UDK 681.32:534.6-07
Originalni znanstveni rad

Primljeno: 2. 1. 1988.

Milan STAMENKOVIĆ
VVTŠ KoV JNA, Zagreb

DIGITALNO PREDSTAVLJANJE I ANALIZA GOVORA U VREMENSKOJ DOMENI

SAŽETAK

U radu su prikazane tradicionalne tehnike digitalizacije (PCM, DPC, DM, LPC) i analize govornog signala u vremenu. Osim teorijskih načela, razmotreni su rezultati vlastitih istraživanja o razumljivosti digitaliziranog govora s aspekta amplitudne kvantizacije. Većina tehnika digitalizacije ilustrirana je praktičnim primjerima.

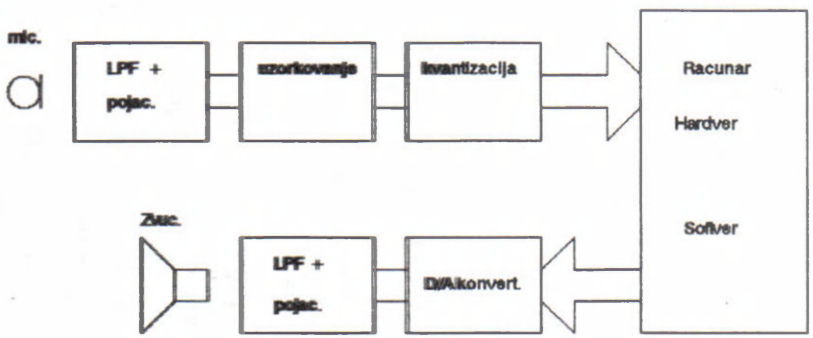
1. Uvod

Iako su prva digitalna računala nastala u razdoblju 1945—1947, masovno se primjenjuju za obradu govornog signala tek od 70-ih godina. Razlog je tome bila njihova velika cijena i mala procesna moć te nedovoljna memorija. Međutim, razvojem tehnologije VLSI (Very Large Scale Integration — vrlo visoki stupanj integracije) integriranih krugova, u posljednjih desetak godina pojavila su se takva računala koja su po svojim dimenzijama više puta manja od standardnog sonografa, a po performansama daleko iznad računarskih sistema koji su 60-ih godina zajedno s pratećom opremom zauzimali čitave sale. Posljednih desetak godina, kada je nastupila era PC (Personal Computer — personalno računalo) analogni sistemi sigurno ustupaju mjesto novoj generaciji opreme. Potpuno se prešlo na digitalnu obradu, uključujući zamjenu analognog magnetofona digitalnim. Iako se digitalna obrada u svojim fundamentima oslanja na analognu, prisutna su načela vezana za pojave amplitudne i vremenske diskretizacije govornog signala i izrazito numerički aspekt obrade.

Osim pregleda najčešće primjenjivanih tehnika digitalizacije i analize u vremenu, ovaj rad sadrži rezultate vlastitih istraživanja. U drugom dijelu prikazani su načini digitalizacije i reprezentacije govora (PCM, DM, ADM, LPC, ADPCM), a u trećem tehnike digitalne analize (kratkovremenska energija, autokorelacija, prolasci kroz nulu itd.). O rezultatima eksperimenata raspravljano je nakon iznošenja načela pojedine tehnike.

2. Digitalizacija govornog signala u vremenskoj domeni

Principijelna shema sistema za digitalnu obradu signala prikazana je na slici 1.



sl. 1 Principijelna shema računarske digitalne obrade govornog signala

Prvi je korak u digitalnoj obradi digitalizacija analognog signala. Digitalizacija govornog signala uključuje diskretizaciju u vremenu, tj. uzorkovanje i diskretizaciju po amplitudi, tj. kvantizaciju. Ako se uzorkuje u ekvidistantnim vremenskim razmacima tada govorimo o uniformnom uzorkovanju, odnosno o neuniformnom ako to nije. Isto tako, kvantizacija može biti linearna kada se amplituda signala linearno skalira odnosno nelinearna, kada skaliranje nije linearno (npr. logaritamsko). Praktične potrebe nametnule su pojavu tehnika digitalnog kodiranja s minimalnom redundancijom, čime se postiže manje računsko i memorijsko opterećenje računala, odnosno kanala. Na slici 2 (Proakis83) prikazane su tradicionalne tehnike digitalizacije svrstane prema zauzeću pamćenja, odnosno prema broju bita potrebnih za digitalizaciju 1 sekunde razumljivog govora približno iste kvalitete.

Tehnika digitalizacije	Kvantizacija vrsta	Br. bita	br. bita za 1 sec. govora
PCM	linearna	8—12	64000—96000
Log PCM	logaritamska	7—8	56000—64000
DPCM	logaritamska	4—6	32000—48000
ADPCM	adaptivna	3—4	24000—32000
DM	binarna	1	34000—64000
ADM	adaptivna bin.	1	16000—32000
LPC			800—7200

sl. 2 Tehnike digitalizacije govornog signala u vremenu

U daljem tekstu sa x(t) bit će označen analogni signal, sa x_n uzorkovani x(t), s $\tilde{x}_n$ kvantovani uzorkovani signal, a sa $\hat{\tilde{x}}_n$ pretpostavljena (predikcijska) uzorkovana vrijednost x_n.

2.1. Kvantizacija amplitude uzoraka (PCM, log PCM)

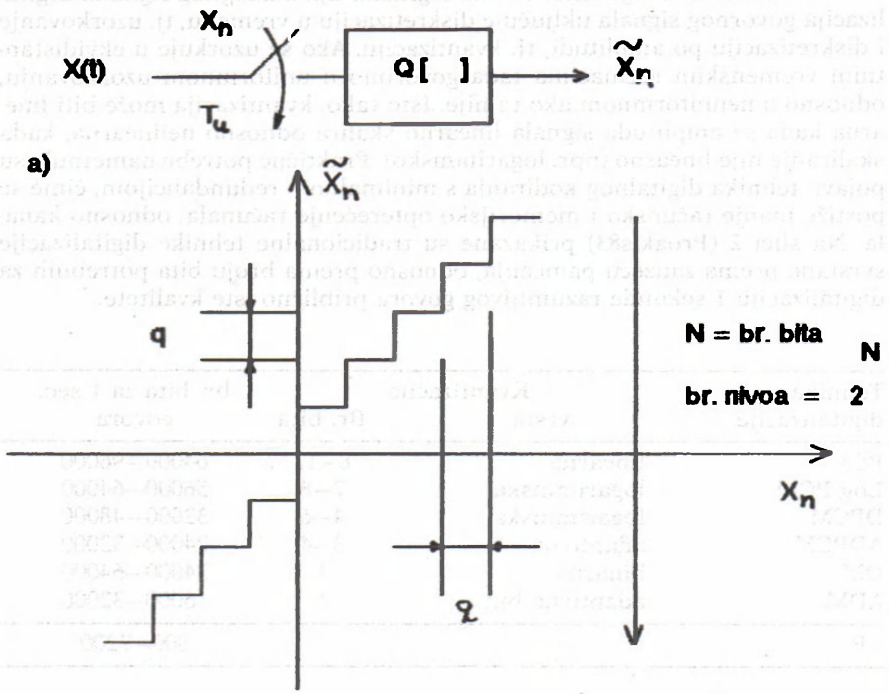
Osnovni je način digitalizacije uniformno uzorkovanje i linearno kvantiranje poznat kao PCM (Pulse Code Modulation — impulsno kodna modulacija) i shematski je predstavljen na slici 3. Govorni se signal najprije frekvencijski ograničava na F_g, uzorkuje X_n i linearno diskretizira na 2^N nivoa (X_n).

Postavljaju se pitanja:

a) koliki mora biti period uzorkovanja T_u, odnosno frekvencija

$$F_u = \frac{1}{T_u};$$

b) koliko je bita dovoljno za kvantizaciju?



sl. 3 Uzorkovanje a) i linearna kvantizacija b)

Teorem o uzorkovanju govori da period T_u mora biti dva puta manji od perioda maksimalne frekvencije frekvencijskog spektra signala. Iako se frekvencijski spektar glasova proteže i više od 10 KHz (npr. za sibilante), razumljivost govora odredena je s prva tri formanta koji su uglavnom locirani ispod 3 KHz. To znači da je dovoljno za F odabrati:

$$F_u = \frac{1}{T_u} > F_u = 6\ \text{KHz} \qquad \ldots(1)$$

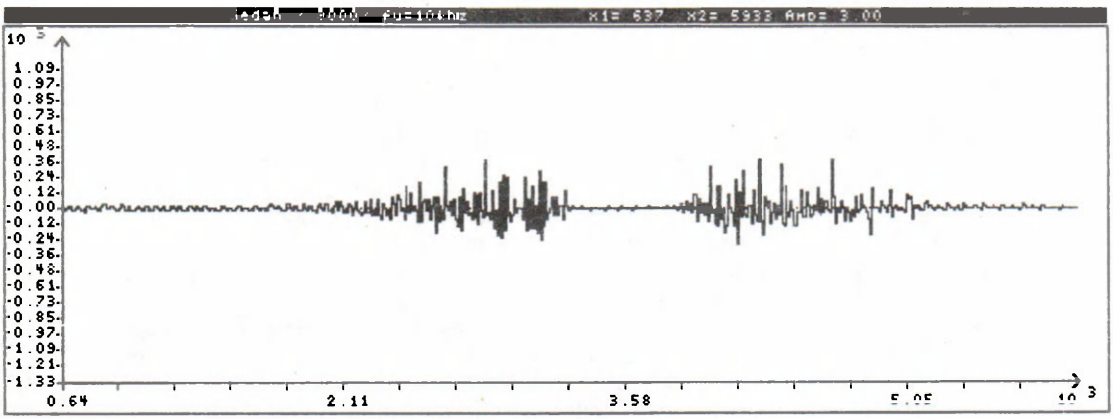
Budući da realni filtri bitno zaostaju po svojim osobinama od proračunatih, minimalnu frekvenciju uzorkovanja treba povisiti na 7—8 KHz. Broj bita kvantizacije odreduje dinamiku signala. Ako je kvantizacija linearna, tada svaki bit nosi 6 dB dinamike što se može lako pokazati:

$$D = 10 \cdot \log \frac{P_1}{P_2} = 10 \cdot \text{Log} \left(\frac{2^{NH}}{2^{N}} \right)^2 = 10 \cdot \log 4 = 6\ \text{dB} \qquad \ldots(2)$$

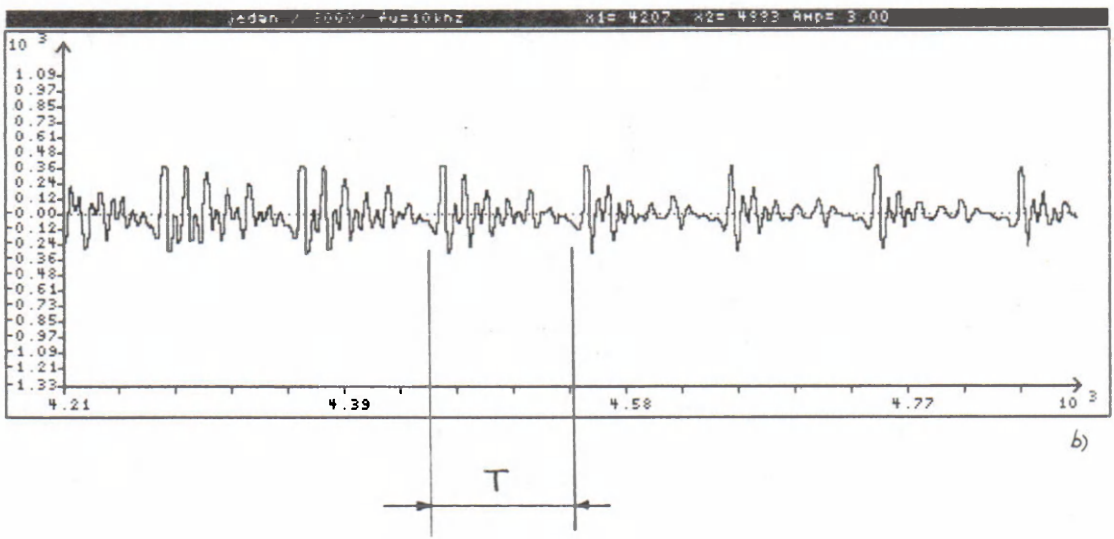
Istraživanja su pokazala (Rabiner78) da je donja granica dinamike za razumljivost govora oko 36 dB, odnosno da je neophodan broj bita:

$$N = 36/6 = 6 \text{ bita}$$

Greška kvantizacije nivoa sa 6 bita jest 1/2 = 1.6% što ne zadovoljava u fonetskim istraživanjima. Budući da za preciznija mjerenja greška kvantizacije mora biti manja od 0.5%, minimalan je broj bita 8 (dinamika 48 dB, greška kvantizacije 0.39%). U literaturi se ipak nalazi (Witten82, Schafer79) 11 bi-



a)



b)

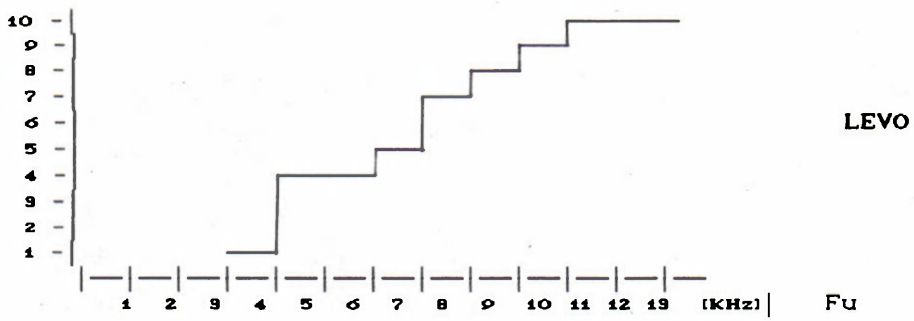
sl. 4 a) riječ »jedan« b) kvaziperiod vokala »a« (PCM)

ta za minimalni broj bita kvantizacije da bi se osigurala dovoljna rezolucija i za tiše glasovne intervale, tj. da ne bi došlo do odsijecanja signala.

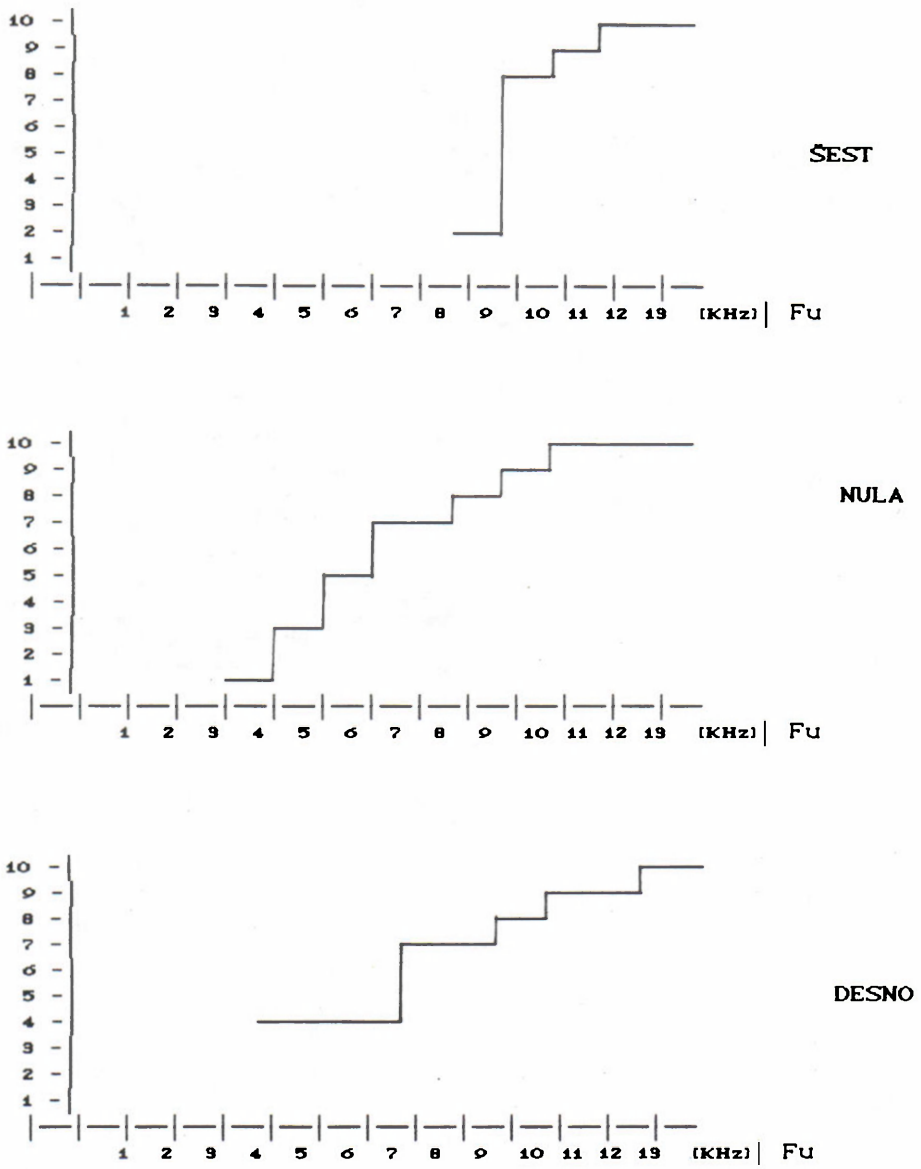
Na slici 4 prikazana je digitalizirana riječ »jedan«, a u izdvojenom dijelu vide se prigušeni kvaziperiodi vokala »a«. Budući da je $F_u = 10KHz$, a perioda ima približno 80 vremenskih uzoraka, to znači da je u ovom slučaju osnovni period tona vokala »a« Ta—80/10000 = 0.008 s, odnosno F_a = 120 Hz. Gornja frekvencija niskopropusnog filtra bila je 4 KHz.

Razumljivost digitaliziranog govora s aspekta broja bita kvantizacije i F_u praktično je provjerena, a rezultati su prikazani na slici 5. Oprema je bila ista kao i u gornjem primjeru. U računalo su unesene riječi »levo«, »desno«, »šest« i »nula« pomoću 8-bitnog A/D konvertora i s frekvencijama uzorkovanja 1—13 KHz (uz odgovarajuće Fg niskopropusnih filtera), a povratak u analogni oblik izvršen je pomoću D/A konvertora. Od ispitanika je (10 studenata) traženo da pismeno identificiraju ono što čuju. Razmak između riječi bio je oko 1 sec. Eksperiment je nadmašio očekivanja prepoznavanja riječi jer su u pojedinim slučajevima one bile prepoznate i bitno ispod teorijskog minimuma (npr. »nula« već i pri 3 KHz). U interpretaciji ovog rezultata mora se uzeti u obzir da slušaoci objektivno nisu mogli čuti fizički signal jer je on bio u prevelikoj mjeri izobličen, već da je prepoznavanje uključivalo određenu lingvističku predikciju našeg jezika i intuiciju ispitanika. Ovaj rezultat upućuje na složenost procesa percepcije zvučnih doživljaja čije tumačenje nadilazi okvire ovog rada. Lako je objasniti da riječi koje sadrže sibilante i frikative (»šest«, »desno«) zahtijevaju višu frekvenciju uzorkovanja radi identifikacije, budući da je energija takvih fonema skoncentrirana u gornjem dijelu spektra (iznad 3 KHz).

Linearna kvantizacija zorno prikazuje digitalizaciju signala, međutim, za govorni signal bolje je koristiti logaritamsku skalu jer se na taj način dobiva finija rezolucija za tiše glasovne intervale uz isti brojbita kvantizacije. Takav način kvantizacije poznat je kao Log PCM.



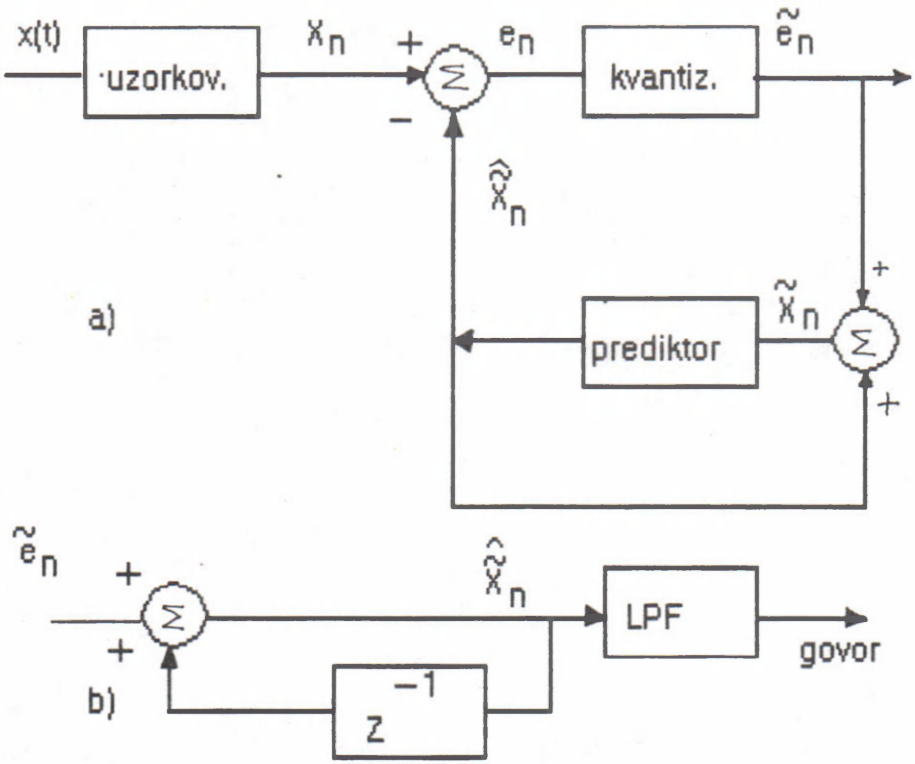
* Kompletnu opremu za govorna istraživanja ustupila je RO Velebit, OOUR Informatika, Zagreb.



sl. 5 Razumljivost govora u ovisnosti od frekvencije uzorkovanja (kvantizacija je 8 bita)

2.2 *Diferencijalna kvantizacija — DPCM*

Za razliku od PCM digitalizacije gdje se svaki uzorak govora nezavisno promatra, tehnike diferencijalne kvantizacije nastoje iskoristiti znatnu korelaciju govornog signala. Osnovna je motivacija smanjenje broja bita za digitalizaciju. Amplitudne su razlike između uzastopnih uzoraka u prosjeku manje nego amplitude pojedinih uzoraka, što dopušta manji broj bita za digitalizaciju govora iste kvalitete. U općem slučaju diferencijalno je kvantiranje predstavljeno na slici 6. Govorni signal x(t) nakon uzorkovanja postaje X_n.



sl. 6 DPCM a) enkoder b) dekoder

Razlika e_n (greška) koja se kvantuje zapravo predstavlja razliku stvarnog uzorka x_n i pretpostavljenog $\hat{x}_n$. Pretpostavljeni uzorak x_n dobija se prediktorom p−og reda odnosno:

$$\hat{x}_n = \sum_{i=1}^{P} a_i \cdot x_{n-i} \qquad ...(3)$$

Koeficijenti predikcije ai odreduju se (Witten82) tako da bude minimalna srednja kvadratna greška:

$$\min\left\{E=(x_n-\sum_{i=1}^{P}a_i\cdot x_{n-i})^2\right\} \qquad ...(4)$$

Znači, izračunata se greška

$$e_n = x_n-\hat{\tilde{x}}_n=x_n-\sum_{k=1}^{r}a_k\cdot\tilde{x}_{n-k} \qquad ...(5)$$

zatim kvantuje i postaje e_n. Kvantovanje e_n u $\tilde{e}_n$ obavlja se s greškom kvantizacije q_n :

$$\begin{aligned}\tilde{e}_n - e = \tilde{e}_n - (x_n - \hat{\tilde{x}}_n &= \tilde{e}_n + \hat{\tilde{x}}_n - x_n \\ &= \tilde{x}_n - x_n \\ &= q_n\end{aligned} \qquad ...(6)$$

Važno je uočiti da se greška kvantizacije zbog prediktora na akumulira :

$$\begin{aligned}q_n &= \tilde{e}_n - e_n \\ &= \tilde{e}_n - (x_n - \hat{\tilde{x}}_n) \\ &= \tilde{x}_n - x_n\end{aligned} \qquad ...(7)$$

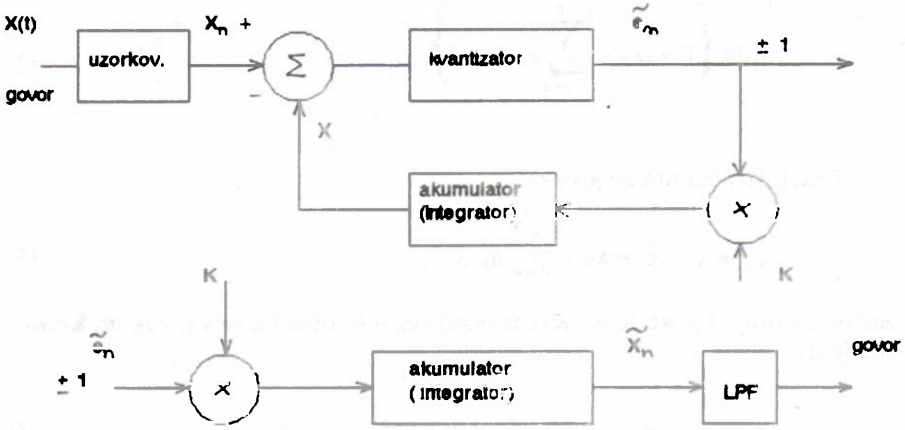
$$=> \tilde{x}_n = x_n + q_n$$

Opisani način digitalizacije naziva se DCPM (Diferential Pulse Code Modulation).

2.3 Binarna kvantizacija — DM, ADM

Ako uzorak signala predstavljamo samo jednim bitom, tada govorimo o binarnoj kvantizaciji. Budući da jednim bitom nije moguće neposredno predstaviti sam uzorak, obično se predstavlja razlika uzastopnih uzoraka ili neki drugi uzajamni odnos. DM (Delta Modulation) je jedna od najpoznatijih tehnika binarne kvantizacije. U stvari, DM je poseban slučaj DPCM gdje

se greška e_n kvantuje s 1 bitom uz prediktor prvog reda. Principijelna shema DM prikazana je na slici 7.



sl. 7 a) DM enkoder b) DM dekoder

Budući da je:

$$\hat{\bar{x}}_n = \bar{x}_{n-1} = \hat{\bar{x}}_{n-1} + \bar{e}_{n-1} \qquad ...(8)$$

tada je:

$$q_n = \bar{e}_n - e_n \qquad ...(9)$$

$$= \bar{e}_n - (x_n - \hat{\bar{x}}_n)$$

odakle je:

$$\hat{\bar{x}}_n = x_{n-1} + q_{n-1} \qquad ...(10)$$

Uočimo da izraz (8) predstavlja integrator s ulazom e_n.

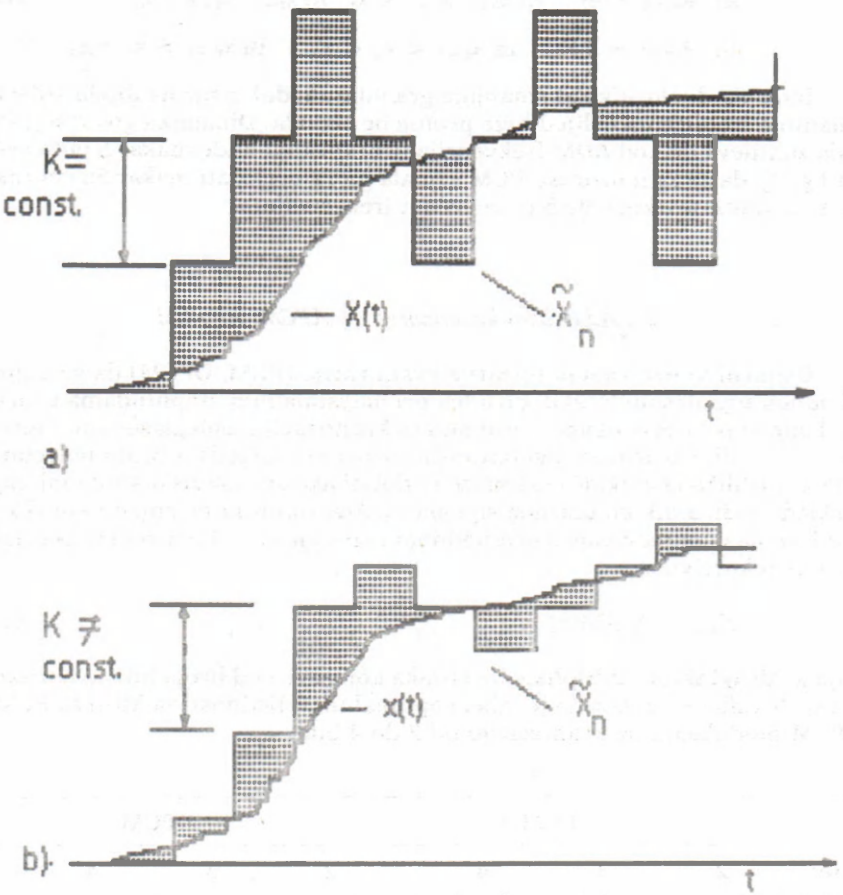
Znači, ako je uzorak x_n veći od x_{n-1}, tada je $e_n = 1$, odnosno 0:

$$\tilde{e}_n = \begin{cases} +1 & \text{za } x_n > x_{n-1} \\ -1 & \text{za } x_n \leq x_{n-1} \end{cases} \qquad ...(11)$$

Digitalizirani signal pomoću DM odlikuje se dvjema vrstama izobličenja:

a) granularna, koja su posljedica koraka integracije
b) oblikovna, koja su posljedica dinamike signala.

Na slici 8 prikazan je taj problem. Ako je korak K velik, bit će izražena izobličenja stacionarnih dijelova, a ako je mali, tada neće biti moguće slijedi-



sl. 8 Osjenčani dio srazmjeran je grešci diskretizacije
a) DM b) ADM

ti brze promjene. Osjenčani dio jednak je grešci rekonstruiranog signala. Rješenje je ovog problema u uvođenju promjenjivog koraka koji slijedi promjenu signala, tj. koraka većeg za brzu promjenu, a manjeg za sporiju kako je ilustrirano na slici 8 b). Takav način digitalizacije naziva se ADM (Adaptive Delta Modulation). U literaturi (Proaksis83) mogu se naci algoritmi za adaptivnu promjenu koraka, a ovdje je iznesena samo osnovna ideja:

$$a)\ K_{n+1} = \frac{K_n}{2} \quad \text{za } x_{n+1}, x_{n-1} < x_n \text{ ili } x_{n+1}, x_{n-1} > x_n \qquad ...(12)$$

$$b)\ K_{n+1} = 2\,K_n \quad \text{za } x_{n+1} < x_n < x_{n-1} \text{ ili } x_{n-1} > x_n > x_{n-1}$$

Izraz na dijelu slike a) smanjuje granularna, dok izraz na dijelu slike b) smanjuje izobličenja uslijed brze promjene signala. Dinamika govornog signala zahtijeva da kod ADM frekvencija uzorkovanja bude makar 5 puta veća od Fg., tj. da se za n uzoraka PCM signala mora osigurati makar 5n uzoraka e_n, tj. izvršiti uzorkovanje 5 puta većom frekvencijom.

2.4. Adaptivna kvantizacija – APCM, ADPCM

Osnovni je nedostatak linearne kvantizacije (PCM, DPCM) da se najbolji odnos signal/šum (s/š) dobiva tek pri maksimalnim amplitudama uzorka x_n. Time se povećava ukupan broj bita za kvantizaciju tiših glasovnih intervala ako se želi ostvariti zadovoljavajuća razina s/š. Ušteda u broju potrebnih bita za približno isti odnos s/š može se dobiti ako se naponski korak mijenja u skladu s dinamikom ulaznog signala x_n. Algoritam za promjenu koraka K svodi se na uspoređivanje s prethodnom razinom x_{n-1}. Veličina koraka definira se rekurzivno:

$$K_{n+1} = K_n \cdot M(n) \qquad ....(13)$$

gdje je M(n) faktor multiplikacije koraka koji zavisi od broja bita kvantizacije i x_n. Na slici 9. prikazana je tabela optimalnih vrijednosti za M(n) za PCM i DPCM modulaciju uz kvantizaciju od 2 do 4 bita.

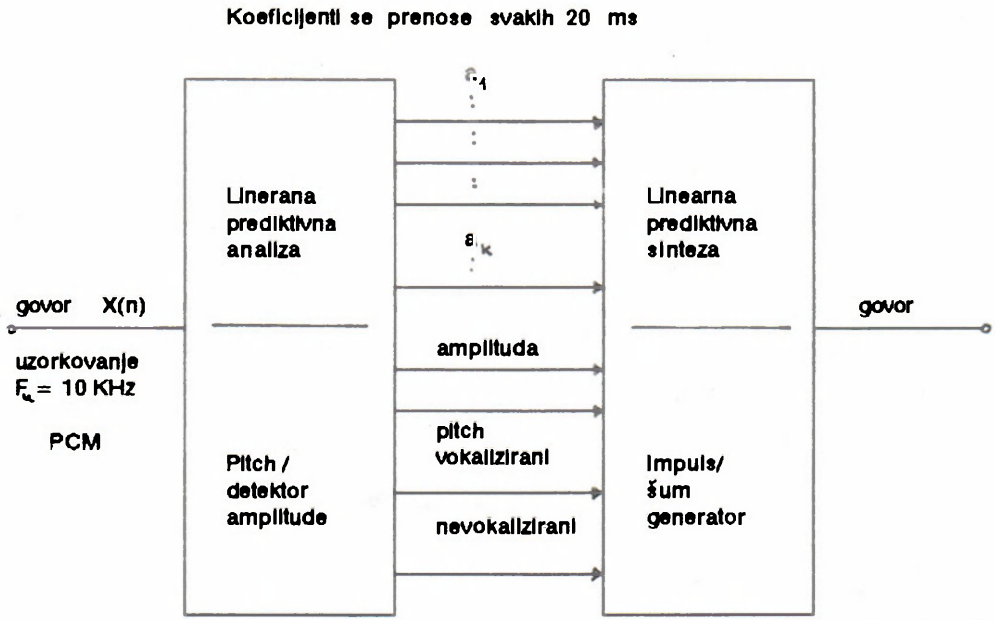
	PCM			DPCM		
bita:	2	3	4	2	3	4
M(1)	0.60	0.85	0.80	0.80	0.90	0.90
M(2)	2.20	1.00	0.80	1.60	0.90	0.90
M(3)		1.00	0.80		1.25	0.90
M(4)		1.50	0.80		1.70	0.90
M(5)			1.20			1.20
M(6)			1.60			1.60
M(7)			2.00			2.00
M(8)			2.40			2.40

sl. 9 Faktori multiplikacije koraka prilagođeni za govorni signal (Jayant 74)

Adaptivni mehanizam za DM modulaciju već je ranije opisan.

2.5 Linearna predikcije govora — LPC

Tehnika linearnog prediktivnog kodiranja (LPC — Linear Predictive Code) relativno je novijeg datuma. U osnovi, LPC pripada grupi tehnika digitalizacije u vremenu, međutim, u analizi govora uglavnom se interpretira u frekvencijskom prostoru. Za razliku od opisanih načina digitalizacije koji se primjenjuju na opći analogni signal, LPC je isključivo motivirana modeliranjem vokalnog trakta i predstavljanjem govornog signala. Na slici 10 prikazana je shema LPC digitalizacije i sinteze govora.



sl. 10 Principijelna shema LPC kodiranja

Koeficijenti a_i zapravo su koeficijenti prediktora p-tog reda koji se ovdje nazivaju koeficijenti refleksije (Markel76) i izračunavaju prema kriteriju minimalnog kvadrata greške kao i kod DPCM.

$$e_n = x_n - \sum_{k=1}^{P} a_k \cdot x_{n-k} \qquad ...(14)$$

Međutim, umjesto signala greške e_n, sada se prenosi informacija o osnovnom tonu i njegovoj amplitudi i informacija je li riječ o vokaliziranom ili nevokaliziranom segmentu govora. Parametri: koeficijenti predikcije a_1 ... a_k, osnovni ton, vokaliziranost/nevokaliziranost čine opis govornog segmenta trajanja 15 — 25 ms. Zbog mnogo dužeg intervala prijenosa parametara (oko

100 puta nego kod PCM), ukupan broj bita za prijenos 1 sekunde govora ne prelazi brojku od nekoliko tisuća.

Primjenjujući Z transformaciju (Rabiner79) na (14) dobivamo slijedeći izraz:

$$E(Z) = X(Z) - \sum_{k=1}^{P} a_k Z^{-k} X(Z) = (1 - \sum_{k=1}^{P} a_k Z^{-k}) \cdot X(Z) \qquad ...(15)$$

A nakon sređivanja:

$$X(Z) = \frac{1}{1 - \sum_{k=1}^{P} a_k Z^{-k}} E(Z) \qquad ...(16)$$

Produkciju govora možemo modelirati kao pobudni izvor (oscilator) čiji valni oblici prolaze mehaničkim rezonatorom (vokalnim traktom) pri čemu se mijenja njegova prijenosna funkcija. Promatrajući vokalni trakt kao niz digitalnih formantnih filtera, njegova prijenosna funkcija može se prikazati (Witten82) kao:

$$\frac{1}{1 - b_1 Z^{-1} + b_2 Z^{-2}} \qquad ...(17)$$

gdje b_1 i b_2 određuju položaj i širinu formantnih rezonanci. Ako se uzme digitalni filter prvog reda:

$$\frac{1}{1 - b_1 Z^{-1}} \qquad ...(18)$$

tada je ostvariva spektralna kompenzacija —6dB po oktavi. Proizvod od (18) može se napisati u obliku:

$$\frac{1}{1 - c_1 Z^{-1} - c_2 Z^{-2} - ... - c_q Z^{-q}} \qquad ...(19)$$

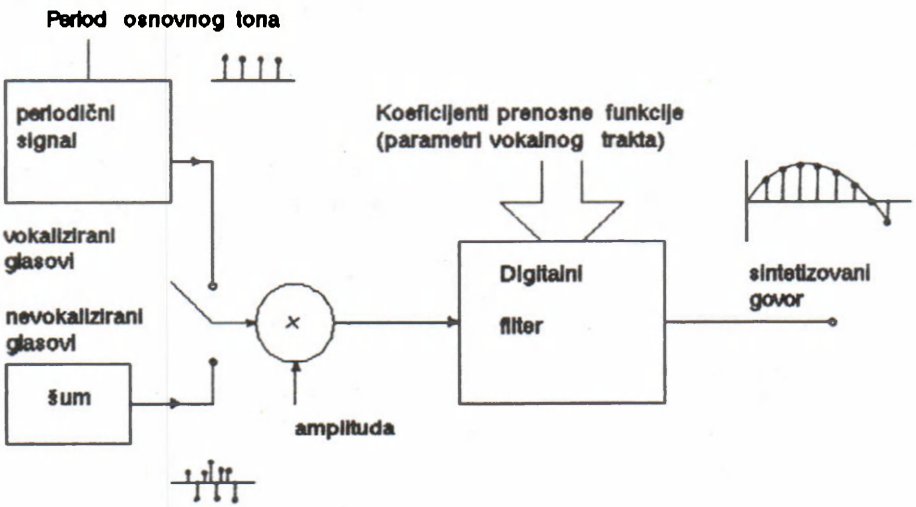
gdje je q dvostruki broj formanata uvećan za jedan, dok se c_i izračunava prema pozicijama i širinama formanata te parametrom spektralne kompenzacije. Na kraju dobivamo Z transformaciju:

$$X(Z) = \frac{1}{1 - \sum_{k=1}^{q} c_k Z^{-k}} I(Z) \qquad ...(20)$$

gdje je I(Z) transformacija izvora. Po obliku, izraz (20) veoma je sličan izrazu (16). Ako su p i q jednaki, tada koeficijenti linearne predikcije a čine polinom p-tog reda koji je isti kao i polinom koji se dobija množenjem polinoma po-

put (17) koji predstavljaju pojedinačne formante. Isto tako, prediktivna se greška E(z) iz (16) u (20) pojavljuje kao I(z), odnosno pobudni signal. Zbog opisane analogije moguće je signal greške parametarski predstavljati pomoću frekvencije i amplitude umjesto vremenskog uzorkovanja i amplitudne kvantizacije. Na kraju se uočava način na koji se odvajaju osobine izvora od prijenosne funkcije filtra vokalnog trakta: parametri pobude mogu se izvesti iz signala greške a filter (koji predstavlja model vokalnog trakta) odreden je koeficijentima linearne predikcije. Zbog fizikalnog značenja, koeficijenti predikcije katkad se nazivaju i koeficijenti refleksije. Broj koeficijenata predikcije uzima se između 10 i 15, što odgovara sintezi od 3 i 7 formanata respektivno.

Zorna slika sinteze govora LPC tehnikom prikazana je na slici 11.



sl. 11 LPC sinteza govora

Prema tome je li segment vokaliziran ili nevokaliziran (odreduje se 1 bitom), signal je pobude (tj. e_n) periodičan ili neperiodičan, odnosno bijeli šum.

3. Analiza govornog signala u vremenskoj domeni

Digitalizirani govorni signal rijetko se kada promatra u svom osnovnom obliku, tj. u odgovarajućem digitalnom kodu (PCM... LPC). Fonetičara zanima reprezentacija nekih parametara koji karakteriziraju govorni signal prema zadatom modelu. Na primjer, gruba slika o izgovorenoj riječi može se

dobiti promatranjem PCM slike poput one na sl. 4. Međutim, najčešće to nije dovoljno, već se traže drugi parametri obilježja. U vremenskoj analizi u pravilu se analizira uvijek unutar segmenata (prozora) tipično 5—40 ms budući da je u tom intervalu govorni signal kvazistacionaran (zbog inercije govornog aparata). Matematička je definicija prozora (funkcije segmentacije) najčešće:

$$w(n) = \begin{cases} 1 & 0 \leqslant n \leqslant N-1 \\ 0 & \text{inače} \end{cases} \qquad ...(21)$$

Kod definiranja prozora moramo biti oprezni da ne bismo izgubili potrebnu informaciju, što ćemo kasnije pokazati. Radi bolje preglednosti u nastavku teksta početni indeks digitaliziranog signala bit će 0, a i-ti digitalizirani uzorak (govorni signal) bit će označen x(i).

3.1 Kratkovremenska energija

Najjednostavnija je reprezentacija digitaliziranog govornog signala x(n) energija:

$$E = \sum_{n=0}^{K-1} x(n)^2 \qquad ...(22)$$

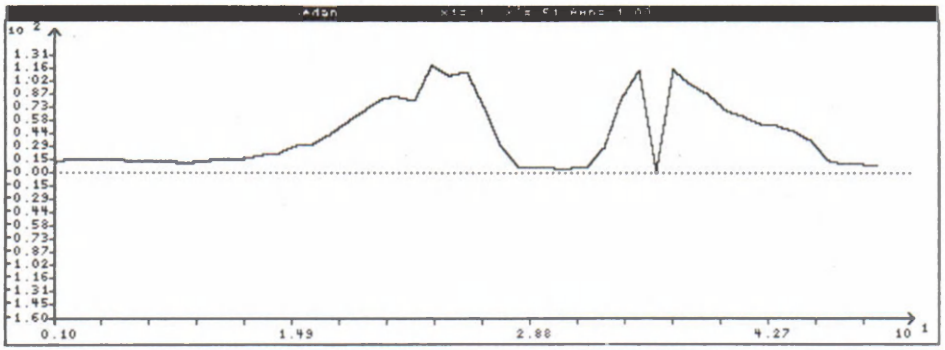
Međutim, srednja energija cijelog signala (riječi, rečenice) ne daje značajnu distinktivnu informaciju, što je prikazano na slici 12. Umjesto prema izrazu (22), energija se uglavnom analizira preko funkcije kratkovremenske energije:

$$E(n) = \sum_{m=0}^{N-1} [w(m) \cdot x(n-m)]^2 \qquad ...(23)$$

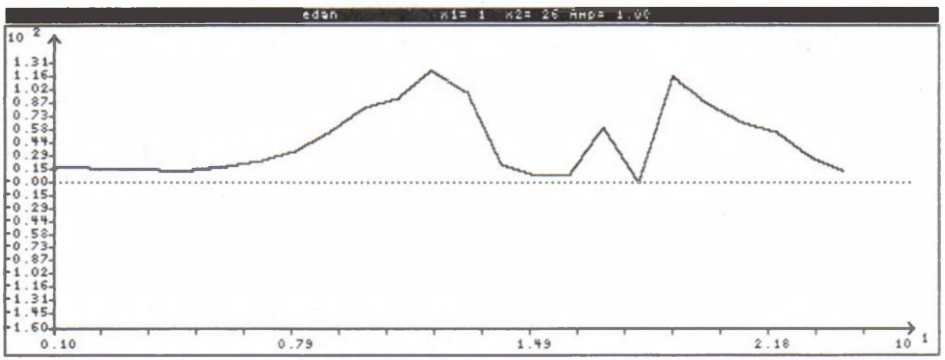
gdje je w(n) prozor kojim se odabire N uzastopnih uzoraka.

RIJEČ	ENERGIJA	RIJEČ	ENERGIJA
nula	6.135 E6	pet	1.425 E6
jedan	3.533 E6	šest	1.160 E6
dva	1.000 E7	sedam	2.530 E6
tri	2.946 E5	osam	5.020 E6
četiri	2.125 E6	devet	3.434 E6

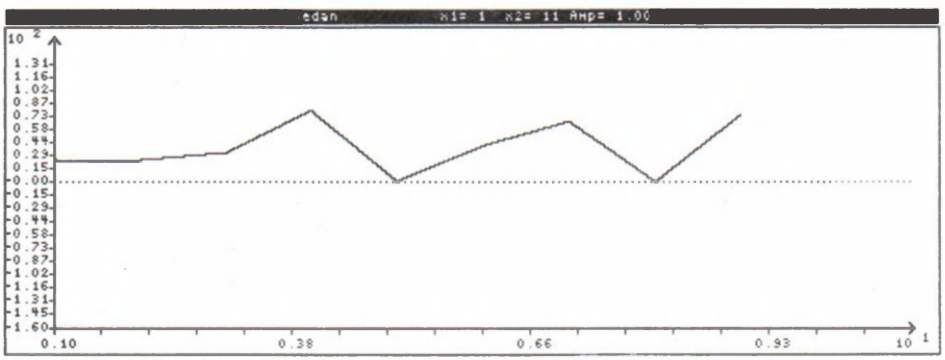
sl. 12 Energija brojki 0—9



a)



b)



c)

sl. 13 Kratkovremenska energija riječi »JEDAN«
a) N= 100 b) N= 200 c) N= 500

Znači, E(n) predstavlja sumu energije posljednjih N uzoraka do x(n). Međutim, takva interpretacija može biti smislena samo u slučaju pravilnog odabira širine prozora w (tj. N). Ako je w premali, E(n) će se veoma brzo mijenjati slijedeći detalje valnog oblika, čime se ne dobija ništa bitno u odnosu na već postojeću sliku PCM signala. Ako je w prevelik, na primjer nekoliko perioda osnovnog tona, E(n) će veoma malo varirati i neće biti moguće uočiti varijacije segmenata govora (vokalizirani, nevokalizirani, akcent itd.), tj. bit će veoma sličan E. Praktično, vremenska je širina prozora 10—20 ms, a odatle slijedi da B (širina odnosno broj uzoraka koje zahvaća prozor) treba da bude unutar intervala:

$$\frac{10}{F_u} \leqslant B \leqslant \frac{20}{F_u} \qquad ...(24)$$

gdje je F_u frekvencija uzorkovanja u KHz. Tako dobivamo da je vrijednost za B između 100 i 200 za F = 10 KHz. Izraz (22) jako naglašava intenzivne segmente govora tako da su tiši glasovni intervali praktično zanemareni. Zato se često upotrebljava nešto izmijenjena formulacija kratkovremenske energije:

$$E(n) = \sum_{m=0}^{N-1} [w(m) \cdot x(n-m)] \qquad ...(25)$$

gdje se umjesto sume kvadrata uzima suma apsolutnih vrijednosti. Na slici 13 prikazana je funkcija kratkovremenske energije za riječ »jedan« i to za B = 100, B = 200 i B = 500 (izgled iste riječi u PCM-u prikazan je na slici 4a).

3.2 Broj prolazaka kroz nulu

Prelazak iz vremenske u frekvencijsku domenu digitalne obrade signala praćen je značajnim brojem računskih operacija (DFT, WFT, FFT itd.) koje se bez posebnih signal-procesora ne mogu obaviti u realnom vremenu. Alternativni način grube procjene frekvencije unutar vremenske domene može se ostvariti mjerenjem prolazaka signala kroz nulu. Osnovna ideja prikazana je na slici 14.

Ako je N broj prolazaka kroz nulu u intervalu T, tada je frekvencija pravilnog kosinusnog (sinusnog) signala sa slike 14:

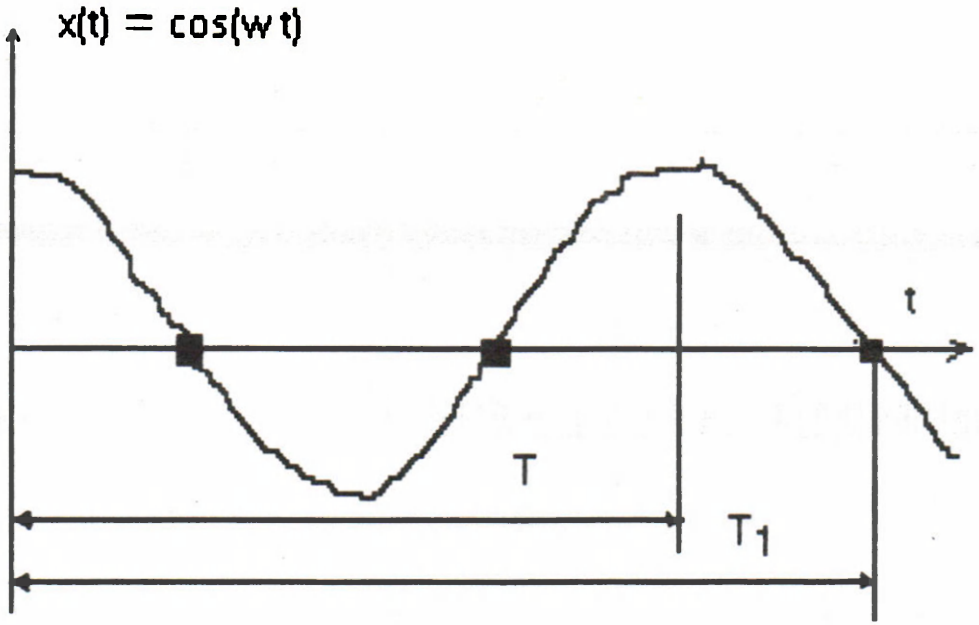
$$F = \frac{1}{T} = \frac{1}{2} N_z \qquad ...(26)$$

Odnosno za interval vremena T1 broj prolazaka kroz nulu određuje srednju frekvenciju unutar T:

$$N_z(T_1) = 2 \cdot \overline{F} \qquad ...(27)$$

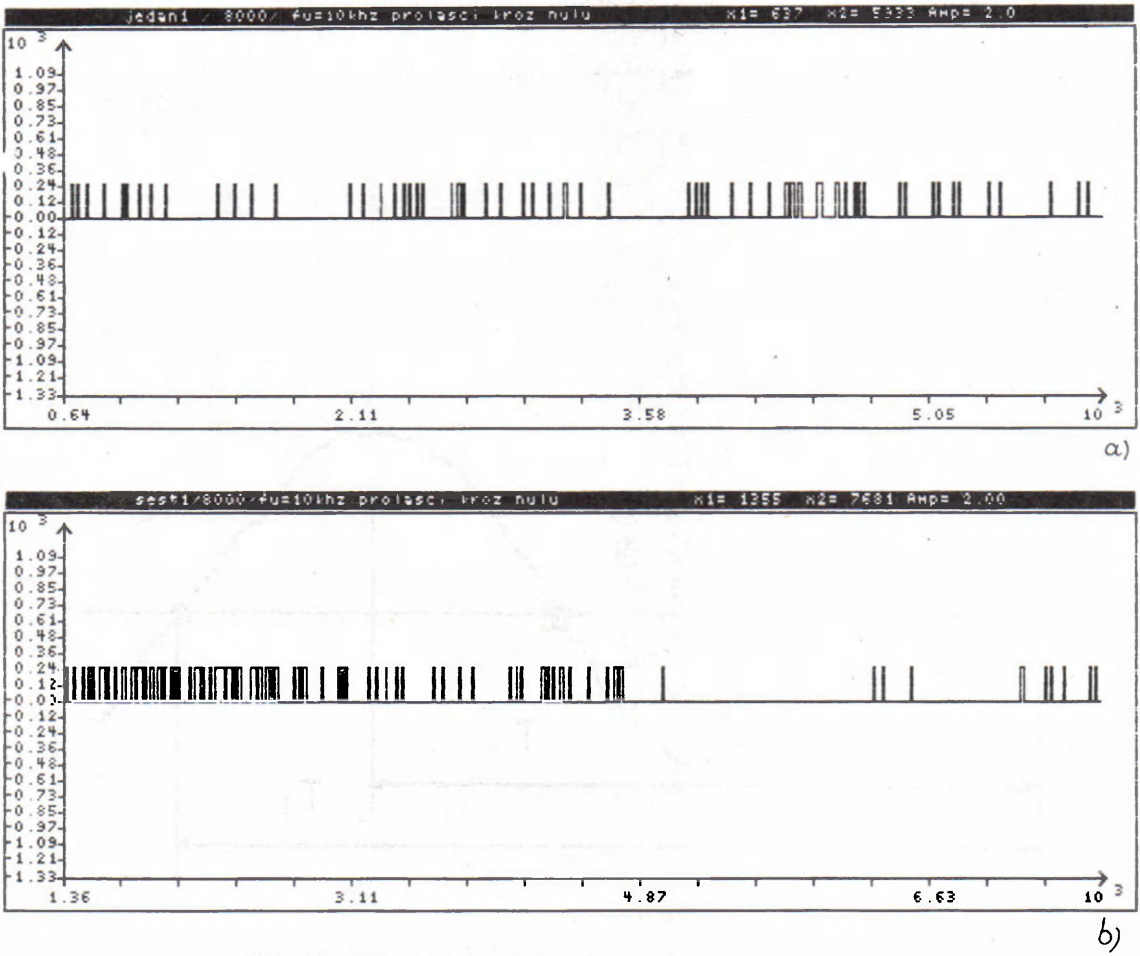
Za digitalni je signal funkcija prolaza kroz nulu definirana:

$$Z(n) = \begin{cases} 1 & \text{za } \operatorname{sign}(x(n)) \neq \operatorname{sign}(x(n-1)) \\ 0 & \text{inače} \end{cases} \qquad \ldots(28)$$



sl. 14 Kosinusni (sinusni) signal u toku periode dva puta prode kroz nulu

Informacija o N_z za govorni signal nema isti fizikalni smisao kao kod sinusnog signala (27), međutim, broj prolaza kroz nulu implicitno navodi na vokalizirani (manji broj) ili nevokalizirani (veći broj prolaza kroz nulu) segment govora. Zajedno s mjerenjem energije, informacija o N_z često se primjenjuje za automatsku segmentaciju govornog niza, određivanje početka ili kraja riječi, sloga itd. Na slici 15 vide se funkcije prolaza kroz nulu za riječi »jedan« i »šest«. Lako se uočava da je N »gušća« oko »s« i »š« nego na mjestima »e« ili »a«. Slike su dobijene prema izrazu (28), a tehnika je digitalizacije bila PCM.



sl. 15 Funkcija prolaska signala kroz nulu
a) »jedan« b) »šest«

3.3 Kratkovremenska autokorelacija

Funkcija autokorelacije (Autocorrelation Function) upotrebljava se za prikaz osnovnih parametara signala (energija, periodičnost itd.) i definirana je na sljedeći način:

$$ACF(m) = \lim_{n \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n) \cdot x(n+m) \qquad \ldots(29)$$

Karakteristike x(n) izravno se manifestiraju u funkciji ACF. Npr. ako je x periodična, tada je periodična i ACF(n) tj.:

$$x(n+T) = x(n) \quad \Rightarrow \quad ACF(m) = ACF(m+T) \qquad \ldots(30)$$

Za m=0, ACF(m) fizikalno predstavlja srednju energiju signala. Dalje, ako ACF(m) brzo opada, to pokazuje visok stupanj korelacije uzastopnih uzoraka. Za analizu govornog signala definira se kratkovremenska korelacija koja se primjenjuje nad stacionarnim segmentima signala. Označimo segment od N uzoraka:

$$x_l(n) = x(n+l), \qquad 0 <= n <= N-1 \qquad \ldots(31)$$

gdje je l početni indeks segmenta. Kratovremenska je autokorelacija tada:

$$ACF(m) = \frac{1}{N} \sum_{m=0}^{N'-1} x_l(m) \cdot {}_l(n+m), \quad 0 <= m <= Mo-1 \qquad \ldots(32)$$

gdje Mo označava broj pomaka autokorelacije. Ako želimo promatrati periodičnost signala periode T, tada Mo mora biti $Mo/F_u > T$. Izraz (32) interpretiramo kao autokorelaciju signala od N uzoraka koji počinje na l-tom uzorku. Ako je N´= M—m, tada se izračunavaju uzorci s indeksima:

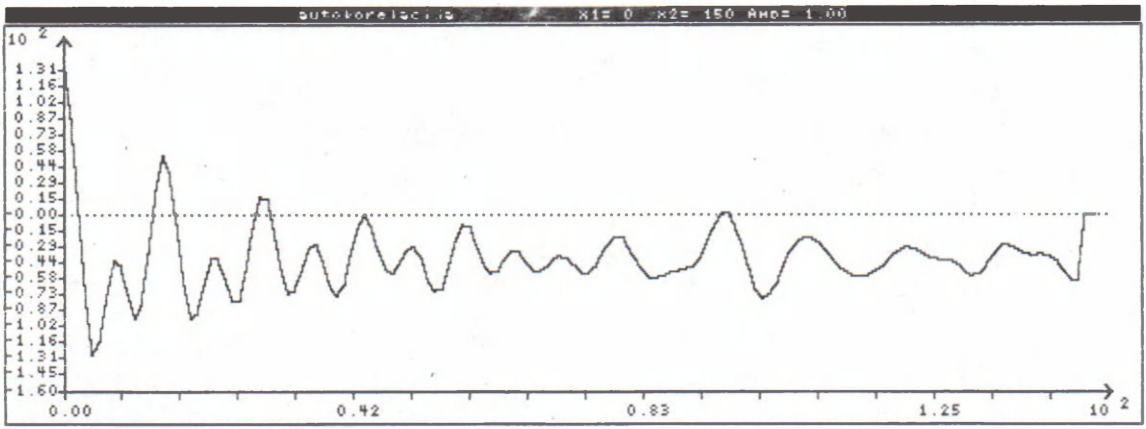
$$l \leqslant n \leqslant N+1-l \qquad \ldots(33)$$

ako je N´= N + l—1, tada indeksi prelaze gornje područje. U prvom slučaju signal se umnožava vremenskim prozorom koji sve vrijednosti x(n) gdje je indeks n izvan dimenzija prozora svode na 0. Izravno izračunavanje ACF(m) od 0 < = m < = Mo—1 zahtijeva Mo*N operacija, što nije zanemarivo. Pomoću funkcije kratkovremenske autokorelacije razvijeni su mnogi algoritmi za izračunavanje frekvencije osnovnog tona (Pitch Period) (Sondhi79), koeficijenata predikcije te prepoznavanja i sinteze govora. Odabir veličine N obično pada u ekvivalentni vremenski interval 20 — 40 ms, što npr. znači da za F_u = 10Khz (PCM), N treba da bude između 200 i 400.

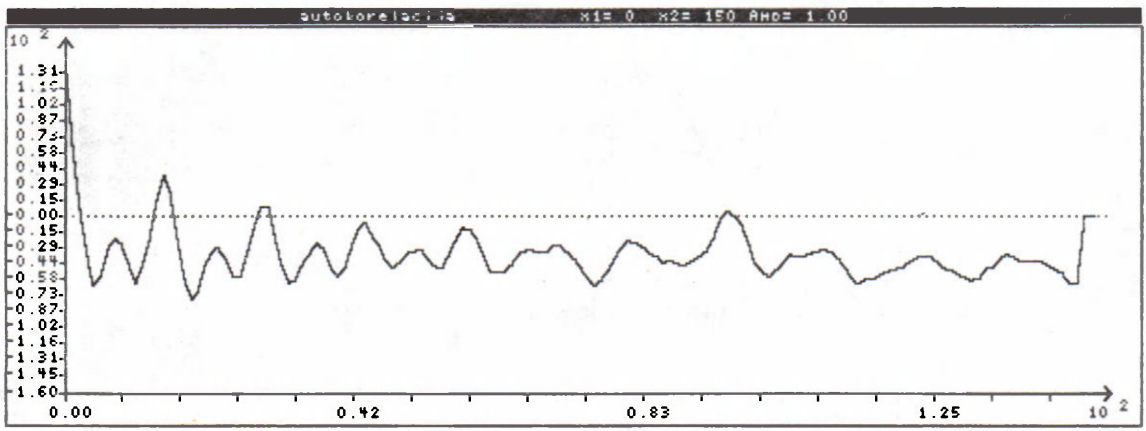
Osim izraza (32) često se upotrebljava autokorelacija govornog signala amplitudno kvantiranog samo jednim bitom (Clipping — odsijecanje) koji predstavlja predznak x(n) tj.

$$CACF(m) = \sum_{m=0}^{K-1} sign(x(k)) \cdot sign(x(k+m)) \qquad \ldots(34)$$

Na slici 16 vide se normirane ACF(m) i CACF(m) za vokalni segment sa slike 4b) uz Mo = 150.



a)



b)

sl. 16 Autokorelacija a) ACF b) CACF

Podsjetimo se da ACF(0) predstavlja ukupnu energiju segmenta.

Zaključak

Od prikazanih načina digitalizacije govornog signala za fonetska istraživanja najpogodniji su PCM i LPC. Iako po količini bita potrebnih za kodiranje govora PCM najviše zahtijeva, zbog jednostavnosti dalje obrade i zorne slike signala uređaji za PCM (A/D, D/A konvertori) obavezan su dio svakog digitalnog fonetskog laboratorija. LPC tehnika omogućuje i frekvencijsko-formantnu analizu, određivanje frekvencije osnovnog tona i široko polje eksperimentiranja s različitim modelima vokalnog trakta, već zamjenjuje klasični FFT digitalnu, frekvencijsku analizu. U opisanim eksperimentima pokazano je da teorija digitalizacije signala u smislu razumljivosti digitaliziranog govora nije potpuna jer nedostaje matematički model zvučne percepcije ispitanika. Iako su opisane samo osnovne digitalne analize govora, jasno se uočava numerički aspekt digitalne obrade i naslućuju mogućnosti koje su bile nedostupne klasičnoj analognoj analizi.

+

LITERATURA

Jayant N. S. (1974), Digital Coding of Speech Waveforms: PCM, DPCM, and DM Quantiziers, IEEE, vol. 62, str. 611—632, New York: IEE Press

Markel J. D. i *Gray A. H.* (1976), Linear Prediction of Speech, New York: Springer—Verlag

Proakis G. John (1983), Digital Communications, Singapore: McGraw—Hill

Rabiner R. L. i *Schafer W. Ronald* (1978), Digital Processing of Speech Signals, Englewood Cliffs: Prentice—Hall

Schafer W. R. i *Rabiner R. L.* (1979), Digital Representation of Speech Signals, (Schafer W. R. i Markel D. J, izdavači), str. 82—97, New York: IEEE Press

Sondhi M. M. (1979), New Methods of Pitch Extraction, (Schafer W. R. i Markel D. J, izdavači), str. 153—157, New York: IEEE Press

Witten H. Ian (1982), Principles of Computer Speech, London: Academic Press

Milan STAMENKOVIĆ
Zagreb

Digital Representation and Analysis of Speech Signals in Time Domain

SUMMARY

The paper presents several time domain digital signal processing and analysis techniques for speech representation. Both theoretic principles and the results of speech inteligibility experiments are included. Most of presented digital methods are discussed through practical examples.