούς της απόδοσης και την υπολογιστικής πολυπλοκότητας. Για το λόγο αυτό στη μελέτη δίνουμε έμφαση στην οπτικοποίηση των χαρακτηριστικών που εξάγονται από διαφορετικά δίκτυα σε χώρους χαμηλής διάστασης ως σχετικά μέτρα της αποτελεσματικότητας διαφορετικών αρχιτεκτονικών. Στο πλαίσιο της προτεινόμενης αρχιτεκτονικής συνελικτικού αναδρομικού νευρωνικού δικτύου αφαιρούμε το στρώμα ταξινόμησης



(α) Cepstral features (β) RCNN features

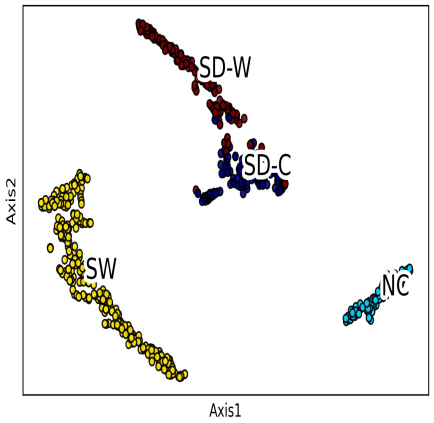
Σχήμα 6.14: Υλοποίηση τεχνικής PCA και t-SNE για την οπτικοποίηση ενός διανύσματος cepstral χαρακτηριστικών και ενός διανύσματος βαθιών χαρακτηριστικών deep features με τις τέσσερις κλάσεις να αναπαριστούν: SW: clicks φυσητήρων, NC: Απουσία βιοσήματος, SD: clicks : Clicks ζωνοδέλφινων, SD-W: whistles Ζωνοδέλφινων.

και εξάγουμε το διάνυσμα χαρακτηριστικών, στη συνέχεια εφαρμόζουμε PCA για να μειώσουμε τη διάσταση του χώρου χαρακτηριστικών από 256 σε 10 και μειώνουμε περαιτέρω τη διάσταση από 10 σε 2 με την τεχνική t-SNE. Τέλος οπτικοποιούμε στο Ευκλείδιο επίπεδο τα εξαγόμενα χαρακτηριστικά. Τα αποτελέσματα απεικονίζονται στο σχήμα 6.5.

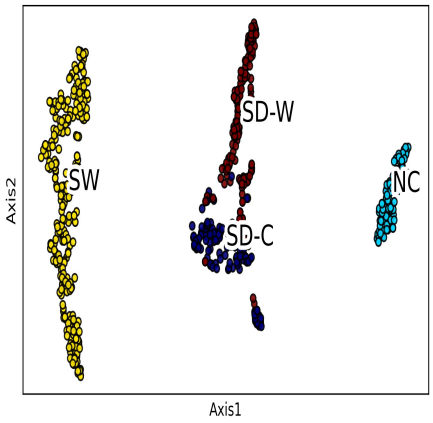
Παρατηρούμε ότι η τεχνική βαθιάς μάθησης αναπτύσσει βελτιωμένη ικανότητα να διαχωρίσει το χώρο χαρακτηριστικών σε σχέση με την παραδοσιακή τεχνική μηχανικής μάθησης που ταξινομεί cepstral χαρακτηριστικά με SVM classifiers. Διαπιστώνουμε πως το υβριδικό δίκτυο ομαδοποιεί επιτυχημένα διαφορετικές κλάσεις σε διακριτές τοποθεσίες ενώ στην περίπτωση των αλληλεπικαλυπτόμενων υποκλάσεων διακρίνει τα δεδομένα διαχέοντας τα αρκετά διακριτά στον ίδιο χώρο. Στο πλαίσιο της έρευνας που έγινε για την επιλογή βέλτιστης αρχιτεκτονικής, η συστηματική χρήση των τεχνικών PCA και t-SNE λειτούργησε βοηθητικά ως κριτηρίο για την επιλογή παραμέτρων/στρωμάτων/δικτύων πριν την επικύρωση των βέλτιστων αρχιτεκτονικών.

6.9.2 Προβολή διανύσματος χαρακτηριστικών δίκτυου ResNet Bidirectional LSTM σε Ευκλείδιο επίπεδο συναρτήσει της μείωσης διάστασης με P.C.A.

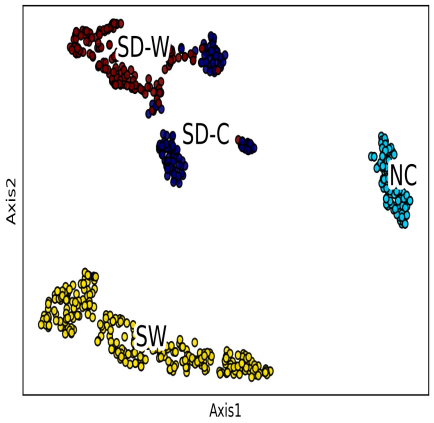
Στην υποενότητα αυτή θα σχολιάσουμε τη συμπεριφορά ενός υβριδικού δικτύου ResNet-Bidirectional LSTM καθώς συμπίεζουμε σε διαφορετική διάσταση με PCA το διάνυσμα των χαρακτηριστικών πριν εφαρμόσουμε την τεχνική t-SNE. Στα παρακάτω σχήματα απεικονίστηκε μια έκδοση του διανύσματος χαρακτηριστικών μετά από συμπίεση PCA μεταβλητής διάστασης και εφαρμογή της t-SNE στο συμπιεσμένο διάνυσμα χαρακτηριστικών.



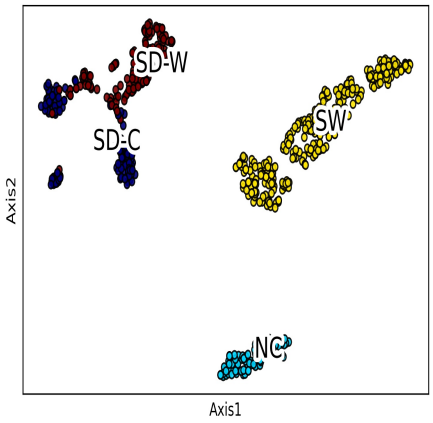
(α) PCA dimension = 3



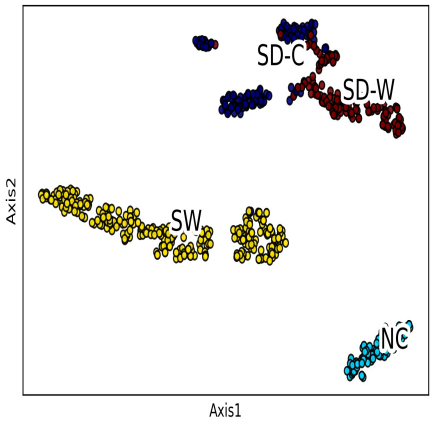
(β) PCA dimension = 5



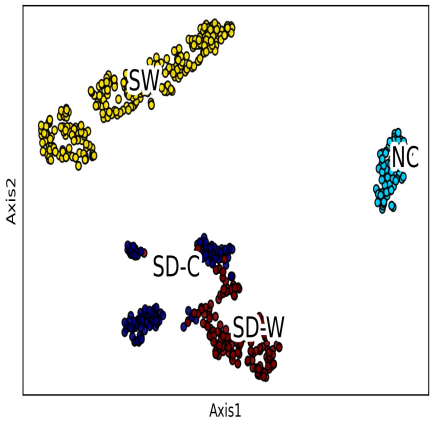
(α) PCA dimension = 10



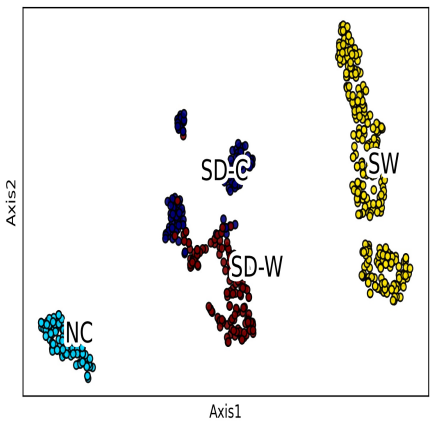
(β) PCA dimension = 15



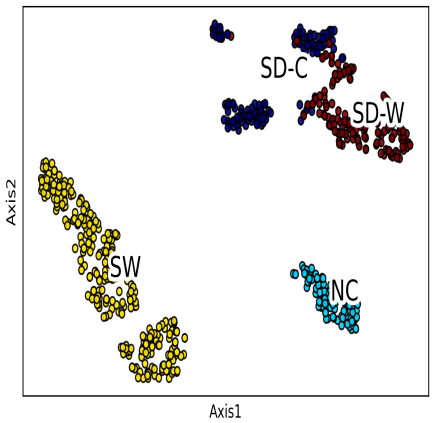
(α) PCA dimension = 20



(β) PCA dimension = 30



(α) PCA dimension = 40



(β) PCA dimension = 50

Παρατηρούμε από τα παραπάνω σχήματα ότι όταν η διάσταση PCA μειώνεται άρα διατηρούμε μικρότερη μεταβλητότητα στο συμπιεσμένο διάνυσμα χαρακτηριστικών τόσο πιο επιμήκεις είναι οι γεωμετρικές δομές των clusters των ηχητικών κλάσεων. Επίσης παρατηρούμε ότι όσο μικραίνει η διάσταση PCA τόσο περισσότερο ομογενοποιείται η κλάση χαρακτηριστικών SW. Συγκεκριμένα, παρατηρείται για διάστασεις PCA=50, 40, 30, 20, 15 μια κλάση SW με δύο διακριτές γεωμετρικές υποομάδες ενώ για μικρότερες διαστάσεις: PCA=10, 5, 3 φαίνεται να υπάρχει συγχώνευση των δύο κλάσεων σε μία. Η γεωμετρική αυτή διαπίστωση εξηγείται από το γεγονός ότι τα ηχητικά σήματα φυσητήρων έχουν πραγματοποιηθεί με διαφορετικά υδρόφωνα, διαφορετικών συχνοτήτων δειγματοληψίας σε διαφορετικές γεωγραφικές περιοχές σχηματίζοντας τελικά δυο γεωμετρικές υπο-ομάδες που με την περαιτέρω συμπύκνωση συγχωνεύονται αφού το διάνυσμα χαρακτηριστικών προβάλλεται σε λιγότερες διαστάσεις.

Ένα συμπέρασμα που δεν πρέπει να περάσει απαρατήρητο είναι η δυνατότητα των δικτύων να αναγνωρίζουν τελικά πρότυπα υπερβαίνοντας χαρακτηριστικά που δεν σχετίζονται άμεσα με τα βιοακουστικά features. Η μηχανή μαθαίνει τελικά πρότυπα που σχετίζονται όχι με τις ιδιότητες των οργάνων και τον περιβάλλοντα θόρυβο αλλά κυρίως με τη φύση των βιοσημάτων. Έτσι τα ηχητικά ζωνοδέλφινων και τα ηχητικά φυσητήρων που έχουν συλλεχθεί από τα ίδια υδρόφωνα στην Πύλο με συχνότητα δειγματοληψίας 100 kHz διαχωρίζονται πολύ ικανοποιητικά σε διακριτές και απομακρυσμένες περιοχές του Ευκλείδειου χάρτη ενώ ηχητικά φυσητήρων από όργανα διαφορετικής δειγματοληψίας σε Σούγια και Πύλο συγκλίνουν τελικά σε ένα cluster.

6.10 Συμπεράσματα - Συζήτηση

Στη μελέτη που προηγήθηκε, ένα υβριδικό CRNN δίκτυο αποτελούμενο από ένα δίκτυο ResNet και διάφορες RNN εκδοχές αρχιτεκτονικών προτείνονται για την ταξινόμηση βιοσημάτων οδοντοκητών που έχουν συλλεχθεί στην Ελληνική Τάφρο από δύο είδη οδοντοκητών, φυσητήρες και ζωνοδέλφινα. Κάθε μια από τις αρχιτεκτονικές που περιγράφηκαν παραπάνω επιτυγχάνει να επιλύσει το πρόβλημα ταξινόμησης σημάτων σε τρεις κλάσεις επιβεβαιώνοντας ευρήματα ερευνών για την αποτελεσματικότητα αρχιτεκτονικών επιβλεπόμενης μάθησης στην επίλυση βιοακουστικών προβλημάτων αναγνώρισης. Η κύρια όμως συνεισφορά της έρευνας είναι ότι στα πλαίσια ενός περισσότερο σύνθετου προβλήματος, όπου η ακουστική κατηγορία των εκφωνήσεων δελφινιών διαιρείται περαιτέρω σε δύο μερικώς αλληλεπικαλυπτόμενες υποκλάσεις, σαφή πλεονεκτήματα καταγράφονται στη χρησιμοποίηση υβριδικών τεχνικών βαθιάς μάθησης σε σχέση τόσο με παραδοσιακές τεχνικές μηχανικής μάθησης όσο και σε σχέση με μοντέλα βάσης ResNet, LSTM ή και παράλληλες αρχιτεκτονικές.

Κεφάλαιο 7

Επίλογος

7.1 Μελλοντικές επεκτάσεις

Στη μελέτη που προηγήθηκε αναπτύξαμε μια σειρά από διαφορετικές αρχιτεκτονικές βαθιών δικτύων και συγκρίναμε τις μεταξύ τους αποδόσεις σε ένα πρόβλημα βιοακουστικής. Δοκιμάσαμε επίσης την επίδραση διαφορετικών μορφών φασματογραφημάτων στα αποτελέσματα χρησιμοποιώντας είτε φασματογραφήματα διαφορετικής κλίμακας (γραμμικά, λογαριθμικά, mel) είτε scalograms. Διαπιστώσαμε ότι η εφαρμογή φίλτρων PCEN μειώνει το σφάλμα εκπαίδευσης του δικτύου τονίζοντας τα μεταβατικά βιοακουστικά events έναντι του background θορύβου. Στη συνέχεια συγκρίναμε τεχνικές παραδοσιακής μάθησης που χρησιμοποιούν ταξινομητές SVM, KNN σε cepstral χαρακτηριστικά παρατηρώντας υψηλές αποδόσεις σε πιο απλά προβλήματα ταξινόμησης. Σε πιο σύνθετα προβλήματα αναδείχτηκε το πλεονέκτημα της χρησιμοποίησης υβριδικών δικτύων βαθιάς μάθησης. Η βασική εναλλακτική στη χρήση χρονοσυχνοτικών αναπαραστάσεων βρίσκει έκφραση στην απευθείας χρήση κυματομορφών στο στρώμα εισόδου νευρωνικών δικτύων. Σημαντικό πλεονέκτημα προς μια τέτοια κατεύθυνση είναι η αποφυγή μιας διαδικασίας βημάτων προεπεξεργασίας για τη δημιουργία φασματογραφημάτων και της απομάκρυνσης θορύβου από αυτά. Μια επέκταση σε παρόμοια κατεύθυνση προκρίνεται και από το γεγονός ότι σε δημοσιευμένες έρευνες δίκτυα που χρησιμοποιούν εισόδους κυματομορφών επιτυγχάνουν αποδόσεις συγκρίσιμες με αυτές των 2Δ εισόδων.

7.2 Για την προστασία της Ελληνικής Τάφρου

Η Ελληνική Τάφρος φιλοξενεί στις υποθαλάσσιες χαράδρες και στους γκρεμούς της ένα από τα σπουδαιότερα καταφύγια κητωδών στη Μεσόγειο. Είναι αξιοσημείωτο ότι τμήμα της Τάφρου επίσημα αναγνωρίζεται ως σημαντική περιοχή για θαλάσσια θηλαστικά (Important Marine Mammal Area). Η διάσταση αυτή έχει ιδιαίτερη σημασία αν αναλογιστεί κανείς ότι οι φυσητήρες της Μεσογείου όπως και κάποια είδη δελφινιών αξιολογούνται ως κινδυνεύοντα είδη (Endangered) ενώ τα ζωνοδέλφινα αξιολογούνται ως ευάλωτα (Vulnerable). Οι σημαντικότεροι κίνδυνοι που αντιμετωπίζουν τα θαλάσσια θηλαστικά είναι κατά βάση ανθρωπογενείς. Οι εκβρασμοί κητωδών συχνά αποδίδονται σε συγκρούσεις τους με ταχύπλοα ή γρήγορα επι-

βατηγά πλοία ενώ ο ανθρωπογενής υποθαλάσσιος θόρυβος επίσης συνιστά πιθανή απειλή οι επιπτώσεις του οποίου δεν έχουν ακόμη προσδιοριστεί επαρκώς. Νομίζουμε ότι η παρούσα έρευνα θα μπορούσε να εμπλουτιστεί σε υλικό πεδίου ώστε να συμπεριλάβει ήχους περισσότερων κητωδών της Ελληνικής Τάφρου και της Μεσογείου. Θα ήταν δυνατή τότε η ανάπτυξη ενός εργαλείου αναγνώρισης κητωδών ικανό να ανιχνεύσει σε πραγματικό χρόνο την παρουσία θηλαστικών, να χαρτογραφήσει τις συνήθειες τους και να συμβάλλει στην προστασία της θαλάσσιας βιοποικιλότητας.

Bibliography

[1] Magnus Wahlberg, Alexandros Frantzis, Paraskevi Alexiadou, Peter Madsen, and Berte Møhl, "Click production during breathing in a sperm whale (physeter macrocephalus) (l)," *The Journal of the Acoustical Society of America*, vol. 118, 2006.

[2] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, vol. 8, 1991.

[3] Frantzis A. Rendell L., "Mediterranean sperm whales, physeter macrocephalus: The precarious state of a lost tribe.," *Advances in Marine Biology*, vol. 75, 2016.

[4] Andre et al., "Sperm whale long-range echolocation sounds revealed by antares, a deep-sea neutrino telescope.," *Scientific Reports, Nature*, vol. 7, 2017.

[5] P Madsen, Roger Payne, N Kristiansen, Magnus Wahlberg, Iain Kerr, and B Møhl, "Sperm whale sound production studied with ultrasound time/depth-recording tags," *The Journal of experimental biology*, vol. 205, pp. 1899–906, 2002.

[6] Peter Bermant, Michael Bronstein, Robert Wood, Shane Gero, and David Gruber, "Deep machine learning techniques for the detection and classification of sperm whale bioacoustics," *Scientific Reports*, vol. 9, 2019.

[7] L.D. Landau and Lifshitz E.M., *Fluid Mechanics - 2nd ed. Course of theoretical physics*, Pergamon Press, 1959.

[8] L. Rabiner and R Schafer, *Theory and Applications of Digital Speech Processing*, Pearson, 2010.

[9] Sands M. Feynman R, Leighton R.B., *Feynman Lectures on Physics*, Addison-Wesley Publishing Company, 1963.

[10] Crawford F.S., *Waves, Berkeley physics course - volume 3*, McGraw-Hill Book Company, 1968.

[11] Kapitza P., "Viscosity of liquid helium below the lamda points," *Nature*, 1938.

[12] Landau L., "Theory of the superfluidity of helium ii," *Phys. Rev., American Physical Society*, 1941.

[13] Finn B. Jensen, William A. Kuperman, Michael B. Porter, and Henrik Schmidt, *Computational Ocean Acoustics*, Springer Publishing Company, Incorporated, 2011.

[14] Stefan Huggenberger, Michel Andre, and Helmut H. A. Oelschlager, "The nose of the sperm whale: overviews of functional design, structural homologies and evolution," *Journal of the Marine Biological Association of the United Kingdom*, vol. 96, 2014.

[15] B.Mohl, "Sound transmission in the nose of the sperm whale physeter catodon. a post mortem study.," *J.Comp Physiol A*, 2001.

[16] Rendell L. Gero S, Whitehead H, "Individual, unit and vocal clan level identity cues in sperm whale codas.," *R Soc Open Sci.*, 2016.

[17] A. Frantzis and P. Alexiadou, "Male sperm whale (physeter macrocephalus) coda production and coda-type usage depend on the presence of conspecifics and the behavioural context.," *Canadian Journal of Zoology*, 2008.

[18] Norris KS. Cranford TW, Amundin M, "Functional morphology and homology in the odontocete nasal complex: implications for sound generation.," *J Morphol.*, 1996.

[19] Walter Zimmer, Peter Tyack, Mark Johnson, and Peter Madsen, "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," *The Journal of the Acoustical Society of America*, vol. 117, 2005.

[20] Annalisa Berta, Eric Ekdale, and Ted Cranford, "Review of the cetacean nose: Form, function, and evolution," *The Anatomical Record*, vol. 297, 2014.

[21] Varvara Kandia and Y. Stylianou, "Detection of sperm whale clicks based on the teager–kaiser energy operator," *Applied Acoustics*, vol. 67, 2006.

[22] Julie Oswald, Shannon Rankin, Jay Barlow, and Marc Lammers, "A tool for real-time acoustic species identification of delphinid whistles," *The Journal of the Acoustical Society of America*, vol. 122, 2007.

[23] R. Gonzalez and R. Woods, *Digital Image Processing*, Pearson, 2017.

[24] R. Policar, "The story of wavelets," *IMACS/IEEE CSCC'99 Proceedings*, 1999.

[25] Alan V. Oppenheim, Alan S. Willsky, and S. Hamid Nawab, *Signals and Systems*, Prentice-Hall, 1996.

[26] Szeliski R., *Computer Vision, Algorithms and Applications*, Springer-Verlag London Limited, 2011.

[27] S. Haykin, *Neural Networks and Learning Machines*, Pearson International Edition, 1999.

[28] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016.

[29] V. Vapnik, "Complete statistical theory of learning," *Automation and Remote Control*, vol. 80, 2019.

[30] Vapnik V., *The Nature of Statistical Learning Theory*, Springer-Verlag New York, Inc., 1995.

[31] Bernhard Schlkopf, Alexander J. Smola, and Francis Bach, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, The MIT Press, 2018.

[32] Laurens van der Maaten and Geoffrey Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[33] D. Makropoulos, A. Tsiami, A. Prospathopoulos, D. Kassis, A. Frantzis, E. Skarsoulis, and P. Maragos, "Deep learning techniques for the detection and classification of sperm whale and striped dolphin bioacoustic patterns," Proc. Mar. and Inl. Wat.Res.Symp., 2022.

[34] D. Makropoulos, A. Tsiami, A. Prospathopoulos, D. Kassis, A. Frantzis, E. Skarsoulis, G. Piperakis, and P. Maragos, "Convolutional recurrent neural networks for the classification of cetacean bioacoustic patterns," Submitted for IEEE International Conference on Acoustics, Speech and Signal Processing in October 2022.

[35] M. Roch, M. Soldevilla, R. Hoenigman, S. Wiggins, and J. A. Hildebrand, "Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes.," *Canadian Acoustics*, vol. 36, 2008.

[36] Nousek-McGregor A. Brown J.C., Smaragdis P, "Automatic identification of individual killer whales," *The Journal of the Acoustical Society of America*, vol. 128, 2010.

[37] A. Walker, R. B. Fisher, and N. Mitsakakis, "Singing maps: classification of whale-song units using a self-organizing feature mapping algorithm," *DAI Research Paper*, 1996.

[38] Ann N. Allen, Matt Harvey, Lauren Harrell, Aren Jansen, Karlina P. Merkens, Carrie C. Wall, Julie Cattiau, and Erin M. Oleson, "A convolutional neural network for automated detection of humpback whale song in a diverse, long-term passive acoustic dataset," *Frontiers in Marine Science*, vol. 8, 2021.

[39] M.N. Nystuen, J. amd Anagnostou, E.N. Anagnostou, and Papadopoulos A., "Monitoring greek seas using passive underwater acoustics," *Journal of Atmospheric and Oceanic Technology*, vol. 32, no. 2, pp. 334–349, 2015.

[40] A. Frantzis, P. Alexiadou, and K.C. Gkikopoulou, "Sperm whale occurrence, site fidelity and population structure along the hellenic trench (greece, mediterranean sea).," *Aquatic Conserv: Mar. Freshw. Ecosyst.,*, vol. 24, pp. 83–102, 2014.

[41] Emmanuel Skarsoulis, George Piperakis, Emmanuel Orfanakis, Panagiotis Papadakis, Despoina Pavlidi, Michael Kalogerakis, Paraskevi Alexiadou, and Alexandros Frantzis, "A real-time acoustic observatory for sperm-whale localization in the eastern mediterranean sea," *Frontiers in Marine Science*, vol. 9, 2022.

[42] Vincent Lostanlen, Justin Salamon, Mark Cartwright, Brian Mcfee, Andrew Farnsworth, Steve Kelling, and Juan Bello, "Per-channel energy normalization: Why and how," *IEEE Signal Processing Letters*, 2018.