

Χαρακτηρισμός υποθαλάσσιων ακουστικών σημάτων με χρήση βίο-μαθηματικού μοντέλου των ψυχο-ακουστικών μηχανισμών των Μεγάπττερων φαλαινών.

Κώστας Σμαραγδάκης
Υποψήφιος διδάκτορας, Πανεπιστήμιο Κρήτης & ΙΤΕ/ΙΥΜ
Ν. Πλαστήρα 100, 70013, Ηράκλειο Κρήτης
kesmarag @ tem.uoc.gr

Μιχάλης Ταρουδάκης
Καθηγητής, Πανεπιστήμιο Κρήτης & ΙΤΕ/ΙΥΜ
Ν.Πλαστήρα 100, 70013, Ηράκλειο Κρήτης
taroud @ math.uoc.gr

ΠΕΡΙΛΗΨΗ

Στην εργασία παρουσιάζεται η αρχή μιας προσπάθειας χαρακτηρισμού τυπικών υποβρύχιων ακουστικών σημάτων χαμηλής συχνότητας χρησιμοποιώντας κατάλληλο βίο-μαθηματικό μοντέλο το οποίο μιμείται τους ψυχο-ακουστικούς μηχανισμούς του ακουστικού νεύρου των Μεγάπττερων φαλαινών (*Humpback whales*). Το ακουστικό μοντέλο που χρησιμοποιείται, αντιστοιχίζει ένα ακουστικό σήμα με μια αναπαράσταση (*visual representation*) η οποία περιέχει όλη την πληροφορία του σήματος και παράλληλα δίνει έμφαση σε κρίσιμα χαρακτηριστικά του. Για να ληφθεί αυτή η πληροφορία γίνεται φιλτράρισμα του σήματος σύμφωνα με την μορφολογία και τα χαρακτηριστικά του αυτιού της φάλαινας (*outer-middle ear filtering*), φασματική ανάλυση (*spectral analysis*) καθώς και μοντελοποίηση της κωδικοποίησης που λαμβάνει χώρα στους νευρώνες στους οποίους καταλήγει το ακουστικό της νεύρο (*neural encoding*).

Characterization of underwater acoustic signals, using a bio-mathematical model of the psycho-acoustic mechanisms of Humpback whales.

ABSTRACT

The paper presents a preliminary stage of a method for the characterization of typical low frequency acoustic signals using an appropriate biomimetic toolbox of psychoacoustic functions of Humpback whales. The acoustic model presented here corresponds to a proper visual representation which includes the

whole signal information and gives the critical feature data. The above visual representation is capable to simulate the way that hearing system of Humpback whale is stimulated. In order to achieve the appropriate feature extraction we model all hearing functions of whales, starting from the transformation in the outer and middle parts of the ear (outer-middle ear filtering), followed by a model for the passage within the inner part (spectral analysis) and ending with the modeling of the coding in the brain (neural encoding).

Εισαγωγή

Υπάρχουν σημαντικές ομοιότητες ανάμεσα στους μηχανισμούς παραγωγής και αντίληψης ηχητικών σημάτων από τις φάλαινες με τους αντίστοιχους που συναντάμε στους ανθρώπους και γενικά στα θηλαστικά. Επιπλέον η μορφολογία των ακουστικών οργάνων των Μεγάπττερων φαλαινών, τους εξασφαλίζουν ένα μεγαλύτερο εύρος απόκρισης συχνοτήτων και μια δυνατότητα διάκρισης μεταξύ των ήχων πολύ μεγαλύτερη από τη δική μας. Σε αυτό σημαντική επίδραση φαίνεται να έχει η πυκνότερη κατανομή ακουστικών νεύρων σε σχέση με τα χερσαία θηλαστικά.

Παρά το γεγονός ότι η φύση των ήχων που παράγουν και ο τρόπος που τους επεξεργάζονται παραμένει μυστήριο έχει αποδειχθεί ότι οι Μεγάπτερες φάλαινες μπορούν να ακούσουν ήχους πολύ χαμηλής συχνότητας της τάξης των 30 Hz έως και πολύ υψηλής 20 – 30 kHz [1]. Η βασική παραδοχή στην προηγούμενη πρόταση είναι ότι όλα τα είδη μπορούν να ακούν τουλάχιστον το εύρος των συχνοτήτων που μπορούν να παράγουν. Επίσης ένα σημαντικό γνώρισμα του ακουστικού συστήματος των Μεγάπττερων φαλαινών αποτελεί το γεγονός ότι έχουν στενότερη κρίσιμη ζώνη συχνοτήτων θορύβου, σε σχέση με τα άλλα θηλαστικά, που θα μπορούσε να τις αποπροσανατολίσει και να τους αποκόψει χρήσιμες ακουστικές πληροφορίες.

Στη παρούσα εργασία θα γίνει η θεωρητική παρουσίαση του μαθηματικού μοντέλου που θα εφαρμοστεί πρακτικά σε μελλοντικές εργασίες με κύριο στόχο να μελετηθεί αφενός μεν κατά πόσο μπορούν ήχοι που παράγονται από ανθρώπινη δραστηριότητα είναι ικανοί να αποπροσανατολίσουν τις Μεγάπτερες φάλαινες μπερδεύοντας τις και αφετέρου κατά πόσο μπορούν να διακρίνουν ψυχο-ακουστικά τις διαφορές σχεδόν όμοιων σημάτων. Απώτερος σκοπός είναι να μελετηθεί η κατασκευή ενός βιο-μιμητικού μοντέλου το οποίο θα χαρακτηρίζει επαρκώς ακουστικά σήματα δίνοντας έμφαση σε κρυμμένα χαρακτηριστικά τους, αντικαθιστώντας κλασσικές μεθόδους όπως μετασχηματισμούς Fourier και κυμματιδίων (wavelets), μιμούμενοι τον τρόπο ακουστικής αντίληψης των Μεγάπττερων φαλαινών.

1 Βιομαθηματικό μοντέλο ακουστικού συστήματος των Μεγάπττερων φαλαινών.

Οι διαδικασίες λήψης και επεξεργασίας ήχων από τις Μεγάπτερες φαλαινές πραγματοποιούνται από τρεις διακριτές διαδοχικές διαδικασίες. Σε αυτή την ενότητα θα αντιστοιχίσουμε κάθε μια από τις φυσικές διαδικασίες με ένα κατάλληλο μαθηματικό μοντέλο. Τα τρία αυτά μοντέλα θα προσομοιώνουν όλους τους μετασχηματισμούς των σημάτων από τη λήψη τους έως και την επεξεργασία τους

στους νευρώνες του εγκεφάλου της Μεγάπτερης φάλαινας καθώς και τον τρόπο με τον οποίο γίνεται η αυτόματη συνεχής εκπαίδευση της που έχει ως στόχο την κατηγοριοποίηση των ήχων που λαμβάνει και την άντληση πληροφορίας.

1.1 Μηχανισμός παραγωγής ήχων από τις Μεγάπτερες φάλαινες

Προτού ασχοληθούμε με τη λειτουργία ακοής των Μεγάπτερων φαλαινών κρίνεται απαραίτητο να παρουσιάσουμε μερικά στοιχεία του συστήματός παραγωγής ήχων από τις ίδιες.

Ένα τυπικό μοντέλο χαρακτηρισμού της παραγωγής ήχων στα θηλαστικά αποτελεί αυτό της πηγής - φίλτρου (source-filter model) [2]. Τη πηγή στις φάλαινες αποτελεί ένας μηχανισμός ανατάραξης του αέρα (air turbulence) ενώ το φίλτρο προσομοιώνει το μετασχηματισμό του σήματος που παρήγαγε η πηγή κατά τη μεταφορά του κατά μήκος της φωνητικής οδού (vocal tract). Η μορφολογία της φωνητικής οδού μεταβάλλεται κατά τη διάρκεια της εκπομπής ήχων από τις φάλαινες αλλάζοντας με αυτό τον τρόπο τη μορφή του φίλτρου. Παρόλα αυτά, τα φίλτρα παρουσιάζουν κάποια κοινά χαρακτηριστικά. Συγκεκριμένα παρουσιάζουν ενίσχυση σε p διακριτές συχνότητες οι οποίες ονομάζονται formants. Υπάρχουν ενδείξεις ότι η τιμή του p είναι μεταξύ του 4 και του 6.

Κάθε formant μπορεί να εκφραστεί με ένα διακριτό φίλτρο δεύτερης τάξης που δρα πάνω στη σειρά των παλμών που παράγονται στη πηγή. Έτσι, εάν θέλουμε να περιγράψουμε την απόκριση του φίλτρου της φωνητικής οδού θα χρειαστούμε p φίλτρα δεύτερης τάξης καθώς και ένα φίλτρο πρώτης τάξης που προσομοιώνει την φασματική αποζημίωση (spectral compensation) λόγω σφαλμάτων επικάλυψης μεταξύ των φίλτρων. Συμφωνά με την θεωρία επεξεργασίας σήματος το συνολικό φίλτρο θα είναι τάξης $2p + 1$ και θα ακολουθεί τη σχέση:

$$X(\omega) = \frac{1}{1 - \sum_{k=1}^{2p+1} c_k e^{-i\omega k}} S(\omega) \quad (1)$$

Όπου $X(\omega)$ το φάσμα του ήχου που θα διαδοθεί στο περιβάλλον και $S(\omega)$ η συνάρτηση διέγερσης της πηγής στο πεδίο συχνοτήτων η οποία μπορεί να θεωρηθεί ως σειρά από παλμούς λευκού θορύβου κατά το διάστημα λειτουργίας της. Σημειώνουμε ότι οι συντελεστές c_k προσδιορίζουν πλήρως τη θέση και το πλάτος των formants.

Χαρακτηριστικό γνώρισμα των συντελεστών c_k αποτελεί το γεγονός ότι μεταβάλλονται πολύ αργά με μεταβολές της τάξης των $10ms$ κατά την διαδικασία παραγωγής ήχων από τις Μεγάπτερες φάλαινες. Μάλιστα χαρακτηριστικά στην ανθρώπινη ομιλία παρατηρείται αλλαγή σε χρόνους που σχετίζονται με το χρόνο μετάβασης μεταξύ διαδοχικών συλλαβών. Στο μοντέλο μας θα υποθέσουμε παράθυρο χρόνου ίσο με $50ms$ το οποίο έχει εφαρμοστεί με επιτυχία από τον Mercano [3].

1.2 Γραμμική ανάλυση φάσματος των ήχων που λαμβάνει η Μεγάπτερη φάλαινα.

Σε αυτό το πρώτο στάδιο επεξεργασίας των ήχων, οι φάλαινες αντιλαμβάνονται ήχους από το περιβάλλον πάνω από ένα σκαλοπάτι έντασης και τους επεξεργά-

ζονται σε διαδοχικά στοιχειώδη τμήματα (frames) της τάξης των 50ms.

Στη συνέχεια αξιοποιώντας την μορφολογία του εξωτερικού και μεσαίου τμήματος του αυτού τους καταφέρνουν να ομαλοποιήσουν τους ήχους κάθε τμήματος ήχου έτσι ώστε να ενισχύσουν τα κοινά χαρακτηριστικά που παρουσιάζουν τα σήματα που εκπέμπουν οι ίδιες. Ουσιαστικά αυτό που γίνεται αποτελεί μια πρώτη εξομάλυνση της κυματομορφής ανακτώντας τη μορφή της απόκρισης του φίλτρου της φωνητικής οδού, ενισχύοντας τις ωφέλιμες συχνότητες (formants) και ως συνέπεια τα σήματα να γίνονται απλούστερα, εξαλείφοντας μη ωφέλιμες πολύπλοκες δομές που προέρχονται από θόρυβο.

Η παραπάνω διαδικασία έχει μοντελοποιηθεί στο παρελθόν με χρήση Αντιληπτικής γραμμικής προσέγγισης σήματος (Perceptual Linear Prediction - PLP) [4] και έχει βρει εφαρμογές για τη μελέτη του ακουστικού συστήματος του συνόλου των θηλαστικών και μάλιστα στις τεχνικές αναγνώριση φωνής στους ανθρώπους.

Έστω $x[n]$ το σήμα που φτάνει στη φάλαινα για κάθε διακριτό χρόνο n . Το εξωτερικό τμήμα του αυτιών της λαμβάνει το ερέθισμα και φιλτράρει το σήμα με στόχο να ανακτήσει τη απόκριση του φίλτρου της ακουστικής οδού προσμένοντας ότι το σήμα που έλαβε προέρχεται από φάλαινα. Χρησιμοποιώντας την παραδοχή ότι το σήμα που προέρχεται από ήχους φαλαινών έχει χρονική συνοχή και δεν γίνονται τυχαίες μεταβάσεις είμαστε σε θέση να εκφράσουμε το σήμα το διακριτό χρόνο n ως γραμμικό συνδυασμό του ίδιου στους $2p+1$ προηγούμενους χρόνους με σχετικά μικρό σφάλμα προσέγγισης $e[n]$, δηλαδή

$$x'[n] = \sum_{k=1}^{2p+1} a_k x[n-k] \quad (2)$$

όπου p εκφράζει τον αριθμό των formants που ορίσαμε πριν.

Είναι φανερό ότι σφάλμα e της προσέγγισης ικανοποιεί τη ακόλουθη σχέση στο πεδίο του χρόνου :

$$e[n] = x[n] - \sum_{k=1}^{2p+1} a_k x[n-k] \quad (3)$$

Παίρνοντας το μετασχηματισμό Fourier των δύο μερών έχουμε:

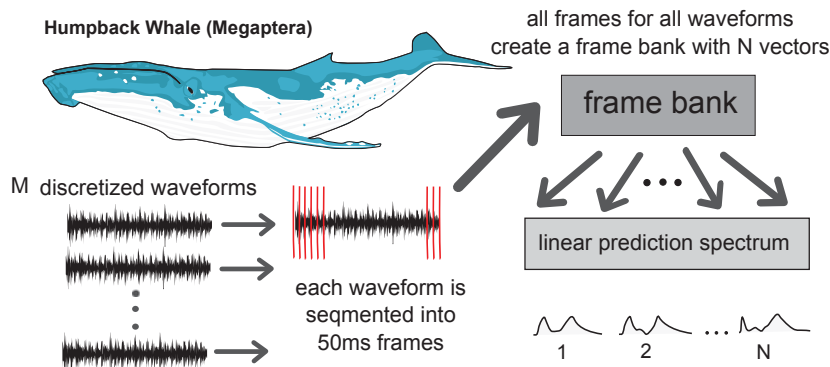
$$X(\omega) = \frac{1}{1 - \sum_{k=1}^{2p+1} a_k e^{-ik\omega}} E(\omega) \quad (4)$$

Παρατηρούμε ότι η παραπάνω σχέση έχει ίδια μορφή με την σχέση 1 του μοντέλου πηγής - φίλτρου με το E και εδώ να περιγράφει λευκό θόρυβο. Μένει να βρούμε το κατάλληλο διάνυσμα των συντελεστών α το οποίο ελαχιστοποιεί το σφάλμα προσέγγισης $e[n]$ το οποίο επιτυγχάνουμε λύνοντας το παρακάτω πρόβλημα βελτιστοποίησης :

$$\alpha' = \arg \min_{\alpha} \|e\|_2 \quad (5)$$

Ως αποτέλεσμα έχουμε μια προσέγγιση των συντελεστών c_k από τους αντίστοιχους συντελεστές α'_k .

Το παραπάνω φάσμα που προσεγγίζει το φίλτρο της φωνητικής οδού ονομάζεται φάσμα γραμμικής προσέγγισης (linear prediction spectrum) και αποτελεί ένδειξη της πραγματικής πληροφορίας που λαμβάνεται ως είσοδο στο ακουστικό νεύρο η οποία θα υποστεί στη συνέχεια φασματική ανάλυση. Όπως έγινε αντιληπτό οι φάλαινες έχουν μηχανισμούς τέτοιους που βελτιστοποιούν την αντίληψη τους σε ήχους που παράγουν οι ίδιες ωστόσο παραμένει άγνωστο πως αυτό το μοντέλο χαρακτηρίζει ξένες ως προς αυτές κυματομορφές. Στο σχήμα 1.1 παρουσιάζεται γραφικά η παραπάνω διαδικασία.



Σχήμα 1.1 Μαθηματικό μοντέλο που περιγράφει την επεξεργασία στο εξωτετικό - μεσαίο τμήμα του αυτιού της Μεγάπτερης φάλαινας.

1.3 Φασματική ανάλυση στο εσωτερικό τμήμα (cochlear filtering) των αυτιών της Μεγάπτερης φάλαινας.

Η φασματική ανάλυση προσομοιώνει τη λειτουργία του εσωτερικού τμήματος του αυτιού (ακουστικού νεύρου). Ως είσοδο λαμβάνει τη ακολουθία φασμάτων γραμμικής προσέγγισης των frames διάρκειας 50ms που αποτελούν ένα σήμα. Με άλλα λόγια λαμβάνει τα αποτελέσματα του μοντέλου που περιγράφηκε παραπάνω και παράγει μια ακολουθία από προσαρμοσμένα φάσματα που το καθένα θα περιγράφεται από L συχνότητες της σχεδόν λογαριθμικής κλίμακας ERB [5] η οποία ορίζεται βάση τα χαρακτηριστικά του ακουστικού κοχλίου (cochlear) ως :

$$ERB = B + f_c/Q \tag{6}$$

όπου B είναι το ελάχιστο πλάτος του φίλτρου, f_c η κεντρική συχνότητα του φίλτρου και Q ο συντελεστής ποιότητας (quality factor). Έχει βρεθεί ότι στις Μεγάπτερες φάλαινες κατάλληλες τιμές είναι $B = 10$ και $Q = 10$ [6].

Ο κοχλίας συμπεριφέρεται ως ένα σύνολο από κατάλληλα Gammatone φίλτρα [7] τα οποία διαχωρίζουν τις συχνότητες τοπολογικά κατά μήκος της βασικής μεμβράνης του ακουστικού νεύρου (basilar membrane) η οποία έχει μήκος 58mm. Κάθε μεμονωμένο φίλτρο προσδιορίζεται από μια κεντρική συχνότητα f_c και ένα πλάτος (bandwidth) Δf . Η παρακάτω σχέση περιγράφει τη μορφή των φίλτρων που χρησιμοποιούνται

$$g(t) = at^{n-1}e^{-2\pi\Delta f t} \cos(2\pi f_c t + \phi) \tag{7}$$

όπου f_c η κεντρική συχνότητα, ϕ η φάση, a το πλάτος, n η τάξη, Δf το πλάτος του φίλτρου και t ο χρόνος.

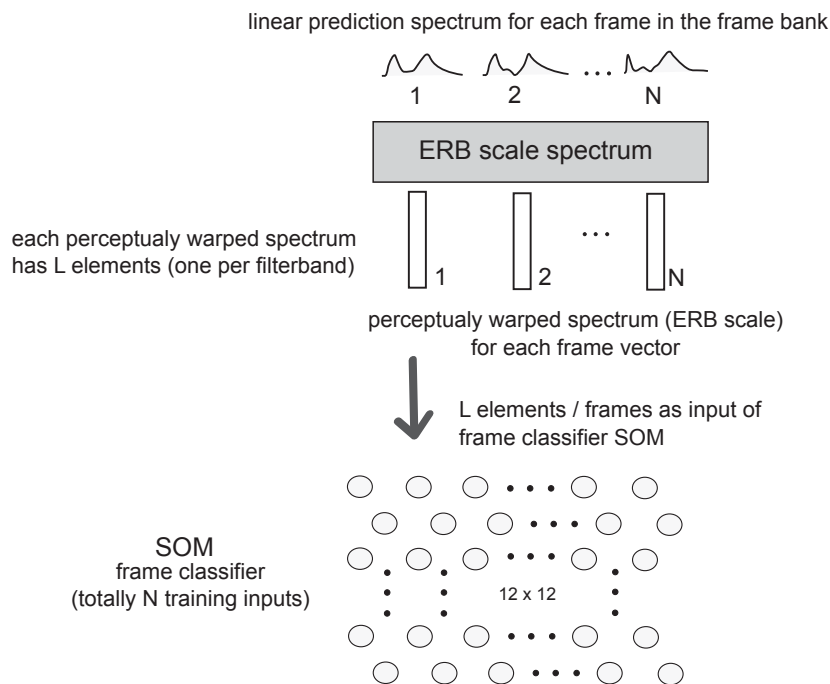
Σημειώνουμε ότι οι κεντρικές συχνότητες δεν ακολουθούν γραμμική κλίμακα κατά μήκος της βασικής μεμβράνης αντιθέτως ακολουθούν την καμπύλη

$$f_c(x) = 240(10^{0.0362x} - 0.958) \tag{8}$$

όπου το x εκφράζει την απόσταση κατά μήκος της βασικής μεμβράνης. Οι βασισμένες σε ανθρώπους μελέτες σχετικά με την απόσταση διαδοχικών φίλτρων έχουν δώσει ως αποδεκτή τιμή τα $0.89mm$. Γνωρίζοντας ότι τα φίλτρα έχουν την ίδια μορφή στις φάλαινες και παράλληλα ότι είναι πιο πυκνά σε αυτές μας κάνει να σκεφτούμε τη χρήση μιας λίγο μικρότερης απόστασης.

Ως αποτέλεσμα της ανάλυσης που περιγράφηκε σε αυτή την ενότητα παίρνουμε το χαρακτηρισμό κάθε τμήματος ενός ήχου με ένα διάνυσμα L στοιχείων, του οποίου το κάθε στοιχείο θα δηλώνει την απόκριση του κάθε ήχου διάρκειας $50ms$ σε κάθε ένα από τα L Gammatone φίλτρα.

1.4 Νευρωνική κωδικοποίηση και κατηγοριοποίηση των τμημάτων των ήχων που καταλήγουν στο εγκεφαλικό στέλεχος.



Σχήμα 1.2 Self Organizing Map (SOM) για τη κατηγοριοποίηση των ηχητικών τμημάτων (frames) στον εγκέφαλο της Μεγάπτερης φάλαινας.

Σε αυτό το στάδιο γίνεται η περιγραφή με χρήση ενός Self Organizing Map (SOM) [8] νευρωνικού δικτύου η εκ των υστέρων επεξεργασία των φασμάτων τα οποία κωδικοποιήθηκαν στο κοχλία, στους νευρώνες του εγκεφάλου. Το συγκεκριμένο μοντέλο έχει εφαρμοστεί επιτυχώς για το χαρακτηρισμό εγγενών στις φάλαινες σημάτων [9].

Τα SOMs είναι Νευρωνικά δίκτυα τα οποία εκφράζουν με μαθηματικό τρόπο την διαδικασία αυτο-εκμάθησης χωρίς επίβλεψη (unsupervised learning) που είναι χαρακτηριστικό γνώρισμα κάθε είδους. Η μόνη εκ' των προτέρων πληροφορία που απαιτούν είναι τα δεδομένα εισόδου σε αντίθεση των κλασικών νευρωνικών δικτύων που απαιτούν ζεύγη δεδομένων-αποτελεσμάτων για την εκπαίδευση τους.

Ξεκινώντας από M_1 ήχους από φάλαινες και από ένα μικρότερο αριθμό M_2 από απλές υποθαλάσσιες πηγές και αναλύοντας το καθένα από αυτά σε στοιχειώδη τμήματα των $50ms$ παράγουμε N_1 frames από εγγενείς ήχους φάλαινων και αντίστοιχα N_2 από ανθρωπογενείς ήχους. Το σύνολο των $N = N_1 + N_2$ frames αποτελεί τη τράπεζα των ηχητικών τμημάτων που θα χρησιμοποιηθούν για την παραγωγή δεδομένων εισόδου του SOM και στη συνέχεια θα το εκπαιδεύσουν.

Στο εκπαιδευμένο νευρωνικό δίκτυο που θα προκύψει θα έχει πραγματοποιηθεί διαχωρισμός των N διανυσμάτων εισόδου σε ένα τοπολογικό πλέγμα 2 διαστάσεων όπως φαίνεται στο 1.2.

Παρότι δεν θα παρουσιάσουμε λεπτομέρειες της εκπαίδευσης ενός SOM κρίνεται απαραίτητο να εκθέσουμε με συντομία τις διαδικασίες στο τυχαίο βήμα εκπαίδευσης k . Στο βήμα αυτό οι νευρώνες στα SOMs οι οποίοι έχουν τιμές που ανήκουν στο διανυσματικό χώρο διάστασης ίσης με το διάνυσμα εισόδου s_k του δικτύου, περιγράφονται με τα διανύσματα w_{ij} , όπου i, j οι συντεταγμένες του κάθε νευρώνα στο πλέγμα. Στη συνέχεια επιλέγεται ο νευρώνας του πλέγματος που είναι πλησιέστερος ως προς την ευκλείδεια νόρμα στο διάνυσμα των δεδομένων, δηλαδή:

$$w_{winner} = \arg \min_{w_{ij}} \|s_k - w_{ij}\|_2 \quad (9)$$

Λέμε ότι ο παραπάνω νευρώνας πυροδοτείται και ακολουθεί η προσαρμογή του ιδίου αλλά και τον γειτονικών νευρώνων στο πλέγμα ώστε το δίκτυο να ταιριάζει ακόμη καλύτερα στο διάνυσμα δεδομένων. Η προσαρμογή γίνεται με τον ακόλουθο τρόπο:

$$w_{ij}^{new} = w_{ij} + \alpha_k h(w_{winner}, w_{ij}) \|s_k - w_{ij}\|_2 \quad (10)$$

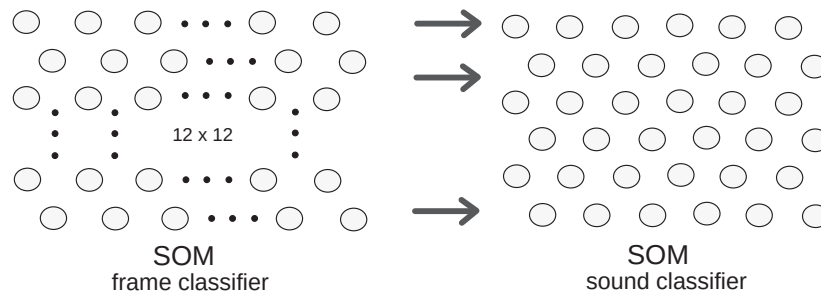
όπου $0 < \alpha_k < 1$ ο ρυθμός εκμάθησης στο k στάδιο εκπαίδευσης και $h(\cdot)$ η απόσταση δύο νευρώνων στο πλέγμα. Σημειώνουμε ότι το α_k θα τείνει προοδευτικά στο μηδέν προσομοιώνοντας τη πραγματική λειτουργία της μάθησης στην οποία όταν υπάρχει ήδη μεγάλη έκθεση σε ερεθίσματα (εμπειρία) ο εγκέφαλος δεν μπαίνει σε διαδικασία εκμάθησης.

Η χρήση $12 \times 12 = 144$ νευρώνων φαίνεται καλή επιλογή για το μοντέλο μας αν λάβουμε υπόψιν ότι προηγούμενες υλοποιήσεις που αξιοποιούσαν μόνο ήχους φάλαινων απαιτούσαν $10 \times 10 = 100$ νευρώνες [9,10]. Στο σχήμα 1.2 βλέπουμε τα στάδια κωδικοποίησης που περιγράφηκαν.

Αντικείμενο μελέτης μελλοντικών εργασιών είναι να διερευνήσουμε τη κατανομή των frames που αντιστοιχούν στους ανθρωπογενείς ήχους στους νευρώνες

του νευρωνικού πλέγματος και το κατά πόσον έχουμε τοπολογικό διαχωρισμό μεταξύ εγγενών και μη, ηχητικών frames.

2 Χρήση του μοντέλου για χαρακτηρισμό γενικών ακουστικών σημάτων.



Σχήμα 2.1 *Self Organizing Map (SOM)* με στόχο το χαρακτηρισμό ήχων στον εγκέφαλο της Μεγάπτερης φάλαινας.

Παρά το γεγονός ότι το νευρωνικό δίκτυο που περιγράφηκε παραπάνω, ομαδοποιεί τα frames σε 144 διακριτές κατηγορίες, οι ήχοι που φτάνουν στις φάλαινες αποτελούνται από μια σειρά από frames που το καθένα μπορεί να περιγραφεί με τον παραπάνω τρόπο. Επιπλέον δεν έχουν όλα τα σήματα την ίδια διάρκεια και έτσι είναι δύσκολο σε αυτό το μοντέλο να χαρακτηρίσουμε ολόκληρους ήχους. Για το λόγο αυτό έχει χρησιμοποιηθεί επιτυχώς ένα δεύτερο SOM [8] το οποίο έχει ως είσοδο στατιστικά χαρακτηριστικά του αλλαγής διατάξης των frames πάνω στο πρώτο νευρωνικό δίκτυο καθώς εξελίσσεται ο ήχος. Με αυτό το τρόπο το δεύτερο νευρωνικό δίκτυο είναι ικανό να κατηγοριοποιήσει ολόκληρες κυματομορφές.

Στο παρελθόν έχει γίνει χαρακτηρισμός ήχων που παράγουν οι φάλαινες με χρήση συνολικά $5 \times 5 = 25$ νευρώνων στο δεύτερο SOM. Ο αριθμός αυτός προκύπτει έτσι ώστε να υπερκαλύπτει τη γκάμα των φανερά διαφορετικών ήχων που παράγουν οι φάλαινες. Με το τρόπο αυτό κατάφεραν να κατηγοριοποιήσουν τους διάφορους ήχους και να γίνει αντιληπτό το πως μια φάλαινα αντιλαμβάνεται τους εγγενείς σε αυτήν ήχους. Η ιδέα στο μοντέλο μας είναι να επεκτείνουμε αυτή τη διαδικασία λαμβάνοντας υπόψιν και ήχους από το περιβάλλον της και αυτό θα το επιτύχουμε αυξάνοντας το πλήθος των διαθέσιμων νευρώνων που αποσκοπούν στην κατηγοριοποίηση σε πλήθος $6 \times 6 = 36$ όπως φαίνεται στο σχήμα 2.1. Με αυτό το τρόπο θέλουμε να δούμε συνολικά εάν μια φάλαινα μπορεί να αντιληφθεί έναν ανθρωπογενή ήχο σαν δικό της ή σαν ήχο άλλης φάλαινας. Δηλαδή, να εξετάσουμε εάν ένας εξωτερικός ήχος είναι ικανός να την αποπροσανατολίσει. Επίσης θέλουμε να δούμε για παρόμοιους ανθρωπογενείς ήχους εάν υπάρχει σύγκληση του χαρακτηρισμού τους, δηλαδή για μικρές μεταβολές σε ένα ήχο εάν διεγερθούν αντίστοιχα ίδιοι ή κοντινοί νευρώνες του πλέγματος και κατά επέκταση τη μελέτη της ευαισθησίας που παρουσιάζει η φάλαινα στο να καταλαβαίνει τις μικρές κριμένες διαφορές γενικών ακουστικών σημάτων.

3 Ευχαριστίες

Η παρούσα εργασία πραγματοποιήθηκε στο πλαίσιο του έργου ΠΕΦΥΚΑ της Δράσης ΚΡΗΠΙΣ της ΓΓΕΤ. Το έργο συγχρηματοδοτείται από την Ελλάδα και το Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης της Ευρωπαϊκής Ένωσης στο Πλαίσιο του ΕΣΠΑ και του Ε.Π. Ανταγωνιστικότητα και Επιχειρηματικότητα.

4 Αναφορές

- [1] D.R. Ketten «The cetacean ear: Form, frequency and evolution» in *Marine mammal sensory system* J.A.Thomas, R.A. Kastelein and A.Y. Supin. Plenum press, pp. 53-75, New York (1992)
- [2] K. Johnson «Acoustic and auditory phonetics» *Phonetica (International Journal of Phonetic Science)* 61(1), pp 56-58 (2004)
- [3] E. Mercado, J.N. Schneider, A.A Pack and L.M. Herman «Sound production by singing humpback whales» *J.Acoust. Soc. Am.* 127(4), pp 2678-2691 (2010)
- [4] H. Hernansky «Perceptual linear predictive (PLP) analysis of speech» *J.Acoust. Soc. Am.* 87(4), pp 1738-1752 (1990)
- [5] B.C.J. Moore and B.R. Glasberg «Suggested formulae for calculating auditory-filter bandwidths and excitation patterns» *J.Acoust. Soc. Am.* 74(3), pp 750-753 (1983)
- [6] D.R. Ketten «Functional analysis of whale ears: Adaptions for underwater hearing» *I.E.E.E proc. Under. Acoust.* vol 1, pp 264-270 (1994)
- [7] R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice «An efficient auditory filterbank based on the gammatone function» *APU report*, 2341, (1987).
- [8] T. Kohonen «Self-Organized formation of topologically correct feature maps» *Biological Cybernetics*, vol 43(1), pp. 59-69, (1982)
- [9] S.O. Murray, E. Mercado and H.L. Roitblat «Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalization» *J.Acoust. Soc. Am.* 104, pp 1679-1688 (1998)
- [10] S.O. Murray, E. Mercado and H.L. Roitblat «The neural network classification of false killer whale (*Pseudorca crassidens*) vocalization» *J.Acoust. Soc. Am.* 104, pp 1679-1688 (1998)

