

**FAKULTET ELEKTROTEHNIKE
I RAČUNARSTVA**

**Zavod za elektroničke sustave
i obradbu informacija**

DIGITALNA OBRADA GOVORA
(Interna zavodska skripta)

Davor Petrinović

Zagreb, 08. 02. 2002

SADRŽAJ

1. UVOD.....	1
2. PREGLED PRIMJENA POSTUPAKA DIGITALNE OBRADJE GOVORA	2
2.1 KODIRANJE GOVORNOG SIGNALA.....	2
2.1.1 Prvi koderi govora temeljeni na reprezentaciji valnog oblika.....	2
2.1.2 Prvi koderi temeljeni na parametarskoj reprezentaciji govornog signala.....	2
2.1.3 Moderni koderi za primjene u mobilnim komunikacijama.....	3
2.1.4 Vojna primjena kodera govornog signala.....	4
2.1.5 Današnji trendovi u području kodiranja govora.....	4
2.1.6 Usporedba kvalitete poznatijih kodera govornog signala.....	4
2.2 SUSTAVI ZA TELEKONFERENCIJE I UDALJENO UČENJE.....	5
2.3 PRIMJENA DIGITALNE OBRADJE GOVORA U SINTEZI I PREPOZNAVANJU GOVORA.....	6
2.3.1 Automatsko prepoznavanje govora i detekcija riječi.....	6
2.3.2 Digitalna sinteza govora.....	7
2.3.3 Sustavi za dijalog s računalom.....	8
2.3.4 Prepoznavanje govornika.....	9
2.3.5 Automatsko prepoznavanje jezika.....	11
2.4 SUSTAVI ZA TRANSFORMACIJE GOVORNE PORUKE.....	11
2.5 SUSTAVI ZA POMOĆ LJUDIMA S TJELESNIM OŠTEĆENJIMA.....	13
2.6 ZAKLJUČAK.....	13
2.7 POTREBNA ZNANJA ZA DIGITALNU OBRADU GOVORA.....	14
3. OSNOVE PROCESA NASTAJANJA GOVORA.....	15
4. NAUKE O GOVORU I JEZIKU.....	17
4.1 PODJELA GLASOVA U HRVATSKOM JEZIKU.....	17
4.1.1 Samoglasnici.....	17
4.1.2 Glasnici.....	18
4.1.3 Suglasnici.....	18
4.1.4 Mjesto tvorbe glasnika i suglasnika.....	19
4.1.5 Preostala fonetska svojstva hrvatskog jezika.....	19
4.1.6 Složeniji fonetski elementi.....	20
5. AKUSTIČKA TEORIJA NASTAJANJA GOVORA.....	21
5.1 ŠIRENJE ZVUKA.....	21
5.2 JEDNOLIKA CIJEV BEZ GUBITAKA.....	22
5.3 UTJECAJ GUBITAKA U VOKALNOM TRAKTU.....	27
5.4 UTJECAJ ZRAČENJA NA USNICAMA.....	32
5.5 POBUDNI SIGNAL VOKALNOG TRAKTA.....	35
6. AKUSTIČKI GOVORNI MODEL.....	39
6.1 MODEL IZVOR-SUSTAV ZA FORMIRANJE GOVORNOG SIGNALA.....	39
6.2 MODEL SA SPOJENIM CIJEVIMA BEZ GUBITAKA.....	39
6.3 ŠIRENJE ZVUČNIH VALOVA U SPOJENIM CIJEVIMA.....	40
6.4 RUBNI UVJETI MODELA SA SPOJENIM CIJEVIMA BEZ GUBITAKA.....	43
7. VREMENSKI DISKRETNI AKUSTIČKI MODEL.....	46
7.1 VEZA AKUSTIČKOG MODELA I VREMENSKI DISKRETNIH FILTERA.....	46
7.2 PRIJENOSNA FUNKCIJA VREMENSKI DISKRETNOG MODELA U Z-DOMENI.....	50
8. DIGITALNI MODEL GOVORNOG SIGNALA.....	55
8.1 MODEL IZVOR-SUSTAV.....	55
8.2 VOKALNI TRAKT.....	56
8.3 ZRAČENJE NA USNICAMA.....	58
8.4 POBUDNI SIGNAL.....	59
8.5 POTPUNI DIGITALNI MODEL.....	61

8.6 ZAKLJUČAK	62
9. POSTUPCI LINEARNE PREDIKCIJE GOVORNIH SIGNALA	63
9.1 UVOD U LINEARNU PREDIKCIJU	63
9.2 PRIMJER PREDIKCIJE GOVORNOG SIGNALA	64
9.3 TEORETSKE POSTAVKE ANALIZE LINEARNOM PREDIKCIJOM	66
9.4 AUTOKORELACIJSKA METODA	72
9.5 METODA KOVARIJANCE	81
9.6 FAKTOR POJAČANJA LPC FILTRA	83
9.7 POSTUPCI RJEŠAVANJA SUSTAVA LPC JEDNADŽBI.....	85
9.7.1 Durbinov rekurzivni postupak za rješavanje autokorelacijskih jednadžbi.....	86
9.7.2 Primjer izračunavanja prediktora Durbinovim algoritmom.....	88
9.7.3 Računanje prediktora za metodu kovarijance Cholesky dekompozicijom	89
9.8 PREDIKCIJA MREŽASTOM (LATTICE) STRUKTUROM	93
9.8.1 Burgov algoritam	98
9.9 USPOREDBA POSTUPAKA LINEARNE PREDIKCIJE	99
9.10 SIGNAL PREDIKCIJSKE POGREŠKE.....	103
9.11 IZRAZI ZA PREDIKCIJSKU POGREŠKU	107
9.12 OVISNOST PREDIKCIJSKOG DOBITKA O TIPU GLASA.....	108
9.13 ALTERNATIVNI SKUPOVI KOEFICIJENATA ZA DEFINIRANJE PREDIKTORA.....	111
9.13.1 Odnos parametara linearnog prediktora i modela s cijevima bez gubitaka	111
9.13.2 Odnos koeficijenata linearnog prediktora i PARCOR koeficijenata.....	113
9.13.3 LAR koeficijenti.....	114
9.14 KVANTIZACIJA PARAMETARA PREDIKTORA	114
9.15 FREKVENCIJSKA INTERPRETACIJA GREŠKE PREDIKCIJE	119
9.15.1 Interpretacija srednje kvadratne pogreške predikcije u frekvencijskoj domeni.....	120
10. HOMOMORFNA OBRADA GOVORNOG SIGNALA	124
10.1 UVOD.....	124
10.2 SUSTAVI HOMOMORFNI ZA KONVOLUCIJU.....	124
10.3 SVOJSTVA KOMPLEKSNOG KEPSTRA	128
10.4 POSTUPCI IZRAČUNAVANJA KEPSTRA.....	131
10.4.1 Kepstar i kompleksni kepstar govornog signala	134
10.5 PRIMJENA KEPSTRALNE ANALIZE ZA ODREĐIVANJE FUNDAMENTALNE FREKVENCije	150

PREDGOVOR

Ova skripta je namijenjena studentima predmeta "Digitalna obrada govora" u svrhu lakšeg savladavanja gradiva i praćenja laboratorijskih vježbi. Sadržaj skripte pokriva uvod u područje digitalne obrade govora i daje fizikalne osnove postupka formiranja govora kao i osnovne akustičke i digitalne matematičke modele za reprezentaciju govornog signala. Velika pažnja je posvećena postupku linearne predikcije, koji je jedan od glavnih alata u analizi i modeliranju govora. U uvodnom dijelu skripte su opisane i specifičnosti hrvatskog jezika, te su dane neke osnovne klasifikacije glasova. Za potpuno razumijevanje izložene problematike, potreban je i praktični rad kroz laboratorijske vježbe iz ovog predmeta koje su pokrivena posebnom skriptom.

Gradivo ove skripte pripremljeno je kroz seminarske radove studenata šk. god. 1998/99 na osnovu odabranih poglavlja iz knjige "*Digital Processing of Speech Signals*", L.R. Rabiner, R.W. Schafer, 1978, koja se koristi kao službeni udžbenik iz ovog predmeta na većem broju sveučilišta u svijetu. Iako je ovo jedan od prvih udžbenika iz ovog područja, način izlaganja je dobro prilagođen znanjima studenata stečenim na predmetima "Teorija mreža i linija", "Sustavi i signali" i "Digitalna obrada signala". Velik dio materijala u ovoj skripti potiče iz navedene knjige, pa se radi toga ova skripta može koristiti isključivo kao interna zavodska skripta namijenjena studentima koji su upisali ovaj predmet, dok je svako njeno umnažanje ili distribucija u cjelini ili dijelovima zabranjeno. Osnovni materijali su prošireni većim brojem primjera i ilustracija, kojima se nastoje naglasiti najvažnije postavke izložene teorije. Obzirom da ova skripta pokriva samo jedan dio tematike izložene u knjizi, zainteresiranim studentima se preporuča da svoja znanja prošire čitanjem ove i drugih knjiga iz ovog područja.

1. UVOD

Govor je signal koji nosi informaciju (akustički valni oblik), čija je osnovna namjena komunikacija. Obzirom da je govor najprirodniji način komunikacije između ljudskih jedinki, njegov značaj nije umanjen ni u doba potpune informatizacije svih aspekata ljudskog života. Makar većina nas to nije svjesna, mnoštvo sustava koje nas danas okružuje temeljeno je na suvremenim govornim tehnologijama koje u sebi sadrže algoritme digitalne obrade govora. Glavne primjene postupaka digitalne obrade govora su u slijedećim područjima:

- kodiranje govora u svrhu učinkovitijeg prijenosa ili pohrane,
- telekonferencijski sustavi i udaljeno učenje
- prepoznavanje govora, diktiranog ili prirodnog s vezanim izgovorom riječi,
- prepoznavanje govornika,
- prepoznavanje jezika,
- sustavi za detekciju riječi,
- sinteza govora,
- sustavi za dijalog čovjeka i računala,
- sustavi za raznovrsne transformacije govornog signala kao što su promjena visine gasa, brzine izgovora itd.,
- sustavi za pomoć osobama oštećena vida, sluha ili govora,
- sustavi za poboljšanje kvalitete govora,
- analiza govora u svrhu dijagnostike raznovrsnih oboljenja.

U narednim poglavljima biti će dan kratak pregled većine gore navedenih primjena, kako bi se stekao osjećaj o širini ovog područja, kao i o trenutnom stanju napretka. Zbog složenosti većine navedenih sustava, kao i zbog njihove ovisnosti o konkretnim jezicima i narječjima, oni su i dan danas predmet intenzivnog istraživanja, kako u akademskoj zajednici, tako i u industriji i malom poduzetništvu.

2. PREGLED PRIMJENA POSTUPAKA DIGITALNE OBRADE GOVORA

2.1 Kodiranje govornog signala

Uobičajeno se u laičkom razmišljanju pod pojmom kodiranja podrazumijevaju sustavi koji provode određene transformacije na signalu u svrhu njegove zaštite, tj. da nitko osim osobe koja ima pripadni dekoder nije u stanju poruku vratiti nazad u originalni oblik. Međutim pojam kodiranja govora ima u stvari drugo značenje, tj. radi se o postupku pretvorbe analognog govornog signala u digitalni oblik, koji je u današnje vrijeme mnogo podesniji kako za pohranu, tako i za prijenos. Naravno, čim je signal transformiran u digitalnu domenu, nad tim nizovima digitalnih podataka se uistinu i može primijeniti bilo kakav postupak kriptografske zaštite. Međutim uobičajen postupak kodiranja i dekodiranja govora se ne bavi problemom kriptografije, već se isključivo bavi čim učinkovitijem postupkom digitalne reprezentacije. Zbog najšire primjene, kodiranje govora predstavlja jedno od najznačajnijih područja digitalne obrade govora. U nastavku će biti ukratko nabrojani sustavi za kodiranje govora kroz njihov povijesni razvoj.

2.1.1 Prvi koderi govora temeljeni na reprezentaciji valnog oblika

Jedan od glavnih pokretača razvoja digitalne obrade govora bili su javni telekomunikacijski sustavi koji su 70-tih godina postajali sve glomazniji i složeniji. Klasični pristup temeljen na analognom prijenosu govornog signala više nije mogao zadovoljiti rastuće potrebe, pa je digitalizacija sustava telekomunikacija bila nužna. Digitalizacija je prvo provedena u telefonskim centralama (javnim i kućnim), a postepeno su i analogni telefoni zamjenjivani s digitalnim. U to vrijeme digitalizacija govornog signala bila je temeljena na reprezentaciji valnog oblika, a poznati standardi kodiranja su tzv. **PCM** i **ADPCM**, s potrebnim brzinama prijenosa od 64 kbit/s za PCM, odnosno 32 kbit/s za osnovni mod rada ADPCM koder. Ti standardi se vode pod oznakama **G.711 (PCM)** i **G.721, G.723, G.726, G.727 (ADPCM)** i standardizirani su od međunarodnog tijela za standardizaciju telekomunikacijskog sektora **ITU-T** (*The Telecommunication Standardization Sector of the International Telecommunication Union*) Bitna značajka tih sustava kodiranja je u tome da su to algoritmi koji nastoje valni oblik govornog signala čim je moguće bolje reprezentirati i prenijeti na prijemnu stranu u svrhu vjerne reprodukcije. U tom smislu, kod tih koderi nije specijalno korištena činjenica da se radi o govornom signalu koji ima vrlo specifična svojstva, već se ti koderi kao takvi mogu koristiti i za prijenos drugih signala (npr. muzike).

2.1.2 Prvi koderi temeljeni na parametarskoj reprezentaciji govornog signala

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, pa je zahtjev za učinkovitim sažimanjem bio od presudne važnosti. Drugi značajan problem mobilnih digitalnih komunikacija jest nepouzdanost i mala kvaliteta prijenosnog kanala, tj. velika vjerojatnost pogreški u prijenosu, kao i povremeni totalni prekidi kanala. Radi toga, predloženi standardi su morali biti imuni na takove probleme u prijenosu. Nažalost, umjesto jednog svjetskog standarda, razvijena su tri najznačajnija sustava mobilne digitalne telefonije: sjeverno američki standard **IS54 VSELP** standardiziran 1989 od tijela **TIA** (*Telecommunication Industry Association*), japanski standard **JDC-VSELP** standardiziran od strane **RCR** (*Research and Development Center for Radio Systems*) pod oznakom **RCR STD-27B** i europski standard **GSM** temeljen na

RPE-LTP koderu standardiziranom 1987 od strane *Groupe Special Mobile of CEPT*. Svi ti sustavi su već bili posebno prilagođeni govornom signalu, tj. visoka učinkovitost sažimanja ostvarena je upravo na račun činjenice da govorni signal u sebi sadrži popriličnu količinu redundantne informacije. Ako se 'bitna' informacija razdvoji od 'nebitne', te ako se 'nebitni' dio opiše modelom, a 'bitni' kvantizira, kodira i prenese na prijemnu stranu, moguće je ostvariti istu kvalitetu reprodukcije uz mnogo manju brzinu prijenosa digitalne informacije. Potrebna brzina prijenosa tih sustava jest: 7.95 kbit/s za IS54 i 13 kbit/s za GSM, dok je kvaliteta samo neznatno niža od one ostvarive klasičnim PCM i ADPCM sustavima.

2.1.3 Moderni koderi za primjene u mobilnim komunikacijama

Dodatni razvoj u području učinkovitog kodiranja govornog signala početkom 90-tih godina, rezultirao je razvojem niza novih standarda, kojima se uz očuvanje iste kvalitete, potrebna brzina prijenosa smanjuje s faktorom dva, ili se pak uz istu brzinu prijenosa ostvaruje veća kvaliteta. Ti su koderi poznati pod nazivima "*Half-rate*" odnosno "*Enhanced full rate*" koderi. Tako je 1994 predložena zamjena originalnog GSM koder sa half-rate GSM standardom oznake **ETSI-TCH-HS** i brzine prijenosa od 5.6kbit/s. Isto tako originalni IS54 standard zamijenjen je s novim sjeverno-američkim standardom **IS96 QCELP** u okviru novog standarda mobilne telefonije temeljene na **CDMA** pristupu (*Code Division Multiple Access*). Za razliku od originalnog VSELP koder, novi QCELP koder ima skokovito promjenljivu brzinu prijenosa (0.8, 2, 4 ili 8.5 kbit/s) zavisno o sadržaju signala koji se kodira (govor ili pauza). Taj standard međutim nije u potpunosti ispunio očekivanja, pogotovo kada je u ulaznom govornom signalu bila prisutna značajna količina pozadinskog šuma. Za novi japanski standard **JDC Half-Rate** odabran je **PSI-CELP** koder brzine prijenosa od 3.45 kbit/s i kvalitete usporedive s full-rate standardom. Treba napomenuti da je dodatna učinkovitost sažimanja ostvarena na račun značajnog povećanja kompleksnosti kodiranja.

U okviru ITU-T organizacije, predložen je također čitav niz novih koder. Tako je u svrhu zamjene zastarjelih PCM i ADPCM standarda, 1992 i 1994 predložen novi standard **G.728** brzine prijenosa 16kbit/s temeljen na **LD-CELP** koderu malog kašnjenja. Isto tako, kao novi standard za mobilnu komunikaciju, je 1996 godine predložen standard **G.729** temeljen na **CS-ACELP** koderu brzine prijenosa od 8 kbit/s. Kao dio standardizacije video-telefona, koji pored govora prenašaju i sliku, 1995 je predložen novi standard za kodiranje govora oznake **G.723** koji ima varijabilnu brzinu prijenosa (5.3 ili 6.3 kbit/s zavisno o signalu). Kod svih do sada spomenutih koder, frekvencijski pojas govornog signala koji je bio kodiran i prenašan na prijemnu stranu jest približno od 200Hz do 3.2kHz. Taj pojas je od presudnog značaja za razumljivost govorne poruke, no kvaliteta signala je ipak značajno narušena tako uskim frekvencijskim pojasom. Radi toga, drugi trend u standardizaciji novih koder ide u smjeru povećanja kvalitete proširenjem pojasa. Ti koderi su poznati pod nazivom "*Wide-band*" koderi. Tako je ITU-T predložio nov standard oznake **G.722** kod kojeg je pojas signala proširen na 50Hz – 7 kHz.

Još jedan sustav temeljen na digitalnom kodiranju govora je sustav satelitskih komunikacija za vezu s brodovima na pučini **Inmarsat-M** (*International Maritime Satellite Corporation*). Za potrebe tog sustava 1990 predložen je koder **IMBE** tipa potrebne brzine prijenosa od samo 4.15 kbit/s. Za razliku od većine prije spomenutih koder koji su temeljeni na CELP shemi (*Code Excited Linear Predictor*), IMBE koder je temeljen na sinusoidalnoj reprezentaciji govornog signala.

2.1.4 Vojna primjena kodera govornog signala

Razvoj svih ovih sustava i novih kodera vjerojatno ne bi bio toliko brz, da u pozadini civilne primjene digitalnih govornih komunikacija nije bio i vojni aspekt, kao jedan od glavnih izvora financiranja većine navedenih razvojnih projekata. Američka vlada i ministarstvo obrane još je vrlo davno prepoznalo velik značaj digitalne reprezentacije govornog signala, što pored pojednostavljenja njegovog prijenosa omogućava i njegovo šifriranje u svrhu zaštite informacije. Tako je već 1970 započet rad na standardizaciji kodera oznake **FS1015** koji je bio temeljen na klasičnoj "*LPC Vocoder*" shemi, brzine prijenosa od 2.4 kbit/s. Ovaj koder poznat je i pod oznakom **LPC-10**. Standardiziran je konačno 1984 od strane **DoD** (*Department of Defense*) i kasnije NATO-a. Treba svakako napomenuti, da za razliku od prije opisivanih kodera kod kojih se nastojala manje ili više doseći kvaliteta klasičnih analognih telefonskih veza, kod ovog kodera jedina težnja je bila na razumljivosti. Radi toga reproducirani govorni signal je poprilično neprirodan (zvuči sintetički). Brzina prijenosa od 2.4 kbit/s bila je određena brzinom tadašnjih modem uređaja korištenih za prijenos digitalne informacije. Razvojem novih modema koji su radili na brzini od 4.8 kbit/s, javila se potreba za novim standardom kod kojeg bi se riješio navedeni problem male kvalitete originalnog kodera. Tako je 1991 predložen novi standard **FS1016** temeljen na **CELP** koderu brzine prijenosa od 4.8 kbit/s. Po kvaliteti, ovaj koder je i dalje nešto lošiji od GSM ili IS54 kodera, ali ima značajno nižu brzinu prijenosa, pogotovo u usporedbi sa GSM koderom.

Sredinom devedesetih, odlučeno je da se predloži novi standard za brzine prijenosa od 2.4kbit/s koji bi trebao u potpunosti zamijeniti FS1015 i FS1016. Kao cilj, traženo je da novi standard ima kvalitetu barem jednaku onoj ostvarivoj s FS1016 koderom koji radi na 4.8 kbit/s, a sa posebnim naglaskom na robusnost na pogreške u prijenosu i pozadinsku buku koja u vojnim primjenama može biti vrlo značajna (oklopno vozilo, tenk, lovački avion itd.). Kao novi standard odabran je koder **MELP** tipa.

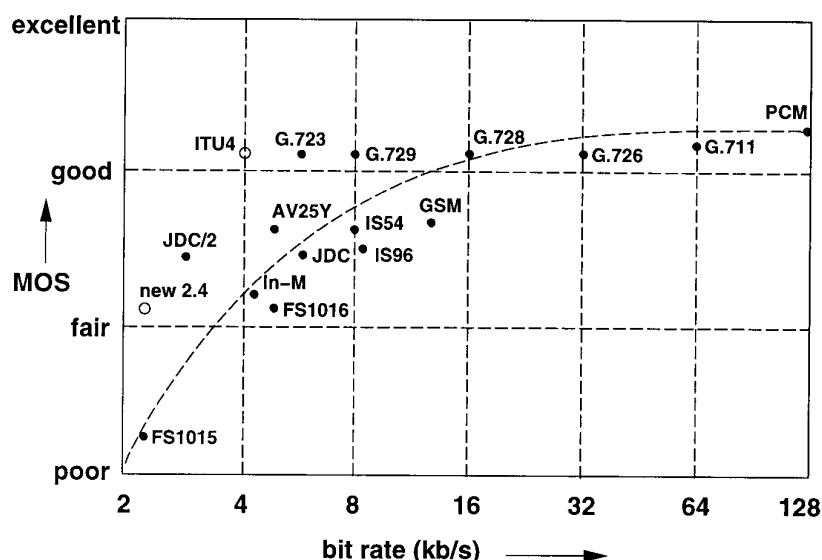
2.1.5 Današnji trendovi u području kodiranja govora

I pored činjenice što se kod današnjih kodera faktori sažimanja približavaju teoretskom maksimumu, i dalje su istraživanja u ovom području vrlo intenzivna. Jedan od važnih aspekata kod stvarne primjene govornih tehnologija jest i složenost algoritma, tj. problem vezan uz potrebnu procesnu moć procesora na kojem će se provoditi dotične obrade. Taj problem je direktno vezan i s problemom utroška električke energije, jer je ta proporcionalna potrebnoj procesnoj moći. Radi toga, kod baterijski napajanih prijenosnih uređaja, kod kojih je problem potrošnje također vrlo značajan, velika pažnja se posvećuje projektiranju učinkovitih algoritama kodiranja govora, koji osiguravaju pogodan kompromis između sažimanja i složenosti.

Razvoj algoritama za kodiranje govora se i dalje nastavlja, a dodatno je potaknut i potrebom za konačnom definicijom i implementacijom novog standarda, koji bi služio kao osnova za mobilne sustave tzv. treće generacije.

2.1.6 Usporedba kvalitete poznatijih kodera govornog signala

Na slici 2.1-1 su prikazani usporedni rezultati testiranja subjektivne kvalitete svih navedenih kodera u idealnim uvjetima (čisti govorni signal, bez pogreški u prijenosu). Kao mjerilo subjektivne kvalitete korištena je tzv. **MOS** mjera (*Mean Opinion Score*), koja ocjenjuje kvalitetu reproduciranog signala sa: "*poor*"-loše, "*fair*"-prihvatljivo, "*good*"-dobro, "*excellent*"-izvrsno. Na x-osi je označena potrebna brzina prijenosa u logaritamskom mjerilu.



Slika 2.1-1 Usporedba subjektivne kvalitete većeg broja kodera govora

2.2 Sustavi za telekonferencije i udaljeno učenje

Jednu specifičnu primjenu algoritama za kodiranje govora predstavljaju sustavi za telekonferencije kao i sustavi za udaljeno učenje (engl. *teleconferencing and distance learning*). U tu grupu možemo ubrojiti i sustave za prijenos govornog signala putem Internet-a (engl. *Voice over IP*) u svrhu zamjene ili nadopune klasičnih telekomunikacijskih sustava. Ono što je zajedničko za sve gore navedene sustave je činjenica da se govorni signal pretvara u digitalni oblik primjenom postupka kodiranja govora, te se zatim ta digitalna informacija prebacuje kroz digitalnu mrežu koja može biti bilo **ISDN** mreža u slučaju klasičnih telekonferencijskih sustava, ili pak **Internet** ili **ATM** računalna mreža u slučaju korištenja suvremenijih telekonferencijskih sustava temeljenih na PC tehnologiji i računalnim mrežama. Postupci kodiranja govora i slike, te razmjene digitalne informacije kod takvih sustava su propisani standardima, što osigurava inter-operabilnost, tj. primjenu raznolikih krajnjih uređaja proizvedenih od širokog spektra proizvođača, bilo da se radi o samostalnim uređajima ili pak programskim rješenjima temeljenim na osobnom računalu s pripadnom multimedijском podrškom (kamera, zvučna kartica, ISDN codec, mrežni priključak itd.). Pojava osobnog računala iznimno visokih procesnih mogućnosti je takve sustave iz domene skupih i specijaliziranih uređaja prebacila u domenu svakodnevne upotrebe kako u uredima tako i u kući. Potrebna infrastruktura za ostvarenje telekonferencije ili učenja na daljinu je postala minimalna, te je kao takva postala dostupna običnim korisnicima ograničenog budžeta. Istraživanja u ovom području su vrlo intenzivna, jer pored samog tehničkog problema, potrebno je nastavne materijale prilagoditi takvom načinu predavanja, a pored toga vrlo je značajno metodiku nastave prilagoditi novom mediju.

2.3 Primjena digitalne obrade govora u sintezi i prepoznavanju govora

Druge dvije također vrlo važne grane područja digitalne obrade govora predstavljaju automatsko prepoznavanje govora i sinteza govora. Ta dva područja su vrlo vezana, jer im je osnovna namjena prirodan način komunikacije čovjeka i računala. Specifičnosti obje primjene biti će detaljno objašnjenje u nastavku.

2.3.1 Automatsko prepoznavanje govora i detekcija riječi

Namjena automatskog prepoznavanja govora jest da govornu poruku izrečenu od strane ljudskog govornika pretvori u tekstualni oblik. U nekim slučajevima, radi se o komandama koje računalo mora prepoznati i zavisno o komandni izvršiti određenu akciju. Te su komande najčešće izolirane riječi iz relativno malog skupa komandi (npr. 100 različitih riječi). U drugom slučaju, računalo mora zamijeniti tajnicu, tj. mora prepoznati i pohraniti riječi koje mu diktira korisnik. Ako su riječi izolirane prilikom diktata, tada se to svodi na problem prepoznavanja izoliranih riječi. U ovom slučaju je skup riječi mnogo širi nego kod komandi, pa je i složenost sustava za prepoznavanje veća. Najsloženiji problem prepoznavanja govora predstavlja prepoznavanja spojenog (prirodnog) govora s riječima iz praktički neograničenih rječnika. Jedan od velikih problema sustava za prepoznavanje jest i zavisnost o govorniku. Od određenih sustava se traži da budu neosjetljivi na govornika, tj. da za sve govornike rade jednako dobro, dok se drugi pak adaptiraju za svakog pojedinog govornika. Namjena adaptacije je podešavanje pohranjenih statističkih modela načinu izgovora dotičnog govornika. Svi takvi sustavi su direktno vezani uz jezik koji se koristi, jer univerzalni sustavi koji bi radili za bilo koji jezik nisu niti približno tako dobri kao oni koji su projektirani za svaki jezik nezavisno. Razlog leži u činjenici što se učinkovitost prepoznavanja postiže ugradnjom fonetičkih i lingvističkih pravila u sustav prepoznavanja, a koja su naravno različita za svaki jezik. I pored svih navedenih problema, današnji sustavi za engleski jezik trenirani za dotičnog korisnika mogu postići točnost prepoznavanja od 95% za prirodni (vezani) izgovor s riječima iz vrlo velikih rječnika. Slično kao i kod kodiranja, riječ je o vrlo složenim algoritmima, tako da su mogućnosti sustava za prepoznavanje ovisne o procesnim mogućnostima sklopovske platforme. Prilikom razvoja novih verzija procesora opće namjene iz Intel-ove porodice, projektanti imaju u vidu potrebe za visokom procesnom moći algoritama za prepoznavanje govora, tako da su određene sklopovske pod-jedinice novih procesora projektirane upravo sa specijalnom zadaćom učinkovite izvedbe takvih algoritama. Danas postoje već brojni komercijalni programi za prepoznavanja govora za PC osobna računala. Isto tako, sustavi za prepoznavanje govora se ugrađuju u velik broj krajnjih produkata. Tako npr. jedna od budućih masovnih primjena prepoznavanja govora su "pametni" mobilni telefoni koji prepoznaju govorne komande, čime se omogućava rad s telefonom bez ruku. Umjesto da se prilikom uspostavljanja veze ručno bira broj, dovoljno je reći ime osobe koju se naziva i telefon će automatski prepoznati o kome se radi te iz memorije pročitati unaprijed pohranjeni telefonski broj dotične osobe i nazvati ga. Ako taj broj još ne postoji u imeniku, biranje se može provesti izgovorom niza željenih znamenki telefonskog broja. Drugi primjer masovne primjene su "pametni" automobili, kod kojih se upravljanje i komunikacija s ugrađenom elektroničkom opremom također provodi korištenjem glasovnih komandi. Takva elektronička oprema može obuhvaćati klasičnu opremu kao što su FM radio, CD, kazetofon, klima-uređaj, grijanje, ventilacija, kao i raznovrsni električki podesivi sustavi (prozori, zrcala, sjedala, rasvjeta itd.). Pored takvih klasičnih sustava, u novije vrijeme se u aute ugrađuju sustavi automatske satelitske navigacije koji omogućavaju nalaženje puta do cilja, zatim sustavi telefonske / Internet veze koji auto

pretvaraju u mobilni ured, kao i svi mogući sustavi elektronike za zabavu (TV, DVD, video igre itd.). Obzirom da upravljanje tako složenim sustavima zahtijeva i odvlači previše pažnje vozaču automobila, čija je ipak glavna zadaća sama vožnja, primjena postupaka upravljanja govorom je jedino smisljeno rješenje.

Jedna specifična primjena sustava za automatsko prepoznavanje su sustavi za detekciju riječi. Od takvih sustava se ne traži kompletna transkripcija razgovora, već isključivo detekcija određenih ključnih riječi u razgovoru. Najznačajnija primjena takvih sustava je u praćenju velikog broja telefonskih linija u svrhu nadzora. Da bi se olakšao posao ljudskim operaterima, računalo simultano prati velik broj linija i "čeka" pojavu ključnih riječi na bilo kojoj od aktivnih linija. U slučaju da se takve riječi pojave, snimljeni razgovor na toj liniji se prosljeđuje ljudskom operateru na definitivnu provjeru sadržaja. U slučaju da se očekuje točno određena osoba u razgovoru, tada se takvi sustavi mogu kombinirati i s automatskim prepoznavanjem govornika, čime se može dodatno povećati vjerojatnost sigurne detekcije "interesantnog" razgovora. Takvi sustavi su danas vrlo značajni, jer klasični postupci praćenja u modernim komunikacijskim sustavima više nisu primjenjivi, prvenstveno zbog enormnog broja korisnika koji istovremeno koriste takve sustave. Sustavi za detekciju se uobičajeno projektiraju za relativno mali skup ključnih riječi, jer moraju istovremeno osigurati i vrlo veliku vjerojatnost detekcije pravih riječi, kao i vrlo malu vjerojatnost lažne aktivacije tokom razgovora.

2.3.2 Digitalna sinteza govora

Naravno da je u svrhu prirodne komunikacije s računalom nužan i drugi smjer, tj. računalo mora biti sposobno da bilo kakvu informaciju koju želi prenijeti korisniku može pretvoriti u govornu poruku. I ovo područje se vrlo intenzivno razvijalo zadnjih 20-30 godina, tako da je danas već u prilično zreloj fazi. Kod najnovijih sustava za sintezu govora, prirodnost izgovora je toliko dobra da je već vrlo teško razlučiti da li se radi o živoj osobi ili o računalu. Ulogu sustava za automatsku sintezu govora najlakše je usporediti s zadaćom spikera na TV ili radio stanici, koji napisani tekst mora pročitati na pravilni način, tj. korištenjem svih pravila izgovora, naglasaka, intonacije itd. Važno je naglasiti da ljudski čitač prilikom čitanja provodi i interpretaciju teksta, tj. napisana poruka na papiru neće biti pročitana doslovce nego će pojedine oznake brojke ili kratice biti pretvorene u riječi. Tako npr. pisana poruka oblika: *"Jučer, 29.10.01 u 13:22 izbio je požar u Petrinjskoj ul. br. 22"*, biti će izgovorena na slijedeći način: *"Jučer, dvadeset devetog listopada dvije tisuće i prve godine u trinaest sati i dvadeset i dvije minute izbio je požar u Petrinjskoj ulici na kućnom broju dvadeset i dva"*. Usporedbom ove dvije verzije iste poruke, očito je koliko je složena zadaća automatskog sustava za sintezu. Programi za sintezu starijih generacija taj bi gore navedeni tekst izgovorili na slijedeći način: *"Jučer zarez dvadeset i devet točka deset točka nula jedan točka u trinaest dvotočka dvadeset i dva izbio je požar u Petrinjskoj ul točka br točka dvadeset i dva"*. Naravno da bi takav način sinteze bio vrlo naporan za slušača, a ponekad bi čak i smisao poruke bio pogrešno prenesen. Interpretacija (razumijevanje smisla poruke) vrlo je važna i za intonaciju. Pravilna intonacija zahtijeva točno prepoznavanje strukture rečenice, tj. što je imenica, pridjev, prilog, glagol, gdje je početak, gdje kraj itd. Očito da je u tu svrhu fonetička i lingvistička pravila pojedinog jezika potrebno opisati na matematički egzaktni način, čime se omogućava implementacija takovih pravila u sustavima za sintezu i prepoznavanje. Zbog tih činjenica je u posljednje vrijeme značajno promijenjen pristup i u tim znanstvenim područjima, te se sve više pažnje posvećuje "računalnoj" fonetici i lingvistici. Sustavi za automatsku sintezu se danas susreću u nizu primjena, od komercijalnih programa za PC računala, pa sve do dječjih igračaka koje danas sve manje ili više nešto pričaju.

2.3.3 Sustavi za dijalog s računalom

Objedinjavanjem sustava za prepoznavanje i sustava za sintezu, te dodavanjem sustava za umjetnu inteligenciju i baze podataka, dobivaju se sustavi za dijalog, koji također postaju jako popularni u novije vrijeme. Krajnji cilj takvih sustava je da u potpunosti zamijene ljudskog operatera u raznovrsnim uslugama informacija, kao što su npr. informacije o redu vožnje ili letenja na kolodvorima ili aerodromima, informacije o tel. brojevima, ili bilo kakve usluge korisnicima koje se nude od tzv. 'call-centara' velikih firmi (npr. servis, održavanje itd.). Jedna od interesantnih primjena je informacijski pult, gdje bi turist u stranom gradu razgovorom s računalom koje se nalazi "na cesti" saznao bilo kakvu potrebnu informaciju (od smještaja, restorana, kulturnih i povijesnih znamenitosti, snalaženja po gradu itd.). Zadatke koje mora obaviti takav sustav za dijalog se mogu svrstati u slijedeće glavne cjeline:

- snimanje izgovorene rečenice ili upita korisnika
- automatsko prepoznavanje (transkripcija govorne poruke u tekstualni zapis)
- određivanje smisla snimljene rečenice i detekcija svih ključnih riječi i informacija
- provjera kompletnosti upita
- zahtjev za upotpunjenje nedostajućih informacija
- prema potrebi provodi se i verifikacija smisla upita sa konačnom potvrdom tipa DA/NE
- pretraživanje baze podataka i određivanje traženog odgovora
- konstrukcija rečenice s odgovorom
- sinteza odgovora u govornu poruku
- reprodukcija odgovora

Obavljanje određene transakcije najčešće zahtijeva višestruke prolaze kroz gore opisane korake, tako dugo dok se kroz dijalog ne prikupe sve potrebne informacije da bi računalo moglo dati suvisao odgovor. Iz gornje diskusije je vidljivo da se radi o vrlo složenim sustavima, čije se znanje i inteligencija formiraju postupcima "učenja ili treninga". U tu svrhu se koriste govorne baze sa uobičajenim razgovorima za konačnu primjenu. Tako npr. ako se radi o projektiranju sustava za automatske informacije o redu letenja aviona, tada se prilikom učenja sustava moraju koristiti snimke razgovora stvarnih korisnika i živih operatera upravo te tematike. Iz tih snimki se zatim određuje skup riječi korišten u konverzacijama, te se sustav za automatsko prepoznavanje trenira da korektno prepozna samo riječi iz tog skupa. Iako to djeluje malo neobično, u takvim usko tematski specijaliziranim razgovorima fond riječi je prilično mali (par tisuća ili manje). Iz svih prikupljenih transkripcija razgovora potrebno je odrediti i tipove upita upućene od različitih korisnika, jer se ista stvar može upitati na velik broj različitih načina. Objedinjavanjem svih tih podataka formiraju se statistički modeli koji se koriste za prepoznavanje svih elemenata dijaloga :

- osnovnih fonemskih grupa,
- riječi,
- i rečenica

Treniranje sustava, tj. određivanje parametara tih statističkih modela, provodi se direktno na osnovu podataka određenih iz govorne baze za učenje. Pri svemu tome treba voditi računa i o činjenici da jedan određeni postotak riječi i rečenica ipak neće biti obuhvaćen modelima. Sustav mora biti robustan na takve slučajeve i tražiti od korisnika da preformulira pitanje, ako se postavljeni upit ne uklapa niti u jedan postojeći model. Za kvalitetno treniranje sustava potrebna je velika količina snimljenog govornog materijala (stotine sati), jer su čak i tada neke rijetke riječi ili neuobičajeni tipovi upita nedovoljno zastupljeni za pouzdano učenje

statističkih modela. U svrhu vrednovanja takvih automatskih sustava uobičajeno se koristi kvocijent prosječnog trajanja transakcije obavljene s živim operaterom u odnosu na trajanje transakcije obavljene s računalom. I kod najboljih sustava taj kvocijent je i dalje u korist živih operatera. Velik problem je i u tome što se korisnici ponašaju potpuno drugačije kada shvate da je s druge strane veze računalno, pa i baza razgovora s živim operaterima nije dobar reprezentant stvarnih dijaloga koji će se pojaviti u konverzaciji s računalom. Radi toga modeli se uobičajeno moraju ponovno trenirati sa stvarnim razgovorima snimljenim nakon prve primjene takvih sustava. Ipak, zbog male pouzdanosti, takvi sustavi se danas koriste u eksperimentalne svrhe, prvenstveno za primjene gdje konačni rezultat (odgovor sustava) ipak nije od presudne važnosti i neće uzrokovati značajnu štetu ili probleme korisniku u slučaju da odgovor nije ono što je on tražio. Druga mogućnost je da se u slučajevima kada nije moguće ostvariti konstruktivni dijalog korisnika s računalom, da se tada on prosloji na živog operatera.

U slučaju kada se sustavi za dijalog primjenjuju u informacijskim pultovima, tada se oni obično kombiniraju i s jednom dodatnom tehnologijom, a to je tzv. "agent". Smisao agenta je da pored audio komunikacije, postoji i vizualna komunikacija čovjeka i računala, tj. video kamera snima korisnika i interpretira njegove izraze lica, dok se na ekranu računala animira lik virtualne osobe "računala" s kojom se provodi komunikacija. Animacija takvog virtualnog lika mora biti u skladu sa sustavom za sintezu govora, tako da se dobije sinkronizirani audio-vizualni doživljaj koji u potpunosti odgovara stvarnom razgovoru. Pokazalo se da sa takvim pristupom olakšava i ubrzava razgovor, jer pored funkcionalnosti takav sustav je zabavan i lakše prihvatljiv za korisnika.

2.3.4 Prepoznavanje govornika

Jedno specifično područje digitalne obrade govora se bavi problemom automatskog prepoznavanja govornika. Tipične primjene su sustavi koji moraju na osnovu izgovora točno određene (ili pak bilo kakve) govorne sekvence automatski odrediti o kojem govorniku iz konačnog skupa govornika se radi. Za svakog potencijalnog kandidata iz tog konačnog skupa govornika izračunava se vjerojatnost da je snimljena govorna sekvenca izgovorena upravo od strane tog govornika. Sortiranjem takve liste vjerojatnosti dobivaju se najvjerojatniji kandidati za prepoznavanje. Sustav radi dobro, ako je vjerojatnost pravog govornika mnogo veća od vjerojatnosti bilo kojeg drugog krivog govornika iz te baze (tj. na toj listi). Nekad, zbog inherentne sličnosti glasa dva različita govornika takva diskriminacija nije moguća, nego se oba javljaju na vrhu liste kao najbolji potencijalni kandidati s vrlo bliskim vjerojatnostima prepoznavanja. U takvim slučajevima zamjene su moguće i neizbježne, jer su objektivno karakteri ta dva glasa vrlo slični. Razlikovanje između govornika treba prvenstveno temeljiti na onim parametrima koji su fizikalno uvjetovani, tj. koji su određeni biofizikalnim značajkama govornika (oblik i duljina vokalnog trakta), a ne na onim parametrima koji se mogu vježbom lažno učiniti sličnim originalnom govorniku (kao što to rade profesionalni imitatori). Također treba voditi računa o pouzdanoj negativnoj detekciji glasova govornika koji se ne nalaze u bazi autoriziranih korisnika. Svi takvi glasovi moraju biti klasificirani kao "nepoznati", tj. prilikom usporedbe tog glasa s glasovima autoriziranih kandidata u bazi, vjerojatnost svakog kandidata iz baze mora biti ispod traženog minimalnog praga potrebnog za detekciju autoriziranog govornika. Pouzdanost isključenja "nepoznatih" govornika može se osigurati na više načina:

- samo autorizirani korisnik poznaje rečenicu koju mora izgovoriti, i/ili
- ta rečenica (*password*) se mijenja prilikom svakog ulaza prema nekom unaprijed dogovorenom pravilu koje znaju samo autorizirani korisnici, i/ili

- radi jednostavnosti, traženu rečenicu računalo ispisuje na terminalu, te korisnik mora izgovoriti baš tu rečenicu, čime se eliminira mogućnost korištenja ilegalno unaprijed snimljene rečenice pravog govornika, ili pak uvježbavanje pravilne imitacije već korištenog *password*-a.

Slični mehanizmi se koriste i kod konvencionalnih autorizacija korištenjem *login/password*-a na računalu. Treba voditi računa i o činjenici da se glas istog govornika mijenja sa starenjem. Te su promjene vrlo izražene u djetinjstvu, kao i u starijoj dobi kada dolazi do značajnih promjena biofizičkih parametara. Isto tako, do promjene glasa može doći uslijed bolesti kao što su prehlade, upale grla itd. Sustavi za identifikaciju moraju biti projektirani tako da uzimaju u obzir i takve modifikacije glasa, jer se u protivnom statistički modeli moraju nanovo trenirati na novi glas. Osnovni principi koji se koriste kod prepoznavanja govornika su vrlo slični postupcima prepoznavanja govora i temeljeni su na neovisnim statističkim modelima za svakog govornika, čiji se parametri određuju postupcima treninga na osnovi višestrukih izgovora različitih sekvenci koje se koriste prilikom autorizacije. Najveća pouzdanost sustava se postiže ako su prilikom treniranja statističkih modela korištene iste rečenice koje će biti korištene i prilikom prepoznavanja. U slučaju da se prilikom autorizacije koristi proizvoljna govorna sekvenca, tada je problem automatskog prepoznavanja govornika mnogo složeniji.

Takvi sustavi se mogu koristiti kao jedan od modaliteta provjere identiteta prilikom ulaza u osiguranu zonu, kombinirajući ih sa drugim modalitetima identifikacije kao što su: unos šifre na tastaturi, identifikacija primjenom chip-kartice, otisak prsta, slika retine, digitalizirani potpis, itd. Maksimalna pouzdanost identifikacije postiže se isključivo kombinacijom odabranih ili svih navedenih modaliteta.

Druga popularna primjena sustava za automatsko prepoznavanje govornika jest u programima za automatsko prepoznavanje govora. Kao što je prije diskutirano, kvaliteta prepoznavanja govora se može značajno uvećati, ako se statistički modeli riječi ili fonetskih grupa adaptiraju (podese) na način izgovora upravo tog konkretnog korisnika. Radi toga, takvi programi za svakog potencijalnog korisnika vode konfiguracijsku datoteku u kojoj su pohranjeni parametri specifični za tog govornika. Na početku korištenja programa za automatsko prepoznavanje govora, potrebno je računalu dati do znanja o kojem korisniku se radi, da bi program koristio pripadajuću konfiguracijsku datoteku. Taj odabir govornika može biti proveden manualno (odabirom opcije u izborniku programa), ili pak što je mnogo atraktivnije automatski, tako da program samostalno prepozna o kojem govorniku se radi, te na osnovu toga odabere pripadne parametre za automatsko prepoznavanje govora.

Treća vrlo značajna primjena je u forenzičke svrhe, tj. kao dokaz identiteta u sudskim sporovima. Isto tako, kao što je već diskutirano u poglavlju o sustavima za detekciju riječi, automatsko prepoznavanje govornika se može koristiti u sustavima za praćenje razgovora u svrhu detekcije "interesantnog" materijala.

Zadnja primjena koju ćemo spomenuti je u sustavima za automatsku transkripciju govornih emisija ili snimljenih razgovora u kojima je sudjelovao veći broj govornika. U tom slučaju pored prepoznavanja govora, potrebno je odrediti koji dio teksta pripada pojedinom govorniku u tom snimljenom materijalu.

Valja napomenuti, da su sve navedene aplikacije vrlo zahtjevne i naporne ako ih moraju provoditi živi operateri, pa je stoga značaj automatskih sustava utoliko veći. Međutim, zbog konačnih mogućnosti i pouzdanosti takvih automatskih sustava, oni se često koriste samo kao alat, koji smanjuje količinu posla živom slušaču, dok se kritične odluke ipak moraju prepustiti ljudskom uhu.

2.3.5 Automatsko prepoznavanje jezika

Jedna od novijih primjena digitalne obrade govora je u sustavima za automatsko prepoznavanje jezika, na osnovu snimljene govorne sekvence nepoznatog sadržaja. Takvi sustavi se integriraju u sustave za automatski dijalog koji mogu provoditi razgovor s potencijalnim korisnikom na bilo kojem jeziku, tj. na materinjem jeziku tog korisnika. Obzirom da sustavi za dijalog izrazito ovise o korištenom jeziku i obzirom da su direktno projektirani i trenirani za svaki pojedini jezik, univerzalni sustav za dijalog se u biti sastoji od N paralelnih sustava od kojih je svaki projektiran za jedan specifični jezik. Da bi bilo moguće odabrati jedan od tih sustava potrebna je povratna veza od strane korisnika, koji na neki način mora sustavu dati do znanja koji jezik bi želio koristiti u dijalogu. Do sada se to obavljalo direktnim odabirom iz nekog izbornika, ili pak ako se radi o sustavu implementiranom na klasičnoj telefonskoj infrastrukturi, tada se odabir obavljao DTMF signalima, tj. biranjem na tastaturi telefona. Novi pristup detekciji jezika temeljen je na direktnom prepoznavanju jezika na osnovu glasa korisnika. U tom postupku se pokušavaju oponašati mogućnosti ljudskog slušača, koji je kadar prepoznati jezik govornika i bez da poznaje riječi dotičnog jezika. To znači da čovjek može i bez razumijevanja što govori dotični govornik, prepoznati da li se radi o ovom ili onom jeziku. Pri tome se koristimo značajkama kao što su:

- fonetske karakteristike jezika (glasovi koji se javljaju u govoru)
- melodičnost i način izgovora (intonacija)

Analogno sustavima za automatsko prepoznavanje govora i govornika, moguće je sa svaki jezik formirati statističke modele koji opisuju akustičke značajke govora na tom jeziku. Cilj je ostvariti sustave koji su u mogućnosti odrediti jezik sugovornika, na osnovu što kraće govorne sekvence. Brzina detekcije jezika je značajna jer je detekciju potrebno obaviti prije početka samog dijaloga.

Takvi sustavi detekcije jezika, koji su neovisni o samoj govornoj poruci, mogu se kombinirati sa rezultatima prepoznavanja govora da bi se povećala njihova učinkovitost. To se radi tako da se snimljena govorna poruka koja se koristi za detekciju jezika propusti kroz svih N sustava za prepoznavanje govora, tj. da se prepoznavanje govora provede uz hipotezu svakog od N potencijalnih jezika. Obzirom da se prilikom prepoznavanja dobivaju vjerojatnosti svake izgovorene riječi, odnosno rečenice, moguće je za svaku hipotezu jezika dobiti vjerojatnosti prepoznate govorne sekvence. Onaj jezik koji rezultira najvećom vjerojatnošću prepoznavanja govorne poruke jest vjerojatno stvarni jezik korišten od strane govornika.

2.4 Sustavi za transformacije govorne poruke

Jedno posebno područje digitalne obrade govora predstavlja primjena u sustavima za transformacije govorne poruke. Tipične transformacije govora mogu biti:

- promjena spola govornika (muško u ženski i obratno)
- promjena visine glasa
- promjena brzine izgovora (ubrzavanje ili usporavanje)
- poboljšanje kvalitete govora (npr. uklanjanje signala smetnje ili šuma)

U nastavku će biti ukratko opisane tipične primjene takvih sustava. Prva primjena je kod ronioca s bocama za disanje. U slučaju kada se roni na većim dubinama, tada se pored komprimiranog zraka u bocama nalazi i plin helij, koji sprječava komplikacije koje mogu nastupiti pri ronjenju uz veliki tlak. Međutim, helij značajno mijenja akustička svojstva medija (fluida) u vokalnom traktu, što uzrokuje velike promjene u glasu (glas postaje visok,

... nešto kao Pajo patak). Da bi se olakšala komunikacija ronioca moguće je uz primjenu sustava za transformaciju govora vratiti govor u normalni oblik.

Drugi primjer primjene je kod pilota lovačkih aviona, koji također dišu zrak pod pritiskom i koriste specijalna "aktivna" odjela koja vanjskim pritiskom kompenziraju povišeni tlak u plućima. Pritisak zraka i pritisak odijela se dinamički mijenja zavisno o trenutnim akceleracijama u avionu, da spriječi bježanje krvi iz glave u noge. Uslijed povišenog tlaka, mijenja se oblik vokalnog trakta, te producirani govor nema iste značajke kao i govor u normalnim uvjetima. I u ovom slučaju je moguće određenim transformacijama, vratiti takav izobličen govor u prirodni oblik, te poboljšati kvalitetu komunikacije.

Interesantne primjene su u sustavima za promjenu identiteta govornika, gdje je moguće zadržati sadržaj i brzinu izgovora govorne poruke, a kompletno promijeniti karakter glasa govornika. Takvi sustavi se koriste npr. prilikom svjedočenja zaštićenih svjedoka u sudskim sporovima, ili pak u raznim primjenama kod kojih se iz sigurnosnih razloga mora zaštititi pravi identitet govornika. Još jedna primjena takvih sustava je u zabavnoj industriji, kao npr. modifikacija glasa animiranog lika u crtanim filmovima, ili pak raznovrsne modifikacije glasa pjevača u muzičkoj produkciji.

U određenim slučajevima poželjno je zadržati istu boju glasa, ali se želi unaprijed snimljenu govornu poruku reproducirati većom ili manjom brzinom od stvarne. Veća brzina reprodukcije pogodna je kod pretraživanja većih govornih baza u svrhu nalaženja interesantnog materijala. Kod jeftinijih sustava reprodukcije audio-vizualnog zapisa (kao što su npr. klasični video-rekorderi), kod kojih ne postoje sustavi za transformaciju brzine reprodukcije audio zapisa, prilikom ubrzane reprodukcije se uobičajeno audio zapis ne reproducira, već isključivo samo slika. Ugradnjom takovog sustava za ubranu / usporenu reprodukciju govornog ili audio signala, moguće je ostvariti reprodukciju kompletnog audio-vizualnog zapisa u svim brzinama reprodukcije.

Zadnja grupa sustava za transformaciju govora su sustavi za poboljšanje kvalitete govora. Kvaliteta i razumljivost govora mogu biti ugroženi različitim izvorima smetnji ili izobličenja. Tipični izvori su:

- kvantizacijski šum uzrokovan kodiranjem govora
- okolni šum (buka) pribrojen govornom signalu kao smetnja
- drugi govornik ili govornici koji govore u pozadini
- revebracije (jeke) koje se javljaju prilikom snimanja u zatvorenim prostorijama s lošim akustičkim svojstvima
- jeke uzrokovane preslušavanjima u raznim točkama prijenosnog kanala
- pojava mikrofonijske (akustičke povratne veze) prilikom snimanja u prostorijama u kojima se signal s mikrofona reproducira na sustavu ozvučenja

Način otklanjanja svakog od navedenih izvora smetnji je specifičan za svaki od navedenih sustava, a provodi se specifičnim uređajima kao što su:

- uređaj za poništenje akustičke povratne veze (engl. *acoustic feedback canceller*)
- uređaj za poništenje jeke (engl. *echo canceller*)
- uređaj za povećanje kvalitete govora (engl. *speech enhancement device*)

Postupci i uređaji koji se primjenjuju za jedan dio gore navedenih transformacija su univerzalni za sve tipove audio signala, tj. nisu projektirani specifično za govor, dok je drugi dio učinkovit jedino u slučaju kada se primjenjuje na govorni signal.

2.5 Sustavi za pomoć ljudima s tjelesnim oštećenjima

Digitalna obrada govora primjenjuje se i u raznovrsnim sustavima za pomoć ljudima s tjelesnim oštećenjima. Tako npr. ljudi koji su nijemi mogu koristiti prijenosne uređaje za sintezu govora, kod kojih se željena poruka odabire i formira primjenom namjenske tipkovnice, te sintetizira i reproducira primjenom takvih uređaja. Kod osoba koje su gluhe, mogu se primjenjivati sustavi za automatsko prepoznavanje govora, koji prihvaćenu govornu poruku mogu prepoznati i prikazati u tekstualnom obliku na zaslonu prijenosnog uređaja.

Postoji velik problem u učenju govora kod osoba koje su gluhe od rođenja. U takvim slučajevima, potreban je vrlo veliki trud specijaliziranog osoblja koje radi s djecom, da se i pored činjenice da dijete ne čuje svoj vlastiti izgovor postigne zadovoljavajuća kvaliteta govora. U tu svrhu se danas koriste programi koji provode analizu izgovora, te vizualnim putem daju direktnu povratnu vezu govorniku kako poboljšati izgovor. Na ovaj način, svaka osoba može raditi sama i bez direktne pomoći specijaliziranog osoblja za obuku. Ovakvi programi su prilagođeni uzrastu korisnika, tako da učenje čine interesantnim i zabavnim, jer više liče na igru nego na program za učenje.

Kod slijepih osoba, sva vizualna komunikacija se mora obaviti zvučnim ili taktilnim putem. Da se takvim osobama omogući čitanje normalnih knjiga ili tiska, tekst u grafičkom obliku se skenira (digitalizira), te prevodi u tekstualni zapis primjenom programa za automatsko prepoznavanje teksta (engl. *optical character recognition*). Takav tekst se tada može reproducirati primjenom sustava za sintezu govora, čime se zapravo ostvaruje postupak automatskog čitanja otisnutog teksta.

Digitalna obrada govora se primjenjuje i u tzv. umjetnim pužnicama, tj. kod osoba koje su oštećenog sluha, ali im je živčani dio slušnog sustava zdrav i upotrebljiv. Naziv umjetna pužnica i nije najprimjereniji, jer se u stvari radi o direktnoj električkoj stimulaciji živčanih završetaka u prirodnoj pužnici primjenom specijalne elektrode s velikim broj izvoda koji stimuliraju živce duž cijele pužnice. Takvom električkom stimulacijom se može zamijeniti uloga uha kod kojeg se stimulacija tih živaca provodi akustičkim putem pomicanjem vrlo malih dlačica koje se nalaze unutar pužnice. Sustav za električku pobudu stimulatora je minijaturan i često odvojen od glavnog dijela sustava u kojem se provodi pretvorba govornog signala snimljenog mikrofonom u niz impulsa određenih amplituda i frekvencija koji se prosljeđuju do pojedinih izvoda elektrode. Danas je operativni zahvat ugradnje umjetne pužnice vrlo rasprostranjen u svijetu i brojnim osobama je primjenom takvih uređaja vraćen sluh.

Digitalna obrada govora se koristi i u sustavima za dijagnostiku raznovrsnih poremećaja sluha ili govora. Pri tome je primjenom određenih postupaka moguće točno odrediti mjesto i tip oboljenja, te pratiti razvoj bolesti, odnosno terapijski tijek.

2.6 Zaključak

Kroz ova uvodna poglavlja željelo se ilustrirati koliko je široko područje digitalne obrade govora i koliko su brojne primjene u kojima se primjenjuju postupci digitalne obrade. Jednako tako su široke i profesije u kojima su nužna znanja iz digitalne obrade govora, kao što su telekomunikacije (žične i bežične), digitalni radio i televizija, multimedijски sustavi, telekonferencijski sustavi i učenje na daljinu, sudstvo, forenzika, informacijske i obavještajne službe (tajne i javne), vojna primjena, primjena u medicini i defektologiji, itd. Zbog složenosti ukupne problematike, poznavanje svih opisanih primjena je praktički nemoguće, tako da će se istraživači u pravilu profilirati u jednom ili više relativno uskih područja obrade govora zavisno o vlastitim interesima i željama.

2.7 Potrebna znanja za digitalnu obradu govora

U svrhu lakšeg savladavanja gradiva predmeta digitalne obrade signala nužna su određena predznanja. Najbitnija znanja su iz područja signala i sustava, te digitalne obrade signala, tj. vremenski diskretnih sustava. Ta znanja obuhvaćaju osnovne pojmove iz modeliranja linearnih vremenski nepromjenjivih sustava, kao i pripadne transformacije signala i sustava. U slučaju kontinuiranih sustava tu se misli na reprezentaciju sustava korištenjem linearnih diferencijalnih jednadžbi u vremenskoj domeni, odnosno reprezentaciju signala i sustava u frekvencijskoj domeni primjenom Laplace-ove transformacije. U slučaju vremenski diskretnih signala i sustava, radi se o modeliranju pomoću jednadžbi diferencija, odnosno ekvivalentni modeli u frekvencijskoj domeni temeljeni na Z transformaciji. Specijalni slučaj Laplace-ove i Z transformacije su Fourier-ova transformacija, odnosno vremenski diskretna Fourier-ova transformacija koje opisuju frekvencijske karakteristike takvih sustava i signala u stacionarnom stanju. U okviru gradiva biti će opisivan akustički model vokalnog trakta temeljen na spojenim cijevima bez gubitaka. Takav model ima električki dual, koji je temeljeni na propagaciji signala duž električnih linija bez gubitaka. Radi toga, kompletna teorija koja se koristi u modeliranju i analizi električnih linija je direktno primjenjiva i u ovom području. Osnova znanja iz filtracije u vremenski diskretnoj domeni primjenom rekurzivnih i nerekurzivnih filtara su također vrlo značaja, jer su glavni digitalni modeli vokalnog trakta temeljeni na upravo takvim filtrima. Obzirom da će govorni signal biti modeliran i analiziran u obje domene (kontinuiranoj i vremenski diskretnoj), bitno je poznavati osnove postavke otipkavanja i rekonstrukcije signala.

Za primjene u automatskom prepoznavanju govora, govornika, jezika ili detekciji riječi, pored ovih osnovnih znanja koja su prvenstveno iz domene obrade signala, potreba su i znanja iz slučajnih procesa i sustava, te modeliranja takvih stohastičkih sustava. Jedan od glavnih alata u tom modeliranju su skriveni Markovljevi modeli (engl. *Hidden Markov Model*) temeljeni na diskretnim ili kontinuiranim opservacijama.

Područje digitalne obrade govora je posebno interesantno upravo zbog činjenice da kombinira praktične aspekte digitalne obrade signala sa matematičkim znanjima iz područja modeliranja slučajnih procesa i sustava. Ono što posebno stimulira istraživače u tom području je i činjenica što sva ta teorija ima i vrlo realnu primjenu u svakodnevnom životu.

3. OSNOVE PROCESA NASTAJANJA GOVORA

Prilikom komunikacije primjenom govora, poruka se u apstraktnom obliku pojavljuje u mozgu govornika. Ona se zatim pretvara u skup neuronskih signala koji upravljaju postupkom artikulacije. Micanje artikulatora (jezika, usnica, glasnica, itd.) formira zvučni signal koji sadrži željenu informaciju originalne poruke.

U fizikalnom procesu nastajanja govora sudjeluju pluća govornika koja se pod djelovanjem mišića prsnog koša stišću i potiskuju zrak kroz vokalni trakt. Vokalni trakt se u širem smislu sastoji od slijedećih osnovnih dijelova.

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

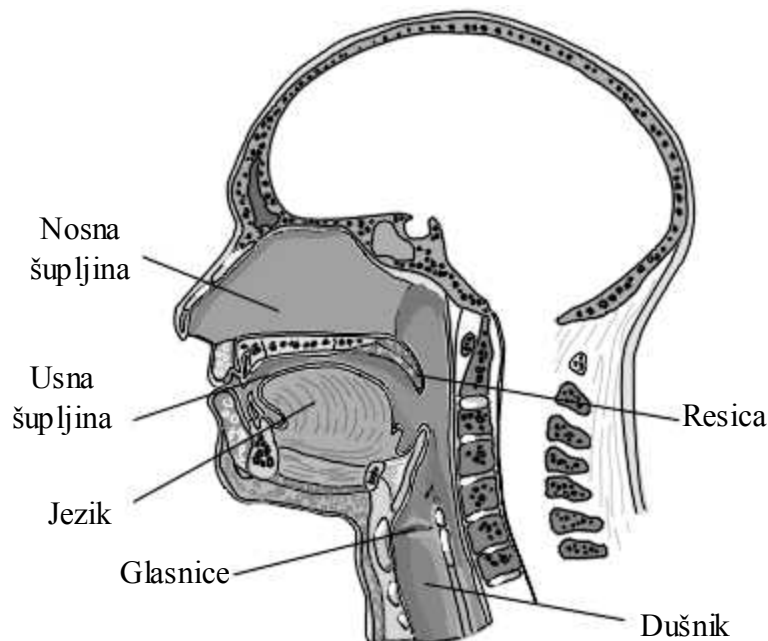
U užem smislu, vokalni trakt ne sadrži nosnu šupljinu (nazalni trakt), ali obzirom da i nazalni trakt sudjeluje u formiranju određene klase glasova i on će biti tretiran dijelom vokalnog trakta.

Kao što će u narednim poglavljima biti detaljnije diskutirano, glasnice su vrlo značajan organ u procesu formiranja govora. Glasnice se ponašaju kao mehanički oscilator (pisak), koji prelazi u stanje relaksacijskih oscilacija uslijed struje zraka iz pluća koja kroz njih prolazi. Na frekvenciju njihovog titranja utječu brojni parametri, a među najznačajnijim su pritisak zraka iz pluća na ulazu u glasnice i napetost samih glasnica, koju je moguće svjesno upravljati. Takvim periodičkim titranjem, glasnice formiraju periodičku struju zraka koja zatim prolazi kroz ostatak vokalnog trakta. U slučaju da su glasnice potpuno opuštene, neće doći do oscilacija i struja zraka iz pluća će neometano prolaziti kroz vokalni trakt.

Vokalni trakt se ponaša kao svojevrsan filter, koji će spektralno obojiti pobudni signal. Slično kao što se geometrijom cijevi kod orgulja određuje ton (visina i spektralni sastav) signala koji se formira, tako će i geometrijski oblik vokalnog trakta određivati koje se spektralne komponente signala pojačavaju, a koje prigušuju. Za razliku od orgulja kod kojih su cijevi brojne i fiksne geometrije (nezavisne za svaki ton), vokalni trakt je temeljen na jednoj cijevi, ali čiji se oblik mijenja zavisno o položaju artikulatora. Pored karakteristika pobudnog signala vokalnog trakta, na formiranje glasa utječu

- položaj jezika
- položaj usana
- položaj čeljusti
- položaj resice

Naravno da je zbog fizičkih ograničenja navedenih artikulatora, ukupni zbir glasova koje je moguće proizvesti vokalnim traktom ograničen. Isto tako brzina prijelaza artikulatora iz jednog stanja u drugo je također ograničena. Kod većine jezika, glasovi i riječi su u povijesti nastajale na način da budu jednostavno ostvarive (izgovorljive) uz navedena fizikalna ograničenja. Taj princip minimalnog napora je značajno utjecao na vjerojatnost pojedinih kombinacija glasova u riječima. Tako npr. pojavljivanje glasova "s" i "z" u nastavku jednog odmah iza drugog, "sz", je vrlo malo vjerojatno u hrvatskom jeziku, jer se taj par glasova vrlo teško izgovara u slijedu. Međutim ubacivanje samoglasnika "a" ili "u" između ova dva glasa značajno olakšava izgovor, pa se tako "saz" nalazi u riječi "sazrijeti" a "suz" u riječi "suza". S druge strane, glasovi su prilikom formiranja jezika odabirani tako da su perceptualno različiti, tj. da je ljudskom slušaču moguće bez velikog napora odrediti primljenu poruku, pa čak i kada su uvjeti prijenosa vrlo loši (npr. okolna buka, velika udaljenost itd.).



Slika 2.7-1 Presjek i osnovni dijelovi vokalnog trakta koji sudjeluju u produkciji govornog signala

4. NAUKE O GOVORU I JEZIKU

Proučavanje govora kao jednog od glavnih čimbenika ljudske komunikacije jest multidisciplinarni problem, te se različita područja znanosti bave pojedinim aspektima ovog problema. Tako se lingvistika bavi studijom pravila jezika i njihovog utjecaja na ljudsku komunikaciju, dok se fonetika bavi studijom i klasifikacijom glasova u govoru.

Poznavanje lingvistike i fonetike je od velikog značaja kod većeg broja primjena digitalne obrade govora, kao što su npr. sinteza i prepoznavanja govora.

4.1 Podjela glasova u hrvatskom jeziku

Na osnovu konfiguracije i otvora vokalnog trakta, glasovi u hrvatskom jeziku se dijele na tri osnovne grupe. To su:

- otvorni glasovi ili samoglasnici ili vokali,
- poluotvorni glasovi ili glasnici ili sonanti i
- zatvorni glasovi ili suglasnici ili konsonanti.

Sonanti i konsonanti se nekad u širem smislu ubrajaju u istu grupu, tj. u suglasnike.

4.1.1 Samoglasnici

Vokali ili samoglasnici čine prvu grupu i karakterizira ih činjenica da su to glasovi najveće energije, kod kojih je vokalni trakt većim dijelom otvoren, a tokom cijelog trajanja izgovora glasnice titraju. Osnovna namjena vokala jest povezivanje suglasnika u slogove, tj. formiranje izgovorljivih riječi. Vokali često sami za sebe nose malo informacije, tj. u većini slučajeva prepoznavanje govorne poruke je moguće i kada su iz riječi kompletno izbačeni svi samoglasnici. Tako npr. iz poruke: "dns sm bl škl", nije teško prepoznati da je originalna poruka bila: "danas sam (ili smo) bila (ili bili) (u) školi". Pri "dešifriranju" ove poruke, koristili smo se upravo statističkim informacijama o vjerojatnosti pojavljivanja određenih glasova u nizu, kao i lingvističkim pravilima formiranja rečenica. I pored toga što u rečenici ima čak osam samoglasnika (a a/o i a/i u o i), moguća su samo dva značenja ove rečenice. Postoje više podjela samih vokala. Prva podjela je po mjestu tvorbe, a prikazana je u tablici 4.1-1. Mjesto tvorbe je mjesto najvećeg suženja vokalnog trakta. Tako se npr. prilikom izgovora samoglasnika "u" stražnji dio jezika podiže i približava stražnjem (mekom) nepcu. Kao što će biti pokazano u narednim poglavljima, pozicija i oblik takvog suženja u vokalnom traktu direktno će određivati njegovu frekvencijsku karakteristiku a samim time i karakter glasa koji se proizvodi.

Tablica 4.1-1 Podjela samoglasnika po mjestu tvorbe

Tip glasa	Glas	Mjesto tvorbe
prednji	i e	prednji dio jezika i prednje nepce
srednji	a	sredina jezika
stražnji	o u	stražnji dio jezika i stražnje nepce

Duga podjela vokala je prema veličini otvora između jezika i nepca i prikazana je u tablici 4.1-2.

Tablica 4.1-2 Podjela samoglasnika po veličini otvora

Tip glasa	Glas
najviši ili najotvoreniji	a
srednji	e o
najniži ili najzatvoreniji	i u

4.1.2 Glasnici

Druga grupa glasova u hrvatskom jeziku su glasnici ili sonanti. Svi glasnici su također zvučni glasovi jer pri njihovom izgovoru glasnice titraju. Ovi glasovi su prikazani u tablici 4.1-3, a spadaju u grupu polu-otvornih glasova, jer se uslijed približavanja ili dodirivanja pojedinih organa (artikulatora) u vokalnom traktu, otvor za prolaz zraka sužava ili djelomično zatvara. Naravno da je u usporedbi s vokalima, taj preostali otvor značajno manji.

Tablica 4.1-3 Glasnici u hrvatskom jeziku

j	l	lj	m	n	nj	r	v
----------	----------	-----------	----------	----------	-----------	----------	----------

Interesantno je uočiti da sonant "r" može biti i samoglasnik, tj. u nekim riječima on preuzima ulogu povezivanja konsonanata u slogove, odnosno riječi (npr. riječ "prst"). U povijesti hrvatskog jezika istu ulogu su imali i neki drugi sonanti.

4.1.3 Suglasnici

Zadnja grupa glasova su zatvorni glasovi ili konsonanti. Kod tih glasova se prolaz zračnoj struji potpuno zatvara ili ostaje samo uski tjesnac kroz koji se ona tare. Neki od konsonanata su zvučni, dok su preostali bezvučni. Zavisno o tipu prepreke (suženja) u vokalnom traktu dijele se u tri osnovne grupe:

- zatvorni ili praskavi ili eksplozivni konsonanti
- tjesnačni ili strujni ili frikativni konsonanti
- složeni glasovi, ili zatvorno-tjesnačni, ili afrikati

U tablici 4.1-4 prikazani su svih 18 konsonanata podijeljeni po zvučnosti i tipu prepreke. Kod zatvornih konsonanata, struja zraka je potpuno zatvorena na određenom mjestu vokalnog trakta. Dolaskom zraka iz pluća, raste tlak iza navedene prepreke, te nakon što postigne dovoljnu razinu uzrokuje naglo otvaranje prepreke i formiranje impulsnog zvučnog signala. Za razliku od zatvornih, kod tjesnačnih konsonanata je vokalni trakt otvoren cijelo vrijeme, ali je prolaz vrlo uzak, te struja zraka prolazi kroz takav tjesnac vrlo velikom brzinom. Uslijed te velike brzine i uslijed trenja o stijenke tog tjesnaca, dolazi do turbulencija u zračnoj struji i do formiranja glasa "šumovitog" karaktera.

Zadnja grupa konsonanata su složeni konsonanti, tj. oni koji nastaju kao kombinacija zatvornih i tjesnačnih konsonanata. Tako npr. glas "č" je vrlo sličan istovremenom izgovoru glasova "t" i "š". Parovi glasova koji formiraju navedene složene konsonante prikazani su u trećem i četvrtom redu tablice 4.1-4.

Interesantno je uočiti da se konsonanti praktički uvijek javljaju u zvučno / bezvučnom paru. Tako su npr. oblik i vremenske promjene vokalnog trakta prilikom izgovora glasova "b" i "p" vrlo slične, dok je jedina bitna razlika u modu pobude, tj. kod "b" glasnice titraju, dok kod "p" ne. Isto vrijedi i za preostale parove prikazane u prva dva reda tablice 4.1-4.

Tablica 4.1-4 Parovi zvučnih i bezzvučnih suglasnika

tip glasa	zatvorni ili praskavi ili eksplozivni			složeni glasovi ili zatvorno-tijesnačni ili afrikati			tijesnačni ili strujni ili frikativni			
zvučni	b	d	g	dž	đ	-	v	z	ž	-
bezzvučni	p	t	k	č	ć	c	f	s	š	h
				dž	d'z'					
				tš	t's'	ts				

4.1.4 Mjesto tvorbe glasnika i suglasnika

Slično kao što je u tablici 4.1-1 napravljena podjela samoglasnika po mjestu tvorbe, to isto je napravljeno u tablici 4.1-5 za sve sonante i konsonante. Ova podjela nije ekskluzivna, tj. neki se glasovi javljaju u dvije kategorije. Tako je npr. glas "m" istovremeno dvousnjeni glas, kao i nazalni glas, jer prilikom prvog dijela izgovora glasa "m" struja zraka prolazi kroz nosni trakt i izlazi na nosnicama, dok su usnice potpuno zatvorene. Tek na kraju izgovora dolazi do naglog otvaranja usnica. Slično je i za glas "n" koji je istovremeno i zubni i nazalni glas.

Tablica 4.1-5 Podjela glasnika i suglasnika po mjestu tvorbe

Tip glasa	Glas	Mjesto zapreke (tvorbe)
dvousnjeni ili bilabijalni	b p m	zapreka su obje usne
usnnozubni 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 ć dž đ ž 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 spuštana
piskavi ili sibilantni	s z c	stvara se piskav šum

Kao što je vidljivo iz tablice, mjesto prepreke u vokalnom traktu može biti na raznim mjestima i između različitih organa (usnice, zubi, zubno meso, jezik, nepce).

4.1.5 Preostala fonetska svojstva hrvatskog jezika

Iako u hrvatskom jeziku ima samo 30 različitih glasova, u stvarnom govoru se pojavljuje daleko veći broj modifikacija istih. Način izgovora pojedinog glasa ovisi značajno o njegovom kontekstu, tj. o glasovima koji se nalaze lijevo i desno od njega. Ova pojava se

naziva koartikulacija. Zbog toga, kvalitetni sintetički govor nije moguće dobiti jednostavnim spajanjem 30 diskretnih glasova. Isto tako važno je naglasiti da prijelazi iz jednog glasa prema drugom nisu nagli (skokoviti), već su vrlo postepeni i određeni postepenim prijelazom artikulatora iz početne pozicije koja odgovara prvom glasu, prema novoj poziciji koja odgovara slijedećem glasu. Dakle pri tome vokalni trakt prelazi kroz niz među-stanja, što uzrokuje formiranje niza prijelaznih glasova relativno kratkog trajanja. Izbacivanje tih prijelaza značajno narušuje prirodnost sintetičkog govora. U svrhu rješavanja ovog problema, kod jednostavnijih sintetizatora govora, se kao osnovni elementi sinteze koriste parovi glasova (fonema) ili tzv. "difoni" koji su izvađeni (snimljeni) iz stvarnog govora. Na taj način se unutar tog osnovnog elementa nalaze i navedeni prijelazi između glasova.

Za razliku od hrvatskog jezika koji ima samo 30 fonema, u engleskom jeziku ih ima čak 42. Posebno je velika razlika u broju samoglasnika. Tako npr. dok u hrvatskom ima samo 5 samoglasnika, u engleskom ih je čak 11. Pored samih samoglasnika, postoje i tzv. "diftonzi" koji predstavljaju prijelazne samoglasnike, koji na početku izgovora odgovaraju jednom, a na kraju drugom samoglasniku. Tako npr. u riječi "bay" (zaljev), diftong započinje u samoglasniku "e" a završava sa "i", ili npr. u "boy" (dječak), započinje u "o", a završava sa "i". Naravno taj prijelaz je vrlo postepen, pa je to i razlog zašto se navedeni prijelazni glasovi tretiraju kao posebna kategorija fonema.

Sve navedene podjele i klasifikacije ne treba tretirati previše strogo, jer obzirom da govor nastaje kao rezultat prirodnog živog bića, a ne nekog stroja ili mehaničke naprave točno određenih i ponovljivih dimenzija, varijacije između pojedinih govornika, dijalekata i jezika su vrlo velike, što problem digitalne obrade govora čini posebno složenim.

U hrvatskom jeziku, govorna informacija je potpuno sadržana u zvučnosti pobudnog signala (glasnice titraju ili ne) i u frekvencijskoj karakteristici vokalnog trakta. Visina glasa ne predstavlja nikakvu informaciju, tj. neovisno o visini (osnovnoj frekvenciji) zvučnih glasova, poruka je uvijek ista. Međutim to nije istina za sve jezike, jer npr. u mandarinskom kineskom jeziku, smisao (značenje) riječi ovisi i o visini izgovora, što predstavlja dodatni problem kod sustava za automatsko prepoznavanje govora.

4.1.6 Složeniji fonetski elementi

Spajanjem glasova dobivaju se slogovi. Slog je skup glasova koji se izgovara jednim izdisajem. Količina ili kvantitet sloga određena je vremenom potrebnim da se slog izgovori. Naravno da do varijacija u trajanju sloga dolazi upravo varijacijama trajanja samoglasnika, dok su trajanja suglasnika većinom stalna. Tako je npr. trajanje samoglasnika "u" različito u riječima "muž" i "muževi".

Pored trajanja, drugi mogući način modifikacije izgovora pojedinog glasa u riječi je modifikacija amplitude. Tako se isticanjem pojedinog samoglasnika izgovarajući ga većom snagom formira akcent. U hrvatskom jeziku postoje četiri tipa akcenta, koja su navedena u tablici 4.1-6 zajedno s nekoliko primjera za svaki od tipova.

Tablica 4.1-6 Tip akcenta u hrvatskom jeziku

tip akcenta	primjer
kratkosilazni	brat, ginuti, govor, istina
dugosilazni	budim, dragi, glad, meso
kratkouzlazni	gora, loza, voda, pero
dugouzlazni	glava, hvaliti, pitati, trava

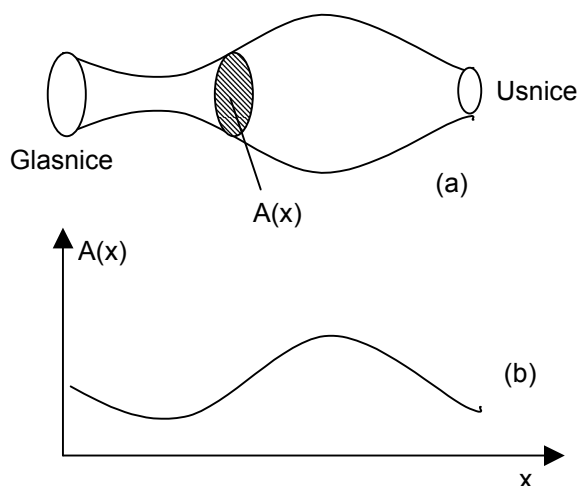
5. AKUSTIČKA TEORIJA NASTAJANJA GOVORA

5.1 Širenje zvuka

Pod pojmom zvuka se podrazumijevaju vibracije. Vibracije su uzrok nastanka zvučnih valova koji se nakon toga šire titranjem čestica medija kroz koji putuje. Zbog toga osnovu za opisivanje nastanka i širenja zvuka u vokalnog trakta čine zakoni fizike. To se posebno odnosi na osnovne zakone očuvanja mase, očuvanja količine gibanja i očuvanja energije zajedno sa zakonima termodinamike i mehanike fluida koji se svi mogu primijeniti na zrak. Zrak je osnovni medij širenja zvuka kod govora, a može se svrstati u grupu stlačivih fluida niske viskoznosti. Koristeći ove fizikalne principe moguće je postaviti skup parcijalnih diferencijalnih jednačbi koje opisuju gibanje zraka u govornom sustavu. Sama formulacija, a i rješenje ovih diferencijalnih jednačbi je iznimno složeno izuzev u slučaju velikih pojednostavljenja oblika vokalnog trakta i gubitaka energije u njemu. Detaljna akustička teorija mora uzeti u obzir slijedeće utjecaje:

- vremenska promjenjivost oblika vokalnog trakta
- gubici zbog toplinske vodljivosti i viskoznog trenja na stjenkama vokalnog trakta
- mekoća stjenki vokalnog trakta
- zračenje zvuka na usnama
- akustička veza usne i nosne šupljine
- izvor zvuka u vokalnog traktu

Potpuna i detaljna akustička teorija koja bi uzimala u obzir sve gore navedene efekte nadmašuje opseg ove skripte, a preciznije, kao takva još ni ne postoji. Treba se zadovoljiti pojednostavljenim matematičkim modelom uz zanemarenja nekih od gore navedenih faktora.



Slika 5.1-1 (a) Model vokalnog trakta; (b) Odgovarajuća funkcija površine poprečnog presjeka $A(x,t)$

Najjednostavnija fizikalna konfiguracija kojom se može opisati proces nastanka govora je prikazana na slici 5.1-1 a). Vokalni trakt je modeliran kao cijev nejednolikog, vremenski promjenjivog poprečnog presjeka. Za frekvencije čije su valne duljine dugačke u usporedbi s dimenzijama vokalnog trakta (manje od otprilike 4000 Hz), moguće je pretpostaviti da se zvučni val kroz vokalni trakt širi kao plošni val koji putuje duž osi cijevi. Daljnje pojednostavljenje uvodi se pretpostavkom da nema gubitaka uslijed viskoznosti ili toplinske

vodljivosti bilo u unutrašnjosti cijevi (u samom fluidu) ili na njenim stjenkama. Koristeći ove pretpostavke i zakone očuvanja mase, momenta i energije, Portnoff je pokazao da zvučni valovi u cijevi zadovoljavaju slijedeći par jednažbi:

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial(u/A)}{\partial t} \quad (5.1-1)$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \frac{\partial(pA)}{\partial t} + \frac{\partial A}{\partial t} \quad (5.1-2)$$

gdje su:

- $p=p(x,t)$ promjena zvučnog tlaka na mjestu x u trenutku t
- $u=u(x,t)$ promjena brzine protoka volumena zraka na mjestu x u trenutku t
- ρ gustoća zraka u cijevi
- c brzina širenja zvuka u zraku
- $A=A(x,t)$ prostorna funkcija cijevi, tj. iznos površine poprečnog presjeka okomitog na os cijevi kao funkcija udaljenosti x duž cijevi i vremena t

Jednažbe (5.1-1) i (5.1-2) nije moguće analitički riješiti osim u slučaju najjednostavnijih konfiguracija. Međutim, moguće ih je riješiti numerički. Potpuno rješenje diferencijalnih jednažbi iziskuje nalaženje tlaka i brzine protoka zraka za sve vrijednosti x i t u prostoru ograničenom glasnica i usnama. Da bi se sustav mogao riješiti, moraju biti definirani rubni uvjeti za oba kraja cijevi. Na strani usnica, rubni uvjet je određen zračenjem zvuka iz cijevi, tj. pretvorbom brzine protoka volumena zraka na mjestu usnica u varijacije zvučnog tlaka koje se zatim šire prostorom. Rubni uvjet na strani glasnica uvjetuje sama priroda uzbude.

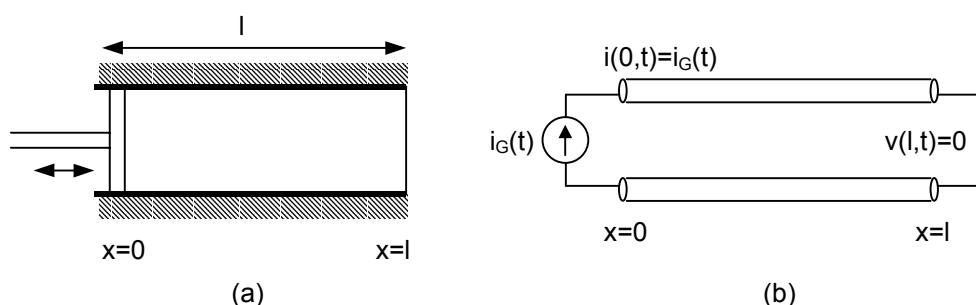
Pored rubnih uvjeta, mora također biti poznata i funkcija površine poprečnog presjeka $A(x,t)$. Na slici 5.1-1 b) prikazana je funkcija površine poprečnog presjeka cijevi prikazane na slici 5.1-1 a) u određenom vremenskom trenutku. Za trajne glasove realno je pretpostaviti da se funkcija $A(x,t)$ ne mijenja značajno u vremenu, dok se međutim kod tranzijentnih glasova to ne može pretpostaviti. Detaljna mjerenja funkcije $A(x,t)$ je iznimno teško provesti, čak i kod trajnih glasova. Jedan pristup takvim mjerenjima je upotrebom Rontgenskih zraka za dobivanje "filma" koji prikazuje oblik i promjene oblika vokalnog trakta kroz vrijeme. Fant i Perkell su radili takva mjerenja, međutim zbog štetnosti x-zraka ovakva mjerenja mogu biti samo vrlo ograničenog opsega. Drugi pristup je zaključivanje o obliku vokalnog trakta na osnovi provedenih akustičkih mjerenja. Sondhi i Gopinath su opisali jedan takav pristup koji se temelji na uzbudi vokalnog trakta pomoću vanjskog izvora. Oba ova pristupa su korisna za prikupljanje znanja o dinamici nastanka govora, ali nisu direktno primjenjivi na modeliranje govornih signala (npr. u svrhu prijenosa). Atal je opisao istraživanja usmjerena ka dobivanju $A(x,t)$ direktno iz snimljenih govornih signala.

Potpuno rješenje jednažbi (5.1-1) i (5.1-2) je jako složeno čak i kada je $A(x,t)$ točno određena. Na sreću, nije neophodno rješavanje jednažbi u najopćenitijim uvjetima da bi se dobio uvid u samu prirodu govornih signala. Za dobivanje rješenja moguće je koristiti čitav niz razumnih pojednostavljenja i aproksimacija.

5.2 Jednolika cijev bez gubitaka

Koristan uvid u prirodu govornih signala može se dobiti razmatrajući vrlo jednostavan model u kojem se pretpostavlja da je prostorna funkcija vokalnog trakta $A(x,t)$ konstantna i u

vremenu i u prostoru (vremenski nepromjenjiva s jednolikim presjekom). Ova je konfiguracija približno točna za neutralni samoglasnik /UH/. Najprije će biti ispitan ovaj najjednostavniji model, a zatim kasnije i oni realniji i složeniji.



Slika 5.2-1 (a) Jednolika cijev bez gubitaka s idealnim zaključenjima; (b) odgovarajuća analogija s električnom linijom bez gubitaka

Slika 5.2-1 a) prikazuje cijev jednolikog presjeka pobuđenu idealnim izvorom brzine protoka volumena zraka. Idealni izvor predstavlja klip koji se pod utjecajem vanjske sile može slobodno gibati lijevo-desno na bilo koji odabrani način i to neovisno o varijacijama tlaka u cijevi. Daljnja pretpostavka je da na otvorenom kraju cijevi nema promjene u tlaku već samo varijacije u brzini protoka volumena zraka. Ovo su očito velika zanemarenja koja je nemoguće postići u stvarnosti, međutim opravdava ih činjenica da razmatranje ovog modela kao i osnovne karakteristike rješenja dobivenih ovim postupkom imaju mnogo toga zajedničkog s puno realnijim modelima. Štoviše, pokazat će se da općenitiji modeli mogu biti sagrađeni spajanjem više ovakvih jednolikih cijevi.

Ako je funkcija $A(x,t)=A$ konstantna, tada parcijalne diferencijalne jednačbe (5.1-1) i (5.1-2) prelaze u oblik:

$$-\frac{\partial p}{\partial x} = \frac{\rho}{A} \frac{\partial u}{\partial t} \quad (5.2-1)$$

$$-\frac{\partial u}{\partial x} = \frac{A}{\rho c^2} \frac{\partial p}{\partial t} \quad (5.2-2)$$

Može se pokazati da rješenje jednačbi (5.2-1) i (5.2-2) ima oblik:

$$u(x,t) = (u^+(t-x/c) - u^-(t+x/c)) \quad (5.2-3)$$

$$p(x,t) = \frac{\rho c}{A} (u^+(t-x/c) + u^-(t+x/c)) \quad (5.2-4)$$

U jednačbama (5.2-3) i (5.2-4) funkcije $u^+(t-x/c)$ i $u^-(t+x/c)$ mogu biti interpretirane kao valovi koji putuju u pozitivnom odnosno negativnom smjeru osi x . Međusobni odnos među tim suprotno putujućim valovima određen je rubnim uvjetima.

Iz teorije električnih linija bez gubitaka, izrazi koji povezuju napon $v(x,t)$ i struju $i(x,t)$ na jednolikoj (homogenoj) liniji bez gubitaka su:

$$-\frac{\partial v}{\partial x} = L \frac{\partial i}{\partial t} \quad (5.2-5)$$

$$-\frac{\partial i}{\partial x} = C \frac{\partial v}{\partial t} \quad (5.2-6)$$

gdje su L i C induktivitet, odnosno kapacitet po jedinici dužine. Može se uočiti da je teorija jednolikih električnih linija bez gubitaka direktno primjenjiva na akustičku cijev jednolikog presjeka ako usvojimo analogiju prikazanu u tablici 5.2-1.

Tablica 5.2-1 Analogija između akustičkih i električkih veličina

Akustička veličina	Analogna električka veličina
p – tlak	v – napon
u – brzina protoka volumena zraka	i – struja
ρ/A – akustički induktivitet	L – induktivitet
$A/(\rho c^2)$ – akustički kapacitet	C – kapacitet

Uz ove analogije se akustička cijev jednolikog presjeka ponaša jednako kao jednolika električna linija bez gubitaka zaključena kratkim spojem ($v(l,t)=0$) s jedne strane, a pobuđena idealnim strujnim izvorom s druge ($i(0,t)=i_G(t)$). Ovo je prikazano na slici 5.2-1 b).

Jako je korisno promotriti ponašanje linearnih sustave kao što su električne linije i krugovi u frekvencijskoj domeni. Analogno tome, mogu se napraviti slične transformacije za slučaj jednolike cijevi bez gubitaka. Prikaz ovog modela u frekvencijskoj domeni dobiva se ako se pretpostavi slijedeći rubni uvjet za $x=0$:

$$u(0,t) = u_G(t) = U_G(\Omega)e^{j\Omega t} \quad (5.2-7)$$

Drugim riječima, cijev je pobuđena kompleksnom eksponencijalnom funkcijom promjene brzine protoka volumena zraka, kružne frekvencije Ω i kompleksne amplitude $U_G(\Omega)$. Pošto su jednadžbe (5.2-1) i (5.2-2) linearne, pretpostavljena rješenja za pozitivno i negativno putujuće valove $u^+(t-x/c)$ i $u^-(t+x/c)$ također moraju biti istog oblika kao i pobuda, tj. :

$$u^+(t-x/c) = K^+ e^{j\Omega(t-x/c)} \quad (5.2-8)$$

$$u^-(t+x/c) = K^- e^{j\Omega(t+x/c)} \quad (5.2-9)$$

Ako sada ovakva pretpostavljena rješenja uvrstimo u izraze (5.2-3) i (5.2-4) koji definiraju tlak i brzinu protoka duž cijevi moguće je uz poznavanje rubnih uvjeta odrediti nepoznate konstante K^+ i K^- . Rubni uvjet na desnom kraju cijevi gdje su usne glasi:

$$p(l,t) = 0 \quad (5.2-10)$$

dok je rubni uvjet na lijevom kraju gdje su glasnice definiran izrazom (5.2-7). Nakon određivanja konstanti K^+ i K^- i njihovog uvrštavanja u izraze (5.2-8) i (5.2-9) dobiva se rješenje u stacionarnom stanju:

$$p(x,t) = jZ_0 \frac{\sin[\Omega(l-x)/c]}{\cos[\Omega l/c]} U_G(\Omega) e^{j\Omega t} \quad (5.2-11)$$

$$u(x,t) = \frac{\cos[\Omega(l-x)/c]}{\cos[\Omega l/c]} U_G(\Omega) e^{j\Omega t} \quad (5.2-12)$$

gdje je:

$$Z_0 = \frac{\rho c}{A} \quad (5.2-13)$$

karakteristična akustička impedancija cijevi.

Alternativni pristup koji se koristi u nastavku omogućava izbjegavanje odvojenog rješavanja za valove koji putuju u jednom odnosno drugom smjeru, tako da se rješenja za $p(x,t)$ i $u(x,t)$ direktno pretpostave u obliku kompleksnih eksponencijala čija kompleksna amplituda ovisi o poziciju unutar cijevi:

$$p(x,t) = P(x,\Omega) e^{j\Omega t} \quad (5.2-14)$$

$$u(x,t) = U(x,\Omega) e^{j\Omega t} \quad (5.2-15)$$

Uvrštenjem ovih pretpostavljenih rješenja u jednačbe (5.1-1) i (5.1-2) vremenska ovisnost definirana članom $e^{j\Omega t}$ se gubi, jer se krati s lijeve i desne strane parcijalne diferencijalne jednačbe. Radi toga dobivaju se obične diferencijalne jednačbe koje povezuju kompleksne amplitude tlaka i brzine protoka kao funkcije pozicije x unutar cijevi.

$$-\frac{dP(x,\Omega)}{dx} = ZU(x,\Omega) \quad (5.2-16)$$

$$-\frac{dU(x,\Omega)}{dx} = YP(x,\Omega) \quad (5.2-17)$$

gdje je:

$$Z = j\Omega \frac{\rho}{A} \quad (5.2-18)$$

akustička impedancija po jedinici duljine, a

$$Y = j\Omega \frac{A}{\rho c^2} \quad (5.2-19)$$

akustička admitancija po jedinici duljine. Diferencijalne jednačbe (5.2-16) i (5.2-17) imaju rješenje oblika:

$$P(x,\Omega) = Ae^{\gamma x} + Be^{-\gamma x} \quad (5.2-20)$$

$$U(x,\Omega) = Ce^{\gamma x} + De^{-\gamma x} \quad (5.2-21)$$

gdje je:

$$\gamma = \sqrt{ZY} = j\Omega/c \quad (5.2-22)$$

Nepoznati koeficijenti mogu se izračunati uvrštavajući rubne uvjete

$$P(l, \Omega) = 0 \quad (5.2-23)$$

$$U(0, \Omega) = U_G(\Omega) \quad (5.2-24)$$

Dobiveni rezultat je naravno isti kao u jednadžbama (5.2-11) i (5.2-12) koje izražavaju odnos između amplitude sinusoidalnog izvora brzine protoka volumena zraka na ulazu cijevi i tlaka odnosno brzine protoka volumena zraka u bilo kojoj točki u cijevi. Vrlo je interesantno promotriti odnos između brzine protoka zraka na usnama i brzine protoka na izvoru (odnos sa početka na kraj cijevi). Tako iz izraza (5.2-12) slijedi:

$$u(l, t) = U(l, \Omega)e^{j\Omega t} = \frac{1}{\cos(\Omega l / c)} U_G(\Omega)e^{j\Omega t} \quad (5.2-25)$$

Omjer kompleksnih amplituda $U(l, \Omega)$ i $U_G(\Omega)$

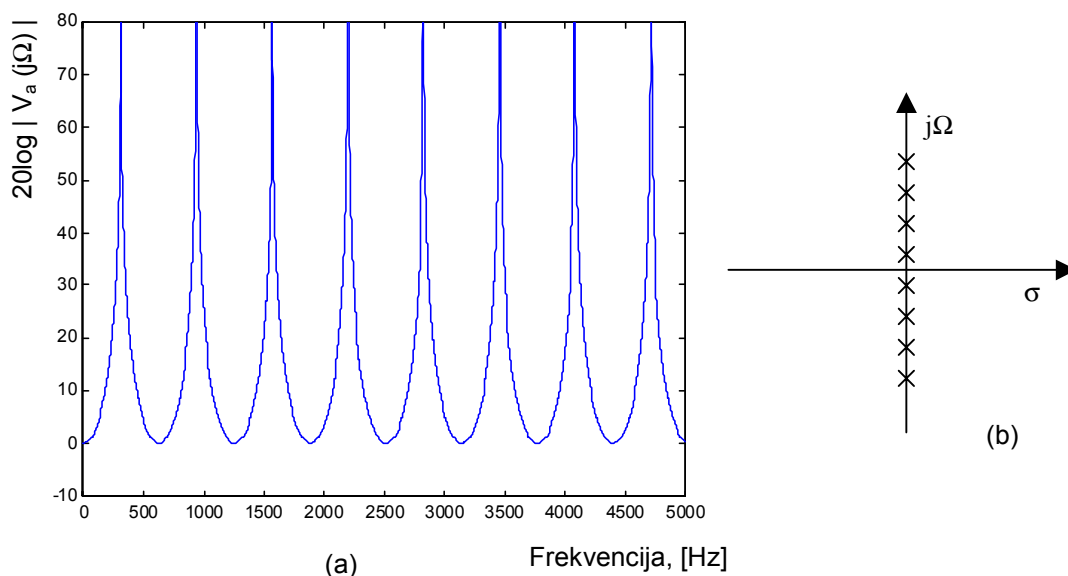
$$\frac{U(l, \Omega)}{U_G(\Omega)} = V_a(j\Omega) = \frac{1}{\cos(\Omega l / c)} \quad (5.2-26)$$

definira frekvencijsku karakteristiku cijevi koji povezuje ulaznu i izlaznu brzinu protoka volumena zraka. Ova funkcija je nacrtana na slici 5.2-2 a) za vrijednosti $l=0.175$ m i $c=350$ m/s. Zamjenom Ω sa s/j dobivamo Laplaceovu transformaciju ili prijenosnu funkciju sustava.

$$V_a(s) = \frac{2e^{-sl/c}}{1 + e^{-s2l/c}} \quad (5.2-27)$$

Izjednačavanjem nazivnika s nulom može se primijetiti kako $V_a(s)$ ima beskonačan broj jednoliku razmaknutih polova na $j\Omega$ osi, tj. :

$$s_n = \pm j \left[\frac{(2n+1)\pi c}{2l} \right] \quad n=0, 1, 2, \dots \quad (5.2-28)$$



Slika 5.2-2 (a) Frekvencijska karakteristika; (b) pozicije polova za jednoliku cijev bez gubitaka

Polozicije polova su prikazane na slici 5.2-2 b). Polovi prijenosne funkcije linearnog vremenski nepromjenjivog sustava su vlastite frekvencije sustava. Polovi također odgovaraju rezonantnim frekvencijama sustava. Kada se radi o obradi govora, rezonantne frekvencije se još nazivaju i formantne frekvencije. Slični se efekti rezonancije javljaju i kod složenijih oblika vokalnog trakta.

Treba napomenuti da poznavanje frekvencijske karakteristike sustava omogućava nalaženje odziva ne samo za sinusoidalni ulazni signal, već za bilo kakav pobudni signal uz korištenje Fourierove analize. Jednadžba (5.2-26) se može interpretirati i kao omjer Fourierove transformacije protoka na mjestu usnica i Fourierove transformacije pobudnog signala, tj. protoka na mjestu glasnica. Zbog toga je frekvencijska karakteristika akustičkog sustava jako pogodan način njegovog opisa.

5.3 Utjecaj gubitaka u vokalnog trakta

Izrazi za širenje zvuka u vokalnog trakta koji su do sada korišteni bili su izvedeni uz pretpostavku da ne postoje energetska gubitci. U stvarnosti, ti gubitci postoje i oni su posljedica trenja između zraka i stjenke cijevi, toplinske vodljivosti kroz stjenke i vibracije samih stjenki. Da bi i ovi utjecaji bili uključeni u jednadžbama širenja zvuka u vokalnog trakta potrebno je postaviti nove parcijalne diferencijalne jednadžbe koje uključuju i navedene gubitke. Rješavanje takvih jednadžbi je mnogo složenije nego u idealiziranom slučaju bez gubitaka, a glavni problem je u činjenici što su gubitci frekvencijski zavisni. Zbog toga se često koristi pojednostavljeni model koji kreće od jednadžbi za idealnu cijev bez gubitaka u frekvencijskoj domeni, a utjecaji gubitaka se modeliraju promjenom akustičke impedancije, odnosno admitancije. U nastavku ovog poglavlja bit će predstavljeni takvi modeli kojima se opisuju sva tri tipa gubitaka.

Prvo će biti razmotren gubitak koji nastaje uslijed vibracija stjenki vokalnog trakta. Različiti zračni tlak na različitim mjestima unutar trakta uzrokuje i različitu silu na stjenke. U slučaju da su stjenke elastične poprečni presjek cijevi $A(x,t)$ će se mijenjati u ovisnosti o tlaku u cijevi $p(x,t)$. U slučaju da su promjene tlaka vrlo male, rezultirajuće promjene površine poprečnog presjeka mogu se razmatrati kao male promjene "nominalne" površine, tj. :

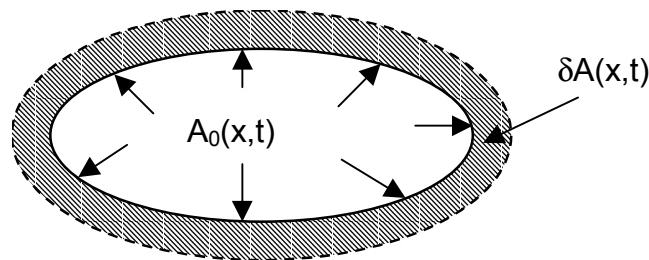
$$A(x,t) = A_0(x,t) + \delta A(x,t) \quad (5.3-1)$$

gdje su : $A_0(x,t)$ nominalna površina poprečnog presjeka
 $\delta A(x,t)$ mala promjena površine poprečnog presjeka

Na slici 5.3-1 je ilustrirana pojava povećanja nominalnog presjeka $A_0(x,t)$ za promjenu $\delta A(x,t)$. Zbog mase i elastičnosti stjenki vokalnog trakta veza između promjene površine $\delta A(x,t)$ i promjene tlaka $p(x,t)$ može biti predstavljena sljedećom diferencijalnom jednadžbom :

$$m_w \frac{d^2(\delta A)}{dt^2} + b_w \frac{d(\delta A)}{dt} + k_w (\delta A) = p(x,t) \quad (5.3-2)$$

gdje su : m_w masa po jedinici dužine stjenke vokalnog trakta
 b_w prigušenje po jedinici dužine stjenke vokalnog trakta
 k_w ukrućenost po jedinici dužine stjenke vokalnog trakta



Slika 5.3-1 Ilustracija promjene poprečnog presjeka vokalnog trakta uslijed vibracija stjenki

Ako se izraz za poprečni presjek $A(x,t)$ iz izraza (5.3-1) uvrsti u parcijalne diferencijalne jednačbe koje opisuju odnos brzine protoka $u(x,t)$ i tlaka $p(x,t)$ duž vokalnog trakta, prikazane u jednačbama (5.1-1) i (5.1-2), tada slijedi:

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial \left(\frac{u}{A_0(x,t) + \delta A(x,t)} \right)}{\partial t} \quad (5.3-3)$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \frac{\partial (p(A_0(x,t) + \delta A(x,t)))}{\partial t} + \frac{\partial (A_0(x,t) + \delta A(x,t))}{\partial t} \quad (5.3-4)$$

Za male promjene poprečnog presjeka $\delta A(x,t)$ moguće je $u/A(x,t)$ zamijeniti sa $u/A_0(x,t)$, odnosno $pA(x,t)$ zamijeniti sa $pA_0(x,t)$, što vodi do jednostavnije jednačbe:

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial \left(\frac{u}{A_0(x,t)} \right)}{\partial t} \quad (5.3-5)$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \frac{\partial (p(A_0(x,t)))}{\partial t} + \frac{\partial (A_0(x,t))}{\partial t} + \frac{\partial (\delta A(x,t))}{\partial t} \quad (5.3-6)$$

Radi određivanja utjecaja gubitaka uslijed elastičnosti stjenki, pogodno je razmotriti odnose brzine protoka i tlaka u frekvencijskoj domeni. Za slučaj elastične cijevi čiji je nominalni presjek $A_0(x,t)$ vremenski nepromjenjiv, tj. $A_0(x,t) \equiv A_0(x)$ i koja je pobuđena sa brzinom protoka $u(0,t)$ oblika kompleksne eksponencijale, tj. za:

$$u(0,t) = U_G(\Omega)e^{j\Omega t} \quad (5.3-7)$$

rješenje za brzinu protoka $u(x,t)$ i tlak $p(x,t)$ će također biti oblika kompleksnih eksponencijala iste frekvencije Ω , tj.

$$p(x,t) = P(x,\Omega)e^{j\Omega t} \quad (5.3-8)$$

$$u(x,t) = U(x,\Omega)e^{j\Omega t} \quad (5.3-9)$$

Takvo rješenje je moguće pretpostaviti zbog linearnosti i vremenske nepromjenjivosti parcijalnih diferencijalnih jednačbi.

Uvrštavajući pretpostavljena rješenja (5.3-8) i (5.3-9) u jednačbe (5.3-2), (5.3-5) i (5.3-6) slijedi:

$$-\frac{\partial P(x, \Omega)}{\partial x} = Z(x, \Omega)U(x, \Omega) \quad (5.3-10)$$

$$-\frac{\partial U(x, \Omega)}{\partial x} = Y(x, \Omega)P(x, \Omega) + Y_w(x, \Omega)P(x, \Omega) \quad (5.3-11)$$

gdje su :

$$Z(x, \Omega) = j\Omega \frac{\rho}{A_0(x)} \quad (5.3-12)$$

$$Y(x, \Omega) = j\Omega \frac{A_0(x)}{\rho c^2} \quad (5.3-13)$$

$$Y_w(x, \Omega) = \frac{1}{j\Omega m_w(x) + b_w(x) + \frac{k_w(x)}{j\Omega}} \quad (5.3-14)$$

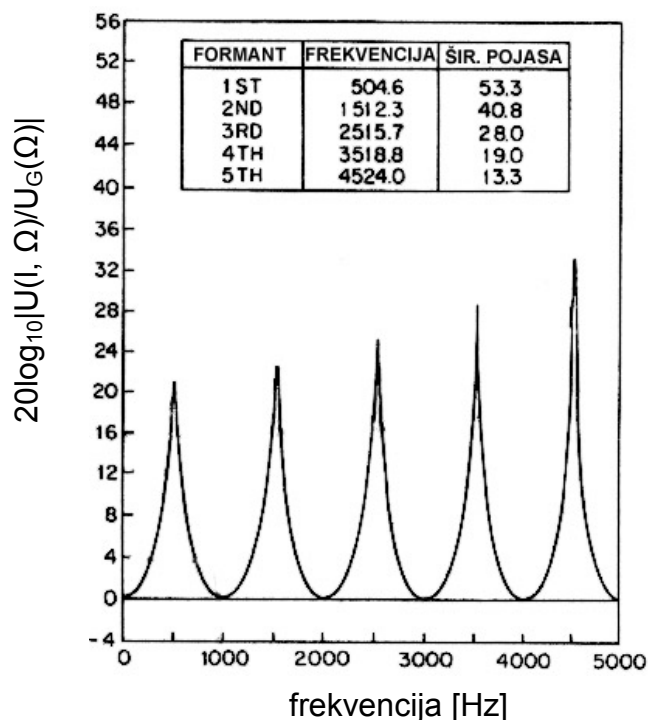
U slučaju elastične cijevi vremenski nepromjenjivog nominalnog poprečnog presjeka izrazi su identični kao i za cijev bez gubitaka (izrazi (5.2-16) do (5.2-19)) osim za novi član Y_w . Razlika je i u tome što je sada $A_0(x)$ funkcija od x , dok je prije bila konstanta A . Uz eksperimentalno određene parametre m_w , b_w i k_w za stvarno tkivo, te uz rubne uvjete :

$$u(0, t) = U_G(\Omega)e^{j\Omega t}, \quad p(l, t) = 0 \quad (5.3-15)$$

moćno je odrediti novu frekvencijsku karakteristiku koja je prikazana slikom 5.3-2. Ona prikazuje odnos brzine protoka volumena zraka na usnama (kraj cijevi) i brzine protoka volumena zraka na glasnicama (početak cijevi), te naravno uzima u obzir gubitke uslijed vibracija stjenki.

Dobivena frekvencijska karakteristika ima oblik:

$$V_a(j\Omega) = \frac{U(l, \Omega)}{U_G(\Omega)} \quad (5.3-16)$$



Slika 5.3-2 Frekvencijska karakteristika cijevi poprečnog presjeka 5.0 cm^2 i duljine 17.5 cm sa vibrirajućim stjenkama i bez drugih gubitaka

Iz grafa je vidljivo da frekvencijska karakteristika više nije beskonačna na frekvencijama polova: 500 Hz , 1500 Hz , 2500 Hz itd. U pripadajućoj tablici na slici 5.3-2 prikazane se centralne frekvencije formanata kao i širine pojasa. Širina pojasa se definira kao razlika dvije frekvencije (lijevo i desno od maksimuma) na kojima amplitudna karakteristika ima 3 dB manju vrijednost nego u maksimumu. Vidljivo je da su centralne frekvencije više od centralnih frekvencija za slučaj bez gubitaka. Nadalje, širine pojasa na rezonantnim frekvencijama više nisu nula kao što je slučaj za idealne rezonatore kod cijevi bez gubitaka, već poprimaju neki konačan iznos. Uočavamo da je učinak vibracija stjenki najveći na niskim frekvencijama. Takvo ponašanje se moglo i očekivati, obzirom da stjenke vokalnog trakta relativno velike mase mogu pratiti varijacije tlaka samo na nižim frekvencijama, dok se pri visokim frekvencijama ponašaju kao kruta stjenka. Ponašanje ilustrirano u ovom primjeru na slici 5.3-2 odnosi se na elastičnu cijev s vremenski nepromjenjivim poprečnim presjekom koji nije niti funkcija od x , tj. $A_0(x)=A$. Međutim ovakvo ponašanje je tipično i za druge oblike funkcije $A_0(x)$ i može se zaključiti da gubici uslijed elastičnosti stjenki uzrokuju povišenje centralnih frekvencija formanata i proširenje širine pojasa koje je posebno izraženo na niskim frekvencijama.

Učinci viskoznog trenja i toplinske vodljivosti kroz stjenke vokalnog trakta su manje izraženi i manje značajni od gore navedenih učinaka uslijed vibracija stjenki. Pokazalo se da se gubici uslijed viskoznog trenja mogu uspješno modelirati u frekvencijskoj domeni modifikacijom izraza za akustičku impedanciju $Z(x, \Omega)$ i to dodavanjem realnog frekvencijski ovisnog dijela kao što je prikazano u slijedećem izrazu.

$$Z(x, \Omega) = \frac{S(x)}{[A_0(x)]^2} \sqrt{\Omega \rho \mu / 2} + j\Omega \frac{\rho}{A_0(x)} \quad (5.3-17)$$

gdje su : $S(x)$... opseg vokalnog trakta na mjestu x ,

μ koeficijent trenja i

ρ gustoća zraka u cijevi

Slično tome gubitak uslijed toplinske vodljivosti uzima se u obzir dodavanjem realnog frekvencijski ovisnog dijela samo ovaj puta akustičkoj admitanciji.

Novodobiveni izraz glasi :

$$Y(x, \Omega) = \frac{S(x)(\eta - 1)}{\rho c^2} \sqrt{\frac{\lambda \Omega}{2c_p \rho}} + j\Omega \frac{A_0(x)}{\rho c^2} \quad (5.3-18)$$

gdje su : c_p konstanta specifične topline uz konstantan tlak

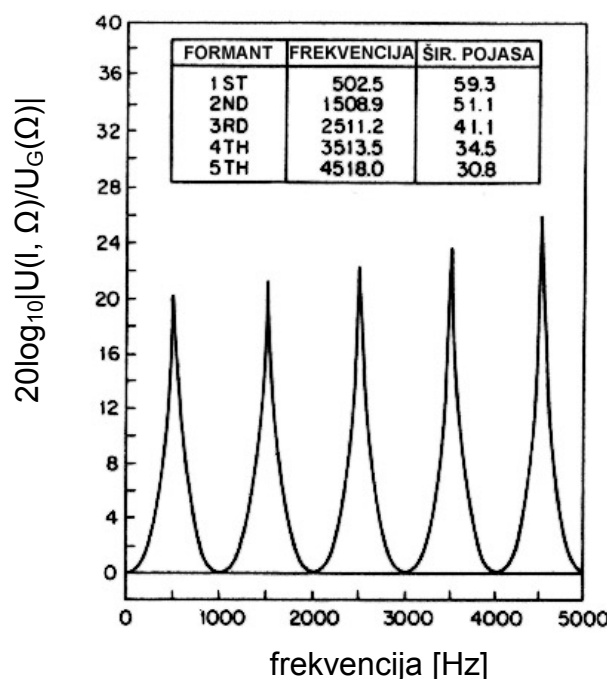
η omjer c_p i konstante specifične topline uz konstantan volumen c_v

λ koeficijent toplinske vodljivosti

Iz gornje diskusije može se zaključiti da su gubici uslijed viskoznog trenja proporcionalni realnom dijelu od $Z(x, \Omega)$, dok su gubici uslijed toplinske vodljivosti proporcionalni realnom dijelu od $Y(x, \Omega)$. Iz izraza (5.3-17) i (5.3-18) je vidljivo da su u oba slučaja ti gubici (realni dijelovi) proporcionalni drugom korijenu frekvencije, tj: $\Omega^{-1/2}$.

Uzimajući u obzir izraze za akustičku impedanciju $Z(x, \Omega)$, (5.3-17) i admitanciju $Y(x, \Omega)$, (5.3-18), te izraz (5.3-14) za dodatnu akustičku admitanciju Y_w moguće je numerički riješiti sustav parcijalnih diferencijalnih jednačbi (5.3-10) i (5.3-11).

Rezultirajuća frekvencijska karakteristika uz rubni uvjet $p(l, t) = 0$ dan je slikom 5.3-3. Ponovo su centralne frekvencije i širine pojasa prikazane u tablici. Uspoređujući ovu sliku i pripadnu tablicu sa prethodnom primjećujemo da centralne frekvencije opadaju dodavanjem trenja i toplinskih gubitaka, dok se širine pojasa povećavaju.



Slika 5.3-3 Frekvencijska karakteristika cijevi poprečnog presjeka 5.0 cm^2 i duljine 17.5 cm sa vibrirajućim stjenkama, trenjem i toplinskim gubitcima

Iz gore navedenih izraza te posebno iz njima pripadnih grafova i tablica zaključujemo da se viskozno trenje i toplinski gubitci povećavaju sa porastom frekvencije te stoga imaju najveći učinak na visokim frekvencijama, dok je učinak vibracija stjenki najizraženiji na niskim frekvencijama. Vibracije stjenki nastoje podići rezonantne frekvencije dok ih viskozno trenje i toplinski gubici nastoje smanjiti. Utjecaj trenja i toplinskih gubitaka je zanemarujući u usporedbi sa utjecajem elastičnih stjenki za frekvencije ispod 3-4kHz, tako da se za veći broj analiza širenja zvuka u vokalnom traktu mogu koristiti izrazi (5.3-10) do (5.3-14) kod kojih su ti gubitci zanemareni. U slijedećem poglavlju biti će pokazano da su gubitci uslijed zračenja na usnicama puno značajniji izvor visokofrekvencijskih gubitaka, što je dodatno opravdanje da se u modelima formiranja govora ignoriraju gubitci uslijed trenja i toplinske vodljivosti.

5.4 Utjecaj zračenja na usnicama

U dosadašnjem razmatranju pokazano je kako gubitci unutar vokalnog trakta utječu na svojstva širenja zvuka kroz vokalni trakt, a prilikom svih analiza pretpostavljalo se da je rubni uvjet na usnicama zadan sa $p(l,t)=0$, što na žalost ne odgovara stvarnosti. U ekvivalentnom modelu s električnom linijom, takvo zaključenje odgovara idealnom kratkom spoju na kraju linije. Takvo idealno zaključenje je gotovo nemoguće postići u stvarnom akustičkom sustavu, jer bi zahtijevalo da na kraju vokalnog trakta smije postojati promjena u iznosu brzine protoka, ali ne i promjena u tlaku. U stvarnosti na kraju vokalnog trakta postoji otvor između usnica (ili otvor nosnica u slučaju nazalnih glasova). Na slici 5.4-1 a) prikazan je pojednostavljeni model koji simulira zračenje na usnicama. Glava se modelira s idealnom kuglom koja na određenom mjestu ima kružni otvor koji simulira otvor usnica. Potrebno je odrediti odnos između tlaka i brzine volumena zraka na kraju trakta kako bi bilo moguće točno odrediti granični uvjet na usnama. Pomoću sfernog modela sa slike 5.4-1 a) to je vrlo teško pa je potrebno pretpostaviti da je površina zračenja (otvor usnica) vrlo mala u odnosu na cijelu sferu (glava). Koristeći to pojednostavljenje, moguće je koristiti model sa slike 5.4-1 b) kod kojeg cijev promjera a završava kao kružni otvor u beskonačnoj plohi.



Slika 5.4-1 a) zračenje iz kružnog otvora na kugli, b) zračenje iz cijevi koja završava kao otvor na beskonačnoj plohi

Koristeći model sa slike 5.4-1 b) dolazimo do izraza za odnos između kompleksne amplitude tlaka i brzine protoka volumena zraka na usnicama u stacionarnom stanju.

$$Z_L(\Omega) = \frac{P(l,\Omega)}{U(l,\Omega)} \quad (5.4-1)$$

gdje je Z_L impedancija zračenja ili teret zračenja

Analogna električna shema ovog tereta zračenja odgovara paralelnom spoju otpora zračenja R_r i zavojnice zračenja L_r . Stoga izraz za impedanciju zračenja izgleda ovako :

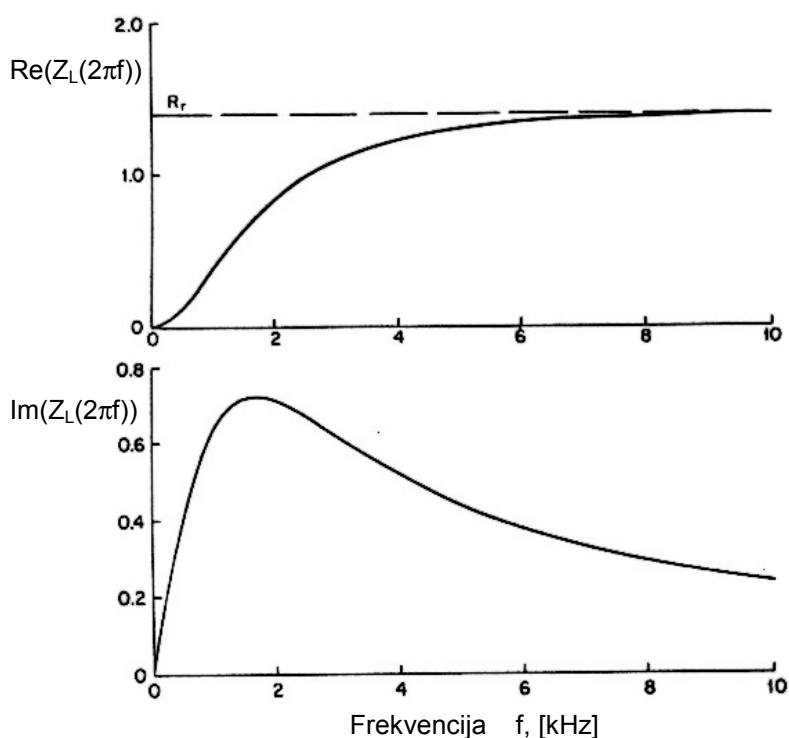
$$Z_L(\Omega) = \frac{j\Omega L_r R_r}{R_r + j\Omega L_r} \quad (5.4-2)$$

$$R_r = \frac{128}{9\pi^2}, \quad L_r = \frac{8a}{3\pi c} \quad (5.4-3)$$

gdje su : a promjer otvora i

c brzina širenja zvuka

Karakteristika ovog tereta zračenja utječe i na širenje zvučnog vala unutar samog vokalnog trakta, jer je rubni uvjet na desnom kraju cijevi definiran izrazom (5.4-1). Iz izraza (5.4-2) možemo zaključiti da je za vrlo niske frekvencije pobudnog signala impedancija zračenja $Z_L(\Omega) \approx 0$, što je identično kratkom spoju, tj. idealiziranom slučaju koji je do sada bio razmatran. Za srednje frekvencije kod kojih je $\Omega L_r \ll R_r$ impedancija zračenja $Z_L(\Omega) \approx j\omega \Omega L_r$. Na visokim frekvencijama kod kojih vrijedi $\Omega L_r \gg R_r$, $Z_L(\Omega) \approx R_r$. Na slikama 5.4-2 prikazani su realni i imaginarni dio od $Z_L(\Omega)$ kao funkcija od Ω , za tipične vrijednosti parametara. Energetski gubici uslijed zračenja su proporcionalni realnom dijelu od $Z_L(\Omega)$. Stoga možemo primijetiti da će za cijeli sustava koji obuhvaća i vokalni trakt i zračenje na usnicama, gubici uslijed zračenja na usnicama biti izraženiji na visokim frekvencijama. Ako uz rubni uvjet definiran izrazom (5.4-1), ponovno numerički riješimo sustav diferencijalnih jednadžbi (5.3-10) do (5.3-14), za slučaj jednolike vremenski nepromjenjive cijevi s elastičnim stjenkama i gubitcima uslijed trenja i toplinske vodljivosti dobiva se frekvencijska karakteristika prikazana na slici 5.4-3. Ona prikazuje frekvencijsku karakteristiku uz pobudu $U(0,t) = U_G(\Omega)e^{j\Omega t}$, no ovaj puta uz $p(l,t) \neq 0$.

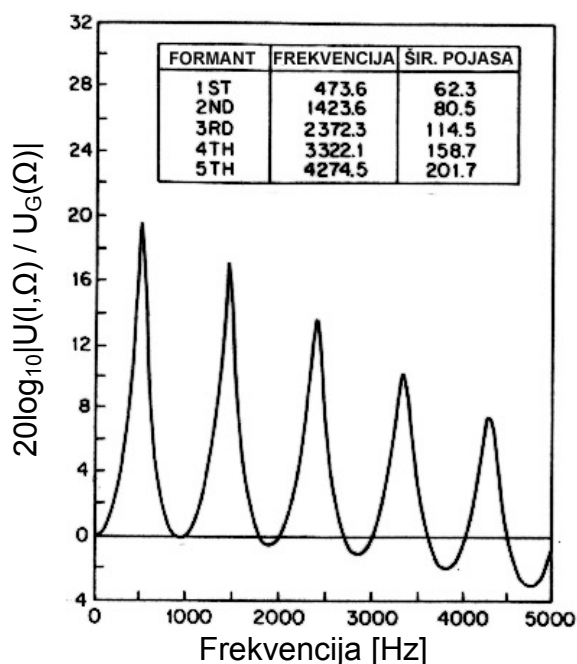


Slika 5.4-2 Realni i imaginarni dio impedancije Z_L , kao funkcija frekvencije f

Ta karakteristika je definirana kvocijentom između kompleksne amplitude brzine protoka volumena zraka na kraju i na početku cijevi, tj. :

$$V_a(j\Omega) = \frac{U(l, \Omega)}{U_G(\Omega)} \quad (5.4-4)$$

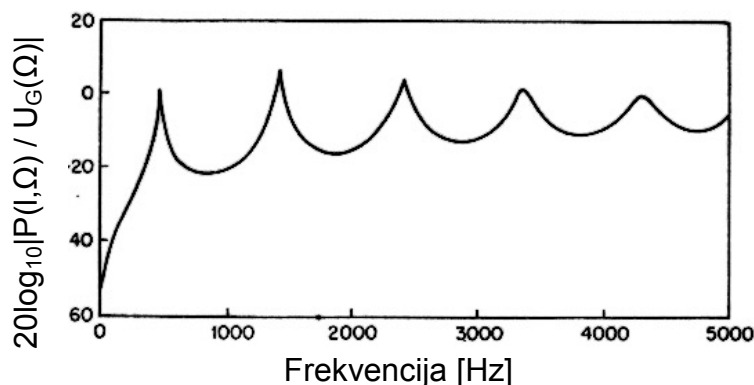
Sada su dakle pored svih dosad spomenutih gubitaka uzeti u obzir i gubitci uslijed zračenja na usnicama, koje je modelirano pomoću cijevi koja završava kao otvor u beskonačnoj plohi. Usporedbom podataka u tablici na slici 5.4-3 sa onima na slici 5.3-3 može se uočiti da se utjecaj gubitaka uslijed zračenja na usnicama manifestira kao sniženje centralnih frekvencija formantata i proširenje širine pojasa formanta. Taj utjecaj je posebno izražen na visokim frekvencijama. Širina prvoga formanta (prva rezonantna frekvencija) je uglavnom određena vibracijom stjenki. Širine drugog i trećeg formanta su kombinacija utjecaja vibracija stjenki i zračenja. Na visokim frekvencijama dominira utjecaj zračenja koji na tim frekvencijama nadvlada utjecaje vibracija stjenki, trenja i toplinske vodljivosti.



Slika 5.4-3 Frekvencijska karakteristika jednolike cijevi sa gubiticima uslijed vibrirajućih stjenki, trenja, toplinske vodljivosti, kao i uslijed zračenja na usnicama $p(l, t) \neq 0$

Obzirom da je ljudsko uho kao i veći broj mikrofona osjetljivo na varijacije zvučnog tlaka (a ne na varijacije brzine protoka zraka) interesantno je odrediti frekvencijsku karakteristiku cjelokupnog sustava, ali definiranu kao kvocijent kompleksne amplitude tlaka na usnicama $P(l, \Omega)$ (kraj cijevi) i kompleksne amplitude brzine protoka volumena zraka na glasnicama $U(0, \Omega)$ (početak cijevi). Ta frekvencijska karakteristika biti će označena sa $H_a(j\Omega)$ i definirana je sljedećim relacijama:

$$H_a(j\Omega) = \frac{P(l, \Omega)}{U_G(\Omega)} = \frac{P(l, \Omega)}{U(l, \Omega)} \cdot \frac{U(l, \Omega)}{U_G(\Omega)} = Z_L(\Omega) \cdot V_a(\Omega) \quad (5.4-5)$$



Slika 5.4-4 Prijenosna funkcija vokalnog trakta definirana kao kvocijent tlaka na usnama i brzine protoka volumena zraka na glasnicama za jednoliku cijev sa svim gubitcima

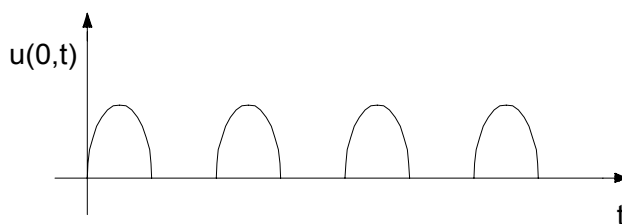
Slika 5.4-4 prikazuje prijenosnu funkciju vokalnog trakta, $20\log_{10}(|H_a(j\Omega)|)$, koja uključuje sve gubitke kao i utjecaj zračenja iz cijevi zaključene beskonačnom plohom. Usporedbom ove slike sa slikom 5.4-3 na kojoj je prikazana frekvencijska karakteristika $V_a(j\Omega)$, vidljivo je izdizanje (pojačanje) visokih frekvencija, kao i nula u prijenosnoj funkciji na frekvenciji $\Omega=0$.

5.5 Pobudni signal vokalnog trakta

U prethodnom poglavlju je bilo objašnjeno kako se pojave širenja zvučnog vala i pojave radijacije ili zračenja na usnicama mogu opisati osnovnim fizikalnim zakonitostima. Radi potpunog opisa akustičkog procesa nastajanja govora potrebno je razmotriti i mehanizme stvaranja zvučnog vala u vokalnom traktu.

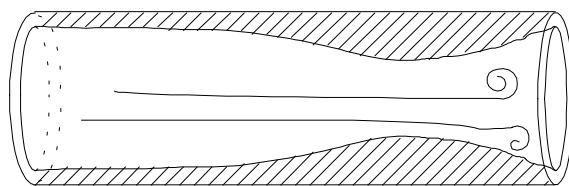
Postoje tri glavna mehanizma nastajanja zvučnog vala:

1. Tok zraka iz pluća je moduliran titranjem glasnica, što rezultira kvazi-periodičkom pobudom u obliku niza pulseva.



Slika 5.5-1 Kvalitativni prikaz brzine protoka volumena zraka na glasnicama $u(0,t)$

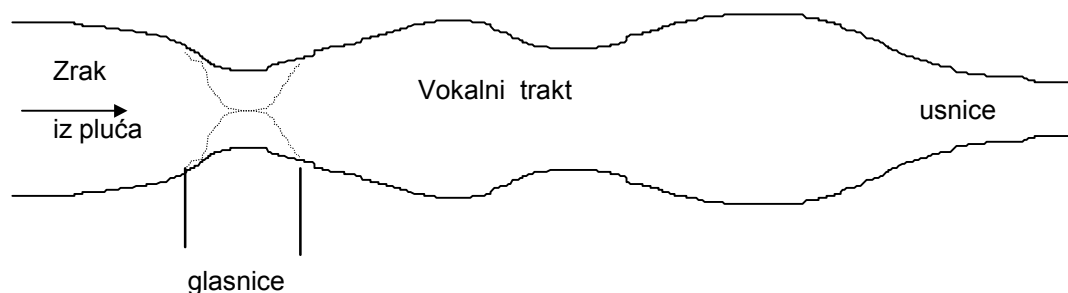
2. Zračna struja koja izlazi iz pluća, prolazeći kroz uske prolaze u vokalnom traktu postaje turbulentna što uzrokuje formiranje šumovitog zvuka (glasa).



Slika 5.5-2 Pojava turbulencija na mjestu suženja u vokalnom traktu

3. Dolazi do formiranja povišenog tlaka iza potpuno zatvorene prepreke u vokalnom traktu. Naglo otvaranje ove prepreke uzrokuje otpuštanje tlaka i nastaje pobudni signal u obliku naglog tranzijenta (eksplozije).

Detaljni pobudni model govornog signala u govornom sustavu uključuje pred-glasnički sustav (pluća, bronhije i trachea), glasnice i vokalni trakt. Doista, model koji je cjelovit u svim potrebnim detaljima je također potpuno sposoban modelirati čak i proces disanja, a ne samo proces nastajanja govora. Model koji je baziran na klasičnoj mehanici i mehanici fluida zbog svoje složenosti nije pogodan za primjenu. Međutim, kvalitativna diskusija osnovnih principa generiranja glasa bit će korisna za formiranje jednostavnijih modela koji se često koriste kao osnova za obradu govora.



Slika 5.5-3 Shematski prikaz vokalnog sustava

Titranje glasnica je moguće objasniti uz pomoć blok sheme vokalnog trakta na slici 5.5-3. Glasnice predstavljaju prepreku toku zraka iz pluća prema vokalnom traktu. Zbog pritiska u plućima, zrak iz pluća prolazi kroz otvor glasnica (engl. *glottis*). Prema Bernoullijevom zakonu, tlak na mjestu prepreke je niži nego lijevo i desno od nje (zato jer je protok veći). Uz pravilnu napetost glasnica ovaj sniženi tlak će privući i spojiti glasnice i potpuno zatvoriti tok zraka, kao što je označeno crtkanim linijama na slici 5.5-3. Kada tlak iza zatvorenih glasnica ponovo dovoljno naraste, glasnice se ponovo razmiču i propuštaju zrak prema vokalnom traktu. Ovaj se proces ciklički ponavlja, tj. glasnice prelaze u stanje trajnog titranja.

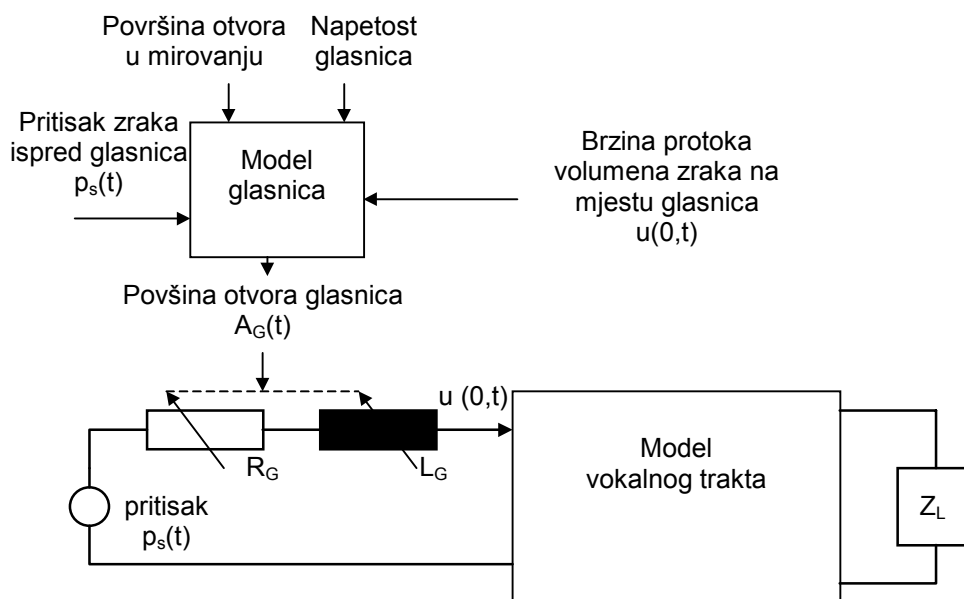
Frekvencija titranja glasnica (otvaranja i zatvaranja) određena je:

- tlakom u plućima,
- napetosti i krutosti glasnica i
- površinom otvora glasnica u opuštenom položaju.

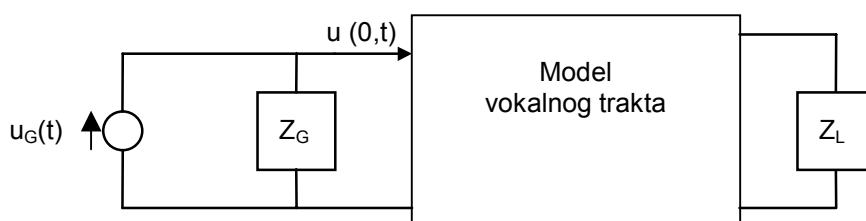
Ovo su kontrolni parametri modela ponašanja glasnica. Ovakav model mora uključivati i utjecaje vokalnog trakta, budući da varijacije tlaka u vokalnom traktu utječu na varijacije tlaka na mjestu glasnica.

U ekvivalentnom modelu vokalnog trakta korištenjem električke mreže, utjecaj vokalnog trakta na titranje glasnica može se nadomjestiti korištenjem ekvivalentnog tereta koji je

spojen na električki izvor koji modelira titranje glasnica. Blok shema tog ekvivalentnog električkog modela glasnica prikazana je na slici 5.5-4.



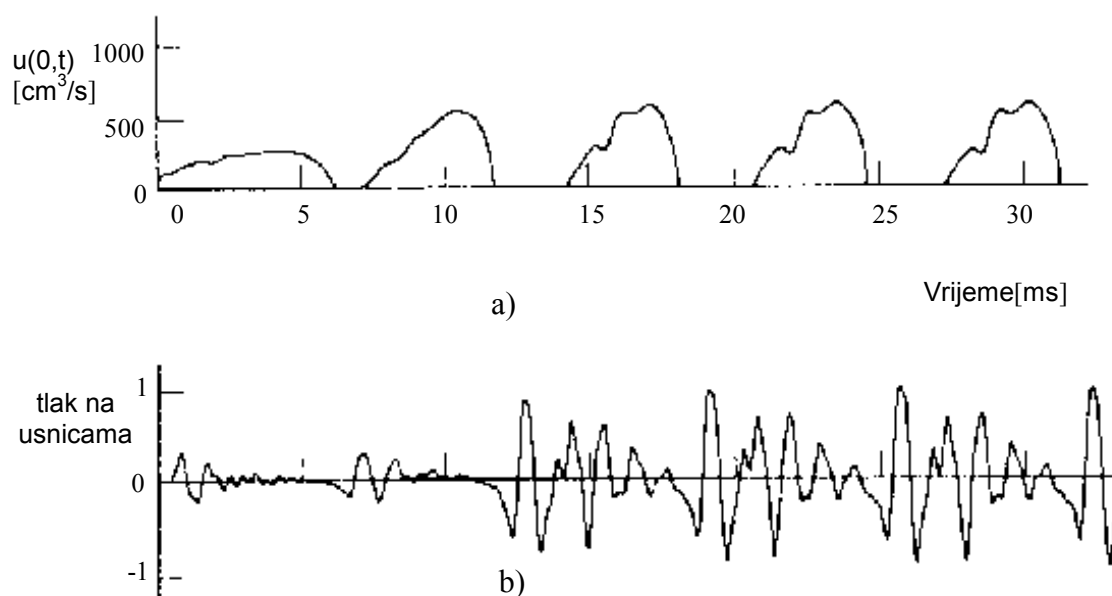
Slika 5.5-4 Model glasnica



Slika 5.5-5 Pojednostavljeni model glasnica

Matematički model glasnica opisan je skupom složenih nelinearnih diferencijalnih jednadžbi. Veza ovih diferencijalnih jednadžbi sa parcijalnim diferencijalnim jednadžbama koje opisuju vokalni trakt može biti predstavljena pomoću vremenski promjenjivog akustičkog otpora i induktiviteta. Ovi impedancijski elementi su funkcije od $1/A_G(t)$. Npr. kada je $A_G(t)=0$ (glasnice zatvorene), impedancija je beskonačna i brzina protoka zraka je jednaka nuli, tj. $u(0,t)=0$. Prema tome zrak iz pluća je automatski sasječen u pulseve. Primjer signala generiranih pomoću ovakvog modela prikazan je na slici 5.5-6. Slika 5.5-6 a) prikazuje brzinu protoka volumena zraka na glasnicama, a 5.5-6 b) tlak na usnama za konfiguraciju vokalnog trakta kod izgovora glasa "a".

Iako su jednadžbe koje opisuju širenje akustičkog vala u vokalnom traktu i zračenje na usnicama linearne, zbog ovisnosti površine otvora glasnica o protoku zraka na ulazu vokalnog trakta, cjelokupni sustav je ipak nelinearan. Na sreću, to međudjelovanje vokalnog trakta i glasnica nije jako izraženo, pa se kod većine postupaka analize govora to međudjelovanje zanemaruje. Ovakvim zanemarenjem cjelokupni sustav se razdvaja na dva neovisna linearna podsustava: pobudni podsustav i prijenosni podsustav kao što je prikazano na slici 5.5-5. U ovom slučaju $u_G(t)$ je izvor brzine protoka čiji je valni oblik prikazan na slici 5.5-6 a). Akustička impedancija na glasnicama Z_G dobivena je linearizacijom odnosa između tlaka i brzine protoka na glasnicama. Ova impedancija je slijedećeg oblika:



Slika 5.5-6 Brzina protoka volumena zraka na glasnicama a), i tlak na usnicama b)

$$Z_G(\Omega) = R_G + j\Omega L_G \quad (5.5-1)$$

gdje su R_G i L_G konstante. Uz ovakvu konfiguraciju rubni uvjet na glasnicama u frekvencijskoj domeni postaje:

$$U(0, \Omega) = U_G(\Omega) - P(0, \Omega) / Z_G(\Omega) \quad (5.5-2)$$

Akustička impedancija glasnica $Z_G(\Omega)$ utječe na širenje rezonantnih karakteristika vokalnog trakta, a njezin utjecaj je najznačajniji na niskim frekvencijama kod kojih $Z_G(\Omega)$ ima malu vrijednost ($\Omega L_G \ll R_G$). Pri višim frekvencijama impedancija je dominantno određena induktivnim članom, tako da modul $|Z_G(\Omega)|$ raste i sve veći dio $u_G(t)$ ulazi u vokalni trakt, a sve manji u Z_G . Može se zaključiti da gubitak uslijed elastičnosti stjenki vokalnog trakta i gubitak uslijed akustičke impedancije glasnica utječu na širine nisko-frekvencijskih formanata, dok zračenje na usnicama, gubici zbog trenja i gubici zbog toplinske vodljivosti utječu na širine visoko-frekvencijskih formanata.

Bezvučni glasovi u govoru nastaju uslijed turbulencija zraka, do kojih dolazi kada brzina protoka volumena zraka na mjestu suženja prijeđe kritični iznos. Ovakav pobudni signal se uspješno modelira sa izvorom slučajnog šuma koji se postavlja na mjesto suženja, pri čemu je amplituda ovog šumovitog signala nelinearno vezana s brzinom protoka na mjestu suženja. Kod bezvučnih frikativa glasnice su opuštene i ne titraju, dok kod zvučnih frikativa pobudni signal ima i periodični i šumovit karakter, jer glasnice titraju. Do pojave turbulencija kod zvučnih frikativa dolazi u trenucima kada brzina protoka na mjestu suženja poprima maksimalni iznos, tj. u vrhovima glotalnih pulseva.

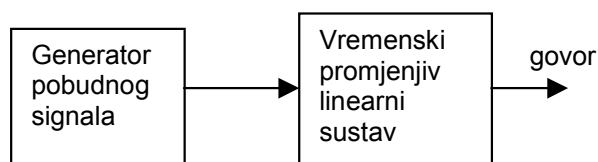
Zadnji tip glasova u govoru su eksplozivni glasovi koji se formiraju u dvije faze. U prvoj fazi vokalni trakt je na određenom mjestu potpuno zatvoren i tlak iza prepreke postepeno raste. U drugoj fazi dolazi do naglog otvaranja prepreke i zrak velikom brzinom izlazi i uzrokuje turbulencije koje se manifestiraju kao eksplozivni glas.

6. AKUSTIČKI GOVORNI MODEL

U prošlom poglavlju diskutirani su osnovni detalji vezani uz akustičku teoriju nastajanja govora. Uz takav detaljni model koji opisuje generiranje (engl. *generation*), širenje (engl. *propagation*) i zračenje (engl. *radiation*) zvuka moguće je u principu odrediti rješenje za izlazni valni oblik govornog signala uz korištenje prikladno odabranih vrijednosti pobude i parametara vokalnog trakta. Primjenom takvih postupaka moguće je izvršiti sintezu govornog signala (umjetni govor), koji će imati vrlo prirodni karakter, tj. zvučat će kao stvarni govorni signal. Međutim, za mnoge je primjene takav detaljan i složeni model nepraktičan i nepotreban. U takvim slučajevima može se koristiti jednostavniji pristup modeliranju govornih signala temeljen na akustičkoj teoriji, koji će biti opisan u nastavku.

6.1 Model izvor-sustav za formiranje govornog signala

Slika 6.1-1 prikazuje pojedini blok dijagram nastajanja govora, koji predstavlja osnovu velikog broja modela korištenih za obradbu ili analizu govora. Svim je ovim modelima zajedničko da je pobuda odijeljena od vokalnog trakta i svojstava zračenja. Utjecaji vokalnog trakta i zračenja ujedinjani su u vremenski promjenjivom linearnom sustavu čija je namjena modeliranje rezonantnih karakteristika opisanih u prošlom poglavlju.

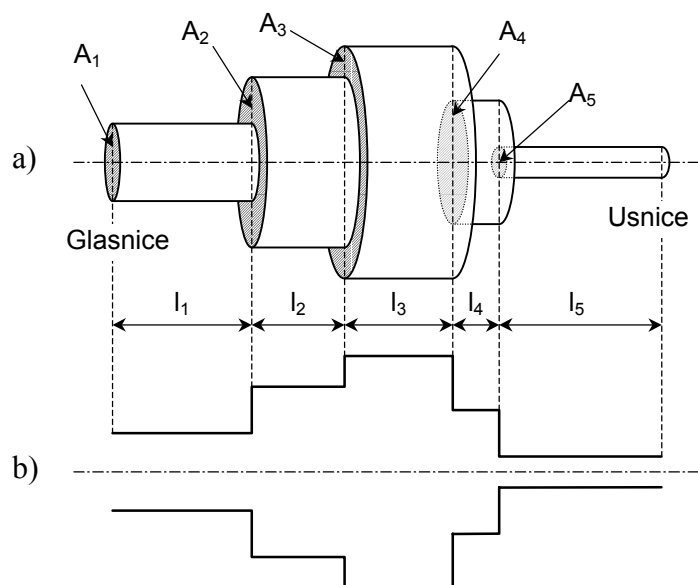


Slika 6.1-1 Model “izvor-sustav” kod stvaranja govora

Generator pobudnog signala na svom izlazu ima ili niz pulseva (tzv. glotalnih pulseva), ili slučajno promjenjivi signal (šum). Parametri izvora i sustava su tako odabrani da rezultirajući izlazni signal ima željene karakteristike koje su što sličnije prirodnom govoru. Ako je to ostvarivo, opisani model može poslužiti kao temelj za obradbu govora. U narednim poglavljima biti će opisani modeli koji se temelje na upravo opisanom principu.

6.2 Model sa spojenim cijevima bez gubitaka

Vrlo široko upotrebljavan model formiranja govora temeljen je na pretpostavci da se vokalni trakt u određenom trenutku vremena može opisati pomoću niza međusobno spojenih cijevi bez gubitaka, kao što je prikazano na slici 6.2-1. Svaka od tih cijevi ima stalan poprečni presjek A_k i dužinu l_k . Površine poprečnog presjeka cijevi i njihove dužine odabrane su tako da aproksimiraju funkciju površine poprečnog presjeka vokalnog trakta, $A(x)$. Ako se koristi mnogo kratkih cijevi, može se očekivati da će rezonantne frekvencije (*formanti*) spojenih uniformnih cijevi biti bliske onima kod cijevi s kontinuirano promjenjivom funkcijom površine poprečnog presjeka, $A(x)$. Međutim, budući da ova aproksimacija zanemaruje gubitke uslijed trenja, toplinske vodljivosti i vibracije stjenki cijevi, može se očekivati da će se širine frekvencijskih pojaseva formantata ovog pojednostavljenog modela razlikovati od širina kod detaljnog modela koji uključuje spomenute gubitke. Neke od tih gubitaka je ipak moguće uzeti u obzir i kod ovog modela, a radi se o utjecaju glotalne impedancije na početku prve cijevi, kao i utjecaju impedancije zračenja na kraju zadnje cijevi, čime se povećava točnost aproksimacije takvog modela.

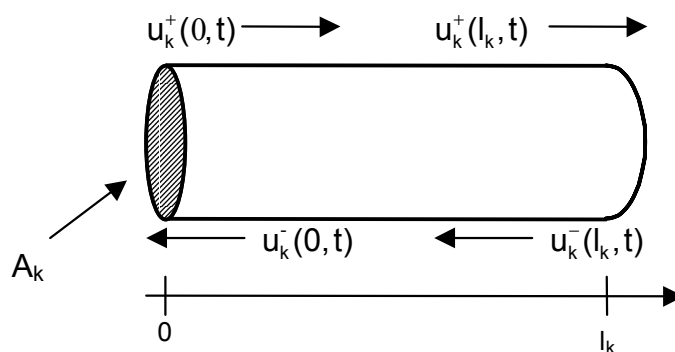


Slika 6.2-1 Model vokalnog trakta sa spojenim cijevima bez gubitaka:
a) izgled modela, b) pojednostavljena skica

Za sadašnju raspravu, mnogo je važnija činjenica da modeli sa cijevima bez gubitaka omogućuju pogodniji prijelaz s vremenski kontinuiranih modela na vremenski diskretne. Prema tome, u nastavku će se razmatrat upravo modeli oblika poput onog sa slike 6.2-1.

6.3 Širenje zvučnih valova u spojenim cijevima

Budući da cijevi sa slike 6.2-1 prema pretpostavci nemaju gubitaka, širenje zvuka kroz k -tu cijev prikazanu na slici 6.3-1 s površinom poprečnog presjeka A_k i dužinu l_k može biti opisano slijedećim relacijama za tlak i brzinu protoka volumena zraka:



Slika 6.3-1 k -ta cijev bez gubitaka

$$p_k(x,t) = \frac{\rho c}{A_k} [u_k^+(t - x/c) + u_k^-(t + x/c)] \quad (6.3-1)$$

$$u_k(x,t) = u_k^+(t - x/c) - u_k^-(t + x/c) \quad (6.3-2)$$

gdje x predstavlja udaljenost od lijevog kraja k -te cijevi ($0 \leq x \leq l_k$) (slika 6.3-1), a u_k^+ i u_k^- su pozitivno odnosno negativno putujući val kroz k -tu cijev.

Odnos između putujućih valova u susjednim cijevima možemo odrediti primjenom fizikalnog principa da tlak i brzina protoka volumena zraka moraju biti kontinuirani i u prostoru i vremenu svugdje u sustavu. Iz ovoga slijede granični uvjeti, koji se primjenjuju na oba kraja cijevi.

Ako se razmotri spoj između k -te i $(k+1)$ -ve cijevi, kao što je prikazano na slici 6.3-2, primjenom uvjeta kontinuiteta na spoju cijevi, dobiva se:

$$p_k(l_k, t) = p_{k+1}(0, t) \quad (6.3-3)$$

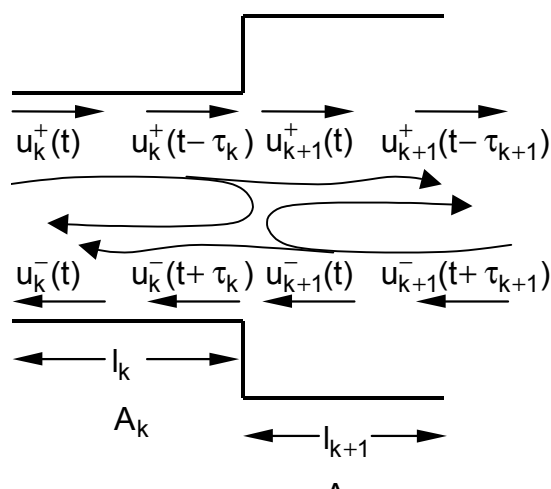
$$u_k(l_k, t) = u_{k+1}(0, t) \quad (6.3-4)$$

Uvrštavanjem jednadžbi (6.3-1) i (6.3-2) u jednadžbe (6.3-3) i (6.3-4), dobiva se:

$$\frac{A_{k+1}}{A_k} [u_k^+(t - \tau_k) + u_k^-(t + \tau_k)] = u_{k+1}^+(t) + u_{k+1}^-(t) \quad (6.3-5)$$

$$u_k^+(t - \tau_k) - u_k^-(t + \tau_k) = u_{k+1}^+(t) - u_{k+1}^-(t) \quad (6.3-6)$$

gdje je $\tau_k = l_k / c$ vrijeme potrebno za prolazak zvučnog vala kroz k -tu cijev. Iz slike 6.3-2 može se uočiti da se dio vala koji se širi kroz k -tu cijev u pozitivnom smjeru, kada dosegne rub cijevi djelomično prenosi dalje na desno, dok se dio reflektira (na lijevo).



Slika 6.3-2 Ilustracija spoja između dvije cijevi bez gubitaka

Slično tomu, val koji ide u negativnom smjeru kroz $(k+1)$ -vu cijev dijelom se prenosi na lijevo u k -tu cijev, dok se dio ponovno reflektira nazad u desno. Ako se $u_{k+1}^+(t)$ i $u_k^-(t + \tau_k)$ izraze preko $u_{k+1}^-(t)$ i $u_k^+(t - \tau_k)$, moći će se odrediti ponašanje originalnog i reflektiranog vala u cijelom sustavu. Ako se iz jednadžbe (6.3-6) izrazi $u_k^-(t + \tau_k)$ i uvrsti u jednadžbu (6.3-5), dobiva se slijedeći izraz:

$$u_{k+1}^+(t) = \left(\frac{2A_{k+1}}{A_{k+1} + A_k} \right) u_k^+(t - \tau_k) + \left(\frac{A_{k+1} - A_k}{A_{k+1} + A_k} \right) u_{k+1}^-(t). \quad (6.3-7)$$

Oduzimanjem jednadžbe (6.3-6) od jednadžbe (6.3-5), slijedi izraz:

$$u_k^-(t + \tau_k) = -\left(\frac{A_{k+1} - A_k}{A_{k+1} + A_k}\right)u_k^+(t - \tau_k) + \left(\frac{2A_k}{A_{k+1} + A_k}\right)u_{k+1}^-(t). \quad (6.3-8)$$

Iz jednadžbe (6.3-7) može se zaključiti da veličina

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} \quad (6.3-9)$$

određuje koliki dio negativno putujućeg vala $u_{k+1}^-(t)$ se reflektira na granici dviju cijevi. Zbog toga se veličina r_k naziva faktorom refleksije za k -ti spoj (spoj k -te i $(k+1)$ -ve cijevi). S obzirom da su sve površine poprečnih presjeka pozitivne veličine, lako se pokazuje da vrijedi:

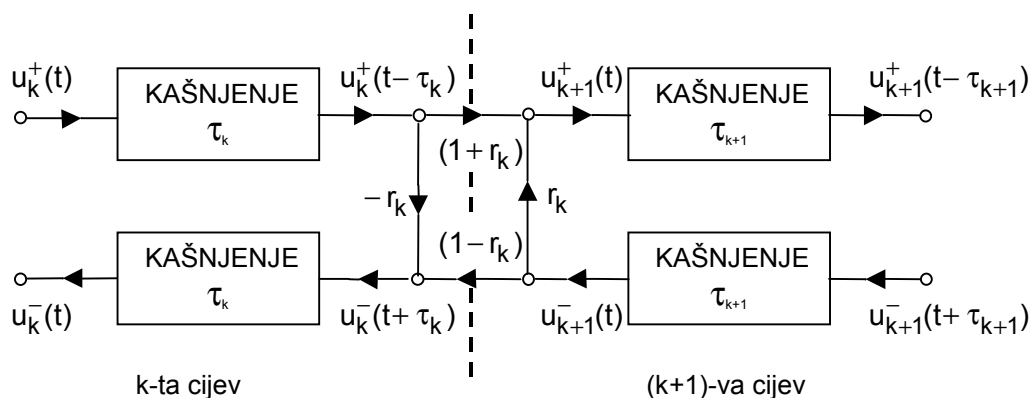
$$-1 \leq r_k \leq 1. \quad (6.3-10)$$

Koristeći se definicijom faktora refleksije r_k iz izraza (6.3-9) moguće je jednadžbe (6.3-7) i (6.3-8) zapisati u slijedećem obliku:

$$u_{k+1}^+(t) = (1 + r_k)u_k^+(t - \tau_k) + r_k u_{k+1}^-(t) \quad (6.3-11)$$

$$u_k^-(t + \tau_k) = -r_k u_k^+(t - \tau_k) + (1 - r_k)u_{k+1}^-(t) \quad (6.3-12)$$

Ove jednadžbe mogu se prikazati i grafički pomoću blok dijagrama na slici 6.3-3. Operacije množenja i zbrajanja u jednadžbama (6.3-11) i (6.3-12), prikazane su u blok dijagramu prema konvencijama za grafove toka signala za prikaz množenja i zbrajanja. Jasno je da se svaki spoj dvije susjedne cijevi modela vokalnog trakta prikazanog na slici 6.2-1, može prikazati sustavom poput onog na slici 6.3-3. Ovakav model nam daje rješenje za tlak i brzinu protoka na rubovima svakog segmenta. Ta činjenica ne predstavlja ograničenje budući da je najzanimljiviji odnos upravo onaj između izlaza zadnje cijevi i ulaza prve cijevi. Tako će npr. model s pet cijevi bez gubitaka prikazan na slici 6.2-1, imati pet grupa unaprednih (engl. *forward*) i unazadnih (engl. *backward*) kašnjenja i četiri spoja, svaki opisan svojim faktorom refleksije.



Slika 6.3-3 Spoj između dvije cijevi bez gubitaka prikazan pomoću grafa toka signala

Radi potpunog opisa ovog modela potrebno je diskutirati i rubne uvjete na prvom odnosno zadnjem segmentu sustava, kao što će biti prikazano u slijedećem poglavlju.

6.4 Rubni uvjeti modela sa spojenim cijevima bez gubitaka

Pretpostavimo da se model vokalnog trakta sastoji od N cijevi (segmenata) označenih brojevima 1 do N , gdje početak prve cijevi odgovara glasnicama, dok kraj N -te cijevi odgovara usnicama. U tom slučaju će rubni uvjet na usnicama određivati odnos između tlaka $p_N(l_N, t)$ i brzine protoka $u_N(l_N, t)$ na izlazu N -te cijevi, sa tlakom i brzinom protoka na početku otvorenog prostora kroz koji se zvučni val širi zračenjem s usnica.

Prije je već pokazano, da su u slučaju pobude s kompleksnom eksponencijalnom, kompleksna amplituda tlaka i brzine protoka volumena zraka na usnicama povezane s impedancijom zračenja Z_L , tj. prema izrazu:

$$P_N(l_N, \Omega) = Z_L U_N(l_N, \Omega) \quad (6.4-1)$$

Ako za prvu ruku ignoriramo činjenicu da je Z_L kompleksna veličina i pretpostavimo da je realna, tada se izraz (6.4-1) može u vremenskoj domeni napisati kao:

$$\frac{\rho c}{A_N} (u_N^+(t - \tau_N) + u_N^-(t + \tau_N)) = Z_L (u_N^+(t - \tau_N) - u_N^-(t + \tau_N)) \quad (6.4-2)$$

Ako iz ovog izraza izrazimo $u_N^-(t + \tau_N)$ kao funkciju od $u_N^+(t - \tau_N)$, tada slijedi:

$$u_N^-(t + \tau_N) = -r_L u_N^+(t - \tau_N) \quad (6.4-3)$$

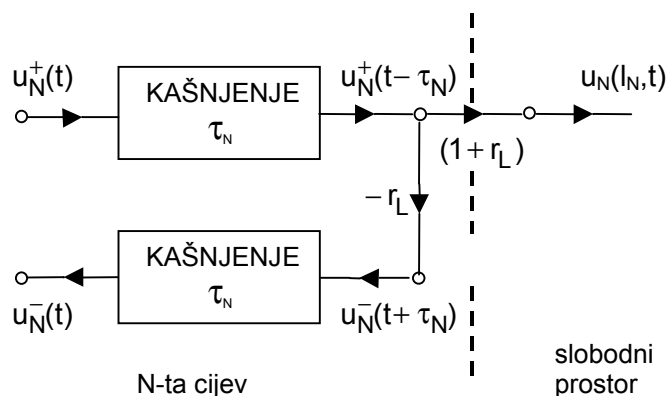
gdje je faktor refleksije na usnicama r_L definiran slijedećim izrazom:

$$r_L = \frac{(\rho c)/A_N - Z_L}{(\rho c)/A_N + Z_L} \quad (6.4-4)$$

Prema tome brzina protoka na ulazu u otvoreni prostor $u_N(l_N, t)$ odgovara razlici između pozitivno putujućeg vala $u_N^+(t - \tau_N)$ i njegove reflektirane vrijednosti $u_N^-(t + \tau_N)$, tj. prema:

$$\begin{aligned} u_N(l_N, t) &= u_N^+(t - \tau_N) - u_N^-(t + \tau_N) \\ &= (1 + r_L) u_N^+(t - \tau_N) \end{aligned} \quad (6.4-5)$$

Opisana pojava refleksije na kraju zadnje cijevi određena gornjim izrazima ilustrirana je i grafički na slici 6.4-1.



Slika 6.4-1 Rubni uvjet na usnicama (spoj zadnje cijevi i slobodnog prostora)

U slučaju da je zaključna impedancija Z_L kompleksan broj (što i jest istina), može se pokazati da tada i dalje vrijedi izraz (6.4-4) za faktor refleksije r_L , jedino što je tada i on kompleksan broj. U tom slučaju se izraz (6.4-3) mora zapisati u frekvencijskoj domeni, što vodi na diferencijalnu jednadžbu koja povezuje $u_N^-(t + \tau_N)$ i $u_N^+(t - \tau_N)$.

Analiza za rubni uvjet na glasnicama je vrlo slična prethodnom slučaju. Uz pretpostavku da se pobudni signal može odvojiti od vokalnog trakta, tj. modelirati kao paralelni spoj idealnog strujnog izvora $u_G(t)$ i glotalne impedancije Z_G koji su spojeni na ulaz prve cijevi kao što je prikazano na slici 5.5-5, tada su prema izrazu (5.5-2) tlak i brzina protoka na ulazu u prvu cijev u frekvencijskoj domeni vezani slijedećim izrazom:

$$U_1(0, \Omega) = U_G(\Omega) - P_1(0, \Omega) / Z_G \quad (6.4-6)$$

Ako ponovno pretpostavimo da je impedancija Z_G realna, tada se dobiva slijedeći izraz u vremenskoj domeni:

$$u_1^+(t) - u_1^-(t) = u_G(t) - \frac{\rho c}{A_1} \left(\frac{u_1^+(t) + u_1^-(t)}{Z_G} \right) \quad (6.4-7)$$

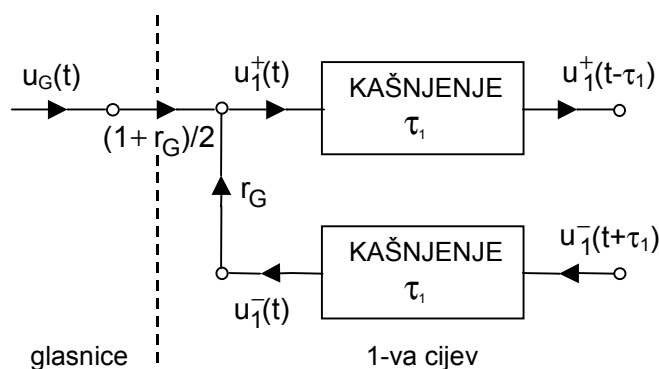
Ako izrazimo $u_1^+(t)$ kao funkciju od $u_1^-(t)$ i pobude $u_G(t)$, tada slijedi:

$$u_1^+(t) = \left(\frac{1 + r_G}{2} \right) u_G(t) + r_G u_1^-(t) \quad (6.4-8)$$

gdje je faktor refleksije na glasnicama r_G definiran slijedećim izrazom:

$$r_G = \frac{Z_G - (\rho c) / A_1}{Z_G + (\rho c) / A_1} \quad (6.4-9)$$

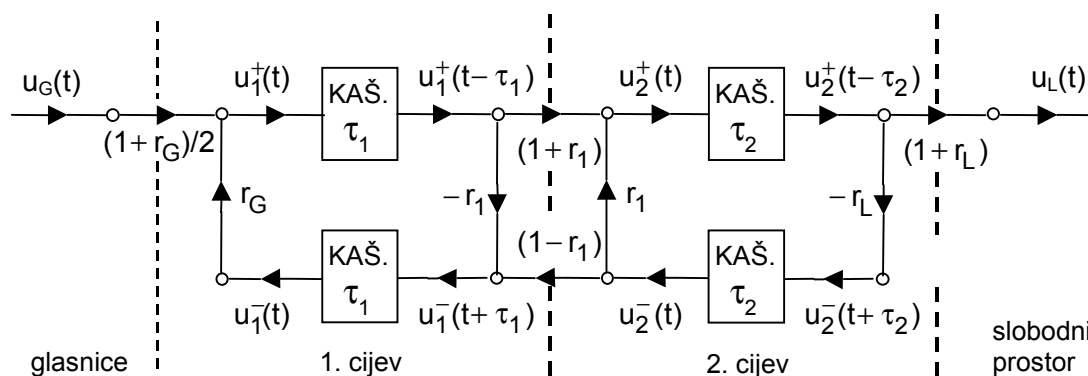
Odnosi opisani jednadžbama (6.4-8) i (6.4-9) mogu se predočiti i grafički, kao što je prikazano na slici 6.4-2.



Slika 6.4-2 Rubni uvjet na glasnicama (spoj glasnica i prve cijevi)

Analogno kao i kod rubnog uvjeta na usnicama i ovdje izraz (6.4-9) za faktor refleksije r_G vrijedi čak i za kompleksnu glotalnu impedanciju Z_G . U tom slučaju se izraz (6.4-8) mora zapisati u frekvencijskoj domeni, što vodi na diferencijalnu jednadžbu koja povezuje $u_1^+(t)$ sa $u_1^-(t)$ i $u_G(t)$.

Sada kada su poznati svi elementi ovog modela, moguće je ilustrirati cjelokupni sustav na primjeru sustava s dvije spojene cijevi, kao što je prikazano na slici 6.4-3.



Slika 6.4-3 Ukupni model za sustav s dvije spojene cijevi

Izlaz tog modela predstavlja brzina protoka volumena zraka na usnicama $u_L(t) = u_2(l_2, t)$, dok je ulaz pobudni signal $u_G(t)$. U slučaju pobudnog signala oblika kompleksne eksponencijale moguće je odrediti frekvencijsku karakteristiku cijelog sustava, tj. omjer kompleksne amplitude brzine protoka na izlazu $U_L(\Omega)$ i amplitude na ulazu $U_G(\Omega)$. Raspisivanjem svih izraza opisanih u prethodnim poglavljima u frekvencijskoj domeni, frekvencijska karakteristika dvocijevnog sustava se nalazi kao:

$$\begin{aligned} V_a(\Omega) &= \frac{U_L(\Omega)}{U_G(\Omega)} \\ &= \frac{0.5(1+r_G)(1+r_L)(1+r_1) \cdot e^{-j\Omega(\tau_1+\tau_2)}}{1+r_1r_G \cdot e^{-j\Omega 2\tau_1} + r_1r_L \cdot e^{-j\Omega 2\tau_2} + r_Lr_G \cdot e^{-j\Omega 2(\tau_1+\tau_2)}} \end{aligned} \quad (6.4-10)$$

Potrebno je uočiti da član $e^{-j\Omega(\tau_1+\tau_2)}$ koji se javlja u brojniku ove frekvencijske karakteristike pomnožen s konstantom predstavlja kašnjenje cjelokupnog sustava od glasnice do usnica u iznosu od $(\tau_1+\tau_2)$. Ova frekvencijska karakteristika se može transformirati u prijenosnu funkciju sustava $V_a(s)$ jednostavnom zamjenom $j\Omega$ sa s , kao što slijedi:

$$V_a(s) = \frac{0.5(1+r_G)(1+r_L)(1+r_1) \cdot e^{-s(\tau_1+\tau_2)}}{1+r_1r_G \cdot e^{-s2\tau_1} + r_1r_L \cdot e^{-s2\tau_2} + r_Lr_G \cdot e^{-s2(\tau_1+\tau_2)}} \quad (6.4-11)$$

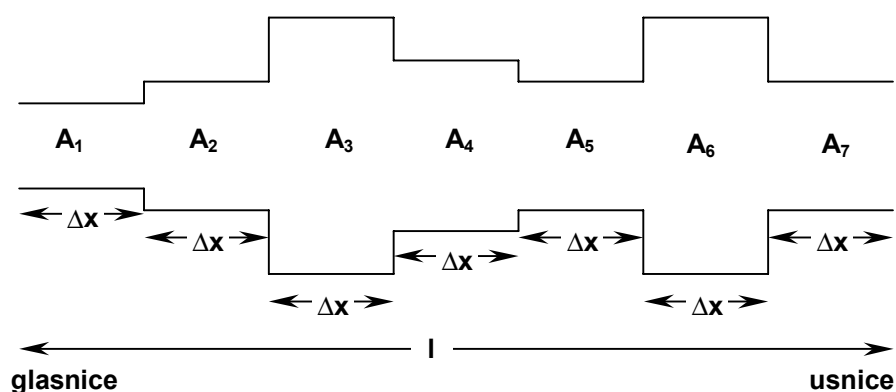
Polovi prijenosne funkcije $V_a(s)$ određuju rezonantne karakteristike modela sa spojenim cijevima bez gubitaka. Nazivnik ove funkcije je malo neobičan, tj. umjesto da se u nazivniku nalazi običan polinom u varijabli s čije korijene je lako naći, kompleksna frekvencija s se javlja u eksponentu. Poznato je da se eksponencijalna funkcija može raspisati kao beskonačni red potencija, što drugim riječima znači da je polinom u nazivniku beskonačnog stupnja i da ima beskonačno mnogo korijena. Međutim, pokazalo se da se uz pogodno odabrane dužine segmenata i njihove poprečne presjeke dobiva prijenosna funkcija $V_a(s)$ koja dobro aproksimira stvarnu akustičku prijenosnu funkciju za veći broj samoglasnika.

7. VREMENSKI DISKRETNİ AKUSTIČKI MODEL

U prethodim poglavljima pokazano je kako je moguće korištenjem relativno jednostavnog matematičkog modela opisati proces nastajanja govornog signala. Uveden je model s spojenim cijevima kojima se aproksimira stvarni oblik vokalnog trakta, te su određene prijenosne funkcije takvih modela. Cijela analiza je provedena uz pretpostavku da su svi signali u sustavu vremenski kontinuirani. Sa stanovišta stvarne primjene, mnogo je interesantnije odrediti sličan model, ali za slučaj vremenski diskretnih signala. Za bilo kakvu digitalnu obradu potrebno je otipkati govorni signal. Modeliranje u vremenski diskretnoj domeni vodit će na određivanje vremenski diskretne pobude koja propuštena kroz vremenski diskretni rekurzivni filter s pogodno odabranim koeficijentima producira izlazni vremenski diskretni signal koji se što bolje poklapa sa stvarnim otipkanim govornim signalom. U narednim poglavljima biti će ilustrirano kako je moguće prijeći na taj vremenski diskretni model.

7.1 Veza akustičkog modela i vremenski diskretnih filtara

Oblik prijenosne funkcije sustava sa spojenim cijevima bez gubitaka $V_a(s)$ upućuje na sličnost sa vremenski diskretnim filtrima, koji se u žargonu često nazivaju i digitalnim filtrima. Da bi to zornije prikazali razmotrimo sustav od N cijevi bez gubitaka od kojih je svaka dužine $\Delta x = L/N$ gdje je L ukupna duljina vokalnog trakta. Takav sustav prikazan je na slici 7.1-1 za $N=7$.



Slika 7.1-1 Primjer sustava od 7 povezanih cijevi bez gubitaka

Širenje zvučnog vala kroz takav sustav može se modelirati pomoću blok dijagrama prikazanih na slikama 6.3-3, 6.4-1 i 6.4-2. Vrijeme propagacije signala kroz svaku cijev je jednako i iznosi $\tau = \Delta x / c$ gdje je c brzina širenja zvuka. Interesantno je promotriti što se zbiva kada se takav sustav pobudi s delta funkcijom, tj. $u_G(t) = \delta(t)$. Taj se pobudni impuls širi kroz niz cijevi, a na svakom spoju biva djelomično propušten i djelomično reflektiran unazad. Može se pokazati da će impulsni odziv takvog sustava biti oblika:

$$v_a(t) = \alpha_0 \delta(t - N\tau) + \sum_{k=1}^{\infty} \alpha_k \delta(t - N\tau - 2k\tau) \quad (7.1-1)$$

Iz formule (7.1-1) vidljivo je da će prvi impuls do izlaza stići za $N\tau$ sekundi, a svaki sljedeći impuls dolaziti će na izlaz sa dodatnim kašnjenjem od 2τ sekundi što je posljedica refleksija na granicama cijevi. Pošto svaki reflektirani val mora na svom putu prema izlazu najmanje dva puta proći kroz svaku cijev (unazad, pa zatim ponovno unaprijed), svi reflektirani valovi se pojavljuju s vremenskim kašnjenjima koja su cjelobrojni višekratnici od

2 τ . Prebacivanje impulsnog odziva (7.1-1) u Laplace-ovu domenu dobiva se prijenosna funkcija takvog sustava slijedećeg oblika

$$V_a(s) = \sum_{k=0}^{\infty} \alpha_k e^{-s(N+2k)\tau} = e^{-sN\tau} \sum_{k=0}^{\infty} \alpha_k e^{-s2k\tau} \quad (7.1-2)$$

Zajednički faktor $e^{-sN\tau}$ predstavlja kašnjenje signala zbog prolaska kroz svih N cijevi, dok će desna suma u izrazu (7.1-2) biti označena sa $\hat{V}_a(s)$, tj. :

$$\hat{V}_a(s) = \sum_{k=0}^{\infty} \alpha_k e^{-s2k\tau} \quad (7.1-3)$$

Upravo taj član određuje rezonantna svojstva sustava, jer je zajednički faktor $e^{-sN\tau}$ jednak za sve sustave sa istim brojem segmenata, tj. ne ovisi o poprečnim presjecima cijevi. To pokazuje da se model $V_a(s)$ može rastaviti na dva dijela: na čisto kašnjenje od $N\tau$ sekundi i sustav $\hat{V}_a(s)$ s nultim kašnjenjem i impulsnim odzivom $\hat{v}_a(t) = v_a(t + N\tau)$, kao što je prikazano na slici 7.1-2 a). Frekvencijska karakteristika $\hat{V}_a(\Omega)$ tog sustava s nultim kašnjenjem je:

$$\hat{V}_a(\Omega) = \sum_{k=0}^{\infty} \alpha_k e^{-j\Omega 2k\tau} \quad (7.1-4)$$

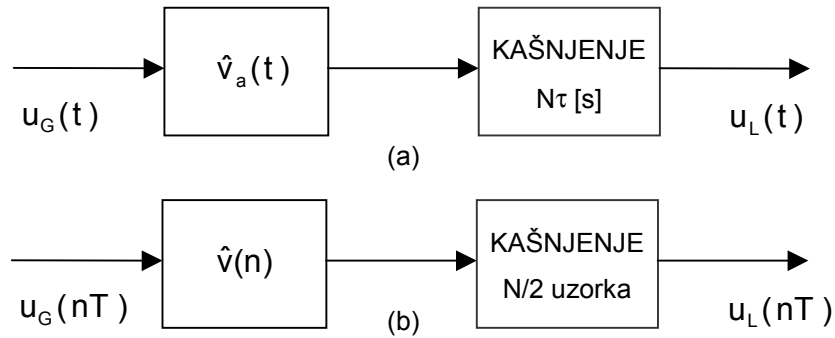
Lako je pokazati da je $\hat{V}_a(\Omega)$ periodična po Ω , i to s periodom $2\pi/2\tau$, tj. vrijedi:

$$\hat{V}_a\left(\Omega + \frac{2\pi}{2\tau}\right) = \hat{V}_a(\Omega) \quad (7.1-5)$$

Svojstvo periodičnosti frekvencijske karakteristike je karakteristično za vremenski diskretne sustave, kod kojih taj period iznosi $2\pi/T$, gdje je T period frekvencije otipkavanja. Ako bi se prilikom vremenske diskretizacije originalnog kontinuiranog modela $V_a(s)$ odabrao period otipkavanja $T=2\tau$, tada takva vremenska diskretizacija ne bi unijela nikakav gubitak informacije, tj. frekvencijske karakteristike vremenski kontinuiranog i diskretnog sustava bile bi jednake. To naravno vrijedi samo pod uvjetom da je ulazni signal imao ograničen spektar, tj. da nije imao frekvencijske komponente iznad polovine frekvencije otipkavanja. Uz tako odabran period otipkavanja, kašnjenje od $N\tau$ sekundi odgovara pomaku od $N/2$ uzorka, dok ekvivalentni vremenski diskretni filter ima impulsnim odziv $\hat{v}(n)$ koji se dobiva otipkavanjem vremenski kontinuiranog impulsnog odziva $\hat{v}_a(t)$ s periodom T , tj.:

$$\hat{v}(n) = \hat{v}_a(nT) = \begin{cases} \alpha_n & n \geq 0 \\ 0 & n < 0 \end{cases} \quad (7.1-6)$$

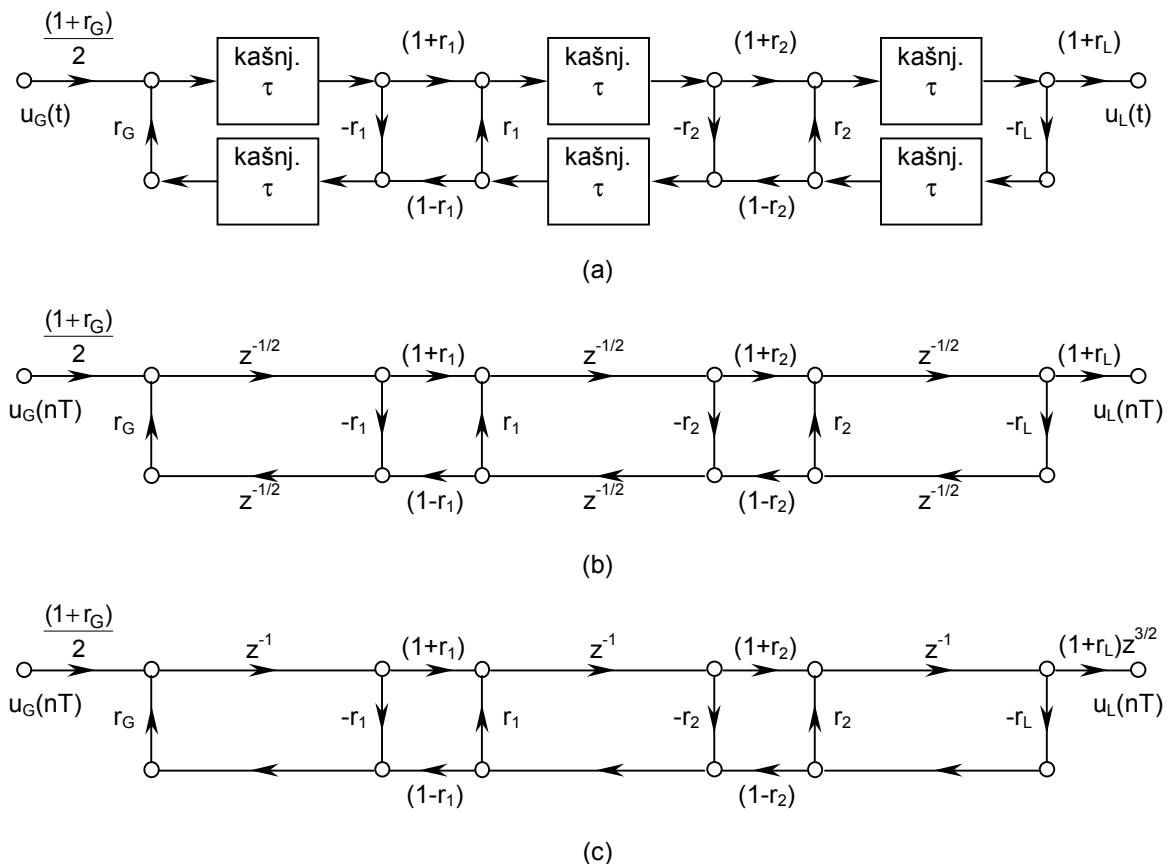
Taj vremenski diskretni sustav prikazan je na slici 7.1-2 b), a opisana transformacija vremenski kontinuiranog sustava u diskretni se naziva metoda jednakog impulsnog odziva. U slučaju kada je broj cijevi N paran broj, tada se odziv $u_L(nT)$ dobiva običnim pomakom odziva sustava $\hat{v}(n)$ na pobudu $u_G(nT)$ za $N/2$ koraka. U slučaju da je N neparan, mora se provoditi interpolacija, no vrlo često se to kašnjenje u oba slučaja zanemaruje, jer je za većinu primjena nebitno.



Slika 7.1-2 Blok dijagram akustičkog modela sa spojenim cijevima bez gubitaka u vremenski kontinuiranoj (a) i diskretnoj (b) domeni

Prema pravilima koja vrijede za transformaciju s jednakim impulsnim odzivom, z-transformacija od $\hat{v}(n)$, označena sa $\hat{V}(z)$, može se odrediti direktno na osnovu $\hat{V}_a(s)$ supstitucijom e^{sT} sa z odnosno:

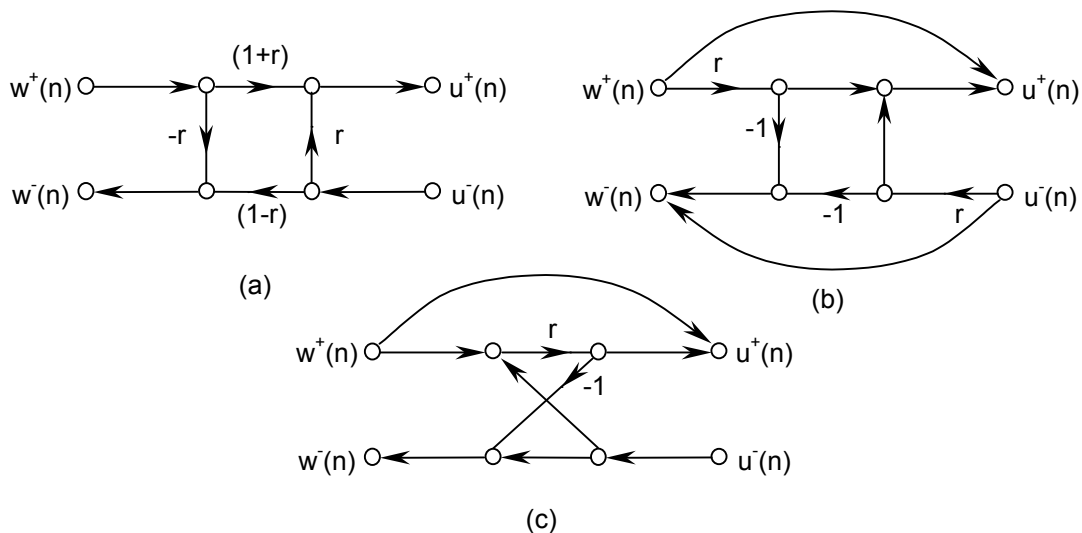
$$\hat{V}(z) = \sum_{k=0}^{\infty} \alpha_k z^{-k} \tag{7.1-7}$$



Slika 7.1-3 (a) Shema toka signala za model vokalnog trakta sa cijevima bez gubitaka; (b) ekvivalentan vremenski diskretni sustav; (c) ekvivalentan vremenski diskretni sustav sa cjelobrojnim kašnjenjima u ljestvičastoj strukturi

Schema toka signala za vremenski diskretni akustički model može se dobiti analognim načinom iz toka signala vremenski kontinuiranog modela. Vremenski kontinuirane varijable u svim čvorovima zamjenjuju se odgovarajućim diskretnim nizovima. Također svako kašnjenje od τ sekundi zamjenjuje se sa pomakom od pola koraka, budući je $\tau=T/2$. Primjer je dan na slici 7.1-3 za sustav sa tri cijevi. Kašnjenje signala kroz jednu cijev je na slici 7.1-3 b) predstavljeno prijenosnom funkcijom $z^{-1/2}$.

Pomak za $1/2$ uzorka zahtjeva interpolaciju na sredini između dva uzorka. Promatranjem ljestvičaste strukture na slici 7.1-3 b) može se uočiti da, ukoliko se kašnjenja iz donjih grana prebaci u pripadajuće gornje grane, kašnjenje unutar bilo kojeg zatvorenog kruga u strukturi ostaje nepromijenjeno. Ukupno kašnjenje od ulaza do izlaza je sada izmijenjeno, ali to nema nekog većeg značenja u praksi, a teoretski bi se moglo ispraviti umetanjem sklopa za prethođenje za $N/2$ koraka na samom izlazu, kao što je prikazanom sklopom $z^{N/2}$ na slici 7.1-3 c). Ovakvom strukturom izbjegnute su poteškoće vezane uz interpolaciju u strukturi 7.1-3 b). Još jedna prednost strukture 7.1-3 c) je u činjenici da se takav sustav može zapisati u obliku jednadžbi diferencija koje se onda mogu koristiti za iterativno izračunavanje izlaznih uzoraka na osnovu zadane pobude.



Slika 7.1-4 Varijante spoja cijevi bez gubitaka sa 4 množenja (a), sa 2 množenja (b) i sa samo 1 množenjem (c)

Digitalne mreže kao ona na slici 7.1-3 c) mogu se upotrijebiti za izračunavanje sintetičkog govornog signala iz odgovarajuće odabranog pobudnog signala. Kod takvih primjena struktura mreže određuje složenost operacija potrebnih za proračun svakog izlaznog uzorka. U modelu sa slike 7.1-3 na svakom spoju dviju cijevi potrebno je izvršiti 4 množenja i 2 zbrajanja dakle ukupno $4N$ množenja i $2N$ zbrajanja za N cijevi. Kako je množenje vremenski najzahtjevnije bilo bi pogodno kada bi se koristile strukture koje uključuju manji broj množenja. Do takvih struktura može se doći izmjenama strukture svakog spojnog mjesta prikazanog na slici 7.1-4 a) prema pravilima o transformacijama toka signala. Spojno mjesto susjednih cijevi definirano je jednadžbama diferencija :

$$u^+(n) = (1+r)w^+(n) + ru^-(n) \quad (7.1-8)$$

$$w^-(n) = -rw^+(n) + (1-r)u^-(n) \quad (7.1-9)$$

koje su grafički predočene strukturom 7.1-4 a). Modifikacijama ovih jednadžbi diferencija, spojno mjesto se može izvesti i uz upotrebu dva množenja, ili čak sa samo jednim množenjem. U oba izraza (7.1-8) i (7.1-9) pojavljuju se članovi $rw^+(n)$ i $ru^-(n)$, pa ih je dovoljno izračunati samo u jednom izrazu, te zatim iskoristiti u drugom, što vodi na strukturu 7.1-4 b). Slično tome ako bi se u oba izraza grupirali članovi koji su množeni sa r , dobiva se slijedeći par jednadžbi diferencija:

$$u^+(n) = w^+(n) + r(u^-(n) + w^+(n)) \quad (7.1-10)$$

$$w^-(n) = u^-(n) - r(u^-(n) + w^+(n)) \quad (7.1-11)$$

Iz izraza (7.1-10) i (7.1-11) je vidljivo da je moguće prvo zbrojiti $w^+(n)$ i $u^-(n)$, sumu pomnožiti sa r , te zatim taj član iskoristiti u obje jednadžbe, što vodi na strukturu prikazanu na slici 7.1-4 c).

7.2 Prijenosna funkcija vremenski diskretnog modela u z-domeni

U prošlom poglavlju pokazano je kako je moguće odrediti prijenosnu funkciju vremenski diskretnog modela govornog trakta $V(z)$. Impulsni odziv vremenski kontinuiranog modela $v_a(t)$ otipkan je s periodom $T=2\tau$, te su tako dobiveni uzorci $v(n)$ z-transformacijom prebačeni u z-domenu, čime je dobiven $V(z)$. Konkretno, $V(z)$ se nalazi tako da se $\hat{V}(z)$ iz izraza (7.1-7) pomnoži sa $z^{-N/2}$.

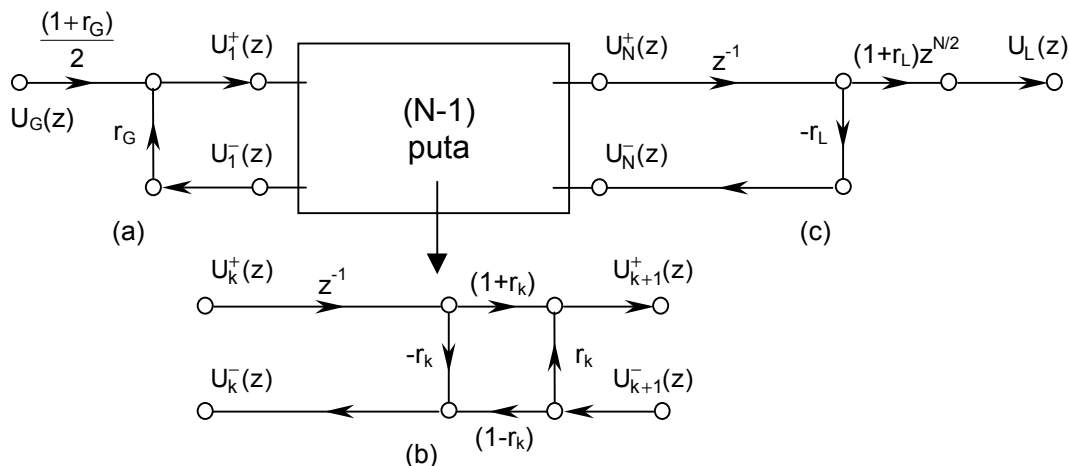
Tako dobiveni $V(z)$ je funkcija uzoraka impulsnog odziva, no mnogo je interesantnije odrediti $V(z)$ kao funkciju koeficijenata refleksije r_G, r_1 do r_{N-1} i r_L . Obzirom da je za općenit slučaj impulsni odziv $v_a(t)$ beskonačnog trajanja (signal se beskonačno puta reflektira na svim spojevima) impulsni odziv diskretnog modela $v(n)$ će također biti beskonačan, pa suma u izrazu (7.1-7) stvarno mora ići do beskonačnosti. Po tom svojstvu beskonačnog trajanja, takav impulsni odziv je sličan impulsnom odzivu rekurzivnog digitalnog filtra (IIR filter). Poznato je da se prijenosna funkcija IIR filtera može prikazati u kompaktnoj formi u obliku razlomljene racionalne funkcije od z , koja u brojniku i nazivniku ima polinom u varijabli z konačnog stupnja. Po analogiji sa IIR filterima, interesantno je i za vremenski diskretni model vokalnog trakta odrediti takvu 'kompaktnu' prijenosnu funkciju, jer je iz tog oblika moguće odrediti brojna svojstva modela (npr. centralne frekvencije i širine rezonantnih karakteristika itd.). Očekuje se da bi koeficijenti polinoma u brojniku i nazivniku trebali biti isključivo funkcija koeficijenata refleksije.

Dijagram toka signala vremenski diskretnog modela vokalnog trakta koji je prikazan na slici 7.1-3 c) može se rastaviti na kaskadu više segmenata, kao što je prikazano na slici 7.2-1. Prva sekcija te kaskade prikazana na 7.2-1 a) određena je rubnim uvjetom na glasnicama. Iza nje slijede $N-1$ sekcija 7.2-1 b) koje su određene koeficijentima refleksije na svakom od $N-1$ spojeva cijevi. Zadnja sekcija je prikazana na 7.2-1 c), a određena je rubnim uvjetom na usnicama. Na slici 7.2-1 uvedene su i oznake za z-transformacije signala u svim čvorovima te strukture. Prijenosna funkcija $V(z)$ ovog modela definirana je kao kvocijent $U_L(z)$ i $U_G(z)$, pa je prema tome potrebno nekako dovesti u vezu te dvije veličine. Do ove veze je najlakše doći raspisivanjem prijenosnih funkcija sekcija od izlaza prema ulazu. Analizom strukture na 7.2-1 c) slijedi par jednadžbi:

$$U_N^-(z) = -r_L z^{-1} U_N^+(z), \quad U_L(z) = (1 + r_L) z^{N/2} z^{-1} U_N^+(z) \quad (7.2-1)$$

Ako bi iz ove dvije jednadžbe izrazili $U_N^+(z)$ i $U_N^-(z)$, kao funkciju izlaza sustava $U_L(z)$ dobiva se slijedeći odnos:

$$U_N^+(z) = z^{1-N/2} \frac{1}{1+r_L} U_L(z), \quad U_N^-(z) = -z^{-N/2} \frac{r_L}{1+r_L} U_L(z) \quad (7.2-2)$$



Slika 7.2-1 Blok shema vremenski diskretnog modela vokalnog trakta u z -domeni razbijena na kaskadu sekcija

U nastavku ove analize će uvijek par signala $U_k^+(z)$ i $U_k^-(z)$ biti izražen kao funkcija ulaznih signala slijedeće sekcije $U_{k+1}^+(z)$ i $U_{k+1}^-(z)$, pa je zbog toga je pogodno svaki par signala ujediniti u dvodimenzionalne vektore, koji će biti označeni kao $\mathbf{U}_k(z)$, odnosno $\mathbf{U}_{k+1}(z)$. Koristeći takav način označavanja, jednadžba (7.2-2) se može zapisati kao:

$$\mathbf{U}_N(z) = \begin{bmatrix} U_N^+(z) \\ U_N^-(z) \end{bmatrix} = z^{1-N/2} \frac{1}{1+r_L} \begin{bmatrix} 1 \\ -r_L z^{-1} \end{bmatrix} U_L(z) \quad (7.2-3)$$

Slično opisanom postupku, analiza sekcije 7.2-1 b) daje slijedeći par jednadžbi:

$$U_{k+1}^+(z) = (1+r_k)z^{-1}U_k^+(z) + r_k U_{k+1}^-(z), \quad U_k^-(z) = -r_k z^{-1}U_k^+(z) + (1-r_k)U_{k+1}^-(z) \quad (7.2-4)$$

Ako izrazimo $U_k^+(z)$ i $U_k^-(z)$ kao funkciju $U_{k+1}^+(z)$ i $U_{k+1}^-(z)$ dobiva se par jednadžbi u matricnoj formi oblika:

$$\mathbf{U}_k(z) = \begin{bmatrix} U_k^+(z) \\ U_k^-(z) \end{bmatrix} = \frac{z}{1+r_k} \begin{bmatrix} 1 & -r_k \\ -r_k z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} U_{k+1}^+(z) \\ U_{k+1}^-(z) \end{bmatrix} = \frac{z}{1+r_k} \mathbf{Q}_k \cdot \mathbf{U}_{k+1}(z) \quad (7.2-5)$$

Izraz (7.2-5) se primjenjuje unazad preko svih $N-1$ sekcija b) sve do prve sekcije koja povezuje $\mathbf{U}_1(z)$ sa $\mathbf{U}_2(z)$ uz pomoć matrice \mathbf{Q}_1 . Da bi konačno doveli u vezu $U_L(z)$ i $\mathbf{U}_G(z)$, potrebno je izraziti ulazni signal $U_G(z)$ kao funkciju $\mathbf{U}_1(z)$, što se postiže analizom ulazne sekcije 7.2-1 a), koja daje slijedeću jednadžbu:

$$U_1^+(z) = r_G U_1^-(z) + \frac{1+r_G}{2} U_G(z) \quad (7.2-6)$$

odnosno u matričnoj formi:

$$\mathbf{U}_G(z) = \frac{2}{1+r_G} [1 \quad -r_G] \mathbf{U}_1(z) \quad (7.2-7)$$

Kombiniranjem izraza (7.2-7), (7.2-5) i (7.2-3) konačno slijedi željena veza:

$$\mathbf{U}_G(z) = \frac{2}{(1+r_G)(1+r_L) \prod_{k=1}^{N-1} (1+r_k)} z^{N/2} [1 \quad -r_G] \left(\prod_{k=1}^{N-1} \mathbf{Q}_k \right) \begin{bmatrix} 1 \\ -r_L z^{-1} \end{bmatrix} \mathbf{U}_L(z) \quad (7.2-8)$$

Na osnovu izraza (7.2-8) moguće je zaključiti da se općenita prijenosna funkcija vremenski diskretnog modela sa cijevima bez gubitaka može se prikazati u obliku :

$$V(z) = \frac{0.5(1+r_G)(1+r_L) \prod_{k=1}^{N-1} (1+r_k) z^{-N/2}}{D(z)} \quad (7.2-9)$$

gdje je $D(z)$ polinom od z^{-1} koji se dobije umnoškom matrica :

$$D(z) = [1 \quad -r_G] \begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} \cdots \begin{bmatrix} 1 & -r_{N-1} \\ -r_{N-1} z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ -r_L z^{-1} \end{bmatrix} \quad (7.2-10)$$

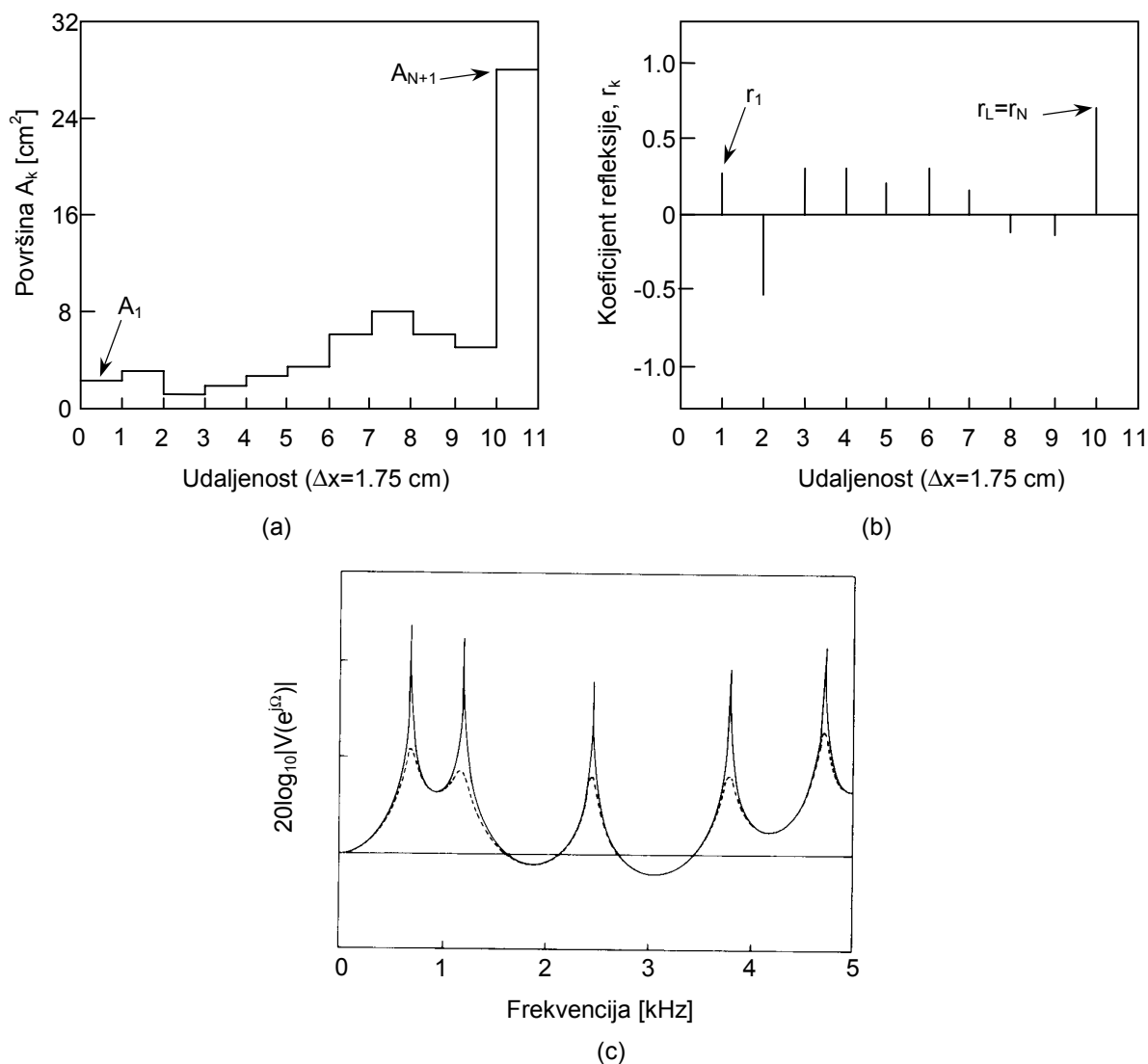
Obzirom da su svi elementi matrica \mathbf{Q}_k ili konstanta ili konstanta množena sa z^{-1} nije teško pokazati da će produkt matrica iz izraza (7.2-10) biti polinom N -tog stupnja u varijabli z^{-1} , tj. oblika:

$$D(z) = 1 - \sum_{k=1}^N \alpha_k z^{-k} \quad (7.2-11)$$

Razmatrajući jednadžbe (7.2-9) i (7.2-11) može se zaključiti da model sa cijevima bez gubitaka ima kašnjenje proporcionalno broju segmenata (cijevi), te da prijenosna funkcija nema nule nego isključivo polove (engl. *all-pole system*). U brojniku se nalazi konstanta pomnožena sa $z^{-N/2}$ što odgovara običnom pomaku za $N/2$ koraka, tj. brojnik će biti jednak nuli samo u trivijalnom slučaju kada z teži u beskonačno. Za razliku od toga, polovi postoje i njihove pozicije su određene korijenima polinoma $D(z)$, čime su ujedno određene i rezonantne (formantne) karakteristike modela vokalnog trakta.

Gubitci u modelu vokalnog trakta sa idealnim cijevima bez gubitaka određeni su isključivo sa zaključnim impedancijama Z_G , odnosno Z_L . Na osnovu njih su izračunati faktori refleksije r_G i r_L , koji figuriraju u modelu na slikama 7.1-3 i 7.2-1. U slučaju kad su ovi faktori po modulu jednaki jedan, tada ne postoje realni gubitci u modelu, pa se signal u vokalnom traktu beskonačno puta reflektira sa ulaza na izlaz bez gubitka energije. U tom slučaju će sve rezonantne karakteristike biti neprigušene, tj. širine formanata biti će jednake nuli. Radi jednostavnosti, moguće je pretpostaviti da je ulazni faktor refleksije r_G jednak 1, te zatim gubitke regulirati isključivo sa r_L . Ovu zaključnu impedanciju odnosno pripadni faktor refleksije na usnicama moguće je modelirati tako da se na sustav od N cijevi doda još jedna cijev beskonačne dužine, poprečnog presjeka A_{N+1} , kojom se modelira slobodni prostor ispred usnica. Obzirom da je ova cijev beskonačne dužine, neće biti niti povratnog vala, što je u skladu s blok shemom 7.2-1 c). Odabirom površine A_{N+1} , moguće je podesiti da taj zadnji faktor refleksije $r_N=r_L$ uzrokuje gubitke odnosno širine formanata koje su bliske stvarnim.

Ukoliko se, na primjer, odabere $A_{N+1}=\infty$ tada je $r_N=r_L=1$ što predstavlja akustički kratki spoj na izlazu i samim time model bez gubitaka.



Slika 7.2-2 Poprečni presjeci deset segmenata za samoglasnik 'a' zaključeni sa zaključnim segmentom površine $A_{N+1}=30\text{cm}^2$ (a) koeficijenti refleksije za 10 segmentnu cijev, (b); frekvencijska karakteristika 10 segmentne cijevi (crtkana krivulja odgovara uvjetima pod b), dok puna krivulja odgovara slučaju $r_L=1$), (c)

Izbor broja segmenata ovisi o odabranoj frekvenciji otipkavanja f_s vremenski diskretnog modela. Obzirom da je frekvencijska karakteristika modela sa cijevima bez gubitaka periodična, modelom je moguće aproksimirati ponašanje vokalnog trakta samo u frekvencijskom pojasu od 0 do $f_s/2$. Period otipkavanja $T=1/f_s$, već je prije odabran da iznosi $T=2\tau$, gdje je τ vrijeme propagacije kroz jednu cijev u jednom smjeru. Ukoliko model ima N segmenata, a ukupna duljina trakta iznosi l , tada je $\tau=l/(cN)$. Ukoliko se za primjer odabere $l=0.175$ m i $c=350$ m/s proizlazi da je

$$f_s = \frac{1}{T} = \frac{1}{2\tau} = \frac{Nc}{2l} = N \cdot 1000 \text{ [Hz]} \quad (7.2-12)$$

To znači da uz tipičnu dužinu vokalnog trakta i poznatu frekvenciju otipkavanja, broj segmenata N se nalazi tako da se f_s podijeli sa 1kHz. Budući je polinom u nazivniku prijenosne funkcije $D(z)$ stupnja N , to znači da sustav može imati najviše $N/2$ konjugirano kompleksnih parova polova od kojih svaki par formira jednu rezonantnu karakteristiku u pojasu 0 do $f_s/2$. Tako npr. ukoliko je $f_s=10$ kHz, tada je frekvencijski pojas signala od 0 do 5 kHz, a obzirom da je $N=10$, to ujedno povlači da će biti 5 rezonantnih karakteristika u tom frekvencijskom području. Što je manja ukupna duljina vokalnog trakta l , to je manji broj rezonancija po kHz frekvencijskog pojasa signala, pa i broj segmenata N uz istu frekvenciju otipkavanja može biti manji.

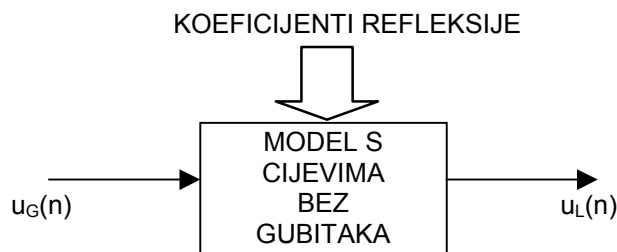
Slika 7.2-2 prikazuje primjer određivanja frekvencijske karakteristike vremenski diskretnog modela uz poznate površine poprečnih presjeka, za $N=10$ i $f_s=10$ kHz. Na slici 7.2-2 a) prikazane su površine poprečnih presjeka segmenata za samoglasnik 'a'. Slika 7.2-2 b) prikazuje odgovarajuće koeficijente refleksije. Uz odabrani zaključni segment $A_{N+1}=A_{11}=30\text{cm}^2$, koeficijent refleksije na usnicama iznosi $r_L=r_N=0.714$. Najveći koeficijenti refleksije po modulu nalaze se na spojevima gdje je promjena površine najveća. Slika 7.2-2 c) prikazuje frekvencijske karakteristike za $r_L=1$ (puna linija) i $r_L=0.714$ (crtkano). Usporedbom crtkane karakteristike i karakteristike detaljnijih modela vokalnog trakta može se primijetiti da su oni dosta slični odnosno da model vokalnog trakta sa cijevima bez gubitaka, a s gubitkom samo na usnicama dosta dobro aproksimira stvarnu frekvencijsku karakteristiku vokalnog trakta.

8. DIGITALNI MODEL GOVORNOG SIGNALA

8.1 Model izvor-sustav

Već je u poglavlju 6.1 najavljeno da će se analiza procesa formiranja govornog signala ograničiti na analizu modela tipa izvor-sustav (slika 6.1-1), čija je prijenosna funkcija sa ulaza na izlaz što sličnija stvarnoj prijenosnoj funkciji vokalnog trakta. Tim se modelom ne pokušavaju modelirati fizikalne pojave unutar samog vokalnog trakta, već isključivo obzirom na njegove krajeve. Zato se za takve modele kaže da su ekvivalentni s krajeva (engl. *terminal analog model*). U tu grupu modela ubraja se i model s spojenim cijevima bez gubitaka analiziran u poglavljima 6 i 7. Dakle, radi se o linearnom sustavu čiji izlaz ima željena govornu slična svojstva, kada se njime upravlja nizom parametara koji su na neki način povezani s procesom nastanka govora. Od posebnog su značaja vremenski diskretni modeli ekvivalentni s krajeva, opisani u poglavlju 7, jer je takvim modelima moguće predstavljanje otipkanih govornih signala.

Za generiranje signala sličnog govornom signalu, pobudni signal i rezonantna svojstva linearnog sustava moraju se mijenjati s vremenom. Pokazano je da je ta promjena spora, pa se stoga može pretpostaviti da će opća svojstva pobudnog signala i vokalnog trakta ostati nepromijenjena za većinu glasova kroz period od 10 do 20 ms. Dakle, može se zaključiti da se model govornog signala ekvivalentan s krajeva sastoji od vremenski sporo promjenjivog linearnog sustava koji je pobuđen signalom čija se svojstva također sporo mijenjaju od kvazi-periodičnih impulsa za zvučne glasove do slučajnog šuma za bezvučne glasove.



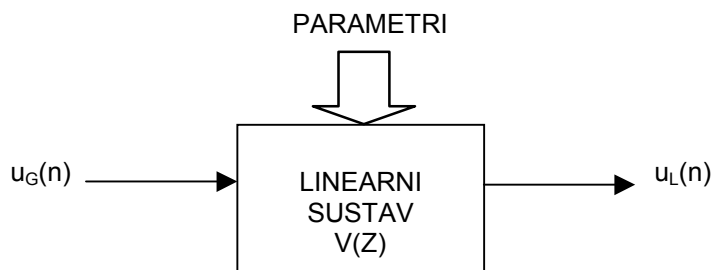
Slika 8.1-1 Blok shema modela s cijevima bez gubitaka

Tako npr. kod vremenski diskretnog modela sa spojenim cijevima bez gubitaka koji je prikazan na slici 8.1-1, parametri modela su površine poprečnih presjeka segmenata cijevi, odnosno njima pripadni koeficijenti refleksije. Promjenama tih parametara kroz vrijeme modelira se vremenska zavisnost vokalnog trakta. Korištenjem vremenski diskretnog modela kakav je prikazan na slici 7.1-3 c) moguće je odrediti odziv sustava na bilo kakvu željenu pobudu. U prethodnom poglavlju je također pokazano da je prijenosna funkcija $V(z)$ tog modela oblika:

$$V(z) = \frac{G}{1 - \sum_{k=1}^N \alpha_k z^{-k}} \quad (8.1-1)$$

gdje parametri G i $\{\alpha_k\}$ ovise o površinama poprečnih presjeka segmenata A_k uzduž vokalnog trakta (Opaska: ignorirano je kašnjenje u brojniku $V(z)$). Prijenosna funkcija $V(z)$ odgovara prijenosnoj funkciji digitalnog rekurzivnog filtra koji ima samo polove. Ista željena prijenosna funkcija može se ostvariti raznim izvedbama takvog digitalnog filtra, npr.: direktna, kaskadna, paralelna ili mrežasta forma. Struktura prikazana na slici 7.1-3 c) samo je

jedna od takvih realizacija, koje su sve ekvivalentne po prijenosnoj funkciji sa ulaza na izlaz. Stoga možemo zaključiti da će što se izlaza tiče, svaki sustav s ovom prijenosnom funkcijom davati jednak izlazni signal. (Ovo nije u potpunosti točno za vremenski promjenjive sustave, ali razlike mogu biti smanjene pažljivom implementacijom.) Prema tome, vremenski diskretan model ekvivalentan s krajeva poprima opći oblik kao što je prikazano na slici 8.1-2., a njegova stvarna 'unutarnja' izvedba može biti odabrana prema želji.



Slika 8.1-2 Blok shema općenitog vremenski diskretnog modela

Potpuni model, osim prijenosne funkcije vokalnog trakta mora uključivati i odgovarajuću reprezentaciju promjenjive pobudne funkcije $u_G(n)$ i efekte zračenja na usnicama. U nastavku ovog poglavlja svaka komponenta modela biti će posebno proučena i zatim uključena u potpuni model.

8.2 Vokalni trakt

Rezonantne karakteristike govornog signala, ili tzv. formantne karakteristike određene su polovima prijenosne funkcije $V(z)$, koja odgovara prijenosnoj funkciji digitalnog filtra bez nula. Takav model vokalnog trakta koji ima samo polove omogućuje vrlo dobru reprezentaciju efekata vokalnog trakta za većinu glasova. No, prema akustičkoj teoriji, prijenosne funkcije nazala i frikativa sadrže i polove i nule. Da bi se ostvarilo točno modeliranje i za te glasove potrebno je: ili stvarno dodati nulu u prijenosnu funkciju $V(z)$, ili pak povisiti broj polova u nazivniku. Ova druga mogućnost se češće primjenjuje, jer opći oblik $V(z)$ ostaje isti, a korištenjem većeg broja polova je do neke mjere ipak moguće modelirati nulu koja se tada nalazi između dva susjedna pola.

Kako su koeficijenti nazivnika prijenosne funkcije $V(z)$ u izrazu (8.1-1) realni, korijeni polinoma u nazivniku biti će ili realni ili u konjugirano kompleksnim parovima. Rezonantnu karakteristiku vokalnog trakta u s -domeni možemo reprezentirati s parom polova s_k i s_k^* prema izrazu:

$$s_k, s_k^* = -\sigma_k \pm j2\pi f_k \quad (8.2-1)$$

gdje su f_k i $2\pi f_k$ centralna frekvencija formanta u [Hz], odnosno u [rad/s]. Širina frekvencijskog pojasa rezonantne karakteristike vokalnog trakta približno je jednaka $2\sigma_k$ [rad/s]. Odgovarajući konjugirano kompleksni par polova u z -domeni je:

$$\begin{aligned} z_k, z_k^* &= e^{-\sigma_k T} e^{\pm j2\pi f_k T} \\ &= e^{-\sigma_k T} \cos(2\pi f_k T) \pm j e^{-\sigma_k T} \sin(2\pi f_k T) \end{aligned} \quad (8.2-2)$$

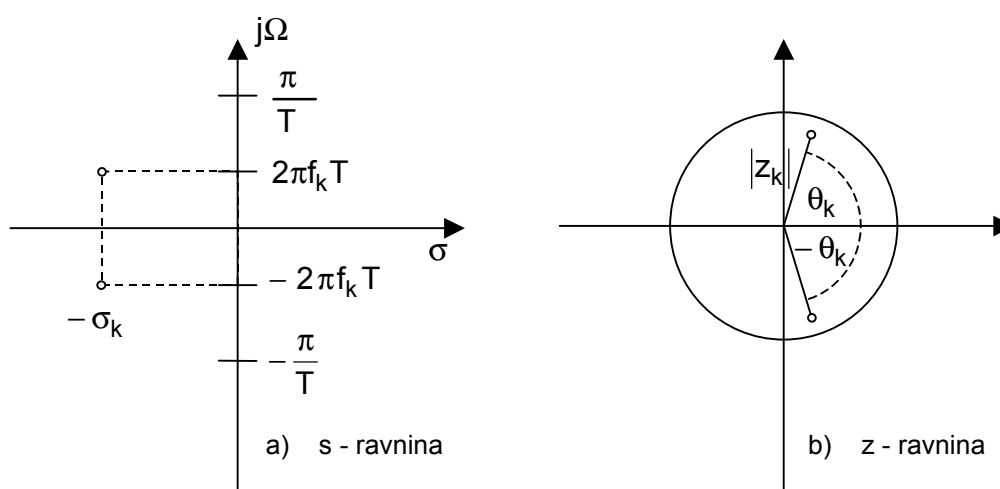
gdje je T period frekvencije otipkavanja. U z -ravnini, udaljenost pola do ishodišta definira širinu frekvencijskog pojasa:

$$|z_k| = e^{-\sigma_k T}, \quad (8.2-3)$$

dok će kut pola u z-ravnini ovisiti o f_k prema izrazu:

$$\theta_k = 2\pi f_k T \quad (8.2-4)$$

Uz poznatu prijenosnu funkciju $V(z)$, centralne frekvencije formanta f_k i njihove širine mogu se odrediti faktorizacijom nazivnika $V(z)$ te primjenom izraza (8.2-3) i (8.2-4), koji povezuju vremenski kontinuiranu i diskretnu domenu. Kao što je ilustrirano na slici 8.2-1 a), sve vlastite frekvencije vokalnog trakta nalaze se u lijevoj poluravnini s-ravnine, budući da se radi o stabilnom sustavu. Stoga je $\sigma_k > 0$, te vrijedi da je $|z_k| < 1$, tj. svi odgovarajući polovi diskretnog sustava moraju biti unutar jedinične kružnice da bi uvjet stabilnosti bio zadovoljen (slika 8.2-1 b)).



Slika 8.2-1 Položaj polova u a) s-domeni; i b) z-domeni

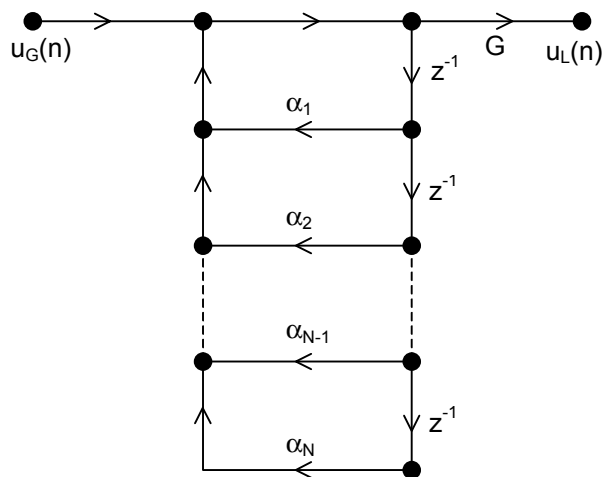
Kao što je pokazano u prethodnim poglavljima, model s cijevima bez gubitaka vodi na prijenosnu funkciju prema izrazu (8.1-1). Može se pokazati da će svi polovi pripadne prijenosne funkcije $V(z)$ biti unutar jedinične kružnice ako su površine poprečnih presjeka modela s cijevima bez gubitaka pozitivni brojevi. Isto tako može se pokazati da vrijedi i obrat, tj. da se svaka stabilna prijenosna funkcija $V(z)$ oblika kao u (8.1-1) može realizirati modelom s cijevima bez gubitaka.

Jedan od načina implementacije prijenosne funkcije $V(z)$ je korištenjem ljestvičaste strukture kao što je prikazano na slici 7.1-3 c). Drugi pristup implementaciji je upotreba standardnih struktura digitalnih filtara. Može se, na primjer, koristiti direktna realizacija prijenosne funkcije $V(z)$ kako je prikazano na slici 8.2-2. U slučaju kaskadne izvedbe, prijenosna funkcija $V(z)$ se razbija na kaskadu sustava 2. reda, tj.:

$$V(z) = \prod_{k=1}^M V_k(z) \quad (8.2-5)$$

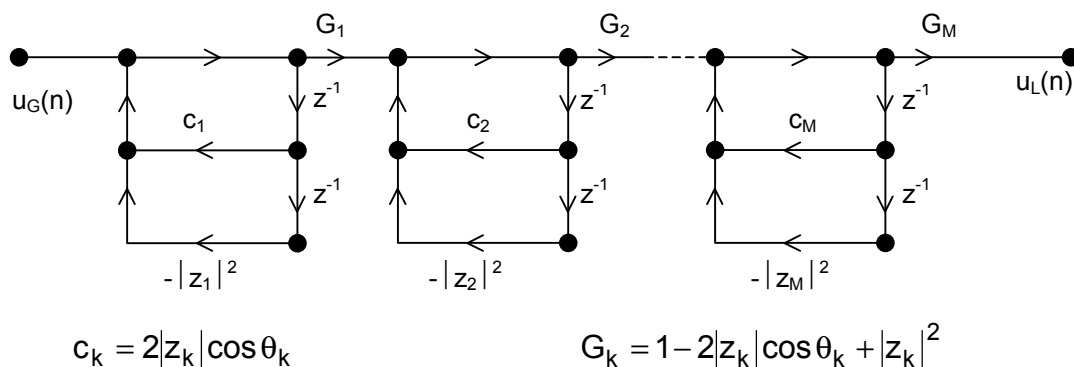
gdje je M prvi cijeli broj veći od $N/2$, a prijenosna funkcija jedne kaskade $V_k(z)$ je:

$$V_k(z) = \frac{(1 - 2|z_k| \cos(2\pi f_k T) + |z_k|^2)}{(1 - 2|z_k| \cos(2\pi f_k T) z^{-1} + |z_k|^2 z^{-2})} \quad (8.2-6)$$



Slika 8.2-2 Direktna realizacija prijenosne funkcije modela bez nula

Brojnik funkcije $V_k(z)$ odabran je tako da produkt kaskada ima isto ukupno pojačanje kao i polazna prijenosna funkcija. U slučaju kada je sustav pobuđen signalom frekvencije 0, ($z=1$), prijenosna funkcija k -te kaskade na toj frekvenciji iznosi $V_k(z)=1$. Kaskadna realizacija modela prikazana je na slici 8.2-3.



Slika 8.2-3 Kaskadna realizacija prijenosne funkcije modela bez nula

Postoji još jedan način realizacije $V(z)$, a to je rastavljanje $V(z)$ na parcijalne razlomke na osnovu kojih se izvodi paralelna realizacija modela.

8.3 Zračenje na usnicama

Do sada je razmatrana prijenosna funkcija $V(z)$ koja povezuje brzinu protoka volumena zraka na ulazu vokalnog trakta i brzinu protoka na usnicama. Obzirom da je ljudsko uho osjetljivo na promjenu tlaka, a ne na brzinu protoka, trebalo bi prikazati odnos brzine protoka volumena zraka na ulazu vokalnog trakta i tlaka na usnicama. U tom slučaju nužno je uzeti u obzir utjecaj zračenja na usnicama. Za slučaj analognog modela, odnos tlaka i brzine protoka u frekvencijskoj domeni dan je izrazom:

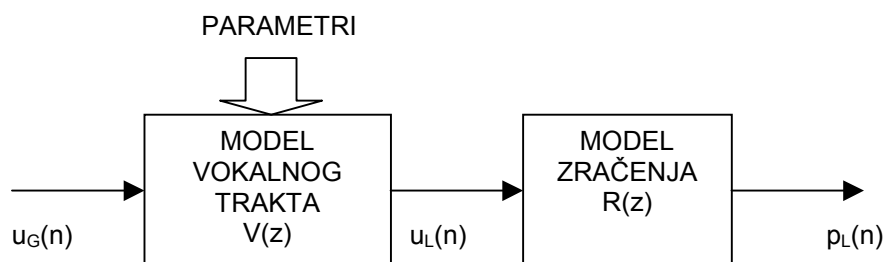
$$P(l, \Omega) = Z_L(\Omega)U(l, \Omega) \tag{8.3-1}$$

gdje je Z_L impedancija zračenja. Željeni izraz u z -domeni trebao bi imati sličan oblik:

$$P_L(z) = R(z)U_L(z) \tag{8.3-2}$$

U poglavlju 5.4 analiziran je utjecaj zračenja na usnicama za analogni model, te je pokazano su tlak i brzina protoka na usnicama vezani operacijom visoko-propusne filtracije, tj. niske frekvencije su potisnute. Zapravo se može reći da je niskim frekvencijama tlak približno jednak derivaciji brzine protoka volumena zraka. Stoga, da bi se odredio vremenski diskretni model ovog odnosa na usnicama treba upotrijebiti metodu diskretizacije koja izbjegava pojavu aliasinga. Poznato je da transformacija primjenom metode jednakih impulsnih odziva nije primjenjiva za visoko-propusne filtre. Ako bi se međutim primijenio postupak bilinearne transformacije, može se pokazati da se zadovoljavajuća aproksimacija efekta zračenja na usnicama ostvaruje prvom diferencijom:

$$R(z) = R_0(1 - z^{-1}) \quad (8.3-3)$$



Slika 8.3-1 Analogni model govornog signala uz efekt zračenja

Utjecaj zračenja na usnicama uzima se u obzir slaganjem u kaskadu postojećeg modela vokalnog trakta i modela zračenja, kao na slici 8.3-1.

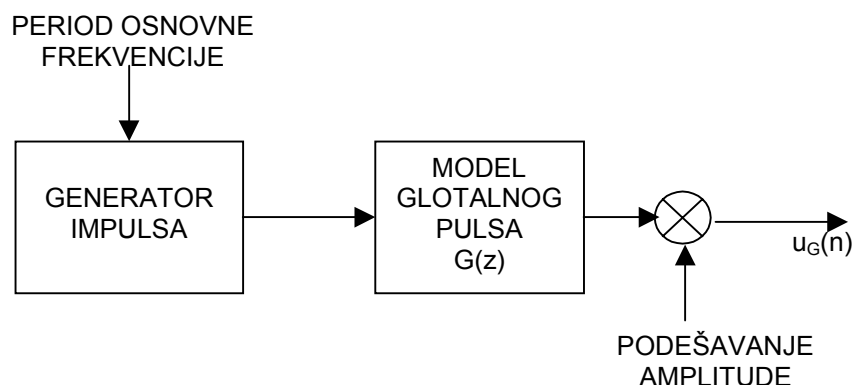
8.4 Pobudni signal

Da bi se upotpunio digitalni model formiranja govornog signala, potrebno je razmotriti i način generiranja pobudnog signala za sustav vokalnog trakta s dodanim modelom zračenja na usnicama. U poglavlju 5.5 diskutirano je da se većina govornih glasova može se podijeliti na zvučne i bezzvučne glasove. Grubo rečeno, za generiranje zvučnog glasa, na ulaz vokalnog trakta treba dovesti kvazi-periodične impulse, dok se za bezzvučni glas kao izvor signala koristi slučajni šum.

Tipični valni oblik pobudnog signala za zvučne glasove prikazan je na slici 5.5-6 a). Pobudni signal se sastoji od niza pulseva na pravilnom razmaku, a ti se pulsevi nazivaju glotalnim pulsevima. Uobičajeni način generiranja zvučnog pobudnog signala prikazan je na slici 8.4-1. Generator jediničnih impulsa daje na svom izlazu niz jediničnih impulsa u pravilnom razmaku koji odgovaraju periodu osnovne frekvencije titranja glasnica (engl. *pitch period*). Ovaj signal pobuđuje linearni sustav čiji impulсни odziv $g(n)$ odgovara otipkanom valnom obliku glotalnog pulsa. Pojačanje na izlazu, A_v , određuje intenzitet zvučne pobude.

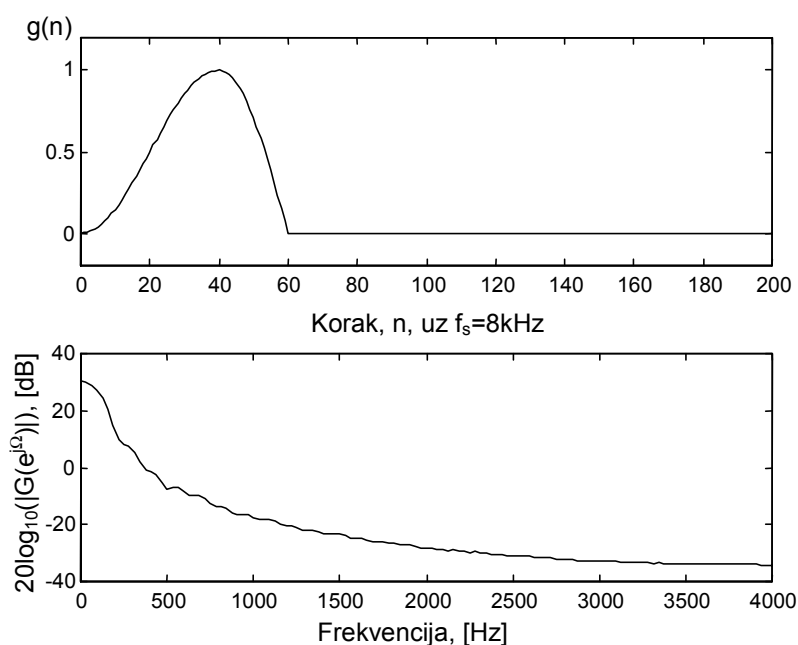
Sam oblik funkcije $g(n)$ i nije tako kritičan tako dugo dok njegova Fourierova transformacija ima željena svojstva, tj. ima oblik prijenosne funkcije niskopropusnog filtra. Tako je pokazano da prirodni glotalni puls može biti zamijenjen umjetno generiranim signalom oblika:

$$\begin{aligned} g(n) &= \frac{1}{2} [1 - \cos(\pi n / N_1)] & 0 \leq n \leq N_1 \\ &= \cos(\pi(n - N_1) / 2N_2) & N_1 \leq n \leq N_1 + N_2 \\ &= 0 & \text{inače} \end{aligned} \quad (8.4-1)$$



Slika 8.4-1 Generiranje pobudnog signala za zvučne glasove

Impulsni odziv modela glotalnog pulsa $g(n)$ prema izrazu (8.4-1) i njegova frekvencijska karakteristika, $G(e^{j\Omega})$ prikazane su na slici 8.4-2, za tipične vrijednosti $N_1=40$ i $N_2=20$. Ovakav sintetički pobudni signal je vrlo sličan stvarnom pobudnom signalu koji je prikazan na 5.5-6 a). Na slici 8.4-2 je vidljivo da frekvencijska karakteristika modela glotalnog pulsa ima željeni nisko-propusni karakter.



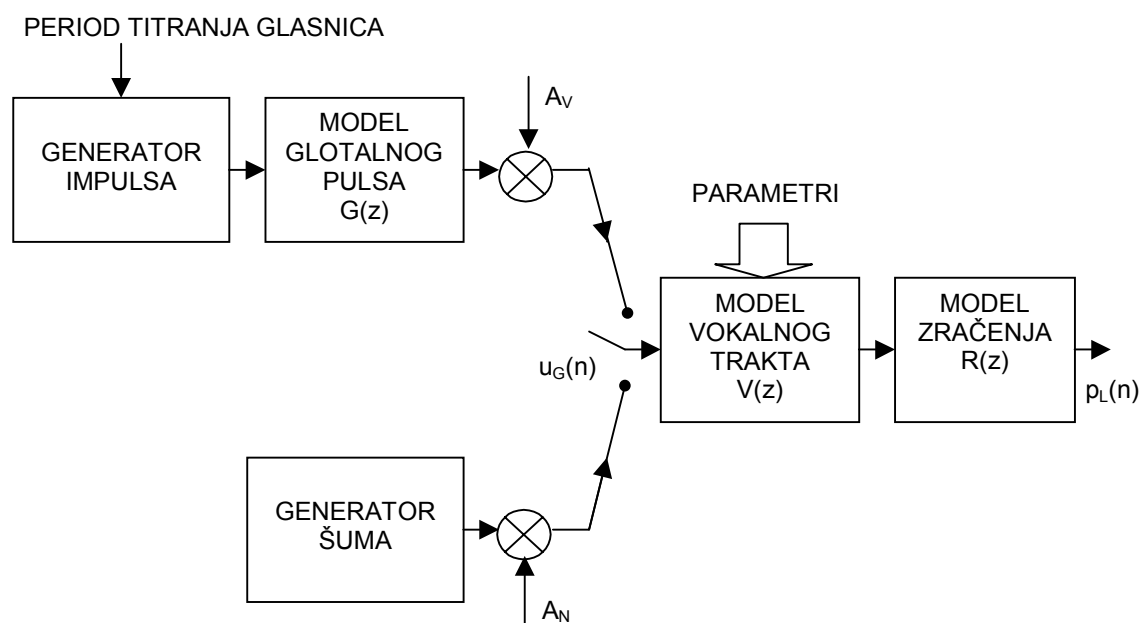
Slika 8.4-2 Aproksimacija glotalnog pulsa i modul njegove Fourier-ova transformacije

Model prema izrazu (8.4-1) odgovara vremenski diskretnom filtru s konačnim impulsnim odzivom, ili tzv. FIR filtru, pa je i njegova z -transformacija $G(z)$ jednaka polinomu u varijabli z^{-1} čiji su koeficijenti jednaki uzorcima $g(n)$. Takav $G(z)$ ima samo nule (nema polove), a već je kod izvoda prijenosne funkcije vokalnog trakta $V(z)$ diskutirano da je all-pole model interesantniji. Pokazalo se da se frekvencijska karakteristika slična onoj prikazanoj na slici 8.4-2 može ostvariti i IIR filtrom drugog reda koji ima samo polove.

Model pobudnog signala za bezvučni glas je jednostavniji. Potreban je samo izvor šuma te pojačanje na izlazu za kontrolu intenziteta bezvučne pobude. Za diskretne sustave se kao generator šuma koristi generator slučajnih brojeva. Funkcija gustoće razdiobe generatora slučajnih brojeva neće biti od presudnog značaja na kvalitetu sintetiziranog govora.

8.5 Potpuni digitalni model

Spajanjem svih prethodno opisanih dijelova u jednu cjelinu dobiva se potpuni digitalni model formiranja govornog signala (slika 8.5-1).



Slika 8.5-1 Potpuni model formiranja govornog signala

Promjena pobudnog signala postiže se preklapanjem sklopke za odabir pobude između generatora zvučnog i bezvučnog signala.

U nekim slučajevima se model glotalnog pulsa i model zračenja spajaju u jedan sustav. Kasnije će biti pokazano da se u slučaju analize govornog signala postupkom linearne predikcije, model glotalnog pulsa, model zračenja i model vokalnog trakta udružuju u jedan sustav, te su predstavljeni jedinstvenom prijenosnom funkcijom bez nula:

$$H(z) = G(z)V(z)R(z) \quad (8.5-1)$$

Dakle, slika 8.5-1 je samo jedan općenit prikaz koji je moguće na razne načine modificirati.

Pitanje koje se samo po sebi nameće tiče se ograničenja ovakvog modela. Sigurno je da je konačni model daleko od parcijalnih diferencijalnih jednadžbi od kojih se krenulo, ali na sreću ni jedno od zanemarenja ne utječe bitno na primjenjivost ovog modela.

Što se tiče utjecaja vremenske varijabilnosti parametara, u kontinuiranim glasovima kao što su samoglasnici (vokali), parametri se mijenjaju vrlo sporo i model doista dobro opisuje stvarni proces. Kod tranzijentnih glasova model nije toliko dobar, ali još uvijek zadovoljava. Kod kreiranja modela krenulo se od pretpostavke kvazi-stacionarnosti govornog signala, tj. pretpostavke da ako se govor promatra u dovoljno kratkim vremenskim odsječcima (između 10 i 20 ms), da će tada spektralna i statistička svojstva signala unutar cijelog intervala biti stalna, tj. da su parametri modela konstantni za takve vremenske periode. Tada prijenosna funkcija $V(z)$ zaista opisuje ponašanje govornog sustava čiji se parametri mijenjaju vrlo sporo s vremenom.

Drugo ograničenje modela je nedostatak nula potrebnih za reprezentaciju nazala i frikativa. Ovo ograničenje više utječe na nazale, ali se može riješiti dodavanjem nula u model. Treće ograničenje se odnosi na zvučne frikative kod kojih jednostavni model

preklapanja pobude zvučna/bezvučna nije dovoljan, jer pobudni signal za te glasove ima istovremeno i periodični i šumovit karakter. Jednostavno zbrajanje ove dvije pobude također nije dovoljno jer do pojave turbulencija i šumovitog zvuka dolazi samo u vrhovima periodičnog pobudnog signala. Zato je razvijen složeniji model koji se može upotrijebiti kada je to potrebno. Zadnje ograničenje ovog jednostavnog modela je u činjenici da period titranja glasnica mora biti cjelobrojni višekratnik perioda frekvencije otipkavanja (cijeli broj uzoraka kašnjenja), ali postoje i složeniji modeli kod kojih je i to ograničenje moguće eliminirati

8.6 Zaključak

U poglavljima 4,6,7 i 8 naglasak je bio na slijedećim temama: fizikalne osnove procesa formiranja govora, te vremenski kontinuirani i vremenski diskretni modeli za oponašanje ovog procesa. Poznavanje osnova akustičke teorije je nužno, da bi bilo moguće razumijevanje modela koji iz nje slijede. Diskutirane teme trebale bi dati uvid u osnovne karakteristike govornog signala, kao i temeljne modele koji se mogu koristiti prilikom njegove digitalne obrade.

Modeli predloženi u poglavljima 7 i 8 biti će osnova za postupke obrade govornog signala diskutirane u slijedećim poglavljima. Ove modele je moguće promatrati na dva načina: prvi se naziva postupkom analize govora, a drugi postupkom sinteze govora. U postupcima analize govornog signala težište je na algoritmima pomoću kojih je na osnovu snimljenog i otipkanog govornog signala moguće odrediti parametre diskutiranih modela na čijem se izlazu prema pretpostavci nalazi upravo taj signal. U postupcima sinteze govornog signala, model se koristi da bi se na osnovu pravilno odabranih parametara (npr. iz analize) na njegovom izlazu dobio sintetički govorni signal. Ova dva pristupa govornom modelu se često pojavljuju i istovremeno u većem broju primjena.

Uz poznavanje ovih osnovnih principa i modela, u narednim poglavljima biti će diskutirano kako je postupke digitalne obrade signala moguće primijeniti na obradu govornog signala.

9. POSTUPCI LINEARNE PREDIKCIJE GOVORNIH SIGNALA

9.1 Uvod u linearnu predikciju

Metoda linearne predikcije je jedna od najdjelotvornijih metoda analize govora. Ova metoda je postala prevladavajuća tehnika određivanja (estimacije) osnovnih parametara govora, npr.: osnovne frekvencije titranja glasnica, formanata, kratkotrajnih spektralnih svojstava, funkcije površine poprečnog presjeka vokalnog trakta, te za predstavljanje govora kod prijenosa ili pohrane korištenjem malog broja bita. Važnost ove metode je u mogućnosti pronalaženja izuzetno točnih estimacija parametara, te u relativno velikoj brzini izračunavanja, tj. maloj numeričkoj složenosti postupka. U ovom poglavlju bit će predstavljena formulacija osnovne ideje metode linearne predikcije, te diskusija o važnijim aspektima korištenja ove metode u stvarnim aplikacijama za obradu govora.

Osnovna ideja metode linearne predikcije je u pretpostavci da se uzorak signala može aproksimirati linearnom kombinacijom prethodnih uzoraka. Unikatni skup "najboljih" koeficijenata prediktora može se odrediti minimizacijom sume kvadrata razlike stvarnih uzoraka govora i uzoraka dobivenih linearnom predikcijom na ograničenom vremenskom intervalu. (Koeficijenti prediktora su težinski koeficijenti koji se koriste u linearnoj kombinaciji prethodnih uzoraka.)

Teoretska osnova postupka linearne predikcije je blisko povezana s osnovnim govornim modelom koji je opisan u prethodnim poglavljima, gdje je pokazano da govor može biti modeliran kao izlaz iz linearnog, vremenski promjenjivog sustava pobuđenog kvazi-periodičnim nizom impulsa (za vrijeme zvučnog govora), ili slučajnim šumom (za vrijeme bezvučnog govora). Metoda linearne predikcije omogućuje robustnu, pouzdanu i točnu metodu određivanja parametara koji karakteriziraju takav linearni, vremenski promjenjivi sustav.

U ovom poglavlju dan je općeniti pregled analize linearnom predikcijom, i bit će pokazano kako osnovna ideja linearne predikcije vodi ka skupini od nekoliko tehnika analize koje se mogu koristiti za određivanje parametara govornog modela. Ovaj općeniti skup tehnika analize linearnom predikcijom se često naziva kodiranje linearnom predikcijom (engl. *Linear Predictive Coding, LPC*).

Tehnike i metode linearne predikcije su već dugo poznate u inženjerskoj literaturi. Tako npr. linearna predikcija se koristi u području automatske regulacije i upravljanja gdje je poznata pod nazivom "estimacija sustava", ili pak u području teorije informacija gdje se naziva "identifikacija sustava". Ovaj zadnji naziv je posebno ilustrativan, jer ukazuje na činjenicu da je korištenjem LPC postupaka moguće odrediti koeficijente prediktora, koji jednoznačno opisuju (identificiraju) nepoznat sustav na osnovu njegovog izlaza, te ga modeliraju digitalnim rekurzivnim filtrom bez nula.

Sa stanovišta primjene u obradi govora, pojam linearne predikcije se odnosi na niz praktično ekvivalentnih formulacija problema modeliranja govornog signala. Razlike između ovih formulacija su prvenstveno u načinu gledanja na problem, a u nekim slučajevima razlike se odnose i na detalje proračuna kojim se određuju parametri prediktora. Konkretno, značajnije formulacije analize linearnom predikcijom primijenjene u obradi govora su:

1. metoda kovarijance (engl. *the covariance method*)
2. metoda autokorelacije (engl. *the autocorrelation formulation*)
3. metoda korištenjem mrežaste strukture (engl. *the lattice method*)
4. metoda inverznog filtra (engl. *the inverse filter formulation*)

5. metoda određivanja spektra (engl. *the spectral estimation formulation*)
6. metoda maksimalne sličnosti (engl. *the maximum likelihood formulation*)
7. metoda unutrašnjeg produkta (engl. *inner product formulation*)

U nastavku će biti detaljno razmotrene samo prve tri metode s obzirom da su ostale u biti jednake jednoj od te prve tri. Postupak linearne predikcije je vrlo važan u području obrade govornog signala, prvenstveno zbog činjenice što se osnovni model linearne predikcije vrlo dobro poklapa s modelima formiranja govornog signala diskutiranim u prethodnim poglavljima. Najveće težište ovog poglavlja biti će na diskusiji kako je moguće pouzdano odrediti parametre govornog modela primjenom postupka linearne predikcije, a biti će najavljene i moguće primjene ovog postupka u obradi govora.

9.2 Primjer predikcije govornog signala

Interesantno je proučiti učinkovitost postupka linearne predikcije govornog signala i prije njegove formalne definicije. Na slici 9.2-1 a) prikazan je kratki odsječak valnog oblika govornog signala $s(n)$, za glas 'u' otipkan s frekvencijom $f_s=8\text{kHz}$. Slikom 9.2-1 ilustrirano je kako se postupkom linearne predikcije izračunava predikcija uzorka $s(5)$ na osnovu $p=10$ uzoraka koji mu prethode, tj. na osnovu $s(4)$, $s(3)$, ... $s(-4)$, $s(5)$. Točnije rečeno predikcija uzorka $s(n)$, označena sa $\tilde{s}(n)$ se izračunava kao linearna kombinacija uzoraka $s(n-1)$, $s(n-2)$, ... $s(n-p)$, gdje je p stupanj (ili red) prediktora. Koeficijenti te linearne kombinacije neka su označeni sa α_1 do α_p , pa se predikcija $\tilde{s}(n)$ nalazi kao:

$$\tilde{s}(n) = \sum_{k=1}^p \alpha_k s(n-k) \quad (9.2-1)$$

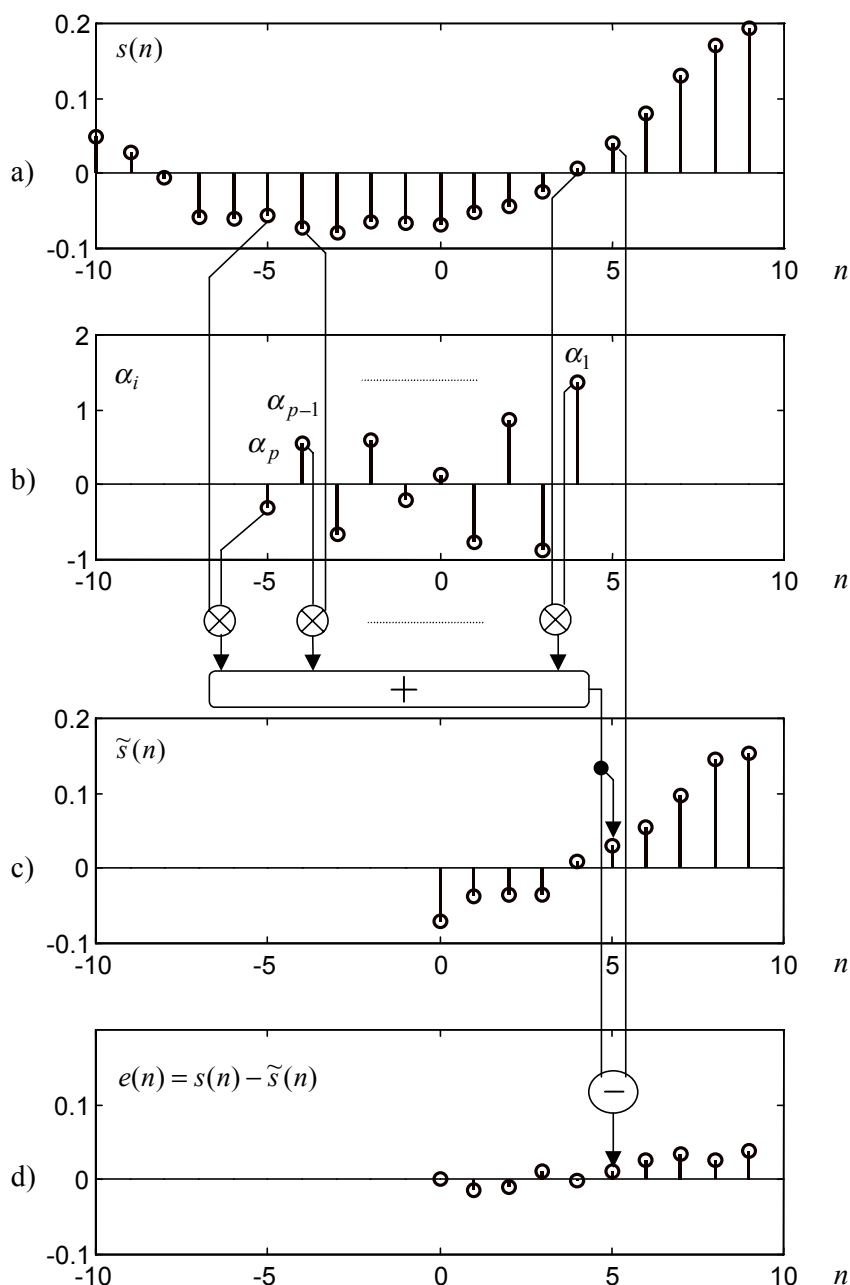
Koeficijenti prediktora su prikazani na slici 9.2-1 b) u vremenskom položaju za predikciju $n=5$ -tog uzorka. Za prvu ruku pretpostavimo da su ti koeficijenti unaprijed poznati i da su zadani tablicom 9.2-1.

Tablica 9.2-1 Koeficijenti linearnog prediktora za glas 'u'

α_{10}	α_9	α_8	α_7	α_6	α_5	α_4	α_3	α_2	α_1
-0.312	0.550	-0.664	0.585	-0.212	0.128	-0.779	0.866	-0.865	1.363

Predikcija $\tilde{s}(n)$ za uzorke $n=0$ do 9 , izračunata primjenom skupa koeficijenata iz tablice 9.2-1, prikazana je na slici 9.2-1 c) i uistinu pokazuje sličnost s originalnim signalom $s(n)$ prikazanim na 9.2-1 a). Razlika između stvarne vrijednosti n -tog uzorka i njegove predikcije $e(n)=s(n)-\tilde{s}(n)$ se naziva predikcijskom pogreškom, a prikazana je na slici 9.2-1 d). Na primjer, uzorak signala na indeksu $n=5$ iznosi $s(5)=0.0409$, predviđeni uzorak iznosi $\tilde{s}(5)=0.0297$, a predikcijska pogreška iznosi $e(5)=0.0112$.

Koeficijenti prediktora su isti (stalni) za predikciju svih uzoraka unutar nekog promatranog intervala, a cilj 'dobre' predikcije je odabrati koeficijente prediktora tako da pogreška predikcije svih uzoraka bude što je moguće manja. Kod klasičnih postupaka linearne predikcije, kao mjera kvalitete predikcije koristi se suma kvadrata predikcijske pogreške na tom promatranom intervalu. Promatrani interval redovito je višestruko duži od reda prediktora, pa stoga odabir prediktora po kriteriju najmanje kvadratne predikcijske pogreške predstavlja određeno kompromisno rješenje koje je "podjednako dobro" ili "podjednako loše" za sve uzorke tog intervala.

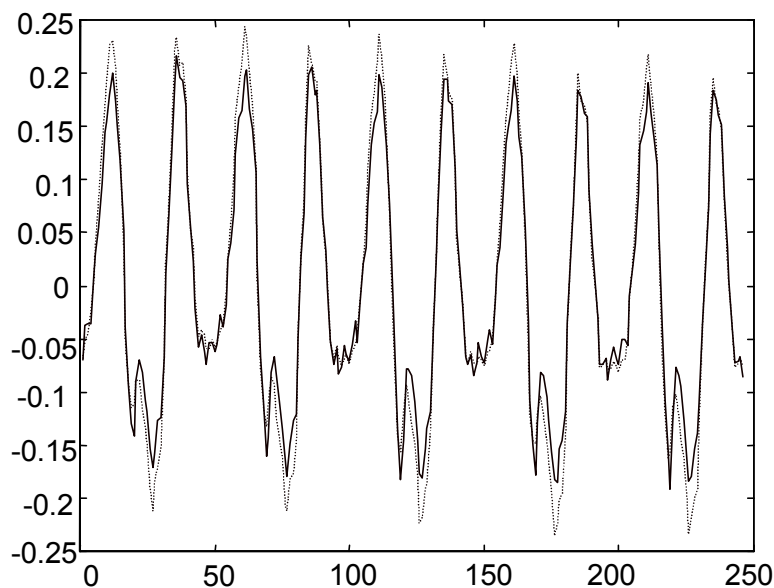


Slika 9.2-1 a) uzorci promatranog signala, b) koeficijenti linearnog prediktora, c) linearna predikcija signala, d) pogreška predikcije

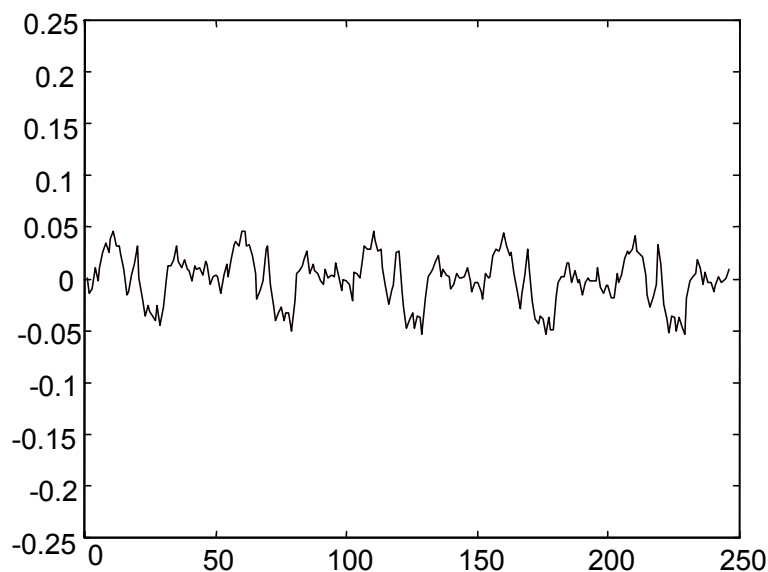
Na slici 9.2-2 prikazani su originalni govorni signal za glasu 'u' i njegova linearna predikcija dobivena koeficijentima prediktora iz tablice 9.2-1. Na slici je vidljivo vrlo dobro poklapanje ova dva signala, tj. predikcijska pogreška prikazana slikom 9.2-3 vrlo je mala.

Za signale kod kojih je energija predikcijske pogreške značajno manja od energije polaznog signala kažemo da su korelirani, tj. imaju spektar koji nije ravan nego ima izražene maksimume. Postupak linearne predikcije je posebno učinkovit za takvu klasu signala, a govorni signal se obzirom na izraženu formantnu strukturu svakako ubraja u tu klasu (energija signala je u spektru lokalizirana u svega nekoliko rezonantnih karakteristika, tj. maksimuma). Glava tema narednih poglavlja bit će kako na osnovu snimljenih uzoraka govornog signala odrediti koeficijente "dobrih" prediktora koji "minimiziraju" predikcijsku

pogrešku, te kako tako dobivene koeficijente prediktora dovesti u vezu s parametrima digitalnog modela vokalnog trakta predloženog u poglavljima 7 i 8.



Slika 9.2-2 Valni oblik promatranog signala (crtkano) i njegove linearne predikcije (punom linijom), glas 'u', $f_s=8\text{kHz}$



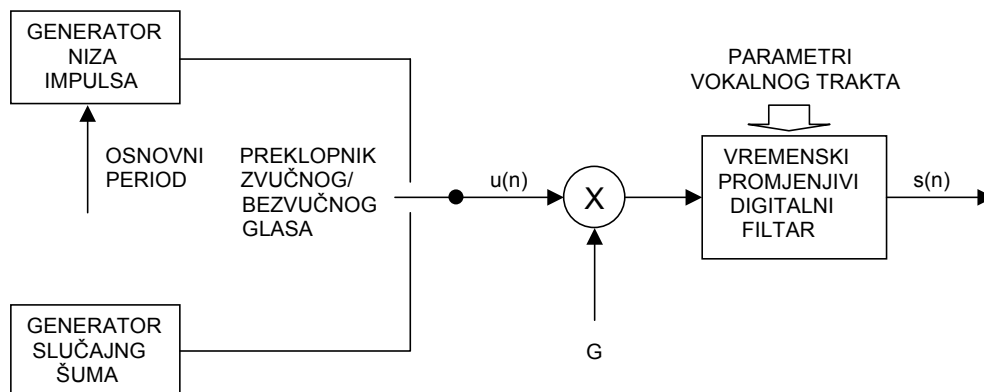
Slika 9.2-3 Pogreška predikcije za signal na slici 9.2-2

9.3 Teoretske postavke an alize linearnom predikcijom

U poglavlju 8.5 uveden je potpuni digitalni model formiranja govornog signala, s blok shemom prikazanom na slici 8.5-1. U svrhu diskusije postupka linearne predikcije pogodno je taj model pojednostaviti, kao što je prikazano na slici 9.3-1. Kod ovog pojednostavljenog modela su utjecaji oblika glotalnog pulsa, prijenosne funkcije vokalnog trakta i zračenja na usnicama združeni u jednom sustavu i predstavljeni s vremenski promjenjivim digitalnim filtrom čija prijenosna funkcija $H(z)$ ima oblik:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (9.3-1)$$

Takav sustav je pobuđen nizom jediničnih impulsa za zvučne glasove ili sekvencom slučajnog šuma za bezvučne glasove. Prema tome parametri ovog modela su: položaj sklopke za odabir tipa pobude (zvučni ili bezvučni glas), period osnovne frekvencije titranja glasnica za zvučne glasove, faktor pojačanja G i koeficijenti digitalnog filtra, a_k . Svi ovi parametri su, naravno, sporo promjenjivi s vremenom.



Slika 9.3-1 Blok shema jednostavnog modela za formiranje govora

Zvučnost i period osnovne frekvencije mogu se odrediti korištenjem više različitih metoda, ali i pomoću onih temeljenih na analizi linearnom predikcijom. Kao što je već diskutirano, ovakav model koji ima samo polove predstavlja vrlo dobru reprezentaciju za zvučne ne-nazalne glasove. U slučaju modeliranja nazalnih glasova ili zvučnih frikativa detaljna akustička teorija zahtijeva da prijenosna funkcija modela ima i nule, a ne samo polove, no biti će pokazano da će i ti glasovi biti reprezentirani dovoljno dobro ako je stupanj modela p dovoljno visok. Glavna prednost ovog pojednostavljenog modela s filtrom bez nula je u tome da se parametar pojačanja G i koeficijenti a_k digitalnog filtra mogu na vrlo jednostavan i računski učinkovit način odrediti upravo korištenjem metode analize linearnom predikcijom.

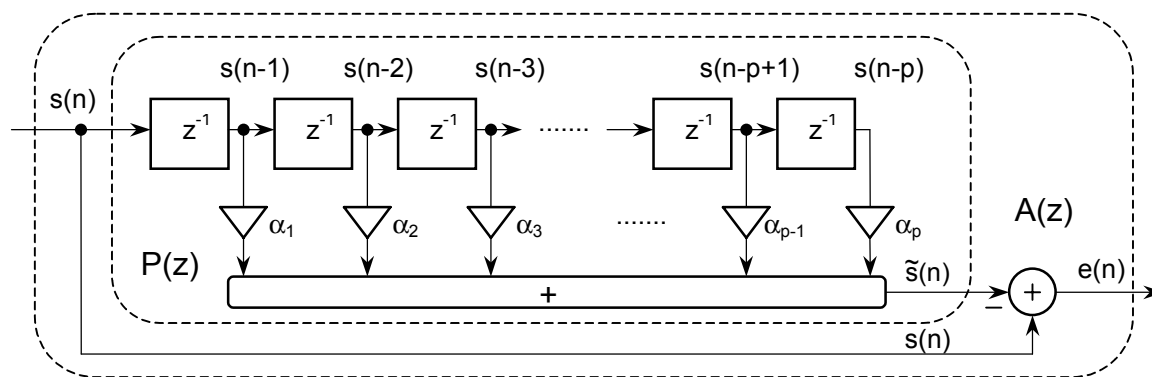
Za sustav prikazan slikom 9.3-1 i izrazom (9.3-1) možemo pisati

$$\begin{aligned} GU(z) &= S(z) - a_1 S(z)z^{-1} - \dots - a_p S(z)z^{-p} \\ Gu(n) &= s(n) - a_1 s(n-1) - \dots - a_p s(n-p) \end{aligned} \quad (9.3-2)$$

Iz izraza (9.3-2) vidimo da su uzorci govora $s(n)$ i pobude $u(n)$ povezani jednostavnom jednadžbom diferencija:

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (9.3-3)$$

U prošlom poglavlju je uveden pojam linearnog prediktora s koeficijentima predikcije α_k , čija je blok shema prikazana na slici 9.3-2. Jednadžba diferencija za ovaj prediktor dana u izrazu (9.2-1) vrlo je slična jednadžbi diferencija (9.3-3) pojednostavljenog modela sa slike 9.3-1.



Slika 9.3-2 Blok shema linearnog prediktora

Prijenosna funkcija linearnog prediktora p -tog reda je polinom $P(z)$:

$$P(z) = \sum_{k=1}^p \alpha_k z^{-k} \quad (9.3-4)$$

Pogreška predikcije $e(n)$ definirana je kao :

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k) \quad (9.3-5)$$

Iz relacije (9.3-5) može se uočiti da je niz uzoraka pogreške predikcije $e(n)$ u stvari izlaz sustava $A(z)$ prikazanog na slici 9.3-2 čija prijenosna funkcija glasi

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^p \alpha_k z^{-k} = 1 - P(z) \quad (9.3-6)$$

Ako bi iz izraza (9.3-5) izrazili $s(n)$ kao funkciju prethodnih uzoraka $s(n-k)$ i pogreške predikcije $e(n)$ slijedi:

$$s(n) = \sum_{k=1}^p \alpha_k s(n-k) + e(n) \quad (9.3-7)$$

Uspoređujući izraze (9.3-3) i (9.3-7) vidljivo je da su vrlo slični. Uz pretpostavku da je govorni signal stvarno nastao kao izlaz modela $H(z)$ iz izraza (9.3-1) i uz pretpostavku da su koeficijenti prediktora α_k jednaki koeficijentima modela a_k , tada će pogreška predikcije $e(n)$ biti identički jednaka pobudnom signalu modela $Gu(n)$. To drugim riječima znači da kada bi govorni signal $s(n)$ propustili kroz sustav $A(z)$, da bi se na njegovom izlazu dobio pobudni signal modela, pa se stoga $A(z)$ često naziva inverzni filter sustava $H(z)$ tj. :

$$H(z) = \frac{G}{A(z)} \quad (9.3-8)$$

Osnovni problem analize linearnom predikcijom je određivanje skupa koeficijenata prediktora $\{\alpha_k\}$ direktno iz uzoraka otipkanog govornog signala na takav način da se postigne dobro poklapanje kratkotrajnih spektralnih svojstava govornog signala i modela prema izrazu (9.3-8) u čijem nazivniku figuriraju koeficijenti $\{\alpha_k\}$. Zbog vremenski promjenjive prirode govornog signala, koeficijenti prediktora moraju biti određeni iz kratkog segmenta govornog signala unutar kojeg su spektralna svojstva signala stalna. Osnovna ideja linearne predikcije je pronalaženje takvih koeficijenata prediktora koji će minimizirati srednju kvadratnu

pogrešku predikcije na tom promatranom segmentu govornog signala. Rezultirajući optimalni parametri $\{\alpha_k\}$ će zatim biti prihvaćeni kao koeficijenti polinoma u nazivniku modela $H(z)$, tj. pretpostavit će se da su parametri modela $\{a_k\}$ upravo jednaki tim određenim koeficijentima prediktora $\{\alpha_k\}$.

Možda nije odmah uočljivo da će ovakav pristup dati zadovoljavajuće rezultate, ali on se može opravdati na nekoliko načina. Prvo, uz pretpostavku da stvarno vrijedi uvjet da je $\alpha_k = a_k$, tada je $e(n) = Gu(n)$. Za zvučni govor ovo bi značilo da se $e(n)$ sastoji od niza impulsa, tj. $e(n)$ bi uglavnom bio jednak nuli i imao bi malu energiju. Stoga, čini se da je pronalaženje parametara $\{\alpha_k\}$ koji minimiziraju energiju pogreške predikcije u skladu s gore izloženim opažanjima.

Drugo objašnjenje je čisto matematičke prirode. Uz pretpostavku da vremenski nepromjenjiv sustav $H(z)$ prema izrazu (9.3-1) s poznatim koeficijentima $\{a_k\}$ pobudimo s jediničnim impulsom ili sa stacionarnim bijelim šumom, te ako na uzorcima njegovog odziva provedemo postupak linearne predikcije, tada će tako dobiveni optimalni koeficijenti prediktora $\{\alpha_k\}$ biti jednaki polaznim koeficijentima sustava $\{a_k\}$. Identička jednakost vrijedi u slučaju pobude s jediničnim impulsom, dok u slučaju pobude sa stacionarnim bijelim šumom koeficijenti $\{\alpha_k\}$ teže prema $\{a_k\}$, a poklapanje je to bolje što je broj uzoraka signala veći.

Treće opravdanje za korištenje minimalne srednje kvadratne pogreške predikcije kao osnove za određivanje parametara modela je čisto praktične prirode, tj. uz tako definiranu ciljnu funkciju optimalni koeficijenti prediktora se nalazi kao rješenje običnog skupa linearnih jednadžbi. Još važnije je da rezultirajući parametri veoma dobro modeliraju kratkotrajna spektralna svojstva govornog signala kako će biti pokazano.

Postupak određivanja koeficijenata prediktora provodi se nad vremenskim segmentom govornog signala konačnog trajanja. Taj segment biti će označen sa $s_n(m)$, a radi se o dijelu signala $s(n)$ u okolini indeksa n , gdje je ta okolina određena rasponom indeksa m , tj. vrijedi:

$$s_n(m) = s(m+n) \quad (9.3-9)$$

Ta operacija izdvajanja segmenta signala ilustrirana je na slici 9.3-3, gdje se indeks m kreće od m_1 do m_2

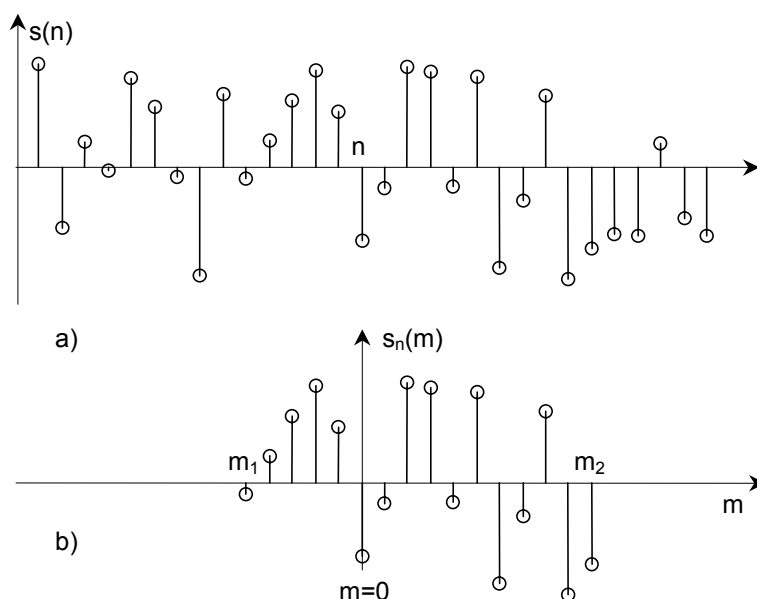
Srednja vrijednost pogreške predikcije na tako odabranom segmentu signala $s_n(m)$ definirana je kao :

$$E_n = \sum_m e_n^2(m) \quad (9.3-10)$$

$$= \sum_m (s_n(m) - \tilde{s}_n(m))^2 \quad (9.3-11)$$

$$= \sum_m \left[s_n(m) - \sum_{k=1}^p \alpha_k s_n(m-k) \right]^2 \quad (9.3-12)$$

Granice sumacija u izrazima (9.3-10) do (9.3-12) su privremeno ostavljene nedefinirane, tj. naznačeno je samo da se sumacija provodi preko indeksa m . Raspon ove sumacije ovisiti će o tipu predikcijskog postupka, ali u svakom slučaju radi se o konačnom rasponu. Stroga definicija srednje vrijednosti zahtijeva da se izrazi (9.3-10) do (9.3-12) podijele se brojem uzoraka u tom konačnom intervalu, no taj faktor skale neće utjecati na rješenje, tj. optimalni skup koeficijenata α_k koji minimiziraju E_n ne ovisi o tom faktoru skale.



Slika 9.3-3 Govorni signal $s(n)$ a), Izdvojeni segment $s_n(m)$ u okolini indeksa n b)

Vrijednosti parametara α_k koji minimiziraju E_n prema relaciji (9.3-12) mogu se odrediti tako da se parcijalne diferencijalne jednadžbe E_n po nepoznatim koeficijentima izjednače s nulom, tj. $\partial E_n / \partial \alpha_i = 0$, za $i=1, 2, \dots, p$. Lako je pokazati da iz ovog uvjeta slijedi sustav od p linearnih jednadžbi po nepoznanicama α_1 do α_p oblika:

$$\sum_m s_n(m-i)s_n(m) = \sum_{k=1}^p \alpha_k \sum_m s_n(m-i)s_n(m-k), \quad 1 \leq i \leq p \quad (9.3-13)$$

Primjer izvoda ovog sustava biti će prikazan za slučaj linearnog prediktora drugog reda ($p=2$). Srednja kvadratna pogreška predikcije E_n definirana je kao :

$$\begin{aligned} E_n &= \sum_m [s_n(m) - \alpha_1 s_n(m-1) - \alpha_2 s_n(m-2)]^2 \\ &= \sum_m (s_n^2(m) + \alpha_1^2 s_n^2(m-1) + \alpha_2^2 s_n^2(m-2) - 2\alpha_1 s_n(m)s_n(m-1) \\ &\quad - 2\alpha_2 s_n(m)s_n(m-2) + 2\alpha_1 \alpha_2 s_n(m-1)s_n(m-2)) \end{aligned} \quad (9.3-14)$$

Izjednačavanjem parcijalnih diferencijalnih jednadžbi s nulom slijedi:

$$\frac{\partial E_n}{\partial \alpha_1} = \sum_m [2\alpha_1 s_n^2(m-1) - 2s_n(m)s_n(m-1) + 2\alpha_2 s_n(m-1)s_n(m-2)] = 0 \quad (9.3-15)$$

$$\frac{\partial E_n}{\partial \alpha_2} = \sum_m [2\alpha_2 s_n^2(m-2) - 2s_n(m)s_n(m-2) + 2\alpha_1 s_n(m-1)s_n(m-2)] = 0 \quad (9.3-16)$$

odnosno nakon dijeljenja s 2 i separacije nepoznanica i konstanti :

$$\sum_m s_n(m)s_n(m-1) = \alpha_1 \sum_m s_n(m-1)s_n(m-1) + \alpha_2 \sum_m s_n(m-1)s_n(m-2) \quad (9.3-17)$$

$$\sum_m s_n(m)s_n(m-2) = \alpha_1 \sum_m s_n(m-1)s_n(m-2) + \alpha_2 \sum_m s_n(m-2)s_n(m-2) \quad (9.3-18)$$

Usporedbom ovog sustava s općom formom danom u izrazu (9.3-13) vidljivo je poklapanje. Sustav jednadžbi moguće je i kompaktnije zapisati ako se definira:

$$\phi_n(i,k) = \sum_m s_n(m-i)s_n(m-k) \quad (9.3-19)$$

tada izraz (9.3-13) možemo kraće pisati kao :

$$\sum_{k=1}^p \alpha_k \phi_n(i,k) = \phi_n(i,0) \quad , \quad i = 1,2,\dots,p \quad (9.3-20)$$

ili u matričnom obliku:

$$\begin{bmatrix} \phi_n(1,1) & \phi_n(1,2) & \dots & \phi_n(1,p) \\ \phi_n(2,1) & \phi_n(2,2) & \dots & \phi_n(2,p) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \phi_n(p,1) & \phi_n(p,2) & \dots & \phi_n(p,p) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \phi_n(1,0) \\ \phi_n(2,0) \\ \dots \\ \phi_n(p,0) \end{bmatrix} \quad (9.3-21)$$

Rješavanjem ovog sustava jednadžbi slijedi skup optimalnih koeficijenata prediktora $\{\alpha_k\}$ koji minimizira srednju kvadratnu pogrešku predikcije E_n za segment $s_n(m)$. Naravno da i ovi koeficijenti ovise o indeksu n za koji je proveden postupak, ali to neće biti posebno naglašavano dodavanjem indeksa na koeficijente α_k .

Korištenjem izraza (9.3-12) i (9.3-13) može se pokazati da se minimalna srednja kvadratna vrijednost pogreške predikcije uz koeficijente α_k koji zadovoljavaju sustav (9.3-21), može odrediti kao:

$$E_n = \sum_m s_n^2(m) - \sum_{k=1}^p \alpha_k \sum_m s_n(m)s_n(m-k) \quad (9.3-22)$$

odnosno uz korištenje izraza (9.3-19) za $\phi_n(i,k)$ u kompaktnoj formi oblika:

$$E_n = \phi_n(0,0) - \sum_{k=1}^p \alpha_k \phi_n(0,k) \quad (9.3-23)$$

Dakle, ukupna minimalna pogreška sastavljena je od fiksne komponente $\phi_n(0,0)$ i od komponenata koje ovise o koeficijentima predikcije. Ovaj fiksni dio $\phi_n(0,0)$ je u stvari jednak energiji segmenta ulaznog signala $s_n(m)$.

Da bi se odredili optimalni koeficijenti predikcije, treba prvo odrediti pomoćne varijable $\phi_n(i,k)$ i to za one kombinacije i i k koje se javljaju su sustavu jednadžbi (9.3-21), tj. za $1 \leq i \leq p$ i $0 \leq k \leq p$. Nakon toga treba još jedino riješiti sustav (9.3-21) čime se nalaze optimalni koeficijenti prediktora $\{\alpha_k\}$. Iako iz izloženog izgleda da je analiza linearnom predikcijom u principu vrlo jednostavna, ipak su detalji izračunavanja $\phi_n(i,k)$ i proizlazećih rješenja jednadžbi prilično komplicirani i zahtijevaju daljnju diskusiju.

Do sada još uvijek nije fiksiran raspon sumacije po indeksu m , no valja napomenuti da je on jednak u svim izrazima u kojima figurira suma po m . Naravno, zbog vremenske promjenjivosti spektralnih svojstava govornog signala, taj raspon mora biti konačan. Postoje dva osnovna pristupa odabiru ovog raspona i kao što će biti prikazano u narednim poglavljima, dva postupka linearne predikcije će se razlikovati samo po tome kako je odabran

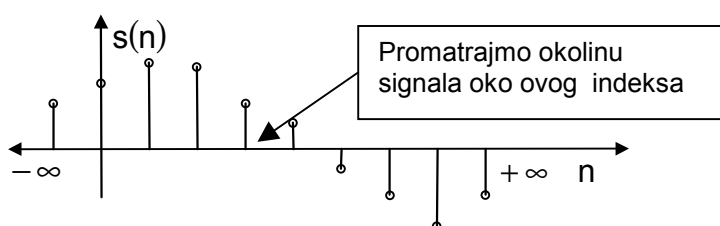
taj raspon, odnosno na koji način se provodi izdvajanje segmenta $s_n(m)$ iz govornog signala $s(n)$.

9.4 Autokorelacijska metoda

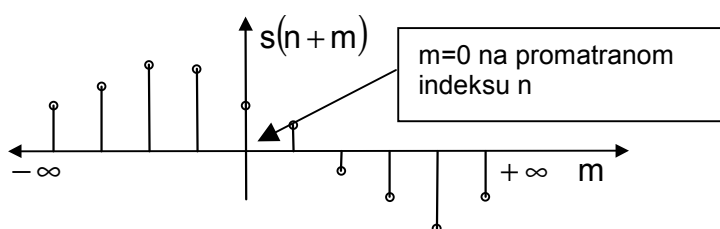
Jedan od pristupa za određivanje granica sumacije pogreške predikcije (engl. *short time average prediction error*) polazi od pretpostavke da je segment signala $s_n(m)$ jednak nuli izvan intervala $0 \leq m \leq N-1$. To se na prikladan način može zapisati kao

$$s_n(m) = s(m+n)w(m), \quad (9.4-1)$$

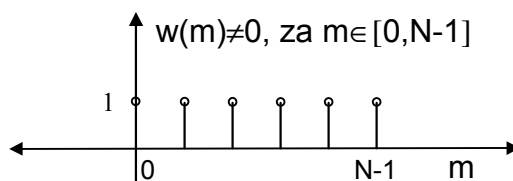
gdje je $w(m)$ vremenski otvor konačne dužine (npr. Hammingov vremenski otvor) jednak nuli izvan intervala $0 \leq m \leq N-1$.



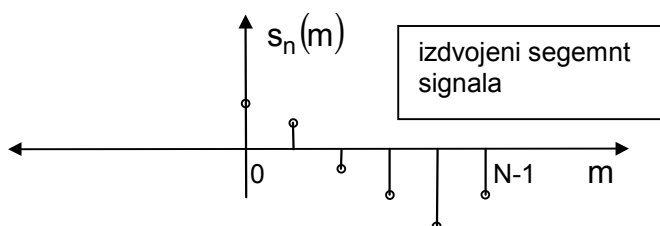
Slika 9.4-1 Signal beskonačnog trajanja



Slika 9.4-2 Signal beskonačnog trajanja s pomakom



Slika 9.4-3 Pravokutni vremenski otvor dužine N uzoraka



Slika 9.4-4 Segment signala izdvojen pomoću vremenskog otvora

Slikama 9.4-1 do 9.4-4 ilustrirano je izdvajanje segmenta signala, pri čemu je radi jednostavnosti korišten pravokutni vremenski otvor. Budući da $s_n(m)$ različit od nule samo na intervalu $0 \leq m \leq N-1$, odgovarajuća pogreška predikcije $e_n(m)$ (engl. *prediction error*) za prediktor p -tog reda biti će različita od nule samo na intervalu $0 \leq m \leq N-1+p$ (... vidi izraz (9.3-5)). Srednja kvadratna pogreška predikcije E_n se dakle u ovom slučaju može odrediti izrazom (9.4-2).

$$E_n = \sum_{m=0}^{N+p-1} e_n^2(m) \quad (9.4-2)$$

Alternativno se može reći da treba sumirati sve vrijednosti različite od nule na intervalu od $-\infty$ do $+\infty$.

Pogreška predikcije je u pravilu velikog iznosa na početku intervala (preciznije, za $0 \leq m \leq p-1$), zato što se pokušava provesti predikcija signala na temelju uzoraka koji su na silu postavljeni na nulu. Pogreška predikcije također je velika na kraju intervala (preciznije, za $N \leq m \leq N+p-1$), jer se pokušava provesti predikcija uzoraka jednakih nuli na temelju uzoraka koji su različiti od nule. Jedan od načina kojim je moguće riješiti ovaj problem je primjena vremenskog otvora $w(m)$ koji prigušuje signal prema rubovima segmenta $s_n(m)$, kao što je npr. Hamming-ov vremenski otvor.

Granice sumacije u izrazu (9.4-2) za srednju kvadratnu predikcijsku pogrešku su identične su onima u izrazu (9.4-3) za $\phi_n(i,k)$ koji definira elemente matrice i slobodnog stupca linearnog sustava jednadžbi iz prošlog poglavlja.

$$\phi_n(i,k) = \sum_{m=0}^{N+p-1} s_n(m-i)s_n(m-k), \quad \text{za } 1 \leq i \leq p, \quad 0 \leq k \leq p. \quad (9.4-3)$$

Ovaj izraz je moguće modificirati tako da se provede supstitucija varijable m sa m^* prema izrazu: $m^* = m - i$, čime se dobiva:

$$\phi_n(i,k) = \sum_{m^*=-i}^{N-1+p-i} s_n(m^*)s_n(m^*+i-k) \quad (9.4-4)$$

Budući da je $s_n(m^*)$ jednak nuli izvan intervala $0 \leq m^* \leq N-1$, jednostavno se može pokazati da se granice sumacije u izrazu (9.4-4) sužuju s obzirom da vrijedi:

- za donju granicu $s_n(m^*) = 0, m^* \leq 0$;
- analogno za gornju $s_n(m^*+i-k) = 0, m^*+i-k \geq N$.

Na osnovu toga novi raspon indeksa je: $0 \leq m^* \leq N-1-(i-k)$, odnosno:

$$\phi_n(i,k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k), \quad 1 \leq i \leq p, \quad 0 \leq k \leq p \quad (9.4-5)$$

Nadalje, može se uočiti da je oblik izraza (9.4-5) za $\phi_n(i,k)$ identičan vremenski kratkotrajnoj autokorelacijskoj funkciji (engl. *short-time autocorrelation function*) izračunatoj za pomak $(i-k)$, tj. vrijedi:

$$\phi_n(i,k) = R_n(i-k), \quad (9.4-6)$$

pri čemu je:

$$R_n(j) = \sum_{m=0}^{N-1-j} s_n(m)s_n(m+j). \quad (9.4-7)$$

S obzirom da je $R_n(j)$ parna funkcija po pomaku j , vrijedi izraz (9.4-8),

$$\phi_n(i,k) = R_n(|i-k|), \quad i = 1,2,\dots,p; \quad k = 0,1,2,\dots,p \quad (9.4-8)$$

Zamjenom izraza $\phi_n(i,k)$ s autokorelacijskom funkcijom $R_n(|i-k|)$ prema izrazu (9.4-8), linearni sustav jednadžbi koje definiraju optimalni prediktor poprima oblik:

$$\sum_{k=1}^p \alpha_k R_n(|i-k|) = R_n(i), \quad 1 \leq i \leq p, \quad (9.4-9)$$

gdje su koeficijenti α_k traženi koeficijenti optimalnog prediktora kojeg želimo odrediti. Analogno tome, minimalna vrijednost srednje kvadratne pogreške predikcije prediktora reda p koja se postiže primjenom optimalnih koeficijenta α_k koji zadovoljavaju sustav (9.4-9) može se odrediti na slijedeći način:

$$E_n = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k). \quad (9.4-10)$$

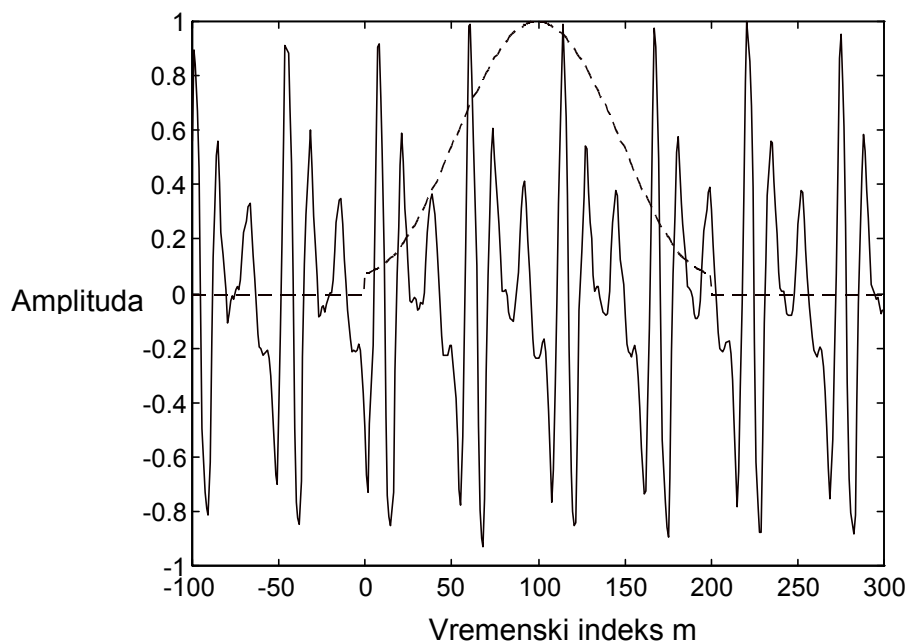
Jednadžbe dane izrazom (9.4-9) mogu se zapisati i u matričnoj formi:

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \dots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \dots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \dots & R_n(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \dots & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \dots \\ \dots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \dots \\ \dots \\ R_n(p) \end{bmatrix}. \quad (9.4-11)$$

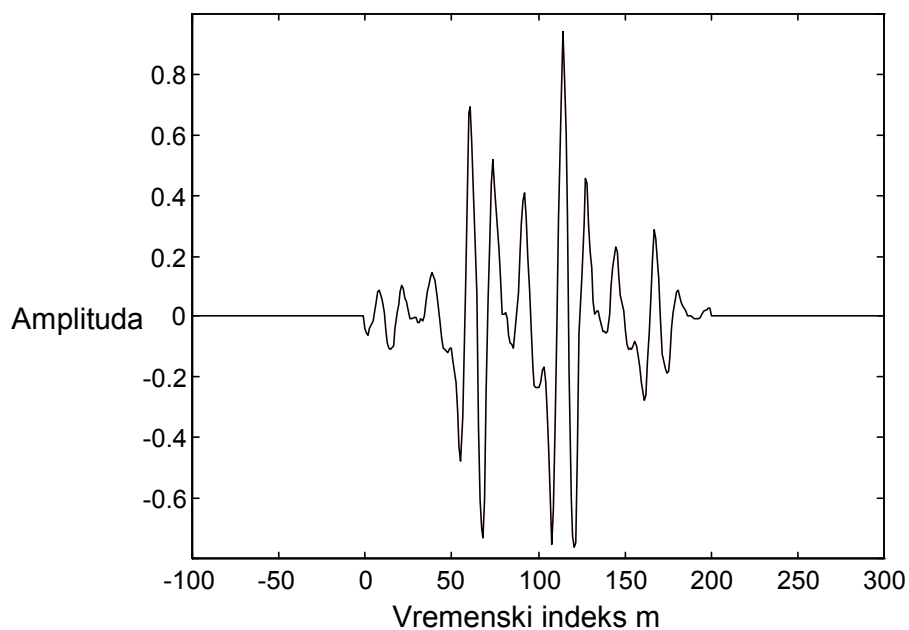
Matrica autokorelacijskih vrijednosti, dana izrazom (9.4-11), je Toeplitz-ove forme (simetrična po obje glavne dijagonale) dimenzija $p \times p$. U ovom slučaju, su čak i vrijednosti elemenata duž svake pojedine desno položene dijagonale jednake. Zbog ovih svojstava, gornji sustav se može riješiti po koeficijentima α_k mnogo učinkovitije nego kod klasičnih postupaka (npr. množenje inverznom matricom), što će biti pokazano u slijedećim poglavljima.

U nastavku ovog poglavlja bit će prikazan primjer određivanja koeficijenata prediktora klasičnom metodom, te pripadne prijenosne funkcije $H(z)$ za prediktor reda $p=6$.

Na slici 9.4-5 prikazan je odsječak govornog signala i Hammingov vremenski otvor koji će poslužiti za izdvajanje segmenta signala. Izdvojeni signal prikazan je slikom 9.4-6.

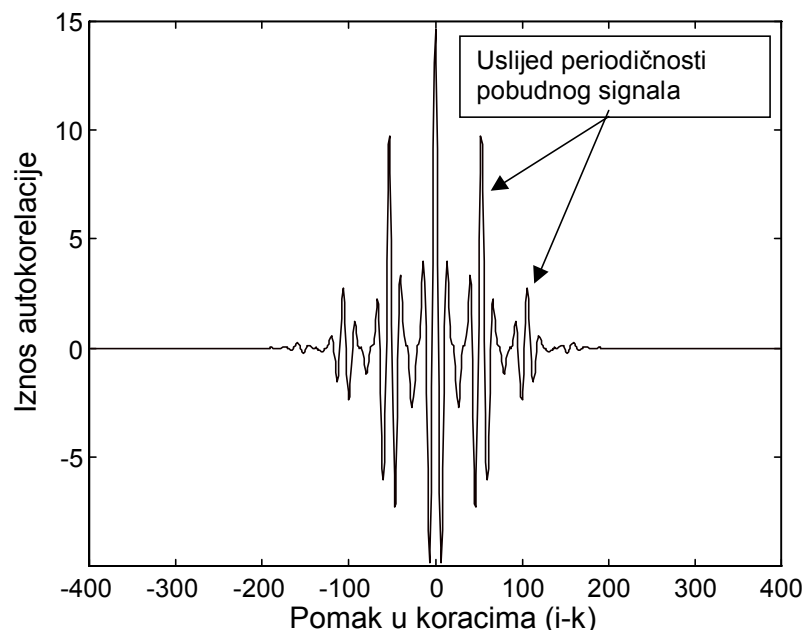


Slika 9.4-5 Signal i Hammingov vremenski otvor



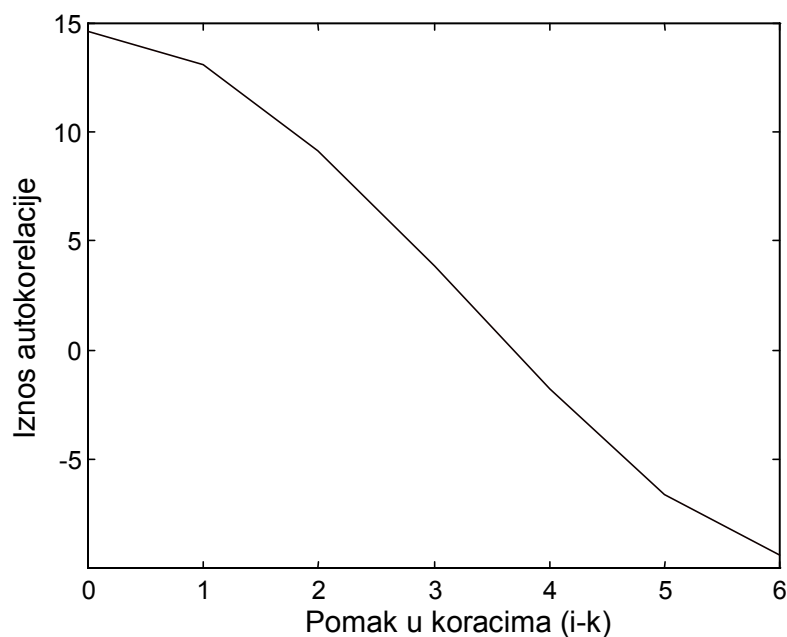
Slika 9.4-6 Segment signala izdvojen vremenskim otvorom ($N=200$)

Autokorelacija signala pomnoženog vremenskim otvorom prikazana je na slici 9.4-7. Treba uočiti simetričnost oko nule (parnost). Izraženi šiljci predstavljaju mjesta dobrog poklapanja signala i njegove pomaknute verzije. Naime, šiljci se pojavljuju na iznosima pomaka (engl. *autocorrelation lag*) koji odgovaraju osnovnom periodu pobudnog signala vokalnog trakta i njegovim višekratnicima.



Slika 9.4-7 Autokorelacija signala pomnoženog vremenskim otvorom

Autokorelacija signala pomnoženog s vremenskim otvorom (za pomake od 0 do p) prikazana je na slici 9.4-8, a neposredno iza slike slijedi tablica 9.4-1 sa numeričkim vrijednostima autokorelacijskih koeficijenata.



Slika 9.4-8 Autokorelacija signala pomnoženog vremenskim otvorom (pomaci od 0 do p)

Tablica 9.4-1 Autokorelacija za pomake od 0 do p

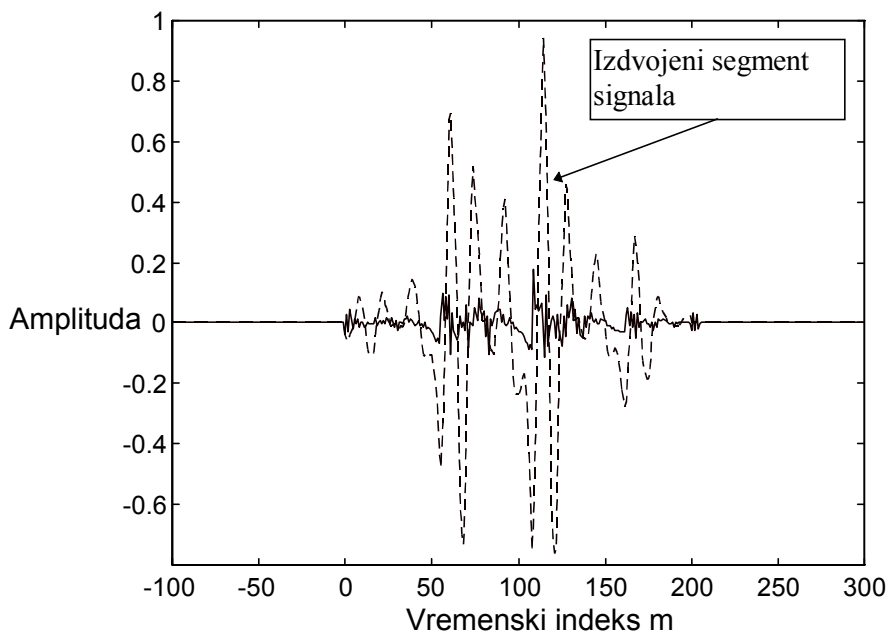
$(i-k)$	0	1	2	3	4	5	6
$R(i-k)$	14.61	13.10	9.16	3.88	-1.80	-6.61	-9.39

Tih sedam brojeva popisanih u tablici jednoznačno određuju prediktor reda p kojim je modeliran vokalni trakt. Moguće je sada formirati sustav jednačbi prema relaciji (9.4-10) i riješiti ga po nepoznicama α_k . U tablici 9.4-2 popisane su vrijednosti koeficijenata α_k dobivenih rješavanjem linearnog sustava jednačbi.

Tablica 9.4-2 Koeficijenti linearnog prediktora

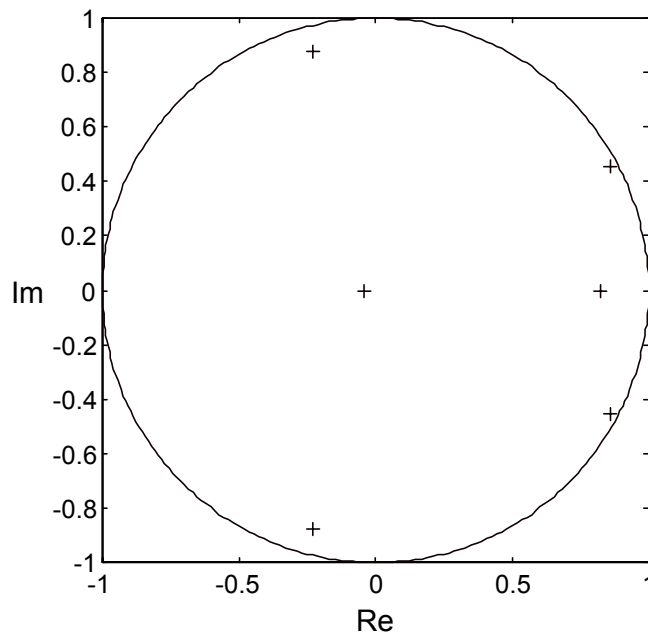
α_1	α_2	α_3	α_4	α_5	α_6
2.06	-1.96	1.72	-1.52	0.58	0.03

Na slici 9.4-9 prikazan je signal pomnožen vremenskim otvorom (crtkanom linijom) i pogreška predikcije (punom linijom).



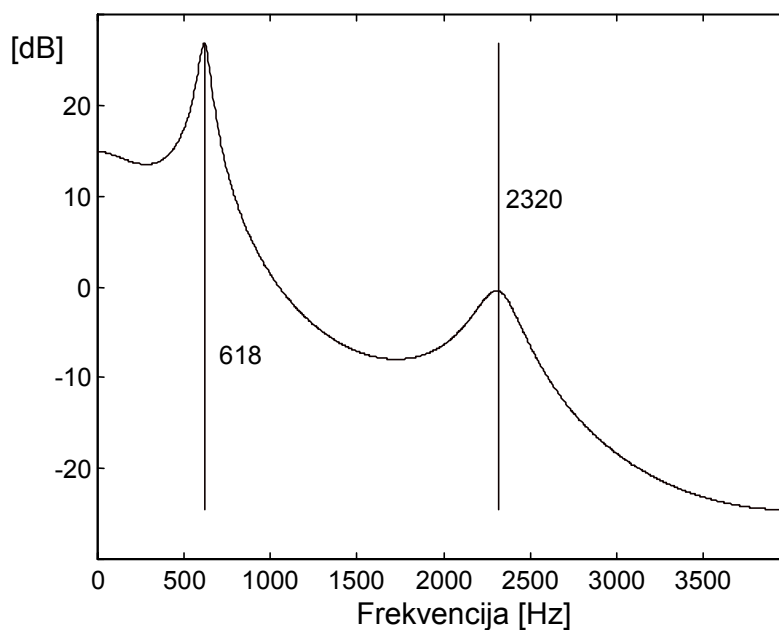
Slika 9.4-9 Izdvojeni segment signala (crtkano) i pogreška predikcije (punom linijom)

Sumarna kvadratna pogreška predikcije iznosi $E_n=0.27$. Odgovarajuća prijenosna funkcija $H(z)$ modela govornog sustava za izračunate koeficijente α_k ima šest polova: dva realna i dva para konjugirano kompleksnih polova. Njihov položaj u z -ravnini prikazan je slikom 9.4-10. Sustav je stabilan jer se svi polovi nalaze unutar jedinične kružnice. Autokorelacijska metoda uvijek rezultira stabilnim prediktorom. Svaki konjugirano kompleksni par polova određuje jednu rezonantnu karakteristiku, koja se poklapa s jednim od formanta govornog signala. Polovi smješteni blizu jedinične kružnice imaju veliki Q -faktor i daju izražene maksimume u prijenosnoj funkciji, tj. odgovaraju uskim formantima.



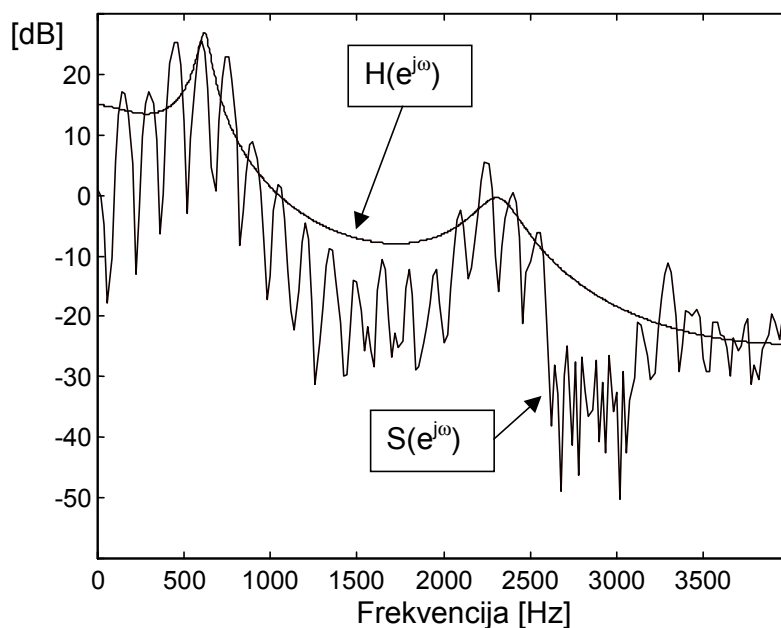
Slika 9.4-10 Pozicije polova u z-ravnini

Na slici 9.4-11 prikazana je amplitudno frekventijska karakteristika $H(z)$ s naznačenim frekvencijama polova.



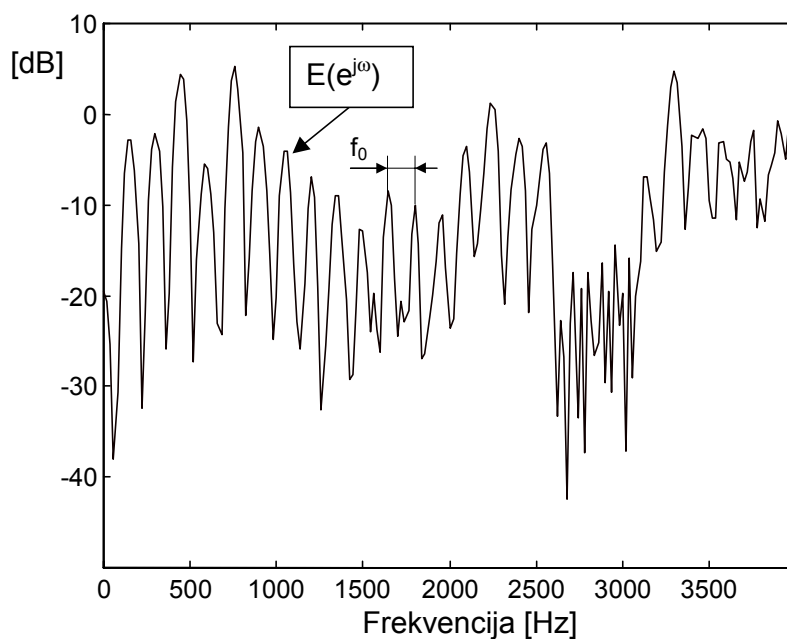
Slika 9.4-11 Frekventijska karakteristika od $H(z)$ i frekvencije polova

Na slici 9.4-12 prikazani su frekventijska karakteristika prediktora $H(e^{j\omega})$ i spektar ulaznog signala $S(e^{j\omega})$. Bitno je uočiti kako je frekventijskom karakteristikom prediktora modeliran spektar ulaznog signala, tj. prediktor opisuje njegov sporo-promjenjivi dio uzduž frekventijske osi. Radi toga se za prediktor kaže da opisuje spektralnu ovojnici govornog signala (engl. *spectral envelope*).



Slika 9.4-12 Frekvencijska karakteristika od $H(z)$ i spektar ulaznog signala

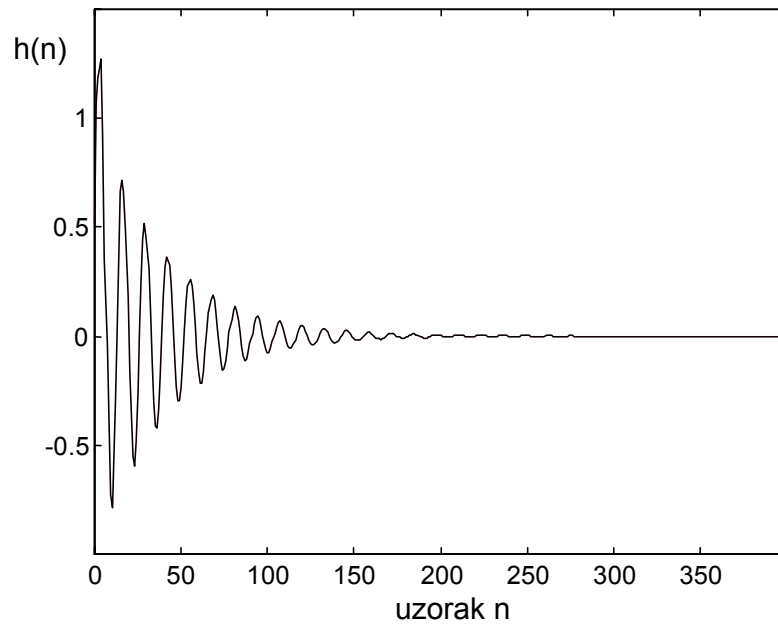
Na slici 9.4-13 prikazan je spektar rezidualnog signala $E(e^{j\omega})$. To je spektar pogreške predikcije koji se izračunava kao razlika spektra signala i amplitudno frekvencijske karakteristike prediktora u logaritamskoj mjeri, tj. u dB. Drugim riječima, taj spektar zapravo odgovara kvocijentu modula spektra signala i modula frekvencijske karakteristike $H(z)$. Razmak između latica na slici 9.4-13 odgovara fundamentalnoj (osnovnoj) frekvenciji titranja glasnica f_0 .



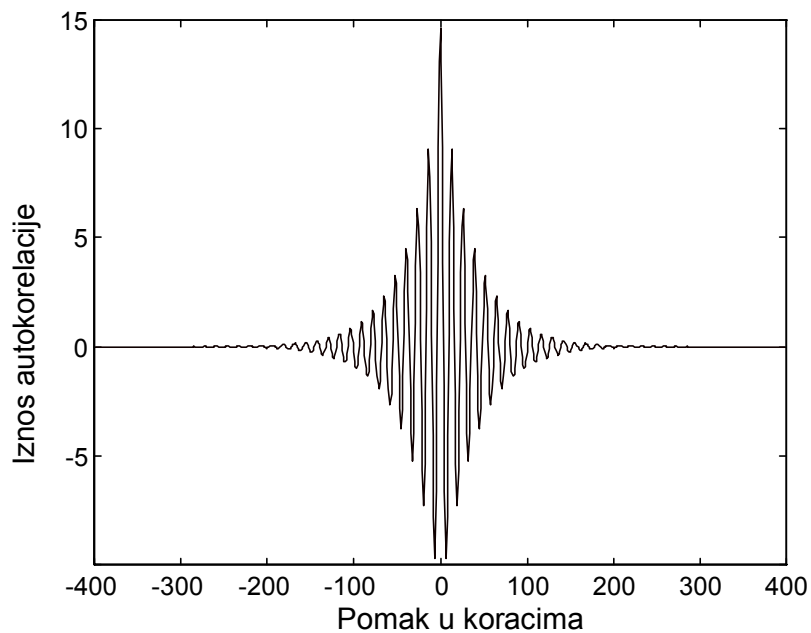
Slika 9.4-13 Spektar rezidualnog signala

Usporedbom slika 9.4-12 i 9.4-13 bitno je uočiti da je spektar pogreške predikcije mnogo 'ravniji' nego spektar ulaznog signala, tj. vrhovi svih latica su približno na istoj visini (unutar 15 dB). Zbog te činjenice, može se reći da filter $A(z)$ na čijem se izlazu nalazi signal $e(n)$ provodi spektralno izbjeljivanje signala. To izbjeljivanje je to bolje što je red prediktora viši.

Na slici 9.4-14 prikazan je impulsni odziv $h(n)$ sustava $H(z)$, a na slici 9.4-15 je prikazana njegova autokorelacija.



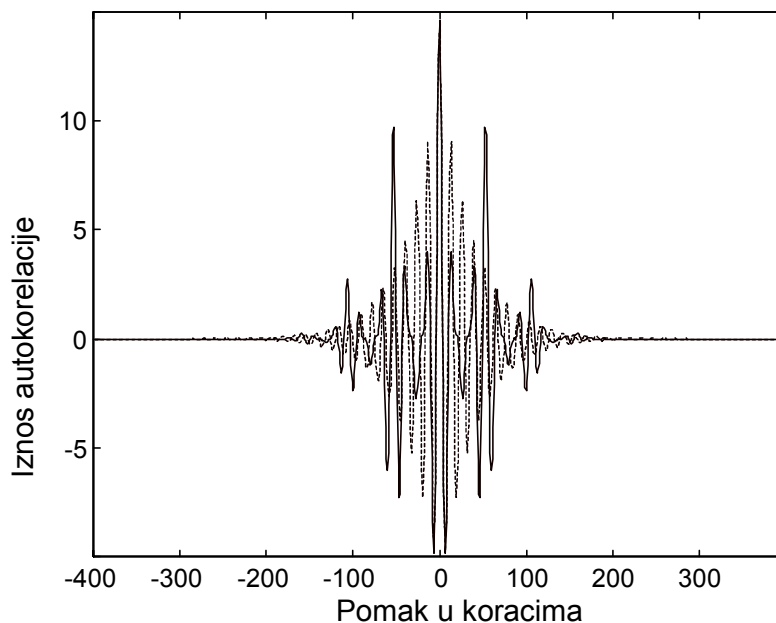
Slika 9.4-14 Impulsni odziv $h(n)$ sustava $H(z)$



Slika 9.4-15 Autokorelacija impulsnog odziva sustava $h(n)$

Na slici 9.4-16 prikazana je autokorelacija impulsnog odziva sustava $h(n)$ crtkanom linijom i autokorelacija ulaznog signala pomnoženog vremenskim otvorom punom linijom. Te dvije autokorelacije su identički jednake za pomake od $-p$ do $+p$, dok se izvan tog područja razlikuju. Ova činjenica je vrlo važna za samu ideju modeliranja primjenom autokorelacijskog postupka, tj. metodom autokorelacije odredili smo linearni sustav $H(z)$ čija

će autokorelacija impulsnog odziva biti identički jednaka autokorelaciji izdvojenog segmenta signala za pomake od $-p$ do $+p$.



Slika 9.4-16 Autokorelacija signala pomnoženog vremenskim otvorom (punom linijom) i autokorelacija impulsnog odziva sustava $h(n)$ (crtkano)

9.5 Metoda kovarijance

Drugi osnovni pristup definiranju govornog segmenta $s_n(m)$ i granica sumiranja je da se ograniči indeks m preko kojeg se usrednjava pogreška predikcije, te se uz tako odabrane granice sumacije odrede i granice za $\phi_n(i,k)$. Prema tome, ako je definirano

$$E_n = \sum_{m=0}^{N-1} e_n^2(m) \quad (9.5-1)$$

tada je raspon indeksa po m za E_n isti kao i za $\phi_n(i,k)$, odnosno

$$\phi_n(i,k) = \sum_{m=0}^{N-1} s_n(m-i)s_n(m-k) \quad \begin{array}{l} 1 \leq i \leq p \\ 0 \leq k \leq p \end{array} \quad (9.5-2)$$

U ovom slučaju, ako promijenimo indekse sumiranja uvođenjem pomoćne varijable m' , (dakle, $m'=m-i$, ili $m'=m-k$), tada $\phi_n(i,k)$ možemo izraziti na dva načina i to :

$$\phi_n(i,k) = \sum_{m'=-i}^{N-i-1} s_n(m')s_n(m'+i-k) \quad \begin{array}{l} 1 \leq i \leq p \\ 0 \leq k \leq p \end{array} \quad (9.5-3)$$

odnosno,

$$\phi_n(i,k) = \sum_{m'=-k}^{N-k-1} s_n(m')s_n(m'+k-i) \quad \begin{array}{l} 1 \leq i \leq p \\ 0 \leq k \leq p \end{array} \quad (9.5-4)$$

Iako izrazi (9.5-3) i (9.5-4) izgledaju vrlo slično kao izraz (9.4-5), vidimo da granice sumiranja nisu iste. Izrazi (9.5-3) i (9.5-4) zahtijevaju vrijednosti $s_n(m)$ izvan intervala $0 \leq m \leq N-1$. Konkretno, za izračunavanje $\phi_n(i,k)$ za sve potrebne kombinacije od i i k , trebaju se koristiti vrijednosti $s_n(m)$ unutar intervala $-p \leq m \leq N-1$. Za razliku od autokorelacijskog postupka, kod kojeg se u izračunavanju $\phi_n(i,k)$ pretpostavljalo da je signal $s_n(m)$ jednak nuli za $m < 0$, kod metode kovarijance se za izračunavanje $\phi_n(i,k)$ koriste i zadnjih p uzoraka iz prethodnog segmenta (okvira analize) ($-p \leq m \leq -1$). Zbog toga se kod metode kovarijance ne treba koristiti vremenski otvor koji prigušuje signal prema rubovima segmenta, već se može koristiti pravokutni otvor kod kojeg svi uzorci signala s jednakom težinom sudjeluju u izračunavanju prediktora. Ovaj pristup vodi ka funkciji koja baš i nije autokorelacijska funkcija, već odgovara kroskorelacijskoj funkciji dvaju vrlo sličnih, ali ne i identičnih segmenta govornog signala konačne duljine (razlika je upravo u rubnim uzorcima). Iako razlike između izraza (9.5-3) i (9.5-4) i izraza (9.4-5) izgledaju kao sitni računski detalji, jednadžbe linearnog sustava

$$\sum_{k=1}^p \alpha_k \phi_n(i,k) = \phi_n(i,0) \quad i = 1, 2, \dots, p \quad (9.5-5)$$

imaju bitno različita svojstva koja značajno utječu na metodu izračunavanja prediktora kao i na svojstva rezultirajućeg optimalnog prediktora. U matičnom obliku, ove jednadžbe izgledaju ovako

$$\begin{bmatrix} \phi_n(1,1) & \phi_n(1,2) & \phi_n(1,3) & \dots & \phi_n(1,p) \\ \phi_n(2,1) & \phi_n(2,2) & \phi_n(2,3) & \dots & \phi_n(2,p) \\ \phi_n(3,1) & \phi_n(3,2) & \phi_n(3,3) & \dots & \phi_n(3,p) \\ \dots & \dots & \dots & \dots & \dots \\ \phi_n(p,1) & \phi_n(p,2) & \phi_n(p,3) & \dots & \phi_n(p,p) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \dots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \phi_n(1,0) \\ \phi_n(2,0) \\ \phi_n(3,0) \\ \dots \\ \phi_n(p,0) \end{bmatrix} \quad (9.5-6)$$

U ovom slučaju, budući da je $\phi_n(i,k) = \phi_n(k,i)$, (vidi izraze (9.5-3) i (9.5-4)), matrica $\{\phi_n(i,k)\}$ dimenzije $p \times p$, koja sadrži kros-korelacije za sve pomake i i k je simetrična, ali nije Toeplitz strukture. Zapravo, može se pokazati da su svi elementi u desno ležećim dijagonalama povezani izrazom :

$$\phi_n(i+1, k+1) = \phi_n(i, k) + s_n(-i-1)s_n(-k-1) - s_n(N-1-i)s_n(N-1-k) \quad (9.5-7)$$

Korištenjem gornjeg izraza moguće je ubrzati postupak izračunavanja matrice $\{\phi_n(i,k)\}$, tj. dovoljno je izračunati samo prvi stupac matrice $\{\phi_n(i,1), i=1,2,\dots,p\}$ korištenjem originalnog izraza (9.5-3) ili (9.5-4), te zatim sve elemente na i ispod glavne dijagonale izračunati primjenom izraza (9.5-7). Elementi iznad glavne dijagonale se nalaze zrcaljenjem oko glavne dijagonale. Prednost takvog pristupa je u tome što direktna primjena izraza (9.5-3) ili (9.5-4) zahtijeva cca. N množenja i N zbrajanja za svaki element matrice $\phi_n(i,k)$, dok se izraz (9.5-7) izvodi pomoću samo dva množenja i tri zbrajanja za sve elemente osim onih u prvom stupcu.

Postupak određivanja optimalnog prediktora opisan u ovom poglavlju koji minimizira predikcijsku pogrešku prema izrazu (9.5-1), a temeljen na izračunavanju $\phi_n(i,k)$, poznat je

pod nazivom "metoda kovarijance" zato jer matrica vrijednosti $\{\phi_n(i,k)\}$ ima svojstva kovarijancijske matrice.

Pokazano je da se različitim definiranjem segmenta govornog signala koji se analizira mogu dobiti dva različita sustava linearnih jednadžbi koji određuju optimalni linearni prediktor. Za autokorelacijsku metodu, signal je izdvojen vremenskim otvorom širine N uzoraka i koeficijenti sustava jednadžbi $\phi_n(i,k)$ se nalaze vremenski kratkotrajnom autokorelacijskom funkcijom. Rezultirajuća matrica je Toeplitz strukture i vodi ka jednom načinu rješavanja koeficijenata predikcije. Za metodu kovarijance, pretpostavlja se da je signal poznat za skup vremenskih indeksa unutar intervala $-p \leq m \leq N-1$. Dovoljno je poznavati signal isključivo unutar tog intervala, jer samo ti uzorci figuriraju u izrazima za $\phi_n(i,k)$. Rezultirajuća korelacijska matrica (matrica linearnog sustava) je u ovom slučaju simetrična, ali nije Toeplitz strukture. Slično kao i kod autokorelacijske metode i ovdje postoji učinkovitiji postupak rješavanja sustava jednadžbi koji iskorištava činjenicu da je matrica sustava simetrična.

Važno je napomenut da se optimalni prediktori za te dvije metode razlikuju, a imaju i različita svojstva, što će biti diskutirano u narednim poglavljima.

9.6 Faktor pojačanja LPC filtra

Razumno je očekivati da se iznos pojačanja G može odrediti usporedbom energije signala s energijom linearno predviđenih uzoraka. Ta tvrdnja je doista istinita ako se za signal pobude LPC sustava uvedu odgovarajuće pretpostavke.

Koristeći jednadžbe (9.3-3) i (9.3-5) moguće je povezati faktor pojačanja G sa signalom pobude i pogreškom predikcije. Signal pobude može se izraziti kao:

$$Gu(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (9.6-1)$$

dok se signal predikcijske pogreške, $e(n)$ može izraziti kao:

$$e(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k) \quad (9.6-2)$$

U slučaju jednakosti koeficijenata prediktora i koeficijenata modela vokalnog trakta, to jest za $a_k = \alpha_k$, vrijedi:

$$e(n) = Gu(n) \quad (9.6-3)$$

odnosno, ulazni signal je proporcionalan signalu pogreške predikcije, a konstanta proporcionalnosti je faktor pojačanja G .

Jednadžba (9.6-3) je samo približna (to jest, vrijedi samo u slučaju jednakosti idealnih i stvarnih parametara linearnog prediktora), pa općenito nije moguće pouzdano izraziti G direktno preko signala pogreške. Zbog toga se uvodi jednostavna pretpostavka koja samo zahtjeva da energija signala pogreške bude jednaka energiji signala pobude, to jest:

$$G^2 \sum_{m=0}^{N-1} u^2(m) = \sum_{m=0}^{N-1} e^2(m) = E_n \quad (9.6-4)$$

Da bi bilo moguće povezati faktor pojačanja G s poznatim veličinama: koeficijentima α_k i korelacijskim koeficijentima, potrebno je uvesti neke pretpostavke o signalu pobude $u(n)$. Interesantna su dva slučaja pobude. Za zvučni govor u slučaju kada se analiza provodi sinkrono s glotalnim pulsevima i to tako da je točno jedan osnovni period titranja glasnica sadržan unutar okvira analize, tada se može pretpostaviti da je $u(n) = \delta(n)$, to jest pobuda je jedinični uzorak u koraku $n = 0$. Da bi ta pretpostavka bila valjana, spektralna svojstva valnog oblika glotalnog pulsa zvučnog govora moraju biti integrirana u cjelokupni model vokalnog sustava $H(z)$, tj. oba utjecaja (valni oblik pobudnog signala i frekvencijska karakteristika vokalnog trakta) se modeliraju vremenski promjenjivim sustavom $H(z)$. To zahtjeva da red prediktora, p , bude dovoljno visok da obuhvati efekte vokalnog trakta i glasnica. Za bezvučni govor najbolje je pretpostaviti da je $u(n)$ stacionaran bijeli šum, srednje vrijednosti jednake nuli i jedinične varijance.

Na osnovi ovih pretpostavki i koristeći jednadžbu (9.6-4) možemo odrediti faktor pojačanja G . Za zvučni govor pobuda sustava $H(z)$ je jednaka $G\delta(n)$, pa je stoga odziv na tu pobudu jednak impulsnom odzivu tog sustava $h(n)$, odnosno vrijedi:

$$h(n) = \sum_{k=1}^p \alpha_k h(n-k) + G\delta(n) \quad (9.6-5)$$

Može se pokazati da autokorelacijska funkcija impulsnog odziva $h(n)$ definirana kao:

$$\tilde{R}(m) = \sum_{n=0}^{\infty} h(n)h(m+n) \quad (9.6-6)$$

zadovoljava jednadžbe:

$$\tilde{R}(m) = \sum_{k=1}^p \alpha_k \tilde{R}(|m-k|) \quad m=1, 2, \dots, p \quad (9.6-7)$$

i

$$\tilde{R}(0) = \sum_{k=1}^p \alpha_k \tilde{R}(k) + G^2 \quad (9.6-8)$$

Obzirom da su jednadžbe (9.6-7) i (9.6-8) identične jednadžbi (9.4-9), slijedi da:

$$\tilde{R}(m) = R_n(m) \quad 1 \leq m \leq p \quad (9.6-9)$$

Ukupna energija u signalu $R(0)$ i impulsnom odzivu $\tilde{R}(0)$ mora biti jednaka, pa se korištenjem jednadžbi (9.4-10), (9.6-4), i (9.6-8) dobiva:

$$G^2 = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k) = E_n \quad (9.6-10)$$

Interesantno je primijetiti da jednadžba (9.6-9), kao i zahtjev da energije impulsnog odziva i signala budu jednake, zapravo zajedno zahtijevaju da prvih $p+1$ koeficijenata autokorelacijske funkcije impulsnog odziva modela budu identični prvim $p+1$ koeficijentima autokorelacijske funkcije signala govora.

U slučaju bezvučnog govora, korelacijske funkcije su definirane kao statistička očekivanja. Pretpostavlja se da je na ulazu sustava bijeli šum, srednje vrijednosti nula i jedinične varijance, to jest:

$$E[u(n)u(n-m)] = \delta(m) \quad (9.6-11)$$

Ako se sustav pobudi slučajnim signalom $Gu(n)$, a odziv sustava se nazove $g(n)$, tada je

$$g(n) = \sum_{k=1}^p \alpha_k g(n-k) + Gu(n) \quad (9.6-12)$$

Ako $\tilde{R}(m)$ označava autokorelacijsku funkciju signala $g(n)$, tada je

$$\begin{aligned} \tilde{R}(m) = E[g(n)g(n-m)] &= \sum_{k=1}^p \alpha_k E[g(n-k)g(n-m)] + E[Gu(n)g(n-m)] \\ &= \sum_{k=1}^p \alpha_k \tilde{R}(m-k) \quad m \neq 0 \end{aligned} \quad (9.6-13)$$

jer je $E[u(n)g(n-m)] = 0$ za $m > 0$ zbog nekoreliranosti $u(n)$ s nijednim signalom koji prethodi $u(n)$. Za $m = 0$ dobije se:

$$\tilde{R}(0) = \sum_{k=1}^p \alpha_k \tilde{R}(k) + GE[u(n)g(n)] = \sum_{k=1}^p \alpha_k \tilde{R}(k) + G^2 \quad (9.6-14)$$

jer je $E[u(n)g(n)] = E[u(n)(Gu(n) + \text{članovi prije } n)] = G$. Pošto energija odziva na pobudu $Gu(n)$ mora biti jednaka energiji signala, dobije se:

$$\tilde{R}(m) = R_n(m) \quad 0 \leq m \leq p \quad (9.6-15)$$

ili

$$G^2 = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k) \quad (9.6-16)$$

tj. isto kao što je bio u slučaju impulsne pobude zvučnog govora.

9.7 Postupci rješavanja sustava LPC jednadžbi

Obzirom da se postupak analize govornog signala primjenom linearne predikcije često koristi u sustavima koji rade u stvarnom vremenu, potrebno je postupak izračunavanja prediktora provesti što je moguće učinkovitije. Iako je u tu svrhu moguće primijeniti bilo koji poznati postupak rješavanja linearnog sustava jednadžbi, postoje značajne razlike u njihovoj učinkovitosti, tj. neki su jednostavniji i brži, dok su drugi složeniji i sporiji. Zbog posebnih svojstava matrice koeficijenata sustava, rješenje je moguće naći puno učinkovitije nego u općenitom slučaju. U ovom poglavlju biti će predstavljena i detaljno analizirana dva takva učinkovita postupka, od kojih se prvi odnosi na autokorelacijsku metodu, dok se drugi primjenjuje u slučaju upotrebe metode kovarijance. Diskutirat će se i svojstva dobivenih prediktora.

9.7.1 Durbinov rekurzivni postupak za rješavanje autokorelacijskih jednadžbi

Matrični oblik sustava jednadžbi za izračunavanje koeficijenata prediktora autokorelacijskom metodom ima oblik:

$$\sum_{k=1}^p \alpha_k R_n(|i-k|) = R_n(i) \quad 1 \leq i \leq p \quad (9.7-1)$$

Koristeći svojstva Toeplitz-ove strukture koju posjeduje matrica sustava jednadžbi, razvijeno je nekoliko učinkovitih rekurzivnih algoritama za rješavanje gornjeg sustava jednadžbi. Premda su Levinson-Robinson-ovi algoritmi najpopularniji i najpoznatiji, najučinkovitiji od poznatih postupaka za rješavanje ovakvog sustava jednadžbi je Durbin-ov rekurzivni algoritam koji se može opisati na sljedeći način (zbog jednostavnosti zapisa izostavljen je indeks n kod autokorelacijske funkcije R i pogreške predikcije E):

$$E^{(0)} = R(0) \quad (9.7-2)$$

$$k_i = \left(R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j) \right) / E^{(i-1)} \quad 1 \leq i \leq p \quad (9.7-3)$$

$$\alpha_i^{(i)} = k_i \quad (9.7-4)$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad 1 \leq j \leq i-1 \quad (9.7-5)$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)} \quad (9.7-6)$$

Jednadžbe od (9.7-3) do (9.7-6) rješavaju se rekurzivno za $i=1,2,\dots,p$ pa je konačno rješenje dano u obliku:

$$\alpha_j = \alpha_j^{(p)} \quad 1 \leq j \leq p \quad (9.7-7)$$

Može se uočiti da su u postupku izračunavanja koeficijenata za prediktor reda p također dobiveni i svi koeficijenti za prediktore nižeg reda, tj. $\alpha_j^{(i)}$ je j -ti koeficijent prediktora reda i .

Za ilustraciju gornjeg postupka, može se razmotriti primjer izračunavanja koeficijenata prediktora drugog reda. Izvorna matrična jednadžba ima sljedeći oblik:

$$\begin{bmatrix} R(0) & R(1) \\ R(1) & R(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \end{bmatrix}$$

Koristeći izraze (9.7-2) do (9.7-7) dobivamo:

$$E^{(0)} = R(0)$$

$$k_1 = R(1)/R(0)$$

$$\alpha_1^{(1)} = R(1)/R(0)$$

$$E^{(1)} = \frac{R^2(0) - R^2(1)}{R(0)}$$

$$k_2 = \frac{R(2)R(0) - R^2(1)}{R^2(0) - R^2(1)}$$

$$\alpha_2^{(2)} = k_2 = \frac{R(2)R(0) - R^2(1)}{R^2(0) - R^2(1)}$$

$$\alpha_1^{(2)} = \frac{R(1)R(0) - R(1)R(2)}{R^2(0) - R^2(1)}$$

$$\alpha_1 = \alpha_1^{(2)}$$

$$\alpha_2 = \alpha_2^{(2)}$$

Treba uočiti da vrijednost $E^{(i)}$ u izrazu (9.7-6) predstavlja pogrešku predikcije za prediktor reda i . Moguće je stoga, u svakom koraku rekurzivnog postupka pratiti pogrešku predikcije za prediktor reda i . Isto tako, ako se koeficijenti autokorelacije $R(k)$ zamijene skupom normaliziranih autokorelacijskih koeficijenata, tj. $r(k) = R(k)/R(0)$, tada rješenje matricne jednadžbe ostaje nepromijenjeno. Međutim, pogreška $E^{(i)}$ se onda interpretira kao normalizirana pogreška predikcije. Ako se ta normalizirana pogreška označi sa $V^{(i)}$ tada vrijedi:

$$V^{(i)} = \frac{E^{(i)}}{R(0)} = 1 - \sum_{k=1}^i \alpha_k r(k) \quad (9.7-8)$$

gdje je:

$$0 < V^{(i)} \leq 1 \quad i \geq 0 \quad (9.7-9)$$

Obzirom da je $R(0)$ jednak energiji izdvojenog segmenta govornog signala, $V^{(i)}$ definira koliko je puta energija predikcijske pogreške manja od ulazne energije signala. Normalizirana pogreška predikcije prediktora nultog reda $V^{(0)}$ jednaka je 1, jer se predikcija uopće ne provodi, dok se povećanjem reda prediktora $V^{(i)}$ smanjuje prema nuli, tj. vrijedi $V^{(i+1)} \leq V^{(i)}$. U postupku predikcije poželjno je što je moguće više smanjiti energiju predikcijske pogreške u odnosu na ulaznu energiju. Dobitak uslijed linearne predikcije se često izražava u dB, prema izrazu:

$$PG^{(i)} = -10 \log_{10}(E^{(i)} / R(0)) = -10 \log_{10}(V^{(i)}), \quad [dB] \quad (9.7-10)$$

Može se pokazati da se normalizirana pogreška predikcije za $i=p$ može zapisati i u obliku:

$$V^{(p)} = \prod_{i=1}^p (1 - k_i^2) \quad (9.7-11)$$

gdje su k_i pomoćne varijable dobivene u Durbin-ovom rekurzivnom postupku koje su uvijek u slijedećim granicama:

$$-1 \leq k_i \leq 1 \quad (9.7-12)$$

Ovaj uvjet za parametre k_i je važan jer predstavlja nužan i dovoljan uvjet da svi korijeni polinoma $A(z)$ budu unutar jedinične kružnice čime se osigurava stabilnost sustava $H(z)$. Treba dodati da kod metode kovarijance koja će biti opisana u narednim poglavljima ne postoji ovakav uvjet koji bi garantirao stabilnost sustava.

9.7.2 Primjer izračunavanja prediktora Durbinovim algoritmom

Durbinov algoritam uzima u obzir činjenicu da su elementi unutar svih desno položenih dijagonala matrice autokorelacije jednaki (Toeplitz-ova struktura), te je za njegovu izvedbu potrebno poznavati samo autokorelacijske koeficijente od $R(0)$ do $R(p)$. Provedbom gore opisanog rekurzivnog algoritma dobivamo izlaznu matricu koeficijenata linearnog prediktora u kojoj su pohranjeni koeficijenti ne samo za prediktor reda p nego i za sve prediktore od prvog do p -tog reda. Za konkretan primjer linearnog prediktora 10. reda, dobiva se matrica prikazana u tablici 9.7-1. Treba napomenuti da su elementi na glavnoj dijagonali upravo jednaki koeficijentima k_1 do k_{10} .

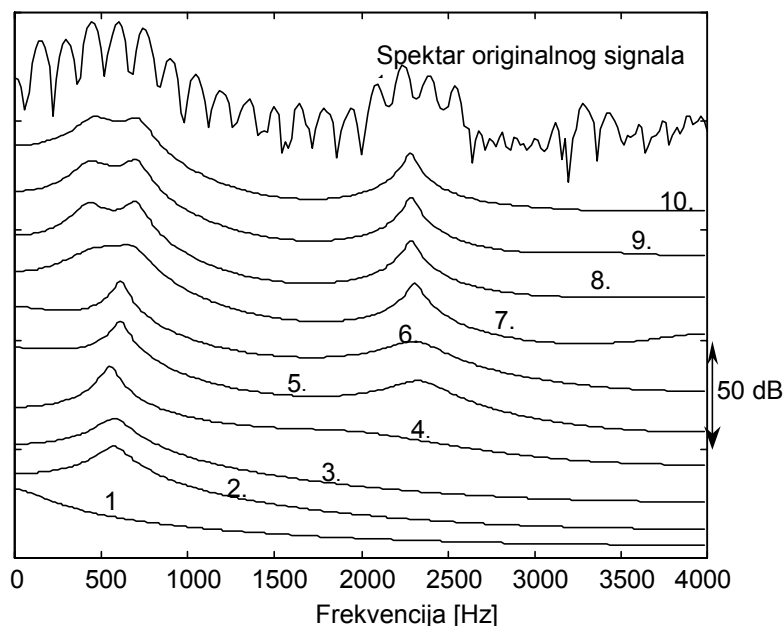
Pored matrice koeficijenata linearnog prediktora ovim algoritmom se dobiva i vektor pogreške predikcije E čijim je praćenjem moguće utvrditi dovoljan red prediktora. Brzina s kojom se $E^{(i)}$ smanjuje s povećanjem reda prediktora nije jednaka, tj. ovisi o i , a naravno ovisi i o glasu čija se predikcija provodi. Ispočetka se $E^{(i)}$ smanjuje relativno naglo, dok nakon nekog stupnja i dobitak uslijed povećanja reda prediktora više nije toliko značajan.

Tablica 9.7-1 Koeficijenti linearnog prediktora za prediktore reda 1. do 10.

RED	$\alpha_1^{(i)}$	$\alpha_2^{(i)}$	$\alpha_3^{(i)}$	$\alpha_4^{(i)}$	$\alpha_5^{(i)}$	$\alpha_6^{(i)}$	$\alpha_7^{(i)}$	$\alpha_8^{(i)}$	$\alpha_9^{(i)}$	$\alpha_{10}^{(i)}$
1.	0.75	-	-	-	-	-	-	-	-	-
2.	1.41	-0.89	-	-	-	-	-	-	-	-
3.	1.44	-0.93	0.03	-	-	-	-	-	-	-
4.	1.43	-0.62	-0.46	0.34	-	-	-	-	-	-
5.	1.52	-0.74	-0.62	0.72	-0.27	-	-	-	-	-
6.	1.44	-0.52	-0.81	0.50	0.19	-0.30	-	-	-	-
7.	1.46	-0.53	-0.84	0.55	0.23	-0.40	0.07	-	-	-
8.	1.47	-0.60	-0.81	0.64	0.09	-0.49	0.30	-0.16	-	-
9.	1.52	-0.69	-0.65	0.61	-0.11	-0.24	0.49	-0.62	0.31	-
10.	1.52	-0.70	-0.65	0.61	-0.11	-0.23	0.48	-0.63	0.33	-0.01

Na slici 9.7-1 prikazane su amplitudno frekvencijske karakteristike $|H(e^{j\Omega})|$ koje su dobivene korištenjem linearnih prediktora od prvog do desetog reda (prvi red je na dnu, a zatim redovi rastu prema gore) međusobno pomaknute za 20 dB radi bolje preglednosti, a na samom vrhu prikazan je spektar signala koji se modelira. Lako se uočava da sa povećanjem reda prediktora prijenosne funkcije $H(z)$ sve bolje slijede formantnu strukturu spektra originalnog signala. Iako je pogreška predikcije do sada bila izražavana samo u vremenskoj domeni, kao suma kvadrata razlike stvarnog i predviđenog signala, ta se ista pogreška može definirati i u spektralnoj domeni. Bit će pokazano da se to svodi na izračunavanje integrala kvocijenta spektra snage signala, $|S_n(e^{j\Omega})|^2$ i kvadrata amplitudno frekvencijske karakteristike

modela, $|H(e^{j\omega})|^2$. Uistinu, vidljivo je da se povećanjem reda predikcije prijenosna funkcija modela po obliku sve više približava obliku spektra signala tj. sve ga bolje modelira.



Slika 9.7-1 Prikaz prijenosnih funkcija za različite redove prediktora

9.7.3 Računanje prediktora za metodu kovarijance Cholesky dekompozicijom

Kod metode kovarijance, sustav jednažbi kojeg treba riješiti je oblika:

$$\sum_{k=1}^p \alpha_k \phi_n(i, k) = \phi_n(i, 0) \quad i = 1, 2, \dots, p \quad (9.7-13)$$

ili u matričnom obliku

$$\Phi \alpha = \psi \quad (9.7-14)$$

gdje je Φ pozitivno definitna simetrična matrica koja u i -tom retku i j -tom stupcu ima element $\phi_n(i, j)$, dok su α i ψ stupci (vektori) sa elementima α_i odnosno $\phi_n(i, 0)$ (i predstavlja indeks retka). Sustav jednažbi, dan izrazom (9.7-13) može biti učinkovito riješen, budući da je matrica Φ simetrična i pozitivno definitna matrica. Metoda rješavanja koja se može primijeniti na takve matrice naziva se dekompozicija Choleskog, a ponekad se naziva i metoda drugog korijena. Osnovna ideja te metode je rastavljanje matrice Φ u produkt tri matrice, tj., prema izrazu:

$$\Phi = \mathbf{V} \mathbf{D} \mathbf{V}^T \quad (9.7-15)$$

gdje je \mathbf{V} donja trokutasta matrica (svi elementi glavne dijagonale imaju iznos 1), dok je \mathbf{D} dijagonalna matrica. Oznaka T označava matričnu transpoziciju. Elementi matrice \mathbf{V} i \mathbf{D} se lako i jednoznačno određuju direktno iz izraza (9.7-15). Raspisivanjem tog izraza za element $\phi_n(i, j)$ u i -tom retku i j -tom stupcu matrice Φ dobiva se slijedeći odnos:

$$\phi_n(i, j) = \sum_{k=1}^j v_{ik} d_k v_{jk} \quad 1 \leq j \leq i-1 \quad (9.7-16)$$

Obzirom da su svi dijagonalni elementi matrice \mathbf{V} jednaki 1, izdvajanjem zadnjeg člana sume u gornjem izrazu slijedi:

$$V_{ij}d_j = \phi_n(i,j) - \sum_{k=1}^{j-1} V_{ik}d_k V_{jk} \quad 1 \leq j \leq i-1 \quad (9.7-17)$$

Specijalno za dijagonalne elemente $\phi_n(i,i)$, izraz je još jednostavniji i slijedi :

$$\phi_n(i,i) = \sum_{k=1}^i V_{ik}d_k V_{ik} \quad (9.7-18)$$

odnosno,

$$d_i = \phi_n(i,i) - \sum_{k=1}^{i-1} V_{ik}^2 d_k \quad i \geq 2 \quad (9.7-19)$$

Inicijalna vrijednost za $i=1$ dana je sa:

$$d_1 = \phi_n(1,1) \quad (9.7-20)$$

Da ilustriramo korištenje jednadžbi (9.7-15) do (9.7-20) razmotriti ćemo primjer sa $p = 4$ i matricnim elementima kraće označenim sa $\phi_{ij} = \phi_n(i,j)$. Jednadžba (9.7-15) je prema tome oblika :

$$\begin{bmatrix} \phi_{11} & \phi_{21} & \phi_{31} & \phi_{41} \\ \phi_{21} & \phi_{22} & \phi_{32} & \phi_{42} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{43} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ V_{21} & 1 & 0 & 0 \\ V_{31} & V_{32} & 1 & 0 \\ V_{41} & V_{42} & V_{43} & 1 \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix} \begin{bmatrix} 1 & V_{21} & V_{31} & V_{41} \\ 0 & 1 & V_{32} & V_{42} \\ 0 & 0 & 1 & V_{43} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Da izračunamo d_1 do d_4 i V_{ij} , počinjemo sa jednadžbom (9.7-20). Za $i = 1$ dobivamo prvi dijagonalni element matrice \mathbf{D} , tj., :

$$d_1 = \phi_{11}$$

Koristeći jednadžbu (9.7-17) za $i = 2,3,4$ računamo V_{21} , V_{31} i V_{41} , tj. određujemo prvi stupac matrice \mathbf{V} , iz čega slijedi:

$$\begin{aligned} V_{21}d_1 &= \phi_{21} & , & & V_{31}d_1 &= \phi_{31} & , & & V_{41}d_1 &= \phi_{41} \\ V_{21} &= \phi_{21}/d_1 & , & & V_{31} &= \phi_{31}/d_1 & , & & V_{41} &= \phi_{41}/d_1 \end{aligned}$$

Sada korištenjem rezultata iz prethodnog koraka, primjenom jednadžbe (9.7-19), za $i = 2$ nalazimo slijedeći dijagonalni element matrice \mathbf{D} :

$$d_2 = \phi_{22} - V_{21}^2 d_1$$

Ponovnom primjenom jednadžbe (9.7-17) za $i = 3$ i 4 nalazimo elemente drugog stupca matrice \mathbf{V} :

$$\begin{aligned}V_{32}d_2 &= \phi_{32} - V_{31}d_1V_{21} \\V_{42}d_2 &= \phi_{42} - V_{42}d_1V_{21}\end{aligned}$$

odnosno:

$$\begin{aligned}V_{32} &= (\phi_{32} - V_{31}d_1V_{21})/d_2 \\V_{42} &= (\phi_{42} - V_{42}d_1V_{21})/d_2\end{aligned}$$

Jednadžba (9.7-19) se sada koristi uz $i = 3$ za izračunavanje d_3 , zatim se jednadžba (9.7-17) koristi uz $i = 4$ za izračunavanje V_{34} , te se konačno jednadžba (9.7-19) koristi uz $i = 4$ za izračunavanje d_4 . Iz ovog je vidljivo da se u tom postupku naizmjenice nalazi po jedan dijagonalni element matrice \mathbf{D} i po jedan stupac matrice \mathbf{V} .

Nakon što se odrede matrice \mathbf{V} i \mathbf{D} izračunavanje vektora $\boldsymbol{\alpha}$ je vrlo jednostavno i provodi se primjenom dvokoračnog postupka. Kombiniranjem izraza (9.7-14) i (9.7-15) dobivamo :

$$\mathbf{VDV}^T \boldsymbol{\alpha} = \boldsymbol{\psi} \quad (9.7-21)$$

što može biti prikazano i u obliku

$$\mathbf{VY} = \boldsymbol{\psi} \quad (9.7-22)$$

gdje je \mathbf{Y} pomoćni stupac koji određuje rješenje sustava po $\boldsymbol{\alpha}$, tj. :

$$\mathbf{DV}^T \boldsymbol{\alpha} = \mathbf{Y} \quad (9.7-23)$$

odnosno, :

$$\mathbf{V}^T \boldsymbol{\alpha} = \mathbf{D}^{-1} \mathbf{Y} \quad (9.7-24)$$

Zbog donje trokutaste forme matrice \mathbf{V} , određivanje pomoćnog stupca \mathbf{Y} na osnovu matrice \mathbf{V} i stupca $\boldsymbol{\psi}$ prema izrazu (9.7-22) vrlo je jednostavno i moguće je provesti korištenjem rekursivne formule oblika :

$$Y_i = \psi_i - \sum_{j=1}^{i-1} V_{ij} Y_j \quad p \geq i \geq 2 \quad (9.7-25)$$

s početnom jednakosti:

$$Y_1 = \psi_1 \quad (9.7-26)$$

Nakon određivanja pomoćnog stupca \mathbf{Y} , provodi se množenje s \mathbf{D}^{-1} , kao što zahtijeva izraz (9.7-24), no obzirom da je matrica \mathbf{D} dijagonalna, ovo se svodi na obično dijeljenje elemenata stupca \mathbf{Y} s dijagonalnom elementima matrice \mathbf{D} (prvi s prvim, drugi s drugim itd.). Zbog gornje trokutaste forme matrice \mathbf{V}^T i izraz (9.7-24) se može rekursivno riješiti po elementima stupca $\boldsymbol{\alpha}$ koristeći slijedeći izraz:

$$\alpha_i = Y_i / d_i - \sum_{j=i+1}^p V_{ji} \alpha_j \quad 1 \leq i \leq p-1 \quad (9.7-27)$$

s početnom jednakosti:

$$\alpha_p = Y_p / d_p \quad (9.7-28)$$

Treba napomenuti da indeks i u jednadžbi (9.7-27) ide unazad od $i = p-1$ do $i = 1$. Da ilustriramo upotrebu izraza (9.7-25) do (9.7-28), nastavljamo prethodni primjer i najprije određujemo vrijednosti Y_i pretpostavljajući da su \mathbf{V} i \mathbf{D} već određeni i poznati. U matričnom obliku imamo jednadžbu :

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ V_{21} & 1 & 0 & 0 \\ V_{31} & V_{32} & 1 & 0 \\ V_{41} & V_{42} & V_{43} & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \Psi_4 \end{bmatrix}$$

Iz jednadžbi (9.7-26) i (9.7-25) dobivamo :

$$\begin{aligned} Y_1 &= \Psi_1 \\ Y_2 &= \Psi_2 - V_{21}Y_1 \\ Y_3 &= \Psi_3 - V_{31}Y_1 - V_{32}Y_2 \\ Y_4 &= \Psi_4 - V_{41}Y_1 - V_{42}Y_2 - V_{43}Y_3 \end{aligned}$$

Sa izračunatim vrijednostima Y_i rješavamo jednadžbu (9.7-24) koja je oblika

$$\begin{bmatrix} 1 & V_{21} & V_{31} & V_{41} \\ 0 & 1 & V_{32} & V_{42} \\ 0 & 0 & 1 & V_{43} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = \begin{bmatrix} 1/d_1 & 0 & 0 & 0 \\ 0 & 1/d_2 & 0 & 0 \\ 0 & 0 & 1/d_3 & 0 \\ 0 & 0 & 0 & 1/d_4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} Y_1/d_1 \\ Y_2/d_2 \\ Y_3/d_3 \\ Y_4/d_4 \end{bmatrix}$$

Iz jednadžbi (9.7-28) i (9.7-27) dobivamo :

$$\begin{aligned} \alpha_4 &= Y_4/d_4 \\ \alpha_3 &= Y_3/d_3 - V_{43}\alpha_4 \\ \alpha_2 &= Y_2/d_2 - V_{32}\alpha_3 - V_{42}\alpha_4 \\ \alpha_1 &= Y_1/d_1 - V_{21}\alpha_2 - V_{31}\alpha_3 - V_{41}\alpha_4 \end{aligned}$$

što ujedno predstavlja konačno rješenje za linearni sustav jednadžbi metode kovarijance.

Upotreba postupka dekompozicije Choleskog vodi ka vrlo jednostavnom izrazu za minimalnu pogrešku predikcije izraženu kao funkcija pomoćnog stupca \mathbf{Y} i matrice \mathbf{D} . Prisjećamo se da izraz za minimalnu pogrešku predikcije E_n u slučaju upotrebe metode kovarijance, glasio:

$$E_n = \phi_n(0,0) - \sum_{k=1}^p \alpha_k \phi_n(0,k) \quad (9.7-29)$$

ili u matričnoj notaciji :

$$E_n = \phi_n(0,0) - \boldsymbol{\alpha}^T \boldsymbol{\Psi} \quad (9.7-30)$$

Prema izrazu (9.7-24), $\boldsymbol{\alpha}^T$ je moguće zamijeniti sa $\mathbf{Y}^T \mathbf{D}^{-1} \mathbf{V}^{-1}$ pa slijedi :

$$E_n = \phi_n(0,0) - \mathbf{Y}^T \mathbf{D}^{-1} \mathbf{V}^{-1} \boldsymbol{\Psi} \quad (9.7-31)$$

Koristeći jednadžbu (9.7-22) dobivamo

$$E_n = \phi_n(0,0) - \mathbf{Y}^T \mathbf{D}^{-1} \mathbf{Y} \quad (9.7-32)$$

odnosno :

$$E_n = \phi_n(0,0) - \sum_{k=1}^p Y_k^2 / d_k \quad (9.7-33)$$

Prema tome, minimalna srednja kvadratna pogreška predikcije E_n , koja se postiže uz primjenu optimalnog prediktora, može biti određena direktno iz stupca \mathbf{Y} i dijagonalne matrice \mathbf{D} . Nadalje, izraz (9.7-33) se može koristiti za izračunavanje vrijednosti E_n i za prediktore svih nižih redova, (od 1 do $p-1$) jednostavnim skraćivanjem sume u izrazu (9.7-33) do željenog stupnja p . Na osnovu toga je moguće pratiti kako se energija predikcijske pogreške smanjuje s povećanjem reda prediktora, te odabrati potrebni red.

9.8 Predikcija mrežastom (*lattice*) strukturom

Dva do sada opisana pristupa izračunavanju koeficijenata linearnog prediktora su: metoda kovarijance i metoda autokorelacije koje se obje provode u dva osnovna koraka:

- 1) računanje matrice korelacijskih vrijednosti (\mathbf{R} ili Φ) i
- 2) izračunavanje rješenja sustava linearnih jednačini.

Iako se navedene metode često i uspješno koriste u primjenama vezanim uz obradu govora, postoji još jedan pristup problemu linearne predikcije koji se naziva predikcija mrežastom strukturom (engl. *lattice method*). Kod ove metode su oba koraka izračunavanja prediktora objedinjena u formi rekurzivnog postupka. Oba pristupa imaju zajedničku osnovu, a njihovu vezu je najlakše ilustrirati primjenom Durbinovog algoritma. U i -tom koraku tog rekurzivnog postupka, Durbinov algoritam rezultira s koeficijentima optimalnog linearnog prediktora i -tog stupnja $\{\alpha_j^{(i)} | j = 1, 2, \dots, i\}$. Koristeći te koeficijente možemo definirati prijenosnu funkciju inverznog filtra i -tog reda:

$$A^{(i)}(z) = 1 - \sum_{j=1}^i \alpha_j^{(i)} z^{-j} \quad (9.8-1)$$

Ako je ovaj filter pobuđen sa segmentom govornog signala $s_n(m) = s(n+m)w(m)$, tada će na njegovom izlazu biti pogreška predikcije, $e_n^{(i)}(m) = e^{(i)}(n+m)$ gdje je:

$$e^{(i)} = s(m) - \sum_{j=1}^i \alpha_j^{(i)} s(m-j) \quad (9.8-2)$$

$$= s(m) - \tilde{s}_f^{(i)}(m) \quad (9.8-3)$$

Radi jednostavnosti zapisa namjerno je ispušten indeks n koji označava segment signala i predikcijske pogreške u okolini uzorka n . Predikcija uzorka $s(m)$ primjenom prediktora reda i označena je sa $\tilde{s}_f^{(i)}(m)$ (značenje sufiksa f biti će objašnjeno kasnije). Ako na izraz (9.8-2) primijenimo z -transformaciju vrijedi:

$$E^{(i)}(z) = A^{(i)}(z)S(z) \quad (9.8-4)$$

U Durbin-ovom postupku pokazano je kako se na osnovu koeficijenata prediktora stupnja $(i-1)$ i koeficijenta k_i određuju koeficijenti prediktora i -tog stupnja tj. vrijedi:

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad \text{za } j=1,2,\dots,i-1. \quad (9.8-5)$$

Ako bi ovaj izraz uvrstili u prijenosnu funkciju inverznog filtra (9.8-1) tada slijedi:

$$\begin{aligned} A^{(i)}(z) &= 1 - \alpha_i^{(i)} z^{-i} - \sum_{j=1}^{i-1} \alpha_j^{(i)} z^{-j} \\ &= 1 - k_i z^{-i} - \sum_{j=1}^{i-1} \alpha_j^{(i)} z^{-j} \\ &= 1 - k_i z^{-i} - \sum_{j=1}^{i-1} (\alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}) z^{-j} \end{aligned} \quad (9.8-6)$$

odnosno, rastavljanjem na dvije sume :

$$A^{(i)}(z) = \left(1 - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} z^{-j} \right) - k_i z^{-i} \left(1 - \sum_{j=1}^{i-1} \alpha_{i-j}^{(i-1)} z^{i-j} \right) \quad (9.8-7)$$

U prvom članu prepoznamo prijenosnu funkciju prediktora reda $(i-1)$, tj. $A^{(i-1)}(z)$, dok ćemo drugi član malo modificirati, tako da indeks sumacije j , zamijenimo sa $j'=i-j$, pa slijedi:

$$\begin{aligned} A^{(i)}(z) &= A^{(i-1)}(z) - k_i z^{-i} \left(1 - \sum_{j'=(i-1)}^1 \alpha_{j'}^{(i-1)} z^{j'} \right) \\ &= A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}) \end{aligned} \quad (9.8-8)$$

Treba naglasiti da prijenosna funkcija prediktora $A^{(i-1)}$ u drugom članu ima argument z^{-1} , a ne z kao što je to u prvom članu. Uvrstavajući izraz (9.8-8) u izraz (9.8-4) dobivamo:

$$E^{(i)}(z) = A^{(i-1)}(z) S(z) - k_i z^{-i} A^{(i-1)}(z^{-1}) S(z) \quad (9.8-9)$$

Prvi dio izraza (9.8-9) je očito z -transformacije pogreške predikcije za prediktor reda $(i-1)$. Drugi dio izraza se može malo modificirati tako da izlučimo $k_i z^{-1}$, pa slijedi :

$$\begin{aligned} E^{(i)}(z) &= E^{(i-1)}(z) - k_i z^{-1} \left(z^{-(i-1)} A^{(i-1)}(z^{-1}) S(z) \right) \\ &= E^{(i-1)}(z) - k_i z^{-1} B^{(i-1)}(z) \end{aligned} \quad (9.8-10)$$

Član u zagradi označen je sa $B^{(i-1)}(z)$, tj. općenito vrijedi:

$$B^{(i)}(z) = z^{-i} A^{(i)}(z^{-1}) S(z) \quad (9.8-11)$$

Smisao ovog člana lakše je shvatiti ako se odredi inverzna z -transformacija od $z^{-1} B^{(i-1)}(z)$ koji nakon raspisivanja glasi:

$$\begin{aligned}
z^{-1}B^{(i-1)}(z) &= z^{-i}A^{(i-1)}(z^{-1})S(z) \\
&= z^{-i}\left(1 - \alpha_1^{(i-1)}z - \alpha_2^{(i-1)}z^2 - \dots - \alpha_{(i-1)}^{(i-1)}z^{(i-1)}\right)S(z) \\
&= z^{-i}S(z) - \alpha_1^{(i-1)}z^{-(i-1)}S(z) - \dots - \alpha_{(i-1)}^{(i-1)}z^{-1}S(z)
\end{aligned} \tag{9.8-12}$$

Nakon prebacivanja u vremensku domenu slijedi:

$$\begin{aligned}
b^{(i-1)}(m-1) &= s(m-i) - \sum_{j=1}^{(i-1)} \alpha_j^{(i-1)}s(m-(i-j)) \\
&= s(m-i) - \tilde{S}_b^{(i-1)}(m-i)
\end{aligned} \tag{9.8-13}$$

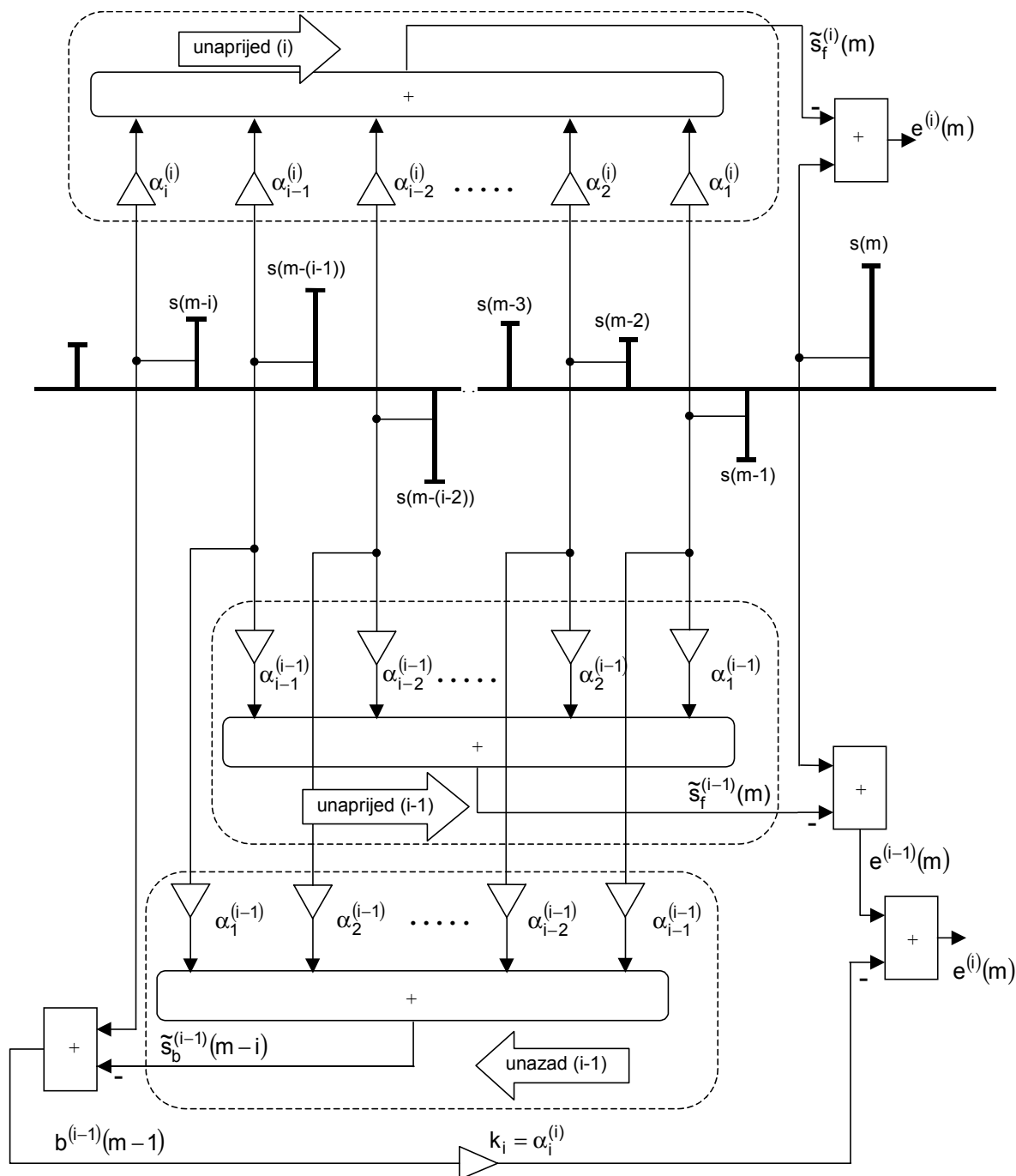
Iz izraza (9.8-13) može se prepoznati da $b^{(i-1)}(m-1)$ predstavlja pogrešku predikcije unazadnog prediktora reda $(i-1)$ (engl. *backward prediction error sequence*) koji provodi predikciju uzorka $s(m-i)$ na osnovu $i-1$ uzoraka koji slijede, tj. na osnovu $\{s(m-i+j)\}$, za $j=1,2,\dots,i-1$. Ta unazadna predikcija označena je sa $\tilde{S}_b^{(i-1)}(m-i)$, gdje sufiks $_b$ dolazi od *backward* (unazad). Treba naglasiti da ovaj unazadni prediktor koristi iste koeficijente $\{\alpha_j^{(i-1)}, j=1,2,\dots,i-1\}$ kao i normalni unaprijedni prediktor reda $(i-1)$, koji provodi predikciju uzorka $s(m)$ označenu sa $\tilde{S}_f^{(i-1)}(m)$ na osnovu $i-1$ uzoraka koji mu prethode, tj. na osnovu $\{s(m-j)\}$, za $j=1,2,\dots,i-1$ (sufiks $_f$ dolazi od *forward* (unaprijed)). Taj unaprijedni prediktor odgovara prvom članu u izrazima (9.8-8) i (9.8-10), a njegova pogreška predikcije označena je sa $e^{(i-1)}(m)$. Interesantno je uočiti da se predikcija u oba slučaja provodi na osnovu istih uzoraka signala i na osnovu istih koeficijenata prediktora, jedino što se u prvom slučaju predviđa uzorak $s(m-i)$, a u drugom slučaju uzorak $s(m)$. Razlika je i u tome što je redoslijed koeficijenata obrnut, tj. u prvom slučaju $\alpha_1^{(i-1)}$ množi uzorak $s(m-i+1)$, dok u drugom slučaju taj se koeficijent množi sa $s(m-1)$.

Uz ovako definiranu unaprijednu i unazadnu pogrešku predikcije prediktora reda $(i-1)$, lako je pokazati da se inverznom z -transformacijom izraza (9.8-10) dobiva slijedeća jednadžba diferencija za pogrešku predikcije unaprijednog prediktora reda i , $e^{(i)}(m)$:

$$e^{(i)}(m) = e^{(i-1)}(m) - k_i b^{(i-1)}(m-1) \tag{9.8-14}$$

Drugim riječima, pogrešku predikcije unaprijednog prediktora i -tog reda moguće je odrediti na osnovu unaprijedne i unazadne pogreške prediktora reda $(i-1)$ i na osnovu poznavanja koeficijenta $k_i = \alpha_i^{(i)}$.

Opisana zakonitost ilustrirana je i na slici 9.8-1, gdje su prikazana sva tri prediktora: unaprijedni prediktor reda (i) (na vrhu), unaprijedni prediktor reda $(i-1)$ (u sredini) i unazadni prediktor reda $(i-1)$ (na dnu). Pogrešku predikcije $e^{(i)}(m)$ moguće je dobiti bilo kao izlaz prvog prediktora, bilo kombiniranjem pogreški $e^{(i-1)}(m)$ i $b^{(i-1)}(m-1)$ prema izrazu (9.8-14).



Slika 9.8-1 Ilustracija predikcije unaprijed i predikcije unazad koristeći prediktore reda $(i-1)$

Slična zakonitost vrijedi i za unazadnu pogrešku prediktora reda i , označenu sa $b^{(i)}(m)$. To je moguće pokazati uvrštavanjem izraza (9.8-8) u (9.8-11) nakon čega dobivamo:

$$\begin{aligned}
 B^{(i)}(z) &= z^{-i}A^{(i-1)}(z^{-1})S(z) - z^{-i}k_i z^i A^{(i-1)}(z)S(z) \\
 &= z^{-1}z^{-(i-1)}A^{(i-1)}(z^{-1})S(z) - k_i A^{(i-1)}(z)S(z)
 \end{aligned}
 \tag{9.8-15}$$

odnosno :

$$B^{(i)}(z) = z^{-1}B^{(i-1)}(z) - k_i E^{(i-1)}(z)
 \tag{9.8-16}$$

Prebacivanjem u vremensku domenu slijedi da se unazadna pogreška predikcije prediktora reda i , također može odrediti kombiniranjem unaprijedne i unazadne pogreške prediktora reda $(i-1)$, tj. :

$$b^{(i)}(m) = b^{(i-1)}(m-1) - k_i e^{(i-1)}(m) \quad (9.8-17)$$

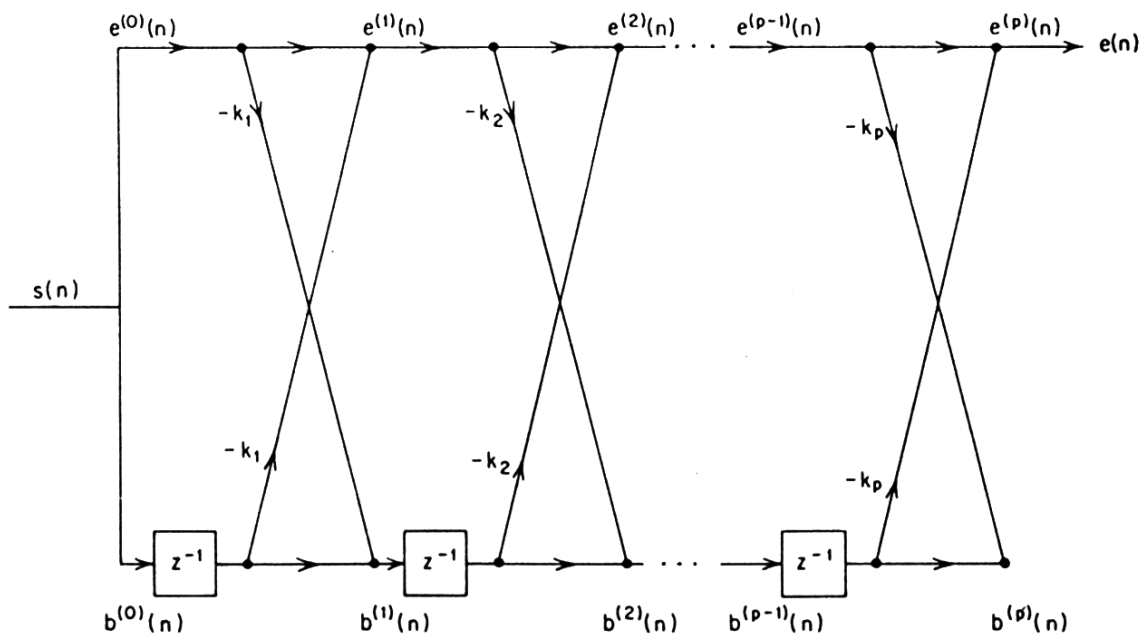
Izrazi (9.8-14) i (9.8-17) predstavljaju svojevrsnu rekurzivnu formulu kojom je moguće na osnovu predikcijske pogreške prediktora nižeg reda odrediti unaprijednu i unazadnu predikcijsku pogrešku prediktora prvog višeg reda. Kao i kod svake rekurzivne formule, potrebno je definirati inicijalnu vrijednost za slučaj $i=0$, tj. za slučaj kada se predikcija uopće ne provodi. Tada vrijedi:

$$e^{(0)}(m) = b^{(0)}(m) = s(m) \quad (9.8-18)$$

tj. predikcijske pogreške su jednake ulaznom signalu $s(m)$.

Rekurzivne formule (9.8-14) i (9.8-17) možemo prikazati i pomoću grafa toka signala na slici 9.8-2. Takva struktura se naziva mrežastom (engl. *lattice*) strukturom. Ako mrežu proširimo na p sekcija, tada izlaz iz posljednje gornje grane na slici 9.8-2 predstavlja unaprijednu pogrešku predikcije prediktora p -tog reda. Slika 9.8-2 u stvari prikazuje još jednu moguću izvedbu inverznog filtra (engl. *prediction error filter*) čija je prijenosna funkcija $A(z)$ dana izrazom (9.8-1) uz $i=p$.

Potrebno je naglasiti da je mrežasta struktura na slici 9.8-2 direktna posljedica primjene Durbin-ovog algoritma i parametara k_i dobivenih izrazima (9.7-2) do (9.7-7). Koeficijenti predikcije $\{\alpha_j^{(p)}, j=1,2,\dots,p\}$ nigdje ne figuriraju eksplicitno na slici 9.8-2, već indirektno preko parametara k_i . Itakura je pokazao da se optimalni parametri k_i mogu odrediti direktno na osnovu unaprijedne i unazadne pogreške predikcije prediktora reda $i-1$, a zbog prirode mrežaste strukture taj se postupak može provesti rekurzivno za sve koeficijente $\{k_i, i=1,2,\dots,p\}$ bez da se eksplicitno izračunaju koeficijenti $\{\alpha_j^{(p)}, j=1,2,\dots,p\}$.



Slika 9.8-2 Blok dijagram izvedbe inverznog filtra $A(z)$ mrežastom (lattice) strukturom

To se provodi korištenjem slijedećeg izraza:

$$k_i = \frac{\sum_{m=0}^{N-1} e^{(i-1)}(m)b^{(i-1)}(m-1)}{\left\{ \sum_{m=0}^{N-1} \left(e^{(i-1)}(m) \right)^2 \sum_{m=0}^{N-1} \left(b^{(i-1)}(m-1) \right)^2 \right\}^{1/2}} \quad (9.8-19)$$

Ovaj izraz je oblika normalizirane kros-korelacijske funkcije koja mjeri sličnost unaprijedne i unazadne pogreške predikcije prediktora reda $(i-1)$. Zbog toga se koeficijenti k_i uobičajeno nazivaju koeficijentima parcijalne korelacije ili kraće PARCOR koeficijentima, a nekad se i nazivaju samo k -parametri.

Izraz (9.8-19) zamjenjuje izraz (9.7-3) u Durbin-ovom algoritmu, a ako je baš potrebno koeficijenti prediktora, $\alpha_j^{(p)}$, se i dalje mogu izračunati rekurzivno iz koeficijenata k_i kao i prije. Određivanje optimalnih koeficijenata prediktora prema izrazu (9.8-19) se naziva PARCOR analiza. Takav pristup predstavlja alternativu klasičnim postupcima kod kojih se u svrhu izračunavanja prediktora mora provesti inverzija matrice. Rezultati koji se dobivaju ovakvim postupkom su identični onima koji bi se dobili autokorelacijskom metodom, tj. niz PARCOR koeficijenata je određen tako da minimizira srednju kvadratnu pogrešku predikcije unaprijed. Najvažnije od svega je da ovakav pristup otvara jednu cijelu novu klasu postupaka temeljenih na mrežastoj strukturi.

9.8.1 Burgov algoritam

U dosadašnjim razmatranjima koeficijenti prediktora k_i određivani su tako da minimiziraju isključivo srednju kvadratnu pogrešku unaprijedne predikcije. Burg je predložio modifikaciju tog postupka, kod koje se minimizira suma srednje kvadratne pogreške unaprijedne i unazadne predikcije sa slike 9.8-2, tj. prema izrazu:

$$\tilde{E}^{(i)} = \sum_{m=0}^{N-1} \left[\left(e^{(i)}(m) \right)^2 + \left(b^{(i)}(m) \right)^2 \right] \quad (9.8-20)$$

Uvrstimo li izraze (9.8-14) i (9.8-17) u izraz (9.8-20), te ako odredimo parcijalnu derivaciju po koeficijentu k_i dobivamo:

$$\begin{aligned} \frac{\partial \tilde{E}^{(i)}}{\partial k_i} = & -2 \sum_{m=0}^{N-1} \left[e^{(i-1)}(m) - k_i b^{(i-1)}(m-1) \right] b^{(i-1)}(m-1) \\ & - 2 \sum_{m=0}^{N-1} \left[b^{(i-1)}(m-1) - k_i e^{(i-1)}(m) \right] e^{(i-1)}(m) \end{aligned} \quad (9.8-21)$$

Izjednačimo li parcijalnu derivaciju s nulom i riješimo po k_i slijedi:

$$k_i = \frac{2 \sum_{m=0}^{N-1} \left[e^{(i-1)}(m) b^{(i-1)}(m-1) \right]}{\sum_{m=0}^{N-1} \left[e^{(i-1)}(m) \right]^2 + \sum_{m=0}^{N-1} \left[b^{(i-1)}(m-1) \right]^2} \quad (9.8-22)$$

Slično kao i kod Durbin-ovog algoritma, ako su koeficijenti k_i izračunati korištenjem izraza (9.8-22) tada i dalje vrijedi: $-1 \leq k_i \leq 1$, što garantira stabilnost LPC filtra. Naravno

treba imati u vidu da će se ovi k_i koeficijenti razlikovati od onih određenih izrazom (9.8-19), koji su u stvari identični koeficijentima koji se dobivaju autokorelacijskim postupkom.

Na kraju ponovimo ukratko koji su koraci korišteni za izračunavanje optimalnih koeficijenata predikcije i k-parametara:

1)	Inicijalizacija: $e^{(0)}(m) = b^{(0)}(m) = s(m)$.
2)	Izračunavanje $k_1 = \alpha_1^{(1)}$ prema izrazu (9.8-22)
3)	Određivanje pogrešaka predikcije unaprijed $e^{(1)}(m)$ i unazad $b^{(1)}(m)$ izrazima (9.8-14) i (9.8-17)
4)	Postavi $i=2$.
5)	Određivanje $k_i = \alpha_i^{(i)}$ izrazom (9.8-22).
6)	Određivanje $\alpha_j^{(i)}$ za $j=1,2,\dots,i-1$ iz izraza (9.7-5), (prema potrebi)
7)	Određivanje $e^{(i)}(m)$ i $b^{(i)}(m)$ iz izraza (9.8-14) i (9.8-17).
8)	Povećaj i za jedan.
9)	Ako je i manje ili jednako p vrati se na korak 5).
10)	Inače procedura je završena.

Vidjeli smo da postoji nekoliko razlika između izvedbe predikcije mrežastom strukturom i izvedbe metodama kovarijance ili autokorelacije, no osnovna razlika je u tome što se mrežastom metodom predikcijski koeficijenti računaju direktno iz uzoraka govora bez posrednog računanja korelacijskih funkcija. Istovremeno metoda jamči stabilan filter i bez korištenja vremenskog otvora koji je nužan kod metode autokorelacije. Iz tih je razloga formulacija linearne predikcije mrežastom strukturom ostala važan i opravdan pristup u analizi govora linearnom predikcijom.

9.9 Usporedba postupaka linearne predikcije

U prethodnim poglavljima već su razmotrene teoretske razlike u formulacijama jednadžbi linearne prediktivne analize za autokorelacijsku metodu, za metodu kovarijance i metodu pomoću mrežaste (lattice) strukture. U ovom poglavlju razmotrit će se razlike vezane uz praktičnu primjenu jednadžbi prediktivne analize. U to su uključena razmatranja numeričke složenosti postupaka, stabilnost rješenja, kao i pitanje kako odabrati red prediktora i duljinu govornog segmenta korištenog za analizu. Prvo će biti razmotren problem numeričke složenosti postupka određivanja prediktorskih koeficijenata iz valnog oblika govora.

Dva glavna aspekta pri proračunu prediktorskih koeficijenata su potrebna količina memorije i ukupan broj množenja. Tablica 9.9-1 prikazuje potrebne proračunske zahtjeve za kovarijancijsku, autokorelacijsku i metodu s mrežastom strukturom. Što se tiče memorijskog prostora za metodu kovarijance je u osnovi potrebno N_1 lokacija za podatke, i reda veličine $p^2/2$ lokacija za korelacijsku matricu, gdje je N_1 broj uzoraka analiziranog segmenta govora, a p red predikcije. Za autokorelacijsku metodu potrebno je N_2 lokacija i za podatke i za vremenski otvor, a broj lokacija potreban za pohranu autokorelacijske matrice iznosi $p+1$. Za analizu mrežastom strukturom potrebno je $3N_3$ lokacija za podatke i unaprijednu i unazadnu pogrešku predikcije (engl. *forward and backward prediction errors*). Treba istaknuti da broj uzoraka signala: N_1 za kovarijancijsku, N_2 za autokorelacijsku i N_3 za mrežastu metodu, ne

moraju nužno biti jednaki. Ovo pitanje će biti razmotreno kasnije u ovom poglavlju. Dakle, po pitanju potrebnog memorijskog prostora (uz pretpostavku da su N_1 , N_2 i N_3 sumjerljivi) kovarijancijska i autokorelacijska metoda trebaju nešto manje prostora nego metoda pomoću mrežaste strukture.

Tablica 9.9-1 Usporedba tri LPC postupka obzirom na utrošak memorije i numeričku složenost

		Kovarijancijska metoda	Autokorelacijska metoda	Mrežasta (lattice) metoda
izračunavanje prediktora		(Cholesky dekompozicija)	(Durbin-ova metoda)	(Burg-ova metoda)
memorijski prostor	Podaci	N_1	N_2	$3N_3$
	Matrica	$\sim p^2/2$	$\sim p$	-
	Vremenski otvor	0	N_2	-
broj množenja	Vremenski otvor	0	N_2	-
	Korelacija	$\sim N_1 p$	$\sim N_2 p$	-
	Rješavanje matrice	$\sim p^3$	$\sim p^2$	$5N_3 p$

Numerička složenost izražena u broju množenja, za svaku od te tri metode prikazana je pri dnu tablice 9.9-1. Za metodu kovarijance, proračun korelacijske matrice zahtijeva oko $N_1 p$ množenja, dok samo rješavanje matrice jednadžbe (koristeći Cholesky dekompoziciju) zahtijeva broj množenja proporcionalan sa p^3 . Za autokorelacijsku metodu proračun korelacijske matrice zahtijeva oko $N_2 p$ množenja, dok je rješavanje matrice jednadžbe jednostavnije nego kod metode kovarijance i zahtijeva oko p^2 množenja. Dakle ako su N_1 i N_2 približno jednaki, a $N_1 \gg p$, $N_2 \gg p$, tada autokorelacijska metoda zahtijeva nešto manji broj operacija nego kovarijancijska metoda. Dakako, pošto je u većini problema vezanih uz govor broj množenja potrebnih da se izračunaju elementi korelacijske matrice daleko veći od broja množenja potrebnih da se riješi matična jednadžba, trajanje proračuna za obje ove metode je približno jednako. Za mrežastu strukturu potrebno je ukupno $5N_3 p$ množenja da se izračuna skup koeficijenata parcijalne korelacije (engl. *PARCOR coefficients*), koji se još često naziva skup k-koeficijenata. Dakle kod metode mrežaste strukture rješavanje LPC jednadžbi je najsporije. Naravno, na umu treba imati neke druge prednosti mrežaste metode kada se razmatra mogućnost njene upotrebe.

Drugi problem pri usporedbi ovih triju formulacija je stabilnost dobivenog sustava.

$$H(z) = \frac{G}{A(z)} \quad (9.9-1)$$

Taj sustav je stabilan ako svi polovi leže strogo unutar jedinične kružnice u z-ravnini. Polovi sustava $H(z)$, su nule polinoma u nazivniku, $A(z)$, gdje:

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k} \quad (9.9-2)$$

Kao što je već prije navedeno, za autokorelacijsku metodu sve nule $A(z)$ uvijek leže unutar jedinične kružnice - dakle, $H(z)$ je garantirano stabilan. Potrebno je uočiti da ova teoretska garancija stabilnosti za autokorelacijsku metodu neće biti ostvarena u praksi ako autokorelacijska funkcija nije izračunata sa dovoljnom preciznošću. Tako npr. u slučaju izvedbe na procesorima s cjelobrojnom aritmetikom zaokruživanje pri računanju autokorelacije može dovesti do toga da matrica bude blizu singulariteta, tj. da nema dobro definiran inverz (engl. *ill-conditioned*). Markel i Gray su pokazali da ti neželjeni efekti mogu biti umanjeni tako da se spektar govora prije primjene LPC postupaka učini što je moguće ravnijim. Taj postupak poravnavanja spektra se provodi filtracijom ulaznog govornog signala filtrom za pred-naglašavanje (engl. *preemphasis filter*). U slučaju upotrebe takvog pred-filtra mogu biti korištene i riječi sa manjim bojem bita, a da rezultirajući prediktorski polinom uglavnom uvijek ostane stabilan. Durbinov algoritam daje dobar test stabilnosti, jer je nužno i dovoljno da parametri k_i zadovoljavaju uvjet:

$$-1 \leq k_i \leq 1 \quad (9.9-3)$$

Dakle, ako u procesu određivanja prediktorskih koeficijenata $\{\alpha_i\}$, bilo koji od k_i prekrši jednadžbu (9.9-3) to znači da postoje korijeni $A(z)$ koje leže izvan jedinične kružnice.

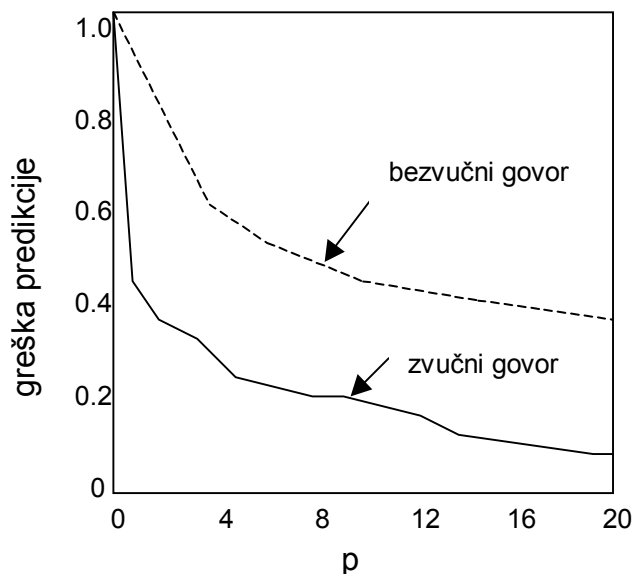
Za metodu kovarijance stabilnost prediktorskog polinom se ne može garantirati. Naravno, u praksi, ako je broj uzoraka u segmentu dovoljno velik, tada će dobiveni prediktorski polinom gotovo uvijek biti stabilan. To je zbog činjenice što za velike brojeve uzoraka u analiziranom segmentu, kovarijancijska i autokorelacijska metoda daju gotovo identične rezultate.

Za metodu pomoću mrežaste strukture prediktorski polinom je garantirano stabilan jer su prediktorski koeficijenti dobiveni iz koeficijenata parcijalne korelacije koji, po definiciji, zadovoljavaju jednadžbu (9.9-3). Nadalje, stabilnost je očuvana čak i kada je proračun izveden uz korištenje aritmetike s konačnom duljinom riječi.

U slučaju korištenja LPC postupaka kod kojeg stabilnost nije garantirana, potrebno je odrediti polove sustava $H(z)$ i provjeriti da li se svi polovi nalaze unutar jedinične kružnice. Za polove sa radijusom $r > 1$ moguće je provesti postupak stabilizacije, kojim se taj pol reflektira unutar kružnice na radijus $1/r$, a uz isti kut ω . Takvim postupkom dobiva se sustav $H'(z)$ s jednakom amplitudno frekvencijskom karakteristikom kao i sustav $H(z)$, ali kod kojeg su svi polovi unutar jedinične kružnice.

Druga dva aspekta u usporedbi tih triju formulacija LPC postupaka su izbor reda polinoma $A(z)$, koji je označen sa p i izbor širine vremenskog otvora, tj. duljine segmenta govornog signala, N , nad kojim se provodi LPC postupak. Izbor za p ovisi najviše o frekvenciji otipkavanja i u osnovi ne ovisi o tipu LPC postupka koji se koristi. Sustav $H(z)$ modelira združeno djelovanje frekvencijske karakteristike vokalnog trakta, spektra pobudnog signala, kao i frekvencijske karakteristike zračenja na usnicama. Što se tiče utjecaja vokalnog trakta može se pretpostaviti da će govorni signal otipkan s frekvencijom otipkavanja f_s imati u prosjeku $f_s/1000$ formantnih (rezonantnih) karakteristika, tj. jednu po svakom kHz frekvencije otipkavanja. Tako npr. uz $f_s=10\text{kHz}$, u području od 0 do $f_s/2=5\text{kHz}$ govorni signal će imati u prosjeku oko 5 formanata. Obzirom da je svaki formant opisan jednim konjugirano kompleksnim parom polova, potrebni red LPC sustava za vjernu reprezentaciju ovih karakteristika iznosi $p=2 \times 5=10$. Na ovo je potrebno dodati još tri do četiri pola za reprezentaciju spektra pobudnog signala i frekvencijske karakteristike zračenja na usnicama. Dakle za $f_s=10\text{kHz}$ potreban iznos za p je oko 13 do 14. Potvrda ovog zaključka vidljiva je na slici 9.9-1 koja prikazuje normalizirane efektivne vrijednosti pogreške predikcije u odnosu na red prediktora p i to za zvučni i bezvučni govor uz frekvenciju otipkavanja 10 kHz. Iako se pogreška predikcije lagano smanjuje kako p raste, za p reda 13-14 krivulja se u osnovi

izravnava i pokazuje samo malo smanjenje kako p dalje raste. Iz ove slike se može primijetiti da je normalizirana pogreška predikcije za bezvučni govor znatno viša nego za zvučni govor. To je naravno bilo očekivano jer model za bezvučni govor nije ni približno tako točan kao za zvučni govor.



Slika 9.9-1 Promjena efektivne vrijednosti predikcijske pogreške sa redom prediktora

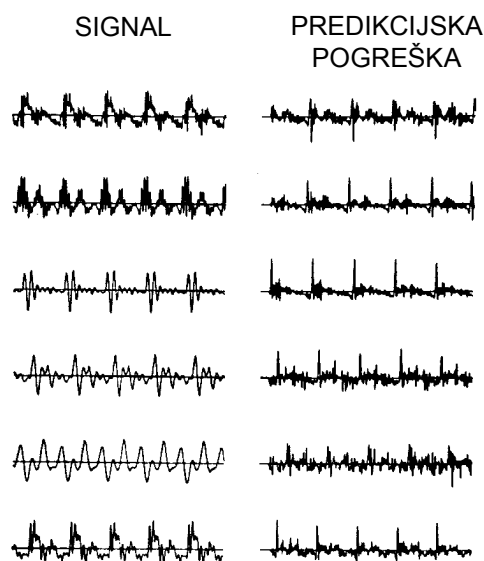
Pravilni izbor duljine segmenta N vrlo je značajan za većinu sustava temeljenih na postupku linearne predikcije (LPC). Naravno, pogodno je odabrati N što je moguće manji jer je ukupno računsko opterećenje, za sve tri metode, u osnovi proporcionalno s N . Za autokorelacijsku metodu je pokazano da N mora biti reda veličine nekoliko perioda osnovne frekvencije titranja glasnica da se osiguraju pouzdani rezultati. Kako se u autokorelacijskoj metodi koristi množenje segmenta vremenskim otvorom, duljina odsječka mora biti dovoljno velika da efekti odsijecanja, odnosno prigušenja rubova otvorom (engl. *tapering effects*) ne utječu ozbiljnije na rezultate. Dakle, u LPC primjenama autokorelacijske metode koriste se širine segmenta od $N=100$ do $N=400$ uzoraka (pri $f_s=10\text{kHz}$), a kod većine sustava broj uzoraka N je bliži gornjoj granici. I za metodu kovarijance i za metodu mrežastom strukturom izbor duljine segmenta je određen sa sljedećih nekoliko razmatranja. Kako nema potrebe za množenjem s vremenskim otvorom, nema stvarnog ograničenja na minimalnu širinu odsječka. Ako se analiza provede na osnovu uzoraka govornog signala između dva glotalna pulsa, tj. unutar jedne periode osnovne frekvencije (npr. korištenjem analize sinkrone s osnovnom frekvencijom glasnica), tada se već i sa širinama N reda veličine $2p$ mogu ostvariti zadovoljavajući rezultati. To znači da vremenski otvor mora biti tako pozicioniran u odnosu na govorni signal da u njemu bude sadržano prigušeno istitravanje vokalnog trakta na pobudni glotalni puls, dakle signal koji prema kraju intervala analize eksponencijalno trne. Naravno ako se koristi tako malen N i ako se glotalni puls (engl. *pitch pulse*) pojavi unutar otvora analize (signal prvo raste a zatim pada) neće se dobiti zadovoljavajući rezultati. Dakle u većini praktičnih sustava u kojima nije moguće upotrijebiti postupak sinkroniziran sa osnovnom frekvencijom glasnica, širina vremenskog otvora analize za kovarijancijsku i mrežastu metodu odabere se na isti način kao kod autokorelacijske metode.

9.10 Signal predikcijske po greške

Nusproizvod LPC analize je određivanje signala pogreške $e(n)$, definiranog kao:

$$e(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k) = Gu(n) \quad (9.10-1)$$

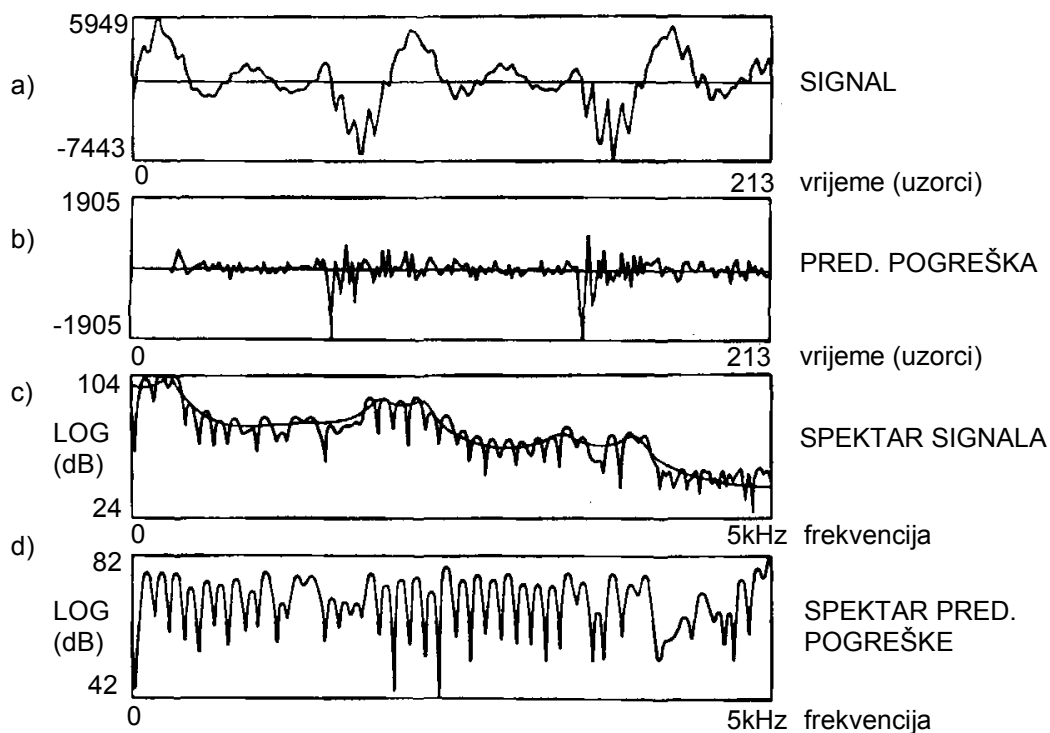
Ako je stvarni govorni signal generiran sustavom koji se može dobro modelirati vremenski promjenjivim prediktorom reda p , tada je $e(n)$ jednako dobra aproksimacija pobudnog signala tog sustava.



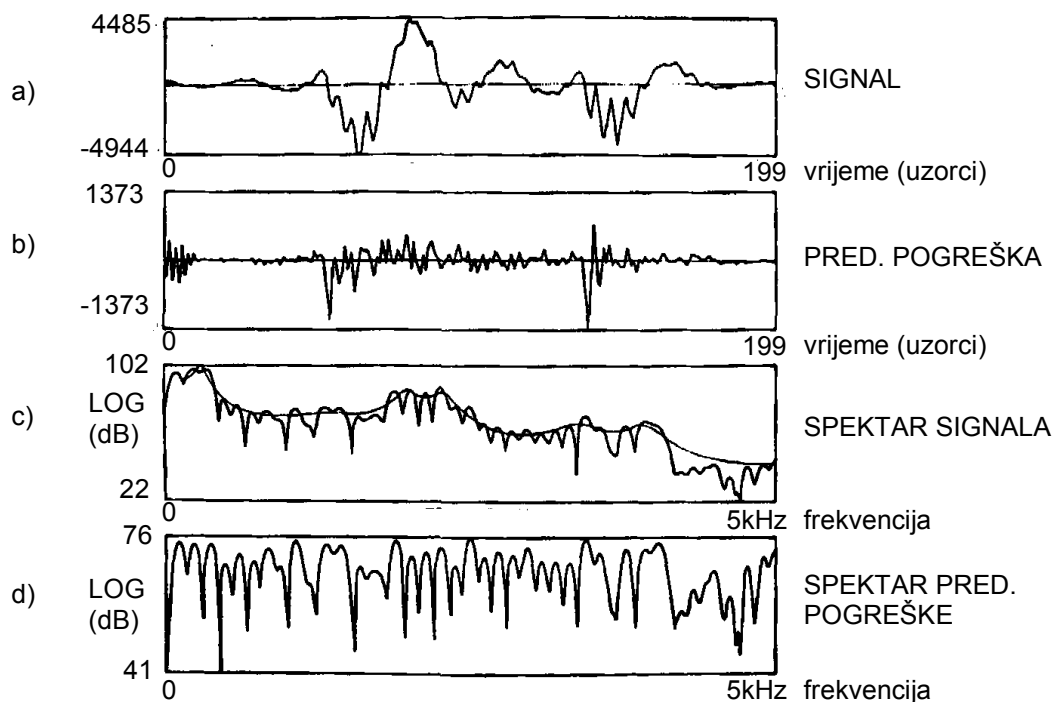
Slika 9.10-1 Primjer signala i predikcijske pogreške samoglasnika i , e , a , o , u i y

Takvim razmišljanjem može se očekivati da će predikcijska pogreška biti velika (za zvučni govor) na početku svakog pitch perioda. Stoga se pitch period može odrediti pomoću pozicija uzoraka $e(n)$ koji imaju veliku amplitudu, to jest pitch period se može definirati kao vremenski razmak između parova uzoraka $e(n)$ čija amplituda prelazi određeni prag. Inače, pitch period se može procijeniti i izračunavanjem autokorelacijske funkcije signala predikcijske pogreške $e(n)$ i detekcijom najvišeg vrha u odgovarajućem opsegu pomaka. Korisnost signala predikcijske pogreške $e(n)$ u određivanju pitch perioda može se tumačiti i time da je spektar signala predikcijske pogreške približno ravan, iz čega slijedi da su efekti formantata uklonjeni iz signala predikcijske pogreške, što omogućava točniju estimaciju pitch perioda.

Za ilustraciju prirode signala predikcijske pogreške, slika 9.10-1 prikazuje niz odsječaka valnih oblika nekoliko samoglasnika i pripadajućih signala predikcijskih pogrešaka $e(n)$. Za ove jednostavne zvukove samoglasnika, signal predikcijske pogreške $e(n)$ ima oštre pulseve na razmacima koji odgovaraju pitch periodu tih samoglasnika.



Slika 9.10-2 Tipični signali i spektri za LPC metodu kovarijance, za muškog govornika, $p = 14$, $N = 200$

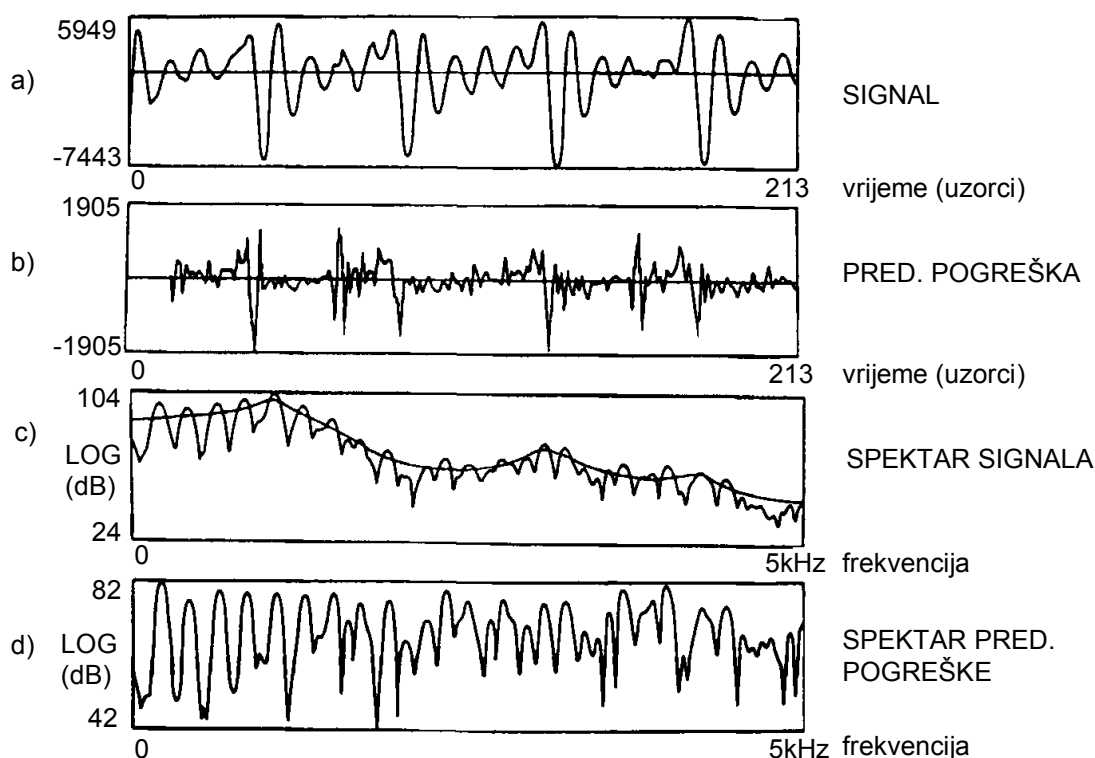


Slika 9.10-3 Tipični signali i spektri za LPC autokorelacijsku metodu sa Hamming-ovim vremenskim otvorom, za muškog govornika, $p = 14$, $N = 200$

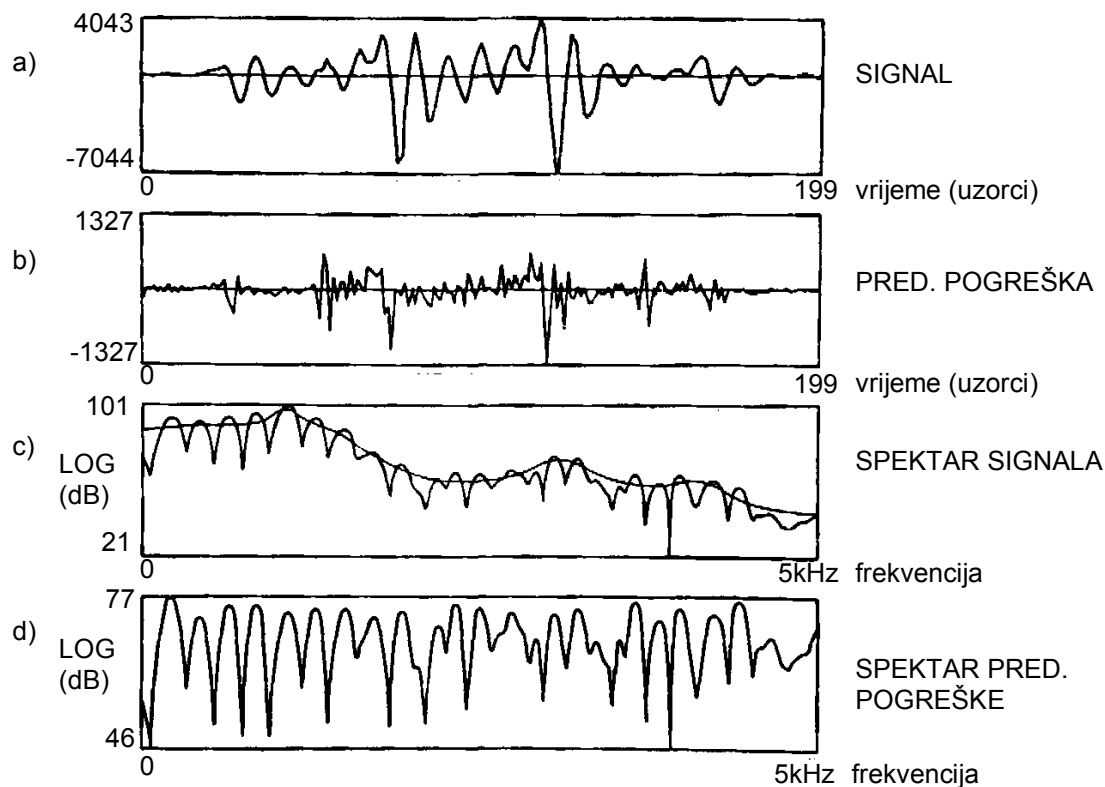
Dodatni primjeri signala predikcijske pogreške $e(n)$ prikazani su na slikama 9.10-2 do 9.10-5. Na svakoj od ovih slika (a) dio prikazuje odsječak signala koji se analizira, dio (b) prikazuje rezultirajući signal predikcijske pogreške $e(n)$, dio (c) prikazuje modul DFT-a signala iz (a) dijela slike izražen u [dB] sa superponiranim modulom frekvencijske karakteristike $H(e^{j\omega T})$ sustava $H(z)$, dok dio (d) prikazuje amplitudni spektar signala predikcijske pogreške $e(n)$, također izražen u [dB].

Na slikama 9.10-2 i 9.10-3 prikazana je analiza 20 ms samoglasnika *i*, izgovorenog od muškog govornika (LRR), metodom kovarijance odnosno autokorelacijskom metodom (s Hamming-ovim vremenskim otvorom). Na slikama se vidi da signal predikcijske pogreške ima oštar puls na početku svakog pitch perioda a pripadajući spektar je dosta ravan, ali pokazuje efekt češlja zbog periodičnosti pobudnog signala. Treba primijetiti veliku predikcijsku pogrešku na početku segmenta, na slici 9.10-3, analiziranog autokorelacijskom metodom. Do velikog iznosa pogreške dolazi zbog toga što se pokušava predvidjeti uzorke signala različitih od nule na osnovi nuliranih uzoraka lijevo od intervala $0 \leq m \leq 199$. Oblik Hammingovog vremenskog otvora s prigušenjem prema rubovima nije potpuno učinkovit pri smanjenju te pogreške.

Slike 9.10-4 i 9.10-5 prikazuju slične rezultate za 20 ms samoglasnika *a*, izgovorenog od ženske govornice. Unutar analiziranog intervala nalazi se oko pet kompletnih pitch perioda. Tako na slici 9.10-4 signal pogreške ima velik broj oštih vrhova u intervalu analize metodom kovarijance. Međutim, korištenjem Hammingovog vremenskog otvora u autokorelacijskoj metodi dolazi do smanjivanja amplitude pitch pulseva na rubovima analiziranog intervala, pa su stoga i vrhovi u signalu predikcijske pogreške smanjene amplitude prema rubovima.



Slika 9.10-4 Tipični signali i spektri za LPC metodu kovarijance, za žensku govornicu, $p = 14$, $N = 200$



Slika 9.10-5 Tipični signali i spektri za LPC autokorelacijsku metodu sa Hamming-ovim vremenskim otvorom, za žensku govornicu, $p = 14$, $N = 200$

Promatrajući ponašanje signala predikcijske pogreške prikazanog na prethodnim slikama 9.10-2 do 9.10-5, može se doći do zaključka da je signal predikcijske pogreške dobar kandidat za signal iz kojeg bi se mogao jednostavno odrediti pitch period. Na nesreću, situacija nije tako čista za druge primjere zvučnog govora. Makhoul i Wolf su pokazali da za glasove koji nisu bogati harmonicima, npr. likvide kao r, l ili nazale kao m, n, vrhovi u signalu predikcijske pogreške nisu uvijek jako oštri ili jasno vidljivi. Osim toga, na spoju između zvučnih i bezvučnih glasova oznake periodičnosti u signalu predikcijske pogreške često potpuno nestanu.

Na kraju, iako se signal predikcijske pogreške $e(n)$ čini idealnim kandidatom za detekciju pitch perioda, postoje teškoće pri lociranju glotalnih pulseva za mnoštvo zvučnih glasova, pa se stoga u ovoj primjeni ne može potpuno osloniti samo na signal predikcijske pogreške.

9.11 Izrazi za predikcijsku pogrešku

Normalizirana srednja kvadratna predikcijska pogreška za autokorelacijsku metodu je definirana kao:

$$V_n = \frac{\sum_{m=0}^{N+p-1} e_n^2(m)}{\sum_{m=0}^{N-1} s_n^2(m)} \quad (9.11-1)$$

gdje je $e_n(m)$ izlaz filtra predikcijske pogreške koji odgovara segmentu govora $s_n(m)$ pozicioniranom na vremenskom indeksu n . Za metodu kovarijance odgovarajuća definicija je:

$$V_n = \frac{\sum_{m=0}^{N-1} e_n^2(m)}{\sum_{m=0}^{N-1} s_n^2(m)} \quad (9.11-2)$$

Ako definiramo $\alpha_0 = -1$, niz predikcijske pogreške može se izraziti kao:

$$e_n(m) = -\sum_{k=0}^p \alpha_k s_n(m-k) \quad (9.11-3)$$

Uvrštenjem jednadžbe (9.11-3) u jednadžbe (9.11-1) ili (9.11-2) (ovisno o korištenoj metodi) i koristeći izraz (9.3-19), slijedi:

$$V_n = \sum_{i=0}^p \sum_{j=0}^p \alpha_i \frac{\phi_n(i,j)}{\phi_n(0,0)} \alpha_j \quad (9.11-4)$$

a uvrštenje (9.3-20) u (9.11-4) daje:

$$V_n = -\sum_{i=0}^p \alpha_i \frac{\phi_n(0,i)}{\phi_n(0,0)} \quad (9.11-5)$$

Drugi izraz za V_n dobiven je Durbinovim algoritmom, to jest:

$$V_n = \prod_{i=1}^p (1 - k_i^2) \quad (9.11-6)$$

Gornji izrazi nisu svi ekvivalentni i mogu biti tumačeni ovisno o detaljima korištene metode linearne predikcije. Na primjer, jednadžba (9.11-6) bazirana je na Durbinovom algoritmu i vrijedi samo za autokorelacijsku i lattice metodu. Nadalje, pošto lattice metoda eksplicitno ne zahtijeva računanje korelacijske funkcije, jednadžbe (9.11-4) i (9.11-5) nisu direktno primjenjive na lattice metodu. U tabeli 9.11-1 sažeti su gornji izrazi za normaliziranu srednju kvadratnu pogrešku predikcije i prikazana je valjanost izraza za pojedinu metodu.

Tablica 9.11-1 Izrazi za normaliziranu pogrešku

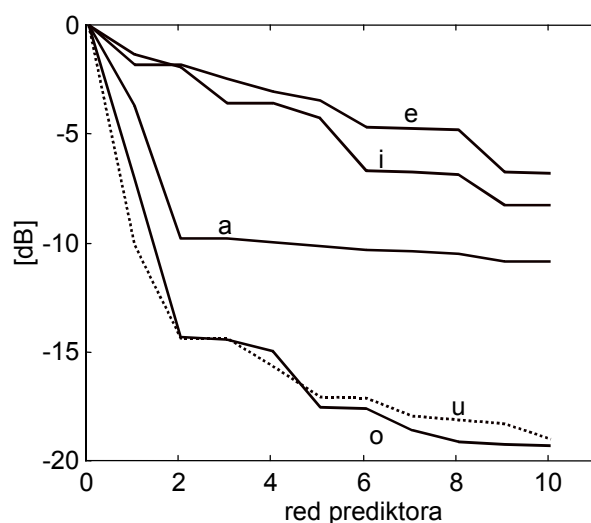
Izraz za predikcijsku pogrešku	Metoda kovarijance	Autokorelacijska metoda	Lattice metoda
$V_n = \frac{\sum_m e_n^2(m)}{\sum_m s_n^2(m)}$	vrijedi	vrijedi*	vrijedi
$V_n = \sum_i \sum_j \alpha_i \frac{\phi_n(i,j)}{\phi_n(0,0)} \alpha_j$	vrijedi	vrijedi**	ne vrijedi
$V_n = \sum_i \alpha_i \frac{\phi_n(i,j)}{\phi_n(0,0)}$	vrijedi	vrijedi**	ne vrijedi
$V_n = \prod_i (1 - k_i^2)$	ne vrijedi	vrijedi	vrijedi

* u slučaju da $s_n(m)$ predstavlja segment signala pomnožen s vremenskim otvorom

** ... u slučaju da se $\Phi(i,j)$ zamijeni sa $R(|i - j|)$

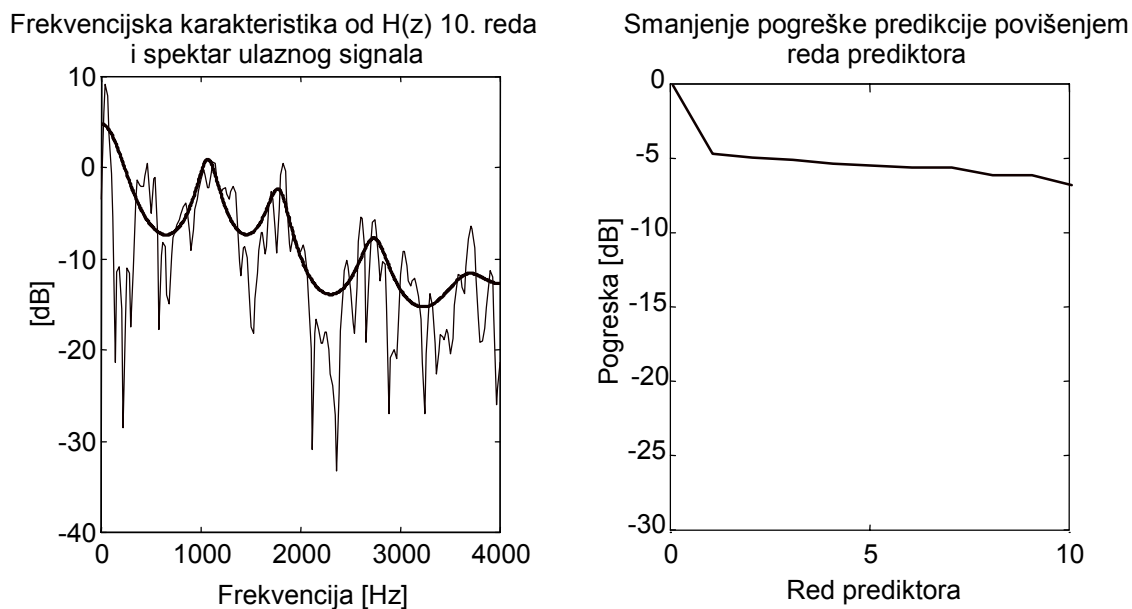
9.12 Ovisnost predikcijskog dobitka o tipu glasa

Na slici 9.12-1 prikazane su funkcije ovisnosti normalizirane pogreške predikcije, $10\log_{10}(V^{(i)})$, o redu predikcije za samoglasnike ‘a’, ‘e’, ‘i’, ‘o’, ‘u’ na kojima je vidljiva različita brzina opadanja pogreške predikcije. Krivulje za glasove ‘o’ i ‘u’ imaju znatno brži pad nego glasovi ‘a’, ‘i’, a osobito ‘e’.



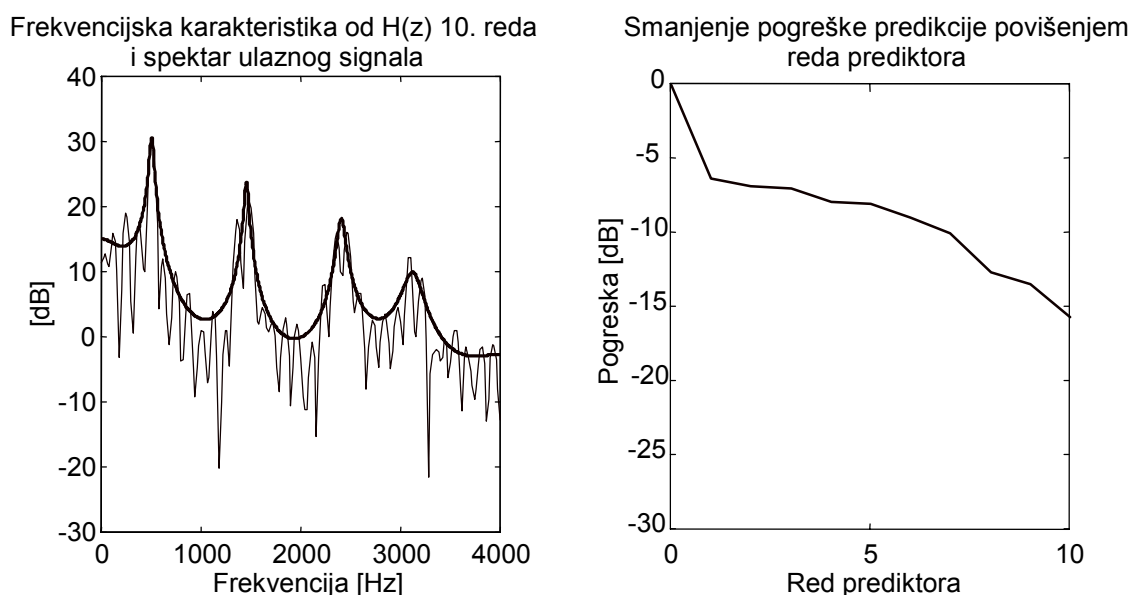
Slika 9.12-1 Normalizirana greška predikcije kao funkcija reda prediktora za samoglasnike ‘a’, ‘e’, ‘i’, ‘o’, ‘u’

Razlog ove pojave leži u činjenici da glasovi 'o' i 'u' imaju jako izražen samo jedan ili dva formanta koji se dobro modeliraju već sa prediktorom drugog ili četvrtog reda, što izrazito smanjuje grešku predikcije i sa niskim redom. S druge strane glas 'e' ima vrlo "rogati" spektar sa četiri vrlo izražena formanta kojeg nedovoljno dobro modelira čak i prediktor 10. reda.



Slika 9.12-2 Frekvencijska karakteristika prediktora i normalizirana greška predikcije kao funkcija reda prediktora za suglasnik 't'

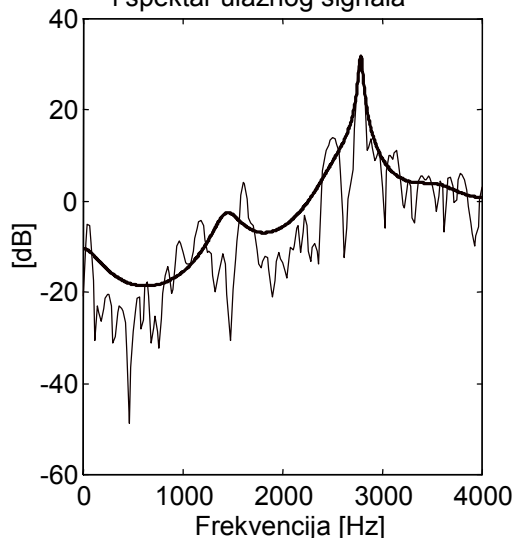
Također uočljiva je i pojava lomova unutar svake pojedine krivulje nakon kojih je opadanje pogreške predikcije znatno blaže, te nije potrebno povećanja reda prediktora za taj određeni glas.



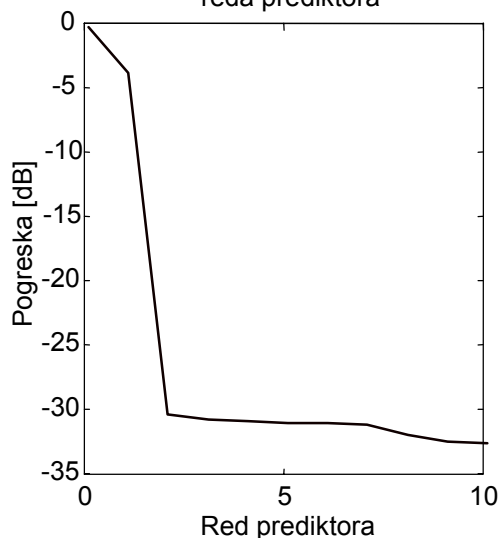
Slika 9.12-3 Frekvencijska karakteristika prediktora i normalizirana greška predikcije kao funkcija reda prediktora za suglasnik 'd'

Slike 9.12-2 do 9.12-5 prikazuju spektre govornih odsječaka i prijenosne funkcije prediktora 10. reda s lijeve strane, odnosno promjenu greške predikcije zavisno o redu prediktora s desne strane. Prve dvije slike (9.12-2 i 9.12-3) su za 'd' i 't' kao primjer zvučnog/bezvučnog para zatvornih (praskavih) glasova. Druge dvije slike (9.12-4 i 9.12-5) su za glasove 'ž' i 'š' kao primjer zvučnog/bezvučnog para tjesnačnih glasova (frikativa). Analizirajući ove slike moguće je uočiti slične pojave kao kod samoglasnika.

Frekvencijska karakteristika od H(z) 10. reda i spektar ulaznog signala

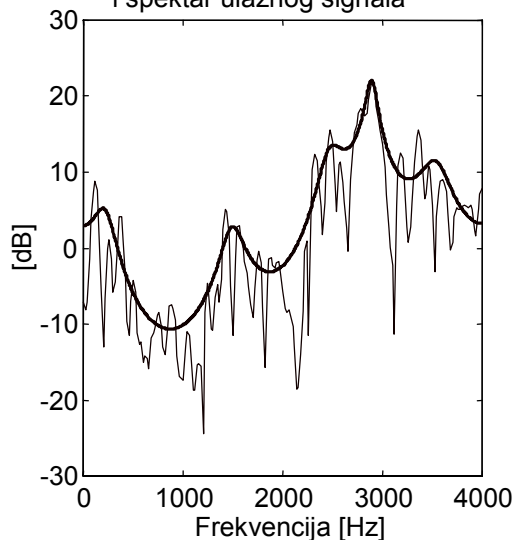


Smanjenje pogreške predikcije povišenjem reda prediktora

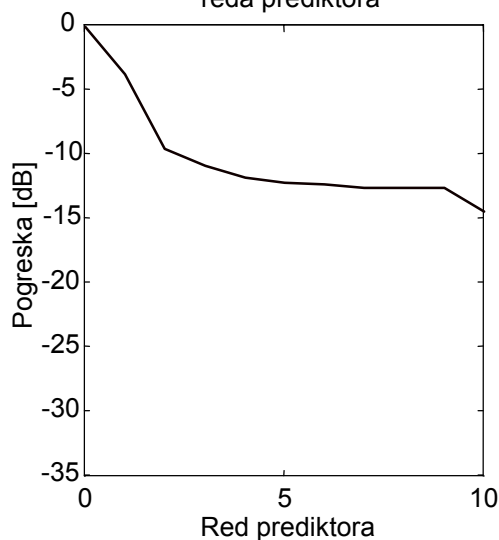


Slika 9.12-4 Frekvencijska karakteristika prediktora i normalizirana greška predikcije kao funkcija reda prediktora za suglasnik 'š'

Frekvencijska karakteristika od H(z) 10. reda i spektar ulaznog signala



Smanjenje pogreške predikcije povišenjem reda prediktora



Slika 9.12-5 Frekvencijska karakteristika prediktora i normalizirana pogreška predikcije kao funkcija reda predikt. za suglasnik 'ž'

Suglasnici 't' i 'd' imaju u spektru nekoliko podjednako izraženih formanata što se očituje u postupnom padu greške predikcije s povećanjem reda prediktora. S druge strane suglasnik 'š' ima jedan jako izražen formant na frekvenciji 2783 Hz koji se uspješno modelira već i prediktorom 2. reda, što uzrokuje nagli pad pogreške predikcije vidljiv na slici

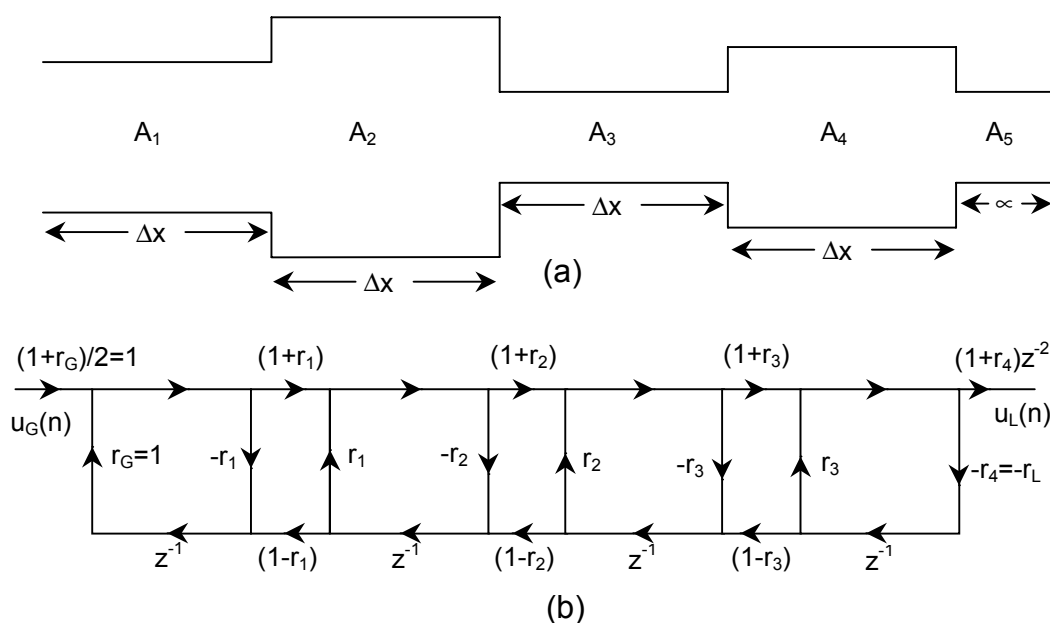
9.12-4. Općenito, može se zaključiti da dobitak odnosno kvaliteta predikcije linearnog prediktora ovisi o “obojenosti” spektra što je termin za dinamiku (razvedenost po amplitudi) amplitudno frekvencijske karakteristike pojedinog glasa, te o broju i izraženosti formanta u spektru. Što je broj formanta manji i što su oni izraženiji, to je lakše modelirati glas čak i s prediktorom niskog reda (npr. 'o', 'u', 'š'), a energija predikcijske pogreške je značajno manja od ulazne energije signala, te se ostvaruje velik dobitak predikcije. S druge strane, ako je broj formanta velik, te ako su oni podjednako izraženi, potrebno je koristiti prediktore višeg reda koji uspješno modeliraju sve formante (npr. 'e', 'd'). Općenito će za zvučne glasove, dobitak uslijed predikcije biti veći nego kod bezvučnih glasova, kao što je vidljivo na paru 'd' i 't'.

9.13 Alternativni skupovi koeficijenata za definiranje prediktora

9.13.1 Odnos parametara linearnog prediktora i modela s cijevima bez gubitaka

U poglavlju 6 i 7 razmatran je model formiranja govora kojeg čini niz od N akustičkih cijevi bez gubitaka, shematski prikazanih na slici 9.13-1 a). Ekvivalentni vremenski diskretni model prikazan je na slici 9.13-1 b), a koeficijenti refleksije r_k definirani su kao omjer površina poprečnih presjeka dvije susjedne cijevi bez gubitaka tj. prema izrazu:

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k} \quad (9.13-1)$$



Slika 9.13-1 Model s 4 cijevi bez gubitaka koji završava beskonačno dugom cijevi; (a) i pripadajući graf toka signala uz beskonačnu impedanciju glasnica, (b)

Prijenosna funkcija vremenski diskretnog modela izvedena je u poglavlju 7.2 i dana je u konačnim izrazima (7.2-9) do (7.2-11). Sustav sa N spojenih cijevi je u općenitom slučaju opisan sa $N-1$ faktora refleksije r_1 do r_{N-1} na spojevima svake dvije cijevi, kao i zaključnim faktorima refleksije na glasnica r_G i usnicama r_L . Ako se radi jednostavnosti pretpostavi da

je koeficijent refleksije na glasnicama $r_G=1$, tj. da je impedancija glasnica beskonačna, te ako se koeficijent refleksije na usnicama r_L označi sa r_N , tada se prijenosna funkcija vremenski diskretnog modela (7.2-9) svodi na slijedeći oblik:

$$V(z) = \frac{\prod_{k=1}^N (1+r_k)z^{-N/2}}{D(z)} \quad (9.13-2)$$

gdje $D(z)$ zadovoljava slijedeću polinomialnu rekurziju:

$$D_0(z) = 1 \quad (9.13-3)$$

$$D_k(z) = D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}) \quad (9.13-4)$$

$$D(z) = D_N(z) \quad (9.13-5)$$

Ovu rekurziju je lako dokazati analizom izraza (7.2-10) za specijalni slučaj $r_G=1$. Svi ovi izrazi dosta podsjećaju na razmatranje mrežastih struktura u poglavlju 9.8, gdje je pokazano da polinom koji definira prijenosnu funkciju inverznog filtra:

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k} \quad (9.13-6)$$

a koji je dobiven analizom pomoću linearne predikcije, može biti određen sljedećom rekurzijom:

$$A^{(0)}(z) = 1 \quad (9.13-7)$$

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}) \quad (9.13-8)$$

$$A(z) = A^{(p)}(z) \quad (9.13-9)$$

gdje se parametri $\{k_i\}$ nazivaju PARCOR koeficijentima. Usporedbom izraza (9.13-3) do (9.13-5) sa izrazima (9.13-7) do (9.13-9) proizlazi da prijenosna funkcija sustava:

$$H(z) = \frac{G}{A(z)} \quad (9.13-10)$$

dobivena analizom pomoću linearne predikcije ima isti oblik kao i prijenosna funkcija sustava modela s cijevima bez gubitaka koji se sastoji od p spojenih cijevi. Uz pretpostavku da je:

$$r_i = -k_i \quad (9.13-11)$$

tada je očito da će polinomi $D(z)$ i $A(z)$ biti identični, tj.:

$$D(z) = A(z) \quad (9.13-12)$$

Koristeći izraze (9.13-1) i (9.13-11) može se lako pokazati da odnos između površina poprečnih presjeka ekvivalentnog modela sa cijevima bez gubitaka i PARCOR koeficijenata glasi:

$$A_{i+1} = \left(\frac{1-k_i}{1+k_i} \right) A_i \quad (9.13-13)$$

Može se uočiti da PARCOR koeficijenti definiraju samo omjer između površina presjeka dviju susjednih cijevi. Na ovaj način površine modela s cijevima bez gubitaka nisu apsolutno određene, jer će svaka normalizacija (množenje svih površina s istim faktorom) dati novi model s istom prijenosnom funkcijom. Uz definiciju apsolutnog iznosa površina presjeka bilo kojeg segmenta (npr. zadnjeg), sve se ostale mogu odrediti korištenjem koeficijenata k_i i izraza (9.13-13).

Treba naglasiti da funkcija poprečnog presjeka dobivena korištenjem izraza (9.13-13) ne predstavlja pravu funkciju površine presjeka ljudskog vokalnog trakta. Međutim, pokazano je u literaturi da ako se prije analize linearnom predikcijom provede prednaglašavanje govornog signala (visoko-propusna filtracija), da će tada utjecaji spektra glotalnog pulsa i utjecaji zračenja na usnicama biti otklonjeni, pa će tako dobivene funkcije površina biti vrlo slične stvarnom obliku vokalnog trakta prilikom izgovora.

9.13.2 Odnos koeficijenata linear nog prediktora i PARCOR koeficijenata

Postupak za određivanje koeficijenata prediktora $\{\alpha_j^{(p)}, j=1,2,\dots,p\}$ na osnovu PARCOR koeficijenata $\{k_i, i=1,2,\dots,p\}$ već je ustvari objašnjen u okviru Durbin-ovog algoritma za određivanje rješenja autokorelacijske metode. Koeficijenti prediktora mogu se dobiti iz PARCOR koeficijenata korištenjem slijedeće rekurzije:

$$\alpha_i^{(i)} = k_i \quad (9.13-14)$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad 1 \leq j \leq i-1 \quad (9.13-15)$$

Ako se izrazi (9.13-14) i (9.13-15) izračunavaju za $i=1, \dots, p$, tada koeficijenti prediktora α_1 do α_p slijede iz zadnjeg koraka rekurzije prema izrazu:

$$\alpha_j = \alpha_j^{(p)} \quad 1 \leq j \leq p \quad (9.13-16)$$

Slično tome, skup PARCOR koeficijenata može se odrediti iz skupa LPC koeficijenata prediktora koristeći obrnutu rekurziju oblika:

$$k_i = \alpha_i^{(i)} \quad (9.13-17)$$

$$\alpha_j^{(i-1)} = \frac{\alpha_j^{(i)} + \alpha_i^{(i)} \alpha_{i-j}^{(i)}}{1 - k_i^2} \quad 1 \leq j \leq i-1 \quad (9.13-18)$$

Za razliku od prethodne rekurzije, varijabla i u ovom slučaju ide od p prema dolje do 1, a kao inicijalna vrijednost postavlja se:

$$\alpha_j^{(p)} = \alpha_j \quad 1 \leq j \leq p \quad (9.13-19)$$

9.13.3 LAR koeficijenti

Postoji veći broj jedan-na-jedan transformacija parametara prediktora u neki novi skup parametara, koji za određene primjene imaju bolja svojstva nego originalni parametri. Jedan takav skup ekvivalentnih parametara, koji se mogu odrediti iz PARCOR koeficijenata su LAR parametri (engl. *Log Area Ratio coefficients*) koji su definirani slijedećim izrazom:

$$g_i = \log\left(\frac{A_{i+1}}{A_i}\right) = \log\left(\frac{1-k_i}{1+k_i}\right), \quad 1 \leq i \leq p \quad (9.13-20)$$

Parametri g_i su jednaki logaritmu omjera površina poprečnih presjeka susjednih segmenata kod modela s cijevima bez gubitaka, gdje taj model ima identičnu prijenosnu funkciju kao i LPC model opisan koeficijentima k_i . Utvrđeno je da su g_i parametri vrlo pogodni za kvantizaciju, jer imaju relativno ravnu karakteristiku spektralne osjetljivosti. To znači da će neovisno o apsolutnom iznosu pojedinog parametra, promjena njegovog iznosa uslijed kvantizacije prouzročiti podjednaku spektralnu pogrešku, tj. razliku između originalne i kvantizirane prijenosne funkcije modela.

Iz g_i parametra je moguće inverznom transformacijom odrediti k_i koeficijente korištenjem slijedećeg izraza:

$$k_i = \frac{1-e^{g_i}}{1+e^{g_i}}, \quad 1 \leq i \leq p \quad (9.13-21)$$

9.14 Kvantizacija parametara prediktora

Jedna od najznačajnijih primjena linearne predikcije jest područje kodiranja govora za potrebe digitalnih komunikacijskih sustava s malom brzinom prijenosa (npr. LPC vokoder), ili pak za potrebe digitalne pohrane glasa. Slika 9.14-1 prikazuje blok dijagram LPC vokodera. Vokoder čine tri osnovna dijela:

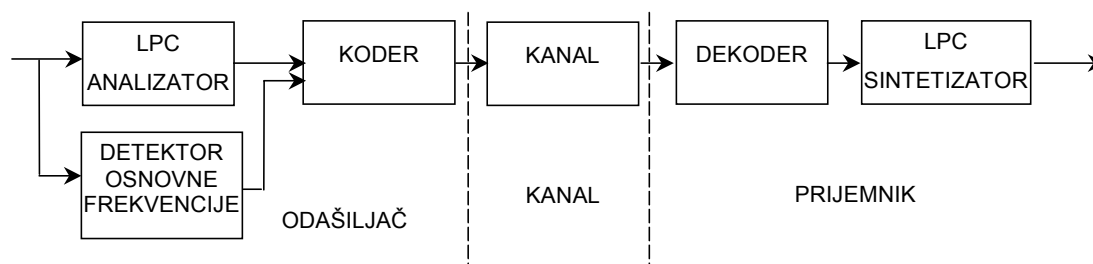
- 1) odašiljač ili predajnik koji provodi LPC analizu, određuje karakter govornog signala (zvučan ili bezvučan), određuje period osnovne frekvencije titranja glasnica (pitch-period) za zvučne glasove i konačno kodira sve određene parametre za potrebe prijenosa,
- 2) kanal kojim se parametri šalju,
- 3) prijemnik koji dekodira parametre i iz njih obrnutim postupcima sintetizira govor.

Postupci analize i sinteze su već izloženi u osnovnim crtama u prošlim poglavljima no o postupcima kvantizacije i kodiranja parametara još nije bilo riječi. Upravo u ovom poglavlju analizirati će se osnovni postupci kodiranja i dekodiranja, te utvrditi koja skupina parametara za opis vokalnog trakta je najpogodnija za kodiranje ako se želi ostvariti što veća kvaliteta uz što manju brzinu prijenosa. Radi pojednostavljenja, koristi se pretpostavka da je digitalni kanal kojim se vrši prijenos idealan (ne unosi pogreške).

Osnovni parametri Vokodera koje je potrebno odrediti u postupku LPC analize, te zatim kodirati i prenijeti na prijemnu stranu su:

- 1) skup od p LPC koeficijenata, odnosno koeficijenata prediktora $P(z)$,
- 2) period osnovne frekvencije titranja glasnica (pitch-period),
- 3) binarna informacija o zvučnosti (zvučni ili bezvučni glas),
- 4) i parametar pojačanja (informacija o energiji pobudnog signala).

Postupci pravilnog kodiranja pitch-perioda, binarne informacije o zvučnosti glasa, kao i parametra pojačanja su relativno jednostavni. Za kodiranje pitch-perioda uglavnom je dovoljno 6 bita (odabir jednog od 64 moguća perioda), za odluku o zvučnosti 1 bit, a za pojačanje 5 bita raspodijeljenih na logaritamskoj skali (jedan od 32 moguća nivoa).



Slika 9.14-1 Blok dijagram LPC vokodera

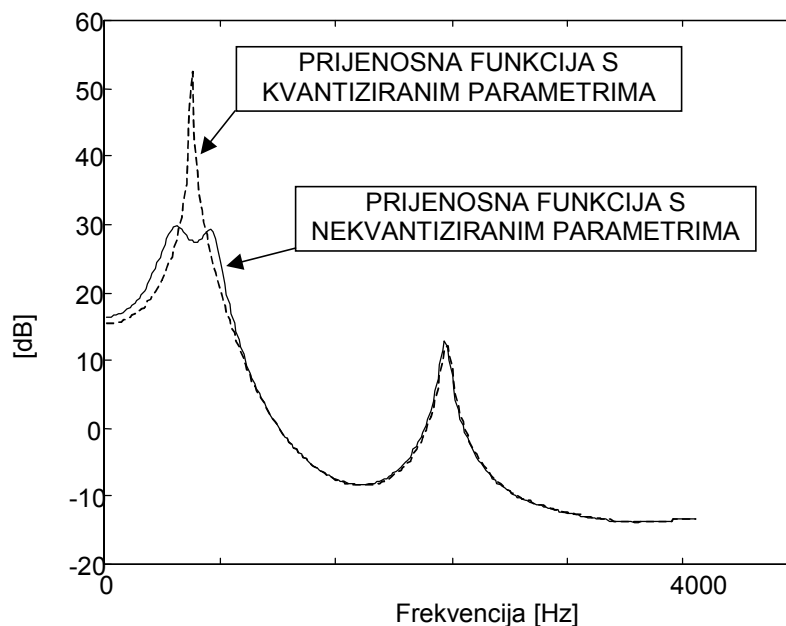
Što se tiče kvantizacije parametara LPC prediktora, stvar je složenija. Iako prva ideja koja se sama po sebi nameće je direktna kvantizacija koeficijenata prediktora, takav pristup se ne preporučuje. Uslijed kvantizacije koeficijenata dolazi do promjene prijenosne funkcije prediktora, odnosno promjene pozicije polova LPC filtra $H(z)$. Ta promjena može uzrokovati izlazak pojedinih polova van jedinične kružnice, tj. nestabilnost LPC filtra. Sigurna stabilnost LPC filtra zahtjeva relativno visoku točnost kvantizacije (8-10 bita po LPC koeficijentu). Razlog ovome jest u činjenici da male promjene koeficijenata prediktora vode u relativno velike promjene položaja polova, pa se direktna kvantizacija koeficijenata prediktora općenito izbjegava.

Navedeni zaključci biti će ilustrirani na jednom malom primjeru. Idealni α_k koeficijenti dobiveni LPC analizom govornog signala prikazani su u prvom retku tablice 9.14-1. za jedan konkretni slučaj. Prvo je provedena normalizacija ovih koeficijenata, tj. dijeljenje s koeficijentom najvećim po apsolutnoj vrijednosti. Zatim je provedena kvantizacija sa $B=6$ bitnim uniformnim kvantizatorom, tako što su normalizirani koeficijenti pomnoženi sa 2^{B-1} i zatim zaokruženi na najbliže cijele brojeve. Pretpostavljajući idealni prijenosni kanal na prijamnoj se strani vrši dekodiranje pristigle poruke, dijeleći cjelobrojne koeficijente sa 2^{B-1} , te množeći ih s faktorom skale koji je korišten kod normalizacije na predajnoj strani. Rezultati svakog pojedinog koraka opisanog postupka prikazani su u tablici 9.14-1.

Tablica 9.14-1 Ilustracija direktne kvantizacije α -koeficijenata s 6-bitnim uniformnim kvantizatorom

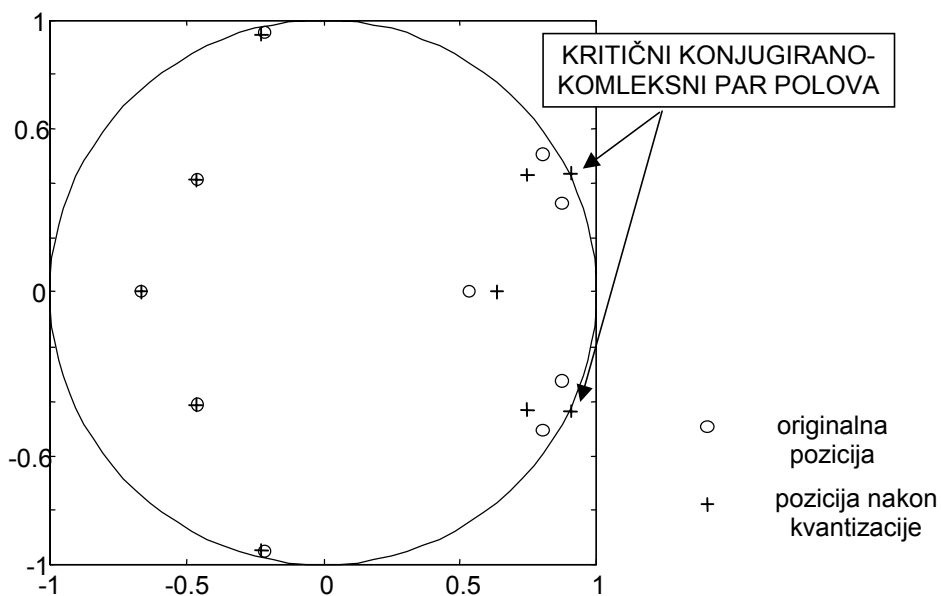
	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}
NEKVANT. KOEFICIJENTI	1.899	-1.198	0.619	-0.744	-0.203	0.715	0.073	-0.286	-0.130	0.104
NORMALIZIR. KOEFICIJENTI	1.000	-0.630	0.325	-0.392	-0.107	0.376	0.038	-0.150	-0.068	0.054
ZAOKRUŽENI	32	-20	10	-13	-3	12	1	-5	-2	2
KOEFICIJENTI NA PRIJEMU	1.000	-0.625	0.312	-0.406	-0.093	0.375	0.031	-0.156	-0.062	0.062
KRAJNI REZULTAT KODIRANJA	1.899	-1.187	0.593	-0.771	-0.178	0.712	0.059	-0.296	-0.118	0.118

Uspoređujući na kraju prijenosne funkcije LPC filtra $H(z)$ prikazane na slici 9.14-2, koje odgovaraju nekvantiziranim, odnosno kvantiziranim koeficijentima prediktora uočavaju se velika odstupanja.



Slika 9.14-2 Prikaz odstupanja prijenosne funkcije LPC filtra uzrokovanog kvantizacijom koeficijenata prediktora

Objašnjenje ove velike razlike između idealne i kvantizirane karakteristike najbolje daje slika 9.14-3 na kojoj su prikazani položaji polova LPC filtra prije i poslije kvantizacije α_k koeficijenata. Vidljivo je da jedan konjugirano-kompleksni par polova koji odgovara kvantiziranim koeficijentima čak izlazi iz jedinične kružnice, što potpuno potvrđuje gornje zaključke.



Slika 9.14-3 Utjecaj kvantizacije na poziciju polova LPC filtra

Ovaj zaključak otvara pitanje koji su parametri najpogodniji za kvantizaciju i prijenos. Najprirodniji sljedeći kandidati su korijeni polinoma prediktora i koeficijenti refleksije. U slučaju kvantizacije korijena prediktora, pozicija svakog konjugirano kompleksnog para polova kodira se u polarnim koordinatama, tj. tako da se posebno kodira kut pola, a posebno njegov radijus. Kut pola određuje centralnu frekvenciju formanta, dok njegov radijus određuje širinu pojasa. Stabilnost je vrlo lako osigurati, tako da se prilikom kvantizacije

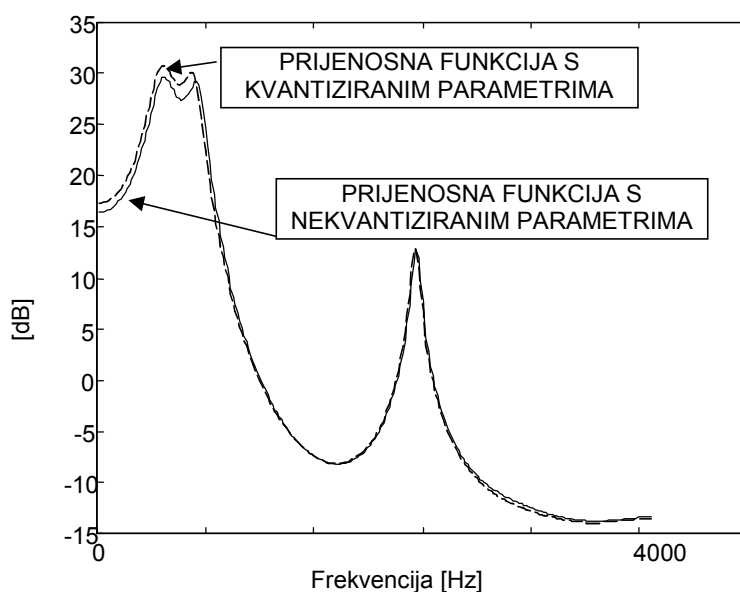
osigura da radijus pola nikada ne dosegne jedinicu. Koristeći opisani pristup u literaturi je pokazano da je 5 bita po korijenu (tj. 5 bita za centralnu frekvenciju pola i 5 bita za njegovu širinu pojasa) dovoljno da sačuva kvalitetu sintetiziranog govora gotovo jednaku govoru sintetiziranom korištenjem idealnih, nekvantiziranih parametara. Glavni nedostatak korištenja ovog skupa parametara je prevelika složenost njihovog izračunavanja. Obzirom da ne postoji analitički izraz za određivanje korijena polinoma stupnja većeg od 4, potrebno je primjenjivati složene numeričke postupke određivanja korijena.

Uz opisane postupke kodiranja i uz 12. red LPC analize ukupni broj bita za kodiranje svih parametara jednog okvira analize iznosi $(12 \times 5 + 6 + 5 + 1) = 72$ bita po okviru. Prema, tome ukupna potrebna brzina prijenosa iznosi $72F_r$ bita u sekundi gdje je F_r učestalost analize koja se izražava u broju okvira u sekundi. Tipične vrijednosti za F_r su 100, 67, 50 i 33 okvira/s što daje brzine prijenosa od 7200, 4800, 3600 i 2400 bita u sekundi.

Sljedeća skupina parametara koja pokazuje dobra kvantizacijska svojstva i kod koje je lako moguće osigurati stabilnost LPC filtra su PARCOR koeficijenti, k_i . Uvjet stabilnosti za ove parametre je $|k_i| < 1$, što je jednostavno očuvati i nakon kvantizacije. Primjena ovih parametara ilustrirana je na istom primjeru danom u tablici 9.14-1. U prvom koraku postupka su na osnovu idealnih koeficijenata prediktora α_1 do α_{10} izračunati idealni PARCOR koeficijenti k_1 do k_{10} . U ovom slučaju normalizaciju nije potrebno provoditi jer su svi koeficijenti po modulu već manji od 1, ($|k_i| < 1$).

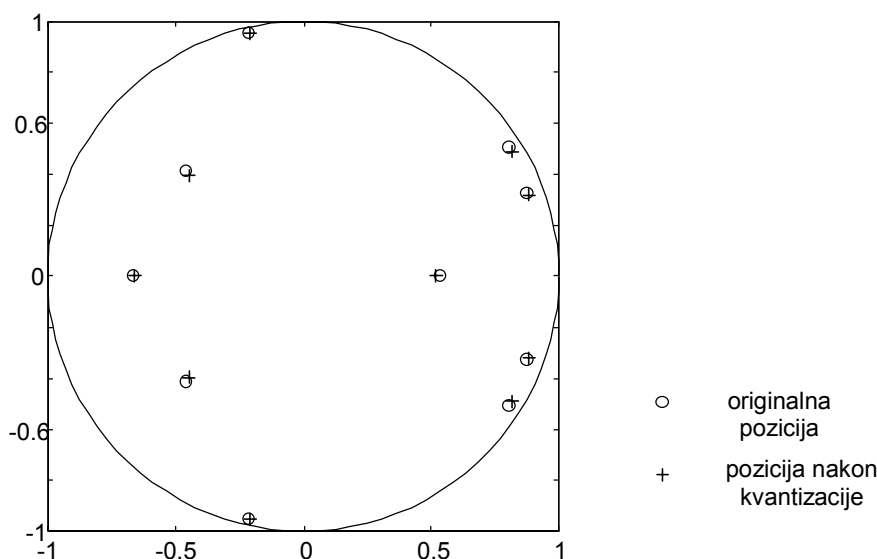
Tablica 9.14-2 Prikaz dobivenih rezultata prilikom kvantizacije k-parametara 6-bitnim uniformnim kvantizatorom

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}
NEKVANT. KOEFICIJENTI	-0.898	0.8954	-0.059	0.431	-0.612	-0.082	0.530	0.287	-0.067	-0.104
ZAOKRUŽENI	-29	29	-2	14	-20	-3	17	9	-2	-3
KRAJNI REZULTAT KODIRANJA	-0.906	0.906	-0.062	0.437	-0.625	-0.093	0.531	0.281	-0.062	-0.093



Slika 9.14-4 Prikaz prijenosne funkcije LPC filtra prije i nakon kvantizacije PARCOR koeficijenata

Provodeći identičan postupak kvantizacije sa 6-bitnim uniformnim kvantizatorom dobivena je tablica 9.14-2 i slike 9.14-4 i 9.14-5 na kojima se vide neznatna odstupanja polova uslijed provedene kvantizacije, a time i dobro poklapanje prijenosnih funkcija.



Slika 9.14-5 Prikaz polova prijenosne funkcije LPC filtra prije i nakon kvantizacije PARCOR koeficijenata

U literaturi je pokazano da je statistička razdioba ovih PARCOR koeficijenata nejednolika, tj. da su neke vrijednosti češće (vjerojatnije) od drugih. Ovo predstavlja problem kod projektiranja kvantizatora, kojeg je moguće riješiti korištenjem dodatne nelinearne transformacije. Ako se želi postići minimalna spektralna osjetljivost na promjene parametara uslijed kvantizacije, tada se pokazalo da je optimalna transformacija k -parametara sljedećeg oblika:

$$g_i = \log\left(\frac{A_{i+1}}{A_i}\right) = \log\left(\frac{1-k_i}{1+k_i}\right), \quad 1 \leq i \leq p \quad (9.14-1)$$

Dakle, optimalni parametri za kvantizaciju s uniformnim kvantizatorom, koji ima jednolik razmak između kvantizacijskih nivoa su LAR parametri. Kao što je već opisano u poglavlju 9.13.3, LAR parametri su jednaki logaritmu omjera poprečnih presjeka A_i dva susjedna segmenata modela sa cijevima bez gubitaka koji modelira vokalni trakt. Promatrajući izraz (9.14-1), lako je pokazati da se područje $-1 \leq k_i \leq 1$ ovom nelinearnom transformacijom preslikava u $-\infty \leq g_i \leq \infty$. Utvrđeno je da g_i koeficijenti imaju približno jednoliku razdiobu, i malu među-parametarsku korelaciju, što ih čini vrlo pogodnim za kvantizaciju i digitalni prijenos. Kvantizacijom LAR parametara uz 5 do 6 bita po parametru, postiže se kvaliteta signala takva da ga više nije moguće razlučiti od signala koji nastaje sintezom sa idealnim (nekvantiziranim) parametrima.

Bez obzira koji se parametri kvantizirali, jasno je da se događa bespovratni gubitak dijela informacije. Točnost koju je potrebno ostvariti prilikom kvantizacije određena je s perceptualnim zahtjevima, tj. utjecaj kvantizacije ne smije biti čujan. Naravno, kod projektiranja koda i njihove međusobne usporedbe značajno je ostvariti mogućnost objektivnog, a ne subjektivnog vrednovanja. U tu svrhu najčešće se koristi numerički parametar spektralnog izobličenja koji opisuje koliko se dobro poklapaju prijenosne funkcije dva filtra: onog sa idealnim koeficijentima i onog sa kvantiziranim. U stvari se radi o

efektivnoj vrijednosti razlike između amplitudno frekvencijskih karakteristika ta dva filtra, $|H(e^{j\omega})|$ i $|\hat{H}(e^{j\omega})|$, u logaritamskom mjerilu, tj. prema izrazu:

$$SD = \sqrt{\frac{1}{2\pi} \int_0^{2\pi} \left[20 \log_{10} |H(e^{j\omega})| - 20 \log_{10} |\hat{H}(e^{j\omega})| \right]^2 d\omega}, \quad [\text{dB}] \quad (9.14-2)$$

Vidljivo je da se usrednjavanje provodi preko cijelog frekvencijskog područja (0 do 2π). Ako se postupkom kvantizacije osigura da prosječno spektralno izobličenje ne bude veće od 1 dB, tada ljudskim uhom nije moguće razlučiti razliku između govora koji nastaje sintezom s originalnim parametrima, i onoga koji se sintetizira na osnovu kvantiziranih parametara. Takav slučaj se u literaturi naziva 'transparentnim kodiranjem'. Postoji još jedan dodatni uvjet za ostvarenje transparentnog kodiranja. Ljudsko uho je posebno osjetljivo na povremena velika izobličenja, tj. čak i kada je srednje spektralno izobličenje manje od 1 dB, učestala pojava okvira s izobličenjem većim od 2 ili 4 dB može biti posebno čujna i ometajuća. Radi toga se dodatno zahtijeva da postotak okvira analize kod kojih je spektralno izobličenje prouzročeno kvantizacijom veće od 2dB bude manji od 2%, dok onih sa izobličenjem većim od 4dB uopće ne smije biti.

U svim gore navedenim postupcima kodiranja, parametri se kodiraju pomoću PCM postupaka kodiranja, tj. svaki za sebe nezavisno i to tako da se nekom intervalu ulazne realne veličine pridijeli neki kod. U slučaju korištenja ADPCM tehnika, gdje bi se ti parametri kodirali zavisno jedan o drugom, mogle bi se postići dodatne uštede u brzini prijenosa.

9.15 Frekvencijska interpretacija greške predikcije

Do ovog poglavlja, metode linearne predikcije su uglavnom bile prezentirane pomoću jednadžbi diferencija i pomoću korelacijskih funkcija, dakle njihove reprezentacije u vremenskoj domeni. Dakako na samom početku diskusija vezanih uz postupke linearne predikcije naglašeno je da koeficijenti linearnog prediktora predstavljaju koeficijente nazivnika funkcije sustava kojom se modeliraju zajednički efekti frekvencijske karakteristike vokalnog trakta, oblika pobudnog pulsa vokalnog trakta (engl. *glottal pulse*) i zračenja na usnicama. Dakle, uz poznati skup prediktorskih koeficijenata moguće je odrediti frekvencijsku karakteristiku modela govornog sustava jednostavnom evaluacijom prijenosne funkcije $H(z)$ za $z = e^{j\omega}$, tj. za z na jediničnoj kružnici:

$$H(e^{j\omega}) = \frac{G}{1 - \sum_{k=1}^p \alpha_k e^{-j\omega k}} = \frac{G}{A(e^{j\omega})} \quad (9.15-1)$$

U amplitudno frekvencijskoj karakteristici sustava, $|H(e^{j\omega})|$ moguće je očekivati izdignuća (rogove) na formantnim frekvencijama, baš kao što bi se i dobilo vremenski kratkotrajnom Fourier-ovom analizom promatranog odsječka. Dakle linearna prediktivna analiza se može promatrati i kao metoda određivanja kratkotrajnog spektra. Uistinu takve tehnike su široko korištene izvan područja obrade govora samo za ovu ulogu. U narednom poglavlju biti će predstavljena interpretacija srednje kvadratne pogreške predikcije u frekvencijskoj domeni.

9.15.1 Interpretacija srednje kvadratne pogreške predikcije u frekvencijskoj domeni

Ako se određivanje koeficijenata prediktora provede autokorelacijskim LPC postupkom, tada se srednja kvadratna pogreška predikcije može izraziti u vremenskoj domeni kao:

$$E_n = \sum_{m=0}^{N+p-1} e_n^2(m) \quad (9.15-2)$$

ili u frekvencijskoj domeni (korištenjem Parsevalovog teorema) kao:

$$E_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} |S_n(e^{j\omega})|^2 |A(e^{j\omega})|^2 d\omega \quad (9.15-3)$$

gdje je $S_n(e^{j\omega})$ Fourier-ova transformacija segmenta govora $s_n(m)$, a $A(e^{j\omega})$ je frekvencijska karakteristika inverznog filtra $A(z)$:

$$A(e^{j\omega}) = 1 - \sum_{k=1}^p \alpha_k e^{-j\omega k} \quad (9.15-4)$$

Obzirom da je frekvencijska karakteristika modela govornog sustava dana izrazom:

$$H(e^{j\omega}) = \frac{G}{A(e^{j\omega})} \quad (9.15-5)$$

tada se jednadžba (9.15-3) može izraziti i kao:

$$E_n = \frac{G^2}{2\pi} \int_{-\pi}^{\pi} \frac{|S_n(e^{j\omega})|^2}{|H(e^{j\omega})|^2} d\omega \quad (9.15-6)$$

Obzirom da je podintegralna funkcija u jednadžbi (9.15-6) pozitivna, slijedi da je minimizacija E_n ekvivalentna minimizaciji integrala odnosa spektra snage segmenta govora i kvadrata amplitudno frekvencijske karakteristike linearnog sustava, tj. modela govornog trakta.

U poglavlju 9.4 pokazano je da su autokorelacijska funkcija $R_n(m)$, segmenta govora $s_n(m)$ i autokorelacijska funkcija $R(m)$ impulsnog odziva $h(m)$ pripadnog LPC sustava $H(z)$, jednake u prвих $(p+1)$ vrijednosti, tj. za $m=0,1,2,\dots,p$. Dakle, kad $p \rightarrow \infty$ autokorelacijske funkcije su identične za sve vrijednosti indeksa m . Obzirom da je Fourierova transformacija autokorelacije $R_n(m)$ jednaka spektru snage $|S_n(e^{j\omega})|^2$ segmenta govornog signala $s_n(m)$ i obzirom da se Fourier-ovom transformacijom autokorelacije impulsnog odziva sustava $h(m)$ dobiva kvadrat amplitudno frekvencijske karakteristike sustava $|H(e^{j\omega})|^2$, ekvivalencija u vremenskoj domeni se može zapisati i u frekvencijskoj domeni kao:

$$\lim_{p \rightarrow \infty} |H(e^{j\omega})|^2 = |S_n(e^{j\omega})|^2 \quad (9.15-7)$$

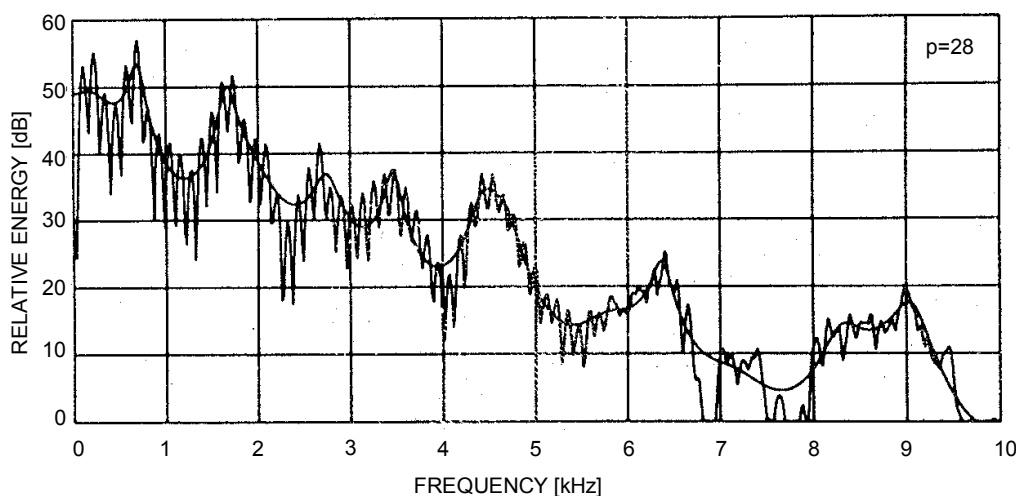
To znači da je uz dovoljno velik red LPC modela p , moguće aproksimirati spektar signala s filtrom bez nula $H(z)$ i to uz proizvoljno malu pogrešku modeliranja.

Prema jednadžbi (9.15-7) je vidljivo da je uz dovoljno visok stupanj prediktora p , modul spektra govornog signala jednak modulu amplitudno frekvencijske karakteristike sustava $H(z)$, tj. $|H(e^{j\omega})| = |S_n(e^{j\omega})|$. Međutim ova ekvivalencija ne povlači nužno za sobom

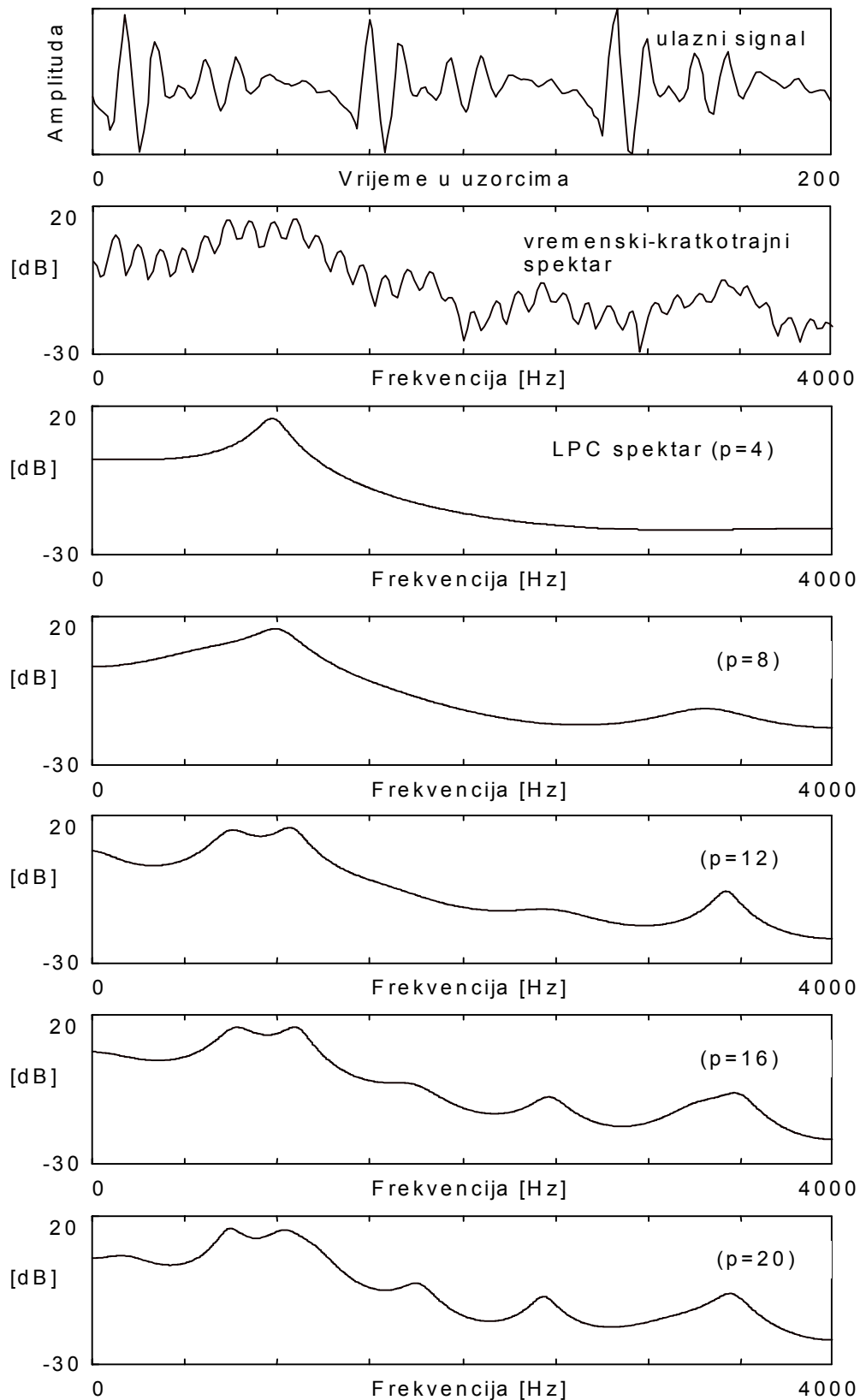
identičnost frekvencijskih karakteristika $H(e^{j\omega})=S_n(e^{j\omega})$, tj. LPC modelom se modelira isključivo modul spektra signala, ali ne i njegova faza. Obzirom da je sustav $H(z)$ stabilan sa svim polovima unutar jedinične kružnice i obzirom da $H(z)$ nema nule, $H(e^{j\omega})$ je prijenosna funkcija minimalne faze, što općenito ne mora vrijediti za $S_n(e^{j\omega})$.

U svrhu ilustracije mogućnosti spektralnog modeliranja linearnom predikcijom, slika 9.15-1 prikazuje usporedbu između spektra signala i njegovog LPC modela u dB mjerilu, tj. između $10\log_{10}(|H(e^{j\omega})|^2)$ i $10\log_{10}(|S_n(e^{j\omega})|^2)$. Spektar signala je dobiven FFT analizom segmenta govora trajanja 20 ms (otipkanog sa 20kHz), pomnoženog Hamming-ovim otvorom. Promatrani segment dobiven je na osnovu glasa za samoglasnik /ae/. Segment je obrađen autokorelacijskim LPC postupkom sa redom prediktora $p=28$, a frekvencijska karakteristika LPC filtra $H(z)$ prikazana je na slici 9.15-1 zajedno s spektrom signala. Harmonijska (periodička) struktura spektra signala je jasno vidljiva na slici. Na slici je vidljivo i jedno vrlo važno svojstvo spektralnog modeliranja primjenom LPC postupka. Riječ je o tome da se frekvencijska karakteristika LPC modela mnogo bolje poklapa sa spektrom signala u područjima sa velikom energijom signala (oko vrhova spektra) nego u područjima sa niskom energijom signala (oko udolina u spektru signala). Opisano svojstvo prepoznatljivo je i u jednadžbi (9.15-6) jer područja gdje $|H(e^{j\omega})| < |S_n(e^{j\omega})|$ više pridonose ukupnoj pogrešci nego područja u kojima $|H(e^{j\omega})| > |S_n(e^{j\omega})|$. Dakle kriterij pogreške LPC spektra favorizira dobro poklapanje u blizini vrhova spektra, dok poklapanje u spektralnim udolinama nije ni približno tako dobro.

Gornje razmatranje pokazuje da se redom linearne prediktivne analize p , može kontrolirati stupanj glatkoće rezultirajućeg spektra LPC modela. To je prikazano na slici 9.15-2 koja prikazuje ulazni segment govora, Fourier-ovu transformaciju tog segmenta i amplitudno frekvencijsku karakteristiku LPC modela za različite redove prediktora. Jasno je da kako p raste, sve više spektralnih detalja ostaje sačuvano. Prilikom analize govora LPC postupkom, LPC filter svojom frekvencijskom karakteristikom mora reprezentirati samo produkt spektra glotalnog pulsa, frekvencijske karakteristike vokalnog trakta i zračenja na usnicama, ali ne i spektar pobudnog signala. Prema tome, kao što je već diskutirano, red LPC filtra mora biti tako odabran da formantna struktura i opći oblik spektra budu što vjernije reprezentirani, dok spektralna struktura pobudnog signala mora preostati u rezidualnom signalu predikcije.



Slika 9.15-1 LPC spektar 28-polnog prediktora u usporedbi sa spektrom polaznog segmenta govornog signala



Slika 9.15-2 Spektar za samoglasnik /a/ otipkan uz $f_s=8\text{kHz}$ i frekvencijske karakteristike LPC modela za nekoliko različitih redova prediktora p

Treba naglasiti da je u okviru dosadašnjih razmatranja pretpostavljeno da se izračunavanje koeficijenata prediktora provodi autokorelacijskim LPC postupkom, jer je samo u tom slučaju Fourierova transformacija vremenski kratkotrajne autokorelacijske funkcije signala jednaka kvadratu amplitude vremenski kratkotrajne Fourierove transformacije promatranog segmenta govornog signala. Međutim, to ne sprečava da se $H(e^{j\omega})$ koristi u svrhu spektralnog modeliranja čak i kad su prediktorski koeficijenti dobiveni metodom kovarijance.

10. HOMOMORFNA OBRADA GOVORNOG SIGNALA

10.1 Uvod

Jedna od osnovnih pretpostavki vezana uz obradu govornog signala je da se govor može prikazati kao izlaz iz linearnog, vremenski promjenljivog sustava čija se svojstva sporo mijenjaju s vremenom. To vodi prema osnovnom principu analize govora koji kaže da ako se promatraju dovoljno kratki segmenti govornog signala, da se tada svaki segment može učinkovito modelirati kao izlaz iz linearnog, vremenski invarijantnog sustava pobuđenog bilo kvazi-periodičnim impulsima bilo slučajnim šumom (engl. *random noise signal*). Problem govorne analize predstavlja određivanje parametara govornog modela kao i određivanje njihovih promjena u vremenu. Pošto konvolucija pobude i impulsnog odziva linearnog, vremenski invarijantnog sustava predstavlja njegov izlaz (govorni signal), ovaj se problem može promatrati i kao problem razdvajanja konvolviranih komponenti, što je poznato pod nazivom dekonvolucija. Dekonvolucija se može razmatrati sa stanovišta vremenski kratkotrajne Fourier-ove analize, ali i korištenjem koncepta homomorfne filtracije, što će biti objašnjeno u nastavku.

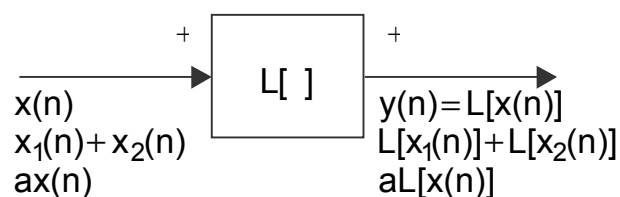
10.2 Sustavi homomorfni za konvoluciju

Homomorfni sustavi za konvoluciju podliježu općem principu superpozicije. Ovaj se princip za uobičajene linearne sustave izražava sljedećim uvjetima:

$$\begin{aligned} L[x(n)] &= L[x_1(n) + x_2(n)] = \\ &= L[x_1(n)] + L[x_2(n)] = \\ &= y_1(n) + y_2(n) = y(n) \end{aligned} \quad (10.2-1)$$

$$L[ax(n)] = aL[x(n)] = ay(n)$$

gdje je L linearni operator. Princip superpozicije kaže da ukoliko je ulazni signal sastavljen od linearne kombinacije elementarnih signala, tada je izlaz linearna kombinacija pripadnih izlaza linearnog operatora za svaki od tih ulaza. To je prikazano slikom 10.2-1 gdje simbol $+$ na ulazu i izlazu bloka $L[]$ označava da aditivna kombinacija na ulazu proizvodi aditivnu kombinaciju na izlazu.



Slika 10.2-1 Prikaz sustava koji zadovoljava princip superpozicije

Vrlo dobar primjer primjene principa superpozicije je izraz za konvoluciju gdje se odziv linearnog sustava na pobudu $x(n)$ nalazi kao superpozicija pomaknutih i skaliranih impulsnih odziva $h(n)$ prema izrazu:

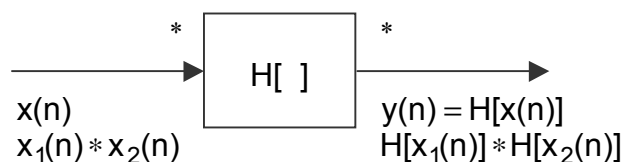
$$y(n) = \sum_{k=-\infty}^{\infty} h(n-k)x(k) = h(n) * x(n) \quad (10.2-2)$$

Simbol $*$ će u nastavku označavati operaciju vremenski diskretne linearne konvolucije. Po analogiji s principom superpozicije za obične linearne sustave može se definirati grupa sustava $H[]$ koja podliježe općem principu superpozicije kod kojeg je zbrajanje zamijenjeno operacijom konvolucije (može se lako pokazati da konvolucija ima ista algebarska svojstva kao i zbrajanje). Takvi sustavi moraju zadovoljavati slijedeći uvjet:

$$\begin{aligned} H[x(n)] &= H[x_1(n) * x_2(n)] = \\ &= H[x_1(n)] * H[x_2(n)] = \\ &= y_1(n) * y_2(n) = y(n) \end{aligned} \quad (10.2-3)$$

Oni bi morali zadovoljavati i dugo svojstvo principa superpozicije, tj. množenje s konstantom, ali to svojstvo nije značajno za primjene koje će se razmatrati u nastavku.

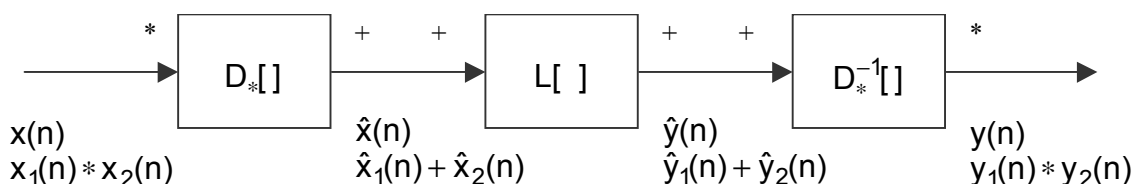
Sustavi koji imaju svojstvo dano izrazom (10.2-3) nazivaju se sustavi homomorfni za konvoluciju. Porijeklo ovakvog naziva dolazi iz činjenice da se ovakva transformacija može prikazati kao homomorfna transformacija u smislu linearnog vektorskog prostora. Primjer takvog sustava prikazan je slikom 10.2-2, gdje je operacija konvolucije $*$ označena eksplicitno na ulazu i na izlazu iz sustava.



Slika 10.2-2 Prikaz sustava homomorfnog za konvoluciju

Homomorfni filtar je u osnovi homomorfni sustav koji ima svojstvo da jedna (željena) komponenta prolazi kroz sustav nepromijenjena, dok je neželjena komponenta odstranjena. Npr., ako je u izrazu (10.2-3) neželjena komponenta $x_1(n)$, može se zahtijevati da izlaz $H[x_1(n)]$ koji odgovara ulazu $x_1(n)$ bude jedinični uzorak, dok izlaz $H[x_2(n)]$ koji odgovara ulazu $x_2(n)$ bude što bolja aproksimacija $x_2(n)$. Tada će konvolucija ta dva izlazna signala biti jednaka $H[x_2(n)]$ čime je odvojena samo željena komponenta. Ovo je u potpunosti analogno situaciji kod običnih linearnih sustava, gdje se pojavljuje problem razdvajanja željenog signala iz aditivne kombinacije signala i šuma.

Važan aspekt teorije homomorfni sustava je da se bilo koji takav sustav može prikazati kao kaskada tri homomorfna sustava kao što je predočeno slikom 10.2-3 za slučaj sustava homomorfnog za konvoluciju.



Slika 10.2-3 Kaskadni oblik sustava za homomorfnu dekonvoluciju

U prvi sustav ulazi signal koji predstavlja konvoluciju dvije komponente i transformira se u aditivnu kombinaciju odgovarajućih izlaza. Drugi sustav je uobičajeni linearni sustav koji podliježe principu superpozicije iskazanom izrazima (10.2-1). Treći sustav je inverzan prvom, što znači da transformira aditivnu kombinaciju signala nazad u konvolucijsku kombinaciju izlaza. Prednost ovakvog kaskadnog rastava homomorfni sustava je u činjenici

da je karakteristika ukupnog sustava određena središnjim linearnim sustavom koji se projektira za željenu primjenu. Suprotno tome sustav $D_*[\]$ koji se naziva karakteristični sustav za homomorfnu dekonvoluciju se ne mijenja (uvijek je jednak), a podliježe generaliziranom principu superpozicije gdje je ulazna operacija konvolucija, a izlazna operacija obično zbrajanje. Svojstva karakterističnog sustava su definirana na sljedeći način:

$$\begin{aligned} D_*[x(n)] &= D_*[x_1(n) * x_2(n)] = \\ &= D_*[x_1(n)] + D_*[x_2(n)] = \\ &= \hat{x}_1(n) + \hat{x}_2(n) = \hat{x}(n) \end{aligned} \quad (10.2-4)$$

Isto tako, inverzni karakteristični sustav $D_*^{-1}[\]$ je također fiksni sustav, a definiran na sljedeći način:

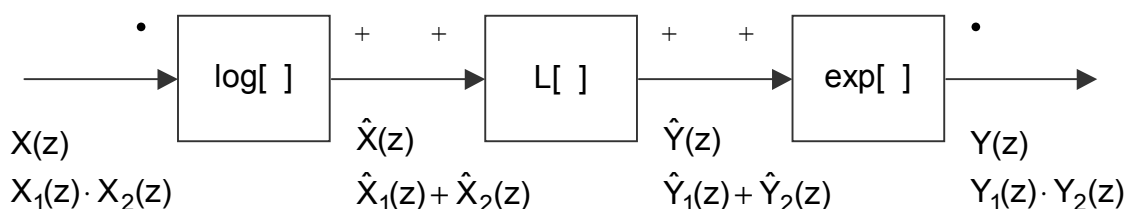
$$\begin{aligned} D_*^{-1}[\hat{y}(n)] &= D_*^{-1}[\hat{y}_1(n) + \hat{y}_2(n)] = \\ &= D_*^{-1}[\hat{y}_1(n)] * D_*^{-1}[\hat{y}_2(n)] = \\ &= y_1(n) * y_2(n) = y(n) \end{aligned} \quad (10.2-5)$$

Za određivanje karakterističnog sustava za dekonvoluciju potrebno je krenuti od ulaznog signala $x(n)$ koji predstavlja konvoluciju signala $x_1(n)$ i $x_2(n)$:

$$x(n) = x_1(n) * x_2(n), \quad (10.2-6)$$

Taj signal moguće je prikazati i u z-domeni kao produkt z-transformacija ulaznih signala:

$$X(z) = X_1(z) \cdot X_2(z) \quad (10.2-7)$$

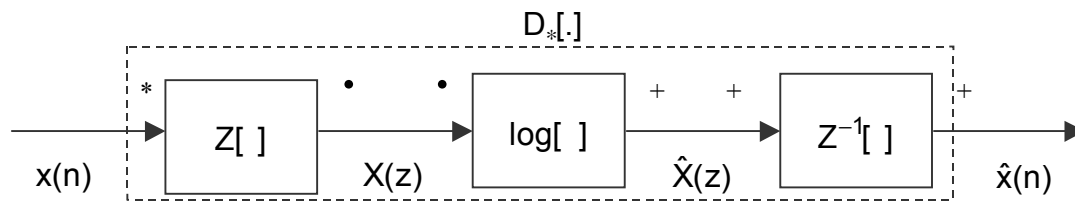


Slika 10.2-4 Prikaz sustava homomorfnog za konvoluciju u frekvencijskoj domeni

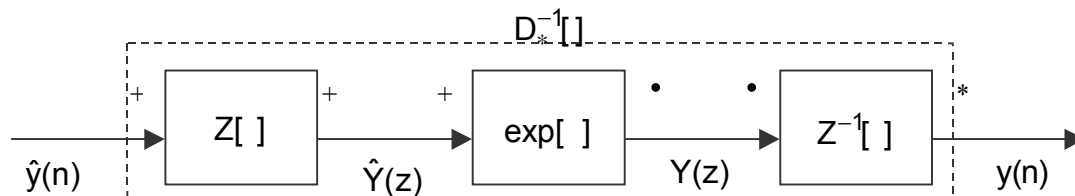
Iz izraza (10.2-4) proizlazi da z-transformacija izlaza karakterističnog sustava mora biti aditivna kombinacija dvaju z-transformacija. Prema tome, karakteristični sustav za dekonvoluciju u frekvencijskoj domeni mora imati svojstvo da ukoliko je ulazni signal produkt z-transformacija, tada izlaz mora biti zbroj transformiranih ulaza. Jedna mogućnost realizacije takvog sustava prikazana je na slici 10.2-4. Ona se temelji na činjenici da se logaritam produkta može prikazati kao zbroj logaritama množitelja, tj.:

$$\begin{aligned} \hat{X}(z) &= \log[X(z)] = \log[X_1(z) \cdot X_2(z)] = \\ &= \log[X_1(z)] + \log[X_2(z)] \end{aligned} \quad (10.2-8)$$

Ukoliko želimo prikazati signale u obliku vremenskih nizova, umjesto njihove reprezentacije u frekvencijskoj domeni kao na slici 10.2-4, karakteristični sustav se može prikazati kao što je učinjeno na slici 10.2-5. Slično tome, inverzni karakteristični sustav može se prikazati slikom 10.2-6.



Slika 10.2-5 Prikaz karakterističnog sustava za homomorfnu dekonvoluciju



Slika 10.2-6 Prikaz inverznog karakterističnog sustava za homomorfnu dekonvoluciju

Ova dva prikaza temeljeni su na valjanosti izraza (10.2-8), tj. logaritam mora biti definiran tako da vrijedi da je logaritam produkta jednak sumi logaritama. To je trivijalno za realne pozitivne veličine, međutim, Z-transformacija je u osnovi kompleksna veličina te se postavlja pitanje jednoznačnosti logaritma kompleksnog broja. Pošto se u stvarnim aplikacijama umjesto Z-transformacije koristi Fourier-ova transformacija, tj. Z-transformacija se promatra samo za vrijednosti na jediničnoj kružnici (za $z = e^{j\omega}$), dovoljno je osigurati da izraz (10.2-8) vrijedi u tim uvjetima. U tom slučaju se kompleksni logaritam može definirati kao:

$$\hat{X}(e^{j\omega}) = \log |X(e^{j\omega})| + j \cdot \arg [X(e^{j\omega})] \quad (10.2-9)$$

Iako je u ovom izrazu korišten prirodni logaritam, može se koristiti logaritam po bilo kojoj drugoj bazi. S realnim dijelom nema nekih osobitih poteškoća, dok se problemi jednoznačnosti pojavljuju pri definiranju imaginarnog dijela, koji je jednak kutu Z-transformacije na jediničnoj kružnici. Jedan od pristupa ovom problemu je zahtjev da fazni kut bude kontinuirana neparna funkcija od ω , čime je zadovoljen izraz (10.2-8). Naime fazni kut u izrazu (10.2-9) se mijenja u intervalu od $-\pi$ do π , tj. postoje lomovi i višeznačnost funkcije pa se tako definiran logaritam ne može direktno primijeniti. Rješenje problema se može postići postupkom nadopunjavanja osnovne vrijednosti faze (npr. $-\pi$ do π) cjelobrojnim višekratnicima od 2π . To znači da na svim mjestima gdje faza ima diskontinuitet reda veličine 2π , potrebno je pribrojiti cjelobrojni višekratnik od 2π koji osigurava kontinuiranost funkcije faze. Treba također osigurati da i na početku ($\omega=0$) i na kraju ($\omega=\pi$, $\omega=-\pi$) funkcija faze bude jednaka nuli.

Uz zadovoljene uvjete za računanje kompleksnog logaritma prema izrazu (10.2-8) i uz zamjenu blokova $Z[\]$ i $Z^{-1}[\]$ sa Fourier-ovom odnosno inverznom Fourier-ovom transformacijom, može se izračunati inverzna Fourier-ova transformacija kompleksnog logaritma Fourier-ove transformacije ulaza koja onda predstavlja izlaz karakterističnog sustava za homomorfnu dekonvoluciju, prema izrazu:

$$\hat{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{X}(e^{j\omega}) e^{jn\omega} d\omega \quad (10.2-10)$$

Izlaz karakterističnog sustava, $\hat{x}(n)$, se naziva kompleksni kepstar zbog činjenice što se u izrazu (10.2-9) za $\hat{X}(e^{j\omega})$ koristi kompleksni logaritam. Pored ovog kompleksnog kepstra postoji i običan kepstar koji se uobičajeno naziva samo kepstar (engl. *cepstrum*). Taj pojam je u literaturi prihvaćen za označavanje inverzne Fourier-ove transformacije logaritma spektra snage signala (dakle određen je isključivo modulom frekvencijske karakteristike dok se faza ignorira). Za razliku od kompleksnog kepstra $\hat{x}(n)$ taj običan kepstar $c(n)$ je definiran slijedećim izrazom:

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |X(e^{j\omega})| e^{jn\omega} d\omega \quad (10.2-11)$$

pri čemu sekvenca $c(n)$ predstavlja parni dio kompleksnog kepstara $\hat{x}(n)$. Valja naglasiti da iako se koristi termin "kompleksni kepstar", da su obje sekvence $\hat{x}(n)$ i $c(n)$ realne sekvence, jer su određene kao inverzna Fourier-ova transformacija simetričnih spektara.

U prethodnom izlaganju definiran je karakteristični sustav za homomorfnu dekonvoluciju, kao i temeljni kanonski oblik svih sustava homomorfni za konvoluciju. Svi sustavi ove grupe razlikuju se samo po linearnom dijelu sustava. Izbor linearnog sustava nužno ovisi o svojstvima njegovih ulaznih signala. Prema tome, pri realizaciji linearnog sustava neophodno je uzeti u obzir svojstva izlaznog signala karakterističnog sustava, tj. potrebno je proučiti svojstva kompleksnog kepstara za tipične ulazne signale, što će biti tema slijedećeg poglavlja.

10.3 Svojstva kompleksnog kepstra

Da bi se odredila svojstva kompleksnog kepstara, dovoljno je promotriti slučaj racionalne Z-transformacije. Najopćenitiji oblik razlomljene racionalne funkcije u varijabli Z koji predstavlja bilo prijenosnu funkciju linearnog sustava, bilo Z-transformaciju signala $x(n)$, može se prikazati slijedećim izrazom:

$$X(z) = \frac{Az^r \prod_{k=1}^{M_i} (1 - a_k z^{-1}) \prod_{k=1}^{M_o} (1 - b_k z)}{\prod_{k=1}^{N_i} (1 - c_k z^{-1}) \prod_{k=1}^{N_o} (1 - d_k z)} \quad (10.3-1)$$

gdje su sve vrijednosti veličina a_k, b_k, c_k, d_k po modulu manje od 1.

Prema tome, izrazi $(1 - a_k z^{-1})$ i $(1 - c_k z^{-1})$ odgovaraju nulama i polovima unutar jedinične kružnice, dok izrazi $(1 - b_k z)$ i $(1 - d_k z)$ odgovaraju nulama i polovima izvan jedinične kružnice. Faktor z^r predstavlja pomak vremenskog ishodišta, dok A predstavlja konstantu pojačanja. Uz pretpostavku da vrijedi (10.2-8), kompleksni logaritam od $X(z)$ je:

$$\hat{X}(z) = \log[A] + \log[z^r] + \sum_{k=1}^{M_i} \log(1 - a_k z^{-1}) + \sum_{k=1}^{M_0} \log(1 - b_k z) - \sum_{k=1}^{N_i} \log(1 - c_k z^{-1}) - \sum_{k=1}^{N_0} \log(1 - d_k z) \quad (10.3-2)$$

Ako izraz (10.3-2) evaluiramo na jediničnoj kružnici može se uočiti da izraz $\log[e^{jr\omega}]$ pridonosi samo imaginarnom dijelu kompleksnog logaritma. S obzirom da ovaj izraz nosi samo informaciju o pomaku vremenskog ishodišta, može se izostaviti iz proračuna, te se stoga može zanemariti u diskusiji o svojstvima kompleksnog kepstara. Koristeći činjenicu da logaritam izraza $1-x$ može biti zapisan kao red potencija u varijabli x , tj.

$$\log(1-x) = -(x + \frac{x^2}{2} + \frac{x^3}{3} + \dots + \frac{x^n}{n} + \dots)$$

Iako je pokazati da kompleksni kepstar razlomljene racionalne funkcije $X(z)$ ima slijedeći oblik:

$$\hat{x}(n) = \begin{cases} \log[A] & n = 0 \\ \sum_{k=1}^{N_i} \frac{c_k^n}{n} - \sum_{k=1}^{M_i} \frac{a_k^n}{n} & n > 0 \\ \sum_{k=1}^{M_0} \frac{b_k^{-n}}{n} - \sum_{k=1}^{N_0} \frac{d_k^{-n}}{n} & n < 0 \end{cases} \quad (10.3-3)$$

Izraz (10.3-3) omogućava uočavanje nekoliko važnih svojstava kompleksnog kepstara. Kao prvo, kompleksni kepstar nikad nije jednak nuli, tj. neograničenog (beskonačnog) je trajanja i za pozitivne i za negativne indekse n , pa čak i kada je polazni $x(n)$ kauzalan, stabilan i konačnog trajanja. Nadalje, očigledno je da je kompleksni kepstar padajući niz koji je po modulu ograničen s :

$$|\hat{x}(n)| < \beta \frac{\alpha^{|n|}}{|n|} \quad \text{za} \quad |n| \rightarrow \infty \quad (10.3-4)$$

gdje je α maksimum apsolutnih vrijednosti veličina a_k, b_k, c_k i d_k , a β je konstanta proporcionalnosti.

Ako $X(z)$ nema ni polove ni nule izvan jedinične kružnice, tj. $b_k = d_k = 0$, tada je :

$$\hat{x}(n) = 0 \quad \text{za} \quad n < 0 \quad (10.3-5)$$

Signali (sustavi) koji zadovoljavaju navedeni uvjet se nazivaju signali (sustavi) s minimalnom fazom (engl. *minimum phase*). Općenito vrijedi da se nizovi oblika danog izrazom (10.3-5) mogu u potpunosti opisati realnim dijelom svoje Fourier-ove transformacije. Stoga se kompleksni kepstar $\hat{x}(n)$ signala s minimalnom fazom može prikazati direktno na osnovu logaritma modula Fourier-ove transformacije signala $x(n)$, tj. na osnovu $\text{Re}(\hat{X}(e^{j\omega})) = \log|X(e^{j\omega})|$. Ovo se lako može dokazati jer vrijedi da je realni dio Fourier-ove transformacije u stvari jednak Fourier-ovoj transformaciji parnog dijela sekvence. Pošto je običan kepstar $c(n)$ jednak parnom dijelu kompleksnog kepstra $\hat{x}(n)$, tj. vrijedi:

$$c(n) = \frac{\hat{x}(n) + \hat{x}(-n)}{2} \quad (10.3-6)$$

tada se kombiniranjem izraza (10.3-5) i (10.3-6) može lako pokazati da je:

$$\hat{x}(n) = \begin{cases} 0 & n < 0 \\ c(n) & n = 0 \\ 2c(n) & n > 0 \end{cases} \quad (10.3-7)$$

Stoga se kompleksni kepstar sekvenci s minimalnom fazom može se odrediti izračunavanjem običnog kepstra $c(n)$ i potom primjenom izraza (10.3-7). Sljedeći važan rezultat za sekvence s minimalnom fazom je da se kompleksni kepstar može izračunati rekurzivno iz samog ulaznog signala, a formula za to je:

$$\hat{x}(n) = \begin{cases} 0 & n < 0 \\ \log(x(0)) & n = 0 \\ \frac{x(n)}{x(0)} - \sum_{k=0}^{n-1} \binom{k}{n} \hat{x}(k) \frac{x(n-k)}{x(0)} & n > 0 \end{cases} \quad (10.3-8)$$

Slični rezultati se mogu dobiti u slučaju kada $X(z)$ nema ni polove ni nule unutar jedinične kružnice. Takve signale (sustave) nazivamo signalima (sustavima) s maksimalnom fazom (engl. *maximum phase*). U ovom slučaju iz izraza (10.3-3) slijedi da je

$$\hat{x}(n) = 0 \quad \text{za} \quad n > 0 \quad (10.3-9)$$

Ukoliko upotrijebimo izraze (10.3-6) i (10.3-9) dolazimo do sličnog rezultata kao i u izrazima (10.3-7) i (10.3-8) s jedinom razlikom u tome što $n < 0$ i $n > 0$ mijenjaju mjesto.

$$\hat{x}(n) = \begin{cases} 0 & n > 0 \\ c(n) & n = 0 \\ 2c(n) & n < 0 \end{cases} \quad (10.3-10)$$

$$\hat{x}(n) = \begin{cases} 0 & n > 0 \\ \log(x(0)) & n = 0 \\ \frac{x(n)}{x(0)} - \sum_{k=n+1}^0 \binom{k}{n} \hat{x}(k) \frac{x(n-k)}{x(0)} & n < 0 \end{cases} \quad (10.3-11)$$

Posebno interesantan slučaj je kada se ulazni signal sastoji od niz impulsa na pravilnom razmaku N_p , tj. prema izrazu:

$$p(n) = \sum_{r=0}^M \alpha_r \delta(n - rN_p) \quad (10.3-12)$$

čija je Z-transformacija dana sljedećim izrazom:

$$P(z) = \sum_{r=0}^M \alpha_r z^{-rN_p} \quad (10.3-13)$$

Iz izraza (10.3-13) evidentno je da je $P(z)$ u stvari polinom po varijabli z^{-N_p} , a ne z^{-1} . Prema tome, $P(z)$ se može izraziti kao produkt korijenih faktora oblika $(1 - az^{-N_p})$ i $(1 - bz^{N_p})$, te se lako može vidjeti da će kompleksni kepstar, $\hat{p}(n)$, biti različit od nule samo za cjelobrojne višekratnike od N_p . Ako radi jednostavnosti pretpostavimo da se pobudni signal sastoji od samo dva impulsa :

$$p(n) = \delta(n) + \alpha\delta(n - N_p) \quad \text{za} \quad 0 < \alpha < 1, \quad (10.3-14)$$

tada su:

$$P(z) = 1 + \alpha z^{-N_p} \quad (10.3-15)$$

i

$$\hat{P}(z) = \log(1 + \alpha z^{-N_p}) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\alpha^n}{n} z^{-nN_p} \quad (10.3-16)$$

Dakle, kompleksni kepstar $\hat{p}(n)$ je beskonačni niz impulsa razmaknutih za N_p :

$$\hat{p}(n) = \sum_{r=1}^{\infty} (-1)^{r+1} \frac{\alpha^r}{r} \delta(n - rN_p) \quad (10.3-17)$$

Činjenica da je kompleksni kepstar niza jednoliko razmaknutih impulsa također jednoliko razmaknuti niz impulsa na istom razmaku, je veoma važan rezultat za analizu govora. Važno je uočiti da amplituda tih impulsa teži prema 0 kako r raste, pod uvjetom da je korijen unutar jedinične kružnice ($\alpha < 1$). Interesantna je i činjenica da ne postoji impuls u $n=0$, tj. $\delta(n)$, već tek u $n=N_p$ i zatim u svim njegovim cjelobrojnim višekratnicima ($r=2,3,\dots$).

10.4 Postupci izračunavanja kepstra

U prethodnom poglavlju bilo je prikazano kako se mogu formirati karakteristični sustav homomorfni za konvoluciju, kao i njegov inverz. Ako sada razmatranje ograničimo samo na ulazne signale koji imaju konačnu energiju, tada njihova z -transformacija ima područje konvergencije koje uključuje jediničnu kružnicu. To drugim riječima znači da takvi signali imaju definiranu Fourier-ovu transformaciju. U tom slučaju moguće je z -transformaciju u karakterističnom sustavu za dekonvoluciju i njegovom inverzu zamijeniti Fourier-ovom transformacijom, kao što je već najavljeno. Za konačno duge ulazne signale, karakteristični sustav za homomorfnu dekonvoluciju opisan je slijedećim izrazima:

$$X(e^{j\omega}) = \sum_{n=0}^{N-1} x(n)e^{-j\omega n} \quad (10.4-1)$$

$$\hat{X}(e^{j\omega}) = \log[X(e^{j\omega})] = \log|X(e^{j\omega})| + j\arg[X(e^{j\omega})] \quad (10.4-2)$$

$$\hat{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{X}(e^{j\omega}) e^{j\omega n} d\omega \quad (10.4-3)$$

Izrazom (10.4-1) dana je Fourierova transformacija ulaznog signala. Izrazom (10.4-2) dan je izraz za kompleksni logaritam Fourierove transformacije, a izrazom (10.4-3) dana je inverzna Fourierova transformacija kompleksnog logaritma Fourierove transformacije ulaznog signala. Da bi sa ovim izrazima ((10.4-1) do (10.4-3)) mogli jednoznačno definirati kompleksni kepar, potrebno je prvo definirati kompleksni logaritam Fourierove transformacije. Prvi zahtjev koji mora biti osiguran je da kompleksni kepar realnog ulaznog signala također bude realni signal. Inverznom Fourier-ovom transformacijom dobiva se realni signal po uvjetom da je realni dio transformata bio parna funkcija, a imaginarni dio neparna funkcija od ω . To znači da je za realan kompleksni kepar potrebno osigurati da logaritam modula Fourier-ove transformacije signala bude paran po ω , odnosno da faza Fourier-ove transformacije bude neparna po ω . Nadalje se može pokazati da je dovoljan uvjet za jednoznačnost kompleksnog logaritma taj da faza bude kontinuirana periodička funkcija od ω sa periodom 2π . Taj uvjet kontinuiteta također mora vrijediti i za $\hat{X}(e^{j\omega})$ da bi bio valjana Fourierova transformacija.

Problem koji se javlja u gornjim izrazima (10.4-1) do (10.4-3) je potreba za izračunavanjem integrala. Taj problem se rješava aproksimacijom gornjih izraza primjenom diskretne Fourier-ove transformacije (DFT). DFT konačno dugog ulaznog signala možemo predstaviti kao otipkanu verziju Fourierove transformacije. Nadalje, DFT možemo vrlo učinkovito izračunati koristeći brzu Fourier-ovu transformaciju (FFT).

Dakle pristup izračunavanja kompleksnog keptra se sastoji u zamijeni Fourier-ove transformacije DFT-om. U tom slučaju izrazi (10.4-1) do (10.4-3) prelaze u slijedeće:

$$X_p(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi kn}{N}} \quad 0 \leq k \leq N-1 \quad (10.4-4)$$

$$\hat{X}_p(k) = \log[X_p(k)] \quad 0 \leq k \leq N-1 \quad (10.4-5)$$

$$\hat{x}_p(n) = \frac{1}{N} \sum_{k=0}^{N-1} \hat{X}_p(k) e^{j\frac{2\pi kn}{N}} \quad 0 \leq n \leq N-1 \quad (10.4-6)$$

Izrazom (10.4-6) opisan je inverzni DFT kompleksnog logaritma DFT-a ulaznog signala konačne duljine. Sufiks $_p$ upućuje da rezultirajući signali nisu potpuno jednaki onima dobivenim izrazima (10.4-1) do (10.4-3). To je stoga što je kompleksni logaritam $\hat{X}_p(k)$ korišten u izračunavanju pomoću DFT-a otipkana verzija izraza (10.4-2), te konačni izraz (10.4-6) u sebi sadrži pojavu aliasing-a, odnosno preklapanja spektra. Diskretizacija spektra uzrokuje periodičnost keptra, tj. kompleksni kepar izračunat pomoću DFT-a možemo dobiti periodičkim ponavljanjem kompleksnog keptra izračunatog korištenjem prave Fourierove transformacije svakih N uzoraka, tj. prema slijedećem izrazu:

$$\hat{x}_p(n) = \sum_{r=-\infty}^{\infty} \hat{x}(n+rN) \quad (10.4-7)$$

Blok shema karakterističnog sustava za dekonvoluciju za izračunavanje kompleksnog kepstra primjenom DFT-a prikazana je na slici 10.4-1.



Slika 10.4-1 Postupak izračunavanja kompleksnog kepstra

Kod kompleksnog kepstra koristi se kompleksni logaritam. Međutim kod običnog kepstra računa se samo običan logaritam modula Fourier-ove transformacije, kao što je dano izrazom (10.4-8).

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(|X(e^{j\omega})|\right) e^{j\omega n} d\omega, \quad -\infty < n < \infty \quad (10.4-8)$$

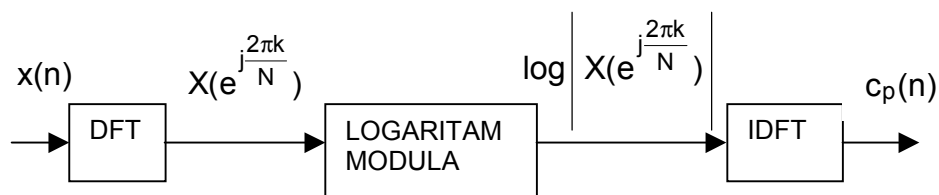
I u ovom slučaju može se koristiti aproksimacija stvarnog kepstra korištenjem DFT-a umjesto Fourier-ove transformacije. To znači da se kepstar $c_p(n)$ nalazi kao inverzni DFT logaritma modula DFT-a ulaznog niza, tj. prema izrazu:

$$c_p(n) = \frac{1}{N} \sum_{k=0}^{N-1} \log(|X_p(k)|) e^{j\frac{2\pi}{N}nk}, \quad 0 \leq n \leq N-1 \quad (10.4-9)$$

I u ovom slučaju zbog pojave aliasinga uslijed otipkavanja Fourier-ove transformacije aproksimacija kepstra $c_p(n)$ i stvarni kepstar su vezani slijedećim izrazom:

$$c_p(n) = \sum_{r=-\infty}^{\infty} c(n+rN) \quad (10.4-10)$$

Na slici 10.4-2 prikazana je blok shema postupka određivanja kepstra $c_p(n)$.



Slika 10.4-2 Postupak izračunavanja kepstra korištenjem DFT-a

Zbog opisane pojave aliasinga kod korištenja DFT-a, potrebno je računanje DFT-a provesti sa velikim brojem uzoraka N . Velika rezolucija otipkavanja Fourier-ove transformacije potrebna je i kod izračunavanja kompleksnog logaritma, gdje je potrebno osigurati kontinuitet fazne karakteristike. Korištenje velikog broja uzorka ($N=512$ ili više) ne predstavlja problem pošto je diskretnu Fourier-ovu transformaciju moguće realizirati pomoću brzih algoritama kao što je FFT.

10.4.1 Kepstar i kompleksni kepstar govornog signala

Postupci vremenski kratkotrajne analize govornog signala, kod kojih se analiza provodi nad segmentima konačnog trajanja, lako se mogu primijeniti i na homomorfnu obradu. Model formiranja govornog signala temeljen je na vremenski sporo promjenjivom linearnom sustavu pobuđenim s periodičnim nizom impulsa ili sa slučajnim šumom. Za dovoljno kratki segment govora može se pretpostaviti sa su parametri sustava bili stalni. Stoga za kratki zvučni segment govora moguće je pretpostaviti da je nastao pobuđivanjem linearnog vremenski nepromjenjivog sustava periodičnim nizom impulsa $p(n)$ perioda N_p .

$$\begin{aligned} s(n) &= p(n) * g(n) * v(n) * r(n) = p(n) * h_v(n) \\ &= \sum_{r=-\infty}^{\infty} h_v(n - rN_p) \end{aligned} \quad (10.4-11)$$

Segment govora $s(n)$ dobiva se periodičkim ponavljanjem impulsnog odziva $h_v(n)$ svakih N_p uzoraka. Impulsni odziv linearnog sustava $h_v(n)$ obuhvaća valni oblik glotalnog pulsa $g(n)$, impulsni odziv vokalnog trakta $v(n)$ i impulsni odziv zračenja na usnicama $r(n)$.

Analogno tome, kratki segment bezvučnog govora se dobiva pobuđivanjem linearnog sustava s impulsnim odzivom $h_u(n)$ sa slučajnim šumom $u(n)$.

$$s(n) = u(n) * v(n) * r(n) = u(n) * h_u(n) \quad (10.4-12)$$

U ovom slučaju impulsni odziv linearnog sustava $h_u(n)$ obuhvaća impulsni odziv vokalnog trakta i impulsni odziv zračenja na usnicama. Izraze (10.4-11) i (10.4-12) moguće je zapisati i u z-domeni:

$$\begin{aligned} H_v(z) &= G(z) \cdot V(z) \cdot R(z) \\ H_u(z) &= V(z) \cdot R(z) \end{aligned} \quad (10.4-13)$$

Za analitičko izračunavanje kompleksnog kepstara govornog signala, potrebno je poznavati ove tri prijenosne funkcije $G(z)$, $V(z)$ i $R(z)$. Prijenosna funkcija vokalnog trakta $V(z)$ ima slijedeći opći oblik:

$$V(z) = \frac{Az^{-M} \prod_{k=1}^{M_i} (1 - a_k z^{-1}) \prod_{k=1}^{M_o} (1 - b_k z)}{\prod_{k=1}^{N_i} (1 - c_k z^{-1})} \quad (10.4-14)$$

Za sve zvučne glasove osim nazalnih, $V(z)$ ima samo polove, tj. $a_k=0$ i $b_k=0$ za sve k . Kod nazalnih glasova i kod bezvučnih glasova prijenosna funkcija sadrži i polove i nule. Neke od tih nula se mogu nalaziti i izvan jedinične kružnice, ali zbog razloga stabilnosti svi polovi c_k moraju biti unutar kružnice. Pošto je impulsni odziv $v(n)$ realan, polovi i nule se ili pojavljuju u konjugirano kompleksnim parovima, ili su realni.

Efekt zračenja na usnicama moguće je u grubo modelirati visoko propusnom funkcijom oblika:

$$R(z) \approx 1 - z^{-1} \quad (10.4-15)$$

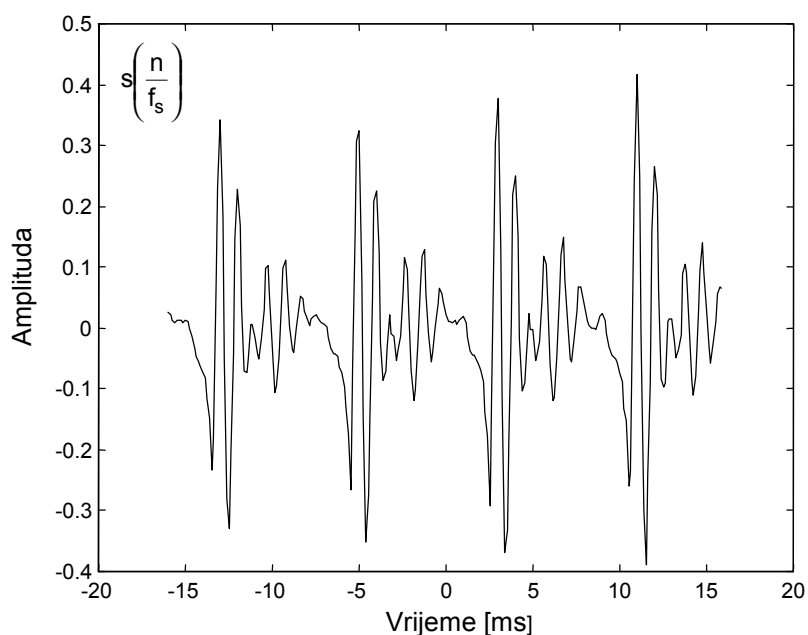
Kod zvučnih glasova ukupna prijenosna funkcija ovisi i o z-transformaciji valnog oblika glotalnog pulsa $g(n)$, koji je vremenski niz konačnog trajanja pa se stoga i može modelirati kao FIR sustav:

$$G(z) = B \prod_{k=1}^{L_i} (1 - \alpha_k z^{-1}) \prod_{k=1}^{L_o} (1 - \beta_k z) \quad (10.4-16)$$

gdje su α_k i β_k nule koje se nalaze unutar odnosno izvan jedinične kružnice.

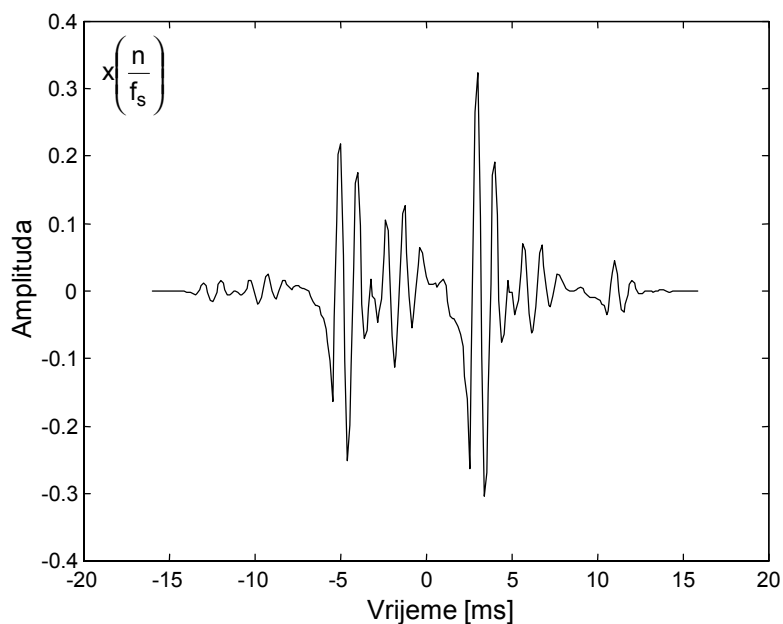
Uz ovako zadane prijenosne funkcije i uz poznavanje svojstava kompleksnog kepstra koja su diskutirana u poglavlju 10.3, moguće je predvidjeti očekivani oblik kompleksnog kepstra. Za zvučne glasove, ukupna prijenosna funkcija $H_v(z)$ imat će polove unutar jedinične kružnice i nule unutar i izvan jedinične kružnice, pa se stoga može zaključiti da se ne radi o funkciji minimalne faze. Zbog tog razloga kompleksni kepstar će biti različit od nule i lijevo i desno od $n=0$. Pošto se kod zvučnih glasova pobuda može predstaviti periodičnim nizom jediničnih impulsa na razmaku N_p , može se očekivati da se u kompleksnom kepstru zvučnog segmenta pojave šiljci na indeksu N_p i njegovim cjelobrojnim višekratnicima. Slična svojstva vrijede i za običan kepstar $c(n)$, koji predstavlja parni dio kompleksnog kepstra $\hat{x}(n)$, i koji će upravo zbog tog svojstva parnosti biti simetričan oko nule.

Ova svojstva će biti ilustrirana na primjeru kepstra $c_p(n)$ govornog odsječka glasa 'e' otipkanog s frekvencijom otipkavanja $f_s=8000$ Hz, prikazanog na slici 10.4-3. Kepstralna analiza je provedena korištenjem DFT-a u $N=256$ točaka, prema blok shemi prikazanoj na slici 10.4-2.

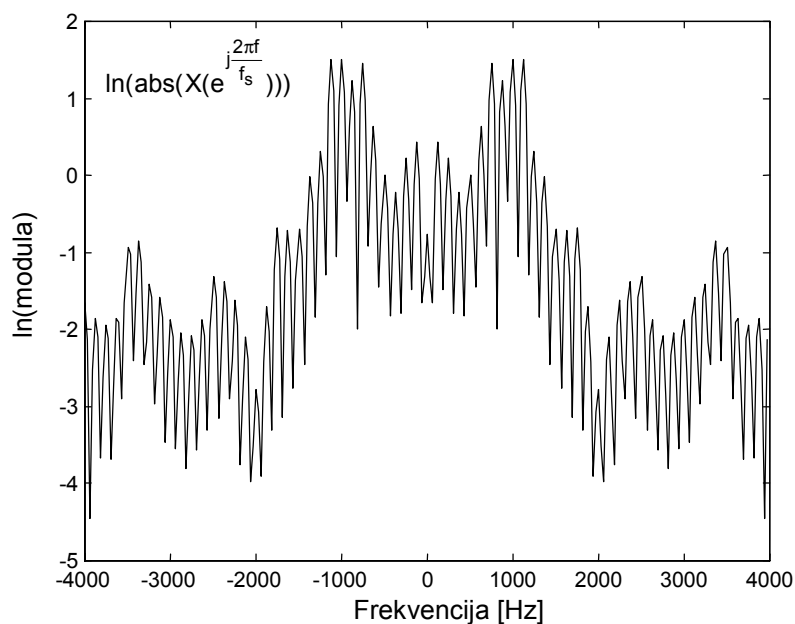


Slika 10.4-3 Valni oblik govornog signala za samoglasnik 'e'

Prije izračunavanja DFT-a potrebno je signal pomnožiti s vremenskim otvorom. Kod obrade govornih signala vrlo se često koristi Hamming-ov vremenski otvor, no moguće je koristiti i druge vremenske otvore s dobrim spektralnim svojstvima. U slučaju korištenja pravokutnog vremenskog otvora, dolazi do pojave spektralnog rasipanja, gdje spektralne komponente velikih amplituda prekrivaju svojim bočnim laticama spektralne komponente malih amplituda. Ovime se potpuno gubi formantna struktura, jer DFT vrlo loše aproksimira stvarni spektar, zbog konvolucije sa sporo padajućim spektrom pravokutnog vremenskog otvora. Signal sa slike 10.4-3 nakon množenja s Blackman-ovim vremenskim otvorom prikazan je na slici 10.4-4.



Slika 10.4-4 Signal pomnožen s Blackman-ovim vremenskim otvorom

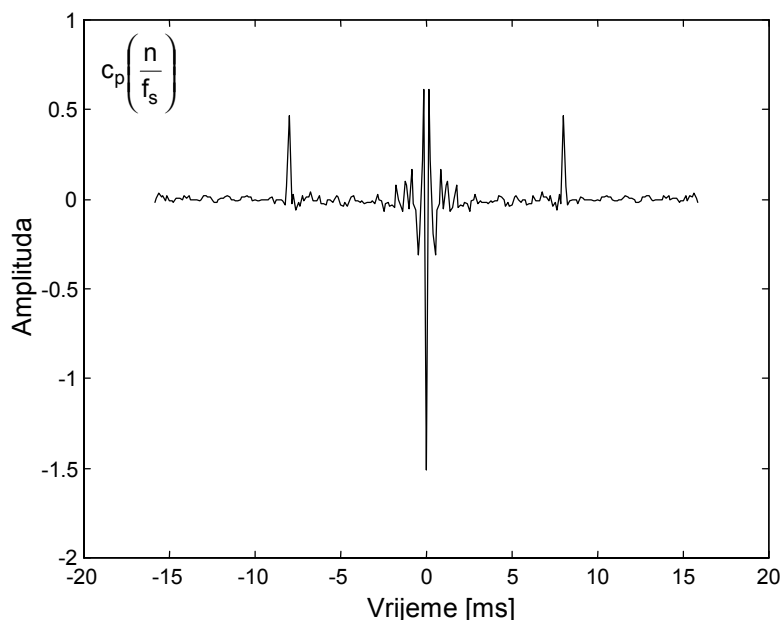


Slika 10.4-5 Prirodni logaritam modula DFT-a

Prirodni logaritam modula diskretne Fourier-ove transformacije signala sa slike 10.4-4 prikazan je na slici 10.4-5. Pojava periodičnih latica na slici 10.4-5 je posljedica periodičnosti pobudnog signala vokalnog trakta. DFT idealnog pobudnog signala vokalnog trakta $p(n)$ jednak je ovome prikazanom na slici 10.4-5, s tom razlikom da su sve latice iste visine. Amplitude latica prikazanih na slici 10.4-5 određene su prijenosnom funkcijom $H_V(z)$ za

$z = e^{j\frac{2\pi k}{N}}$, tj. otipkanom prijenosnom funkcijom u N ekvidistantnih točaka na jediničnoj kružnici. Zbog toga moguće je zaključiti da je sporo promjenljiva spektralna ovojnica određena prijenosnom funkcijom $H_V(z)$, dok su brze oscilacije određene DFT-om pobudnog signala $p(n)$. Razmak između svake dvije susjedne laticе odgovara osnovnoj frekvenciji

titranja glasnica. U primjeru na slici 10.4-5 moguće je prebrojiti 32 latice na intervalu od 0 do 4000 Hz, pa se lako može procijeniti da osnovna frekvencija iznosi $f_0=125$ Hz. Nakon logaritmiranja modula DFT-a ova dva dijela (ovojnica i brzo oscilirajuće latice) vezana su aditivnom superpozicijom. Ako za kratko ignoriramo činjenicu da slika 10.4-5 prikazuje modul spektra signala prikazan u logaritamskom mjerilu, nego ga interpretiramo kao običan vremenski promjenljiv valni oblik, tada u tom valnom obliku lako prepoznamo dva dijela: prvi sporo promjenljivi dio i drugi brzo oscilirajući dio. Drugim riječima, nisko odnosno visoko propusnom filtracijom ovog signala bilo bi moguće razdvojiti ta dva dijela. Isto tako ako bi izračunali Fourier-ovu transformaciju tog signala, sporo promjenljive komponente nalazile bi se na frekvencijama blizu nule, dok bi se brzo promjenljiva komponenta pojavila kao šiljci na mjestima koja odgovaraju cjelobrojnim višekratnicima frekvencije tih brzih oscilacija. Ovo je i bila originalna zamisao autora koji je prvi predložio postupke dekonvolucije temeljene na logaritmu modula DFT-a. Pojam kepstra upravo proizlazi iz činjenice da se ovdje radi o svojevrsnom spektru od spektra, pa je autor predložio naziv **kepstar** koji se formira od riječi **spekstar** tako da se izvorne redoslijed prva četiri slova. Kasnije je ovaj inženjerski pristup upotpunjen matematičkim modelom homomorfne obrade koji se temelji na z-transformaciji i kompleksnom kepstru, koji je opisan u prošlom poglavlju. Iako se u karakterističnom sustavu za dekonvoluciju nakon logaritmiranja koristi inverzna diskretna Fourier-ova transformacija (izrazi (10.4-6) odnosno (10.4-9)), a ne diskretna Fourier-ova transformacija kao što je to bilo originalno predloženo, gore opisana svojstva za slučaj Fourier-ove transformacije logaritma modula DFT-a signala biti će identična. DFT i inverzni DFT se razlikuju isključivo u faktoru skale (1 za DFT, odnosno $1/N$ za inverzni) i u fazama baznih funkcija transformacije ($f_k(n) = e^{-j\frac{2\pi kn}{N}}$ za DFT, odnosno $g_n(k) = e^{j\frac{2\pi kn}{N}}$ za inverzni). Kepstar $c_p(n)$ dobiven inverznom DFT transformacijom prirodnog logaritma modula DFT-a signala sa slike 10.4-4 prikazan je na slici 10.4-6.

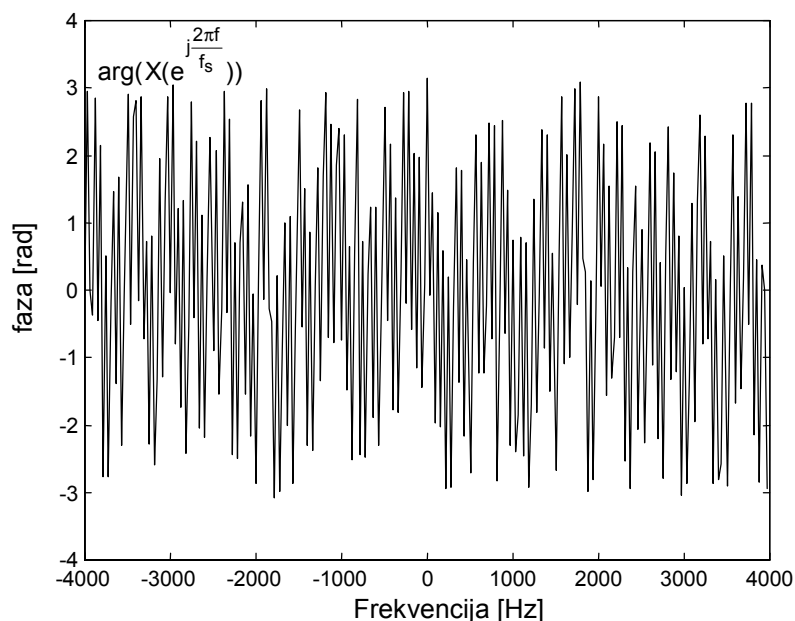


Slika 10.4-6 Kepstar glasa 'e'

Brzo padajući dio kepstra oko nule (približno ± 3 ms) proizlazi od sporo promjenljive spektralne ovojnice DFT-a, tj. jednoznačno opisuje modul prijenosne funkcije $H_v(z)$, koja je određena valnim oblikom glotalnog pulsa, impulsnim odzivom vokalnog trakta i utjecajem zračenja na usnicama. S druge strane, pulsevi na pozicijama -8 ms i $+8$ ms su posljedica

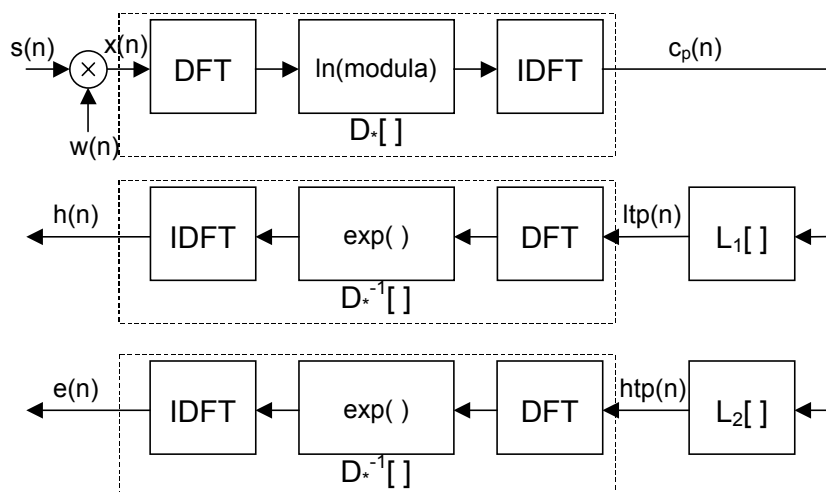
periodičnosti pobudnog signala vokalnog trakta osnovne frekvencije $f_0=125$ Hz. Vrlo značajno svojstvo koje je vidljivo na slici 10.4-6 jest činjenica da su te dvije komponente kepstra razdvojene u vremenu, tj. dio blizak nuli (engl. *low time part*) je određen sa $h_v(n)$, dok je preostali dio koji je daleko od nule (engl. *high time part*) određen sa $p(n)$. Vremenska dislokacija ta dva aditivno vezana dijela će omogućiti njihovo razdvajanje običnim (jednostavnim) linearnim sustavom (vremenskim otvorom).

U slučaju korištenja kompleksnog kepstra, potrebno je izračunati i fazu DFT-a. Za signal sa slike 10.4-4, osnova vrijednost faze ($-\pi$ do π) prikazana je na slici 10.4-7.



Slika 10.4-7 Faza DFT-a segmenta govornog signala glasa "e"

Kao što je već opisano potrebno je osigurati kontinuiranost ove fazne karakteristike, tj. na svim mjestima gdje je fazni skok sumjerljiv sa 2π , potrebno je fazi prikazanoj na slici 10.4-7 pribrojiti cjelobrojni višekratnik od 2π . Ovaj postupak nije jednostavan, jer linearni fazni član uslijed vremenskog pomaka glotalnog pulsa u odnosu na poziciju vremenskog otvora može imati vrlo veliki nagib, pa je vrlo teško odrediti pozicije stvarnih 'preskoka'. Stvar se dodatno komplicira kod zvučnih glasova gdje uslijed periodičnosti pobudnog signala velik broj komponenata DFT-a može imati modul blizak nuli. Faza tih komponenata određena je šumom i numeričkim pogreškama, a ne fazom pobudnog signala $p(n)$ i fazom prijenosne funkcije $H_v(z)$. To su razlozi zbog kojih se kompleksni kepstar rijetko koristi, pa se većina postupaka numeričke homomorfne analize govornog signala ograničava na običan kepstar $c_p(n)$ prikazan na slici 10.4-6.



Slika 10.4-8 Blok shema sustava za homomorfnu dekonvoluciju govora korištenjem DFT-a

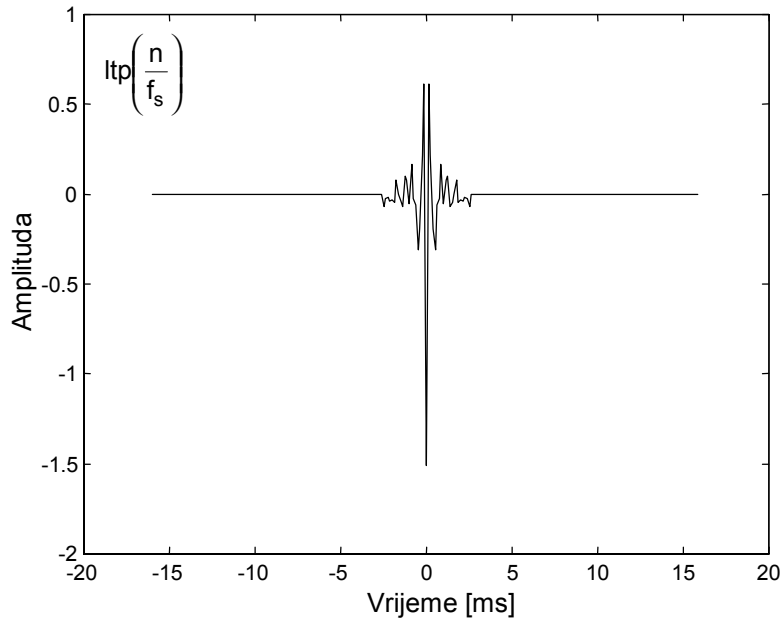
Korištenjem opisanog svojstva razdvajanja pobudnog signala i prijenosne funkcije linearnog sustava u kepralnoj domeni moguće je formirati cijeli sustav za homomorfnu dekonvoluciju. Blok shema takvog sustava prikazana je na slici 10.4-8. Govorni signal $s(n)$ množi se s vremenskim otvorom $w(n)$, te se na tako izdvojenom segmentu govornog signala primjenjuje karakteristični sustav za homomorfnu dekonvoluciju $D^*[\]$, čiji je izlaz keprstar $c_p(n)$. Na keprstar se primjenjuju dva linearna operatora $L_1[\]$ i $L_2[\]$, te se ta dva dijela zatim propuštaju kroz inverzni karakteristični sustav $D^{*-1}[\]$. Ovaj postupak rezultira s dva izlazna signala $h(n)$ i $e(n)$. Uz pravilno odabrane linearne operatore $L_1[\]$ i $L_2[\]$, $h(n)$ će biti jednak impulsnom odzivu linearnog sustava, dok će $e(n)$ biti jednak njegovoj pobudi. Iz diskusija o svojstvima kepra $c_p(n)$ prikazanog na slici 10.4-6, lako se može zaključiti da se operator $L_1[\]$, koji treba primijeniti na $c_p(n)$ radi izdvajanja impulsnog odziva sustava, može realizirati običnim kepralnim vremenskim otvorom $l_1(n)$, prema izrazu (10.4-17).

$$ltp(n) = L_1[c_p(n)] = c_p(n) \cdot l_1(n), \quad \text{gdje} \quad l_1(n) = \begin{cases} 1 & \text{za } |n| < n_0 \\ 0 & \text{za } |n| \geq n_0 \end{cases} \quad (10.4-17)$$

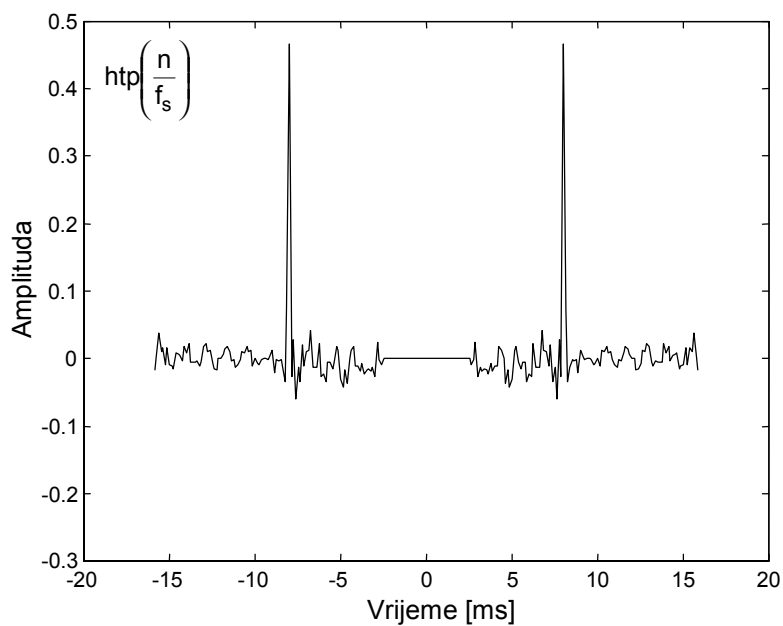
Opisanim postupkom se izdvaja n_0-1 uzoraka kepra lijevo i desno od nule. Ovaj dio kepra se naziva '*low time part*' ili *ltp* dio, jer odgovara dijelu s 'niskim' vremenskim indeksima. Rubni indeks n_0 se odabire tako da ne zahvati pulseve uslijed periodičnosti pobude, tj. $n_0 < f_s/f_0$, gdje je f_s frekvencija otipkavanja, a f_0 osnovna frekvencija titranja glasnica. Operator $L_2[\]$, je komplementaran operatoru $L_1[\]$, tj. on propušta ono što $L_1[\]$ ne propušta, pa vrijedi izraz:

$$htp(n) = L_2[c_p(n)] = c_p(n) \cdot l_2(n), \quad \text{gdje} \quad l_2(n) = \begin{cases} 0 & \text{za } |n| < n_0 \\ 1 & \text{za } |n| \geq n_0 \end{cases} \quad (10.4-18)$$

Izlaz ovog operatora se naziva '*high time part*' ili *htp* dio kepra jer odgovara 'visokim' vremenskim indeksima. Na slikama 10.4-9 i 10.4-10 prikazani su *ltp* i *htp* dio kepra sa slike 10.4-6, uz rubni indeks $n_0=20$, što odgovara intervalu ± 2.5 ms oko nule.

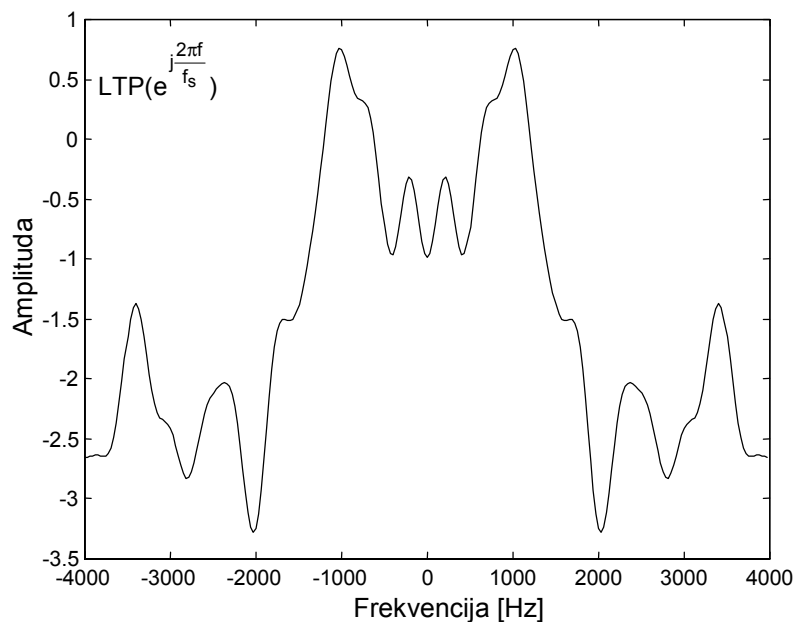


Slika 10.4-9 Nisko-vremenski dio kepstra (ltp)



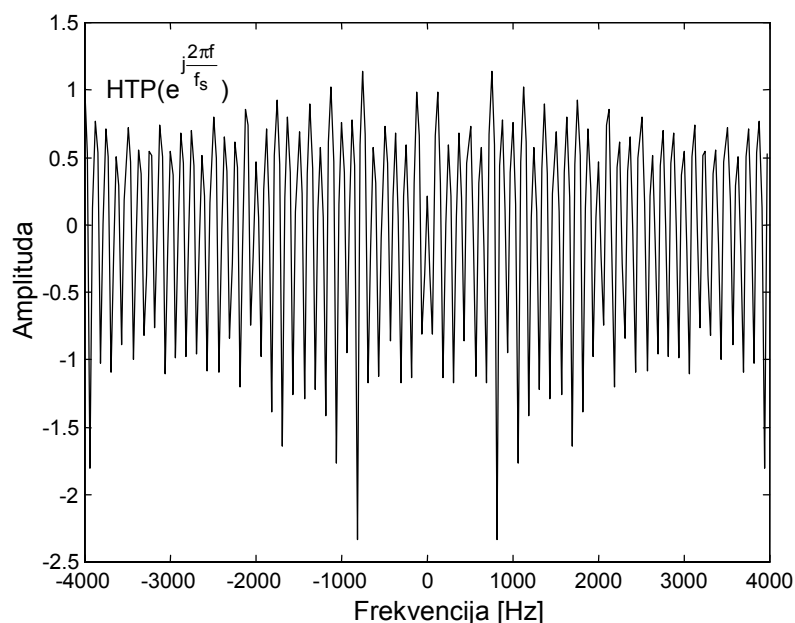
Slika 10.4-10 Visoko-vremenski dio kepstra (htp)

Ova dva dijela se zatim obrađuju inverznim karakterističnim sustavom $D_*^{-1}[\]$, a prvi korak u toj obradi je izračunavanje DFT-a ta dva signala. Pošto su oba signala parna (simetrična oko nule) DFT će biti realni (imaginarni dijelovi su jednaki nula). DFT od ltp dijela i DFT od htp dijela kepstra prikazani su na slikama 10.4-11 i 10.4-12.



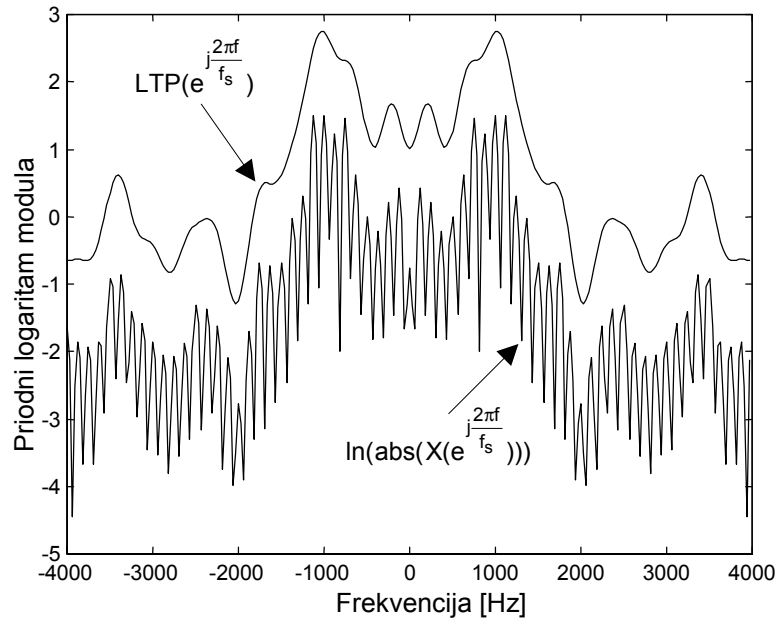
Slika 10.4-11 DFT ltp dijela kepstra

Na ovim slikama je vidljiv rezultat željene operacije razdvajanja dvije konvolvirane komponente (pobudnog signala i impulsnog odziva sustava) na sumu dvije DFT transformacije. Pošto je suma operatora $L_1[]$ i $L_2[]$ jednaka jediničnom operatoru, suma DFT-a od ltp dijela i DFT-a od htp dijela kepstra je identički jednaka logaritmu modula DFT-a koji je prikazan na slici 10.4-5.



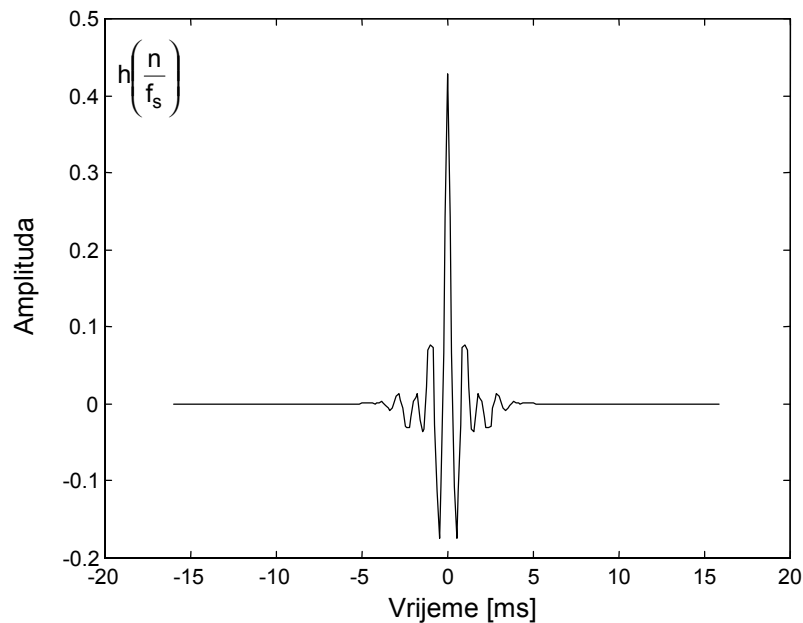
Slika 10.4-12 DFT htp dijela kepstra

Radi bolje ilustracije činjenice da ltp dio kepstra jednoznačno opisuje sporo promjenljivu spektralnu ovojnicu govornog signala, na slici 10.4-13 prikazan je prirodni logaritam modula DFT-a signala sa slike 10.4-4, zajedno s DFT-om ltp dijela kepstra. Radi jasnijeg prikaza DFT ltp dijela je podignut u odnosu na pravu poziciju za 2 prema gore.

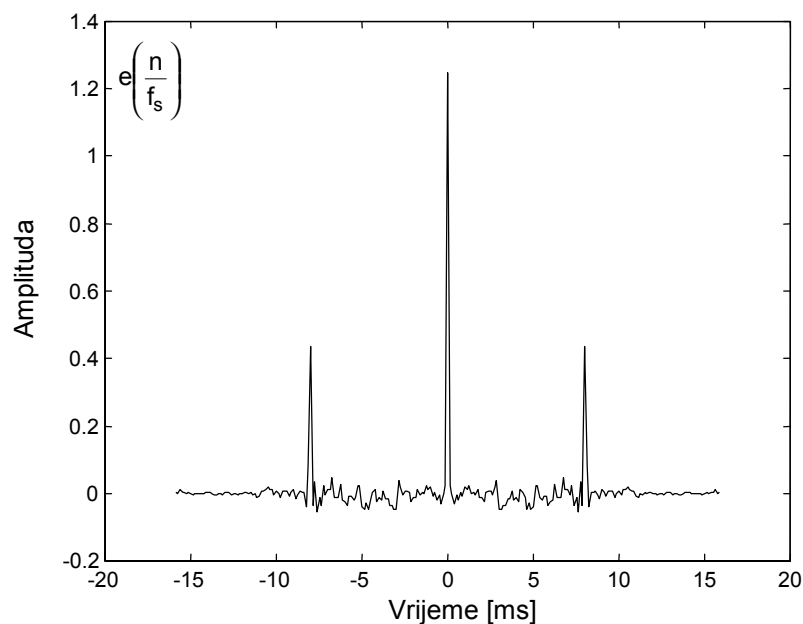


Slika 10.4-13 Modeliranje spektralne ovojnice ltp dijelom kepstra

Završni korak u ovom postupku homomorfne dekonvolucije je eksponenciranje ovih DFT-a, i izračunavanje inverzne DFT transformacije, što rezultira signalima $h(n)$ i $e(n)$. Ovi signali su također parni, jer kod korištenja običnog kepstra (za razliku od kompleksnog kepstra), DFT od ltp i htp dijela kepstra su realni. Pobudni signal $e(n)$ je vrlo sličan idealnom pobudnom signalu za zvučne glasove (niz jediničnih impulsa), s time da amplituda ovih pulseva opada u skladu s oblikom primijenjenog vremenskog otvora $w(n)$.

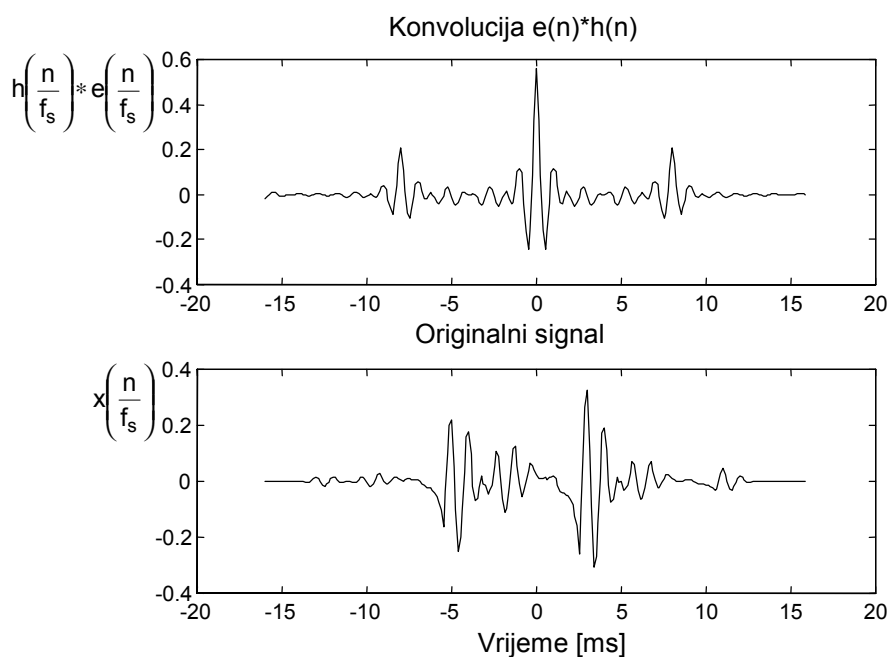


Slika 10.4-14 Impulsni odziv linearnog sustava, $h(n)$



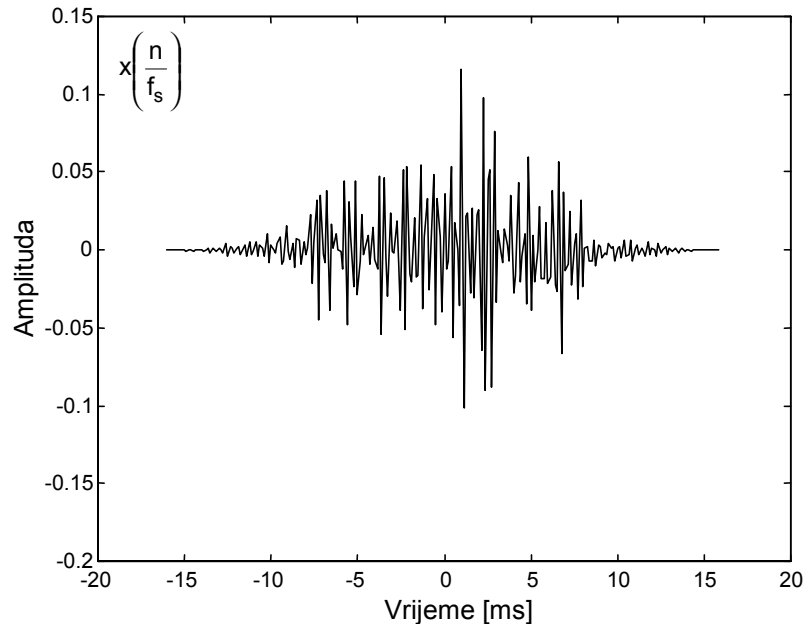
Slika 10.4-15 Pobudni signal linearnog sustava $e(n)$

Konvolucijom ova dva signala $e(n)$ i $h(n)$ trebao bi se dobiti početni govorni segment $x(n)$ izdvojen vremenskim otvorom. Međutim, ovo vrijedi samo u slučaju korištenja kompleksnog kepstra $\hat{x}_p(n)$. U slučaju korištenja običnog kepstra $c_p(n)$, dobiva se signal prikazan u gornjoj polovici slike 10.4-16 koji se očito razlikuje od signala $x(n)$ prikazanog u donjoj polovici iste slike. Ako bi se izračunao DFT od ova dva signala u istom broju točaka N koji je korišten i kod kepstralne analize, pokazalo bi se da su moduli ta dva DFT-a identički jednaki, tj. da se ta dva signala razlikuju isključivo po fazi. Ovakav rezultat se i mogao očekivati obzirom da je kepstar $c_p(n)$ izračunat isključivo na osnovu modula DFT-a signala $x(n)$, dok je faza ignorirana.

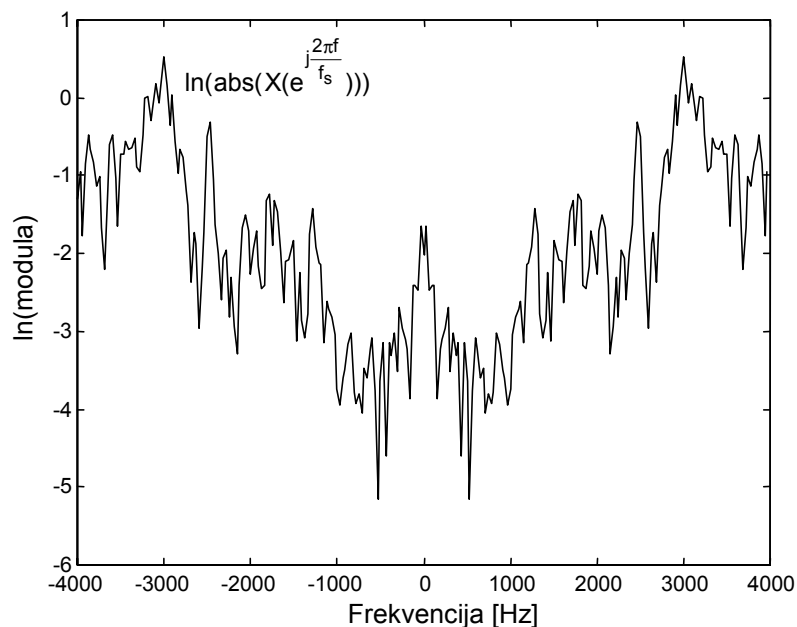


Slika 10.4-16 Konvolucija signala $e(n)$ i $h(n)$ i početni signal $x(n)$

Svi opisani postupci se mogu primijeniti i na bezvučni govorni segment. U slijedećem primjeru koristit će se glas 'š', kao primjer bezvučnog frikativa. Parametri kepralne analize identični su onima za glas 'e', tj. $f_s=8000\text{Hz}$, $N=256$ i Blackman-ov vremenski otvor. Izdvojeni govorni segment pomnožen s vremenskim otvorom prikazan je na slici 10.4-17.



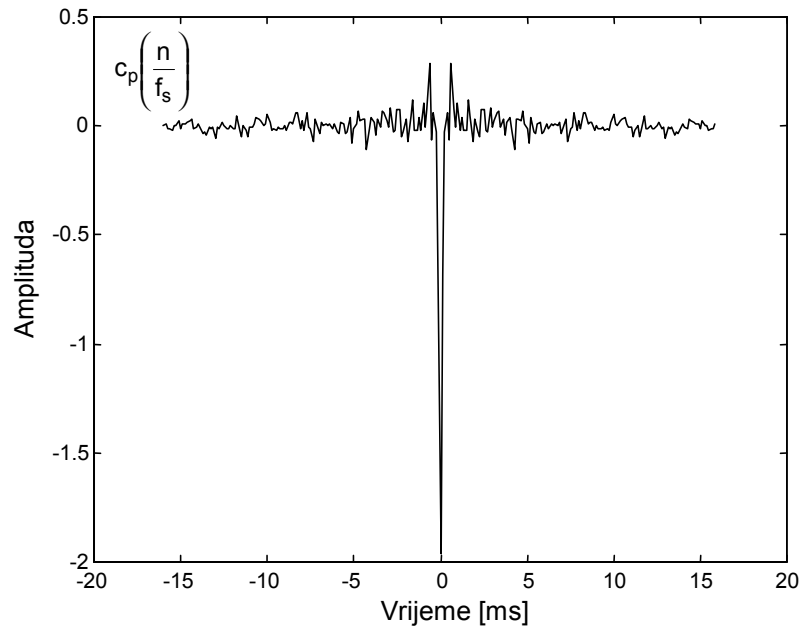
Slika 10.4-17 Bezvučni segment pomnožen s Blackman-ovim vremenskim otvorom



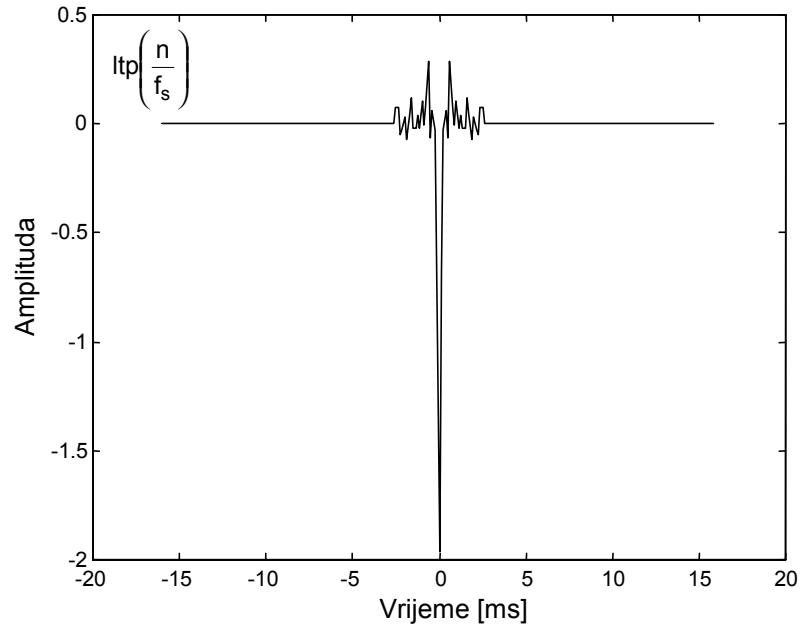
Slika 10.4-18 Prirodni logaritam modula DFT-a

Prirodni logaritam modula diskretne Fourier-ove transformacije signala $x(n)$ sa slike 10.4-17, prikazan je na slici 10.4-18. U ovom spektru je vidljiva formantna struktura, s značajnim formantom na frekvenciji 3000 Hz, i par slabijih formanata na nižim frekvencijama. Za razliku od zvučnog segmenta kod kojeg je 'fina' spektralna struktura bila pravilna i periodična, kod bezvučnog glasa 'š', brze oscilacije su nepravilne (šumovite) i

aperiodske. Zbog te činjenice u kepstru $c_p(n)$ prikazanom na slici 10.4-19 više ne postoje izraženi maksimumi osim onog za $n=0$. Taj nulti kepstralni koeficijent $c_p(0)$ nosi informaciju o energiji ulaznog signala, i nalazi se kao integral prirodnog logaritma modula spektra normiran sa 2π .

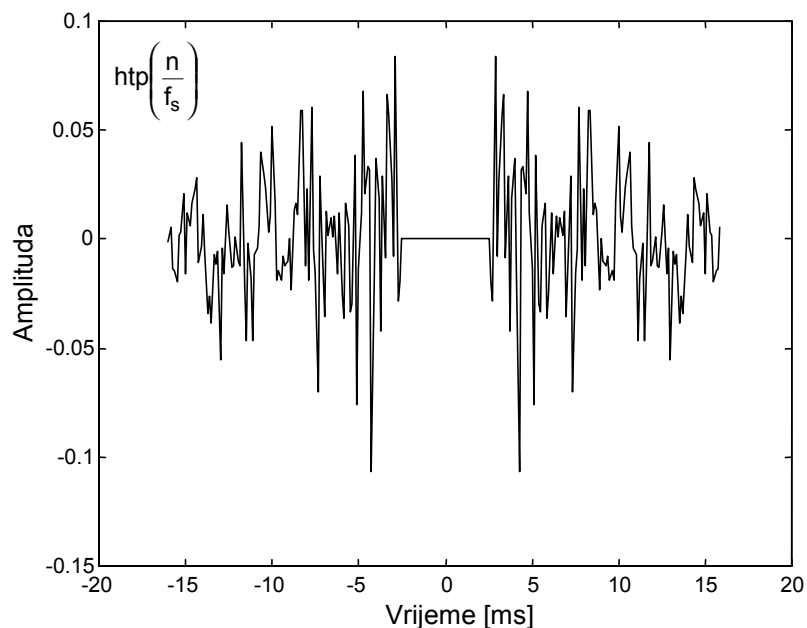


Slika 10.4-19 Kepstar bezvučnog glasa 'š'

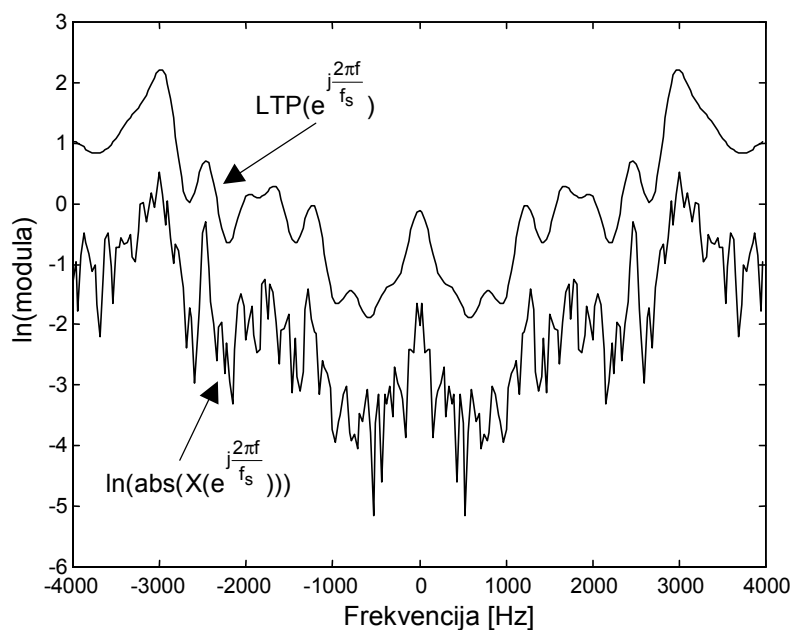


Slika 10.4-20 Nisko-vremenski dio kepstra (ltp) glasa "š"

Nakon razdvajanja na nisko-vremenski i visoko-vremenski dio operatorima $L_1[]$ i $L_2[]$ uz $n_0=20$, dobivaju se $ltp(n)$ i $htp(n)$ prikazani na slikama 10.4-20 i 10.4-21. Nisko-vremenski dio pokazuje brzo padajući karakter jednako kao i kod zvučnih glasova, dok se visoko-vremenski dio potpuno razlikuje, tj. pokazuje šumovit karakter bez ikakvih izraženih maksimuma.

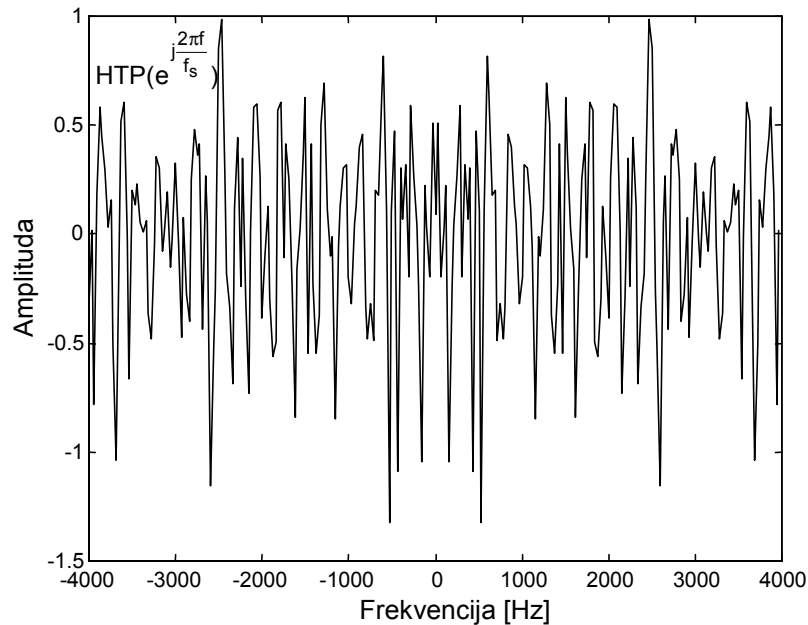


Slika 10.4-21 Visoko-vremenski dio kepra (htp) glasa "š"

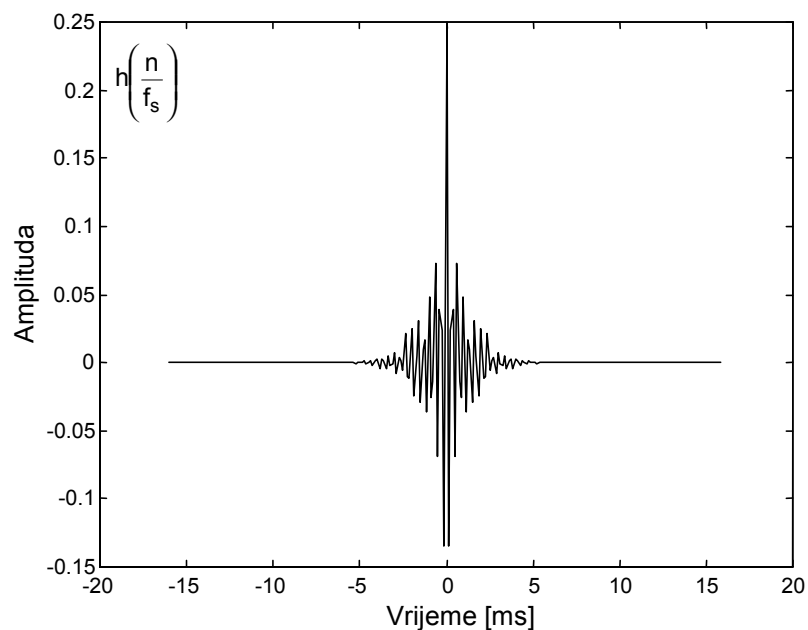


Slika 10.4-22 Modeliranje spektralne ovojnice ltp dijelom kepra

Primjenom diskretne Fourier-ove transformacije na $ltp(n)$ dobiva se spektralna ovojnica signala koja je prikazana na slici 10.4-22 zajedno s prirodnim logaritmom modula DFT-a signala $x(n)$ pomnoženog s vremenskim otvorom. Iz ovog primjera je vidljivo da ltp dio kepra i kod bezvučnih glasova dobro modelira spektralnu ovojnicu, tj. formantnu strukturu govornog signala. DFT visoko-vremenskog dijela $htp(n)$, prikazan je na slici 10.4-23 i ima oblik spektra bijelog šuma. To znači da su sve spektralne komponente sličnih amplituda, ali ne postoji nikakva pravilnost ili periodičnost. Potrebno je ponovno naglasiti, da u slučaju korištenja običnog kepra $c_p(n)$, koji je paran, sve komponente DFT-a htp i ltp dijela imaju nultu-fazu, tj. realne su. Ova činjenica imat će utjecaja na oblik pobudnog signala $e(n)$, koji se nalazi na osnovu spektra prikazanog na slici 10.4-23.

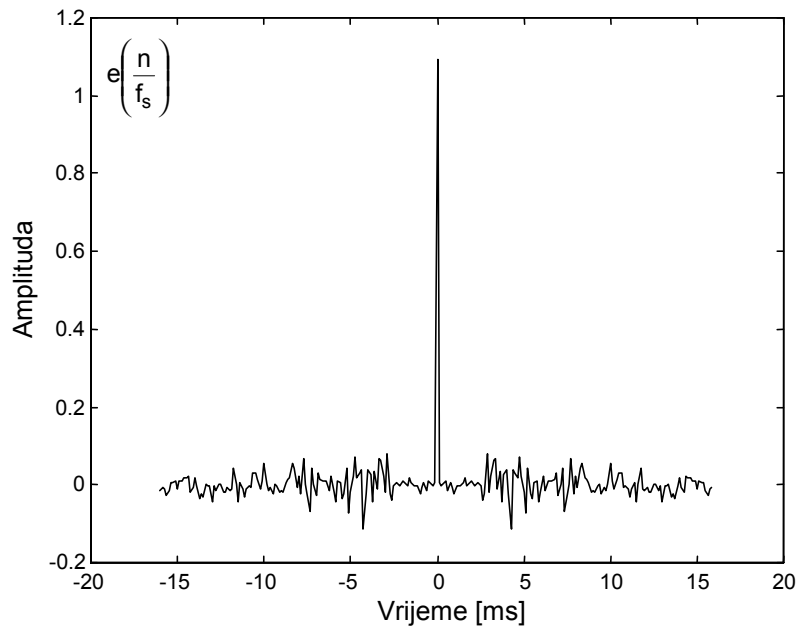


Slika 10.4-23 DFT htp dijela kepstra

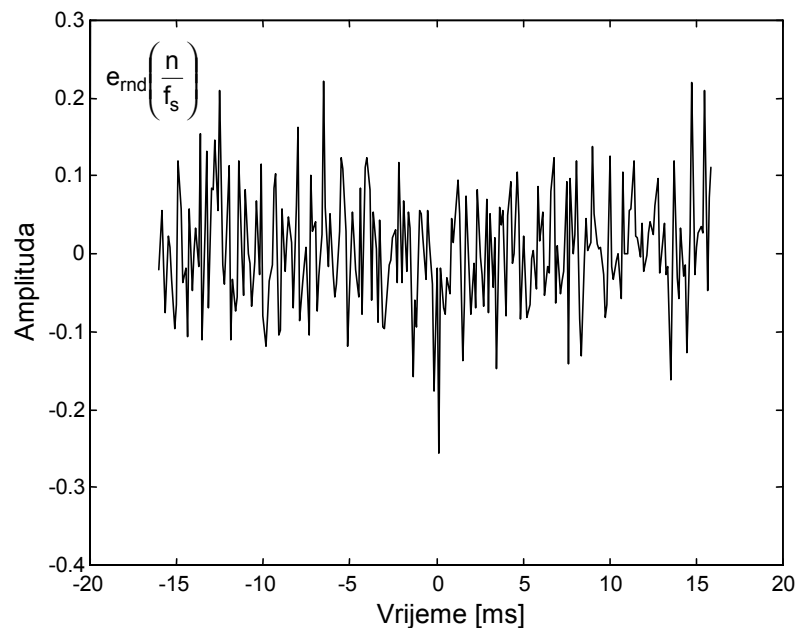
Slika 10.4-24 Impulsni odziv linearnog sustava, $h(n)$

U skladu s blok shemom obrade na slici 10.4-8, eksponenciranjem i inverznom diskretnom Fourier-ovom transformacijom DFT-a od ltp i htp dijela kepstra dobivaju se impulsni odziv linearnog sustava $h(n)$ prikazan na slici 10.4-24 i pobudni signal sustava $e(n)$ prikazan na slici 10.4-25. Impulsni odziv sustava $h(n)$ odgovara očekivanjima i u slučaju bezvučnog glasa obuhvaća samo impulsni odziv vokalnog trakta i impulsni odziv zračenja na usnicama. U njemu su vidljiva visoko-frekvencijska istitravanja vlastitim frekvencijama sustava, od kojih je najizraženija frekvencija na mjestu najznačajnijeg formanta (3000Hz). S druge strane, pobudni signal sustava $e(n)$ ne odgovara početnim pretpostavkama, tj. kod bezvučnih glasova korištena je pretpostavka da je pobudni signal sustava $u(n)$ u izrazu (10.4-12) slučajni šum, a ne jedinični impuls kao što je prikazano na slici 10.4-25. Do ove

pojave dolazi zbog toga što sve komponente DFT-a htp dijela imaju nultu-fazu, pa se idealno pribrajaju na vremenskom indeksu $n=0$ (koherentne su).



Slika 10.4-25 Pobudni signal linearnog sustava $e(n)$



Slika 10.4-26 Pobudni signal sa slučajnom fazom $e_{\text{rnd}}(n)$

Ovaj problem moguće je riješiti promjenom fazne karakteristike pobudnog signala. Kao što je već napomenuto kod primjera zvučnog glasa, u slučaju korištenja običnog kepstra konvolucija $e(n)*h(n)$ nije identički jednaka signalu $x(n)$, već su ta dva signala jednaka isključivo po modulu DFT-a. To drugim riječima znači da se faza svakog od ta tri signala, $e(n)$, $h(n)$ i $x(n)$, može po želji mijenjati, a da i dalje vrijedi jednakost modula DFT-a. U skladu s ovim pravilom, može se provesti modifikacija faze DFT-a htp dijela, koja bi rezultirala s pobudnim signalom koji je sličniji bijelom šumu.

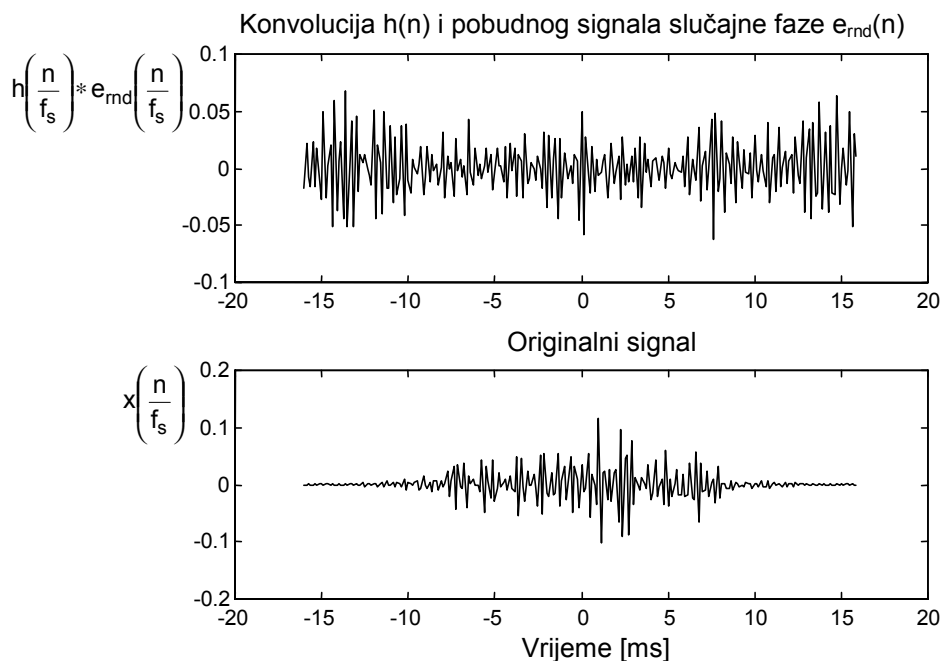
$$\begin{aligned} \text{HTP}(k) &= \text{HTP} \left(e^{j \frac{2\pi k}{N}} \right) = \sum_{n=0}^{N-1} \text{htp}(n) \cdot e^{-j \frac{2\pi nk}{N}} = \\ &= \sum_{n=0}^{N-1} c_p(n) \cdot l_2(n) \cdot e^{-j \frac{2\pi nk}{N}} = \sum_{n=n_0}^{N-n_0} c_p(n) \cdot e^{-j \frac{2\pi nk}{N}} \end{aligned} \quad (10.4-19)$$

$$e(n) = \frac{1}{N} \sum_{k=0}^{N-1} e^{\text{HTP}(k)} \cdot e^{j \frac{2\pi nk}{N}} \quad (10.4-20)$$

$$e_{\text{rnd}}(n) = \frac{1}{N} \sum_{k=0}^{N-1} e^{\text{HTP}(k)} \cdot e^{j\phi_{\text{rnd}}(k)} \cdot e^{j \frac{2\pi nk}{N}} \quad (10.4-21)$$

$$\phi_{\text{rnd}}(k) = \begin{cases} 0 & \text{za } k = 0 \\ 2\pi \text{rnd}(k) & \text{za } 0 < k < N/2 \\ 0 & \text{za } k = N/2 \\ -\phi_{\text{rnd}}(N-k) & \text{za } N/2 < k < N \end{cases} \quad (10.4-22)$$

Izrazi (10.4-19) i (10.4-20) opisuju postupak izračunavanja pobude $e(n)$ prikazane na slici 10.4-25. U izrazu (10.4-21) vidljiv je postupak modifikacije faze, tj. svakoj eksponenciranoj komponenti $\text{HTP}(k)$ dodaje se slučajni fazni pomak $\phi_{\text{rnd}}(k)$, koji će razbiti sinfaznost na nultom uzorku ($n=0$). Pošto izlazni signal $e_{\text{rnd}}(n)$ mora biti realan, potrebno je osigurati antisimetričnost faze oko $\omega=0$, kao što se to vidi iz izraza (10.4-22). Slučajna faza $\phi_{\text{rnd}}(k)$ se izračunava na osnovu niza od $N/2-1$ slučajnih brojeva $\text{rnd}(k)$ na intervalu $[0,1]$, s jednolikom razdiobom.



Slika 10.4-27 Konvolucija signala $e_{\text{rnd}}(n)$ i $h(n)$ i početni signal $x(n)$

Ovakav postupak modifikacije fazne karakteristike rezultira pobudnim signalom $e_{md}(n)$ prikazanim na slici 10.4-26, čiji je modul DFT-a identički jednak modulu DFT-a signala $e(n)$. Drugim riječima, iako se valni oblici ta dva signala značajno razlikuju, razlika je sadržana isključivo u faznom članu.

Na gornjoj polovici slike 10.4-27 prikazan je signal koji se dobiva konvolucijom modificiranog pobudnog signala $e_{md}(n)$ i impulsnog odziva sustava $h(n)$. Ovaj se signal po valnom obliku razlikuje od početnog signala $x(n)$ koji je prikazan u donjoj polovici iste slike, ali moduli DFT transformacija u N točaka ova dva signala su identički jednaki. Korištenje kompleksnog kepstra u svrhu očuvanja faznih odnosa kod zvučnih glasova može imati opravdanje, međutim kod bezvučnih glasova fazna karakteristika s perceptualnog stanovišta ne igra nikakvu ulogu. To znači da faza bezvučnog glasa može biti proizvoljno podešena, a da se taj glas i dalje čuje kao isti glas.

Teoretska analiza i primjeri dani u ovom poglavlju pokazali su da je primjenom homomorfne dekonvolucije govornog signala moguće odrediti aproksimacije pojedinih komponenti konvolucije, tj. rastaviti govorni signal na pobudu i impulsi odziv linearnog sustava. Međutim, kod većine primjena homomorfne obrade govora, ne traži se cjelokupni postupak dekonvolucije. Na primjer, u slučaju određivanja osnovne frekvencije titranja glasnica ili određivanja frekvencija formantata, dovoljno je provesti prvi dio postupka, tj. određivanje kepstra. Primjeri kepstra zvučnog i bezvučnog glasa prikazani na slikama 10.4-6 i 10.4-19 sugeriraju mogućnost određivanja informacije o zvučnosti govornog segmenta isključivo na osnovnu kepstra. Također na osnovu pozicije izraženih maksimuma u kepstru zvučnih glasova moguće je direktno odrediti frekvenciju titranja glasnica. Pošto ltp dio kepstra jednoznačno opisuje spektralnu ovojniciu izdvojenog govornog segmenta, tj. njegovu formantnu strukturu, moguće je odrediti frekvencije formantata postupcima traženja pozicija lokalnih maksimuma u DFT-u ltp dijela. U slijedećem poglavlju biti će opisani postupci određivanja osnovne frekvencije titranja glasnica korištenjem kepstra.

10.5 Primjena kepstralne analize za određivanje fundamentalne frekvencije

Kao što se moglo uočiti iz razmatranja u prethodnom poglavlju, homomorfna obrada predstavlja vrlo učinkovit postupak za određivanje osnovne frekvencije titranja glasnica (engl. *pitch detection*). Tako se u kepstru zvučnog govornog segmenta na indeksu koji odgovara osnovnom periodu titranja glasnica može uočiti šiljak, dok u kepstru bezvučnog govornog segmenta takav šiljak ne postoji. Upravo ovo svojstvo kepstra može se koristiti kao osnova za određivanje zvučnosti govornog segmenta kao i za određivanje osnovnog perioda zvučnih govornih segmenata.

Postupci kod procedure za određivanje osnovne frekvencije temeljene na kepstru relativno su jednostavni. Nakon izračunavanja kepstra na način objašnjen u prethodnom poglavlju, u okolini očekivanog osnovnog perioda u kepstru se traži šiljak. Ako je veličina šiljka iznad nekog prethodno postavljenog praga, ulazni govorni segment je vjerojatno zvučni, pri čemu pozicija šiljka predstavlja dobru estimaciju osnovnog perioda. Ako je veličina šiljka ispod nivoa, ulazni govorni segment je vjerojatno bezvučni. Promjena tipa pobude (zvučno / bezvučno) i promjena osnovnog perioda titranja glasnica kroz vrijeme može se procijeniti računanjem vremenski kratkotrajnog kepstra temeljenog na vremenski kratkotrajnoj Fourier-ovoj transformaciji. Tipično se keprar računa jednom svakih 10-20 ms pošto se parametri pobude u normalnom govoru ne mijenjaju naglo.

Ovi primjeri mogu navesti na pomisao da i jednostavan algoritam može rezultirati izuzetno dobrim estimacijama osnovne frekvencije i zvučnosti. Na žalost, kao što je čest slučaj u obradi govora, postoje brojni specijalni slučajevi i kompromisi o kojima treba voditi računa kod dizajna algoritama za određivanje osnovne frekvencije temeljenih na kepstru. Iako postoji čitav niz razvijenih algoritama koji se koriste s uspjehom, umjesto detaljnog opisa nekog od njih treba ipak radije naglasiti neke od osnovnih problema koji se javljaju tokom osmišljanja takvog algoritma.

Kao prvo, prisutnost snažnog šiljka u kepstru u rasponu od 3 do 20 ms je vrlo jaka indikacija da je govorni segment zvučan. Međutim, odsutnost šiljka ili postojanje šiljka malog iznosa ne mora nužno značiti da je govorni segment bezzvučan. Drugim riječima, veličina pa čak i samo postojanje šiljka u kepstru za zvučni segment govora ovisi o nizu faktora, uključujući dužinu vremenskog otvora koji se primjenjuje na ulaznom signalu te formantnu strukturu ulaznog signala. Može se pokazati da je najveći iznos šiljka jednak jedinici, a postiže se samo u slučaju da se osnovni period ni malo ne mijenja tokom cijelog segmenta. Međutim, u normalnom govoru se tako nešto teško može dogoditi čak i u slučaju da se pravokutnim otvorom zahvati cjeli broj perioda. Pravokutni vremenski otvor se inače u praksi rijetko koristi zbog slabijih rezultata estimacije, dok je za slučaj Hamming-ovog otvora jasno da i širina otvora kao i relativna pozicija otvora prema signalu značajno utječu na veličinu šiljka u kepstru. Kao primjer za granični slučaj, neka je vremenski otvor kraći od dva osnovna perioda. Jasno je da u ovom slučaju ne bi trebalo očekivati bilo kakvu izraženu periodičnost bilo u spektru bilo u kepstru. Zbog toga se u praksi dužina vremenskog otvora obično uzima tako da su unutar govornog segmenta ograničenog otvorom nalaze barem dvije pune periode, uzimajući u obzir i opadanje (engl. *tapering*) vremenskog otvora prema rubovima. Tako se za obradu govora muškog govornika dubljeg glasa tipično uzima vremenski otvor trajanja oko 40 ms, dok se za više glasove (veća osnovna frekvencija) mogu koristiti proporcionalno užu vremenski otvori. Poželjno je, naravno, zadržati trajanje otvora što kraćim da bi se smanjili utjecaji promjene parametara govora unutar intervala koji se analizira. Što je vremenski otvor duži, veće su promjene parametara od početka do kraja intervala, pa je veće i odstupanje od početno pretpostavljenog stacionarnog modela signala za analizu. Jedan od pristupa određivanju dužine otvora (da ne bude niti prekratak niti predugačak) jest prilagodba veličine otvora tokom obrade na temelju prethodnih (ili usrednjenih) estimacija osnovne frekvencije.

Drugi slučaj kod kojeg signal može značajno odstupati od modela jest kada je signal iznimno frekvencijski ograničen. Granični slučaj za to jest primjer čistog sinusnog signala. Logaritam spektra takvog signala rezultira samo jednim šiljkom pa pošto u njemu nema periodičkih oscilacija, neće biti ni šiljka u kepstru. U realnom govoru su npr. zvučni praskavi glasovi u principu jako frekvencijski ograničeni, bez jasno definirane harmonijske strukture na frekvencijama iznad nekoliko stotina Hz. U tim slučajevima niti nema izraženog šiljka u kepstru. Na sreću, za sve slučajeve osim onih s najmanjim osnovnim periodom, ipak se pojavljuju šiljci u kepstru, ali na mjestima gdje su sve druge komponente gotovo nestale, pa se za određivanje šiljaka može koristiti relativno nizak prag detekcije (reda veličine 0.1).

U većini slučajeva se uz adekvatnu dužinu vremenskog otvora osnovna frekvencija i zvučnost mogu relativno pouzdano odrediti na temelju pozicije i amplitude šiljaka kepstra. U preostalim slučajevima kada kepstar ne daje jasne informacije, pouzdanost određivanja osnovne frekvencije i zvučnosti može se povećati korištenjem dodatnih informacija kao npr. prolaza kroz nulu, energije i sl., ili uz forsiranje blažih promjena estimacija. Dodatna logika potrebna za obradu specijalnih slučajeva često zahtjeva značajnu količinu koda u programskim implementacijama, ali se s obzirom na dobivenu korist isplati.