

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

## **ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Κωδικοποίηση πολυκαναλικών ηχητικών  
σημάτων βασιζόμενη στην μοντελοποίηση  
ενδο-ωτικών παραμέτρων.**

---

**Χατζηγιάννη Βασιλική**

**Επιβλέπων καθηγητής: Μιχαήλ Παρασκευάς**

Αντίριο Ιανουάριος 2016

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή

Αντίρριο, Ημερομηνία

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Ονοματεπώνυμο, Υπογραφή
2. Ονοματεπώνυμο, Υπογραφή
3. Ονοματεπώνυμο, Υπογραφή

## Ευχαριστίες

---

Αρχικά θέλω να ευχαριστήσω όλους τους καθηγητές μου για τις πολύτιμη συμβολή τους στην ολοκλήρωση των προπτυχιακών σπουδών μου που ολοκληρώνονται με την παρούσα εργασία. Πολλές ευχαριστίες οφείλω στους γονείς μου που χωρίς την αμέριστη υλική και ηθική τους υποστήριξη θα ήταν αδύνατον να καταφέρω τον σκοπό μου. Επίσης θέλω να ευχαριστήσω τον καθηγητή μου κ. Μιχαήλ Παρασκευά για την καθοδήγηση και την επίβλεψη της παρούσας πτυχιακής εργασίας

## Περιεχόμενα

Ευχαριστίες .....	3
ΚΕΦΑΛΑΙΟ 1 : Ψυχοακουστική Θεωρία .....	6
1.1 Εισαγωγή .....	6
1.2 Ανατομία του αυτιού .....	6
1.2.1 Εξωτερικό Αυτί .....	7
1.2.2 Μέσο Αυτί.....	8
1.2.3 Εσωτερικό Αυτί.....	9
1.3 Ακουστική Επικάλυψη.....	10
1.4 Χωρική Ακουστική Αντίληψη .....	13
1.5 Ενδωτικές Παράμετροι.....	14
1.6 Η διαδικασία της ισοστάθμισης.....	16
ΚΕΦΑΛΑΙΟ 2ο : Συμπίεση και Κωδικοποίηση .....	19
2.1 Αναλογικά και Ψηφιακά Σήματα .....	19
2.2 Συμπίεση.....	21
2.3 Κατηγορίες Συμπίεσης Δεδομένων .....	22
2.3.1 Μη απωλεστική συμπίεση .....	22
2.3.2 Απωλεστική συμπίεση.....	22
2.4 Αλγόριθμοι Συμπίεσης Δεδομένων.....	23
2.4.1 Αλγόριθμος RLE .....	23
2.4.2 Αλγόριθμος του Huffman .....	24
2.4.3 Συμπίεση LZW .....	26
2.5 Κωδικοποίηση Δεδομένων.....	26
2.5.1 Ψηφιακή Παράσταση Ήχου .....	26
2.5.2 Άλλα πρότυπα .....	28
ΚΕΦΑΛΑΙΟ 3ο : ΠΡΟΤΥΠΟ MPEG-1 (AUDIO LAYER 3) - MP3.....	32
3.1 Ανάπτυξη του προτύπου .....	32
3.2 Τεχνικά χαρακτηριστικά του προτύπου .....	36
3.2.1 Αποκωδικοποίηση Ήχου.....	36
3.2.2 Ρυθμός Bit .....	37
3.2.3 Ποιότητα ήχου.....	38
3.3 Η δομή ενός αρχείου MP3 .....	40
3.3.1 Σχεδιαστικοί περιορισμοί.....	41

3.3.2	Ετικέτες ID3 .....	41
3.3.3	Εξομάλυνση Έντασης ήχου.....	42
3.4	Αδειοδότηση και Ευρισιτεχνίες του MP3.....	42
3.5	Εναλλακτικές τεχνολογίες .....	46
ΚΕΦΑΛΑΙΟ 4ο : ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ.....		46
4.1	ΚΩΔΙΚΑΣ MATLAB .....	56
A)	Επανακβάντιση Ηχητικού σήματος με 8 bits ευκρίνειας ( χωρίς την παρουσία dither και NSF) .....	57
B)	Επανακβάντιση με 8 bits ευκρίνειας ( με την παρουσία RPDF dither και NSF 3ης τάξης) .....	57
Γ)	Επανακβάντιση με 4 bits ευκρίνειας ( χωρίς dither και NSF) .....	58
	Συνάρτηση QuantiseAudio.....	59
	Συνάρτηση dithmake.....	61
ΒΙΒΛΙΟΓΡΑΦΙΑ .....		63

# ΚΕΦΑΛΑΙΟ 1 : Ψυχοακουστική Θεωρία

## 1.1 Εισαγωγή

Η αντίληψη είναι η διαδικασία με την οποία οι άνθρωποι αισθάνονται, επιλέγουν, οργανώνουν και ερμηνεύουν πληροφορίες (π.χ. στην όραση, στον ήχο και στην ακοή) για να σχηματίσουν μια υποκειμενική εικόνα του κόσμου έτσι ώστε να αναγνωρίσουν, να ανακτήσουν και να αποκριθούν στην πληροφορία. Έτσι η αντίληψη είναι στενά συνυφασμένη με τη γνώση. Το ακουστικό σύστημα αντίληψης είναι ένα υψηλής περιπλοκότητας δομημένο σύστημα, που αποτελείται ανατομικά από διάφορα περιφερειακά υποσυστήματα όπως και από εκατομμύρια συνδεδεμένα κεντρικά νεύρα.

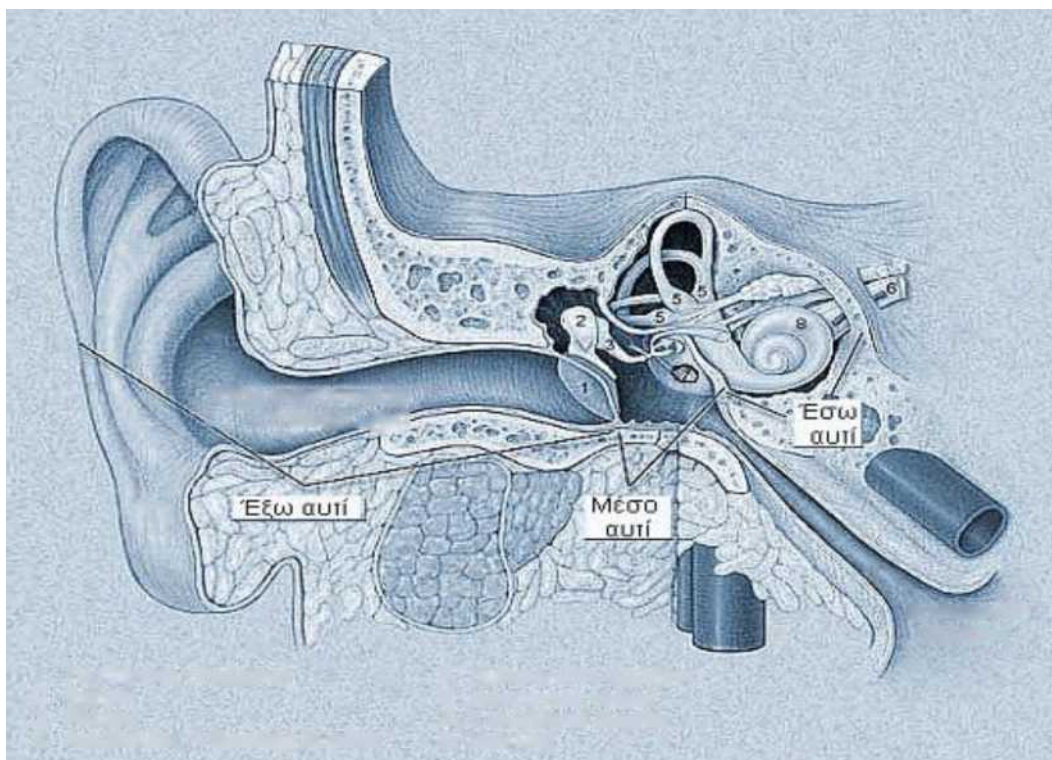
Η μελέτη της αντίληψης παραδοσιακά ξεκινά με ένα μοντέλο της αντιληπτικής διαδικασίας, που είναι μια σειρά βημάτων που ξεκινούν από την άφιξη ενός εξωτερικού ερεθίσματος ακολουθούμενη από την επεξεργασία του ερεθίσματος από τα όργανα-δέκτες και τα κεντρικά ακουστικά νεύρα, οδηγώντας στην τελική μετατροπή της πληροφορίας σε αντίληψη και γνώση. Η ψυχοακουστική (και η γνωστική ψυχολογία) είναι οι κυριότεροι κλάδοι που παρέχουν την λειτουργική/συμπεριφορική ένωση του ερεθίσματος με την αντίληψη/γνώση. Το φυσιολογικό επίπεδο της ανάλυσης επικεντρώνεται στη σχέση μεταξύ του ερεθίσματος και της προκαλούμενης επεξεργασίας σε διάφορα δομικά στάδια.

## 1.2 Ανατομία του αυτιού

Το ανθρώπινο σύστημα ακοής αποτελείται από δύο υποσυστήματα :

- (α) Το περιφερειακό, στο οποίο γίνεται η μετατροπή των ηχητικών κυμάτων σε ηλεκτροχημικά σήματα.
- (β) Το 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 πραγματοποιείται η προ-επεξεργασία του ακουστικού σήματος και ο μετασχηματισμός του σε ηλεκτροχημικό σήμα.



**Σχήμα 1.1 : Ανατομία ανθρώπινου αυτιού**

### **1.2.1 Εξωτερικό Αυτί**

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. Παράλληλα, επιτελεί και ένα πιο σημαντικό και περίπλοκο ρόλο, αυτόν του προσδιορισμού της θέσης στο χώρο μιας ακουστικής πηγής (source localization). Η χρησιμότητα του εξωτερικού ακουστικού καναλιού εστιάζεται σε δυο βασικά χαρακτηριστικά: (**α**) προστατεύει το μέσο αυτί και το τύμπανο, (**β**) «τοποθετεί» το εσωτερικό αυτί πολύ κοντά στον εγκέφαλο, μειώνοντας έτσι το απαιτούμενο μήκος των νεύρων, εξασφαλίζοντας ευκολότερη μετάβαση των παραγόμενων ασθενών ηλεκτρικών σημάτων προς τον εγκέφαλο. Παράλληλα το εξωτερικό ακουστικό κανάλι επιδρά σημαντικά στην απόκριση συχνότητας του συνολικού ακουστικού συστήματος. Ουσιαστικά συμπεριφέρεται σαν ένας ανοικτός σωλήνας συντονισμού με μέσο μήκος 2 cm που αντιστοιχεί στο 1/4 του μήκους κύματος για συχνότητες κοντά στα 4 kHz. Ενισχύει δηλαδή τις συχνότητες στην περιοχή των 4 kHz και εξηγείται με αυτό τον τρόπο η υψηλή επιλεκτικότητα του ανθρώπινου ακουστικού συστήματος για ηχητικά σήματα με συχνοτικό περιεχόμενο κοντά στην παραπάνω τιμή. Προς επιβεβαίωση των παραπάνω, το έντονο βύθισμα που παρατηρείται στο απόλυτο κατώφλι ακουστότητας εντοπίζεται σε αυτή τη συχνοτική περιοχή.

### 1.2.2 Μέσο Αυτί

Το μέσο αυτί ανατομικά αποτελείται από το τύμπανο, από τρία μικρά και πολύ σκληρά οστά ή ακουστικά οστάρια (σφύρα, άκμονας και αναβολέας) και από την ευσταχιανή σάλπιγγα. Λειτουργεί σαν ακουστικό-μηχανικός μετατροπέας, προσαρμόζοντας κατάλληλα τις διαφορετικές εμπεδήσεις του αέρα και του λεμφικού υγρού που βρίσκεται στο εσωτερικό αυτί. Ο όρος προσαρμογή εμπεδήσεων αναφέρεται στη μετατροπή των ταλαντώσεων του αέρα, οι οποίες έχουν μικρές δυνάμεις και μεγάλες απομακρύνσεις, σε κινήσεις του λεμφικού υγρού, με αντίστοιχα μεγάλες δυνάμεις και μικρές απομακρύνσεις, χωρίς ενεργειακές απώλειες. Η απόκριση συχνότητας του μέσου αυτιού παρουσιάζει τη μορφή υψηλοδιαβατού φίλτρου με κλίση 9-12 dB/οκτάβα κάτω από τα 1000 Hz και με ελαφριά ενίσχυση στην περιοχή των μεσαίων συχνοτήτων, γεγονός που συντελεί στην αυξημένη ευαισθησία της ακοής στην περιοχή αυτή. Μια σημαντική ιδιότητα του μέσου αυτιού,

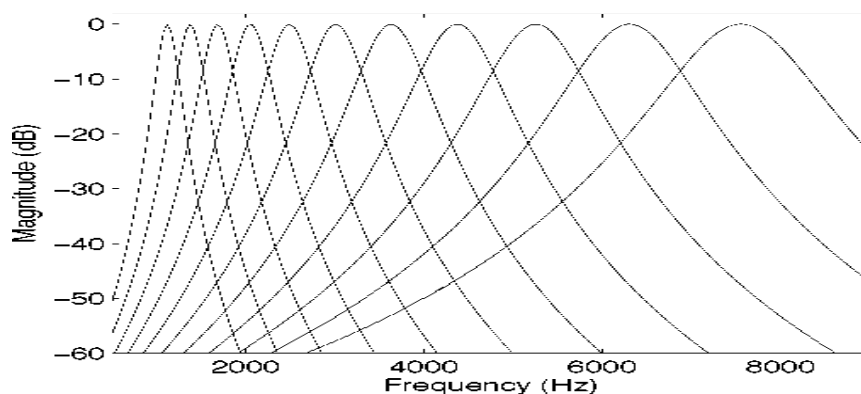


είναι η μη-γραμμική συμπεριφορά του για απότομα υψηλές στάθμες του ηχητικού σήματος, που αποσκοπεί στην προστασία του εσωτερικού αυτιού από επικίνδυνες υπερφορτώσεις. Τέλος, ο ρόλος της ευσταγγιανής σάλπιγγας είναι να εξισορροπεί την ατμοσφαιρική πίεση στις δυο πλευρές του τυμπάνου όπως αυτή μεταβάλλεται από διάφορες εξωτερικές δραστηριότητες (π.χ., χρήση ανελκυστήρα, καταδύσεις κλπ) που ουσιαστικά μεταβάλλουν τη θέση ισορροπίας του τυμπάνου.

### 1.2.3 Εσωτερικό Αυτί

Αποτελεί το πιο ενδιαφέρον και περισσότερο περίπλοκο τμήμα του ανθρώπινου ακουστικού συστήματος όπου μετατρέπει την μηχανική ενέργεια των δονήσεων που φθάνουν από το τύμπανο και το μέσο αυτί, σε ηλεκτροχημικό σήμα και αποτελείται από τον κοχλία και από τον λαβύρινθο. Ο κοχλίας έχει μήκος περίπου 32 mm, αποτελείται από 2.5 στροφές και περιέχει δυο διαφορετικά οργανικά υγρά (ενδόλυμφος και περίλυμφος), τα οποία διαχωρίζονται από τη βασική μεμβράνη. Επάνω στη μεμβράνη βρίσκεται το όργανο του, που συγκροτείται από 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 λειτουργίας της ακοής αφού επιτελεί τη συχνοτική 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 του «αναλυτή συχνοτήτων». Για πολύ χαμηλές συχνότητες η ταλάντωση της βασικής μεμβράνης μηδενίζεται στο τέρμα του κοχλίου, ενώ για υψηλότερες νωρίτερα και για την περίπτωση ύπαρξης περισσότερων από μία συχνοτήτων, παρατηρούνται μέγιστα πλάτους σε περισσότερες θέσεις. Όπως αναφέρθηκε παραπάνω, το όργανο του Corti 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 με αυτόν τον τρόπο την ηλεκτρική αντίσταση του κυττάρου, το οποίο με ηλεκτροχημικές διαδικασίες, παράγει παλμούς ηλεκτρικού ρεύματος που μεταφέρονται στον εγκέφαλο από ένα 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 κυμαίνεται από 1500 έως 2000, ενώ ο συνολικός αριθμός των νευρικών ινών είναι περίπου 30000. Κάθε αισθητήριο κύτταρο κωδικοποιεί μια στενή περιοχή συχνοτήτων της τάξης των μερικών εκατοντάδων Hz, ενώ περισσότερο πολύπλοκη κωδικοποίηση γίνεται από τις νευρικές ίνες και εν συνεχεία στον ακουστικό φλοιό από τα ανώτερα ακουστικά κέντρα του εγκεφάλου. Πρέπει να σημειωθεί ότι υπάρχουν δυο ειδών αισθητήρια κύτταρα, τα οποία δρουν «ανταγωνιστικά», όπως επίσης και δυο ειδών νευρικές ίνες που εξασφαλίζουν αμφίδρομη επικοινωνία με τον εγκέφαλο, δημιουργώντας έτσι ένα περίπλοκο μη-γραμμικό σύστημα με ανάδραση. Μη-γραμμική συμπεριφορά θεωρείται 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 1000 Hz. Όπως προέκυψε από την παραπάνω περιγραφή και για τους σκοπούς της ανάλυσης ηχητικών σημάτων, η λειτουργία του εσωτερικού αυτιού μοντελοποιείται ικανοποιητικά από μια τράπεζα παράλληλων ζωνοδιαβατών φίλτρων, όπως φαίνεται και στο Σχήμα 2.2



**Σχήμα 1.2: Προσομοίωση της απόκρισης συχνότητας του εσωτερικού αυτιού με χρήση ζωνοδιαβατών φίλτρων.**

### 1.3 Ακουστική Επικάλυψη

Επάνω στη βασική μεμβράνη επιτελείται διαχωρισμός των φασματικών συνιστωσών

του ακουστικού σήματος, οφειλόμενος στη μεταβαλλόμενη σκληρότητα της. Συνεπώς, ένα ακουστικό σήμα μιας συχνότητας (τονικό σήμα) διεγείρει μια συγκεκριμένη περιοχή της μεμβράνης, με πλάτος απομάκρυνσης ανάλογο με την ένταση του σήματος. Όπως είναι φυσικό η απομάκρυνση της βασικής μεμβράνης μετατοπίζεται και σε γειτονικές περιοχές, οι οποίες αντιστοιχούν σε άλλες συχνότητες. Αν ταυτόχρονα με το προηγούμενο σήμα, διεγείρει το ακουστικό σύστημα και άλλο ηχητικό σήμα, με παραπλήσια συχνότητα, αλλά με μικρότερη ένταση, τότε η διασπορά της ταλάντωσης της βασικής μεμβράνης που οφείλεται στο ισχυρό σήμα θα υπερκαλύψει την ταλάντωση της μεμβράνης που οφείλεται στο λιγότερο ισχυρό σήμα. Συνεπώς, ο δεύτερος ήχος παρά το γεγονός ότι διεγείρει το αυτί δεν γίνεται ακουστός ή μειώνεται η ακουστότητα του. Η διαδικασία αυτή αναφέρεται ως φαινόμενο επικάλυψης (masking effect) και παρουσιάζει σημαντικό ερευνητικό ενδιαφέρον σε διάφορες εφαρμογές επεξεργασίας ψηφιακού ήχου (π.χ. συμπίεση ηχητικών δεδομένων). Αποτελεί χαρακτηριστικό φαινόμενο της καθημερινής εμπειρίας, αφού η ακουστότητα ενός σήματος (π.χ. ομιλία) μπορεί να μειωθεί ή να εξαφανισθεί εντελώς, όταν εμφανισθεί ταυτόχρονα ένας άλλος ισχυρότερος ήχος (π.χ. θόρυβος) με παρόμοιο συχνοτικό περιεχόμενο. 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

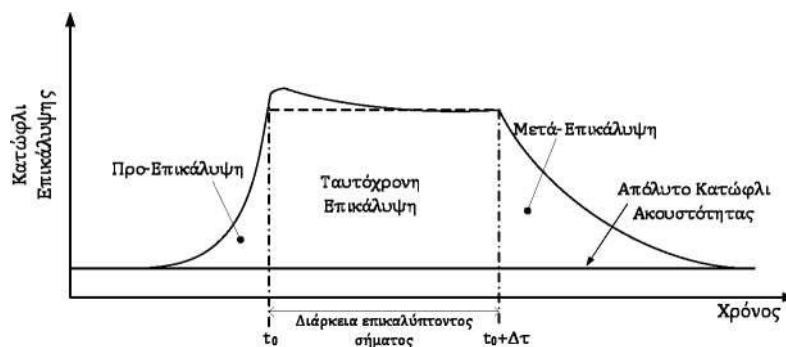
Σχήμα 2.4:

**(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 συχνοτικό περιεχόμενο.

**(β) Μετά-επικάλυψη (forward masking):** Για την δημιουργία του φαινομένου της μετά-επικάλυψης (στην βιβλιογραφία αναφέρεται επίσης ως post-stimulus masking ή

postmasking) το υπό-εξέταση σήμα παρουσιάζεται αμέσως μετά την παύση του επικαλύπτοντος σήματος. Με την ενέργεια αυτή, το δημιουργηθέν κατώφλι επικάλυψης δεν μειώνεται απότομα, αλλά μεσολαβεί ένα χρονικό διάστημα (της τάξης των μερικών εκατοντάδων ms) έως ότου εκφυλιστεί στο απόλυτο κατώφλι ακουστότητας. Διάφορες παράμετροι που διαμορφώνουν ή καθορίζουν το δημιουργηθέν κατώφλι επικάλυψης είναι: η χρονική καθυστέρηση του υπό εξέταση σήματος, η στάθμη και η χρονική διάρκεια του επικαλύπτοντος σήματος, η χρονική διάρκεια, η φάση και το συχνοτικό περιεχόμενο του υπό εξέταση σήματος καθώς και η συχνοτική διασπορά των δύο 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 (στην βιβλιογραφία επίσης αναφέρεται ως pre-stimulus masking ή premasking). 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 συμμετέχοντων 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 (μέγιστη τιμή) πριν την παρουσία του επικαλύπτοντος σήματος. Επιπλέον, στην εργασία υπονοείται σαφώς ότι το φαινόμενο της προ-επικάλυψης καθορίζεται σε σημαντικό βαθμό από ανώτερες εγκεφαλικές λειτουργίες.



**Σχήμα 1.3: Σχηματικό διάγραμμα για την απεικόνιση και το χαρακτηρισμό χρονικών περιοχών για τις διάφορες περιπτώσεις της ακουστικής επικάλυψης.**

### 1.4 Χωρική Ακουστική Αντίληψη

Ο τρόπος με τον οποίο αντιλαμβανόμαστε ένα ακουστικό γεγονός, που εξελίσσεται στο χώρο, παρουσιάζει παρόμοια χαρακτηριστικά με τον τρόπο που αντιλαμβανόμαστε, μέσω του οπτικού μας συστήματος, μια οπτική εικόνα. Τα διαφορετικά αντικείμενα τα οποία είναι μέρος της χωρικής ακουστικής εικόνας, αναφέρονται ως ακουστικά γεγονότα. Για παράδειγμα, τα ακουστικά γεγονότα σε μια μουσική παράσταση είναι τα διαφορετικά μουσικά όργανα τα οποία συμμετέχουν σε αυτή. Σε μια προσπάθεια να κατανοήσουμε πως το ακουστικό σύστημα διακρίνει την κατεύθυνση μιας πηγής, οι ιδιότητες των σημάτων στις εισόδους των ακουστικών καναλιών, πρέπει να ληφθούν υπόψη. Χωρίς βλάβη της γενικότητας, τα σήματα στην είσοδο των ακουστικών καναλιών μπορεί να θεωρηθούν ως φιλτραρισμένες εκδόσεις του αρχικού σήματος. Τα φίλτρα τα οποία μοντελοποιούν τη διαδρομή του σήματος από την πηγή στις εισόδους των ακουστικών καναλιών, αναφέρονται και ως συναρτήσεις HRTF (Head Related Transfer Functions).



**Σχήμα 1.4: (a) Μοντελοποίηση των διαφορετικών διαδρομών από την πηγή στις ακουστικές εισόδους με χρήση συναρτήσεων HRTF, (b) Συσχέτιση της γωνίας  $\phi$  με τις αποστάσεις  $d_1$  και  $d_2$  και τη δημιουργία διαφορετικών χρόνων άφιξης. Επιπρόσθετα, η σκίαση του κεφαλιού επιφέρει διαφορές στις στάθμες των σημάτων εισόδου ως συνάρτηση της γωνίας  $\phi$ .**

Το Σχήμα 2.4, απεικονίζει τις δύο (αριστερά-δεξιά) συναρτήσεις HRTFs,  $h_1$  και  $h_2$ , για μια πηγή σε γωνία  $\varphi$ . Για διαφορετική θέση της πηγής, αντίστοιχες συναρτήσεις HRTFs πρέπει να χρησιμοποιηθούν για την μοντελοποίηση των σημάτων κατά την είσοδο τους στα ακουστικά κανάλια. Η διαφορά του μήκους των διαδρομών από την πηγή στα ακουστικά κανάλια,  $d_1 - d_2$ , μπορεί να εκφραστεί ως συνάρτηση της γωνίας  $\varphi$ . Εξ αιτίας του διαφορετικού μήκους των διαδρομών  $d_1$  και  $d_2$ , υπάρχουν διαφορετικοί χρόνοι άφιξης για τα σήματα στις ακουστικές εισόδους. Η πιο απλή σχέση (στην βιβλιογραφία αναφέρεται ως sine law [Blauert (1997)]) που περιγράφει τις διαφορές των διαδρομών ανάμεσα στις δύο ακουστικές εισόδους, προτάθηκε από τον Hornbostel [von Hornbostel (1920)] και δίνεται ακολούθως:

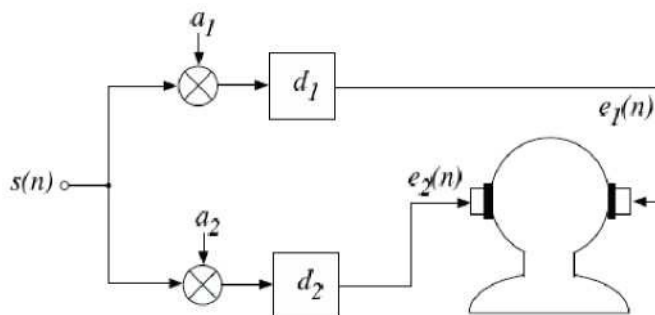
$$\Delta d = k \cdot \sin \varphi, k = 21 \text{ cm}$$

όπου το  $k$  περιγράφει την απόσταση ανάμεσα σε δυο μικρόφωνα τα όποια μοντελοποιούν τις δυο ακουστικές εισόδους. Επειδή το καμπύλο τμήμα από το κεφάλι δεν λαμβάνεται υπόψη, επιλέγεται μια μεγαλύτερη τιμή για το  $k$  σε σχέση με την πραγματική απόσταση των δυο ακουστικών εισόδων. Ένας ακόμη περιορισμός που εισάγει η παραπάνω έκφραση, είναι ότι δεν λαμβάνει υπόψη την σκίαση της κεφαλής, δηλαδή την επίδραση στις εντάσεις των σημάτων κατά την είσοδο τους στα ακουστικά κανάλια.

## 1.5 Ενδοωτικές Παράμετροι

Ως αποτέλεσμα των διαφορετικών διαδρομών ανάμεσα στην πηγή και τις δύο ακουστικές εισόδους, διαφορετικοί χρόνοι άφιξης του ηχητικού σήματος καταγράφονται στις δυο ακουστικές εισόδους, οι οποίοι καθορίζουν την ενδο-ωτική χρονική διαφορά (inter-aural time difference, ITD). Αντίστοιχα, η σκίαση της κεφαλής δημιουργεί διαφορετικές στάθμες στις εντάσεις του ηχητικού σήματος στις δυο ακουστικές εισόδους, οι οποίες διαμορφώνουν την ενδο-ωτική στάθμη διαφοράς (interaural level difference, ILD). Γενικότερα φαινόμενα διάχυσης, ανακλάσεων και συντονισμού που οφείλονται στο κεφάλι, τον κορμό καθώς και το εξωτερικό αυτί, έχουν ως αποτέλεσμα οι παράμετροι ILD και ITD, εκτός από την γωνία  $\varphi$ , να εξαρτώνται και από το είδος του ηχητικού σήματος. Από την προηγούμενη συζήτηση προκύπτει ότι οι παράμετροι ILD και ITD είναι ιδιότητες των ακουστικών σημάτων εισόδου και

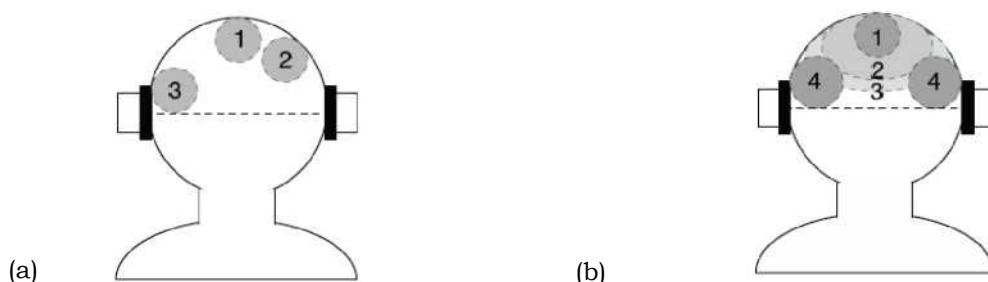
παρέχουν στο ακουστικό σύστημα την απαραίτητη πληροφορία για τον προσδιορισμό της κατεύθυνσης μιας πηγής. Συνεπώς, ένα ζευγάρι τιμών ILD-ITD μπορεί να συσχετισθεί με την προσδιορισμό της διεύθυνσης μιας πηγής. Στην περίπτωση αναπαραγωγής ηχητικών σημάτων μέσω ακουστικών, τα σήματα στις ακουστικές εισόδους μπορούν να θεωρηθούν (ιδανικά) ισοδύναμα με αυτά που μεταφέρονται από την πηγή στους ηλεκτροακουστικούς μετατροπείς των ακουστικών. Οπότε είναι εφικτή η ανεξάρτητη αξιολόγηση των παραμέτρων ILD και ITD με τη χρήση ακουστικών σε κατάλληλες πειραματικές διατάξεις. Στο Σχήμα 2.5, απεικονίζεται η πειραματική διάταξη για την δημιουργία συναφών ηχητικών σημάτων (όπου ο δείκτης  $n$  αναφέρεται σε ακολουθίες διακριτού χρόνου) στις δύο ακουστικές εισόδους (αριστερά-δεξιά),  $e_1(n)$  και  $e_2(n)$ , από ένα αρχικό ηχητικό σήμα (πηγή)  $s(n)$ . Η παράμετρος ITD καθορίζεται από τις καθυστερήσεις  $d_1, d_2$  και είναι ίση με  $d_2 - d_1$ , ενώ η παράμετρος ILD καθορίζεται από τους συντελεστές κλίμακας  $a_1$  και  $a_2$  και εκφράζεται σε dB από την σχέση  $20 \log(a_2/a_1)$ .



**Σχήμα 1.5 :** Πειραματική διάταξη για την δημιουργία σημάτων για την αριστερή και δεξιά ακουστική είσοδο με συγκεκριμένο ITD και ILD.

Στο Σχήμα 2.6(a) απεικονίζονται τα αντιληπτά ακουστικά γεγονότα για διαφορετικές τιμές των ITD και ILD για δυο συναφή ηχητικά σήματα τα οποία αναπαράγονται με βάση την πειραματική διάταξη του Σχήματος 2.5. Στην περίπτωση που τα δύο σήματα είναι συναφή (coherent), τότε έχουν την ίδια στάθμη (ILD=0) και καμία χρονική καθυστέρηση (ITD=0)

και το ακουστικό γεγονός εντοπίζεται στο κέντρο ανάμεσα στις δύο ακουστικές εισόδους, όπως φαίνεται στην Περιοχή 1 στο Σχήμα 2.6(a). Αυξάνοντας την ηχητική στάθμη σε μια πλευρά (π.χ. δεξιά), το ακουστικό γεγονός μετατοπίζεται αντίστοιχα όπως φαίνεται στην Περιοχή 2 στο Σχήμα 2.6(a). Στην ιδιάζουσα περίπτωση που μόνο το ηχητικό σήμα στην



**Σχήμα 1.6: (a) Καθορισμός της θέσης του ακουστικού γεγονότος στο επάνω μέρος της κεφαλής, όπως προκύπτει από τις παραμέτρους ILD και ITD (b) Το εύρος του ακουστικού γεγονότος αυξάνει (Περιοχές 1-3) καθώς η παράμετρος IC μειώνεται, έως ότου δύο διακριτά ακουστικά γεγονότα να δημιουργηθούν (Περιοχή 4).**

αριστερή πλευρά είναι ενεργό, το ακουστικό γεγονός περιορίζεται στην Περιοχή 3 (Σχήμα 2.6(a)). Αντίστοιχα η παράμετρος ITD μπορεί να χρησιμοποιηθεί με παρόμοιο τρόπο για τον έλεγχο της θέσης του ακουστικού γεγονότος.

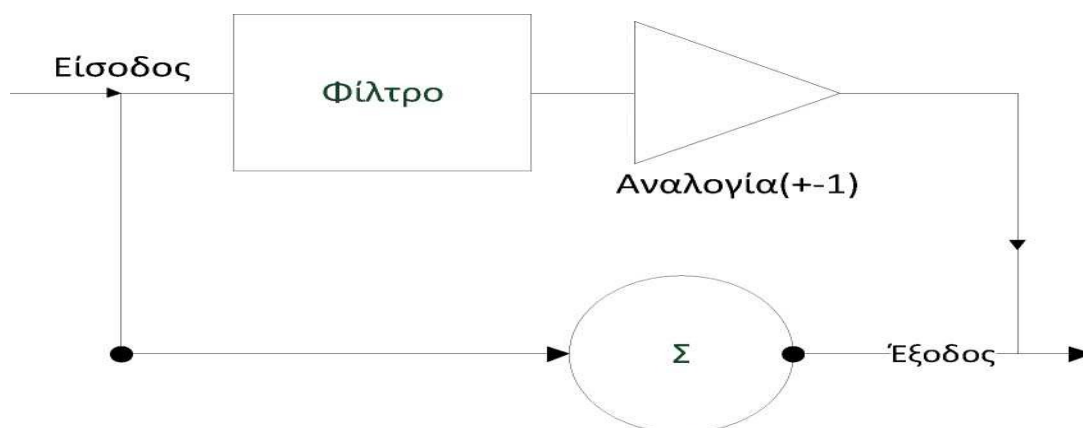
Εκτός από τις παραμέτρους ITD και ILD, μια ακόμη παράμετρος χρησιμοποιείται, η οποία αποτελεί μια μέτρηση του βαθμού ομοιότητας ανάμεσα στα δυο σήματα στις ακουστικές εισόδους και αναφέρεται ως ενδο-ωτική συνάφεια (inter-aural coherence, IC).

## 1.6 Η διαδικασία της ισοστάθμισης

Μια σημαντική και σχετική με το φιλτράρισμα περιοχή της επεξεργασίας σήματος



είναι η ισοστάθμιση . Αντίθετα με το φιλτράρισμα, η ισοστάθμιση δεν είναι διαδικασία κατά την οποία γίνεται αφαίρεση συχνοτικών στοιχείων αλλά αύξηση ή μείωση αυτών ώστε να επιτευχθεί το επιθυμητό αποτέλεσμα. Η διαδικασία της ισοστάθμισης μπορεί να μοντελοποιηθεί ως η διαδικασία της πρόσθεσης ή αφαίρεσης μιας φιλτραρισμένης έκδοσης σήματος από το ίδιο το σήμα όπως φαίνεται στο σχήμα.



**Σχήμα 1.7 : Διαδικασία Φιλτραρίσματος**

Προσθέτοντας τη φιλτραρισμένη έκδοση δίνει μια αύξηση στις συχνότητες που έχει επιλέξει το φίλτρο τη στιγμή που αφαιρώντας τη φιλτραρισμένη έξοδο μειώνει το πλάτος των συχνοτικών στοιχείων στο εύρος συχνοτήτων του φίλτρου. Το φίλτρο μπορεί να είναι ένα απλό ανωδιαβατό η κατωδιαβατό, που οδηγεί σε είτε έλεγχο των ψηλών είτε των χαμηλών συχνοτήτων, ή μπορεί να είναι ζωνοδιαβατό ώστε να δίνει μια απόκριση σε σχήμα Gauss. Οι συχνότητες αποκοπής των φίλτρων μπορούν να είναι είτε προκαθορισμένες η μεταβαλλόμενες ανάλογα με την εφαρμογή. Ακόμα το εύρος ζώνης των ζωνοδιαβατών φίλτρων και λιγότερο συχνά οι κλίσεις των κατωδιαβατών και ανωδιαβατών φίλτρων μπορεί να ποικίλουν.

Ένας ισοσταθμιστής στον οποίο όλοι οι παράμετροι του φίλτρου μπορούν να μεταβληθούν ονομάζεται παραμετρικός ισοσταθμιστής. Ωστόσο, στην πράξη, πολλές εφαρμογές χρησιμοποιούν ένα μέρος των πιθανών ελέγχων για απλότητα και ευκολία χρήσης. Τυπικά σ'αυτές τις περιπτώσεις μόνο οι συχνότητες αποκοπής των ζωνοπερατών φίλτρων, και σε κάποιες περιπτώσεις τα

κατωδιαβατά και ανωδιαβατά φίλτρα μεταβάλλονται. Η εναλλακτική χρήση της δομής του ισοσταθμιστή χρησιμοποιεί μια τράπεζα φίλτρων που βρίσκονται σε παραπλήσιες θέσεις ως προς την κάλυψη όλου του εύρους συχνοτήτων και αποτελούνται από παράλληλα ζωνοπερατά φίλτρα με σταθερό εύρος αλλά μεταβλητό κέρδος/αποκοπή[6]. Αυτό οδηγεί στο γραφικό ισοσταθμιστή με τυπικό εύρος ζώνης των μεμονωμένων φίλτρων που κυμαίνεται από 1/3 έως 1 οκτάβα. Για παραμετρικούς ισοσταθμιστές το εύρος ζώνης μπορεί να γίνει σχετικά μικρό. Επειδή όλοι οι ισοσταθμιστές χρησιμοποιούν φίλτρα έχουν τις ίδιες επιδράσεις με τα φίλτρα στα πεδία του χρόνου και συχνότητας. Αυτό γίνεται ιδιαίτερα εμφανές όταν χρησιμοποιείται ισοστάθμιση στενού εύρους ζώνης, καθώς το σχετικό φίλτρο μπορεί να ‘‘κουδουνίσει’’ για ένα σημαντικό χρονικό διάστημα και κατά την αύξηση και κατά τη μείωση. Οι ισοσταθμιστές έτσι έχουν ευρεία χρήση στην επεξεργασία ήχου. Παρ’όλ’αυτά, παρά τη χρησιμότητά τους, πρέπει να χρησιμοποιούνται με επιφύλαξη, πρώτον για να αποφευχθούν υπερβολές του χαρακτήρα του ήχου, εκτός κι αν αυτό είναι επιθυμητό, και δεύτερον ώστε να αποφευχθούν μη επιθυμητές παρεμβολές μεταξύ διαφορετικών ζωνών συχνοτήτων του ισοσταθμιστή. Σαν παράδειγμα μπορούμε να σκεφτούμε το αποτέλεσμα της πρόσθεσης ψηλών, μεσαίων και χαμηλών συχνοτήτων σε ένα δοσμένο σήμα. Επειδή υπάρχουν αναπόφευκτες παρεμβολές μεταξύ των αποκρίσεων συχνότητας του ισοσταθμιστή, το συνολικό αποτέλεσμα είναι να πάρουμε το ίδιο φάσμα με το αρχικό. Ότι έχει συμβεί είναι το κέρδος να είναι μεγαλύτερο. Σημειωτέον ότι αυτό μπορεί να γίνει αν μια συγκεκριμένη ζώνη συχνοτήτων αυξηθεί και, έπειτα, επειδή το αποτέλεσμα είναι λίγο υπερβολικό άλλες ζώνες συχνοτήτων ρυθμιστούν για την εξισορρόπηση. Κατά την ενεργητική ακύρωση θορύβου (Active noise cancellation), το ακουστικό σήμα προφιλτράρεται (prefiltering) από τον ισοσταθμιστή (equalizer) για την επακόλουθη μετάδοσή του στο ακουστικό σύστημα(π.χ. ακουστικά).

## ΚΕΦΑΛΑΙΟ 2ο : Συμπύεση και Κωδικοποίηση

### 2.1 Αναλογικά και Ψηφιακά Σήματα

Τα μεγέθη και δεδομένα διακρίνονται σε δύο κατηγορίες. Τα αναλογικά και ψηφιακά δεδομένα. Τα αναλογικά δεδομένα είναι τα δεδομένα που παίρνουν συνεχόμενες τιμές ενώ τα ψηφιακά αυτά που οι τιμές διακρίνονται σε δύο καταστάσεις 0 και 1.

Ο πραγματικός κόσμος είναι ένας αναλογικός κόσμος. Όλα τα μεγέθη παίρνουν τιμές με άπειρη ακρίβεια. Π.χ. το ηλεκτρικό σήμα τάσης όπου κάθε στιγμή η τιμή του μπορεί να απαιτεί άπειρα δεκαδικά ψηφία για να περιγραφεί και μπορεί η τάση στα άκρα μιας πηγής να πάρει οποιαδήποτε τιμή από 0 έως  $X$  volt. Ο ψηφιακός κόσμος των υπολογιστών, αντίθετα, μπορεί να χειριστεί μόνο πεπερασμένο αριθμό δεδομένων και με πεπερασμένη ακρίβεια. Οι υπολογιστές και γενικά τα πληροφοριακά συστήματα χρησιμοποιούν ψηφιακά δεδομένα. Για να μπορέσουμε να μεταφέρουμε τα δεδομένα μας λοιπόν επιβάλλεται η μετατροπή των αναλογικών σε ψηφιακά σήματα και το αντίθετο. Ο όρος ψηφιακό σήμα αναφέρεται σε περισσότερες από μια έννοιες. Μπορεί να αναφέρεται σε ένα σήμα διακριτού χρόνου το οποίο μπορεί να πάρει συγκεκριμένες (διακριτές) τιμές στον άξονα του χρόνου. Σε αυτή τη περίπτωση μιλάμε για ένα σήμα το οποίο παράγεται μέσω μιας μεθόδου ψηφιακής διαμόρφωσης και θεωρείται περισσότερο ως αναλογικό σήμα (δηλ. ένα επεξεργασμένο αναλογικό σήμα για τη μετατροπή του σε ψηφιακό).

Ένα αναλογικό σήμα είναι μια χρονικά μεταβαλλόμενη τιμή δεδομένων ή όπως είναι ο ορισμός του, μια ομαλά μεταβαλλόμενη τιμή ηλεκτρικής τάσης ή έντασης ρεύματος (δηλ. ένα σήμα με πλάτος χρονικά μεταβαλλόμενο) η οποία μπορεί να περιγραφεί από μια μαθηματική συνάρτηση, με το χρόνο να αποτελεί την ανεξάρτητη και την τιμή του σήματος, κάθε χρονική στιγμή, την εξαρτημένη μεταβλητή. Ένα διακριτό σήμα είναι το αποτέλεσμα που παίρνουμε μέσω της μεθόδου της δειγματοληπτικής

μείωσης από το αρχικό αναλογικό σήμα : δηλαδή, η τιμή των δεδομένων σημειώνεται ανά τακτά χρονικά διαστήματα (π.χ. μικροδευτερόλεπτο) και όχι συνεχώς (όπως είναι εκ φύσεως τα μηχανικά κύματα). Αν οι ατομικές τιμές του σήματος αντί να μετρηθούν επακριβώς, επάνω στον άξονα του χρόνου, είναι εναρμονισμένες με κάποια ορισμένη ακρίβεια, τότε η ροή δεδομένων που προκύπτει είναι το ψηφιακό σήμα. Η διαδικασία προσέγγισης αυτής της ακρίβειας (δηλ. μιας συγκεκριμένης τιμής), μέσα από ένα σταθερό αριθμό ψηφίων (δηλ. bit) ονομάζεται ψηφιοποίηση. Σε γενικές γραμμές, ένα ψηφιακό σήμα είναι ένα ψηφιοποιημένο σήμα διακριτού χρόνου. Το διακριτό σήμα είναι το αποτέλεσμα της επεξεργασίας ενός αναλογικού σήματος με τη μέθοδο της δειγματοληπτικής μείωσης. Η Ψηφιακή Επανάσταση έχει αυξήσει κατακόρυφα τη χρήση ψηφιακών σημάτων. Οι περισσότερες-αν όχι όλες, οι σύγχρονες συσκευές ειδικότερα αυτές που συνδέονται στους υπολογιστές, χρησιμοποιούν ψηφιακά σήματα για την αναπαράσταση σημάτων τα οποία παραδοσιακά αναπαριστώνταν ως σήματα συνεχούς χρόνου. Κινητά τηλέφωνα, συσκευές αναπαραγωγής βίντεο και ήχου, ψηφιακές φωτογραφικές μηχανές είναι μερικά παραδείγματα. Σχεδόν σε όλες τις εφαρμογές, τα ψηφιακά σήματα αναπαριστώνται μέσω των δυαδικών αριθμών, έτσι ώστε να μπορούν να μετρηθούν σε bit. Επειδή 7 bits (δυαδικά ψηφία) μπορούν να καταγράψουν 128 διακριτές τιμές (0 έως 127), αυτό το σύστημα είναι το πιο ικανό από οποιοδήποτε άλλο για να εκφραστεί ένα τεράστιο πλήθος τιμών. Τόσο στους υπολογιστές όσο και σε οποιοδήποτε άλλο ψηφιακό σύστημα, η κυματομορφή του σήματος εναλλάσσεται μεταξύ δύο επιπέδων τάσης (0 και 4,8V) οι οποίες αναπαριστούν αντίστοιχα τις δύο τιμές του δυαδικού συστήματος(0 και 1) .Έτσι, αναφερόμαστε σε αυτή τη κυματομορφή ως ψηφιακό σήμα. Παρ' όλο που πρόκειται για μια αναλογική κυματομορφή τάσεως, το ονομάζουμε ψηφιακό διότι εναλλάσσεται μεταξύ δύο σταθερών καταστάσεων. Το σήμα του ρολογιού είναι ένα ειδικό ψηφιακό σήμα το οποίο χρησιμοποιείται για τον συγχρονισμό των ψηφιακών κυκλωμάτων.Στην διπλανή εικόνα φαίνεται η συγκεκριμένη κυματομορφή. Οι λογικές αλλαγές ενεργοποιούνται είτε απο την αύξηση είτε από την μείωση του σήματος. Στο διπλανό διάγραμμα βλέπουμε ένα παράδειγμα του πρακτικού παλμού. Έτσι, όταν λέμε: Αύξηση του σήματος : εννοούμε την διαδικασία της μετάβασης από χαμηλή σε υψηλή τάση. Μείωση του σήματος : εννοούμε την διαδικασία της μετάβασης από υψηλή σε χαμηλή τάση.

## 2.2 Συμπίεση

Με τον όρο συμπίεση δεδομένων (data compression) εννοούμε τη μετατροπή ενός ψηφιακού αρχείου σε μικρότερο αρχείο (που περιέχει μικρότερο αριθμό μπιτ) με τρόπο ώστε να είναι δυνατή η επαναμετατροπή του συμπιεσμένου αρχείου στο αρχικό. Για το σκοπό αυτό χρησιμοποιούνται πολλές μέθοδοι, οι οποίες χωρίζονται σε δύο μεγάλες κατηγορίες: τις μη απωλεστικές και τις απωλεστικές. Αν θεωρήσουμε τα δεδομένα μας σαν μια ακολουθία τότε η συμπίεση ορίζεται όπως παρακάτω: Συμπίεση (compression) μιας ακολουθίας δεδομένων, ονομάζουμε την ελάττωση του μεγέθους της ακολουθίας, ώστε να χρειάζεται λιγότερος χώρος για την αποθήκευση ή τη μετάδοσή της. Η διαδικασία της συμπίεσης εφαρμόζεται συστηματικά στα υπολογιστικά συστήματα που χρησιμοποιούν και επεξεργάζονται μεγάλο όγκο ψηφιακών δεδομένων. Υπάρχουν πολλοί αλγόριθμοι για να καταφέρουμε να μειώσουμε το μέγεθος μίας ακολουθίας. Ένας αλγόριθμος συμπίεσης μπορεί να επιφέρει μεγάλη ελάττωση του μήκους σε ακολουθίες δεδομένων με κάποιο ιδιαίτερο χαρακτηριστικό, ενώ ο ίδιος να είναι αναποτελεσματικός για άλλες ακολουθίες, αφήνοντάς τις στο ίδιο μήκος ή ακόμα και μεγαλώνοντάς το σε μερικές περιπτώσεις. Όταν έχουμε την ακολουθία των δεδομένων σε συμπιεσμένη μορφή, πρέπει να εφαρμοστεί η αντίστροφη διαδικασία της *αποσυμπίεσης* (decompression, extraction) προκειμένου τα δεδομένα να μπορούν να χρησιμοποιηθούν και πάλι. Η διαδικασία αυτή της αποσυμπίεσης των δεδομένων καθορίζει και τις κατηγορίες των μεθόδων συμπίεσης. Έτσι, υπάρχουν δύο κατηγορίες αλγορίθμων συμπίεσης, οι *απωλεστικοί* (lossy) και οι μη *απωλεστικοί* (lossless) αλγόριθμοι.

Στους απωλεστικούς αλγορίθμους, όταν γίνει η συμπίεση και μετά ακολουθήσει αποσυμπίεση των δεδομένων, η τελική ακολουθία των δεδομένων διαφέρει από την αρχική. Αντίθετα στους μη απωλεστικούς αλγορίθμους, η διαδικασία συμπίεσης και αποσυμπίεσης επαναφέρει την αρχική ακολουθία. Αν πρέπει να μεταφερθούν δεδομένα με απόλυτη ακρίβεια, πιστότητα, χωρίς να αλλοιωθεί το περιεχόμενό τους, πρέπει να εφαρμοστεί μια μη απωλεστική μέθοδος συμπίεσης. Εφαρμογές ή συστήματα που μεταδίδουν αναλλοίωτες πληροφορίες από το ένα μέσο στο άλλο, για παράδειγμα κάρτες δικτύου ή modem, χρησιμοποιούν τεχνικές μη απωλεστικής συμπίεσης. Υπάρχουν όμως εφαρμογές όπου η μικρή διαφοροποίηση από την αρχική

μορφή των δεδομένων δεν επιφέρει σημαντικές αλλαγές. Έτσι, οι περισσότερες εφαρμογές που έχουν να κάνουν με σύνθετες μορφές δεδομένων όπως είναι ο ήχος, η εικόνα, το video, όπου το τελικό αποτέλεσμα αξιολογείται από τον ανθρώπινο παράγοντα (αυτί, μάτι), μπορούν να κάνουν απωλεστική συμπίεση χωρίς πολλές φορές να υπάρχουν εμφανείς αλλοιώσεις στην ποιότητα των δεδομένων. Πολλά αρχεία περιέχουν μεγάλα τμήματα τα οποία επαναλαμβάνονται, όπως ένα αρχείο κειμένου μπορεί να περιέχει πολλές φορές μια λέξη π.χ. Test οπότε αρκεί η λέξη να τοποθετηθεί σε έναν πίνακα με λέξεις και να αντικαθίσταται με έναν αριθμό που δείχνει τη θέση της στον πίνακα, ή περιέχουν δεδομένα που η απώλειά τους δεν προκαλεί σοβαρή μεταβολή του περιεχομένου.

Η συμπίεση μπορεί να μειώσει τον όγκο των δεδομένων που στέλνονται ή αποθηκεύονται, με την ελαχιστοποίηση του ενυπάρχοντος πλεονασμού. Ο πλεονασμός παρουσιάζεται κατά τη δημιουργία των δεδομένων. Με τη διαδικασία της συμπίεσης η μεταφορά και η αποθήκευση γίνονται με πιο αποδοτικό τρόπο, ενώ παράλληλα διατηρείται η ακεραιότητα των δεδομένων.

## **2.3 Κατηγορίες Συμπίεσης Δεδομένων**

### **2.3.1 Μη απωλεστική συμπίεση**

Στην μη απωλεστική συμπίεση (lossless compression) διατηρείται η ακεραιότητα των δεδομένων. Τα αρχικά δεδομένα και τα δεδομένα μετά τη συμπίεση και την αποσυμπίεση είναι ακριβώς τα ίδια, επειδή σε αυτές τις μεθόδους ο αλγόριθμος συμπίεσης και ο αλγόριθμος αποσυμπίεσης είναι ακριβώς αντίστροφοι. Κατά τη διαδικασία δε χάνεται κανένα μέρος των δεδομένων. Τα πλεονάζοντα δεδομένα κωδικοποιούνται κατά τη συμπίεση και αποκωδικοποιούνται κατά την αποσυμπίεση. Αυτοί οι μέθοδοι χρησιμοποιούνται όταν δεν πρέπει να χαθεί ούτε ένα μπιτ δεδομένων όπως στην περίπτωση ενός αρχείου κειμένου ή ενός προγράμματος. Τα γνωστά προγράμματα 7z, bz, zip και rar χρησιμοποιούν μη απωλεστική συμπίεση.

### **2.3.2 Απωλεστική συμπίεση**

Η απώλεια δεδομένων μπορεί να μην είναι αποδεκτή σε αρχεία κειμένου ή ενός

προγράμματα, είναι όμως αποδεκτή σε εικόνες και ταινίες. Ο λόγος είναι ότι τα μάτια μας και τα αφτιά μας δεν μπορούν να διακρίνουν πολύ μικρές αλλαγές. Για τέτοιες περιπτώσεις είναι κατάλληλες οι απωλεστικές μέθοδοι συμπίεσης (lossy data compression). Οι μέθοδοι αυτές είναι οικονομικότερες και απαιτούν λιγότερο χρόνο και χώρο όταν πρέπει να σταλούν εκατομμύρια μπιτ εικόνων και βίντεο το δευτερόλεπτο.

Χαρακτηριστικό παράδειγμα απωλεστικής συμπίεσης εικόνας είναι η μέθοδος JPEG (Joint Photographic Experts Group) για βίντεο η μέθοδος MPEG (Moving Pictures Experts Group) και για ήχο το πρότυπο mp3.

## 2.4 Αλγόριθμοι Συμπίεσης Δεδομένων

### 2.4.1 Αλγόριθμος RLE

Ο αλγόριθμος RLE (Run Length Encoding) αποτελεί μια από τις απλούστερες τεχνικές συμπίεσης. Σύμφωνα με τη μέθοδο αυτή, διατρέχεται η ακολουθία των bytes που αποτελούν τα δεδομένα προς συμπίεση και εντοπίζονται οι διαδοχικές επαναλήψεις του ίδιου χαρακτήρα. Στη συνέχεια αντικαθίστανται οι συνεχόμενες επαναλήψεις με το πλήθος τους, ακολουθούμενο από τον χαρακτήρα.



Σχήμα 2.1 : Παράδειγμα κωδικοποίησης RLE

Αν συμπίεσουμε την ακολουθία AAABBEBEEEAAOBD με τη μέθοδο RLE, τότε η συμπιεσμένη μορφή της ακολουθίας θα είναι 3A3B3E2A3D, που έχει μήκος 10 χαρακτήρες και όχι 14 όπως η αρχική. Παρατηρούμε ότι αποδίδει ικανοποιητικά για ακολουθίες δεδομένων που έχουν συχνές επαναλήψεις χαρακτήρων. Αν για παράδειγμα συμπίεσουμε την ακολουθία «ABCDE» μήκους 5 bytes με τον

αλγόριθμο RLE, η συμπιεσμένη μορφή της είναι «1A1B1C1D1E», με μήκος 10. Στην περίπτωση αυτή όχι μόνο δεν έγινε συμπίεση, αλλά διπλασιάστηκε το μήκος της ακολουθίας. Για να αποφευχθούν αυτές οι περιπτώσεις, ο RLE δε συμπιέζει τις ακολουθίες μη συνεχόμενων χαρακτήρων. Στην πράξη ο αλγόριθμος διατρέχει την ακολουθία των δεδομένων και αναζητά διαδοχικές επαναλήψεις του ίδιου χαρακτήρα. Όταν βρεθεί μεμονωμένος χαρακτήρας που δεν επαναλαμβάνεται, ή διπλή επανάληψη του, μένει ως έχει, καθώς δεν επωφελούμαστε από την κωδικοποίηση RLE που απαιτεί 2 χαρακτήρες. Όταν όμως βρεθούν διαδοχικές επαναλήψεις ενός χαρακτήρα, τις αντικαθιστά με ένα ζεύγος (*αριθμός, χαρακτήρας*), όπου «αριθμός» είναι το σύνολο των συνεχόμενων εμφανίσεων του χαρακτήρα.

#### 2.4.2 Αλγόριθμος του Huffman

Αυτή η τεχνική συμπίεσης παρουσιάστηκε από τον D. Huffman το 1952 και επιτυγχάνει μεγάλα ποσοστά συμπίεσης εκμεταλλευόμενη τη στατιστική ανάλυση της ακολουθίας δεδομένων. Η ανάλυση αυτή είναι πιο σύνθετη από την απλή ανίχνευση των συνεχόμενων επαναλήψεων που κάνει ο RLE. Σύμφωνα με την κωδικοποίηση του προτύπου ASCII, σε κάθε χαρακτήρα αντιστοιχεί ένα byte, που είναι ο κωδικός ASCII του χαρακτήρα αυτού. Αν η κωδικοποίηση γίνεται με τον νέο κώδικα Unicode, χρειάζονται 2 bytes για κάθε χαρακτήρα. Το συνολικό μέγεθος της ακολουθίας σε bytes είναι ίσο με το πλήθος των χαρακτήρων επί 1 ή 2 bytes ανά χαρακτήρα (ανάλογα με την κωδικοποίηση). Στην κωδικοποίηση Huffman, επιτυγχάνεται συμπίεση καθώς ο κώδικας που αντιστοιχεί σε κάθε χαρακτήρα δεν έχει σταθερό μήκος αλλά μεταβλητό. Η κωδικοποίηση για τους πιο συχνά εμφανιζόμενους χαρακτήρες είναι μικρότερη από ό,τι για τους λιγότερο συχνά εμφανιζόμενους, με αποτέλεσμα το συνολικό μέγεθος να είναι μικρότερο, καθώς για τους συχνότερους χρησιμοποιούνται μόνο 2-3 bits και όχι 8 ή 16.

Ο αλγόριθμος συμπίεσης Huffman αποτελείται από τα εξής στάδια:

Βήμα 1 : Μετράμε τη συχνότητα εμφάνισης του κάθε χαρακτήρα στην ακολουθία

Βήμα 2: Ταξινομούμε τις συχνότητες εμφάνισης σε μια λίστα κατά φθίνουσα τάξη



Βήμα 3: Κατασκευάζουμε ένα «δένδρο» για την κωδικοποίηση ξεκινώντας με τους συχνότερα εμφανιζόμενους χαρακτήρες

Βήμα 4: Αντιστοιχίζουμε τα δυαδικά '0' και '1' σε κάθε κόμβο του δέντρου: Αρχίζοντας από την ρίζα του δέντρου, προσθέτουμε '0' για κάθε αριστερό παιδί και '1' για κάθε δεξί. Οι χαρακτήρες που κωδικοποιούνται είναι τα φύλλα στην βάση του δέντρου. Αρχίζοντας από την κορυφή (ρίζα) του δέντρου, διατρέχοντας το μοναδικό μονοπάτι προς κάθε φύλλο, συλλέγουμε 0 ή 1 και ορίζουμε τον κώδικα για το χαρακτήρα που αντιστοιχεί στο φύλλο αυτό.

Έστω ότι θέλουμε να κωδικοποιήσουμε τη λέξη «ΑΛΛΟΣ».

Βήμα 1,2: Δημιουργούμε μια λίστα με τους χαρακτήρες ταξινομώντας τους με σειρά εμφάνισης. Έτσι έχουμε τη λίστα {«Λ»,«Α»,«Ο»,«Σ»} αφού το «Λ» εμφανίζεται δύο φορές και οι υπόλοιποι χαρακτήρες από μία (τους οποίους και τοποθετούμε στη λίστα με τη σειρά εμφάνισης τους).

Βήμα 3: Ξεκινώντας από τη ρίζα, δημιουργούμε δύο παιδιά: το «Λ» (το γράμμα που εμφανίζεται πρώτο στη λίστα) και έναν εσωτερικό κόμβο. Στη συνέχεια από τον κόμβο αυτό δημιουργούμε πάλι δύο παιδιά, ένα με τον επόμενο χαρακτήρα στη λίστα, δηλαδή το «Α» και έναν εσωτερικό κόμβο. Διαγράφουμε το «Λ» από τη λίστα. Επαναλαμβάνουμε την διαδικασία μέχρι να τελειώσουν όλοι οι χαρακτήρες της λίστας. Τελικά όλοι οι χαρακτήρες «Λ», «Α», «Ο», και «Σ» αποθηκεύονται σε φύλλα του δέντρου.

Βήμα 4: Αντιστοιχίζουμε τα δυαδικά 0 και 1, διατρέχοντας το μοναδικό μονοπάτι προς κάθε φύλλο-χαρακτήρα. Τελικά, η ακολουθία «ΑΛΛΟΣ» που είχε αρχικό μήκος  $5 \cdot 8 = 40$  bits, συμπιεσμένη γίνεται: 1000110111, συνολικού μήκους 11 bits, άρα περίπου 4 φορές μικρότερη. Βέβαια, στην πράξη, στο τέλος κάθε ακολουθίας προστίθεται και το σχήμα της κωδικοποίησης, ώστε να μπορεί να γίνει η αποσυμπίεση. Η ακολουθία κωδικοποίησης που παράγεται με τον προηγούμενο αλγόριθμο είναι μοναδική, ώστε να μπορεί να γίνεται η αποσυμπίεση. Έτσι, αν διαθέτουμε το σχήμα της κωδικοποίησης (πίνακας στο πλάι) και την κωδικοποιημένη ακολουθία, τότε μπορούμε να κάνουμε εύκολα την αποσυμπίεση κωδικοποιημένης ακολουθίας 1000110111. Διαβάζεται από αριστερά προς τα δεξιά το πρώτο 1. Δεν υπάρχει χαρακτήρας που να αντιστοιχεί σε αυτό. Διαβάζεται και το 0. Ο χαρακτήρας Α αντιστοιχεί στο 10, άρα αντικαθίσταται το

10 με αυτόν. Στη συνέχεια διαβάζουμε το 0, και ο μοναδικός χαρακτήρας που αρχίζει με 0 είναι ο Λ. Όμοια και για τα υπόλοιπα γράμματα.

### 2.4.3 Συμπίεση LZW

Πολλά αρχεία, ιδιαίτερα αρχεία χαρακτήρων, έχουν συμβολοσειρές που επαναλαμβάνονται συχνά, για παράδειγμα το άρθρο «τον». Για την αναπαράσταση της λέξης αυτής με χρήση του κώδικα ASCII χρειάζονται 5 bytes, αν συμπεριλάβουμε τα κενά πριν και μετά τη λέξη. Αν αντιστοιχίσουμε έναν αριθμό σε ολόκληρη τη λέξη (π.χ. το 256 που έχει 2 bytes μήκος), τότε με 2 bytes αντικαθιστούμε παντού τα 5 bytes του άρθρου «τον».

Αυτή είναι η προσέγγιση που ακολουθείται στον αλγόριθμο LZW, ο οποίος παρουσιάστηκε από τους J. Ziv και A. Lempel το 1977, και τροποποιήθηκε από τον T. Welch το 1984.

Ο αλγόριθμος ξεκινά φτιάχνοντας ένα λεξικό όλων των ακολουθιών χαρακτήρων (συμβολοσειρών) που εμφανίζονται στο υπό συμπίεση αρχείο. Στο λεξικό αυτό, που ονομάζεται και πίνακας αναφορών, αποθηκεύεται για κάθε συμβολοσειρά ο αριθμός των εμφανίσεων της και δίπλα ένας μοναδικός κωδικός που αντιστοιχεί σε αυτή. Στη φάση της συμπίεσης αντικαθιστούμε κάθε συμβολοσειρά με τον κωδικό της, ο οποίος έχει πολύ μικρότερο μέγεθος από αυτή. Επιπλέον, στο τέλος προσθέτουμε και το λεξικό, ώστε να μπορεί να γίνει η αποσυμπίεση. Για την αποσυμπίεση γίνεται κατ' αρχήν ανάγνωση του λεξικού και αντικατάσταση των κωδικών με τις αντίστοιχες αρχικές συμβολοσειρές. Ο αλγόριθμος LZW χρησιμοποιείται με επιτυχία για αρχεία με μεγάλη κανονικότητα και συχνές επαναλήψεις μεγάλων συμβολοσειρών. Τα αρχεία τύπου «zip» δημιουργούνται με συμπίεση βασισμένη στον αλγόριθμο LZW.

## 2.5 Κωδικοποίηση Δεδομένων

### 2.5.1 Ψηφιακή Παράσταση Ήχου

Η ψηφιοποίηση του αναλογικού ήχου γίνεται με την περιοδική λήψη δειγμάτων από το αναλογικό σήμα πολλές φορές το δευτερόλεπτο, η οποία λέγεται

*δειγματοληψία* (sampling). Ο αριθμός των δειγμάτων που παίρνουμε ανά δευτερόλεπτο, ώστε ο ψηφιακός ήχος να έχει την ίδια ποιότητα με τον αναλογικό, καθορίζεται από τη μέγιστη συχνότητα που εμφανίζει ο αναλογικός ήχος μας. Ο αριθμός αυτός πρέπει να είναι τουλάχιστο ίσος με το διπλάσιο της μέγιστης συχνότητας του ήχου, σύμφωνα με το θεώρημα του Shannon. Το ανθρώπινο αυτί αντιλαμβάνεται ήχους συχνοτήτων από 20Hz έως 20KHz. Έτσι, για να έχουμε πιστή αναπαραγωγή του αναλογικού ήχου χρειάζονται πάνω από 40.000 δείγματα ανά δευτερόλεπτο. Το πλήθος των δειγμάτων ή αλλιώς ο *ρυθμός* (rate) ή *συχνότητα δειγματοληψίας* (sampling frequency), δεν είναι το μόνο στοιχείο που καθορίζει την κωδικοποίηση του ψηφιακού ήχου. Το κάθε δείγμα αντικατοπτρίζει την ένταση του ήχου για τη στιγμή της δειγματοληψίας στην οποία αντιστοιχεί. Στον αναλογικό ήχο, οποιαδήποτε τιμή έντασης είναι επιτρεπτή. Στον ψηφιακό ήχο όμως, το *πλήθος των bits* (number of bits), που χρησιμοποιούμε για την αποθήκευση του κάθε δείγματος, καθορίζει και τον αριθμό των διαφορετικών τιμών εντάσεως που μπορεί να εμφανιστεί. Συνεπώς, κοντινές, αλλά διαφορετικές τιμές αναλογικής έντασης αντιστοιχούν στην ίδια ψηφιακή τιμή. Το φαινόμενο αυτό λέγεται *κβαντισμός* (quantization) των σταθμών έντασης του ήχου. Είναι φανερό ότι όσο πιο πολλά bits χρησιμοποιούμε για την αποθήκευση του κάθε δείγματος, τόσο πιο πιστή αναπαράσταση του αναλογικού ήχου πετυχαίνουμε. Αυτός ο τρόπος κωδικοποίησης λέγεται *παλμοκωδική κωδικοποίηση* (Pulse Code Modulation). Ήχος σε ποιότητα CD χρειάζεται δείγματα των 16 bits (2 bytes) και ρυθμό δειγματοληψίας 44,1 KHz (χιλιάδες δείγματα ανά δευτερόλεπτο) για κάθε κανάλι. Έτσι για ψηφιακό στερεοφωνικό ήχο (2 κανάλια ήχου) ποιότητας CD έχουμε 176.400 bytes/δευτερόλεπτο· ένα τραγούδι διάρκειας πέντε λεπτών καταλαμβάνει περίπου 50MB. Αν οι ανάγκες μας σε ποιότητα ήχου είναι μεγάλες, τότε η παλμοκωδική κωδικοποίηση έχει πολύ μεγάλες αποθηκευτικές απαιτήσεις. Σε συνήθεις εφαρμογές πολυμέσων δεν έχουμε απαιτήσεις στερεοφωνικού ήχου ή ακουστικών συχνοτήτων άνω των 4-5KHz (συχνότητα δειγματοληψίας 11KHz) και αφιερώνουμε 1 byte για κάθε δείγμα. Έτσι για το παραπάνω τραγούδι των 5 λεπτών χρειαζόμαστε το 1/16 των 50MB.

Το επόμενο βήμα για να μειώσουμε περισσότερο το μέγεθος των ηχητικών δεδομένων είναι να μην κωδικοποιούμε ξεχωριστά το κάθε δείγμα δίνοντας του π.χ. 8 bits. Είναι πιο οικονομικό να βρίσκουμε τη διαφορά του κάθε δείγματος με

το προηγούμενο και να κωδικοποιούμε αυτήν. Επειδή ο αριθμός των δειγμάτων ανά δευτερόλεπτο είναι μεγάλος, η πιθανότητα δυο διαδοχικά δείγματα να έχουν κοντινές τιμές έντασης είναι μεγάλη, άρα έχουμε να κωδικοποιήσουμε μικρές διαφορές και δεσμεύουμε λίγα bits, συνήθως 2-3 αντί των 8. Η κωδικοποίηση αυτή λέγεται διαφορική παλμοκωδική (Differential Pulse Code Modulation). Μια παραλλαγή της μεθόδου αυτής που επιτρέπει την ύπαρξη μεγαλύτερων διαφορών λέγεται προσαρμοστική-διαφορική παλμοκωδική (Adaptive Differential Pulse Code Modulation) και χρησιμοποιείται στα αρχεία .wav των Windows. Όλες αυτές οι τεχνικές αντιμετωπίζουν το ηχητικό σήμα σαν μαθηματική κυματομορφή, αγνοώντας τις ιδιαιτερότητες του ανθρώπινου αυτιού. Έτσι, όλες οι συχνότητες αντιμετωπίζονται ισοδύναμα. Στην πράξη όμως το ανθρώπινο αυτί είναι περισσότερο ευαίσθητο σε κάποια ζώνη συχνοτήτων από κάποια άλλη. Νέες τεχνικές κωδικοποίησης που βασίζονται σε αυτή την απλή παρατήρηση, επιτυγχάνουν πολύ μεγαλύτερα ποσοστά συμπίεσης, χωρίς να υπάρχει ιδιαίτερη απώλεια σε ποιότητα ήχου. Το πρότυπο MPEG-3 (Motion Picture Expert Group) πετυχαίνει συμπίεση έως και 12 φορές ως προς την αρχική μορφή χωρίς εμφανείς απώλειες στην ποιότητα.

Τα αρχεία .mp3 είναι κωδικοποιημένα με το πρότυπο αυτό. Αν όμως δεχθούμε συμβιβασμούς στην ακουστική ποιότητα, π.χ. για μονοφωνικό ήχο ποιότητας τηλεφωνικής μετάδοσης, τότε επιτυγχάνεται συμπίεση μέχρι και 100 φορές.

### 2.5.2 Άλλα πρότυπα

Παρά το γεγονός της η ριζική και ραγδαία εξέλιξη του MP3 το οποίο έχει γνωρίσει ευρεία αποδοχή μεταξύ χρηστών, εντούτοις, 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 MP3 Pro, το OGG και το AC-3 Dolby Digital. WMA (Windows Media Audio) Η μεταφορά, η αποθήκευση και η χρήση ακουστικού υλικού με βάση την μορφή απωλεστικών συμπιεσμένων ηχητικών αρχείων μέσω υπολογιστή, ωθεί τον μεγαλύτερο κατασκευαστή

λειτουργικών συστημάτων μα μην μείνει έξω από το παιχνίδι. Το πρότυπο Windows Media Audio (WMA) 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 επακόλουθο 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.

### **MP3 Pro**

Τον Ιανουάριο του 2001 στη CES, παρουσιάστηκε από την Coding Technologies η τεχνολογία Mp3 Pro, 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.

### **MP3 Surround**

Το 2004 το Ινστιτούτο Fraunhofer IIS παρουσίασε μία πολυκαναλική έκδοση του MP3, το MP3 Surround το οποίο βασίζεται στην τεχνολογία Binaural Cue Coding της Agere.

Η τεχνολογία αυτή, επιτρέπει την μείξη σημάτων από πολλά κανάλια σε δύο, με στόχο την δημιουργία ενός σήματος συμβατού με τον απλό MP3 codec, ενώ κωδικοποιεί μία σειρά από παραμέτρους που περιγράφουν πλήρως το ηχητικό πεδίο surround. Τέτοιες παράμετροι είναι, οι χρονικές διαφορές μεταξύ των καναλιών, οι διαφορές στάθμης μεταξύ των καναλιών και η συσχέτιση μεταξύ των καναλιών.

### **AAC (Advance Audio Coding)**

Το πρότυπο ACC 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 ACC, ως ένα πραγματικό αριστούργημα κωδικοποίησης 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 ACC 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. Αρκετά σημεία αυτό, οφείλουμε να αναφέρουμε ότι το πρότυπο ACC παρέχει καλύτερη απόδοση από το MP3, ενώ το 2003 η έκδοση του παρουσιάζεται συμβατή με τις προδιαγραφές του Mpeg ούτως ώστε το πρότυπο να αναφέρεται και ως Mpeg-4 ACC.

### **RA-Real Audio**

Το Real Audio ως κλειστό πρότυπο, δημιουργήθηκε και υποστηρίχθηκε από την εταιρία Real Network με σκοπό την αναπαραγωγή ήχων στο Διαδίκτυο χωρίς να προηγείται κατέβασμα των ήχων στο σκληρό δίσκο του υπολογιστή. Το πρότυπο Real Audio, είναι αρκετά δημοφιλές και αυτό εξαιτίας της ελεύθερης διάθεσης του λογισμικού ανάγνωσης των αρχείων ήχου τέτοιου τύπου υποστηρίζοντας, μεγάλη συμπίεση και κατακανόνα χαμηλή ποιότητα ήχου.

### **OGG Vorbis**

Ο codec Ogg Vorbis 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 ΜΟΟΤ για τον μετασχηματισμό του σήματος από το πεδίο του χρόνου στο πεδίο της συχνότητας, καθώς και μία εναλλακτική 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. 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.

### **AC3 Dolby Digital**

Ένα από τα πιο διαδεδομένα πρότυπα για τον ψηφιακό πολυκάναλο ήχο είναι το AC3, που εκτός σημαντικού απροόπτου άρχισε να γίνεται το διεθνές πρότυπο για την συμπίεση ηχητικών δεδομένων. Στο ψηφιακό σύστημα ήχου AC3, 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) δύο συνοδευτικά κανάλια περιβάλλοντος ήχου και
- (γ) ένα κανάλι για τις υπόλοιπες συχνότητες (σύστημα 3/2/.1).

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.

## ΚΕΦΑΛΑΙΟ 3ο : ΠΡΟΤΥΠΟ MPEG-1 (AUDIO LAYER 3) - MP3

Το MPEG-1 Audio Layer 3 (Ήχος τύπου MPEG-1 3ου επιπέδου), γνωστό και ως MP3 (προφέρεται έμ-πι-θρί), είναι μία δημοφιλής ψηφιακή κωδικοποίηση ήχου. Επίσης είναι μορφή αρχείου απωλεστικής συμπίεσης και αλγόριθμος, ο οποίος είναι σχεδιασμένος να μειώνει δραστικά το μέγεθος των δεδομένων που απαιτούνται για την αναπαραγωγή του ήχου, ο οποίος όμως ακούγεται σαν πιστή αναπαραγωγή του αρχικού ασυμπίεστου ήχου από τους περισσότερους ακροατές. Εφευρέθηκε από μία ομάδα Γερμανών μηχανικών του Ιδρύματος Fraunhofer, οι οποίοι εργάστηκαν στα πλαίσια του προγράμματος EUREKA 147 DAB το οποίο έκανε έρευνα πάνω στο ψηφιακό πρόγραμμα ραδιοφώνου, και τυποποιήθηκε με βάση το πρότυπο ISO/IEC το 1991. Το MP3 είναι μια μορφή ψηφιακού συμπιεσμένου αρχείου ειδικά για την αποθήκευση ήχου. Παρέχει τη δυνατότητα της αναπαράστασης ήχου κωδικοποιημένου με μορφή Pulse Code Modulation (PCM) (διαμόρφωση με βάση κωδικούς παλμών) δεσμεύοντας πολύ λιγότερο χώρο (για δεδομένα) από τις άμεσες μεθόδους. Αυτό γίνεται χρησιμοποιώντας ψυχοακουστικά μοντέλα για να απορρίψει τμήματα ή περιοχές του ηχητικού φάσματος, που δεν ακούει το ανθρώπινο αυτί, και καταγράφοντας την υπόλοιπη πληροφορία με αποτελεσματικό τρόπο. Παρόμοιες μέθοδοι χρησιμοποιούνται από το JPEG, μια μορφή συμπίεσης εικόνων με απώλειες.

### 3.1 Ανάπτυξη του προτύπου

Η κωδικοποίηση ήχου τύπου MPEG-1 2ου επιπέδου άρχισε ως σχέδιο DAB (Digital Audio Broadcast) το οποίο διηύθυνε ο Egon Meier-Engelen του Γερμανικού κέντρου αεροδιαστημικής στην Γερμανία. Το σχέδιο χρηματοδοτούσε η Ευρωπαϊκή Ένωση ως μέρος του ερευνητικού προγράμματος EUREKA, γνωστού και ως EU-147. Το πρόγραμμα αυτό διήρκεσε από το 1987 έως και το 1994. Το 1991 υπήρχαν δύο διαθέσιμες προτάσεις: Το Musicam (γνωστό και ως 2ο επίπεδο) και το ASPEC (Adaptive Spectral Perceptual Entropy Coding). Η μέθοδος Musicam, όπως είχε προταθεί από την Philips (Ολλανδική εταιρία), την CCETT (Γαλλική εταιρεία τηλεπικοινωνιών), και το Institut für Rundfunktechnik (Γερμανικό ινστιτούτο τηλεπικοινωνιών), επιλέχθηκε εξαιτίας της απλότητας και της ικανοποιητικής



αντιμετώπισης σφαλμάτων, καθώς και για το ότι απαιτείται σχετικά χαμηλή υπολογιστική ισχύς για την κωδικοποίηση συμπιεσμένου ήχου υψηλής ποιότητας. Η μορφοποίηση Musicam, η οποία βασιζόταν στην κωδικοποίηση υπο-συχνοτήτων του ήχου, ήταν το κλειδί στην δημιουργία της βάσης της μορφής συμπίεσης MPEG Audio (ρυθμοί δειγματοληψίας, δομή των frames ("πλαίσια"), κεφαλίδες, και αριθμός δειγμάτων ανά πλαίσιο). Η τεχνολογία και οι ιδέες ενσωματώθηκαν πλήρως στον ορισμό του προτύπου μορφής ISO MPEG Audio Layer I (πρώτου επιπέδου), στο Επίπεδο II και πιο πολύ στο Επίπεδο III (δηλαδή το MP3). Υπό την εποπτεία του καθηγητή Müssman (Πανεπιστήμιο του Ανόβερου) η επεξεργασία του προτύπου έγινε με ευθύνη του Leon van de Kerkhof (Επίπεδο I) και του Gerhard Stoll (Επίπεδο II). Μία ομάδα εργασίας αποτελούμενη από τους Leon Van de Kerkhof (Ολλανδία), Gerhard Stoll (Γερμανία), Leonardo Chiariglione (Ιταλία), Yves-François Deheroy (Γαλλία) και Karlheinz Brandenburg (Γερμανία), χρησιμοποιώντας ιδέες από το Musicam και το ASPEC και προσθέτοντας μερικές δικές τους, δημιούργησε το MP3, το οποίο σχεδιάστηκε για να επιτυγχάνει ποιότητα ήχου στα 128 Kbit/δευτ. όμοια με του MP2 στα 192 Kbit/δευτ., μειώνοντας δηλαδή τον όγκο των δεδομένων που απαιτούνταν και διατηρώντας σταθερή την ποιότητα του ήχου. Όλοι οι αλγόριθμοι εγκρίθηκαν το 1991 και οριστικοποιήθηκαν το 1992 ως μέρος του προτύπου MPEG-1, του πρώτου της σειράς προτύπων από την ομάδα MPEG από το οποίο προέκυψε το διεθνές πρότυπο ISO/IEC 11172-3, που δημοσιεύθηκε το 1993. Περαιτέρω εργασία πάνω στο MPEG Audio ολοκληρώθηκε το 1994 σαν μέρος της δεύτερης σειράς προτύπων MPEG, με το MPEG-2, πιο επίσημα γνωστό και ως διεθνές πρότυπο ISO/IEC 13818-3, που δημοσιεύθηκε για πρώτη φορά το 1995. Η αποδοτικότητα της συμπίεσης των κωδικοποιητών συχνά ορίζεται με βάση το bit rate (ρυθμό αποθήκευσης/ανάγνωσης bit ανά δευτερόλεπτο), επειδή η συμπίεση εξαρτάται από το εύρος των bits και την συχνότητα δειγματοληψίας του προς συμπίεση σήματος. Παρόλα αυτά, συχνά δημοσιεύονται ρυθμοί συμπίεσης που χρησιμοποιούν τις παραμέτρους της δειγματοληψίας του CD ως αναφορά (44.1 Khz, 2 κανάλια και 16 bit ανά κανάλι, ή 2x16 bit). Μερικές φορές χρησιμοποιούνται οι παράμετροι των ρυθμών δειγματοληψίας της ψηφιακής κασέτας (DAT, Digital Audio Tape), δηλαδή 48 Khz & 2x16 bit. Οι ρυθμοί συμπίεσης με αυτές τις παραμέτρους είναι υψηλότεροι, γεγονός που αποδεικνύει τον προβληματικό όρο "ρυθμός συμπίεσης" για τους κωδικοποιητές με απώλειες. Δηλαδή, ενώ χρησιμοποιούμε έναν αλγόριθμο για να μειώσουμε το μέγεθος ενός αρχείου ήχου "συμπιέζοντας" τα δεδομένα, τελικά

δημιουργούμε ένα μεγαλύτερο αρχείο χρησιμοποιώντας τις παραμέτρους αυτές. Ο Karlheinz Brandenburg χρησιμοποίησε το κομμάτι "Tom's Diner" από το CD της Σούζαν Βέγκα (Suzanne Vega) για να αξιολογήσει τον αλγόριθμο συμπίεσης του MP3. Το τραγούδι αυτό επιλέχθηκε εξαιτίας της απλότητας και της απαλής μουσικής που έχει, κάνοντας πιο απλό να ακουστούν ατέλειες της συμπιεσμένης μορφής κατά την αναπαραγωγή. Κάποιοι αστεειυόμενοι αναφέρουν την Suzanne Vega ως "μητέρα του MP3". Επίσης κάποια σημαντικά αποσπάσματα από μουσικά όργανα (τρίγωνο, ακορντεόν, μεταλλόφωνο) από το CD αναφοράς EBU V3/SQAM και χρησιμοποιήθηκαν από επαγγελματίες μηχανικούς ήχου για να αξιολογηθεί η υποκειμενική ποιότητα των μορφών ήχου του MPEG. Μία υλοποίηση ενός λογισμικού εξομοίωσης (για αναφορά) γραμμένου στη γλώσσα προγραμματισμού C γνωστό και ως ISO 11172-5, αναπτύχθηκε από τα μέλη της επιτροπής του ISO MPEG Audio ώστε να δημιουργηθούν αρχεία συμβατά με το MPEG Audio (επιπέδου 1, 2, και 3). Αυτό το πρόγραμμα κατάφερε να παρουσιάσει σε μερικά λειτουργικά συστήματα την πρώτη ζωντανή αποκωδικοποίηση συμπιεσμένου ήχου. (Δηλαδή το λογισμικό έκανε ανάγνωση του συμπιεσμένου αρχείου, και ταυτόχρονα αποσυμπίεση και αναπαραγωγή του. Ενώ νωρίτερα είχαν γίνει δοκιμές ώστε πρώτα να αποσυμπιέζεται ολόκληρο το αρχείο και στη συνέχεια να αναπαράγεται). Στην πραγματικότητα όμως το λογισμικό αυτό εξομείωνε τον τρόπο λειτουργίας του υλικού (δηλαδή των μικροτσιπ) το οποίο θα εκτελούσε αυτή την εργασία. Άλλες υλοποιήσεις άμεσης κωδικοποίησης από τους κωδικοποιητές του MPEG Audio ήταν διαθέσιμοι για χρήση στην ψηφιακή εκπομπή σήματος για καταναλωτικούς δέκτες.

Αργότερα, τον Ιούλιο του 1994, η ομάδα Fraunhofer κυκλοφόρησε το πρώτο λογισμικό που κωδικοποιούσε MP3 το οποίο ονομάστηκε L3enc. Η επέκταση αρχείου .mp3 επιλέχθηκε από την ομάδα Fraunhofer στις 14 Ιουλίου του 1995 (νωρίτερα τα αρχεία είχαν επέκταση .bit). Με το πρώτο λογισμικό που μπορούσε να αναπαράγει MP3 το Winplay 3 (το οποίο κυκλοφόρησε στις 9 Σεπτεμβρίου του 1995) πολλοί χρήστες είχαν την δυνατότητα να κωδικοποιούν και να αναπαράγουν MP3 στους υπολογιστές τους. Εξ αιτίας των σχετικά μικρών σκληρών δίσκων (περίπου 500MB) που υπήρχαν εκείνη την εποχή στους προσωπικούς υπολογιστές η τεχνολογία αυτή ήταν απαραίτητη για να αποθηκευθούν κομμάτια μουσικής με φυσικό ήχο και φωνή σε έναν υπολογιστή. (Σε αντίθεση με αρχεία τύπου tracker και midi τα οποία εκτελούσαν μουσικά κομμάτια χρησιμοποιώντας μονό δείγματα ήχου από μουσικά

όργανα) Τον Οκτώβριο του 1993, αρχεία τύπου MP2 εμφανίστηκαν στο Ιντερνετ και συχνά "έπαιζαν" χρησιμοποιώντας τον Xing MPEG Audio Player, και αργότερα με ένα πρόγραμμα για Unix που είχε δημιουργήσει το Tobias Banding και λεγόταν MAPlay, το οποίο κυκλοφόρησε για πρώτη φορά στις 22 Φεβρουαρίου του 1994. (Το MAPlay μεταφέρθηκε και στο λειτουργικό σύστημα των Windows). Αρχικά το μόνο πρόγραμμα κωδικοποίησης MP2 ήταν το Xing Encoder, μαζί με το πρόγραμμα CDDA2WAV, που επεξεργάζεται μουσικά CD και μετατρέπει τα κομμάτια τους σε αρχείο κυματομορφής (waveform). Η ιστοσελίδα IUMA (Internet Underground Music Archive, Αρχείο "υπόγειας" μουσικής στο Διαδίκτυο) γενικά αναγνωρίζεται ως η αρχή της μουσικής επανάστασης στο Διαδίκτυο. Η IUMA ήταν η πρώτη μουσική ιστοσελίδα υψηλής πιστότητας και φιλοξενούσε χιλιάδες εγκεκριμένα μουσικά κομμάτια πριν το MP3 ή ο Παγκόσμιος Ιστός γίνουν δημοφιλή. Από το πρώτο μισό του 1995 έως και τα τέλη της δεκαετίας του 1990, αρχεία MP3 άρχισαν κυκλοφορούν ευρέως στο Διαδίκτυο. Η επιτυχία του MP3 οφειλόταν κυρίως στην επιτυχία εταιριών λογισμικού και των προγραμμάτων τους όπως το Winamp της Nullsoft (που κυκλοφόρησε το 1997), το mpg123 και το Napster (που κυκλοφόρησε το 1999). Αυτά τα προγράμματα έκαναν για τον απλό χρήστη πολύ εύκολη την διαδικασία της αναπαραγωγής, δημιουργίας, μοιράσματος και της συλλογής MP3 αρχείων.

Αντιπαραθέσεις που είχαν να κάνουν με την ανταλλαγή αρχείων MP3 μέσω ισότιμων δικτύων (peer-to-peer) είναι συνηθισμένες τα τελευταία χρόνια, κυρίως επειδή η υψηλή συμπίεση που επιτυγχάνεται επιτρέπει το μοίρασμα και ανταλλαγή αρχείων που σε άλλη περίπτωση θα ήταν πολύ μεγάλα και ογκώδη (σε χώρο που απαιτούν ως δεδομένα) για να μοιραστούν εύκολα. Κάποιες μεγάλες δισκογραφικές εταιρείες αντέδρασαν υποβάλλοντας μηνύσεις εναντίον της εταιρείας Napster, λόγω της μεγάλης διάδοσης των MP3 μέσω του Διαδικτύου, για να προστατεύσουν τα πνευματικά τους δικαιώματα. (Δείτε επίσης και τον όρο πνευματική ιδιοκτησία)

Οι εμπορικές online υπηρεσίες πώλησης μουσικής (όπως το Online κατάστημα της Apple iTunes) συνήθως προτιμούν άλλους τύπους αρχείων, οι οποίοι υποστηρίζουν το DRM (Digital Rights Management, διαχείριση ψηφιακών δικαιωμάτων), για να ελέγξουν και να περιορίσουν την χρήση της ψηφιακής μουσικής. Η χρήση αυτών των τύπων αρχείων που υποστηρίζουν το DRM είναι μία προσπάθεια να αποτραπεί η παραβίαση υλικού το οποίου τα δικαιώματά του είναι προστατευμένα, αλλά

υπάρχουν διάφοροι μέθοδοι για την παραβίαση των περισσότερων μεθόδων προστασίας. Τέτοιες μέθοδοι είναι παράνομες σε πολλές χώρες. Μερικές, όμως, online υπηρεσίες πώλησης μουσικής (όπως το eMusic και το DJTunes.com) χρησιμοποιούν τον τύπο MP3, κυρίως λόγω της συμβατότητας με τα φορητά συστήματα αναπαραγωγής μουσικής (τα λεγόμενα MP3 players).

### 3.2 Τεχνικά χαρακτηριστικά του προτύπου

Το πρότυπο MPEG-1 δεν συμπεριλαμβάνει ακριβείς προδιαγραφές για έναν κωδικοποιητή MP3. Από την άλλη μεριά, ο αλγόριθμος και η μορφή του αρχείου ορίζονται ικανοποιητικά. Όσοι υλοποιούν το πρότυπο θεωρείται ότι θα επινοήσουν δικούς τους αλγόριθμους, ικανούς να αφαιρέσουν τμήματα της πληροφορίας στο αρχικό ηχητικό κομμάτι. Ως αποτέλεσμα, υπάρχουν πολλοί διαφορετικοί κωδικοποιητές MP3, ο καθένας από τους οποίους δημιουργεί αρχεία διαφορετικής ποιότητας. Συγκριτικές δοκιμές είναι ευρέως διαθέσιμες, ώστε να είναι εύκολο για έναν πιθανό χρήστη ενός κωδικοποιητή να αναζητήσει την καλύτερη επιλογή. Πρέπει εδώ να σημειωθεί ότι ένας κωδικοποιητής που έχει δυνατότητα να δημιουργεί αρχεία σε υψηλότερα bit rates (βλ. παρακάτω) (όπως ο LAME, που είναι ευρέως διαδεδομένος για την κωδικοποίηση σε υψηλά bit rates) δεν είναι απαραίτητα τόσο καλός στην κωδικοποίηση με χαμηλότερους ρυθμούς bit rate.

#### 3.2.1 Αποκωδικοποίηση Ήχου

Η αποκωδικοποίηση από την άλλη μεριά, είναι ένα προσεκτικά σχεδιασμένο πρότυπο. Οι περισσότεροι αποκωδικοποιητές είναι "σύμμορφοι με τη ροή των bits" (bitstream compliant), που σημαίνει ότι το αποσυμπιεσμένο αποτέλεσμα που παράγουν από κάποιο αρχείο MP3 θα είναι το ίδιο (μέσα σε ένα ανεκτό βαθμό στρογγυλοποίησης) όπως το αποτέλεσμα που ορίζεται μαθηματικά από το έγγραφο του προτύπου ISO/IEC. Το αρχείο MP3 έχει μία σταθερή μορφή που αποτελείται από 384, 576 ή 1152 δείγματα (ανάλογα με την έκδοση και το επίπεδο του MPEG) και όλα τα "πλαίσια", έχουν σχετική πληροφορία στην κεφαλίδα (32 bit) και την

υπόλοιπη πληροφορία (9, 17, ή 32 bytes, ανάλογα με την έκδοση του MPEG και αν είναι στερεοφωνικός ή μονοφωνικός ο ήχος). Η πληροφορία της κεφαλίδας και του υπόλοιπου μέρους βοηθά τον αποκωδικοποιητή να αποκωδικοποιήσει σωστά τα δεδομένα. Για αυτό τον λόγο οι αποκωδικοποιητές συγκρίνονται συνήθως με βάση την υπολογιστική τους απόδοση (δηλαδή πόση μνήμη και χρόνο από τον επεξεργαστή ενός υπολογιστή απαιτούν για την διαδικασία της αποκωδικοποίησης).

### 3.2.2 Ρυθμός Bit

Ο ρυθμός bit (bit rate) είναι κυμαινόμενος για τα αρχεία MP3. Ο γενικός κανόνας είναι ότι όσο μεγαλύτερο ρυθμό Bit έχει ένα αρχείο τόσο περισσότερη πληροφορία περιλαμβάνεται από τον αρχικό ήχο, και έτσι είναι ποιοτικότερο το αποτέλεσμα κατά την αναπαραγωγή. Στις πρώτες μέρες της κωδικοποίησης των MP3 χρησιμοποιούνταν σταθερός ρυθμός bit για όλο το αρχείο. Οι διαθέσιμοι ρυθμοί Bit για το MPEG-1 επιπέδου 3 είναι 32, 40, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256 και 320 kbit/s, και οι διαθέσιμες συχνότητες δειγματοληψίας είναι 32, 44.1 και 48 Khz. Η πιο συνηθισμένη είναι αυτή των 44.1Khz (και κατα σύμπτωση είναι ίδια με αυτή του CD), ενώ τα 128Kbit έχει γίνει ο συνηθισμένος ρυθμός bit για ένα "αρκετά καλό" αποτέλεσμα. Αν και τα 192Kbit άρχισαν να γίνονται όλο και πιο δημοφιλή στα δίκτυα ανταλλαγής αρχείων (peer-to-peer), κυρίως λόγω της μεγαλύτερης διαθεσιμότητας σε ευρυζωνικές ταχύτητες Ιντερνετ. Το MPEG-2 και το ανεπίσημο MPEG-2.5 συμπεριλαμβάνουν πρόσθετους ρυθμούς bit 8, 16, 24, 32, 40, 48, 56, 64, 80, 96, 112, 128, 144, 160 kbit/δευτ και προσφέρουν και χαμηλότερες συχνότητες δειγματοληψίας (8, 11.025, 12, 16, 22.05 και 24 kHz)

Επίσης είναι δυνατό να χρησιμοποιηθούν κυμαινόμενοι ρυθμοί bit (Variable bit rates ή VBR). Τα αρχεία MP3 χωρίζονται σε "πλαίσια", κάθε ένα από τα οποία έχει το δικό του ρυθμό bit, και έτσι είναι δυνατό να αλλαχθεί δυναμικά ο ρυθμός καθώς το αρχείο κωδικοποιείται. Αυτή η τεχνική κάνει δυνατή τη χρήση περισσότερων bit για κομμάτια του ήχου με υψηλότερη δυναμική (περισσότερη κίνηση στον ήχο), και λιγότερα bit σε σημεία με μικρότερη δυναμική, βελτιώνοντας περισσότερο την ποιότητα και μειώνοντας τον χώρο που απαιτείται για την αποθήκευσή τους. Για παράδειγμα, ένα μέρος που αποτελείται από καθαρούς τόνους μπορεί να κωδικοποιηθεί στα 48Kbit/δευτ, καταλαμβάνοντας λιγότερο χώρο χωρίς κάποια εμφανή διαφορά, ενώ ένα μέρος που παίζεται από μια πλήρη συμφωνική ορχήστρα

κωδικοποιείται στα 224Kbit/δευτ για να το αναπαραστήσει με μεγαλύτερη πιστότητα. Αν και αρχικά δεν γινόταν αυτό, πολλοί κωδικοποιητές χρησιμοποιούν αυτή την τεχνική σε μεγαλύτερο ή μικρότερο βαθμό.

Ρυθμοί bit εκτός του τύπου μέχρι και 640Kbit/δευτ μπορούν να επιτευχθούν με τον κωδικοποιητή LAME (χρησιμοποιώντας την επιλογή freeformat, δηλαδή, ελεύθερη μορφοποίηση) αλλά λιγότερα προγράμματα αναπαραγωγής μπορούν να παίξουν αυτά τα αρχεία. Ο Gabriel Bouvigne, ένας βασικός προγραμματιστής στην ανάπτυξη του LAME, παρείχε την παρακάτω πληροφορία σχετικά με το freeformat.

Το freeformat είναι συμβατό με το πρότυπο MP3. Οι αποκωδικοποιητές απαιτείται να μπορούν αν αποκωδικοποιήσουν μέχρι και τα 320Kb/δευτ, αλλά η αποκωδικοποίηση υψηλότερων ρυθμών bit δεν είναι απαραίτητη. Πρακτικά, αυτό σημαίνει ότι λίγοι αποκωδικοποιητές υποστηρίζουν παραπάνω από 320Kbit/δευτ.

### 3.2.3 Ποιότητα ήχου

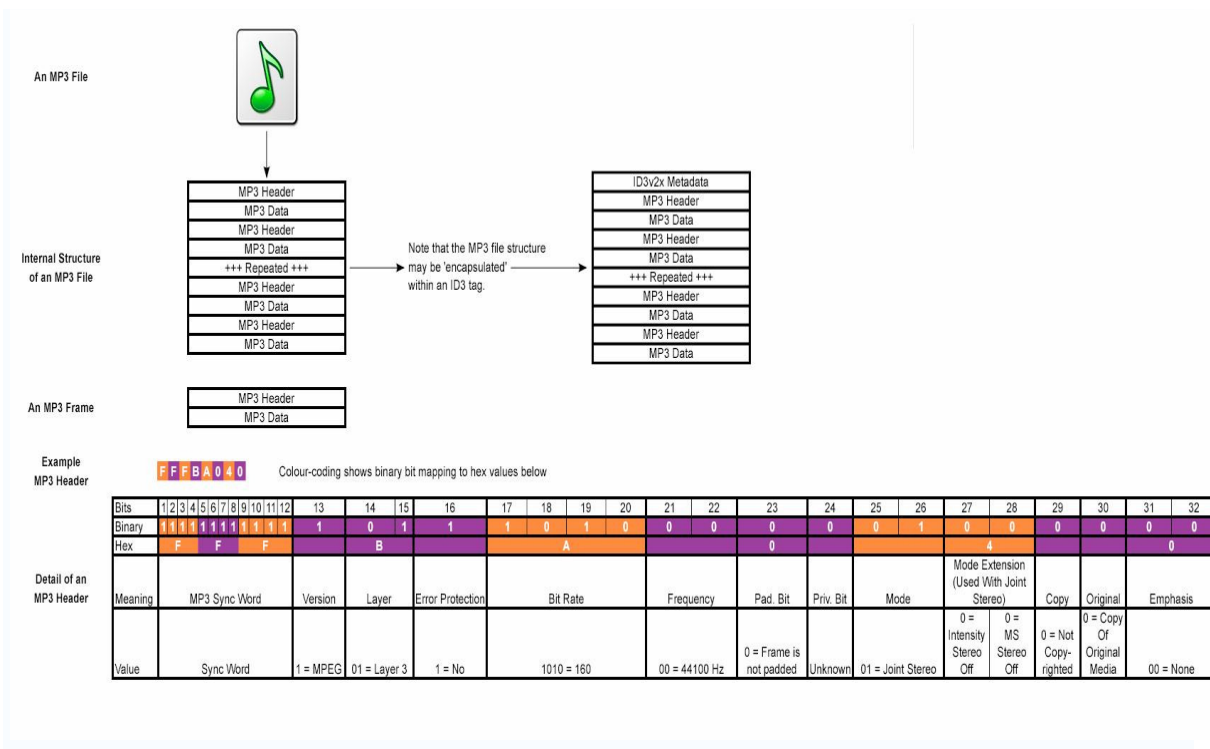
Επειδή το MP3 είναι μορφή αρχείου απωλεστικής συμπίεσης, είναι δυνατόν να παρέχει έναν αριθμό από διαφορετικές επιλογές για τους ρυθμούς bit που χρησιμοποιεί, δηλαδή τον αριθμό των bit κωδικοποιημένης πληροφορίας τα οποία αναπαριστούν κάθε δευτερόλεπτο ήχου. Τυπικά, οι ρυθμοί αυτοί είναι μεταξύ των 128 και 320 kbit/δευτ. Αντίθετα ο ασυμπίεστος ήχος όπως αποθηκεύεται σε έναν ψηφιακό δίσκο (CD) έχει ρυθμό bit 1411.2 kb/ δευτ (16bit ανα δείγμα X 44100 δείγματα το δευτερόλεπτο X 2 κανάλια). Αρχεία MP3 τα οποία κωδικοποιήθηκαν με μικρότερο ρυθμό bit σε γενικές γραμμές θα αναπαράγουν τον ήχο σε χαμηλότερη ποιότητα. Με πολύ χαμηλό ρυθμό bit, "Προϊόντα συμπίεσης" (δηλ. ήχοι που δεν υπήρχαν στον αρχικό ήχο) μπορεί να ακούγονται στην αναπαραγωγή. Ένα καλό παράδειγμα των προϊόντων συμπίεσης είναι τα χειροκροτήματα: είναι δύσκολο να συμπίεστούν γιατί είναι εντελώς τυχαία και έχουν οξείς ήχους. Για αυτό τα προϊόντα συμπίεσης μπορεί να ακουστούν σαν κουδουνίσματα ή ηχώ που προηγείται του κανονικού ήχου. Η ποιότητα επίσης έχει εξαρτάται και από την ποιότητα του προγράμματος κωδικοποίησης και την δυσκολία της μετατροπής του σήματος που κωδικοποιείται (συμπιέζεται). Επειδή το πρότυπο του MP3 δίνει αρκετή ελευθερία

στους αλγόριθμους κωδικοποίησης, διαφορετικοί κωδικοποιητές μπορεί να αποφέρουν διαφορετικές ποιότητες, ακόμα και έχοντας παρόμοιους ρυθμούς bit. Για παράδειγμα, σε μία δημόσια δοκιμή ακρόασης τον Ιούλιο του 2003 που έγινε σε δύο κωδικοποιητές στα 128Kbps ο ένας πέτυχε 3,66 βαθμούς σε κλίμακα 1 έως 5 ενώ ο άλλος μόλις 2,22.

Η ποιότητα είναι άμεσα συσχετιζόμενη με την επιλογή κωδικοποιητή και των παραμέτρων του. Ενώ με τους παλαιότερους κωδικοποιητές στα 128kbps η ποιότητα ήταν ανάμεσα στο ενοχλητικό και το ανεκτό, οι πιο καινούργιοι καταφέρουν να παρέχουν καλύτερη ποιότητα σε αυτούς τους ρυθμούς Bit, στατιστικά όχι με μεγάλες διαφορές από την ποιότητα που προσφέρει το AAC (τον διάδοχο του MP3 από τεχνικής απόψεως). Το 1998 όμως το MP3 στα 128Kbps παρείχε ποιότητα ανάλογη του AAC στα 96Kbps και του MP2 στα 192Kbps. Το όριο στο οποίο το MP3 ακούγεται χωρίς να ξεχωρίζει από τον αρχικό ήχο, μπορεί να εκτιμηθεί περίπου στα 128Kbps χρησιμοποιώντας καλούς κωδικοποιητές σε ένα τυπικό κομμάτι μουσικής. Αυτό αποδεικνύεται από την καλή του απόδοση στην παραπάνω δοκιμή, αλλά πιθανώς κάποια συγκεκριμένα πιο "δύσκολα" κομμάτια να απαιτούν 192Kbps ή και περισσότερα. Όπως και με όλες τις μορφές απωλεστικής συμπίεσης, κάποια δείγματα δεν είναι δυνατόν να κωδικοποιηθούν ώστε να μην γίνονται αντιληπτά από όλους τους χρήστες. Μια εναλλακτική απεικόνιση της κωδικοποίησης είναι η χρήση του VBR (κυμαινόμενου ρυθμού bit). Αυτό στοχεύει σε μια σταθερή ποιότητα ήχου και μεταβάλλει ανάλογα τον ρυθμό Bit. Οι χρήστες που γνωρίζουν ότι κάποια συγκεκριμένη "ρύθμιση ποιότητας" είναι "διαφανής" για τα αυτιά τους (δηλαδή δεν μπορούν να ξεχωρίσουν τη διαφορά μεταξύ αρχικού ήχου και MP3) μπορούν να χρησιμοποιούν αυτή τη ρύθμιση σε όλα τα κομμάτια της μουσικής τους και δεν υπάρχει λόγος να ανησυχούν ότι χρειάζεται να κάνουν δοκιμές σε κάθε κομμάτι για να επιλέξουν τις κατάλληλες ρυθμίσεις. Σε χαμηλότερους ρυθμούς Bit η ποιότητα του MP3 πέφτει απότομα και είναι μακράν πίσω από την απόδοση της ποιότητας του AAC στα 32Kbps όπως φάνηκε σε μία ακουστική δοκιμή (06/2004) Είναι επίσης σημαντικό να σημειωθεί ότι η αντίληψη της ποιότητας ενός κωδικοποιημένου ήχου μπορεί να επηρεαστεί από το περιβάλλον της ακρόασης (θόρυβος στο περιβάλλον), την προσοχή του ακροατή, και την εκπαίδευσή του (να έχει δηλαδή "ευαίσθητο αυτί" ώστε να κατανοεί τις διαφορές).

### 3.3 Η δομή ενός αρχείου MP3

Ένα αρχείο MP3 αποτελείται από πολλάπλά πλαίσια (frames) τα οποία αποτελούν την κεφαλίδα του αρχείου και τα δεδομένα. Αυτή η αλληλουχία από πλαίσια ονομάζεται στοιχειώδης ροή (elementary stream). Τα πλαίσια είναι αυτόνομα στοιχεία. Κάποιος θα μπορούσε να αφαιρέσει κάποια πλαίσια από το αρχείο και ένα πρόγραμμα αναπαραγωγής MP3 θα μπορούσε να το "παίζει". Τα δεδομένα του MP3 είναι το πραγματικό ωφέλιμο μέρος. Στο διάγραμμα φαίνεται ότι η κεφαλίδα του MP3 αποτελείται από μία "λέξη" συγχρονισμού η οποία χρησιμοποιείται για να προσδιορίσει την έναρξη ενός έγκυρου πλαισίου. Ακολουθεί ένα bit που επισημαίνει ότι αυτό είναι το πρότυπο MPEG και άλλα δύο Bit που επισημαίνουν ότι χρησιμοποιείται το επίπεδο 3, δηλαδή το MPEG-1 επιπέδου 3 ή πιο απλά MP3. Μετά από αυτό οι τιμές θα διαφοροποιούνται ανάλογα με το αρχείο MP3. Το πρότυπο ISO/IEC 11172-3 ορίζει το εύρος των τιμών για κάθε ενότητα της κεφαλίδας μαζί με την προδιαγραφή της κεφαλίδας. Τα περισσότερα αρχεία MP3 σήμερα περιέχουν μεταδεδομένα (metadata) τύπου ID3 που προηγούνται ή ακολουθούν τα πλαίσια του MP3. Στο σχήμα 3.1 παρουσιάζεται η δομή ενός τυπικού MP3 αρχείου :





## Σχήμα 3.1 : Δομή MP3 αρχείου

### 3.3.1 Σχεδιαστικοί περιορισμοί

Υπάρχουν διάφοροι περιορισμοί οι οποίοι στους οποίους υπόκειται το αρχείο MP3 και δεν μπορούν να ξεπεραστούν από κανέναν κωδικοποιητή. Νεότερα πρότυπα κωδικοποίησης όπως το Vorbis και το AAC δεν έχουν πια αυτούς τους περιορισμούς. Με τεχνικούς όρους, το MP3 περιορίζεται με τους παρακάτω τρόπους:

- Ο ρυθμός Bit περιορίζεται στο μέγιστο στα 320Kb/δευτ (παρόλο που κάποιοι κωδικοποιητές μπορούν να δημιουργήσουν αρχεία με υψηλότερους ρυθμούς υπάρχει πολύ μικρή ή καθόλου υποστήριξη για αυτά τα αρχεία).
- Η ανάλυση του χρόνου μπορεί να είναι πολύ χαμηλή για κάποια σήματα με υψηλές συχνότητες για μικρό διάστημα, προκαλώντας προβλήματα σε κάποιους κρουστικούς ήχους.
- Η ανάλυση των συχνοτήτων υπόκειται σε περιορισμούς και αυτό περιορίζει την αποτελεσματικότητα της κωδικοποίησης
- Για συχνότητες πέραν των 15.5/15.8 Khz δεν υπάρχει συντελεστής κλίμακας.
- Το joint stereo επεξεργάζεται πλαίσιο προς πλαίσιο
- Η καθυστέρηση της κωδικοποίησης/αποκωδικοποίησης δεν ορίζεται, το οποίο σημαίνει ότι υπάρχει έλλειψη επίσημης πρόβλεψης για αναπαραγωγή κομματιών χωρίς κενά ανάμεσά τους. Παρόλα αυτά κάποιοι κωδικοποιητές όπως ο LAME μπορούν να προσθέσουν επιπλέον "μεταδεδομένα" που επιτρέπουν στα προγράμματα αναπαραγωγής να γνωρίζουν για αυτό και να προσφέρουν αναπαραγωγή χωρίς κενά.

Παρόλα αυτά, ένας καλά ρυθμισμένος κωδικοποιητής μπορεί να αποδώσει ανταγωνιστικά ακόμα και με αυτούς τους περιορισμούς.

### 3.3.2 Ετικέτες ID3

Μία ετικέτα μέσα σε ένα συμπιεσμένο αρχείο ήχου, είναι μια ενότητα του αρχείου που περιέχει μεταδεδομένα (metadata) όπως ο τίτλος, ο καλλιτέχνης, το άλμπουμ, ο αριθμός του τραγουδιού και άλλες πληροφορίες που σχετίζονται με το τραγούδι. Μέχρι το 2006, οι πιο διαδεδομένοι τύποι μορφών ετικετών είναι οι ID3v1 και ID3v2, και πρόσφατα παρουσιάστηκε το APEv2. Το APEv2 αρχικά είχε αναπτυχθεί για το

αρχείο τύπου MPC. Η ετικέτα APEv2 μπορεί να συνυπάρχει μαζί με τις ετικέτες ID3 στο ίδιο αρχείο, αλλά μπορεί επίσης να χρησιμοποιηθεί και αυτόνομα. Η δυνατότητα επεξεργασίας των ετικετών στα αρχεία MP3 είναι συχνά ενσωματωμένη στα προγράμματα αναπαραγωγής και επεξεργασίας MP3, αλλά υπάρχουν και προγράμματα ειδικά για την επεξεργασία των ετικετών, με περισσότερες δυνατότητες, όπως η μαζική αλλαγή ετικετών σε πολλά αρχεία, ή η αντιγραφή μέρος του ονόματος ενός αρχείου σε κάποια ετικέτα και αντίστροφα.

### 3.3.3 Εξομάλυνση Έντασης ήχου

Επειδή η ψηφιακοί δίσκοι (CD) και άλλες πηγές ηχογραφούνται και παράγονται σε διαφορετικές εντάσεις ήχου, είναι χρήσιμο να αποθηκεύεται η πληροφορία της έντασης του ήχου ενός αρχείου στην ετικέτα ώστε κατά την διάρκεια της αναπαραγωγής, η ένταση του ήχου να μπορεί να ρυθμίζεται δυναμικά.

Έχουν προταθεί μερικά πρότυπα για την κωδικοποίηση της αύξησης του ήχου ενός MP3. Η ιδέα είναι να εξομαλυνθεί η μέση ένταση ενός αρχείου ήχου (όχι οι απότομες αλλαγές της), έτσι ώστε η ένταση να μην αλλάζει μεταξύ των συνεχόμενων κομματιών. Αυτό δεν πρέπει να συγχέεται με την δυναμική συμπίεση ορίων (DRC, dynamic range compression) η οποία είναι μία μορφή εξομάλυνσης ήχου που χρησιμοποιείται κατά την διάρκεια παραγωγής μουσικής. Η πιο δημοφιλής και διαδεδομένη λύση για την αποθήκευση της αύξησης του ήχου κατά την αναπαραγωγή είναι απλά γνωστή ως "Replay Gain" (αύξηση ήχου στην αναπαραγωγή). Συνήθως, η μέση ένταση και η πληροφορία αύξησης ή μείωσης της για το κομμάτι του ήχου αποθηκεύεται στην ετικέτα που περιέχει τα μεταδεδομένα (metadata tag). Κάποιος χρήστης μπορεί κατεβάσει λογισμικό από το ιντερνέτ για να κάνει αλλαγές του είδους.

## 3.4 Αδειοδότηση και Ευρεσιτεχνίες του MP3

Ένας μεγάλος αριθμός οργανισμών διεκδίκησε την ιδιοκτησία των ευρεσιτεχνιών που απαιτούνται για την υλοποίηση του προτύπου MP3 (κωδικοποίηση ή/και αποκωδικοποίηση). Αυτές οι διεκδικήσεις οδήγησαν στη λήψη ενός αριθμού νομικών μέτρων και νομικών απειλών από διάφορες πηγές, έχοντας ως αποτέλεσμα την

αβεβαιότητα σχετικά με το τί είναι απαραίτητο για να παραχθούν προϊόντα που υποστηρίζουν το MP3 και να είναι νόμιμα, σε χώρες που επιτρέπουν τις ευρεσιτεχνίες λογισμικού. Οι διάφορες ευρεσιτεχνίες, που διεκδικούνται από τους (πολλούς) κατόχους τους, έχουν και διαφορετικές ημερομηνίες λήξης, οι οποίες βρίσκονται ανάμεσα στο 2007 και το 2017 στις ΗΠΑ. Όμως, οι ευρεσιτεχνίες στις ΗΠΑ μπορούν να διαρκέσουν μέχρι 20 χρόνια, και οι προδιαγραφές του MP3 παρουσιάστηκαν το 1991, οπότε αν τα δικαστήρια των ΗΠΑ εφήρμοζαν το νόμο, καμία ευρεσιτεχνία δεν θα μπορούσε να υφίσταται για το MP3 πέρα από το 2011. Στις ΗΠΑ οποιαδήποτε ευρεσιτεχνία διεκδικεί την κάλυψη των βασικών χαρακτηριστικών του MP3 μετά το 2012 θα πρέπει (σύμφωνα με το νόμο) να απορριφθεί ως μη ισχύουσα ευρεσιτεχνία, εξ αιτίας του ότι ήδη έχουν εκδοθεί οι προδιαγραφές περισσότερο από ένα χρόνο από την κατάθεση της ευρεσιτεχνίας. Αν έχει εκδοθεί ακόμα νωρίτερα (όπως για παράδειγμα σε δημόσια προσχέδια), η τελευταία ημερομηνία θα είναι ακόμα νωρίτερα. Παρόλα αυτά, είναι ασαφές αν τα δικαστήρια των ΗΠΑ θα το επιβάλουν αυτό. Παρόμοια είναι η κατάσταση και σε άλλες χώρες που επιτρέπουν ευρεσιτεχνίες λογισμικού. Η εταιρεία Thomson Consumer Electronics διεκδικεί την αδειοδότηση των ευρεσιτεχνιών του MPEG-1/2 Επίπεδου 3 σε πολλές χώρες, συμπεριλαμβανομένων και των ΗΠΑ, Ιαπωνίας, Καναδά και των χωρών της Ευρωπαϊκής ένωσης. Η Thompson επιβάλλει ενεργά αυτές τις ευρεσιτεχνίες. Λόγω των διαφορετικών πρακτικών στις χώρες της Ευρώπης, όταν κατοχυρώνουν ευρεσιτεχνίες για εφευρέσεις που υλοποιούνται με υπολογιστές με βάση την Ευρωπαϊκή σύμβαση ευρεσιτεχνιών, είναι ασαφές αν τα Εθνικά δικαστήρια μπορούν να υπεραμυνθούν αυτών των ευρεσιτεχνιών. Για πρόσφατες πληροφορίες σχετικά με το ίδρυμα Fraunhofer και τις ευρεσιτεχνίες της Thomson όπως και για τους όρους αδειοδότησης και τα τέλη χρήσης, δείτε την ιστοσελίδα τους [mp3licensing.com](http://mp3licensing.com). Η άδειες για το MP3 απέφεραν έσοδα €100 εκ. για το Ινστιτούτο Fraunhofer το 2005. Τον Σεπτέμβριο του 1998 το Ίδρυμα Fraunhofer έστειλε μία επιστολή σε αρκετούς παραγωγούς λογισμικού για MP3 δηλώνοντας ότι απαιτείται άδεια για την "διανομή ή/και πώληση κωδικοποιητών ή αποκωδικοποιητών". Η επιστολή ανέφερε ότι "μη αδειοδοτημένα προϊόντα παραβιάζουν τα δικαιώματα των ευρεσιτεχνιών του Ιδρύματος και της Thomson. Για την παραγωγή, πώληση ή/και διανομή προϊόντων που χρησιμοποιούν το πρότυπο MPEG Επίπεδο-3, απαιτείται να αποκτήσετε άδεια για τη χρήση αυτών των ευρεσιτεχνιών από εμάς". Τα θέματα αυτά σχετικά με τις ευρεσιτεχνίες καθυστέρησαν αρκετά την ανάπτυξη ελευθερου λογισμικού (χωρίς

άδεια για χρήση δηλαδή) και εστίασαν περισσότερο το ενδιαφέρον για τη δημιουργία και την προώθηση ως πιο δημοφιλή εναλλακτικών μορφών αρχείων και κωδικοποίησης όπως το WMA και το Ogg Vorbis. Η Microsoft, δημιουργός των λειτουργικών συστημάτων Windows, επέλεξε να αποχωριστεί από το MP3 και να δημιουργήσει τη δική της, "ιδιωτική" μορφή Windows Media, για να αποφύγει θέματα αδειοδότησης που είχαν σχέση με τις ευρεσιτεχνίες. Μέχρι να λήξουν οι προθεσμίες των ευρεσιτεχνιών, προγράμματα κωδικοποίησης και αναπαραγωγής δημιουργούν θέματα παραβίασης δικαιωμάτων σε όσες χώρες αναγνωρίζονται αυτές οι ευρεσιτεχνίες. Παρόλους τους περιορισμούς λόγω ευρεσιτεχνιών, η ύπαρξη της μορφής MP3 συνεχίζεται. Οι λόγοι για τους οποίους το MP3 είναι τόσο δημοφιλές φαίνεται να είναι οι εξής:

- Η οικειότητα με την μορφή του αρχείου.
- Η μεγάλη ποσότητα μουσικών αρχείων που είναι διαθέσιμες στην μορφή αυτή.
- Η μεγάλη ποικιλία ήδη διαθέσιμου λογισμικού και υλικού που εκμεταλλεύεται τη συγκεκριμένη μορφή αρχείου
- Η έλλειψη περιορισμών που έχουν να κάνουν με ψηφιακά δικαιώματα (μουσικής), που κάνει τα αρχεία MP3 εύκολα να επεξεργαστούν, να αντιγραφούν και να διανεμηθούν μέσω δικτύου
- Η πλειοψηφία των οικιακών χρηστών, οι οποίοι δεν γνωρίζουν ή δεν ενδιαφέρονται για τη διαμάχη των ευρεσιτεχνιών, και που συχνά δεν εξετάζουν τέτοια ζητήματα όταν επιλέγουν τη μορφή μουσικού αρχείου, που θα χρησιμοποιήσουν για προσωπική χρήση.

Επιπρόσθετα, οι κάτοχοι των ευρεσιτεχνιών αρνήθηκαν να ζητήσουν δικαιώματα από αποκωδικοποιητές ανοιχτού κώδικα, επιτρέποντας έτσι να αναπτυχθούν πολλοί αποκωδικοποιητές [MP3]. Εκτός αυτού, ενώ έγιναν προσπάθειες να αποθαρρυνθούν όσοι διένειμαν κωδικοποιητές, η Thomson δήλωσε ότι όσα άτομα χρησιμοποιούν δωρεάν κωδικοποιητές δεν απαιτείται να πληρώσουν δικαιώματα. Έτσι, ενώ τα θέματα των δικαιωμάτων υφίσταντο για τις εταιρείες που προσπαθούσαν να χρησιμοποιήσουν το MP3, δεν επηρέασαν επί της ουσίας τους χρήστες, επιτρέποντας στην μορφή αυτή να γίνει δημοφιλέστερη. Η εταιρεία Sisvel S.p.A και η θηγατρική της στις Η.Π.Α. Audio MPEG, Inc. είχαν μηνύσει παλαιότερα την Thomson για παραβίαση ευρεσιτεχνιών σχετικών με την τεχνολογία του MP3, αλλά αυτές οι διαφορές επιλύθηκαν τον Νοέμβριο του 2005, οπότε η Sisvel έδωσε άδεια στην

Thomson για τις ευρεσιτεχνίες της. Επίσης, η Motorola πρόσφατα υπέγραψε συμφωνία με την Audio MPEG για τη αδειοδότηση των σχετικών με το MP3 ευρεσιτεχνιών. Το Σεπτέμβριο του 2006 Γερμανοί αξιωματούχοι κατέσχεσαν συσκευές MP3 από το περίπτερο της Sandisk στην Έκθεση "IFA Show" στο Βερολίνο, αφού μία Ιταλική εταιρεία ευρεσιτεχνιών πέτυχε την εφαρμογή ασφαλιστικών μέτρων εκ μέρους της Sisvel και εναντίον της Sandisk σε μία διαμάχη σχετικά με την αδειοδότηση των δικαιωμάτων. Τα ασφαλιστικά μέτρα ακυρώθηκαν αργότερα από δικαστή του Βερολίνου αλλά η ακύρωση αυτή ανακλήθηκε την ίδια μέρα από έναν άλλο δικαστή του ίδιου δικαστηρίου, "δημιουργώντας μία βεντέτα τύπου άγριας δύσης στις ευρεσιτεχνίες στην Γερμανία" όπως ανέφερε ο σχολιαστής. Στις 16 Φεβρουαρίου του 2007, η εταιρεία Texas MP3 Technologies μήνυσε την Apple, την Samsung Electronics και την Sandisk για παραβίαση ευρεσιτεχνιών σχετικά με τις φορητές συσκευές αναπαραγωγής MP3. Η μήνυση κατατέθηκε στο Marshall του Τέξας. Αυτή η περιοχή είναι συνηθισμένη για τις μηνύσεις που γίνονται σχετικά με θέματα καταπάτησης ευρεσιτεχνιών, γιατί οι δίκες γίνονται πολύ γρήγορα και οι ένορκοι συχνά παίρνουν το μέρος του ενάγοντα. Η Texas MP3 Technologies υποστήριξε παραβίαση της ευρεσιτεχνίας των Η.Π.Α με αριθμό 7065417, που κατοχυρώθηκε τον Ιούνιο του 2006 στην εταιρεία κατασκευής μικροτσιπ πολυμέσων SigmaTel και κάλυπτε "ένα φορητό σύστημα αναπαραγωγής ήχου MPEG και μία μέθοδο για την αναπαραγωγή δεδομένων ήχου συμπιεσμένου με την μέθοδο MPEG" Η Alcatel-Lucent διεκδικεί, επίσης, την ιδιοκτησία αρκετών ευρεσιτεχνιών σχετικών με το MP3 και την κωδικοποίηση/αποκωδικοποίηση του. Τον Νοέμβριο του 2006, (πριν από την συγχώνευση των 2 εταιρειών), η Alcatel κατέθεσε αγωγή εναντίον της Microsoft ισχυριζόμενη παραβίαση επτά ευρεσιτεχνιών της. Στις 23 Φεβρουαρίου του 2007 ένα δικαστήριο του Σαν Ντιέγκο υπεραμύνθηκε της αγωγής και επιδίκασε στην Alcatel-Lucent αποζημίωση - ρεκόρ των 1.52 δισ. δολλαρίων για ζημίες που υπέστη η εταιρεία. Η Microsoft δήλωσε ότι θα ασκήσει έφεση κατά της απόφασης, υποστηρίζοντας ότι η απόφαση του ομοσπονδιακού δικαστηρίου είναι "αστήρικτη, βασισμένη στον Νόμο και στην πραγματικότητα" καθώς η Microsoft είχε ήδη πληρώσει 16 εκ. δολάρια για την άδεια των δικαιωμάτων από το ίδρυμα Fraunhofer IIS το οποίο, υποστηρίζει, είναι ο αναγνωρισμένος αδειοδότης που αναγορίζει η βιομηχανία. Μία βδομάδα αργότερα, ο δικαστής της περιφέρειας των Η.Π.Α. Rudi Brewster αποφάσισε ότι όσα υποστήριζε η Alcatel-Lucent's σχετικά με MP3 δεν ισχύουν. Η Alcatel-Lucent σχεδιάζει να ασκήσει έφεση. Εν συντομία, επειδή η

Thomson το Ίδρυμα Fraunhofer, η Sisvel (καθώς και η θυγατρική της στις ΗΠΑ Audio MPEG), η Texas MP3 Technologies και η Alcatel-Lucent διεκδικούν τον νομικό έλεγχο όλων των σχετικών με το MP3 ευρεσιτεχνιών, η νομική υπόσταση του MP3 και των σχετικών ευρεσιτεχνιών είναι ασαφής σε όσες χώρες επιτρέπουν τις ευρεσιτεχνίες λογισμικού.

### 3.5 Εναλλακτικές τεχνολογίες

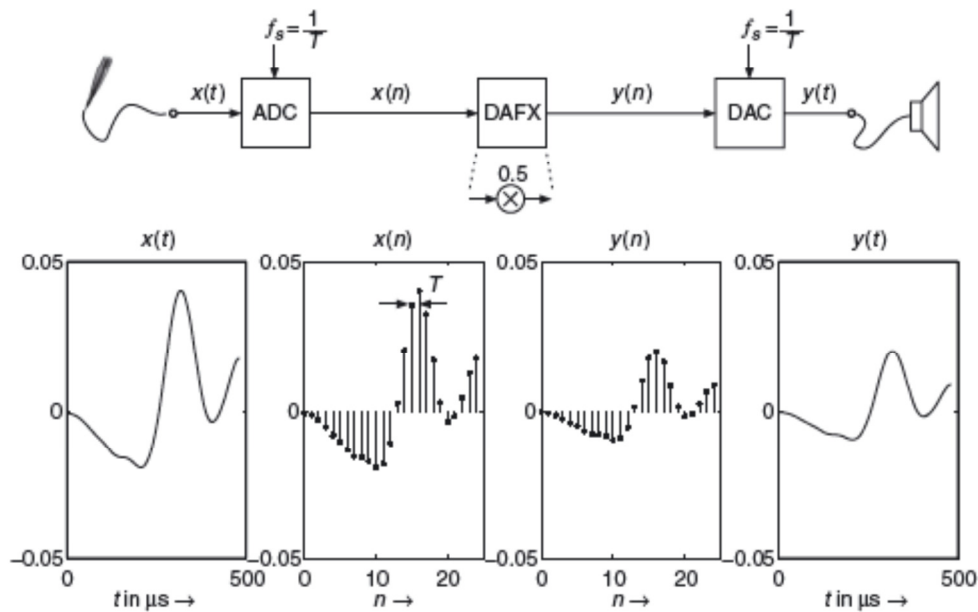
Υπάρχουν διάφοροι άλλοι κωδικοποιητές (codecs) είτε απωλεστικής είτε μη απωλεστικής συμπίεσης. Μεταξύ αυτών τα mp3PRO, AAC & MP2 είναι όλα μέλη της ίδιας οικογένειας τεχνολογίας σαν του MP3 και πάνω κάτω βασίζονται στα ίδια ψυχοακουστικά μοντέλα. Το ίδρυμα Fraunhofer κατέχει πολλές από τις βασικές ευρεσιτεχνίες πάνω στις οποίες βασίζονται αυτοί οι κωδικοποιητές, ενώ άλλοι είναι ιδιοκτησία των Dolby Labs, Sony, Thomson Consumer Electronics, και AT&T.

Το 2005 σε μία ακουστική δοκιμή, η οποία συνέκρινε την απόδοση του κωδικοποιητή MP3 LAME με πιο νέες μορφές συμπίεσης ήχου στα 128Kbit/δευτ., παρατηρήθηκε ότι δεν υπήρχε εμφανής στατιστική διαφορά μεταξύ των LAME, [Ogg Vorbis], αρκετών κωδικοποιητών AAC και του WMA. Όμως, σε μία δοκιμή στα 32Kbit/δευτ., φάνηκε ότι το MP3 ήταν εμφανώς χειρότερο από τις πιο νέες μορφές συμπίεσης ήχου σε χαμηλότερους ρυθμούς bit.

## ΚΕΦΑΛΑΙΟ 4ο : ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

Στο πειραματικό μέρος της εργασίας θα ασχοληθούμε αρχικά με την διαδικασία της κβάντισης ενός ηχητικού σήματος με τη χρήση του λογισμικού Matlab.

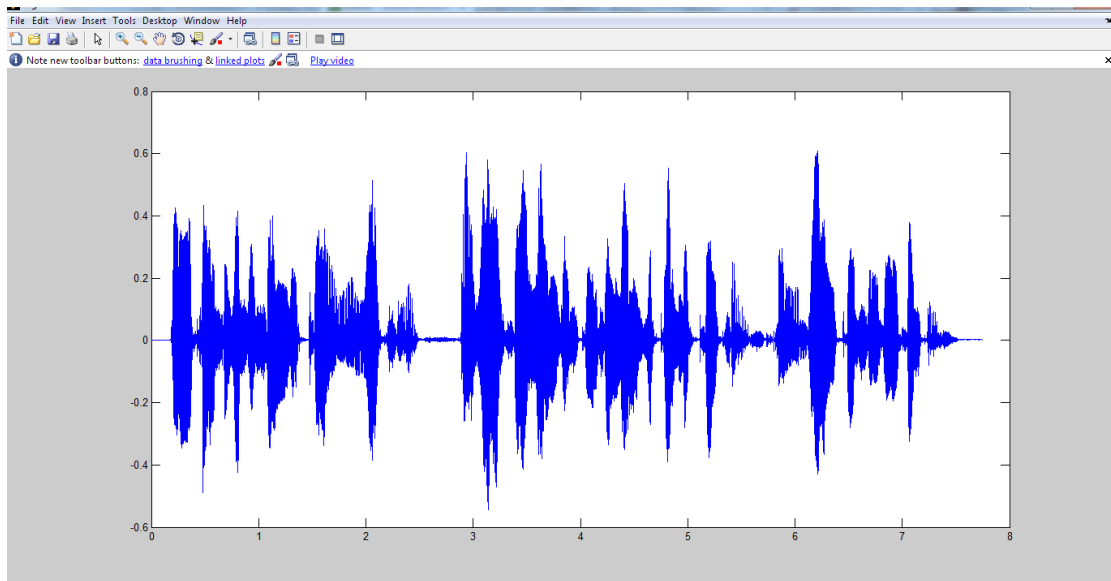
Στο Σχήμα 4.1 απεικονίζονται οι διαδικασίες της δειγματοληψίας και κβαντισμού από ένα ADC μετατροπέα, η δυνατότητα ψηφιακής επεξεργασίας (π.χ., audio effects) και η ανασύνθεση από ένα DAC μετατροπέα για την τελική αναπαραγωγή



**Σχήμα 4.1 : Διάγραμμα ADC-DAC μετατροπών**

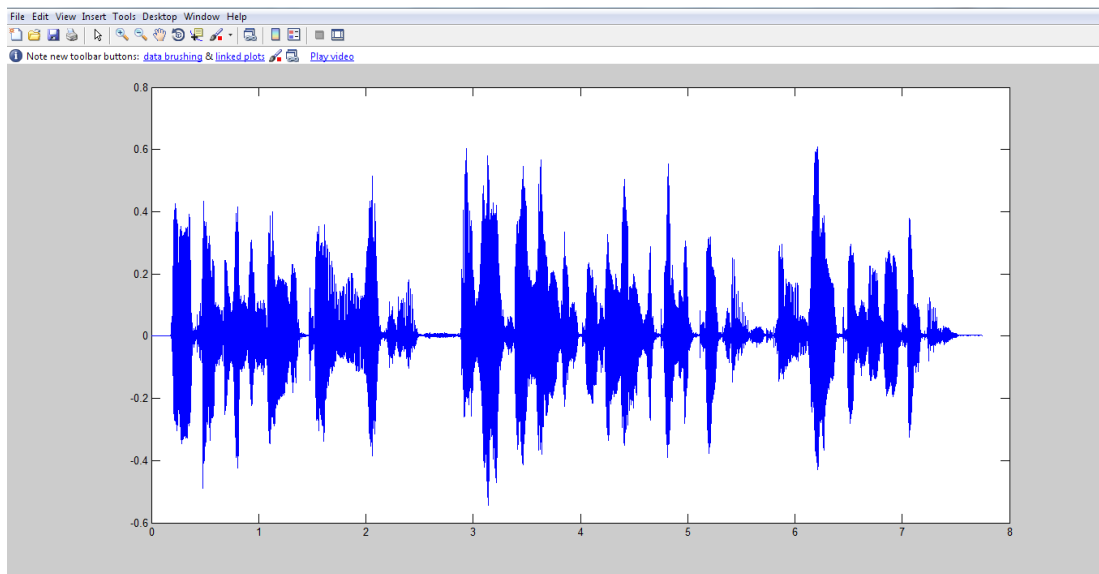
Αρχικά με την εντολή `[y,fs,nbits]=wavread('MAN_SPEA.wav');` μετατρέπουμε ένα ηχητικό σήμα τύπου .wav σε ψηφιακή μορφή μέσω της συνάρτησης wavread. Η συνάρτηση wavread δέχεται ως όρισμα το αρχείο ήχου (εδώ 'MAN\_SPEA.wav') και επιστρέφει το ψηφιακό σήμα ( y ), την συχνότητα δειγματοληψίας (fs) και το πλήθος των bits ανά δείγμα (bits/sample) που χρησιμοποιήθηκαν για την κωδικοποίηση του αρχικού ηχητικού σήματος(nbits). σημειώνεται ότι το σήμα y είναι σε μορφή διανύσματος γραμμής και για το λόγο αυτό το μετατρέπουμε σε μορφή διανύσματος στήλης με την εντολή `y1=y1(:,1);` .

Ακολούθως με την εντολή `plot(y1)` δημιουργούμε τη γραφική απεικόνιση Σχήμα 2 του σήματος σε σχέση με τον χρόνο.



**Σχήμα 4.2 : Αρχικό σήμα ήχου με 16 bits κβαντισμού**

Δημιουργούμε το κατάλληλο χρονικό «διάνυσμα» το οποίο περιέχει τις ακριβείς χρονικές στιγμές δειγματοληψίας του ψηφιακού σήματος, λαμβάνοντας υπόψιν ότι η συχνότητα δειγματοληψίας είναι ίση με  $f_s=44100$  Hz και θεωρώντας αρχική τιμή χρόνου την  $t=0$ . Προσδιορίζουμε τη χρονική διάρκεια του σήματος με την εντολή `max(t)`. Ακολούθως απεικονίζουμε πάλι την κυματομορφή, ως συνάρτηση του χρόνου, σε νέο figure. Σχήμα 3.



**Σχήμα 4.3 : Αρχικό σήμα σε πραγματικό χρόνο**



Υπολογίζουμε τις την μέγιστη και την ελάχιστη τιμή του πλάτους του σήματος με τις εντολές  $\max(y1)$  και  $\min(y1)$  αντίστοιχα και έχουμε:

$$\max(y1) = 0.6096$$

$$\min(y1) = -0.5429$$

Μετατρέπουμε τις τιμές του σήματος εισόδου σε ακέραιες. Επειδή το σήμα ήταν αποθηκευμένο σε αρχείο **.wav**, είναι ήδη κβαντισμένο (με ευκρίνεια κβαντισμού 16bit), αρα αρκεί ο πολλαπλασιασμός των τιμών του με την μέγιστη (απολύτως) τιμή της αναπαράστασης προσημασμένων ακεραίων μήκους **2 byte** (32768). Ακολούθως υπολογίζουμε εκ νέου τη μέγιστη και ελάχιστη τιμή του σήματος και παίρνουμε :

$$\max(y1) = 19977$$

$$\min(y1) = -17791$$

Στη συνέχεια επανακβαντίζουμε το σήμα εισόδου  $y1$  με ευκρίνεια κβαντισμού 8 bit, χωρίς χρήση dither και τεχνικής μορφοποίησης θορύβου (noise---shaping) με τη βοήθεια της συνάρτησης QuantiseAudio. Η συνάρτηση QuantiseAudio έχει γενική μορφή :

**[OutSignal]=QuantiseAudio(InSignal,N,up\_limit,down\_limit,NSFlag,DitherFlag)**

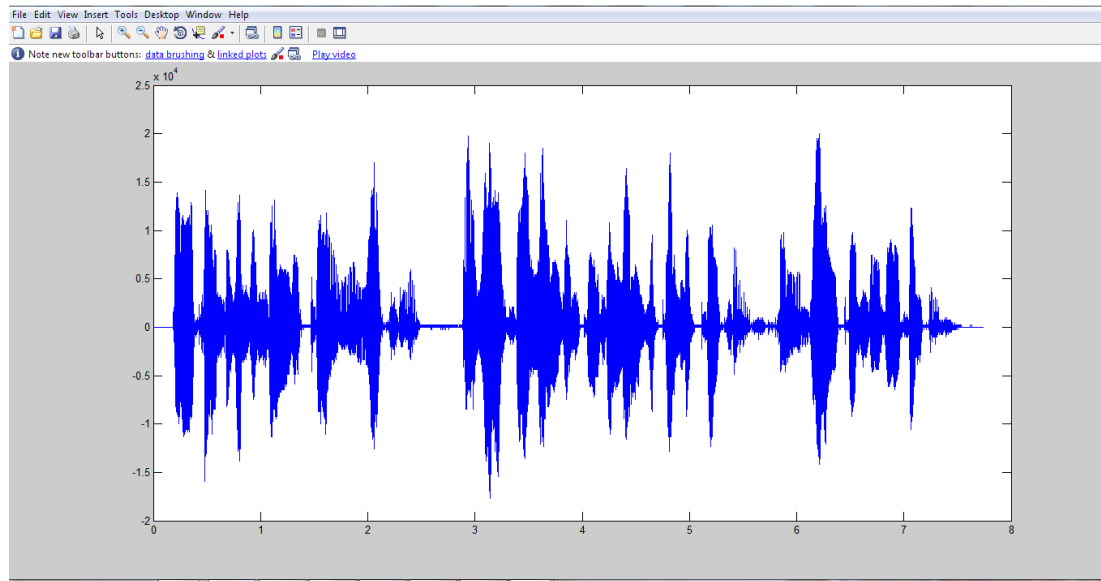
Δέχεται ως ορίσματα , ένα σήμα (InSignal) , το πλήθος των bits της ευκρίνειας κβαντισμού (N) , το εύρος των τιμών του κβαντιστή (μέσω των ακραίων τιμών up\_limit και down\_limit) , το είδος της τεχνικής μορφοποίησης θορύβου (NSFlag) καθώς και την τιμή της σημαίας DitherFlag με την οποία καθορίζουμε την ύπαρξη και το είδος του dither. Στην έξοδο προκύπτει το κβαντισμένο σήμα OutSignal.

Καλούμε την QuantiseAudio για το σήμα  $y1$  χρησιμοποιώντας αρχικά ως ορίσματα :  $sp=y1$  ,  $N=8$  ,  $up\_limit = 32767$  ,  $down\_limit = -32768$  ,  $NSFlag=0$  και  $DitherFlag=0$  και επανακβαντίζουμε το σήμα δίχως τη χρήση Noise-shapping και dither ( οι αντίστοιχες παράμετροι έχουν τεθεί ως 0)

**RQsp=QuantiseAudio(sp,8,32767,-32768,0,0);**

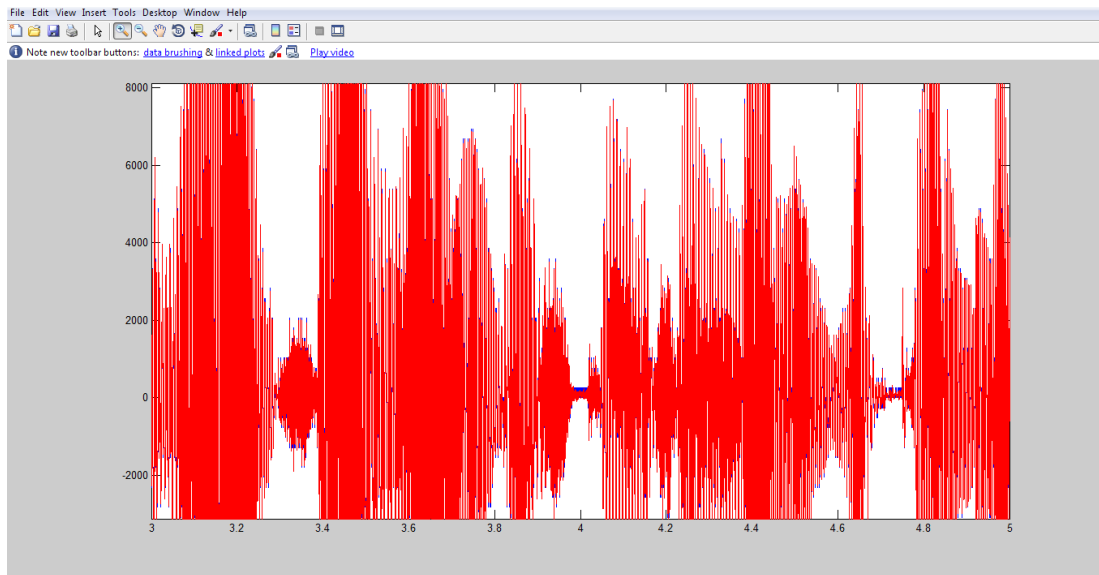
Απεικονίζουμε το επανακβαντισμένο σήμα ως συνάρτηση του χρόνου με την εντολή :

**plot(t, RQsp)** (Σχημα 4).

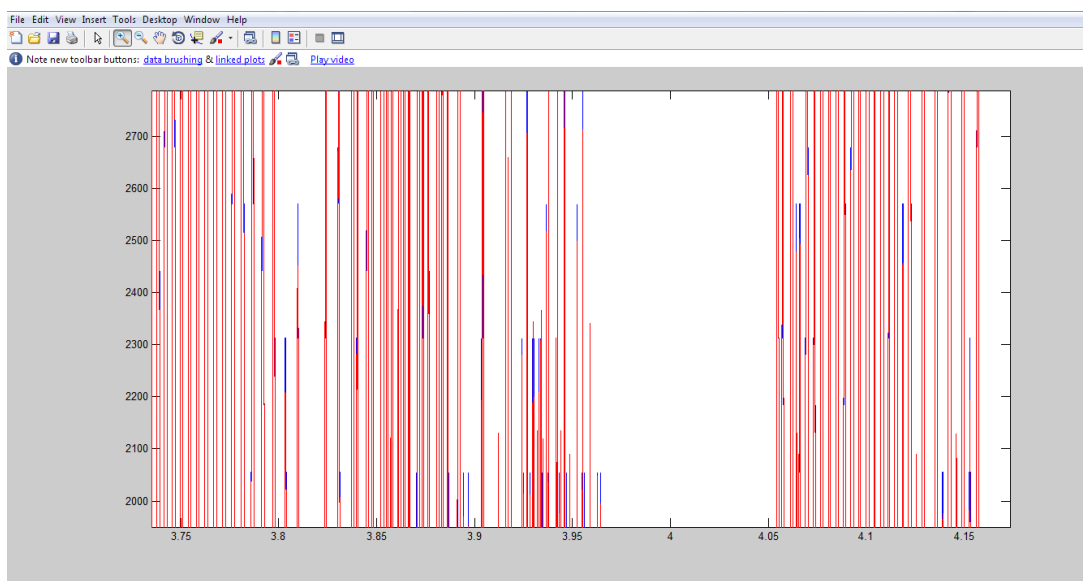


**Σχήμα 4.4 : Επανακβαντισμένο σήμα με 8 bits ευκρίνειας κβαντισμού**

Προκειμένου να εντοπίζουμε τις διαφορές των σημάτων που προκύπτουν από την κβάντιση με τα διαφορετικά επίπεδα ευκρίνειας ( το σήμα sp(κόκκινο) με 16 bits και το RQsp(μπλε) με 8 bits) , τα εμφανίζουμε στο ίδιο διάγραμμα (Σχήμα 5 και Σχήμα 6)



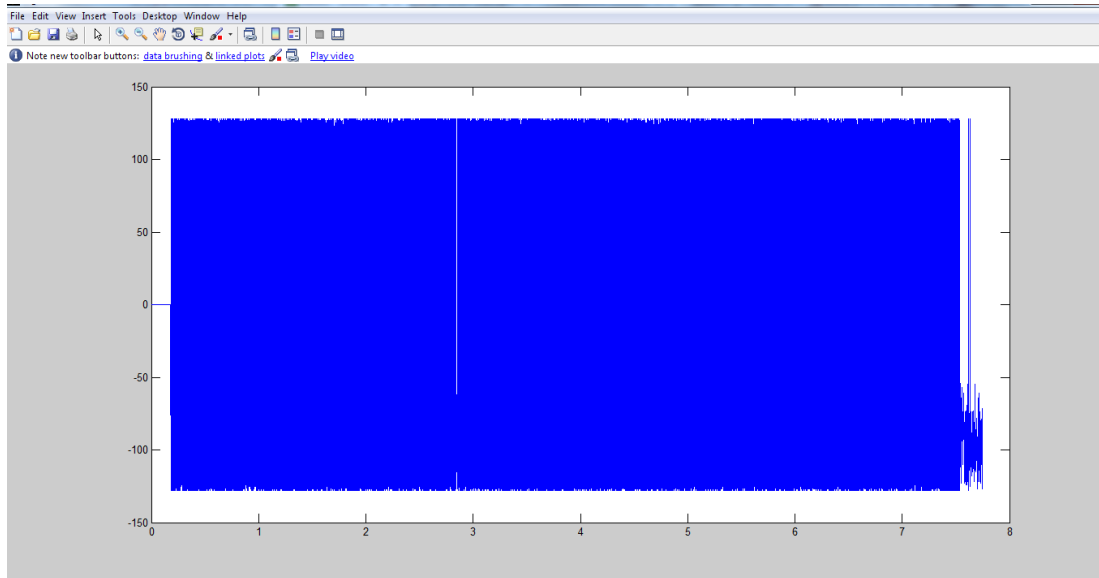
Σχήμα 4.5 : Συγκριτικό διάγραμμα του σήματος με 8 και 16 bits



Σχήμα 4.6 : Εστίαση σημάτων με 8 και 16 bits

Παρατηρούμε ότι το σήμα με που κβαντίστηκε με επίπεδο ευκρίνειας 16 bits ( $y_1$ ) είναι καλύτερης ποιότητας (μεγαλύτερης πιστότητας) από το αντίστοιχο με επίπεδο ευκρίνειας 8 bits ( $RQ_{sp}$ ). Υπολογίσαμε το θόρυβο κβαντισμού (quantization noise) ως τη διαφορά μεταξύ των σημάτων 8 και 16 bit:  $e = RQ_{sp} - y_1$ ;

Το διάγραμμα του θορύβου κβαντισμού  $e$  ως συνάρτηση του χρόνου δίνεται στο σχήμα 7



**Σχήμα 4.7 : Θορυβος κβαντισμού**

Τέλος με τις εντολές :

**wavwrite(e/32768,44100,8,'QuantError8bit.wav');**

**wavwrite(RQsp/32768,44100,8,'Requantised8bit.wav');**

δημιουργούμε αρχεία ήχου που περιέχουν το σφάλμα κβάντισης ('**QuantError8bit.wav**') και το επανακβαντισμένο σήμα με ευκρίνεια 8 bits ('**Requantised8bit.wav**').

Ως επόμενο βήμα εξετάζουμε την επίδραση θορύβου dither που προσθέτουμε στο σήμα με σκοπό την βελτίωση της ποιότητας του. Συγκεκριμένα, η προσθήκη του dither στην είσοδο ενός κβαντιστή προκαλεί μια σειρά από σημαντικές βελτιώσεις στη συνολική λειτουργία του: (α) αλλάζει την μορφή του σφάλματος κβαντισμού και το καθιστά ανεξάρτητο από το σήμα εισόδου, (β) αφαιρεί κάθε είδους αρμονικής παραμόρφωσης η οποία εμφανίζεται κατά την μετατροπή σημάτων πολύ μικρού πλάτους, δίνοντας στον θόρυβο κβαντισμού χαρακτηριστικά όμοια με τα χαρακτηριστικά λευκού θορύβου και (γ) βελτιώνει τη διακριτική ικανότητα του

κβαντιστή, αυξάνοντας κατά υποκειμενικό (κι όχι ποσοτικό) τρόπο τη δυναμική του περιοχής. Γενικά, η εξομάλυνση των παραμορφώσεων λόγω κβαντισμού με προσθήκη dither θα πρέπει να υλοποιείται σε όλες τις επεξεργασίες του ψηφιακού σήματος όπου απαιτείται επανακβαντισμός ή και αριθμητική αποκοπή του (truncation), όπως για παράδειγμα κατά την μεταβολή της στάθμης ή το ψηφιακό φιλτράρισμα του σήματος σε μία ψηφιακή συσκευή.

Με βάση τα παραπάνω, η χρήση dither έχει σαν αποτέλεσμα την αποδιαμόρφωση του σφάλματος κβαντισμού από το σήμα εισόδου και την μείωση της αρμονικής παραμόρφωσης που δημιουργείται κατά τον κβαντισμό σημάτων μικρού πλάτους, με ταυτόχρονη αύξηση το επιπέδου του θορύβου κβαντισμού. Παρόλο που σε τυπικές ηχητικές εφαρμογές (κβαντισμός 16bit, συχνότητα δειγματοληψίας 44.1kHz) η στάθμη θορύβου είναι αρκετά χαμηλή (περίπου ίση με -124dB) κι έχει χαρακτηριστικά λευκού θορύβου, εντούτοις για σήματα χαμηλής έντασης, το επίπεδο του θορύβου ξεπερνά το ελάχιστο κατώφλι ακοής και ο θόρυβος κβαντισμού καθίσταται ακουστός. Η ακουστότητα του θορύβου αυξάνει λόγω της μεταβλητής ευαισθησίας του ανθρωπίνου ακουστικού οργάνου με τη συχνότητα, με αποτέλεσμα ο θόρυβος κβαντισμού να είναι περισσότερο ακουστός από τη συχνότητα των 4kHz και πάνω. Για το λόγο αυτό, σε τυπικές ηχητικές εφαρμογές, ο θόρυβος κβαντισμού μορφοποιείται κατάλληλα στο πεδίο της συχνότητας, χρησιμοποιώντας τεχνικές μορφοποίησης θορύβου (Noise-Shaping). Οι τεχνικές αυτές υλοποιούνται μέσω ανατροφοδοτούμενων ψηφιακών φίλτρων που λειτουργούν ως κλάδοι ανατροφοδότησης γύρω από έναν κβαντιστή.

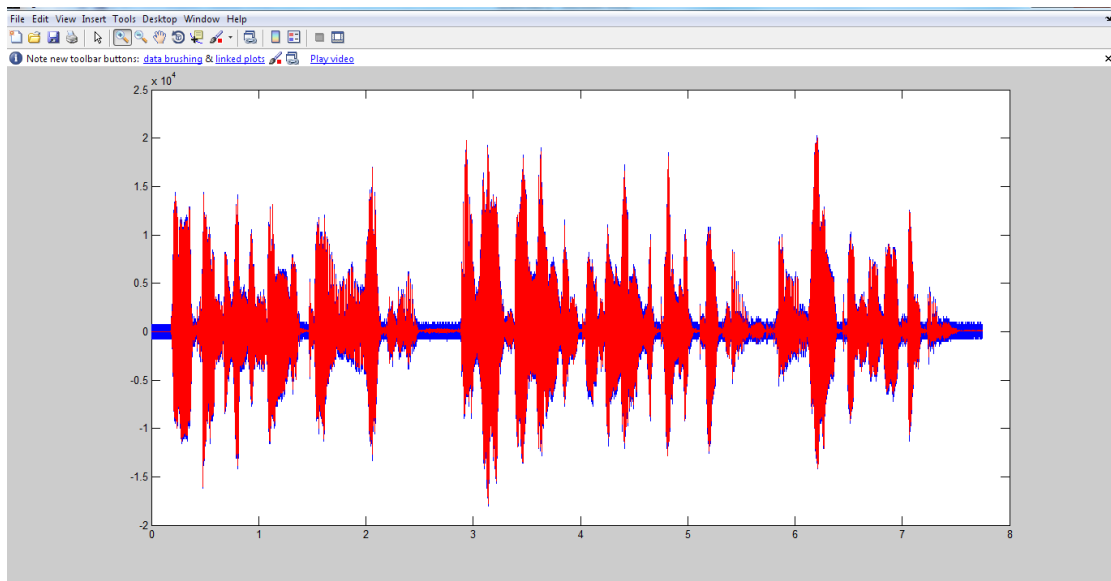
Για την προσθήκη του dither θορύβου, αλλάζουμε τις μεταβλητές **NSFlag**

και **DitherFlag** κατά την κλήση της συνάρτησης **QuantiseAudio** που υλοποιεί τον επανακβαντισμό του σήματος. Συγκεκριμένα η εντολή διαμορφώνεται ως εξής :

**RQsp=QuantiseAudio(sp,4,32767,-32768,3,1)**; όπου οι παράμετροι τέθηκαν ως

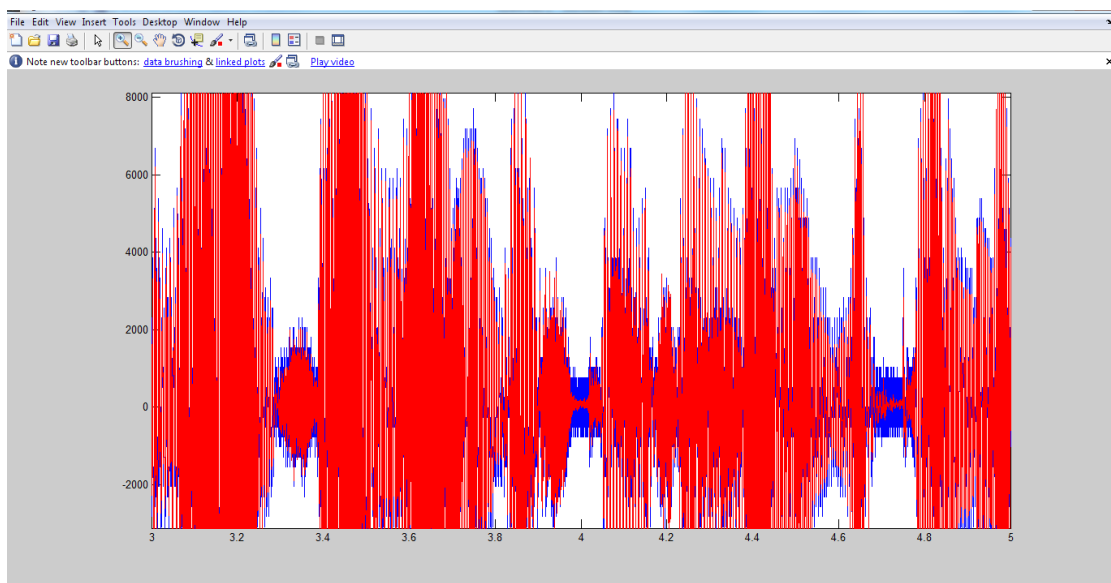
NSFlag=3 και DitherFlag=1 που αντιστοιχεί σε RPDF dither θόρυβο.

Τα διαγράμματα του σήματος που προέκυψαν δίνονται στα σχήματα που ακολουθούν:



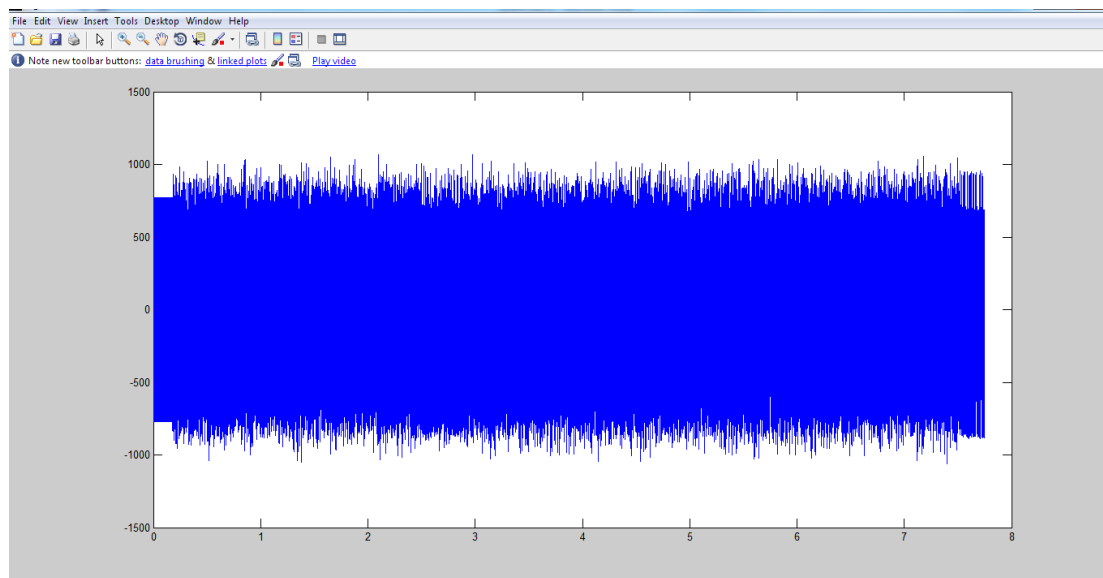
**Σχήμα 4.8 : Συγκριτικό διάγραμμα σημάτων παρουσία dither**

Με κόκκινο χρώμα εμφανίζεται το σήμα με ευκρίνεια 16 bits ενώ με μπλέ χρώμα , το επανακβαντισμένο σήμα με ευκρίνεια 8 bits παρουσία και θορύβου dither. Για να γίνει εμφανέστερη η παρουσία του ενισχυμένου θορύβου στο σήμα παρουσιάζουμε το διάγραμμα με μεγαλύτερη λεπτομέρεια (zoom in):



**Σχήμα 4.9 : Εστίαση συγκριτικών διαγραμμάτων**

Το διάγραμμα του θορύβου κβαντισμού  $e$  ως συνάρτηση του χρόνου δίνεται στο σχήμα 10:



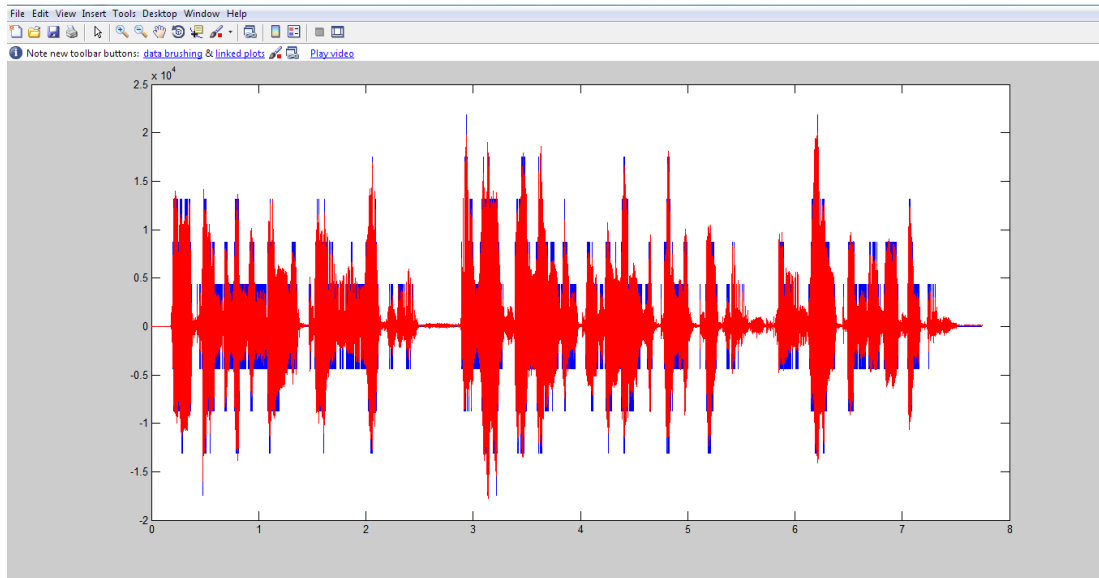
**Σχήμα 4.10 : Θόρυβος κβαντισμού παρουσία dither**

Το επίπεδο του θορύβου κβάντισης είναι ασφαλώς όπως ήταν αναμενόμενο , αυξημένο λόγω και της παρουσίας του dither σε αυτή την περίπτωση.

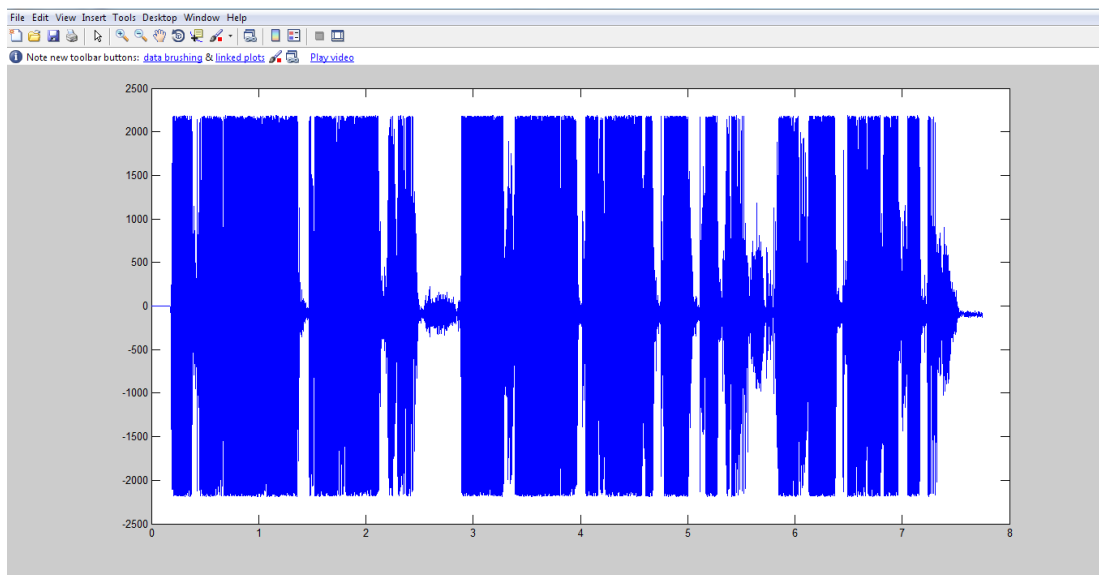
Δημιουργούμε αρχεία ήχου που περιέχουν το σφάλμα κβάντισης ('**QuantError8bit2.wav**') και το επανακβαντισμένο σήμα με ευκρίνεια 8 bits παρουσία dither ('**Requantised8bit2.wav**').

Τέλος επαναλαμβάνουμε τον κώδικα επανακβαντισμού του ηχητικού σήματος με ευκρίνεια 4 bits και χωρίς την χρήση θορύβου dither και φίλτρου Noise Shaping.

Στην περίπτωση αυτή πλέον ο θόρυβος κβαντισμού και η αλλοίωση που αυτός επιφέρει στο σήμα είναι εμφανής και στα διαγράμματα που ακολουθούν αλλά και γίνεται "ενοχλητικός" ακουστικά στα αρχεία ήχου τύπου wav ,('QuantError8bit3.wav' και '**Requantised8bit3.wav**') που δημιουργήσαμε.



**Σχήμα 4.11 : Συγκριτικό διάγραμμα σημάτων με 16 και 4 bits ευκρίνειας**



**Σχήμα 4.12 : Θόρυβος κβαντισμού για επανακβαντισμένο σήμα με 4 bits ευκρίνειας**

Ο κώδικας Matlab για τα τρία σενάρια που υλοποιήσαμε παρατίθεται αμέσως παρακάτω.

## 4.1 ΚΩΔΙΚΑΣ MATLAB



## A) Επανακβάντιση Ηχητικού σήματος με 8 bits ευκρίνειας (χωρίς την παρουσία dither και NSF)

```
clc
clear all

[speech,fs,N]=wavread('MAN_SPEA.wav');

y1=speech(:,1);
npoints=length(y1)
figure
plot(y1)

t=[0:1:npoints-1]/fs;
figure
plot(t,y1)

max(y1)
min(y1)

y1=y1*32768;

max(y1)
min(y1)

RQsp=QuantiseAudio(y1,8,32767,-32768,0,0);

figure
plot(t,RQsp)
hold on
plot(t,y1,'r')
hold off

e=RQsp-y1;
figure
plot(t,e);

sound(e./32768,fs)

wavwrite(e/32768,44100,16,'QuantError8bit.wav');

wavwrite(RQsp/32768,44100,16,'Requantised8bit.wav');
```

## B) Επανακβάντιση με 8 bits ευκρίνειας ( με την παρουσία RPDF dither και NSF 3ης τάξης)

```
clc
clear all
```

```

[speech, fs, N]=wavread('MAN_SPEA.wav');

y1=speech(:,1);
npoints=length(sp)
figure
plot(y1)

t=[0:1:npoints-1]/fs;
figure
plot(t,y1)

max(y1)
min(y1)

y1=y1*32768;

max(y1)
min(y1)

RQsp=QuantiseAudio(y1,8,32767,-32768,3,1);

figure
plot(t,RQsp)
hold on
plot(t,y1,'r')
hold off

e=RQsp-y1;
figure
plot(t,e);

sound(e./32768,fs)

wavwrite(e/32768,44100,16,'QuantError8bit2.wav');

wavwrite(RQsp/32768,44100,16,'Requantised8bit2.wav');

```

## Γ)Επανακβάντιση με 4 bits ευκρίνειας (χωρίς dither και NSF)

```

clc
clear all

[speech, fs, N]=wavread('MAN_SPEA.wav');

y1=speech(:,1);
npoints=length(y1)
figure

```

```

plot(y1)

t=[0:1:npoints-1]/fs;
figure
plot(t,y1)

max(y1)
min(y1)

y1=y1*32768;

max(y1)
min(y1)

RQsp=QuantiseAudio(y1,8,32767,-32768,0,0);

figure
plot(t,RQsp)
hold on
plot(t,y1,'r')
hold off

e=RQsp-y1;
figure
plot(t,e);

sound(e./32768,fs)

wavwrite(e/32768,44100,16,'QuantError8bit3.wav');

wavwrite(RQsp/32768,44100,16,'Requantised8bit3.wav');

```

## Συνάρτηση QuantiseAudio

```

function
[OutSignal]=QuantiseAudio(InSignal,N,up_limit,down_limit,NSFlag,DitherFlag);
% function y =
QuantiseAudio(x,N,UpLimit,DownLimit,NSFlag,DitherFlag).
% This function quantises the input signal x using N bit resolution.
% The quantiser output range is [DownLimit UpLimit]. During the
% quantisation, controlled dither can be added and noise-shaping
% techniques can be applied, using the following options:
%
%   DitherFlag   = 0      No dither
%                = 1      RPDF dither
%                = 2      TPDF dither
%                = 3      HighPass TPDF dither
%   NSFlag       = 0      No noise-shaping
%                = 2      2nd order noise-shaping
%                = 3      3rd order noise-shaping
%

```

```

% Initialization
OutSignal=[];
dither=[];

% Find the number of points to be calculated
npoints=max(size(InSignal));

% Calculate the PCM quantization step
LSB=(up_limit-down_limit)/(2^N-1);

% Calculate dither sequence if desired
if (DitherFlag==0)
    dither=zeros(size(InSignal));
    InSignal=InSignal+dither;
else
    dither=dithmake(DitherFlag,LSB,npoints);
    InSignal=InSignal+dither';
end

% Perform quantization process without NoiseShaping
if (NSFlag==0)

    OutSignal=LSB*floor((InSignal/LSB)+1/2);
    i=find(OutSignal>up_limit);
    Umatrix=ones(size(i))*up_limit;
    OutSignal(i)=Umatrix;

    i=find(OutSignal<down_limit);
    Umatrix=ones(size(i))*down_limit;
    OutSignal(i)=Umatrix;
end

% Perform quantization process 2nd Order NoiseShaping
if (NSFlag==2)
    f1=0;
    f2=0;
    for i=1:1:npoints,
        w=2*f1-f2;
        OutSignal(i)=LSB*floor(((InSignal(i)-w)/LSB)+1/2);

        if (OutSignal(i)>up_limit);
            OutSignal(i)=up_limit;
        end

        if (OutSignal(i)<down_limit);
            OutSignal(i)=down_limit;
        end

        f=OutSignal(i)-InSignal(i)+w;

        f2=f1;
        f1=f;
    end
    OutSignal=OutSignal';
end
end

```

```

% Perform quantization process 3rd Order NoiseShaping
if (NSFlag==3)
    f1=0;
    f2=0;
    f3=0;
    for i=1:1:npoints,
        w=3*f1-3*f2+f3;
        OutSignal(i)=LSB*floor(((InSignal(i)-w)/LSB)+1/2);

        if (OutSignal(i)>up_limit);
            OutSignal(i)=up_limit;
        end

        if (OutSignal(i)<down_limit);
            OutSignal(i)=down_limit;
        end

        f=OutSignal(i)-InSignal(i)+w;
        f3=f2;
        f2=f1;
        f1=f;
    end
    OutSignal=OutSignal';

end

```

## Συνάρτηση dithmake

```

function [dith]=dithmake(dithtype,LSB,npoints);

% function [A]=dithmake(dithtype,dithseed,npoints)
%
% Η Συνάρτηση αυτή αποτελεί γεννήτρια dither npoints σημείων. Οι
% υπολογισθείσες τιμές του dither αποθηκεύονται στο διάλυσμα A. Η
% παράμετρος dithtype καθορίζει τον τύπο του dither.
%
% dithtype    = 1, RPDF           με εύρος [-LSB/2 LSB/2]
%             = 2, TPDF           με εύρος [-LSB LSB]
%             = 3, HighPass RPDF με εύρος
%
% Φλώρος Ανδρέας
% Εργαστήριο Ενσύρματης Επικοινωνίας
% Audio Group
% Last Modified: 12/01/1999.

% Initialization
a1=3453;
a2=2945;
m=65536;
c1=1;
c2=1;
dithran1=1531;

```

```

dithran2=18531;
dithtemp1=0;
dithtemp2=0;
dithtemp=0;
dith=[];

% ===== RPDF Dither Generator =====
%
%           *       *       *
%
%           -LSB/2   0       LSB/2
if (dithtype==1)
  for i=1:npoints,
    dithran1=mod((dithran1*a1+c1),m);
    dithtemp1=LSB*(dithran1-32767)/65536;
    dith=[dith dithtemp1];
  end
end
% =====

% ===== TPDF Dither Generator =====
%
%
%           *
%           *
%           *
%           *
%           *
%           -LSB   -LSB/2   0       LSB/2   LSB
if (dithtype==2)
  for i=1:npoints,
    dithran1=mod((dithran1*a1+c1),m);
    dithran2=mod((dithran2*a2+c2),m);
    dithtemp1=LSB*(dithran1-32767)/65536;
    dithtemp2=LSB*(dithran2-32767)/65536;
    dithtemp=(dithtemp1+dithtemp2)/2;
    dith=[dith dithtemp];
  end
end
% =====

% ===== High Pass TPDF Dither Generator =====
%
%
%           *
%           *
%           *
%           *
%           *
%           -LSB   -LSB/2   0       LSB/2   LSB
if (dithtype==3)
  for i=1:npoints,
    dithran1=mod((dithran1*a1+c1),m);
    dithtemp1=LSB*(dithran1-32767)/65536;
    dithtemp=(dithtemp1-dithtemp2)/2;
    dithtemp2=dithtemp;
    dith=[dith dithtemp];
  end
end
% =====

```

## **BIBΛΙΟΓΡΑΦΙΑ**

[1] Jeroen Breebaart, Christof Faller , Spatial Audio Processing MPEG Surround and Other Applications , John Wiley & Sons

[2] Θωμας Σ. Ζαρούχας , Μοντελοποίηση και επεξεργασία ηχητικών δεδομένων για αναπαραγωγή σε χώρους με αντήχηση , Διδακτορική Διατριβή , Πανεπιστήμιο Πατρών

[3] Andreas Spanias ,Ted Painter ,Venkatraman Atti , Audio Signal Processing and coding, WILEY-INTERSCIENCE

[4] Yuli You , Audio Coding Theory and Applications , Springer

[5] Παρασκευάς Μ , Μουρτζόπουλος Γ. ,Συμπύεση Ηχητικών Σημάτων από την ADPCM στο MPEG-4, Ομάδα Ήχου, Εργαστήριο Ενσύρματης Τηλεπικοινωνίας Τμήμα Ηλεκτρολόγων Μηχανικών & Τεχνολογίας Υπολογιστών Πανεπιστήμιο Πατρών

[5] Ιστοσελίδα Wikipedia, <http://www.wikipedia.org>

[6] Ιστοσελίδα Matlab, <http://www.mathworks.org>

[7] Ιστοσελίδα [http://www.inf.teilam.gr/OLD/course\\_material](http://www.inf.teilam.gr/OLD/course_material)

[8] Ιστοσελίδα [www.ece.upatras.gr/gr/personnel/faculty.html?id=180](http://www.ece.upatras.gr/gr/personnel/faculty.html?id=180)

[9] Ιστοσελίδα [www.csd.uoc.gr/~hy430/HY430-2004\\_1.ppt](http://www.csd.uoc.gr/~hy430/HY430-2004_1.ppt)

[10] Ιστοσελίδα [http://users.ionio.gr/~floros/lessons/avarts\\_dap\\_Lab02.pdf](http://users.ionio.gr/~floros/lessons/avarts_dap_Lab02.pdf)

