

SVEUČILIŠTE U ZAGREBU  
**FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA**

Višemedijske komunikacije

# **1. DOMAĆA ZADAĆA**

Vladimir-Mario Obelić

# Sadržaj

Uvod.....	1
1    Vremenska i frekvencijska analiza govora .....	1
1.1    Odgovori.....	1
2    Simulacija kodera LPC .....	5
2.1    Odgovori.....	6
3    Usporedba formata za kodiranje zvuka.....	9
3.1    Odgovori.....	9
Zaključak.....	10

# Uvod

Ovom vježbom proći ćemo kroz osnove kodiranja zvuka, govora i glazbe te vidjeti neke zakonitosti koje uočavamo analizirajući, sintetizirajući i kodirajući različite signale. Presudno je korištenje programske potpore za analizu i sintezu zvuka kao i za njegovo kodiranje. U vježbi koristiti ćemo alat „Audacity“ otvorenog koda koji se često koristi upravo za neke jednostavnije manipulacije i analize signala. Nadalje koristiti ćemo se i implementacijom jednog koda u javi.

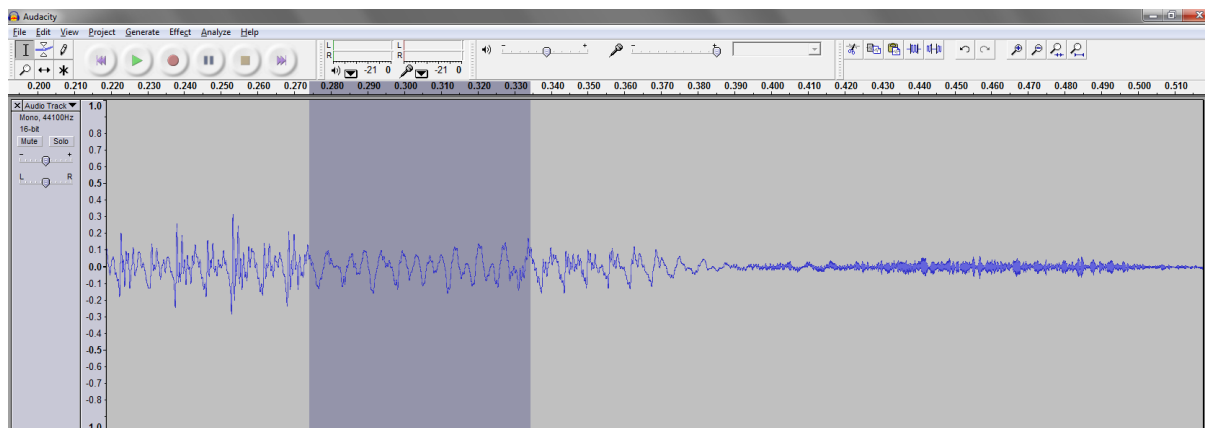
## 1 Vremenska i frekvencijska analiza govora

Vremensku i frekvencijsku analizu govora vršiti ćemo u Audacity alatu za analizu i obradu zvuka. Konkretno koristi se verzija alata 1.2.6 kako je priloženo u materijalima predmeta.

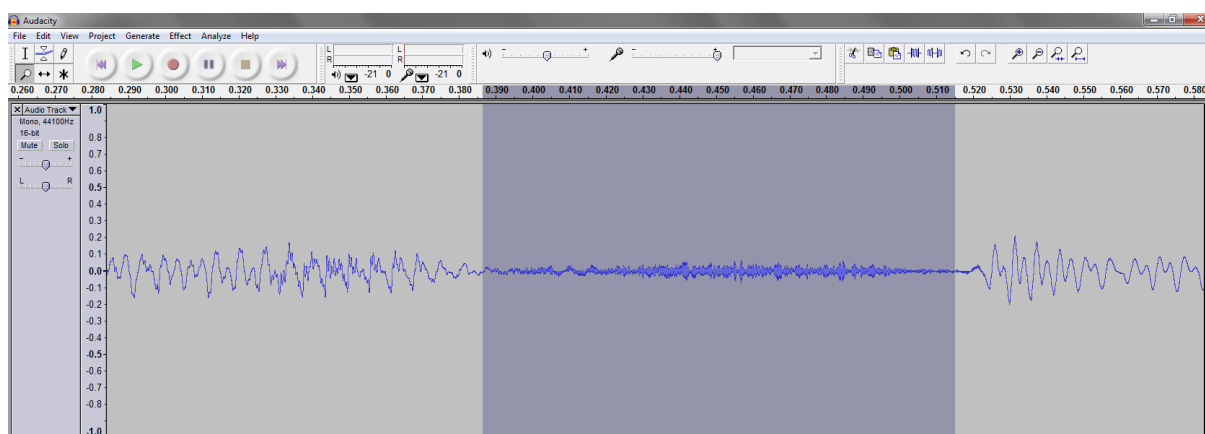
### 1.1 Odgovori

1. Pokretanje Audacity alata
2. Snimanje vlastitog govornog signala vrši se otvaranjem programskog alata, te pritiskom lijeve tipke miša na „record“ gumb. Sve što je potrebno dalje je aktivirati mikrofonski uređaj na računalo i snimiti svoj glas. Podrazumijeva se da računalo posjeduje priključen mikrofonski uređaj. U Audacity-u se otvara novi prozor sa snimljenim materijalom u „waveform“ obliku. U tom obliku pretpostavlja se da se snimljeni materijal nalazi u vremenskoj domeni. Od preostalih oblika u kojima možemo prikazati snimljeni materijal, značajna je frekvencijska domena.
3. Traženje glasova u vremenskoj domeni je jednostavno. Preslušavanjem snimljenog materijala odredimo gdje se po prilici nalazi glas kojeg tražimo, te kada nađemo mjesto koje nas zanima jednostavno ga klikom miša označimo i dalje koristimo „zoom-in“. Konkretno tražilo se naći po jedan zvučni i jedan bezzvučni glas. U mom slučaju našao sam zvučni glas „N“ i bezzvučni glas „S“ što je vidljivo na slikama 1 i 2. Glavna razlika između ova dva zapisa je upravo u periodičnosti prikazanog signala, te pronalaženju pravilnosti. Vidimo da je zvučni signal pravilan tj periodičan s obzirom da se radi o glasu kojeg proizvode naše glasnice kada titraju. S druge strane, bezzvučni glas formiran je jedino strujom glasa kroz naš govorni organ i kako tu nema

karakterističnog titranja, u zapisanom signalu jednostavno nema pravilnosti, koliko god zoomirali u vremenskoj domeni.

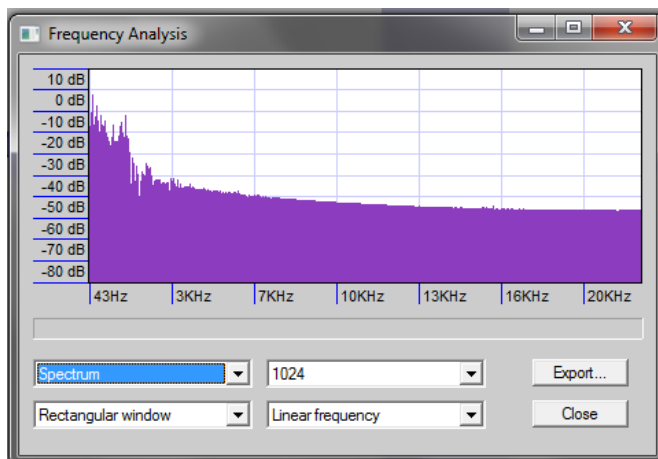


Slika 1. Zvučni glas "N"

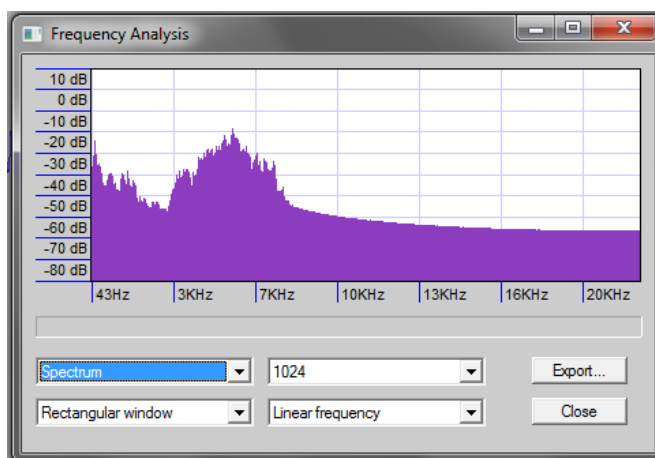


Slika 2. Bezvučni glas "S"

4. Prikaz signala odabranih glasova u frekvencijskoj domeni vidljiv je na slikama 3 i 4. Vidimo da se frekvencijskom prikazu signala glasa "N" na slici 2 neke frekvencije ističu dok se kod prikaza frekvencijskog spektra signala glasa „S“ na slici 3 značajni broj frekvencija. Zaključujemo da se na temelju frekvencijske analize signala glasa „N“ mogu približno odrediti neki značajniji formanti dok se to kod glasa „S“ ne može dobiti.

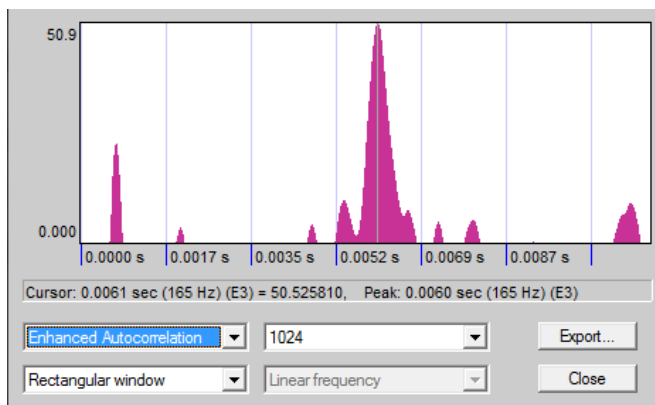


Slika 3. Zvučni glas "N" u frekvencijskoj domeni

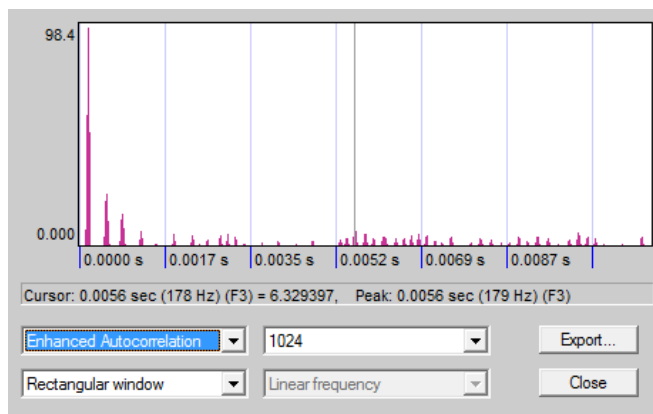


Slika 4. Bezvučni glas "S" u frekvencijskoj domeni

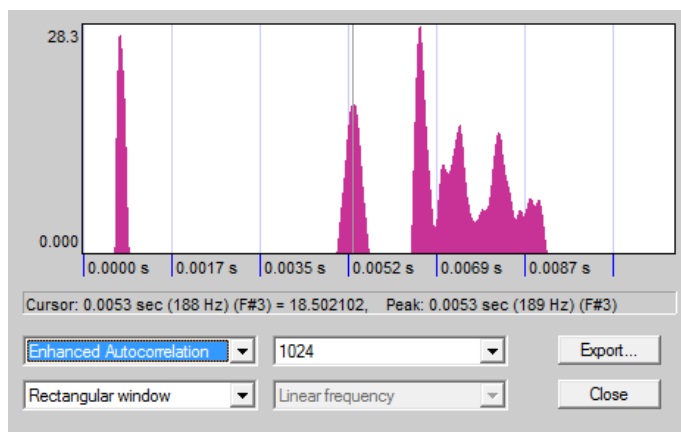
5. Određivanje formanta kod signala može se pojednostaviti u frekvencijskoj analizi korištenjem autokorelacije. Uočavamo tako značajne formante koje se pojavljuju kod signala glasa „N“ (Slika 5) na frekvencijama 1400Hz, 493Hz, 165Hz, 89Hz. Uočavamo i „problem“ kod određivanja formanta na signalu kod glasa „S“ (Slika 6). Analizom glasa „A“ (Slika 7) pak dobivamo frekvencije 1414Hz, 189Hz, 151Hz, 142Hz, 123Hz.



**Slika 5. Formanti glasa "N"**



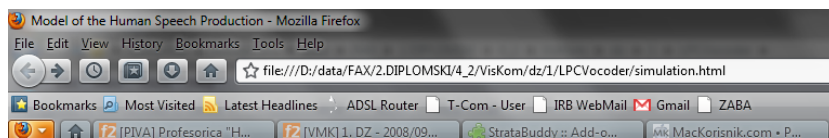
**Slika 6. Formanti glasa "S"**



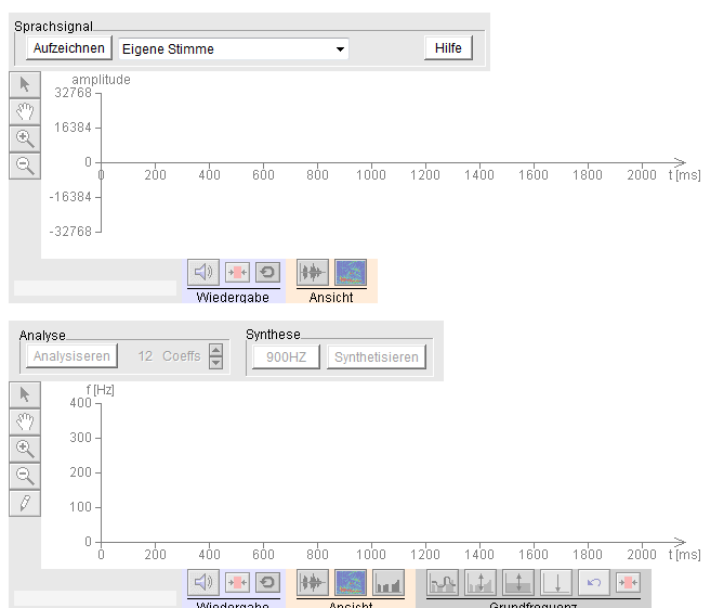
**Slika 7. Formanti glasa "A"**

## 2 Simulacija koder LPC

LPC koder (*Linear Predictive Coding*) koristi se kod modeliranja i proizvodnje govora. Ono što karakterizira ovaj koder je da se koriste jednostavni principi kod analize postojećeg govornog signala. Sam koder temelji se na principima na kojima govorni organi čovjeka funkcioniraju, te se upravo zbog toga koriste parametri poput određivanja da li je izgovoreno zvučni ili bezzvučni glas, te frekvencija kojom naše glasnice titraju. Pomoću tih parametara moguće je sintetizirati ljudski govor. Ovdje koristimo implementaciju koder LPC u javi. Točnije radi se o java appletu kojega pokrećemo u web pregledniku odlaskom na adresu stranice na kojoj se sam applet nalazi. Također moguće je koristiti arhiviranu kopiju appleta na računalu kao u našem slučaju. Dakle pokrećemo `simulation.html`, otvara se stranica sa java appletom gdje je implementiran koder LPC. Na slici Slika 8. Java applet sa LPC koderom vidimo kako izgleda početno pokretanje appleta. Vidljive su dvije sekcije. Prva sekcija je za analizu signala koji možemo samostalno snimiti, ili kojeg možemo pokrenuti iz padajućeg izbornika. Sljedeća sekcija služi sintetizaciji signala temeljenog na onom signalu kojeg smo prethodno analizirali, te postavljanjem LPC parametara.



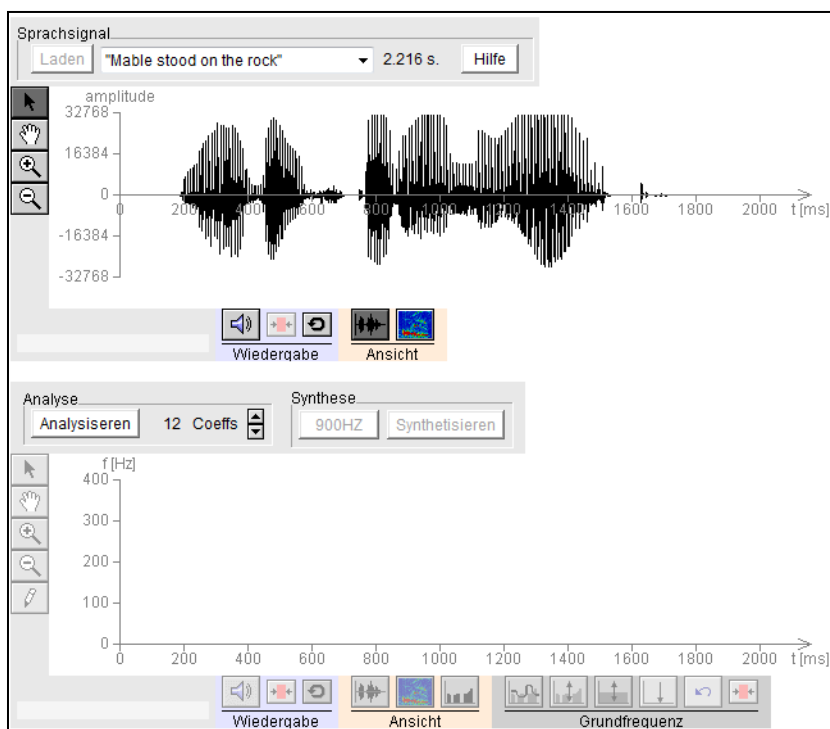
### Model of the Human Speech Production



[Close the simulation](#)

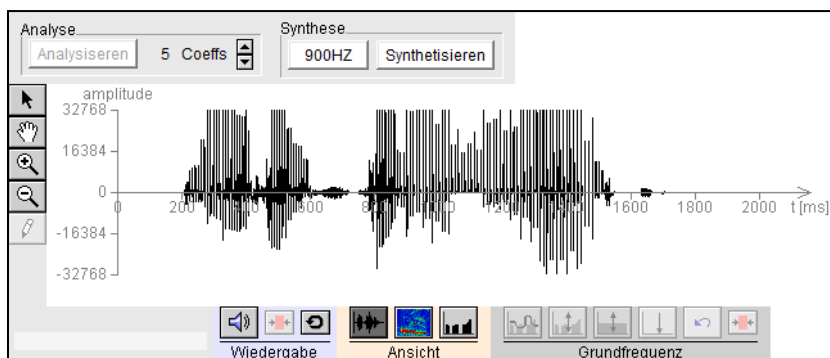
Slika 8. Java applet sa LPC koderom

## 2.1 Odgovori



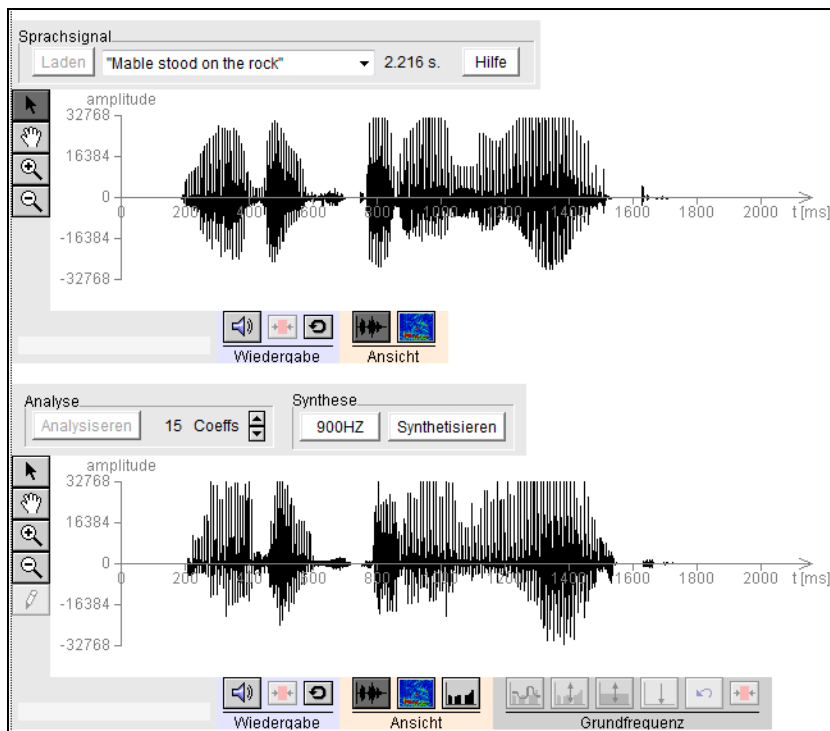
Slika 9. Vremenska analiza fraze "Mable stood on the rock" u simulatoru kodera LPC

4. Broj koeficijenata utječe na samu kvalitetu zvuka. Prilikom korištenja manjeg broja koeficijenata sintetizirani glas postaje sve manje prepoznatljiv i razumljiv te dolazi do pojave šumova i pucketanja. Povećanjem broja koeficijenata u spektru nestaju neke frekvencije, odnosno smanjenjem broja koeficijenata dolazi do pojave nekih novih frekvencija. Upravo te nove frekvencije koje se pojavljuju smanjenjem broja koeficijenata prilikom analize dovodi do tih šumova koje smo opazili preslušavanjem sintetiziranog glasa. Na slikama Slika 10 i Slika 11 vidimo usporedbu grafova u vremenskoj domeni korištenjem različitih koeficijenata. Vidimo kako je na slici Slika 11 sintetizirani signal glasa vjerodostojniji odnosno sličniji originalnom uzorku, dakle veće kvalitete.



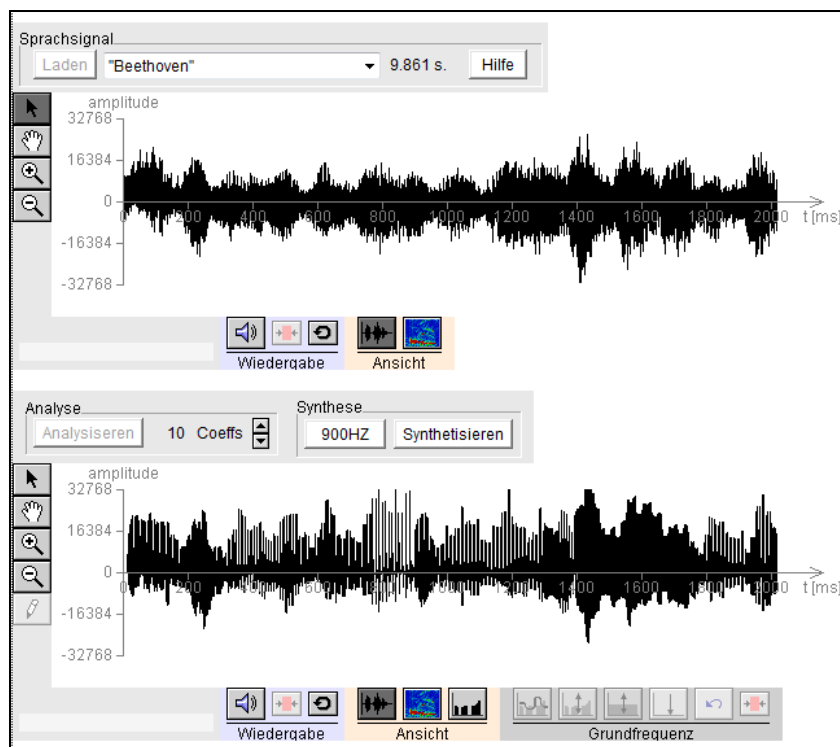
Slika 10. Sintetizirani uzorak glasa sa prethodne slike korištenjem 5 koeficijenata u analizi



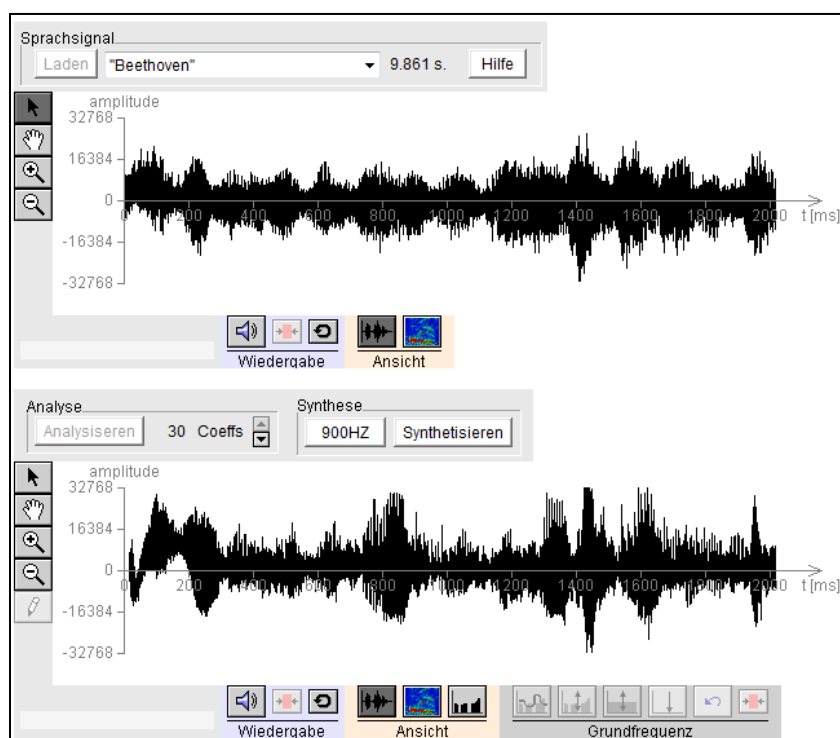


Slika 11. Sintetizirani uzorak glasa korištenjem 15 koeficijenata u analizi

5. Izračun brzine (*bitrate*): Svaki vremenski okvir od 22.5ms za 10 korištenih koeficijenata sadrži ukupno  $n = 54$  bita. Jedna sekunda sadrži  $k = 1000/22.5 = 44.444$  okvira. Prema tome imamo  $B = k * n = 44.444 * 54 = 2.4$  kbit/s
6. Razlika između kodiranja govornog signala i glazbenog signala je očigledna. Do takve razlike dolazi zbog karakteristika i svojstava samog LPC koder. LPC koder je dizajniran kako bi simulirao konfiguraciju govornih organa koristeći se parametrima koji direktno utječu na govorni signal. Glazbeni signal u sebi sadrži frekvencije i amplitude koje ljudski govorni organ ne može stvarati, te prema tome bez obzira kakve parametre LPC koderu zadajemo ne može se efikasno sintetizirati glazbeni signal LPC koderom. Na slikama Slika 12 i Slika 13 vidljivo je da ma kako veliki koeficijent za LPC koder zadajemo prilikom analize, sintetizirani signal je prilično različit od originala, odnosno imamo šumove. Za snimljeni materijal, koji je govorni imamo situaciju istu kao na slikama Slika 10 i Slika 11. Povećanjem koeficijenata za LPC koder dobivamo vjerodostojniju sintetizaciju originalnog zvuka.



Slika 12. Sintetizirani glazbeni uzorak "Beethoven" s 10 koeficijenata



Slika 13. Sintetizirani glazbeni uzorak "Beethoven" s 30 koeficijenata

### 3 Usporedba formata za kodiranje zvuka

U paketu audio-samples.rar nalaze se datoteke različitog formata, odnosno korištenog kodiranja, u kojima je zapisan isti zvučni zapis. Ono što nas zanima je analiza svake od pojedinih datoteka, te usporedba i performanse svakog od korištenih kodiranja. Za preslušavanje svake pojedine datoteke ponovo ćemo koristiti alat Audacity. Postupak je jednostavan. Otvaramo željenu datoteku i pokrećemo reprodukciju lijevim klikom miša na „play“ tipku. Preslušavamo svaku pojedinu datoteku i bilježimo dojmove. Potrebno je odrediti dojmove za parametre MOS i DMOS. Respektivno prvi parametar služi određivanu općeg dojma, dok je drugi parametar isključivo za dojam kakvu degradaciju osjetimo prilikom reprodukcije.

#### 3.1 Odgovori

2. Rezultati su dani u Tablica 1. Usporedba formata za kodiranje zvuka. Dodatno potrebno je još izračunati brzinu u kbit/s. Za taj jednostavni proračun potrebno je poznavati veličinu pojedine datoteke, te trajanje zapisa u sekundama. Brzina koju tražimo omjer je te veličine datoteke, u bitovima, i trajanja signala.

Naziv datoteke	MOS	DMOS	Veličin [kB]	Brzina [kbit/s]
Fox.aiff	4	4	578	$578 \cdot 8 / 17.88 = 258$
Fox_gsm.wav	3	4	79	35
Fox_adpcm.wav	3	4	195	87
Fox_pcm.wav	4	4	386	172
Fox.mp3	4	4	70	31

Tablica 1. Usporedba formata za kodiranje zvuka

3. Sam zapis u aiff datoteci ipak nije idealne kvalitete pa ne dobiva odlične subjektivne ocjene. Zaključke temeljimo na relativnom odnosu kvaliteta između zapisa u originalnoj datoteci i ostalim formatima. Mp3 format najbolje se nosi kada gledamo omjer veličine datoteke i subjektivnog dojma kvalitete istog. GSM kodiranjem postiže se također vrlo mala veličina datoteke međutim na trošak kvalitete s obzirom da je datoteka dobila najlošiji subjektivni dojam. Ipak, GSM je učinkovitiji od adpcm formata s obzirom da su i adpcm i gsm sličnog subjektivnog dojma, no gsm i dalje prednjači u maloj veličini datoteke. PCM format nosi sličnu kvalitetu originalu no sa najvećom veličinom datoteke ako ne ubrojimo original.

## **Zaključak**

Moguće je saznati veoma puno korištenjem praktičnih vježbi, te analiziranjem i modifikacijom signala u za to predviđenim alatima. Ljudski govorni organ složene je građe no upoznati smo sa osnovnim principom rada istoga, te je to bilo primijenjeno kod dizajniranja koda predviđenih za baratanje i sintetiziranje upravo ljudskog glasa. Također vidimo kako se u stvarnosti ponašaju različiti koderi, te kolika je njihova efikasnost u ovisnosti o tome za koju svrhu se koriste.