

# Digitalna obrada govora

---

**Dunić, Ante**

**Undergraduate thesis / Završni rad**

**2023**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Rijeka, Faculty of Maritime Studies, Rijeka / Sveučilište u Rijeci, Pomorski fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:187:277403>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2025-03-27**



**Sveučilište u Rijeci, Pomorski fakultet**  
University of Rijeka, Faculty of Maritime Studies

*Repository / Repozitorij:*

[Repository of the University of Rijeka, Faculty of Maritime Studies - FMSRI Repository](#)



**uniri** DIGITALNA  
KNJIŽNICA



SVEUČILIŠTE U RIJECI  
POMORSKI FAKULTET

**ANTE DUNIĆ**

**DIGITALNA OBRADA GOVORA**

**ZAVRŠNI RAD**

Rijeka, 2023.

SVEUČILIŠTE U RIJECI  
POMORSKI FAKULTET

**DIGITALNA OBRADA GOVORA**  
**DIGITAL SPEECH PROCESSING**

**ZAVRŠNI RAD**

Kolegij: Osnove elektroničkih komunikacija

Mentor: dr.sc. Zoran Mrak

Student: Ante Dunić

Studijski program: Elektroničke i informatičke tehnologije u pomorstvu

JMBAG: 0112065495

**Rijeka, rujan 2023.**

Student: Ante Dunić

Studijski program: Elektroničke i informatičke tehnologije u pomorstvu

JMBAG: 0112065495

## IZJAVA O SAMOSTALNOJ IZRADI ZAVRŠNOG RADA

Kojom izjavljujem da sam završni rad s naslovom:

DIGITALNA OBRADA GOVORA

izradio samostalno pod mentorstvom

dr. sc. Zoran Mrak

U radu sam primijenio metodologiju izrade stručnog/znanstvenog rada i koristio literaturu koja je navedena na kraju završnog rada. Tude spoznaje, stavove, zaključke, teorije i zakonitosti koje sam izravno ili parafrazirajući naveo u završnem radu na uobičajen, standardan način citirao sam i povezao s fusnotama i korištenim bibliografskim jedinicama, te nijedan dio rada ne krši bilo čija autorska prava. Rad je pisan u duhu hrvatskog jezika.

Student

Ante Dunić

Ime i prezime studenta

ANTE DUNIĆ

Student: Ante Dunić

Studijski program: Elektroničke i informatičke tehnologije u pomorstvu

JMBAG:0112065495

IZJAVA STUDENTA – AUTORA  
O JAVNOJ OBJAVI OBRANJENOG ZAVRŠNOG RADA

Izjavljujem da kao student – autor završnog rada dozvoljavam Pomorskom fakultetu Sveučilišta u Rijeci da ga trajno javno objavi i besplatno učini dostupnim javnosti u cjelovitom tekstu u mrežnom digitalnom repozitoriju Pomorskog fakulteta.

U svrhu podržavanja otvorenog pristupa završnim radovima trajno objavljenim u javno dostupnom digitalnom repozitoriju Pomorskog fakulteta, ovom izjavom dajem neisključivo imovinsko parvo iskorištavanja bez sadržajnog, vremenskog i prosternog ograničenja mog završnog rada kao autorskog djela pod uvjetima *Creative Commons* licencije CC BY Imenovanje, prema opisu dostupnom na <http://creativecommons.org/licenses/>.

Student – autor

  
\_\_\_\_\_

## SAŽETAK

Ovaj završni rad bavi se proučavanjem metoda i tehnika digitalne obrade govora te prikaza govornih signala. Opisane su Z-transformacije i Fourierove transformacije kao tehnike analize govornih signala kojima se određuju određene karakteristike tog signala, te njihov grafički prikaz u vidu spektrogramске analize glasovnog signala. Obradene su jednostavnije metode digitalnog enkodiranja poput PCM-a, DPCM-a i ADPCM-a. Prikazan je i razvoj LPC-a i CELP-a kao tehnologija za rekreiranje i kompresiju govornog signala. Za sam kraj obradene su moderne tehnike digitalne obrade govora poput VoIP-a, HMM-a i ANN-ova te njihova uloga u svakidašnjem životu ljudi.

Ključne riječi: analiza govora, digitalna obrada govora, govorni signali

## SUMMARY

This thesis studies the methods and techniques of digital speech signal processing and representation of speech signals. Z-transformations and Fourier's transformations are described as techniques for analysis of speech signals that determine specific characteristics of those signals, and allow for a graphical representation in form of a spectrogram. Then it elaborates elementary methods of digital speech encoding such as PCM, DPCM and ADPCM. Furthermore, LPC and CELP are represented as technologies for recreating and compressing of speech signals. Lastly, VoIP, HMM and ANNs are depicted as modern techniques for digital speech processing and their role in everyday human lives is described.

Keywords: speech analysis, digital speech processing, speech signals

# SADRŽAJ

<b>SAŽETAK .....</b>	<b>I</b>
<b>SUMMARY .....</b>	<b>I</b>
<b>SADRŽAJ .....</b>	<b>II</b>
<b>1. UVOD.....</b>	<b>1</b>
<b>2. GOVOR I NAŠA PERCEPCIJA.....</b>	<b>3</b>
2.1. NASTAJANJE GOVORA .....	3
2.2. KARAKTERISTIKE GOVORNIH SIGNALA.....	4
2.3. PERCEPCIJA GOVORA.....	4
<b>3. TEMELJNE TEHNIKE ANALIZE GOVORNIH SIGNALA.....</b>	<b>5</b>
3.1. Z-TRANSFORMACIJA .....	5
3.2. INVERZNA Z-TRANSFORMACIJA .....	5
3.3. DISKRETNA FOURIEROVA TRANSFORMACIJA.....	6
3.4. BRZA FOURIEROVA TRANSFORMACIJA.....	6
<b>4. PRIKAZ GOVORNIH SIGNALA .....</b>	<b>8</b>
4.1. PRIKAZ SIGNALA U VREMENSKOJ DOMENI .....	8
4.2. PRIKAZ SIGNALA U FREKVENCIJSKOJ DOMENI .....	9
4.3. SPEKTROGRAMSKA ANALIZA .....	9
<b>5. METODE DIGITALNE OBRADJE GOVORA.....</b>	<b>12</b>
5.1. PULSNO-KODNA MODULACIJA (PCM) .....	12
5.1.1. Diferencijalna pulsno-kodna modulacija (DPCM).....	16
5.1.2. Adaptivno-diferencijalna pulsno-kodna modulacija (ADPCM).....	16
5.2. LINEARNO PREDIKTIVNO KODIRANJE (LPC) .....	16
5.2.1. Kodno-pobuđena linearna predikcija (CELP) .....	18

<b>6. PRIMJENA DIGITALNE OBRADNE GOVORNIH SIGNALA .....</b>	<b>21</b>
6.1. GOVOR PUTEV INTERNET PROTOKOLA (VOIP) .....	21
6.2. SKRIVENE MARKOVLJEVE MREŽE (HMM) .....	22
6.3. UMJETNE NEURALNE MREŽE.....	23
6.4. DRAGON DICTATE.....	24
<b>7. ZAKLJUČAK .....</b>	<b>25</b>
<b>LITERATURA.....</b>	<b>26</b>
<b>KAZALO KRATICA .....</b>	<b>27</b>
<b>POPIS ILUSTRACIJA.....</b>	<b>28</b>



## 1. UVOD

Osnova komunikacije ljudskog društva te ono što ga uvelike razlikuje od ostatka živoga svijeta jest govor. Mozak govornika vokalnom traktu šalje signale kako bi proizveo zvučne valove, glasove, stvarajući riječi te prenoseći željenu poruku. Slušatelj te valove prima svojim osjetilom sluha, pretvarajući ih u električni impuls koji se putem slušnog živca šalje do mozga koji primljenu poruku obrađuje i daje joj smisao. Potreba za govornom komunikacijom i prijenosom informacija govorom dovela je do razvoja raznih tehnologija za zapis, obradu i prijenos govornih signala.

Temeljne tehnike analize govornih signala jesu Z-transformacija, inverzna Z-transformacija, diskretna Fourierova transformacija (engl. *Discrete Fourier Transform*, DFT) te brza Fourierova transformacija (engl. *Fast Fourier Transform*, FFT). Z-transformacija omogućava analizu i dizajniranje diskretnih sustava, dok DFT koristeći FFT metodu za izračun služi za analizu frekvencijske domene diskretnih signala. Analiza signala omogućuje grafički prikaz istih u vremenskoj i/ili frekvencijskoj domeni, odnosno do spektrogramске analize govornih signala. Na taj način se mogu odrediti određene karakteristike tog signala poput visine glasa, šumova ili harmonika, te postojanje obrazaca unutar govornog signala.

Osim spektrogramске analize, tehnike analize govornih signala omogućuju i njihovu digitalnu obradu različitim metodama. Pulsno-kodna modulacija (engl. *Pulse Code Modulation*, PCM) digitalno enkodira analogni signal uzorkovanjem i kvantizacijom u diskretnim intervalima. Diferencijalna pulsno-kodna modulacija (engl. *Differential Pulse Code Modulation*, DPCM) enkodira signal koristeći razliku među vrijednostima uzastopnih uzoraka, smanjujući redundanciju podataka i potrebe za podatkovnom pohranom u odnosu na PCM. Adaptivno-diferencijalna pulsno-kodna modulacija (engl. *Adaptive Differential Pulse Code Modulation*, ADPCM) poboljšava brzinu prijenosa bez ugrožavanja kvalitete prenesenog signala prilagođavajući faze predikcije i kvantizacije samim karakteristikama govora.

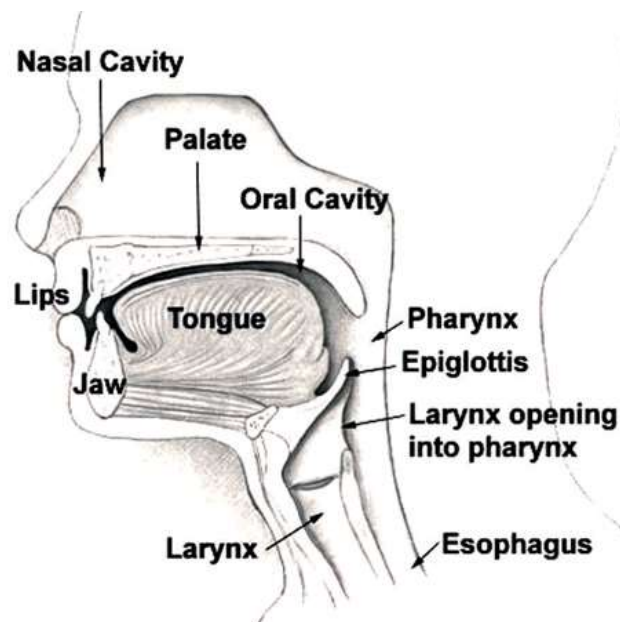
Linearno prediktivno kodiranje (engl. *Linear Predictive Coding*, LPC) metoda je obrade govornih i audio signala koja kreira izlazni signal analizirajući karakteristike govornog signala te predviđanjem određuje parametre linearnog prediktora. Kodno-pobuđena linearna predikcija (engl. *Code Excited Linear Prediction*, CELP) enkodira govorne signale odabirući predefinirane ekscitacijske signale iz kodnih knjiga i prilagođavajući parametre filtera da što vjernije odgovaraju originalnom signalu, omogućavajući efikasnu kompresiju signala uz dobro očuvanu kvalitetu.

Do danas, metode digitalne obrade govora poprilično su se razvile, revolucionarizirajući neke životne sfere, a posebno telekomunikacije. Govor putem Internet protokola (engl. *Voice over Internet Protocol*, VoIP) tehnologija je koja omogućuje prijenos govornih signala putem Interneta koristeći kodiranje govornih signala za kompresiju. Skrivena Markovljeva mreža i umjetne neuronske mreže modeli su koji se, između ostalog, koriste za sintezu i prepoznavanje ljudskog govora. Skrivena Markovljeva mreža (engl. *Hidden Markov Models*, HMM) statistički su modeli koji prikazuju parametre govora kao sekvence sa skrivenim stanjima koji generiraju prepoznatljive karakteristike govora. Umjetne neuronske mreže (engl. *Artificial Neural Networks*, ANN) računalni su modeli inspirirani strukturom mreže neurona ljudskog mozga koji kroz prilagođavanje koeficijenta težine tijekom procesa učenja omogućavaju analizu i obradu govornih signala.

## 2. GOVOR I NAŠA PERCEPCIJA

### 2.1. NASTAJANJE GOVORA

Proces nastanka ljudskog govora započinje kognitivnom namjerom te kulminira proizvodnjom zvučnih valova. Prvenstveno, mozak odlučuje što će iskomunicirati samim time uključujući ideju, formulaciju te jezik kojim će se koristiti pri toj radnji. Respiratorni sustav zatim, djelujući kao glavni izvor napajanja samog govora, podupire prolaz zraka iz pluća stvarajući strujanje koje se diže do naših glasnica. Glasnice potom vibriraju pri kontaktu sa strujanjem zraka proizvodeći fonaciju ili bolje rečeno, vibraciju [1]. Proizvedeni zvuk potom nastavlja svoj put duž ždrijela te dolazi do rezonantne kutije koju u glavnini tvore usna i ždrijelna šupljina, te ponekad i nosna šupljina. Mijenjajući svoj oblik, veličinu i napetost stijenke uslijed kretnji grkljana, jezika, nepca, usnica i zuba potaknutih koordiniranim radom brojnih mišićnih skupina, dolazi do oblikovanja fonema, najmanjih glasovnih jedinica svakog jezika. Samoglasnici (vokali) primarno nastaju u grkljanu titranjem glasnica, dok položaj usnica, zuba i jezika određuju koji je to karakteristični vokal. Kod tvorbe suglasnika (konsonanata) zrak slobodno prolazi kroz grkljan te glasnice ne titraju. Oni nastaju u rezonantnoj kutiji u kojoj jezik priječi put prolasku zraka na različitim lokacijama (grlo, zubi, usne, nepce...) te na taj način stvara različite suglasnike. [1]



**Slika 1. Anatomija ljudskog vokalnog trakta**

Izvor: <https://training.seer.cancer.gov/head-neck/anatomy/overview.html>

## 2.2. KARAKTERISTIKE GOVORNIH SIGNALA

Govorni signali su složeni akustični valovi koji se promatraju i odvijaju u vremenskoj domeni s oscilacijama amplitude i frekvencije koje zatim pridodaju ritmičku karakteristiku govoru svakog zasebnog pojedinca. Ritmika govora može sadržavati niz čimbenika poput visine, duljine te intenziteta govora. Kada glasnice vibriraju periodično počinju proizvoditi harmonike. Tokom prolaska navedenih harmonika kroz vokalni trakt, specifične frekvencije su pojačane rezultirajući vrhovima koje nazivamo formantima koji su izrazito bitni za razlikovanje zvukova samoglasnika. Veoma bitno svojstvo je temeljna frekvencija ( $f_0$ ), primarna frekvencija na kojoj glasnice vibriraju koja je relativna percepciji visine glasa pojedinca.

## 2.3. PERCEPCIJA GOVORA

Opažanje govora složeni je auditorni proces. Zvučni valovi prvenstveno ulaze kroz uho slušatelja putujući kroz zvučni kanal ne bi li naposljetku stigli do bubnjića. Unutar srednjeg uha, vibracije bubnjića amplificiraju tri ušne košćice: čekić, nakovanj i stremen. Pojačane vibracije su zatim prosljeđene pužnici, organu ispunjenom tekućinom u unutarnjem uhu. Unutar pužnice trepetljikave stanice se kreću u skladu sa oscilacijama tekućine proizvodeći električne signale. Električni signali zatim putuju preko auditornih živaca do mozga u cilju dolaska do slušne kore gdje su ti signali dekodirani vodeći do percepcije i tumačenja govora. Nadalje, tumačenje jezika nije isključivo temeljeno na samom zvučnom valu. Naime, čimbenici poput znanja, konteksta te predviđanja posjeduju određenu težinu u tom istom procesu [2]. Primjerice čak i u prisustvu buke, naš mozak može razaznati riječi te premostiti praznine u zvuku oslanjajući se na kontekst i prijašnje znanje.

## 3. TEMELJNE TEHNIKE ANALIZE GOVORNIH SIGNALA

### 3.1. Z-TRANSFORMACIJA

Z-transformacija je temeljni matematički alat u domeni digitalne obrade govora služeći kao most između vremenske i frekvencijske domene prikaza govornih signala i sustava. Navedena pretvorba igra ključnu ulogu u analizi i izvedbi digitalnih sustava pružajući uvid u njihova svojstva i obrasce.

Jezgreno, Z-transformacija je diskretni pandan Laplaceovoj transformaciji korištenoj u obradi kontinuiranih vremenskih signala [3]. Pretvara niz diskretnih vremenskih signala u složenu funkciju kompleksne varijable najčešće označene sa  $Z$ . Za dani diskretni vremenski signal  $x(n)$ , Z-transformacija tog signala  $X(z)$  definirana je sljedećom formulom:

$$X(z) = \sum_{n=-\infty}^{\infty} x[n] z^{-n}$$

Bitno je napomenuti domenu konvergencije za Z-pretvorbu koja definira raspon  $z$  vrijednosti za koje Z-transformacija konvergira. Domena konvergencije nudi kritične informacije o stabilnosti vezanog sustava.

### 3.2. INVERZNA Z-TRANSFORMACIJA

Inverzna Z-transformacija omogućava rekonstrukciju originalnog diskretnog vremenskog signala na temelju njegove Z-transformacije. Uključuje računski utvrđenu sumu Z-pretvor pomnoženu sa eksponentima  $z$  varijable za različite vrijednosti  $n$  varijable. Matematički ona je definirana kao sljedeći izraz:

$$x[n] = \frac{1}{2\pi j} \oint_C X(z) z^{n-1} dz$$

gdje je integral procijenjen duž zatvorene krivulje  $C$  unutar domene konvergencije, a varijabla  $j$  predstavlja imaginarni broj. Z-transformacija je široko primjenjiv matematički koncept koji potkuje analizu i izvedbu digitalnih sustava obrade signala. Svojstvo neprimjetnog sjedinjenja prikaza vremenske i frekvencijske domene dopušta inženjerima i istraživačima detaljniji uvid u ponašanje sustava, njegovu stabilnost te frekvencijski odaziv istoga. [3]

### 3.3. DISKRETNA FOURIEROVA TRANSFORMACIJA

Diskretna Fourierova transformacija (*engl. Discrete Fourier Transform*), u daljnjem tekstu DFT, je temeljna tehnika digitalne obrade signala služeći kao podoban alat za analizu i razumijevanje frekvencijskog sadržaja diskretnih vremenskih signala. DFT pruža diskretni pandan na kontinuiranu Fourierovu transformaciju dozvoljavajući rastavljanje diskretnog signala na njegove sastavne sinusoidalne komponente. DFT pretvara niz  $N$  kompleksnih brojeva  $x[n]$  u drugi niz  $N$  kompleksnih brojeva  $X[k]$  [4]. Matematički, DFT je definiran sa sljedećom formulom:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\left(\frac{2\pi}{N}\right)kn}$$

DFT dijeli duboku povezanost sa kontinuiranom Fourierovom transformacijom budući da obje tehnike raščlanjuju frekvencijske komponente prisutne u jednom signalu. Dok se kontinuirana Fourierova transformacija izvodi na kontinuiranim signalima, DFT diskretizira ujedno i signal i njegovu frekvencijsku domenu čineći ovaj oblik transformacije pogodnom za digitalnu analizu i izračun. Izračun DFT-a uključuje procjenu prijašnje navedene sume za svaku vrijednost  $k$  rezultirajući nizom uzoraka frekvencijske domene  $X[k]$ . Niz algoritama poput brze Fourierove transformacije učinkovito izračunavaju DFT koristeći simetrije i redukcije samog računalnog procesa. Brza Fourierova transformacija znatno smanjuje potrebu za računalnim resursima čineći ju pogodnom za korištenje u stvarnom vremenu. DFT je temeljni alat ka razumijevanju frekvencijskih karakteristika signala. Pretvorba signala iz vremenske u frekvencijsku domenu omogućuje identifikaciju najzastupljenijih frekvencija, harmonika i šumova.

### 3.4. BRZA FOURIEROVA TRANSFORMACIJA

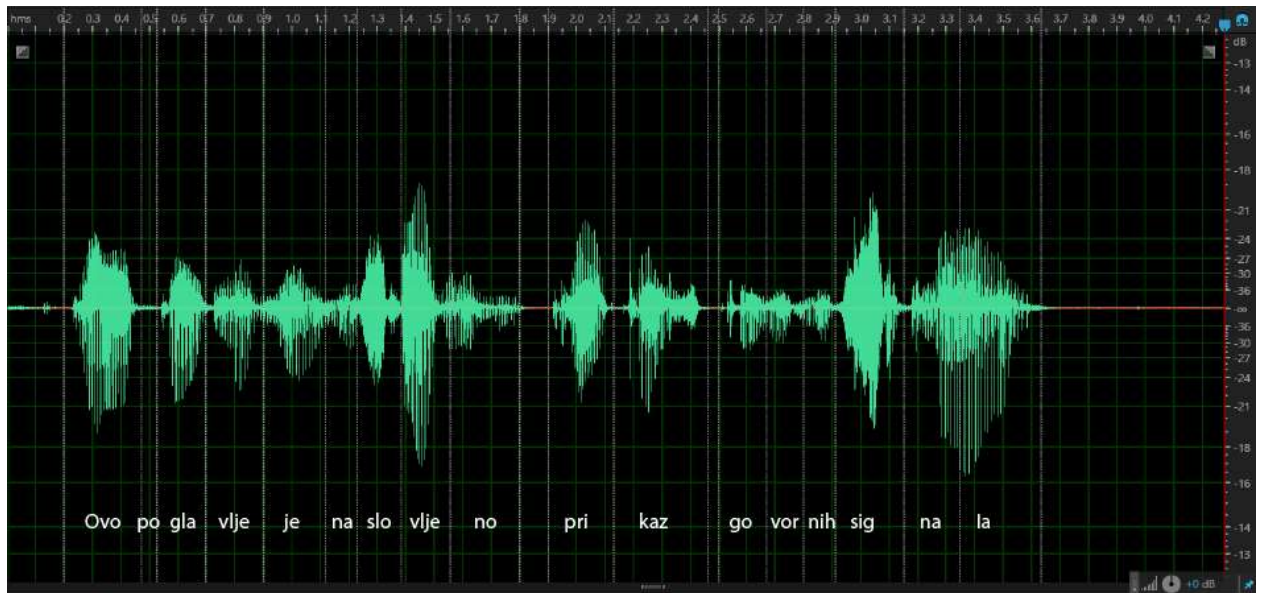
Brza Fourierova transformacija (*engl. Fast Fourier Transform*), u daljnjem tekstu FFT, unaprijeđeni je izračun DFT-a koji pruža učinkovito sredstvo za analizu frekvencijskih komponenti signala [5]. Ideja iza razvoja FFT-a rodila se iz potrebe računalnih resursa za brži izračun DFT-a, a rezultat je značajno smanjenje potrebnih aritmetičkih operacija za izračun DFT-a. U spektralnoj analizi, FFT dopušta izoliranje frekvencijske informacije signala rastavljajući ih na sastavne komponente te pomažući u zadacima identifikacije prevalentnih frekvencija i uočavanju periodičnih obrazaca. Najpoznatiji primjer FFT-a je Turkey Cooley algoritam koji

funkcionira na temelju razdvajanja bilo kojeg DFT skupa  $N$  na dva skupa  $N_1$  i  $N_2$  pri transformaciji signala. Ovakvim postupkom se postiže smanjenje potrebnih aritmetičkih operacija za jedan signal i dopušta bržu provedbu DFT transformacije. [4]

## 4. PRIKAZ GOVORNIH SIGNALA

### 4.1. PRIKAZ SIGNALA U VREMENSKOJ DOMENI

Govor kao signal ovisan o vremenu ima zvučni zapis koji odražava kontinuirane promjene u pritisku zraka pri prolasku vremena. Svaki izgovor, pauza, predah ili ton glasa koji je proizveden se može promatrati kao jedinstveni valni uzorak, samim time pretvarajući vremensku domenu istoga u statičku sliku vokalnog iskaza. Valni oblik karakteriziran je svojom amplitudom i trajanjem; vršni iznosi amplitude prikazuju visoko-energetska, odnosno glasna područja poput samoglasnika, dok niski iznosi amplitude označavaju suglasnike ili pauze. Bilo da se pokušava otkriti početne i krajnje točke izgovorene riječi, analizirati ritam i kadencu govora ili pak profilirati buku ili šumove, prikaz u vremenskoj domeni nudi podroban prikaz izvornog valnog oblika signala.



**Slika 2. Prikaz govornog signala fraze “Ovo poglavlje je naslovljeno prikaz govornih signala” u vremenskoj domeni**

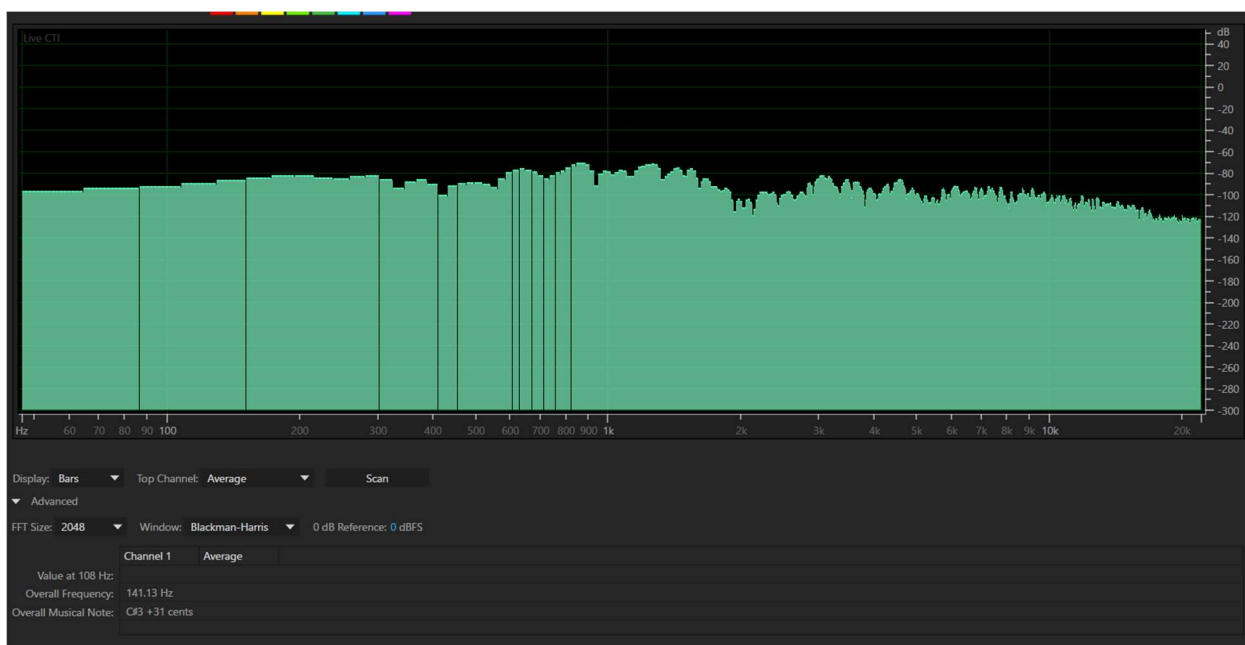
Izvor: Izradio student pomoću softverskog programa Adobe Audition

Na priloženom primjeru može se vidjeti zapis autorovog glasa za navedenu frazu. Ordinata amplitude predstavlja jačinu glasa u decibelima, a apscisa vrijeme sa trajanjem govornog signala od 3.9 sekundi.



## 4.2. PRIKAZ SIGNALA U FREKVENCIJSKOJ DOMENI

Kada se amplitude frekvencija govornog signala grafički prikažu u odnosu na njihove odgovarajuće frekvencije, dobije se spektralni profil valnog oblika koji se može smatrati jedinstvenim otiskom svakog zasebnog zvuka govora. Taj prikaz pruža neosporivo bitan uvid u karakteristike govora i njegovih obrazaca [6]. Koristeći računalne metode, posebice FTT, spektralna analiza izvedena je učinkovitije i brže čineći proces analize u stvarnom vremenu mogućim. U ovom primjeru je prikazana frekvencijska domena iste fraze sa Slike 2. Razlika između dvije je što os apscise u Slici 3. predstavlja frekvenciju. Iz priloženog se vidi da je prevladavajuća frekvencija 137.77 Hz odgovarajući noti C#3.



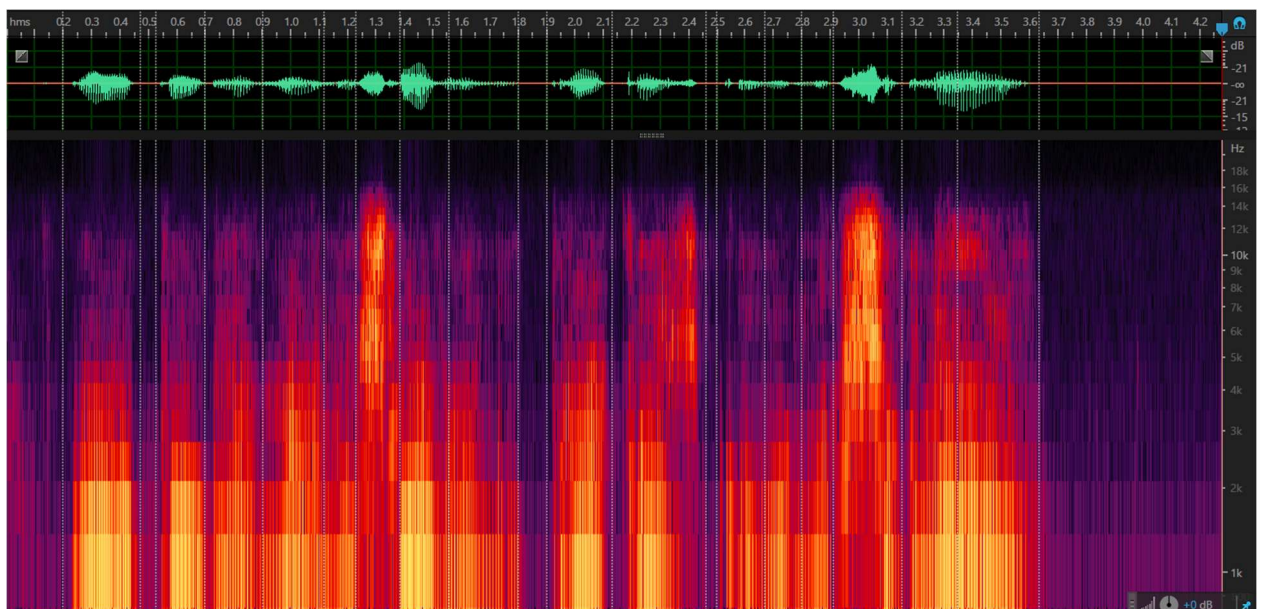
**Slika 3. Prikaz govornog signala fraze “Ovo poglavlje je naslovljeno prikaz govornih signala” u frekvencijskoj domeni**

Izvor: Izradio student pomoću softverskog programa Adobe Audition

## 4.3. SPEKTROGRAMSKA ANALIZA

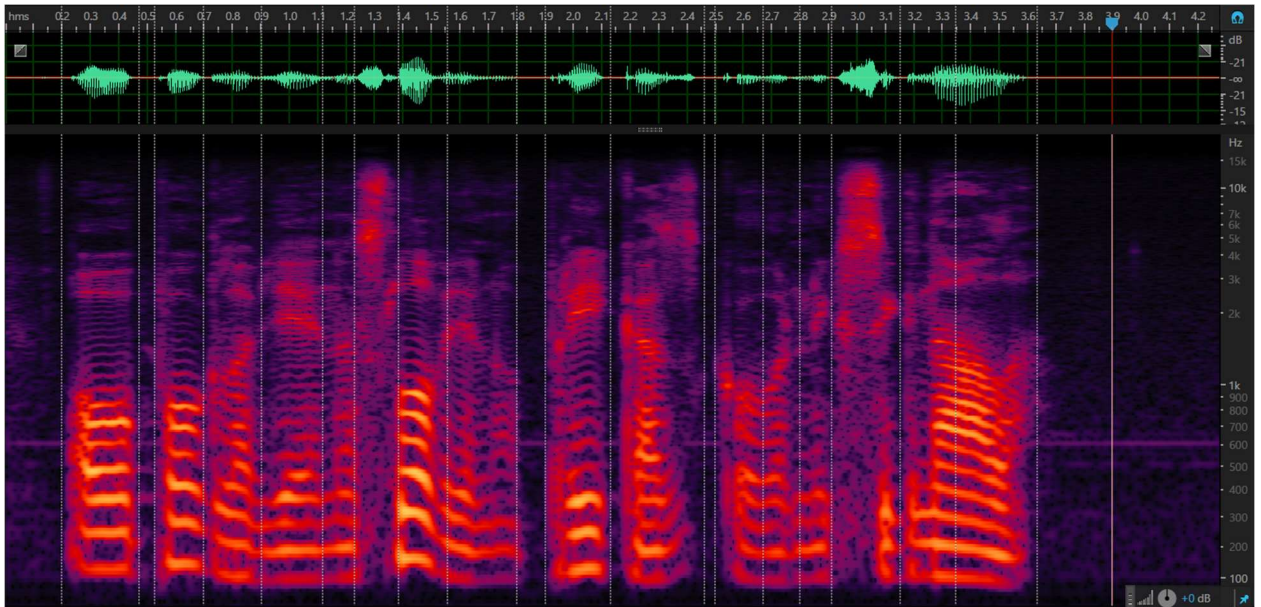
Prethodna potpoglavlja koja su bila fokusirana na razumijevanje signala primarno sa stajališta vremenske ili primarno frekvencijske domene služe kao temelj za razumijevanje spektrogramске analize, područja koje premošćuje jaz između prijašnje opisanih domena pružajući vremensko-frekvencijski prikaz signala. Spektrogram je grafički prikaz spektralnog sastava jednog signala u ovisnosti o vremenu. Spektrogramska analiza značajno se oslanja na metodi kratkotrajne

Fourierove transformacije (*engl. Short-Time Fourier Transform*), u daljnjem tekstu STFT. Koncept je relativno jednostavan: metodom uzorkovanja signal se podijeli u segmente kraćeg iznosa vremena te se analizira frekvencijski sadržaj svakog zasebnog elementa. Lokalni spektri tih segmenata posloženi jedan pored drugoga formiraju spektrogram. STFT i sam spektrogram prikazuju lokalne frekvencijske karakteristike signala pružajući detaljizirani pogled promjene tih karakteristika kroz vrijeme. Ni STFT nije bez manjkavosti, a najznačajnija je problem rezolucije u vremensko-frekvencijskoj domeni. Kraći vremenski segmenti mogu prikazati nagle promjene u frekvenciji, ali istovremeno i propustiti postojanje obrasca u dužim vremenskim segmentima. U kontrastu, dulji vremenski segmenti nam pružaju detaljni prikaz frekvencije no istovremeno postoji šansa za propustom obrazaca brzih promjena [7]. Primjer opisanog je prikazan usporedbom Slike 3. i Slike 4. Važna metoda u spektrogramskoj analizi jest sposobnost namještanja točnog iznosa vremenskog segmenta krojenog prema potrebama područja primjene i karakteristikama samog signala. U kontekstu analize govornih signala ona omogućuje određivanje visine glasa, harmonika i šumova, te čak i promjene u položaju grkljanja tijekom govora čime se definira boja glasa.



**Slika 4. Spektrogramska analiza govornog signala fraze “Ovo poglavlje je naslovljeno prikaz govornih signala” pri nižoj rezoluciji**

Izvor: Izradio student pomoću softverskog programa Adobe Audition



**Slika 5. Spektrogramska analiza govornog signala fraze “Ovo poglavlje je naslovljeno prikaz govornih signala” pri višoj rezoluciji**

Izvor: Izradio student pomoću softverskog programa Adobe Audition

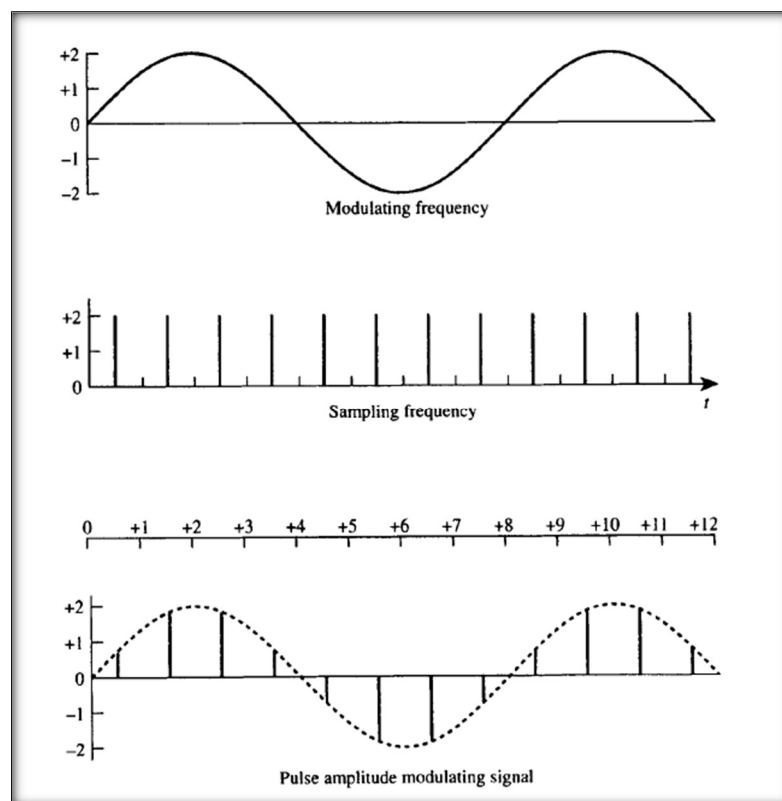
## 5. METODE DIGITALNE OBRADJE GOVORA

### 5.1. PULSNO-KODNA MODULACIJA (PCM)

Pulsno-kodna modulacija (*engl. Pulse Code Modulation*), u daljnjem tekstu PCM, široko je prihvaćen elementarni oblik digitalne reprezentacije analognih signala. Proces započinje uzorkovanjem analognog signala kojim se dobivaju diskretne vrijednosti datog signala. Analogni signal se konvertira u odgovarajući niz uzoraka s jednakim vremenskim odmakom (Slika 6.). Rezultat uzorkovanja jest pulsno-amplitudna modulacija (*engl. Pulse Amplitude Modulation*, PAM) signala karakteristična po uzastopnim izlaznim intervalima koji izgledaju poput sekvence amplituda impulsa izvedenih iz izvornog signala. Analogni signal  $s(t)$  uzorkuje se u zadanim frekvencijskim intervalima ( $f_s$ ) svakih  $T_s$  ( $1/f_s$ ) sekundi, čime se dobiva jednačba:

$$s_\delta(t) = \sum_{n=-\infty}^{\infty} s(nT_s)\delta(t - nT_s)$$

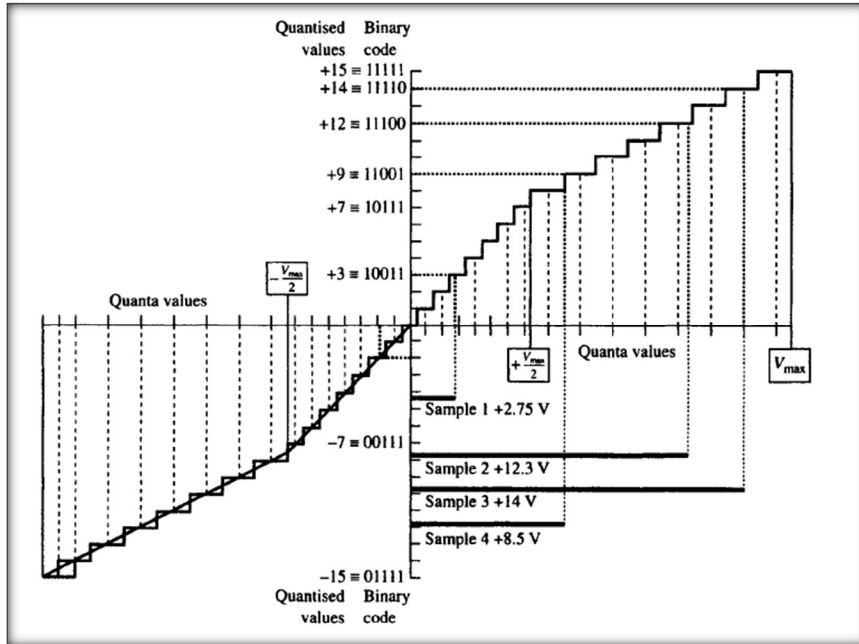
u kojoj je  $s_\delta(t)$  idealizirani uzorkovani signal, a  $\delta(t - nT_s)$  je delta funkcija smještena u vremenu  $t = nT_s$ . Prilikom uzorkovanja signala bitno je pridržavati se Nyquist-Shannon teorema uzorkovanja kako bi se izbjegla distorzija signala (alias-efekt). Teorem tvrdi da za prihvatljivu reprezentaciju analognog signala u digitalnoj domeni, on mora biti uzorkovan u stopi dvostruko većoj od njegove najviše frekvencije prisutne u signalu prije pretvorbe [7].



**Slika 6. Postupak uzorkovanja**

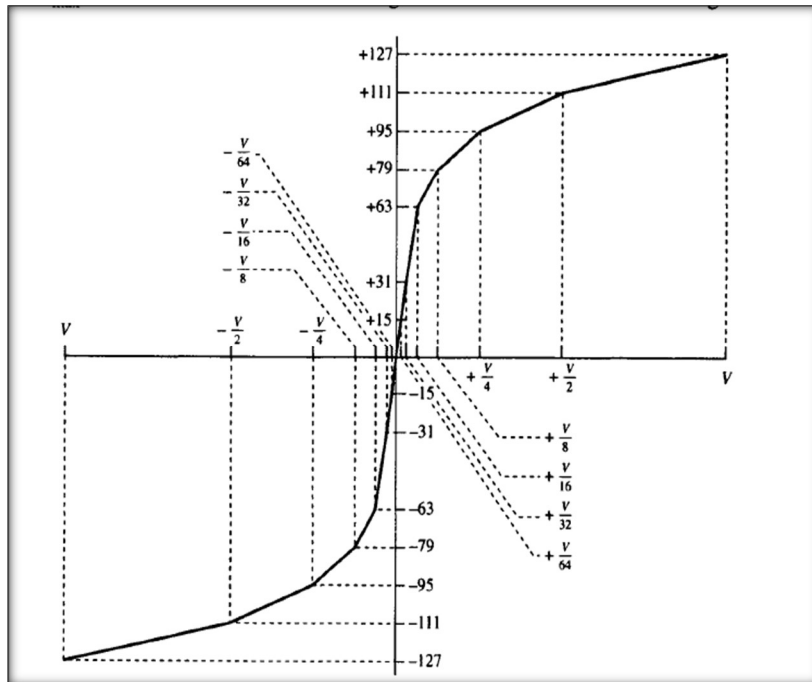
Izvor: „Analogue and Digital Communications“ – Graham Smillie, Newnes 2002

Diskretne vrijednosti dobivene procesom uzorkovanja se potom kvantiziraju. Kvantizacija uključuje aproksimaciju svake od uzorkovanih vrijednosti na najbližu vrijednost unutar predefiniраниh vrijednosti zadanog raspona (Slika 7., Slika 8.). Unatoč tome što ovaj korak može uvesti potencijalne neželjene šumove ili stupanj pogreške unutar postupka, poznatije pod imenom kvantizacijski šum, pažljiv dizajn i oprezan pristup procesu može osigurati da su ti šumovi smanjeni do prihvatljive razine za željenu primjenu.



**Slika 7. Kvantizacija uzorkovanih vrijednosti**

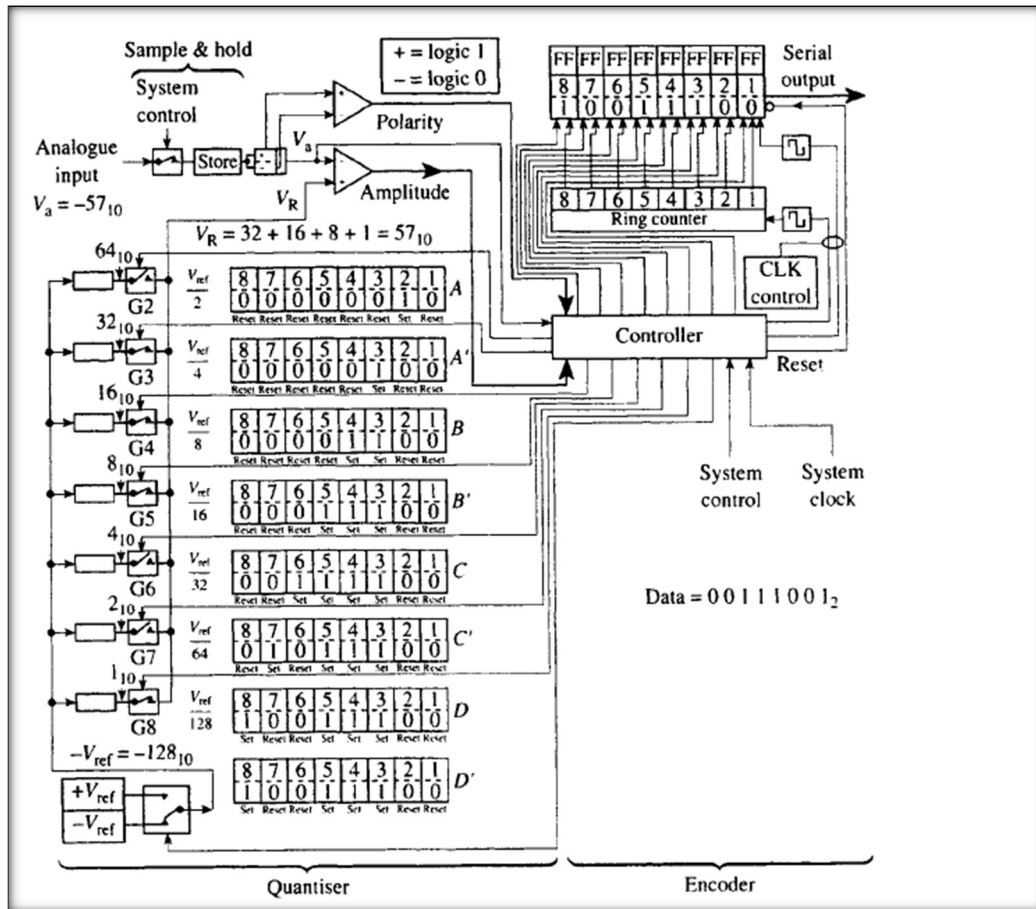
Izvor: „Analogue and Digital Communications“ – Graham Smillie, Newnes 2002



**Slika 8. Aproksimacija signala na najbliže vrijednosti**

Izvor: „Analogue and Digital Communications“ – Graham Smillie, Newnes 2002

Nakon kvantizacije signala slijedi enkodiranje tijekom kojeg dolazi do digitalizacije analognog signala. Svako kvantiziranoj vrijednosti pridružuje se binarni kod u rasponu od 0 do  $n$ , gdje je  $n$  jednak jednoj od potencija broja 2, ovisno o zahtijevanoj preciznosti. Povećanjem vrijednosti  $n$  smanjuje se razlika među susjednim razinama kvantizacije čime se smanjuje kvantizacijski šum, ali se povećava veličina digitalnog podatka. (Slika 9.)



**Slika 9. Enkodiranje signala**

Izvor: „Analogue and Digital Communications“ – Graham Smillie, Newnes 2002

Digitalizacija signala olakšala je i potaknula razvoj niza naprednih modulacijskih tehnika te učinkovitijih tehnika za kontrolu prijenosa signala, samim time sveobuhvatno optimizirajući područje telekomunikacija [8].

### 5.1.1. Diferencijalna pulsno-kodna modulacija (DPCM)

Tijekom procesa PCM-a, razlike među uzastopnim ulaznim signalima jesu minimalne te je predviđanje vrijednosti slijedećeg uzorka iz prethodnih uzoraka učestala procedura tijekom kodiranja govora. Diferencijalna pulsno-kodna modulacija (*engl. Differential Pulse Code Modulation*), u daljnjem tekstu DPCM, kvantizira razliku među originalnim i predviđenim signalima što dovodi do smanjenja broja bitova za pohranu uzorka u odnosu na PCM. Analogni signal se uzorkuje jednako kao u PCM-u te se zatim pohranjuje u prediktoru. Prediktor šalje pohranjeni signal diferencijatoru koji uspoređuje trenutni uzorak s prethodnim, te tu razliku među signalima šalje na kvantizaciju i enkodiranje. Potom se diferentni signal prenosi do prijemnika gdje se dekvantizira i pridružuje uzorku pohranjenom na prediktoru, nakon čega se šalje filteru koji rekonstruira originalni ulazni signal. Unatoč tome što DPCM smanjuje brzinu prijenosa podataka (*engl. bitrate*), korištena metoda kvantizacije dovodi do toga da veći uzorci signala imaju bolji omjer signala i šuma, čineći kvalitetu glasa boljom u usporedbi s manjim uzorcima. [9]

### 5.1.2. Adaptivno-diferencijalna pulsno-kodna modulacija (ADPCM)

Budući da većina glasovnih signala proizvedenih ljudskim glasom jesu mali uzorci, kvaliteta glasa trebala bi biti održana i na tim razinama. Za tu potrebu razvijena je adaptivno-diferencijalna pulsno-kodna modulacija (*engl. Adaptive Differential Pulse Code Modulation*), u daljnjem tekstu ADPCM.

Za razliku od DPCM-a, faza predikcije i kvantizacije u ADPCM-u su adaptivne, odnosno prilagođavaju se kako bi odgovarale karakteristikama govora koji se kodira. ADPCM prilagođava razinu kvantizacije diferentnog/razlikovnog signala koji se stvara tijekom DPCM-a. Ukoliko je diferentni signal nizak, ADPCM smanjuje veličinu razine kvantizacije, odnosno povećava razinu kvantizacije kad je diferentni signal visok [10].

## 5.2. LINEARNO PREDIKTIVNO KODIRANJE (LPC)

Linearno prediktivno kodiranje (*engl. Linear Predictive Coding*), u daljnjem tekstu LPC, digitalna metoda kodiranja analognog signala. Temeljena je na pretpostavci da se trenutni uzorak bilo kojeg signala može predvidjeti pomoću linearne kombinacije konačnog broja prošlih uzoraka.



Budući da svi iznosi prošlih uzoraka ne doprinose jednako iznosu trenutnog uzorka, oni se prije zbrajanja moraju pomnožiti odgovarajućim težinskim koeficijentima – parametrima ili koeficijentima linearnog prediktora. Broj tih koeficijenata određuje red prediktora. Cilj LPC analize je određivanje parametara linearnog prediktora tako da razlika između originalnih uzoraka i onih dobivenih predikcijom bude minimalna kako bi se što više uštedilo na prostoru. Parametri prediktora računaju se u okviru analize signala konačnog trajanja (20-30 ms), budući da ne mogu biti konstantni u dužem vremenskom intervalu. Oni se najčešće računaju pomoću metode autokorelacije. Dio signala se izdvaja te se zatim izračunava vrijednost autokorelacijske funkcije iz čijih se dobivenih rezultata izračunavaju vrijednosti koeficijenata prediktora rješavanjem linearnog sustava jednažbi:

$$\sum_{j=1}^p \alpha_j \cdot R_n(|i-j|) = R_n(i), \quad \text{za } i = 1, 2, \dots, p$$

koji prikazan u matričnom obliku izgleda:

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

Tijekom analize zvučnog signala prvo se određuje je li trenutni segment zvuka zvučan (samoglasnici) ili bezvučan (suglasnici). Potom generator šuma kreira uzorke  $f(n)$  za bezvučne glasove koji će služiti kao ulazni signal za simulator vokalnog trakta (engl. *vocal-tract simulator*), dok se zvučni glasovi kreiraju u uzorke putem generatora impulsnog niza (engl. *pulse-train generator*).

Ukoliko se izlazne vrijednosti označe sa  $s(n)$ , tada za ulazne uzorke bezvučnih glasova  $f(n)$  izlazni uzorak ovisi o deset prethodnih izlaznih vrijednosti:

$$s(n) = \sum_{i=1}^{10} \alpha_i s(n-i) + G \cdot f(n),$$

Gdje su  $\alpha_i$  koeficijenti prediktora dobiveni autokorelacijom, a  $G$  je faktor pojačanja dobiven formulom:

$$G = R(0) - \sum_{j=1}^{10} \alpha_j R(j)$$

Analizom se određuje i duljina perioda glasa  $P$  (engl. *pitch period*) tj. trajanja vibracije glasnica tijekom produkcije glasa, pomoću funkcije prosječne amplitudalne razlike (engl. *average magnitude difference function*, AMDF) definirane kao:

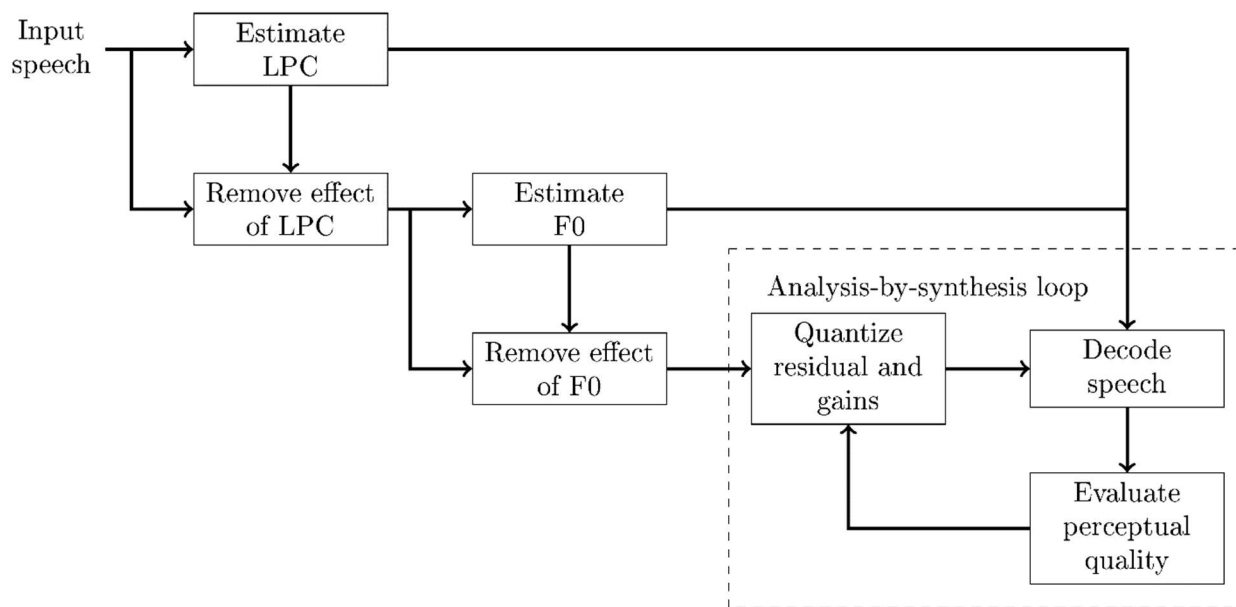
$$AMDF(P) = \frac{1}{N} \sum_{k=0}^{N-1} |f_k(n) - f_{k-P}(n)|$$

Budući da je duljina perioda glasa kod ljudi ograničena, AMDF se izračunava za ograničeni raspon mogućih vrijednosti  $P = [2.5, 19.5]$  ms, tj. za signal uzorkovan na stopi od 8kHz  $AMDF(P) = [20, 160]$ . Vrijednosti niže od zadanog praga bit će zabilježene kao bezvučne te rekonstruirane kao bijeli šum. Izlazni signali se potom šalju do dekodera koji rekreira originalni ulazni signal pomoću IIR filtera. Obzirom na zvučnost signala dekođer šalje pobudni signal za zvučne signale ili šum za bezvučne. Impuls zvučnih signala sastoji se od 40 uzoraka koji su spremljeni lokalno u dekoderu, dok šum za bezvučne signale stvara pseudo-nasumični generator brojeva. Potom se pitch period zvučnih signala upotrebljava kako bi se odredilo treba li impuls skratiti ili produžiti. Posljednji korak dekodiranja jest prolazak ekscitacijskog signala kroz filter kako bi se proizveo sintetizirani glasovni signal. Ukoliko ulazni govorni signal ne može biti označen kao zvučan ili bezvučan, LPC model neuspješno definira okvir zvučnosti zbog čega dolazi do izobličenja u reprodukciji izlaznog signala. Suprotno tome, ukoliko su ti okviri prestrogo definirani, rezultat je izrazito drugačiji od prirodnog ljudskog glasa, odnosno zvuči robotizirano [11].

### 5.2.1. Kodno-pobuđena linearna predikcija (CELP)

Kodno-pobuđena linearna predikcija (engl. *Code Excited Linear Prediction*), u daljnjem tekstu CELP, jedna je od najprepoznatljivijih tehnika obrade govora danas iz koje je izveden niz načina kodiranja poput AMR-a, EVS-a, G.718 te Opusa. CELP metoda bazirana je na *source-*

*filter* modelu govora unutar kojega linearna predikcija služi za modeliranje filtracijskih promjena vokalnog trakta opisanih u prošlom poglavlju. U svrhu modeliranja šuma ova metoda potpomognuta je korištenjem kodne knjige vektora pobuda (engl. *codebook of excitation vectors*). Uzorci knjige uspoređuju se sa stvarnim govorom te se šalje indeks najboljeg podudaranja. Uz korištenje kodne knjige, CELP metoda uvodi dvije vrste predviđanja koje zamjenjuju proces analize zvučnih i bezzvučnih zvukova. Redundancija govornog signala postiže se kroz kratkotrajno predviđanje (engl. *short-time prediction*, STP) i dugotrajno predviđanje (engl. *long-time prediction*, LTP). STP provodi LPC analizu signala nad nekoliko uzastopnih uzoraka nakon kojeg LTP utvrđuje korelaciju između uzoraka udaljenih za jedan period glasa. CELP koristi analizu sintezom (engl. *analysis-by-synthesis*) gdje se najbolje podudaranje u kodnoj knjižici određuje metodom zatvorene petlje (engl. *closed-loop system*), a ostali parametri otvorenom petljom (engl. *open-loop system*). Metoda optimizacije zatvorenom petljom koristi parametre kako bi se stvorio signal tijekom kodiranja koji se potom preciznije podešavaju kako bi generirali što vjerniju konstrukciju. U sustavu otvorene petlje parametri koji reprezentiraju signal dobivaju se izravno iz ulaznog signala te se zatim kvantiziraju i koriste za sintezu zvuka.



**Slika 10. Objektni dijagram CELP enkodera**

Izvor: [https://speechprocessingbook.aalto.fi/Transmission/Code-excited\\_linear\\_prediction\\_CELP.html](https://speechprocessingbook.aalto.fi/Transmission/Code-excited_linear_prediction_CELP.html)

Ulazni signal dijeli se na vremenske okvire duljine 20 do 30 ms i podokvire duljine 5 do 7.5 ms. STP se provodi za svaki okvir kako bi se dobili koeficijenti linearnog predviđanja, dok se LTP provodi nad podokvirima. Kad su svi koeficijenti izračunati, filteri započinju određivati pobudni niz. Za svaki podokvir traži se najbolji uzorak iz kodne knjige koji započinje stvaranjem generiranog skupa filtriranih pobudnih nizova s odgovarajućim pojačanjima, te se potom za svaki izračuna pogreška u odnosu na uzorke i odabere onaj s najmanjom. U konačnici se indeks kodne knjižice, pojačanje, LTP parametri i LPC kodiraju, pakiraju i šalju kao CELP skup podataka. Obzirom da svi govorni segmenti koriste isti skup predložaka iz kodne knjižice, preneseni govor je puno prirodniji u odnosu na LPC. Glavna mana CELP metode je što je specijalizirana isključivo za kodiranje i reprezentaciju govora te daje jako loše rezultate kada je ulazni signal glazba [12].

## 6. PRIMJENA DIGITALNE OBRADJE GOVORNIH SIGNALA

### 6.1. GOVOR PUTEM INTERNET PROTOKOLA (VOIP)

Govor putem Internet protokola (engl. *Voice over Internet Protocol*), u daljnjem tekstu VoIP, predstavlja inovativan pristup u telekomunikacijskim sustavima koji je drastično promijenio okruženje te industrije. Za razliku od tradicionalnih sustava telefonije koji se pouzdaju u zadužene mreže signalnih kabela ili mobilnih mreža, VoIP dopušta prijenos glasovnih podataka putem Interneta. Učinkovitost VoIP prijenosa temeljena je na kodecima koje koristi za kompresiju i dekompresiju govornih podataka, a najistaknutiji su G.711, G.729 i GSM [23]. Temeljna sastavnica VoIP-a jest kvaliteta usluge (engl. *Quality of Service*, QoS). Uzevši u obzir da čimbenici poput kašnjenja, šumova i gubitaka podatkovnih paketa mogu uvelike utjecati na kvalitetu poziva, potrebno je riješiti ili smanjiti probleme s propusnošću, raspoloživošću usluge, gubitkom paketa, kašnjenjem i kolebanjem kašnjenja [13].

Prednosti usvajanja VoIP-a su brojne. Moderni VoIP sustavi često mogu biti integrirani s drugim digitalnim uslugama poput elektroničke pošte, sustava za video konferencije i pozive te automatskog sustava povratnih informacija korisnika. Nadalje, VoIP nudi naprednija svojstva u usporedbi s tradicionalnim sustavima telefonije uključujući automatski prijepis glasovne na elektroničku poštu, virtualne mobilne brojevi te video pozive. Osim toga, istaknuta prednost VoIP-a je njegova prilagodljivost. Kroz jedan životni ciklus kompanije te neophodnog rasta iste, VoIP sustavi se mogu pravilno prilagoditi novom obujmu korisnika i mreže koju kompanija zastupa bez većih promjena na već uspostavljenoj infrastrukturi.

Oslanjanje na Internet za prijenos svojih podataka predstavlja još jedan izazov VoIP sustavu. Budući da su prijetnje sigurnosti digitalne okoline u današnje vrijeme sve češće, time je sve prisutnija potreba za sigurnom virtualnom okolinom. Prisluski vanja, napadi uskraćivanja usluga (engl. *Distributed Denial of Service*, DDoS) te napadi posrednika (engl. *Man-in-the-middle*, MITM) su samo neki od potencijalnih prijetnji ovakvim sustavima. U naporima eliminacije ovih sigurnosnih rizika, neophodno je enkriptirati glasovne podatke u prijenosu te osigurati uporabu prikladnih sigurnosnih protokola diljem mreže.

## 6.2. SKRIVENE MARKOVLJEVE MREŽE (HMM)

Markovljev lanac ili Markovljev proces je stohastički model koji opisuje niz događaja u kojima vjerojatnost svakoga ovisi o promatranom stanju iz prijašnjeg događaja tj. procjena trenutnog izlaza događaja uvjetuje vjerojatnost sljedećeg. Markovljevi procesi temelj su općih stohastičkih simulacijskih metoda poznatijih pod nazivom Monte Carlo Markovljevi lanci. Ti lanci mogu simulirati izrazito kompleksne raspodjele vjerojatnosti s primjenom u područjima termodinamike, statistike, fizike, obrade signala i govora, informacijske teorije i sl. Skrивene Markovljeve mreže (engl. *Hidden Markov Models*), u daljnjem tekstu HMM, popularan je statistički alat za modeliranje sustava s velikim količinama podataka. U svrhu obrade prirodnog jezika (engl. *Natural Language Processing*, NLP), HMM se s velikim uspjehom koristi u područjima identifikacije izgovorenih riječi te raščlanjenja istih na njene fonemske sastavnice [14]. HMM je statistički Markovljev model unutar kojeg se pretpostavlja da sustav koji se obrađuje jest Markovljev proces, u sljedećim frazama opisan sa  $X$ , s nepromatranim tj. skrivenim stanjima. Po definiciji HMM-a zahtjeva se da postoji i promatrano stanje  $Y$  čiji izlazi su utjecani od strane izlaza  $X$  na poznati način. Budući da  $X$  ne može biti promatran direktno, cilj je učiti o mehanizmima  $X$  promatranjem izlaza  $Y$ . HMM imaju dodatan uvjet da izlazi  $Y$  u vremenu  $t=t_0$  moraju biti utjecani isključivo od strane izlaza na  $X$  u vremenu  $t=t_0$ . Navedeni izlazi  $X$  i  $Y$  u vremenu  $t < t_0$  moraju biti uvjetno nezavisni naspram  $Y$  u vremenu  $t=t_0$  uz uvjet da je  $X$  promatran u istom vremenu. Metodologija HMM slijedi u daljnje opisanim frazama sa  $X_t$  i  $Y_t$  koji su kontinuirani stohastički procesi. Navedeni par  $(X_t, Y_t)$  je HMM ako vrijedi:

- $X_t$  je Markovljev proces čije ponašanje nije moguće direktno promatrati,
- $P(Y_{t_0} \in A \mid \{X_t \in B_t\}_{t \leq t_0}) = P(Y_{t_0} \in A \mid X_{t_0} \in B_{t_0})$  što vrijedi za svaki  $t_0$ , svaki skup  $A$  te svaki podskup  $\{B_t\}_{t < t_0}$ .

Stanja procesa  $X_t$  se zovu skrivena stanja, a jednakost pod drugom točkom objašnjenja HMM-a se zove emisijska vjerojatnost (engl. *emission probability*). Primjenom HMM-a u digitalnoj obradi govora postižu se dva veoma bitna cilja: prepoznavanje i izolacija govora. Dekompozicija signala putem HMM-a vrši se koristeći paralelne HMM-ove. U takvom sustavu, svaki HMM je zaslužan za jednu od sastavnica na koje će se govorni signal dekompozirati. U slučaju dekompozicije u svrhu elimiranja šuma, skupovi korišteni u dekompoziciji su učestali modeli riječi i skup modeliranih šumova. Vjerojatnosti promatranja procijenjene su na temelju

izlaza govornog modela i izlaza modeliranog šuma. Prepoznavanje govora provedeno je sa jednim skupom HMM-a, dok je šum prepoznat pomoću drugog skupa.

### 6.3. UMJETNE NEURALNE MREŽE

Umjetne neuronske mreže (*engl. Artificial Neural Network*), u daljnjem tekstu ANN, predstavljaju prekretnicu između biološke inspiracije i računalnih mogućnosti. Začetno modelirani na temelju zapletene mreže neurona u ljudskom mozgu, ANN su se izrodile kao prevalentni alat u nizu računalnih zadataka, posebice u području prepoznavanja obrazaca i složenog nelinearnog modeliranja podataka. ANN pripadaju jedinstvenom skupu modela strojnog učenja. Prednost ovih sustava je mogućnost oblikovanja i učenja na temelju pridruženih izvora podataka te prijašnjih iteracija obrade istih. Praktična primjena ANN-ova u kontekstu ovog rada je prepoznavanje riječi u ljudskom govoru, jezika te čak i prevalentne emocije govornika budući da ANN mogu iterativno doći do znanja o postojanju specifičnih obrazaca govora koji bi u suprotnom istraživačima i stručnjacima promakli. Funkcionalnost ANN-a temeljena je na neuronima, umreženim živčanim stanicama koje prenose informacije (živčani impuls) s jedne na drugu. Svaki neuron ima mogućnost primiti nekoliko ulaznih signala od susjednih neurona, ali može prenijeti isključivo jedan na sljedeći. Uvjetovanje tih veza izvedeno je pomoću težinskih koeficijenata po kojima svaka veza ima vlastitu numeričku težinu. Iterativnost ANN-a dopušta skladištenje i promjenu težinskih koeficijenata čime ANN-ovi uče. Obrada signala i aktivacija neurona je prikazana pomoću sljedeće formule:

$$X = \sum_{i=1}^n X_i W_i$$
$$Y = \begin{cases} +1 & \text{ako je } X \geq 0, \\ -1 & \text{ako je } X < 0. \end{cases}$$

gdje je  $X$  prosjek mrežnih ulaza,  $X_i$  vrijednost unosa,  $W_i$  težina unosa,  $n$  broj neurona na ulazu te  $Y$  koji predstavlja izlaz neurona. Navedeni neuroni su razvrstani u slojeve: ulazni sloj, niz skrivenih slojeva te izlazni sloj. Prilagođavajući koeficijente težine koji povezuju navedene neurone tokom procesa učenja mreže, ti neuroni imaju sposobnost savladavanja procjene nelinearnih funkcija čineći ih izrazito pogodnim za zadatke gdje komplicirani uzorci trebaju biti izvedeni na temelju

velikih skupova podataka. Proces učenja za kompleksnije neuralne mreže zahtijeva ogromne količine sortiranih podataka i računalnih resursa. Nadalje, razumijevanje i tumačenje prirode crnih kutija ovakvih mreža još uvijek ostaje izazov učestalo stvarajući probleme pri tumačenju pojedinačnih odluka ANN-ova u određenim trenucima [15].

#### 6.4. DRAGON DICTATE

Dragon Dictate, učestalo referiran kao samo Dragon, predstavlja značajan skok u razvoju softvera prepoznavanja govora. Inicijalno razvijen kao specijalizirani alat za područje medicine, preciznost ovog alata je ubrzo nakon začeća dovela do široke raspostranjenosti u nizu ostalih domena od prava do odgojno-obrazovnih sustava. Najzamjetnije svojstvo Dragona je njegov napredan algoritam s mogućnošću prilagodbe i učenja na temelju govora svakog pojedinačnog korisnika. Ovakav progresivni oblik strojnog učenja osigurava da navedeni softver daje precizne rezultate sa rastućom krivuljom točnosti pri kontinuiranoj upotrebi. Kontinuiranom upotrebom sustav je u stanju obraditi čak i raznovrsne dijalekte ili specifičan izgovor nekoliko jezika (engleski, nizozemski, francuski, njemački, talijanski, španjolski, japanski) čineći ovaj softver univerzalno primjenjivim alatom. Ne spominjući mogućnost prijepisa, Dragon sa sobom donosi još funkcionalnosti sa sposobnošću neprimjetne integracije sa nizom softverskih okruženja poput programa za obradu teksta ili Internet preglednika. Korisnici upoznati s ovim softverom u stanju su putem isključivo glasovnih naredbi sastavljati elektroničku poštu, izrađivati izvještaje te čak pretraživati Internet. Sa strane korisničke dostupnosti (engl. *Accessibility*), utjecaj Dragona je neosporiv. Korisnicima koji posjeduju tjelesne poteškoće (tetraplegija, sindrom karpalnog tunela...) ili oboljenjima poput multiple skleroze, Dragon nudi prihvatljivu alternativu tradicionalnim kompjuterskim sučeljima. Nadalje, u dobu karakteriziranom naglim rastom u području rada na daljinu i virtualnih međudjelovanja, Dragon olakšava jednostavniju i učinkovitiju komunikaciju zaobilazeći ograničenja pisanja na tipkovnici. Naravno, Dragon Dictate ima i svoja ograničenja. Čimbenici poput okolne buke, preklapajućeg razgovora te specifične složenosti dijalekata mogu ponekad utjecati na preciznost. Postoji i svojstven period učenja koji je potreban za prihvatljivu implementaciju ovog softvera u njegovo okruženje tokom kojeg perioda su korisnici potaknuti na upoznavanje sa zasebnim glasovnim naredbama te tehnikama izgovora u svrhu pospješenja učinkovitosti istoga [16].



## 7. ZAKLJUČAK

Ubrzanim razvojem tehnologije u današnjici, brza i neometana komunikacija postala je potreba više nego htijenje. Digitalna obrada govora omogućila je smanjenje potrebnog podatkovnog prostora za prijenos govornih signala na velikim udaljenostima. Razvivši se iz jednostavnijih metoda poput PCM-a, DPCM i ADPCM izrazito su poboljšali kvalitetu prijenosa i izlaznog signala temeljeći se na razlikama kvantizacijskih razina umjesto cijelokupnih uzoraka. LPC i CELP uvelike su potpomogli u smanjenju obujma podataka koristeći sustave predikcije u kombinaciji s kodnim knjigama pobudnih signala, odnosno riječi. CELP metoda otišla je i korak dalje u kompresiji uz proces analize putem sinteze.

Kompliciraniji zahtjevi za obradu govora doveli su do razvitka matematičkih i računalnih modela poput skrivenih Markovljevih modela i umjetnih neuronskih mreža. Oni su bili ključni u poboljšanju preciznosti i učinkovitosti sustava digitalne obrade govora. Govor putem Internet protokola revolucionarizirao je područje telekomunikacija i omogućio prijenos glasovnih signala uz minimalne gubitke kvalitete.

Unatoč tome što je područje digitalne obrade govora relativno novija tehnologija, u kratko vrijeme svog postojanja uspjela je postati nezamjenjiv proces koji utječe na gotovo svaki aspekt svakidašnjeg života ljudi. Realno je za očekivati da je budućnost digitalne obrade obećavajuća te da će dovesti do daljnjeg unaprjeđenje interakcije ljudi i računala, same komunikacije te uporabe govora u različitim sferama društva i života.

## LITERATURA

- [1] "Kompendij anatomije čovjeka: za studente opće medicine i stomatologije - Dio 2: Pregled građe glave, vrata i leđa" - Zdenko Križan, Školska knjiga 1989
- [2] "Speech Physiology, Speech Percetion and Acoustic Phonetics" – Philip Libereman, Sheila E. Blumstein; Cambridge University Press, 1988
- [3] "Z-Transform Theory and the FTDT Method" – Dennis M. Sullivan, IEEE Transactions on antennas and propagation, Vol. 44, No.1 January 1996
- [4] "Digital Signal Processing – Fundamentals and Applications" – Li Tan, Jean Jiang, Academic Press Elsevier, 2013
- [5] "Speech Analysis in Time and Frequency Domain" - Solly Joy, Sabitha Upadhya, International Journal of Engineering Research and Technology (IJERT) 2015
- [6] "Spectral Analysis of Signals" – Petre Stoica, Randolph L. Moses, Prentice Hall 2005
- [7] „Wireless Communications & Networking – Chapter 4 An Overview of Digital Communication and Transmission“, Vijay K. Garg, Elsevier 2007
- [8] "Analogue and Digital Communication Techniques" – Grahame Smillie, Newnes 2002, Elsevier Science 1999
- [9] "Modern Digital and Analog Communication Systems Third edition" - B.P. Lathi, Oxford University Press 1998
- [10] „Adaptive, Differential Pulse Code Modulation for Speech Processing“ – Peter Cummiskey, Newark College of Engineering, D. Eng, Sc., 1973
- [11] "Linear Predictive Coding" – Jeremy Bradbury, Kluwer Academic 2000
- [12] "Speech Coding with Code-Excited Linear Prediction" – Tom Bäckström, Springer 2017
- [13] "Voice over Internet Protocol (VoIP)" – Bur Goode, Proceedings of the IEEE, Vol. 90, No.9, September 2002
- [14] "Workbook For: Nuance Dragon Professional Individual" – Nuance, Dragon Speech Recognition Guide v15, August 2015
- [15] "Hidden Markov Models" – Phil Blunsom, 2004
- [16] "Neural Networks and Speech Processing" – D.P. Morgan et al, Kluwer Academic 1991

## KAZALO KRATICA

Kratika	Puni naziv na engleskom jeziku	Tumačenje na hrvatskom jeziku
ADPCM	Adaptive Differential Pulse Code Modulation	Adaptivno-diferencijalna pulsno-kodna modulacija
AMDF	Average Magnitude Difference Function	Funkcija prosječne amplitudalne razlike
ANN	Artificial Neural Network	Umjetne neuronske mreže
CELP	Code Excited Linear Production	Kodno-pobuđena linearna predikcija
DDoS	Distributed Denial of Service	Napadi uskraćivanja usluge
DFT	Discrete Fourier Transform	Diskretna Fourierova transformacija
DPCM	Differential Pulse Code Modulation	Diferencijalna pulsno-kodna modulacija
FFT	Fast Fourier Transform	Brza Fourierova transformacija
HMM	Hidden Markov Model	Skrivene Markovljeve mreže
LPC	Linear Predictive Coding	Linearno prediktivno kodiranje
LTP	Long-Time Prediction	Dugotrajno predviđanje
MITM	Man-in-the-middle	Napad posrednika
NLP	Natural Language Processing	Obrada prirodnog jezika
PAM	Pulse Amplitude Modulation	Pulsno-amplitudna modulacija
PCM	Pulse Code Modulation	Pulsno-kodna modulacija
QoS	Quality of Service	Kvaliteta usluge
STFT	Short-Time Fourier Transform	Kratkotrajna Fourierova transformacija
STP	Short-Time Prediction	Kratkotrajno predviđanje
VoIP	Voice over Internet Protocol	Govor putem Internet protokola

## POPIS ILUSTRACIJA

Slika 1. Anatomija ljudskog vokalnog trakta.....	3
Slika 2. Prikaz govornog signala fraze “Ovo poglavlje je naslovljeno prikaz govornih signala” u vremenskoj domeni.....	8
Slika 3. Prikaz govornog signala fraze “Ovo poglavlje je naslovljeno prikaz govornih signala” u frekvencijskoj domeni.....	9
Slika 4. Spektrogramska analiza govornog signala fraze “Ovo poglavlje je naslovljeno prikaz govornih signala” pri nižoj rezoluciji .....	10
Slika 5. Spektrogramska analiza govornog signala fraze “Ovo poglavlje je naslovljeno prikaz govornih signala” pri višoj rezoluciji.....	11
Slika 6. Postupak uzorkovanja.....	13
Slika 7. Kvantizacija uzorkovanih vrijednosti .....	14
Slika 8. Aproksimacija signala na najbliže vrijednosti.....	14
Slika 9. Enkodiranje signala .....	15
Slika 10. Objektni dijagram CELP enkodera.....	19