



Diplomski studij

Informacijska i  
komunikacijska tehnologija:

Obradba informacija  
Telekomunikacije i informatika

# Višemedijske komunikacije

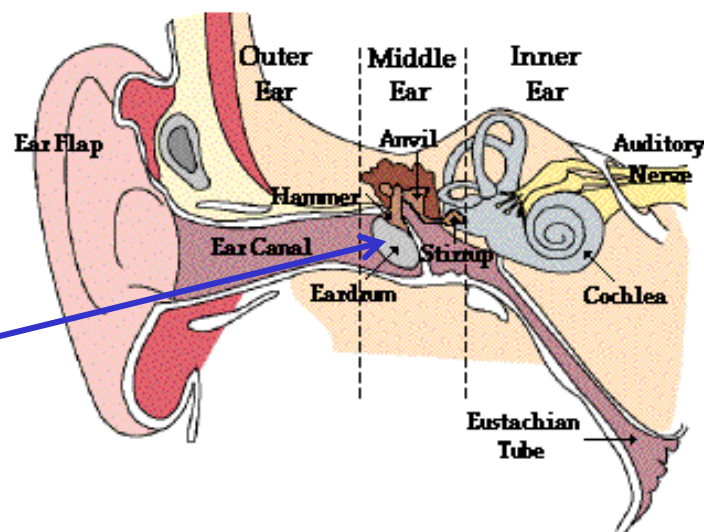
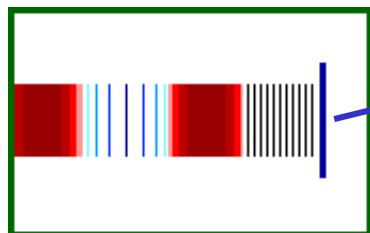
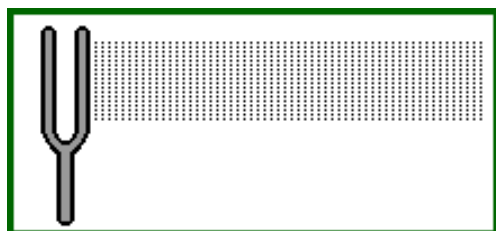
4.

Informacijska svojstva i  
kodiranje zvuka

- Fizikalne osnove, svojstva i percepcija zvuka
- Digitalni zvuk u umreženom računalnom sustavu
- Kriteriji usporedbe i mjerenje kvalitete koda
- Kodiranje govora
  - Karakteristike govora važne za kodiranje
  - Koderi govora
    - Koderi valnog oblika: PCM, ADPCM
    - Koderi zasnovani na modelu: LPC, CELP
- Koderi zvuka u frekvencijskoj domeni
  - Osnovni percepcijski koder
  - MPEG audio

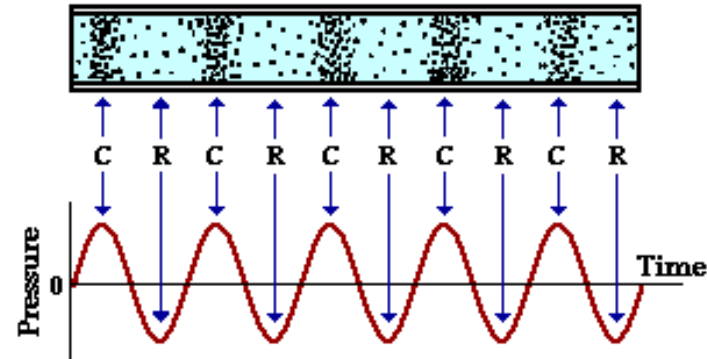
- Longitudinalni val titranja čestica zraka

- Izvor zvuka stvara titranje čestica (promjene tlaka)
- Titranje se širi brzinom  $v = \lambda f \approx 340 \text{ m/s}$  (ovisno o temperaturi)
- Lokalno titranje tlaka uzrokuje titranje struktura uha
- Pužnica ima više od 10000 „detektora“ osjetljivih na razne frekvencije; oni pretvaraju zvuk u žičane signale



- Vremenskoj domena: valni oblik

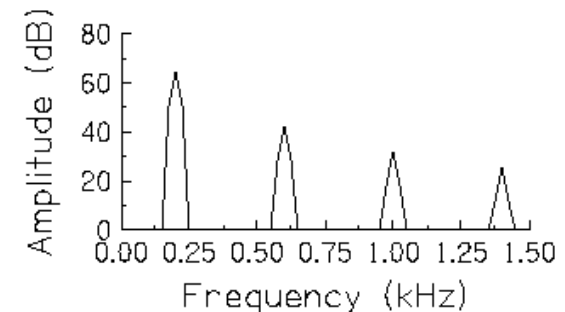
- amplituda
- faza
- frekvencija




- Frekvencijska domena: frekvencijski spektar

- harmonici (višečetnici osnovne frekvencije  $f$ ,  $2f$ ,  $3f$ , ..)
- za govor, formanti (rezonantne frekvencije govornog trakta), npr:

– a	F1 660	F2 1700	F3 2400 Hz
– e	F1 530	F2 1850	F3 2500 Hz
– u	F1 300	F2 870	F3 2250 Hz



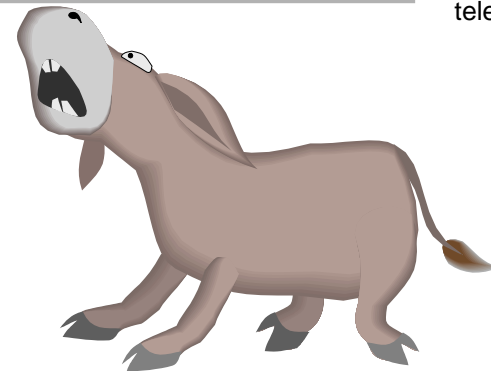
- Frekvencija ( $f$ ) [Hz] → **visina**:  $20 \log (f/f_r)$  [Mel]
  - **Jednake rastuće korake *frekvencije* primjećujemo kao sve manje i manje korake rasta *visine zvuka***
    - Čujno područje 20 Hz do 20 kHz
- Amplituda ( $A$ ) [Pa] → Intenzitet ( $I$ ) [ $\text{W}/\text{m}^2$ ] → **glasnoca**:  $20 \log (I/I_r)$  [dB] ( $I_r$  je granica čujnosti)
  - **Jednake rastuće korake *amplitude* (intenziteta) primjećujemo kao sve manje i manje korake rasta *jačine* odnosno, *glasnoće zvuka***
    - granica čujnosti 0 dB ( $1 \cdot 10^{-12} \text{ W}/\text{m}^2$ )
    - šapat 25 dB
    - uredska buka 50 dB
    - razgovor 60-65 dB
    - prosječni kućni stereo uređaj ~90 dB
    - granica boli 120 dB
    - ozljeda bubnjića 130dB



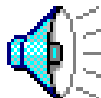
**dinamički raspon  
ljudskog sluha  
120 dB**

- Prirodni - analogni

- ljudski glas (govor, pjevanje)
- glasanje životinja
- glazbeni instrumenti
- prirodni zvukovi, npr. šum mora, vjetar, grmljavina,...



- Sintetički – obično digitalni



- oponašanje nekog od prirodnih zvukova, računalna glazba, ...
- Digitalni oblik je pogodan za pohranu, obradu i prijenos u računalnim sustavima
- Prirodni zvukovi se stoga digitaliziraju
- Svrha kodiranja digitalnog zvuka: što manji zapis bez osjetnog gubitka kvalitete

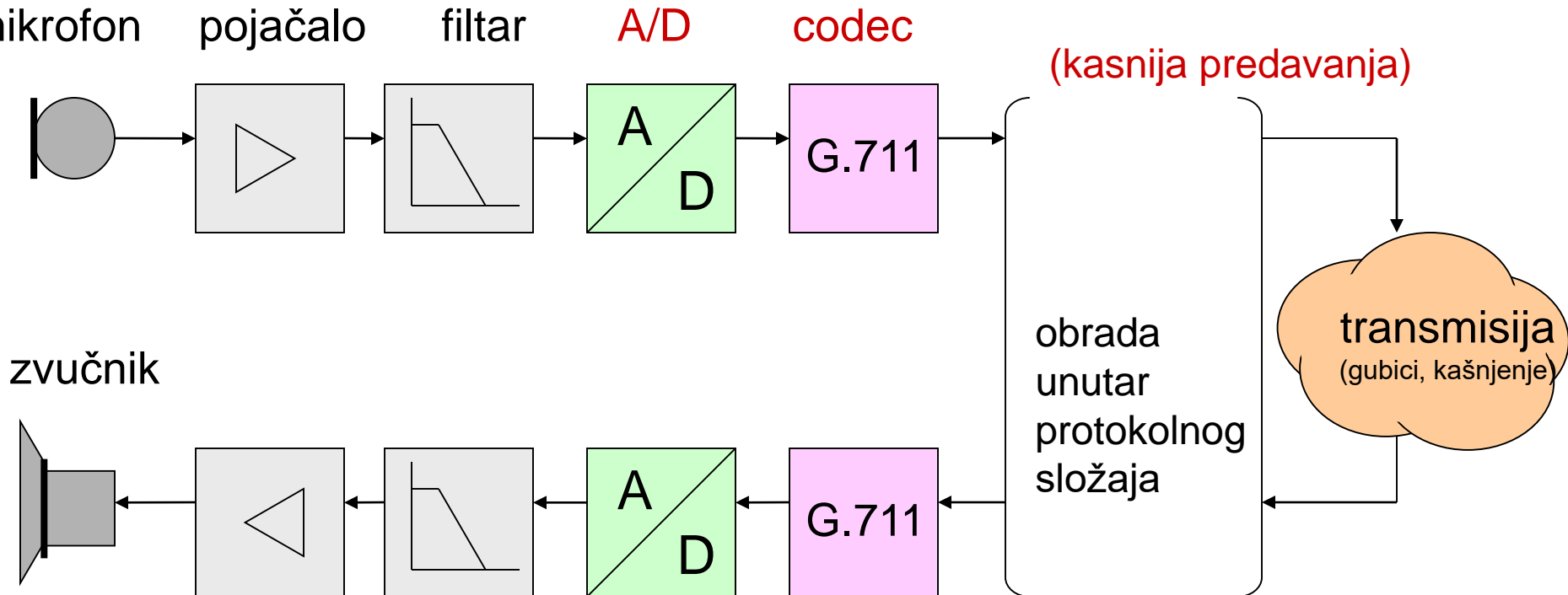
# Digitalni zvuk u umreženom računalnom sustavu

(ovo predavanje)

A/D

codec

(kasnija predavanja)



- **codec**, coder + decoder = uređaj koji kodira i dekodira, odn. komprimira i dekomprimira, audio i video

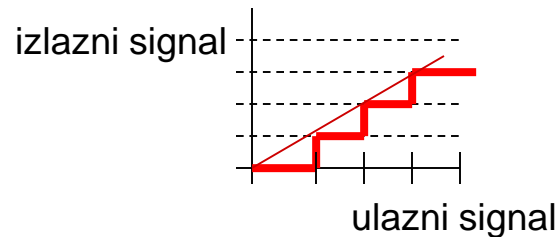
# Postupak digitalizacije zvuka



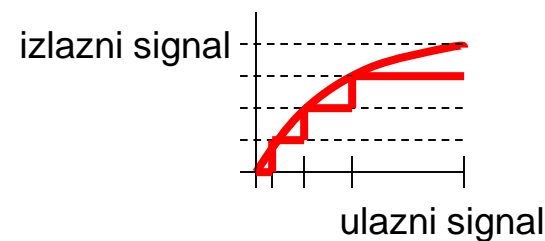
Zavod za  
telekomunikacije

- 1. uzorkovanje
- 2. kvantizacija
  - više ulaznih vrijednosti preslikava se na istu izlaznu vrijednost čime se gubi mogućnost točne rekonstrukcije
  - razlika stvarne i kvantizirane vrijednosti je kvantizacijska pogreška
  - uz broj bita  $k$  dobiva se  $2^k$  razina kvantizacije
  - općenito vrijedi: veći broj bita po uzorku → manje izobličenje

## *linearna* kvantizacija



## *nelinearna* kvantizacija



(npr. PCM,  
A- ili  $\mu$ -zakon),  
**Manja**  
**Kvantizacijska**  
**pogreška!**

kvantizacija može biti *fiksna* (zadane razine) ili *adaptivna*  
(adaptivno određene razine kvantizacije)

- 3. kodiranje



# Vrste koder s obzirom na namjenu



Zavod za  
telekomunikacije

- **uskopojasni** (telefonski kanal;  $B = 4$  kHz)
  - prikaz zvuka s visokom točnošću:  
 $f_u = 8$  kHz, 16 bita po uzorku (linearno)  $\rightarrow 128$  kbit/s
  - analogna telefonija: 300 Hz – 3400 Hz  
 $f_u = 8$  kHz, 8 bita po uzorku (nelinarno,  $A/\mu$ )  $\rightarrow 64$  kbit/s
- **širokopojasni** (AM radio, ISDN; videokonferencija;  $B = 7$  kHz)
- **zvuk visoke kvalitete** (FM radio, televizija;  $B = 15$  kHz)
  - kanal 30 Hz – 15 kHz  
 $f_u = 32$  kHz, 16 bita po uzorku
- **zvuk visoke vjernosti** (CD kvaliteta;  $B = 20$  kHz)
  - svaki kanal 20 Hz – 20 kHz  
 $f_u = 44.1$  kHz, 16 bita po uzorku

# Kriteriji usporedbe codec-a



Zavod za  
telekomunikacije

- brzina, izlaz, *bitrate* (bit/s)
- kvaliteta
  - **objektivna mjerila** (“klasične” metode, izobličenje signala i SNR, nisu dobra mjerila za ljudsku percepciju rekonstruiranog signala)
  - **subjektivna mjerila** (često važnija od objektivnih!)
- kašnjenje
  - algoritamsko kašnjenje u koderu na izvoru - koliko traje kodiranje
  - kod dekodiranja - koliko traje dekodiranje?
  - sinkronizacija s ostalim medijima u višemedijskoj aplikaciji
- otpornost na gubitke
  - posebno važno za prijenos preko mreže
- primjena na ostale zvukove koji nisu govor, npr. fax i modemske signale, te glazbu
- složenost (hw/sw)
- cijena izvedbe

# Subjektivna mjerila kvalitete



Zavod za  
telekomunikacije

- Opća ocjena kvalitete - mišljenje korisnika
  - Mean Opinion Score (MOS)
  - Degradation MOS (DMOS)

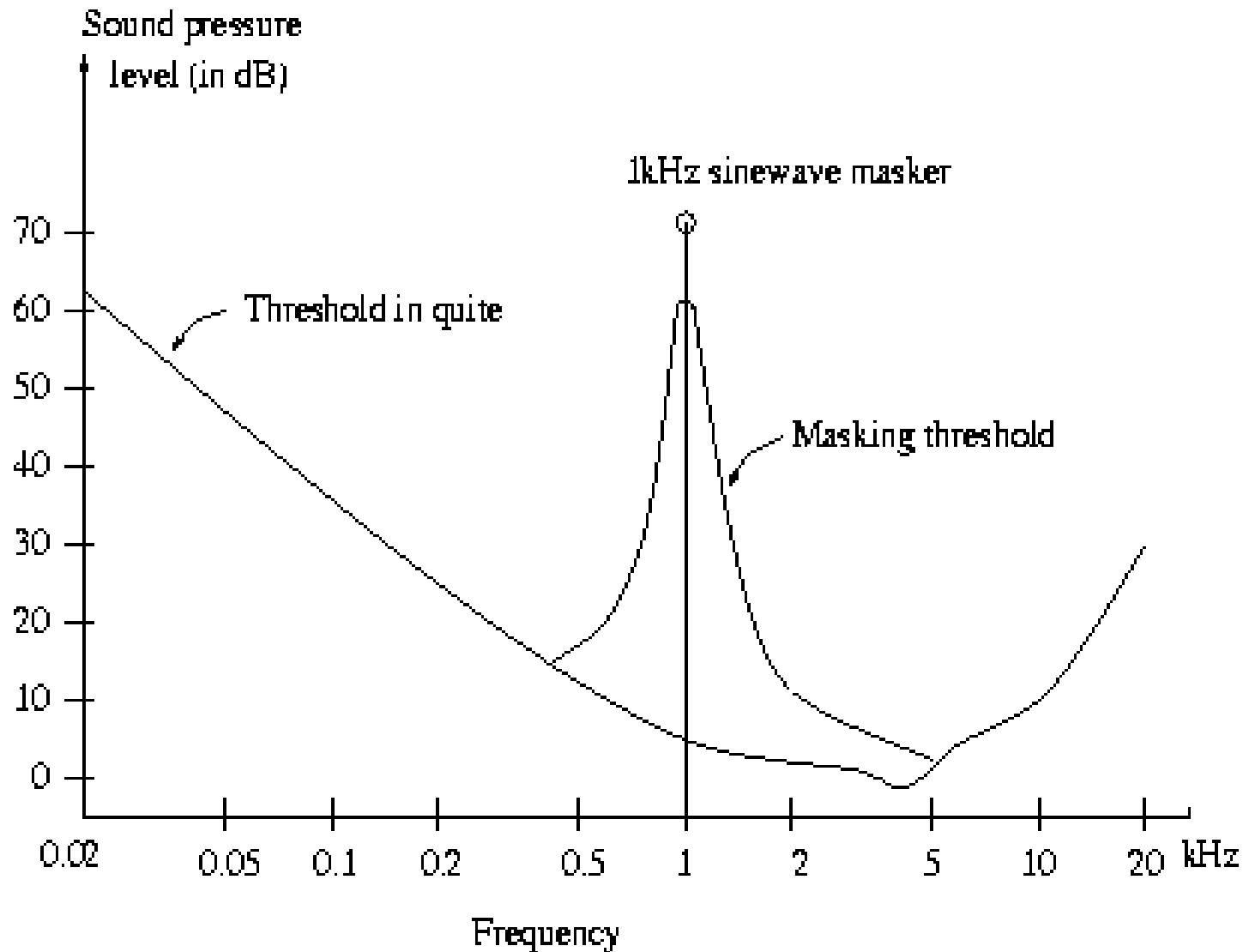
Ocjena	MOS	DMOS	Procjena napora razumijevanja
5	izvrsno	nečujno pogoršanje	bez napora
4	dobro (toll quality)	čujno pogoršanje, ali ne smeta	bez posebnog napora
3	prihvatljivo	primjetno pogoršanje, malo smeta	osrednji napor
2	slabo	podnošljivo, ali smeta	priličan napor
1	loše	izraženo pogoršanje, jako smeta	neprepoznatljivo bez izrazitog napora

- Ocjena razumljivosti govora
  - dijagnostički testovi s parovima riječi koje slično zvuče, 90% točnost smatra se “toll quality”
  - Dynamic Rhyme Test (DRT) – 96 parova engleskih riječi, npr. *dune/tune, chair/care, moon/noon*, .. slušačima se nudi popis i pita ih se koju riječ su čuli

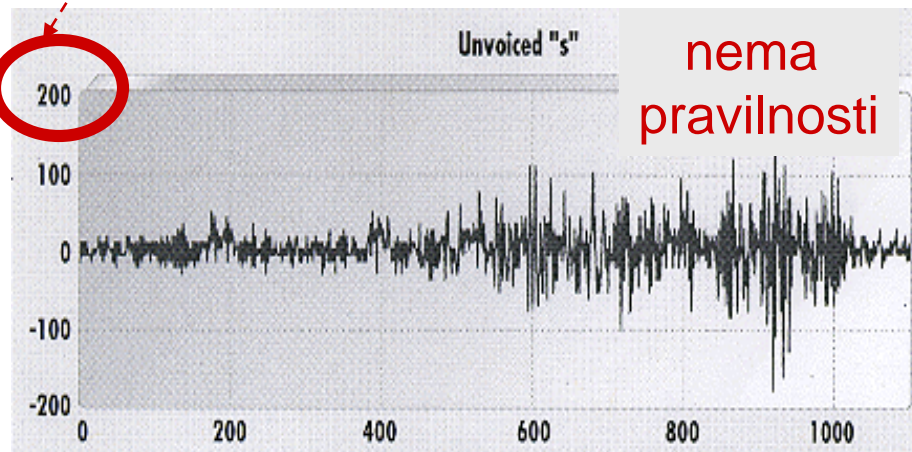
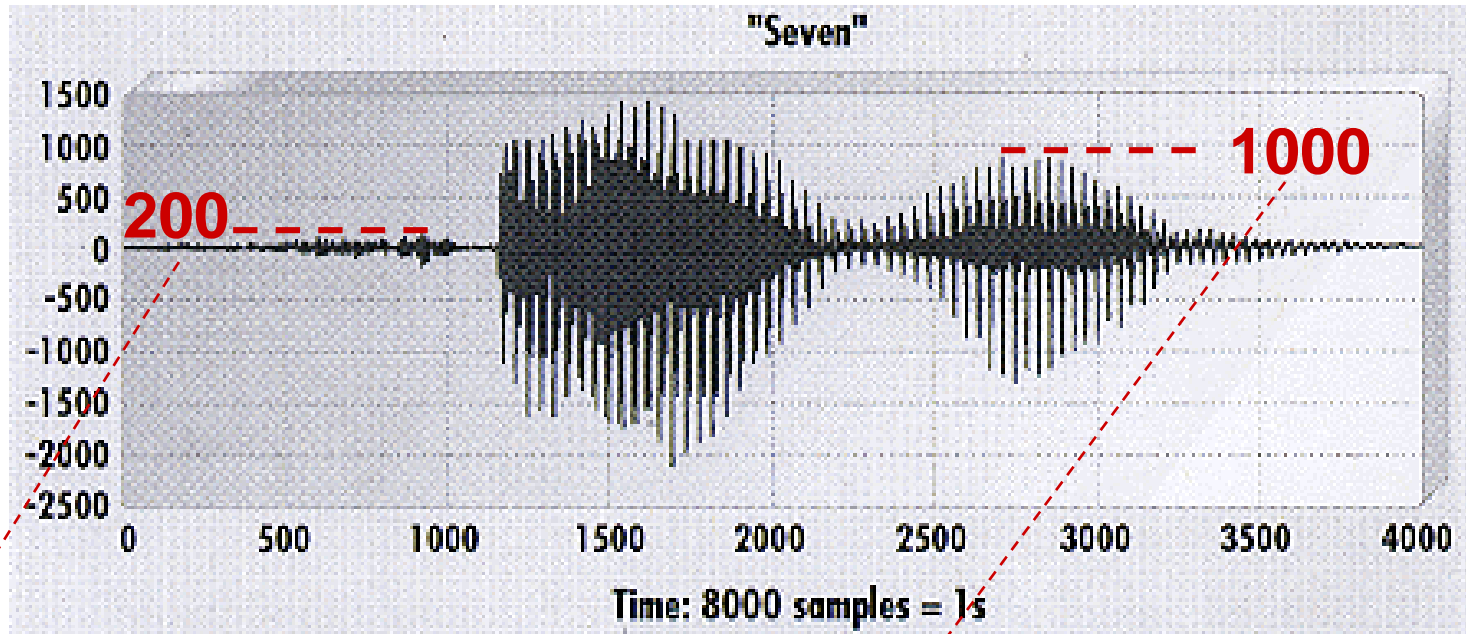
- novije objektivne metode temelje se na poznavanju ljudskog slušnog sustava – računaju izobličenje zvučnog signala s *percepcijskim* težinskim faktorima
  - ideja: izobličenja koje uho “više čuje” ima veći težinski faktor od onog manje primjetnog ili neprimjetnog
- Perceptual Evaluation of Speech Quality (**PESQ**)
  - ITU-T preporuka P.862
  - metoda procjene subjektivne kvalitete govornih kodeka
  - algorithm predviđa subjektivnu ocjenu kvalitete degradiranog uzorka govora
  - izlaz iz algoritma je procijenjena vrijednost MOS
- Perceptual Evaluation of Audio Quality (**PEAQ**)
  - ITU-R preporuka BS.1387
  - algoritam za procjenu kvalitete audia

- Izvor: ljudski govorni organi
  - Zrak iz pluća prolaskom kroz govorne organe (dušnik, grkljan, glasnice, šupljine usta i nosa) stvara glas
    - **zvučni glasovi** (zrak izaziva titranje glasnica)
    - **bezvučni glasovi** (nema titranja glasnica)
  - Frekv. raspon 60 Hz - 8 kHz, dinamički raspon 40 dB
- Prijamnik: ljudski slušni organi
  - Čujno područje 20Hz – 20 kHz, dinamički raspon 120 dB
  - Za razumljivost najvažnije 2 - 5 kHz (500-2000Hz)
  - Kvaliteta se ocjenjuje subjektivno
  - Poznata anomalija: efekt *maskiranja*: jedan zvučni signal prekriva drugi (ovisno o relativnima glasnoćama i frekvencijama)

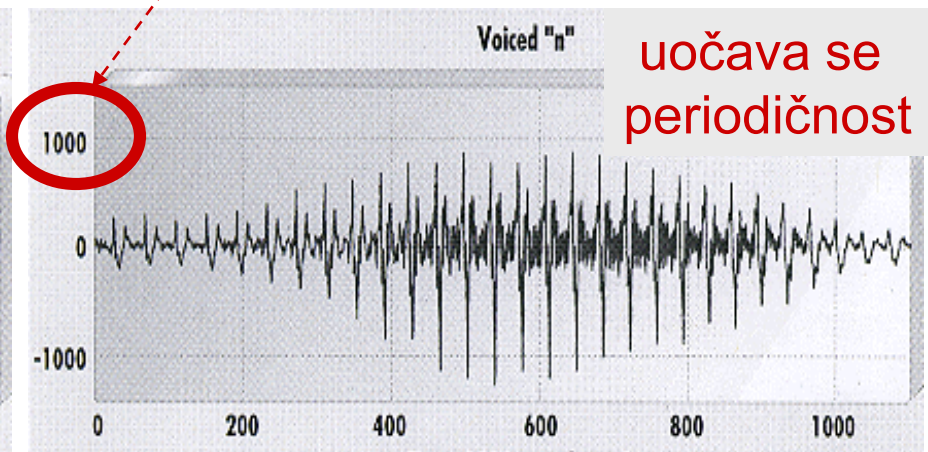
# Maskiranje zvuka



# Vremenska analiza govora



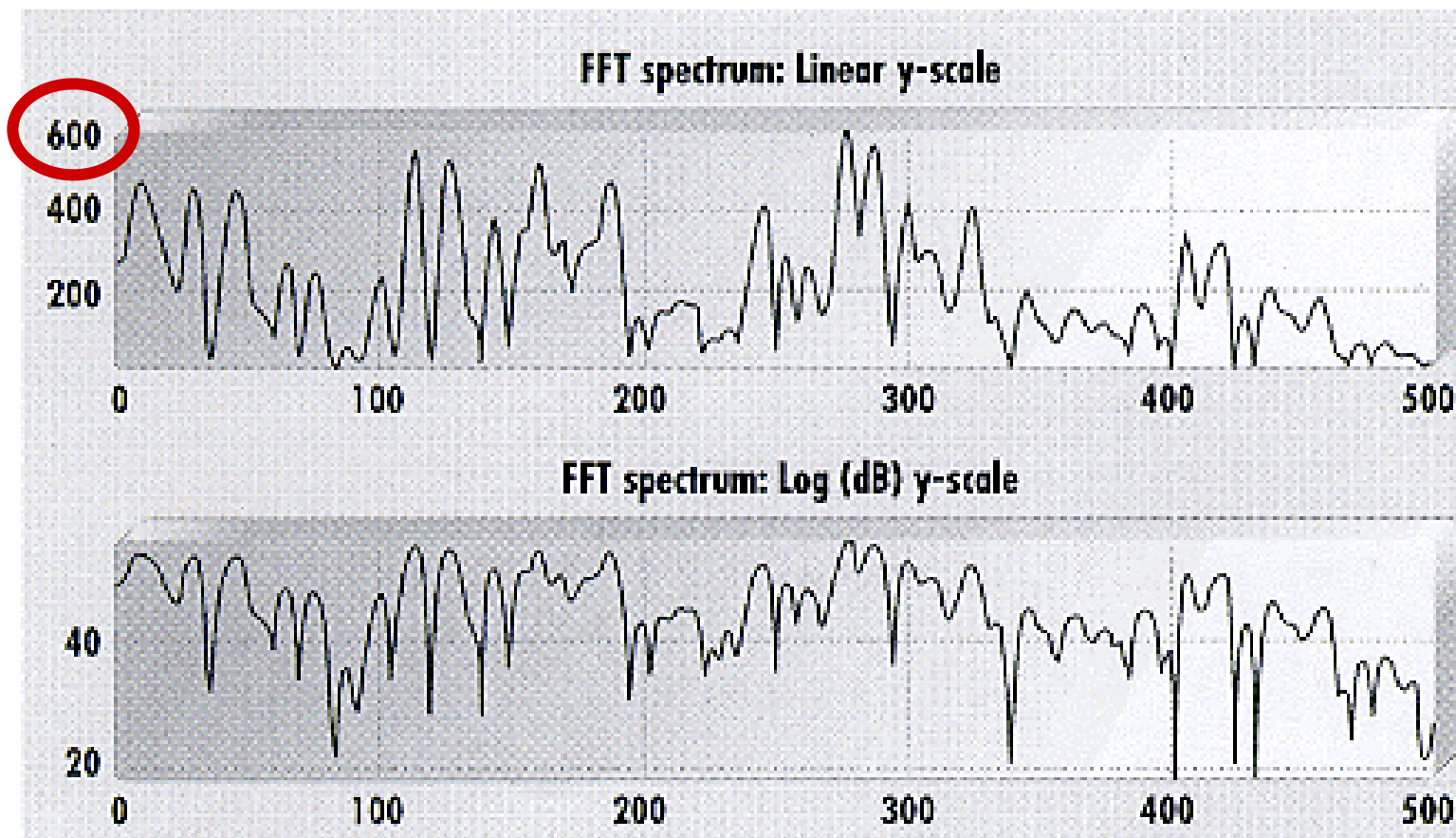
bezvučni suglasnik "s"



zvučni suglasnik "n"

# Frekvencijska analiza govora (1)

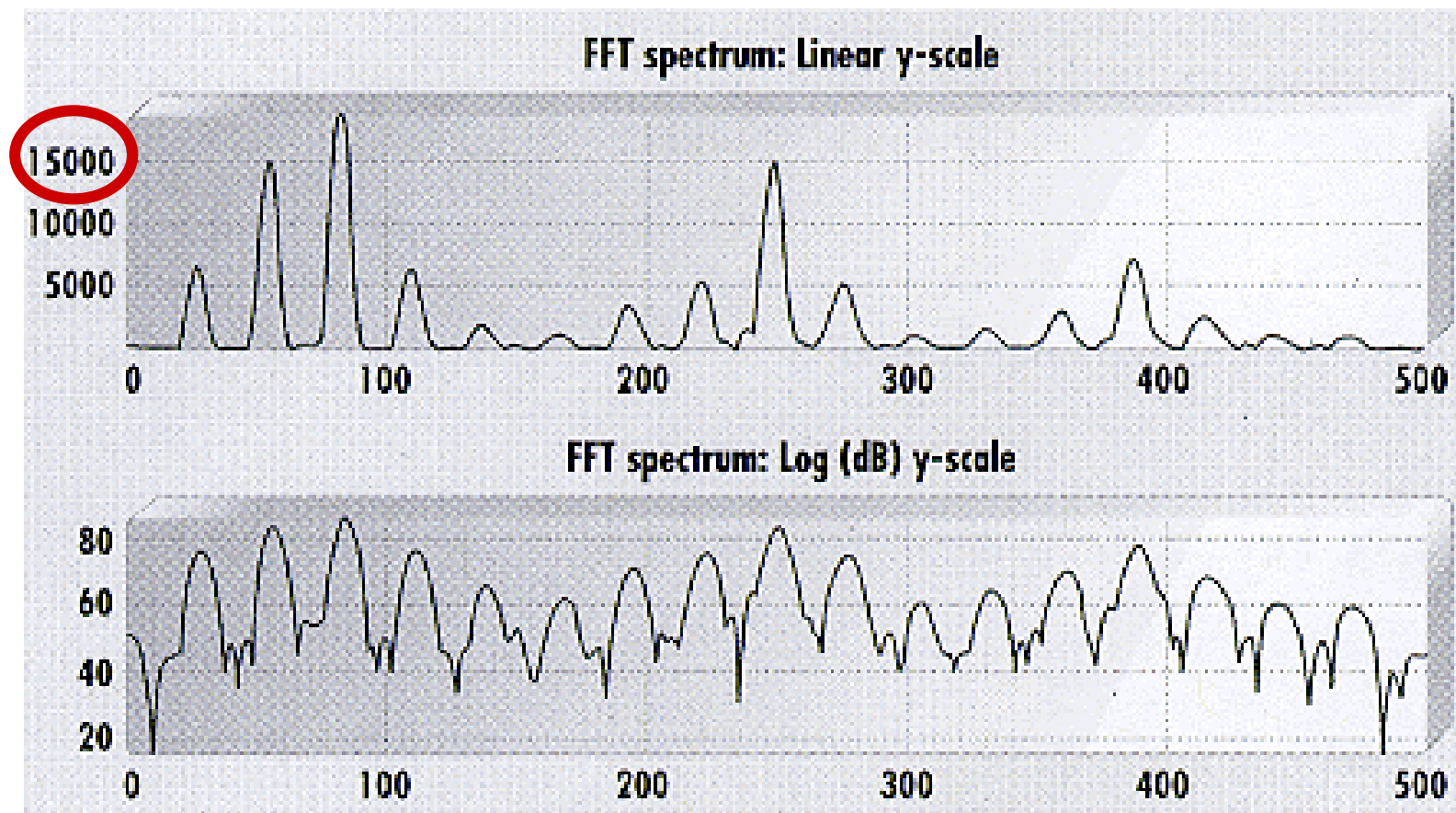
- Odsječak 30 ms **bezvučni suglasnik "s"**





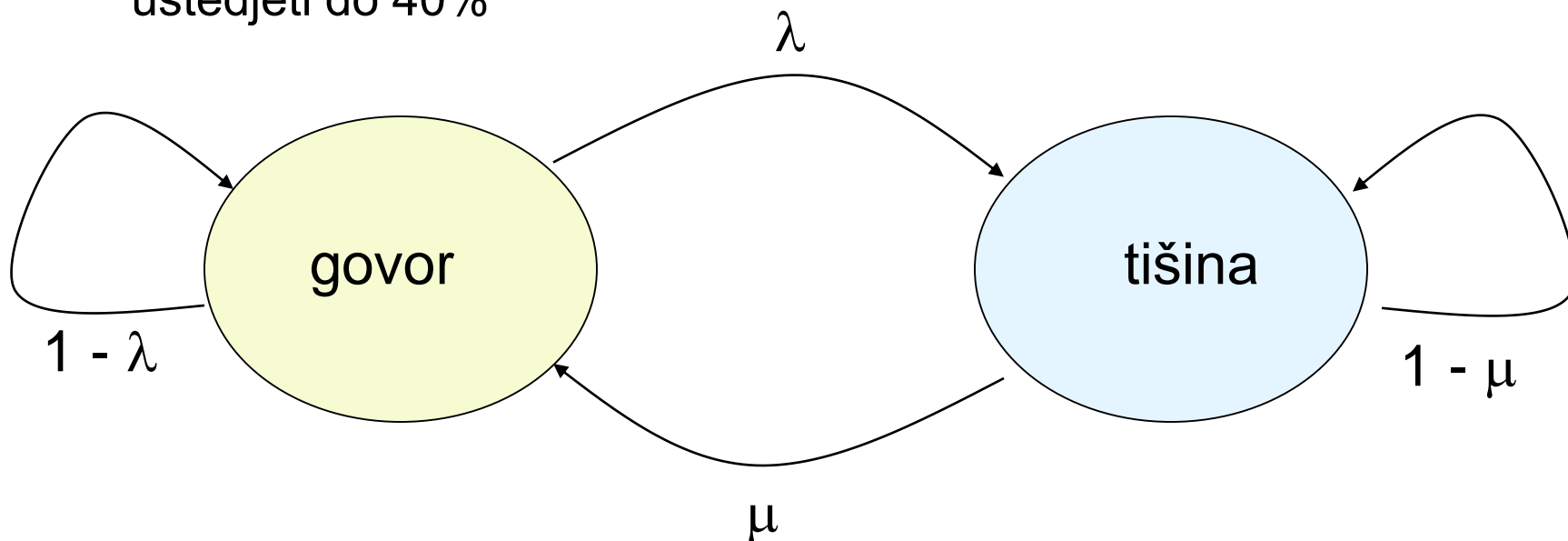
# Frekvencijska analiza govora (2)

- Odsječak 30 ms **zvučni suglasnik “n”**
- Uočljive rezonantne frekvencije - **formanti**



# Model govora u vremenu

- model govora u vremenu može se opisati “on-off” modelom
- intervali govora prosječno traju 800 ms - 1.2 s
- intervali tišine (između pojedinih glasova, riječi i rečenica) prosječno traju 1 – 1.6 s
  - ako se na izlaz koderu ne šalje ništa u intervalima tišine, može se uštedjeti do 40%

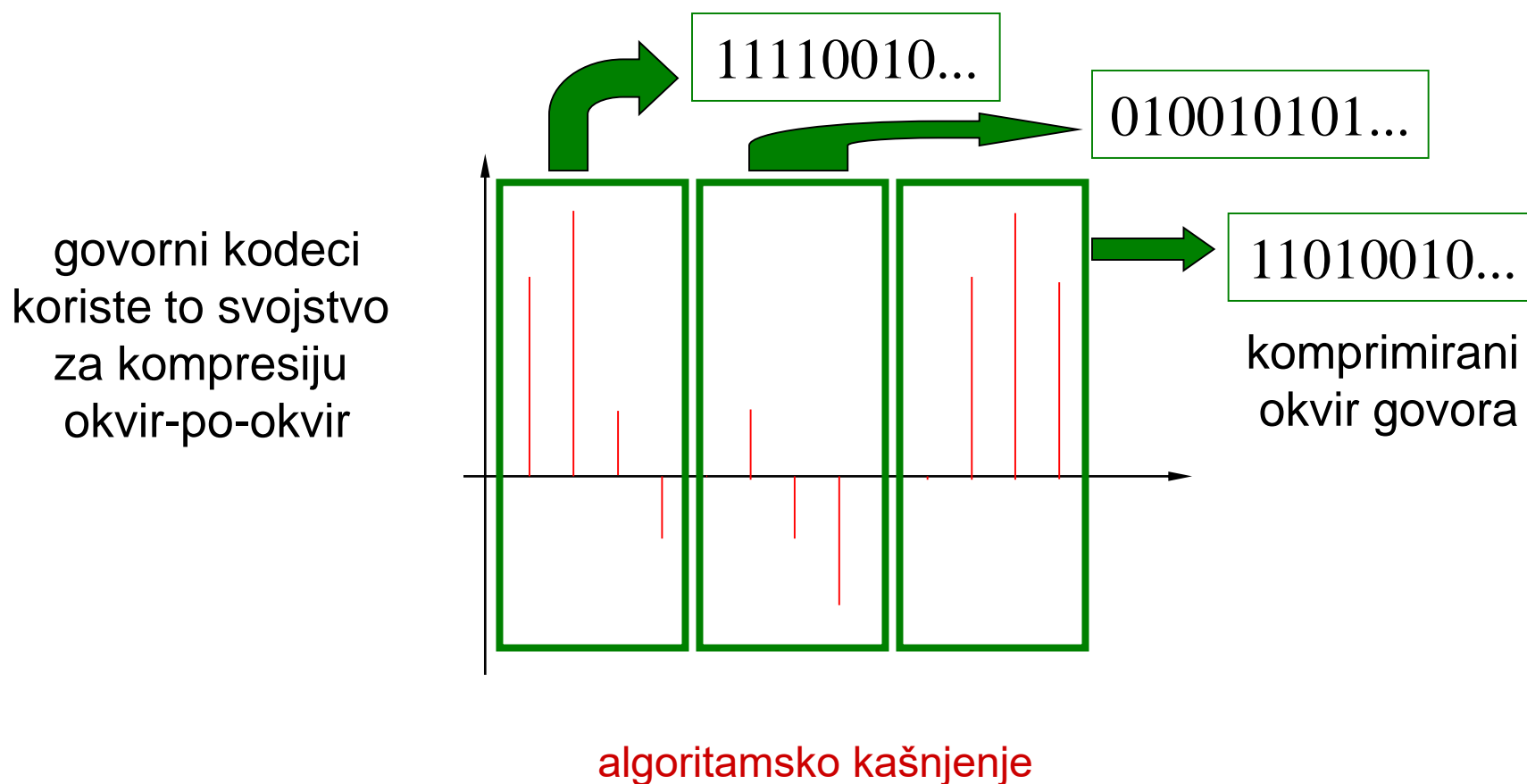


# Vremenska svojstva govornog signala



Zavod za  
telekomunikacije

- govorni signal je nestacionaran (mijenja se u vremenu), ali u manjim vremenskim odsječcima (**okvirima** od 20 - 30 ms) može se promatrati kao stacionaran



# Ideje za konstrukciju codeca



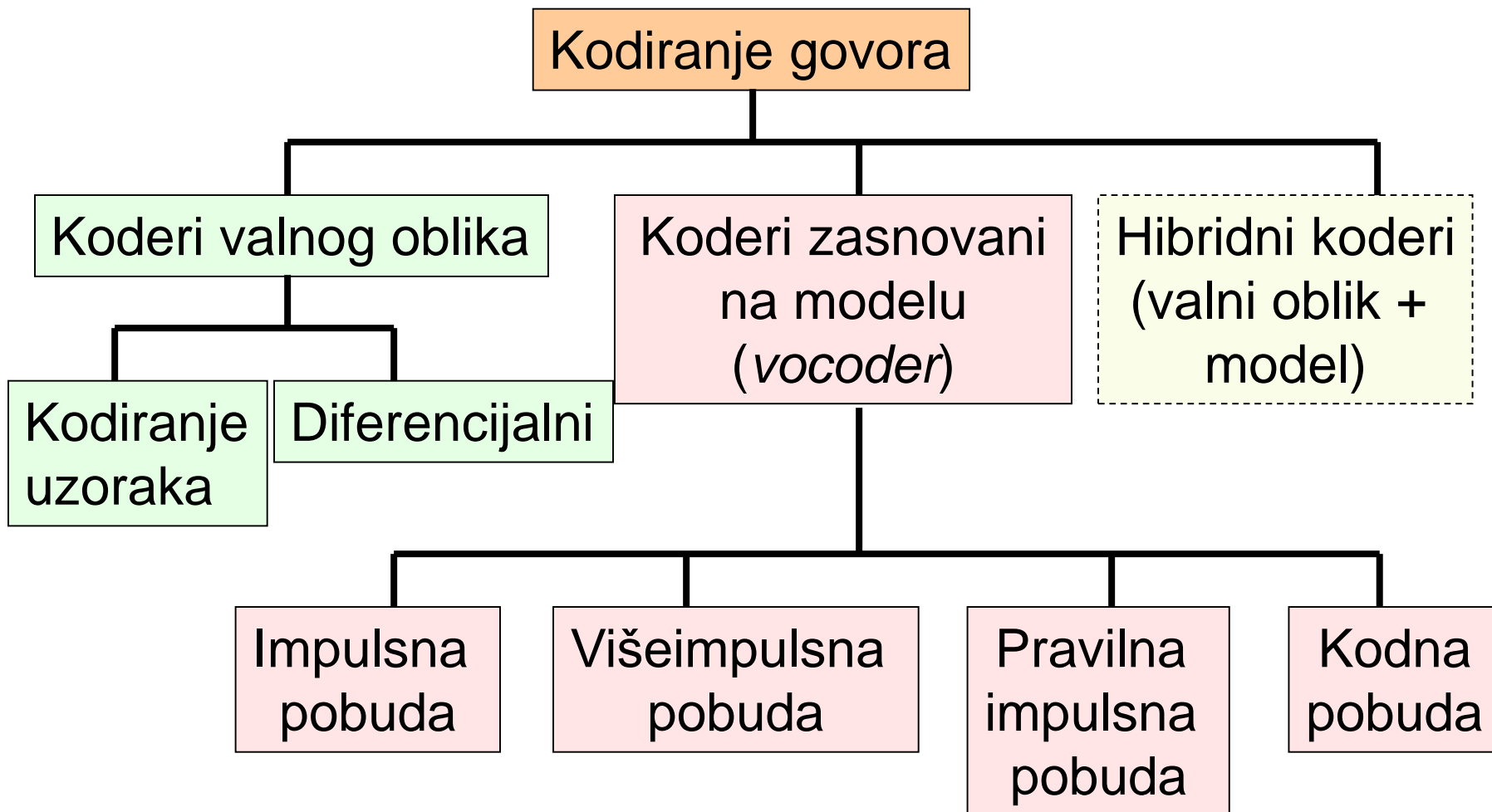
Zavod za  
telekomunikacije

- Svojstvo govornog signala je da ima veću vjerojatnost poprimanja manjih vrijednosti nego većih vrijednosti
  - Uniformna kvantizacija nije optimalna
  - “isplati se” točnije kodirati manje vrijednosti od većih - nelinearna kvantizacija daje bolju kvalitetu uz jednak broj bita po uzorku
- Postoji visoka korelacija između uzastopnih uzoraka i uzastopnih okvira
  - Uklanjanjem redundancije u signalu može se sažeti zapis
- Na temelju poznavanja svojstava govora, tj. fizioloških karakteristika govornog trakta, može se napraviti model
  - Parametri modela se računaju na temelju stvarnih uzoraka
  - Prenose se samo parametri, a govor se rekonstruira (sintetizira) na temelju modela
  - U najnovijim koderima (npr. MPEG-4 SA) ova ideja proširuje se i na druge zvukove

# Koderi govora

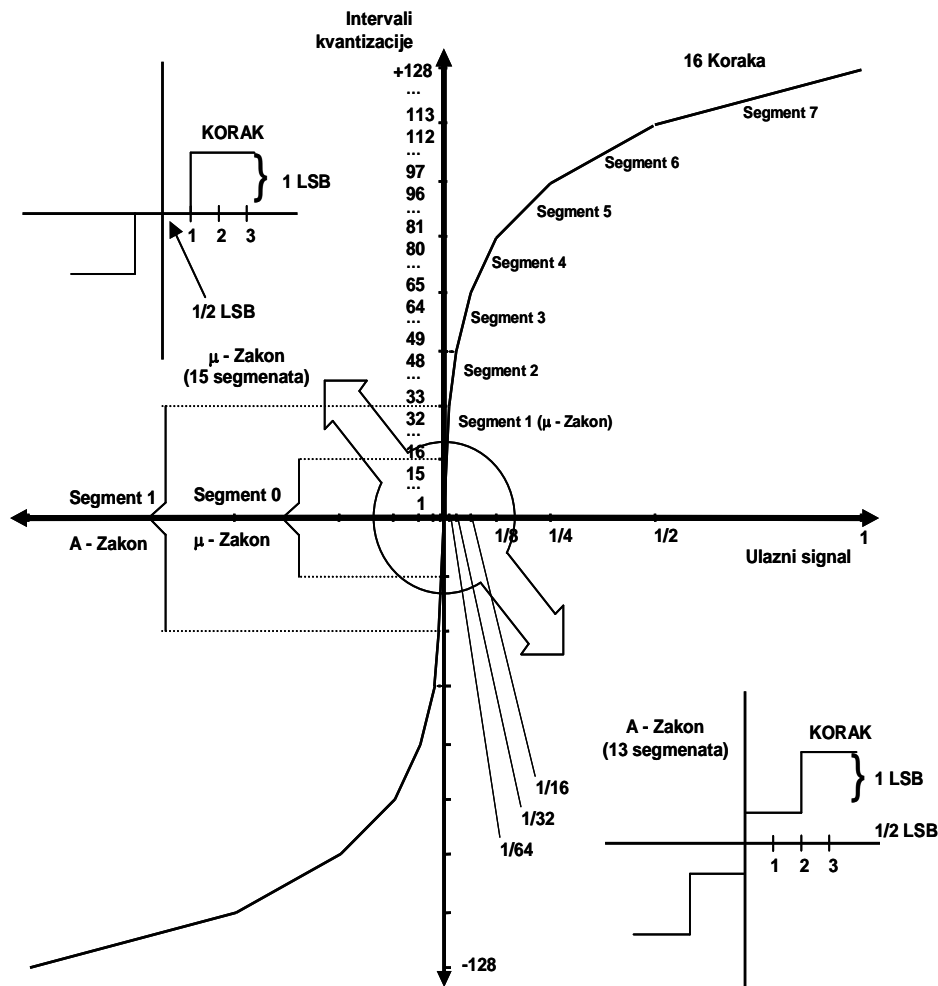


Zavod za  
telekomunikacije

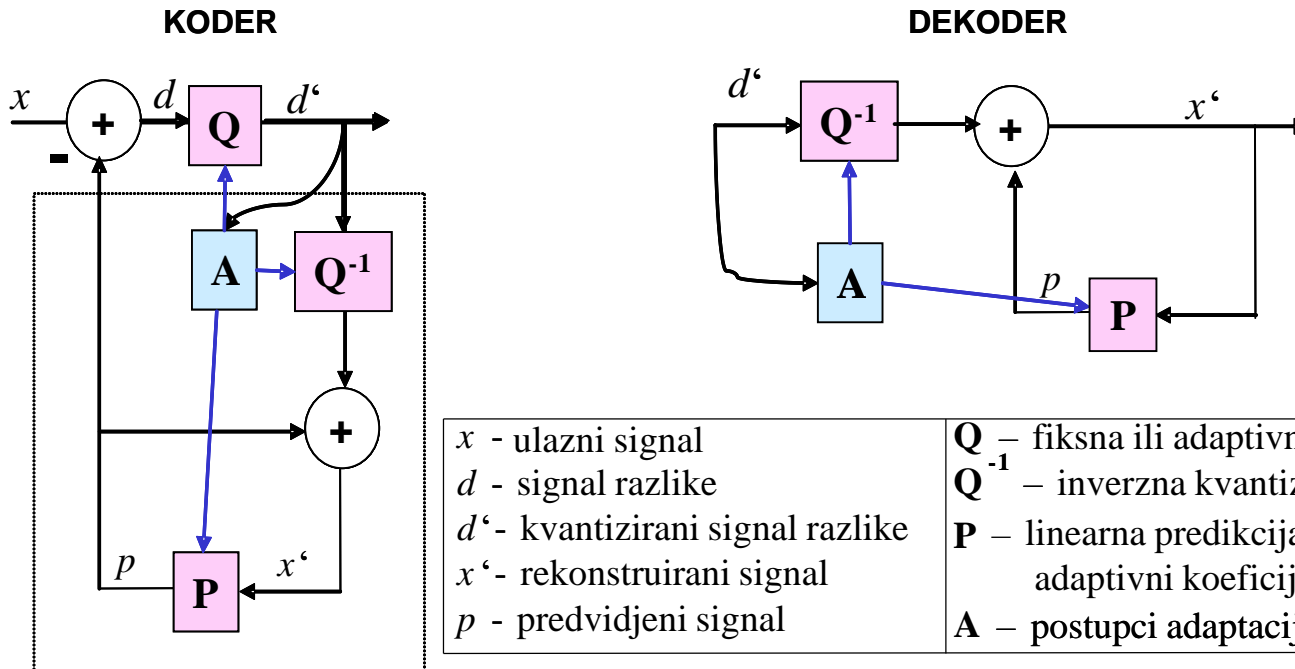


- Veće brzine, dobra kvaliteta, razvijeni za fiksnu i (kasnije dorađeni) za mobilnu telefoniju
- Pulsno-kodna modulacija (**PCM**)
  - preporuka *ITU-T G.711 Pulse Code Modulation for voice frequencies (PCM)*
- Adaptivni diferencijalni PCM (**ADPCM**)
  - preporuka *ITU-T G.726 Adaptive Differential Pulse Code Modulation (ADPCM)*; sadrži zastarjelu preporuku G.721 (originalna norma)
  - proširenje je *ITU-T G.727 5-, 4-, 3- and 2 bits per sample embedded Adaptive Differential Pulse Code Modulation (ADPCM)*

- Uzorkovanje na 8 kHz, nelinearna kvantizacija po logaritamskoj karakteristici prema A-zakonu (Europa) ili  $\mu$ -zakonu (SAD, Japan)
- Prednosti:
  - jednostavan
  - visoka kvaliteta (MOS 4.3)
  - malo kašnjenje (1 uzorak)
- Nedostaci:
  - 64 kbit/s nije malo
  - nema mehanizme za kontrolu i ispravljanje pogrešaka (nije dobar kandidat za internetsku telefoniju)
- Primjena: već desetljećima u uporabi u fiksnoj telefonskoj mreži



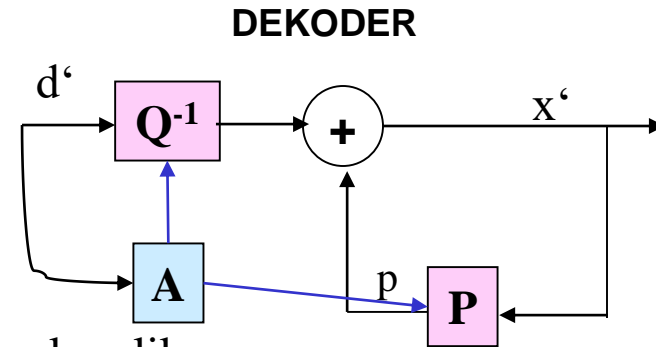
# Princip diferencijalnog koder (ADPCM)



$$p_n = a_1 x'_{n-1} + a_2 x'_{n-2} + \dots + a_k x'_{n-k}$$

U koderu je sadržan dekoder, te se računa razlika između signala kojeg bi dekoder predvidio i stvarnog signala; ova razlika se kvantizira i šalje dekoderu.





x - ulazni signal  
d - signal razlike

d'- kvantizirani signal razlike

$x'$ - rekonstruirani signal

p - predvidjeni signal

## Q – fiksna ili adaptivna kvantizacija

## Q<sup>-1</sup> – inverzna kvantizacija

**P** – linearna predikcija (fiksni ili adaptivni koeficijenti)

## A – postupci adaptacije

$$p_n = a_1 x'_{n-1} + a_2 x'_{n-2} + \dots + a_k x'_{n-k}$$

U koderu je sadržan dekodeer, te se računa razlika između signala kojeg bi dekodeer predvidio i stvarnog signala; ova razlika se kvantizira i šalje dekodeeru.

# ADPCM - svojstva

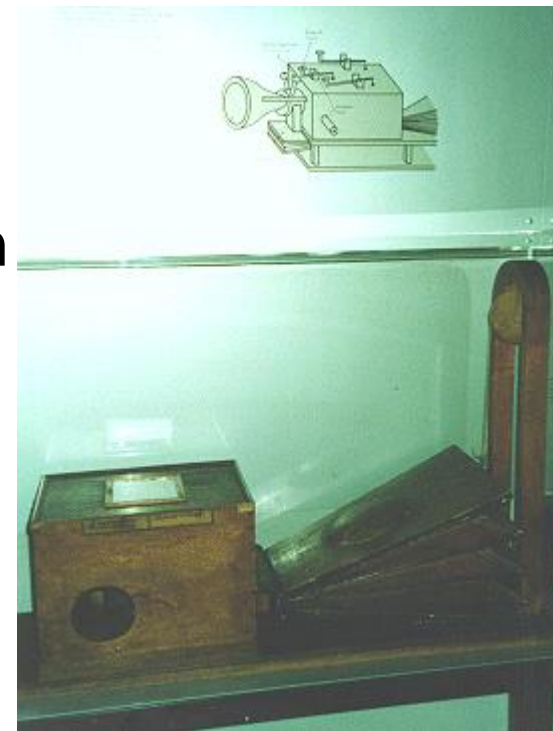


Zavod za  
telekomunikacije

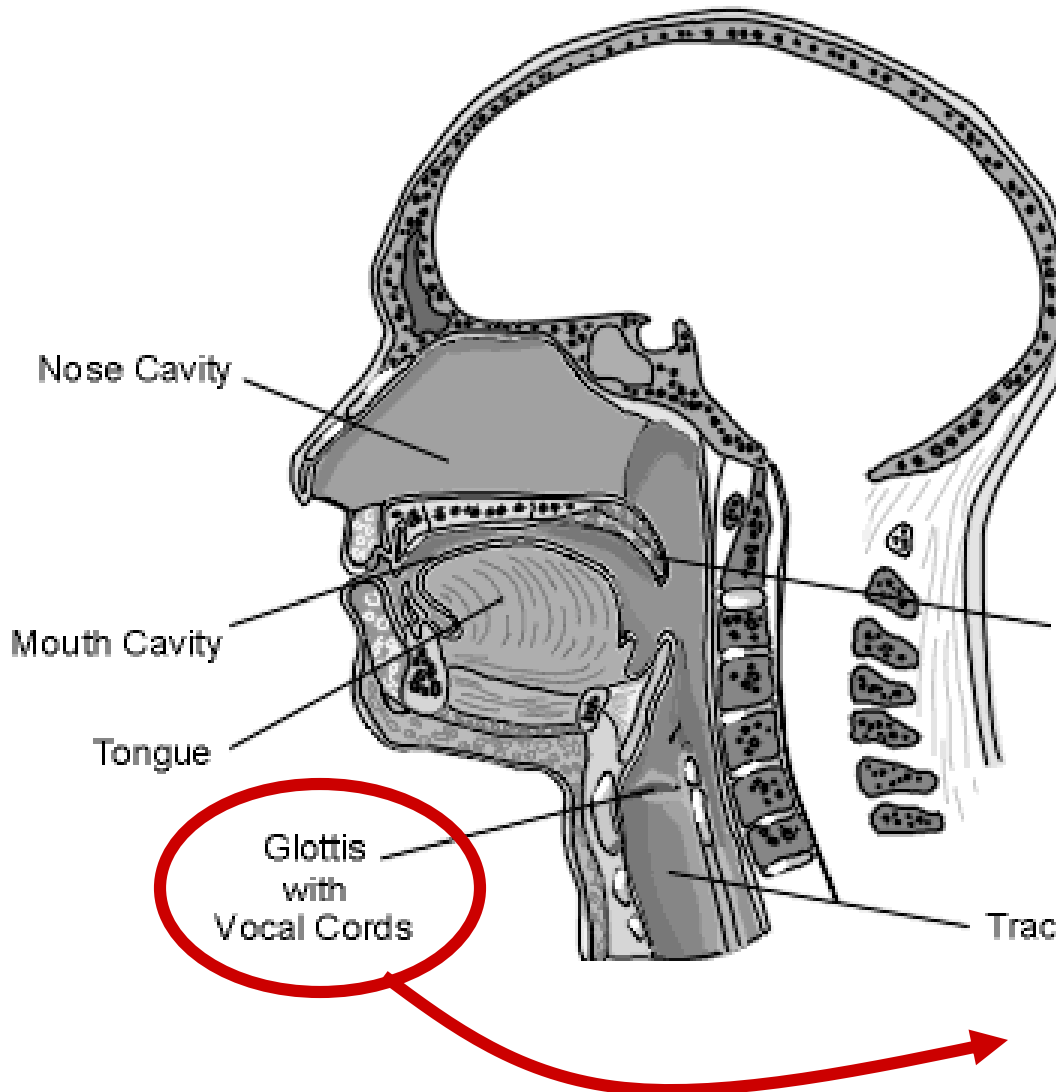
- osim tipične brzine od 32 kbit/s, ovisno o broju bita za kodiranje greške, norma specificira i brzine 40 kbit/s (5 bita), 24 kbit/s (3 bita), 16 kbit/s (2 bita)
- prednosti:
  - nema algoritamskog kašnjenja
  - prenosi i modemske i fax signale bez degradacije
- nedostaci:
  - visoka brzina (postoje bolja rješenja na manjim brzinama), osjetljiv na gubitke
- primjena: kućni bežični telefon - norma DECT (Digital European Cordless Telephony)
- primjena i u širokopojasnom koderu; preporuka *ITU-T G.722 Wideband (7 kHz) audio codec by Subband ADPCM (SB-ADPCM)*
  - 64 (56, 48) kbit/s
  - primjena: ISDN aplikacije, telekonferencija

- Ideja: koder i dekoder imaju isti (parametrizirani) model govornog trakta
  - Parametri modela se računaju za okvire uzoraka govora
  - Dekoderu se prenose parametri modela (a ne uzorci govora) te se govor sintetizira na odredištu
  - Princip analize/sinteze
- postižu se vrlo male brzine
- prvi koderi, npr. LPC-10, su bili lošije kvalitete, razvijeni za sustave ograničene namjene, npr. robotika, sigurna telefonija
- noviji koderi, npr. CELP na malim brzinama postižu dobru kvalitetu, ali su računski složeniji

- Kempelen Farkas / Wolfgang von Kempelen (1734-1804)
  - „govorni stroj“ – govorni trakt simuliran pomoću cijevi
  - Prvi zabilježeni pokušaj umjetne proizvodnje govora
- Homer Dudley, Bell Labs, 1939
  - „*Channel vocoder*“ (voice coder)
  - Model govornog trakta s filterima
  - Prethodnik današnjih kodera govora, postavlja osnovne principe



# Ljudski govorni organi

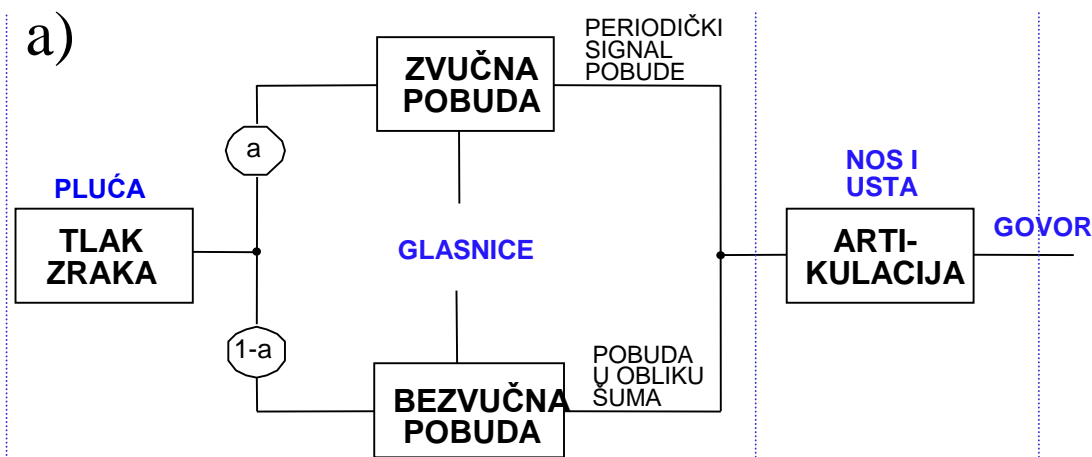


# Model proizvodnje govora



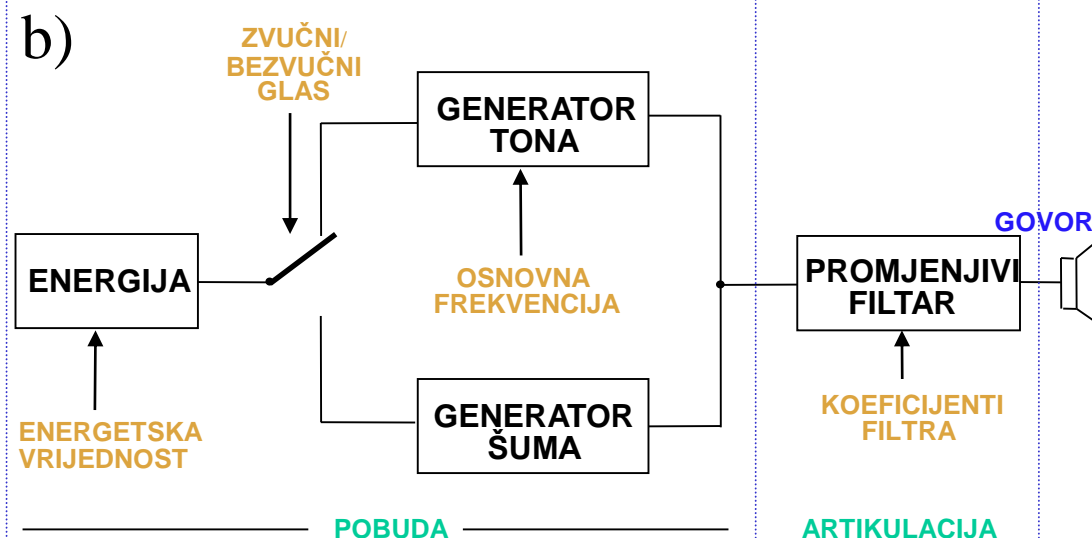
Zavod za  
telekomunikacije

a) blok dijagram  
ljudskih  
govornih organa



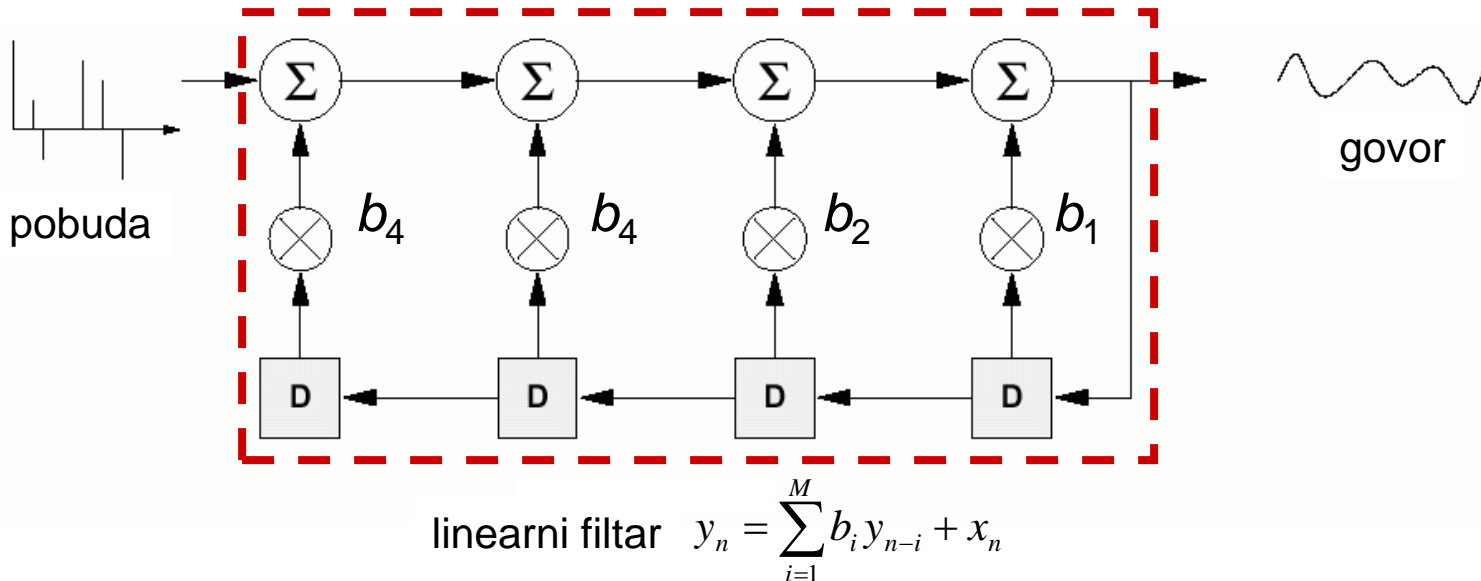
b) blok dijagram  
dekodera  
zasnovanog  
na modelu

(LPC - Linear  
Predictive Coder)



# Model govornog trakta

- Govorni trakt se modelira linearnim filtrom s nizom koeficijenata
- Signal se može prikazati kao izlaz linearnog filtra uz zadanu *pobudu*

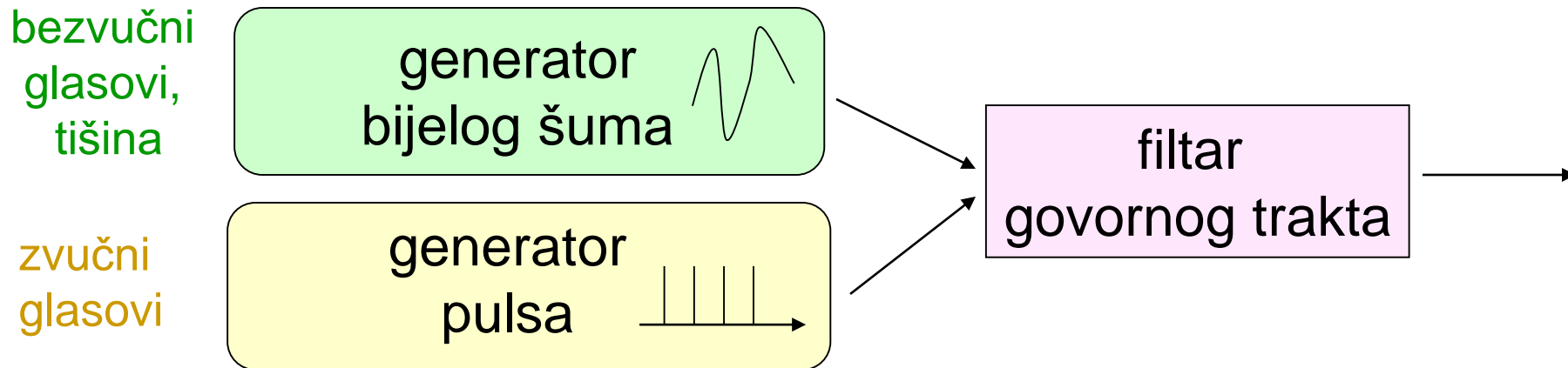


- Koder izračunava parametre filtra i pobude i šalje ih dekoderu

# Linear Predictive Coding (LPC)



Zavod za  
telekomunikacije



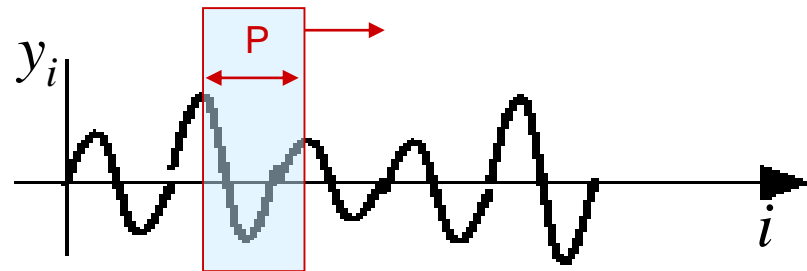
- Pobuda: periodički impulsni signal
- Parametri modela:
  - Frekvencija pobude 6 bita
  - Jačina pobude 5 bita
  - Zvučni/bezvučni glas 1 bit
  - Koeficijenti filtra 42 bita (10 koeficijenata)
- Npr, **LPC-10** na 2.4 kbit/s
  - Frekvencija uzorkovanja: 8kHz
  - Duljina okvira: 180 uzoraka = 22.5 ms



# LPC - računanje parametara modela

- Frekvencija pobude
  - Average Magnitude Difference Function (AMDF)

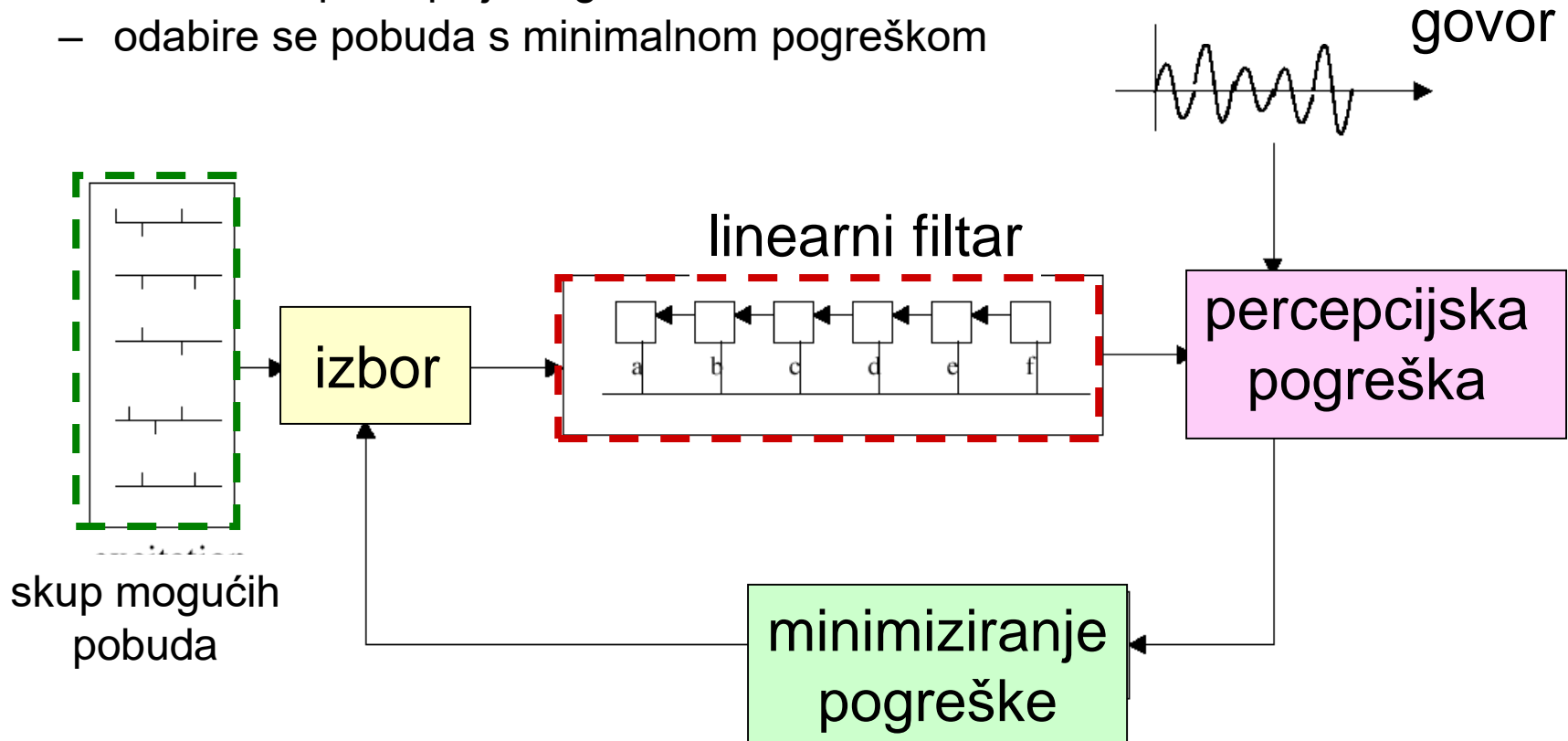
$$AMDF(P) = \frac{1}{N} \sum_{i=k_0+1}^{k_0+N} |y_i - y_{i-P}|$$



- AMDF izračunava prosjek razlika signala u zvučnom okviru udaljenih za neki period P
- Zvučni/bezvučni glas
  - AMDF nema jasnih minimuma za bezvučne glasove
  - Bezvučni glasovi: manja amplituda signala
- Jačina pobude
  - korijen srednje vrijednosti kvadrata signala
- Koeficijenti filtra
  - Metoda najmanjeg kvadrata pogreške

# Code Excited Linear Prediction (CELP)

- Zajednički rječnik kodova (code-book) u koderu i dekoderu
- Dekoderu se šalje indeks (kôd) pobude
- Analiza-sintezom služi za određivanje pobude
  - pretražuje se skup mogućih pobuda i za svaku provodi sinteza
  - računa se percepcijska greška
  - odabire se pobuda s minimalnom pogreškom



# Primjeri koderu zasnovanih na modelu (1)



Zavod za  
telekomunikacije

- preporuka *ITU-T G.728 Low Delay CELP (LD-CELP)*
  - 16 kbit/s, MOS 4, algoritamsko kašnjenje samo 0,625 ms
- preporuka *ITU-T G.729 Conjugate Structure Algebraic CELP (CS-ACELP)*
  - 8 kbit/s, MOS 4, kašnjenje 15 ms
- preporuka *ITU-T G.723.1 Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s*
  - 5.3 i 6.3 kbit/s, MOS 3.8
  - pogodan za internetsku telefoniju

# Primjeri koda zasnova na modelu (2)

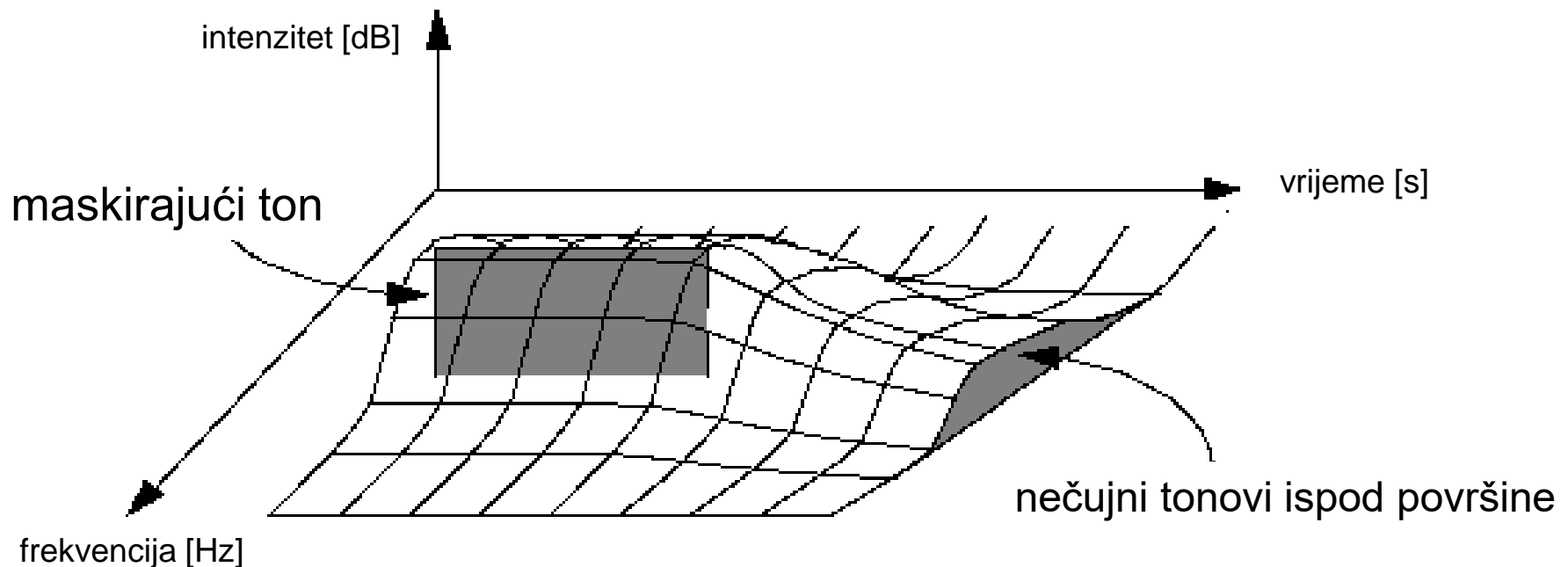


Zavod za  
telekomunikacije

- ETSI GSM 06.10: Full Rate codec na 13 kbit/s
  - u uporabi u većini GSM 900 and PCS 1800 mreža
  - Regular Pulse Excitation LPC with Long Term Prediction (RPE-LTP) koder
- ETSI GSM 06.60: GSM Enhanced Full Rate na 12.2 kbit/s
  - Algebraic Code-Excited Linear Prediction (ACELP)
  - osnova i za sjevernoamerički TDMA IS-136, kao i za ITU-T G.729
- ETSI GSM 06.20: GSM Half Rate na 5.6 kbit/s
  - Vector-Sum Excited Linear Prediction (VSELP)

# Koderi zvuka u frekvencijskoj domeni

- Koderi zvuka u frekvencijskoj domeni nisu ograničeni na govor; imaju dobra svojstva za bilo kakve zvukove npr. glazbu
  - Koriste pod-pojasno kodiranje (podjela na frekvencijske pod-pojase prije kodiranja)
- Koriste efekt maskiranja
  - uz jaki signal na nekoj frekvenciji uho ne može čuti slabiji signal na bliskoj frekvenciji
  - osim u zadanom trenutku, maskiranje ima utjecaj i u vremenu (prije/poslije)

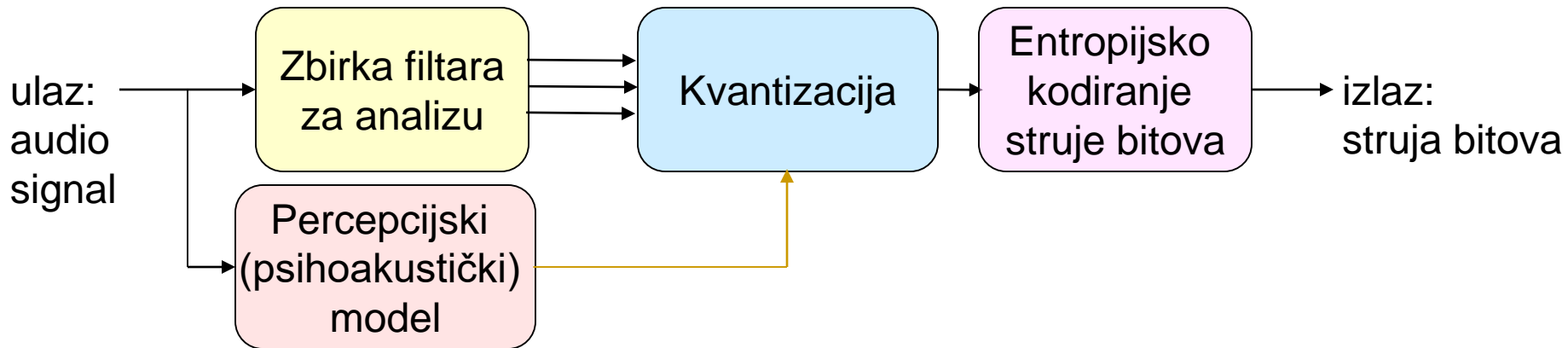


# Osnovni percepcijski koder

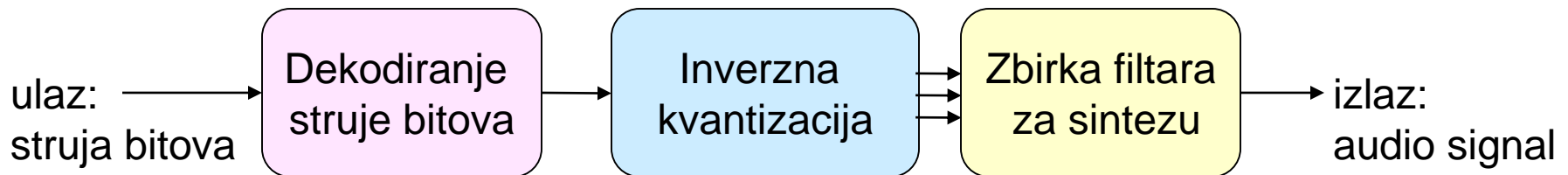


Zavod za  
telekomunikacije

## Shema koder



## Shema dekodera

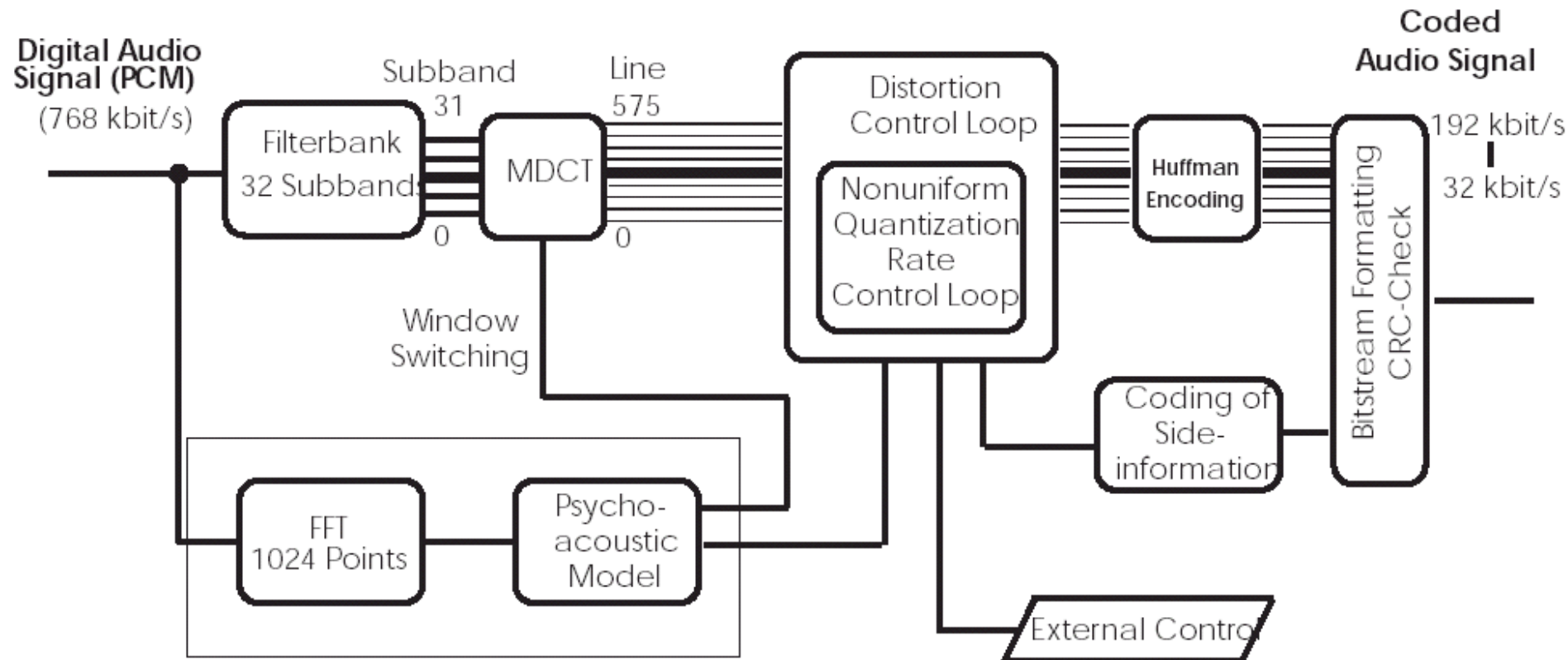


- MPEG - Moving Picture Expert Group ([ISO/IEC JTC1/SC29/WG11](#))
- MPEG-1
  - Dva audio kanala
  - $f_u = 44.1$  kHz (isto kao CD), 32 kHz, 48 kHz (isto kao DAT)
  - Brzine od 8-16 kbit/s do 320 kbit/s
  - MPEG Audio Layer I, II, III: razine kodiranja rastuće složenosti
  - Koristi se percepcijsko kodiranje
- MPEG-2
  - Isti osnovni koder kao MPEG-1
  - Pet audio kanala + niskofrekventni kanal
  - Uz MPEG-1 još i  $f_u = 16$  kHz, 22.05 kHz, 24 kHz
  - AAC – Advanced Audio Codec, dodan kasnije
- MPEG-4
  - AAC, dva koderi za govor, strukturirani audio, ...

# MPEG Audio Layer 3 – MP3 koder



Zavod za  
telekomunikacije



- Literatura: K. Brandenburg “MP3 and AAC Explained”, AES 17th international conference on High Quality Audio Coding, Florence, Italy, September 1999.



# Pregled koder govora i zvuka

Kodiranje	kb/s	MOS	uporaba
LPC-10	2.4	2.3	robotika, sigurna telefonija
G.711	64.0	4.5	telefonija (A, u-law)
G.722	64.0		7kHz codec (subband)
G.723.1	5.3/6.3	3.8	videotelefon (ostatak kanala za video)
G.726	16-40	ovisi	niska kompleksnost (ADPCM)
G.726	32	4.1	niska kompleksnost (ADPCM)
G.728	16.0	4.0	malo kašnjenje
G.729	8.0	4.0	mobilna telefonija
GSM	13.0	3.5	GSM (mobilna telefonija, Europa)
GSM EFR	12.2	4.0	GSM 2.5G
GSM HR	5.6	3.5	GSM 2.5G
IS 54/136	7.95	3.5	TDMA (Sj. Amerika mobile (stari std.))
IS 641	7.4	4.0	TDMA (Sj. Amerika mobile (novi std.))
MPEG L3	56-128.0	N/A	CD stereo
DVI	32.0	toll-quality	(Intel, Microsoft)
16 bit/44.1 kHz	1411		compact disc

# LPC demo

---



Zavod za  
telekomunikacije

- Pokazati na predavanju