

Τμήμα  
Μηχανικών  
Πληροφορικής τ.ε.  
Τεχνολογικό Εκπαιδευτικό Ίδρυμα  
Δυτικής Ελλάδας

## ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Τεχνικές Συμπίεσης Ψηφιακού Ήχου με Χρήση  
Ψυχοακουστικών Μοντέλων Ακοής

---

Αναστασία Αθαν. Τζάλα

Σταυρούλα Αθαν. Πολυδώρου

Επιβλέπων Καθηγητής : Μιχάλης Παρασκευάς

Αντίρριο, Απρίλιος 2015

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή  
Αντίρριο, 29/4/2015

**ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ**

1. Παρασκευάς Μιχάλης
2. Τσακανίκας Βασίλειος
3. Μπόρας Ιωσήφ

## Περιεχόμενα

ΕΙΣΑΓΩΓΙΚΕΣ ΕΝΝΟΙΕΣ .....	6
1.1 ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ.....	6
1.1.2 ΑΚΟΥΣΤΙΚΗ-ΨΥΧΟΑΚΟΥΣΤΙΚΗ .....	8
1.1.3 ΕΦΑΡΜΟΣΜΕΝΗ ΨΥΧΟΑΚΟΥΣΤΙΚΗ.....	9
1.1.4 ΨΥΧΟΑΚΟΥΣΤΙΚΑ ΠΕΙΡΑΜΑΤΑ .....	9
1.1.5 ΨΥΧΟΑΚΟΥΣΤΙΚΑ ΠΕΙΡΑΜΑΤΑ - ΜΕΘΟΔΟΙ ΕΚΤΕΛΕΣΗΣ - ΑΠΟΤΕΛΕΣΜΑΤΑ ....	10
1.2 ΣΥΝΕΙΣΦΟΡΑ ΕΡΓΑΣΙΑΣ .....	11
1.3 ΠΑΡΟΥΣΙΑΣΗ ΚΑΙ ΔΟΜΗ .....	11
ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ ΑΚΟΥΣΤΙΚΗΣ ΕΠΙΚΑΛΥΨΗΣ – ΜΕΘΟΔΟΙ ΤΕΧΝΙΚΩΝ ΗΧΗΤΙΚΗΣ ΣΥΜΠΙΕΣΗΣ .....	13
2.1 ΕΙΣΑΓΩΓΗ .....	13
2.2 ΑΝΑΤΟΜΙΑ ΑΚΟΥΣΤΙΚΟΥ ΣΥΣΤΗΜΑΤΟΣ.....	13
2.2.1 ΕΞΩΤΕΡΙΚΟ ΑΥΤΙ .....	14
2.2.2 ΜΕΣΟ ΑΥΤΙ.....	15
2.2.3 ΕΣΩΤΕΡΙΚΟ ΑΥΤΙ .....	16
2.3 ΚΡΙΣΙΜΕΣ ΠΕΡΙΟΧΕΣ ΣΥΧΝΟΤΗΤΑΣ .....	18
2.4 ΑΠΟΛΥΤΟ ΚΑΤΩΦΛΙ ΑΚΟΥΣΤΟΤΗΤΑΣ.....	20
2.5 ΦΑΙΝΟΜΕΝΟ ΑΚΟΥΣΤΙΚΗΣ ΕΠΙΚΑΛΥΨΗΣ .....	21
2.6 ΣΥΝΟΨΗ ΚΕΦΑΛΑΙΟΥ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ.....	23
ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ ΜΕ ΧΡΗΣΗ ΨΥΧΟΑΚΟΥΣΤΙΚΟΥ ΜΟΝΤΕΛΟΥ.....	24
3.1 ΣΤΑΔΙΑ ΨΗΦΙΟΠΟΙΗΣΗΣ.....	24
3.1.1 ΔΕΙΓΜΑΤΟΛΗΨΙΑ.....	27
3.1.2 ΠΡΟΒΛΗΜΑΤΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ .....	28
3.1.3 ΚΒΑΝΤΟΠΟΙΗΣΗ .....	30
3.2 ΑΛΓΟΡΙΘΜΟΙ ΣΥΜΠΙΕΣΗΣ ΣΤΟ ΠΕΔΙΟ ΤΟΥ ΧΡΟΝΟΥ- ΜΕΘΟΔΟΙ ΚΩΔΙΚΟΠΟΙΗΣΗΣ .....	32
3.2.1 ΠΡΟΣΑΡΜΟΖΩΜΕΝΗ ΠΑΛΜΟΚΩΔΙΚΗ ΔΙΑΜΟΡΦΩΣΗ PCM .....	32
3.2.2 MU-LAW PCM ΚΑΙ A-LAW PCM.....	33
3.2.3 ΔΙΑΦΟΡΙΚΗ ΠΑΛΜΟΚΩΔΙΚΗ ΔΙΑΜΟΡΦΩΣΗ DPCM.....	34
3.2.4 ΔΙΑΜΟΡΔΩΣΗ ΔΕΛΤΑ (Delta Modulation - DM).....	35
3.2.5 ΔΙΑΦΟΡΙΚΗ ΠΡΟΣΑΡΜΟΖΩΜΕΝΗ ΠΑΛΜΟΚΩΔΙΚΗ ΔΙΑΜΟΡΦΩΣΗ ADPCM.....	36
3.2.6 ΠΡΟΣΑΡΜΟΖΩΜΕΝΗ ΠΑΛΜΟΚΩΔΙΚΗ ΚΩΔΙΚΟΠΟΙΗΣΗ LPC .....	38
ΜΕΛΕΤΗ ΤΕΧΝΙΚΩΝ ΣΥΜΠΙΕΣΗΣ ΨΗΦΙΑΚΟΥ ΗΧΟΥ.....	39
4.1 ΕΙΣΑΓΩΓΗ .....	39

4.2 ΣΥΜΠΙΕΣΗ ΨΗΦΙΑΚΟΥ ΗΧΟΥ .....	39
4.3 ΠΡΟΤΥΠΑ MPEG.....	42
4.3.1 Το Πρότυπο MPEG - 1.....	42
4.3.2 Το Πρότυπο MPEG – 2.....	45
4.3.2.1 Το Πρότυπο MPEG - 2 LSF.....	46
4.3.2.2 Το Πρότυπο MPEG - 2 BC .....	46
4.3.2.3 Το πρότυπο MPEG - 2 AAC .....	47
4.4 Το Πρότυπο MPEG - 4.....	49
4.5 Το Πρότυπο MPEG - 7.....	49
4.5 ΑΛΛΑ ΠΡΟΤΥΠΑ ΨΗΦΙΑΚΟΥ ΗΧΟΥ .....	50
4.5.1 WMA (Windows Media Audio).....	50
4.5.2 MP3 Pro.....	51
4.5.3 AAC (Advance Audio Coding).....	51
4.5.4 OGG Vorbis .....	52
4.5.5 AC-2 DOLBY DIGITAL .....	52
4.5.6 AC-3 DOLBY DIGITAL .....	53
4.5.7 ATRAC .....	54
4.5.8 PAC .....	54
4.6 ΤΥΠΟΙ ΑΡΧΕΙΩΝ ΨΗΦΙΑΚΟΥ ΗΧΟΥ .....	55
ΥΛΟΠΟΙΗΣΗ ΤΟΥ ΨΥΧΟΑΚΟΥΣΤΙΚΟΥ ΜΟΝΤΕΛΟΥ ΜΕ ΧΡΗΣΗ MATLAB.....	57
5.1 ΕΙΣΑΓΩΓΗ .....	57
5.2 Η ΑΠΟΔΕΙΞΗ ΤΗΣ ΕΝΝΟΙΑΣ : ASP_mp3.m ΜΕΣΩ ΤΟΥ MATLAB .....	58
5.2.1 Η ΤΡΑΠΕΖΑ ΦΙΛΤΡΩΝ ΔΥΟ ΚΑΝΑΛΙΩΝ .....	58
5.2.2 ΤΡΑΠΕΖΑ ΦΙΛΤΡΩΝ ΔΥΟ ΚΑΝΑΛΙΩΝ QMF.....	63
5.2.3 ΤΡΑΠΕΖΑ ΦΙΛΤΡΩΝ 32 ΚΑΝΑΛΙΩΝ PSEUDO-QMF .....	65
5.2.4 ΤΡΑΠΕΖΕΣ ΦΙΛΤΡΩΝ και ΕΠΙΚΑΛΥΠΤΟΜΕΝΟΙ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ .....	69
5.2.5 Perceptual Audio Coding (αντληπτική ακουστική κωδικοποίηση) .....	72
5.3 ΥΛΟΠΟΙΗΣΗ ΚΑΙ ΑΠΕΙΚΟΝΙΣΗ ΔΙΑΓΡΑΜΑΤΩΝ ΣΕ ΔΙΑΦΟΡΕΣ ΤΙΜΕΣ ΒΙΤΡΑΤΕ ....	82
5.3.1 ΠΕΡΙΟΔΟΓΡΑΜΜΑΤΑ ΑΡΧΙΚΟΥ ΗΧΟΥ ΣΕ ΔΙΑΦΟΡΕΣ ΤΙΜΕΣ ΒΙΤΡΑΤΕ .....	83
5.3.2 ΑΠΟΣΠΑΣΜΑΤΑ ΑΠΟ ΤΗΝ SUB-BAND #2.....	84
5.3.3 ΦΑΣΜΑΤΟΓΡΑΦΗΜΑΤΑ ΑΡΧΙΚΟΥ ΗΧΟΥ .....	85
5.3.4 ΠΕΡΙΟΔΟΓΡΑΜΜΑΤΑ ΣΤΟ ΣΤΑΔΙΟ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΚΑΙ ΚΒΑΝΤΟΠΟΙΗΣΗΣ ..	86
5.3.5 ΔΙΑΓΡΑΜΜΑ ΓΙΑ ΤΟ ΕΛΑΧΙΣΤΟ ΚΑΤΩΦΛΙ ΑΚΡΟΑΣΗΣ .....	87
5.3.6 ΦΑΣΜΑΤΟΓΡΑΦΗΜΜΑΤΑ ΣΥΜΠΙΕΣΜΕΝΟΥ ΗΧΟΥ .....	88

5.3.7 ΑΠΟΣΠΑΣΜΑΤΑ ΣΥΜΠΙΕΣΜΕΝΟΥ ΗΧΟΥ ΑΠΟ ΤΗΝ SUB-BAND #2 .....	89
5.3.8 ΑΠΟΣΠΑΣΜΑΤΑ ΑΠΟ ΤΗΝ SUB-BAND #20 .....	90
5.3.9 ΠΕΡΙΟΔΟΓΡΑΜΜΑΤΑ ΣΥΜΠΙΕΣΜΕΝΟΥ ΗΧΟΥ .....	91
5.3.10 ΓΡΑΦΙΚΗ ΠΑΡΑΣΤΑΣΗ ΚΛΙΜΑΚΑΣ .....	92
5.4 ΣΥΜΠΕΡΑΣΜΑ .....	93
ΑΝΑΦΟΡΕΣ .....	93

# ΚΕΦΑΛΑΙΟ 1

## ΕΙΣΑΓΩΓΙΚΕΣ ΕΝΝΟΙΕΣ

### 1.1 ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ

Είναι γεγονός ότι στην σύγχρονη αυτή εποχή η τεχνολογική πρόοδος και η εφαρμογή των τεχνολογικών εξελίξεων στο χώρο των τηλεπικοινωνιών οδήγησε σε νέα επιτεύγματα. Τα επιτεύγματα αυτά έχουν σημειωθεί στο χώρο αυτό έχουν δημιουργήσει μια νέα και εντυπωσιακή πραγματικότητα. Οι ολοένα αυξανόμενοι ρυθμοί της ζωής μας δημιούργησαν την ανάγκη για κατανάλωση και τεχνολογικές κατακτήσεις. Οι νέες δυνατότητες επικοινωνίας και ψυχαγωγίας έχουν προσφερθεί στον σύγχρονο άνθρωπο και στην σύσφιξη διαπροσωπικών σχέσεων όσο και την ευκαιρία για την δημιουργική αξιοποίηση ελεύθερου χρόνου. Δημιουργείται έτσι μεγάλη δυνατότητα ενημερωτικών ή ψυχαγωγικών επιλογών είτε μέσω δορυφορικών πραγμάτων (ραδιοφώνου, τηλεόρασης) είτε μέσω διεθνών ηλεκτρονικών υπολογιστών.

Σημαντικά τεχνολογικά επιτεύγματα όπως η Ψηφιακή Ραδιοφωνία (Digital Audio Broadcast) η οποία παρέχει ακουστική ποιότητα ίση με αυτήν του compact disc (CD) και η τηλεόραση Υψηλής Ευκρίνειας (HDTV) έχει ήδη ολοκληρωθεί η πειραματική τους χρήση. Οικονομικά και πολιτικά θεωρούνται εμπόδια για την ολοκληρωμένη χρήση τους. Η ψηφιακή τεχνολογία του ήχου υπερέρχει σε σύγκριση με την αναλογική τεχνολογία τόσο στην ποιότητα όσο και στην ευκολία διανομής του επεξεργασμένου σήματος. Αναμφίβολα, ένα από τα βασικά μειονεκτήματα της ψηφιακής τεχνολογίας είναι η υψηλή παροχή των ψηφιακών δεδομένων που απαιτούνται για την αποθήκευση της ακουστικής πληροφορίας.

Στόχος της πτυχιακής είναι να γίνει μια έρευνα στην εύρεση της κατάλληλης συχνότητας ώστε να έχουμε υψηλή ευκρίνεια στον ήχο και το συμπιεσμένο αρχείο να έχει όσο τον δυνατόν μικρότερο χώρο αποθήκευσης. Συγκεκριμένα, η εύρεση της βέλτιστης συχνότητας θα βασιστεί σε ποιοτικά και ποσοτικά κριτήρια. Τα ποιοτικά κριτήρια στηρίζονται στο ακουστικό μέρος ενώ τα ποσοτικά στα διαγράμματα.

### 1.1.1 ΕΝΔΕΙΚΤΙΚΕΣ ΛΕΙΤΟΥΡΓΙΕΣ ΤΟΥ ΑΚΟΥΣΤΙΚΟΥ ΣΥΣΤΗΜΑΤΟΣ

Τα εισερχόμενα στο αυτί ακουστικά κύματα προκαλούν δονήσεις στο ακουστικό τύμπανο που μεταδίδονται στη βασική μεμβράνη, η οποία βρίσκεται στο εσωτερικό αυτί και συντονίζεται σε διαφορετικό σημείο ανάλογα με την συχνότητα του διεγείροντος ακουστικού κύματος. Οι ταλαντώσεις και τα τοπικά μέγιστα της βασικής μεμβράνης θέτουν σε διέγερση ένα μεγάλο αριθμό ευαίσθητων τριχοειδών κυττάρων, που βρίσκονται τοποθετημένα κατά μήκος της και οι ηλεκτρικοί παλμοί που παράγονται από τα κύτταρα αυτά μεταδίδονται στα ανώτερα ακουστικά κέντρα από ένα περίπλοκο μη-γραμμικό δίκτυο νευρικών ινών.

Η συχνοτική ανάλυση που επιτελείται από το παραπάνω σύστημα αντιστοιχεί σε ένα σύστημα παράλληλων και επικαλυπτόμενων ζωνοδιαβατών φίλτρων. Όπως έχει βρεθεί από πειράματα, [Παρ95] κάθε ένα από αυτά τα φίλτρα έχει εύρος ζώνης πάνω από ένα τρίτο της οκτάβας της κεντρικής του συχνότητας και κλίση περισσότερο απότομη για την κάτω από την κεντρική συχνότητα περιοχή. Έχει βρεθεί όμως, ότι για καθαρούς τόνους ή αρμονικές στην περιοχή των 4 με 5 kHz, η συχνοτική επιλεκτικότητα του ακουστικού συστήματος μπορεί να είναι στενότερη από αυτήν του ενός τρίτου της οκτάβας φίλτρου. Έχει βρεθεί επίσης, ότι υπάρχει και ένας χρόνος ολοκλήρωσης της τάξης των 150 ms, κάτω από τον οποίο πολύ βραχείς και υψηλής στάθμης ήχοι με μικρότερο από το ένα τρίτο της οκτάβας εύρος ζώνης δεν διαχωρίζονται από μεγαλύτερης διάρκειας και χαμηλότερης στάθμης ήχους της ίδιας ενέργειας και συχνοτικού περιεχομένου. Η χρονοσυχνοτική ανάλυση του αυτιού οδηγεί και σε φαινόμενα επικάλυψης (masking effects), κατά τα οποία παραμορφώσεις και θόρυβοι είναι δυνατό να επικαλυφθούν και να μην είναι ακουστοί, από ισχυρότερα σήματα της ίδιας ή χαμηλότερης ζώνης συχνοτήτων, όταν διεγείρουν το σύστημα ακοής στο ίδιο περίπου διάστημα χρόνου.

Ο μηχανισμός της επικάλυψης είναι τόσο σημαντικός, ώστε όλοι οι ερευνητές να πιστεύουν ότι η επίτευξη της κωδικοποίησης υψηλής πιστότητας με πολύ χαμηλό ρυθμό παροχής δεδομένων (bit rate) (π.χ. 96 ή 64 kbits/s), προϋποθέτει τον σχεδιασμό αλγορίθμων βασισμένων στην προσομοίωση της λειτουργίας αυτής. Έτσι, σχεδόν όλοι οι σύγχρονοι αλγόριθμοι υποκειμενικής κωδικοποίησης, (Perceptual Audio Coders - PACs), υπολογίζουν το κατώφλι επικάλυψης θορύβου (Noise Masking Threshold - NMT), και κατόπιν μορφοποιούν τις συνιστώσες του θορύβου κβαντισμού (quantization noise) σύμφωνα με το κατώφλι, έτσι ώστε να μην είναι ακουστές. Ο όρος υποκειμενική κωδικοποίηση,

χρησιμοποιείται για να υποδηλωθεί η προσομοίωση του ανθρώπινου ακουστικού συστήματος από - το σύστημα κβαντισμού και κωδικοποίησης. Το κατώτερο όριο για την παροχή πληροφορίας ανά δείγμα, που επιτυγχάνουν οι υποκειμενικοί κωδικοποιητές υπολογίστηκε από τον Johnston , ο οποίος έδειξε ότι η χωρητικότητα πληροφορίας του ανθρώπινου ακουστικού συστήματος είναι περίπου 2 bits ανά δείγμα, για διάφανη (transparent) κωδικοποίηση. Το όριο αυτό επιτυγχάνεται με τεχνικές κωδικοποίησης Υποκειμενικής Εντροπίας (Perceptual Entropy techniques), οι οποίες παρέχουν ποιότητα CD για παροχή 128 kbits/s ανά κανάλι και ποιότητα σχεδόν ίδια με του CD για παροχή 64 kbits/s ανά κανάλι, και στις δύο περιπτώσεις για σήμα πλήρους εύρους ζώνης 20 kHz.

## 1.1.2 ΑΚΟΥΣΤΙΚΗ-ΨΥΧΟΑΚΟΥΣΤΙΚΗ

Η ακουστική (Acoustics) [Μητ11] είναι η επιστήμη της συμπεριφοράς των ηχητικών κυμάτων. Μελετά την παραγωγή, διάδοση, και λήψη των εν λόγω κυμάτων από το ανθρώπινο αυτί. Είναι επίσης η επιστήμη που ασχολείται με όλα γενικά τα μηχανικά κύματα, τα οποία εκτείνονται πέρα από το ακουστικό φάσμα του ανθρώπινου αυτιού και ως εκ τούτου δεν γίνονται αντιληπτά. Τέτοια κύματα είναι τα χαμηλής συχνότητας κύματα που παράγονται από τους σεισμούς (υποηχητικά κύματα) και τα υψηλής συχνότητας ή υπερηχητικά κύματα καθώς και οι υψίσυχνες ταλαντώσεις των ατόμων στα στερεά σώματα. Η Ακουστική είναι εξαιρετικά σημαντική στο λόγο και στην ακρόαση (ραδιοφωνικά studios, θεατρικές αίθουσες, συνεδριακοί χώροι) και στη μουσική (παραγωγή και ακρόαση) σε studios ηχογράφησης και αίθουσες συναυλιών.

Η ψυχοακουστική (Psychoacoustics) [link1] αποτελεί τον κλάδο της ακουστικής ο οποίος μελετά τον υποκειμενικό τρόπο με τον οποίο το ανθρώπινο σύστημα ακοής αντιλαμβάνεται τους διάφορους ήχους, συμπεριλαμβανόμενης και της μουσικής (συνειδητοποίηση της ηχητικής πληροφορίας σε γνωστικό επίπεδο). Με την ψυχοακουστική εμπλέκονται και πολλές άλλες επιστήμες όπως η ψυχολογία, οι νευροεπιστήμες, η γλωσσολογία ως μέσο έρευνας της ανθρώπινης συμπεριφοράς.



### 1.1.3 ΕΦΑΡΜΟΣΜΕΝΗ ΨΥΧΟΑΚΟΥΣΤΙΚΗ

Η ψυχοακουστική αυτή την στιγμή εφαρμόζεται σε πολλά πεδία της ανάπτυξης λογισμικού, όπου οι προγραμματιστές αποτυπώνουν πειράματα μαθηματικών μοντέλων που έχουν αποδειχθεί στην ψηφιακή επεξεργασία σήματος, όπου πολλοί κωδικοποιητές συμπίεσης ήχου - όπως το mp3 - χρησιμοποιούν ένα ψυχοακουστικό πρότυπο για την αύξηση του λόγου συμπίεσης [Μου09]. Επίσης, εφαρμόζεται κατά το σχεδιασμό των υψηλής ποιότητας συστημάτων ήχου για ακριβή αναπαραγωγή της μουσικής σε θέατρα και σπίτια καθώς και για αμυντικά συστήματα, όπου οι επιστήμονες έχουν πειραματιστεί με περιορισμένη επιτυχία στην δημιουργία νέων ακουστικών όπλων, τα οποία εκπέμπουν που μπορούν να περιορίσουν τις βλάβες ή το θάνατο. Επιπλέον, στις μέρες μας η ψυχοακουστική έχει εφαρμογές στην μουσική, όπου οι μουσικοί και οι καλλιτέχνες συνεχίζουν να δημιουργούν νέες ακουστικές εμπειρίες συγκαλύπτοντας τις συχνότητες των ανεπιθύμητων μέσων. Τέλος, μια άλλη εφαρμογή είναι στο σχεδιασμό των μικρών ή κατώτερης ποιότητας ηχείων, τα οποία χρησιμοποιούν το φαινόμενο της χαμμένης χαμηλής συχνότητας για να δώσει το εφέ των χαμηλών συχνοτήτων του μπάσου όταν το σύστημα, εξαιτίας του συγκεκριμένου ορίου συχνοτήτων τις οποίες καλύπτει, δεν μπορεί να τις αναπαράγει.

### 1.1.4 ΨΥΧΟΑΚΟΥΣΤΙΚΑ ΠΕΙΡΑΜΑΤΑ

Τα ψυχοακουστικά πειράματα αναπτύσσονται με σκοπό να κατανοήσουν οι επιστήμονες το πώς αντιλαμβάνεται τον ήχο το ανθρώπινο αυτί [Μου09]. Για το λόγο αυτό έχουν αναπτυχθεί αρκετά ψυχοακουστικά πειράματα τα οποία μελετούν τις διαφορετικές ιδιότητες της αντίληψης κάποιων ηχητικών φαινομένων. Οι μέθοδοι που ακολουθούνται για την εκτέλεση των ψυχοακουστικών πειραμάτων ποικίλουν ανάλογα με το είδος του πειράματος. Επιπλέον, για την όσο το δυνατόν καλύτερη κατανόηση των συμπερασμάτων και το αποτελεσμάτων χρειάζεται ένα μεγάλο δείγμα από αυτά τα πειράματα. Τα πειράματα αυτά για την καλύτερη και ακριβέστερη δυνατή διεξαγωγή τους γίνονται μέσα σε κατάλληλα διαμορφωμένους χώρους χρησιμοποιώντας τον κατάλληλο εξοπλισμό τόσο σε υλικό όσο και σε λογισμικό ή ακόμα και στο διαδίκτυο. Επίσης, όπως αναφέρθηκε, τα πειράματα αυτά έχουν ως σκοπό να χαρτογραφήσουν την ικανότητα της ακουστικής των ανθρώπων, να εξελίξουν κάποιους αλγορίθμους συμπίεσης ή ακόμη να μελετήσουν την ακουστική

ικανότητα ενός χώρου. Ένα παράδειγμα για το πώς συνδέονται τα πειράματα αυτά μεταξύ τους είναι η συμπίεση των ηχητικών δεδομένων με τον κωδικοποιητή mp3. Ουσιαστικά οι ερευνητές μελετούν τις δυνατότητες του ανθρώπινου αυτιού και καταγράφουν τις συχνότητες εκείνες που δεν αντιλαμβάνεται το ανθρώπινο αυτί. Με τον τρόπο αυτό κατάφεραν να δημιουργήσουν έναν κωδικοποιητή συμπίεσης, το γνωστό σε όλους mp3, ο οποίος έχει ως σκοπό να σβήσει από το ηχητικό αρχείο αυτές τις συχνότητες χωρίς να αντιλαμβάνεται ο μέσος χρήστης την αλλοίωση που υπέστη ο ήχος.

Η πλατφόρμα των ψυχοακουστικών πειραμάτων αναπτύχθηκε με την χρήση τριών γνωστών ψυχοακουστικών φαινομένων. Τα ψυχοακουστικά αυτά πειράματα δεν εξελίσσονται σε κάποιον ειδικά διαμορφωμένο χώρο αλλά ούτε χρησιμοποιούν κάποιο εξειδικευμένο λογισμικό όπως συμβαίνει σε διάφορα εργαστήρια όπου γίνονται οι μελέτες.

Αν και υπάρχουν αρκετά ψυχοακουστικά πειράματα τα οποία μπορούν να ενσωματωθούν σε αυτή την εγκατάσταση, ένα σημαντικό κριτήριο για την επιλογή των πειραμάτων είναι η δυνατότητα έλεγχου λίγων παραμέτρων με μικρή σχετικά δυναμική περιοχή. Αυτό το κριτήριο το λαμβάνουμε υπ' όψιν λόγω των περιορισμένων διαστάσεων του χώρου της εγκατάστασης και κατ' επέκταση της περιορισμένης απόστασης μεταξύ του συμμετέχοντος και της περιοχής προσέγγισης.

### **1.1.5 ΨΥΧΟΑΚΟΥΣΤΙΚΑ ΠΕΙΡΑΜΑΤΑ - ΜΕΘΟΔΟΙ ΕΚΤΕΛΕΣΗΣ - ΑΠΟΤΕΛΕΣΜΑΤΑ**

Τα ψυχοακουστικά πειράματα γίνονται με σκοπό την καταγραφή της ευκρίνειας και του τρόπου αντίληψης του ήχου από τους ανθρώπους. Για την διεξαγωγή αυτών των πειραμάτων απαιτούνται να υπάρχουν κάποιες κατάλληλες προϋποθέσεις για την διεξαγωγή τους (π.χ. ιδανικές συνθήκες ακρόασης, μονωμένοι χώροι όπου δεν υπάρχουν παρεμβολές από τρίτους, χρήση ακουστικών κ.α.). Πέρα όμως από τα υλικά τα οποία πολλές φορές είναι αναγκαία, έχουν αναπτυχθεί και κατάλληλα συστήματα λογισμικού από ειδικούς τα οποία εξομοιώνουν με το καλύτερο δυνατό τρόπο αυτά τα πειράματα προκειμένου να μπορέσουν να χαρτογραφήσουν και να εξηγήσουν στην συνέχεια τα αποτελέσματα που παράγονται.

## 1.2 ΣΥΝΕΙΣΦΟΡΑ ΕΡΓΑΣΙΑΣ

Στα πλαίσια της παρούσας πτυχιακής υλοποιήθηκε ένα μοντέλο ακουστικής επικάλυψης για την επεξεργασία ηχητικών σημάτων. Οι τεχνικές κωδικοποίησης ακουστικών σημάτων κυρίως υλοποιούμενες στο πεδίο του χρόνου και χωρίς την εκμετάλλευση ψυχοακουστικών ιδιοτήτων του αυτιού με περισσότερο χαρακτηριστικές την Διαφορική Παλμοκωδική Διαμόρφωση (Differential Pulse Code Modulation - DPCM) και την Προσαρμοζόμενη Διαφορική Παλμοκωδική Διαμόρφωση (Adaptive Differential Pulse Code Modulation – ADPCM) αξιοποιούν τη χρονική στασιμότητα των ηχητικών σημάτων. Συγκεκριμένα, η χρονική στασιμότητα αξιοποιείται κωδικοποιώντας τη διαφορά ανάμεσα σε διαδοχικά δείγματα και όχι κάθε δείγμα αυτόνομα όπως η PCM, ενώ επιπλέον η ADPCM προσαρμόζει δυναμικά το βήμα κβαντισμού σύμφωνα τις στιγμιαίες μεταβολές του σήματος εισόδου.

## 1.3 ΠΑΡΟΥΣΙΑΣΗ ΚΑΙ ΔΟΜΗ

Η πτυχιακή είναι διαρθρωμένη σε πέντε κεφάλαια. Κάθε κεφάλαιο αποτελεί μία εννοιολογικά αυτοτελή ενότητα. Στην αρχή κάθε κεφαλαίου υπάρχει μία σύντομη περίληψη, η οποία αναφέρεται σε θέματα που αφορούν το αντίστοιχο κεφάλαιο.

Το παρόν κεφάλαιο γίνεται εισαγωγή της πτυχιακής, στην οποία αναφέρονται οι λόγοι που καθιστούν χρήσιμες και απαραίτητες τις τεχνικές συμπίεσης σημάτων μουσικής.

Στο **Κεφάλαιο 2** παρουσιάζονται επιγραμματικά τα βασικά χαρακτηριστικά της ανατομίας, οι μηχανισμοί λειτουργίας του ανθρώπινου ακουστικού συστήματος. Οι θεμελιώδεις αρχές που σχετίζονται με την ευρύτερη περιοχή της ψυχοακουστικής, η δομή δηλαδή και η λειτουργία του εξωτερικού και στην συνέχεια του μέσου αυτιού. Επιπρόσθετα, ακολουθεί αναλυτική παρουσίαση του εσωτερικού αυτιού, το οποίο παρουσιάζει μη γραμμική συμπεριφορά και επιτελεί τις πλέον σημαντικές ακουστικές λειτουργίες. Τέλος, παρουσιάζει το φαινόμενο της ακουστικής επικάλυψης.

Στο **Κεφάλαιο 3** αναφέρεται στην διαδικασία ψηφιοποίησης του ήχου. Συγκεκριμένα αναλύεται η δειγματοληψία καθώς και τα προβλήματα που μπορεί να προκαλέσει. Επίσης, παρουσιάζεται ο καθοριστικός ρόλος της κινητοποίησης. Επιπλέον, γίνεται πλήρης ανάλυση στις μεθόδους κωδικοποίησης. Πιο συγκεκριμένα αναφέρεται οι εξής : Παλμοκωδική κωδικοποίηση (PCM), η διαφορική Παλμοκωδική κωδικοποίηση (DPCM), γραμμική προβλεπόμενη κωδικοποίηση (LPC), (Mu-Law- PCM) λογαριθμική Παλμοκωδική

κωδικοποίηση (a-Law-PCM) γραμμική Παλμοκωδική κωδικοποίηση, ADPCM (επέκταση/ παραλλαγή DPCM).

Στο **Κεφάλαιο 4** θα δοθούν οι βασικές έννοιες των τεχνικών συμπίεσης του ψηφιακού ήχου. Αρχικά θα γίνει μια παρουσίαση των προτύπων MPEG και ACC3 δίνοντας βάση στην χρησιμότητα τους. Επιπλέον, γίνεται αναφορά σε άλλους τύπους αρχείων ψηφιακού ήχου με αντίστοιχες δυνατότητες.

Στο **Κεφάλαιο 5** γίνεται υλοποίηση του ψυχοακουστικού μοντέλου με χρήση MATLAB.

## ΚΕΦΑΛΑΙΟ 2

# ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ ΑΚΟΥΣΤΙΚΗΣ ΕΠΙΚΑΛΥΨΗΣ – ΜΕΘΟΔΟΙ ΤΕΧΝΙΚΩΝ ΗΧΗΤΙΚΗΣ ΣΥΜΠΙΕΣΗΣ

## 2.1 ΕΙΣΑΓΩΓΗ

Παρουσιάζεται ένα σημαντικός αριθμός από ακουστικά μοντέλα επικάλυψης, ο σχεδιασμός των οποίων βασίζεται δυο εναλλακτικές μεθοδολογίες:

α) Υλοποίηση ακουστικών μοντέλων επικάλυψης που χρησιμοποιούν βασικές αρχές της νευρο- φυσιολογίας του ανθρώπινου ακουστικού συστήματος.

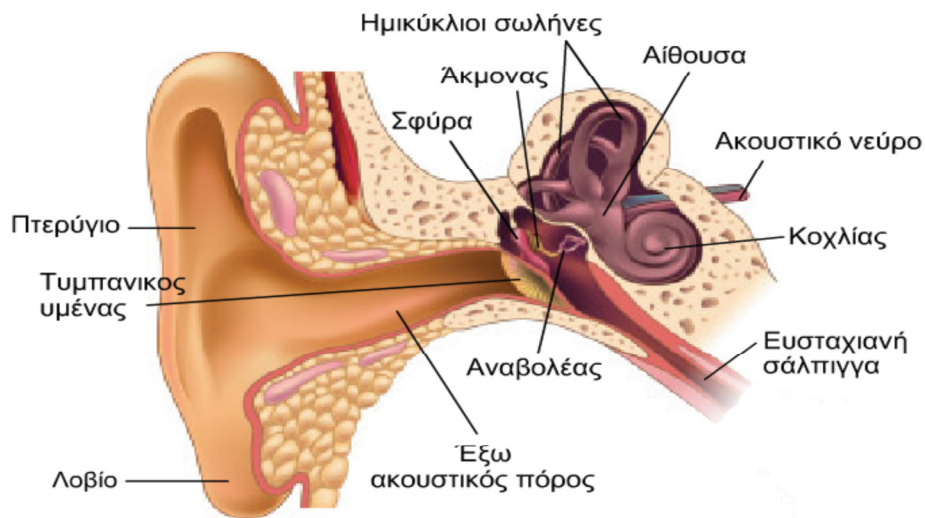
β) Υλοποίηση ακουστικών μοντέλων επικάλυψης που χρησιμοποιούν βασικές αρχές της ψυχοακουστικής.

Στα ακουστικά μοντέλα για την περίπτωση α) γίνεται κύρια χρήση της νευρο-φυσιολογίας του ακουστικού συστήματος, ενώ στην περίπτωση β) γίνεται κύρια χρήση των αποτελεσμάτων που προκύπτουν από ψυχοακουστικής μετρήσεις ή πειράματα. Σε σύγκριση με τις ψυχοακουστικής μεθόδους, που εννοούν την πλήρη ενσωμάτωση του ακουστικού συστήματος στην εκάστοτε μέτρηση, οι μέθοδοι που έχουν ως κύρια βάση την νευρο-φυσιολογία του ακουστικού συστήματος, επιτρέπουν κάθε φορά την εξέταση μονό συγκεκριμένων τμημάτων. Ακουστικά μοντέλα που είναι βασισμένα σε αποτελέσματα νευρο-φυσιολογικών μετρήσεων είναι πολύ λεπτομερή, αλλά ωστόσο δεν παρέχουν μια περιεκτική εικόνα της ακουστικής επεξεργασίας και επιπλέον είναι συχνά δύσκολο να ενσωματωθούν σε εφαρμογές που εμπίπτουν στην επιστημονική περιοχή του μηχανικού ήχου.

## 2.2 ΑΝΑΤΟΜΙΑ ΑΚΟΥΣΤΙΚΟΥ ΣΥΣΤΗΜΑΤΟΣ

Το ανθρώπινο σύστημα ακοής [Παρ95] αποτελείται από δύο υποσυστήματα: (α) το περιφερειακό, στο οποίο γίνεται η μετατροπή των ηχητικών κυμάτων σε ηλεκτροχημικά σήματα και (β) στο νευρικό, (που περιλαμβάνει και τα ανώτερα κέντρα ακοής του εγκεφάλου), στο οποίο γίνεται η κύρια επεξεργασία του παραγόμενου ηλεκτροχημικού σήματος και δημιουργείται το αίσθημα της ακοής.

Το περιφερειακό σύστημα αποτελείται από το εξωτερικό, μέσο και εσωτερικό αυτί, στα οποία γίνεται η προ-επεξεργασία του ακουστικού σήματος και ο μετασχηματισμός του σε ηλεκτροχημικό σήμα. Η δομή του συστήματος αυτού είναι σε γενικές γραμμές γραμμική, ωστόσο υπάρχουν και μη γραμμικά στοιχεία με αποτέλεσμα να δημιουργούνται μη-γραμμικά φαινόμενα στο ακουστικό σήμα. Στο σχήμα 2.1, δείχνεται η μορφή του περιφερειακού ακουστικού συστήματος του ανθρώπου.



Σχήμα 2.1: Ανατομία του ανθρώπινου αυτιού.

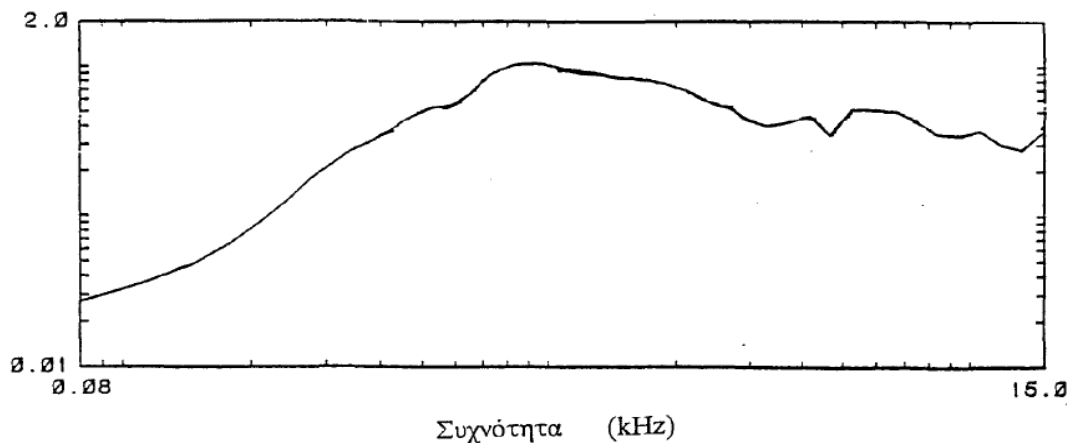
## 2.2.1 ΕΞΩΤΕΡΙΚΟ ΑΥΤΙ

Το εξωτερικό αυτί, αποτελείται από το πτερύγιο, το εξωτερικό ακουστικό κανάλι και τον ακουστικό πόρο και λειτουργεί σαν ένα γραμμικό ακουστικό φίλτρο. Ο προφανής του ρόλος είναι να συγκεντρώνει και να κατευθύνει την ενέργεια του ακουστικού κύματος προς το μέσο αυτί. Επιφυλάσσει όμως και έναν δεύτερο, πιο σημαντικό και περίπλοκο ρόλο, αυτόν του προσδιορισμού της θέσης στο χώρο μίας ακουστικής πηγής (source localization) [Παρ95]. Ο ακουστικός σωλήνας έχει διπλή χρησιμότητα: (α) προστατεύει το μέσο αυτί και το τύμπανο και (β) φέρνει το εσωτερικό αυτί πολύ κοντά στο εγκέφαλο, μειώνοντας το μήκος των νεύρων, εξασφαλίζοντας έτσι εύκολη διόδευση στα παραγόμενα ασθενή ηλεκτρικά σήματα προς τον εγκέφαλο. Ο ακουστικός σωλήνας επιδρά ισχυρά στην απόκριση

συχνότητας του συνολικού ακουστικού συστήματος. Συμπεριφέρεται σαν ένας ανοικτός σωλήνας συντονισμού με μέσο μήκος 2,25 cm, μέση διάμετρο 0,75 cm και τέταρτο μήκους κύματος, συχνότητας περίπου 4 kHz. Ενισχύει δηλαδή, τις συχνότητες στην περιοχή των 4 kHz και εξηγείται έτσι η υψηλή επιλεκτικότητα του αυτιού για ήχους που βρίσκονται στην περιοχή αυτή. Προς επιβεβαίωση των παραπάνω, το έντονο βύθισμα που παρατηρείται στο απόλυτο κατώφλι ακουστότητας (βλ. Ενότητα 2.4) εντοπίζεται σε αυτή την συχνοτική περιοχή.

### 2.2.2 ΜΕΣΟ ΑΥΤΙ

Το μέσο αυτί, ανατομικά αποτελείται από το τύμπανο, από τρία μικρά και πολύ σκληρά οστά, (σφύρα, άκμονας και αναβολέας), και από την ευσταγγιανή σάλπιγγα. Λειτουργεί σαν ακουστό-μηχανικός μετατροπέας, προσαρμόζοντας τις διαφορετικές εμπεδήσεις του αέρα και του λεμφικού υγρού που βρίσκεται στο εσωτερικό αυτί. Με τον όρο προσαρμογή εμπεδήσεων εννοούμε τη μετατροπή των ταλαντώσεων του αέρα, οι οποίες έχουν μικρές δυνάμεις και μεγάλες απομακρύνσεις, σε κινήσεις του λεμφικού υγρού, με μεγάλες δυνάμεις και μικρές απομακρύνσεις, χωρίς ενεργειακές απώλειες. Ο συντελεστής μετατροπής έχει υπολογιστεί σε 1 προς 20 [Παρ95] και οφείλεται στη σχέση μηκών των τριών οστών. Για τους άνδρες ο βέλτιστος λόγος μετασχηματισμού επιτυγχάνεται στην περιοχή 1.000 - 3.000 Hz. Γενικά, η απόκριση συχνότητας του μέσου αυτιού παρουσιάζει τη μορφή υψηλοδιαβατού φίλτρου με κλίση 9-12 dB/οκτάβα κάτω από τα 1.000 Hz και με ελαφριά ενίσχυση στην περιοχή των μεσαίων συχνοτήτων (σχήμα 2.2) , γεγονός που συντελεί στην αυξημένη ευαισθησία της ακοής σ' αυτήν την περιοχή.



Σχήμα 2.2: Απόκριση συχνότητας μέσου αυτιού

Μια άλλη σημαντική ιδιότητα του μέσου αυτιού, είναι η μη-γραμμική συμπεριφορά του για απότομα υψηλές στάθμες του ηχητικού κύματος, που αποσκοπεί στην προστασία του εσωτερικού αυτιού από επικίνδυνες υπερφορτώσεις. Τέλος, ο ρόλος της ευσταγγιανής σάλπιγγας είναι να εξισορροπεί την ατμοσφαιρική πίεση στις δύο πλευρές του τύμπανου, έτσι ώστε να διεγείρεται μόνο από ακουστικά σήματα και να μην επηρεάζεται από την ατμοσφαιρική πίεση.

Αρκετές ικανοποιητικές προσομοιώσεις του μέσου αυτιού έχουν αναπτυχθεί (κυρίως με ηλεκτρικά ανάλογα), καθώς επίσης και συσκευές κλινικών μετρήσεων στην ιατρική, οι οποίες εξετάζουν την τυχόν παθολογική λειτουργία της ακοής.

### 2.2.3 ΕΣΩΤΕΡΙΚΟ ΑΥΤΙ

Πρόκειται για το πιο ενδιαφέρον και περίπλοκο τμήμα του ακουστικού συστήματος. Μετατρέπει την μηχανική ενέργεια των δονήσεων, που φθάνουν από το τύμπανο και το μέσο αυτί, σε ηλεκτροχημικό σήμα. Αποτελείται από τον κοχλία και από τον λαβύρινθο, ο οποίος δεν είναι όργανο ακοής αλλά ισορροπίας. Ο κοχλίας έχει μήκος 32 mm, αποτελείται από 2,5 στροφές και είναι γεμάτος από δύο διαφορετικά οργανικά υγρά (ενδόλυμφος και περίλυμφος), τα οποία διαχωρίζονται από τη βασική μεμβράνη. Επάνω στη μεμβράνη βρίσκεται το όργανο του Corti, που συγκροτείται από τα τριχοειδή κύτταρα και μετατρέπει τις μηχανικές ταλαντώσεις της βασικής μεμβράνης σε ηλεκτρικό σήμα, που οδηγείται στο ακουστικό νεύρο. Ο ρόλος της βασικής μεμβράνης είναι θεμελιώδης για το μηχανισμό της ακοής, αφού επιτελεί τη συχνотική ανάλυση του ηχητικού σήματος. Λόγω της μεταβλητής ως

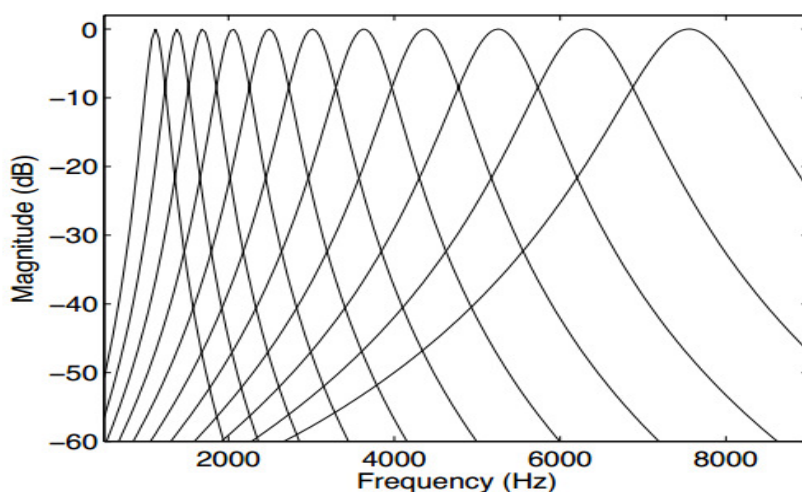


προς το μήκος διατομής της, εμφανίζει σκληρότητα που μειώνεται εκθετικά, με αποτέλεσμα να συντονίζεται σε διαφορετικά σημεία για διαφορετικές συχνότητες. Η ιδιότητα αυτή προσδίδει στη βασική μεμβράνη το χαρακτηρισμό "αναλυτής συχνοτήτων". Για πολύ χαμηλές συχνότητες η ταλάντωση της βασικής μεμβράνης μηδενίζεται στο τέρμα του κοχλία, ενώ για υψηλότερες ωριότερα. Αν στο σήμα υπάρχουν περισσότερες από μία συχνότητες, τότε θα σχηματισθούν μέγιστα πλάτους σε περισσότερες θέσεις.

Το όργανο του Corti [Παρ95] περιέχει τα πολύ σημαντικά αισθητήρια κύτταρα (hair-cells), που καταλήγουν σε ένα τριχίδιο. Οι κάθετες απομακρύνσεις της βασικής μεμβράνης μετατρέπονται σε ακτινικές κινήσεις των τριχιδίων, μεταβάλλοντας έτσι την ηλεκτρική αντίσταση του κυττάρου, το οποίο με ηλεκτροχημικές διαδικασίες, παράγει παλμούς ηλεκτρικού ρεύματος, που μεταφέρονται στον εγκέφαλο από έναν μεγάλο αριθμό (περίπου 20 για κάθε αισθητήρα), νευρικών ινών. Το φαινόμενο της δημιουργίας των παλμοσειρών αυτών, μπορεί να χαρακτηριστεί σαν "διαδικασία ανόρθωσης". Το πλήθος των παλμών ανά δευτερόλεπτο καθορίζεται από τη στάθμη ηχητικής πίεσης του διεγείροντος σήματος και προσδιορίζει την ένταση του υποκειμενικού αισθήματος της ακοής.

Ο αριθμός των αισθητηρίων κυττάρων είναι 1.500 έως 2.000, ενώ ο συνολικός αριθμός των νευρικών ινών είναι περίπου 30.000. Κάθε αισθητήριο κύτταρο κωδικοποιεί μία στενή περιοχή συχνοτήτων της τάξης των μερικών εκατοντάδων Hz, ενώ περισσότερο πολύπλοκη κωδικοποίηση γίνεται από τις νευρικές ίνες και στη συνέχεια στον ακουστικό φλοιό από τα ανώτερα ακουστικά κέντρα του εγκεφάλου. Πρέπει να σημειωθεί ότι υπάρχουν δύο ειδών αισθητήρια κύτταρα, τα οποία δρουν "ανταγωνιστικά", όπως επίσης και δύο ειδών νευρικές ίνες, που εξασφαλίζουν αμφίδρομη επικοινωνία με τον εγκέφαλο, δημιουργώντας έτσι ένα περίπλοκο μη-γραμμικό σύστημα με ανάδραση. Μη-γραμμική συμπεριφορά πιστεύεται ότι έχουν και τα αισθητήρια κύτταρα, εξαρτώμενη από την πίεση του ακουστικού σήματος, γεγονός που εξηγεί τη μεγάλη δυναμική περιοχή και την πολύ λεπτή επιλεκτικότητα συχνοτήτων του αυτιού, της τάξης των 3 Hz στα 1.000 Hz.

Όπως προέκυψε από την παραπάνω συζήτηση και για τους σκοπούς της ανάλυσης ακουστικών σημάτων, η λειτουργία του εσωτερικού αυτιού μοντελοποιείται ικανοποιητικά από μία τράπεζα παράλληλων ζωνοδιαβατών φίλτρων, όπως φαίνεται στο σχήμα 2.3.



Σχήμα 2.3: Προσομοίωση της απόκρισης συχνότητας του εσωτερικού αυτιού με χρήση ζωνοδιαβατών φίλτρων.

## 2.3 ΚΡΙΣΙΜΕΣ ΠΕΡΙΟΧΕΣ ΣΥΧΝΟΤΗΤΑΣ

Επιπλέον έρευνες έχουν αποδείξει την ύπαρξη κρίσιμων περιοχών συχνότητας (Critical Bands - CB) [Παρ95], για την αίσθηση του ακουστικού ερεθίσματος από τον άνθρωπο, καθώς επίσης και τη σχέση ανάμεσα στις περιοχές αυτές και στα μέγιστα ταλάντωσης της βασικής μεμβράνης. Μέσα σε μία κρίσιμη περιοχή συχνότητας το επίπεδο ακουστότητας εξαρτάται μόνο από την ενεργή (RMS) τιμή της ακουστικής πίεσης του ηχητικού σήματος. Μία κρίσιμη περιοχή συχνότητας αντιστοιχεί σε απόσταση 1,3 mm επάνω στη βασική μεμβράνη και ονομάζεται επίσης και κλίμακα bark . Η ιδέα των critical bands προτάθηκε αρχικά από τον Fletcher (1940). Αυτός υπέθεσε ότι το τμήμα του φάσματος του θορύβου που επικαλύπτει έναν τόνο, είναι μέρος του φάσματος που βρίσκεται κοντά στον τόνο. Επικάλυψη επιτυγχάνεται όταν η ισχύς ενός τόνου και η ισχύς ενός τμήματος από ένα φάσμα θορύβου που βρίσκεται κοντά στον τόνο, επικαλύπτονται.

Κρίσιμες περιοχές συχνοτήτων ορίζονται να έχουν τέτοιο εύρος ζώνης συχνοτήτων έτσι ώστε να παράγεται η ίδια ακουστική ισχύς από τον τόνο και από τον θόρυβο, ο οποίος μόλις επικαλύπτει τον τόνο. Η υπόθεση αυτή του Fletcher, μας δίνει τη δυνατότητα να καθορίσουμε το εύρος κάθε κρίσιμης περιοχής συχνότητας, ωστόσο έχουν αναπτυχθεί και άλλες τεχνικές υπολογισμού, που προσφέρουν περισσότερη ακρίβεια. Στον Πίνακα 2.1, δείχνεται μία λίστα με τις κρίσιμες περιοχές, όπου κάθε μία απ' αυτές ορίζεται από την κεντρική της συχνότητα και το εύρος ζώνης.

Αριθμός Κρίσιμης Περιοχής Συχνότητας (Bark)	Κεντρική Συχνότητα (Hz)	Εύρος Ζώνης (Hz)	Αριθμός Κρίσιμης Περιοχής Συχνότητας (Bark)	Κεντρική Συχνότητα (Hz)	Εύρος Ζώνης (Hz)
1	50	100	14	2.150	320
2	150	100	15	2.500	380
3	250	100	16	2.900	450
4	350	100	17	3.400	550
5	450	110	18	4.000	700
6	570	120	19	4.800	900
7	700	140	20	5.800	1.100
8	840	150	21	7.000	1.300
9	1.000	160	22	8.500	1.800
10	1.170	190	23	10.500	2.500
11	1.370	210	24	13.500	3.500
12	1.600	240	25	19.500	8.000
13	1.850	280			

**Πίνακας 2.1 Χαρακτηριστικά μεγέθη των κρίσιμων περιοχών συχνότητας.**

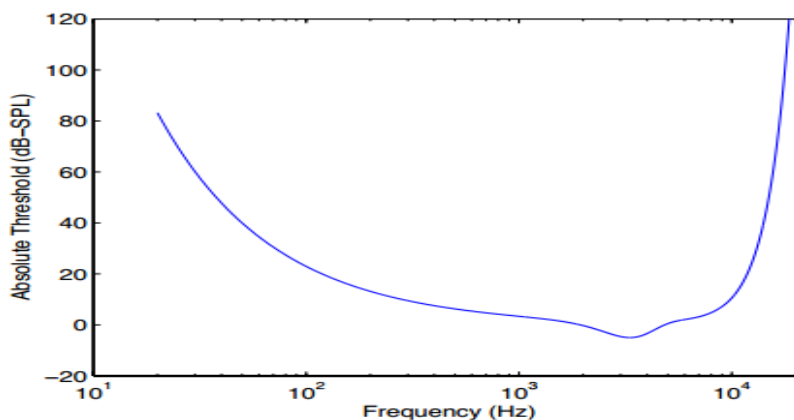
Χρησιμοποιώντας όρους επεξεργασίας σήματος, οι κρίσιμες περιοχές συχνότητας είναι μία τράπεζα φίλτρων, που διαχωρίζει το φάσμα συχνοτήτων σε περιοχές με διαφορετικό εύρος ζώνης. Η κεντρική συχνότητα και το εύρος ζώνης κάθε κρίσιμης περιοχής συχνότητας έχουν επιλεγεί έτσι ώστε να προσεγγίζουν τη λειτουργία του ανθρώπινου ακουστικού συστήματος. Έχει δοθεί μεγαλύτερη ανάλυση στις χαμηλές συχνότητες, οι οποίες έχουν αναλυθεί σε περισσότερες κρίσιμες περιοχές συχνότητας και λιγότερη στις υψηλές συχνότητες, επειδή έτσι γίνεται και από το ανθρώπινο ακουστικό σύστημα.

## 2.4 ΑΠΟΛΥΤΟ ΚΑΤΩΦΛΙ ΑΚΟΥΣΤΟΤΗΤΑΣ

Το απόλυτο κατώφλι ακουστότητας υποδηλώνει συναρτήσει της συχνότητας, τη στάθμη της ηχητικής πίεσης ενός θορύβου στενής ζώνης ή ενός τονικού σήματος η οποία είναι μόλις ακουστή. Συνεπώς, το απόλυτο κατώφλι ακουστότητας μπορεί να θεωρηθεί ως το κατώτερο όριο της δυναμικής περιοχής του ανθρωπινού ακουστικού συστήματος. Εκτός από την έντονη συχνοτική εξάρτηση που παρουσιάζει το απόλυτο κατώφλι ακουστότητας, εξαρτάται από το εύρος ζώνης υπό εξέταση σήματος καθώς και από την χρονική του διάρκεια. Δεδομένου ότι το κατώφλι ακουστότητας προσδιορίζεται μονοσήμαντα για κάθε μια ανθρωπινή οντότητα (λόγω της ξεχωριστής συχνοτικής εξάρτησης), πρακτικά χρησιμοποιείται ως αναφορά μια μέση τιμή του απόλυτου κατωφλίου ακουστότητας για ένα μεγάλο αριθμό δειγμάτων (ακροατών) [Παρ95]. Σε μια προσπάθεια να εξηγηθεί ο μηχανισμός του απόλυτου κατωφλίου ακουστότητας με βάση τη φυσιολογία του ανθρώπινου ακουστικού συστήματος, η λειτουργία του προσομοιώνεται ως ένα είδος θορύβου που υπερτίθεται στο υπό εξέταση σήμα και εν γένει το επικαλύπτει. Υπό αυτή την έννοια, η αναπτυσσόμενη εσωτερικά διέγερση των ακουστικών νεύρων αποτελεί την κύρια πηγή θορύβου, η οποία είναι ομοιόμορφα κατανομημένη σε όλο το συχνοτικό φάσμα που καλύπτει το ακουστικό σύστημα. Με βάση την ακόλουθη αναλυτική έκφραση όπου  $f$  σε kHz,

$$L_{TQ} = 3.64 \cdot f^{-0.8} - 6.5 \cdot e^{0.6 \cdot (f-3.3)^2} + f^4 \cdot 10^{-3}$$

το απόλυτο κατώφλι ακουστότητας ( $L_{TQ}$  σε dB) φαίνεται στο Σχήμα 2.4.



Σχήμα 2.4: Το απόλυτο κατώφλι ακουστότητας συναρτήσει της συχνότητας.

## 2.5 ΦΑΙΝΟΜΕΝΟ ΑΚΟΥΣΤΙΚΗΣ ΕΠΙΚΑΛΥΨΗΣ

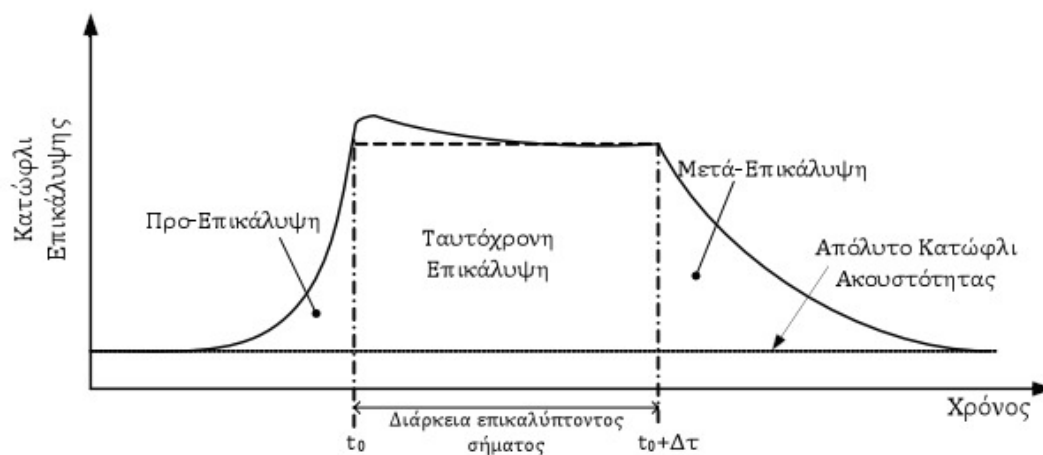
Επάνω στην βασική μεμβράνη επιτελείται διαχωρισμός των φασματικών συνιστωσών του ακουστικού σήματος, οφειλόμενος στη μεταβαλλόμενη σκληρότητα της. Έτσι, ακουστικό σήμα μίας συχνότητας (καθαρός τόνος), διεγείρει μία συγκεκριμένη περιοχή της μεμβράνης, με πλάτος απομάκρυνσης ανάλογο με την ένταση του σήματος. Όπως είναι φυσικό η απομάκρυνση αυτή της βασικής μεμβράνης "απλώνεται" και σε γειτονικές περιοχές, οι οποίες αντιστοιχούν σε άλλες συχνότητες. Αν ταυτόχρονα με τα προηγούμενα σήμα, διεγείρει το αυτί και άλλο ακουστικό σήμα, με παραπλήσια συχνότητα, αλλά με μικρότερη ένταση, τότε η διασπορά της ταλάντωσης της βασικής μεμβράνης, που οφείλεται στο ισχυρότερο σήμα θα υπερκαλύψει την ταλάντωση της μεμβράνης, που οφείλεται στο λιγότερο ισχυρό σήμα. Έτσι, ο δεύτερος ήχος, παρά το γεγονός ότι διεγείρει το αυτί δεν γίνεται ακουστός ή μειώνεται η ακουστότητα του. Το φαινόμενο αυτό ονομάζεται φαινόμενο επικάλυψης (masking effect) [Zap10] και εμφανίζει σημαντικό ενδιαφέρον, ειδικά σε εφαρμογές συμπίεσης ακουστικών δεδομένων. Πρόκειται για ένα φαινόμενο γνώριμο σε όλους από την καθημερινή εμπειρία, αφού η ακουστότητα ενός σήματος (π.χ. ομιλίας) μπορεί να μειωθεί ή να εξαφανισθεί εντελώς, όταν εμφανισθεί ταυτόχρονα ένας άλλος ισχυρότερος ήχος (π.χ. θόρυβος) με παρόμοιο συχνотικό περιεχόμενο.

Μιλώντας με όρους επεξεργασίας σήματος, το φαινόμενο αυτό περιγράφεται σαν η μεταβολή της απόκρισης συχνότητας από τη μορφή στενής ζώνης, σε άλλη ευρύτερης ζώνης που επικαλύπτει γειτονικές φασματικές περιοχές. Ανάλογα με την χρονική θέση του υπό-εξέταση σήματος (test signal) σε σχέση με το επικαλύπτον σήμα (masker), το φαινόμενο της ακουστικής επικάλυψης διαχωρίζεται σε τρεις βασικές κατηγορίες (βλ. Σχήμα 2.5):

α) *ταυτόχρονη επικάλυψη* (simultaneous masking): Στην περίπτωση αυτή το υπό-εξέταση καθώς και το επικαλύπτον σήμα παρουσιάζονται ταυτόχρονα στα πλαίσια μιας ελεγχόμενης ακρόασης (listening test). Ιδιαίτερη σημασία για το συγκεκριμένο φαινόμενο έχει η ασυμμετρία της επικάλυψης (asymmetry of masking) που σχετίζεται με την διαφορετική ικανότητα επικάλυψης που παρουσιάζει ένα σήμα θορύβου στενής ζώνης σε σχέση με ένα τονικό σήμα. Πιο συγκεκριμένα, ένα σήμα θορύβου στενής ζώνης με κεντρική συχνότητα αυτή ενός τονικού σήματος είναι πιο αποτελεσματικό (σε σχέση με το τονικό σήμα) στην επικάλυψη ενός υπό-εξέταση τονικού σήματος με παρόμοιο συχνотική περιεχόμενο.

β) *μετά- επικάλυψη* (forward masking): Για την δημιουργία του φαινομένου της μετά-επικάλυψης το υπό εξέταση σήμα παρουσιάζεται αμέσως μετά την παύση του επικαλύπτοντος σήματος. Με την ενεργεία αυτή, το δημιουργηθέν κατώφλι επικάλυψης δεν μειώνεται απότομα, αλλά μεσολαβεί ένα χρονικό διάστημα (της τάξης των μερικών εκατοντάδων ms) έως ότου εκφυλιστεί το απόλυτο κατώφλι ακουστότητας. Διαφορές παράμετροι που διαμορφώνουν ή καθορίζουν το δημιουργηθέν κατώφλι επικάλυψης είναι: η χρονική καθυστέρηση του υπό εξέταση σήματος, η χρονική διάρκεια, η φάση και το συχνοτικό περιεχόμενο του υπό εξέταση σήματος καθώς και η συχνοτική διασπορά των δυο σημάτων.

γ) *προ-επικάλυψη* (backward masking): Το υπό-εξέταση σήμα εμφανίζεται λίγο πριν την έναρξη του επικαλύπτοντος σήματος για την δημιουργία του φαινομένου της προ-επικάλυψης. Παρά το γεγονός ότι σημαντικές έρευνες σχετικά με το φαινόμενο της προ-επικάλυψης έχουν παρουσιαστεί τα τελευταία χρόνια, ο μηχανισμός λειτουργίας της συγκεκριμένης διαδικασίας δεν έχει κατανοηθεί σε βάθος. Σε αντίθεση με την ταυτόχρονη και τη μετά- επικάλυψη η διαδικασία της προ-επικάλυψης εξαρτάται σε σημαντικό βαθμό από την εμπειρία των συμμετεχόντων ακροατών σε σχετικές μετρήσεις ή ελεγχόμενες ακροάσεις. Διάφοροι ερευνητές έχουν επισημάνει ότι το φαινόμενο της προ-επικάλυψης μπορεί να παρατηρηθεί για ένα χρονικό διάστημα 20ms (μεγίστη τιμή) πριν την παρουσία του επικαλύπτοντος σήματος. Επιπλέον, το φαινόμενο της προ-επικάλυψης καθορίζεται σε σημαντικό βαθμό από ανώτερες εγκεφαλικές λειτουργίες.



Σχήμα 2.5: Σχηματικό διάγραμμα για την απεικόνιση και το χαρακτηρισμό χρονικών περιοχών για τις διάφορες περιπτώσεις της ακουστικής επικάλυψης.

## 2.6 ΣΥΝΟΨΗ ΚΕΦΑΛΑΙΟΥ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό αναλύθηκε η ανατομία του ανθρώπινου ακουστικού συστήματος. Συγκεκριμένα, με τη χρησιμοποίηση των εξελιγμένων μοντέλων που περιγράφουν λειτουργίες του εσωτερικού ή του εξωτερικού αυτιού, επιτυγχάνεται βελτιστοποίηση των ψηφιακών ηχητικών συστημάτων, αφού η λειτουργία τους προσεγγίζει τον τρόπο που ο ανθρώπινος δέκτης αναλύει το ηχητικό σήμα. Είναι προφανές ότι με την τελειοποίηση του μοντέλου και το φαινόμενο της επικάλυψης προβλέπεται ανάπτυξη σε διάφορες τομείς όπως για παράδειγμα σε ιατρικές εφαρμογές, στην ακριβέστερη αντικειμενική αξιολόγηση της πιστότητας των ηχητικών συστημάτων, κλπ.

## ΚΕΦΑΛΑΙΟ 3

### ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ ΜΕ ΧΡΗΣΗ ΨΥΧΟΑΚΟΥΣΤΙΚΟΥ ΜΟΝΤΕΛΟΥ

#### 3.1 ΣΤΑΔΙΑ ΨΗΦΙΟΠΟΙΗΣΗΣ

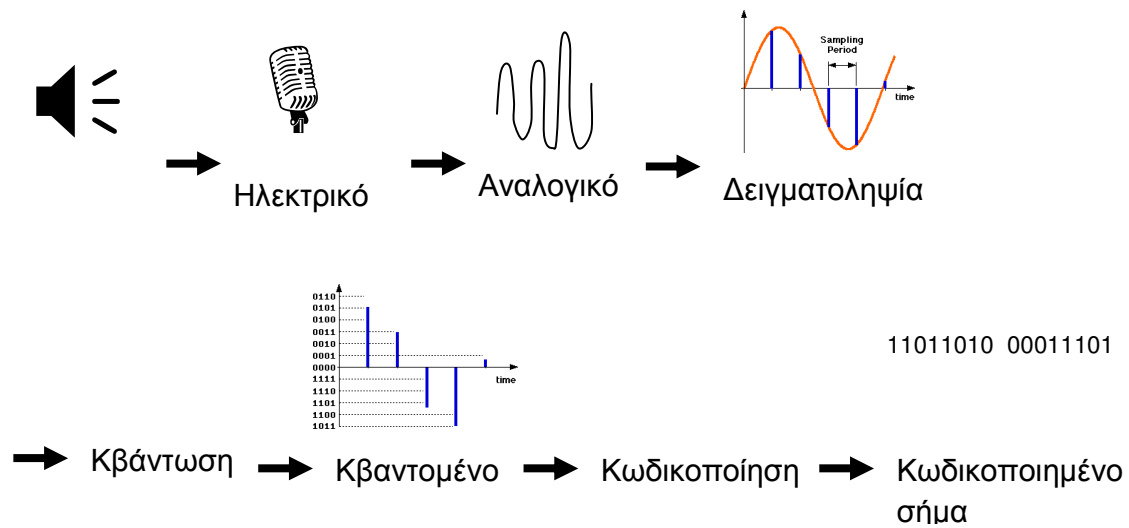
Κατά την δεκαετία του '90 η δύναμη της ψηφιακής τεχνολογίας (H/Y) πραγματοποιεί την επεξεργασία, την αποθήκευση και την μεταφορά ενός αντικειμένου. Χαρακτηριστικό παράδειγμα είναι το δυαδικό σύστημα. Το δυαδικό σύστημα δηλώνει ότι οι πληροφορίες που εισάγουμε στο H/Y μετατρέπονται σε 2 ψηφία ,το 0 και το 1 (bits) αντιπροσωπεύοντας την παροχή η μη παροχή ρεύματος στον υπολογιστή καθώς είναι η μόνη λειτουργία που αντιλαμβάνεται ο H/Y. Αυτό έχει σαν αποτέλεσμα να διευκολύνει τον χρήστη τόσο στην επεξεργασία όσο και στην αποθήκευση εικόνων αλλά και μεταφορά κειμένου.

Για να πραγματοποιηθεί η ψηφιοποίηση είναι αναγκαία η μετατροπή από αναλογικό (analog) σε ψηφιακό (digital) με την χρήση των μετατροπέων ADC (Analog to Digital Converters) και DAC (Digital to Analog Converters) [Χρυ06].Οι βασικές λειτουργίες της ψηφιοποίησης είναι τρεις. Η δειγματοληψία αρχικού σήματος , ο κβαντισμός των τιμών του σήματος διακριτού χρόνου που προκύπτει από την δειγματοληψία και η κωδικοποίηση. Ειδικότερα κατά την δειγματοληψία λαμβάνονται δείγματα του πλάτους του αναλογικού σήματος σε τακτά χρονικά διαστήματα με αποτέλεσμα να δημιουργείται ένα στιγμιότυπο του σήματος της συγκεκριμένης τιμής , συνόλου του οποίου τα δείγματα διαμορφώνουν το ψηφιοποιημένο σήμα. Σε αντίθεση με την δειγματοληψία στον κβαντισμό οι διαδοχικές τιμές της στάθμης (πλάτους) του σήματος του διακριτού χρόνου  $s_d$  (nTs) μετατρέπονται σε διακριτές (ψηφικές) τιμές , πραγματοποιείται δηλαδή μια απεικόνιση της μορφής  $S(nTs) = Q\{s_d(nTs)\}$  όπου  $Q\{s_d(nTs)\}$  είναι η κβαντισμένη τιμή. Σύμφωνα με την μετατροπή το σήμα είναι διακριτό τόσο ως προς τον χρόνο όσο και ως προς το πλάτος που καλείται ψηφιακό.

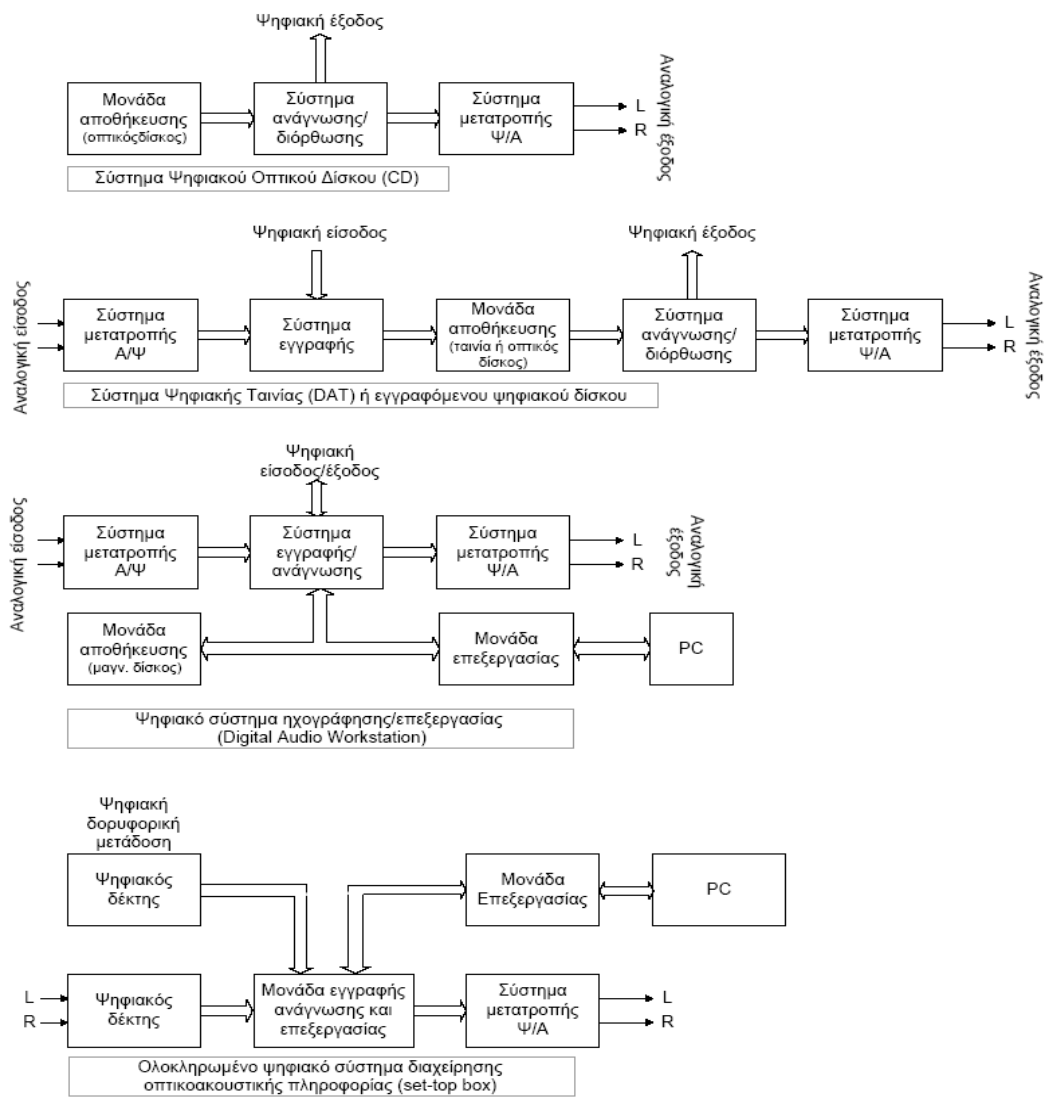
Το αναλογικό σήμα ακουστικής πίεσης μετατρέπεται σε ψηφιακό με την χρήση υποσυστημάτων εφαρμογής Α/Ψ (Analog to Digital- Converters) και το αντίστροφο υλοποιείται με την χρήση Ψ/Α μετατροπών (Digital to Analog- Converters DACs) όπου οι κβαντισμένες τιμές μετατρέπονται σε αναλογικές. Επειδή οι διαδικασίες κβαντισμού και δειγματοληψίας σε τέτοια υποσυστήματα πραγματοποιούνται ταυτόχρονα η υλοποίηση των μετατροπέων και στις δύο φάσεις παρουσιάζουν τόσο θεωρητικά όσο και πρακτικά



προβλήματα. Για να υπάρξει σταθερή μετατροπή είναι απαραίτητη η ικανοποίηση συγκεκριμένων συνθηκών δειγματοληψίας και κβαντισμού όπως αναφέρεται παρακάτω.



Σχήμα 3.2: Διαδικασία Ψηφιοποίησης Ηχητικού Σήματος.

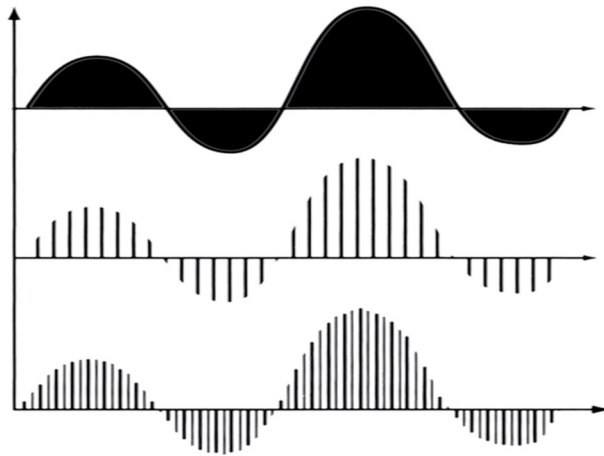


Σχήμα 3.2: Διαδικασία Ψηφιοποίησης Ηχητικού σήματος

### 3.1.1 ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Δειγματοληψία είναι η μετατροπή ενός αναλογικού σήματος  $sc(t)$  (σήμα του οποίου η ανεξάρτητη χρονική μεταβλητή  $t$  παίρνει συνεχείς τιμές) σε διακριτό σήμα [Χρυ06]. Συγκεκριμένα έχει την εξής μορφή  $sc(t) \rightarrow sd(nTs)$  όπου  $n=1,2 \dots$  ο αριθμός του δείγματος και  $Ts$  (sec) η περίοδος της Δειγματοληψίας. Χαρακτηριστικό ρόλο παίζει η περίοδος της δειγματοληψίας στην μετατροπή ενός σήματος συνεχούς χρόνου σε διακριτού χρόνου. Το δείγμα αναφέρεται σε 1 τιμή του συνεχούς σήματος σε μια συγκεκριμένη χρονική στιγμή. Δειγματολήπτης ορίζεται ως ένα σύστημα το οποίο εξάγει δείγματα (σε ίσα χρονικά διαστήματα) από ένα συνεχές σήμα.

Η συχνότητα (Hz μέτρηση) δειγματοληψίας υποδηλώνει πόσα δείγματα έχουν ληφθεί από τον δειγματολήπτη σε διάρκεια ενός δευτερολέπτου. Όσο μεγαλύτερη είναι η συχνότητα ψηφιοποίηση τόσο καλύτερη ποιότητα αναπαραγωγής έχουμε αλλά ταυτόχρονα με ριζική αύξηση του χώρου αποθήκευσης.



Σχήμα 3.3: (α) Αναλογικό σήμα, (β) Ψηφιακό με χαμηλή συχνότητα, (γ) Ψηφιακό με υψηλή συχνότητα δειγματοληψίας.

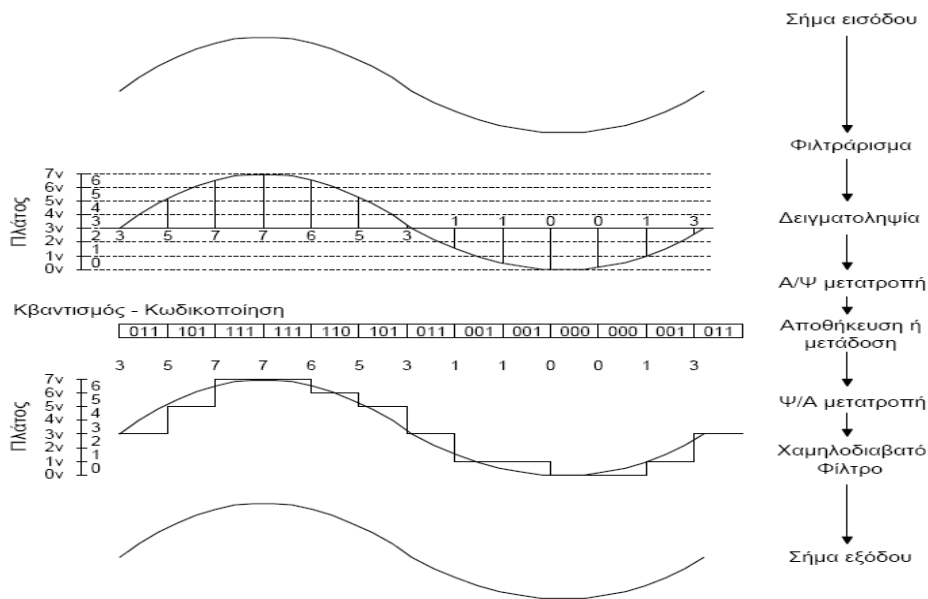
Σύμφωνα με το θεώρημα Nyquist απαραίτητη προϋπόθεση για τη μετατροπή ενός αναλογικού ηχητικού σήματος σε ψηφιακό είναι η δειγματοληψία. Είναι αναγκαίο η συχνότητα δειγματοληψίας να είναι μεγαλύτερη ή ίση του διπλάσιου της μέγιστης συχνότητας του αρχικού αναλογικού σήματος,  $f_s > 2f$  ή  $T_s \leq T/2$ . Για να ισχύει το παραπάνω πρέπει το αρχικό σήμα να αναζωπυρωθεί στο μισό της συχνότητας δειγματοληψίας περνώντας από ένα ιδανικό βαθυπέρατου φίλτρου.

Οι συχνότητες δειγματοληψίας που χρησιμοποιούνται σήμερα για την ψηφιακή εγγραφή μουσικής, κυμαίνονται στα 22 με 44 KHZ. Επιτρέποντας να περάσει ένα φάσμα ακουστικών συχνοτήτων από 11 έως 22 KHZ. Κατά συνέπεια ο ήχος που γίνεται αντιληπτός από το ανθρώπινο αυτί δεν ξεπερνά τα 18 με 22 KHZ γι' αυτό θα πρέπει να δειγματολογούμε με συχνότητα περίπου 44 KHZ (CD) με επακόλουθο ένα πιστό ψηφιακό αντίγραφο του αναλογικού σήματος.

### 3.1.2 ΠΡΟΒΛΗΜΑΤΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Η ανακάλυψη του επιστήμονα H.Nyquist [Χρυ06] ότι η μέγιστη συχνότητα αναλογικού σήματος μπορεί να αποδοθεί χωρίς αλλοίωση είναι το μισό της συχνότητας δειγματοληψίας ωθεί το φαινόμενο της αναδίπλωσης (aliasing). Συγκεκριμένα συχνότητες μεγαλύτερες από τα μισά της συχνότητας ψηφιοποίησης εμφανίζονται λανθασμένες σαν χαμηλές συχνότητες με επακόλουθο το φαινόμενο αυτό του ελλιπούς ρυθμού δειγματοληψίας.

Με στόχο την αποκοπή των συχνοτήτων μεγαλύτερων από την μέγιστη συχνότητα δειγματοληψίας για αποφυγή του φαινομένου της αναδίπλωσης, το ηλεκτρονικό σήμα μετά το στάδιο της προενίσχυσης τροφοδοτείται σε ένα ιδανικό βαθυπέρατο φίλτρο ( low-pass ). Στην συνέχεια κατά το στάδιο της δειγματοληψίας τροποποιείται το αναλογικό ηλεκτρικό σήμα της εισόδου σε δυαδικούς αριθμούς και αποθηκεύεται στη μνήμη του Η/Υ, ολοκληρώνοντας έτσι το μετασχηματισμό από το αναλογικό (analog) σε ψηφιακό (digital).



Σχήμα 3.4: Διαδικασία μετατροπής αναλογικού σε ψηφιακή μορφή και αντίστροφα.

Σε αντίθεση με την διαδρομή της καταγραφής η απόδοση ενός ψηφιακού ηχητικού σήματος ακολουθεί αντίστροφη διαδρομή βημάτων. Δηλαδή η τροφοδότηση του σήματος υλοποιείται από τη μνήμη του Η/Υ σε ένα μετατροπέα ψηφιακού σε αναλογικού σήματος. Όσον αφορά τον τρόπο λειτουργίας και τη συχνότητα δειγματοληψίας ο μετατροπέας παράγει ένα ηλεκτρικό σήμα που αποτελεί μια προσεγγιστική μορφή του ηλεκτρικού σήματος το οποίο ενισχύεται και στη συνέχεια τροφοδοτείται στην είσοδο του ηχείου, αποδίδοντας το σήμα.

Τα προβλήματα κατά την ψηφιοποίηση του ήχου οφείλονται στο γεγονός πως δεν υπάρχει ένα τέλειο βαθυπέρατο φίλτρο που να μπορεί να αποκόψει τέλεια συχνότητες πάνω από ένα επιθυμητό όριο, ενώ η χρήση τέτοιου φίλτρου διαμορφώνει μικρές παραμορφώσεις στο σήμα στο οποίο και εκφράζεται.

Επιπροσθέτως κατά την διαδικασία της Ψ/Α μετατροπής εξαιτίας των κβαντισμένων τιμών πλάτος του αναπαραγόμενου σήματος, διαμορφώνονται συχνότητες υψηλότερες της μέγιστης συχνότητας όπου και φιλτράρονται με ένα παρόμοιο φίλτρο εξομάλυνσης (smoothing filter).

### 3.1.3 ΚΒΑΝΤΟΠΟΙΗΣΗ

Το επόμενο βήμα με μεγάλο καθορισμό στην ποιότητα του ήχου είναι αυτό της κβάντωσης, προκειμένου να αναπαρασταθεί κάθε δείγμα με την μορφή μιας δυαδικής σειράς από bits καθώς πρέπει στο συνεχώς μεταβαλλόμενο πλάτος της τάσης του αναλογικού σήματος να τεθεί μία διακριτή τιμή.

Η κβάντωση σε συνδυασμό με την δειγματοληψία δημιουργεί την ψηφιοποίηση, καθώς είναι συμπληρωματικές διαδικασίες [Χρυ06]. Η κβαντοποίηση συμβάλλει στον σχεδιασμό του συστήματος του ψηφιακού ήχου ενώ το format του ψηφιακού ήχου διαμορφώνεται ανάλογα με τις συνθήκες κβάντωσης. Βασική παράμετρος της φάσης αυτής είναι το μέγεθος του δείγματος (samplingsize) με αποτέλεσμα όσα περισσότερα bits χρησιμοποιούνται να δημιουργείται μεγαλύτερη ακρίβεια στην περιγραφή.

Για παράδειγμα αν έχουμε 8 bits τότε μπορούν να χρησιμοποιηθούν  $2^8=256$  διαφορετικές τιμές, ενώ για 16 bits μπορούν να χρησιμοποιηθούν  $2^{16}= 65.536$  διαφορετικές τιμές, επομένως μεγαλύτερη πιστότητα και ακρίβεια.

Ένα χαρακτηριστικό πρόβλημα κατά την ηχογράφηση και την αναπαραγωγή είναι ο θόρυβος της κβαντοποίησης. Το φαινόμενο αυτό προέρχεται από το γεγονός ότι τα δείγματα του αναλογικού σήματος θα χρησιμοποιηθούν ακέραια, άρα η μετατροπή τους σε ψηφιακά δείγματα συνεπάγεται στην στρογγυλοποίηση του δείγματος. Είναι αναγκαίο να επισημανθεί ότι στην περίπτωση αυτή το ψηφιοποιημένο σήμα απαρτίζεται από άθροισμα του αναλογικού σήματος καθώς και από ένα σήμα θορύβου το οποίο και μετριέται σε αρνητικά decibel (db = μονάδα μέτρησης έντασης του ήχου )

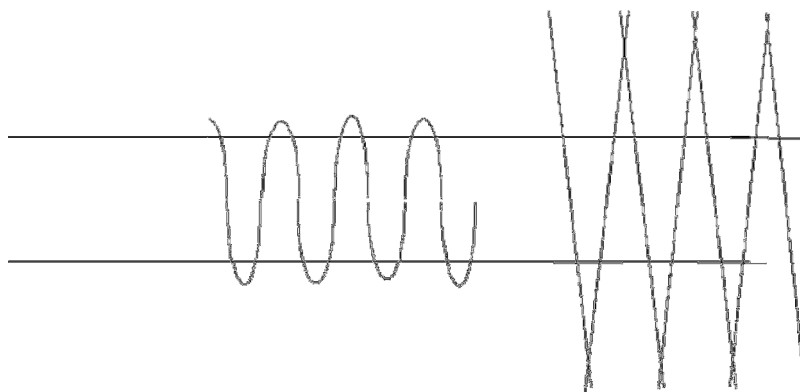
Το σφάλμα αυτό εκφράζεται από τον λόγο σήματος προς θόρυβο (Signal to Noise Ratio, SNR )  $S/E \text{ (db)} = 6.02n + 176$  όπου το n είναι ο αριθμός των bits της λέξης στο format του ψηφιακού ήχου. Όσο μεγαλύτερος είναι ο αριθμός των db τόσο μικρότερος είναι ο θόρυβος, ενώ τα περισσότερα συστήματα σήμερα κυμαίνονται στα 90db SNR όπου η παραμόρφωση μετριέται επί τοις εκατό του αρχικού σήματος. Η ευκρίνεια του κβαντισμού και του λόγου σήματος προς θόρυβο παρουσιάζεται στον παρακάτω πίνακα όπου ενδεικτικές τιμές SNR αντιστοιχούν σε διαφορετικές τάξεις μετατροπής.

Τάξη N(bits)	SNR (db)
8	49.8
12	73.8
16	97.8
18	109.8

**Πίνακας 1 Ευκρίνειας Κβαντισμού και SNR.**

Συμπερασματικά, ο θόρυβος που παράγουν τα ψηφιακά ηχητικά συστήματα, εμφανίζεται κατά την διαδικασία αναπαραγωγής του κβαντισμένου ηχητικού σήματος, έτσι ώστε να διαμορφώνεται ένας ανεπιθύμητος χαρακτήρας.

Ένας άλλος σημαντικός παράγοντας που επηρεάζει την ποιότητα του ήχου, είναι ο ψαλιδισμός (clipping) του σήματος, όπου κατά την εγγραφή του σήματος η μέγιστη στάθμη του σήματος εισόδου υπερβαίνει την μέγιστη στάθμη κβαντισμού διαμορφώνοντας κάποια διαστρέβλωση. Συγκεκριμένα, απεικονίζεται ένα ψαλιδισμένο σήμα με κοψίματα στην κορυφή και στη βάση με επακόλουθο ο ψαλιδισμός να μειώνεται αφού μειωθεί το μέγεθος του σήματος. Όπως φαίνεται παρακάτω:



**Σχήμα 3.5: Ψαλιδισμός (Clipping).**

## 3.2 ΑΛΓΟΡΙΘΜΟΙ ΣΥΜΠΙΕΣΗΣ ΣΤΟ ΠΕΛΙΟ ΤΟΥ ΧΡΟΝΟΥ-ΜΕΘΟΔΟΙ ΚΩΔΙΚΟΠΟΙΗΣΗΣ

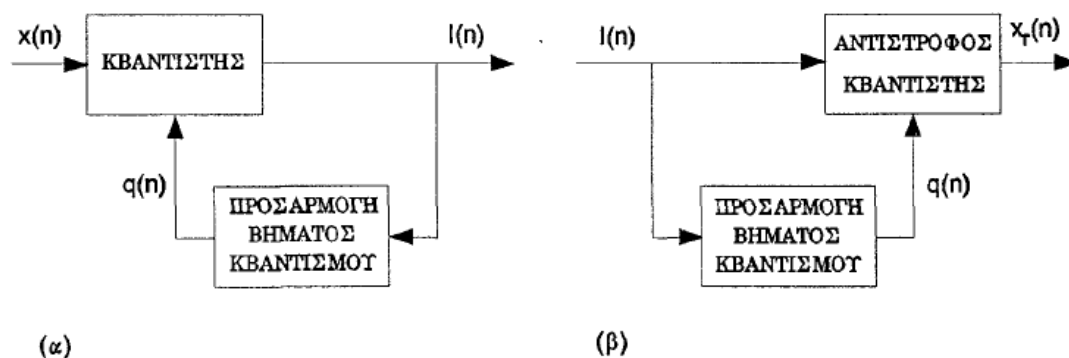
Ο συνδυασμός της δειγματοληψίας και της κβαντοποίησης ως συμπληρωματικές διαδικασίες αποκαλείται ψηφιοποίηση, ενώ για την εκμετάλλευση των πλεονεκτημάτων τους απαραίτητη προϋπόθεση η χρησιμοποίηση μιας διαδικασίας κωδικοποίησης (encoding process) για την μετατροπή του διακριτού συνόλου των τιμών των δειγμάτων σε μια πιο κατάλληλη μορφή. Κατά το στάδιο αυτό, η ένταση του κάθε δείγματος αντιστοιχεί σε ένα σταθερό δυαδικό αριθμό 0,1 και διατίθεται για την αποθήκευση κάθε δείγματος και καλείται εύρος δείγματος.

Αποτελούν την κατηγορία τεχνικών συμπίεσης [Παρ95], με σημαντική αξία αλλά χαμηλές επιδόσεις, όχι τόσο ποιότητας, όσο παροχής δεδομένων. Βρίσκονται σε εμπορική χρήση για αρκετά χρόνια, ενώ πολλές από αυτές έχουν καθιερωθεί σαν πρότυπα από τους διεθνείς οργανισμούς τυποποίησης και ISO. Παρέχουν συντελεστές συμπίεσης από 1,5 μέχρι 2,5 bits/δείγμα και πολύ μικρή πολυπλοκότητα κωδικοποιητή και αποκωδικοποιητή. Η παλαιότερη είναι η PCM, που χρησιμοποιείται στα CD και DAT (Digital Audio Tape), με 16 bits μήκος λέξης, για καταναλωτικές εφαρμογές ήχου και 18, 20 ή 24 bits μήκος λέξης, για επαγγελματικές εφαρμογές ήχου και είναι το σημείο αναφοράς της ακουστικής ποιότητας και της παροχής για τις νεώτερες τεχνικές.

### 3.2.1 ΠΡΟΣΑΡΜΟΖΩΜΕΝΗ ΠΑΛΜΟΚΩΔΙΚΗ ΔΙΑΜΟΡΦΩΣΗ PCM

Η προσαρμοζόμενη παλμοκωδική διαμόρφωση (Adaptive Pulse Code Modulation - APCM) [Παρ95] είναι μία μέθοδος που εφαρμόζεται τόσο σε ομοιόμορφους όσο και σε ανομοιόμορφους κβαντιστές. Το βασικό της χαρακτηριστικό είναι ότι μεταβάλλει το βήμα του κβαντισμού ανάλογα με τη μεταβατική συμπεριφορά του σήματος. Στις χρονικές στιγμές όπου το σήμα εισόδου μεταβάλλεται γρήγορα, δηλαδή όταν η κυματομορφή του παρουσιάζει απότομη κλίση, το βήμα κβαντισμού αυξάνεται, προκειμένου να ελαχιστοποιηθεί το σφάλμα κβαντισμού. Αντίθετα, σε στιγμές όπου το σήμα παρουσιάζει μικρές διακυμάνσεις πλάτους, το βήμα κβαντισμού μειώνεται, προκειμένου να διατηρηθεί υψηλή ευκρίνεια και χαμηλός θόρυβος κβαντισμού. Στην πράξη αυτό επιτυγχάνεται με τη χρήση ενός ενισχυτή μεταβλητής απολαβής.





Σχήμα 3.6: Λειτουργικό διάγραμμα Προσαρμοζόμενης Παλμοκωδικής Διαμόρφωσης:  
 (α) Κωδικοποιητής και (β) Αποκωδικοποιητής.

Η βασική ιδέα του συστήματος αυτού βασίζεται σε μία βαθμίδα προσαρμοζόμενης ανάδρασης, όπως δείχνεται στο σχήμα 3.6. Το σήμα εισόδου  $x(n)$  στον κωδικοποιητή κβαντίζεται και κωδικοποιείται παράγοντας το σήμα εξόδου  $I(n)$ , το οποίο επεξεργάζεται από μία λογική μονάδα προσαρμογής του βήματος κβαντισμού και προκύπτει ένα σήμα οδήγησης  $q(n)$  του κβαντιστή, με το οποίο προσαρμόζεται το βήμα κβαντισμού. Αντίστοιχα, στον αποκωδικοποιητή το λαμβανόμενο σήμα  $I(n)$  επεξεργάζεται από έναν αντίστροφο κβαντιστή, δίνοντας μία προσέγγιση  $x_r(n)$  του αρχικού σήματος εισόδου στον κωδικοποιητή. Το βήμα κβαντισμού του αντίστροφου κβαντιστή προκύπτει με την επεξεργασία του σήματος  $I(n)$  από τη μονάδα προσαρμογής του βήματος κβαντισμού.

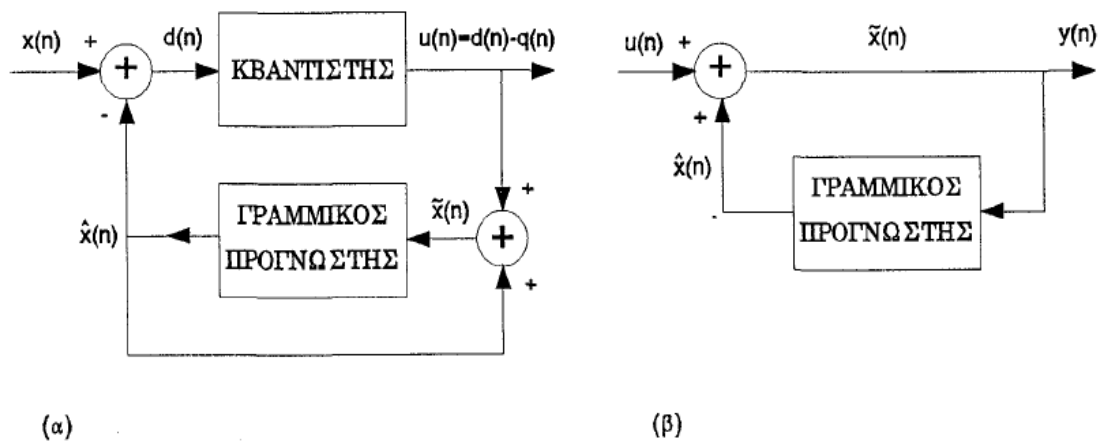
### 3.2.2 MU-LAW PCM ΚΑΙ A-LAW PCM

Η διαφορά της λογαριθμικής κωδικοποίησης ήχου σε σχέση με την γραμμική βρίσκεται στο γεγονός ότι κατά την λογαριθμική αντιστοίχιση, αντιστοιχούνται ολοένα και περισσότερες στάθμες σε χαμηλές συχνότητες και λιγότερες στις υψηλές, ενώ ταυτόχρονα πραγματοποιείται τόσο καλύτερη αναπαράσταση όσο και καλύτερη συμπίεση του σήματος με αυτό της παλμοκωδικής, η οποία δεν πραγματοποιεί καμία συμπίεση ούτως ώστε να προκύπτει ένα αρχείο ήχου κωδικοποιημένο χωρίς καμία απώλεια. Κατά συνέπεια, στην κατηγορία αυτή της λογαριθμικής κωδικοποίησης ήχου εντάσσονται οι Mu-Law PCM και A-Law PCM [Χρυ06], δύο μέθοδοι που έχουν τυποποιηθεί από το ITU-T (International

Telecommunication Union-Telecommunication Standardization Sector) στο πρότυπο G711, Pulse Code Modulation of noise Frequencies το οποίο και περιγράφει λεπτομερές τα δύο είδη κωδικοποίησης. Συγκεκριμένα, η Mu-Law υλοποιείται σε ISDN (Integrated Services Digital Network) δίκτυα της Ιαπωνίας και Β. Αμερικής ενώ η A-Law σε δίκτυα των υπόλοιπων χωρών.

### 3.2.3 ΔΙΑΦΟΡΙΚΗ ΠΑΛΜΟΚΩΔΙΚΗ ΔΙΑΜΟΡΦΩΣΗ DPCM

Η μέθοδος αυτή εκμεταλλεύεται την υψηλή συσχέτιση που υπάρχει συνήθως ανάμεσα σε διαδοχικά δείγματα του ακουστικού σήματος εισόδου, για να επιτύχει μείωση της απαιτούμενης παροχής δεδομένων. Το όνομα της οφείλεται στο γεγονός ότι κβαντίζει και κωδικοποιεί τη διαφορά ανάμεσα σε διαδοχικά δείγματα και όχι τα ίδια τα δείγματα. Επειδή η απόλυτη τιμή της διαφοράς των διαδοχικών δειγμάτων κυμαίνεται σε μικρότερη δυναμική περιοχή από ότι τα πραγματικά δείγματα γι' αυτό το λόγο απαιτούνται λιγότερα bits για την κωδικοποίηση της, προκειμένου να επιτευχθεί ο ίδιος λόγος σήματος προς θόρυβο κβαντισμού [Παρ95]. Το βήμα κβαντισμού μεταξύ διαδοχικών δειγμάτων διαφοράς διατηρείται σταθερό.

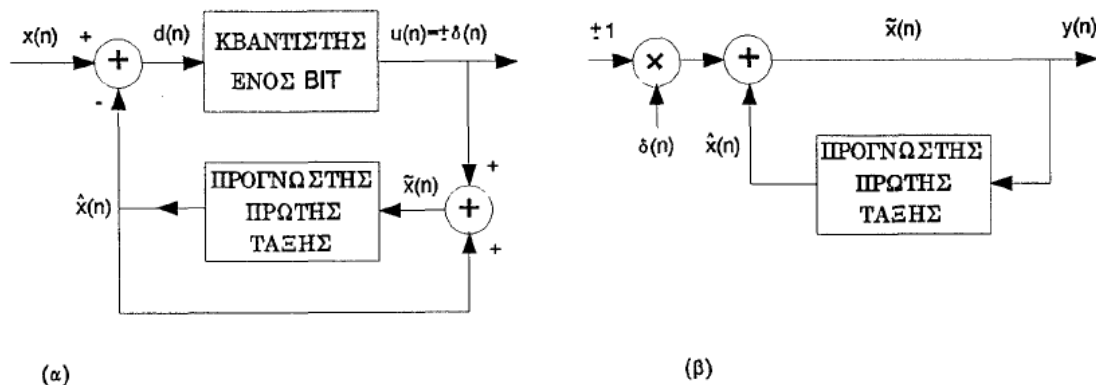


Σχήμα 3.7: Λειτουργικό διάγραμμα Διαφορικής Παλμοκωδικής Διαμόρφωσης: (α) Κωδικοποιητής και (β) Αποκωδικοποιητής.

Όπως φαίνεται στο σχήμα 3.7 (α), το σύστημα κωδικοποίησης DPCM στην ουσία είναι ένα σύστημα κβαντισμού κλειστού βρόχου ή ανάδρασης. Η απλούστερη μορφή της DPCM προκύπτει για μοναδιαία καθυστέρηση και μια παραλλαγή της DPCM αποτελεί η Διαμόρφωση Δέλτα.

### 3.2.4 ΔΙΑΜΟΡΦΩΣΗ ΔΕΛΤΑ (Delta Modulation - DM)

Όταν η διαφορά ανάμεσα σε διαδοχικά δείγματα κβαντίζεται σε δύο μόνο στάθμες και κωδικοποιείται με 1 bit, τότε αναφερόμαστε στη Διαμόρφωση Δέλτα (Delta Modulation - DM) [Παρ95]. Τα συστήματα της Διαμόρφωσης Δέλτα πλεονεκτούν ως προς τα M-ιαδικά PCM και M-ιαδικά DPCM συστήματα στο γεγονός ότι τα κυκλώματα που χρειάζονται για τη διαμόρφωση, τη μετάδοση και την αποδιαμόρφωση είναι πολύ πιο απλά.



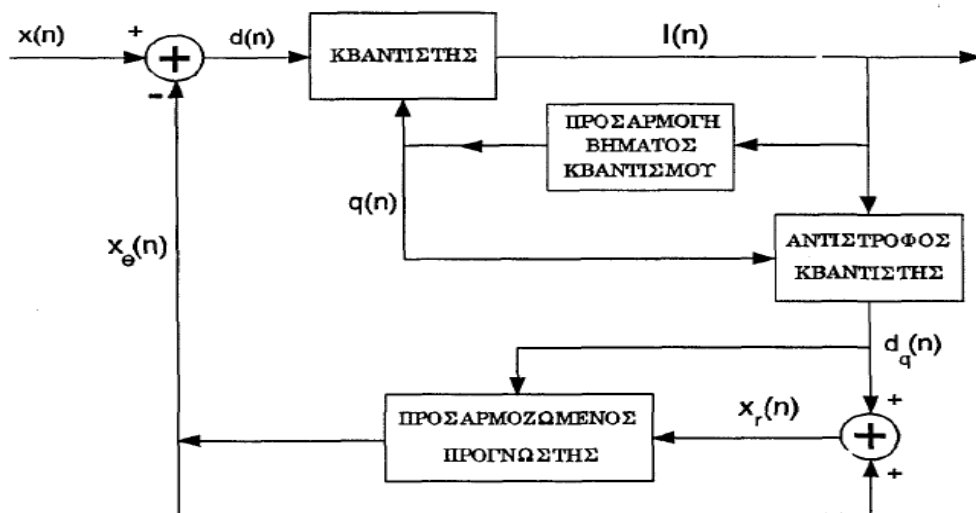
Σχήμα 3.8: Λειτουργικό διάγραμμα Δέλτα Διαμόρφωσης: (α) Κωδικοποιητής και (β) Αποκωδικοποιητής.

Τα κύριο πρόβλημα της Διαμόρφωσης Δέλτα παρουσιάζεται στα σημεία μέγιστης κλίσης του σήματος, δηλαδή στα έντονα μεταβατικά περάσματα, όπου τη λύση προσφέρει η Προσαρμοζόμενη Διαμόρφωση Δέλτα (Adaptive Delta Modulation - ADM). Σύμφωνα με αυτή και χρησιμοποιώντας τεχνικές γραμμικής πρόγνωσης, ανιχνεύονται οι καταστάσεις μεταβολής του σήματος και αυξάνεται το βήμα κβαντισμού, έτσι ώστε να μειώνεται η υπερφόρτωση. Αντίθετα, όταν το σήμα αποκτά μικρά πλάτη μειώνεται το βήμα κβαντισμού ώστε να ελαχιστοποιείται ο θόρυβος κβαντισμού.

Συστήματα βασισμένα σε απλή και σε προσαρμοζόμενη Διαμόρφωση Δέλτα βρίσκονται σε εμπορική χρήση εδώ και πολλά χρόνια και χρησιμοποιούνται για συμπίεση σημάτων εικόνας, ομιλίας καθώς και σε ψηφιακά συστήματα ήχου.

### 3.2.5 ΔΙΑΦΟΡΙΚΗ ΠΡΟΣΑΡΜΟΖΩΜΕΝΗ ΠΑΛΜΟΚΩΔΙΚΗ ΔΙΑΜΟΡΦΩΣΗ ADPCM

Ο αλγόριθμος ADPCM [Παρ95] αποτελεί συγκερασμό των APCM και DPCM τεχνικών και συνδυάζει επιτυχώς τα πλεονεκτήματά τους. Έχει καταχωρηθεί σαν πρότυπο με αριθμό G.721 από την CCITT και εμπορικά χρησιμοποιείται για τα κανάλια ήχου του Αυστραλιανού δορυφόρου και σε πολλές άλλες ηχητικές εφαρμογές, (π.χ. αποθήκευση αρχείων ήχων σε πολυμέσα).



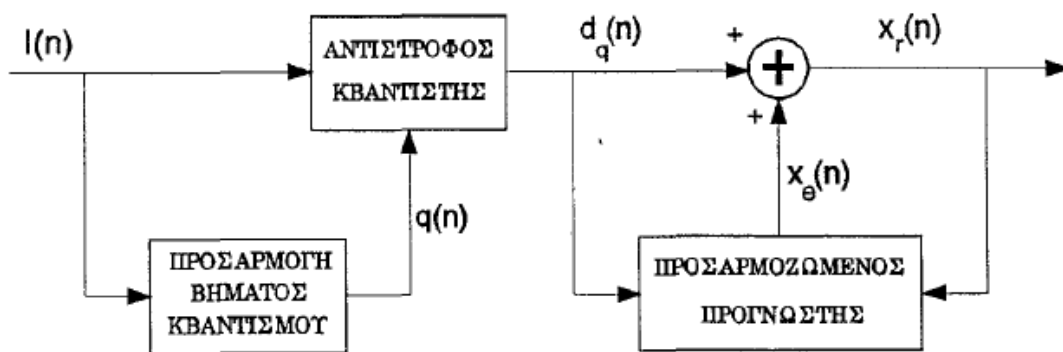
Σχήμα 3.9: Διάγραμμα λειτουργίας ADPCM κωδικοποιητή.

Το βασικό χαρακτηριστικό της τεχνικής ADPCM είναι ότι κωδικοποιεί τη διαφορά ανάμεσα στο παρόν δείγμα  $x(n)$  και σε μία εκτίμηση  $x_e(n)$ , που έχει προκύψει από προηγούμενα δείγματα. Επιπλέον το βήμα κβαντισμού καθορίζεται με βάση τα χαρακτηριστικά της κυματομορφής του σήματος και ειδικά από την ταχύτητα μεταβολής του. Σε εμπορικές εφαρμογές η τεχνική ADPCM έχει απόδοση κατά 50% καλύτερη από την

τεχνική NICAM. Στο σχήμα 3.9 δείχνεται το διάγραμμα λειτουργίας του κωδικοποιητή ADPCM.

Οι λειτουργίες που επιτελούνται από τον κωδικοποιητή ADPCM, συνοπτικά είναι οι εξής. Ο κβαντιστής κβαντίζει τη διαφορά  $d(n)$ , μεταξύ του παρόντος δείγματος εισόδου  $c(n)$  και μίας εκτίμησης  $x_e(n)$ , που γίνεται για το δείγμα εισόδου και βασίζεται σε προηγούμενα δείγματα του σήματος. Το σήμα διαφοράς  $d(n)$  κωδικοποιείται σε μία λέξη  $I(n)$  των 4 bits και μεταδίδεται στην έξοδο. Το βήμα κβαντισμού ελέγχεται από ένα σήμα οδήγησης  $q(k)$ , το οποίο εκφράζει την ταχύτητα μεταβολών του σήματος (αργά ή γρήγορα μεταβατικά). Μεγάλες τιμές του σήματος οδήγησης  $q(n)$ , οδηγούν σε αύξηση του βήματος κβαντισμού, ώστε να διατηρείται σταθερό το μήκος λέξης. Ο Αντίστροφος Κβαντιστής και ο Προσαρμοζόμενος Προγνώστης (Adaptive Predictor), σχηματίζουν τη βαθμίδα πρόγνωσης του παρόντος δείγματος εισόδου, χρησιμοποιώντας τιμές περασμένων δειγμάτων. Ο Προσαρμοζόμενος Προγνώστης, είναι ένα ψηφιακό φίλτρο δύο πόλων και έξη μηδενικών και η λειτουργία του βασίζεται σε τεχνικές γραμμικής πρόγνωσης. Η εκτίμηση  $x_e(n)$  είναι ένα άθροισμα όρων των φίλτρων μόνο πόλων (all- pole) και μόνο μηδενικών (all- zero).

Ο αποκωδικοποιητής ADPCM είναι πολύ απλούστερος από τον κωδικοποιητή και δείχνεται στο σχήμα 3.10.



Σχήμα 3.10: Διάγραμμα λειτουργίας αποκωδικοποιητή ADPCM.

Όπως δείχνεται στο σχήμα 3.10, από τη λαμβανόμενη κωδικοποιημένη παλμοσειρά  $I(n)$ , δημιουργείται το σήμα οδήγησης  $q(n)$  (εξαρτώμενο από την ταχύτητα μεταβολών του σήματος εισόδου), που καθορίζει το βήμα κβαντισμού του αντίστροφου κβαντιστή, έτσι ώστε να αναδημιουργηθεί μία προσέγγιση της διαφοράς  $d_q(n)$ , ανάμεσα στο παρόν δείγμα και στο προηγούμενο. Το σήμα  $d_q(n)$ , οδηγεί τον προσαρμοζόμενο προγνώστη, που παράγει την

εκτίμηση  $x_e(n)$  του προηγούμενου δείγματος. Τέλος, αθροίζεται το σήμα διαφοράς  $d_q(n)$  με το σήμα εκτίμησης  $x_e(n)$  και προκύπτει σαν έξοδος μία προσέγγιση  $x_r(n)$  του σήματος εισόδου του κωδικοποιητή.

Στον Πίνακα 3.2, γίνεται μία σύγκριση ανάμεσα σε μερικές από τις προαναφερθείσες τεχνικές, που έχουν βρει εμπορική χρήση σε ηχητικά συστήματα (για μονοφωνική αναπαραγωγή).

	PCM (CD format)	PCM	NICAM	Προσαρμοζόμενη Δέλτα
Συχνότητα Δειγματοληψίας (kHz)	44,1	32	32	32
Μήκος Λέξης (bits/Δείγμα)	16	14	10,1	10,3
Παροχή Δεδομένων (bits/s)	706	448	323	330

**Πίνακας 2.2 Σύγκριση βασικών παραμέτρων τεχνικών κωδικοποίηση πεδίου χρόνου.**

### 3.2.6 ΠΡΟΣΑΡΜΟΖΩΜΕΝΗ ΠΑΛΜΟΚΩΔΙΚΗ ΚΩΔΙΚΟΠΟΙΗΣΗ LPC

Μια από τις νέες τεχνολογίες κωδικοποίησης ήχου που έχουν συνταχθεί για την ανθρώπινη ομιλία και πραγματοποιούν σημαντικούς βαθμούς συμπίεσης είναι η γραμμική προβλεπόμενη κωδικοποίηση LPC [Χρυ06], με σκοπό ο κωδικοποιητής αυτός να συγκρίνει τα σήματα φωνής που παραλαμβάνει με βάση ένα αναλυτικό μοντέλο φωνής που έχει αποθηκευμένο. Τα χαρακτηριστικά που ταιριάζουν καλύτερα στο αναλυτικό μοντέλο μεταδίδονται, ενώ ο αποκωδικοποιητής χρησιμοποιεί τα χαρακτηριστικά αυτά για να ανασυνθέσει τα φωνητικά σήματα. Παρόλα αυτά, μειονέκτημα της μεθόδου αυτής εντοπίζονται στην αδυναμία της να επεξεργαστεί άλλο σήμα εκτός από την ομιλία, η οποία δημιουργήθηκε για να εξυπηρετήσει την μετάδοση ομιλίας στην κινητή τηλεφωνία.

## ΚΕΦΑΛΑΙΟ 4

### ΜΕΛΕΤΗ ΤΕΧΝΙΚΩΝ ΣΥΜΠΙΕΣΗΣ ΨΗΦΙΑΚΟΥ ΗΧΟΥ

#### 4.1 ΕΙΣΑΓΩΓΗ

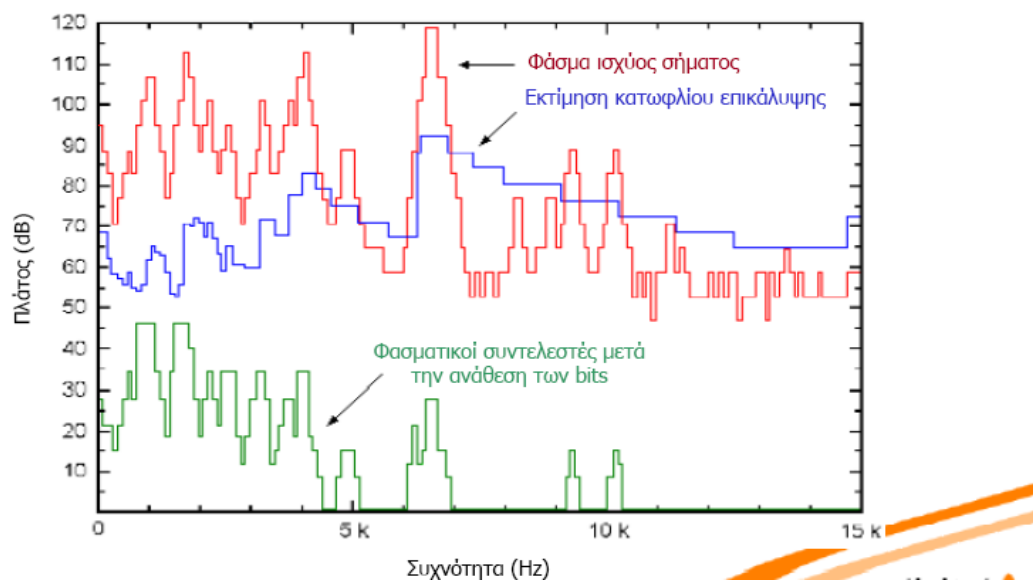
Η εξασφάλιση υψηλών συντελεστών συμπίεσης και ηχητικής ποιότητας σύμφωνα με την έρευνα αναζήτησης τεχνικών αποδοτικής συμπίεσης ηχητικού σήματος στηρίχτηκε στην αποτελεσματική προσομοίωση χαρακτηριστικών της ανθρώπινης ακοής. Το πρότυπο που δημιουργήθηκε ονομάστηκε MPEG-1 την εποχή αυτή βρισκόμαστε τόσο στο MPEG-2 όσο και στο MPEG-4 [Παρ95]. Η διαθεσιμότητα ενός πλήθους από εφαρμογή ευνοεί την εμφάνιση περισσότερων εφαρμογών σε γρήγορο χρονικό διάστημα. Ευνοημένες θεωρούνται οι εφαρμογές που στηρίζονται στην εμφάνιση και διανομή μουσικής μέσω δικτύων και ειδικά του internet, οι εφαρμογές πολυμέσων, οι ραδιοτηλεοπτικές εκπομπές και κατ' επέκταση εφαρμογές εκπαίδευσης, ψυχαγωγίας και επικοινωνίας.

#### 4.2 ΣΥΜΠΙΕΣΗ ΨΗΦΙΑΚΟΥ ΗΧΟΥ

Η αντικατάσταση της αναλογικής τεχνολογίας από την ολοένα αυξανόμενη χρήση της ψηφιακής τεχνολογίας χρησιμοποιείται σε πολλούς τομείς, όπως στις τηλεπικοινωνίες, την ηχογράφηση και αναπαραγωγή της μουσικής. Κατά την διαδικασία της ψηφιοποίησης ηχητικών δεδομένων παράγονται αρχεία μεγάλων μεγεθών, τα οποία είναι δύσκολο να μεταφερθούν ή να υποστούν επεξεργασία, αφού προϋποθέτουν πολύ χρόνο, μεγάλο αποθηκευτικό χώρο.

Οι τεχνολογίες συμπίεσης του ηχητικού σήματος είναι αυτές που θα δώσουν την λύση στο τεχνικό πρόβλημα του χώρου και της ανεπάρκειας στην ταχύτητα μεταφοράς. Οι τεχνολογίες αυτές διακρίνονται σε δύο κατηγορίες, τις μη απωλεστικές (lossless ) και τις απωλεστικές (lossy ) [Χρυ06]. Σκοπός τους είναι να μειώσουν τον όγκο και την ροή των δεδομένων που απαιτούνται για τη κωδικοποίηση των ηχητικών σημάτων. Η διαφορά τους είναι ότι στις μη απωλεστικές η διαδικασία συμπίεσης δεν αλλοιώνει καθόλου την πληροφορία, μετά την αποσυμπίεση η πληροφορία επανέρχεται ακριβώς στην αρχική της μορφή. Γι' αυτό είναι κατάλληλες σε περιπτώσεις που δεν υπάρχει περιθώριο απωλειών.

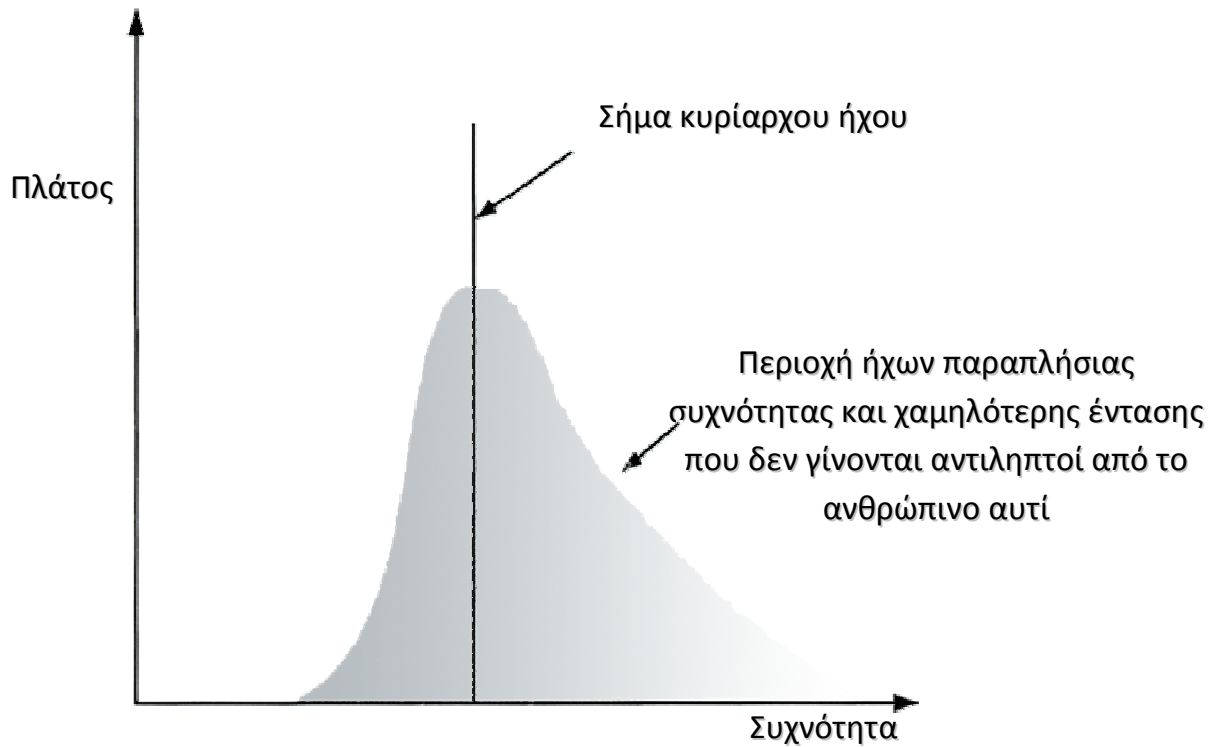
Από την άλλη πλευρά οι απωλεστικές τεχνικές προκαλούν αλλοίωση στα δεδομένα απορρίπτοντας έτσι τα μέρη εκείνα που δε είναι χρήσιμα με βάση κάποιο συγκεκριμένο κριτήριο. Με ακρίβεια δεν υπάρχει μεταβολή αλλά υπερισχύει η έννοια της μείωσης της ποιότητας καθώς οι απωλεστικές είναι περισσότερο αποτελεσματικές επιτυγχάνοντας συμπίεση που φτάνει 24:1 με καλή ποιότητα , ενώ οι μη απωλεστικές σπάνια ξεπερνούν το 4:1. Παρόλα αυτά στις σύγχρονες μεθόδους συμπίεσης εφαρμόζονται διάφορα ψυχοακουστικά μοντέλα αντίληψης βάση των οποίων απορρίπτονται οι ήχοι που δεν μπορούν να γίνουν κατανοητοί από το ανθρώπινο αυτί.



**Σχήμα 4.1: Ψυχοακουστικό Μοντέλο Ακοής.**

Οι σημαντικότεροι μέθοδοι συμπίεσης ψηφιακών ηχητικών αρχείων στηρίζονται στο φαινόμενο της ηχητικής σκίασης (auditory masking ) που αποσκοπεί στην ακουστότητα των ήχων. Πιο συγκεκριμένα συχνότητες χαμηλής έντασης κοντά σε μία συχνότητα υψηλής έντασης δεν γίνονται αντιληπτές από το ανθρώπινο αυτί. Οι συχνότητες που απορρίπτονται από το ανθρώπινο αυτί είναι αυτές που δεν μπορεί να αντιληφθεί σε ένα καθορισμένο διάστημα συχνοτήτων κατά την συμπίεση ψηφιακών ηχητικών δεδομένων. Παρά το γεγονός ότι δεν λαμβάνονται κάποιες ηχητικές πληροφορίες η ποιότητα του ήχου παραμένει σε υψηλά επίπεδα.





Σχήμα 4.2: Φαινόμενο Ηχητικής Σκίασης.

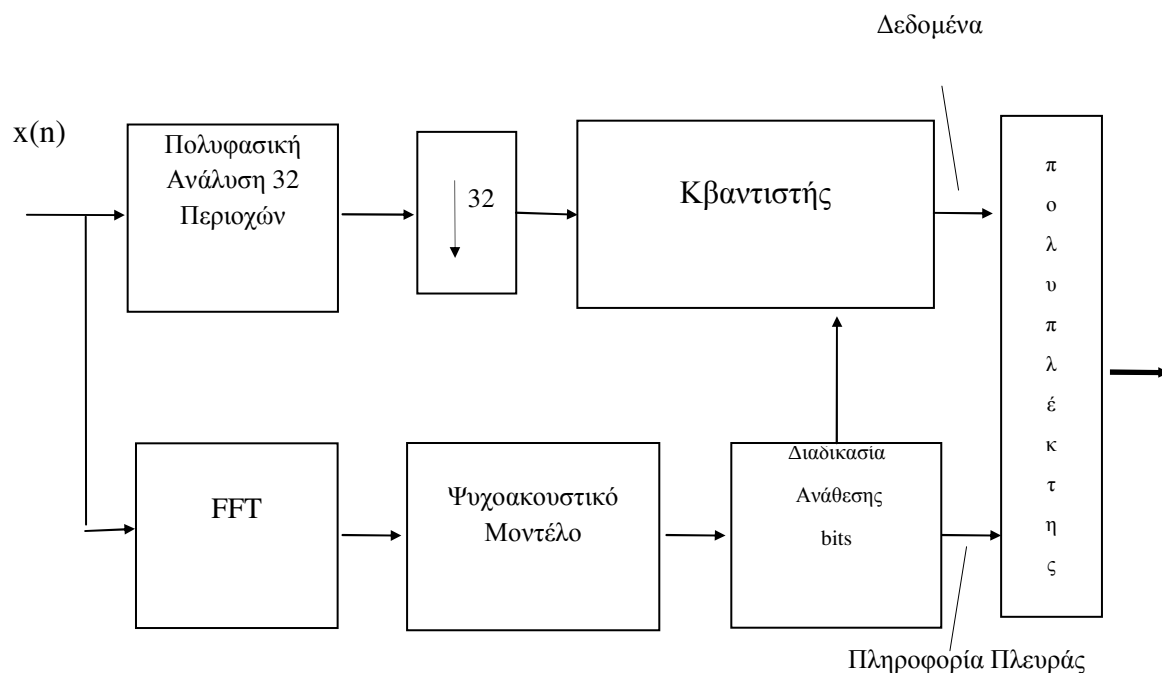
Το φαινόμενο της ηχητική σκίασης, είναι άμεσα συνδεδεμένο τόσο με το πρότυπο AC-3 Dolby Digital (H.P.A) όσο και με το MPEG των οποίων οι διαφορές τους στηρίζονται στο τρόπο υλοποίησης τους. Εντούτοις και τα δύο πρότυπα συμπίεσης ηχητικών δεδομένων, δουλεύουν χωρίζοντας το φάσμα των ακουστικών συχνοτήτων σε υπομπάντες χρησιμοποιώντας περίπλοκους αλγόριθμους και ψυχοακουστικά μέσα για να απορρίψουν τις μη ακουστικές από το ανθρώπινο αυτί συχνοτήτες. Ο ανταγωνισμός μεταξύ των προτύπων AC-3 και MPEG, φαίνεται να ευνοεί το μοντέλο AC-3 το οποίο στην πορεία θα αναδειχθεί ως το διεθνές πρότυπο για την συμπίεσης ψηφιακών ηχητικών δεδομένων.

## 4.3 ΠΡΟΤΥΠΙΑ MPEG

Η MPEG, ως μια συνεργασία Πανεπιστημίων, ερευνητών ινστιτούτων και εταιριών λειτουργεί στα πλαίσια του Διεθνούς Οργανισμού Τυποποίησης γνωστή σαν ISO/IEC [Χρυ06] με στόχο την ανάπτυξη διεθνών πρότυπων για την συμπίεση και αποσυμπίεση, την επεξεργασία και την κωδικοποιημένη αντιπροσώπευση της κίνησης των εικόνων, του ήχου και τους συνδυασμούς τους. Το όνομα MPEG, έχει επικρατήσει όμως, να αναφέρεται και στη οικογένεια των τυποποιήσεων (standards) που δημιουργήθηκαν από την ομάδα MPEG και χρησιμοποιούνται για την μετάδοση οπτικών και ηχητικών δεδομένων σε ψηφιακή συμπίεσμένη μορφή. Στην οικογένεια MPEG, εντάσσονται τα standards MPEG-1 που αφορά την συμπίεση ήχου και εικόνας, το MPEG-2 για την εφαρμογή στην ψηφιακή τηλεόραση και το MPEG-4 ως standard για εφαρμογές επικοινωνίας πολυμέσων. Επίσης, υπάρχει στα σχέδια τους και το MPEG-7, με στόχο την αναπαράσταση περιεχομένου (content representation) για την αναζήτηση πληροφοριών σε εφαρμογές.

### 4.3.1 Το Πρότυπο MPEG - 1

Το πρότυπο ISO/IEC 11172-3 παρουσιάστηκε το 1993 και ήταν ο συγκερασμός των ερευνητικών τεχνικών υποκειμενικής συμπίεσης της μουσικής που αναπτύχθηκαν κατά τα προηγούμενα πέντε χρόνια και ειδικά του ASPEC και του MUSICAM [Παρ95]. Το πρότυπο αυτό περιλαμβάνει ευέλικτες τεχνικές κωδικοποίησης, που περιλαμβάνουν επαρκείς μεθόδους χρονοσυχνοτικής ανάλυσης, κωδικοποίηση εντροπίας, δυναμική ανάθεση των bits, ανομοιόμορφο κβαντισμό και αποδοτικό ψυχοακουστικό μοντέλο. Ο κωδικοποιητής δέχεται σαν είσοδο σήμα PCM με συχνότητα δειγματοληψίας 32, 44.1 και 48 kHz και μπορεί να λειτουργήσει σε μονοφωνικό, στερεοφωνικό, διπλό μονοφωνικό και σε συνδυασμένο στερεοφωνικό τρόπο. Για μονοφωνικό σήμα η παροχή δεδομένων που επιτυγχάνει κυμαίνεται από 32 μέχρι 192 kbits/s και για στερεοφωνικό από 64 μέχρι 384 kbits/s. Η αρχιτεκτονική του αναλύεται σε τρία επίπεδα (Layers) με αυξανόμενη πολυπλοκότητα, καθυστέρηση και ηχητική ποιότητα από το I προς το III.

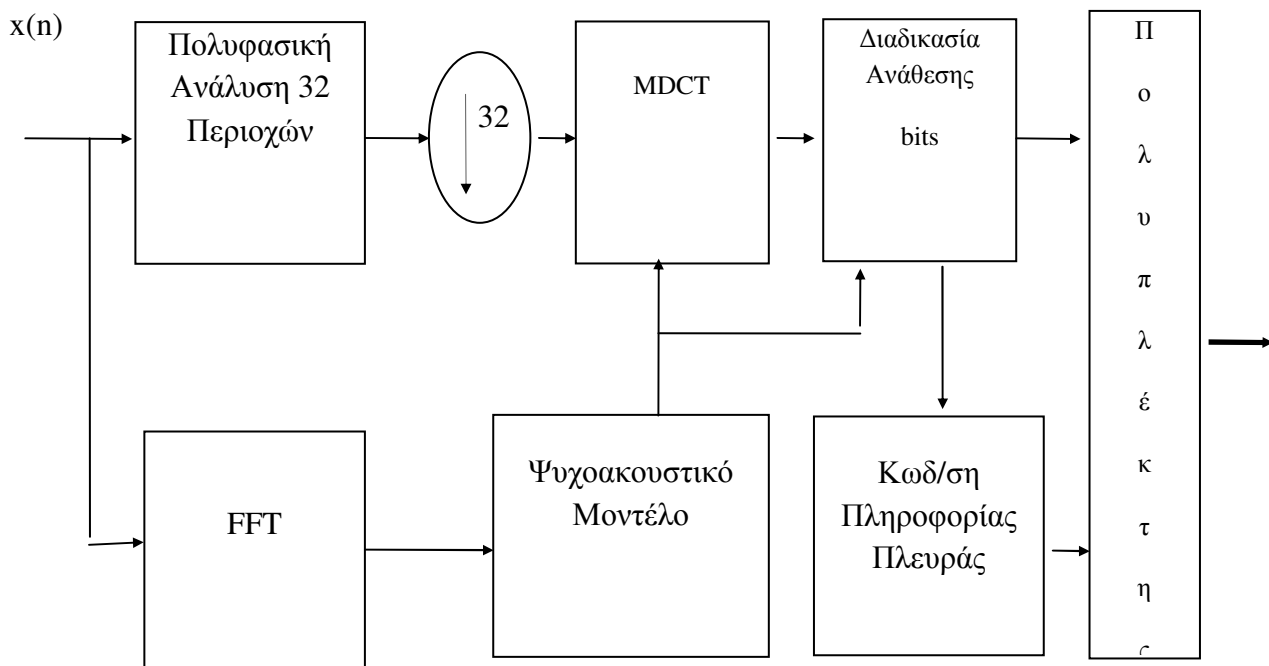


Σχήμα 4.3: Κωδικοποιητής ISO/MPEG Επιπέδου I και II.

Στα επίπεδα I και II και με τη χρήση πολυφασικής τράπεζας φίλτρων, το σήμα αναλύεται σε 32 κρίσιμα υποδειγματισμένες περιοχές. Η φασματική ευκρίνεια του συστήματος ανάλυσης, για συχνότητα δειγματοληψίας 48 kHz, είναι 750 Hz. Για τον υπολογισμό του κατωφλίου ακουστότητας των παρα-μορφώσεων χρησιμοποιείται ο FFT (512 σημεία για το επίπεδο I και 1024 για το επίπεδο II). Στη συνέχεια, τα δείγματα από το κάθε φίλτρο κανονικοποιούνται ως προς τη μέγιστη τιμή, η οποία και μεταδίδεται σαν πληροφορία πλευράς. Κατόπιν, μία επαναληπτική διαδικασία ανάθεσης των bits που χρησιμοποιεί το κατώφλι ακουστότητας, καθορίζει το βήμα κβαντισμού, που είναι διαφορετικό για κάθε φασματική περιοχή, έτσι ώστε να επιτυγχάνονται ταυτόχρονα οι στόχοι της μείωσης της παροχής δεδομένων και της διατήρησης της ηχητικής ποιότητας. Σε κάθε φασματική περιοχή οι συντελεστές κλιμάκωσης κωδικοποιούνται με 6 bits και τα δείγματα με 4 bits.

Το επίπεδο II επιτυγχάνει καλύτερη ακουστική ποιότητα και μειωμένη παροχή δεδομένων σε σχέση με το επίπεδο I, για διάφορους λόγους. Ο πρώτος λόγος είναι ότι το ψυχοακουστικό μοντέλο που υλοποιεί έχει καλύτερη συχνοτική ευκρίνεια λόγω του μεγαλύτερου μήκους FFT. Ο δεύτερος λόγος είναι ότι υιοθετεί περισσότερο ευφυή αλγόριθμο

υπολογισμού και μετάδοσης των συντελεστών κλιμάκωσης, ενώ ο τρίτος είναι η βελτιωμένη ευκρίνεια του κβαντιστή.



Σχήμα 4.4: Κωδικοποιητής ISO/MPEG Επιπέδου III.

Το επίπεδο III με την πρόσθεση διαφόρων σημαντικών μηχανισμών στην αρχιτεκτονική του προερχόμενων από τους αλγορίθμους ASPEC και MUSICAM (Σχήμα 4.4) επιτυγχάνει την υψηλότερη απόδοση. Η βαθμίδα χρονοσυχνοτικής ανάλυσης υλοποιείται από μία υβριδική δομή υποζωνικών φίλτρων και τροποποιημένου μετασχηματισμού συνημιτόνου (MDCT), που προσδίδει στο σύστημα καλύτερη συχνοτική ανάλυση και εγγύτερη προσομοίωση των κρίσιμων περιοχών του ανθρώπινου ακουστικού συστήματος.

Το μήκος του παραθύρου ανάλυσης μεταβάλλεται ανάλογα με τα στατιστικά χαρακτηριστικά του σήματος εισόδου, επιτρέποντας έτσι τον αποτελεσματικό έλεγχο των φαινομένων προ-ηχούς (pre-echo effects). Συγκεκριμένα, σε στιγμές όπου το σήμα παρουσιάζει έντονα μεταβατικά, το μήκος του παραθύρου μειώνεται, έτσι ώστε να ελαχιστοποιηθεί το φαινόμενο της προηχούς, ενώ σε στάσιμες περιόδους του σήματος το παράθυρο μεγαλώνει, ώστε να μειωθεί η ποσότητα της πληροφορίας πλευράς. Καλά σχεδιασμένες τεχνικές ανάθεσης των bits, μη ομοιόμορφος κβαντισμός, διαδικασίες ανάλυσης – σύνθεσης και κωδικοποίηση εντροπίας (Huffman Coding), χρησιμοποιούνται για

τον κβαντισμό των φασματικών συντελεστών και αποδίδουν υψηλότερο κέρδος ακουστικής συμπίεσης χωρίς αισθητό υποβιβασμό της ηχητικής ποιότητας. Ο επαναληπτικός κβαντιστής που χρησιμοποιείται προέρχεται από την τεχνική OCF.

Οι εφαρμογές που εξυπηρετούνται από το πρότυπο MPEG-1 προφανώς είναι διαφορετικές για κάθε επίπεδο. Συγκεκριμένα, για εφαρμογές όπου η χαμηλή παροχή δεδομένων δεν είναι ιδιαίτερα απαιτητή, όπως σε καταναλωτικές εφαρμογές, δηλαδή οικιακές ηχογραφήσεις σε ψηφιακές κασέτες (Digital Compact Cassette – DCC) και σε μαγνητοοπτικούς δίσκους χρησιμοποιείται το επίπεδο I, με τυπική παροχή 192 kbits/s. Το επίπεδο II χρησιμοποιείται τόσο σε οικιακές όσο και σε επαγγελματικές εφαρμογές, δηλαδή σε χρήση studio ή ακόμα και στους δίσκους DVD. Η τυπική παροχή του επιπέδου αυτού κυμαίνεται στην περιοχή των 128 kbits/s.

Το επίπεδο III παρέχει την χαμηλότερη παροχή δεδομένων και ταυτόχρονα υψηλή ακουστική ποιότητα και ως εκ τούτου χρησιμοποιείται σε τηλεπικοινωνιακές εφαρμογές και ειδικά για τη διανομή ηχητικού υλικού μέσα από δίκτυα υπολογιστών (Internet, στενής ζώνης ISDN κλπ) ή σε ραδιοφωνικές εφαρμογές (δορυφορικά κανάλια, ψηφιακό ραδιόφωνο DAB). Η τυπική παροχή του κυμαίνεται στην περιοχή των 64-96 kbits/s.

### 4.3.2 Το Πρότυπο MPEG – 2

Η Ομάδα Ειδικών Κινούμενης Εικόνας ή MPEG - Moving Picture Experts Group είναι μια ομάδα εργασίας ISO / IEC, η οποία είναι υπεύθυνη για την ανάπτυξη των τηλεοπτικών και ακουστικών προτύπων κωδικοποίησης. Ένα από τα πρότυπα της οικογένειας MPEG το οποίο αξιοποιείται για τη δημιουργία τηλεοπτικών σημάτων είναι το MPEG-2 [Σαρ12]. Το πρότυπο αυτό υποστηρίζει εφαρμογές με μεγάλες απαιτήσεις στην ποιότητα του video και προσφέρει μεταβλητούς ρυθμούς μετάδοσης από τη στιγμή που ο βαθμός συμπίεσης μεταβάλλεται αντιστρόφως ανάλογα με την πολυπλοκότητα των πλαισίων που κωδικοποιούνται. Έχει καθιερωθεί σε παγκόσμιο επίπεδο ως το πρότυπο για συμπίεση της ψηφιακής τηλεόρασης αφού παρέχει υψηλό βαθμό συμπίεσης διατηρώντας την εικόνα σε υψηλό επίπεδο ποιότητας. Το πρότυπο MPEG-2 σε συνδυασμό με το πρότυπο Multi Protocol Encapsulation - MPE ενθυλακώνει τα IP δεδομένα σε ροές μεταφοράς οι οποίες αποστέλλονται μέσω δικτυακών πλατφόρμων DVB. Εκτός από την επίγεια ψηφιακή τηλεόραση το πρότυπο MPEG-2 αξιοποιείται ακόμα στην δορυφορική τηλεόραση (Satellite

Digital TV), στην καλωδιακή τηλεόραση (Cable TV) και στους δίσκους DVD - Digital Video Disk.

#### 4.3.2.1 Το Πρότυπο MPEG - 2 LSF

Το πρότυπο ISO/IEC 13818- MPEG-2 LSF [Παρ95] (Lower Sampling Frequency) παρουσιάστηκε το 1994, αποτελεί μία επέκταση του MPEG-1 και στην ουσία μειώνει την παροχή δεδομένων με τη χρήση συχνότητας δειγματοληψίας χαμηλότερης ή ίσης των 32 kHz. Η παροχή που επιτυγχάνει κυμαίνεται στην περιοχή 16-64 kbits/s, χρησιμοποιείται για την κωδικοποίηση φωνή ευρείας ζώνης ή μουσικής μέσης ποιότητας και προορίζεται για χρήση κυρίως σε τηλεδιασκέψεις και γενικά σε εφαρμογές που επιβάλλεται χαμηλή παροχή δεδομένων. Η ποιότητα που επιτυγχάνει στα 64 kbits/s για συνδυασμένο στερεοφωνικό σήμα και με συχνότητα δειγματοληψίας 24 kHz προσεγγίζει την ποιότητα του ASPEC για παροχή 64 kbits/s και για μονοφωνικό σήμα στα 44.1 kHz.

#### 4.3.2.2 Το Πρότυπο MPEG - 2 BC

Το Πρότυπο MPEG-2 BC [Παρ95] (Backward Compatible) στην πραγματικότητα αποτελεί επέκταση του MPEG-1, για εφαρμογές πολυκαναλικής κωδικοποίησης. Έχει συμβατότητα με το MPEG-1 και προσφέρει καλή ηχητική ποιότητα για παροχή δεδομένων από 640 μέχρι 896 kbits/s για πέντε κανάλια πλήρους φάσματος. Τυποποιήθηκε με το ίδιο πρότυπο με το MPEG-2 LSF το 1994 . Χρησιμοποιείται επίσης για την πολυγλωσσική μετάδοση σε δορυφορικά τηλεοπτικά προγράμματα. Ο τρόπος με τον οποίο υλοποιεί την πολυκαναλική κωδικοποίηση βασίζεται στην κωδικοποίηση με τον MPEG-1 αλγόριθμο των σημάτων  $L_C$  και  $R_C$ , όπου:

$$L_C = L + \alpha C + \beta L_S$$

$$R_C = R + \alpha C + \beta R_S$$

L και R είναι τα αρχικά δεξί και αριστερό κανάλια αντίστοιχα, C είναι το κεντρικό κανάλι και  $L_S$  και  $R_S$  είναι το δεξί και αριστερό κανάλια του περιβάλλοντος ήχου.

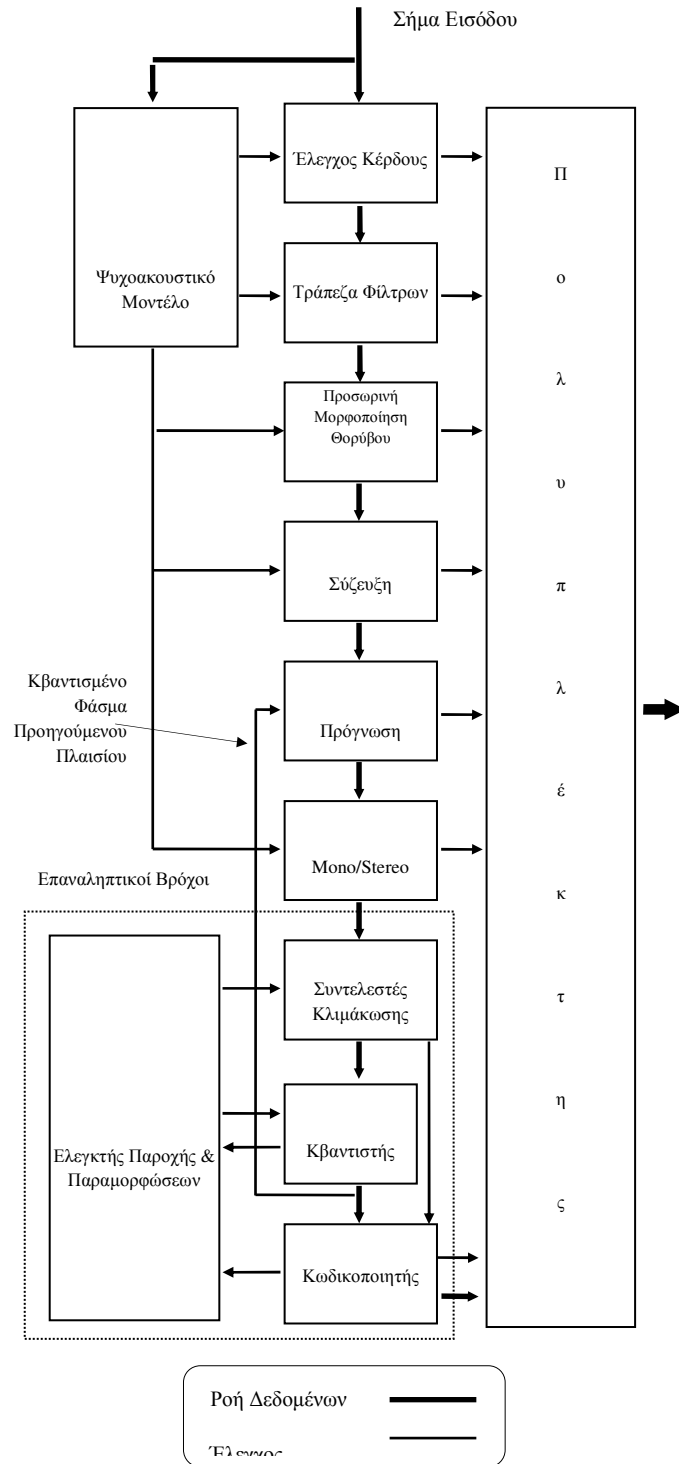
### 4.3.2.3 Το πρότυπο MPEG - 2 AAC

Το πρότυπο MPEG-2 AAC [Παρ95] (Advanced Audio Coding) αποτελεί το πρότυπο αναφοράς για υψηλής ποιότητας κωδικοποίηση ηχητικού σήματος πλήρους φάσματος στα 64 kbits/s. Τυποποιήθηκε τον Απρίλιο του 1997 με το πρότυπο ISO/IEC 13818-7 και μπορεί να κωδικοποιήσει πολυκαναλικό σήμα 5 καναλιών στα 320 kbits/s με σχεδόν άριστη ηχητική ποιότητα. Τα βασικά του χαρακτηριστικά είναι:

- Η τμηματική του σχεδίαση
- Η υποστήριξη πολλών συχνοτήτων δειγματοληψίας στα 32, 44.1 και 48 kHz
- Η ικανότητα μονοφωνικής, στερεοφωνικής και πολυκαναλικής κωδικοποίησης

Το πρότυπο αυτό μπορεί να λειτουργήσει σε τρεις διαφορετικούς τρόπους:

- **Κύριος Τρόπος**, στον οποίο είναι ενεργοποιημένες όλες οι μονάδες, εκτός από τον Έλεγχο Κέρδους. Αποδίδει την υψηλότερη ποιότητα για εφαρμογές που δεν τίθενται περιορισμοί διατιθέμενης μνήμης.
- **Τρόπος Χαμηλής Πολυπλοκότητας**, στον οποίο απενεργοποιείται επιπλέον και το τμήμα της Πρόγνωσης και μειώνεται η τάξη του φίλτρου Προσωρινής Μορφοποίησης του Θορύβου.
- **Τρόπος Κλιμακωτής Παροχής**, στον οποίο και σε σχέση με τον τρόπο χαμηλής πολυπλοκότητας, ενεργοποιείται η βαθμίδα Ελέγχου Κέρδους.



Σχήμα 4.5: Διάγραμμα λειτουργίας του κωδικοποιητή MPEG-2 AAC.

Η φιλοσοφία με την οποία σχεδιάστηκε το πρότυπο MPEG-2 AAC, είναι ότι για συγκεκριμένη επιθυμητή παροχή δεδομένων είναι δυνατό να επιλεγθεί η βέλτιστη δομή



ανάμεσα σε 16 προκαθορισμένες δομές, ώστε να ικανοποιηθούν διάφορα κριτήρια σχετικά με πολυπλοκότητα, καθυστέρηση, αντοχή σε σφάλματα, κλπ.

#### 4.4 Το Πρότυπο MPEG - 4

Το σύστημα αυτό θα αποτελέσει το πρότυπο για τη χρήση των πολυμέσων, ενώ ελπίζεται ότι θα καλύψει το χάσμα ανάμεσα στα ηλεκτρονικά μέσα ενημέρωσης και στις τηλεπικοινωνίες. Θα εξασφαλίζει υψηλό κέρδος συμπίεσης και καθολική προσπελασιμότητα. Οι εφαρμογές που θα ευνοηθούν θα είναι η προσπέλαση βάσεων δεδομένων, οι τηλεπικοινωνίες και γενικά όλες οι διαδραστικές υπηρεσίες [Παρ95]. Θα έχει τη δυνατότητα να χειρίζεται κάθε τύπο ηχητικού υλικού, δηλαδή μουσική υψηλής ποιότητας (Mono, Stereo και Surround), συνθετική μουσική, φωνή ευρείας και στενής ζώνης, καθώς επίσης και συνθετική φωνή.

Για την κωδικοποίηση μουσικής ευρείας ζώνης θα χρησιμοποιεί τον αλγόριθμο MPEG-2 AAC στα 64 kbits/s για κάθε κανάλι, ενώ για μουσική μεσαίας ποιότητας τον αλγόριθμο MPEG-2 LSF.

#### 4.5 Το Πρότυπο MPEG - 7

Το πρότυπο του MPEG-7 [link2] πρωτοπαρουσιάστηκε από την ομάδα ειδικών κινούμενων εικόνων (Moving Pictures Expert Group) και έναντι των προηγούμενων MPEG προτύπων (MPEG-1, MPEG-2, MPEG-4) τα οποία απευθύνονται στην κωδικοποιημένη αναπαράσταση οπτικοακουστικού περιεχομένου, το MPEG-7 επικεντρώνεται στην αναπαράσταση πληροφορίας περί του περιεχομένου, παρά για το ίδιο το περιεχόμενο. Στόχος του MPEG-7 είναι να προσφέρει ένα πλούσιο σύνολο εργαλείων για την περιγραφή πολυμεσικού -μη λεκτικού- περιεχομένου, τα οποία θα βοηθήσουν στην κατηγοριοποίηση, αναζήτηση και ανάκτηση αντίστοιχης μορφής δεδομένων.

Το MPEG-7 καθορίζεται από ένα σύνολο από ψηφιακούς περιγραφείς (descriptors) και σχήματα περιγραφής (description schemes), μέσω των οποίων καθιστάτε εφικτή η περιγραφή του ψηφιακού περιεχομένου μη λεκτικής πληροφορίας όπως ήχοι, εικόνες, βίντεο τρισδιάστατα μοντέλα κλπ. Για κάθε τύπο πληροφορίας υπάρχουν και οι αντίστοιχοι βασικοί

περιγραφείς οι οποίοι συνδυαζόμενοι μεταξύ τους παράγουν πιο σύνθετους περιγραφείς. Ωστόσο, σε περίπτωση που οι υπάρχον περιγραφείς και ο συνδυασμός τους (σχήματα περιγραφής) δεν ικανοποιούν τις ανάγκες χαρακτηρισμού του περιεχόμενου μιας μη λεκτικής πληροφορίας, τότε υπάρχει η δυνατότητα να κατασκευαστεί ένας νέος τύπος περιγραφής, μέσω της γλώσσας ορισμού περιγραφής (Description Definition Language – DDL) που προβλέπει πρότυπο του MPEG-7 και συντάσσεται κατά τα πρότυπα της γλώσσας XML.

Η παραγόμενη πληροφορία που προκύπτει από τους περιγραφείς συντάσσεται, σύμφωνα με πρότυπα που ορίζει το MPEG-7, ακολουθώντας τη δομή της DDL και είναι συνυφασμένη με τα δεδομένα τα οποία περιγράφει. Με αυτό τον τρόπο επιτυγχάνεται γρήγορη και αποτελεσματική αναζήτηση πολυμεσικών δεδομένων, σύμφωνη με τα κριτήρια που έχει θέσει ο χρήστης της μηχανής αναζήτησης και όχι αυτός που δημοσίευσε, με τυχόν λάθος λεκτική περιγραφή, την πληροφορία στο διαδίκτυο.

## **4.5 ΑΛΛΑ ΠΡΟΤΥΠΑ ΨΗΦΙΑΚΟΥ ΗΧΟΥ**

Παρά το γεγονός της ριζικής και της ραγδαίας εξέλιξης του MP3 το οποίο έχει γνωρίσει ευρεία αποδοχή μεταξύ χρηστών, παρόλα αυτά επικρατεί πληθώρα διαφορετικών τεχνολογιών που ολοένα και εξελίσσονται με πρωταρχικό στόχο την καλύτερη συμπίεση ψηφιακού ήχου. Τέτοιες προσπάθειες, έχουν υλοποιηθεί από την εταιρία Microsoft με το δικό της γνωστό και διαδεδομένο πρότυπο Windows Media Audio (WMA), το AAC (Advanced Audio Coding), το AC-3 Dolby Digital, το MP3 Pro και το OGG [Χρυ06].

### **4.5.1 WMA (Windows Media Audio)**

Η μεταφορά, η αποθήκευση και η χρήση ακουστικού υλικού με βάση την μορφή απωλεστικών συμπιεσμένων ηχητικών αρχείων μέσω υπολογιστή, παροτρύνει τον μεγαλύτερο κατασκευαστή λειτουργικών συστημάτων να μην μείνει έξω από το παιχνίδι. Το πρότυπο Windows Media Audio (WMA) [Παρ95] της εταιρία Microsoft, προσφέρει όμοιες δυνατότητες με το MP3, με άριστη ποιότητα τόσο αναπαραγωγής όσο και μεγαλύτερη συμπίεση (64 kbps). Πιο συγκεκριμένα, το WMA αποτελεί ένα σύστημα κωδικοποίησης/αποκωδικοποίησης ήχου, επιτρέποντας την συμπίεση ψηφιακών δεδομένων

ήχου στο 1/20 του αρχικού τους όγκου και την εγγραφή τους σε ένα μόνο δίσκο CD με αποτέλεσμα τα τραγούδια που είναι προστατευμένα να μην μπορούν να μεταδοθούν ελεύθερα. Συμπερασματικά, γι' αυτό ακριβώς το λόγω ο μεγαλύτερος αριθμός δισκογραφικών εταιριών χρησιμοποιεί στα πλαίσια υλοποίησης του έργου τους το πρότυπο αυτό.

#### 4.5.2 MP3 Pro

Τον Ιανουάριο του 2001, παρουσιάστηκε από την Coding Technologies η τεχνολογία Mp3 Pro [Χρυ06], μια βελτιωμένη έκδοση του Mp3 με δυνατότητα να προσφέρει όμοια ποιότητα στο μισό μέγεθος των αρχείων, γεγονός που υλοποιείται με μεγαλύτερη συμπίεση δεδομένων. Συγκεκριμένα, η συμπίεση στα 64kbps και 96kbps, προσφέρει τη ίδια απόδοση ήχου με τα 128kbps και 192kbps του Mp3.

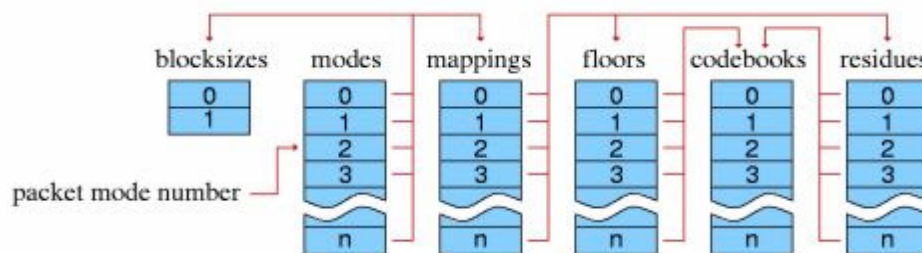
#### 4.5.3 AAC (Advance Audio Coding)

Το πρότυπο AAC [Χρυ06] αναπτύχθηκε για πρώτη φορά το 1997 από το Ινστιτούτο Fraunhofer IIS και χρησιμοποιεί όπως και το MP3 το ψυχοακουστικό μοντέλο επικάλυψης, με σκοπό να καλύψει τα προβλήματα που υπήρχαν σχετικά με την ποιότητα των αρχείων MP3 στους μικρούς αριθμούς μετάδοσης των δεδομένων. Το πρότυπο AAC, ως κωδικοποίηση έχει την ικανότητα να αποδίδει εξαιρετικά υψηλή ποιότητα ήχου σε bitrate 64Kb/sec, επιτρέποντας τόσο την κωδικοποίηση 48 καναλιών ήχου και 16 καναλιών χαμηλής συχνότητας για εφέ όσο και την υποστήριξη πολλών γλωσσών ταυτόχρονα. Παράλληλα, το AAC διακατέχεται από τρεις διαφορετικές όψεις, την «κύρια», την «χαμηλής πολυπλοκότητα» και την «κλιμακούμενη συχνότητα δειγματοληψίας». Η «κύρια» όψη αναφέρεται σε εφαρμογές που η υπολογιστική ισχύει και εφαρμογές δεν είναι περιορισμένη, η «χαμηλής πολυπλοκότητα» σε εφαρμογές που η ισχύος και η μνήμη βρίσκονται σε μεγάλη ζήτηση, ενώ η τελευταία, είναι φτιαγμένη έτσι ώστε οι αποκωδικοποιητές να έχουν ελάχιστες απαιτήσεις σε μνήμη και ισχύ. Υποκειμενικά τεστ που έχουν πραγματοποιηθεί με καλά εκπαιδευμένους ακροατές, έδειξαν ότι η συγκεκριμένη

κωδικοποίηση προσφέρει καλύτερη ποιότητα ήχου από οποιαδήποτε άλλη κωδικοποίηση ήχου με το μισό μόνο bitrate.

#### 4.5.4 OGG Vorbis

Ο codec Ogg Vorbis [Χρυ06] αναπτύχθηκε γύρω από το πρότυπο αρχείων Ogg και βασίζεται στη open source εφαρμογή απωλεστικής συμπίεσης με την ονομασία Vorbis. Ως προς τον τρόπο κωδικοποίησης, ο τρόπος μοιάζει με αυτό του Mp3 ενώ ταυτόχρονα ο Ogg Vorbis χρησιμοποιεί την μέθοδο επεξεργασίας του φάσματος, κατά την οποία κωδικοποιείται το φάσμα βάσης του οποίου η κατανομή είναι σχετικώς ομαλή και με περισσότερα ψηφία, το απομένον φάσμα που η δομή και η χρονική εξέλιξη είναι πολύ πιο πολύπλοκη (σχήμα 4.5). Συμπερασματικά, η τακτική αυτή σε συνδυασμό με την καλή ποιότητα ήχου που προσφέρει, ωθεί το πρότυπο Ogg σε ένα ανταγωνιστικό παιχνίδι ως προς το WMA και MP3.



Σχήμα 4.5: Διάγραμμα βαθμίδων του κωδικοποιητή Ogg Vorbis. Η κωδικοποίηση του φάσματος βάσης και του απομένοντος φάσματος.

#### 4.5.5 AC-2 DOLBY DIGITAL

Ο αλγόριθμος AC-2 (Audio Coding) [Παρ95] της Dolby είναι μία τεχνική κωδικοποίησης βασισμένη σε μαθηματικό μετασχηματισμό. Είναι εμπορικά διαθέσιμος από το 1989 και χαρακτηρίζεται από τη χαμηλή πολυπλοκότητα και καθυστέρηση (7 έως 60 ms) και το μεσαίο κέρδος συμπίεσης. Είναι σχεδιασμένη στη βάση της πλήρους ανεξαρτησίας μεταξύ των καναλιών εισόδου και η παροχή της κυμαίνεται από 128 μέχρι 192 kbits/s, για σήμα εύρους ζώνης 20 kHz. Χρησιμοποιείται κυρίως σε εφαρμογές καλωδιακής τηλεόρασης.

Υλοποιεί τους τροποποιημένους μετασχηματισμούς ημιτόνου και συνημιτόνου (MDST/MDCT) με ακύρωση του φαινομένου της αναδίπλωσης στο πεδίο του χρόνου (Time-Domain Aliasing Cancelation - TDAC). Η διαδικασία ανάθεσης των bits ελέγχεται από μια εκτίμηση της φασματικής περιβάλλουσας (spectral envelope), που επιτρέπει μόνο σε ένα ποσοστό της τάξης του 20% του συνολικού ρυθμού δεδομένων να προσαρμόζεται στα χαρακτηριστικά του σήματος. Για τον κβαντισμό χρησιμοποιείται ομοιόμορφος κβαντιστής και όλες οι συνιστώσες με συχνοτικό περιεχόμενο μεγαλύτερο των 20 kHz μηδενίζονται, δημιουργώντας έτσι ένα ανεπιθύμητο χαρακτηριστικό ζωνοδιαβατού φίλτρου. Η μέθοδος διόρθωσης σφαλμάτων που χρησιμοποιείται είναι η Read-Solomon.

#### 4.5.6 AC-3 DOLBY DIGITAL

Ένα από τα πιο διαδεδομένα πρότυπα για τον ψηφιακό πολυκάναλο ήχο είναι το AC-3 [Παρ95] [Χρυ06], που εκτός απροόπτου άρχισε να γίνεται το διεθνές πρότυπο για την συμπίεση ηχητικών δεδομένων. Η παροχή του κυμαίνεται από 32 μέχρι 640 kbits/s και είναι το πρότυπο κωδικοποίησης ήχου για την τηλεόραση υψηλής ευκρίνειας (HDTV) στη Βόρειο Αμερική, ενώ χρησιμοποιείται στο DVD και στον ψηφιακό κινηματογράφο. Στο ψηφιακό σύστημα ήχου AC-3, ο ήχος κωδικοποιείται σε έξι συνολικά κανάλια. Συγκεκριμένα, υπάρχουν: (α) τρία κανάλια (αριστερό, κεντρικό, δεξί) που αποσκοπούν στο να φέρουν την βασική ηχητική πληροφορία, (β) δυο συνοδευτικά κανάλια περιβάλλοντος ήχου και (γ) ένα κανάλι για τις υπόλοιπες συχνότητες. Ως προς τον τρόπο λειτουργίας τους, τα πέντε πρώτα κανάλια διαχειρίζονται συχνότητες ήχου στο διάστημα 3-20000Hz, ενώ το έκτο συχνότητες 3-120Hz. Συμπερασματικά, ο ρυθμός δειγματοληψία είναι 48KHz μεγαλύτερος από το ρυθμό των 44KHz που χρησιμοποιείται στα CDs, ενώ η συμπίεση των ηχητικών δεδομένων ανέρχεται στην αναλογία 10:1. Ο αλγόριθμός της βασίζεται σε MDCT με χαρακτηριστικά TDAC, με τυπικό μήκος μετασχηματισμού τα 256 δείγματα, που αποδίδει συχνοτική ευκρίνεια 93.75 Hz στα 48 kHz και χρονική 2.66 ms. Η διαδικασία ανάθεσης των bits είναι πλήρως προσαρμοζόμενη στο σήμα εισόδου και υλοποιείται σε δυο στάδια. Στο πρώτο υπολογίζεται ένα πλήρες ψυχοακουστικό μοντέλο και στο δεύτερο πραγματοποιείται μια παραμετρική ανάθεση bits ανά φασματικό συντελεστή. Οι κβαντισμένες τιμές συνδυάζονται ανά δυο ή ανά τρεις σε πακέτα, αυξάνοντας έτσι το κέρδος συμπίεσης.

Για την κωδικοποίηση στερεοφωνικού σήματος η τεχνική συνδυάζει τους συντελεστές μετασχηματισμού δεξιού - αριστερού καναλιού ή αθροίσματος - διαφοράς, ελαχιστοποιώντας έτσι την ακουστότητα των παραμορφώσεων στην περίπτωση της λειτουργίας περιβάλλοντος ήχου (surround sound). Η καθυστέρηση της μεθόδου είναι περίπου 100 ms.

#### 4.5.7 ATRAC

Ο αλγόριθμος ATRAC (Adaptive Transform Acoustic Coding) [Παρ95] παρουσιάστηκε το 1992 από την εταιρεία Sony και χρησιμοποιείται στο MiniDisk. Το βασικό της χαρακτηριστικό είναι η μεταβαλλόμενη συχνοτική της ευκρίνεια ανά φασματική περιοχή, προσομοιώνοντας έτσι καλύτερα τις ιδιότητες του ανθρώπινου ακουστικού συστήματος. Επιπλέον, το μήκος του παραθύρου ανάλυσης μεταβάλλεται ανάλογα με τα στατιστικά του σήματος. Η ανάθεση των bits γίνεται με παρόμοιο τρόπο με την AC-2 και η παροχή δεδομένων που επιτυγχάνει είναι 140 kbits/s ανά κανάλι.

#### 4.5.8 PAC

Ο αλγόριθμος PAC (Perceptual Audio Coding) [Παρ95] βασίζεται σε μία τράπεζα φίλτρων υψηλής συχνοτικής ανάλυσης, που υλοποιείται από έναν MDCT 128-1024 σημείων, επιτρέποντας τον αποτελεσματικό χειρισμό του φαινομένου της προ-ηχούς με τη χρήση μικρότερου παραθύρου. Η απόφαση για την αλλαγή του μήκους του παραθύρου δίνεται ύστερα από μια μέτρηση της Υποκειμενικής Εντροπίας, όπως ακριβώς γίνεται και στο επίπεδο III του MPEG-1. Μια άλλη ομοιότητα ανάμεσα στα δυο πρότυπα είναι το ψυχοακουστικό μοντέλο που υλοποιούν. Ο κβαντισμός των φασματικών τιμών γίνεται από έναν ομοιόμορφο κβαντιστή και η κωδικοποίηση από έναν διαφορικό Huffman κωδικοποιητή. Οι συντελεστές κλιμάκωσης κωδικοποιούνται επίσης διαφορετικά. Η πολυκαναλική εκδοχή του PAC ονομάζεται MPAC.

## 4.6 ΤΥΠΟΙ ΑΡΧΕΙΩΝ ΨΗΦΙΑΚΟΥ ΗΧΟΥ

Ανάλογα με την διαδικασία παραγωγής του, τα αρχεία ήχου διακρίνονται σε δυο κατηγορίες: α) τα αρχεία ήχων κυματομορφής και β) αρχεία MIDI [Χρυ06].

Πιο συγκεκριμένα, τα αρχεία ήχων κυματομορφής, παράγονται με την διαδικασία της ψηφιοποίησης που σκοπό έχουν την παροχή ακριβής ψηφιακής εικόνας της κυματομορφής του ήχου. Έτσι, τα προγράμματα επεξεργασίας ήχων αυτής της κατηγορίας, αναπαριστούν τον ήχο ως κυματομορφή δίνοντας την ευκαιρία στο χρήστη να κόψει, να αντιγράψει, να εισάγει και να τροποποιήσει είτε ένα μέρος είτε ολόκληρη την κυματομορφή. Οι βασικές δυνατότητες αυτών των εργαλείων είναι η:

1. εισαγωγή ήχου
2. αντιγραφή, αποκοπή, επικόλληση ηχητικού τμήματος
3. ρύθμιση έντασης και συχνότητας
4. απαλοιφή θορύβου
5. εισαγωγή διαφορών εφέ
6. τροποποίηση επιπέδων κβάντωσης
7. τροποποίηση χρονικής διάρκειας
8. συμπίεση

Η δεύτερη κατηγορία αρχείων ήχου σχετικά με την διαδικασία παραγωγής τους είναι το αρχείο MIDI, το οποίο δεν καταγράφει το ηχητικό σήμα αλλά τα στάδια που πραγματοποιούνται για να παραχθεί αυτό. Συγκεκριμένα, αντί να αποθηκεύσουμε τους ήχους μετατρέποντας το αναλογικό σήμα σε ψηφιακό αποθηκεύουμε τέσσερις αριθμούς για κάθε νότα που παίζεται, δηλαδή ο πρώτος μας δίνει ποια νότα είναι, ο δεύτερος μας λέει πιο όργανο την παίζει, ο τρίτος μας πληροφορεί πόσο δυνατά παίζεται η νότα και ο τέταρτος μας δίνει το χρόνο. Επειδή η διαδικασία αναπαράστασης μοιάζει με αυτή του κώδικα ASCII, πολλές φορές το MIDI αποκαλείται και ως «κώδικα ASCII της μουσικής». Όσο αφορά τα προγράμματα επεξεργασίας αρχείων MIDI αυτά παρέχουν την δυνατότητα αναπαράστασης του ήχου ως παρτιτούρα καθώς και την δυνατότητα παροχής στο χρήστη, ώστε να τοποθετήσει νότες, παύσεις και άλλα σύμβολα της μουσικής πάνω στο πεντάγραμμο καθορίζοντας τον τρόπο εκτέλεσης της κάθε νότας.

Συμπερασματικά, ανάλογα με την πλατφόρμα και το πρόγραμμα που χρησιμοποιείται για την παραγωγή των ψηφιακών ήχων, τα αρχεία αποθηκεύονται στον ηλεκτρονικό

υπολογιστή με διαφορετική μορφή η οποία αναγνωρίζεται από τη έκταση του αρχείου (σχήμα 4.6).

.rif	RIFF	Αναπτύχθηκε από την Microsoft και υποστηρίζει αρχεία ψηφιακού ήχου WAV, MIDI.
.wav	Wave	Αποτελεί το πρότυπο αποθήκευσης ψηφιακού ήχου και είναι υποσύνολο του προτύπου RIFF.
.mid	MIDI	Διεθνές πρότυπο για την αποθήκευση μουσικών αρχείων MIDI.
.aif	AIFF	Δημιουργήθηκε από την Apple αλλά υποστηρίζεται και από άλλες πλατφόρμες, υποστηρίζει δειγματοληψία 32 bit.
.rmi	RMI	Αναπτύχθηκε από την Microsoft για την υποστήριξη αρχείων MIDI.
.mp3	MPEG-layer3	Πρότυπο συμπίεσης αρχείων ήχου που χρησιμοποιείται κυρίως για την διακίνηση αρχείων μουσικής στο Διαδίκτυο
.wma	WMA	Windows Media Audio από την Microsoft. Στόχο έχει την υποστήριξη απωλεστικής συμπίεσης
.ra	Real Audio	Προορίζεται για την άμεση αναπαραγωγή ήχων μέσω του Διαδικτύου

**Σχήμα 4.6: Τύποι Αρχείων Ήχου.**



## ΚΕΦΑΛΑΙΟ 5

### ΥΛΟΠΟΙΗΣΗ ΤΟΥ ΨΥΧΟΑΚΟΥΣΤΙΚΟΥ ΜΟΜΤΕΛΟΥ ΜΕ ΧΡΗΣΗ MATLAB

#### 5.1 ΕΙΣΑΓΩΓΗ

Το MATLAB είναι ένα σύγχρονο ολοκληρωμένο μαθηματικό λογισμικό πακέτο που χρησιμοποιείται σε πανεπιστημιακά μαθήματα αλλά και ερευνητικές και άλλες εφαρμογές με επιστημονικούς υπολογισμούς (scientific computing). Το όνομα του προέρχεται από τα αρχικά γράμματα των λέξεων MATtrix LABoratory (εργαστήριο πινάκων). Το MATLAB είναι ένα διαδραστικό (interactive) πρόγραμμα για αριθμητικούς υπολογισμούς και οπτικοποίηση δεδομένων (data visualization) με δυνατότητες προγραμματισμού που το καθιστούν ένα ισχυρό και χρήσιμο εργαλείο στις μαθηματικές και φυσικές επιστήμες. Το MATLAB στις αρχικές του εκδοχές δεν έκανε συμβολικούς υπολογισμούς. Στις νεότερες εκδοχές του, το πακέτο περιλαμβάνει εργαλείοθήκες που επιτρέπουν συμβολικούς υπολογισμούς. Όπως υποδηλώνεται και από το όνομα του, το MATLAB είναι ειδικά σχεδιασμένο για υπολογισμούς με πίνακες, όπως η επίλυση γραμμικών συστημάτων, η εύρεση ιδιοτιμών και ιδιοδιανυσμάτων, η αντιστροφή τετραγωνικών πινάκων κλπ. Επιπλέον το πακέτο αυτό είναι εφοδιασμένο με πολλές επιλογές για γραφικά (δηλ. την κατασκευή γραφικών παραστάσεων) και προγράμματα γραμμένα στη δική του γλώσσα προγραμματισμού για την επίλυση άλλων προβλημάτων όπως η εύρεση των ριζών μη γραμμικής εξίσωσης, η επίλυση μη γραμμικών συστημάτων, η επίλυση προβλημάτων αρχικών τιμών με συνήθεις διαφορικές εξισώσεις κτλ. Η γλώσσα προγραμματισμού του MATLAB δίνει την ευχέρεια στον χρήστη να το επεκτείνει με δικά του προγράμματα. Συχνά θα γράφουμε στη MATLAB (εννοώντας τη γλώσσα προγραμματισμού) και όχι το (πακέτο) MATLAB. Το MATLAB είναι σχεδιασμένο για την αριθμητική επίλυση προβλημάτων σε αριθμητική πεπερασμένης ακρίβειας (finite-precision arithmetic), δηλαδή δεν βρίσκει την ακριβή αλλά μια προσεγγιστική λύση ενός προβλήματος. Αυτή είναι και η βασική του διαφορά από τα συστήματα συμβολικών υπολογισμών όπως η Maple και το Mathematica. Ας σημειωθεί ότι ο καλύτερος (και ουσιαστικά ο μόνος) τρόπος εκμάθησης της MATLAB είναι η συστηματική ενασχόληση με αυτή και η διερεύνησή της από τον ίδιο τον χρήστη. Το πακέτο είναι εφοδιασμένο με ένα εκτενές σύστημα βοήθειας όπου κάθε εντολή επεξηγείται

αναλυτικά και με αντιπροσωπευτικά παραδείγματα. Η πιο σημαντική εντολή της MATLAB είναι η `help` (βοήθεια)! Επίσης, στην επίσημη ιστοσελίδα της MATLAB: <http://www.mathworks.com> μπορεί κάποιος να βρει μια πληθώρα πληροφοριών τόσο για αρχάριους όσο και για προχωρημένους.

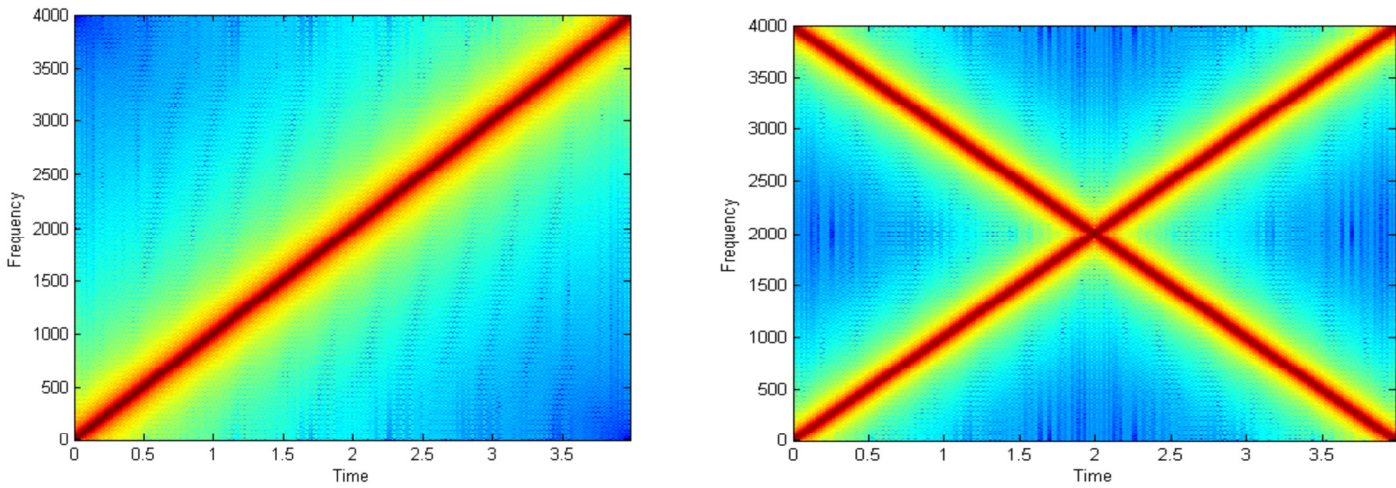
## 5.2 Η ΑΠΟΔΕΙΞΗ ΤΗΣ ΕΝΝΟΙΑΣ : ASP\_mp3.m ΜΕΣΩ ΤΟΥ MATLAB

Σε αυτή την ενότητα, παρέχουμε μια απόδειξη της έννοιας του ακουστικού κανόνα κωδικοποίησης του MPEG-1Layer-I. Αρχίζουμε με την εξέταση μιας τράπεζας φίλτρου δύο καναλιών χρησιμοποιώντας συμβατικά φίλτρα (Ενότητα 5.1.1) και φίλτρα QMF (Ενότητα 5.1.2). Εμείς στη συνέχεια θα επεκτείνουμε τη μελέτη μας σε μια τράπεζα φίλτρου 32-καναλιών PQMF (Ενότητα 5.1.3) και θα δείξουμε πώς μπορεί να εφαρμοστεί αποτελεσματικά χρησιμοποιώντας το block-based που βασίζεται στους επικαλυπτόμενους μετασχηματισμούς (Ενότητα 5.1.4). Με την παροχή τράπεζας φίλτρου ολοκληρώνουμε με έναν αντιληπτικό κβαντιστή για σήματα sub-band (Ενότητα 5.1.5).

### 5.2.1 Η ΤΡΑΠΕΖΑ ΦΙΛΤΡΩΝ ΔΥΟ ΚΑΝΑΛΙΩΝ

Πριν από την εξέταση μιας ολοκληρωμένης τράπεζας φίλτρων M-καναλιού, θα συζητήσουμε πρώτα την περίπτωση 2 καναλιών, καθώς και την επίδραση του αποδεκατισμού και της παρεμβολής σε κάθε κλάδο. Ας δημιουργήσουμε πρώτα ένα chirp signal 4-δευτερολέπτων (0-4 kHz) με έναν ρυθμό δειγματοληψίας 8 kHz, το οποίο θα χρησιμοποιηθεί ως σημείο αναφοράς σε όλη αυτή την ενότητα και θα σχεδιαστεί το φασματογράφημα του (Σχήμα 5.1 αριστερά).

```
Fs=8000;  
input_signal=chirp((0:4*Fs)/Fs,0,4,4000);  
specgram(input_signal,1024,Fs,256);
```



Σχήμα 5.1: Φασματογράφημα chirp signal. Αριστερά: πρωτότυπο σήμα Δεξιά: μετά downsampling- upsampling χωρίς την χρήση φίλτρων ανάλυσης και σύνθεσης.

Εφαρμόζοντας απευθείας downsampling-upsampling με 2 αποτελέσματα αντικαθιστούμε κάθε άλλο δείγμα με το μηδέν. Το αποτέλεσμα είναι ότι προστίθεται τεχνούργημα, σήματος του οποίου το φάσμα είναι η εικόνα τετραγωνισμού καθρέφτη του chirp signal. Δύο αιμοφόρα αγγεία μπορούν να ακουστούν ανά πάσα στιγμή (Σχήμα 5.1 Δεξιά).

```
downsampled=input_signal(1:2:end);
upsampled(1:2:2*length(downsampled))=2*downsampled;
```

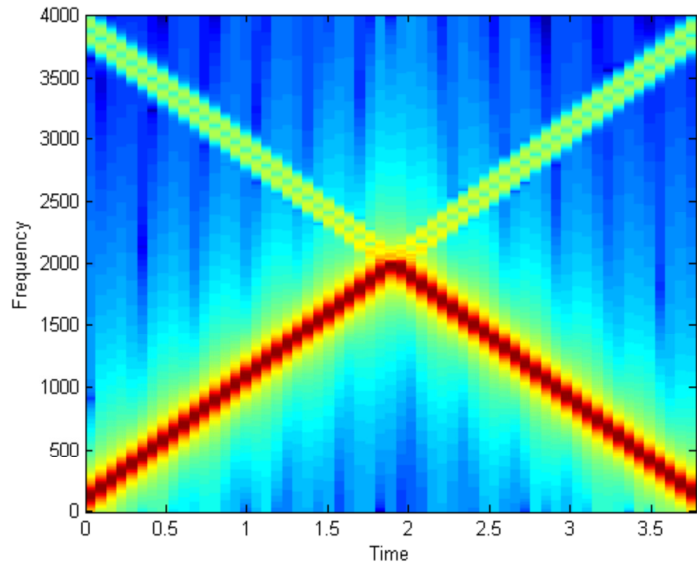
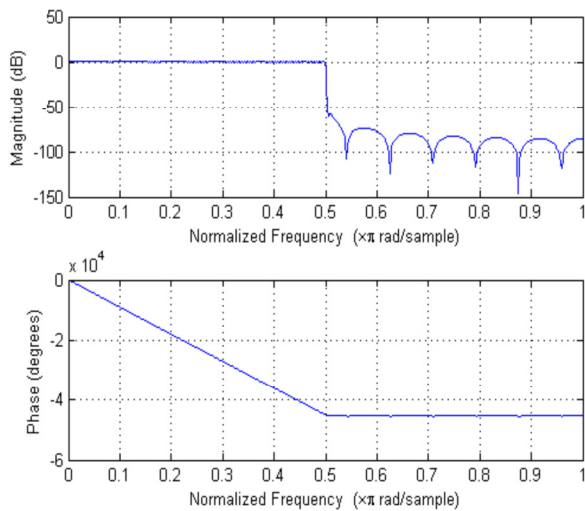
Προσθέτοντας ένα φίλτρο quarter-band μετά από μετατροπή δειγματοληψίας (upsampling) εξαλείφεται η παραμόρφωση σήματος της εικόνας, διασφαλίζοντας ότι εμφανίζεται μόνο μία ημιτονοειδής ανά πάσα στιγμή (Σχήμα 5.2). Κατά το πρώτο μισό του chirp signal, σήματος ημιτονοειδής αποτελεί το ίδιο το chirp signal κατά δεύτερο μισό, σήματος η ημιτονοειδής είναι ένα ψευδώνυμο του σήματος που οφείλεται στην εικόνα τετραγωνισμού καθρέφτη του chirp signal. Σχεδιάζουμε το φίλτρο σύνθεσης ως ένα συμμετρικό FIR, έτσι ώστε να αποφεύγεται η παραμόρφωση φάσεως και θέσαμε τη διάταξή του σε 1000, έτσι ώστε να ληφθεί υψηλής stop-band-rejection (κοντά στο 80 dB). Παρατηρείται ότι τα πρώτα 1000 δείγματα είναι παροδικά.

```

G0=fir1(1000,1/2);
freqz(G0,1);
G0_output=filter(G0,1,upsampled);

specgram(G0_output(1001:end),1024,Fs);

```



Σχήμα 5.2: Αριστερά: Απόκριση συχνότητας του φίλτρου quarter-band ανάλυσης. Δεξιά: Φασματογράφημα \_ chirp signal μετά από downsampling-upsampling , χρησιμοποιώντας μια low-pass σύνθεση μόνο.

Η προσθήκη μιας quarter-band ανάλυσης φίλτρου πριν το downsampling αφαιρεί την παραμόρφωση aliasing, στο δεύτερο μισό του chirp signal (Σχήμα 5.3 αριστερά). Αυτό το φίλτρο σύνθεσης μπορεί να είναι εξαρχής πανομοιότυπο με το φίλτρο ανάλυσης.

```

H0=G0;
H0_output=filter(H0,1,input_signal);

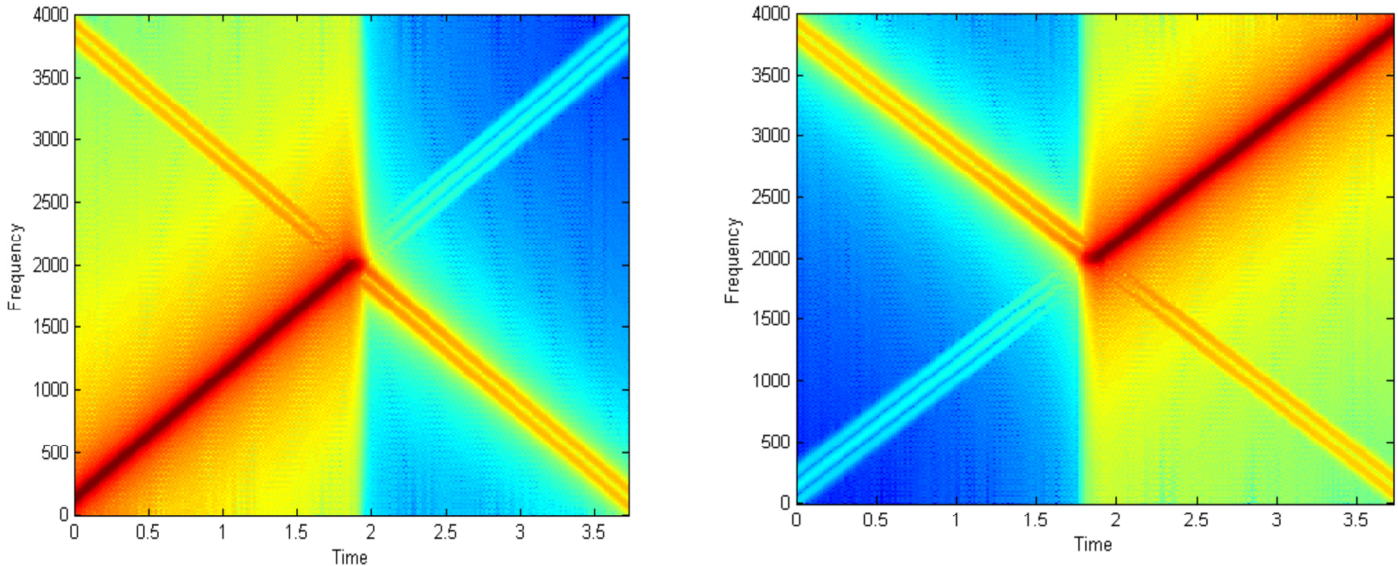
```

```

downsampled=H0_output(1:2:end);
upsampled(1:2:2*length(downsampled))=2*downsampled;
G0_output=filter(G0,1,upsampled);

specgram(G0_output(2001:end),1024,Fs, 256);

```



Σχήμα 5.3: Αριστερά: Φασματογράφημα ενός chirp downsampling-upsampling , με τη χρήση φίλτρων ανάλυσης και σύνθεσης low-pass. Δεξιά: Το ίδιο, με φίλτρα ανάλυσης και σύνθεσης high pass.

Χρησιμοποιώντας HF quarter-band φίλτρα αντί LF επιλέγεται το δεύτερο μισό του chirp (Σχήμα 5.3 Δεξιά).

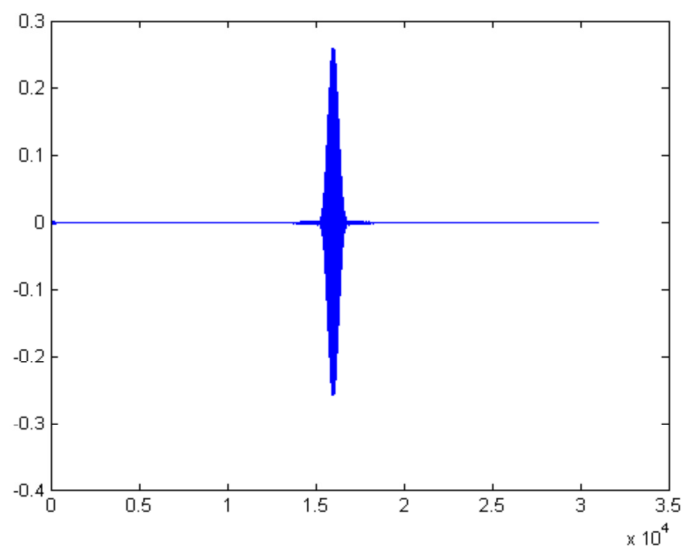
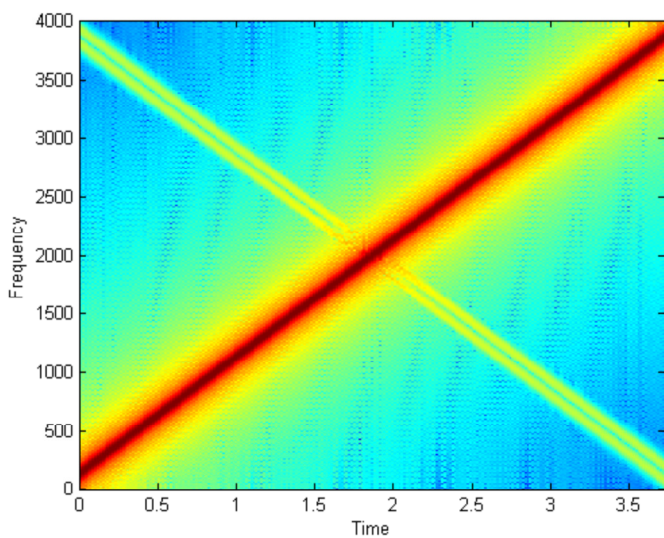
```

G1=fir1(1000,1/2,'high');
H1=G1;
H1_output=filter(H1,1,input_signal);
downsampled=H1_output(1:2:end);
upsampled(1:2:2*length(downsampled))=2*downsampled;
G1_output=filter(G1,1,upsampled);

```

Στο σημείο αυτό, εξετάζεται η έξοδος της υπο-ζώνης κανάλι-2 της τράπεζας φίλτρων, με την προσθήκη των δύο σημάτων που λήφθηκαν παραπάνω (Σχήμα 5.4 αριστερά). Τέλεια ανασυγκρότηση δεν επιτυγχάνεται, όπως φαίνεται στο προηγούμενο φασματογράφημα. Με μια πιο προσεκτική εξέταση της κυματομορφής σφάλματος (Σχήμα 5.4 δεξιά) φαίνεται ότι το μεγαλύτερο μέρος του σφάλματος βρίσκεται στο κέντρο του σήματος (αυτό δεν φαίνεται στο φασματογράφημα), για συχνότητες, δηλαδή, για τις οποίες τα φίλτρα H0 και H1 επικαλύπτονται: σημαντικό aliasing δεν μπορεί να αποφευχθεί σε κάθε ζώνη για αυτές τις συχνότητες.

```
synt_signal=G0_output+G1_output;
% Shift synt_signal to account for the filters delay.
% A symmetric FIR filter of order N (length N+1) brings
% a delay of N/2 samples.
error=synt_signal(1001:end)-input_signal(1:end-1000);
```



Σχήμα 5.4 : Αριστερά: Φασματογράφημα της εξόδου του καναλιού-2 της τράπεζας φίλτρου.  
Δεξιά: Λάθος σήμα κυματομορφής.

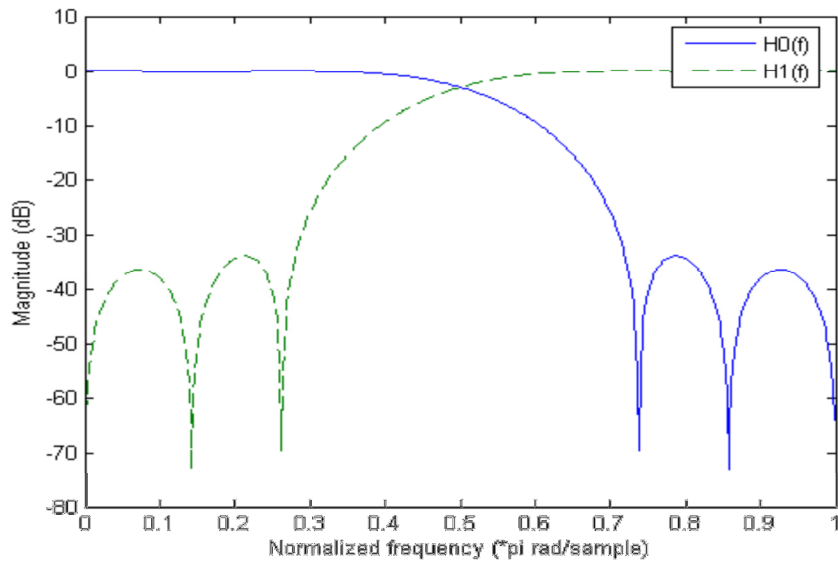
## 5.2.2 ΤΡΑΠΕΖΑ ΦΙΛΤΡΩΝ ΔΥΟ ΚΑΝΑΛΙΩΝ QMF

Η τέλεια αναδημιουργία είναι δυνατή ακόμη και με τα μη ιδανικά φίλτρα (δηλαδή ακόμη και με κάποια εξομάλυνση σε κάθε ζώνη), με την προϋπόθεση ότι η συνολική εξομάλυνση ακυρώνεται κατά την προσθήκη των low-pass και high-pass sub-band καναλιών. Αυτό, όπως θα φανεί, αποδίδεται κατά κύριο λόγο στα φίλτρα ανάλυσης και σύνθεσης. Τα φίλτρα QMF τύπου Johnson's παρέχουν μια λύση για μια σχεδόν τέλεια ανασυγκρότηση. Σε αυτό το παράδειγμα, χρησιμοποιούμε φίλτρα Johnson's "B-12" QMF (Σχήμα 5.5).

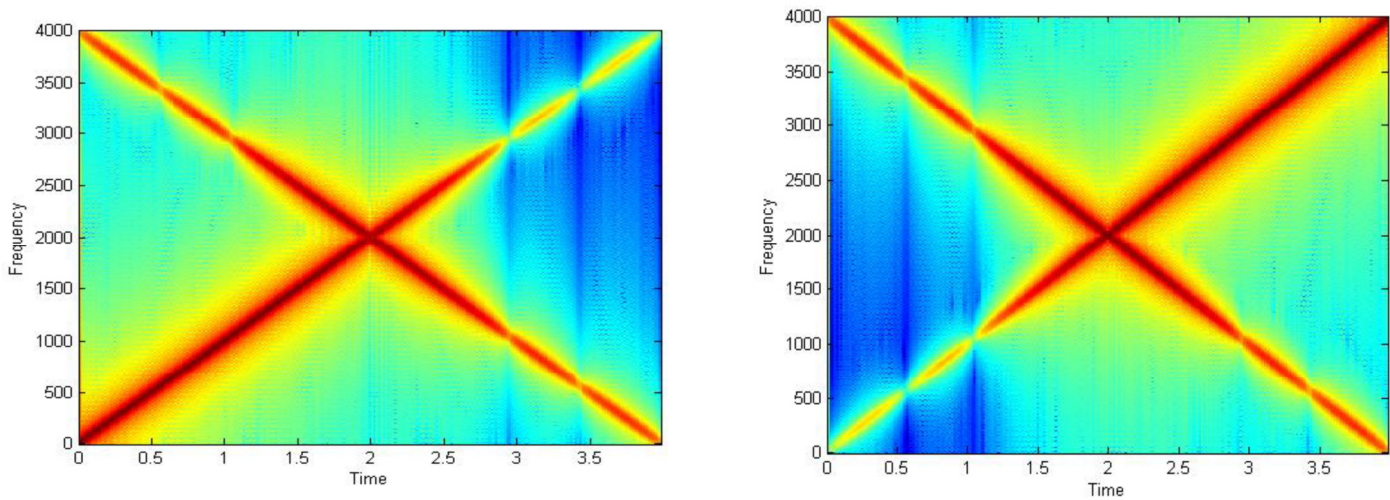
```
H0_QMF=[-0.006443977 0.02745539 -0.00758164 ...  
        -0.0913825 0.09808522 0.4807962];  
H0_QMF=[H0_QMF flip1r(H0_QMF)];  
[H0,w]=freqz(H0_QMF,1);  
  
H1_QMF=H0_QMF.*[1 -1 1 -1 1 -1 1 -1 1 -1];  
[H1,w]=freqz(H1_QMF,1);
```

Τα φίλτρα QMF τύπου Johnson's δεν είναι πολύ επιλεκτικά στη συχνότητα (επειδή η διαταγή τους είναι χαμηλή), επομένως θα επιτρέψουν σημαντική εξομάλυνση σε κάθε ζώνη. Αυτό επιβεβαιώνεται από το πέρασμα του chirp signal μέσω της αντίστοιχης τράπεζας φίλτρων (Σχήμα 5.6).

```
% LF band  
H0_output=filter(H0_QMF,1,input_signal);  
subband_0=H0_output(1:2:end);  
upsampled(1:2:2*length(subband_0))=2*subband_0;  
G0_QMF=H0_QMF;  
G0_output=filter(G0_QMF,1,upsampled);  
  
% HF band  
H1_output=filter(H1_QMF,1,input_signal);  
subband_1=H1_output(1:2:end);  
upsampled(1:2:2*length(subband_0))=2*subband_1;  
G1_QMF=-H1_QMF;  
G1_output=filter(G1_QMF,1,upsampled);  
  
synt_signal=G0_output+G1_output;
```



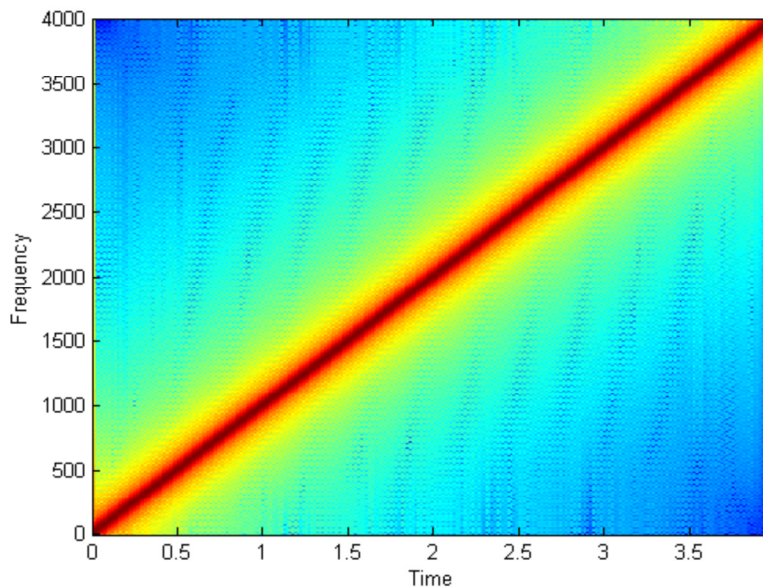
Σχήμα 5.5: Αριστερά: Φασματογράφημα της εξόδου καναλιού-2 της τράπεζας φίλτρων.  
 Δεξιά: Λάθος σήμα κυματομορφής.



Σχήμα 5.6: Αριστερά: Φασματογράφημα της εξόδου του καναλιού LF της τράπεζας φίλτρου QMF. Δεξιά: Φασματογράφημα της εξόδου του καναλιού HF. Παρατηρείται σημαντική εξομάλυνση.



Ωστόσο, τέλεια ανασυγκρότηση έχει πλέον επιτευχθεί, διότι τα φίλτρα ανάλυσης και σύνθεσης QMF είναι τέτοια ώστε η εξομάλυνση σε κάθε ζώνη συνοψίζεται σε μηδέν (Σχήμα 5.7).



Σχήμα 5.7: Φασματογράφημα της εξόδου των 2καναλιών QMF τράπεζα φίλτρων.

### 5.2.3 ΤΡΑΠΕΖΑ ΦΙΛΤΡΩΝ 32 ΚΑΝΑΛΙΩΝ PSEUDO-QMF

Μπορούμε τώρα να δημιουργήσουμε μία τράπεζα φίλτρων 32 καναλιών PQMF, όπως εφαρμόζεται στο MPEG1 Layer-I πρότυπο, και ελέγχουμε τη δυνατότητα τέλει ανασυγκρότησής του.

#### Η λειτουργία MATLAB περιλαμβάνει:

- `HN = PQMF32_prototype`

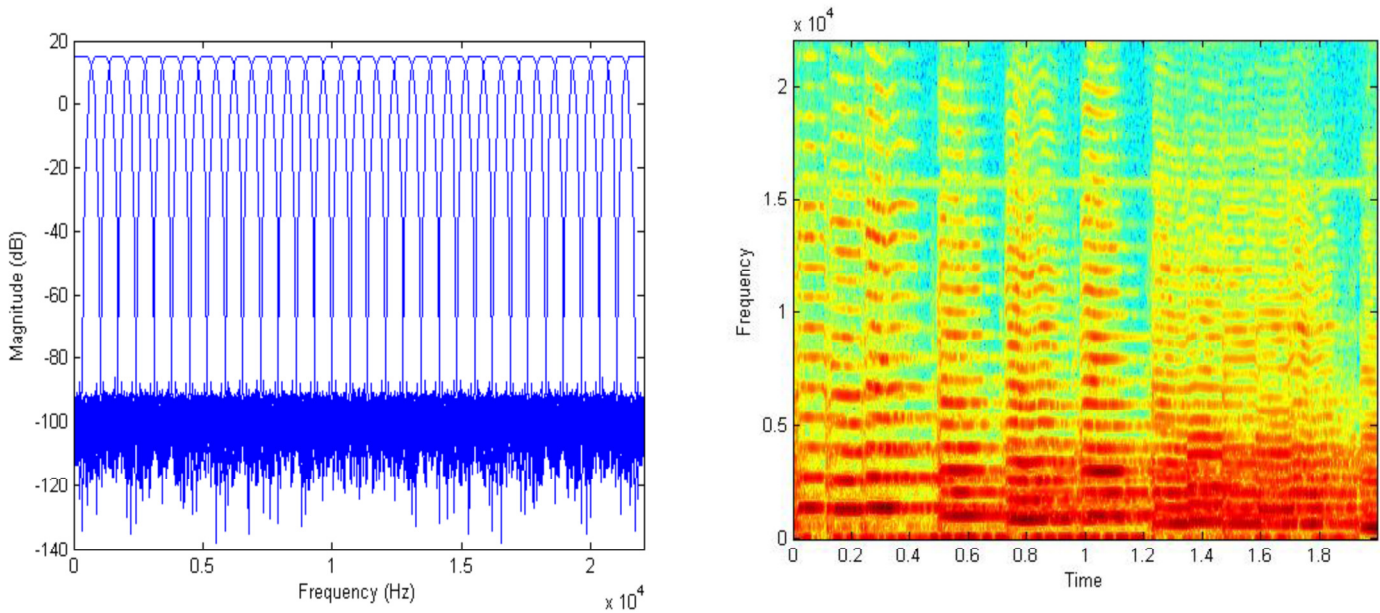
Επιστρέφει στο `hn` η κρουστική απόκριση του πρωτοτύπου χαμηλής διέλευσης συμμετρικού φίλτρου μήκους 512 για την κατασκευή τράπεζας φίλτρου 32-καναλιού PQMF. Αυτό το φίλτρο χρησιμοποιείται στον κωδικοποιητή MPEG-1 Layer-I. Η διέλευση ζώνης του είναι  $F_s / 2/32/2 = 344,5$  Hz και πληροί την προϋπόθεση PR.

Μπορούμε πρώτα να ανακατασκευάσουμε τράπεζα φίλτρου από το πρωτότυπο φίλτρο και να εξετάσουμε τις αποκρίσεις συχνότητας όλων των φίλτρων (Σχήμα 5.8 αριστερά). Σαφώς, τα φίλτρα PQMF δεν είναι ιδανικά φίλτρα διέλευσης ζώνης με εύρος ζώνης 689Hz, αλλά η ανταπόκρισή τους συμπίπτει μόνο με δύο γειτονικά τους φίλτρα (εξ ου και το όνομά τους "ψευδό" QMF φίλτρα). Το συνολικό εύρος ζώνης του κάθε φίλτρου είναι πράγματι περίπου δύο φορές 689 Hz. Παρατηρείται ότι το κέρδος φίλτρου είναι 15 dB, δηλαδή  $20 * \log_{10}(32) / 2$ . Ως αποτέλεσμα, περνώντας δύο φορές μέσα από το φίλτρο παράγεται ένα κέρδος 32, το οποίο αντισταθμίζει τον αποδεκατισμό των 32. Στις επόμενες γραμμές, είναι λοιπόν αναγκαίο να πολλαπλασιαστούν σήματα υποζώνης ανερχόμενης δειγματοληψίας από 32.

```
% Load the prototype lowpass filter
hn=PQMF32_prototype;

% Build 32 cosine modulated filters centered on normalized
% frequencies Fi=(2*i+1)/64 *1/2
PQMF32_Gfilters = zeros(32, 512);
for i = 0:31
    t2 = ((2*i+1)*pi/(2*32))*((0:511)+16);
    PQMF32_Gfilters(i+1,:) = hn.*cos(t2);
end
PQMF32_Hfilters=flip1r(PQMF32_Gfilters);

for i = 1:32
    [H,W]=freqz(PQMF32_Hfilters(i,:),1,512,44100);
    plot(W,20*log10(abs(H))); hold on
end
xlabel('Frequency (Hz)'); ylabel('Magnitude (dB)');
set(gca,'xlim',[0 22050]);
ylabel('Magnitude (dB)');
hold off;
```



Σχήμα 5.8: Αριστερά: Απόκριση συχνότητας μεγέθους των 32 φίλτρων ανάλυσης μιας τράπεζας PQMF. Δεξιά: Φασματογράφημα ενός ηχητικού σήματος που περιέχει μερικές σημειώσεις παίζοντας βιολί.

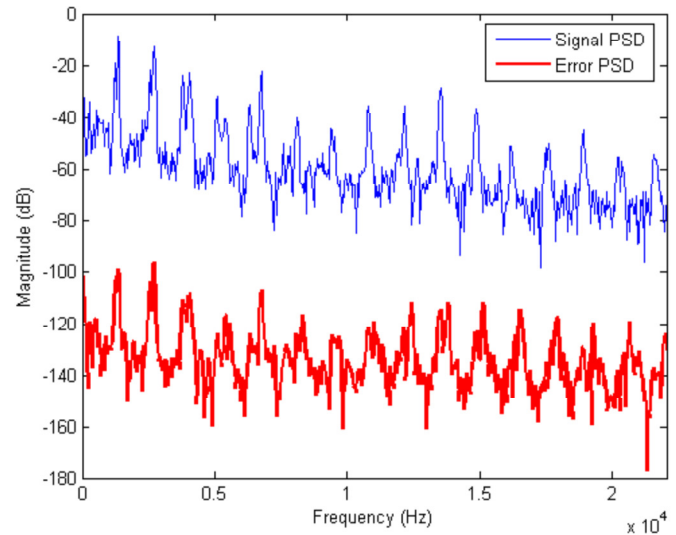
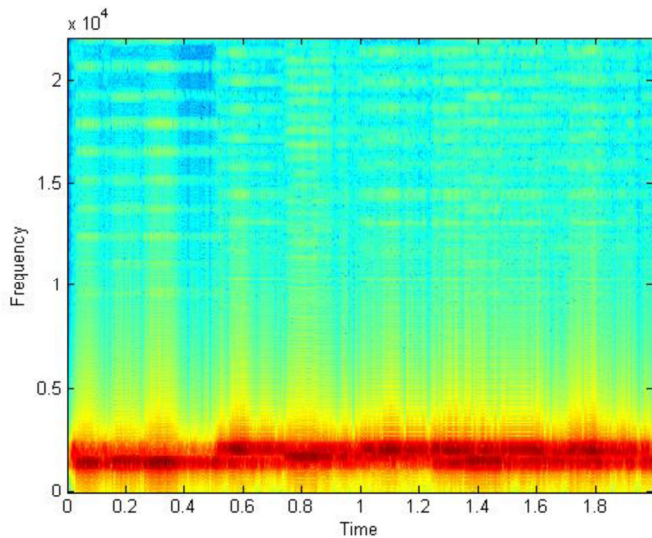
Ας ελέγξουμε τώρα την έξοδο της τράπεζας φίλτρων PQMF όταν τροφοδοτείται με 2 δευτερόλεπτα μονοφωνικού σήματος του βιολιού που επιλέγεται σε 44.100 Hz (Σχήμα 5.8 δεξιά). Όπως φάνηκε από την ακρόαση της υπό-ζώνης 3, τα απομονωμένα σήματα υπό-ζώνης είναι κατά πολύ εξομαλυμένα, διότι κάθε φίλτρο PQMF δεν είναι ιδανικό (Σχήμα 5.9 αριστερά).

```
[input_signal,Fs]=wavread('violin.wav');

output_signal=zeros(size(input_signal));
for i=1:32
    Hi_output=filter(PQMF32_Hfilters(i,:),1,input_signal);
    subband_i=Hi_output(1:32:end);
    upsampled_i(1:32:32*length(subband_i))=subband_i;

    % synthesis filters are the symmetric of the analysis
    % filters, which ensures linear phase in each sub-band
    Gi_output=filter(PQMF32_Gfilters(i,:),1,upsampled_i);
    output_signal=output_signal+Gi_output;

    if i==3
        G3_output=Gi_output;
    end;
end;
```



Σχήμα 5.9: Αριστερά: Φασματογράφημα της υπό-ζώνης 3. Δεξιά: Περιοδόγραμμα των αυθεντικών και λανθασμένων σφαλμάτων, γύρω από το δείγμα 11000.

Η τράπεζα φίλτρων QMF, ωστόσο, εξασφαλίζει την εξομάλυνση στις παρακείμενες ζώνες και ακυρώνεται όταν προστίθενται υπό-ζώνη. Η ισχύς του σφάλματος ανασυγκρότησης είναι περίπου 85 dB κάτω από αυτήν του σήματος (Σχήμα 5.9 δεξιά). Παρατηρείται ότι η έξοδος καθυστερείται από τα δείγματα 511 (δεδομένου ότι τα φίλτρα ανάλυσης και σύνθεσης έχουν μια καθυστέρηση της τάξεως του  $511/2$ ).

### Η λειτουργία MATLAB περιλαμβάνει:

- `snr = snr(signal,signal_plus_noise, max_shift, show-plot)`

Ο τύπος αυτός επιστρέφει την αναλογία σήματος-θορύβου που υπολογίζεται από τα σήματα εισόδου. Το `Max_shift` παρέχει τη μέγιστη χρονική μετατόπιση (σε δείγματα) μεταξύ του `signal` και του `signal_plus_noise`. Η πραγματική χρονική μετατόπιση, που λαμβάνεται από το μέγιστο της εγκάρσιας συσχέτισης μεταξύ των δύο σημάτων, λαμβάνεται υπόψη για την εκτίμηση του θορύβου. Αν το `showplot` έχει καθοριστεί, τότε το `signal`, `signal_plus_noise` και το λάθος απεικονίζονται και το SNR αποτυπώνεται στο διάγραμμα.

```
error=output_signal(512:end)-input_signal(1:end-511);
[signal_psd,w]=periodogram(input_signal(11001:12024),...
                            hamming(1024));
[error_psd,w]=periodogram(error(11001:12024),...
                            hamming(1024));
plot(w/pi*22050,10*log10(signal_psd));
hold on;
plot(w/pi*22050,10*log10(error_psd),'r','linewidth',2);
hold off;

snr_PQMF=snr(input_signal(1:end-511),output_signal(512:end),0)
```

$snr_{PQMF} = 84.2443$

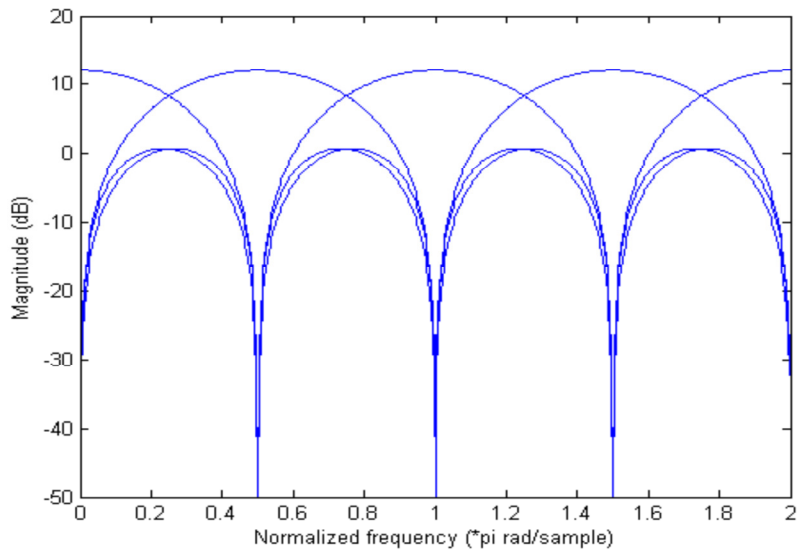
## 5.2.4 ΤΡΑΠΕΖΕΣ ΦΙΛΤΡΩΝ και ΕΠΙΚΑΛΥΠΤΟΜΕΝΟΙ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ

Θα δείξουμε τώρα πώς τα φίλτρα ανάλυσης και σύνθεσης μπορούν να εφαρμοστούν χρησιμοποιώντας μετατροπές εφαρμογών (μπλοκ). Όταν το μήκος των φίλτρων που χρησιμοποιείται στα φίλτρα υποζωνών (sub-band) M-καναλιών είναι δείγματα M, οι ενέργειες αυτές μειώνουν, σε ένα γραμμικό μετασχηματισμό διαδοχικών μη-επικαλυπτόμενων πλαισίων δείγματος M, το σήμα εισόδου. Ένα DFT 4 δειγμάτων, για παράδειγμα, μπορεί να εφαρμόσει μια τράπεζα φίλτρων 4 καναλιών όπου οι κρουστικές αποκρίσεις των φίλτρων υπό-ζώνης αποτελούν τον χρόνο αντίστροφης από τις γραμμές μεγέθους των πλεγμάτων 4x4 DFT. Εφαρμόζοντάς το στο chirp είναι απλό.

```
Fs=8000;
input_signal=chirp((1:4*Fs)/Fs,0,4,4000);
for i=1:length(input_signal)/4
    % creating a column vector with 4 samples
    input_frame=input_signal(4*(i-1)+1:4*(i-1)+4)';
    % producing one sample in each downsampled sub-band, i.e.
    % band-pass filtering and downsampling all sub-bands in
    % one operation.
    subbands=fft(input_frame);
    % producing four samples of the filter bank output, i.e.
    % upsampling, band-pass filtering all sub-bands, and
    % summing them in one operation.
    output_frame=ifft(subbands);
    % storing the output column vector in the output signal
    output_signal(4*(i-1)+1:4*(i-1)+4)=output_frame';
end;
%soundsc(output_signal,Fs);
```

Προφανώς, αυτό θα επιστρέψει το αρχικό chirp. Εφόσον τα υποκείμενα φίλτρα έχουν σύνθετους συντελεστές, κάθε σήμα υπό-ζώνης είναι σύνθετο. Επιπλέον, αυτό το είδος τράπεζας φίλτρου δεν επιλέγει τη συχνότητα, όπως φαίνεται στο Σχήμα 5.10. Η επικάλυψη της συχνότητας μεταξύ γειτονικών ζωνών είναι περίπου το μισό του κύριου λοβού στη διέλευση ζωνών (όπως στην προηγούμενη ενότητα για την PQMF), αλλά οι πλευρικοί λοβοί είναι πολύ υψηλοί. Αυτό δεν το κάνει ένα καλό υποψήφιο για κωδικοποίηση υπο-ζώνης.

```
tmp=[1 ; exp(-j*pi/2) ; exp(-j*pi) ; exp(-j*3*pi/2)];
DFT_matrix_4x4=vander(tmp);
for i=1:4
    [H,W]=freqz(flip1r(DFT_matrix_4x4(i,:)),1,'whole');
    plot(W/pi,max(20*log10(abs(H)),-50)); hold on
end
```



Σχήμα 5.10: Το μέγεθος απόκρισης συχνότητας των 4 καναλιών DFT τράπεζας φίλτρου. Παρατηρείται ότι δείχνουμε τις απαντήσεις σε πλήρη  $[0,1]$  φάσμα συχνοτήτων, καθώς αυτές οι απαντήσεις συχνοτήτων δεν αποτελούν δύο όψεις.

Γενικά, το μήκος των κρουστικών αποκρίσεων των φίλτρων ανάλυσης και σύνθεσης χρησιμοποιείται σε κωδικοποιητές υπό-ζωνών υπερβαίνοντας τον αριθμό  $M$  καναλιών. Οι ενέργειες φιλτραρίσματος, ωστόσο, μπορούν ακόμα να εφαρμοστούν όσο ο πολλαπλασιασμός των  $L$ -δειγμάτων ταιριάζει με  $L \times M$  ή  $M \times L$  πλέγματα/πίνακες. Για παράδειγμα, η τράπεζα φίλτρων 32καναλιών QMF που εισήχθη στην προηγούμενη ενότητα, στην οποία το μήκος  $N$  της κρουστικής απόκρισης του κάθε φίλτρου είναι 512 δείγματα, μπορεί να εφαρμοστεί αποτελεσματικά με τον τρόπο που φαίνεται παρακάτω (ο οποίος είναι παρόμοιος με την εφαρμογή DFT - based τράπεζας φίλτρου, με την προσθήκη επικάλυψης).

```

[input_signal,Fs]=wavread('violin.wav');
% Block-based sub-band filtering
input_frame=zeros(512,1);
output_signal=zeros(size(input_signal));
for i=1:(length(input_signal)-512+32)/32
    % Overlap input_frames (column vectors)
    input_frame=input_signal((i-1)*32+1:(i-1)*32+512);

    % Analysis filters and downsampling
    % Since PQMF H filters are the time-reversed G filters,
    % we use the G filters matrix to simulate analysis
    % filtering
    subbands_frame_i = PQMF32_Gfilters*input_frame;

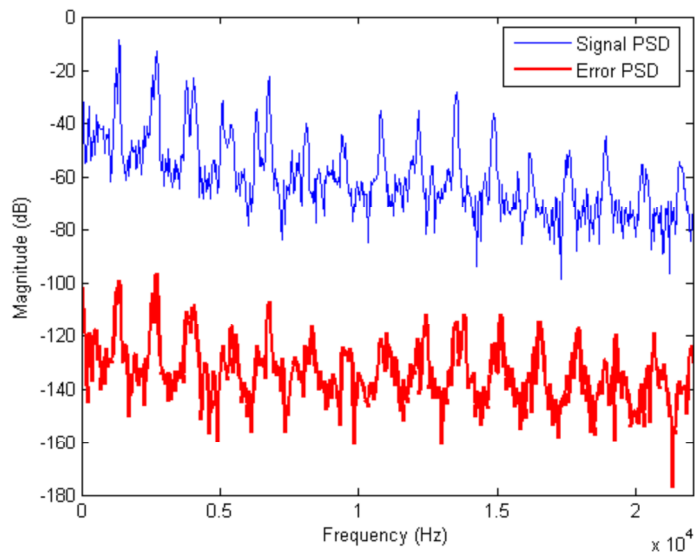
    % Synthesis filters
    output_frame = PQMF32_Gfilters'*subbands_frame_i;

    % Overlap output_frames (with delay of 511 samples)
    output_signal((i-1)*32+1:(i-1)*32+512)= ...
        output_signal((i-1)*32+1:(i-1)*32+512)+output_frame;
end

error=output_signal(512:end)-input_signal(1:end-511);

```

Προφανώς έχουμε τα ίδια αποτελέσματα όπως και πριν (Σχήμα 5.11).



Σχήμα 5.11: Περιοδόγραμμα των πρωτότυπων και σήματα σφάλματος, γύρω από το δείγμα 11000.

## 5.2.5 Perceptual Audio Coding (αντιληπτική ακουστική κωδικοποίηση)

Η διαδικασία φιλτραρίσματος υπό-ζώνης (sub-band) μετασχηματίζει τα αρχικά δείγματα σε συχνότητα  $f_s$  δειγματοληψίας σε 32 παράλληλες υπό-ζώνες που επιλέγονται σε  $F_s/32$ . Χρειάζεται ακόμα ένα βήμα κβαντοποίησης και κωδικοποίησης για να παραχθεί γενική συμπίεση σε ρυθμό bit.

Η υπό-ζώνη κάνει ομοιόμορφη κβαντοποίηση των δειγμάτων και δεν επιτρέπει πολλή διαφάνεια στους χαμηλούς ρυθμούς bit όπως φαίνεται στη δοκιμή δείγματος 4-bits ανά υπό-ζώνη. Το σήμα που προκύπτει δίνεται στο Σχήμα 5.12.

```
% Build the PQMF H and G filters
hn=PQMF32_prototype;
PQMF32_Gfilters = zeros(32, 512);
for i = 0:31
    t2 = ((2*i+1)*pi/(2*32))*((0:511)+16);
    PQMF32_Gfilters(i+1,:) = hn.*cos(t2);
end

[input_signal,Fs]=wavread('violin.wav');

% Block-based sub-band analysis filtering
input_frame=zeros(512,1);
output_signal=zeros(size(input_signal));

n_frames=(length(input_signal)-512+32)/32;
for i=1:n_frames

    % Overlap input_frames (column vectors)
    input_frame=input_signal((i-1)*32+1:(i-1)*32+512);

    % Analysis filters and downsampling
    % NB: we put sub-band signals in columns
    subbands(i,:) = (PQMF32_Gfilters*input_frame)';

    % Uniform quantization on 4 bits, using a mid-thread
    quantizer in [-1,+1]
    n_bits = 4;
    alpha = 2^(n_bits-1);
    quantized_subbands(i,:) = ...
        (floor(alpha*subbands(i,.)+0.5))/alpha; % mid-thread

    % Synthesis filters
    output_frame = PQMF32_Gfilters'*quantized_subbands(i,:)';

    % Overlap output_frames (with delay of 511 samples)
    output_signal((i-1)*32+1:(i-1)*32+512)= ...
        output_signal((i-1)*32+1:(i-1)*32+512)+output_frame;
```



Το σήμα παραγωγής είναι έντονα διαστρεβλωμένο. Το SNR πέφτει κάτω από 10.3 dB.

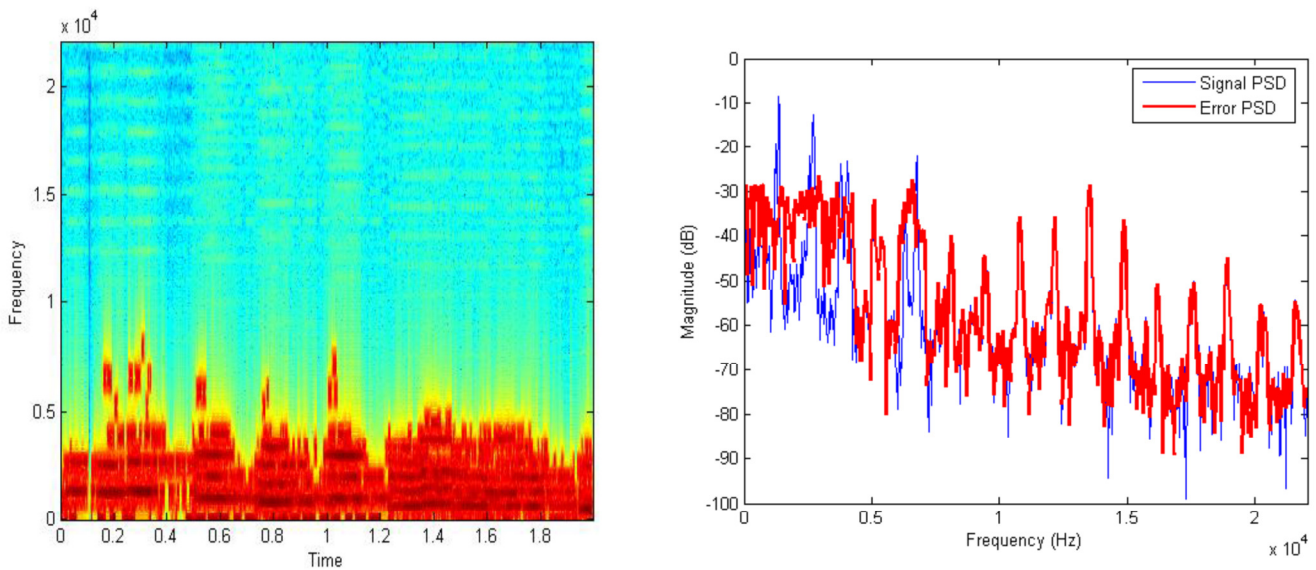
```
% NB: no delay compensation required, as the first sub-band
% samples produced by the lapped transform correspond to the
% 512th original samples.

error=output_signal-input_signal;

[signal_psd,w]=periodogram(input_signal(11001:12024),...
    hamming(1024));
[error_psd,w]=periodogram(error(11001:12024),...
    hamming(1024));
plot(w/pi*22050,10*log10(signal_psd));
hold on;
plot(w/pi*22050,10*log10(error_psd),'r','linewidth',2);
hold off;

snr_4bits=snr(input_signal(512:end-512),...
    output_signal(512:end-512),0)
```

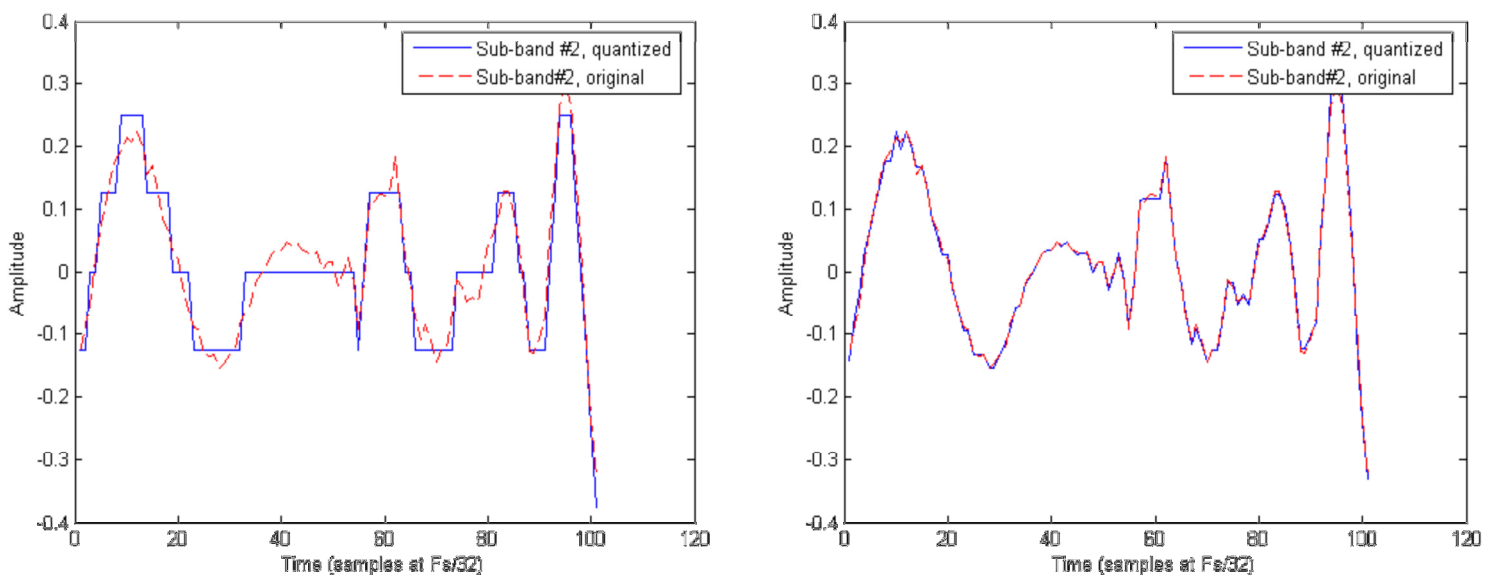
$snr_{4bits} = 10.3338$



Σχήμα 5.12: Η κωδικοποίηση υπό-ζώνης με την χρήση κβαντοποίησης 4-bit. Αριστερά: Φασματογράφημα. Δεξιά: Το περιόδγραμμα των αρχικών και των εσφαλμένων σημάτων (υπολογίζεται εδώ γύρω στα 11000 δείγματα) είναι σχεδόν πανομοιότυπα, με εξαίρεση τις χαμηλές συχνότητες.

Μπορεί κανείς να δει ότι αυτό το σταθερό  $[-1, +1]$  εύρος κβαντισμού δεν αντιπροσωπεύει επαρκώς τη διακύμανση της στάθμης του σήματος υπό-ζώνης σε όλη την υπο-ζώνη, καθώς και στο χρόνο (Σχήμα 5.13 Αριστερά).

```
plot(quantized_subbands(100:200,2)); hold on;
plot(subbands(100:200,2),'--r'); hold off;
```



Σχήμα 5.13: Απόσπασμα από το sub-band # 2 πριν και μετά την κβαντοποίηση. Αριστερά: Η χρήση σταθερών 4-bit ομοιόμορφης κβαντοποίησης. Δεξιά: Χρησιμοποιεί ομοιόμορφη κβαντοποίηση 4-bit προσαρμοσμένη στο επίπεδο.

Ένα προφανές μέσο για την ενίσχυση της ποιότητάς της είναι να εφαρμοστεί ένας συντελεστής κλίμακας σε κάθε κβαντιστή υπό-ζώνης. Όπως και στον κωδικοποιητή MPEG-1 Layer-I, υπολογίζουμε ένα νέο παράγοντα κλίμακας κάθε 12 δείγματα υπό-ζώνης (δηλαδή κάθε  $32 \times 12 = 384$  δείγματα στο αρχικό ποσοστό του δείγματος). Τα σφάλματα κβαντισμού που προκύπτουν είναι πολύ μειωμένα (Σχήμα 5.13: Δεξιά).

```

% Adaptive quantization per blocks of 12 frames
n_frames=fix(n_frames/12)*12;
for k=1:12:n_frames

    % Computing scale factors in each 12 samples sub-band
    % chunk
    [scale_factors,tmp]=max(abs(subbands(k:k+11,:)));

    % Adaptive uniform quantization on 4 bits, using a
    % mid-thread quantizer in [-Max,+Max]
    for j=1:32 % for each sub-band

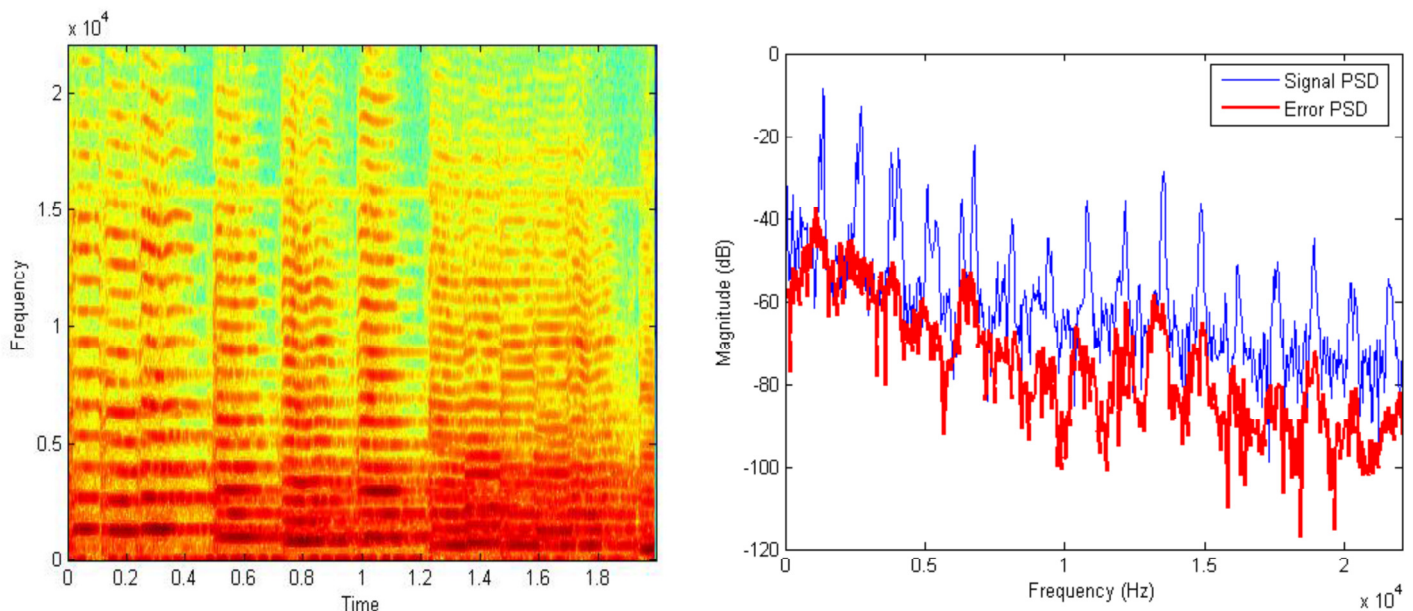
        n_bits = 4;
        alpha = 2^(n_bits-1)/scale_factors(j);
        quantized_subbands(k:k+11,j) = ... % mid-thread
            (floor(alpha*subbands(k:k+11,j)+0.5))/alpha;

    end;

end;

```

Το σήμα το οποίο προκύπτει είναι τώρα πολύ υψηλότερης ποιότητας, όπως φαίνεται στο Σχήμα 5.14. Η συνολική SNR έχει αυξηθεί σε 25 dB.



Σχήμα 5.14: Κωδικοποίηση υπό-ζώνης χρησιμοποιώντας ομοιόμορφη κβαντοποίηση 4-bit.

Αριστερά: Φασματογράφημα. Δεξιά: Περιοδόγραμμα των αρχικών και εσφαλμένων σημάτων, γύρω στα 11000 δείγματα.

```
snr_4bits_scaled=snr(input_signal(512:end-512),...  
output_signal(512:end-512),0)
```

*snr\_4bits\_scaled = 25.0629*

Ο τελευταίος καθορισμός, που είναι μακράν ο πιο αποτελεσματικός και προκύπτει από τα χρόνια έρευνας ήχου, συνίσταται στην αποδοχή περισσότερου θορύβου κβαντισμού (με την κατανομή των λιγότερων bits, και με αυτόν τον τρόπο έχουμε την αποδοχή χαμηλότερου SNR) στις ζώνες συχνότητας όπου δεν θα ακουστεί, και χρησιμοποιώντας αυτά τα επιπλέον bits για περισσότερα αντιληπτικά bands. Η απαιτούμενη αντιληπτική πληροφορία παρέχεται από ένα ψυχο-ακουστικό μοντέλο.

Για κάθε 512 frame που λαμβάνεται από το σήμα εισόδου, το MPEG-1 Layer-I ψυχο-ακουστικό μοντέλο υπολογίζει μια συνολική οριακή τιμή επικάλυψης, που λαμβάνεται από την πρώτη ανίχνευση εξέχοντα τονικά και μάσκες θορύβου, ξεχωριστά, και συνδυάζοντας τα επιμέρους όρια τους. Το μέγιστο αυτού του σφαιρικού κατώτατου ορίου και του απόλυτου ακουστικού κατώτατου ορίου λαμβάνεται έπειτα ως το τελικό κατώτατο όριο (Σχήμα 5.15 Κορυφή).

## Η λειτουργία MATLAB περιλαμβάνει:

- `function [SMR, min_threshold_subband, masking_threshold] =  
MPEG1_psycho_acoustic_model1(frame)`

Υπολογίζει το όριο κάλυψης (σε dB) που αντιστοιχεί σε ψυχο-ακουστικό μοντέλο # 1. Χρησιμοποιείται σε μορφή MPEG-1 Audio (CF ISO / CEI πρότυπο 11172-3: 1993 (F)). Η Είσοδος μήκους του πλαισίου θα πρέπει να είναι 512 δείγματα. Το `min_threshold_subband` επιστρέφει το ελάχιστο της συγκάλυψης ορίου (`masking threshold`), σε καθεμία από τις 32 υπο-ζώνες. Η `SMR` επιστρέφει 27 σήμα-προς-μάσκα (`signal-to-mask`) αναλογίες (σε dB). `SMR (28-32)` δεν χρησιμοποιούνται.

```

frame=input_signal(11001:11512);
[SMR, min_threshold, frame_psd_dBSPL]= ...
    MPEG1_psycho_acoustic_model1(frame);
% NB: the power levels returned by this function assume that a
full-scale signal (in [-1,+1]) corresponds to 96 DB SPL

f = (0:255)/512*44100;
auditory_threshold_dB = 3.64*((f/1000).^0.8) - ...
    6.5*exp(-0.6.*((f/1000)-3.3).^2) + 0.001*((f/1000).^4);
plot(f, frame_psd_dBSPL, f, min_threshold, 'r', ...
    f, auditory_threshold_dB, 'k');
hold off;
axis([0 22050 -20 100]);
legend('Signal PSD', 'Min. threshold per sub-band',...
    'Absolute threshold');
xlabel('Frequency (Hz)'); ylabel('Magnitude (dB)');

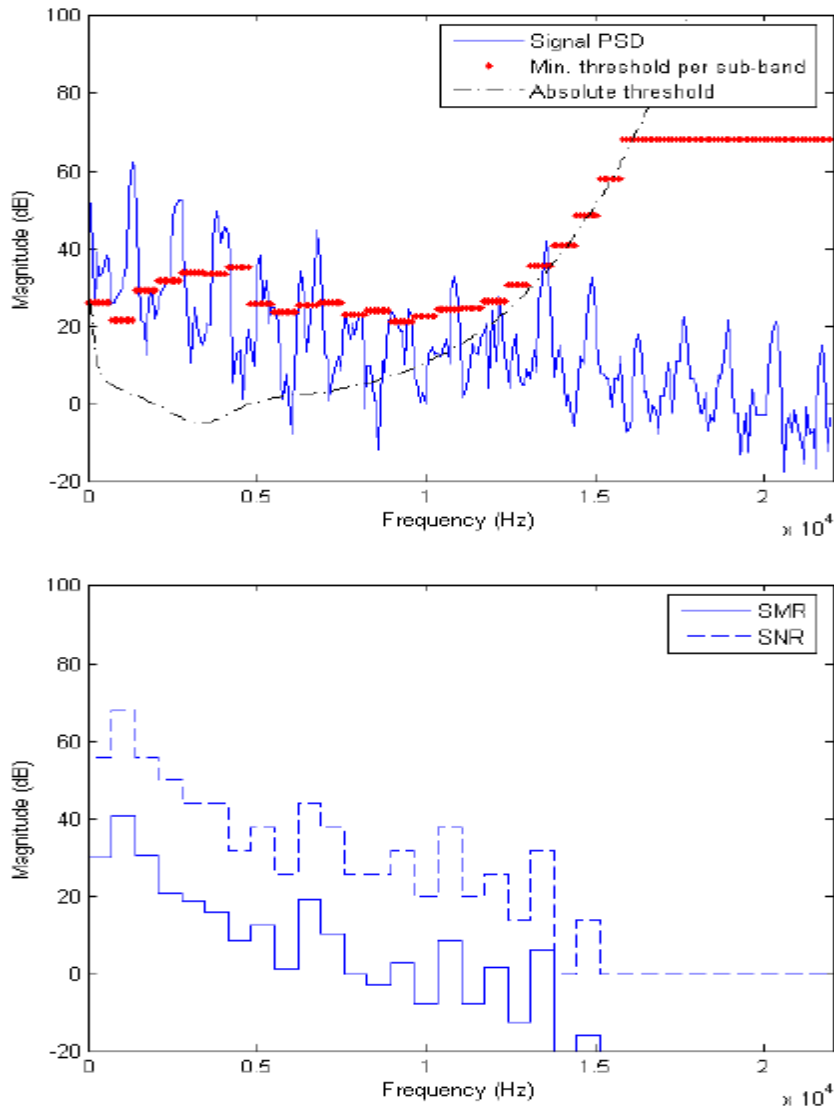
```

Οι αναλογίες Signal-to-Mask (SMR) υπολογίζονται για κάθε ζώνη με τον τρόπο αυτό: σαν αναλογία του μεγίστου σήματος PSD προς το ελάχιστο του ορίου κάλυψης για κάθε υπό-ζώνη. Η κατανομή των bit εκτελείται από έναν επαναληπτικό αλγόριθμο που δίνει προτεραιότητα στην υπό-ζώνη με υψηλότερο SMR. Η κατάληξη SNR σε κάθε υπό-ζώνη πρέπει να είναι μεγαλύτερη ή όση με το SMR, ώστε να προωθηθεί το επίπεδο θορύβου κάτω από το κατώτατο όριο (Σχήμα 5.15).

## Η λειτουργία MATLAB περιλαμβάνει:

- `function [N_bits,SNR]= MPEG1_bit_allocation(SMR,bit_rate)`

Εφαρμόζει έναν απλουστευμένο αλγόριθμο κατανομής bit. Το SMR είναι η αναλογία signal-to-mask σε κάθε υπό-ζώνη, όπως καθορίζεται το ψυχοακουστικό πρότυπο MPEG-1. Ο ρυθμός bit είναι στα kbits/s. Το N\_bits είναι ο αριθμός των bits σε κάθε υπό-ζώνη. Το SNR είναι το μέγιστο SNR σε κάθε υπο-ζώνη μετά από την κβαντοποίηση, δηλαδή το SNR που υποθέτει κάθε υπό-ζώνη περιέχει μια μεγάλη έκταση. Το N\_bits και το SNR θέτονται στο μηδέν για sub-band 28 έως 32.



Σχήμα 5.15: Το ψυχοακουστικό πρότυπο MPEG-1 #1. Κορυφή: το κατώτατο όριο ανά υποζώνη. Κάτω: οι αναλογίες signal-to-mask (SMR) και το μέγιστο (θεωρητικό) SNR που έλαβαν μετά από την κατανομή των bit. Το SNR είναι συστηματικά πάνω από το SMR, ώστε να ακούγεται ο θόρυβος κβαντισμού.

Παρατηρήστε ότι αυτό πραγματοποιείται μόνο από το γεγονός ότι έχουμε διαθέσιμη αρκετά bits για να κινητοποιήσουμε τα frames. Όταν το ποσοστό bit rate μειώνεται, το SNR μπορεί να είναι χαμηλότερο από το SMR.

Ας εξετάσουμε τελικά αυτό στο πλήρες σήμα, που προσθέτει την αντιληπτική κατανομή bit στην προσαρμοστική ομοιόμορφη κβαντοποίηση. Παρατηρήστε ότι έχουμε απλοποιήσει το σύστημα κβαντοποίησης εδώ, σε σύγκριση με το MPEG-1, με την εξέταση κβαντοποίησης με οποιοδήποτε αριθμό των bits στο εύρος [0 ... 16].

```
% Adaptive quantization per blocks of 12 frames
n_frames=fix(n_frames/12)*12;
for k=1:12:n_frames

    % Computing scale factors in each 12 samples sub-band
    % chunk
    [scale_factors,tmp]=max(abs(subbands(k:k+11,:)));

    % Computing SMRs 13
    frame=input_signal(176+(k-1)*32:176+(k-1)*32+511);
    SMR = MPEG1_psycho_acoustic_model1(frame);

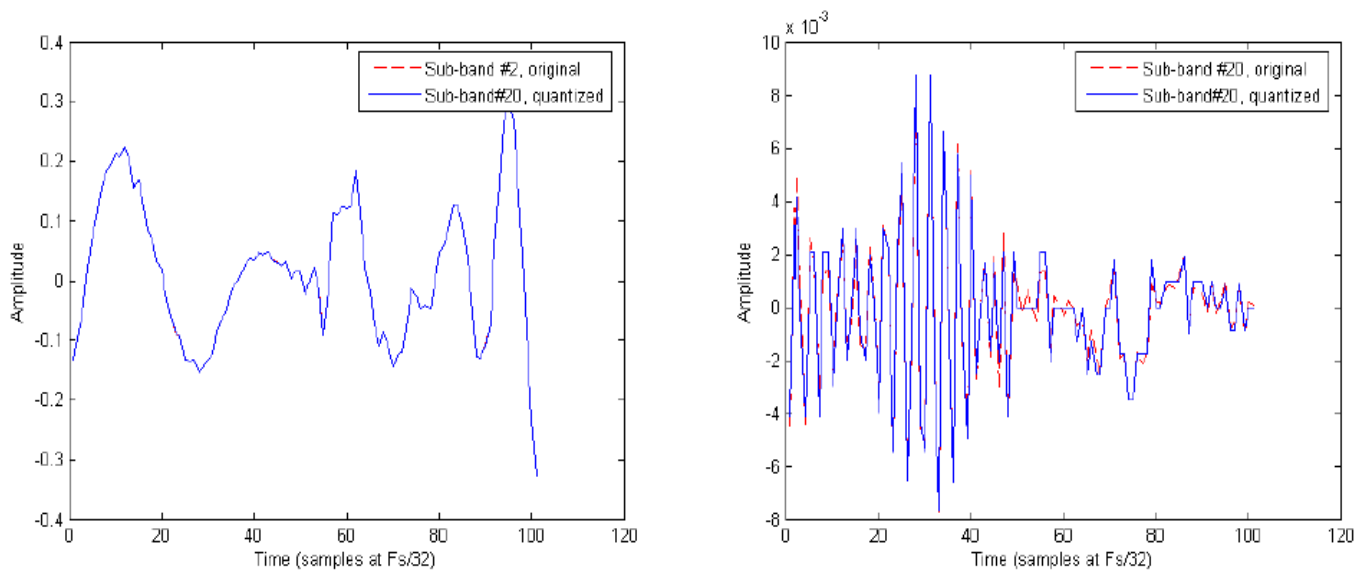
    N_bits = MPEG1_bit_allocation(SMR, 192000);

    % Adaptive perceptual uniform quantization, using a mid-
    % thread quantizer in [-Max,+Max]
    for j=1:32 % for each sub-band

        if N_bits(j)~=0
            codes= uencode(subbands(k:k+11,j),N_bits(j),...
                scale_factors(j),'signed');
            quantized_subbands(k:k+11,j) = ...
                udecode(codes,N_bits(j),scale_factors(j));
        else
            quantized_subbands(k:k+11,j) = 0;
        end;
    end;
end;
```

Ο θόρυβος κβαντοποίησης είναι τώρα πολύ μικρός σε μερικές προεξέχουσες υπό-ζώνες, όπως στην υπό-ζώνη #2 (Σχήμα 5.16 Αριστερά). Επίσης, είναι σημαντικότερος ο θόρυβος κβαντοποίησης σε άλλες υπό-ζώνες, όπως στην υπό-ζώνη #20 (Σχήμα 5.16 Δεξιά).

\*\*Το πλαίσιο εισόδου για το ψυχο-ακουστικό μοντέλο καθυστερείται από 176 δείγματα, δεδομένου ότι θα πρέπει να μπει στη μέση του τοπικού block των 12 δειγμάτων υπό-ζώνης, και το πρώτο δείγμα υπό-ζώνης αντιστοιχεί στο αρχικό δείγμα 256 (στην πραγματικότητα, το δείγμα 512, αλλά και το φίλτρο ανάλυσης εισάγει μια καθυστέρηση 256 δείγματα). Έτσι, το κέντρο του πρώτου πλαισίου θα πρέπει να είναι σε αρχικό δείγμα  $256 + 11 * 32/2 = 432$ , και το αρχικό αυτού του πλαισίου πέφτει στο δείγμα  $11 * 32/2 = 176$ .



Σχήμα 5.16: Αποσπάσματα από τις sub-band #2 και #20.

Το γενικό SNR έχει αυξηθεί σε 36,6 dB. Το αντιληπτό SNR είναι πολύ υψηλότερο, δεδομένου ότι το μεγαλύτερο μέρος του θορύβου δεν μπορεί να ακουστεί. Το σήμα που προκύπτει είναι υψηλής ποιότητας ήχος (Σχήμα 5.17).

```
snr_scaled_perceptual=snr(input_signal(512:end-512),...
    output_signal(512:end-512),0)
```

*snr\_scaled\_perceptual = 36.6457*

Η τιμή προς πληρωμή είναι ότι οι παράγοντες κλίμακας και ο αριθμός των bits ανά υπό-ζώνη πρέπει τώρα να αποθηκευτούν, σε κάθε 12 frames (δηλαδή κάθε υπό-ζώνη 12\*32). Στο MPEG-1 οι παράγοντες κλίμακας εκφράζονται στο dB και κβαντοποιούνται σε κάθε 6 bit. Αυτό προέρχεται από το γεγονός ότι το αντί αντιλαμβάνεται την ένταση του ήχου ως ενέργεια και έχει περίπου 96 dB της δυναμικής ακρόασης με ένα κατώτατο όριο ευαισθησίας περίπου 1 dB. Με 6 dB το βήμα κβαντοποίησης είναι 96/64 dB και το λάθος βρίσκεται στο [-96/64/2, +96/64/2], το οποίο είναι κάτω από το κατώτατο όριο 1 dB. Υποθέτοντας ότι 4 bits απαιτούνται για την κωδικοποίηση αριθμού bits που χρησιμοποιούνται σε καθεμία από τις πρώτες 27 υπό-ζώνες (sub-band 28 έως 32 αγνοούν το MPEG 1), αυτό οδηγεί σε 27 \* 10 =



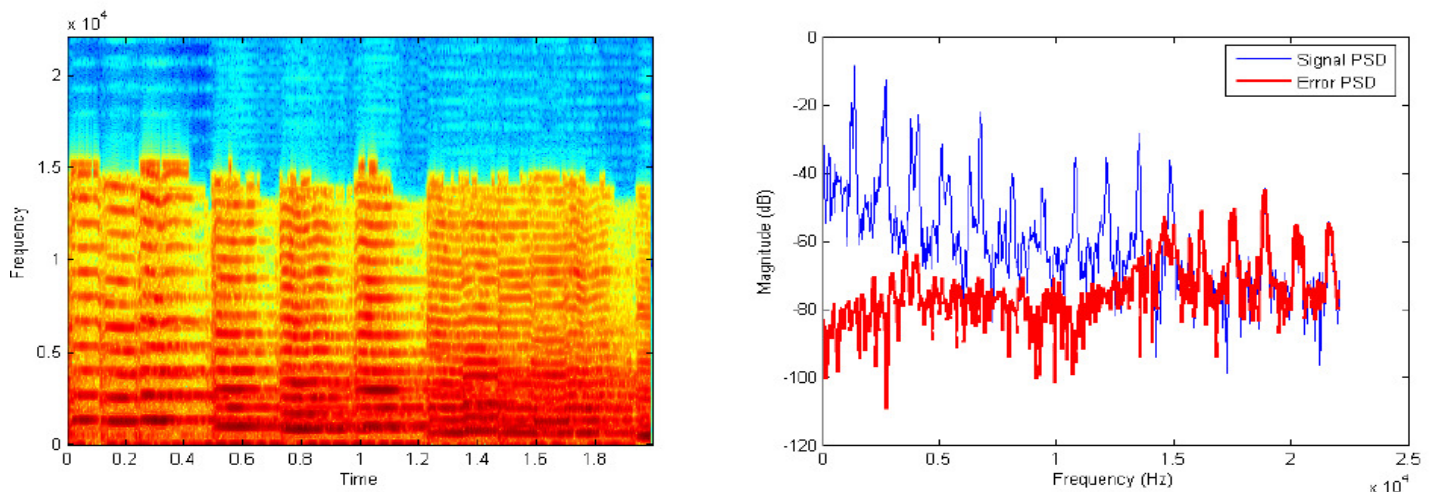
270 bits ανά 12 frames. Είναι εύκολο να ληφθεί ο αριθμός των bits που χρησιμοποιούνται για το τελευταίο μπλοκ των 12 frames στην ακουστική μας δοκιμή, καθώς και το σχετικό bit-rate.

Στην πράξη, το MPEG-1 δεν επιτρέπει καν όλες τις ακέραιες τιμές στο πεδίο [1,16] για τον αριθμό των bits σε κάθε ζώνη, η οποία καθιστά πιθανό να το κβαντοποιήσει λιγότερο έπειτα σε 4 bits. Ο συνολικός αριθμός των bits που χρησιμοποιούνται για την κατανομή δυαδικών ψηφίων μπορεί στη συνέχεια να μειωθεί σε 88 αντί του  $4 * 27 = 108$ .

```
bits_per_block=sum(N_bits)*12+270  
bit_rate=bits_per_block*44100/384
```

*bits\_per\_block = 1674*

*bit\_rate = 1.9225e+005*

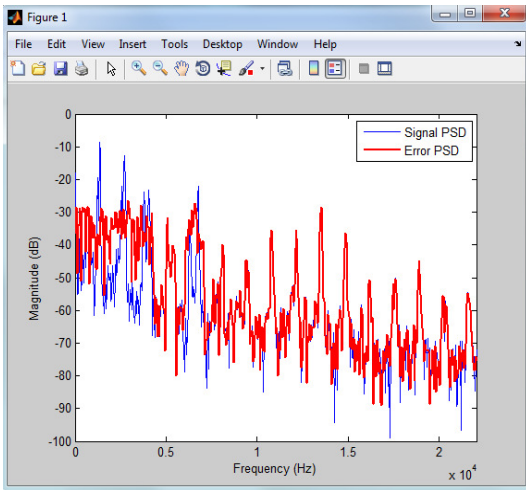


Σχήμα 5.17: Κωδικοποίηση υπό-ζωνών που χρησιμοποιούν την προσαρμοστική ομοιόμορφη κβαντοποίηση. Αριστερά: Φασματογράφημα. Δεξιά: Περιοδόγραμμα των αρχικών και λάθος σημάτων, γύρω στα 11000 δείγματα.

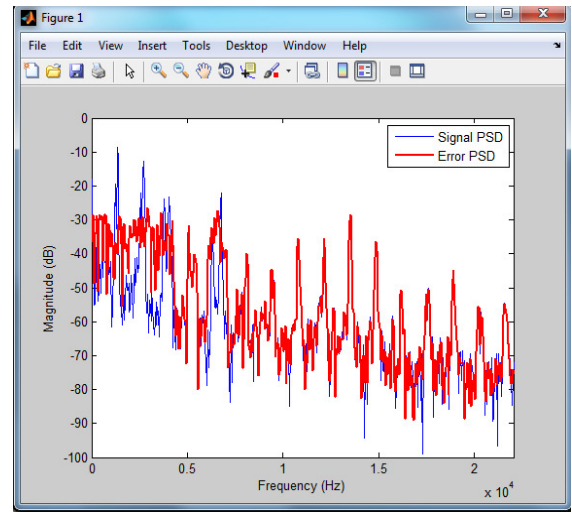
### 5.3 ΥΛΟΠΟΙΗΣΗ ΚΑΙ ΑΠΕΙΚΟΝΙΣΗ ΔΙΑΓΡΑΜΑΤΩΝ ΣΕ ΔΙΑΦΟΡΕΣ ΤΙΜΕΣ ΒΙΤΡΑΤΕ

Σε αυτό το κεφάλαιο της πτυχιακής μας πραγματοποιείται το πρακτικό κομμάτι το οποίο βασίζεται στη μέτρηση της ακουστικής ποιότητας NMR. Μετά την υλοποίηση του κώδικα τον τρέξαμε με διαφορές τιμές bitrate όπως 32,64,128,192. Με αποτέλεσμα να απεικονίζονται τα αντίστοιχα διαγράμματα στον αρχικό ήχο, στην δειγματοληψία και στην κβαντοποίηση και τέλος στο συμπιεσμένο ήχο. Στην συνέχεια, στο επόμενο κεφάλαιο γίνεται αναλυτική απεικόνιση των διαγραμμάτων.

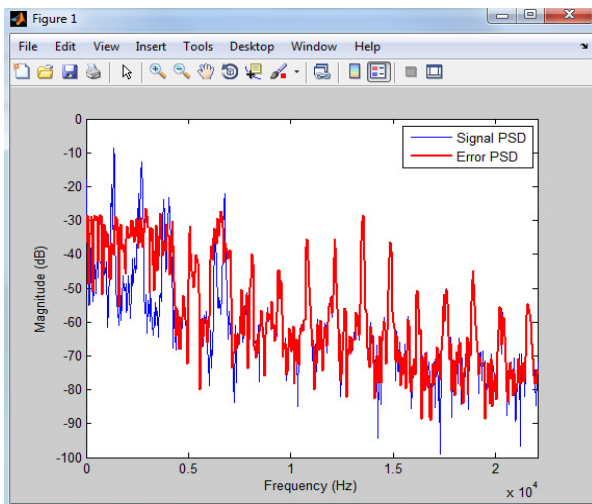
### 5.3.1 ΠΕΡΙΟΔΟΓΡΑΜΜΑΤΑ ΑΡΧΙΚΟΥ ΗΧΟΥ ΣΕ ΔΙΑΦΟΡΕΣ ΤΙΜΕΣ BITRATE



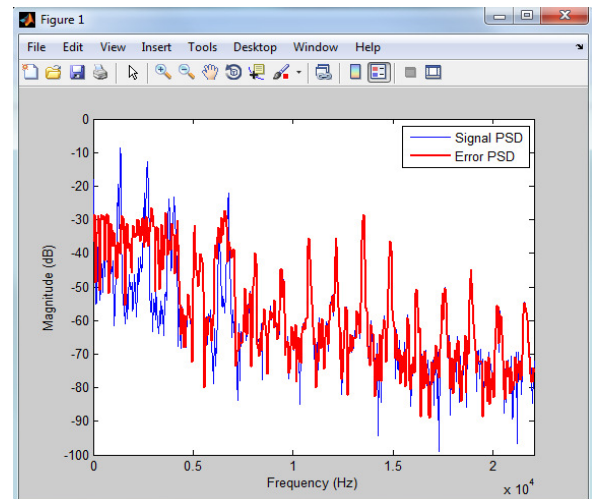
32



64



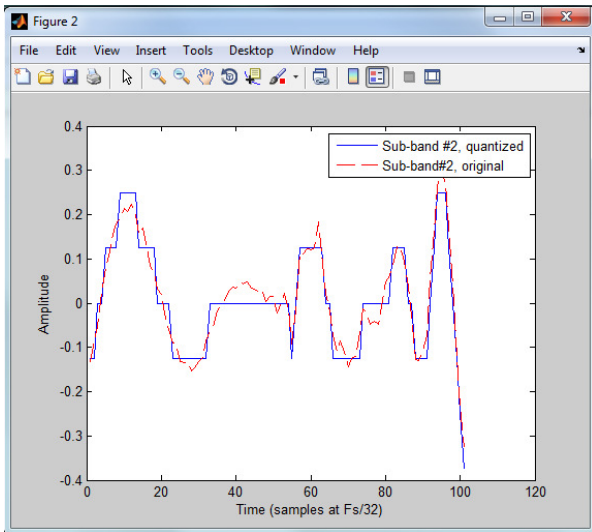
128



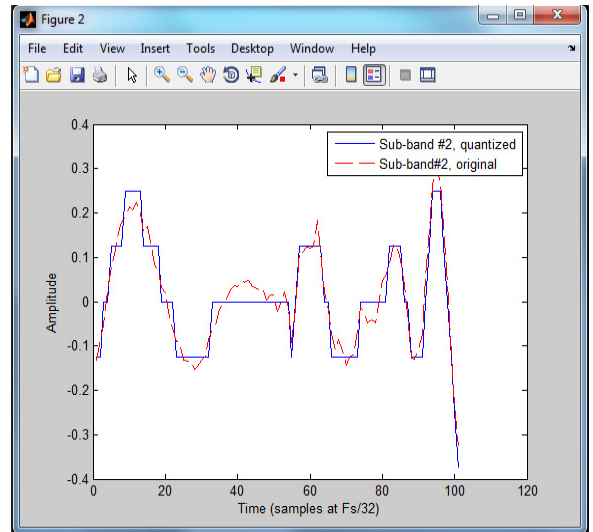
192

Στα συγκεκριμένα περιοδογράμματα απεικονίζεται ο αρχικός ήχος. Το μπλε χρώμα αντιστοιχεί στο σήμα ενώ το κόκκινο χρώμα στο σφάλμα. Τέλος, στόχος μας είναι το σήμα να είναι πιο ισχυρό από το σφάλμα.

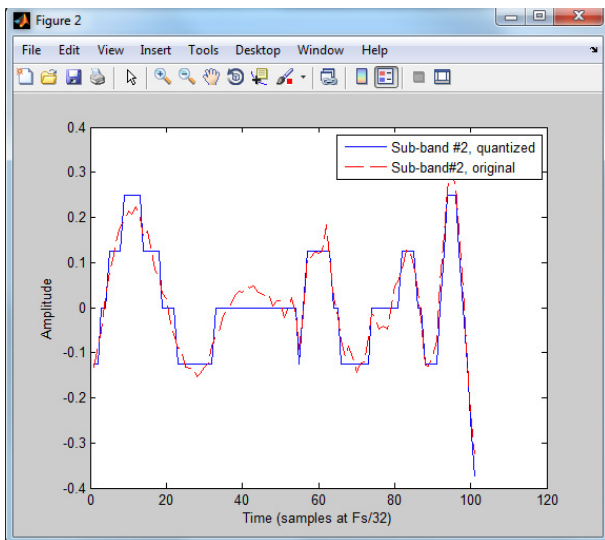
### 5.3.2 ΑΠΟΣΠΑΣΜΑΤΑ ΑΠΟ ΤΗΝ SUB-BAND #2



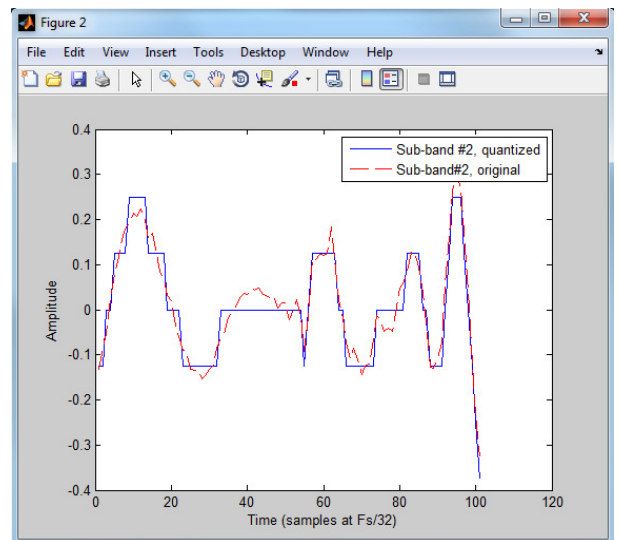
32



64



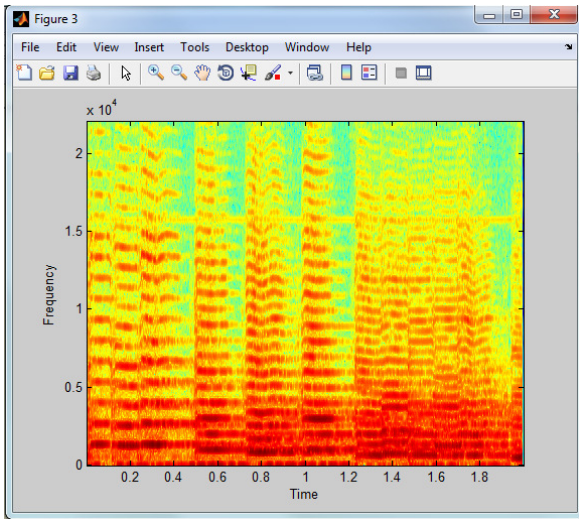
128



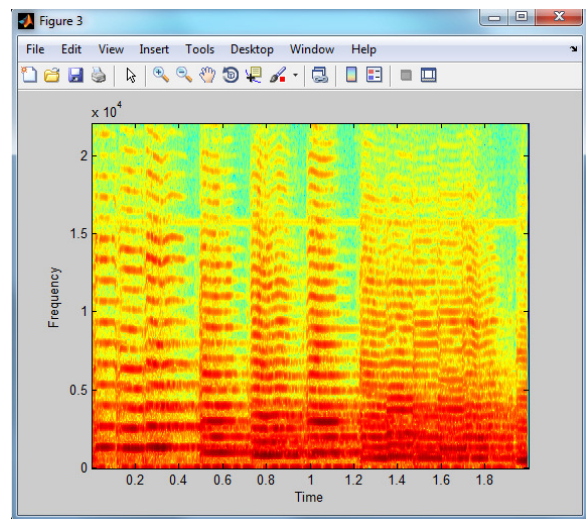
192

Στις συγκεκριμένες γραφικές παραστάσεις απεικονίζεται το μπλε και το κόκκινο σήμα. Το κόκκινο σήμα προκύπτει από την δειγματοληψία ενώ το μπλε σήμα προκύπτει όταν κάνουμε την κβάντιση για αυτό και είναι σε συγκεκριμένες στάθμες.

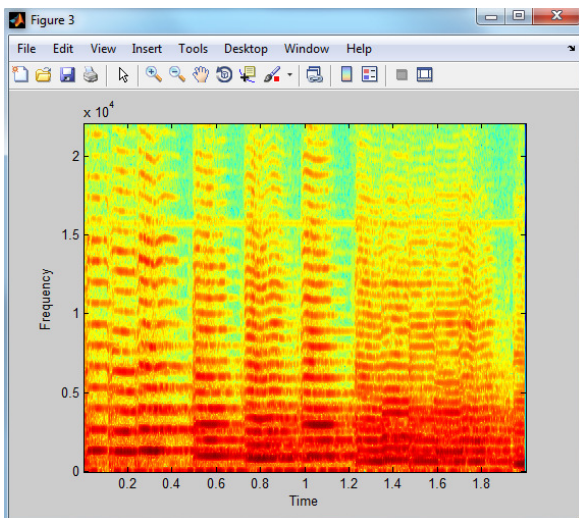
### 5.3.3 ΦΑΣΜΑΤΟΓΡΑΦΗΜΑΤΑ ΑΡΧΙΚΟΥ ΗΧΟΥ



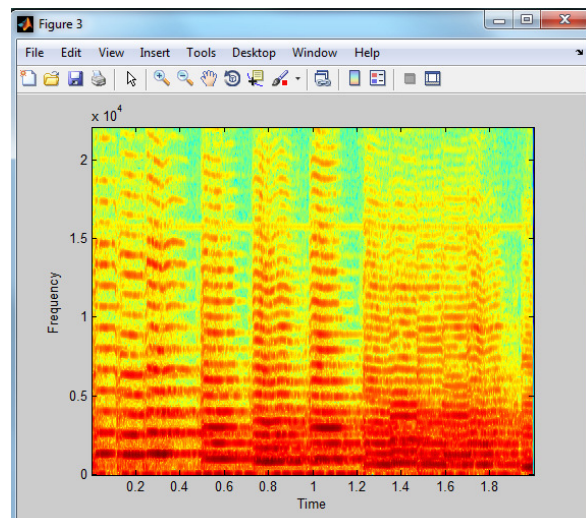
32



64



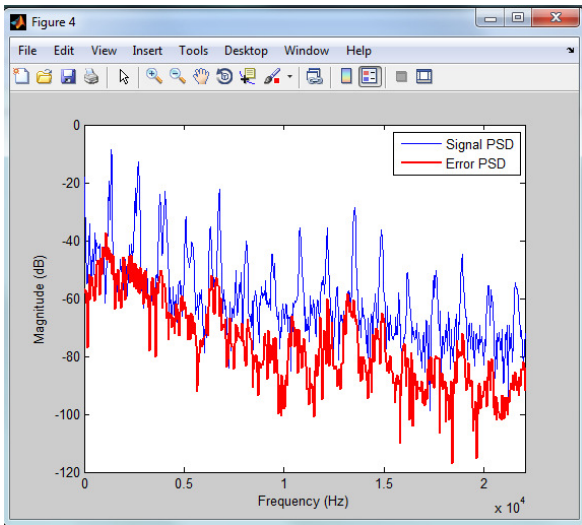
128



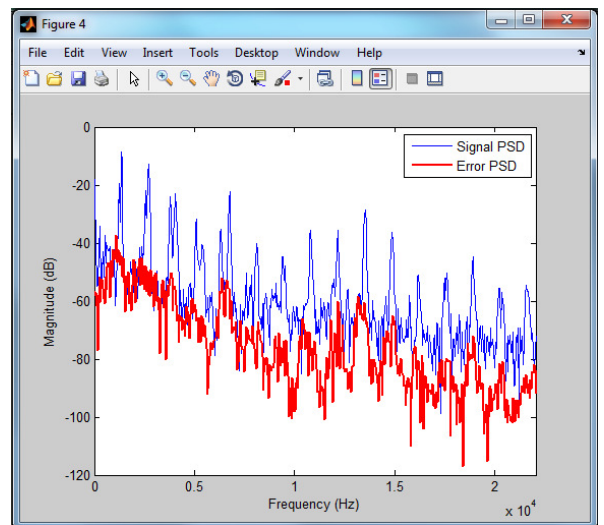
192

Το φασματογράφημα απεικονίζει ένα διάγραμμα συχνότητας σε συνάρτηση με το χρόνο και υπάρχει χρωματική κωδικοποίηση της έντασης.

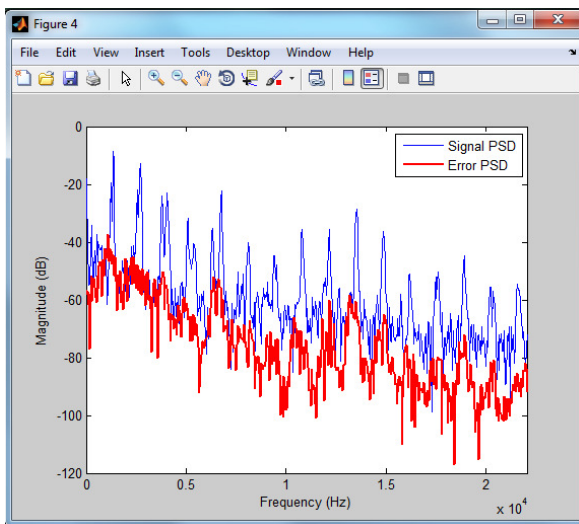
### 5.3.4 ΠΕΡΙΟΔΟΓΡΑΜΜΑΤΑ ΣΤΟ ΣΤΑΔΙΟ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ ΚΑΙ ΚΒΑΝΤΟΠΟΙΗΣΗΣ



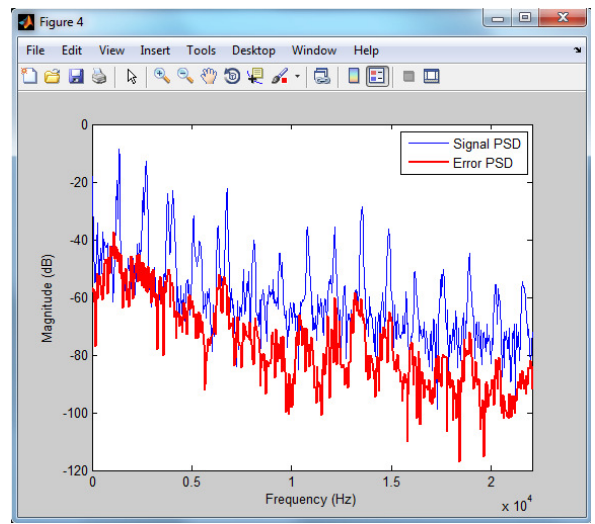
32



64



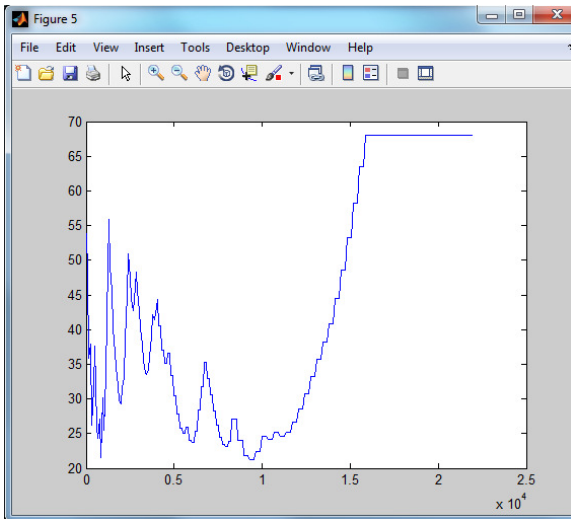
128



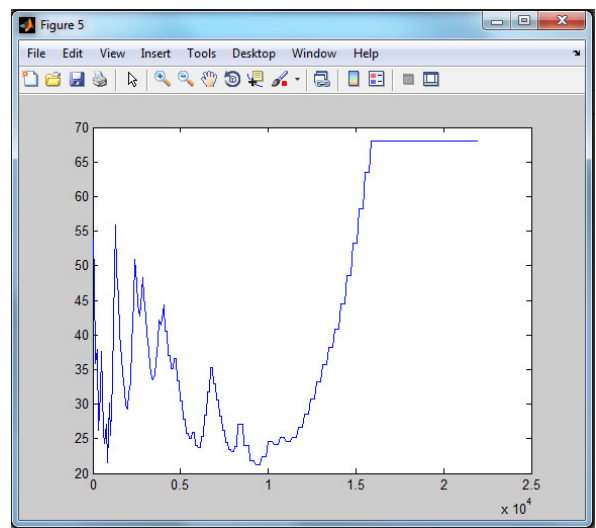
192

Τα περιοδογράμματα αυτά παρουσιάζουν την δειγματοληψία και την κβαντοποίηση τα οποία έχουν προκύψει μετά από τον αρχικό ήχο (figure 1). Συμπεραίνουμε ότι έχει μειωθεί το σφάλμα και έχει αυξηθεί ο ήχος.

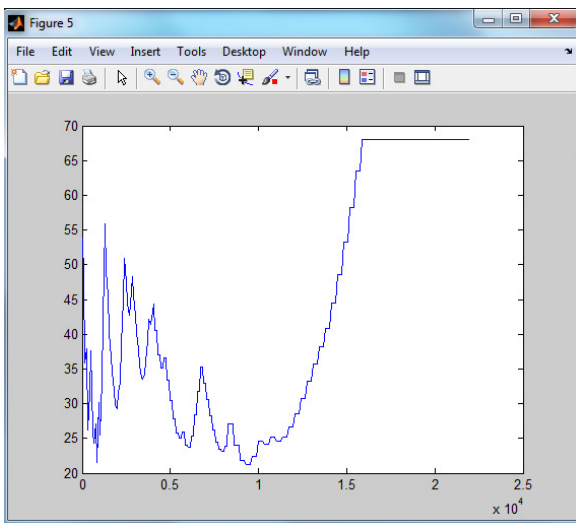
### 5.3.5 ΔΙΑΓΡΑΜΜΑ ΓΙΑ ΤΟ ΕΛΑΧΙΣΤΟ ΚΑΤΩΦΛΙ ΑΚΡΟΑΣΗΣ



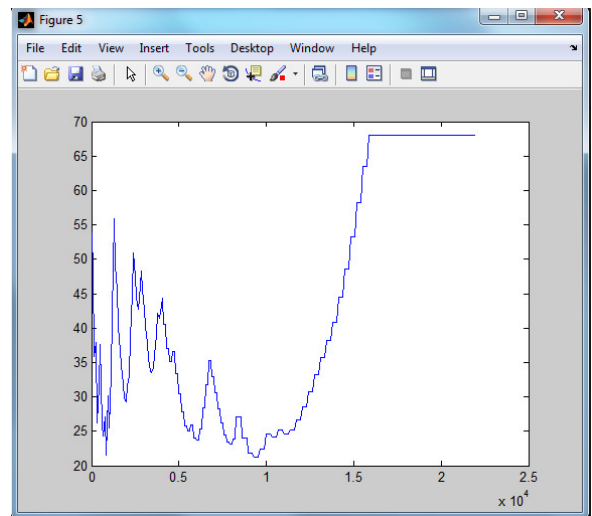
32



64



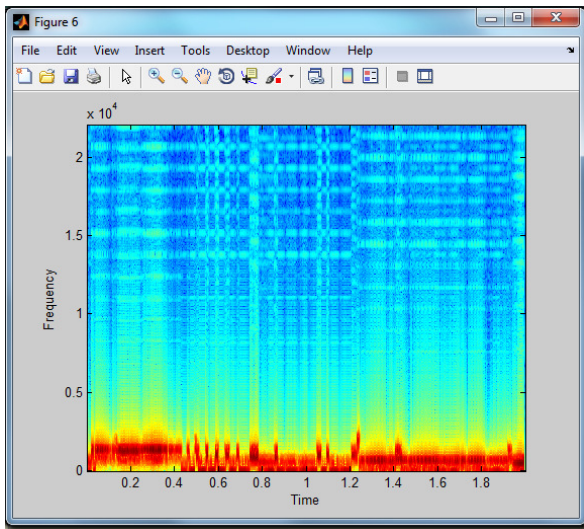
128



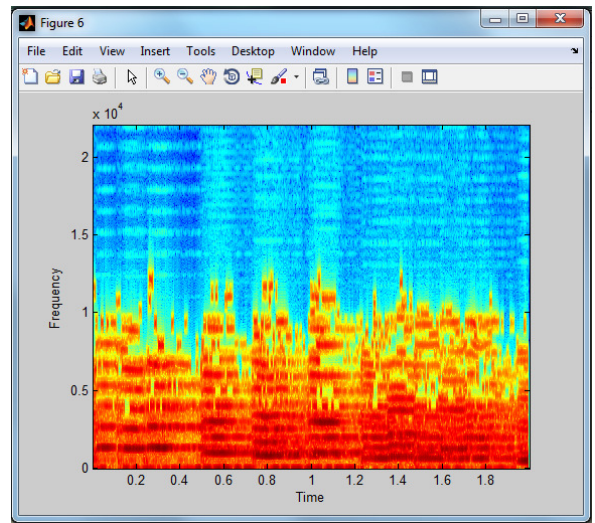
192

Στις γραφικές παραστάσεις αυτές απεικονίζεται το ελάχιστο κατώφλι ακουστότητας. Οι ήχοι που τοποθετούνται κάτω από αυτό το κατώφλι δεν είναι απαραίτητο να κωδικοποιηθούν επειδή δεν μπορούν να γίνουν αντιληπτοί.

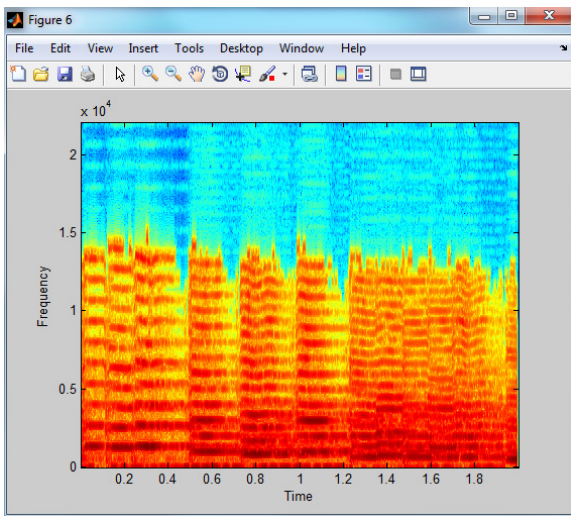
### 5.3.6 ΦΑΣΜΑΤΟΓΡΑΦΗΜΜΑΤΑ ΣΥΜΠΙΕΣΜΕΝΟΥ ΗΧΟΥ



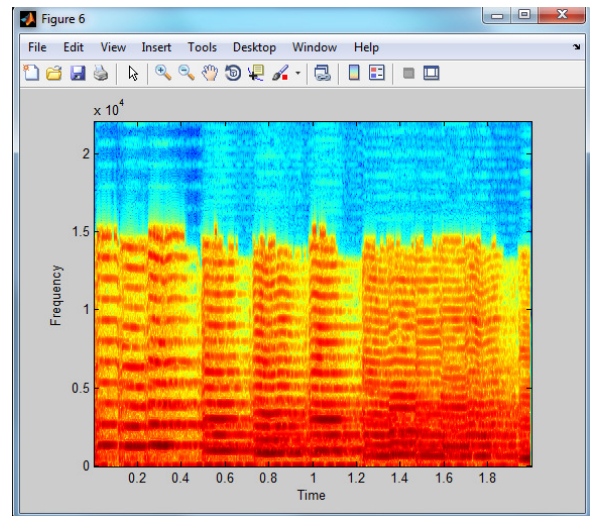
32



64



128

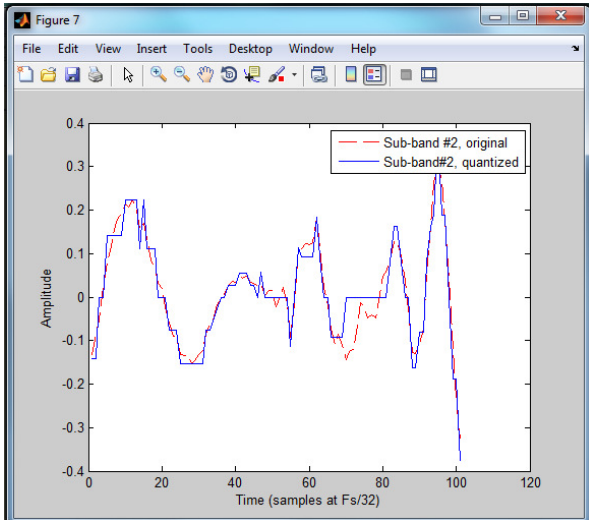


192

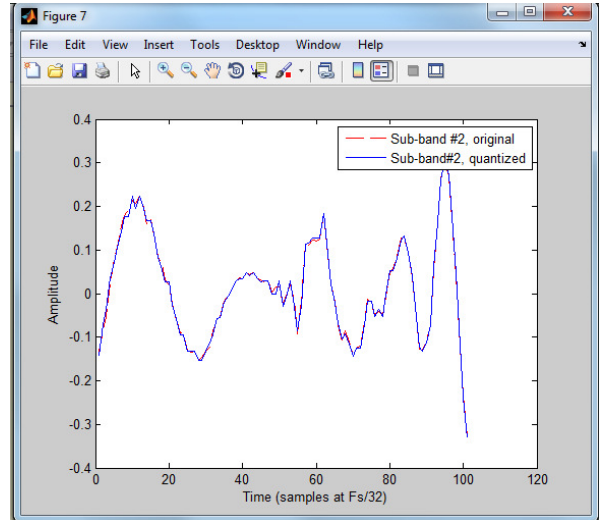
Στο figure 6 γίνεται αναπαράσταση του συμπιεσμένου ήχου όπου προέρχεται από την γραφική παράσταση figure 3(αρχικός ήχος). Όσο πιο ερυθρό (κόκκινο) είναι το χρώμα τόσο περισσότερη ένταση έχουμε. Στόχος μας είναι να συμπιέσουμε το ερυθρό χρώμα όσο το δυνατόν πιο χαμηλά.



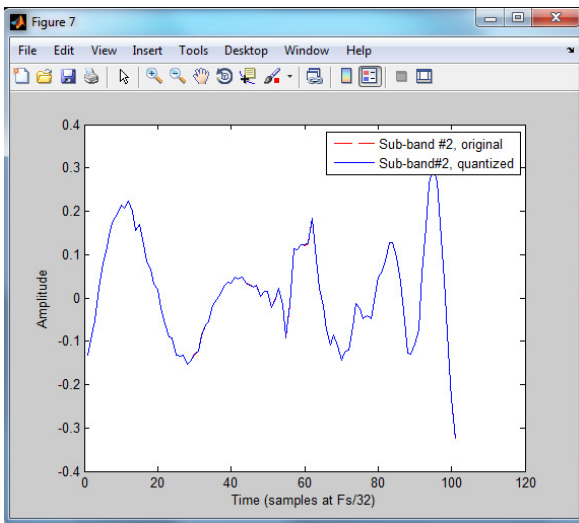
### 5.3.7 ΑΠΟΣΠΑΣΜΑΤΑ ΣΥΜΠΙΕΣΜΕΝΟΥ ΗΧΟΥ ΑΠΟ ΤΗΝ SUB-BAND #2



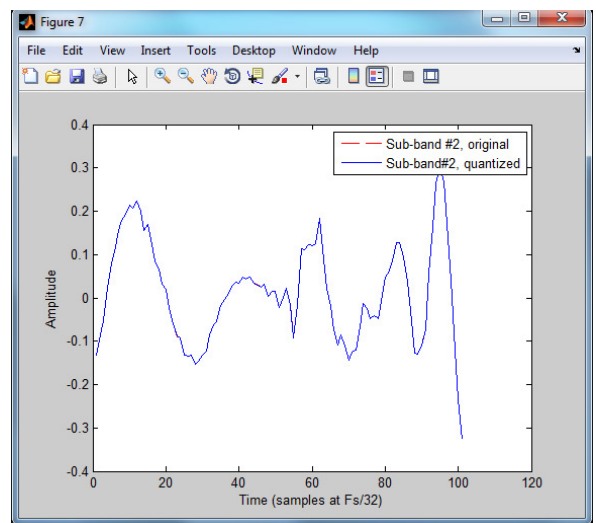
32



64



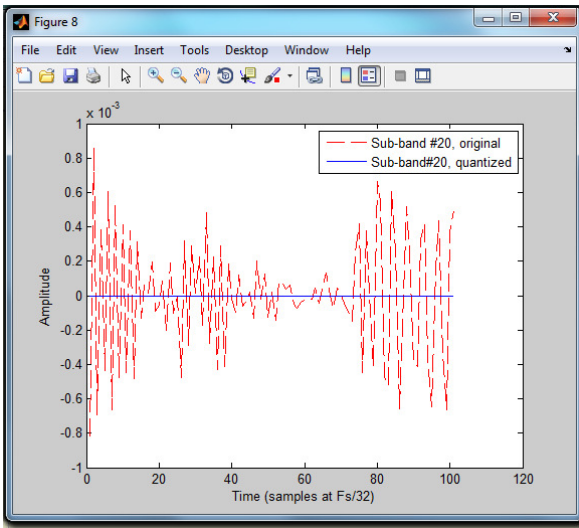
128



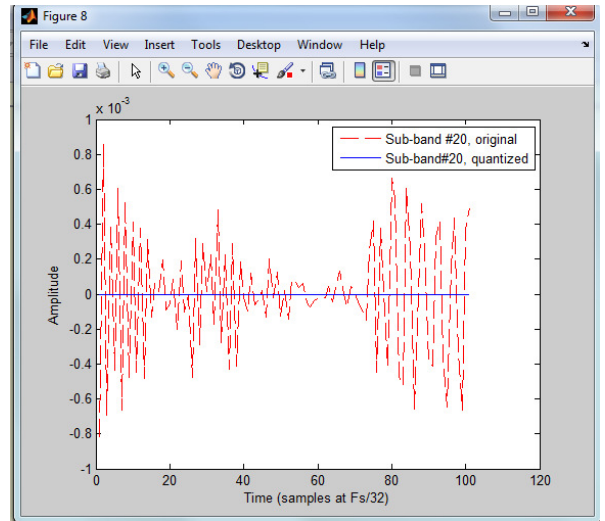
192

Το figure 7 είναι το διάγραμμα του συμπιεσμένου ήχου που προκύπτει από το figure 2 (το σήμα που προκύπτει μετά από την δειγματοληψία και την κβάντιση). Το μπλε σήμα υπερκαλύπτει το κόκκινο. Καλύτερη γραφική παράσταση είναι όταν το σήμα πλησιάζει όσο το δυνατόν στο γράφημα της κβάντισης.

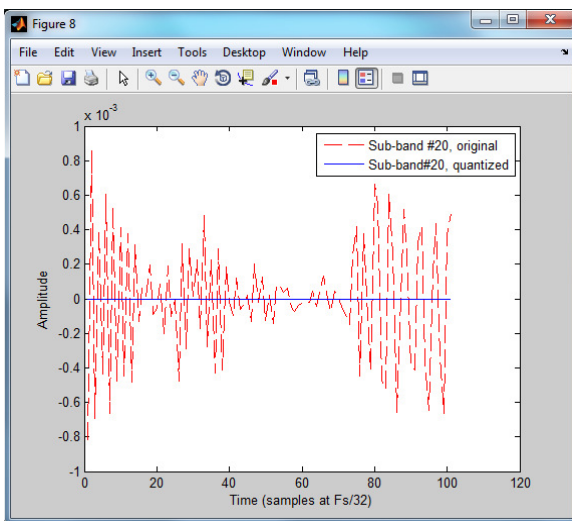
### 5.3.8 ΑΠΟΣΠΑΣΜΑΤΑ ΑΠΟ ΤΗΝ SUB-BAND #20



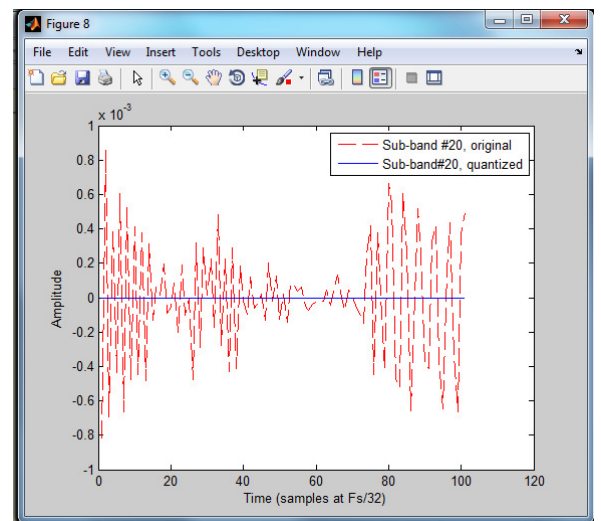
32



64



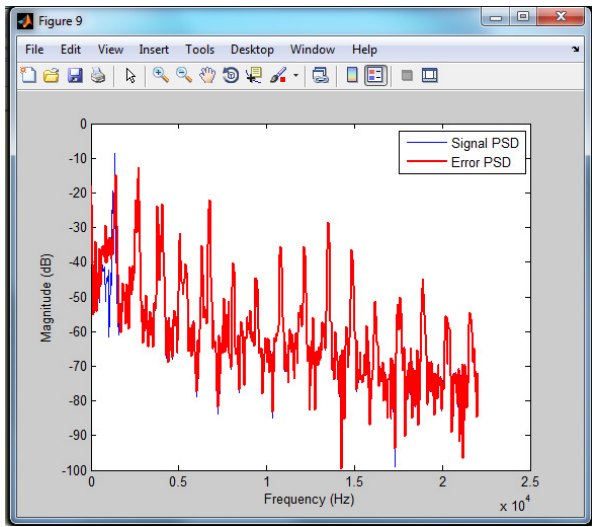
128



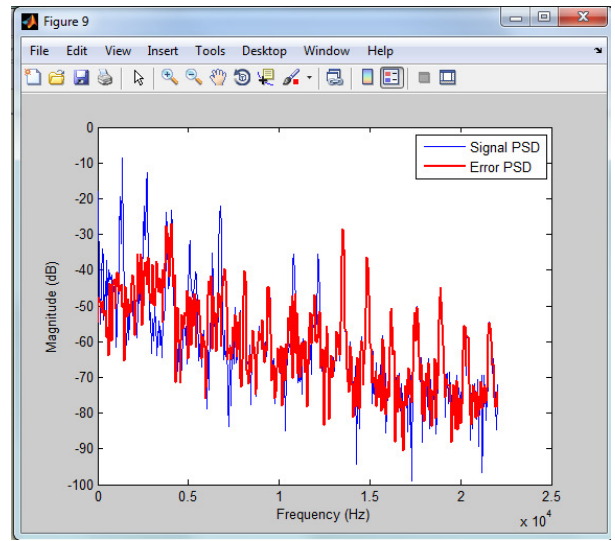
192

Το διάγραμμα απεικονίζει μια γραφική παράσταση σήματος σε συνάρτηση με το χρόνο. Όσο μεγαλύτερη τιμή δίνουμε στο sub-band τόσο υψηλή συχνότητα έχουμε.

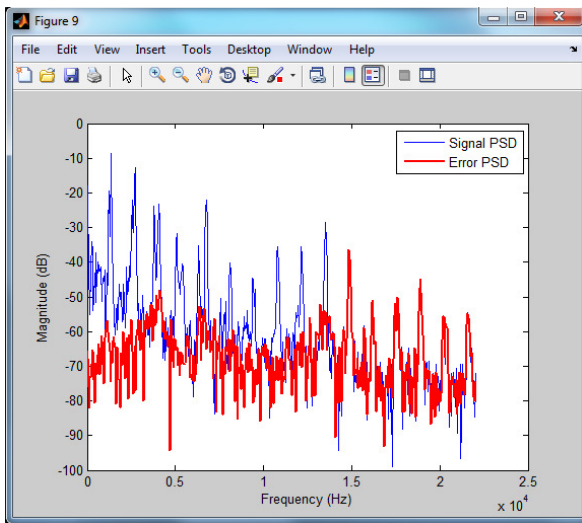
### 5.3.9 ΠΕΡΙΟΔΟΓΡΑΜΜΑΤΑ ΣΥΜΠΙΕΣΜΕΝΟΥ ΗΧΟΥ



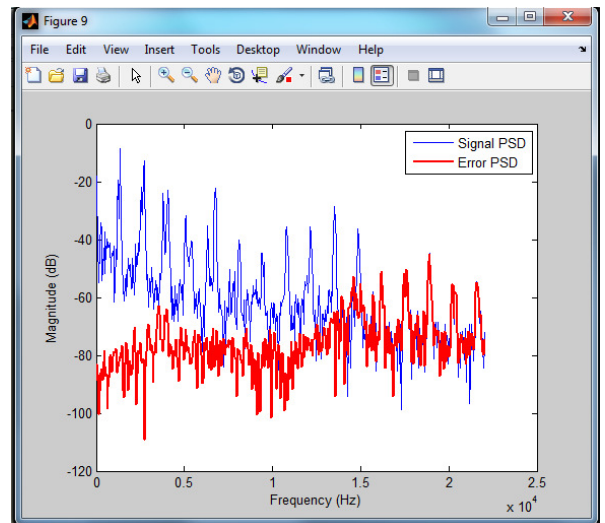
32



64



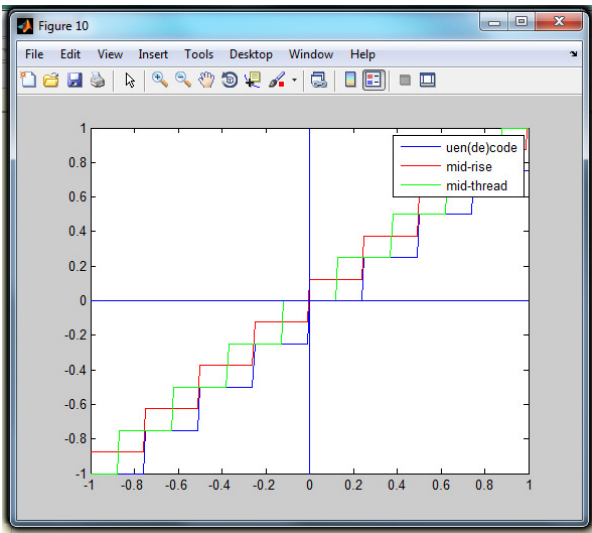
128



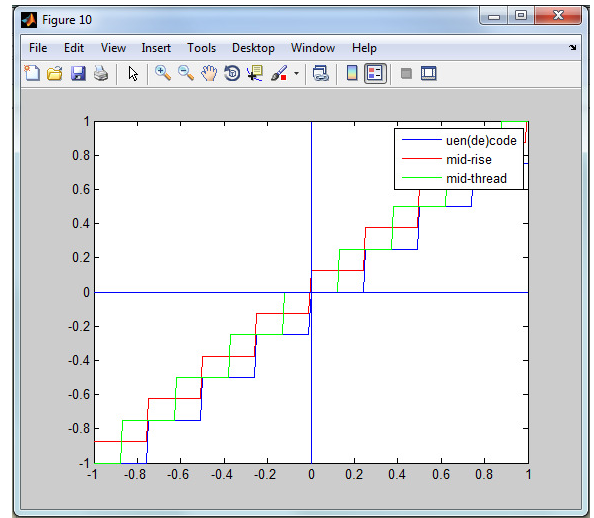
192

Το περιοδόγραμμα (figure 1) μας αναπαριστά τον αρχικό μας ήχο, στην συνέχεια έχουμε το figure 4 όπου είναι το στάδιο της δειγματοληψίας και της κβαντοποίησης και καταλήγουμε στο figure 9 όπου είναι και το τελικό δηλαδή ο συμπιεσμένος ήχος. Στις υψηλές συχνότητες δεν έχω καλό SNR σε αντίθεση με τις χαμηλές συχνότητες όπου το SNR είναι καλύτερο και έχει υψηλή ευκρίνεια.

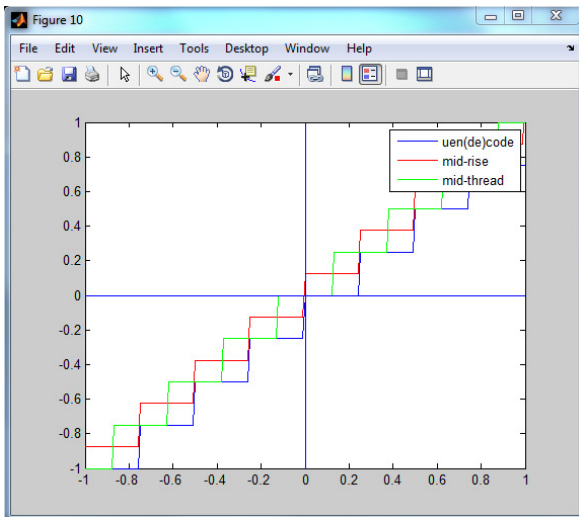
### 5.3.10 ΓΡΑΦΙΚΗ ΠΑΡΑΣΤΑΣΗ ΚΛΙΜΑΚΑΣ



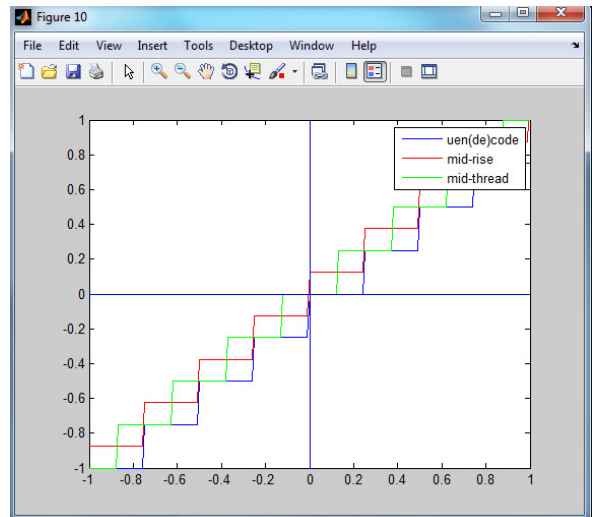
32



64



128



19

## 5.4 ΣΥΜΠΕΡΑΣΜΑ

Βασιζόμενοι στο περιοδόγραμμα συμπιεσμένου ήχου (figure 9) συμπεραίνουμε ότι η καλύτερη δυνατή ποιότητα ήχου είναι η τιμή bitrate 192. Όπως παρατηρούμε στην τιμή bitrate 32 το σφάλμα υπερκαλύπτει το σήμα, στην τιμή bitrate 64 αρχίζει να διαχωρίζεται το σήμα με το σφάλμα αντίστοιχα και στην τιμή bitrate 128. Ενώ στην τιμή bitrate 192 το σήμα έχει διαχωριστεί από το σφάλμα και αυτό παρατηρείται επίσης στο ακουστικό μέρος. Όμως, υπάρχουν εφαρμογές που οι υπόλοιπες τιμές bitrate μπορούν να θεωρηθούν πιο κατάλληλες σε σχέση με την τιμή 192.

## ΑΝΑΦΟΡΕΣ

[Ζαρ10] ΖΑΡΟΥΧΑΣ ΘΩΜΑΣ, «Μοντελοποίηση και Επεξεργασία Ηχητικών Δεδομένων για Αναπαραγωγή σε Χώρους με Αντήχηση » Διδακτορική Διατριβή, Πάτρα Ιούνιος 2010.

[Μητ11] ΜΗΤΡΑΚΗΣ ΜΑΡΙΟΣ, « Ακουστική Μελέτη και Βελτίωση του control room του studio Ηχογραφήσεων στο Α.Τ.Ε.Ι Ρεθύμνου » Πτυχιακή εργασία, Ρέθυμνο 2011.

[Μου09] ΜΟΥΣΑΣ ΧΡΗΣΤΟΣ, « ISEE Διαδραστική Εγκατάσταση για την εκτέλεση Ψυχοακουστικών πειραμάτων » Πτυχιακή εργασία, Κέρκυρα Ιούνιος 2009.

[Παρ95] ΠΑΡΑΣΚΕΥΑΣ ΜΙΧΑΗΛ, « Διαφορική Υποκειμενική Ακουστική Κωδικοποίηση (ΔΥΑΚ): μια Βελτιωμένη Τεχνική Ηχητικής Κωδικοποίησης » Διδακτορική Διατριβή, Πάτρα Ιούλιος 1995.

[Σαρ12] ΣΑΡΑΝΤΙΔΗ ΙΩΑΝΝΗ, « Μελέτη παραμέτρων σχεδιασμού σύγχρονων ασύρματων δικτύων για την υποστήριξη ψηφιακών τηλεοπτικών μεταδόσεων » Διπλωματική εργασία, Πάτρα Ιούλιος 2012.

[Χρυ06] ΧΡΥΣΑΝΘΟΥ ΜΑΡΙΑ, « ΔΙΑΔΙΚΑΣΙΑ ΨΗΦΙΟΠΟΙΗΣΗΣ ΗΧΗΤΙΚΩΝ ΑΡΧΕΙΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΕΣ ΨΗΦΙΟΠΟΙΗΜΕΝΟΥ ΗΧΟΥ » Πτυχιακή εργασία, Κέρκυρα 2006.

[link1] Wikipedia link:

<http://el.wikipedia.org/wiki/%CE%A8%CF%85%CF%87%CE%BF%CE%B1%CE%BA%CE%BF%CF%85%CF%83%CF%84%CE%B9%CE%BA%CE%AE>

[link2] Link:

[http://www.ipet.gr/digitech2/index.php?option=com\\_content&task=view&id=89&Itemid=57](http://www.ipet.gr/digitech2/index.php?option=com_content&task=view&id=89&Itemid=57)

