



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Μοντελοποίηση επεξεργασίας σημάτων ομιλίας στο
ακουστικό νευρικό σύστημα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΗΛΙΑ ΑΛΕΒΙΖΟΥ

Επιβλέπων : Καραγιάννης Γεώργιος
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2008



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Μοντελοποίηση επεξεργασίας σημάτων ομιλίας στο ακουστικό νευρικό σύστημα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΗΛΙΑ ΑΛΕΒΙΖΟΥ

Επιβλέπων : Καραγιάννης Γεώργιος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 15η Απριλίου 2008.

.....
Καραγιάννης Γεώργιος
Καθηγητής Ε.Μ.Π.

.....
Μαραγκός Πέτρος
Καθηγητής Ε.Μ.Π.

.....
Κόλλιας Στέφανος
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2008

(Υπογραφή)

.....

ΑΛΕΒΙΖΟΣ ΗΛΙΑΣ

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

© 2008 - All rights reserved

Περίληψη

Τα ακουστικά σήματα διέρχονται από ένα πλήθος ενδιάμεσων σταδίων νευρωνικής επεξεργασίας πριν φτάσουν στον ακουστικό φλοιό όπου θεωρείται ότι λαμβάνουν χώρα ανώτερα στάδια λεκτικής, συντακτικής και σημασιολογικής ανάλυσης. Κάθε στάδιο επιβάλλει τους δικούς του μετασχηματισμούς, πολλοί από τους οποίους μας είναι ακόμα άγνωστοι. Ο τρόπος που τελικά αναδύεται το νόημα κατόπιν αυτής της επεξεργασίας παραμένει ανοιχτό ζήτημα. Ένα γνωσιακό μοντέλο αντίληψης του λόγου που θα επιχειρούσε να συλλάβει αυτή τη διαδικασία θα έπρεπε να βασιστεί σε μία “εσωτερική” νευρωνική αναπαράσταση των σημάτων ομιλίας. Για να είναι χρήσιμη μία τέτοια αναπαράσταση θα πρέπει να διατηρεί τα σημαντικά χαρακτηριστικά τέτοιων σημάτων και με αυτά να τροφοδοτεί το γνωσιακό μοντέλο. Αυτή η διπλωματική παρουσιάζει μία επισκόπηση της παρούσας έρευνας στην περιοχή της ακουστικής νευρωνικής επεξεργασίας και μοντελοποίησης καθώς και μία απόπειρα υλοποίησης ενός μοντέλου που θα μπορούσε να λειτουργήσει ως εμπρόσθιο τμήμα σε ένα γνωσιακό οπίσθιο τμήμα αντίληψης του λόγου.

Λέξεις - κλειδιά : ακουστική μοντελοποίηση, επεξεργασία λόγου, ακουστική νευρωνική αναπαράσταση, γνωσιακό μοντέλο

Abstract

Auditory signals pass through a number of intermediate stages of neural processing before they reach the auditory cortex where higher levels of lexical, syntactic and semantic analysis are assumed to take place. Every stage imposes its own transformation, many of which are still unknown. A cognitive model of speech perception attempting to capture this process should be based on an “internal” neural representation of speech signals. In order for such a representation to be useful, it should preserve the important features of these signals and provide them to the cognitive model. This Diploma Thesis presents an overview of the current research in the area of auditory neural processing and modeling as well as an attempt to implement a model that could serve as a front-end for a cognitive back-end of speech perception.

Keywords : auditory modeling, speech processing, auditory neural representation, cognitive model

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κ. Αθανάσιο Πρωτόπαπα, ερευνητή στο Ινστιτούτο Επεξεργασίας Λόγου (Ι.Ε.Λ.), για την ευκαιρία που μου έδωσε να ασχοληθώ στη διπλωματική μου με το χώρο των γνωσιακών επιστήμων, κάτι που θεωρούσα μάλλον δύσκολο να καταφέρω εντός της σχολής. Επίσης, θα ήθελα να τον ευχαριστήσω για την καθοδήγηση του επί των θεμάτων της διπλωματικής και όχι μόνο. Τέλος, θα ήθελα να ευχαριστήσω και τον κ. Γεώργιο Καραγιάννη, καθηγητή Ε.Μ.Π. και διευθυντή του Ι.Ε.Λ., για την επίβλεψη της διπλωματικής από την πλευρά του Ε.Μ.Π.

“A sudden silence in the middle of a conversation suddenly brings us back to essentials: it reveals how dearly we must pay for the invention of speech.”

Emil Cioran

Περιεχόμενα

1	Εισαγωγή	19
1.1	Αντικείμενο της εργασίας	19
1.2	Οργάνωση της εργασίας	20
2	Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα	23
2.1	Ανατομική περιγραφή	23
2.1.1	Από τη σύλληψη μέχρι την πρώτη κωδικοποίηση σε νευρικά σήματα	23
2.1.1.1	Το έξω ους	24
2.1.1.2	Το μέσο ους	25
2.1.1.3	Το έσω ους	26
2.1.2	Από τον κοχλία στον ακουστικό φλοιό	28
2.1.2.1	Από τον κοχλία στο σύμπλεγμα κοχλιακών πυρήνων . . .	28
2.1.2.2	Από το σύμπλεγμα κοχλιακών πυρήνων στο άνω ελαιώδες σύμπλεγμα	29
2.1.2.3	Ο πλευρικός λημνίσκος	31
2.1.2.4	Το κάτω διδύμιο	32
2.1.2.5	Το έσω γονατώδες σώμα	32
2.1.2.6	Ο ακουστικός φλοιός	34
2.1.2.7	Ο μηχανισμός ανάδρασης του ακουστικού συστήματος . .	36
2.2	Λειτουργική περιγραφή	39
2.2.1	Μετασχηματισμοί των σημάτων στο πρώιμο, περιφερικό ακουστικό σύστημα	41
2.2.1.1	Τοπική αναπαράσταση	41
2.2.1.2	Χρονική αναπαράσταση	43
2.2.1.3	Μικτή χρονοτοπική αναπαράσταση	43
2.2.1.4	Πλευρική αναστολή	45
2.2.1.5	Αποκρίσεις σε τόνους διαμορφωμένους κατά συχνότητα (FM)	46
2.2.1.6	Αποκρίσεις σε τόνους διαμορφωμένους κατά πλάτος (AM)	46
2.2.1.7	Αποκρίσεις σε ενάρξεις (onset)	47
2.2.2	Μετασχηματισμοί των σημάτων στο κεντρικό ακουστικό σύστημα . .	48
2.2.2.1	Η υπόθεση της τοπικής αναπαράστασης (place representation)	48
2.2.2.2	Η αναπαράσταση βέλτιστης έντασης (best-intensity)	49
2.2.2.3	Η κατηγορική αναπαράσταση (categorical representation)	50

2.2.2.4	Αναπαράσταση πολλαπλών επιπέδων (multiscale representation)	50
2.2.2.5	Αποκρίσεις σε απλούς τόνους	50
2.2.2.6	Αποκρίσεις σε τόνους διαμορφωμένους κατά πλάτος (AM)	53
2.2.2.7	Αποκρίσεις σε τόνους διαμορφωμένους κατά συχνότητα (FM)	54
2.2.2.8	Αποκρίσεις σε πιο περίπλοκα ερεθίσματα	56
2.2.2.9	Αποκρίσεις σε άλλα ερεθίσματα	59
3	Υπολογιστική και μαθηματική μοντελοποίηση του ακουστικού συστήματος	63
3.1	Μοντέλα βασισμένα σε ψυχοακουστικά δεδομένα	64
3.1.1	Mel-frequency cepstrum coefficients	64
3.1.2	Αναπαράσταση με βάση την PLP (perceptual linear predictive) τεχνική	65
3.1.3	Το μοντέλο των Tchorz και Kollmeier	68
3.2	Μοντέλα βασισμένα σε νευροφυσιολογικά δεδομένα	71
3.2.1	Μοντελοποίηση του περιφερικού ακουστικού συστήματος	72
3.2.1.1	Η αναπαράσταση EIH (Ensemble Interval Histogram)	72
3.2.1.2	Κοχλιόγραμμα και συσχετισιόγραμμα	73
3.2.2	Μοντελοποίηση του κεντρικού ακουστικού συστήματος	80
3.2.2.1	Το φασματόγραμμα διαμόρφωσης	80
3.2.2.2	Το μοντέλο του Shamma	81
3.3	Ανασκόπηση	89
4	Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου	91
4.1	Ανάλυση του μοντέλου εσωτερικής αναπαράστασης	93
4.2	Υλοποίηση του συστήματος	95
4.2.1	Φασματόγραμμα αρχικού σήματος	95
4.2.2	Επεξεργασία του φασματογράμματος	95
4.2.2.1	Η κύρια συνάρτηση, multiresScan	97
4.2.2.2	Η συνάρτηση scanSpecGabor	99
4.2.2.3	Η συνάρτηση extractpatch	100
4.2.2.4	Η συνάρτηση processpatch	102
4.2.2.5	Η συνάρτηση findpeaks	103
4.2.2.6	Η συνάρτηση conjpeaks	106
4.2.2.7	Η συνάρτηση classifypeaks	107
4.2.2.8	Βοηθητικές συναρτήσεις	108
4.2.3	Απεικόνιση των αποτελεσμάτων	111
4.3	Έλεγχος του συστήματος	113
4.3.1	Απόκριση σε απλούς τόνους	113
4.3.2	Απόκριση σε ευρυζωνικές μπάρες “θορύβου”	114
4.3.3	Fm sweep	115
4.3.4	Πραγματικά σήματα ομιλίας	117

ΠΕΡΙΕΧΟΜΕΝΑ

5	Επίλογος	131
5.1	Σύνοψη και συμπεράσματα	131
5.2	Μελλοντικές επεκτάσεις	132

Κατάλογος Σχημάτων

2.1	Ανατομία του ωτός, [30]	24
2.2	Η μετατροπή των κυμάτων πίεσης του αέρα σε μηχανικές και εν συνεχεία σε υδραυλικές ταλαντώσεις, [31]	25
2.3	Η βασική μεμβράνη, [32]	26
2.4	Κοχλίας	27
2.5	Μεταφέροντας το ηχητικό σήμα από τον κοχλία στο ακουστικό νεύρο, [35]	28
2.6	Φωτογραφίες από ηλεκτρονικό μικροσκόπιο σάρωσης	29
2.7	Το κεντρικό νευρικό σύστημα, [12]	30
2.8	Πρόσθιες όψεις εγκεφαλικού στελέχους, [10]	30
2.9	Τομή ανώτερου προμήκους μυελού, [10]	31
2.10	Τομή κατώτερης γέφυρας, [10]	32
2.11	Τομή μέσης γέφυρας, [10]	33
2.12	Τομή κατώτερου μεσεγκεφάλου, [10]	33
2.13	Ο θάλαμος, [10]	34
2.14	Προς τον θάλαμο, [10]	35
2.15	Φωτογραφίες ακουστικού φλοιού, [10]	35
2.16	Ο ακουστικός φλοιός, [38]	37
2.17	Η διαδρομή των ηχητικών σημάτων στο νευρικό σύστημα, [39]	38
2.18	Γραφικές παραστάσεις του ρυθμού εκφόρτισης συναρτήσεως της χαρακτηριστικής συχνότητας για διάφορα επίπεδα πίεσης, [18]	42
2.19	Γραφικές παραστάσεις του ρυθμού εκφόρτισης συναρτήσεως της χαρακτηριστικής συχνότητας για ίνες χαμηλού SR, [7]	44
2.20	Γραφικές παραστάσεις της κυρίαρχης συχνότητας κλειδώματος συναρτήσεως της ΧΣ, [18]	45
2.21	Υπολογισμός του ALSR, [18]	46
2.22	Χάρτης συχνότητας - έντασης, [18]	49
2.23	Χώρος ανάλυσης στην αναπαράσταση πολλαπλών επιπέδων, [18]	51
2.24	Χρονική εξέλιξη του υποδεκτικού πεδίου ενός νευρώνα με ΧΣ 2.8 kHz στον ΑΙ γάτας, [7]	52
2.25	Παραδείγματα καμπυλών συντονισμού από νευρώνες στην πλευρική ζώνη rhesus πιθήκου, [26]	54
2.26	Η σχέση μεταξύ περιοχών του εγκεφάλου και συχνότητας ενός FM sweep, [27]	55
2.27	Χάρτης ευαισθησίας συχνοτήτων σε τμήμα ανθρώπινου ακουστικού φλοιού, [24]	56

2.28	Χρονική εξέλιξη της δραστηριότητας του φλοιού σε ένα FM sweep, [24] . . .	57
2.29	Φασματικά κύματα	58
2.30	Συνολική δραστηριότητα των νευρώνων ως απόκριση σε ακολουθίες FM sweep με διάφορους ρυθμούς επανάληψης, [17]	59
2.31	Εξάρτηση της επιλεκτικότητας κατεύθυνσης συναρτήσεως του ρυθμού επανάληψης, [17]	60
2.32	Αποκρίσεις σε ακολουθίες από κλικ, [14]	61
3.1	Η δομή του ακουστικού συστήματος, [7]	63
3.2	Συστοιχία φίλτρων που ακολουθεί την κλίμακα mel, [3]	65
3.3	Σύγκριση συνηθισμένου και mfcc φασματογράμματος	66
3.4	Σχηματικό διάγραμμα των σταδίων επεξεργασίας στην PLP αναπαράσταση, [11]	67
3.5	Σύγκριση LP με PLP , [11]	69
3.6	Δομή του μοντέλου των Tchorz και Kollmeier, [25]	70
3.7	Παράδειγμα αναπαράστασης με βάση το μοντέλο των Tchorz και Kollmeier, [25]	71
3.8	Η γενική δομή του μοντέλου EIH, [6]	72
3.9	Αποκρίσεις της συστοιχίας φίλτρων στο μοντέλο EIH, [6]	73
3.10	Διαδικασία υπολογισμού του EIH, [6]	74
3.11	Σύγκριση EIH και απλού φάσματος για ένα φωνήεν, [6]	74
3.12	Κοχλιόγραμμα και συσχετισιόγραμμα, [23]	75
3.13	Σχηματικό διάγραμμα του μοντέλου του κοχλία, [23]	75
3.14	Ενδεικτικές συναρτήσεις μεταφοράς φίλτρου μετάδοσης και φίλτρου μετατροπής στη βασική μεμβράνη, [23]	77
3.15	Σύγκριση φασματογράμματος και κοχλιογράμματος, [23]	77
3.16	Κοχλιόγραμμα του “tea” από το “grease”, [23]	78
3.17	Σύστημα υπολογισμού συσχετισιογράμματος, [23]	79
3.18	Παραδείγματα συσχετισιογραμμάτων, [23]	79
3.19	Μοντέλο υπολογισμού φασματογράμματος διαμόρφωσης, [8]	80
3.20	Σύγκριση φασματογράμματος διαμόρφωσης με κανονικό φασματόγραμμα, [8]	82
3.21	Σχηματικό διάγραμμα του μοντέλου του Shamma, [2]	83
3.22	Παραδείγματα εξόδων του πρώιμου σταδίου του μοντέλου του Shamma, [2]	83
3.23	Φασματικό κύμα, [4]	84
3.24	Διαδικασία εξαγωγής συναρτήσεων μεταφοράς $T(\Omega)$ και $T(\omega)$, [4]	85
3.25	Παραδείγματα RF και IR αποκρίσεων, [18]	85
3.26	Παραδείγματα STRF	86
3.27	Φλοιική αναπαράσταση στο μοντέλο του Shamma για σήμα τριών τόνων, [2]	87
3.28	Φλοιική αναπαράσταση στο μοντέλο του Shamma για καθοδικό και ανοδικό φασματικό κύμα, [2]	88
3.29	Φλοιική αναπαράσταση στο μοντέλο του Shamma για σήμα ομιλίας, [2]	88
4.1	Σχηματική αναπαράσταση των σταδίων επεξεργασίας του ακουστικού συστήματος, [7]	92
4.2	Αντιληπτικοί χάρτες, [7]	93
4.3	Η συστοιχία φίλτρων της συνάρτησης mfcc.	96

4.4	Σχηματικό διάγραμμα του συστήματος επεξεργασίας.	96
4.5	Παράδειγμα εφαρμογής της συνάρτησης <code>extractpatch</code>	101
4.6	Παραδείγματα Gabor φίλτρων και οι μετασχηματισμοί Fourier που τους αντιστοιχούν, [5].	103
4.7	Παράδειγμα εφαρμογής της συνάρτησης <code>processpatch</code>	104
4.8	Παράδειγμα εφαρμογής της συνάρτησης <code>findpeaks</code>	106
4.9	Παραδείγματα προτύπων, όπως προκύπτουν από τη συνάρτηση <code>createTemplate</code>	109
4.10	Παραδείγματα εφαρμογής των συναρτήσεων <code>fmsweep</code> και <code>energybar</code>	111
4.11	Στιγμιότυπο από βίντεο που παράγει η συνάρτηση <code>playmon</code>	113
4.12	Απόκριση του συστήματος σε σήμα δύο τόνων 500 και 3000 Hz	114
4.13	Απόκριση του συστήματος κατά την έναρξη μίας έκρηξης θορύβου	115
4.14	Απόκριση του συστήματος εντός μίας έκρηξης θορύβου	116
4.15	Απόκριση του συστήματος κατά την παύση μίας έκρηξης θορύβου	116
4.16	Απόκριση του συστήματος σε καθοδικό <code>fm sweep</code>	117
4.17	Απόκριση του συστήματος σε ανόδικο <code>fm sweep</code>	118
4.18	Φασματόγραμμα της φράσης “come home right away”.	119
4.19	Στιγμιότυπο 1 από την απόκριση του συστήματος για τη φράση “come home right away”.	119
4.20	Στιγμιότυπο 2 από την απόκριση του συστήματος για τη φράση “come home right away”.	120
4.21	Στιγμιότυπο 3 από την απόκριση του συστήματος για τη φράση “come home right away”.	120
4.22	Στιγμιότυπο 4 από την απόκριση του συστήματος για τη φράση “come home right away”.	121
4.23	Στιγμιότυπο 5 από την απόκριση του συστήματος για τη φράση “come home right away”.	121
4.24	Στιγμιότυπο 6 από την απόκριση του συστήματος για τη φράση “come home right away”.	122
4.25	Στιγμιότυπο 7 από την απόκριση του συστήματος για τη φράση “come home right away”.	122
4.26	Φασματόγραμμα της φράσης “we’ve done our part”.	123
4.27	Στιγμιότυπο 1 από την απόκριση του συστήματος για τη φράση “we’ve done our part”.	124
4.28	Στιγμιότυπο 2 από την απόκριση του συστήματος για τη φράση “we’ve done our part”.	124
4.29	Στιγμιότυπο 3 από την απόκριση του συστήματος για τη φράση “we’ve done our part”.	125
4.30	Στιγμιότυπο 4 από την απόκριση του συστήματος για τη φράση “we’ve done our part”.	125
4.31	Στιγμιότυπο 5 από την απόκριση του συστήματος για τη φράση “we’ve done our part”.	126
4.32	Φασματόγραμμα της φράσης “πάνω στον Περίφημο”.	127
4.33	Στιγμιότυπο 1 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.	127

4.34	Στιγμιότυπο 2 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.	128
4.35	Στιγμιότυπο 3 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.	128
4.36	Στιγμιότυπο 4 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.	129
4.37	Στιγμιότυπο 5 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.	129
4.38	Στιγμιότυπο 6 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.	130
4.39	Στιγμιότυπο 7 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.	130

Κεφάλαιο 1

Εισαγωγή

Οι ακριβείς μηχανισμοί μέσω των οποίων ο εγκέφαλος επεξεργάζεται τα αισθητηριακά αντιλήμματα είναι σε μεγάλο βαθμό άγνωστοι προς το παρόν, ιδιαίτερα όσον αφορά στο ακουστικό σύστημα. Ωστόσο, με την άνοδο των νευροεπιστημών και την εμφάνιση νέων τεχνικών τα τελευταία χρόνια, έχει καταστεί δυνατή η εξέτασή του με όλο και μεγαλύτερη λεπτομέρεια. Καινούρια πειραματικά δεδομένα εμφανίζονται με ταχείς ρυθμούς, προερχόμενα από διάφορες τεχνικές αλλά και είδη. Προφανώς, χωρίς κάποια θεωρία που θα ενοποιεί όλα αυτά τα δεδομένα και θα τα θέτει υπό ένα κοινό πλαίσιο ερμηνείας, αυτά θα παραμείνουν απλά δεδομένα. Είναι λοιπόν αναμενόμενο ότι η συσσώρευση πειραματικών δεδομένων θα επιφέρει σιγά σιγά (και ήδη έχει αρχίσει να το κάνει) την εμφάνιση θεωριών και των αντίστοιχων μοντέλων τους που θα επιχειρούν να συλλάβουν τα βασικά χαρακτηριστικά και τις ιδιότητες που διέπουν την επεξεργασία στον εγκέφαλο των ηχητικών σημάτων, αλλά και πιο συγκεκριμένα των σημάτων λόγου.

Μέχρι πρόσφατα, η έρευνα στο πεδίο της επεξεργασίας σημάτων ομιλίας περιοριζόταν κυρίως στο ίδιο το σήμα, όπως αυτό καταγράφεται, και στα χαρακτηριστικά που ενδεχομένως διαθέτει, με συνήθη σκοπό την αναγνώριση των λέξεων που περιέχουν. Προς αυτή την κατεύθυνση έχει αναπτυχθεί η λεγόμενη τεχνολογία αυτόματης αναγνώρισης φωνής (ASR, Automatic Speech Recognition), όπως για παράδειγμα τα συστήματα που είναι βασισμένα σε κρυφά μαρκοβιανά μοντέλα (HMMs, Hidden Markov Models). Συνήθως, αυτές οι προσεγγίσεις δεν λαμβάνουν υπόψιν τους τον τρόπο που ο εγκέφαλος επεξεργάζεται, αντιλαμβάνεται και αποκωδικοποιεί τα εισερχόμενα σήματα. Στις περιπτώσεις όπου επιχειρείται κάτι τέτοιο, η απόπειρα βασίζεται περισσότερο στη διαίσθηση του μηχανικού και στις ελλιπείς γνώσεις περί του περιφερικού ακουστικού συστήματος (αλλά πληρέστερες εν σχέσει με άλλα, πιο “εσωτερικά” στάδια επεξεργασίας), όπως στις περιπτώσεις των συντελεστών mfcc και της PLP τεχνικής. Τα νέα δεδομένα όμως επιτρέπουν πλέον μία αρχική, έστω και άχαρη ή και αφελή, χαρτογράφηση και ψηλάφηση αυτής της περιοχής.

1.1 Αντικείμενο της εργασίας

Στην ιδανική περίπτωση, θα ήταν επιθυμητό να έχουμε ένα γνωσιακό μοντέλο αντίληψης του λόγου το οποίο θα μπορούσαμε να τροφοδοτήσουμε με πραγματικά σήματα ομιλίας και να λάβουμε αποτελέσματα όσο το δυνατό πλησιέστερα με τα αποτελέσματα ψυχοακουστι-

κών πειραμάτων. Η παρούσα εργασία έχει πολύ πιο ταπεινές φιλοδοξίες. Το αντικείμενο της είναι η ανάπτυξη ενός συστήματος λογισμικού το οποίο ενδεχομένως θα μπορούσε να λειτουργήσει ως εμπρόσθιο τμήμα σε ένα τέτοιο γνωσιακό μοντέλο. Γίνεται η υπόθεση ότι τα ανώτερα στάδια επεξεργασίας του λόγου, που φαίνεται να εντοπίζονται κυρίως στον φλοιό του εγκεφάλου, βασίζονται σε κάποια χαμηλότερα στάδια. Αυτά με τη σειρά τους έχουν ως κύρια λειτουργία να εξάγουν από τα αρχικά σήματα αυτά τα χαρακτηριστικά που είναι απαραίτητα για την ανώτερη επεξεργασία. Το εμπρόσθιο τμήμα που αναπτύσσεται σε αυτή την εργασία τοποθετείται ακριβώς σε αυτό το ενδιάμεσο επίπεδο και έχει σκοπό να παρέχει μία κατάλληλη “εσωτερική” αναπαράσταση των ηχητικών σημάτων. Ένα γνωσιακό μοντέλο που θα επιχειρούσε να μοντελοποιήσει ανώτερα στάδια στα οποία τελικά αναδύεται το νόημα του αρχικού σήματος, θα μπορούσε να βασιστεί σε αυτή την αναπαράσταση.

Το σύστημα ουσιαστικά αποτελείται από συστοιχίες ανιχνευτών, καθέναν με συγκεκριμένα χαρακτηριστικά, με βάση τα οποία προσπαθούν να εντοπίσουν ορισμένες ιδιότητες του σήματος ομιλίας οι οποίες θεωρούμε ότι είναι σημαντικές για την κατανόησή του. Δέχεται στην είσοδό του το αρχικό σήμα μαζί με κάποιες επιπλέον παραμέτρους και παράγει στην έξοδό του ένα σύνολο από χάρτες ενεργοποίησης των παραπάνω ανιχνευτών. Έντονη απόκριση ενός ανιχνευτή σηματοδοτεί την παρουσία στο σήμα της ιδιότητας που ο συγκεκριμένος ανιχνευτής είναι εξειδικευμένος να εντοπίζει. Θεωρείται ότι αυτοί οι χάρτες ενεργοποίησης που παράγονται από τις αποκρίσεις των ανιχνευτών θα μπορούσαν να τροφοδοτήσουν ένα επόμενο στάδιο ακουστικής επεξεργασίας ώστε να καταλήξουμε στην αποκωδικοποίηση του μηνύματος.

Επειδή η έξοδος του συστήματος είναι αρκετά πολύπλοκη ώστε να είναι δυνατή η εξαγωγή συμπερασμάτων με απλή επισκόπησή της, παρέχεται και ένα βοηθητικό σύστημα για την απεικόνιση των αποτελεσμάτων. Το βοηθητικό σύστημα λαμβάνει στην είσοδό του τα αποτελέσματα του κυρίως συστήματος και στο τέλος παράγεται ένα αρχείο βίντεο. Στο βίντεο απεικονίζονται οι χάρτες ενεργοποίησης μαζί με το αρχικό σήμα και το φασματόγραμμα του καθώς οι ανιχνευτές μετατοπίζουν τα παράθυρα ανάλυσής τους από τη μία χρονική στιγμή στην επόμενη. Με αυτόν τον τρόπο μπορεί να ελεγχθεί η απόδοση του συστήματος.

1.2 Οργάνωση της εργασίας

Ο βασικός κορμός της εργασίας αποτελείται από τα επόμενα τρία κεφάλαια. Στο 2^ο κεφάλαιο γίνεται μία περιγραφή της διαδρομής που ακολουθούν τα ηχητικά σήματα στον εγκέφαλο, από τη σύλληψή τους στο αυτί μέχρι τον ακουστικό φλοιό. Στο πρώτο τμήμα του κεφαλαίου παρουσιάζονται κάποιες βασικές πληροφορίες ανατομίας του ακουστικού νευρικού συστήματος. Στο δεύτερο γίνεται μία μετατόπιση της οπτικής προς μία λειτουργική περιγραφή και στους μετασχηματισμούς που υφίστανται τα σήματα καθώς διέρχονται από τις ανατομικές δομές που περιγράφονται προηγουμένως. Το 3^ο κεφάλαιο παρουσιάζει διάφορες απόπειρες υπολογιστικής και μαθηματικής μοντελοποίησης του ακουστικού συστήματος που έχουν ήδη επιχειρηθεί. Ορισμένες από αυτές βασίζονται σε ψυχοακουστικά δεδομένα ενώ οι πιο πρόσφατες διαθέτουν και νευροφυσιολογικό υπόβαθρο. Τέλος, στο 4^ο κεφάλαιο περιγράφεται το θεωρητικό μοντέλο στο οποίο βασίζεται η εργασία, η υλοποίησή του καθώς και ο έλεγχός του με τεχνητά σήματα αλλά και με πραγματικά σήματα ομιλίας. Η εργασία κλείνει με το 5^ο κεφάλαιο όπου παρουσιάζονται τα συμπεράσματα και τυχόν

Κεφάλαιο 1. Εισαγωγή

μελλοντικές προεκτάσεις.

Κεφάλαιο 2

Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα

Τα σήματα λόγου, ως ηχητικά σήματα, διέρχονται από τα ίδια στάδια επεξεργασίας όπως και οποιοδήποτε άλλο ηχητικό σήμα αν και φαίνεται ότι υπάρχει σημαντική διαφοροποίηση στα ανώτερα επίπεδα (φλοιός). Μετά την όραση, η ακοή είναι η αίσθηση με τη μεγαλύτερη διακριτική ικανότητα και ευαισθησία στο ανθρώπινο είδος. Είναι αξιοσημείωτο το γεγονός ότι τα εύρη συχνοτήτων και εντάσεων (χαρακτηριστικά με βάση τα οποία επιτελείται η ανάλυση ήχων) στα οποία είναι πιο ευαίσθητο το ανθρώπινο αυτί είναι ακριβώς αυτά όπου εμπίπτει το μεγαλύτερο ποσοστό της πληροφορίας που περιέχεται σε ένα σήμα λόγου. Οι διακυμάνσεις του αέρα που συνιστούν ένα σήμα ήχου υφίστανται μία σειρά πολύπλοκων, μη γραμμικών μετασχηματισμών από τη στιγμή που συλλαμβάνονται από το αυτί μέχρι την τελική τους νευρωνική αναπαράσταση στον ακουστικό φλοιό του εγκεφάλου. Αυτό το κεφάλαιο επικεντρώνεται στην περιγραφή αυτής της διαδρομής μέσα στο νευρικό σύστημα με βάση τις ελλιπείς γνώσεις που έχουμε στη διάθεσή μας μέχρι στιγμής. Αρχικά, γίνεται μία ανατομική περιγραφή των διαφόρων δομών που λαμβάνουν μέρος σε αυτή τη διαδικασία και κατόπιν μία εκτενέστερη λειτουργική περιγραφή που θα παράσχει τα θεωρητικά και πειραματικά θεμέλια πάνω στα οποία θα αναπτυχθεί εν συνεχεία η παρούσα εργασία.

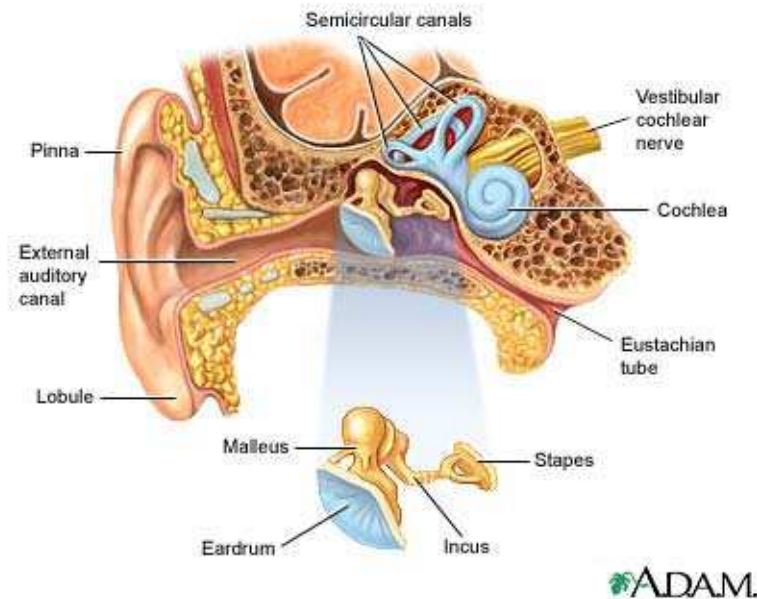
2.1 Ανατομική περιγραφή

Αυτή η ενότητα βασίστηκε στα [1], [7], [10] και [12].

2.1.1 Από τη σύλληψη μέχρι την πρώτη κωδικοποίηση σε νευρικά σήματα

Τα κύματα πίεσης που συλλαμβάνει το αυτί δεν μετατρέπονται άμεσα σε νευρικά σήματα αλλά διέρχονται αρχικά από ορισμένες μετατροπές μηχανικής και υδραυλικής φύσης πριν παρουσιαστούν στους προσαγωγούς νευρώνες που θα τα οδηγήσουν τελικά στα ενδότερα του εγκεφάλου. Η έκταση αυτών των μετασχηματισμών δεν είναι αμελητέα, όπως γίνεται κατανοητό και με μία απλή επισκόπηση της ανατομίας του αυτιού στο σχήμα 2.1.

Η ύπαρξη ακουστικού μηχανισμού και στις δύο πλευρές του κεφαλιού είναι σημαντική για τον εντοπισμό της προέλευσης των ήχων. Άτομα που έχουν χάσει την αίσθηση της ακοής από το ένα αυτί διατηρούν την ικανότητα να αντιλαμβάνονται κανονικά ήχους αλλά αντιμετωπίζουν δυσκολίες στο να καταλάβουν ποια είναι η πηγή τους, ιδιαίτερα όταν βρίσκονται σε περιβάλλον με πολλές τέτοιες πηγές από διάφορες κατευθύνσεις.



Σχήμα 2.1: Ανατομία του ωτός, [30]

Το αυτί μπορεί να διαχωριστεί στο έξω, μέσο και έσω ους η λειτουργία των οποίων επεξηγείται στις επόμενες ενότητες.

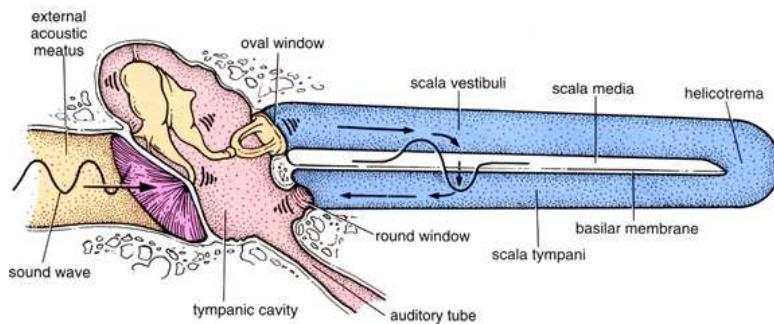
2.1.1.1 Το έξω ους

Το έξω ους αποτελείται από το ωτικό πτερύγιο (pinna) και από τον έξω ακουστικό πόρο (external auditory meatus). Το ωτικό πτερύγιο εντοπίζει τους ήχους και τους κατευθύνει στο εσωτερικό του αυτιού, όντας ελαφρά πιο ευαίσθητο σε ήχους προερχόμενους από μπροστά σε σχέση με όσους έρχονται από πίσω. Το σχήμα και το μέγεθος του πτερυγίου καθορίζουν και το είδος των σημάτων που μπορεί να συλλάβει. Ήχοι με πολύ μεγάλο μήκος κύματος που ξεπερνούν τις διαστάσεις του πτερυγίου δεν είναι δυνατό να εντοπιστούν.

Ο έξω ακουστικός πόρος έχει μήκος περίπου 2.5 εκ. σε έναν ενήλικα άνδρα και είναι κλειστός από τη μία μεριά λειτουργώντας ως ένας συντονιστής (quarter-wave resonator). Η χαμηλότερη συχνότητα συντονισμού του είναι περίπου 3.44 kHz. Αυτό το χαρακτηριστικό του συντονισμού σε υψηλές συχνότητες είναι χρήσιμο στην αντίληψη δασέων ήχων (fricatives) καθώς ένα μεγάλο ποσοστό της ενέργειας που βοηθάει στον διαχωρισμό μεταξύ τους βρίσκεται σε περιοχές άνω των 2 kHz.

2.1.1.2 Το μέσο ους

Το μέσο ους αποτελείται από την τυμπανική μεμβράνη (tympanic membrane), που αποτελεί και το σύνορο διαχωρισμού από το έξω ους, και από τρία οστάρια (ossicles), τη σφύρα (malleus), τον άκμονα (incus) και τον αναβολέα (stapes). Η τυμπανική μεμβράνη δονείται, ανταποκρινόμενη σε μικρές μεταβολές πίεσης για ένα αρκετά μεγάλο εύρος συχνοτήτων (βλ. σχήμα 2.2). Σε χαμηλές συχνότητες, όλες οι περιοχές της μεμβράνης δονούνται με την ίδια συχνότητα σε αντίθεση με τις υψηλές συχνότητες όπου κάθε περιοχή μπορεί να είναι ευαίσθητη σε διαφορετικά εύρη συχνοτήτων. Η αλυσίδα των τριών οσταρίων που ακολουθεί λειτουργεί ως γέφυρα μεταξύ της μεμβράνης και του κοχλίου του έσω ωτός. Αυτή η γέφυρα αιωρείται εντός της κοιλότητας του μέσου ωτός ώστε να μπορεί να δονείται ανεμπόδιση, περιβαλλόμενη από αέρα και στηριζόμενη μέσω κάποιων συνδέσμων. Οι ταλαντώσεις της μεμβράνης μεταφέρονται στα τρία οστάρια των οποίων οι μηχανικές ταλαντώσεις περνάνε με τη σειρά τους στην ωσειδή θυρίδα (oval window) του κοχλίου. Σε αυτό το στάδιο έχουμε επομένως την πρώτη μετατροπή από κύματα πίεσης του αέρα σε μηχανικές ταλαντώσεις, με τον όλο μηχανισμό να είναι εξαιρετικά καλά προσαρμοσμένος στις συχνότητες που είναι σημαντικές για τα σήματα λόγου.



Σχήμα 2.2: Η μετατροπή των κυμάτων πίεσης του αέρα σε μηχανικές και εν συνεχεία σε υδραυλικές ταλαντώσεις, [31]

Η κύρια λειτουργία του μέσου ωτός είναι αυτή ενός μετασχηματιστή που αυξάνει την πίεση των ηχητικών σημάτων κατά 30 dB περίπου. Η αύξηση της πίεσης είναι απαραίτητη εφόσον ο κοχλίας στο έσω ους είναι γεμάτος με υγρό, γεγονός που δημιουργεί διαφορά αντίστασης (impedance mismatch) σε σχέση με τον αέρα του μέσου και έξω ωτός. Λόγω αυτής της διαφοράς στην αντίσταση των δύο μέσων, ένα μεγάλο μέρος της ενέργειας που προσπίπτει στον κοχλίο ανακλάται, με αποτέλεσμα να καθίσταται αναγκαία μία αύξηση της πίεσης ώστε να μην χανθεί μέρος του σήματος κατά την μετάβασή του από το ένα τμήμα του αυτιού στο επόμενο.

Η ενίσχυση των ήχων επιτυγχάνεται εν μέρει λόγω της κατασκευής και του τρόπου σύνδεσης των οσταρίων, με τον άκμονα να λειτουργεί σαν μοχλός, μεταφέροντας την πίεση από τη μακριά σφύρα στον κατά πολύ μικρότερο αναβολέα. Ωστόσο, η ενίσχυση που τελικά προκύπτει με αυτόν τον τρόπο είναι της τάξης των 5 dB. Το μεγαλύτερο μέρος της προέρχεται από τη σχέση των εμβαδών της τυμπανικής μεμβράνης και της ωσειδούς θυρίδας. Η τυμπανική μεμβράνη είναι περίπου 0.85 mm^2 , με ένα 65% αυτής να συμμετέχει

στις ταλαντώσεις, ενώ το εμβαδόν της ωοειδούς θυρίδας είναι 0.03 mm^2 . Εφόσον η δύναμη που εξασκείται ισούται με το γινόμενο της επιφάνειας επί της πίεσης ($F = pA$) και αυτή η δύναμη παραμένει σταθερή, συνεπάγεται ότι η πίεση που θα ασκείται τελικά στην ωοειδή θυρίδα θα είναι ενισχυμένη κατά $\frac{0.55}{0.33} = 18$ φορές ή 25 dB.

Μία άλλη λειτουργία του μέσου ωτός είναι η προστασία από δυνατούς ήχους μέσω του ακουστικού αντανακλαστικού (acoustic reflex). Ήχοι άνω των 90 dB προκαλούν τη σύσπαση του μυός του αναβολέα (stapedius muscle) ο οποίος προφανώς είναι συνδεδεμένος με τον αναβολέα. Η σύσπαση του προκαλεί τη μεταβολή της γωνίας ζεύξης του αναβολέα με την ωοειδή θύρα μειώνοντας έτσι την εφαρμοζόμενη πίεση ή, κατά μία άλλη θεωρία, την σκλήρυνση της αλυσίδας των τριών οσταρίων με το ίδιο αποτέλεσμα και πάλι. Η αντίδραση αυτού του μυός παρουσιάζει μία καθυστέρηση μερικών millisecond που έχει ως συνέπεια την αδυναμία προστασίας από απότομους, δυνατούς ήχους. Η έκθεση σε δυνατούς ήχους μακράς διαρκείας κουράζει τον μυ του αναβολέα η αποτελεσματικότητα του οποίου αρχίζει να μειώνεται σταδιακά, επιτρέποντας την είσοδο των ήχων χωρίς μείωση της ισχύος τους.

2.1.1.3 Το έσω ους

Το έσω ους αποτελείται από τον κοχλία (cochlea), την αίθουσα (vestibule) και τους ημικυκλικούς σωλήνες (semicircular canals). Τα αισθητικά όργανα της ακοής βρίσκονται εντός του κοχλία στον οποίο και θα επικεντρωθούμε ενώ η αίθουσα και οι ημικυκλικοί σωλήνες είναι όργανα ισορροπίας. Ο κοχλίας είναι ένα σπειροειδές όργανο, σαν σαλιγκάρι, που περιστρέφεται γύρω από έναν κεντρικό άξονα (modiolus) σχεδόν τρεις φορές. Αν τμήσουμε κάθετα τον ξεδιπλωμένο κοχλία μπορούμε να παρατηρήσουμε την εσωτερική του δομή, όπως φαίνεται στο σχήμα 2.4. Το εσωτερικό του το διασχίζουν δύο μεμβράνες, η βασική μεμβράνη (basilar membrane) και η μεμβράνη του Reissner (Reissner's membrane) οι οποίες τον διαμερίζουν σε τρία τμήματα. Τα δύο από αυτά, η αιθουσαία κλίμακα (scala vestibuli) και η τυμπανική κλίμακα (scala timpani) περιέχουν εντός τους ένα υγρό πλούσιο σε Na^+ , την έξω λέμφο (perilymph). Το τρίτο διαμέρισμα ονομάζεται κοχλιακός πόρος (scala media) και στο εσωτερικό του κυκλοφορεί ένα άλλο υγρό με υψηλή συγκέντρωση K^+ , η έσω λέμφος (endolymph). Η βασική μεμβράνη διατρέχει τον κοχλία από τη βάση του, όπου είναι πιο στενή και σκληρή, μέχρι την κορυφή του όπου γίνεται πιο πλατιά και χαλαρή (βλ. σχήμα 2.3).



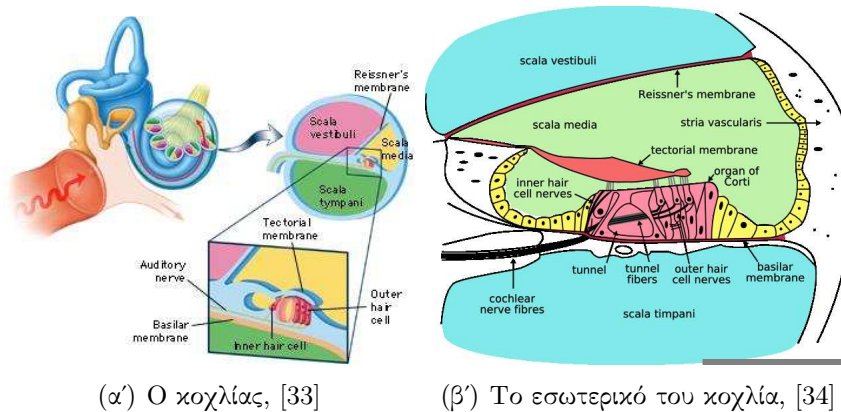
Σχήμα 2.3: Η βασική μεμβράνη, [32]

Αυτή η ιδιότητα της βασικής μεμβράνης την κάνει να έχει μεγαλύτερη απόκριση στις χαμηλές συχνότητες στα σημεία όπου είναι πιο πλατιά, κοντά στην κορυφή, ενώ οι υψηλές συχνότητες διεγείρουν περισσότερο τα στενότερα τμήματά της στη βάση. Το όργανο ακοής που μετατρέπει τις ταλαντώσεις σε ηλεκτρικά σήματα είναι το όργανο του Corti (organ of Corti) που στηρίζεται επί της βασικής μεμβράνης. Κάθε όργανο του Corti διαθέτει τα λεγόμενα στυλοειδή κύτταρα (pillar cells) τα οποία ενώνονται στην κορυφή τους σχηματίζοντας τη διάδο του Corti (tunnel of Corti). Δίπλα από τα στυλοειδή κύτταρα,

Κεφάλαιο 2. Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα

προς την πλευρά του κοχλιακού άξονα, υπάρχει ένα εσωτερικό τριχοφόρο κύτταρο (inner hair cell) που φέρει κάποια κύτταρα - “τρίχες” ή στερεοβλεφαρίδες (stereocilia). Από την άλλη πλευρά υπάρχουν τρία ή τέσσερα εξωτερικά τριχοφόρα κύτταρα (outer hair cell) που φέρουν και αυτά στερεοβλεφαρίδες. Πάνω από τη δίοδο του Corti και τα τριχοφόρα κύτταρα βρίσκεται η οροφιαία μεμβράνη (tectorial membrane) η οποία συνδέεται με τα υψηλότερα των εξωτερικών τριχοφόρων κυττάρων μέσω των στερεοβλεφαρίδων τους.

Ο τρόπος που τα ηχητικά σήματα μετατρέπονται σε ηλεκτρικά δεν είναι γνωστός με ακρίβεια. Οι μηχανικές ταλαντώσεις του αναβολέα στο μέσο ους εισέρχονται στον κοχλία μέσω της ωοειδούς θυρίδας όπου μετατρέπονται σε υδραυλικές ταλαντώσεις της έσω λέμφου και ανέρχονται την αιθουσαία κλίμακα. Στη συνέχεια, πάλι μέσω της έσω λέμφου, κατέρχονται την τυμπανική κλίμακα και εξέρχονται από την στρογγυλή θυρίδα (round window) (βλ. σχήμα 2.2). Αυτά τα κύματα θέτουν σε κίνηση και την βασική μεμβράνη η οποία με τη σειρά της μετατοπίζει τις στερεοβλεφαρίδες των εξωτερικών τριχοφόρων κυττάρων με αποτέλεσμα να ανοίγουν τα κανάλια ιόντων καλίου των στερεοβλεφαρίδων. Αυτό επιτρέπει στα ιόντα καλίου της έσω λέμφου να εισέλθουν στα εσωτερικά τριχοφόρα κύτταρα και να τα εκπολώσουν. Αυτά εν συνεχεία απελευθερώνουν διαβιβαστές στη σχισμή μεταξύ αυτών και των προσαγωγών νευρικών ιών, παράγοντας κατ’ αυτόν τον τρόπο δυναμικά δράσης που μεταφέρουν την πληροφορία στο ακουστικό νεύρο.

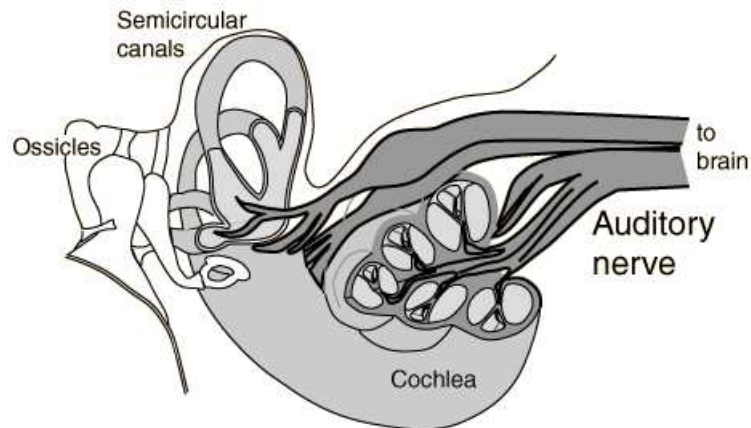


(α') Ο κοχλίας, [33]

(β') Το εσωτερικό του κοχλίας, [34]

Σχήμα 2.4: Κοχλίας

Οι προσαγωγές νευρικές ίνες του ακουστικού νεύρου που νευρώνουν το όργανο του Corti προέρχονται από το σπειροειδές γάγγλιο (spiral ganglion) που βρίσκεται εντός του οστεώδους κοχλιακού άξονα. Αποτελείται από διπολικά κύτταρα που συνδέουν τα τριχοφόρα κύτταρα με το εγκεφαλικό στέλεχος (brainstem). Το ακουστικό νεύρο περιλαμβάνει επίσης και απαγωγές ίνες που μεταφέρουν σήματα από το στέλεχος στον κοχλία. Το μεγαλύτερο ποσοστό κυττάρων του ακουστικού νεύρου είναι γαγγλιακά κύτταρα Τύπου I και Τύπου II. Το 95% των προσαγωγών ιών είναι κύτταρα Τύπου I, οι περιφερειακοί νευρίτες των οποίων σχηματίζουν συνάψεις με τα εσωτερικά τριχοφόρα κύτταρα ενώ το υπόλοιπο 5% είναι κύτταρα Τύπου II που νευρώνουν εξωτερικά τριχοφόρα κύτταρα. Επομένως, το μεγαλύτερο ποσοστό της ακουστικής πληροφορίας μεταφέρεται στον εγκέφαλο μέσω αξόνων Τύπου I. Η λειτουργία των ιών Τύπου II παραμένει ακόμα άγνωστη ([18]). Οι απαγωγές νευρικές ίνες από το εγκεφαλικό στέλεχος στον κοχλία σχηματίζουν συνάψεις,



Σχήμα 2.5: Μεταφέροντας το ηχητικό σήμα από τον κοχλία στο ακουστικό νεύρο, [35]

κυρίως χολινεργικής φύσης, με τα εξωτερικά τριχοφόρα κύτταρα αλλά και με προσαγωγές ίνες των εσωτερικών τριχοφόρων κυττάρων. Τα εξωτερικά τριχοφόρα κύτταρα έχουν επίσης τη δυνατότητα να υφίστανται μεταβολές του μήκους τους κατόπιν ηλεκτρικής διέγερσης ή εφαρμογής ακετυλοχολίνης, γεγονός που υποδεικνύει, σε συνδυασμό με τα παραπάνω, ότι αυτά τα κύτταρα επιτελούν λειτουργίες ενίσχυσης στον κοχλία και ελέγχου του κέρδους.

2.1.2 Από τον κοχλία στον ακουστικό φλοιό

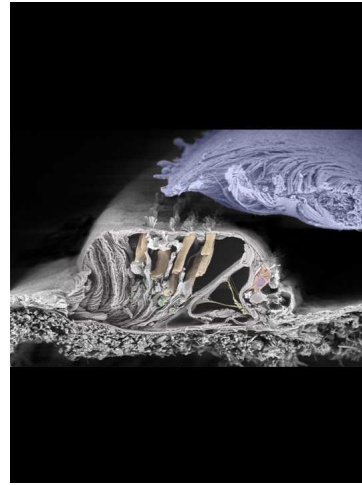
Από τη στιγμή που ένα ηχητικό σήμα αποκτά την αρχική του νευρωνική κωδικοποίηση στον κοχλία και μέχρι να φτάσει στον ακουστικό φλοιό πρέπει να διέλθει από αρκετούς σταθμούς μέσα στον εγκέφαλο που συνήθως ονομάζονται πυρήνες (nuclei), δηλαδή ομάδες νευρώνων που μοιράζονται μία κοινή λειτουργία. Η άνοδος προς τον ακουστικό φλοιό πραγματοποιείται φυσικά μέσω δύο οδών που εκκινούν από τον κοχλία κάθε αυτιού και σε πολλά σημεία της πορείας τους αλληλεπιδρούν. Ένα ιδιαίτερο χαρακτηριστικό του ακουστικού συστήματος είναι, όπως έχει ήδη αναφερθεί, ότι διαθέτει και νευρικές οδούς που κατέρχονται από το κεντρικό νευρικό σύστημα προς τα αισθητήρια όργανα. Στη συνέχεια αυτής της ενότητας γίνεται μία περιγραφή, σταθμό προς σταθμό, των βημάτων που ακολουθούν τα κωδικοποιημένα ηχητικά σήματα μέχρι τον προορισμό τους στον φλοιό. Το σχήμα 2.7 δείχνει μία εικόνα του κεντρικού νευρικού συστήματος. Μερικές από τις δομές που απεικονίζονται σε αυτό το σχήμα παίζουν σημαντικό ρόλο στην επεξεργασία των ακουστικών σημάτων.

2.1.2.1 Από τον κοχλία στο σύμπλεγμα κοχλιακών πυρήνων

Οι περίπου 30.000 νευρικές ίνες από τον κοχλία, μαζί με αυτές από τους ημικυκλικούς σωλήνες, συνενώνονται για να σχηματίσουν το ακουστικό νεύρο (auditory nerve ή vestibulocochlear nerve) ή όγδοο κρανιακό νεύρο (CN VIII) (βλ. σχήμα 2.5). Μέσω του έσω ακουστικού πόρου (internal auditory meatus) εισέρχονται στο εγκεφαλικό στέλεχος



(α') Ο κοχλίας ινδικού χοιριδίου, [36]



(β') Ανθρώπινο όργανο του Corti, [37]

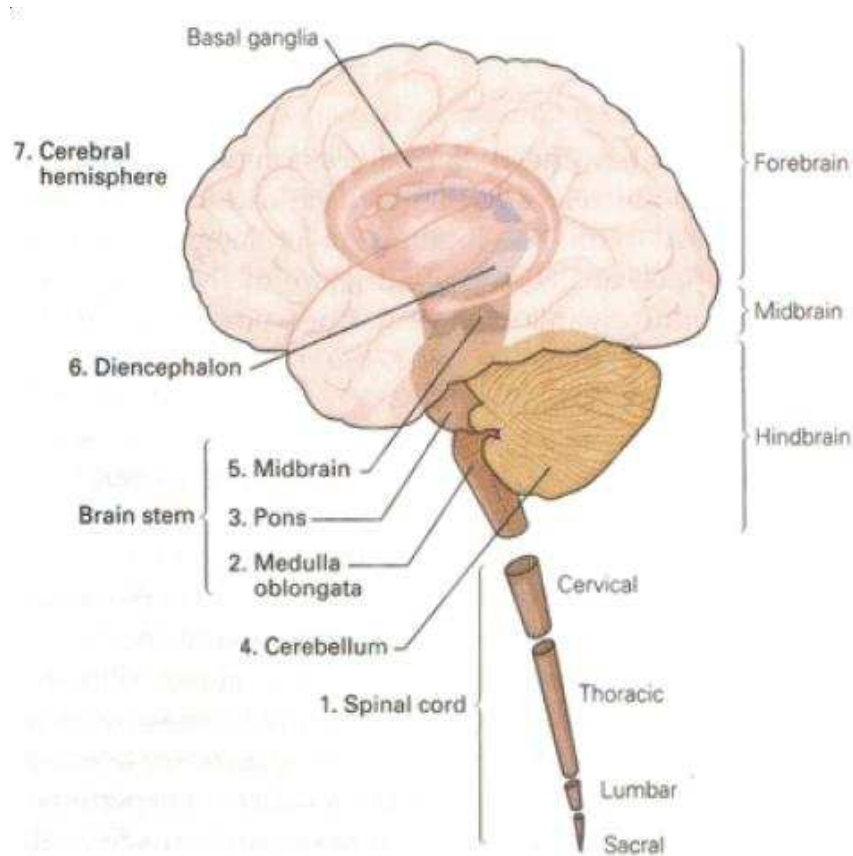
Σχήμα 2.6: Φωτογραφίες από ηλεκτρονικό μικροσκόπιο σάρωσης

(brainstem), στα ανώτερα επίπεδα του προμήκους μυελού (medulla), όπου σχηματίζουν συνάψεις στο σύμπλεγμα κοχλιακών πυρήνων (cochlear nuclear complex). Στο σχήμα 2.8 παρουσιάζονται μερικές πρόσθιες όψεις του εγκεφαλικού στελέχους και της παρεγκεφαλίδας με σημειωμένες τις δομές που αναφέρθηκαν προηγουμένως.

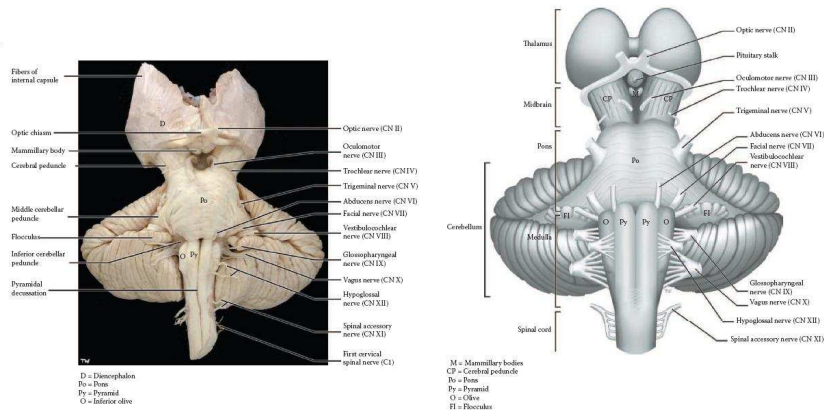
Το σύμπλεγμα κοχλιακών πυρήνων βρίσκεται στην επιφάνεια του στελέχους, κοντά στο σύνορο μεταξύ του προμήκους μυελού και της γέφυρας (pons). Οι Τύπου I ίνες του ακουστικού νεύρου προβάλλουν επί του εμπρόσθιου και του οπίσθιου κοχλιακού πυρήνα (ventral/dorsal cochlear nuclei) (βλ. σχήμα 2.9) όπου συνδέονται με διάφορους τύπους κυττάρων, όπως πολυπολικά, σφαιρικά, γιγάντια κτλ. Οι προβολές του ακουστικού νεύρου σε αυτούς τους πυρήνες πραγματοποιούνται με έναν κοχλιοτοπικό, επομένως και τονοτοπικό, τρόπο. Είναι αξιοσημείωτο ότι, από τους χιλιάδες νευρώνες των κοχλιακών πυρήνων, μόνο μερικοί από αυτούς συνδέονται με ίνες του ακουστικού νεύρου. Οι περίπου 30.000 άξονες του ακουστικού νεύρου είναι αρκετά λιγότεροι από τους νευρώνες των κοχλιακών πυρήνων γεγονός που δεν αποκλείει φυσικά τη δυνατότητα σύνδεσης μεταξύ αυτών των νευρώνων. Οι Τύπου I προσαγωγοί χρησιμοποιούν γλουταμινικό οξύ (glutamate) ως διεγερτικό νευροδιαβιβαστή αλλά υπάρχουν ενδείξεις ότι λαμβάνει χώρα και ανασταλτική δράση επί των νευρώνων του κοχλιακού συμπλέγματος.

2.1.2.2 Από το σύμπλεγμα κοχλιακών πυρήνων στο άνω ελαιώδες σύμπλεγμα

Ο επόμενος σταθμός των νευρικών σημάτων είναι το άνω ελαιώδες σύμπλεγμα (superior olivary complex) στο οποίο καταλήγουν οι περισσότερες (αλλά όχι όλες) ίνες που αναχωρούν από το κοχλιακό σύμπλεγμα. Αποτελείται κυρίως από δύο πυρήνες, την πλευρική άνω ελαία (ΠΑΕ) (lateral superior olive, LSO) και την μέση άνω ελαία (ΜΑΕ) (medial superior olive, MSO) και βρίσκεται στα κάτω επίπεδα της γέφυρας, κοντά στο έβδομο νεύρο,



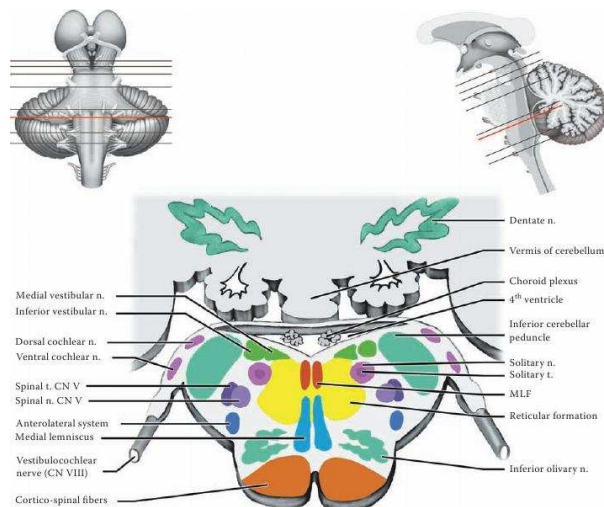
Σχήμα 2.7: Το κεντρικό νευρικό σύστημα, [12]



(α) Φωτογραφία πρόσθιας όψης του (β') Πρόσθια όψη στελέχους με κρανιακά νεύρα

Σχήμα 2.8: Πρόσθιες όψεις εγκεφαλικού στελέχους, [10]

Κεφάλαιο 2. Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα

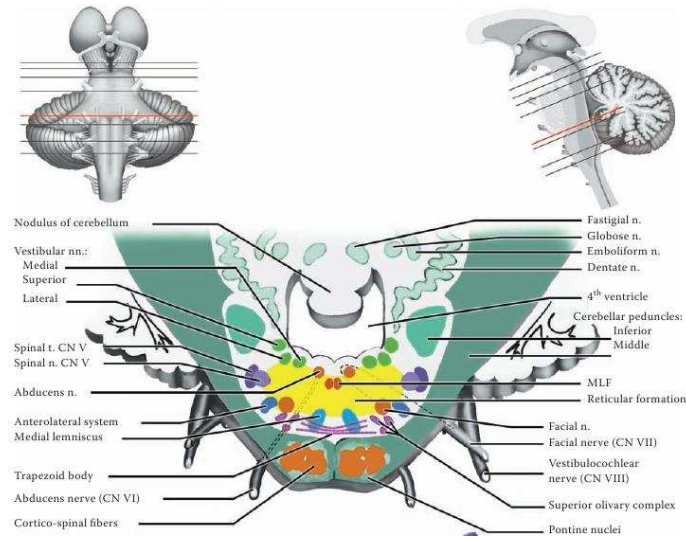


Σχήμα 2.9: Τομή ανώτερου προμήκους μυελού, [10]

αυτό του προσώπου (Facial nerve, CN VII). Οι νευρώνες της ΠΑΕ έχουν την ικανότητα να εντοπίζουν διαφορές έντασης μεταξύ των δύο αυτιών ενώ αυτοί της ΜΑΕ μπορούν να εντοπίζουν χρονικές διαφορές από το ίδιο αυτί. Στον άνθρωπο, το άνω ελαιώδες σύμπλεγμα αποτελείται κυρίως από την ΜΑΕ η οποία λαμβάνει σήματα από τον εμπρόσθιο κοχλιακό πυρήνα. Οι ίνες που έρχονται από το κοχλιακό σύμπλεγμα δεν είναι απαραίτητο να εκκινούν από την ίδια πλευρά. Μερικές από αυτές περνούν στην άλλη πλευρά μέσω μίας δομής που ονομάζεται τραπεζοειδές σώμα (trapezoid body) (βλ. σχήμα 2.10). Τα κύτταρα της ΠΑΕ διεγείρονται από το ομόπλευρο αυτί ενώ αναστέλλονται από το ετερόπλευρο αυτί μέσω των ινών που διαπερνούν το τραπεζοειδές σώμα. Μερικές από τις ίνες του τραπεζοειδούς σώματος συνεχίζουν προς τον πλευρικό λημνίσκο (lateral lemniscus) αντί να τερματίζουν στο ετερόπλευρο ελαιώδες σύμπλεγμα. Θεωρείται ότι η κύρια λειτουργία του άνω ελαιώδους συμπλέγματος είναι ο εντοπισμός της προέλευσης των ήχων εφόσον είναι η πρώτη δομή στην οποία συγκλίνουν πληροφορίες και από τα δύο αυτιά στα οποία ένα ίδιο ηχητικό σήμα φθάνει πάντα με κάποια χρονική διαφορά. Αυτή τη διαφορά μπορεί να τη χρησιμοποιήσει το συγκεκριμένο σύμπλεγμα για να υπολογίσει την προέλευση.

2.1.2.3 Ο πλευρικός λημνίσκος

Οι ίνες που ανέρχονται από το άνω ελαιώδες σύμπλεγμα, είτε από την ίδια πλευρά είτε από την αντίθετη, διαμέσου του τραπεζοειδούς σώματος, σχηματίζουν μία οδό, τον πλευρικό λημνίσκο (βλ. σχήμα 2.11). Οι δύο λημνίσκοι διασυνδέονται μεταξύ τους σε διάφορα σημεία κατά την άνοδό τους πριν καταλήξουν να τροφοδοτούν με ακουστική πληροφορία το κάτω διδύμιο (inferior colliculus) του μεσεγκεφάλου. Με βάση δεδομένα από γάτες, φαίνεται ότι ο πλευρικός λημνίσκος αποτελείται από τρεις πυρήνες, έναν οπίσθιο, έναν διάμεσο και έναν εμπρόσθιο. Ο οπίσθιος λαμβάνει διωτικά (binaural) δεδομένα από το άνω ελαιώδες σύμπλεγμα ενώ οι άλλοι δύο μονοωτικά από το τραπεζοειδές σώμα και το κοχλιακό σύμπλεγμα. Ο οπίσθιος πυρήνας του λημνίσκου προβάλλει επίσης και στο άνω



Σχήμα 2.10: Τομή κατώτερης γέφυρας, [10]

διδύμιο (superior colliculus).

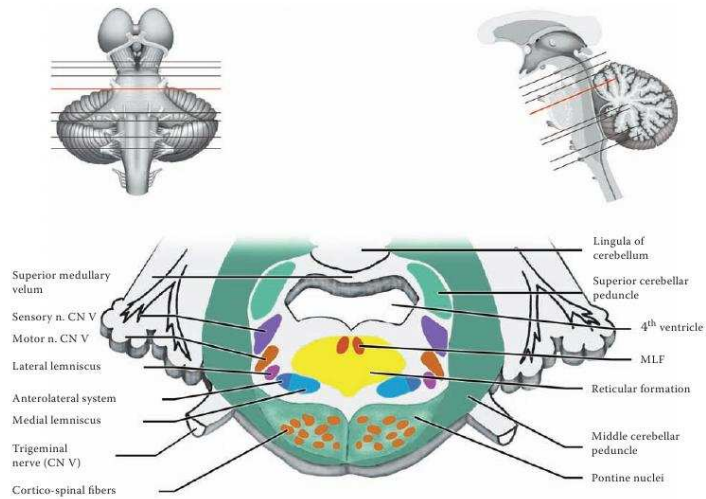
2.1.2.4 Το κάτω διδύμιο

Οι περισσότεροι άξονες του πλευρικού λημνίσκου τερματίζουν στο κάτω διδύμιο που αποτελείται από έναν κεντρικό πυρήνα και δύο φλοιούς, έναν εξωτερικό φλοιό και έναν οπίσθιο φλοιό. Στο επίπεδο του κάτω διδυμίου έχουμε και πάλι συνδέσεις μεταξύ των δύο πλευρών. Ο εξωτερικός φλοιός του διδυμίου λαμβάνει ακουστικά και σωματαιοσθητικά ερεθίσματα και προβάλλει στο μέσο τμήμα του έσω γονατώδους σώματος (medial geniculate body) στον θάλαμο. Ο οπίσθιος φλοιός δέχεται συνάψεις από τον ακουστικό φλοιό και προβάλλει στο οπίσθιο τμήμα του μέσου γονατώδους σώματος. Το κάτω διδύμιο αποστέλλει επίσης πληροφορίες και στο άνω διδύμιο όπου κατασκευάζονται και συγκρίνονται ακουστικοί και οπτικοί χάρτες. Εξ' αυτού του λόγου θεωρείται ότι μία σημαντική λειτουργία του κάτω διδυμίου είναι η χαρτογράφηση του ακουστικού χώρου η οποία μπορεί να γίνεται με αρκετά πολύπλοκο τρόπο και με βάση διάφορα ακουστικά χαρακτηριστικά, όπως η ένταση του σήματος ή η βέλτιστη συχνότητα. Επειδή σχεδόν όλες οι ανερχόμενες (αλλά και κατερχόμενες) νευρικές οδοί τελικά προβάλλουν στο κάτω διδύμιο θεωρείται ότι αυτό αποτελεί ένα σημείο σύγκλισης ακουστικών πληροφοριών που έχουν προηγουμένως υποστεί παράλληλη επεξεργασία σε άλλους πυρήνες του εγκεφαλικού στελέχους.

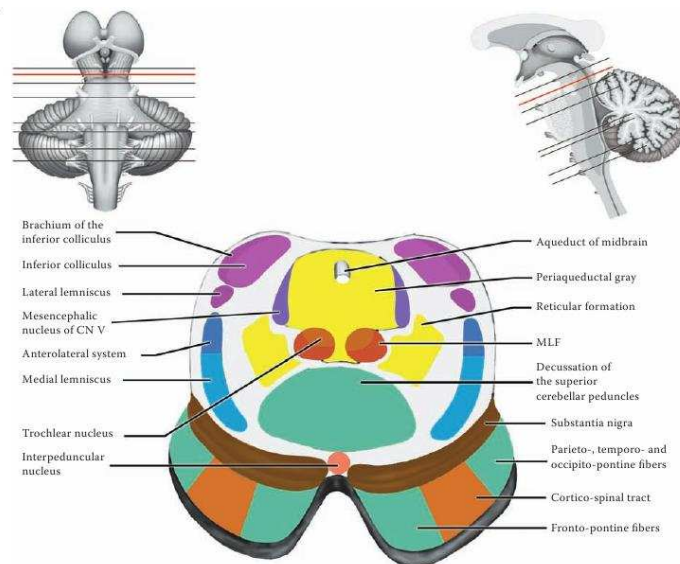
2.1.2.5 Το έσω γονατώδες σώμα

Η επόμενη στάση των ηχητικών σημάτων είναι στο έσω γονατώδες σώμα του θαλάμου (βλ. σχήμα 2.13) Από τον θάλαμο περνάνε σχεδόν όλες οι αισθητικές οδοί στην πορεία τους προς το τμήμα του φλοιού που τους αντιστοιχεί. Ο θάλαμος μπορεί να διαχωριστεί σε επιμέρους τμήματα καθένα από τα οποία είναι υπεύθυνο για ένα συγκεκριμένο αισθητικό

Κεφάλαιο 2. Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα

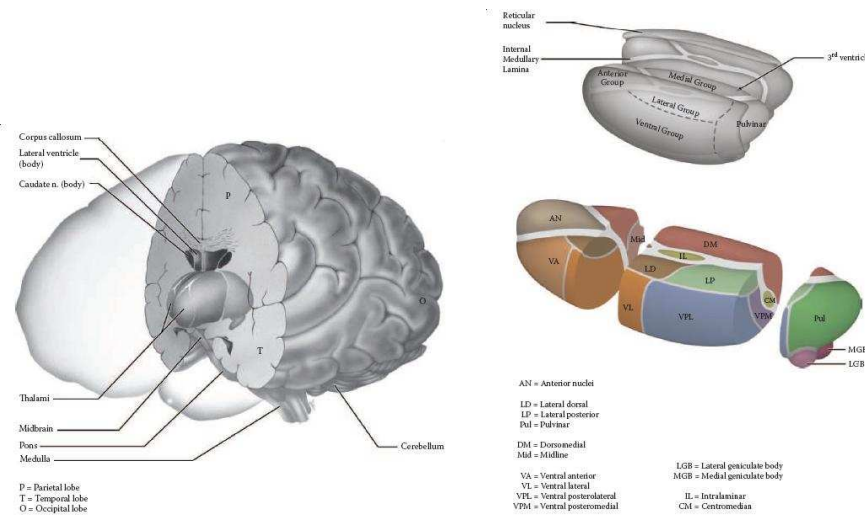


Σχήμα 2.11: Τομή μέσης γέφυρας, [10]



Σχήμα 2.12: Τομή κατώτερου μεσεγκεφάλου, [10]

σύστημα. Μόνο το οσφρητικό σύστημα παρακάμπτει τον θάλαμο μεταβιβάζοντας πληροφορίες απευθείας στον έσω κροταφικό λοβό.



(α') Η θέση του θαλάμου στον εγκέφαλο (β') Οι πυρήνες του θαλάμου

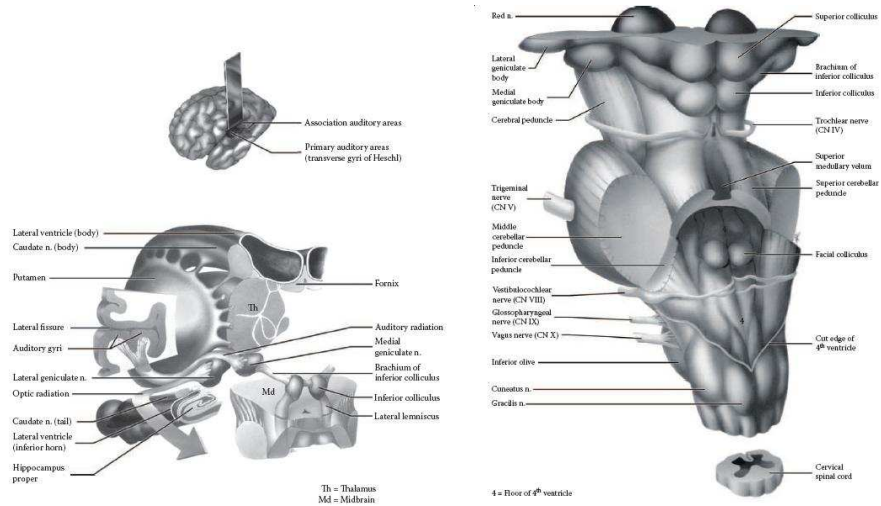
Σχήμα 2.13: Ο θάλαμος, [10]

Το κάτω διδύμιο συνδέεται με το έσω γονατώδες σώμα μέσω του βράχιου (brachium) (βλ. σχήμα 2.14). Το γονατώδες σώμα είναι σαν μία μικρή προεξοχή του θαλάμου, αποτελείται από τρεις κύριους πυρήνες, τον πρόσθιο, τον μέσο και τον οπίσθιο και παρουσιάζει αμφίδρομες συνδέσεις με τον ακουστικό φλοιό. Κύτταρα του πρόσθιου τμήματος προβάλλουν κατά κύριο λόγο στον πρωτοταγή ακουστικό φλοιό (AI) και λαμβάνουν σήματα από τον κεντρικό πυρήνα και τον οπίσθιο φλοιό του κάτω διδυμίου. Αυτά που είναι κοντύτερα στο μέσο τμήμα προβάλλουν στον δευτεροταγή ακουστικό φλοιό (AII) και δέχονται εισόδους από τον κεντρικό και εξωτερικό φλοιό του κάτω διδυμίου, από σωματισθητικές και οπτικές πηγές αλλά και από βαθύτερα στρώματα του άνω διδυμίου. Ο οπίσθιος πυρήνας τέλος δέχεται πληροφορίες από φλοιικές περιοχές του κάτω διδυμίου και από άλλους πυρήνες του εγκεφαλικού στελέχους και προβάλλει σε περιοχές που περιβάλλουν τον πρωτοταγή ακουστικό φλοιό. Σε αυτό το σημείο φαίνεται ότι ήδη το ηχητικό σήμα έχει υποστεί σημαντική επεξεργασία. Έχουν βρεθεί κύτταρα του έσω γονατώδους σώματος που αποκρίνονται σε πολύπλοκα σήματα αλλά όχι σε απλούς τόνους (π.χ. συγκεκριμένα “κελαηδήματα” σε ορισμένα είδη πτηνών).

2.1.2.6 Ο ακουστικός φλοιός

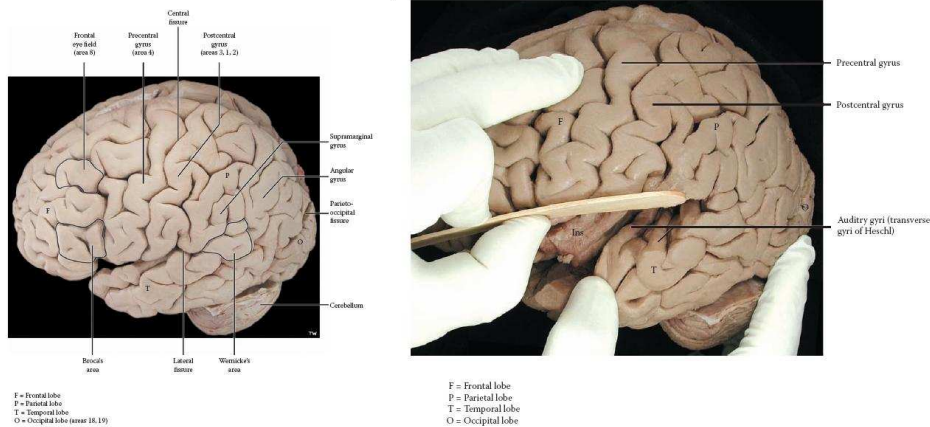
Στο επόμενο και τελευταίο στάδιο, οι νευρώνες του μέσου γονατώδους σώματος προβάλλουν στον ακουστικό φλοιό. Ο πρωτοταγής ακουστικός φλοιός (AI) βρίσκεται στον κροταφικό λοβό, στην περιοχή 41 που ονομάζεται και εγκάρσιες έλικες του Heschl (transverse gyri of Heschl), εντός της πλευρικής σχισμής (lateral fissure) (βλ. σχήμα 2.15).

Κεφάλαιο 2. Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα



(α') Από το κάτω διδύμο στο μέσο γονα- (β') Οπίσθια όψη στελέχους χωρίς την τώδες σώμα παρεγκεφαλίδα

Σχήμα 2.14: Προς τον θάλαμο, [10]



(α') Φωτογραφία ημισφαιρίου

(β') Οι εγκάρσιες έλικες του Heschl

Σχήμα 2.15: Φωτογραφίες ακουστικού φλοιού, [10]

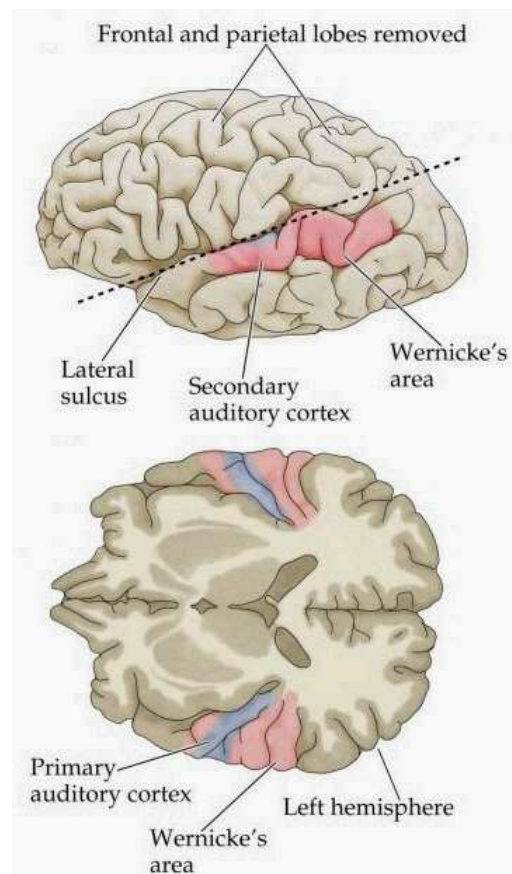
Ο ΑΙ αποτελείται από έξι στρώματα, όπως και άλλες αισθητικές περιοχές του φλοιού, με υψηλή πυκνότητα κοκκοειδών κυττάρων, με το στρώμα I να συνορεύει με τη χοριοειδή μήνιγγα (pia mater) και το στρώμα VI με τη λευκή ουσία (white matter). Περίπου το ένα τέταρτο των κυττάρων του στις γάτες είναι GABA-εργικά και επομένως ανασταλτικά (ο GABA είναι ανασταλτικός νευροδιαβιβαστής). Είναι αμφίδρομα συνδεδεμένος με τους πυρήνες του μέσου γονατώδους σώματος και λαμβάνει σήματα και από περιοχές του θαλάμου. Από τον πρόσθιο πυρήνα του μέσου γονατώδους σώματος προβάλλουν προσαγωγοί στο στρώμα IV του ΑΙ. Ο ΑΙ είναι οργανωμένος κατά τονοτοπικό τρόπο, όπου συγκεκριμένες περιοχές - στήλες του είναι ευαίσθητες σε συγκεκριμένα εύρη συχνοτήτων. Κάθε ισοσυχνοτική στήλη περιλαμβάνει νευρώνες που είτε διεγείρονται και από τα δύο αυτιά είτε νευρώνες που διεγείρονται από το ετερόπλευρο αυτί και αναστέλλονται από το ομόπλευρο. Κάποιες μελέτες έχουν δείξει ότι ενδεχομένως και άλλες περιοχές του ακουστικού φλοιού παρουσιάζουν οργάνωση σε τονοτοπικούς άξονες με διάφορους βαθμούς συχνοτικής διακριτικής ικανότητας, όπως ο δευτεροταγής ακουστικός φλοιός, ΑΙΙ. Αυτές οι περιοχές εκτείνονται γύρω από τον ΑΙ και λαμβάνουν και αυτές δεδομένα από το μέσο γονατώδες σώμα. Μεταξύ του ΑΙ και του ΑΙΙ σχηματίζονται αμφίδρομες συνδέσεις καθώς επίσης και μεταξύ των αντίστοιχων φλοιικών περιοχών των δύο ημισφαιρίων. Γενικά, ο βαθμός διασυνδεσιμότητας του ακουστικού φλοιού είναι αρκετά υψηλός αλλά ακριβείς λεπτομέρειες δεν είναι προς το παρόν γνωστές. Κάποιες μελέτες έχουν δείξει ότι υπάρχει μία τάση για σύνδεση μεταξύ φλοιικών περιοχών με παρόμοιες χαρακτηριστικές συχνότητες ([7],p.98).

Ασυμμετρίες μεταξύ των δύο ημισφαιρίων, όπως αυτή της περιοχής του Wernicke δεν είναι ασυνήθιστες στον ακουστικό φλοιό η κύρια περιοχή του οποίου είναι μεγαλύτερη στο αριστερό ημισφαίριο σε σχέση με το δεξί. Σημαντική διαφοροποίηση παρατηρείται και στην κυτταρική οργάνωση των δύο ημισφαιρίων με το αριστερό ημισφαίριο να παρουσιάζει αυξημένο αριθμό από μεγαλύτερα κύτταρα, εντονότερα μυελνωμένα και με εκτενέστερες διασυνδέσεις. Παρ' όλα αυτά δεν είναι βέβαιο ότι αυτές οι ανατομικές ασυμμετρίες αντικατοπτρίζουν και λειτουργικές ασυμμετρίες όπως μία εξειδίκευση του αριστερού ημισφαιρίου στην επεξεργασία ταχέως μεταβαλλόμενων ήχων (π.χ. στον λόγο) και του δεξιού ημισφαιρίου σε βραδύτερους αλλά με καλύτερη συχνοτική διακριτική ικανότητα ήχους.

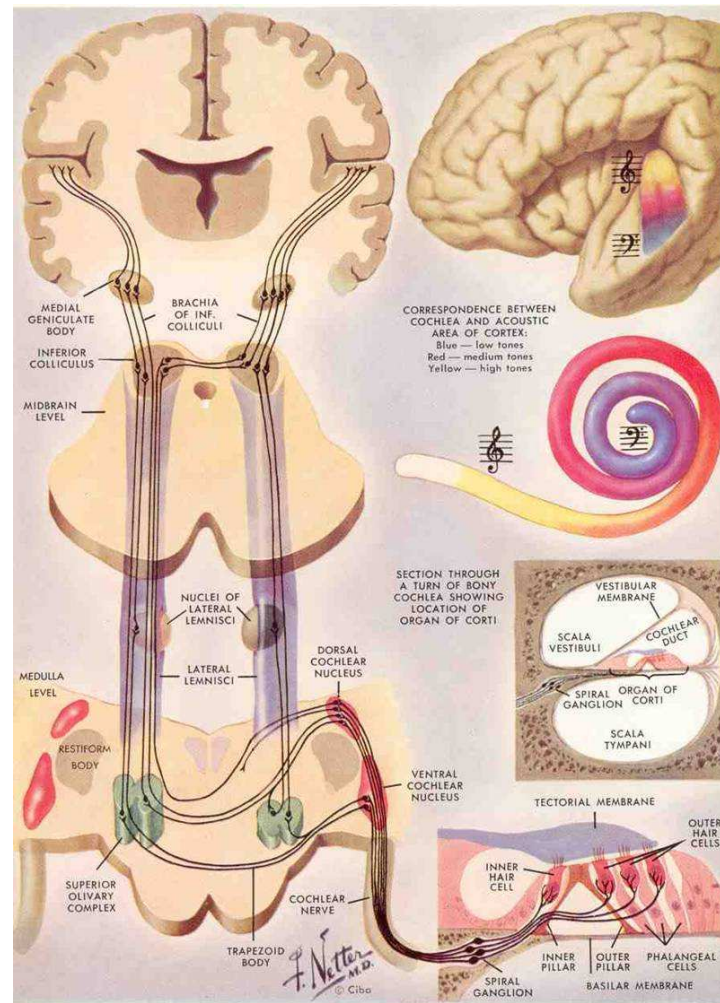
Η όλη διαδρομή των ηχητικών σημάτων μέσα στο νευρικό σύστημα μπορεί να συνοψιστεί στο σχήμα 2.17.

2.1.2.7 Ο μηχανισμός ανάδρασης του ακουστικού συστήματος

Ένα άλλο χαρακτηριστικό του ακουστικού συστήματος, όπως και των υπόλοιπων αισθητηριακών συστημάτων, είναι ότι διαθέτει και οδούς ανάδρασης που κατέρχονται από τα ανώτερα επίπεδα προς τα κατώτερα. Φλοιικές περιοχές προβάλλουν στο ομόπλευρο μέσο γονατώδες σώμα και στο κάτω διδύμιο. Το κάτω διδύμιο στέλνει απαγωγούς τόσο στο ελαιώδες σύμπλεγμα όσο και σε κοχλιακούς πυρήνες σχηματίζοντας τη λεγόμενη ελαιοκοχλιακή ανάδραση. Άλλες ίνες από το ελαιώδες σύμπλεγμα σχηματίζουν την ελαιοκοχλιακή δεσμίδα (olivocochlear bundle) και φτάνουν σε ομόπλευρα εσωτερικά τριχοφόρα κύτταρα ή σε ετερόπλευρα εξωτερικά τριχοφόρα κύτταρα του κοχλία. Αυτές οι κατερχόμενοι οδοί μπορούν να διαμορφώσουν την επεξεργασία των ακουστικών σημάτων σε διάφορα επίπεδα από το γονατώδες σώμα μέχρι και τον κοχλία. Φαίνεται ότι είναι σημαντικές στο φιλτράρισμα της ακουστικής πληροφορίας, ακόμα και στο στάδιο του κοχλία. Μία τέτοια λειτουργία μπορεί να διευκολύνει την επιτυχία του ανθρώπινου ακουστικού μηχανισμού



Σχήμα 2.16: Ο ακουστικός φλοιός, [38]



Σχήμα 2.17: Η διαδρομή των ηχητικών σημάτων στο νευρικό σύστημα, [39]

κατά το λεγόμενο φαινόμενο του cocktail party κατά το οποίο μπορεί να διαχωριστεί ο “θόρυβος” (πραγματικός ή άλλοι αδιάφοροι ομιλητές) από τον λόγο του ομιλητή που μας ενδιαφέρει και στον οποίο εστιάζουμε χωρίς καμμία σύγχυση. Σε μελέτες σε νυχτερίδες και αρουραίους έχει βρεθεί ότι τέτοιου είδους συνδέσεις ανάδρασης ενισχύουν την απόκριση νευρώνων στον θάλαμο και στο κάτω διδύμιο που έχουν το μέγιστό τους στις ίδιες τιμές μίας δεδομένης παραμέτρου ενώ αναστέλλουν νευρώνες με μέγιστο σε διαφορετικές τιμές αυτής της παραμέτρου ([7],p.102).

2.2 Λειτουργική περιγραφή

Οι μετασχηματισμοί που υφίσταται ένα ακουστικό σήμα ξεκινάνε ήδη από το έξω ους, πριν ακόμα γίνει η πρώτη νευρωνική κωδικοποίηση. Λόγω της δομής του ωτικού πτερυγίου και του έξω ακουστικού πόρου, συχνότητες μεταξύ 2.5 και 5 kHz ενισχύονται κατά σχεδόν 20 dB ενώ στο μέσο ους, λόγω των χαρακτηριστικών του αντίστασης, παρατηρείται εξασθένηση του ενεργειακού περιεχομένου σε συχνότητες κάτω των 500 Hz. Επομένως, η λειτουργία του έξω και μέσου ωτός μπορεί να μοντελοποιηθεί ως ένα ζωνοπερατό φίλτρο. Ο κοχλίας είναι μία συσκευή συντονισμού που αποσυνθέτει το σήμα στις φασματικές συνιστώσες του μέσω των εσωτερικών και εξωτερικών τριχοφόρων κυττάρων του. Όπως έχει ήδη αναφερθεί, τα εσωτερικά τριχοφόρα κύτταρα είναι αυτά που αποπολώνουν τους νευρώνες του ακουστικού νεύρου μεταφέροντας έτσι το σήμα προς τον εγκέφαλο ενώ τα εξωτερικά έχουν κυρίως ρόλο ενίσχυσης, ελέγχου του κέρδους και πιθανώς της περαιτέρω βελτίωσης της ικανότητας συντονισμού του κοχλίου. Ωστόσο, θα πρέπει να σημειωθεί ότι ο κοχλίας δεν είναι ένας απλός Fourier αναλυτής εφόσον η βασική μεμβράνη έχει και μη γραμμικές ιδιότητες. Η νευρωνική αναπαράσταση που προκύπτει από τον κοχλίο και μεταδίδεται στις ίνες του ακουστικού νεύρου είναι τοπική, με την έννοια ότι κάθε (περιορισμένο) εύρος συχνοτήτων ενεργοποιεί εντονότερα τον κοχλίο σε συγκεκριμένες περιοχές του και αντιστοίχως και τις ίνες του ακουστικού νεύρου που νευρώνουν αυτές τις περιοχές. Έτσι, κάθε περιοχή συχνοτήτων αποκτά τον “δικό” της τόπο αναπαράστασης στον κοχλίο και στο ακουστικό νεύρο, με τη βάση του κοχλίου να είναι πιο ευαίσθητη σε υψηλές συχνότητες, άνω των 10 kHz, την κορυφή να ανταποκρίνεται σε συχνότητες κάτω του 1 kHz και στις ενδιάμεσες περιοχές διαβαθμίζονται οι υπόλοιπες συχνότητες. Η χαρακτηριστική συχνότητα (characteristic frequency, CF) μίας ίνας είναι η συχνότητα στην οποία απαιτείται η ελάχιστη αύξηση των επιπέδων πίεσης (sound pressure level, SPL) ώστε να επιτευχθεί μία συγκεκριμένη, π.χ. 20%, αύξηση στη δραστηριότητα εκφόρτισης της ίνας. Σε αυτή τη συχνότητα θεωρούμε ότι η ίνα αντιδρά βέλτιστα. Διατηρώντας σταθερό το ποσοστό νευρωνικής δραστηριότητας που πρέπει να επιτευχθεί και μεταβάλλοντας το επίπεδο πίεσης συναρτήσει της συχνότητας λαμβάνουμε την συχνοτική καμπύλη συντονισμού (frequency tuning curve, FTC) μίας ίνας. Με βάση τα παραπάνω, αυτό το στάδιο ακουστικής επεξεργασίας μπορεί να μοντελοποιηθεί ως μία σειρά από ζωνοπερατά, μερικώς επικαλυπτόμενα φίλτρα που καλύπτουν όλο το εύρος συχνοτήτων που μπορούν να γίνουν αντιληπτές.

Η τονοτοπική οργάνωση που παρατηρείται στον κοχλίο εμφανίζεται και στα υπόλοιπα στάδια επεξεργασίας του ήχου στο νευρικό σύστημα με τον κάθε πυρήνα να μπορεί να διαχωριστεί σε ισοσυχνοτικές ζώνες, περιοχές που αντιδρούν εντονότερα σε τόνους συγκεκριμένων συχνοτήτων. Το πλάτος μίας συγκεκριμένης φασματικής συνιστώσας αντικατοπτρίζεται από το ρυθμό εκφόρτισης της ζώνης που της αντιστοιχεί. Πέρα από τη

συχνοτική ανάλυση, το ακουστικό σύστημα έχει την ικανότητα να εξάγει πληροφορίες και με βάση άλλες μεταβλητές του σήματος, όπως το πλάτος του και η χρονική του εξέλιξη. Μία τέτοια εναλλακτική αναπαράσταση που φαίνεται να λαμβάνει χώρα στο νευρικό σύστημα είναι η χρονική, αυτή δηλαδή που βασίζεται στη χρονική εξέλιξη των εκφορτίσεων. Σε χαμηλές συχνότητες, η δραστηριότητα των ινών του ακουστικού νεύρου δεν είναι τυχαία αλλά συμβαίνει ανά συγκεκριμένα χρονικά διαστήματα (φάσεις) ακολουθώντας το ρυθμό με τον οποίο εκφορτίζονται τα τριχοφόρα κύτταρα. Αυτό το φαινόμενο ονομάζεται κλείδωμα φάσης (phase locking) και παρατηρείται σε όλα τα σπονδυλωτά. Επομένως, ένας άλλος τρόπος νευρωνικής αναπαράστασης του ήχου μπορεί να επιτευχθεί με βάση τα χρονικά μοτίβα εξέλιξης της δραστηριότητας εκφόρτισης. Ωστόσο, η πληροφορία που κωδικοποιείται κατ' αυτόν τον τρόπο θα πρέπει να μετασχηματίζεται σε μία άλλη μορφή στα ανώτερα στάδια του νευρικού συστήματος καθώς η ικανότητα για κλείδωμα φάσης μειώνεται δραματικά, ήδη από το κοχλιακό σύμπλεγμα, με τα 100 Hz να αποτελούν το άνω όριο στον ακουστικό φλοιό. Παρ' όλα τα παραπάνω ευρήματα, δεν είναι βέβαιο ότι ο τονοτοπικός τρόπος οργάνωσης αντανακλά και μία λειτουργική οργάνωση όπου η επεξεργασία γίνεται με βάση τη συχνότητα εφόσον οι τονοτοπικοί άξονες που σχηματίζονται θα μπορούσαν να προκύπτουν απλώς από τον τρόπο ανάπτυξης των νευρικών ινών καθώς περνούν από το ένα στάδιο στο επόμενο. Επίσης, δεν έχουν εντοπιστεί ακόμα ανατομικές δομές οργανωμένες κατά τρόπο παρόμοιο με αυτό των ισοσυχνοτικών ζωνών που να κωδικοποιούν όμως άλλα χαρακτηριστικά των ηχητικών σημάτων εκτός της συχνότητας. Καθώς η έρευνα στρέφεται προς συγκεκριμένες κλάσεις κυττάρων παρόμοιας λειτουργικότητας που απαντώνται στις ακουστικές νευρικές οδούς, φαίνεται ότι όντως αυτές οργανώνονται σε ανατομικές δομές με βάση τη λειτουργία που επιτελούν. ([7],p.82).

Τα τελευταία χρόνια γίνεται όλο και πιο εμφανές ότι η αντιμετώπιση των σύνθετων ακουστικών σημάτων ως ένα απλό γραμμικό άθροισμα συχνοτικών συνιστωσών είναι ανεπαρκής για να περιγράψει την απόκριση των νευρώνων σε τέτοια ερεθίσματα. Αρχικά χρησιμοποιήθηκαν απλοί ημιτονοειδείς τόνοι ως ερέθισμα για να μετρηθούν οι ιδιότητες απόκρισης των ακουστικών νευρώνων. Τα πειραματικά αποτελέσματα όμως υποδεικνύουν ότι εμπλέκονται και μη γραμμικοί μηχανισμοί με την απόκριση των ινών να εξαρτάται από πολλαπλούς παράγοντες ταυτόχρονα. Επομένως, καθίσταται δύσκολη η εξαγωγή συμπερασμάτων και προβλέψεων με βάση τις αποκρίσεις σε απλά ερεθίσματα. Για αυτόν το λόγο, η έρευνα έχει στραφεί πλέον στη μελέτη της συμπεριφοράς των ακουστικών νευρώνων όταν αυτοί διεγείρονται από πιο πολύπλοκα ερεθίσματα, πολλές φορές δυναμικά μεταβαλλόμενα. Ένα άλλο χαρακτηριστικό του ακουστικού συστήματος που επίσης καθιστά περίπλοκη την ανάλυσή του είναι η παρουσία πολλαπλών, παράλληλων οδών επεξεργασίας που εκκινούν από το κοχλιακό σύμπλεγμα με την κάθemia να αποτελεί και ένα ξεχωριστό κανάλι επεξεργασίας. Προς το παρόν, υπάρχει ένα σημαντικό κενό στις γνώσεις μας για το πλήθος και το είδος των διαφόρων κλάσεων κυττάρων που συμμετέχουν σε αυτές τις παράλληλες ακουστικές οδούς, στον τρόπο που αυτά συνδέονται μεταξύ τους και στο πώς τελικά τερματίζουν στον ακουστικό φλοιό.

Όπως θα φανεί στη συνέχεια και παρά τις ελλειπείς γνώσεις που διαθέτουμε, το περιφερικό ακουστικό σύστημα είναι σε μεγάλο βαθμό δομημένο, παρέχοντας τη νευρωνική κωδικοποίηση στα επόμενα στάδια. Κάθε τέτοιο στάδιο επεξεργάζεται το εισερχόμενο σήμα εξάγοντας από αυτό συγκεκριμένες ιδιότητες τις οποίες μεταβιβάζει στη συνέχεια στα ανώτερα στάδια για περαιτέρω επεξεργασία.

Τα παραπάνω, καθώς και όσα θα ακολουθήσουν σε αυτή την ενότητα, έχουν βασιστεί σε μεγάλο βαθμό στο [18] αλλά και στο [7].

2.2.1 Μετασχηματισμοί των σημάτων στο πρώιμο, περιφερικό ακουστικό σύστημα

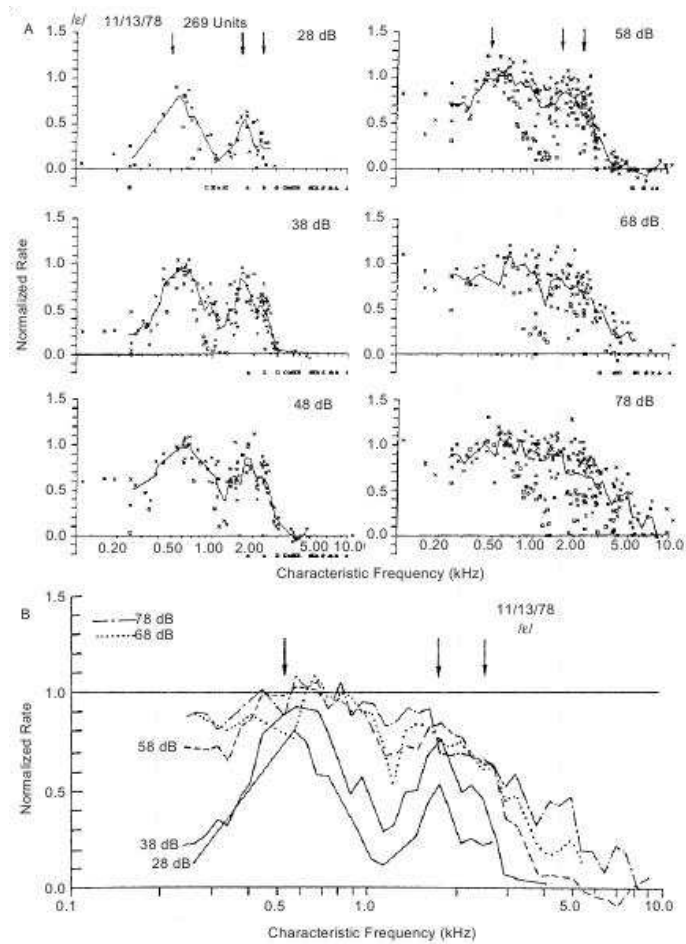
Όπως αναφέρθηκε παραπάνω, δύο βασικοί τρόποι νευρωνικής αναπαράστασης ενός ηχητικού σήματος είναι ο τοπικός (place representation) και ο χρονικός (temporal representation). Η βασική μεμβράνη, ως μία σειρά παράλληλων ζωνοπερατών φίλτρων διατεταγμένων χωρικά, παρέχει την τονοτοπική κωδικοποίηση μέσω του μέσου ρυθμού εκφόρτισης. Ταυτόχρονα όμως, η έξοδος αυτών των φίλτρων ακολουθεί ένα συγκεκριμένο χρονικό μοτίβο εκφορτίσεων παράγοντας έτσι το φαινόμενο του κλειδώματος φάσης.

2.2.1.1 Τοπική αναπαράσταση

Τα φωνήεντα είναι μία κλάση φθόγγων που έχουν χρησιμοποιηθεί κατ'επανάληψη σε φυσιολογικά (physiological) και συμπεριφορικά πειράματα μιας και είναι σχετικά σταθερά στο χρόνο, με καλή περιοδικότητα και με φάσμα δομημένο και σχετικά απλό, τέτοιο ώστε να είναι δυνατή και η σύνθεση τεχνητών φωνηέντων στα οποία υπάρχει πλήρης έλεγχος των χαρακτηριστικών τους. Στα φωνήεντα, η πρώτη και η δεύτερη μορφική (formant) περιέχουν την περισσότερη πληροφορία που επιτρέπει την αναγνώρισή τους. Ορισμένες μελέτες έχουν επιχειρήσει να διερευνήσουν το κατά πόσο αυτές οι δύο πρώτες μορφικές μπορούν να εντοπιστούν στην νευρωνική αναπαράσταση του σήματος με βάση την υπόθεση της τονοτοπικότητας.

Σε πειράματα που έχουν γίνει σε γάτες (κεφ.8 από [7], [28]) έχει βρεθεί ότι οι ίνες του ακουστικού νεύρου που έχουν χαρακτηριστική συχνότητα (CF) κοντά στη συχνότητα κάποιας μορφικής παρουσιάζουν σαφώς εντονότερη δραστηριότητα εκφόρτισης, αποτέλεσμα που υποστηρίζει την υπόθεση της τονοτοπικότητας. Αυτά τα πειραματικά αποτελέσματα προήλθαν από ίνες με χαμηλό κατώφλι, με υψηλό ρυθμό αυθόρμητης δραστηριότητας (high spontaneous rate, high SR) και για χαμηλές ή μέσaiες τιμές SPL. Για υψηλές τιμές SPL όμως, αυτές οι ίνες εισέρχονται σε κατάσταση κορεσμού ενώ όσες έχουν CF σε άλλες περιοχές του φάσματος αυξάνουν και αυτές τη δραστηριότητά τους με αποτέλεσμα να αρχίσει να χάνεται η ικανότητα διάκρισης των μορφικών με βάση το κριτήριο της τονοτοπικότητας. Μερικά τέτοια αποτελέσματα παρουσιάζονται στο σχήμα 2.18.

Ωστόσο, οι μετρήσεις σε ίνες που έχουν χαμηλό ρυθμό αυθόρμητης δραστηριότητας (και ταυτόχρονα υψηλότερο κατώφλι και ευρύτερο δυναμικό εύρος) έδειξαν ότι αυτές οι ίνες διατηρούν την ικανότητα διαχωρισμού των μορφικών ακόμα και για υψηλές τιμές SPL (βλ. σχ.2.19). Ένα ενδιαφέρον εύρημα επί αυτού του θέματος αφορά στη συμπεριφορά των νευρώνων τύπου chopper στον κοχλιακό πυρήνα. Ο μέσος ρυθμός εκφόρτισης σε αυτούς τους νευρώνες εμφανίζει τα μέγιστα του στις θέσεις των μορφικών, ακόμα και για επίπεδα πίεσης όπου οι ίνες υψηλού SR έχουν κορεστεί. Σε τέτοια υψηλά επίπεδα πίεσης, η συμπεριφορά των νευρώνων chopper ακολουθούσε αυτή των ιών του AN χαμηλού SR ενώ σε χαμηλότερα επίπεδα πίεσης ήταν παρόμοια με αυτή των ιών υψηλού SR. Βάσει αυτού του ευρήματος διατυπώθηκε η υπόθεση του επιλεκτικού ακούσματος (selective listening) κατά το οποίο το κεντρικό νευρικό σύστημα “επιλέγει” να στρέψει την προσοχή του στις ίνες



Σχήμα 2.18: Γραφικές παραστάσεις του ρυθμού εκφόρτισης συναρτήσει της χαρακτηριστικής συχνότητας για διάφορα επίπεδα πίεσης, [18]

υψηλού SR όταν τα επίπεδα πίεσης είναι χαμηλά και στις ίνες χαμηλού SR όταν τα επίπεδα πίεσης αυξάνουν. Επίσης, το δυναμικό εύρος των ινών του AN είναι μεγαλύτερο κατά την αρχή της εκφόρτισης (onset) ενώ τα προηγούμενα αποτελέσματα αφορούν σε σταθερά φωνήεντα (steady-state). Αυτό το γεγονός θα μπορούσε να λειτουργεί επικουρικά στη διάκριση των μορφικών. Ακόμα όμως και στην περίπτωση που η δομή των μορφικών δεν είναι άμεσα ορατή μέσω μίας τοπολογικής αντιστοίχισης, η κατανομή του μέσου ρυθμού εκφόρτισης σε όλο τον πληθυσμό των ινών του AN μεταβάλλεται ανάλογα με το φωνήεν, διατηρώντας έτσι την ικανότητα διάκρισης με βάση τα γενικότερα χαρακτηριστικά δραστηριότητας. Άλλες μελέτες (κεφ.8 από [7]) έχουν δείξει ότι είναι δυνατός ο διαχωρισμός δύο φωνηέντων που διαφέρουν στη συχνότητα μίας μόνο μορφικής με βάση τους μέσους ρυθμούς εκφόρτισης, ακόμα και για επίπεδα SPL στα 70 dB. Τέλος, πρέπει να σημειωθεί ότι τέτοιου είδους πειράματα γίνονται συνήθως σε αναισθητοποιημένα ζώα. Η αναισθησία ενδέχεται να επηρεάζει κάποιες οδούς ανάδρασης για τις οποίες υπάρχουν ενδείξεις ότι παίζουν ένα ρυθμιστικό ρόλο επί των ινών που ίσως τις βοηθάει να προσαρμόσουν το ρυθμό

εκφόρτισής τους σε υψηλά επίπεδα πίεσης.

Η έρευνα για άλλες κλάσεις φωνημάτων είναι πιο περιορισμένη. Για τα δασέα φωνήματα έχει βρεθεί ότι ο μέσος ρυθμός εκφόρτισης των πιο ευαίσθητων ινών του ΑΝ ακολουθεί επίσης ένα τονοτοπικό μοτίβο. Οι ίνες που δραστηριοποιούνται εντονότερα είναι αυτές που αντιστοιχούν στις φασματικές περιοχές όπου το κάθε δασύ φώνημα παρουσιάζει μέγιστα. Η τοπική αναπράσταση των δασέων διευκολύνεται από το γεγονός ότι αυτά εκφέρονται σε χαμηλά επίπεδα πίεσης ενώ, από την άλλη, η χρονική τους αναπράσταση είναι πιο δύσκολη καθώς η ενέργειά τους συγκεντρώνεται κυρίως σε περιοχές συχνοτήτων όπου δεν υπάρχει η ικανότητα phase-locking.

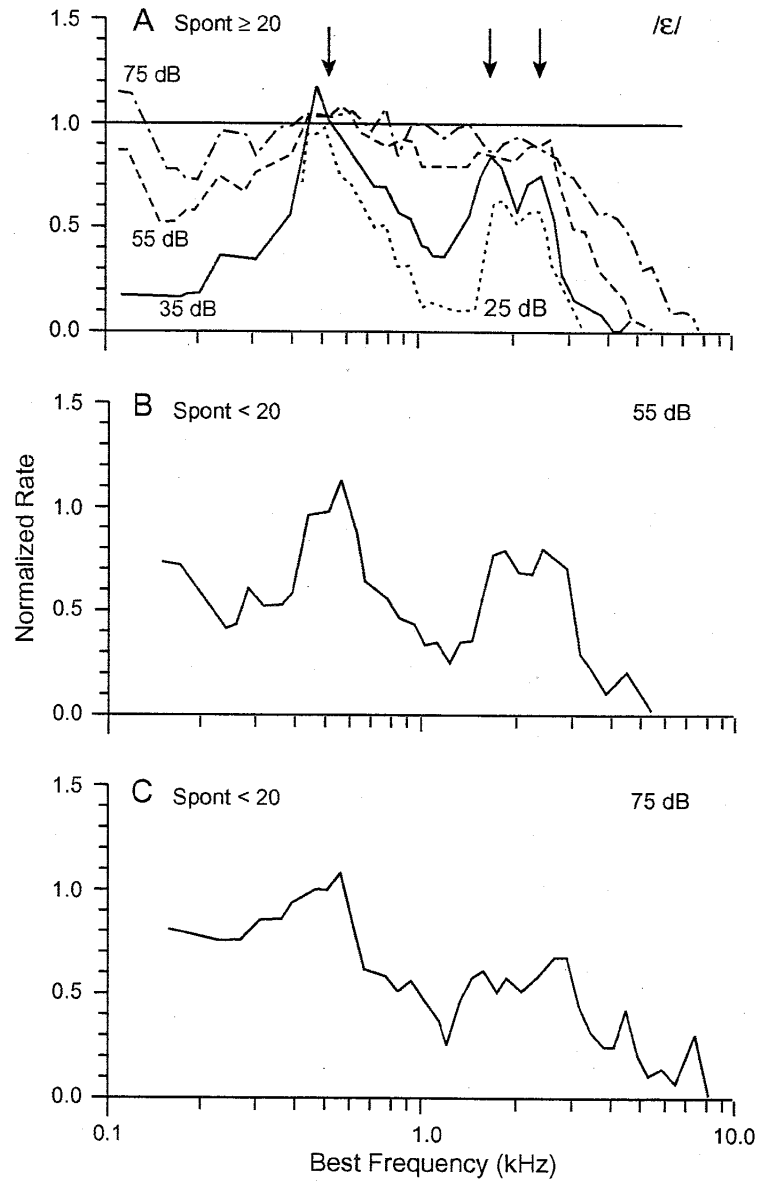
2.2.1.2 Χρονική αναπράσταση

Οι ίνες του ΑΝ έχουν την ικανότητα να συγχρονίζουν (phase-lock) τη δραστηριότητά τους με διάφορα χαρακτηριστικά των ηχητικών σημάτων. Ιδιαίτερα για τα σήματα ομιλίας των οποίων το φάσμα καταλαμβάνει τις περιοχές κάτω των 5 kHz, αυτή η συμπεριφορά είναι αρκετά διαδεδομένη. Οι ίνες με χαρακτηριστική συχνότητα κοντά σε μία μορφική κλειδώνουν στη συχνότητα της μορφικής. Ίνες με ΧΣ κάτω από την πρώτη μορφική ή μεταξύ της πρώτης και της δεύτερης κλειδώνουν είτε στην πιο κοντινή στη ΧΣ αρμονική είτε στη διαμόρφωση του ύψους (pitch) της φωνής που είναι το αποτέλεσμα της αλληλεπίδρασης μεταξύ αρμονικών. Όσες ίνες έχουν ΧΣ πάνω από τη δεύτερη μορφική κλειδώνουν είτε στη δεύτερη μορφική είτε πάλι σε διαμορφώσεις του ύψους της φωνής. Σε γενικές γραμμές, οι συχνότητες των μορφικών κυριαρχούν και καταστέλλουν την απόκριση σε άλλα χαρακτηριστικά. Στο σχήμα 2.20 παρουσιάζονται μερικές γραφικές παραστάσεις για τέσσερα φωνήεντα με την κυρίαρχη συχνότητα κλειδώματος συναρτήσει της ΧΣ από τις οποίες είναι εύκολο να εξαχθούν η θεμελιώδης αρμονική και οι δύο πρώτες μορφικές.

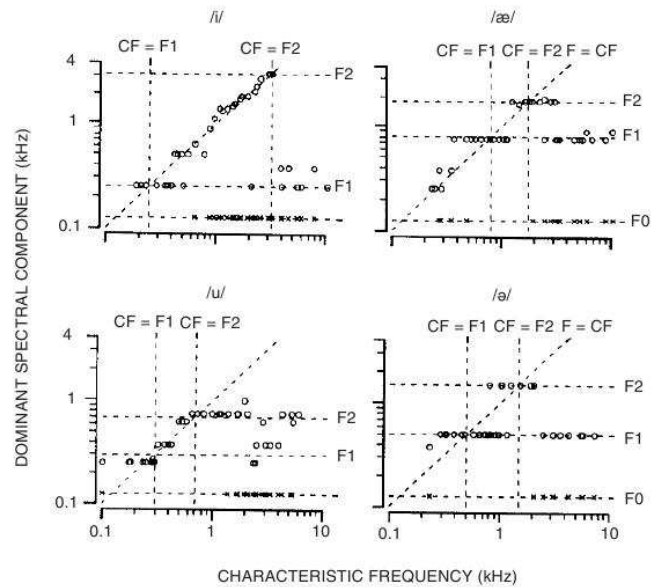
Σημαντικό ρόλο στην χρονική κωδικοποίηση φαίνεται ότι παίζουν τα θυσανωτά (bushy) κύτταρα του κοχλιακού πυρήνα. Κάθε θυσανωτό κύτταρο νευρώνεται συνήθως από λίγες, χαμηλής ΧΣ, ίνες του ΑΝ οι οποίες τυλίγονται γύρω από αυτό σχηματίζοντας μεγάλες συνάψεις. Ο τρόπος διασύνδεσης των ινών του ΑΝ με τα σφαιρικά κύτταρα (είδος θυσανωτών κυττάρων) επιτρέπει σε αυτά να κωδικοποιούν χρονικές πληροφορίες και να τις μεταδίδουν σε πιο κεντρικές περιοχές του ακουστικού συστήματος. Τα σφαιρικά κύτταρα λαμβάνουν επίσης και ανασταλτικές εισόδους, κάτι που ενδεχομένως οξύνει ακόμα περισσότερο την διακριτική τους ικανότητα.

2.2.1.3 Μικτή χρονοτοπική αναπράσταση

Μία πιο στιβαρή αναπράσταση που συνδυάζει τόσο πληροφορίες τοπικότητας όσο και χρονικότητας και παραμένει σχετικά ανεξάρτητη από παράγοντες όπως η ένταση του ερεθίσματος και ο θόρυβος υποβάθρου είναι αυτή του μέσου εντοπισμένου συγχρονισμένου ρυθμού (average localized synchronized rate, ALSR) που πρότειναν οι Young και Sachs ([28], κεφ.10 από [7]). Αρχικά υπολογίζονται ιστογράμματα των νευρωνικών αποκρίσεων των ινών σε ένα ερέθισμα στα οποία κατόπιν εφαρμόζεται ένας μετασχηματισμός Fourier για να βρεθεί ο βαθμός στον οποίο η κάθε ίνα είναι κλειδωμένη στις διάφορες συχνότητες. Μέσω αυτών των υπολογισμών εξάγονται πληροφορίες χρονικότητας και, όπως ήταν αναμενόμενο, βρέθηκε ότι οι ίνες κλειδώνουν σε αρμονικές που είναι πλησίον της ΧΣ τους. Στο επόμενο βήμα, λαμβάνεται ο μέσος όρος των παραπάνω αποκρίσεων για κάθε αρμονική.



Σχήμα 2.19: Γραφικές παραστάσεις του ρυθμού εκφόρτισης συναρτήσει της χαρακτηριστικής συχνότητας για ίνες χαμηλού SR, [7]



Σχήμα 2.20: Γραφικές παραστάσεις της κυρίαρχης συχνότητας κλειδώματος συναρτήσεως της ΧΣ, [18]

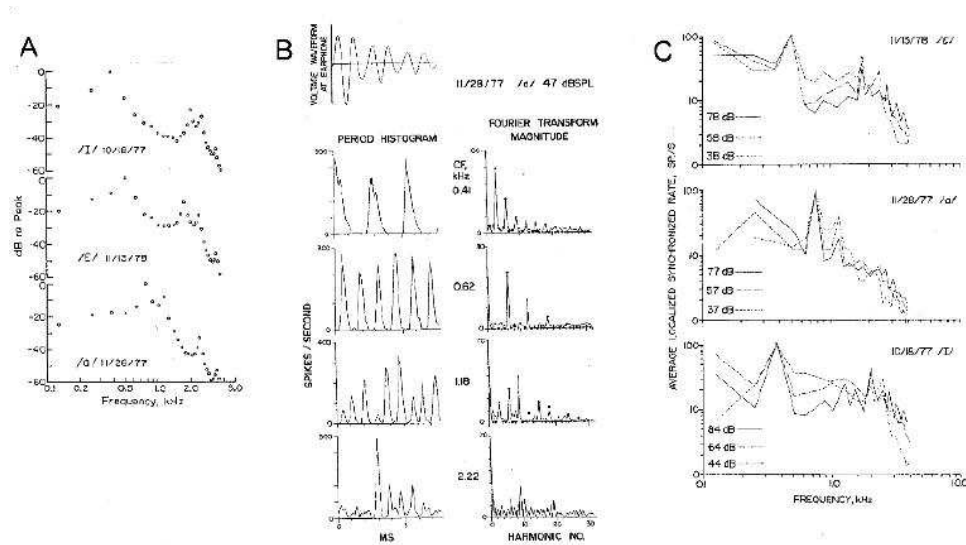
Για κάθε αρμονική, αυτός ο μέσος όρος υπολογίζεται μέσα σε μία γειτονιά ιών (τοπικότητα) των οποίων οι ΧΣ κινούνται σε ένα εύρος συχνοτήτων γύρω από την αρμονική, π.χ. μίση οκτάβα πάνω και κάτω από αυτήν.

$$ALSR(k) = \frac{1}{M_k} \sum_{l \in C_k} R_{kl}$$

όπου R_{kl} είναι η k συνιστώσα της l ίνας, C_k το σύνολο των ιών εντός ενός συγκεκριμένου εύρους ΧΣ και M_k το πλήθος τους. Η παραπάνω διαδικασία εξαγωγής του ALSR συνοψίζεται στο σχήμα 2.21. Η ALSR αναπαράσταση μπορεί να υπολογιστεί και για άφωνους ήχους, μπορεί να αναπαραστήσει τις μεταβάσεις των μορφικών και διατηρεί τη φασματική πληροφορία για δύο ταυτόχρονα φωνήεντα με διαφορετικό ύψος. Παρ' όλα αυτά τα πλεονεκτήματα όμως, δεν υπάρχουν ενδείξεις προς το παρόν, με βάση ανατομικά και φυσιολογικά δεδομένα, ότι ένας τέτοιος μετασχηματισμός λαμβάνει όντως χώρα ή ότι οι πληροφορίες που "αποθηκεύονται" σε αυτόν είναι διαθέσιμες και σε ανώτερα επίπεδα του νευρικού συστήματος.

2.2.1.4 Πλευρική αναστολή

Το φαινόμενο της πλευρικής αναστολής (lateral inhibition) έχει παρατηρηθεί στον αμφιβληστροειδή όπου και ενισχύει την αντίληψη ακμών και κορυφών, δηλαδή χαρακτηριστικών με απότομες μεταβάσεις. Λειτουργία αναστολής παρουσιάζεται και στη βασική μεμβράνη όπου το σημείο που έχει τη μέγιστη απόκριση αναστέλλει τη δραστηριότητα των διπλών του. Ένα δίκτυο πλευρικής αναστολής (lateral inhibition network, LIN), όπως έχει



Σχήμα 2.21: Υπολογισμός του ALSR, [18]

προταθεί από τον Shamma, μπορεί να παρέχει πληροφορίες για το φασματικό περιεχόμενο ενός ερεθίσματος, εντοπίζοντας “αχμές” στη νευρωνική δραστηριότητα των ινών του AN. Ο εντοπισμός τους μπορεί να γίνει μέσω μίας πράξης παραγωγίσισης η οποία μηδενίζει την παρόμοια δραστηριότητα και οξύνει τις αντιθέσεις.

2.2.1.5 Αποκρίσεις σε τόνους διαμορφωμένους κατά συχνότητα (FM)

Οι ίνες του AN αντιδρούν σε FM τόνους κυρίως με βάση τη στιγμιαία συχνότητα του τόνου, όπως θα αντιδρούσαν και σε έναν απλό ημιτονοειδή τόνο ίδιας συχνότητας. Οι αποκρίσεις τους δεν παρουσιάζουν ασυμμετρίες ή προτιμήσεις κατεύθυνσης σε μία FM σάρωση (sweep), αναλόγως με το αν αυτό κινείται προς τα πάνω ή προς τα κάτω. Η μόνη διαφορά που παρατηρείται αφορά στην μετατόπιση της συχνότητας μέγιστης απόκρισης λόγω προσαρμογής σε συχνότητες που προηγήθηκαν. Αυτή η συμμετρία στις αποκρίσεις έρχεται σε αντίθεση με τη συμπεριφορά που επιδεικνύουν οι νευρώνες του ακουστικού φλοιού, όπως θα δούμε παρακάτω, οι οποίοι μπορεί να είναι αρκετά επιλεκτικοί σε διάφορα χαρακτηριστικά μίας FM σάρωσης. Ομοίως, αρκετοί νευρώνες του κοχλιακού συμπλέγματος έχουν αποκρίσεις σε FM τόνους που συμβαδίζουν με αυτές σε απλά ημίτονα αν και σε κάποιους έχουν παρατηρηθεί προτιμήσεις κατεύθυνσης της σάρωσης.

2.2.1.6 Αποκρίσεις σε τόνους διαμορφωμένους κατά πλάτος (AM)

Οι κύριες παράμετροι ενός AM τόνου που επηρεάζουν την απόκριση των ακουστικών νευρώνων είναι ο ρυθμός και το βάθος της διαμόρφωσης. Οι νευρώνες αντιδρούν διαμορφώνοντας τις εκφορτίσεις τους. Ένας τρόπος περιγραφής της απόκρισής τους είναι μέσω των λεγόμενων συναρτήσεων μεταφοράς διαμόρφωσης (modulation transfer function, MTF).

Κεφάλαιο 2. Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα

Για ένα συγκεκριμένο βάθος διαμόρφωσης του ερεθίσματος, οι MTF μάς δίνουν το βάθος διαμόρφωσης της απόκρισης (ή τον λόγο των δύο βαθών σε dB) συναρτήσει του ρυθμού διαμόρφωσης.

Οι ίνες του AN έχουν σταθερό μέσο ρυθμό εκφόρτισης, ανεξάρτητο του ρυθμού διαμόρφωσης, και οι MTF τους είναι βαθυπερατές για όλα τα βάθη διαμόρφωσης με τη συχνότητα αποκοπής να αυξάνεται καθώς αυξάνεται η ΧΣ της ίνας. Το μέγιστο μίας MTF μάς δίνει το ρυθμό διαμόρφωσης όπου παρατηρείται η εντονότερη διαμόρφωση των νευρωνικών εκφορτίσεων και ονομάζεται βέλτιστη συχνότητα διαμόρφωσης (best modulation frequency, BMF). Τυπικές τιμές BMF για τις ίνες του AN είναι μεταξύ 400 και 1500 Hz. Για ερεθίσματα μίας μορφικής, η διαμόρφωση της απόκρισης γίνεται πιο έντονη καθώς αυξάνεται η ένταση του ερεθίσματος, μέχρι ένα σημείο κορεσμού όπου φτάνει στο μέγιστό της και στη συνέχεια αρχίζει να μειώνεται.

Οι περισσότεροι τύποι νευρώνων του κοχλιακού συμπλέγματος συγχρονίζουν τη δραστηριότητά τους με την περιβάλλουσα του σήματος καλύτερα σε σχέση με τις ίνες του AN, ιδιαίτερα τα κύτταρα τύπου onset και chopper. Οι MTF αυτών των κυττάρων είναι συνήθως ζωνοπερατές, με μεγάλο κέρδος στις περιοχές μεγίστου και συχνότητα γονάτου (η συχνότητα όπου παρατηρείται μείωση από το μέγιστο κατά 3 dB) συνήθως κάτω από 1.5 kHz. Ωστόσο, παρουσιάζονται αρκετές διαφοροποιήσεις των MTF ανάλογα με τον τύπο κυττάρου του κοχλιακού πυρήνα στον οποίο αναφερόμαστε. Οι διαφοροποιήσεις αυτές αφορούν σε ένα πλήθος χαρακτηριστικών όπως το μέγεθος του κέρδους, η βέλτιστη συχνότητα διαμόρφωσης (BMF), η συχνότητα αποκοπής και τα επίπεδα πίεσης SPL. Μάλιστα, οι MTF ορισμένων κυττάρων του οπίσθιου κοχλιακού πυρήνα μεταβάλλουν και το σχήμα τους, όντας βαθυπερατές για χαμηλά SPL και ζωνοπερατές για υψηλά SPL. Ένα άλλο ενδιαφέρον χαρακτηριστικό των νευρώνων του κοχλιακού πυρήνα είναι ότι δεν επηρεάζονται σημαντικά από τη συχνότητα του φέροντος με αποτέλεσμα η διαμόρφωση στην απόκριση να είναι αρκετά διαδεδομένη επί του τονοτοπικού άξονα. Τέλος, αξίζει να σημειωθεί ότι η ένταση της διαμόρφωσης στον κοχλιακό πυρήνα δεν κωδικοποιείται στον μεσό ρυθμό εκφόρτισης, ο οποίος είναι ανεξάρτητος του ρυθμού διαμόρφωσης, αλλά στο χρονικό μοτίβο εκφορτίσεων.

2.2.1.7 Αποκρίσεις σε ενάρξεις (onset)

Η απότομη απελευθέρωση ενέργειας, όπως στα εκρηκτικά (plosive) φωνήματα, δημιουργεί χαρακτηριστικές κάθετες μπάρες ενέργειας στα φασματογράμματά τους. Οι ίνες του AN ανταποκρίνονται έντονα κατά την έναρξη των σημάτων, εντονότερα σε σχέση με την απόκρισή τους κατά το σταθερότερο τμήμα του ήχου που ακολουθεί μετά την έναρξη. Ο μέσος ρυθμός εκφόρτισής τους κωδικοποιεί την ένταση του ερεθίσματος για ένα μεγάλο εύρος τιμών. Τα κύτταρα του κοχλιακού πυρήνα αντιδρούν επίσης με παρόμοιο τρόπο σε ενάρξεις ήχων. Κύτταρα τύπου octopus δέχονται εισόδους από ίνες του AN με ένα μεγάλο εύρος ΧΣ. Επομένως, δεν είναι καλά συντονισμένα σε συγκεκριμένες συχνότητες και μοντελοποιούνται καλύτερα ως ευρυζωνικά φίλτρα με καλή απόκριση σε απότομες ενάρξεις ήχων. Με βάση τα προηγούμενα, φαίνεται αρκετά πιθανό η πληροφορία για την έναρξη των ήχων να περνάει τελικά στα επόμενα στάδια επεξεργασίας.

2.2.2 Μετασχηματισμοί των σημάτων στο κεντρικό ακουστικό σύστημα

Το κεντρικό ακουστικό σύστημα περιλαμβάνει το κάτω διδύμιο, το μέσο γονατώδες σώμα στον θάλαμο και τον ακουστικό φλοιό. Οι γνώσεις που υπάρχουν αυτή τη στιγμή για την επεξεργασία των ηχητικών σημάτων σε αυτά τα στάδια του ακουστικού συστήματος υστερούν σημαντικά εν σχέσει με αυτές που είναι διαθέσιμες για το περιφερικό σύστημα, όπως περιγράφηκαν στην προηγούμενη ενότητα. Η έρευνα έχει επικεντρωθεί κυρίως στις αποκρίσεις νευρώνων του ακουστικού φλοιού σε διάφορα ερεθίσματα. Η προσκόλληση αυτή στον ακουστικό φλοιό δεν είναι αδικαιολόγητη εφόσον είναι γενικώς αποδεκτό ότι ο φλοιός είναι η περιοχή των ανώτερων λειτουργιών στον άνθρωπο και είναι πολύ πιθανό σε αυτόν να βρίσκονται οι κρίσιμες περιοχές για την κατανόηση και αντίληψη του λόγου. Ωστόσο, πρέπει να σημειωθεί μία ιδιαιτερότητα ειδικά του ακουστικού συστήματος. Η επεξεργασία που υφίσταται ένα σήμα ήχου στα κατώτερα στάδια, πριν φτάσει στον φλοιό, είναι πολύ πιο εκτεταμένη σε σχέση με τα υπόλοιπα αισθητικά συστήματα. Μάλιστα, έχει διατυπωθεί η υπόθεση (με μάλλον μικρή αποδοχή προς το παρόν) ότι η κύρια λειτουργία του ακουστικού φλοιού ενδεχομένως να είναι ο εντοπισμός της προέλευσης του ήχου και όχι η σημασιολογική, συντακτική ή άλλου είδους ανώτερη ανάλυση (κεφ.5 από [7]). Τα παραπάνω, δεδομένου και του παρόντος ελλείμματος γνώσεων, μάλλον καθιστούν παρακινδυνευμένες ενδεχόμενες εικασίες περί της ακριβούς λειτουργίας των επιμέρους σταθμών του ακουστικού συστήματος, συμπεριλαμβανομένου και του φλοιού. Παρ' όλα αυτά, και εν αναμονή των επόμενων λεπτομερέστερων μελετών, η λειτουργική περιγραφή που ακολουθεί αφορά σε μεγάλο βαθμό στον ακουστικό φλοιό.

Η τοποτική οργάνωση που παρατηρείται στο περιφερικό ακουστικό σύστημα διατηρείται και στους επόμενους πυρήνες, μέχρι και τον φλοιό, με τη διαφορά ότι στον φλοιό έχουμε έναν μετασχηματισμό σε δύο διαστάσεις. Στον ΑΙ, οι νευρώνες οργάνωνται σε ισοσυχνοτικές περιοχές, δηλαδή περιοχές όπου οι νευρώνες έχουν περίπου την ίδια ΧΣ. Ο λόγος πίσω από αυτή τη συμπεριφορά ίσως να είναι ότι ο φλοιός δημιουργεί και άλλους άξονες ανάλυσης πέρα από αυτόν της ΧΣ έτσι ώστε η τελική κωδικοποίηση να είναι αρκετά πιο λεπτομερειακή και πολύπλοκη. Η υπόθεση αυτή στηρίζεται και σε δεδομένα από τον τρόπο λειτουργίας του οπτικού συστήματος στο οποίο η αρχική αναπαράσταση στον αμφιβληστροειδή απλώς διατηρεί το σχήμα ενός αντικειμένου αλλά στον φλοιό η ανάλυση γίνεται με βάση ποικίλες παραμέτρους όπως, για παράδειγμα, η κατεύθυνση μίας ακμής ή το πάχος της. Τα νευροφυσιολογικά δεδομένα που έρχονται στην επιφάνεια σιγά σιγά είναι υποστηρικτικά μίας τέτοιας υπόθεσης, όπως θα φανεί στη συνέχεια. Στις επόμενες υποενότητες γίνεται μία παρουσίαση μερικών υποθέσεων περί της αναπαράστασης των ηχητικών σημάτων στο κεντρικό ακουστικό σύστημα καθώς και μερικών δεδομένων από νευροφυσιολογικές μελέτες για τον τρόπο απόκρισης του σε διάφορα ερεθίσματα ([18]).

2.2.2.1 Η υπόθεση της τοπικής αναπαράστασης (place representation)

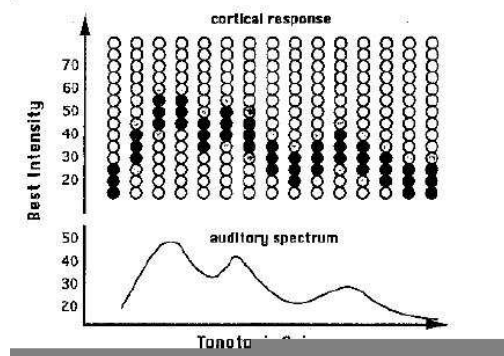
Η υπόθεση ότι οι αποκρίσεις του κεντρικού ακουστικού συστήματος οργανώνονται κατά περιοχές και μπορούν να προβλεφθούν και υπολογισθούν συνθέτοντας τις επιμέρους αντιδράσεις σε απλούς τόνους κατεύθυνε αρχικά την έρευνα ακριβώς στη μελέτη της συμπεριφοράς των νευρώνων σε τέτοια απλά ερεθίσματα. Τα αποτελέσματα που προέκυψαν

Κεφάλαιο 2. Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα

όμως άρχισαν να εγείρουν υποψίες ότι η προηγούμενη υπόθεση είναι μάλλον υπεραπλουστευτική. Τελικά, ένα μεγάλο κέρδος από αυτές τις πρώιμες μελέτες είναι ακριβώς το γεγονός ότι έγινε πλέον κατανοητό πως μία απλή τοπική αναπαράσταση των ήχων είναι εξαιρετικά ανεπαρκής για να περιγράψει την πολύπλοκη συμπεριφορά που επιδεικνύουν οι πυρήνες του κεντρικού ακουστικού συστήματος. Οι τονικές αποκρίσεις που παρατηρούνται παρουσιάζουν συχνά έντονα στοιχεία μη γραμμικότητας, με περιοχές διέγερσης και αναστολής, με αρκετές ασυμμετρίες και με ΧΣ και εύρη ζώνης να κυμαίνονται σε ένα αρκετά ευρύ φάσμα τιμών. Παρομοίως, οι συναρτήσεις ρυθμού-επιπέδου (rate-level function) πολλές φορές δεν είναι μονοτονικές, με τις τιμές κατωφλίου, κορεσμού και δυναμικού εύρους να καλύπτουν και αυτές ένα μεγάλο φάσμα. Τα παραπάνω ευρήματα αποκλείουν την πιθανότητα μίας απλής τοπικής, νευρωνικής αναπαράστασης ενώ ορισμένοι ερευνητές ([18]) αμφισβητούν ακόμα και τη σημασία των ισοσυχνοτικών περιοχών του φλοιού ως λειτουργικές μονάδες.

2.2.2.2 Η αναπαράσταση βέλτιστης έντασης (best-intensity)

Σύμφωνα με αυτή την αναπαράσταση, συστοιχίες νευρώνων οργανώνονται σε μία πρώτη φάση σε ισοσυχνοτικές ζώνες με βάση τις ΧΣ τους και στη συνέχεια κάθε τέτοια ζώνη οργανώνεται με βάση την ένταση. Κατ' αυτόν τον τρόπο προκύπτουν ορισμένοι διδιάστατοι νευρωνικοί χάρτες συχνότητας - έντασης που κωδικοποιούν το φάσμα του ερεθίσματος, όπως αυτός του σχήματος 2.22. Τέτοιοι χάρτες όμως έχουν εντοπιστεί μόνο σε ένα είδος νυχτερίδας και όχι και σε άλλα είδη. Δεν έχουν βρεθεί χωρικά οργανωμένοι χάρτες έντασης ενώ και η βέλτιστη ένταση ενός νευρώνα μεταβάλλεται ανάλογα με το ερέθισμα που του παρουσιάζεται. Γενικά, οι αποκρίσεις των νευρώνων του ΑΙ συναρτήσει της έντασης απλών τόνων παρουσιάζουν σαφώς μεγαλύτερη πολυπλοκότητα από αυτήν που υπαινίσσεται η αναπαράσταση βέλτιστης έντασης. Ακόμα και αν ισχύει η υπόθεση αυτής της αναπαράστασης θα πρέπει να ισχύει σε μία αρκετά πιο περίπλοκη εκδοχή της.



Σχήμα 2.22: Χάρτης συχνότητας - έντασης, [18]

2.2.2.3 Η κατηγορική αναπαράσταση (categorical representation)

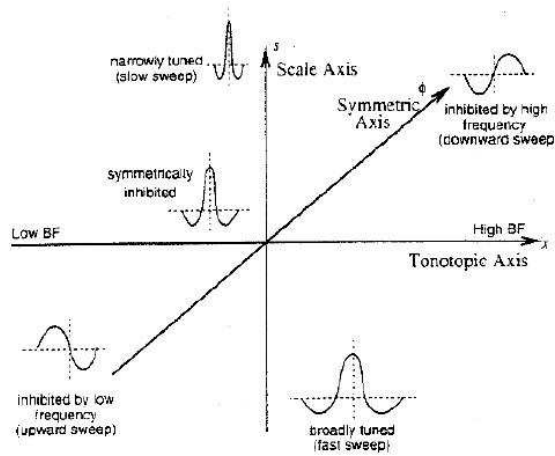
Μία ακραία εκδοχή της τοπικής αναπαράστασης μπορεί να καταλήξει στη λεγόμενη κατηγορική αναπαράσταση που βασίζεται στην ιδέα του “χυττάρου της γιαγιάς”, μία ιδέα που προέρχεται από τη μελέτη του οπτικού συστήματος όπου κάποιες φορές γίνεται η υπόθεση ύπαρξης περιοχών εξαιρετικά επιλεκτικών σε συγκεκριμένα πρόσωπα (π.χ. σε αυτό της γιαγιάς μου). Επεκτείνοντας αυτή την ιδέα και στο ακουστικό σύστημα, διατυπώνεται η υπόθεση ότι υπάρχουν νευρώνες ή μικρές ομάδες νευρώνων που είναι επιλεκτικοί όχι απλώς σε μερικά γενικά φασματικά χαρακτηριστικά, όπως η συχνότητα, η ένταση ή η κατεύθυνση μίας σάρωσης, αλλά σε πολύ συγκεκριμένες διαμορφώσεις του φάσματος όπως ένα φώνημα ή και μία συλλαβή. Ακόμα και αν αυτή η υποθεση ισχύει σε κάποιο βαθμό για ορισμένα είδη, όπως κάποια είδη νυχτερίδας που διαθέτουν εξειδικευμένους μηχανισμούς ηχοεντοπισμού, οι μελέτες των τελευταίων χρόνων στα υπόλοιπα είδη δεν έχουν δείξει κάτι παρόμοιο αλλά, αντιθέτως, φαίνεται ότι η κωδικοποίηση των ήχων επιμερίζεται σε πληθυσμούς νευρώνων που λειτουργούν παράλληλα για να αποσυνθέσουν και να αναλύσουν τα σήματα εισόδου με ένα συνδυασμένο χρονοσυχνοτικό τρόπο.

2.2.2.4 Αναπαράσταση πολλαπλών επιπέδων (multiscale representation)

Η υπόθεση που φαίνεται να κερδίζει έδαφος τελευταία, υποστηριζόμενη και από νευροφυσιολογικά και ψυχοακουστικά πειραματικά δεδομένα, είναι αυτή της ανάλυσης σε πολλαπλά επίπεδα. Η χρήση του όρου “multiscale” γίνεται γιατί θεωρείται ότι, πέρα από τον τονοτοπικό άξονα, η ανάλυση επιτελείται και σε έναν δεύτερο άξονα εύρους ζώνης. Για κάθε συχνότητα, η ίδια ανάλυση επαναλαμβάνεται κατ’ επανάληψη για διάφορα εύρη ζώνης με αποτέλεσμα να προκύπτει μία αναπαράσταση σε πολλαπλά επίπεδα διακριτικής ικανότητας, όπου το γενικό προφίλ του φάσματος κωδικοποιείται από νευρώνες με μεγάλα εύρη ζώνης ενώ οι λεπτομέρειές του από νευρώνες με μικρότερα εύρη ζώνης. Ένας τρίτος άξονας ανάλυσης είναι και αυτός της ασυμμετρίας. Η απόκριση ενός νευρώνα δεν είναι κατ’ ανάγκη συμμετρική περί την ΧΣ του έτσι ώστε αυτός να παρουσιάζει επιλεκτικότητα σε ενδεχόμενες ασυμμετρίες στο ηχητικό σήμα. Με βάση τα παραπάνω, προκύπτει ένα επίπεδο ανάλυσης τριών διαστάσεων, όπως φαίνεται στο σχήμα 2.23. Τέλος, υπάρχει και η διάσταση του χρόνου όπου και εκεί φαίνεται πως λαμβάνει χώρα μία παρόμοια επεξεργασία σε διάφορες χρονικές κλίμακες. Υπάρχουν νευρώνες που μπορούν να παρακολουθούν τις ταχείες μεταβολές του φάσματος και άλλοι που είναι πιο εξειδικευμένοι στις αργές διαμορφώσεις. Η ανάλυση πολλαπλών επιπέδων δεν είναι χαρακτηριστικό μόνο του ακουστικού συστήματος καθώς έχει παρατηρηθεί και στον οπτικό φλοιό, γεγονός που ενισχύει την πεποίθηση ότι πρόκειται για έναν γενικευμένο και ουσιαστικό τρόπο νευρωνικής κωδικοποίησης.

2.2.2.5 Αποκρίσεις σε απλούς τόνους

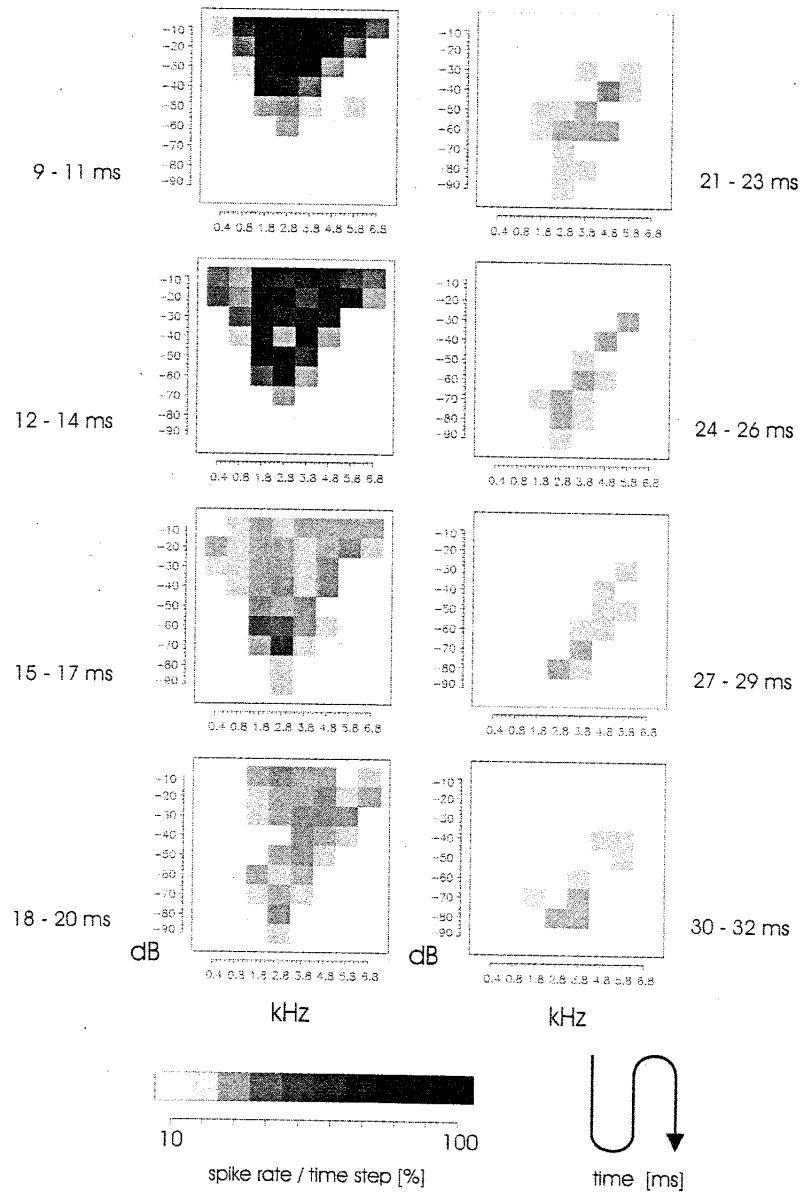
Παρότι οι αποκρίσεις σε απλούς τόνους δεν είναι ιδιαίτερα χρήσιμες για την πρόβλεψη των αποκρίσεων σε πιο πολύπλοκους ήχους μπορούν ωστόσο να δώσουν μία εικόνα της δυναμικής συμπεριφοράς του φλοιού. Τα διάφορα χαρακτηριστικά των σημάτων λόγου εξελίσσονται σε διαφορετικές χρονικές κλίμακες. Πιο συγκεκριμένα, στην κλίμακα του millisecond έχουμε την πληροφορία για τις συχνοτικές συνιστώσες, στην κλίμακα των 2-



Σχήμα 2.23: Χώρος ανάλυσης στην αναπαράσταση πολλαπλών επιπέδων, [18]

10 ms τις απότομες μεταβολές (transient), στην κλίμακα των 100 ms τις μεταβάσεις των μορφικών και τα φωνήματα και τέλος, στις μερικές εκατοντάδες ms γίνεται η ανάλυση συλλαβών και λέξεων. Από αυτή την οπτική, έχει σημασία πώς συμπεριφέρονται οι νευρώνες με το πέρασμα του χρόνου.

Στη συνέχεια παρουσιάζονται τα αποτελέσματα μελέτης σε νευρώνες του ακουστικού φλοιού γατών. Για απλούς τόνους 10 διαφορετικών συχνοτήτων και 10 διαφορετικών εντάσεων μετρήθηκαν τα υποδεκτικά πεδία μεμονωμένων νευρώνων με βάση τα PSTH (poststimulus-time histogram). Από αυτές τις μετρήσεις είναι δυνατό να κατασκευαστούν χρονικές αλληλουχίες των καμπυλών συντονισμού των νευρώνων. Μία τέτοια αλληλουχία φαίνεται στο σχήμα 2.24 όπου παρουσιάζονται μερικά χρονικά παράθυρα της απόκρισης ενός νευρώνα για τις διάφορες τιμές συχνότητας και έντασης των τόνων. Αρχικά ο νευρώνας αντιδρά για ένα ευρύ φάσμα συχνοτήτων και για υψηλές εντάσεις αλλά στη συνέχεια η δραστηριότητα του υποδεκτικού του πεδίου μεταφέρεται πλησιέστερα προς την ΧΣ του σε χαμηλότερες εντάσεις. Ένα έντονο χαρακτηριστικό είναι και η εμφάνιση μίας διαγωνίου στο υποδεκτικό πεδίο προς το τέλος του ερεθίσματος στην περιοχή υψηλών συχνοτήτων. Σε άλλους νευρώνες, τέτοιες διαγωνίσι εμφανίζονται στην περιοχή χαμηλών συχνοτήτων ενώ μερικοί παρουσιάζουν συμμετρία του υποδεκτικού τους πεδίου. Παρόμοιες ασυμμετρίες εμφανίζονται και για χρονοσυχνοτικά υποδεκτικά πεδία, για τομές δηλαδή σε μία συγκεκριμένη ένταση με τον χρόνο και τη συχνότητα να είναι οι συντεταγμένες του πεδίου. Η συγχρονισμένη δραστηριότητα νευρώνων που διαθέτουν τέτοιες διαγωνίσι, τόσο στο χώρο έντασης - συχνότητας όσο και σε αυτόν χρόνου - συχνότητας, θα μπορούσαν να καταλήξουν σε μία αρκετά απλή αναπαράσταση μερικών δυναμικά εξελισσόμενων χαρακτηριστικών των σημάτων λόγου. Για παράδειγμα, μία μετάβαση, όπως σε μία μορφική, μεταφράζεται σε ένα σύντομο και χρονικά ακριβές νευρωνικό γεγονός. Επίσης, είναι εμφανές ότι η έντονη αρχική δραστηριότητα του νευρώνα για υψηλές εντάσεις έχει μικρή διάρκεια ενώ έχουμε μία σταθερή σχεδόν ενεργοποίησή του γύρω από τη ΧΣ του για μεσαίες και χαμηλές εντάσεις. Αυτή η παρατήρηση συμφωνεί και με προηγούμενες μελέτες



Σχήμα 2.24: Χρονική εξέλιξη του υποδεκτικού πεδίου ενός νευρώνα με ΧΣ 2.8 kHz στον ΑΙ γάτας, [7]

στις οποίες είχαμε μία μείωση της καθυστέρησης έναρξης (onset latency) και της διάρκειας της απόκρισης καθώς αυξανόταν η ένταση του ερεθίσματος (κεφ. 9, [7]). Μία εξάρτηση επίσης είχε παρατηρηθεί και με την κλίση της έναρξης του ερεθίσματος.

Το συμπέρασμα που προκύπτει από τέτοιες μελέτες είναι ότι τα υποδεκτικά πεδία των νευρώνων του φλοιού παρουσιάζουν έντονη χρονική εξάρτηση. Μάλιστα, οι ιδιότητές τους στο πεδίο του χρόνου και της συχνότητας είναι σε τέτοιο βαθμό αλληλοεξαρτώμενες ώστε να μην είναι γόνιμη η ανάλυση σε κάθε επιμέρους πεδίο ξεχωριστά. Πρόκειται για ιδιότητες μη διαχωρίσιμες. Αν η φλοιική δραστηριότητα παρατηρηθεί στο επίπεδο ολόκληρων πληθυσμών από νευρώνες τότε αναδύεται ένα πολύπλοκο αλλά σταδιακό και με νόημα χρονικό μοτίβο εξέλιξης δραστηριότητας. Τα υποδεκτικά πεδία ολόκληρων πληθυσμών νευρώνων διαφέρουν από αυτά των μεμονωμένων νευρώνων στο ότι λειτουργούν σε μικρότερες χρονικές κλίμακες και με μεγαλύτερη ομοιογένεια.

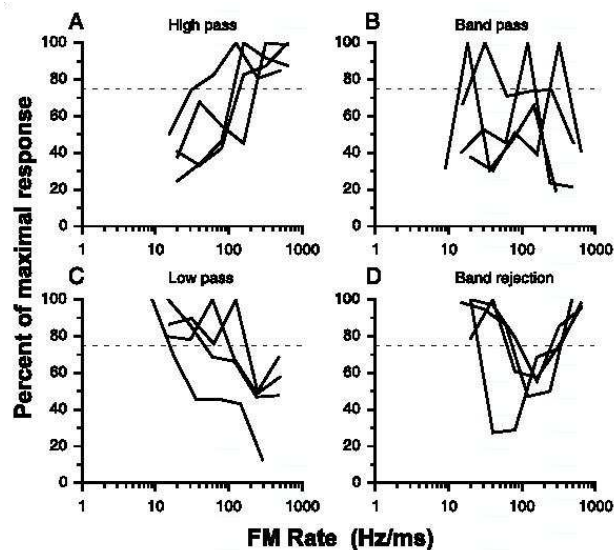
2.2.2.6 Αποκρίσεις σε τόνους διαμορφωμένους κατά πλάτος (AM)

Οι νευρώνες του κάτω διδυμίου έχουν βαθυπερατές MTF (συναρτήσεις μεταφοράς διαμόρφωσης) για χαμηλά επίπεδα SPL οι οποίες γίνονται ζωνοπερατές σε υψηλότερα SPL ενώ οι BMF (βέλτιστες συχνότητες διαμόρφωσης) τους είναι συνήθως στην περιοχή κάτω των 200 Hz, όπως έχουν καταγραφεί σε ορισμένα είδη, π.χ. σε ινδικά χοιρίδια. Σε σχέση με τον κοχλιακό πυρήνα, οι BMF στο κάτω διδύμιο κυμαίνονται σε χαμηλότερες τιμές. Ένα ιδιαίτερο χαρακτηριστικό ορισμένων νευρώνων του κάτω διδυμίου είναι ότι οι MTF τους προκύπτουν παρόμοιες, είτε υπολογίζονται με βάση τη συγχρονισμένη δραστηριότητά τους είτε με βάση τον μέσο ρυθμό εκφόρτισης. Αυτή η παρατήρηση, σε συνδυασμό με το γεγονός ότι η ικανότητα κλειδώματος φάσης μειώνεται στα πιο κεντρικά στάδια των ακουστικών οδών, υποδεικνύει ότι ενδεχομένως στο κάτω διδύμιο λαμβάνει χώρα μία επανακωδικοποίηση των πληροφοριών συγχρονισμού και διαμόρφωσης. Ανοιχτό είναι ακόμα το θέμα της ύπαρξης ή όχι τοπογραφικών χαρτών στο κάτω διδύμιο στους οποίους να απεικονίζεται η πληροφορία διαμόρφωσης. Σε μερικά είδη υπάρχουν ενδείξεις υπέρ της ύπαρξης μίας τέτοιας οργάνωσης ενώ σε άλλα είδη δεν έχει εντοπιστεί κάτι.

Στον ΑΙ, η μεγάλη πλειοψηφία των νευρώνων έχουν ζωνοπερατές MTF με BMF μεταξύ 3 και 100 Hz. Ένα μικρότερο ποσοστό έχει βαθυπερατές MTF με πολύ μικρή συχνότητα αποκοπής. Κάποια τοπογραφική οργάνωση με βάση την BMF δεν έχει ανιχνευθεί. Οι περισσότεροι νευρώνες του ακουστικού φλοιού περιορίζουν την ευαισθησία τους σε διαμορφώσεις της περιβάλλουσας σε ρυθμούς κάτω των 20 Hz. Η δυνατότητα άμεσης κωδικοποίησης, μέσω κλειδώματος φάσης, ξεκινάει από τα 4000 Hz στο ΑΝ, πέφτει στα 800 με 1000 Hz στο κάτω διδύμιο και παρουσιάζει περαιτέρω μείωση στον ακουστικό φλοιό με το άνω όριο να είναι τα 100 Hz. Ωστόσο, υπάρχουν ενδείξεις ότι ο ακουστικός φλοιός δεν χρησιμοποιεί μόνο ένα μηχανισμό κωδικοποίησης των διαμορφώσεων AM. Σε χαμηλούς ρυθμούς AM, κάτω από 100 Hz, οι νευρώνες συγχρονίζουν-κλειδώνουν τις εκφορτίσεις τους με τη διαμορφωμένη περιβάλλουσα αλλά σε υψηλότερους ρυθμούς, όπου δεν υπάρχει πλέον αυτή η δυνατότητα, η κωδικοποίηση γίνεται πιο έμμεσα, μέσω του μέσου ρυθμού εκφόρτισης, με κάθε κύτταρο να ανταποκρίνεται επιλεκτικά σε ένα μικρό εύρος συχνοτήτων φέροντος και διαμόρφωσης (κεφ. 10, [7]).

2.2.2.7 Αποκρίσεις σε τόνους διαμορφωμένους κατά συχνότητα (FM)

Όπως έχει ήδη αναφερθεί προηγουμένως, στο περιφερικό ακουστικό σύστημα, οι αποκρίσεις σε FM ερεθίσματα δεν παρουσιάζουν ιδιαιτερότητες και σχεδόν μπορούν να προβλεφθούν από τις αποκρίσεις σε απλούς τόνους. Οι νευρώνες αντιδρούν στις συχνότητες ενός FM sweep που αντιστοιχούν σε μια περιοχή γύρω από την ΧΣ τους ενώ για τις υπόλοιπες συχνότητες του sweep παραμένουν αδρανείς. Η κατάσταση όμως αλλάζει σημαντικά στο κεντρικό ακουστικό σύστημα με το βαθμό ασυμμετρίας των αποκρίσεων να γίνεται εντονότερος καθώς προχωράμε από το κάτω διδύμιο στον φλοιό. Τα πειραματικά ευρήματα έχουν επιβεβαιώσει επανειλημμένα την ύπαρξη κυττάρων που είναι επιλεκτικά σε συγκεκριμένες κατευθύνσεις και ρυθμούς ενός FM sweep. Μερικοί νευρώνες μάλιστα αποκρίνονται σε FM συχνότητες που είναι εκτός του εύρους ζώνης τους, όπως αυτό καθορίζεται με βάση απλούς τόνους. Γενικότερα, δεν παρατηρείται κάποια εξάρτηση μεταξύ της ΧΣ (για απλούς τόνους) ενός νευρώνα και της προτίμησής του για τον FM ρυθμό αλλά έχει παρατηρηθεί σχέση μεταξύ ΧΣ και κατεύθυνσης με τις περιοχές υψηλών ΧΣ να προτιμούν καθοδικά sweep και αυτές με χαμηλές ΧΣ ανοδικά ([27]). Όσον αφορά την επιλεκτικότητα στον ρυθμό, οι νευρώνες μπορούν να κατηγοριοποιηθούν σε βαθυπερατούς, υψιπερατούς, ζωνοπερατούς και ζωνοφρακτικούς (βλ. σχήμα 2.25). Η σχέση μεταξύ ρυθμού και κατεύθυνσης μπορεί να είναι πολύπλοκη και ασύμμετρη. Κάποιοι νευρώνες μπορεί να είναι υψιπερατός για ανοδική κατεύθυνση (upward) του sweep αλλά βαθυπερατός όταν η κατεύθυνση είναι καθοδική. Άλλοι νευρώνες παρουσιάζουν δραστηριότητα μόνο σε μία κατεύθυνση. Νευρώνες με μεγάλο εύρος ζώνης αντιδρούν καλύτερα σε ταχέως κινούμενα sweep. Επίσης, η επιλεκτικότητα κατεύθυνσης μειώνεται καθώς αυξάνεται ο ρυθμός του FM sweep.



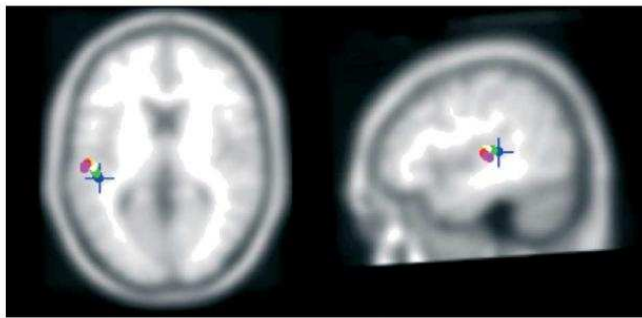
Σχήμα 2.25: Παραδείγματα καμπυλών συντονισμού από νευρώνες στην πλευρική ζώνη rhesus πιθήκου, [26]

Αξιοσημείωτο είναι το γεγονός πως υπάρχουν νευρώνες που αντιδρούν αρκετά έντο-

Κεφάλαιο 2. Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα

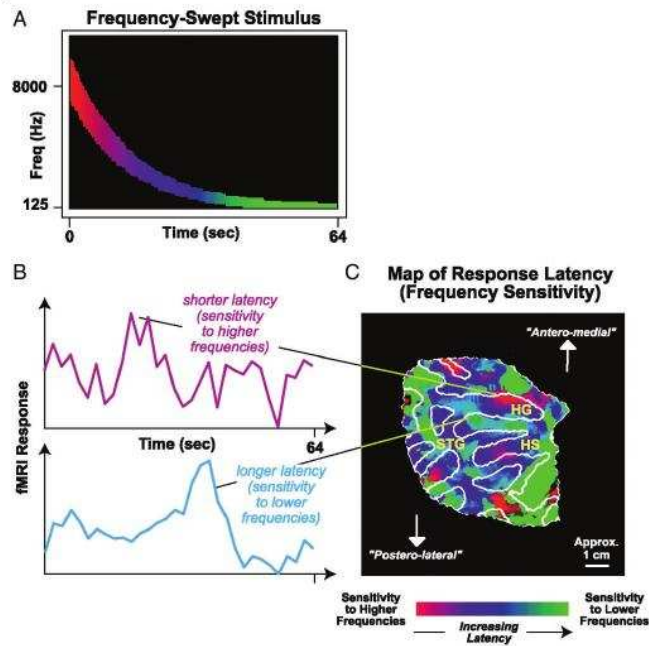
να σε FM σήματα αλλά καθόλου σε απλούς τόνους. Αυτό το φαινόμενο παρατηρείται σε μεγαλύτερο βαθμό σε περιοχές εκτός του πρωτοταγούς φλοιού ([26]), κάτι που συμφωνεί με τη γενικότερη πεποίθηση ότι, καθώς απομακρυνόμαστε από τις πρωτοταγείς αισθητικές περιοχές, οι νευρώνες προτιμούν όλο και πολυπλοκότερα ερεθίσματα. Σε αυτές τις πιο εξωτερικές περιοχές είναι εντονότερα και τα φαινόμενα μη γραμμικότητας στις αποκρίσεις. Για παράδειγμα, στην πλευρική ζώνη (lateral belt) rhesus πιθήκων, η βέλτιστη στιγμιαία συχνότητα εξαρτάται από το ρυθμό του sweep αλλά όχι με έναν γραμμικό και προβλέψιμο τρόπο. Στον AI έχουν εντοπιστεί τέτοιες γραμμικές εξαρτήσεις. Από τα παραπάνω, φαίνεται πως, στους πιθήκους rhesus τουλάχιστον, η πλευρική ζώνη είναι ένα ανώτερο στάδιο επεξεργασίας από τον AI.

Άλλες μελέτες έχουν εστιάσει στην καταγραφή και χαρτογράφηση των ενδεχόμενων τοπογραφικών χαρτών που μπορεί να υπάρχουν στον φλοιό. Τα αποτελέσματα έχουν αποκαλύψει μία μη γραμμική εξάρτηση μεταξύ συχνότητας και τοποθεσίας στον εγκέφαλο ([27]). Στο σχήμα 2.26 φαίνεται η πορεία ενεργοποίησης περιοχών του εγκεφάλου κατά την παρουσίαση FM sweep. Τα “κρύα” χρώματα, όπως το μπλε, αντιστοιχούν σε υψηλές συχνότητες, κατά την έναρξη του ερεθίσματος, και τα “θερμά” χρώματα σε χαμηλές συχνότητες. Η εικόνα δεν προέρχεται από ένα μόνο υποκείμενο του πειράματος αλλά από τον συνδυασμό αποτελεσμάτων από όλα τα υποκείμενα της εν λόγω μελέτης.



Σχήμα 2.26: Η σχέση μεταξύ περιοχών του εγκεφάλου και συχνότητας ενός FM sweep, [27]

Πιο λεπτομερείς μελέτες ([24]) έχουν αποκαλύψει την ύπαρξη χαρτών και συγκεκριμένων διαδρομών ενεργοποίησης που ακολουθούνται στον φλοιό κατά την παρουσίαση sweep. Ο ανθρώπινος ακουστικός φλοιός διαθέτει τουλάχιστον 8 περιοχές που είναι ευαίσθητες είτε σε χαμηλές είτε σε υψηλές συχνότητες. Στο σχήμα 2.27 φαίνονται ευδιάκριτα μερικές τέτοιες περιοχές. Καθώς το σήμα σαρώνει τις συχνότητες, διαφορετικές περιοχές αντιδρούν με διαφορετική καθυστέρηση, επομένως και σε διαφορετική συχνότητα. Το μέρος του φλοιού που απεικονίζεται περιλαμβάνει την έλικα του Heschl (Heschl's gyrus, HG), την αύλακα του Heschl (Heschl's sulcus, HS) και την άνω κροταφική έλικα (superior temporal gyrus, STG). Εξαιρετικό ενδιαφέρον παρουσιάζει επίσης και η χρονική εξέλιξη της δραστηριότητας αυτών των περιοχών. Στην εικόνα 2.28 μπορούμε να δούμε εννέα στιγμιότυπα ενεργοποίησης του φλοιού. Σε κάθε επιμέρους εικόνα, οι άσπρες περιοχές αποτελούν μέτωπα κύματος υπό την έννοια ότι πρόκειται για περιοχές που παρουσίασαν δραστηριότητα



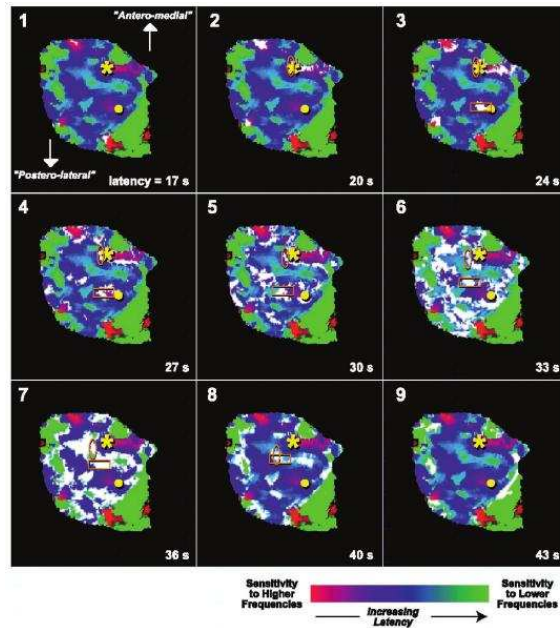
Σχήμα 2.27: Χάρτης ευαισθησίας συχνοτήτων σε τμήμα ανθρώπινου ακουστικού φλοιού, [24]

εντός ενός χρονικού παραθύρου 6.4 ms γύρω από τη χρονική στιγμή που είναι σημειωμένη σε κάθε περίπτωση. Είναι προφανής η κίνηση των μετώπων από τις κόκκινες περιοχές, που αντιστοιχούν σε υψηλές συχνότητες, προς τις πράσινες των χαμηλότερων συχνοτήτων. Η έλλειψη και το παραλληλόγραμμο που είναι σημειωμένα στις εικόνες ιχνηλατούν την κίνηση δύο μεμονωμένων μετώπων τα οποία έχουν διαφορετικές αφετηρίες αλλά τελικά καταλήγουν στο ίδιο σημείο. Η συγκεκριμένη μελέτη μπόρεσε να εντοπίσει έξι τέτοιες διαδρομές που ήταν παρούσες κατ' επανάληψη στα υποκείμενα. Ο αριθμός των διαδρομών πιθανότατα είναι μεγαλύτερος αφού κάποιες άλλες δεν συμπεριλήφθηκαν λόγω της μεταβλητότητας τους από υποκείμενο σε υποκείμενο, κάτι που μπορεί να οφείλεται και σε ιδιοσυγκρασιακούς παράγοντες των υποκείμενων, αν λάβουμε υπόψιν και την πλαστικότητα του εγκεφάλου.

2.2.2.8 Αποκρίσεις σε πιο περίπλοκα ερεθίσματα

Η στροφή προς όλο και πιο περίπλοκα ερεθίσματα για τη μελέτη του ακουστικού φλοιού υπαγορεύτηκε από την συσσώρευση όλο και περισσότερων πειραματικών ευρημάτων που παρείχαν ενδείξεις ότι οι φλοιικοί νευρώνες δεν αποδομούν ένα σήμα με βάση μόνο τα φασματικά χαρακτηριστικά του. Αντιθέτως, είναι οι γενικότερες χρονοφασματικές ιδιότητες του που διεγείρουν εντονότερα τους νευρώνες.

Ένα από τα πιο περίπλοκα ερεθίσματα που έχουν χρησιμοποιηθεί είναι τα δυναμικά φασματικά κύματα (dynamic ripple spectra). Τα στατικά φασματικά κύματα είναι απλώς ευρυζωνικός θόρυβος το φασματικό περιεχόμενο του οποίου όμως διαμορφώνεται ημιτονικά



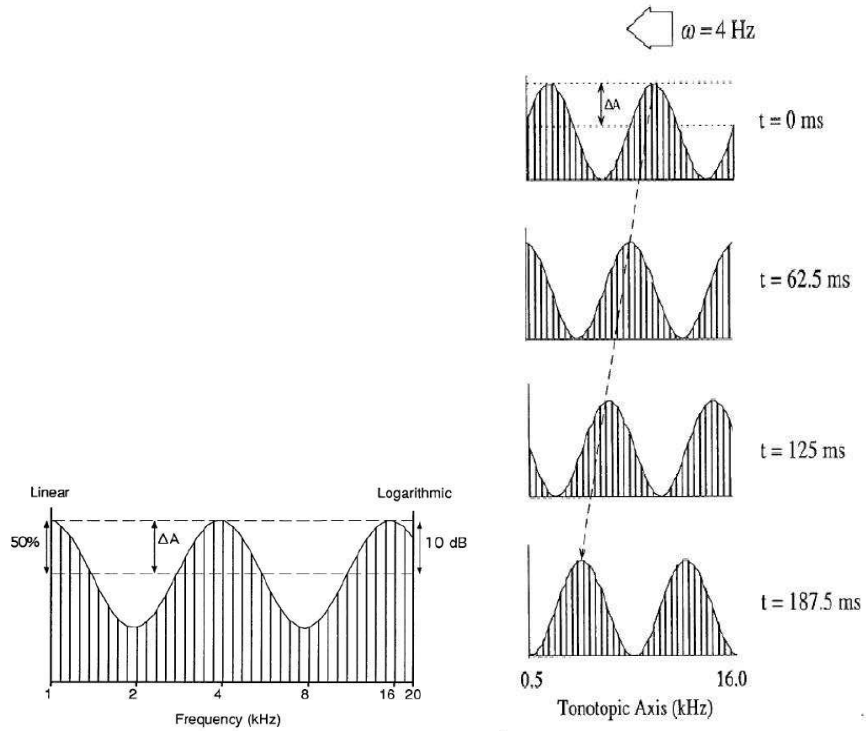
Σχήμα 2.28: Χρονική εξέλιξη της δραστηριότητας του φλοιού σε ένα FM sweep, [24]

με ένα ρυθμό Ω (cycles/octave) (βλ. σχήμα 2.29(α')). Ονομάζονται στατικά γιατί το φάσμα τους δεν μεταβάλλεται με τον χρόνο. Τα δυναμικά ή κινούμενα φασματικά κύματα μετακινούνται στο χρόνο με ένα ρυθμό ω και έχουν εξίσωση :

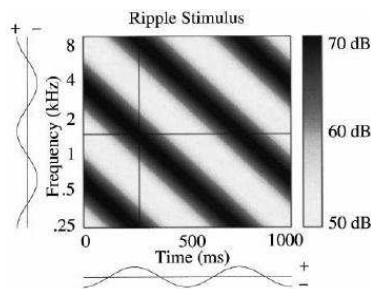
$$S(t, x) = 1 + A \sin(2\pi(\omega t + \Omega x) + \Phi)$$

Ένα παράδειγμα κινούμενου κύματος και ένα τυπικό φασματόγραμμα ενός δυναμικού κύματος παρουσιάζονται στο σχήμα 2.29(β') και 2.29(γ') αντίστοιχα. Τυπικές τιμές για τους ρυθμούς ω και Ω , για τους οποίους οι νευρώνες αντιδρούσαν βέλτιστα, είναι $\omega < 25$ Hz (με μερικούς να φτάνουν μέχρι 100) και $\Omega < 1.6$ cycles/octaves (με μερικούς να φτάνουν μέχρι 4) αντίστοιχα. Τα αποτελέσματα από τέτοιου είδους πειράματα επιβεβαιώνουν την υπόθεση της αναπαράστασης σε πολλαπλά επίπεδα διακριτικής ικανότητας. Περισσότερες λεπτομέρειες θα δοθούν στο επόμενο κεφάλαιο.

Παρόμοια ερεθίσματα είναι και οι λεγόμενες ακολουθίες από FM sweep (trains of FM sweeps), δηλαδή FM sweep που επαναλαμβάνονται με συγκεκριμένο ρυθμό. Μία μελέτη που χρησιμοποίησε τέτοια ερεθίσματα είναι αυτή των Orduna et al ([17]). Οι παράμετροι που μελετήθηκαν ήταν ο ρυθμός επανάληψης των sweep και η κατεύθυνσή τους. Για ρυθμούς επανάληψης μέχρι 10 sweeps/sec οι νευρώνες αντιδρούσαν συγχρονισμένα με το ερέθισμα, ακολουθώντας την περιοδικότητά του. Για υψηλότερους ρυθμούς όμως η απόκρισή τους περιοριζόταν κατά την έναρξη (onset) του ερεθίσματος (βλ. σχήμα 2.30, [17], [14]). Ο βέλτιστος ρυθμός επανάληψης ήταν γύρω στα 4 με 6 sweeps/sec. Μόνο ένα πολύ μικρό ποσοστό (0.4%) βρέθηκε να αντιδρά κατά την παύση (offset) του ερεθίσματος. Αρκετοί νευρώνες (57%) επέδειξαν επίσης προτιμήσεις κατεύθυνσης με τους περισσότερους από αυτούς να είναι πιο ευαίσθητοι σε καθοδικά sweep.



(α) Παράδειγμα στατικού φασματικού κύματος, [21] (β) Παράδειγμα κινούμενου φασματικού κύματος, [13]

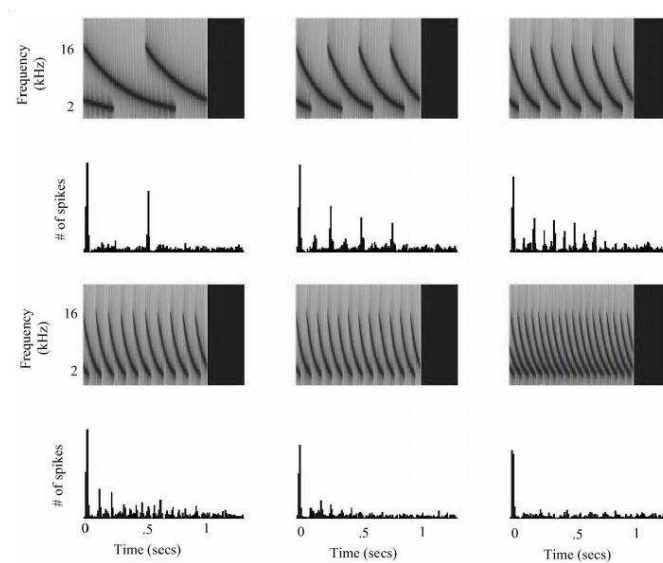


(γ) Τυπικό φασματόγραμμα κινούμενου φασματικού κύματος, [4]

Σχήμα 2.29: Φασματικά κύματα

Κεφάλαιο 2. Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα

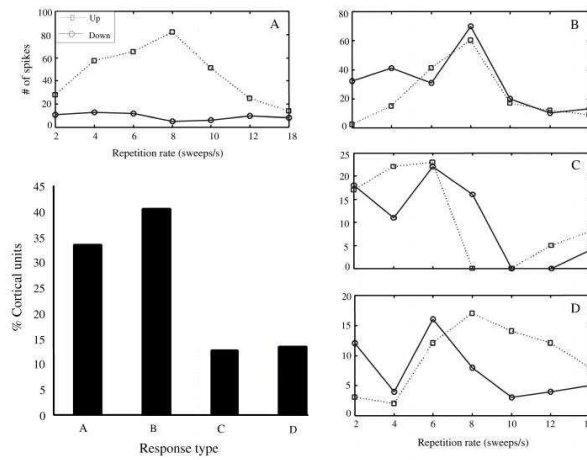
Ενδιαφέρουσα είναι η αλληλεπίδραση μεταξύ κατεύθυνσης και ρυθμού επανάληψης. Σε γενικές γραμμές, σε χαμηλούς και μεσαίους ρυθμούς επανάληψης, υπήρχαν περισσότεροι νευρώνες που αντιδρούσαν καλύτερα σε καθοδικά sweep ενώ σε υψηλότερους ρυθμούς, άνω των 10 sweeps/sec, παρατηρείται μία αντιστροφή με το μεγαλύτερο ποσοστό να προτιμάει ανοδικά sweep. Μερικοί νευρώνες διατηρούν την επιλεκτικότητα κατεύθυνσης για όλους τους ρυθμούς επανάληψης, άλλοι αντιστρέφουν την επιλεκτικότητά τους και κάποιοι άλλοι εμφανίζουν επιλεκτικότητα μόνο για μία περιοχή τιμών ρυθμού επανάληψης. Ενδεικτικές συναρτήσεις νευρωνικής δραστηριότητας συναρτήσει του ρυθμού επανάληψης φαίνονται στο σχήμα 2.31.



Σχήμα 2.30: Συνολική δραστηριότητα των νευρώνων ως απόκριση σε ακολουθίες FM sweep με διάφορους ρυθμούς επανάληψης, [17]

2.2.2.9 Αποκρίσεις σε άλλα ερεθίσματα

Στον ακουστικό φλοιό έχουν εντοπιστεί νευρώνες που έχουν την ιδιότητα να κωδικοποιούν την έναρξη (onset) των ερεθισμάτων ([9]). Οι νευρώνες αυτοί είναι εξειδικευμένοι στη σύλληψη των ενάρξεων και όχι απλώς μία υποπερίπτωση κάποιας άλλης κλάσης νευρώνων. Φιλοξενούνται σε περιοχές που είναι χωρικά διαχωρισμένες από άλλες περιοχές νευρώνων διαφορετικής λειτουργικότητας, όπως αυτοί που αντιδρούν καλύτερα σε σταθερά στο χρόνο (steady-state) σήματα. Οι αποκρίσεις των onset νευρώνων παρουσιάζουν εξάρτηση από διάφορες παραμέτρους του σήματος, όπως το SPL και ο χρόνος ανόδου (rise time). Ωστόσο, προσεκτικότερη ανάλυση αποκαλύπτει ότι η απόκρισή τους είναι ουσιαστικά μία ολοκλήρωση του ρυθμού μεταβολής της πίεσης εντός ενός χρονικού παραθύρου. Το παράθυρο ολοκλήρωσης έχει αφετηρία την έναρξη του τόνου και διάρκεια αντιστρόφως ανάλογη του ρυθμού μεταβολής και της ευαισθησίας του νευρώνα σε απότομες μεταβολές. Πρόσφατα ([19]), εντοπίστηκαν στον ΑΙ γατών νευρώνες που σηματοδοτούν την παύση

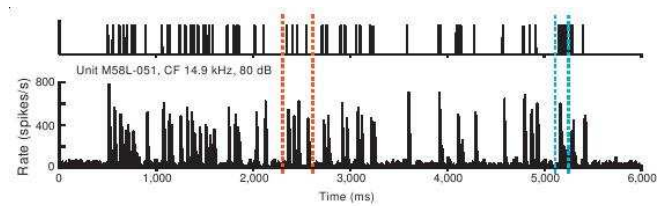


Σχήμα 2.31: Εξάρτηση της επιλεκτικότητας κατεύθυνσης συναρτήσει του ρυθμού επανάληψης, [17]

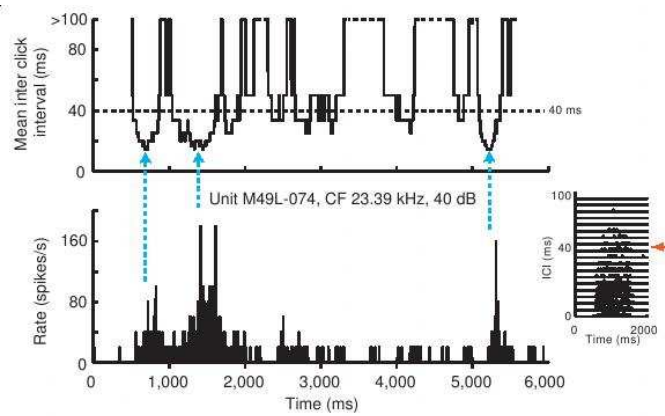
(offset) του ερέθισματος. Οι αποκρίσεις κατά το τέλος του ερέθισματος συνοδεύονταν συνήθως από αντίστοιχες αποκρίσεις και κατά την έναρξη έτσι ώστε αυτά τα κύτταρα να χαρακτηρίζονται ως ON-OFF κύτταρα. Το υποδεκτικό πεδίο (frequency receptive field, FRF) του κυττάρου κατά την έναρξη μπορεί να διέφερε αρκετά από αυτό κατά την παύση αλλά και τα δύο παρέμεναν σταθερά για διάφορες τιμές έντασης. Η ύπαρξη τέτοιων νευρώνων υποδεικνύει ότι οι παύσεις των ερεθισμάτων κωδικοποιούνται ενεργητικά στον ΑΙ και όχι με κάποιον έμμεσο τρόπο.

Οι Lu et al ([14]) μελέτησαν τις αποκρίσεις του ακουστικού φλοιού πιθήκων σε ακολουθίες από κλικ (click). Τα αποτελέσματα έδειξαν ότι μπορεί να γίνει ένας διαχωρισμός σε δύο κλάσεις νευρώνων. Στην πρώτη, οι νευρώνες έχουν τη δυνατότητα να συγχρονίζονται με το ερέθισμα. Για χρονικές αποστάσεις μεταξύ διαδοχικών κλικ κοντά στα 100 ms, το χρονικό μοτίβο εκφόρτισης αυτών των νευρώνων αντανάκλα τις χρονικές αποστάσεις στο ερέθισμα. Όταν οι αποστάσεις μεταξύ των κλικ είναι κάτω από 15 ms τότε η αντίδραση αποτελείται από ένα onset ακολουθούμενο από αδράνεια. Όπως φαίνεται και στο σχήμα 2.32(α'), όταν η απόσταση μεταξύ των κλικ είναι αρκετά μεγάλη, οι νευρώνες μπορούν να ακολουθήσουν το ερέθισμα, ικανότητα που χάνουν καθώς η απόσταση μειώνεται. Η δεύτερη κλάση νευρώνων είναι αυτοί που δεν μπορούν να συγχρονιστούν με το ερέθισμα. Παρ' όλα αυτά, μπορούν να κωδικοποιήσουν ακόμα και σήματα με μικρές αποστάσεις μεταξύ των κλικ. Αυτό το επιτυγχάνουν με το να παρουσιάζουν έντονη αλλά μη συγχρονισμένη δραστηριότητα στα γρήγορα κλικ με την ένταση της δραστηριότητας να μειώνεται, μέχρι ένα σημείο μονοτονικά, καθώς το διάστημα μεταξύ των κλικ αυξάνεται (βλ. σχήμα 2.32(β')). Με αυτόν τον τρόπο, οι συγκεκριμένοι νευρώνες παρουσιάζουν ευαισθησία στις γρήγορες ακολουθίες από κλικ (ευαισθησία εξαρτώμενη από την ταχύτητα της ακολουθίας) αλλά μειωμένη απόκριση για αποστάσεις άνω των 40 ms.

Κεφάλαιο 2. Η διαδρομή των σημάτων λόγου στο νευρικό σύστημα



(α') Απόκριση συγχρονισμένων νευρώνων σε ακολουθία από κλικ



(β') Απόκριση μη συγχρονισμένων νευρώνων σε ακολουθία από κλικ

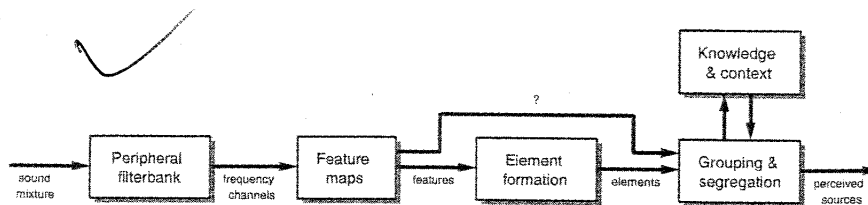
Σχήμα 2.32: Αποκρίσεις σε ακολουθίες από κλικ, [14]

Κεφάλαιο 3

Υπολογιστική και μαθηματική μοντελοποίηση του ακουστικού συστήματος

Μετά τη μελέτη του ακουστικού συστήματος σε επίπεδο ανατομίας και νευροφυσιολογίας το λογικά επόμενο βήμα είναι η προσπάθεια μαθηματικής και υπολογιστικής μοντελοποίησής του. Ένα μοντέλο του ακουστικού συστήματος, ακόμα και αν μοντελοποιεί μόνο μερικά από τα χαρακτηριστικά του, μπορεί να είναι εξαιρετικά χρήσιμο στην περαιτέρω κατανόηση και εμπάθυνση των γνώσεών μας περί του τρόπου που ο ήχος (και ο λόγος ειδικότερα) γίνεται τελικά αντιληπτός. Αν μη τι άλλο, προσφέρει τη δυνατότητα να ελεγχθεί η θεωρία που υπάρχει από πίσω του να το στηρίζει και τα όρια ισχύος της.

Αρκετές τέτοιες προσπάθειες έχουν γίνει στο παρελθόν, αρχικά βασισμένες κυρίως σε ψυχοακουστικά δεδομένα αλλά στη συνέχεια και σε νευροφυσιολογικά. Φυσικά, το εύρος της επιτυχίας τους περιοριζόταν και περιορίζεται ακόμα από το σημαντικό έλλειμμα γνώσης που υπάρχει όσον αφορά στους ακριβείς νευροβιολογικούς μηχανισμούς που εμπλέκονται στην επεξεργασία των ηχητικών σημάτων. Ωστόσο, προϊόντος του χρόνου, τα μοντέλα που προτείνονται γίνονται όλο και πιο πολύπλοκα και εκλεπτυσμένα, με την ελπίδα ότι κάποια στιγμή θα αρχίσει να μας αποκαλύπτεται η “εσωτερική αναπαράσταση” του λόγου. Σε γενικές γραμμές, η δομή του ακουστικού συστήματος που είναι λίγο ως πολύ γενικώς αποδεκτή απεικονίζεται στο σχήμα 3.1 (βλ. κεφ.24, [7]). Η συστοιχία φίλτρων του κοχλίου



Σχήμα 3.1: Η δομή του ακουστικού συστήματος, [7]

και του AN αποτελούν ένα πρώτο στάδιο μετασχηματισμού η έξοδος του οποίου τροφοδοτεί τα πιο εσωτερικά συστήματα. Αυτά με τη σειρά τους εξάγουν τα σημαντικά χαρακτηριστικά του σήματος ομιλίας σε μία ενδιάμεση αναπαράσταση η οποία και αποτελεί την είσοδο σε ένα γνωσιακό μοντέλο που χειρίζεται αυτά τα χαρακτηριστικά με βάση προϋπάρχουσα γνώση και συμφραζόμενα ώστε να προκύψει τελικά το νόημα. Αξίζει να σημειωθεί ότι η διαδικασία πολύ πιθανό να μην είναι τόσο ευθεία και bottom-up όσο παρουσιάζεται εδώ αλλά να υπάρχουν και top-down βρόχοι ανάδρασης. Η ενδιάμεση αναπαράσταση είναι χρήσιμη γιατί, από τη μία, μπορεί να μοντελοποιηθεί με άμεσο τρόπο και, από την άλλη, είναι ευέλικτη παρέχοντας αρκετά περιθώρια ελευθερίας στα ανώτερα, γνωσιακά επίπεδα επεξεργασίας για τα οποία και υπάρχουν περισσότερες αμφιβολίες. Είναι ενδιαφέρον πάντως ότι δεν υπάρχει καθολική ομοφωνία ως προς το ποια ακριβώς χαρακτηριστικά είναι τα σημαντικά για την κατανόηση του λόγου και ποια η σημασία του καθενός. Ακόμα μεγαλύτερη ασυμφωνία υπάρχει ως προς τα τελικά στάδια αυτής της δομής. Ο κύριος όγκος της έρευνας έχει επικεντρωθεί στο περιφερικό ακουστικό σύστημα αλλά παρατηρείται μία στροφή προς τη μελέτη του κεντρικού ακουστικού συστήματος που θα επιτρέψει να συμπληρωθεί όλη η αλυσίδα μετασχηματισμών.

Στη συνέχεια αυτού του κεφαλαίου παρουσιάζονται ορισμένα μοντέλα που έχουν προταθεί, ξεκινώντας από τα απλούστερα και προχωρώντας προς τα πιο πρόσφατα και πολύπλοκα.

3.1 Μοντέλα βασισμένα σε ψυχοακουστικά δεδομένα

Σημαντική ώθηση στα μοντέλα που βασίζονται σε ψυχοακουστικά δεδομένα δόθηκε από την ανάπτυξη αυτόματων συστημάτων αναγνώρισης φωνής (Automatic Speech Recognition, ASR). Για να μπορέσουν οι αλγόριθμοι τέτοιων συστημάτων να είναι αποδοτικοί έπρεπε το αρχικό σήμα να τροφοδοτείται στην είσοδό τους σε μία συμπιεσμένη μορφή. Επίσης, το ακουστικό σύστημα δεν χρησιμοποιεί όλη την πληροφορία που είναι διαθέσιμη μέσω του αρχικού σήματος λόγου για να το αποκωδικοποιήσει. Υπάρχει ένας πλεονασμός αρχικής πληροφορίας από την οποία το ακουστικό σύστημα μπορεί και απομονώνει αυτά τα χαρακτηριστικά που θα του επιτρέψουν να επιτύχει την αναγνώριση. Προς αυτή την κατεύθυνση, αναπτύχθηκαν αναπαραστάσεις του ηχητικού σήματος οι οποίες λάμβαναν υπόψιν τους αποτελέσματα από ψυχοακουστικά πειράματα ώστε να μειωθεί ο όγκος της προς επεξεργασία πληροφορίας.

3.1.1 Mel-frequency cepstrum coefficients

Μία αναπαράσταση που χρησιμοποιείται αρκετά συχνά ως εμπρόσθιο τμήμα (front-end) σε συστήματα ASR είναι αυτή των mel-frequency cepstrum coefficients (mfcc) ([3]). Σε μία σειρά πειραμάτων αναγνώρισης μονοσυλλαβικών, CVC λέξεων συγκρίθηκε η απόδοση της αναπαράστασης σε mfcc με αυτή από άλλες αναπαραστάσεις, όπως linear-frequency cepstrum coefficients και LPC (Linear Predictive Coding)¹. Τελικά, τα ποσοστά αναγνώ-

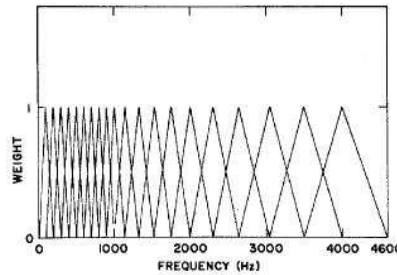
¹Τεχνική μοντελοποίησης που αναπαριστά το σύστημα παραγωγής φωνής υπό τη μορφή ενός all-pole φίλτρου.

ρισης με βάση τους συντελεστές mfcc ήταν τα υψηλότερα, γεγονός που μπορεί να αποδοθεί στην καλύτερη αναπαράσταση των αντιληπτικά κρίσιμων περιοχών του φάσματος.

Για να προκύψει η αναπαράσταση σε συντελεστές mfcc, το αρχικό σήμα τμηματοποιείται σε πλαίσια και πολλαπλασιάζεται με ένα παράθυρο Hamming. Στη συνέχεια, κάθε πλαίσιο περνάει από μία συστοιχία φίλτρων η οποία ακολουθεί την ψυχοακουστική κλίμακα mel^2 , όπως φαίνεται στο σχήμα 3.2. Ακολουθεί υπολογισμός της ενέργειας από την έξοδο κάθε φίλτρου και του λογαρίθμου αυτής X_k , $k = 1, 2, \dots, Q$, όπου Q ο αριθμός των φίλτρων της συστοιχίας. Οι συντελεστές mfcc τελικά υπολογίζονται από την εξίσωση :

$$MFCC_i = \sum_{k=1}^Q X_k \cos i(k - \frac{1}{2}) \frac{\pi}{20}$$

όπου $i = 1, 2, \dots, M$, M ο αριθμός των συντελεστών. Έχει ενδιαφέρον να παρατηρήσουμε τη διαφορά μεταξύ ενός συνηθισμένου φασματογράμματος (σχήμα 3.3(α')) και ενός φασματογράμματος που ανακατασκευάζεται από mfcc συντελεστές (σχήμα 3.3(β')). Είναι εμφανές ότι το mfcc φασματογράμμα διατηρεί τα γενικά χαρακτηριστικά του αρχικού φασματογράμματος, που πιθανώς είναι και τα σημαντικά για την αποκωδικοποίηση του μηνύματος, και μειώνει τις πολλές λεπτομέρειες και τον “θόρυβο” (όπως τις αρμονικές), που πιθανότερα αφορούν σε ιδιοσυγκρασιακά χαρακτηριστικά του ομιλητή.

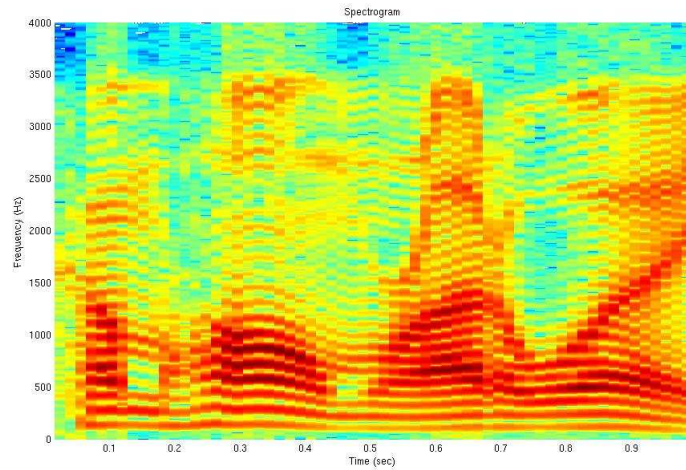


Σχήμα 3.2: Συστοιχία φίλτρων που ακολουθεί την κλίμακα mel, [3]

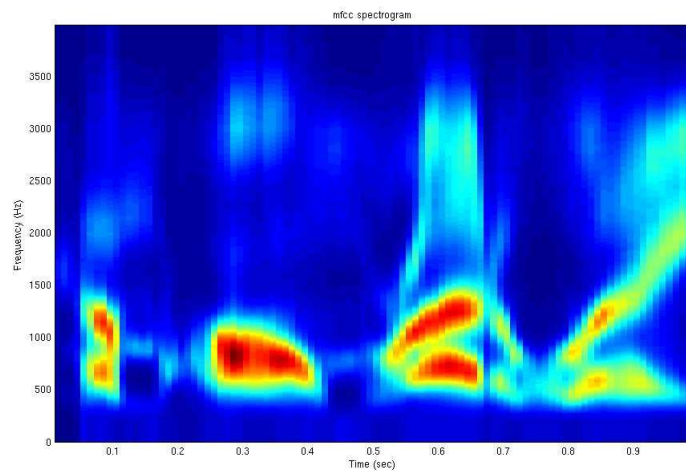
3.1.2 Αναπαράσταση με βάση την PLP (perceptual linear predictive) τεχνική

Στην PLP αναπαράσταση γίνεται προσπάθεια να ενταχθούν περισσότερα ψυχοακουστικά φαινόμενα στους μετασχηματισμούς του ηχητικού σήματος. Όπως και στην περίπτωση της mfcc αναπαράστασης, εφόσον το ανθρώπινο ακουστικό σύστημα έχει μεταβλητή διακριτική ικανότητα στις διάφορες συχνότητες, η ανάλυση δεν γίνεται επί μίας γραμμικής κλίμακας συχνοτήτων αλλά με βάση μία κλίμακα κρίσιμων ζωνών (critical bands). Εδώ επιλέγεται

²Με βάση αυτή την κλίμακα γίνεται στην πραγματικότητα η διάκριση από το ανθρώπινο αυτί μεταξύ των διαφορετικών συχνοτήτων. Όπως φαίνεται και από το σχήμα 3.2, στις χαμηλές συχνοότητες η διακριτική ικανότητα είναι σαφώς καλύτερη.



(α') Συνηθισμένο φασματόγραμμα (MATLAB)

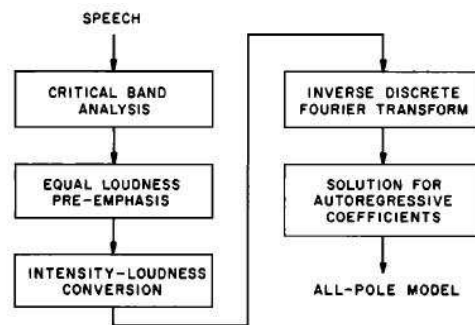


(β') Φασματόγραμμα με βάση συντελεστές mfcc, όπως υπολογίζεται από το Auditory Toolbox του Malcolm Slaney, <http://cobweb.ecn.purdue.edu/malcolm/interval/1998-010/>

Σχήμα 3.3: Σύγκριση συνηθισμένου και mfcc φασματογράμματος

Κεφάλαιο 3. Υπολογιστική και μαθηματική μοντελοποίηση του ακουστικού συστήματος

η κλίμακα Bark³ αντί για την mel της mfcc αναπαράστασης. Στη συνέχεια, εφαρμόζεται προέμφαση στο σήμα με μία καμπύλη η οποία προσομοιώνει την ευαισθησία στα επίπεδα SPL της ανθρώπινης ακοής για τις διάφορες συχνότητες του φάσματος. Ένας ακόμα μετασχηματισμός επιβάλλεται ώστε να προσομοιωθεί και ο νόμος κυβικής ρίζας πραγματικής - αντιληπτής έντασης που διέπει το ακουστικό σύστημα και προσομοιώνει τη μη γραμμική σχέση μεταξύ της πραγματικής έντασης του ήχου και αυτής που τελικά γίνεται αντιληπτή. Τέλος, από το σήμα που έχει προκύψει κατόπιν των παραπάνω μετασχηματισμών υπολογίζονται οι LPC συντελεστές ενός μοντέλου χωρίς μηδενικά (all-pole). Όλη η αλυσίδα αυτών των σταδίων επεξεργασίας φαίνεται σχηματικά στο σχήμα 3.4.



Σχήμα 3.4: Σχηματικό διάγραμμα των σταδίων επεξεργασίας στην PLP αναπαράσταση, [11]

Το πρώτο βήμα είναι η τμηματοποίηση του σήματος εισόδου με παράθυρα Hamming της μορφής

$$W(n) = 0.54 + 0.46 \cos \frac{2\pi n}{N-1}$$

όπου N το μήκος του. Συνήθης διάρκεια του παραθύρου είναι τα 20 ms. Σε κάθε παραθυρωμένο πλαίσιο εφαρμόζεται ο μετασχηματισμός Fourier (συνήθως FFT) και το φάσμα ισχύος υπολογίζεται από το πραγματικό και φανταστικό του μέρος ως :

$$P(\omega) = \Re[S(\omega)]^2 + \Im[S(\omega)]^2$$

Για την ανάλυση σε κρίσιμες ζώνες, αρχικά αναδιπλώνεται ο άξονας συχνοτήτων σε έναν άξονα Bark μέσω της εξίσωσης :

$$\Omega(\omega) = 6 \ln \frac{\omega}{1200\pi} + \left[\left(\frac{\omega}{1200\pi} \right)^2 + 1 \right]^{0.5}$$

Το προκύπτον αναδιπλωμένο φάσμα συνελλίσσεται με μία καμπύλη κρίσιμων ζωνών $\Psi(\Omega)$

³ Αυτή η κλίμακα, αν και περίπου λογαριθμικής μορφής, όπως η mel, βασίζεται στα φαινόμενα ηχοκάλυψης που παρατηρούνται στο ακουστικό σύστημα όπου ένας τόνος ενδέχεται να μην είναι αντιληπτός αν η συχνότητά του εμπίπτει εντός της κρίσιμης ζώνης ενός άλλου τόνου που είναι ταυτοχρόνως παρών.

που δίνεται από :

$$\Psi(\Omega) = \begin{cases} 0 & \text{για } \Omega < -1.3 \\ 10^{2.5(\Omega+0.5)} & \text{για } -1.3 \leq \Omega \leq -0.5 \\ 1 & \text{για } -0.5 < \Omega < 0.5 \\ 10^{-1.0(\Omega-0.5)} & \text{για } 0.5 \leq \Omega \leq 2.5 \\ 0 & \text{για } \Omega > 2.5 \end{cases}$$

Η συνέλιξη μάς δίνει :

$$\Theta(\Omega_i) = \sum_{\Omega=-1.3}^{2.5} P(\Omega - \Omega_i)\Psi(\Omega)$$

Η χαμηλή διακριτική ικανότητα της $\Theta(\Omega)$ επιτρέπει τη δειγματοληψία της σε διαστήματα του 1 Bark. Η προσομοίωση της καμπύλης ίσης έντασης (equal-loudness curve) επιτυγχάνεται με προέμφαση με βάση την προσέγγιση :

$$\Xi[\Omega(\omega)] = E(\omega)\Theta[\Omega(\omega)]$$

όπου

$$E(\omega) = \frac{(\omega^2 + 56.8x10^6)\omega^4}{(\omega^2 + 6.3x10^6)^2(\omega^2 + 0.38x10^9)(\omega^6 + 9.58x10^{26})}$$

Η παραπάνω καμπύλη αντιστοιχεί στα 40 dB και έχει κλίση 12 dB/oct μεταξύ 0 και 400 Hz, 0 dB/oct μεταξύ 400 και 1200 Hz, 6 dB/oct μεταξύ 1200 και 3100 Hz, 0 dB/oct μεταξύ 3100 και 5000 Hz και -18 dB/oct από 5000 Hz και πάνω. Το προκύπτον φάσμα συμπιέζεται με βάση την εξίσωση :

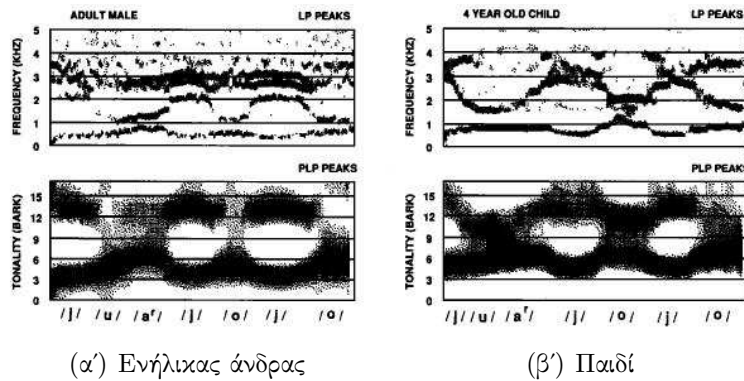
$$\Phi(\Omega) = \Xi(\Omega)^{0.33}$$

Με αυτόν τον τρόπο προσομοιώνεται η μη γραμμικότητα μεταξύ πραγματικής και αντιληπτής έντασης. Τέλος, εφαρμόζεται ένας αντίστροφος μετασχηματισμός Fourier και υπολογίζονται οι LPC συντελεστές. Πειράματα έδειξαν ότι ένα all-pole μοντέλο τάξης 5 δίνει τα καλύτερα ποσοστά επιτυχούς αναγνώρισης.

Ένα σημαντικό πλεονέκτημα της PLP αναπαράστασης είναι ότι απορρίπτει από το φάσμα την πληροφορία που εξαρτάται από τα ιδιαίτερα χαρακτηριστικά του ομιλητή διατηρώντας αυτά που είναι κρίσιμα για την αποκωδικοποίηση του μηνύματος. Στο σχήμα 3.5 φαίνονται τα φασματογράμματα της ίδιας φράσης, όπως εκφωνείται από ένα παιδί και από έναν ενήλικα άνδρα, έτσι όπως προκύπτουν από την απλή LP μέθοδο σε σχέση με αυτά από την PLP. Είναι εμφανές ότι τα LP φασματογράμματα διαφέρουν σε σημαντικό βαθμό τόσο όσο προς τη θέση όσο και ως προς τον αριθμό των μορφικών ενώ τα PLP φασματογράμματα παρουσιάζουν αξιοσημείωτη ομοιομορφία.

3.1.3 Το μοντέλο των Tchorz και Kollmeier

Οι Tchorz και Kollmeier ([25]) πρότειναν ένα μοντέλο που επιχειρεί να ενσωματώσει ακόμα περισσότερα φαινόμενα που έχουν παρατηρηθεί σε ψυχοακουστικά, αλλά και νευροφυσιολογικά, πειράματα. Για παράδειγμα, τέτοια φαινόμενα είναι αυτά της ηχοχάλυψης (masking) και της βραχυπρόθεσμης προσαρμογής (short-term adaptation). Έχει παρατηρηθεί ότι η



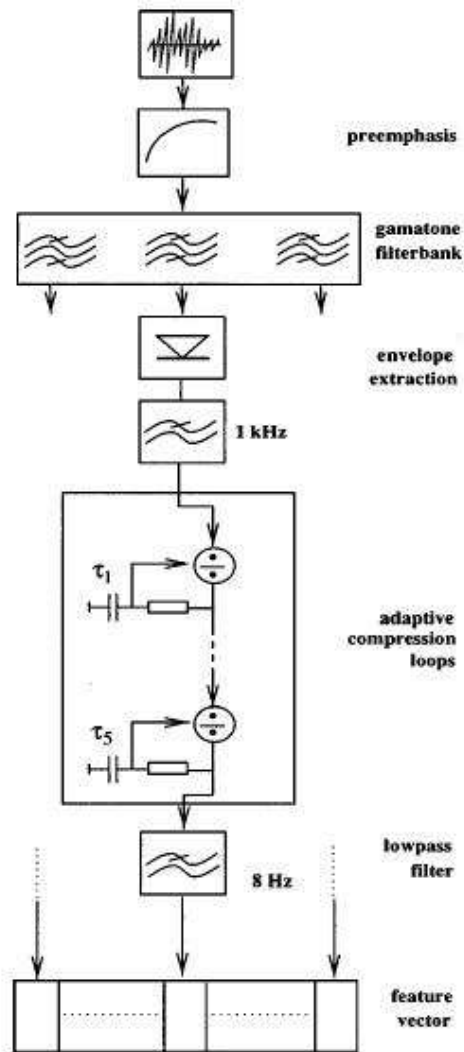
(α') Ενήλικας άνδρας

(β') Παιδί

Σχήμα 3.5: Σύγκριση LP με PLP , [11]

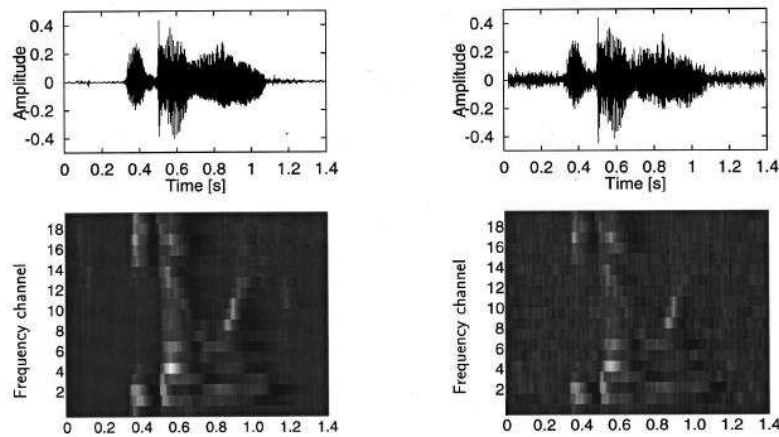
απόκριση των ινών του ακουστικού νεύρου μειώνεται καθώς ο χρόνος περνάει, όταν ένα ερέθισμα παραμένει σταθερό. Μετά την παύση του ερεθίσματος, έπεται μία περίοδος χαλάρωσης κατά την οποία η ίνα παρουσιάζει ρυθμό εκφόρτισης χαμηλότερο από τον αυθόρμητο της. Αυτά τα φαινόμενα αναφέρονται με τον όρο βραχυπρόθεσμη προσαρμογή. Επίσης, έχει βρεθεί ότι μεγάλο μέρος της ακουστικής πληροφορίας διατηρείται στις χαμηλές συχνότητας διαμορφώσεις πλάτους. Ακόμα και με πολύ μειωμένη φασματική πληροφορία, είναι δυνατό να επιτευχθεί αναγνώριση λέξεων αν έχει διατηρηθεί η πληροφορία διαμόρφωσης. Το μοντέλο που παρουσιάζεται στη συνέχεια παράγει μία αναπαράσταση που είναι συμβατή με τα παραπάνω φαινόμενα.

Η γενική δομή του μοντέλου φαίνεται στο σχήμα 3.6. Αρχικά εφαρμόζεται προέμφαση στο σήμα για να προσομοιωθεί η συνάρτηση μεταφοράς του έξω ωτός και στη συνέχεια το σήμα που προκύπτει περνάει μέσα από μία συστοιχία gammatone φίλτρων. Χρησιμοποιούνται 19 φίλτρα από 300 μέχρι 4000 Hz, κατανεμημένα γραμμικά σε κλίμακα ERB (Equivalent Rectangular Bandwidth, ισοδύναμο ορθογώνιο εύρος ζώνης) και με κρουστικές αποκρίσεις παρόμοιες με αυτές που έχουν παρατηρηθεί πειραματικά. Το επόμενο στάδιο περιλαμβάνει ανόρθωση ημίσεος κύματος και ένα βαθυπερατό φίλτρο με συχνότητα αποκοπής στα 1000 Hz, πάνω από την οποία υπάρχει και απώλεια της ικανότητας κλειδώματος φάσης από το ακουστικό νεύρο. Με αυτό το φίλτρο εξάγεται η περιβάλλουσα αλλά χάνονται οι λεπτομέρειες των υψηλών συχνοτήτων. Κατόπιν υπάρχει ένα στάδιο συμπίεσης του πλάτους το οποίο είναι προσαρμοστικό (adaptive) και όχι στατικό. Αποτελείται από πέντε βρόχους προσαρμογής και κάθε βρόχος περιλαμβάνει ένα βαθυπερατό φίλτρο και έναν διαιρέτη που διαιρεί το σήμα εισόδου με την έξοδο του φίλτρου. Οι σταθερές χρόνου των φίλτρων είναι διαφορετικές και κυμαίνονται από 5 ως 500 ms. Οι ταχείες, σε σχέση με τις σταθερές χρόνου, μεταβολές στο σήμα εισόδου υφίστανται απλά ένα γραμμικό μετασχηματισμό αλλά οι αργές μεταβολές έναν σχεδόν λογαριθμικό, της μορφής $Y = X^{\frac{1}{32}}$. Αυτός ο μετασχηματισμός ενισχύει τις ενάρξεις και τις παύσεις αλλά αποδυναμώνει τα χρονικά στατικά χαρακτηριστικά. Αυτά τα φίλτρα λειτουργούν και ως ένα είδος μνήμης για το σύστημα, γεγονός που επιτρέπει την μοντελοποίηση φαινομένων όπως η σκίαση και η βραχυπρόθεσμη προσαρμογή. Τέλος, υπάρχει και ένα άλλο βαθυπερατό φίλτρο με συχνότητα αποκοπής στα 8 Hz το οποίο εξασθενεί τις ταχείες μεταβολές της περιβάλλου-



Σχήμα 3.6: Δομή του μοντέλου των Tchorz και Kollmeier, [25]

σας. Παράδειγμα αναπαράστασης που προκύπτει από αυτό το μοντέλο φαίνεται στο σχήμα 3.7. Στο δεξί τμήμα, έχει προστεθεί στο αρχικό σήμα θόρυβος 10 dB παρόμοιος με αυτόν που συναντάται στην ομιλία (speech simulating noise). Οι ενάρξεις και οι παύσεις είναι πιο έντονες ενώ έχει μειωθεί η επίδραση του σχετικά στατικού θορύβου. Σε μία σειρά πειραμάτων αναγνώρισης λέξεων ανεξαρτήτως ομιλητή που πραγματοποιήθηκαν, το παρόν μοντέλο επέδειξε μεγαλύτερη ανοχή στον θόρυβο σε σχέση με την mfcc αναπαράσταση.



Σχήμα 3.7: Παράδειγμα αναπαράστασης με βάση το μοντέλο των Tchorz και Kollmeier, [25]

3.2 Μοντέλα βασισμένα σε νευροφυσιολογικά δεδομένα

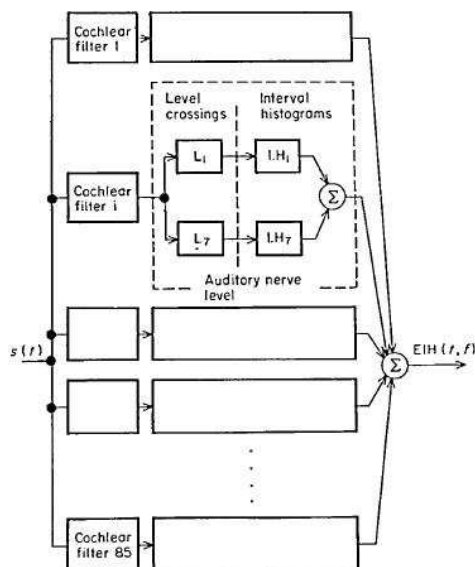
Η λεπτομερέστερη μελέτη του ακουστικού συστήματος, αυτή τη φορά σε ένα χαμηλότερο επίπεδο, επέτρεψε την ανάπτυξη μοντέλων βασισμένων σε νευροφυσιολογικά δεδομένα. Αυτά τα μοντέλα προσπαθούν να συλλάβουν με μαθηματικό τρόπο τους μετασχηματισμούς που υφίσταται ένα ηχητικό σήμα στα διάφορα στάδια του κατά τη διαδρομή του μέσα στο νευρικό σύστημα. Οι αναπαραστάσεις που προκύπτουν με αυτόν τον τρόπο μπορούν να μας βοηθήσουν να κατανοήσουμε με βάση ποια χαρακτηριστικά του αρχικού σήματος ο εγκέφαλος το αποσυνθέτει για να το επεξεργαστεί και τελικά να αναδυθεί το νόημα του. Όπως έχει αναφερθεί επανειλημμένα, η έρευνα έχει επικεντρωθεί μέχρι στιγμής στη μελέτη του πρώιμου, περιφερικού ακουστικού συστήματος με αποτέλεσμα τα περισσότερα μοντέλα να έχουν αναπτυχθεί ακριβώς για αυτό το στάδιο. Στη συνέχεια, παρουσιάζονται μερικά από αυτά.

3.2.1 Μοντελοποίηση του περιφερικού ακουστικού συστήματος

Τα πειράματα που έχουν γίνει στον κοχλία και στο ακουστικό νεύρο έχουν δείξει ότι, σε ένα πρώτο στάδιο, συμπεριφέρονται σαν μία συστοιχία ζωνοπερατών φίλτρων κατά μήκος της βασικής μεμβράνης. Όλα τα μοντέλα που θα παρουσιαστούν αποδέχονται αυτό το εύρημα αλλά μπορεί να διαφέρουν στα επόμενα στάδιά τους.

3.2.1.1 Η αναπαράσταση EIH (Ensemble Interval Histogram)

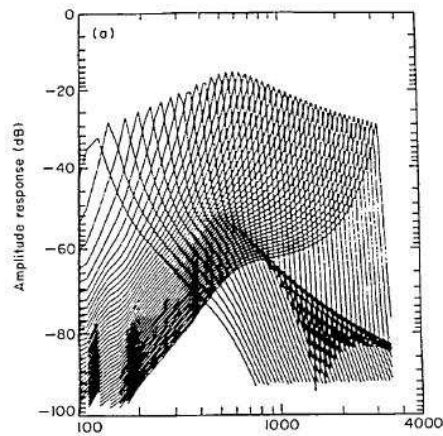
Ένα από αυτά ([6]), η λεγόμενη αναπαράσταση EIH, χρησιμοποιεί μία προσέγγιση που βασίζεται στις χρονικές πληροφορίες που μπορούν να εξαχθούν από το σήμα. Θεωρεί ότι μόνο η χρονική πληροφορία μεταδίδεται στο κεντρικό νευρικό σύστημα μέσω του ακουστικού νεύρου χωρίς αναφορά στη θέση της νευρικής ίνας επί της βασικής μεμβράνης. Το μοντέλο φαίνεται σχηματικά στην εικόνα 3.8. Το πρώτο του στάδιο αντιστοιχεί στον κοχλί-



Σχήμα 3.8: Η γενική δομή του μοντέλου EIH, [6]

α, πριν το ακουστικό νεύρο, και αποτελείται από μία λογαριθμικά κατανομημένη συστοιχία φίλτρων. Τα φίλτρα καλύπτουν τις συχνότητες από 200 μέχρι 3200 Hz και συνολικά είναι 85. Η απόκριση μερικών από αυτά παρουσιάζεται στο σχήμα 3.9.

Οι ίνες του ακουστικού νεύρου μοντελοποιούνται ως μία σειρά από ανιχνευτές οι οποίοι έπονται των φίλτρων. Κάθε ανιχνευτής δέχεται ως είσοδο την έξοδο του φίλτρου που του αντιστοιχεί και αποτελείται από μία σειρά επιπέδων έντασης λογαριθμικά κατανομημένων επί του δυναμικού εύρους του σήματος. Το σήμα αρχικά ανορθώνεται (half-wave rectification) και κάθε ανιχνευτής καταγράφει το πλήθος των φορών που το σήμα ξεπέρασε καθένα από τα επίπεδα έντασης. Η καταγραφή γίνεται σε ένα είδος ιστογράμματος όπου η τετμημένη αντιστοιχεί σε άξονα συχνότητας που υπολογίζεται ως το αντίστροφο του χρονικού

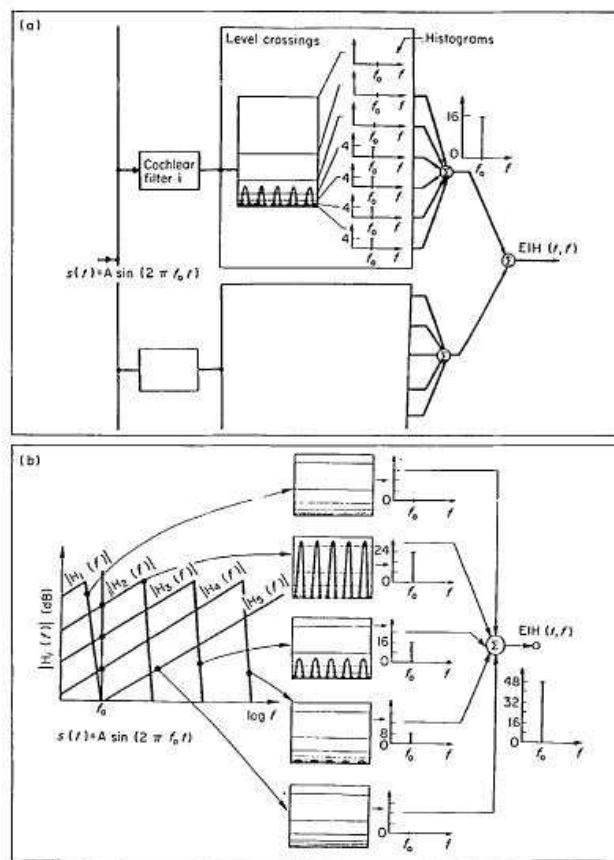


Σχήμα 3.9: Αποκρίσεις της συστοιχίας φίλτρων στο μοντέλο ΕΙΗ, [6]

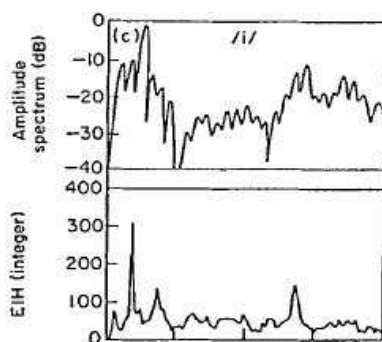
διαστήματος μεταξύ δύο διαδοχικών μεμονωμένων καταγραφών. Μόνο τα 20 πιο πρόσφατα χρονικά διαστήματα λαμβάνονται υπόψιν. Στη συνέχεια, τα ιστογράμματα από όλα τα επίπεδα έντασης συγκεντρώνονται αθροιστικά σε ένα ενιαίο ιστόγραμμα που αντιπροσωπεύει τη δραστηριότητα εκφόρτισης της συγκεκριμένης ίνας. Η παραπάνω διαδικασία φαίνεται στο σχήμα 3.10 όπου ως είσοδο έχουμε έναν απλό τόνο της μορφής $s(t) = A \sin(2\pi f_0 t)$. Στο σχήμα 3.10a έχουμε ένα φίλτρο με ΧΣ ίση με f_0 . Φαίνεται ότι αν αυξήσουμε την ένταση του ερεθίσματος τότε περισσότερα επίπεδα στους ανιχνευτές θα ενεργοποιούνται και επομένως και η έξοδος τελικά θα είναι μεγαλύτερη. Στο σχήμα 3.10b βλέπουμε την απόκριση από διάφορα γειτονικά φίλτρα όπου φαίνεται ότι δεν αποκρίνεται μόνο ένα φίλτρο με την κατάλληλη ΧΣ στον απλό τόνο αλλά και μερικά γειτονικά του, αν και με μικρότερη ένταση. Λόγω της μορφής των φίλτρων, πιο πιθανό είναι να αποκριθούν φίλτρα που έχουν ΧΣ μεγαλύτερη από τη συχνότητα του τόνου. Επίσης, μία ισχυρή συνιστώσα f_1 μπορεί να κυριαρχεί στην απόκριση μερικών φίλτρων με μεγαλύτερη ΧΣ όπως θα μπορούσε να συμβαίνει στην περίπτωση κάποιας μορφικής. Ένα άλλο χαρακτηριστικό της αναπαράστασης είναι ότι το χρονικό παράθυρο που γίνεται η καταγραφή δεν έχει την ίδια διάρκεια για όλα τα φίλτρα. Εφόσον χρησιμοποιούνται τα τελευταία 20 παράθυρα, τα φίλτρα με υψηλές ΧΣ θα έχουν συντομότερα παράθυρα και καλή χρονική διακριτική ικανότητα ενώ αυτά με χαμηλές ΧΣ καλή συχνοτική διακριτική ικανότητα. Ένα παράδειγμα απόκρισης σε φωνήεν, σε σύγκριση με το συνηθισμένο φάσμα, παρουσιάζεται στο σχήμα 3.11. Το χαρακτηριστικό του ΕΙΗ είναι ότι παρέχει καλή ανάλυση για την περιβάλλουσα και τις λεπτομέρειες των χαμηλών μορφικών αλλά κυρίως για την περιβάλλουσα σε υψηλότερα μορφικές. Ακόμα, η συγκεκριμένη αναπαράσταση, σε πειράματα αναγνώρισης λέξεων, επέδειξε ανοχή στην παρουσία λευκού θορύβου σε σχέση με το απλό φάσμα.

3.2.1.2 Κοχλιογράμμα και συσχετισιόγραμμα

Μία άλλη αναπαράσταση που επικεντρώνεται και αυτή στα χρονικά χαρακτηριστικά του σήματος είναι αυτή του κοχλιογράμματος (cochleagram) και, σε ένα επόμενο βήμα, του συσχετισιογράμματος (correlogram), όπως φαίνεται στο σχήμα 3.12 ([23]). Οι συγκεκρι-



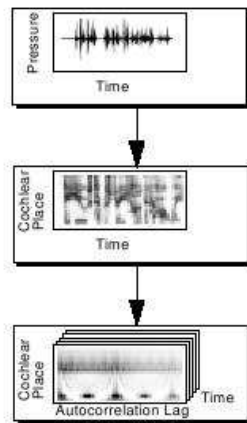
Σχήμα 3.10: Διαδικασία υπολογισμού του EIH, [6]



Σχήμα 3.11: Σύγκριση EIH και απλού φάσματος για ένα φωνήεν, [6]

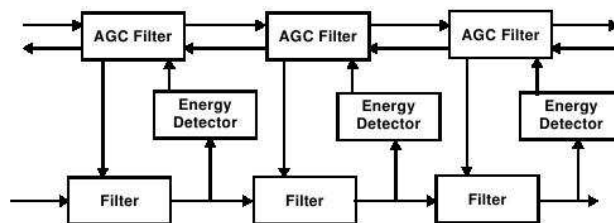
Κεφάλαιο 3. Υπολογιστική και μαθηματική μοντελοποίηση του ακουστικού συστήματος

μένες αναπαραστάσεις βασίζονται σε πιο λεπτομερή, μη γραμμικά μοντέλα του κοχλία. Στο κοχλιόγραμμα, το αρχικό μονοδιάστατο ηχητικό σχήμα μετασχηματίζεται σε έναν διδιάστατο χάρτη που απεικονίζει τον ρυθμό εκφόρτισης ως συνάρτηση του χρόνου και της θέσης στη βασική μεμβράνη. Το συσχετισιόγραμμα εισάγει μία επιπλέον διάσταση, αυτή της περιοδικότητας.



Σχήμα 3.12: Κοχλιόγραμμα και συσχετισιόγραμμα, [23]

Ένα μπλοκ διάγραμμα του μοντέλου του κοχλία που χρησιμοποιείται για να παραχθεί το κοχλιόγραμμα δίνεται στο σχήμα 3.13. Η διάδοση των κυμάτων κατά μήκος της βασικής μεμβράνης προσομοιώνεται από μία σειρά φίλτρων. Τα εσωτερικά τριχοφόρα κύτταρα που μετατρέπουν τη μετατόπιση της βασικής μεμβράνης σε νευρωνικές εκφορτίσεις μοντελοποιούνται από μη γραμμικούς ανιχνευτές. Τέλος, υπάρχει κι ένας μηχανισμός αυτόματης ρύθμισης του κέρδους (automatic gain control) που συμπιέζει σήματα με μεγάλη μεταβλητότητα σε ένα πιο περιορισμένο δυναμικό εύρος. Σκοπός είναι να παραχθεί τελικά μία αναπαράσταση που θα μας δίνει το στιγμιαίο ρυθμό εκφόρτισης των ιών του ακουστικού νεύρου. Το κοχλιόγραμμα που προκύπτει μοιάζει με το φασματόγραμμα εφόσον έχει έναν άξονα του χρόνου και για κάθε χρονική στιγμή έχουμε τις στιγμιαίες αποκρίσεις μίας σειράς ιών με διαφορετικές ΧΣ.



Σχήμα 3.13: Σχηματικό διάγραμμα του μοντέλου του κοχλία, [23]

Σύμφωνα με μία προσέγγιση (την WKB), η μετάδοση ενός κύματος από μία περιοχή x_1 της βασικής μεμβράνης σε μία άλλη x_2 περιλαμβάνει ένα πολλαπλασιασμό του εισερχόμενου

κύματος με την ποσότητα :

$$\exp(i \int_{x_1}^{x_2} k(\omega, x) dx)$$

Η διάδοση του κύματος κατ' αυτόν τον τρόπο μπορεί να μοντελοποιηθεί ως μία σειρά φίλτρων με συνάρτηση μεταφοράς :

$$H_i(\omega, x) = e^{ik(\omega, x_i) dx}$$

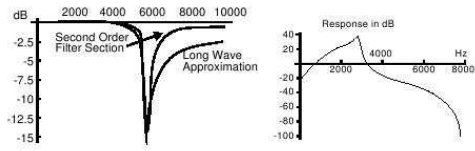
για κάθε τμήμα dx της βασικής μεμβράνης στη θέση x_i . Η γραφική παράσταση μίας ενδεικτικής τέτοιας συνάρτησης φαίνεται στο σχήμα 3.14(α'). Η συχνότητα συντονισμού της μεταβάλλεται από 20 kHz στη βάση της μεμβράνης μέχρι 20 Hz στην κορυφή της. Η μετατροπή της πίεσης σε μετατόπιση της βασικής μεμβράνης επιτυγχάνεται με μία επιπλέον συστοιχία ζωνοπερατών φίλτρων όπως αυτό του σχήματος 3.14(β'). Τα εσωτερικά τριχοφόρα κύτταρα και οι νευρώνες του ακουστικού νεύρου μετατρέπουν τη μηχανική κίνηση της βασικής μεμβράνης σε νευρωνικές εκφορτίσεις με τρόπο που μπορεί να μοντελοποιηθεί ως ένας ανορθωτής ημίσεος κύματος (half wave rectifier) ή με μία συνάρτηση κορεσμού :

$$\frac{1}{2}(1 + \tanh(x + a))$$

Τα μοντέλα των τριχοφόρων κυττάρων λειτουργούν ως ανιχνευτές, υπό την έννοια ότι διατηρούν στην έξοδό τους τόσο την πληροφορία για τη μέση ενέργεια του σήματος, μέσω της ανόρθωσης, όσο και τη χρονική πληροφορία. Όπως έχει περιγραφεί στο προηγούμενο κεφάλαιο, το ακουστικό σύστημα διαθέτει και κατερχόμενες οδούς ανάδρασης που νευρώνουν τον κοχλία και κυρίως τα εξωτερικά τριχοφόρα κύτταρα. Αυτή η ανάδραση προσομοιώνεται με ένα μοντέλο αυτόματης ρύθμισης του κέρδους το οποίο επιτρέπει την έντονη απόκριση στην αρχή ενός ερεθίσματος αλλά τη μείωσή της, λόγω προσαρμογής (adaptation), καθώς περνάει ο χρόνος και το ερέθισμα επιμένει. Η προσαρμογή αυτή μπορεί να επηρεάζει ακόμα και τις μηχανικές ιδιότητες της βασικής μεμβράνης μεταβάλλοντας την εξίσωση διάδοσης. Ένα παράδειγμα κοχλιογράμματος σε σχέση με το αντίστοιχο φασματόγραμμα φαίνεται στο σχήμα 3.15. Στην εικόνα 3.15(α') έχουμε φασματόγραμμα στενής (narrowband) και ευρείας (wideband) ζώνης και στην 3.15(β') το κοχλιόγραμμα από δύο ελαφρώς διαφορετικά μοντέλα με το κάτω να είναι κάπως πιο σύγχρονο. Παρατηρείται σημαντική ομοιότητα μεταξύ φασματογράμματος και κοχλιογράμματος με το κοχλιόγραμμα να είναι ελαφρώς πιο ευαίσθητο στις ενάρξεις. Αν κοιτάξουμε σε μικρότερη χρονική κλίμακα, όπως στην εικόνα 3.16, όπου φαίνεται το κοχλιόγραμμα του "rea" από το "grease", βλέπουμε ότι, εκτός από τις μορφικές, είναι εμφανείς και οι γλωττικοί παλμοί καθώς και οι κορυφές της κυματομορφής, χρονική πληροφορία που μπορεί να είναι πολύ χρήσιμη στην αποκωδικοποίηση του σήματος.

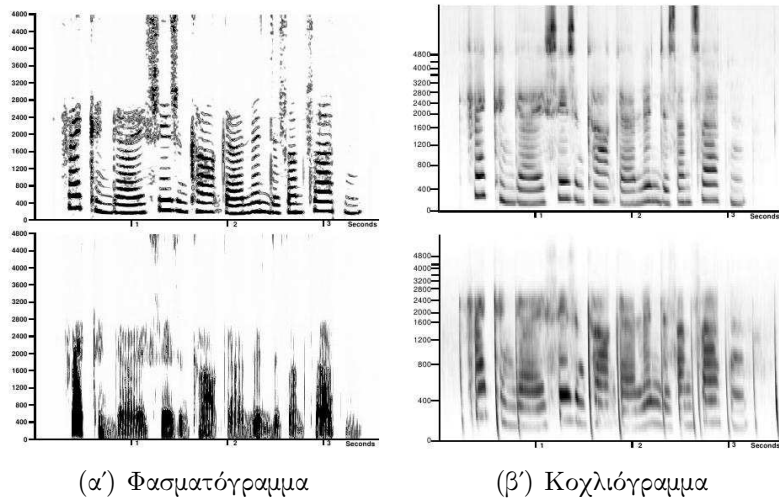
Το επόμενο βήμα είναι ο υπολογισμός του συσχετισιογράμματος το οποίο μας δίνει ακόμα περισσότερες πληροφορίες για τα χρονικά χαρακτηριστικά του σήματος και τις τυχόν περιοδικότητες του. Αυτό επιτυγχάνεται εισάγοντας μία ακόμα διάσταση, αυτή της περιοδικότητας. Η πράξη που χρησιμοποιείται για να μετρηθεί η περιοδικότητα είναι η αυτοσυσχέτιση. Για κάθε θέση στον κοχλία και για κάθε χρονική στιγμή υπολογίζεται η αυτοσυσχέτιση του ρυθμού εκφόρτισης για διάφορες τιμές της καθυστέρησης (time-lag). Η οπτικοποίηση των αποτελεσμάτων γίνεται υπό τη μορφή μίας σειράς διαδοχικών πλαισίων (video) όπου σε κάθε χρονική στιγμή έχουμε μία εικόνα με τον κάθετο άξονα να

Κεφάλαιο 3. Υπολογιστική και μαθηματική μοντελοποίηση του ακουστικού συστήματος



(α') Ενδεικτική συνάρτηση μεταφοράς φίλτρου μετάδοσης επί της βασικής μεμβράνης
 (β') Ενδεικτική συνάρτηση μεταφοράς φίλτρου μετατροπής επί της βασικής μεμβράνης

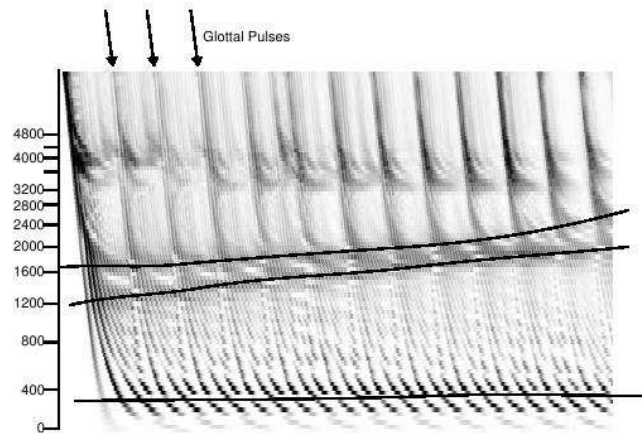
Σχήμα 3.14: Ενδεικτικές συναρτήσεις μεταφοράς φίλτρου μετάδοσης και φίλτρου μετατροπής στη βασική μεμβράνη, [23]



(α) Φασματογράμμα

(β) Κοχλιογράμμα

Σχήμα 3.15: Σύγκριση φασματογράμματος και κοχλιογράμματος, [23]



Σχήμα 3.16: Κοχλιόγραμμα του “tea” από το “grease”, [23]

αντιστοιχεί σε συχνότητα (ή θέση στον κοχλία), τον οριζόντιο στην χρονική καθυστέρηση της αυτοσυσχέτισης και η ένταση σε κάθε σημείο της εικόνας αναπαρίσταται σε διαβαθμίσεις του γκρι.

Το σύστημα που εκτελεί τους παραπάνω υπολογισμούς φαίνεται στο σχήμα 3.17. Το ηχητικό σήμα εισέρχεται στον κοχλία που το αναλύει σε συχνотικά κανάλια και στη συνέχεια υπολογίζονται οι αυτοσυσχετίσεις με το τελικό αποτέλεσμα να τροφοδοτεί τα επόμενα, υψηλότερα κέντρα επεξεργασίας. Η αυτόσυσχέτιση ενός σήματος $x(t)$ δίνεται από τον τύπο

$$R_{xx}(\tau) = \int_{-\infty}^{+\infty} x(t)x(t - \tau)dt$$

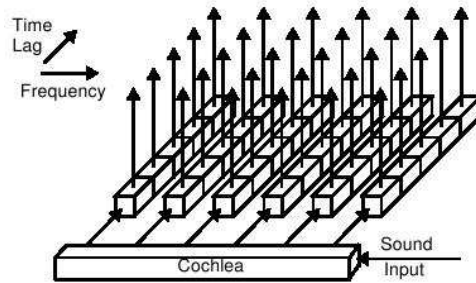
Στην περίπτωση του συσχετισιογράμματος η συσχέτιση υπολογίζεται εντός χρονικών παραθύρων $w(t)$ περί την κάθε χρονική στιγμή :

$$C_f(\tau, t) = \int_0^{+\infty} x_f(t - s)x_f(t - s - \tau)w(s)ds$$

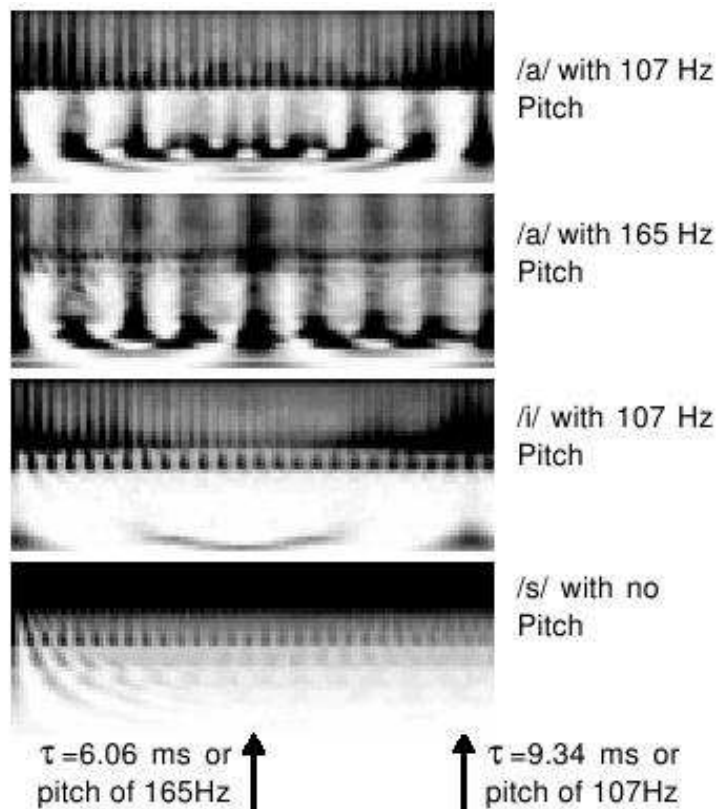
όπου $x_f(t)$ είναι ο ρυθμός εκφόρτισης στη θέση της βασικής μεμβράνης με ΧΣ ίση με f .

Παραδείγματα συσχετισιογραμμάτων παρουσιάζονται στο σχήμα 3.18. Σε ένα συσχετισιογράμμα, οι κάθετες, επαναλαμβανόμενες μπάρες υποδεικνύουν συγχρονισμένη δραστηριότητα πολλών καναλιών με την ίδια περιοδικότητα T , κάτι που αποτελεί ένδειξη για ύψος (pitch) με συχνότητα $\frac{1}{T}$. Υψηλότερο ύψος έχει ως αποτέλεσμα οι κάθετες μπάρες να μετακινούνται προς τα αριστερά και να έρχονται πιο κοντά η μία στην άλλη (βλ. δύο ο πρώτα συσχετισιογράμματα). Οριζόντιες μπάρες υποδεικνύουν έντονη δραστηριότητα σε συγκεκριμένες ζώνες συχνοτήτων, ένδειξη για την ύπαρξη μορφικών. Μεταβολή των μορφικών θα έχει ως αποτέλεσμα την μετακίνηση και των οριζόντιων μπαρών στο συσχετισιογράμμα (βλ. πρώτο και τρίτο συσχετισιογράμματα). Για τους άφωνους ήχους που δεν παρουσιάζουν περιοδικότητες το συσχετισιογράμμα δεν προσφέρει πληροφορίες (βλ. τέταρτο συσχετισιογράμμα).

Κεφάλαιο 3. Υπολογιστική και μαθηματική μοντελοποίηση του ακουστικού συστήματος



Σχήμα 3.17: Σύστημα υπολογισμού συσχετισιογράμματος, [23]



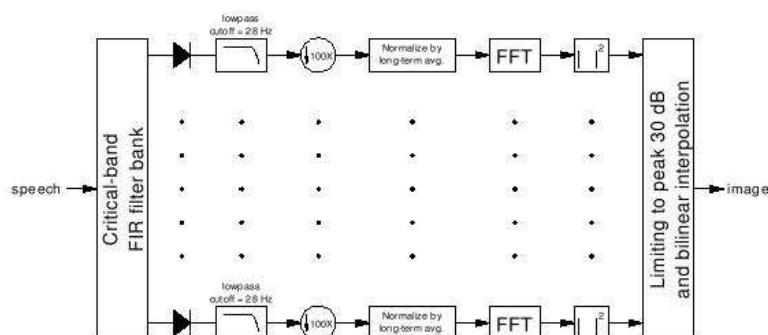
Σχήμα 3.18: Παραδείγματα συσχετισιογραμμάτων, [23]

3.2.2 Μοντελοποίηση του κεντρικού ακουστικού συστήματος

Όπως έχει αναφερθεί επανειλημμένως, η μελέτη του κεντρικού ακουστικού συστήματος βρίσκεται ακόμα σε πολύ πρώιμο στάδιο. Ως αποτέλεσμα, είναι πολύ λίγες οι προσπάθειες μοντελοποίησής του. Σε αυτή την ενότητα θα εστιάσουμε σε μία από τις πιο ολοκληρωμένες τέτοιες προσπάθειες, στο μοντέλο που έχει προταθεί από τον Shamma και τους συνεργάτες του ([2]).

3.2.2.1 Το φασματογράμμα διαμόρφωσης

Μία απλούστερη αναπαράσταση αφορά στο φασματογράμμα διαμόρφωσης ([8]). Όπως θα είναι προφανές και από την ονομασία της, σε αυτή την αναπαράσταση δίνεται έμφαση στις αργές μεταβολές κάποιων γενικών χαρακτηριστικών του φάσματος της ομιλίας που αφορούν κυρίως σε φαινόμενα διαμόρφωσης. Θεωρείται ότι οι ταχείες μεταβολές και οι λεπτομέρειες του φάσματος δεν προσφέρουν σημαντική πληροφορία για την κατανόηση του σήματος ομιλίας εφόσον εξαρτώνται σε μεγάλο βαθμό από τον ομιλητή και το περιβάλλον. Αντιθέτως, ψυχοακουστικά πειράματα (κεφ. 3, [7]) έχουν δείξει ότι οι αργές διαμόρφωσης της περιβάλλουσας του σήματος, στην περιοχή των 4-16 Hz, περιέχουν ένα σημαντικό ποσοστό αυτής της πληροφορίας. Νευροφυσιολογικά πειράματα (βλ. κεφάλαιο 1) επίσης έχουν εντοπίσει νευρώνες στον ΑΙ που είναι ευαίσθητοι σε διαμορφώσεις κάτω των 20 Hz. Το μοντέλο που προτείνεται εδώ παρέχει μία αναπαράσταση ακριβώς αυτών των συχνοτήτων διαμόρφωσης που υπάρχουν στο σήμα.



Σχήμα 3.19: Μοντέλο υπολογισμού φασματογράμματος διαμόρφωσης, [8]

Η διαδικασία υπολογισμού του φασματογράμματος διαμόρφωσης φαίνεται στο σχήμα 3.19. Μία συστοιχία FIR φίλτρων αναλύει το σήμα εισόδου σε κανάλια που ακολουθούν την κλίμακα κρίσιμων ζωνών. Μέσω ανόρθωσης ημίσεος κύματος και βαθυπερατού φιλτραρίσματος, με συχνότητα αποκοπής στα 28 Hz, προκύπτει η περιβάλλουσα. Κατόπιν ακολουθεί downsampling κατά ένα παράγοντα 100 και κανονικοποίηση με βάση τη μέση τιμή της περιβάλλουσας. Τέλος, οι διαμορφώσεις προκύπτουν μέσω ενός FFT και τετραγωνισμό των συντελεστών του FFT για τα 4 Hz. Ένα παράδειγμα φασματογράμματος διαμόρφωσης σε σύγκριση με κανονικό φασματογράμμα, για καθαρή και θορυβώδη ομιλία, φαίνεται στο σχήμα 3.20. Τα κύρια χαρακτηριστικά του φασματογράμματος διαμόρφωσης

φαίνεται να παρουσιάζουν μία σταθερότητα μεταξύ καθαρής και θορυβώδους ομιλίας ενώ αυτά του κανονικού φασματογράμματος χάνονται μέσα στον θόρυβο.

3.2.2.2 Το μοντέλο του Shamma

Πρόκειται για μία πιο ολοκληρωμένη πρόταση που επιχειρεί να μοντελοποιήσει διάφορα στάδια της επεξεργασίας των ακουστικών σημάτων από το νευρικό σύστημα ([2]). Αποτελείται από δύο στάδια, ένα πρώιμο ακουστικό στάδιο και το φλοιικό στάδιο. Το πρώιμο στάδιο έχει αρκετές ομοιότητες με μερικά από τα μοντέλα που έχουν παρουσιαστεί προηγουμένως. Το φλοιικό στάδιο προέκυψε κατόπιν εκτεταμένων νευροφυσιολογικών πειραμάτων στον ακουστικό φλοιό κουναβιών. Δύο φαινόμενα στα οποία επικεντρώνεται ιδιαίτερα και προσπαθεί να ενσωματώσει αυτό το μοντέλο είναι η προοδευτική μείωση της ικανότητας των νευρώνων για κλειδώμα φάσης, καθώς προχωράμε από το ακουστικό νεύρο στον φλοιό, και οι ευαισθησίες του ΑΙ σε περίπλοκα ερεθίσματα, τα λεγόμενα φασματικά κύματα (βλ. κεφάλαιο 2).

Το πρώιμο στάδιο παρέχει μία αναπαράσταση των μετασχηματισμών που υφίσταται το ακουστικό σήμα από τον κοχλία μέχρι τον μεσεγκέφαλο. Όπως και στα προηγούμενα μοντέλα, έχουμε μία συστοιχία 128 αλληλοεπικαλυπτόμενων, ζωνοπερατών φίλτρων σταθερού Q. Οι κεντρικές συχνότητες των φίλτρων κατανέμονται λογαριθμικό σε ένα εύρος 5.3 οκτάβων. Αν η χροστική απόκριση των φίλτρων είναι $h(t, x)$ τότε η έξοδος των κοχλιακών φίλτρων θα είναι :

$$y_{coch}(t, x) = h(t, x) \otimes s(t)$$

Στη συνέχεια, η έξοδος $y_{coch}(t, x)$ μετατρέπεται σε δραστηριότητα του ΑΝ περνώντας από το στάδιο των εσωτερικών τριχοφόρων κυττάρων. Εκεί, επιτελείται αρχικά μία παραγωγή στο πεδίο του χρόνου μέσω υψιπερατού φιλτραρίσματος και κατόπιν μία μη γραμμική συμπίεση και ένα επιπλέον βαθυπερατό φιλτράρισμα ώστε να προσομοιωθεί η μείωση της ικανότητας κλειδώματος φάσης. Το σήμα που προκύπτει θα είναι :

$$y_{AN}(t, x) = g(\vartheta_t y_{coch}(t, x)) \otimes w(t)$$

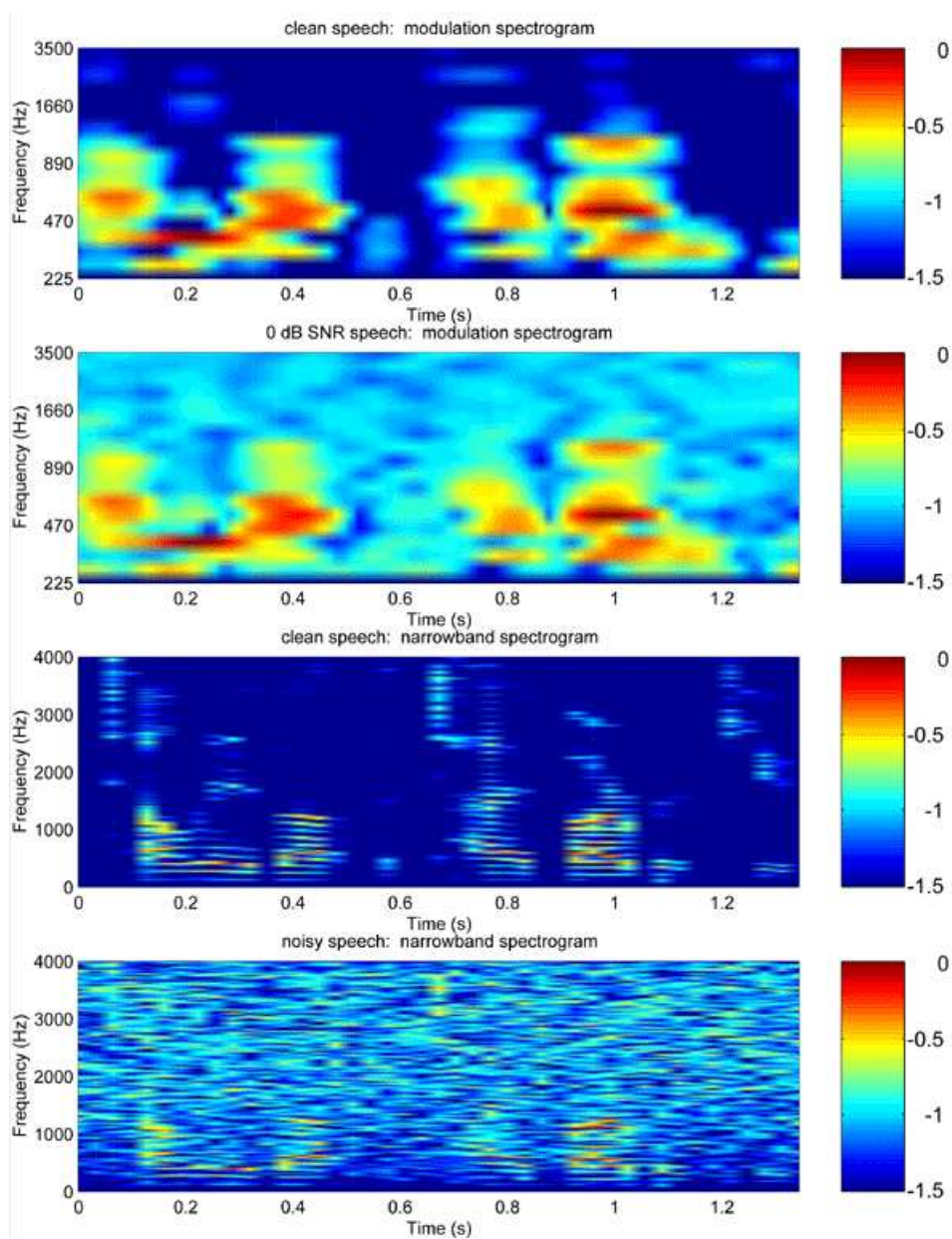
Το επόμενο στάδιο είναι ένα LIN (lateral inhibitory network) δίκτυο που θεωρείται ότι υπάρχει στον κοχλιακό πυρήνα και επιτελεί παραγωγή στο πεδίο της συχνότητας και ανόρθωση ημίσεος κύματος.

$$y_{LIN}(t, x) = \max(\vartheta_x y_{AN}(t, x), 0)$$

Τέλος, εισάγεται και ένας παράγοντας ολοκλήρωσης $\mu(t, \tau) = e^{-\frac{t}{\tau}} u(t)$ με σταθερά χρόνου $\tau = 8 \text{ ms}$ που προσομοιώνει την περαιτέρω απώλεια κλειδώματος φάσης στον μεσεγκέφαλο.

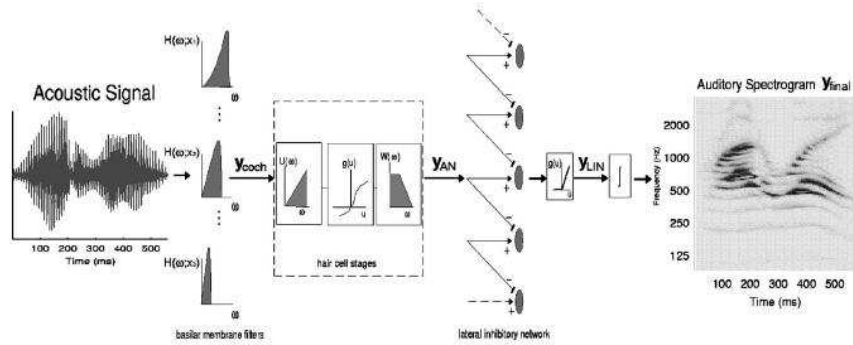
$$y_{final}(t, x) = y_{LIN}(t, x) \otimes_t \mu(t, \tau)$$

Μία σχηματική αναπαράσταση του πρώιμου σταδίου φαίνεται στο σχήμα 3.21. Παραδείγματα της αναπαράστασης που προκύπτει από το πρώιμο στάδιο παρουσιάζονται στο σχήμα 3.22. Στο σχήμα 3.22(α') έχουμε ως είσοδο ένα σήμα τριών απλών τόνων και βλέπουμε την έξοδο του LIN δικτύου και την τελική έξοδο. Είναι εμφανές ότι στην τελική έξοδο

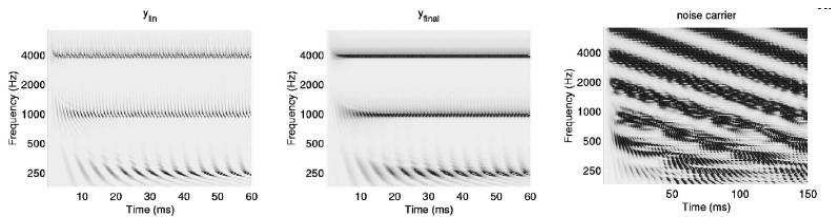


Σχήμα 3.20: Σύγκριση φασματογράμματος διαμόρφωσης με κανονικό φασματόγραμμα, [8]

Κεφάλαιο 3. Υπολογιστική και μαθηματική μοντελοποίηση του ακουστικού συστήματος



Σχήμα 3.21: Σχηματικό διάγραμμα του μοντέλου του Shamma, [2]



(α') Η έξοδος του LIN και η τελική έξοδος για (β') Τελική έξοδος για σήμα τριών τόνων φασματικό κύμα

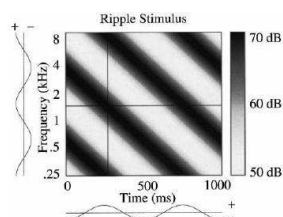
Σχήμα 3.22: Παραδείγματα εξόδων του πρώιμου σταδίου του μοντέλου του Shamma, [2]

παρατηρείται κλείδωμα φάσης μόνο για τον χαμηλότερο τόνο ενώ στους άλλους δύο διατηρείται η περιβάλλουσα μόνο. Παρομοίως, στο φασματικό κύμα, στις υψηλές συχνότητες, διατηρούνται μόνο οι αργές διαμορφώσεις του θορύβου ενώ στις συχνότητες κάτω των 1000 Hz έχουμε και τις ταχύτερες αποκρίσεις στους τόνους του ερεθίσματος.

Το φλοιικό στάδιο προσομοιώνει τις αποκρίσεις των νευρώνων σε ανώτερα επίπεδα και κυρίως στον ΑΙ. Αυτοί οι νευρώνες επιδεικνύουν μία αρκετά πολύπλοκη συμπεριφορά, επιλεκτική σε διάφορες παραμέτρους του ηχητικού σήματος, τόσο χρονικές όσο και συχνοτικές. Η υπόθεση που γίνεται εδώ είναι ότι η επεξεργασία του σήματος γίνεται, τουλάχιστον σε υπολογιστικό επίπεδο, και πάλι μέσω μίας συστοιχίας φίλτρων καθένα από τα οποία είναι συντονισμένο σε συγκεκριμένα εύρη τιμών των παραμέτρων που αναφέρθηκαν πριν. Η ευαισθησία των φίλτρων σε αυτές τις παραμέτρους αποδίδεται μαθηματικά με την έννοια του χρονοσυχνοτικού πεδίου απόκρισης (spectrotemporal response field, STRF), κατά κάποιο τρόπο το ανάλογο της χροστικής απόκρισης στα συνήθη γραμμικά συστήματα ενώ ταυτόχρονα φαίνεται να είναι και το αντίστοιχο των υποδεκτικών πεδίων του οπτικού συστήματος. Κάθε κύτταρο χαρακτηρίζεται από το STRF του και η έξοδός του σε ένα ερέθισμα είναι η συνέλιξη αυτού του STRF με το ερέθισμα. Στη συνέχεια επεξηγείται πώς προκύπτει η έννοια του STRF.

Έστω ένα ερέθισμα της μορφής (βλ. σχήμα 3.23):

$$S(t, x) = 1 + A \sin(2\pi(\omega t + \Omega x) + \Phi)$$

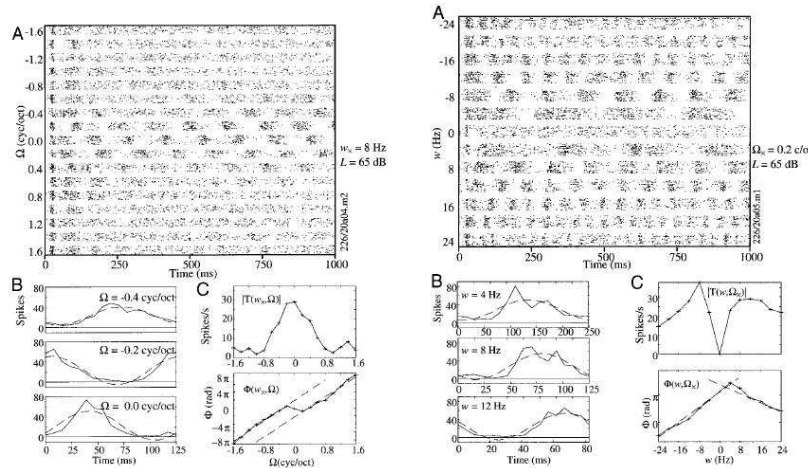


Σχήμα 3.23: Φασματικό κύμα, [4]

Παρουσιάζοντας τέτοια ερεθίσματα για διάφορες τιμές του Ω και του ω μπορούμε να λάβουμε δύο συναρτήσεις μεταφοράς. Αν διατηρήσουμε σταθερό το Ω , δηλαδή την πυκνότητα των κυμάτων κατά τον άξονα της συχνότητας, και μεταβάλλουμε την ταχύτητα κίνησής - επανάληψής τους ω , οι αποκρίσεις του νευρώνα μάς δίνουν τη χρονική συνάρτηση μεταφοράς $T(\omega)$ (temporal transfer function). Διατηρώντας το ω σταθερό και μεταβάλλοντας το Ω λαμβάνουμε την κυματική συνάρτηση μεταφοράς $T(\Omega)$ (ripple transfer function). Η διαδικασία εξαγωγής των συναρτήσεων μεταφοράς $T(\Omega)$ και $T(\omega)$ φαίνεται στο σχήμα 3.24. Αν εφαρμόσουμε έναν αντίστροφο μετασχηματισμό Fourier σε κάθε μία από τις $T(\Omega)$ και $T(\omega)$ προκύπτουν το πεδίο απόκρισης $RF(x)$ και η χρονική χροστική απόκριση $IR(t)$ αντίστοιχα (βλ. σχήμα 3.25). Οι αποκρίσεις RF και IR μπορούν να έχουν μία ποικιλία σχημάτων με διάφορα χαρακτηριστικά (π.χ. πλάτος ή ασυμμετρίες). Για κάθε RF υπάρχουν νευρώνες με διαφορετικό IR που επεξεργάζονται τα δυναμικά χαρακτηριστικά του φάσματος σε διάφορες χρονικές κλίμακες. Αντίστοιχα, μέσω των διαφορετικών RF έχουμε ανάλυση και σε πολλαπλά επίπεδα συχνότητας και εύρους ζώνης. Επομένως, έχουμε

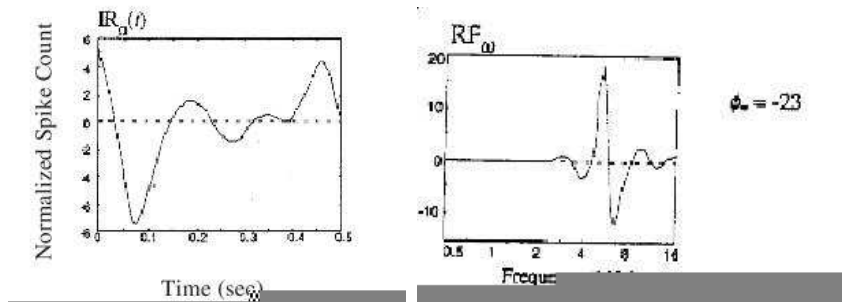
Κεφάλαιο 3. Υπολογιστική και μαθηματική μοντελοποίηση του ακουστικού συστήματος

μία αναπαράσταση σε πολλαπλές κλίμακες, τόσο στον άξονα της συχνότητας όσο και σε αυτόν του χρόνου (βλ. και παρ. 2.2.2.4).



(α) Εξαγωγή της $T(\Omega)$ από τις νευρωνικές αποκρίσεις σε διάφορες τιμές του Ω
 (β) Εξαγωγή της $T(\omega)$ από τις νευρωνικές αποκρίσεις σε διάφορες τιμές του ω

Σχήμα 3.24: Διαδικασία εξαγωγής συναρτήσεων μεταφοράς $T(\Omega)$ και $T(\omega)$, [4]

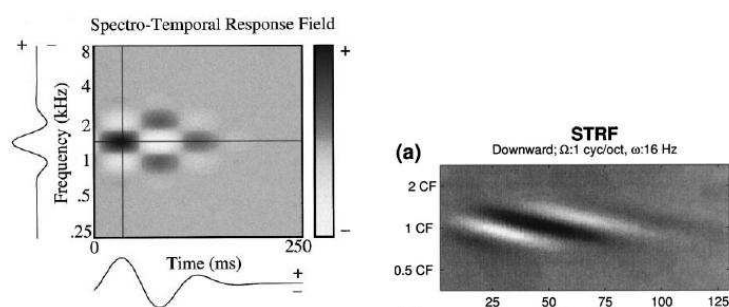


(α) Χρονική κρουστική απόκριση (β) Πεδίο απόκρισης

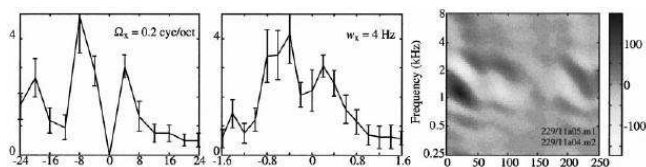
Σχήμα 3.25: Παραδείγματα RF και IR αποκρίσεων, [18]

Αν πάρουμε τον διδιάστατο αντίστροφο μετασχηματισμό Fourier της γενικότερης συνάρτησης μεταφοράς $T(\omega, \Omega)$ προκύπτει τελικά το STRF του νευρώνα που μελετάται. Παραδείγματα από STRF φαίνονται στο σχήμα 3.26. Το STRF θα μπορούσαμε να πούμε ότι περιγράφει πλήρως ένα νευρώνα υπό την έννοια ότι αποτελεί ένα είδος κρουστικής απόκρισης του όπως στα γραμμικά συστήματα. Η απόκριση ενός νευρώνα σε ένα ερέθισμα θα είναι η συνέλιξη του STRF του και του ερεθίσματος. Η απόκριση αυτή θα είναι μέγιστη αν το ερέθισμα διαθέτει χαρακτηριστικά τα οποία είναι συντονισμένο να συλλαμβάνει το

STRF του νευρώνα. Κάθε STRF αναπτύσσεται γύρω από μία κεντρική συχνότητα και εμφανίζει προτιμήσεις στους ρυθμούς διαμόρφωσης τόσο στον άξονα της συχνότητας όσο και σε αυτόν του χρόνου. Επίσης, μπορεί να συντονίζεται βέλτιστα σε συγκεκριμένη κατεύθυνση. Για παράδειγμα, στο σχήμα 3.26(β'), ο νευρώνας προτιμάει ερεθίσματα με καθοδική κατεύθυνση και με χρονικές τιμές διαμόρφωσης γύρω στα 16 Hz και συχνοτικές στο 1 cyc/oct. Ένα οποιοδήποτε φάσμα με συγκεκριμένα χρονοσυχνοτικά χαρακτηριστικά μπορεί να συντεθεί ως άθροισμα επιμέρους δυναμικών φασματικών κυμάτων που θα αποτελούν και τις συνιστώσες του. Αν οι νευρώνες ικανοποιούν την αρχή της υπέρθεσης τότε η απόκρισή τους θα είναι απλώς το άθροισμα των αποκρίσεων σε κάθε συνιστώσα ξεχωριστά. Πειράματα που έγιναν για να διαπιστωθεί αν ικανοποιείται η αρχή της υπέρθεσης έδωσαν μερικά υποστηρικτικά αποτελέσματα.



(α') Εξιδανικευμένο STRF, [4] (β') Μαθηματικό, καθοδικό STRF με $\Omega=1$ cyc/oct και $\omega=16$ Hz, [2]



(γ') Πραγματικό STRF με τις αντίστοιχες $T(\Omega)$ και $T(\omega)$, [4]

Σχήμα 3.26: Παραδείγματα STRF

Στο μοντέλο του Shamma, τα πραγματικά STRF, όπως αυτό του σχήματος 3.26(γ'), αντικαθίστανται από μαθηματικά υπολογισμένα STRF. Συγκεκριμένα, ένα STRF υπολογίζεται ως εξής :

$$STRF = \Re\{h_{IRT}(t) \cdot h_{IRS}(x)\}$$

όπου

$$h_{IRS}(x; \Omega, \phi) = h_{irs}(x; \Omega, \phi) + j\hat{h}_{irs}(x; \Omega, \phi)$$

$$h_{IRT}(t; \omega, \theta) = h_{irt}(t; \omega, \theta) + j\hat{h}_{irt}(t; \omega, \theta)$$

όπου $\hat{h}(\cdot)$ δηλώνει μετασχηματισμό Hilbert και οι $h_{irs}(\cdot)$ και $h_{irt}(\cdot)$ προέρχονται από ημιτονική παρεμβολή συναρτήσεων Gabor $h_s(\cdot)$ και $h_t(\cdot)$ αντίστοιχα. Οι $h_s(\cdot)$ και $h_t(\cdot)$ αντιπροσωπεύουν φασματικές και χρονικές συναρτήσεις που καθορίζουν την ευαισθησία

Κεφάλαιο 3. Υπολογιστική και μαθηματική μοντελοποίηση του ακουστικού συστήματος

του STRF στους αντίστοιχους άξονες. Η έξοδος του νευρώνα σε ένα ερέθισμα $y(t, x)$ δίνεται από τη συνέλιξη :

$$r_c \Downarrow (\Uparrow)(t, x; \omega_c, \Omega_c, \theta_c, \Phi_c) = y(t, x) \otimes_{tx} \Re\{h_{IRT}^{(*)}(t; \omega_c, \theta_c) \cdot h_{IRS}(x; \Omega_c, \phi_c)\}$$

όπου \Downarrow και \Uparrow δηλώνουν την προτίμηση κατεύθυνσης και $*$ δηλώνει τον συζυγή. Αυτή η απόκριση ονομάζεται φλοιική αναπαράσταση.

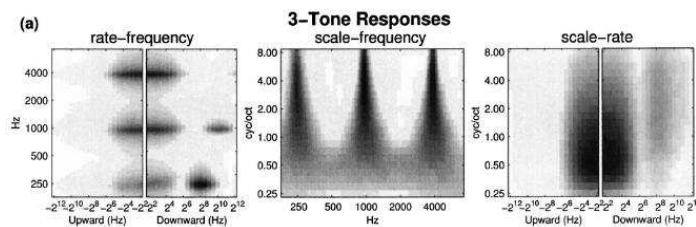
Για την οπτική αναπαράσταση της απόκρισης, χρησιμοποιούνται τα μεγέθη

$$|z \Downarrow (t, x; \omega_c, \Omega_c)|$$

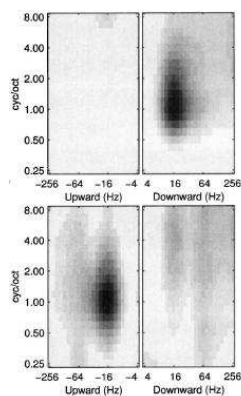
και

$$|z \Uparrow (t, x; \omega_c, \Omega_c)|$$

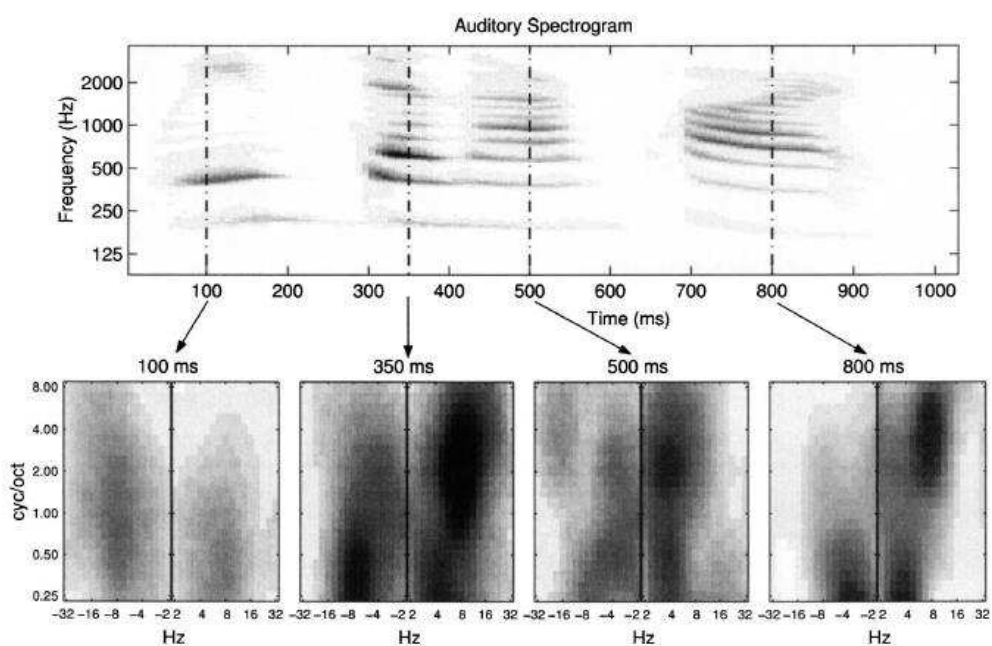
που αντιστοιχούν στη μέγιστη απόκριση στην καθοδική και ανοδική κατεύθυνση αντίστοιχα (για περισσότερες λεπτομέρειες βλ. [2]). Για να μειωθούν οι διαστάσεις, αρχικά αφαιρείται η εξάρτηση από το χρόνο μέσω ολοκλήρωσης της απόκρισης στο χρόνο, αν το σήμα είναι χρονικά στατικό, ή για ένα παράθυρο χρόνου αν το σήμα μεταβάλλεται οπότε και προκύπτει μία ακολουθία πλαισίων. Από το προκύπτον σήμα μπορούν να εξαχθούν τρεις διαφορετικές αναπαραστάσεις, ανάλογα με το ποια από τις τρεις υπόλοιπες μεταβλητές θα επιλέξουμε να αφαιρέσουμε, πάλι μέσω ολοκλήρωσης. Αυτές οι τρεις αναπαραστάσεις φαίνονται στο σχήμα 3.27 για το σήμα τριών τόνων. Στην αριστερή αναπαράσταση έχει αφαιρεθεί η εξάρτηση από το Ω και είναι εμφανές το κλειδώμα φάσης για $x = 250$ Hz και $\omega = 250$ Hz όπως και για $x = 1000$ Hz και $\omega = 1000$ Hz. Στο μεσαίο σχήμα έχει αφαιρεθεί η εξάρτηση από την συχνότητα ω . Παρατηρούμε ότι για μικρό εύρος Ω οι τρεις τόνοι είναι δύσκολο ως αδύνατο να διαχωριστούν αλλά γίνονται ευδιάκριτοι καθώς το Ω αυξάνει. Τέλος, στο δεξί σχήμα έχει αφαιρεθεί η εξάρτηση από την κεντρική συχνότητα x , όπου και πάλι φαίνεται το κλειδώμα φάσης στα 250 Hz. Στο σχήμα 3.28 φαίνεται η scale-rate φλοιική αναπαράσταση για ένα καθοδικό και ένα ανοδικό φασματικό κύμα με $\omega = 16$ Hz και $\Omega = 1$ cyc/oct και για τα δύο. Είναι εμφανής η ασυμμετρία της απόκρισης στα κατάλληλα σημεία (ω, Ω), ανάλογα με την κατεύθυνση του κύματος. Τέλος, στο σχήμα 3.29 έχουμε μερικά πλαίσια από την φλοιική αναπαράσταση από ένα σήμα ομιλίας. Οι ασυμμετρίες στην απόκριση και πάλι συλλαμβάνουν ορισμένα χαρακτηριστικά του φασματογράμματος. Για παράδειγμα, η ανοδική κίνηση του μορφικής στο πλαίσιο των 100 ms αντικατοπτρίζεται στην εντονότερη ενεργοποίηση των ανοδικών φίλτρων.



Σχήμα 3.27: Φλοιική αναπαράσταση στο μοντέλο του Shamma για σήμα τριών τόνων, [2]



Σχήμα 3.28: Φλοιική αναπαράσταση στο μοντέλο του Shamma για καθοδικό και ανοδικό φασματικό κύμα, [2]



Σχήμα 3.29: Φλοιική αναπαράσταση στο μοντέλο του Shamma για σήμα ομιλίας, [2]

3.3 Ανασκόπηση

Η παρουσίαση των προηγούμενων μοντέλων εσχευμένα ακολούθησε μία πορεία από τα απλούστερα και συνήθως παλαιότερα προς τα πιο σύγχρονα και συνθετότερα. Αυτή η πορεία, που λίγο - πολύ ακολουθεί και την ιστορική εξέλιξη, αντανακλάει και τη σταδιακή μετατόπιση του ερευνητικού ενδιαφέροντος προς προσεγγίσεις που επιχειρούν να βασιστούν και να ενσωματώσουν στα μοντέλα τους νευροφυσιολογικά δεδομένα. Η πρώτη κατηγορία μοντέλων που παρουσιάστηκαν είναι αυτά που βασίζονται σε ψυχοακουστικά δεδομένα. Μέχρι στιγμής, οι απόπειρες μοντελοποίησης του ακουστικού συστήματος βασίζονται κυρίως σε τέτοιου είδους αναπαραστάσεις.

Η απλούστερη είναι αυτή που βασίζεται στους συντελεστές mfcc και περνάει το σήμα από μία συστοιχία φίλτρων, λογαριθμικά κατανεμημένων με βάση την ψυχοακουστική κλίμακα mel. Παρά την απλότητά του, το συγκεκριμένο μοντέλο επιδεικνύει αρκετά καλά ποσοστά επιτυχούς αναγνώρισης λέξεων σε συστήματα ASR, τουλάχιστον για απλές λέξεις από τον ίδιο ομιλητή και χωρίς θόρυβο. Η PLP τεχνική, με την προσομοίωση περισσότερων ψυχοακουστικών φαινομένων, όπως οι καμπύλες ευαισθησίας έντασης - συχνότητας και ο νόμος κυβικής ρίζας, επιτυγχάνει μία κανονικοποίηση του φασματογράμματος καθιστώντας το σχετικά ανεξάρτητο από τον ομιλητή. Επίσης, με τη χρήση της κλίμακας Bark, καταφέρνει να προσομοιώσει και φαινόμενα ηχοκάλυψης συγχωνεύοντας κορυφές στο φάσμα που βρίσκονται αρκούντως κοντά η μία στην άλλη. Το μοντέλο των Tchorz και Kollmeier έχει μία λογική παρόμοια με αυτή των προηγούμενων μοντέλων. Δανείζεται κάποιες ιδέες από αυτά (π.χ. η αρχική συστοιχία φίλτρων) αλλά εισάγει και ένα στάδιο προσαρμογής μέσω μίας σειράς προσαρμοστικών βρόχων συμπίεσης. Αυτοί οι βρόχοι ενισχύουν τις απότομες μεταβολές του φάσμογράμματος ενώ αποδυναμώνουν τις αργές διακυμάνσεις προσομοιώνοντας κατ' αυτόν τον τρόπο τα φαινόμενα προσαρμογής. Το συγκεκριμένο μοντέλο έχει καλή απόδοση σε περιπτώσεις όπου υπάρχει και θόρυβος εφόσον μπορεί και συμπίεζει το σχετικά σταθερό υπόβαθρο θορύβου σε ένα σήμα με αποτέλεσμα να επιτυγχάνει καλύτερα ποσοστά αναγνώρισης λέξεων σε σχέση με την mfcc αναπαράσταση.

Τα μοντέλα αυτής της κατηγορίας χρησιμοποιούνται ευρέως σε συστήματα ASR αλλά είναι εξαιρετικά αμφίβολο αν τελικά έχουν να προσφέρουν κάτι ουσιαστικό στην κατανόηση του τρόπου με τον οποίο ο άνθρωπος αντιλαμβάνεται τον ήχο και ειδικότερα την ομιλία. Το οπίσθιο τμήμα τους που αναλαμβάνει να επιτελέσει την αναγνώριση λέξεων με βάση τις ιδιότητες που προκύπτουν από το εμπρόσθιο τμήμα βασίζεται σε δομές (π.χ. τα κρυφά μαρκοβιανά μοντέλα) οι οποίες προέρχονται κυρίως από το χώρο της στατιστικής, των πιθανοτήτων και των μαθηματικών γενικότερα. Δεν λαμβάνουν υπόψιν τους τους πραγματικούς μηχανισμούς επεξεργασίας του εγκεφάλου. Επομένως, η όποια επιτυχία τους στο πεδίο της αυτόματης αναγνώρισης λέξεων δεν μπορεί να αποτελεί κριτήριο εγκυρότητας των εμπρόσθιων τμημάτων, τουλάχιστον όχι όσον αφορά στη μοντελοποίηση των πραγματικών μηχανισμών επεξεργασίας του λόγου. Από την άλλη, αυτού του είδους τα μοντέλα, στην προσπάθειά τους να προσομοιώσουν κάποια φαινόμενα, αυτό που κάνουν είναι απλά να εισάγουν κάποιους μαθηματικούς μετασχηματισμούς στην όλη αλυσίδα των σταδίων τους ώστε να επιτύχουν το ζητούμενο. Κατά κάποιον τρόπο, κάνουν λήψη του ζητούμενου χωρίς τα φαινόμενα να αναδύονται μέσα από τη λειτουργία δομών που όντως παρατηρούνται στον εγκέφαλο.

Αυτό το πρόβλημα αποφεύγεται με τη χρήση δεδομένων από νευροφυσιολογικά πειρά-

ματα με βάση τα οποία κατασκευάζονται τα μοντέλα της επόμενης κατηγορίας. Το EIH μοντέλο χρησιμοποιεί χρονική πληροφορία για να προσομοιώσει τα μοτίβα ενεργοποίησης των νευρικών ινών του ακουστικού νεύρου και να κατασκευάσει μία αναπαράσταση υπό τη μορφή φάσματος που διατηρεί μία καλή διακριτική ικανότητα στις χαμηλές συχνότητες αλλά είναι αρκετά πιο εξομαλυμένο στις υψηλές. Αυτό το μοντέλο χρησιμοποιήθηκε και πάλι σε πειράματα αναγνώρισης λέξεων επιδεικνύοντας σχετικά καλή ανοχή στον θόρυβο. Το κοχλιόγραμμα επιχειρεί να μοντελοποιήσει τις νευρωνικές εκφορτίσεις επί του κοχλία βασισζόμενο στις ιδιότητες που είναι γνωστές για αυτόν. Η έξοδος που προκύπτει αναπαριστά αυτές τις εκφορτίσεις υπό τη μορφή φασματογράμματος το οποίο όμως, σε αντίθεση με τα κοινά φασματογράμματα, εμπεριέχει σημαντικό όγκο χρονικής πληροφορίας. Το κοχλιόγραμμα χρησιμοποιείται για να παραχθεί το συσχετισιόγραμμα σκοπός του οποίου είναι ο εντοπισμός περιοδικότητων του ηχητικού σήματος στις οποίες φαίνεται να είναι ευαίσθητο το ακουστικό νευρικό σύστημα. Αν και είναι αληθές ότι υπάρχουν νευρώνες στο ακουστικό σύστημα που κλειδώνουν σε περιοδικότητες του σήματος, δεν έχει αποδειχθεί ακόμα ότι το σήμα αναπαρίσταται σε κάποιο στάδιο υπό τη μορφή ενός πλήρους συσχετισιογράμματος. Επομένως, η ύπαρξη του παραμένει προς το παρόν υπόθεση. Η πιο ολοκληρωμένη πρόταση αυτή τη στιγμή φαίνεται να είναι το μοντέλο που έχει αναπτυχθεί από τον Shihab Shamma και τους συνεργάτες του. Το πλεονέκτημα αυτού του μοντέλου είναι ότι έχει βασιστεί σε δεδομένα από εκτεταμένες καταγραφές της απόκρισης φλοιικών νευρώνων σε ηχητικά ερεθίσματα. Με βάση αυτά τα δεδομένα, προτείνεται ένα μοντέλο στο οποίο οι νευρώνες λειτουργούν ως ανιχνευτές συντονισμένοι σε συγκεκριμένα χαρακτηριστικά του σήματος. Ένα πλήθος τέτοιων νευρώνων σε πολλαπλά επίπεδα διακριτικής ικανότητας συνιστούν το φλοιικό στάδιο επεξεργασίας του ήχου πριν αυτός περάσει στα ανώτερα επίπεδα σημασιολογικής ανάλυσης.

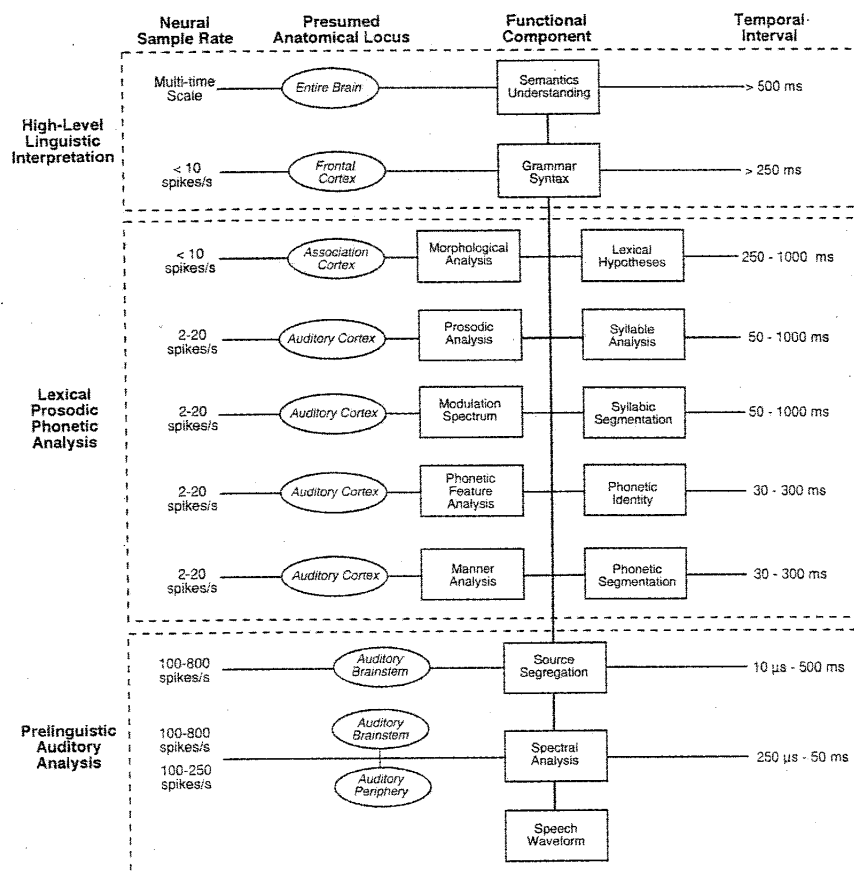
Μία δυσκολία στις προσεγγίσεις που βασίζονται σε νευροφυσιολογικά δεδομένα έγκειται στην λήψη των δεδομένων η οποία γίνεται σε αναισθητοποιημένα ζώα. Η χρήση ανθρώπων ως υποκειμένων στα πειράματα περιορίζεται σαφώς σε όσα χρησιμοποιούν μη επεμβατικές απεικονιστικές τεχνικές οι οποίες όμως δεν παρέχουν το επιθυμητό επίπεδο διακριτικής ικανότητας που μπορεί να φτάνει ακόμα και σε μεμονωμένους νευρώνες. Από την άλλη, δεν υπάρχουν σοβαροί λόγοι να θεωρήσουμε σημαντικές τις νευροανατομικές διαφορές μεταξύ του ανθρώπου και άλλων θηλαστικών, ειδικά αν η ανάλυσή μας κινείται σε ένα επίπεδο που προηγείται των ανώτερων σταδίων επεξεργασίας του λόγου. Εξαιρεση ενδεχομένως αποτελούν κάποια είδη με εξαιρετικά εξειδικευμένο ακουστικό σύστημα, όπως οι νυχτερίδες που το χρησιμοποιούν ως σύστημα ηχοεντισμού. Σημαντική δυσκολία υπάρχει και στον τρόπο εξέτασης της επιτυχίας τέτοιων μοντέλων. Όπως αναφέρθηκε παραπάνω, η χρήση συστημάτων ASR δεν ενδείκνυται. Ο πιο ευθύς τρόπος θα ήταν η εκτέλεση ψυχοακουστικών πειραμάτων και η διερεύνηση του κατά πόσο οι αποκρίσεις των υποκειμένων συμφωνούν με αυτές του μοντέλου. Εδώ βέβαια τίθεται το ζήτημα του τι συνιστά μία τέτοια συμφωνία, κάτι που μπορεί να μην είναι και τόσο προφανές. Μία τέτοια συζήτηση όμως μπορεί να μας μεταφέρει στο πεδίο της επιστημολογίας. Τέτοιου είδους συζητήσεις μέσα στον χώρο των γνωσιακών επιστημών ίσως να μην είναι και τόσο σπάνιες στο μέλλον.

Κεφάλαιο 4

Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου

Κατόπιν όλων των παραπάνω, το εύλογο ερώτημα είναι πάνω σε ποια χαρακτηριστικά των σημάτων ομιλίας θα έπρεπε να δομηθεί ένα γνωσιακό μοντέλο αντίληψης του λόγου. Ποια είναι αυτά τα στοιχεία που χρησιμοποιεί το ακουστικό σύστημα για να αποκωδικοποιήσει το μήνυμα που μεταφέρει ένα σήμα λόγου; Με βάση τις περιορισμένες γνώσεις και τα λίγα δεδομένα που είναι διαθέσιμα σήμερα θα μπορούσαμε να ισχυριστούμε ότι το ακουστικό σύστημα δομείται κατά έναν ιεραρχικό τρόπο όπου σε κάθε στάδιο έχουμε και ένα πιο αφαιρετικό επίπεδο ανάλυσης, ξεκινώντας από την απλή φασματική ανάλυση στο περιφερικό σύστημα και φτάνοντας μέχρι τους κανόνες της γραμματικής και τη σημασιολογία στον φλοιό. Αυτός ο τρόπος οργάνωσης παρουσιάζεται σχηματικά στο σχήμα 4.1. Στα κατώτερα στάδια του περιφερικού συστήματος και του εγκεφαλικού στελέχους επιτελείται μία ανάλυση που είναι κοινή σε όλα τα ηχητικά σήματα. Πρόκειται για μία κυρίως φασματική ανάλυση που επιτυγχάνει τον διαχωρισμό των πηγών και την αυξημένη ανοχή στον θόρυβο μέσω νευρώνων με ταχείες αποκρίσεις. Η φωνητική και προσωδιακή ανάλυση φαίνεται να λαμβάνει χώρα στον ακουστικό φλοιό ενώ η λεξική επεκτείνεται και στις συνειρμικές περιοχές του φλοιού. Τα ανώτερα επίπεδα της γραμματικής και της σημασιολογίας ενεργοποιούν αρκετά μεγάλες και διαφορετικές περιοχές του εγκεφάλου με αρκετά υψηλούς χρόνους επεξεργασίας.

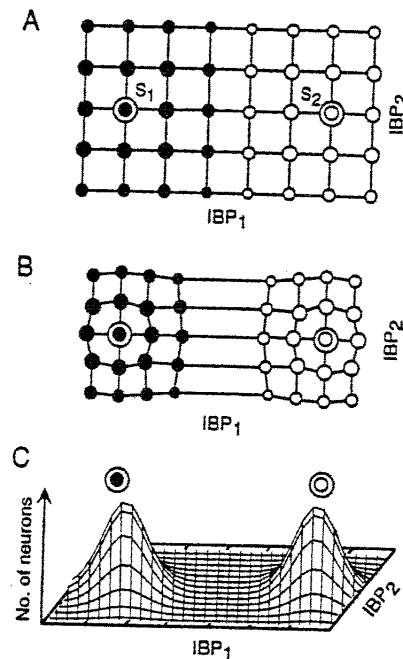
Μία οργάνωση του ακουστικού συστήματος όπως αυτή που παρουσιάστηκε προηγουμένως υποδεικνύει ότι ένα γνωσιακό μοντέλο που θα ήθελε να κινηθεί στα ανώτερα επίπεδα της σημασιολογικής, γραμματικής και λεξικής ανάλυσης θα έπρεπε να βασιστεί στις πληροφορίες που παρέχονται από τα κατώτερα επίπεδα. Το είδος αυτών των πληροφοριών και ο τρόπος παροχής τους παραμένουν προς το παρόν ανοιχτά ζητήματα. Ωστόσο, με βάση τα όσα έχουν αναφερθεί στα προηγούμενα κεφάλαια, θα μπορούσε να γίνει μία πρώτη προσπάθεια μοντελοποίησης αυτής της διαδικασίας. Στην παρούσα εργασία υιοθετείται η προσέγγιση του Nobuo Suga (βλ. κεφάλαιο 11, [7]).



Σχήμα 4.1: Σχηματική αναπαράσταση των σταδίων επεξεργασίας του ακουστικού συστήματος, [7]

4.1 Ανάλυση του μοντέλου εσωτερικής αναπαράστασης

Το μοντέλο που προτείνεται, βασίζεται στην υπόθεση της ύπαρξης νευρώνων στο νευρικό σύστημα (κυρίως στον ακουστικό φλοιό) που λειτουργούν ως φίλτρα συντονισμένα σε συγκεκριμένες παραμέτρους του ηχητικού σήματος, τα λεγόμενα IBP (information bearing parameters) φίλτρα. Όπως είδαμε στην προηγούμενη ενότητα, μία τέτοια υπόθεση αποτελεί και τη βάση του μοντέλου του Shamma. Αυτοί οι νευρώνες συνδυάζονται κατά τέτοιο τρόπο ώστε να σχηματίζουν αντιληπτικούς χάρτες, όπως φαίνεται στο σχήμα 4.2. Στο σχήμα A φαίνεται η κωδικοποίηση δύο σημάτων S_1 και S_2 σε ένα χώρο δύο παραμέτρων όπου οι μεγαλύτεροι κύκλοι υποδηλώνουν συχνότερη εμφάνιση της κάθε παραλλαγής των δύο βασικών προτύπων. Μετά από την έκθεση στα δύο πρότυπα ο χώρος μεταβάλλεται όπως στο σχήμα B, με τις αποστάσεις μεταξύ των παραλλαγών του ίδιου σήματος να μειώνονται και αυτές μεταξύ παραλλαγών διαφορετικών σημάτων να αυξάνονται. Αυτή η λειτουργία του “αντιληπτικού μαγνήτη” είναι πιθανό να επιτελείται όπως στο σχήμα C. Νευροβιολογικά δεδομένα υποστηρίζουν ότι ερεθίσματα που εμφανίζονται συχνά κωδικοποιούνται από περισσότερους νευρώνες σε σχέση με τα σπανιότερα. Φυσικά, το νευρικό σύστημα θα πρέπει να διαθέτει πολλούς τέτοιους, ενδεχομένως πολυδιάστατους, χάρτες επί διαφόρων παραμέτρων και διακριτικών επιπέδων.



Σχήμα 4.2: Αντιληπτικοί χάρτες, [7]

Οι νευρώνες που συνθέτουν τους παραπάνω χάρτες ή φίλτρα IBP θεωρείται ότι είναι εξειδικευμένοι νευρώνες, συντονισμένοι σε συγκεκριμένα χαρακτηριστικά ή συνδυασμούς

χαρακτηριστικών του ηχητικού σήματος. Τέτοιες παράμετροι θα μπορούσαν να είναι οι σχετικές με AM και FM διαμόρφωση (ρυθμός και βάθος διαμόρφωσης), το πλάτος, η συχνότητα, η διάρκεια κτλ. Στο ακουστικό σύστημα έχουν εντοπιστεί νευρώνες που είναι ευαίσθητοι σε τέτοιες παραμέτρους ή σε συνδυασμούς αυτών. Βασικά στοιχεία ενός φασματογράμματος που είναι κρίσιμα για την αντίληψη του λόγου είναι οι κάθετες μπάρες από εκρήξεις θορύβου (NB, noise bursts) (συνήθως στην περιοχή πάνω από 2 kHz), οι οριζόντιες μπάρες σταθερής συχνότητας (CF, constant frequency) και οι FM μεταβάσεις (συνήθως στην περιοχή κάτω των 3 kHz). Σημαντικές είναι επίσης και οι αργές AM διαμορφώσεις. Το κεντρικό ακουστικό σύστημα διαθέτει νευρώνες που είναι ευαίσθητοι σε αυτές τις παραμέτρους όπως και νευρώνες που παρουσιάζουν ασυμμετρίες, π.χ. επιλεκτικότητα στην κατεύθυνση ενός FM sweep, όπως έχει ήδη αναφερθεί. Επιπλέον, σε πολλά είδη έχουν βρεθεί νευρώνες, ιδιαίτερα σε περιοχές εκτός του AI, που αποκρίνονται βέλτιστα σε συνδυασμούς αυτών των χαρακτηριστικών. Σημαντικοί είναι οι συνδυασμοί CF/CF και FM/FM. Οι περισσότεροι CF/CF νευρώνες είναι ευαίσθητοι σε συνδυασμούς δύο συχνοτήτων αλλά μερικοί παρουσιάζουν ευαισθησία και σε συνδυασμούς τριών συχνοτήτων. Οι περιοχές του εγκεφάλου που διαθέτουν τέτοιους νευρώνες ενδεχομένως παίζουν σημαντικό ρόλο στον εντοπισμό συνδυασμών από μορφικές. Ως γνωστό, τέτοιοι συνδυασμοί είναι τα φωνήεντα. Οι FM/FM νευρώνες είναι ιδιαίτερα ευαίσθητοι στη χρονική απόσταση μεταξύ των δύο FM ήχων. Επίσης, υπάρχουν νευρώνες με τη δυνατότητα εντοπισμού ήχων με τρία ή και τέσσερα FM συστατικά. Πολλοί από τους FM/FM νευρώνες αντιδρούν καλύτερα σε FM ερεθίσματα παρά σε απλούς τόνους σταθερής συχνότητας. Οι περιοχές με τέτοιους νευρώνες θα έχουν ως βασική λειτουργία την επεξεργασία συνδυασμών από μεταβάσεις. Ακόμα, όπως έχει αναφερθεί και προηγουμένως, έχει επιβεβαιωθεί και η ύπαρξη νευρώνων ευαίσθητων τόσο σε ενάρξεις όσο και σε παύσεις ήχων, αν και φαίνεται ότι οι onset νευρώνες κυριαρχούν, γεγονός που μπορεί να γίνει κατανοητό από μία εξελικτική σκοπιά. Από τα παραπάνω μπορούμε να συνάγουμε ότι τα διάφορα χαρακτηριστικά ενός ηχητικού σήματος εντοπίζονται και υφίστανται επεξεργασία διαμέσου παράλληλων οδών σε διαφορετικές περιοχές του φλοιού για να σχηματιστούν τελικά οι αντιληπτικοί χάρτες.

Όπως θα παρουσιαστεί αναλυτικότερα στη συνέχεια, το μοντέλο που υλοποιείται στην παρούσα εργασία αποτελείται ουσιαστικά από συστοιχίες ανιχνευτών που επεξεργάζονται το φασματόγραμμα για να εντοπίσουν χαρακτηριστικά όπως αυτά που αναφέρθηκαν παραπάνω. Η ανάλυση γίνεται σε διάφορα επίπεδα χρονικής και συχνοτικής διακριτικής ικανότητας. Κάθε ανιχνευτής “κοιτάει” ένα συγκεκριμένο παράθυρο του φασματογράμματος που έχει μία κεντρική συχνότητα για μία συγκεκριμένη χρονική στιγμή. Για κάθε τέτοιο ζεύγος κεντρικής συχνότητας και χρονικής στιγμής η ανάλυση πραγματοποιείται για διάφορα εύρη ζώνης και χρονικά εύρη. Προς το παρόν, το μοντέλο διαθέτει έξι είδη ανιχνευτών, συμμετρικών ανά δύο. Το πρώτο ζεύγος ανιχνευτών είναι οι onset και offset που κωδικοποιούν ενάρξεις και παύσεις ήχων αντίστοιχα. Σκοπός τους είναι ο εντοπισμός των ευρυζωνικών εκρήξεων θορύβου που εμφανίζονται στο φασματόγραμμα ως κάθετες μπάρες ενέργειας καθώς και της έναρξης και λήξης ακουστικών γεγονότων. Το επόμενο ζεύγος είναι οι upwards και downwards ανιχνευτές που συλλαμβάνουν FM μεταβάσεις οι οποίες εμφανίζονται ως κεκλιμένες μπάρες ενέργειας. Αυτοί οι ανιχνευτές είναι επιλεκτικοί ως προς την κατεύθυνση της μετάβασης, όπως υποδεικνύουν και τα ονόματά τους. Τέλος, έχουμε τους high και low ανιχνευτές που εντοπίζουν φασματικές ακμές κατά τη διάσταση της συχνότητας και οι οποίες παρουσιάζονται στο φασματόγραμμα ως οριζόντιες μπάρες

σταθερής συχνότητας. Όπως και οι FM ανιχνευτές, και αυτοί παρουσιάζουν ευαισθησία αναλόγως με το αν η ενέργεια εμφανίζεται στο άνω ή κάτω ήμισυ του παραθύρου που εξετάζεται. Στην επόμενη ενότητα παρουσιάζεται πιο αναλυτικά η υλοποίηση του συστήματος.

4.2 Υλοποίηση του συστήματος

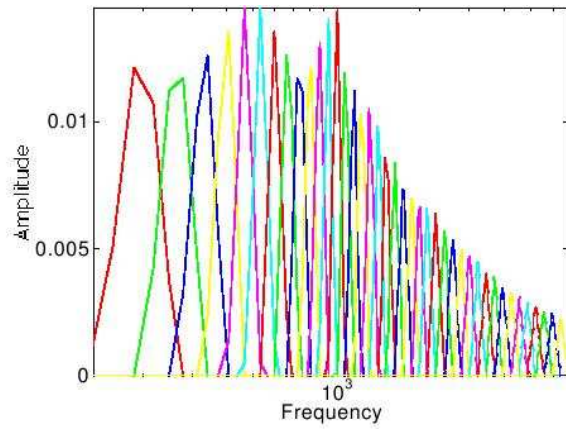
4.2.1 Φασματόγραμμα αρχικού σήματος

Η υλοποίηση του συστήματος έγινε στο περιβάλλον του MATLAB. Το πρώτο βήμα κατά την ανάλυση του σήματος είναι η αναπαράστασή του υπό τη μορφή φασματογράμματος. Θα ήταν επιθυμητό το φασματόγραμμα να είναι όσο το δυνατό ανεξάρτητο από τα ιδιосуγκρασιακά χαρακτηριστικά του ομιλητή που δεν έχουν σχέση με το μήνυμα αυτό καθαυτό που μεταφέρει ένα σήμα ομιλίας, όπως οι αρμονικές. Όπως είδαμε στο προηγούμενο κεφάλαιο, η PLP αναπαράσταση το επιτυγχάνει εν μέρει αυτό. Ωστόσο, το φασματόγραμμα που παράγεται με αυτή την αναπαράσταση είναι αρκετά παραμορφωμένο διατηρώντας μόνο τα πολύ γενικά χαρακτηριστικά του αρχικού φασματογράμματος. Για αυτό το λόγο, επιλέχθηκε η mfcc αναπαράσταση η οποία απομακρύνει από το αρχικό φασματόγραμμα τις πλεονάζουσες λεπτομέρειες αλλά διατηρεί αρκετά καλά τη γενική του μορφή. Με αυτή την επιλογή, αποφεύγουμε επίσης τις δυσκολίες που προκύπτουν από το κανονικό φασματόγραμμα το οποίο θα μπορούσε να αποπροσανατολίσει τους ανιχνευτές καθιστώντας τους ευαίσθητους στις διάφορες τοπικές διακυμάνσεις του. Μία τέτοια προσέγγιση δεν θα ήταν επιτρεπτή αν ο σκοπός του συστήματος ήταν και η αναγνώριση του ομιλητή καθώς σε αυτήν την περίπτωση οι λεπτομέρειες που τώρα απορρίπτουμε και που αποδίδουν σε μεγάλο βαθμό τη χροιά του ομιλητή θα μας ήταν αναγκαίες. Ωστόσο, στην παρούσα εργασία μάς ενδιαφέρει το γενικό περίγραμμα του φασματογράμματος που θεωρούμε ότι παραμένει σχετικά αμετάβλητο από ομιλητή σε ομιλητή για τον ίδιο ήχο.

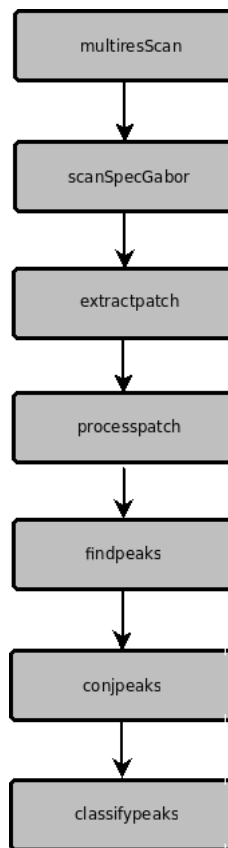
Για την υλοποίηση της mfcc αναπαράστασης χρησιμοποιήθηκε η αντίστοιχη συνάρτηση που υπάρχει στο Auditory Toolbox του Malcolm Slaney ([22]), με μικρές τροποποιήσεις. Όπως αναφέρεται και στο Technical Report του Toolbox, η συνάρτηση αυτή χρησιμοποιεί μία συστοιχία 40 φίλτρων με αποκρίσεις αυτές που φαίνονται στο σχήμα 4.3. Το σήμα περνάει αρχικά από παραθύρωση με παράθυρα Hamming, υπολογίζεται το πλάτος του FFT, γίνεται μετατροπή στις εξόδους της συστοιχίας φίλτρων και υπολογίζεται ο μετασχηματισμός συνημιτόνου. Ως είσοδο δέχεται το αρχικό σήμα, το ρυθμό δειγματοληψίας, το ρυθμό επανάληψης των πλαισίων και το μέγεθος τους. Στην έξοδο δίνει τους mel-frequency cepstral συντελεστές, το πλάτος του FFT, τις εξόδους της συστοιχίας φίλτρων, τις εξόδους της συστοιχίας μέσω μετασχηματισμού συνημιτόνου των συντελεστών cepstrum, την ανακατασκευασμένη, ομαλή απόκριση συχνότητας μέσω παρεμβολής των εξόδων στις συστοιχίες, και τα διανύσματα συχνότητας και χρόνου που αντιστοιχούν στο ανακατασκευασμένο φασματόγραμμα.

4.2.2 Επεξεργασία του φασματογράμματος

Αφού υπολογιστεί το mfcc φασματόγραμμα, οι συστοιχίες ανιχνευτών που αναφέρθηκαν παραπάνω το επεξεργάζονται για να εντοπίσουν σημαντικά γεγονότα σε αυτό. Σε αυτή την ενότητα περιγράφονται οι συναρτήσεις που χρησιμοποιούνται για αυτόν τον σκοπό (βλ.



Σχήμα 4.3: Η συστοιχία φίλτρων της συνάρτησης mfcc.



Σχήμα 4.4: Σχηματικό διάγραμμα του συστήματος επεξεργασίας.

σχήμα 4.3). Η `multiresScan` απλά εκκινεί τη διαδικασία επεξεργασίας και προετοιμάζει τις παραμέτρους που θα περάσουν στα επόμενα στάδια. Η `scanSpecGabor` εκκινεί με τη σειρά της τη διαδικασία επεξεργασίας αλλά για ένα μόνο παράθυρο ανάλυσης συγκεκριμένου εύρους ζώνης και χρονικού εύρους. Η `multiresScan` καλεί την `scanSpecGabor` επαναληπτικά για διάφορα μεγέθη παραθύρων ανάλυσης. Με τη σειρά της, η `scanSpecGabor` καλεί επαναληπτικά τις επόμενες συναρτήσεις ώστε να καλυφθεί όλο το φασματόγραμμα. Αρχικά, η `extractpatch` αποκόπτει από το φασματόγραμμα ένα συγκεκριμένο τμήμα του ώστε να επικεντρωθεί σε αυτό η περαιτέρω επεξεργασία. Η `processpatch` μετασχηματίζει αυτό το τμήμα σε μία άλλη διδιάστατη αναπαράσταση η οποία ιδανικά θα πρέπει να αποτελείται από ζεύγη συζυγών κορυφών. Η ακριβής θέση τους μας επιτρέπει να εξαγάγουμε συμπεράσματα σχετικά με το περιεχόμενο του αρχικού τμήματος. Οι κορυφές της νέας αναπαράστασης εντοπίζονται από την `findpeaks` ενώ η `conjpeaks` διατηρεί τα έγκυρα μόνο ζεύγη συζυγών κορυφών. Τέλος, η `classifypeaks` κατηγοριοποιεί τα εναπομείναντα ζεύγη ανάλογα με τον προσανατολισμό και την έντασή των κορυφών τους.

4.2.2.1 Η κύρια συνάρτηση, `multiresScan`

Η κύρια συνάρτηση που εκκινεί τη διαδικασία επεξεργασίας του φασματογράμματος ονομάζεται `multiresScan`. Οι είσοδοι και οι έξοδοί της παρουσιάζονται στη συνέχεια.

Είσοδοι :

1. **spec** : Το φασματόγραμμα, όπως έχει υπολογιστεί προηγουμένως από την `mfcc` συνάρτηση.
2. **F** : Το διάνυσμα συχνοτήτων του φασματογράμματος. Δίνεται και αυτό από την `mfcc` συνάρτηση.
3. **T** : Το διάνυσμα χρόνου του φασματογράμματος. Δίνεται και αυτό από την `mfcc` συνάρτηση.
4. **bws** : Ένα διάνυσμα με τα εύρη ζώνης στα οποία θα γίνει η ανάλυση.
5. **timews** : Ένα διάνυσμα με τα χρονικά εύρη στα οποία θα γίνει η ανάλυση.
6. **fscale** : Ένα αλφαριθμητικό, είτε 'linear' είτε 'mel', σχετικά με την κατανομή των κεντρικών συχνοτήτων στον άξονα της συχνότητας. Προς το παρόν, χρησιμοποιείται μόνο η επιλογή 'linear'.
7. **stepf** : Ένα διάνυσμα ή ένα βαθμωτό μέγεθος. Αν είναι διάνυσμα, οι κεντρικές συχνοτήτες των παραθύρων ανάλυσης είναι αυτές που καθορίζονται από το διάνυσμα. Αν είναι αριθμός, αυτός ο αριθμός καθορίζει την απόσταση μεταξύ των κεντρικών συχνοτήτων.
8. **stept** : Ένα διάνυσμα ή ένα βαθμωτό μέγεθος. Αν είναι διάνυσμα, οι χρονικές θέσεις των παραθύρων ανάλυσης είναι αυτές που καθορίζονται από το διάνυσμα. Αν είναι αριθμός, αυτός ο αριθμός καθορίζει την απόσταση μεταξύ των χρονικών θέσεων.

Έξοδοι :

1. **scannedspec** : Πρόκειται για μία δομή cell (πίνακας που μπορεί να περιέχει πίνακες διαφόρων διαστάσεων), διαστάσεων $length(timews) \cdot length(bws)$. Κάθε στοιχείο της δομής αυτής αντιστοιχεί σε ένα συγκεκριμένο ζεύγος εύρους ζώνης και χρονικού εύρους και είναι και αυτό μία δομή με έξι πίνακες. Κάθε πίνακας αντιστοιχεί σε έναν ανιχνευτή. Οι πίνακες αυτοί μάς δίνουν την απόκριση των ανιχνευτών στις διάφορες κεντρικές συχνότητες και χρονικές στιγμές.
2. **spF** : Μία δομή cell του MATLAB , διαστάσεων $length(timews) \cdot length(bws)$. Κάθε στοιχείο της δομής είναι ένα διάνυσμα που μας δίνει τις κεντρικές συχνότητες όπου έγινε ανάλυση για κάθε ζεύγος εύρους ζώνης και χρονικού εύρους.
3. **spT** : Μία δομή cell, διαστάσεων $length(timews) \cdot length(bws)$. Κάθε στοιχείο της δομής είναι ένα διάνυσμα που μας δίνει τις χρονικές θέσεις όπου έγινε ανάλυση για κάθε ζεύγος εύρους ζώνης και χρονικού εύρους.

Τα ορίσματα εισόδου *spec*, *F* και *T* πρέπει να δοθούν από τον χρήστη. Τα υπόλοιπα είναι προαιρετικά. Αν δεν δοθούν τίθενται αυτόματα στις τιμές $bws = [200 : 200 : F(end)/3]$, $timews = [0.01 \ 0.02 \ 0.04]$, $fscaler = 'linear'$, $stepf = 200$, $stept = 0.005$. Για κάθε συνδυασμό εύρους ζώνης και χρονικού εύρους, δηλαδή για κάθε συνδυασμό τιμών που προκύπτει από τα διανύσματα *bws* και *timews*, η *multiresScan* καλεί την *scanSpecGabor* που παρουσιάζεται στην επόμενη υποενότητα.

Στη συνέχεια παρουσιάζεται ένα παράδειγμα χρήσης αυτής της συνάρτησης. Έστω ότι έχουμε το σήμα *signal* που έχει ρυθμό δειγματοληψίας $F_s = 8000$ και διάρκεια 1 sec. Αρχικά λαμβάνουμε το *mfcc* φασματογράμμα ως εξής :

$$[ceps \ frresp \ fb \ fbrec \ Smfcc \ Fmfcc \ Tmfcc] = mfcc(signal, F_s, 1000, 512)$$

όπου χρησιμοποιούμε $frameRate = 1000$ και $windowSize = 512$. Οι έξοδοι που μας ενδιαφέρουν είναι οι *Smfcc*, *Fmfcc* και *Tmfcc*. Καλούμε την *multiresScan* ως εξής :

$$[scannedspec \ spF \ spT] = multiresScan(Smfcc, Fmfcc, Tmfcc, bws, timews, 'linear', stepf, stept)$$

όπου οι υπόλοιπες είσοδοι θα μπορούσαν να είναι

$$bws = [400 \ 600 \ 800 \ 1000 \ 1200]$$

$$timews = [0.01 \ 0.02 \ 0.04 \ 0.08]$$

$$stepf = (250 : 250 : 3500)$$

$$stept = (0.01 : 0.01 : 1)$$

Τα παραπάνω σημαίνουν ότι η ανάλυση του φασματογράμματος θα ξεκινάει από τα 250 Hz και θα φτάνει μέχρι τα 3500 με βήματα των 250. Επίσης, θα ξεκινάει από τη χρονική στιγμή 0.01 sec μέχρι το 1 sec με βήματα του 0.01. Τα παράθυρα θα έχουν εύρη ζώνης 400, 600, 800, 1000 και 1200 Hz στον άξονα της συχνότητας και 0.01, 0.02, 0.04 και 0.08 sec σε αυτόν του χρόνου.

4.2.2.2 Η συνάρτηση `scanSpecGabor`

Η `scanSpecGabor` αναλύει το φασματόγραμμα για ένα δεδομένο μέγεθος παραθύρου ανάλυσης τόσο στο χρόνο όσο και στη συχνότητα.

Είσοδοι :

1. **spec** : Το φασματόγραμμα, όπως προηγουμένως.
2. **F** : Το διάνυσμα συχνοτήτων, όπως προηγουμένως.
3. **T** : Το διάνυσμα χρόνου, όπως προηγουμένως.
4. **df** : Το εύρος ζώνης του παραθύρου ανάλυσης.
5. **dt** : Το χρονικό εύρος του παραθύρου ανάλυσης.
6. **stepf** : Όπως προηγουμένως, οι κεντρικές συχνότητες των παραθύρων ανάλυσης. Αν είναι βαθμωτό μέγεθος και $f_{scale} = 'mel'$ η κατανομή των κεντρικών συχνοτήτων γίνεται με βάση την κλίμακα mel ενώ αν $f_{scale} = 'linear'$ γίνεται γραμμικά με απόσταση μεταξύ των κεντρικών συχνοτήτων ίση με `stepf`.
7. **stept** : Όπως προηγουμένως, οι χρονικές στιγμές των παραθύρων ανάλυσης. Αν είναι βαθμωτό μέγεθος η απόσταση μεταξύ των χρονικών στιγμών θα είναι ίση με `stept`.
8. **scale** : Ένα αλφαριθμητικό, είτε `'linear'` είτε `'mel'` , όπως προηγουμένως. Αν δεν δοθεί τίθεται αυτόματα σε `'linear'`.

Έξοδοι :

1. **scannedspec** : Πρόκειται για μία δομή cell έξι πινάκων, όσοι και οι ανιχνευτές. Κάθε πίνακας περιέχει την απόκριση του αντίστοιχου ανιχνευτή στα διάφορα χρονικά σημεία και συχνότητες όπως καθορίζονται από τα `stepf` και `stept`. Οι διαστάσεις τους είναι $length(spF) \cdot length(spT)$.
2. **spF** : Διάνυσμα που περιέχει τις κεντρικές συχνότητες όπου έγινε η ανάλυση. Δεν συμπίπτει απαραίτητα με το `stepf` εφόσον για κάποιες από τις ακραίες τιμές του `stepf` ενδέχεται τα παράθυρα ανάλυσης να βγαίνουν εκτός των ορίων του φασματογραμματος. Αυτές οι τιμές απορρίπτονται.
3. **spT** : Διάνυσμα που περιέχει τις χρονικές στιγμές όπου έγινε η ανάλυση. Δεν συμπίπτει απαραίτητα με το `stept`.

Η συνάρτηση σαρώνει το φασματόγραμμα και σε κάθε σημείο του, όπως αυτά καθορίζονται από τα διανύσματα `stepf` και `stept`, εξάγει ένα μικρό τμήμα του (ένα patch) μέσω της συνάρτησης `extractpatch`. Το τμήμα που εξάγεται έχει εύρος ζώνης `df` και χρονικό εύρος `dt`. Στη συνέχεια επεξεργάζεται από την `processpatch` και γίνεται ο χαρακτηρισμός του από τις `findpeaks`, `conjpeaks` και `classifypeaks`. Με βάση αυτόν τον χαρακτηρισμό ανατίθεται η κατάλληλη τιμή στις εξόδους των ανιχνευτών. Αυτές οι συναρτήσεις παρουσιάζονται παρακάτω.

Ένα παράδειγμα χρήσης αυτής της συνάρτησης είναι :

$$[\text{scannedspec } spF \text{ } spT] = \text{scanSpecGabor}(Smfcc, Fmfcc, Tmfcc, 800, 0.08, \text{stepf}, \text{stept}, 'linear')$$

Το παράθυρο της ανάλυσης θα έχει εύρος 800 Hz και 0.08 δευτερόλεπτα. Οι *stepf* και *stept* θα είναι όπως πριν, με τη διαφορά ότι το παράθυρο ανάλυσης δεν θα πρέπει να είναι εκτός ορίων του φασματογράμματος. Η κλήση της μέσα από τη *multiresScan* το εξασφαλίζει αυτό.

4.2.2.3 Η συνάρτηση *extractpatch*

Η *extractpatch* αποσπά από ένα φασματογράμμα ένα τμήμα - παράθυρό του ώστε αυτό να περάσει μετά από κατάλληλη επεξεργασία.

Είσοδοι :

1. **spec** : Το φασματογράμμα, όπως προηγουμένως.
2. **F** : Το διάνυσμα συχνοτήτων, όπως προηγουμένως.
3. **T** : Το διάνυσμα χρόνου, όπως προηγουμένως.
4. **cf** : Η κεντρική συχνότητα του παραθύρου.
5. **ct** : Η χρονική του στιγμή.
6. **df** : Το εύρος ζώνης του.
7. **dt** : Το χρονικό του εύρος.

Έξοδοι :

1. **patch** : Το παράθυρο του φασματογράμματος που εξήχθη.
2. **pF** : Διάνυσμα συχνοτήτων του παραθύρου.
3. **pT** : Διάνυσμα χρόνου του παραθύρου.

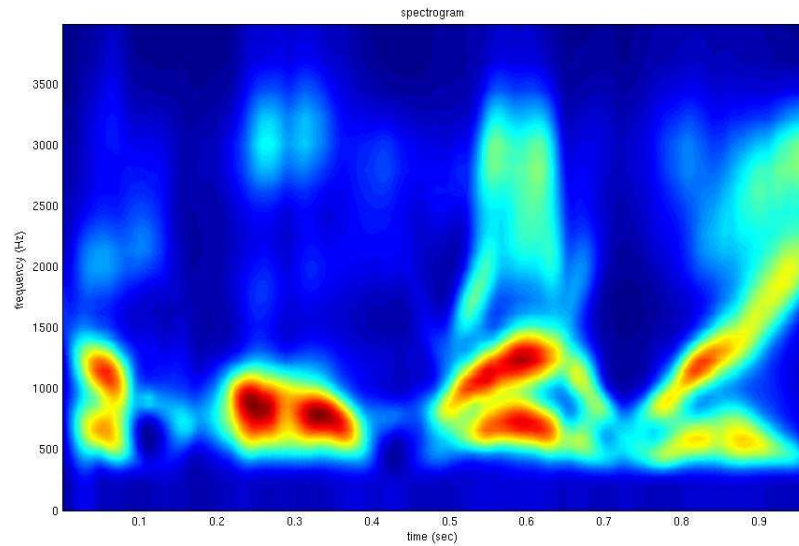
Η συνάρτηση κάνει κάποιους ελέγχους αν οι τιμές *cf*, *ct*, *df*, *dt* που δόθηκαν είναι έγκυρες, δηλαδή αν το παράθυρο που ορίζεται από αυτές είναι μέσα στα όρια του φασματογράμματος. Κατόπιν, απομονώνει τις τιμές του πίνακα *spec* που αντιστοιχούν στο παράθυρο στον πίνακα *patch*.

Ένα παράδειγμα χρήσης της είναι :

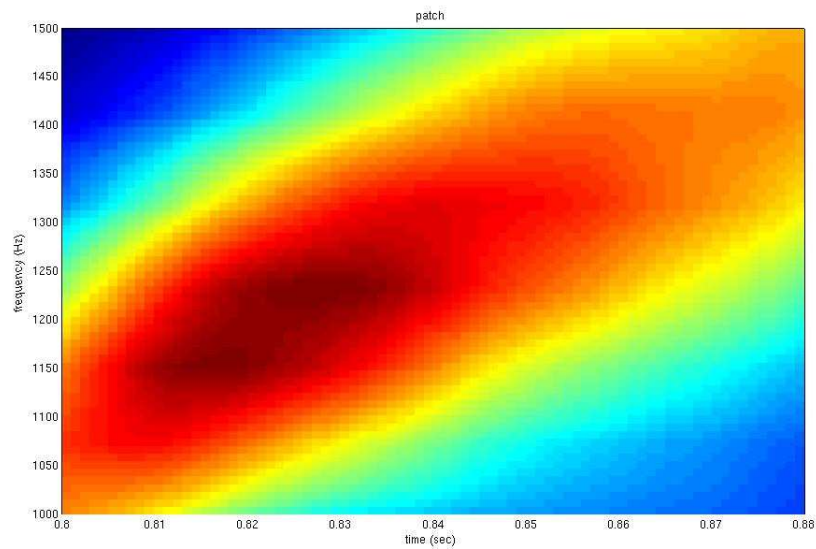
$$[\text{patch } pF \text{ } pT] = \text{extractpatch}(Smfcc, Fmfcc, Tmfcc, 1250, 0.84, 500, 0.08)$$

Η παραπάνω κλήση θα απομονώσει ένα τμήμα γύρω από τη συχνότητα 1250 Hz και τη στιγμή 0.84 δευτερόλεπτα. Εφόσον το εύρος ζώνης είναι 500 Hz και το χρονικό εύρος 0.08 δευτερόλεπτα, το παράθυρο θα εκτείνεται από 1000 μέχρι 1500 Hz και από τα 0.8 ως τα 0.88 δευτερόλεπτα. Στο σχήμα 4.5 φαίνεται το αποτέλεσμα της κλήσης αυτής της συνάρτησης με τα ορίσματα που δόθηκαν στο παράδειγμα.

Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου



(α') Φασματόγραμμα σήματος



(β') Παράθυρο που εξάγεται γύρω από τα 1250 Hz και τα 0.84 δευτερόλεπτα

Σχήμα 4.5: Παράδειγμα εφαρμογής της συνάρτησης extractpatch

4.2.2.4 Η συνάρτηση processpatch

Η processpatch δέχεται ως είσοδο το παράθυρο που έχει εξαχθεί μέσω της extractpatch και το επεξεργάζεται κατάλληλα ώστε να εντοπιστούν τυχόν ενδιαφέροντα χαρακτηριστικά του. Η επεξεργασία που επιτελείται βασίζεται στην εργασία των Ezzat et al ([5]). Αρχικά, από το παράθυρο $P_{ij}(f, t)$ αφαιρείται η μέση τιμή του $\frac{1}{df \cdot dt} \sum_{f,t} P_{ij}(f, t)$. Στη συνέχεια, το παράθυρο πολλαπλασιάζεται με ένα 2Δ Γκαουσιανό παράθυρο $W(f, t)$ γύρω από το κέντρο του f_0, t_0

$$W(f, t) = \frac{1}{2\pi\sigma_f\sigma_t} e^{-\frac{1}{2}\left(\frac{(f-f_0)^2}{\sigma_f^2} + \frac{(t-t_0)^2}{\sigma_t^2}\right)}$$

Τα σ_f και σ_t συνήθως είναι περίπου στο 1/3 του ύψους και του πλάτους του παραθύρου. Στη συνέχεια εφαρμόζεται ένας 2Δ μετασχηματισμός Fourier μεγέθους $N_H \cdot N_W$.

$$R_{ij}(\Omega, \omega) = \sum_f \sum_t W(f, t) P_{ij}(f, t) e^{-j2\pi \frac{\Omega}{N_H} f} e^{-j2\pi \frac{\omega}{N_W} t}$$

Η παραπάνω εξίσωση μπορεί να γραφτεί και ως :

$$R_{ij}(\Omega, \omega) = \sum_f \sum_t P_{ij}(f, t) G_{\Omega, \omega}^*(f, t)$$

όπου

$$G_{\Omega, \omega}(f, t) = W(f, t) e^{j2\pi\left(\frac{\Omega}{N_H} f + \frac{\omega}{N_W} t\right)}$$

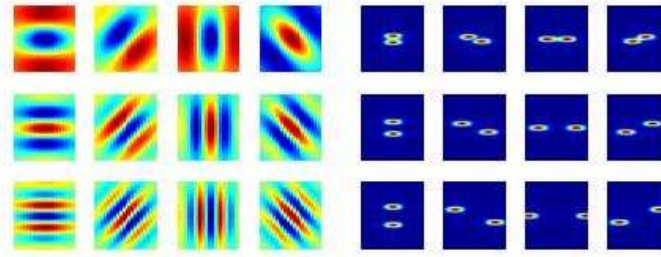
Η παραπάνω εξίσωση είναι αυτή ενός Gabor φίλτρου. Επομένως, το μέγεθος $R_{ij}(\Omega, \omega)$ μπορεί να θεωρηθεί ότι είναι η προβολή του παραθύρου $P_{ij}(f, t)$ επί μίας συστοιχίας Gabor φίλτρων $G_{\Omega, \omega}(f, t)$. Ο μετασχηματισμός Fourier ενός Gabor φίλτρου αποτελείται από ζεύγη συζυγών Γκαουσιανών κορυφών, ο προσανατολισμός και η απόσταση των οποίων από την αρχή των αξόνων εξαρτάται από τα χαρακτηριστικά του φίλτρου. Παραδείγματα Gabor φίλτρων με τους μετασχηματισμούς Fourier φαίνονται στο σχήμα 4.6. Ακολουθώντας αυτή τη διαδικασία για κάθε παράθυρο του αρχικού φασματογράμματος και εντοπίζοντας το είδος των κορυφών που τελικά προκύπτουν, μπορούμε να διακρίνουμε αν το παράθυρο περιέχει κάποια κάθετη, οριζόντια ή κεκλιμένη μπάρα ενέργειας, τα χαρακτηριστικά δηλαδή που μας ενδιαφέρουν. Η συνάρτηση processpatch λαμβάνει ένα παράθυρο του φασματογράμματος και παράγει αυτές τις Γκαουσιανές, συζυγείς κορυφές.

Είσοδοι :

1. **patch** : Το παράθυρο προς επεξεργασία.
2. **pF** : Το διάνυσμα συχνοτήτων του παραθύρου.
3. **pT** : Το διάνυσμα χρόνου του παραθύρου.

Έξοδοι :

1. **gabfft** : Τα ζεύγη Γκαουσιανών, συζυγών κορυφών.
2. **gabor** : Το Gabor φίλτρο που αντιστοιχεί στο παράθυρο patch.



Σχήμα 4.6: Παραδείγματα Gabor φίλτρων και οι μετασχηματισμοί Fourier που τους αντιστοιχούν, [5]

Ένα παράδειγμα χρήσης της συγκεκριμένης συνάρτησης είναι :

$$[gabfft \text{ gabor}] = \text{processpatch}(\text{patch}, pF, pT)$$

Αν εφαρμόσουμε τη συνάρτηση στο παράθυρο που πήραμε προηγουμένως, στο σχήμα 4.5(β'), λαμβάνουμε τα αποτελέσματα που φαίνονται στο σχήμα 4.7. Είναι εμφανές ότι ο προσανατολισμός που έχουν οι δύο κορυφές που προέκυψαν μπορεί να διακρίνει το ανοδικό fm sweep του παραθύρου.

4.2.2.5 Η συνάρτηση `findpeaks`

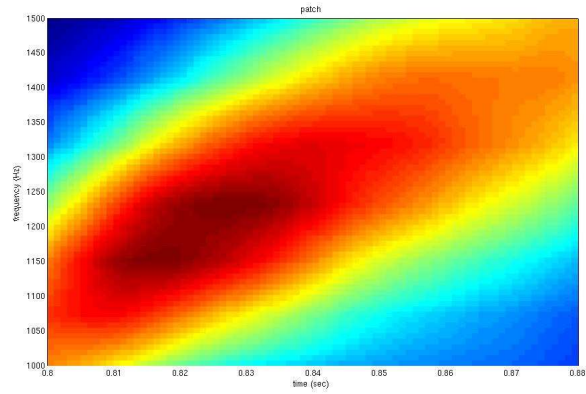
Από τη στιγμή που διαθέτουμε την αναπαράσταση υπό τη μορφή Γκαουσσισιανών, συζυγών κορυφών, το επόμενο βήμα είναι να εντοπίσουμε αυτές τις κορυφές και τις θέσεις τους. Η συνάρτηση `findpeaks` έχει αυτόν ακριβώς το σκοπό.

Είσοδοι :

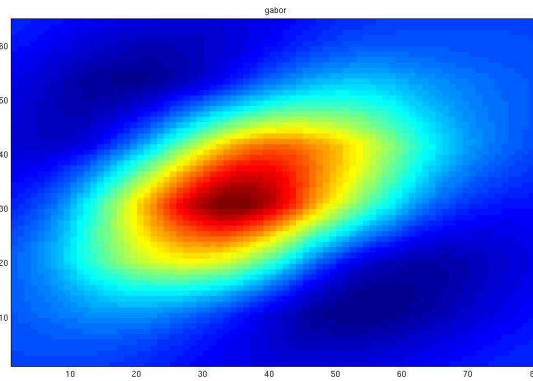
1. **gabfft** : Οι Γκαουσσισιανές κορυφές, όπως προέκυψαν από την `processpatch`.
2. **method** : Αλφαριθμητικό που δηλώνει τη μέθοδο που θα χρησιμοποιηθεί για τον εντοπισμό των κορυφών. Μπορεί να είναι μία εκ των 'tophat', 'threshold' ή 'regmax'. Η tophat χρησιμοποιεί τον ομώνυμο μετασχηματισμό (παρ. 10.7, [29]). Η threshold χρησιμοποιεί απλή κατωφλίωση. Τα καλύτερα αποτελέσματα επιτυγχάνονται με τη regmax η οποία χρησιμοποιεί τη συνάρτηση `imregionalmax` του MATLAB ([15]).
3. **noisethr** : Το κατώφλι θορύβου. Περιοχές με τιμές κάτω από αυτό το κατώφλι αποκλείονται.
4. **strelsize** : Το μέγεθος του δομικού στοιχείου (structuring element) των μορφολογικών τελεστών όταν χρησιμοποιείται η μέθοδος tophat.

Έξοδοι :

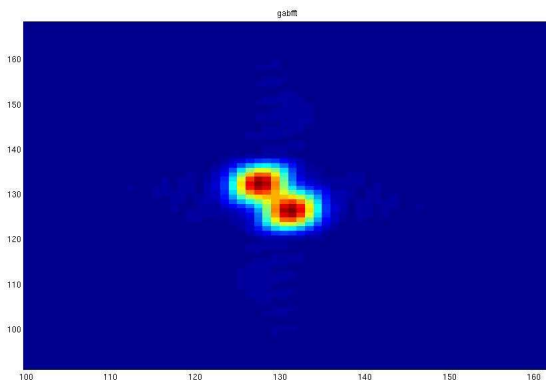
1. **peakslabel** : Εικόνα με τις κορυφές που εντοπίστηκαν. Τα εικονοστοιχεία κάθε κορυφής σημειώνονται με ένα δικό τους αριθμό (label).



(α') Παράθυρο φασματογράμματος



(β') Gabor φίλτρο για το παράθυρο



(γ') Συζυγείς, Γκαουσιανές κορυφές για το παράθυρο

Σχήμα 4.7: Παράδειγμα εφαρμογής της συνάρτησης processpatch

Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου

2. **components** : Δομή cell διαστάσεων $2 \cdot N$, όπου N ο αριθμός των κορυφών που εντοπίστηκαν. Κάθε στήλη του cell αντιστοιχεί και σε μία κορυφή. Τα στοιχεία $components\{1,i\}$ είναι πίνακες που περιέχουν τις x συντεταγμένες των εικονοστοιχείων που αποτελούν την κορυφή i ενώ τα $components\{2,i\}$ είναι πίνακες με τις y συντεταγμένες αυτών των εικονοστοιχείων.
3. **centroids** : Πίνακας διαστάσεων $2 \cdot N$, όπου N ο αριθμός των κορυφών που εντοπίστηκαν. Περιέχει τις συντεταγμένες των κέντρων των κορυφών. Τα στοιχεία $centroids(1,i)$ περιέχουν τη x συντεταγμένη του κέντρου της κορυφής i και τα $centroids(2,i)$ την y .
4. **peakints** : Διάνυσμα N στοιχείων που περιέχει την ένταση των κορυφών που εντοπίστηκαν.

Για να χρησιμοποιηθεί η `imregionalmax`, η αρχική εικόνα μετατρέπεται σε μία άλλη με ένα συγκεκριμένο αριθμό (εδώ 10) διακριτών επιπέδων έντασης. Η ένταση κάθε εικονοστοιχείου μετατρέπεται σε αυτήν του κοντινότερου από τα διαφορετικά επίπεδα. Αφού εντοπιστούν οι κορυφές, στη συνέχεια υπολογίζονται οι συντεταγμένες των στοιχείων που ανήκουν σε αυτές, οι συντεταγμένες των κέντρων τους και η έντασή τους.

Συνεχίζοντας με το προηγούμενο παράδειγμα, η κλήση της συνάρτησης ως εξής :

$$[peaklabel\ comp\ centrs\ peakints] = findpeaks(gabfft, 'regmax')$$

θα πρέπει να εντοπίσει τις δύο συζυγείς κορυφές που αντιστοιχούν στο ανοδικό `fm sweep`. Στο σχήμα 4.8 φαίνεται το `peaklabel` που προκύπτει (με ένα ζουμ) με τα εικονοστοιχεία της κάθε κορυφής σημειωμένα με διαφορετικό χρώμα (`label`). Τα `componenets` θα είναι :

$$components\{1,1\} = [3\ 4\ 5\ 4\ 5]$$

$$components\{2,1\} = [-1\ -1\ -1\ 0\ 0]$$

$$components\{1,2\} = [-3\ -2\ -3\ -2\ -1]$$

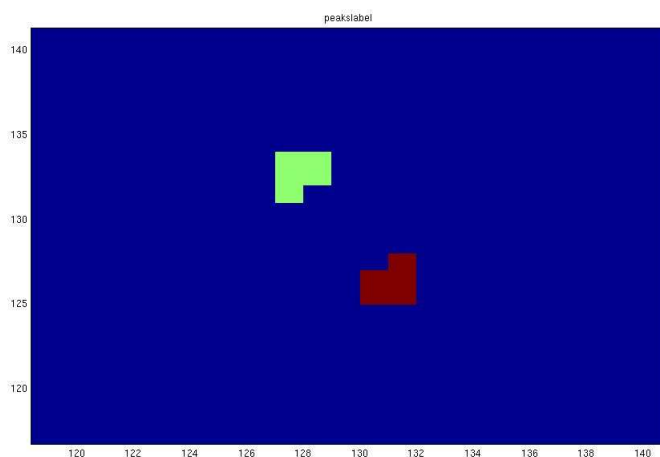
$$components\{1,2\} = [2\ 2\ 3\ 3\ 3]$$

Τα `centroids` :

$$centroids = \begin{bmatrix} 4.200 & -2.200 \\ -0.600 & 2.600 \end{bmatrix}$$

Τέλος, οι εντάσεις των κορυφών είναι :

$$peakints = [0.9349\ 0.9349]$$



Σχήμα 4.8: Παράδειγμα εφαρμογής της συνάρτησης findpeaks

4.2.2.6 Η συνάρτηση conjpeaks

Η συνάρτηση conjpeaks εξετάζει τις κορυφές που εντόπισε η findpeaks και βρίσκει τα ζεύγη συζυγών κορυφών.

Είσοδοι :

1. **centrs** : Τα κέντρα των κορυφών, το centroids από την findpeaks.
2. **distthres** : Ένα κατώφλι απόστασης μεταξύ δύο κορυφών.
3. **peakints** : Οι εντάσεις των κορυφών, όπως προκύπτουν από την findpeaks.

Έξοδοι :

1. **validpeaks** : Μία δομή cell διαστάσεων $2 \cdot M$, όπου M ο αριθμός των ζευγών συζυγών κορυφών που εντοπίστηκαν. Το $\text{validpeaks}\{1,i\}$ είναι ένας πίνακας $2 \cdot 2$ με τις συντεταγμένες των κορυφών του ζεύγους i και το $\text{validpeaks}\{2,i\}$ είναι η μέση τιμή των εντάσεων των δύο κορυφών.
2. **dists** : Πίνακας $N \cdot N$, όπου N ο αριθμός των κορυφών. Η σημασία του εξηγείται παρακάτω.

Ο αλγόριθμος εντοπίζει τα ζεύγη συζυγών κορυφών με βάση τις αποστάσεις μεταξύ των κορυφών. Για κάθε κορυφή, υπολογίζεται η απόσταση της από όλες τις άλλες κορυφές. Στην πραγματικότητα, οι αποστάσεις αυτές είναι οι αποστάσεις από τις συζυγείς όλων των άλλων κορυφών και όχι από τις ίδιες τις κορυφές. Με αυτόν τον τρόπο, αν υπάρχει κάποια άλλη κορυφή συζυγής αυτής που εξετάζεται, η απόσταση που θα προκύψει θα είναι μηδέν ή πολύ κοντά στο μηδέν. Η ίδια διαδικασία επαναλαμβάνεται για όλες τις κορυφές και έτσι προκύπτει ο πίνακας dists με διαστάσεις $N \cdot N$. Αφού κατασκευαστεί αυτός ο πίνακας, εντοπίζεται το ζεύγος που έχει τη μικρότερη απόσταση σε αυτόν και όλες οι αποστάσεις

που αντιστοιχούν στις κορυφές αυτού του ζεύγους τίθενται σε μία “άπειρη” τιμή. Αν αυτή η μικρότερη απόσταση είναι μικρότερη και από το κατώφλι `distthres`, τότε θεωρείται ότι έχει εντοπιστεί ένα έγκυρο ζεύγος το οποίο και προστίθεται στο `validpeaks`. Αυτό επαναλαμβάνεται μέχρι να εξαντληθούν όλες οι κορυφές ή μέχρι η ελάχιστη απόσταση στον πίνακα `dists` να γίνει μεγαλύτερη του κατωφλίου.

Για το παράδειγμά μας :

$$[validpeaks\ dists] = conjpeaks(centrs, 5, peakints)$$

$$validpeaks\{1, 1\} = \begin{bmatrix} -2.200 & 4.200 \\ 2.600 & -0.600 \end{bmatrix}$$

$$validpeaks\{2, 1\} = 0.9349$$

$$dists = \begin{bmatrix} Inf & 2.8284 \\ 2.8284 & Inf \end{bmatrix}$$

4.2.2.7 Η συνάρτηση `classifypeaks`

Η συνάρτηση `classifypeaks` εξετάζει τα έγκυρα ζεύγη κορυφών που εντόπισε η `conjpeaks` και τους αποδίδει ένα χαρακτηρισμό ο οποίος θα καθορίσει και την έξοδο των ανιχνευτών.

Είσοδοι :

1. **validpeaks** : Τα έγκυρα ζεύγη κορυφών, όπως προκύπτουν από την `conjpeaks`.
2. **patch** : Το παράθυρο του φασματογράμματος που εξετάζεται.
3. **vertlim** : Ένα όριο για χαρακτηρισμό ως κάθετη μπάρα ενέργειας. Προαιρετικό.
4. **horizlim** : Ένα όριο για χαρακτηρισμό ως οριζόντια μπάρα ενέργειας. Προαιρετικό.
5. **onoffthr** : Ένα κατώφλι για χαρακτηρισμό ως onset ή offset στην περίπτωση κάθετης μπάρας ενέργειας. Προαιρετικό.
6. **lowhighthr** : Ένα κατώφλι για χαρακτηρισμό ως low ή high στην περίπτωση οριζόντιας μπάρας ενέργειας. Προαιρετικό.
7. **balthr** : Ένα κατώφλι ασυμμετρίας, όπως εξηγείται παρακάτω.

Έξοδοι :

1. **peakstype** : Μία δομή cell διαστάσεων $2 \cdot M$, όπου M ο αριθμός των ζευγών συζυγών κορυφών. Για κάθε τέτοιο ζεύγος, το `peakstype\{1,i\}` είναι ένα αλφαριθμητικό που μας δίνει τον χαρακτηρισμό του ζεύγους και το `peakstype\{2,i\}` είναι η ένταση του ζεύγους. Ο χαρακτηρισμός μπορεί να είναι ένας από τους εξής : `onset`, `offset`, `verticalbar`, `lowhorizontal`, `highhorizontal`, `horizontalbar`, `downsweep`, `upsweep`, `undefined`.

Για τον χαρακτηρισμό ενός ζεύγους, αρχικά υπολογίζεται η κλίση της ευθείας που ενώνει τις κορυφές του ζεύγους. Αν η κλίση είναι μικρότερη του `vertlim` έχουμε `verticalbar` ενώ αν είναι μεγαλύτερη από `horizlim` έχουμε `horizontalbar`. Διαφορετικά, αν είναι θετική έχουμε `downsweep` και αν είναι αρνητική έχουμε `upsweep`. Αν τίποτα από τα παραπάνω δεν ισχύει, το αποτέλεσμα είναι `undefined`. Στην περίπτωση του `verticalbar`, συγκρίνουμε το παράθυρο με ένα πρότυπο `onset` (παράθυρο που το μισό, δεξί του τμήμα είναι πλήρως ενεργοποιημένο και το αριστερό μηδέν) και ένα πρότυπο `offset` (το αντίστροφο του `onset`). Αν το λάθος που προκύπτει είναι μικρότερο από `onoffthr` και η ασυμμετρία προς τη μία πλευρά έντονη (π.χ. για το `onset` θα πρέπει η μέση τιμή του δεξιού μισού προς τη μέση τιμή του αριστερού να είναι μεγαλύτερη από `balthr`), τότε ο χαρακτηρισμός θα είναι αναλόγως είτε `onset` είτε `offset`. Αλλιώς θα είναι απλά `verticalbar`. Αντιστοίχως γίνεται και ο χαρακτηρισμός ως `lowhorizontal` ή `highhorizontal` στην περίπτωση του `horizontalbar`.

Με κλήση της συνάρτησης για το παράδειγμά μας :

$$peakstype = classifypeaks(validpeaks, patch, 0.3, 1.2)$$

έχουμε

$$peakstype = \begin{bmatrix} 'upsweep' \\ 0.9349 \end{bmatrix}$$

4.2.2.8 Βοηθητικές συναρτήσεις

Στη συνέχεια θα παρουσιαστούν μερικές βοηθητικές συναρτήσεις οι οποίες χρησιμοποιούνται από τις συναρτήσεις που περιγράφηκαν προηγουμένως.

Η **quantintens** δέχεται ως είσοδο μία εικόνα και έναν αριθμό που αντιπροσωπεύει τον αριθμό των δυνατών επιπέδων τιμών που θέλουμε να έχει η εικόνα στην έξοδο. Κάθε εικονοστοιχείο της εικόνας λαμβάνει τιμή ίση με αυτή ενός εκ των επιπέδων. Το εύρος των τιμών στην εικόνα εξόδου είναι από 0 ως 1. Η συνάρτηση αυτή χρησιμοποιείται από την `findpeaks`.

Είσοδοι :

1. **image** : Η αρχική εικόνα.
2. **levelsno** : Το πλήθος των επιπέδων τιμών που θα έχει η τελική εικόνα.

Έξοδοι :

1. **quantim** : Η τελική, “χβαντοποιημένη” εικόνα.

Η **createTemplate** δημιουργεί ένα πρότυπο παράθυρο τύπου `onset`, `offset`, `highhorizontal` ή `lowhorizontal`. Χρησιμοποιείται από την `classifypeaks`.

Είσοδοι :

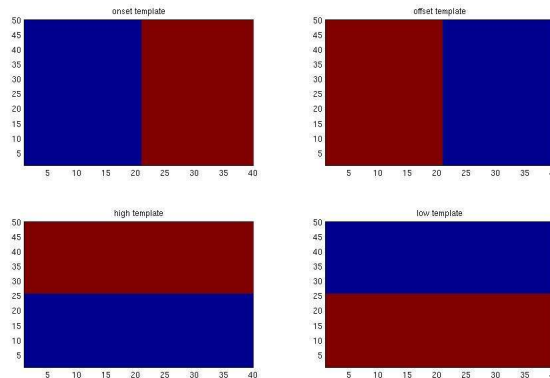
1. **type** : Ο τύπος του προτύπου. Ένα αλφαριθμητικό εκ των ‘`onset`’, ‘`offset`’, ‘`high`’, ‘`low`’.
2. **rows** : Ο αριθμός των σειρών του προτύπου.
3. **cols** : Ο αριθμός των στηλών του προτύπου.

Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου

Έξοδοι :

1. **templ** : Το πρότυπο που προκύπτει, διαστάσεων $rows \cdot cols$.

Στο σχήμα 4.9 παρουσιάζονται μερικά παραδείγματα προτύπων, όλα διαστάσεων $50 \cdot 40$.



Σχήμα 4.9: Παραδείγματα προτύπων, όπως προκύπτουν από τη συνάρτηση `createTemplate`

Η **templerror** εξετάζει δύο εικόνες ίδιων διαστάσεων και υπολογίζει το λάθος που υπάρχει μεταξύ τους. Αν δύο εικόνες είναι πανομοιότυπες το λάθος θα είναι ίσο με μηδέν ενώ αν είναι ακριβώς αντίθετες (π.χ. τα πρότυπα `onset` και `offset`) θα είναι ίσο με τη μονάδα. Χρησιμοποιείται από την `classifypeaks`.

Είσοδοι :

1. **input** : Η μία εικόνα, συνήθως αυτή που είναι προς σύγκριση.
2. **templ** : Η δεύτερη εικόνα, συνήθως κάποιο πρότυπο επί του οποίου γίνεται η σύγκριση.

Έξοδοι :

1. **error** : Το λάθος μεταξύ των δύο εικόνων.

Αν θέλουμε η ανάλυση να γίνεται στον άξονα `mel`, θα πρέπει, εκτός από τις κεντρικές συχνότητες, και τα εύρη ζώνης να ακολουθούν αυτή την κλίμακα. Σε αυτό βοηθάει η συνάρτηση **`bwmel2hz`** η οποία δέχεται δύο μονοδιάστατα διανύσματα τιμών, ίσου μεγέθους. Το ένα αντιστοιχεί στις κεντρικές συχνότητες σε `Hz` και το άλλο σε εύρος ζώνης σε `mel`. Η έξοδος μάς δίνει ένα άλλο διάνυσμα ίδιου μεγέθους, με τα εύρη ζώνης σε `Hz`.

Είσοδοι :

1. **cf** : Το διάνυσμα κεντρικών συχνοτήτων.
2. **melbw** : Το διάνυσμα με τα εύρη ζώνης σε `mel`.

Έξοδοι :

1. **hzbw** : Το διάνυσμα με τα εύρη ζώνης σε Hz.

Αν γνωρίζουμε τις κεντρικές συχνότητες και επιθυμούμε και τα εύρη ζώνης να ακολουθούν την κλίμακα mel, μπορούμε να δώσουμε στην παραπάνω συνάρτηση τις κεντρικές συχνότητες στην είσοδο cf και ένα διάνυσμα με σταθερό εύρος ζώνης σε mel στην είσοδο melbw. Με αυτόν τον τρόπο, το διάνυσμα που θα επιστραφεί θα έχει εύρη ζώνης για κάθε κεντρική συχνότητα που όμως θα ακολουθούν την κλίμακα mel. Η συνάρτηση χρησιμοποιείται από την `scanSpecGabor`, όταν το όρισμά της `scale` είναι 'mel'. Ακολουθεί ένα παράδειγμα χρήσης της συνάρτησης :

$$hzbw = bwmel2hz([500\ 1000\ 1500\ 2000], [200\ 200\ 200\ 200])$$

Η έξοδος είναι :

$$hzbw = 212.3987\ 300.8982\ 389.3976\ 477.8971$$

Η **energybar** δημιουργεί ένα σήμα που στο φασματόγραμμά του περιέχει μία μπάρα ενέργειας με συγκεκριμένα χαρακτηριστικά.

Είσοδοι :

1. **starttime** : Η χρονική στιγμή που ξεκινάει η μπάρα (sec).
2. **endtime** : Η χρονική στιγμή που τελειώνει η μπάρα (sec). Πρέπει να ισχύει $starttime < endtime$.
3. **startfreq** : Η συχνότητα που ξεκινάει η μπάρα (Hz).
4. **endfreq** : Η συχνότητα που τελειώνει η μπάρα (Hz). Πρέπει να ισχύει $startfreq < endfreq$.
5. **duration** : Η χρονική διάρκεια ολόκληρου του σήματος (sec). Πρέπει να ισχύει $duration > endtime$.
6. **F_s** : Η συχνότητα δειγματοληψίας του σήματος (Hz). Πρέπει να ισχύει $endfreq < \frac{F_s}{2}$.

Έξοδοι :

1. **eb** : Το προκύπτον σήμα.

Παρομοίως, η **fmsweep** δημιουργεί ένα FM sweep.

Είσοδοι :

1. **duration** : Η χρονική διάρκεια του σήματος (sec).
2. **startfreq** : Η συχνότητα που ξεκινάει το sweep (Hz). Πρέπει να ισχύει $startfreq < \frac{F_s}{2}$.
3. **endfreq** : Η συχνότητα που τελειώνει το sweep (Hz). Πρέπει να ισχύει $endfreq < \frac{F_s}{2}$. Αν $endfreq > startfreq$ το sweep θα είναι ανοδικό, διαφορετικά θα είναι καθοδικό.

Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου

4. **F_s** : Η συχνότητα δειγματοληψίας του σήματος (Hz).

Έξοδοι :

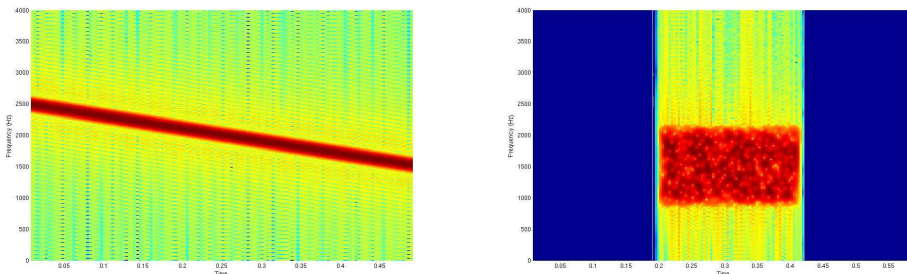
1. **sweep** : Το προκύπτον σήμα.

Οι δύο παραπάνω συναρτήσεις χρησιμοποιούνται για να κατασκευαστούν κάποια απλά σήματα με τα οποία θα δοκιμαστεί το σύστημα στη συνέχεια. Παραδείγματα χρήσης τους είναι :

$$\text{sweep} = \text{fmsweep}(0.5, 2500, 1500, 8000)$$

$$\text{eb} = \text{energybar}(0.2, 0.4, 1000, 2000, 0.6, 8000)$$

Τα φασματογράμματα των παραπάνω σημάτων φαίνονται στο σχήμα 4.10.



(α') Φασματογράμμα sweep, όπως αυτό προκύπτει από τη συνάρτηση fmsweep.

(β') Φασματογράμμα σήματος μπάρας ενέργειας, όπως αυτό προκύπτει από τη συνάρτηση energybar.

Σχήμα 4.10: Παραδείγματα εφαρμογής των συναρτήσεων fmsweep και energybar.

4.2.3 Απεικόνιση των αποτελεσμάτων

Η απεικόνιση των αποτελεσμάτων που προκύπτουν από την παραπάνω επεξεργασία γίνεται υπό τη μορφή βίντεο. Κάθε ανιχνευτής διαθέτει ένα 2Δ “χάρτη” για κάθε χρονικό εύρος με τις κεντρικές συχνότητες στον άξονα y και το εύρος ζώνης στον άξονα x. Σε κάθε χρονική στιγμή, ένας τέτοιος χάρτης παρουσιάζει ένα συγκεκριμένο μοτίβο ενεργοποίησης ανάλογα με την απόκριση των ανιχνευτών αυτού του τύπου στις διάφορες κεντρικές συχνότητες και εύρη ζώνης. Μία ακολουθία από τέτοιες εικόνες, ξεκινώντας από την αρχή του σήματος, μπορεί να παρουσιαστεί ως ένα βίντεο.

Ωστόσο, η δομή που έχουν τα δεδομένα, όπως προκύπτουν από τη συνάρτηση multiresScan δεν είναι κατάλληλη για μία τέτοια αναπαράστασή τους. Σε αυτό το σημείο παρεμβάλλεται η συνάρτηση **bwtime2typetime** η οποία μετατρέπει την έξοδο scanned-spec της multiresScan σε μία άλλη δομή που καθιστά πιο εύκολη την απεικόνιση σε βίντεο. Πιο συγκεκριμένα, η έξοδος της **bwtime2typetime** είναι μία δομή cell με διαστάσεις $1 \cdot T$, όπου T είναι το πλήθος των χρονικών ευρών. Κάθε στοιχείο της δομής αντιστοιχεί σε ένα χρονικό εύρος και περιέχει μία άλλη δομή cell διαστάσεων $D \cdot F$, όπου D είναι το πλήθος

των ανιχνευτών και F το πλήθος των χρονικών στιγμών στις οποίες έγινε ανάλυση. Τέλος, κάθε τέτοια υποδομή περιέχει ένα “χάρτη” ενεργοποίησης, όπως αναφέρθηκε παραπάνω, με διαστάσεις $(N + 1) \cdot (M + 1)$, όπου N το πλήθος των κεντρικών συχνοτήτων και M το πλήθος των ευρών ζώνης. Η αύξηση κατά 1 των διαστάσεων γίνεται επειδή κατά την απεικόνιση χρησιμοποιείται η συνάρτηση `pcolor` του MATLAB η οποία αγνοεί την τελευταία σειρά και την τελευταία στήλη.

Είσοδοι :

1. **scspec** : Το `scannespec` της `multiresScan`.
2. **scF** : Το `spF` της `multiresScan`.
3. **scT** : Το `spT` της `multiresScan`.

Έξοδοι :

1. **typetime** : Η έξοδος της `bwtime2typetime`, όπως περιγράφηκε προηγουμένως.

Η συνάρτηση που τελικά υλοποιεί την απεικόνιση των αποτελεσμάτων είναι η **playmov**.

Είσοδοι :

1. **x** : Το αρχικό σήμα.
2. **Fs** : Ο ρυθμός δειγματοληψίας του x .
3. **spec** : Το φασματόγραμμά του.
4. **F** : Το διάνυσμα συχνοτήτων του φασματογράμματος.
5. **T** : Το διάνυσμα χρόνου του φασματογράμματος.
6. **bws** : Το διάνυσμα με τα εύρη ζώνης, όπως στην `multiresScan`.
7. **timews** : Το διάνυσμα με τα χρονικά εύρη, όπως στην `multiresScan`.
8. **typetime** : Η έξοδος της `bwtime2typetime`.
9. **scF** : Ένα στοιχείο του `scF` της `bwtime2typetime`, π.χ. το `scF{1,1}`.
10. **scT** : Ένα στοιχείο του `scT` της `bwtime2typetime`, π.χ. το `scT{1,1}`.
11. **movienam** : Ένα αλφαριθμητικό για το όνομα του αρχείου βίντεο που παράγεται στο τέλος.

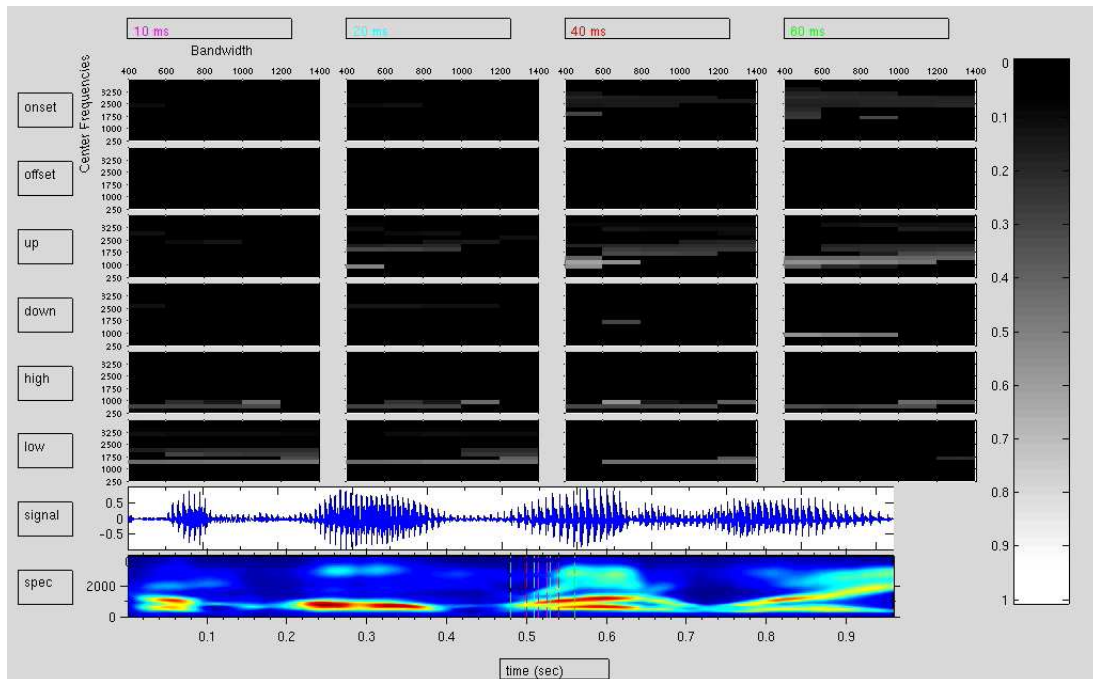
Η συνάρτηση δεν έχει εξόδους. Το αποτέλεσμα αποθηκεύεται υπό τη μορφή ενός αρχείου βίντεο.

Ένα παράδειγμα κλήσης της συνάρτησης είναι :

```
playmov(signal, Fs, Smfcc, Fmfcc, Tmfcc, bws, timews,  
        typetime, scF1, 1, scT1, 1, 'movie')
```

Στο σχήμα 4.11 φαίνεται ένα στιγμιότυπο από το βίντεο που παράγεται από την παραπάνω κλήση.

Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου



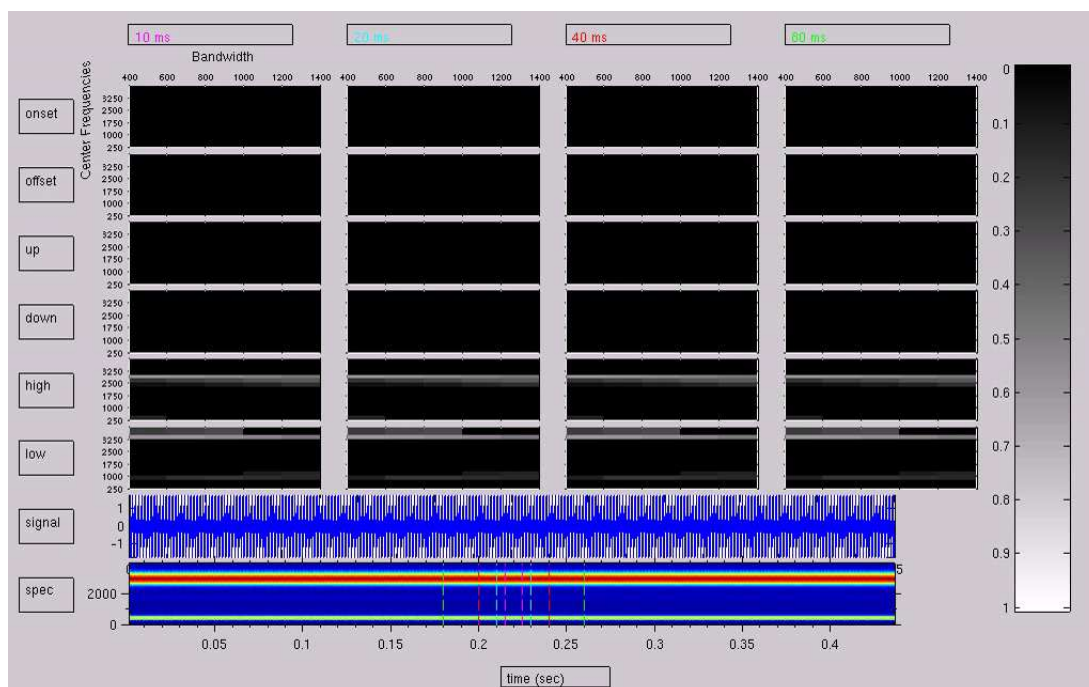
Σχήμα 4.11: Στιγμιότυπο από βίντεο που παράγει η συνάρτηση playmon.

4.3 Έλεγχος του συστήματος

Σε αυτή την ενότητα θα παρουσιαστούν τα αποτελέσματα που δίνει το σύστημα για διάφορες εισόδους, ξεκινώντας από απλά σήματα και φτάνοντας μέχρι κανονικά σήματα ομιλίας.

4.3.1 Απόκριση σε απλούς τόνους

Οι απλοί ημιτονοειδείς τόνοι οι οποίοι εμφανίζονται στο φασματογράμμα ως λεπτές οριζόντιες μπάρες ενέργειας θα πρέπει να ενεργοποιούν κυρίως ή και κατ' αποκλειστικότητα τους ανιχνευτές τύπου low και high. Για να διαπιστωθεί αν όντως ισχύει αυτό, πραγματοποιείται μία δοκιμή με ένα σήμα δύο τόνων, 500 και 3000 Hz με ρυθμό δειγματοληψίας 8000 Hz. Το αποτέλεσμα φαίνεται στο σχήμα 4.12. Η απόκριση είναι ίδια και για όλα τα υπόλοιπα στιγμιότυπα του βίντεο. Η εντονότερη μπάρα ενέργειας στα 3000 Hz προκύπτει λόγω της χρήσης mfcc φασματογράμματος. Όπως φαίνεται από το σχήμα, οι αποκρίσεις είναι οι αναμενόμενες. Ο τόνος των 500 Hz ενεργοποιεί τους high ανιχνευτές της κεντρικής συχνότητας των 250 Hz μόνο για εύρος ζώνης 400 Hz γιατί στα υπόλοιπα εύρη ζώνης το παράθυρο βγαίνει εκτός ορίων του φασματογράμματος με αποτέλεσμα να μην υπολογίζεται κάποια απόκριση για αυτές τις περιπτώσεις. Επίσης, φαίνεται πως οι ανιχνευτές γύρω από τα 3000 Hz αντιδρούν εντονότερα εφόσον η αντίστοιχη μπάρα ενέργειας είναι πιο έντονη (λόγω mfcc). Μπορεί ακόμα να παρατηρηθεί ότι μεγαλύτερος αριθμός ανιχνευτών περί τα 3000 Hz ενεργοποιούνται αφού η μπάρα ενέργειας σε αυτή την περιοχή έχει μεγαλύτερο πάχος - εύρος ζώνης.



Σχήμα 4.12: Απόκριση του συστήματος σε σήμα δύο τόνων 500 και 3000 Hz

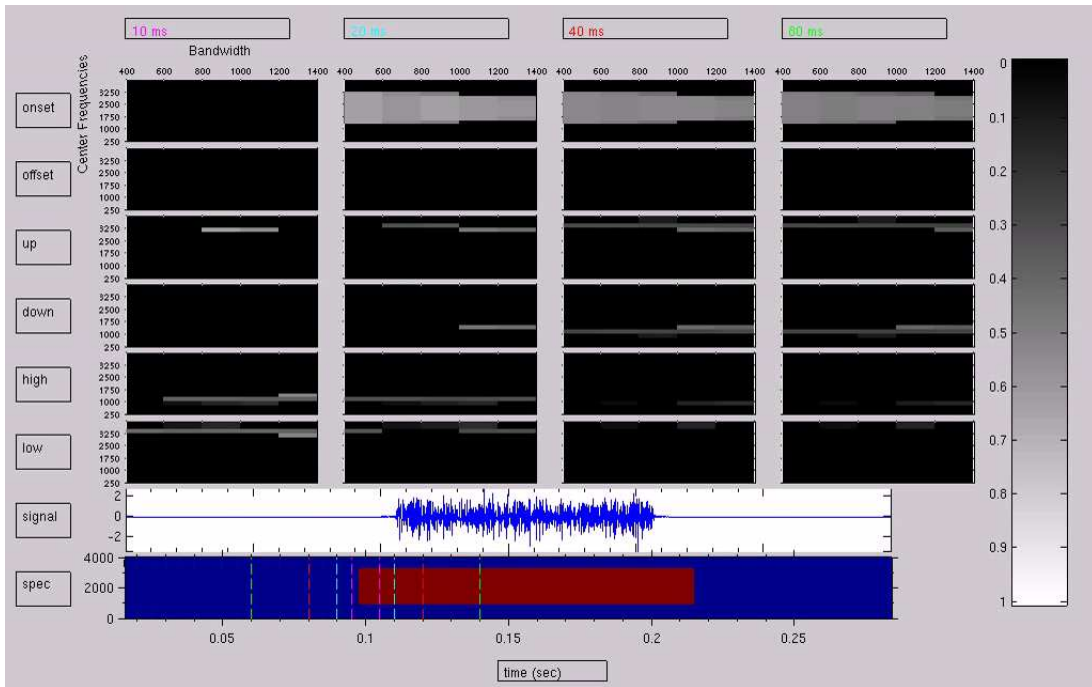
4.3.2 Απόκριση σε ευρυζωνικές μπάρες “θορύβου”

Δραστηριότητα που κατανέμεται σε ένα μεγάλο εύρος συχνοτήτων και που προσομοιώνει εκρήξεις θορύβου που παρουσιάζονται συχνά σε πραγματικά σήματα θορύβου θα πρέπει να ενεργοποιεί κυρίως του ανιχνευτές onset και offset στην αρχή και στο τέλος της δραστηριότητας αντίστοιχα. Επίσης, αναμένεται και απόκριση των low και high ανιχνευτών στις υψηλές και στις χαμηλές συχνότητες της μπάρας αντίστοιχα. Στα σχήματα 4.13, 4.14 και 4.15 φαίνεται η απόκριση του συστήματος σε έναν τέτοιο τεχνητό θόρυβο που καταλαμβάνει τις συχνότητες από 1000 μέχρι 3250 Hz περίπου και εκτείνεται από τα 0.1 μέχρι τα 0.215 δευτερόλεπτα περίπου. Στο πρώτο στιγμιότυπο τα παράθυρα ανάλυσης βρίσκονται γύρω από την αρχή της δραστηριότητας, στο δεύτερο όλα είναι εντός της και στο τρίτο γύρω από το τέλος της.

Όπως ήταν αναμενόμενο, στο πρώτο στιγμιότυπο (σχ. 4.13) εμφανίζεται έντονη ενεργοποίηση των onset ανιχνευτών για τις συχνότητες που καταλαμβάνει η μπάρα, με την εξαίρεση του παραθύρου των 10 ms το οποίο όμως έχει ήδη “βυθιστεί” κατά το μεγαλύτερο μέρος του εντός του θορύβου. Για τα μικρά παράθυρα των 10 και 20 ms παρατηρείται απόκριση και των low και high ανιχνευτών εφόσον σε αυτά ένα μεγάλο τμήμα τους καταλαμβάνεται ήδη από δραστηριότητα. Κάτι που εκ πρώτης όψεως είναι μη αναμενόμενο αφορά στην απόκριση των up και down ανιχνευτών, κυρίως στα παράθυρα με μεγαλύτερο χρονικό εύρος. Αυτό οφείλεται στο παράθυρο που βλέπουν οι ανιχνευτές και στο πώς αυτό ερμηνεύεται από αυτούς. Για παράδειγμα, στην περίπτωση του παραθύρου των 0.08 ms με κεντρική συχνότητα στα 3250 Hz, όπου έχουμε απόκριση up ανιχνευτών, το παράθυρο

Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου

είναι κενό εκτός από το κάτω, δεξί τεταρτημόριό του. Ένα παρόμοιο παράθυρο θα έβλεπαν οι ανιχνευτές και στην περίπτωση που υπήρχε ένα ανόδικο fm sweep κοντά στην περιοχή αυτού του τεταρτημορίου. Η ενεργοποίηση των up και down ανιχνευτών στις “γωνίες” της περιοχής δραστηριότητας δεν είναι απαραίτητως ανεπιθύμητη, εφόσον με αυτόν τον τρόπο είναι δυνατό να εντοπιστούν πιο ευκρινώς τα όρια της.



Σχήμα 4.13: Απόκριση του συστήματος κατά την έναρξη μίας έκρηξης θορύβου

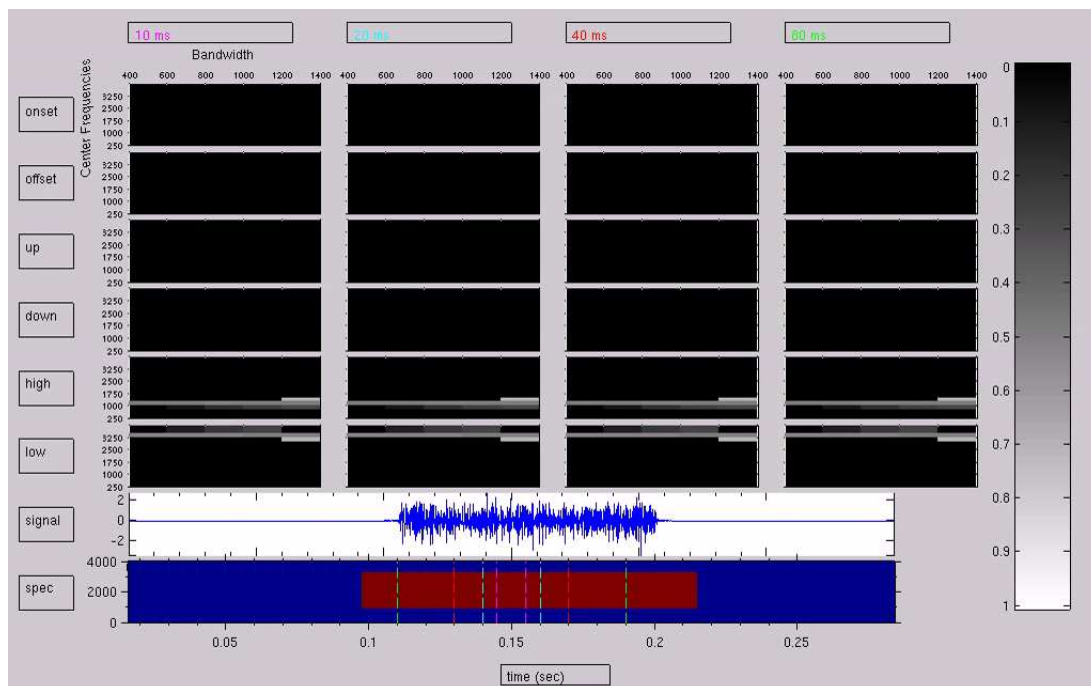
Όταν τα παράθυρα ανάλυσης βρίσκονται εντός της περιοχής δραστηριότητας, όπως στο δεύτερο στιγμιότυπο (σχ. 4.14), η μόνη ενεργοποίηση του παρατηρείται είναι στους low και high ανιχνευτές στα άνω και κάτω όρια της περιοχής, όπως είναι αναμενόμενο.

Τέλος, όταν τα παράθυρα ανάλυσης εστιάζουν γύρω από το τέλος του θορύβου (σχ. 4.15) οι offset ανιχνευτές αντιδρούν εντονότερα. Κατά τ' άλλα, οι high και low αντιδρούν όπως στο πρώτο στιγμιότυπο ενώ οι up και down με έναν συμμετρικό τρόπο σε σχέση με το πρώτο στιγμιότυπο.

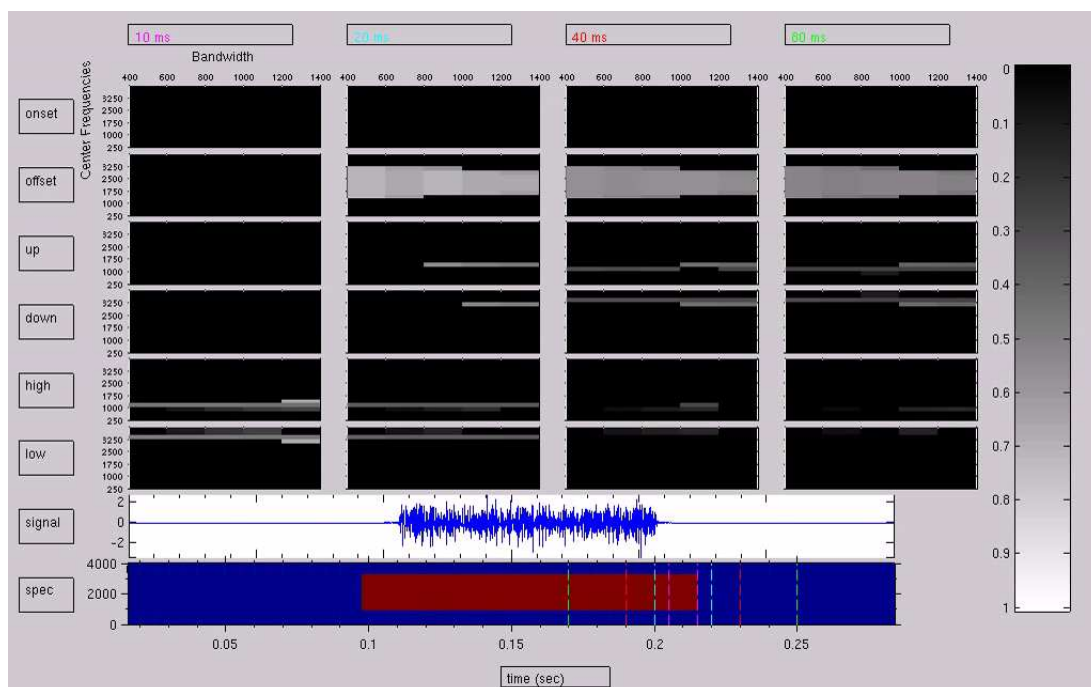
4.3.3 Fm sweep

Ένα ακόμα χαρακτηριστικό που μας ενδιαφέρει είναι και οι ανοδικές και καθοδικές μεταβάσεις που εμφανίζονται στα φασματογράμματα ομιλίας με τη μορφή κεκλιμένων μπαρών ενέργειας. Οι ανιχνευτές up και down θα πρέπει να είναι ικανοί να εντοπίζουν τέτοιες μεταβάσεις. Στη συνέχεια, παρουσιάζεται η απόκριση του συστήματος σε ιδανικά ανοδικά και καθοδικά fm sweep.

Στο σχήμα 4.16 έχουμε ένα καθοδικό fm sweep που σαρώνει τις συχνότητες από 2500 μέχρι 1500 Hz εντός περίπου 100 ms. Η παρουσία της κατάβασης διακρίνεται έντονα



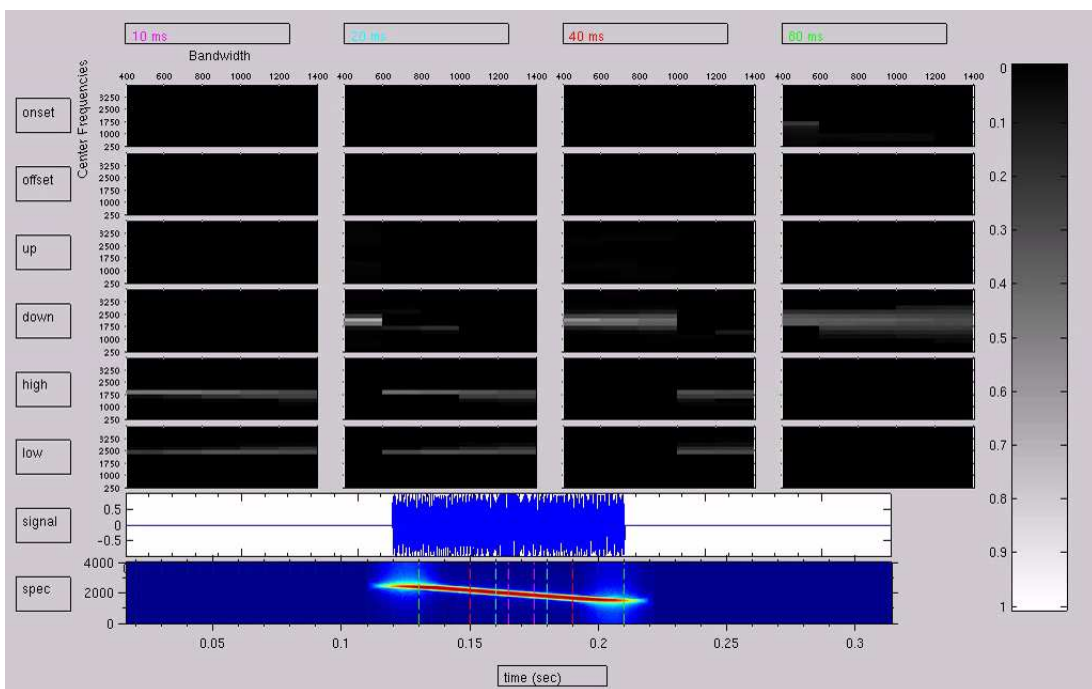
Σχήμα 4.14: Απόκριση του συστήματος εντός μίας έκρηξης θορύβου



Σχήμα 4.15: Απόκριση του συστήματος κατά την παύση μίας έκρηξης θορύβου

Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου

από την ενεργοποίηση των down ανιχνευτών για τις κατάλληλες συχνότητες, κυρίως για μεγάλα σε χρονική διάρκεια παράθυρα των 40 και 80 ms, αφού το sweep έχει μεγάλη διάρκεια. Αντιθέτως, στα παράθυρα των 10 και 20 ms κυριαρχεί η απόκριση των high και low ανιχνευτών. Επειδή η κατάβαση δεν είναι ταχεία και η κλίση της σχετικά μικρή, σε αυτά τα μικρά παράθυρα, μοιάζει περισσότερο με οριζόντια μπάρα, ειδικά όταν το εύρος ζώνης του παραθύρου αυξάνεται. Για παράδειγμα, στα 40 ms, από ένα εύρος ζώνης και μετά (1000 Hz) η απόκριση μετατοπίζεται από τους down ανιχνευτές στους low/high. Στα 20 ms, αυτή η μετατόπιση ξεκινάει από ακόμα μικρότερο εύρος ζώνης (600 Hz) ενώ στα 10 ms οι ανιχνευτές αντιλαμβάνονται την μπάρα μόνο ως οριζόντια. Γενικότερα, παρατηρείται μία συμπληρωματικότητα μεταξύ των αποκρίσεων των down ανιχνευτών και των low/high.



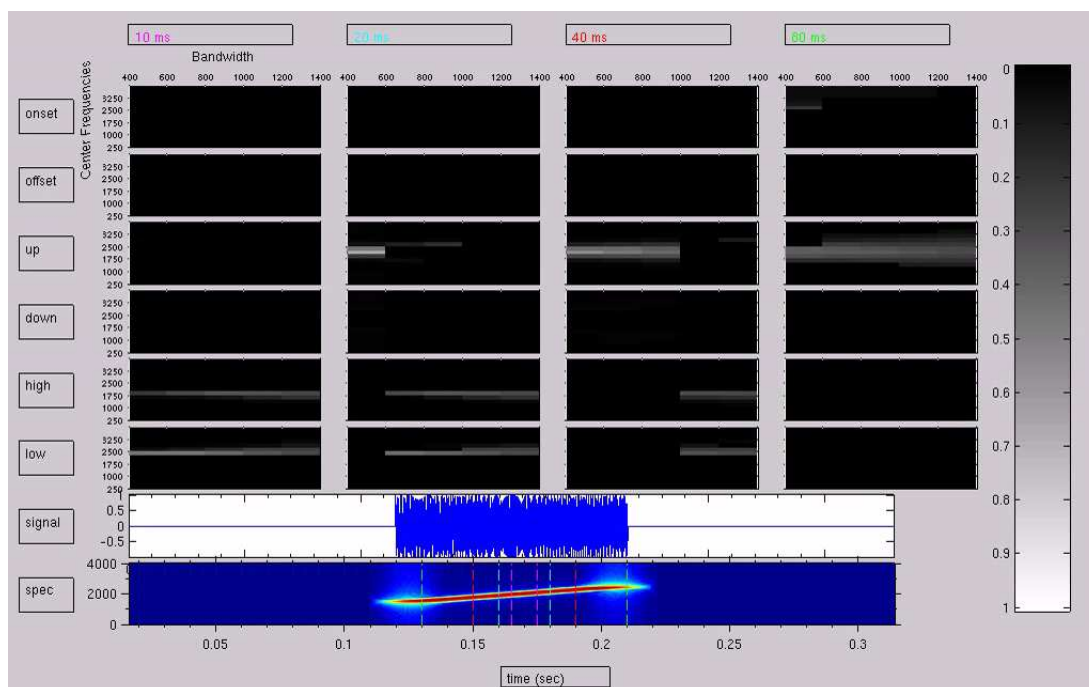
Σχήμα 4.16: Απόκριση του συστήματος σε καθοδικό fm sweep

Στο σχήμα 4.17 έχουμε ένα fm sweep με τα ίδια χαρακτηριστικά όπως προηγουμένως αλλά είναι ανοδικό. Η απόκριση του συστήματος είναι παρόμοια όπως και στην περίπτωση της κατάβασης με τη διαφορά φυσικά ότι εδώ έχουμε ενεργοποίηση των up ανιχνευτών. Επίσης, παρατηρείται και πάλι η συμπληρωματικότητα των αποκρίσεων μεταξύ των down ανιχνευτών και των low/high.

4.3.4 Πραγματικά σήματα ομιλίας

Σε αυτή την ενότητα εξετάζεται η απόκριση του συστήματος σε πραγματικά σήματα ομιλίας.

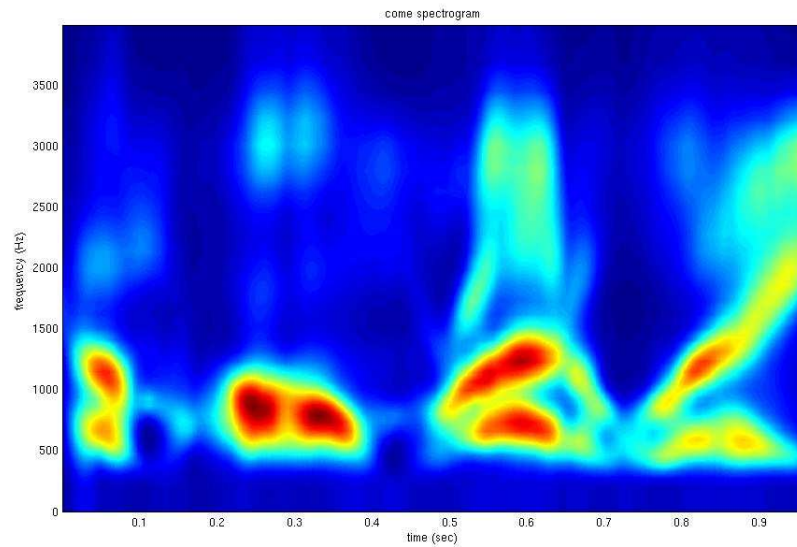
Το πρώτο σήμα που εξετάζεται είναι η φράση “come home right away” που έχει δειγματοληπτηθεί στα 8000 Hz. Το φασματόγραμμα της φαίνεται στο σχήμα 4.18 ενώ μερικά



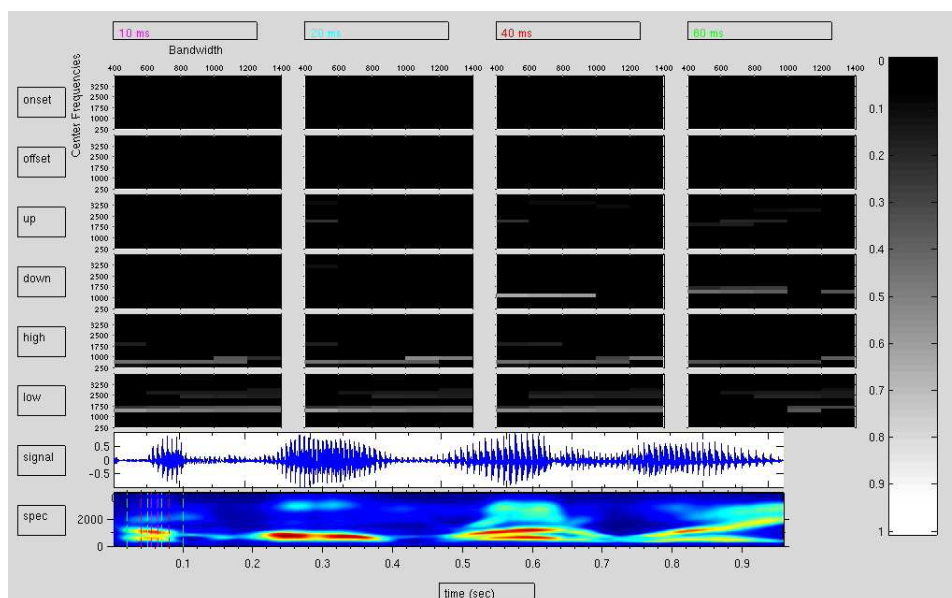
Σχήμα 4.17: Απόκριση του συστήματος σε ανόδικο fm sweep

στιγμιότυπα από την απόκριση του συστήματος παρουσιάζονται στα σχήματα 4.19 ως 4.25. Στο πρώτο στιγμιότυπο παρατηρούμε ότι το σύστημα εντοπίζει την κατάβαση που υπάρχει κοντά στα 1000 Hz αλλά ασθενώς και την αμυδρή ανάβαση στα 2000 Hz. Η κατάβαση στα 1000 Hz γίνεται αντιληπτή ως οριζόντια μπάρα για κάποιες συχνότητες στο άνω όριο της. Στο δεύτερο στιγμιότυπο, παρόλο που δεν υπάρχουν καθαρές αναβάσεις ή καταβάσεις, μπορούμε να δούμε ότι η περιοχή δραστηριότητας έχει ένα “τριγωνικό” σχήμα στην αρχή της, γεγονός που εντοπίζεται από τους ανιχνευτές καθώς ενεργοποιούνται οι up για συχνότητες λίγο πάνω από τα 1000 Hz και οι down για λίγο χαμηλότερες συχνότητες. Στο τρίτο στιγμιότυπο έχουμε κυρίως οριζόντιες μπάρες που εντοπίζονται. Στο τέταρτο, βλέπουμε και πάλι ότι εντοπίζεται η ψευδοκατάβαση που υπάρχει λόγω του σχήματος της περιοχής δραστηριότητας. Στο πέμπτο στιγμιότυπο, υπάρχει ασθενής απόκριση των onset ανιχνευτών στην ευρυζωνική μπάρα ενέργειας στην περιοχή 2500-3000 Hz. Επίσης, γίνονται αντιληπτές και οι δύο αναβάσεις στα 1500-2000 και 1000-1500 Hz. Έχει ενδιαφέρον να παρατηρήσουμε ότι η πρώτη ανάβαση που έχει μεγαλύτερη κλίση και είναι ταχύτερη εντοπίζεται κυρίως από τα παράθυρα των 20 και 40 ms ενώ η δεύτερη που είναι πιο αργή από τα παράθυρα των 40 και 80 ms. Τέλος, η οριζόντια μπάρα στα 500 περίπου Hz ενεργοποιεί και αυτή τους αντίστοιχους high ανιχνευτές. Στο έκτο στιγμιότυπο, το μεγαλύτερο ενδιαφέρον έχει η κατάβαση στα 500-1000 Hz στην οποία ανταποκρίνονται οι down ανιχνευτές. Στο έβδομο στιγμιότυπο, υπάρχει μία έντονη ανάβαση που εντοπίζεται. Επίσης, εμφανίζεται και η συμπληρωματικότητα που αναφέρθηκε παραπάνω μεταξύ up/down και low/high ανιχνευτών. Στα μικρά χρονικά παράθυρα, όπου η ανάβαση μπορεί να μην είναι τόσο καθαρή, οι ανιχνευτές την αντιλαμβάνονται συχνά ως οριζόντια μπάρα.

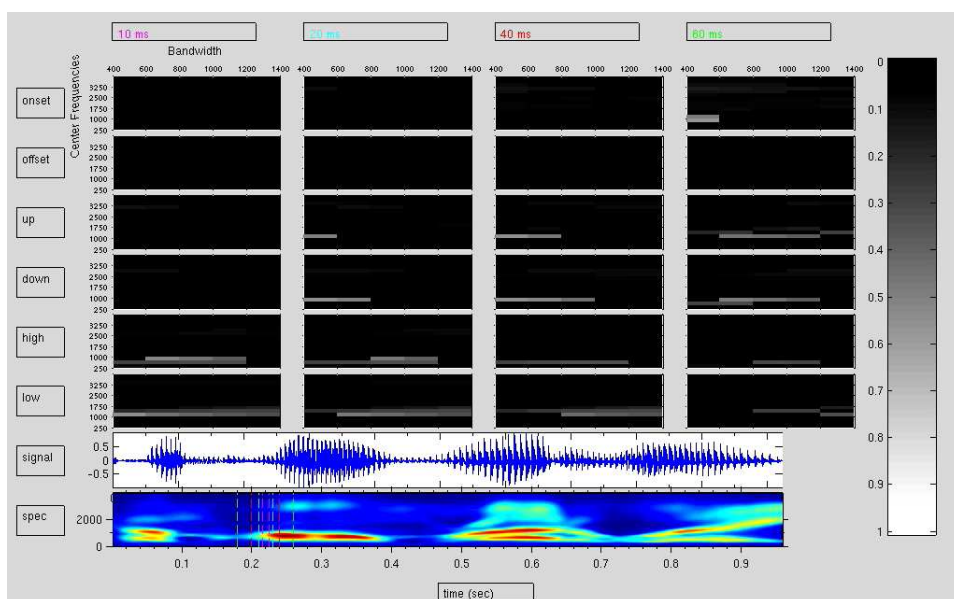
Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου



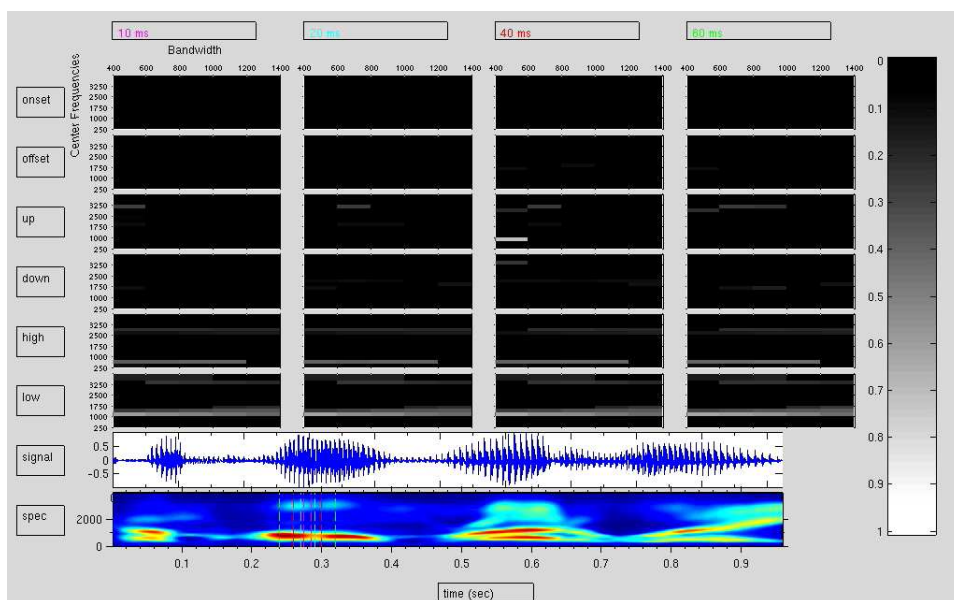
Σχήμα 4.18: Φασματόγραμμα της φράσης “come home right away”.



Σχήμα 4.19: Στιγμιότυπο 1 από την απόκριση του συστήματος για τη φράση “come home right away”.

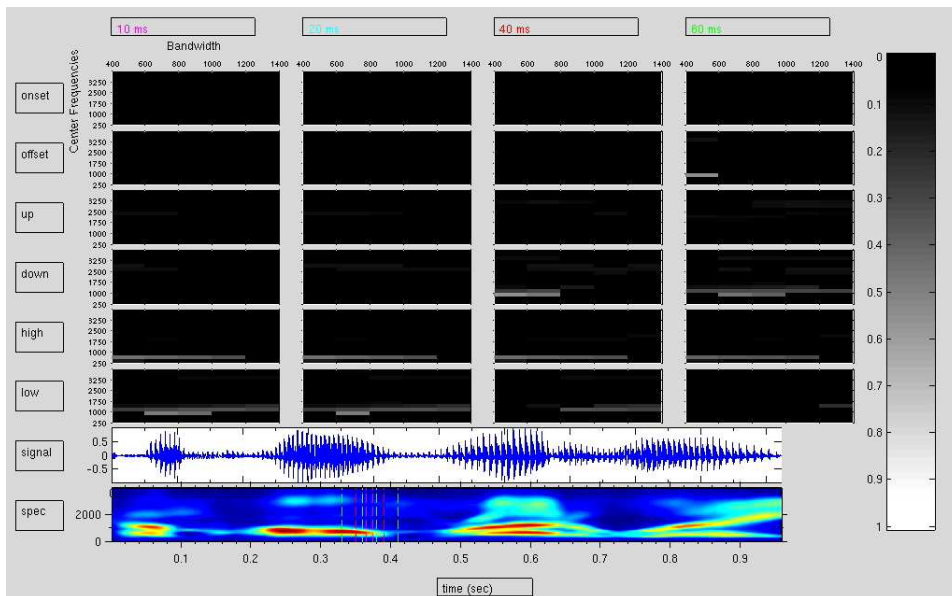


Σχήμα 4.20: Στιγμιότυπο 2 από την απόκριση του συστήματος για τη φράση “come home right away”.

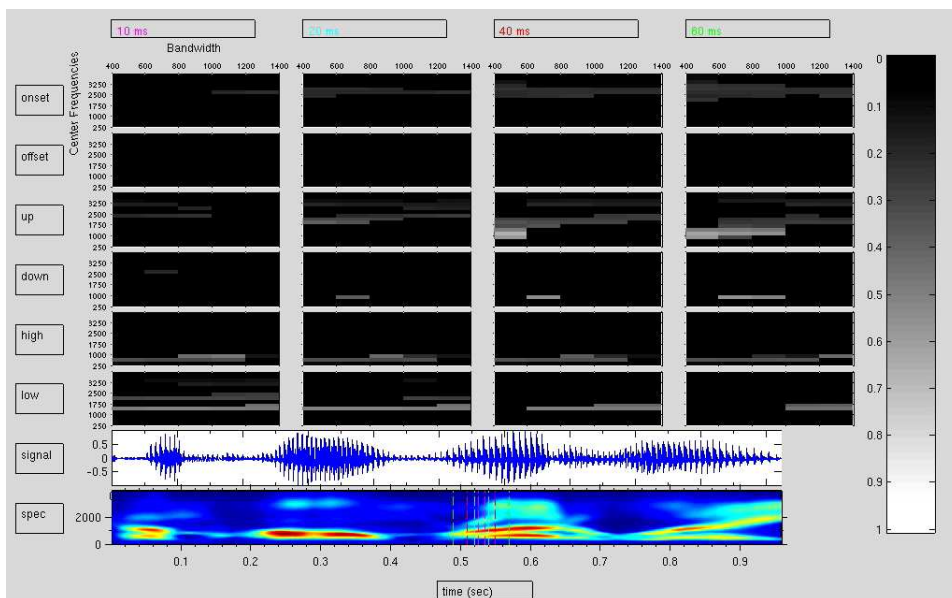


Σχήμα 4.21: Στιγμιότυπο 3 από την απόκριση του συστήματος για τη φράση “come home right away”.

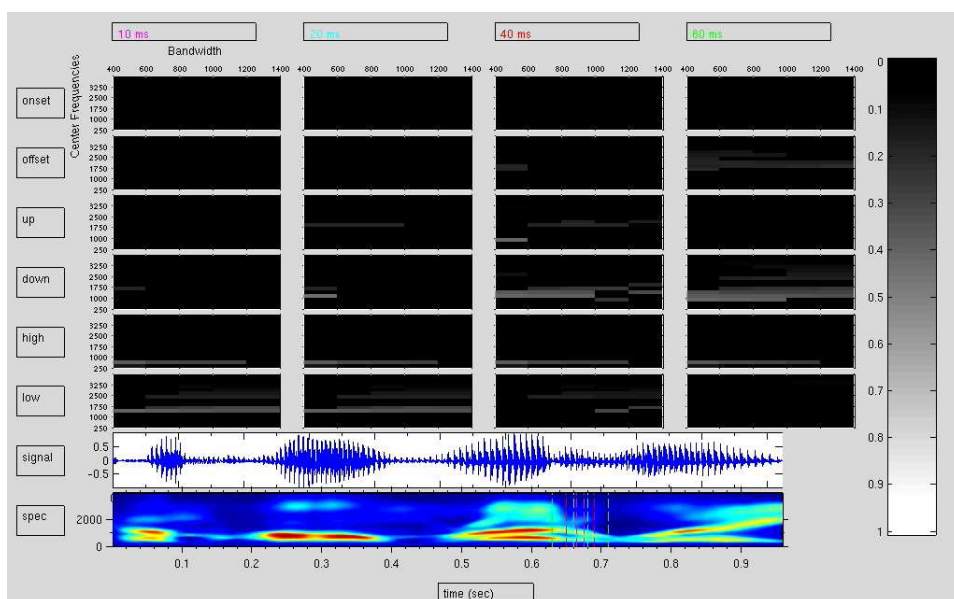
Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου



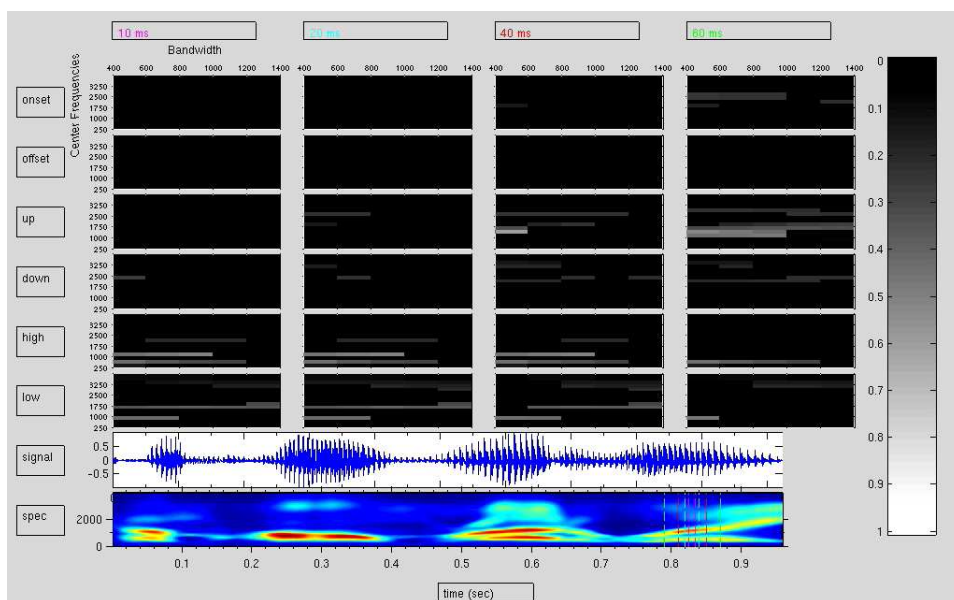
Σχήμα 4.22: Στιγμιότυπο 4 από την απόκριση του συστήματος για τη φράση “come home right away”.



Σχήμα 4.23: Στιγμιότυπο 5 από την απόκριση του συστήματος για τη φράση “come home right away”.



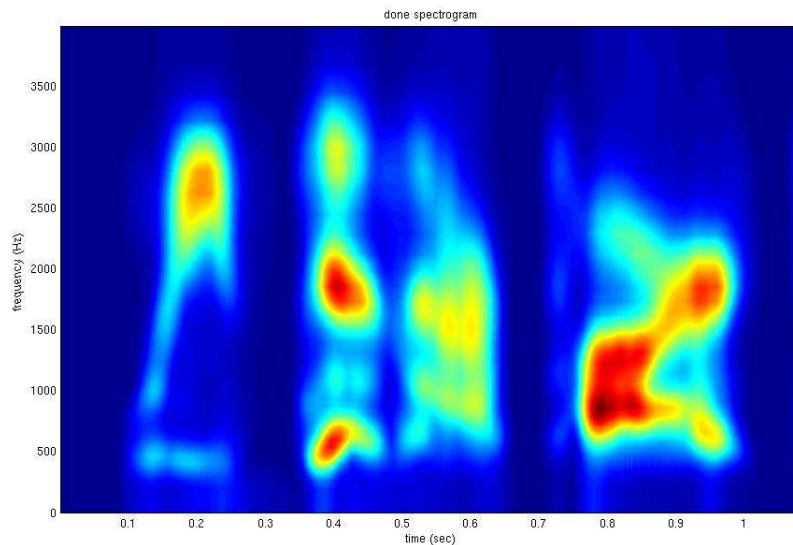
Σχήμα 4.24: Στιγμιότυπο 6 από την απόκριση του συστήματος για τη φράση “come home right away”.



Σχήμα 4.25: Στιγμιότυπο 7 από την απόκριση του συστήματος για τη φράση “come home right away”.

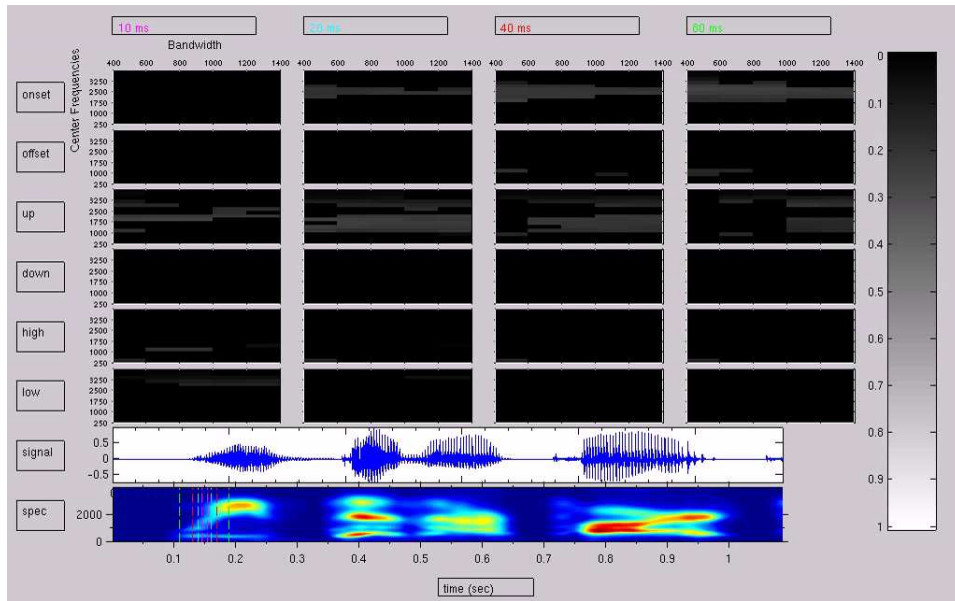
Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου

Το επόμενο σήμα ομιλίας που εξετάζουμε εκφέρεται αυτή τη φορά από γυναίκα και είναι η φράση “we’ve done our part”. Όπως και προηγουμένως, παρουσιάζονται πέντε στιγμιότυπα από την επεξεργασία αυτής της φράσης (σχήματα 4.27 ως 4.31). Στο πρώτο στιγμιότυπο, έχουμε μία σχετικά απότομη ανάβαση η οποία συλλαμβάνεται από τα παράθυρα των 10, 20 και 40 ms ενώ φαίνεται και η έναρξη της δραστηριότητας στα 2500-3000 Hz από τους onset ανιχνευτές. Λίγο παρακάτω εμφανίζεται αμυδρά και η οριζόντια μπάρα στα 500 Hz μέσω των low ανιχνευτών. Στο δεύτερο στιγμιότυπο, παρατηρούμε ότι η ανάβαση στην περιοχή των 500 Hz γίνεται αντιληπτή μόνο από το παράθυρο των 80 ms, όσο περίπου και η διάρκεια της ανάβασης, ενώ στα άλλα παράθυρα φαίνεται ως οριζόντια μπάρα. Το ίδιο συμβαίνει και με την ιδιόμορφη κατάβαση που υπάρχει κοντά στα 2000 Hz. Στο τρίτο στιγμιότυπο είναι έντονη η απόκριση των offset ανιχνευτών ενώ ταυτόχρονα παρατηρείται και δραστηριότητα των down στην περιοχή κυρίως άνω των 2000 Hz για τα μεγάλα χρονικά παράθυρα και των low για τα μικρότερα. Στο τέταρτο στιγμιότυπο, η κατάβαση που υπάρχει γίνεται μόλις αισθητή στο παράθυρο των 80 ms ενώ κυριαρχούν οι low και down αποκρίσεις που αντιστοιχούν στην έκρηξη θορύβου στις χαμηλότερες συχνότητες. Στο τελευταίο στιγμιότυπο συλλαμβάνονται η κατάβαση και η ανάβαση που υπάρχουν αλλά ταυτόχρονα εμφανίζεται και δραστηριότητα low και high για τη συγκρέτωση ενέργειας στην περιοχή των 1500-2000 Hz.

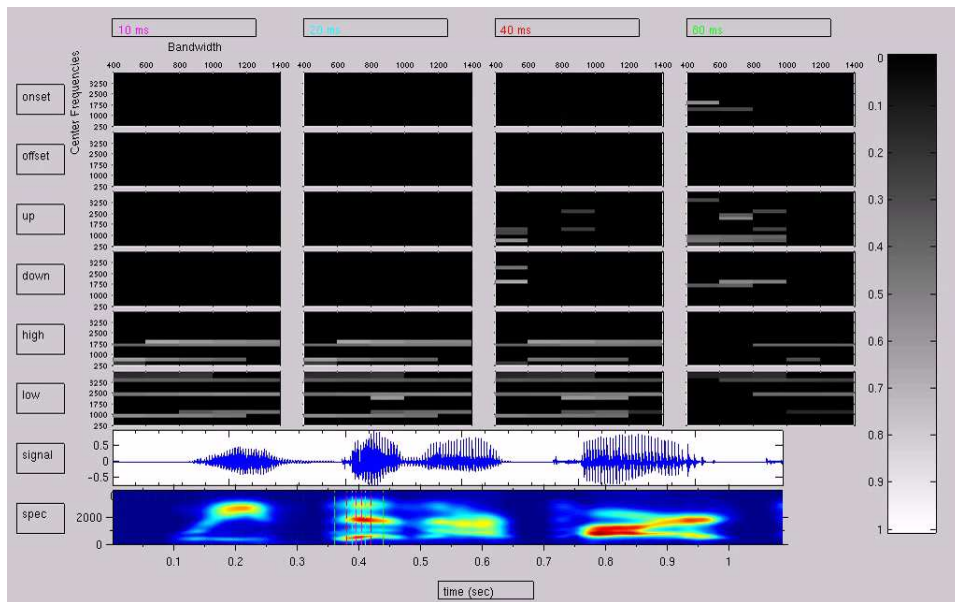


Σχήμα 4.26: Φασματόγραμμα της φράσης “we’ve done our part”.

Τέλος, δοκιμάζουμε το σύστημα με τη φράση “πάτησε με το ποντίκι πάνω στον Περίφημο για να ξεκινήσει” που εκφέρεται και πάλι από γυναίκα και έχει δειγματοληπτηθεί στα 22050 Hz. Στη συνέχεια θα παρουσιαστούν στιγμιότυπα (σχήματα 4.33 ως 4.39) για ένα τμήμα αυτής της φράσης από το δεύτερο δευτερόλεπτό της (1 ως 2 δευτ.) που αντιστοιχεί περίπου στο “πάνω στον Περίφημο”. Στο πρώτο στιγμιότυπο έχουμε δύο ανοδικές μεταβάσεις γύρω από τα 1000 Hz. Εντοπίζονται και οι δύο, με τη μεγαλύτερη ευκρίνεια

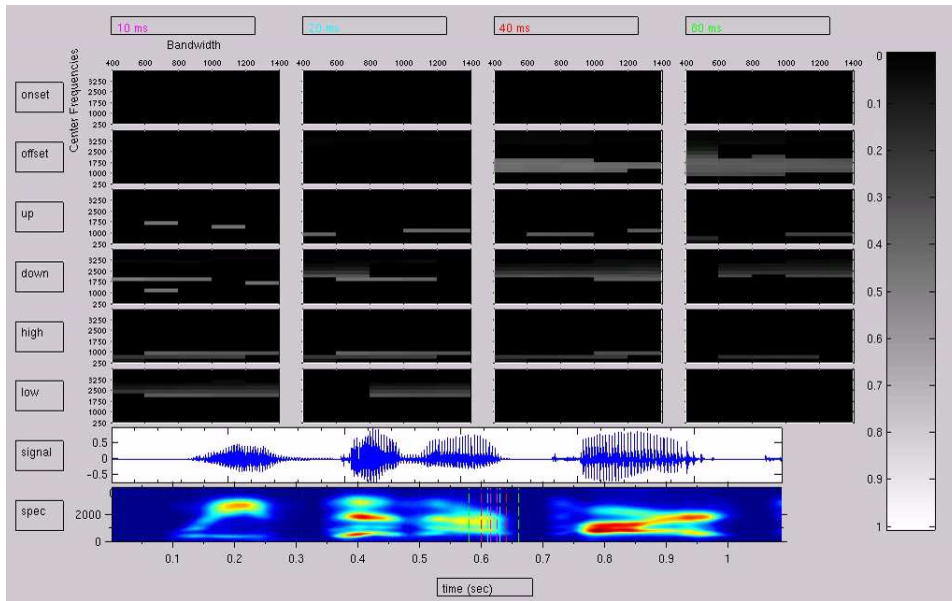


Σχήμα 4.27: Στιγμιότυπο 1 από την απόκριση του συστήματος για τη φράση “we’ve done our part”.

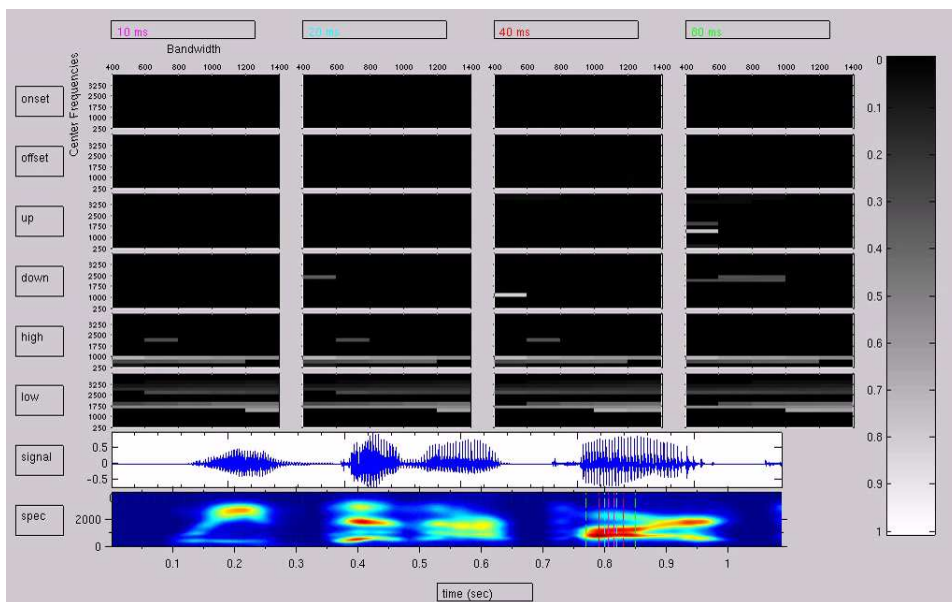


Σχήμα 4.28: Στιγμιότυπο 2 από την απόκριση του συστήματος για τη φράση “we’ve done our part”.

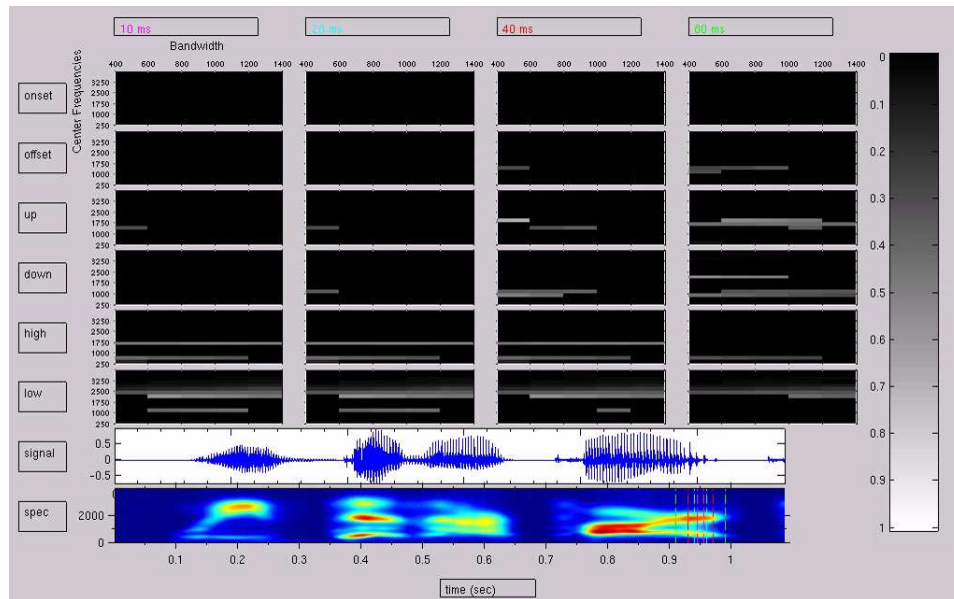
Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου



Σχήμα 4.29: Στιγμιότυπο 3 από την απόκριση του συστήματος για τη φράση “we’ve done our part”.



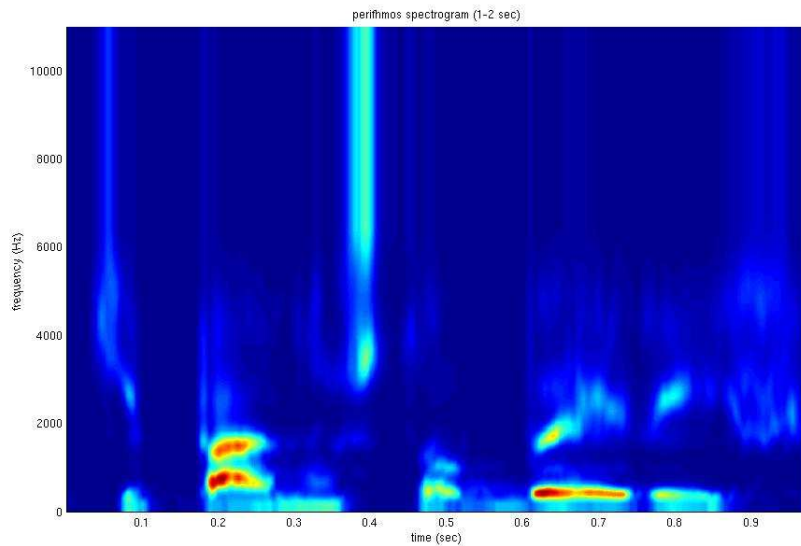
Σχήμα 4.30: Στιγμιότυπο 4 από την απόκριση του συστήματος για τη φράση “we’ve done our part”.



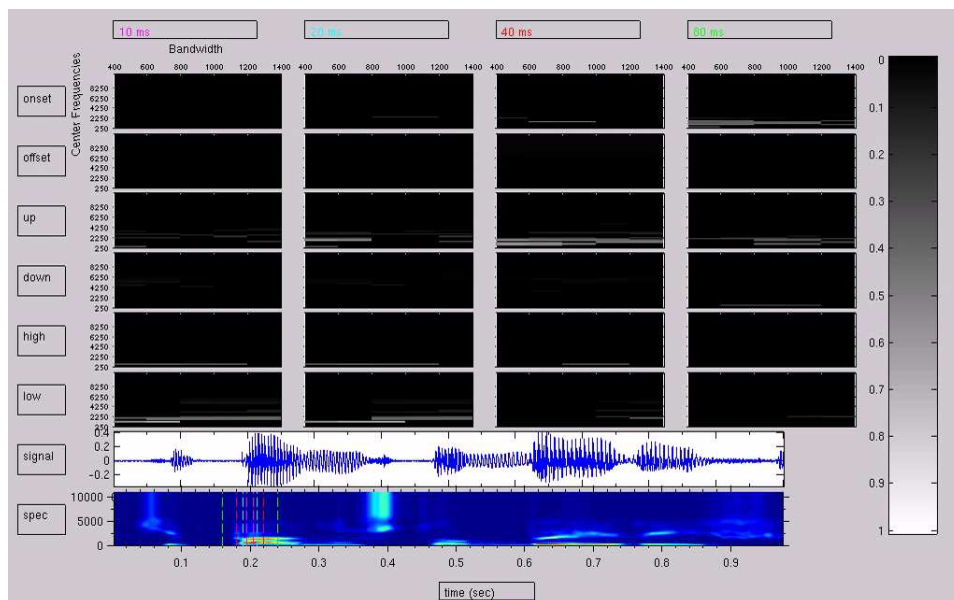
Σχήμα 4.31: Στιγμιότυπο 5 από την απόκριση του συστήματος για τη φράση “we’ve done our part”.

στο παράθυρο των 40 ms. Στο επόμενο στιγμιότυπο, η υψηλότερη από τις δύο μεταβάσεις συνεχίζει να είναι ανοδική αλλά η χαμηλότερη έχει μετατραπεί σε καθοδική. Αυτή η μεταβολή αντανακλάται και στους ανιχνευτές. Αυτοί των χαμηλών συχνοτήτων μετατοπίζουν τη δραστηριότητά τους από τους up στους down. Στο τρίτο στιγμιότυπο, τα παράθυρα ανάλυσης αρχίζουν να εισέρχονται στην έκρηξη θορύβου που υπάρχει στην περιοχή άνω των 5000 Hz. Η είσοδος αυτή σηματοδοτείται από την ενεργοποίηση των onset ανιχνευτών. Στο σημείο που υπάρχει αυτή η έκρηξη θορύβου, σε χαμηλότερες συχνότητες, κοντά στα 4000 Hz, υπάρχει και μία σχετικά απότομη ανάβαση. Όπως φαίνεται από το τέταρτο στιγμιότυπο, αυτή η ανάβαση συλλαμβάνεται από τα παράθυρα των 10 και 20 ms, μία από τις λίγες περιπτώσεις όπου το παράθυρο των 10 ms αντιλαμβάνεται μία μετάβαση ως τέτοια και όχι ως οριζόντια μπάρα. Ταυτόχρονα, υπάρχει και απόκριση των down ανιχνευτών σε υψηλότερες συχνότητες η οποία οφείλεται στο σχήμα που έχει το κάτω όριο της έκρηξης θορύβου (εμφανίζει όντως καθοδική τάση) που βρίσκεται από πάνω. Στο πέμπτο στιγμιότυπο, το σύστημα καταφέρνει να εντοπίσει την μικρή κατάβαση που υπάρχει στις πολύ χαμηλές συχνότητες ενώ στο έκτο η σαφώς εμφανέστερη ανάβαση ενεργοποιεί τους up ανιχνευτές. Επίσης, η οριζόντια μπάρα που υπάρχει χαμηλότερα εντοπίζεται και αυτή, κυρίως από τους low ανιχνευτές. Το τελευταίο στιγμιότυπο είναι σχεδόν ίδιο με το έκτο με τη διαφορά ότι είναι χαμηλότερης ενέργειας. Επομένως και η απόκριση των ανιχνευτών παρουσιάζεται ασθενέστερη.

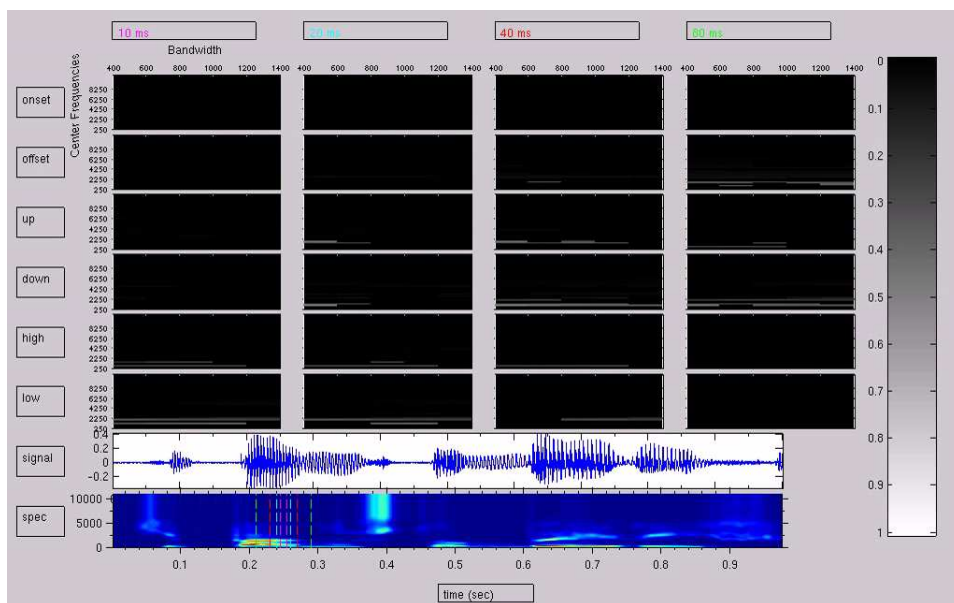
Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου



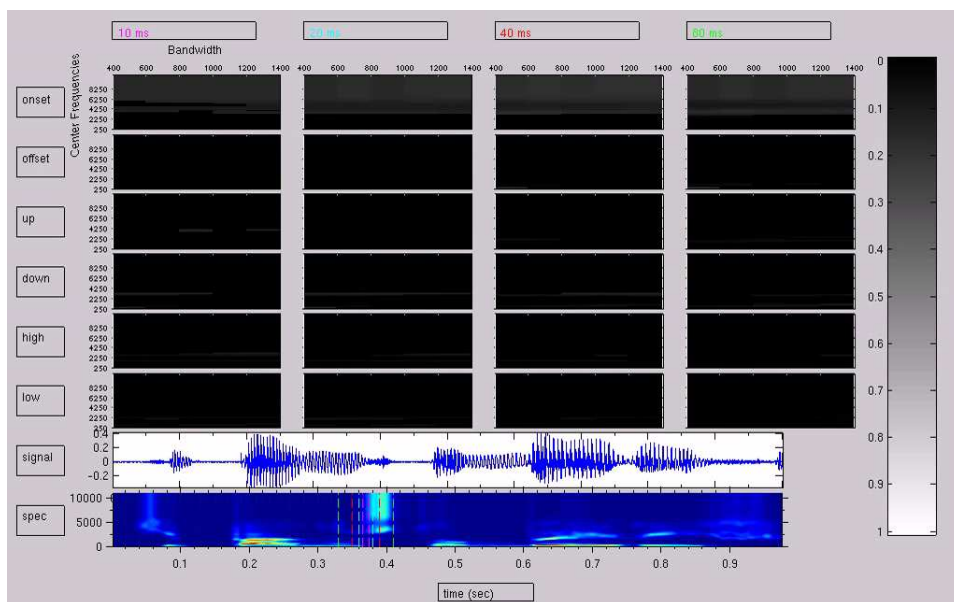
Σχήμα 4.32: Φασματογράμμα της φράσης “πάνω στον Περίφημο”.



Σχήμα 4.33: Στιγμιότυπο 1 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.

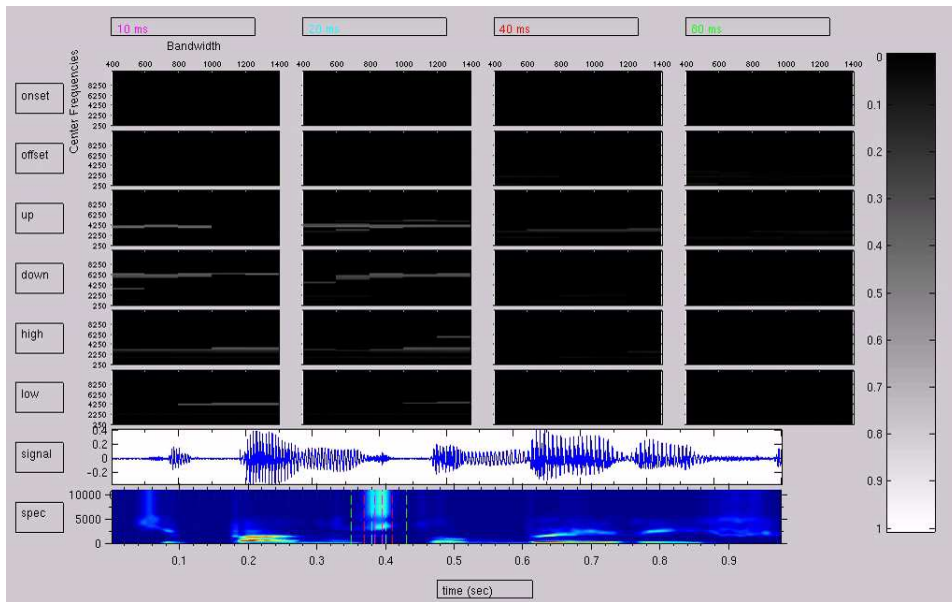


Σχήμα 4.34: Στιγμιότυπο 2 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.

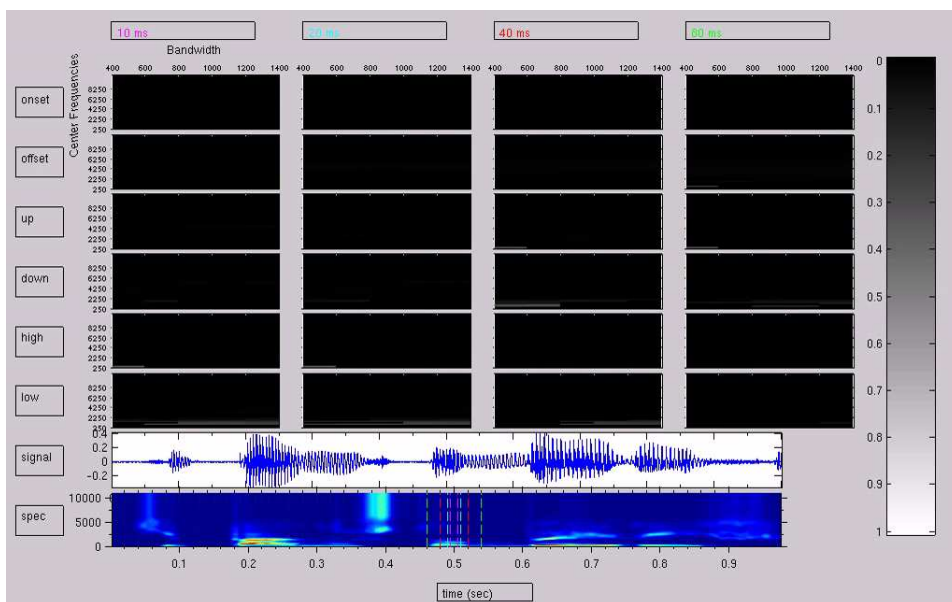


Σχήμα 4.35: Στιγμιότυπο 3 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.

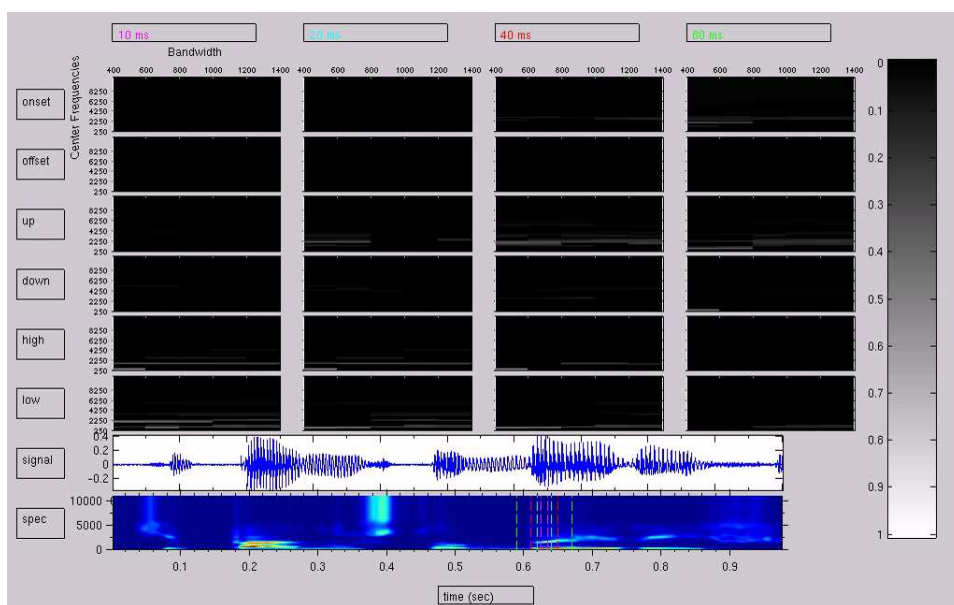
Κεφάλαιο 4. Μία αναπαράσταση κατάλληλη για είσοδο σε ένα γνωσιακό μοντέλο αντίληψης του λόγου



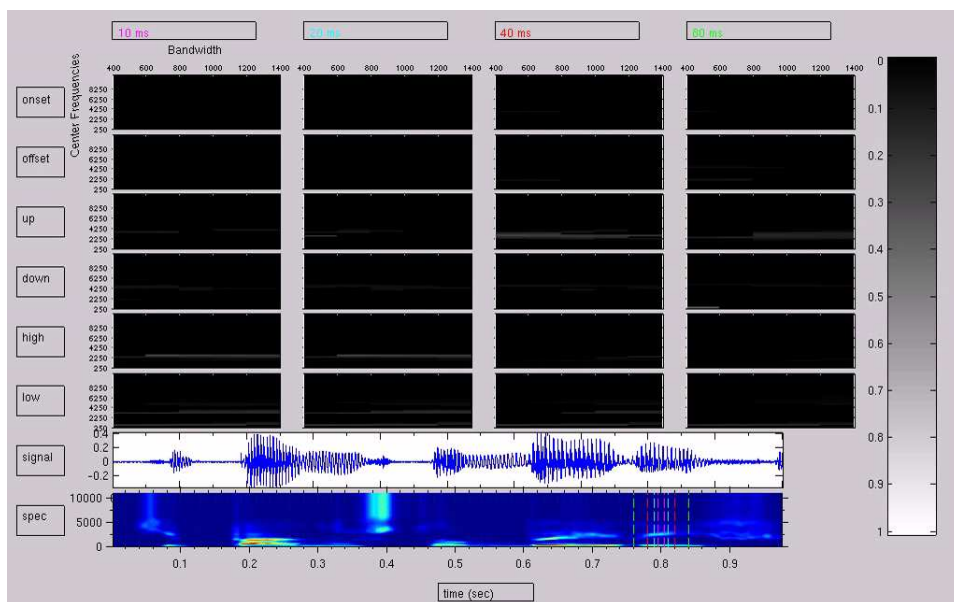
Σχήμα 4.36: Στιγμιότυπο 4 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.



Σχήμα 4.37: Στιγμιότυπο 5 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.



Σχήμα 4.38: Στιγμιότυπο 6 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.



Σχήμα 4.39: Στιγμιότυπο 7 από την απόκριση του συστήματος για τη φράση “πάνω στον Περίφημο”.

Κεφάλαιο 5

Επίλογος

5.1 Σύνοψη και συμπεράσματα

Η μελέτη του ακουστικού νευρικού συστήματος σε επίπεδο μεμονωμένων νευρώνων ή ομάδων νευρώνων έχει αρχίσει να αποκαλύπτει σιγά σιγά το βαθμό της πολυπλοκότητας του. Διάφορα στάδια ανάλυσης των ηχητικών σημάτων συμμετέχουν στη διαδικασία αποκωδικοποίησης τους, αποτελούμενα από νευρώνες διαφόρων ειδών και συμμετέχοντας σε ποικίλες μεταξύ τους αλληλεπιδράσεις. Κάποιες απλοϊκές αντιλήψεις του παρελθόντος, όπως ότι το ακουστικό σύστημα ενδεχομένως να είναι ένας απλός, γραμμικός αναλυτής Fourier, φαίνεται ότι είναι πλέον σε μεγάλο βαθμό ανεπαρκείς για να ερμηνεύσουν τα δεδομένα που προκύπτουν από νευροφυσιολογικά πειράματα τα τελευταία χρόνια. Στα προηγούμενα κεφάλαια έγινε μία επισκόπηση της μέχρι τώρα έρευνας στο πεδίο της επεξεργασίας των ηχητικών σημάτων από το ακουστικό νευρικό σύστημα. Η μετατόπιση της προσοχής σε όλο και πολυπλοκότερα ερεθίσματα και στη μελέτη του ακουστικού φλοιού έχει αναδείξει ακριβώς τα όσα αναφέρθηκαν προηγουμένως. Επίσης, περιγράφηκε και μία προσπάθεια για την ανάπτυξη ενός συστήματος λογισμικού με σκοπό τη μοντελοποίηση της νευρωνικής επεξεργασίας των ηχητικών σημάτων στα πρώιμα στάδιά της.

Είναι εμφανές ότι ακόμα υπάρχει ένα σημαντικό έλλειμμα γνώσης, όχι τόσο στις ανατομικές δομές που συμμετέχουν στην αντίληψη του ήχου όσο στο είδος των μετασχηματισμών που αυτές επιβάλλουν στα ηχητικά σήματα και στις μεταξύ τους αλληλεπιδράσεις και διασυνδέσεις. Σε γενικές γραμμές, όπως φαίνεται και από το σχήμα 2.17, οι απολήξεις των τριχοφόρων κυττάρων του κοχλίου σχηματίζουν τις ίνες του ακουστικού νεύρου. Αυτές καταλήγουν στο κοχλιακό σύμπλεγμα του εγκεφαλικού στελέχους και ανέρχονται κατόπιν μέσω των πλευρικών λημνίσκων για να φτάσουν στο κάτω διδύμιο του μεσεγκεφάλου. Ο επόμενος σταθμός είναι το μέσο γονατώδες σώμα του θαλάμου απ' όπου οι ακουστικές οδοί προβάλλουν τελικά στον ακουστικό φλοιό. Σε πολλά σημεία αυτής της διαδρομής, οι δύο παράλληλες οδοί επικοινωνούν μεταξύ τους. Επίσης, υπάρχουν και σημεία ανάδρασης όπου ανώτερα στάδια προβάλλουν σε κατώτερα.

Ενώ οι ανατομικές δομές, όπως παρουσιάστηκαν προηγουμένως, είναι λίγο πολύ γνωστές, μας διαφεύγει προς το παρόν ο ακριβής ρόλος που έχει η καθεμία. Γνωρίζουμε ότι ο κοχλίας επιτελεί κάποιο είδος φασματικής ανάλυσης, εφόσον είναι οργανωμένος κατά τονοτοπικό τρόπο αλλά είναι πολύ πιθανό ότι διαθέτει και επιπλέον λειτουργίες. Όσον αφορά

στον φλοιό, είναι γνωστό ότι ο πρωτοταγής ακουστικός φλοιός έχει και αυτός μία τονοτοπική οργάνωση. Άλλες περιοχές του ακουστικού φλοιού όμως παρουσιάζουν προτιμήσεις σε πολυπλοκότερα ερεθίσματα, αποτελώντας ενδεχομένως ανώτερα στάδια επεξεργασίας. Για τα στάδια που παρεμβάλλονται μεταξύ του κοχλίου και του φλοιού, τα στοιχεία που διαθέτουμε είναι ακόμα φτωχότερα.

Οι απόπειρες μοντελοποίησης του ακουστικού συστήματος, λόγω της παραπάνω κατάστασης, είναι εκ των πραγμάτων περιορισμένες. Οι περισσότερες εμφανίστηκαν πριν αρκετά χρόνια και είχαν ως στόχο τη μοντελοποίηση του περιφερικού ακουστικού συστήματος για το οποίο και υπάρχουν πληρέστερες γνώσεις. Ταυτόχρονα, προσπαθούσαν να ενσωματώσουν και δεδομένα από ψυχοακουστικά πειράματα, όπως οι κρίσιμες ζώνες, οι καμπύλες ίσης έντασης και τα φαινόμενα σκίασης και προσαρμογής. Παρ' όλο που αυτά τα μοντέλα είχαν επιτυχία στο πεδίο της αυτόματης αναγνώρισης ομιλίας, φαίνεται ότι πλέον είναι αναγκαία η μοντελοποίηση και των άλλων σταδίων του κεντρικού ακουστικού συστήματος ώστε να ερμηνευθούν τα νέα δεδομένα που προκύπτουν από νευροφυσιολογικά πειράματα.

Προς αυτή την κατεύθυνση κινήθηκε η παρούσα εργασία. Δανειζόμενη από τις εργασίες του Shamma ([2]) και των συνεργατών του την ιδέα της αναπαράστασης σε πολλαπλά επίπεδα διακριτικής ικανότητας, τόσο στον άξονα του χρόνου όσο και σε αυτόν της συχνότητας, και την ιδέα της συστοιχίας Gabor φίλτρων από τους Ezzat et al ([5]), υλοποιήθηκε ένα σύστημα που προσπαθεί να συλλάβει τα βασικά χαρακτηριστικά των σημάτων ομιλίας. Το σύστημα αποτελείται από συστοιχίες Gabor ανιχνευτών που αναλύουν το σήμα εισόδου σε διάφορα επίπεδα διακριτικής ικανότητας ώστε να εντοπιστούν εκείνες οι παράμετροι που φέρουν πληροφορία. Με αυτόν τον τρόπο, είναι δυνατό να κατασκευαστούν αντιληπτικοί χάρτες, όπως προτείνεται από τον Nobuo Suga (κεφάλαιο 11, [7]). Αυτοί οι χάρτες θα μπορούσαν να αποτελέσουν το πρώτο στάδιο ενός γνωσιακού μοντέλου αντίληψης του λόγου, παρέχοντας σε αυτό μία αναπαράσταση του αρχικού σήματος που διατηρεί τις απαραίτητες πληροφορίες για την αποκωδικοποίηση του. Όπως παρουσιάστηκε και στο προηγούμενο κεφάλαιο, το παρόν μοντέλο έχει την ικανότητα να εντοπίζει αυτές τις πληροφορίες, ίσως όχι με τον ιδανικό πάντα τρόπο. Υπάρχουν πολλές παράμετροί του που θα έπρεπε να εξεταστούν με προσοχή.

5.2 Μελλοντικές επεκτάσεις

Το μοντέλο που παρουσιάστηκε σε αυτή την εργασία είναι μία πολύ πρώιμη προσπάθεια, βασισμένη σε πολλές υποθέσεις. Κάποια στοιχεία ίσως αποδειχθεί στο μέλλον ότι πρέπει να αφαιρεθούν. Για παράδειγμα, η ύπαρξη των offset ανιχνευτών δεν έχει επιβεβαιωθεί οριστικά από νευροφυσιολογικά πειράματα. Ακόμα, τα χρονικά παράθυρα των 10 ms ίσως να μην έχουν κάποια πραγματική αξία. Γενικότερα, οι τιμές των παραμέτρων που έχουν χρησιμοποιηθεί κατά τον έλεγχο του συστήματος είναι ως ένα βαθμό αυθαίρετες, βασισμένες σε κάποιες πολύ γενικές γνώσεις περί του ποιες είναι σημαντικές στην αντίληψη του λόγου και σε μερικά πρόσφατα πειράματα. Η διεξαγωγή νέων πειραμάτων για να υπολογιστούν με μεγαλύτερη ακρίβεια οι τιμές αυτών των παραμέτρων θα επέτρεπε την περαιτέρω βελτιστοποίηση του μοντέλου.

Ένα άλλο ενδιαφέρον ερώτημα αφορά στην ύπαρξη ή μη και άλλου είδους ανιχνευτών που έχουν παραλειφθεί εδώ. Για παράδειγμα, μία επέκταση του συστήματος θα μπορούσε να γίνει προς την κατεύθυνση της ενσωμάτωσης πληροφοριών για τις αργές διαμορφώσεις

πλάτους της περιβάλλουσας ενός σήματος ομιλίας. Όπως είδαμε, τέτοιες συχνότητες διαμόρφωσης μεταξύ 4 και 16 Hz είναι πολύ σημαντικές για την κατανόηση της ομιλίας ενώ υπάρχουν και νευρώνες ευαίσθητοι σε AM διαμορφώσεις. Ο εντοπισμός τέτοιων διαμορφώσεων θα μπορούσε να γίνει μέσω του φασματογράμματος διαμόρφωσης (βλ. παράγραφο 3.2.2.1). Επειδή όμως το φάσμα δεν αποκαλύπτει πάντα όλες τις λεπτομέρειες ενός σήματος, θα μπορούσαν να χρησιμοποιηθούν και τελεστές διαχωρισμού ενέργειας (ή τελεστές Teager) για αυτόν τον σκοπό. Άλλοι ανιχνευτές που ίσως θα ήταν χρήσιμο να προστεθούν είναι οι τύπου CF/CF και FM/FM για τον εντοπισμό συνδυασμών από οριζόντιες και κεκλιμένες μπάρες ενέργειας. Αυτό μπορεί να γίνει με τροποποίηση του υπάρχοντος συστήματος, εφόσον, όσο περισσότερες μπάρες περιέχει ένα παράθυρο, τόσο πιο απομακρυσμένες είναι οι συζυγείς, Γκαουσιανές κορυφές που προκύπτουν στη συνέχεια, όπως φαίνεται και στο σχήμα 4.6.

Φυσικά, η βασική υπόθεση περί της ύπαρξης αντιληπτικών χαρτών από συστοιχίες φίλτρων θα πρέπει τελικά να ελεγχθεί με ένα πραγματικό γνωσιακό μοντέλο για να διαπιστωθεί η εγκυρότητα της και τα όρια ισχύος της. Σε μία τέτοια προσπάθεια όμως, ελέγχεται όχι μόνο το παρόν μοντέλο αλλά και το ίδιο το γνωσιακό μοντέλο, εφόσον ούτε σε αυτή την περιοχή υπάρχουν πειραματικά επιβεβαιωμένες θεωρίες. Μάλιστα, εκεί ίσως υπάρχει ακόμα μεγαλύτερη ασάφεια και ασυμφωνία μεταξύ των ερευνητών. Υπό αυτές τις συνθήκες, οι πιθανότητες επιτυχίας μίας τέτοιας δοκιμής μάλλον προδιαγράφονται αρκετά περιορισμένες. Παρ' όλα αυτά, μέσω τέτοιων δοκιμών και ενδεχόμενων αποτυχιών μπορούμε να ελπίζουμε ότι κάποια στιγμή θα μας αποκαλυφθεί ο τρόπος με τον οποίο είναι δομημένος ο εγκέφαλος ώστε να μας επιτρέπει τη γλωσσική επικοινωνία.

Βιβλιογραφία

- [1] Borden GJ, Harris KS, Raphael LJ, 2003, *Speech science primer*, 4th ed., Lippincott Williams & Wilkins.
- [2] Chi T, Ru P, Shamma SA, 2005, *Multiresolution spectrotemporal analysis of complex sounds*, J Acoust Soc Am, 118:887-906.
- [3] Davis SB, Mermelstein P, 1980, *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*, IEEE Transactions on Acoustics, Speech and Signal Processing, 28:357-366.
- [4] Depireux DA, Simon JZ, Klein DJ, Shamma SA, 2001, *Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex*, J Neurophysiol, 85:1220-1234.
- [5] Ezzat T, Bouvrie J, Pozzio T, 2007, *Spectro-temporal analysis of speech using 2-D Gabor filters*, to appear in Interspeech, Antwerp, Belgium, 2007.
- [6] Ghitza O, 1986, *Auditory nerve representation as a front-end for speech recognition in a noisy environment*, Computer Speech and Language, 1:109-130.
- [7] Greenberg S, Ainsworth WA (Ed.), 2006, *Listening to speech : an auditory perspective*, Lawrence Erlbaum.
- [8] Greenberg S, Kingsbury B, 1997, *The modulation spectrogram : In pursuit of an invariant representation of speech*, in Proc. ICASSP, 1647-1650.
- [9] Heil P, 1997, *Auditory cortical onset responses revisited. II. Response strength*, J Neurophysiol, 77:2642-2660.
- [10] Hendelman W, 2006, *Atlas of functional neuroanatomy*, 2nd ed., CRC Press.
- [11] Hermansky H, 1990, *Perceptual linear predictive (PLP) analysis of speech*, J Acoust Soc Am, 87:1738-1752.
- [12] Kandel ER, Schwarzh JH, Jessell TM, 2000, *Principles of neural science*, 4th ed., McGraw-Hill.
- [13] Kowalski N, Depireux DA, Shamma SA, 1996, *Analysis of dynamic spectra in ferret primary auditory cortex. I. Response characteristics of single unit responses to moving ripple spectra*, J Neurophysiol, 76:3503-3523.

- [14] Lu T, Liang L, Wang X, 2001, *Temporal and rate representations of time-varying signals in the auditory cortex of awake primates*, Nat Neurosci, 11:1131-1138.
- [15] *MATLAB Help*
- [16] Oppenheim AV, Schafer RW, 1999, *Discrete-time signal processing*, 2nd ed., Prentice Hall.
- [17] Orduna I, Mercado E, Gluck M, Merzenich M, 2001, *Spectrotemporal sensitivities in rat auditory cortical neurons*, Hear Res, 160:47-57.
- [18] Palmer A, Shamma S, 2004, *Physiological representations of speech*, In Greenberg S, Ainsworth WA, Popper AN, Fay RR (Ed.), *Speech processing in the auditory system* (Chapter 4), Springer.
- [19] Qin L, Chimoto S, Sakai M, Wang J, Sato Y, 2007, *Comparison between offset and onset responses of primary auditory cortex ON-OFF neurons in awake cats*, J Neurophysiol, 97:3421-3431.
- [20] Quatieri TF, 2002, *Discrete-time speech processing : principles and practice*, Prentice Hall.
- [21] Shamma SA, Versnel H, Kowalski N, 1995, *Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusodally ripple spectra*, Aud Neurosci, 1:233-254.
- [22] Slaney M, *Auditory Toolbox*, <http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/>
- [23] Slaney M, Lyon R, 1993, *On the importance of time - A temporal representation of sound*, In Cooke M, Beet S, Crawford M (Ed.), *Visual representations of speech signals* (Chapter 5), John Wiley & Sons.
- [24] Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM, 2004, *Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity*, J Neurophysiol, 91:1282-1296.
- [25] Tchorz J, Kollmeier B, 1999, *A model of auditory perception as front end for automatic speech recognition*, J Acoust Soc Am, 106:2040-2050.
- [26] Tian B, Rauschecker JB, 2004, *Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey*, J Neurophysiol, 92:2993-3013.
- [27] Weisz N, Wienbruch C, Hoffmeister S, Elbert T, 2004, *Tonotopic organization of the human auditory cortex probed with frequency-modulated tones*, Hear Res, 191:49-58.
- [28] Young ED, Sachs MB, 1979, *Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers*, J Acoust Soc Am, 66:1381-1403.

BIBΛΙΟΓΡΑΦΙΑ

- [29] Μαραγκός Π, 2005, *Ανάλυση εικόνων και όραση υπολογιστών*, Εκδόσεις ΕΜΠ.
- [30] <http://www.nlm.nih.gov/medlineplus/ency/images/ency/fullsize/1092.jpg>.
- [31] http://anatomy.iupui.edu/courses/histo_D502/D502f04/lecture.f04/Earf04/cochshake.jpg.
- [32] <http://www.sickkids.ca/cochlearimplant/section.asp?s=Physicians&ssID=6704&ss= Physiology+of+normal+hearing&ssID=6733>.
- [33] <http://www.brainconnection.com/med/medart/tif/cochlea-chambers.tif>.
- [34] <http://scienceblogs.com/clock/upload/2006/06/Cochlea.jpeg>.
- [35] http://www.santafevisions.com/csf/images/overviews/hearing/2_ear_structure/7_anerv.gif.
- [36] http://www.cite-sciences.fr/english/ala_cite/expo/explora/image/diaporama/images/30-n701.jpg.
- [37] <http://www.medfarm.uu.se/forskning/bild.php?typ=forskningsprogram&id=379>.
- [38] <http://www.cosmoscan.pe.kr/moving%20music/auditory%20area.jpg>.
- [39] <http://emcap.iua.upf.es/BasicsAuditorySystem.jpg>.