

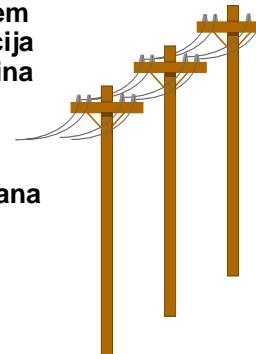
<h2>Digitalna obrada govora</h2> <p>Davor Petrinović, dr.sc.</p>	<h3>Digitalna obrada govornih signala</h3> <ul style="list-style-type: none"> • Koja je namjena govora • komunikacija • Govor je signal koji nosi informaciju (akustički valni oblik) 	<h3>Proces nastajanja govora u komunikaciji</h3> <ul style="list-style-type: none"> • poruka u apstraktnom obliku se pojavljuje u mozgu • poruka se pretvara u skup neuronskih signala koji upravljaju postupkom artikulacije (pomicanje jezika, usnica, glasnica, itd.) • micanje artikulatora formira zvučni signal koji sadrži informaciju originalne poruke 
<h3>Elementi govora</h3> <ul style="list-style-type: none"> • Informacija koja se prenosi govorom je vremenski diskretne prirode • nastaje spajanjem elemenata iz konačnog skupa simbola 	<h3>Elementi govora</h3> <ul style="list-style-type: none"> • Ti osnovni elementi se nazivaju FONEMI ili GLASOVI • Svaki jezik ima svoj raznolik skup fonema, (tipično 30 do 50 fonema) • npr. engleski jezik ima oko 42 fonema 	<h3>Glavne primjene digitalne obrade govora (I dio)</h3> <ul style="list-style-type: none"> • kodiranje govora u svrhu učinkovitijeg prijenosa ili pohrane, • prepoznavanje govora, diktiranog ili prirodnog s vezanim izgovorom riječi, • prepoznavanje govornika, • prepoznavanje jezika, • sustavi za detekciju riječi, • sinteza govora, ...
<h3>Glavne primjene digitalne obrade govora ... (II dio)</h3> <ul style="list-style-type: none"> • sustavi za dijalog čovjeka i računala, • sustavi za raznovrsne transformacije govornog signala kao što su promjena visine gasa, brzine izgovora itd., 	<h3>Glavne primjene digitalne obrade govora ... (III dio)</h3> <ul style="list-style-type: none"> • sustavi za pomoć osobama oštećena vida, sluha ili govora, • sustavi za poboljšanje kvalitete govora, • analiza govora u svrhu dijagnostike raznovrsnih oboljenja 	<h3>Potrebna znanja za digitalnu obradu govora</h3> <ul style="list-style-type: none"> •vremenski diskretni signali i sustavi •transformacije reprezentacija signala i sustava <ul style="list-style-type: none"> –z-transformacija –Fourier-ova transformacija –diskretna Fourier-ova transformacija •osnove digitalnih filtera FIR, IIR •problem otiskivanja <ul style="list-style-type: none"> –teorem otiskivanja –decimacija –otiskivanje valnog oblika –interpolacija –otiskivanje valnog oblika

Kodiranje govora

- Što je kodiranje govora ?
 - Da li je to kriptografska zaštita ?
 - ... Ne, riječ je isključivo o učinkovitoj digitalnoj reprezentaciji
- Gdje se primjenjuje ?
 - U svim suvremenim sustavima temeljenim na digitalnom zapisu ili prijenosu
 - danas su analogni sustavi praktički izbačeni iz upotrebe

Količina informacije u govoru

- Središnji problem teorije informacija je potrebna brzina prijenosa informacije
 - gruba procjena je vezana uz fizička ograničenja pomicanja artikulatora



Količina informacije u govoru

- tipična brzina govora iznosi 10 fonema u sekundi
- pošto je za kodiranje fonema dovoljno 6 bita
- oko 60bita/s je dovoljno za prijenos govorne informacije



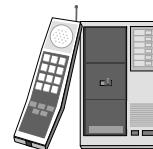
Količina informacije u govoru

- međutim, donja granica stvarnog informacijskog sadržaja govora je mnogo viša
- navedena procjena ne uključuje identitet govornika, emocionalno stanje, brzinu izgovora, glasnoću, itd.

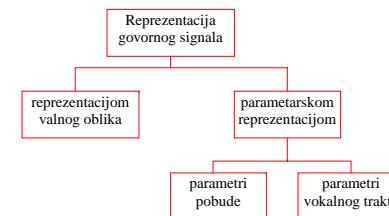


Sustavi za prijenos, pohranu i obradu govora

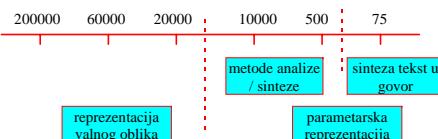
- očuvanje sadržaja poruke govornog signala
- reprezentacija govornog signala u obliku koji je pogodan za prijenos, pohranu ili neku vrstu obrade, ali takva reprezentacija ne smije značajno narušiti sadržaj govorne poruke



Reprezentacija govora



Potrebna brzina prijenosa (bit / sekundi)



Sustavi sa reprezentacijom valnog oblika signala

- pojavili se već početkom 70-tih godina
- ... složenost analognih centrala postala je prevelika i prijetila je zagrušenjem sustava
- Najznačajniji standardi iz tog doba su PCM (64 kbit/s) i ADPCM (32 kbit/s)
- Većina uredskih telefonskih instalacija u USA je i danas temeljena na tim standardima

Sustavi sa reprezentacijom valnog oblika signala

- navedeni koderi su standardizirani od strane međunarodnog tijela za standardizaciju telekomunikacijskog sektora ITU-T (The Telecommunication Standardization Sector of the International Telecommunication Union)
- označe G.711 (PCM) i G.721, G.726, G.727 (ADPCM)

Sustavi sa reprezentacijom valnog oblika signala

- to su algoritmi koji nastoje valni oblik govornog signala čim je moguće bolje reprezentirati i prenijeti na prijemnu stranu
- nije specijalno korištena činjenica da se radi o govornom signalu koji ima vrlo specifična svojstva,
- ti se koderi kao takvi mogu koristiti i za prijenos drugih signala (npr. muzike).

Prvi sustavi s parametarskom reprezentacijom

- Krajem 80-tih godina, ... dodatni zamah razvoju postupaka učinkovitog kodiranja govornog signala dao je razvoj mobilnih komunikacija
- Kod mobilnih sustava, cijena kanala je direktno proporcionalna korištenoj brzini prijenosa
- ⇒ zahtjev za učinkovitim sažimanjem je od presudne važnosti

Prvi sustavi s parametarskom reprezentacijom

- značajan problem mobilnih digitalnih komunikacija
 - nepouzdanost i mala kvaliteta prijenosnog kanala, tj. velika vjerojatnost pogreški u prijenosu, kao i povremeni totalni prekidi kanala.
- zbog toga, ... potreba za algoritmima koji su imuni na takove probleme u prijenosu
- pored svega ... zahtjev na jednostavnost i malu potrošnju električne energije

Prvi sustavi s parametarskom reprezentacijom

- nažalost ... nije оформljen svjetski standard
 - svatko je gurao svoj sustav, a glavni "igrači" su:
- sjeverno američki standard IS54 VSELP
 - 1989 TIA (Telecommunication Industry Association) - brzina prijenosa 7.95 kbit/s

Prvi sustavi s parametarskom reprezentacijom

- japanski standard JDC-VSELP
 - standardiziran od strane RCR (Research and Development Center for Radio Systems) pod oznakom RCR STD-27B
- europski standard GSM, RPE-LTP koder
 - 1987 Groupe Special Mobile of CEPT 13 kbit/s

Prvi sustavi s parametarskom reprezentacijom

- svi navedeni sustavi su posebno prilagođeni govornom signalu,
- visoka učinkovitost sažimanja ostvarena na račun činjenice da govorni signal u sebi sadrži značajnu količinu redundantne (nebitne) informacije
- UŠTEDA u brzini prijenosa...
 - razdvajanjem 'bitne' informacije od 'nebitne', te opisom 'nebitnog' dijela modelom, a kvantizacijom, kodiranjem i prijenosom samo 'bitnog' dijela

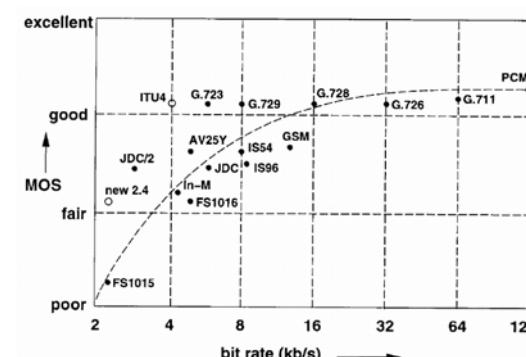
Noviji koderi govornog signala

- početkom 90 godina, ...
 - želja za udvostručenjem broja telefonskih kanala prepunjavanjem brzine prijenosa uz očuvanje kvalitete (Half Rate, HS Coders)
 - ili, povećanje kvalitete uz istu brzinu prijenosa (Enhanced Full Rate, EFR Coders)
 - primjer takvih kodera:
 - half-rate GSM standard ETSI-TCH-HS, 1994, brzine prijenosa od 5.6 kbit/s.
 - novi sjeverno-američki standard, IS96 QCELP
 - novi japanski standard JDC Half-Rate PSI-CELP koder brzine prijenosa od 3.45 kbit/s

Noviji koderi govornog signala

- dodatna učinkovitost sažimanja ostvarena na račun značajnog povećanja kompleksnosti kodiranja
- Danas zbog visoke konkurenkcije... naglasak je na kvaliteti reproduciranog signala, dok je brzina prijenosa od sekundarnog prioriteta
- ... težnje za proširenjem spektralnog pojasa sa 3.5 kHz na 7 kHz, tzv. Wide Band Coders

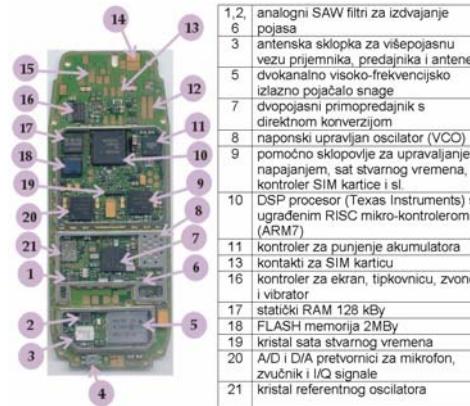
Ovisnost kvalitete i brzine kodera (Mean Opinion Score)



Izvedba u mobilnim telefonima

- Mobilni telefoni imaju namjenski procesor koji je zadužen za kodiranje i dekodiranje govora
- Riječ je o procesoru za digitalnu obradbu signala (DSP)
- Mobiteli novijih generacija imaju i do 4 procesora (2 DSP i 2 mikro-kontrolera)
- Koder je izведен programski na DSPu (Real Time !)

Izvedba tipičnog mobitela



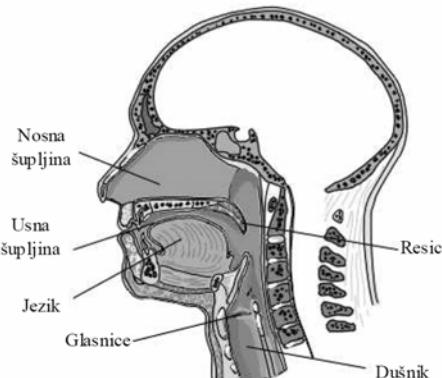
Parametarska reprezentacija

- Govorni signal je predstavljen modelom, a ne valnim oblikom!
- Model je opisan parametrima.
- Parametri modela se određuju vremenski kratkotrajnim postupcima.
- Parametri se kvantiziraju, kodiraju i pretvaraju u digitalni niz željene brzine prijenosa.
- Na dekoderu se provodi obrnuti postupak.

Parametarska reprezentacija

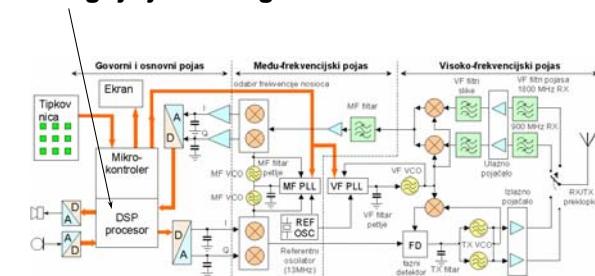
- Pogreške parametarske reprezentacije:
 - uslijed neusuglašenosti modela i stvarnog fizikalnog procesa,
 - uslijed pogrešaka prilikom estimacije parametara modela,
 - uslijed kvantizacijskih pogrešaka parametara.
- Povećanje kvalitete
 - Hibridnim postupcima

Anatomija govornog sustava



Izvedba tipičnog mobitela

- Struktura mobitela
- ... gdje je koder govora ?



Modeliranje govora

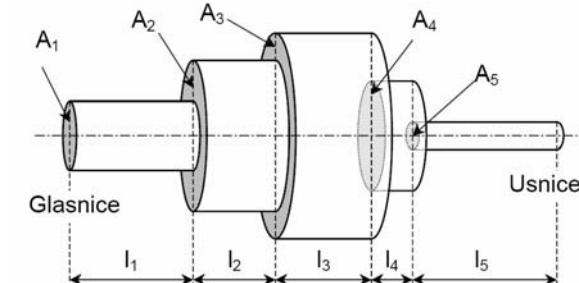
- Model mora biti blizak sa fizikalnim procesom nastajanja govornog signala
 - Tokom 60-tih i 70-tih godina detaljno su istraživani akustički modeli govora:
 - proces nastajanja signala,
 - širenje kroz vokalni trakt,
 - zračenje na usnicama i širenje zvučnog vala otvorenim prostorom.

Model sa spojenim cijevima

- Pokazalo se da je oblik vokalnog traka moguće modelirati skupom cijevi, čiji presjek varira od glasnica do usnica.
 - Već i sa relativno malim brojem cijevi postiže se dobra aproksimacija akustičke prijenosne funkcije.
 - Takov pojednostavljeni model je omogućio određivanje analitičkih rješenja odziva za zadalu pobudu i zadan oblik vokalnog traka.

Model sa spojenim cijevima

- Svaki glas ima karakteristični skup cijevi, koji aproksimira oblik vokalnog traka

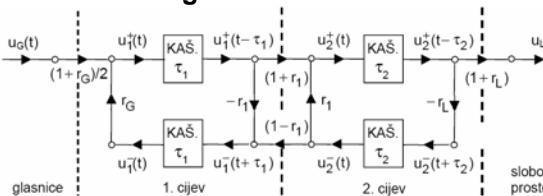


Model sa spojenim cijevima

- Na spojevima cijevi dolazi do djelomičnog propuštanja i djelomične refleksije zvučnog vala.
- Refleksija je to jača, što je razlika u presjecima veća.
- Koeficijent refleksije:
 $r_k = (A_{k+1} - A_k) / (A_{k+1} + A_k)$
- Koeficijenti refleksije određuju prijenosnu funkciju modela

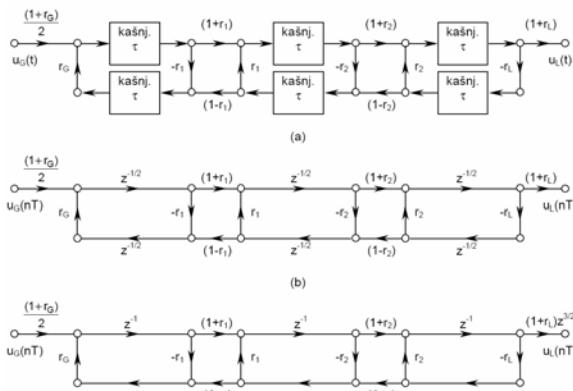
Prijenosna funkcija i blok dijagram modela sa cijevima

- Prijenosna funkcija vremenski kontinuiranog modela:



$$V_a(s) = \frac{0.5(1+r_G)(1+r_L)(1+r_1) \cdot e^{-s(t_1+t_2)}}{1+r_1r_G \cdot e^{-s^2t_1} + r_1r_L \cdot e^{-s^2t_2} + r_Lr_G \cdot e^{-s^2(t_1+t_2)}}$$

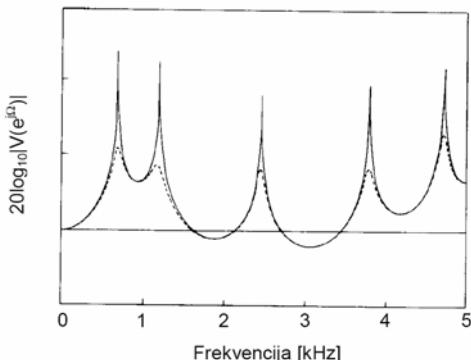
Vremenski diskretni model



Vremenski diskretni model

- Prijenosna funkcija vremenski diskretnog modela vokalnog trakta ima oblik običnog IIR filtra.
- Red filtra je jednak broju cijevi.
- Filtar ima samo polove.
- U brojniku je obično kašnjenje.
- Prijenosna funkcija ima oblik skupa rezonatora, koji se zovu FORMANTI.
- Svaki konjugirano kompleksni par kojeg opisuje jedan formant.

Dobivena prijenosa funkcija



Karakteristike formanata glasa "a"

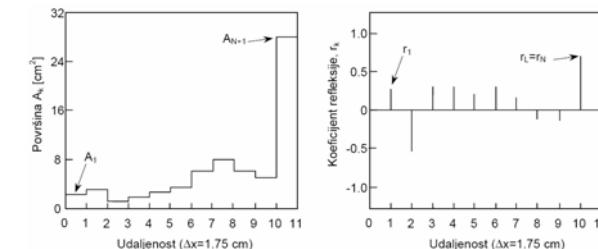
FORMANT	CENTRALNA FREKVENCIJA (Hz)	ŠIRINA FORMANTA (Hz)
prvi	650.3	94.1
drugi	1075.7	91.4
treći	2463.1	107.4
četvrti	3558.3	198.7
peta	4631.3	89.8

Model sa spojenim cijevima

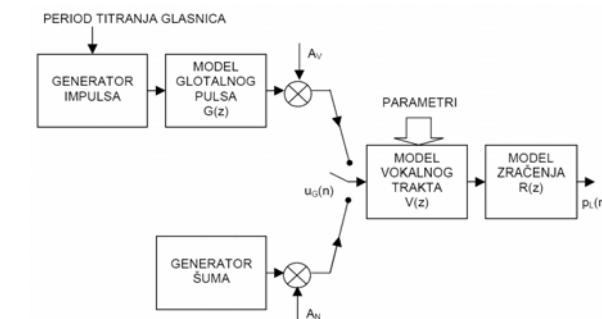
- U slučaju da su dužine svih cijevi jednake, tada je vrijeme širenja zvučnog vala kroz svaki segment jednako.
- Takov model se može opisati u vremenski diskretnoj domeni bez gubitka točnosti uslijed otiskivanja.
- Period otiskivanja se odabire da bude jednak dvostrukom vremenu širenja kroz jednu cijev.

Primjer modela sa spojenim cijevima

- Model sa 10 spojenih cijevi
- Model potpuno opisan koeficijentima refleksije u intervalu [-1, 1]



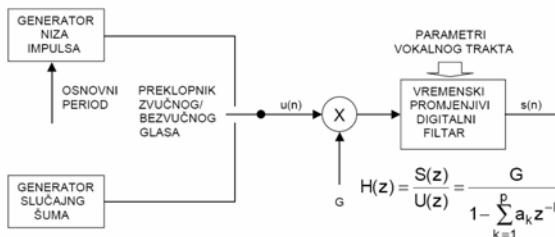
Vremenski diskretni model Izvor - Sustav



Vremenski diskretni model

Izvor - Sustav

- Utjecaj oblika glotalnog pulsa, vokalnog trakta i zračenja na usnicama se združuje u filtru $H(z)$



Određivanje parametara modela

- Kako odrediti model?
- Na raspolaganju nam je odziv modela, ali nemamo informaciju o pobudi (signal na glasnicama).
- Kako identificirati model i njegove parametre isključivo na osnovu odziva?
- Rješenje, ... pretpostavimo da je pobuda spektralno ravna (jedinični impuls ili bijeli šum).

Nedostaci opisanog modela

- problem promjene parametara govornog trakta kroz vrijeme ... to je jače izraženo kod eksplozivnih glasova (b p d t g k)
- ... u modelu se koriste diskretni sustavi bez nula, što nije dovoljno za nazalne glasove (m n), a stvara i manje probleme kod frikativa

Nedostaci opisanog modela

- metoda izmjene zvučni / bezvučni nije dovoljna za zvučne frikative (v z ž) ... moraju se koristiti složeniji modeli
- razmak pobudnih impulsa mora biti cijelobrojni višekratnik perioda otiskavanja T

Linearna predikcija

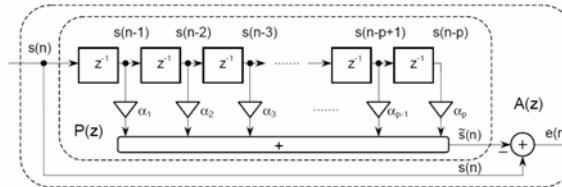
- Određivanje koeficijenata linearne prediktora:
- ... npr. autokorelacijskim postupkom

$$R_n(j) = \sum_{m=0}^{N-1-j} s_n(m)s_n(m+j)$$

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \dots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \dots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \dots & R_n(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \dots & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \vdots \\ R_n(p) \end{bmatrix}$$

Linearna predikcija

- Modeliranje trenutnih uzoraka signala na osnovu linearne kombinacije konačnog broja prethodnih uzoraka.
- Koeficijenti prediktora određuju težine pojedinih zakašnjelih uzoraka.

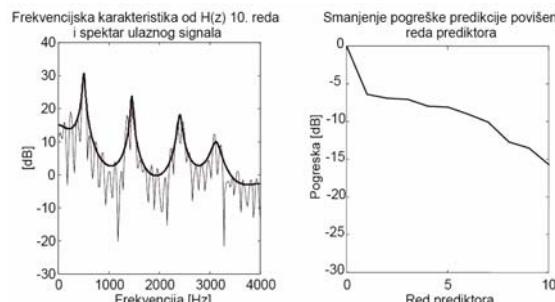


Linearna predikcija – LPC model

- Zadatak linearne predikcije:
 - Za zadani signal konačnog trajanja odrediti skup koeficijenata prediktora takav da je srednja kvadratna pogreška predikcije najmanja moguća.
- Ako ovaj postupak primjenimo na uzorcima izdvojenog segmenta govornog signala, ...
 - Dobivene koeficijente linearne prediktora postavljamo kao koeficijente nazivnika IIR filtra modela vokal. trakta.

LPC model

- Točnost modeliranja ovisi o redu prediktora

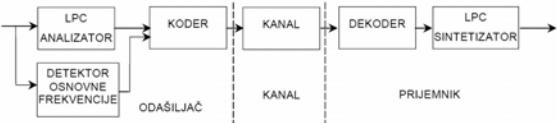


LPC model

- LPC model opisuje vremenski kratkotrajne korelacije u govornom signalu (1-2 ms)
- Imaju oblik istitravanja
- Određuje "gruba" spektralna svojstva govornog signala, tj. spektralnu ovojnicu
- Dugotrajne korelacije određene su svojstvima pobude (titranje glasnica), s periodom 3-15 ms

VOCODER

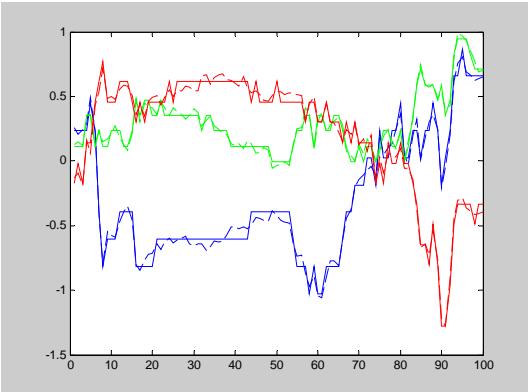
- Vocoder je najjednostavniji koder govornog signala.
- Pobuda se modelira isključivo kao niz impulsa ili bijeli šum
- Kodira se informacija o zvučnosti, periodu, energiji i sam LPC model



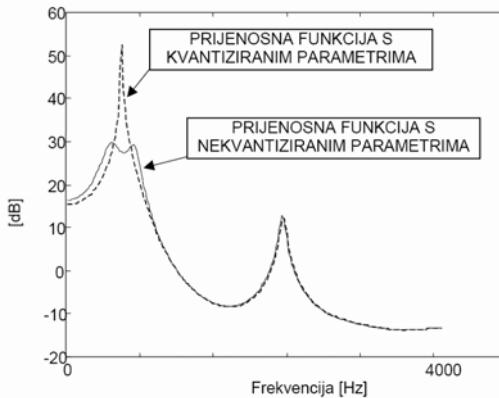
Kvantizacija LPC modela

- Red LPC modela je tipično 10 za frekvenciju otiskivanja 8kHz.
- Filter opisan sa 10 koeficijenata.
- Parametri filtra se određuju 50 do 100 puta u sekundi.
- Potrebno ih je kvantizirati i kodirati s čim moguće manjim brojem bita.
- Kvantizacija koeficijenata filtra uzrokuje promjenu frekvencijske karakteristike!

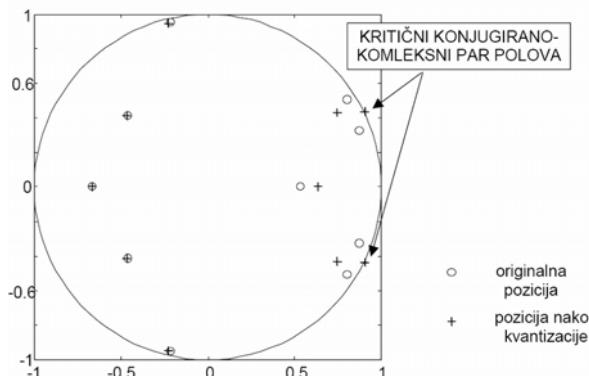
Kvantizacija koeficijenata modela



Kvantizacija LPC modela



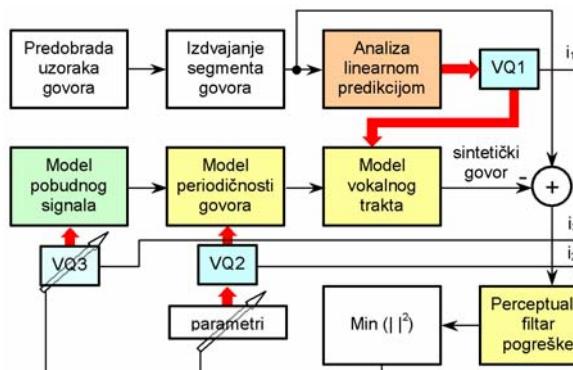
Utjecaj na položaj polova filtra



Općeniti ABS model kodera

- Kako poboljšati kvalitetu vokodera?
 - Potrebno je ipak malo preciznije opisati pobudni signal LPC modela.
 - Potrebno je ugraditi i perceptualna svojstva ljudskog slušnog sustava.
 - Dugotrajne korelacije je također pogodno opisati modelom.
- Parametri modela pobudnog signala i modela dugotrajnih korelacija se podešavanju u zatvorenoj petlji sinteze ... [Analysis-By-Synthesis](#)

Blok dijagram ABS kodera govora



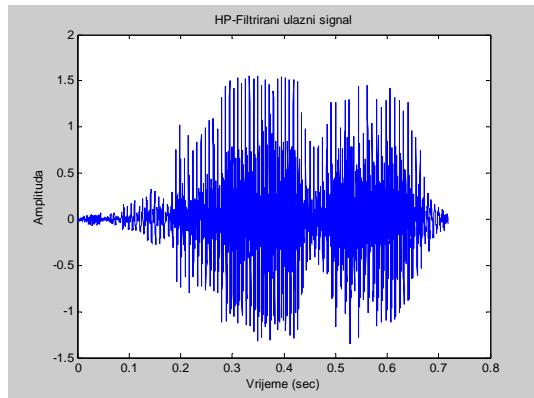
Kvantizacija LPC modela

- Razvijene brojne alternativne reprezentacije koeficijenata LPC filtra koje imaju bolja svojstva za kvantizaciju (k-koef., LAR, ArSin, LSP)
- Razvijeni postupci vektorske kvantizacije vektora LPC parametara
 - Izuzetno visoka učinkovitost.
 - Transparentno (nečujno) kodiranje uz svega 2 do 2.5 bita po koeficijentu.

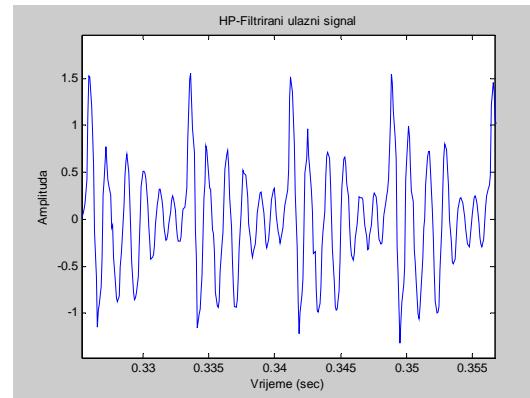
Određivanje periodičnosti pobude

- Model traži informaciju o periodu titranja glasnica – [Pitch analiza](#)
- Kako odrediti taj period?
 - Autokorelacija
 - Metode u spektralnoj domeni
 - [Kepstralna analiza](#)
- Kepstralna analiza je temeljena na homomorfnoj obradi signala
 - vrlo jednostavan postupak,
 - visoka učinkovitost uslijed FFT-a.

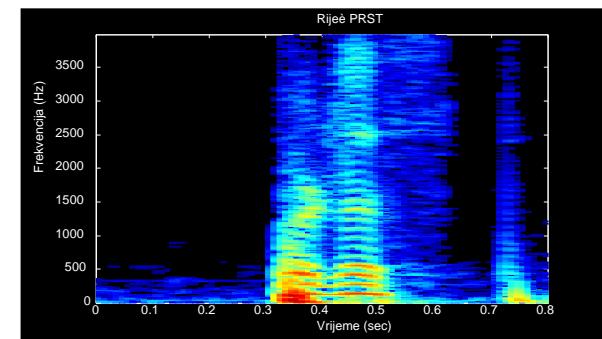
Izgovor riječi "glava"



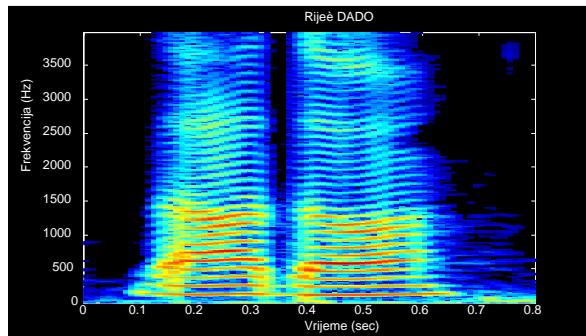
Detalj samoglasnika "a"



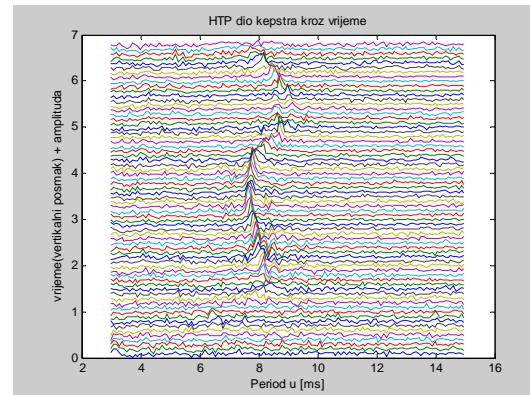
Spektrogram riječi "prst"



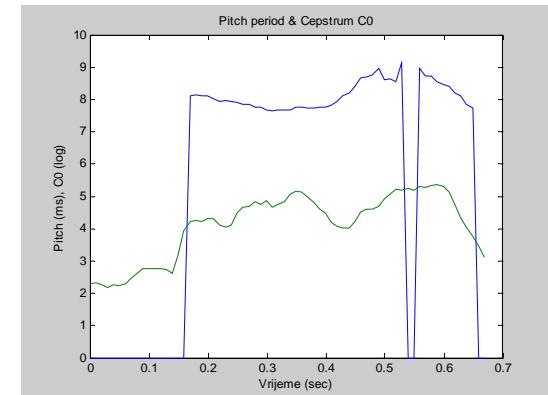
Spektrogram riječi "dado"



Pitch analiza primjenom kepstra

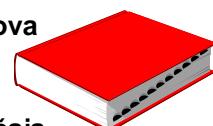


Period titranja glasnica i energija



Znanstvena područja koja se bave govorom

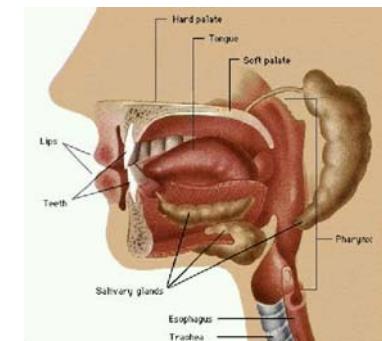
- Studija pravila jezika i njihovog utjecaja na ljudsku komunikaciju ... **LINGVISTIKA**
- Studija i klasifikacija glasova u govoru **FONETIKA**
- Poznavanje lingvistike i fonetike je od velikog značaja kod sinteze i prepoznavanja govora



Vokalni trakt - Anatomija

- prostor između glasnica, glottis
- pharynx ili ždrijelo (veza usta i jednjaka)
- usna šupljina (jezik, stražnje, srednje i prednje nepce, nadzubno meso, zubi i usne)
- velum ili resica zatvara usnu šupljinu prema nosnoj
- nosna šupljina koja završava s nosnicama

Presjek vokalnog trakta



... na formiranje glasa utječu

- položaj jezika
- položaj usana
- položaj čeljusti
- položaj resice
- otvor glasnica
- pritisak u plućima



.... dijele se i prema veličini otvora između jezika i nepca

najviši ili najotvoreniji	a
srednji	e o
najniži ili najzatvoreniji	i u

Klasifikacija fonema u hrvatskom jeziku

- Otvorni glasovi ili samoglasnici ili **vokali**
- Poluotvorni glasovi ili glasnici ili **sonanti**
- Zatvorni glasovi ili suglasnici ili **konsonanti**

Samoglasnici (vokali) ... svi su zvučni glasovi

prednji	i	e	prednji dio jezika i prednje nepce
srednji	a		sredina jezika
stražnji	o	u	stražnji dio jezika i stražnje nepce

Podjela glasnika i suglasnika prema mjestu tvorbe

dvostruchi ili bilabijalni	b p m	zapreka su obje usne
usnenozubni ili labiodentalni	f v	donja usna i gornji zubi
zubni ili dentalni	d t n c z s	zubi i jezik
prednjonepčani ili palatalni	j lj nj č ď dž đ š	jezik i prednje nepce
stražnjonepčani ili velarni	k g h	zadnji dio jezika i stražnje nepce
tekući ili likvidni	l r	vrh jezika dodiruje prednje nepce
nosni ili nazalni	m n	resica je spuštena
piskavi ili sibilantni	s z c	stvara se piskav šum

Poluotvorni glasovi ili glasnici ili sonanti

- j l lj m n nj r v
- uslijed približavanja ili dodirivanja pojedinih organa otvor za prolaz zraka se sužava ili djelomično zatvara
- svi glasnici su također zvučni glasovi
- r može biti i samoglasnik, a u povijesti hrvatskog jezika mogli su biti i drugi sonanti

Zatvorni glasovi ili suglasnici ili konsonanti

zvučni	b	d	g	dž	đ	-	v	z	ž	-
bezvučni	p	t	k	č	ć	c	f	s	š	h
				dž	d'z'					
				tš	t's'	ts				
zatvorni ili praskavi ili eksplozivni							složeni glasovi ili zatvorno-tjesnačni (afrikativi)		tjesnačni ili strujni ili frikativni	

Slogovi

- Spajanjem glasova dobivaju se **slogovi**
- **Slog je skup glasova koji se izgovara jednim izdisajem**

Količina ili kvantitet sloga

- vrijeme koje je potrebno da se slog izgovori
- određen je dužinom samoglasnika u slogu (muž / muževi)
- dodatni parametar koji se mora uzeti u obzir kod sinteze na osnovu fonema

Akcent pri izgovoru

- akcent je isticanje pojedinog samoglasnika izgovarajući ga većom snagom

kratkosilazni	brat, ginuti, govor, istina
dugosilazni	budim, dragi, glad, meso
kratkouz lazni	gora, loza, voda, pero
dugouz lazni	glava, hvaliti, pitati, trava