

Επεξεργασία Ήχου Φωνής
6^η Διάλεξη

ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΗΧΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Ψηφιακή Επεξεργασία Σήματος σε ηχητικές εφαρμογές

- Στο κεφάλαιο αυτό υπενθυμίζεται η βασική θεωρία της Ψηφιακής Επεξεργασίας Σήματος και αναλύονται κάποιες βασικές κατηγορίες ψηφιακών εφέ.
- Η μονάδα επεξεργασία του ηχητικού σήματος μπορεί να αποδοθεί με την έννοια ενός συστήματος, όπου η είσοδος $x(n)$ είναι το πριν την επεξεργασία ηχητικό σήμα (*dry signal*) και η έξοδος $y(n)$ την τελική μορφή του επεξεργασμένου ηχητικού σήματος (*wet signal*).



FIR και IIR Φίλτρα-Συστήματα

- **FIR**

- Στις περιπτώσεις αυτές έχουμε Γραμμικά Χρονικά Αναλλοίωτα Συστήματα και Πεπερασμένη Κρουστική Απόκριση (Finite Impulse Response) το οποίο αναπαριστάται μέσω της εξίσωσης διαφορών

$$y(n) = b_0x(n) + b_1x(n-1) + \dots + b_qx(n-q) \quad \text{ή} \quad y(n) = \sum_{k=0}^q b_kx(n-k)$$

q ακέραιος η τάξη του φίλτρου, b_i οι συντελεστές του φίλτρου που καθορίζουν την δράση του φίλτρου στο σήμα.

- **IIR (Infinite Impulse Response)**

- Στην περίπτωση που η απόκριση δεν είναι πεπερασμένη, οι εξισώσεις διαφορών γίνονται

$$a_0y(n) + a_1y(n-1) + \dots + a_py(n-p) = b_0x(n) + b_1x(n-1) + \dots + b_qx(n-q) \quad \text{ή} \quad \sum_{k=0}^p a_ky(n-k) = \sum_{k=0}^q b_kx(n-k)$$

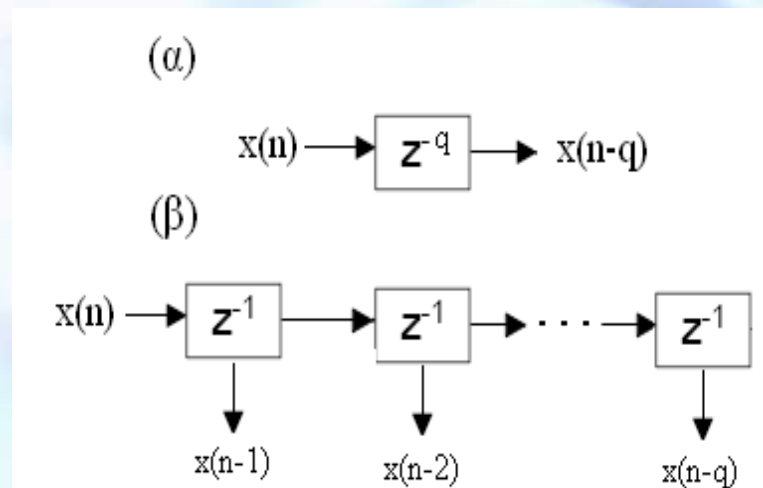
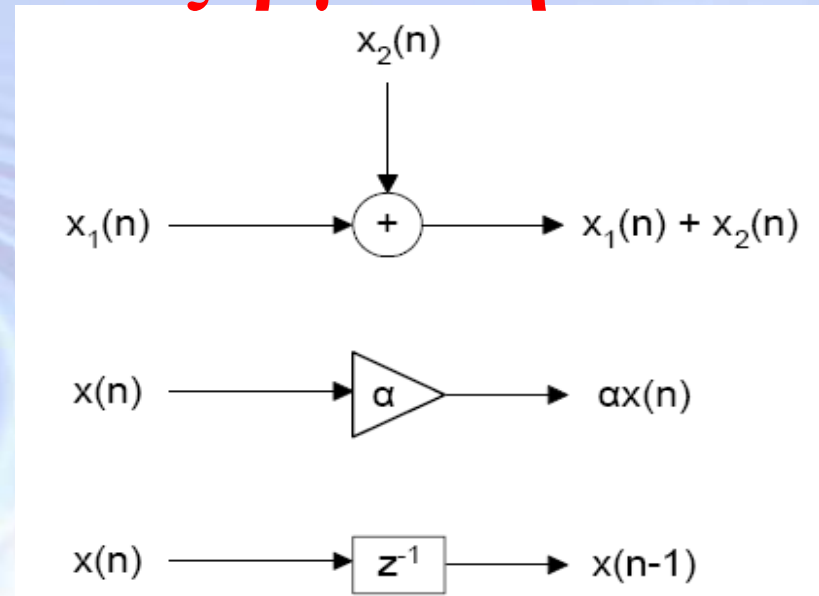
- Δηλαδή ή έξοδος n χρειάζεται ένα αριθμό προηγούμενων εξόδων. Στην περίπτωση αυτή θέλουμε τις αρχικές συνθήκες δηλαδή για $n=0$ π.χ για μηδενικές αρχικές συνθήκες ισχύει

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

Δομικά στοιχεία του επεξεργαστή

- Εξίσωση διαφορών:
 - Άθροιση
 - Πολλαπλασιασμός
 - Καθυστέρηση

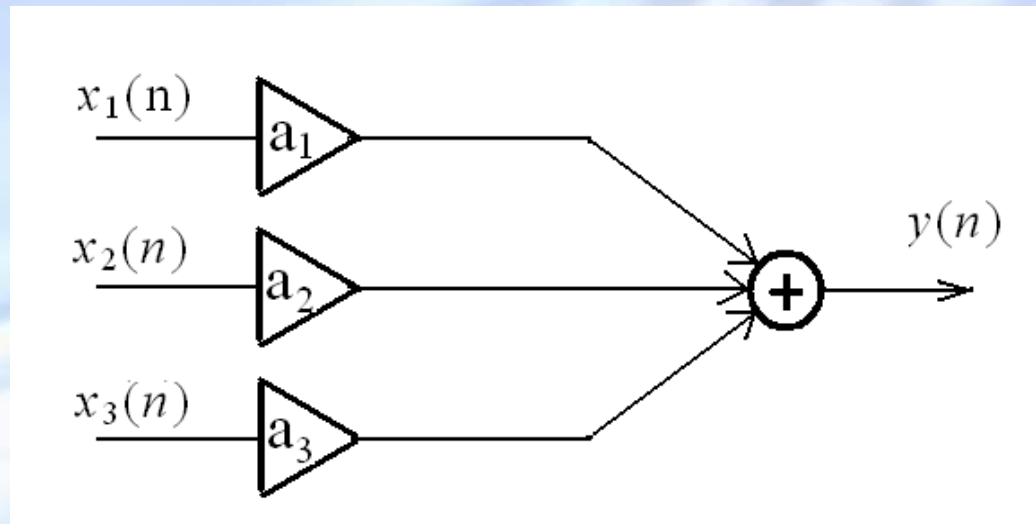
Το στοιχείο καθυστέρησης συγκρατεί στην μνήμη ενδεχομένως q στοιχεία και δίνει σαν έξοδο το στοιχείο q θέσεις ποιο πίσω



Μίξη Ψηφιακού Σήματος

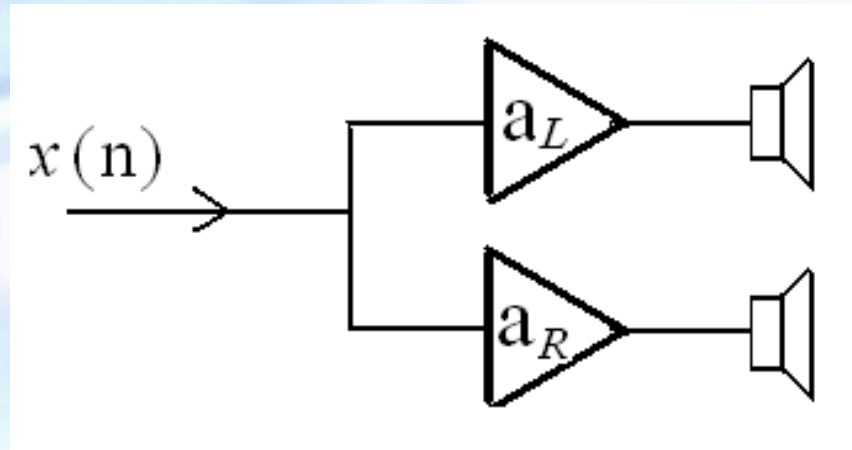
- Η διαδικασία της μίξης έχει ως σκοπό την ένωση πολλών ηχητικών σημάτων σε ένα κανάλι ορίζοντας ταυτόχρονα την σχετική αναλογία του κάθε σήματος.

$$y(n) = a_1x_1(n) + a_2x_2(n) + a_3x_3(n)$$



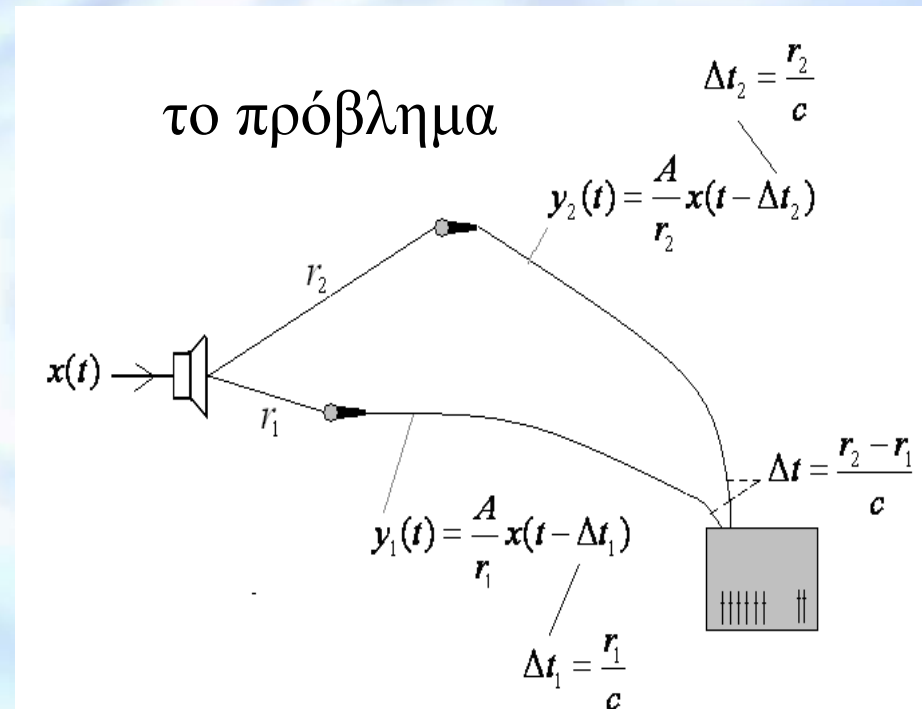
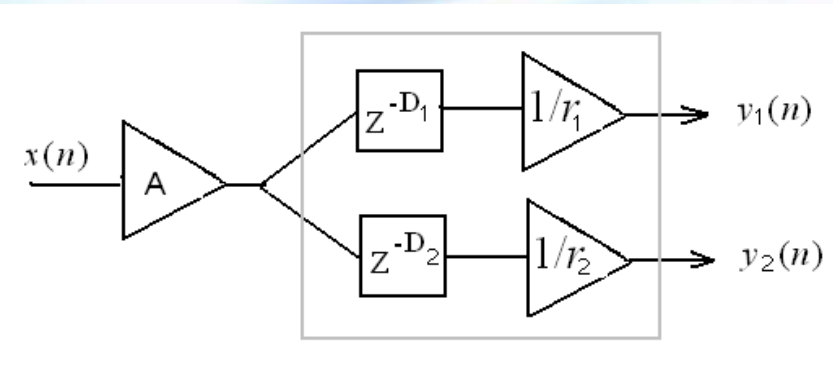
Τροποποίηση ακουστικού ειδώλου

- Δύο ηχεία τοποθετούνται μπροστά από τον ακροατή με τέτοιο τρόπο ώστε να σχηματίζεται ένα ισόπλευρο τρίγωνο.
- Η στερεοφωνική αναπαράσταση επιτυγχάνεται με τη διαφοροποίηση του σήματος στο κάθε ηχείο είτε ως προς την ένταση είτε ως προς το χρόνο. Ορίζεται η τιμή μιας παραμέτρου α τ. ω. $-1 \leq \alpha \leq 1$, και προκύπτουν τα α_L και α_R ως $\alpha_L = 1 - \alpha$ και $\alpha_R = 1 + \alpha$. Η στερεοφωνική εικόνα μετατοπίζεται προς το ηχείο με τη μεγαλύτερη ένταση.



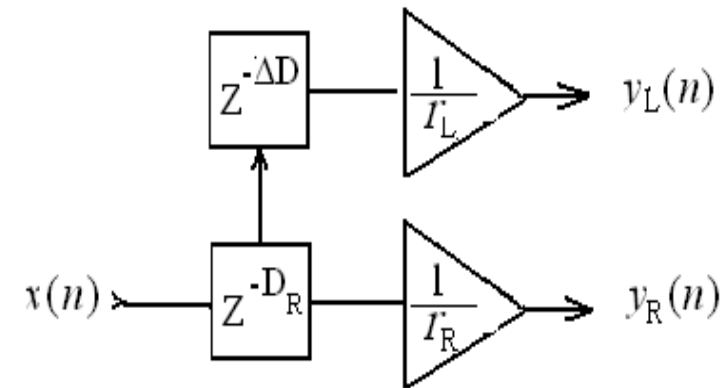
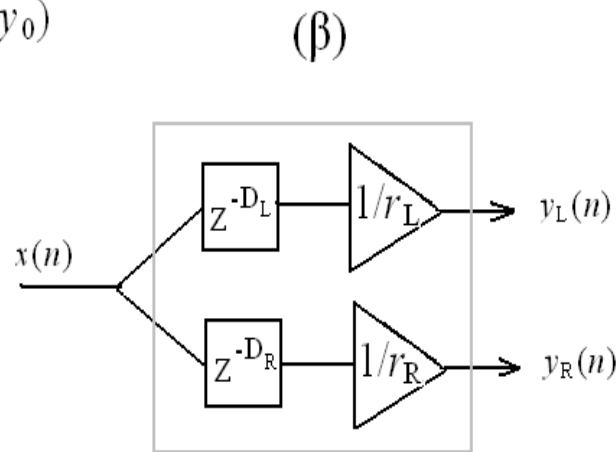
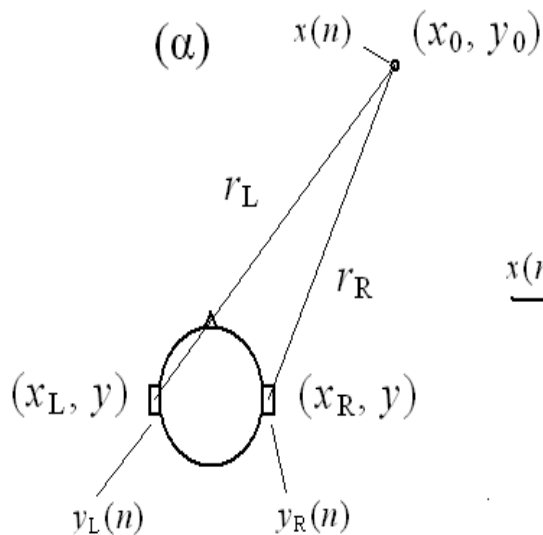
Μοντελοποίηση της Διάδοσης του Ήχου

- Από θεωρία ακουστικής η ηχητική πίεση $p \sim \frac{1}{r}$
- Για δύο σημεία r_1, r_2 ισχύει $\frac{p_1}{p_2} = \frac{r_2}{r_1}$
- Και στα 2 θα φτάνει ήχος με σχετική καθυστέρηση $\Delta t = \frac{r_2 - r_1}{c}$
- Αν με 2 μικρόφωνα μεταφέρω το σήμα σ' ένα μίκτη με δεδομένη γραμμικότητα θα είναι $y_1(t) = \frac{A}{r_1} x(t - \Delta t_1)$ και $y_2(t) = \frac{A}{r_2} x(t - \Delta t_2)$
- Ψηφιοποιημένα
- $y_1(n) = \frac{A}{r_1} x(n - D_1)$ και $y_2(n) = \frac{A}{r_2} x(n - D_2)$
- Η υλοποίησή της
-



Τροποποίηση Ακουστικού Ειδώλου

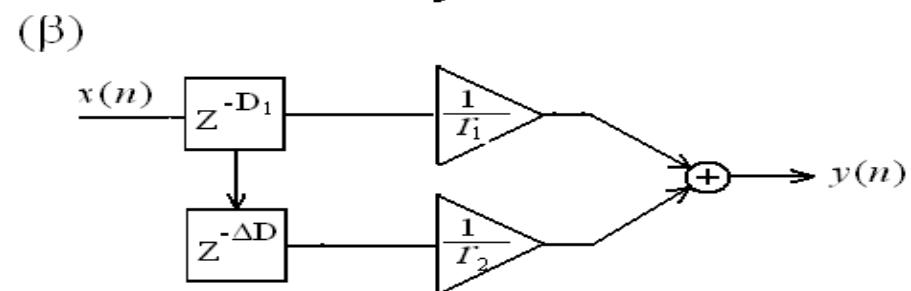
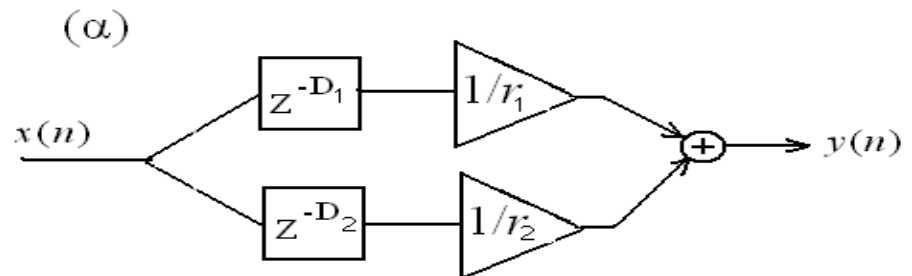
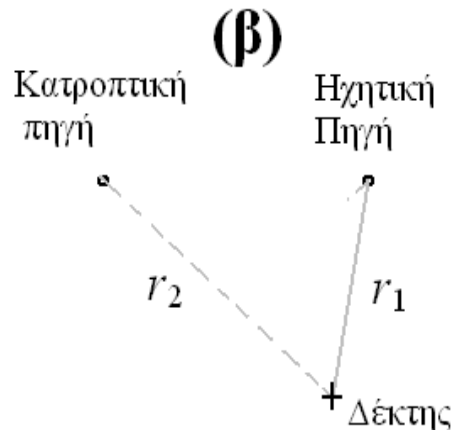
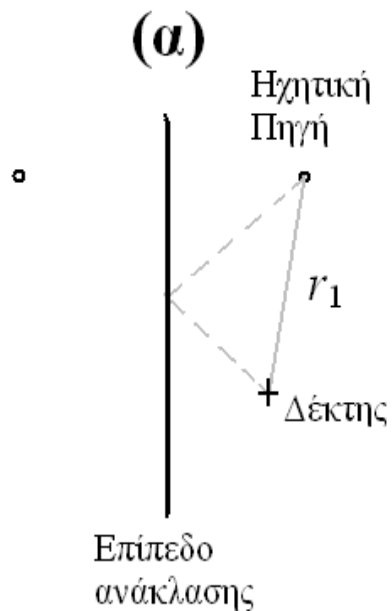
- Ο ήχος που φτάνει στο κάθε αυτί είναι διαφορετικός λόγω τις διαφοράς δρόμου αφού $r_L = \sqrt{(x_L - x_0)^2 + (y - y_0)^2}$ και $r_R = \sqrt{(x_R - x_0)^2 + (y - y_0)^2}$ και αν θεωρήσουμε χρονικές καθυστερήσεις D_L και D_R και τελικά η ψηφιακή προσομοίωση του συστήματος θα ήταν $y_L(n) = \frac{1}{r_L} x(n - D_L)$ και $y_R(n) = \frac{1}{r_R} x(n - D_R)$
- Αν $\Delta D = D_L - D_R$ και η παραπάνω σχέση γράφεται $y_L(n) = y_R(n - \Delta D)$ και $y_R(n) = \frac{1}{r_R} x(n - D_R)$
- Για σχετικά μικρότερο Buffer.



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

Ανάκλαση του Ήχου

- Κάτι ανάλογο με το προηγούμενο γίνεται και στην περίπτωση που έχουμε ανάκλαση από σταθερό εμπόδιο .
- Η έξοδος που θα αναμένουμε από ανάκλαση όπως στο σχήμα
$$y(n) = \frac{1}{r_1} x(n - D_1) + \frac{1}{r_2} x(n - D_2)$$
- Πάλι $D_1 = \frac{r_1}{cTs}$, $D_2 = \frac{r_2}{cTs}$ δηλαδή $\frac{r_1}{r_2} = \frac{D_1}{D_2}$
- Αριστερά το πρόβλημα δεξιά η υλοποίησή του



σχετική καθυστέρηση: $\Delta D = D_2 - D_1$

Ανάκλαση του Ήχου (παράδειγμα)

Μια ηχητική πηγή εκπέμπει μπροστά από ένα επίπεδο ανάκλασης με ένα δέκτη. Η απόσταση του άμεσου και του ανακλώμενου ήχου είναι $r_1=0,312\text{m}$ και $r_2=0,468\text{m}$ 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.

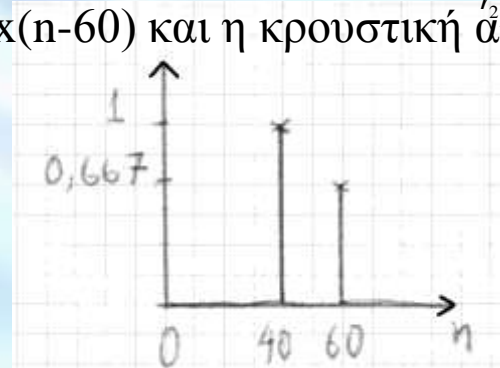
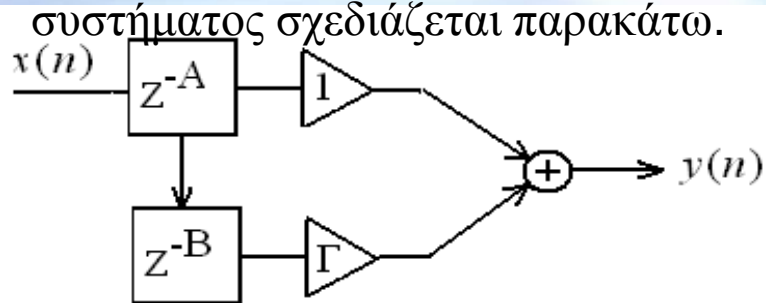
Απάντηση Κατ' αρχάς υπολογίζουμε τις καθυστερήσεις του άμεσου και του ανακλώμενου ήχου

$$D_1 = \frac{0.312}{344 \cdot (1/44100)} = 40 \text{ και } D_2 = \frac{0.468}{344 \cdot (1/44100)} = 60.$$

Επειδή ο ανακλώμενος ήχος φτάνει καθυστερημένα σε σχέση με τον άμεσο χρησιμοποιούμε 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 $\Gamma = \frac{r_1}{r_2} = \frac{0,312}{0,468} = 0.667$.

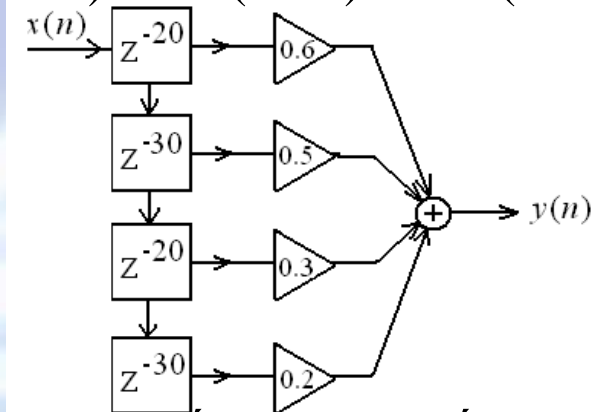
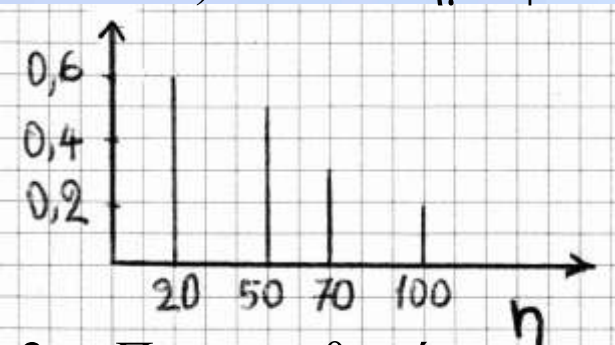
Η εξίσωση διαφοράς είναι $y(n)=x(n-40)+0.667x(n-60)$ και η κρουστική απόκριση του συστήματος σχεδιάζεται παρακάτω.



2 Παραδείγματα

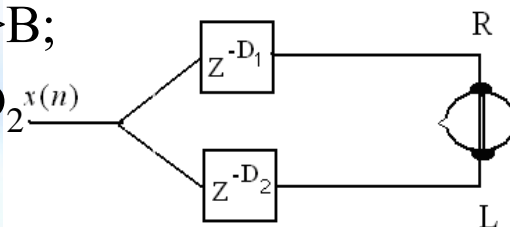
1. Σχεδιάστε ένα σύστημα το οποίο θα είχε την παρακάτω κρουστική απόκριση. Χρησιμοποιήστε όσο το δυνατόν μικρότερο buffer για τα στοιχεία καθυστέρησης.

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.

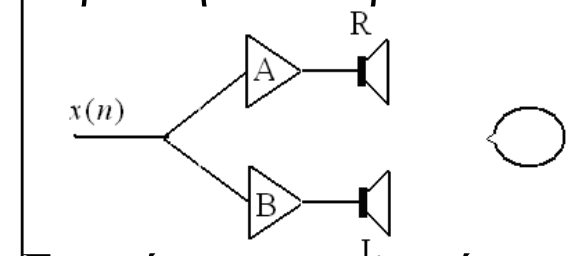


2. Που τοποθετείται το ακουστικό είδωλο σε σχέση με τον ακροατή στο παρακάτω σχήμα αν α) $D_1 > D_2$ και β) $A > B$;

Απάντηση: α) Η σχέση $D_1 > D_2$ υποδηλώνει ότι ο ήχος φτάνει

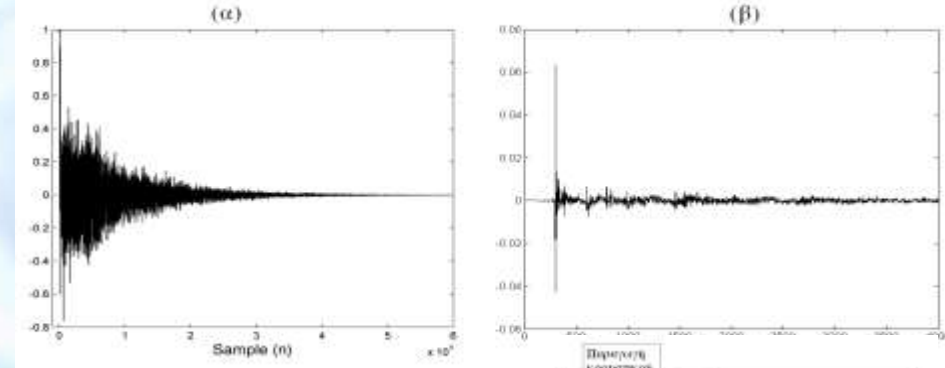
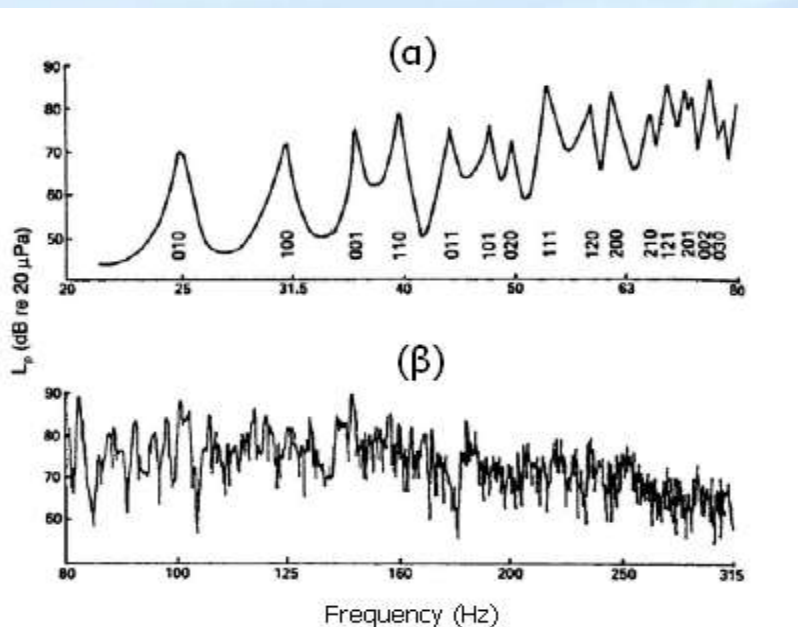


καθυστερημένα στο δεξί αντί σε σχέση με το αριστερό. Επομένως ο ακροατής αντιλαμβάνεται το ακουστικό είδωλο στα αριστερά του. β) Αφού $A > B$ πάει να πει ότι το δεξί ηχείο παίζει πιο δυνατά από το αριστερό και επομένως το ακουστικό είδωλο τοποθετείται δεξιά σε σχέση με τον ακροατή.



Αντήχηση

- Η διατήρηση ενός ήχου (εκθετική απόσβεση) σ' ένα κλειστό χώρο μετά την ακαριαία διακοπή του.
- 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.
- Η αντήχηση ενός χώρου περιγράφεται από την *κρουστική απόκριση* του. Πειραματικά επιτυγχάνεται με την δημιουργία κάποιου κρότου την καταγραφή του ήχου σε κάποιο σημείο με την χρήση μετασχηματισμού Fourier έχουμε την φασματική απόκριση του χώρου.

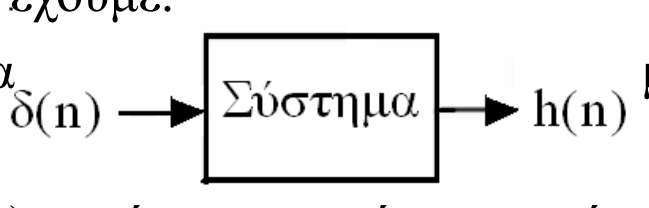


Κλειστός Χώρος - Γραμμικό Σύστημα

• Η κρουστική απόκριση ενός συστήματος είναι η αντίδρασή του αν είσοδος σε αυτό είναι ο «μοναδιαίος» παλμός. Δηλαδή η συνάρτηση δ

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

• Στην περίπτωση αυτή η έξοδος του συστήματος $h(n)$ δίνει πληροφορία για οσοδήποτε περίπλοκη είσοδο έχουμε.

• Έστω το σύστημα  με $h(n)$ η απόκριση του συστήματος

• Αν δεν έχουμε $\delta(n)$ και έχουμε εν γένει, μια είσοδο της μορφής $x(n)$

• 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:

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

• Η έξοδός του είναι $y(n) = 1h(n) + 2h(n-1) + 3h(n-2) + 2h(n-3) + 1h(n-4)$,

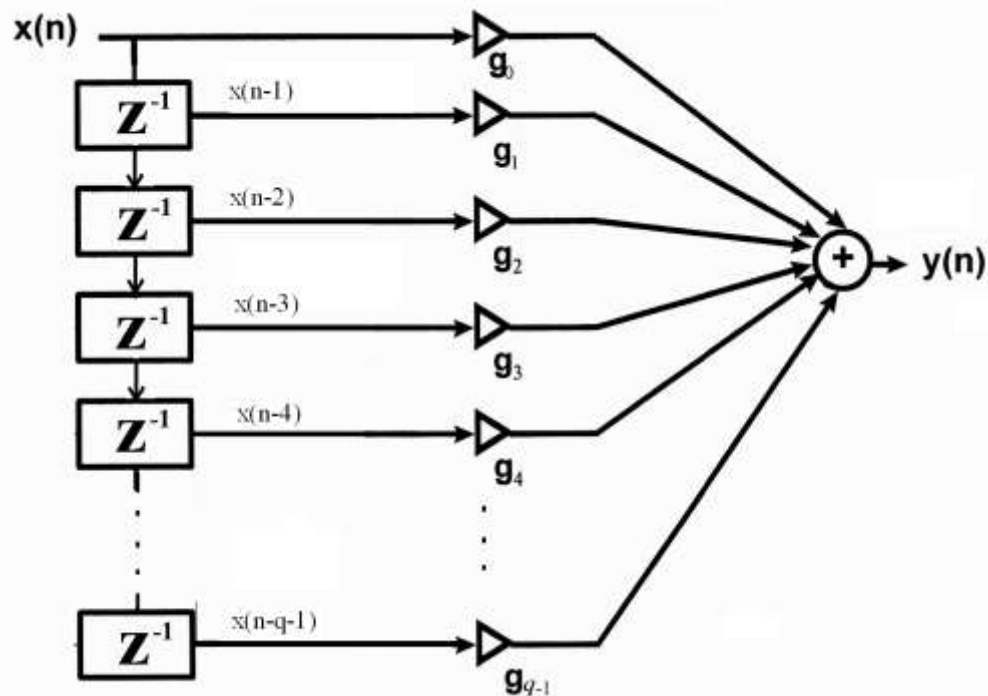
• Στην γενική περίπτωση: Σύστημα με απόκριση $h(n)$ και είσοδο $x(n)$ θα έχει έξοδο την συνέλιξη της απόκρισης με την είσοδο $y(n) = h(n) \cdot x(n) = \sum_{k=0}^{\infty} h(k)x(n-k)$

• Αν λοιπόν η απόκριση είναι ένα σήμα g δειγμάτων γράφεται

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

Δημιουργία Ψηφιακής Αντήχησης

- Έχοντας υπόψη τα προηγούμενα παρατηρούμε ότι το σήμα εξόδου στην ουσία είναι μια σειρά από ανακλάσεις με ένταση g_i . Άρα θέλαμε να φτιάξουμε ένα φίλτρο το οποίο θα έχει την ίδια συμπεριφορά με ένα χώρο δοσμένης κρουστικής απόκρισης μήκους q δειγμάτων χρειαζόμαστε q στοιχεία πολλαπλασιασμού (τρίγωνα) και $q-1$ στοιχεία καθυστέρησης (z^{-1}).
- Ακόμα 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 $N+q-1$.
- Η προσθήκη αντήχησης βάσει της χρήσης της συνέλιξης λέγεται “convolution reverb”



Παράδειγμα

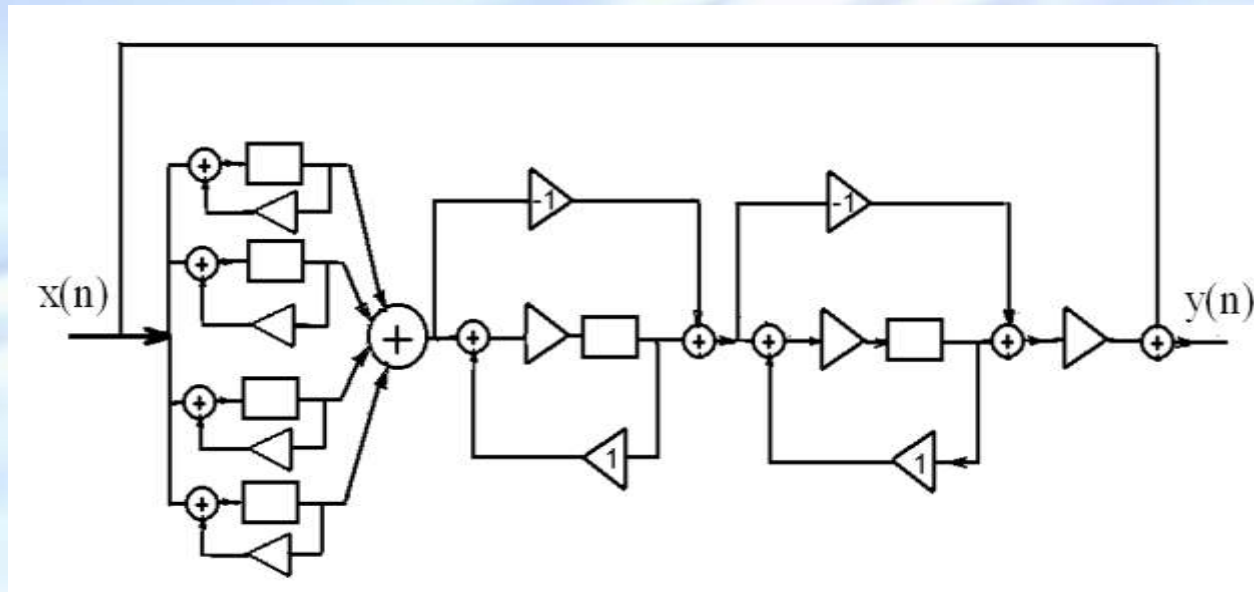
- Η κρουστική απόκριση ενός χώρου έχει χρόνο αντήχησης $RT_{60}=2$ sec και χρησιμοποιείται για τη λειτουργία ενός convolution reverb με συχνότητα δειγματοληψίας $F_s=44100$. Θεωρώντας ότι το σήμα είναι μονοφωνικό
 - a) Πόσοι πολλαπλασιασμοί πρέπει να γίνονται ανά περίοδο δειγματοληψίας για την υλοποίηση της εφαρμογής;
 - b) Να βρεθεί ο κύκλος λειτουργίας T_{inst} του επεξεργαστή που απαιτείται για την υλοποίηση της εφαρμογής.
 - c) 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 $44100 \cdot 2 = 88200$ δείγματα. Αυτό σημαίνει ότι απαιτούνται 88200 πολλαπλασιασμοί ανά περίοδο δειγματοληψίας.
- b) 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
$$T_{inst} = \frac{T_s}{88200} \Rightarrow T_{inst} = \frac{2.27 \cdot 10^{-5}}{88200} = 2,57 \cdot 10^{-10} = 0,257n \text{ sec}$$
- c) 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.

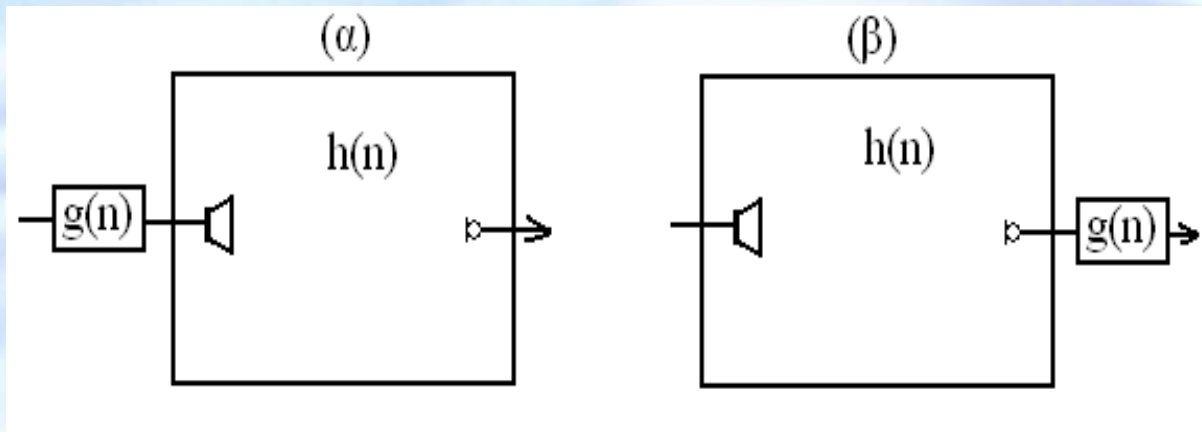
IIR Φίλτρα για Αντήχηση

- **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 $\frac{1}{0.257} \cdot 10^9 = 3,89GHz$!!!! Πράγμα που σημαίνει ακόμα και για σύγχρονο υπολογιστή σχεδόν αποκλειστική απασχόληση.
- Έχουν προταθεί IIR φίλτρα τα οποία κάνουν την ίδια δουλειά τη μορφής του σχήματος



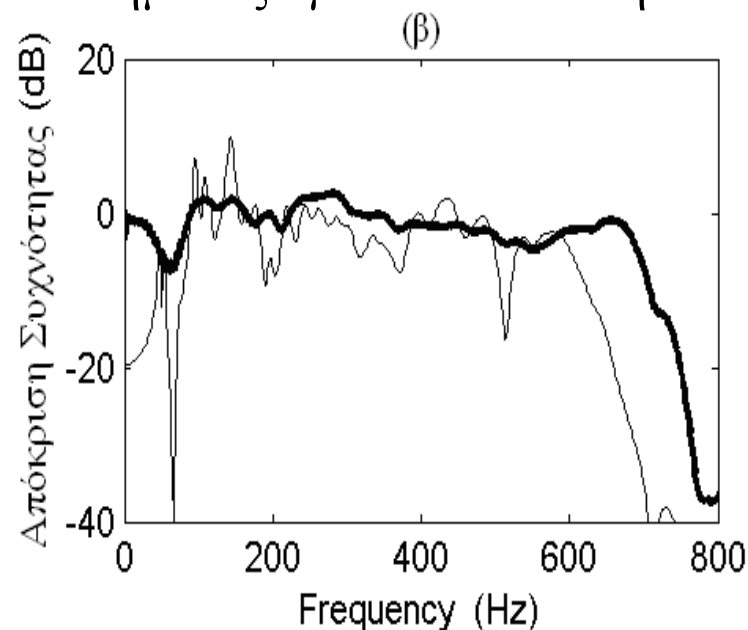
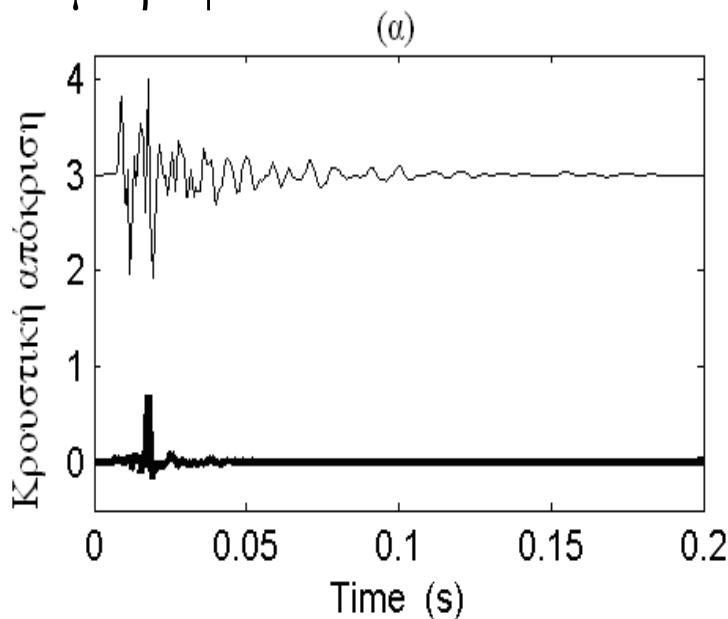
Ισοστάθμιση και Αφαίρεση Αντήχησης

- Η αντήχηση μπορεί να είναι επιθυμητή (συναυλίες- συμφωνική μουσική αλλά και ανεπιθύμητη(συνεδριακά κέντρα ομιλίες).
- Ισοστάθμιση κάνει το αντίστροφο της προηγούμενης κουβέντας μας. Την από-συνέλιξη δηλαδή του σήματος με την Κρουστική Απόκριση.
- Έστω $h(n)$ η Κ.Α και $H(f)$ η αντίστοιχη απόκριση συχνότητας θέλουμε $h(n) * g(n) = \delta(n)$ και $H(f) \cdot G(f) = 1$.
- Η ισοστάθμιση αυτή μπορεί να γίνει πριν την έξοδο (κινηματογράφος) είτε μετά την λήψη ηχογράφηση μουσικής ή ομιλίας, τηλεδιάσκεψη, κινητή τηλεφωνία κτλ.



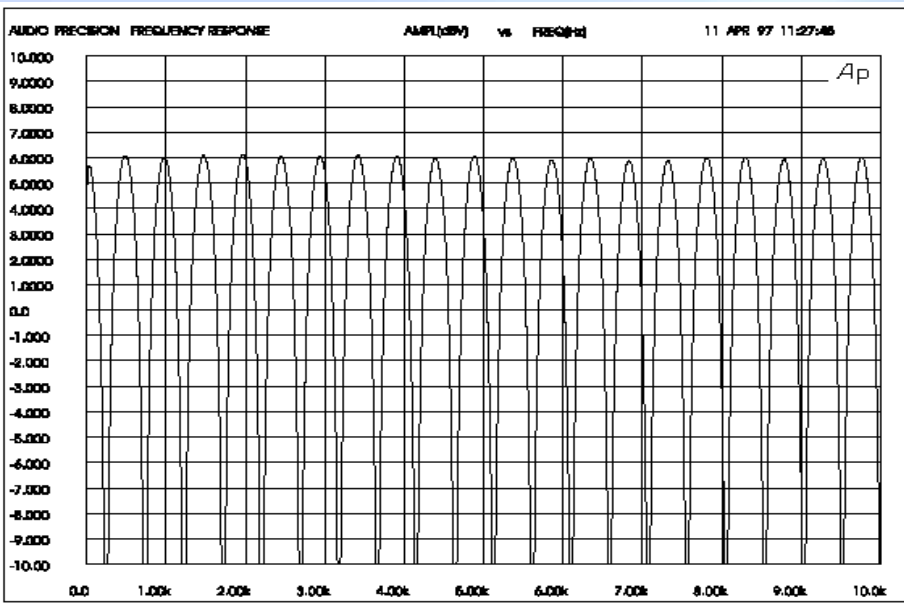
Ισοστάθμιση και Αφαίρεση Αντήρησης

- Ο ορισμός του αντίστροφου φίλτρου $g(n)$ δεδομένης της Κ.Α $h(n)$ δεν είναι πάντα δυνατός, γιατί το αντίστροφο φίλτρο θα πρέπει να είναι αιτιατό και ευσταθές, και η αντιστροφή του μπορεί να μην οδηγήσει σε τέτοιο, βρίσκεται φίλτρο $g(n)$ που αν και δεν κάνει την τέλεια αντιστροφή, είναι βέλτιστο βάσει κάποιων κριτηρίων.
- Στο σχήμα βλέπουμε την Κ.Α $h(n)$ (α) και τη Σ.Α. $|H(f)|$ (β) όπως μετρήθηκαν σε κλειστό χώρο. Για την ισοστάθμιση έγινε χρήση ενός FIR φίλτρου $g(n)$ που επέδρασε πριν την αναπαραγωγή. Με την πιο παχιά γραμμή φαίνεται στο το αποτέλεσμα της ισοστάθμισης. Στο (α) το σύνθετο σύστημα απέκτησε μια κρουστική απόκριση που μοιάζει με ένα καθυστερημένο μοναδιαίο παλμό, πράγμα που οφείλεται στο χρονικό διάστημα που χρειάζεται ο ήχος για να ταξιδέψει από το ηχείο στο μικρόφωνο. Η Σ.Α. του σύνθετου συστήματος έγινε πιο επίπεδη.

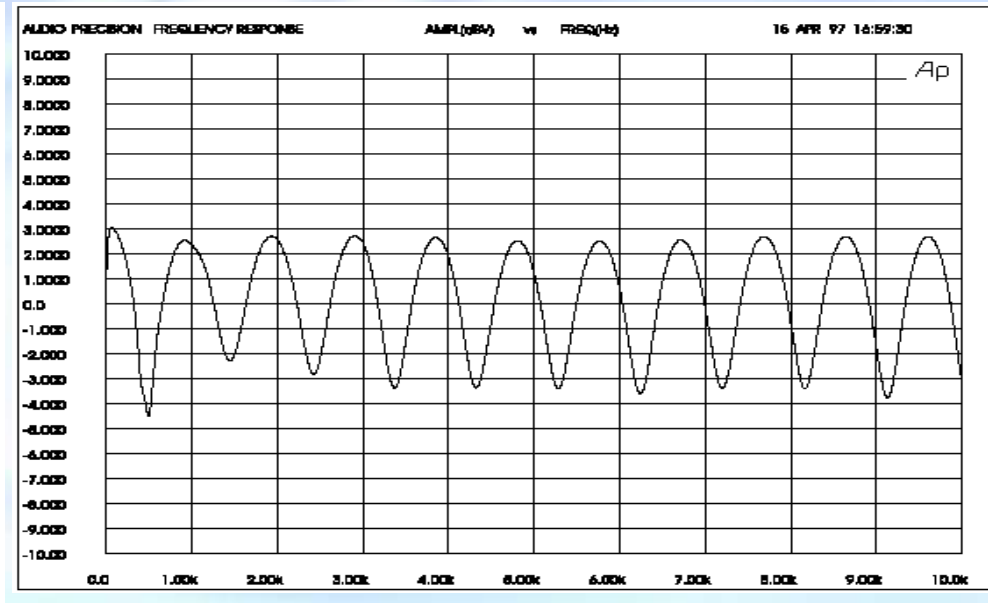


Παραμόρφωση Φίλτρων Τύπου Κτένας

- Όταν σήμα που έρχεται καθυστερημένο, συμβάλει με το αρχικό έχουμε παραμόρφωση των φίλτρων τύπου κτένας (combfilter), πχ όταν το ίδιο πρόγραμμα παίζεται από δύο ηχεία, ο ήχος από αυτό που είναι κοντύτερα σε εμάς, αλλοιώνεται από αυτό που είναι μακρύτερα, ή όταν λαμβάνουμε την ίδια πηγή με δύο μικρόφωνα σε διαφορετική απόσταση μεταξύ τους, αλλά ακόμα και όταν επιστρέφουμε την έξοδο κάποιου ψηφιακού effects processor στην κονσόλα μας και την μιξάρουμε με το dry σήμα. Τα φίλτρα τύπου κτένας, δεν μπορούν να αντιμετωπιστούν με τα equalizers, αλλά μπορούν να μειωθούν αισθητά με τον συγχρονισμό των σημάτων μεταξύ τους.



Στο αρχικό σήμα έχει προστεθεί ο εαυτός του στη ίδια στάθμη, με καθυστέρηση 2msec. Μέγιστη αύξηση 6dB αρκετά μεγάλη μέγιστη μείωση.



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.