



Αρχές Ψηφιακής Τηλεόρασης ***Ενότητα 4 – Συμπίεση ήχου***

Καθηγητής Δρ. Ευάγγελος Πάλλης
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών
Ελληνικό Μεσογειακό Πανεπιστήμιο

Περιεχόμενα

- Σύνοψη και μαθησιακά αποτελέσματα/στόχοι
- Εισαγωγή στην Ενότητα 4
- Ψηφιοποίηση ήχου
 - Βασικά στάδια ψηφιοποίησης
 - Ψηφιοποίηση στερεοφωνικού ήχου
 - Πολυκαναλικός ήχος
- Συμπίεση ήχου
 - Ακουστικό φάσμα και ευαισθησία
 - Επικάλυψη
 - Επικάλυψη στο πεδία συχνότητας
 - Επικάλυψη στο πεδίο του χρόνου
 - Κβαντισμός και θόρυβος κβαντισμού
- Κωδικοποίηση ήχου
 - Κωδικοποίηση υπομπάντας
 - Παράδειγμα επικάλυψης
 - Δημιουργία πλαισίου
 - Κωδικοποίηση μετασχηματισμού
 - Υβριδικό μοντέλο κωδικοποίησης
- Πρότυπα κωδικοποίησης ήχου MPEG

Περιγραφή Ενότητας 4

Σύνοψη

Στην Ενότητα 4 εξετάζουμε τα θέματα που αφορούν στη συμπίεση και την κωδικοποίηση του ήχου, ο οποίος αποτελεί ένα απαραίτητο τμήμα του τηλεοπτικού σήματος. Η συμπίεση του ήχου είναι σε μεγάλο βαθμό εφικτή με την εκμετάλλευση ιδιοτήτων της ανθρώπινης ακοής και συγκεκριμένα των ψυχοακουστικών χαρακτηριστικών, όπως για παράδειγμα στην επικάλυψη ορισμένων ήχων από άλλους οι οποίοι είναι κοντινοί είτε από πλευράς συχνότητας είτε χρονικά.

Προαπαιτούμενη γνώση

Η Ενότητα 4 είναι σε μεγάλο βαθμό αυτόνομη. Σε κάθε περίπτωση, οι μέθοδοι συμπίεσης και κωδικοποίησης του ήχου είναι συναφείς με τη συμπίεση και την κωδικοποίηση του βίντεο (βλ. Ενότητα 3), και για αυτό το λόγο ενδείκνυται η κατανόηση των εννοιών αυτών αλλά και εν γένει της ανάγκης για συμπίεση. Επίσης, ενδείκνυται, η γνώση της γενικής εικόνας του συστήματος και των επιμέρους βαθμίδων που το αποτελούν τόσο στον πομπό όσο και στον δέκτη, ώστε να γίνει αντιληπτή η σημασία της λειτουργικότητας που εξετάζουμε σε αυτή την Ενότητα.

Μαθησιακά αποτελέσματα

Με την επιτυχή ολοκλήρωση της Ενότητας 4 ο φοιτητής / τρια θα είναι σε θέση να:

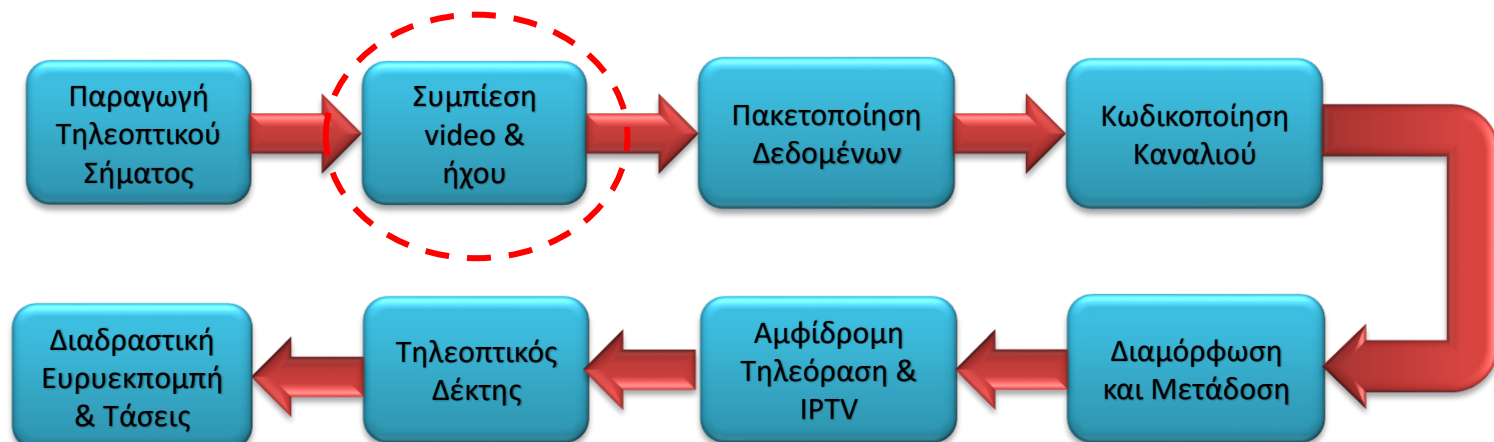
- Κατανοεί τον τρόπο λειτουργίας και οργάνωσης των τηλεοπτικών συστημάτων συμπίεσης πληροφορίας ήχου και τη χρήση τους στην παραγωγή, μετάδοση και λήψη ακουστικής πληροφορίας σε τηλεοπτικές υπηρεσίες.
- Γνωρίζει τις βασικές αρχές συμπίεσης/κωδικοποίησης του σήματος ήχου και τους μηχανισμούς απωλεστικής και μη-απωλεστικής αφαίρεσης πληροφορίας, με βάση ψυχοακουστικά χαρακτηριστικά, την επικάλυψη στο πεδίο του χρόνου και της συχνότητας.
- Εφαρμόζει μεθόδους αξιολόγησης και εργαλεία ανάλυσης των επιδόσεων των μηχανισμών συμπίεσης ήχου, καθώς και τεχνικές βέλτιστης συμπίεσης για την επίτευξη υψηλής ποιότητας σήματος στερεοφωνικού και πολυκαναλικού ήχου.
- Αναλύει και υπολογίζει τα βασικά χαρακτηριστικά συμπίεσης του σήματος ήχου, και των μηχανισμών κωδικοποίησης του σε τηλεοπτικά συστήματα μετάδοσης με βάση τα πρότυπα MPEG.
- Προτείνει λύσεις σε θέματα λειτουργίας και συντήρησης των μονάδων συμπίεσης του σήματος video, και τρόπους αξιολόγησης της παρεχόμενης ποιότητας.

Εισαγωγή

Εισαγωγή στην Ενότητα 4

Τα συστήματα ήχου στα τηλεοπτικά συστήματα πρέπει να καλύπτουν την απόκριση του ανθρώπινου αυτιού, το οποίο έχει δυναμικό εύρος ακοής περί τα 140 dB, με εύρος ζώνης από τις πολύ χαμηλές συχνότητες (κάποιες δεκάδες Hz) μέχρι περίπου τα 15 KHz με 20 kHz. Σε αυτή την Ενότητα θα δούμε ότι το σήμα του ήχου έχει σημαντικές απαιτήσεις αναφορικά με το εύρος ζώνης (όχι τόσο μεγάλες όσο το σήμα video), που σημαίνει ότι και για το σήμα του ήχου απαιτείται συμπίεση και κωδικοποίηση.

Η συμπίεση, αντίστοιχα με αυτά τα οποία είδαμε στην Ενότητα 3, στηρίζεται σε **μη απωλεστικές** και **απωλεστικές** μεθόδους, δηλαδή στην αφαίρεση α) της πλεονάζουσας πληροφορίας και β) της πληροφορίας η οποία δεν είναι μεν πλεονάζουσα αλλά η απώλειά της δεν γίνεται εύκολα αντιληπτή από τον χρήστη.



Εισαγωγή στην Ενότητα 4

Οι μέθοδοι συμπίεσης του ήχου εκμεταλλεύονται το ψυχοακουστικό μοντέλο, το οποίο αξιοποιεί κάποια βασικά χαρακτηριστικά του ανθρώπινου αυτιού, εκ' των οποίων το πιο σημαντικό είναι η **κάλυψη (masking)** η οποία συμβαίνει στο πεδίο του χρόνου αλλά και στο πεδίο της συχνότητας.

Πιο συγκεκριμένα:

- **Κάλυψη στο πεδίο του χρόνου:** ένας ήχος είναι πολύ κοντινός χρονικά σε κάποιον άλλο και το χρονικό διάστημα που μεσολαβεί μεταξύ τους είναι πολύ μικρό. Ο μεταγενέστερος χρονικά ήχος, δεν θα γίνει αντιληπτός (υπό ορισμένες συνθήκες), αφού θα επικαλυφθεί από αυτόν που προηγήθηκε.
- **Κάλυψη στο πεδίο της συχνότητας:** δύο ήχοι γειτνιάζουν στο πεδίο της συχνότητας και τα φάσματά τους έχουν πολύ μικρή απόσταση μεταξύ τους, με αποτέλεσμα αυτός που έχει την υψηλότερη ένταση να επισκιάζει τον άλλο. Για την πιο εύκολη διαχείριση του φάσματος, το φάσμα χωρίζεται σε επιμέρους 32 τμήματα (32 υπομπάντες).

Όπως είδαμε και στην περίπτωση του video, είναι δυνατή η χρήση του Διακριτού Μετασχηματισμού Συνημιτόνου (Discrete Cosine Transform - DCT) με στόχο την περαιτέρω συμπίεση, μόνο που εδώ χρησιμοποιούμε Τροποποιημένο Διακριτό Μετασχηματισμό Συνημιτόνου (Modified Discrete Cosine Transform - MDCT), ο οποίος έχεις κάποιες διαφοροποιήσεις στη διαχείριση των άκρων των διαστημάτων όπου εφαρμόζεται.

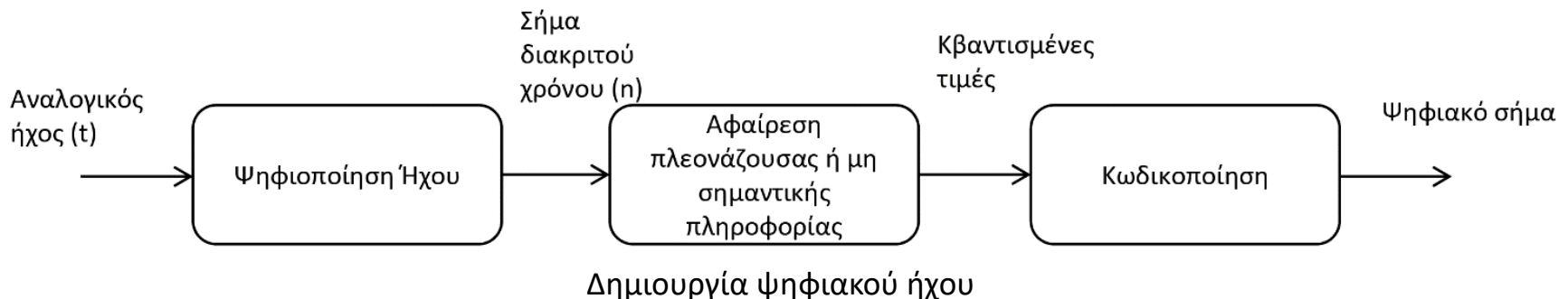
Εισαγωγή στην Ενότητα 4

Τα βασικά σημεία τα οποία θα εξετάσουμε στην Ενότητα 4 είναι:

- η ψηφιοποίηση του ήχου (δειγματοληψία και κβαντισμός),
- η αφαίρεση της πλεονάζουσας ή της μη σημαντικής (όσον αφορά την ποιότητα) πληροφορίας, και στη συνέχεια
- η κωδικοποίηση του ηχητικού σήματος.

Αναφορικά με τα πρότυπα τα οποία χρησιμοποιούνται, βασικό ρόλο παίζει το MPEG και μάλιστα το τρίτο κάθε φορά τμήμα του εκάστοτε προτύπου (part 3), το οποίο ασχολείται με τα θέματα του ήχου.

- Το MPEG-1 audio αποτελείται από τρία επίπεδα (layers) I, II και III. Το επίπεδο III δίνει το γνωστό πρότυπο mp3 (MPEG-1, layer III). Το MPEG-2 υιοθέτησε τα τρία επίπεδα του MPEG-1 όσον αφορά τον ήχο και το επίπεδο II επεκτάθηκε για να δημιουργήσει το επίπεδο IIMC (multichannel).

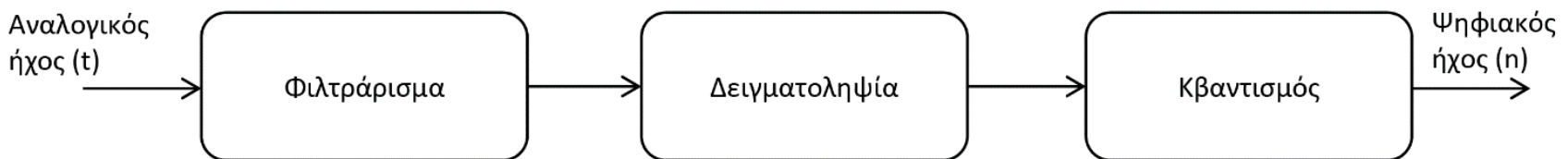


Ψηφιοποίηση ήχου

Βασικά στάδια ψηφιοποίησης

Τα βασικά βήματα της ψηφιοποίησης του ήχου είναι το φιλτράρισμα, η δειγματοληψία και ο κβαντισμός:

- **Φιλτράρισμα:** Αρχικά, το εύρος ζώνης του αναλογικού σήματος περιορίζεται με χρήση ενός Φίλτρου Χαμηλών Συχνοτήτων (ΦΧΣ). Η συχνότητα αποκοπής του οποίου βρίσκεται μεταξύ 15 KHz έως και 20 KHz.
- **Δειγματοληψία:** Στη συνέχεια γίνεται η μετατροπή του σήματος από σήμα συνεχούς χρόνου σε σήμα διακριτού χρόνου, με συχνότητα δειγματοληψίας 32 kHz, 44.1 kHz, 48 ή και 96 kHz. Η δειγματοληψία στα 44.1 KHz αντιστοιχεί στην ποιότητα του CD, ενώ η δειγματοληψία στα 48 KHz ή τα 96 KHz δίνουν ποιότητα αντίστοιχη αυτής του τηλεοπτικού στούντιο.



Τα βασικά στάδια της ψηφιοποίησης του ήχου

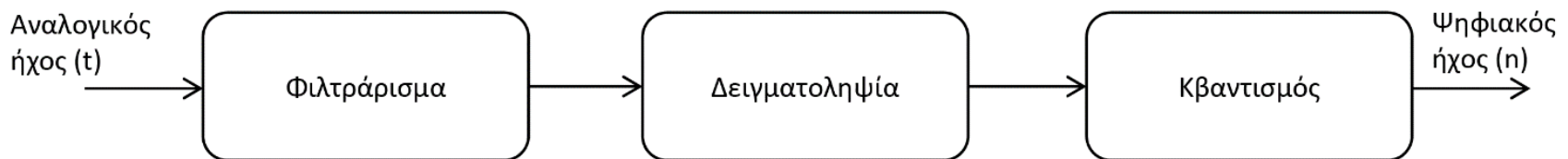
Βασικά στάδια ψηφιοποίησης

- **Κβαντισμός:** Μετά τη δειγματοληψία γίνεται ο κβαντισμός των τιμών των δειγμάτων, αντιστοίχιση δηλαδή των συνεχών τιμών σε διακριτές στάθμες. Η ποιότητα του κβαντισμού εξαρτάται από το πλήθος των bits που χρησιμοποιούνται για την αποτύπωση των τιμών. Εάν το πλήθος των bits που θα χρησιμοποιηθούν κατά την κβάντηση είναι N , τότε το πλήθων με τις στάθμες κβαντισμού θα είναι 2^N . Π.χ. στην περίπτωση κβαντισμού με χρήση 16 bits, δημιουργούνται 2^{16} στάθμες τιμών.

Ρυθμός δεδομένων (C_m) για δειγματοληψία μονοφωνικού ήχου στα 48 KHz και κβάντηση με χρήση 16 bits.

$$C_m = 48 \cdot 10^3 \cdot 16 = 768000 = 768 \text{ Kbps} \quad (1)$$

- Ο ρυθμός αυτός αυξάνεται στην περίπτωση πολυφωνικού ήχου, ο οποίος αποτελείται περισσότερα κανάλια (π.χ. δύο κανάλια στο στερεοφωνικό).



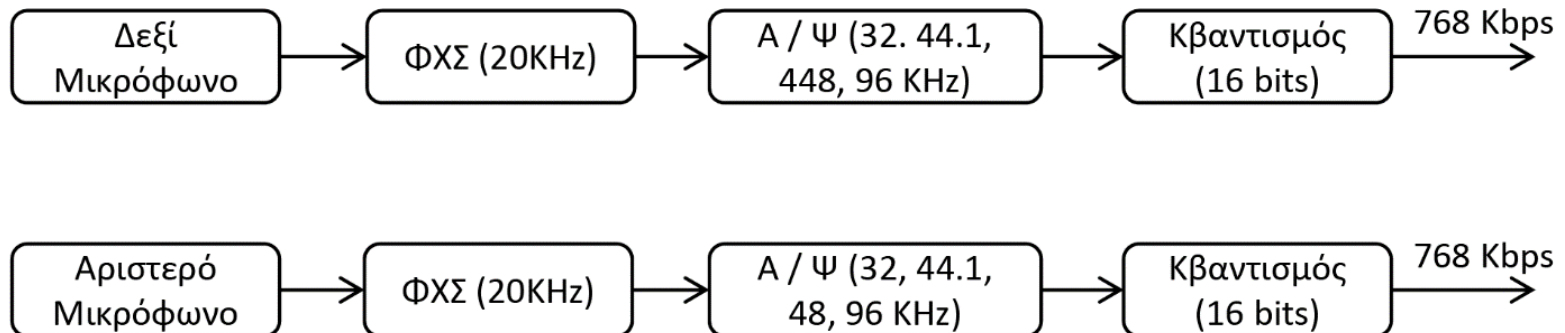
Τα βασικά στάδια της ψηφιοποίησης του ήχου

Ψηφιοποίηση στερεοφωνικού ήχου

Ο στερεοφωνικός ήχος αποτελείται από δύο κανάλια, είναι δηλαδή σύνθεση δύο ήχων που προκύπτουν από δύο συσκευές λήψης οι οποίες λειτουργούν ταυτόχρονα, το δεξί και το αριστερό μικρόφωνο.

- Ο ρυθμός δεδομένων για τον στερεοφωνικό ήχο είναι ο διπλάσιος του μονοφωνικού (προηγούμενη περίπτωση), λόγω της ύπαρξης των δύο καναλιών. Η νέα τιμή που υπολογίζουμε για τον στερεοφωνικό ήχο είναι η C_s :

$$C_s = 2 * 768 = 1.536 \text{ Kbps} = 1,5 \text{ Mbps} \quad (2)$$



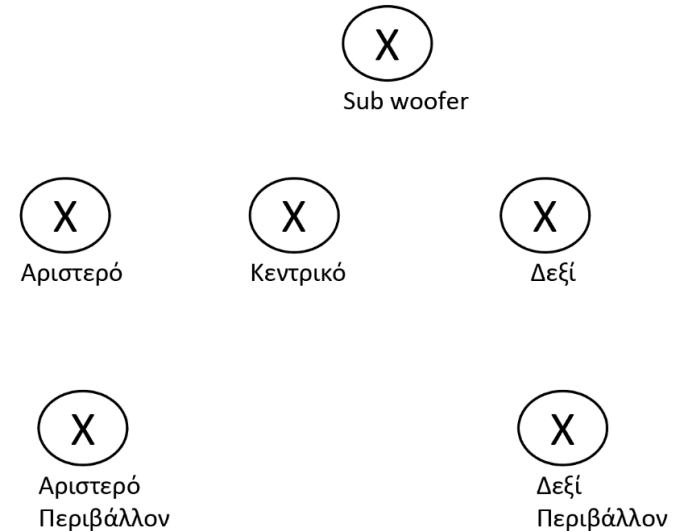
Δημιουργία στερεοφωνικού ήχου

Πολυκαναλικός ήχος

Στην περίπτωση του πολυκαναλικού ήχου, η κωδικοποίηση και η συμπίεση γίνονται και μεταξύ των καναλιών που αποστέλλονται.

- Τα κανάλια που μεταδίδονται αξιολογούνται για συσχετιζόμενη πληροφορία που δεν συνεισφέρει στο τελικό ηχητικό αποτέλεσμα. Αυτή η διαδικασία γίνεται, στο MPEG2 layerII MC (πολυκαναλικός ήχος Multi Channel) και στο Dolby Digital 5.1 Surround.

Μια σχετική αναπαράσταση του περιβάλλοντος **λήψης πολυκαναλικού ήχου** με βάση το μοντέλο 5.1 απεικονίζεται στο σχήμα. Σε αυτό το μοντέλο μεταδίδονται τα επόμενα κανάλια: αριστερό, κέντρο, δεξί, αριστερό περιβάλλον (surround), δεξί περιβάλλον (surround) και το ενισχυτικό των χαμηλών συχνοτήτων (Low-Frequency Enhancement, LFE) κανάλι για χρήση σε sub-woofer.



Δημιουργία πολυκαναλικού ήχου

Συμπίεση ήχου

Ακουστικό Φάσμα και ευαισθησία

Η ροή δεδομένων που προκύπτει από την ψηφιοποίηση του ήχου πρέπει να συμπιεστεί και ο αντίστοιχος ρυθμός να μειωθεί από τα 1.5 Mbps (στην περίπτωση του στερεοφωνικού ήχου) σε μια τιμή μεταξύ 100 Kbps και 400 Kbps (δηλ. ποσοστό συμπίεσης από 16:1 για τη χαμηλή ποιότητα, έως και 4:1 για την πιο υψηλή ποιότητα).

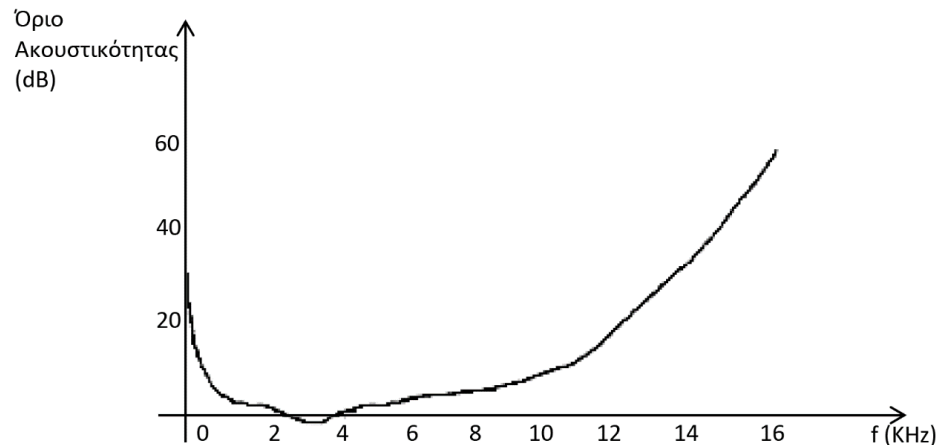
Μεγάλη σημασία στη δραστική συμπίεση του ηχητικού σήματος έχει η ευαισθησία και εν γένει η συμπεριφορά του αυτιού στις διάφορες συχνότητες.

- **Συχνότητες του ήχου:** Το εύρος ζώνης των ήχων που γίνονται αντιληπτοί από το ανθρώπινο αυτί είναι, από τα 0 KHz έως και τα 20 KHz. Οι τιμές αυτές μεταβάλλονται αφού με το πέρασμα των χρόνων το φάσμα των συχνοτήτων που αντιλαμβανόμαστε περιορίζεται. Ήχοι εκτός αυτού του εύρους ζώνης δεν γίνονται αντιληπτοί από τον άνθρωπο.
- **Ένταση του ήχου:** Το όριο αυτό της έντασης, κάτω από το οποίο ένας ήχος δεν γίνεται αντιληπτός, διαφοροποιείται σε σχέση με τη συχνότητα. Η μέγιστη ευαισθησία του ανθρώπου στον ήχο είναι στο εύρος μεταξύ 3 KHz και 4 KHz. Στο συγκεκριμένο διάστημα η ευαισθησία του αυτιού μεγιστοποιείται και εκτός αυτού μειώνεται με σχετικά γρήγορο ρυθμό και προς τις χαμηλότερες και προς τις μεγαλύτερες συχνότητες.

Ακουστικό Φάσμα και ευαισθησία

Στο σχήμα απεικονίζεται η καμπύλη ευαισθησίας του ανθρώπινου αυτιού, σε σχέση με τη συχνότητά τους (το διάγραμμα έχει προκύψει από μετρήσεις ευρείας κλίμακας και αξιοπιστίας).

- Παρατήρηση: το όριο «ακουστικότητας» στις συχνότητες περί τα 3KHz παρουσιάζει το ολικό ελάχιστο και αυξάνει στις υπόλοιπες. Αυτό σημαίνει ότι στις συχνότητες αυτές μπορούμε να ακούμε ήχους με την ελάχιστη ένταση, ενώ για να μπορέσουμε να ακούσουμε ήχους στις υπόλοιπες συχνότητες απαιτείται μεγαλύτερη ένταση.



Καμπύλη ευαισθησίας ανθρώπινου αυτιού σε σχέση με τις συχνότητες των ήχων

Ακουστικό Φάσμα και ευαισθησία

Με βάση την παραπάνω συζήτηση προκύπτει ότι κάποιοι ήχοι δεν γίνονται αντιληπτοί από τον άνθρωπο και συνεπώς δεν χρειάζεται να κωδικοποιούνται.

Τέτοιοι ήχοι είναι:

- Οι ήχοι των οποίων το φασματικό περιεχόμενο (οι συχνότητες δηλαδή από τις οποίες αποτελούνται) είναι εκτός του εύρους συχνοτήτων 20 Hz έως και 20 KHz. Για αυτούς τους ήχους χρησιμοποιούμε το **Φίλτρο Χαμηλών Συχνοτήτων (ΦΧΣ)**, το οποίο περιορίζει το φασματικό περιεχόμενο του προς μετάδοση ήχου μεταξύ 20Hz και 20 KHz. Απαιτείται προσοχή στην επιλογή της συχνότητας δειγματοληψίας ώστε να είναι κατ' ελάχιστον το διπλάσιο της μέγιστης τιμής της συχνότητας του ηχητικού σήματος που διέρχεται από το φίλτρο. Στην περίπτωση, για παράδειγμα, όπου το ΦΧΣ επιτρέπει τη διέλευση συχνοτήτων μέχρι τα 20 KHz, η συχνότητα δειγματοληψίας πρέπει να είναι κατ' ελάχιστον 40 KHz ώστε να ικανοποιείται το **Θεώρημα του Shannon** (η συχνότητα αυτή είναι γνωστή και ως **συχνότητα Nyquist**).
- Οι ήχοι των οποίων η ένταση είναι κάτω από το ελάχιστο όριο ακοής για τις αντίστοιχες συχνότητες στις οποίες εκτείνονται. Για αυτούς τους ήχους γίνεται χρήση μιας **βαθμίδα σύγκρισης της έντασης** του εισερχόμενου ήχου με το όριο ακουστικότητας, στην αντίστοιχη συχνότητα. Αυτό σημαίνει ότι εάν η ένταση κάποιου ήχου δεν υπερκαλύπτει αυτό το όριο, ο ήχος μπορεί να αγνοείται.

Επικάλυψη στο πεδίο της συχνότητας

Έστω ότι έχουμε ένα ήχο ο οποίος έχει συχνότητα f_0 kHz με σταθερό πλάτος, και παράλληλα υπάρχουν και άλλοι ήχοι με συχνότητες γύρω από την f_0 . Για να γίνουν αντιληπτοί οι άλλοι ήχοι πρέπει να υπερβούν κάποιο συγκεκριμένο πλάτος (διαφορετικά επικαλύπτονται). **Το φαινόμενο αυτό ονομάζεται επικάλυψη στο πεδίο της συχνότητας, δηλαδή, ο ήχος στη συχνότητα f_0 επικαλύπτει τους ήχους στις γύρω συχνότητες.**

Από μετρήσεις στο πλαίσιο πειραμάτων της ανθρώπινης ακοής, προκύπτουν οι επόμενες παρατηρήσεις:

- Η μορφή της επικάλυψης εξαρτάται από τη συχνότητα του επικαλύπτοντος ήχου.
- Όσο πιο υψηλή είναι η συχνότητα του επικαλύπτοντος ήχου τόσο μεγαλύτερο το (συχνοτικό) εύρος της επικάλυψης.

Συμπέρασμα: ήχοι οι οποίοι επικαλύπτονται δεν χρειάζεται να κωδικοποιούνται και να αποστέλλονται άρα πρέπει να αγνοούνται κατά τη συμπίεση.

Επικάλυψη στο πεδίο της συχνότητας

Το φαινόμενο της επικάλυψης στο πεδίο της συχνότητας φαίνεται σε αυτό το σχήμα, στο οποίο διακρίνονται οι καμπύλες επικάλυψης.

- Παρατηρούμε ότι εάν έχουμε έναν ήχο του οποίου η συχνότητα αντιστοιχεί στη συχνότητα του κεντρικού σημείου μιας καμπύλης και η ένταση του ήχου αντιστοιχεί στην κορυφή της καμπύλης, τότε ο ήχος αυτός επικαλύπτει το σύνολο των (τυχόν) ήχων οι οποίοι αντιστοιχούν, όσον αφορά τη συχνότητα και την ένταση στην επιφάνεια κάτω από τη συγκεκριμένη καμπύλη.

Για να μπορέσει να γίνει αντιληπτός ένας ήχος (εάν η συχνότητά του αντιστοιχεί στο εύρος συχνοτήτων που καλύπτει η καμπύλη) πρέπει η ένταση αυτού να υπερβαίνει τον ορίζοντα της καμπύλης.



Επικάλυψη στο πεδίο του χρόνου

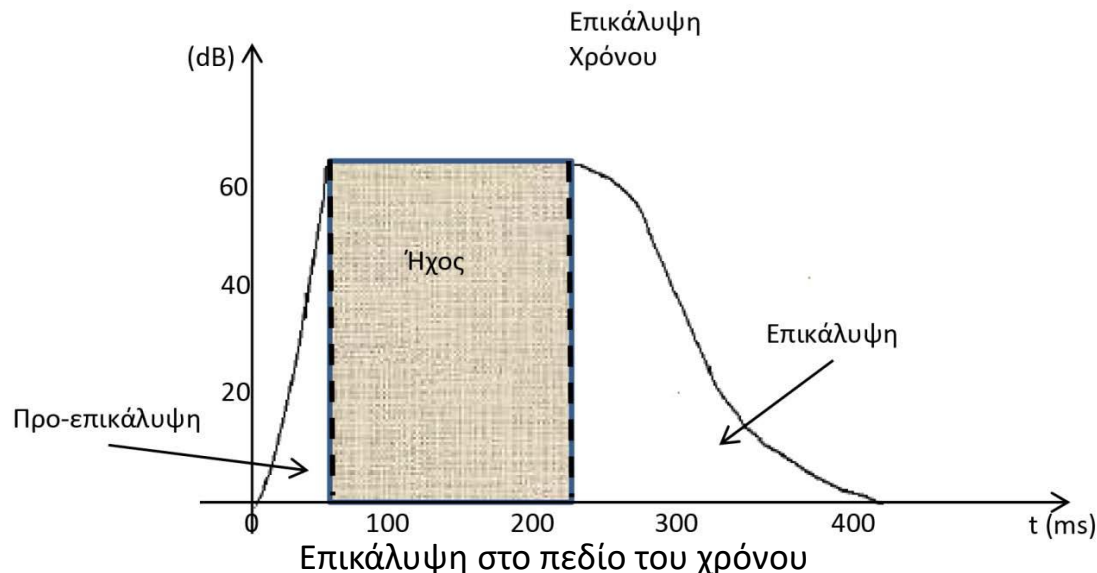
Επικάλυψη στο πεδίο του χρόνου έχουμε όταν ένας έντονος ήχος (μεγάλης έντασης) επικαλύπτει άλλους ήχους (δεν επιτρέπει να γίνουν αντιληπτοί) οι οποίοι χρονικά είναι πριν και μετά από αυτόν (εφόσον η ένταση αυτών των ήχων δεν ξεπερνάει κάποιο κατώφλι).

- Η επικάλυψη πριν το χρονικό διάστημα που αναπαράγεται ο ήχος οφείλεται στην περιορισμένη ευαισθησία του ανθρώπινου αυτιού αλλά και τον τρόπο με τον οποίο μεταδίδονται τα σήματα από το αυτί στον εγκέφαλο.

Επικάλυψη στο πεδίο του χρόνου

Το σχήμα σε αυτή τη διαφάνεια παρουσιάζει το φαινόμενο της επικάλυψης στο πεδίο του χρόνου. Δείχνει έναν ήχο (γκρι επιφάνεια) ο οποίος εκτείνεται, έχοντας σταθερή ένταση, στο χρονικό διάστημα που αντιστοιχεί στη χρονική στιγμή των 80 ms και στη χρονική στιγμή των 220 ms.

- Ο ήχος αυτός προκαλεί επικάλυψη πριν (οπότε έχουμε το φαινόμενο της προεπικάλυψης) και (κυρίως) μετά (οπότε έχουμε την επικάλυψη ή κυρίως επικάλυψη, ή μετα-επικάλυψη).
- Η επικάλυψη αφορά το χρονικό διάστημα από την έναρξη του χρόνου μέχρι τα 400 ms.



Κβαντισμός και θόρυβος κβαντισμού

Ο κβαντισμός καθορίζεται από το πλήθος των bits, έστω N , που αξιοποιούνται για την αναπαράσταση των τιμών του σήματος. Στην περίπτωση του ήχου, το πλήθος των bits που συνήθως χρησιμοποιείται είναι 16 (παλιότερα 8 bits), και καθορίζει το πλήθος από τις στάθμες κβαντισμού 2^{16} **και συνεπώς και τον θόρυβο κβαντισμού**. Ο θόρυβος κβαντισμού προκύπτει από την ψηφιοποίηση ενός αναλογικού σήματος και, όπως έχουμε δει σε προηγούμενες Ενότητες, ο λόγος σήματος προς θόρυβο (Signal to Noise Ratio) είναι περίπου $6N$, όπου N το πλήθος των bits που χρησιμοποιούνται για τον κβαντισμό.

Αυτό σημαίνει ότι:

- Με χρήση 8 bits έχουμε σηματοθορυβικό λόγο 48 dB
- Με χρήση 16 bits έχουμε σηματοθορυβικό λόγο 96 dB

Η διαφορά των δύο επιπέδων είναι 48 dB η οποία αντιστοιχεί στο γινόμενο $6 \cdot 8 = 48$ όπου 8 είναι τα επιπλέον χρησιμοποιούμενα bits.

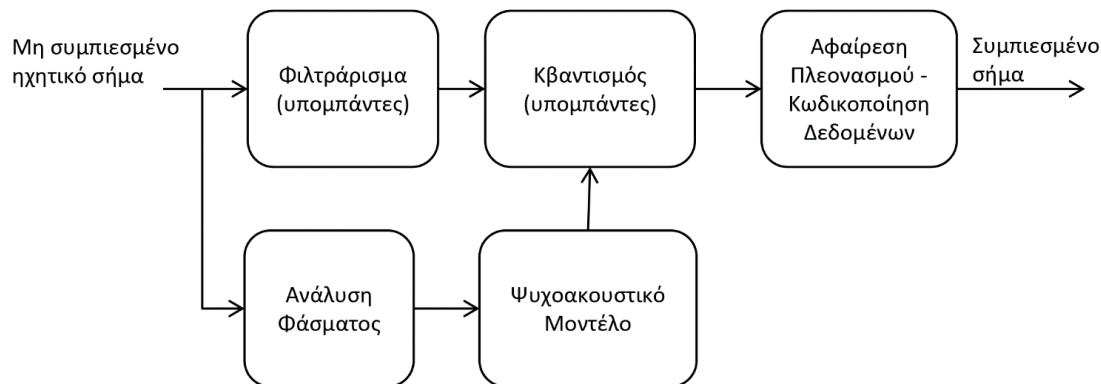
Όπως είπαμε προηγουμένως, το εύρος της ακοής του ηχητικού σήματος από το ανθρώπινο αυτί είναι περίπου 140 dB γεγονός το οποίο **δικαιολογεί την επιλογή ακόμη μεγαλύτερης ακρίβειας κβαντισμού (δηλ. περισσότερων bits)**.

Κωδικοποίηση ήχου

Βήματα κωδικοποίησης ήχου

Στο σχήμα διακρίνονται οι δύο διακριτές διαδρομές στον κωδικοποιητή:

- Στην μία διαδρομή (κάτω σκέλος), το σήμα αφού φιλτραριστεί, οδηγείται στον **αναλυτή φάσματος**, όπου μετασχηματίζεται κατά Διακριτό Μετασχηματισμό Φουριέ (ΔΜΦ), ο οποίος υλοποιείται μέσω του αλγορίθμου του Ταχέως Μετασχηματισμού Φουριέ. Αφού πραγματοποιηθεί ο Ταχύς Μετασχηματισμός Φουριέ (Fast Fourier Transform, FFT) ξεχωρίζουν τα τμήματα του ηχητικού σήματος και τα εύρη των συχνοτήτων που αυτά ανήκουν.
- Στην άλλη διαδρομή (άνω σκέλος), το ηχητικό σήμα υφίσταται **φιλτράρισμα** και χωρίζεται σε **υπομπάντες** (οι οποίες αφορούν συγκεκριμένες και περιορισμένες περιοχές συχνοτήτων). Οι υπομπάντες αυτές είναι εν γένει πιο εύκολα διαχειρίσιμες, όσον αφορά τη θεώρηση των φαινομένων κάλυψης, από το συνολικό ηχητικό σήμα (το οποίο εκτείνεται στο συνολικό εύρος ζώνης).



Βήματα κωδικοποίησης ηχητικού σήματος

Βήματα κωδικοποίησης ήχου

Με χρήση του ψυχοακουστικού μοντέλου και των φαινομένων της επικάλυψης συχνότητας, μπορούν να εντοπιστούν τα τμήματα του ηχητικού σήματος τα οποία δύναται να παραλειφθούν. Συγκεκριμένα:

- Μια υπομπάντα μπορεί να επικαλύπτεται από κάποια άλλη (επειδή η ένταση του ήχου σε αυτήν την υπομπάντα είναι πιο χαμηλή από το όριο). Αυτό σημαίνει ότι **το περιεχόμενο της συγκεκριμένης υπομπάντας δεν χρειάζεται να μεταδοθεί**.
- Μια υπομπάντα μπορεί να μην επικαλύπτεται οριακά. Ο **κβαντισμός** στα δείγματα της υπομπάντας αυτής **μπορεί να γίνει όχι στις αρχικές τιμές αλλά στις διαφορές των πραγματικών τιμών** από τα αντίστοιχα όρια ακοής (για τη συγκεκριμένη υπομπάντα).
- Επιπλέον **η ακρίβεια της κβάντισης μπορεί να διαφέρει**, δίνοντας, για παράδειγμα, έμφαση (με την ανάθεση περισσότερων bits) στις χαμηλότερες συχνότητες και μικρότερη (μικρότερο πλήθος bits κβαντισμού) στις υψηλότερες.

Η απόφαση αναφορικά με το αν θα παραλείψουμε μια υπομπάντα ή εάν αυτή θα κβαντιστεί με μικρότερη ακρίβεια, εξαρτάται από την εφαρμογή των ψυχοακουστικών μοντέλων, η οποία λαμβάνει υπόψη της και το φασματικό περιεχόμενο του σήματος. **Ο κβαντισμός ελέγχεται από τον κβαντιστή κάθε υπομπάντας.**

Ο κβαντισμός ακολουθείται από **περιορισμό του πλεονασμού**, ο οποίος γίνεται με τη βοήθεια ειδικής κωδικοποίησης. Με το πέρας αυτών των διαδικασιών είναι διαθέσιμο το συμπιεσμένο ψηφιακό ηχητικό σήμα.

Κωδικοποίηση υπομπάντας

Στο πρότυπο MPEG (και πιο συγκεκριμένα τα επίπεδα I και II) το ηχητικό σήμα διέρχεται από μια συστοιχία φίλτρων, τα οποία χωρίζουν το σήμα σε υπομπάντες των 750 Hz.

Για κάθε υπομπάντα υπάρχει ξεχωριστός κβαντιστής, ο οποίος ελέγχεται από το μπλοκ του ΔΜΦ (ο οποίος εφαρμόζει τον αλγόριθμο του Ταχέως Μετασχηματισμού Φουριέ) και το ψυχοακουστικό μοντέλο.

- Ο κβαντιστής έχει τη δυνατότητα να απορρίψει εξολοκλήρου την υπομπάντα ή να προσαρμόσει (περιορίσει) την ακρίβεια του κβαντισμού.

Σύμφωνα με το πρότυπο MPEG Layer II, ο ΔΜΦ λαμβάνει χώρα κάθε 24 milliseconds και εφαρμόζεται σε 1.024 δείγματα. **Αυτό σημαίνει ότι η πληροφορία που δίνεται στο ψυχοακουστικό μοντέλο αλλάζει κάθε 24 ms.**

Στο χρονικό διάστημα αυτό των 24 ms οι υπομπάντες πραγματοποιούν συμπίεση με βάση την πληροφορία που έχουν λάβει από το μπλοκ του ψυχοακουστικού μοντέλου. **Αυτό σημαίνει ότι ο μηχανισμός προκαλεί μια καθυστέρηση.**

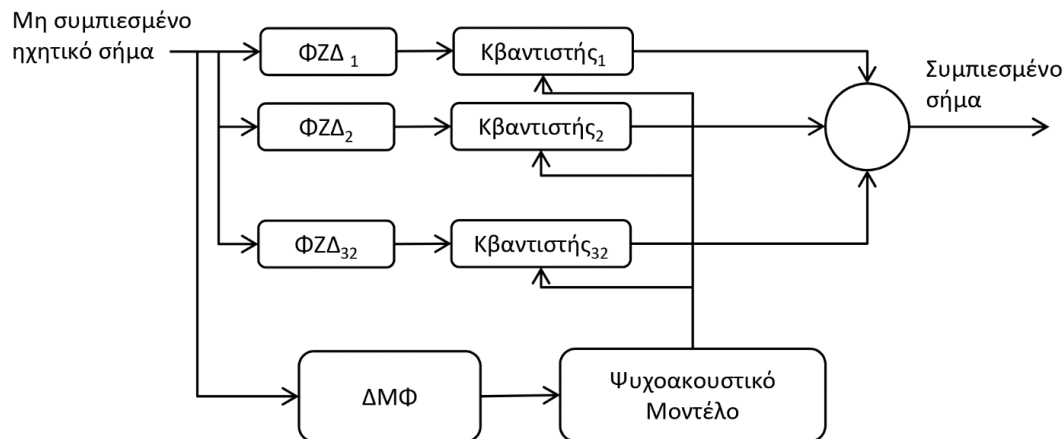
Κωδικοποίηση υπομπάντας

Στο σχήμα διακρίνεται ο μηχανισμός σύμφωνα με τον οποίο το σήμα με τη βοήθεια Φίλτρων Ζώνης Διέλευσης (ΦΖΔ) χωρίζεται στις 32 υπομπάντες, και για καθεμία από τις οποίες λαμβάνει χώρα ο κβαντισμός.

Η απόφαση για τον κβαντισμό λαμβάνεται από τον παράλληλο κλάδο που περιλαμβάνει το **Διακριτό Μετασχηματισμό Φουριέ και το ψυχοακουστικό μοντέλο**. Ο ΔΜΦ πραγματοποιείται με:

- 512 σημεία στην περίπτωση του MPEG Layer I
- 1024 σημεία στην περίπτωση των MPEG Layer II

Οι διαδικασίες αυτές λαμβάνουν χώρα σε χρονικό διάστημα των 24 ms, με βάση τη δειγματοληψία του ηχητικού σήματος. **Το σύνολο των συνιστωσών του ηχητικού σήματος στις εξόδους των κβαντιστών αποτελούν το συμπιεσμένο σήμα.**



Κωδικοποίηση ηχητικού σήματος στις υπομπάντες

Κωδικοποίηση υπομπάντας

Η ανάθεση του πλήθους των bits για τον κβαντισμό (άρα και η ποιότητα του κβαντισμού) είναι διαφορετική για κάθε υπομπάντα. Ο κβαντισμός να γίνεται

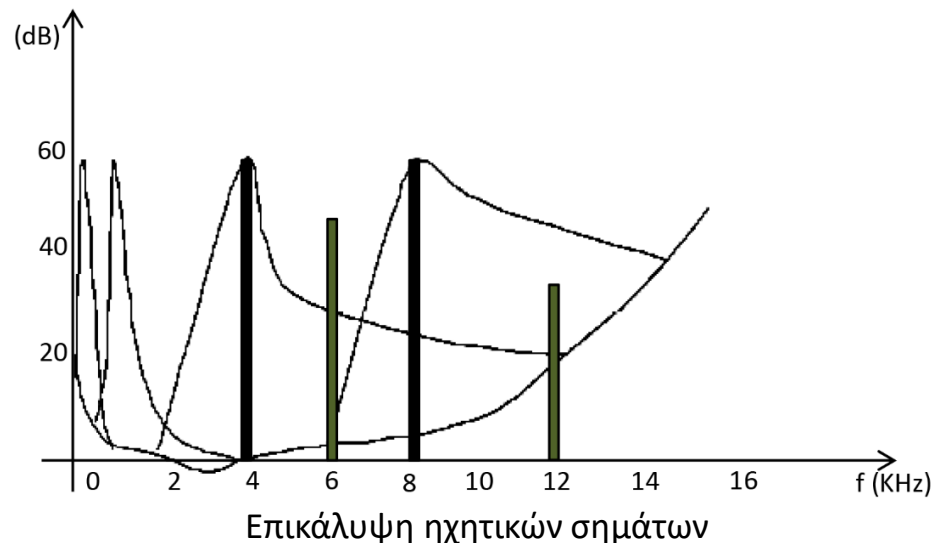
- Με μεγαλύτερη λεπτομέρεια στις πιο χαμηλές συχνότητες (π.χ. στα 3 KHz όπου εμφανίζεται και το μέγιστο) και
- Με μικρότερη λεπτομέρεια στις πιο υψηλές συχνότητες.

Σημείωση: οι υπομπάντες εξετάζονται για το κατά πόσο περιέχουν αρμονικές σημάτων που ανήκουν σε επόμενες υπομπάντες, δηλ. κατά πόσο, οι επικαλυπτόμενοι ήχοι αποτελούν αρμονικές ή είναι ξεχωριστοί ήχοι. **Στην περίπτωση των αρμονικών οι ήχοι αυτοί μπορούν να εξαλειφθούν πλήρως.**

Παράδειγμα επικάλυψης

Έστω ότι έχουμε δύο απλά ηχητικά σήματα (τόνους) με υψηλή ένταση τα οποία βρίσκονται στα **4 KHz** και στα **8 KHz**, όπως προκύπτει από το Διακριτό Μετασχηματισμό Φουριέ. Επιπλέον υπάρχουν άλλα δύο ηχητικά σήματα (τόνοι) στα **6 KHz** και στα **12 KHz**, όπως παρουσιάζεται στην εικόνα.

- Θα εξετάσουμε κατά πόσο υπάρχει επικάλυψη μεταξύ των σημάτων, ποια πρέπει κωδικοποιηθούν και ποια θα μπορούσαν να αγνοηθούν.



Παράδειγμα επικάλυψης

Διακρίνουμε τις δύο περιπτώσεις:

- **Περίπτωση 1:** Η ένταση του σήματος στη συχνότητα 6 KHz παρατηρούμε ότι υπερβαίνει το κατώφλι επικάλυψης. Το τμήμα της επικάλυψης οφείλεται στον ήχο ο οποίος έχει συχνότητα 4 KHz. Στην περίπτωση αυτή περιορίζεται το πλήθος των επιπέδων κβαντισμού.
- **Περίπτωση 2:** Σε επόμενη υπομπάντα παρατηρούμε ότι το σήμα στα 12KHz έχει ένταση η οποία είναι κάτω από το κατώφλι επικάλυψης. Αυτό σημαίνει ότι ο ήχος αυτό δεν θα γίνει αντιληπτός από τον ακροατή, αφού επικαλύπτεται από τον ήχο στα 8 KHz ο οποίος ανήκει σε προηγούμενη υπομπάντα. Ως αποτέλεσμα, το περιεχόμενο της τρέχουσας υπομπάντας δεν θα γίνει αντιληπτό, άρα μπορεί να αγνοηθεί.

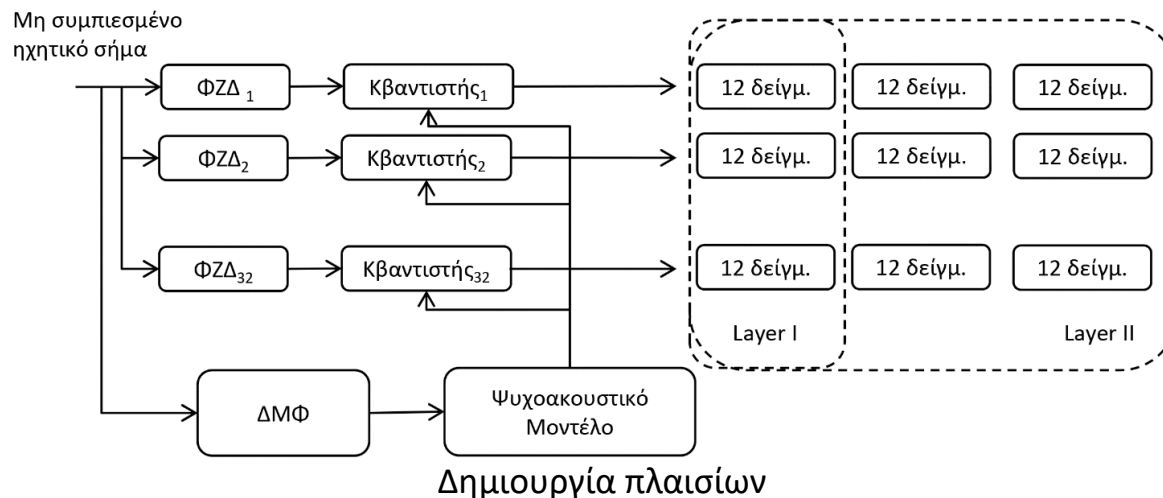
Συμπέρασμα: Παρατηρούμε ότι εκ των τεσσάρων ήχων (τόνων), ο ένας μπορεί να αγνοηθεί ενώ άλλος ένας κωδικοποιείται με μικρότερο πλήθος bits.

Δημιουργία πλαισίου

Στην κωδικοποίηση του MPEG τα δείγματα συνδυάζονται σε πλαίσια.

- Ένα πλαίσιο του επιπέδου I (layer I) αποτελείται από τμήμα 12 δειγμάτων για κάθε υπομπάντα.
- Ένα πλαίσιο του επιπέδου II (layer II) αποτελείται από 3 τμήματα των 12 δειγμάτων (έκαστο) για κάθε υπομπάντα.

Για κάθε τμήμα των 12 δειγμάτων, εντοπίζεται το δείγμα με τη μέγιστη τιμή. Η τιμή αυτή χρησιμοποιείται για την κλιμάκωση των υπόλοιπων δειγμάτων και εφαρμόζεται σε όλα τα 12 δείγματα του μπλοκ **για αφαίρεση του φαινομένου της κάλυψης**.

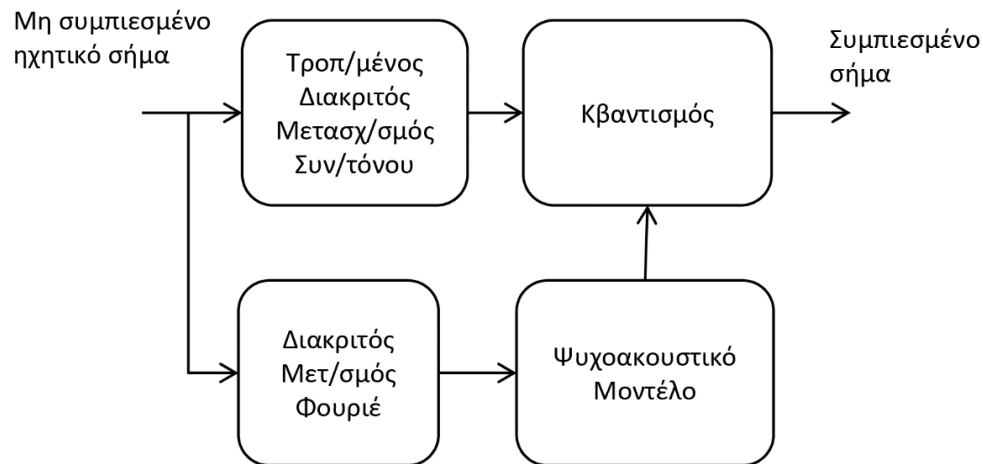


Κωδικοποίηση μετασχηματισμού

Η κωδικοποίηση μετασχηματισμού, μετασχηματίζει την πληροφορία από το πεδίο του χρόνου στο πεδίο της συχνότητας με χρήση του Διακριτού Μετασχηματισμού Φουριέ (Discrete Fourier Transform).

- Χρησιμοποιεί τον Τροποποιημένο Διακριτό Μετασχηματισμό Συνημιτόνου (Modified Discrete Fourier Transform, MDCT) ώστε να επεξεργαστεί το ηχητικό σήμα και να δώσει 256 ή 512 τιμές στο πεδίο της συχνότητας. Αυτό γίνεται με τον ίδιο τρόπο που λειτουργεί η κωδικοποίηση που βασίζεται στις υπομπάντες.
- Ο Ταχύς Μετασχηματισμός Φουριέ (Fast Fourier Transform, FFT) εκτελείται με μεγάλη ακρίβεια και ταχύτητα στο πεδίο της συχνότητας.

Σύμφωνα με το ψυχοακουστικό μοντέλο, οι τιμές που προκύπτουν από τον Τροποποιημένο Διακριτό Μετασχηματισμό Φουριέ υφίστανται κβαντισμό. **Η μέθοδος αυτή υπερέχει σε σχέση με τη μέθοδο που διαχωρίζει το φασματικό περιεχόμενο του ηχητικού σήματος σε υπομπάντες στο γεγονός ότι μπορεί να επιτύχει μεγαλύτερη ανάλυση στο πεδίο της συχνότητας.**



Κωδικοποίηση μετασχηματισμού

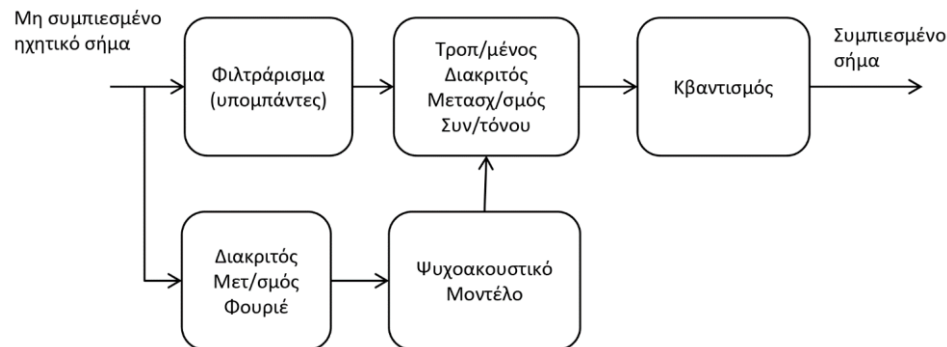
Υβριδικό μοντέλο κωδικοποίησης

Εκτός από τους δύο προηγούμενους τρόπους κωδικοποίησης (δηλ. κωδικοποίηση υπομπάντας και κωδικοποίηση μετασχηματισμού) υπάρχει, επίσης, η δυνατότητα συνδυασμού αυτών των δύο μηχανισμών, που είναι γνωστός ως **υβριδική κωδικοποίηση**.

Σε αυτόν τον τρόπο μετασχηματισμού το φιλτράρισμα σε υπομπάντες πραγματοποιείται πριν τον μετασχηματισμό MDCT.

- Αυτό σημαίνει ότι γίνεται ένας διαχωρισμός σε υπομπάντες και στη συνέχεια εφαρμόζεται ο μετασχηματισμός MDCT σε κάθε υπομπάντα για καλύτερη ανάλυση, όπως απεικονίζεται παρακάτω.

Τα δεδομένα που προκύπτουν από τον μετασχηματισμό MDCT υποβάλλονται σε συμπίεση με χρήση του ψυχοακουστικού μοντέλου το οποίο τροφοδοτείται από την έξοδο του Ταχέως Μετασχηματισμού Φουριέ. Αυτός είναι ο τρόπος κωδικοποίησης χρησιμοποιείται στο MPEG – layer 3 (δηλαδή στα .mp3).



Υβριδικό μοντέλο επικάλυψης

Πρότυπα κωδικοποίησης ήχου MPEG

Πρότυπα κωδικοποίησης ήχου MPEG

Οι τεχνικές κωδικοποίησης α) με διαίρεση σε μπάντες και β) με μετασχηματισμό, χρησιμοποιήθηκαν στα πρώτα πρότυπα του MPEG που είχαν να κάνουν με τη συμπίεση του ήχου (πρότυπο MPEG-1) αρχές της δεκαετίας του 1990 (ISO/IEC 11172-3), το οποίο αποτελείται από τρία επίπεδα (layers) I, II και III για τον ήχο . **Το επίπεδο III δίνει το γνωστό πρότυπο mp3 (MPEG-1, layer III).**

Το MPEG-2 υιοθέτησε τα τρία επίπεδα του MPEG-1 όσον αφορά τον ήχο και το επίπεδο II επεκτάθηκε για να δημιουργήσει το επίπεδο II MC, το οποίο είναι και το πολυκαναλικό (multichannel).

- Το πρότυπο ISO/IEC 13818-3 του MPEG-2 για τον ήχο υιοθετήθηκε το 1994. Το MPEG-2 χρησιμοποιείται στην τηλεόραση τυπικής ευκρίνειας (εικόνας και ήχου) αλλά επεκτάθηκε και για την τηλεόραση υψηλής ευκρίνειας. Αυτό είχε ως αποτέλεσμα το MPEG-3 να εγκαταλειφθεί και να μην προχωρήσει περαιτέρω. Έκτοτε έχουν υπάρξει διαφοροποιήσεις με κυριότερες το MPEG-2 AAC ISO/IEC 13818-7 (όπου AAC είναι το Advanced Audio Coding) και το σύστημα ήχου του MPEG-4 ISO/IEC 14496-3: AAC και AAC Plus.

Πρότυπα κωδικοποίησης ήχου MPEG

Το πρότυπο **Advanced Audio Coding (AAC)** έχει σχεδιαστεί ως διάδοχος του προτύπου MPEG 1 & 3 Layer 3, και μπορεί να πετύχει καλύτερη ποιότητα με αντίστοιχο ρυθμό.

Το AAC προτυποποιήθηκε από τους οργανισμούς ISO και IEC, ως τμήμα του MPEG-2 (ως **PART 7 και συγκριμένα ως το πρότυπο ISO/IEC 13818-7:1997**) και του MPEG-4.

- Το AAC υποστηρίζει τον συνδυασμό μέχρι και 48 καναλιών, πλήρους εύρους ζώνης (μέχρι δηλαδή και τα 96 kHz) σε μία ροή και επιπλέον 16 κανάλια περιορισμένης συχνότητας (Low Frequency Effects, LFE) μέχρι δηλαδή τα 120 Hz, 16 διαλογικά κανάλια και 16 ροές δεδομένων.

Αναφορικά με το MPEG 2, το Part 7 ήταν ένα νέο διαφορετικό τμήμα αφού πιο πριν ο ήχος εντάσσεται στο Part 3. Το Part 7 δεν είχε συμβατότητα προς τα πίσω. Δεν ήταν, δηλαδή, συμβατό με τα mp1, mp2 και mp3. Για αυτόν τον λόγο αναφερόταν ως μη συμβατό MPEG-2 NBC (Non-Backward Compatible). **Το MPEG-2 Part 7** όρισε τρία προφίλ:

- χαμηλής πολυπλοκότητας (Low-Complexity profile, AAC-LC)
- βασικό προφίλ (Main profile, AAC-Main)
- κλιμακωτού ρυθμού δειγματοληψίας (Scalable Sampling Rate, AAC-SSR)

Πρότυπα κωδικοποίησης ήχου MPEG

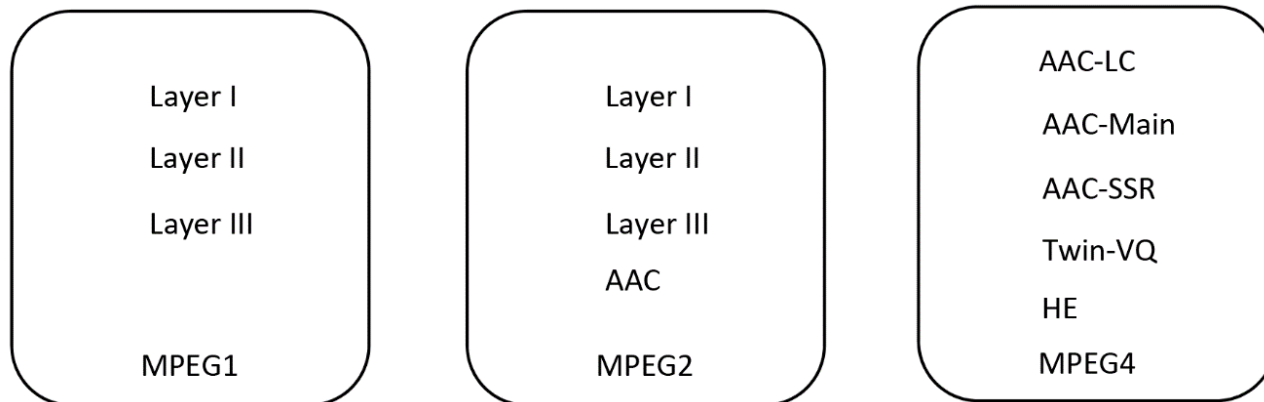
Βασικές επεκτάσεις του AAC σε σχέση με το MP3.

- Υποστηρίζει μεγαλύτερο εύρος ρυθμών δειγματοληψίας (από 8 KHz έως 96 kHz) σε σχέση με το MP3 (το οποίο υποστήριζε από 16 KHz έως και 48 kHz). Αυτό συνεπάγεται καλύτερη απόδοση στα σήματα με τις σχετικά υψηλές συχνότητες, δηλαδή στις συχνότητες άνω των 16 KHz.
- Υποστηρίζει μέχρι και 48 κανάλια ήχου με υψηλή συχνότητα δειγματοληψίας (το MP3 υποστηρίζει δύο κανάλια στο MPEG-1 και μέχρι 5.1 κανάλια στο MPEG-2).
- Υποστηρίζονται διαφορετικοί ρυθμοί κωδικοποίησης και μεταβλητό μήκος πλαισίου (σε αντίθεση με το mp3 όπου ο ρυθμός ήταν προκαθορισμένος).
- Υποστηρίζει απλουστευμένη λειτουργία όσον αφορά το αρχικό φιλτράρισμα των ήχων.
- Υποστηρίζει μήκος μπλοκ των 1.024 ή και 960 δειγμάτων, σε σχέση με την υποστήριξη των τμημάτων με τα 576 δείγματα του MP3. Το γεγονός αυτός υποστηρίζει καλύτερη απόδοση στην περίπτωση ηχητικών σημάτων χωρίς πολλές αλλαγές.
- Βελτιωμένη απόδοση όσον αφορά τη συμπίεση.

Πρότυπα κωδικοποίησης ήχου MPEG

Το MPEG-4 Part 3 περιλαμβάνει τους προηγούμενους κωδικοποιητές MPEG-2 AAC και τους νέους MPEG-4 κωδικοποιητές ήχου, τον MPEG-4 HE (High Efficiency) AAC.

Σε ένα τυπικό σύστημα ψηφιακής τηλεόρασης, η επιλογή της κωδικοποίησης του ήχου δεν είναι αυστηρά προδιαγεγραμμένη και μάλιστα σε πολλές περιπτώσεις διαφοροποιείται. Οι περισσότεροι Audioplayers υποστηρίζουν MPEG-1 Layer III Audio, MPEG-2 Layer III Audio και AAC MPEG-2 και MPEG-4.



Εξέλιξη προτύπων κωδικοποίησης ήχου

Επαναληπτικές ασκήσεις – Κριτήρια αξιολόγησης

Κριτήριο αξιολόγησης 1

Άσκηση

Υπολογίστε τον ρυθμό στερεοφωνικού ηχητικού σήματος στις περιπτώσεις που η δειγματοληψία (16 bits) γίνεται με συχνότητες:

- $f_{s1} = 48 \text{ KHz}$
- $f_{s2} = 96 \text{ KHz}$

Επαναλάβετε για την περίπτωση ηχητικού σήματος 5:1, το οποίο δηλαδή αποτελείται από 5+1 κανάλια, τα ηχητικά σήματα των οποίων θεωρούμε ότι τα διαχειριζόμαστε με ομοιόμορφο τρόπο (αναφορικά με τη δειγματοληψία και την κβάντιση). Τι παρατηρείτε???

Κριτήριο αξιολόγησης 1

Απάντηση / Λύση

Η σχέση υπολογισμού του ρυθμού είναι η επόμενη:

$$C = n_{ch} * f_s * n_q$$

Όπου

- n_{ch} το πλήθος των καναλιών,
- f_s η συχνότητα δειγματοληψίας,
- n_q το πλήθος των bits που χρησιμοποιούνται για τη δειγματοληψία.

Με χρήση της σχέσης αυτής της σχέσης, έχουμε τις επόμενες περιπτώσεις:

- Για το στερεοφωνικό σήμα

$$f_{s1}=48 \text{ KHz}, C = 2 * 48 * 10^3 * 16 = \mathbf{1.536 \text{ Kbps}}$$

$$f_{s2} = 96 \text{ KHz}, C = 2 * 96 * 10^3 * 16 = \mathbf{3.072 \text{ Kbps}}$$

- Για το σήμα 5:1

$$f_{s1}=48 \text{ KHz}, C = 6 * 48 * 10^3 * 16 = \mathbf{4.608 \text{ Kbps}}$$

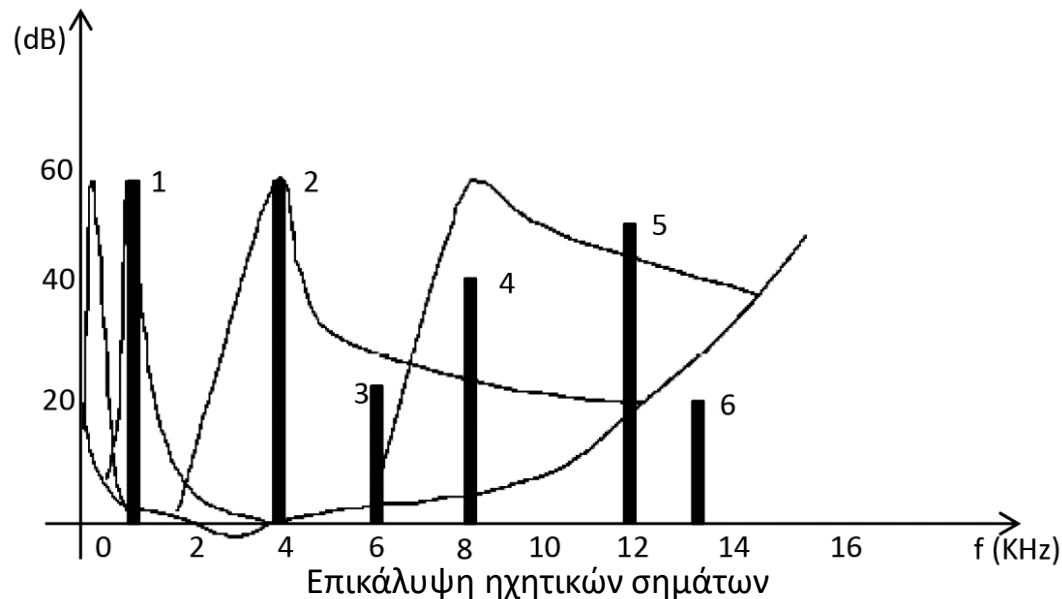
$$f_{s2} = 96 \text{ KHz}, C = 6 * 96 * 10^3 * 16 = \mathbf{9.216 \text{ Kbps}}$$

Παρατηρούμε ότι η αύξηση του ρυθμού δειγματοληψίας αλλά και του πλήθους των καναλιών οδηγεί σε σημαντική αύξηση και του ρυθμού του σήματος.

Κριτήριο αξιολόγησης 2

Άσκηση

Έστω ότι έχουμε ηχητικό σήμα το οποίο πρόκειται να κωδικοποιηθεί σύμφωνα με το πρότυπο του MPEG1 Layer II. Το φασματικό περιεχόμενο του ηχητικού σήματος κάποια χρονική στιγμή (σε κάποιο δηλαδή χρονικό διάστημα μεταξύ δύο διαδοχικών στιγμών δειγματοληψίας) αποτυπώνεται στο σχήμα. Επεξηγήστε ποιο τμήμα του φασματικού περιεχομένου θα κωδικοποιηθεί και ποιο πρέπει να παραλειφθεί. Σε κάποιες περιπτώσεις θα χρειαστεί να κάνετε κάποιες εύλογες εκτιμήσεις αναφορικά με τις πιθανές επικαλύψεις.

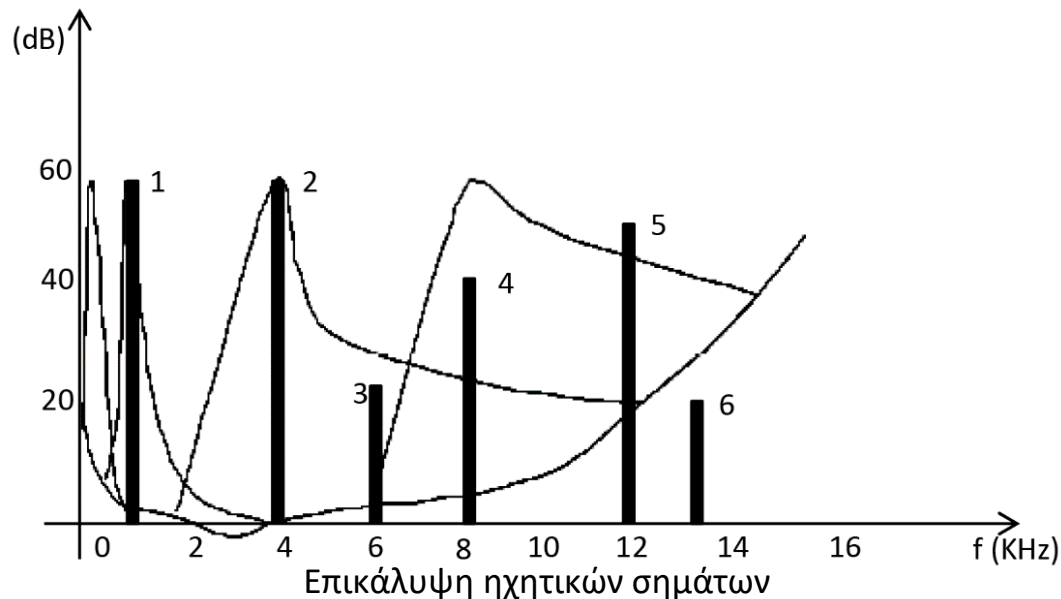


Κριτήριο αξιολόγησης 2

Απάντηση / Λύση

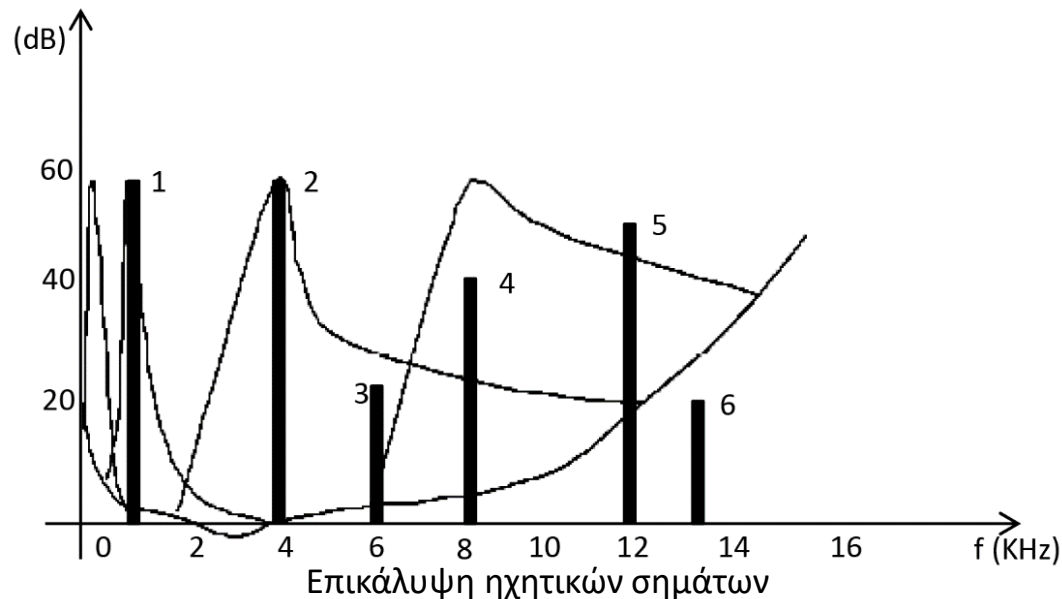
Με βάση το σχήμα, μπορούμε να κάνουμε τις επόμενες παρατηρήσεις, για τα επιμέρους ηχητικά σήματα:

1. Το ηχητικό σήμα 1, περίπου στο 1 KHz, δεν υφίσταται επικάλυψη από τον ήχο που είναι στις πιο χαμηλές συχνότητες.
2. Το ηχητικό σήμα 2, στα 4 KHz, ομοίως δεν υφίσταται επικάλυψη.
3. Το ηχητικό σήμα 3, στα 6 KHz επικαλύπτεται εξολοκλήρου από το φασματικό περιεχόμενο του ηχητικού σήματος 2.



Κριτήριο αξιολόγησης 2

4. **Το ηχητικό σήμα 4**, περίπου στα 8.5 KHz εκτιμούμε ότι υφίσταται μερική μόνο επικάλυψη από τον προηγούμενο ήχο (το ηχητικό σήμα 3). Αυτό σημαίνει ότι θα πρέπει να γίνει κωδικοποίηση με βάση τις διαφορές του συνολικού σήματος, χωρίς το επικαλυπτόμενο τμήμα του.
5. **Το ηχητικό σήμα 5**, στα 12 KHz, υφίσταται μερική επικάλυψη από το φάσμα του προηγούμενου σήματος (ηχητικό σήμα 4), οπότε ομοίως θα γίνει κωδικοποίηση με βάση τις διαφορές του συνολικού σήματος χωρίς το επικαλυπτόμενο τμήμα του.
6. **Το ηχητικό σήμα 6**, στα 13.5 KHz, επικαλύπτεται εξολοκλήρου από το προηγούμενο σήμα (το ηχητικό σήμα 5).



Βιβλιογραφία και βασικές πηγές

Βιβλιογραφία και βασικές πηγές

Προτεινόμενη βιβλιογραφία

1. Παπαδάκης, Α., 2015. Ψηφιακή τηλεόραση. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/5005>
2. Κ. Τσαμούταλος και Π. Σαράντης. Αναλογική και Ψηφιακή Τηλεόραση. Σταμούλης, 2003.
3. Π. Βαφειάδης. Αναλογική-ψηφιακή τηλεόραση και video. 7η έκδοση. Π. Βαφειάδης, 2014.
4. International Standards Organization / International Electrotechnical Commission. Information technology — Generic coding of moving pictures and associated audio information: Audio, International Standards Organization / International Electrotechnical Commission ISO/IEC, 11172-3, 1993.
5. International Standards Organization / International Electrotechnical Commission. Information technology — Generic coding of moving pictures and associated audio information: Audio, International Standards Organization / International Electrotechnical Commission ISO/IEC, 13818-3, 1998.
6. International Standards Organization / International Electrotechnical Commission. Information technology — Generic coding of moving pictures and associated audio information: Advanced Audio Coding, International Standards Organization / International Electrotechnical Commission ISO/IEC, 13818-7, 1998.
7. International Standards Organization / International Electrotechnical Commission. Information technology — Generic coding of moving pictures and associated audio information: MPEG4 Advanced Audio Coding, International Standards Organization / International Electrotechnical Commission ISO/IEC, 14496-3, 2009.
8. Advanced Television Standards Committee (ATSC) Standard. Digital Audio Compression (AC-3, E-AC-3). Advanced Television Standards Committee, 2012.
9. K. Brandenburg. MP3 and AAC Explained, Fraunhofer Institute for Integrated Circuits FhG-IIS A. Erlangen, 1999.

Συναφή επιστημονικά περιοδικά

1. IEEE Communications Magazine, ComSoc
2. IEEE Transactions on Broadcasting
3. International Journal of Digital Television, Intellect