

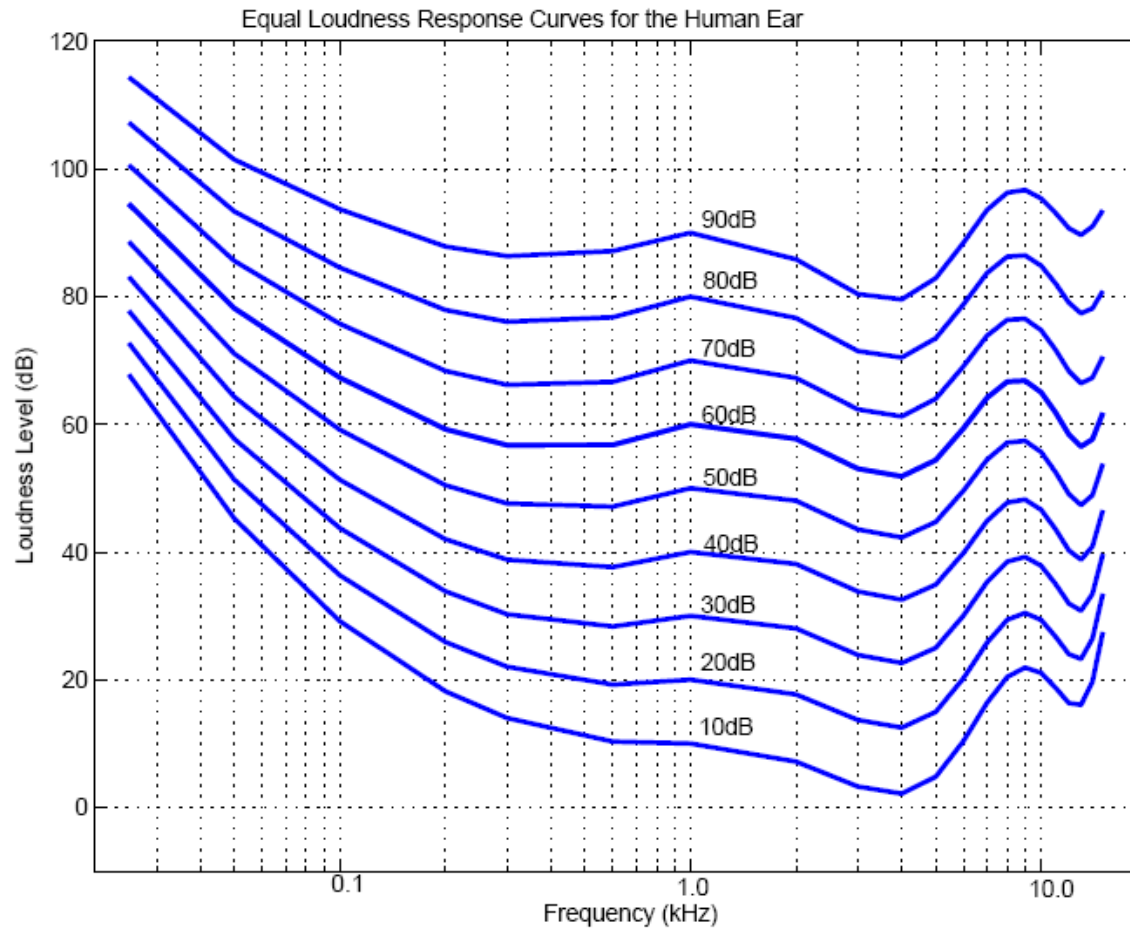
Κωδικοποίηση ήχου

Σύστημα ακοής
MP3 / MP4
Κωδικοποίηση φωνής

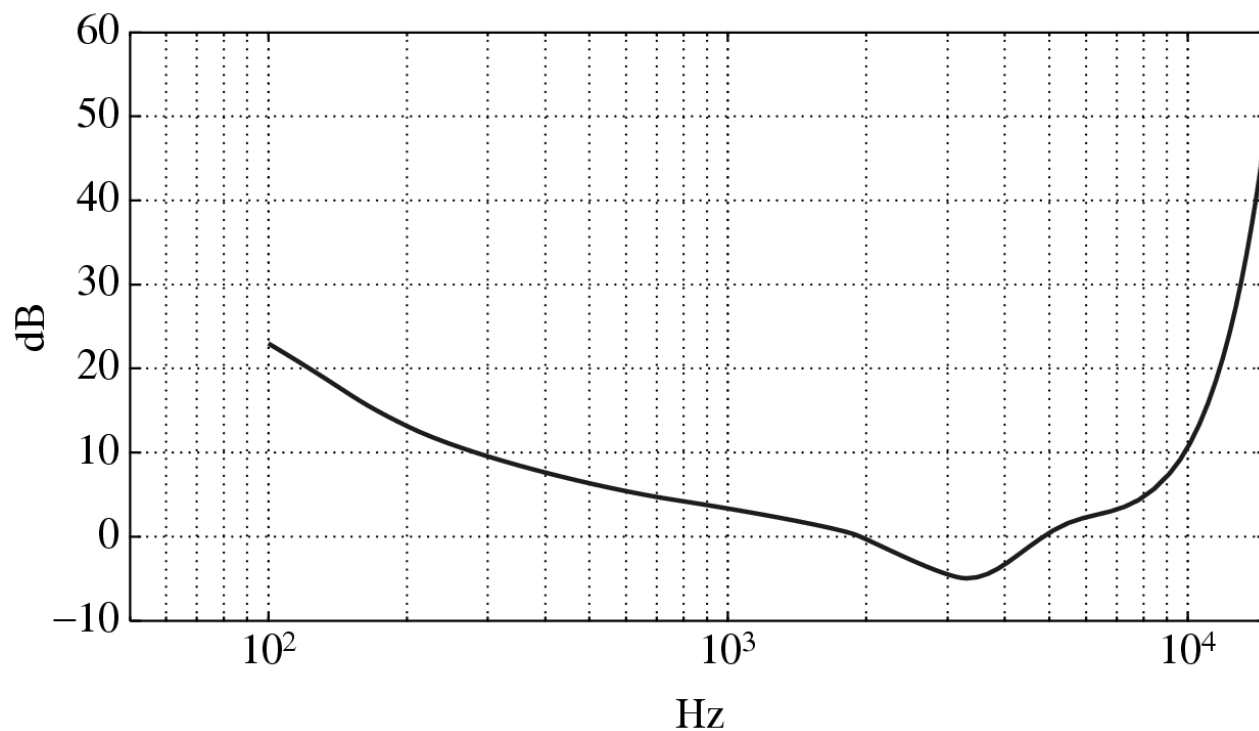
T. Painter and A. Spanias, **Perceptual Coding of Digital Audio**,
Proceedings of the IEEE, pp. 451-513, April 2000.

P. Noll, **MPEG digital audio coding**,
IEEE Signal Processing Magazine, pp. 59-81, Sept. 1997.

Καμπύλες ηχηρότητας Fletcher-Munson

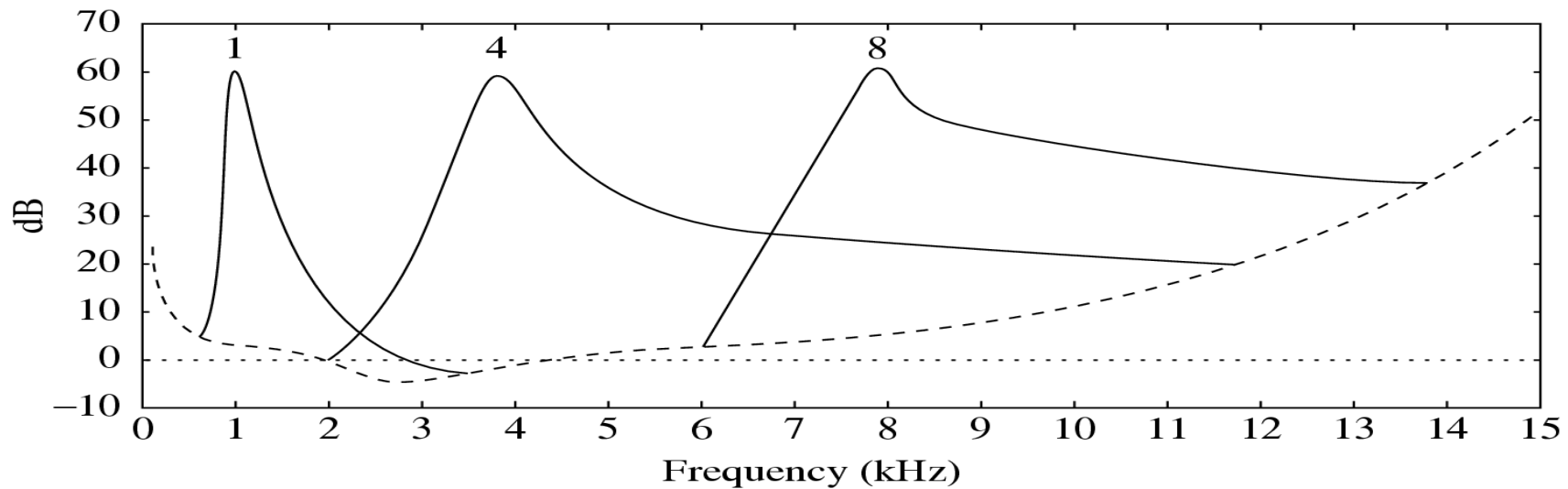
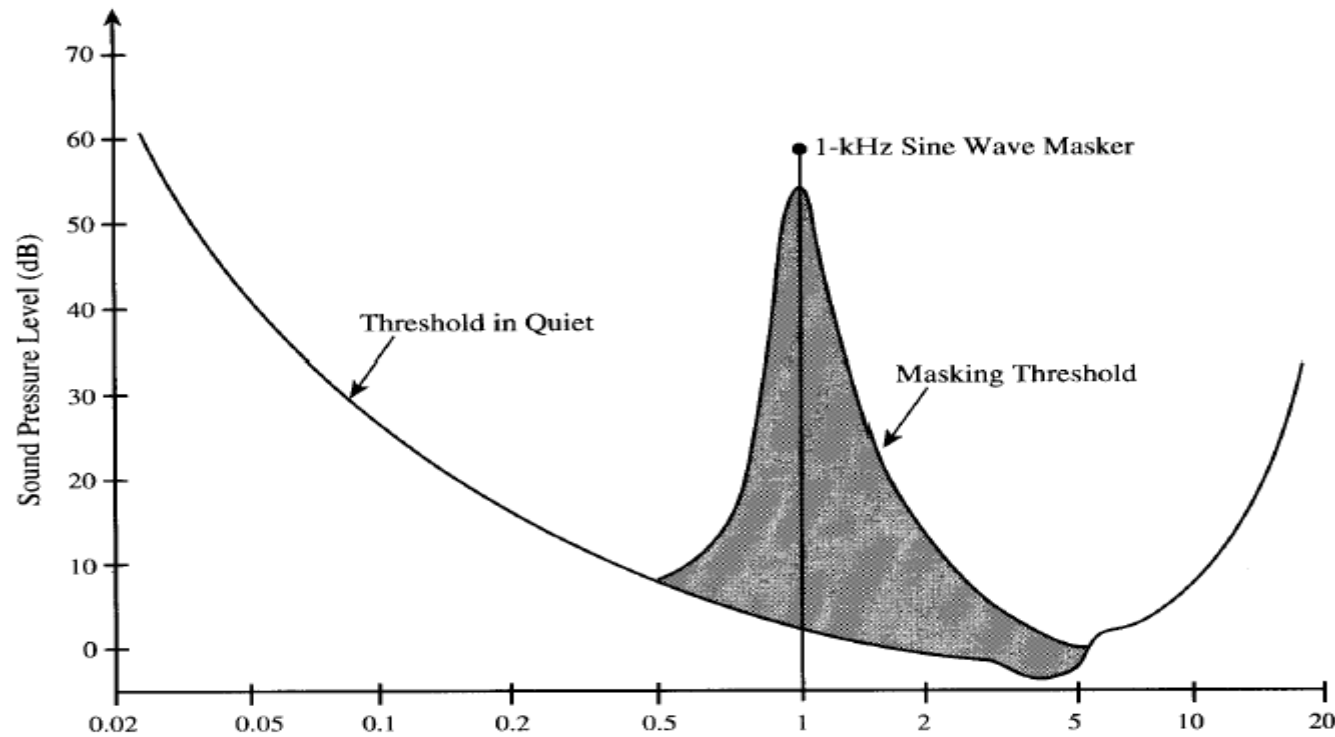


Όριο ακοής



Καθαροί τόνοι, περιβάλλον χωρίς θόρυβο

Ακουστική σκίαση συχνοτήτων



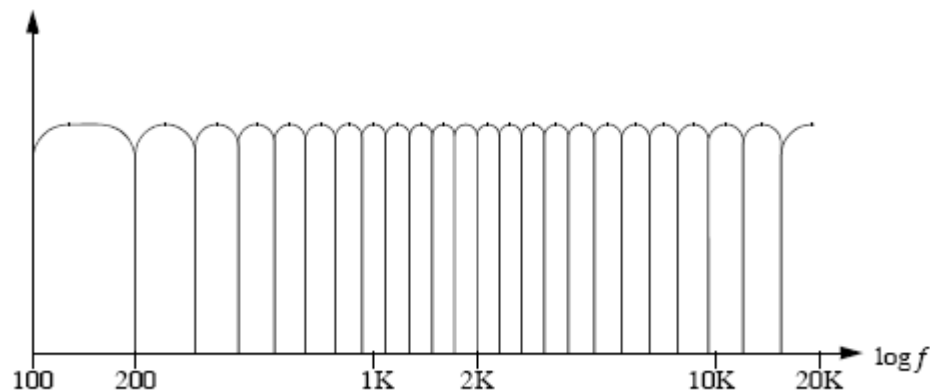
Κρίσιμες ζώνες συχνοτήτων (1/2)

Σύστημα ακοής: επικαλυπτόμενα φίλτρα

Εύρος ζώνης 50 έως 100 Hz για σήματα κάτω των 500 Hz

και μέχρι 5000 Hz για υψηλές συχνότητες

25 κρίσιμες ζώνες



$$b = 13 \arctan(0.76 f) + 3.5 \arctan\left(\frac{f^2}{56.25}\right)$$

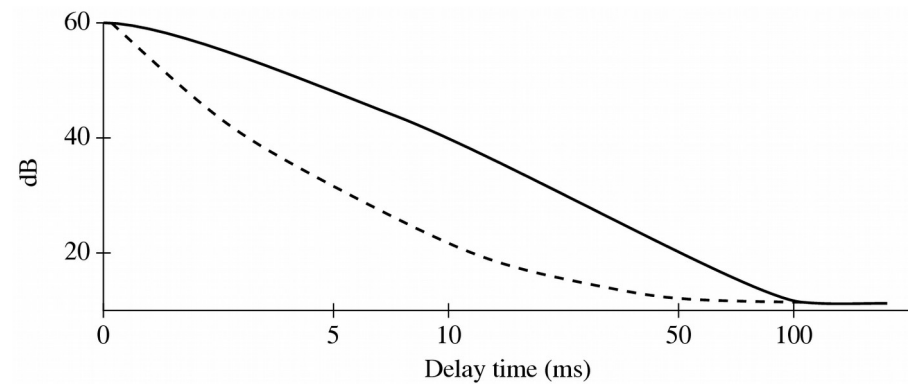
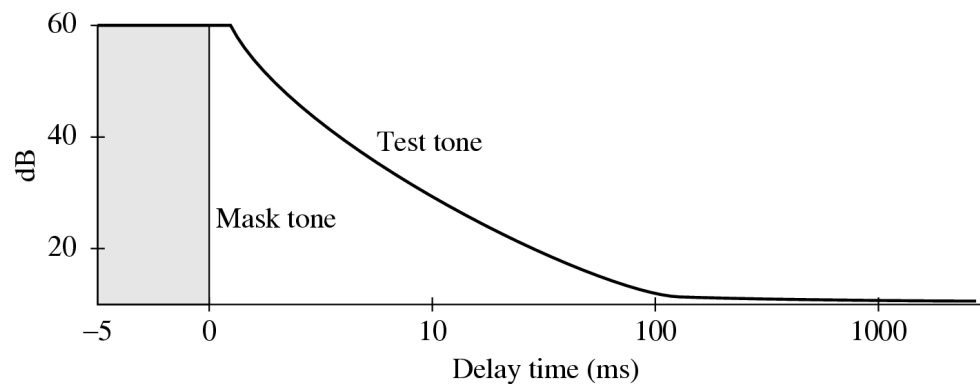
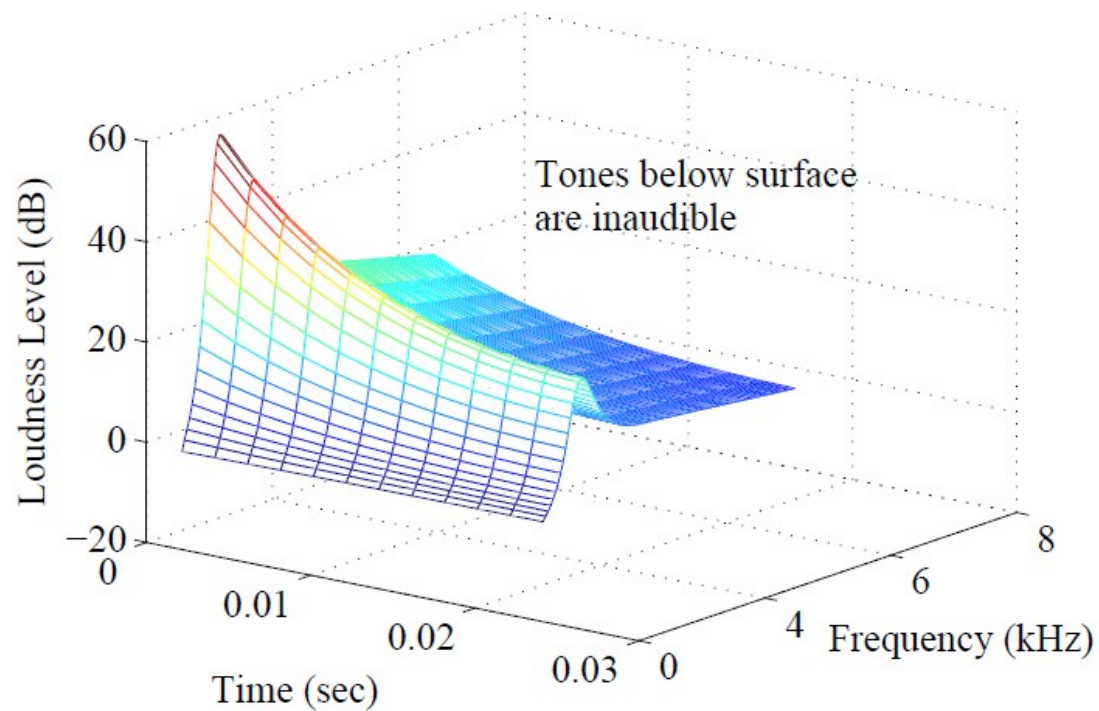
$$\Delta f = 25 + 75(1 + 1.4 f^2)^{0.69}$$

f kHz, Δf Hz

Κρίσιμες ζώνες συχνοτήτων (2/2)

Band No.	Center Freq. (Hz)	Bandwidth (Hz)	Band No.	Center Freq. (Hz)	Bandwidth (Hz)	Band No.	Center Freq. (Hz)	Bandwidth (Hz)
1	50	-100	10	1175	1080-1270	19	4800	4400-5300
2	150	100-200	11	1370	1270-1480	20	5800	5300-6400
3	250	200-300	12	1600	1480-1720	21	7000	6400-7700
4	350	300-400	13	1850	1720-2000	22	8500	7700-9500
5	450	400-510	14	2150	2000-2320	23	10,500	9500-12000
6	570	510-630	15	2500	2320-2700	24	13,500	12000-15500
7	700	630-770	16	2900	2700-3150	25	19,500	15500-
8	840	770-920	17	3400	3150-3700			
9	1000	920-1080	18	4000	3700-4400			

Ακουστική χρονική σκίαση



Χρονική διάρκεια τόνου σκίασης

Τεχνολογίες κλειδιά

Ακουστική σκίαση

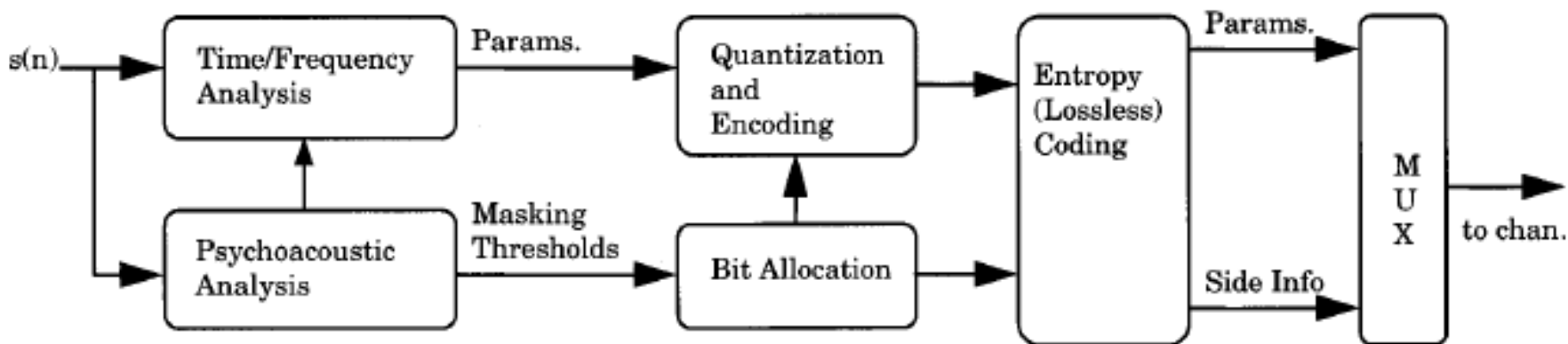
Προσδιορισμός μιας συνάρτησης της συχνότητας:

σήμα / επίπεδο σκίασης

Βήμα κβαντισμού κι επομένως κατανομή bits

Διαφανής συμπίεση = χωρίς αντιληπτή διαφορά

Συμπίεση στο πεδίο του μετασχηματισμού



Συμπίεση MPEG-1

CD-Audio : $2 \times 44100 \text{ samples/sec} \times 16 \text{ bits/sample} = 1.41 \text{ Mbits/sec}$

	Διαφανής ποιότητα		Βαθμός συμπίεσης
Layer I	384 kbits/sec	DCC	4
Layer II	192 kbits/sec	DAB, CD-I, DVD	8
Layer III	128 kbits/sec	ISDN, Internet, Δορυφορικό	12

Δειγματοληψία : 32 kHz, 44.1 kHz, 48 kHz

The diagram illustrates the system architecture with the following components and data flow:

- Input:** A single input line splits into two paths.
- Φίλτρα ανάλυσης (Analysis Filters):** Receives the top input path and outputs six parallel signals to the **Κβαντιστής**.
- FFT:** Receives the bottom input path and outputs a single signal to the **Συνάρτηση μάσκας**.
- Κβαντιστής (Quantizer):** Receives six signals from the **Φίλτρα ανάλυσης** and one signal from the **Κατανομή πόρων και συντελεστών**. It produces six parallel output signals.
- Συνάρτηση μάσκας (Masking Function):** Receives one signal from the **FFT** and one signal from the **Κατανομή πόρων και συντελεστών**. It outputs a single signal to the **Κατανομή πόρων και συντελεστών**.
- Κατανομή πόρων και συντελεστών (Resource and Coefficient Allocation):** Receives one signal from the **Συνάρτηση μάσκας** and one signal from the **Κβαντιστής**. It outputs a single signal to the **Κβαντιστής** and another single output signal.

10

Layers I και II : Φίλτρα ανάλυσης

32 φίλτρα ίσου εύρους : 750 Hz για 48 kHz

Βαθυπερατό φίλτρο και διαμόρφωση μέσω DCT

Ανάλυση σε 32 ζώνες συχνοτήτων

Quadrature-Mirror Filter

$$h_k(n) = h(n) \cos\left(\frac{(2k-1)\pi n}{2M} + \varphi_k\right)$$

$$k = 1, \dots, M=32$$

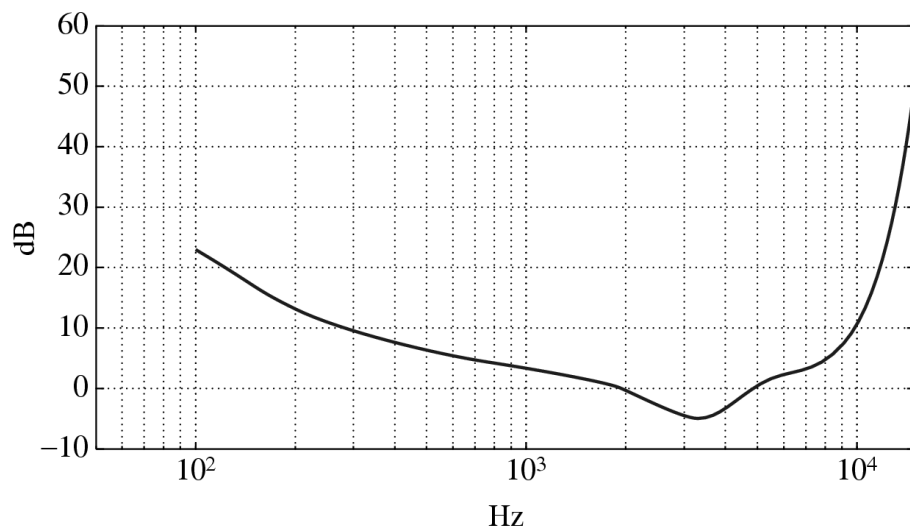
$$n = 1, \dots, N=512$$

Συμπίεση MPEG-1 (Layer II)

Ισχύς

Δυναμική κατανομή bits (128 kbits/sec)

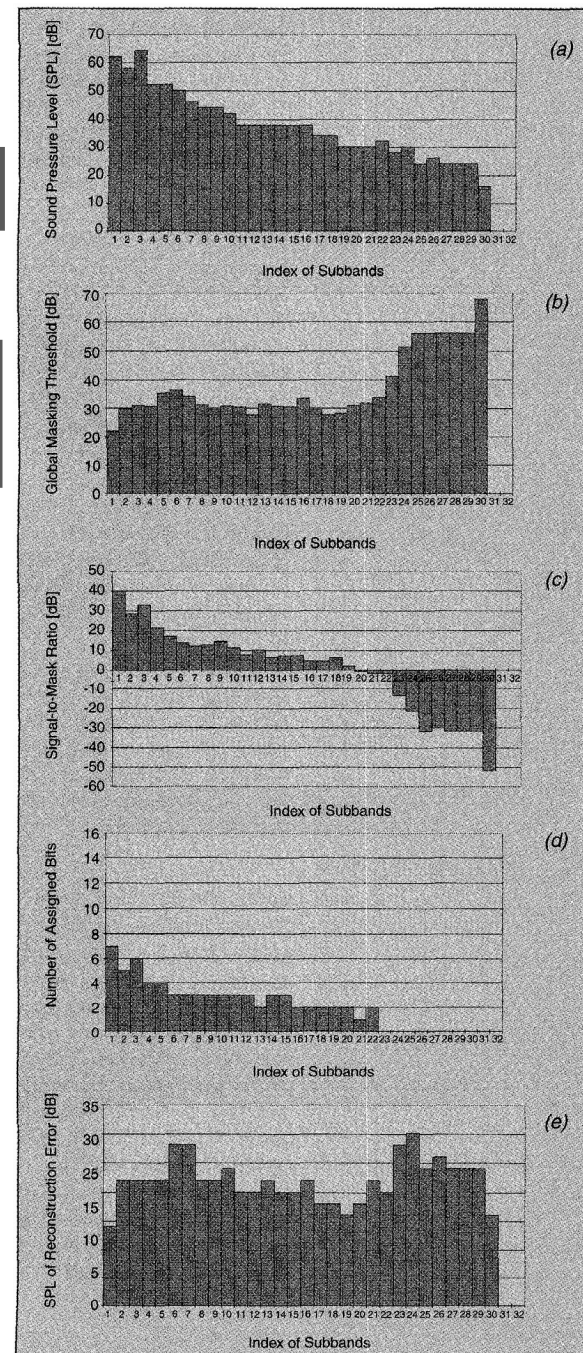
Όριο
σκίασης



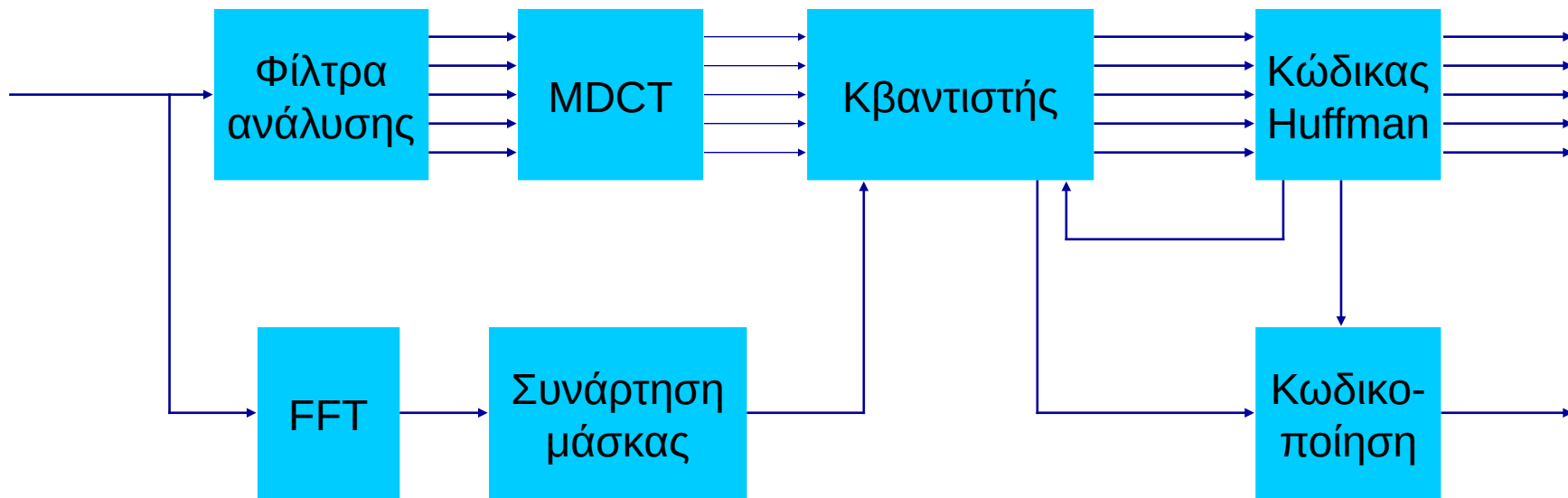
Σήμα/σκιά

Αριθμός bits

Ισχύς σφάλματος



Συμπίεση MPEG-1 (Layer III) / MP3



Τροποποιημένος DCT (50% επικάλυψη μπλοκ, 6/18 συντελεστών)
καλύτερη φασματική ανάλυση

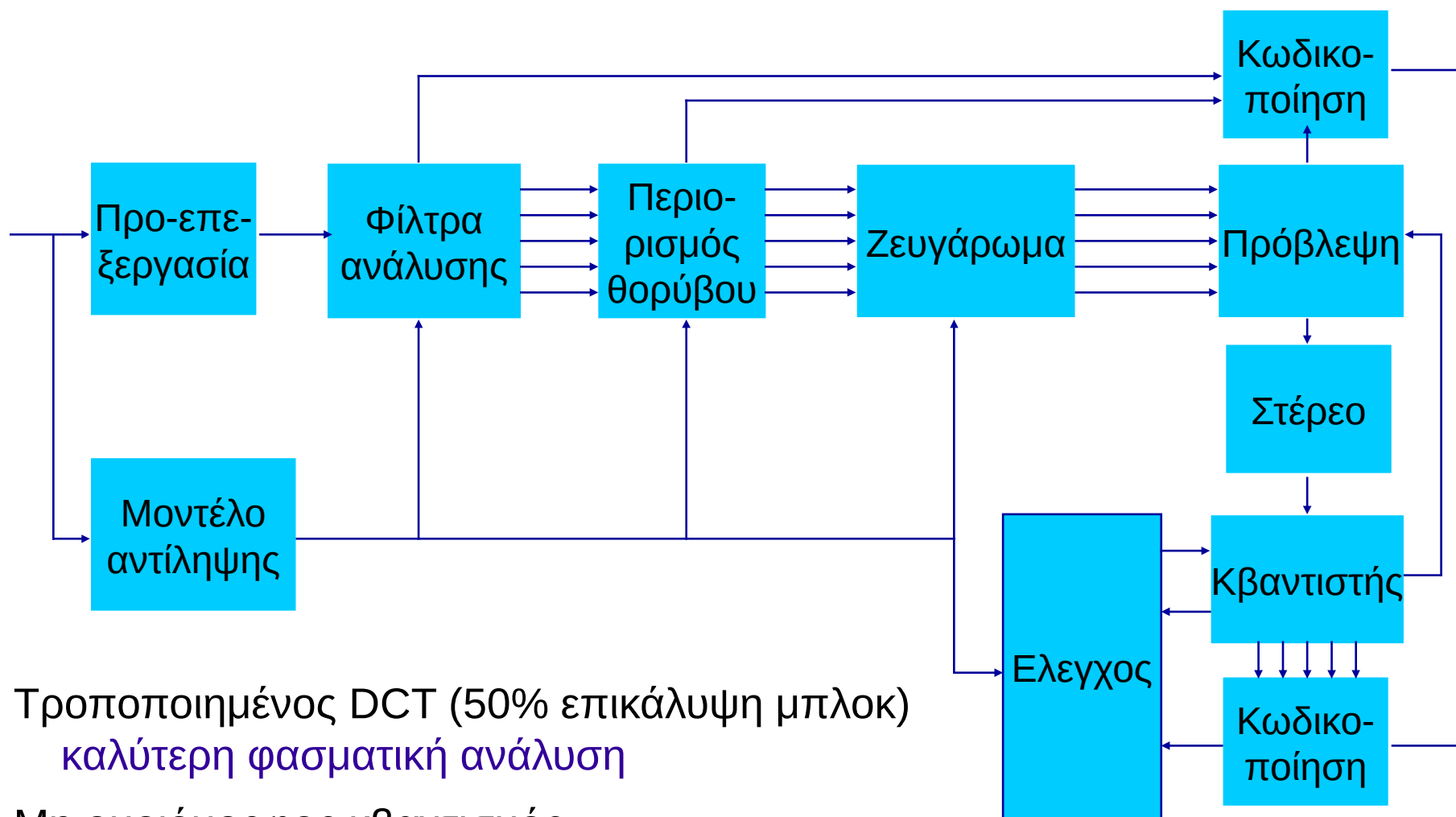
Μη ομοιόμορφος κβαντισμός

32 πίνακες κωδίκων Huffman και διαστήματα «μηδέν»

Συμπίεση MPEG-1 (Layer III) / MP3

Sound Quality	Bandwidth	Mode	Compression Ratio
Telephony	3.0 kHz	Mono	96:1
Better than Short-wave	4.5 kHz	Mono	48:1
Better than AM radio	7.5 kHz	Mono	24:1
Similar to FM radio	11 kHz	Stereo	26 - 24:1
Near-CD	15 kHz	Stereo	16:1
CD	> 15 kHz	Stereo	14 - 12:1

Συμπύεση MPEG-2 AAC (1/2)



Τροποποιημένος DCT (50% επικάλυψη μπλοκ)
καλύτερη φασματική ανάλυση

Μη ομοιόμορφος κβαντισμός

Κωδικοποίηση Huffman

Συμπίεση MPEG-2 AAC (2/2)

Περισσότερες συχνότητες δειγματοληψίας
8 kHz – 96 kHz

Πολυκαναλικά σήματα

Περιορισμός θορύβου μέσω της κατανομής των
σφαλμάτων κβαντισμού στο χρόνο χρησιμοποιώντας
πρόβλεψη στο πεδίο των συχνοτήτων



Τύποι κωδικοποίησης ήχου MPEG-4

- Συνθετική φωνή (TTS)
200-1200 bits/sec
κείμενο, παράμετροι προσωδίας, ρυθμού ομιλίας, ...
διεθνές αλφάβητο φωνημάτων
- Συνθετική μουσική (SAOL)
δομή ορχήστρας, περιγραφή συνθετικής μουσικής
- Παραμετρική κωδικοποίηση φωνής/ήχου
2-4 kbits/sec στα 8 kHz
4-16 kbits/sec στα 8 kHz ή 16 kHz
- Κωδικοποίηση φωνής με γραμμική πρόβλεψη (CELP)
6-24 kbits/sec στα 8 kHz ή 16 kHz
- Κωδικοποίηση ήχου (MPEG-4 AAC)



Συμπίεση MPEG-4 AAC

Βασικό σχήμα : MPEG-2 AAC

Αντικατάσταση θορύβου μέσω γένεσης σημάτων
τύπου θορύβου στον αποκωδικοποιητή

Κλιμάκωση του δυαδικού κώδικα

24 kbits/s (mono), 40 kbits/s (stereo), 56 kbits/s (stereo)

Διανυσματικός κβαντισμός για υψηλή συμπίεση

Μεγαλύτερη αντοχή σε σφάλματα μετάδοσης



Συμπίεση MPEG-4 AAC / MP4

Αντικατάσταση ζωνών τύπου «θορύβου» μέσω γένεσης σημάτων θορύβου στον αποκωδικοποιητή (παραμετρική μοντελοποίηση)

Μακροχρόνια πρόβλεψη

χρήσιμη για στάσιμα αρμονικά σήματα

Διανυσματικός κβαντισμός μετά από διεμπλοκή και στάθμιση

Έκδοση για μικρή καθυστέρηση (20 ms)

Ισχυρή προστασία από σφάλματα μετάδοσης

Κλιμάκωση ποιότητας με αριθμητική κωδικοποίηση (έκδοση 2)

Τυπικός στόχος : 24 kbits/sec/channel

Βασικές παράμετροι σημάτων φωνής

	Συχνότητες Hz	Δειγματοληψία kHz	bits/sample	kbits/sec
Φωνή τηλεφώνου	200-3400	8	8	64
Φωνή ευρείας ζώνης	50-7000	16	8	128

- Υψηλός ρυθμός : 32 kbits/sec
- Μέσος ρυθμός : 8 kbits/sec
- Χαμηλός ρυθμός : 4 kbits/sec
- Πολύ χαμηλός ρυθμός : 2 kbits/sec

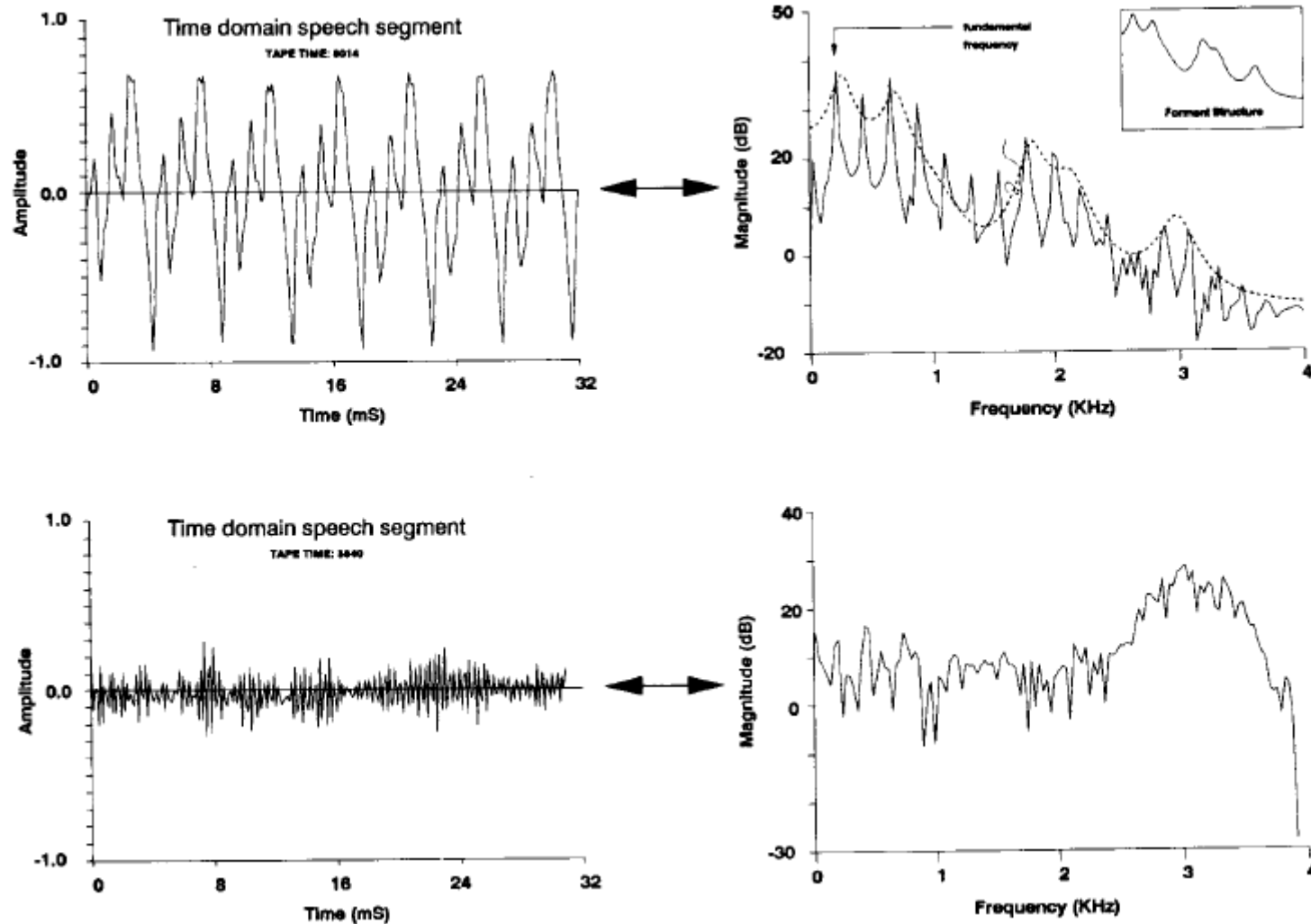
Χαρακτηριστικά σήματος φωνής (1/2)

Μη στάσιμο σήμα, με στάσιμα χαρακτηριστικά
για μικρά διαστήματα, 5-20 msec

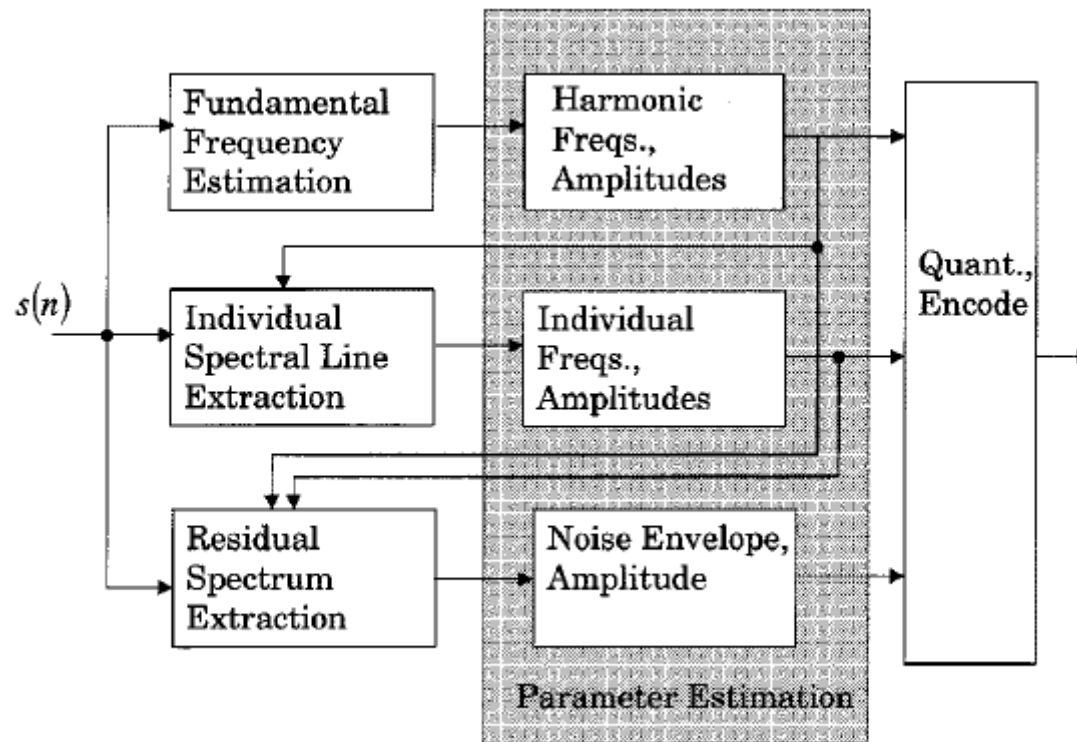
- Φωνήεντα : περιοδικά σήματα
διαμορφωμένες καθαρές συχνότητες (αρμονικές)
- Σύμφωνα : ευρεία ζώνη, σήματα τύπου θορύβου

Συχνότητες συντονισμού της φωνητικής κοιλότητας

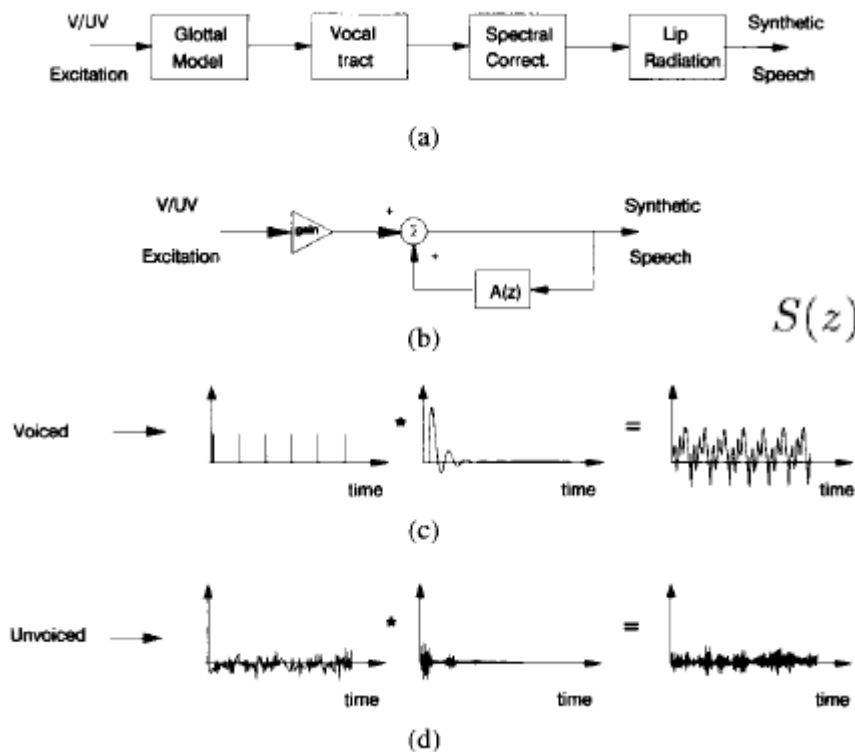
Χαρακτηριστικά σήματος φωνής (2/2)



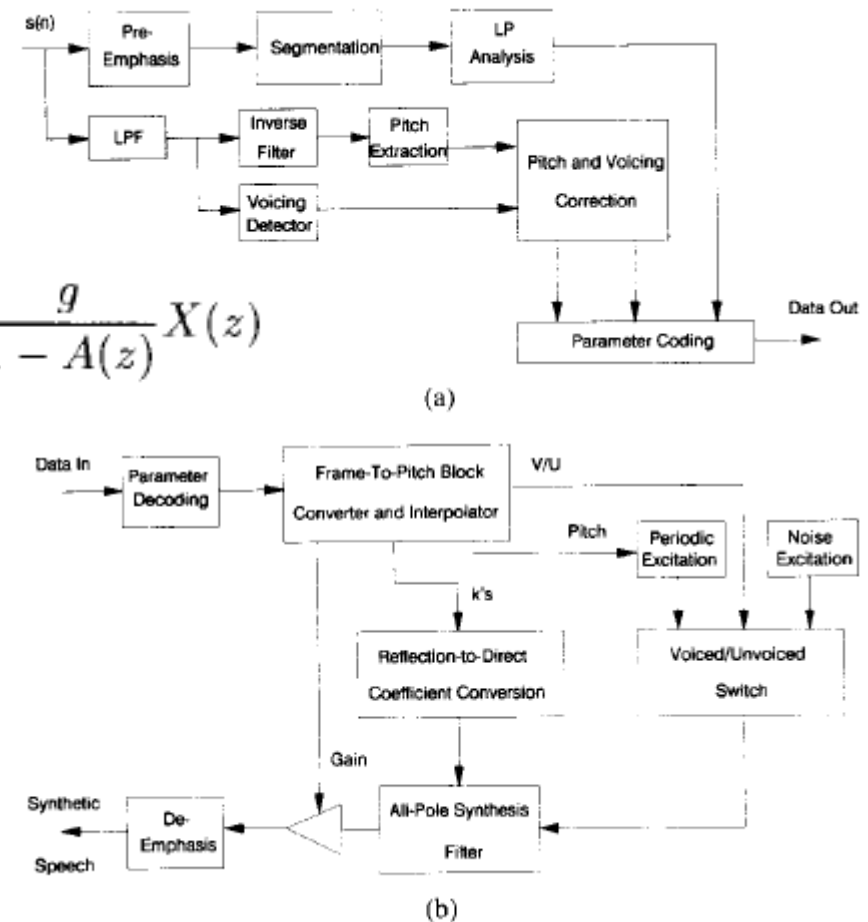
Παραμετρική κωδικοποίηση ήχου



Κωδικοποίηση γραμμικής πρόβλεψης (LPC)



$$S(z) = \frac{g}{1 - A(z)} X(z)$$

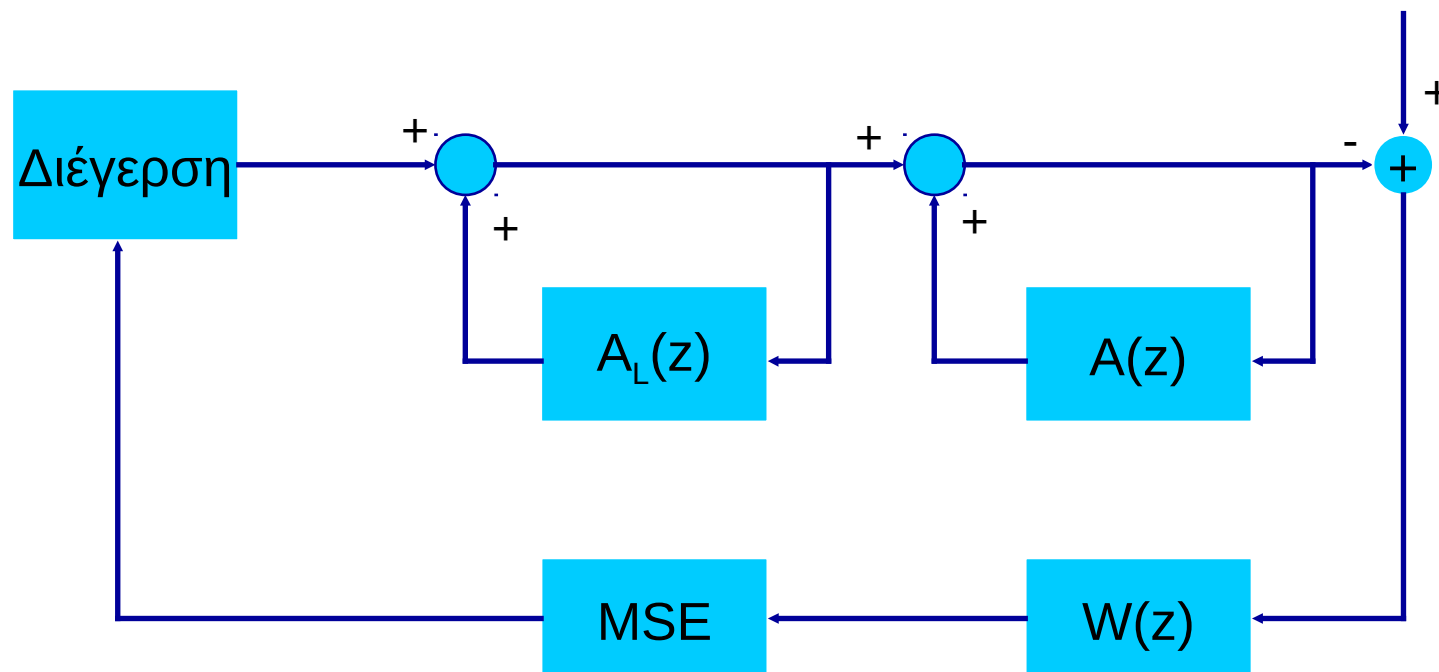


Μοντέλο διέγερσης 2 καταστάσεων
Κέρδος, θεμελιώδης συχνότητα και
απόφαση φωνήεν (ναι/όχι), θόρυβος

LPC-10
2,4 kbits/sec

Κωδικοποίηση με ανάλυση μέσω σύνθεσης

Η διέγερση προσδιορίζεται με βελτιστοποίηση κλειστού βρόχου



Βραχυχρόνιο φίλτρο : διαμόρφωση φωνής

Μακροχρόνιο φίλτρο : θεμελιώδης συχνότητα

Στάθμιση σφάλματος : ανθρώπινη ακοή

Διέγερση μέσω κωδικοποιημένων σημάτων διάρκειας 5 ms

Λεξικό 1024 διανυσμάτων με 40 συνιστώσες