

ΤΕΙ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΠΟΛΥΜΕΣΩΝ

σημειώσεις για το μάθημα

ΕΠΕΞΕΡΓΑΣΙΑ ΗΧΟΥ ΦΩΝΗΣ

Στεφανάκης Νικόλαος
Φουντουλάκης Αντώνιος

Περιεχόμενα

1. Ήχος - η Βασική Θεωρία	4
1.1 Κύματα γενικά	4
1.2 Ηχητικά Κύματα	6
1.3 Ήχος και άνθρωπος	11
1.4 Ήχος και θόρυβος.....	16
1.5 Διάδοση του ήχου, ανάκλαση, αντήχηση	16
2. ΘΕΩΡΙΑ ΨΗΦΙΑΚΟΥ ΗΧΟΥ	19
2.1 Εισαγωγή	19
2.2 Βασικές Αρχές Λειτουργίας	21
2.3 Αριθμητική αναπαράσταση σταθερού σημείου και βήμα κβαντισμού	25
2.4 Δυναμική Περιοχή Ηχητικού Συστήματος	27
2.5 Dither	34
2.6 Υπερδειγματοληψία.....	40
3. ΨΗΦΙΑΚΟ ΣΤΟΥΝΤΙΟ	44
3.1 Γενικές Δομές και Κατηγορίες Συστημάτων	44
3.2 Γενικές Απαιτήσεις από τις Συσκευές	48
3.3 Υλοποίηση Μεθόδων - Συστημάτων σε Υλικό	49
3.4 Υλοποίηση Μεθόδων και Συστημάτων σε Λογισμικό.....	51
4. ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΗΧΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ	55
4.1 Η ψηφιακή επεξεργασία σήματος σε ηχητικές εφαρμογές	55
4.2 Δομικά στοιχεία του επεξεργαστή	56
4.3 Μίξη ψηφιακού σήματος	58
4.4 Τροποποίηση ακουστικού ειδώλου με διαφοροποίηση της έντασης....	58

4.5 Διάδοση του ήχου	59
4.6 Τροποποίηση ακουστικού ειδώλου με χρήση και στοιχείων καθυστέρησης.....	62
4.7 Ανάκλαση του Ήχου	63
4.8 Προσθήκη Αντήχησης	67
4.9 Ισοστάθμιση και Αφαίρεση Αντήχησης	74
4.9.1.1 Συγχρονισμός Ηχείων	76
4.9.1.2 Παραμόρφωση φίλτρων τύπου κτένας.....	76
4.9.1.3 Υπολογισμός Συχνοτήτων του Comb Filter.....	78
5. EQUALISERS.....	80
5.1 Τυπικό equalizer 3 περιοχών τύπου κονσόλας	80
5.2 Το EQ ως ισοσταθμιστής.....	81
5.3 Κύκλωμα crossover στα ηχεία	83
5.4 Φίλτρα A, B, C	84
5.5 Ψηφιακά φίλτρα.....	85
6. Δυναμική Επεξεργασία Σήματος.....	88
6.1 Συμπιεστής (Compressor).....	88
6.2 Expander.....	94
6.3 Multi-band compressor	95
6.4 Noise Gates.....	96

1. Ήχος - η Βασική Θεωρία

1.1 Κύματα γενικά

Ο ήχος από πλευράς φυσικής είναι ένα μηχανικό κύμα. Η διάδοση δηλαδή μιας διαταραχής στον αέρα ή άλλα ελαστικά μέσα η οποία τελικά μπορεί να ερεθίσει το ανθρώπινο αυτί.

Ας αρχίσουμε να μιλάμε πιο συγκεκριμένα και να θέσουμε τις βασικές αρχές για να μπορέσουμε να εξετάσουμε τα βασικά χαρακτηριστικά του ήχου και να μπορέσουμε να τον επεξεργαστούμε. Όπως προαναφέρθηκε ο ήχος είναι μηχανικό κύμα.

Κύμα είναι μια διαταραχή στην κατάσταση ισορροπίας η οποία διαδίδεται από μια περιοχή του χώρου σε μια άλλη. Αν αυτή η διαταραχή χρειάζεται κάποιο φυσικό μέσο για να μεταδοθεί τότε μιλάμε για **μηχανικό κύμα**. Αν δεν χρειάζεται κάποιο φυσικό μέσο για να μεταδοθεί, αλλά πρόκειται για διαταραχή σε κάποιο πεδίο, βασικά την ένταση του ηλεκτρικού ή μαγνητικού πεδίου, μιλάμε για τα **Ηλεκτρομαγνητικά κύματα**, ή άλλες ποιό εξωτικές μορφές όπως τα βαρυτικά κύματα. Αν η διαταραχή από την κατάσταση ισορροπίας είναι παράλληλη με την διεύθυνση μετάδοσης έχουμε τα **διαμήκη κύματα** ενώ αν είναι κάθετη τα **εγκάρσια κύματα**.

Αν η διαταραχή αυτή είναι περιοδική έχουμε τα **περιοδικά κύματα** ενώ ο χρόνος στον οποίο έχουμε μια πλήρη επανάληψη της διαταραχής ονομάζεται **περίοδος** του κύματος, ενώ το αντίστροφο του (πόσες επαναλήψεις πραγματοποιούνται στην μονάδα του χρόνου) **συχνότητα** του κύματος. Τα κύματα περιγράφονται μαθηματικά σαν μια συνάρτηση δύο μεταβλητών, δηλαδή το μέγεθος της διαταραχής σαν συνάρτηση της θέσης και του χρόνου.

$$\Psi = f(\vec{r}, t), \quad (1.1)$$

και είναι λύσεις της διαφορικής εξίσωσης

$$\frac{\partial^2}{\partial t^2} \Psi = c^2 \nabla^2 \Psi \quad (1.2)$$

Στην ειδική περίπτωση που η διαταραχή είναι μια αρμονική ταλάντωση τότε έχουμε τα αρμονικά-ημιτονοειδή - κύματα που για την μονοδιάστατη περίπτωση (το κύμα μεταδίδεται πάνω σε μία γραμμή) είναι

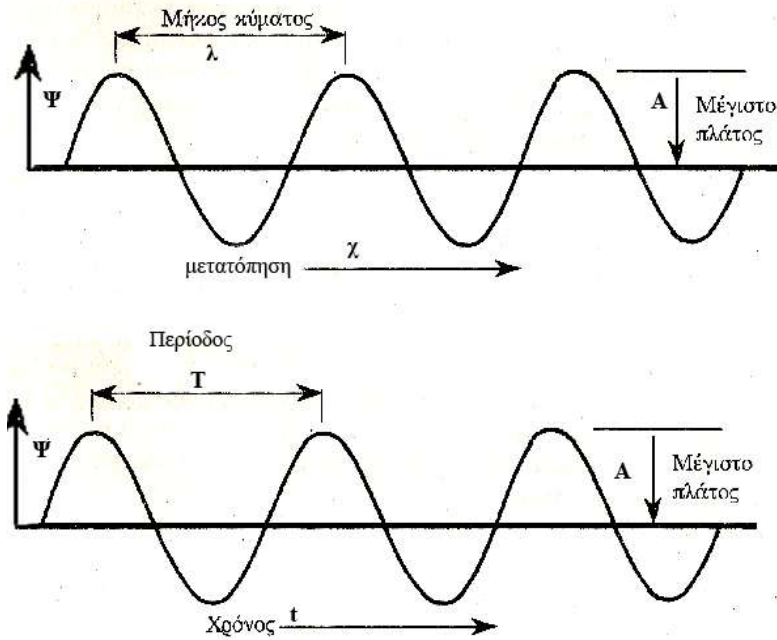
$$\Psi(x, t) = A \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right), \quad (1.3)$$

όπου A το **πλάτος του κύματος**, μια ποσότητα που δείχνει πόσο ισχυρό είναι ένα κύμα, T η περίοδος του κύματος και λ το **μήκος κύματος** του κύματος, η απόσταση που διανύει το κύμα σε χρόνο μιας περιόδου. Για ημιτονοειδή κύματα η ποσότητα η οποία βρίσκεται στο ημίτονο (στην σχέση (1.3) το $2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right)$) ονομάζεται **φάση του κύματος** για να βρίσκονται

δύο ημιτονοειδή κύματα σε φάση θα πρέπει η διαφορά φάσης τους (η διαφορά των ποσοτήτων που βρίσκονται μέσα στο ημίτονο) να είναι ακέραιο πολλαπλάσιο του 2π . Για να περιγράψουμε πόσο γρήγορα μεταφέρεται ένα κύμα εισάγουμε τον όρο της ταχύτητας του κύματος που είναι

$$v = \frac{\lambda}{T} = \frac{\frac{2\pi}{k}}{\frac{2\pi}{\omega}} = \frac{\omega}{k}, \quad (1.4)$$

όπου ω η κυκλική συχνότητα του κύματος και k ο κυματαριθμός ή στην τρισδιάστατη περίπτωση το κυμαάνυσμα του κύματος.

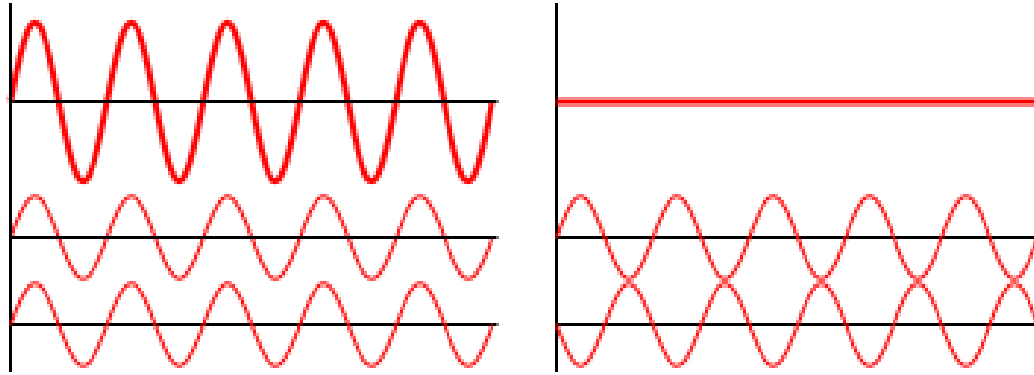


Εικόνα 1 Γραφική αναπαράσταση ενός κύματος για την κατανόηση μήκους κύματος και περιόδου.

Μια βασική ιδιότητα των κυμάτων είναι η αρχή της επαλληλίας δηλαδή όταν 2 ή περισσότερα κύματα συνευρίσκονται, η διαταραχή που προκαλείται σ' ένα σημείο του χώρου μια συγκεκριμένη χρονική στιγμή, είναι ίση με το "άθροισμα" των διαταραχών που θα προκαλούσε το κάθε ένα αν βρισκόταν στον συγκεκριμένο χώρο την συγκεκριμένη χρονική στιγμή δηλ

$$\Psi_{all}(x,t) = \Psi_1(x,t) + \Psi_2(x,t). \quad (1.5)$$

Μια ενδιαφέρουσα περίπτωση εφαρμογής της αρχής της επαλληλίας είναι όταν στο ίδιο χώρο μεταδίδονται σύμφωνα κύματα (με την ίδια συχνότητα) και την ίδια ένταση τότε σ' ένα σημείο του χώρου μπορούμε να έχουμε διπλάσια ένταση (πλάτος), ενισχυτική συμβολή, ή μηδενική ένταση, καταστρεπτική συμβολή. Αποδεικνύεται ότι το γεγονός αυτό εξαρτάται από την διαφορά φάσης των δύο κυμάτων που στην πρώτη περίπτωση είναι ακέραιο πολλαπλάσιο του 2π ενώ στην δεύτερη περιπτώ πολλαπλάσιο του π . Αν τα κύματα αυτά προέρχονταν από όμοιες πηγές θα λέγαμε ότι η διαφορά δρόμου των δύο κυμάτων είναι ακέραιο πολλαπλάσιου του μήκους κύματος στην πρώτη περίπτωση και περιττό πολλαπλάσιο του μισού μήκους κύματος στην δεύτερη.



Εικόνα 2 *Αριστερά* ενισχυτική συμβολή τα δύο κύματα (κάτω) είναι σε φάση. *Δεξιά* καταστρεπτική συμβολή τα δύο κύματα έχουν διαφορά φάσης π .

Μια άλλη ιδιότητα των κυμάτων εμφανίζεται όταν ένα κύμα το οποίο διαδίδεται σε ένα μέσο διάδοσης προσπέσει σε διαφορετικό μέσο διάδοσης. Στην περίπτωση αυτή, γενικά, ένα μέρος του κύματος θα περάσει (διάδοσης - διάθλαση), η ενδεχομένως να απορροφηθεί, ενώ ένα άλλο μέρος του θα αλλάξει κατεύθυνση και θα επιστρέψει στο αρχικό μέσο διάδοσης (ανάκλαση).

Μια εφαρμογή των παραπάνω είναι η δημιουργία στάσιμων κυμάτων. Μια τέτοια περίπτωση έχουμε όταν δημιουργείται ένα κύμα σε ένα χώρο του οποίου τα τοιχώματα δεν επιτρέπουν να τα διαπεράσει το κύμα αλλά αυτό ανακλάται και συμβάλει με το προσπίπτον, οπότε σχηματίζονται τα στάσιμα κύματα.

Η εξίσωση που πρέπει να λύσουμε σε μονοδιάστατη περίπτωση γραμμικού μήκους L είναι $\frac{\partial^2 \Psi(x)}{\partial x^2} + k^2 \Psi(x) = 0$ με γενική λύση $\Psi(x) = A \cos kx + B \sin kx$ με A και B σταθερές, αν

λάβουμε υπόψη τις συνοριακές συνθήκες αποδεικνύεται ότι $k = \frac{n\pi}{L} \Leftrightarrow \frac{L}{n} = \frac{\lambda}{2}$.

1.2 Ηχητικά Κύματα

Σαν ήχο χαρακτηρίζουμε τα μηχανικά κύματα που διαδίδονται στον αέρα ή σε άλλα ελαστικά μέσα και μπορούν να διεγείρουν το αισθητήριο όργανο της ακοής. Για να δώσουμε μια γενική περιγραφή ηχητικά κύματα είναι κύματα στην πίεση του ελαστικού μέσου. Με λίγα λόγια ο ήχος παράγεται από την δημιουργία μιας διαταραχής στη ηχητική πηγή, π.χ. μετατόπιση της μεμβράνης ενός τυμπάνου ή μιας χορδής ενός εγχόρδου οργάνου ή των φωνητικών χορδών κ.α, η διαταραχή αυτή μετατοπίζει τα μόρια του αέρα που γειτνιάζουν με την πηγή, και έτσι η διαταραχή μεταδίδεται στον αέρα ή κάποιο ελαστικό μέσο και ερεθίζει το ανθρώπινο αυτί. Για να μπορέσει ένα τέτοιο ηχητικό κύμα να ερεθίσει το ανθρώπινο αυτί θα πρέπει να έχει συχνότητα από 20 Hz μέχρι 20 KHz (20000 Hz) περίπου. Ας υποθέσουμε ότι έχουμε μονοδιάστατη περίπτωση και η αρχική διαταραχή είναι αρμονική, η μετατόπιση κάθε μορίου του αέρα στην θέση x και την χρονική στιγμή t θα είναι της μορφής

$$\Psi = A \sin(\omega t - kx), \quad (1.6)$$

όπου k ο κυματάριθμος, ω η συχνότητα του κύματος και A η μέγιστη απομάκρυνση από την

θέση ισορροπίας, αποδεικνύεται ότι οι διακυμάνσεις πίεσης λόγω του ηχητικού κύματος είναι τις μορφής

$$p = BkA \cos(\omega t - kx), \quad (1.7)$$

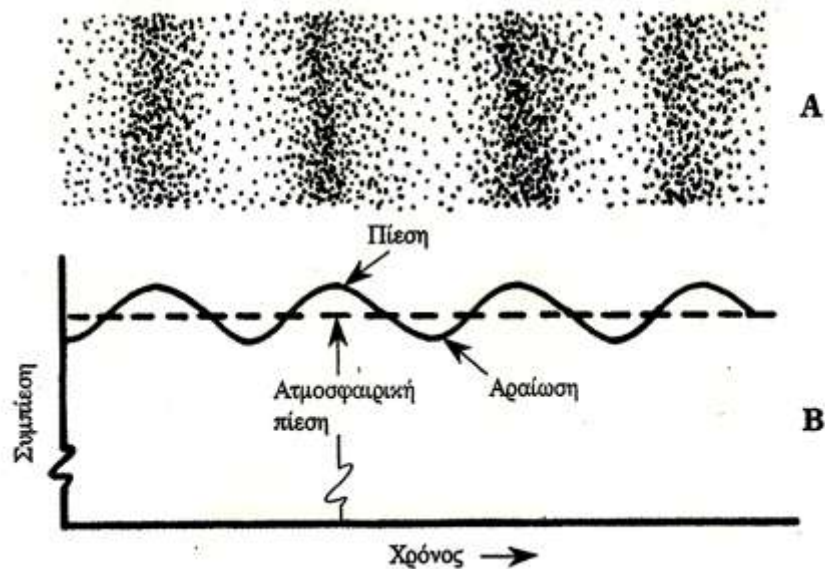
με B το μέτρο ελαστικότητας όγκου του υλικού άρα οι διακυμάνσεις τις πίεση στο κάθε σημείο είναι και αυτές αρμονικές.

Η ταχύτητα του ήχου εξαρτάται από το υλικό στο οποίο διαδίδεται στην γενική της περίπτωση η ταχύτητα του ήχου δίνεται από την σχέση

$$c = \sqrt{\frac{B}{\rho}}, \quad (1.8)$$

με ρ την πυκνότητα του υλικού και B το μέτρο ελαστικότητας όγκου του υλικού ($B = \frac{dP}{dV/V}$).

Ειδικά στον αέρα γίνεται $c = \sqrt{\frac{\gamma RT}{MB}}$ με R σταθερά των ιδανικών αερίων γ ο λόγος των ειδικών θερμοτήτων C_p/C_v T η θερμοκρασία και MB το μοριακό βάρος του αερίου.



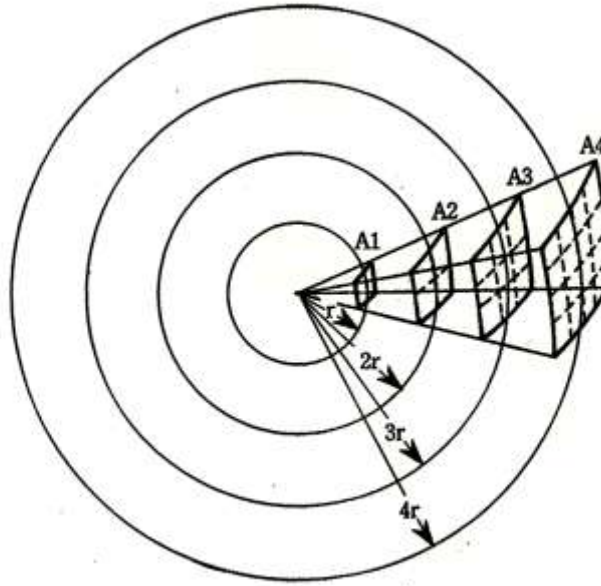
Εικόνα 3 Γραφική αναπαράσταση ενός ηχητικού κύματος.

Αν έχουμε μια σημειακή ηχητική πηγή σε ανοικτό χώρο τότε τα ηχητικά κύματα θα μεταδίδονται ισοτροπικά (αφού δεν υπάρχει κάποια ιδιαίτερη προτίμηση για ακολουθήσουν κάποια συγκεκριμένη διεύθυνση) δηλαδή η ένταση του ήχου σ' όλες τις διευθύνσεις για την ίδια απόσταση από την πηγή θα είναι η ίδια. Το πόσο δυνατός είναι ένας ήχος εξαρτάται από την ένταση του ηχητικού κύματος στο σημείο αυτό. Όταν μιλάμε για ένταση ενός κύματος σε κάποιο σημείο μιλάμε για την ενέργεια που μεταφέρει το κύμα ανά μονάδα χρόνου ανά μονάδα επιφάνειας ($J/s/m^2 = W/m^2$ στο SI). Έχοντας υπόψη ότι η ενέργεια ισοκατανέμεται σε όλη την επιφάνεια που περιλαμβάνει το μέτωπο κύματος (τα σημεία του χώρου με την ίδια ένταση του κύματος) αν σε σημείο με απόσταση r_1 από την ηχητική πηγή η ένταση του ηχητικού κύματος είναι I_1 τότε σε απόσταση r_2 θα είναι I_2 και αφού η ενέργεια

δεν χάνετε πουθενά θα ισχύει

$$P_{all} = I_1 \cdot 4\pi r_1^2 = I_2 \cdot 4\pi r_2^2 \Leftrightarrow \frac{I_1}{I_2} = \left(\frac{r_2}{r_1}\right)^2, \quad (1.9)$$

βλέπουμε λοιπόν ότι στον ελεύθερο χώρο ισχύει ο κανόνας του αντιστρόφου τετραγώνου η ένταση δηλαδή του ήχου μειώνεται με το τετράγωνο της απόστασης.



Εικόνα 4 Σχηματική αναπαράσταση σημείων με την ίδια ένταση του ήχου, ο κάθε ομόκεντρος κύκλος είναι ένα μέτωπο κύματος.

Αποδεικνύεται ότι η μέση ισχύ σ' ένα αρμονικό ηχητικό κύμα είναι τις μορφής

$$I_{RMS} = \frac{1}{2} \omega B k A^2 = \frac{1}{2} \sqrt{\rho B} \omega^2 A^2 = \frac{p_{max}^2}{2\sqrt{\rho B}}, \quad (1.10)$$

όπου ρ η πυκνότητα του μέσου μετάδοσης και p_{max} το πλάτος πίεσης για το ηχητικό κύμα. Διευκρινιστικά η μέση ισχύς που αναφέρουμε παραπάνω είναι η ενέργεια ανά μονάδα επιφάνειας που μεταφέρετε από το κύμα σε χρόνο μίας περιόδου, όταν μιλάμε για ένταση ενός ήχου από εδώ και στο εξής θα μιλάμε για την ένταση αυτή.

Συχνά όταν μιλάμε για το πόσο ισχυρός είναι ένα ήχος ακούμε σαν μονάδα μέτρησης τα decibel (dB). Ας συνδέσουμε την μονάδα αυτή με τις γνωστές μας μονάδες ισχύος και ενέργειας. Τα διάφορα ανθρώπινα αισθητήρια όργανα όπως, αυτιά και μάτια, μπορούν να αντιληφτούν ήχο και φως σε μεγάλο εύρος τιμών, αρκετές τάξεις μεγέθους. Για το λόγο αυτό η αντίληψη του ήχου και φωτός γίνεται λογαριθμικά και όχι γραμμικά. Γι αυτό εισάγουμε ένα άλλο τρόπο προσδιορισμού της έντασης του φωτός και του ήχου αλλά και γενικότερα ποσοτήτων που έχουν τόσο μεγάλο εύρος τιμών το dB. Μια στάθμη σε decibel είναι δέκα φορές ο δεκαδικός λογάριθμος δύο ποσοτήτων. Θα λέγαμε λοιπόν ότι η στάθμη μιας ισχύος W_1 μπορεί να γραφεί σε σχέση με μια ισχύ αναφοράς W_0 ως εξής

$$L_1 = 10 \log_{10} \frac{W_1}{W_0} \quad (1.11)$$

Η χρήση του decibel δεν περιορίζεται μόνο στην ακουστική αλλά έχει μεγάλη εφαρμογή και σε άλλους τομείς της σύγχρονης επιστήμης και ιδιαίτερα στις τηλεπικοινωνίες.

Στην περίπτωση του ήχου μπόουμε να έχουμε μέτρηση σε dB τόσο στάθμης ακουστικής έντασης όσο και ηλεκτρικής, όμως μπορούμε να εκφράσουμε σε dB και στάθμες πίεσης αν αντικαταστήσουμε την (1.10) στην (1.11) θα έχουμε

$$L_1 = 10 \log_{10} \frac{\frac{p_1^2}{2\sqrt{\rho B}}}{\frac{p_2^2}{2\sqrt{\rho B}}} = 10 \log_{10} \left(\frac{p_1}{p_2} \right)^2 = 20 \log_{10} \frac{p_1}{p_2} \quad (1.12)$$

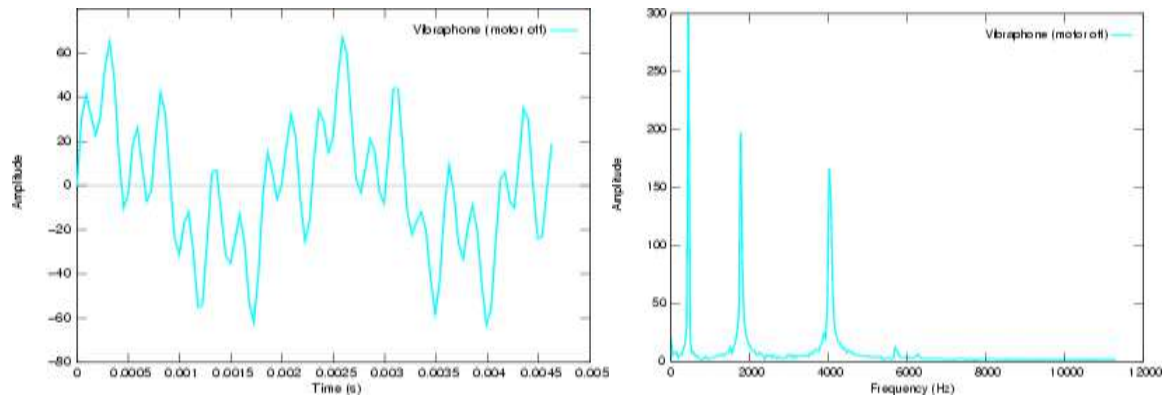
Για τη στάθμη πίεσης ήχου SPL (Sound Pressure Level) πίεση αναφοράς είναι τα 20 $\mu\text{Pa} = 2 \cdot 10^{-5} \text{ N/m}^2$ ενώ για τη στάθμη έντασης ήχου τα 10^{-12} W/m^2 . Ένα μέσο ανθρώπινο αυτί μπορεί να ακούσει ήχους θεωρητικά από 0 dB ενώ τα 120 dB μπορεί να ανεχθούν για μερικά λεπτά ενώ υπάρχει κώφωση για επίπεδα πάνω από 140 με 150 dB στον παρακάτω πίνακα βλέπουμε κάποιες ηχητικές πηγές και πόσα dB είναι η κάθε μια.



Εικόνα 5 Σύγκριση ηχητικής πηγής και στάθμης πίεσης για μερικούς ήχους

Μέχρι τώρα μιλούσαμε για απλά ημιτονοειδή κύματα αν όμως ακούγαμε ένα τέτοιο ήχο θα διαπιστώναμε ότι είναι ένας μόνο τόνος (ένα μονότονο beeeeeer) το οποίο προφανώς δεν μεταφέρει καμία πληροφορία. Για να έχουμε μεταφορά πληροφορίας μέσω ήχων και παραγωγή αυτού που λέμε μουσική θα πρέπει οι κυματομορφές να είναι ποιο σύνθετες. Όσο

πολύπλοκο και να είναι ένα κύμα μπορεί να αναλυθεί σε ημιτονοειδής συνιστώσες και ανάποδα άθροισμα πολλών ημιτόνων ή συνημίτονων μπορεί να δώσει μια ποιο σύνθετη κυματομορφή, άρα ότι είπαμε παραπάνω για τα ημιτονοειδή κύματα ισχύει και για τον γραμμικό συνδυασμό τους τις ποιο σύνθετες κυματομορφές. Φάσμα τώρα μιας κυματομορφής μπορούμε να πούμε ότι είναι οι συχνότητες οι οποίες συμμετέχουν στο σχηματισμό μιας μορφής καθώς και με τι σχετικό βάρος. Έτσι αυτό που έχουμε στο μυαλό μας όταν λέμε φασματική απεικόνιση για ένα κύμα είναι μια γραφική παράσταση των συχνοτήτων σε σχέση με την "συμμετοχή" τους στην κυματομορφή (σχετικό πλάτος της συγκεκριμένης συχνότητας).



Εικόνα 6 Μια κυματομορφή από vibraphone (αριστερά) και το φάσμα της δεξιά

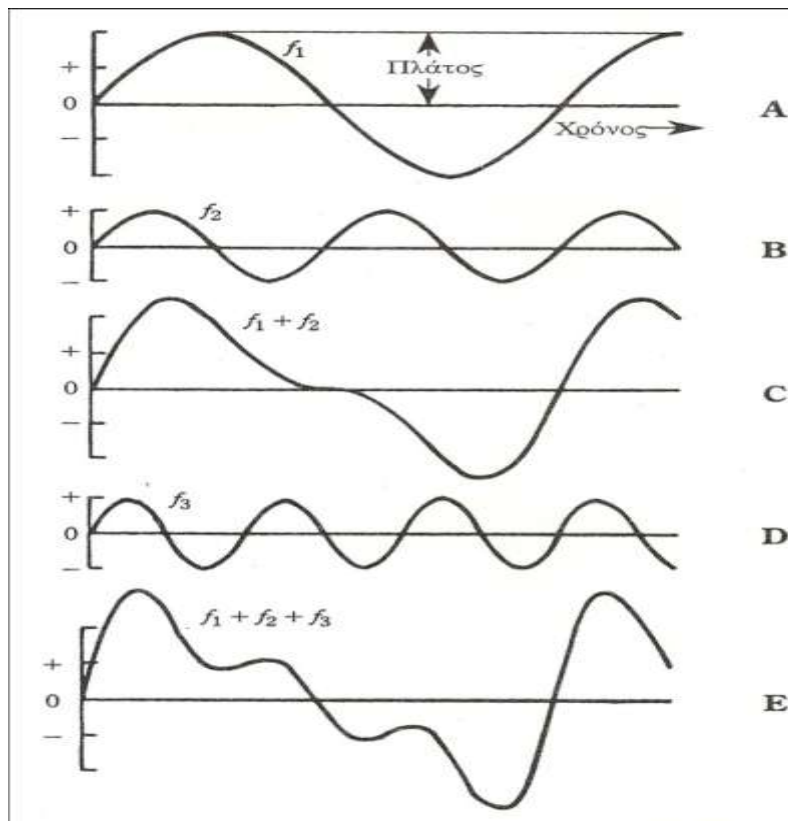
Αν έχουμε τον συνδυασμό κυμάτων όπου οι συχνότητες που συμμετέχουν είναι ακέραιο πολλαπλάσιο μιας θεμελιώδους συχνότητας (σειρές Fourier¹) τότε οι τελευταίες ονομάζονται αρμονικές της θεμελιώδους συχνότητας. Μια τέτοια κυματομορφή θα γράφετε σαν άθροισμα ημιτόνων με συχνότητες οι οποίες είναι ακέραιο πολλαπλάσιο μιας θεμελιώδους της f_0 θα είναι δηλαδή της μορφής:

$$\Psi = \sum_{i=1}^N A_i \sin 2\pi(i f_0 t - \phi_i) \quad (1.13)$$

¹ Μια συνεχή περιοδική συνάρτηση $f(t)$ Μπορεί να γραφεί $f(t) = \frac{\alpha_0}{2} + \sum_{n=1}^{\infty} \alpha_n \cos n\omega_0 t + b_n \sin n\omega_0 t$ με

$$\alpha_0 = \frac{2}{T} \int_{-T/2}^{T/2} f(t) dt, \quad \alpha_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos n\omega_0 t dt \dots n = 0, 1, 2, \dots$$

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin n\omega_0 t dt \dots n = 0, 1, 2, \dots, \quad A_n = \sqrt{\alpha_n^2 + b_n^2}, \quad \tan \phi_n = \frac{b_n}{\alpha_n}$$



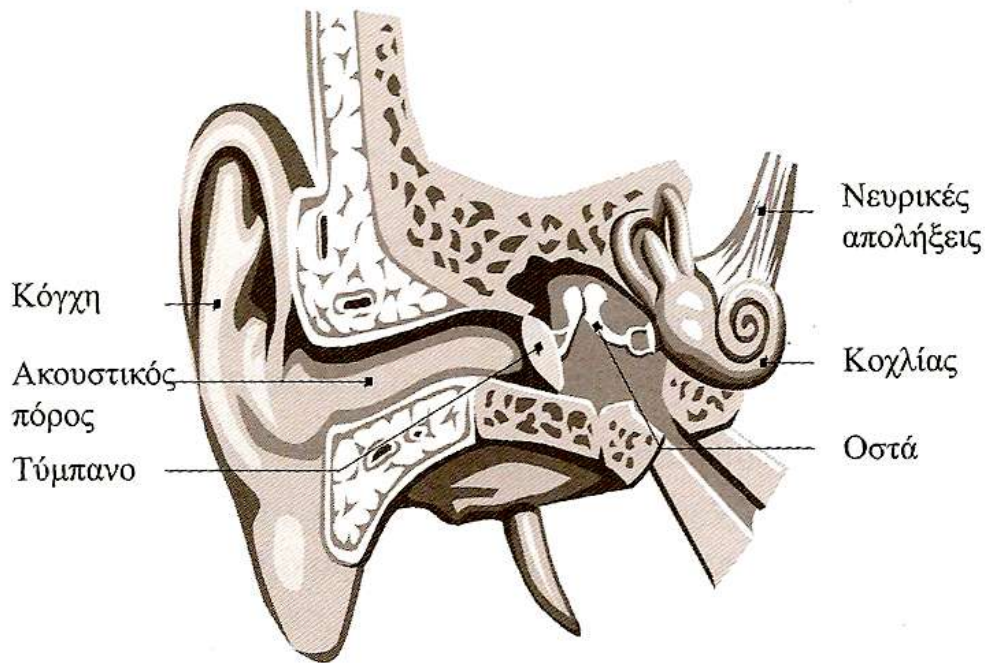
Εικόνα 7 πρόσθεση ημιτονοειδών κυμάτων με $f_2=f_1$, $f_3=3f_1$ τα τρία κύματα είναι σε φάση.

Να συνδέσουμε τα παραπάνω με κάτι σχετικό με την μουσική. Λέμε ότι δύο τόνοι (ημιτονοειδή κύματα) απέχουν μια οκτάβα αν η συχνότητα του ενός είναι διπλάσια από την συχνότητα του άλλου. Για παράδειγμα αν έχουμε μια βασική συχνότητα στα 100 Hz τότε η δεύτερη αρμονική 200 Hz θα είναι μια οκτάβα παραπάνω αλλά αν θέλουμε να ανεβούμε ακόμα μια οκτάβα πρέπει να πάμε στην τέταρτη αρμονική(400Hz) ενώ αν θέλουμε ακόμα μια (3 οκτάβες πάνω από την βασική) θα πρέπει να πάμε στην όγδοη αρμονική (800 Hz).

1.3 Ήχος και άνθρωπος

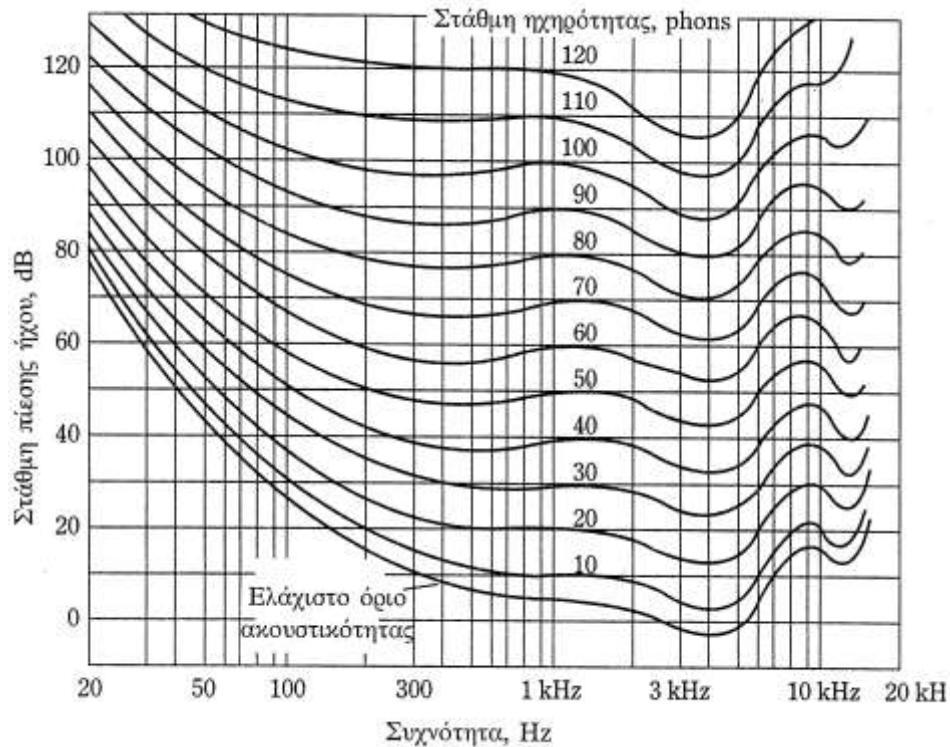
Για να κατανοήσουμε καλύτερα την αντίληψη κάθε ατόμου για τον ήχο ας πούμε λίγα πράγματα για το ανθρώπινο αυτί. Η λειτουργία του είναι παρόμοια με ενός μικροφώνου όμως η καμπύλη απόκρισής του είναι καμπύλη αντί για ευθεία. Αποτελείται από τρία μέρη το εξωτερικό το μέσο και το εσωτερικό αυτί όπως φαίνεται στην **Εικόνα 8**. Το εξωτερικό αυτί συγκεντρώνει το ηχητικό κύμα με την κόγχη και μέσω του ακουστικού πόρου ή λαβυρίνθου (ο οποίος δρα σαν ακουστικός σωλήνας) μεταφέρεται στο μέσο αυτί και προσπίπτει στο τύμπανο, το διεγείρει έτσι ώστε να κάνει ταλαντώσεις, οι ταλαντώσεις αυτές ανιχνεύονται μεταφέρονται μέσω της "σφύρας", του "άκμονα" και του αναβολέα στο εσωτερικό αυτί σε μυϊκούς ιστούς οι οποίοι μειώνουν την ταλάντωση στους δυνατούς ήχους και τελικά η ταλάντωση οδηγείται στον κοχλία όπου υπάρχει η βασική μεμβράνη που φέρει

~ 24000 νευρικές απολήξεις με την μορφή τριχιδίων το καθένα από τα οποία συντονίζεται και σε διαφορετική συχνότητα και το ερέθισμα που είναι μια πιεζοηλεκτρική τάση μεταφέρεται μέσω νευρών στον εγκέφαλο.



Εικόνα 8 Ανθρώπινο αυτί

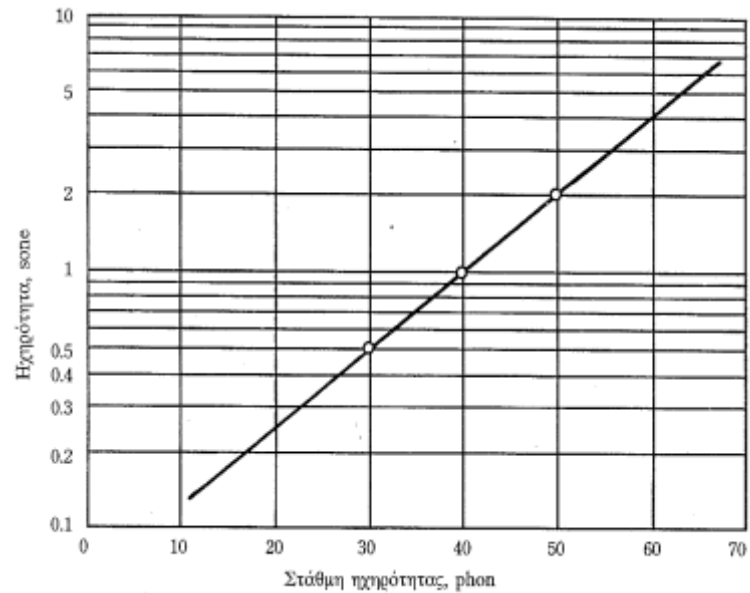
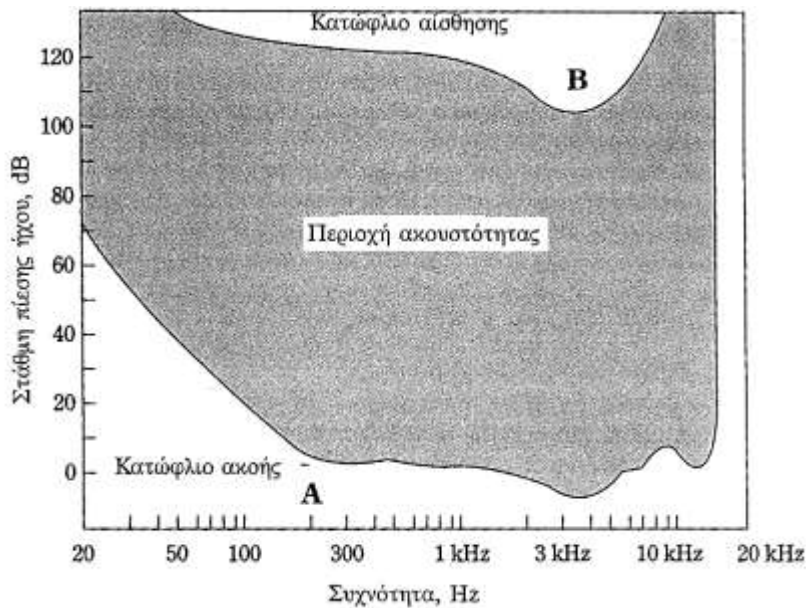
Είναι σημαντικό να δούμε πως αντιλαμβάνεται ο άνθρωπος τον ήχο. Όπως είπαμε το ανθρώπινο αυτί συλλαμβάνει ήχους από 20 Hz μέχρι 20 kHz όμως ήχος ίδιας έντασης δεν γίνεται αντιληπτός από τον άνθρωπο το ίδιο, αν βρίσκονται σε διαφορετικές συχνότητες. Με άλλα λόγια το αυτί δεν έχει την ίδια απόκριση σε όλες τις συχνότητες. Θα μπορούσαμε να πούμε το πόσο δυνατός γίνεται αντιληπτός ένας ήχος ονομάζεται αντιληπτή **ηχηρότητα** ή **ακουστότητα** του ήχου αυτού. Για τον παραπάνω λόγο έχουν κατασκευαστεί τα λεγόμενα διαγράμματα ηχηρότητας, καμπύλες δηλαδή σε διάγραμμα συχνότητας – στάθμης πίεσης στις οποίες έχουμε την ίδια ηχηρότητα, ένα μέσο αυτί αισθάνεται, δηλαδή ότι ο ήχος έχει την ίδια ένταση. Για να έχουμε μια αντικειμενική αίσθηση της ηχηρότητας ενός ήχου ορίζουμε τη στάθμη ηχηρότητας σε phon. Για παράδειγμα, το περίγραμμα ίσης ηχηρότητας που περνά από τη στάθμη πίεσης ήχου 40 dB σε συχνότητα 1000 Hz ονομάζεται περίγραμμα 40 phon. Τα περιγράμματα ίσης-ηχηρότητας phon είναι ένας τρόπος αντιστοιχίσουμε το dB SPL ενός καθαρού τόνου στο αντιληπτό επίπεδο ηχηρότητας phon όπως φαίνεται και στην Εικόνα 9.



Εικόνα 9 Περιγράμματα ίσης ηχηρότητας του ανθρώπινου αυτιού (Robinson & Dadson). Παρατηρείται έλλειψη ευαισθησίας του αυτιού σε χαμηλούς τόνους.

Από τα παραπάνω είναι σχετικά εύκολο να προσδιορίσουμε την περιοχή ακουστότητας του ανθρώπινου αυτιού όπως αυτή φαίνεται στην **Εικόνα 10α**. Το διάγραμμα που παρουσιάζεται είναι στο x άξονα συχνότητα ενώ στον y dB SPL η περιοχή ακουστότητας περιορίζεται από δύο καμπύλες την καμπύλη κατωφλίου (A) και το κατώφλι αίσθησης (B). Παρατηρούμε ότι η μέγιστη ευαισθησία βρίσκεται στα 3 kHz. Κάτω από την καμπύλη (A) δεν γίνεται αντιληπτός ένας ήχος ενώ αν κάποιος ήχος βρίσκεται πάνω από την καμπύλη (B) μπορεί να προκαλέσει από μικρή ενόχληση μέχρι μόνιμο πρόβλημα στο ανθρώπινο αυτί.

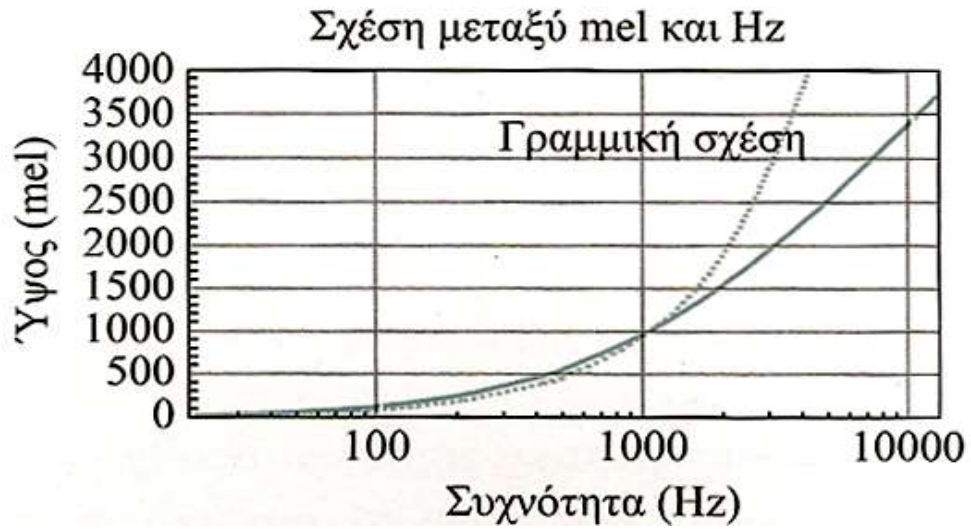
Από πειράματα που πραγματοποιήθηκαν σε εκατοντάδες άτομα συνάγεται το συμπέρασμα ότι αύξηση της SPL κατά 10 dB ο άνθρωπος αισθάνεται ένα ήχο 2 φορές πιο ισχυρό. Για το λόγο αυτό έχει υιοθετηθεί μια μονάδα υποκειμενικής ηχηρότητας το **stone**. **1 stone** είναι η ηχηρότητα που αντιλαμβάνεται ένα άτομο ένα τόνο σταθερής στάθμης 40 phon. Ήχος 2 stone έχει διπλάσια ηχηρότητα ενώ ήχος 0,5 stone την μισή ηχηρότητα στην **Εικόνα 10 β** παρουσιάζετε η σχέση phon-stone. Αξίζει ακόμα να αναφερθεί ότι το πως αντιλαμβανόμαστε ένα σύνθετο ήχο έχει να κάνει και με την διάρκειά του αλλά και με την με το πλάτος του στον χώρο των συχνοτήτων.



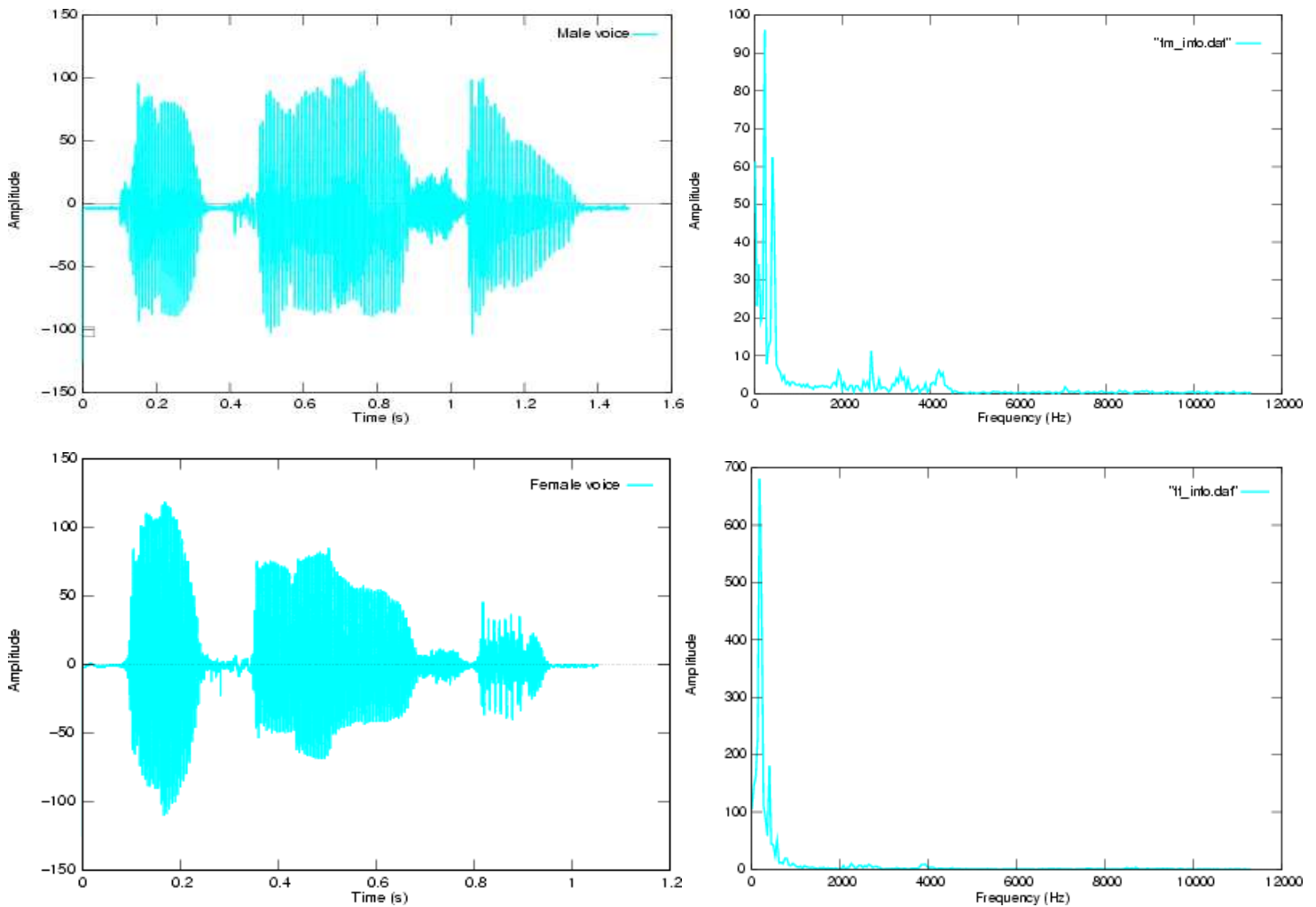
Εικόνα 10 α αριστερά :Η περιοχή ακουστότητας του ανθρώπινου αυτιού **β** δεξιά γραφικά η σχέση φυσικής στάθμης ακουστότητας phon και υποκειμενικής sone.

Δύο άλλα εμπειρικά χαρακτηριστικά του ήχου είναι το **ύψος** και η **χροιά**. Σ' ότι αφορά το **ύψος** έχει κάνει με την συχνότητα του ηχητικού κύματος, δείχνει την υποκειμενική απόκριση του αυτιού στη συχνότητα. Γενικά υψηλοί ήχοι είναι ήχοι μεγάλης συχνότητας ενώ χαμηλοί ήχοι έχουν μικρές συχνότητες, μια υποκειμενική μονάδα μέτρησης του ύψους είναι το mel. Ύψος αναφοράς 1000 mel έχει ορισθεί σαν ύψος ενός τόνου 1000 Hz και SPL 60dB μια εμπειρική καμπύλη συχνότητας ήχου φαίνεται στη **Εικόνα 11**. Παράγοντες που επηρεάζουν τον καθορισμό του ύψους είναι η στάθμη ήχου και η κυματομορφή του.

Χροιά ονομάζουμε τα υποκειμενικά χαρακτηριστικά του ήχου τα οποία κάνουν δυνατό το διαχωρισμό δύο τόνων της ίδιας έντασης και θεμελιώδους συχνότητας αλλά διαφορετικών κυματομορφών. Δηλαδή η χροιά διαφοροποιεί ήχους με διαφορετικό φάσμα. Για παράδειγμα ένας άντρας και μια γυναίκα ενώ θα πουν ακριβώς τα ίδια θα ακουστούν διαφορετικά γιατί το φάσμα των συχνοτήτων που συνθέτουν την ομιλία του καθενός να είναι διαφορετικό. Σχετικό παράδειγμα φαίνεται στην **Εικόνα 12**



Εικόνα 11 Σχέση Ύψους - συχνότητας η συνεχής γραμμή δίνει την σχέση ύψους-συχνότητας ενώ η διακεκομμένη γραμμή την αντίστοιχη γραμμική σχέση.

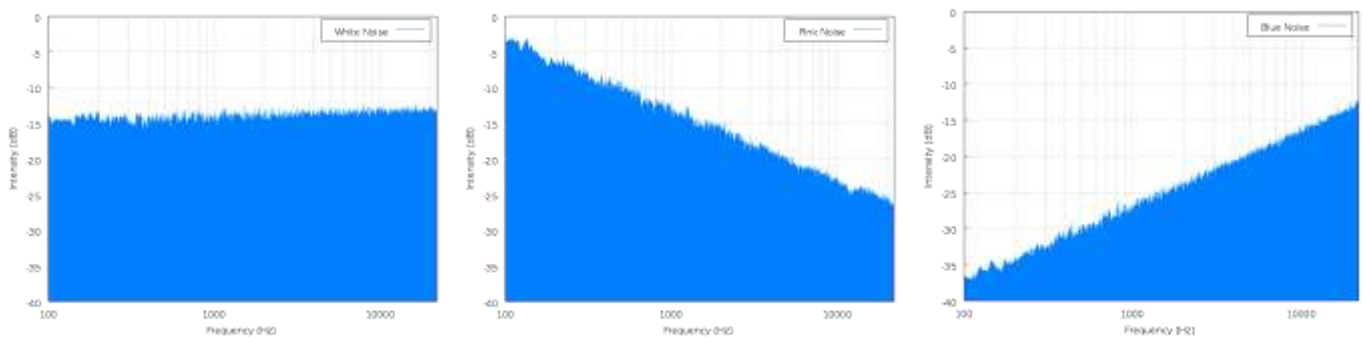


Εικόνα 12 Η κυματομορφή αριστερά και το φάσμα δεξιά μιας αντρικής φωνής πάνω και μιας γυναικείας κάτω.

1.4 Ήχος και θόρυβος

Θα αναφερθούμε εδώ στα φαινόμενα θορύβου. Θα μπορούσαμε απλά να πούμε ότι θόρυβος είναι οποιοσδήποτε δυσάρεστος ή ανεπιθύμητος ήχος, κατ' επέκταση θα λέγαμε γενικότερα ότι είναι μια ανεπιθύμητη διαταραχή σ' ένα επιθυμητό σήμα, και εν γένει είναι ένα σύνθετο απεριοδικό σήμα που η στιγμιαία αυξομειώνεται με τυχαίο τρόπο.

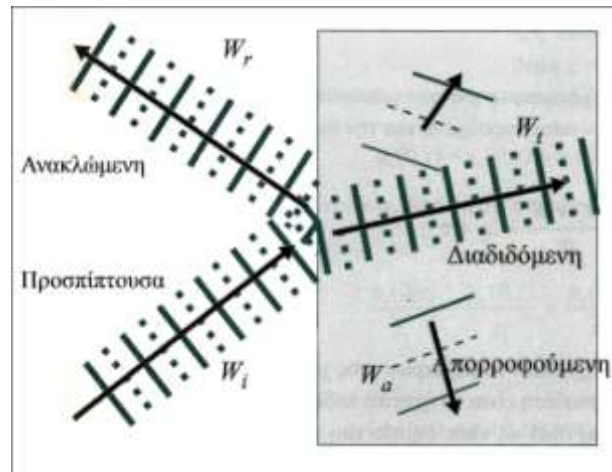
Θα προσπαθήσουμε τώρα να κ θόρυβος είναι οποιοσδήποτε δυσάρεστος ή ανεπιθύμητος ήχος, κατ' επέκταση θα λέγαμε γενικότερα ότι είναι μια ανεπιθύμητη διαταραχή σ' ένα επιθυμητό σήμα, και εν γένει είναι ένα σύνθετο απεριοδικό σήμα που η στιγμιαία αυξομειώνεται με τυχαίο τρόπο. κατηγοριοποιήσουμε τον θόρυβο. Αρχικά ας εξετάσουμε την εξάρτηση του φάσματος του θορύβου από την συχνότητα. Να δούμε δηλαδή τα διάφορα "χρώματα" του θορύβου. Αν για το " power spectral density per unit of bandwidth" θεωρήσουμε ότι είναι ανάλογο του $ps \sim \frac{1}{f^p}$ ανάλογα με την τιμή του p έχουμε και το ανάλογο χρώμα στον θόρυβο. Αν $p=0$ τότε ο θόρυβος έχει την ίδια ένταση σ' όλες τις συχνότητες και τότε έχουμε τον λευκό θόρυβο. Στην περίπτωση που $ps \sim \frac{1}{f}$ έχουμε τον ροζ θόρυβο. Το φάσμα συχνότητας του ροζ θορύβου είναι επίπεδο στο λογαριθμικό διάστημα έχει την ίδια ένταση στις ζώνες που είναι αναλογικά ευρείες. Αυτό σημαίνει ότι ο ροζ θόρυβος θα είχε την ίση ισχύ στο φάσμα συχνότητας από 40 έως 60 Hz όπως στη ζώνη από 4000 έως 6000 Hz. Η πυκνότητα ισχύος, έναντι του άσπρου θορύβου, μειώνεται από 3 DB ανά οκτάβα (πυκνότητα ανάλογη προς $1/f$). Για αυτόν τον λόγο, ο ρόδινος θόρυβος καλείται συχνά « $1/f$ θόρυβος». Στον αντίποδα αν $ps \sim f$ έχουμε τον μπλε θόρυβο. Ακόμα ο καφέ θόρυβος $ps \sim \frac{1}{f^2}$ ο μωβ $ps \sim f^2$.



Εικόνα 13 Από αριστερά προς τα δεξιά το φάσμα λευκού, ροζ και μπλε θορύβου.

1.5 Διάδοση του ήχου, ανάκλαση, αντήχηση

Όταν ένα ηχητικό κύμα πέσει πάνω σε επιφάνεια διαχωρισμού δύο μέσων ένα μέρος από την ενέργεια που μεταφέρει ανακλάται ενώ ένα άλλο απορροφάται ή διαδίδεται στην άλλη πλευρά του μέσου. Για την μελέτη τους ορίσαμε τρεις συντελεστές



Εικόνα 14 Ανάκλαση, απορρόφηση και διάδοση του ήχου

$$\text{Συντελεστή ανάκλασης } a_r = \frac{W_r}{W_i} = \frac{I_r}{I_i}$$

$$\text{Συντελεστής διάδοσης } \tau = \frac{W_t}{W_i} = \frac{I_t}{I_i}$$

$$\text{Συντελεστής απορρόφησης } a_a = \frac{W_a}{W_i} = \frac{I_a}{I_i}$$

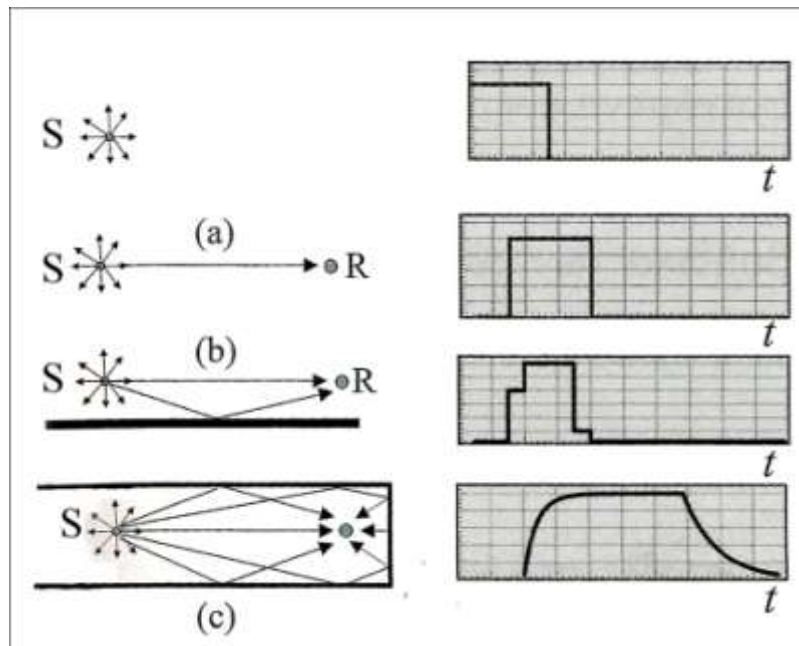
με $a_r + a_a + \tau = 1$ W_a , W_t , W_r η απορροφούμενη, προσπίπτουσα και ανακλώμενη ενέργεια το ίδιο ισχύει και για τις εντάσεις I_x .

Στην περίπτωση ανάκλασης ισχύει γενικά ότι η γωνία πρόσπτωσης του μετώπου κύματος είναι ίση με την γωνία ανάκλασης ενώ για την διάδοση ισχύει ο νόμος του Snell $\frac{\sin \theta_i}{\sin \theta_t} = \frac{c_1}{c_2}$ με θ_i την γωνία πρόσπτωσης και θ_t η γωνία διάθλασης ενώ c_1 η ταχύτητα του ήχου στο πρώτο μέρος και c_2 στο δεύτερο.

Κατά την διέλευση ενός ηχητικού κύματος σε κάποιο μέσο παρατηρούμε φαινόμενα εξασθένισης δηλαδή απώλεια της ενέργειας που μεταφέρει, αυτό συμβαίνει λόγω μεταφοράς ενέργειας σε θερμότητα στο υλικό. Αν προσπέσει ενέργεια I_0 σε υλικό πάχους x τότε η ένταση του ήχου που βγαίνει από αυτό είναι $I_x = I_0 e^{-2ax}$ με a ο συντελεστής απορρόφησης του υλικού στην γλώσσα των επιπέδων ήχου $\Delta L_l = 10 \log e^{-2ax} = 8.7ax = mx$.

Ιδιαίτερο ενδιαφέρον έχει να μελετήσουμε την συμπεριφορά του ήχου σε κλειστούς χώρους. Στην περίπτωση αυτή για να φτάσει ένας ήχος σ' ένα δέκτη έχει υποστεί ανακλάσεις και απορρόφηση τόσο από το μέσο διάδοσης όσο και από τους τοίχους. Παρατηρούμε ότι σε κλειστούς χώρους ο ήχος που ακούμε αλλάζει σε σχέση με τον αρχικά παραχθέντα (Εικόνα 15). Η λειτουργία ενός κλειστού χώρου λειτουργεί σαν ένα κύκλωμα LC. Ο ήχος υφίσταται απορρόφηση (αντίσταση) ενώ λόγω των πολλαπλών ανακλάσεων

έχουμε μεταφορά του ήχου χρονικά καθώς η ηχητική ενέργεια αποθηκεύεται στο χώρο (πυκνωτής) και όχι την απότομη μηδένισή του. Σ' ένα κλειστό χώρο αν μία ηχητική πηγή σταματήσει να εκπέμπει ένα ήχο αυτός δεν θα σταματήσει ακαριαία αλλά θα σβήσει εκθετικά. Χρόνος αντήχησης (T_{60}) ορίζεται ο χρόνος που απαιτείται ώστε η στάθμη του ήχου (L_p) να μειωθεί κατά 60 dB μετά το μηδενισμό του πρωτογενούς ήχου. Ένας προσεγγιστικός τρόπος υπολογισμού του χρόνου αντήχησης είναι ο $T_{60} = \frac{0.161V}{A_s}$ με V τον όγκο του χώρου και A_s απορρόφηση του δωματίου ($A_s = \sum a_i s_i$ a_i και s_i ο συντελεστής απορρόφησης και το εμβαδόν κάθε τοίχου)



Εικόνα 15 Εκπομπή ηχητικής ενέργειας από πηγή στον ελεύθερο χώρο (α) πάνω από ανακλαστική επιφάνεια (β) και σε κλειστό χώρο (γ). Οι παραστάσεις δεξιά δίνουν την ηχητική πίεση σε σημείο κοντά στην πηγή και τον δέκτη αντίστοιχα.

2. ΘΕΩΡΙΑ ΨΗΦΙΑΚΟΥ ΗΧΟΥ

2.1 Εισαγωγή

Τα ψηφιακά ηχητικά συστήματα εκτελούν όλες τις λειτουργίες ηλεκτροακουστικών συσκευών (μετατροπή, επεξεργασία, αποθήκευση, αναπαραγωγή, ανάλυση, μέτρηση, κλπ.), σε ηχητικά σήματα που εμφανίζονται στο ψηφιακό πεδίο.

Τα πρώτα τέτοια ψηφιακά ηχητικά συστήματα (Digital Audio Systems) που επέτρεπαν καταγραφή και αναπαραγωγή ήχου με ικανοποιητική πιστότητα εμφανίστηκαν στα τέλη της δεκαετίας του '60 και εξελίχθηκαν στις ΗΠΑ (από τον T. G. Stookham, Πανεπιστήμιο Utah), στην Ιαπωνία (από την εταιρεία NHK) και στην Ευρώπη (από την εταιρεία Philips). Από τα μέσα της δεκαετίας του '70 άρχισε η εμπορική διάθεσή τους και από τότε τα συστήματα αυτά εμφανίζουν εντυπωσιακή εξάπλωση τόσο σε επαγγελματικές, όσο και σε οικιακές εφαρμογές. Τα σημαντικά πλεονεκτήματα των συστημάτων αυτών σε σύγκριση με τα αντίστοιχα αναλογικά είναι κυρίως:

- 1 Η υψηλότερη σχέση σήματος προς θόρυβο και η αυξημένη δυναμική περιοχή λειτουργίας.
- 2 Η ευκολότερη αποθήκευση και μετάδοση του ηχητικού σήματος και η επικοινωνία των συσκευών με τα ψηφιακά τηλεπικοινωνιακά και υπολογιστικά συστήματα.
- 3 Η ευκολία στην χρήση, το χαμηλότερο κόστος παραγωγής και το μικρότερο μέγεθος.
- 4 Η αέναη δυνατότητα διατήρησης της ηχογραφημένης πληροφορίας σε ψηφιακή μορφή.
- 5 Η αυξημένη δυνατότητα επεξεργασίας και εφαρμογής υπολογιστικών τεχνικών

Σαν κύριο μειονέκτημα μπορεί να θεωρηθεί η ανάγκη μετατροπής του εγγενώς αναλογικού ηχητικού σήματος σε ψηφιακό (και η αντίστοιχη μετατροπή του ψηφιακού σε αναλογικό κατά την αναπαραγωγή), διαδικασία που εμφανίζει πάντα ορισμένες παραμορφώσεις.

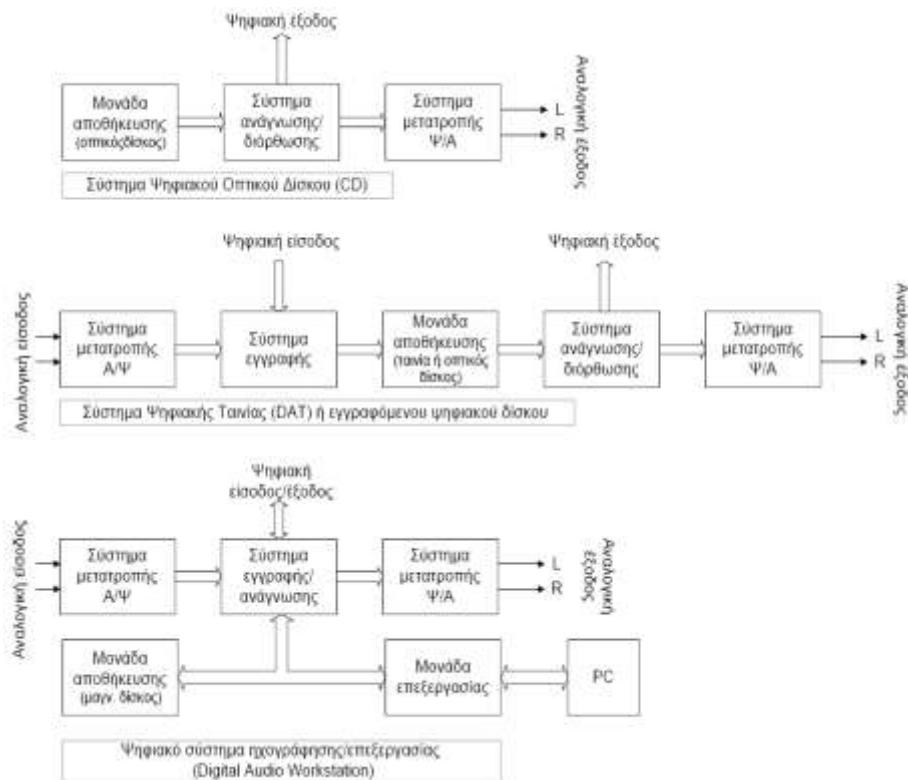
Τα διάφορα ψηφιακά υποσυστήματα που συνδυάζονται σε πολυπλοκότερες συσκευές ανήκουν στις παρακάτω κατηγορίες (Σχήμα 2.1).

- 1 Μετατροπείς Αναλογικού σε Ψηφιακό (Α/Ψ) και Ψηφιακού σε Αναλογικό (Ψ/Α) ηχητικού σήματος (Analog to Digital, A/D ή ADC και Digital to Analog, D/A Converter ή DAC).
- 2 Ψηφιακά υποσυστήματα κωδικοποίησης και μετάδοσης ηχητικών σημάτων (PCM, ADPCM κλπ).
- 3 Υποσυστήματα αποθήκευσης ψηφιακών ηχητικών δεδομένων (μαγνητική ταινία, σκληρός δίσκος, Οπτικός Δίσκος CD, ταινία DAT, κλπ).
- 4 Ειδικοί ψηφιακοί επεξεργαστές ήχου (audio processors).
- 5 Υποσυστήματα λογισμικού επεξεργασίας και διαχείρισης ηχητικών δεδομένων (π.χ. προγράμματα διαχείρισης ηχητικών αρχείων σε υπολογιστές).
- 6 Συστήματα(προγράμματα) συμπίεσης ήχου.

Οι τεχνολογικές εξελίξεις και βελτιώσεις των παραπάνω υποσυστημάτων ανοίγουν το δρόμο σε νέες εφαρμογές και χρήσεις τους, όπως π.χ. τηλεόραση με πολυκαναλικό ψηφιακό

ήχο υψηλής πιστότητας, συστήματα δορυφορικής μετάδοσης και διανομής ψηφιακών μουσικών σημάτων, δίσκοι CD, SACD (Super Audio CD) και DVD (Digital Versatile Disc) με πολυκαναλικό ψηφιακό ήχο υψηλής πιστότητας, οπτικοακουστικές οικιακές συσκευές σε επικοινωνία με PC, συστήματα ψηφιακής ραδιοφωνίας (DAB), καθώς και νέες μέθοδοι σύνθεσης, ηχογράφησης και διανομής της μουσικής (π.χ. μέσω δικτύων).

Όπως αναφέρθηκε και προηγουμένως, η αποθήκευση αναλογικών ηχητικών δεδομένων σε οποιοδήποτε μέσο αποθήκευσης και διανομής ψηφιακής πληροφορίας απαιτεί την μετατροπή των αρχικών αναλογικών δεδομένων σε ψηφιακά. Η μετροπή αυτή γίνεται μέσω ειδικών διατάξεων μετατροπής από αναλογικό σε ψηφιακό (Α/Ψ μετατροπείς - A/D Converters), τα ειδικά χαρακτηριστικά λειτουργίας των οποίων θα αναλυθούν σε επόμενη παράγραφο. Τέτοιες διατάξεις Α/Ψ μετατροπέων είναι ενσωματωμένες σε οποιοδήποτε ψηφιακό ηχητικό σύστημα που δέχεται ως είσοδο αναλογικό ηχητικό σήμα (π.χ. PC, DAT, CD-R, Minidisk κ.λ.π.) και τροφοδοτούν τις λοιπές ψηφιακές βαθμίδες επεξεργασίας/αποθήκευσης με ψηφιακά δεδομένα τα οποία κωδικοποιούνται με διάφορους τρόπους (π.χ. Παλμοκωδική Διαμόρφωση -PCM).



Σχήμα 2.1 Τυπικές υλοποιήσεις ψηφιακών συστημάτων

Η κωδικοποίηση των ψηφιακών δεδομένων είναι απαραίτητη αφενός μεν για τη μετάδοσή τους διά μέσου του διαθέσιμου καναλιού μετάδοσης, αφετέρου δε για την αποθήκευσή τους σε κάποιο ψηφιακό μέσο αποθήκευσης, αποτελεί δε τη δυαδική αναπαράσταση "καναλιού" των δεδομένων. Εκτός όμως από τις δύο παραπάνω λειτουργίες (μετάδοση/αποθήκευση), τα ψηφιακά δεδομένα μπορεί να υποστούν και περαιτέρω επεξεργασία στο ψηφιακό πεδίο.

2.2 Βασικές Αρχές Λειτουργίας

Η διαδικασία μετατροπής ενός αναλογικού ηχητικού σήματος σε ψηφιακό συνίσταται από δύο βασικές λειτουργίες: τη δειγματοληψία του αρχικού σήματος και τον κβαντισμό των τιμών του σήματος διακριτού χρόνου που προκύπτει από τη δειγματοληψία. Πιο συγκεκριμένα, δειγματοληψία είναι η διαδικασία κατά την οποία ένα αναλογικό σήμα $s_c(t)$ (σήμα του οποίου η ανεξάρτητη χρονική μεταβλητή t , παίρνει συνεχείς τιμές) μετατρέπεται σε σήμα διακριτού χρόνου (σήμα το οποίο ορίζεται σε διακριτές χρονικές στιγμές), πραγματοποιείται δηλαδή μια απεικόνιση της μορφής $s_c(t) \rightarrow s_d(nTs)$, όπου $n=1, 2, \dots$ είναι ο αριθμός του δείγματος και Ts (sec) είναι η περίοδος δειγματοληψίας. Έτσι, κατά τη διαδικασία της δειγματοληψίας, το χρονικά συνεχές αναλογικό σήμα μετατρέπεται σε πεπερασμένο αριθμό διαδοχικών τιμών, τα οποία μπορούν να ονομαστούν αναλογικά στιγμιότυπα. Η ονομασία αυτή πηγάζει από το γεγονός ότι το $s_d(nTs)$ μπορεί να πάρει ένα πλήθος από άπειρες δυνατές τιμές, γεγονός που οφείλεται στο ότι το σήμα δεν έχει ακόμα κβαντοποιηθεί. Είναι δηλαδή ένα σήμα διακριτού χρόνου αλλά συνεχούς πλάτους.

Κβαντισμός είναι η διαδικασία κατά την οποία οι διαδοχικές τιμές της στάθμης (πλάτους) του σήματος διακριτού χρόνου $s_d(nTs)$ μετατρέπονται σε διάκριτες (ψηφιακές) τιμές, πραγματοποιείται δηλαδή μια απεικόνιση της μορφής $s_d(nTs) \rightarrow \langle s_d(nTs) \rangle$, όπου $\langle s_d(nTs) \rangle$ είναι η κβαντισμένη τιμή στάθμης του σήματος. Κατά την παραπάνω μετατροπή γίνεται προφανώς κάποιο είδος στρογγυλοποίησης, οπότε προκύπτει ένα σφάλμα $e(n)$ το οποίο καλείται σφάλμα κβαντισμού. Το τελικό σήμα $\langle s_d(nTs) \rangle$, είναι πλέον διάκριτο τόσο ως προς τον χρόνο, όσο και ως προς το πλάτος και καλείται ψηφιακό.

Είναι γνωστό και από άλλα μαθήματα (π.χ. Ψηφιακή Επεξεργασία Σημάτων), ότι τα αναλογικά σήματα της ακουστικής πίεσης ή του ηλεκτρικού σήματος που αντιστοιχεί στην πίεση αυτή, μετατρέπονται σε ψηφιακά με τη χρήση υποσυστημάτων μετατροπής Α/Ψ (Analog-to-Digital Converters, ADC's). Σε τέτοια υποσυστήματα, οι διαδικασίες κβαντισμού και δειγματοληψίας πραγματοποιούνται ταυτόχρονα. Το αντίστροφο επιτυγχάνεται με τη χρήση Ψ/Α μετατροπέων (Digital-to-Analog Converters, DAC's), όπου χρονικά διακριτά δείγματα και οι κβαντισμένες στάθμες μετατρέπονται (κατά προσέγγιση) σε συνεχείς αναλογικές τιμές. Πρακτικά και θεωρητικά προβλήματα εμφανίζονται και στις δύο περιπτώσεις, σε σχέση με την πιστότητα αναπαραγωγής των ακουστικών σημάτων. Για την ακριβή και πιστή μετατροπή (είτε σε αντικειμενική, είτε σε υποκειμενική βάση) πρέπει να ικανοποιούνται ορισμένες συνθήκες δειγματοληψίας και κβαντισμού, οι οποίες αναλύονται στις επόμενες παραγράφους.

2.2.1 Δειγματοληψία ηχητικών σημάτων

Είναι γνωστό από το θεώρημα της δειγματοληψίας (γνωστό και ως θεώρημα του Nyquist) ότι η συχνότητα δειγματοληψίας περιορίζει το συχνοτικό εύρος του αναλογικού σήματος που μπορεί να αναπαρασταθεί σωστά κατά την ψηφιοποίηση του. Συγκεκριμένα, αν F_s (σε κύκλους ανά δευτερόλεπτο, δηλαδή Herz) είναι η τιμή της συχνότητας δειγματοληψίας, η μέγιστη συχνότητα του αναλογικού σήματος για να μην εμφανιστούν φαινόμενα φασματικής αναδίπλωσης στο ψηφιακό σήμα πρέπει να είναι υποδιπλάσια της συχνότητας

δειγματοληψίας, δηλαδή πρέπει να ισχύει $f_{\max} \leq Fs/2$. Αυτή η θεώρηση υπαγορεύει ότι για την σωστή δειγματοληψία ενός ημιτονικού σήματος χρειάζονται τουλάχιστον δύο δείγματα ανά περίοδο.

Από το θεώρημα του Nyquist αποδεικνύεται ότι κατά την απεικόνιση στο πεδίο του χρόνου $s_c(t) \rightarrow s_d(nTs)$, στο πεδίο της συχνότητας προκύπτει μια σχέση μεταξύ του φάσματος του αναλογικού σήματος $S_c(f)$ και του φάσματος του ψηφιακού σήματος $S_d(f)$ της μορφής

$$S_d(f) = \frac{1}{Ts} \sum_{k=-\infty}^{+\infty} S_c(f - kFs) \quad (2.1)$$

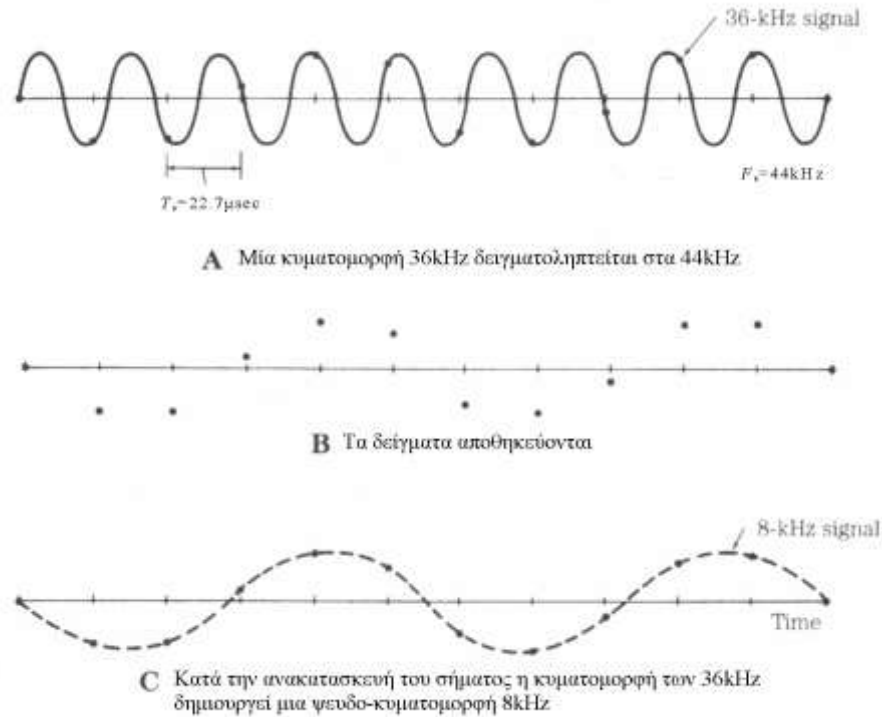
Πρέπει να σημειωθεί ότι οι παραστάσεις $S_c(f)$ και $S_d(f)$ προκύπτουν από το συνεχή και διακριτό μετασχηματισμό Fourier των σημάτων $s_c(t)$ και $s_d(nTs)$ αντίστοιχα. Η σχέση (2.1) μας λέει ότι ο μετασχηματισμός Fourier του σήματος $s_d(nTs)$ είναι η περιοδική αναπαράσταση του μετασχηματισμού του σήματος $s_c(t)$ στα άπειρα ακέραια πολλαπλάσια της συχνότητας δειγματοληψίας. Η σχέση (2.1) εκφράζει επίσης με μαθηματικό τρόπο την δημιουργία του φαινομένου της φασματικής αναδίπλωσης, δηλαδή την εμφάνιση συχνοτικών συνιστωσών στο ψηφιακό σήμα οι οποίες προέρχονται από μεγαλύτερες συχνότητες του αναλογικού σήματος.

Ας θεωρήσουμε ότι $Fs=44$ kHz και ας θέσουμε $f=8$ kHz στην (2.1). Εξετάζουμε το φαινόμενο λαμβάνοντας υπόψη για το k μόνο τις τιμές -1,0 και 1.

Η σχέση τώρα μας λέει ότι

$$\begin{aligned} S_d(8\text{kHz}) &= \frac{1}{Ts} [S_c(8\text{kHz} - 44\text{kHz}) + S_c(8\text{kHz}) + S_c(8\text{kHz} + 44\text{kHz})] \Rightarrow \\ S_d(8\text{kHz}) &= \frac{1}{Ts} [S_c(-32\text{kHz}) + S_c(8\text{kHz}) + S_c(54\text{kHz})]. \end{aligned} \quad (2.2)$$

Γίνεται λοιπόν φανερό ότι εκτός από τις συνιστώσες του αναλογικού σήματος στα 8kHz, συνιστώσες των 32kHz και των 54kHz θα εμφανιστούνε στο ψηφιακό σήμα, εκτός αν έχουν ληφθεί μέτρα έτσι ώστε το αναλογικό -προς ψηφιοποίηση- σήμα να μην έχει τέτοιες συνιστώσες, δηλαδή να έχει γίνει χρήση κατάλληλου φίλτρου χαμηλής διέλευσης συχνοτήτων. Ένα ζήτημα ενδεχομένως να δημιουργείται σχετικά με το αρνητικό πρόσημο συχνότητας (-32kHz). Θα πρέπει να θυμηθεί ο αναγνώστης ότι ο μετασχηματισμός Fourier ενός σήματος έχει μιγαδικές τιμές και ότι αν το σήμα $g(t)$ είναι πραγματικό (όπως όλα τα σήματα που εξετάζονται στο μάθημα) θα ισχύει $G(-f)=G^*(f)$, όπου το ‘*’ εκφράζει μιγαδική συζυγία και $G(f)$ είναι ο μετασχηματισμός Fourier συνεχούς χρόνου του σήματος $g(t)$.



Σχήμα 2.2 Τυπική περίπτωση του φαινομένου της φασματικής αναδίπλωσης.

Στο σχήμα 2.2 φαίνεται ένα απλό παράδειγμα φασματικής αναδίπλωσης όπου κατά την ψηφιοποίηση ενός ημιτονικού σήματος συχνότητας 36kHz με συχνότητα δειγματοληψίας 44kHz δημιουργείται μια ψευδο-συχνότητα 8kHz στο ψηφιακό σήμα. Ο περιορισμός του συχνοτικού εύρους του αναλογικού σήματος είναι σημαντικότερος για τη σωστή ψηφιοποίηση του σήματος.

Από όσα αναφέρθηκαν προηγουμένως και γνωρίζοντας ότι τα ακουστικά σήματα γίνονται αντιληπτά μέχρι περίπου τη συχνότητα των 20kHz, προκύπτει ότι η συχνότητα δειγματοληψίας F_s (Hz) θα πρέπει να είναι τουλάχιστον ίση με 40kHz. Παρόλα αυτά, λόγω των αρχικών τεχνικών προδιαγραφών των συστημάτων video στα οποία γινόταν αρχικά η αποθήκευση των ψηφιακών ηχητικών δεδομένων, η συχνότητα δειγματοληψίας για τα οικιακά ψηφιακά ηχητικά συστήματα, ορίστηκε στα 44.1kHz. Στην πράξη όμως τα ακουστικά σήματα δεν περιορίζονται στη μέγιστη τιμή των 22.05kHz που απαιτεί η παραπάνω συχνότητα δειγματοληψίας, γι' αυτό και απαιτείται προ-φιλτράρισμα του αρχικού αναλογικού σήματος πριν οδηγηθεί στη βαθμίδα δειγματοληψίας, με στόχο την αποκοπή συχνοτήτων μεγαλύτερων από τη μέγιστη συχνότητα που στο συγκεκριμένο σύστημα ικανοποιεί τη συχνότητα δειγματοληψίας. Αυτό επιτυγχάνεται με τη χρήση κατωδιαβατού (low-pass) φίλτρου που συχνά ονομάζεται και φίλτρο anti-aliasing. Ομοίως, κατά τη διαδικασία της Ψ/Α μετατροπής, λόγω των κβαντισμένων (διακριτών) τιμών πλάτους του αναπαραγώμενου σήματος, δημιουργούνται συχνότητες υψηλότερες της πραγματικής μέγιστης συχνότητας του σήματος, οι οποίες αποκόπτονται με παρόμοιο φίλτρο που συχνά ονομάζεται φίλτρο ανακατασκευής ή εξομάλυνσης reconstruction/smoothing filter).

2.2.2 Κβαντισμός ηχητικών σημάτων

Όπως έχει ήδη αναφερθεί, το δεύτερο στάδιο στη διαδικασία μετατροπής ενός αναλογικού σήματος σε ψηφιακό είναι ο κβαντισμός των τιμών πλάτους του σήματος διακριτού χρόνου που προκύπτει κατά τη δειγματοληψία του αρχικού αναλογικού σήματος. Θεωρητικά, η αναλογική κυματομορφή του ακουστικού σήματος διακριτού χρόνου $s(n)$ έχει άπειρο αριθμό τιμών πλάτους. Κατά την Α/Ψ μετατροπή, οι τιμές αυτές αντιστοιχίζονται σε πεπερασμένο αριθμό σταθμών (τιμών πλάτους), το πλήθος των οποίων εξαρτάται από την τάξη του κβαντιστή. Η διαδικασία αυτή μπορεί να περιγραφεί μαθηματικά ως

$$\hat{s}(n) = Q\{s(n)\} \quad (2.3)$$

Στην ανάλυση που ακολουθεί θα εννοείται ότι ο αριθμός δείγματος n αντιστοιχεί στην $t=nTs$ τιμή του αναλογικού σήματος $s_c(t)$ και ο δείκτης d θα εξυπακούεται, δηλαδή

$$s_d(nTs) = s(n) \quad (2.4)$$

και

$$\langle s_d(nTs) \rangle = Q\{s(n)\} = \hat{s}(n) \quad (2.5)$$

Το $s(n)$ θα υπονοεί δηλαδή σήμα διακριτού χρόνου και συνεχούς πλάτους (αναλογικό στιγμιότυπο) και το $\hat{s}(n)$ δείγμα διακριτού χρόνου και διακριτού πλάτους (ψηφιακό σήμα).

Υπάρχουν διάφοροι τύποι κβαντιστών που υλοποιούν την παραπάνω γενική σχέση, με πιο κοινό τον ομοιόμορφο, γραμμικό κβαντιστή, ο οποίος στρογγυλεύει την αναλογική τιμή της εισόδου του στην πλησιέστερη στάθμη (rounding)

Κάθε δυνατή έξοδος του κβαντιστή (στάθμη) μπορεί να κωδικοποιηθεί δυαδικά από μία ψηφιακή λέξη μήκους N bit, οπότε το συνολικό πλήθος των σταθμών εξόδου είναι 2^N . Το N είναι σημαντικός παράγοντας για την ποιότητα ψηφιακής αναπαράστασης της ηχητικής πληροφορίας. Το N αναφέρεται επίσης ως ευκρίνεια κβαντισμού και ως τάξη του κβαντιστή. Οι στάθμες μεταξύ διαδοχικών σταθμών κβαντισμού απέχουν μεταξύ τους απόσταση Δ , η οποία καλείται βήμα κβαντισμού (ή αλλιώς τιμή του Ελάχιστου Σημαντικού Ψηφίου -Least Significant Bit, LSB) και δίνεται από τη σχέση:

$$\Delta = LSB = \frac{s_{\max}}{2^{N-1}} \quad (2.6)$$

όπου s_{\max} το μέγιστο πλάτος εισόδου που μπορεί να κβαντιστεί σωστά από το συγκεκριμένο κβαντιστή. Στην πράξη, το βήμα κβαντισμού Δ ορίζεται σαν η αύξηση της τάσης εισόδου που απαιτείται να εφαρμοστεί στην είσοδο ενός κβαντιστή για την αλλαγή του λιγότερο σημαντικού ψηφίου (LSB) της εξόδου του.

Παράδειγμα: Μία ψηφιακή κάρτα ήχου δέχεται στην είσοδό της αναλογικό σήμα $\pm 5V$. Κατά την ψηφιοποίηση του σήματος χρησιμοποιείται κβαντιστής τάξης $N=12$ bit. Ποιά είναι η ελάχιστη μεταβολή της τάσης σε Volt που απαιτείται για την αλλαγή του λιγότερου σημαντικού ψηφίου (LSB); Από πόσους σταθμούς κβαντοποίησης διέρχεται ένα ημιτονικό σήμα πλάτους 0.1 V;

Είναι $\Delta = \frac{5V}{2^{12-1}} = \frac{5V}{2^{11}} = 0.00244V = 2,44mV$. Το ημιτονικό σήμα έχει πλάτος από κορυφή σε

κορυφή (p-p) $2 \cdot 0,1V = 0,2V = 200mV$. Επομένως διέρχεται από $200/2,44 = 82$ σταθμούς κβαντισμού.

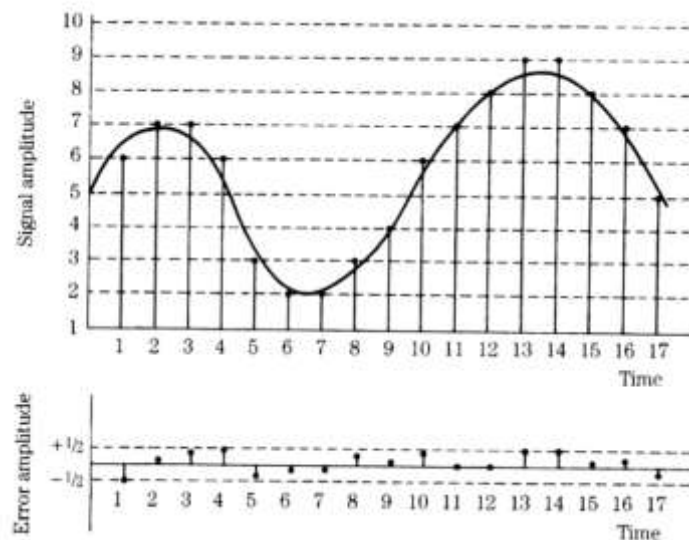
Όπως φαίνεται, ο κβαντισμός είναι μία μη γραμμική διαδικασία καθώς, εάν τα δείγματα ενός σήματος διακριτού χρόνου κβαντιστούν, η εύρεση των αρχικών τιμών πλάτους του σήματος είναι αδύνατη. Επιπλέον, η μετατροπή αυτή εμφανίζει σφάλμα το οποίο είναι δυνατόν να μετρηθεί, αλλά και είναι ακουστό. Αν θεωρήσουμε ότι κατά τον ιδανικό κβαντισμό ενός αναλογικού σήματος διακριτού χρόνου $s(n)$ παράγεται ένα σήμα $\hat{s}(n)$ και εμφανίζεται σφάλμα κβαντισμού $e(n)$, τότε (βλ. σχήμα 1.3)

$$e(n) = \hat{s}(n) - s(n) \quad (2.7)$$

Η μορφή της συνάρτησης $e(n)$ έχει μεγάλη σημασία στην πιστότητα αναπαραγωγής ενός ψηφιακού ηχητικού συστήματος και γενικά εξαρτάται από την στιγμιαία τιμή πλάτους του σήματος $s(n)$ σε σχέση με το βήμα κβαντισμού Δ του μετατροπέα. Για την ανάλυση της μορφής του σφάλματος κβαντισμού ως μία στατιστική διαδικασία είναι απαραίτητο να γίνουν οι ακόλουθες παραδοχές:

1. Η σειρά των τιμών του σφάλματος $e(n)$ είναι στατιστικά στάσιμη
2. Είναι στατιστικά ασυσχέτιστη με το σήμα $s(n)$
3. Τα δείγματά της είναι επίσης μεταξύ τους ασυσχέτιστα και προσεγγίζουν τη συμπεριφορά του λευκού θορύβου
4. Η συνάρτηση πυκνότητας πιθανότητας του λάθους είναι ομοιόμορφη

Αυτές οι παραδοχές γενικά ισχύουν όταν το πλάτος του σήματος εισόδου είναι σχετικά μεγάλο και επομένως κατά τον κβαντισμό του αξιοποιείται ένας μεγάλος αριθμός από κβαντικές στάθμες. Οι παραδοχές αυτές είναι σημαντικές για τον ορισμό της δυναμικής περιοχής του ψηφιακού κυκλώματος, κάτι που συζητιέται σε επόμενη παράγραφο.



Σχήμα 2.3 Σφάλμα κβαντισμού κατά τον κβαντισμό ενός ημιτονικού σήματος

2.3 Αριθμητική αναπαράσταση σταθερού σημείου και βήμα κβαντισμού

Γενικά, η αριθμητική αναπαράσταση ψηφιακών δεδομένων υλοποιείται βασικά με δύο

τρόπους: (α) με αριθμητική σταθερού σημείου (fixed-point number representation) και (β) με αριθμητική κινητής υποδιαστολής (floating-point number representation). Η επιλογή του κατάλληλου τύπου αριθμητικής αναπαράστασης που θα χρησιμοποιηθεί στα πλαίσια κάποιας συγκεκριμένης εφαρμογής εξαρτάται βασικά από διάφορους παράγοντες. Μόνο η αναπαράσταση με αριθμητική σταθερού σημείου εξετάζεται εδώ.

Η αριθμητική αναπαράσταση σταθερού σημείου ψηφιακών ηχητικών δειγμάτων με εύρος τιμών από $-s_{\max}$ έως s_{\max} προκύπτει με κανονικοποίηση των τιμών των δειγμάτων εισόδου και με προσέγγιση της προκύπτουσας τιμής από ζυγισμένο άθροισμα πεπερασμένου πλήθους δυνάμεων του 2. Συγκεκριμένα, εάν $\bar{s}(k)$ το ψηφιακό ηχητικό δείγμα που αντιστοιχεί στην k -οστή περίοδο δειγματοληψίας του αρχικού αναλογικού σήματος, η κανονικοποιημένη τιμή που αντιστοιχεί στο παραπάνω δείγμα δίνεται από τη σχέση\

$$\hat{s}(k) = \frac{\hat{s}_d(k)}{|s_{\max}|} \quad (2.8)$$

και το νέο εύρος τιμών του κανονικοποιημένου σήματος είναι από -1 έως 1 . Στη συνέχεια, οι προκύπτουσες κανονικοποιημένες τιμές εκφράζονται ως ζυγισμένο άθροισμα πεπερασμένου μήκους, δηλαδή:

$$\hat{s}(k) = -b_0 + b_1 2^{-1} + b_2 2^{-2} + \dots + b_{N-1} 2^{-(N-1)} \quad (2.9)$$

όπου οι N σε πλήθος τιμές "ζυγίσματος" $b_0, b_1, \dots, b_i, \dots, b_{N-1}$ αποτελούν τα bit της δυαδικής αναπαράστασης του συγκεκριμένου δείγματος (με δυνατές τιμές 1 ή 0). Το bit b_0 αποτελεί το Πλέον Σημαντικό Ψηφίο (Most Significant Bit, MSB), ενώ το b_{N-1} το Λιγότερο Σημαντικό Ψηφίο (Least Significant Bit, LSB). Για θετικές τιμές της κανονικοποιημένης τιμής, το b_0 είναι 0, καθώς για αρνητικές είναι 1. Προφανώς, το εύρος των δυνατών τιμών των κανονικοποιημένων δειγμάτων που προκύπτει κάνοντας χρήση της σχέσης (2.9) είναι

$$-1 \leq \hat{s}(k) \leq 1 - 2^{-(N-1)} \quad (2.10)$$

Η μικρότερη αρνητική τιμή (-1) δημιουργείται για $b_0=1$ και όλα τα άλλα bits ίσα με μηδέν, δηλαδή για ψηφιακή λέξη της μορφής $[1000\dots000]$. Η μεγαλύτερη θετική τιμή δημιουργείται αντίστοιχα για ψηφιακές λέξεις της μορφής $[0111\dots111]$.

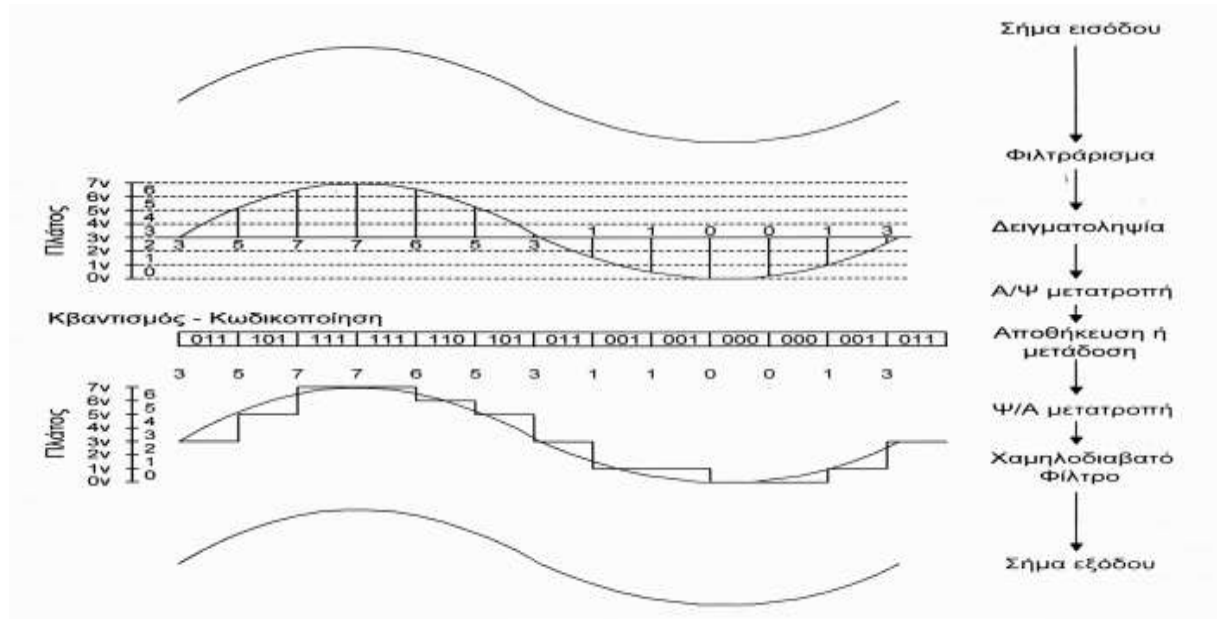
Από όσα έχουν ειπωθεί εως τώρα είναι δυνατόν να επαναδιατυπώσουμε την τιμή του βήματος κβαντισμού Δ με βάση τις κανονικοποιημένες τιμές του σήματος εισόδου από -1 έως $1 - 2^{-(N-1)}$. Αυτή προκύπτει αν διαιρέσουμε το εύρος των δυνατών τιμών του ψηφιακού ηχητικού δείγματος με το πλήθος των διαθέσιμων σταθμών κβαντισμού. Από τη σχέση (1.3.3) φαίνεται ότι το εύρος δυνατών τιμών, δηλαδή, η απόσταση από την ελάχιστη αρνητική τιμή στη μέγιστη θετική είναι $2 - 2^{-(N-1)}$. Για μεγάλες σχετικά τιμές του N , ο όρος $2^{-(N-1)}$ γίνεται πολύ μικρός οπότε μπορούμε να τον αγνοήσουμε και να θεωρήσουμε ότι το εύρος τιμών είναι ίσο με 2. Οπότε, το βήμα κβαντισμού γράφεται

$$\Delta = \frac{2}{2^N} = 2^{-(N-1)} \quad (2.11)$$

Το βήμα κβαντισμού Δ λέγεται και διακριτική ικανότητα, καθότι καθορίζει την ποιότητα της ανάλυσης κατά την ψηφιοποίηση του ηχητικού σήματος. Όσο πιο μεγάλο το μήκος λέξης (δηλαδή το N) τόσο πιο μικρό το Δ και άρα τόσο καλύτερη η ανάλυση που προσφέρει το κύκλωμα κβαντοποίησης. Ο κβαντιστής γίνεται λοιπόν μεγάλος παράγοντας για τον

καθορισμό της ποιότητας του ψηφιοποιημένου ηχητικού σήματος. Το βήμα Δ του κβαντιστή καθορίζει άμεσα με την ποιότητα του αναλογικού σε ψηφιακό μετατροπέα και αυτός με τη σειρά του καθορίζει την ποιότητα του ηχητικού σήματος κατά την αναπαραγωγή του. Η ανάλυση ως εδώ κλείνει τη βασική θεωρία για την όλη διαδικασία μετατροπής του αναλογικού σήματος σε ψηφιακό και αντίστροφα. Η όλη διαδικασία απεικονίζεται συνοπτικά και στο Σχήμα 2.4.

Παράδειγμα Κάρτα ήχου δουλεύει στα ± 5 V αν κάνουμε ψηφιοποίηση στα 8 bit. Πόσες στάθμες υπάρχουν στα 0-5 V. Ποια η ελάχιστη μεταβολή τάσης για την αλλαγή του LSB. Αν έχω σήμα πλάτους 1.5 V πόσες στάθμες εκμεταλλεύεται. Αν έχω στιγμιαία τάση -1.2 V πως θα είναι αυτή αποθηκευμένη ψηφιακά.



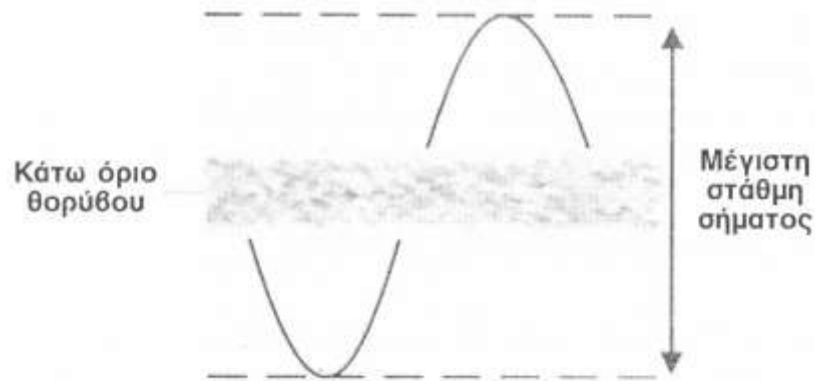
Σχήμα 2.4 Διαδικασία μετατροπής αναλογικού σήματος σε ψηφιακό και αντίστροφα

2.4 Δυναμική Περιοχή Ηχητικού Συστήματος

Η δυναμική περιοχή είναι από τα πιο σημαντικά χαρακτηριστικά ενός ηχητικού συστήματος. Η δυναμική περιοχή ορίζεται τόσο για αναλογικά όσο και για ψηφιακά συστήματα.

2.4.1 Δυναμική περιοχή και SNR σε ηχητικά συστήματα

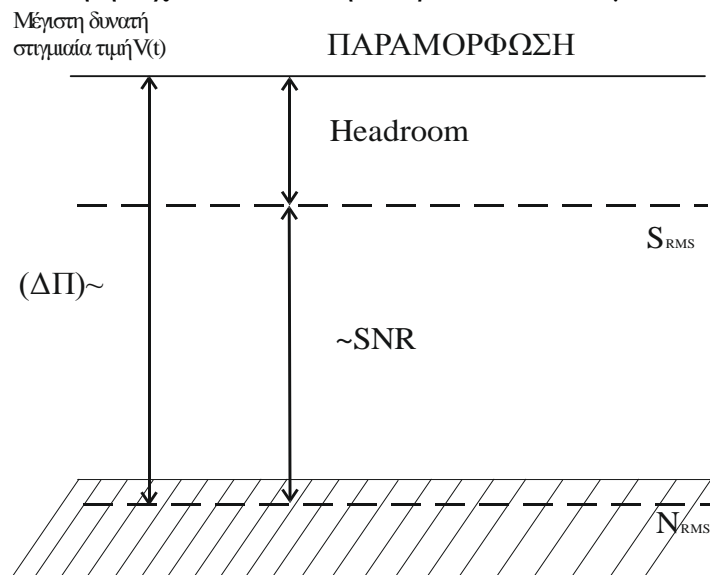
Υπό μία έννοια, δυναμική περιοχή είναι η διαφορά, σε dB, μεταξύ των πιο δυνατών και πιο ήσυχων περασμάτων του προγράμματος που μπορούν να αποδοθούν με ευκρίνεια από το ηχοσύστημα. Εκφράζει τη μέγιστη διαφορά σε ακουστές στάθμες του προγράμματος. Κατ' ουσία υπάρχουν δύο απόλυτες στάθμες σε κάθε ηλεκτρονικό σύστημα, η μέγιστη στάθμη και το κάτω όριο θορύβου. Και τα δύο φαίνονται στο Σχήμα 2.5.



Σχήμα 2.5 Μέγιστη στάθμη σήματος και κάτω όριο θορύβου

Είναι φανερό ότι ένα σήμα πολύ μικρής στάθμης κινδυνεύει να ‘πνιγεί’ μέσα στο θόρυβο, ενώ από την άλλη, ένα σήμα πολύ μεγάλης στάθμης κινδυνεύει να παραμορφώσει (ψαλιδιστεί). Για να μεταδοθεί επομένως με ευκρίνεια μία ηχητική πληροφορία, θα πρέπει να κινείται μεταξύ της ‘Σκύλας’ του θορύβου και της ‘Χάρυβδης’ της παραμόρφωσης.

Ποιες όμως είναι οι αιτίες του θορύβου σε ένα ηλεκτρονικό σύστημα; Σε ένα αναλογικό ηχοσύστημα, ο θόρυβος προέρχεται κυρίως από παρεμβολές που δημιουργούνται από τα ενεργά στοιχεία του κυκλώματος, όπως οι τελεστικοί ενισχυτές και τα τρανζίστορ, τα οποία τροφοδοτούνται με τάσεις πολλές φορές άνω των 20 V. Ακόμα όμως και σε ένα παθητικό σύστημα, όπως το δυναμικό μικρόφωνο, ο ηλεκτρονικός θόρυβος είναι παρών και οφείλεται στην τυχαία κίνηση των ηλεκτρονίων στην αντίσταση του πηνίου (θερμικός θόρυβος). Δεν υπάρχει επομένως απόλυτη ησυχία ούτε στα ηλεκτρονικά κυκλώματα.



Σχήμα 2.6 Δυναμική περιοχή και headroom σε ένα αναλογικό ηχητικό σύστημα

Ο κλασικός συμβολισμός για την επεξήγηση της δυναμικής περιοχής φαίνεται στο σχήμα 2.6. Η δυναμική περιοχή τυπικά εκτείνεται από τη μέση τετραγωνική τιμή του θορύβου μέχρι την απόλυτη μέγιστη στάθμη. Πρέπει να σημειωθεί ωστόσο ότι ένας τέτοιος ορισμός της δυναμικής περιοχής, αν και έχει σημασία για τεχνικούς σκοπούς, δεν

αντιπροσωπεύει την υποκειμενικά αισθητή δυναμική περιοχή, όπως αυτή ορίζεται για το ανθρώπινο αισθητήριο όργανο της ακοής. Συγκεκριμένα, όταν ακούμε η ένταση ενός ήχου δεν εξαρτάται τόσο από τις μέγιστες στιγμιαίες τιμές της ηχητικής πίεσης, αλλά από τη μέση τετραγωνική της τιμή. Η μέση τετραγωνική τιμή της πίεσης είναι βέβαια ανάλογη της αντίστοιχης τιμής της τάσης με την οποία αναπαρίσταται το ηχητικό σήμα. Θεωρώντας ένα αναλογικό ηχητικό σύστημα, η μεταβολή της τάσης είναι συνεχούς χρόνου και επομένως η μέση τετραγωνική τιμή ενός ηχητικού σήματος $s(t)$ ορίζεται από τη σχέση²

$$S_{RMS} = \left(\frac{1}{T} \int_0^T s^2(t) dt \right)^{1/2}. \quad (2.12)$$

Γενικά, η μέση τετραγωνική τιμή ενός δυναμικού μεγέθους εκφράζει ενέργεια. Είναι αξιοσημείωτο ότι το σύστημα ακοής του ανθρώπου, σε αναλογία με τη σχέση (2.12), εκτελεί κάτι σαν ολοκλήρωση κατά την επεξεργασία του ακουστικού σήματος. Είναι λοιπόν φανερό ότι αν θέλουμε ένα αντιπροσωπευτικό μέγεθος για τη δυναμική περιοχή ενός ηχοσυστήματος θα πρέπει να συγκρίνουμε ενέργειες (μέσες τετραγωνικές τιμές του σήματος) και όχι στιγμιαίες απόλυτες στάθμες. Για αυτό το λόγο, ορίζεται η σχέση σήματος προς θόρυβο (Signal to Noise Ratio, SNR)

$$SNR_{(dB)} = 20 \log_{10} \frac{S_{RMS}}{N_{RMS}} \quad (2.13)$$

όπου S_{RMS} και N_{RMS} είναι η μέση τετραγωνική τιμή του σήματος εισόδου και του θορύβου αντίστοιχα. Αυτή η σχέση δίνει τιμές σε dB μικρότερες από αυτές της δυναμικής περιοχής (βλ. πχ. Σχήμα 2.6), γεγονός που οφείλεται στο ότι η μέση τετραγωνική τιμή ενός σήματος είναι πάντα μικρότερη από την μέγιστη απόλυτη τιμή της συνάρτησης $s(t)$ κάτι που μαθηματικά μπορούμε να το παραστήσουμε ως

$$S_{RMS} < \max \{|s(t)|\} \quad (2.14)$$

Για παράδειγμα, για ένα ημιτονικό σήμα $s(t) = V_0 \sin \omega t$, το πλάτος σχετίζεται με τη μέση τετραγωνική τιμή μέσω της σχέσης

$$S_{RMS} = \frac{V_0}{\sqrt{2}} \quad (2.15)$$

κάτι που φαίνεται και από το Σχήμα 2.7.

Ένα μέγεθος που έχει σημασία είναι ο λόγος της μέγιστης στιγμιαίας τιμής προς τη μέση τετραγωνική τιμή του σήματος $\frac{\max \{|s(t)|\}}{S_{RMS}}$. Ο λόγος αυτός εκφράζει ένα μέγεθος το

οποίο σχετίζεται με τη μεταβλητότητα του σήματος, το πόσο δηλαδή μεγαλύτερες είναι οι μέγιστες στιγμιαίες τιμές του πλάτους από τη μέση τετραγωνική του τιμή. Η αναμενόμενη τιμή του συγκεκριμένου λόγου για ηχητικά σήματα ομιλίας είναι γύρω στο 8:1 ενώ για

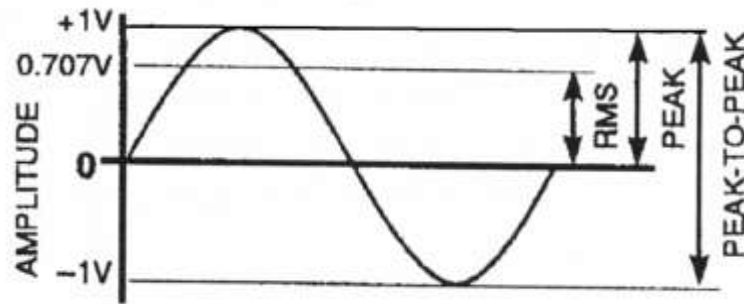
² Ανάλογα ορίζεται και η μέση τετραγωνική ισχύς ενός σήματος, η οποία είναι ίση με την ποσότητα S_{RMS}^2 , δηλαδή τη μέση τετραγωνική τιμή εις το τετράγωνο.

κλασσική μουσική γύρω στο 30:1.³ Φαίνεται λοιπόν ότι η διαφορά στη στατιστική μεταξύ των ηχητικών σημάτων αντανακλάται σε διαφορετική σχέση μεταξύ δυναμικής περιοχής και μέσης τετραγωνικής τιμής. Για παράδειγμα, η παραδοχή για σήμα ομιλίας μας υπαγορεύει ότι η δυναμική περιοχή του συστήματος θα πρέπει να είναι τέτοια ώστε $(\Delta\Pi)=8S_{RMS}$ ενώ για σήμα κλασσικής μουσικής θα πρέπει $(\Delta\Pi)=30S_{RMS}$ κοκ. Η διαφοροποίηση αυτή μας αναγκάζει να θεωρήσουμε ένα μέγεθος το οποίο το ονομάζουμε headroom του οποίου η τιμή σε dB είναι τέτοια ώστε $(\Delta\Pi)=SNR+Headroom$ (Βλ. Σχήμα 2.6.)

Αν λοιπόν ξέρουμε τη φύση του σήματος εισόδου, το headroom είναι ένα μέγεθος που μπορούμε να το ορίζουμε εμείς ώστε να ικανοποιείται η σχέση

$$Headroom = 20 \log_{10} \left(\frac{\max \{|s(t)|\}}{S_{RMS}} \right) \quad (2.16)$$

Ο διαχωρισμός της μέγιστης απόλυτης στιγμιαίας τιμής από τη μέση τετραγωνική τιμή του σήματος διαφαίνεται και στα χαρακτηριστικά των επαγγελματικών ηχητικών συστημάτων τα οποία ορίζουν το μέγιστο πλάτος τάσης του συστήματος (peak level) και τη νόμιμη στάθμη λειτουργίας (nominal level, RMS level). Σε ένα επαγγελματικό αναλογικό ηχητικό σύστημα η τιμή του headroom είναι της τάξης των 20dB. Αυτό το headroom επιτρέπει στα μέγιστα στιγμιαία πλάτη του σήματος να διατηρούν ένα λόγο 10:1 σε σχέση με τη μέση τετραγωνική του τιμή, κάτι που είναι απαραίτητο για μουσικά σήματα.



Σχήμα 2.7 Σχέση μεταξύ μέγιστης στιγμιαίας τιμής (peak), πλάτους από κορυφή σε κορυφή (peak to peak, p-p) και μέσης τετραγωνικής τιμής (RMS) για ένα ημιτονικό σήμα. $\left(\frac{1}{\sqrt{2}} = 2^{-1/2} = 0.707\right)$

Παράδειγμα: Ένα αναλογικό σύστημα ήχου λειτουργεί σε στάθμες 10Vp-p. Ποια η νόμιμη στάθμη λειτουργίας σε Volt δεδομένου ότι χρειαζόμαστε headroom 20dB;

Καταρχάς, παρατηρούμε ότι τα 10 V ορίζονται από κορυφή σε κορυφή (p-p) που σημαίνει ότι το σήμα μας θα κινείται μεταξύ ± 5 V (το πλάτος θα είναι δηλαδή 5V). Από τη σχέση (2.16) βλέπουμε ότι $20 \log_{10} \frac{5}{V_{RMS}} = 20\text{dB}$ επομένως, $\frac{5}{V_{RMS}} = 10^{20/20} = 10$ και τελικά $V_{RMS}=0.5$ V. Δηλαδή, η νόμιμη στάθμη λειτουργίας είναι ίση με το 10% της μέγιστης απόλυτης στάθμης.

³ Τα ρεαλιστικά σήματα όπως τα σήματα ομιλίας και μουσικής ακολουθούν θεωρητικά την κανονική κατανομή. Στην ουσία η παραδοχή είναι ότι η πιθανότητα τα δείγματα να πέσουν πέρα από την περιοχή που ορίζουν οι λόγοι είναι μικρότερη από ένα πολύ μικρό αριθμό.

2.4.2 Δυναμική περιογή και SNR σε ψηφιακά ηχητικά συστήματα

Ο παράγοντας του θορύβου σε ένα ψηφιακό κύκλωμα είναι το σφάλμα κβαντισμού $e(n) = \hat{s}(n) - s(n)$. Το σφάλμα αυτό χαρακτηρίζεται από μια μέση τετραγωνική τιμή E_{RMS} η οποία πρέπει να καθοριστεί με βάση τις παραδοχές που έγιναν στην παράγραφο 2.2.2. Για τον υπολογισμό του SNR χρειάζεται βέβαια και ο καθορισμός του αριθμητή της σχέσης (1.4.2), η οποία φυσικά εκφράζει τη μέση τετραγωνική τιμή S_{RMS} του σήματος εισόδου $s(n)$.

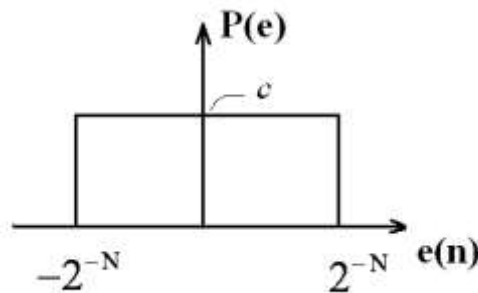
Υπολογισμός του E_{RMS} : Κατ' αρχάς παρατηρούμε ότι το $e(n)$ θα είναι μέγιστο όταν η πραγματική τιμή του αναλογικού στιγμιότυπου $s(n)$ πέφτει ακριβώς ανάμεσα σε δύο διαδοχικούς σταθμούς κβαντισμού. Δεδομένου ότι η απόσταση μεταξύ δύο σταθμών κβαντισμού είναι ίση με Δ , φαίνεται ότι το $e(n)$ θα κινείται μεταξύ $-\Delta/2$ και $\Delta/2$, δηλαδή

$$-\frac{\Delta}{2} \leq e(n) \leq \frac{\Delta}{2} \quad (2.17)$$

Θεωρώντας ότι το σήμα εισόδου είναι κανονικοποιημένο μεταξύ -1 και 1 , μπορώ να αντικαταστήσω το Δ σύμφωνα με τη σχέση (2.17) και επομένως θα έχω

$$-2^{-N} \leq e(n) \leq 2^{-N} \quad (2.18)$$

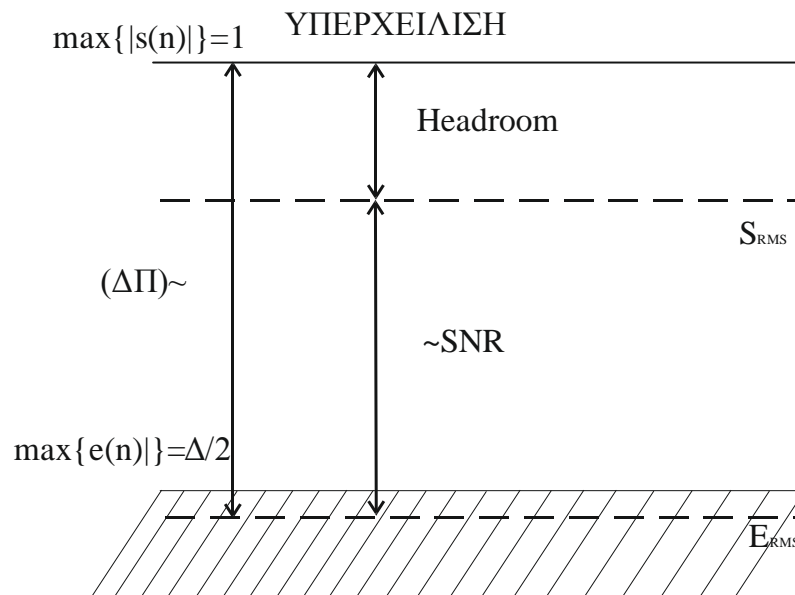
Όταν το σήμα εισόδου κινείται σε φυσιολογικές στάθμες, η μορφή του σφάλματος κβαντισμού είναι τέτοια ώστε να θεωρήσουμε ότι σε αυτό το εύρος το $e(n)$ παίρνει όλες τις δυνατές τιμές με την ίδια πιθανότητα. Με αυτά σαν δεδομένα σχεδιάζουμε το Σχήμα 2.8 όπου φαίνεται το διάγραμμα κατανομής πυκνότητας πιθανότητας του $e(n)$. Ο κατακόρυφος άξονας εκφράζει πιθανότητα και ο οριζόντιος πλάτος. Είναι προφανές ότι η τιμή του $P(e)$ είναι σταθερή, δηλαδή $P(e) = c$. Υπενθυμίζεται ότι το εμβαδό του ορθογωνίου που σχηματίζεται από το Σχήμα 2.8 θα πρέπει να είναι ίσο με 1. Ζητείται από τον αναγνώστη να αποδείξει ότι η τιμή του c είναι 2^{N-1} .



Σχήμα 2.8 Διάγραμμα κατανομής πιθανότητας του σφάλματος κβαντοποίησης.

Για τη δεδομένη κατανομή πιθανότητας του $e(n)$ η RMS τιμή του μπορεί εύκολα να υπολογιστεί με κατάλληλη ολοκλήρωση ότι είναι ίση με

$$E_{RMS} = \frac{\Delta}{2\sqrt{3}} = \frac{2^{-N}}{\sqrt{3}} \quad (2.19)$$



Σχήμα 2.9 Δυναμική περιοχή και SNR σε ένα ψηφιακό κύκλωμα με σήμα εισόδου $s(n)$ και σφάλμα κβαντοποίησης $e(n)$.

Ορισμός του S_{RMS} : Για το σήμα εισόδου $s(n)$ θα θεωρήσουμε ότι είναι ένα ημιτονικό σήμα και ότι αυτό το σήμα εισέρχεται στο ψηφιακό κύκλωμα αξιοποιώντας όλη τη δυναμική περιοχή του συστήματος. Αυτό σημαίνει ότι το πλάτος του είναι ίσο με $V_0=1$ και άρα η RMS τιμή του θα είναι ίση με $S_{RMS}=1/\sqrt{2}$. Αντικαθιστώντας αυτήν την τιμή στην (2.13) και λαμβάνοντας υπόψιν της έκφραση του E_{RMS} μπορούμε να δούμε ότι

$$SNR(dB) = 20 \log_{10} \frac{1/\sqrt{2}}{2^{-N}/\sqrt{3}} = 20 \log_{10} 2^N + 20 \log_{10} \frac{\sqrt{3}}{\sqrt{2}} = 6,02N + 1,76 \text{ dB} \quad (2.20)$$

Από τη σχέση αυτή φαίνεται λοιπόν ότι η προσθήκη κάθε ενός μπιτ (δηλαδή η μετάβαση από το N στο $N+1$) έχει ως αποτέλεσμα την αύξηση του SNR κατά 6 dB, που σε γραμμική κλίμακα αντιστοιχεί σε διπλασιασμό του SNR ($20 \log_{10} 2 = 6,02$). Με άλλα λόγια, αύξηση της τάξης του κβαντιστή οδηγεί σε υποδιπλασιασμό της στάθμης του θορύβου και επομένως σε διπλασιασμό της δυναμικής περιοχής.

Στο Σχήμα 2.9 απεικονίζεται η σχέση μεταξύ Δυναμικής περιοχής (ΔΠ), Headroom και SNR σε ένα ψηφιακό σύστημα.

Στον παρακάτω πίνακα φαίνεται ο λόγος σήματος προς θόρυβο για διαφορές τιμές της τάξης του κβαντιστή. Η θεώρηση είναι για ημιτονικό σήμα.

Τάξη N (bits)	SNR (dB)
8	49,8
12	73,8
16	97,8
18	109,8

Πίνακας.2.1 SNR σε σχέση με την ευκρίνεια κβαντισμού για ημιτονικό σήμα εισόδου.

Όσον αφορά το σταθερό όρο 1,76 στη σχέση (2.20) είναι ένας αριθμός που εξαρτάται από τις παραδοχές για τα στατιστικά χαρακτηριστικά του σφάλματος κβαντοποίησης και του σήματος εισόδου. Αν αλλάξουν αυτές οι παραδοχές, θα αλλάξει και η σχέση σήματος προς θόρυβο. Για τις υποθέσεις που κάναμε για το σήμα εισόδου βλέπουμε ότι χρειάζεται headroom ίσο με $20\log_{10}\sqrt{2}=3\text{dB}$. Για πιο ρεαλιστικά σήματα εισόδου, σαν αυτά που συζητηθήκαν προηγουμένως, το headroom θα πρέπει να αυξηθεί και κατά συνέπεια, το SNR θα μειωθεί ανάλογα.

Παράδειγμα: Θέλουμε να ηχογραφήσουμε και να μεταδώσουμε στερεοφωνικό ψηφιακό σήμα με συχνότητα δειγματοληψίας $F_s=48\text{kHz}$. Το μέγιστο bit rate που μας επιτρέπει το κανάλι μετάδοσης είναι 1940kbps. Πόσο το μέγιστο SNR που μπορώ να έχω για ημιτονικό σήμα δεδομένου ότι έχω κβαντιστή μεταβλητής τάξης N ; Πόσο το μέγιστο SNR υπό την παραδοχή ότι το σήμα είναι κλασσικής μουσικής (θεωρήστε ότι ο λόγος της μέγιστης απόλυτης τιμής προς τη μέση τετραγωνική τιμή του σήματος για κλασσική μουσική είναι ίσος με 30:1).

Αν N η μέγιστη τάξη του κβαντιστή για τα παραπάνω δεδομένα θα ισχύει: $48000 \times 2 \times N = 1940000$. Άρα, $N = 1940000 / 96000 = 20,21$. Επειδή η τάξη του κβαντιστή πρέπει να είναι ακέραιος αριθμός συμπεραίνουμε ότι $N=20$. Για ημιτονικό σήμα, από τη σχέση (2.20) βλέπουμε ότι $\text{SNR} = 6,02N + 1,76 = 122,16\text{dB}$.

Για σήμα κλασσικής μουσικής βλέπουμε ότι χρειαζόμαστε $\text{headroom} = 20\log_{10}30 = 29,54\text{dB}$. Δεδομένου ότι για ημιτονικό σήμα το headroom είναι μόνο 3 dB, αυτό σημαίνει ότι χρειαζόμαστε $29,54 - 3 = 26,54$ dB περισσότερο headroom, άρα η τιμή του SNR θα μειωθεί ανάλογα $\text{SNR} = 122,16 - 26,54 = 96,52$ dB.

Παράδειγμα: Για μουσική ικανοποιητικής πιστότητας απαιτούμε $\text{SNR} = 90$ dB. Αν η συχνότητα δειγματοληψίας είναι $F_s = 44,1\text{kHz}$, προσδιορίστε τη μνήμη που απαιτείται για διαφύλαξη ενός, μέσης διάρκειας 3 min στερεοφωνικού μουσικού κομματιού. Θεωρήστε ένα headroom 10dB.

Για ένα headroom ίσο με 10dB, σημαίνει ότι θα έχω 7dB λιγότερα από ότι μου υπαγορεύει η σχέση(2.20). Θα είναι επομένως $\text{SNR} = 6,02N + 1,76 - 7 = 6,02N - 5,24$ και επομένως

$$N = \frac{90 + 5,24}{6,02} = 15,82$$
. Άρα το N θα πρέπει να είναι ίσο με 16. Για τα τρία λεπτά (180 sec) θα

χρειαζόμαστε επομένως

Μνήμη = $44100 \cdot 16 \cdot 2 \cdot 180 = 254016000 = 254016\text{kbits}$

2.4.3 Ορισμός του dB-FS (Full Scale)

Συχνά ο ορισμός του πλάτους και της μέσης τετραγωνικής τιμής ενός ψηφιακού σήματος μας δίνεται εκφρασμένο σε dB-FS. Το μέγεθος αυτό ορίζεται με τέτοιο τρόπο ώστε η τιμή 0 dB-FS να αντιστοιχεί στο μέγιστο δυνατό πλάτος του σήματος, όταν δηλαδή το σήμα αξιοποιεί όλο το δυναμικό εύρος του συστήματος. Για παράδειγμα, μια ημιτονική κυματομορφή χαρακτηρίζεται πλήρους πλάτους (full scale) όταν το πλάτος V_0 είναι ίσο με

1, δεδομένου ότι χρησιμοποιούμε κανονικοποιημένη κλίμακα. Από εκεί και πέρα, ισχύουν όλες οι ιδιότητες των λογαρίθμων και οι χαρακτηριστικές τιμές της συνάρτησης $f(x)=20\log_{10}x$ που αναφέρονται στο τέλος των σημειώσεων.

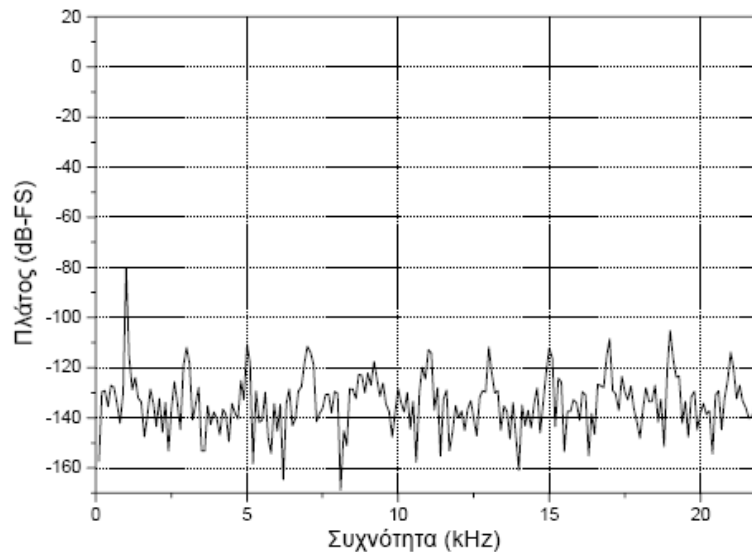
Παράδειγμα: Η είσοδος μιας ψηφιακής κάρτας ήχου λειτουργεί για στάθμες αναλογικού σήματος $\pm 5V$. Εκφράστε σε dB-FS τις τιμές του πλάτους και της μέσης τετραγωνικής τιμής του ψηφιακού σήματος που θα προκύψει, θεωρώντας ότι στην είσοδο έχουμε ημιτονικό σήμα πλάτους 0.5 V. Από πόσους σταθμούς κβαντισμού διέρχεται το ψηφιοποιημένο ημιτονικό σήμα αν $N=16$ bits;

Για το πλάτος θα ισχύει προφανώς $20\log_{10}(0,5/5)=20\log_{10}(1/10)=-20\text{dB-FS}$ και η RMS τιμή του σήματος είναι $-20-3 = -23\text{dB-FS}$. Το ημιτονικό σήμα αξιοποιεί το 1/10 του δυναμικού εύρους του κβαντιστή, άρα διέρχεται από $2^{16}/10 = 6554$ σταθμούς κβαντισμού.

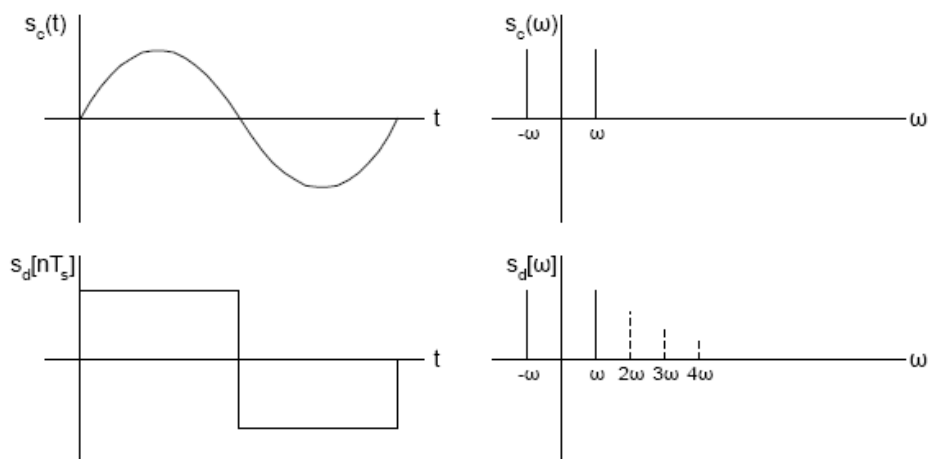
2.5 Dither

Στο σημείο αυτό πρέπει να τονιστεί ότι σε αντίθεση με τα αναλογικά συστήματα, τα προβλήματα μετατροπής παρουσιάζονται στις χαμηλές στάθμες σήματος, όπως π.χ. στα ήσυχα περάσματα της μουσικής ή στο τέλος της διάρκειας μιας μουσικής νότας, καθώς όσο μικραίνει το πλάτος του σήματος εισόδου, τόσο το σφάλμα κβαντισμού δεν έχει την θεωρητική μορφή του λευκού θορύβου, αλλά αποκτά συσχετισμένο προς το σήμα εισόδου χαρακτήρα (δηλ. μορφή αρμονικής παραμόρφωσης) και η παραπάνω ανάλυση δεν ισχύει (Σχήμα 2.10). Μάλιστα, συχνά συμβαίνει το πλάτος του σήματος να είναι μικρότερο της ελάχιστης στάθμης κβαντισμού του μετατροπέα. Στις περιπτώσεις αυτές ενδέχεται να εμφανιστούν και άλλα ανεπιθύμητα φαινόμενα παραμορφώσεων (Σχήμα 2.11). Όπως φαίνεται στο παρακάτω Σχήμα, ελλιπής κβαντισμός του σήματος προκαλεί παραμόρφωση με αποτέλεσμα το φάσμα του παραγόμενου σήματος να εμφανίζει νέες αρμονικές που ενδέχεται να μην ικανοποιούν την συνθήκη δειγματοληψίας και να παράγουν αναδιπλούμενες συνιστώσες (aliasing). Αυτή η μορφή αρμονικής παραμόρφωσης σε χαμηλής στάθμης σήματα, προσδίδει συχνά τη χαρακτηριστική τραχύτητα στην ακουστική χροιά των ήχων που αναπαράγονται από ψηφιακά συστήματα, κυρίως χαμηλής ποιότητας/τιμής (π.χ. κάρτες ήχου για PC με ευκρίνεια κβαντισμού 8 bit). Ο ελλιπής κβαντισμός περιορίζει επίσης την ακουστική ευκρίνεια στην αναπαραγωγή μουσικών ήχων - κάτι που δεν είναι σύνηθες σε αντίστοιχα αναλογικά συστήματα. Συμπερασματικά, ο θόρυβος που παράγουν τα ψηφιακά ηχητικά συστήματα εμφανίζεται κατά το στάδιο του κβαντισμού (ή της αναπαραγωγής του κβαντισμένου σήματος) και έχει ιδιάζοντα και ανεπιθύμητο χαρακτήρα. Για τη λύση των προβλημάτων αυτών έχουν προταθεί κατά καιρούς τεχνικές που βασίζονται στη χρήση μη γραμμικών floating point μετατροπέων, που όμως δεν τυγχάνουν ευρείας χρήσης για μουσικά σήματα. Προφανώς, θα ήταν χρήσιμο αν η τάξη N (bits) του κβαντιστή αυξηθεί πάνω από τα 16 bit, αλλά αυτό, όπως θα συζητηθεί αργότερα, δεν ήταν τεχνικά εφικτό την δεκαετία του '80 όταν ορίστηκαν οι προδιαγραφές του συστήματος CD. Οι περιορισμοί αυτοί πλέον δεν ισχύουν και έτσι στο σύστημα DVD

(και σε άλλα ψηφιακά ηχητικά συστήματα) είναι δυνατή η χρήση σήματος κβαντισμένου με ευκρίνεια μέχρι και 24 bit. Στην περίπτωση συστημάτων 16 bit, μια λύση που ακολουθείται κατά την αρχική Α/Ψ μετατροπή του μουσικού σήματος (ηχογράφηση) βασίζεται στην προσθήκη θορύβου χαμηλής στάθμης (dither), που έχει τιμές πλάτους συγκρίσιμες της στάθμης κβαντισμού Δ και η επίδρασή του θα συζητηθεί εκτενώς σε αυτό το κεφάλαιο. Το αποτέλεσμα της προσθήκης αυτής όπως θα δούμε, αναιρεί πολλά από τα προβλήματα που συζητήθηκαν παραπάνω, έτσι ώστε να εμφανίζεται ένα περισσότερο ικανοποιητικό ακουστικό αποτέλεσμα, αυξάνει όμως τη στάθμη του θορύβου, ο οποίος έχει πλέον τον επιθυμητό χαρακτήρα του λευκού θορύβου.



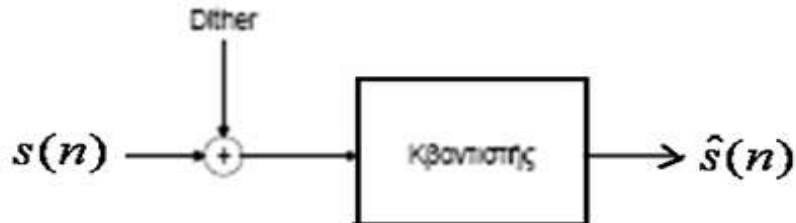
Σχήμα 2.10: Μορφή φάσματος ημιτονικού σήματος πλάτους -80dB-FS , συχνότητας 1kHz, $N=16$ bit



Σχήμα 2.11: Επίδραση του κβαντισμού στο φασματικό περιεχόμενο σημάτων χαμηλής στάθμης

2.5.1 Κβαντισμός και Dither

Το Dither είναι ουσιαστικά μια διαδικασία κατά την οποία εισάγουμε στο σήμα μας θόρυβο κατάλληλου πλάτους και στατιστικών χαρακτηριστικών. Το πλάτος αυτό συνήθως είναι της τάξης του $\frac{1}{2}$ και 1LSB, και ο θόρυβος έχει συνήθως χαρακτηριστικά λευκού θορύβου (παρόμοια δηλαδή με την αρχική υπόθεση που κάναμε για το σφάλμα κβαντισμού).



Σχήμα 2.12: Προσθήκη θορύβου Dither κατά την Α/Ψ μετατροπή

Τα οφέλη από τη χρήση Dither φαίνονται στα διαγράμματα του σχήματος 1.13. Στην αριστερή στήλη φαίνονται τα χαρακτηριστικά του σήματος εισόδου χωρίς τη χρήση Dither και στη δεξιά με τη χρήση του Dither. Σχήμα 2.13 Α: ένα ημιτονικό σήμα συχνότητας 1 kHz εισέρχεται στον κβαντιστή με πλάτος ίσα ίσα 2 LSB. Με το που θα εισέλθει στον κβαντιστή θα πάρει τη μορφή που φαίνεται στο διάγραμμα Β. Στο διάγραμμα C φαίνεται το σφάλμα κβαντισμού και στο D απεικονίζεται η φασματική πυκνότητα του ψηφιακού σήματος (είναι ένα διάγραμμα έντασης ανά συχνότητα). Φαίνεται λοιπόν ότι όσο λιγοστεύουν οι σταθμοί κβαντισμού από όπου διέρχεται το σήμα, αυτό παίρνει μια μορφή που μοιάζει με τετραγωνικό παλμό, γεγονός που έχει ως αποτέλεσμα να εμφανίζονται συχνότητες αρμονικές της αρχικής. Στο σχήμα D βλέπουμε ότι ναι μεν υπερिशύει η αρχική συχνότητα του 1 kHz αλλά εμφανίζονται και αρμονικές στα 3, 5, 7, 9 kHz και πάει λέγοντας. Τα ίδια μεγέθη με τη χρήση Dither πλάτους 2 LSB p-p (από κορυφή σε κορυφή) φαίνονται τώρα στη δεξιά στήλη. Βλέπουμε λοιπόν ότι ο θόρυβος από το Dither υπερτίθεται πάνω στην ημιτονική κυματομορφή αλλάζοντας την απόκριση του κβαντιστή στο διάγραμμα F. Το σφάλμα κβαντισμού έχει αποκτήσει και αυτό διαφορετική μορφή (διάγραμμα G) και από τη φασματική πυκνότητα του ψηφιακού σήματος στο διάγραμμα Η φαίνεται ότι οι αρμονικές συχνότητες έχουν εξουδετερωθεί. Στη θέση τους βλέπουμε ένα ομοιόμορφα κατανομημένο λευκό θόρυβο, ο οποίος είναι λιγότερο ενοχλητικός από την περίπτωση με τις δυνατές αρμονικές συνιστώσες.

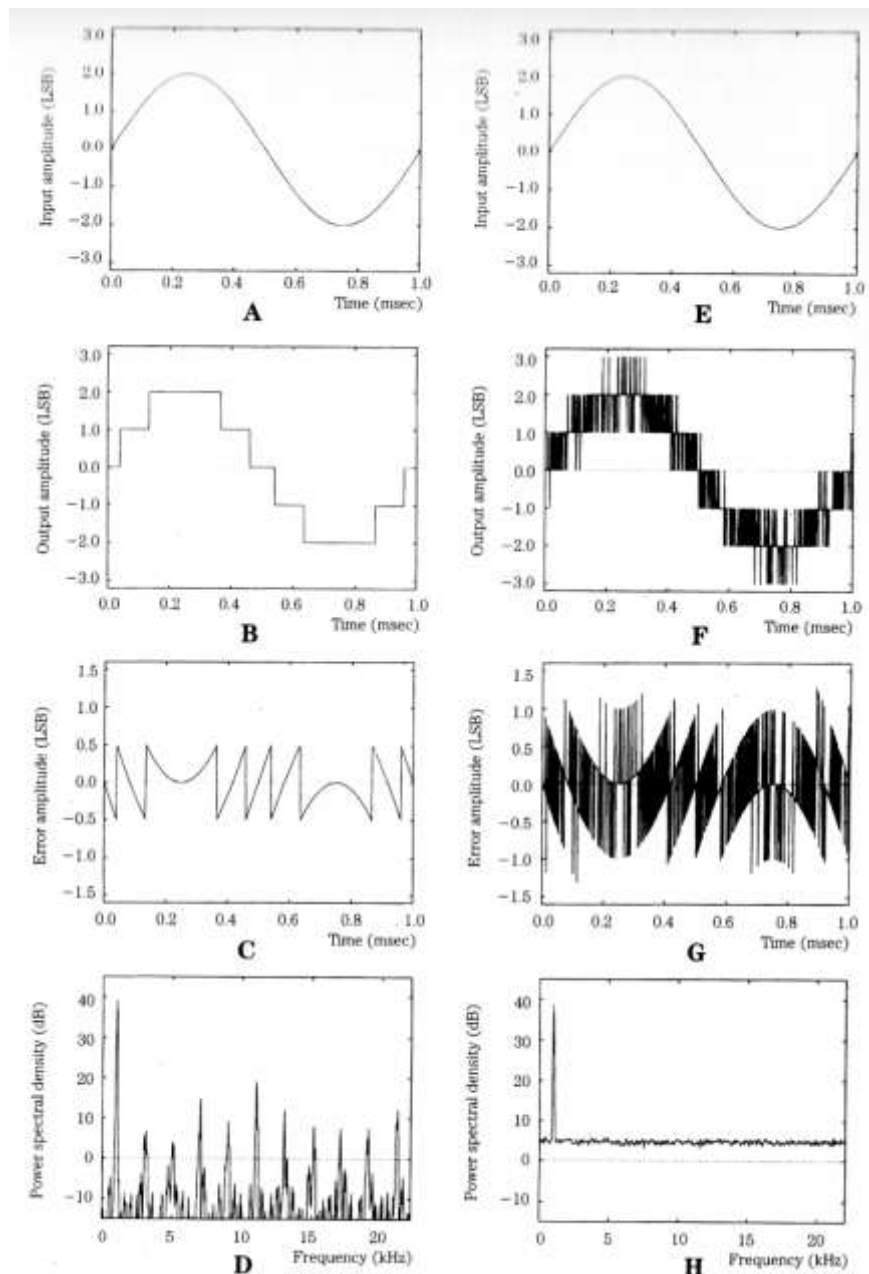
Με το Dither καταφέρνουμε λοιπόν να βελτιώνουμε υποκειμενικά τη χροιά του ψηφιακού σήματος. Η προσθήκη του dither στην είσοδο ενός κβαντιστή μπορούμε λοιπόν να πούμε ότι προκαλεί τις εξής σημαντικές βελτιώσεις στη συνολική λειτουργία του:

(α) αλλάζει την μορφή του σφάλματος κβαντισμού και το καθιστά ανεξάρτητο από το σήμα εισόδου,

(β) αφαιρεί κάθε είδους αρμονικής παραμόρφωσης η οποία εμφανίζεται κατά τη μετατροπή σημάτων πολύ μικρού πλάτους, δίνοντας στον θόρυβο κβαντισμού χαρακτηριστικά όμοια με τα χαρακτηριστικά λευκού θορύβου και

(γ) βελτιώνει τη διακριτική ικανότητα του κβαντιστή, αυξάνοντας κατά υποκειμενικό (κι

όχι ποσοτικό) τρόπο τη δυναμική του περιοχή. 0,5

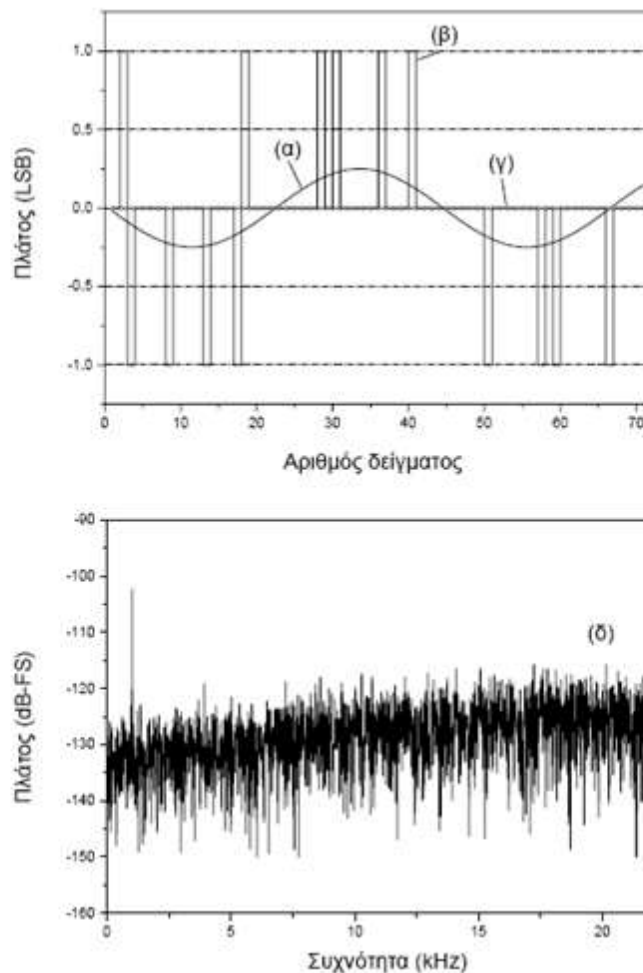


Σχήμα 2.13: Επίδραση της προσθήκης Dither στη μορφή και τη φασματική πυκνότητα του ψηφιακού σήματος.

Ο λόγος που λέμε κατά υποκειμενικό και όχι κατά ποσοτικό έχει να κάνει με το ότι με την προσθήκη Dither το επίπεδο του θορύβου αυξάνεται, δε μειώνεται. Πράγματι, αν υπολογίσουμε τη μέση τετραγωνική ισχύ του σφάλματος κβαντισμού χωρίς Dither θα δούμε ότι αυτή κυμαίνεται στο διάγραμμα D στα 0 dB περίπου, ενώ στο διάγραμμα H η στάθμη του θορύβου είναι αυξημένη κατά 5 dB (σχεδόν διπλάσια). Το πλεονέκτημα λοιπόν του

Dither είναι ότι αλλάζει τη μορφή του σφάλματος κβαντισμού, όχι ότι μειώνει το επίπεδο του θορύβου. Το Dither αυξάνει το επίπεδο θορύβου του κβαντιστή, και αυτή αύξηση είναι ανάλογη του πλάτους του Dither που χρησιμοποιείται (πχ αν είναι 1LSB p - p , 2 LSB p - p κλπ).

Παράδειγμα: Η επίδραση του dither στη συνολική απόδοση ενός κβαντιστή μπορεί εύκολα να παρουσιασθεί εάν θεωρήσουμε έναν 16bit κβαντιστή με dither πλάτους 2 LSB p - p . Στην είσοδο του κβαντιστή εφαρμόζουμε ένα ημιτονοειδές σήμα πλάτους LSB/4 και συχνότητας 1kHz (Σχήμα 2.14(α)). Όπως φαίνεται από το σχήμα αυτό, ένας απλός κβαντιστής (χωρίς τη χρήση dither) δεν μπορεί να κωδικοποιήσει ένα τόσο μικρού πλάτους σήμα, και επομένως θα έδινε μηδενική έξοδο. Αντίθετα, στην περίπτωση κβαντιστή με dither, οι νέες τιμές του σήματος εισόδου που εφαρμόζεται στον κβαντιστή (αρχικό σήμα + dither – Σχήμα 2.14(β)), υποχρεώνει την έξοδο του κβαντιστή να είναι μη μηδενική σε κάποιες περιόδους δειγματοληψίας (Σχήμα 2.14(γ)). Λόγω των στατιστικών χαρακτηριστικών του dither, η έξοδος του κβαντιστή περιέχει σε μεγάλο βαθμό την πληροφορία του αρχικού σήματος (Σχήμα 2.14(δ)).

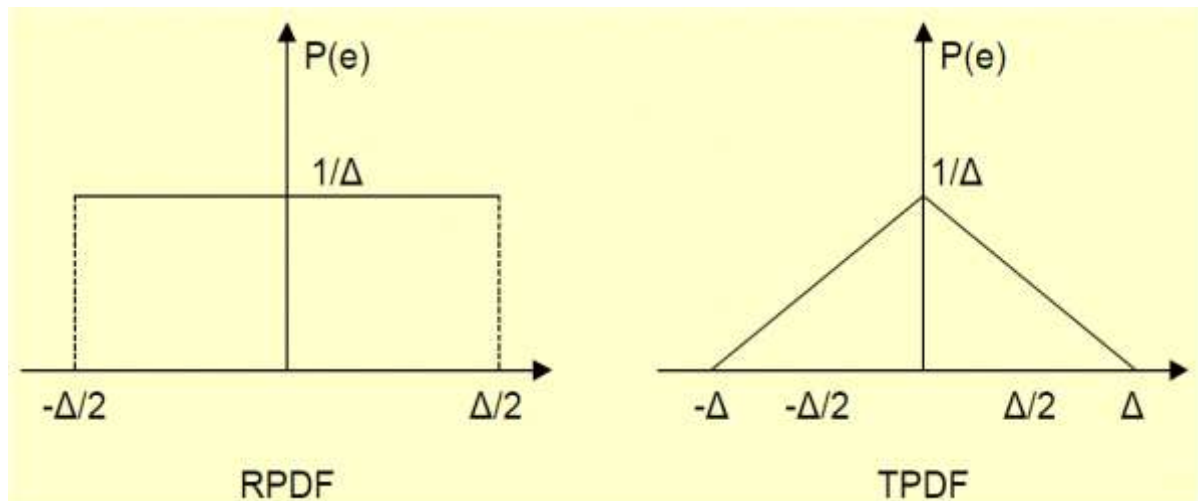


Σχήμα 2.14: Παράδειγμα επίδρασης του θορύβου dither σε ημιτονικό σήμα 1kHz πλάτους -102.3dB-FS

Από το προηγούμενο παράδειγμα είναι ξεκάθαρο ότι η χρήση του dither επιβάλλεται κατά την Α/Ψ μετατροπή ενός ηχητικού σήματος. Έτσι, τα περισσότερα ψηφιακά συστήματα λειτουργούν σε συχνότητα δειγματοληψίας 44,1kHz και χρησιμοποιούν γραμμικό κβαντισμό τουλάχιστο 16bit με προσθήκη θορύβου dither κατά την ηχογράφιση. Τα συστήματα αυτά είναι δυνατόν να επιτύχουν ικανοποιητική απόδοση που όμως εξαρτάται σε μεγάλο βαθμό από τις τεχνικές προδιαγραφές των Α/Ψ και Ψ/Α μετατροπέων που χρησιμοποιούνται.

2.5.2 Είδη Dither

Οι διάφορες κατηγορίες dither δημιουργούνται με κριτήριο τη συνάρτηση πυκνότητας πιθανότητας (Probability Density Function, PDF) του θορύβου dither, αλλά και του φασματικού τους περιεχομένου. Συγκεκριμένα, στα ψηφιακά συστήματα χρησιμοποιούνται διεξοδικά γεννήτριες dither ομοιόμορφης κατανομής (Rectangular PDF, RPDF) πλάτους 1LSB p-p (ή 1Δ p-p) και η τριγωνική κατανομή (Triangular PDF, TPDF) πλάτους 2LSB p-p (ή 2Δ p-p), οι οποίες φαίνονται στο Σχήμα 2.15. Η ομοιόμορφη κατανομή προσθέτει θόρυβο ισχύος $\Delta^2/12$, οπότε η συνολική ισχύς μαζί με τον θόρυβο κβαντισμού είναι $2\Delta^2/12$. Η τριγωνική κατανομή υλοποιείται με το άθροισμα δύο ομοιόμορφων κατανομών, οπότε η συνολική ισχύς θορύβου στον κβαντιστή είναι $3\Delta^2/12$. Έχει αποδειχθεί ότι ο θόρυβος dither με TPDF κατανομή και πλάτος 2 LSB p-p παρουσιάζει την καλύτερη στατιστική συμπεριφορά και ελαχιστοποιεί τη συσχέτιση του θορύβου κβαντισμού με το σήμα. Το μειονέκτημα της εφαρμογής dither για μια TPDF είναι μία αύξηση του επιπέδου του θορύβου της τάξης των 6dB.

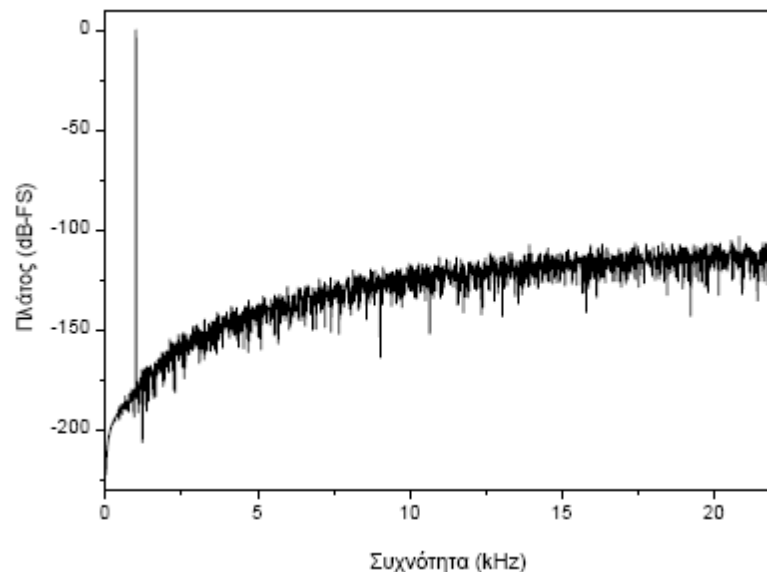


Σχήμα 2.15 Τυπικές μορφές συναρτήσεων πυκνότητας θορύβου dither RPDF πλάτους 1LSB p-p (δεξιά) TPDF πλάτους 2LSB p-p.

2.5.2 Μορφοποίηση θορύβου

Όπως αναφέρθηκε προηγουμένως, η χρήση dither έχει σαν αποτέλεσμα την αποδιαμόρφωση του σφάλματος κβαντισμού από το σήμα εισόδου και την μείωση της

αρμονικής παραμόρφωσης που δημιουργείται κατά τον κβαντισμό σημάτων μικρού πλάτους, με ταυτόχρονη αύξηση το επιπέδου του θορύβου κβαντισμού. Παρόλο που σε τυπικές ηχητικές εφαρμογές (κβαντισμός 16bit, συχνότητα δειγματοληψίας 44.1kHz) η στάθμη θορύβου είναι αρκετά χαμηλή κι έχει χαρακτηριστικά λευκού θορύβου, εντούτοις για σήματα χαμηλής έντασης, το επίπεδο του θορύβου ξεπερνά το ελάχιστο κατώφλι ακοής και ο θόρυβος κβαντισμού καθίσταται ακουστός. Η ακουστότητα του θορύβου αυξάνει λόγω της μεταβλητής ευαισθησίας του ανθρωπίνου ακουστικού οργάνου με τη συχνότητα, με αποτέλεσμα ο θόρυβος κβαντισμού να είναι λιγότερο ακουστός από τη συχνότητα των 4kHz και πάνω. Επίσης, κατά καιρούς έχουν εμφανιστεί και υλοποιήσεις φίλτρων με συντελεστές δεκαδικούς αριθμούς, τα οποία προσπαθούν να δώσουν στο θόρυβο κβαντισμό τη μορφή του ελαχίστου κατωφλίου ακοής (βλ. 'Εγχειρίδιο Ακουστικής' κεφ 3). Για το λόγο αυτό, σε τυπικές ηχητικές εφαρμογές, ο θόρυβος κβαντισμού μορφοποιείται κατάλληλα στο πεδίο της συχνότητας, χρησιμοποιώντας τεχνικές μορφοποίησης θορύβου (Noise-Shaping). Η τεχνικές αυτές υλοποιούνται μέσω ανατροφοδοτούμενων ψηφιακών φίλτρων που λειτουργούν ως κλάδοι ανατροφοδότησης γύρω από έναν κβαντιστή. Στο Σχήμα 2.16 φαίνεται η επίδραση που έχει η εφαρμογή της τεχνικής Noise-shaping κατά τον κβαντισμό μιας 0dB-FS ημιτονικής κυματομορφής 1kHz με χρήση κατάλληλου αναδρομικού φίλτρου. Με σωστή επιλογή λοιπόν του φίλτρου, μπορούμε να διαμορφώσουμε το θόρυβο και να μορφοποιήσουμε το επίπεδό του ώστε η ενέργεια του θορύβου να μετατοπιστεί προς τις υψηλές συχνότητες όπου γίνεται λιγότερο ακουστός.



Σχήμα 2.16: Φάσμα εξόδου κβαντιστή για είσοδο ημιτονικού σήματος πλάτους 0dB-FS, συχνότητας 1kHz και χρήση Noise-Shaping τεχνικής.

2.6 Υπερδειγματοληψία

Οι τεχνικές υπερδειγματοληψίας (oversampling) εμφανίστηκαν σε συστήματα στα τέλη της δεκαετίας του '80 και είναι δυνατόν να χρησιμοποιηθούν είτε στην Α/Ψ είτε στην Ψ/Α

μετατροπή, βασίζονται δε στην αύξηση της συχνότητας δειγματοληψίας (ή της ροής δεδομένων) κατά R φορές, όπου αν π.χ. $R=2$, τότε υλοποιείται $2x$ oversampling, αν $R=4$, $4x$ oversampling κλπ. Όπως θα εξηγηθεί παρακάτω, τα κύρια πλεονεκτήματα αυτής (της θεωρητικά περιττής) αύξησης είναι:

1. Η ελάττωση των παραμορφώσεων από το αναλογικό antialiasing φίλτρο εισόδου ή εξόδου του μετατροπέα.

2. Η ελάττωση του θορύβου σφάλματος κβαντισμού, κάτι που συμπληρωματικά διευκολύνεται με τη χρήση τεχνικών μορφοποίησης φάσματος του σφάλματος (Noise-Shaping).

3. Η δυνατότητα ελάττωσης της τάξης του κβαντιστή του μετατροπέα, χωρίς ταυτόχρονη ελάττωση της σχέσης SNR, κάτι που με τη χρήση ειδικών μεθόδων διαμόρφωσης του προς μετατροπή σήματος, επιτρέπει κβαντισμό χαμηλής τάξης μέχρι και $N = 1$ bit (μετατροπείς Σίγμα-Δέλτα).

4. Η απλούστευση της κατασκευής και ρύθμισης του μετατροπέα και η ελάττωση του κόστους κατασκευής του.

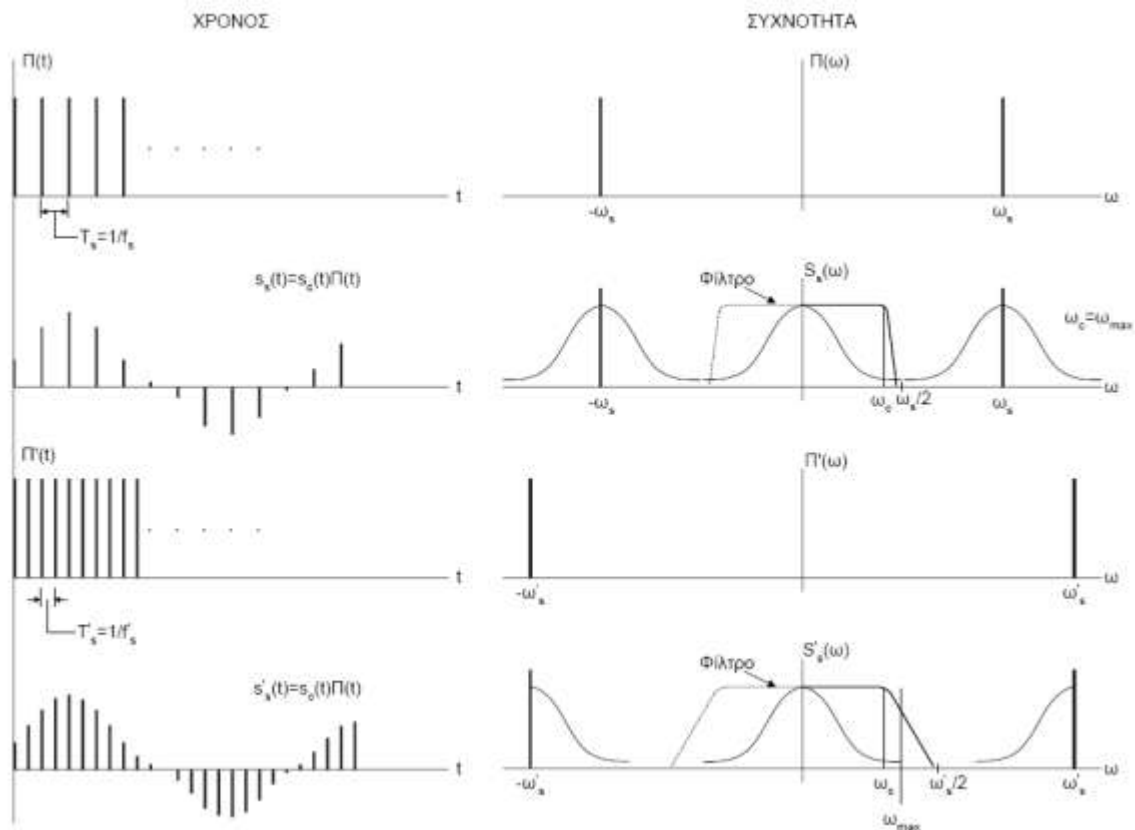
Για να εξηγηθούν τα παραπάνω, θα πρέπει να αναλυθεί η θεωρία της τεχνικής αυτής. Με υπερδειγματοληψία R -φορές, λαμβάνονται δείγματα του σήματος με μια νέα συχνότητα $f'_s = Rf_s$ (Hz), π.χ. εάν $R=4$ και $f_s=44,1\text{kHz}$, τότε $f'_s = 176,4\text{kHz}$. Έτσι η νέα μέγιστη συχνότητα η οποία είναι δυνατόν να εμφανίζεται στο «υπερδειγματοληπτούμενο» σήμα f'_{\max} θα είναι επίσης αυξημένη κατά R φορές, δηλαδή $f'_{\max} = Rf_{\max}$. Όπως φαίνεται και από το Σχήμα 2.17, η νέα αυτή συνθήκη επιτρέπει τη χρήση ομαλότερων (χαμηλότερης τάξης) αναλογικών anti-aliasing φίλτρων που δε χρειάζεται να έχουν τόσο υψηλό ρυθμό αποκοπής. Επίσης, βασιζόμενοι στην παραδοχή ο θόρυβος κβαντισμού είναι λευκός, τότε εμφανίζει και σταθερή φασματική πυκνότητα. Στο «υπερδειγματοληπτούμενο» σήμα το φάσμα καλύπτει R -φορές μεγαλύτερη περιοχή και άρα εμφανίζει ελάττωση του πλάτους του κατά R -φορές σε συγκεκριμένη συχνότητα, π.χ. αν $R=4$, τότε:

$$(\bar{e}')^2 = \frac{(\bar{e})^2}{4} \quad (2.21)$$

και επομένως

$$10 \log \frac{(\bar{e})^2}{(\bar{e}')^2} = -6\text{dB} \quad (2.22)$$

όπου $(\bar{e})^2$ και $(\bar{e}')^2$ είναι η πυκνότητα φασματικής ισχύος του σφάλματος κβαντισμού χωρίς και με υπερδειγματοληψία αντίστοιχα. Άρα σε αυτό το παράδειγμα εμφανίζεται μια θεωρητική βελτίωση της τάξης των 6 dB και γενικά, κάθε διπλασιασμός του R αυξάνει την τιμή SNR κατά 3dB περίπου. Επιπλέον, οι περισσότερες περιπτώσεις χρήσης υπερδειγματοληψίας συνδυάζονται με Noise-Shaping τεχνικές έτσι ώστε το σφάλμα κβαντισμού να είναι ελάχιστο στην περιοχή ακουστικών συχνοτήτων (20Hz-20kHz) και να αυξάνει στις υψηλότερες συχνότητες, οπότε τελικά, επιτυγχάνεται μεγαλύτερη ωφέλιμη ελάττωση στο θόρυβο κβαντισμού. (Σχήμα 2.18).



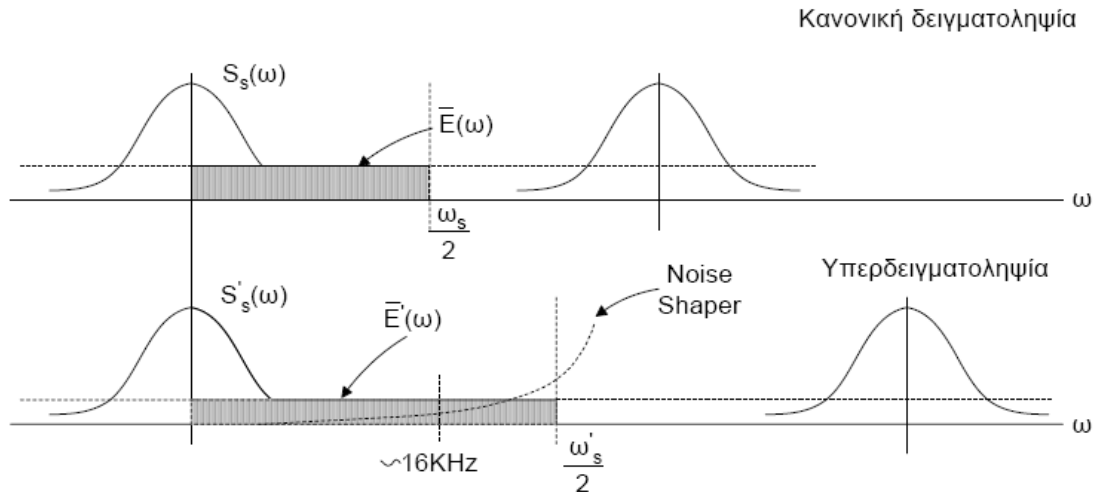
Σχήμα 2.17: Παράδειγμα εφαρμογής υπερδειγματοληψίας. Φαίνεται η δυνατότητα χρήσης φίλτρου αναλογικού φίλτρου με πιο ομαλή κλίση και έτσι αποφεύγονται πολλά από τα προβλήματα που εισάγονται από τη χρήση φίλτρων μεγάλου ρυθμού αποκοπής συχνότητων.

Το τρίτο πλεονέκτημα της τεχνικής υπερδειγματοληψίας είναι η αύξηση της «ενεργού τάξης κβαντισμού» ενός δεδομένου μετατροπέα. Έστω N η τάξη ενός κβαντιστή όπου πριν τον κβαντισμό εφαρμόζουμε δειγματοληψία παράγοντα R . Θα αποδείξουμε ότι η ποιότητα αναπαραγωγής είναι ισοδύναμη με ένα δεύτερο κβαντιστή τάξης $N' > N$ για τον οποίο δεν έχει γίνει χρήση υπερδειγματοληψίας. Αυτό προκύπτει από τη γνωστή σχέση

$$SNR_{(dB)} = 20 \log_{10} \frac{S_{RMS}}{N_{RMS}} = 6,02N + 1,76, \text{ όπου } N_{RMS} \text{ είναι η μέση τετραγωνική τιμή του}$$

σφάλματος κβαντισμού. Όπως είπαμε προηγουμένως, η υπερδειγματοληψία ελαττώνει την πυκνότητα φασματικής ισχύος του σφάλματος κβαντισμού με αποτέλεσμα η εντός ακουστών συχνοτήτων (20Hz - 20kHz) ενέργεια του σφάλματος κβαντισμού να ελαττώνεται. Αν λοιπόν N_{RMS} η χωρίς υπερδειγματοληψία RMS τιμή του σφάλματος κβαντισμού τότε κατά την εφαρμογή υπερδειγματοληψίας αντιστοιχεί μια νέα τιμή N'_{RMS} τέτοια ώστε $N'_{RMS} < N_{RMS}$. Από την προηγούμενη σχέση μπορούμε να θεωρήσουμε ότι αυτή η μείωση αντιστοιχεί σε μια αύξηση της τάξης κβαντισμού από N σε N' , τέτοια ώστε

$$N' - N = \Delta N = 0,5 \log_2 R \quad (2.23)$$



Σχήμα 2.18: Παράδειγμα εφαρμογής υπερδειγματοληψίας. Το σφάλμα κβαντισμού εξαπλώνεται συχνοτικά με αποτέλεσμα τη μείωση της φασματικής του πυκνότητας. Εν τέλει, η ενέργεια του σφάλματος κβαντισμού που είναι εντός του ακουστικού εύρους συχνοτήτων μειώνεται και αυτό ισοδυναμεί σε αύξηση του SNR. Τα οφέλη ενισχύονται από τη χρήση noise-shaping τεχνικών οι οποίες διαμορφώνουν το φάσμα του σφάλματος κβαντισμού ώστε να είναι λιγότερο αισθητός υποκειμενικά.

Από την εξίσωση (2.23) οδηγούμαστε σε μια νέα έκφραση του λόγου σήματος προς θόρυβο, λαμβάνοντας υπόψιν τη χρήση υπερδειγματοληψίας παράγοντα R

$$SNR = 6,02(N + 0,5 \log_2 R) + 1,76 \quad (2.24)$$

Από τη σχέση αυτή φαίνεται ότι διπλασιασμός της συχνότητας δειγματοληψίας αντιστοιχεί σε αύξηση της τάξης κβαντισμού κατά μισό bit. Στη λογική αυτή βασίζεται η τεχνολογία μετατροπέων οι οποίοι λειτουργούν με πολύ μικρές τάξεις κβαντισμού και μεγάλες συχνότητες δειγματοληψίας. Ένα χαρακτηριστικό παράδειγμα είναι οι μετατροπείς “Δέλτα” και “Σίγμα-Δέλτα” οι οποίοι επιτυγχάνουν ικανοποιητική ποιότητα αναλογικής σε ψηφιακή μετατροπή και αντίστροφα με διαμόρφωση σήματος μόνο σε 1 bit. Οι τεχνολογίες αυτές βασίζονται στη χρήση μεγάλου παράγοντα υπερδειγματοληψίας και σε βέλτιστες τεχνικές μορφοποίησης θορύβου (Noise Shaping).

Παράδειγμα: Να βρεθεί η συχνότητα δειγματοληψίας με την οποία ένας κβαντιστής τάξης $N=12$ bit επιτυγχάνει την ίδια ποιότητα μετατροπής με κβαντιστή τάξης $N'=16$ bit και συχνότητα δειγματοληψίας 44100 Hz.

Απάντηση: Η τάξη του διαθέσιμου κβαντιστή είναι κατά 4 bit μικρότερη από αυτή που θέλουμε να επιτύχουμε. Από τη σχέση (2.23) φαίνεται λοιπόν ότι πρέπει να χρησιμοποιηθεί παράγοντας δειγματοληψίας R τέτοιος ώστε $0.5 \log_2 R = 4$. Επομένως θα ισχύει $\log_2 R = 8$ και επομένως $R = 2^8 = 256$. Η συχνότητα δειγματοληψίας εν τέλει θα πρέπει να είναι ίση με $256 \cdot 44100 = 11,29$ MHz περίπου.

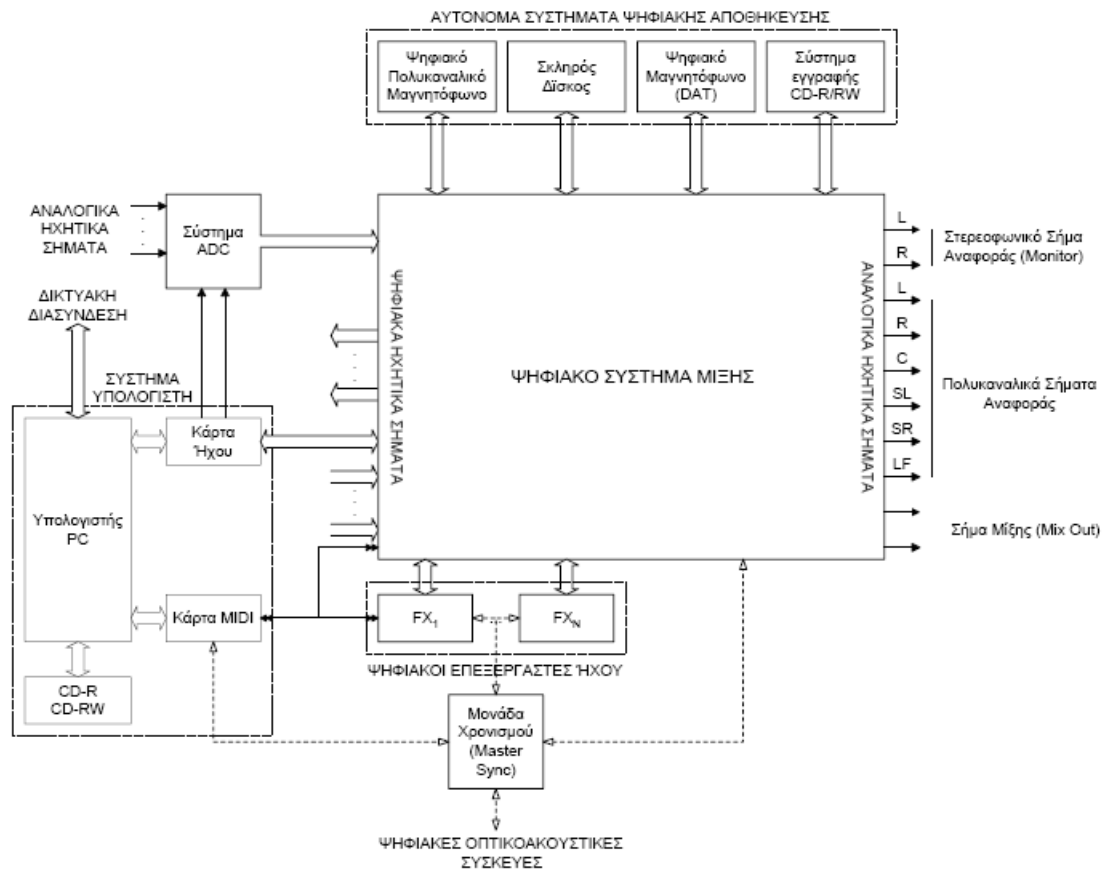
3. ΨΗΦΙΑΚΟ ΣΤΟΥΝΤΙΟ

3.1 Γενικές Δομές και Κατηγορίες Συστημάτων

Ο ψηφιακός ήχος μπορεί να αντιμετωπισθεί σαν μία σειρά από ψηφιακά δείγματα, κωδικοποιημένα με τους τρόπους που περιγράφηκαν, τα οποία, ανάλογα με τις συγκεκριμένες εφαρμογές, θα πρέπει να αποθηκευθούν, να υποστούν επεξεργασία (π.χ. τεμαχισμό, ισοστάθμιση, μίξη, κλπ) και να κωδικοποιηθούν πριν από την τελική τους διανομή στην μορφή που θα μπορεί να τα δεχθεί ο χρήστης (π.χ. σαν στερεοφωνικό κομμάτι μουσικής σε οπτικό δίσκο CD, σαν πολυκαναλικό κομμάτι μουσικής σε οπτικό δίσκο DVD, σαν αρχείο ήχου μέσω του Διαδικτύου, κλπ). Το Κεφάλαιο αυτό θα ασχοληθεί με τις τεχνικές και τις τεχνολογίες που επιτρέπουν την υλοποίηση τέτοιων διαδικασιών, αρχίζοντας από το στάδιο της παραγωγής του ψηφιακού ήχου, δηλαδή της ηχοληψίας/επεξεργασίας (χρησιμοποιώντας την παραδοσιακή ονομασία, το στάδιο αυτό μπορεί να θεωρηθεί σαν αυτό της «ψηφιακής κωδικοποίησης του ηχητικού σήματος») και συνεχίζοντας στο στάδιο της ψηφιακής αναπαραγωγής κατά την ακρόαση (δηλαδή της «αποκωδικοποίησης του ψηφιακού ηχητικού σήματος»). Οι τεχνολογίες αυτές βασίζονται σε μεθόδους Ψηφιακής Επεξεργασίας Σήματος, Ψηφιακών Ηλεκτρονικών ή και Ψηφιακών Επικοινωνιών, αλλά λόγω των ιδιαιτέρων χαρακτηριστικών των ηχητικών σημάτων και του δέκτη (ακροατή), συνήθως βασίζονται σε εξειδικευμένες υλοποιήσεις οι οποίες αναλύονται στις επόμενες παραγράφους.

Το στάδιο της κωδικοποίησης του σήματος παραδοσιακά υλοποιείται σε κατάλληλους επαγγελματικούς χώρους (studio ηχοληψίας), όπου λόγω της εκτεταμένης χρήσης της ψηφιακής τεχνολογίας, τα ηχητικά σήματα μετά το στάδιο της καταγραφής από μικρόφωνο, ψηφιοποιούνται και τυγχάνουν επεξεργασίας στο ψηφιακό πεδίο. Έτσι, οι διάφορες βαθμίδες των συσκευών που χρησιμοποιούνται σε αυτούς τους χώρους, γίνονται σε όλο και μεγαλύτερο βαθμό πλήρως ψηφιακές, αποτελούμενες από αυτόνομες συσκευές, όπως φαίνεται και στο Σχήμα 3.1. Οι ψηφιακές αυτές συσκευές λειτουργούν με κατάλληλο υλικό (hardware), ικανό να υλοποιήσει τους απαιτούμενους αλγόριθμους σε «πραγματικό χρόνο», έτσι ώστε ο χρήστης – μηχανικός ήχου να μπορεί να ακούει χωρίς διακοπές ή καθυστερήσεις το επεξεργασμένο ηχητικό υλικό μέσα από την βαθμίδα αναφοράς (monitor) του συστήματος (Σχήμα 3.1). Συχνά, οι συσκευές αυτές ελέγχονται από υπολογιστή ο οποίος μπορεί, κάνοντας χρήση κατάλληλου λογισμικού ελέγχου, να πραγματοποιήσει τις διάφορες ρυθμίσεις των επιμέρους συσκευών. Με την αύξηση των ταχυτήτων επεξεργασίας που επιτυγχάνουν οι υπολογιστές, και δεδομένης της διάδοσης της χρήσης περιφερειακών υποσυστημάτων για ψηφιακή ηχοληψία/αναπαραγωγή σε υπολογιστές, όλο και περισσότερες λειτουργίες που εκτελούν οι αυτόνομες συσκευές του Σχήματος 3.1 ανατίθενται στον υπολογιστή, σε μορφή υλικού (hardware) που τοποθετείται υπό μορφή καρτών επέκτασης σε αυτόν, είτε και ειδικού λογισμικού, το οποίο πλέον είναι δυνατό να εκτελέσει σε «πραγματικό χρόνο» τις περισσότερες από τις διεργασίες που απαιτούνται (δηλαδή ψηφιακή ηχοληψία–αναπαραγωγή, τεμαχισμός, επεξεργασία σήματος, κ.λ.π.).

Έτσι, είναι πλέον δυνατό, το πλήρως ψηφιακό studio να βασίζεται σε υπολογιστή ψηφιακό σταθμό εργασίας ήχου (Digital Audio Workstation – D.A.W.), εξοπλισμένο με υποσυστήματα (κάρτες) ήχου, στον οποίο εκτελείται κατάλληλο λογισμικό που προσομοιώνει παραδοσιακές λειτουργίες των αυτόνομων συσκευών ηχητικής επεξεργασίας, όπως φαίνεται στο Σχήμα 3.2. Η προσέγγιση αυτή μάλιστα ακολουθείται πλέον και από ημι-επαγγελματικά ή/και ερασιτεχνικά ή ακόμη και σε κάποιο απλοποιημένο βαθμό, από τους απλούς χρήστες υπολογιστών



Σχήμα 3.1: Σύστημα ψηφιακής εγγραφής και επεξεργασίας βασισμένο σε αυτόνομες συσκευές

Στο Κεφάλαιο αυτό θα δοθεί η γενική περιγραφή των λειτουργιών της κονσόλας μίξης, των ψηφιακών επεξεργαστών και των υπολογιστών που χρησιμοποιούνται σε σταθμοί εργασίας ήχου.

(α) Ψηφιακή κονσόλα μίξης

Κατ' αναλογία των λειτουργιών μίας αναλογικής κονσόλας μίξης, μία ψηφιακή κονσόλα που είτε υλοποιείται σαν ανεξάρτητη συσκευή (Σχήμα 3.1), είτε υλοποιείται από λογισμικό, εκτελεί τις ακόλουθες βασικές λειτουργίες:

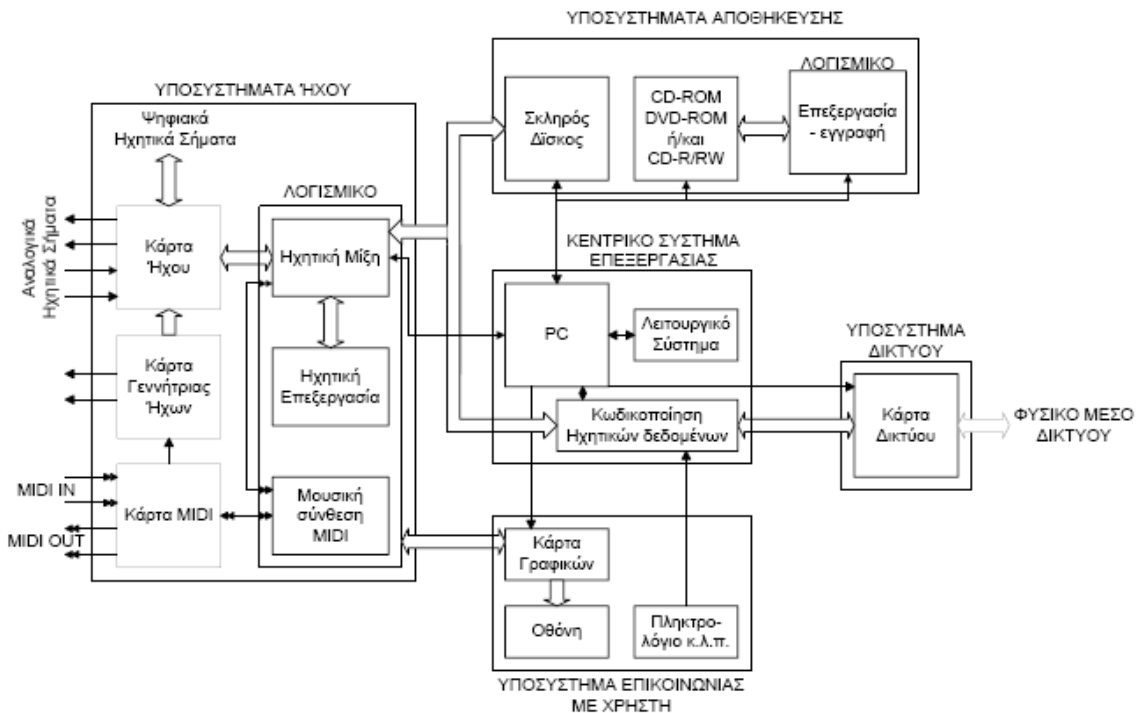
(1) την άθροιση πολλαπλών ηχητικών καναλιών -σημάτων εισόδου (π.χ. 2 – 64), ή και άλλων σημάτων τα οποία είναι αποθηκευμένα σε τοπικό μέσο, σε συγκεκριμένο αριθμό καναλιών εξόδου (π.χ. 1-8).

(2) την προσθήκη κατάλληλου (χρονικά σταθερού ή μεταβλητού από τον μηχανικό ήχου)

κέρδους στο κάθε κανάλι

(3) την προσθήκη κατάλληλης ηχητικής επεξεργασίας στο κάθε κανάλι (π.χ. φιλτράρισμα, τεχνητή αντήχηση, κ.λ.π.)

(4) τη διανομή των επεξεργασμένων ή μη ηχητικών καναλιών σε επιλεγμένες ομάδες ή μεμονωμένα κανάλια εξόδου



Σχήμα 3.2: Σύστημα ψηφιακής εγγραφής και επεξεργασίας ήχου βασισμένο σε υπολογιστή

Η ουσιαστική διαφοροποίηση των παραπάνω λειτουργιών στις ψηφιακές κονσόλες σε σχέση με τις συμβατικές αναλογικές, βρίσκεται στο γεγονός ότι όλες οι παραπάνω επεξεργασίες εκτελούνται στο ψηφιακό πεδίο, για πλήρως ψηφιακά κανάλια και σήματα. Αυτό το γεγονός επιτρέπει στις ψηφιακές κονσόλες να εμφανίζουν τα εξής πλεονεκτήματα σε σχέση με αναλογικές:

(i) επιτυγχάνεται ακριβέστερη επεξεργασία (π.χ. με ακρίβεια περιόδου δειγματοληψίας) ή/και υλοποιούνται επεξεργασίες που δεν μπορεί να πραγματοποιηθούν στο αναλογικό πεδίο

(ii) ο έλεγχος/αποθήκευση/ανάκληση των παραμέτρων όλων των παραπάνω επεξεργασιών μπορεί να γίνει επίσης ψηφιακά και με μεγάλη ακρίβεια

(iii) οι διαστάσεις της κονσόλας δεν είναι απαραίτητα εξαρτημένες από τον αριθμό των καναλιών εισόδου / εξόδου (αφού δεν απαιτείται διακριτό κύκλωμα για το κάθε κανάλι) και έτσι το μέγεθος της κονσόλας μπορεί να είναι μικρό και επίσης να παρουσιάζει βέλτιστα χαρακτηριστικά εργονομίας.

Βαθμίδα	Τοπολογία/στοιχεία	
	Αναλογική Υλοποίηση	Ψηφιακή Υλοποίηση
Μεταβολή κέρδους	Τελεστικός ενισχυτής, μεταβλητό ποτενσιόμετρο	Πολλαπλασιαστής, ψηφιακό ποτενσιόμετρο
Άθροιση σημάτων	Τελεστικός ενισχυτής	Άθροιστής, καθυστέρηση / μνήμη RAM, πολυπλέκτης
Φίλτρο ισοστάθμισης	Τελεστικός ενισχυτής, μεταβλητά ποτενσιόμετρα, αντιστάσεις	Εξίσωση διαφοράς, πολλαπλασιαστής, ψηφιακό ποτενσιόμετρο, άθροιστής, καθυστέρηση / μνήμη RAM, πολυπλέκτης
Διακόπτης επιλογής σήματος	Διακόπτης πολλαπλών θέσεων	Ψηφιακός διακόπτης, απο-πολυπλέκτης

Πίνακας 3.1: Απαραίτητα στοιχεία για την υλοποίηση της ηχητικής επεξεργασίας με αναλογικά και ψηφιακά μέσα.

Σε αντίθεση με τις αναλογικές υλοποιήσεις, όπου αποκλειστικά κυκλώματα εκτελούν συγκεκριμένες διεργασίες (π.χ. προ-ενίσχυση, άθροιση, ισοστάθμιση), στο ψηφιακό πεδίο, όλες αυτές οι διεργασίες υλοποιούνται από την καθιερωμένη μεθοδολογία της Ψηφιακής Επεξεργασίας Σημάτων (ΨΕΣ-DSP), χρησιμοποιώντας τις γνωστές βαθμίδες που εκτελούν τις αντίστοιχες σχέσεις, όπως φαίνεται για παράδειγμα στον Πίνακα 2.1. Έτσι, το ουσιαστικότερο πρόβλημα στην υλοποίηση των ψηφιακών μικτών είναι της βέλτιστης κατανομής της υπολογιστικής ισχύος των επεξεργαστών ΨΕΣ που χρησιμοποιούνται (οι οποίοι, συνήθως, είναι γενικού σκοπού, εκτός από ελάχιστες εξαιρέσεις υλοποιήσεων με επεξεργαστές ειδικής χρήσης), ώστε να επιτυγχάνεται η ταυτόχρονη εκτέλεση όλων των απαιτούμενων εντολών του κώδικα (μέχρι 10^9 εντολές/δευτερόλεπτο), ο οποίος ελέγχει όλες τις εντολές ελέγχου από τον χρήστη (π.χ. αλλαγές στις ρυθμίσεις των ποτενσιομέτρων), επιλέγει και ελέγχει τα κατάλληλα σήματα εισόδου, εκτελεί τις απαραίτητες αριθμητικές πράξεις και μεταφέρει τα επεξεργασμένα δεδομένα στην βαθμίδα εξόδου

Ταυτόχρονα, πολλά επιμέρους προβλήματα θα πρέπει να αντιμετωπισθούν στον σχεδιασμό των διαφόρων βαθμίδων. Για παράδειγμα, στην βαθμίδα εξόδου, η άθροιση των καναλιών μπορεί να προκαλέσει την αδυναμία του επεξεργαστή να επεξεργαστεί το εύρος των αριθμητικών τιμών που προκύπτουν, ενώ κάτι αντίστοιχο μπορεί να προκύψει και στις ενδιάμεσες βαθμίδες επεξεργασίας (π.χ. κατά την ισοστάθμιση). Έτσι, σε αντίθεση με τις παραδοσιακές αναλογικές υλοποιήσεις όπου ο περιορισμός των τιμών πλάτους σε κάθε βαθμίδα του κυκλώματος μπορεί να επιτευχθεί προοδευτικά (και χωρίς σημαντικές ηχητικές παραμορφώσεις), στα αντίστοιχα ψηφιακά υποσυστήματα, απαιτείται ο επανακβαντισμός του σήματος. Η ακριβής μέθοδος υλοποίησης των παραπάνω βαθμίδων θα συζητηθεί παρακάτω.

(β) Ψηφιακοί επεξεργαστές ήχου

Οι επεξεργαστές αυτοί υλοποιούνται είτε σε υλικό (hardware), είτε σε λογισμικό (software) και διαχωρίζονται στις εξής κύριες περιοχές εφαρμογών:

(ι) ισοσταθμιστές/φίλτρα (equalisers, filters): Ενισχύουν ή περιορίζουν επιλεκτικά

φασματικές περιοχές του σήματος. Όπως θα δούμε υλοποιούνται με εξισώσεις διαφοράς που περιγράφουν «παραμετρικά» ή «γραφικά» φίλτρα/τράπεζες φίλτρων, 1ης-4ης τάξης, με τοπολογίες IIR ή FIR. Σε ειδικές περιπτώσεις υλοποιούνται και φίλτρα υψηλής τάξης (π.χ. μέχρι και 3000 συντελεστών), σε τοπολογίες IIR ή FIR συχνά υπό μορφή φασματικού γινομένου (μέσω χρήσης σχέσεων FFT).

(ii) μετατροπείς δυναμικής περιοχής (dynamic range modifiers, compressors, limiters, expanders): υλοποιούν μη-γραμμικές σχέσεις εισόδου/εξόδου (π.χ. λογαριθμικές εξισώσεις) που επιτρέπουν την επιλεκτική συμπίεση, περιορισμό ή επέκταση του πλάτους του επεξεργασμένου σήματος.

(iii) μονάδες προσθήκης καθυστέρησης, αντήχησης και σχετικών διαμορφώσεων (delay lines, reverberators, chorus/flanging processors): υλοποιούν ειδικές σχέσεις εξισώσεων διαφοράς που προσθέτουν καθυστερημένες χρονικά εκδοχές του αρχικού σήματος με αυτό. Ο αριθμός, πυκνότητα και σχετικό κέρδος των ανακλάσεων μπορεί να προέρχεται από αλγόριθμο προσομοίωσης ακουστικής κλειστών χώρων, ή ακόμα οι παραπάνω παράμετροι μπορεί να διαμορφώνονται χρονικά για την υλοποίηση ηχητικών εφέ.

(iv) μονάδες τροποποίησης ακουστικού ειδώλου: υλοποιούν ειδικές σχέσεις μεταξύ των καναλιών του σήματος εισόδου ή και εξισώσεις διαφοράς για την περιγραφή ειδικών φίλτρων, τα οποία παράγουν την εντύπωση ότι η πηγή βρίσκεται σε συγκεκριμένη γωνία ως προς τον ακροατή. Μια μέθοδος τροποποίησης ακουστικού ειδώλου είναι και η γνωστή σε όλους μας στερεοφωνία.

v) μετατροπείς συχνότητας δειγματοληψίας: υλοποιούν κυρίως ψηφιακά χαμηλοδιαβατά φίλτρα ή/και τράπεζες φίλτρων (εξισώσεις διαφοράς) για την απόρριψη παρασιτικών ειδώλων που προκύπτουν από την διαφορετική της αρχικής συχνότητα δειγματοληψίας.

(γ) Σταθμοί εργασίας ήχου

Με την συνεχώς αυξανόμενη υπολογιστική ισχύ που προσφέρεται από τους προσωπικούς υπολογιστές, τα συστήματα αυτά επιτρέπουν πλέον την ηχογράφιση, αποθήκευση, επεξεργασία, διανομή και αναπαραγωγή των ηχητικών δεδομένων. Τα συνηθέστερα περιβάλλοντα στα οποία εγκαθίστανται τέτοιες εφαρμογές είναι τα Windows της Microsoft και το Macintosh της Apple, ενώ πέραν της αρχικής μετατροπής του σήματος από αναλογικό σε ψηφιακό (και το αντίστροφο), διεργασίες οι οποίες εκτελούνται από ειδική κάρτα ήχου, όλες οι υπόλοιπες επεξεργασίες μπορούν πλέον να υλοποιηθούν σε λογισμικό. Το Soundforge είναι ένας σταθμός εργασίας ήχου που βασίζεται σε λογισμικό.

3.2 Γενικές Απαιτήσεις από τις Συσκευές

Για τη λειτουργία των παραπάνω συσκευών ή αλγορίθμων σε «πραγματικό χρόνο», κύρια απαίτηση είναι η δυνατότητα της επεξεργασίας των ηχητικών δειγμάτων με την κατάλληλη ταχύτητα και αριθμητική ακρίβεια, έτσι ώστε όλες οι απαραίτητες διεργασίες (ανάκλιση δείγματος από την βαθμίδα εισόδου, αριθμητικές πράξεις υλοποίησης αλγορίθμου, αποστολή δείγματος στην βαθμίδα εξόδου, κλπ), να εκτελούνται στα χρονικά και αριθμητικά πλαίσια που ορίζει η κάθε ηχητική εφαρμογή. Συγκεκριμένα, όλες οι παραπάνω διεργασίες θα πρέπει να υλοποιούνται μέσα στα χρονικά όρια της μίας περιόδου

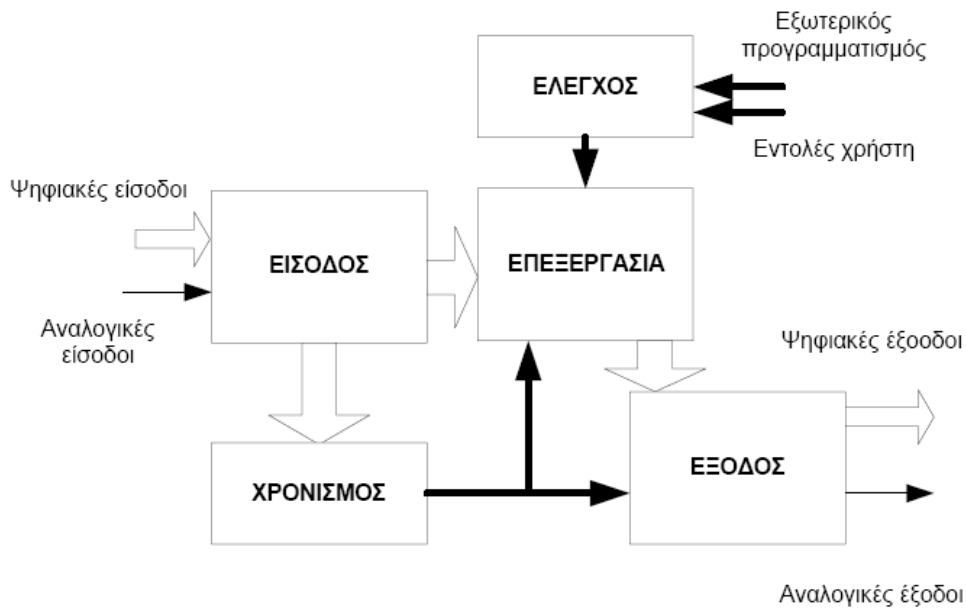
δειγματοληψίας, ενώ η αριθμητική ακρίβεια των πράξεων θα εξαρτάται τόσο από την αριθμητική ακρίβεια των δειγμάτων (όπως εισάγονται από το στάδιο εισόδου και αποστέλλονται στο στάδιο εξόδου), όσο και από την αριθμητική ακρίβεια των παραμέτρων που καθορίζουν την υλοποίηση του συγκεκριμένου αλγορίθμου (π.χ. των συντελεστών ενός ψηφιακού φίλτρου). Οι βασικές σχέσεις που χρησιμοποιούνται για την υλοποίηση των παραπάνω εφαρμογών και οι γενικές υπολογιστικές τους απαιτήσεις δίνονται στον Πίνακα 3.3, ενώ η ακριβής μορφή κάποιων εκ των αλγορίθμων θα περιγραφεί αργότερα.

	Δομή αλγορίθμου	Υπολογιστική πολυπλοκότητα	Ειδικές απαιτήσεις σε υλικό (πέραν του επεξεργαστή ΨΕΣ)
ισοσταθμιστές/φίλτρα	εξίσωση διαφοράς	μέτρια	καμμία
μετατροπείς δυναμικής περιοχής	εξισώσεις τροποποίησης τιμής δείγματος εισόδου	χαμηλή	καμμία
μονάδες προσθήκης καθυστέρησης, αντήχησης και σχετικών διαμορφώσεων	εξισώσεις διαφοράς με αποθήκευση τιμών σε μνήμη	υψηλή	μνήμη RAM
μονάδες τροποποίησης ακουστικού ειδώλου	εξίσωση διαφοράς ή συνέλιξη (FFT)	χαμηλή έως υψηλή	καμμία
μετατροπείς συχνότητας δειγματοληψίας	εξίσωση διαφοράς ή συνέλιξη (FFT)	μέτρια έως υψηλή	ειδική δομή φίλτρου ή επεξεργαστής FFT
φίλτρα καταστολής παραμορφώσεων	εξίσωση διαφοράς ή συνέλιξη (FFT)	υψηλή	ειδική δομή φίλτρου ή επεξεργαστής FFT
μονάδες κωδικοποίησης/ αποκωδικοποίησης δεδομένων	μετασχηματισμοί FFT ή τράπεζες φίλτρων	υψηλή	επεξεργαστής FFT

Πίνακας 3.3: Βασική δομή αλγορίθμων και ειδικές απαιτήσεις στις διάφορες κατηγορίες ψηφιακών επεξεργαστών ήχου

3.3 Υλοποίηση Μεθόδων - Συστημάτων σε Υλικό

Η πλέον συνήθης υλοποίηση σε υλικό (hardware) των τεχνικών και των συσκευών που παρουσιάστηκαν στις προηγούμενες παραγράφους, βασίζεται στην χρήση μικροεπεξεργαστών ΨΕΣ ειδικού σκοπού (DSP chips), οι οποίοι διασυνδέονται με κατάλληλα περιφερειακά, όπως φαίνεται στο Σχήμα 2.3. Τα βασικά υποσυστήματα μιας τέτοιας συσκευής είναι: (α) εισόδου/εξόδου και χρονισμού δεδομένων (αναλογικά ή και ψηφιακά), (β) κυρίως επεξεργασίας δεδομένων (βαθμίδα ΨΕΣ), (γ) ελέγχου, αποθήκευσης παραμέτρων και επικοινωνίας με χρήστη ή και υπολογιστή.



Σχήμα 3.3: Υλοποίηση σε υλικό (Hardware)

3.3.1 Βαθμίδα εισόδου εξόδου

Το υποσύστημα εισόδου/εξόδου και χρονισμού των δεδομένων βασίζεται στην χρήση βαθμίδων εισόδου-εξόδου αναλογικού ή και ψηφιακού ηχητικού σήματος και προαιρετικού εξωτερικού συγχρονισμού (wordclock), οι οποίες υλοποιούνται με την χρήση καταλλήλων μικροεπεξεργαστών (chips).

3.3.2 Βαθμίδα επεξεργασίας δεδομένων

Η κυρίως βαθμίδα επεξεργασίας δεδομένων (ΨΕΣ), αποτελείται από κύκλωμα ειδικού σκοπού το οποίο υλοποιεί τον αντίστοιχο αλγόριθμο εφαρμογής της συσκευής (π.χ. ένα ψηφιακό φίλτρο). Δεδομένου ότι οι περισσότεροι αλγόριθμοι ΨΕΣ, έχουν απαιτήσεις για γρήγορες υλοποιήσεις αριθμητικών πράξεων πολλ/σμού και άθροισης (Multiply/Accumulate – MAC), από την αρχή της δεκαετίας του '80 έχουν παρουσιαστεί στην αγορά επεξεργαστές ειδικού σκοπού ΨΕΣ (DSP chips) βασισμένοι σε ανάλογη αρχιτεκτονική (τύπου Harvard), που πλέον προσφέρουν υψηλές ταχύτητες επεξεργασίας, χαμηλό κόστος και ευέλικτο προγραμματισμό από τον χρήστη. Πέρα από τις προ-υλοποιημένες βαθμίδες πολλαπλασιασμού και άθροισης (MAC), οι επεξεργαστές περιέχουν τοπικές μνήμες ROM ή RAM, στις οποίες αποθηκεύονται οι συντελεστές των φίλτρων και μνήμες RAM για την κυκλική αποθήκευση μερικών αποτελεσμάτων ή την υλοποίηση στοιχείων καθυστέρησης (θα δούμε τη σχέση αυτών των μνημών με την υλοποίηση του αλγόριθμου επεξεργασίας και στο κεφάλαιο 3.5). Βασική προϋπόθεση της επιλογής επεξεργαστή, αποτελούν οι υπολογιστικές απαιτήσεις της συγκεκριμένης εφαρμογής, έτσι ώστε το συνολικό σύστημα να έχει την δυνατότητα λειτουργίας σε συνθήκες «πραγματικού χρόνου» (real-time). Έτσι, στο διάστημα μίας περιόδου δειγματοληψίας των δεδομένων, η βαθμίδα ΨΕΣ θα πρέπει να έχει την δυνατότητα υλοποίησης τόσο όλων των αριθμητικών πράξεων που περιγράφουν τον αλγόριθμο, όσο και την εκτέλεση των υπολοίπων βοηθητικών λειτουργιών (π.χ. ανάγνωσης δεδομένων, παραμέτρων, κλπ.). Η εκτέλεση των αριθμητικών αυτών πράξεων γίνεται σε δείγματα που οι τιμές τους είναι εκφρασμένες είτε

κατά την μέθοδο της σταθερής υποδιαστολής (fixed-point), είτε της κινητής υποδιαστολής (floating-point), ανάλογα με τις δυνατότητες του συγκεκριμένου επεξεργαστή. Στην περίπτωση που η ταχύτητα του επεξεργαστή δεν επαρκεί, τότε συχνά χρησιμοποιούνται βαθμίδες ΨΕΣ υλοποιημένες με περισσότερους του 1 επεξεργαστή, οι οποίοι λειτουργώντας συνδυαστικά, εκτελούν ο καθένας τμήμα του προς εκτέλεση αλγορίθμου.

Μία ακόμη σημαντική παράμετρος για την επιλογή του συγκεκριμένου επεξεργαστή είναι και το εύρος τιμών σε bit που επιτρέπει τόσο για τα δεδομένα εισόδου/εξόδου, για τα ενδιάμεσα αποτελέσματα των αριθμητικών πράξεων που μεταφέρονται μέσω του διαύλου δεδομένων (data bus), όσο και για την δυνατότητα εκτέλεσης πράξεων κινητής ή σταθερής υποδιαστολής, δεδομένου ότι διαφορετικές εφαρμογές έχουν διαφορετικές απαιτήσεις ως προς αυτά τα χαρακτηριστικά.

3.3.3 Βαθμίδες αποθήκευσης εντολών, παραμέτρων κι ελέγχου

Για την λειτουργία της συσκευής απαιτείται η αποθήκευση τόσο των εντολών του προγράμματος (Program), όσο και των ενδιάμεσων αποτελεσμάτων (Data) και παραμέτρων σε τοπικές μνήμες (ROM, RAM) οι οποίες και ανακαλούν κατάλληλες εντολές ελέγχου (Control), διευθύνσεις (Address) και δεδομένα (Data) για την εκτέλεση του αλγορίθμου. Οι μνήμες αυτές συνήθως υλοποιούνται σαν εσωτερικές ενός επεξεργαστή ΨΕΣ ειδικού σκοπού (on-chip memory), εξωτερικές συνοδευτικές του επεξεργαστή μνήμες (external memory), ή ακόμη είναι δυνατό και να υλοποιούνται και από τελείως ανεξάρτητες βαθμίδες. Η δυνατότητα υλοποίησης ενός αλγορίθμου εξαρτάται εν μέρει μόνο από την ταχύτητα του επεξεργαστή, αφού το μέγεθος των μνημών αυτών, καθώς και η ταχύτητα προσπέλασης εντολών και δεδομένων, παίζουν επίσης πολύ σημαντικό ρόλο.

3.4 Υλοποίηση Μεθόδων και Συστημάτων σε Λογισμικό

Με την εξάπλωση της χρήσης Σταθμών Εργασίας Ήχου (DAW), τόσο σε επαγγελματικές, όσο και σε ερασιτεχνικές εφαρμογές και την ολοένα αυξανόμενη ταχύτητα των επεξεργασιών στους προσωπικούς υπολογιστές, είναι πλέον εφικτή η υλοποίηση αλγορίθμων και τεχνικών για την επεξεργασία ψηφιακών δεδομένων σε τέτοιο υπολογιστικό περιβάλλον αξιοποιώντας τον επεξεργαστή του υπολογιστή και χωρίς την χρήση άλλου εξειδικευμένου υλικού πέραν της κάρτας ήχου του υπολογιστή (δες προηγούμενες παραγράφους). Τέτοιες υλοποιήσεις μπορούν να διαχωριστούν σε: (α) προγράμματα που λειτουργούν σε μη-πραγματικό χρόνο, όπου τα ψηφιακά ηχητικά δεδομένα τυγχάνουν επεξεργασίας μέσω παραμέτρων που ορίζει ο χρήστης, αποθηκεύονται σε ένα νέο αρχείο το οποίο κατόπι είναι δυνατό να σταλεί προς ηχητική αναπαραγωγή στην κάρτα ήχου με τον συνήθη τρόπο και (β), προγράμματα που λειτουργούν σε πραγματικό χρόνο, όπου τα ψηφιακά ηχητικά δεδομένα τυγχάνουν επεξεργασίας μέσω παραμέτρων που ορίζει ο χρήστης με ταυτόχρονη και αδιάκοπη ηχητική αναπαραγωγή από την κάρτα ήχου (ή από άλλη περιφερειακή έξοδο του υπολογιστή).

Εδώ θα πρέπει να επισημανθεί και η ύπαρξη υβριδικών συστημάτων υλικού / λογισμικού, όπου με την χρήση ειδικών καρτών επεξεργασίας ήχου οι οποίες συνδέονται και λειτουργούν υπό τον έλεγχο υπολογιστών, υλοποιούνται τμήματα του προς εκτέλεση

αλγορίθμου στους επεξεργαστές της κάρτας (συνήθως τεχνολογίας ΨΕΣ ειδικού σκοπού που περιγράφηκε προηγουμένως), ενώ ο υπολογιστής εκτελεί το υπόλοιπο τμήμα των επεξεργασιών (συνήθως της επικοινωνίας με τον χρήστη μέσω του γραφικού περιβάλλοντος της εφαρμογής). Με την αύξηση των δυνατοτήτων των υπολογιστών, προβλέπεται ότι αυτή η υβριδική προσέγγιση θα περιοριστεί κυρίως σε ειδικές εφαρμογές. Η περιγραφή των μεθοδολογιών που ακολουθούνται σε κάθε περίπτωση θα επικεντρωθεί σε προσωπικούς υπολογιστές που λειτουργούν σε περιβάλλον Microsoft Windows, λόγω της εκτεταμένης χρήσης τους στις παραπάνω εφαρμογές. Η επικοινωνία του προγράμματος με τον χρήστη επιτυγχάνεται μέσω ενός γραφικού περιβάλλοντος αλληλοεπίδρασης (graphic interface), το οποίο είναι δυνατό να υλοποιηθεί με κατάλληλα εργαλεία σχεδιασμού, όπως το Microsoft Developer Studio, σε γλώσσες VisualC++ ή Visual Basic, έτσι κατά την δραστηριοποίηση της εφαρμογής να παρουσιάζεται στην οθόνη του υπολογιστή ένα παράθυρο διαλόγου (dialog box) στο οποίο ο χρήστης μπορεί να επιλέγει τα δεδομένα (αρχεία ήχου) εισόδου/εξόδου και τις τιμές των επιθυμητών παραμέτρων (δες Σχήμα 3.30). Για την ανάπτυξη του κώδικα εφαρμογής, ακολουθείται η συνήθης επιλογή της χρήσης συναρτήσεων της γλώσσας προγραμματισμού C ή και C++.

(α) προγράμματα μη πραγματικού χρόνου

Η συνήθης διαδικασία υλοποίησης ενός τέτοιου προγράμματος ακολουθεί τα παρακάτω βήματα, όπως φαίνεται και στο Σχήμα 2.4. Από το γραφικό περιβάλλον αλληλοεπίδρασης αναγνωρίζονται τα επιλεγμένα ονόματα αρχείων εισόδου /εξόδου και οι τιμές παραμέτρων

(1) Έλεγχος και αναγνώριση αρχείων:

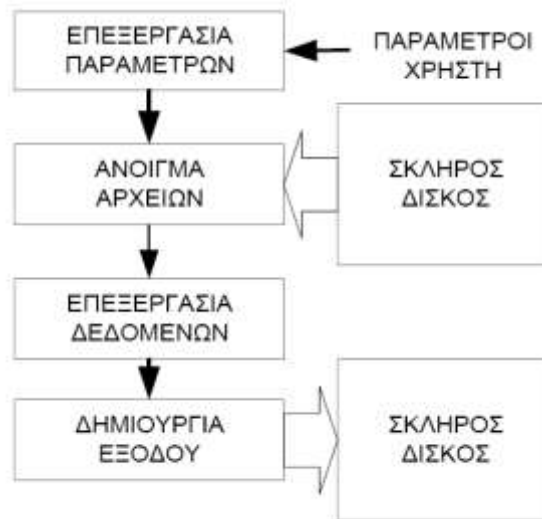
Ελέγχεται ο τύπος των αρχείων που έχουν επιλεγεί, έτσι ώστε τυχόν δεδομένα σε μορφή που δεν υποστηρίζονται από την εφαρμογή να μην τύχουν περεταίρω επεξεργασίας (π.χ. να είναι τύπου .WAV). Αναγνωρίζονται και αποθηκεύονται τα χαρακτηριστικά των αρχείων αυτών (διάρκεια, αριθμός δειγμάτων, αριθμός καναλιών, συχνότητα δειγματοληψίας, ακρίβεια κβαντισμού).

(2) Ανάγνωση και επεξεργασία δεδομένων:

Αρχίζει η ανάγνωση δεδομένων εισόδου, που αναλόγως της δομής του αλγορίθμου είναι δυνατόν να επιβάλει την χρήση τοπικής μνήμης (buffer). Εκτελείται ο αλγόριθμος χρησιμοποιώντας κατά προτίμηση αριθμητική κινητής υποδιαστολής (floating point) και τα επεξεργασμένα δεδομένα αποθηκεύονται σε τοπική μνήμη ή το αρχείο εξόδου.

(3) Δημιουργία αρχείου εξόδου:

Δημιουργείται η επικεφαλίδα του αρχείου εξόδου και τα επεξεργασμένα δεδομένα μεταφέρονται και αποθηκεύονται στο αρχείο αυτό.

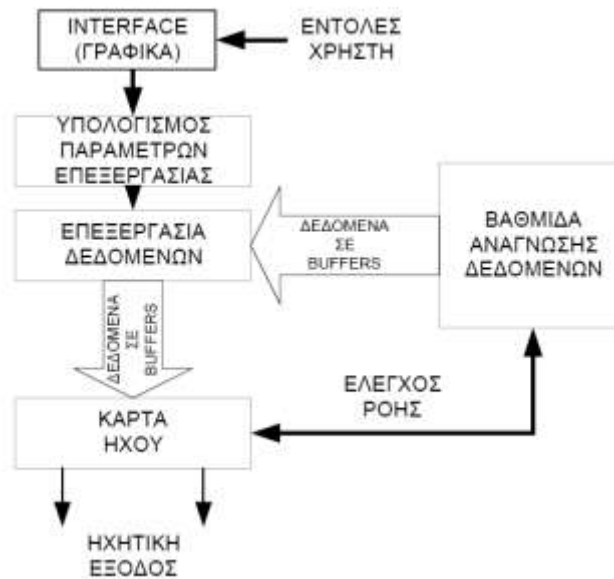


Σχήμα 3.4: Συνήθης διαδικασία υλοποίησης προγράμματος ηχητικής επεξεργασίας

(β) προγράμματα πραγματικού χρόνου

Η προσέγγιση αυτή ακολουθείται πλέον στην υλοποίηση τεχνικών που βασίζονται σε αλγορίθμους που υλοποιούνται σε επίπεδο δείγματος (π.χ. εξισώσεις διαφοράς, αριθμητικές πράξεις επί των δεδομένων, μεταβολής κέρδους, πρόσθεσης αντήχησης, ανακλάσεων, κλπ.). Για την υλοποίηση τέτοιων αλγορίθμων, απαιτείται η συνεχής και αδιάκοπη σύγχρονη ανάγνωση δειγμάτων από το αρχείο εισόδου, η επεξεργασία των δεδομένων με τις εκάστοτε τιμές παραμέτρων που ορίζει ο χρήστης μέσω του γραφικού περιβάλλοντος αλληλοεπίδρασης (interface) και η αποστολή των δεδομένων στην έξοδο (κάρτα ήχου) όπου τα επεξεργασμένα δεδομένα μετατρέπονται σε ήχο (Σχήμα 3.4). Το τελευταίο αυτό στάδιο πρακτικά αποδεικνύεται και το πλέον προβληματικό σε εφαρμογές σε περιβάλλον PC Windows, δηλαδή χρησιμοποιώντας τις παραδοσιακές εντολές πολυμέσων waveIn, waveOut του περιβάλλοντος Windows audio DDI που περιλαμβάνεται στις εντολές MCI (Media Control Interface), αφού απαιτείται η διαρκής τροφοδοσία της τοπικής μνήμης της κάρτας (buffer) με δεδομένα τα οποία και πρέπει να μεταφέρονται στα στάδια εξόδου, χωρίς καμία διακοπή ή καθυστέρηση, η οποία και δημιουργεί ηχητική παραμόρφωση στο αναπαραγόμενο σήμα (εμφανίζονται είτε χρονικά κενά, είτε στιγμιαίοι υψίσυχοι κρουστικοί ήχοι – clicks). Τυπικά, μία τέτοια εφαρμογή θα ακολουθούσε σε μεγάλο βαθμό την μεθοδολογία και τα βήματα (1)-(3) της προηγούμενης παραγράφου, αλλά σε αυτήν την περίπτωση, οι διαρκώς μεταβαλλόμενες από τον χρήστη τιμές των παραμέτρων (μέσω του γραφικού περιβάλλοντος), θα μεταφέρονται σε τακτά χρονικά διαστήματα (π.χ. κάθε 1/30 s) στην κυρίως βαθμίδα επεξεργασίας δεδομένων και θα ανανεώνουν τις ανάλογες βαθμίδες του αλγορίθμου. τσι, μετά την εκκίνηση του προγράμματος, την αναγνώριση και αποθήκευση των χαρακτηριστικών του αρχείου, εκτελείται η διαδικασία κατανομής τοπικής μνήμης (buffers) και το άνοιγμα της επικοινωνίας με την κάρτα ήχου. Το σύνηθες μέγεθος του τοπικού buffer της κάρτας ήχου είναι 1470 δείγματα (33.3 ms στα 44.1 kHz), το οποίο αρχικά γεμίζει από δεδομένα. Κατά την διάρκεια της αναπαραγωγής των δεδομένων αυτών, εκτελείται η επεξεργασία και η πλήρωση ενός δεύτερου, βοηθητικού buffer, ο οποίος μεταφέρεται στον πρώτο buffer, με την επιστροφή ενός bit ενημέρωσης από την κάρτα. Η

διαδικασία αυτή επαναλαμβάνεται συνεχώς μέχρι την διακοπή λειτουργίας του προγράμματος από τον χρήστη, ή μέχρι το τέλος των δεδομένων εισόδου (Σχήμα 3.4). Σαν πηγή των προβλημάτων στην αδιάκοπη λειτουργία ενός τέτοιου προγράμματος (που προαναφέρθηκαν), εντοπίζεται στην ανακρίβεια των περισσότερων καρτών ήχου για PC στο να ενημερώσουν την εφαρμογή για την ολοκλήρωση της αναπαραγωγής του περιεχομένου της τοπικής μνήμης (buffer).



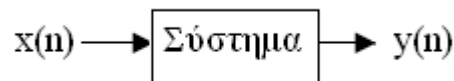
Σχήμα 3.4: Υλοποίηση προγραμμάτων πραγματικού χρόνου

4. ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΗΧΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Με τη συνεχή αύξηση της ισχύος των επεξεργαστών και την πρόοδο στην ψηφιακή αναπαράσταση του ηχητικού σήματος, η λειτουργία των ηχητικών εφέ έχει αρχίσει να βασίζεται ολοένα και περισσότερο στην ψηφιακή τεχνολογία. Η υψηλή οργάνωση επεξεργασίας του ψηφιακού σήματος έχει επιτρέψει την υλοποίηση διεργασιών που ήταν αδύνατες με τα αναλογικά μέσα. Επιπλέον, το κόστος έχει μειωθεί αισθητά με αποτέλεσμα το home studio να έχει γίνει πιο φτηνή λύση για το μουσικό από ότι η επίσκεψη στο επαγγελματικό studio. Στο κεφάλαιο αυτό υπενθυμίζεται η βασική θεωρία της Ψηφιακής Επεξεργασίας Σήματος η οποία είναι η βάση για την υλοποίηση της επεξεργασίας ηχητικού σήματος και αναλύονται κάποιες βασικές κατηγορίες ψηφιακών εφέ όπως τροποποίηση του ακουστικού ειδώλου, προσθήκη ηχούς, προσθήκη αντήχησης και ψηφιακή ισοστάθμιση.

4.1 Η ψηφιακή επεξεργασία σήματος σε ηχητικές εφαρμογές

Η βασική θεωρία ψηφιακής επεξεργασία σήματος όπου το σήμα και το σύστημα θεωρείται διακριτού χρόνου ορίζει επακριβώς τις βάσεις για τη μελέτη της επεξεργασίας του ψηφιακού ηχητικού σήματος. Η μονάδα επεξεργασία του ηχητικού σήματος μπορεί να αποδοθεί με την έννοια του *συστήματος*, όπου η είσοδος $x(n)$ αναπαριστά το πριν την επεξεργασία ηχητικό σήμα (*dry signal*) και η έξοδος $y(n)$ την τελική μορφή του επεξεργασμένου ηχητικού σήματος (*wet signal*).



Σχήμα 4.1: Αναπαράσταση μιας μονάδας ηχητικής επεξεργασίας μίας εισόδου και μίας εξόδου

α) FIR φίλτρα (Finite Impulse Response)

Η εξέταση του συστήματος θα γίνει σε πρώτη φάση θεωρώντας ένα Γραμμικό Χρονικά Αναλλοίωτο σύστημα (ΓΧΑ), και συγκεκριμένα, σύστημα με *πεπερασμένη κρουστική απόκριση* (FIR φίλτρο) το οποίο μπορεί να αναπαρασταθεί από εξισώσεις διαφοράς με σταθερούς συντελεστές ως

$$y(n) = b_0x(n) + b_1x(n-1) + \dots + b_qx(n-q) \quad (4.1)$$

Η εξίσωση διαφοράς εκφράζει τη έξοδο του συστήματος τη χρονική στιγμή n συναρτήσει της τρέχουσας τιμής, $x(n)$, και των παρελθοντικών τιμών $x(n-1)$, $x(n-2)$, ..., $x(n-q)$ του σήματος εισόδου και μπορεί να αναπαρασταθεί σε πιο συνοπτική μορφή μέσω της σχέσης

$$y(n) = \sum_{k=0}^q b_k x(n-k) \quad (4.2)$$

το q εδώ είναι ένας ακέραιος αριθμός που αναπαριστά την τάξη του φίλτρου. Τα b_0 , b_1 , b_2 , ..., b_q λέγονται συντελεστές του φίλτρου και είναι αυτοί που καθορίζουν την επίδραση του συστήματος στο εισερχόμενο προς επεξεργασία ηχητικό σήμα. Για παράδειγμα, με κατάλληλη επιλογή της τιμής των συντελεστών b_j μπορούμε να ορίσουμε αν το σύστημά

μας θα είναι ένα φίλτρο χαμηλής ή υψηλής διέλευσης συχνοτήτων, ή αν θα λειτουργεί σε μονάδα προσθήκης αντήχησης, λαμβάνοντας υπόψη την ακουστική ενός μικρού ή μεγάλου κλειστού χώρου. Σε ένα τέτοιο FIR φίλτρο οι συντελεστές περιγράφουν πλήρως την κρουστική απόκριση του συστήματος. Η κρουστική απόκριση του συστήματος της σχέσης (4.2) είναι μια ακολουθία q δειγμάτων η οποία ταυτίζεται με το διάνυσμα $h(n)=[b_0, b_1, b_2, \dots, b_q]$. Η κρουστική απόκριση του συστήματος ορίζει με μοναδικό τρόπο τα χρονικά χαρακτηριστικά της αλληλεπίδρασης σήματος-συστήματος. Η κρουστική απόκριση $h(n)$ μπορεί μέσω του μετασχηματισμού Fourier να μας δώσει την απόκριση συχνότητας $H(\omega)$ του συστήματος (το ω αναπαριστά συχνότητα), η οποία καθορίζει με μοναδικό τρόπο τα φασματικά χαρακτηριστικά του συστήματος.

β) IIR φίλτρα (Infinite Impulse Response)

Η εξίσωση διαφοράς είναι δυνατόν εκτός από τις καθυστερημένες τιμές του σήματος εισόδου να λαμβάνει υπόψη και τις καθυστερημένες τιμές του σήματος εξόδου $y(n)$. Η εξίσωση διαφοράς σε μια τέτοια περίπτωση γράφεται

$$a_0 y(n) + a_1 y(n-1) + \dots + a_p y(n-p) = b_0 x(n) + b_1 x(n-1) + \dots + b_q x(n-q) \quad (4.3)$$

όπου οι αριθμοί p και q μπορεί εν γένει να είναι διαφορετικοί. Ένα σύστημα του οποίου η έξοδος εξαρτάται από προηγούμενες τιμές της εξόδου λέγεται άπειρης κρουστικής απόκρισης (IIR φίλτρο). Σε πιο συνοπτική μορφή, ένα IIR σύστημα μπορεί να γραφεί ως

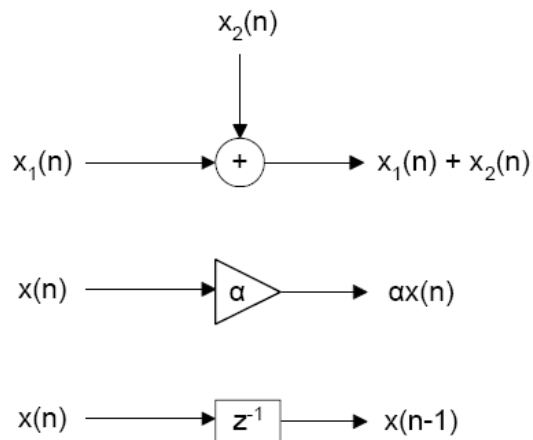
$$\sum_{k=0}^p a_k y(n-k) = \sum_{k=0}^q b_k x(n-k) \quad (4.4)$$

Είναι φανερό ότι για την υλοποίηση της παραπάνω παράστασης χρειάζεται να γίνει μια σύμβαση σχετικά με τις αρχικές συνθήκες οι οποίες καθορίζουν τις αρχικές τιμές εισόδου και εξόδου $x(n)$ και $y(n)$. Αν δηλαδή θεωρήσουμε ότι το σήμα εισόδου αρχίζει να διέρχεται από το σύστημα τη χρονική στιγμή $n=0$, η εξισώσεις (4.3) και (4.4) μας λένε ότι θα πρέπει να έχουμε ορίσει κάποιες τιμές για τα $x(-1), x(-2), \dots, x(-q)$ καθώς και για τα $y(-1), y(-2), \dots, y(-p)$ για να μπορέσουμε να υπολογίσουμε τις τιμές της εξόδου για τις χρονικές στιγμές $n=0$ και μετά. Οι αρχικές συνθήκες εν γένει σε μία εξίσωση διαφοράς ορίζονται ανάλογα με κάποιες παραδοχές σχετικά με το σύστημα, εμείς όμως θα εξετάζουμε τέτοιες εξισώσεις διαφοράς για μηδενικές αρχικές συνθήκες και μηδενικές αρχικές τιμές της εισόδου που σημαίνει πολύ απλά ότι

$$y(-1) = y(-2) = \dots = y(-p) = x(-1) = x(-2) = \dots = x(-q) = 0 \quad (4.5)$$

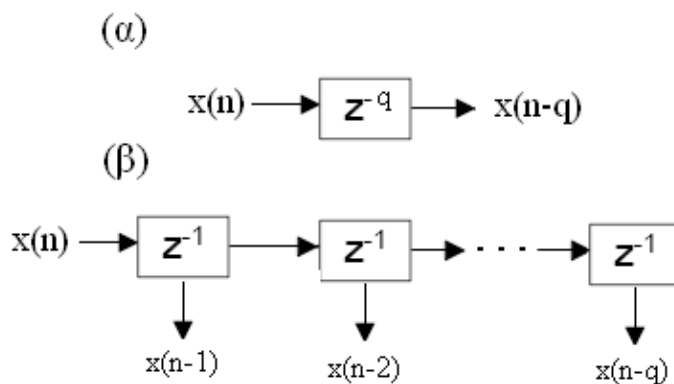
4.2 Δομικά στοιχεία του επεξεργαστή

Η χρήση εξισώσεων διαφοράς (όπως προαναφέρθηκε), βασίζεται σε τρεις επιμέρους διεργασίες, οι οποίες θεωρούνται τα «δομικά στοιχεία» μιας τέτοιας υλοποίησης, δηλαδή στα στοιχεία της άθροισης, του πολλαπλασιασμού και της καθυστέρησης (Σχήμα 4.2)



Σχήμα 4.2: Επιμέρους διεργασίες αλγορίθμων επεξεργασίας σήματος

Με αυτά τα στοιχεία μπορούμε να παραστήσουμε με πολύ εποπτικό τρόπο τη δομή του αλγόριθμου. Λίγο εξοικείωση χρειάζεται ενδεχομένως με το στοιχείο καθυστέρησης z^{-1} το οποίο συγκρατεί το δείγμα εισόδου στη μνήμη και αποδίδει στην έξοδο την τιμή του προηγούμενου δείγματος. Χρήση πολλών τέτοιων στοιχείων καθυστέρησης σε σειρά μπορούν να επιτύχουν μεγαλύτερες τιμές καθυστέρησης μέσω ενός στοιχείου z^{-q} που φαίνεται και στο το σχήμα 4.3(α). Στο σχήμα 4.3(β) φαίνεται ότι μπορούμε να στήσουμε έτσι τα στοιχεία καθυστέρησης ώστε ανά πάσα στιγμή να έχουμε διαθέσιμες όλες τις παρελθοντικές τιμές του σήματος εισόδου, μέχρι κάποια τάξη q . Τα στοιχεία καθυστέρησης λοιπόν μπορούν να χρησιμοποιηθούν για να αποστέλλουν στην έξοδό τους τις παρελθοντικές τιμές του σήματος εισόδου $x(n-1)$, $x(n-2)$, ..., $x(n-q)$ ή και του σήματος εξόδου $y(n-1)$, $y(n-2)$, ..., $y(n-p)$ που είναι απαραίτητες για την υλοποίηση των εξισώσεων διαφοράς που αναφέρθηκαν προηγουμένως.



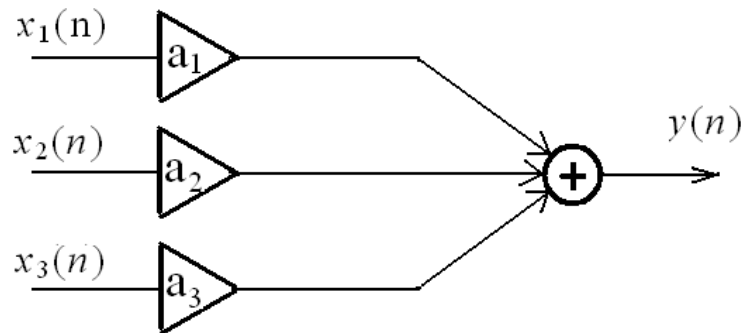
Σχήμα 4.3: Διατάξεις πολλών στοιχείων καθυστέρησης σε σειρά.

Ο συνδυασμός όλων των παραπάνω στοιχείων μπορεί να συνθέσει εν τέλει ένα οποιαδήποτε σύστημα το οποίο υλοποιεί της εξισώσεις διαφοράς (4.4) ή (4.2). Σε μετέπειτα φάση θα δούμε πως αυτά τα στοιχεία υλοποιούν μια σημαντική κατηγορία ψηφιακών ηχητικών εφέ όπως μονάδες equalizer, προσθήκη ηχούς και προσθήκη αντήχησης. Προς το παρόν, θα δούμε πως οι λειτουργίες της κονσόλας μπορούν να προσομοιωθούν με απλό τρόπο μέσω των στοιχείων πρόσθεσης και πολλαπλασιασμού.

4.3 Μίξη ψηφιακού σήματος

Μια από τις πιο βασικές διεργασίες που λαμβάνει χώρα σε μια ψηφιακή κονσόλα είναι η μίξη των ηχητικών σημάτων. Συνήθως, ένα ξεχωριστό κανάλι αφιερώνεται σε κάθε μουσικό όργανο. Η διαδικασία της μίξης που φαίνεται στο Σχήμα 4.4 έχει ως σκοπό την ένωση όλων των ηχητικών σημάτων σε ένα κανάλι, ορίζοντας ταυτόχρονα την σχετική αναλογία του κάθε σήματος (ή μουσικού οργάνου) μέσω των στοιχείων πολλαπλασιασμού a_1 , a_2 και a_3 τα οποία καθορίζουν το λεγόμενο ‘κέρδος’ του κάθε ηχητικού σήματος. Οι τιμές των a_1 , a_2 και a_3 καθορίζονται από τον ηχολήπτη με βάση αισθητικά κριτήρια, αλλά πρέπει να προσέχουμε ώστε το τελικό σήμα $y(n)$ να μην υπερβαίνει το δυναμικό εύρος του συστήματος (δηλαδή να μην έχουμε υπερχειλίση). Από το σχήμα βλέπουμε ότι έχουμε μια διεργασία με πολλές εισόδους και μία έξοδο (Multiple Input – Single Output). Για το σήμα εξόδου θα ισχύει προφανώς

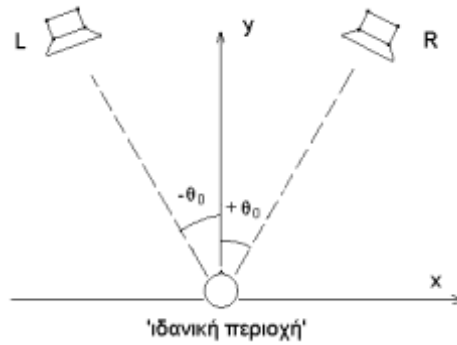
$$y(n) = a_1x_1(n) + a_2x_2(n) + a_3x_3(n) \quad (4.6)$$



Σχήμα 4.4: Συνένωση τριών ηχητικών σημάτων σε ένα κανάλι.

4.4 Τροποποίηση ακουστικού ειδώλου με διαφοροποίηση της έντασης

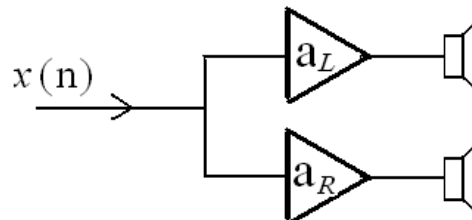
Η στερεοφωνική αναπαραγωγή (stereophony) με δύο πανομοιότυπα ηχεία κατάλληλα τοποθετημένα είναι σίγουρα η πιο κοινή διάταξη ηχητικής αναπαραγωγής που χρησιμοποιείται σήμερα. Ιδανικά, τα δύο ηχεία τοποθετούνται μπροστά από τον ακροατή με τέτοιο τρόπο ώστε ακροατής και ηχεία να σχηματίζουν ένα ισόπλευρο τρίγωνο (βλ. Σχήμα (4.5)). Η περιοχή ακρόασης όπου η ηχητική απεικόνιση είναι ιδανική περιορίζεται σε ένα μικρό τμήμα που ονομάζεται “ιδανική περιοχή”. Η στερεοφωνική αναπαραγωγή επιτυγχάνεται με τη διαφοροποίηση του σήματος στο κάθε ηχείο είτε ως προς την ένταση είτε ως προς το χρόνο.



Σχήμα 4.5: Μια τυπική διάταξη στερεοφωνικής αναπαραγωγής.

Η δημιουργία στερεοφωνικής εικόνας με διαφοροποίηση της έντασης είναι η πιο συνηθισμένη περίπτωση επίτευξης στερεοφωνίας. Αυτή η διαδικασία φαίνεται στο Σχήμα 4.5 για ένα ηχητικό σήμα $x(n)$. Στην ουσία, ο ηχολήπτης ορίζει την τιμή μιας παραμέτρου α τέτοιας ώστε $-1 \leq \alpha \leq 1$, από την οποία τα α_L και α_R προκύπτουν ως

$\alpha_L = 1 - \alpha$ και $\alpha_R = 1 + \alpha$. Η στερεοφωνική εικόνα μετατοπίζεται προς το ηχείο με τη μεγαλύτερη ένταση. Για ακραίες τιμές του α (-1 ή 1), είναι φανερό ότι η στερεοφωνική εικόνα ταυτίζεται με το αριστερό ή το δεξί ηχείο.



Σχήμα 4.6 Στερεοφωνική μίξη με διαφοροποίηση της έντασης.

4.5 Διάδοση του ήχου

Όπως αναφέρθηκε και στο πρώτο κεφάλαιο η ένταση του ηχητικού κύματος είναι ανάλογη του τετραγώνου της ακουστικής πίεσης και αντιστρόφως ανάλογη του τετραγώνου της απόστασης από μια ηχητική πηγή. Άρα η ακουστική πίεση θα είναι αντιστρόφως ανάλογη της απόστασης δηλαδή

$$p \sim \frac{1}{r} \quad (4.7)$$

και για δύο σημεία r_1, r_2 ισχύει

$$\frac{p_1}{p_2} = \frac{r_2}{r_1}. \quad (4.8)$$

Ακόμα αν δύο όμοια ηχητικά σήματα από πηγές με αποστάσεις r_1 και r_2 , συμβάλουν σε ένα σημείο του χώρου τότε είναι σαν να φτάνουν στο σημείο αυτό με διαφορά χρόνου

$$\Delta t = \frac{r_2 - r_1}{c} \quad (4.9)$$

Η διαφορά χρόνου αυτή, είναι υπαίτια για τη δημιουργία ηχητικών εφέ και φασματικών χρωματισμών που σε κάποιες περιπτώσεις είναι επιθυμητά, για τον ηθελημένο χρωματισμό

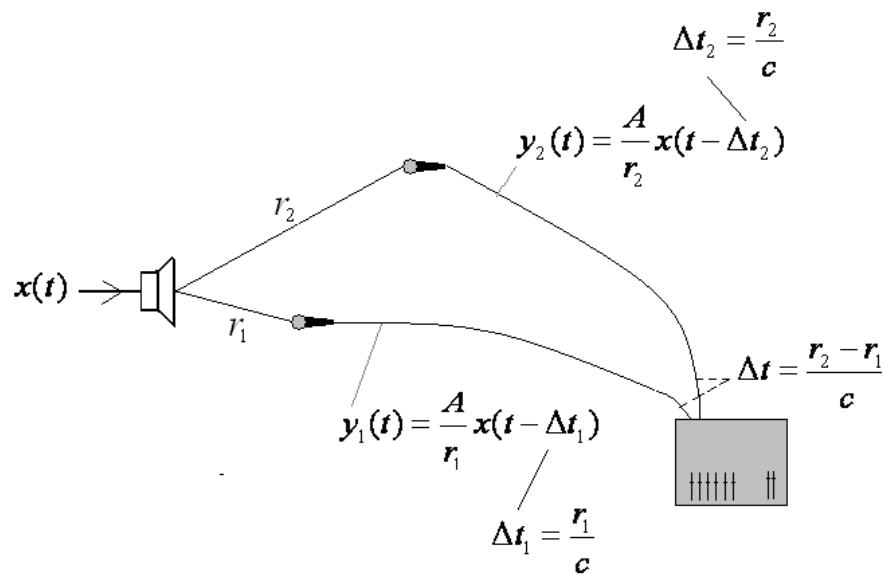
και την προσθήκη ηχούς στο σήμα, και σε άλλες περιπτώσεις ανεπιθύμητα, οπότε και πρέπει να αντιμετωπίζονται με τεχνικές ισοστάθμισης.

Στο σχήμα 4.7 έχουμε την περίπτωση που δύο μικρόφωνα βρίσκονται εκτεθειμένα σε μία ηχητική πηγή η οποία βρίσκεται σε απόσταση r_1 και r_2 από το καθένα. Ας υποθέσουμε ότι η ηχητική πηγή είναι ένα ηχητικό σύστημα αναπαραγωγής μουσικής με CD. Είμαστε σε θέση επομένως να γνωρίζουμε το ηλεκτρικό σήμα $x(t)$ που φεύγοντας από το CD player τροφοδοτεί τον ενισχυτή ο οποίος τροφοδοτεί το ηχείο. Το ηχείο είναι ένας ηλεκτρακουστικός μετατροπέας ο οποίος μετατρέπει την ηλεκτρική διέγερση σε ακουστική, με κατάλληλη κίνηση του διαφράγματος. Αντίστοιχα, το μικρόφωνο είναι ένας άλλος ηλεκτρακουστικός μετατροπέας ο οποίος μετατρέπει την ακουστική διέγερση σε ηλεκτρική. Αν το ηχείο και τα μικρόφωνα είναι απολύτως γραμμικά, μπορούμε να υποθέσουμε ότι η μεταβολή της **ηλεκτρικής τάσης** $y_1(t)$ και $y_2(t)$ μπορεί να περιγραφεί πλήρως από τη συνάρτηση $x(t)$ μέσω των σχέσεων

$$y_1(t) = \frac{A}{r_1} x(t - \Delta t_1) \text{ και } y_2(t) = \frac{A}{r_2} x(t - \Delta t_2) \quad (4.10)$$

όπου $\Delta t_1 = \frac{r_1}{c}$ και $\Delta t_2 = \frac{r_2}{c}$ είναι ο χρόνος που απαιτείται για τη διάδοση του ήχου από το ηχείο στα μικρόφωνα.

Οι εξισώσεις εδώ απαιτούνε την απόλυτα γραμμική αναλογία μεταξύ ηλεκτρικών και ακουστικών μεγεθών. Η ηχητική πίεση σε κάθε μικρόφωνο είναι μια εξασθενημένη και καθυστερημένη εκδοχή της ηχητικής πίεσης μπροστά από το ηχείο, γεγονός που εξασφαλίζει ότι το ηλεκτρικό σήμα σε κάθε μικρόφωνο θα είναι η ίδια εξασθενημένη και καθυστερημένη εκδοχή του ηλεκτρικού σήματος στην είσοδο του ηχείου. Στις εξισώσεις εμφανίζεται επίσης η σταθερά A η οποία μπορεί να εξαρτάται από ένα σωρό παράγοντες όπως η ισχύς του ενισχυτή, η απόδοση του ηχείου, η πυκνότητα του αέρα κ.α.



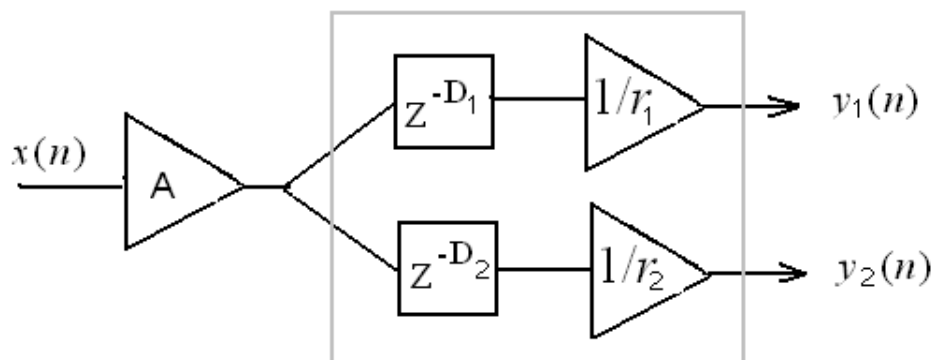
Σχήμα 4.7 Σχέσεις μεταξύ του ηχητικού σήματος εισόδου στην είσοδο του ηχείου και στην είσοδο δύο μικροφώνων σε διαφορετική απόσταση.

Εάν τα δύο σήματα από τα μικρόφωνα προστεθούν σε μία κονσόλα όπως για παράδειγμα στο Σχήμα 4.7, τότε θα δημιουργηθούν διάφορα ενδιαφέροντα φαινόμενα. Τα δύο σήματα $y_1(t)$ και $y_2(t)$ είναι όπως λέμε *συσχετισμένα* γεγονόσ που έχει να κάνει ότι το ένα είναι μια εξασθενημένη και καθυστερημένη εκδοχή του άλλου. Μεταξύ τους παρουσιάζεται *διαφορά φάσης* που έχει ως αποτέλεσμα το φασματικό χρωματισμό του αθροίσματός τους $y(t) = y_1(t) + y_2(t)$ μέσα στη μίξη. Καθοριστικό ρόλο σε αυτά τα φαινόμενα παίζει η σχετική καθυστέρηση του ήχου $\Delta t = \frac{r_2 - r_1}{c}$, που εκφράζει το πόσο καθυστερημένο είναι το σήμα στο ένα μικρόφωνο σε σχέση με το άλλο. Παρόμοια φαινόμενα συσχέτισης εμφανίζονται και στην περίπτωση που έχουμε δύο συσχετισμένες ηχητικές πηγές, όπως για παράδειγμα στην στερεοφωνική διάταξη ή κατά το φαινόμενο της ανάκλασης. Όπως θα δούμε σε μετέπειτα φάση, υπάρχουν κατάλληλες τεχνικές *ισοστάθμισης* οι οποίες μπορούν να αντισταθμίζουν ως ένα βαθμό αυτά τα φαινόμενα όταν είναι ανεπιθύμητα.

Η ψηφιακή αναπαράσταση της παραπάνω διάταξης πηγάζει άμεσα από τα μεγέθη και τις υποθέσεις που έχουν γίνει ήδη για το σήμα στην αναλογική του μορφή. Συγκεκριμένα, θεωρώντας διακριτό χρόνο μπορούμε να εκφράσουμε τα ψηφιακά σήματα $y_1(n)$ και $y_2(n)$ συναρτήσει του $x(n)$ ως

$$y_1(n) = \frac{A}{r_1} x(n - D_1) \text{ και } y_2(n) = \frac{A}{r_2} x(n - D_2) \quad (4.11)$$

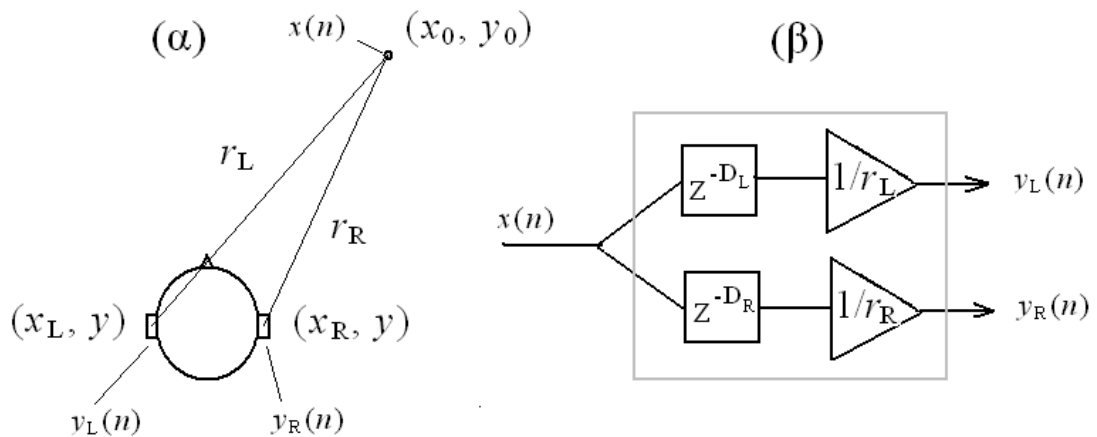
Λίγη προσοχή θέλει κατά τον ορισμό της τιμής των διακριτών καθυστερήσεων D_1 και D_2 οι οποίες προκύπτουν ως $D_1 = \frac{\Delta t_1}{T_s}$ και $D_2 = \frac{\Delta t_2}{T_s}$ όπου $T_s (=1/F_s)$ είναι η περίοδος δειγματοληψίας και με το \wedge εννοούμε ακέραια τιμή (τα D_1, D_2 πρέπει να είναι ακέραιοι αριθμοί δειγμάτων). Στο παρακάτω σχήμα φαίνεται πως μοντελοποιείται το φαινόμενο διάδοσης του ήχου από το ηχείο στα μικρόφωνα με χρήση των στοιχείων καθυστέρησης και πολλαπλασιασμού.



Σχήμα 4.8: Μοντελοποίηση της διάδοσης του ήχου για την περίπτωση μιας πηγής και δύο δεκτών.

4.6 Τροποποίηση ακουστικού ειδώλου με χρήση και στοιχείων καθυστέρησης

Εναλλακτικός τρόπος για τη δημιουργία στερεοφωνικής εικόνας (πέρα από τη διαφοροποίηση της έντασης που είπαμε προηγουμένως) είναι με τη μοντελοποίηση της διαφοράς χρονικής άφιξης στα αυτιά του ακροατή, όπως φαίνεται στο Σχήμα 3.9. Ένας ήχος ο οποίος έρχεται από τα δεξιά, φθάνει πρώτα στο δεξί αυτί και μετά στο αριστερό (διαωτική διαφορά χρονικής άφιξης). Επιπλέον, επειδή ταξιδεύει μεγαλύτερη απόσταση μέχρι να φτάσει στο αριστερό αυτί από ότι στο δεξί, παρουσιάζεται εξασθενημένος στο αριστερό αυτί. Πρέπει να σημειωθεί ότι αυτή η εξασθένηση δεν είναι μόνο αποτέλεσμα της μεγαλύτερης απόστασης διάδοσης, αλλά και των φαινομένων σκίασης που προκαλεί η κεφαλή με αποτέλεσμα η εξασθένηση να είναι ακόμα μεγαλύτερη από αυτήν που υπαγορεύει ο νόμος της αντίστροφης απόστασης. Για περισσότερα σχετικά με αυτό το θέμα αλλά και γενικότερα για τον τρόπο αντίληψης της κατεύθυνσης ο σπουδαστής παραπέμπετε στο κεφάλαιο 3 του βιβλίου 'Εγχειρίδιο Ακουστικής' (F. Alton Everest) που διατίθεται στα πλαίσια του μαθήματος.



Σχήμα 4.9: Διαφοροποίηση της κατεύθυνσης του ήχου με βάση τη θεωρία διάδοσης του ήχου. (α) Γεωμετρικό μοντέλο, (β) ψηφιακό μοντέλο.

Στην παρούσα φάση, το απλό μοντέλο που παρουσιάζεται στο Σχήμα 4.9 λαμβάνει υπόψιν μόνο την εξασθένηση του ήχου λόγω απόστασης. Η ηχητική πηγή θεωρείται σε ένα σημείο με συντεταγμένες (x_0, y_0) και τα αυτιά του ακροατή βρίσκονται σε συντεταγμένες (x_L, y) και (x_R, y) . Είναι φανερό η ομοιότητα του συγκεκριμένου συστήματος με αυτό που περιγράφηκε στην προηγούμενη παράγραφο. Οι αποστάσεις r_L και r_R μπορούν να υπολογιστούν με βάση τις συντεταγμένες ως

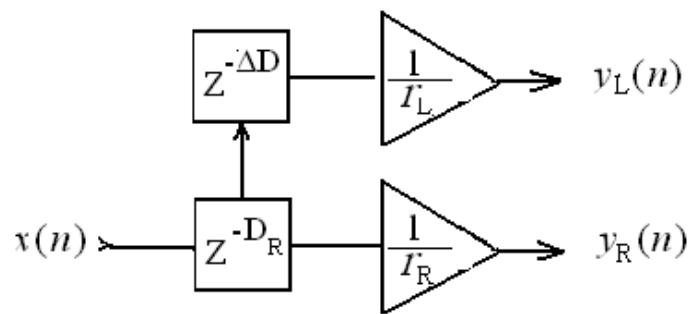
$$r_L = \sqrt{(x_L - x_0)^2 + (y - y_0)^2} \text{ και } r_R = \sqrt{(x_R - x_0)^2 + (y - y_0)^2} \quad (4.12)$$

ενώ οι χρονικές καθυστερήσεις D_L και D_R με βάση τις σχέσεις απόστασης δια ταχύτητας του ήχου που ορίστηκαν προηγουμένως. Μπορούμε να παρατηρήσουμε ότι για τα στοιχεία καθυστέρησης z^{-D_L} και z^{-D_R} απαιτείται μνήμη σε bits ανάλογη του αθροίσματος $D_L + D_R$. Τα σήματα εξόδου $y_L(n)$ και $y_R(n)$ σχετίζονται με το σήμα εισόδου $x(n)$ μέσω των σχέσεων

$$y_L(n) = \frac{1}{r_L} x(n - D_L) \text{ και } y_R(n) = \frac{1}{r_R} x(n - D_R) \quad (4.13)$$

Το παραπάνω σύστημα μπορεί να απλοποιηθεί κάνοντας χρήση ενός κοινού στοιχείου καθυστέρησης τάξης D_R σύμφωνα με το Σχήμα 4.10. Αυτή η υλοποίηση είναι εκτός των άλλων οικονομικότερη καθότι απαιτεί μικρότερο buffer για τη δημιουργία της καθυστέρησης στο πιο μακρινό αυτί μέσω του στοιχείου $z^{-\Delta D}$, όπου τώρα $\Delta D = D_L - D_R$, με το Δ να συμβολίζει διαφορά. Συγκεκριμένα, τα στοιχεία $z^{-\Delta D}$ και z^{-D_R} θα απαιτούν τώρα D_L σε αριθμό στοιχεία καθυστέρησης και από την τοπολογία του Σχήματος 4.10 φαίνεται ότι η σχέση εισόδου εξόδου του συστήματος μπορούν να γραφούν πάλι

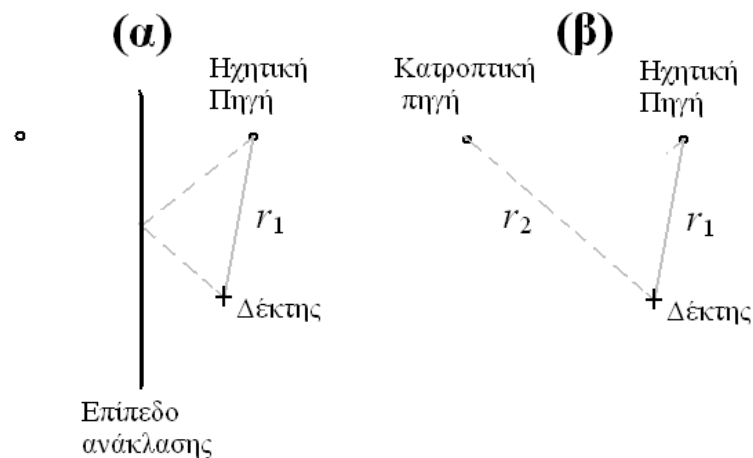
$$y_L(n) = y_R(n - \Delta D) \text{ και } y_R(n) = \frac{1}{r_R} x(n - D_R) \quad (4.14)$$



σχετική καθυστέρηση: $\Delta D = D_L - D_R$

Σχήμα 4.10 Απλούστευση του συστήματος του σχήματος 2.12(β) με ταυτόχρονη μείωση του απαιτούμενου buffer για τα στοιχεία καθυστέρησης

4.7 Ανάκλαση του Ήχου



Σχήμα 4.11: Ανάκλαση του ήχου σε ένα σκληρό επίπεδο (α) και ισοδύναμη αναπαράσταση με κατροπτική πηγή (β).

Ένα από τα πιο συχνά φαινόμενα που παρατηρείται κατά τη διάδοση του ήχου σε κλειστούς χώρους είναι οι ανακλάσεις. Γενικά, το ηχητικό σήμα σε ένα δέκτη αποτελείται από το άθροισμα του άμεσου ήχου και του ανακλώμενου ήχου. Ο άμεσος ήχος είναι το αποτέλεσμα της διάδοσης του ανεμπόδιστου ήχου αν θεωρήσουμε ότι δεν υπάρχουν ανακλάσεις. Στο Σχήμα 4.11(α) ο άμεσος ήχος αναπαρίσταται από την γκριζα συνεχόμενη γραμμή και ο ανακλώμενος ήχος από τη γκριζα διακεκομμένη γραμμή. Αν θεωρήσουμε ότι το επίπεδο ανάκλασης είναι ένας τοίχος ο οποίος εκτείνεται στο άπειρο και ότι ο τοίχος αποτελείται από πολύ σκληρό υλικό, τότε μπορούμε να αφαιρέσουμε το επίπεδο ανάκλασης και να θεωρήσουμε ότι ο ανακλώμενος ήχος είναι ακριβώς σα να προέρχεται από μια δεύτερη ηχητική πηγή η οποία είναι η κατοπτρική της κύριας ηχητικής πηγής σε σχέση με το επίπεδο ανάκλασης (Σχήμα 4.11(β)). Αυτό σημαίνει ότι η κατοπτρική ηχητική πηγή μοντελοποιεί επακριβώς την σχετική εξασθένηση και σχετική καθυστέρηση του ανακλώμενου ήχου σε σχέση με τον άμεσο ήχο. Από το Σχήμα 2.15(α) φαίνεται ότι το σήμα στο δέκτη μπορεί να περιγραφεί μέσω της εξίσωσης

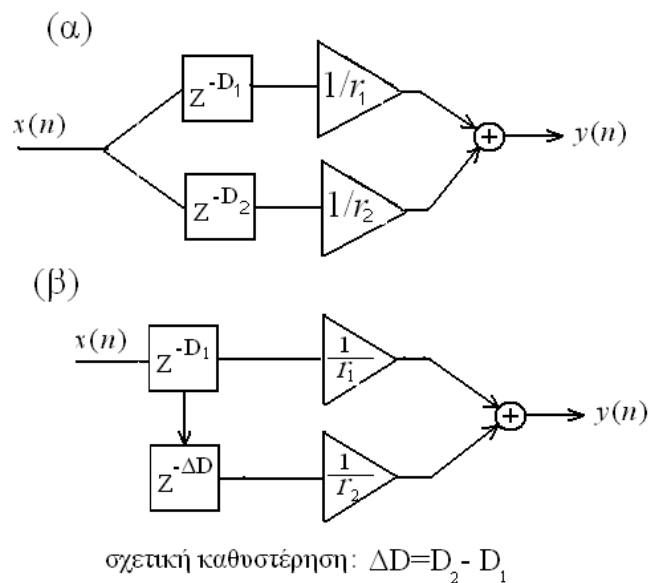
$$y(n) = \frac{1}{r_1} x(n - D_1) + \frac{1}{r_2} x(n - D_2) \quad (4.15)$$

όπου r_1 και r_2 οι αποστάσεις από την κύρια πηγή στο δέκτη και από την κατοπτρική πηγή στο δέκτη αντίστοιχα και $D_1 = \frac{r_1}{cTs}$, $D_2 = \frac{r_2}{cTs}$, οι αντίστοιχες καθυστερήσεις τους.

Παρατηρήστε ότι η θεώρηση της κατοπτρικής πηγής, η οποία μοντελοποιεί μια τέλεια ανάκλαση, οδηγεί στην πολύ απλή σχέση μεταξύ καθυστέρησης και απόστασης διάδοσης

$$\frac{r_1}{r_2} = \frac{D_1}{D_2} \quad (4.16)$$

Το σύστημα στο Σχήμα 4.12(α) μπορεί να υλοποιηθεί με τη χρήση λιγότερου buffer όπως και πριν, σύμφωνα με το σχήμα (β).



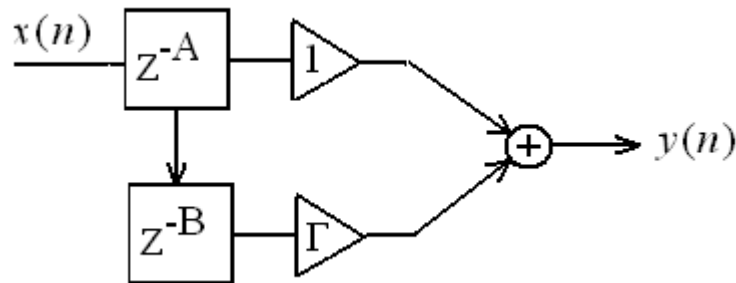
Σχήμα 4.12: (α) Αναπαράσταση της ανάκλασης με στοιχεία πολλαπλασιασμού και καθυστέρησης και (β), ισοδύναμη αναπαράσταση με ελαχιστοποίηση του buffer για τα στοιχεία καθυστέρησης.

Η υποκειμενική αντίδραση της ανθρώπινης ακοής στην ανάκλαση εξαρτάται τόσο από

τη σχέση μεταξύ της έντασης του άμεσου ήχου και του ανακλώμενου (που ουσιαστικά εκφράζεται από το λόγο $(\frac{r_1}{r_2})$ όσο και από τη σχετική καθυστέρηση $\Delta D = D_2 - D_1$.⁴

Λεπτομέρειες γύρω από αυτό το θέμα αναφέρονται στο κεφ. 3 του συγγράμματος 'Εγχειρίδιο Ακουστικής'.

Παράδειγμα: Μια ηχητική πηγή εκπέμπει μπροστά από ένα επίπεδο ανάκλασης με ένα δέκτη. Η απόσταση του άμεσου και του ανακλώμενου ήχου είναι 0,312m και 0,468m αντίστοιχα. Η ταχύτητα του ήχου είναι $c=344\text{m/sec}$ και η συχνότητα δειγματοληψίας $F_s=44,1\text{ kHz}$. Βρείτε τις τιμές των καθυστερήσεων A , B και του πολλαπλασιαστή Γ που μοντελοποιούν την ανάκλαση στο παρακάτω σχήμα και γράψτε τις εξισώσεις διαφοράς που δίνουν την έξοδο $y(n)$ συναρτήσει της εισόδου $x(n)$. Σχεδιάστε την κρουστική απόκριση του συστήματος.



Απάντηση: Κατ' αρχάς υπολογίζουμε τις καθυστερήσεις του άμεσου και του ανακλώμενου ήχου. $D_1 = \frac{0.312}{344 \cdot (1/44100)} = 40$ και $D_2 = \frac{0.468}{344 \cdot (1/44100)} = 60$.

Στο γράφημα δε μας ορίζεται ποιος από τους δύο κλάδους αντιπροσωπεύει τον άμεσο ήχο και ποιος τον ανακλώμενο ήχο. Επειδή ο ανακλώμενος ήχος πρέπει να φτάνει καθυστερημένα σε σχέση με τον άμεσο ήχο, καταλαβαίνουμε ότι θα πρέπει να χρησιμοποιήσουμε τον κάτω κλάδο για τον ανακλώμενο ήχο, όπου το διαθέσιμο στοιχείο καθυστέρησης z^{-B} προσδίδει την επιπλέον αυτή καθυστέρηση. Η τιμή του A είναι επομένως 40 δείγματα και για το B θα ισχύει $B = \Delta D = 60 - 40 = 20$ δείγματα.

Παρατηρούμε ότι η τιμή του πολλαπλασιαστή του άμεσου ήχου είναι ίση με ένα. Επειδή η ένταση του ανακλώμενου ήχου θα πρέπει να είναι μικρότερη από αυτή του άμεσου ήχου, καταλαβαίνουμε ότι θα πρέπει η τιμή του Γ να είναι μικρότερη της μονάδας. Για να διατηρηθεί λοιπόν η αναλογία έντασης μεταξύ των δύο ήχων, καταλήγουμε ότι $\Gamma = \frac{r_1}{r_2} = \frac{0,312}{0,468} = 0.667$.

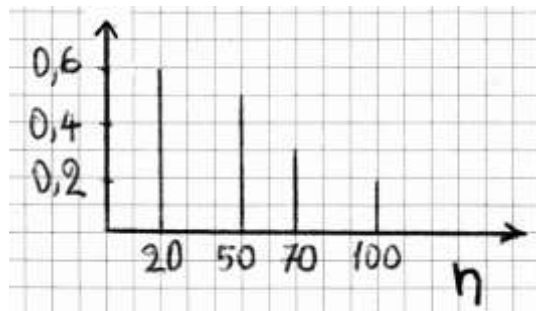
⁴ Χρησιμοποιώντας της γλώσσα των decibels, η εξασθένηση του ανακλώμενου ήχου σε dB μπορεί να υπολογιστεί

$$\text{μέσω της σχέσης} \cdot 20 \log_{10} \frac{r_1}{r_2}$$

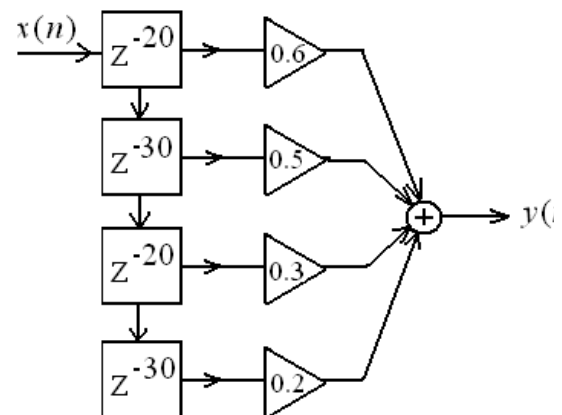
Η εξίσωση διαφοράς είναι $y(n)=x(n-40)+0.667x(n-60)$ και η κρουστική απόκριση του συστήματος σχεδιάζεται παρακάτω.



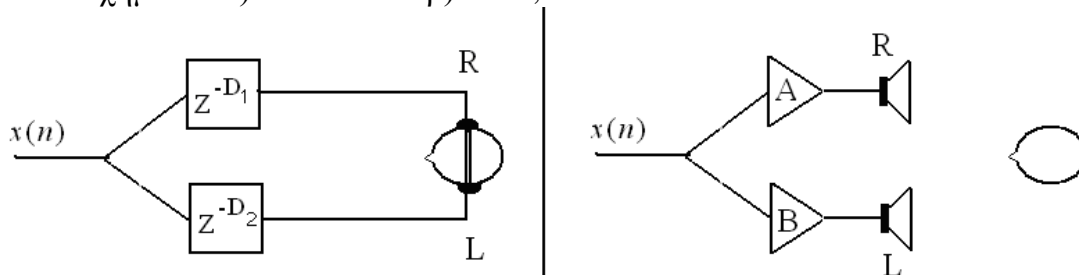
Παράδειγμα: Σχεδιάστε ένα σύστημα το οποίο θα είχε την παρακάτω κρουστική απόκριση. Χρησιμοποιήστε όσο το δυνατόν μικρότερο buffer για τα στοιχεία καθυστέρησης.



Απάντηση: Οι σχετικές διαφορές άφιξης από τον ένα παλμό στον άλλο είναι $D_1=20$, $\Delta D_2=50-20=30$, $\Delta D_3=70-50=20$ και $\Delta D_4=100-70=30$. Οι τιμές του κέρδους σε κάθε κλάδο καθυστέρησης είναι 0.6, 0.5, 0.3 και 0.2. Η εξίσωση διαφοράς του συστήματος είναι $y(n)=0.6x(n-20)+0.5x(n-50)+0.3x(n-70)+0.2x(n-100)$. Το σύστημα φαίνεται στο διπλανό σχήμα.



Παράδειγμα: Που τοποθετείται το ακουστικό είδωλο σε σχέση με τον ακροατή στο παρακάτω σχήμα αν α) $D_1 > D_2$ και β) $A > B$;

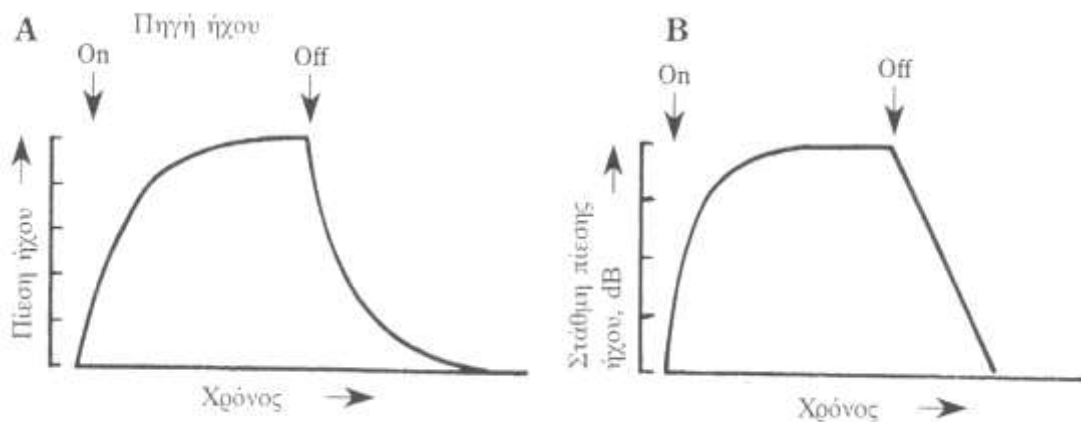


Απάντηση: α) Η σχέση $D_1 > D_2$ υποδηλώνει ότι ο ήχος φτάνει καθυστερημένα στο δεξί αυτί σε σχέση με το αριστερό. Επομένως ο ακροατής αντιλαμβάνεται το ακουστικό είδωλο στα αριστερά του. β) Αφού $A > B$ πάει να πει ότι το δεξί ηχείο παίζει πιο δυνατά από το αριστερό και επομένως το ακουστικό είδωλο τοποθετείται δεξιά σε σχέση με τον ακροατή.

4.8 Προσθήκη Αντήχησης

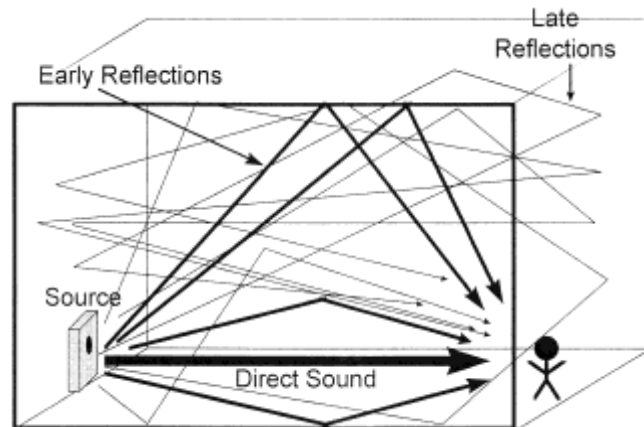
Αντήχηση είναι το χαρακτηριστικό γνώρισμα ενός χώρου όπου όταν μια πηγή ήχου διακοπεί απότομα ο ήχος δεν διακόπτεται αλλά συνεχίζει να υφίσταται για ένα μικρό ή μεγάλο χρονικό διάστημα μέχρι εν τέλει να εξασθενήσει τελείως. Την αντήχηση ενός χώρου μπορούμε να την αντιληφθούμε επίσης διεγείροντας το χώρο με ένα κρουστικό ήχο όπως ένα παλαμάκι ή μια πιστολιά (η τελευταία μέθοδος χρησιμοποιείται συχνά κατά την ακουστική μελέτη κλειστών χώρων, με άσφαιρα πιστόλια). Στο σχήμα 4.13 φαίνεται η καμπύλη εξασθένησης του ήχου από τη στιγμή διακοπής της ηχητικής πηγής. Η καμπύλη εξασθένησης του ήχου είναι εκθετική σε γραμμική κλίμακα (Σχήμα Α), γεγονός που εξασφαλίζει ότι σε λογαριθμική κλίμακα η εξασθένηση είναι ευθεία γραμμή (Σχήμα Β).

Χρόνος αντήχησης λέγεται το χρονικό διάστημα που μεσολαβεί από το σταμάτημα της πηγής μέχρι ο ήχος να εξασθενήσει 60 dB. Επειδή $20 \log 1000 = 60$, καταλαβαίνουμε ότι μείωση 60 dB σημαίνει εξασθένηση της ηχητικής πίεσης στο 0.1% της αρχικής της τιμής (1/1000).



Σχήμα 4.13: Η ανάπτυξη και εξασθένηση του ήχου μέσα σε ένα δωμάτιο: Α) Η κατακόρυφη κλίμακα είναι σε γραμμικές μονάδες πίεσης ήχου, Β) η κατακόρυφη κλίμακα είναι σε λογαριθμικές μονάδες (dB).

Η αντήχηση και η ανάκλαση έχουν στενή σχέση μεταξύ τους. Ουσιαστικά, η αντήχηση δημιουργείται από πολλές διαφορετικές ανακλάσεις που καταφθάνουν σε διαφορετικά χρονικά διαστήματα. Επειδή οι ανακλάσεις είναι πολλές και απέχουν πολύ μικρό χρονικό διάστημα μεταξύ τους, η αντήχηση γίνεται αντιληπτή σαν μια ομαλή εξασθένηση του ήχου. Όπως και στην περίπτωση της ανάκλασης που εξετάστηκε προηγουμένως, όσο πιο πολύ καθυστερεί μια ανάκλαση τόσο περισσότερο εξασθενημένη καταφθάνει στο δέκτη. Κατά τη μελέτη της αντήχησης συνηθίζεται να τη διαχωρίζουμε χρονικά σε τρία διαφορετικά τμήματα, όπως φαίνεται και από το Σχήμα 4.14: Τον άμεσο ήχο (direct sound), τις πρώτες ανακλάσεις (early reflections) και το εναπομένον τμήμα της αντήχησης (καθυστερημένες ανακλάσεις -late reflections)



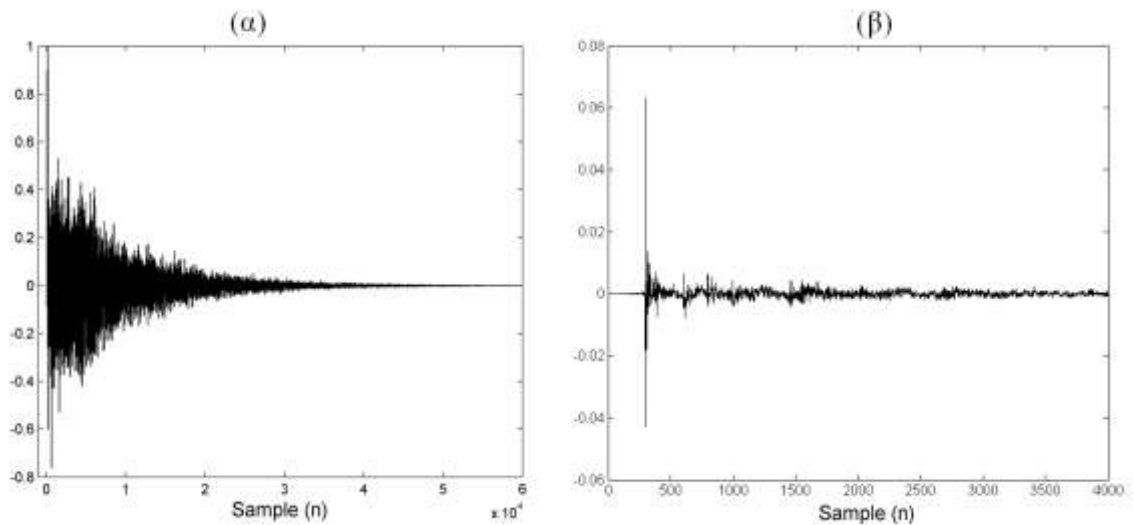
Σχήμα 4.14: Αναπαράσταση της δημιουργίας αντήχησης σε ένα κλειστό χώρο

Όπως φαίνεται και από το σχήμα 4.14, τμήματα του ήχου φτάνουν στον ακροατή αφού έχουν ανακλαστεί σε περισσότερες από μία επιφάνειες του χώρου. Όσο πιο πολλές είναι οι επιφάνειες του χώρου με τις οποίες ανακλάται ο ήχος, τόσο με λιγότερη ενέργεια καταφθάνει ο ανακλώμενος ήχος με αποτέλεσμα να έχουμε την ομαλή αυτή εξασθένηση που είπαμε προηγουμένως.

Η αντήχηση ενός χώρου περιγράφεται από την *κρουστική απόκριση* (impulse response) του χώρου. Η κρουστική απόκριση δεν είναι τίποτα άλλο από το ηχητικό σήμα που θα ηχογραφήσουμε με ένα μικρόφωνο σε κάποια θέση του χώρου όταν σε κάποια άλλη θέση διεγείρουμε το χώρο με ένα κρουστικό ήχο, δηλαδή με κάποιο ήχο πολύ μικρής διάρκειας όπως μια πιστολιά ή το σκάσιμο ενός μπαλονιού (βλ. σχήμα 4.15).

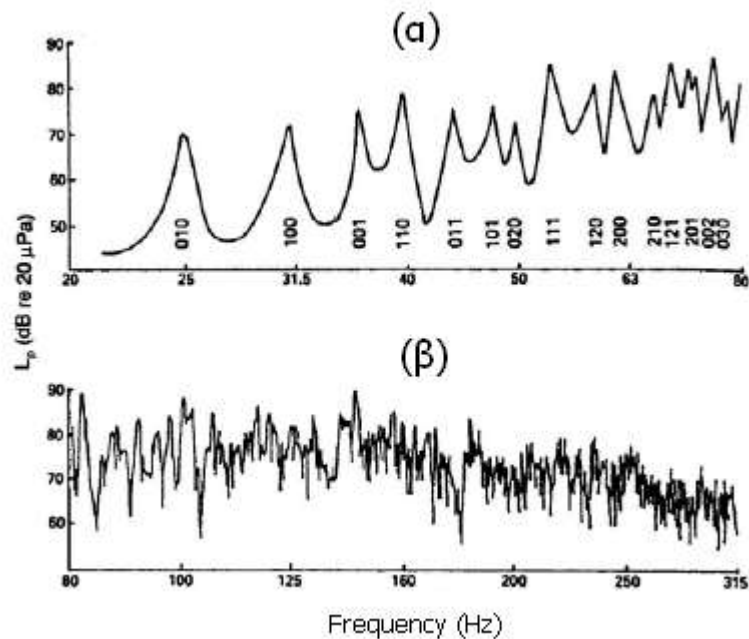


Σχήμα 4.15: Μέτρηση της κρουστικής απόκρισης ενός κλειστού χώρου



Σχήμα 4.16. Κρουστική απόκριση ενός (α) μεγάλου κλειστού χώρου με πολύ αντήχηση και (β) ενός μικρού κλειστού χώρου με λίγη αντήχηση.

Από την κρουστική απόκριση ενός χώρου παίρνουμε την απόκριση συχνότητας του χώρου κάνοντας χρήση του μετασχηματισμού Fourier. Μια τυπική απόκριση συχνότητας ενός κλειστού χώρου φαίνεται στο σχήμα 4.17. Όπως παρατηρεί κανείς, ένας κλειστός χώρος εισάγει μεγάλες συχνοτικές διακυμάνσεις, ενισχύοντας και εξασθενώντας επιλεκτικά κάποιες συχνότητες. Το φαινόμενο αυτό είναι γνωστό και ως φασματικός χρωματισμός, και είναι ιδιαίτερα αισθητό στις χαμηλές συχνότητες και στους μικρούς σχετικά χώρους.



Σχήμα 4.17: Τυπική απόκριση συχνότητας ενός κλειστού χώρου.

4.8.1 Ο κλειστός χώρος σαν γραμμικό σύστημα

Την έννοια της κρουστικής απόκρισης την έχουμε δει και σε προηγούμενη παράγραφο. Είναι η απόκριση ενός συστήματος όταν αυτό διεγερθεί με έναν απειροστό παλμό, εάν στην

είσοδό του δηλαδή θεωρήσουμε ένα σήμα πάρα πολύ μικρής χρονικής διάρκειας. Στο διακριτό χρόνο, ο απειροστός παλμός προσεγγίζεται από ένα σήμα που συμβολίζεται με $\delta(n)$ και το οποίο έχει όλες τις τιμές του μηδενικές εκτός από μία. Συγκεκριμένα, ο μοναδιαίος παλμός, όπως λέγεται, ορίζεται ως εξής:

$$\delta(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \quad (4.17)$$

Είναι περίεργο πως ένα τόσο απλό σήμα μπορεί να είναι τόσο χρήσιμο και σημαντικό στην ανάλυση και μοντελοποίηση συστημάτων. Ουσιαστικά, η σημασία της απόκρισης μοναδιαίου παλμού οφείλεται στο ότι αν γνωρίζουμε πώς το σύστημα θα αντιδράσει στο μοναδιαίο παλμό, τότε μπορούμε να ξέρουμε πως θα αντιδράσει σε οποιοδήποτε σήμα που εμφανιστεί στην είσοδό του. Η πληροφορία λοιπόν που παρέχεται από την κρουστική απόκριση αξιοποιείται για να κατασκευαστεί η έξοδος του συστήματος σε πιο πολύπλοκα σήματα, όπως σήματα φωνής και μουσικά σήματα.



Σχήμα 4.18. Διεγείροντας ένα σύστημα με το μοναδιαίο δείγμα λαμβάνουμε στην έξοδο την κρουστική απόκριση του συστήματος.

Το πώς γίνεται αυτό θα το δείξουμε παρακάτω. Προς το παρόν, ας δούμε λίγο την έννοια του μετατοπισμένου κατά k δείγματα μοναδιαίου παλμού $\delta(n-k)$

$$\delta(n-k) = \begin{cases} 1, & n = k \\ 0, & n \neq k \end{cases} \quad (4.18)$$

Επειδή μιλάμε για γραμμικά συστήματα, αμετάβλητα προς τη χρονική μετατόπιση (ΓΧΑ), ξέρουμε ότι αν η απόκριση του συστήματος είναι $h(n)$ για είσοδο $\delta(n)$, τότε για είσοδο $\delta(n-k)$ η έξοδος του συστήματος θα είναι $h(n-k)$. Με άλλα λόγια, αν ο μοναδιαίος παλμός καθυστερήσει k δείγματα να εισαχθεί στο σύστημα, η κρουστική απόκριση του συστήματος θα εμφανιστεί αυτούσια αλλά θα καθυστερήσει και αυτή k δείγματα να εμφανιστεί. Με βάση λοιπόν την έννοια του κρουστικού παλμού και του ΓΧΑ συστήματος, μπορούμε να εκφράσουμε οποιοδήποτε διακριτό σήμα σαν άθροισμα κατάλληλα ζυγισμένων μετατοπισμένων κρουστικών παλμών. Έστω για παράδειγμα ένα σήμα πέντε τιμών $x(n)=[1 \ 2 \ 3 \ 2 \ 1]$, όπου το n παίρνει τις τιμές 0,1,2,3 και 4. Παρατηρούμε ότι το $x(n)$ μπορεί να γραφεί και ως:

$$x(n) = \delta(n) + 2\delta(n-1) + 3\delta(n-2) + 2\delta(n-3) + \delta(n-4). \quad (4.19)$$

Αφού λοιπόν το σύστημα είναι γραμμικό και αμετάβλητο ως προς τη μετατόπιση, η έξοδος $y(n)$ του συστήματος για είσοδο $x(n)$ μπορεί να εκφραστεί ως:

$$y(n) = 1h(n) + 2h(n-1) + 3h(n-2) + 2h(n-3) + 1h(n-4) \quad (4.20)$$

όπου $h(n)$ είναι η κρουστική απόκριση του συστήματος.

Πρέπει να σημειωθεί ότι το σύστημα το οποίο ορίζεται κάθε φορά μέσω της κρουστικής απόκρισης $h(n)$ εξαρτάται τόσο από τη θέση διέγερσης (θέση πυροβολισμού) όσο και από τη θέση καταγραφής της κρουστικής απόκρισης (θέση του μικροφώνου). Γίνεται λοιπόν φανερό ότι ο ορισμός του συστήματος εδώ περιγράφει την ακουστική ζεύξη μεταξύ δύο σημείων στον κλειστό χώρο. Αν η θέση πυροβολισμού ή η θέση ηχογράφησης αλλάξει, θα αλλάξει και η κρουστική απόκριση του συστήματος.

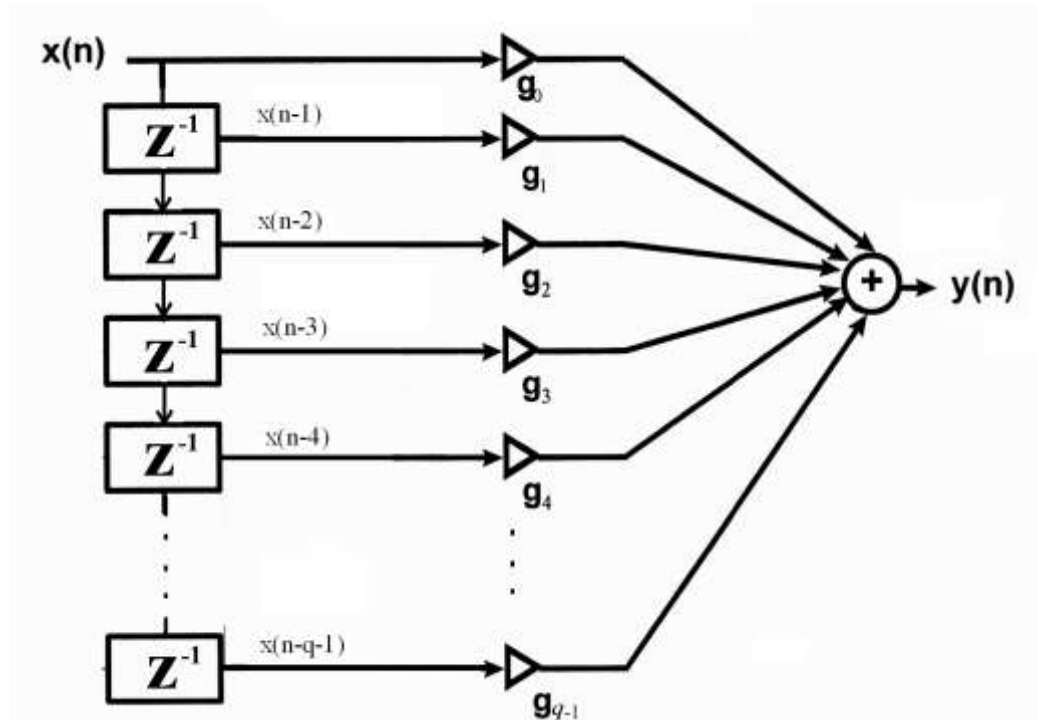
Γνωρίζοντας λοιπόν την κρουστική απόκριση του χώρου $h(n)$, μπορούμε θεωρητικά να καθορίσουμε επακριβώς το σήμα που θα κατέγραφε το μικρόφωνό μας αν στη θέση διέγερσης είχαμε ένα οποιοδήποτε σύνθετο σήμα $x(n)$ αντί για πυροβολισμό. Συγκεκριμένα, το σήμα $y(n)$ που θα κατέγραφε το μικρόφωνο μπορεί να υπολογιστεί μέσω της χρήσης της συνέλιξης (convolution), σύμφωνα με τη σχέση

$$y(n) = h(n) \cdot x(n) = \sum_{k=0}^{\infty} h(k)x(n-k) \quad (4.21)$$

Κάνοντας λοιπόν την παραπάνω πράξη προσθέτουμε ουσιαστικά την αντήχηση του χώρου στο ηχητικό μας σήμα. Είναι προφανές ότι όσο πιο μεγάλος ο χρόνος αντήχησης που θέλουμε να αποκτήσει το σήμα μας, τόσο μεγαλύτερο και το μήκος της κρουστικής απόκρισης που θα χρησιμοποιηθεί. Αν λοιπόν υποθέσουμε ότι η κρουστική απόκριση του συστήματος είναι ένα σήμα διάρκειας q δειγμάτων $h(n)=[g_0 \ g_1 \ \dots \ g_{q-1}]$, τότε η παραπάνω πράξη μπορεί να γραφεί με τη μορφή εξίσωσης διαφοράς ενός FIR φίλτρου ως (βλ. παράγραφο 4.1),

$$y(n) = \sum_{k=0}^{q-1} g_k x(n-k) = g_0 x(n) + g_1 x(n-1) + \dots + g_{q-1} x(n-q+1) \quad (4.22)$$

Βλέπουμε λοιπόν ότι το σήμα εξόδου $y(n)$ εξαρτάται από την παροντική τιμή του σήματος εισόδου $x(n)$, και από τις παρελθοντικές τιμές $x(n-1)$, $x(n-2)$, ..., $x(n-q+1)$, ακριβώς όπως είδαμε στην περίπτωση της ανάκλασης, με τη διαφορά ότι τώρα έχουμε προφανώς παραπάνω από μία ανακλάσεις. Η ένταση της κάθε ανάκλασης καθορίζεται από την τιμές $g_0, g_1, g_2, \dots, g_{q-1}$ οι οποίες προκύπτουν από το αντίστοιχο δείγμα της κρουστικής απόκρισης. Στο Σχήμα 4.19 φαίνεται πως μπορούμε να αναπαραστήσουμε την πράξη της συνέλιξης κάνοντας χρήση από τα δομικά στοιχεία της πρόσθεσης, της καθυστέρησης και του πολλαπλασιασμού. Πρέπει ο σπουδαστής να παρατηρήσει ότι για μήκος κρουστικής απόκρισης q δειγμάτων χρειαζόμαστε q στοιχεία πολλαπλασιασμού (τρίγωνα) και $q-1$ στοιχεία καθυστέρησης (z^{-1}). Επίσης, αν το πλήθος δειγμάτων που απαρτίζουν το σήμα εισόδου είναι N και το πλήθος δειγμάτων της κρουστικής απόκρισης είναι q , τότε, το προϊόν της συνέλιξης της εισόδου $x(n)$ με την κρουστική απόκριση $h(n)$ θα είναι $N+q-1$. Βάζοντας δηλαδή μια είσοδο σε ένα σύστημα η έξοδος που θα πάρουμε θα υποστεί επιμήκυνση ίση με το πλήθος δειγμάτων της κρουστικής απόκρισης μείον 1.



Σχήμα 4.19. Προσθήκη αντήχησης με χρήση της πράξης της συνέλιξης. Για κρουστική απόκριση q δειγμάτων απαιτούνται q πολλαπλασιασμοί και $q-1$ στοιχεία καθυστέρησης.

Ένας επεξεργαστής για να ανταπεξέλθει στις απαιτήσεις της συνέλιξης θα πρέπει να προλαβαίνει να κάνει q πολλαπλασιασμούς ανά περίοδο δειγματοληψίας. Επίσης, θα πρέπει να διαθέτει αρκετό μεγάλη χωρητικότητα buffer ώστε ανά πάσα στιγμή να μπορεί να ανακαλεί όλες τις $q-1$ παρελθοντικές τιμές του σήματος εισόδου $x(n-1)$, $x(n-2)$, ..., $x(n-q-1)$ αλλά και για να αποθηκεύει όλες τις q τιμές των συντελεστών g_0, g_1, \dots, g_{q-1} . Γίνεται λοιπόν φανερό ότι οι απαιτήσεις μνήμης και ταχύτητας κύκλου λειτουργίας αυξάνονται σημαντικά για μεγάλους χρόνους αντήχησης. Για αυτό το λόγο, τα reverb plugins, τα οποία είναι διαδεδομένα για την προσθήκη αντήχησης σε ένα ψηφιακό σταθμό εργασίας, αυξάνουν σημαντικά τις απαιτήσεις σε CPU και μνήμη RAM στους προσωπικούς υπολογιστές. Η προσθήκη αντήχησης με βάση την χρήση της συνέλιξης λέγεται *convolution reverb*.

Παράδειγμα: Η κρουστική απόκριση ενός χώρου έχει χρόνο αντήχησης $RT_{60}=2$ sec και χρησιμοποιείται για τη λειτουργία ενός convolution reverb με συχνότητα δειγματοληψίας $F_s=44100$. Θεωρώντας ότι το σήμα είναι μονοφωνικό

A) Πόσοι πολλαπλασιασμοί πρέπει να γίνονται ανά περίοδο δειγματοληψίας για την υλοποίηση της εφαρμογής;

B) Να βρεθεί ο κύκλος λειτουργίας T_{inst} του επεξεργαστή που απαιτείται για την υλοποίηση της εφαρμογής.

Γ) Η τάξη κβαντισμού του ηχητικού σήματος και των συντελεστών της κρουστικής απόκρισης είναι 16 και 12 bit αντίστοιχα. Να υπολογιστεί το μέγεθος της μνήμης του buffer που απαιτείται για την υλοποίηση της συνέλιξης.

Απάντηση: A) Είναι προφανές ότι η κρουστική απόκριση θα απαρτίζεται από

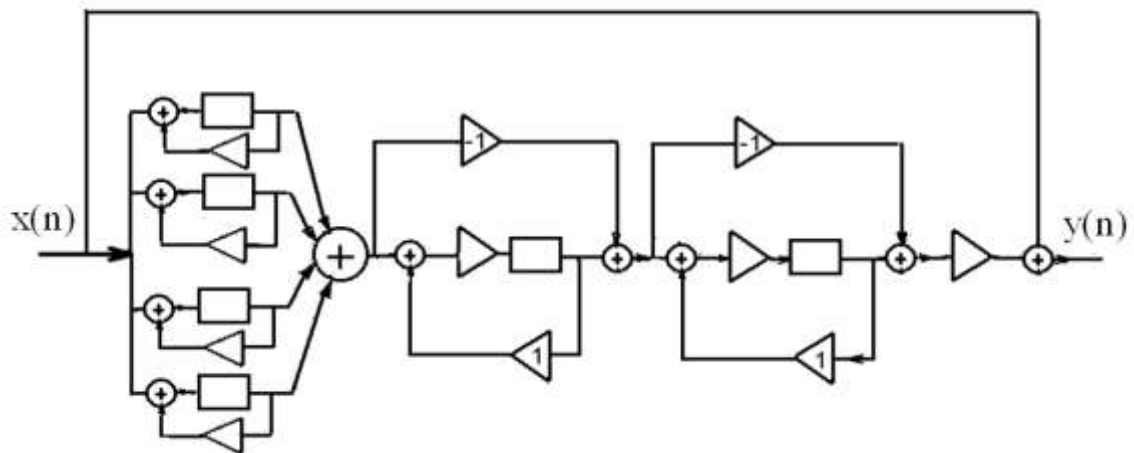
44100·2=88200 δείγματα. Αυτό σημαίνει ότι απαιτούνται 88200 πολλαπλασιασμοί ανά περίοδο δειγματοληψίας.

Β) Ως κύκλος λειτουργίας ορίζεται ο χρόνος που απαιτείται για την υλοποίηση μιας πράξης από τον επεξεργαστή. Η περίοδος δειγματοληψίας είναι $T_s=2.27 \cdot 10^{-5}$ sec σε αυτό το διάστημα ο επεξεργαστής θα πρέπει να προλαβαίνει να κάνει 88200 πολλαπλασιασμούς. Άρα,

$$T_{inst} = \frac{T_s}{88200} \Rightarrow T_{inst} = \frac{2.27 \cdot 10^{-5}}{88200} = 2,57 \cdot 10^{-10} = 0,257 nsec .$$

Γ) Χρειαζόμαστε $88199 \cdot 16=1411184$ bits για την αποθήκευση των παρελθοντικών τιμών του σήματος εισόδου $x(n-1), x(n-2), \dots, x(n-88199)$ και $88200 \cdot 12=1058400$ bits για τους συντελεστές των φίλτρων. Επομένως συνολικά χρειαζόμαστε μνήμη 2,47 Mbits.

Παρατήρηση: Από το παραπάνω παράδειγμα μπορούμε να δούμε ότι για την προσθήκη χρόνου αντήχησης 2 sec υπό τη μορφή ενός FIR φίλτρου απαιτείται ταχύτητα επεξεργαστή $\frac{1}{0.257} \cdot 10^9 = 3,89 GHz (!)$. Αν και θεωρητικά ένας σημερινός επεξεργαστής μπορεί να ανταπεξέλθει σε αυτές τις απαιτήσεις, είναι φανερό ότι θα πρέπει να διατίθεται σχεδόν εξολοκλήρου για την πραγματοποίηση της εφαρμογής σε πραγματικό χρόνο. Επειδή λοιπόν οι υπολογιστικές απαιτήσεις γίνονται απαγορευτικές για μεγάλους χρόνους αντήχησης, έχουν προταθεί αλγόριθμοι προσθήκης αντήχησης που χρησιμοποιούν IIR δομές αντί για FIR. Όπως είδαμε στο κεφάλαιο 4.1, σε μια τυπική IIR δομή η έξοδος εξαρτάται και από παρελθοντικές τιμές του σήματος εξόδου ($y(n)$), εκτός από παρελθοντικές τιμές του σήματος εισόδου ($x(n)$). Αυτό έχει το πλεονέκτημα ότι μπορούμε να κατασκευάζουμε ένα σύστημα με κρουστική απόκριση μεγάλης διάρκειας με χρήση πολύ λιγότερων στοιχείων καθυστέρησης και με πολύ μικρότερες υπολογιστικές απαιτήσεις. Ένα τέτοιο κύκλωμα φαίνεται και στο Σχήμα 3.20 όπου φαίνεται η χρήση κλάδων ανατροφοδότησης του σήματος.



Σχήμα 4.20: Κατασκευή ενός IIR φίλτρου που προορίζεται για προσθήκη αντήχησης στο ηχητικό σήμα. Τα κουτάκια αναπαριστούν στοιχεία καθυστέρησης ενώ τα τρίγωνα στοιχεία πολλαπλασιασμού.

4.9 Ισοστάθμιση και Αφαίρεση Αντήχησης

Η χαρακτηριστική επίδραση ενός κλειστού χώρου σε ένα ηχητικό κύμα το οποίο διαδίδεται μέσα σε αυτόν μπορεί να είναι τόσο επιθυμητή όσο και ανεπιθύμητη. Για παράδειγμα, σε μία συναυλία συμφωνικής μουσικής η αντήχηση που εισάγει ένας προσεκτικά σχεδιασμένος χώρος είναι επιθυμητή. Για την ακρίβεια, η επίδραση του χώρου είναι αναπόσπαστο κομμάτι της όλης αισθητικής και για αυτό κάποιοι μουσικοί χώροι είναι τόσο διάσημοι για την ακουστική τους όσο και οι μαέστροι οι οποίοι επιδίδονται μέσα σε αυτούς. Πολλές φορές ωστόσο, η επίδραση του χώρου σε ένα ηχητικό σήμα γίνεται ανεπιθύμητη. Ένας χώρος για παράδειγμα μπορεί να αλλοιώνει τόσο πολύ την ομιλία με αποτέλεσμα αυτή να μη γίνεται κατανοητή από κάποιον ακροατή. Εξαιτίας αυτής της επίδρασης που μπορεί να έχει ένας κλειστός χώρος σε ένα ηχητικό σήμα, έχουν προταθεί και εφαρμοστεί τεχνικές που, βασισμένες στην ψηφιακή επεξεργασία σήματος, αποσκοπούν στη μείωση ή και αφαίρεση της επίδρασης αυτής.

Ιδανικά, η ισοστάθμιση μπορεί να επιτύχει εξολοκλήρου αφαίρεση της επίδρασης του χώρου, πράξη που στην ουσία ισοδυναμεί με την πράξη της από-συνέλιξης. Στην απλούστερή της μορφή, δηλ. για μία πηγή και έναν δέκτη σε έναν κλειστό χώρο, η διαδικασία μπορεί να διατυπωθεί ως εξής: Έστω ότι η κρουστική απόκριση του χώρου για συγκεκριμένες θέσεις της πηγής και του δέκτη συμβολίζεται ως $h(n)$ και η αντίστοιχη απόκριση συχνότητας ως $H(f)$. Τότε, σκοπός της ισοστάθμισης είναι η εύρεση ενός αντίστροφου φίλτρου σχεδιασμένο ώστε η κρουστική του απόκριση $g(n)$ και η συχνοτική του απόκριση $G(f)$ να ικανοποιούν τις σχέσεις

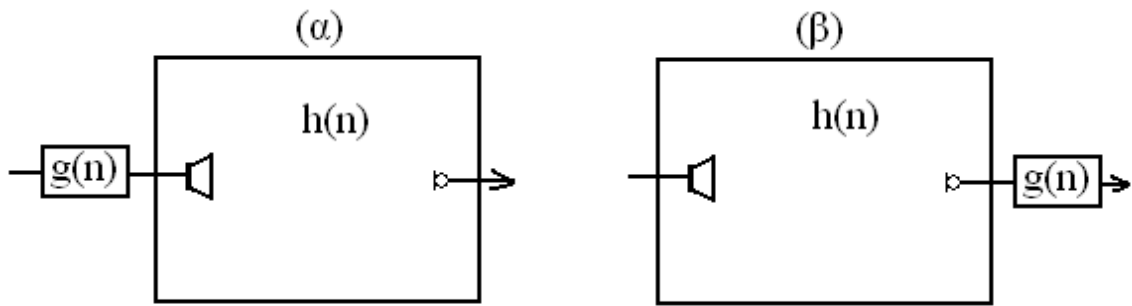
$$h(n) * g(n) = \delta(n) \quad (4.23)$$

και

$$H(f) \cdot G(f) = 1. \quad (4.24)$$

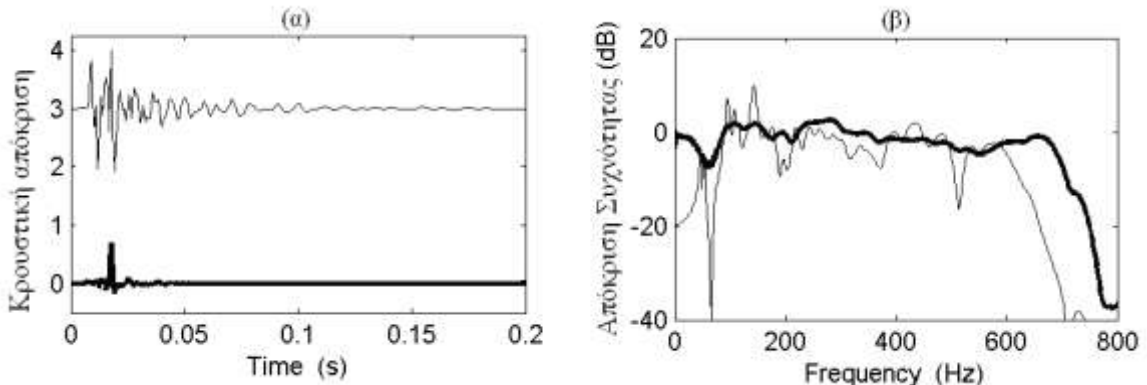
Ανάλογα με την εφαρμογή της ισοστάθμισης πριν ή μετά την αναπαραγωγή του ηχητικού σήματος σε έναν κλειστό χώρο διακρίνονται δύο κυρίες κατηγορίες εφαρμογών. Η πρώτη αφορά εφαρμογές όπως μία ζωντανή συναυλία ή την αναπαραγωγή του ηχητικού σήματος σε έναν κινηματογράφο, όπου η ισοστάθμιση επιτυγχάνεται με την προ-επεξεργασία του ηχητικού σήματος πριν αυτό αναπαραχθεί στον κλειστό χώρο από τα ηχεία, έτσι ώστε να αντισταθμιστούν οι παραμορφώσεις τις οποίες θα εισάγει ο χώρος στο ηχητικό σήμα (σχήμα 4.21(α)). Αντιθέτως, η δεύτερη κατηγορία αφορά εφαρμογές όπως η ηχογράφηση μουσικής ή ομιλίας, η τηλεδιάσκεψη και η κινητή τηλεφωνία, όπου η ισοστάθμιση επιτυγχάνεται με την επεξεργασία του καταγεγραμμένου σήματος για την απομάκρυνση των παραμορφώσεων οι οποίες εισήχθησαν κατά τη διάδοσή του σήματος στον κλειστό χώρο (σχήμα 4.21(β)).

Γενικά, ο ορισμός ενός αντίστροφου φίλτρου $g(n)$, δεδομένης της κρουστικής απόκρισης $h(n)$ δεν είναι πάντα δυνατός. Αυτό συμβαίνει διότι ενώ το αντίστροφο φίλτρο θα πρέπει να είναι αιτιατό και ευσταθές, η αντιστροφή ενός συστήματος μπορεί να οδηγήσει σε ένα φίλτρο που δεν τηρεί αυτές τις προϋποθέσεις και επομένως δεν είναι πρακτικά υλοποιήσιμο. Μπορούμε ωστόσο να βρούμε ένα φίλτρο $g(n)$ το οποίο αν και δεν κάνει τέλεια αντιστροφή, καταφέρνει ωστόσο να είναι βέλτιστο σε σχέση με κάποια κριτήρια.



Σχήμα 4.21: (α) Ισοστάθμιση με προ-επεξεργασία του σήματος πριν την αναπαραγωγή του στο χώρο και (β) ισοστάθμιση με επεξεργασία του καταγεγραμμένου σήματος μετά την αναπαραγωγή του στο χώρο.

Στο σχήμα 4.22 με τη λεπτή γραμμή φαίνεται στο (α) η κρουστική απόκριση ($h(n)$) και στο (β) η συχνοτική απόκριση ($|H(f)|$) όπως μετρήθηκαν για μια πηγή και ένα δέκτη μέσα σε ένα κλειστό χώρο. Για την ισοστάθμιση έγινε χρήση ενός FIR φίλτρου $g(n)$ το οποίο επέδρασε πριν την αναπαραγωγή από το ηχείο, σύμφωνα δηλαδή με τη διάταξη του σχήματος 4.21(α). Με την πιο παχιά γραμμή φαίνεται στο (α) το αποτέλεσμα της ισοστάθμισης ($h(n) * g(n)$) στο πεδίο του χρόνου και στο (β), στο πεδίο της συχνότητας ($|H(f) \cdot G(f)|$). Παρατηρούμε από το σχήμα (α) ότι το τώρα το σύνθετο σύστημα αντίστροφο φίλτρο – κλειστός χώρος απέκτησε μια κρουστική απόκριση που μοιάζει με ένα καθυστερημένο μοναδιαίο παλμό που θα θέλαμε να επιτύχουμε σύμφωνα με την εξίσωση (4.23). Το ότι ο παλμός είναι καθυστερημένος οφείλεται στο χρονικό διάστημα που χρειάζεται ο ήχος για να ταξιδέψει από το ηχείο στο μικρόφωνο. Από την άλλη, η συχνοτική απόκριση του σύνθετου συστήματος έγινε πιο επίπεδη, μειώνοντας σημαντικά τις διακυμάνσεις που ήταν εμφανείς πριν την ισοστάθμιση.



Σχήμα 4.22: (α) Κρουστική απόκριση πριν (λεπτή γραμμή) και μετά (παχιά γραμμή) την ισοστάθμιση (β) Συχνοτική απόκριση πριν (λεπτή γραμμή) και μετά (παχιά γραμμή) την ισοστάθμιση.

Η δυνατότητα ψηφιακής αναπαράστασης σημάτων και συστημάτων έδωσε νέες δυνατότητες στην αντιστάθμιση της επίδρασης που έχει ένας κλειστός χώρος στο ηχητικό σήμα. Η ανάγκη για ισοστάθμιση ωστόσο δεν εμφανίζεται μόνο σε κλειστούς χώρους, αλλά και στο ελεύθερο πεδίο (ελεύθερο πεδίο εννοούμε χώρο χωρίς ανακλάσεις), όταν για παράδειγμα το ηχητικό σήμα παράγεται από περισσότερες από μία ηχητικές πηγές. Στην επόμενη παράγραφο παρατίθεται ένα κείμενο από το ιντερνετ το οποίο περιγράφει ένα πρόβλημα που εμφανίζεται όταν δύο ηχεία που τροφοδοτούνται από το ίδιο σήμα

βρίσκονται σε διαφορετική απόσταση από κάποιο σημείο του χώρου με αποτέλεσμα να έχουμε φασματικούς χρωματισμούς σε αυτό το σημείο. Η ισοστάθμιση στο συγκεκριμένο σημείο του χώρου μπορεί να επιτευχθεί με χρήση απλών στοιχείων καθυστέρησης ώστε το σήμα από τα δύο ηχεία να φτάνει ταυτόχρονα στο σημείο που μας ενδιαφέρει.

4.9.1 Ο ρόλος της ψηφιακής καθυστέρησης, στο στήσιμο μικρομεγαφωνικής εγκατάστασης.

Ο περισσότερο καταληπτός και κατανοητός ήχος υφίσταται, όταν δύο άνθρωποι συζητούν πρόσωπο με πρόσωπο και αυτός είναι στην πλειοψηφία των περιπτώσεων ο ήχος που αναζητάμε, όταν στήνουμε ένα ηχητικό σύστημα για ζωντανή εμφάνιση (P.A.). Στην δύσκολη αυτή αποστολή έρχονται να μας βοηθήσουν τα ψηφιακά μηχανήματα καθυστέρησης (θα τα αναφέρουμε ddl από δω και πέρα χάριν συντομίας), που τα τελευταία χρόνια έχουν προσιτές τιμές, λόγω της ανάπτυξης των DSP.

Τρεις είναι οι διαφορετικοί στόχοι που εξυπηρετούν τα ddl. Ο πιο κοινός και ο σημαντικότερος είναι ο συγχρονισμός των ηχείων μεταξύ τους, για τον έλεγχο της υπερβάλλουσας αντήχησης και ηχούς, ο δεύτερος έχει να κάνει με τον περιορισμό του φαινομένου των φίλτρων τύπου κτένας (comb filter) και ο τελευταίος με την τοποθέτηση της ακουστικής εικόνας, έτσι ώστε ο ήχος να φαίνεται ότι έρχεται από τους εκτελεστές και όχι από τα ηχεία.

4.9.1.1 Συγχρονισμός Ηχείων

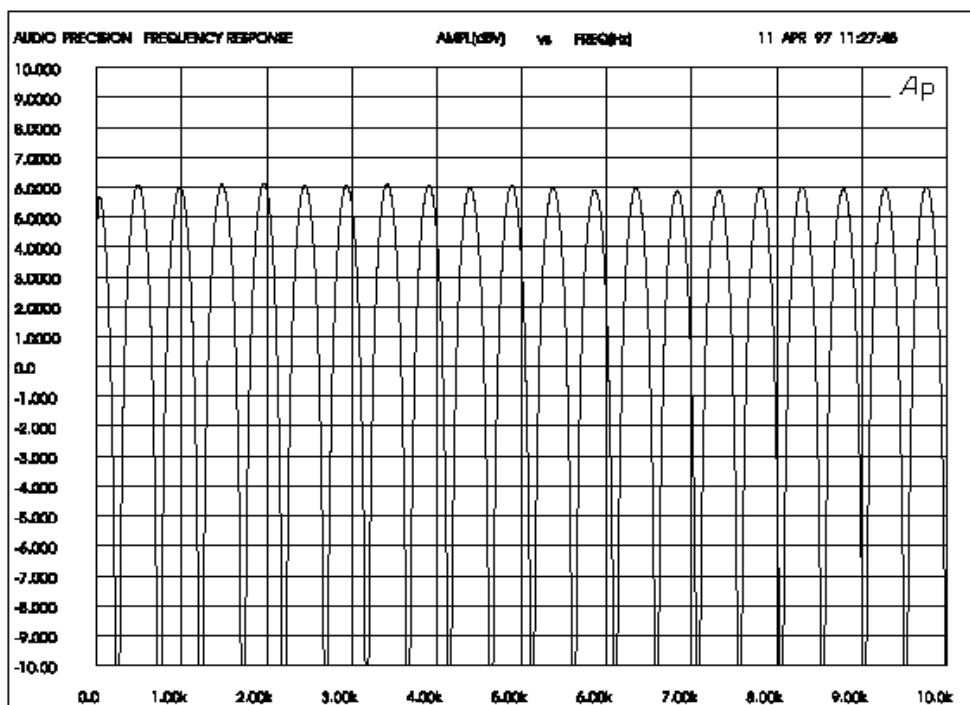
Η ταχύτητα του ήχου στον αέρα, ως γνωστόν είναι περίπου 344m/sec. Αυτό σημαίνει ότι ο ήχος για να διανύσει απόσταση ενός μέτρου, χρειάζεται 2,9msec. Τα ηλεκτρικά σήματα όμως, ταξιδεύουν σχεδόν ένα εκατομμύριο φορές γρηγορότερα, δηλαδή πρακτικά χωρίς καμία καθυστέρηση. Έτσι, όταν το ίδιο σήμα τροφοδοτήσει ηχεία τοποθετημένα σε απόσταση μεταξύ τους, ο ήχος που αυτά παράγουν, διανύοντας διαφορετικές αποστάσεις, λόγω του φαινομένου της αντήχησης και των πολλαπλών εικονικών ανακλάσεων, φτάνει στον ακροατή με μειωμένη καθαρότητα και καταληπτότητα. Σκοπός μας είναι να καθυστερήσουμε το σήμα που φτάνει στο ηχείο που είναι κοντότερα τοποθετημένο στον ακροατή. Αυτό το πετυχαίνουμε τροφοδοτώντας με έναν πολύ σύντομο παλμό το σύστημα των ηχείων και ρυθμίζοντας το ddl έτσι, ώστε όταν παίζουν και τα δύο ηχεία ταυτόχρονα, να ακούμε μόνον έναν παλμό να αναπαράγεται. Ο παλμός αυτός μπορεί να παραχθεί από κάποιο μηχανήμα μέτρησης, από ένα phase checker, ή ακόμα και από ένα κρουστό σε ένα μικρόφωνο. Θα πρέπει να λάβουμε πολύ σοβαρά υπόψη μας, ότι όλα τα ψηφιακά μηχανήματα εισάγουν λόγω των μετατροπών τους (converters), μία καθυστέρηση στο σήμα που κυμαίνεται από 0,9-5msec. Για τον λόγο αυτό, πρέπει όλα τα ψηφιακά μηχανήματα που είναι συνδεδεμένα στην αλυσίδα του σήματος, να είναι σε λειτουργία και όχι bypass όταν εκτελούμε αυτήν την εργασία του συγχρονισμού του συστήματος.

4.9.1.2 Παραμόρφωση φίλτρων τύπου κτένας

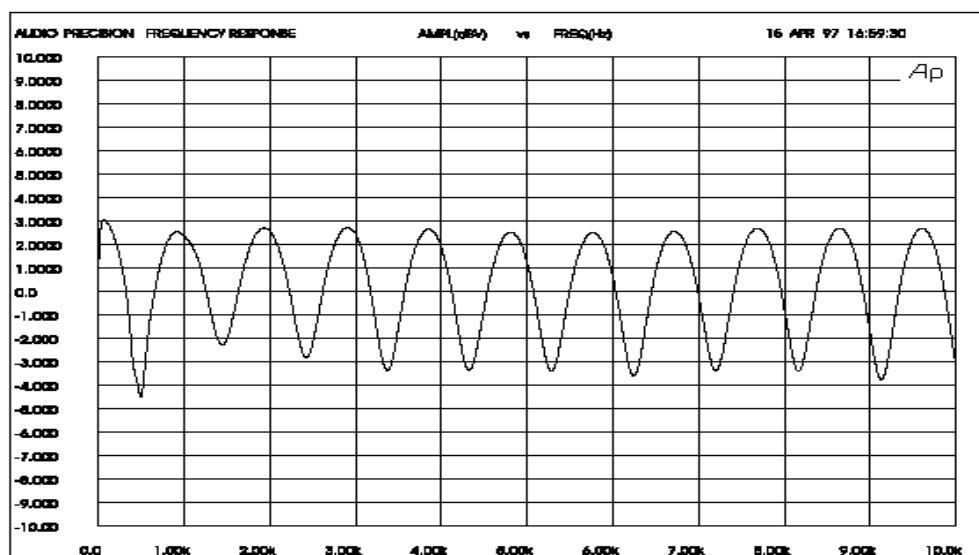
Η παραμόρφωση των φίλτρων τύπου κτένας, προκύπτει όταν ένα σήμα που έρχεται καθυστερημένο, προστίθεται στο αρχικό. Το γράφημα του αποτελέσματος της πρόσθεσης των δύο αυτών ηχητικών κυμάτων, μοιάζει με τα δόντια μίας κτένας και για αυτό έχει ονομαστεί comb filter. Υπάρχουν αρκετές περιπτώσεις που μπορούν να προκαλέσουν comb

filter. Παραδείγματος χάρη, όταν το ίδιο πρόγραμμα παίζεται από δύο ηχεία, ο χροιά που ακούμε από αυτό που είναι κοντύτερα σε εμάς, αλλοιώνεται από αυτό που είναι μακρύτερα, ή όταν λαμβάνουμε την ίδια πηγή με δύο μικρόφωνα σε κοντινή αλλά διαφορετική απόσταση μεταξύ τους, αλλά ακόμα και όταν επιστρέφουμε την έξοδο κάποιου ψηφιακού effects processor στην κονσόλα μας και την μιζάρουμε με το dry σήμα.

Στα παρακάτω διαγράμματα συχνότητας-έντασης παρατηρούμε ότι η επίδραση του comb filter όσον αφορά την στάθμη, εξαρτάται από την στάθμη του καθυστερημένου σήματος.



Σχήμα 4.23: Στο αρχικό σήμα έχει προστεθεί ο εαυτός του στη ίδια στάθμη, με καθυστέρηση 2msec. Η μέγιστη αύξηση είναι +6dB και η μέγιστη μείωση είναι $-\infty$ dB.



Σχήμα 4.24: Στο αρχικό σήμα έχει προστεθεί ο εαυτός του με διαφορά στάθμης -10dB και καθυστέρηση 2msec. Η μέγιστη αύξηση είναι +2,5dB και η μέγιστη μείωση είναι -4dB.

4.9.1.3 Υπολογισμός Συχνοτήτων του Comb Filter

Οι συχνότητες των εξάρσεων και των ακυρώσεων που παρουσιάζονται, εξαρτώνται από τον χρόνο καθυστέρησης (την διαφορά μεταξύ του απευθείας σήματος και του καθυστερημένου). Η συχνότητα της πρώτης ακύρωσης, δίνεται από τον τύπο

$$f = 1 / (2 \cdot \Delta t), \quad (3.9.3)$$

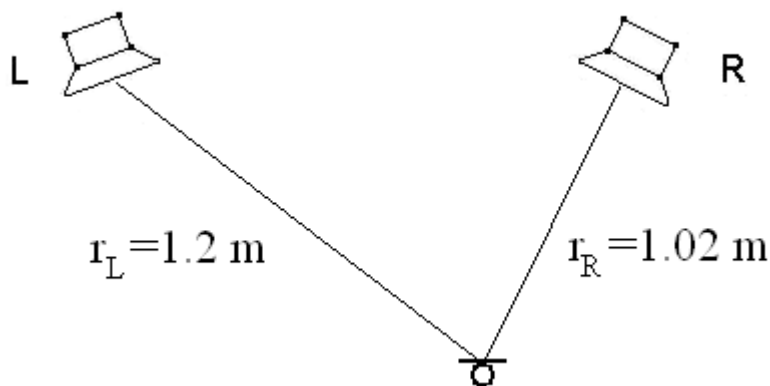
όπου Δt είναι ο χρόνος καθυστέρησης σε sec. Οι ακυρώσεις απέχουν μεταξύ τους $(1/\Delta t)$ Hz και οι εξάρσεις δημιουργούνται στην μέση μεταξύ δύο ακυρώσεων.

Ο παρακάτω πίνακας μας δείχνει, ότι οι συχνότητες αυξομειώσης έρχονται κοντύτερα μεταξύ τους, όσο το Δt αυξάνεται.

ΧΡΟΝΟΣ ΚΑΘΥΣΤ.=0,002sec		ΧΡΟΝΟΣ ΚΑΘΥΣΤ.=0,003sec		ΧΡΟΝΟΣ ΚΑΘΥΣΤ.=0,004sec	
Συχνότητα Ακύρωσης(Hz)	Συχνότητα Έξαρσης(Hz)	Συχνότητα Ακύρωσης(Hz)	Συχνότητα Έξαρσης(Hz)	Συχνότητα Ακύρωσης(Hz)	Συχνότητα Έξαρσης(Hz)
250	500	167	333	125	250
750	1000	500	667	375	500
1250	1500	833	1000	625	750
1750	2000	1167	1333	875	1000
2250	2500	1500	1667	1125	1250
2750	3000	1833	2000	1375	1500
3250	3500	2187	2333	1625	1750
3750	4000	2500	2667	1875	2000
4250	4500	2833	3000	2125	2250

Τα φίλτρα τύπου κτένας, δεν μπορούν να αντιμετωπιστούν με τα equalizers, αλλά μπορούν να μειωθούν αισθητά με τον συγχρονισμό των σημάτων μεταξύ τους.

Παράδειγμα: Πόση είναι η απόσταση μεταξύ διαδοχικών συχνοτικών ακυρώσεων στην παρακάτω διάταξη χωρίς χρήση στοιχείου ddl; Προτείνεται τη θέση και την ρύθμιση στοιχείου ddl για την εξάλειψη των φαινομένων χρωματισμού τύπου κτένας. Η ταχύτητα του ήχου είναι 344 m/sec.



Απάντηση: Παρατηρούμε ότι τα δύο ηχεία βρίσκονται σε διαφορετική απόσταση από το σημείο λήψης. Επειδή είναι το δεξί ηχείο που είναι πιο κοντά, θα πρέπει σε αυτό να εφαρμοστεί το στοιχείο ψηφιακής καθυστέρησης ddl . Ο χρόνος που χρειάζεται ο ήχος για να φτάσει από το αριστερό ηχείο στο σημείο λήψης είναι ίσος με $D_L = 1.2 / 344 = 3,49 msec$ ενώ για το δεξί ηχείο είναι ίσος με $D_R = 1.02 / 344 = 2,97 msec$.

Το ddl θα πρέπει επομένως να ρυθμιστεί στα $3,49 - 2,97 = 0,52 msec$. Η απόσταση μεταξύ διαδοχικών ακυρώσεων είναι $1/\Delta t = 1923 Hz$.

Παράδειγμα: Θεωρείστε μια απλή περίπτωση ανάκλασης όπως για παράδειγμα στην περίπτωση της παραγράφου 4.7 Περιγράψτε τι συμβαίνει, όσον αφορά φαινόμενα χρωματισμού φίλτρου χτένας, όσο η πηγή και ο δέκτης απομακρύνονται από το επίπεδο ανάκλασης.

Απάντηση: Στην περίπτωση της ανάκλασης, επειδή ο άμεσος ήχος καταφτάνει νωρίτερα από τον ανακλώμενο έχουμε φαινόμενα χρωματισμού φίλτρου χτένας, ακριβώς όπως και στην περίπτωση των δύο ηχείων που περιγράφηκε προηγουμένως. Επειδή η απόσταση μεταξύ διαδοχικών ακυρώσεων είναι αντιστρόφως ανάλογη της χρονικής διαφοράς άφιξης των δύο ήχων ($1/\Delta t$), είναι προφανές ότι όσο απομακρύνουμε την πηγή από το επίπεδο ανάκλασης, η χρονική διαφορά αύξησης αυξάνεται και επομένως η απόσταση μεταξύ διαδοχικών ακυρώσεων ελαττώνεται (βλ. παρακάτω σχήμα). Ένα άλλο χαρακτηριστικό που πρέπει να σχολιάσουμε αφορά τη μέγιστη μείωση που έχουμε στις συχνότητες ακύρωσης. Όσο απομακρύνουμε πηγή και δέκτη από το επίπεδο ανάκλασης, ο ανακλώμενος ήχος φτάνει σε όλο και μικρότερη ένταση σε σχέση με τον άμεσο ήχο, επομένως η μείωση που έχουμε στις συχνότητες ακύρωσης ελαττώνεται (βλ ως παράδειγμα σχήματα 3.23 και 3.24).



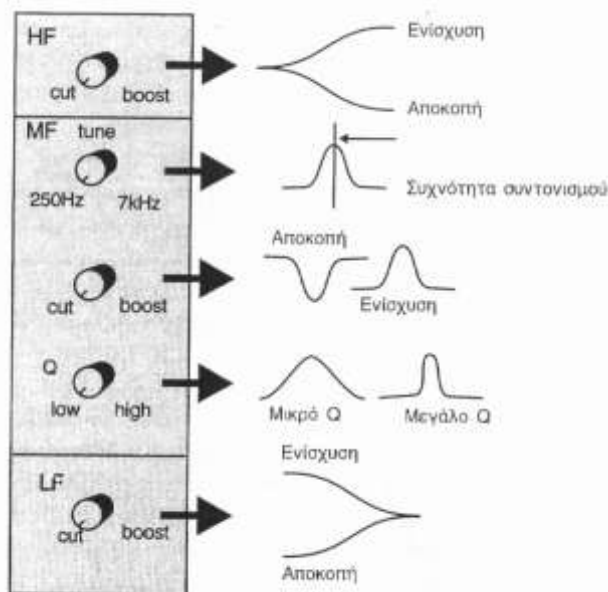
Σχήμα 4.25: Απομακρυνόμενοι από το επίπεδο ανάκλασης ο ανακλώμενος ήχος φτάνει όλο και καθυστερημένος και πιο εξασθενημένος με αποτέλεσμα οι εξάρσεις και οι ακυρώσεις στο φάσμα να γίνονται όλο και πιο πυκνές, ενώ η μέγιστη μείωση στις συχνότητες ακύρωσης γίνεται όλο και πιο μικρή.

5. EQUALISERS

‘Equalizers’, ή ‘EQ’, ή ‘ισοσταθμιστές’, ή ‘εξισορροπητές’ ή ‘ρυθμιστές ελέγχου τόνου’ ή απλά ‘φίλτρα’, είναι μερικές από τις ονομασίες που επικρατούν για το πολύτιμο αυτό εργαλείο του μουσικού και του ηχολήπτη. Ένα equalizer είναι ένα σύστημα που έχει ως σκοπό να μεταβάλλει τα φασματικά χαρακτηριστικά ενός ηχητικού σήματος. Η μορφή και ο τρόπος κατασκευής του equalizer μεταβάλλεται ανάλογα με τη χρήση του. Στο κεφάλαιο αυτό αναφέρονται οι βασικοί τύποι ισοσταθμιστών και επεξηγείται η χρήση τους ανάλογα με την περίπτωση.

5.1 Τυπικό equalizer 3 περιοχών τύπου κονσόλας

Όταν θέλουμε να τονίσουμε ή να εξασθενίσουμε κάποιες φασματικές συνιστώσες ενός ηχητικού σήματος ώστε να του δώσουμε κάποια τεχνητά χαρακτηριστικά. Πχ, αν εφαρμόσουμε ένα φίλτρο το οποίο εξασθενεί με τη συχνότητα, κάνουμε τον ήχο μας μουντό. Αυτό μπορεί να δώσει την αίσθηση ότι ο ήχος έρχεται από κάποιο άλλο δωμάτιο διαπερνώντας κάποιον τοίχο ή ότι έρχεται από κάποιο μακρινό μέρος, κάτι που θα είχε ως αποτέλεσμα την εξασθένηση των υψηλών συχνοτήτων. Από την άλλη, τονίζοντας τις μεσαίες συχνότητες και αποκόπτοντας τις χαμηλές και τις υψηλές, μπορούμε σε ένα σήμα φωνής να δώσουμε τα χαρακτηριστικά χροιά μιας τηλεφωνικής γραμμής. Μια τυπική μορφή ενός equalizer, όπως συναντάται στα κανάλια εισόδου μίας κονσόλας φαίνεται στο Σχήμα 5.1.



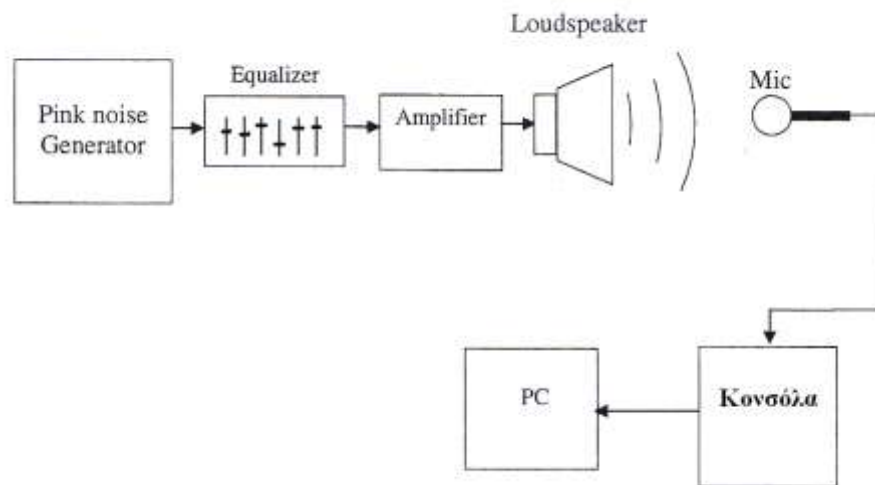
Σχήμα 5.1. Τυπική μορφή ενός equalizer όπως εμφανίζεται στα κανάλια εισόδου μίας κονσόλας

Το συγκεκριμένο equalizer είναι τριών περιοχών. Βλέπουμε ότι υπάρχει διαφορετική βαθμίδα για τις χαμηλές (LF), τις μεσαίες (MF) και τις υψηλές συχνότητες (HF). Με κατάλληλη χρήση του περιστροφικού ποτενσιόμετρου μπορούμε λοιπόν να ενισχύσουμε ή

να εξασθενήσουμε την κάθε περιοχή συχνοτήτων. Στις μεσαίες συχνότητες βλέπουμε ότι μπορούμε να ρυθμίσουμε περισσότερες παραμέτρους. Συγκεκριμένα, μπορούμε μέσω του ποτενσιόμετρου tune να ορίσουμε την κεντρική συχνότητα στην οποία θέλουμε να επέμβουμε. Για το συγκεκριμένο equalizer βλέπουμε ότι μπορούμε να κινηθούμε οπουδήποτε από τα 250Hz έως τα 7kHz. Μία άλλη παράμετρος που έχουμε στη διάθεσή μας είναι το εύρος ζώνης συχνοτήτων (Q) στο οποίο θέλουμε να επέμβουμε. Ρυθμίζοντας λοιπόν το Q μπορούμε να επέμβουμε σε μεγάλο εύρος συχνοτήτων (μικρό Q) ή σε στενό εύρος συχνοτήτων (μεγάλο Q), με αποτέλεσμα το φίλτρο μας να γίνεται πιο μυτερό ή πλατύ. Η παράμετρος Q λέγεται και *παράγοντας επιλεκτικότητας* του φίλτρου.

5.2 Το EQ ως ισοσταθμιστής

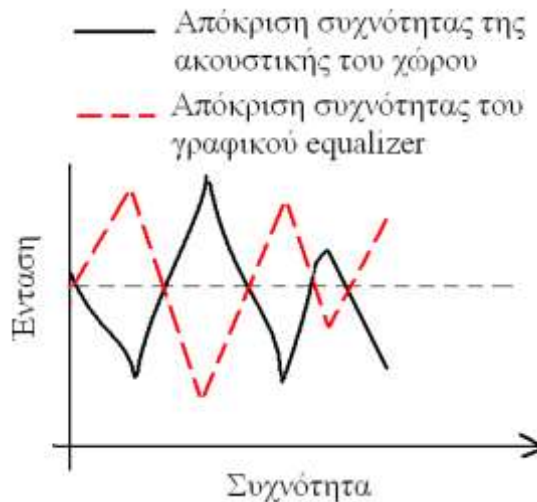
Αυτή είναι και η πιο συχνή χρήση τέτοιων επεξεργαστών. Ένα ηχητικό σήμα, ιδιαίτερα κατά τη λήψη του με μικρόφωνο είναι δυνατόν να αποκτήσει πλασματικά φασματικά χαρακτηριστικά, τα οποία δεν ακούγονται ‘φυσικά’ κατά την αναπαραγωγή. Για παράδειγμα, όταν ένας τραγουδιστής φέρνει το μικρόφωνο πολύ κοντά στο στόμα, δημιουργείται το γνωστό ως ‘proximity effect’ το οποίο έχει ως αποτέλεσμα να ενισχύονται πολύ τα μπάσα, οπότε και ο ηχολήπτης πρέπει να τα αφαιρέσει με κατάλληλη χρήση του ρυθμιστή χαμηλών συχνοτήτων. Για μια τέτοια περίπτωση, ένα equalizer σαν αυτό του Σχήματος 5.1 θα μπορούσε να κάνει τη δουλειά. Ωστόσο, υπάρχουν και άλλες περιπτώσεις που η αντιστάθμιση είναι πιο περίπλοκης φύσης. Εκεί χρειάζεται ένας άλλος τύπος equalizer, το γραφικό equalizer.



Σχήμα 5.2. Τυπική συνδεσμολογία κατά τη ρύθμιση του γραφικού equalizer

Τα γραφικά equalizer χρησιμοποιούνται για την αντιστάθμιση της αλλοίωσης που προκαλεί στις αναπαραγόμενες από τα ηχεία συχνότητες η ακουστική ενός χώρου. Το γραφικό equalizer παρεμβάλλεται μεταξύ του ηλεκτρονικού μέσου αναπαραγωγής (κονσόλα, μίκτης, cd player, κλπ) και ενισχυτή, με σκοπό να μεταβάλει τα φασματικά χαρακτηριστικά του σήματος, πριν αυτό αναπαραχθεί από τα ηχεία, ώστε κατά τη διάδοση του ηχητικού σήματος να αντισταθμιστεί όσο το δυνατόν περισσότερο η επίδραση του χώρου. Η συνήθης συνδεσμολογία που χρησιμοποιείται για τη ρύθμιση του γραφικού equalizer φαίνεται στο Σχήμα 5.2.

Στο Σχήμα 5.2 φαίνεται βασικά η συνδεσμολογία κατά τη φάση ρύθμισης του γραφικού Equalizer. Συγκεκριμένα, στη θέση του ηλεκτρονικού μέσου αναπαραγωγής βρίσκεται μία γεννήτρια ροζ θορύβου. Στο σημείο όπου θέλουμε να κάνουμε ισοστάθμιση βάζουμε ένα μικρόφωνο αναφοράς. Το μικρόφωνο οδηγεί το ηχητικό σήμα σε έναν υπολογιστή ο οποίος διαθέτει κάποιο φασματικό αναλυτή. Η φύση του ροζ θορύβου και η κατασκευή του φασματικού αναλυτή είναι τέτοια ώστε αν δεν έχουμε φασματική αλλοίωση από την ακουστική του χώρου, τότε θα δούμε μια ευθεία γραμμή παράλληλη στον άξονα των συχνοτήτων στο φασματικό αναλυτή. Στην πραγματικότητα ωστόσο, ανάλογα με τη φύση του χώρου θα δούμε σίγουρα μια μικρή ή μεγάλη απόκλιση από την ιδανική ευθεία, όπως για παράδειγμα η μαύρη καμπύλη με μέγιστα και ελάχιστα στο διάγραμμα συχνότητας έντασης του Σχήματος 5.3. Το γραφικό Equalizer είναι προφανές ότι θα πρέπει να εξασθενήσει τις περιοχές συχνοτήτων που έχουν ενισχυθεί από την επίδραση του χώρου και αντίστοιχα, να ενισχύσει τις περιοχές που έχουν εξασθενήσει. Βλέπουμε δηλαδή ότι οι ρυθμίσεις του equalizer είναι κατά κάποιο τρόπο αντικατροπικές σε σχέση με την απόκριση του χώρου, ως προς μία ευθεία γραμμή. (βλ. Σχήμα 5.3). Ο στόχος σε κάθε περίπτωση είναι η επίτευξη μιας επίπεδης απόκρισης συχνότητας (flat frequency response).

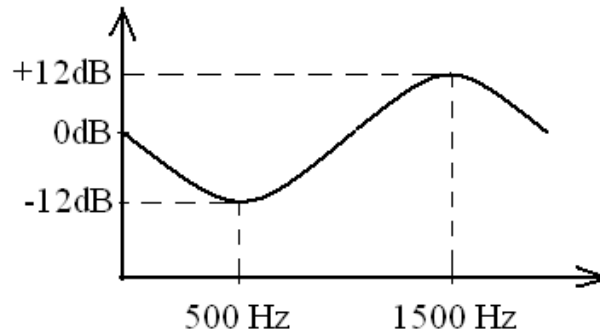


Σχήμα 5.3. Ισοστάθμιση με τη χρήση γραφικού equalizer

Μετά τη ρύθμιση του equalizer, οι ρυθμίσεις μένουν ως έχουν και στη θέση της γεννήτριας ροζ θορύβου μπαίνει η κονσόλα, το cd player ή οποιοδήποτε άλλο μέσο προορίζεται για την ηχητική αναπαραγωγή στο χώρο.

Για τη συγκεκριμένη μέθοδο θα πρέπει να αναφέρουμε ωστόσο μερικά μειονεκτήματα. Κατ' αρχάς, οι συχνότητες είναι σωστές μόνο στο σημείο και στη γειτονική περιοχή που ήταν τοποθετημένο το μικρόφωνο. Σε άλλα σημεία του χώρου πιθανόν να υπάρχει ακόμα ακουστική απόκλιση, ενώ δεν είναι απίθανο τα πράγματα να έχουν χειροτερέψει σε σχέση με πριν την ισοστάθμιση. Επίσης, τα ακουστικά χαρακτηριστικά της αίθουσας μπορούν να αλλάξουν δριμύτατα αν πχ. εισέλθει κόσμος ή ακόμα και αν αλλάξει η θερμοκρασία. Τέλος, η ισοστάθμιση με τη συγκεκριμένη μέθοδο είναι κατώτερης ποιότητας από αυτήν που μπορούμε να επιτύχουμε με την ψηφιακή τεχνική ισοστάθμισης που παρουσιάστηκε στην ενότητα 4.9. Παρόλα αυτά, χρησιμοποιείται ακόμα και σήμερα διότι βασίζεται εξολοκλήρου σε αναλογικό εξοπλισμό και έτσι αποφεύγεται η ανάγκη για μετατροπή του σήματος από αναλογικό σε ψηφιακό και αντίστροφα.

Παράδειγμα: Ένα ημιτονικό σήμα συχνότητας 500 Hz και πλάτους 2 V διέρχεται μέσα από ένα equalizer του οποίου η απόκριση συχνότητας φαίνεται στο παρακάτω σχήμα. Να βρεθεί το πλάτος του σήματος στην έξοδο του equalizer. Ποια είναι η αντίστοιχη τιμή του πλάτους για ένα ημιτονικό σήμα 1500 Hz το οποίο εισέρχεται με πλάτος 1 V στο ίδιο equalizer;



Απάντηση: Από το σχήμα φαίνεται ότι το σήμα συχνότητας 500 Hz εξερχόμενο από το equalizer εξασθενεί κατά 12 dB. Αυτό σημαίνει ότι αν V_{in} και V_{out} είναι το πλάτος του σήματος στην είσοδο και στην έξοδο του EQ τότε θα ισχύει:

$$20 \log_{10} \frac{V_{out}}{V_{in}} = -12 \text{ dB} ,$$

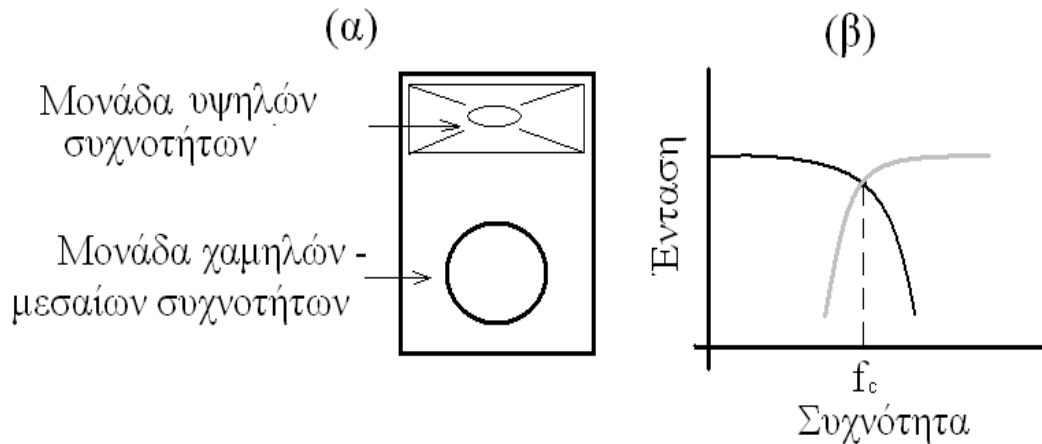
άρα, $\log_{10} \frac{V_{out}}{V_{in}} = -\frac{12}{20}$, και επομένως, $\frac{V_{out}}{V_{in}} = 10^{-\frac{12}{20}} \approx 0.25$. Τελικά βλέπουμε ότι

$V_{out} = 0.25 \cdot V_{in} = 0.5$ Volt. Αντίστοιχα, χρησιμοποιώντας τη γνωστή ιδιότητα των λογαρίθμων βλέπουμε ότι για το σήμα συχνότητας 1500 Hz θα ισχύει $\frac{V_{out}}{V_{in}} = \frac{1}{0.25} = 4$ και επομένως $V_{out} = 4$

Volt.

5.3 Κύκλωμα crossover στα ηχεία

Η αναπαραγωγή των διαφορετικών περιοχών συχνοτήτων στα ηχεία γίνεται συνήθως από διαφορετικές μονάδες. Άλλη μονάδα είναι υπεύθυνη για την αναπαραγωγή των χαμηλών συχνοτήτων (sub woofer), άλλη για την αναπαραγωγή των υψηλών συχνοτήτων (tweeter) ενώ συχνά χρησιμοποιείται ξεχωριστή μονάδα και για τις μεσαίες συχνότητες (mid-range). Υπάρχει επομένως η ανάγκη για τη χρήση φίλτρων με τα οποία διαχωρίζεται το συχνοτικό περιεχόμενο του προγράμματος, ώστε η κατάλληλα φιλτραρισμένη εκδοχή του ηχητικού σήματος να πηγαίνει στην κατάλληλη βαθμίδα του μεγαφώνου. Ένα παράδειγμα φαίνεται στο σχήμα 5.4 όπου φαίνεται ένα ηχείο δύο δρόμων το οποίο χρησιμοποιεί δύο διαφορετικές μονάδες, μία περιοχής χαμηλών - μεσαίων συχνοτήτων και μία υψηλής. Για ένα τέτοιο ηχείο χρειάζεται ένα low pass φίλτρο (χαμηλής διέλευσης) και ένα high pass φίλτρο (υψηλής διέλευσης). Η απόκριση συχνότητας των δύο φίλτρων απεικονίζεται στο σχήμα 5.4(β). Είναι φανερό ότι εκεί που μειώνεται η απόκριση του ενός φίλτρου θα πρέπει να ξεκινάει η απόκριση του άλλου ώστε να μη δημιουργείται συχνοτικό κενό. Η συχνότητα που συναντώνται τα δύο φίλτρα ονομάζεται συχνότητα crossover και συμβολίζεται με f_c .



Σχήμα 5.4. Χρήση crossover για διαχωρισμό της συχνοτικής περιοχής του ηχητικού σήματος κατά την αναπαραγωγή του από το ηχείο.

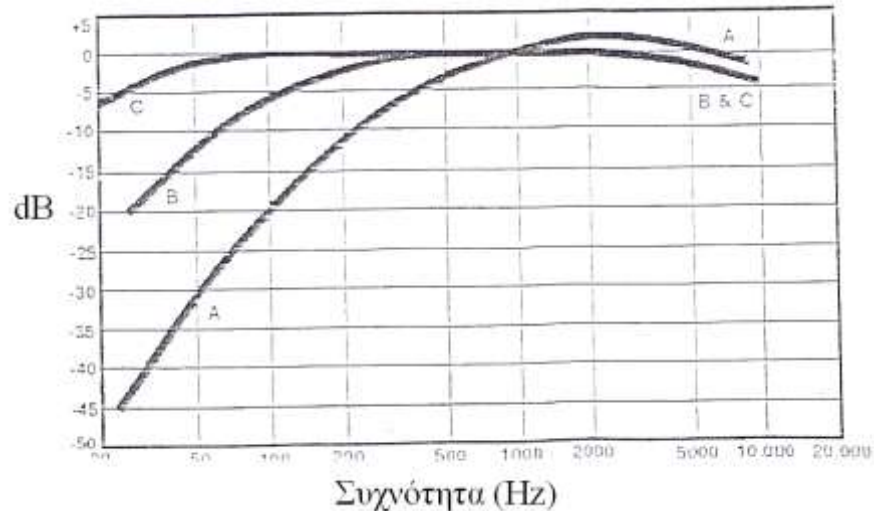
Ο συχνοτικός διαχωρισμός του ηχητικού σήματος είναι δυνατόν να γίνεται:

- Μετά τον ενισχυτή, οπότε τα φίλτρα που χρησιμοποιούνται είναι συνήθως παθητικά και κατάλληλα κατασκευασμένα ώστε να επεξεργαστούνε σήμα ισχύος, γεγονός που τα διαφοροποιεί από τα άλλα φίλτρα στα οποία δεν υπάρχει η απαίτηση να δουλεύουν σε μεγάλα ρεύματα.

- Πριν τον ενισχυτή, οπότε θα πρέπει να χρησιμοποιηθεί διαφορετικός ενισχυτής για κάθε μονάδα συχνοτήτων. Σε μια τέτοια περίπτωση, η συχνότητα crossover είναι δυνατόν να ρυθμίζεται από τον ηχολήπτη ανάλογα με το είδος της μουσικής και το είδος του εξοπλισμού που χρησιμοποιείται.

5.4 Φίλτρα A, B, C

Λόγω της μεταβλητής ευαισθησίας της ανθρώπινης ακοής με τη συχνότητα, απαιτείται συχνά η χρήση τυποποιημένων φίλτρων τα οποία εξασθενούν ή ενισχύουν κάποιες περιοχές συχνοτήτων, προσομοιώνοντας έτσι τη χαρακτηριστική λειτουργία της ακοής. Αυτά είναι τα λεγόμενα φίλτρα βάρους A, B και C των οποίων η απόκριση φαίνεται στο σχήμα 5.5. Τα φίλτρα αυτά είναι απαραίτητα όταν κάνουμε ηχητικές μετρήσεις που έχουν ως σκοπό να αποτιμήσουν το πόσο υποκειμενικά αισθητός είναι κάποιος θόρυβος σε μας. Μπορούμε να παρατηρήσουμε ότι οι καμπύλες απόκρισης αυτών των φίλτρων είναι αντίστροφες των καμπύλων ίσης ηχηρότητας που περιγράφονται στην παράγραφο 1.3(Εικόνα 10α).

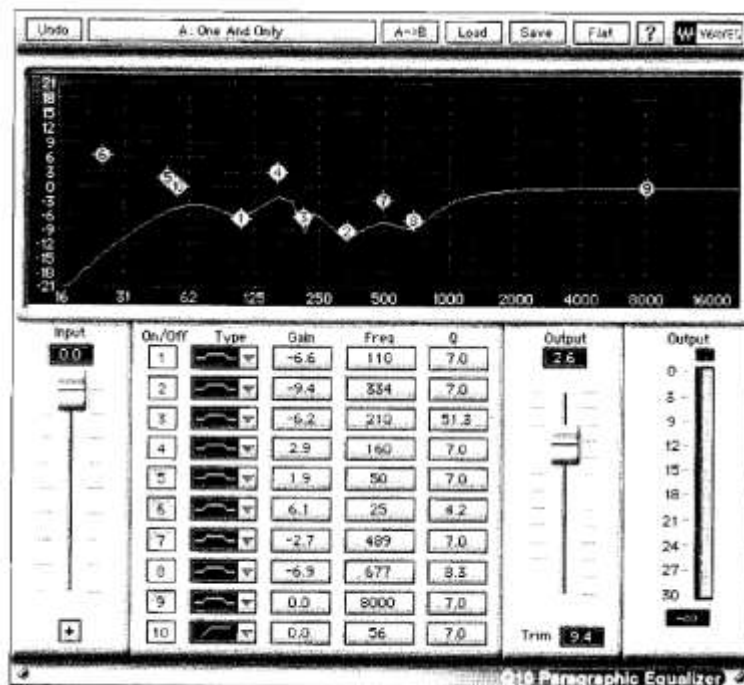


Σχήμα 5.5. Απόκριση συχνότητας των φίλτρων βάρους A, B, C..

Είναι προφανές ότι αφού δεν είμαστε ευαίσθητοι στις χαμηλές συχνότητες, ένας θόρυβος ο οποίος περιέχει μεγάλη ενέργεια στις χαμηλές, δεν θα μας είναι τόσο ενοχλητικός. Αντίθετα, ένας θόρυβος που περιέχει μεγάλη ενέργεια στις μεσαίες συχνότητες θα είναι πολύ περισσότερο ενοχλητικός. Ως εκ τούτου, χρησιμοποιώντας τις συναρτήσεις βάρους A, B και C, πετυχαίνουμε την κατάλληλη ζύγιση του προς μέτρηση θορύβου ώστε η στάθμη που θα υπολογιστεί να είναι αντιπροσωπευτική της μεταβλητής αυτής ευαισθησίας της ακοής. Ανάλογα με την καμπύλη βάρους που χρησιμοποιείται, το αποτέλεσμα της μέτρησης εκφράζεται σε μονάδες dB(A), dB(B) και dB(C). Η επιλογή του φίλτρου αν θα είναι A, B ή C έχει να κάνει με το είδος και τη στάθμη του προς μέτρηση θορύβου. Για παράδειγμα, στις υψηλές στάθμες θορύβου που δημιουργούνται από κινητήρες αεροσκαφών χρησιμοποιείται η καμπύλη C. Η καμπύλη αυτή χαρακτηρίζεται από πιο επίπεδη απόκριση συχνότητας σε σχέση με τις άλλες καμπύλες, γεγονός που συμφωνεί με το ότι στις υψηλές στάθμες η απόκριση του αυτιού τείνει να γίνει αρκετά πιο επίπεδη (flat). Σε αυτό το φαινόμενο βασίζεται η επιλεκτική δυνατότητα χρήσης, σε κάποιους ενισχυτές, ειδικού φίλτρου (loudness) που έχει ως σκοπό να ενισχύσει τις χαμηλές ή και τις υψηλές συχνότητες όταν η στάθμη ακρόασης μουσικού υλικού είναι χαμηλή.

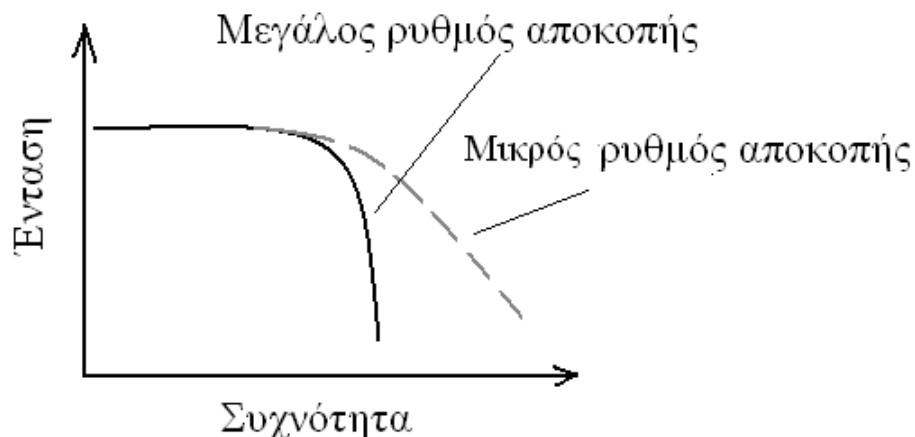
5.5 Ψηφιακά φίλτρα

Σχεδόν όλες οι παραπάνω διεργασίες που περιγράφηκαν υποθέτοντας τη χρήση αναλογικών equalizer μπορούν σήμερα να υλοποιηθούν με τη χρήση ψηφιακής τεχνολογίας. Το ψηφιακό equalizer βασίζεται στη χρήση τεχνικών ψηφιακής επεξεργασίας σήματος (DSP). Αν και η σκοπιμότητα των ψηφιακών equalizer δεν αλλάζει σε σχέση με τα αναλογικά, πρέπει να σημειωθεί η σημαντικά αυξημένη ευελιξία που παρέχεται από την ψηφιακή αναπαράσταση του σήματος. Για παράδειγμα, το εικονιζόμενο equalizer επιτρέπει την ανεξάρτητη ρύθμιση 10 διαφορετικών συχνοτικών περιοχών όπου μπορεί να μεταβάλλεται τόσο η κεντρική συχνότητα λειτουργίας, όσο και ο βαθμός ενίσχυσης-εξασθένησης αλλά και το εύρος ζώνης συχνοτήτων (Q) κάθε περιοχής.



Σχήμα 5.6. Παραμετρικό ψηφιακό equalizer 10 περιοχών σε μορφή λογισμικού της σειράς Diamond plugins της εταιρείας Waves.

Με ένα ψηφιακό equalizer μπορούμε θεωρητικά να καθορίζουμε την απόκριση συχνότητας με οποιοδήποτε θεμιτό τρόπο, κάτι που είναι αδύνατον να επιτευχθεί με αναλογικά μέσα. Επιπλέον πλεονέκτημα είναι η δυνατότητα υλοποίησης φίλτρων χαμηλής διέλευσης με μεγάλο ρυθμό αποκοπής, που είναι απαραίτητο κατά την ψηφιοποίηση του σήματος για την αποφυγή των φαινομένων συχνοτικής αλλαγής (aliasing) και το οποίο επιτυγχάνεται με χρήση τεχνικών υπερδειγματοληψίας (κεφ. 2.6). Το μειονέκτημα της αναλογικής τεχνολογίας κατά την υλοποίηση τέτοιων φίλτρων είναι καταρχάς το αυξημένο κόστος κατασκευής τους και επιπλέον, η εισαγωγή παραμορφώσεων στο ηχητικό σήμα. Αντίθετα, το μόνο μειονέκτημα του ψηφιακού φίλτρου είναι η αυξημένη τάξη του φίλτρου, κάτι που όμως δεν αποτελεί πρόβλημα στους σύγχρονους επεξεργαστές.

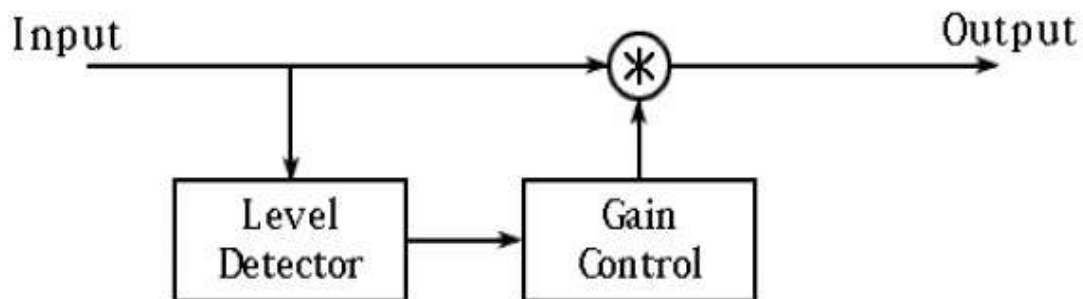


Σχήμα 5.7. Ορισμός του ρυθμού αποκοπής σε ένα φίλτρο χαμηλής διέλευσης συχνοτήτων.

Στις εφαρμογές αυτές υλοποιούνται σχέσεις εξισώσεων διαφοράς που περιγράφηκαν στην παράγραφο 4.1. Μπορούν λοιπόν να κατασκευαστούν «παραμετρικά» ή «γραφικά» φίλτρα / τράπεζες φίλτρων, τυπικά 1ης-4ης τάξης, με τοπολογίες IIR ή FIR που επιτρέπουν την επιλεκτική μεταβολή του φάσματος των επεξεργασμένων σημάτων. Σε ειδικές περιπτώσεις υλοποιούνται και φίλτρα υψηλής τάξης (π.χ. μέχρι και 3000 συντελεστών), σε τοπολογίες IIR ή FIR συχνά υπό μορφή φασματικού γινομένου (μέσω χρήσης του μετασχηματισμού Fourier).

6. Δυναμική Επεξεργασία Σήματος

Στις εφαρμογές αυτές υλοποιούνται μη – γραμμικές σχέσεις εισόδου / εξόδου (π.χ. λογαριθμικές εξισώσεις) που επιτρέπουν τον επιλεκτικό έλεγχο της δυναμικής περιοχής του σήματος εισόδου, με συμπίεση, περιορισμό (συστολή) ή επέκταση του πλάτους του επεξεργασμένου σήματος. Η χρήση τέτοιων συστημάτων είναι πολύτιμη σε διάφορες εφαρμογές όπως στη μουσική παραγωγή, σε εφαρμογές συμπίεσης ψηφιακού σήματος, στην αποθορυβοποίηση (de-noising), στην προστασία από την υπεροδήγηση κυκλωμάτων όπως ενισχυτών και ηχείων και στα ακουστικά βαρηκοΐας.



Σχήμα 6.1: Block διάγραμμα συστήματος μεταβολής και ελέγχου δυναμικής περιοχής

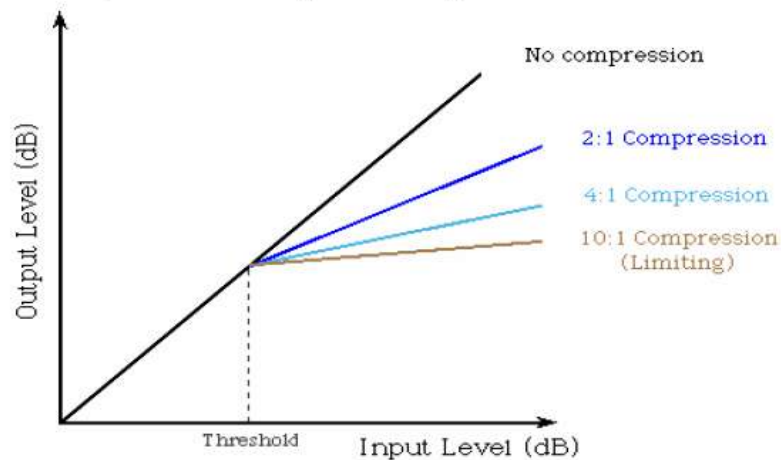
Μια απλοϊκή αναπαράσταση της λειτουργίας των δυναμικών επεξεργαστών φαίνεται στο Σχήμα 6.1, Γίνεται λοιπόν χρήση ενός πλευρικού βρόγχου ο οποίος υπολογίζει την στάθμη του ηχητικού σήματος εισόδου, την απεικονίζει σε μία άλλη τιμή (κέρδος) με βάση προκαθορισμένη και επιλεγμένη από τον χρήστη μη-γραμμική σχέση εισόδου-εξόδου και τελικά, με την χρήση αυτής της τιμής κέρδους $g(t)$, μεταβάλλει τοπικά το σήμα εισόδου $x(t)$, σύμφωνα με την σχέση

$$y(t) = g(t) \cdot x(t) \quad (6.1)$$

6.1 Συμπιεστής (Compressor)

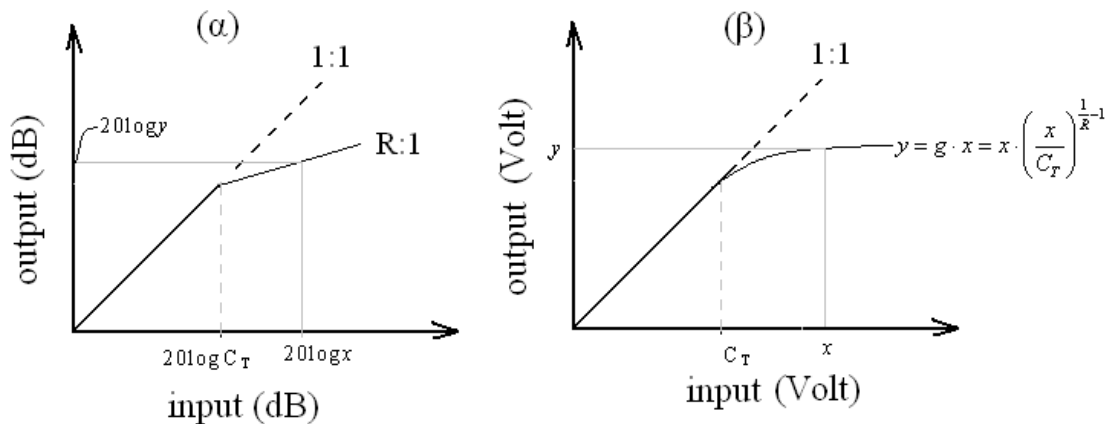
Ο συμπιεστής έχει ως σκοπό τον περιορισμό της δυναμικού εύρους του σήματος. Η στατική τοπική απεικόνιση του πλάτους εισόδου σε κέρδος γίνεται τυπικά μέσα από καμπύλες όπως φαίνονται στο Σχήμα 6.2. Οι καμπύλες αυτές εκφράζουν την τιμή του κέρδους $g(t)$ μέσω του λόγου συμπίεσης R και του κατωφλίου συμπίεσης L_T . Το κατώφλι συμπίεσης ορίζει την τιμή πλάτους του σήματος εισόδου πάνω από την οποία επενεργεί η μη-γραμμική μεταβολή κέρδους σύμφωνα με την σχέση (6.1).

Compressor Input/Output Characteristic



Σχήμα 6.2: Σχέσεις εισόδου-εξόδου για διάφορους λόγους συμπίεσης R .

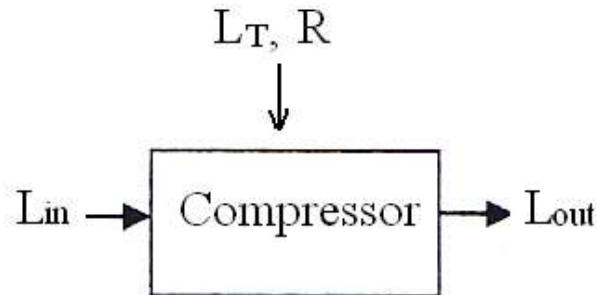
Οι τιμές R και L_T ορίζονται από τον χρήστη, ανάλογα με τα ηχητικά δεδομένα και την συγκεκριμένη εφαρμογή. Μια επεξήγηση για τη λειτουργία του συμπιεστή μπορεί να είναι η εξής: όταν το πλάτος του σήματος εισόδου (L_{in}) δεν ξεπερνάει το κατώφλι συμπίεσης L_T τότε ο συμπιεστής πρακτικά δε λειτουργεί και το σήμα διέρχεται από το συμπιεστή ανεπηρέαστο. Στην αντίθετη περίπτωση, αν το σήμα εισόδου είναι M dB πάνω από την τιμή L_T , τότε το πλάτος στην έξοδο (L_{out}) θα είναι μόνο $\frac{M}{R}$ dB πάνω από το $20\log C_T$, όπου R είναι ο λόγος συμπίεσης. Παρατηρούμε από το Σχήμα 6.3(α) ότι η συμπίεση του σήματος έχει γραμμική σχέση στη λογαριθμική κλίμακα εισόδου – εξόδου (όταν δηλαδή στους άξονες έχουμε dB). Θα πρέπει να έχουμε όμως κατά νου ότι η σχέση αυτή είναι εκθετική στη γραμμική κλίμακα, όταν δηλαδή στους άξονες έχουμε γραμμικές μονάδες όπως τάση σε Volt (Σχήμα 6.3(β)).



Σχήμα 6.3: Σχέση εισόδου-εξόδου (α) σε λογαριθμική κλίμακα και (β) σε γραμμική κλίμακα για ένα συμπιεστή με κατώφλι συμπίεσης $L_T=20\log C_T$ και λόγο συμπίεσης $R:1$ Με x και y συμβολίζεται αντίστοιχα η είσοδος και η έξοδος σε γραμμικές μονάδες (πχ. Volt).

Λαμβάνοντας υπόψη όλα τα παραπάνω είναι εύκολο τώρα να κατασκευάσουμε τη μαθηματική σχέση η οποία θα δίνει τη στάθμη εξόδου L_{out} του ηχητικού σήματος συναρτήσει της στάθμης εισόδου L_{in} , του κατώφλι συμπίεσης L_T και του λόγου συμπίεσης R . Συγκεκριμένα, εφόσον η στάθμη εισόδου ξεπερνάει το κατώφλι συμπίεσης, η στάθμη εξόδου L_{out} μπορεί να υπολογιστεί μέσω του τύπου

$$L_{out} = L_T + \frac{L_{in} - L_T}{R} \quad (6.2)$$



Σχήμα 6.4: Η στάθμη εξόδου L_{out} του σήματος μπορεί εύκολα να υπολογιστεί εάν ξέρουμε τη στάθμη εισόδου L_{in} , το κατώφλι συμπίεσης L_T και το λόγο συμπίεσης R .

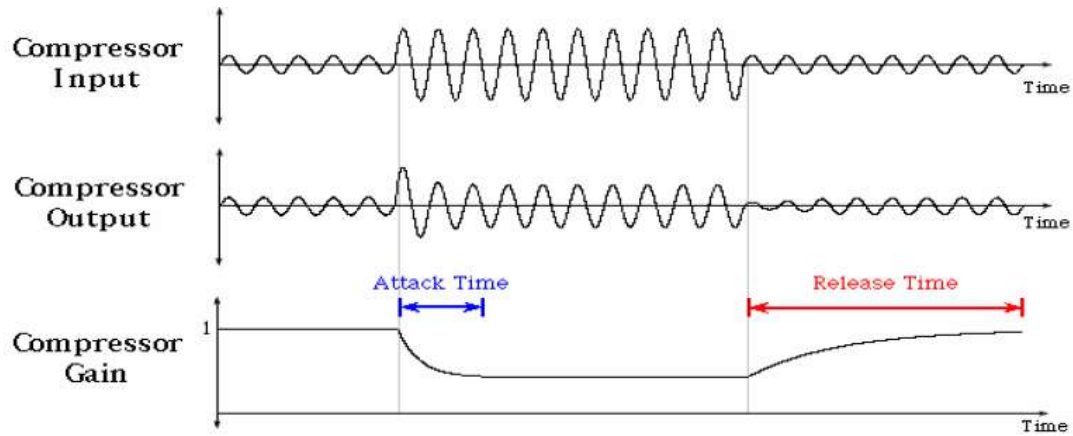
Από την τελευταία εξίσωση βλέπουμε ότι για μεγάλους λόγους συμπίεσης R ο όρος $\frac{L_{in} - L_T}{R}$ γίνεται πολύ μικρός και επομένως η στάθμη εξόδου L_{out} πρακτικά συμπίπτει με το κατώφλι συμπίεσης. Ένα σύστημα που λειτουργεί με πολύ μεγάλο λόγο συμπίεσης (τυπικά 10:1 και άνω) είναι το *limiter*. Το *limiter* παρεμβάλλεται συνήθως μεταξύ της εξόδου της κονσόλας και του τελικού ενισχυτή, έχοντας ως σκοπό να προστατεύει τον ενισχυτή και τα ηχεία από απότομες μη φυσιολογικές εξάρσεις του σήματος, όπως πχ. από πτώση μικροφώνου, οι οποίες θα μπορούσαν να οδηγήσουν έως και σε καταστροφή των μονάδων χαμηλών συχνοτήτων των ηχείων.

Παράδειγμα: Να βρεθεί η στάθμη εξόδου L_{out} ενός ηχητικού σήματος το οποίο εισέρχεται σε ένα συμπιεστή στα α) 0 dB και β) -16 dB. Για το συμπιεστή θεωρήστε κατώφλι συμπίεσης $L_T = -10$ dB και λόγο συμπίεσης $R = 5:1$.

Απάντηση: α) Παρατηρούμε ότι η στάθμη του σήματος είναι 10 dB πάνω από το κατώφλι συμπίεσης. Ο λόγος συμπίεσης $R = 5:1$ μας λέει ότι αυτά τα 10 dB θα απεικονίζονται σε μια δυναμική περιοχή μόνο $10:5 = 2$ dB. Με άλλα λόγια, η στάθμη του σήματος θα αυξηθεί μόνο 2 dB πάνω από το κατώφλι L_T . Επομένως, η στάθμη του σήματος εξόδου θα είναι $-10 + 2 = -8$ dBu. Αυτή η τιμή μπορεί επίσης να προκύψει απευθείας με απλή αντικατάσταση στη σχέση 5.1.

β) Τα -16 dB είναι κάτω από το κατώφλι συμπίεσης επομένως ο συμπιεστής δε θα επιδράσει στο σήμα. Η στάθμη εξόδου θα παραμείνει -16 dB.

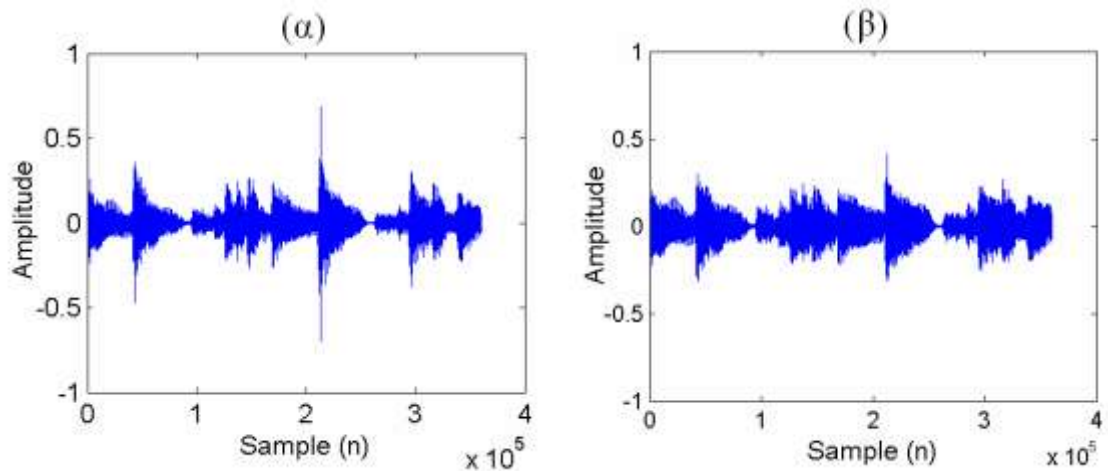
Για την προοδευτική μετάβαση του κέρδους από τις καταστάσεις γραμμικής/μη-γραμμικής επεξεργασίας, τα παραπάνω στατικά χαρακτηριστικά συμπληρώνονται και με «χρονικά χαρακτηριστικά», όπου η αρχή (*attack*) και το τέλος (*release*) της μη-γραμμικής επεξεργασίας, εξομαλύνονται με τοπική εκθετική καμπύλη μετάβασης, όπως φαίνεται και στο Σχήμα 6.5.



Σχήμα 6.5: Ορισμός του attack και του release σε υλοποίηση συμπίεσης.

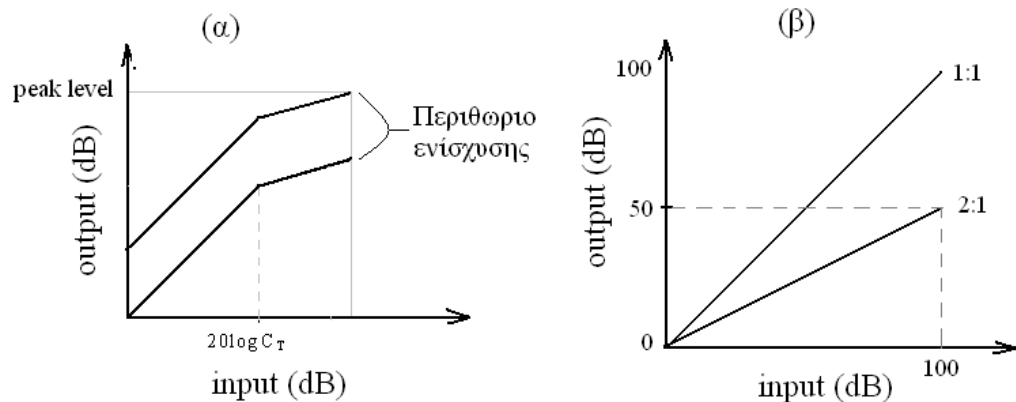
Βλέπουμε ότι η παράμετρος του attack καθορίζει το χρονικό διάστημα που μεσολαβεί, από τη στιγμή που η στάθμη του σήματος περάσει το κατώφλι συμπίεσης, μέχρι να ξεκινήσει η συμπίεση του σήματος. Από την άλλη, η παράμετρος release καθορίζει το χρονικό διάστημα που θα περάσει μέχρι ο κομπρέσορας να επιστρέψει στη γραμμική περιοχή λειτουργίας. Η ρύθμιση των παραμέτρων αυτών καθορίζεται με βάση τα δυναμικά χαρακτηριστικά του σήματος. Πολύ μικρές τιμές του attack και του release time θα είναι ενδεχομένως επιθυμητές σε σήματα με απότομες εξάρσεις οι οποίες δε διαρκούν μεγάλο χρονικό διάστημα. Ωστόσο η απότομη μεταβολή του κέρδους μπορεί να κάνει τη μετάβαση από ένα χαμηλό σε ένα υψηλό πέρασμα να μην ακούγεται φυσική. Από την άλλη, αύξηση του attack time, επιτρέπει μια πιο ομαλή ελάττωση της στάθμης του σήματος και κάνει το πέρασμα να ακούγεται πιο φυσικό. Ωστόσο, μεγάλη τιμή του release time ενδεχομένως να παρατείνει ανεπιθύμητα τη δράση του συμπιεστή σε χρονικά τμήματα όπου η στάθμη του σήματος έχει ήδη επανέλθει σε χαμηλές στάθμες, όπως φαίνεται και στο Σχήμα 6.5.

Κάποιες μονάδες συμπίεσης, όπως για παράδειγμα limiters που χρησιμοποιούνται για ραδιοφωνική μετάδοση, λειτουργούν λαμβάνοντας υπόψη το στιγμιαίο πλάτος του σήματος (peak level detection). Άλλες μονάδες λαμβάνουν υπόψη τη μέση τετραγωνική τιμή του σήματος (rms detection). Κατά τη λειτουργία σε peak level detection, ιδιαίτερα σε συνδυασμό με ένα υψηλό λόγο συμπίεσης, απαγορεύουμε εντελώς στο σήμα εξόδου υπερβεί προκαθορισμένες στάθμες, ούτε καν για κλάσματα του δευτερολέπτου. Από την άλλη, κατά τη λειτουργία υπό rms detection, η στάθμη του σήματος υπολογίζεται για ένα μεγαλύτερο χρονικό παράθυρο, το οποίο έχει ως αποτέλεσμα να επιτρέπεται στο σήμα να υπερβεί την προκαθορισμένη στάθμη για κάποιο χρονικό διάστημα, πριν ο συμπιεστής ελαττώσει τη στάθμη του. Είναι προφανές ότι οι παράμετροι attack και release time έχουν νόημα περισσότερο για συμπιεστές που λειτουργούν υπό καθεστώς rms detection παρά για peak detection.



Σχήμα 6.6: Επίδραση της συμπίεσης σε ένα ηχητικό σήμα φωνής (α) πριν την επεξεργασία και (β) μετά την επεξεργασία, όπου εκτός από συμπίεση έχει εφαρμοστεί και μια ήπια ενίσχυση του σήματος.

Η χρήση συμπίεστη γίνεται τόσο για αισθητικούς όσο και για πρακτικούς λόγους. Στη ροκ μουσική για παράδειγμα ο συμπίεστης χρησιμοποιείται για να γίνει ο ήχος πιο συμπαγής. Το μπάσο είναι ένα δημοφιλές compressor όργανο. Για ηχητικό σήμα φωνής ένας compressor είναι σχεδόν πάντα απαραίτητος. Δεδομένου ότι μετά την συμπίεση το σήμα έχει περιορισμένη δυναμική περιοχή, είναι συνήθως χρήσιμη και εφικτή η σταθερή αύξηση του συνολικού κέρδους. Το πώς δημιουργείται αυτή η δυνατότητα περαιτέρω ενίσχυσης του σήματος φαίνεται στο Σχήμα 6.7(α)



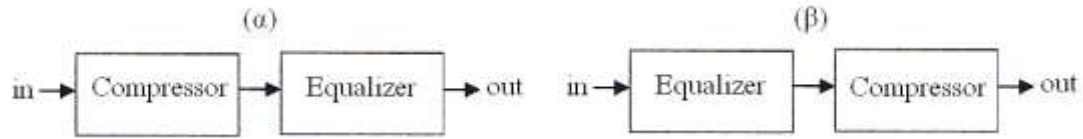
Σχήμα: 6.7 (α) Όσο πιο χαμηλό το κατώφλι συμπίεσης και όσο πιο υψηλός ο λόγος συμπίεσης τόσο μεγαλώνει το περιθώριο ενίσχυσης του σήματος. (β) χαμηλώνοντας εντελώς το κατώφλι συμπίεσης και με ένα ήπιο λόγο συμπίεσης 2:1 είναι δυνατόν να «χωρέσουμε» δυναμικό εύρος 100 dB σε κασετόφωνο δυναμικής περιοχής μόνο 50 dB.

Παράδειγμα: Έστω ότι ένα ηχητικό σήμα εισέρχεται σε ένα ηχητικό σύστημα στη μέγιστη επιτρεπόμενη νόμιμη στάθμη λειτουργίας των $L_0 = +4$ dB. Να βρεθεί το περιθώριο ενίσχυσης του σήματος σε dB που μπορεί να αποδώσει ένας κομπρέσορας που συνδέεται πριν το ηχητικό σύστημα με λόγο συμπίεσης $R = 4:1$ και κατώφλι συμπίεσης $L_T = -8$ dB.

Απάντηση: Είναι προφανές ότι το ηχητικό σήμα περνώντας μέσα από τον συμπίεστη εξέρχεται με μειωμένο δυναμικό εύρος και άρα η στάθμη εισόδου στο ηχητικό σύστημα θα είναι μικρότερη από πριν. Συγκεκριμένα, αφού $L_T = -8$ dB η στάθμη εξόδου από τον κομπρέσορα θα είναι $L_{out} = -8 + \frac{4 - (-8)}{4} = -8 + 3 = -5$ dB. Αφού η στάθμη του σήματος κατά την είσοδο στο ηχητικό σύστημα επιτρέπεται να είναι έως τα +4 dB, το περιθώριο

ενίσχυσης σε dB θα είναι ίσο με $\text{Gain}=+4-(-5)=9\text{dB}$.

Παράδειγμα: Θεωρείστε ένα equalizer ρυθμισμένο να προσδίδει αποκοπή 6 dB στα 800 Hz και έναν RMS συμπιεστή με παραμέτρους $L_T=0\text{dB}$ και $R=4:1$. Υπολογίστε την RMS στάθμη (σε dB) με την οποία θα εξέρχεται ένα ημιτονικό σήμα συχνότητας 800 Hz και RMS τιμής 8dB από την σε σειρά συνδεσμολογία α) compressor – equalizer και β) equalizer – compressor. Τι παρατηρείται;



Απάντηση: Στην περίπτωση (α), η στάθμη με την οποία εξέρχεται το σήμα από τον συμπιεστή είναι $L_{out} = 0 + \frac{(8-0)}{4} = 2\text{dB}$. Περνώντας από το equalizer η στάθμη μειώνεται

κατά 6 dB επομένως η στάθμη εξόδου στην πρώτη περίπτωση είναι $2-6=-4\text{dB}$. Στην περίπτωση (β) παρατηρούμε ότι το σήμα θα εξέρχεται στα $8-6=2\text{dB}$ από το equalizer. Η στάθμη του σήματος από την έξοδο του συμπιεστή θα είναι εν τέλει $L_{out} = 0 + \frac{(2-0)}{4} = 0.5\text{dB}$.

Παρατηρούμε ότι αλλάζοντας τη σειρά των μονάδων equalizer και compressor αλλάζει και η στάθμη εξόδου του σήματος από την αλυσίδα. Αυτό το φαινόμενο οφείλεται στην μη γραμμική επεξεργασία που υπόκειται το σήμα από το δυναμικό επεξεργαστή και μας υπαγορεύει ότι θα πρέπει να περιμένουμε διαφορετικά αποτελέσματα ανάλογα με το σε ποια θέση βάζουμε ένα δυναμικό επεξεργαστή σε μια αλυσίδα από ηχητικά εφέ.

Μια πολύ βασική χρησιμότητα του συμπιεστή εμφανίζεται όταν ενώ έχουμε ένα ηχητικό σήμα μεγάλου δυναμικού εύρους, ο εξοπλισμός που διατίθεται για την μετάδοση ή ηχογράφηση του σήματος έχει περιορισμένο δυναμικό εύρος. Για παράδειγμα, αν υποθέσουμε ότι σε μια συναυλία το ηχητικό γεγονός καταλαμβάνει ένα δυναμικό εύρος 100 dB και εμείς θέλουμε να το ηχογραφήσουμε σε ένα κασετόφωνο όπου η διαθέσιμη δυναμική περιοχή είναι μόνο 50 dB. Για να χωρέσουμε λοιπόν τα 100 dB στα 50 dB διαθέσιμου δυναμικού εύρους μπορούμε να χρησιμοποιήσουμε ένα μικρό λόγο συμπίεσης $R=2:1$ και να χαμηλώσουμε εντελώς το κατώφλι συμπίεσης όπως φαίνεται και στο σχήμα 6.7(β) Επειδή όμως κατά την μετάδοση - ηχογράφηση έχουμε περιορισμό του δυναμικού εύρους του σήματος, πρέπει να γίνεται κάτι κατά την αναπαραγωγή του συγκεκριμένου αρχείου που να επιτρέπει αποκατάσταση του αρχικού δυναμικού εύρους του σήματος. Αυτή επέκταση του δυναμικού εύρους γίνεται εφικτή με τη χρήση expander.

Παράδειγμα: Τι συντελεστή συμπίεσης όγκου δεδομένων μπορεί να παρέχει ένας συμπιεστής ψηφιακού σήματος με παραμέτρους $L_T=-24\text{dBFS}$ και $R=4:1$; Θεωρείστε τάξη κβαντισμού 16 bit για το ψηφιακό σήμα και ότι το ψηφιακό σήμα πριν την είσοδο στον συμπιεστή καταλαμβάνει όλη τη διαθέσιμη δυναμική περιοχή. Επίσης, ο συμπιεστής λαμβάνει υπόψη peak στάθμες.

Απάντηση: Ένα ψηφιακό σήμα το οποίο διέρχεται μέσα από ένα τέτοιο συμπιεστή θα υποστεί περιορισμό του δυναμικού του εύρους. Έχουμε δει ότι η δυναμική περιοχή σχετίζεται άμεσα με την τάξη κβαντισμού, και επομένως καταλαβαίνουμε ότι αφού θα ελαττωθεί το δυναμικό εύρος του σήματος, θα χρειάζονται λιγότερα bits για τη μετάδοση ή αποθήκευσή του. Αφού το ηχητικό σήμα καταλαμβάνει όλη τη δυναμική περιοχή σημαίνει

ότι κάποιες χρονικές στιγμές τα πλάτη του αγγίζουν τα 0 dBFS. Κατά την έξοδο λοιπόν από το συμπίεστη, τα μέγιστα πλάτη του σήματος δε θα υπερβαίνουν τα

$$L_{out} = -24 + \frac{0 - (-24)}{4} = -24 + 6 = -18\text{dBFS}.$$

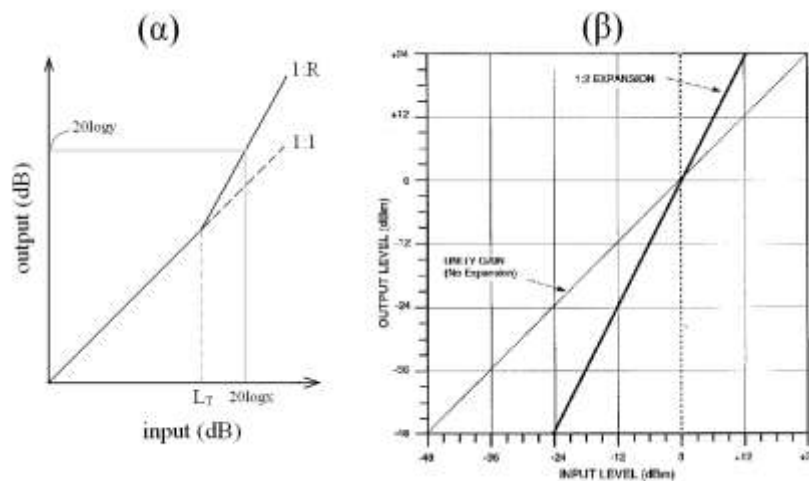
Γνωρίζουμε όμως από προηγούμενο κεφάλαιο ότι κάθε επιπλέον bit αντιστοιχεί σε 6 dB δυναμικής περιοχής. Αυτό σημαίνει ότι χρειαζόμαστε $18:6=3$ bit λιγότερα για την μετάδοση ή αποθήκευση του σήματος, επομένως, $16-3=13$ bit επαρκούν. Ο λόγος συμπίεσης όγκου δεδομένων θα είναι επομένως $\frac{N}{N'} = \frac{16}{13} = 1,23$ (ή 1,23:1).

5.2 Expander

Μία μονάδα expander κάνει θεωρητικά το αντίθετο από ότι ένας compressor: οδηγεί σε επέκταση του δυναμικού εύρους του σήματος λαμβάνοντας υπόψη παρόμοιες παραμέτρους όπως κάποιο κατώφλι ενεργοποίησης (ή κατώφλι επέκτασης) L_T και κάποιο λόγο επέκτασης 1:R. Η στάθμη εξόδου L_{out} ενός ηχητικού σήματος δεδομένης της στάθμης εισόδου μπορεί να αναπαρασταθεί μαθηματικά σύμφωνα με τον τύπο

$$L_{out} = L_T + (L_{in} - L_T) \cdot R \quad (6.3)$$

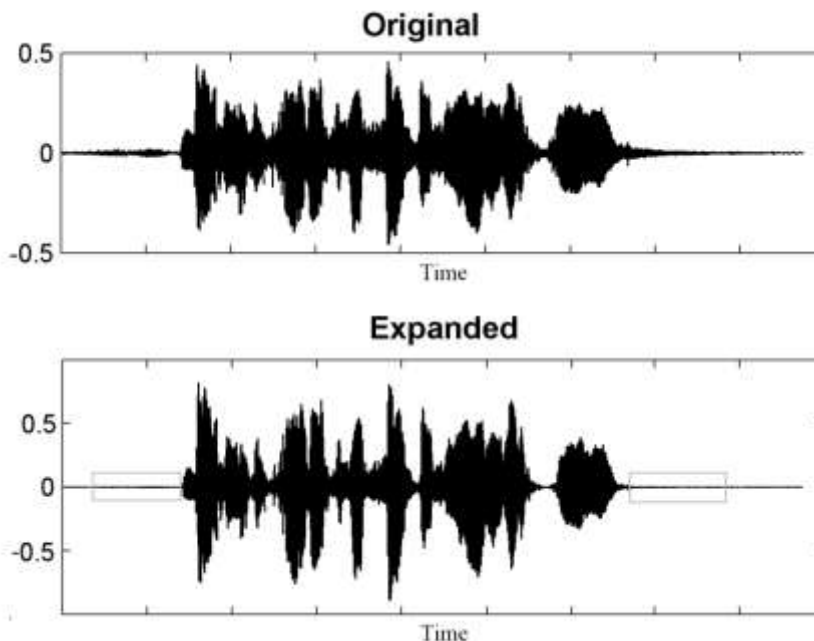
Στο σχήμα 6.8 φαίνονται δύο χαρακτηριστικές περιπτώσεις λειτουργίας expander. Στο σχήμα (α) ο expander επιδρά από μια στάθμη του σήματος (L_T) και πάνω ενισχύοντας περεταίρω τις υψηλές στάθμες και αφήνοντας ανεπηρέαστες τις χαμηλές. Στο σχήμα (β) ο expander επιδρά σε όλο το δυναμικό εύρος του σήματος εισόδου. Παρατηρούμε ότι τα 36 dB δυναμικού εύρους του σήματος εισόδου (από -24 έως 12) αντικατροπίζονται σε 72 dB δυναμικού εύρους στην έξοδο (από -48 έως 24). Ο expander σε μια τέτοια περίπτωση επιδρά ενισχύοντας τις υψηλές στάθμες και εξασθενώντας τις χαμηλές στάθμες. Για παράδειγμα, ένα σήμα που εισέρχεται στα -12 dB εξέρχεται στα -24 dB ενώ ένα σήμα που εισέρχεται στα 12 dB εξέρχεται στα 24 dB. Παρατηρούμε λοιπόν ότι η δουλειά που θα έκανε ένα τέτοιο expander είναι ακριβώς η αντίθετη από τη δουλειά που κάνει ο compressor του σχήματος 6.7(β).



Σχήμα 6.8: Χαρακτηριστικές καμπύλες εισόδου – εξόδου expander

Ένα παράδειγμα από τη χρήση ενός τέτοιου expander σε ένα σήμα φωνής φαίνεται στο σχήμα 6.9. Στο πάνω διάγραμμα φαίνεται η αρχική κυματομορφή του σήματος και στο κάτω

διάγραμμα φαίνεται η κυματομορφή του σήματος κατά την έξοδό του από το expander. Παρατηρούμε ότι οι κορυφές (εξάρσεις) του σήματος έχουν γίνει ακόμα πιο απότομες και πιο ευδιάκριτες από πριν. Από την άλλη, χρονικές περιοχές με μικρή στάθμη σήματος έχουν καταπιεστεί ακόμα πιο πολύ, με αποτέλεσμα σχεδόν να εξαφανίζεται το σήμα (οι περιοχές αυτές περικλείονται από τα γκρι ορθογώνια πλαίσια στο διάγραμμα). Υπό ένα τέτοιο καθεστώς λειτουργίας ένα expander μπορεί να επιτελέσει λειτουργία αποθορυβοποίησης (de-noising), αν και η κατασκευή του δεν είναι εξαρχής ορισμένη για αυτό το σκοπό.

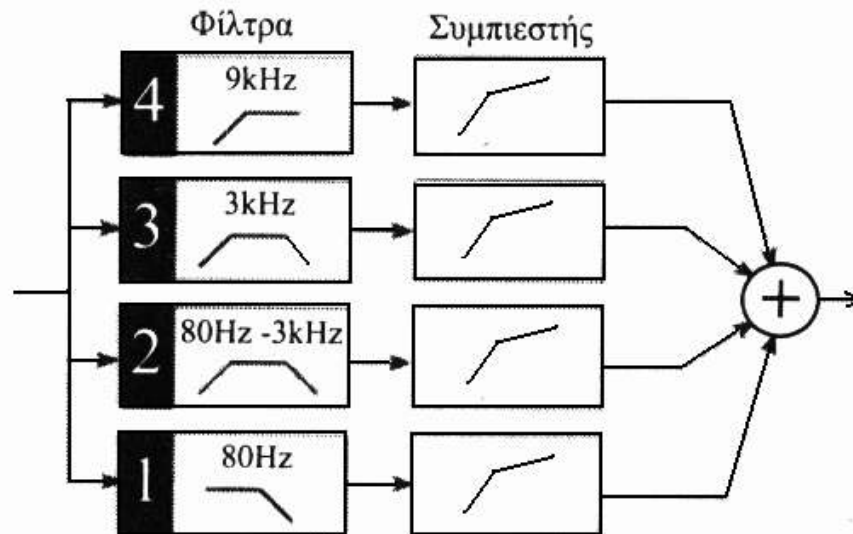


Σχήμα 6.9: Κυματομορφή σήματος πριν και μετά την έξοδο από μονάδα expander

6.3 Multi-band compressor

Κατά την δυναμική επεξεργασία ενός ηχητικού σήματος θα θέλαμε πιθανόν να επέμβουμε σε κάποιες συγκεκριμένες συχνοτικές περιοχές, αφήνοντας ανεπηρέαστες κάποιες άλλες. Ας υποθέσουμε ότι έχουμε ένα μονοφωνικό αρχείο το οποίο προήλθε από την ηχογράφηση μπάσου και βιολιού ταυτόχρονα. Δυστυχώς δεν έχουμε σε ξεχωριστά κανάλια το μπάσο και το βιολί, αλλά είναι και τα δύο μέσα σε ένα κανάλι. Κατά την ακρόαση του αρχείου δημιουργείται η επιθυμία να κάνουμε τον ήχο του μπάσου πιο συμπαγή, αλλά δε θα θέλαμε με αυτήν την πράξη να επηρεάσουμε και το βιολί. Υπάρχει άραγε κάποιος τρόπος να κάνουμε συμπίεση στο δυναμικό εύρος του μπάσου χωρίς να επηρεαστεί ο ήχος του βιολιού; Ευτυχώς, τα δύο αυτά μουσικά όργανα έχουν αρκετά διαφορετικό φάσμα. Το μπάσο περιέχει κυρίως χαμηλές συχνότητες ενώ το βιολί κυρίως μεσαίες και χαμηλές. Θα μπορούσαμε λοιπόν να επιτύχουμε το ζητούμενο ως εξής: Αν περάσουμε το ηχητικό σήμα από ένα φίλτρο χαμηλής διέλευσης συχνοτήτων, θα πάρουμε ένα σήμα το οποίο θα περιέχει κυρίως τις χαμηλές συχνότητες του μπάσου και ελάχιστο βιολί. Από την άλλη, αν περάσουμε το σήμα από ένα φίλτρο υψηλής διέλευσης το οποίο αφήνει ανεπηρέαστες τις μεσαίες και χαμηλές συχνότητες και κόβει τις χαμηλές, θα πάρουμε στην έξοδό ένα σήμα που περιέχει κυρίως το βιολί και ελάχιστο μπάσο. Είναι προφανές ότι τα δύο αυτά φίλτρα θα πρέπει να επενεργούν σε παράλληλη διάταξη. Στην

έξοδο του φίλτρου χαμηλής διέλευσης μπορούμε τώρα να συνδέσουμε ένα συμπιεστή και να ορίσουμε τις επιθυμητές παραμέτρους για την συμπίεση του μπάσου. Αν φανταστούμε περισσότερα από δύο τέτοια φίλτρα τα οποία επενεργούνε παράλληλα, σε σειρά το καθένα με κάποιο συμπιεστή, παίρνουμε το σχήμα του multi-band compressor που φαίνεται στο σχήμα 6.10.



Σχήμα 6.10 Τυπική μορφή ενός multi-band compressor τεσσάρων συχνοτικών περιοχών.

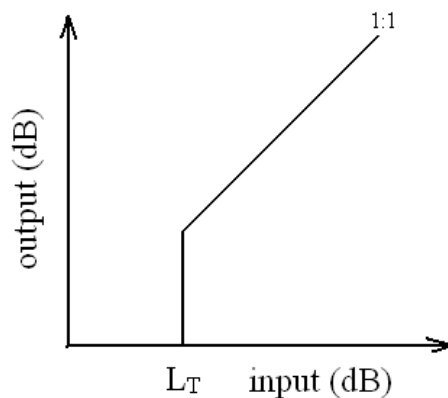
Στο σχήμα 6.10 φαίνεται μια συσκευή η οποία διαχωρίζει το σήμα σε 4 διαφορετικές συχνοτικές περιοχές ή μπάντες (εξού και ο όρος multi-band). Κάθε μπάντα έχει το δικό της κομπρέσορα. Σε αρκετές τέτοιες εφαρμογές, ο χρήστης μπορεί να ορίζει μόνοι του τη θέση κάθε μπάντας στο συχνοτικό φάσμα, δηλαδή σε ποια συχνότητα θα ξεκινάει και θα τελειώνει κάθε μπάντα. Παρατηρείστε ότι η πρώτη μπάντα ορίζεται από ένα φίλτρο χαμηλής διέλευσης, η τελευταία από ένα φίλτρο υψηλής διέλευσης, ενώ για τις ενδιάμεσες μπάντες χρειάζονται φίλτρα διέλευσης ζώνης συχνοτήτων (ζωνοπερατά φίλτρα). Προφανώς, η αρχή και το τέλος κάθε μπάντας θα πρέπει να είναι τέτοιο ώστε να καλύπτεται όλο το ακουστικό φάσμα. Μετά από τη δυναμική επεξεργασία, τα σήματα από κάθε μπάντα προστίθενται μεταξύ τους και έτσι ανακατασκευάζεται ένα σήμα πλήρους συχνοτικού φάσματος.

6.4 Noise Gates

Ένας από τους πιο απλούς αλλά και πιο χρήσιμους δυναμικούς επεξεργαστές είναι τα λεγόμενα noise gates. Αυτές οι διατάξεις υπολογίζουνε τη στάθμη του σήματος και επιτρέπουνε στο σήμα να διέλθει μόνο εφόσον η στάθμη του ξεπερνάει κάποιο καθορισμένο κατώφλι L_T . Σε περίπτωση που η στάθμη του σήματος είναι μικρότερη από το κατώφλι L_T , το noise gate “κλείνει”, επιβάλλοντας απόλυτη σιωπή. Το σκεπτικό για την κατασκευή μιας τέτοιας συσκευής προέρχεται από τη επιθυμία μείωσης της στάθμης θορύβου σε μικροφωνικές διατάξεις και γενικότερα, για την αποφυγή ανεπιθύμητης διαρροής σήματος σε πολυφωνικές διατάξεις ηχογράφησης.

Το noise gate για παράδειγμα είναι χρήσιμο κατά την ηχογράφιση ενός σετ από drums. Για να ηχογραφηθεί ένα drumset θα πρέπει να τοποθετήσουμε ένα ξεχωριστό μικρόφωνο σε κάθε στοιχείο της drums, με σκοπό από τη μία να καλύψουμε όλες τις ακουστικές πηγές του σετ, αλλά και να μπορούμε να επέμβουμε ξεχωριστά σε κάθε τύμπανο ή σε κάθε πιατίι κατά το στάδιο της ηχητικής επεξεργασίας. Ας υποθέσουμε λοιπόν ότι έχουμε τοποθετήσει ένα μικρόφωνο στη μπότα (kick drum) και ένα μικρόφωνο στο ταμπούρο (snare drum) τα οποία είναι δύο από τα πιο βασικά στοιχεία του ρυθμού στη ροκ μουσική. Όταν ο drummer χτυπάει το ταμπούρο, τότε αναπόφευκτα ένα μέρος του σήματος διαρρέει στο μικρόφωνο της μπότας. Αντίστοιχα, όταν ο drummer χτυπάει τη μπότα, ένα μέρος του σήματος διαρρέει στο μικρόφωνο του ταμπούρου. Αν λοιπόν θέλαμε να βάλουμε ένα equalizer στο σήμα που προέρχεται από την μπότα με σκοπό να τονίσουμε κάποιες συχνότητες της μπότας, τότε αναπόφευκτα θα τονίζαμε τις ίδιες συχνότητες και στο σήμα του ταμπούρου που έχει διαρρεύσει στο κανάλι αυτό. Προφανώς, κάτι αντίστοιχο θα συνέβαινε αν θέλαμε να επέμβουμε στον ήχο του ταμπούρου. Δημιουργείται λοιπόν η ανάγκη για απομόνωση των ακουστικών πηγών, κάτι που δεν είναι εφικτό λόγω του φαινομένου της διαρροής που μόλις περιγράψαμε. Εδώ λοιπόν έρχεται το noise gate να βοηθήσει την κατάσταση προσφέροντας ένα είδος τεχνητής απομόνωσης στις ακουστικές πηγές. Συνδέοντας ένα noise gate στην έξοδο του μικροφώνου της μπότας και ρυθμίζοντας κατάλληλα το κατώφλι L_T , μπορούμε να επιβάλλουμε στο gate να “ανοίγει” μόνο όταν ο drummer βαράει τη μπότα και να μένει κλειστό κατά την υπόλοιπη διάρκεια. Αυτό βασίζεται στο ότι το ηλεκτρικό σήμα που αντιστοιχεί στη μπότα θα είναι ισχυρότερο από το σήμα που έχει διαρρεύσει από το ταμπούρο ή από τα άλλα στοιχεία της drums. Αντίστοιχα, το μικρόφωνο του ταμπούρου θα πρέπει να “ανοίγει” μόνο όταν διέρχεται ένα ισχυρό σήμα το οποίο προέρχεται από το χτύπημα του ταμπούρου. Έτσι θα μπορούσαμε λοιπόν να απομονώσουμε κάθε στοιχείο της drums, επιβάλλοντας ρυθμίσεις οι οποίες δε θα επηρεάζουμε ηχητικές συνιστώσες διαφορετικές από αυτές για τις οποίες διατίθεται το κάθε μικρόφωνο.

Η χαρακτηριστική εισόδου-εξόδου μιας μονάδας noise gate φαίνεται στο σχήμα 6.11. Εκτός από το κατώφλι L_T , ο χρήστης μπορεί επίσης να ρυθμίζει χρονικά χαρακτηριστικά όπως το attack και το release, τα οποία επιδρούν με τρόπο παρόμοιο που ισχύει για τις μονάδες compressor που περιγράψαμε προηγουμένως.



Σχήμα 6.11: Τυπικό διάγραμμα εισόδου – εξόδου μιας μονάδας noise gate.