

**Τμήμα
Μηχανικών
Πληροφορικής τ.ε.**

Τεχνολογικό Εκπαιδευτικό Ίδρυμα
Δυτικής Ελλάδας

Πτυχιακή Εργασία

Στο πλαίσιο του Προγράμματος Μεταπτυχιακών Σπουδών

«Τεχνολογίες και Συστήματα Ευρυζωνικών Εφαρμογών και Υπηρεσιών»

**«Η εξέλιξη των προτύπων κωδικοποίησης ήχου.
Από τον ήχο MPEG-1 στον ήχο MPEG-H»**

**Δήμητρα Μαρκουτσά
Α.Μ.: Μ3**

Επιβλέποντες:

Επικ. Καθ. Μιχάλης Παρασκευάς

Δρ. Θωμάς Ζαρούχας

Αντίρριο, Ιούνιος 2015

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή
Αντίρριο, 16 Ιουνίου 2016

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Παρασκευάς Μιχαήλ
2. Ζαρούχας Θωμάς
3. Ασαρίδης Ηλίας

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα πτυχιακή εργασία εκπονήθηκε στο πλαίσιο του Προγράμματος Μεταπτυχιακών Σπουδών, «Τεχνολογίες και Συστήματα Ευρυζωνικών Εφαρμογών και Υπηρεσιών» του Τμήματος Μηχανικών Πληροφορικής του Τεχνολογικού Εκπαιδευτικού Ινστιτούτου Δυτικής Ελλάδας, υπό την επίβλεψη του Επικ. Καθ. Μιχάλη Παρασκευά και του Δρ. Θωμά Ζαρούχα. Θα ήθελα λοιπόν να ευχαριστήσω θερμά τους επιβλέποντες της πτυχιακής εργασίας μου, για την πολύτιμη βοήθεια και καθοδήγησή τους.

Επίσης, θα ήθελα να ευχαριστήσω θερμά την οικογένειά μου και τους φίλους μου για την αγάπη και την υποστήριξή τους όλα αυτά τα χρόνια.

Δήμητρα Μαρκουτσά

Πίνακας Περιεχομένων

Ευχαριστίες	3
Περίληψη	11
Abstract	12
Κεφάλαιο 1 Βασικές Αρχές Επεξεργασίας Ηχητικών Σημάτων.....	13
1.1 Εισαγωγή	13
1.2 Αναπαράσταση Ηχητικών Σημάτων.....	13
1.3 Κβάντιση.....	16
1.4 Εισαγωγή στην Ψυχοακουστική	16
1.4 Ποιοτική Μέτρηση Κωδικοποιήσεων Ήχου	18
Κεφάλαιο 2 Πρότυπα Κωδικοποίησης Ήχου	21
2.1 Εισαγωγή	21
2.2 Το πρότυπο MPEG -1	22
2.1.1 Κύρια χαρακτηριστικά του πρότυπου MPEG-1	22
2.1.2 Απεικόνιση χρόνου στη συχνότητα.....	24
2.1.3 Ψυχοακουστικό μοντέλο του MPEG-1	26
2.1.3.1. Ψυχοακουστικό μοντέλο 1	27
2.1.3.2. Ψυχοακουστικό μοντέλο 2	28
2.1.4 Δομή ηχητικού σήματος κατά το πρότυπο MPEG-1	29
2.1.5 Στερεοφωνική κωδικοποίηση ήχου	31
2.2 Το πρότυπο MPEG-2	33
2.2.1 Εισαγωγή στον ήχο πολλών καναλιών	33
2.2.2 Το πρότυπο MPEG-2 AAC	35
2.2.3 Κβαντισμός και κωδικοποίηση	39
2.2.4 Κωδικοποίηση χωρίς θόρυβο	43
2.2.5 Πολύπλεξη συμβολοσειράς.....	46
2.2.6 Εκτίμηση πολυπλοκότητας αποκωδικοποιητή.....	47
2.3 Το πρότυπο MPEG-4	49
2.3.1 Στόχοι και λειτουργίες.....	50
2.3.2 Επιδόσεις του προτύπου MPEG-4.....	52
2.3.3 Προστασία πνευματικών δικαιωμάτων	53
2.4 Το Πρότυπο MPEG – D	54

2.5	Το Πρότυπο MPEG – H	55
2.5.1	Απαιτήσεις προτάσεων για τον τρισδιάστατο ήχο	55
2.5.2	Μορφή εισόδου κωδικοποιητή τρισδιάστατου ήχου	57
2.5.2.1	Είσοδος καναλιού.....	57
2.5.2.2	Είσοδος αντικειμένου.....	57
2.5.2.3	Higher Order Ambisonics HOA είσοδος.....	60
2.5.3	Προφίλ τρισδιάστατου ήχου.....	61
2.6	Σύγκριση προτύπων MPEG-1 και MPEG-2 με το πρότυπο MPEG-4	63
Κεφάλαιο 3	Υλοποίηση.....	66
3.1	Εισαγωγή	66
3.2	Υλοποίηση MPEG-4 AAC με τη συνάρτηση Audiowrite	67
3.3	Συμπεράσματα	72
3.4	Υλοποίηση MPEG-1.....	87
Βιβλιογραφία	93

Κατάλογος Εικόνων

Εικόνα 1 Μεταβολές πίεσης σε σχέση με το χρόνο	14
Εικόνα 2 Ψηφιακή και αναλογική αναπαράσταση ηχητικού σήματος.....	15
Εικόνα 3 Κβαντιστής μηδενικής στάθμης και κβαντιστής μη μηδενικής στάθμης	16
Εικόνα 4 Η φυσιολογία του αυτιού [10]	18
Εικόνα 5 Σχηματικό διάγραμμα κωδικοποιητή MPEG-1	23
Εικόνα 6 Σχηματικό διάγραμμα αποκωδικοποιητή MPEG-1	24
Εικόνα 7 Σχηματικό διάγραμμα του φίλτρου κωδικοποίησης (ανάλυσης) του MPEG-1, επίπεδο III.	25
Εικόνα 8 Σχηματικό διάγραμμα φίλτρου αποκωδικοποίησης (σύνθεσης) του MPEG-1, επίπεδο III.....	25
Εικόνα 9 Σχηματικό διάγραμμα ψυχοακουστικού μοντέλου 1.....	28
Εικόνα 10 Σχηματικό μοντέλο ψυχοακουστικού μοντέλου 2	29
Εικόνα 11 Δομή του πλαισίου δεδομένων του ήχου MPEG-1 για τα επίπεδα I,II και III	30
Εικόνα 12 Κεφαλίδα ήχου MPEG-1.....	30
Εικόνα 13 Δομή πακέτων ήχου για τα 3 επίπεδα του προτύπου MPEG-1.....	31
Εικόνα 14 Τοποθέτηση ηχείων για αναπαραγωγή ήχου 5 καναλιών.....	34
Εικόνα 15 Διάγραμμα MPEG-2 AAC κωδικοποιητή [9]	37
Εικόνα 16 Διάγραμμα MPEG-2 AAC αποκωδικοποιητή.....	39
Εικόνα 17 Λογικό διάγραμμα εσωτερικού βρόχου επανάληψης MPEG-2 AAC.....	41
Εικόνα 18 Λογικό διάγραμμα εξωτερικού βρόχου επανάληψης MPEG-2 AAC	42
Εικόνα 19 Αλγόριθμοι και εφαρμογές τους ανάλογα με τις συχνότητες που υποστηρίζουν.....	51
Εικόνα 20 Σχηματικό διάγραμμα MPEG-H 3D ήχου	61
Εικόνα 21 Δομικό διάγραμμα αποκωδικοποιητή 3D ήχου σε σύνδεση με ένα MPEG Surround αποκωδικοποιητή	62
Εικόνα 22 Κωδικοποιητής MPEG-4 CELP για σήματα ομιλίας.....	63
Εικόνα 23 Κωδικοποιητής MPEG-4 γενικής μορφής.....	64
Εικόνα 24 Γράφημα τυχαίου σήματος	73
Εικόνα 25 Το αρχικό σήμα (πάνω) και το κωδικοποιημένο σήμα (κάτω).....	74
Εικόνα 26 Διαφορά αρχικού σήματος με το κωδικοποιημένο σήμα στα πρώτα δείγματα	74
Εικόνα 27 Διαφορά αρχικού και κωδικοποιημένου σήματος ως προς τα τελευταία δείγματα	75
Εικόνα 28 Γράφημα μουσικού σήματος.....	79
Εικόνα 29 Το αρχικό σήμα (πάνω) και το κωδικοποιημένο σήμα (κάτω).....	80
Εικόνα 30 Διαφορά αρχικού σήματος με το κωδικοποιημένο σήμα στα πρώτα δείγματα	80
Εικόνα 31 Διαφορά αρχικού και κωδικοποιημένου σήματος ως προς τα τελευταία δείγματα	81
Εικόνα 32 Σύγκριση των τελευταίων δειγμάτων fade-out των δυο κωδικοποιήσεων (πάνω 192 kbps, κάτω 128 kbps).....	82
Εικόνα 33 Δείγμα fade-out για ρυθμό δεδομένων 96 kbps	82
Εικόνα 34 Διαφορά αρχικού σήματος με τα κωδικοποιημένα σήματα για ρυθμούς δεδομένων 128, 192 και 96 kbps αντίστοιχα	83
Εικόνα 35 Γράφημα σήματος ορχηστρικής μουσικής.....	84

Εικόνα 36 Σήματα διαφοράς του αρχικού σήματος με τα κωδικοποιημένα κατά MPEG-4 AAC για ρυθμούς δεδομένων 128, 192 και 96 kbps αντίστοιχα	86
Εικόνα 37 Τοπικά μέγιστα στο πεδίο της συχνότητας	87
Εικόνα 38 Τονικά στοιχεία συναρτήσεως της συχνότητας.....	88
Εικόνα 39 Τονικά και μη τονικά στοιχεία συναρτήσεως της συχνότητας.....	88
Εικόνα 40 Μη τονικά στοιχεία σε σχέση με τα απόλυτα κατώφλια.....	89
Εικόνα 41 Τονικά στοιχεία σε σχέση με τα απόλυτα κατώφλια.....	89
Εικόνα 42 Τονικά και μη τονικά στοιχεία σε σχέση με τα απόλυτα κατώφλια.....	90
Εικόνα 43 Κατώφλια επικάλυψης τονικών στοιχείων	90
Εικόνα 44 Κατώφλια επικάλυψης για τονικά και μη τονικά στοιχεία	91
Εικόνα 45 Συνολικό ελάχιστο κατώφλι επικάλυψης	91

Κατάλογος Πινάκων

Πίνακας 1 Κλίμακα ακουστικής παραμόρφωσης	20
Πίνακας 2 Μέγιστος ρυθμός δεδομένων ανά συχνότητα δειγματοληψίας	38
Πίνακας 3 Χαρακτηριστικά κωδίκων.....	45
Πίνακας 4 Αριθμός εντολών AAC ανά στάδιο	48
Πίνακας 5 Απαιτήσεις μνήμης AAC.....	49
Πίνακας 6 Σύγκριση απαιτήσεων των δυο προφίλ.....	49
Πίνακας 7 Αξιολογήσεις κωδικοποιήσεων MPEG με βάση το εργαλείο κωδικοποίησης, τον αριθμό καναλιών και το ρυθμό δεδομένων.....	53
Πίνακας 8 Κωδικοποιήσεις MPEG και χαρακτηριστικά	65
Πίνακας 9 Αρχεία που υποστηρίζονται από την audiodwrite.....	67
Πίνακας 10 Μέγιστος αριθμός καναλιών ανά κωδικοποίηση	68
Πίνακας 11 Πεδίο ορισμού δεδομένων	68
Πίνακας 12 Μορφή και bits εξόδου ανά δείγμα	70

Συντμήσεις Όρων

ASIC	Application Specific Integrated Circuit
ASPEC	Adaptive Spectral Perceptual Entropy Coding
BMLD	Binaural Masking Level Difference
CELP	Code Excited Linear Prediction
CRC	Cyclic Redundancy Code
CQF	Conjugate Quadrature Filters
DAB	Digital Audio Broadcasting
DRC processing	Dynamic Range Control processing
DVB	Digital Video Processing
FEC error protection	Forward error correction error protection
HOA	Higher Order Ambisonics
IMDCT	Inverse Modified Discrete Cosine Transform
IPMP	Intellectual Property Management and Protection
ITU-R	International Telecommunication Union Radiocommunication Sector
ITU-T	International Telecommunication Union Telecommunication Sector
LTP	Long Term Prediction
MDCT	Modified Discrete Cosine Transform
MPEG	Moving Pictures Expert Group
OTDAC	Oddly Stacked Time Domain Aliasing Cancellation
PCM	Pulse-Code Modulation
PNS	Perceptual Noise Substitution
PQMF	Pseudo Quadrature Mirror Filter
QMF	Quadrature Mirror Filter
SA	Structured Audio
SAOC 3D	Spatial Audio Object Coding
SASBF	Structured Audio Sample Bank Format
SCFSI	Scale Factor Selection Information
SNR	Signal to Noise Ratio
SPL	Sound Pressure Level

TTS interface

Text To Speech interface

Twin VQ

Transform-domain Weighted Interleave Vector Quantization

ΠΕΡΙΛΗΨΗ

Η εξέλιξη της τεχνολογίας έφερε τον πλανήτη στην ψηφιακή εποχή. Τα αναλογικά μέσα πέρασαν σε δεύτερη μοίρα λόγω της ευκολίας και των δυνατοτήτων που προσφέρουν τα ψηφιακά μέσα. Η εισαγωγή του CD σαν εξέλιξη των αναλογικών μέσων, όπως οι δίσκοι και οι κασέτες σηματοδότησε την αρχή δημιουργίας προτύπων κωδικοποίησης ήχου. Η παλμοκωδική διαμόρφωση (Pulse Code Modulation – PCM) του προτύπου CD παρά την ιστορική της σημασία, ήταν δεσμευμένη από τον περιορισμό του υψηλού ρυθμού δεδομένων που απαιτούσε. Το πρότυπο MPEG-1 ήρθε για να φέρει τη συμπίεση στον ψηφιακό ήχο. Εκτός της συμπίεσης σαν πρότυπο σχεδιάστηκε ώστε να εξυπηρετεί πολλές εφαρμογές όπως την ψηφιακή μορφή των παλιών αναλογικών μέσων, τη ψηφιακή αναμετάδοση ήχου και την εξυπηρέτηση των ηχητικών αναγκών σε διαδραστικές εφαρμογές και στην ψηφιακή τηλεόραση. Η πρόκληση σε αυτές τις εφαρμογές ήταν η βελτιστοποίηση της ποιότητας με ταυτόχρονη μείωση του ρυθμού δεδομένων κατά το δυνατόν. Αυτό οδήγησε στη δημιουργία μιας σειράς προτύπων MPEG που σκοπό είχαν τόσο την εξυπηρέτηση των παραπάνω αναγκών όσο και την αύξηση των δυνατοτήτων καθώς και τη συμβατότητα κάθε προτύπου με τα προηγούμενα πρότυπα ώστε να μην απαιτείται και ένας παλαιότερου τύπου αποκωδικοποιητής για την αναπαραγωγή ήχου κωδικοποιημένου με βάση παλαιότερο πρότυπο. Η εξέλιξη αυτή των προτύπων οδήγησε από το πρώτο και απλούστερο MPEG-1 στο πιο σύγχρονο και πιο πολύπλοκο πρότυπο MPEG-H που υποστηρίζει ακόμα και τρισδιάστατο ήχο.

ABSTRACT

The evolution of technology brought to the planet the digital era. The analog media of sound reproduction were sidelined, due to the ease of use and the capabilities that digital media offered for sound reproduction. The introduction CD as an evolution of analog means of sound storage and reproduction, such as cassettes and LPs, signaled the beginning of standardizing ways to digitize sound. The CD standard uses Pulse Code Modulation (PCM) which, despite its historical significance, was bound by the restriction that a high bitrate is required. The MPEG-1 standard brought compression to the digital world. Apart from compression, MPEG as a standard was designed to be able to serve a lot of applications such as the digital form of the old analog media, digital broadcasting of audio signals and digital sound for interactive applications and digital television. The challenge in all of these applications was to optimize the quality of sound produced while reducing the bitrate as much as possible. This led to the creation of series of MPEG standards that served these needs and also increased the set of features that the previous standards supported, always keeping the compatibility of each standard to the previous ones. This means that there is no need to have a decoder from a previous standard to decode an audio signal that has been compressed with a previous standard. The evolution of these standards led from the first and simplest MPEG-1 standard to the most contemporary and complex MPEG-H standard which is the first standard to support 3D sound.

Κεφάλαιο 1

Βασικές Αρχές Επεξεργασίας Ηχητικών Σημάτων

1.1 Εισαγωγή

Ο ήχος είναι η αίσθηση που προκαλείται λόγω της διέγερσης των αισθητηρίων οργάνων της ακοής από μεταβολές πίεσης του ατμοσφαιρικού αέρα. Τα ηχητικά κύματα παράγονται από σώματα που εκτελούν μηχανικές ταλαντώσεις και μεταφέρουν μηχανική ενέργεια. Στον αέρα και υπό κανονικές συνθήκες, η ταχύτητα του ήχου υπολογίζεται στα 331,5 μέτρα το δευτερόλεπτο.

Η αντίληψη του ήχου αποτελεί βασική αίσθηση σε πολλούς οργανισμούς και πραγματοποιείται μέσω της ακοής. Η ακοή στον άνθρωπο εκτείνεται για ήχους με συχνότητα μεταξύ 20 Hz και 20.000 Hz. Ο άνθρωπος με την πρόοδο της τεχνολογίας επεδίωξε την αναπαράσταση του ήχου με τεχνητά μέσα για επικοινωνία, ψυχαγωγία και άλλους λόγους. Αρχικά, η αναπαράσταση των ηχητικών σημάτων γινόταν με αναλογικά μέσα. Η ψηφιακή εποχή όμως προσδιόρισε τη μετάβαση σε ψηφιακά πρότυπα αναπαράστασης ήχου και εισήγαγε την έννοια της κωδικοποίησης των ηχητικών σημάτων με βάση κάποια πρότυπα.

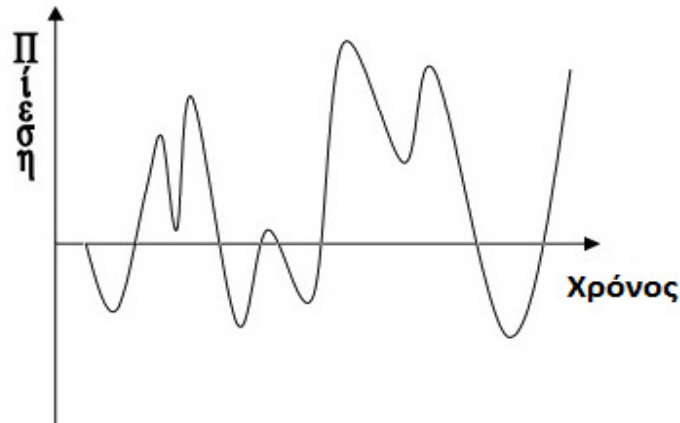
Στόχος αυτής της πτυχιακής εργασίας είναι η παρουσίαση της εξέλιξης των προτύπων κωδικοποίησης και συμπίεσης ήχου. Η ιστορική αναδρομή και ανάλυση των προτύπων από το πρότυπο κωδικοποίησης ήχου MPEG-1 ως και το σύγχρονο πρότυπο τρισδιάστατου ήχου MPEG-H. Επίσης, η υλοποίηση των προτύπων MPEG-1 και MPEG-4 και η παρουσίαση των χαρακτηριστικών των παραγόμενων σημάτων στο πεδίο της συχνότητας και στο πεδίο του χρόνου.

1.2 Αναπαράσταση Ηχητικών Σημάτων

Κάθε ήχος αποτελείται από ένα σύνολο μεταβολών πίεσης. Μια σχεδίαση της σχέσης της πίεσης με το χρόνο είναι η εικόνα 1.

Ένας απλός τρόπος δημιουργίας ενός τέτοιου διαγράμματος είναι να προσαρμόσουμε ένα κομμάτι μεταλλικό σύρμα σε ένα λεπτό φύλλο χαρτί και να τοποθετήσουμε ένα μαγνήτη σχετικά κοντά. Όταν αρχίσουμε να μιλάμε μπροστά στο χαρτί τότε οι μεταβολές της πίεσης που οφείλονται στη φωνή μας θα αναγκάσουν το χαρτί να κινείται εμπρός και πίσω μεταβάλλοντας τη θέση του μεταλλικού σύρματος προς το μαγνήτη. Η κίνηση αυτή θα

δημιουργήσει ένα μεταβαλλόμενο ηλεκτρικό ρεύμα, το οποίο θα είναι ανάλογο με τις μεταβολές της πίεσης του αέρα που δημιουργεί η φωνή μας. Ο μηχανισμός αυτός λειτουργεί και αντίστροφα. Ένα μεταβαλλόμενο ηλεκτρικό ρεύμα μπορεί να μεταβάλλει το μαγνητικό πεδίο του μαγνήτη μεταβάλλοντας και τη θέση του σύρματος ως προς το μαγνήτη και προκαλώντας κατά συνέπεια την κίνηση του χαρτιού. Αυτή η διάταξη προσομοιάζει τους πρώτους φωνογράφους, τις πρώτες συσκευές καταγραφής και αναπαραστάσης αναπαραγωγής ήχου.



Εικόνα 1 Μεταβολές πίεσης σε σχέση με το χρόνο

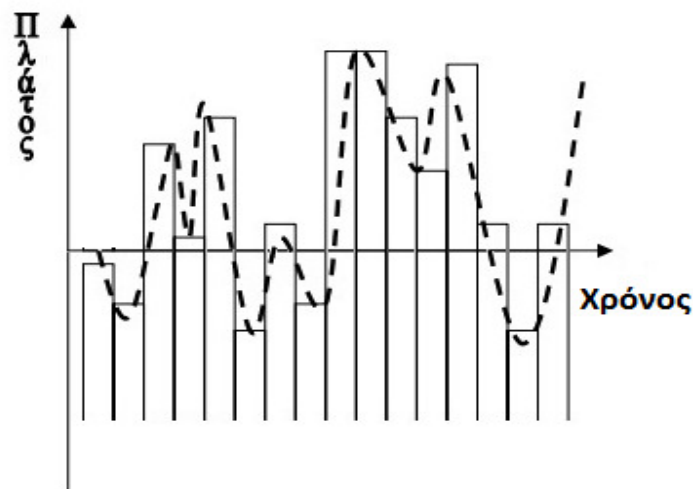
Στο φωνογράφο η βελόνα που κατέγραφε τον ήχο ήταν συνδεδεμένη με ένα είδος φωναγωγού με τρόπο ώστε να λαμβάνει μηχανικά τις δονήσεις που προκαλούσε ο ήχος που περνούσε μέσα σ' αυτόν. Η βελόνα παλλόταν κάθετα προς την επιφάνεια του κυλίνδρου και διαμόρφωνε το αλουμινόχαρτο δημιουργώντας "κορυφές και κοιλίες". Ο κύλινδρος περιστρεφόταν με το χέρι, με την βοήθεια μιας μανιβέλας, που ήταν προσαρμοσμένη στον άξονά του. Αργότερα ο Έντισον αντί για κυλίνδρους επικαλυμμένους με αλουμινόχαρτο άρχισε να χρησιμοποιεί κέρινους κυλίνδρους στις συσκευές του. Αυτό είχε ως αποτέλεσμα τη δυνατότητα διαγραφής μιας ηχογράφησης, αφαιρώντας την επιφάνεια του κέρινου κυλίνδρου με την χρήση αιχμηρής λεπίδας.

Με βάση αυτή την πρώτη συσκευή, ο επίσης Αμερικανός Emil Berliner (1851-1929), επινόησε δέκα χρόνια αργότερα μια άλλη συσκευή, το γραμμόφωνο (Gramophone), που αντί για κέρινο κύλινδρο χρησιμοποιούσε μια κυκλική πλάκα (δίσκο) από μίγμα ρητίνης γομαλάκας (shellac) για να καταγράψει τον ήχο.

Όπως είναι δυνατό να μετατρέψουμε τις μεταβολές πίεσης, που προέρχονται από έναν ήχο, σε αναλογικό σήμα, με ανάλογο τρόπο είμαστε σε θέση να μετατρέψουμε ένα μεταβαλλόμενο σήμα σε μια σειρά από διακριτές τιμές και αντίστροφα. Επειδή όμως η αναλογική και η διακριτή μορφή ενός ήχου είναι διαφορετικές, πάντα χάνουμε πληροφορία κατά την μετατροπή του συνεχούς σήματος σε διακριτό. Είναι σημαντικό λοιπόν να ελαχιστοποιηθεί η

πληροφορία που χάνεται κατά αυτή τη διαδικασία και τι μπορούμε να κάνουμε για να συμβεί αυτό.

Σε ψηφιακή μορφή ένα ηχητικό σήμα παριστάνεται από μια σειρά από αριθμούς, τους οποίους ονομάζουμε **δείγματα** και αντιστοιχούν στην πίεση του αέρα ή στην ηλεκτρική τάση σε διαδοχικές χρονικές στιγμές. Οι δυο αναπαραστάσεις απεικονίζονται στο παραπάνω διάγραμμα από το οποίο προκύπτουν και τα δυο κύρια προβλήματα της δειγματοληψίας ενός ηχητικού σήματος.



Εικόνα 2 Ψηφιακή και αναλογική αναπαράσταση ηχητικού σήματος

Η ψηφιακή αναπαράσταση αποτελείται από ορθογώνια σταθερού πλάτους, δηλαδή ίσης χρονικής διάρκειας. Αυτό είναι το πρώτο πρόβλημα, καθώς όσο πλατύτερο είναι το ορθογώνιο τόσο περισσότερη πληροφορία χάνεται. Το δεύτερο πρόβλημα είναι ότι οι τιμές του ψηφιακού σήματος είναι περιορισμένες. Κατά συνέπεια το σήμα λαμβάνει τιμές σε ένα πεπερασμένο σύνολο κάτι που σημαίνει ότι μπορεί και να υπάρχει σφάλμα ανάμεσα στην αναλογική τιμή και το δείγμα που λαμβάνεται. Αυτά τα προβλήματα είναι οι πηγές απωλειών στη μετατροπή ενός ήχου από αναλογική σε ψηφιακή μορφή. Οι αιτίες αυτές μπορούν να αντιμετωπιστούν αλλά δεν μπορούν ποτέ να εξαλειφθούν. Πρέπει λοιπόν, αναγνωρίζοντας τη δεδομένη ύπαρξή τους, να φροντίζουμε ώστε να διατηρούνται οι απώλειες σε ανεκτά επίπεδα.

Το κύριο πλεονέκτημα της ψηφιακής επεξεργασίας ήχου, είναι ότι η αντιγραφή ψηφιακών σημάτων δεν εισάγει επιπλέον θόρυβο στο σήμα. Από την άλλη μεριά η αναλογική επεξεργασία ήχου προσθέτει θόρυβο κατά τη διάρκεια της διαδικασίας αντιγραφής ενός σήματος από το ένα μέσο στο άλλο.

1.3 Κβάντιση

Στον ψηφιακό ήχο, λόγω της φύσης των υπολογιστών, πρέπει το συνεχές σήμα των τιμών του πλάτους να μετατραπεί σε διακριτή αναπαράσταση που μπορεί να αποθηκευτεί στον υπολογιστή. Η αποθήκευση πεπερασμένων ψηφίων (bits) πληροφορίας, οδηγεί σε αποθήκευση πληροφορίας με πεπερασμένη ακρίβεια. Ο θόρυβος κβάντισης είναι η κύρια πηγή παραμόρφωσης στη διαδικασία κωδικοποίησης ηχητικών σημάτων. Αυτή η παραμόρφωση έχει επίδραση στο αποτέλεσμα, με αποτέλεσμα να πρέπει να συμβιβαστεί η ποιότητα του ήχου με τον αριθμό των bits που αποθηκεύονται για κάθε δείγμα.

Τα δείγματα αποθηκεύονται με τη μορφή δυαδικών αριθμών. Οι αριθμητικές πράξεις και ο χειρισμός των bits είναι ευκολότερος για τον υπολογιστή μιας και η λογική λειτουργίας του είναι δυαδική. Έτσι μόνο με λογικές πράξεις και την αριστερή και δεξιά ολίσθηση εκτελούνται όλες οι επιθυμητές επεξεργασίες των δεδομένων.

Η κβάντιση είναι η χαρτογράφηση των συνεχών τιμών πλάτους σε κωδικοποιήσεις που αναπαριστώνται από πεπερασμένο αριθμό bits. Στη διανυσματική κβάντιση μια ομάδα συνεχών τιμών πλάτους κβαντίζεται σε μια ενοποιημένη κωδικοποίηση. Η διανυσματική κβάντιση είναι αποδοτική τακτική σε χαμηλούς ρυθμούς δειγματοληψίας.

Μια κβάντιση που καταλαμβάνει μήκος N bit αναπαριστά μέχρι 2^N κωδικοποιήσεις στάθμης. Η αποκβάντιση, είναι η χαρτογράφηση των N -bit κωδικοποιήσεων σε πλάτος σήματος. Τα πλάτη των σημάτων μπορεί να είναι τόσο θετικά όσο και αρνητικά, έτσι πρέπει να υπάρχουν κωδικοποιήσεις τόσο για θετικές όσο και για αρνητικές τιμές. Τυπικά, οι κβαντιστές είναι συμμετρικοί, έχουν δηλαδή ίσο αριθμό θετικών και αρνητικών σταθμών. Οι κβαντιστές επίσης διαιρούνται σε μηδενικής στάθμης και μη μηδενικής στάθμης, ανάλογα με το εάν υπάρχει κωδικοποιημένη στάθμη που να αντιστοιχεί στο μηδέν. Οι κβαντιστές μηδενικής στάθμης έχουν 2^1-1 κωδικοποιήσεις, ενώ οι κβαντιστές μη μηδενικής στάθμης έχουν 2^N κωδικοποιήσεις. Οι σχέσεις εισόδου-εξόδου των κβαντιστών φαίνεται στο παρακάτω σχήμα.



Εικόνα 3 Κβαντιστής μηδενικής στάθμης και κβαντιστής μη μηδενικής στάθμης

1.4 Εισαγωγή στην Ψυχοακουστική

Η Ψυχοακουστική είναι η επιστήμη που μελετά τη στατιστική σχέση μεταξύ ακουστικών ερεθισμάτων και της αίσθησης της ακοής και χρησιμοποιείται στο σχεδιασμό κωδικοποιητών αντίληψης. Η γενική ιδέα είναι ο θόρυβος κβαντισμού να τοποθετηθεί σε περιοχές του φάσματος του σήματος που επηρεάζει λιγότερο την πιστότητα του σήματος, ώστε ο ρυθμός

παροχής δεδομένων να μειώνεται χωρίς ακουστική παραμόρφωση. Η γνώση του πως λειτουργεί η ανθρώπινη ακοή μπορεί να βοηθήσει στο σχεδιασμό κωδικοποιήσεων ήχου.

Ο ήχος φτάνει στο ανθρώπινο αυτί σαν κύμα πίεσης. Η μονάδα μέτρησης της πίεσης είναι το Pascal, όπου $1 \text{ Pa} = 1 \text{ N/m}^2$. Οι σχετικές τιμές της πίεσης του ήχου για ηχητικές εφαρμογές κυμαίνονται ανάμεσα σε 10^{-5} Pa , όπου είναι το κάτω όριο της ανθρώπινης ακοής και τα 10^2 Pa που αντιστοιχεί στο όριο του πόνου.

Ορίζεται μια πλατιά περιοχή πιέσεων ήχου που συνήθως αναφέρεται σε λογαριθμική κλίμακα.

$$SPL = 10 \log\left(\frac{p}{p_0}\right)^2 \quad (1)$$

όπου $p_0 = 20 \mu\text{Pa}$ που είναι η ηχητική πίεση που αντιστοιχεί στο όριο ακοής για τονικές συχνότητες γύρω στα 2 KHz.

Επίσης ορίζεται η ένταση (I) του ήχου, που είναι η ισχύς ανά μονάδα επιφάνειας και είναι ανάλογο του p^2 .

$$SPL = 10 \log\left(\frac{I}{I_0}\right) \quad (2)$$

Η ένταση I μετριέται σε μονάδες W/m^2 . Η τιμή αναφοράς σε ένταση ήχου είναι $I_0 = 10^{-12} \text{ W/m}^2$ και αναφέρεται σε κύμα με πίεση αναφοράς p_0 .

Η επικάλυψη ηπιότερων ήχων από δυνατότερους είναι ένα φαινόμενο που αντιμετωπίζει κάθε άνθρωπος καθημερινά. Χαρακτηριστικό παράδειγμα είναι μια συζήτηση στο δρόμο που διακόπτεται, όταν από το δρόμο περνά ένα θορυβώδες φορτηγό, μιας και οποιαδήποτε συνομιλία θα επικαλυφθεί από την ισχύ του θορύβου της μηχανής του φορτηγού. Αυτό το φαινόμενο ονομάζεται **επικάλυψη στη συχνότητα** (tonal masking). Αυτό γίνεται λόγω της φυσιολογίας του αυτιού που δεν μπορεί να ακολουθήσει δυο ήχους γειτονικών συχνοτήτων αν υπάρχει μεγάλη διαφορά στην έντασή τους. Επικαλύψεις υπάρχουν και σε μη ταυτόχρονους ήχους όπως ένα δυνατό φωνήεν που επικαλύπτει ένα επακόλουθο σύμφωνο. Το φαινόμενο αυτό ονομάζεται **επικάλυψη στον χρόνο** (temporal masking).

Το ανθρώπινο αυτί έχει μια συγκεκριμένη φυσιολογία. Το έξω αυτί προσλαμβάνει τους ήχους μέσω του τυμπάνου και τους διοχετεύει στο μέσο αυτί. Το μέσο αυτί μεταφράζει την πίεση που μεταβιβάζεται από το τύμπανο σε κινήσεις υγρών στον κοχλία του εσωτερικού αυτιού. Ο κοχλίας μεταφράζει τις κινήσεις των υγρών σε ηλεκτρικά σήματα.

Το σύστημα ακοής χωρίζεται σε δυο περιοχές όπου γίνεται η επεξεργασία των ηχητικών ερεθισμάτων:

1. Η περιφερειακή περιοχή, όπου τα ερεθίσματα προ-επεξεργάζονται, αλλά διατηρούν το χαρακτήρα τους.
2. Τα αισθητήρια κελιά, που δημιουργούν την αίσθηση της ακοής με τη χρήση νευρικής επεξεργασίας.

Οι ανακλάσεις γίνονται στην περιφερειακή περιοχή (πτερύγιο, ακουστικό πόρος) και οι επικαλύψεις ηχητικών σημάτων στο εξωτερικό αυτί και το κανάλι προς το μέσο αυτί. Η αισθητηριακή επεξεργασία γίνεται στο εσωτερικό αυτί.



Εικόνα 4 Η φυσιολογία του αυτιού [10]

Υπάρχει ένα κάτω όριο ακουστότητας, το οποίο ονομάζεται κατώφλι ακουστότητας το οποίο καθορίζει την ένταση πάνω από την οποία ένα ηχητικό σήμα μπορεί να γίνει αντιληπτό απουσία άλλων ήχων. Ακόμα ορίζεται και ένα όριο στην αντίληψη διαφορών τονικού ύψους των ήχων. Αυτά τα όρια επηρεάζονται από διάφορες παραμέτρους όπως η ηλικία, η βιολογική προδιάθεση, η μουσική εκπαίδευση κ.α. , ενώ αυτά τα όρια εξαρτώνται και από το περιεχόμενο του ηχητικού σήματος.

Ο ήχος διαμορφώνεται από τον ανθρώπινο εγκέφαλο με έναν τρόπο που δεν είναι πλήρως γνωστός. Ο θάλαμος του εγκεφάλου, ελέγχει τα ηχητικά σήματα ώστε να αναγνωρίσει πιθανό κίνδυνο και να πυροδοτήσει αντανακλαστική αντίδραση. Σε περίπτωση που ανιχνεύσει κίνδυνο αποστέλλεται σήμα στην αμυγδαλή του εγκεφάλου που παράγει το συναίσθημα του φόβου και καθορίζεται η συμπεριφορά. Στην πλειονότητα όμως των περιπτώσεων, όπου δεν ανιχνεύεται κίνδυνος, το σήμα προχωρά στην παρεγκεφαλίδα όπου και αποκωδικοποιείται πλήρως. Στον κροταφικό λοβό γίνεται η πρόσληψη της μελωδίας και του τονικού ύψους του σήματος καθώς επίσης και του ηχοχρώματος και της χωρικής τοποθέτησης του. Το δεξί ημισφαίριο κατασκευάζει ένα περίγραμμα μελωδικής γραμμής και το αριστερό καταχωρεί τις τονικές σχέσεις, προσλαμβάνει και επεξεργάζεται την ομιλία.

1.4 Ποιοτική Μέτρηση Κωδικοποιήσεων Ήχου

Η κωδικοποίηση ήχου βασίζεται στην εξισορρόπηση των περιορισμών ρυθμού δεδομένων και πολυπλοκότητας συστημάτων σε σχέση με τις υψηλές απαιτήσεις για ήχο υψηλής ποιότητας. Η ποιότητα ήχου είναι το ζητούμενο στην κωδικοποίηση του ήχου, όμως είναι δύσκολο να τεθούν αντικειμενικά κριτήρια για την αξιολόγηση της. Αντικειμενικά μέτρα ποιότητας ήχου είναι ο λόγος ήχου προς θόρυβο ή η παραμόρφωση και είναι αντικειμενικά και μετρήσιμα όμως αγνοούν την ποιότητα που αντιλαμβάνεται το ανθρώπινο αυτί. Η αντιλαμβανόμενη

ποιότητα, μετράται βάσει μοντέλων ακοής τα οποία όμως είναι υποκειμενικό το κατά πόσο αντιπροσωπεύουν την αντιλαμβανόμενη ποιότητα. Συνεπώς, η ποιότητα ήχου βασίζεται ακόμα κατά κύριο λόγο σε επίσημες δοκιμασίες ακοής για να μετρηθεί η ποιότητα ήχου ενός συστήματος, παρά το γεγονός ότι και τα αντικειμενικά μέτρα θεωρούνται ικανοποιητικά. Το ανθρώπινο αυτί είναι ο κύριος κριτής της ποιότητας ήχου και συνεπώς είναι το τελικό και κυριότερο μέτρο.

Η ανεπάρκεια των απλών αντικειμενικών μέτρων ποιότητας έγινε σαφής από τη δεκαετία του '80 και το «θαύμα των 13db». Σε αυτό δυο επεξεργασμένα σήματα με SNR 13 db παρουσιάζονται στο κοινό. Στο ένα εισάγεται λευκός θόρυβος, ενώ στο άλλο θόρυβος σχεδιασμένος με βάση μοντέλα αντίληψης. Το αποτέλεσμα είναι ο λευκός θόρυβος να θεωρηθεί ενοχλητικός σαν συνεχόμενο θρόισμα. Ο ειδικά διαμορφωμένος θόρυβος προκαλεί παραμόρφωση που είναι ελαφρώς αισθητή ή καθόλου αισθητή καθώς υπήρχε μερική ή ολική επικάλυψη των απλών σημάτων. Το SNR και των δυο σημάτων είναι το ίδιο όμως η ποιότητα που αντελήφθησαν οι ακροατές είναι τελείως διαφορετική, με το σήμα με το λευκό θόρυβο να θεωρείται χειρότερης ποιότητας. Με αυτή τη δοκιμή κατέστη σαφές ότι τα αντικειμενικά μέτρα ποιότητας δεν αντικατοπτρίζουν την αντιλαμβανόμενη από τον άνθρωπο ποιότητα.

Η ποιότητα της κωδικοποίησης εξαρτάται τόσο στο ηχητικό υλικό που κωδικοποιείται, όσο και στο ρυθμό δεδομένων που χρησιμοποιείται. Διαφορετικά σήματα απαιτούν κωδικοποιητές με διαφορετικά χαρακτηριστικά. Λόγω αυτής της εξάρτησης της ποιότητας από το σήμα πηγής, κάθε αξιολόγηση ποιότητας απαιτεί μια κρίσιμη βάση αξιολόγησης. Η μείωση του ρυθμού δεδομένων μειώνει την ποιότητα της κωδικοποίησης παράγοντας ευρήματα παραμόρφωσης. Συνεπώς, κάθε αξιολόγηση ποιότητας θα πρέπει να γίνει συνεκτιμώντας τους ρυθμούς δεδομένων του κάθε κωδικοποιητή, καθώς όσο αυξάνεται ο ρυθμός δεδομένων τόσο καλύτερη αναμένεται να είναι η ποιότητα της κωδικοποίησης.

Η ποιοτική αξιολόγηση είναι απαραίτητη για τη συνολική αξιολόγηση ενός κωδικοποιητή ηχητικών σημάτων, αλλά είναι κρίσιμη και για το σχεδιασμό και τη ρύθμιση των διαφόρων επιπέδων. Ο σχεδιασμός ενός κωδικοποιητή ήχου απαιτεί πολλές αποφάσεις και επιλογές οι οποίες οριστικοποιούνται μετά από ελεγχόμενες ακροάσεις. Η εξάρτηση από τις ελεγχόμενες ακροάσεις αύξησε τη ζήτηση για εναλλακτικές μορφές αυτής της διαδικασίας. Οι επιλογές σχεδιασμού δίνονται με βάση απλές αντικειμενικές μετρήσεις ή άτυπες δοκιμασίες ακοής με λίγα αντικείμενα.

Η ποιότητα ήχου σε ένα σύστημα κωδικοποίησης συνδέεται με τη διαφορά ανάμεσα στην αισθανόμενη έξοδο του εξεταζόμενου συστήματος και ενός γνωστού σήματος αναφοράς. Η αξιολόγηση μπορεί να αφορά σήματα που είναι πιστά αντίγραφα του σήματος αναφοράς, αλλά και κάποια που έχουν πολύ μικρή ομοιότητα με το αρχικό σήμα.

Η διαφάνεια είναι μια βασική έννοια στην αξιολόγηση ποιότητας. Εάν ειδικοί ακροατές σημάτων στην αναγνώριση παραμορφώσεων δεν μπορούν να διαχωρίσουν το αρχικό και το υπό εξέταση σήμα, τότε αναφερόμαστε στο σύστημα κωδικοποίησης ως διαφανές. Το σήμα αναφοράς και το εξεταζόμενο παρέχονται σε τυχαία σειρά στους ακροατές. Εάν τα συστήματα κωδικοποίησης είναι διαφανή τότε οι ακροατές θα επιλέγουν λάθος κατά 50%.

Εάν ένα σύστημα είναι διαφανές το περιθώριο κωδικοποίησης αναφέρεται στο πόσο μακριά είναι ο κωδικοποιητής από την εισαγωγή αδυναμιών που είναι αισθητές στο αυτί. Ο έλεγχος γίνεται με ακουστικές δοκιμές όπου ο ρυθμός δεδομένων μειώνεται μέχρι οι ακροατές να αντιληφθούν τη διαφορά.

Εάν ένα σύστημα δεν είναι διαφανές, μας ενδιαφέρει πόσο ενοχλητικές είναι οι αδυναμίες για διάφορα δοκιμαστικά σήματα. Έτσι αποφασίζεται εάν ένας κωδικοποιητής είναι επαρκής για κάποια εφαρμογή. Γενικά αναζητείται το σύστημα να είναι διαφανές ή σχεδόν διαφανές. Σε αυτές τις περιπτώσεις μας ενδιαφέρει η αξιολόγηση των μικρών αδυναμιών με βάση την πενταβάθμια κλίμακα αδυναμιών που δείχνεται στον Πίνακα 1.

Ποιότητα	Αδυναμίες
5 Εξαιρετική	5 Ανεπαίσθητη
4 Καλή	4 Αισθητή, αλλά όχι ενοχλητική
3 Ικανοποιητική	3 Ελαφρά ενοχλητική
2 Φτωχή	2 Ενοχλητική
1 Κακή	1 Πολύ ενοχλητική

Πίνακας 1 Κλίμακα ακουστικής παραμόρφωσης

Κεφάλαιο 2

Πρότυπα Κωδικοποίησης Ήχου

2.1 Εισαγωγή

Η ψηφιακή κωδικοποίηση ήχου έγινε αναγκαία με την πρόοδο της τεχνολογίας και την αύξηση των απαιτήσεων του κοινού. Η διαδικασία εξελίχθηκε παράλληλα με την κωδικοποίηση της εικόνας και οι απαιτήσεις για υψηλότερη ποιότητα και στις δυο τεχνολογίες ήταν συνεχής. Η ποιότητα και η ομοιομορφία της δομής των κωδικοποιήσεων απαιτούσε τη σύσταση μιας επιτροπής προτυποποίησης των κωδικοποιήσεων. Η επιτροπή αυτή ιδρύθηκε στα πλαίσια του Τεχνικού Συμβουλίου Τεχνολογίας Πληροφορικής (Joint Technical Committee, JTC 1) με τη συνδρομή του Διεθνούς Οργανισμού Προτυποποίησης (International Standardization Organization, ISO) και της Διεθνούς Ηλεκτροτεχνικής Επιτροπής (International Electrotechnical Commission, IEC). Η επιτροπή ονομάστηκε Moving Pictures Expert Group (MPEG) και από το τέλος της δεκαετίας του «80 επιφορτίστηκε με την ανάπτυξη προτύπων για την κωδικοποιημένη αναπαράσταση κινούμενων εικόνων, ηχητικών σημάτων και του συνδυασμού τους.

Το πρώτο πρότυπο που αναπτύχθηκε από την επιτροπή ονομάστηκε MPEG-1 μετά από 3 χρόνια ερευνών. Το MPEG-1 Audio αφορά σε ηχητικά σήματα και αποτελεί το πρώτο διεθνές πρότυπο που καθορίζει το ψηφιακό format για υψηλής ποιότητας ήχο. Στόχος του προτύπου ήταν η μείωση του ρυθμού δεδομένων χωρίς τη μείωση της ποιότητας. Πρότυπα που δημιουργήθηκαν πριν το MPEG-1 αφορούσαν σε εφαρμογές μόνο για φωνητικά σήματα (ομιλία) ή απέδιδαν μέτρια ποιότητα ήχου. Η επιτυχία των προτύπων MPEG οδήγησε στην αποδοχή συμπιεσμένου ήχου υψηλής ποιότητας σε μια ευρεία γκάμα εφαρμογών, όπως η ψηφιακή μετάδοση και οι διαδικτυακές εφαρμογές. Το πιο ευρέως γνωστό πρότυπο MPEG είναι το MP3 (MPEG Layer III). Η εισαγωγή των προτύπων MPEG άλλαξε δραστικά το τοπίο στη ψηφιακή μετάδοση μουσικής.

Το πρότυπο MPEG-1 πρώτης φάσης είναι ένας κωδικοποιητής ενός ή δυο καναλιών σε στερεοφωνική μορφή με ρυθμούς δειγματοληψίας 32, 44.1 και 48 kHz. Στη δεύτερη φάση δόθηκε έμφαση στην υποστήριξη ήχου πολλών καναλιών και στην επέκταση σε χαμηλότερες συχνότητες δειγματοληψίας και ρυθμού bits.

Το πρότυπο MPEG-2 αποτελείται από δυο πρότυπα το MPEG-2 BC και το MPEG-2 AAC. Σε αντίθεση με το MPEG-2 BC που περιορίζεται από τη συμβατότητα με τα προηγούμενα επίπεδα του MPEG-1 (Backward Compatibility), το MPEG-2 AAC (Advanced Audio Coding) δεν έχει περιορισμούς και μπορεί να παρέχει καλύτερη επίδοση κωδικοποίησης.

Το τελευταίο ολοκληρωμένο πρότυπο είναι το MPEG-4 για κανάλια πολύ χαμηλού ρυθμού δεδομένων και χρησιμοποιείται σε διαδικτυακές και φορητές εφαρμογές.

2.2 Το πρότυπο MPEG -1

Η κωδικοποίηση MPEG-1 είναι ένα πρότυπο συμπίεσης για συγχρονισμένο ήχο και βίντεο σε ένα ρυθμό δεδομένων 1.5 Mb/s. Περιλαμβάνει προδιαγραφές για τα συστήματα, το βίντεο και τον ήχο. Το πρότυπο περιγράφει τη σύνταξη της κωδικοποιημένης ροής bit και τη διαδικασία αποκωδικοποίησης. Επίσης, υποδείγματα προγραμματισμού και ένα σύνολο διανυσμάτων δειγμάτων για την αξιολόγηση της συμμόρφωσης του αποκωδικοποιητή παρέχονται επίσης από τις προδιαγραφές του προτύπου. Η δομή του MPEG-1 κωδικοποιητή ήχου δεν είναι επιβεβλημένη από τις προδιαγραφές του προτύπου και η περιγραφή του γίνεται συμβουλευτικά από το πρότυπο. Το γεγονός ότι η σύνταξη και η αποδιαμόρφωση ακολουθούν υποχρεωτικά μια συγκεκριμένη προτυποποιημένη διαδικασία εξασφαλίζει την πλήρη συμβατότητα με όλες τις συσκευές, όμως η ανάπτυξη του κωδικοποιητή επαφίεται στους σχεδιαστές του συστήματος επιτρέποντας μεγάλη ελευθερία εντός των ορίων των προδιαγραφών του προτύπου. Το πρότυπο MPEG-1 περιγράφει έναν αλγόριθμο κωδικοποίησης με βάση αντιληπτικά κριτήρια σχεδιασμένο για γενικά ηχητικά σήματα.

Εξειδικευμένα μοντέλα πηγών, όπως για σήματα ομιλίας, δεν υπάρχουν. Η μόνη υπόθεση εργασίας είναι ότι τα στατιστικά του σήματος εισόδου είναι σχεδόν στάσιμα, δηλαδή θεωρούμε ότι δεν αλλάζουν σημαντικά στο πέρασμα του χρόνου. Το ηχητικό σήμα αναπαρίσταται από το φασματικό του περιεχόμενο ανά δείγμα και κωδικοποιείται με βάση τα μοντέλα αντίληψης. Στόχος είναι να παρέχει ένα σχήμα κωδικοποίησης που με βάση την αντίληψη δεν θα έχει ακουστές απώλειες.

2.1.1 Κύρια χαρακτηριστικά του προτύπου MPEG-1

Οι συχνότητες δειγματοληψίας που υποστηρίζονται από το πρότυπο MPEG-1 είναι 32, 44.1 και 48 kHz. Τα κανάλια μπορεί να είναι ένα ή δύο. Υποστηρίζεται μονοφωνικός ήχος με ένα κανάλι, διπλός μονοφωνικός ήχος για δυο κανάλια που είναι ανεξάρτητα και στερεοφωνικός ήχος για στερεοφωνικά κανάλια που διαμοιράζει τα διαθέσιμα bit στα δυο κανάλια, αλλά δεν εκμεταλλεύεται κάποιο άλλο χωρικό μοντέλο αντίληψης. Ακόμα υποστηρίζονται τύποι κοινού στερεοφωνικού ήχου που εκμεταλλεύονται τη συσχέτιση μεταξύ των δυο στερεοφωνικών καναλιών.

Οι ρυθμοί δεδομένων ποικίλουν μεταξύ 32 και 224 kb/s ανά κανάλι, προσφέροντας λόγους συμπίεσης από 2,7 έως 24:1 ανάλογα με το ρυθμό δειγματοληψίας. Επίσης, δίνεται η δυνατότητα για επιπλέον σταθερούς ρυθμούς δεδομένων.

Το πρότυπο MPEG-1 ορίζει 3 επίπεδα (Layers) με αυξανόμενα υψηλότερη ποιότητα ήχου και ελαφρώς αυξανόμενη πολυπλοκότητα. Τα πρώτα δυο επίπεδα μοιράζονται τη βασική δομή κωδικοποίησης που έχει τις ρίζες της σε ένα παλαιότερο αλγόριθμο που ονομάζεται MUSICAM (Masking pattern adapted Universal Subband Integrated Coding And Multiplexing) [34] με το

τρίτο επίπεδο να είναι σημαντικά διαφορετικό. Το τρίτο δημιουργήθηκε από τη συνένωση του προτύπου ASPEC (Adaptive Spectral Perceptual Entropy Coding) [35] με τα πρώτα δυο επίπεδα του MPEG-1, ώστε ο αποκωδικοποιητής του επιπέδου III να μπορεί να αποκωδικοποιεί συμβολοσειρές που παρήχθησαν με τα επίπεδα I και II.

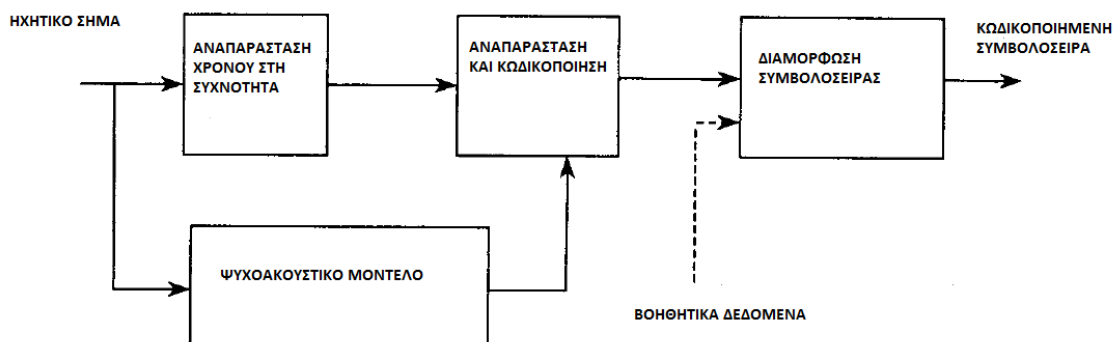
Το επίπεδο I είναι το απλούστερο και λειτουργεί με ρυθμούς δεδομένων μεταξύ 32 και 224 kb/s ανά κανάλι, και το προτιμότερο είναι ο ρυθμός δεδομένων λειτουργίας να υπερβαίνει τα 128 kb/s. Το επίπεδο I βρίσκει εφαρμογή στην ψηφιακή κασέτα στα 192 kb/s ανά κανάλι.

Το επίπεδο II είναι μέτριας πολυπλοκότητας και χρησιμοποιεί ρυθμούς δεδομένων μεταξύ 32 και 192 kb/s ανά κανάλι. Στα 128 kb/s η ποιότητα θεωρείται πολύ καλή. Η κύρια εφαρμογή του επιπέδου II είναι η μετάδοση ψηφιακού ήχου.

Το επίπεδο III είναι η πολυπλοκότερη και καλύτερη σε ποιότητα, με ρυθμούς δεδομένων να κυμαίνονται μεταξύ 32 και 160 kb/s ανά κανάλι. Το επίπεδο III αποδίδει εξαιρετική ποιότητα για ρυθμούς δεδομένων άνω των 128 kb/s ανά κανάλι. Η πιο σημαντική εφαρμογή του επιπέδου III ήταν παλαιότερα η μετάδοση ήχου σε ISDN γραμμές και σήμερα σε διαδικτυακές εφαρμογές. Η τροποποίηση του επιπέδου III σε χαμηλότερες συχνότητες δειγματοληψίας οδήγησε στο format του MP3.

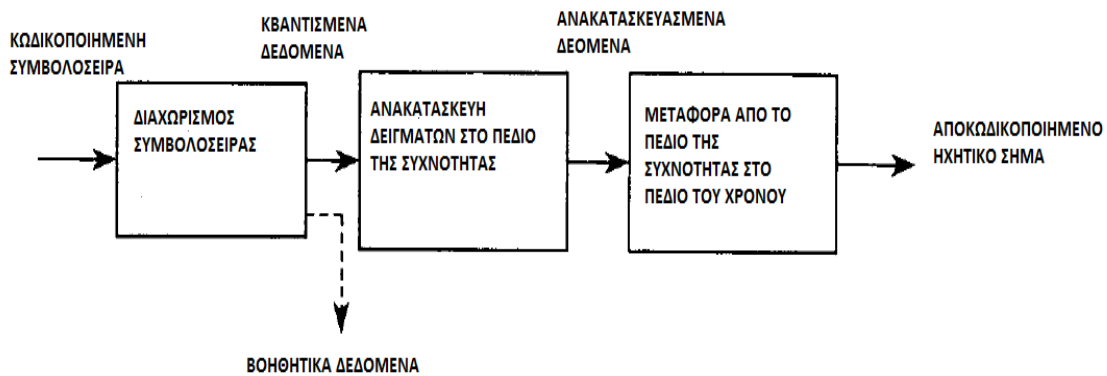
Κοινό σημείο παρά τις διαφορές είναι η υλοποίηση σε ένα chip πραγματικού χρόνου αποκωδικοποιητών και για τα 3 επίπεδα. Επίσης, και τα τρία επίπεδα δίνουν τη δυνατότητα να συμπεριληφθούν δεδομένα στη συμβολοσειρά της κωδικοποίησης.

Τα κύρια δομικά στοιχεία του σχήματος κωδικοποίησης MPEG-1 περιλαμβάνουν ένα στάδιο αναπαράστασης από το πεδίο του χρόνου στο πεδίο της συχνότητας ακολουθούμενο από ένα στάδιο μετάφρασης σε bit. Το σήμα εισόδου τροφοδοτεί ένα ψυχοακουστικό μοντέλο, του οποίου η έξοδος καθορίζει την ακρίβεια του σταδίου αναπαράστασης. Το στάδιο διαμόρφωσης της συμβολοσειράς συνδυάζει την αναπαράσταση των κβαντισμένων δεδομένων με συμπληρωματικές πληροφορίες και προαιρετικά βοηθητικά δεδομένα.



Εικόνα 5 Σχηματικό διάγραμμα κωδικοποιητή MPEG-1

Ο αποκωδικοποιητής μεταφράζει τη συμβολοσειρά, αποκαθιστά τα κβαντισμένα φασματικά στοιχεία του σήματος και τελικά ανακατασκευάζει την αναπαράσταση του ηχητικού σήματος στο πεδίο του χρόνου από την αναπαράσταση του στο πεδίο της συχνότητας.

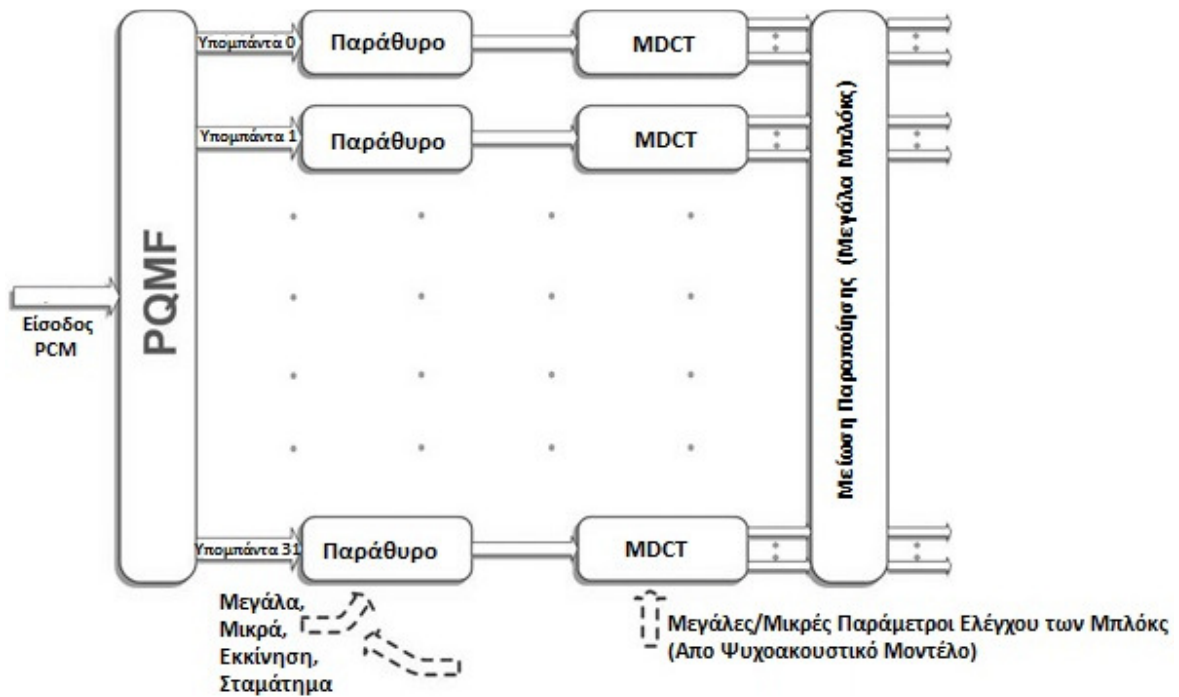


Εικόνα 6 Σχηματικό διάγραμμα αποκωδικοποιητή MPEG-1

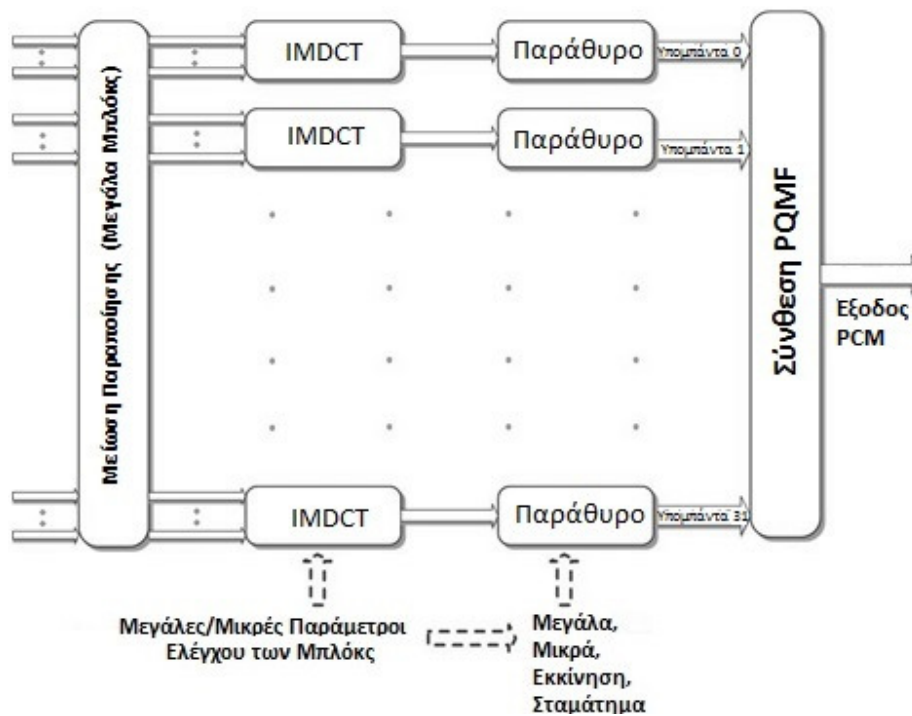
2.1.2 Απεικόνιση χρόνου στη συχνότητα

Ένα φίλτρο PQMF (Pseudo-Quadrature Mirror Filter) [36] είναι το βασικό κομμάτι της μετατροπής από το πεδίο του χρόνου στο πεδίο της συχνότητας, για όλα τα επίπεδα MPEG. Το φίλτρο διαιρεί το φάσμα συχνοτήτων σε 32 ισομερείς συχνοτικές (υπο)περιοχές (subbands). Στα επίπεδα I και II η έξοδος του PQMF αναπαριστά τα φασματικά δεδομένα του σήματος προς κβάντιση. Η ανάλυση συχνότητας στα επίπεδα I και II είναι 750 Hz για ρυθμό δειγματοληψίας 48 kHz. Στο επίπεδο III και προκειμένου να αυξηθεί η ανάλυση του φίλτρου, το PQMF ακολουθείται από έναν τροποποιημένο μετασχηματισμό συνημιτόνου (MDCT, Modified Discrete Cosine Transform) [37] με 18 γραμμές συχνότητας, οπότε προκύπτουν 576 συνολικά κανάλια συχνότητας.

Το υβριδικό φίλτρο του επιπέδου III, από τις 32 υποπεριοχές συχνοτήτων του φίλτρου PQMF, σχηματίζει πακέτα 36 δειγμάτων για κάθε συχνότητα, με επικάλυψη 50%. Τα δείγματα αυτά πολλαπλασιάζονται από ένα παράθυρο ημιτόνου και επεξεργάζονται από τον μετασχηματισμό MDCT. Επιπρόσθετα, με την παραμόρφωση της συχνότητας που εισάγεται από το PQMF, ο μετασχηματισμός MDCT εισάγει και παραμόρφωση στο χρόνο, η οποία όμως εξουδετερώνεται σε συνεχόμενα πακέτα χρόνου όταν δεν υπάρχει κβάντιση στη φάση επικάλυψης – άθροισης (overlap-add) της αποκωδικοποίησης. Για το λόγο αυτό ο μετασχηματισμός MDCT ονομάζεται Oddly-stacked Time-Domain Aliasing Cancellation (OTDAC) [38]. Με σκοπό τη μείωση των ανεπιθύμητων στοιχείων που εισάγονται από την επικάλυψη συχνοτήτων στο PQMF, για μεγάλα πακέτα το φίλτρο του επιπέδου III πολλαπλασιάζει την έξοδο του MDCT με συντελεστές που μειώνουν την παραμόρφωση της συχνότητας του σήματος.



Εικόνα 7 Σχηματικό διάγραμμα του φίλτρου κωδικοποίησης (ανάλυσης) του MPEG-1, επίπεδο III.



Εικόνα 8 Σχηματικό διάγραμμα φίλτρου αποκωδικοποίησης (σύνθεσης) του MPEG-1, επίπεδο III

Στην αποκωδικοποίηση γίνεται η αντίστροφη διαδικασία μείωσης της παραμόρφωσης, η οποία εφαρμόζεται πριν από το φίλτρο IMDCT, ώστε να παρέχει τα σωστά δείγματα των υποπεριοχών συχνοτήτων στο φίλτρο σύνθεσης PQMF. Ένα καθαρό ημιτονικό σήμα στο οποίο γίνεται επεξεργασία από το υβριδικό φίλτρο PQMF/MDCT χωρίς μείωση της παραμόρφωσης, μπορεί να παράγει ακουστές παραμορφώσεις (audible artifacts) της τάξης των -12 dB σε σχέση με το αρχικό σήμα. Η παραμόρφωση μπορεί να μειώνεται, όμως δεν εξαφανίζεται εντελώς, καθώς μπορεί να υπάρχουν εναπομείναντα στοιχεία παραμόρφωσης. Στο στάδιο της σύνθεσης, το φίλτρο IMDCT εφαρμόζεται πριν την ανακατασκευή PQMF. Μετά την αντίστροφη κβάντιση των φασματικών στοιχείων και όπου είναι εφικτό, στη στερεοφωνική επεξεργασία του σήματος εφαρμόζεται η αντίστροφη μείωση της παραμόρφωσης. Έπειτα, το φίλτρο IMDCT εφαρμόζεται ακολουθούμενο από τη διαδικασία παραθυροποίησης, όπου τα παράθυρα δειγμάτων που εφαρμόζονται είναι τα ίδια με τα παράθυρα στη φάση της ανάλυσης, δηλαδή της κωδικοποίησης. Το πρώτο μισό του τρέχοντος παραθυροποιημένου πακέτου επικαλύπτεται και προστίθεται με το δεύτερο μισό των παραθυροποιημένων δειγμάτων του προηγούμενου πακέτου, δηλ. εφαρμόζεται επικάλυψη μεταξύ των διαδοχικών παραθύρων ίση με 50%.

2.1.3 Ψυχοακουστικό μοντέλο του MPEG-1

Ο στόχος του προτύπου MPEG είναι να παρέχει κωδικοποιημένο σήμα με ποιότητα τέτοια ώστε οι απώλειες λόγω συμπίεσης να μην είναι αντιληπτές από τον ακροατή. Αυτό σημαίνει να μην είναι δυνατός ο διαχωρισμός του κανονικού (γνήσιου) από το κωδικοποιημένο σήμα. Ο στόχος είναι να μειωθεί ο ρυθμός δεδομένων όσο το δυνατόν περισσότερο, κάτι που για να συμβεί εκμεταλλευόμαστε ψυχοακουστικές αρχές και μοντέλα προσομοίωσης της ανθρώπινης ακοής. Κατά την κωδικοποίηση, το σήμα εισόδου αναλύεται ανά παράθυρο και υπολογίζεται η ικανότητα επικάλυψης των στοιχείων του σήματος. Για κάθε παράθυρο δειγμάτων υπολογίζονται αρχικά οι φασματικοί συντελεστές και το όριο (κατώφλι) επικάλυψης. Ακολουθώντας, τα bits που είναι διαθέσιμα κατανέμονται μέσω μίας διαδικασίας (bit allocation) που λαμβάνει υπόψη το φάσμα του σήματος και το κατώφλι επικάλυψης και με στόχο να ελαχιστοποιείται το ακουστικό σφάλμα. Η διαδικασία κωδικοποίησης μπορεί να μην είναι προκαθορισμένη από το πρότυπο, καθώς περιγράφονται δυο ψυχοακουστικά μοντέλα στις προδιαγραφές του προτύπου και παρέχεται μεγάλος βαθμός ελευθερίας στην εφαρμογή του επιθυμητού ψυχοακουστικού μοντέλου. Σε υψηλούς ρυθμούς δεδομένων το ψυχοακουστικό μοντέλο μπορεί να παρακαμφθεί (αγνοηθεί) πλήρως, αφήνοντας τον διαμοιρασμό των διαθέσιμων πόρων (bits) σε μια επαναληπτική διαδικασία τοποθέτησης βασισμένη στην ισχύ των φασματικών συντελεστών του σήματος.

Η είσοδος στο ψυχοακουστικό μοντέλο είναι η χρονική αναπαράσταση του ηχητικού σήματος σε κάποιο χρονικό διάστημα και οι αντίστοιχες έξοδοι είναι οι λόγοι σήματος προς επικάλυψη (Signal to Masking Ratio, SMR) για τους διαχωρισμούς συχνοτήτων του κωδικοποιητή. Με βάση αυτές τις πληροφορίες, η τοποθέτηση των bits (bit allocation) και η μορφοποίηση του θορύβου κβαντισμού (noise shaping) υπολογίζεται για κάθε παράθυρο δεδομένων εισόδου. Η ακριβής συχνοτική αναπαράσταση του σήματος εισόδου υπολογίζεται από ένα διακριτό μετασχηματισμό Fourier παράλληλα με την μετατροπή του ηχητικού σήματος από το χρόνο

στη συχνότητα. Η έξοδος του PQMF ή του υβριδικού φίλτρου μπορεί να χρησιμοποιηθεί για να απλοποιηθεί η δομή του αλγορίθμου. Η εκτίμηση του κατωφλίου επικάλυψης έχει σκοπό τη μέγιστη ακρίβεια της αναπαράστασης του σήματος. Ανεπαρκής ανάλυση στη συχνότητα και πιθανή παραμόρφωση στη συχνότητα, μπορούν να επηρεάσουν ανεπανόρθωτα τις αξιολογήσεις του ψυχοακουστικού μοντέλου.

Το πρώτο βήμα και στα δυο ψυχοακουστικά μοντέλα είναι να ευθυγραμμιστούν χρονικά τα ηχητικά δεδομένα που χρησιμοποιούνται από το στάδιο ψυχοακουστικών μοντέλων με τα δεδομένα του ηχητικού σήματος. Αυτή η διαδικασία λαμβάνει υπόψη της την καθυστέρηση λόγω του φίλτρου PQMF και τη χρονική μετατόπιση που χρειάζεται έτσι ώστε το παράθυρο ψυχοακουστικής ανάλυσης να εστιάζει στο τρέχον πακέτο (παράθυρο) δεδομένων προς κωδικοποίηση. Για παράδειγμα, στο επίπεδο I η καθυστέρηση από το φίλτρο είναι 256 δείγματα και το πακέτο προς κωδικοποίηση έχει μήκος 384 δείγματα. Το παράθυρο ανάλυσης που εφαρμόζεται στο επίπεδο I στο ψυχοακουστικό μοντέλο 1 έχει μήκος 512 δείγματα. Η απόκλιση που πρέπει να εφαρμοστεί για να είναι χρονικά ευθυγραμμισμένο είναι $256 + (512 - 384)/2 = 320$ δείγματα.

2.1.3.1. Ψυχοακουστικό μοντέλο 1

Το παράθυρο ανάλυσης είναι ένα παράθυρο Hanning μήκους N ίσο με 512 δείγματα για το επίπεδο I και 1024 δείγματα για τα επίπεδα II και III. Η επικάλυψη μεταξύ συνεχόμενων παραθύρων είναι N/16. Αφού τα επίπεδα II και III χρησιμοποιούν μήκος διανύσματος δειγματοληψίας ίσο με 1152 δείγματα, το γεγονός ότι το παράθυρο ανάλυσης δειγμάτων έχει μήκος 1024 δείγματα σημαίνει ότι δεν καλύπτει όλα τα δεδομένα ήχου στο διάνυσμα. Αυτό έχει ως αποτέλεσμα εάν συμβεί κάποια αλλαγή στο κομμάτι του διανύσματος που δεν καλύπτεται η αλλαγή τότε δεν θα επηρεάσει το ψυχοακουστικό μοντέλο 1 καθώς θα είναι εκτός του παραθύρου ανάλυσης. Τελικά όμως το μήκος των 1024 δειγμάτων για παράθυρο ανάλυσης κρίνεται σαν ένας επαρκής συμβιβασμός.



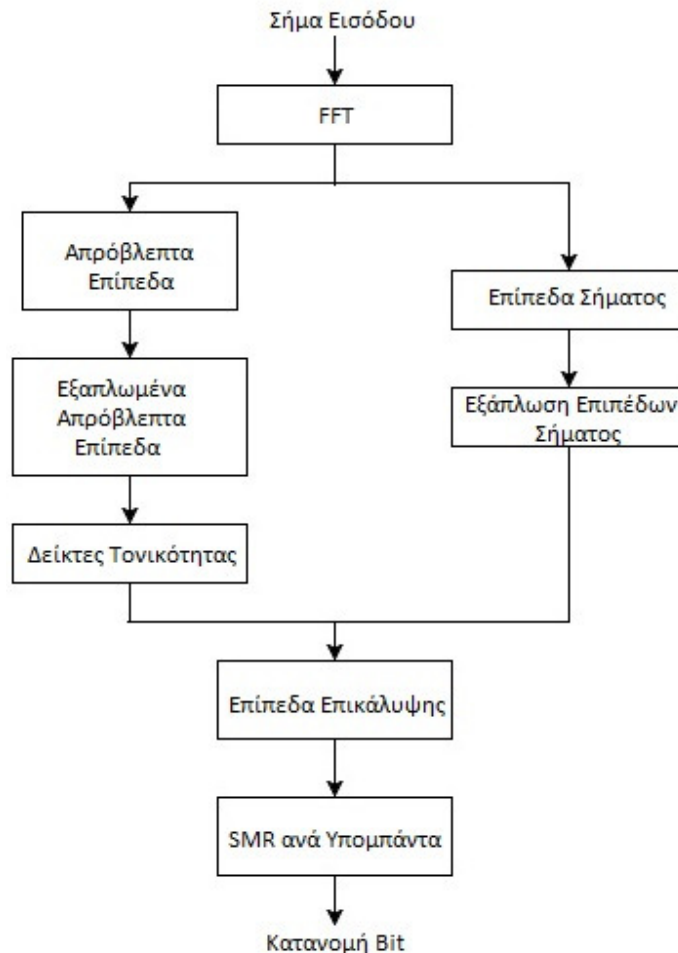
Εικόνα 9 Σχηματικό διάγραμμα ψυχοακουστικού μοντέλου 1

2.1.3.2. Ψυχοακουστικό μοντέλο 2

Το ψυχοακουστικό μοντέλο 2 σε διαδικασία και υπολογισμούς διαφέρει σημαντικά από το μοντέλο 1. Υπάρχουν δυο παράλληλα μονοπάτια, το πρώτο αφορά τον υπολογισμό της ενέργειας επικάλυψης και το δεύτερο τον υπολογισμό του δείκτη τονικότητας (tonal index).

Το μοντέλο 2 εφαρμόζει επίσης παράθυρο Hanning για το μετασχηματισμό Fourier FFT στα χρονικά ευθυγραμμισμένα δείγματα του σήματος. Το μέγεθος του παραθύρου του FFT είναι 1024 για όλα τα επίπεδα, όμως για τα επίπεδα II και III αυτό το μοντέλο υπολογίζει δυο FFT για κάθε διάνυσμα. Τα κέντρα των πακέτων εισόδου FFT είναι ευθυγραμμισμένα με το κέντρο του πρώτου και του δεύτερου μισού του πακέτου του ηχητικού σήματος. Το μοντέλο 2 χρησιμοποιεί την έξοδο του μετασχηματισμού FFT για τον υπολογισμό των καμπυλών επικάλυψης και τους αντίστοιχους λόγους για τις υποπεριοχές (subbands) του κωδικοποιητή. Επιλέγεται ο υψηλότερος από τους δυο λόγους σήματος προς επικάλυψη (SMR, Signal to

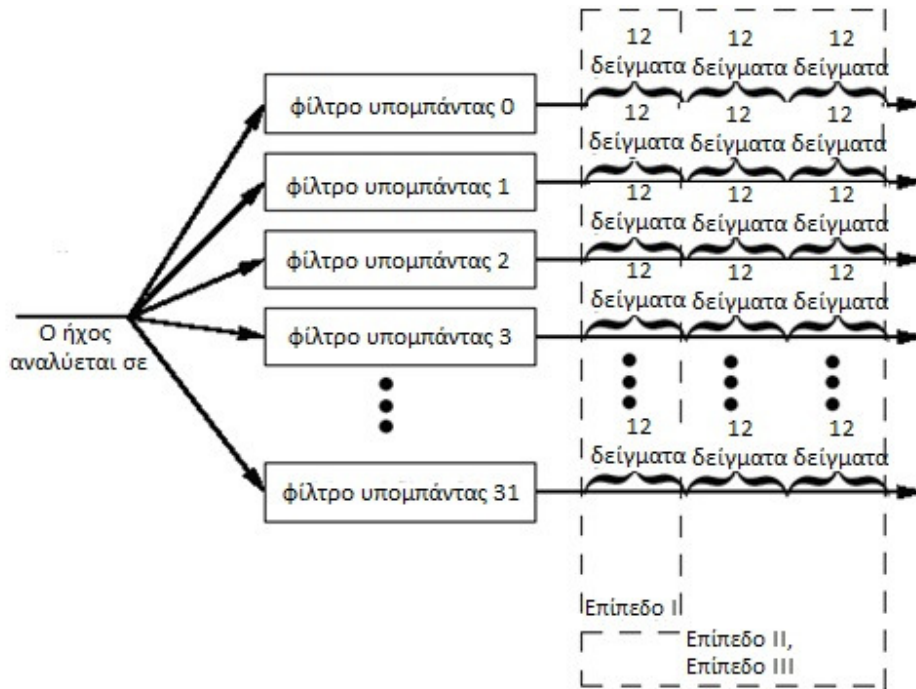
Masking Ratio) ή το χαμηλότερο από τα δυο όρια επικάλυψης (masking thresholds) για κάθε υποπεδίο συχνοτήτων.



Εικόνα 10 Σχηματικό μοντέλο ψυχοακουστικού μοντέλου 2

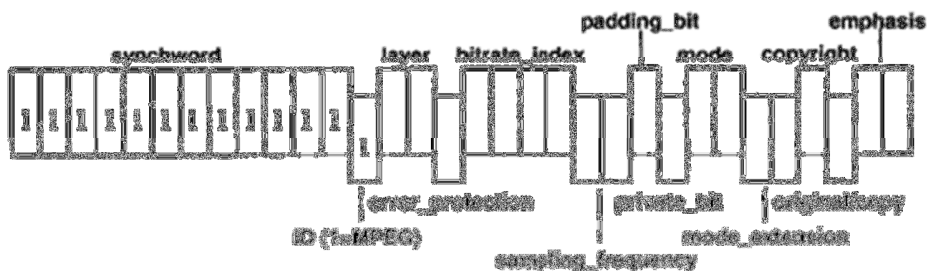
2.1.4 Δομή ηχητικού σήματος κατά το πρότυπο MPEG-1

Μια συμβολοσειρά ήχου MPEG περιέχει περιοδικά πακέτα (packets) με κεφαλίδες (headers) για να αναγνωριστούν τα κωδικοποιημένα δεδομένα. Στο επίπεδο I το μέγεθος πακέτου αποτελείται από 12 επί 32, δηλαδή 384 δείγματα. Στα επίπεδα II και III τρία τμήματα των 12 δειγμάτων για κάθε υποπεδίο συχνοτήτων ενώνονται σε ένα πλαίσιο μήκους 1152 δειγμάτων. Η γενική δομή του πλαισίου δεδομένων MPEG-1 για τα επίπεδα I, II και III δείχνεται στην Εικόνα 11.



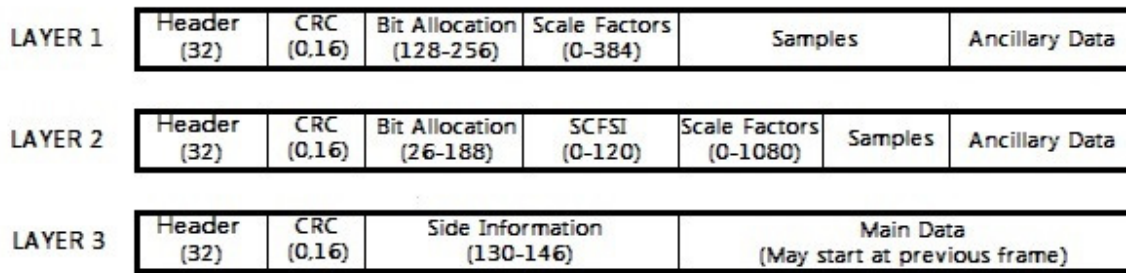
Εικόνα 11 Δομή του πλαισίου δεδομένων του ήχου MPEG-1 για τα επίπεδα I, II και III

Μια κεφαλίδα (header) MPEG-1 ήχου έχει 32 bit μήκος. Αρχικά υπάρχουν 12 bit αποτελούμενα από «1» που ονομάζονται λέξη χρονισμού, αμέσως μετά υπάρχει το bit MPEG ID, που αν είναι «1» υποδηλώνει συμβολοσειρά MPEG-1 και το χαρακτηριστικό επιπέδου που αποτελείται από δυο bits. Το bit προστασίας σφάλματος, τα 4 bits του δείκτη bitrate, τα δυο bits της συχνότητας δειγματοληψίας, τα bits μορφοποίησης, το bit ιδιωτικότητας και μορφής (στερεοφωνική, μικτή στερεοφωνική, δυο καναλιών, μονού καναλιού) και άλλων ειδών bits συμπληρώνουν την κεφαλίδα. Η δομή της κεφαλίδας δίνεται στην Εικόνα 12.



Εικόνα 12 Κεφαλίδα ήχου MPEG-1

Η δομή του πακέτου για τα τρία διαφορετικά επίπεδα (I, II και III) φαίνεται στην Εικόνα 13.



Εικόνα 13 Δομή πακέτων ήχου για τα 3 επίπεδα του προτύπου MPEG-1

Η δομή των πακέτων αλλάζει μετά την κεφαλίδα και τον έλεγχο ακεραιότητας CRC. Στο επίπεδο I ακολουθεί η κατανομή των bits, οι παράγοντες κλίμακας, τα δείγματα και τέλος τα βοηθητικά δεδομένα. Στο επίπεδο II ακολουθεί η κατανομή των bits, το πεδίο SCFSI που είναι ένα τεχνικό πεδίο με flags για τον έλεγχο του πακέτου, οι παράγοντες κλίμακας, τα δείγματα και τα βοηθητικά δεδομένα. Στο επίπεδο III ακολουθούν βοηθητικά δεδομένα και τα κυρίως δεδομένα που μπορεί να έχουν ξεκινήσει σε προηγούμενο πλαίσιο. Αυτό κάνει προφανές ότι το επίπεδο III έχει πιο απλή δομή που παρέχει όμως περισσότερη ευελιξία.

2.1.5 Στερεοφωνική κωδικοποίηση ήχου

Παρά το γεγονός ότι στα περισσότερα στερεοφωνικά σήματα υπάρχει πολύ μικρή συσχέτιση μεταξύ της χρονικής αναπαράστασης του αριστερού και του δεξιού καναλιού, εντούτοις υπάρχει ισχυρή συσχέτιση στα αντίστοιχα φάσματα ισχύος. Σχετικά με το στερεοφωνικό ήχο είναι γνωστό ότι το ανθρώπινο αυτί έχει μειωμένη ικανότητα να αναγνωρίσει την τοποθεσία της πηγής του ήχου στις υψηλές συχνότητες.

Γενικά, η κοινή κωδικοποίηση αριστερού και δεξιού καναλιού παρέχει υψηλότερα κέρδη κωδικοποίησης, λόγω της ισχυρής συσχέτισης στα φάσματα ισχύος, όπως αναφέρθηκε παραπάνω. Ωστόσο υπάρχουν και περιπτώσεις στις οποίες η στερεοφωνική κωδικοποίηση απαιτεί ρυθμούς δεδομένων που υπερβαίνουν το διπλάσιο της κωδικοποίησης ενός μονού καναλιού. Συνεπώς, κάποιες αδυναμίες στην κωδικοποίηση μονού καναλιού γίνονται αισθητές από το ανθρώπινο αυτί μόνο όταν ένα στερεοφωνικό σήμα κωδικοποιείται σαν διπλό κανάλι.

Αυτό το φαινόμενο σχετίζεται με τις διαφορές ανάμεσα στο όριο επικάλυψης που υπολογίστηκε όταν το σήμα ηχογραφήθηκε σαν σήμα μονού καναλιού και το όριο επικάλυψης όταν το σήμα παρουσιάζεται σε συνθήκες ακρόασης από δυο ανθρώπινα αυτιά. Η διαφορά ονομάζεται στάθμη διαφοράς αμφιωτικής επικάλυψης (Binaural Masking Level Difference, BMLD) [39, 40] και εμφανίζεται κυρίως στις χαμηλές συχνότητες. Παράδειγμα αυτού του φαινομένου σχετίζεται με το «πρόβλημα του cocktail party» (Cocktail party effect) [41], όπου ο ακροατής μπορεί να εστιάσει σε μια συζήτηση παρά την ύπαρξη ενός ισχυρότερου θορύβου στο περιβάλλον. Εάν ο ακροατής καλύψει το ένα του αυτί, τότε είναι δυσκολότερο να παρακολουθήσει τη συζήτηση. Αυτό συμβαίνει επειδή η επικάλυψη του ήχου της συζήτησης είναι χαμηλότερη όταν χρησιμοποιούνται και τα δυο αυτιά από όταν χρησιμοποιείται μόνο το

ένα. Η εφαρμογή αυτών των ευρημάτων έδειξε ότι οι αδυναμίες στην κωδικοποίηση λόγω επικάλυψης σε μονό κανάλι ήχου μπορούν να εξαφανιστούν όταν χρησιμοποιούνται δυο κανάλια.

Η ιδέα στην κοινή στερεοφωνική κωδικοποίηση είναι η κατάλληλη περιστροφή του στερεοφωνικού πεδίου για κάθε κρίσιμη συχνότητα (critical frequency) [42] στην κύρια κατεύθυνση του στερεοφωνικού σήματος εισόδου. Η εφαρμογή αυτής της ιδέας γίνεται με την μεταφορά παράπλευρων δεδομένων ώστε να μεταφερθούν και οι παράμετροι κατεύθυνσης. Αυτό μερικές φορές μπορεί να σημαίνει μεγάλη ποσότητα παράπλευρης πληροφορίας, χωρίς κέρδος στην κωδικοποίηση. Απλοποιημένες προσεγγίσεις αυτής της εφαρμογής είναι η στερεοφωνική κωδικοποίηση μέσου/αθροίσματος (M/S) και η στερεοφωνική κωδικοποίηση έντασης (Intensity Stereo Coding) [43].

Η στερεοφωνική κωδικοποίηση μέσου/αθροίσματος εστιάζει στην αφαίρεση πλεονασματικών στοιχείων και είναι μια διαδικασία με συγκεκριμένα αποτελέσματα αν αγνοήσουμε τα σφάλματα κβάντισης και κωδικοποίησης. Στην κωδικοποίηση M/S μόνο δυο κατευθύνσεις λογίζονται. Αντί της μετάδοσης ανεξάρτητα του αριστερού (L) και του δεξιού (R) καναλιού, μεταδίδονται ο κανονικοποιημένος μέσος όρος των σημάτων $M=L+R$ καθώς και η διαφορά τους $S=L-R$. Αυτά τα σήματα ονομάζονται επίσης μέσο σήμα (M) και σήμα διαφοράς (S). Το κέρδος από την εφαρμογή αυτής της μεθόδου εξαρτάται από το σήμα. Το μέγιστο κέρδος επιτυγχάνεται όταν το αριστερό και το δεξί κανάλι είναι ίδια ή διαφέρουν στην φάση κατά γωνία π (rad). Η στερεοφωνική κωδικοποίηση μέσου/αθροίσματος μπορεί να εφαρμοστεί σε όλο το φάσμα του σήματος από τη στιγμή που δεν επιφέρει απώλεια πληροφορίας και μάλιστα διατηρεί τα χωρικά χαρακτηριστικά του σήματος. Η ανακατασκευή των σημάτων L και R γίνεται από τις σχέσεις $L=(M+S)/2$ και $R=(M-S)/2$.

Η στερεοφωνική κωδικοποίηση έντασης εστιάζει στην αφαίρεση πλεονασματικών στοιχείων ιδιαίτερα στις υψηλές συχνότητες, στις οποίες το σήμα κωδικοποιείται με μειωμένη χωρική ανάλυση. Η προσέγγιση αυτή βασίζεται στο γεγονός ότι το ανθρώπινο σύστημα ακοής ανιχνεύει την κατεύθυνση του ήχου στις υψηλές συχνότητες βασιζόμενο στη σχετική διαφορά έντασης μεταξύ των σημάτων που καταφθάνουν στα δυο αυτιά. Στη στερεοφωνική κωδικοποίηση έντασης μόνο ένα κανάλι που προκύπτει από το συνδυασμό αριστερού και δεξιού καναλιού μεταδίδεται για κάθε κρίσιμη συχνότητα. Οι πληροφορίες κατεύθυνσης μεταδίδονται χωρίς παράγοντες κλίμακας για το αριστερό και το δεξί κανάλι. Η στερεοφωνική κωδικοποίηση έντασης διατηρεί την ενέργεια του στερεοφωνικού ήχου, αλλά κάποια από τα στοιχεία του σήματος μπορεί να μη μεταφερθούν σωστά οδηγώντας σε απώλεια χωρικών δεδομένων. Γενικά, η απώλεια χωρικών δεδομένων θεωρείται προτιμότερη έναντι των αδυναμιών κωδικοποίησης λόγω απώλειας bits. Αυτός είναι ο λόγος που η κωδικοποίηση έντασης εφαρμόζεται κυρίως σε χαμηλούς ρυθμούς δεδομένων. Η στερεοφωνική κωδικοποίηση έντασης είναι μια προσέγγιση που εφαρμόζεται αποκλειστικά στις υψηλές συχνότητες. Εάν χρησιμοποιηθεί σε χαμηλές συχνότητες τότε μπορεί να οδηγήσει σε σημαντική παραμόρφωση και μεγάλη απώλεια χωρικών δεδομένων. Υποθέτοντας ότι η κωδικοποίηση έντασης εφαρμόζεται στο άνω μισό μέρος του φάσματος, μπορεί να επιτευχθεί μια εξοικονόμηση 20% του ρυθμού δεδομένων του συστήματος. Η επέκταση σε πολυκαναλικά συστήματα αναπτύχθηκε αργότερα με προσεγγίσεις, όπως ο συνδυασμός καναλιών και η δυναμική συνακρόαση ή η γενικευμένη κωδικοποίηση έντασης.

2.2 Το πρότυπο MPEG-2

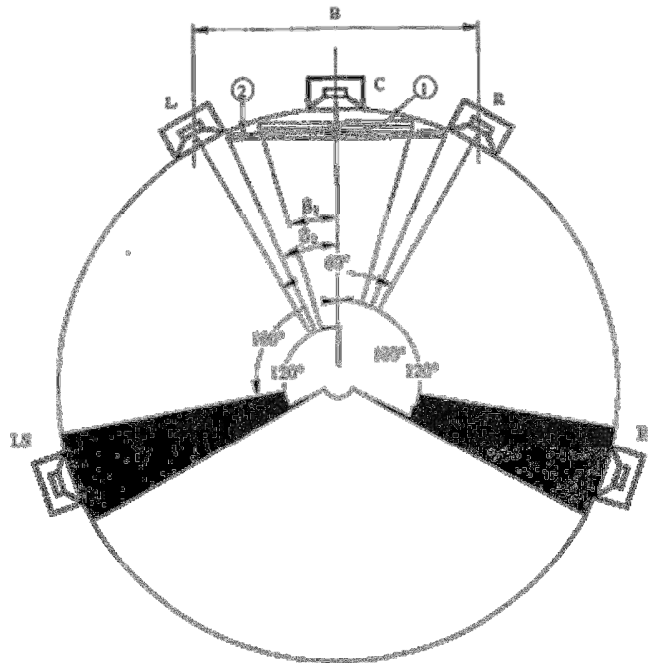
Οι κωδικοποιήσεις MPEG-2 αναπτύχθηκαν σαν επέκταση της κωδικοποίησης MPEG-1 για την επίτευξη χαμηλότερων ρυθμών δεδομένων και για εφαρμογές πολλών καναλιών (multichannel). Αρχικά αναπτύχθηκαν οι κωδικοποιήσεις MPEG-2 LSF (Low Sampling Frequencies) και MPEG-2.5, στις οποίες χρησιμοποιούνται χαμηλότερες συχνότητες δειγματοληψίας ώστε να μειωθεί ο ρυθμός δεδομένων. Η μορφή δεδομένων MP3 συνήθως εφαρμόζεται σαν MPEG-2.5 επέκταση του επιπέδου III της MPEG-1. Η κωδικοποίηση MPEG-2 BC (Backwards Compatible) επέκτεινε την MPEG-1 σε πολυκάναλη λειτουργία, ενώ διατήρησε την προς τα πίσω συμβατότητα ώστε τα σήματα με MPEG-2 BC μπορούν να αναπαραχθούν σε υπάρχοντα συστήματα MPEG-1. Η κωδικοποίηση MPEG-2 AAC (Adaptive Audio Coding) είναι η μετεξέλιξη της MPEG-2, η οποία χρησιμοποιώντας όλες τις προηγμένες τεχνικές κωδικοποίησης ήχου φτάνει σε υψηλότερης ποιότητας κωδικοποίησης πολυκάναλων συστημάτων χωρίς να έχει τις δεσμεύσεις συμβατότητας με την MPEG-1.

2.2.1 Εισαγωγή στον ήχο πολλών καναλιών

Το επίκεντρο του ήχου MPEG-2 τέθηκε γύρω από τον πολυκάναλο ήχο (multichannel audio) και τη χωρική αναπαράσταση του ήχου (spatial audio). Η εξέλιξη του πολυκάναλου ήχου πυροδοτήθηκε από την πρόοδο της βιομηχανίας του κινηματογράφου, κάνοντας τον πολυκάναλο ήχο να μετακινηθεί προς το στερεοφωνικό, τετραφωνικό, περιβάλλοντα (surround) ήχο, κτλ.

Η βιομηχανία του κινηματογράφου αγκάλιασε τα πολυκάναλα formats ήχου καθώς προωθούσαν την ευελιξία και την πιο πλήρη εμπειρία ήχου που παρείχαν. Στη δεκαετία του '80 εισάχθηκε στην αγορά το CD και ο ψηφιακός στερεοφωνικός ήχος. Εκείνη την εποχή λίγοι καλλιτέχνες ασχολήθηκαν με τον τετραφωνικό ήχο και πολύ μικρό κοινό είχε πρόσβαση σε συστήματα με πάνω από δυο κανάλια για οικιακή χρήση. Στη δεκαετία του '90, που έγινε η μετάβαση των οπτικοακουστικών συστημάτων από τις αίθουσες κινηματογράφου στα σαλόνια των σπιτιών, ο ρυθμός θέσπισης καινούριων standards κωδικοποίησης ήχου και βίντεο αυξήθηκε. Στην αρχή της δεκαετίας του '90 η εστίαση της τυποποίησης από τον μονοφωνικό και τον στερεοφωνικό ήχο στράφηκε σε πολυκάναλα ηχητικά σήματα. Τα σημερινά τεχνολογικά δεδομένα με τη σταθερή αύξηση του bandwidth μεταφοράς και τη δυνατότητα αποθήκευσης έκαναν τον πολυκάναλο ήχο μια πιο ρεαλιστική και προσιτή επιλογή για αναπαραγωγή ήχου.

Το MPEG-2 είναι ένα από τα πρώτα πρότυπα κωδικοποίησης ήχου που περιείχε τη ρύθμιση για format ήχου 5.1. Αυτό το format περιέχει 5 κανάλια πλήρους φάσματος 20 kHz, και ένα κανάλι χαμηλής συχνότητας (subwoofer) για ενίσχυση της εμπειρίας ήχου σε πολύ χαμηλές συχνότητες, δηλαδή κάτω από 200 Hz.



Εικόνα 14 Τοποθέτηση ηχείων για αναπαραγωγή ήχου 5 καναλιών

Η παραπάνω σχεδίαση της τοποθέτησης των ηχείων καταδεικνύει γιατί ο ήχος 5.1 συχνά αναφέρεται σαν ήχος 3/2/.1, μιας και τα 3 ηχεία τοποθετούνται μπροστά από τον ακροατή και 2 πλάγια και πίσω. Το ηχείο χαμηλών συχνοτήτων τυπικά τοποθετείται μπροστά από τον ακροατή αν και αυτό δεν ορίζεται ρητά από το πρότυπο, καθώς η ανθρώπινη ακοή δεν προσπαθεί να τοποθετήσει στο χώρο ένα μονό χαμηλής συχνότητας ηχητικό κανάλι. Σκοπός του ηχείου χαμηλών συχνοτήτων είναι η ενίσχυση των ηχείων χαμηλών συχνοτήτων χωρίς την υπερφόρτωση των κύριων καναλιών.

Η μετάβαση από το στερεοφωνικό ήχο στον πολυκάναλο προσθέτει απαιτήσεις στο σύστημα, τόσο για την αποθήκευση όσο και την μεταφορά των μέσων. Το format του CD που δειγματοληπτείται με PCM στα 44.1 kHz και κβαντίζεται ομοιόμορφα με 16 bits ανά δείγμα, απαιτεί ρυθμό δεδομένων 3598 Mb/s για ήχο 5.1. Μια ώρα διάρκεια από πολυκάναλη μουσική σε μορφή CD απαιτεί 1.62 GB, δηλαδή πολύ περισσότερο από τη χωρητικότητα ενός δίσκου CD των 800 MB. Είναι λογικό λοιπόν οι απαιτήσεις για εφαρμογές μετάδοσης διαδικτυακού ήχου και διανομή μουσικής μέσω του διαδικτύου να περιέχουν προκλήσεις σε σχέση με τη χωρητικότητα ή και το bandwidth.

Όπως στον μονοφωνικό και στον στερεοφωνικό ήχο, έτσι και στον πολυκάναλο ήχο η πρόκληση είναι η μείωση του ρυθμού δεδομένων χωρίς θυσίες στην ποιότητα του ήχου. Η διαμόρφωση PCM είναι μια καθιερωμένη και γνωστή μέθοδος κωδικοποίησης που προσφέρει χαμηλής πολυπλοκότητας υλοποιήσεις, αλλά απαιτεί πολύ υψηλό bandwidth για να παρέχει υψηλής ποιότητας ηχητικά σήματα. Τα ηχητικά σήματα σε μορφή CD μπορεί να υποφέρουν από υποβάθμιση της ποιότητας. Τα επίπεδα διακριτότητας για CD και τυπικούς μετατροπείς των 16-bits δείχνουν ότι αισθητός θόρυβος κβαντισμού μπορεί να εισαχθεί σε μεσαίες

συχνότητες. Η επιπλοκή είναι ότι όσο και να μοιάζει δαπανηρό, απαιτείται η αύξηση από 16 στα 24 bits της ακρίβειας του PCM. Επίσης, επιπλέον της αυξημένης ακρίβειας δείγματος, μπορεί να αυξηθεί και ο ρυθμός δειγματοληψίας από 44.1 και 48, στα 96kHz και 192 kHz. Αυτό οδηγεί σε διαφορετικές επιλογές εξισορρόπησης της απαίτησης της αγοράς για ήχο υψηλής ποιότητας με τους ισχύοντες όμως περιορισμούς στα τεχνικά ζητήματα.

Σε υψηλές συχνότητες δειγματοληψίας, οι περιορισμοί που εγείρονται είναι ακόμα πιο δεσμευτικοί και απαγορεύουν τον πολυκάναλο ήχο ακόμα και για πιο δυναμικές τεχνολογίες υψηλότερων δυνατοτήτων. Εάν αναλογιστούμε τον πλήρη ήχο 5 καναλιών δειγματοληπτημένο στα 96 kHz με ακρίβεια 24 bits και ρυθμό δεδομένων 11.52 Mb/s, λαμβάνουμε ένα αποτέλεσμα που υπερβαίνει τα 9.6 Mb/s του DVD-Audio.

2.2.2 Το πρότυπο MPEG-2 AAC

Η κωδικοποίηση MPEG-2 AAC (Advanced Audio Coding), ξεκίνησε το 1994 σαν μια άλλη προσπάθεια της επιτροπής MPEG-2 ήχου για υψηλότερης ποιότητας πρότυπο πολλών καναλιών με προς τα πίσω συμβατότητα (backward compatibility). Το πρότυπο τελειοποιήθηκε το 1997. Η κωδικοποίηση AAC χρησιμοποίησε όλες τις τεχνικές κωδικοποίησης ήχου που ήταν διαθέσιμες την εποχή της ανάπτυξης του προτύπου για να παρέχει πολύ υψηλής ποιότητας ήχο πολλών καναλιών.

Ο στόχος της ανάπτυξης της MPEG-2 AAC ήταν η απόδοση εξαιρετικής ποιότητας ήχου σε ρυθμούς δεδομένων της τάξης των 384 kb/s και χαμηλότερους για 5 πλήρους bandwidth ηχητικά κανάλια. Δοκιμασίες που έγιναν το 1996 από το BBC στο Ηνωμένο Βασίλειο και το NHK στην Ιαπωνία, έδειξαν ότι η MPEG-2 AAC ικανοποιεί τις απαιτήσεις ποιότητας της ITU-R στα 320 kb/s για 5 πλήρους bandwidth κανάλια. Η κωδικοποίηση MPEG-2 AAC αποτελεί τον πυρήνα της MPEG-4, που παρέχει ήχο υψηλής ποιότητας με χαμηλή καθυστέρηση, φυσικό ήχο και φορητά προφίλ ηχητικής δικτύωσης. Οι προδιαγραφές της AAC είναι το αποτέλεσμα συλλογικής προσπάθειας εταιριών ανά τον κόσμο για την ανάπτυξη της ηχητικής κωδικοποίησης. Οι AAC προδιαγραφές δέχθηκαν πολλές τροποποιήσεις και διορθώσεις πριν την πρώτη κατάθεσή τους. Για να οριστεί ένα σύστημα AAC η επιτροπή ήχου JTC επέλεξε μια σταδιακή προσέγγιση, στην οποία το πλήρες σύστημα αποσυντίθεται σε μια σειρά αυτόνομων τμημάτων, όπου ένα τμήμα ορίζεται σαν ένα τμήμα κωδικοποίησης που μπορεί να χρησιμοποιηθεί σαν χωριστό στοιχείο του συνολικού συστήματος. Το μοντέλο αναφοράς της AAC περιέγραψε τα χαρακτηριστικά κάθε τμήματος και πως αυτά συνεργάζονται. Κάθε οπτική του μοντέλου αναφοράς εκτιμήθηκε με πειράματα που αποτελούνται από ελεγχόμενες ακροάσεις με οδηγίες που ορίζονται από την ITU-R BS.

Τα τμήματα της AAC είναι:

- Έλεγχος κέρδους
- Φίλτρα
- Πρόβλεψη
- Κβάντιση και Κωδικοποίηση
- Κωδικοποίηση χωρίς θόρυβο
- Πολύπλεξη συμβολοσειράς

- Διαμόρφωση παραθύρου θορύβου
- Στερεοφωνική κωδικοποίηση μέσου/ακραίου
- Στερεοφωνική κωδικοποίηση έντασης

Για να είναι δυνατή η επιλογή μεταξύ ποιότητας και απαιτήσεων μνήμης/επεξεργαστικής ισχύος, το σύστημα AAC προσφέρει τρία προφίλ: Κύριο προφίλ, Χαμηλής πολυπλοκότητας προφίλ και Προφίλ κλιμακούμενου ρυθμού δειγματοληψίας.

Στη ρύθμιση κύριου προφίλ, το σύστημα AAC παρέχει την καλύτερη ποιότητα ήχου σε κάθε ρυθμό δεδομένων. Με εξαίρεση το εργαλείο επεξεργασίας, όλα τα μέρη των τμημάτων AAC μπορούν να χρησιμοποιηθούν. Η μνήμη και η επεξεργαστική ισχύς που απαιτείται σε αυτή τη ρύθμιση είναι υψηλότερη από τη μνήμη και την επεξεργαστική ισχύ που απαιτείται στο προφίλ χαμηλής πολυπλοκότητας. Το κύριο προφίλ μπορεί να αποκωδικοποιήσει μια συμβολοσειρά που έχει κωδικοποιηθεί από προφίλ χαμηλής πολυπλοκότητας.

Στο προφίλ χαμηλής πολυπλοκότητας, η πρόβλεψη και η προεπεξεργασία δεν χρησιμοποιούνται και η τάξη της διαμόρφωσης παραθύρου θορύβου είναι περιορισμένη. Ενώ η ποιοτική επίδοση του προφίλ χαμηλού είναι πολύ υψηλή, η μνήμη και η υπολογιστική ισχύς που απαιτούνται μειώνεται σημαντικά με αυτή τη ρύθμιση.

Στο προφίλ κλιμακούμενου ρυθμού δειγματοληψίας, ο έλεγχος κέρδους είναι απαραίτητος. Η προεπεξεργασία που γίνεται από τον έλεγχο κέρδους αποτελείται από ένα Συζυγές Τετράγωνο Φίλτρο (CQF, Conjugate Quadrature Filter), ανιχνευτές κέρδους και τροποποιητές κέρδους. Η πρόβλεψη και η διαμόρφωση θορύβου παραθύρου, καθώς και το bandwidth περιορίζονται σε αυτό το προφίλ. Το προφίλ κλιμακούμενου ρυθμού δειγματοληψίας έχει χαμηλότερη πολυπλοκότητα από το κύριο και το χαμηλής πολυπλοκότητας προφίλ και μπορεί να παρέχει ένα σήμα κλιμακούμενης συχνότητας.

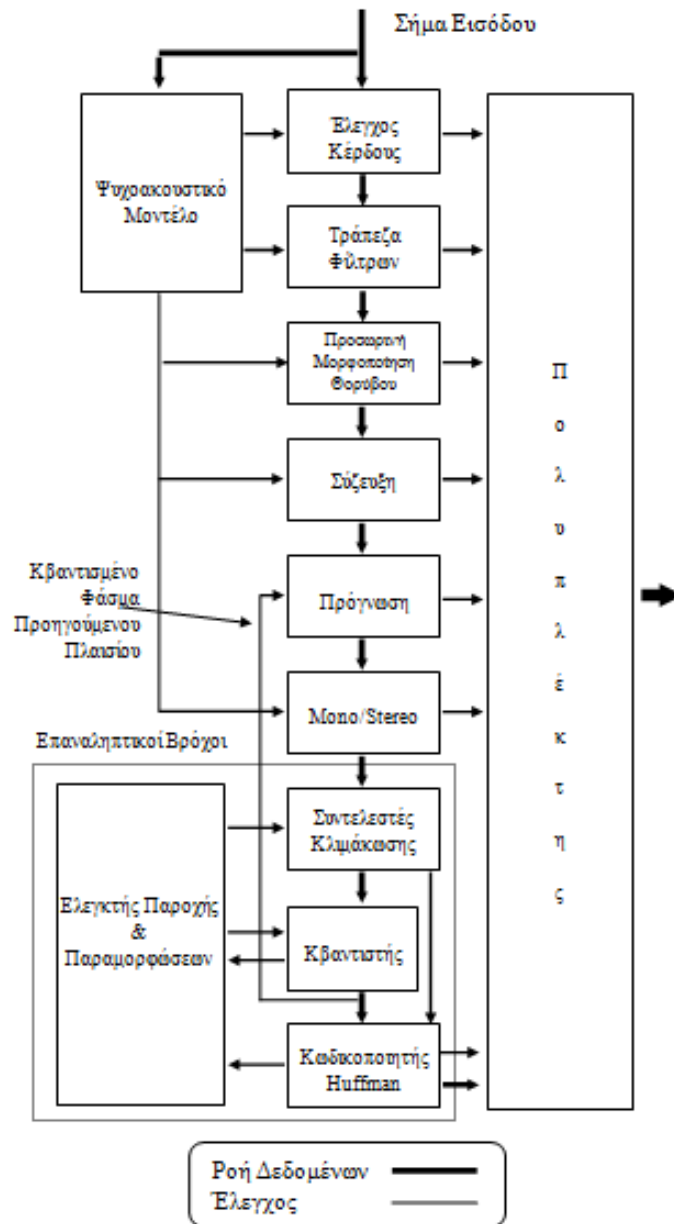
Ο κωδικοποιητής AAC φαίνεται στην Εικόνα 15. Πρώτα, ένα φίλτρο βασισμένο στον MDCT χρησιμοποιείται για την αποσύνθεση του σήματος εισόδου σε υποδειγματοληπτημένα φασματικά στοιχεία. Σε ρυθμό δειγματοληψίας 48 kHz το φίλτρο AAC επιτρέπει διακριτότητα συχνοτήτων 23 Hz και χρονική ανάλυση 2.6 ms. Με βάση το σήμα εισόδου υπολογίζεται μια χρονικά εξαρτημένη εκτίμηση των ορίων επικάλυψης. Ένα μοντέλο αντίληψης παρόμοιο με το ψυχοακουστικό μοντέλο MPEG χρησιμοποιείται και για το σύστημα AAC.

Μετά το φίλτρο ανάλυσης, η διαμόρφωση παραθύρου θορύβου κάνει ένα φιλτράρισμα στις φασματικές τιμές, για παράδειγμα αντικαθιστά τις παραμέτρους φάσματος με υπόλοιπα πρόβλεψης. Η τεχνική του παραθύρου διαμόρφωσης θορύβου επιτρέπει στον κωδικοποιητή να ελέγχει τη δομή του θορύβου κβάντισης ακόμα και σε ένα απλό χρονικό παράθυρο του φίλτρου.

Σε σήματα πολλών καναλιών, η στερεοφωνική κωδικοποίηση έντασης μπορεί να εφαρμοστεί με μόνο την ενεργειακή περιβάλλουσα (envelope) να μεταδίδεται. Η στερεοφωνική κωδικοποίηση έντασης επιτρέπει τη μείωση της χωρικής πληροφορίας που μεταδίδεται και είναι πολύ αποτελεσματική σε χαμηλούς ρυθμούς δεδομένων.

Η πρόβλεψη στο πεδίο του χρόνου χρησιμοποιείται για την αξιοποίηση των συσχετίσεων μεταξύ των υποδειγματοληπτημένων (sub-sampled) στοιχείων με επόμενα παράθυρα

δεδομένων κάτι που οδηγεί σε αυξημένη μείωση των πλεονασμών (redundancy) για στατική ανάλυση.



Εικόνα 1515 Διάγραμμα MPEG-2 AAC κωδικοποιητή [9]

Αντί της μετάδοσης αριστερού και δεξιού καναλιού, μπορεί να μεταδοθεί το κανονικοποιημένο (normalized) άθροισμα και η διαφορά τους. Η αναπτυγμένη στερεοφωνική κωδικοποίηση αθροίσματος/διαφοράς χρησιμοποιείται στον κωδικοποιητή AAC σε χαμηλούς ρυθμούς δεδομένων.

Τα φασματικά στοιχεία κβαντίζονται και κωδικοποιούνται με στόχο τη διατήρηση του θορύβου κβάντισης κάτω από το κατώφλι επικάλυψης. Αυτό το βήμα γίνεται χρησιμοποιώντας το στάδιο ανάλυσης-σύνθεσης και χρησιμοποιώντας επιπλέον εργαλεία συμπίεσης χωρίς θόρυβο. Ένας μηχανισμός «αποθήκης bit» (bit reservoir) παρόμοιος με αυτόν που χρησιμοποιήθηκε στο MPEG επίπεδο III, επιτρέπει ένα τοπικά μεταβλητό ρυθμό δεδομένων, ώστε να ικανοποιηθούν οι απαιτήσεις του σήματος από κάθε παράθυρο δεδομένων στο επόμενο. Τελικά, ένας μορφοποιητής συμβολοσειράς χρησιμοποιείται για να οργανώσει τη συμβολοσειρά, η οποία αποτελείται από τα κβαντισμένα και κωδικοποιημένα φασματικά στοιχεία και τις παραμέτρους ελέγχου.

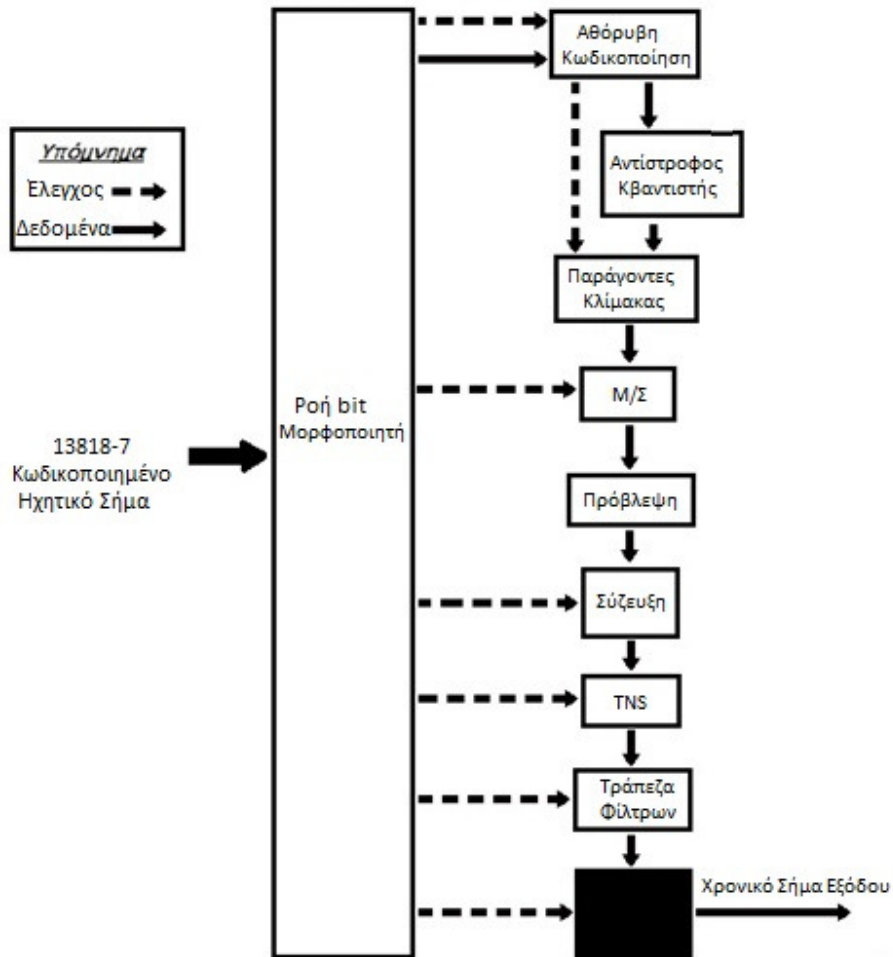
Το σύστημα MPEG-2 AAC υποστηρίζει μέχρι 48 ηχητικά κανάλια. Οι βασικές ρυθμίσεις περιέχουν μονοφωνικό, δύο καναλιών και 5 καναλιών συν ένα χαμηλών συχνοτήτων. Επιπλέον των βασικών ρυθμίσεων, δεκαέξι πιθανές προγραμματιζόμενες ρυθμίσεις μπορούν να οριστούν στον κωδικοποιητή.

Οι ρυθμοί δειγματοληψίας που υποστηρίζονται από το σύστημα AAC ποικίλουν από τα 8 kHz ως τα 96 kHz. Στον παρακάτω πίνακα φαίνεται ο μέγιστος ρυθμός δεδομένων ανά κανάλι που εξαρτάται από το ρυθμό δειγματοληψίας και το μέγεθος του buffer αποθήκης bit των 6144 bits ανά κανάλι.

Συχνότητα Δειγματοληψίας (Hz)	Μέγιστος ρυθμός δεδομένων ανά κανάλι (kb/s)
96.000	576
88.200	329,2
64.000	384
48.000	288
44.100	264,6
32.000	192
24.000	144
22.050	132,3
16.000	96
12.000	72
11.025	66,25
8.000	48

Πίνακας 2 Μέγιστος ρυθμός δεδομένων ανά συχνότητα δειγματοληψίας

Το διάγραμμα του αποκωδικοποιητή MPEG-2 AAC είναι το παρακάτω.



Εικόνα 1616 Διάγραμμα MPEG-2 AAC αποκωδικοποιητή

2.2.3 Κβαντισμός και κωδικοποίηση

Ο κύριος στόχος του σταδίου κβάντισης και κωδικοποίησης είναι να κβαντίσει τα φασματικά δεδομένα με τέτοιο τρόπο που ο θόρυβος κβάντισης να ικανοποιεί τις απαιτήσεις του ψυχοακουστικού μοντέλου. Την ίδια στιγμή, ο αριθμός των bits που απαιτούνται για να κωδικοποιήσουμε το κβαντισμένο σήμα πρέπει να είναι κάτω από ένα όριο, συνήθως ο μέσος αριθμός bits που διατίθενται για κάποιο πακέτο ηχητικών δεδομένων. Η τιμή εξαρτάται από τη συχνότητα δειγματοληψίας και τον επιθυμητό ρυθμό δεδομένων. Στην AAC, μια αποθήκη δεδομένων δίνει τη δυνατότητα ελέγχου της βραχυπρόθεσμης κατανομής των bits μεταξύ διαδοχικών πακέτων ήχου. Η εκπλήρωση των απαιτήσεων του ψυχοακουστικού μοντέλου και η διατήρηση του αριθμού των τοποθετημένων bits κάτω από ένα συγκεκριμένο αριθμό,

συνδέονται με τη διαδικασία της κβάντισης. Ζητούμενο, είναι η συμπεριφορά όταν δεν μπορούν να ικανοποιηθούν ταυτόχρονα και οι δυο περιορισμοί ή όταν δεν απαιτείται να χρησιμοποιηθούν όλα τα όρια των περιορισμών.

Στρατηγική καθορισμένη για βέλτιστη κβάντιση δεν υπάρχει, η μόνη απαίτηση είναι η συμβολοσειρά να παράγεται με μορφή που να συνάδει με την AAC. Μια πιθανή στρατηγική είναι η χρήση δυο εμφωλευμένων βρόχων (nested loops). Αυτή η τεχνική χρησιμοποιήθηκε στην επίσημη δοκιμασία της AAC. Ένα σημαντικό θέμα είναι η ακριβής ρύθμιση μεταξύ ψυχοακουστικού μοντέλου και της διαδικασίας κβάντισης, που μπορεί να θεωρηθεί σαν ένα από τα μυστικά της κωδικοποίησης ήχου, μιας και απαιτεί εμπειρία και γνώση.

Τα κύρια χαρακτηριστικά της διαδικασίας κβάντισης AAC είναι :

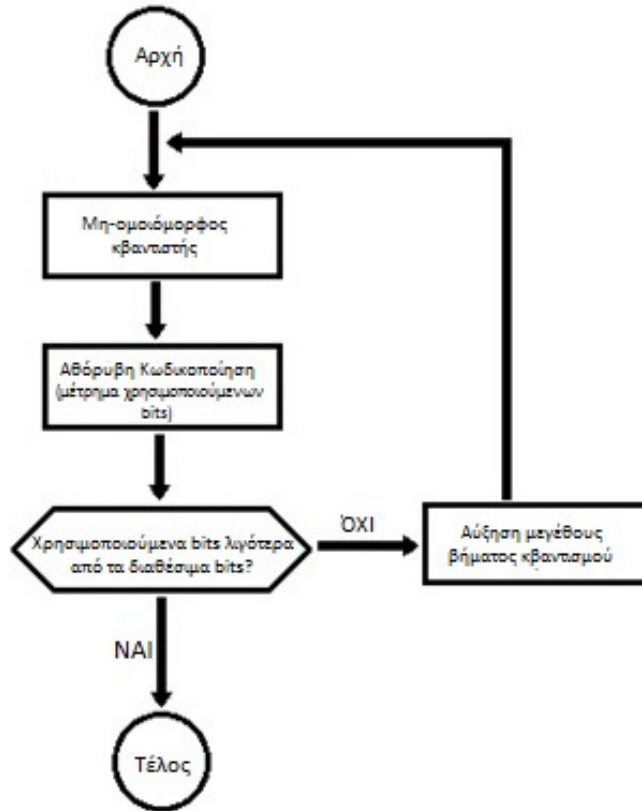
- Μη-ομοιόμορφη κβάντιση
- Κωδικοποίηση Huffman των φασματικών τιμών με διαφορετικούς πίνακες
- Μορφοποίηση θορύβου με ενίσχυση ομάδων φασματικών τιμών. Η πληροφορία γύρω από την ενίσχυση αποθηκεύεται στις τιμές των συντελεστών κλίμακας
- Κωδικοποίηση Huffman με διαφορετικούς παράγοντες κλίμακας

Η χρήση ενός μη ομοιόμορφου κβαντιστή (non-uniform quantizer) δεν επαρκεί για τις απαιτήσεις του ψυχοακουστικού μοντέλου. Οι απαιτήσεις καλύπτονται επαρκώς μόνο εάν έχουμε τη δυνατότητα να μορφοποιήσουμε το θόρυβο κβάντισης σε τμήματα παρόμοια με τις κρίσιμες περιοχές συχνοτήτων (critical bands) για το ανθρώπινο σύστημα ακοής. Αφού το σύστημα AAC προσφέρει μια σχετικά υψηλή διακριτότητα συχνότητας για μεγάλα πακέτα σε ρυθμό δειγματοληψίας 48 kHz, είναι δυνατή η δημιουργία ομάδων φασματικών τιμών που αντικατοπτρίζουν το bandwidth των κρίσιμων περιοχών συχνοτήτων.

Η αντίστροφη ενίσχυση παραγόντων κλίμακας πρέπει να εφαρμοστεί στον αποκωδικοποιητή. Ο πρώτος παράγοντας κλίμακας αντιπροσωπεύει το καθολικό μέγεθος του βήματος κβαντιστή (quantization step) και κωδικοποιείται σε PCM τιμή που ονομάζεται καθολικό κέρδος (global gain). Όλοι οι παράγοντες κλίμακας που ακολουθούν κωδικοποιούνται διαφορικά (differential coding), χρησιμοποιώντας μια ειδική κωδικοποίηση Huffman.

Η απόφαση του ποιος παράγοντας κλίμακας συχνότητας πρέπει να ενισχυθεί, αφήνεται μέχρι ενός σημείου στον αποκωδικοποιητή. Τα όρια που υπολογίζονται από το ψυχοακουστικό μοντέλο είναι τα πιο σημαντικά κριτήρια, αλλά όχι τα μόνα, από τη στιγμή που μόνο ένας περιορισμένος αριθμός bits μπορεί να χρησιμοποιηθεί. Η επαναληπτική διαδικασία που θα περιγραφεί παρακάτω είναι μια διαδικασία μορφοποίησης του θορύβου κβαντισμού (noise shaping). Αυτή η μέθοδος είναι γνωστό ότι παράγει πολύ καλή ποιότητα ήχου. Δύο εμφωλευμένοι επαναληπτικοί βρόχοι, ένας εσωτερικός και ένας εξωτερικός χρησιμοποιούνται για τον καθορισμό της βέλτιστης κβάντισης. Η περιγραφή που δίνεται εδώ είναι απλοποιημένη ώστε να βοηθηθεί η κατανόησή της. Σκοπός του εσωτερικού βρόχου επανάληψης είναι η αλλαγή του βήματος κβάντισης μέχρι τα φασματικά δεδομένα να μπορούν να κωδικοποιηθούν με το διαθέσιμο αριθμό bits. Γι' αυτό το σκοπό επιλέγεται ένα αρχικό μέγεθος βήματος κβάντισης, τα φασματικά δεδομένα κβαντίζονται και μετρίεται ο απαραίτητος αριθμός bits για να κωδικοποιηθούν τα κβαντισμένα δεδομένα. Εάν ο αριθμός είναι υψηλότερος από τα διαθέσιμα bits, τότε το μέγεθος βήματος κβάντισης αυξάνεται και η όλη διαδικασία

επαναλαμβάνεται. Το λογικό διάγραμμα του εσωτερικού βρόχου επανάληψης φαίνεται στο παρακάτω σχήμα.



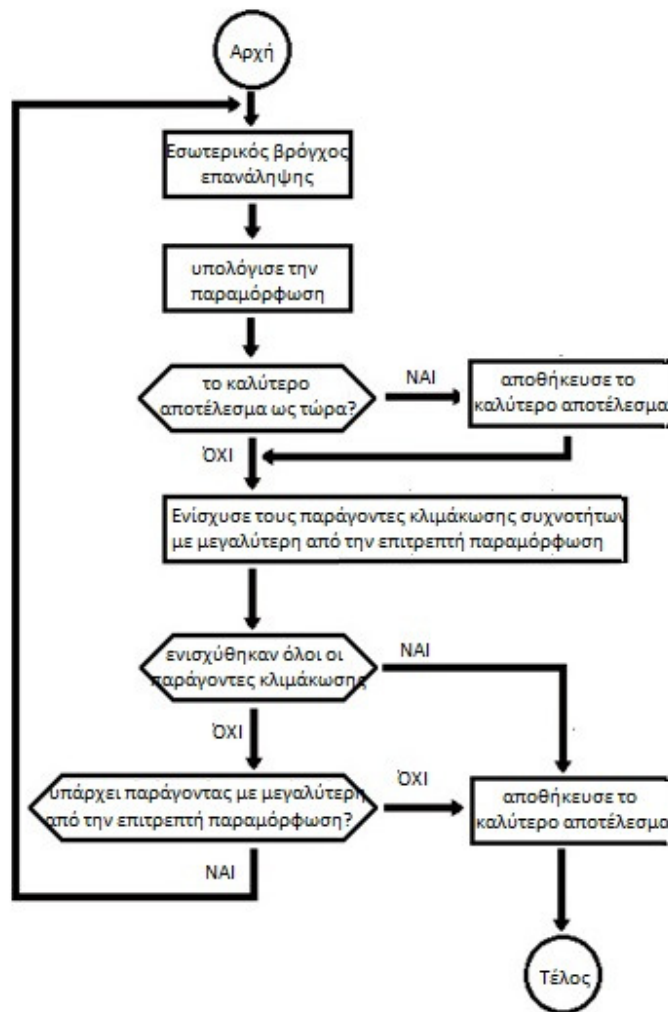
Εικόνα 1717 Λογικό διάγραμμα εσωτερικού βρόχου επανάληψης MPEG-2 AAC

Ο σκοπός του εξωτερικού βρόχου επανάληψης είναι η ενίσχυση των παραγόντων κλίμακας συχνοτήτων με τέτοιο τρόπο ώστε οι απαιτήσεις του ψυχοακουστικού μοντέλου να ικανοποιούνται όσο το δυνατόν περισσότερο. Εφαρμόζονται τα ακόλουθα βήματα:

1. Αρχικά, κανένας παράγοντας κλίμακας συχνοτήτων δεν ενισχύεται.
2. Καλείται ο εσωτερικός βρόχος.
3. Για κάθε παράγοντα κλίμακας συχνοτήτων υπολογίζεται η παραμόρφωση που προκαλείται από την κβάντιση.
4. Η πραγματική παραμόρφωση συγκρίνεται με την επιτρεπτή παραμόρφωση που υπολογίζεται από το ψυχοακουστικό μοντέλο.
5. Εάν το αποτέλεσμα είναι το καλύτερο αποτέλεσμα μέχρι στιγμής, αποθηκεύεται. Αυτό είναι σημαντικό καθώς η επαναληπτική διαδικασία δεν είναι απαραίτητο ότι συγκλίνει.
6. Οι παράγοντες κλιμάκωσης συχνοτήτων με πραγματική παραμόρφωση υψηλότερη από την επιτρεπτή παραμόρφωση ενισχύονται. Σε αυτό το σημείο μπορούν να

εφαρμοστούν διαφορετικές μέθοδοι καθορισμού των παραγόντων κλιμάκωσης συχνοτήτων που θα ενισχυθούν.

7. Εάν όλοι οι παράγοντες κλιμάκωσης συχνοτήτων ενισχύθηκαν, τότε η επαναληπτική διαδικασία σταματά. Έτσι επιτυγχάνεται το βέλτιστο αποτέλεσμα.
8. Εάν δεν υπάρχει ούτε ένας παράγοντας κλιμάκωσης συχνοτήτων που να είναι ανώτερος από την επιτρεπτή παραμόρφωση, τότε η επαναληπτική διαδικασία σταματά επίσης.
9. Αλλιώς, η διαδικασία επαναλαμβάνεται από το βήμα 2 με νέες τιμές ενίσχυσης.



Εικόνα 1818 Λογικό διάγραμμα εξωτερικού βρόχου επανάληψης MPEG-2 AAC

Υπάρχουν συνθήκες που δεν αναφέρονται παραπάνω, οι οποίες προκαλούν τερματισμό του εξωτερικού βρόχου επανάληψης. Αφού τα ενισχυμένα του φάσματος απαιτούν περισσότερα bits για κωδικοποίηση, ενώ ο αριθμός των διαθέσιμων bits είναι σταθερός, το βήμα κβάντισης πρέπει να τροποποιηθεί στον εσωτερικό βρόχο επανάληψης για να μειωθεί ο αριθμός των χρησιμοποιούμενων bits. Αυτός ο μηχανισμός μεταφέρει bits από φασματικές περιοχές που

δεν απαιτούνται σε περιοχές που είναι απαραίτητα. Για τον ίδιο λόγο το αποτέλεσμα μετά από μια ενίσχυση στον εξωτερικό βρόχο μπορεί να είναι χειρότερο από πριν, ώστε το καλύτερο αποτέλεσμα να πρέπει να αποκατασταθεί μετά από τον τερματισμό της διαδικασίας επανάληψης. Η κβάντιση και η διαδικασία κωδικοποίησης για μικρά πακέτα είναι παρόμοια με αυτή για μεγάλα πακέτα, αλλά η ομαδοποίηση και η εναλλαγή πρέπει να ληφθεί υπ' όψη. Το λογικό διάγραμμα του εξωτερικού βρόχου επανάληψης φαίνεται στο παρακάτω σχήμα.

2.2.4 Κωδικοποίηση χωρίς θόρυβο

Η είσοδος στο τμήμα κωδικοποίησης χωρίς θόρυβο είναι ένα σύνολο 1024 κβαντισμένων φασματικών συντελεστών. Το πρώτο βήμα της μεθόδου της χωρίς θόρυβο συμπίεσης του δυναμικού εύρους μπορεί να εφαρμοστεί στο φάσμα. Μέχρι τέσσερις συντελεστές μπορούν να κωδικοποιηθούν χωριστά σαν πλάτη με μια τιμή +1 για θετικές τιμές και -1 για αρνητικές τιμές. Οι ψαλιδισμένοι συντελεστές κωδικοποιούνται σαν ακέραια πλάτη και η απόσταση από τη βάση του πίνακα συντελεστών χρησιμοποιείται για να οριστεί η θέση τους. Οι πληροφορίες για τη μεταφορά των ψαλιδισμένων συντελεστών κοστίζουν μερικά bits, η συμπίεση χωρίς θόρυβο εφαρμόζεται μόνο εάν συμβάλει στη μείωση των bits.

Η κωδικοποίηση χωρίς θόρυβο τεμαχίζει το σύνολο των 1024 κβαντισμένων φασματικών συντελεστών σε τμήματα, ώστε ένας απλός κώδικας Huffman να χρησιμοποιείται για να κωδικοποιεί κάθε τμήμα. Για λόγους απόδοσης κωδικοποίησης, τα όρια των τμημάτων μπορούν να είναι μόνο στα όρια των παραγόντων κλίμακας συχνοτήτων, ώστε για κάθε τμήμα του φάσματος να πρέπει να μεταδοθεί το μήκος του τμήματος, σε όρους του αριθμού των παραγόντων κλίμακας συχνοτήτων που περιέχει, καθώς και ο αριθμός του κώδικα Huffman που χρησιμοποιείται.

Η τμηματοποίηση (segmentation) είναι δυναμική και τυπικά διαφέρει από πακέτο σε πακέτο, ώστε ο αριθμός των bits που χρειάζονται για την αναπαράσταση του πλήρους συνόλου των κβαντισμένων φασματικών συντελεστών να μειώνεται. Αυτό γίνεται χρησιμοποιώντας έναν άπληστο αλγόριθμο ένωσης (Greedy Algorithm), που ξεκινά με τον μέγιστο δυνατό αριθμό τμημάτων, το καθένα από τα οποία χρησιμοποιεί ένα Huffman κώδικα με τον ελάχιστο δυνατό δείκτη. Τα τμήματα ενώνονται, εάν το συνολικό ενοποιημένο τμήμα οδηγεί σε ένα χαμηλότερο συνολικό αριθμό bits, με ενοποιήσεις που φέρνουν τη μέγιστη μείωση αριθμού bits να γίνονται πρώτα. Εάν τα τμήματα προς ένωση δεν έχουν τον ίδιο κώδικα Huffman, τότε χρησιμοποιείται ο κώδικας με τον υψηλότερο δείκτη.

Τα τμήματα συχνά περιέχουν μόνο συντελεστές με τιμή μηδέν. Για παράδειγμα, εάν η είσοδος ήχου είναι ζωνοπεριορισμένη στα 20 kHz ή χαμηλότερα, τότε οι υψηλότεροι συντελεστές είναι μηδέν. Τέτοια τμήματα κωδικοποιούνται με τον μηδενικό κώδικα και δεν απαιτείται να σταλεί κάποια κωδική λέξη Huffman.

Εάν η ακολουθία παραθύρων είναι 8 μικρά παράθυρα, τότε το σύνολο των 1024 συντελεστών είναι στην πραγματικότητα ένας πίνακας 8 επί 128 συντελεστών συχνοτήτων, που αναπαριστούν την εξέλιξη χρόνου συχνότητας του σήματος κατά τη διάρκεια των 8 παραθύρων. Παρά το γεγονός ότι ο μηχανισμός τμηματοποίησης είναι αρκετά ευέλικτος για να αναπαραστήσει τα 8 μηδενικά τμήματα, η τμηματοποίηση και η εναλλαγή παρέχουν καλύτερη

απόδοση κωδικοποίησης. Οι συντελεστές που έχουν σχέση με συνεχή μικρά παράθυρα μπορούν να ομαδοποιηθούν, ώστε να μοιράζονται συντελεστές κλίμακας με όλους τους παράγοντες κλίμακας συχνοτήτων μέσα στην ομάδα. Επιπλέον, οι συντελεστές μιας ομάδας μπορούν να εναλλαχθούν, με εναλλαγή των συντελεστών κλίμακας συχνοτήτων και των παραθύρων. Συγκεκριμένα, έστω ότι το σύνολο πριν την εναλλαγή των 1024 συντελεστών σ λαμβάνουν δείκτες όπως παρακάτω:

σ [ο] [π] [κ] [δ]

όπου: ο ο δείκτης της ομάδας
π ο δείκτης του παραθύρου
κ ο δείκτης παράγοντα κλίμακας συχνοτήτων σε ένα παράθυρο
δ ο δείκτης συντελεστών σε παράγοντα κλίμακας συχνοτήτων

Μετά την εναλλαγή των συντελεστών οι δείκτες γίνονται:

σ [ο] [κ] [π] [δ]

Αυτή η εναλλαγή δίνει το πλεονέκτημα του συνδυασμού όλων των μηδενικών τμημάτων λόγω του περιορισμού συχνοτήτων σε κάθε ομάδα.

Το κωδικοποιημένο φάσμα χρησιμοποιεί έναν κβαντιστή για κάθε παράγοντα κλίμακας συχνοτήτων. Το μέγεθος βήματος για κάθε έναν από αυτούς τους κβαντιστές ορίζεται σαν ένα σύνολο συντελεστών κλίμακας και ένα καθολικό κέρδος που κανονικοποιεί τους συντελεστές κλίμακας. Για να γίνει αύξηση της συμπίεσης, οι παράγοντες κλίμακας που σχετίζονται με παράγοντες κλίμακας συχνοτήτων που έχουν μόνο μηδενικούς συντελεστές δεν μεταδίδονται. Τόσο το καθολικό κέρδος όσο και οι παράγοντες κλίμακας κβαντίζονται σε βήματα των 1.5 dB. Το καθολικό κέρδος κωδικοποιείται σαν ένας 8-bit χωρίς πρόσημο ακέραιος και οι παράγοντες κλίμακας κωδικοποιούνται διαφορετικά σχετικά με την προηγούμενη τιμή του παράγοντα κλίμακας συχνοτήτων και μετά κωδικοποιούνται κατά Huffman. Το δυναμικό εύρος του καθολικού κέρδους είναι επαρκές για την αναπαράσταση πλήρους κλίμακας τιμές για μια 24-bit PCM ηχητική πηγή.

Η κωδικοποίηση Huffman [44] χρησιμοποιείται για να αναπαρασταθούν N-αδες κβαντισμένων συντελεστών, με κώδικα Huffman από 12 κώδικες. Οι φασματικοί συντελεστές στις N-αδες δεδομένων κατατάσσονται από το χαμηλό στο υψηλό και το μέγεθος της N-αδας είναι 2 ή 4 συντελεστές. Η μέγιστη απόλυτη τιμή των κβαντισμένων συντελεστών που μπορεί να αναπαρασταθεί από κάθε κώδικα Huffman και ο αριθμός συντελεστών σε κάθε N-αδα για κάθε κώδικα φαίνεται στον παρακάτω πίνακα. Υπάρχουν δυο κώδικες για κάθε μέγιστη απόλυτη τιμή, με καθεμία να αναπαριστά μια διακριτή συνάρτηση κατανομής πιθανότητας. Για οικονομία αποθήκευσης κώδικα οι περισσότεροι κώδικες αναπαριστούν μη προσημασμένες τιμές. Γι' αυτούς τους κώδικες το πλάτος των συντελεστών κωδικοποιείται κατά Huffman και το bit προσήμου για κάθε μη μηδενικό συντελεστή προστίθεται στην κωδική λέξη.

Δείκτης Κώδικα	Μέγεθος N-αδας	Μέγιστη απόλυτη τιμή	Προσημασμένες τιμές
0		0	
1	4	1	Ναι
2	4	1	Ναι
3	4	2	Όχι
4	4	2	Όχι
5	2	4	Ναι
6	2	4	Ναι
7	2	7	Όχι
8	2	7	Όχι
9	2	12	Όχι
10	2	12	Όχι
11	2	16	Όχι

Πίνακας 3 Χαρακτηριστικά κωδίκων

Δυο κώδικες απαιτούν ειδική επισήμανση, ο κώδικας 0 και ο κώδικας 11. Ο κώδικας 0 σημαίνει ότι όλοι οι συντελεστές του τμήματος είναι μηδέν. Ο κώδικας 11 αναπαριστά κβαντισμένους συντελεστές με απόλυτη τιμή μεγαλύτερη ή ίση του 16. Εάν το πλάτος ενός ή δυο συντελεστών είναι μεγαλύτερο του 16, ένας ειδικός μηχανισμός κωδικοποίησης αποφυγής χρησιμοποιείται για την αναπαράσταση αυτών των τιμών. Το πλάτος των συντελεστών είναι περιορισμένο σε τιμές μικρότερες από 16 και η αντίστοιχη 2-αδα είναι κωδικοποιημένη κατά Huffman. Τα bits προσήμου, αν χρειαστεί, προστίθενται στην κωδική λέξη. Για κάθε πλάτος συντελεστή μεγαλύτερο ή ίσο του 16 προστίθεται ένας κωδικός αποφυγής:

Κωδικός αποφυγής = <πρόθεμα_αποφυγής> <διαχωριστικό_αποφυγής> <λέξη_αποφυγής>

Όπου: <πρόθεμα_αποφυγής> είναι η ακολουθία N δυαδικών «1»

<διαχωριστικό_αποφυγής> είναι ένα δυαδικό «0»

<λέξη_αποφυγής> είναι ένας N+4 μη προσημασμένος ακέραιος με μορφή MSB (Most Significant Bit) πρώτα

N, είναι ένας αριθμός αρκετά μεγάλος ώστε το πλάτος των κβαντισμένων συντελεστών να είναι ίσο με $2^{(N+4)} + \langle \text{λέξη}_{\text{αποφυγής}} \rangle$ (3)

2.2.5 Πολύπλεξη συμβολοσειράς

Η συμβολοσειρά του συστήματος MPEG-2 AAC συντάσσεται με πολύ ευέλικτο τρόπο. Ορίζονται δυο επίπεδα, το χαμηλότερο και πιο κοντά στα δεδομένα επίπεδο που καθορίζει τα ηχητικά δεδομένα, ενώ το υψηλότερο καθορίζει τον μηχανισμό μεταφοράς τους. Η μεταφορά των δεδομένων γίνεται με διαφορετικό τρόπο για κάθε εφαρμογή και έτσι τα ηχητικά δεδομένα πρέπει και είναι επαρκή για απλές εφαρμογές, όπως συμπίεση για αποθήκευση σε συσκευές.

Η δομή της συμβολοσειράς είναι η παρακάτω:

1. Ακολουθία μεταφοράς
2. Πακέτο δεδομένων

Αυτή η δομή επαναλαμβάνεται μέχρι να συμπληρωθούν τα δεδομένα.

Η δομή του πακέτου δεδομένων είναι η παρακάτω:

1. Στοιχείο προγραμματισμού ρύθμισης
2. Ηχητικό στοιχείο από τα: μονό κανάλι, στερεοφωνικό ζεύγος, χαμηλής συχνότητας
3. Ζεύξη πολλών καναλιών
4. Στοιχείο συμπλήρωσης, που ρυθμίζει το ρυθμό δεδομένων ώστε τα κανάλια να έχουν σταθερό ρυθμό
5. Στοιχείο τερματισμού, που συμβολίζει το τέλος ενός πακέτου

Από αυτά δεν είναι όλα τα στοιχεία απαραίτητα ούτε εμφανίζονται πάντα μόνο μια φορά σε κάθε θέση που υποδεικνύεται.

Το σύστημα AAC διαθέτει ένα buffer δεδομένων που επιτρέπει στο ρυθμό δεδομένων να μεταβάλλεται όταν αυτό απαιτείται από το ηχητικό σήμα. Έτσι, το μέγεθος κάθε πακέτου δεν είναι σταθερό. Συνεπώς, η συμβολοσειρά AAC δεν έχει σταθερό μέγεθος κεφαλίδας που χρησιμοποιείται σαν ακολουθία μεταφοράς. Οι κεφαλίδες πρέπει να έχουν συγκεκριμένη στοίχιση, ώστε να επιτρέπονται παρεμβάσεις στη συμβολοσειρά ανάμεσα στα πακέτα.

Το στοιχείο προγραμματισμού ρύθμισης, είναι ένα στοιχείο ρύθμισης που χαρτογραφεί τα κανάλια ήχου στα εξωτερικά κανάλια, ώστε η κωδικοποίηση πολλών καναλιών να είναι κατά το δυνατόν ευέλικτη.

Τα τρία πιθανά ηχητικά στοιχεία είναι το μονοφωνικό κανάλι, το στερεοφωνικό ζεύγος και ένα κανάλι subwoofer. Κάθε ηχητικό κανάλι ονομάζεται με 4 bits, ώστε να μπορούν να αναπαρασταθούν μέχρι 16 από κάθε ηχητικό στοιχείο στη συμβολοσειρά και να χρησιμοποιηθούν στα κανάλια εξόδου.

Το στοιχείο ζεύξης πολλών καναλιών, είναι ο μηχανισμός κωδικοποίησης ηχητικών στοιχείων κοινών σε δύο ή και παραπάνω κανάλια.

Το στοιχείο δεδομένων, είναι μια συμβολοσειρά που μπορεί να συνεχιστεί σε αυθαίρετο αριθμό πακέτων και περιέχει μια απαρίθμηση, ώστε ο αποκωδικοποιητής να μπορεί να το απομονώσει κατά την αποκωδικοποίηση. Οι συμβολοσειρές του στοιχείου δεδομένων, είναι μέχρι 16 όπως και οι κωδικοί των ηχητικών στοιχείων.

Το στοιχείο συμπλήρωσης, είναι ένας μηχανισμός συμπλήρωσης bits, ώστε να μπορεί ο αποκωδικοποιητής να αυξήσει το ρυθμό της συμπιεσμένης συμβολοσειράς, ώστε να γεμίζει ένα κανάλι σταθερού ρυθμού.

Το στοιχείο τερματισμού σηματοδοτεί το τέλος του πακέτου. Είναι απαραίτητο να κάνει τη συμβολοσειρά χρηστική. Μπορούν να προστεθούν bits συμπλήρωσης μετά το πέρας του πακέτου και μέχρι το επόμενο ώστε να ξεκινάει ακριβώς στα όρια ενός byte.

Τα στοιχεία δεν είναι όλα απαραίτητα στη δημιουργία της συμβολοσειράς και έτσι παρέχεται ευελιξία και ελάχιστη σπατάλη στη συμβολοσειρά.

2.2.6 Εκτίμηση πολυπλοκότητας αποκωδικοποιητή

Η πολυπλοκότητα του αποκωδικοποιητή είναι σημαντική, καθώς καθορίζει την πολυπλοκότητα όλης της υλοποίησης, ειδικά για το κύριο και χαμηλής πολυπλοκότητας προφίλ. Η εκτίμηση της πολυπλοκότητας είναι δύσκολο να ποσοτικοποιηθεί. Μέτρα ποσοτικοποίησης είναι οι εντολές μηχανής, οι καταστάσεις αποθήκευσης ανάγνωσης /εγγραφής (RAM), οι καταστάσεις ανάγνωσης μόνο (ROM) και καθορίζονται για κάθε τμήμα του αποκωδικοποιητή.

Ένας τρόπος για να είναι πιο εύκολη η ταξινόμηση των αποτελεσμάτων, είναι να υποθέσουμε ότι συχνότητα δειγματοληψίας είναι 48 kHz, το μήκος λέξης είναι 16 bits ανά δείγμα και ο ρυθμός δεδομένων είναι 64 kb/s ανά κανάλι, ενώ κάθε πακέτο υποθέτουμε ότι περιέχει 1024 τιμές συχνότητας.

Υπάρχουν δυο κατηγορίες αποκωδικοποιητών AAC ως προς την υλοποίηση. Οι αποκωδικοποιητές λογισμικού που χρησιμοποιούν γενικού σκοπού επεξεργαστές και οι αποκωδικοποιητές υλικού που χρησιμοποιούν ολοκληρωμένα ASIC.

Το ελάχιστο μέγεθος buffer εισόδου είναι 6144 bits. Αυτό προκύπτει από την πρόσθεση της αποθήκης bit του κωδικοποιητή και το μέγιστο ρυθμό δεδομένων ανά κανάλι. Η έξοδος του αποκωδικοποιητή, αν υποθεθεί ότι χρησιμοποιείται 16-bit PCM με διπλό σύστημα buffer, απαιτεί ένα μήκος 1024 bits με μήκος λέξης 16-bits buffer. Ο συνολικός αριθμός των λέξεων μήκους 16 bits για τον buffer εισόδου/εξόδου (RAM) του αποκωδικοποιητή είναι $384 + 1024 = 1408$.

Η αποκωδικοποίηση μιας κωδικής λέξης κατά Huffman γίνεται με το να διατρέξουμε ένα δέντρο κωδικοποίησης Huffman από τον κόμβο της ρίζας μέχρι τον κόμβο φύλλο. Κατά μέσο όρο απαιτούνται 10 εντολές για κάθε αποκωδικοποίηση bit Huffman. Από τη στιγμή που κατά μέσο όρο υπάρχουν 1365 bits ανά πακέτο, ο αριθμός εντολών ανά πακέτο είναι 13650. Η αποκωδικοποίηση Huffman απαιτεί την αποθήκευση του δένδρου και την τιμή που αντιστοιχεί στην κωδική λέξη. Το συνολικό buffer που απαιτείται είναι 995 λέξεις των 16 bits (ROM).

Η αντίστροφη κβάντιση γίνεται μέσω ενός βοηθητικού πίνακα. Αν υποθέσουμε ότι μόνο 854 φασματικοί συντελεστές με 20 kHz bandwidth πρέπει να κβαντιστούν αντίστροφα και να αλλάξουν κλίμακα με κάποιο συντελεστή κλίμακας, ο 16-bit ROM buffer που απαιτείται είναι μεγέθους 256 λέξεων και περιέχει 1708 εντολές.

Αν υποθέσουμε ότι έχουμε μόνο 672 φασματικούς συντελεστές που θα χρειαστούν πρόβλεψη και έναν δεύτερης τάξης προγνώστη (predictor), ο αριθμός των εντολών για τον προγνώστη είναι 66 και ο συνολικός αριθμός εντολών για κάθε πακέτο είναι 44352. Ο απαιτούμενος buffer αποθήκευσης είναι μεγέθους 4032 16-bit λέξεων.

Η μορφοποίηση παραθύρου θορύβου στο κύριο προφίλ χρησιμοποιεί ένα φίλτρο τάξης 20 με 672 φασματικούς συντελεστές. Ο αριθμός εντολών ανά πακέτο είναι 13630. Στο χαμηλής πολυπλοκότητας προφίλ η μορφοποίηση παραθύρου θορύβου χρησιμοποιεί φίλτρο τάξης 12 με συνολικά 8130 εντολές ανά πακέτο. Η μορφοποίηση παραθύρου θορύβου δεν απαιτεί buffers αποθήκευσης με μεγάλο μέγεθος.

Το πολύ απλό τμήμα που εκτελεί την πινακοποίηση δυο καναλιών ενός στερεοφωνικού ζεύγους δεν απαιτεί επιπλέον χώρο αποθήκευσης. Κατά τον υπολογισμό μέσης τιμής/ διαφοράς για ένα 20 kHz bandwidth σήμα, ο συνολικός αριθμός εντολών του στερεοφωνικού ζεύγους που απαιτείται είναι 854.

Η στερεοφωνική κωδικοποίηση έντασης δεν απαιτεί επιπλέον μνήμη, ούτε ROM ούτε RAM. Η πολυπλοκότητα της στερεοφωνικής κωδικοποίησης έντασης αντισταθμίζεται από τη μείωση κατά μια του αριθμού αντίστροφων κβαντίσεων για κάθε συντελεστή που κωδικοποιείται κατά την στερεοφωνική κωδικοποίηση έντασης.

Το φίλτρο IMDCT μήκους 1024, απαιτεί περίπου 20000 εντολές ανά πακέτο, ενώ για το φίλτρο IMDCT μήκους 128 ο συνολικός αριθμός εντολών για κάθε 8 μικρά πακέτα είναι 24576. Η συνολική RAM που απαιτείται είναι 1536 λέξεις και η συνολική ROM μαζί με τους συντελεστές παραθύρων είναι 2270 λέξεις. Οι απαιτήσεις αποθήκευσης είναι για μήκος λέξης μεταξύ 16 και 24 bits ανάλογα με το στάδιο του φίλτρου.

AAC αριθμός εντολών	Μονό κανάλι	5 κανάλια
Κωδικοποίηση Huffman	13.657	68.285
Αντίστροφη κβάντιση	1.708	8.540
Πρόβλεψη	44.352	221.760
Μορφοποίηση παραθύρου θορύβου	13.850	69.250
Υπολογισμός Μέσης τιμής/διαφοράς		1.708
Φίλτρο IMDCT	24.576	122.880
Συνολικά	98.143	492.423

Πίνακας 4 Αριθμός εντολών AAC ανά στάδιο

AAC 16-bit RAM	Μονό κανάλι	5 κανάλια
Buffer εισόδου	384	1.920
Buffer εξόδου	1.024	5.120
Buffer λειτουργίας	2.048	10.240
Πρόβλεψη	4.032	20.160
Φίλτρο IMDCT	1.024	5.120
Συνολικά	8.512	42.560

Πίνακας 5 Απαιτήσεις μνήμης AAC

Σύγκριση των δυο προφίλ	Κύριο προφίλ	Προφίλ χαμηλής πολυπλοκότητας
Εντολές ανά πακέτο	492.423	242.063
RAM	42.560	22.400
ROM	3.545	3.545

Πίνακας 6 Σύγκριση απαιτήσεων των δυο προφίλ

2.3 Το πρότυπο MPEG-4

Το MPEG-4 είναι ακόμα ένα πρότυπο ISO/IEC που αρχικά προτάθηκε το 1992. Οι καινοτομίες που προσέφερε ήταν η οπτικοακουστική κωδικοποίηση σε χαμηλούς ρυθμούς bit, η επεκτασιμότητα και η υβριδική κωδικοποίηση, δηλαδή σημάτων ήχου και βίντεο. Το τελικό επίπεδο της MPEG-4 δημοσιεύτηκε το 1998. Ένα δεύτερο επίπεδο ολοκληρώθηκε το 1999. Μετά από αυτή έχουν προστεθεί διορθώσεις και κάθε τόσο προστίθενται νέα χαρακτηριστικά, που όμως έχουν τον περιορισμό ότι πρέπει να είναι συμβατά με τις παλαιότερες εκδόσεις.

Το πρότυπο MPEG-4 χρησιμοποιείται σε εφαρμογές όπως ενσύρματες, ασύρματες, streaming, ψηφιακή μετάδοση, διαδραστικά πολυμέσα και υψηλής ποιότητας ήχο/βίντεο. Αντί της προτυποποίησης ενός πλήρους προτύπου όπως στα MPEG-1 και MPEG-2, το MPEG-4 καθορίζει κάποια εργαλεία που χρησιμοποιούνται σαν στοιχεία σε διαφορετικούς αλγορίθμους κωδικοποίησης. Καθορίζονται επίσης διαφορετικά προφίλ που αντιπροσωπεύουν διαφορετικά εργαλεία ανά συγκεκριμένη εφαρμογή.

Το πρότυπο MPEG-4 περιέχει και γενική κωδικοποίηση ήχου όπως φωνή, συνθετικό ήχο και μετατροπή κειμένου σε ήχο. Τα κύρια χαρακτηριστικά του είναι επεκτασιμότητα, ειδικά εφέ, χειρισμοί ήχου και σύνθεση τρισδιάστατου ήχου. Η επεκτασιμότητα του MPEG-4 επιτρέπει τη δυναμική ρύθμιση του ρυθμού δεδομένων και της χωρητικότητας του καναλιού. Η ευελιξία

που προκύπτει από αυτή τη δυνατότητα βοηθάει στις εφαρμογές με κανάλια μετάδοσης μεταβλητής χωρητικότητας, όπως το διαδίκτυο και η κινητή τηλεφωνία.

Το πρότυπο MPEG-4 καθορίζει τις παραμέτρους κωδικοποίησης των στοιχείων ήχου, στατικής και κινούμενης εικόνας (βίντεο) και οπτικοακουστικής πληροφορίας σαν πολυμεσικά δεδομένα. Τα δεδομένα αυτά μπορεί να είναι πολυδιάστατα, φυσικά ή συνθετικά. Το βίντεο και ο ήχος μπορούν να είναι ανεξάρτητα ή και να συντίθενται, επίσης μπορούν να είναι φυσική καταγραφή ή να έχουν δημιουργηθεί από επεξεργασία σε υπολογιστή.

Αυτός ο συνδυασμός απαιτεί και έναν μηχανισμό συγχρονισμού και πολύπλεξης των δεδομένων βίντεο και ήχου, για να μπορούν τόσο να μεταφερθούν όσο και να αναπαραχθούν από τα κανάλια πολυμέσων. Επίσης παρέχουν προστασία της πνευματικής ιδιοκτησίας όπως καθορίζεται από το πρότυπο.

Η δομή του προτύπου αποτελείται από διαφορετικά μέρη που το καθένα καθορίζει ένα μέρος του προτύπου. Το πρότυπο ISO/IEC 14496 οργανώνεται ως εξής:

ISO/IEC 14496-1: Συστήματα

ISO/IEC 14496-2: Εικόνα

ISO/IEC 14496-3: Ήχος

ISO/IEC 14496-6: Πλαίσιο μεταφοράς σύνθετων πολυμέσων

ISO/IEC 14496-10: Προηγμένη κωδικοποίηση βίντεο

2.3.1 Στόχοι και λειτουργίες

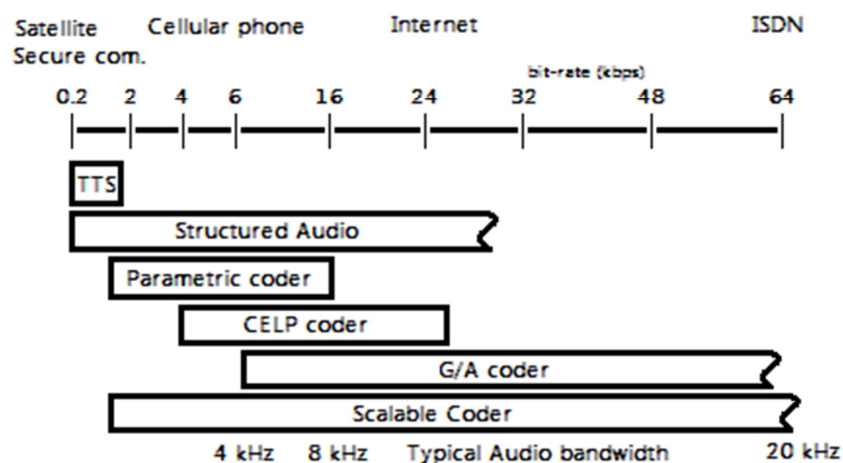
Οι στόχοι του προτύπου MPEG-4 είναι υψηλότεροι λόγω των πολλών ειδών εφαρμογών που έχει σκοπό να καλύψει. Έτσι οι κύριες απαιτήσεις είναι η ευελιξία πρόσβασης τόσο στα κωδικοποιημένα δεδομένα και τα υποσύνολά τους, καθώς και η αποδοτικότητα της κωδικοποίησης λόγω του μεγέθους των δεδομένων που πρέπει να μεταφερθούν, και όλα αυτά ενώ πρέπει να μειώνονται τα σφάλματα μετάδοσης.

Όσον αφορά στον ήχο υπάρχουν και πιο συγκεκριμένοι στόχοι, όπως η κωδικοποίηση φωνής για τηλεπικοινωνιακές εφαρμογές, η εύκολη πρόσβαση στα δεδομένα, ώστε να ικανοποιηθούν οι διαφορετικές και διαφοροποιούμενες απαιτήσεις των διαφορετικών καναλιών μεταφοράς και η ανθεκτικότητα σε σφάλματα σε περιβάλλοντα που τείνουν να δημιουργούν αλλοιώσεις. Επιπλέον, η εύκολη πρόσβαση στα δεδομένα διευκολύνει και το χειρισμό τους κάνοντας πιο εύκολη την επεξεργασία των πολυμέσων με υπολογιστή.

Το MPEG-4 αφορά σε δυο είδη ήχου, τον συνθετικό και τον φυσικό ήχο. Ο συγχρονισμός και η μίξη φυσικού και συνθετικού ήχου ονομάζεται υβριδική κωδικοποίηση συνθετικού/φυσικού ήχου. Η διεπαφή TTS (Text To Speech) του ήχου MPEG-4 ορίζει το πρωτόκολλο μεταφοράς για επεξεργασμένη ομιλία, όπου το σύστημα TTS μεταφράζει πληροφορίες κειμένου σε ομιλία για να μεταφερθεί σε τηλεφωνικές γραμμές που μεταφέρουν σήματα ομιλίας. Τα συστήματα TTS υποστηρίζουν ρυθμούς δεδομένων μεταξύ 200 b/s και 1.2 kb/s. Η διεπαφή SA (Structured Audio) ορίζει τη μεταφορά επεξεργασμένου ήχου γενικά. Υποστηρίζει εξαιρετικά χαμηλούς

ρυθμούς δεδομένων και διάδραση στο λήπτη. Η συμβολοσειρά SA καθορίζει ένα σύνολο αλγορίθμων σύνθεσης που περιγράφουν τη δημιουργία ήχου και ένα σύνολο παραμέτρων ελέγχου που περιγράφουν ποιον ήχο θα δημιουργηθούν. Ο παραμετρικός κωδικοποιητής φωνής πετυχαίνει καλή ποιότητα ήχου μεταξύ 2 και 4 kb/s. Χαμηλότεροι ρυθμοί δεδομένων όπως 1.2 kb/s επιτυγχάνονται με μεταβλητό ρυθμό κωδικοποίησης. Ρυθμοί δεδομένων μεταξύ 4 και 24 kb/s μπορούν να επιτευχθούν με τον κωδικοποιητή CELP (Code-Excited Linear Prediction) [45]. Δυο ρυθμοί δειγματοληψίας υποστηρίζονται, στα 8 και 16 kHz, όπου ο πρώτος χρησιμοποιείται για στενής ζώνης κωδικοποίηση ομιλίας και ο δεύτερος για ευρείας ζώνης κωδικοποίηση ομιλίας.

Ο ήχος γενικά καλύπτει ρυθμούς δεδομένων μεταξύ 6 kb/s για ηχητικά σήματα με bandwidth 4 kHz, και 300 kb/s ανά κανάλι για σήματα με bandwidth πάνω από 20 kHz για μονοφωνικό σε πολλών καναλιών ήχο. Έτσι, επεκτείνεται και η λειτουργικότητα των προτύπων MPEG-1 και MPEG-2. Η επεκτασιμότητα των δεδομένων που απαιτείται από το πρότυπο επιτρέπει στα δεδομένα να προσπελαστούν σε συμβολοσειρές με μικρότερους ρυθμούς δεδομένων και παρ' όλα αυτά να αποκωδικοποιούνται σε ένα σήμα που μπορεί να αναπαραχθεί. Με χρήση λοιπόν αυτής της ιδιότητας πιο απλοί αποκωδικοποιητές/κωδικοποιητές μπορούν να αποκωδικοποιήσουν/κωδικοποιήσουν συμβολοσειρές, οι οποίες στις προηγούμενες κωδικοποιήσεις θα απαιτούσαν μεγαλύτερους σε μέγεθος και πολυπλοκότητα αποκωδικοποιητές/κωδικοποιητές.



Εικόνα 1919 Αλγόριθμοι και εφαρμογές τους ανάλογα με τις συχνότητες που υποστηρίζουν

Η κωδικοποίηση MPEG-4 εισάγει και διάφορα εργαλεία που βοηθούν στην ποιότητα του ήχου για διάφορες εφαρμογές. Όσον αφορά στην ομιλία, τα εργαλεία αυτά αφορούν ειδικά στην κωδικοποίηση φωνής και είναι ο γραμμικός προγνώστης CELP και η παραμετρική κωδικοποίηση ομιλίας HVCX (Harmonic Vector eXcitation Coding).

Όσον αφορά στα σήματα ήχου γενικά, εκτός των βασικών κωδικοποιητών ήχου MPEG-2 AAC περιλαμβάνεται και η Twin VQ (Transform-domain weighted interleave Vector Quantization),

που είναι μια σταθμισμένη τεχνική διανυσματικού κβαντισμού για συμπίεσης ηχητικών σημάτων. Ακόμα, υπάρχουν εργαλεία που βελτιώνουν την επίδοση του MPEG-2 AAC, όπως η αντικατάσταση θορύβου PNS (Perceptual Noise Substitution) και η πρόβλεψη LTP (Long Term Prediction).

Η επεκτασιμότητα προωθείται από τη δυνατότητα που δίνεται να αποκωδικοποιούνται τα δεδομένα σε διαφορετικούς ρυθμούς δεδομένων. Έτσι, μια συμβολοσειρά μπορεί πλέον να αποκωδικοποιηθεί σε πολύ διαφορετικά bandwidth ανάλογα με τις απαιτήσεις.

Τα εργαλεία σύνθεσης που προσφέρονται βοηθούν στη δημιουργία τεχνητών ήχων είτε με την επεξεργασία φυσικών ήχων, είτε με την εξαρχής σύνθεση τους με υπολογιστικά μέσα. Τέτοια εργαλεία είναι τα SA, SASBF και MIDI.

Σε αυτό το πρότυπο υποστηρίζεται και ο συνδυασμός σημάτων βίντεο και ήχου για τη δημιουργία οπτικοακουστικών σκηνών σε ένα συνολικό αποτέλεσμα. Αυτή η επεξεργασία μπορεί να γίνει σε πολλά επίπεδα, καθώς μπορούν να υπάρξουν ακόμα και μίξεις ηχητικών σημάτων στο κομμάτι του ήχου του τελικού αποτελέσματος. Στο streaming μπορεί να γίνει ειδικός χειρισμός στο πως τα ηχητικά σήματα μεταδίδονται.

Μια άλλη σημαντική εργαλειοθήκη αφορά στην προστασία από σφάλματα, που είναι χρήσιμη ή ακόμα και απαραίτητη, σε μερικά περιβάλλοντα καναλιών που είναι επιρρεπή στο θόρυβο. Αυτά τα εργαλεία βελτιώνουν την ανθεκτικότητα της τελικής κωδικοποίησης στο θόρυβο. Τα εργαλεία που αφορούν στη μείωση του θορύβου είναι η εμπρός προστασία σφάλματος FEC (Forward Error Correction) και ο κυκλικός κώδικας πλεονασμού CRC (Cyclic Redundancy Code).

2.3.2 Επιδόσεις του προτύπου MPEG-4

Οι επιδόσεις της κωδικοποίησης MPEG-4 κρίθηκαν όπως και για όλες τις άλλες κωδικοποιήσεις με δοκιμασίες που κρίνουν την αντικειμενική απόδοση των αλγορίθμων. Στις δοκιμασίες αυτές συμπεριλήφθηκαν κωδικοποιητές του προτύπου MPEG-2 και ITU-T. Τα συγκριτικά αποτελέσματα των δοκιμασιών με βάση την ITU-R κλίμακα 5 βαθμών δυσκολίας φαίνονται στον πίνακα 7.

Εργαλείο Κωδικοποίησης	Αριθμός καναλιών	Ρυθμός δεδομένων	Κλίμακα αξιολόγησης	Τυπική ποιότητα
AAC	5	320 kb/s	Δυσκολία	4,6
95 MPEG LII BC	5	640 kb/s	Δυσκολία	4,6
AAC	2	128 kb/s	Δυσκολία	4,8
AAC	2	96 kb/s	Δυσκολία	4,4
MPEG-1 LII	2	192 kb/s	Δυσκολία	4,3
MPEG-1 LIII	2	128 kb/s	Δυσκολία	4,1
AAC	1	24 kb/s	Ποιότητα	4,2

CELP	1	6 kb/s + 18 kb/s	Ποιότητα	3,7
TwinVQ	1	6 kb/s + 18 kb/s	Ποιότητα	3,6
AAC	1	18 kb/s	Ποιότητα	3,2
G.723	1	6,3 kb/s	Ποιότητα	2,8
CELP Ευρείας συχνότητας	1	18,2 kb/s	Ποιότητα	2,3
BSAC	2	96 kb/s	Ποιότητα	4,4
BSAC	2	80 kb/s	Ποιότητα	3,7
BSAC	2	64 kb/s	Ποιότητα	3,0
AAC-LD (20ms)	1	64 kb/s	Ποιότητα	4,4
G.722	1	32 kb/s	Ποιότητα	4,2
AAC-LD (30ms)	1	32 kb/s	Ποιότητα	3,4
CELP Στενής συχνότητας	1	6 kb/s	Ποιότητα	2,5
TwinVQ	1	6 kb/s	Ποιότητα	1,8
HILN	1	16 kb/s	Ποιότητα	2,8
HILN	1	6 kb/s	Ποιότητα	1,8

Πίνακας 7 Αξιολογήσεις κωδικοποιήσεων MPEG με βάση το εργαλείο κωδικοποίησης, τον αριθμό καναλιών και το ρυθμό δεδομένων

2.3.3 Προστασία πνευματικών δικαιωμάτων

Η προστασία των πνευματικών δικαιωμάτων ήταν πάντα σημαντική απαίτηση και στην ανάπτυξη του προτύπου MPEG-4 αυτό λήφθηκε σοβαρά υπ' όψη. Στην κατοχύρωση των πνευματικών δικαιωμάτων αφιερώνεται ένα μεγάλο τμήμα των προδιαγραφών του MPEG-4. Ο προσδιορισμός των πνευματικών δικαιωμάτων των πολυμεσικών αντικειμένων γίνεται μέσω του προτύπου διεπαφής διαχείρισης και προστασίας πνευματικής ιδιοκτησίας IPMP (Intellectual Property Management and Protection Extension) [46] που αναπτύχθηκε καθαρά για την MPEG-4. Η ευρεία γκάμα εφαρμογών που στοχεύει να καλύψει η MPEG-4, δηλαδή εφαρμογές όπως διαδικτυακή ομιλία χαμηλών ρυθμών δεδομένων μέχρι και υψηλής πιστότητας ήχο και βίντεο, υποστηρίζονται ακόμα και με ένα κατάλληλο υποσύνολο των προφίλ MPEG-4. Το επίπεδο προστασίας αλλάζει ανάλογα με το περιεχόμενο, την πολυπλοκότητα και τα μοντέλα αγοράς. Τα συστήματα IPMP δεν καθορίζονται σε επίπεδο προτύπου, όμως η διεπαφή IPMP ορίζεται συγκεκριμένα σαν επέκταση της αρχιτεκτονικής συστημάτων MPEG-4 μέσω στοιχείων περιγραφής (IPMP-D) και στοιχειωδών συμβολοσειρών (IPMP-ES). Τα δικαιώματα της MPEG-4 AAC διατίθενται από τα Dolby Laboratories και τα δικαιώματα της MPEG-4 βίντεο από την εταιρεία MPEG LA, LLC.

2.4 Το Πρότυπο MPEG – D

Στόχος του προτύπου MPEG-D είναι η βελτίωση της διαδραστικότητας (interactivity) καθώς και η προσωποποίηση (personalization) των χαρακτηριστικών του ηχητικού σήματος στις προτιμήσεις του χρήστη. Η MPEG-D εμφανίζεται σε τρεις εκδοχές: την MPEG Surround ή Spatial Audio Coding, την Spatial Audio Object Coding (SAOC) και την Unified speech and audio coding.

Η MPEG Surround ή Spatial Audio Coding, υποστηρίζει παραμετρική κωδικοποίηση για ηχητικά σήματα πολλών καναλιών.

Η Spatial Audio Object Coding (SAOC) επιτρέπει την αποθήκευση και την μεταφορά μεμονωμένων ηχητικών αντικειμένων σε ηχητικές μίξεις, διατηρώντας την επιλογή να προσαρμόσει ο χρήστης τη μίξη με βάση το προσωπικό του γούστο. Αυτό γίνεται με αλλαγή της ρύθμισης αναπαραγωγής από τη στερεοφωνική σε περιβάλλοντα ήχο (surround) ή και ρύθμιση για ακουστικά. Επίσης, υποστηρίζει την παραμετρική κωδικοποίηση ξεχωριστών ηχητικών αντικειμένων και παρέχει τη δυνατότητα επιλογής ανάμεσα σε διάφορες ρυθμίσεις αναπαραγωγής (1.0, 2.0, 5.0 κ.α.). Η κωδικοποίηση SAOC μεταδίδει ηχητικά αντικείμενα κοινών μονοφωνικών ή στερεοφωνικών σημάτων ώστε να μπορούν να αναπαραχθούν ξεχωριστά τα ηχητικά αντικείμενα μιας ηχητικής σκηνής. Στην SAOC περιέχονται οι αλγόριθμοι OLD (Object Level Differences), IOC (Inter-Object Cross Coherences) και DCLD (Downmix Channel Level Differences). Ο αποκωδικοποιητής SAOC μετατρέπει την αναπαράσταση παραμέτρου SAOC σε αναπαράσταση παραμέτρου MPEG Surround που αποκωδικοποιείται στη συνέχεια από ένα MPEG Surround αποκωδικοποιητή. Η ηχητική πληροφορία της μουσικής σκηνής αποκωδικοποιείται πάντα με βάση τις προσωποποιημένες επιλογές του ακροατή.

Οι εφαρμογές της κωδικοποίησης SAOC είναι διάφορες, όπως η εικονική μίξη ηχητικών σημάτων, το διαδραστικό gaming, οι τηλεπικοινωνιακές εφαρμογές και άλλες. Αναλυτικότερα:

- Στην εικονική μίξη τα μουσικά όργανα μπορούν να συνδυαστούν σαν να εκτελούσαν ταυτόχρονα κάποιο κομμάτι. Ακόμα, μπορεί να ρυθμιστεί το επίπεδο αναπαραγωγής των διαλόγων σε ταινίες και άλλες μεταδόσεις ώστε να είναι πιο εύληπτοι.
- Στο διαδραστικό gaming, ο SAOC είναι ένας αποδοτικός τρόπος αποθήκευσης και υπολογισμού της αναπαραγωγής ήχου. Η μετακίνηση στην εικονική σκηνή συνεπάγεται την αλλαγή των παραμέτρων αναπαραγωγής του αντικειμένου. Οι πολλαπλοί παίκτες σε ένα διαδικτυακό παιχνίδι χρησιμοποιούν τη μεταφορά SAOC για την αναπαράσταση όλων των ηχητικών αντικειμένων που είναι ξένα προς το τερματικό κάθε παίκτη.
- Στις τηλεπικοινωνιακές υποδομές που είναι μονοφωνικές, τα τερματικά που υποστηρίζουν SAOC λαμβάνουν πολλές πηγές ήχου, δηλαδή αντικείμενα, και παράγουν μονοφωνικά σήματα, που μεταφέρονται με υπάρχοντες κωδικοποιητές φωνής. Οι παράπλευρες πληροφορίες μεταφέρονται με ενσωματωμένο τρόπο. Τα τερματικά SAOC βελτιώνουν το σήμα προσθέτοντας χωρικό διαχωρισμό σε διαφορετικά ηχεία.

2.5 Το Πρότυπο MPEG – H

Η κωδικοποίηση MPEG-H είναι ένα σύνολο προτύπων που ακόμα είναι υπό ανάπτυξη. Το σύνολο αυτών των προτύπων περιέχει ένα πρότυπο κωδικοποίησης βίντεο, ένα πρότυπο κωδικοποίησης ήχου και δυο πρότυπα ελέγχου συμμόρφωσης.

Η δομή του προτύπου MPEG-H αποτελείται συγκεκριμένα από τα παρακάτω μέρη:

- MPEG-H μέρος 1: Μεταφορά δεδομένων MPEG (MMT, Media Transport)
- MPEG-H μέρος 2: Κωδικοποίηση βίντεο υψηλής απόδοσης (HEVC, High Efficiency Video Coding)
- MPEG-H μέρος 3: 3D ήχος
- MPEG-H μέρος 4: MMT λογισμικό αναφοράς
- MPEG-H μέρος 5: HEVC λογισμικό αναφοράς
- MPEG-H μέρος 6: 3D ήχος λογισμικό αναφοράς
- MPEG-H μέρος 7: MMT έλεγχος συμμόρφωσης
- MPEG-H μέρος 8: HEVC έλεγχος συμμόρφωσης
- MPEG-H μέρος 9: 3D ήχος έλεγχος συμμόρφωσης
- MPEG-H μέρος 10: MMT FEC κώδικες
- MPEG-H μέρος 11: MMT κωδικοποίηση σύνθεσης
- MPEG-H μέρος 12: HEVC image μορφή αρχείου
- MPEG-H μέρος 13: MMT οδηγίες εφαρμογής

Η κωδικοποίηση MPEG-H παρέχει εκτός του βίντεο και ένα ολοκληρωμένο σύστημα ήχου επιπέδου «home theatre» που φτάνει μέχρι και τον τρισδιάστατο ήχο με χρήση πολλών ηχείων, ώστε να επιτύχει την καλύτερη δυνατή ηχητική εμπειρία. Αυτό γίνεται με την τοποθέτηση των ηχείων σε κατάλληλα σημεία σε σχέση με τη θέση και σε κατάλληλο ύψος του χρήστη. Αυτή η τοπολογία των ηχείων δημιουργεί την αίσθηση ότι ο χρήστης μπορεί να εντοπίσει την κατεύθυνση από την οποία προέρχεται ο ήχος ακόμα και την απόσταση. Η τοποθέτηση ηχείων σε διάφορα σημεία δεν είναι υποχρεωτική, καθώς το πρότυπο μπορεί να παράξει τρισδιάστατο ήχο ακόμα και από ένα μόνο ηχείο φορητής συσκευής ή και ακουστικά που να το υποστηρίζουν.

2.5.1 Απαιτήσεις προτάσεων για τον τρισδιάστατο ήχο

Το τελικό πρότυπο τρισδιάστατου ήχου πρέπει να δέχεται τόσο εισόδους που παρέχονται ανά κανάλι, όσο και εισόδους που παρέχονται ανά ηχητικό αντικείμενο. Επίσης, πρέπει να έχει μια κεντρική δομή, να αξιοποιεί την υπάρχουσα τεχνολογία MPEG, όπου αυτό είναι εφικτό, με τις καινοτομίες να εισάγονται όταν παρέχουν σημαντική βελτίωση ή παρέχει λειτουργικότητες που δεν υποστηρίζονταν πριν.

Οι κύριες απαιτήσεις του προτύπου που πρέπει να πληρούνται απαραίτητα είναι οι παρακάτω:

- Υψηλή ποιότητα: Η ποιότητα του ήχου πρέπει όχι μόνο να ξεπερνά τα προηγούμενα πρότυπα, αλλά να είναι κατάλληλη για σύγχρονες εφαρμογές υψηλής ποιότητας ώστε να παρέχουν ήχο κατά το δυνατόν μέγιστης ευκρίνειας.
- Εντοπισμός θέσης και εμπειρία ήχου: Η δυνατότητα του χρήστη να μπορεί να «αισθάνεται» την κατεύθυνση και την απόσταση του ήχου είναι σημαντική, ώστε η εμπειρία ήχου που αντιλαμβάνεται ο χρήστης να είναι βέλτιστη.
- Αναπαραγωγή σε συστήματα με λιγότερα ηχεία: Η αναπαράσταση του ήχου θα πρέπει να μπορεί να γίνει σε μικρότερο αριθμό ηχείων από τη βασική ρύθμιση του προγράμματος κωδικοποίησης. Το σήμα στα λιγότερα ηχεία θα πρέπει να έχει τη μέγιστη αντικειμενική ποιότητα σε σχέση με το σήμα αναφοράς.
- Ελαστική τοποθέτηση ηχείων: Η αναπαράσταση του ήχου πρέπει να μπορεί να αποκωδικοποιηθεί και αναπαραχθεί σε μια εγκατάσταση ηχείων με ηχεία σε διαφορετικές και πιθανόν λιγότερες τοποθεσίες, με τη βέλτιστη δυνατή ποιότητα.
- Καθυστέρηση: Η καθυστέρηση θα πρέπει να είναι πολύ χαμηλή για να υποστηρίζει δεδομένα πραγματικού χρόνου. Η μονομερής αλγοριθμική καθυστέρηση δεν θα πρέπει να ξεπερνά το 1 δευτερόλεπτο.
- Τρισδιάστατος ήχος: Θα πρέπει να δέχεται σήματα PCM ανά κανάλι τουλάχιστον 22 πλήρους bandwidth καναλιών και 2 χαμηλής συχνότητας που ρυθμίζονται για άμεση αναπαραγωγή. Είναι δυνατή η αποδοχή διακριτών ηχητικών αντικειμένων PCM με τις σχετικές πληροφορίες αναπαράστασης, θέσης, σκηνής. Είναι αποδεκτή η χρήση PCM σημάτων που αναπαριστούν Higher Order Ambisonics (ηχοτόπια υψηλής τάξης) [47].
- Ακρόαση με ακουστικά: Το πρότυπο θα πρέπει να υποστηρίζει αναπαράσταση για αμφιωτική ακρόαση, όπως γίνεται στα ακουστικά. Η ακρόαση με ακουστικά θα πρέπει να έχει χαρακτηριστικά κατά το δυνατόν πιο κοντά στην ακρόαση με ηχεία. Ακόμα, ο αποκωδικοποιητής θα πρέπει να μπορεί να αναπαράγει τα δεδομένα με τρόπο που καθορίζεται από το χρήστη με βάση τις προτιμήσεις ακρόασής του.

Επιπλέον των παραπάνω υπάρχουν και δευτερεύουσες απαιτήσεις που θεωρείται πλεονέκτημα να πληρούνται για την πρόταση που τα περιλαμβάνει.

- Υπολογιστική πολυπλοκότητα: Η πολυπλοκότητα θα πρέπει να μην είναι απαγορευτική για το σενάριο της εφαρμογής. Παράδειγμα είναι η μετάδοση μέσω διαδικτύου, που ενώ η πολυπλοκότητα του κωδικοποιητή δεν υπόκειται σε αυστηρούς περιορισμούς, η πολυπλοκότητα του αποκωδικοποιητή θα πρέπει να είναι χαμηλή.
- Διαδραστικότητα: Θα πρέπει να είναι δυνατή η τροποποίηση μιας ηχητικής σκηνής σε σχέση με την κωδικοποιημένη αναπαράσταση.

Οι εφαρμογές της MPEG-H 3D κωδικοποίησης ήχου περιλαμβάνουν Home Theatre με υποστήριξη ακόμα και περισσότερων από 20 ηχείων, προσωπική τηλεόραση υψηλής ανάλυσης και αρκετών ηχείων, τηλεόραση για smartphone με υποστήριξη τρισδιάστατου ήχου και προσομοίωση της αναπαράστασης 22.2 καναλιών και αποκωδικοποιητές χαμηλότερης πολυπλοκότητας για χαμηλότερη κατανάλωση και δυο ή περισσότερα ηχεία στη συσκευή. Επιπλέον, προγράμματα πολλών καναλιών όπως αναπαραγωγή ήχου home theatre, χωρίς τον περιορισμό συγχρονισμού με την εικόνα.

Σε όλες τις εφαρμογές πρέπει να υποστηρίζεται τρισδιάστατος ήχος και καλύτερη εμπειρία ήχου από το τρέχον 5.1 . Οι κύριοι στόχοι είναι υψηλής ποιότητας κωδικοποίηση και μετάδοση χωρίς σφάλματα με επιδόσεις όσον αφορά στον εντοπισμό του ήχου και την ποιότητα της εμπειρίας. Σύμφωνα με όλα τα παραπάνω θα υπάρχει ανάγκη συμπίεσης του συστήματος ήχου πολλών καναλιών και η αναπαραγωγή τους σε άλλους αριθμούς ηχείων.

2.5.2 Μορφή εισόδου κωδικοποιητή τρισδιάστατου ήχου

Οι μορφές εισόδου του ήχου MPEG-H 3D είναι είσοδος καναλιού, είσοδος ηχητικού αντικειμένου και HOA είσοδο. Θα τις εξετάσουμε αναλυτικά στις επόμενες ενότητες.

2.5.2.1 Είσοδος καναλιού

Η είσοδος καναλιού είναι ένα σύνολο σημάτων μονοφωνικών καναλιών, όπου κάθε κανάλι είναι ένα μονοφωνικό .WAV αρχείο. Τα αρχεία .WAV ονοματοδοτούνται με βάση τη σύμβαση:

< όνομα_αντικειμένου > _A < γωνία_αζιμούθιου > _E < γωνία_ανύψωσης > .wav

Οι γωνίες αζιμούθιου είναι γωνίες σε ± 180 μοίρες. Οι θετικές τιμές συμβολίζουν κατεύθυνση προς τα αριστερά. Οι τιμές που λαμβάνουν αναπαρίστανται με τρία ψηφία και πρόσημο.

Οι γωνίες ανύψωσης δίνονται σε γωνίες ± 90 μοίρες, με θετικές τιμές να συμβολίζουν την κατεύθυνση προς τα πάνω. Οι τιμές που λαμβάνουν αναπαρίστανται με δυο ψηφία και πρόσημο.

Στα κανάλια χαμηλής συχνότητας η σύμβαση είναι :

< όνομα_αντικειμένου > _LFE < αριθμός_καναλιού > .wav

όπου ο αριθμός καναλιού είναι 1 ή 2.

Το αρχείο .WAV γίνεται δεκτό αν οι γωνίες αζιμούθιου και ανύψωσης είναι εντός των ορίων που καθορίζονται για κάθε ηχείο. Τα αρχεία .WAV εισόδου χαρτογραφούνται σε αντίστοιχα ηχεία χωρίς τροποποίηση του περιεχομένου. Το περιεχόμενο βρίσκεται σε 48 kHz και 24 bits.

2.5.2.2 Είσοδος αντικειμένου

Η είσοδος αντικειμένου μεταφέρεται σαν αρχείο μεταδεδομένων και ένα σύνολο μονοφωνικών ηχητικών σημάτων, όπου κάθε ηχητικό σήμα αναπαρίσταται σε μονοφωνικό .WAV αρχείο. Το περιεχόμενο μπορεί να είναι τόσο ένα αντικείμενο όσο και ένα ηχητικό σήμα καναλιού. Τα αρχεία είναι δειγματοληπτημένα στα 48 kHz και 24 bit.

Η σύμβαση ονοματοδοσίας είναι:

< όνομα_αντικειμένου > _ < κωδικός_αριθμός_αντικειμένου > .wav

όπου ο κωδικός αριθμός αντικειμένου είναι ένας τριών ψηφίων αριθμός που μετριέται από το μηδέν.

Μαζί με τα δεδομένα του αντικειμένου υπάρχει και το αρχείο μεταδεδομένων (<όνομα_αντικειμένου>.OAM) που περιγράφει τα δεδομένα για τη συσχέτιση ηχητικών σκηνών που περιέχουν τόσο κανάλια όσο και αντικείμενα. Αυτό το αρχείο περιέχει τον αριθμό των αντικειμένων μιας σκηνής και τον αριθμό και τα ονόματα όλων των αρχείων των σημάτων καναλιών που ανήκουν σε αυτή τη σκηνή. Το αρχείο ξεκινά με κεφαλίδα που παρέχει τη συνολική πληροφορία στην περιγραφή της σκηνής. Την κεφαλίδα ακολουθεί μια σειρά πεδίων περιγραφών καναλιών και αντικειμένων. Στα ηχητικά σήματα καναλιών απαιτείται μόνο το όνομα του αρχείου.

Οι ιδιότητες του αντικειμένου πρέπει να είναι μεταβλητές με το χρόνο. Έτσι στα πεδία περιγραφής του αντικειμένου περιέχεται μια χρονική σφραγίδα (timestamp) και ένας δείκτης ηχητικού αρχείου. Τα πεδία περιγραφής των χρονικών σφραγίδων όλων των αντικειμένων πλην του πρώτου, πρέπει να είναι αύξοντα σε σχέση με τη χρονική σφραγίδα του προηγούμενου αντικειμένου του ίδιου δείκτη αντικειμένου στο αρχείο. Για μια συγκεκριμένη χρονική σφραγίδα οι περιγραφές αντικειμένων όλων των αντικειμένων στη σκηνή θα εισαχθούν με αύξουσα σειρά των δεικτών αντικειμένων τους.

Οι περιγραφές αντικειμένων σε μια σκηνή επαναλαμβάνονται με περίοδο 1024 δειγμάτων, ακόμα και εάν δεν υπάρχει αλλαγή στις ιδιότητες. Η επανάληψη απλοποιεί την υλοποίηση και αποτελεί δείκτη του σημείου γύρω από το οποίο οι θέσεις πρέπει να παραχθούν.

Η κεφαλίδα του αρχείου πρέπει να ακολουθείται τουλάχιστον από τα πεδία περιγραφής καναλιού <number_of_channel_signals> με τα αντίστοιχα ονόματα των WAV αρχείων, και <number_of_object_signals> με το αντίστοιχο πρώτο δείγμα των WAV αρχείων.

Η ενδεικτική μορφή που ορίζεται από το πρότυπο είναι:

Σύνταξη	Αριθμός bytes	Μορφή δεδομένων
<pre>description_file () { scene_description_header () while (end_of_file == 0) { for (i=0; i< number_of_object_signals >;i++){ object_metadata (i) } } }</pre>		

Σύνταξη	Αριθμός bytes	Μορφή δεδομένων
<pre> scene_description_header () { format_id_string format_version if(format_version >2){ hasDynamicObjectPriority } number_of_channel_signals number_of_object_signals description_string for (i=0; i<number_of_channel_signals; i++) { channel_file_name } for (i=0; i<number_of_object_signals; i++) { object_description } } </pre>	<p>4</p> <p>2</p> <p>2</p> <p>2</p> <p>2</p> <p>32</p> <p>64</p> <p>64</p>	<p>Char</p> <p>Unsigned int</p> <p>Unsigned int</p> <p>Unsigned int</p> <p>Unsigned int</p> <p>Char</p> <p>Char</p> <p>Char</p>

Η σημαία **hasDynamicObjectPriority** καθορίζει εάν υπάρχει προτεραιότητα δυναμικού αντικειμένου στα μεταδεδομένα αντικειμένου.

Σύνταξη	Αριθμός bytes	Μορφή δεδομένων
<pre> object_metadata () { sample_index object_index position_azimuth position_elevation position_radius gain_factor if (format_version > 1) { spread } if (format_version > 2) { if (hasDynamicObjectPriority) { dynamic_object_priority } } } </pre>	<p>8</p> <p>2</p> <p>4</p> <p>4</p> <p>4</p> <p>4</p> <p>4</p> <p>4</p> <p>4</p>	<p>Unsigned int</p> <p>Unsigned int</p> <p>32-bit float</p> <p>32-bit float</p> <p>32-bit float</p> <p>32-bit float</p> <p>32-bit float</p> <p>32-bit float</p> <p>32-bit float</p>

sample_index: χρονική σφραγίδα δείγματος, που αφορά στην περιγραφή του αντικειμένου. Το πρώτο δείγμα έχει Δείκτη_δείγματος = 0.

object_index: αριθμός αντικειμένου συγκεκριμένου σήματος. Το πρώτο αντικείμενο έχει δείκτη_αντικειμένου = 0 .

position_radius: θέση του αντικειμένου, μετριέται σε μέτρα

gain_factor: πολλαπλασιαστής κέρδους αντικειμένου

spread: Παράμετρος που καθορίζει τη γωνιακή έκταση της περιοχής στην οποία κατανέμεται η ενέργεια του ηχητικού στοιχείου, μετριέται σε μοίρες.

dynamic_object_priority: Αυτό το πεδίο παίρνει τιμές από 0 έως 7. Το αντικείμενο μπορεί να απορριφθεί εάν έχει προτεραιότητα μικρότερη του 7. Εάν κάποια αντικείμενα απορριφθούν, προτιμώνται αυτά που έχουν τη χαμηλότερη προτεραιότητα.

Με αυτό τον τρόπο κάθε αντικείμενο αποκτά θέσεις (αζιμούθιο, ανύψωση, ακτίνα) στις διάφορες χρονικές σφραγίδες. Ο σκοπός του αναπαραγωγέα είναι να υπολογίσει τα βέλτιστα σήματα ηχείων για τη θέση του χρήστη, ώστε να έχει την κατά το δυνατόν πιο ρεαλιστική εμπειρία.

2.5.2.3 Higher Order Ambisonics HOA είσοδος

Η είσοδος σ' αυτή την περίπτωση είναι ένα σύνολο σημάτων μονοφωνικών καναλιών, σε .WAV αρχεία, των 32 bits με ρυθμό δειγματοληψίας 48 kHz. Το περιεχόμενο κάθε αρχείου είναι πραγματικό σήμα συντελεστών στο πεδίο του χρόνου που ονομάζεται στοιχείο HOA.

Από αυτό προκύπτει η περιγραφή πεδίου ήχου:

$$p(k, r, \theta, \phi) = \sum_{n=0}^N \sum_{m=-n}^n i^n B_n^m(k) j_n(kr) Y_n^m(\theta, \phi) \quad (4)$$

Οι πραγματικοί συντελεστές HOA δίνονται από τον αντίστροφο μετασχηματισμό Fourier:

$$b_n^m(t) = i\mathcal{F}_t \{ B_n^m(k) \} \quad (5)$$

Οι $Y_n^m(\theta, \phi)$ είναι οι σφαιρικές αρμονικές που αντιστοιχούν στους πραγματικούς συντελεστές HOA.

Η $j_n(kr)$ είναι η N-οστής τάξης σφαιρική συνάρτηση Bessel.

Ο αναπαραγωγέας παρέχει τα σήματα εξόδου για μια σφαιρική διάταξη ηχείων, όπου εάν τα ηχεία δεν είναι σε αυτή τη διάταξη πρέπει να γίνει μετασχηματισμός των σημάτων πριν αναπαραχθούν στα ηχεία.

Τα μονοφωνικά .WAV αρχεία ακολουθούν την παρακάτω σύμβαση ονοματοδοσίας :

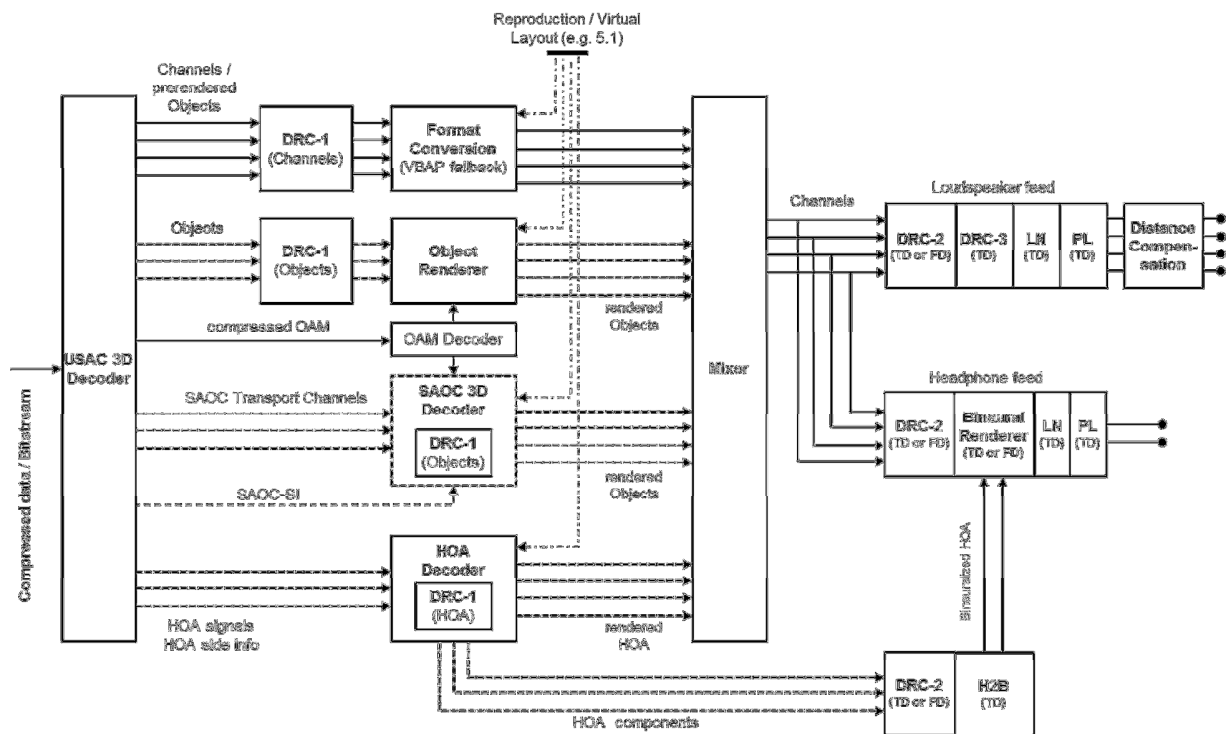
$$\langle \text{όνομα_αρχείου} \rangle _ \langle N \rangle _ \langle n \rangle \langle \mu \rangle \langle \pm \rangle .wav$$

Όπου N είναι η τάξη των HOA και $n \pm$ είναι ένα στοιχείο τάξης n και $\mu = \text{abs}(m)$, $\pm = \text{sign}(m)$ όπου m είναι ο δείκτης αζιμουθιακής συχνότητας.

2.5.3 Προφίλ τρισδιάστατου ήχου

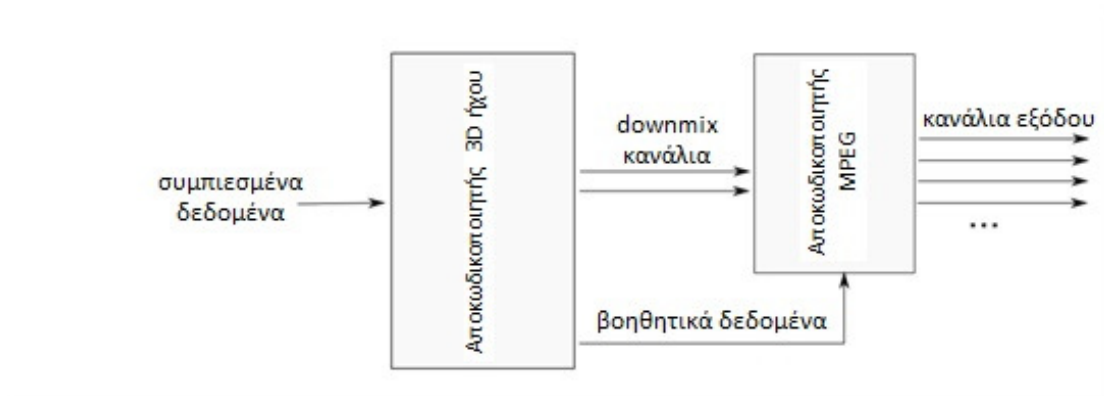
Το κύριο προφίλ του MPEG-H 3D ήχου περιέχει μια σειρά από εργαλεία αποκωδικοποιητή:

- MPEG-H 3D ήχου πυρήνα αποκωδικοποιητή
- Higher Order Ambisonics αναπαραγωγέας
- SAOC 3D αναπαραγωγέας
- Στατικά μεταδεδομένα αντικειμένου (MAE) και αναπαραγωγή
- Δυναμικά μεταδεδομένα αντικειμένου (OAM) και αναπαραγωγή
- Γενική αναπαραγωγή ηχείων/Αλλαγή μορφής
- Surround αναπαραγωγή ηχείων/Αλλαγή μορφής
- Αναπαραγωγή για δυο αυτιά στο πεδίο χρόνου ή/και συχνότητας
- Αναπαραγωγή για ακουστικά
- Μεταδεδομένα έντασης
- Επεξεργασία DRC



Εικόνα 2020 Σχηματικό διάγραμμα MPEG-H 3D ήχου

Η έξοδος του πυρήνα αποκωδικοποιητή 3D ήχου των καναλιών και αντικειμένων μπορούν να επεξεργαστούν περαιτέρω από το MPEG Surround (MPS). Η σχετική σύνδεση του αποκωδικοποιητή 3D ήχου με τον MPS αποκωδικοποιητή είναι η παρακάτω:



Εικόνα 211 Δομικό διάγραμμα αποκωδικοποιητή 3D ήχου σε σύνδεση με ένα MPEG Surround αποκωδικοποιητή

Η ρύθμιση εισερχόμενης συμβολοσειράς αποφασίζει ποια εργαλεία πυρήνα αποκωδικοποιητή είναι ενεργά και εάν μέρη του κωδικοποιητή (codec) του πυρήνα λειτουργούν σε QMF στο πεδίο της συχνότητας. Εάν το τελευταίο στάδιο είναι σε πεδίο QMF, τότε η επεξεργασία του codec ηχητικού πυρήνα γίνεται στο πεδίο της συχνότητας. Αλλιώς, η επεξεργασία γίνεται στο πεδίο του χρόνου.

Εάν η επεξεργασία του codec ηχητικού πυρήνα γίνεται στο πεδίο του χρόνου και κάποιο πακέτο αναπαραγωγής περιεχομένου και μεταεπεξεργασίας περιεχομένου απαιτεί πράξεις στο πεδίο συχνότητας, τότε ο μίκτης λειτουργεί στο πεδίο της συχνότητας. Αλλιώς, λειτουργεί στο πεδίο του χρόνου.

Εάν η επεξεργασία του codec ηχητικού πυρήνα γίνεται στο πεδίο του χρόνου και το επόμενο πακέτο αναπαραγωγής απαιτεί λειτουργία στο πεδίο συχνότητας, τότε πρέπει να εφαρμοστεί QMF ανάλυση και έξοδος πεδίου συχνότητας για το πακέτο αναπαραγωγής.

Εάν το πακέτο αναπαραγωγής λειτουργεί στο πεδίο συχνότητας και ο μίκτης λειτουργεί στο πεδίο του χρόνου, τότε εφαρμόζεται QMF σύνθεση μετά τον αναπαραγωγέα και το αποτέλεσμα είναι στο πεδίο του χρόνου.

Εάν ο μίκτης λειτουργεί στο πεδίο του χρόνου και κάποιο πακέτο σε περιεχόμενο μεταεπεξεργασίας απαιτεί λειτουργία στο πεδίο της συχνότητας, τότε μετατρέπεται σε πεδίο συχνότητας μετά το μίκτη και μετά τη λειτουργία στο πεδίο συχνότητας μετατρέπεται σε πεδίο χρόνου.

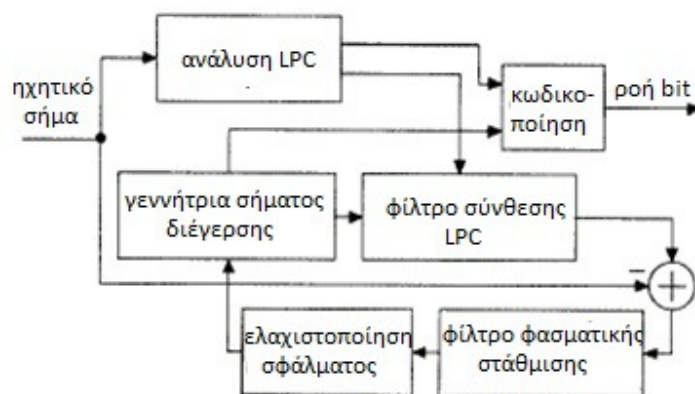
Στο περιεχόμενο εξόδου όλα τα πακέτα λειτουργούν στο πεδίο του χρόνου.

2.6 Σύγκριση προτύπων MPEG-1 και MPEG-2 με το πρότυπο MPEG-4

Οι κωδικοποιήσεις MPEG-1 και MPEG-2 βασίστηκαν στη βελτίωση του αρχικού σήματος ήχου μέσω της συχνοτικής αναπαράστασης, η οποία είναι πιο αποδοτική από την απλή δειγματοληψία PCM που λειτουργεί στο πεδίο του χρόνου. Αυτό συμβαίνει επειδή τα διαδοχικά πλαίσια επεξεργασίας του σήματος στη συχνότητα παρουσιάζουν στατιστική στασιμότητα, πράγμα το οποίο δεν ισχύει για τα διαδοχικά πλαίσια επεξεργασίας του σήματος στο χρόνο. Πέρα από τη στατιστική ανάλυση των σημάτων για την αφαίρεση των στατιστικών πλεονασμών από αυτά, επιστρατεύτηκαν επίσης και τα ψυχοακουστικά μοντέλα αντίληψης της ανθρώπινης ακοής. Αυτά βοήθησαν στην εστίαση της κωδικοποίησης στα χαρακτηριστικά εκείνα του σήματος στα οποία οι παραμορφώσεις αντανακλούν περισσότερο αρνητικά στην εμπειρία των ακροατών. Ακόμα επιτεύχθηκε η μείωση των στατιστικών μη-συσχετίσεων σε χαμηλούς ρυθμούς δεδομένων βελτιώνοντας την ποιότητα του ήχου. Στα πιο πολλά είδη ηχητικών σημάτων έχουμε πολύ καλή ποιότητα ήχου για συμπίεση σε ρυθμούς δεδομένων 128 kb/s.

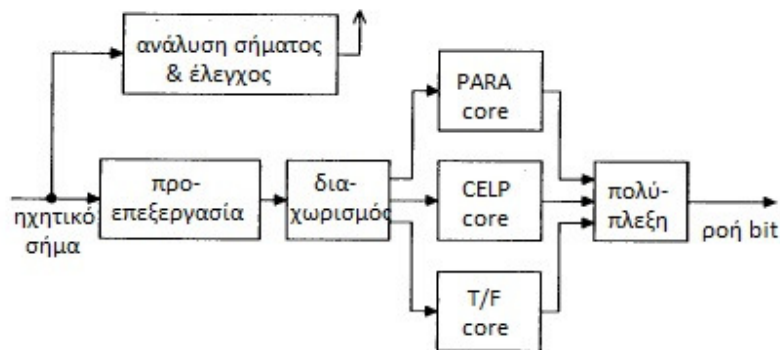
Οι κωδικοποιητές ομιλίας (vocoders), όπως το CELP μοντελοποιούν το φωνητικό σήμα και αποδίδουν καλά ακόμα και σε ρυθμούς δεδομένων 32 kb/s ή και χαμηλότερους. Εάν όμως κάποιο σήμα ομιλίας έχει διαφορά από τα φωνητικά μοντέλα που υποστηρίζονται, τότε το αποτέλεσμα έχει πολλές αδυναμίες. Ειδικά αν χρησιμοποιηθεί κωδικοποιητής ομιλίας σε σήματα που δεν πληρούν τις τυπικές υποθέσεις των μοντέλων φωνής, π.χ. μουσική, τότε η κωδικοποίηση είναι εξαιρετικά ανεπιτυχής. Συνεπώς, κάθε κωδικοποιητής έχει ειδικές απαιτήσεις και είναι σχεδιασμένος για ειδικές κλάσεις ηχητικών σημάτων, οπότε καλό είναι να χρησιμοποιείται μόνο γι' αυτές.

Η δομή ενός κωδικοποιητή MPEG-4 για σήματα ομιλίας φαίνεται στο παρακάτω σχήμα.



Εικόνα 2222 Κωδικοποιητής MPEG-4 CELP για σήματα ομιλίας

Μια άλλη εκδοχή κωδικοποιητή MPEG-4 είναι η παρακάτω.



Εικόνα 2323 Κωδικοποιητής MPEG-4 γενικής μορφής

Οι τρεις κύριοι αλγόριθμοι που εμφανίζονται στα παραπάνω διαγράμματα κωδικοποιητών είναι:

- Η χαρτογράφηση χρόνου/συχνότητας (Time/Frequency, T/F). Αυτή η κωδικοποίηση υπήρχε και στις κωδικοποιήσεις MPEG-1 και MPEG-2 σαν εργαλείο γενικών ηχητικών σημάτων. Η διαφορά στην MPEG-4 είναι η προσθήκη των προαναφερθέντων εργαλείων βελτίωσης αποδοτικότητας σε χαμηλούς ρυθμούς δεδομένων.
- Η CELP κωδικοποίηση. Αφορά αποκλειστικά στην κωδικοποίηση σημάτων φωνής, με κύριο στοιχείο ένα απλό μοντέλο αντίληψης στο οποίο η φασματική περιβάλλουσα του θορύβου κβαντισμού ακολουθεί τη φασματική περιβάλλουσα του σήματος εισόδου.
- Η παραμετρική αναπαράσταση. Αυτή χρησιμοποιείται για την τροποποίηση των ηχητικών σημάτων χρονικά και σε πλάτος και χρησιμοποιείται για εφαρμογές πολύ χαμηλών ρυθμών δεδομένων όπου είναι πιο αποδοτική από τις άλλες δυο.

Η κωδικοποίηση ανάλογα με τα χαρακτηριστικά του σήματος εισόδου αν γίνει χωριστά βελτιώνει την απόδοση. Ο μόνος περιορισμός είναι ο αλγόριθμος που επιλέγεται να είναι ο κατάλληλος για την κλάση του σήματος. Έτσι κάτι που μπορεί να βελτιωθεί στην κωδικοποίηση θα ήταν η δυνατότητα για αυτόματη αναγνώριση του βέλτιστου αλγορίθμου κωδικοποίησης για ξεχωριστή κωδικοποίηση.

Μια σημαντική αλλαγή ανάμεσα στο πρότυπο MPEG-4 και στα προηγούμενα πρότυπα είναι η μη αναπαράσταση του συμπιεσμένου σήματος χωρίς πολύπλεξη και λέξη συγχρονισμού. Στην MPEG-4 το μόνο που χρησιμοποιείται είναι πακέτα πληροφορίας ρυθμίσεων και δεδομένα κάθε κωδικοποιητή. Στην MPEG-4 καθορίζεται ένα «Flex-Mux» σύστημα για τις ανάγκες πολύπλεξης των λειτουργιών MPEG-4 και για την επεκτασιμότητα.

Πρότυπο	Ρυθμός Δειγματοληψίας Ήχου (kHz)	Ρυθμός bits συμπιεσμένου σήματος (kbits/sec)	Κανάλια	Έγκριση Προτύπου
MPEG-1 Layer I	32 , 44.1, 48	32 – 448	1-2 κανάλια	1992
MPEG-1 Layer II	32 , 44.1, 48	32 – 384	1-2 κανάλια	1992
MPEG-1 Layer III	32 , 44.1, 48	32 – 320	1-2 κανάλια	1993
MPEG-2 Layer I	32 , 44.1, 48	32 – 448 για δυο κανάλια	1-5.1 κανάλια	1994
	16, 22.05, 24	32 – 256 για δυο κανάλια		
MPEG-2 Layer II	32 , 44.1, 48	32 – 384 για δυο κανάλια	1-5.1 κανάλια	1994
	16, 22.05, 24	8 – 160 για δυο κανάλια		
MPEG-2 Layer III	32 , 44.1, 48	32 – 384 για δυο κανάλια	1-5.1 κανάλια	1994
	16, 22.05, 24	8 – 160 για δυο κανάλια		
MPEG-2 AAC	8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48, 64, 88.2, 96	Ορίζεται από 23-ψήφιο ακέραιο	1 -48 κανάλια	1997
MPEG-4	8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48, 64, 88.2, 96	Ορίζεται από 23-ψήφιο ακέραιο	1 -48 κανάλια	1999

Πίνακας 8 Κωδικοποιήσεις MPEG και χαρακτηριστικά

Κεφάλαιο 3

Υλοποίηση

3.1 Εισαγωγή

Η αναδρομή στην ιστορική εξέλιξη καθώς και στην αρχιτεκτονική των προτύπων MPEG κατέδειξε την πρόοδο που υπήρξε στην απόδοση και στις δυνατότητες των τεχνικών κωδικοποίησης ηχητικών σημάτων. Κάποια χαρακτηριστικά όμως των κωδικοποιήσεων MPEG ή καλύτερα των υλοποιήσεων των κωδικοποιήσεων MPEG δεν είναι ορατά παρά μόνο με απτά παραδείγματα υλοποιήσεων.

Οι κωδικοποιήσεις MPEG όμως είναι πολύπλοκα συστήματα και δεν είναι απλή η συγγραφή του κώδικα που τα υλοποιεί. Ακόμα και οι υλοποιήσεις σε MATLAB είναι περιορισμένες. Στο κεφάλαιο αυτό θα γίνει η χρήση μιας συνάρτησης του MATLAB για την κωδικοποίηση σύμφωνα με το πρότυπο MPEG-4 ενός τυχαίου συστήματος και αφού εξαχθεί το ηχητικό αποτέλεσμα, θα εξεταστεί η τροποποίηση που προκαλεί η κωδικοποίηση στο σήμα εισόδου. Η συνάρτηση για την παραγωγή του τυχαίου σήματος είναι αυτή που ακολουθείται στην υλοποίηση που χρησιμοποιείται στο επόμενο πρακτικό κομμάτι.

Στη συνέχεια θα χρησιμοποιηθεί μια υλοποίηση σε MATLAB του αλγορίθμου κωδικοποίησης MPEG-1 επίπεδο I με βάση το ψυχοακουστικό μοντέλο I, που αναπτύχθηκε από τον Fabien Petitcolas [48]. Σε αυτή την υλοποίηση παράγονται πολλά γραφήματα, όπως:

- τα τοπικά μέγιστα στο πεδίο των συχνοτήτων,
- τα τονικά στοιχεία στο πεδίο συχνοτήτων,
- τα τονικά και μη τονικά στοιχεία στο πεδίο συχνοτήτων,
- τα μη τονικά στοιχεία και το απόλυτο κατώφλι επικάλυψης,
- τα τονικά στοιχεία και το απόλυτο κατώφλι επικάλυψης,
- τα τονικά στοιχεία με επικάλυψη όσων είναι πολύ κοντά μεταξύ τους και
- το κατώτατο κατώφλι επικάλυψης στο πεδίο συχνότητας.

Η ανάλυση εδώ γίνεται στο πεδίο της συχνότητας για να φανούν τα χαρακτηριστικά της κωδικοποίησης, το οποίο άλλωστε είναι και το σύνηθες πεδίο επεξεργασίας στις τεχνικές ηχητικής κωδικοποίησης.

3.2 Υλοποίηση MPEG-4 AAC με τη συνάρτηση Audiowrite

Η ενσωματωμένη συνάρτηση audiowrite του MATLAB υποστηρίζεται από την έκδοση R2012b και εντεύθεν. Η συνάρτηση audiowrite δέχεται σαν είσοδο έναν πίνακα ηχητικών δεδομένων y , με ρυθμό δειγματοληψίας F_s σε ένα αρχείο filename. Η μορφή του ονόματος του αρχείου καθορίζει και τη μορφή των δεδομένων εξόδου. Ανάλογα με την επέκταση του αρχείου καθορίζεται η μορφή στην οποία θα κωδικοποιηθούν τα δεδομένα ήχου y .

Η πιο απλή σύνταξη της εντολής είναι η εξής: audiowrite(filename, y, F_s).

Σε αυτή τη σύνταξη χρησιμοποιούνται μόνο τα ορίσματα που εξηγήθηκαν παραπάνω, δηλαδή το όνομα του αρχείου (filename), τα δεδομένα ήχου (y) και ο ρυθμός δειγματοληψίας (F_s).

Το όνομα του αρχείου που θα αποθηκευτούν τα κωδικοποιημένα δεδομένα καθορίζεται σαν συμβολοσειρά που περιέχει την επέκταση του αρχείου. Το μονοπάτι του αρχείου μπορεί να είναι απόλυτο ή σχετικό. Εάν δεν καθορίζεται μονοπάτι τότε σαν φάκελος προορισμού ορίζεται το τρέχον μονοπάτι.

Τα είδη αρχείων που υποστηρίζονται είναι τα παρακάτω:

Πλατφόρμες	Μορφή αρχείου
Όλες οι πλατφόρμες	WAVE (.wav)
	OGG (.ogg)
	FLAC (.flac)
Windows και Mac	MPEG-4 AAC (.m4a, .mp4)

Πίνακας 9 Αρχεία που υποστηρίζονται από την audiowrite

Στα Windows όταν αποθηκεύονται AAC αρχεία, το audiowrite συμπληρώνει την αρχή και το τέλος του σήματος εξόδου με επιπλέον δείγματα σιγής, δηλαδή μηδενικής έντασης. Ο κωδικοποιητής AAC των Windows προσθέτει ένα απότομο fade-in και fade-out στον ήχο. Αυτό έχει σαν αποτέλεσμα να αποθηκεύονται περισσότερα δείγματα από τα δεδομένα εισόδου στον υπολογιστή.

Τα δεδομένα εισόδου ορίζονται σαν $M \times N$ πίνακας, όπου M είναι ο αριθμός ηχητικών δειγμάτων και N είναι ο αριθμός των ηχητικών καναλιών. Εάν κάποια διάσταση είναι μοναδιαία η συνάρτηση υποθέτει ότι πρόκειται για τον αριθμό καναλιών και θεωρεί διάσταση δειγμάτων την άλλη. Ο μέγιστος αριθμός καναλιών ανά μορφή αρχείου φαίνεται στον παρακάτω πίνακα.

Μορφή αρχείου	Μέγιστος αριθμός καναλιών
WAVE (.wav)	256
OGG (.ogg)	255
FLAC (.flac)	8
MPEG-4 AAC (.m4a, .mp4)	2

Πίνακας 10 Μέγιστος αριθμός καναλιών ανά κωδικοποίηση

Το πεδίο ορισμού των δεδομένων ανάλογα με το είδος δεδομένων φαίνεται στον παρακάτω πίνακα.

Είδος δεδομένων γ	Πεδίο ορισμού
UInt8	$0 \leq \gamma \leq 255$
Int16	$-32768 \leq \gamma \leq +32768$
Int32	$-2^{32} \leq \gamma \leq 2^{32}-1$
Single	$-1.0 \leq \gamma \leq +1.0$
Double	$-1.0 \leq \gamma \leq +1.0$

Πίνακας 11 Πεδίο ορισμού δεδομένων

Εάν τα δεδομένα ξεπερνούν αυτά τα όρια, ψαλιδίζονται και αναπαρίστανται με τη μέγιστη ή την ελάχιστη αντίστοιχα τιμή που είναι εντός ορίων.

Η συχνότητα δειγματοληψίας σε hertz, των ηχητικών δεδομένων γ , ορίζεται σαν ένας θετικός αριθμός μεγαλύτερος του 0. Οι τιμές του F_s περιορίζονται σε ακέραιες τιμές. Κατά την κωδικοποίηση σε MPEG-4 AAC σε πλατφόρμες Windows, η συνάρτηση audiowrite υποστηρίζει μόνο ρυθμούς παροχής δεδομένων 44100 και 48000, όπως ήταν αναμενόμενο από τη θεωρία.

Εκτός των απαραίτητων ορισμάτων της συνάρτησης υπάρχουν και κάποια προαιρετικά ορίσματα που δεν απαιτούνται, βοηθούν όμως να καθοριστεί καλύτερα το λαμβανόμενο κωδικοποιημένο σήμα εξόδου.

Το όρισμα «Bits Per Sample», με τιμές 8, 16, 24, 32 και 64, καθορίζει τον αριθμό των bits εξόδου ανά δείγμα. Αυτό το όρισμα διατίθεται μόνο για την κωδικοποίηση WAVE (.wav) και FLAC (.flac), για την οποία είναι δυνατές μόνο οι τιμές 8, 16 και 24. Ένα παράδειγμα σύνταξης είναι το 'BitsPerSample', 32. Η προκαθορισμένη τιμή που χρησιμοποιείται αν δεν οριστεί διαφορετικά είναι 16.

Το όρισμα «BitRate», με τιμές 64, 96, 128, 160, 192, 256 και 320, καθορίζει το ρυθμό παροχής δεδομένων ανά δευτερόλεπτο που χρησιμοποιείται για συμπιεσμένα αρχεία. Μη ακέραιες τιμές στρογγυλοποιούνται. Σε πλατφόρμες Windows 7 και μεταγενέστερες οι έγκυρες τιμές είναι 96, 128, 160 και 192. Η επιλογή αυτή είναι διαθέσιμη μόνο για την κωδικοποίηση MPEG-4, με παράδειγμα σύνταξης 'BitRate', 96. Η προκαθορισμένη τιμή είναι 128 αν δεν οριστεί διαφορετικά. Είναι λογικό ότι όσο μεγαλύτερη είναι η τιμή αυτή, τόσο καλύτερη είναι η ποιότητα του παραγόμενου ήχου, αλλά και τόσο μεγαλύτερο το μέγεθος του αντίστοιχου αρχείου.

Το όρισμα «Quality», με τιμές στο πεδίο 0 έως 100, καθορίζει την ποιότητα για την κωδικοποίηση OGG. Η κωδικοποίηση OGG δεν ανήκει στα πρότυπα MPEG (όπως και το Flac), αλλά είναι ένα ανοικτό και ελεύθερο δικαιωμάτων format ηχητικής κωδικοποίησης. Η αναφορά γίνεται για συγκριτικούς λόγους. Όσο μεγαλύτερη είναι η τιμή, τόσο καλύτερη είναι η ποιότητα και μικρότερη η συμπίεση, ενώ αντίστοιχα όσο μικρότερη είναι η τιμή, τόσο χειρότερη είναι η ποιότητα και μεγαλύτερη η συμπίεση. Η σύνταξή της είναι της μορφής 'Quality',25. Η προκαθορισμένη τιμή είναι 75.

Για πληρότητα στα μεταδεδομένα υπάρχουν και τα ορίσματα τίτλου «Title», καλλιτέχνη «Artist» και σχολίων «Comment», που δέχονται όλα δεδομένα συμβολοσειράς.

Ένα παράδειγμα κώδικα με τα προαιρετικά χαρακτηριστικά είναι το παρακάτω.

```
load handel.mat % Φόρτωση αρχείου δεδομένων

filename = 'handel.flac'; % Ονομασία αρχείου και καθορισμός μορφής σαν flac
audiowrite(filename,y,Fs,'BitsPerSample',24,...
'Comment','This is my new audio file.');
```

Ο πίνακας που δείχνει ανάλογα με την κωδικοποίηση τον αριθμό των bits εξόδου ανά δείγμα και τη μορφή δεδομένων εξόδου είναι ο παρακάτω.

Μορφή Αρχείου	Είδος δεδομένων γ	Bits ανά δείγμα εξόδου	Είδος δεδομένων εξόδου
WAVE(.wav)	Uint8, int16, int32, single, double	8	Uint8
		16	Int16
		24	Int32
	Uint8, int16, int32 single, double single, double	32	Int32
		32	Single
		64	Double

FLAC(.flac)	UInt8, int16, int32, single, double	8	Int8
		16	Int16
		24	Int32
MPEG-4(.m4a, .mp4) OGG(.ogg)	UInt8, int16, int32, single, double	N/A	Single

Πίνακας 12 Μορφή και bits εξόδου ανά δείγμα

Οι πληροφορίες για το αρχείο δεδομένων που προκύπτει μετά την κωδικοποίηση μπορούν να αντληθούν από τη συνάρτηση `audioinfo(filename)`. Ένα παράδειγμα εντολής και της αντίστοιχης εξόδου είναι το παρακάτω.

```
info = audioinfo(filename)

info =
  Filename: 'S:\handel.flac'
  CompressionMethod: 'FLAC'
  NumChannels: 1
  SampleRate: 8192
  TotalSamples: 73113
  Duration: 8.9249
  Title: []
  Comment: 'This is my new audio file.'
  Artist: []
  BitsPerSample: 24
```

Για τη σύγκριση του κωδικοποιημένου σήματος με το αρχικό θα χρησιμοποιηθεί η συνάρτηση `audioread`, η οποία διαβάζει τα δεδομένα του κωδικοποιημένου αρχείου και επιστρέφει τα δείγματα όπως αυτά έχουν διαμορφωθεί από την κωδικοποίηση. Το παράδειγμα της σύνταξης που θα χρησιμοποιηθεί είναι `[y, Fs] = audioread(filename)`;

Ακόμα θα χρησιμοποιηθεί η εντολή `whos` που θα μας δώσει τα στοιχεία των πινάκων που περιέχουν τα δείγματα τόσο πριν όσο και μετά την κωδικοποίηση.

Τέλος, η αναπαραγωγή των δυο σημάτων σε ηχητικό αποτέλεσμα μπορεί να γίνει με τη συνάρτηση `sound(x, fs)`, η οποία όμως για το συγκεκριμένο σήμα δεν έχει αισθητή διαφορά.

Ο κώδικας της υλοποίησης είναι ο παρακάτω:

```
% Δημιουργία τυχαίου σήματος
fs = 44100; % Ρύθμιση συχνότητας δειγματοληψίας
f = [1500 1700 1800 2000 10000 11000 12000 13000];
a = [ 1 2 1 .5 2 1.5 3 1];
t = 0:1/fs:1;
x = randn(size(t));
for i = 1:length(f),
    x = x + a(i) * sin(2 * pi * f(i) * t);
end
x = x / max(abs(x));
% Σχεδιασμός τυχαίου σήματος
plot(x);
figure;
% Εμφάνιση στοιχείων δειγμάτων
whos x
% Αναπαραγωγή ηχητικού σήματος
sound(x,fs);
% Καθορισμός ονόματος και είδους αρχείου κωδικοποίησης
filename = 'random_signal.m4a';
% Αποθήκευση κωδικοποιημένου σήματος
audiowrite(filename,x,fs);
% Ανάγνωση κωδικοποιημένου σήματος και εξαγωγή κωδικοποιημένων δειγμάτων
[y,fs] = audioread(filename);
% Πληροφορίες κωδικοποιημένου σήματος
info = audioinfo(filename)
% Εμφάνιση στοιχείων δειγμάτων
whos y
% Αναπαραγωγή ηχητικού σήματος
sound(y,fs);
% Σχεδιασμός σημάτων
subplot(2,1,1), plot(x)
subplot(2,1,2), plot(y)
axis ([0 45060 -1.1 1.1]) % Ορισμός αξόνων
```

Αρχικά, γίνεται η δημιουργία ενός τυχαίου σήματος που για λόγους συνέχειας επιλέχθηκε να είναι αυτό που χρησιμοποιείται στη υλοποίηση του Fabien Petitcolas. Η συχνότητα δειγματοληψίας ορίζεται στα 44100 Hz και δημιουργείται ένα σήμα με εναλλασσόμενα πλάτη που πολλαπλασιάζονται τιμές ημιτόνου που εξαρτώνται από εναλλασσόμενες τιμές συχνότητας και σε καθένα από αυτά προστίθεται ένα τυχαίος ακέραιος. Το αποτέλεσμα όλων των παραπάνω διαιρείται με τη μέγιστη απόλυτη τιμή του διανύσματος, ώστε τα αποτελέσματα να κανονικοποιηθούν στη μονάδα. Έπειτα, γίνεται σχεδιασμός του σήματος για να γίνει αντιληπτή πρακτικά η μορφή του και εμφανίζονται τα στοιχεία των δειγμάτων όπως ο αριθμός του, το είδος τους και το συνολικό μέγεθος που καταλαμβάνουν. Ακόμα γίνεται και αναπαραγωγή του ηχητικού σήματος.

Στην κωδικοποίηση πρώτα καθορίζεται το όνομα του αρχείου, ώστε από την επέκτασή του να καθοριστεί ο τρόπος κωδικοποίησης. Με τη συνάρτηση `audiowrite` και ορίσματα το όνομα αρχείου, τα δείγματα του τυχαίου σήματος και σταθερή τη συχνότητα δειγματοληψίας κωδικοποιείται σύμφωνα με την MPEG-4. Έπειτα, ανακτώνται τα κωδικοποιημένα δείγματα με τη συνάρτηση `audioread` και την ίδια συχνότητα δειγματοληψίας. Αφού γίνει η προβολή των πληροφοριών της κωδικοποίησης, γίνεται προβολή των στοιχείων των δειγμάτων για να εξαχθούν συσχετιστικά συμπεράσματα με το αρχικό σήμα. Έπειτα, γίνεται αναπαραγωγή του κωδικοποιημένου σήματος.

Τέλος, γίνεται αναπαράσταση των δυο σημάτων σε κοινό γράφημα, ώστε να γίνουν παρατηρήσεις πάνω στη σύγκριση των δυο κυματομορφών.

3.3 Συμπεράσματα

Το αρχικό σήμα σύμφωνα με την `whos` παράγει 44101 δείγματα τύπου `double` μεγέθους 352808 bytes.

```
Name Size Bytes Class Attributes
x 1x44101 352808 double
```

Το κωδικοποιημένο σήμα παράγει σύμφωνα με τη `whos` 45055 δείγματα τύπου `double` μεγέθους 360440 bytes.

```
Name Size Bytes Class Attributes
y 45055x1 360440 double
```

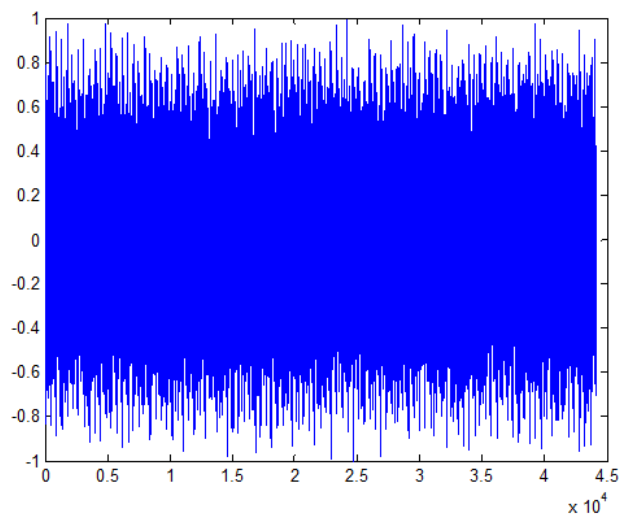
Συνεπώς, το αποτέλεσμα αυτό μας δείχνει ότι κατά την κωδικοποίηση προστίθενται κάποια δείγματα όπως αναμενόταν από τη θεωρία και σαφώς αυξάνεται και ο χώρος που καταλαμβάνεται από αυτά.

Η audioinfo που μας δίνει τις πληροφορίες του κωδικοποιημένου σήματος μας δίνει το ακόλουθο αποτέλεσμα.

```
info =  
  Filename: 'C:\Users\Dimitra\Desktop\random_signal.m4a'  
  CompressionMethod: 'AAC'  
  NumChannels: 1  
  SampleRate: 44100  
  TotalSamples: 45055  
  Duration: 1.0217  
  Title: []  
  Comment: []  
  Artist: []  
  BitRate: 128
```

Έτσι, γίνεται κατανοητό ότι πρόκειται για ένα σήμα AAC, ενός καναλιού, όπως ορίζεται από την ενός καναλιού είσοδο, με ρυθμό δειγματοληψίας 44100 Hz, 45055 δειγμάτων, διάρκειας 1.0217 second και ρυθμού δεδομένων 128 kb/s.

Το γράφημα του τυχαίου σήματος στο πεδίο του χρόνου είναι το παρακάτω:

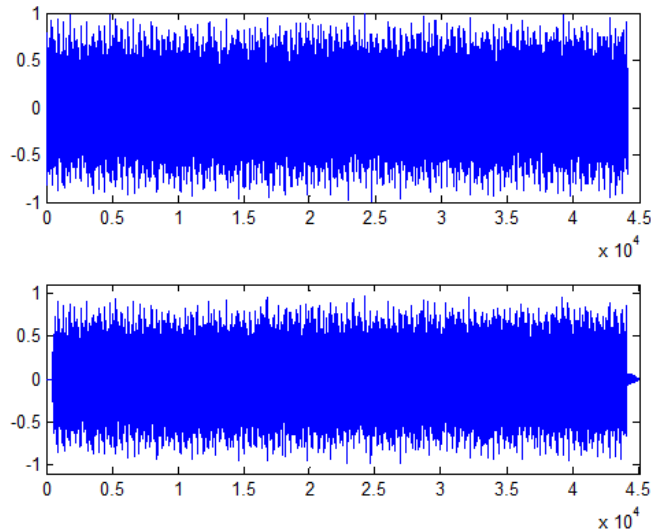


Εικόνα 2424 Γράφημα τυχαίου σήματος

Από το γράφημα είναι προφανές ότι πρόκειται για ένα τελείως τυχαίο σήμα, που μοιάζει με θόρυβο και δεν ακολουθεί κάποια συγκεκριμένη μορφή που θα του έδινε ειδικές ιδιότητες,

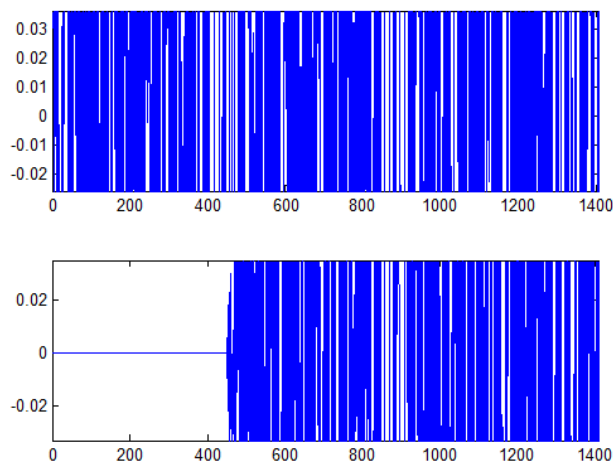
κάνοντας τα συμπεράσματα ειδικά και μη γενικεύσιμα. Το σήμα ξεκινά από το μηδέν και φτάνει ως τα 44101 δείγματα.

Το αρχικό σήμα σε αντιπαράθεση με το κωδικοποιημένο φαίνεται στο παρακάτω γράφημα.



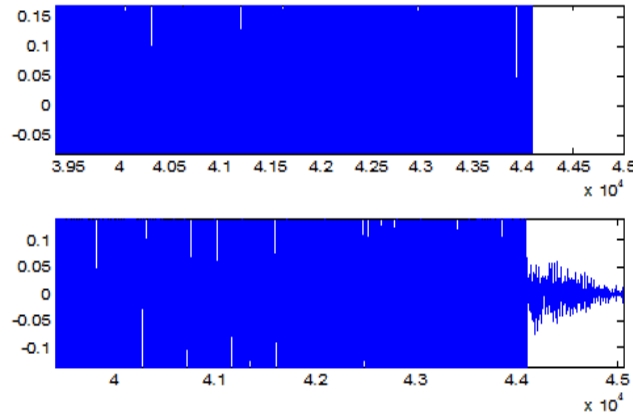
Εικόνα 2525 Το αρχικό σήμα (πάνω) και το κωδικοποιημένο σήμα (κάτω)

Είναι προφανές ακόμα και από τη γενική, συνολική εικόνα ότι τα δυο σήματα δεν είναι ταυτόσημα. Στο κωδικοποιημένο σήμα υπάρχουν επιπλέον δείγματα τα οποία από ότι είδαμε από τις προηγούμενες συναρτήσεις είναι $45055 - 44101 = 954$ δείγματα. Με μια μεγέθυνση του διαγράμματος μπορούμε να δούμε ότι στην έναρξη του σήματος προστίθενται κάποια δείγματα ηρεμίας με μηδενικό πλάτος, που αν μετρηθούν είναι 448. Ακόμα, φαίνεται η σταδιακή αύξηση του πλάτους του κωδικοποιημένου σήματος που αναφέρεται στη θεωρία της συνάρτησης ως fade-in.



Εικόνα 2626 Διαφορά αρχικού σήματος με το κωδικοποιημένο σήμα στα πρώτα δείγματα

Ακόμα, τα δείγματα που προστίθενται στο τέλος του σήματος φαίνονται στην παρακάτω μεγέθυνση, όπου το fade-out είναι εμφανές καθώς στο σημείο στο οποίο σταματούν τα δείγματα του αρχικού σήματος έχουμε δείγματα συνεχώς μειούμενου πλάτους στο κωδικοποιημένο σήμα.



Εικόνα 2727 Διαφορά αρχικού και κωδικοποιημένου σήματος ως προς τα τελευταία δείγματα

Η υλοποίηση αυτή μπορεί να γίνει και για σήματα μουσικής στα οποία αν διαφοροποιηθεί ο ρυθμός παροχής δεδομένων διαφοροποιούνται τα στοιχεία του κωδικοποιημένου σήματος. Η υλοποίηση για σήματα μουσικής που λαμβάνει υπ' όψη της το ρυθμό παροχής δεδομένων, μέσω του ορίσματος BitRate είναι η παρακάτω:

```
%Σήμα Μουσικής
[x,fs] = audioread('S_GRUS.wav');
% Σχεδιασμός σήματος μουσικής
figure;
subplot(2,1,1)
plot(x(:,1))
title('Left channel')
subplot(2,1,2)
plot(x(:,2))
title('Right channel')
% Εμφάνιση στοιχείων δειγμάτων
whos x
% Αναπαραγωγή ηχητικού σήματος
sound(x,fs);
% Αποθήκευση κωδικοποιημένου σήματος
audiowrite('test128.mp4',x,fs,'BitRate',128);
% Ανάγνωση κωδικοποιημένου σήματος και εξαγωγή κωδικοποιημένων δειγμάτων
[x128,fs] = audioread('test128.mp4');
% Πληροφορίες κωδικοποιημένου σήματος
info = audioinfo('test128.mp4')
% Εμφάνιση στοιχείων δειγμάτων
whos x128
% Αναπαραγωγή ηχητικού σήματος
sound(x128,fs);
```

```
% Σχεδιασμός σημάτων
subplot(2,1,1), plot(x)
subplot(2,1,2), plot(x128)
axis ([0 length(x128) -1.1 1.1]) % Ορισμός αξόνων
figure;
% Αποθήκευση κωδικοποιημένου σήματος
audiowrite('test192.mp4',x,fs,'BitRate',192);
% Ανάγνωση κωδικοποιημένου σήματος και εξαγωγή κωδικοποιημένων δειγμάτων
[x192,fs] = audioread('test192.mp4');
% Πληροφορίες κωδικοποιημένου σήματος
info = audioinfo('test192.mp4')
% Εμφάνιση στοιχείων δειγμάτων
whos x192
% Αναπαραγωγή ηχητικού σήματος
sound(x192,fs);
% Σχεδιασμός σημάτων
subplot(2,1,1), plot(x)
subplot(2,1,2), plot(x192)
axis ([0 length(x192) -1.1 1.1]) % Ορισμός αξόνων
figure;
% Αποθήκευση κωδικοποιημένου σήματος
audiowrite('test96.mp4',x,fs,'BitRate',96);
% Ανάγνωση κωδικοποιημένου σήματος και εξαγωγή κωδικοποιημένων δειγμάτων
[x96,fs] = audioread('test96.mp4');
% Πληροφορίες κωδικοποιημένου σήματος
info = audioinfo('test96.mp4')
% Εμφάνιση στοιχείων δειγμάτων
whos x96
% Αναπαραγωγή ηχητικού σήματος
sound(x96,fs);
% Σχεδιασμός σημάτων
subplot(2,1,1), plot(x)
subplot(2,1,2), plot(x96)
axis ([0 length(x96) -1.1 1.1]) % Ορισμός αξόνων
figure;
%Το σήμα διαφοράς (σε ένα κανάλι), το x είναι stereo
d1=x(:,1)-x128(449:500449,1);
d2=x(:,1)-x192(449:500449,1);
d3=x(:,1)-x96(449:500449,1);
%Αναπαραγωγή των σημάτων διαφοράς
sound(d1,fs)
sound(d2,fs)
sound(d3,fs)
%Απεικόνιση των σημάτων διαφοράς
subplot(3,1,1), plot(d2)
subplot(3,1,2), plot(d1,'r')
subplot(3,1,3), plot(d3,'k')
figure;
%Σήμα Ορχηστρικής Μουσικής
[x,fs] = audioread('ORCHESTR.wav');
% Σχεδιασμός ορχηστρικού σήματος
figure;
subplot(2,1,1)
plot(x(:,1))
title('Left channel')
```

```
subplot(2,1,2)
plot(x(:,2))
title('Right channel')
% Εμφάνιση στοιχείων δειγμάτων
whos x
% Αναπαραγωγή ηχητικού σήματος
sound(x,fs);
% Αποθήκευση κωδικοποιημένου σήματος
audiowrite('test128.mp4',x,fs,'BitRate',128);
% Ανάγνωση κωδικοποιημένου σήματος και εξαγωγή κωδικοποιημένων δειγμάτων
[x128,fs] = audioread('test128.mp4');
% Πληροφορίες κωδικοποιημένου σήματος
info = audioinfo('test128.mp4')
% Εμφάνιση στοιχείων δειγμάτων
whos x128
% Αναπαραγωγή ηχητικού σήματος
sound(x128,fs);
% Σχεδιασμός σημάτων
subplot(2,1,1), plot(x)
subplot(2,1,2), plot(x128)
axis ([0 length(x128) -1.1 1.1]) % Ορισμός αξόνων
figure;
% Αποθήκευση κωδικοποιημένου σήματος
audiowrite('test192.mp4',x,fs,'BitRate',192);
% Ανάγνωση κωδικοποιημένου σήματος και εξαγωγή κωδικοποιημένων δειγμάτων
[x192,fs] = audioread('test192.mp4');
% Πληροφορίες κωδικοποιημένου σήματος
info = audioinfo('test192.mp4')
% Εμφάνιση στοιχείων δειγμάτων
whos x192
% Αναπαραγωγή ηχητικού σήματος
sound(x192,fs);
% Σχεδιασμός σημάτων
subplot(2,1,1), plot(x)
subplot(2,1,2), plot(x192)
axis ([0 length(x192) -1.1 1.1]) % Ορισμός αξόνων
figure;
% Αποθήκευση κωδικοποιημένου σήματος
audiowrite('test96.mp4',x,fs,'BitRate',96);
% Ανάγνωση κωδικοποιημένου σήματος και εξαγωγή κωδικοποιημένων δειγμάτων
[x96,fs] = audioread('test96.mp4');
% Πληροφορίες κωδικοποιημένου σήματος
info = audioinfo('test96.mp4')
% Εμφάνιση στοιχείων δειγμάτων
whos x96
% Αναπαραγωγή ηχητικού σήματος
sound(x96,fs);
% Σχεδιασμός σημάτων
subplot(2,1,1), plot(x)
subplot(2,1,2), plot(x96)
axis ([0 length(x96) -1.1 1.1]) % Ορισμός αξόνων
figure;
% Το σήμα διαφοράς (σε ένα κανάλι), το x είναι stereo
d1=x(23950:1298745,1)-x128(22529:1297324,1);
d2=x(23950:1298745,1)-x192(22529:1297324,1);
```

```
d3=x(23950:1298745,1)-x96(22529:1297324,1);
%Αναπαραγωγή των σημάτων διαφοράς
sound(d1,fs)
sound(d2,fs)
sound(d3,fs)
%Απεικόνιση των σημάτων διαφοράς
subplot(3,1,1), plot(d2)
subplot(3,1,2), plot(d1,'r')
subplot(3,1,3), plot(d3,'k')
```

Το αρχικό σήμα μουσικής για ρυθμό παροχής δεδομένων 192 σύμφωνα με την whos παράγει 500001 δείγματα για δύο κανάλια τύπου double μεγέθους 8000016 bytes.

```
Name Size Bytes Class Attributes
x 500001x2 8000016 double
```

Το κωδικοποιημένο σήμα παράγει σύμφωνα με τη whos 500735 δείγματα τύπου double μεγέθους 8011760 bytes.

```
Name Size Bytes Class Attributes
y 500735x2 8011760 double
```

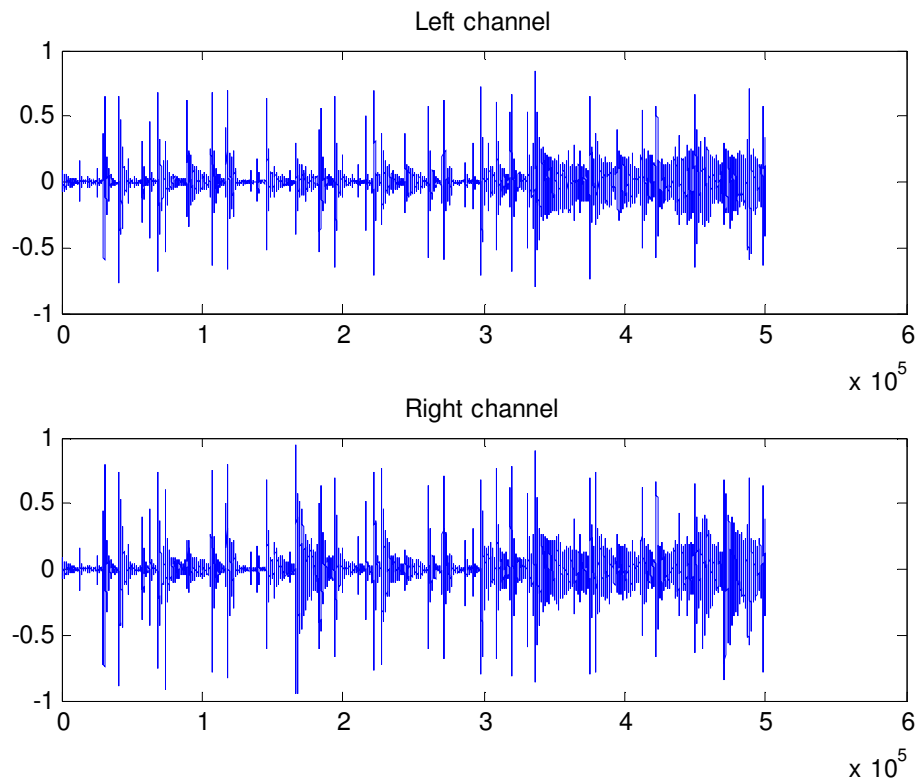
Συνεπώς, το αποτέλεσμα αυτό μας δείχνει και πάλι ότι κατά την κωδικοποίηση προστίθενται κάποια δείγματα όπως αναμενόταν από τη θεωρία.

Η audioinfo που μας δίνει τις πληροφορίες του κωδικοποιημένου σήματος μας δίνει το ακόλουθο αποτέλεσμα.

```
info =
  Filename: 'C:\Users\Dimitra\Desktop\random_signal.m4a'
  CompressionMethod: 'AAC'
  NumChannels: 2
  SampleRate: 44100
  TotalSamples: 500735
  Duration: 11.3545
  Title: []
  Comment: []
  Artist: []
  BitRate: 192
```

Έτσι, γίνεται κατανοητό ότι πρόκειται για ένα σήμα AAC, δυο καναλιών, όπως ορίζεται από την δυο καναλιών είσοδο, με ρυθμό δειγματοληψίας 44100 Hz, 500735 δειγμάτων, διάρκειας 11.3545 second και ρυθμό δεδομένων 192 kb/s.

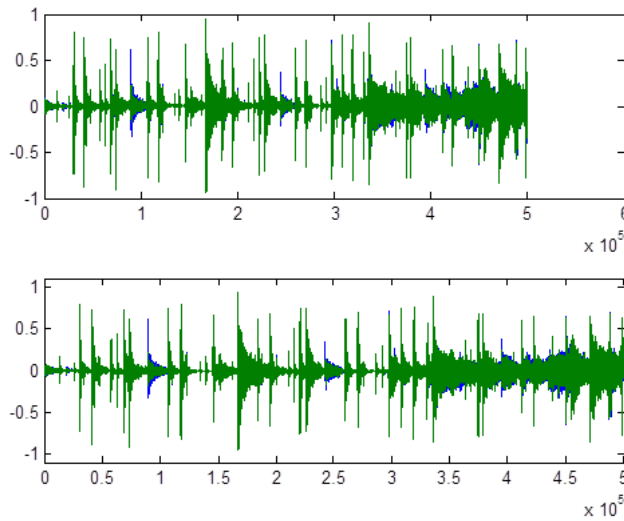
Το γράφημα του μουσικού σήματος είναι το παρακάτω:



Εικόνα 2828 Γράφημα αριστερού και δεξιού καναλιού μουσικού σήματος

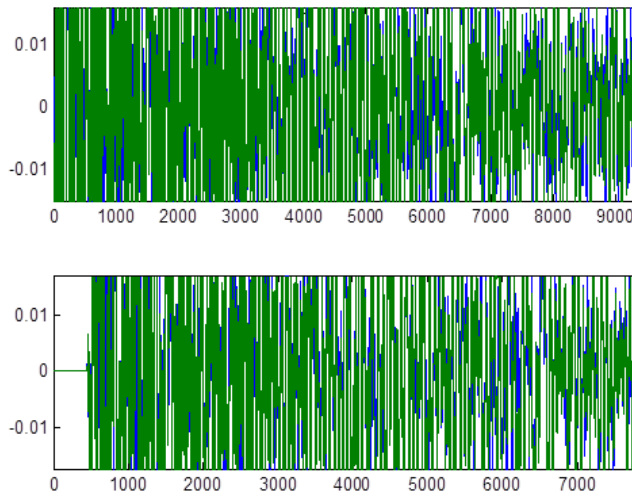
Από το γράφημα είναι προφανές ότι πρόκειται για ένα μουσικό σήμα δυο καναλιών, αριστερού και δεξιού καναλιού το καθένα από τα οποία απεικονίζεται σε ξεχωριστό γράφημα με βάση τα δείγματα που περιλαμβάνει.

Το αρχικό σήμα σε αντιπαράθεση με το κωδικοποιημένο φαίνεται στο παρακάτω γράφημα.



Εικόνα 2929 Το αρχικό σήμα (πάνω) και το κωδικοποιημένο σήμα (κάτω)

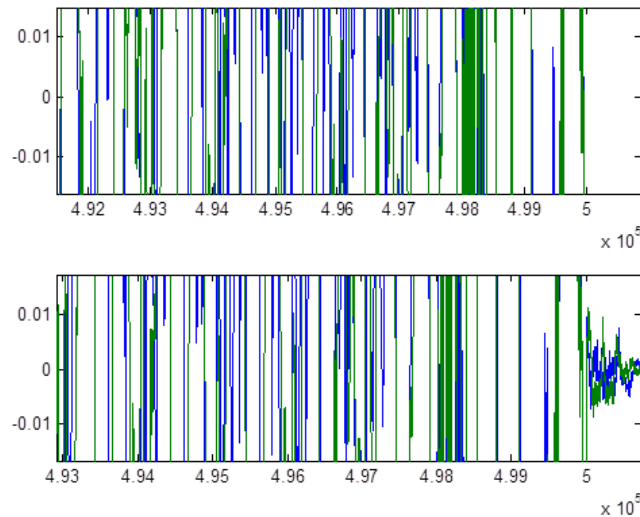
Στο κωδικοποιημένο σήμα υπάρχουν επιπλέον δείγματα τα οποία από ότι είδαμε από τις προηγούμενες συναρτήσεις είναι $500735 - 500001 = 734$ δείγματα. Με μια μεγέθυνση του διαγράμματος μπορούμε να δούμε ότι στην έναρξη του σήματος προστίθενται κάποια δείγματα ηρεμίας με μηδενικό πλάτος, που αν μετρηθούν είναι 448. Ακόμα, φαίνεται και πάλι η σταδιακή αύξηση του πλάτους του κωδικοποιημένου σήματος που αναφέρεται στη θεωρία της συνάρτησης ως fade-in.



Εικόνα 3030 Διαφορά αρχικού σήματος με το κωδικοποιημένο σήμα στα πρώτα δείγματα

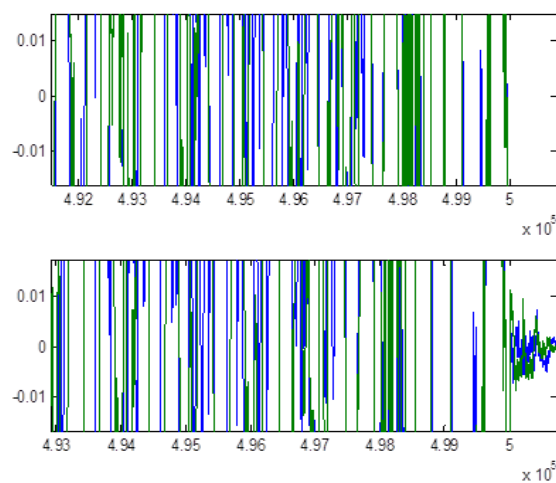
Ακόμα, τα δείγματα που προστίθενται στο τέλος του σήματος φαίνονται στην παρακάτω μεγέθυνση, όπου το fade-out είναι εμφανές καθώς στο σημείο στο οποίο σταματούν τα

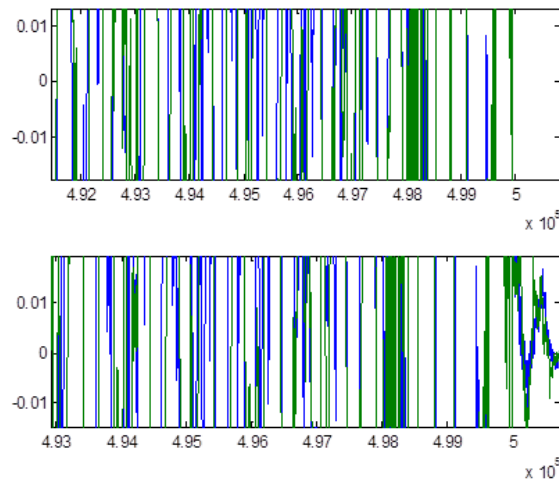
δείγματα του αρχικού σήματος έχουμε δείγματα συνεχώς μειούμενου πλάτους στο κωδικοποιημένο σήμα.



Εικόνα 3131 Διαφορά αρχικού και κωδικοποιημένου σήματος ως προς τα τελευταία δείγματα

Από την παραπάνω ανάλυση είναι προφανές ότι τα ποσοτικά χαρακτηριστικά της κωδικοποίησης δεν εξαρτώνται από το ρυθμό δεδομένων καθώς τα παραγόμενα κωδικοποιημένα σήματα ποσοτικά όσον αφορά στα δείγματα είναι πανομοιότυπα. Η διαφορά είναι ποιοτική. Επειδή δεν είναι δυνατόν να εξεταστούν όλα τα δείγματα ένα προς ένα θα παρατηρήσουμε τη διαφοροποίηση στα τελευταία δείγματα fade-out που χρησιμοποιούνται στις δυο κωδικοποιήσεις. Στο παρακάτω γράφημα φαίνονται και τα δύο κωδικοποιημένα σε σχέση με το αρχικό τους σήμα όσο και το ένα σε σχέση με το άλλο ως προς το fade-out.

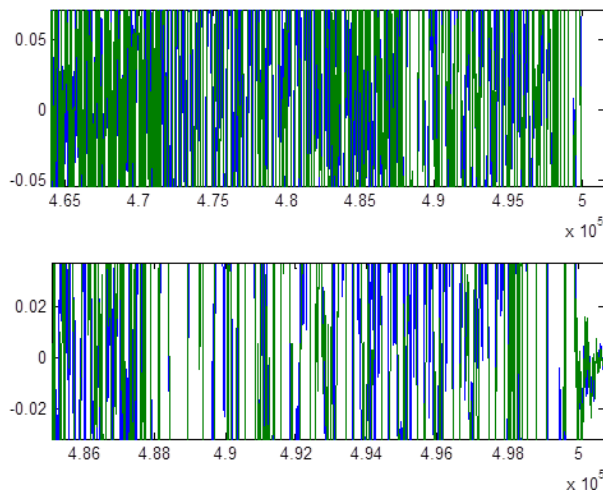




Εικόνα 3232 Σύγκριση των τελευταίων δειγμάτων fade-out των δυο κωδικοποιήσεων (πάνω 192 kbps, κάτω 128 kbps)

Παρατηρώντας τα δυο σχήματα είναι προφανές ότι καλύτερη ποιότητα είναι η κωδικοποίηση με ρυθμό δεδομένων 192 kbps. Αυτό τεκμηριώνεται από το γεγονός ότι το fade-out παρά την ταλάντωση δεν ξεπερνά σε πλάτος το 0,01 για 192 kbps, αλλά για 128 kbps ξεπερνά το 0,01 σημαντικά.

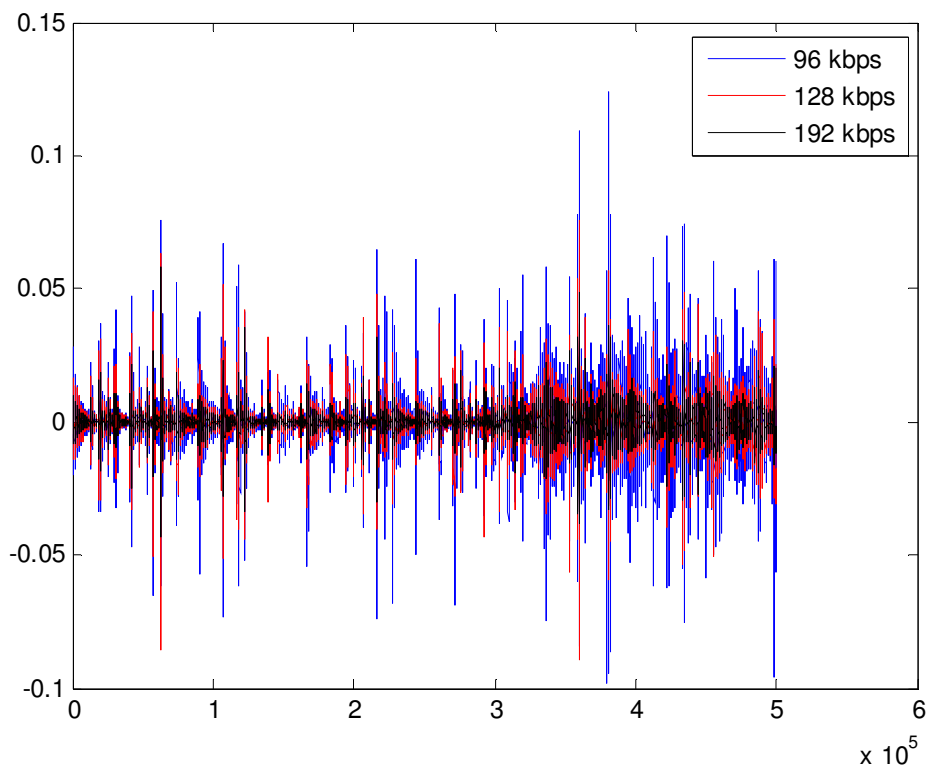
Όταν ο ρυθμός δεδομένων υποδιπλασιάζεται, δηλαδή γίνεται 96kbps η απόκριση στο fade-out χειροτερεύει και η ταλάντωση φτάνει σε πλάτος τα 0,02 που σημαίνει ότι αυξάνεται αναλογικά με τη μείωση του ρυθμού δεδομένων.



Εικόνα 33 Δείγμα fade-out για ρυθμό δεδομένων 96 kbps

Η ηχητική διαφορά ανάμεσα στα τρία κωδικοποιημένα σήματα με τους διαφορετικούς ρυθμούς δεδομένων είναι απειροελάχιστη. Τα κωδικοποιημένα σήματα όταν αναπαράγονται όπως στον κώδικα, με βάση τα κβαντισμένα δείγματα και τη συχνότητα δειγματοληψίας παράγουν ένα ηχητικό αποτέλεσμα πολύ κοντά στο αρχικό σήμα και με μεταξύ τους διαφορές που είναι δύσκολο να γίνουν αντιληπτές από τον κοινό ακροατή και σε χαμηλές σχετικά εντάσεις.

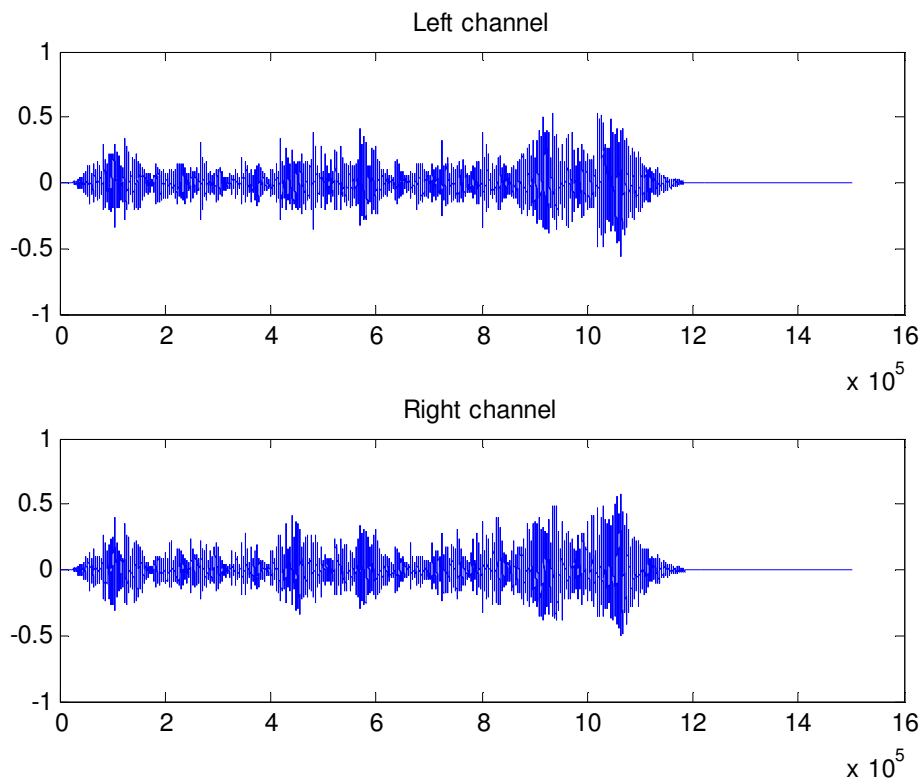
Η διαφορά των δειγμάτων του αρχικού μουσικού σήματος με τα κωδικοποιημένα για τους διαφορετικούς ρυθμούς δεδομένων φαίνονται στο παρακάτω σχήμα, αφού αφαιρεθούν τα αρχικά δείγματα σιωπής που προσθέτει ο αλγόριθμος MPEG. Στο κόκκινο γράφημα φαίνεται η διαφορά του αρχικού σήματος με την κωδικοποίηση MPEG-4 AAC με ρυθμό δεδομένων 128 kbps, που ακολουθεί το γράφημα του αρχικού σήματος και σαν ηχητικό αποτέλεσμα είναι αρκετά κοντά στο αρχικό σήμα. Το ίδιο συμβαίνει και με το μαύρο γράφημα που απεικονίζει τη διαφορά ανάμεσα στο αρχικό σήμα και το τελικό παραγόμενο αποτέλεσμα της MPEG-4 AAC για ρυθμό δεδομένων 192 kbps. Το συμπέρασμα επεκτείνεται και στο μπλε γράφημα για ρυθμό δεδομένων 96 kbps.



Εικόνα 34 Κοινή απεικόνιση της διαφοράς του αρχικού σήματος με τα κωδικοποιημένα σήματα για ρυθμούς δεδομένων 192, 128 και 96 kbps

Η διαφορά όπως είναι εμφανές είναι απειροελάχιστη και δεν είναι άμεσα αντιληπτή από το ανθρώπινο αυτί, συνεπώς όλες αυτές οι κωδικοποιήσεις είναι ικανοποιητικές αν δεν υπάρχει κάποια ιδιαίτερη απαίτηση από την εκάστοτε εφαρμογή. Στην κοινή αυτή απεικόνιση γίνεται σαφές όμως ότι όσο αυξάνει ο ρυθμός δεδομένων τόσο μειώνεται η διαφορά των κωδικοποιημένων σημάτων από το αρχικό. Παρατηρείται ότι η διαφορά είναι μεγαλύτερη για ρυθμό δεδομένων 96 kbps όπως απεικονίζεται στο μπλε γράφημα. Η διαφορά μειώνεται για ρυθμό δεδομένων 128 kbps, στο κόκκινο γράφημα και γίνεται ελάχιστη μεταξύ των τριών για το μεγαλύτερο ρυθμό δεδομένων, δηλαδή το μαύρο γράφημα και 192 kbps.

Το πείραμα επαναλήφθηκε για ένα σήμα ορχηστρικής μουσικής. Τα δυο κανάλια του αρχικού στερεοφωνικού σήματος απεικονίζονται στο παρακάτω γράφημα.



Εικόνα 35 Γράφημα σήματος ορχηστρικής μουσικής

Το αρχικό σήμα μουσικής για ρυθμό παροχής δεδομένων 192 kbps σύμφωνα με την whos παράγει 1500576 δείγματα για δύο κανάλια τύπου double μεγέθους 24009216 bytes.

```
Name Size Bytes Class Attributes
x 1500576x2 24009216 double
```

Το κωδικοποιημένο σήμα παράγει σύμφωνα με τη whos 1501183 δείγματα τύπου double μεγέθους 24018928 bytes.

```
Name Size Bytes Class Attributes
X192 1501183x2 24018928 double
```

Συνεπώς, το αποτέλεσμα αυτό μας δείχνει και πάλι ότι κατά την κωδικοποίηση προστίθενται κάποια δείγματα, συγκεκριμένα προστίθενται 607 δείγματα.

Η audioinfo που μας δίνει τις πληροφορίες του κωδικοποιημένου σήματος μας δίνει το ακόλουθο αποτέλεσμα.

```
info =
```

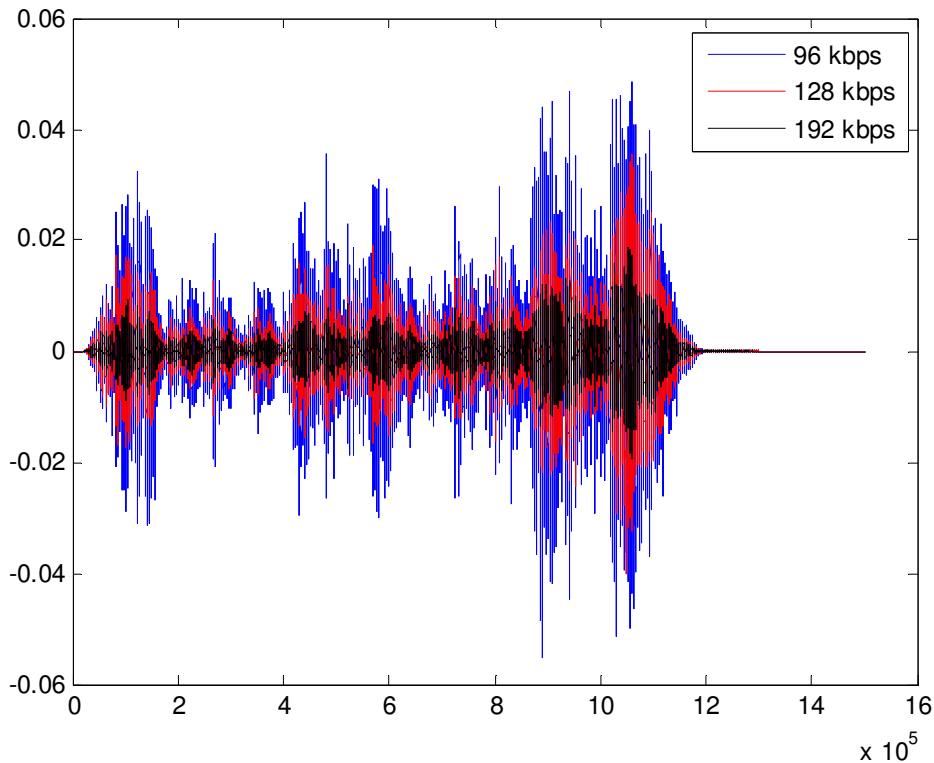
```
Filename: 'C:\Users\Dimitra\Desktop\test192.mp4'
CompressionMethod: 'AAC'
NumChannels: 2
SampleRate: 44100
TotalSamples: 1501183
Duration: 34.0404
Title: []
Comment: []
Artist: []
BitRate: 192
```

Έτσι, γίνεται κατανοητό ότι πρόκειται για ένα σήμα AAC, δυο καναλιών, με ρυθμό δειγματοληψίας 44100 Hz, 1501183 δειγμάτων, διάρκειας 34.0404 second και ρυθμό δεδομένων 192 kb/s.

Τα δείγματα που προστέθηκαν είναι λιγότερα από αυτά που προστέθηκαν στο προηγούμενο σήμα. Αυτό συμβαίνει καθώς το ορχηστρικό σήμα ξεκινά με δείγματα σιωπής, μηδενικού πλάτους τα οποία μετά την κωδικοποίηση MPEG ψαλιδίζονται καθώς το κωδικοποιημένο σήμα ξεκινά με 22529 δείγματα σιωπής, ενώ το αρχικό ξεκινά με 23950 δείγματα σιωπής. Αυτό αποδεικνύει ότι ο αλγόριθμος της κωδικοποίησης MPEG δεν τροποποιεί το αρχικό σήμα δείγμα δείγμα, αλλά επεξεργάζεται μπλοκ δεδομένων, εξάγει τα χαρακτηριστικά τους και τα τροποποιεί. Επίσης τα αρχικά σήματα σιωπής του κωδικοποιημένου σήματος αποτελούνται από τα αρχικά δείγματα που προσθέτει η κωδικοποίηση MPEG και από κάποια τα οποία κωδικοποιήθηκαν και απλά κβαντίστηκαν στη μηδενική στάθμη.

Αφού αφαιρεθούν τα αρχικά δείγματα σιωπής από το αρχικό και τα κωδικοποιημένα παράχθηκε το σήμα διαφοράς του αρχικού σήματος με τα κωδικοποιημένα για διαφορετικούς

ρυθμούς δεδομένων. Αφού αφαιρέθηκαν τα πρώτα 23950 δείγματα του αρχικού σήματος και τα πρώτα 22529 δείγματα από τα κωδικοποιημένα σήματα και λήφθηκαν διανύσματα δειγμάτων ίσου μήκους παράχθηκαν τα παρακάτω σήματα διαφοράς μεταξύ του αρχικού σήματος και του κάθε ενός κωδικοποιημένου χωριστά.



Εικόνα 36 Κοινή απεικόνιση της διαφοράς του αρχικού σήματος με τα κωδικοποιημένα σήματα κατά MPEG-4 AAC για ρυθμούς δεδομένων 192, 128 και 96 kbps

Τα αποτελέσματα και σε αυτό το πείραμα επαληθεύουν τα ευρήματα του προηγούμενου πειράματος. Το σήμα διαφοράς είναι εμφανώς ανάλογο του ρυθμού δεδομένων, ο μικρότερος ρυθμός δεδομένων δίνει σήμα διαφοράς με μεγαλύτερη απόσταση από το αρχικό σε σχέση με τους μεγαλύτερους ρυθμούς δεδομένων. Η ανάλογη σχέση μεταξύ του ρυθμού δεδομένων και της μείωσης της απόστασης από το αρχικό σήμα γίνεται σαφής, καθώς το σήμα διαφοράς της κωδικοποίησης για ρυθμό δεδομένων 96 kbps έχει σταθερά μεγαλύτερες τιμές, από το σήμα διαφοράς της κωδικοποίησης για ρυθμό δεδομένων 128 kbps και το σήμα διαφοράς της κωδικοποίησης για ρυθμό δεδομένων 128 kbps, αποτελείται από εμφανώς μεγαλύτερες τιμές από το σήμα διαφοράς για ρυθμό δεδομένων 192 kbps. Μια ακόμα παρατήρηση είναι ότι το σήμα διαφοράς κρατά τη μορφή και το ηχητικό αποτέλεσμα του αρχικού σήματος που σημαίνει ότι η διαφορά ανάμεσα στο αρχικό και στο κωδικοποιημένο δείγμα εξαρτάται και

από το πλάτος του δείγματος, καθώς όσο μεγαλύτερη κατ' απόλυτη τιμή τόσο πιο πολύ φαίνεται να απέχει από το κωδικοποιημένο αντίστοιχό του.

3.4 Υλοποίηση MPEG-1

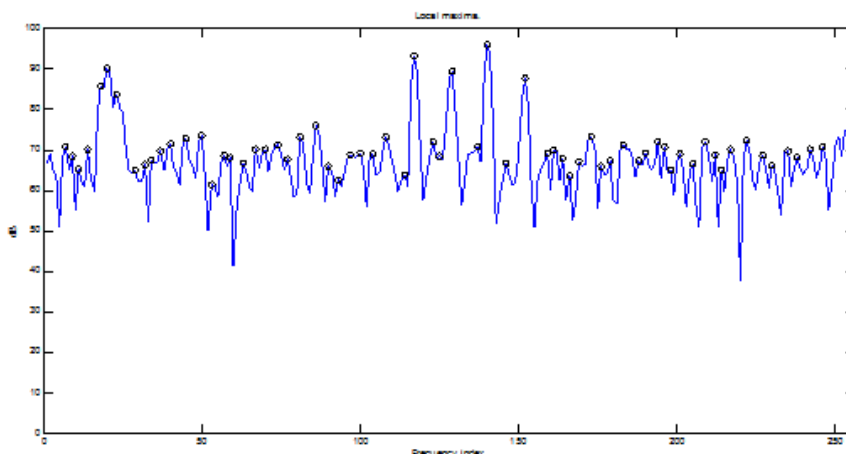
Η υλοποίηση της κωδικοποίησης MPEG-1 με το ψυχοακουστικό μοντέλο 1 ανήκει στον Fabien Petitcolas. Σε αυτή την υλοποίηση παράγονται πολλά γραφήματα στο πεδίο της συχνότητας, ώστε να γίνουν ευκρινή όλα τα χαρακτηριστικά που αναλύονται σε αυτή τη συμπίεση, ώστε να γίνει εκτίμηση των σημάτων και της ποιότητας της συμπίεσης.

Η υλοποίηση αποτελείται από ένα σύνολο συναρτήσεων που υλοποιούν τους υπολογισμούς και παράγουν τα γραφήματα. Κεντρικό αρχείο είναι το Test_MPEG.m, από το οποίο γίνεται η κλήση των συναρτήσεων.

Αρχικά, παράγεται το τυχαίο σήμα με τον ίδιο τρόπο όπως και στην προηγούμενη υλοποίηση. Στη συνέχεια, υπολογίζονται τα κατώφλια κάλυψης, σύμφωνα με τους πίνακες που υπάρχουν στα χαρακτηριστικά του σήματος. Υπολογίζονται οι συντελεστές των φίλτρων υποπεριοχών συχνοτήτων και των παραγόντων κλίμακας.

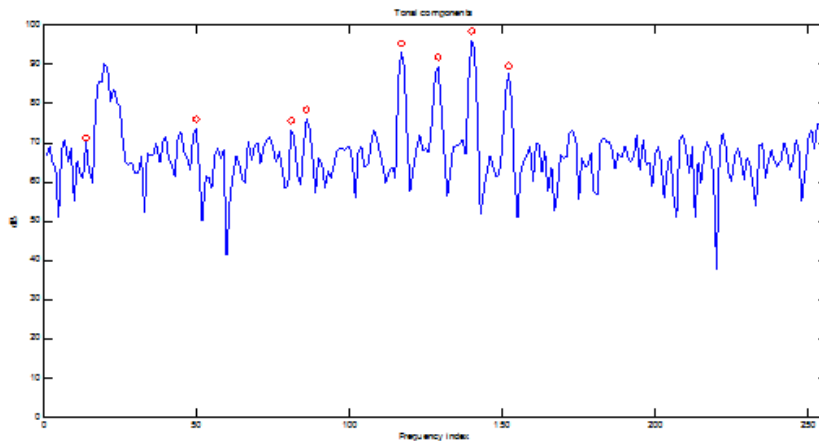
Το επόμενο βήμα είναι η ψυχοακουστική ανάλυση. Αρχικά, υπολογίζεται ο μετασχηματισμός Fast Fourier Transform (FFT) για την ανάλυση του σήματος στο πεδίο της συχνότητας. Στη συνέχεια, μέσω της αντίστοιχης συνάρτησης υπολογίζεται η ηχητική πίεση σε κάθε υποπεριοχή συχνοτήτων. Έπειτα, καλείται η συνάρτηση που εντοπίζει τα τονικά και μη τονικά στοιχεία, ακολουθούμενη από μια συνάρτηση εκκαθάρισης των άσχετων στοιχείων. Ακόμα, καλούνται συναρτήσεις υπολογισμού των ειδικών κατωφλίων και των γενικών κατωφλίων επικάλυψης. Τέλος, υπολογίζεται ο λόγος σήματος προς επικάλυψη.

Τα παραγόμενα γραφήματα εμφανίζονται παρακάτω.



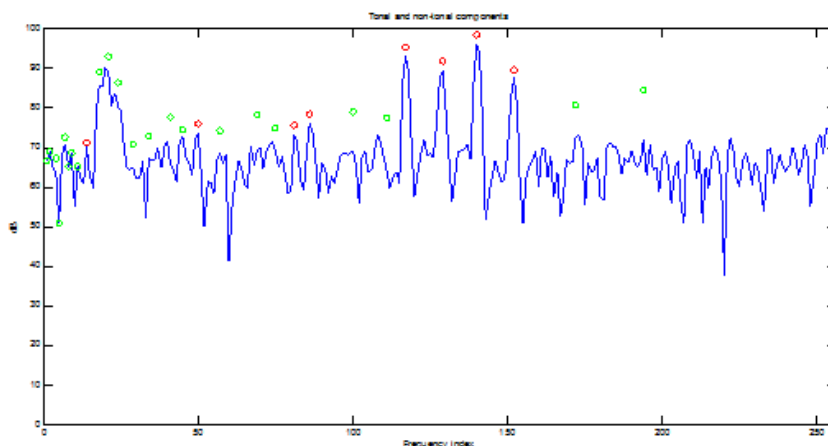
Εικόνα 3337 Τοπικά μέγιστα στο πεδίο της συχνότητας

Στο πρώτο γράφημα φαίνεται το φάσμα του σήματος στη συχνότητα. Σηματοδοτείται με μπλε γραμμή και τα τοπικά μέγιστα σηματοδοτούνται με μαύρους κύκλους. Το φάσμα εκτείνεται από το μηδέν ως τα 250 Hz και είναι φανερό από τα τοπικά μέγιστα ότι το φάσμα είναι ακανόνιστο. Τα τοπικά μέγιστα είναι πολλά και αντιστοίχως τα τοπικά ελάχιστα, ενώ το συχνοτικό περιεχόμενο δε φαίνεται να μειώνεται ή να αυξάνεται με την αύξηση της συχνότητας, κάτι που είναι αναμενόμενο λόγω της τυχαιότητας του σήματος.



Εικόνα 348 Τονικά στοιχεία συναρτήσεως της συχνότητας

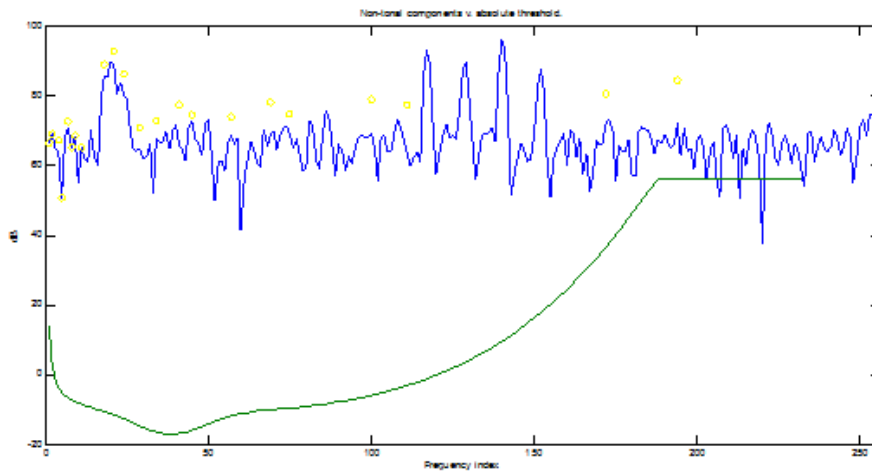
Στο δεύτερο γράφημα, εμφανίζονται τα τονικά στοιχεία, που απεικονίζονται με κόκκινους κύκλους, συναρτήσεως της συχνότητας σε κοινό πλαίσιο με το φάσμα.



Εικόνα 3539 Τονικά και μη τονικά στοιχεία συναρτήσεως της συχνότητας

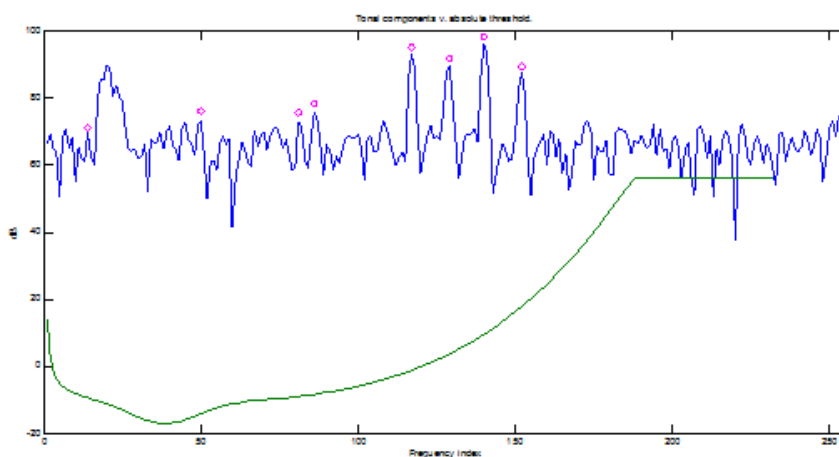
Στο τρίτο γράφημα, προστίθενται τα μη τονικά στοιχεία, τα οποία επισημαίνονται με πράσινους κύκλους, ώστε να φανεί η σχετική τους θέση.

Στο τέταρτο γράφημα, φαίνεται η θέση των μη τονικών στοιχείων, με κίτρινους κύκλους, σε σχέση με τα απόλυτα κατώφλια για κάθε συχνότητα, χαράσσονται με πράσινη γραμμή και παρατηρείται ότι όσο πιο πυκνά είναι τα μη τονικά στοιχεία, τόσο χαμηλότερα είναι τα απόλυτα κατώφλια.



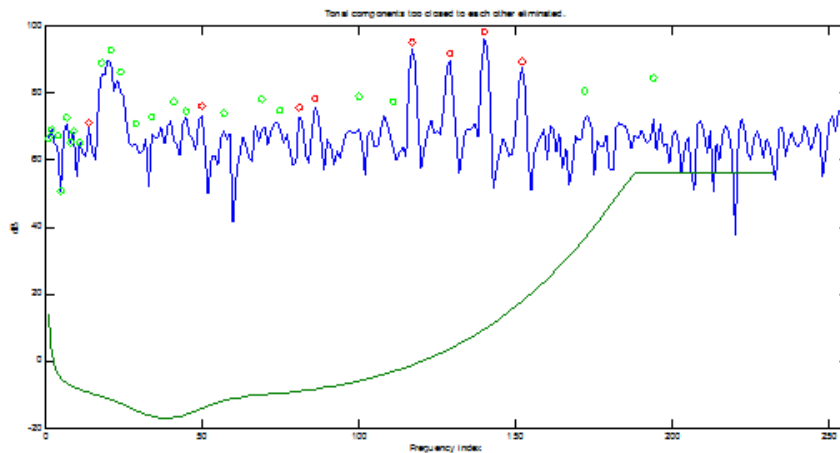
Εικόνα 4036 Μη τονικά στοιχεία σε σχέση με τα απόλυτα κατώφλια

Στο πέμπτο γράφημα, γίνεται παρουσίαση των τονικών στοιχείων, με ροζ κύκλους, σε σχέση με τα απόλυτα κατώφλια, με πράσινη γραμμή.



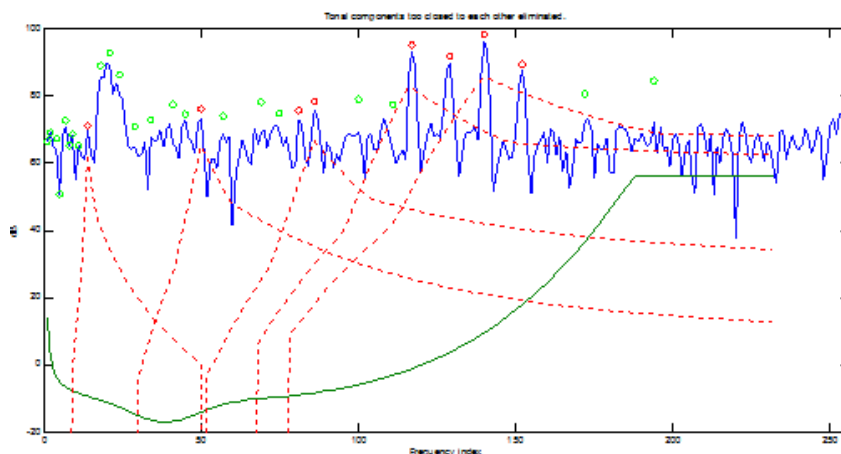
Εικόνα 3741 Τονικά στοιχεία σε σχέση με τα απόλυτα κατώφλια

Στο έκτο γράφημα, φαίνονται τα τονικά στοιχεία (κόκκινοι κύκλοι) και τα μη τονικά στοιχεία (πράσινοι κύκλοι), σε σχέση με τα απόλυτα κατώφλια (πράσινη γραμμή), με την πρόβλεψη εάν κάποια είναι σε πολύ μικρή απόσταση μεταξύ τους να απαλείφονται.



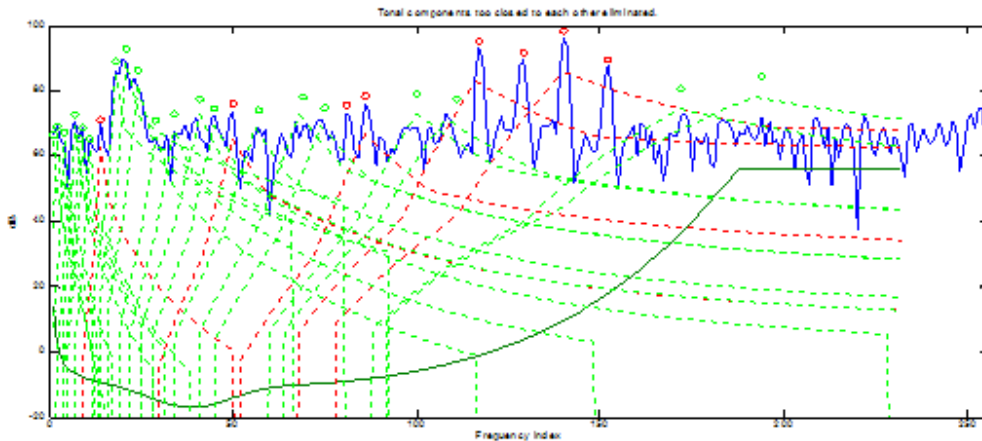
Εικόνα 3842 Τονικά και μη τονικά στοιχεία σε σχέση με τα απόλυτα κατώφλια

Στο έβδομο γράφημα, φαίνονται τα ειδικά κατώφλια επικάλυψης για κάθε τονικό στοιχείο με διακεκομμένη κόκκινη γραμμή που έχει σαν κορυφή το τονικό στοιχείο. Παρατηρείται ότι όταν δύο τονικά στοιχεία είναι κοντά και το κατώφλι επικάλυψης αντιστοιχεί σε υψηλή στάθμη (dB) τότε χαράσσεται κατώφλι επικάλυψης μόνο για το υψηλότερης έντασης τονικό στοιχείο της περιοχής.



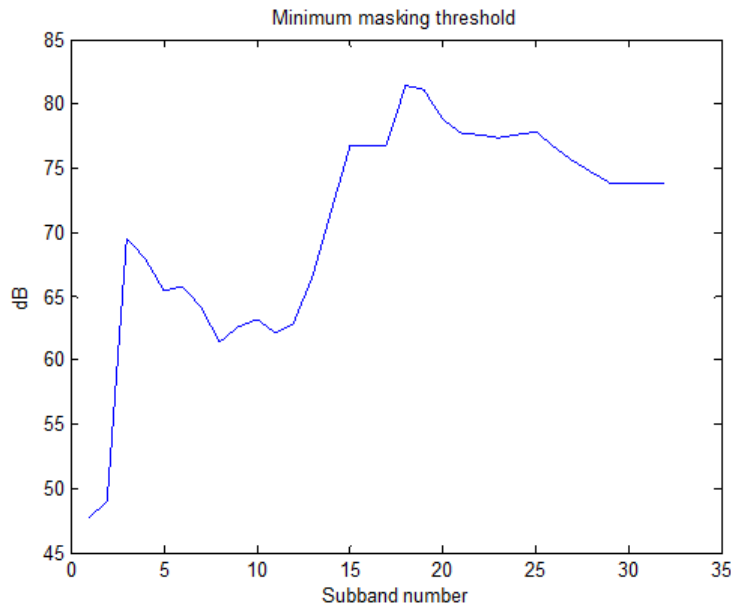
Εικόνα 3943 Κατώφλια επικάλυψης τονικών στοιχείων

Στο όγδοο γράφημα, προστίθενται τα ειδικά κατώφλια επικάλυψης, με πράσινη διακεκομμένη γραμμή, για κάθε μη τονικό στοιχείο. Το συμπέρασμα είναι το ίδιο με τα τονικά στοιχεία και μιας και τα μη τονικά στοιχεία είναι περισσότερα και συνεπώς πιο κοντά το ένα στο άλλο, με αποτέλεσμα να υπάρχει γράφημα κατωφλίου μόνο για τα υψηλότερα μη τονικά στοιχεία κάθε περιοχής.



Εικόνα 404 Κατώφλια επικάλυψης για τονικά και μη τονικά στοιχεία

Το τελικό γράφημα, απεικονίζει το συνολικό ελάχιστο κατώφλι επικάλυψης συναρτήσει των 32 υποπεριοχών συχνότητας, στο οποίο αντιστοιχεί.



Εικόνα 4145 Συνολικό ελάχιστο κατώφλι επικάλυψης

Βιβλιογραφία

- [1] Πρότυπο κωδικοποίησης MPEG mpeg.chiariglione.org (Ανάκτηση: 17 Αυγούστου 2015)
- [2] Πρότυπο κωδικοποίησης MPEG <https://el.wikipedia.org/wiki/MPEG> (Ανάκτηση: 2 Ιουλίου 2015)
- [3] Πρότυπο κωδικοποίησης MPEG en.wikipedia.org/.../Moving_Picture_Experts_Group (Ανάκτηση: 2 Ιουλίου 2015)
- [4] Ηχητικά σήματα, Ν.Σγουρός
<http://users.iit.demokritos.gr/~ntsap/courses/bes04/various/sgouros02.pdf> (Ανάκτηση: 3 Ιουλίου 2015)
- [5] Audio Coding Standards, Chi-Min Liu, Wen-Whei Chang <http://www.mp3-tech.org/programmer/docs/AudioCoding.pdf> (Ανάκτηση: 10 Ιουλίου 2015)
- [6] Audio Signal Processing and Coding, Andreas Spanias, Ted Painter, Venkatraman Atti
- [7] Introduction to Digital Audio Coding and Standards, Marina Bosi, Richard E. Goldberg
- [8] Audio Coding. Theory and Applications, Y. You
- [9] Συμπύεση ηχητικών σημάτων: Από την ADPCM στο MPEG-4, Δρ. Μιχάλης Παρασκευάς, Επικ. Καθηγητής Γιάννης Μουρτζόπουλος
- [10] Ζαρούχας Θωμάς, «Μοντελοποίηση και Επεξεργασία Ηχητικών Δεδομένων για Αναπαραγωγή σε Χώρους με Αντήχηση», Διδακτορική Διατριβή, Πάτρα Ιούνιος 2010.
- [11] Πρότυπο κωδικοποίησης MPEG-1 mpeg.chiariglione.org/standards/mpeg-1 (Ανάκτηση: 17 Σεπτεμβρίου 2015)
- [12] Πρότυπο κωδικοποίησης MPEG-2 mpeg.chiariglione.org/standards/mpeg-2 (Ανάκτηση: 17 Σεπτεμβρίου 2015)
- [13] Πρότυπο κωδικοποίησης MPEG-4 mpeg.chiariglione.org/standards/mpeg-4 (Ανάκτηση: 17 Σεπτεμβρίου 2015)
- [14] Πρότυπο κωδικοποίησης MPEG-H mpeg.chiariglione.org/standards/mpeg-h (Ανάκτηση: 17 Σεπτεμβρίου 2015)
- [15] Η συνάρτηση audiowrite www.mathworks.com/help/matlab/ref/audiowrite.html (Ανάκτηση: 5 Σεπτεμβρίου 2015)
- [16] MPEG for Matlab www.petitcolas.net/fabien/software/mpeg/ (Ανάκτηση: 10 Σεπτεμβρίου 2015)
- [17] Πρότυπο κωδικοποίησης MPEG-1 www.ece.cmu.edu/~ece796/documents/MPEG-1_Audio_CD.doc (Ανάκτηση: 1 Σεπτεμβρίου 2015)
- [18] Universal Multimedia Experiences for tomorrow, Fernando Pereira and Ian Burnett, IEEE SIGNAL PROCESSING MAGAZINE MARCH 2003

- [19] MPEG-D Spatial Audio Object Coding for Dialogue Enhancement (SAOC-DE) , Jouni Paulus, Jurgen Herre, Adrian Murtaza, Leon Terentiv, Harald Fuchs, Sascha Disch, Falko Ridderbusch, AES Convention: 138 (May 2015), Publication Date: May 6, 2015
- [20] A Tutorial on MPEG/Audio Compression, Davis Pan, IEEE Multimedia, Publication Date: 1996
- [21] ISO/MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio, Karlheinz Brandenburg, Gerhard Stoll, JAES Volume 42 Issue 10 pp. 780-792, Publication Date: October 1, 1994
- [22] Overview of MPEG Audio: Current and Future Standards for Low Bit-Rate Audio Coding, Karlheinz Brandenburg, Marina Bosi, JAES Volume 45 Issue ½ pp.4-21, Publication Date: February 1, 1997
- [23] Sound analysis using MPEG compressed audio, Tzanetakis G., Acoustics, Speech and Signal Processing, 2000. ICASSP '00. Proceedings 2000 IEEE International Conference on (Volume:2), Publication Date: June 5, 2000
- [24] Speech recognition on MPEG/Audio encoded files, GL Zick, L Yapp, Multimedia Computing and Systems '97. Proceedings, IEEE International Conference, Publication Date: June 3, 1997
- [25] Subband-domain filtering of MPEG audio signals, C Lanciani, RW Schafer, Acoustics Speech and Signal Processing, 1999. Proceedings, 1999 IEEE International Conference, Publication Date: March 15, 1999
- [26] Theory and Design of Soundfield Reproduction using Continuous Loudspeaker Concept, Yan Jennifer WU, Thushara D. Abhayapala, Department of Information Engineering Research School of Information Sciences & Engineering Australian National University, Audio, Speech, and Language Processing, IEEE Transactions Volume: 17, Issue: 1, Publication Date: December 22, 2008
- [27] "3D sound field recording with high order ambisonics - objective measurements and validation of spherical microphone," S. Moreau, J. Daniel, and S. Bertet, Proc. AES 120th Convention. Audio Engineering Society, Paris, France, May 2006
- [28] "Three-dimensional surround sound systems based on spherical harmonics," M. Poletti, J. Audio Engineering Society, vol. 53, no. 11, pp. 1004–1025, November 2005
- [29] "Comparison of higher-order ambisonics and wave field synthesis with respect to spatial aliasing artifacts," S. Spors and J. Ahrens, Proc. 19th International Congress on Acoustics, Madrid, Spain, Sept 2007
- [30] MPEG-H Audio - The New Standard for Universal Spatial/3D Audio Coding, Herre, Jürgen; Hilpert, Johannes; Kuntz, Achim; Plogsties, Jan, JAES Volume 62 Issue 12 pp. 821-830, Publication Date: January 5, 2015
- [31] ISO/IEC MPEG-2 Advanced Audio Coding, JAES Volume 45 Issue 10 pp. 789-814, Publication Date: October 1, 1997

- [32] MPEG Spatial Audio Object Coding - The ISO/MPEG Standard for Efficient Coding of Interactive Audio Scenes, Herre, Jürgen; Purnhagen, Heiko; Koppens, Jeroen; Hellmuth, Oliver; Engdegård, Jonas; Hilper, Johannes; Villemoes, Lars; Terentiv, Leon; Falch, Cornelia; Hölzer, Andreas; Valero, María Luis; Resch, Barbara; Mundt, Harald; Oh, Hyen-O, JAES Volume 60 Issue 9 pp. 655-673, Publication Date: October 9, 2012
- [33] ISO/IEC 11172-3:1993, Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio, with the permission of ISO
- [34] Πρότυπο κωδικοποίησης MPEG-1 Επιπέδου II https://en.wikipedia.org/wiki/MPEG-1_Audio_Layer_II (Ανάκτηση: 17 Σεπτεμβρίου 2015)
- [35] Κωδικοποίηση ASPEC <http://www.aes.org/e-lib/browse.cfm?elib=5397> (Ανάκτηση: 17 Σεπτεμβρίου 2015)
- [36] H. S. Malvar, "Modulated QMF Filter Banks with Perfect Reconstruction", Electronics Letters , vol. 26, no. 13, pp.906–907 (Equation 13), June 1990
- [37] H. S. Malvar, "Lapped Transforms for Efficient Transform/Subband Coding", IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 38, no. 6, pp. 969-978 (Equation 22), June 1990
- [38] J. P. Princen, A. W. Johnson und A. B. Bradley: Subband/transform coding using filter bank designs based on time domain aliasing cancellation, IEEE Proc. Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2161–2164, 1987
- [39] Διαφορά Επιπέδου Απόκρυψης <http://www.ee.columbia.edu/~dpwe/classes/e6820-2001-01/matlab/MAD/bmld/bmld.htm> (Ανάκτηση: 2 Ιουλίου 2015)
- [40] B.C.J. Moore (1997) An introduction to the psychology of hearing (fourth edition). Academic Press
- [41] Cocktail Party Effect https://en.wikipedia.org/wiki/Cocktail_party_effect (Ανάκτηση: 3 Ιουλίου 2015)
- [42] Critical Band <http://www.psychologyofmusic.co.uk/criticalband.pdf> (Ανάκτηση: 3 Ιουλίου 2015)
- [43] Joint (audio engineering) [https://en.wikipedia.org/wiki/Joint_\(audio_engineering\)](https://en.wikipedia.org/wiki/Joint_(audio_engineering)) (Ανάκτηση: 4 Ιουλίου 2015)
- [44] D. Huffman, A Method for the Construction of Minimum-Redundancy Codes
- [45] B.S. Atal, "The History of Linear Prediction," IEEE Signal Processing Magazine, vol. 23, no. 2, March 2006, pp. 154–161
- [46] Πρότυπο κωδικοποίησης MPEG-4 <http://mpeg.chiariglione.org/standards/mpeg-4/irmp-extensions> (Ανάκτηση: 2 Ιουλίου 2015)

- [47] Jörn Nettingsmeier and David Dohrmann, Preliminary studies on large-scale higher-order Ambisonic sound reinforcement systems, Ambisonics Symposium 2011, Lexington (KY) 2011
- [48] MPEG for Matlab <http://www.petitcolas.net/fabien/software/mpeg/> (Ανάκτηση: 18 Σεπτεμβρίου 2015)
- [49] A Tutorial on MPEG/Audio Compression
https://www.google.gr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=8&cad=rja&uact=8&ved=0ahUKEwj64HglLTMAhWmC8AKHZXpCzMQFghcMAc&url=http%3A%2F%2Fwww.cs.ubc.ca%2F~randeep%2Fdocs%2Fpan95_gakhal.pps&usg=AFQjCNGmkpiECinnER6tZ-PizCDTZjpoAQ (Ανάκτηση: 18 Σεπτεμβρίου 2015)
- [50] M. Bosi, R. Goldberg, Introduction to Digital Audio Coding and Standards, Kluwer Academic Publishers, 2003
- [51] Audio Engineering Society CD-ROM, "Perceptual Audio Coders: What to Listen For", AES 2001
- [52] Brandenburg K.H., "Evaluation of Quality for Audio Encoding at low Bit Rates," presented at the 82nd Convention of the AES, London 1987
- [53] Brandenburg K.H., Herre J., Johnston J.D., Mahieux Y. and Schroeder E., "ASPEC: Adaptive Spectral Entropy Coding of High Quality Music Signals," 90th Convention of the AES, Paris, 1991.
- [54] Brandenburg K.H. and Stoll G., "The ISO/MPEG Audio Codec: A Generic Standard for Coding of High Quality Digital Audio," presented at the 92nd Convention of the AES, Vienna, 1992.
- [55] Brandenburg K.H., Eberlein E., Herre J. and Edler B., "Comparison of Filterbanks for High Quality Audio Coding," in Proc. of ICASSP '92, pp. 1336-1339, 1992.
- [56] ISO/IEC MPEG Audio Coder 11172-3: 1993(E), CH1211 Geneve 20, Switzerland.
- [57] Jayant N.S. and Noll P., "Digital Coding of Waveforms: Principles and Applications to Speech and Video," Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1984.
- [58] Jayant N.S., "High Quality Coding of Telephone Speech and Wideband Audio," IEEE Communications Magazine, pp. 10-20, Jan. 1990.
- [59] Jayant N.S., "Signal Compression: Technology Targets and Research Directions," IEEE Journal on Selected Areas in Commun., Vol. 10, No. 5, pp. 796-818, Jun. 1992.
- [60] Johnston J.D., "Estimation of Perceptual Entropy Using Noise Masking Criteria," in Proc. of ICASSP '88, pp. 2524-2527, 1988.
- [61] Johnston J.D., "Transform Coding of Audio Signals Using Perceptual Noise Criteria," IEEE Trans. on Selected Areas in Commun., Vol. 6, No. 2, Feb. 1988.

- [62] Malvar H.S., "Lapped Transforms for Efficient Transform/Subband Coding," IEEE Trans. on Acoust., Speech and Signal Processing, Vol. ASSP-38, No. 6, pp. 969-978, Jun. 1990.
- [63] M.Paraskevas and J.Mourjopoulos, «A Differential Perceptual Audio Coding Method with Reduced Bitrate Requirements,» IEEE Transactions on Speech and Audio Processing, Vol. 3, No. 6, pp. 490-503, Nov. 1995.
- [64] Princen J.P and Bradley A.B, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation," IEEE Trans. on Acoust., Speech and Signal Processing, Vol. ASSP-34, No. 5, pp. 1153-1161, Oct. 1986.
- [65] Princen J.P, Johnson A.W. and Bradley A.B, "Subband/Transform Coding using Filter Bank Designs Based on Time Domain Aliasing Cancellation," in Proc. of ICASSP '87, pp. 2161-2164, Dallas 1987.
- [66] Rothweiler J.H., "Polyphase Quadrature Filters - A New Subband Coding Technique," in Proc. of ICASSP '83, pp. 1280-1284, 1983.
- [67] Schroeder M., "Model of Hearing," Proceedings of IEEE, Vol 63(9), pp. 1332-1350, Sep. 1975.
- [68] Schroeder M.R., Atal B.S. and Hall J.L., "Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear," J. Acoust. Soc. of Am., Vol. 66, No. 6, pp. 1647-1651, Dec. 1979.
- [69] Stoll G. and Dehery Y.F., "MUSICAM: High Bit-Rate Reduction System Family for Different Applications," CCIR Final Study Group Meetings, Geneva, Oct. 1989.
- [70] Theile G., Stoll G. and Link M., "Low Bit-Rate Coding of High Quality Audio Signals. An Introduction to the MASCAM System," EBU Review - Technical No. 230, pp. 158-181, Aug. 1988.
- [71] Tsutsui K., Suzuki H., Shimoyoshi O., Sonohara M., Akagiri K. and Heddle R., "ATRAC: Adaptive Transform Acoustic Coding for MiniDisc," presented at the 93rd AES Convention, San Francisco, 1992.
- [72] Turgeon A., Soumagne J., Mabillean P., Morissete S. and Pailard B., "A Study of Strategies for the Perceptual Coding of Audio Signals," presented at the 90th AES Convention, Paris 1991.
- [73] Vaidyanathan P.P., "Quadrature Mirror Filter Banks, M-Band Extensions and Perfect Reconstruction Techniques," IEEE ASSP Magazine, pp. 4-20, Jul.1987.
- [74] Zwicker E. and Fastl H., "Psychoacoustics: Facts and Models," Springer - Verlag, Berlin 1990.
- [75] Zwicker E. and Zwicker U.T., "Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System," J. of AES, Vol. 39, No. 3, pp. 115-126, Mar. 1993.