

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

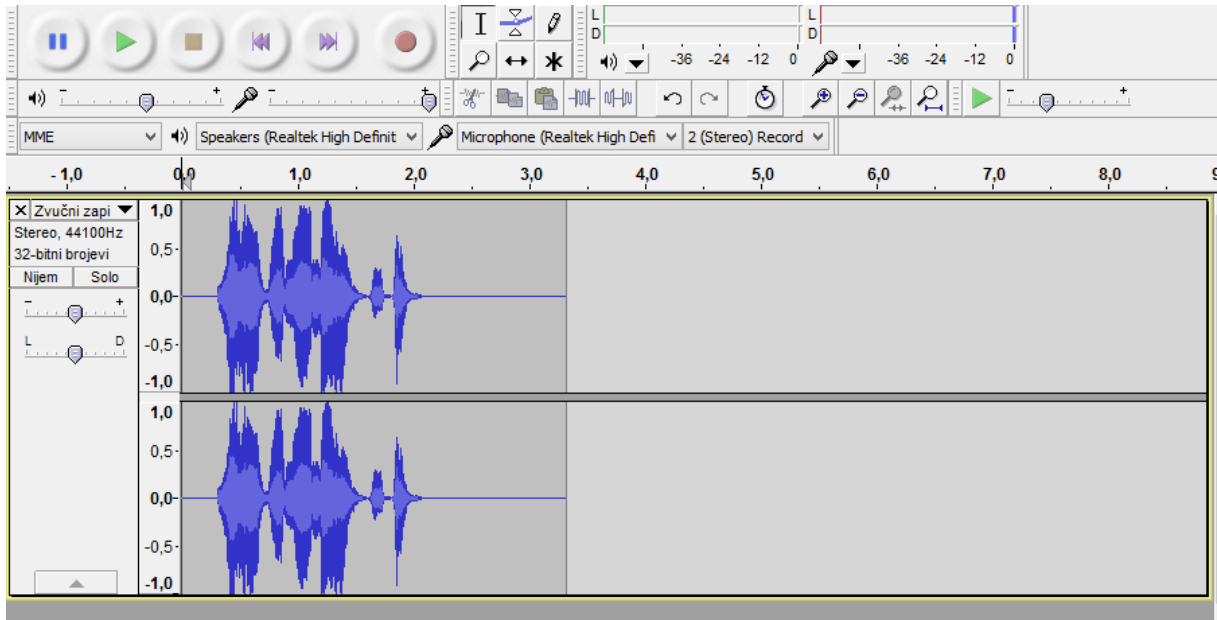
VIŠEMEDIJSKE KOMUNIKACIJE

1. DOMAĆA ZADAĆA

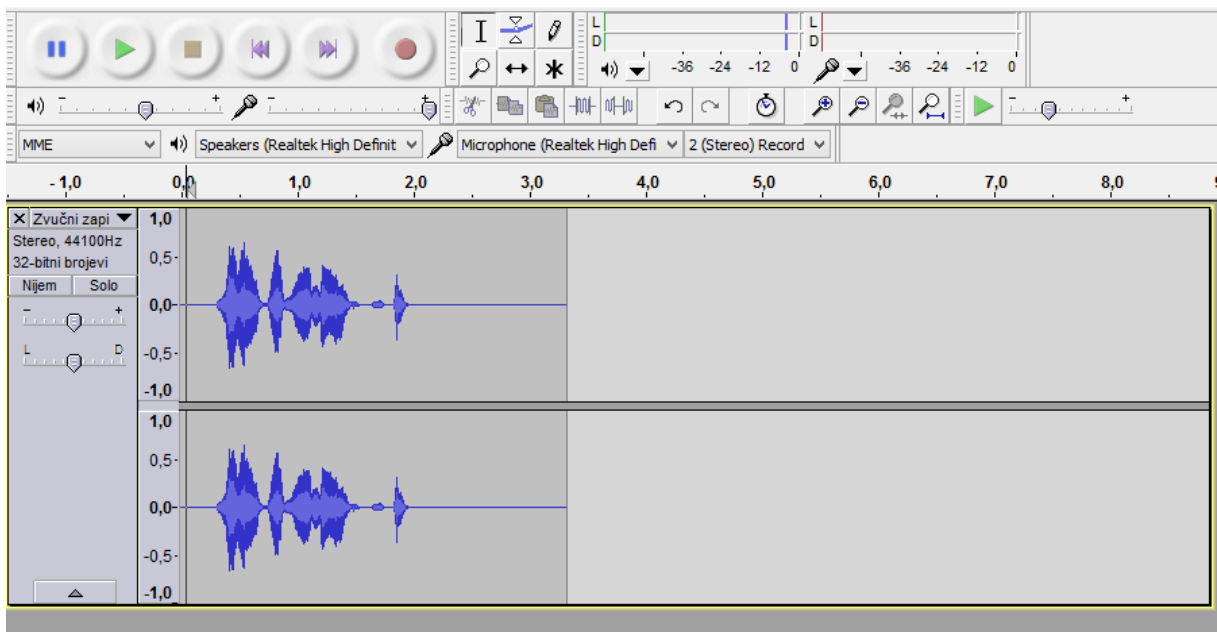
20. ožujak 2015.

1 Vremenska i frekvencijska analiza govora

Za izradu ovoga dijela domaće zadaće potreban nam je alat Audacity koji služi za analizu i obradu zvuka. Na slici 1 prikazana je snimljena rečenica „Moje ime je xxx“ alatom Audacity, dok je na slici 2 također ista rečenica samo je otklonjen šum. Zvuk je prikazan u vremenskoj domeni, dok ćemo za izradu zadaće koristiti još i frekvencijsku domenu koju odabiremo na *Prouči -> Graf dB/Hz*.



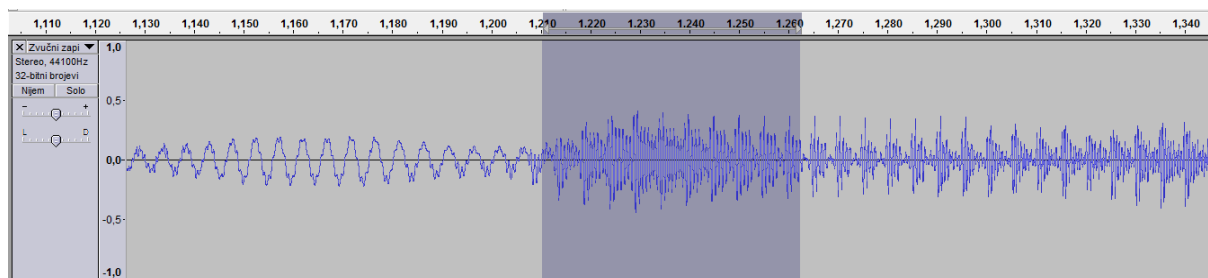
Slika 1. Snimljena rečenica



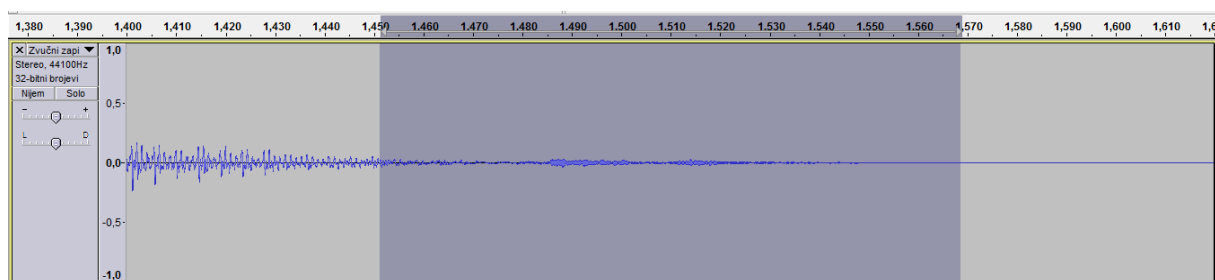
Slika 2. Snimljena rečenica bez šuma

1. U vremenskoj domeni valni oblik zvuka karakteriziran je frekvencijom (valnom duljinom), fazom i amplitudom. Odabrani zvučni glas je „m“ čiji je valni oblik prikazan na slici 2, dok je valni oblik bezvučnog „s“ prikazan na slici 4. Zvučni glas je periodičan

jer je to glas koji proizvode naše glasnice dok titraju, a bezvučni glas nije periodičan jer nastaje strujom glasa kroz naš govorni organ gdje nema karakterističnog titranja.

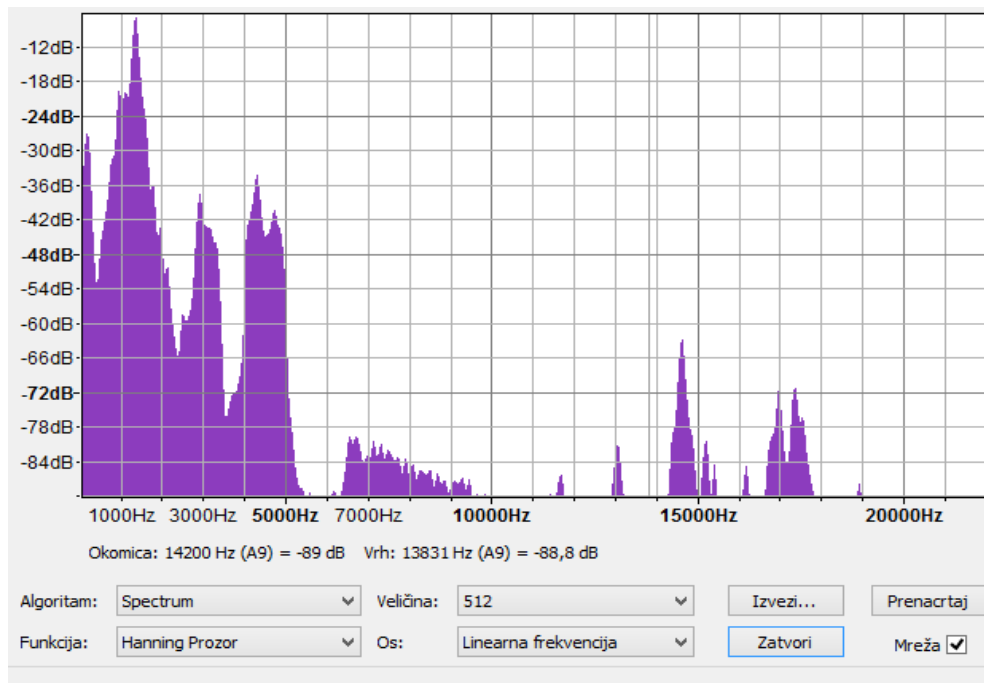


Slika 3. Valni oblik zvučnog glasa „m“

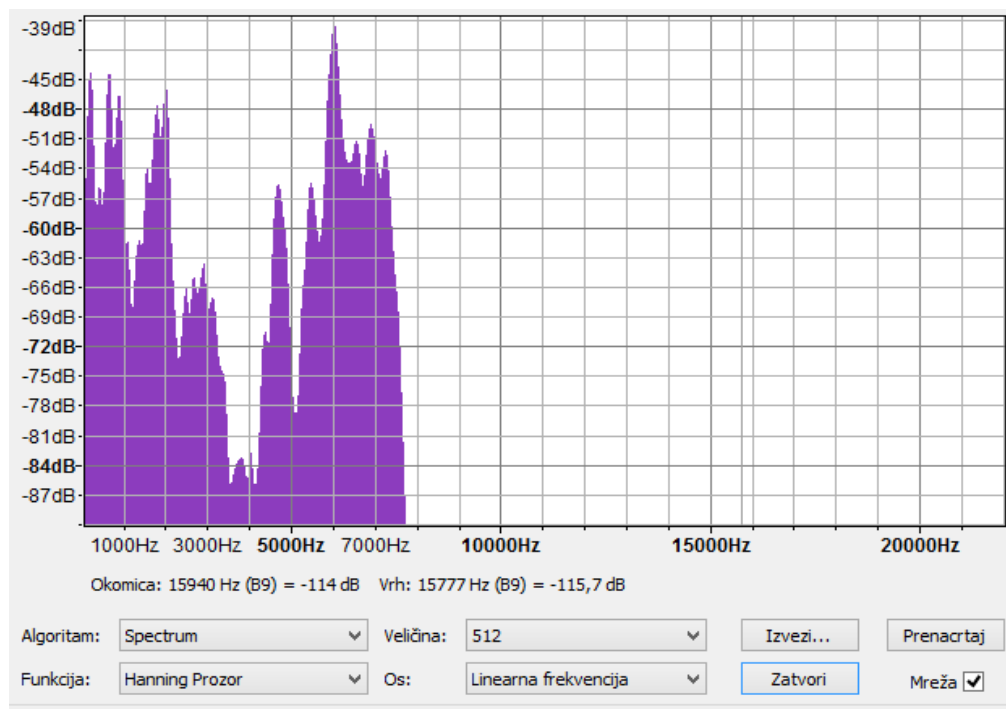


Slika 4. Valni oblik bezvučnog glasa „s“

2. Prikaz signala zvučnog glasa „m“ u frekvencijskoj domeni prikazan je na slici 5, dok je prikaz signala bezvučnog glasa „s“ prikazan na slici 6. Sa slike 5 vidimo da se za zvučni glas „m“ ističu neke frekvencije ili formante, a kod bezvučnog glasa „s“ ne možemo to vidjeti. U frekvencijskoj domeni zvučni glas se može uočiti kroz njegove karakteristične frekvencije ili formante, dok se kod bezvučnih ne može uočiti nikakva pravilnost.



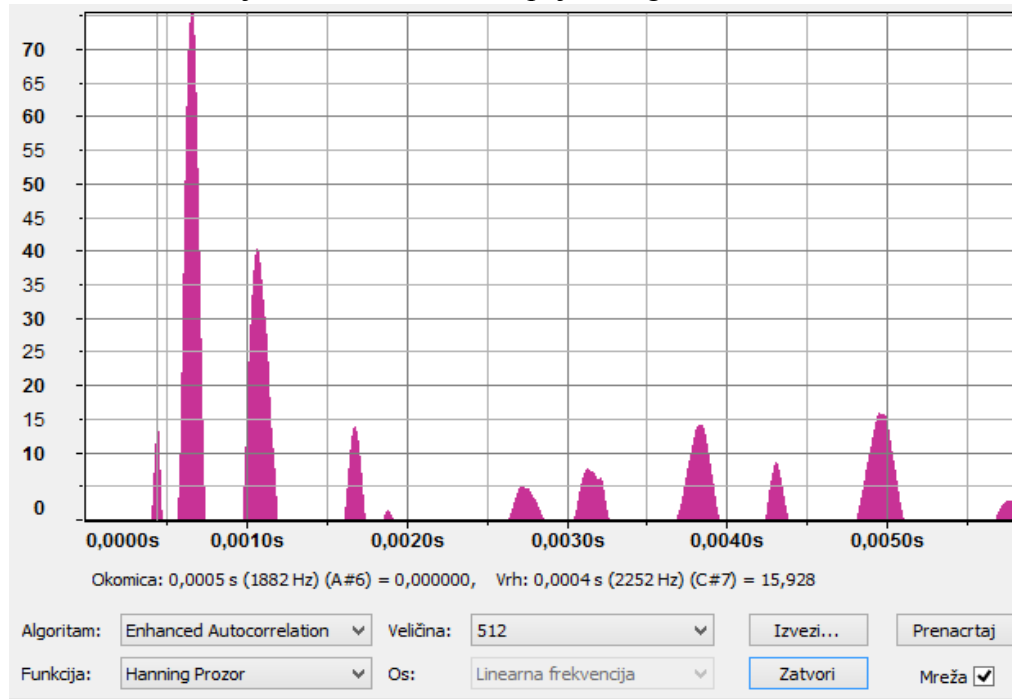
Slika 5. Signal zvučnog glasa „m“ u frekvencijskoj domeni



Slika 6. Signal bezvučnog glasa „s“ u frekvencijskoj domeni

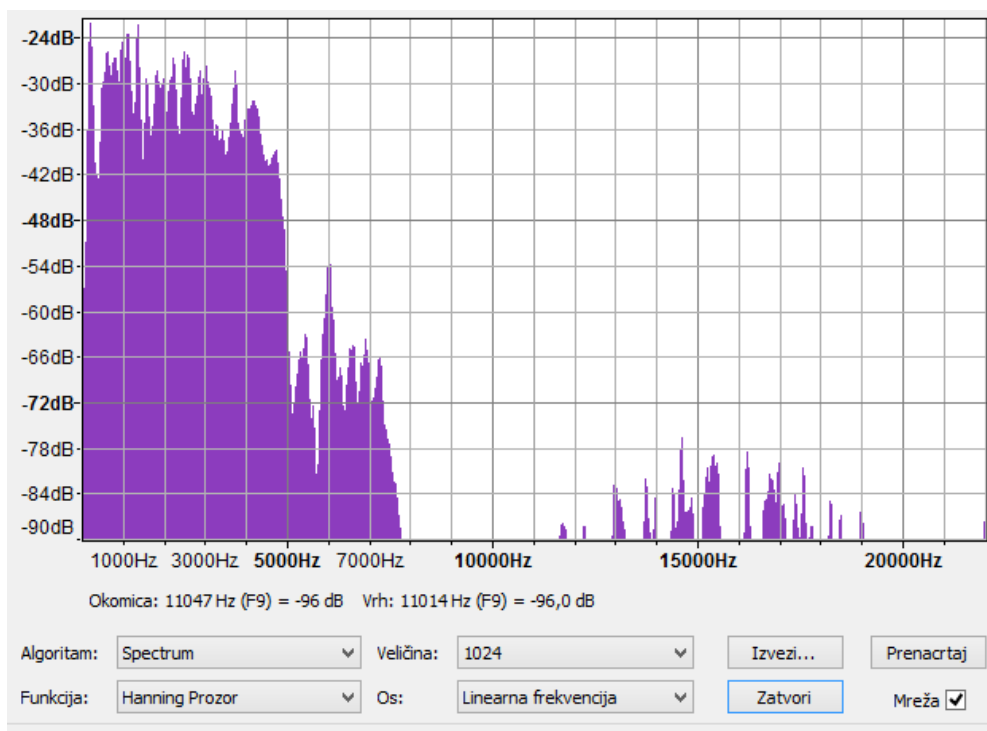
3. Odabrani glas je glas „a“ i njegovi formanti su prikazani na slici 7. Formanti glasa „a“ pojavljuju se na frekvencijama 1507 Hz, 926 Hz, 596 Hz, 367 Hz, 319 Hz, 260 Hz i 201 Hz. Vrh koji se nalazi na najnižoj frekvenciji zove se prvi formant (F1), a na sljedećoj frekvenciji je drugi formant (F2). Ovdje je najniža frekvencija 201 Hz, dok je sljedeća 260 Hz. Vrijednosti iz tablice za ženski glas je prvi formant 215 Hz što je približno jednako snimljenom formantu, do je vrijednost u tablici za drugi formant 669 Hz, što je

puno veća vrijednost od snimljenog, ali u tablici se nalaze prosječne vrijednosti, a svaka osoba ima različite vrijednosti formanta za pojedine glasove.



Slika 7. Formanti glasa „a“

- Na slici 8 prikazan je cijeli govorni signal u frekvencijskoj domeni. Izgled signala je nepravilan i neperiodičan i vidimo da su amplitude signala vidljivo najjače do nekih 3400 Hz što je povezano s rasponom kodiranja ljudskog govora u analognoj telefoniji jer se u tom rasponu (300-3400 Hz) kreće ljudski govor.



Slika 8. Cijeli govorni signal u frekvencijskoj domeni

6. Provođenjem niskopojasnog filtriranja signala sa gornjom graničnom frekvencijom od 3.4kHz kvaliteta govora se ne mijenja, jer je raspon govora od 300 do 3400 Hz, pa ako ograničimo frekvenciju sa 3400 Hz to neće utjecati na signal, te se zbog toga takvo filtriranje koristi kod analognog prijenosa signala.

2 Simulacija koda LPC

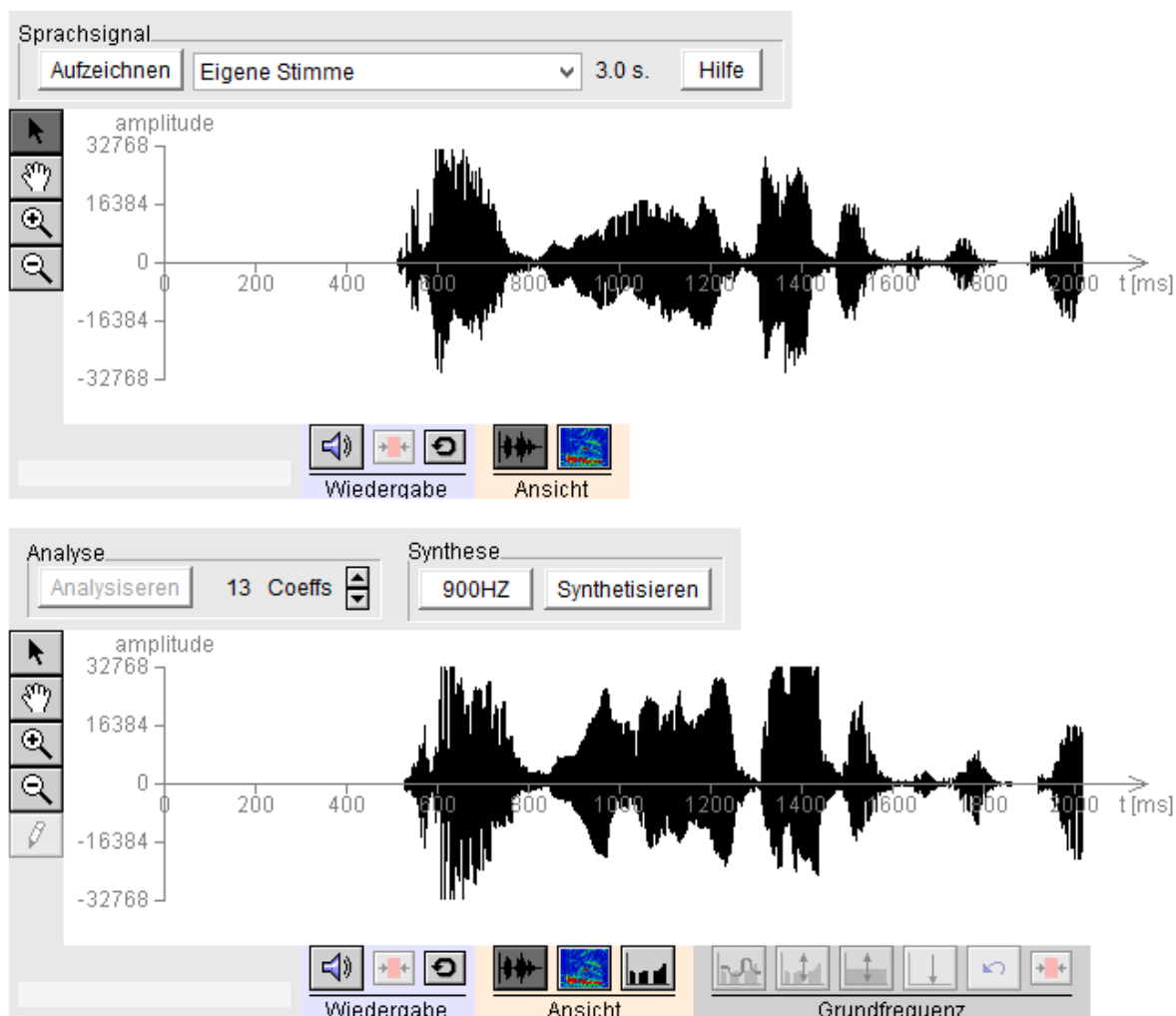
Za izradu ovog dijela domaće zadaće koristimo LPC koder (engl. *Linear Predictive Coding*) koji služi za proizvodnju ljudskog govora. Unutar LPC koda možemo odabrati već neki snimljeni govor ili snimiti vlastitu poruku. Zadatak je obavljen na već snimljenoj poruci "*Mable stood on the rock*". Kasnije smo snimili vlastitu poruku i određivali minimalan broj koeficijenta za zadovoljavajuću kvalitetu zvuka.

1. Povećanjem broja koeficijenta raspon se smanjuje, a amplituda povećava, dok kod smanjenja broja koeficijenta povećava se raspon, a amplituda smanjuje.
2. Broj koeficijenata utječe na frekvencijski spektar analiziranog signala tako što smanjenjem koeficijenta nastaju neke nove frekvencije, dok povećanjem istoga neke frekvencije nestaju. Taj koeficijent utječe na kvalitetu zvuka tako što je veći koeficijent kvaliteta zvuka je bolja ili obrnuto s manjim koeficijentom povećava se šum i pucketanje te kvaliteta zvuka opada što je direktno povezano sa novim frekvencijama. Koeficijenti linearnog filtra LPC koda u praktičnom smislu se koriste kod sintetiziranja govornog signala u procesu nastajanja ljudskog govora.
3. Izračun brzine (engl. *bitrate*) koda LPC. U svakom vremenskom okviru od 22.5 ms određuju se sljedeći parametri:
 - frekvencija pobude (6 bita)
 - jačina pobude (5 bita)
 - zvučni/bezvučni glas (1 bit)
 - koeficijent filtra (42 bita za 10 koeficijenata).

Za jedan vremenski okvir ukupno je potrebno 54 bita. Jedna sekunda sadrži $1000/22.5 = 44.444$ okvira, brzina je 2.4 kbit/s.

Frekvencija uzorkovanja u LPC-u, a i u nekim drugim koderima namijenjenim prvenstveno za ljudski govor je 8 kHz zbog toga što frekvencija uzorkovanja mora biti najmanje dva puta veća od maksimalne frekvencije ljudskog glasa koja iznosi 3400 Hz.

4. Minimalni broj koeficijenta za zadovoljavajući rezultat zvuka je 13. Slika 9 prikazuje originalni i sintetizirani govor u vremenskoj domeni. Razlike u izgledu valnog signala kod sintetiziranog govora su u tome što je amplituda veća.



Slika 9. Originalni i sintetizirani govor

5. Razlika u kvaliteti kodiranja govornog i glazbenog signala s koderom LPC postoji. LPC koder je namijenjen ljudskom govoru, dok glazba sadrži više frekvencije i amplitude koje ljudski govorni organ ne može proizvesti, pa zbog toga LPC koder ne može efikasno sintetizirati zvukove glazbe.
6. Povećanjem broja LPC koeficijenata do 30 ne možemo dobiti zadovoljavajuću kvalitetu sintetiziranog zvuka Beethoven datoteke zbog prije navedenih razloga.

3 Usporedba formata za kodiranje zvuka

U ovom dijelu zadaće uspoređujemo formate za kodiranje zvuka na već snimljenim datotekama i na drugačijim formatima naše datoteke iz prvog zadatka.

1. U tablici 1 nalaze se podaci za usporedbu formata za kodiranje zvuka. Datoteka Fox.aiff je najveća, ali to je i očekivano jer je to original. Fox_gsm.wav datoteka je najmanja, ali zbog toga je kvaliteta najlošija. Druga najveća je Fox.pcm.wav koja ima 385 kb i iako je dosta velika to opravdava s visokom kvalitetom zvuka. Fox_mp3.mp3 datoteka je vrlo mala i dobila je visoke ocjene, što bi se dalo zaključiti da ima najbolji odnos kvalitete i veličine.

Naziv datoteke	MOS	DMOS	Veličina [kb]	Bit rate [kbps]
Fox.aiff (original)	5	5	577	258
Fox_gsm.wav	4	3	39.2	17
Fox_pcm.wav	4	5	385	176
Fox_mp3.mp3	4	4	53	24

Tablica 1. Usporedba formata za kodiranje zvuka

2. U tablici 2 nalaze se podatci za usporedbu formata za kodiranje zvuka nad snimljenom datotekom iz 1. zadatka. Kao i u prijašnjoj tablici, možemo vidjeti da je najbolja kvaliteta originala, ali i da je veličina najveća. Datoteka Moje_mp3.mp3 ima najbolji odnos subjektivnih ocjena i veličine datoteke, dok datoteka Moje_gsm.wav ima najlošiji odnos subjektivnih ocjena i veličine datoteke.

	MOS	DMOS	Veličina [kb]	Bit rate [kbps]
Moje_gsm.wav	3	3	29.1	71
Moje_pcm.wav	4	5	572	1411
Moje_mp3.mp3	4	4	9.91	24

Tablica 2. Usporedba formata za kodiranje zvuka nad snimljenim datotekama

3. Formati za kodiranje zvuka. Za klasičnu telefonsku mrežu koristi se G.711 kodiranje, poznato još i kao PCM (engl. *Pulse Code Modulation*). Za bežični telefon koristi se G.729 kodiranje. Za VoIP se također može koristiti G.729 kodiranje.