

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

VIŠEMEDIJSKE KOMUNIKACIJE

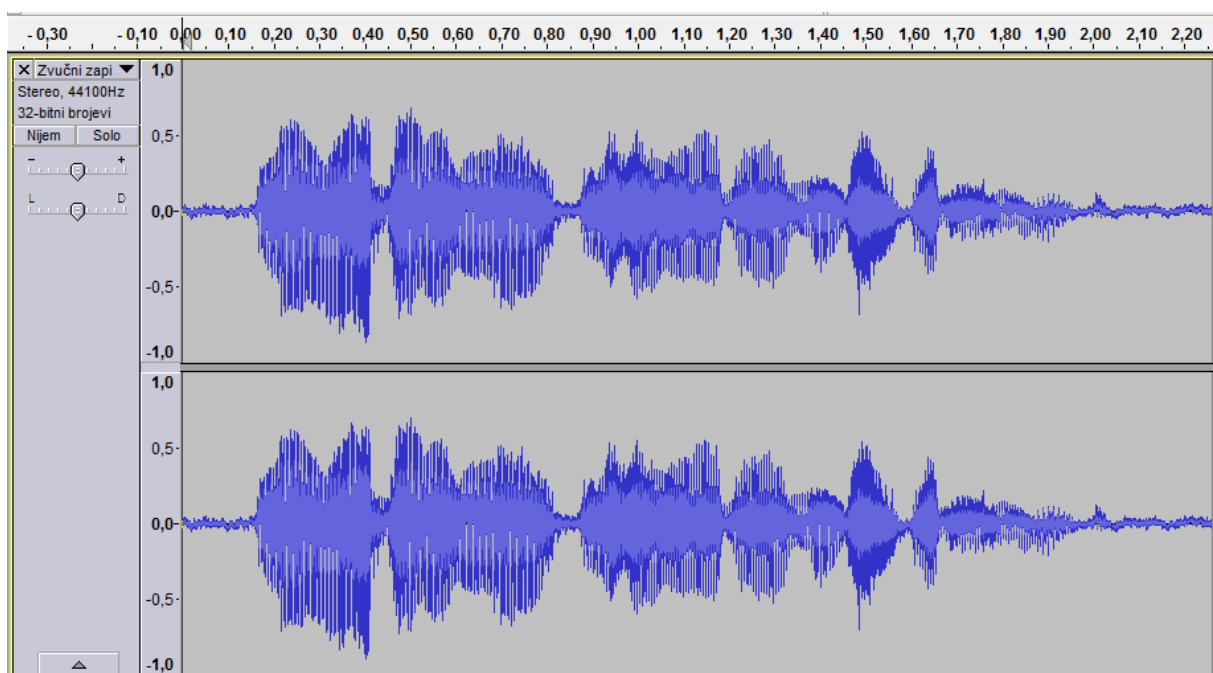
1. DOMAĆA ZADAĆA

Lorena Svržnjak
0036474442

Zagreb, travanj, 2016.

1. Vremenska i frekvencijska analiza govora

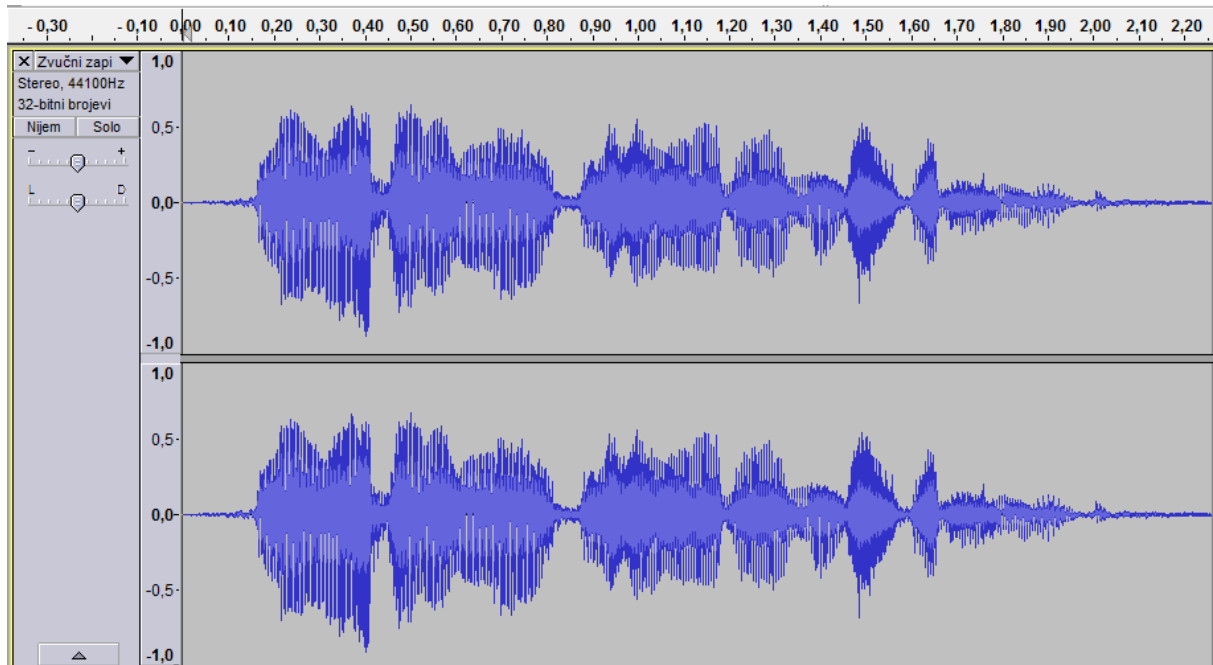
U ovom dijelu zadatka koristi se alat Audacity za analizu zvuka. U gornjem lijevom kutu u prozoru Audacityja klikne se na gumb sa crvenim kružićem „Snimi (R)“ te se snimi govorni signal. Slika 1-1 prikazuje snimljenu rečenicu „Moje ime je Lorena Svržnjak“, a Slika 1-2 istu snimku s provedenom opcijom uklanjanja šuma. Šum se otklanja tako da se na snimci označi dio gdje je samo šum, zatim se klikne na „Učinak -> Uklanjanje šuma...“ nakon čega se otvara prozor u kojem se klikne na gumb „Definicija šuma“. Nakon toga označi se cijela snimka i klikne se na „Učinak -> Ponovi Uklanjanje šuma“ nakon čega je šum uklonjen.



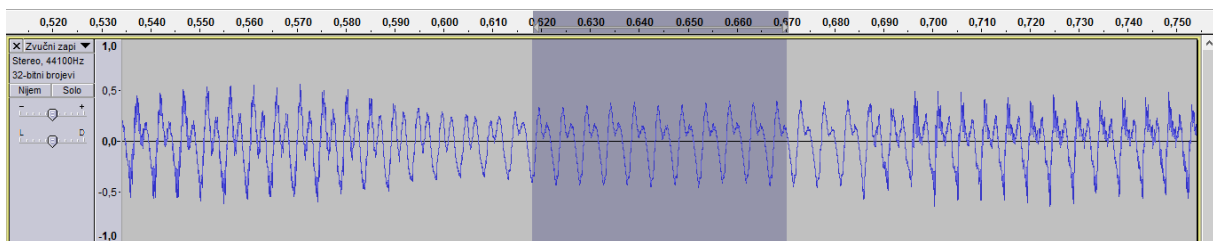
Slika 1-1 Snimljena rečenica sa šumom

1. Valni oblik zvučnog glasa „m“ prikazan je na Slika 1-3, a bezvučnog glasa „s“ na Slika 1-4. Vidljivo je kako zvučni glas ima nekoliko puta veću amplitudu od bezvučnog i periodičan je (obzirom da nastaje titranjem glasnica) za razliku od bezvučnog kod kojeg se ne može ustanoviti nikakva pravilnost (jer je formiran prolaskom struje glasa kroz govorni organ).
2. Da bi snimku prikazali u frekvencijskoj domeni potrebno je kliknuti na „Prouči -> Graf dB/Hz...“. Signal glasa „m“ u frekvencijskoj domeni prikazan je na Slika 1-5, a signal glasa „s“ na Slika 1-6. Ni u frekvencijskoj domeni bezvučnog glasa „s“ ne uočavaju se pravilnosti, dok se kod zvučnog glasa „m“ vide karakteristične frekvencije

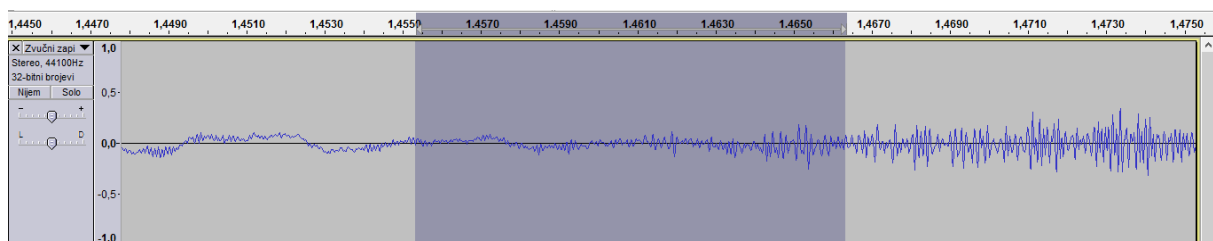
ili formanti koji karakteriziraju zvučne glasove. Formanti su intenzitetski naglašeni dijelovi spektra koji su rezultat rezonancije u rezonantnim šupljinama.



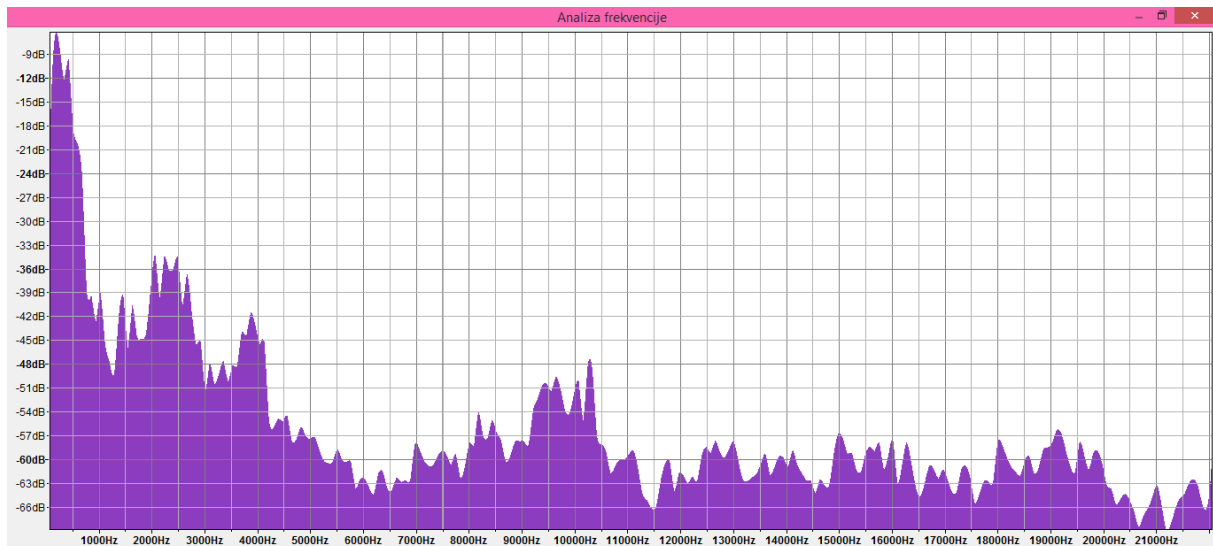
Slika 1-2 Snimljena rečenica s uklonjenim šumom



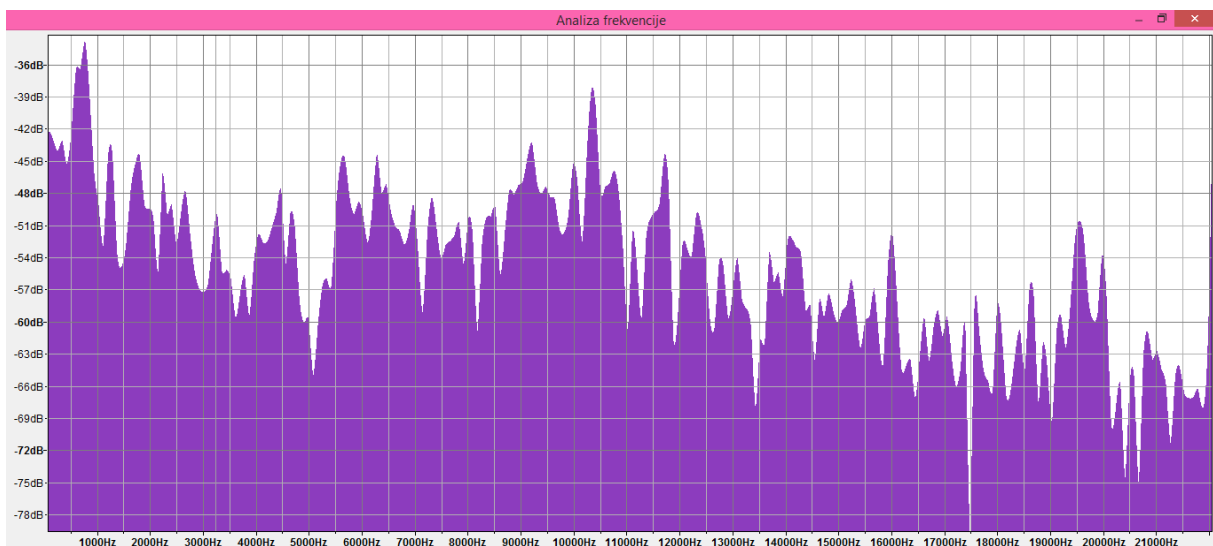
Slika 1-3 Valni oblik zvučnog glasa "m"



Slika 1-4 Valni oblik bezvučnog glasa "s"



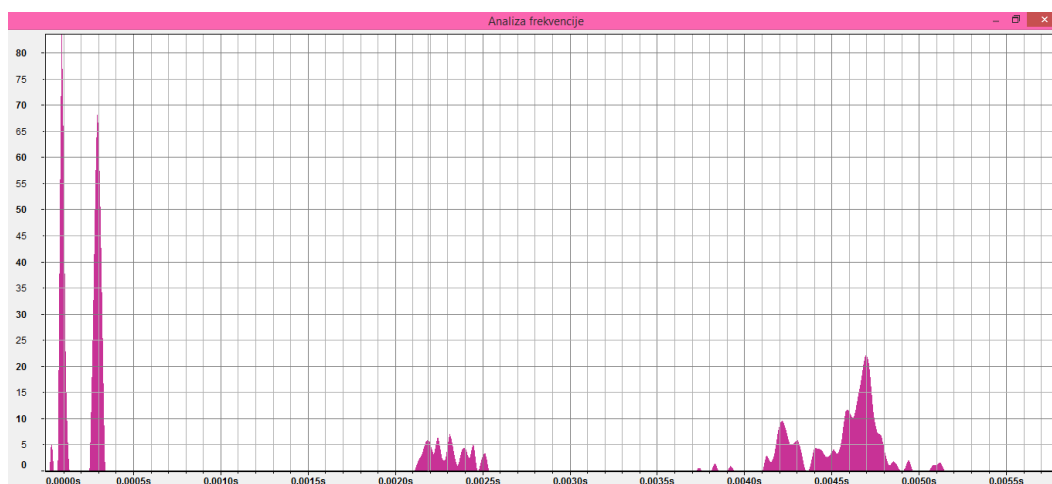
Slika 1-5 Signal glasa "m" u frekvencijskoj domeni



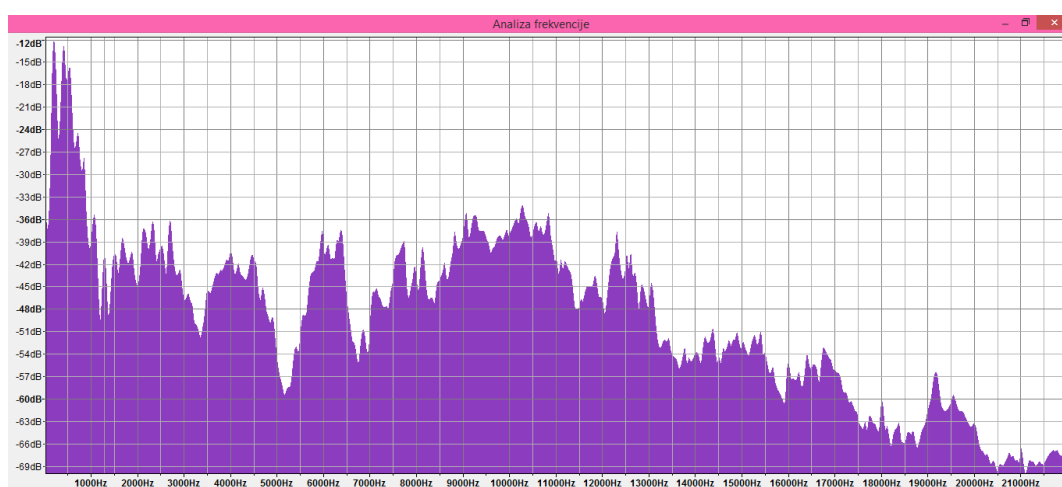
Slika 1-6 Signal glasa "s" u frekvencijskoj domeni

3. Formanti se kod signala mogu lakše uočiti ako u prozoru s frekvencijskom domenom pod „Algoritam“ odaberemo „Enhanced Autocorrelation“. Slika 1-7 prikazuje formante glasa „e“ koji se pojavljuju na frekvencijama 195 Hz, 203 Hz, 218 Hz, 237 Hz, 410 Hz, 457 Hz, 3408 Hz. Prvi formant je onaj koji se nalazi na najnižoj frekvenciji, u ovom slučaju 195 Hz. Prosječni prvi formant iz Tablice 1 za ženski glas „e“ je 214 Hz što je približno snimljenom. Drugi formant nalazi se na sljedećoj frekvenciji od prvog formanta, u ovom slučaju 203. Prosječni drugi formant iz Tablice 1 je 536 Hz što je puno veće od snimljenog.

4.

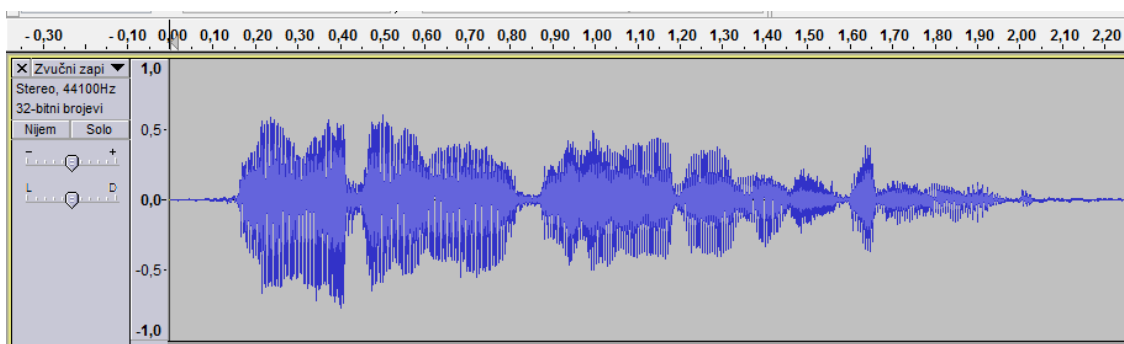


Slika 1-7 Formanti glasa "e"



Slika 1-8 Govorni signal u frekvencijskoj domeni

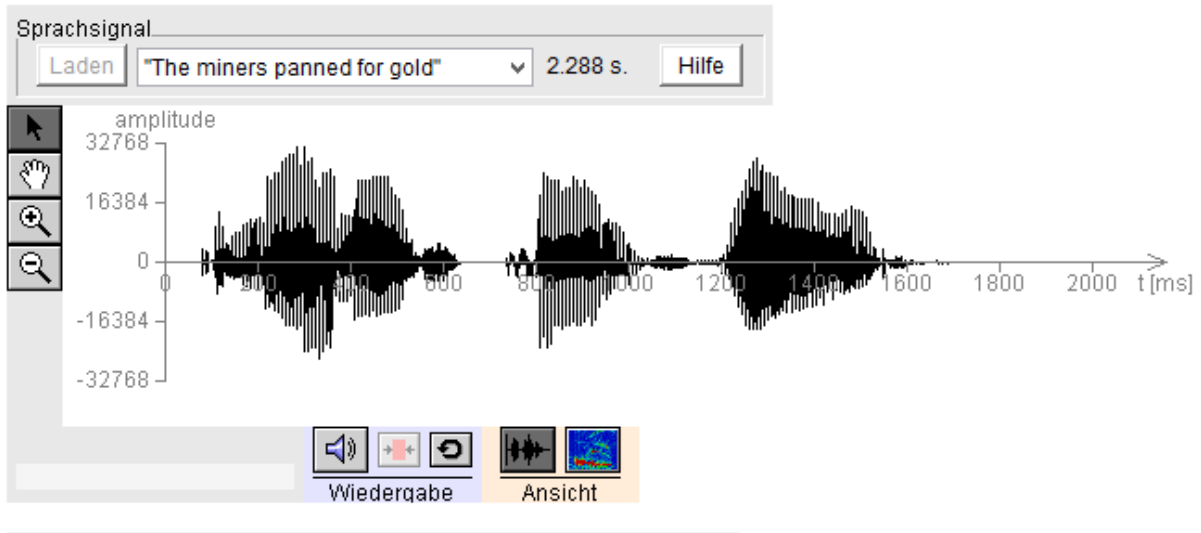
5. Cijeli govorni signal u frekvencijskoj domeni prikazan je na Slika 1-8. Ne uočavaju se pravilnosti, signal je neperiodičan. Frekvencije su vidljivo jače u području do 1000 Hz što je unutar raspona kodiranja ljudskog govora u analognoj telefoniji.
6. Govor je prije provedbe niskopojasnog filtriranja bio malo čišći i više frekvencije su se jasnije čule, dok je nakon niskopojasnog filtriranja govor prigušen.



Slika 1-9 Snimka govora nakon provedbe niskopojasnog filtriranja

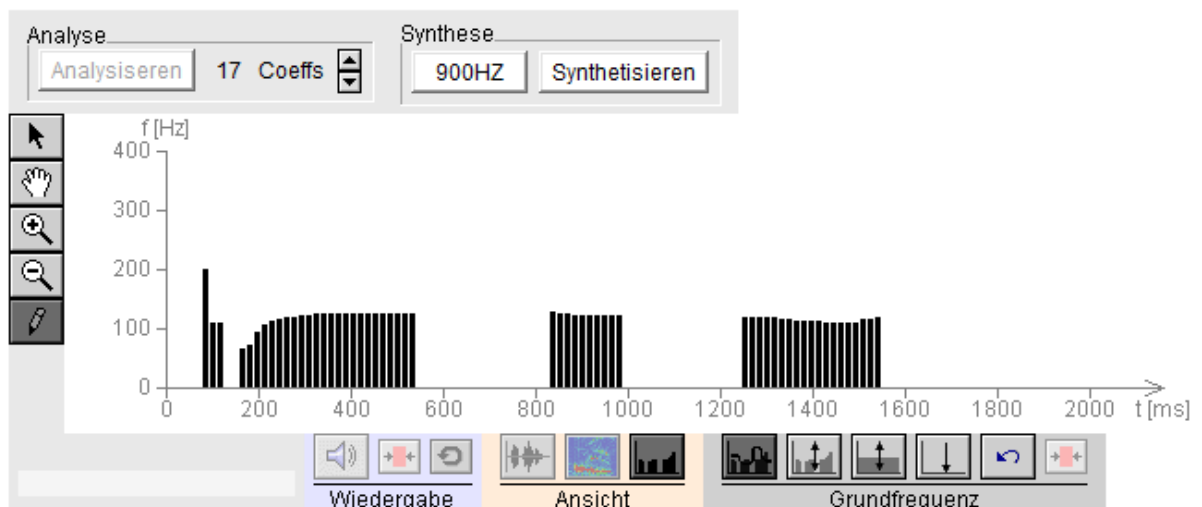
2. Simulacija kodera LPC

U ovom dijelu zadatka koristi se LPC (engl. *Linear Predictive Coding*) koder koji spada u skupinu kodera zasnovanih na modelu. Odabrana govorna fraza koja će se analizirati je „The miners panned for gold“.

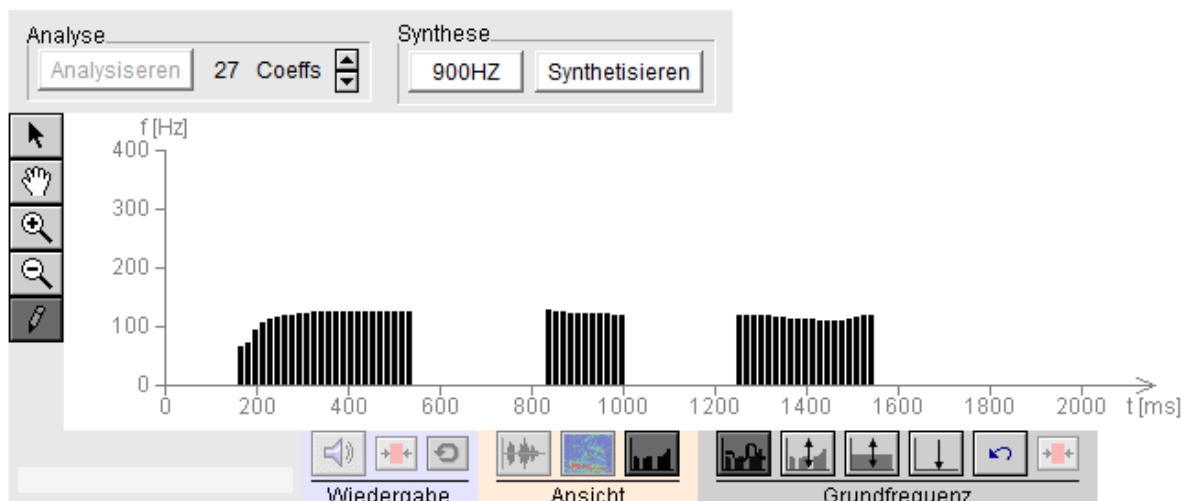


Slika 2-1 Vremenska analiza fraze „The miners panned for gold“

1. Kako povećavamo broj koeficijenata raspon frekvencijskog spektra signala se smanjuje, odnosno neke frekvencije nestaju (i obratno – smanjenjem broja koeficijenata pojavljuju se neke nove frekvencije), dok amplituda uglavnom ostaje ista za sve koeficijente, ali je veća od originalne. Opisano je vidljivo na Slika 2-1 i Slika 2-2.



Slika 2-2 Frekvencijski spektar signala analiziran sa 17 koeficijenata

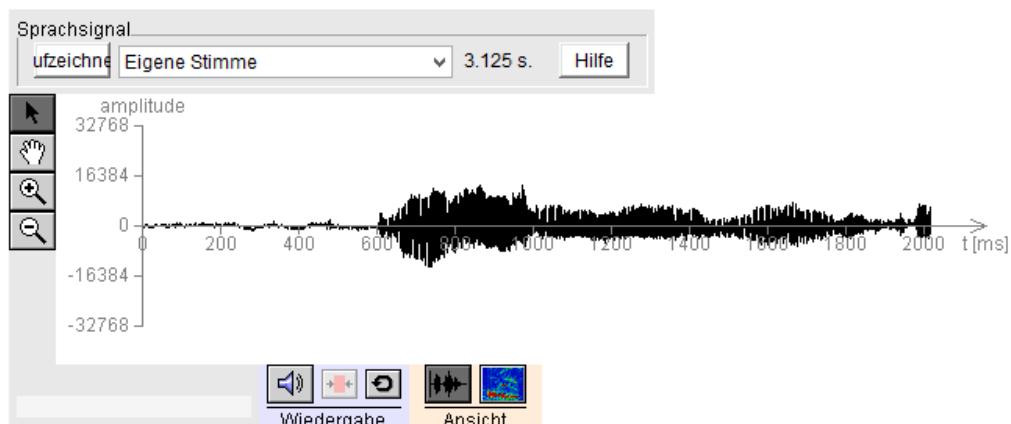


Slika 2-3 Frekvencijski spektar signala analiziran sa 27 koeficijenata

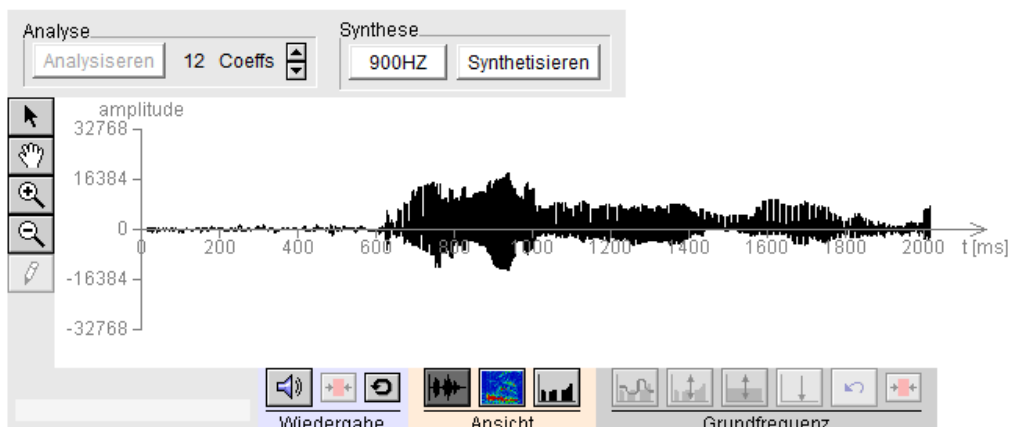


Slika 2-4 Razlika amplitude originalnog i analiziranog signala

2. Kvaliteta zvuka govora bolja je što je veći broj koeficijenata. Smanjenjem broja koeficijenata dolazi do pucketanja i pojave šuma u sintetiziranom zvuku, a razlog tome je da se smanjenjem broja koeficijenata pojavljuju neke frekvencije koje unose šum pa zvuk postaje manje prepoznatljiv. Korištenjem samo jednog koeficijenta rečenica je neprepoznatljiva, šum je u potpunosti prekrije.
3. Svaki vremenski okvir duljine 22.5 ms sadrži parametre: frekvenciju (6 bita) i jačinu pobude (5 bita), koeficijent filtra (10 koeficijenata -> 42 bita) te zvučni ili bezvučni glas (1 bit). Dakle za jedan vremenski okvir potrebno je 54 bita. U jednoj sekundi je $1000/22.5 = 44.444$ okvira pa je brzina jednaka $54 \cdot 44.444 = 2.4 \text{ kbit/s}$. Frekvencija uzorkovanja je 8kHz iz razloga što ona mora biti najmanje dvostruko veća od maksimalne frekvencije ljudskog glasa (3400Hz).
4. Kvaliteta zvuka je zadovoljavajuća kad je sintetiziran s 12 koeficijenata. Slika 2-5 i Slika 2-6 prikazuju originalni i sintetizirani govor. Sintetizirani govor ima malo veće amplitude i neke dodatne frekvencije koje unose šum.



Slika 2-5 Originalni govor



Slika 2-6 Govor sintetiziran s 12 koeficijenata

5. Kvaliteta u kodiranog glazbenog signala puno je lošija od kvalitete kodiranog govornog signala. Glazbeni signal ne možemo aproksimirati govornim zato što sadrži visoke frekvencije i amplitude koje ljudi ne mogu proizvesti.
6. Čak ni sa 30 LPC koeficijenata ne dobije se zadovoljavajuća kvaliteta sintetiziranog zvuka *Beethoven* datoteke.

3. Usporedba formata za kodiranje zvuka

Naziv datoteke	MOS	DMOS	Veličina [kb]	Bit rate [kbps]*
<i>Fox.aiff (original)</i>	3	5	577	258
<i>Fox_gsm.wav</i>	2	3	39.2	17
<i>Fox_pcm.wav</i>	3	5	385	176
<i>Fox_mp3.mp3</i>	3	5	53	24

1. Subjektivna kvaliteta je proporcionalna veličini datoteke osim u slučaju *Fox_mp3.mp3* što znači da ta datoteka ima najbolji omjer veličine i kvalitete.
- 2.

Naziv datoteke	MOS	DMOS	Veličina [kb]	Bit rate [kbps]*
<i>audacity_snimka_pcm.wav</i>	5	5	390	174.5
<i>audacity_snimka_gsm.wav</i>	2	2	19.8	8.85
<i>audacity_snimka_mp3.wav</i>	3	4	6.8	3

Kvaliteta originala je najbolja, ali je i daleko najveće veličine. Mp3 opet ima najbolji omjer veličine i kvalitete, dok je gsm datoteka dosta lošije kvalitete od originala.

3. - klasična telefonska mreža: G.711 (PCM)
- bežični telefon, VoIP: G.729