



**ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**  
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**  
Μεταπτυχιακό Πρόγραμμα Σπουδών

**Γιαννακούρος Αλέξανδρος (Α.Μ. 97506)**  
**MPEG Ψηφιακή Κωδικοποίηση Ήχου**

Εργασία στο μάθημα : Επικοινωνία με Ομιλία  
Διδάσκων : Γεώργιος Κουρουπέτρογλου

Αθήνα 1998

# ΠΕΡΙΕΧΟΜΕΝΑ

## ΕΙΣΑΓΩΓΗ

### 1. ΓΕΝΙΚΑ ΠΕΡΙ ΡΥΘΜΩΝ ΚΩΔΙΚΟΠΟΙΗΣΗΣ

- 1.1. PCM ΡΥΘΜΟΙ bit
- 1.2. ΜΕΙΩΣΗ ΡΥΘΜΟΥ bit
- 1.3. ΕΝΕΡΓΕΙΕΣ MPEG ΚΑΘΙΕΡΩΣΗΣ ΠΡΟΤΥΠΩΝ

### 2. ΤΕΧΝΟΛΟΓΙΕΣ ΚΛΕΙΔΙ ΣΤΗΝ ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ

- 2.1 ΕΙΣΑΓΩΓΙΚΑ
- 2.2 ΑΚΟΥΣΤΙΚΗ ΑΠΟΚΡΥΨΗ (AUDITORY MASKING) ΚΑΙ ΚΩΔΙΚΟΠΟΙΗΣΗ ΑΝΤΙΛΗΨΗΣ (PERCEPTUAL CODING)
- 2.3 ΚΩΔΙΚΟΠΟΙΗΣΗ ΣΤΟ ΧΩΡΟ ΤΩΝ ΣΥΧΝΟΤΗΤΩΝ
- 2.4 ΕΝΑΛΛΑΓΗ/ΜΕΤΑΠΗΔΗΣΗ ΠΑΡΑΘΥΡΟΥ
- 2.5 ΔΥΝΑΜΙΚΟΣ ΚΑΤΑΜΕΡΙΣΜΟΣ bit

### 3. ISO/MPEG-1 ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ

- 3.1 ΕΙΣΑΓΩΓΙΚΑ
- 3.2 ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ
  - ΔΟΜΗ
  - ΣΤΡΩΜΑΤΑ ΚΑΙ ΤΡΟΠΟΙ ΛΕΙΤΟΥΡΓΙΑΣ
  - ΚΩΔΙΚΟΠΟΙΗΣΗ ΣΤΕΡΕΟΦΩΝΙΚΟΥ ΠΛΕΟΝΑΣΜΟΥ
  - ΨΥΧΟΑΚΟΥΣΤΙΚΑ ΜΟΝΤΕΛΑ
- 3.3 ΣΤΡΩΜΑΤΑ I ΚΑΙ II
  - ΤΡΑΠΕΖΑ ΦΙΛΤΡΩΝ
  - ΚΒΑΝΤΙΣΗ
  - ΠΑΡΑΔΕΙΓΜΑΤΑ ΚΩΔΙΚΟΠΟΙΗΣΗΣ
  - ΑΠΟΚΩΔΙΚΟΠΟΙΗΣΗ
- 3.4 ΣΤΡΩΜΑ III
  - ΕΝΑΛΛΑΣΣΟΜΕΝΗ ΤΡΑΠΕΖΑ ΦΙΛΤΡΩΝ
  - ΚΒΑΝΤΙΣΗ ΚΑΙ ΚΩΔΙΚΟΠΟΙΗΣΗ
- 3.5 ΔΟΜΗ ΠΛΑΙΣΙΟΥ ΚΑΙ ΠΟΛΥΠΛΟΚΗ ΔΟΜΗ
  - ΔΟΜΗ ΠΛΑΙΣΙΟΥ
  - ΠΟΛΥΠΛΟΚΗ ΔΟΜΗ
- 3.6 ΥΠΟΚΕΙΜΕΝΙΚΗ ΠΟΙΟΤΗΤΑ (MPEG-1; ΣΤΕΡΕΟΦΩΝΙΚΑ ΗΧΗΤΙΚΑ ΣΗΜΑΤΑ)

### 4. MPEG ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ ΜΕ ΧΑΜΗΛΟΤΕΡΕΣ ΣΥΧΝΟΤΗΤΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

### 5. MPEG ΚΩΔΙΚΟΠΟΙΗΣΗ ΠΟΛΥΚΑΝΑΛΙΚΟΥ ΗΧΟΥ

- 5.1 ΑΝΑΠΑΡΑΣΤΑΣΕΙΣ ΠΟΛΥΚΑΝΑΛΙΚΟΥ ΗΧΟΥ
- 5.2 MPEG-2/ΗΧΟΥ ΠΟΛΥΚΑΝΑΛΙΚΗ ΚΩΔΙΚΟΠΟΙΗΣΗ
  - ΑΝΑΔΡΟΜΙΚΑ ΣΥΜΒΑΤΗ MPEG-2 ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ
  - MPEG-2 ΠΡΟΧΩΡΗΜΕΝΗ ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ
- 5.3 ΑΝΑΔΡΟΜΙΚΗ ΣΥΜΒΑΤΟΤΗΤΑ ΜΕΣΩ ΤΑΥΤΟΧΡΟΝΗ ΕΚΠΙΟΜΠΗΣ
- 5.4 ΥΠΟΚΕΙΜΕΝΙΚΕΣ ΔΟΚΙΜΕΣ (MPEG-2; ΠΟΛΥΚΑΝΑΛΙΚΑ ΣΗΜΑΤΑ ΗΧΟΥ)

## **6. MPEG-4 ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ**

### 6.1 ΕΙΣΑΓΩΓΙΚΑ

#### 6.2 ΣΚΟΠΟΣ ΚΑΙ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥ MPEG-4 ΠΡΟΤΥΠΟΥ

6.2.1 ΚΩΔΙΚΟΠΟΙΗΜΕΝΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΑΝΤΙΚΕΙΜΕΝΩΝ ΜΕΣΟΥ

6.2.2 ΣΥΝΘΕΣΗ ΑΝΤΙΚΕΙΜΕΝΩΝ ΜΕΣΟΥ

6.2.3 ΠΕΡΙΓΡΑΦΗ ΚΑΙ ΣΥΓΧΡΟΝΙΣΜΟΣ ΤΩΝ ΚΑΤΕΥΘΥΝΟΜΕΝΩΝ

ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΑΝΤΙΚΕΙΜΕΝΑ ΜΕΣΟΥ

6.2.4 ΠΑΡΑΔΟΣΗ ΚΑΤΕΥΘΥΝΟΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

6.2.5 ΑΛΛΗΛΕΠΙΔΡΑΣΗ ΜΕ ΑΝΤΙΚΕΙΜΕΝΑ ΜΕΣΟΥ

6.2.6 ΔΙΑΧΕΙΡΙΣΗ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗ ΤΗΣ ΠΝΕΥΜΑΤΙΚΗΣ  
ΙΔΙΟΚΤΗΣΙΑΣ

#### 6.3 ΑΝΑΛΥΤΙΚΗ ΤΕΧΝΙΚΗ ΠΕΡΙΓΡΑΦΗ ΤΟΥ MPEG-4 ΠΡΟΤΥΠΟΥ

##### 6.3.1 DMIF

- ΤΟ DMIF ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ

##### 6.3.2 ΑΠΟΠΟΛΥΠΛΕΞΙΑ, ΣΥΓΧΡΟΝΙΣΜΟΣ ΚΑΙ ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΚΑΤΕΥΘΥΝΟΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

- ΑΠΟΠΟΛΥΠΛΕΞΙΑ
- ΣΥΓΧΡΟΝΙΣΜΟΣ ΚΑΙ ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΒΑΣΙΚΩΝ ΡΕΥΜΑΤΩΝ
- ΔΙΑΧΕΙΡΙΣΗ ΠΡΟΣΩΡΙΝΗΣ ΜΝΗΜΗΣ
- ΑΝΑΓΝΩΡΙΣΗ ΧΡΟΝΟΥ

##### 6.3.3 ΠΕΡΙΓΡΑΦΗ ΣΥΝΤΑΞΗΣ

##### 6.3.4 ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΗΤΙΚΩΝ ΑΝΤΙΚΕΙΜΕΝΩΝ

- ΦΥΣΙΚΟΣ ΗΧΟΣ
- ΣΥΝΘΕΤΕΜΕΝΟΣ ΗΧΟΣ

##### 6.3.5 ΑΝΑΓΝΩΡΙΣΗ ΚΑΙ ΠΡΟΣΤΑΣΙΑ ΤΟΥ IPR ΣΥΣΧΕΤΙΣΜΕΝΟΥ ΜΕ ΤΟ ΠΕΡΙΕΧΟΜΕΝΟ

#### 6.4 ΛΙΣΤΑ ΤΩΝ ΚΥΡΙΩΝ ΛΕΙΤΟΥΡΓΙΚΟΤΗΤΩΝ ΠΟΥ ΠΑΡΕΧΟΝΤΑΙ ΑΠΟ ΤΟ MPEG-4 ΣΤΗΝ ΕΚΔΟΣΗ 1 ΓΙΑ ΤΟΝ ΗΧΟ

#### 6.5 ΔΟΚΙΜΕΣ ΕΠΑΛΗΘΕΥΣΗΣ : ΕΛΕΓΧΟΝΤΑΣ ΤΗΝ ΑΠΟΔΟΣΗ ΤΟΥ MPEG

#### 6.6 ΗΧΗΤΙΚΑ ΠΡΟΦΙΛ ΣΤΟ MPEG-4 ΕΚΔΟΣΗ 1

#### 6.7 ΕΚΔΟΣΗ 2 ΤΟΥ MPEG-4/ΗΧΟΥ : ΕΤΟΙΜΟ ΕΝΑ ΧΡΟΝΟ ΜΕΤΑ ΤΗΝ ΕΚΔΟΣΗ 1

## **7. ΕΦΑΡΜΟΓΕΣ**

## **8. ΣΥΜΠΕΡΑΣΜΑΤΑ**

## **9. ΛΕΞΙΚΟ**

## **10. ΣΥΝΤΜΗΣΕΙΣ**

## **11. ΑΝΑΦΟΡΕΣ**

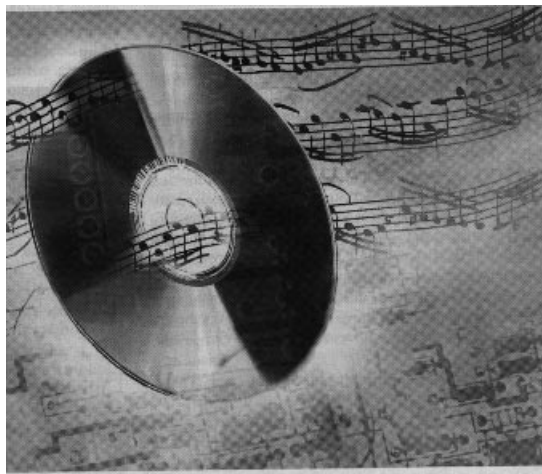
## **12 ΒΙΒΛΙΟΓΡΑΦΙΑ**

# MPEG Ψηφιακή Κωδικοποίηση Ήχου

## Καθορισμός του προτύπου για συμπίεση ήχου υψηλής ποιότητας

### ΕΙΣΑΓΩΓΗ

Η ομάδα MPEG (Moving Pictures Expert Group), η οποία αποτελεί μέλος του Διεθνή Οργανισμού Προτύπων ISO (International Organization of Standards), έχει αναπτύξει μία σειρά οπτικοακουστικών προτύπων, γνωστή ως MPEG-1 και MPEG-2. Αυτά τα οπτικοακουστικά πρότυπα είναι τα πρώτα διεθνή πρότυπα στο χώρο της ακουστικής ψηφιακής συμπίεσης υψηλής ποιότη-



τας (high-quality). Η σειρά MPEG-1 καλύπτει την κωδικοποίηση στερεοφωνικών ακουστικών σημάτων σε υψηλούς ρυθμούς δειγματοληψίας, στοχεύοντας στη διαυγή-άριστη ποιότητα, ενώ αντίθετα η σειρά MPEG-2 προσφέρει επίσης κωδικοποίηση στερεοφωνικών ακουστικών σημάτων σε χαμηλότερους ρυθμούς δειγματοληψίας. Επιπλέον η MPEG-2 εισάγει πολυκαναλική κωδικοποίηση και με αλλά και χωρίς αναδρομική συμβατότητα με την MPEG-1, έτσι ώστε να παρέχει μία βελτιωμένη ακουστική εικόνα για ηχητικές μόνο εφαρμογές και για εξελιγμένα συστήματα τηλεόρασης και video. Η MPEG-2 κωδικοποίηση ήχου χωρίς αναδρομική συμβατότητα, η οποία καλείται MPEG-2 Advanced Audio Coding (AAC), προσφέρει τους υψηλότερους ρυθμούς συμπίεσης.

Τυπικές περιοχές εφαρμογών για ψηφιακό ήχο βασιζόμενο σε MPEG, είναι εκείνες της παραγωγής ακουστικών εφαρμογών (μουσική, ομιλία, ηχητικά εφέ), της διανομής και ανταλλαγής προγραμμάτων, της ψηφιακής εκπομπής ήχου, της ψηφιακής αποθήκευσης και γενικά όλων εκείνων των εφαρμογών που σχετίζονται με πολυμέσα (multimedia). Σε αυτή την εργασία θα περιγράψουμε αρκετά αναλυτικά τις τεχνολογίες κλειδί και τα βασικά χαρακτηριστικά των MPEG-1 και MPEG-2 κωδικοποιητών ήχου. Επίσης θα παρουσιάσουμε μία μικρή ενότητα των επερχόμενων MPEG-4 προτύπων και θα συζητήσουμε κάποιες από τις τυπικές εφαρμογές για MPEG ακουστική συμπίεση.

## 1. ΓΕΝΙΚΑ ΠΕΡΙ ΡΥΘΜΩΝ ΚΩΔΙΚΟΠΟΙΗΣΗΣ

### 1.1 PCM ρυθμοί bit

Τυπικές κατηγορίες ηχητικών σημάτων είναι η τηλεφωνική ομιλία, η ομιλία ευρείας ζώνης και ο ήχος ευρείας ζώνης, τα οποία διαφέρουν όλα στο εύρος ζώνης, στη δυναμική ακτίνα και στις ακουστικές απαιτήσεις της προσφερόμενης ποιότητας. Η ποιότητα της ομιλίας εύρους ζώνης ίδιου με εκείνο που χρησιμοποιείται στην τηλεφωνία είναι αποδεκτό για τηλεφωνικές επικοινωνίες και για κάποιες οπτικοηλεκτρονικές υπηρεσίες. Μεγαλύτερο εύρος συχνοτήτων (7 KHz για ευρείας ζώνης ομιλία) ενδέχεται να είναι απαραίτητο για να βελτιώσει την καθαρότητα και φυσικότητα της ομιλίας. Η ευρείας ζώνης (υψηλής πιστότητας) αναπαράσταση ήχου, συμπεριλαμβανομένου πολυκαναλικού ήχου, χρειάζεται ένα εύρος ζώνης τουλάχιστο 20 KHz. Η συνηθισμένη ψηφιακή διαμόρφωση για αυτή την κατηγορία σημάτων είναι η PCM (Pulse Code Modulation) με τυπικούς ρυθμούς δειγματοληψίας και αναλύσεις μεγεθών (PCM bits ανά δευτερόλεπτο) όπως δίνονται στον πίνακα 1.

ΒΑΣΙΚΕΣ ΠΑΡΑΜΕΤΡΟΙ ΓΙΑ PCM ΚΩΔΙΚΟΠΟΙΗΣΗ ΟΜΙΛΙΑΣ ΚΑΙ ΗΧΗΤΙΚΩΝ ΣΗΜΑΤΩΝ				
Κατηγορία	Εύρος συχνοτήτων σε Hz	Ρυθμός Δειγματοληψίας σε KHz	PCM bits ανά δείγμα	PCM ρυθμός bit σε Kb/s
Ομιλία τηλεφώνου	300-3.400 <sup>1</sup>	8	8	64
Ομιλία ευρείας ζώνης	50-7.000	16	8	128
Ήχος μεσαίας ζώνης	10-11.000	24	16	384
Ήχος ευρείας ζώνης	10-22.000	48 <sup>2</sup>	16	768

ΠΙΝΑΚΑΣ 1.

Το δισκάκι CD (Compact Disk), αποτελεί σήμερα το αδιαμφισβήτητο πρότυπο της αναπαράστασης του ψηφιακού ήχου. Έχοντας ένα ρυθμό δειγματοληψίας 44,1 KHz, προκύπτει στερεοφωνικός καθαρός ρυθμός bit (Πίνακας 2):

$$2 \times 44,1 \times 16 \times 1000 = 1,41 \text{ Mb/s}$$

Πάντως το CD χρειάζεται μία σημαντική επιβάρυνση για γραμμικό κώδικα περιορισμένου μήκους κατά την εφαρμογή, ο οποίος απεικονίζει 8 bits πληροφορίας σε 14 bits, για συγχρονισμό και για διόρθωση λαθών,

καταλήγοντας σε μία 49-bit αναπαράσταση για κάθε 16-bit ακουστικό δείγμα. Γι' αυτό το λόγο, ο ολικός στερεοφωνικός ρυθμός bit είναι :

$$1,41 \times 49/16 = 4,32 \text{ MB/s.}$$

Ο πίνακας 2 συγκρίνει ρυθμούς bit μεταξύ CD και DAT.

CD και DAT ρυθμοί bit (στερεοφωνικά σήματα, δειγματοληπτημένα στα 44,1 KHz. Το DAT επίσης παρέχει ρυθμούς δειγματοληψίας στα 32 KHz και 48 KHz)			
Αποθηκευτικό μέσο	Ακουστικός ρυθμός (Mb/s)	Επιβάρυνση (MB/s)	Ολικός ρυθμός bit (MB/s)
CD	1,41	2,91	4,32
Digital Audio Tape (DAT)	1,41	1,67	3,08

ΠΙΝΑΚΑΣ 2.

Για αρχειοθέτηση και επεξεργασία ακουστικών σημάτων, ρυθμοί δειγματοληψίας διπλάσιοι από εκείνους που αναφέρθηκαν και μεγέθη ανάλυσης μέχρι 24 bit ανά δείγμα είναι υπό συζήτηση. Επιπλέον, η κωδικοποίηση χωρίς απώλεια είναι ένα ενδιαφέρον θέμα, έτσι ώστε να μη θυσιάσουμε την ποιότητα ήχου με κανένα τρόπο [1]. Ο ψηφιακός δίσκος (DVD), με χωρητικότητα 4,7 GB (μονού στρώματος) ή 8,5 GB (διπλού στρώματος) είναι το κατάλληλο αποθηκευτικό μέσο για τέτοιες εφαρμογές.

## 1.2 Μείωση ρυθμού bit

Αν και τα κανάλια και τα δίκτυα υψηλών ρυθμών bit, έχουν γίνει πιο εύκολα προσιτά, οι χαμηλότεροι ρυθμοί bit για κωδικοποίηση ακουστικών σημάτων αποτελούν σίγουρα ένα σημαντικό θέμα.. Τα βασικά κίνητρα για κωδικοποίηση χαμηλότερου ρυθμού bit είναι, η ανάγκη να ελαχιστοποιήσουμε το κόστος μετάδοσης, ή παροχής συμφέρουσας αποθήκευσης, η απαίτηση να μεταδώσουμε μέσω καναλιών περιορισμένης χωρητικότητας (όπως τα κινητά ραδιοφωνικά κανάλια) και η υποστήριξη κωδικοποίησης μεταβλητού ρυθμού σε δίκτυα πακέτων.

Βασικές απαιτήσεις για τη σχεδίαση κωδικοποιητών ήχου χαμηλών ρυθμών bit, είναι πρώτον η εξασφάλιση της υψηλής ποιότητας του ανακατασκευασμένου σήματος με αντοχή στις μεταβολές του φάσματος και των επιπέδων. Στην περίπτωση των στερεοφωνικών και πολυκαναλικών σημάτων, η χωρική ακεραιότητα είναι μία επιπλέον διάσταση της ποιότητας. Δεύτερον,

απαιτείται η αντοχή σε τυχαία (random) και καταιγιστικά (burst) σφάλματα του καναλιού και η μη απώλεια πακέτων. Τρίτον, μεγάλη σημασία έχουν η χαμηλή πολυπλοκότητα και κατανάλωση ισχύος των συστημάτων κωδικοποίησης/αποκωδικοποίησης. Για παράδειγμα, σε εφαρμογές εκπομπής και αναπαραγωγής (playback), η πολυπλοκότητα και η κατανάλωση ισχύος των αποκωδικοποιητών ήχου που χρησιμοποιούνται πρέπει να είναι χαμηλές, δοθέντος ότι οι περιορισμοί στην πολυπλοκότητα του κωδικοποιητή είναι σταθεροποιημένοι. Επιπλέον απαιτήσεις που σχετίζονται με δίκτυα είναι, οι μικρές καθυστερήσεις κωδικοποιητή/αποκωδικοποιητή, η αντοχή σε σφάλματα που προκαλούνται από συζευγμένα συστήματα κωδικοποιτών/αποκωδικοποιτών και μία κομψή και ανεπαίσθητη υποβάθμιση της ποιότητας με αύξηση του ρυθμού σφαλμάτων σε εφαρμογές κινητής ραδιοφωνίας και εκπομπής. Τέλος σε επαγγελματικές εφαρμογές, τα κωδικοποιημένα ρεύματα (stream) bit πρέπει να επιτρέπουν την επεξεργασία (editing), εξασθένιση (fading), μίξη (mixing) και συμπίεση της δυναμικής περιοχής (dynamic range compression).

Έχουμε δει ταχεία ανάπτυξη στις τεχνικές συμπίεσης του ρυθμού bit για ακουστικά σήματα και ομιλία. [2-7]. Η γραμμική πρόβλεψη, η κωδικοποίηση υποζώνης, η κωδικοποίηση μετάδοσης όπως επίσης διάφορες μορφές διανυσματικής κβάντισης και τεχνικές κωδικοποίησης εντροπίας, είναι τεχνικές που έχουν χρησιμοποιηθεί για τη σχεδίαση ικανοποιητικών αλγορίθμων κωδικοποίησης, οι οποίοι επιτυγχάνουν αρκετά μεγαλύτερη συμπίεση από ότι θεωρείτο εφικτό μόλις λίγα χρόνια πριν. Πρόσφατα αποτελέσματα στην κωδικοποίηση ομιλίας και ήχου, υποδεικνύουν πως μια υπέροχη ποιότητα κωδικοποίησης μπορεί να αποκτηθεί με ρυθμούς bit από 0,5 έως 1 bit/δείγμα για ομιλία και ομιλία ευρείας ζώνης και 1 έως 2 bit/δείγμα για ήχο. Για την αποθήκευση και μετάδοση συστημάτων που χειρίζονται πακέτα, επιπλέον εξοικονόμηση μπορεί να πραγματοποιηθεί με τη χρησιμοποίηση κωδικοποίησης μεταβλητού ρυθμού με τη δυνατότητα να προσφέρει μία χρονικά ανεξάρτητη και σταθερής ποιότητας απόδοση.

Οι συμπιεσμένες ψηφιακές αναπαραστάσεις ήχου μπορούν να γίνουν λιγότερο ευαίσθητες σε βλάβες του καναλιού σε σχέση με τις αντίστοιχες αναλογικές, εάν η κωδικοποίηση πηγής και καναλιού εφαρμοσθούν κατάλληλα. Η επέκταση του εύρους ζώνης έχει συχνά αναφερθεί ως μειονέκτημα της ψηφιακής κωδικοποίησης και μετάδοσης, αλλά με τις σημερινές τεχνικές συμπίεσης δεδομένων και πολυεπίπεδης σηματοδότησης (multilevel signaling), το εύρος ζώνης του καναλιού μπορεί να γίνει μικρότερο από εκείνο των αναλογικών συστημάτων. Σε συστήματα εκπομπής, οι μειωμένες απαιτήσεις εύρους ζώνης σε συνδυασμό με την αντοχή σε λάθη των αλγορίθμων κωδικοποίησης, θα επιτρέψει την ικανοποιητική χρήση διαθέσιμων ραδιοφωνικών και τηλεοπτικών καναλιών όπως επίσης και των καναλιών «taboo» τα οποία τώρα μένουν κενά εξαιτίας προβλημάτων interference.

### 1.3 Ενέργειες MPEG καθιέρωσης προτύπων

Ιδιαίτερης σημασίας για ψηφιακό ήχο αποτελεί η εργασία καθιέρωσης προτύπων που επιτελείται στον οργανισμό ISO/IEC, η οποία στοχεύει στην παροχή διεθνών προτύπων, για μια πλατειά περιοχή εφαρμογών που βασίζονται στις επικοινωνίες και την αποθήκευση. Η ομάδα που ασχολείται με αυτό το έργο καλείται MPEG (Moving Pictures Experts Group). Η αρχική προσπάθεια αυτής της ομάδας ήταν η MPEG Phase1 (MPEG-1) πρότυπη κωδικοποίηση IS 11172, η οποία υποστηρίζει ρυθμούς bit περίπου 1,2 Mb/s για εικόνα (με ποιότητα εικόνας συγκρίσιμη με αυτή των σημερινών αναλογικών VCR- Video Cassette Recorders) και 256 Kb/s για δικάναλο ήχο (με ποιότητα ήχου συγκρίσιμη με αυτή των σημερινών CD) [8].

Το πιο πρόσφατο MPEG-2 πρότυπο IS 13818, παρέχει, για την πλευρά της εικόνας, πρότυπα για υψηλής ποιότητας εικόνα (συμπεριλαμβανομένου τηλεόραση υψηλής καθαρότητας High Definition TV –HDTV) με ρυθμούς bit από 3 με 15 Mb/s και παραπάνω. Από την πλευρά του ήχου, καθιερώθηκε πολυκαναλική κωδικοποίηση ήχου με δύο έως πέντε πλήρη εύρη ζώνης. Επιπλέον, για στερεοφωνικό ήχο η περιοχή των ρυθμών δειγματοληψίας επεκτάθηκε σε χαμηλότερες συχνότητες δειγματοληψίας για ρυθμούς bit από ή λιγότερο 64 Kb/s [9]. Το τμήμα IS 13818-7 αυτού του προτύπου θα προσφέρει μία συλλογή από πολύ εύκαμπτα εργαλεία για προχωρημένη κωδικοποίηση ήχου (MPEG-2 AAC -Advanced Audio Coding) για εφαρμογές όπου η συμβατότητα με MPEG-1 δεν είναι σχετική.

Τέλος, η τρέχουσα MPEG-4 εργασία απευθύνεται στον καθορισμό προτύπων οπτικοακουστικής κωδικοποίησης για μία περιοχή εφαρμογών, όπως κινητή πρόσβαση (mobile-access), μικρής πολυπλοκότητας τερματικά πολυμέσα, μέχρι υψηλής ποιότητας πολυκαναλικά ηχητικά συστήματα. Αυτό το πρότυπο θα επιτρέψει την διεπικοινωνία και παγκόσμια προσπέλαση, και θα παρέχει σε μεγάλο βαθμό ευκαμψία και επεκτασιμότητα [10].

## 2. ΤΕΧΝΟΛΟΓΙΕΣ ΚΛΕΙΔΙ ΣΤΗΝ ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ

### 2.1 Εισαγωγικά

Μετά τις προτάσεις για μείωση του εύρους ζώνης των ρυθμών κωδικοποίησης για την ομιλία, ακολούθησαν οι αντίστοιχες προτάσεις για μείωση του εύρους ζώνης της κωδικοποίησης ήχου. Οι διαφορές μεταξύ των σημάτων ομιλίας και ήχου είναι πολλαπλές, πάντως, η κωδικοποίηση ήχου συνεπάγεται υψηλότερους ρυθμούς δειγματοληψίας, καλύτερη ανάλυση πλάτους, υψηλότερη δυναμική ακτίνα, μεγαλύτερες διαφοροποιήσεις στην πυκνότητα ισχύος, στερεοφωνικές και πολυκαναλικές αναπαραστάσεις ηχητικού σήματος και τελικά υψηλότερες προσδοκίες στην ποιότητα. Και πραγματικά, η υψηλή



ποιότητα του CD με την 16 bit/δείγμα μορφοποίηση έχει κάνει τον ψηφιακό ήχο δημοφιλή.

Η κωδικοποίηση ομιλίας και ήχου παρουσιάζουν ομοιότητες στο σημείο, ότι και στις δύο περιπτώσεις, η ποιότητα βασίζεται στις ιδιότητες της ανθρώπινης αίσθησης της ακουστικής ικανότητας. Από την άλλη πλευρά, η ομιλία μπορεί να κωδικοποιηθεί αρκετά ικανοποιητικά διότι ένα μοντέλο παραγωγής ομιλίας είναι διαθέσιμο, ενώ αντίθετα δεν υπάρχει τίποτα αντίστοιχο για ηχητικά σήματα.

Μέτριες μειώσεις στους ρυθμούς bit του ήχου έχουν αποκτηθεί με ακαριαία συμπίεση/αποσυμπίεση (compranding) (π.χ. μία μετατροπή του ομοιόμορφου (uniform) 14-bit PCM σε μία 11-bit μη ομοιόμορφη PCM αναπαράσταση), ή με πρόδρομο προσαρμοστικό PCM (συμπίεση/αποσυμπίεση μπλοκ), όπως εφαρμόζεται σε διάφορες μορφές γειτονικής ταυτόχρονης πολυπλεγμένης κωδικοποίησης συμπίεσης/αποσυμπίεσης ήχου (near-instantaneously compranding audio multiplex (NICAM) coding (ITU-R, Rec 660)). Για παράδειγμα η Βρετανική Τηλεόραση BBC (British Broadcasting Corporation), έχει χρησιμοποιήσει την μορφοποίηση κωδικοποίησης NICAM 728 για ψηφιακή μετάδοση ήχου σε πολλά Ευρωπαϊκά Τηλεοπτικά Δίκτυα. Χρησιμοποιεί δειγματοληψία στα 32 KHz με 14-bit αρχική κβάντιση ακολουθούμενη από μία συμπίεση σε 10-bit μορφοποίηση με βάση ότι μπλοκ 1 ns καταλήγουν σε ολικό στερεοφωνικό ρυθμό bit 728 Kb/s [11]. Τέτοια προσαρμοστικά PCM σχήματα μπορούν να λύσουν το πρόβλημα παρέχοντας μία ικανοποιητική δυναμική περιοχή για κωδικοποίηση ήχου, αλλά δεν είναι επαρκή σχήματα συμπίεσης αφού δεν εκμεταλλεύονται στατιστικές εξαρτήσεις μεταξύ των δειγμάτων και δεν αφαιρούν επαρκώς θόρυβο σήματος (signal irrelevancies).

Στους πρόσφατους αλγόριθμους κωδικοποίησης ήχου, τέσσερις τεχνολογίες κλειδιά παίζουν σημαντικό ρόλο: η κωδικοποίηση αντίληψης (perceptual), η κωδικοποίηση με βάση τη συχνότητα (frequency-domain), μεταπήδηση παραθύρου (window switching) και ο δυναμικός καταμερισμός (allocation) bit. Αυτές οι τεχνολογίες θα καλυφθούν στα επόμενα τμήματα..

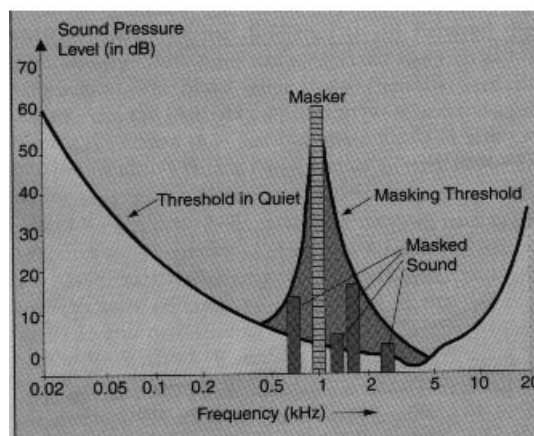
## **2.2 Ακουστική απόκρυψη (auditory masking) και κωδικοποίηση αντίληψης (perceptual coding)**

### Ακουστική απόκρυψη

Ο έσω ους πραγματοποιεί μία μικρών-όρων ανάλυση κρίσιμης ζώνης, όπου οι μετασχηματισμοί από τη συχνότητα στο χώρο συμβαίνουν κατά μήκος της βασιλικής μεμβράνης. Τα φάσματα ισχύος δεν αναπαριστώνται σε γραμμική κλίμακα συχνότητας, αλλά σε περιορισμένες ζώνες συχνότητας οι οποίες λέγονται *κρίσιμες ζώνες*. Το ακουστικό σύστημα μπορεί πρόχειρα να περιγραφεί ως ζωνοπερατή τράπεζα φίλτρων (bandpass filterbank), αποτελούμενη από ισχυρώς επικαλυπτόμενα ζωνοπερατά φίλτρα με εύρος ζώνης της τάξης από 50 έως 100 Hz για σήματα κάτω από 500 Hz, και μέχρι τα 5.000 Hz για σήματα

υψηλών συχνοτήτων. Είκοσι έξι κρίσιμες ζώνες που καλύπτουν συχνότητες μέχρι 24 KHz πρέπει να ληφθούν υπόψη.

*Ταυτόχρονη απόκρυψη* είναι ένα φαινόμενο στο πεδίο των συχνοτήτων, όπου ένα σήμα χαμηλής στάθμης (κρυμμένο-maskee) μπορεί να γίνει μη ακουστικό (masked) από ένα ισχυρότερο σήμα που συμβαίνει ταυτόχρονα (αποκρύβων-masker), εφόσον το κρυμμένο και το αποκρύβων είναι αρκετά κοντά μεταξύ τους στη συχνότητα [12]. Τέτοια απόκρυψη είναι μεγαλύτερη στην κρίσιμη ζώνη στην οποία εντοπίζεται το αποκρύβων σήμα και επιδρά σε χαμηλότερο βαθμό σε γειτονικές ζώνες.. Ένα κατώφλι απόκρυψης (masking threshold) μπορεί να μετρηθεί, και τα σήματα χαμηλού επιπέδου τα οποία είναι κάτω από το κατώφλι δε θα είναι ακουστικά αντιληπτά. Αυτό το κρυμμένο σήμα μπορεί να αποτελείται από εισφορές σημάτων χαμηλού επιπέδου, κβαντισμένου θορύβου, αλλοιωμένη παραμόρφωση ή από σφάλματα μετάδοσης. Το κατώφλι απόκρυψης ,γνωστό από την κωδικοποίηση πηγής και ως κατώφλι μόλις αντιληπτής παραμόρφωσης (Just Noticeable Distortion – JND), μεταβάλλεται με το χρόνο. Εξαρτάται από το επίπεδο πίεσης του ήχου



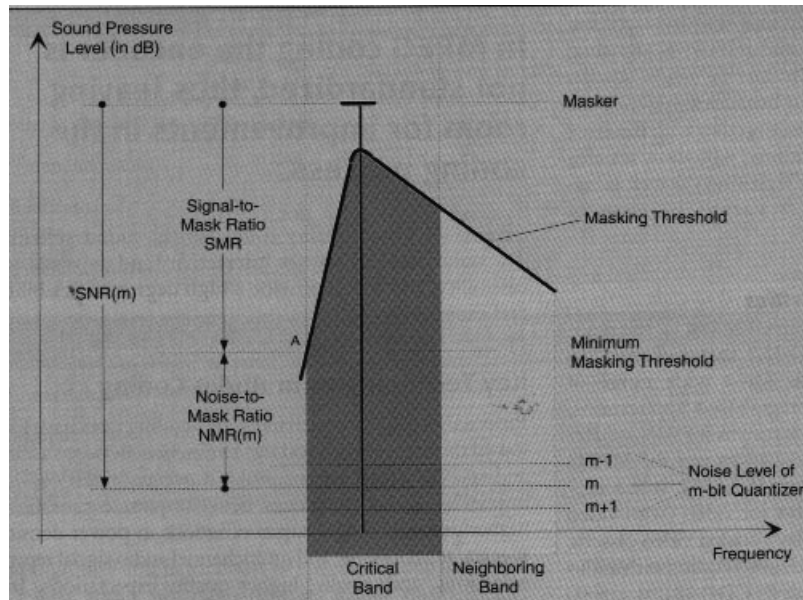
*Σχήμα 1. Κατώφλι σε ηρεμία και κατώφλι απόκρυψης (τα ακουστικά γεγονότα στις γραμμοσκιασμένες περιοχές δε θα είναι ακουστικά αντιληπτά.)*

(Sound Pressure Level –SPL), τη συχνότητα του αποκρύβοντος και τα χαρακτηριστικά του αποκρύβοντος και του κρυμμένου. Ας πάρουμε ένα παράδειγμα για κατώφλι απόκρυψης για το SPL=60 db στενής ζώνης αποκρύβων του σχήματος 1.

Γύρω από το 1KHz τα τέσσερα κρυμμένα σήματα θα κρυφτούν εφόσον τα επίπεδα πίεσης ήχου του καθενός βρίσκονται κάτω από το κατώφλι απόκρυψης. Η κλίση του κατωφλίου απόκρυψης είναι πιο απότομη προς τις χαμηλότερες συχνότητες δηλαδή οι υψηλότερες συχνότητες μπορούν πιο εύκολα να κρυφτούν. Θα πρέπει να σημειωθεί ότι η απόσταση μεταξύ αποκρύβοντος και κατωφλίου απόκρυψης είναι μικρότερη σε πειράματα τόνου απόκρυψης θορύβου (noise-masking-tone) από ότι σε πειράματα θορύβου απόκρυψης τόνου (tone-masking-noise) δηλαδή ο θόρυβος είναι καλύτερος αποκρύπτης από τον

τόνο. Σε MPEG κωδικοποιητές και τα δύο κατώφλια παίζουν ρόλο στον υπολογισμό του κατώφλιου απόκρυψης.

Χωρίς αποκρύβον, ένα σήμα δε θα είναι ακουστικά αντιληπτό αν το επίπεδο πίεσης ήχου του είναι κάτω από το κατώφλι σε ηρεμία, το οποίο εξαρτάται από τη συχνότητα και καλύπτει μία δυναμική περιοχή περισσότερο από 60 db όπως φαίνεται στη χαμηλότερη καμπύλη του σχήματος 1.



*Σχήμα 2. Κατώφλι απόκρυψης και πηλίκo σήμα-προς-μάσκα (SMR)(τα ακουστικά γεγονότα στις γραμμοσκιασμένες περιοχές δε θα είναι ακουστικά αντιληπτά)*

Το ποιοτικό σχέδιο του σχήματος 2 δίνει μερικές ακόμα λεπτομέρειες γύρω από το κατώφλι απόκρυψης. Μέσα σε μία κρίσιμη ζώνη, οι τόνοι κάτω από αυτό το κατώφλι (πιο σκούρα περιοχή) αποκρύπτονται. Η απόσταση μεταξύ του επιπέδου του αποκρύβοντος και του κατωφλίου απόκρυψης καλείται πηλίκo σήμα προς απόκρυψη (SMR Signal-Mask-Ratio). Η μεγαλύτερη τιμή είναι εκείνη στο αριστερό περίγραμμα της κρίσιμης ζώνης (σημείο A Σχήμα 2) και η μικρότερη τιμή του συμβαίνει στην περιοχή συχνοτήτων του αποκρύβοντος και είναι περίπου 6 db στα πειράματα τόνου κρυμμένου από θόρυβο (noise-masking-tone). Υποθέτοντας m-bit κβάντιση ενός ακουστικού σήματος, μέσα σε μία κρίσιμη περιοχή ο κβαντισμένος θόρυβος δε θα είναι αντιληπτός εφόσον το πηλίκo σήμα προς θόρυβο (SNR) είναι υψηλότερο από το SMR. Οι συνεισφορές θορύβου και σήματος έξω από τη συγκεκριμένη κρίσιμη περιοχή επίσης θα αποκρυφτούν – αν και σε μικρότερο βαθμό- εάν τα αντίστοιχα SPL είναι κάτω από το κατώφλι απόκρυψης.

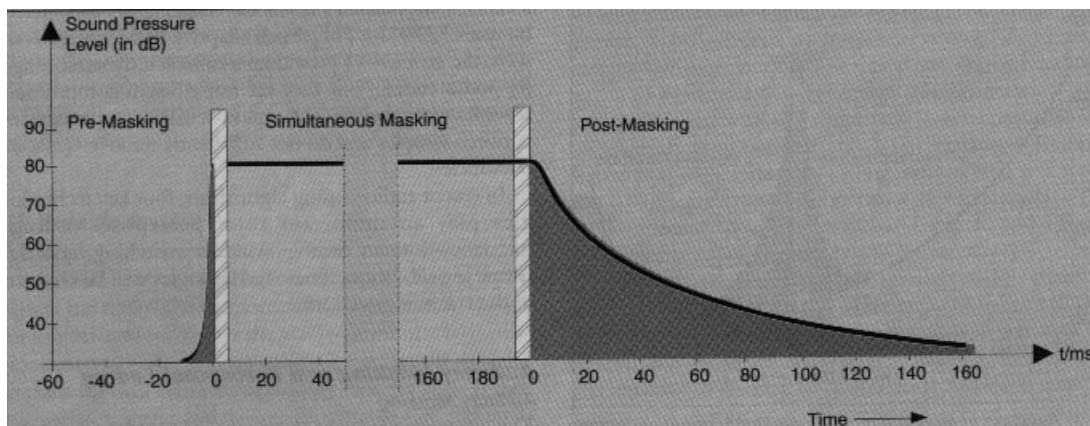
Ορίζοντας  $SNR(m)$  ως το SNR που προκύπτει από μία m-bit κβάντιση, η αισθητή παραμόρφωση για μία δοσμένη υποζώνη υπολογίζεται από το πηλίκo θόρυβος προς απόκρυψη (Noise to Mask Ratio NMR):

$$\text{NMR}(m) = \text{SMR} - \text{SNR}(m) \quad (\text{σε db})$$

Το  $\text{NMR}(m)$  περιγράφει τη διαφορά σε db μεταξύ του SMR και του SNR που αναμένεται από μία m-bit κβάντιση. Η τιμή NMR είναι επίσης η διαφορά (σε db) μεταξύ του επιπέδου του κβαντισμένου θορύβου και του επιπέδου όπου μία παραμόρφωση μπορεί μόλις να γίνει αντιληπτή για μία δοσμένη υποζώνη. Μέσα σε μία κρίσιμη ζώνη, ο κωδικοποιημένος θόρυβος δε θα είναι αντιληπτός εφόσον το  $\text{NMR}(m)$  είναι αρνητικό.

Έχουμε περιγράψει την απόκρυψη με ένα μόνο αποκρύβον. Αν η πηγή σήματος αποτελείται από πολλούς ταυτόχρονους maskers, ο καθένας έχει το δικό του κατώφλι απόκρυψης και ένα συνολικό κατώφλι απόκρυψης μπορεί να υπολογισθεί που περιγράφει το κατώφλι των μόλις αντιληπτών παραμορφώσεων ως μία συνάρτηση της συχνότητας (βλ. επίσης ενότητα ISO/MPEG-1 Κωδικοποίηση Ήχου).

Επιπλέον με την ταυτόχρονη απόκρυψη, το φαινόμενο χρόνου της προσωρινής απόκρυψης παίζει ένα σημαντικό ρόλο στην ανθρώπινη ακουστική αντίληψη. Μπορεί να συμβεί όταν δύο ήχοι εμφανίζονται μέσα σε ένα μικρό χρονικό διάστημα. Ανάλογα με τα μεμονωμένα SPL, ο δυνατότερος ήχος μπορεί να αποκρύψει τον ασθενέστερο, ακόμα και εάν το κρυμμένο προηγείται του αποκρύβοντος (Σχήμα 3).



Σχήμα 3. Προσωρινή απόκρυψη (τα ακουστικά γεγονότα στις γραμμοσκιασμένες περιοχές δε θα είναι ακουστικά αντιληπτά)

Η προσωρινή απόκρυψη μπορεί να βοηθήσει να αποκρύψουμε προ-ηχούς που προκαλούνται από την εξάπλωση ενός ξαφνικού μεγάλου σφάλματος κβάντισης πάνω στο πραγματικό κωδικό μπλοκ (βλ. τμήμα «Μεταπήδηση παραθύρου»). Η διάρκεια μέσα στην οποία εφαρμόζεται προαπόκρυψη είναι σημαντικά μικρότερη από το ένα δέκατο της αντίστοιχης μετα-απόκρυψης, η οποία είναι της τάξης των 50 έως 500 ns. Και οι δύο προ- και μετά- απόκρυψη χρησιμοποιούνται σε MPEG αλγορίθμους κωδικοποίησης ήχου.

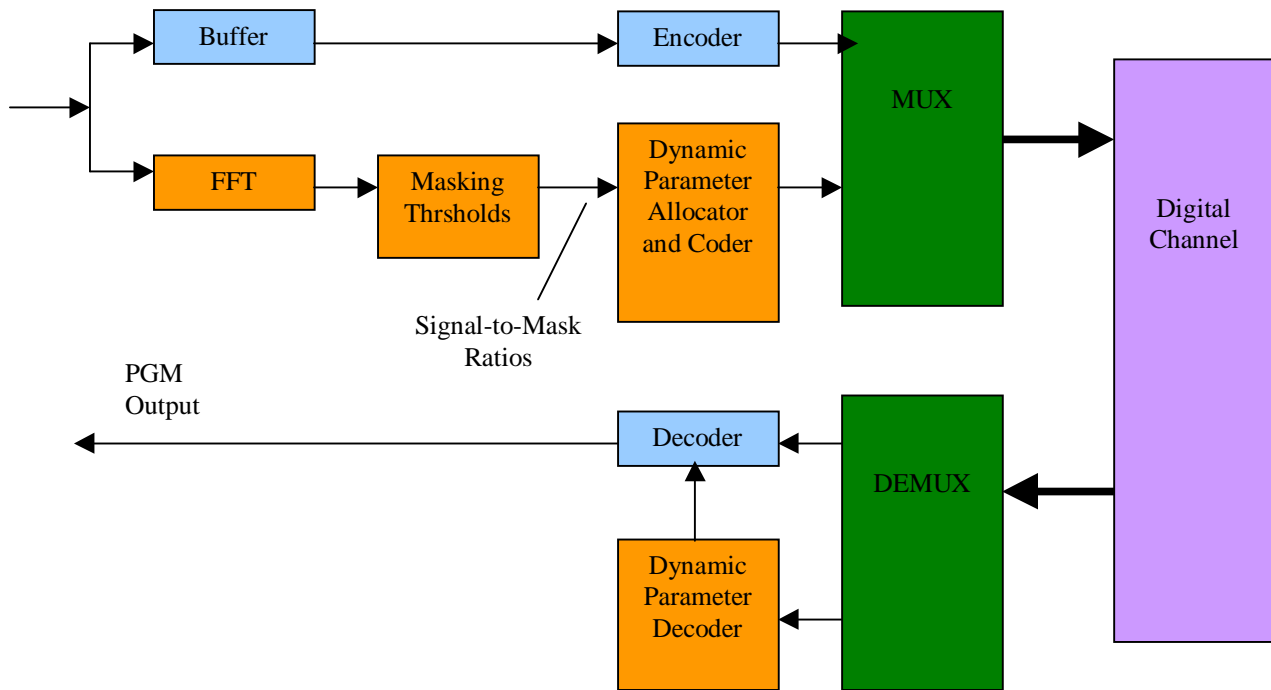
### *Κωδικοποίηση αντίληψης*

Η ψηφιακή κωδικοποίηση σε υψηλούς ρυθμούς bit, διατηρεί την κυματομορφή δηλαδή η κυματομορφή πλάτους σε σχέση με το χρόνο του κωδικοποιημένου σήματος προσεγγίζει εκείνη του σήματος εισόδου. Η διαφορά σήματος μεταξύ κυματομορφών εισόδου και εξόδου γίνεται το βασικό κριτήριο σφάλματος στη σχεδίαση κωδικοποιητών. Οι αρχές κωδικοποίησης κυματομορφών έχουν καλυφθεί λεπτομερώς στο [2]. Σε χαμηλότερους ρυθμούς bit, γεγονότα γύρω από την παραγωγή και αντίληψη ηχητικών σημάτων πρέπει να συμπεριληφθούν στη σχεδίαση κωδικοποιητών, και το κριτήριο σφάλματος πρέπει να μοιάζει ενός σήματος εξόδου το οποίο είναι χρήσιμο στον ανθρώπινο δέκτη, παρά να προτιμούμε ένα σήμα εξόδου που ακολουθεί και διατηρεί την κυματομορφή εισόδου. Βασικά ένας ικανοποιητικός αλγόριθμος κωδικοποίησης πηγής θα (α) απομακρύνει περιττά συστατικά του σήματος πηγής εκμεταλλευόμενος συσχετισμούς μεταξύ των δειγμάτων του και (β) απομακρύνει συστατικά που είναι αντιληπτά άσχετα με το αυτί. Η μη σχετικότητα εμφανίζεται από μόνη της ως ένα άχρηστο μέγεθος ή ανάλυση συχνότητας. Μέρη του σήματος της πηγής που έχουν αποκρυφτεί δεν χρειάζεται να μεταδοθούν.

Η εξάρτηση της ανθρώπινης ακουστικής αντίληψης από τη συχνότητα και από την ακουστική ανοχή σε σφάλματα, η οποία την συνοδεύει, μπορούν (και θα έπρεπε) απευθείας να επηρεάσουν την σχεδίαση κωδικοποιητών. Τεχνικές σχηματοποίησης θορύβου (noise-shaping) μπορούν να δώσουν έμφαση στην κωδικοποίηση θορύβου σε ζώνες συχνότητας όπου ο θόρυβος δεν είναι σημαντικός στην αντίληψη. Σε αυτό το τέλος, η ολίσθηση θορύβου πρέπει να είναι δυναμικά προσαρμοσμένη στο πραγματικό φάσμα εισόδου μικρών-όρων σε συμφωνία με το SMR, το οποίο μπορεί να γίνει με διαφορετικούς τρόπους. Πάντως, τα βάρη συχνότητας που βασίζονται σε γραμμικό φίλτράρισμα, όπως τυπικά γίνεται στην κωδικοποίηση ομιλίας, δεν μπορούν να χρησιμοποιήσουν πλήρως τα αποτελέσματα από την ψυχοακουστική. Επομένως στην κωδικοποίηση ήχου ευρείας ζώνης, οι παράμετροι της σχηματοποίησης θορύβου ελέγχονται δυναμικά με ένα πιο ικανοποιητικό τρόπο ώστε να εκμεταλλευτούν την ταυτόχρονη απόκρυψη και την προσωρινή απόκρυψη.

Το σχήμα 4 απεικονίζει τη δομή ενός κωδικοποιητή που βασίζεται στην αντίληψη, ο οποίος εκμεταλλεύεται την ακουστική απόκρυψη. Η διαδικασία κωδικοποίησης ελέγχεται από το SMR σε αντιπαράθεση με την καμπύλη συχνότητας από την οποία η απαιτούμενη ανάλυση πλάτους (και από τώρα η κατανομή και ο ρυθμός bit) σε κάθε ζώνη συχνότητων προέρχεται. Το SMR τυπικά καθορίζεται από μία υψηλή ανάλυση, ας πούμε, μία 1024 σημείων φασματική ανάλυση βασισμένη σε FFT από το ακουστικό μπλοκ που πρόκειται να κωδικοποιηθεί. Γενικά, κάθε σχήμα κωδικοποίησης μπορεί να χρησιμοποιηθεί, το οποίο μπορεί δυναμικά να ελεγχθεί από τέτοια πληροφορία αίσθησης/αντίληψης. Οι κωδικοποιητές πεδίου συχνότητων (βλέπε επόμενη ενότητα) είναι ιδιαίτερου ενδιαφέροντος, αφού προσφέρουν μία άμεση μέθοδο για σχηματοποίηση θορύβου. Εάν η συχνότητα ανάλυσης αυτών των

κωδικοποιητών είναι αρκετά υψηλή, το SMR μπορεί να προκύψει απευθείας από τα δείγματα υποζώνης ή να μετασχηματίσει συντελεστές χωρίς να τρέξει παράλληλα φασματική ανάλυση βασισμένη σε FFT.



Σχήμα 4. Σχηματικό διάγραμμα κωδικοποιητών με βάση την αντίληψη (ακουστικά γεγονότα στις γκρι περιοχές δε θα είναι αντιληπτά)

Αν ο απαραίτητος ρυθμός bit για μία πλήρη απόκρυψη της παραμόρφωσης είναι διαθέσιμος, το σχήμα κωδικοποίησης θα είναι αισθητά προφανές δηλαδή το αποκωδικοποιημένο σήμα είναι τότε υποκειμενικά μη διακρίσιμο από το σήμα της πηγής. Σε πρακτικές σχεδιάσεις, δεν μπορούμε να φτάσουμε τα όρια της μόλις παρατηρήσιμης παραμόρφωσης, αφού η μετέπειτα επεξεργασία του ακουστικού σήματος από τον τελικό χρήστη και οι πολλαπλές διεργασίες κωδικοποιήσεις/αποκωδικοποιήσεις στους δεσμούς μετάδοσης πρέπει να ληφθούν υπόψη. Πέραν τούτου, η γνώση μας αυτή τη στιγμή γύρω από την ακουστική απόκρυψη είναι πολύ περιορισμένη. Οι γενικεύσεις από αποτελέσματα απόκρυψης, που προέκυψαν από απλούς και σταθερούς maskers και για περιορισμένα εύρη ζώνης, μπορούν να είναι κατάλληλες για τα περισσότερα σήματα πηγών, αλλά μπορούν να αποτύχουν για άλλα.. Επομένως ως μία επιπρόσθετη απαίτηση, χρειαζόμαστε ένα ικανοποιητικό σημάδι ασφαλείας για πρακτικές σχεδιάσεις τέτοιων κωδικοποιητών βασισμένων στην αντίληψη. Θα πρέπει να σημειωθεί ότι το MPEG πρότυπο κωδικοποίησης ήχου είναι ανοικτό για καλύτερα ψυχοακουστικά μοντέλα κατανεμημένα από κωδικοποιη-

τή, αφού τέτοια μοντέλα δεν είναι κανονιστικά στοιχεία του προτύπου (βλ ενότητα «ISO/MPEG-1 Κωδικοποίηση Ήχου»).

### 2.3 Κωδικοποίηση στο χώρο των συχνοτήτων

Ως παράδειγμα δυναμικής σχηματοποίησης θορύβου, μπορεί να χρησιμοποιηθεί η κβάντιση ανατροφοδότησης θορύβου σε σχήματα πρόβλεψης [16,17]. Πάντως, οι κωδικοποιητές του πεδίου των συχνοτήτων με δυναμικό καταμερισμό των bit ( και επομένως συνεισφορές κβάντισης του θορύβου) σε υποζώνες ή οι συντελεστές μετασχηματισμού προσφέρουν ένα ευκολότερο και πιο ακριβή τρόπο ελέγχου της κβάντισης θορύβου [2,14] (βλ επίσης την ενότητα «δυναμικός καταμερισμός bit»).

Σε όλους τους κωδικοποιητές του πεδίου των συχνοτήτων, ο πλεονασμός (τα μη επίπεδα φασματικά χαρακτηριστικά μικρών όρων του σήματος πηγής) και η μη σχετικότητα (σήματα κάτω από τα ψυχοακουστικά κατώφλια) χρησιμοποιούνται, ώστε να μειώσουν το ρυθμό των μεταδιδόμενων δεδομένων με σεβασμό στο PCM. Αυτό επιτυγχάνεται χωρίζοντας το φάσμα της πηγής σε ζώνες συχνοτήτων ώστε να ενεργοποιηθούν γειτονικά μη συσχετισμένα φασματικά συστατικά και κβαντίζοντας αυτά τα συστατικά ξεχωριστά. Υπάρχουν δύο κατηγορίες κωδικοποίησης, η κωδικοποίηση μετάδοσης (transform coding TC) και κωδικοποίηση υποζώνης (subband coding SBC). Η διαφοροποίηση μεταξύ αυτών των δύο κατηγοριών οφείλεται κυρίως σε ιστορικούς λόγους. Και οι δύο χρησιμοποιούν μία τράπεζα φίλτρων ανάλυσης στον κωδικοποιητή ώστε να αποσυνθέσει το σήμα εισόδου μέσα σε υποδειγματοληπτημένα φασματικά συστατικά. Τα φασματικά συστατικά καλούνται δείγματα υποζώνης εάν η τράπεζα φίλτρων έχει χαμηλή ανάλυση συχνότητας, διαφορετικά καλούνται φασματικές γραμμές ή συντελεστές μετασχηματισμού. Αυτά τα φασματικά συστατικά επανασυνδυάζονται στον αποκωδικοποιητή μέσω τραπεζών φίλτρων σύνθεσης.

Στο SBC, το σήμα της πηγής τροφοδοτείται σε μία τράπεζα φίλτρων ανάλυσης, η οποία αποτελείται από  $M$  ζωνοπερατά φίλτρα τα οποία είναι συνεχόμενα στη συχνότητα, έτσι ώστε το σύνολο των σημάτων υποζώνης μπορεί να επανασυνδυαστεί προσθετικά ώστε να παράγει το αρχικό σήμα ή έστω μία κοντινή του έκδοση. Η έξοδος κάθε φίλτρου αποδεκατίζεται κρίσιμα (δηλαδή δειγματοληπτείται στο διπλάσιο του νόμιμου εύρους ζώνης) από ένα παράγοντα ίσο με  $M$ , τον αριθμό των ζωνοπερατών φίλτρων. Αυτός ο αποδεκατισμός καταλήγει σε ένα συνολικό αριθμό δειγμάτων υποζώνης που ισούται με εκείνο στο σήμα της πηγής. Στο δέκτη, ο ρυθμός δειγματοληψίας της κάθε υποζώνης αυξάνεται σε εκείνο του σήματος της πηγής συμπληρώνοντας τον κατάλληλο αριθμό από μηδενικά δείγματα. Τα παρεμβάλλοντα σήματα υποζώνης εμφανίζονται στις εξόδους των ζωνοπερατών φίλτρων της τράπεζας. Η επεξεργασία δειγματοληψίας μπορεί να εισάγει αλλοιωμένη παραμόρφωση, εξαιτίας της ιδιότητας επικάλυψης των υποζωνών. Αν τέλεια φίλτρα, όπως δύο

ζωνών δευτεροβάθμια αντικριστά είδωλα ή πολυφασικά φίλτρα, χρησιμοποιηθούν, οι όροι αλλοίωσης θα μηδενιστούν και το άθροισμα των εξόδων των ζωνοπερατών θα ισούται με το σήμα της πηγής απουσία κβάντισης [18-21]. Με κβάντιση, τα συστατικά αλλοίωσης δε θα μηδενιστούν ιδανικά. Ωστόσο, τα σφάλματα δε θα είναι αντιληπτά στην MPEG κωδικοποίηση ήχου, αν χρησιμοποιηθεί ένας αρκετός αριθμός bit. Πάντως αυτά τα λάθη μπορούν να μειώσουν την αρχική δυναμική περιοχή από 20 bits σε περίπου 18 bits [15].

Στο TC, ένα μπλοκ δειγμάτων εισόδου μετασχηματίζεται γραμμικά μέσω ενός διακριτού μετασχηματισμού, σε ένα σύνολο από γειτονικούς –μη συσχετισμένους συντελεστές μετασχηματισμού. Αυτοί οι συντελεστές κβαντίζονται και μεταδίδονται σε ψηφιακή μορφή στον αποκωδικοποιητή. Στον αποκωδικοποιητή ένας αντίθετος μετασχηματισμός απεικονίζει το σήμα πίσω στο πεδίο του χρόνου. Απουσία σφαλμάτων κβάντισης, η σύνθεση αποφέρει ακριβή ανακατασκευή. Τυπικοί μετασχηματισμοί είναι ο διακριτός μετασχηματισμός Fourier ή ο διακριτός μετασχηματισμός συνημιτόνου (DCT), υπολογισμένος μέσω ενός FFT, και άλλες τροποποιημένες εκδοχές τους. Έχουμε ήδη αναφέρει πως ο αντίστροφος μετασχηματισμός που βασίζεται στον αποκωδικοποιητή μπορεί να θεωρηθεί ως η σύνθεση τράπεζας φίλτρων δηλαδή οι κρουστικές αποκρίσεις των ζωνοπερατών του φίλτρων ισούνται με τις βασικές ακολουθίες του μετασχηματισμού. Οι κρουστικές αποκρίσεις της τράπεζας φίλτρων της ανάλυσης είναι ακριβώς οι χρονικά αντίστροφες εκδόσεις τους. Τα πεπερασμένα μήκη αυτών των αποκρίσεων μπορούν να προκαλέσουν τις αποκαλούμενες περιορισμένες από μπλοκ επιδράσεις. Σύγχρονοι κωδικοποιητές χρησιμοποιούν μία τροποποιημένη DCT (MDCT) τράπεζα φίλτρων όπως προ-τάθηκε από τους Princen και Bradley [20]. Το MDCT τυπικά βασίζεται σε μία επικάλυψη κατά 50% μεταξύ μπλοκ επιτυχημένης ανάλυσης. Χωρίς κβάντιση είναι ελεύθερα από περιοριστικές επιδράσεις μπλοκ, έχουν υψηλότερο κέρδος κωδικοποίησης μετασχηματισμού από το DCT και οι βασικές τους ακολουθίες αντιστοιχούν σε καλύτερες ζωνοπερατές αποκρίσεις. Παρουσία κβάντισης, οι επιδράσεις από περιορισμούς μπλοκ ατονούν λόγω των διπλών κρουστικών αποκρίσεων του φίλτρου που προκαλούνται από την επικάλυψη.

Υβριδικές τράπεζες φίλτρων, όπως συνδυασμοί διακριτού μετασχηματισμού και εφαρμογές τράπεζας φίλτρων, έχουν χρησιμοποιηθεί συχνά στην κωδικοποίηση ομιλίας και ήχου [22,23]. Ένα από τα πλεονεκτήματα είναι ότι διαφορετικές αναλύσεις συχνότητας μπορούν να παραχθούν σε διαφορετικές συχνότητες με ένα εύκαμπτο τρόπο και με χαμηλή πολυπλοκότητα. Για παράδειγμα, μία υψηλή φασματική ανάλυση μπορεί να αποκτηθεί με ένα ικανοποιητικό τρόπο με τη χρήση μιάς σειράς από τράπεζες φίλτρων (με τις μικρές τους καθυστερήσεις) και ένα γραμμικό MDCT μετασχηματισμό ο οποίος χωρίζει κάθε ακολουθία υποζώνης περαιτέρω σε συστατικό συχνότητας ώστε να επιτύχει μια υψηλής συχνότητας ανάλυση. Οι MPEG ακουστικοί κωδικοποιητές χρησιμοποιούν μία προσέγγιση υποζώνης στο Στρώμα I και Στρώμα II και μία υβριδική τράπεζα φίλτρων στο Στρώμα III.

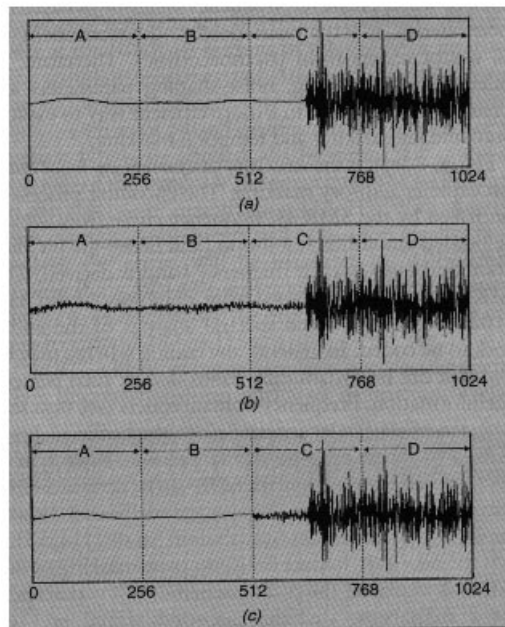


## 2.4 Εναλλαγή/Μεταπήδηση Παραθύρου

Ένα κρίσιμο τμήμα της κωδικοποίησης στο πεδίο των συχνοτήτων των ακουστικών σημάτων είναι η εμφάνιση των προ-ηχών οι οποίες είναι όμοιες με τις επιδράσεις κατά την αντιγραφή στις αναλογικές ταινίες. Θεωρείστε την περίπτωση όπου μία περίοδος σιγής ακολουθείται από κρουστικό ήχο, όπως από καστανιέτες ή τρίγωνα μέσα στο ίδιο μπλοκ κωδικοποίησης. Τέτοια έναρξη («επίθεση») θα προκαλέσει συγκριτικά μεγάλα στιγμιαία σφάλματα κβάντισης. Στο TC, ο αντίστροφος μετασχηματισμός στη διεργασία αποκωδικοποίησης θα διανείμει τέτοια λάθη στα μπλοκ. Ομοίως στο SBC, τα ζωνοπερτά φίλτρα του αποκωδικοποιητή θα εξαπλώσουν τέτοια σφάλματα. Και στις δύο απεικονίσεις οι προ-ηχούς μπορούν να γίνουν διακριτικά αντιληπτές, ειδικά σε χαμηλούς ρυθμούς bit με συγκριτικά υψηλές εισφορές σφαλμάτων. Οι προ-ηχούς μπορούν να αποκρυφτούν από το αποτέλεσμα στο πεδίο του χρόνου της προαπόκρυψης, εάν η εξάπλωση του χρόνου είναι μικρού μήκους (της τάξης των λίγων ms). Επομένως μπορούν να μειωθούν ή να αποφευχθούν με τη χρήση μπλοκ μικρού μήκους. Πάντως ένα μεγαλύτερο ποσοστό του ολικού ρυθμού bit απαιτείται τυπικά για τη μετάδοση πλευρικής πληροφορίας εάν τα μπλοκ είναι μικρότερα.. Μία λύση σε αυτό το πρόβλημα είναι να μεταπηδήσουμε μεταξύ μεγεθών μπλοκ διαφορετικών μηκών όπως προτάθηκε από τον Elder (window switching) [24]. Τυπικά μεγέθη μπλοκ είναι μεταξύ  $N=64$  και  $N=1024$ . Τα μικρά μπλοκ χρησιμοποιούνται μόνο για να ελέγχουν ατέλειες κατά τη διάρκεια μη στατικών περιόδων του σήματος, διαφορετικά ο κωδικοποιητής μεταπηδά πίσω σε μεγάλα μπλοκ. Είναι προφανές πως η επιλογή του μεγέθους του μπλοκ πρέπει να στηριχθεί στην ανάλυση των χαρακτηριστικών του πραγματικού ακουστικού μπλοκ κωδικοποίησης. Το σχήμα 5 δείχνει την επίδραση στην κωδικοποίηση μετασχηματισμού: αν το μέγεθος μπλοκ είναι  $N=1024$  (Σχ. 5β) οι προ-ηχούς είναι καθαρές (ορατές και ) ακουστικά αντιληπτές ενώ ένα μπλοκ μεγέθους  $N=256$  θα μειώσει αυτές τις επιδράσεις επειδή είναι περιορισμένες στο μπλοκ όπου το σήμα επιτίθεται και συμβαίνουν τα αντίστοιχα σφάλματα κβάντισης (Σχ 5γ). Επιπλέον η προαπόκρυψη μπορεί να γίνει αποτελεσματική.

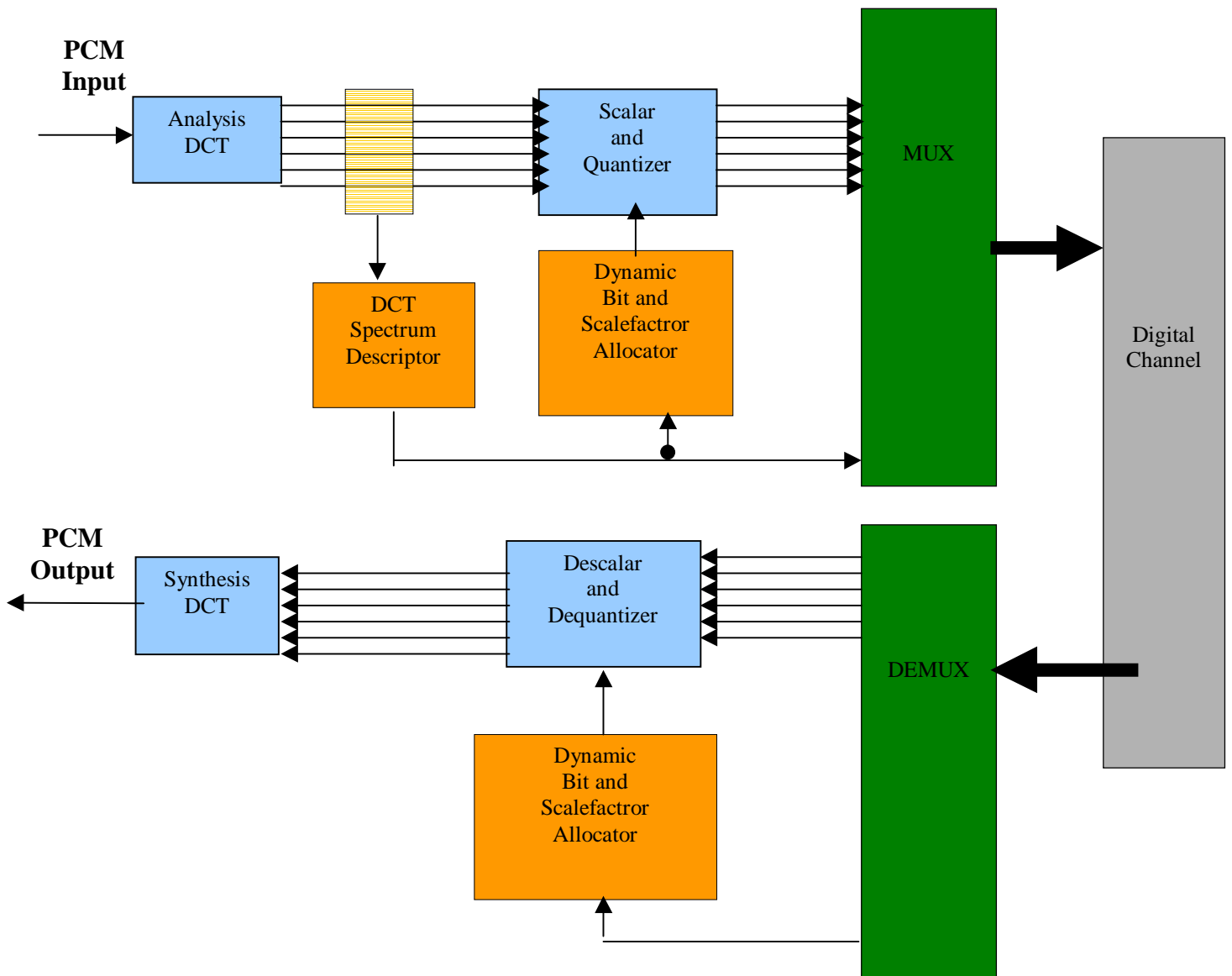
## 2.5 Δυναμικός καταμερισμός bit

Η κωδικοποίηση στο χώρο των συχνοτήτων κερδίζει σημαντικά στις επιδόσεις εάν ο αριθμός των bit που τίθεται σε καθέναν από τους κβαντιστές των συντελεστών μετασχηματισμού, προσαρμόζεται στο φάσμα μικρών-όρων των ακουστικών μπλοκ κωδικοποίησης σε μία βάση μπλοκ με μπλοκ. Στα



*Σχήμα 5. Μεταπήδηση παραθύρου: (α) σήμα πηγής (β) ανακατασκευασμένο σήμα με μήκος μπλοκ  $N=1024$ , (γ) ανακατασκευασμένο σήμα με μήκος μπλοκ  $N=256$ )*

μέσα του 1970 οι Zelinski και Noll εισήγαγαν το δυναμικό καταμερισμό (dynamic allocation) και παρουσίασαν σημαντικές με βάση SNR και υποκειμενικές βελτιώσεις με τον προσαρμοστικό τους μετασχηματισμό κωδικοποίησης (ATC-Adaptive Transform Coding) (βλ. Σχήμα 6) [14-26]. Πρότειναν μία DCT απεικόνιση και ένα δυναμικό αλγόριθμο καταμερισμού bit, ο οποίος χρησιμοποιούσε τους DCT συντελεστές μετασχηματισμού για να υπολογίσει ένα φασματικό φάκελο μικρών-όρων βασισμένο σε DCT. Οι παράμετροι αυτού του φάσματος κωδικοποιούνται και μεταδίδονται από εκεί όπου το φάσμα μικρών-όρων υπολογίστηκε χρησιμοποιώντας γραμμική παρεμβολή στο λογαριθμικό χώρο. Αυτός ο υπολογισμός χρησιμοποιείται τότε για τον υπολογισμό του ιδανικότερου αριθμού bit για κάθε συντελεστή μετασχηματισμού και στον κωδικοποιητή και στον αποκωδικοποιητή. Το ATC είχε ένα αριθμό αδυναμιών, όπως επιδράσεις από περιορισμό του μπλοκ, προ-ηχούς, επισφαλή εκμετάλλευση της απόκρυψης και χαμηλή υποκειμενική ποιότητα σε χαμηλούς ρυθμούς bit. Παρά τις αδυναμίες, βρίσκουμε πολλά από τα χαρακτηριστικά του συνηθισμένου ATC στους περισσότερους πρόσφατους κωδικοποιητές στο πεδίο των συχνοτήτων. Παραδείγματα από τις πολύ εξεζητημένες στρατηγικές καταμερισμού bit που οι ακουστικοί MPEG αλγόριθμοι κωδικοποίησης χρησιμοποιούν περιγράφονται λεπτομερώς στο τμήμα «Στρώματα 1 και 2».



Σχήμα 6. Προσαρμοστικό μετασχηματισμός κωδικοποίησης (ATC-Adaptive Transform Coding)

## ISO/MPEG-1 ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ

### 3.1 Εισαγωγικά

Το MPEG πρότυπο ακουστικής κωδικοποίησης [8,27-29] έχει ήδη γίνει ένα παγκόσμιο πρότυπο σε διάφορα πεδία όπως καταναλωτικά ηλεκτρονικά, επαγγελματική επεξεργασία ήχου, τηλεπικοινωνίες και εκπομπή [30]. Το πρότυπο συνδυάζει χαρακτηριστικά των αλγορίθμων κωδικοποίησης

MUSICAM και ASPEC [31-32]. Βασικά βήματα βελτίωσης προς το MPEG-1 ακουστικό πρότυπο περιγράφονται στο [29,33]. Η MPEG-1 κωδικοποίηση ήχου προσφέρει μία υποκειμενική ποιότητα αναπαραγωγής που είναι ισοδύναμη με την ποιότητα του CD (16-bit PCM) σε στερεοφωνικούς ρυθμούς οι οποίοι δίνονται στον Πίνακα 3 για πολλά είδη μουσικής. Εξαιτίας της υψηλής του δυναμικής περιοχής, το ακουστικό MPEG-1 έχει δυνατότητες να ξεπεράσει την ποιότητα ενός CD [30-34].

Προσεγγιστικοί MPEG-1 ρυθμοί bit για διαυγείς αναπαραστάσεις ηχητικών σημάτων και αντίστοιχοι συντελεστές συμπίεσης (σε σύγκριση με ρυθμούς CD)		
MPEG-1 Κωδικοποίηση ήχου	Προσεγγιστικοί στέρεο ρυθμοί bit για διαυγή ποιότητα	Συντελεστής συμπίεσης
Στρώμα I	384 kb/s	4
Στρώμα II	192 kb/s	8
Στρώμα III	128 kb/s	12

\*Μέσος ρυθμός bit: υποθέτουμε μεταβλητό ρυθμό bit κωδικοποίησης

Πίνακας 3.

### 3.2 Βασικές έννοιες

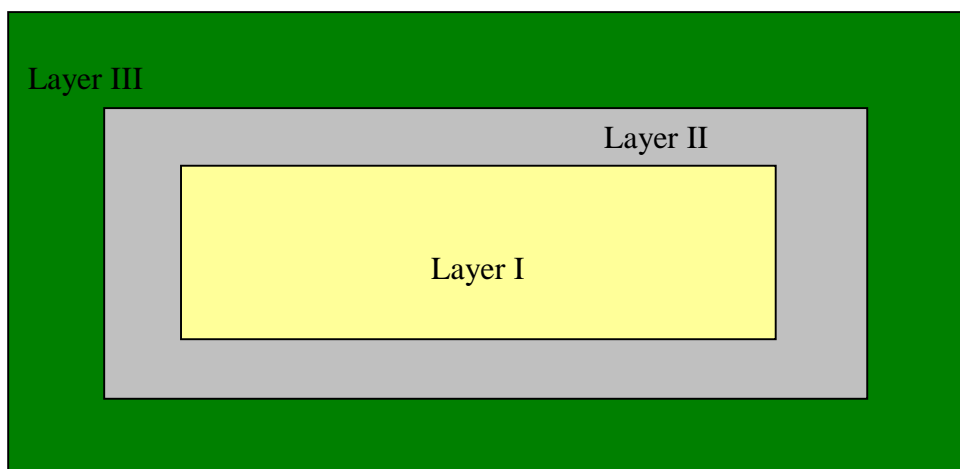
- Δομή

Η βασική δομή των MPEG-1 κωδικοποιητών ήχου ακολουθεί εκείνη των κωδικοποιητών που βασίζονται στην αίσθηση/αντίληψη (βλ. Σχ. 4). Στο πρώτο βήμα το ακουστικό σήμα μετατρέπεται στα φασματικά του συστατικά μέσω μίας τράπεζας φίλτρων ανάλυσης. Τα στρώματα I και II κάνουν χρήση μίας τράπεζας φίλτρων υποζώνης και το στρώμα III απασχολεί μία υβριδική τράπεζα φίλτρων. Κάθε φασματικό συστατικό κβαντίζεται και κωδικοποιείται με το επίτευγμα της διατήρησης του θορύβου κβάντισης κάτω από το κατώφλι απόκρυψης. Ο αριθμός των bit για κάθε υποζώνη και ένας συντελεστής κλίμακας καθορίζονται μπλοκ με μπλοκ. Κάθε μπλοκ έχει 12 (Στρώμα I) ή 36 (Στρώματα II και III) δείγματα υποζώνης (βλ. την ενότητα «Στρώματα I και II»). Ο αριθμός των bit του κβαντιστή προκύπτει από ένα αλγόριθμο δυναμικού καταμερισμού bit (Στρώματα I και II), ο οποίος ελέγχεται από ένα ψυχοακουστικό μοντέλο (βλ. παρακάτω). Οι κωδικές λέξεις υποζώνης, ο συντελεστής κλίμακας και η πληροφορία καταμερισμού bit πολυπλέκονται σε ένα ρεύμα bit, μαζί με μία επικεφαλίδα και με προαιρετικά βοηθητικά δεδομένα. Στον αποκωδικοποιητή η τράπεζα φίλτρων της σύνθεσης ανακατασκευάζει ένα μπλοκ από 32 ακουστικά δείγματα εξόδου από το αποπολυπλεγμένο ρεύμα bit. Το MPEG-1 ήχου υποστηρίζει ρυθμούς δειγματοληψίας των 32, 44.1 και 48

KHz και ρυθμούς bit μεταξύ 32 kb/s (μονο) και 448 kb/s, 384 kb/s και 320 kb/s (στέρεο και Στρώματα I , II και III αντίστοιχα).

- *Στρώματα και τρόποι λειτουργίας*

Το πρότυπο αποτελείται από τρία στρώματα (I ,II και III) αύξουσας πολυπλοκότητας, καθυστέρησης και υποκειμενικής απόδοσης. Από τη σκοπιά του υλικού και του λογισμικού, τα υψηλότερα στρώματα συγχωνεύουν τα κύρια δομικά μπλοκ των χαμηλότερων στρωμάτων (Σχ. 7). Ένας πρότυπος πλήρως MPEG-1 αποκωδικοποιητής ήχου είναι ικανός να αποκωδικοποιήσει ρεύματα bit από όλα και τα τρία στρώματα. Πιο συνηθισμένοι είναι οι αποκωδικοποιητές ακουστικού MPEG-1 Στρώματος X (X= I, II ή III).



Σχήμα 7. Ιεραρχία των Στρωμάτων I,II και III του MPEG-1/Ήχου

- *Κωδικοποίηση στερεοφωνικού πλεονασμού*

Το ακουστικό MPEG-1 υποστηρίζει τέσσερις τρόπους λειτουργίας: μονο, στέρεο, διπλό με δύο ξεχωριστά κανάλια (χρήσιμο για προγράμματα δίγλωσσα) και συνδυασμένο στέρεο. Στον προαιρετικό τρόπο λειτουργίας συνδυασμένου στέρεο, οι εσωκαναλικές εξαρτήσεις εκμεταλλεύονται έτσι ώστε να μειώσουν τον ολικό ρυθμό bit, χρησιμοποιώντας μία τεχνική μη σχετικότητας η οποία καλείται *ένταση στέρεο*. Είναι γνωστό ότι, πάνω από τα 2 KHz και μέσα σε κάθε κρίσιμη ζώνη, το ανθρώπινο ακουστικό σύστημα στηρίζει την αντίληψη του σε στερεοφωνικά ακούσματα περισσότερο στην προσωρινή περιβάλλουσα του ακουστικού σήματος παρά στην τέλεια προσωρινή δομή του. Επομένως ο αλγόριθμος για MPEG συμπίεση ήχου υποστηρίζει ένα τρόπο λειτουργίας στερεοφωνικού πλεονασμού ο οποίος καλείται *κωδικοποίηση έντασης στέρεο* (intensity stereo coding), η οποία μειώνει τον ολικό ρυθμό bit χωρίς να παραβιάζει τη χωρική ακεραιότητα του στερεοφωνικού σήματος.

Σε αυτό τον τρόπο λειτουργίας ο κωδικοποιητής κωδικοποιεί μερικές εξόδους υποζώνης υψηλότερης συχνότητας με ένα μόνο σήμα αθροίσματος

L+R (ή κάποιο γραμμικό συνδυασμό τους) αντί να στέλνει ανεξάρτητα αριστερά (L) και δεξιά (R) σήματα υποζώνης. Ο αποκωδικοποιητής ανακατασκευάζει τα αριστερά και δεξιά κανάλια βασιζόμενος μόνο στο μονό σήμα L+R και στους ανεξάρτητους αριστερούς και δεξιούς συντελεστές κλίμακας του καναλιού. Γι' αυτό το λόγο, το φασματικό σχήμα των αριστερών και δεξιών εξόδων είναι το ίδιο μέσα σε κάθε υποζώνη κωδικοποίησης έντασης αλλά τα μεγέθη είναι διαφορετικά [35]. Ο προαιρετικός τρόπος λειτουργίας του συνδυασμένου στέρεο θα επιδρά μόνο εάν ο απαιτούμενος ρυθμός bit ξεπερνά το διαθέσιμο ρυθμό και θα εφαρμόζεται μόνο σε υποζώνες που αντιστοιχούν σε συχνότητες περίπου 2 KHz και πάνω. Το Στρώμα III έχει μία επιπλέον επιλογή: στον τρόπο λειτουργίας μονο/στερεο (M/S) τα αριστεροκάναλα και δεξιοκάναλα σήματα κωδικοποιούνται ως μεσαία (L+R) και πλευρικά (L-R) κανάλια. Αυτός ο τελευταίος τρόπος λειτουργίας μπορεί να συνδυαστεί με τον τρόπο λειτουργίας συνδυασμένου στέρεο.

- *Ψυχοακουστικά Μοντέλα*

Έχουμε ήδη αναφέρει ότι ο προσαρμοστικός αλγόριθμος καταμερισμού bit ελέγχεται από ένα ψυχοακουστικό μοντέλο. Αυτό το μοντέλο υπολογίζει τα SMR και λαμβάνει υπόψη το φάσμα μικρών όρων του ακουστικού μπλοκ που πρόκειται να κωδικοποιηθεί και τη γνώση γύρω από το θόρυβο απόκρυψης. Αυτό το μοντέλο χρειάζεται μόνο στον κωδικοποιητή, το οποίο κάνει τον αποκωδικοποιητή λιγότερο πολύπλοκο. Αυτή η ασυμμετρία είναι ένα επιθυμητό χαρακτηριστικό για ακουστική αναπαραγωγή (playback) και εφαρμογές εκπομπής ήχου.

Το κανονιστικό μέρος του προτύπου περιγράφει τον αποκωδικοποιητή και την έννοια του κωδικοποιημένου ρεύματος bit, αλλά ο κωδικοποιητής δεν προτυποποιείται, αφήνοντας έτσι χώρο για μία εξελικτική βελτίωση του κωδικοποιητή. Συγκεκριμένα, διαφορετικά ψυχοακουστικά μοντέλα μπορούν να χρησιμοποιηθούν που εκτείνονται από πολύ απλές (ή καθόλου) σε πολύ πολύπλοκες απαιτήσεις βασισμένες στην ποιότητα και εφαρμοσιμότητα. Η πληροφορία γύρω από το φάσμα μικρών όρων μπορεί να προκύψει με διάφορους τρόπους. Για παράδειγμα, ως ένας ακριβής υπολογισμός από φασματική ανάλυση με FFT ακουστικών σημάτων εισόδου, ή, με λιγότερη ακρίβεια απ' ευθείας από τα φασματικά συστατικά όπως στο συνηθισμένο ATC [14] (βλ. Σχ. 6). Οι κωδικοποιητές μπορούν επίσης να βελτιστοποιηθούν για ορισμένες εφαρμογές. Όλοι αυτοί οι κωδικοποιητές μπορούν να χρησιμοποιηθούν με πλήρη συμβατότητα με όλους τους υπάρχοντες MPEG-1 ακουστικούς αποκωδικοποιητές.

Το ενημερωτικό τμήμα του προτύπου δίνει δύο παραδείγματα μοντέλων βασισμένων σε FFT (βλ επίσης [8,29,36]). Και τα δύο μοντέλα αναγνωρίζουν, με διαφορετικούς τρόπους τονικά και μη τονικά φασματικά συστατικά και χρησιμοποιούν τα αντίστοιχα αποτελέσματα των πειραμάτων θορύβου κρυμμένου από τόνο και τόνου κρυμμένου από θόρυβο, για τον υπολογισμό των

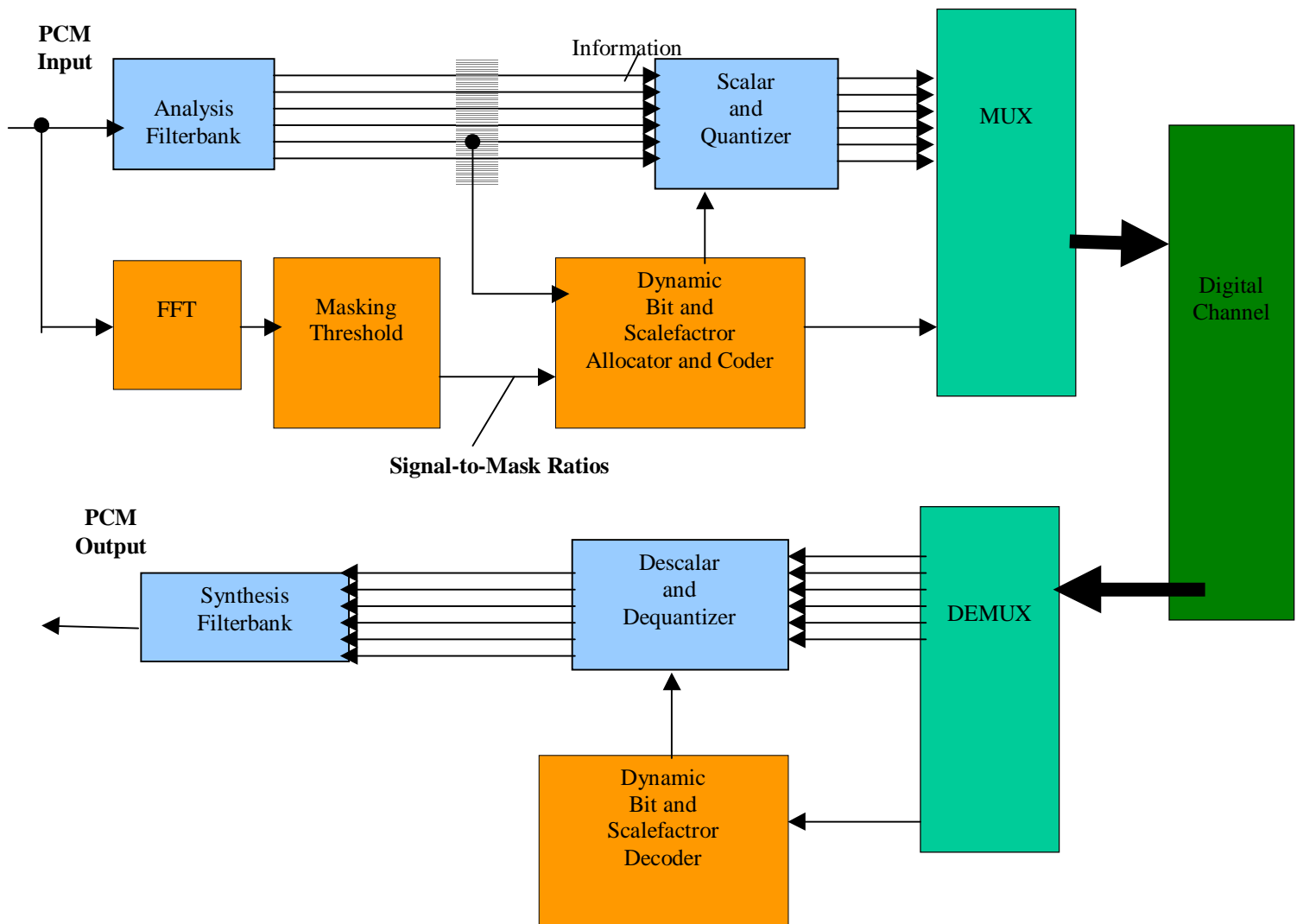
συνολικών κατωφλίων απόκρυψης. Πληροφορίες των πρότυπων πειραματικών αποτελεσμάτων δίνονται και για τα δύο ψυχοακουστικά μοντέλα που περιγράφηκαν [36]. Στο ενημερωτικό τμήμα του προτύπου ένας μετασχηματισμός FFT 512 σημείων προτείνεται για το Στρώμα I και ένας FFT μετασχηματισμός 1024 σημείων για τα Στρώματα II και III. Και στα δύο μοντέλα τα ακουστικά δείγματα εισόδου είναι ζυγισμένα κατά Hann. Το μοντέλο 1, το οποίο μπορεί να χρησιμοποιηθεί για τα Στρώματα I και II, υπολογίζει για κάθε masker το ανεξάρτητο του κατώφλι απόκρυψης, λαμβάνοντας υπόψη τη θέση συχνότητας του, τη ισχύ και την τονική πληροφορία. Το γενικό κατώφλι απόκρυψης αποκτάται ως το άθροισμα όλων των ανεξαρτήτων κατωφλίων απόκρυψης και του απόλυτου κατωφλίου απόκρυψης.. Το SMR είναι τότε το πηλίκο του μέγιστου επιπέδου σήματος μέσα σε μία δοσμένη υποζώνη και της ελαχίστης τιμής του γενικού κατωφλίου απόκρυψης σε αυτή τη δοσμένη υποζώνη (Σχ. 2). Το μοντέλο 2, το οποίο μπορεί να χρησιμοποιηθεί για όλα τα στρώματα, είναι πιο πολύπλοκο. Η τονικότητα θεωρείται όταν μία απλή πρόβλεψη υποδεικνύει ένα υψηλό κέρδος πρόβλεψης, τα κατώφλια απόκρυψης υπολογίζονται κοχλία χώρο, δηλαδή ιδιότητες του έσω ους λαμβάνονται υπόψη με μεγαλύτερη λεπτομέρεια και τελικά στην περίπτωση ενδεχόμενων προ-ηχών το γενικό κατώφλι απόκρυψης προσαρμόζεται κατάλληλα.

### 3.3 Στρώματα I και II

Οι MPEG κωδικοποιητές στρωμάτων I και II έχουν πολύ όμοιες δομές. Ο κωδικοποιητής Στρώματος II επιτυγχάνει μία καλύτερη απόδοση, κυρίως διότι η ολική πληροφορία πλευρικού συντελεστή κλίμακας μειώνεται με την εκμετάλλευση πλεονασμών μεταξύ των συντελεστών κλίμακας. Επιπροσθέτως παρέχεται μία ελαφρώς καλύτερη κβάντιση.

- *Τράπεζα φίλτρων*

Οι κωδικοποιητές Στρώματος I και II απεικονίζουν την ψηφιακή ακουστική είσοδο σε 32 υποζώνες μέσω ισοκατανεμημένων ζωνοπερατών φίλτρων (Σχ. 8 και 9). Μία δομή πολυφασικού φίλτρου χρησιμοποιείται για την απεικόνιση στη συχνότητα. Τα φίλτρα του έχουν 512 συντελεστές. Οι πολυφασικές δομές είναι υπολογιστικά πολύ επαρκείς αφού ένα DCT μπορεί να χρησιμοποιηθεί στη διαδικασία φιλτραρίσματος και είναι μέσης πολυπλοκότητας και χαμηλής καθυστέρησης. Στην αρνητική πλευρά, τα φίλτρα έχουν ίσες αποστάσεις και επομένως οι ζώνες συχνοτήτων δεν ανταποκρίνονται καλά στο διαμερισμό κρίσιμης ζώνης.(βλ προηγούμενη ενότητα της απόκρυψης αντίληψης). Σε ένα ρυθμό δειγματοληψίας 48 KHz, κάθε ζώνη έχει ένα εύρος  $24000/32=750$  Hz. Γι' αυτό το λόγο, σε χαμηλές συχνότητες, μία μονή υποζώνη καλύπτει ένα αριθμό γειτονικών κρίσιμων ζωνών. Τα σήματα υποζώνης ξαναδειγματοληφτούνται (αποδεκατίζονται κρίσιμα) σε ένα ρυθμό 1500 Hz. Η



Σχήμα 8. Δομή του MPEG-1 ακουστικού κωδικοποιητή και αποκωδικοποιητή (Στρώματα I και II)

κρουστική απόκριση της υποζώνης  $k$ ,  $h_{\text{sub}(k)}(n)$  αποκτάται με πολλαπλασιασμό της κρουστικής απόκρισης ενός μονού πρωτότυπου χαμηλοπεραστού φίλτρου,  $h(n)$  με μία συνάρτηση διαμόρφωσης που ολισθαίνει τη χαμηλοπερατή απόκριση στην κατάλληλη περιοχή υποζωνών συχνοτήτων.

$$h_{\text{sub}(k)}(n) = h(n) \cos [(2k-1)/2M + \varphi(k)]$$

$$\begin{aligned} M &= 32, \\ k &= 1, 2, \dots, 32 \\ n &= 1, 2, \dots, 512 \end{aligned}$$

Το πρωτότυπο χαμηλοπερατό φίλτρο  $h(n)$  έχει ένα εύρος ζώνης της τάξης  $750/2=375$  Hz στα 3 db και οι κεντρικές συχνότητες βρίσκονται σε περιττά πολλαπλάσια τριγύρω τους (όλες οι τιμές σε ρυθμό δειγματοληψίας 48 KHz).



Έτσι, οι υποδειγματοληπτιμένες έξοδοι των φίλτρων εκθέτουν μία σημαντική επικάλυψη. Πάντως, η σχεδίαση του πρωτότυπου φίλτρου και η συνυπολογισμός ολισθήσεων φάσης στους όρους συνημιτόνου, καταλήγουν σε μία ακύρωση αλλοίωσης (aliasing cancelation) στην έξοδο της τράπεζας φίλτρων σύνθεσης του αποκωδικοποιητή. Λεπτομέρειες γύρω από τους συντελεστές του πρωτότυπου φίλτρου και των ολισθήσεων φάσης  $\varphi(k)$  δίνονται στο πρότυπο ISO/MPEG. Πληροφορίες περί μίας επαρκούς εφαρμογής της τράπεζας φίλτρων μπορεί να βρεθεί στα [15] και [36] και το κέρδος στα κείμενα προτυποποίησης.

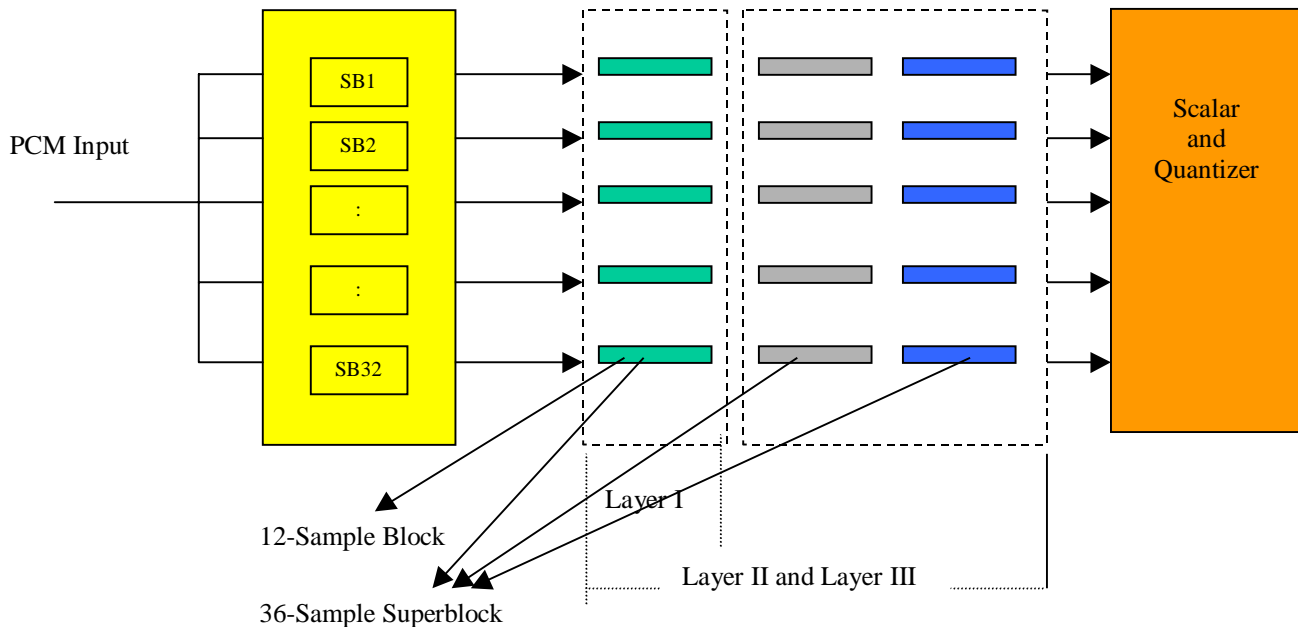
- *Κβάντιση*

Ο αριθμός των επιπέδων του κβαντιστή για κάθε φασματικό συστατικό προκύπτει από ένα κανόνα δυναμικού καταμερισμού bit που ελέγχεται από ένα ψυχοακουστικό μοντέλο. Ο αλγόριθμος καταμερισμού bit συλλέγει ένα ενιαίο μέσου βήματος κβαντιστή μέσα από ένα σύνολο διαθέσιμων κβαντιστών έτσι ώστε και η απαίτηση ρυθμού bit και η απαίτηση απόκρυψης να συμφωνούν. Η επαναληπτική διεργασία ελαχιστοποιεί το NMR σε κάθε υποζώνη. Ξεκινά με τον αριθμό των bits για τα δείγματα και οι συντελεστές κλίμακας μηδενίζονται. Σε κάθε βήμα επανάληψης ο κβαντιστής SNR,  $SNR(m)$ , αυξάνεται για μία m-bit υποζώνη κβαντιστή, παράγοντας τη μεγαλύτερη τιμή του NMR στην έξοδο του κβαντιστή. (Η αύξηση προκύπτει καταμερίζοντας ένα ακόμα bit). Για αυτό το σκοπό, το  $NMR$ ,  $NMR(m) = SMR - SNR(m)$ , υπολογίζεται ως η διαφορά (σε db) μεταξύ του πραγματικού επιπέδου κβάντισης θορύβου και του ελάχιστου γενικού κατωφλίου απόκρυψης. Το πρότυπο παρέχει πίνακες με υπολογισμούς για τον κβαντιστή SNR,  $SNR(m)$  για δεδομένο m.

Η τεχνική συμπίεσης/αποσυμπίεσης χρησιμοποιείται στη διαδικασία κβάντισης, δηλαδή μπλοκ από αποδεκατισμένα δείγματα σχηματίζονται και διαιρούνται με ένα συντελεστή κλίμακας τέτοιο ώστε το δείγμα μεγαλύτερου πλάτους να είναι ενιαίο. Στο Στρώμα I, μπλοκ από 12 αποδεκατισμένα και κλιμακωτά δείγματα σχηματίζονται σε κάθε υποζώνη (και για το αριστερό και για το δεξιό κανάλι) και υπάρχει ένας καταμερισμός bit για κάθε μπλοκ. Σε ρυθμό δειγματοληψίας 48 KHz, 12 δείγματα υποζώνης αντιστοιχούν σε 8ms ακούσματος. Υπάρχουν 32 μπλοκ, καθένα με 12 αποδεκατισμένα δείγματα, αναπαριστώντας  $32 \times 12 = 384$  ακουστικά δείγματα.

Στο στρώμα II σε κάθε υποζώνη ένα υπερμπλοκ 36 δειγμάτων σχηματίζεται από τρία συνεχόμενα μπλοκ των 12 αποδεκατισμένων δειγμάτων που αντιστοιχούν σε 24 ms ακούσματος σε ρυθμό δειγματοληψίας 48 KHz. Υπάρχει ένας καταμερισμός bit για κάθε υπερμπλοκ των 36 δειγμάτων. Όλα τα 32 υπερμπλοκ, καθένα με 36 αποδεκατισμένα δείγματα, αναπαριστούν όλα μαζί,  $32 \times 36 = 1152$  ακουστικά δείγματα. Όπως και στο Στρώμα I, ένας συντελεστής κλίμακας υπολογίζεται για κάθε μπλοκ των 12 δειγμάτων. Μία τεχνική μείωσης πλεονασμού χρησιμοποιείται για τη μετάδοση των συντελεστών κλίμακας : ανάλογα με τη σημαντικότητα των αλλαγών μεταξύ των τριών συνεχόμενων

συντελεστών κλίμακας, ένας, δύο ή και οι τρεις συντελεστές κλίμακας μεταδίδονται, μαζί με ένα 2-μπιτο συντελεστή κλίμακας επιλογής πληροφορίας. Συγκρινόμενος με το Στρώμα I, ο ρυθμός bit για τους συντελεστές κλίμακας μειώνεται περίπου 50% [29]. Το σχήμα 9 υποδεικνύει τη δομή συμπίεσης/από-συμπίεσης μπλοκ .

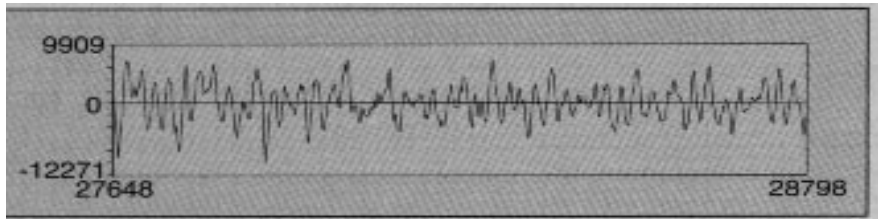


Σχήμα 9. Block Companding σε MPEG-1 ακουστικούς κωδικοποιητές

Τα κλιμακούμενα και κβαντισμένα φασματικά συστατικά υποζώνης μεταδίδονται στο δέκτη μαζί με το συντελεστή κλίμακας, το συντελεστή επιλογής (Στρώμα II) και την πληροφορία καταμερισμού bit. Η κβάντιση με συμπίεση/αποσυμπίεση μπλοκ παρέχει μία πολύ μεγάλη δυναμική περιοχή περισσότερο από 120 db. Για παράδειγμα, στο Στρώμα II οι ενιαίοι κβαντιστές μέσης στάθμης είναι διαθέσιμοι με 3,5,7,9,15,31,...,65535 επίπεδα για υποζώνες χαμηλής ένδειξης (χαμηλές συχνότητες) Στην περιοχή μεσαίων και υψηλών συχνοτήτων οι κβαντιστές έχουν ένα μειωμένο αριθμό επιπέδων. Για παράδειγμα, υποζώνες ένδειξης 24 έως 27 έχουν μόνο κβαντιστές με 3,5 και 65535 (!) επίπεδα. Οι κβαντιστές των 16 bit αποτρέπουν επιδράσεις υπερφόρτωσης. Υποζώνες ένδειξης 28 έως 32 δεν μεταδίδονται καθόλου. Για να μειώσουμε το ρυθμό δειγματοληψίας, οι κωδικές λέξεις τριών επιτυχημένων δειγμάτων υποζώνης που προκύπτουν από κβάντιση με 3-,5- και 9-βημάτων κβαντιστές, τους ανατίθεται μία κοινή κωδική λέξη. Το κέρδος σε ρυθμό bit είναι περίπου 40% [29].

- *Παραδείγματα κωδικοποίησης*

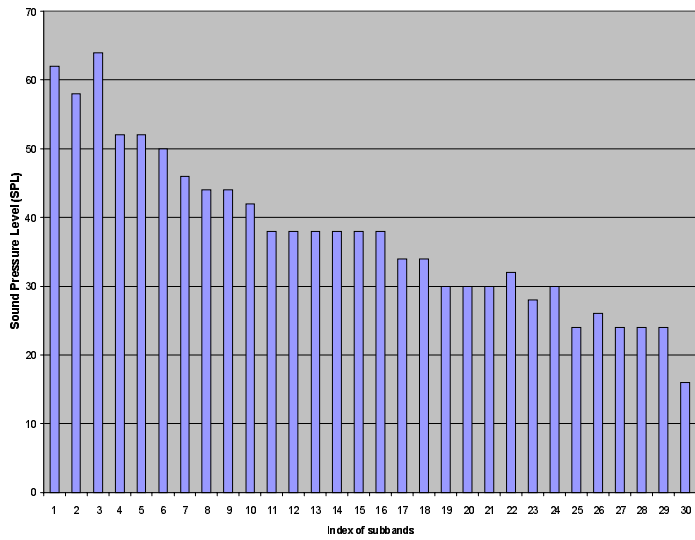
Τα ακόλουθα σχήματα δείχνουν τον τρόπο με τον οποίο το MPEG-1 Στρώμα I κωδικοποιεί τα ακουστικά σήματα. Το σχήμα 10 δείχνει ένα μεμονωμένο ακουστικό μπλοκ 1152 δειγμάτων που πρόκειται να κωδικοποιηθεί.



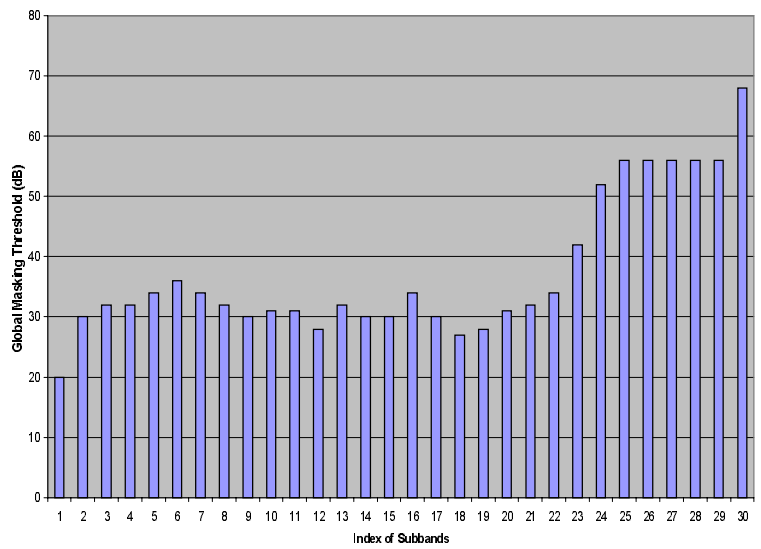
Σχήμα 10. Μπλοκ 1152 δειγμάτων ενός ηχητικού σήματος

Το σχήμα 11 δείχνει τις εξαρτήσεις από τη συχνότητα διαφόρων σημαντικών MPEG παραμέτρων. Οι άξονες συχνότητας διαιρούνται σύμφωνα με τις διακρίσεις υποζώνης. Ο ρυθμός δειγματοληψίας είναι 48 KHz, γι' αυτό το λόγο κάθε ένδειξη υποζώνης αναπαριστά μία υποζώνη εύρους ζώνης 750 Hz. Έχουμε διαλέξει ένα γενικό ρυθμό bit των 128 kb/s.

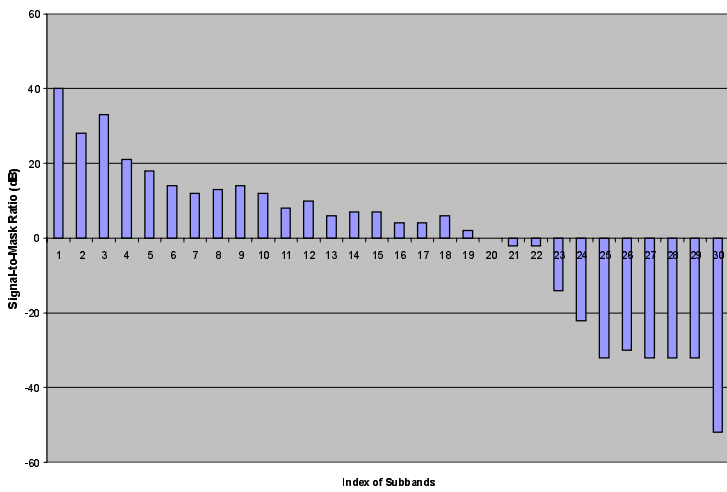
Το σχήμα 11(a) δείχνει την κατανομή συχνότητας του επιπέδου συμπίεσης του ήχου του ακουστικού μπλοκ. Από αυτή την κατανομή και από το κατώφλι σε ηρεμία ένα γενικό κατώφλι απόκρυψης μπορεί να προκύψει (Σχ. 11(b)). Για κάθε υποζώνη, το SMR (σε db) είναι η διαφορά μεταξύ του επιπέδου του masker και της ελάχιστης τιμής του γενικού κατωφλίου απόκρυψης (Σχ. 11(c)). Σημειώστε ότι για υποζώνες ένδειξης 23 και πάνω, η ισχύς του σήματος είναι σημαντικά χαμηλή σε σχέση του γενικού κατωφλίου απόκρυψης. Ανάλογα, τα αντίστοιχα σήματα υποζώνης δεν χρειάζονται μετάδοση. Στο επόμενο βήμα, ο αριθμός των bits ανά κβαντιστή υποζώνης επιλέγεται έτσι ώστε ο θόρυβος κβάντισης να διατηρείται σημαντικά χαμηλά από το γενικό κατώφλι απόκρυψης (Σχ. 11(d)). Επομένως, ο καταμερισμός bit για αυτές τις υποζώνες που πρέπει να μεταδοθούν, μόλις ακολουθεί το SMR. Το φάσμα του σφάλματος επανακατασκευής δίνεται στο σχήμα 11(e) (Να λάβετε υπόψη σας ότι οι τιμές σε db του σχήματος 11(e) καλύπτουν μόνο την περιοχή από 0 έως 35 db). Αν το συγκρίνουμε με το φάσμα του γενικού κατωφλίου απόκρυψης, παρατηρούμε ότι η ισχύς του σφάλματος ανακατασκευής είναι κάτω από το κατώφλι, συνεπώς, έχει αποκρυφτεί. Σημειώστε επίσης ότι το φάσμα του σφάλματος ανακατασκευής είναι πανομοιότυπο με εκείνο του φάσματος εισόδου για υποζώνες 23 και πάνω διότι τα αντίστοιχα σήματα υποζώνης δεν μεταδίδονται.



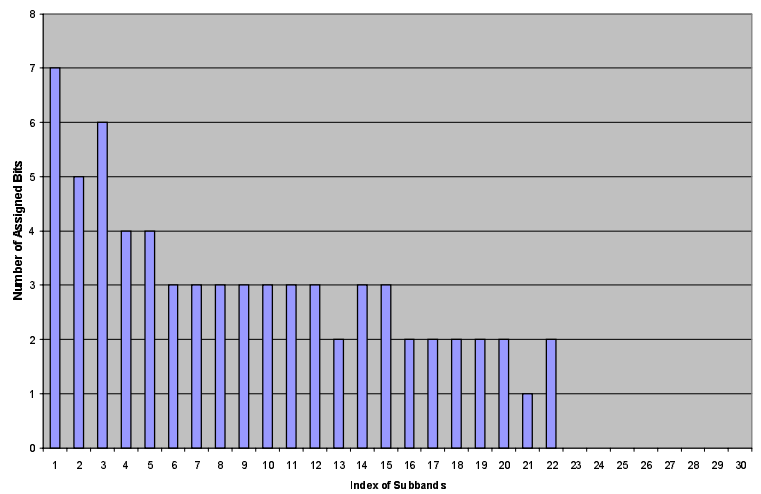
(a)



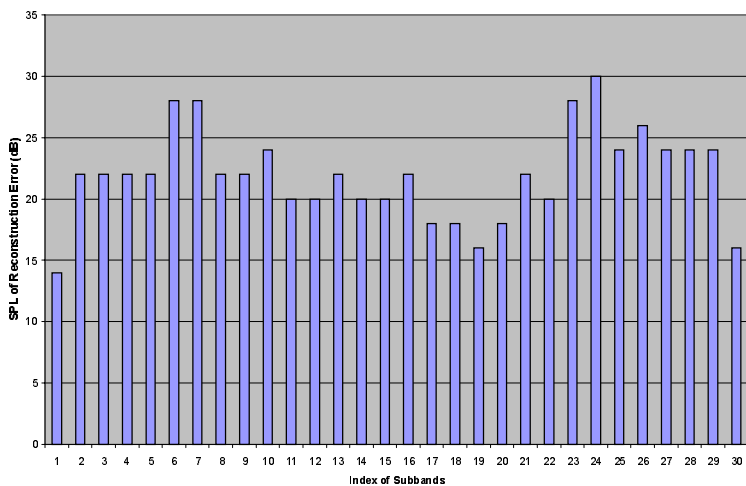
(b)



(c)



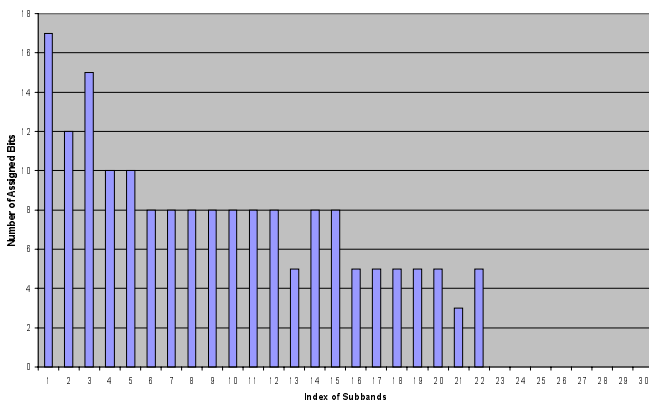
(d)



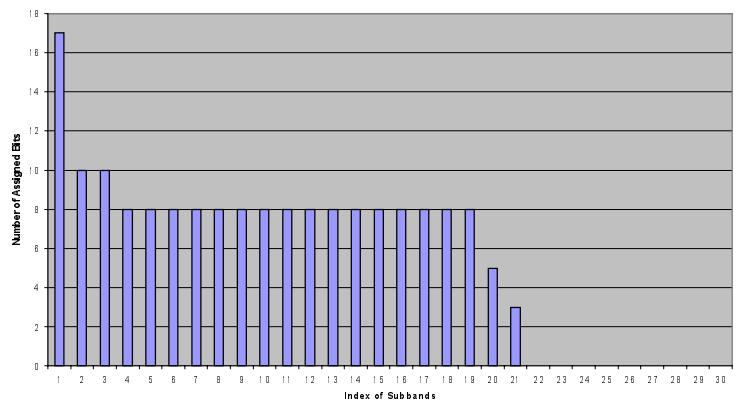
(e)

Σχήμα 11. Κατανομές συχνότητας για διάφορες σημαντικές MPEG παραμέτρους που λήφθηκαν από το ηχητικό μπλοκ του σχήματος 10.

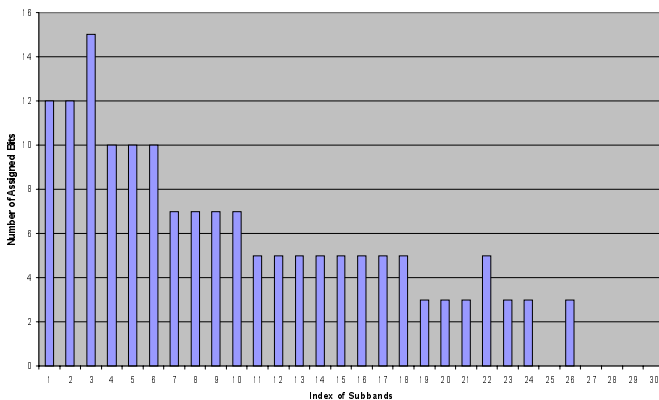
Στο σχήμα 12, τρεις καταμερισμοί bit συγκρίνονται για το ίδιο ακουστικό μπλοκ χρησιμοποιώντας (α) το ψυχοακουστικό μοντέλο 1 του προτύπου (β) το ψυχοακουστικό μοντέλο 2 του προτύπου και (γ) τον χωρίς βάρος καταμερισμό bit. Σε αυτό το παράδειγμα, υπάρχουν καθαρές διαφορές μεταξύ των δύο μοντέλων που προτάθηκαν στο MPEG πρότυπο. Σημειώστε ότι ο υπολογισμός του μη ζυγισμένου καταμερισμού bit δεν εξαρτάται από τα κατώφλια απόκρυψης. Πάντως, ο καταμερισμός bit μοιάζει με εκείνο του μοντέλου 1, εκτός ότι ο μη ζυγισμένος καταμερισμός bit ξοδεύει 3 bits για υποζώνες με ενδείξεις 233 και πάνω όπου η ισχύς σήματος είναι πολύ κάτω από το κατώφλι απόκρυψης.



(a)



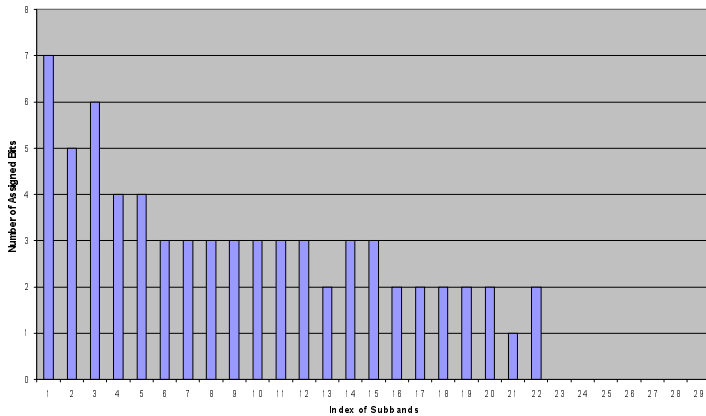
(b)



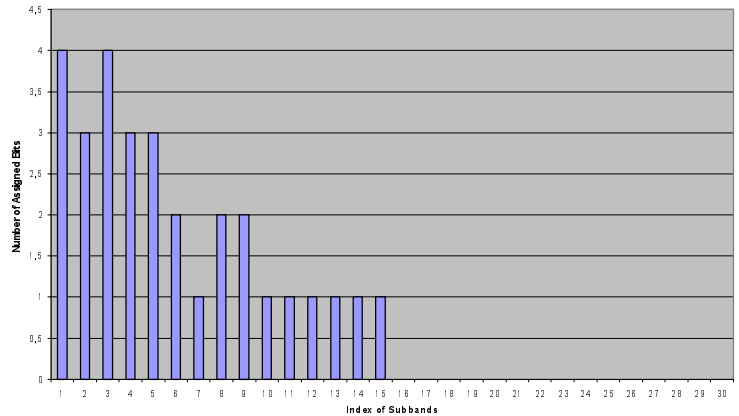
(c)

Σχήμα 12. Καταμερισμοί bit για τρεις κανόνες καταμερισμού που λήφθηκαν από το ηχητικό μπλοκ του Σχήματος 10 για MPEG-1 Στρώματος II κωδικοποίηση με ολικό ρυθμό bit 128 kb/s. (a) Καταμερισμός bit (μοντέλο 1) (b) Καταμερισμός bit (μοντέλο 2) (c) Καταμερισμός bit για μη ζυγισμένο ελάχιστο μέσο τετραγωνικό σφάλμα

Τελικά στο σχήμα 13, συγκρίνουμε τους καταμερισμούς bit του μοντέλου 1 για ρυθμούς bit των 128 kb/s και 64 kb/s πάλι για το ίδιο ακουστικό μπλοκ. Σημειώστε ότι στο χαμηλότερο ρυθμό bit μία χαμηλοπερατή εκδοχή του ακουστικού σήματος ανακατασκευάζεται.



(a)



(b)

Σχήμα 13. Καταμερισμοί bit για MPEG-1 Στρώματος II κωδικοποίηση με ολικούς ρυθμούς bit 128 kb/s και 64 kb/s

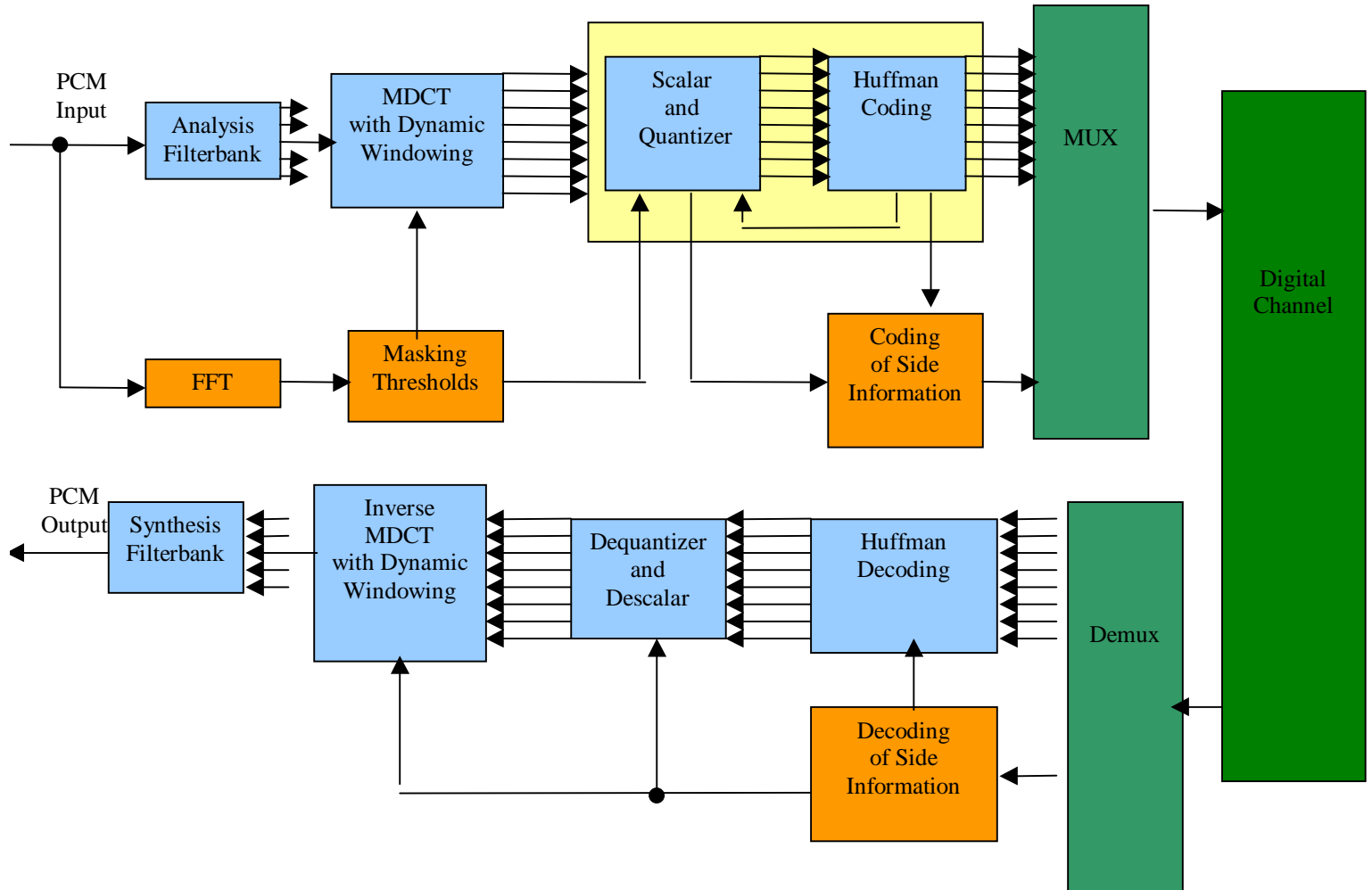
- *Αποκωδικοποίηση*

Η αποκωδικοποίηση είναι άμεση. Οι σειρές της υποζώνης ανακατασκευάζονται με βάση μπλοκ των 12 δειγμάτων υποζώνης, λαμβάνοντας υπόψη τον αποκωδικοποιημένο συντελεστή κλίμακας και την πληροφορία καταμερισμού bit. Αν μία υποζώνη δεν έχει καταμερισμό bit, τα δείγματα σε αυτή την υποζώνη μηδενίζονται. Κάθε φορά που τα δείγματα υποζώνης από όλες τις 32 υποζώνες έχουν υπολογιστεί, κατευθύνονται στην τράπεζα φίλτρων της σύνθεσης και 32 συνεχόμενα 16-bit PCM ακουστικά δείγματα υπολογίζονται. Αν είναι δυνατό, όπως στις αμφίδρομες τηλεπικοινωνίες ή στα συστήματα εγγραφής, η τράπεζα φίλτρων του κωδικοποιητή (ανάλυση) μπορεί να χρησιμοποιηθεί με ανάποδο τρόπο λειτουργίας στη διαδικασία αποκωδικοποίησης.

### 3.4 Στρώμα III

Το Στρώμα III του MPEG-1 προτύπου κωδικοποίησης ήχου εισάγει πολλά νέα χαρακτηριστικά (βλ. Σχ΄. 14), συγκεκριμένα μία υβριδική τράπεζα φίλτρων με δυνατότητα μεταπήδησης. Επιπλέον χρησιμοποιεί μία προσέγγιση ανάλυση-από-σύνθεση, ένα εξελιγμένο έλεγχο προ-ηχούς και μία μη ενιαία κβάντιση με κωδικοποίηση εντροπίας. Μία τεχνική buffer, η οποία καλείται

δεξαμενή bit (bit reservoir), οδηγεί σε παραπέρα εξοικονομήσεις στο ρυθμό bit. Το Στρώμα III είναι το μοναδικό Στρώμα που παρέχει υποχρεωτική (mandatory) υποστήριξη του αποκωδικοποιητή για μεταβλητό ρυθμό bit κωδικοποίησης.

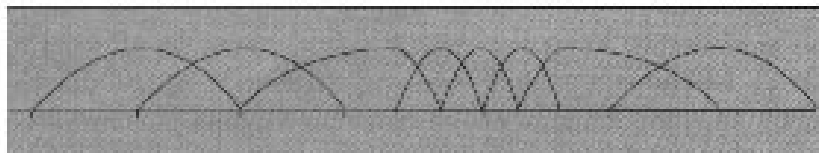


Σχήμα 14. Δομή του MPEG-1 κωδικοποιητή και αποκωδικοποιητή ήχου (Στρώμα III)

- *Εναλλασσόμενη Υβριδική Τράπεζα φίλτρων*

Για να επιτύχουμε μία υψηλότερη ανάλυση συχνότητας πιο κοντά στις διαμερίσεις κρίσιμης ζώνης, τα 32 σήματα υποζώνης υποδιαιρούνται παραπέρα σε περιεχόμενο συχνότητας, εφαρμόζοντας σε κάθε υποζώνη ένα τροποποιημένο DCT μπλοκ μετασχηματισμό 6 σημείων ή 18 σημείων με 50% επικάλυψη. Γι' αυτό το λόγο τα παράθυρα περιέχουν αντίστοιχα, 12 ή 36 δείγματα υποζώνης. Ο μέγιστος αριθμός περιεχομένων συχνότητας είναι  $32 \times 18 = 576$ , καθένα αναπαριστώντας ένα εύρος ζώνης μόνο  $24000/576 = 41.67$  Hz. Ο μπλοκ μετασχηματισμός 18 σημείων εφαρμόζεται κανονικά διότι παρέχει καλύτερη

ανάλυση συχνότητας, ενώ ο μπλοκ μετασχηματισμός 6 σημείων παρέχει καλύτερη ανάλυση χρόνου και εφαρμόζεται στην περίπτωση αναμενόμενων προ-ηχών (βλέπε προηγούμενη ενότητα για μεταπήδηση παραθύρου). Εξ αρχής μία προ-ηχώ θεωρείται όταν συμβαίνει μία στιγμιαία απαίτηση για ένα μεγάλο αριθμό bit. Ανάλογα της φύσης των ενδεχόμενων προ-ηχών όλοι ή ένας μικρότερος αριθμός των μετασχηματισμών μεταπηδούν. Δύο ειδικά MDCT παράθυρα, ένα παράθυρο εκκίνησης και ένα παράθυρο τερματισμού χρειάζονται στην περίπτωση μεταβάσεων μεταξύ μικρών και μεγάλων μπλοκ και ανάποδα έτσι ώστε να διατηρήσουν το γνωστό στο πεδίο του χρόνου ως χαρακτηριστικό ακύρωσης του MDCT [21,24,36]. Το σχήμα 15 δείχνει μία τυπική σειρά από παράθυρα.



Σχήμα 15. Τυπική σειρά παραθύρων σε προσαρμοστική μεταπήδηση παραθύρου

- *Κβάντιση και Κωδικοποίηση*

Τα MDCT δείγματα εξόδου κβαντίζονται ανομοιόμορφα, παρέχοντας έτσι και μικρότερα σφάλματα μέσω των τετραγώνων και απόκρυψης διότι μεγαλύτερα σφάλματα μπορούν να ανεχθούν εάν είναι μεγάλα τα δείγματα που πρόκειται να κβαντιστούν. Η κωδικοποίηση Huffman, βασισμένη σε 32 κωδικούς πίνακες και επιπρόσθετη κωδικοποίηση κατά την εφαρμογή, εφαρμόζεται στην αναπαράσταση των ενδείξεων του κβαντιστή με ικανοποιητικό τρόπο. Ο κωδικοποιητής απεικονίζει τις μεταβλητού μήκους κωδικές λέξεις των κωδικών πινάκων Huffman σε ένα σταθερό ρυθμό bit παρακολουθώντας την κατάσταση μιας δεξαμενής bit. Η δεξαμενή bit εξασφαλίζει ότι ο buffer του αποκωδικοποιητή ούτε υποχειλίζει ούτε υπερχειλίζει όταν ένα ρεύμα bit παρουσιάζεται στον αποκωδικοποιητή με ένα σταθερό ρυθμό.

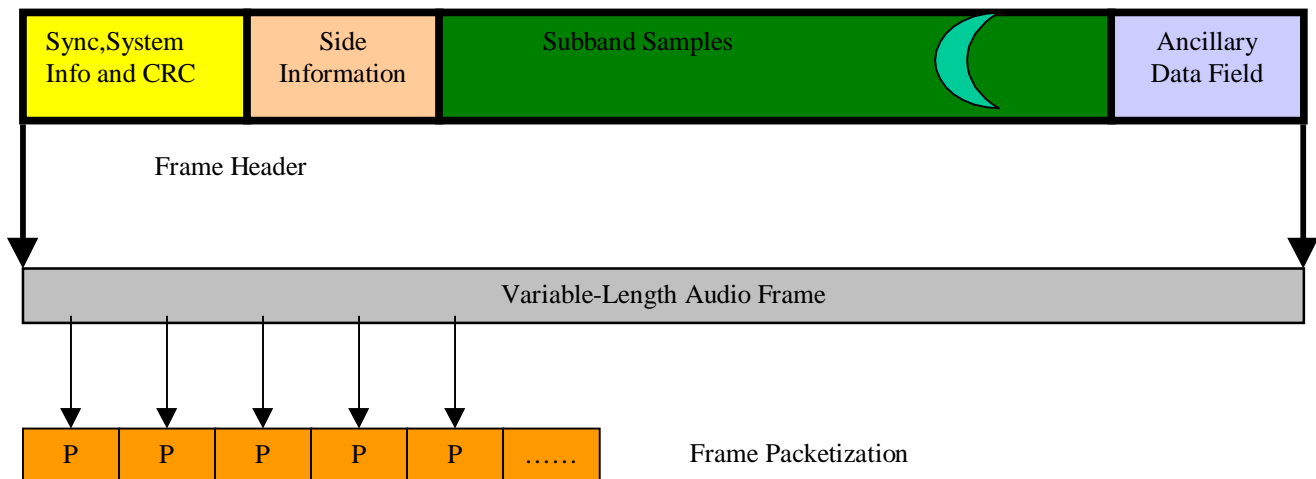
Για να διατηρήσουμε το θόρυβο κβάντισης σε όλες τις κρίσιμες ζώνες κάτω από το γενικό κατώφλι απόκρυψης (καταμερισμός θορύβου) μια επαναληπτική μέθοδος ανάλυσης-από- σύνθεση χρησιμοποιείται με την οποία η διαδικασία της κλιμάκωσης, κβάντισης και κωδικοποίησης των φασματικών δεδομένων ολοκληρώνεται μέσα σε δύο εμφωλιασμένους βρόχους επανάληψης. Η αποκωδικοποίηση ακολουθεί ότι και η διαδικασία κωδικοποίησης.



### 3.5 Δομή πλαισίου και πολύπλοκη δομή

- Δομή πλαισίου

Το σχήμα 16 δείχνει τη δομή πλαισίου του MPEG-1 κωδικοποιημένων σημάτων ήχου για τα δύο Στρώματα I και II. Κάθε πλαίσιο έχει μία επικεφαλίδα. Το πρώτο του μέρος περιέχει 12 bits συγχρονισμού , 20 bits

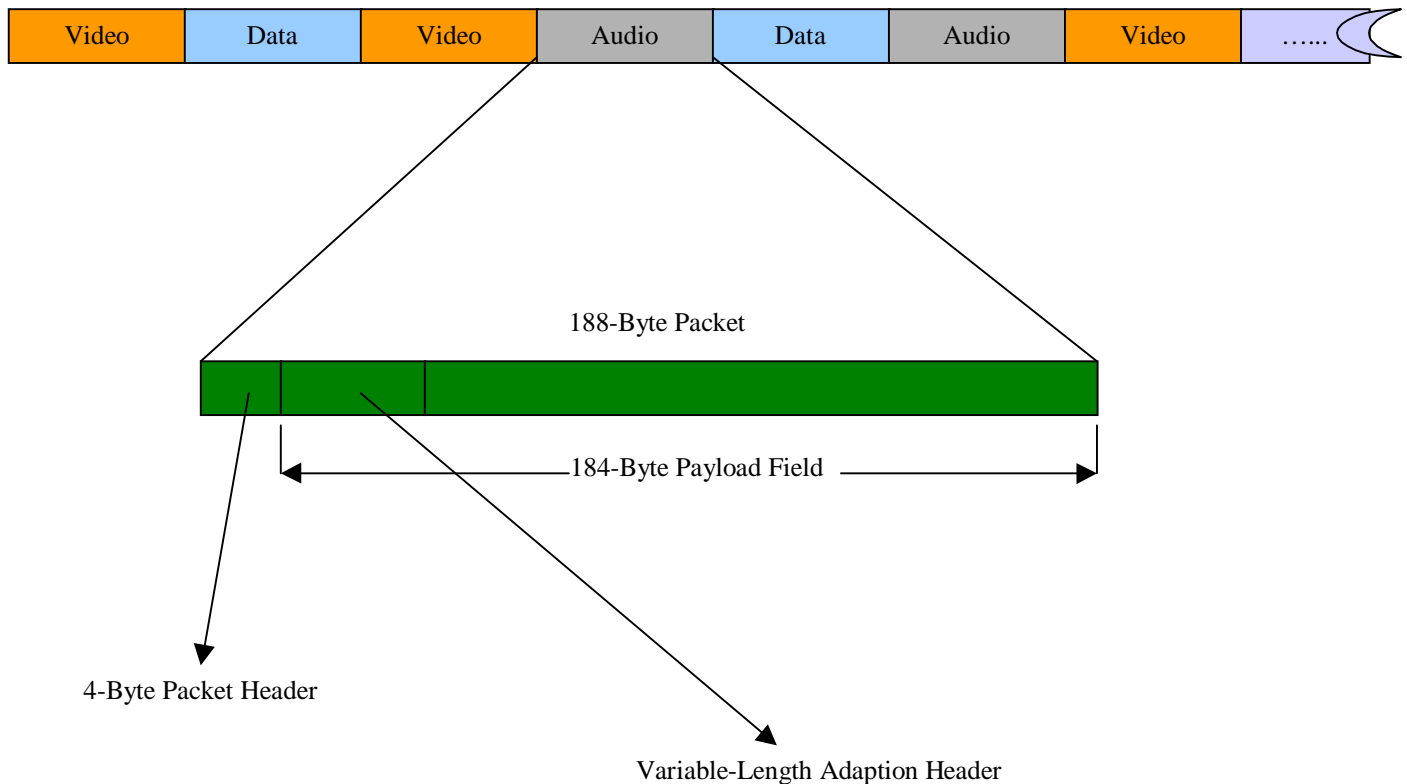


Σχήμα 16. MPEG-1 δομή πλαισίου και πακετάρισμα. Στρώμα I: 384 δείγματα υποζώνης. Στρώμα II : 1152 δείγματα υποζώνης. Πακέτα P: 4 byte επικεφαλίδα, 184 byte πεδίο φορτίου

πληροφορία συστήματος και προαιρετικό 16 bit κυκλικό κώδικα ελέγχου πλεονασμού. Το δεύτερο του μέρος περιέχει πλευρική πληροφορία γύρω από τον καταμερισμό bit και τους συντελεστές κλίμακας (και στο Στρώμα II πληροφορία επιλογής συντελεστή κλίμακας). Ως κύρια πληροφορία ένα πλαίσιο μεταφέρει ένα σύνολο από 32 X 12 δείγματα υποζώνης (που αντιστοιχούν σε PCM ακουστική είσοδο 384 δειγμάτων - ισοδύναμο με 8 ms σε ένα ρυθμό δειγματοληψίας των 48 KHz). Σημειώστε ότι τα πλαίσια είναι αυτόνομα. Κάθε πλαίσιο περιέχει όλη την απαραίτητη πληροφορία για αποκωδικοποίηση. Επομένως κάθε πλαίσιο μπορεί να αποκωδικοποιηθεί ανεξάρτητα από προηγούμενα πλαίσια . Ορίζει ένα σημείο εισόδου για ακουστική αποθήκευση και εφαρμογες επεξεργασίας ήχου. Επίσης σημειώστε ότι τα μήκη των πλαισίων δεν είναι καθορισμένα εξαιτίας του (α) μήκους του πεδίου της κύριας πληροφορίας, το οποίο εξαρτάται στο ρυθμό bit και συχνότητα δειγματοληψίας (β) του πεδίου πλευρικής πληροφορίας το οποίο μεταβάλλεται στο Στρώμα II και (γ) του πεδίου βοηθητικών δεδομένων το μήκος του οποίου δεν καθορίζεται.

- *Πολύπλοκη Δομή*

Το τμήμα συστημάτων του MPEG-1 προτύπου κωδικοποίησης IS11172 ορίζει μία δομή πακέτου για πολυπλεξία ήχου, εικόνας και ρευμάτων bit βοηθητικών δεδομένων σε ένα ρεύμα. Τα μεταβλητού μήκους MPEG πλαίσια σπάνε σε πακέτα. Η δομή πακέτου χρησιμοποιεί πακέτα των 188 byte που αποτελούνται από μια επικεφαλίδα 4 byte ακολουθούμενη από 184 byte ωφέλιμου φορτίου (βλ. Σχ. 17).



*Σχήμα 17. MPEG παράδοση πακέτου*

Η επικεφαλίδα περιλαμβάνει ένα byte συγχρονισμού, ένα πεδίο 13 bit που καλείται αναγνωριστής πακέτου ώστε να πληροφορήσει τον αποκωδικοποιητή για τον τύπο των δεδομένων και επιπλέον πληροφορία. Για παράδειγμα, μία μονάδα ωφέλιμου φορτίου εκκίνησης του δείκτη του 1 bit υποδεικνύει εάν το ωφέλιμο φορτίο ξεκινά με μία επικεφαλίδα πλαισίου. Καμία προκαθορισμένη μίξη ήχου, εικόνας και ρευμάτων bit βοηθητικών δεδομένων δεν απαιτείται. Η μίξη μπορεί να αλλάξει δυναμικά. Υπηρεσίες παρέχονται με ένα πολύ εύκαμπτο τρόπο. Αν επιπλέον πληροφορία επικεφαλίδας απαιτείται όπως για περιοδικό συγχρονισμό ακουστικού και οπτικού χρονισμού μία μεταβλητού μήκους

επικεφαλίδα προσαρμογής μπορεί να χρησιμοποιηθεί ως μέρος του πεδίου ωφέλιμου φορτίου των 184 byte.

Αν και τα μήκη των πλαισίων δεν είναι καθορισμένα, το διάστημα μεταξύ επικεφαλίδων πλαισίου είναι σταθερό (μέσα σε ένα byte) καθ' όλη τη χρήση των bytes συμπλήρωσης. Ο ορισμός των MPEG συστημάτων περιγράφει πως MPEG συμπιεσμένα ρεύματα δεδομένων ήχου και εικόνας πρόκειται να πολυπλεχθούν μαζί ώστε να σχηματίσουν ένα μονό ρεύμα δεδομένων. Η ορολογία και οι θεμελιώδεις αρχές των συστημάτων στρώματος περιγράφονται στο [38]

### **3.6 Υποκειμενική ποιότητα (MPEG-1;Στερεοφωνικά Ηχητικά Σήματα)**

Η διαδικασία δημιουργίας προτύπου περιελάμβανε εκτεταμένες υποκειμενικές δοκιμές και αντικειμενικούς υπολογισμούς παραμέτρων όπως πολυπλοκότητα και ολική καθυστέρηση. Οι MPEG (και ισοδύναμα ITU-R) ακουστικές δοκιμές έγιναν κάτω από πολύ όμοιες και προσεκτικά ορισμένες συνθήκες με περίπου 60 έμπειρους ακροατές, προσεγγιστικά 10 σειρές δοκιμών πραγματοποιήθηκαν και οι περίοδοι εκτελέστηκαν στερεοφωνικά και με ηχεία και με ακουστικά. Για να εντοπίσουμε ακόμα πιο μικρές ατέλειες η 5 σημείων ITU-R κλίμακα ατελειών χρησιμοποιήθηκε σε όλα τα πειράματα. Λεπτομέρειες δίνονται στο [39] και [40]. Κρίσιμα είδη δοκιμών επιλέχθηκαν στις δοκιμές για να υπολογίσουμε τους κωδικοποιητές στην χειρότερη περίπτωση (όχι μέση) απόδοσης. Οι υποκειμενικοί υπολογισμοί οι οποίοι βασίστηκαν σε τυφλά πειράματα τριών ερεθισμάτων, άγνωστης αναφοράς (stimulus/hidden reference/double blind tests), έδειξαν πολύ όμοια και σταθερά αποτελέσματα εκτίμησης. Σε αυτές τις δοκιμές το υποκείμενο ακούει τρία σήματα, A, B και C (τριπλό ερέθισμα). Το A είναι πάντα το ακατέργαστο σήμα πηγής ( η αναφορά) ,B και C, ή C και B είναι η αναφορά με την οποία το σύστημα δοκιμάζεται (κρυμμένη αναφορά) Η επιλογή δεν είναι γνωστή ούτε στα υποκείμενα ούτε στους «μαέστρους» (conductors) της δοκιμής (διπλά τυφλή δοκιμή). Τα υποκείμενα πρέπει να αποφασίσουν εάν το B ή C είναι η αναφορά και πρέπει να βαθμολογήσουν αυτή που απομένει. Το MPEG-1 πρότυπο κωδικοποίησης ήχου έχει δείξει εξαιρετική απόδοση για όλα τα στρώματα στους ρυθμούς που δίνονται στον Πίνακα 3. Θα πρέπει να σημειωθεί πάλι ότι το πρότυπο αφήνει χώρο για βελτιώσεις βασισμένες στον κωδικοποιητή με τη χρήση καλύτερων ψυχοακουστικών μοντέλων. Και πραγματικά, πολλές βελτιώσεις έχουν επιτευχθεί από τότε που υλοποιήθηκαν οι πρώτες υποκειμενικές δοκιμές το 1991.

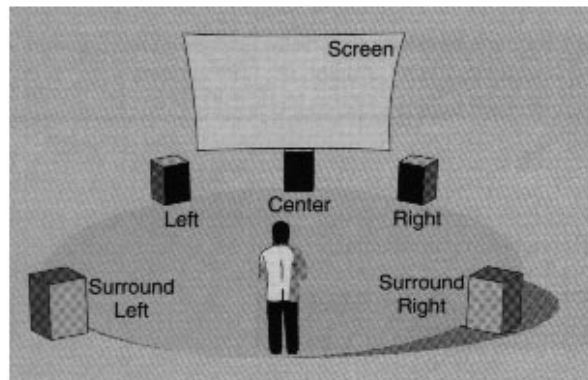
#### **4. MPEG ΑΚΟΥΣΤΙΚΗ ΚΩΔΙΚΟΠΟΙΗΣΗ ΜΕ ΧΑΜΗΛΟΤΕΡΕΣ ΣΥΧΝΟΤΗΤΕΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ**

Έχουμε αναφέρει πιο πάνω ότι το MPEG-1/ήχου υποστηρίζει συχνότητες δειγματοληψίας των 32, 44.1 και 48 KHz. Για εφαρμογές με περιορισμένα εύρη ζώνης (μεσαία ζώνη), έχουν οριστεί χαμηλότερες συχνότητες δειγματοληψίας (16, 22.05 και 24 KHz) στο MPEG-2 για να χαμηλώσουν τους ρυθμούς σε 64 kb/s ανά κανάλι και λιγότερο [9]. Τα αντίστοιχα μέγιστα ακουστικά εύρη ζώνης είναι 7.5, 10.3 και 11.25 KHz. Η σύνταξη, η σημασιολογία και οι τεχνικές κωδικοποίησης του MPEG-1 διατηρούνται εκτός από ένα μικρό αριθμό παραμέτρων (δύο πίνακες στον αποκωδικοποιητή). Επομένως, η κωδικοποίηση μπορεί να στηριχθεί πάλι στα Στρώματα I, II και III. Η επέκταση σε χαμηλότερες συχνότητες δειγματοληψίας οδηγεί σε υψηλότερες αναλύσεις συχνότητας και επομένως σε υψηλότερο κέρδη κωδικοποίησης, εν μέρει εξαιτίας καλύτερων ενσωματώσεων στα κατώφλια απόκρυψης και εν μέρει διότι η πλευρική πληροφορία γίνεται ένα μικρότερο μέρος του ολικού ρυθμού bit. Όπως στην περίπτωση κωδικοποίησης ακουστικών σημάτων ευρείας ζώνης, η καλύτερη ακουστική ποιότητα αποκτάται με το Στρώμα III. Τέλος σημειώνουμε ότι μερικές εφαρμογές χρησιμοποιούν συχνότητες δειγματοληψίας των 8, 11.025 και 12 KHz, οι οποίες είναι έξω από το πρότυπο MPEG-2.

#### **5. MPEG ΚΩΔΙΚΟΠΟΙΗΣΗ ΠΟΛΥΚΑΝΑΛΙΚΟΥ ΗΧΟΥ**

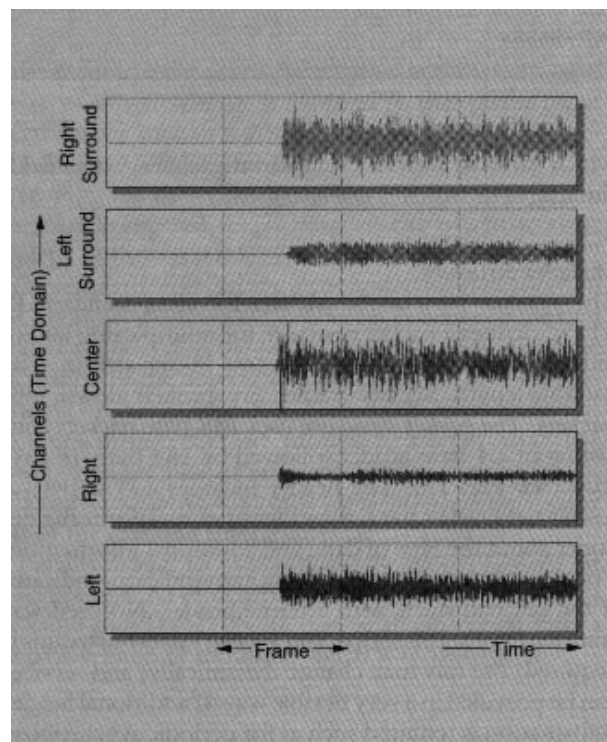
##### **5.1 Αναπαραστάσεις πολυκαναλικού ήχου**

Ένα λογικό επόμενο βήμα στο ψηφιακό ήχο είναι ο ορισμός των πολυκαναλικών ηχητικών συστημάτων αναπαραστάσης, ώστε να δημιουργήσουμε ένα ρεαλιστικό πεδίο περιβάλλοντα ήχου (surround) και για ηχητικές μόνο εφαρμογές και για οπτικοακουστικά συστήματα, συμπεριλαμβανομένου τηλεσυνδιάσκεψης (video conferencing), οπτροφωνίας (videophony), υπηρεσίες πολυμέσων και ηλεκτρονικός κινηματογράφος. Τα πολυκαναλικά συστήματα μπορούν επίσης να παρέχουν πολυφωνικά κανάλια ή επιπρόσθετα κανάλια για οπτικά εξασθενημένα (μία λεκτική περιγραφή του οπτικού τοπίου) και για ακουστικά εξασθενημένα (διάλογος με επαυξημένη κατανοητότητα). Το ITU-R και άλλες διεθνείς ομάδες έχουν προτείνει μία διαμόρφωση ηχείων 5 καναλιών, η οποία αναφέρεται ως 3/2-στέρεο, με ένα αριστερό και ένα δεξιό (L και R), ένα επιπρόσθετο κεντρικό κανάλι (C) και δύο πλευρικά οπίσθια περιβάλλοντα κανάλια (LS και RS) ενισχύοντας τα L και R κανάλια (βλ. Σχ. 18) (ITU-R Rec 775).



Σχήμα 18. 3/2 Πολυκαναλική διαμόρφωση ηχείων

Μία τέτοια διαμόρφωση προσφέρει ένα πεδίο περιβάλλοντα ήχου με μία σταθερή μπροστινή ηχητική εικόνα και μία μεγάλη ακουστική περιοχή. Το Σχήμα 19 δείχνει τέσσερα μπλοκ ενός πεντακάναλου τριγωνικού ηχητικού σήματος ( το οποίο είναι δύσκολο να κωδικοποιηθεί).



Σχήμα 19. Τριγωνική αναπαράσταση ήχου σε πέντε κανάλια (από την κορυφή:  $RS, LS, C, R, L$ )

Τα πολυκαναλικά συστήματα ψηφιακού ήχου υποστηρίζουν p/q αναπαραστάσεις με p μπροστινά και q οπίσθια κανάλια και επίσης παρέχουν τις πιθανότητες μετάδοσης δύο ανεξάρτητων στερεοφωνικών προγραμμάτων και/ή ένα αριθμό καναλιών πολυφωνικών ή παροχής σχολίων. Τυπικοί συνδυασμοί καναλιών δίνονται στον Πίνακα 4.

Μερικές πολυκαναλικές διαμορφώσεις ηχείων		
1 Κανάλι	1/0	Κέντρο (μονοφωνικό)
2 Κανάλια	2/0	Αριστερά, δεξιά (στερεοφωνικό)
3 Κανάλια	3/0	Αριστερά, δεξιά, κέντρο
4 Κανάλια	3/1	Αριστερά, δεξιά, κέντρο, μονοπεριβάλλον
5 Κανάλια	3/2	Αριστερά, δεξιά, κέντρο, περιβάλλον αριστερά, περιβάλλον δεξιά

ΠΙΝΑΚΑΣ 4.

Η ITU-R πρόταση 775 παρέχει ένα σύνολο προς τα κάτω αναμιγμένες εξισώσεις εάν ο αριθμός των ηχείων πρόκειται να μειωθεί (προς τα κάτω συμβατότητα- downwards compatibility). Μία επιπρόσθετη επαύξηση καναλιού χαμηλής συχνότητας (LFE ή subwoofer), ιδιαίτερα χρήσιμη για HDTV εφαρμογές, μπορεί προαιρετικά να προστεθεί σε οποιαδήποτε διαμόρφωση. Το LFE κανάλι επεκτείνει το περιεχόμενο χαμηλής συχνότητας μεταξύ 15 Hz και 120 Hz σε όρους και συχνότητας και επιπέδου. Ένα ή περισσότερα ηχεία μπορεί να τοποθετηθεί ελεύθερα στο δωμάτιο για να αναπαραγάγει αυτό το LFE σήμα. Η βιομηχανία ταινιών χρησιμοποιεί ένα παρόμοιο σύστημα για τα δικά της συστήματα ψηφιακού ήχου. Μία 3/2 διαμόρφωση με πέντε υψηλής ποιότητας πλήρους περιοχής κανάλια συν ένα κανάλι subwoofer συχνά καλείται 5.1 σύστημα.

Για να μειώσουμε τον ολικό ρυθμό bit πλυκαναλικών συστημάτων ηχητικής κωδικοποίησης, πλεονασμοί και μη σχετικότητα, όπως εσωκαναλικές εξαρτήσεις και εσωτερικές επιδράσεις απόκρυψης, αντίστοιχα, μπορούν να χρησιμοποιηθούν. Επιπροσθέτως, συστατικά του πολυκαναλικού σήματος, τα οποία είναι άσχετα με σεβασμό στη χωρική αντίληψη του της στερεοφωνικής παρουσίασης, δηλαδή, εκείνα που δεν συνεισφέρουν στον εντοπισμό των ηχητικών πηγών, μπορούν να αναγνωρισθούν και να αναπαραχθούν σε μία μονοφωνική μορφοποίηση σε παραπέρα μείωση ρυθμών bit. Σύγχρονοι αλγόριθμοι πολυκαναλικής κωδικοποίησης κάνουν χρήση τέτοιων επιδράσεων.

Πάντως, απαιτείται μία προσεχτική σχεδίαση, αλλιώς τέτοια συνδυασμένη κωδικοποίηση μπορεί να παράγει ατέλειες.

## 5.2 MPEG-2/Ήχου πολυκαναλική κωδικοποίηση

Η δεύτερη φάση του MPEG, χαρακτηρισμένη ως MPEG-2, περιλαμβάνει στο κομμάτι του ήχου δύο πολυκαναλικά πρότυπα κωδικοποίησης ήχου, καθένα από τα οποία είναι με πρόδρομα και αναδρομικά συμβατό με το MPEG-1/Ήχου [8,41-44]. Πρόδρομα συμβατό σημαίνει ότι ένας MPEG-2 πολυκαναλικός αποκωδικοποιητής είναι ικανός να αποκωδικοποιήσει σωστά MPEG-1 μονό ή στερεοφωνικά σήματα. Ανάδρομα συμβατό σημαίνει ότι οι υπάρχοντες MPEG-1 στερεοφωνικοί αποκωδικοποιητές, οι οποίοι χειρίζονται μόνο ήχο δύο καναλιών, είναι ικανοί να αναπαράγουν ένα με νόημα βασικό 2/0 στερεοφωνικό σήμα από ένα MPEG-2 πολυκαναλικό ρεύμα bit έτσι ώστε να εξυπηρετήσουν τις ανάγκες των χρηστών με απλό μονό ή στερεοφωνικό εξοπλισμό. Μη ανάδρομα συμβατοί πολυκαναλικοί κωδικοποιητές δε θα είναι ικανοί να τροφοδοτήσουν ένα με νόημα ρεύμα bit σε ένα MPEG-1 στερεοφωνικό αποκωδικοποιητή. Από την άλλη πλευρά, έχουν περισσότερη ελευθερία στην παραγωγή μίας υψηλής ποιότητας αναπαραγωγής ηχητικών σημάτων.

Με αναδρομική συμβατότητα είναι δυνατό να εισάγουμε πολυκαναλικό ήχο οποιαδήποτε στιγμή με ένα ομαλό τρόπο χωρίς να θέσουμε σε αχρηστία στερεοφωνικούς αποκωδικοποιητές δύο καναλιών. Ένα σημαντικό παράδειγμα είναι το Ευρωπαϊκό Σύστημα Εκπομπής Ψηφιακού Ήχου, το οποίο θα απαιτήσει MPEG-1 στερεοφωνικούς αποκωδικοποιητές στην πρώτη γενιά αλλά μπορεί να προσφέρει πολυκαναλικό ήχο αργότερα.

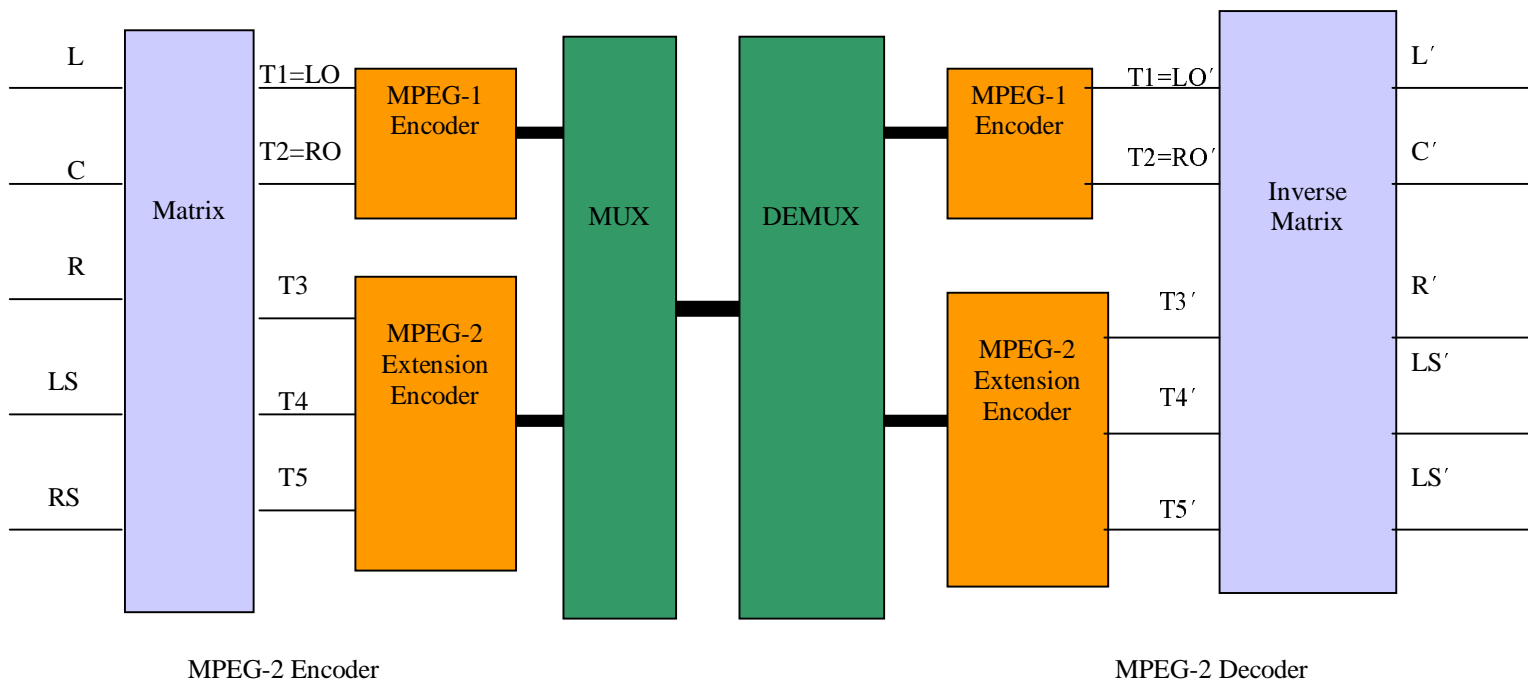
- *Αναδρομικά συμβατή MPEG-2 κωδικοποίηση ήχου.*

Η αναδρομική συμβατότητα υπονοεί τη χρήση πινάκων συμβατότητας. Μία κάτω-μίξη των πέντε καναλιών («matrixing») παραδίδει ένα σωστό βασικό 2/0 στερεοφωνικό σήμα, αποτελούμενο από ένα αριστερό και ένα δεξιό κανάλι, LO και RO αντίστοιχα. Ένα τυπικό σύνολο εξισώσεων είναι:

$$LO = \alpha*(L + \beta*C + \delta*LS) \quad \text{όπου } \alpha = 1/(1+\sqrt{2}) \text{ και } \beta=\delta=\sqrt{2}$$

$$RO = \alpha*(R + \beta*C + \delta*RS)$$

Άλλες επιλογές είναι δυνατές, συμπεριλαμβανομένου LO=L και RO=R. Οι συντελεστές  $\alpha$ ,  $\beta$  και  $\delta$  εξασθενούν τα σήματα για να αποφύγουμε υπερφόρτωση όταν υπολογίζουμε το συμβατό στερεοφωνικό σήμα (LO,RO). Τα LO και RO μεταδίδονται με μορφοποίηση MPEG-1 στα κανάλια μετάδοσης T1 και T2. Τα κανάλια T3, T4 και T5 σχηματίζουν μαζί το πολυκαναλικό σήμα επέκτασης (Σχ. 20). Αυτά πρέπει να επιλεγούν έτσι ώστε ο αποκωδικοποιητής να μπορεί να ξαναυπολογίσει το πλήρες 3/2 στερεοφωνικό πολυκαναλικό σήμα.

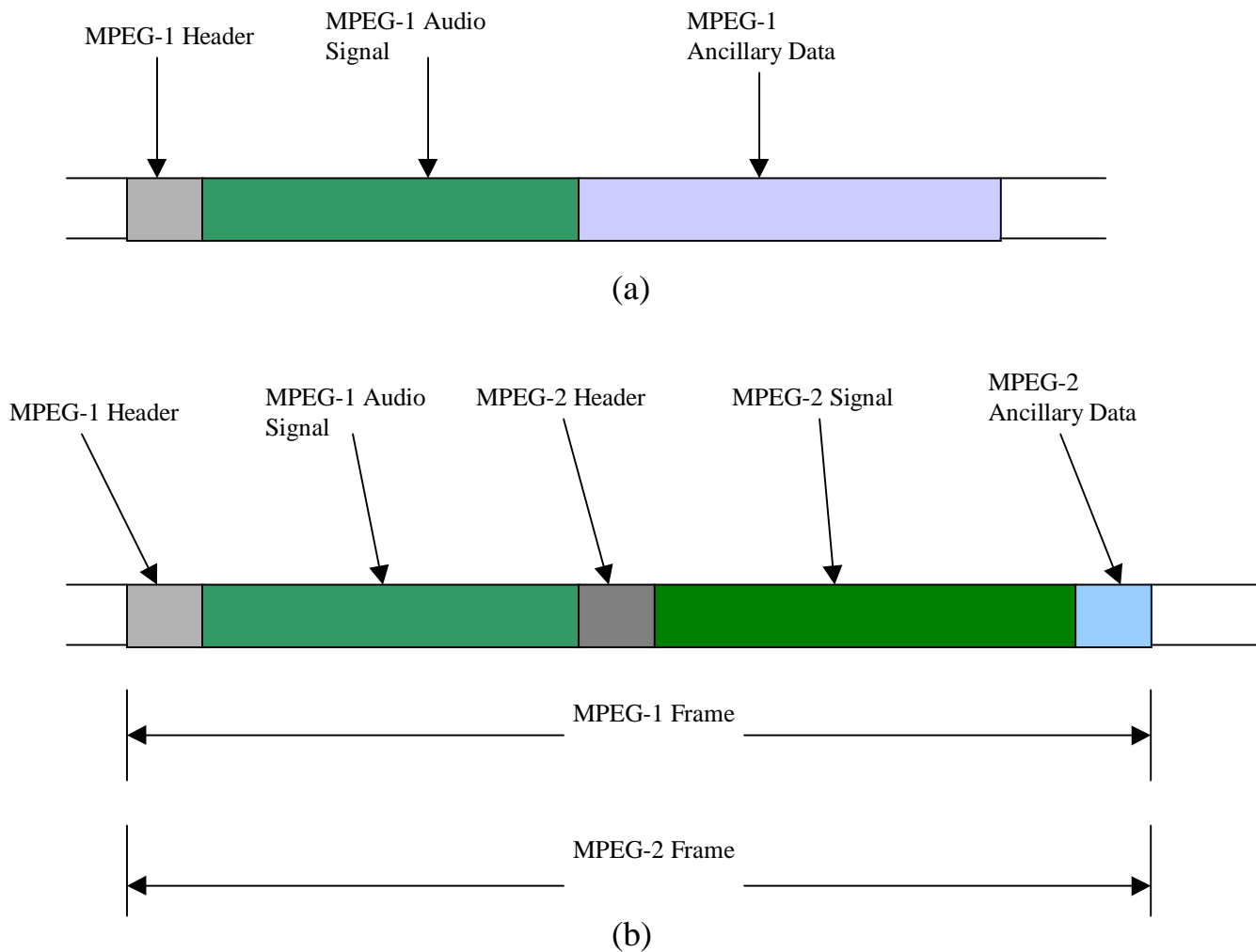


*Σχήμα 20. Συμβατότητα των MPEG-2 πολυκαναλικών ηχητικών ρευμάτων bit*

Εσωκαναλικοί πλεονασμοί και επιδράσεις απόκρυψης λαμβάνονται υπόψη για να βρούμε την καλύτερη επιλογή. Ένα απλό παράδειγμα είναι  $T3=C, T4=LS$ , και  $T5=RS$ . Στο MPEG-2 η πινακοποίηση μπορεί να γίνει με ένα πολύ εύκαμπτο και ακόμα χρονικά ανεξάρτητο τρόπο. Σημειώστε πάντως ότι το περιεχόμενο ήχου του σήματος επέκτασης έχει κιάλας παραδοθεί στο MPEG-1 ηχητικό ρεύμα (σήματα LO και RO). Αυτός ο πλεονασμός μειώνει τον πραγματοποιήσιμο ρυθμό συμπίεσης.

Η αναδρομική συμβατότητα επιτυγχάνεται μεταδίδοντας τα κανάλια LO και RO στην ενότητα δείγματος υποζώνης του MPEG-1 πλαισίου ήχου και όλα τα πολυκαναλικά σήματα επέκτασης ( $T3, T4$  και  $T5$ ) στο πρώτο τμήμα του MPEG-1 πλαισίου δεσμεύονται για βοηθητικά δεδομένα. Αυτό το πεδίο βοηθητικών δεδομένων αγνοείται από τους από τους MPEG-1 αποκωδικοποιητές (βλ. Σχ. 21). Το μήκος του πεδίου βοηθητικών δεδομένων δεν καθορίζεται στο πρότυπο. Αν ο αποκωδικοποιητής είναι τύπου MPEG-1, χρησιμοποιεί την 2/0 μορφοποίηση αριστερού και δεξιού κάτω μίξης σήματος,  $LO'$  και  $RO'$  απευθείας (βλ. Σχ. 22). Αν ο αποκωδικοποιητής είναι τύπου MPEG-2, ξαναυπολογίζει το πλήρες 3/2-στερεοφωνικό πολυκαναλικό σήμα από τα  $L', R', C', LS'$ , και  $RS'$  μέσω «αποπινακοποίησης» του  $LO', RO', T3', T4'$  και  $T5'$  (βλ. Σχ. 20).

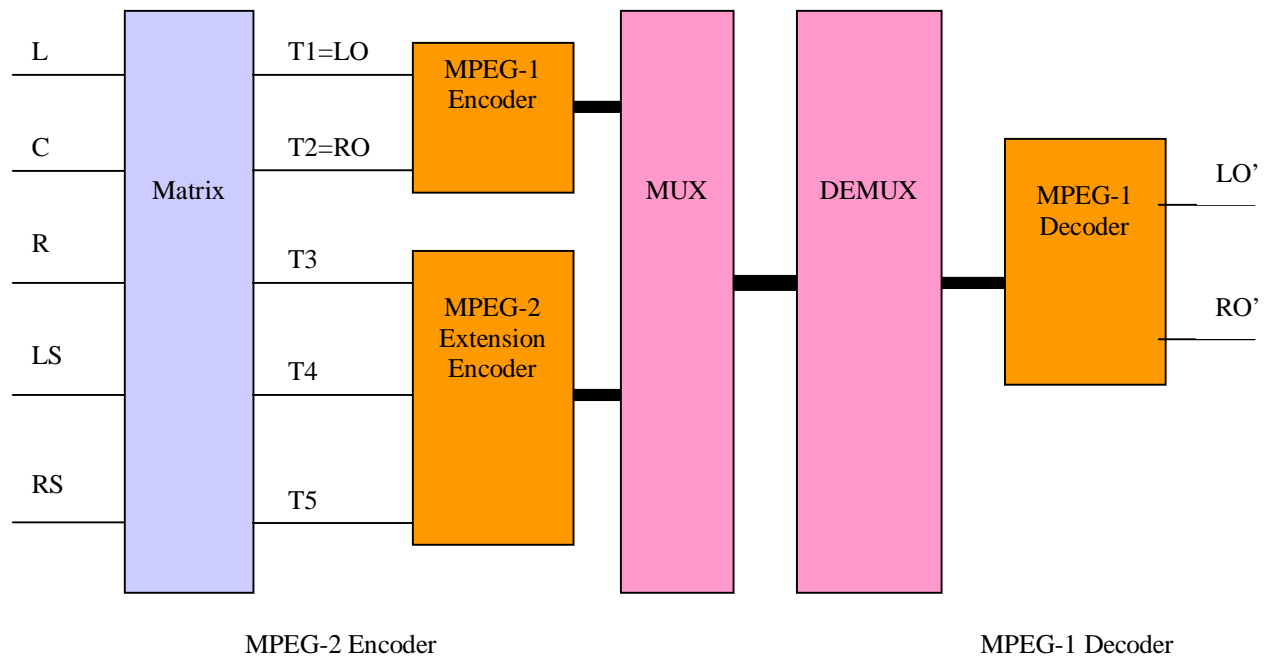




Σχήμα 21. Μορφοποίηση δεδομένων των MPEG ηχητικών ρευμάτων bit.  
 (a) MPEG-1 ηχητικό πλαίσιο (b) MPEG-2 ηχητικό πλαίσιο, συμβατό με  
 MPEG-1 μορφοποίηση

Η πινακοποίηση είναι προφανώς απαραίτητη για να παράσχει αναδρομική συμβατότητα. Πάντως, αν χρησιμοποιηθεί σε σύνδεση με κωδικοποίηση αντίληψης, μπορεί να εμφανιστεί αποκάλυψη του θορύβου κβάντισης [45]. Μπορεί να προκληθεί στη διαδικασία αποπινακοποίησης όταν σχηματιστούν τα σήματα αθροίσματος και αφαίρεσης. Σε ορισμένες περιπτώσεις ένα τέτοιο συστατικό σήμα απόκρυψης αθροίσματος ή διαφοράς μπορεί να εξαφανιστεί σε ένα ορισμένο κανάλι. Εφόσον αυτό το συστατικό υποτίθεται ότι θα απέκρυπτε το θόρυβο κβάντισης σε αυτό το κανάλι, αυτός ο θόρυβος μπορεί να γίνει αντιληπτός ακουστικά. Σημειώστε ότι το σήμα απόκρυψης θα είναι ακόμα παρών στην πολυκαναλική αναπαραστάση, αλλά θα εμφανιστεί σε διαφορετικό

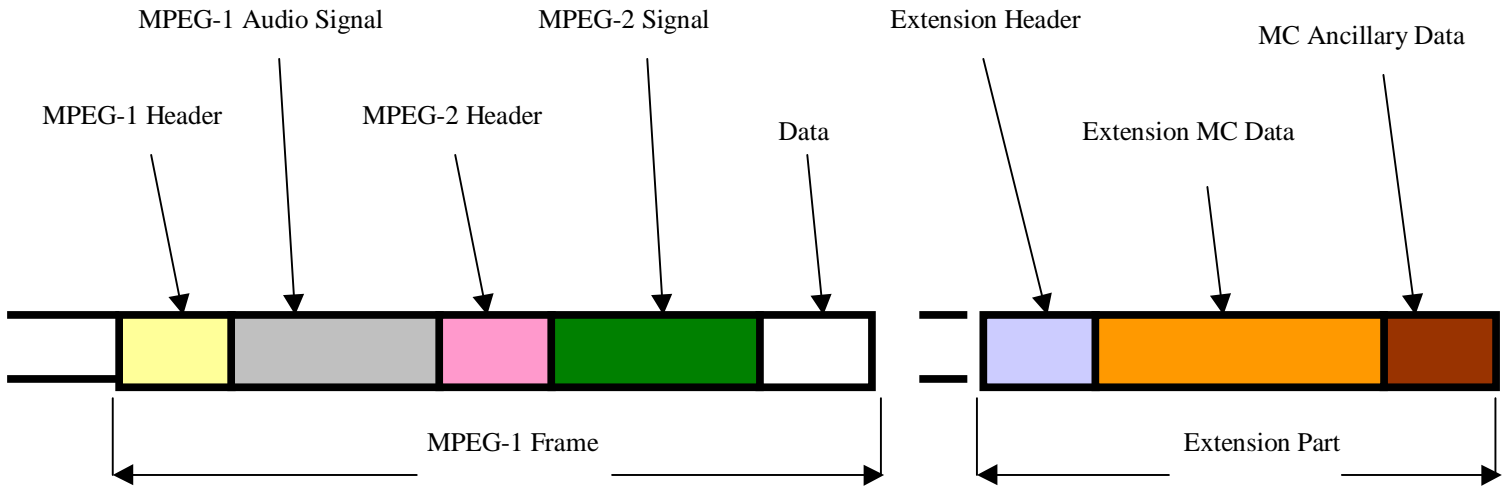
ηχείο. Μέτρα κατά των αποτελεσμάτων της αποκάλυψης περιγράφονται στο [46]. Ως ένα επιπρόσθετο μέτρο η προαιρετική διαμόρφωση μεταβλητού ρυθμού bit του MPEG-2, μπορεί να ανακληθεί να κωδικοποιησει



*Σχήμα 22. MPEG-1 στέρεο αποκωδικοποίηση του MPEG-2 πολυκαναλικού ρεύματος bit*

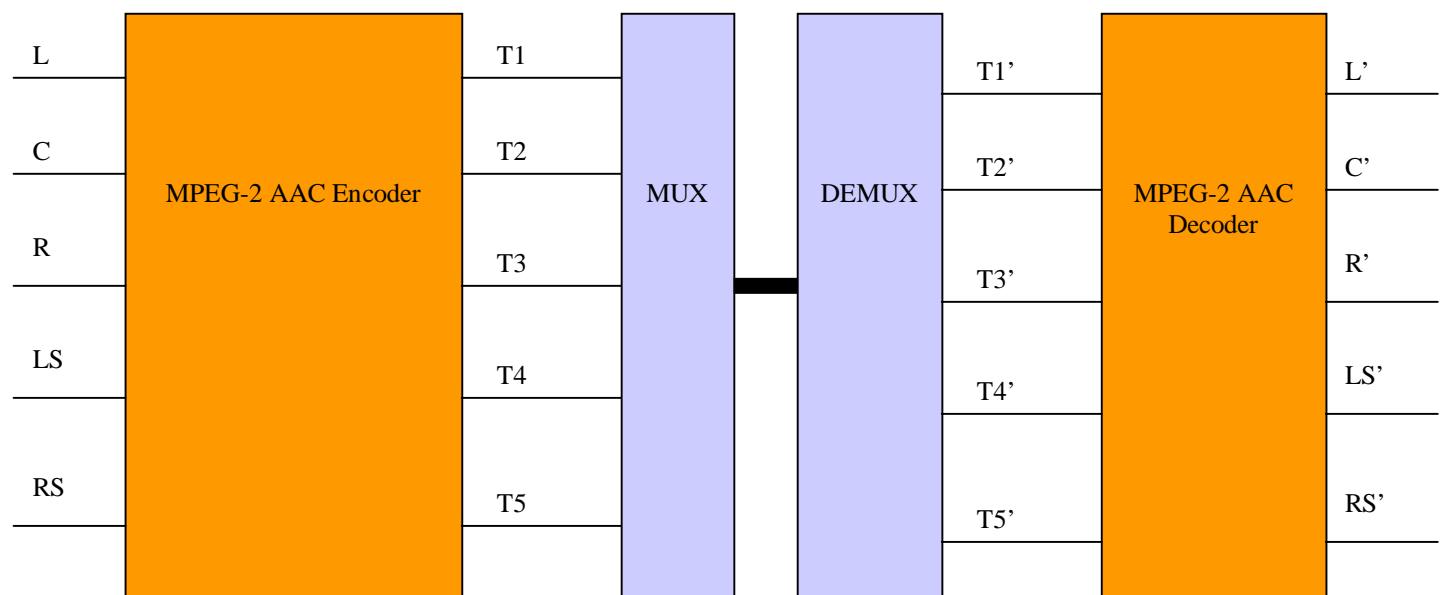
το δύσκολο ηχητικό περιεχόμενο σε ένα στιγμιαία μεγαλύτερο ρυθμό bit.

Οι MPEG-1 αποκωδικοποιητές έχουν ένα περιορισμό στο ρυθμό bit (384 kb/s στο Στρώμα II). Για να ξεπεράσουμε αυτό τον περιορισμό, το πρότυπο MPEG-2 επιτρέπει για ένα δεύτερο ρεύμα bit, το τμήμα επέκτασης, να παρέχει συμβατό πολυκαναλικό ήχο σε υψηλότερους ρυθμούς. Το σχήμα 23 δείχνει τη δομή του ρεύματος bit με επέκταση.



Σχήμα 23. Διαμόρφωση δεδομένων του MPEG-2 ηχητικού ρεύματος bit με τμήμα επέκτασης για πολυκαναλικά δεδομένα

- MPEG-2 προχωρημένη κωδικοποίηση ήχου  
 Ένα δεύτερο πρότυπο μέσα στο MPEG-2 υποστηρίζει εφαρμογές που δεν απαιτούν συμβατότητα με την υπάρχουσα MPEG-1 στερεο διαμόρφωση. Επομένως δεν είναι απαραίτητες η πινακοποίηση και η αποπινακοποίηση και οι αντίστοιχες πιθανές ατέλειες εξαφανίζονται. (βλ. Σχ. 24)

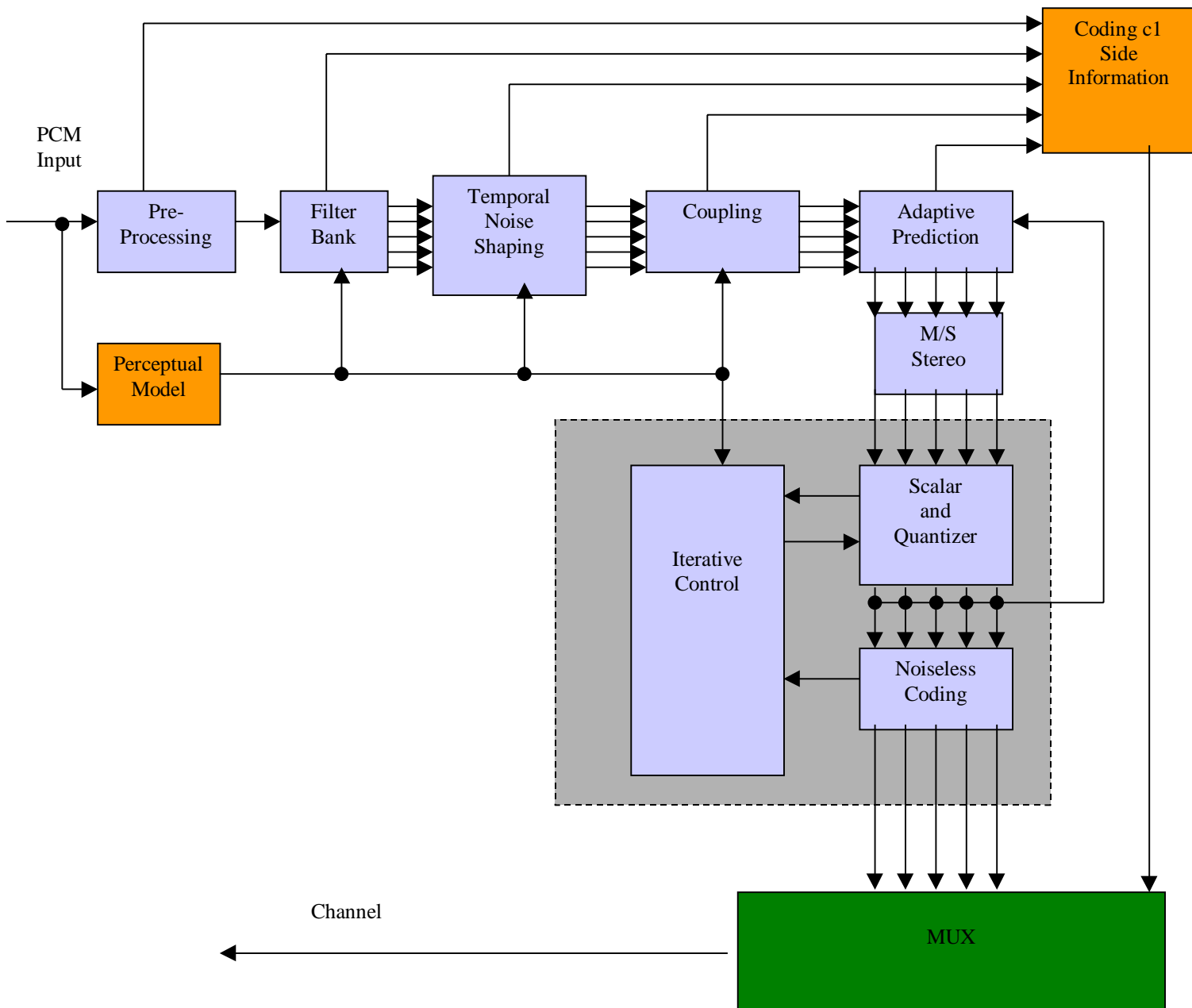


Σχήμα 24. MPEG-2 προχωρημένη κωδικοποίηση ήχου (πολυκαναλική διαμόρφωση)

Τα τελευταία δύο χρόνια έχουμε δει εκτενείς δραστηριότητες στην τελειοποίηση και καθιέρωση προτύπου ενός μη αναδρομικά συμβατού MPEG-2 πολυκαναλικού αλγόριθμου κωδικοποίησης ήχου. Πολλές εταιρίες σε όλο τον κόσμο πρόσφεραν προχωρημένους αλγόριθμους κωδικοποίησης ήχου σε μία προσπάθεια συνεργασίας να προκύψει ένα εύκαμπτο πρότυπο κωδικοποίησης υψηλής ποιότητας [43].

**Εργαλεία.** Το MPEG-2 AAC πρότυπο απασχολεί τράπεζες φίλτρων υψηλής ανάλυσης, τεχνικές πρόβλεψης και κωδικοποίηση για λιγότερο θόρυβο. Βασίζεται σε πρόσφατους υπολογισμούς και ορισμούς εργαλείων (ή μέτρα), καθένα το οποίο έχει επιλεγεί από ένα αριθμό προτεινόμενων. Τα αυτό-περιεχόμενα εργαλεία (self-contained) περιλαμβάνουν μία προαιρετική προεπεξεργασία, μία τράπεζα φίλτρων, ένα μοντέλο αντίληψης, προσωρινή σχηματοποίηση θορύβου, πολυκαναλική κωδικοποίηση έντασης, πρόβλεψη, M/S στέρεο κωδικοποίηση, κβάντιση, κωδικοποίηση χωρίς θόρυβο και ένα πολυπλέκτη ρεύματος bit (βλ. Σχ. 25). Η τράπεζα φίλτρων είναι ένας 1024 γραμμικός τροποποιημένος διακριτός μετασχηματισμός συνημιτόνου και το μοντέλο αντίληψης λαμβάνεται από το MPEG-1 (μοντέλο 2). Το εργαλείο προσωρινής σχηματοποίησης θορύβου ελέγχει τη χρονική εξάρτηση από το θόρυβο κβάντισης, την ένταση και M/S κωδικοποίηση και ο δεύτερης τάξης αναδρομικά προσαρμοστικός προβλεπτής βελτιώνει την επάρκεια κωδικοποίησης. Ο προβλεπτής μειώνει το ρυθμό bit για τα μεταγενέστερα δείγματα υποζώνης σε μία δοσμένη υποζώνη και στηρίζει την πρόβλεψη του στο κβαντισμένο φάσμα του προηγούμενου μπλοκ, το οποίο είναι επίσης διαθέσιμο στον αποκωδικοποιητή (απουσία σφαλμάτων καναλιού). Τελικά για κβάντιση και κωδικοποίηση χωρίς θόρυβο, μία επαναληπτική μέθοδος εφαρμόζεται για να διατηρήσει το θόρυβο κβάντισης σε όλες τις κρίσιμες κάτω από το γενικό κατώφλι απόκρυψης.

**Κατατομές.** Για να εξυπηρετήσουμε διαφορετικές ανάγκες, το πρότυπο παρέχει τρεις κατατομές : (α) η κύρια κατατομή προσφέρει την υψηλότερη ποιότητα, (β) η χαμηλής πολυπλοκότητας κατατομή δουλεύει χωρίς πρόβλεψη και (γ) η κλιμακωτού ρυθμού δειγματοληψίας κατατομή προσφέρει τη χαμηλότερη πολυπλοκότητα. Για παράδειγμα, στην κύρια κατατομή η τράπεζα φίλτρων είναι ένας 1024 γραμμικός MDCT με 50% επικάλυψη (μήκος μπλοκ 2048 δείγματα). Η τράπεζα φίλτρων μπορεί να μεταπηδά σε οκτώ 128 γραμμικά MDCTs (μήκη μπλοκ 256 δείγματα). Γι' αυτό επιτρέπει μία ανάλυση συχνότητας των 23.43 Hz και μία ανάλυση χρόνου των 2.6 ms (και τα δύο σε ένα ρυθμό δειγματοληψίας των 48 KHz). Στην περίπτωση ενός μακρού μήκους μπλοκ το σχήμα του παραθύρου μπορεί να ποικίλει δυναμικά ως μία συνάρτηση του σήματος.



Σχήμα 25. Δομή του MPEG-2 εξελιγμένου κωδικοποιητή ήχου (AAC)

Η χαμηλής πολυπλοκότητας κατατομή δεν απασχολεί προσωρινή σχηματοποίηση θορύβου και πρόβλεψη στο πεδίο του χρόνου (η πρόβλεψη προσθέτει σημαντικά στην πολυπλοκότητα) ενώ στην κατατομή κλιμακωτού ρυθμού δειγματοληψίας χρησιμοποιείται μία υβριδική τράπεζα φίλτρων.

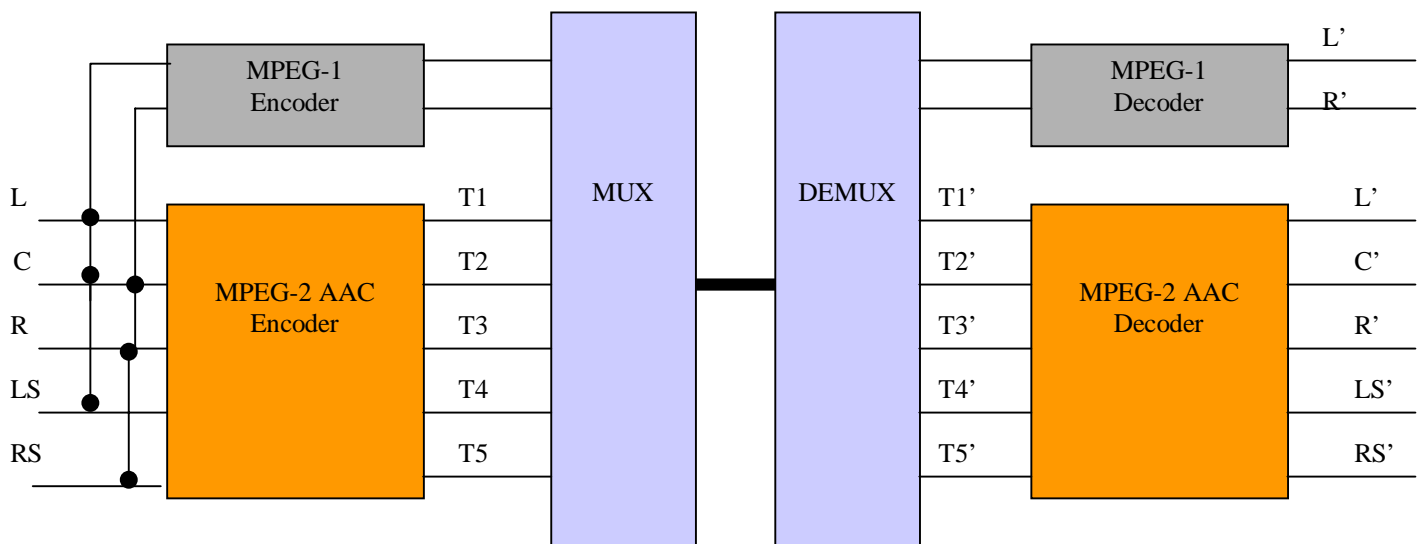
Το MPEG-2 AAC υποστηρίζει μέχρι 46 κανάλια για διάφορες πολυκαναλικές διαμορφώσεις ηχείων και άλλες εφαρμογές. Οι εξ' ορισμού διαμορφώσεις ηχείων είναι το μονοφωνικό κανάλι, το στερεοφωνικό κανάλι και το 5.1 σύστημα (πέντε κανάλια συν LFE κανάλι).

Τα παραπάνω αναφερθέντα επιλεγμένα μέτρα ορίζουν το MPEG-2 AAC πρότυπο ήχου, το οποίο έγινε Διεθνές Πρότυπο τον Απρίλιο του 1997 ως μία

επέκταση του MPEG-2 (ISO/MPEG 13818-7). Μία πιο λεπτομερής περιγραφή του MPEG-2 AAC πολυκαναλικού προτύπου μπορεί να βρεθεί στην βιβλιογραφία [43]. Το πρότυπο προσφέρει υψηλή ποιότητα στους χαμηλότερους δυνατούς ρυθμούς bit μετξύ 320 και 384 kb/s για πέντε κανάλια. Θα βρει πολλές εφαρμογές και για καταναλωτική και για επαγγελματική χρήση.

### 5.3 Αναδρομική συμβατότητα μέσω ταυτόχρονης εκπομπής

Αν οι ρυθμοί των bit δεν μας απασχολούν ιδιαίτερα, μία μετάδοση simulcast μπορεί να χρησιμοποιηθεί, όπου ένα πλήρες MPEG-1 ρεύμα bit πολυπλέκεται με ένα πλήρες μη αναδρομικά συμβατό πολυκαναλικό ρεύμα bit για να υποστηρίξει συμβατότητα προς τα πίσω χωρίς τεχνικές πινακοποίησης (Σχ. 26).



Σχήμα 26. Αναδρομικά συμβατή MPEG-2 πολυκαναλική κωδικοποίηση ήχου (simulcast mode)

### 5.4 Υποκειμενικές δοκιμές (MPEG-2, πολυκαναλικά σήματα ήχου)

Οι πρώτες υποκειμενικές δοκιμές, που πραγματοποιήθηκαν ανεξάρτητα στον Γερμανικό Οργανισμό Τηλεπικοινωνιών και στο BBC κάτω από την προστασία της MPEG-2 διαδικασίας καθιέρωσης προτύπων, έδειξαν μία ικανοποιητική μέση απόδοση των μη αναδρομικά συμβατών και αναδρομικά συμβατών κωδικοποιητών. Οι δοκιμές πραγματοποιήθηκαν με έμπειρο ακουστικό κοινό και κρίσιμα είδη δοκιμών σε χαμηλούς ρυθμούς (320 και 384 kb/s). Πάντως, όλοι τα συστήματα κωδικοποιητών/αποκωδικοποιητών έδειξαν σημαντικές αποκλίσεις από τη διαφάνεια για μερικά από τα είδη δοκιμών

[47,48]. Πρόσφατα ,εκτεταμένες τυπικές υποκειμενικές δοκιμές έχουν πραγματοποιηθεί για να συγκρίνουν διάφορες εκδόσεις του MPEG-2 AAC,σε λειτουργία, αντίστοιχα στα 256 και 320 kb/s και ενός αναδρομικά συμβατού MPEG-2 κωδικοποιητή Στρώματος II,σε λειτουργία στα 640 kb/s [49] (μία έκδοση του 1995 αυτού του τελευταίου κωδικοποιητή χρησιμοποιήθηκε, πάντως τα αποτελέσματα της δοκιμής του δεν απόδωσαν καθόλου μεταγενέστερες βελτιώσεις). Όλοι οι κωδικοποιητές απέδωσαν πολύ καλά με ένα ελαφρό πλεονέκτημα του μη αναδρομικά συμβατού 320 kb/s MPEG-2 AAC κωδικοποιητή συγκρινόμενο με τον αναδρομικά συμβατό 640 kb/s MPEG-2 Στρώματος II κωδικοποιητή. Οι αποδόσεις αυτών των κωδικοποιητών είναι μη διακρίσιμες από το πρωτότυπο με την έννοια του EBU ορισμού της ποιότητας μη διακρισιμότητας [50]. Από αυτές τις υποκειμενικές δοκιμές, έγινε φανερό ότι η ιδέα της αναδρομικής συμβατότητας υπονοεί τη χρήση υψηλότερων ρυθμών bits.

## **6. MPEG-4 ΚΩΔΙΚΟΠΟΙΗΣΗ ΗΧΟΥ**

### **6.1 Εισαγωγικά**

Οι δραστηριότητες μέσα στο MPEG-4 στοχεύουν σε προτάσεις για ένα ευρύ πεδίο δραστηριοτήτων συμπεριλαμβανομένου τα πολυμέσα. (Σημειώστε ότι το MPEG έχει αρχίσει ένα νέο είδος εργασίας το οποίο καλείται MPEG-7 (- Multimedia content description interface). Το MPEG-7 δεν καλύπτει κωδικοποίηση. Η επιτυχία του είναι περισσότερο να προσδιορίσει μία πρότυπη περιγραφή διαφόρων ειδών πληροφορίας πλυμέσων. Μία συνηθισμένη εφαρμογή θα είναι η έρευνα για οπτικό, γραφικό ή ηχητικό υλικό με την έννοια των σημερινών μηχανισμών εύρεσης με βάση κείμενο στο παγκόσμιο δίκτυο WWW). Το MPEG-4 θα προσφέρει υψηλότερους ρυθμούς συμπίεσης και θα συγχωνεύσει όλη την περιοχή ήχου, από την υψηλής πιστότητας κωδικοποίηση ήχου και την κωδικοποίηση ομιλίας μέχρι τη συνθετική ομιλία και τον συνθετικό ήχο, υποστηρίζοντας εφαρμογές από ηχητικά συστήματα υψηλής πιστότητας μέχρι τα κινητής πρόσβασης τερματικά πολυμέσων. Για να αναπαραστήσουμε, να ολοκληρώσουμε και να ανταλλάξουμε κομμάτια οπτικοακουστικής πληροφορίας, το MPEG-4 προσφέρει πρότυπα εργαλεία τα οποία μπορούν να συνδυαστούν για να ικανοποιήσουν συγκεκριμένες απαιτήσεις του χρήστη [51]. Ένας αριθμός τέτοιων διαμορφώσεων μπορεί να προτυποποιηθεί. Μία συντακτική περιγραφή θα χρησιμοποιηθεί για να μεταβιβάζει σε ένα αποκωδικοποιητή την επιλογή των εργαλείων που φτιάχτηκαν από τον κωδικοποιητή. Αυτή η περιγραφή μπορεί επίσης να χρησιμοποιηθεί για να περιγράψει νέους αλγόριθμους και να κατεβάσει τη διαμόρφωση τους στον επεξεργαστή αποκωδικοποίησης για εκτέλεση.

Το σημερινό σύνολο εργαλείων υποστηρίζει συμπίεση ήχου και ομιλίας σε μονοφωνικούς ρυθμούς bit που εκτείνονται από 2 μέχρι 64 kb/s.

Χρησιμοποιούνται τρεις κεντρικοί κωδικοποιητές:

- Ένα παραμετρικό σχήμα κωδικοποίησης για κωδικοποίηση ομιλίας σε χαμηλούς ρυθμούς bit (2 μέχρι 10 kb/s)
- Ένα ανάλυση από σύνθεση σχήμα κωδικοποίησης για μεσαίους ρυθμούς bit (6 μέχρι 16 kb/s)
- Ένα σχήμα κωδικοποίησης βασισμένο σε μετασχηματισμό υποζώνης για ρυθμούς bit κάτω από 64 kb/s.

Οι τρεις κεντρικοί κωδικοποιητές έχουν ολοκληρωθεί σε ένα αποκαλούμενο μοντέλο πιστοποίησης, το οποίο περιγράφει τις λειτουργίες των κωδικοποιητών και αποκωδικοποιητών και το οποίο χρησιμοποιείται για να διεκπεραιώσει προσομοιώσεις και εξιδανικεύσεις. Στο τέλος, το μοντέλο πιστοποίησης θα είναι προσωποποίηση του προτύπου [51].

Επίσης σημειώστε ότι το MPEG-4 θα προσφέρει νέες λειτουργικότητες όπως αλλαγές κλίμακας χρόνου έλεγχου τόνου φωνής (pitch), επεξεργασιμότητα, πρόσβαση σε βάσεις δεδομένων και ικανότητα για κλιμάκωση, η οποία επιτρέπει σε έναν να αφαιρέσει από το μεταδιδόμενο ρεύμα bit ένα υποσύνολο αρκετό για να ενεργοποιήσει ηχητικά σήματα με χαμηλότερο εύρος ζώνης και/ή χαμηλότερη ποιότητα ανάλογα με τη χωρητικότητα του καναλιού ή την πολυπλοκότητα του αποκωδικοποιητή. Το MPEG-4 θα γίνει ένα διεθνές πρότυπο το Νοέμβριο του 1998.

## 6.2 Σκοπός και χαρακτηριστικά του προτύπου MPEG-4

Γενικά το πρότυπο MPEG-4 θα παρέχει ένα σύνολο τεχνολογιών, ώστε να ικανοποιήσει τις ανάγκες των συγγραφέων, των παροχών υπηρεσιών και των χρηστών κατά τον ίδιο τρόπο.

- Για τους συγγραφείς, το MPEG-4 θα καταστήσει δυνατή την παραγωγή περιεχομένου, το οποίο έχει πολύ μεγαλύτερη ικανότητα επαναχρησιμοποίησης, έχει μεγαλύτερη ευελιξία από ότι είναι δυνατό σήμερα με ανεξάρτητες τεχνολογίες όπως ψηφιακή τηλεόραση, κινούμενα γραφικά, σελίδες διαδικτύου (WWW) και οι επεκτάσεις τους. Επίσης, θα είναι δυνατό να διαχειρίζονται και να προστατεύουν καλύτερα τα περιεχόμενα ιδιοκτησίας δικαιωμάτων.
- Για τους παροχείς υπηρεσιών δικτύων το MPEG-4 θα προσφέρει πεντάκαθαρη πληροφορία, η οποία θα μεταφράζεται και θα ερμηνεύεται στα κατάλληλα γηγενή μηνύματα σηματοδότησης καθενός δικτύου με τη βοήθεια σχετικών προτύπων σωμάτων. Ωστόσο το προηγούμενο δεν περιλαμβάνει θέματα της Ποιότητας των Υπηρεσιών, για τις οποίες θα παρέχει ένα γενικό περιγραφέα Ποιότητας Υπηρεσιών για διαφορετικά MPEG-4 μέσα. Οι ακριβείς μεταφράσεις από το σύνολο παραμέτρων



Ποιότητας Υπηρεσιών για κάθε μέσο στο δίκτυο είναι πέρα από το σκοπό του MPEG-4 και αφήνονται να οριστούν από τους παροχείς δικτύου.

- Για τους τελικούς χρήστες, το MPEG-4 θα επιτρέψει πολλές λειτουργίες οι οποίες μπορούν δυναμικά να προσπελαστούν σε ένα μονό συμπαγές τερματικό και σε υψηλότερα επίπεδα αλληλεπίδρασης με περιεχόμενο, μέσα στα όρια που ορίζονται από το συγγραφέα. Υπάρχει ένα κείμενο με εφαρμογές MPEG-4, το οποίο περιγράφει πολλές εφαρμογές τελικού χρήστη συμπεριλαμβανομένου, μεταξύ άλλων, επικοινωνίες σε πραγματικό χρόνο, επιτήρηση και κινητά πολυμέσα.

Το MPEG-4 επιτυχαίνει τους στόχους του παρέχοντας τυποποιημένους τρόπους για να:

1. Αναπαραστήσει μονάδες ακουστικού, οπτικού ή οπτικοακουστικού περιεχομένου, οι οποίες καλούνται «αντικείμενα μέσων» (“media objects”). Αυτά τα αντικείμενα μέσων μπορούν να είναι φυσικής ή συνθετικής προέλευσης. Αυτό σημαίνει ότι μπορούν να καταγραφούν με μία κάμερα ή μικρόφωνο, ή να δημιουργηθούν με ένα υπολογιστή.
2. Περιγράψει τη σύνθεση αυτών των αντικειμένων ώστε να δημιουργήσει σύνθετα αντικείμενα μέσων τα οποία σχηματίζουν οπτικοακουστικά σχήματα.
3. Πολυπλέξει και να συγχρονίσει τα δεδομένα που σχετίζονται με τα αντικείμενα μέσων, έτσι ώστε να μπορούν να μεταδοθούν μέσα από κανάλια δικτύων παρέχοντας μία Ποιότητα Υπηρεσίας (QoS) κατάλληλη για τη φύση των συγκεκριμένων αντικειμένων μέσων.
4. Αλληλεπιδρά με το οπτικοακουστικό σχήμα που δημιουργείται στο τέλος του δέκτη.

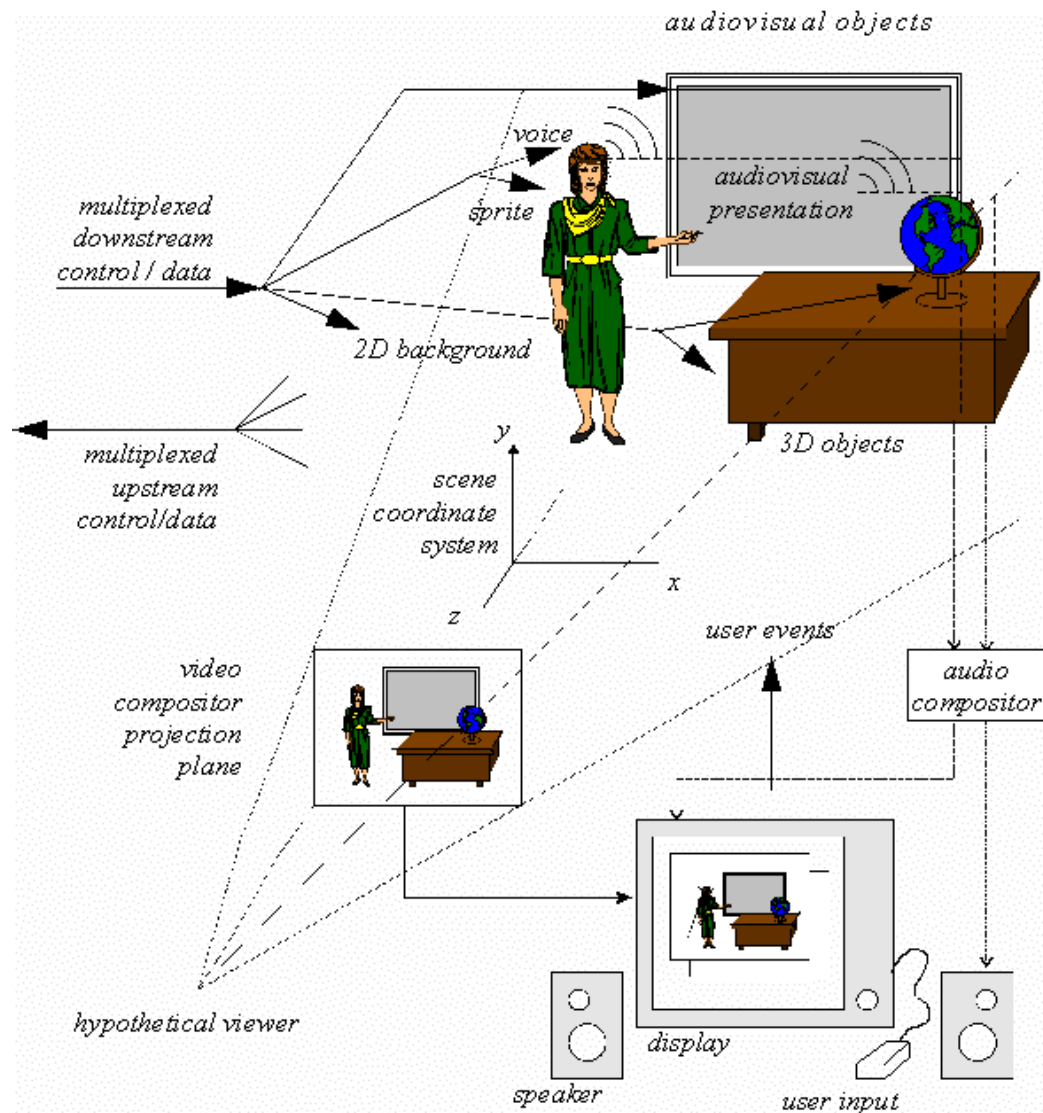
Οι ακόλουθες ενότητες απεικονίζουν τις λειτουργίες του MPEG-4 που περιγράφηκαν παραπάνω, χρησιμοποιώντας το οπτικοακουστικό σχήμα που φαίνεται στο σχήμα 27.

### **6.2.1 Κωδικοποιημένη αναπαράσταση των αντικειμένων μέσου**

Τα οπτικοακουστικά σχήματα συνθέτονται από πολλά αντικείμενα μέσων, τα οποία είναι οργανωμένα με ιεραρχικό τρόπο. Στα φύλλα της ιεραρχίας, βρίσκουμε τα πρωταρχικά αντικείμενα μέσων όπως:

- Ακίνητες εικόνες (π.χ. σταθεροποιημένο φόντο)
- Οπτικά αντικείμενα (π.χ. ένα άτομο που μιλάει - χωρίς το «βάθος»)
- Ακουστικά αντικείμενα (π.χ. η φωνή που σχετίζεται με ένα άτομο)

Το MPEG τυποποιεί ένα αριθμό τέτοιων πρωταρχικών αντικειμένων μέσων, ικανών να αναπαραστήσουν και φυσικού και συνθετικού περιεχομένου



Σχήμα 27. Παράδειγμα ενός MPEG-4 Σκηνικού

είδη, τα οποία μπορούν να είναι είτε δύο ή τριών διαστάσεων. Επιπλέον από τα αντικείμενα μέσω των οποίων αναφέρθηκαν παραπάνω και φαίνονται στο σχήμα 27, το MPEG-4 ορίζει την κωδικοποιημένη αναπαράσταση αντικειμένων όπως:

- Κείμενο και γραφικά
- Ομιλούντα συνθετικά κεφάλια και συσχετισμένα κείμενα που χρησιμοποιούνται να συνθέσουν την ομιλία και να κινήσουν το κεφάλι
- Συνθετικός ήχος

Ένα αντικείμενο μέσου στην κωδικοποιημένη του μορφή αποτελείται από περιγραφικά στοιχεία τα οποία επιτρέπουν να χειριζόμαστε το αντικείμενο σε ένα οπτικοακουστικό σχήμα εξίσου καλά από τα σχετικά δεδομένα ρεύματος, εάν χρειαστεί. Είναι σημαντικό να σημειώσουμε ότι στην κωδικοποιημένη του μορφή, κάθε αντικείμενο μέσου μπορεί να αναπαρασταθεί ανεξάρτητα από ότι το περιβάλλει ή το φόντο-βάθος.

Η κωδικοποιημένη αναπαράσταση των αντικειμένων μέσου είναι τόσο επαρκής όσο είναι δυνατό ενώ λαμβάνουμε υπόψη τις επιθυμητές λειτουργικότητες. Παραδείγματα τέτοιων λειτουργικότητων είναι η αντοχή σε σφάλματα, η εύκολη εξαγωγή και επεξεργασία ενός αντικειμένου, ή να έχουμε ένα αντικείμενο διαθέσιμο σε μία κλιμακούμενη μορφή.

### 6.2.2 Σύνθεση των αντικειμένων μέσου

Το σχήμα 27 δίνει ένα παράδειγμα που φωτίζει τον τρόπο κατά τον οποίο ένα οπτικοακουστικό σκηνικό σε MPEG-4 περιγράφεται καθώς συνθέτεται από ανεξάρτητα αντικείμενα. Το σχήμα περιέχει σύνθετα αντικείμενα μέσων τα οποία μαζί σχηματίζουν πρωταρχικά αντικείμενα μέσων. Τα πρωταρχικά αντικείμενα μέσων αντιστοιχούν σε φύλλα στο περιγραφικό δένδρο ενώ τα σύνθετα αντικείμενα μέσων περιστοιχίζουν ολόκληρα υποδένδρα. Ως ένα παράδειγμα: το οπτικό αντικείμενο που αντιστοιχεί στο άτομο που μιλάει και η αντίστοιχη φωνή είναι δεμένα μαζί για να σχηματίσουν ένα νέο σύνθετο αντικείμενο μέσου, περιέχοντας και τα ακουστικά και τα οπτικά συστατικά ενός ατόμου που μιλάει.

Μια τέτοια ομαδοποίηση επιτρέπει στους συγγραφείς να κατασκευάσουν σύνθετα σχήματα και καθιστά δυνατόν οι καταναλωτές να μεταχειρίζονται (σύνολα από) αντικείμενα τα οποία έχουν νόημα.

Πιο γενικά, το MPEG-4 παρέχει ένα τυποποιημένο τρόπο για να περιγραφεί ένα σκηνικό, επιτρέποντας για παράδειγμα να:

- Τοποθετήσουμε αντικείμενα μέσων οπουδήποτε μέσα σε ένα δοσμένο σύστημα συντεταγμένων
- Εφαρμόζουμε μετασχηματισμούς για να αλλάξουμε τη γεωμετρική ή ακουστική εμφάνιση ενός αντικειμένου μέσου
- Ομαδοποιήσουμε πρωταρχικά αντικείμενα μέσων για να σχηματίσουμε σύνθετα αντικείμενα μέσων
- Εφαρμόσουμε δεδομένα ρεύματος σε αντικείμενα μέσων για να τροποποιήσουμε τις ιδιότητες τους (π.χ. κινούμενη υφή που ανήκει σε ένα αντικείμενο; παράμετροι κίνησης να ζωντανεύουν ένα κινούμενο κεφάλι)
- Αλλάζουμε, με αλληλεπίδραση, τα σημεία θέασης και ακοής του χρήστη οπουδήποτε μέσα στο σκηνικό.

### 6.2.3 Περιγραφή και συγχρονισμός των κατευθυνόμενων δεδομένων για αντικείμενα μέσου

Τα αντικείμενα μέσων μπορούν να βασίζονται σε δεδομένα ρεύματος τα οποία μεταφέρονται σε ένα ή περισσότερα στοιχειώδη ρεύματα. Όλα τα ρεύματα που ενώνονται προς ένα αντικείμενο μέσου αναγνωρίζονται από ένα περιγραφέα αντικειμένων. Αυτό επιτρέπει να χειριζόμαστε ιεραρχικά κωδικοποιημένα δεδομένα τόσο καλά όσο η και ένωση της μετά πληροφορίας γύρω από το περιεχόμενο (object content information) και των δικαιωμάτων πνευματικής ιδιοκτησίας που σχετίζονται με αυτά.

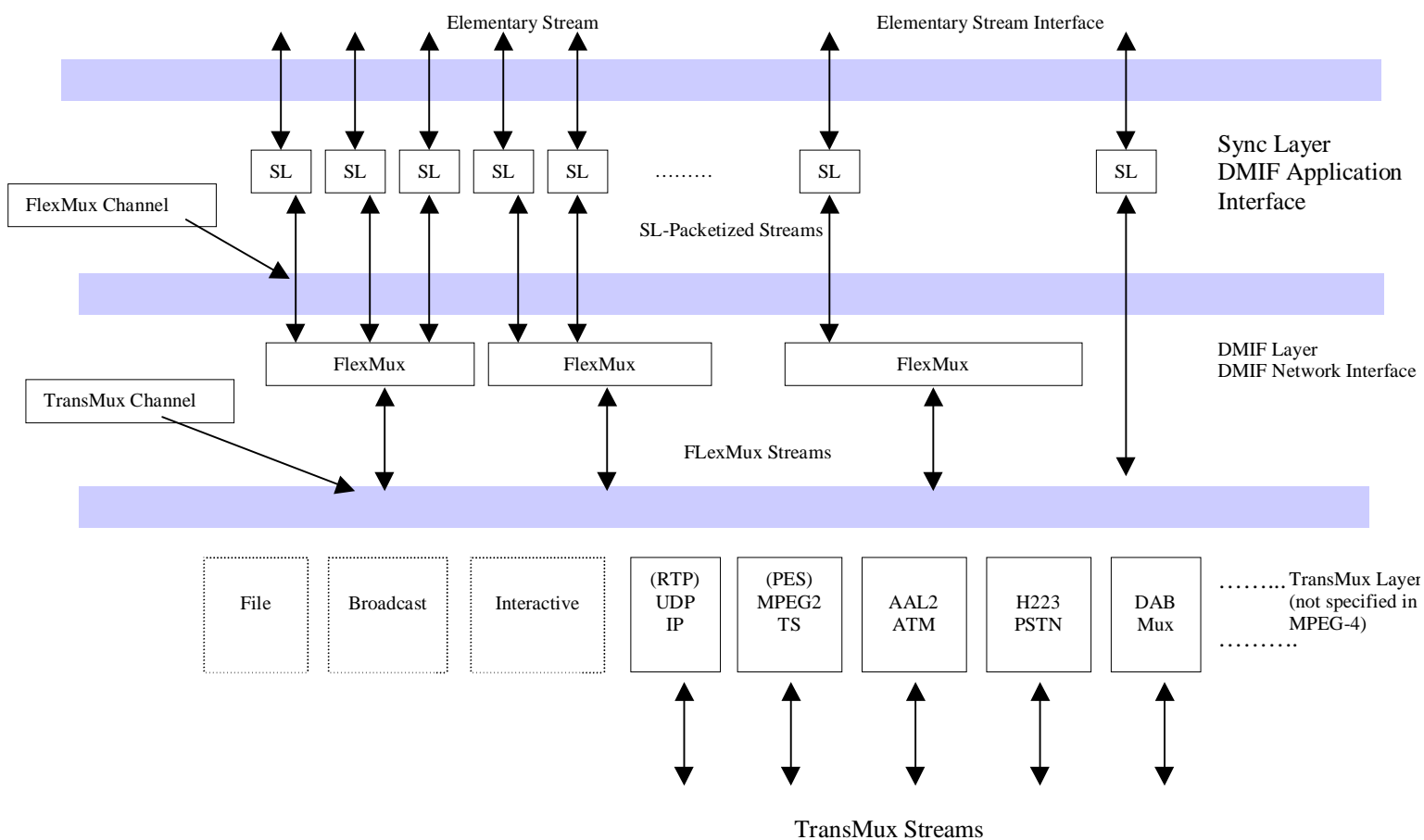
Κάθε ρεύμα χαρακτηρίζεται από μόνο του από ένα σύνολο από περιγραφείς που μεταφέρουν πληροφορία διαμόρφωσης (configuration information), π.χ. να αποφασιστούν οι απαιτούμενοι πόροι του αποκωδικοποιητή και η ακρίβεια της κωδικοποιημένης χρονικής πληροφορίας. Επιπλέον οι περιγραφείς μπορούν να μεταφέρουν υποδείξεις στην Ποιότητα της Υπηρεσίας που ζητά για μετάδοση (π.χ. μέγιστος ρυθμός bit, ρυθμός bit σφάλματος, προτεραιότητα κλπ).

Ο συγχρονισμός των βασικών ρευμάτων επιτυγχάνεται μέσω χρονικού μαρκαρίσματος από μονάδες ανεξάρτητης προσπέλασης μέσα στα βασικά ρεύματα. Η αναγνώριση τέτοιων μονάδων προσπέλασης και το χρονικό μαρκάρισμα επιτυγχάνονται από το στρώμα συγχρονισμού. Ανεξάρτητο από το είδος του μέσου, αυτό το στρώμα επιτρέπει αναγνώριση των μονάδων προσπέλασης (π.χ., οπτικά ή ακουστικά πλαίσια, εντολές περιγραφής σκηنيκού) στα βασικά ρεύματα, ανάκτηση του αντικειμένου μέσου ή περιγραφής του σκηنيκού με βάση το χρόνο και καθιστά δυνατό το συγχρονισμό μεταξύ τους. Η σύνταξη αυτού του στρώματος είναι διαμορφώσιμη με πάρα πολλούς τρόπους, επιτρέποντας τη χρήση σε ένα ευρύ φάσμα συστημάτων.

### 6.2.4 Παράδοση των κατευθυνόμενων δεδομένων

Η συγχρονισμένη παράδοση της πληροφορίας ρεύματος από την πηγή στον προορισμό, εκμεταλλεύομενη διαφορετικές Ποιότητες της Υπηρεσίας όπως διατίθενται από το δίκτυο, καθορίζεται με όρους του προαναφερθέντος στρώματος συγχρονισμού και ενός στρώματος παράδοσης που περιέχει ένα πολυπλέκτη δύο στρωμάτων, όπως φαίνεται στο σχήμα 28.

Το πρώτο στρώμα πολυπλεξίας καθοδηγείται σύμφωνα με τον DMIF ορισμό, μέρος 6 του προτύπου MPEG-4. Αυτός ο πολυπλέκτης μπορεί να ενσωματωθεί από το ορισμένο από την MPEG εργαλείο FlexMux, το οποίο επιτρέπει ομαδοποίηση των Βασικών Ρευμάτων (ES) με μία χαμηλή επιβάρυνση πολυπλεξίας. Η πολυπλεξία σε αυτό το στρώμα μπορεί να χρησιμοποιηθεί για παράδειγμα να ομαδοποιήσει Βασικά Ρεύματα με παρόμοιες απαιτήσεις Ποιότητας Υπηρεσιών, να μειώσει τον αριθμό των συνδέσεων του δικτύου ή να τελειώσει την καθυστέρηση.



Σχήμα 28. Το Μοντέλο MPEG-4 Συστήματος Στρώματος

Το στρώμα “TrunsMux” (Transport Multiplexing) στο σχήμα 28 μοντελοποιεί το στρώμα που προσφέρει υπηρεσίας μεταφοράς οι οποίες ταιριάζουν με τους απαιτούμενες Ποιότητες Υπηρεσιών. Μόνο η διεπαφή σε αυτό το στρώμα ορίζεται στο MPEG-4 ενώ η συγκεκριμένη απεικόνιση των πακέτων δεδομένων και η σηματοδότηση ελέγχου πρέπει να γίνουν σε συνεργασία με τα σώματα που έχουν δικαιοδοσία πάνω στο αντίστοιχο πρωτόκολλο μεταφοράς.

Οποιοδήποτε κατάλληλο υπάρχον πρωτόκολλο μεταφοράς σωρού όπως (RTP)/UDP/IP,(AAL5)/ATM, ή το Ρεύμα Μεταφοράς του MPEG-2 πάνω σε ένα κατάλληλο στρώμα συνδέσμου, μπορούν να γίνουν ένα ειδικό TransMux περιστατικό. Η επιλογή αφήνεται στον τελικό χρήστη/παροχέα υπηρεσιών και επιτρέπει το MPEG-4 να χρησιμοποιείται σε μία μεγάλη ποικιλία από λειτουργικά περιβάλλοντα.

Η χρήση του εργαλείου πολυπλεξίας FlexMux είναι προαιρετική και όπως φαίνεται στο Σχήμα 28, αυτό το στρώμα μπορεί να είναι άδειο αν το

υποκείμενο περιστατικό TransMux παρέχει όλη την απαιτούμενη λειτουργικότητα. Το στρώμα συγχρονισμού πάντως είναι πάντα παρών.

Με βάση το σχήμα 28, μπορεί να είναι δυνατό να:

1. Αναγνωρίσουμε μονάδες προσπέλασης, χρονικές σφραγίδες μεταφοράς και πληροφορία σχετική με το ρολόι και να αναγνωρίσουμε απώλεια δεδομένων
2. Προαιρετικά να εμφωλιάζουμε δεδομένα από διαφορετικά βασικά ρεύματα μέσα σε ρεύματα FlexMux
3. Να μεταβιβάζουμε πληροφορία ελέγχου για να:
  - Υποδείξουμε την απαιτούμενη Ποιότητα Υπηρεσίας για κάθε βασικό ρεύμα και για ρεύμα FlexMux
  - Μεταφράσουμε τέτοιες απαιτήσεις Ποιότητας Υπηρεσιών σε πραγματικούς πόρους του δικτύου
  - Ενώσουμε βασικά ρεύματα σε αντικείμενα μέσων
  - Μεταφέρουμε την απεικόνιση των βασικών ρευμάτων σε κανάλια FlexMux και TransMux

Μέρος των λειτουργικοτήτων ελέγχου θα είναι διαθέσιμες μόνο σε συνδυασμό με μία οντότητα ελέγχου μεταφοράς όπως η δομή DMIF.

### **6.2.5 Αλληλεπίδραση με αντικείμενα μέσων**

Γενικά, ο χρήστης παρατηρεί ένα σκηνικό το οποίο έχει συντεθεί ακολουθώντας τη σχεδίαση του συγγραφέα του σκηνικού. Εξαρτώμενος στο βαθμό ελευθερίας που επιτρέπεται από το συγγραφέα, πάντως, ο χρήστης έχει την δυνατότητα να αλληλεπιδράσει με το σκηνικό. Οι λειτουργίες που μπορεί να επιτραπεί να εκτελέσει ένας χρήστης περιλαμβάνουν:

- Αλλαγή του σημείου θέασης/ακοής του σκηνικού, π.χ. με καθοδήγηση διαμέσου ενός σκηνικού
- Τράβηγμα αντικειμένων του σκηνικού σε μία διαφορετική θέση
- Σκανδαλισμό ενός χείμαρρου από γεγονός κάνοντας «κλικ» σε ένα συγκεκριμένο αντικείμενο, π.χ. ξεκίνημα ή σταμάτημα ενός οπτικού ρεύματος
- Επιλογή της επιθυμητής γλώσσας όταν είναι διαθέσιμες πολλές γλώσσες
- Σκανδαλισμό πιο πολύπλοκων ειδών συμπεριφοράς, π.χ. ένα εικονικό τηλέφωνο κτυπά, ο χρήστης απαντά και εγκαθίσταται ένας δεσμός επικοινωνίας

### **6.2.6 Διαχείριση και αναγνώριση της πνευματικής ιδιοκτησίας**

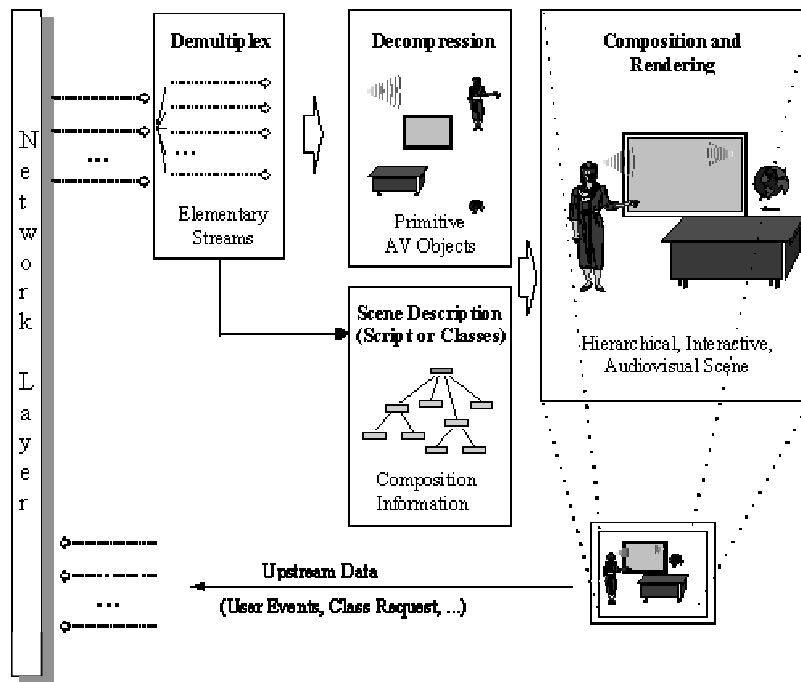
Είναι σημαντικό να έχουμε τη δυνατότητα να αναγνωρίζουμε την πνευματική ιδιοκτησία που κωδικοποιείται σε MPEG-4 αντικείμενα μέσων.

Επομένως, το MPEG εργάζεται με αντιπροσώπους διαφορετικών βιομηχανιών ανάπτυξης στον ορισμό της σύνταξης και των εργαλείων για να υποστηρίξουν αυτό το στόχο. Μία πλήρης ανάπτυξη των απαιτήσεων για την αναγνώριση της πνευματικής ιδιοκτησίας μπορεί να βρεθεί στο ‘Management and Protection of Intellectual Property in MPEG-4’, το οποίο είναι διαθέσιμο από την αρχική σελίδα του MPEG.

Το MPEG-4 συγχωνεύει την αναγνώριση της πνευματικής ιδιοκτησίας αποθηκεύοντας μοναδικούς αναγνωριστές, οι οποίοι προέρχονται από διεθνή συστήματα αρίθμησης (π.χ. ISAN, ISRC, κλπ). Αυτοί οι αριθμοί μπορούν να εφαρμοσθούν για να αναγνωρίσουν ένα προσωρινό χρήστη που κρατά τα δικαιώματα ενός αντικειμένου μέσου. Αφού δεν αναγνωρίζεται όλο το περιεχόμενο από ένα τέτοιο αριθμό, η έκδοση 1 του MPEG-4 προσφέρει τη δυνατότητα να αναγνωρίσουμε την πνευματική ιδιοκτησία από ένα δίτιμο κλειδί (π.χ. «συνθέτης»/«Γιάννης Κωστής»). Επίσης, το MPEG-4 προσφέρει στα άτομα που θέλουν να χρησιμοποιήσουν συστήματα που ελέγχουν την πρόσβαση στην πνευματική ιδιοκτησία, μία τυποποιημένη διεπαφή η οποία ολοκληρώνεται σφικτά στο στρώμα Συστημάτων. Με αυτή τη διεπαφή, ιδιόκτητα συστήματα ελέγχου μπορούν εύκολα να συγχωνευτούν με το τυποποιημένο μέρος του αποκωδικοποιητή.

### **6.3 Αναλυτική τεχνική περιγραφή του προτύπου MPEG\_4**

Όπως φαίνεται στο σχήμα 29, τα ρεύματα που έρχονται από το δίκτυο (ή μία συσκευή αποθήκευσης) ως TransMux Ρεύματα αποπλέκονται σε FlexMux Ρεύματα και περνούν σε κατάλληλους FlexMux αποπλέκτες οι οποίοι επαναφέρουν τα Βασικά Ρεύματα. Αυτό περιγράφεται στην ενότητα 6.3.2. Τα Βασικά Ρεύματα αναλύονται και περνούν στους κατάλληλους αποκωδικοποιητές. Η αποκωδικοποίηση επαναφέρει τα δεδομένα σε ένα AV αντικείμενο από την κωδικοποιημένη του μορφή και εκτελεί τις απαραίτητες λειτουργίες για να ανοικοδομήσει το αρχικό AV αντικείμενο, έτοιμο για παράδοση στην κατάλληλη συσκευή. Τα ακουστικά αντικείμενα αναπαρίστανται στην κωδικοποιημένη τους μορφή η οποία περιγράφεται στην ενότητα 6.3.4 . Τα ανοικοδομημένο AV αντικείμενο γίνεται διαθέσιμο στο στρώμα σύνθεσης για ενδεχόμενη χρήση κατά τη διάρκεια παράδοσης του σκηνικού. Τα αποκωδικοποιημένα AVΟ, μαζί με την πληροφορία περιγραφής του σκηνικού, χρησιμοποιούνται για να συνθέσουν το σκηνικό όπως περιγράφεται από το συγγραφέα. Ο χρήστης μπορεί, στο σημείο που επιτρέπεται από το συγγραφέα, να αλληλεπιδράσει με το σκηνικό το οποίο τελικά παραδίνεται και παρουσιάζεται. Η ενότητα 6.3 περιγράφει αυτή την αλληλεπίδραση.



Σχήμα 29. Κύρια συστατικά ενός MPEG-4 τερματικού (πλευρά δέκτη)

### 6.3.1 DMIF

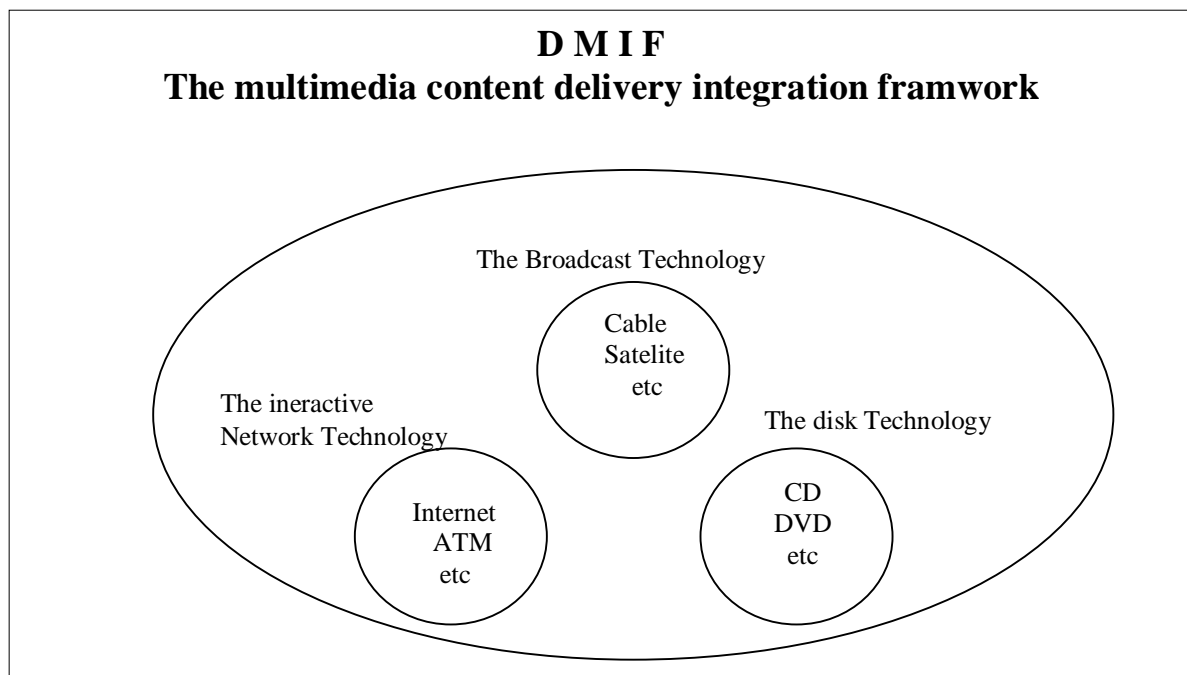
Το DMIF (Delivery Multimedia Integration Framework) είναι ένα πρωτόκολλο συνεδρίασης για τη διαχείριση των πολυμέσων ρεύματος πάνω σε γενικές τεχνολογίες παράδοσης. Γενικά είναι όμοιο με το FTP. Η μόνη (αλλά βασική) διαφορά είναι ότι το FTP επιστρέφει δεδομένα, ενώ το DMIF επιστρέφει δείκτες που δείχνουν που να πάρουμε το δεδομένα (ρεύματος). Όταν το FTP εκτελείται, η πρώτη πράξη που πραγματοποιεί είναι η εγκατάσταση μίας «συνεδρίασης» με την απομακρυσμένη πλευρά. Αργότερα, τα αρχεία επιλέγονται και το FTP στέλνει μία αίτηση για να κατέβουν, το FTP ομότιμα θα επιστρέψει τα αρχεία σε μία ξεχωριστή σύνδεση.

Ομοίως, όταν το DMIF εκτελείται, η πρώτη πράξη που πραγματοποιεί είναι η εγκατάσταση μιας συνεδρίασης με την απομακρυσμένη πλευρά. Αργότερα, τα ρεύματα επιλέγονται και το DMIF στέλνει μία αίτηση για να καθοδηγηθούν, το DMIF ομότιμα θα επιστρέψει τους δείκτες στις συνδέσεις όπου τα ρεύματα θα καθοδηγηθούν (και εγκαθιστούν τη σύνδεση από μόνα τους). Συγκρινόμενο με το FTP, το DMIF είναι και δομή και πρωτόκολλο. Η λειτουργικότητα που παρέχεται από το DMIF εκφράζεται από μία διεπαφή που καλείται DAI (DMIF-Application Interface) και μεταφράζεται σε μηνύματα πρωτοκόλλου. Αυτά τα μηνύματα πρωτοκόλλου μπορούν να διαφέρουν ανάλογα με το δίκτυο στο οποίο λειτουργούν.

Η Ποιότητα της Υπηρεσίας λαμβάνεται υπόψη στη σχεδίαση DMIF, και το DAI επιτρέπει στον DMIF χρήστη να προσδιορίσει τις απαιτήσεις για το επιθυμητό ρεύμα. Τότε εξαρτάται από την υλοποίηση του DMIF να



διασφαλίσουμε ότι οι απαιτήσεις ικανοποιούνται. Ο ορισμός του DMIF παρέχει υποδείξεις για το πώς θα εκτελέσουμε τέτοιες εφαρμογές σε μερικά είδη δικτύου (συμπεριλαμβανομένου το Internet). Το DAI επίσης χρησιμοποιείται για προσπέλαση υλικού εκπομπής και τοπικών αρχείων, αυτό σημαίνει ότι μια μονή, ενιαία διεπαφή καθορίζεται για να προσπελάσουμε περιεχόμενα πολυμέσων ή μια πληθώρα από τεχνολογίες παράδοσης. Ως συνέπεια, είναι πρόβλημα να τονίσουμε ότι η δομή ολοκλήρωσης του DMIF καλύπτει τρεις κύριες τεχνολογίες, την αλληλοεπιδρούμενη τεχνολογία δικτύου, την τεχνολογία εκπομπής και την τεχνολογία δίσκου όπως φαίνονται στο σχήμα 30.



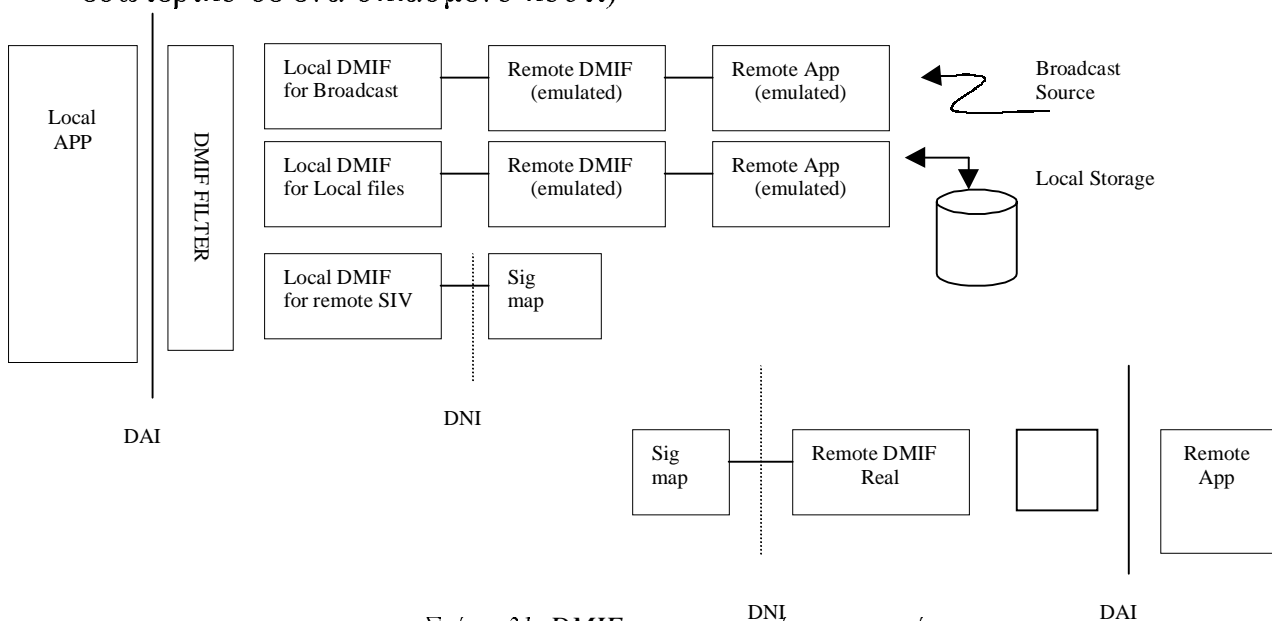
*Σχήμα 30. Το DMIF διευθύνει την ολοκλήρωση παράδοσης τριών κύριων τεχνολογιών*

Η DMIF αρχιτεκτονική είναι τέτοια που εφαρμογές οι οποίες στηρίζονται στο DMIF για επικοινωνία δεν έχουν να ανησυχούν με την υποκείμενη μέθοδο επικοινωνίας. Η υλοποίηση του DMIF φροντίζει για τις λεπτομέρειες της τεχνολογίας παράδοσης παρουσιάζοντας μια απλή διεπαφή στην εφαρμογή. Το σχήμα 30 αναπαριστά την παραπάνω αντίληψη. Μία εφαρμογή προσπελάει δεδομένα διαμέσου της DMIF-Εφαρμογή Διεπαφής (Application Interface), ανεξαρτήτως εάν αυτά τα δεδομένα προέρχονται από μία πηγή εκπομπής, από μία τοπική αποθήκευση ή από ένα απομακρυσμένο επόπτη (server). Σε όλα τα σενάρια η Τοπική Εφαρμογή αλληλεπιδρά μόνο διαμέσου μιας ενιαίας διεπαφής (DAI). Τα διαφορετικά DMIF περιστατικά τότε θα μεταφράσουν τις αιτήσεις της Τοπικής Εφαρμογής σε καθορισμένα μηνύματα για να παραδοθούν στην Απομακρυσμένη Εφαρμογή, λαμβάνοντας υπόψη τις ιδιομορφίες της εμπλεκόμενης τεχνολογίας παράδοσης. Ομοίως, τα δεδομένα που φτάνουν στο τερματικό (από απομακρυσμένους επόπτες, δίκτυα εκπομπής

ή τοπικά αρχεία) ομοιόμορφα παραδίνονται στην Τοπική Εφαρμογή διαμέσου του DAI.

Διαφορετικά καθορισμένα DMIF περιστατικά έμμεσα καλούνται από την Εφαρμογή να διαχειριστούν τις διάφορες καθορισμένες τεχνολογίες παράδοσης, αυτό είναι πάντως προφανές στην Εφαρμογή, η οποία αλληλεπιδρά μόνο με ένα μονό “DMIF φίλτρο”. Αυτό το φίλτρο είναι τότε υπεύθυνο για την διεύθυνση του συγκεκριμένου αρχικού DAI στο σωστό περιστατικό. Το DMIF δεν καθορίζει αυτό το μηχανισμό, απλά υποθέτει ότι υλοποιείται. Αυτό επιπλέον τονίζεται από τα σκιασμένα κουτιά στο σχήμα 31, των οποίων στόχος είναι να αποσαφηνίσουν ποια είναι τα όρια μιας DMIF υλοποίησης, ενώ η DMIF αρχιτεκτονική ορίζει ένα αριθμό από μέτρα, οι αληθινές DMIF υλοποιήσεις χρειάζονται μόνο να διατηρήσουν την εμφάνιση τους σε αυτά τα όρια.

Εννοιολογικά, μία «πραγματική» απομακρυσμένη εφαρμογή που προσπελάζεται διαμέσου ενός δικτύου π.χ. βασισμένο σε IP ή ATM, δεν διαφέρει από μία εφαρμογή απομίμηση απομακρυσμένου παραγωγού η οποία λαμβάνει περιεχόμενο από μία πηγή εκπομπής ή από ένα δίσκο. Πάντως στην προηγούμενη περίπτωση τα μηνύματα που ανταλλάσσονται μεταξύ των δύο οντοτήτων πρέπει να είναι κανονιστικά ορισμένα για να διασφαλίσουμε την εσωτερική λειτουργικότητα (αυτά είναι τα DMIF μηνύματα σηματοδότησης), ενώ στην τελευταία περίπτωση οι διεπαφές μεταξύ των δύο DMIF ταιριών και της προσομοιωμένης Απομακρυσμένης Εφαρμογής είναι εσωτερικές σε μία μονή υλοποίηση και δεν χρειάζεται να ληφθούν υπόψη σε αυτό τον καθορισμό. Σημειώστε ότι για το σενάριο εκπομπής και τοπικής αποθήκευσης το σχήμα δείχνει μία αλυσίδα από «τοπικό DMIF», «Απομακρυσμένο DMIF (προσομοιωμένο)» και «Απομακρυσμένη Εφαρμογή (προσομοιωμένη)», αυτή η αλυσίδα αναπαριστά μόνο ένα ιδεατό μοντέλο και δεν χρειάζεται να αντανακλάται σε πραγματικές υλοποιήσεις (δείχνεται στο σχήμα ολοκληρωτικά εσωτερικό σε ένα σκιασμένο κουτί)



Σχήμα 31. DMIF αρχιτεκτονική επικοινωνίας

Όσον αφορά τα σενάρια της Εκπομπής και Τοπικής Αποθήκευσης, υποτίθεται ότι η (προσομοιωμένη) Απομακρυσμένη εφαρμογή έχει λάβει γνώση για το πώς τα δεδομένα παραδίδονται/ αποθηκεύονται. Αυτό υπονοεί γνώση του είδους της εφαρμογής με την οποία σχετίζονται. Στην περίπτωση του MPEG-4, αυτό πρακτικά εννοεί γνώση των ιδεών όπως την ταυτότητα (ID) του Βασικού Ρεύματος, τον Πρώτο Περιγραφέα Αντικειμένων, το Όνομα Υπηρεσίας. Έτσι, ενώ το Στρώμα DMIF δεν γνωρίζει την εφαρμογή στην οποία παρέχει υποστήριξη, στην ιδιαίτερη περίπτωση των DMIF περιστατικών για Εκπομπή και Τοπική Αποθήκευση αυτή η υπόθεση δεν είναι εντελώς σωστή, εξαιτίας της παρουσίας της (προσομοιωμένης) Απομακρυσμένης Εφαρμογής (η οποία, από την πλευρά της Τοπικής Εφαρμογής, είναι ακόμα μέρος του DMIF Στρώματος). Είναι αξιοσημείωτο ότι αφού η (προσομοιωμένη) Απομακρυσμένη Εφαρμογή έχει λάβει γνώση για το πώς τα δεδομένα παραδίδονται / αποθηκεύονται, ο προσδιορισμός του πώς τα δεδομένα παραδίδονται / αποθηκεύονται είναι κρίσιμος για μία τέτοια DMIF υλοποίηση, η οποία είναι επομένως η «MPEG-4 ενημέρωση συστήματος».

Όταν αντιθέτως λαμβάνουμε υπόψη το σενάριο της Απομακρυσμένης Αλληλεπίδρασης, το DMIF Στρώμα είναι εντελώς ανενήμερο για την εφαρμογή. Μια επιπρόσθετη διεπαφή- η DMIF Διεπαφή Δικτύου (DNI Network Interface)- εισάγεται για να δώσει έμφαση του είδους της πληροφορίας των DMIF ταιριών που πρέπει να ανταλλάξουν. Ένα επιπρόσθετο μέτρο (η «Απεικόνιση Σηματοδότησης» - Signalling mapping- στο σχήμα) φροντίζει για την απεικόνιση των πρωτόγονων DNI σε μηνύματα σηματοδότησης που χρησιμοποιούνται στο συγκεκριμένο δίκτυο. Σημειώστε ότι τα πρωτόγονα DNI καθορίζονται μόνο για σκοπούς πληροφορίας, και δεν χρειάζεται να είναι παρών μια διεπαφή DNI σε μία πραγματική εφαρμογή. Το σχήμα 31 επίσης αναπαριστά καθαρά το DNI ως εσωτερικό στο σκιασμένο κουτί. Αντιθέτως, η σύνταξη των μηνυμάτων που ακολουθούν στο Δίκτυο προσδιορίζεται πλήρως για κάθε συγκεκριμένο δίκτυο που υποστηρίζεται.

Το DMIF επιτρέπει την ταυτόχρονη παρουσία ενός ή περισσότερων DMIF περιστατικών, καθένα προσδιορισμένο για μία συγκεκριμένη τεχνολογία παράδοσης, για να υποστηρίξουμε στις ίδιες πολλαπλές τεχνολογίες παράδοσης τερματικού και ζυγά πολλαπλά σενάρια (εκπομπή, τοπική αποθήκευση, απομακρυσμένη αλληλεπίδραση). Πολλαπλές τεχνολογίες παράδοσης μπορούν να ενεργοποιηθούν από την ίδια εφαρμογή, επομένως αυτές θα μπορούσαν να διαχειρίζονται μεμονωμένα τα δεδομένα που στέλνονται από δίκτυα εκπομπής, τοπικά συστήματα αρχείων και απομακρυσμένα ταίρια που αλληλεπιδρούν.

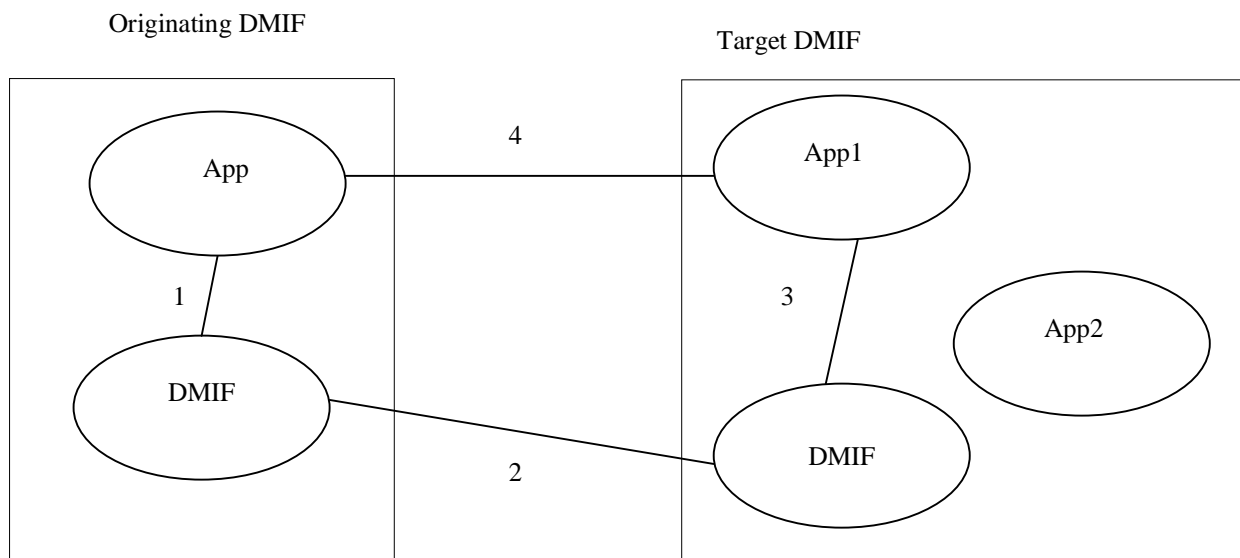
- **Το DMIF Υπολογιστικό Μοντέλο**

Όταν μία εφαρμογή ζητά την ενεργοποίηση μιας υπηρεσίας, χρησιμοποιεί τα πρωτόγονα DAI της Υπηρεσίας, και δημιουργεί μία διάσκεψη υπηρεσίας. Η

DMIF υλοποίηση τότε επικοινωνεί με το αντίστοιχο ταίρι της (αυτό ιδεατά μπορεί να είναι είτε ένα απομακρυσμένο ταίρι, ή ένα τοπικό προσομοιωμένο ταίρι) και δημιουργεί μία διάσκεψη δικτύου με αυτό. Οι διασκέψεις δικτύου έχουν μεγάλη σημασία για όλο το δίκτυο, ενώ αντίθετα οι διασκέψεις υπηρεσιών έχουν τοπική σημασία. Η ένωση μεταξύ τους διατηρείται από το Στρώμα DMIF. Στην περίπτωση των σεναρίων εκπομπής και τοπικής αποθήκευσης, ο τρόπος με τον οποίο η διάσκεψη δικτύου δημιουργείται και διαχειρίζεται είναι έξω από αυτό τον καθορισμό. Αντιθέτως στην περίπτωση του σεναρίου απομακρυσμένης αλληλεπίδρασης, το DMIF χρησιμοποιεί τον φυσικό μηχανισμό σηματοδότησης για αυτό το δίκτυο για να δημιουργήσει και έπειτα να διαχειριστεί τη διάσκεψη δικτύου π.χ. ATM σηματοδότηση. Τα ταίρια της εφαρμογής τότε χρησιμοποιούν αυτή τη διάσκεψη για να δημιουργήσουν διασυνδέσεις οι οποίες χρησιμοποιούνται για να μεταφέρουν τα δεδομένα της εφαρμογής, π.χ. MPEG-4 Βασικά Ρεύματα.

Όταν μία εφαρμογή χρειάζεται ένα κανάλι, χρησιμοποιεί τα πρωτόγονα του DAI του καναλιού, το DMIF μεταφράζει αυτές τις αιτήσεις σε αιτήσεις διασύνδεσης οι οποίες είναι καθορισμένες στην συγκεκριμένη υλοποίησης δικτύου. Στην περίπτωση των σεναρίων Εκπομπής και Τοπικής Αποθήκευσης, ο τρόπος με τον οποίο δημιουργούνται οι διασυνδέσεις και έπειτα διαχειρίζονται είναι έξω από τον σκοπό αυτού του καθορισμού. Αντιθέτως, στην περίπτωση του σεναρίου δικτύου, το DMIF χρησιμοποιεί το φυσικό μηχανισμό σηματοδότησης για αυτό το δίκτυο για να δημιουργήσει αυτές τις διασυνδέσεις. Η εφαρμογή έπειτα χρησιμοποιεί αυτές τις διασυνδέσεις για να παραδώσει την υπηρεσία.

Το σχήμα 32 παρέχει μία όψη υψηλού επιπέδου για μια ενεργοποίηση υπηρεσία και της έναρξης της ανταλλαγής δεδομένων. Η πορεία υψηλού επιπέδου περιλαμβάνει τα ακόλουθα βήματα:



Σχήμα 32. Το DMIF υπολογιστικό μοντέλο

Η Πρωταρχική Εφαρμογή αιτεί την ενεργοποίηση μιας υπηρεσίας στο τοπικό της DMIF Στρώμα-ένα μονοπάτι επικοινωνίας μεταξύ της Πρωταρχικής Εφαρμογής και του τοπικού της DMIF ταιριού εγκαθίσταται στο επίπεδο ελέγχου (1)

Το Πρωταρχικό DMIF ταίρι εγκαθιστά μία διάσκεψη δικτύου με το DMIF ταίρι Στόχο - ένα μονοπάτι επικοινωνίας μεταξύ του Πρωταρχικού DMIF ταιριού και του Στόχου DMIF ταιριού εγκαθίσταται στο επίπεδο ελέγχου (2)  
Το ταίρι Στόχος DMIF αναγνωρίζει την Εφαρμογή Στόχο και προχωρά την αίτηση ενεργοποίησης υπηρεσίας – ένα μονοπάτι επικοινωνίας μεταξύ του Στόχου DMIF ταιριού και της Εφαρμογής Στόχου εγκαθίσταται στο επίπεδο ελέγχου (3)

Το ταίρι Εφαρμογών δημιουργεί κανάλια (αιτήσεις που ρέουν διαμέσου των μονοπατιών επικοινωνίας 1,2, και 3). Τα καταληκτικά κανάλια στο επίπεδο χρήστη (4) θα μεταφέρουν τα πραγματικά δεδομένα που ανταλλάχθηκαν από τις Εφαρμογές.

Το DMIF αναμειγνύεται και στα τέσσερα βήματα που αναφέρθηκαν.

Το DMIF Στρώμα αυτόματα αποφασίζει εάν μία συγκεκριμένη υπηρεσία υποτίθεται ότι πρέπει να παραχθεί από ένα απομακρυσμένο επόπτη σε ένα συγκεκριμένο δίκτυο π.χ. βασισμένο σε IP ή ATM, από ένα δίκτυο εκπομπής ή να μείνει σε μία τοπική συσκευή αποθήκευσης. Η επιλογή βασίζεται στην πληροφορία διεύθυνσης του ταιριού που παρέχεται από την Εφαρμογή σαν μέρος ενός URL που περνά στο DAI.

### **6.3.2 Αποπολυπλεξία, συγχρονισμός και περιγραφή των κατευθυνόμενων (streaming) δεδομένων**

Τα ανεξάρτητα Βασικά Ρεύματα πρέπει να ανακτηθούν στο στρώμα παράδοσης από τα εισερχόμενα δεδομένα από κάποια διασύνδεση του δικτύου ή μια αποθηκευτική συσκευή. Κάθε διασύνδεση δικτύου ή αρχείο θεωρείται ομοιογενώς ένα TransMux κανάλι στο μοντέλο συστήματος MPEG-4. Η αποπολυπλεξία πραγματοποιείται μερικώς ή ολικώς από στρώματα εκτός του σκοπού του MPEG-4, ανάλογα με την εφαρμογή. Το μόνο εργαλείο αποπολυπλεξίας που ορίζεται από το MPEG-4 είναι το FlexMux εργαλείο το οποίο μπορεί προαιρετικά να χρησιμοποιηθεί για χαμηλή καθυστέρηση, για χαμηλή επιβάρυνση πολυπλεξίας και για αποθήκευση των πηγών διασύνδεσης του δικτύου.

Για το σκοπό της ολοκλήρωσης του MPEG-4 σε περιβάλλοντα συστημάτων, η DMIF Διεπαφή Εφαρμογής είναι το σημείο αναφοράς στο οποίο βασικά ρεύματα μπορούν να προσπελαστούν ως σύγχρονα πακεταρισμένα ρεύματα στρώματος. Η DMIF Διεπαφή Δικτύου προσδιορίζει πως SL-πακεταρισμένα ρεύματα (δεν χρησιμοποιείται FlexMux) ή FlexMux Ρεύματα πρόκειται

να ανακτηθούν από το TransMux Στρώμα. Αυτή είναι η διεπαφή στις λειτουργικές μεταφορές που δεν ορίζονται από το MPEG. Το κομμάτι των δεδομένων λαμβάνεται υπόψη εδώ ενώ το κομμάτι ελέγχου έχει να κάνει με το DMIF.

Με τον ίδιο τρόπο που το MPEG-1 και MPEG-2 περιέγραψαν την συμπεριφορά μιας ιδανικής συσκευής αποκωδικοποίησης μαζί με τη σύνταξη του ρεύματος bit και την σημασιολογία, το MPEG-4 ορίζει ένα Μοντέλο Συστήματος Αποκωδικοποιητή. Αυτό επιτρέπει τον ακριβή ορισμό της λειτουργίας του τερματικού χωρίς να κάνει περιττές υποθέσεις γύρω από τις λεπτομέρειες υλοποίησης. Αυτό είναι βασικό για να δώσουμε στους υλοποιητές την ελευθερία να σχεδιάσουν πραγματικά MPEG-4 τερματικά και συσκευές αποκωδικοποίησης με μία ποικιλία από τρόπους. Αυτές οι συσκευές εκτείνονται από τηλεοπτικούς δέκτες οι οποίοι δεν έχουν την ικανότητα να επικοινωνήσουν μέσω αποστολέα με υπολογιστές οι οποίοι είναι πλήρως ικανοί να εκτελέσουν αμφίδρομη επικοινωνία. Κάποιες συσκευές θα λάβουν MPEG-4 ρεύματα πάνω σε ισόχρονα δίκτυα ενώ άλλες θα χρησιμοποιήσουν μη ισόχρονα μέσα (π.χ. το Internet) για να ανταλλάξουν MPEG-4 πληροφορία. Το Μοντέλο Συστήματος Αποκωδικοποιητή παρέχει ένα κοινό μοντέλο πάνω στο οποίο όλες οι υλοποιήσεις των MPEG-4 τερματικών μπορούν να βασιστούν.

Ο προσδιορισμός μιας προσωρινής μνήμης (buffer) και ενός χρονικού μοντέλου είναι βασικός στις συσκευές κωδικοποίησης οι οποίες μπορεί να μην γνωρίζουν εκ των προτέρω ποια είναι η συσκευή τερματικού ή πως θα λάβει το κωδικοποιημένο ρεύμα. Αν και ο καθορισμός του MPEG-4 θα καταστήσει ικανή την συσκευή κωδικοποίησης να πληροφορήσει τη συσκευή αποκωδικοποίησης για τις απαιτήσεις πόρων, μπορεί να μην είναι δυνατό, όπως υποδείχθηκε νωρίτερα, για αυτή τη συσκευή να ανταποκριθεί στον αποστολέα. Είναι επίσης πιθανό, μία MPEG-4 συνεδρίαση να λαμβάνεται ταυτόχρονα από πολλές διαφορετικές συσκευές. Θα είναι πάντως σωστά παραδομένη σύμφωνα με την ικανότητα κάθε συσκευής.

- **Αποπολυπλεξία**

Η αποπολυπλεξία συμβαίνει στο στρώμα παράδοσης το οποίο είναι μοντελοποιημένο ως συστατικό ενός TransMux στρώματος και ενός DMIF στρώματος. Η ανάκτηση των εισερχομένων ρευμάτων δεδομένων από τις διασυνδέσεις του δικτύου ή των μέσων αποθήκευσης αποτελείται από δύο καθήκοντα. Πρώτα, τα κανάλια πρέπει να εντοπιστούν και να ανοιχθούν. Αυτό απαιτεί μία οντότητα ελέγχου μεταφοράς που διαχειρίζεται, μεταξύ άλλων, τους πίνακες που ενώνουν τα κανάλια μεταφοράς με καθορισμένα βασικά ρεύματα. Οι πίνακες απεικόνισης ρευμάτων συνδέουν κάθε ρεύμα σε ένα Σημάδι Ένωσης Καναλιού (Channel Association Tag) το οποίο εξυπηρετεί σαν «χερούλι» στο κανάλι το οποίο μεταφέρει αυτό το ρεύμα. Η ανάλυση των CATS στο πραγματικό κανάλι μεταφοράς όπως επίσης η διαχείριση των συνεδριάσεων και των καναλιών, διευθυνσιοδοτείται από το DMIF μέρος του MPEG-4 προτύπου.

Δεύτερον, τα εισερχόμενα ρεύματα πρέπει να είναι σωστά αποπολυπλεγμένα για να ανακτήσουν SL-πακεταρισμένα ρεύματα από τα κανάλια των κάτω ρευμάτων (εισερχόμενα στο τερματικό υποδοχής) για να περαστούν στο στρώμα συγχρονισμού. Σε εφαρμογές αλληλεπίδρασης, ένα αντίστοιχο στάδιο πολυπλεξίας θα πολυπλέξει τα δεδομένα των άνω ρευμάτων σε κανάλια των πάνω ρευμάτων (εξερχόμενα από το τερματικό υποδοχής)

Ο γενικός όρος “TransMux Layer” χρησιμοποιείται για να αφαιρέσει κάθε υποκείμενη λειτουργικότητα πολυπλεξίας – υπάρχουσα ή μελλοντική- που είναι κατάλληλη για να μεταφέρει MPEG-4 ρεύματα δεδομένων. Σημειώστε ότι αυτό το στρώμα δεν ορίζεται στα περιεχόμενα του MPEG-4. Παραδείγματα είναι το MPEG-2 Ρεύμα Μεταφοράς, H.223, ATM AAL 2, IP/UDP. Το στρώμα Transmux θεωρείται ότι θα παρέχει προστασία και λειτουργικότητα πολυπλεξίας, υποδεικνύοντας ότι αυτό το στρώμα είναι υπεύθυνο για να προσφέρει μία καθορισμένη Ποιότητα Υπηρεσιών. Η λειτουργικότητα της προστασίας περιλαμβάνει εργαλεία προστασίας από σφάλματα και εντοπισμού σφαλμάτων κατάλληλα για το δοσμένο δίκτυο ή μέσο αποθήκευσης.

Σε οποιοδήποτε σενάριο εφαρμογών θα χρησιμοποιηθούν ένα ή περισσότερα καθορισμένα TransMux περιστατικά. Κάθε TransMux αποπλέκτης δίνει πρόσβαση σε TransMux κανάλια. Οι απαιτήσεις για τη διεπαφή των δεδομένων για να προσπελάσουμε ένα TransMux κανάλι είναι οι ίδιες για όλα τα TransMux περιστατικά. Αυτές περιλαμβάνουν την ανάγκη για αξιόπιστο εντοπισμό σφαλμάτων, παράδοση, εάν είναι δυνατό, λανθασμένων δεδομένων με μία κατάλληλη ένδειξη σφάλματος και περιορισμό του φορτίου το οποίο μπορεί να αποτελείται είτε από SL-πακεταρισμένα ρεύματα ή FlexMux ρεύματα. Αυτές οι απαιτήσεις γίνονται περιληπτικές με ένα πληροφοριακό τρόπο στη TransMux Διεπαφή, στο τμήμα Συστημάτων του MPEG-4 προτύπο. Μία προσαρμογή των SL-πακεταρισμένων ρευμάτων πρέπει να καθοριστεί σε κάθε σωρό πρωτοκόλλων μεταφοράς που έχει ενδιαφέρον σύμφωνα με αυτές τις απαιτήσεις και σε συνδυασμό με το σώμα τυποποίησης που έχει την κατάλληλη δικαιοδοσία. Αυτή η διεργασία έχει ξεκινήσει.

Το εργαλείο FlexMux προσδιορίζεται από το MPEG για να παρέχει προαιρετικά μια ευέλικτη, χαμηλής επιβάρυνσης, χαμηλής καθυστέρησης μέθοδο για να αφήνει εσωτερικά τα δεδομένα οποτεδήποτε αυτό δεν υποστηρίζεται επαρκώς από τον υποκείμενο σωρό πρωτοκόλλου. Είναι ιδιαίτερα χρήσιμο όταν το μέγεθος του πακέτου ή η επιβάρυνση του υποκείμενου TransMux περιστατικού είναι μεγάλα, γιατί αλλιώς θα καταλήγαμε σε μία σπατάλη εύρους ζώνης ή αριθμού διασυνδέσεων του δικτύου. Το εργαλείο FlexMux δεν είναι από μόνο του ανθεκτικό σε σφάλματα και μπορεί να χρησιμοποιηθεί είτε σε TransMux κανάλια με υψηλή Ποιότητα Υπηρεσιών ή να δένει Βασικά Ρεύματα που είναι εξίσου ανθεκτικά στα σφάλματα. Το FlexMux απαιτεί αξιόπιστο εντοπισμό σφαλμάτων και επαρκή περιορισμό των FlexMux πακέτων (για τυχαία προσπέλαση και ανάκτηση σφαλμάτων) από τι υποκείμενο στρώμα. Αυτές οι απαιτήσεις αντανακλώνται επίσης στα πρωτόγονα δεδομένα της DMIF

Διεπαφής Εφαρμογής η οποία ορίζει την προσπέλαση δεδομένων σε ανεξάρτητα κανάλια μεταφοράς. Ο FlexMux αποπλέκτης ανακτά SL-πακεταρισμένα ρεύματα από FlexMux ρεύματα.

- **Συγχρονισμός και περιγραφή των βασικών ρευμάτων**

Το στρώμα συγχρονισμού έχει ένα ελάχιστο σύνολο από εργαλεία για έλεγχο σταθερότητας, γεμίσματος, για να μεταβιβάσουν πληροφορία βασισμένη στο χρόνο και να μεταφέρουν χρονικά σφραγισμένες μονάδες προσπέλασης ενός βασικού ρεύματος. Κάθε πακέτο αποτελείται από μία μονάδα προσπέλασης ή ένα τμήμα μιας μονάδας προσπέλασης. Αυτές οι χρονικά σφραγισμένες μονάδες προσπέλασης σχηματίζουν την μόνη εννοιολογική δομή των βασικών ρευμάτων που είναι ορατά σε αυτό το στρώμα. Οι χρονικές σφραγίδες χρησιμοποιούνται για να μεταφέρουν την ονομαστική αποκωδικοποίηση και το χρόνο σύνθεσης για μία μονάδα προσπέλασης. Το στρώμα συγχρονισμού απαιτεί αξιόπιστο εντοπισμό σφαλμάτων και πλαισίωση του κάθε ανεξάρτητου πακέτου από το υποκείμενο στρώμα, το οποίο μπορεί να πραγματοποιηθεί π.χ. χρησιμοποιώντας το FlexMux. Το πώς τα δεδομένα μπορούν να προσπελαστούν από το στρώμα συμπίεσης αναφέρεται περιληπτικά στην πληροφοριακή Διεπαφή Βασικών Ρευμάτων, η οποία μπορεί να βρεθεί στο τμήμα Συστημάτων του MPEG-4 προτύπου. Το στρώμα συγχρονισμού ανακτά τα βασικά ρεύματα από τα SL-πακεταρισμένα ρεύματα.

Για να είναι δυνατό να συσχετίσουμε βασικά ρεύματα με αντικείμενα μέσω μέσα σε ένα σκηνικό, χρησιμοποιούνται περιγραφείς αντικειμένων. Οι περιγραφείς αντικειμένων μεταβιβάζουν πληροφορία με τον αριθμό και τις ιδιότητες των βασικών ρευμάτων που είναι συνδεδεμένα με συγκεκριμένα αντικείμενα μέσω. Οι περιγραφείς αντικειμένων από μόνοι τους μεταβιβάζονται σε ένα ή περισσότερα βασικά ρεύματα, αφού είναι δυνατό να προσθέσουμε και να πετούμε ρεύματα (και αντικείμενα) κατά τη διάρκεια του τμήματος μιας MPEG-4 παρουσίασης. Τα ρεύματα του περιγραφέα αντικειμένων μπορούν να θεωρηθούν ως μία περιγραφή των πόρων ρεύματος για μία παρουσίαση. Ομοίως, η περιγραφή του σκηνικού μεταβιβάζεται επίσης ως ένα βασικό ρεύμα, επιτρέποντας να τροποποιήσουμε το χωροχρονικό περίγραμμα της παρουσίασης με το χρόνο.

- **Διαχείριση προσωρινής μνήμης**

Για να προβλέψουμε πώς ο αποκωδικοποιητής θα συμπεριφερθεί όταν αποκωδικοποιεί τα διάφορα βασικά ρεύματα δεδομένων που σχηματίζουν μία MPEG-4 συνεδρίαση, το Μοντέλο Συστήματος Αποκωδικοποιητή καθιστά ικανό τον κωδικοποιητή να προσδιορίσει και να παρακολουθήσει τους ελάχιστους πόρους προσωρινής μνήμης που χρειάζονται για να αποκωδικοποιηθεί μία συνεδρίαση. Οι απαιτούμενοι πόροι προσωρινής μνήμης μεταβιβάζονται στον αποκωδικοποιητή μέσα σε περιγραφείς αντικειμένων κατά τη διάρκεια της εγκατάστασης της MPEG-4 συνεδρίασης, έτσι ώστε ο αποκωδικο-



ποιητής να μπορεί να αποφασίσει εάν είναι ικανός να χειριστεί αυτή τη συνεδρίαση.

Με τη διαχείριση του πεπερασμένου μεγέθους του χώρου της προσωρινής μνήμης το μοντέλο επιτρέπει σε ένα αποστολέα, για παράδειγμα, να μεταφέρει δεδομένα μη πραγματικού χρόνου προκαταβολικά εάν είναι διαθέσιμος επαρκής χώρος στην πλευρά του δέκτη για να τα αποθηκεύσει. Τα προ-αποθηκευμένα δεδομένα μπορούν τότε να προσπελαστούν όταν χρειάζονται, επιτρέποντας εκείνη τη στιγμή η πληροφορία πραγματικού χρόνου να χρησιμοποιήσει ένα μεγαλύτερο ποσό από τη χωρητικότητα του καναλιού εάν είναι αυτό επιθυμητό.

- **Αναγνώριση χρόνου**

Για λειτουργία σε πραγματικό χρόνο, θεωρούμε ένα χρονικό μοντέλο στο οποίο η καθυστέρηση από τέλος-σε-τέλος από την έξοδο σήματος από ένα κωδικοποιητή έως την είσοδο σήματος σε ένα αποκωδικοποιητή είναι σταθερή. Επιπλέον, τα ρεύματα δεδομένων που εκπέμφθηκαν πρέπει να περιέχουν έμμεση ή άμεση χρονική πληροφορία. Υπάρχουν δύο είδη χρονικής πληροφορίας. Η πρώτη χρησιμοποιείται για τη μεταφορά της ταχύτητας του ρολογιού του κωδικοποιητή, ή τη χρονικά βασισμένη, στον αποκωδικοποιητή. Η δεύτερη, που αποτελείται από χρονικές σφραγίδες προσκολλημένες σε τμήματα των κωδικοποιημένων AV δεδομένων, περιέχει τον επιθυμητό χρόνο αποκωδικοποίησης για τις μονάδες προσπέλασης ή τον χρόνο σύνθεσης και λήξης για τις μονάδες σύνθεσης. Αυτή η πληροφορία μεταβιβάζεται σε SL-πακεταρισμένες επικεφαλίδες που δημιουργούνται στο στρώμα συγχρονισμού. Με αυτή τη χρονική πληροφορία, η απόσταση μέσα στην εικόνα και ο ακουστικός ρυθμός δειγματοληψίας μπορούν να προσαρμοστούν στον αποκωδικοποιητή για να ταιριάξουν με την απόσταση μέσα στην εικόνα και τον ακουστικό ρυθμό δειγματοληψίας για συγχρονισμένη λειτουργία.

Διαφορετικά αντικείμενα μέσω μπορεί να κωδικοποιηθούν από κωδικοποιητές με διαφορετικές χρονικές βάσεις, με την επακόλουθη ελαφρώς διαφορετική ταχύτητα. Είναι πάντα δυνατό να απεικονίσουμε αυτές τις χρονικές βάσεις στη χρονική βάση του τερματικού λήψης. Σε αυτή την περίπτωση, πάντως, καμία πραγματική υλοποίηση ενός τερματικού λήψης δεν μπορεί να αποφύγει την περιστασιακή επανάληψη ή απώλεια των AV δεδομένων, εξαιτίας της προσωρινής αλλοίωσης (σχετική μείωση ή επέκταση της χρονικής τους κλίμακας).

### 6.3.3 Περιγραφή σύνταξης

Το MPEG-4 ορίζει μία συντακτική γλώσσα περιγραφής για να περιγράψει την ακριβή διαδική σύνταξη ενός ρεύματος bit που μεταφέρει δεδομένα για μία αναπαράσταση αντικειμένου μέσου, όπως επίσης και της πληροφορίας περιγραφής του σκηνικού. Αυτό είναι μία απομάκρυνση από την παραδοσιακή προσέγγιση του MPEG με τη χρησιμοποίηση ψεύδο C. Αυτή η γλώσσα είναι

μία επέκταση της C++, και χρησιμοποιείται για να περιγράψει την συντακτική αναπαράσταση των αντικειμένων και τους ορισμούς όλων των κλάσεων των αντικειμένων μέσω και της πληροφορίας περιγραφής σκηνικού με ένα ολοκληρωμένο τρόπο. Αυτό παρέχει ένα σταθερό και ενιαίο τρόπο για να περιγράψουμε τη σύνταξη με μία πολύ ακριβή μορφή, ενώ συντοχρόνως να απλοποιούμε τις δοκιμές συμμόρφωσης των ρευμάτων bit. Εργαλεία λογισμικού μπορούν να χρησιμοποιηθούν για να επεξεργαστούμε τη συντακτική περιγραφή και να δημιουργήσουμε τον απαραίτητο κώδικα για προγράμματα που εκτελούν τον έλεγχο εγκυρότητας.

### 6.3.4 Κωδικοποίηση ηχητικών αντικειμένων

Η MPEG-4 κωδικοποίηση των ακουστικών αντικειμένων παρέχει εργαλεία και αναπαράσταση φυσικών ήχων (όπως ομιλία και μουσική) και για σύνθεση ήχων βασισμένων σε κατασκευασμένες περιγραφές. Η αναπαράσταση για το συνθετικό ήχο μπορεί να προέλθει από κείμενα δεδομένων ή τις αποκαλούμενες περιγραφές οργάνων και με την κωδικοποίηση παραμέτρων για να παρέχουμε αποτελέσματα, όπως αντήχηση και χωρητικότητα. Οι αναπαραστάσεις παρέχουν συμπίεση και άλλες λειτουργικότητες, όπως ικανότητα κλιμάκωσης και αποτελέσματα επεξεργασίας.

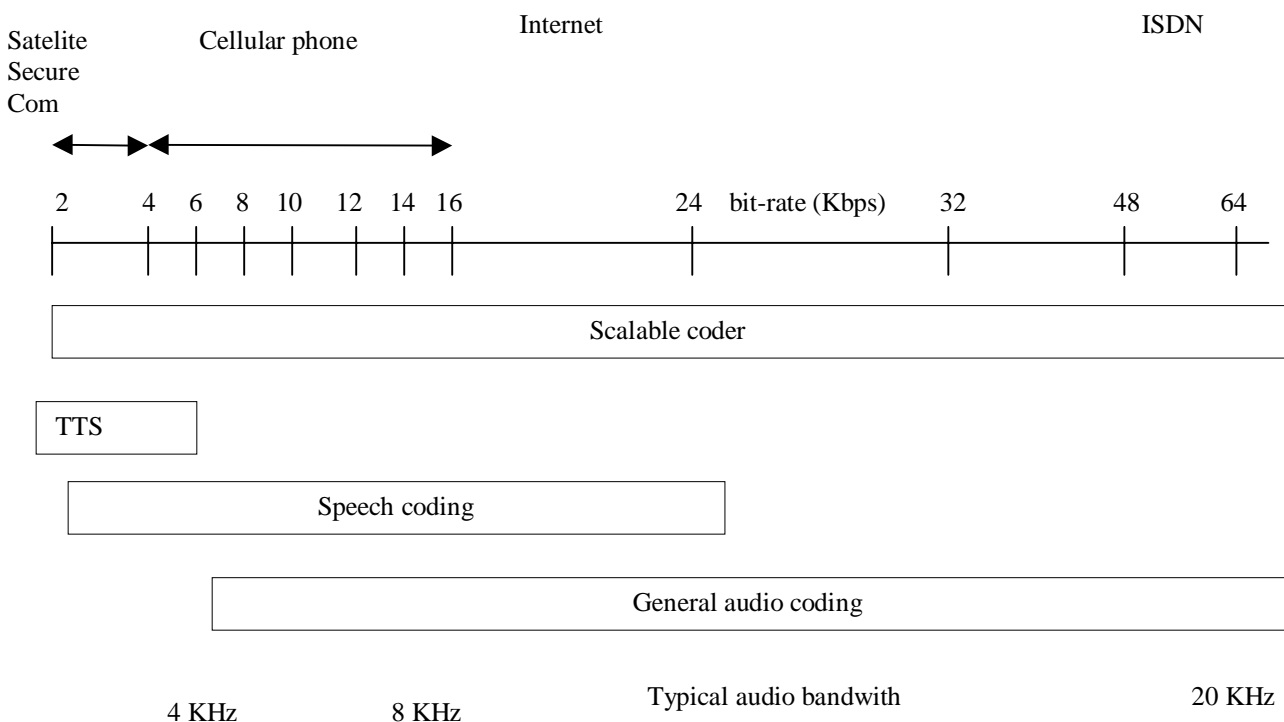
Τα εργαλεία κωδικοποίησης MPEG-4 Ήχου καλύπτοντας μία περιοχή από 64kbit/s έως 24kbit/s έχουν περάσει δοκιμές επαλήθευσης για εφαρμογή AM ψηφιακής εκπομπής ήχου σε συνεργασία με το NADIB (Narrow Band Digital Broadcasting). Με σκοπό την αναγνώριση μιας κατάλληλης μορφοποίησης εκπομπής ψηφιακού ήχου για να παρέχουμε βελτιώσεις πάνω στις υπάρχουσες υπηρεσίες AM διαμόρφωσης, πολλές προτάσεις συστημάτων κωδικοποιητή/ αποκωδικοποιητή που εμπεριέχουν τα MPEG-4 CELP, Twin VQ και AAC εργαλεία έχουν συγκριθεί σε ένα σύστημα αναφοράς AM. Βρέθηκε ότι η υψηλότερη ποιότητα μπορεί να επιτευχθεί μέσα στο ίδιο εύρος ζώνης με τις ψηφιακές τεχνικές και ότι οι κλιμακούμενες διαμορφώσεις κωδικοποιητή πρόσφεραν ανώτερη απόδοση σε εναλλακτικές προτάσεις ταυτόχρονης εκπομπής. Επιπρόσθετες δοκιμές επαλήθευσης πραγματοποιήθηκαν από το MPEG, στις οποίες τα εργαλεία για ομιλία και γενική κωδικοποίηση ήχου συγκρίθηκαν με τα ήδη υπάρχοντα.

- **Φυσικός ήχος**

Το MPEG-4 τυποποιεί την κωδικοποίηση του φυσικού ήχου σε ρυθμούς bit που εκτείνονται από 2kbit/s μέχρι και παραπάνω από 64kbit/s. Όταν επιτρέπεται μεταβλητός ρυθμός κωδικοποίησης, η κωδικοποίηση σε λιγότερο από 2kbit/s, όπως ένας μέσος ρυθμός bit των 1.2kbit/s, υποστηρίζεται επίσης. Η παρουσία του MPEG-2 AAC προτύπου μέσα στο MPEG-4 σύνολο εργαλείων θα παρέχει γενική συμπίεση ήχου στο παραπάνω εύρος ρυθμού bit. Για αυτά, το MPEG-4 πρότυπο ορίζει τη σύνταξη του ρεύματος bit και την επεξεργασία

αποκωδικοποίησης με όρους ενός συνόλου εργαλείων. Για να επιτύχουμε την υψηλότερη ποιότητα ήχου μέσα σε όλη την ακτίνα των ρυθμών bit και ταυτόχρονα να παρέχουμε τις επιπλέον λειτουργικότητες, οι τεχνικές κωδικοποίησης ομιλίας και οι τεχνικές γενικής κωδικοποίησης ήχου ολοκληρώνονται σε μία κοινή δομή:

- Η κωδικοποίηση ομιλίας σε ρυθμούς μεταξύ 2 και 24kbit/s υποστηρίζεται με τη χρήση του HVXC (Harmonic Vector eXcitation Coding) για ένα συνιστώμενο ρυθμό λειτουργίας από 2 έως 4kbit/s και του CELP (Code Excited Linear Predictive) για ρυθμός λειτουργίας από 4 έως 24kbit/s. Επιπλέον το HVXC μπορεί να λειτουργήσει χαμηλότερα σε μέσο ρυθμό 1.2kbit/s στην έκδοση του μεταβλητού ρυθμού bit. Στην κωδικοποίηση CELP, δύο ρυθμοί δειγματοληψίας, 8 και 16 KHz, χρησιμοποιούνται για να υποστηρίξουν ομιλία στενής και ευρείας ζώνης, αντίστοιχα. Οι ακόλουθοι τρόποι λειτουργίας έχουν γίνει το θέμα στη δοκιμή επαλήθευσης: το HVXC στα 2 και 4kbit/s, το CELP στενής ζώνης στα 6, 8.3 και 12kbit/s και το CELP ευρείας ζώνης στα 18kbit/s. Επιπλέον διάφορες κλιμακούμενες διαμορφώσεις έχουν επαληθευτεί.
- Για κωδικοποίηση γενικού ήχου σε ρυθμούς bit έως και παραπάνω από 6kbit/s, εφαρμόζονται τεχνικές κωδικοποίησης μεταφοράς που ονομάζονται Twin VQ και AAC. Τα ηχητικά σήματα σε αυτή την περιοχή τυπικά έχουν συχνότητες δειγματοληψίας ξεκινώντας στα 8KHz.



Σχήμα 33.

Για να επιτρέψουμε βέλτιστη κάλυψη των ρυθμών bit και να επιτρέψουμε δυνατότητα κλιμάκωσης για το τυθμό bit και το εύρος ζώνης, έχει οριστεί μία γενική δομή. Αυτή φαίνεται στο σχήμα 33.

Ξεκινώντας με ένα κωδικοποιητή να λειτουργεί σε χαμηλό ρυθμό bit, προσθέτοντας επαυξήσεις σε ένα γενικό κωδικοποιητή ήχου, και η ποιότητα κωδικοποίησης όπως επίσης και το εύρος ζώνης του ήχου μπορούν να βελτιωθούν.

Η κλιμάκωση του ρυθμού bit, που συχνά αναφέρεται ως εμπολιασμένη κωδικοποίηση, επιτρέπει ένα ρεύμα bit να αναλυθεί σε ένα ρεύμα bit χαμηλότερου ρυθμού το οποίο μπορεί ακόμα να αποκωδικοποιηθεί σε ένα σήμα με νόημα. Η ανάλυση του ρεύματος bit μπορεί να συμβεί είτε κατά τη διάρκεια της μετάδοσης ή στον αποκωδικοποιητή. Η κλιμάκωση του εύρους ζώνης είναι μία ιδιαίτερη περίπτωση κλιμάκωσης ρυθμού bit όπου τμήμα ενός ρεύματος bit που αναπαριστά ένα τμήμα του φάσματος συχνότητας μπορεί να απομακρυνθεί κατά τη διάρκεια της μετάδοσης ή της αποκωδικοποίησης.

Η κλιμάκωση της πολυπλοκότητας του κωδικοποιητή επιτρέπει σε κωδικοποιητές διαφορετικής πολυπλοκότητας να δημιουργήσουν έγκυρα και με νόημα ρεύματα bit. Η κλιμάκωση πολυπλοκότητας αποκωδικοποιητή επιτρέπει σε ένα δοσμένο ρεύμα bit να αποκωδικοποιηθεί από αποκωδικοποιητές διαφορετικών επιπέδων πολυπλοκότητας. Η ποιότητα ήχου, γενικά, σχετίζεται με την πολυπλοκότητα του κωδικοποιητή και αποκωδικοποιητή που χρησιμοποιούνται. Η κλιμάκωση δουλεύει μέσα σε μερικά εργαλεία MPEG-4, αλλά μπορεί επίσης να εφαρμοσθεί σε ένα συνδυασμό από τεχνικές, π.χ. το CELP ως βάση στρώμα και το AAC για τα στρώματα επαύξησης.

Το MPEG-4 στρώμα συστημάτων διευκολύνει τη χρήση και σηματοδότηση διαφορετικών εργαλείων και επομένως επίσης και των συστημάτων κωδικοποιητών / αποκωδικοποιητών σύμφωνα με τα υπάρχοντα πρότυπα, π.χ. MPEG-2 AAC. Καθένας από τους MPEG-4 κωδικοποιητές σχεδιάζεται για να λειτουργήσει με αυθύπαρκτο (stand-alone) τρόπο με τη δική του σύνταξη ρεύματος bit.

Επιπρόσθετες λειτουργικότητες γίνονται κατανοητές και μέσα σε ανεξάρτητους κωδικοποιητές και με μέσα επιπρόσθετων εργαλείων γύρω από τους κωδικοποιητές. Ένα παράδειγμα μιας λειτουργικότητας μέσα σε ένα ανεξάρτητο κωδικοποιητή είναι η αλλαγή της ομιλίας ή του τόνου της ομιλίας μέσα στο HVXC.

- **Συνθετεμένος ήχος (ή συνθετικός ήχος)**

Οι αποκωδικοποιητές είναι επίσης διαθέσιμοι για τη δημιουργία ήχου βασισμένου σε κατασκευασμένες εισόδους. Η είσοδος κειμένου μετατρέπεται σε ομιλία στον TTS (Text to Speech) αποκωδικοποιητή, ενώ πιο γενικοί ήχοι

συμπεριλαμβανομένου τη μουσική μπορεί να έχουν κανονιστικά συντεθεί. Η συνθετική μουσική μπορεί να παραδοθεί σε τρομερά χαμηλούς ρυθμούς bit ενώ ακόμα περιγράφει ένα ακριβές ηχητικό σήμα.

Οι TTS κωδικοποιητές εκτείνονται από 200 bit/s έως 1.2kbit/s, το οποίο επιτρέπει σε ένα κείμενο ή σε ένα κείμενο με μετρικές παραμέτρους (περίγραμμα τόνου φωνής, διάρκεια φωνήματος κλπ) ως τις εισόδους του για να δημιουργήσει κατανοητή συνθετική ομιλία. Υποστηρίζει τη δημιουργία παραμέτρων οι οποίες μπορούν να χρησιμοποιηθούν για να επιτρέψουν συγχρονισμό σε ενωμένη κίνηση προσώπου, διεθνείς γλώσσες για κείμενο και διεθνή σύμβολα για φωνήματα. Επιπρόσθετα βοηθήματα χρησιμοποιούνται για να μεταβιβάζουν πληροφορία ελέγχου μέσα σε κείμενα η οποία προωθείται σε άλλα συστατικά σε συγχρονισμό με το συνεθετημένο κείμενο. Σημειώστε ότι το MPEG-4 παρέχει μια τυποποιημένη διεπαφή για τη λειτουργία ενός TTS κωδικοποιητή (TTSI=Text To Speech Interface) παρά ένα κανονιστικό TTS συνθέτη μόνο του. Τα εργαλεία Κατασκευασμένου Ήχου αποκωδικοποιούν δεδομένα εισόδου και παράγουν ήχους εξόδου. Αυτή η αποκωδικοποίηση καθοδηγείται από μία ειδική γλώσσα σύνθεσης που καλείται SAOL (Structured Audio Orchestra Language), τυποποιημένη ως μέρος του MPEG-4. Αυτή η γλώσσα χρησιμοποιείται για να ορίσει μία «ορχήστρα» φτιαγμένη από «όργανα» (κατεβασμένα στο ρεύμα bit, όχι τοποθετημένα στο τερματικό) τα οποία δημιουργούν και επεξεργάζονται τα δεδομένα ελέγχου. Ένα όργανο είναι ένα μικρό δίκτυο από πρωτόγονα επεξεργασία σήματος που μπορούν να προσομοιώσουν μερικούς ειδικούς ήχους σαν εκείνους ενός φυσικού ακουστικού οργάνου. Το δίκτυο επεξεργασίας σήματος μπορεί να υλοποιηθεί με λογισμικό ή υλικό και περιλαμβάνει και δημιουργία και επεξεργασία ήχων και μεταχείριση των προ-αποθηκευμένων ήχων.

Το MPEG-4 δεν τυποποιεί «μία μέθοδο» σύνθεσης, αλλά μάλλον μία μέθοδο περιγραφής της σύνθεσης. Οποιαδήποτε υπάρχουσα ή μελλοντική μέθοδος σύνθεσης ήχου μπορεί να περιγραφεί με τη SAOL, συμπεριλαμβανομένου κυματομήτρης, FM προσθετικά, φυσική μοντελοποίηση και σπυρωτή σύνθεση, όπως επίσης και μη παραμετρικά υβριδικά αυτών των μεθόδων. Η έλεγχος της σύνθεσης επιτυγχάνεται με το κατέβασμα «σκορ» ή «σεναρίων» στο ρεύμα bit. Ένα σκορ είναι ένα χρονικά διατεταγμένο σύνολο εντολών που αιτεί από διάφορα όργανα σε καθορισμένες χρονικές στιγμές να συνεισφέρουν την έξοδο τους σε μία ολική μουσική παράσταση ή δημιουργία ηχητικών αποτελεσμάτων. Η περιγραφή του σκορ, κατεβασμένη σε μία γλώσσα που καλείται SASL (Structured Audio Score Language), μπορεί να χρησιμοποιηθεί να δημιουργήσει νέους θορύβους και επίσης να συμπεριλάβει επιπρόσθετη πληροφορία ελέγχου για τροποποίηση του υπάρχοντος θορύβου. Αυτό επιτρέπει στον συνθέτη να έχει τελειότερο έλεγχο πάνω στον τελικό συνεθετημένο ήχο. Για την επεξεργασία της σύνθεσης η οποία δεν απαιτεί τέτοιο τέλειο έλεγχο, το εγκατεστημένο MIDI πρωτόκολλο μπορεί επίσης να χρησιμοποιηθεί να ελέγχει την ορχήστρα.

Ο προσεκτικός έλεγχος σε συνδυασμό με τον συνήθη ορισμό οργάνων, επιτρέπει τη δημιουργία ήχων που εκτείνονται από απλά ηχητικά εφέ, όπως βήματα ή κλείσιμο πόρτας, την προσομοίωση ηχητικών θορύβων όπως η βροχόπτωση ή το παίξιμο μουσικής από παραδοσιακά όργανα μέχρι πλήρως συνθετικούς θορύβους για σύνθετα ηχητικά εφέ ή φουτουριστική μουσική. Για τερματικά με λιγότερη λειτουργικότητα και για εφαρμογές οι οποίες δεν απαιτούν τέτοια εξεζητημένη σύνθεση, τυποποιείται επίσης μία «μορφοποίηση τράπεζας κυματοπινάκων». Χρησιμοποιώντας αυτή τη μορφοποίηση, ηχητικά δείγματα για χρήση σε σύνθεση κυματοπίνακα μπορούν να κατέβουν, ως απλή επεξεργασία, όπως φίλτρα, ακουστικές ανακλάσεις, και χορωδιακά εφέ. Σε αυτή την περίπτωση, η υπολογιστική πολυπλοκότητα της απαιτούμενης επεξεργασίας αποκωδικοποίησης μπορεί να καθοριστεί ακριβώς από την επιθεώρηση του ρεύματος bit, το οποίο δεν είναι δυνατό όταν χρησιμοποιείται η SAOL.

### **6.3.5 Αναγνώριση και προστασία του IPR συσχετισμένου με το περιεχόμενο**

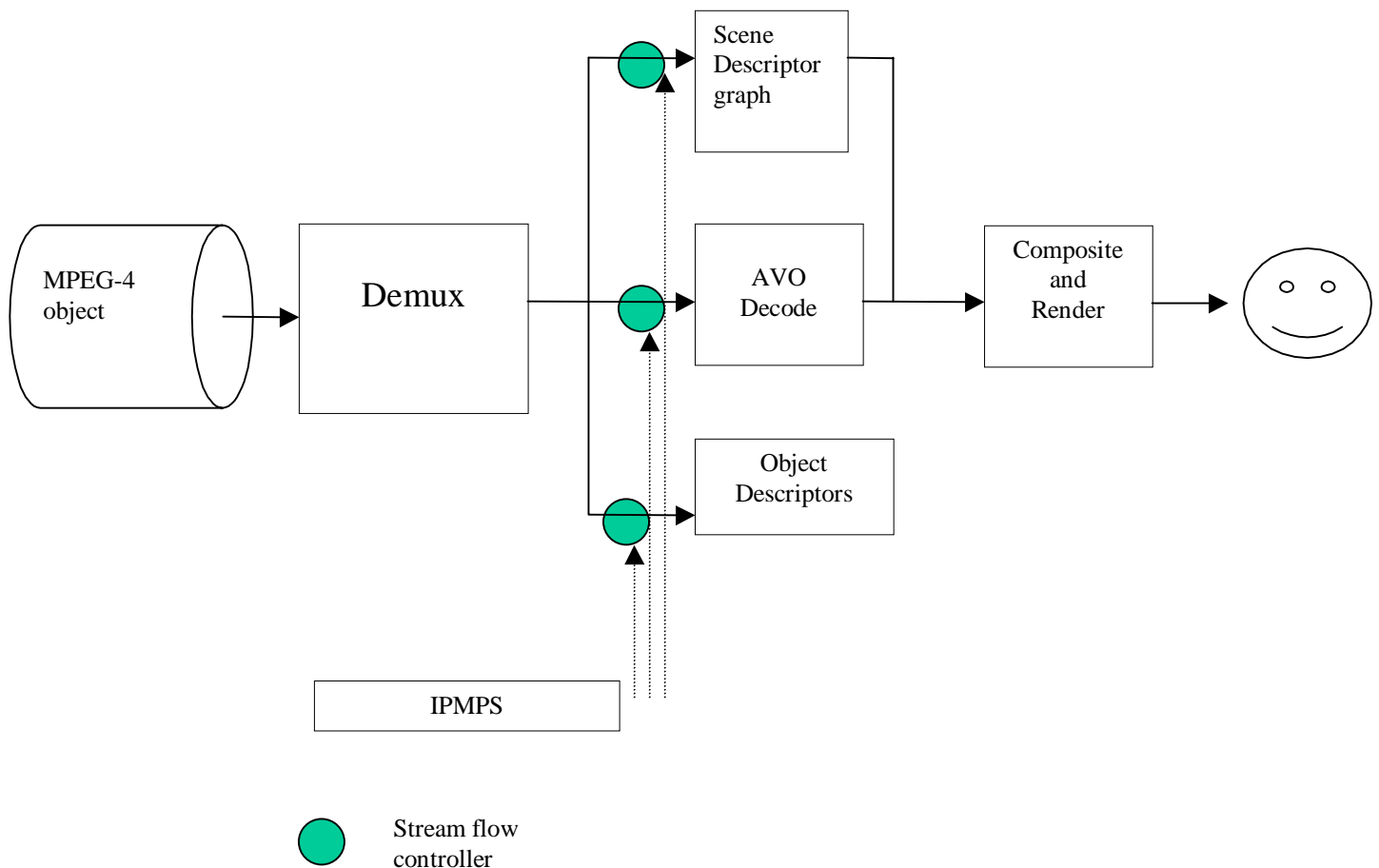
Το MPEG-4 παρέχει μηχανισμούς για προστασία των δικαιωμάτων πνευματικής ιδιοκτησίας (IPR = Intellectual Property Rights), όπως αναφέρθηκε στην ενότητα 1.5. Αυτό επιτυγχάνεται εφοδιάζοντας τα κωδικοποιημένα αντικείμενα μέσω με ένα IPR σύνολο δεδομένων, που μεταφέρει πληροφορία σχετικά με τα περιεχόμενα, το είδος των περιεχομένων και (δείκτες σε) τους κατόχους των δικαιωμάτων. Το σύνολο δεδομένων, αν είναι παρόν, είναι μέρος ενός περιγραφέα βασικού ρεύματος που περιγράφει τα δεδομένα ρεύματος που είναι ενωμένα με ένα αντικείμενο μέσου. Ο αριθμός των συνόλων δεδομένων που πρόκειται να ενωθεί με κάθε αντικείμενο μέσου είναι ευέλικτος. Διαφορετικά αντικείμενα μέσω μπορούν να μοιραστούν τα ίδια σύνολα δεδομένων ή να έχουν ξεχωριστά σύνολα δεδομένων. Η πρόβλεψη των συνόλων δεδομένων επιτρέπει την υλοποίηση μηχανισμών για ακουστική παρακολούθηση, εποπτεία, χρέωση και προστασία αντιγραφής.

Μετά από την αναγνώριση των δικαιωμάτων, κάθε μία από τις εφαρμογές του MPEG-4 έχει ένα σύνολο από απαιτήσεις λαμβάνοντας υπόψη την προστασία της πληροφορίας που διαχειρίζεται. Αυτές οι εφαρμογές μπορούν να έχουν διαφορετικές απαιτήσεις ασφάλειας. Για μερικές εφαρμογές, οι χρήστες ανταλλάσσουν πληροφορία η οποία δεν έχει ουσιαστική αξία αλλά πρέπει παρ' όλα αυτά να προστατευθεί για να διατηρήσει τα δικαιώματα της ησυχίας/μη ενόχλησης. Για άλλες εφαρμογές, η διαχειριζόμενη πληροφορία έχει μεγάλη αξία για το δημιουργό της και/ ή του διανομείς απαιτώντας διαχείριση μεγάλου βαθμού και μηχανισμούς προστασίας. Εδώ εννοείται ότι η σχεδίαση της δομής του IPMP πρέπει να υπολογίσει την πολυπλοκότητα του MPEG-4 προτύπου και την ποικιλία των εφαρμογών του. Αυτή η IPMP δομή αφήνει τις λεπτομέρειες των IPMP σχεδιάσεων συστημάτων στα χέρια των βελτιωτών εφαρμογών. Το επίπεδο και το είδος της διαχείρισης και προστασίας που απαιτείται εξαρτάται

στην αξία του περιεχομένου, την πολυπλοκότητα και τη φιλοσοφία των ενωμένων επιχειρηματικών μοντέλων.

Η προσέγγιση που έγινε επιτρέπει τη σχεδίαση και χρήση IPMP συστημάτων καθορισμένου πεδίου (IPMP-S). Ενώ το MPEG-4 δεν τυποποιεί τα IPMP συστήματα από μόνα τους, εντούτοις τυποποιεί την MPEG-4 IPMP διεπαφή. Αυτή η διεπαφή αποτελείται από IPMP-περιγραφείς (IPMP-D) και IPMP- Βασικά Ρεύματα (IPMP-ES).

Τα IPMP-D και IPMP-ES παρέχουν ένα μηχανισμό επικοινωνίας μεταξύ των IPMP συστημάτων και του MPEG-4 τερματικού. Ορισμένες εφαρμογές μπορεί να απαιτούν πολλαπλά IPMP συστήματα. Όταν τα MPEG-4 αντικείμενα απαιτούν διαχείριση και προστασία, έχουν ενωμένα μαζί τους τα IPMP-D. Αυτά τα IPMP-D υποδεικνύουν ποια IPMP συστήματα πρόκειται να χρησιμοποιηθούν και παρέχουν πληροφορία σε αυτά τα συστήματα για το πώς να διαχειριστούν και να προστατέψουν το περιεχόμενο. (βλ Σχήμα 34).



Σχήμα 34. Η IPMP διεπαφή μέσα σε ένα MPEG-4 σύστημα

Εκτός από το να καθιστά ικανούς τους ιδιοκτήτες της πνευματικής ιδιοκτησίας να διαχειρίζονται και να προστατεύουν τα δικαιώματά τους, το MPEG-4 παρέχει ένα μηχανισμό να αναγνωρίζει αυτά δικαιώματα μέσω του συνόλου

δεδομένων αναγνώρισης πνευματικής ιδιοκτησίας (IPI Data Set). Αυτή η πληροφορία μπορεί να χρησιμοποιηθεί από τα IPMP συστήματα ως είσοδος στη διαδικασία διαχείρισης και προστασίας.

#### **6.4 Λίστα των κύριων λειτουργικοτήτων που παρέχονται από το MPEG-4 Ήχου στην έκδοση 1 για τον ήχο**

Το MPEG-4 Ήχου διευκολύνει μια πλατειά ποικιλία εφαρμογών οι οποίες μπορεί να εκτείνονται από κατανοητή ομιλία μέχρι υψηλής ποιότητας πολυκαναλικό ήχο και από φυσικούς ήχους σε συνθετικούς ήχους. Συγκεκριμένα υποστηρίζει την υψηλώς επαρκή αναπαράσταση των ηχητικών αντικειμένων που αποτελούνται από:

- Σήματα ομιλίας: Η κωδικοποίηση ομιλίας μπορεί να γίνει χρησιμοποιώντας ρυθμούς από 2kbit/s μέχρι 24kbit/s χρησιμοποιώντας τα εργαλεία κωδικοποίησης ομιλίας. Χαμηλότεροι ρυθμοί, όπως ένας μέσος ρυθμός του 1.2kbit/s είναι επίσης δυνατοί όταν επιτρέπεται μεταβλητός ρυθμός κωδικοποίησης. Μικρή καθυστέρηση είναι πιθανή για εφαρμογές επικοινωνίας. Όταν χρησιμοποιούμε τα HVXC εργαλεία, η ομιλία και ο τόνος μπορούν να τροποποιηθούν υπό τον έλεγχο του χρήστη κατά τη διάρκεια της εκτέλεσης (playback). Αν χρησιμοποιούνται τα CELP εργαλεία, μπορεί να επιτευχθεί αλλαγή της ταχύτητας εκτέλεσης με χρήση και επιπλέον εργαλείου για επεξεργασία εφέ.
- Συνθετική ομιλία: Ο ρυθμός των κλιμακούμενων TTS κωδικοποιητών εκτείνεται από 200bit/s έως 1.2kbit/s το οποίο επιτρέπει σε ένα κείμενο ή σε ένα κείμενο με μετρικές παραμέτρους (περίγραμμα τόνου φωνής, διάρκεια φωνήματος κλπ), ως εισόδους του για να δημιουργήσει κατανοητή συνθετική ομιλία. Περιλαμβάνει τις ακόλουθες λειτουργικότητες:
  - Σύνθεση ομιλίας χρησιμοποιώντας την μετρική της πρωτότυπης ομιλίας
  - Έλεγχος συγχρονισμού χειλιών με πληροφορία φωνήματος
  - Λειτουργικότητα με τεχνάσματα: παύση, συνέχιση, πήδημα μπρος/πίσω
  - Υποστήριξη διεθνών γλωσσών και διαλέκτων για κείμενο (μπορεί να σηματοδοτηθεί στο ρεύμα bit ποια γλώσσα και διάλεκτος πρέπει να χρησιμοποιηθεί)
  - Υποστήριξη διεθνών συμβόλων για φωνήματα
  - Υποστήριξη για προσδιορισμό της ηλικίας, γένους και ρυθμού ομιλίας του ομιλητή
  - Υποστήριξη για μεταφορά των παραμέτρων της κίνησης του προσώπου (FAP = Facial Animation Parameter)
- Σήματα γενικού ήχου: Η υποστήριξη για κωδικοποίηση γενικού ήχου που εκτείνεται από πολύ χαμηλούς ρυθμούς μέχρι υψηλής ποιότητας παρέχεται μετατρέποντας τις τεχνικές κωδικοποίησης. Με αυτή τη λειτουργικότητα καλύπτεται μία μεγάλη περιοχή ρυθμών και εύρων ζώνης. Ξεκινά σε ένα



ρυθμό των 6kbit/s και ένα εύρος ζώνης κάτω από 4KHz αλλά επίσης περιλαμβάνει ποιότητα εκπομπής ήχου από μονό μέχρι πολυκαναλικού.

- Συνθετικός ήχος: Η υποστήριξη του συνθετικού ήχου παρέχεται από μία υλοποίηση Αποκωδικοποιητή Κατασκευασμένου Ήχου που επιτρέπει την εφαρμογή πληροφορίας ελέγχου βασισμένης στα σκορ, σε μουσικά όργανα που περιγράφονται σε μία ειδική γλώσσα
- Συνθετικός Ήχος Περιορισμένης Πολυπλοκότητας: αυτό παρέχεται από μία υλοποίηση Αποκωδικοποιητή Κατασκευασμένου Ήχου που επιτρέπει την επεξεργασία μιας τυποποιημένης μορφοποίησης κυματοπίνακα.

Παραδείγματα επιπλέον λειτουργικότητας είναι η αλλαγή του ελέγχου της ταχύτητας και του τόνου για σήματα ομιλίας και η κλιμάκωση με όρους όπως ρυθμός bit, εύρος ζώνης, ανθεκτικότητα σε σφάλματα, πολυπλοκότητα κλπ όπως ορίζονται παρακάτω

- Η λειτουργικότητα της αλλαγής ταχύτητας επιτρέπει την αλλαγή της κλίμακας χρόνου χωρίς να αλλάζουμε τον τόνο κατά τη διάρκεια της διεργασίας αποκωδικοποίησης. Αυτό μπορεί, για παράδειγμα, να χρησιμοποιηθεί για να υλοποιήσουμε μια λειτουργία γρήγορης μετακίνησης προς τα εμπρός “fast forward” (για ψάξιμο βάσης δεδομένων) ή για να προσαρμόσουμε το μήκος μιας ηχητικής ακολουθίας σε μια δοσμένη οπτική ακολουθία, ή για να εξασκηθούμε σε βήματα χορού με μικρότερη ταχύτητα.
- Η λειτουργικότητα αλλαγής του τόνου επιτρέπει την αλλαγή του τόνου χωρίς να αλλάζουμε την κλίμακα χρόνου κατά τη διάρκεια της διεργασίας κωδικοποίησης και αποκωδικοποίησης. Αυτό μπορεί να χρησιμοποιηθεί, για παράδειγμα, για αλλαγή φωνής ή εφαρμογές τύπου Karaoke (εγγραφή φωνής σε μελωδία). Αυτή η τεχνική εφαρμόζεται μόνο σε μεθόδους κωδικοποίησης παραμετρικού και κατασκευασμένου θορύβου.
- Η κλιμάκωση ρυθμού bit επιτρέπει σε ένα ρεύμα bit να αναλυθεί σε ένα ρεύμα bit χαμηλότερου ρυθμού έτσι ώστε ο συνδυασμός να μπορεί ακόμα να αποκωδικοποιηθεί σε ένα σήμα με νόημα. Η ανάλυση του ρεύματος bit μπορεί να συμβεί είτε κατά τη διάρκεια της μετάδοσης ή στον αποκωδικοποιητή.
- Η κλιμάκωση του εύρους ζώνης είναι μία ειδική περίπτωση κλιμάκωσης ρυθμού bit, όπου μέρος ενός ρεύματος bit που αναπαριστά ένα μέρος του φάσματος συχνότητας μπορεί να αφαιρεθεί κατά τη διάρκεια της μετάδοσης ή της αποκωδικοποίησης.
- Η κλιμάκωση της πολυπλοκότητας του κωδικοποιητή επιτρέπει σε κωδικοποιητές διαφορετικής πολυπλοκότητας να δημιουργήσουν έγκυρα και με νόημα ρεύματα bit.
- Η κλιμάκωση της πολυπλοκότητας του αποκωδικοποιητή επιτρέπει σε ένα δοσμένο ρεύμα bit να αποκωδικοποιηθεί από αποκωδικοποιητές

διαφορετικών επιπέδων πολυπλοκότητας. Η ποιότητα ήχου, γενικά, σχετίζεται με την πολυπλοκότητα του κωδικοποιητή και αποκωδικοποιητή που χρησιμοποιούνται.

- Τα ηχητικά εφέ παρέχουν την ικανότητα να επεξεργαστούμε αποκωδικοποιημένα ηχητικά σήματα με πλήρη χρονική ακρίβεια για να επιτύχουμε λειτουργίες για μίξη, αντήχηση και κατανομή χώρου.

## 6.5 Δοκιμές επαλήθευσης: ελέγχοντας την απόδοση του MPEG

Το MPEG πραγματοποιεί δοκιμές επαλήθευσης για να ελέγξει εάν το πρότυπο αποδίδει ότι υπόσχεται.

Οι πρώτες τυπικές δοκιμές στους κωδικοποιητές του MPEG-4 Ήχου ολοκληρώθηκαν, βασισμένες στην συνεργασία μεταξύ του MPEG και του NADIB (Narrowband Digital Audio Broadcasting Group). Αυτές οι δοκιμές διερεύνησαν την απόδοση των κωδικοποιητών ομιλίας και μουσικής δουλεύοντας σε περιοχή ρυθμών από 6kbit/s έως 24kbit/s, συμπεριλαμβανομένου μερικές επιλογές κλιμακωτού κωδικοποιητή. Τα αποτελέσματα δείχνουν ότι μία σημαντική βελτίωση στην ποιότητα μπορεί να προσφερθεί σε σχέση με τη συμβατική αναλογική AM εκπομπή και ότι οι κλιμακωτοί κωδικοποιητές προσφέρουν ανώτερη απόδοση σε λειτουργίες ταυτόχρονης μετάδοσης.

Ετοιμάζονται περισσότερες δοκιμές επαλήθευσης. Για τον πλευρά του ήχου ετοιμάζονται δοκιμές για την ομιλία και το ήχο στο Internet. Οι δοκιμές ήχου και οι δοκιμές ανθεκτικότητας σε σφάλματα θα είναι έτοιμες τον Οκτώβριο, οι υπόλοιπες δοκιμές προγραμματίστηκαν για υλοποίηση το Δεκέμβριο.

## 6.6 Ηχητικά προφίλ στην MPEG-4 Έκδοση 1

Το MPEG-4 παρέχει ένα μεγάλο και πλούσιο σύνολο από εργαλεία για την κωδικοποίηση των οπτικοακουστικών αντικειμένων. Για να επιτρέψουμε αποτελεσματικές υλοποιήσεις του προτύπου, έχουν αναγνωρισθεί υποσύνολα των συνόλων εργαλείων από το MPEG-4 Συστημάτων, Εικόνας και Ήχου, τα οποία μπορούν να χρησιμοποιηθούν για συγκεκριμένες εφαρμογές. Αυτά τα υποσύνολα καλούνται «κατατομές» (“profiles”) και περιορίζουν το σύνολο των εργαλείων που έχει να υλοποιήσει ένας αποκωδικοποιητής. Για καθεμία από αυτές τις κατατομές, έχουν ορισθεί ένα ή περισσότερα επίπεδα, περιορίζοντας την υπολογιστική πολυπλοκότητα. Η προσέγγιση είναι όμοια με το MPEG-2, όπου ο πιο γνωστός συνδυασμός κατατομής/επιπέδου είναι «η κύρια κατατομή @ το κυρίως επίπεδο (“main profile @ main level)”. Ένας συνδυασμός κατατομής@επιπέδου επιτρέπει:

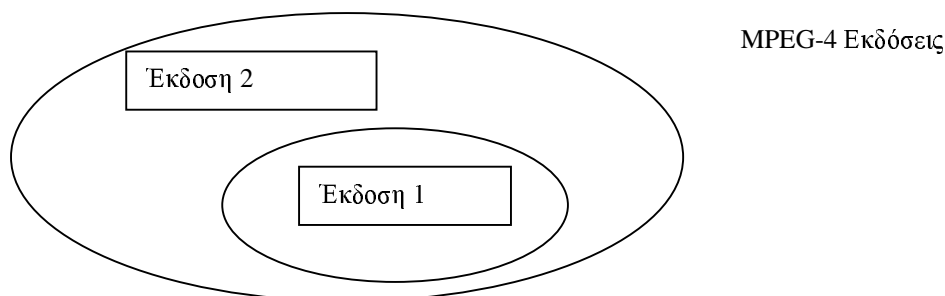
- Ένας κωδικοποιητής κτίστης να υλοποιήσει μόνο το υποσύνολο του προτύπου που χρειάζεται, διατηρώντας την εσωτερική δουλειά με άλλες MPEG-4 συσκευές κτισμένες στον ίδιο συνδυασμό και
- Να ελέγχει εάν οι MPEG-4 συσκευές συμμορφώνονται με το πρότυπο.

Τέσσερις Κατατομές Ήχου έχουν ορισθεί

- Η κατατομή ομιλίας παρέχει HVXC, το οποίο είναι ένας παραμετρικός κωδικοποιητής ομιλίας πολύ χαμηλού ρυθμού bit, έναν CELP στενής ζώνης / ευρείας ζώνης κωδικοποιητή ομιλίας και μία TTS διεπαφή
- Η κατατομή σύνθεσης παρέχει σύνθεση οδηγούμενη από σκορ χρησιμοποιώντας SAOL και κυματοπίνακες και μία TTS διεπαφή για να παράγει ήχο και ομιλία σε πολύ χαμηλούς ρυθμούς bit
- Η κλιμακωτή κατατομή, ένα υπερσύνολο της κατατομής ομιλίας, είναι κατάλληλη για κλιμακούμενη κωδικοποίηση ομιλίας και μουσικής για δίκτυα, όπως το Internet και το NADIB. Οι ρυθμοί bit εκτείνονται από 6kbit/s μέχρι 24kbit/s, με εύρη ζώνης μεταξύ 3.5 και 9 KHz.
- Η κυρίως κατατομή είναι ένα πλούσιο υπερσύνολο όλων των άλλων κατατομών, περιέχοντας εργαλεία για φυσικό και συνθετικό ήχο.

## 6.7 Έκδοση 2 του MPEG-4/Ήχου: Έτοιμο ένα χρόνο μετά την Έκδοση 1

Τον Οκτώβριο του 1998, το πρώτο σύνολο των MPEG-4 προτύπων πάγωσε. Η δουλειά πάνω στο MPEG-4 συνεχίζεται στην Έκδοση 2, η οποία θα προσθέσει εργαλεία στο MPEG-4 πρότυπο. Τα ήδη υπάρχοντα εργαλεία και οι κατατομές από την έκδοση 1 δεν θα αντικατασταθούν στην έκδοση 2. Η τεχνολογία που θα προστεθεί στο MPEG-4 θα είναι της μορφής νέων κατατομών. Το σχήμα 35 παρακάτω δείχνει τη σχέση μεταξύ των δύο εκδόσεων. Η έκδοση 2 είναι μια αναδρομικά συμβατή επέκταση της έκδοσης 1.



Σχήμα 35. Η σχέση μεταξύ των MPEG-4 εκδόσεων

Οι ακόλουθες επιπρόσθετες λειτουργικότητες θα υποστηριχθούν από την MPEG-4 Έκδοση 2:

- Μέθοδοι ελαστικότητας σφαλμάτων

Η έκδοση 2 θα προσθέσει νέα εργαλεία στους αλγόριθμους ήχου για να βελτιώσει την ελαστικότητα τους στα σφάλματα. Υπάρχουν δύο κατηγορίες εργαλείων: Η πρώτη κατηγορία περιέχει αλγόριθμους για να βελτιώσει την αντοχή της πηγής κωδικοποίησης μόνη της, π.χ. η Huffman κωδική λέξη να ξαναζητά AAC. Η δεύτερη κατηγορία αποτελείται από γενικά εργαλεία για προστασία από σφάλματα, επιτρέποντας ίση και άνιση προστασία από σφάλματα των κωδικοποιημένων σχημάτων του MPEG-4 Ήχου. Αφού αυτά τα εργαλεία βασίζονται σε συγκεκριματικούς κώδικες, επιτρέπουν πολύ εύκαμπτη χρήση διαφορετικών επιβαρύνσεων και ικανοτήτων διόρθωσης σφαλμάτων, φιλοξενώντας έτσι πολύ διαφορετικές καταστάσεις του καναλιού.

- Κατανομή περιβάλλοντος χώρου  
Αυτά τα νέα εργαλεία επιτρέπουν παραμετροποίηση των ακουστικών ιδιοτήτων ενός MPEG-4 σκηνικού (π.χ. ένα τρισδιάστατο μοντέλο ενός επιπλωμένου δωματίου ή μία σάλα συναυλιών) δημιουργημένη με τα BIFS εργαλεία περιγραφής σκηνικού. Τέτοιες ιδιότητες είναι, για παράδειγμα, ο χρόνος αντήχησης δωματίου, η ταχύτητα ήχου, οι σταθερές ιδιότητες του υλικού (ανάκλαση, μετάδοση) και η κατευθυντικότητα της πηγής ήχου. Η νέα λειτουργικότητα που έγινε δυνατή με αυτές τις παραμέτρους περιγραφής του σκηνικού περιλαμβάνει εξελιγμένη και φωλιασμένη οπτικοακουστική απόδοση, λεπτομερή ακουστική μοντελοποίηση δωματίου και επαυξημένη αναπαράσταση ήχου στις τρεις διαστάσεις.
- Κωδικοποίηση γενικού ήχου χαμηλής καθυστέρησης  
Αυτή η λειτουργικότητα υποστηρίζει μετάδοση γενικών ηχητικών σημάτων σε εφαρμογές με αμφίδρομη επικοινωνία. Συγκρινόμενη με την Έκδοση 1, θα παρέχεται μία σημαντικά μειωμένη καθυστέρηση κωδικοποίησης/αποκωδικοποίησης με μόνο μία ελαφριά μείωση της επάρκειας κωδικοποίησης.
- Σύνταξη για κανάλι επιστροφής (Backchannel), για προσαρμοστική κωδικοποίηση και εκτέλεση αντικειμένων Ήχου
- Κλιμάκωση μικρού βήματος  
Αυτό το εργαλείο επιτρέπει κλιμακούμενη κωδικοποίηση με πολύ εκλεπτισμένη σπύρωση, π.χ. παραγεμισμένη κωδικοποίηση με πολύ μικρά βήματα ρυθμού bit. Αυτό επιτυγχάνεται από το BSAC (Bit-Sliced Arithmetic Coding) σε συνδυασμό με τα εργαλεία κωδικοποίησης γενικού ήχου της έκδοσης 1.
- Παραμετρική κωδικοποίηση ήχου  
Αυτά τα εργαλεία προσφέρουν την δυνατότητα τροποποίησης της ταχύτητας εκτέλεσης ή του τόνου κατά τη διάρκεια της αποκωδικοποίησης εξαιτίας μίας παραμετρικής αναπαράστασης σήματος χωρίς την ανάγκη της μονάδας επεξεργασίας ειδικών εφέ. Ένας συνδυασμός με το HVXC εργαλείο θα είναι δυνατός, ο οποίος θα είναι επίσης βασισμένος σε μία παραμετρική αναπαράσταση σήματος. Επιπλέον, για εφαρμογές κωδικοποίησης

στηριγμένης σε αντικείμενα που επιτρέπουν επιλογή και/ή εναλλαγή μεταξύ διαφορετικών τεχνικών κωδικοποίησης, αναμένεται μία βελτίωση της ολικής επάρκειας κωδικοποίησης σε συνδυασμό με τα εργαλεία κωδικοποίησης γενικού ήχου της έκδοσης 1.

## **7. ΕΦΑΡΜΟΓΕΣ**

Οι MPEG τεχνολογίες συμπίεσης ήχου θα παίξουν ένα σημαντικό ρόλο στα καταναλωτικά ηλεκτρονικά, στον επαγγελματικό χώρο του ήχου, στις τηλεπικοινωνίες, στην εκπομπή και στα πολυμέσα. Εδώ περιγράφουμε μερικούς τυπικούς χώρους εφαρμογών.

Βασικές εφαρμογές θα στηριχθούν στην παράδοση ψηφιακών ηχητικών σημάτων από επίγεια και δορυφορικά ψηφιακά συστήματα εκπομπής και μετάδοσης όπως γραμμές συνδρομητών, δεσμούς ανταλλαγής προγραμμάτων, κυτταρώδη κινητά ραδιοφωνικά δίκτυα, δίκτυα συνδρομητικής τηλεόρασης και τοπικά δίκτυα [52]. Για παράδειγμα, στα στενής ζώνης δίκτυα ISDN (Integrated Services Digital Networks), οι πελάτες έχουν φυσική πρόσβαση σε ένα ή δύο 64-kb/s B κανάλια και ένα 16-kb/s D κανάλι (το οποίο υποστηρίζει σηματοδότηση αλλά μπορεί επίσης να μεταφέρει πληροφορία του χρήστη). Και άλλες διαμορφώσεις είναι δυνατές συμπεριλαμβάνοντας  $p \times 64 \text{ kb/s}$  ( $p=1,2,3,\dots$ ) υπηρεσίες. Οι ISDN ρυθμοί προσφέρουν χρήσιμα κανάλια για μία πρακτική κατανομή στερεοφωνικών και πολυκαναλικών σημάτων ήχου. Επειδή το ISDN είναι μία αμφίδρομη υπηρεσία, παρέχει επίσης αντίθετα προς το ρεύμα μονοπάτια για μελλοντικές απαιτούμενες και αλληλοεπιδρούμενες οπτικοακουστικές επιτόπου (just in time) υπηρεσίες ήχου. Η ραχοκοκαλιά των ψηφιακών τηλεπικοινωνιακών δικτύων θα είναι ευρείας ζώνης (B-) ISDN με κυψελοειδή δομή. Οι καθυστερήσεις κυψελών και οι απώλειες κυψελών είναι πηγές παραμόρφωσης που πρέπει να ληφθούν υπόψη σε σχεδιάσεις ψηφιακών συστημάτων ήχου [53]. Μία σχετική εφαρμογή είναι η εκπομπή μέσω Internet, η οποία θα χρειαστεί σημαντικούς ρυθμούς συμπίεσης όσο οι προσωπικοί υπολογιστές είναι συνδεδεμένοι στα δίκτυα ραχοκοκαλιές μέσω διαμορφωτών/αποδιαμορφωτών (modem) με συνήθεις ρυθμούς bit μεταξύ 14.4 kb/s και 33 kb/s.

Στο χώρο της ψηφιακής αποθήκευσης σε ψηφιακές ταινίες ήχου και δίσκους (επανεγγραφής), ένας αριθμός από καταναλωτικά προϊόντα βασισμένα στο MPEG έχει πρόσφατα προσέλθει στην αγορά ήχου. Από αυτά τα προϊόντα, το Digital Compact Cassette (DCC) της Philips, ουσιαστικά κάνει χρήση του Στρώματος I του MPEG-1 κωδικοποιητή ήχου, χρησιμοποιώντας το στερεοφωνικό ρυθμό του των 384 kb/s. Ο αλγόριθμος του κωδικοποίησης ήχου καλείται PASC (precision audio subband coding) [15]. Το ανερχόμενο DVD με χωρητικότητα 4,7 GB ελαττώνει την πίεση για εκπληκτικούς συντελεστές συμπίεσης. Θα ανοίξει τις δυνατότητες της αποθήκευσης καναλιών ήχου που έχουν κωδικοποιηθεί με ένα τρόπο χωρίς απώλειες και θα παρέχει την

απαραίτητη χωρητικότητα για διάφορες μορφές πολυκαναλικής κωδικοποίησης. Το DVD θα υποστηρίξει στερεοφωνικό και (τουλάχιστο) 5.1- πολυκαναλικό ήχο. Σε σύνδεση με video, ή έκδοση PAL του DVD (5625/50 TV System) θα χρησιμοποιεί MPEG κωδικοποίηση ήχου με Dolby AC-3 τεχνική μετασχηματισμού κωδικοποίησης ως επιλογή [54-56], ενώ η έκδοση NTSC (525/60 TV System) θα στηριχθεί στο AC-3 με το MPEG ως επιλογή. Ο ολικός ρυθμός bit ήχου είναι 384 kb/s.

Ένας αριθμός αποφάσεων που αφορούν υπηρεσίες όπως την εισαγωγή της ψηφιακής εκπομπής ήχου (DAB digital audio Broadcast), και ψηφιακής εκπομπής εικόνας (DVB digital video broadcast) έχουν παρθεί πρόσφατα. Στην Ευρώπη, μία ερευνητική ομάδα που καλείται Eureka 147 έχει μελετήσει ένα σύστημα DAB το οποίο είναι έτοιμο να αντιμετωπίσει τα προβλήματα της ψηφιακής εκπομπής [57]. Το ITU-R έχει προτείνει το MPEG-1 πρότυπο κωδικοποίησης ήχου αφού πραγματοποίησε εκτεταμένες υποκειμενικές δοκιμές. Το Στρώμα II αυτού του προτύπου χρησιμοποιείται για εκπομπή προγραμμάτων και η έκδοση του Στρώματος III προτείνεται για ερμηνευτικούς δεσμούς σε χαμηλούς ρυθμούς. Ο ρυθμός δειγματοληψίας είναι 48 KHz σε όλες τις περιπτώσεις και το πεδίο βοηθητικών δεδομένων χρησιμοποιείται για δεδομένα που σχετίζονται με το πρόγραμμα (Program-associated data –PAD information) και άλλα δεδομένα. Το σύστημα DAB έχει μία σημαντική επιβάρυνση ρυθμού bit για διόρθωση σφαλμάτων στηριγμένο σε χαλαρούς κώδικες συγκερασμού, για να υποστηρίξει κωδικοποίηση καναλιού προσαρμοσμένης πηγής, δηλαδή μία ανομοιόμορφη προστασία σφαλμάτων που είναι σε συμφωνία με την ευαισθησία μεμονωμένων bit ή μιας ομάδας από bits σε σφάλματα καναλιού [58]. Επιπλέον, τεχνικές απόκρυψης σφαλμάτων εφαρμόζονται ώστε να παρέχουν μία χαριτωμένη υποβάθμιση στην περίπτωση σοβαρών σφαλμάτων. Στις Ενωμένες Πολιτείες Αμερικής κανένα πρότυπο δεν έχει ακόμα οριστεί. Αναλογικές και ψηφιακές εκδόσεις ταυτόχρονης μετάδοσης του ίδιου ακουστικού προγράμματος στην FM γήινη ζώνη (88-108 MHz) είναι ένα σημαντικό θέμα (ενώ η Ευρωπαϊκή λύση στηρίζεται σε νέα κανάλια) [59].

Το δορυφορικό συνδρομητικό τηλεοπτικό σύστημα του Hughes DirectTV και το ADR (Astra Digital Radio) είναι παραδείγματα ψηφιακής εκπομπής βασισμένης σε δορυφόρους που κάνουν χρήση του MPEG-1 Στρώματος II. Ως ένα ακόμα παράδειγμα το σύστημα Eutelsat SaRa θα βασισθεί σε κωδικοποίηση Στρώματος III.

Εξελιγμένα ψηφιακά τηλεοπτικά συστήματα παρέχουν HDTV παράδοση στο κοινό με γήινη μετάδοση και μία ποικιλία από εναλλασσόμενα/αλληλοδιαδοχικά μέσα και προσφέρουν πλήρη κίνηση, υψηλή οπτική ανάλυση και υψηλής ποιότητας πολυκαναλικό περιβάλλοντα ήχο. Ο ολικός ρυθμός bit μπορεί να μεταδοθεί μέσα στο εύρος ζώνης ενός αναλογικού UHF καναλιού τηλεόρασης. Το σύστημα των ΗΠΑ Grand Alliance HDTV και το Ευρωπαϊκό σύστημα DVB κάνουν χρήση του MPEG-2 οπτικού συστήματος συμπίεσης και του MPEG-2 συστήματος μεταφοράς στρώματος, το οποίο χρησιμοποιεί ένα

εύκαμπτο ATM πρωτόκολλο πακέτων με επικεφαλίδες/περιγραφείς για πολυπλεξία ρευμάτων bits ήχου και εικόνας σεωένα ρεύμα με την απαραίτητη πληροφορία για να διατηρήσουν τα συστήματα συγχρονισμένα κατά την αποκωδικοποίηση (βλ Σχ. 17). Τα συστήματα διαφέρουν στον τρόπο με τον οποίο το ηχητικό σήμα συμπίεζεται. Το σύστημα Grand Alliance θα χρησιμοποιήσει τον αλγόριθμο Dolby AC-3, ενώ το Ευρωπαϊκό σύστημα θα χρησιμοποιήσει το MPEG-2/Ήχου.

## **8. ΣΥΜΠΕΡΑΣΜΑΤΑ**

Ο ψηφιακός ήχος χαμηλού ρυθμού bit εφαρμόζεται σε πολλούς διαφορετικούς χώρους, όπως τα καταναλωτικά ηλεκτρονικά, η επαγγελματική επεξεργασία ήχου, οι τηλεπικοινωνίες και η εκπομπή. Η κωδικοποίηση αντίληψης στο χώρο των συχνοτήτων έχει στρώσει το δρόμο για υψηλούς ρυθμούς συμπίεσης στην κωδικοποίηση ήχου. Η MPEG-1 κωδικοποίηση ήχου με τα τρία Στρώματα έχει γίνει ευρέως αποδεκτή ως διεθνές πρότυπο. Κωδικοποιητές λογισμικού, εφαρμογές μονού DSP τσιπ και επεκτάσεις υπολογιστών είναι διαθέσιμα από ένα αριθμό προμηθευτών.

Στην περιοχή της εκπομπής και των κινητών ραδιοφωνικών συστημάτων, οι υπηρεσίες μετακινούνται σε μεταφερόμενες συσκευές και νέα, τρίτης γενιάς κινητά τηλεπικοινωνιακά δίκτυα αναπτύσσονται. Οι κωδικοποιητές για αυτά τα δίκτυα δεν πρέπει μόνο να λειτουργούν σε χαμηλούς ρυθμούς bit αλλά πρέπει να είναι σταθεροί σε σφάλματα έκρηξης και σε περιβάλλοντα απώλειας πακέτων ή κυψελών. Οι τεχνικές απόκρυψης σφαλμάτων παίζουν ένα σημαντικό ρόλο. Εξαιτίας της έλλειψης διαθέσιμου εύρους ζώνης, οι παραδοσιακές τεχνικές κωδικοποίησης καναλιού μπορεί να μην είναι ικανές να βελτιώσουν επαρκώς την αξιοπιστία του καναλιού.

Οι MPEG κωδικοποιητές ήχου ελέγχονται από ψυχοακουστικά μοντέλα που μπορούν να βελτιωθούν, αφήνοντας έτσι χώρο για μία επαναστατική βελτίωση των συστημάτων κωδικοποιητών/αποκωδικοποιητών. Στο μέλλον, θα δούμε νέες λύσεις για κωδικοποίηση. Μία καλύτερη κατανόηση της διπλωματικής αντίληψης και της στερεοφωνικής αναπαράστασης θα οδηγήσει σε νέες προτάσεις.

Ο ψηφιακός πολυκαναλικός ήχος βελτιώνει τις στερεοφωνικές εικόνες και θα είναι μεγάλης σημαντικότητας και για εφαρμογές αποκλειστικά ήχου και για εφαρμογές πολυμέσων. Το MPEG-2/Ήχου προσφέρει και αναδρομικά και μη αναδρομικά συμβατά σχήματα κωδικοποίησης για να εξυπηρετήσει διαφορετικές ανάγκες. Η συνεχόμενη έρευνα θα καταλήξει σε επαυξημένες πολυκαναλικές αναπαραστάσεις κάνοντας καλύτερη χρήση των εσωκαναλικών συσχετίσεων και εσωκαναλικών αποτελεσμάτων απόκρυψης ώστε να κατεβάσει τους ρυθμούς bit ακόμα πιο κάτω. Μπορούμε επίσης να περιμένουμε λύσεις για ειδικές αναπαραστάσεις για ανθρώπους με ελαττώματα στην ακοή ή στην

όραση οι οποίες μπορούν να κάνουν χρήση της πολυκαναλικής διαμόρφωσης με πολλούς τρόπους.

Οι σημερινές δραστηριότητες της ομάδας ειδικών του ISO/MPEG στοχεύουν σε προτάσεις για κωδικοποίηση ήχου που θα προσφέρει υψηλότερους ρυθμούς συμπίεσης και οι οποίες θα συγχωνεύσουν όλη την περιοχή του ήχου από την κωδικοποίηση ήχου υψηλής πιστότητας και κωδικοποίηση ομιλίας μέχρι τη συνθετική ομιλία και το συνθετικό ήχο (ISO/IEC MPEG-4). Το MPEG-4 θα είναι το μελλοντικό πρότυπο πολυμέσων. Εξαιτίας του ότι η βασική ποιότητα ήχου θα είναι πιο σημαντική από τη συμβατότητα με τα υπάρχοντα πρότυπα, αυτή η δραστηριότητα έχει ανοίξει την πόρτα για εντελώς νέες λύσεις.



## 9. ΛΕΞΙΚΟ

Access Unit	Μονάδα προσπέλασης-Μία λογική υποδομή ενός Βασικού Ρεύματος για να διευκολύνει τυχαία προσπέλαση η μεταχείριση ρευμάτων bit	
Adaptive	Προσαρμοστικός	
Allocation	Καταμερισμός	
Animate	Ζωντανεύω, κινώ	
Archiving	Αρχειοθέτηση	
Artifacts	Ατέλειες	
Attenuate	Εξασθενώ	
Backchannel	Κανάλι επιστροφής	
Backwards Compatilby	Αναδρομική Συμβατότητα	
Binaural	Διπλωτικός	
Buffer	Προσωρινή ταχεία μνήμη	
Burst errors	Καταιγιστικά σφάλματα	
Cascade	Συζευγμένος-σε σειρά	
Cochlea	Κοχλίας	
Codec	Κωδικοποιητής/αποκωδικοποιητής	
Companding	Συμπίεση/αποσυμπίεση	
Compound	Σύνθετος	
Conductor	Μαέστρος	
Critical bands	Κρίσιμες ζώνες	
Dynamic allocation	Δυναμικός καταμερισμός	
Dynamic range	Δυναμική Ακτίνα	
Elementary Stream	Βασικό Ρεύμα. Μία ακολουθία δεδομένων που προέρχεται από ένα μονό παραγωγό στο εκπεμπόμενο MPEG-4 Τερματικό και τερματίζει σε ένα μονό παραλήπτη	
Enhanced	Εξελιγμένα	
Enhanced	Προηγμένα-εξελιγμένα	
Envelope	Περιβάλλουσα	
Fading	Εξασθένιση	
Forward	Πρόδρομος	
Frequency domain	Πεδίο συχνοτήτων	
High Fidelity	Υψηλή Πιστότητα	
High Fidelity	Υψηλή πιστότητα	
Interleave	Εμφολιάζω	
Impulse response	Κρουστική απόκριση	
Joint Stereo	Συνδυασμένο στέρεο	

Lossless	Χωρίς απώλειες	
Maskee	(απο)κρυμμένο σήμα	
Masker	Αποκρύβον σήμα	
Masking threshold	Κατώφλι απόκρυψης	
Midtread	Μέσου βήματος- μέσης στάθμης	
Mixing	Μίξη	
Modem	Διαμορφωτής/αποδιαμορφωτής	
Multichannel	Πολυκαναλικός	
Multimedia	Πολυμέσα	
Noise Shaping	Σχηματοποίηση-μορφοποίηση θορύβου	
Overhead	Επιβάρυνση	
Padding	Συμπληρωματικά	
Perceptual	Νοήμων, με αντίληψη	
Performance	Απόδοση	
Pitch	Τόνος φωνής	
Playback	Αναπαραγωγή	
Predictor	Προβλεπτής	
Profile	Κατατομή	
Quantization	Κβάντιση	
Random errors	Τυχαία σφάλματα	
Reverbs	Ακουστικές ανακλάσεις	
Robustness	Αντοχή	
Runlength-limited line code	Γραμμικός κώδικας ο οποίος περιορίζεται από το μήκος κατά την εφαρμογή	
Scalability	Ικανότητα για κλιμάκωση	
Session	Συνεδρίαση	
Simulcast	Ταυτόχρονη μετάδοση	
Spatialization	Κατανομή χώρου	
Standard	Πρότυπο	
Standarization	Καθορισμός προτύπων	
State of the art	Τελευταία λέξη της τεχνολογίας	
Streaming	Κατευθυνόμενο	
Subband	Υποζώνη	
Surround audio	Περιβάλλοντας ήχος	
Transparent	Διαυγής-καθαρός	
Upstream	Αντίθετα προς το ρεύμα	
Video conferencing	Τηλεσυνδιάσκεψη	
Wideband	Ευρεία Ζώνη	
Window switching	Εναλλαγή-μεταπήδηση παραθύρου	
Editing	Επεξεργασία	
Reusability	Ικανότητα επαναχρησιμοποίησης	

## **10. ΣΥΝΤΜΗΣΕΙΣ**

AAC	Advanced Audio Coding (MPEG-2 without backwards compatibility)
AAL	ATM Adaption Layer
ADR	Astra Digital Radio
API	Application Programming Interface
ATC	Adaptive Transform Coding
ATM	Asynchronous Transfer Mode
BBC	British Broadcasting Corporation
BIFS	Binary Format for Scene description
BSAC	Bit-Sliced Arithmetic Coding
CD	Compact Disk
CE	Core Experiment
CELP	Code Excited Linear Prediction
DAB	Digital Audio Broadcast
DAI	DMIF-Application Interface
DAT	Digital Audio Tape
DCC	Digital Compact Cassette
DCT	Discrete Cosine Transform
DMIF	Delivery Multimedia Integration Framework
DNI	DMIF Network Interface
DS	DMIF Signalling
DVB	Digital video Broadcast
DVD	Digital Versatile (or Video) Disk
FAP	Facial Animation Parameters
FFT	Fast Fourier Transform
FlexMux	A Flexible (Content) Multiplex tool
HDTV	High Definition TV
Hi-Fi	High Fidelity
HTTP	Hyper Text Transfer Protocol
ISDN	Integrated Services Digital Networks
IP	Internet Protocol
IPI	Intellectual Property Identification
IPR	Intellectual Property Rights
ISO	International Organization of Standards
JND	Just Noticeable Distortion
LC	Low Complexity
LFE	Low Frequency Enhancement
LPC	Linear Predictive Coding

LTP	Long Term Prediction
MDCT	Modified Discrete Cosine Transform
MIDI	Musical Instrument Digital Interface
MPEG	Moving Pictures Expert Group
NMR	Noise to Mask Ratio
OD	Object Descriptor
PAD	Program-Associated-Data
PASC	Precision Audio Subband Coding
PCM	Pulse Code Modulation
QoS	Quality of Service
RTP	Real Time Transport Protocol
RTSP	Real Time Streaming Protocol
SBC	Subband Coding
SL	Sync(hronization) Layer
SMR	Signal-Mask-Ratio
SPL	Sound Pressure Level
TC	Transform coding
TCP	Transmission Control Protocol
TTS	Text-to-Speech
UDP	User Datagram Protocol
VCR	Video Cassette Recorders
VRML	Virtual Reality Modeling Language
WWW	World Wide Web

## **11. ΑΝΑΦΟΡΕΣ**

1. A.A.M.L. Bruckers et al., "Lossless Coding for DVD Audio", 101<sup>th</sup> Audio Engineering society Convention, Los Angeles, 1996, preprint 4358
2. N.S. Jayant and P. Noll, "Digital Coding of Waveforms: Principles and Applications to Speech and Video" Prentice Hall, 1984
3. A.S. Spanias, "Speech Coding : A Tutorial Review" Proc of the IEEE, Vol. 82, No. 10, pp 1541-1582, Oct 1994
4. N.S. Jayant, J. D. Johnston and Y. Shoham, "Coding of Wideband Speech" Speech Communication 11 (1992), pp. 127-138
5. A. Gersho, "Advances in Speech and Audio Compression" Proc. of the IEEE, Vol. 82, No. 6, pp 900-918, 1994
6. P. Noll, "Wideband Speech and Audio coding" IEEE Commun. Mag., Vol 31, No.11, pp. 34-44. 1993
7. P. Noll, "Digital Audio Coding for Visual Communications" Proc. of the IEEE, Vol. 83, No. 6, June 1995
8. ISO/IEC JTCL/SC29, "Information Technology-Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbits-IS 11172 (Part 3,Audio)," 1992
9. ISO/IEC JTCL/SC29, "Information Technology-Generic Coding of Moving Pictures and Associated Audio Information – IS 13818 (Part 3,Audio)" 1994
10. ISO/MPEG, Doc. N0821, Proposal Package Description-Revision 1.0, Nov. 1994
11. G.T. Hathaway, "A NICAM Digital Stereophonic Encoder" in "Audiovisual Telecommunications (Editor N.D. Nigthingale), Chapman & Hall, 1992, pp. 71-84
12. E. Zwicker and R. Feldtkeller, Das ais Nachrichtenempfänger, Stuttgart: S. Hirzel Verlag, 1967
13. N.S. Jayant, J.D. Johnston and R. Safranek, "Signal compression based on models of human perception" Proc of the IEEE, Vol. 81, No 10, pp 1385-1422, 1993
14. R. Zelinski and P. Noll, "Adaptive Transform Coding of Speech Signals" IEEE Trans. On Acoustics, Speech and Signal Proc, Vol ASSP-25, pp 299-309, August 1977
15. A. Hoogendorn, "Digital compact cassette" Proc. of the IEEE, Vol. 82, No 10, pp. 1479-1489, Oct. 1994
16. P. Noll, "On predictive quantizing schemes" Bell System Technical Journal, vol. 57, pp 1499-1532, 1978
17. J. Makhoul and M. Berouti, "Adaptive noise spectral shaping and entropy coding in predictive coding of speech" IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 27, No 1, pp. 63-73, Feb 1979

- 18.D. Esteban and C. Galand, "Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes", Proc. ICASSP, pp. 191-195,1987
- 19.J.H. Rothweiler, "Polyphase Quadrature Filters, a New Subband Coding Technique", Proc. International Conference ICASSP 83, pp. 1280-1283,1983
- 20.J. Princen and A. Bradlay, "Analysis/Synthesis Filterbank Design Based on Time Domain Aliasing Cancellation", IEEE Trans. On Acoustics,Speech and Signal Process. Vol. ASSP-34, pp. 1153-1161, 1986
- 21.H.S. Malvar, "Signal Processing with Lapped Transforms", Artech House Inc, 1992
- 22.F.S. Yeoh, C.S. Xydeas, "Split-band coding of speech signals using a transform technique" Proc. ICC, 1984, Vol. 3, pp. 1183-1187
- 23.W. Granzow, P. Noll, C. Volmary, "Frequency-domain coding of speech signals" (in German), NTG-Fachbericht No. 94, VDE-Verlag, Berlin, 1986, pp. 150-155
- 24.B. Edler, "Coding of Audio Signals with Overlapping Block Transform and Adaptive Window Functions" (in German), Frequenz, vol 43, pp. 252-256,1989
- 25.M. Iwadare, A. Sugiyama, F. Hazu, A. Hirano, and T. Nishitani, "A 128 kb/s Hi-Fi Audio CODEC Based on Adaptive Transform Coding with Adaptive Block Size" IEEE J. on Sel. Areas in Common, vol 10, No 1, pp. 138-144, Jan 1992
- 26.R. Zelinski and P. Noll, "Adaptive Blockquantisierung von Sprachsignalen" Technical Report No. 181, Heinrich-Hertz-Institut fur Nachrichtentechnik, Berlin,1975
- 27.R.G. van der Waal, K. Brandenburg, and G. Stoll, "Current and future standarization of high-quality digital audio coding in MPEG" Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, N.Y. 1993
- 28.P. Noll and D. Pan "ISO/MPEG Audio Coding", International Journal of High Speed Electronics and Systems, Vol. 8, No 1,pp. 69-118, 1997
- 29.K. Brandenburg and G. Stoll, " the ISO/MPEG-Audio Codec:A Generic Standard for Coding of High Quality digital Audio", Journal of the Audio Engineering Society (AES), vol. 42, No. 10,pp. 780-792, Oct. 1994
- 30.L.M. van de Kerkhof and A.G. Cugini, " The ISO/MPEG audio coding standard" Widescreen Review,1994
- 31.Y.F. Dehery, G. Stoll, L. van de Kerkhof, "MUSICAM Source Coding for Digital Sound", 17<sup>th</sup> International Television Symposium, Montreux, Record pp. 612-617, June 1991
- 32.K. Brandenburg, J. Herre, J.D. Johnston, Y. Mahieux, E.E. Schroeder: "ASPEC:Adaptive Spectral Perceptual Entropy Coding of High Quality Music Signals", 90<sup>th</sup> Audio Engineering Societ Convention, Paris ,preprint 3011,1991

- 33.H.G. Musmann, "The ISO Audio Coding Standard" Proc. IEEE Globecom, Dec. 1990
- 34.R.G. van der Waal,A.W.J. Oomen and F.A. Griffiths, "Performance comparison of CD, noise shaped CD and DCC" 96<sup>th</sup> Audio Engineering Society Convention, Amsterdam ,preprint 3845,1994
- 35.J.Herre, K. Brandenburg, and D. Lederer, "Intensity stereo coding" 96<sup>th</sup> Audio Engineering Society Convention, Amsterdam ,preprint 3799,1994
- 36.D. Pan, "A tutorial on MPEG/Audio Compression" IEEE Trans. On Multimedia, vol. 2, No. 2,1995, pp. 60-74
- 37.K. Brandenburg et al., "Variable data-rate recording on a PC using MPEG-Audio Layer-III", 95<sup>th</sup> Audio Engineering Society Convention, New York,1993
- 38.P.A. Sarginson, "MPEG-2: Overview of the system Layer" BBC Research and Development report ,BBC RD 1996/2, 1996
- 39.T. Ryden, C. Grewin, S. Bergman, " The SR report on the MPEG Audio subjective listening tests in Stockholm April/May 1991", ISO/IEC JTC1/SC29/WG 11:Doc.-No. MPEG 91/010, May 1991
- 40.H. Fuchs, "Report on the MPEG/Audio Subjective listening tests in Hannover", ISO/IEC JTC1/SC29/WG 11:Doc.-No. MPEG 91/331, November 1991
- 41.G. Stoll et al., "Extension of ISO/MPEG-Audio layer II to multi-channel coding: The future standard for broadcasting,telecommunication and multimedia application", 94<sup>th</sup> Audio Engineering Society Convention,Berlin,preprint no. 3550,1993
- 42.B. Grill et al., "Improved MPEG-2 Audio multi-channel encoding", 96<sup>th</sup> Audio Engineering Society Convention, Amsterdam ,preprint 3865,1994
- 43.M. Bosi et al., "ISO/IEC MPEG-2 advanced audio coding" 101<sup>th</sup> Audio Engineering Society Convention,preprint 4382, Los Angeles, 1996
- 44.J.D. Johnston et al., "NBC-Audio-Stereo and multichannel coding methods", 101<sup>th</sup> Audio Engineering Society Convention,preprint 4383, Los Angeles, 1996
- 45.W.R.Th. Ten Kate,et al., "Matrixing of bit rate reduced audio signals" Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 92), vol. 2, pp. II-205-II-208,1992
- 46.W.R.Th. Ten Kate, "Compatibility Matrixing of multi-channel bit-rate-reduced audio signals", 96<sup>th</sup> Audio Engineering Society Convention, Amsterdam ,preprint 3792,1994
- 47.F. Feige and D. Kirby, "Report on the MPEG/Audio Multichannel Formal Subjective Listening Test", ISO/IEC JTC1/SC29/WG 11: Doc. N 0685, March 1994
- 48.D. Meares and D. Kirby, "Brief Subjective Listening Tests on the MPEG-2 Backwards Compatible Multichannel Audio Coders", ISO/IEC JTC1/SC29/WG 11 : August 1994

- 49.ISO/IEC/JTC1/SC29, “Report on the formal subjective listening tests of MPEG-2 NBC multichannel audio coding”, Document N 1371, Oct. 1996
- 50.ITU-R Document TG 10-2/3, Oct. 1991
- 51.IEC/JTC1/SC29, “Description of MPEG-4”, document N 1410, Oct. 1996
- 52.D.S. Burpee and P.W. Shumate, “Emerging residential broadband telecommunications”, Proc. IEEE, Vol. 82, No 4,pp. 604-614, 1994
- 53.N.S. Jayant, “high quality networking of Audio-visual information”, IEEE Commun. Mag.,pp. 84-95,1993
- 54.C. Todd et al., “AC-3: Flexible perceptual coding for audio transmission and storage”, 96<sup>th</sup> Audio Engineering Society Convention, Amsterdam ,preprint 3796,1994
- 55.R. Hopkins, “Choosing an American digital HDTV terrestrial broadcasting system”, Proc. of the IEEE, vol. 82, No. 4, pp. 554-563, 1994
- 56.“The Grand Alliance”, IEEE Spectrum, pp. 36-45, April 1995
- 57.ETSI, European Telecommunication standard, Draft prETS 300 401, Jan 1994
- 58.Ch. Weck, “The error protection of DAB” Audio engineering Society Conference “DAB-The future of radio” London, May 1995
- 59.R.D. Jurgen “Broadcasting with Digital Audio”, IEEE Spectrum,pp. 52-59, March 1996

## **12. ΒΙΒΛΙΟΓΡΑΦΙΑ**

1. Peter Noll “MPEG Digital Audio Coding” IEEE SIGNAL PROCESSING MAGAZINE Sep 1997
2. Rob Koenen “MPEG-4 Overview-(Atlantic City Version)” ISO/IEC JTC1/SC29/WG11 N2459  
(<http://drogo.cselt.stet.it/mpeg/standards/mpeg-4/mpeg-4.htm>)