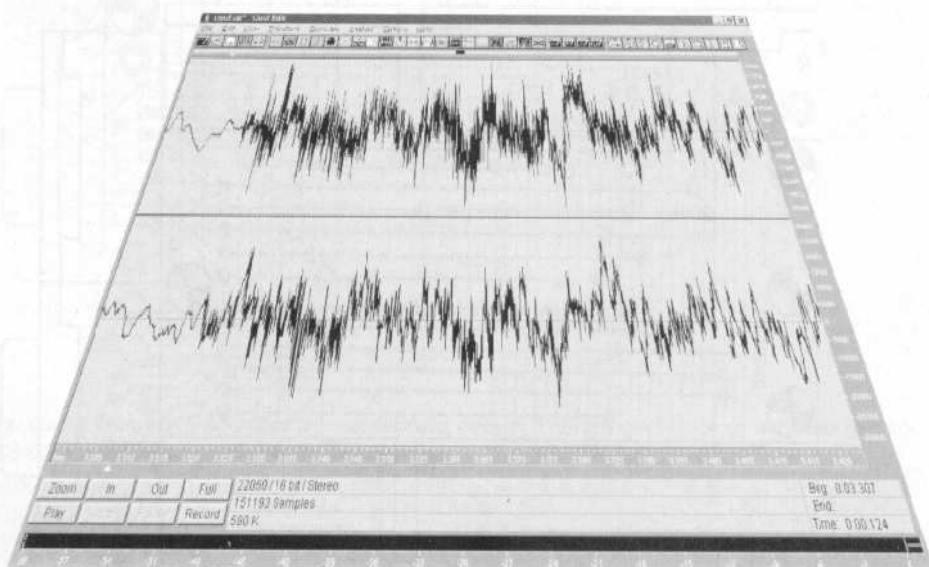


Τα τελευταία χρόνια η μορφή αλλά και η χρήση του υπολογιστή έχει αλλάξει από ένα απλό επεξεργαστή κειμένου και γραφικών σε μια μονάδα που ανάμεσα σε όλα τα άλλα είναι κατάλληλη για εγγραφή και αναπαραγωγή ηχητικών σημάτων και βίντεο. Ιδιαίτερα όσον αφορά την αναπαραγωγή ήχου, τα τελευταία χρόνια έχουν εμφανιστεί αρκετά διαφορετικά πρότυπα μορφοποίησης. Στο άρθρο περιγράφουμε με συντομία τα σπουδαιότερα από αυτά, ενώ δίνουμε ιδιαίτερη βαρύτητα στο δημοφιλέστερο MP3.

Ψηφιακές μορφές ήχου

Περισσότεροι ήχοι από τον υπολογιστή σας



Εισαγωγή

Σήμερα σπάνια θα συναντήσουμε υπολογιστή χωρίς ήχο. Ήχοι υπάρχουν παντού, από τα πολύπλοκα παιχνίδια μέχρι τους ήχους ειδοποίησης που συνοδεύουν τις διάφορες εντολές. Ο ήχος έχει καταστεί σημαντική παράμετρος και στους χρήστες του δικτύου Internet. Σήμερα μπορείτε να ακούσετε ραδιοφωνικούς σταθμούς ή να "φορτώσετε" μουσικά κομμάτια από το δίκτυο. Έτσι, έχουν εμφανιστεί αρκετοί τρόποι μορφοποίησης για την αποθήκευση του ήχου εξαιτίας των πολλών εφαρμογών και της μεγάλης ποικιλίας υπολογιστών που κυκλοφορούν στην αγορά.

Η ψηφιοποίηση

Μετά την επινόηση του CD πριν 20 περί-

που χρόνια, σχεδόν όλοι γνωρίζουν σήμερα τι σημαίνει ο όρος ψηφιακός ήχος. Εν συντομία, το αναλογικό ηχητικό σήμα μεταφράζεται σε ένα μεγάλο αριθμό δυαδικών λέξεων από ένα μετατροπέα αναλογικού σε ψηφιακό (ADC). Η διαδικασία αυτή ονομάζεται κβάντιση ή ψηφιοποίηση. Ο ρυθμός με τον οποίο γίνεται η μετατροπή στην περίπτωση των CD είναι 44100 φορές το δευτερόλεπτο για το αριστερό και άλλο τόσο για το δεξί κανάλι.

Αφού η ευκρίνεια του CD είναι 16 bit, η κβάντιση ενός αναλογικού σήματος έχει σαν αποτέλεσμα μια ροή δεδομένων περίπου 1.4 Mbit/s. Η ροή δεδομένων μαζί με πρόσθετα ψηφία για διόρθωση σφαλμάτων και άλλες πληροφορίες εγγράφονται στο δίσκο CD. Κατά τη διάρκεια της αναπαραγωγής η διαδικασία

λαμβάνει χώρα κατά την αντίθετη φορά. Η ανάγνωση των δεδομένων του CD γίνεται από laser και η ψηφιακή πληροφορία μετατρέπεται σε αναλογικό σήμα από ένα μετατροπέα ψηφιακού σε αναλογικό (DAC).

Τα ηχητικά σήματα σε ένα υπολογιστή επεξεργάζονται με τον ίδιο τρόπο. Τα απαραίτητα εξαρτήματα όπως ο DAC και ο ADC περιέχονται στην κάρτα ήχου, ενώ ο σκληρός δίσκος είναι το μέσο αποθήκευσης. Στον υπολογιστή η ροή δεδομένων αποθηκεύεται σε ένα αρχείο.

Η αποθήκευση των δυαδικών λέξεων χρειάζεται τεράστιο χώρο στο σκληρό δίσκο: για ποιότητα CD, περίπου 1 Mbyte για κάθε έξι δευτερόλεπτα ήχου. Υπάρχουν διάφοροι τρόποι περιορισμού του όγκου των δεδομένων. Ένας από αυτούς είναι η μείωση της συχνότητας δειγματοληψίας από 44.1 KHz σε 10 KHz με συνέπεια όμως τη μείωση του εύρους ζώνης. Άλλος είναι η εγγραφή του σήματος σε μονοφωνική αντί για στερεοφωνική μορφή, που μειώνει την ποιότητα. Ένας τρίτος είναι η μείωση της ευκρίνειας από 16 σε 8 bit που χειροτερεύει το λόγο σήματος προς θόρυβο και αυξάνει την παραμόρφωση. Τέταρτη μέθοδος είναι η συμπίεση του σήματος κατά τη διάρκεια της εγγραφής και η αποσυμπίεσή του κατά τη διάρκεια της αναπαραγωγής.

Σχεδόν κάθε υπολογιστικό σύστημα (Atari, Unix, Intel) διαθέτει τη δικιά του μορφή αρχείων ήχου με αποτέλεσμα τη δημιουργία σύγχυσης ανάλογης με αυτή που επικρατεί στον τομέα της εικόνας. Η κατάσταση έγινε χειρότερη μετά τη εμφάνιση κι άλλων τύπων στο Internet.

Στο πίνακα 1, φαίνονται μερικά από τα πιο δημοφιλή συλ μορφοποίησης μαζί με μια σύντομη περιγραφή.

Συμπίεση ή όχι

Η συμπίεση ενός ηχητικού αρχείου είναι

μια σύνθετη διαδικασία διότι είναι δύσκολο να καθοριστούν κοινά χαρακτηριστικά μεταξύ ενός μεγάλου πλήθους δειγμάτων. Συνεπώς υπάρχουν λίγες μέθοδοι συμπίεσης χωρίς απώλειες για ηχητικά σήματα. Μια από αυτές είναι το πρότυπο ADPCM (στα Windows) για τη δημοφιλή μορφοποίηση WAV.

Ωστόσο είναι δυνατή η συμπίεση των ηχητικών σημάτων, αν αποδεχτούμε μια μικρή απώλεια πληροφορίας όπως και στην επεξεργασία ψηφιακών εικόνων. Στην περίπτωση των ψηφιακών εικόνων, η μέθοδος συμπίεσης θεσπίστηκε από την Joint Photographic Expert Group (JPEG), μια ομάδα της Joint Technical Committee 1 (JCT1) της ISO (International Standard Organization – σύμβουλος των Ηνωμένων Εθνών) και της IEC (International Electrotechnical Commission). Μια άλλη ομάδα της JCT1, η Motion Pictures Expert Group ή MPEG, θέσπισε ένα άλλο πρότυπο για τη συμπίεση ηχητικών δεδομένων: τη μορφοποίηση MPEG. Σήμερα η πιο δημοφιλής μορφοποίηση για τους χρήστες του Internet είναι το MPEG-1 Layer 3 (MP3).

Ήχος σε πραγματικό χρόνο

Στις πρώτες ημέρες της ηχητικής πανδασίας στους υπολογιστές, ένα ηχητικό αρχείο έπρεπε πρώτα να "φορτωθεί" στη μνήμη του υπολογιστή πριν καταστεί δυνατή η αναπαραγωγή του. Αυτό είναι ιδιαίτερα δυσάρεστο κατά τη διάρκεια του "φορτώματος" αρχείων από το Internet.

Με την εμφάνιση των γρήγορων modem έγινε δυνατή η ανάπτυξη τεχνικών που επιτρέπουν την αναπαραγωγή κατά τη διάρκεια αυτής της διαδικασίας. Σε αυτό συνέβαλε η τεχνική με την ονομασία 'AOD' (Audio On Demand), 'real time audio' ή 'streaming audio', ένα προϊόν της εταιρείας RealAudio.

MP3: καλή ποιότητα με υψηλή συμπίεση

Σήμερα το MP3 είναι η πιο δημοφιλής μορφοποίηση ήχου. Σε πολύ σύντομο χρόνο, αυτό το πρωτόκολλο ηχητικής συμπίεσης απέκτησε πολλούς θαυμαστές. Έτσι οι περισσότεροι σύγχρονοι υπολογιστές έχουν προεγκαταστημένο το λογισμικό κωδικοποίησης και αποκωδικοποίησης MP3. Ακόμα, πολλά αρχεία MP3 υπάρχουν στο Internet ενώ κυκλοφορούν ήδη CD με αρχεία MP3. Μερικοί κατασκευαστές ξεκίνησαν τη διάθεση στην αγορά συσκευών walkman MP3 που είναι συμπαγείς φορητές συσκευές αναπαραγωγής CD με ενσωματωμένη μνήμη flash.

Αναπτύχθηκε για το DAB

Σαν συνεργάτης στο πρόγραμμα Eureka 147 (για την ανάπτυξη του συστήματος επίγειων ψηφιακών σταθμών εκπομπής – DAB), το γερμανικό ινστιτούτο Fraunhofer για ολοκλη-

Πίνακας 1. Περίληψη των πιο σημαντικών μορφοποιήσεων ήχου.

AIF:	μια μορφοποίηση που αρχικά αναπτύχθηκε για Macintosh. Είναι δημοφιλής στο Internet και προσφέρει πολλούς διαφορετικούς ρυθμούς δειγματοληψίας και βαθμούς ευκρίνειας. Οι browsers της Netscape και της Microsoft μπορούν άνετα να συνεργαστούν με αυτήν.
AU:	δημιουργήθηκε από τις NeXT και SUN και ήταν δημοφιλής στο διαδίκτυο. Χρησιμοποιείται ακόμα και σήμερα. Υπάρχει ένας αριθμός μεταβλητών, αλλά κανονικά τα δεδομένα συμπιέζονται με μορφή 8 bit σύμφωνα με το πρότυπο u-law. Οι περισσότεροι browsers μπορούν να λειτούργησουν μαζί της.
ES:	μια μορφοποίηση ροής από την EchoCast. Διατίθεται λογισμικό για Windows και Macintosh.
IFF:	μια μορφοποίηση ήχου της Amiga που μπορεί να επεξεργαστεί μόνο μονοφωνικό ήχο 8 bit. Επιτρέπει την ελεύθερη επιλογή ρυθμού δειγματοληψίας.
LCC:	μορφοποίηση υψηλής συμπίεσης (λόγοι μέχρι 1:50) που διατίθεται μόνο για Windows, ενώ τώρα αναπτύσσονται και άλλες εκδόσεις.
MID:	το MIDI δεν είναι πραγματική μορφοποίηση ήχου, αλλά ένα πρότυπο για αντάλλαξη δεδομένων ελέγχου μεταξύ ηλεκτρονικών οργάνων.
MOD:	αυθεντική μορφοποίηση ήχου της Amiga παρόμοια με το MIDI. Ένα αρχείο MOD περιέχει μια συστοιχία από δείγματα και οδηγίες για τον τρόπο αναπαραγωγής τους.
MP3:	η πιο δημοφιλής μορφοποίηση ήχου σήμερα.
RA, RAM, RPM:	μια δημοφιλής μορφοποίηση ροής για το Internet από τη Real Audio.
SND:	η Apple, η Amiga και η Tandy χρησιμοποιούν αυτή την προέκταση για τα αρχεία ήχου. Μερικές μεταβλητές είναι συμβατές με τη μορφοποίηση AU.
STR:	μορφοποίηση για επαγγελματική επεξεργασία ήχου στους Macintosh.
VDO:	μια άλλη μορφοποίηση ροής.
VMD:	μορφοποίηση ροής με την ονομασία Internet Wave. Διατίθεται με δωρεάν κωδικοποιητή – αποκωδικοποιητή για τα Windows.
VOC:	μορφοποίηση φωνής, που αναπτύχθηκε από την Creative Labs κατασκευάστρια των γνωστών καρτών ήχου Soundblaster.
WAV:	η περισσότερο χρησιμοποιούμενη μορφοποίηση ήχου που έγινε γνωστή μετά την εμφάνιση των Windows. Προσφέρει πολλούς ρυθμούς δειγματοληψίας, βαθμούς ευκρίνειας και συντελεστές συμπίεσης.
XDM:	μια μορφοποίηση MPEG από την StreamWorks για ροή ήχου. Διατίθεται ειδικό λογισμικό για τα Windows.

ρωμένα κυκλώματα ανέπτυξε ένα κωδικοποιητή - αποκωδικοποιητή (codec, coder/decoder) για το DAB.

Ο αντίστοιχος αλγόριθμος, λαμβάνοντας υπόψη κάποια χαρακτηριστικά της ανθρώπινης ακοής, καθορίζει ποια τμήματα σε ένα κομμάτι μουσικής μπορούν να γίνουν ή όχι αντιληπτά από τον ακροατή. Με τον τρόπο αυτό, καθορίζει αν η αντίστοιχη πληροφορία πρέπει να συμπεριληφθεί ή όχι στη ροή των δεδομένων. Έτσι επιτυγχάνεται συμπίεση του πλεονασμού, πράγμα που επιτρέπει μια σημαντική μείωση του ποσού των δεδομένων χωρίς υποβάθμιση της ποιότητας του ήχου.

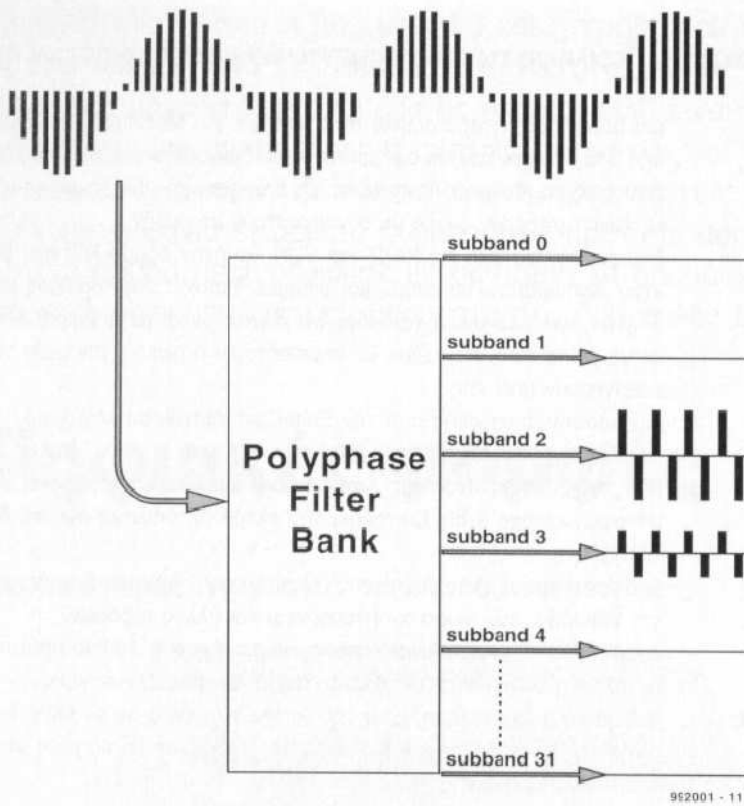
Ο αλγόριθμος βελτιώθηκε από την IIS σε συνεργασία με το πανεπιστήμιο του Erlangen και ονομάστηκε ISO MPEG-1 Layer-3 (IS 11172-3 και IS13818-3).

Layer 1, 2 ή 3

Χωρίς συμπίεση δεδομένων, τα ηχητικά σήματα αναπαριστώνται με δείγματα εύρους 8 ή 16 bit που λαμβάνονται με ρυθμό τουλάχιστον διπλάσιο της υψηλότερης συχνότητας που αυτά εμπεριέχουν.

Όπως αναφέραμε ήδη, η ψηφιοποίηση ενός ηχητικού σήματος έχει σαν αποτέλεσμα ρυθμό δεδομένων περίπου 1.4 Mbit/s. Ένα σύστημα συμπίεσης μπορεί να τον συμπίεσει με λόγο 1:12 χωρίς υποβάθμιση του ήχου. Είναι επίσης δυνατοί λόγοι συμπίεσης μέχρι 1:24, ενώ η ποιότητα του ήχου είναι ανώτερη από αυτή που επιτυγχάνεται με τη μείωση του ρυθμού δειγματοληψίας ή της ευκρίνειας.

Το πρότυπο MPEG-1 περιλαμβάνει τρία στρώματα (Layer) συμπίεσης: Layer 1, Layer 2 και Layer 3. Και τα τρία μπορούν να αναπαραγάγουν ήχο με ποιότητα CD. Στο πίνακα 2 φαίνονται μερικά χαρακτηριστικά των layer.



Σχήμα 1. Κάθε layer MPEG-1 χρησιμοποιεί μια συστοιχία φίλτρων 32 bit. Το αν το σήμα επικαλύπτεται ή όχι καθορίζεται μετά την κβάντιση.

Οι ορισμοί του προτύπου αναφέρονται μόνο στον κωδικοποιητή και τη μορφοποίηση δεδομένων που χρησιμοποιείται. Η πληροφορία αυτή επιτρέπει στους κατασκευαστές να σχεδιάσουν τους ανάλογους αποκωδικοποιητές.

Αν δεν χρειάζεται στερεοφωνική αναπαραγωγή και επιπλέον είναι αποδεκτό ένα περιορισμένο εύρος ζώνης, μπορούν να χρησιμοποιηθούν υψηλότεροι λόγοι συμπίεσης. Ο υψηλότερος, στο Layer 3, χρησιμοποιεί το μικρότερο δυνατό ρυθμό δεδομένων και παράγει την καλύτερη ποιότητα ήχου.

Οι τρεις κώδικες είναι ιεραρχικά συμβατοί, που σημαίνει ότι ο αποκωδικοποιητής του Layer 3 μπορεί να χρησιμοποιηθεί και για τα Layer 1 και 2. Ωστόσο δεν ισχύει το αντίστροφο.

Όσο μεγαλώνει ο αριθμός του Layer, τόσο πιο σύνθετος γίνεται ο αποκωδικοποιητής και μεγαλώνει ο λόγος συμπίεσης που μπορεί να επιτευχθεί.

Στον πίνακα 2 φαίνονται οι πιθανοί λόγοι συμπίεσης που μπορούν να χρησιμοποιηθούν στο Layer 3.

Οι έλεγχοι ακρόασης έδειξαν ότι η απόδοση του Layer 3 παραμένει εξαιρετική με λόγο συμπίεσης 1:12 – που αντιστοιχεί σε ροή δεδομένων 64 Kbit/s για κάθε ηχητικό κανάλι. Σε ορισμένες εφαρμογές το εύρος ζώνης μπορεί να μειωθεί σε 10 KHz, οπότε είναι δυνατή καλή στερεοφωνική αναπαραγωγή με λόγο συμπίεσης 1:24.

Ενιαία δομή

Και τα τρία Layer έχουν την ίδια δομή. Η τεχνική κωδικοποίησης είναι γνωστή σαν perceptual noise shaping. Ο κωδικοποιητής αναλύει το φάσμα του ηχητικού σήματος με τη βοήθεια μιας συστοιχίας φίλτρων (σχήμα 1) και χρησιμοποιεί ένα ψυχοακουστικό μοντέλο για να καθορίσει ευδιάκριτες στάθμες θορύβου. Έτσι η πληροφορία κβαντίζεται και κωδικοποιείται λαμβάνοντας υπόψη δύο σημαντικές συνθήκες: τη μέγιστη ροή δεδομένων και το φαινόμενο επικάλυψης (masking).

Και τα τρία Layer χρησιμοποιούν την ίδια συστοιχία φίλτρων με 32 ζώνες. Επιτρέπουν ρυθμούς δειγματοληψίας 32, 44.1 και 48 KHz με ροή δεδομένων 32 Kbits/s ή υψηλότερη.

Υπόβαθρο

Το MPEG-1 Layer 3 για να πετύχει σημαντική ελάττωση του ψηφιακού εύρους ζώνης χρησιμοποιεί διάφορες τεχνικές και συντομεύσεις. Οι πιο σημαντικές είναι οι εξής:

- Ελάχιστο κατώφλι ακρόασης
- Φαινόμενο επικάλυψης
- Αποθήκευση bytes
- συνδυασμένο stereo
- Κωδικοποίηση Huffman

Ελάχιστο κατώφλι ακρόασης

Η έρευνα έδειξε ότι το ελάχιστο κατώφλι της ανθρώπινης ακοής δεν είναι γραμμικό. Παρουσιάζει μια κορυφή μεταξύ 2 και 5 KHz. Τα χαρακτηριστικά του περιγράφονται από

τους Fletcher και Munson. Δεν είναι απαραίτητο να κωδικοποιείται ήχος με στάθμη χαμηλότερη από αυτό το κατώφλι, αφού ο ακροατής δεν μπορεί να τον ακούσει.

Φαινόμενο επικάλυψης (masking)

Λαμβάνεται υπόψη ότι η ανθρώπινη ακοή δεν μπορεί να αντιληφθεί ασθενείς ήχους που επικαλύπτονται ολικά ή μερικά από άλλους (πολύ) ισχυρότερους.

Η έρευνα έδειξε ότι λόγω της επικάλυψης μερικοί ήχοι δεν χρειάζεται να κωδικοποιηθούν, γεγονός που εξοικονομεί μεγάλο όγκο πληροφορίας. Για το λόγο αυτό όλοι οι κωδικοποιητές MPEG-1 Layer 3 περιέχουν ένα ψυχοακουστικό μοντέλο που ενσωματώνει τα χαρακτηριστικά της ανθρώπινης ακοής.

Αποθήκευση byte

Συχνά εμφανίζεται το φαινόμενο ένα μουσικό πέρασμα να μη μπορεί να κωδικοποιηθεί με το διαθέσιμο ρυθμό δεδομένων. Συνεπώς η ποιότητα του ήχου πρέπει να προσαρμοστεί προσωρινά ώστε να επιτρέπει στη ροή δεδομένων να προσαρμόζεται στη χωρητικότητα του ψηφιακού καναλιού. Το MPEG-1 Layer 3 χρησιμοποιεί έναν απομονωτή που προσφέρει πρόσθετη χωρητικότητα σε τέτοιες περιπτώσεις. Ο απομονωτής αδειάζει όταν ο ήχος κωδικοποιείται με ρυθμό δεδομένων μικρότερο από το διαθέσιμο του καναλιού.

Συνδυασμένο stereo

Πολλά μικρά στερεοφωνικά συγκροτήματα χρησιμοποιούν κοινό (και για τα δύο κανάλια) μεγάλων χαμηλών συχνοτήτων. Ο ακροατής νομίζει ότι ο ήχος προέρχεται κυρίως από τα περιφερειακά μεγάφωνα παρά από αυτό. Η έρευνα έδειξε ότι κάτω από μια ορισμένη συχνότητα το ανθρώπινο αυτί δεν μπορεί να καταλάβει τη διεύθυνση από την οποία έρχεται ο ήχος. Έτσι οι τεχνικές συμπίεσης, εκμεταλλευόμενες αυτό το χαρακτηριστικό, δεν περιλαμβάνουν στερεοφωνική πληροφορία κάτω από αυτή τη συχνότητα, δηλαδή το σήμα κωδικοποιείται σε μονοφωνική μορφή.

Κώδικας Huffman

Η κωδικοποίηση στο MPEG-1 Layer 3 ισοθετεί μια κλασική τεχνική: τον κώδικα Huffman. Χρησιμοποιείται αφού πραγματοποιηθεί η συμπίεση δεδομένων για την κωδικοποίηση της ψηφιακής πληροφορίας. Επομένως δεν αποτελεί σύστημα συμπίεσης, αλλά μια πολύ ικανή τεχνική κωδικοποίησης. Ο αλγόριθμος Huffman δημιουργεί κώδικα μεταβλητού μήκους με ακέραιο αριθμό bits. Στα πιο σημαντικά σήματα διατίθεται σύντομος κώδικας, ενώ στα λιγότερο σημαντικά μεγαλύτερος.

Αφού οι κώδικες Huffman διαθέτουν μια συγκεκριμένη επικεφαλίδα, μπορούν να απο-

κωδικοποιηθούν τέλεια παρά το μεταβλητό τους μήκος. Η αποκωδικοποίηση είναι πολύ γρήγορη με τη βοήθεια πίνακα. Η τεχνική εξοικονομεί περίπου το 20% του χώρου αποθήκευσης.

Η τεχνική Huffman αποτελεί ιδανικό συμπλήρωμα για την συμπίεση κατά MP3. Στα περάσματα που περιέχουν πολλές συχνότητες η εκμετάλλευση του φαινομένου της επικάλυψης παρέχει μια σημαντική μείωση όγκου δεδομένων εξαλείφοντας τα επικαλυπτόμενα σήματα. Στην περίπτωση αυτή δεν υπάρχουν πολλά ταυτόσημα σήματα, οπότε η κωδικοποίηση Huffman έχει χαμηλές επιδόσεις.

Κατά τη διάρκεια των περασμάτων με λίγους διαφορετικούς ήχους δεν υπάρχουν πολλά φαινόμενα επικάλυψης. Τότε ο κώδικας Huffman εξοικονομεί σημαντικό χώρο, αφού υπάρχει μεγάλη ποσότητα πλεονάζουσας πληροφορίας. Τέτοια περάσματα μπορούν επομένως να κωδικοποιηθούν με σύντομες κωδικές λέξεις.

Πίνακας 2. Οι τρόποι λειτουργίας του MPEG-1

Συντελεστής συμπίεσης	μέθοδος
1:4 μέσω	layer 1 το στερεοφωνικό σήμα προκαλεί ροή δεδομένων 384 Kbits/s
1:6 – 1:8 μέσω	layer 2 το στερεοφωνικό σήμα προκαλεί ροή δεδομένων 256-192 Kbits/s
1:10 – 1:12	μέσω layer 3 το στερεοφωνικό σήμα προκαλεί ροή δεδομένων 128-112 Kbits/s



Σχήμα 2: πληροφορίες και διασυνδέσεις από το www.mpeg.org

Πίνακας 3. Δυνατότητες του MPEG-1 Layer 3

Ποιότητα ήχου	Συμπίεση	εύρος ζώνης	λειτουργία	ρυθμός δεδομένων
Τηλεφωνία	1:96	2.5	μονοφωνική	8
Καλύτερη από ραδιοφωνία βραχέων κυμάτων	1:48	4.5	μονοφωνική	16
Καλύτερη από ραδιοφωνία μεσαίων κυμάτων	1:24	7.5	μονοφωνική	32
Ραδιοφωνία FM	1:24 – 1:26	11	στερεοφωνική	56-64
Σχεδόν CD	1:16	15	στερεοφωνική	96
CD	1:12 – 1:14	>15	στερεοφωνική	112-128