

Prepoznavanje Govora

Semplovanje / PCM / WAV format

Priroda zvuka

- Talas, dobijen superpozicijom (prostim sabiranjem) elementarnih talasa – harmonika – koji mogu da se predstave kao obične sinusoide.
- Svaki harmonik ima svoju frekvenciju i amplitudu.
- Demonstracija: [link](#)

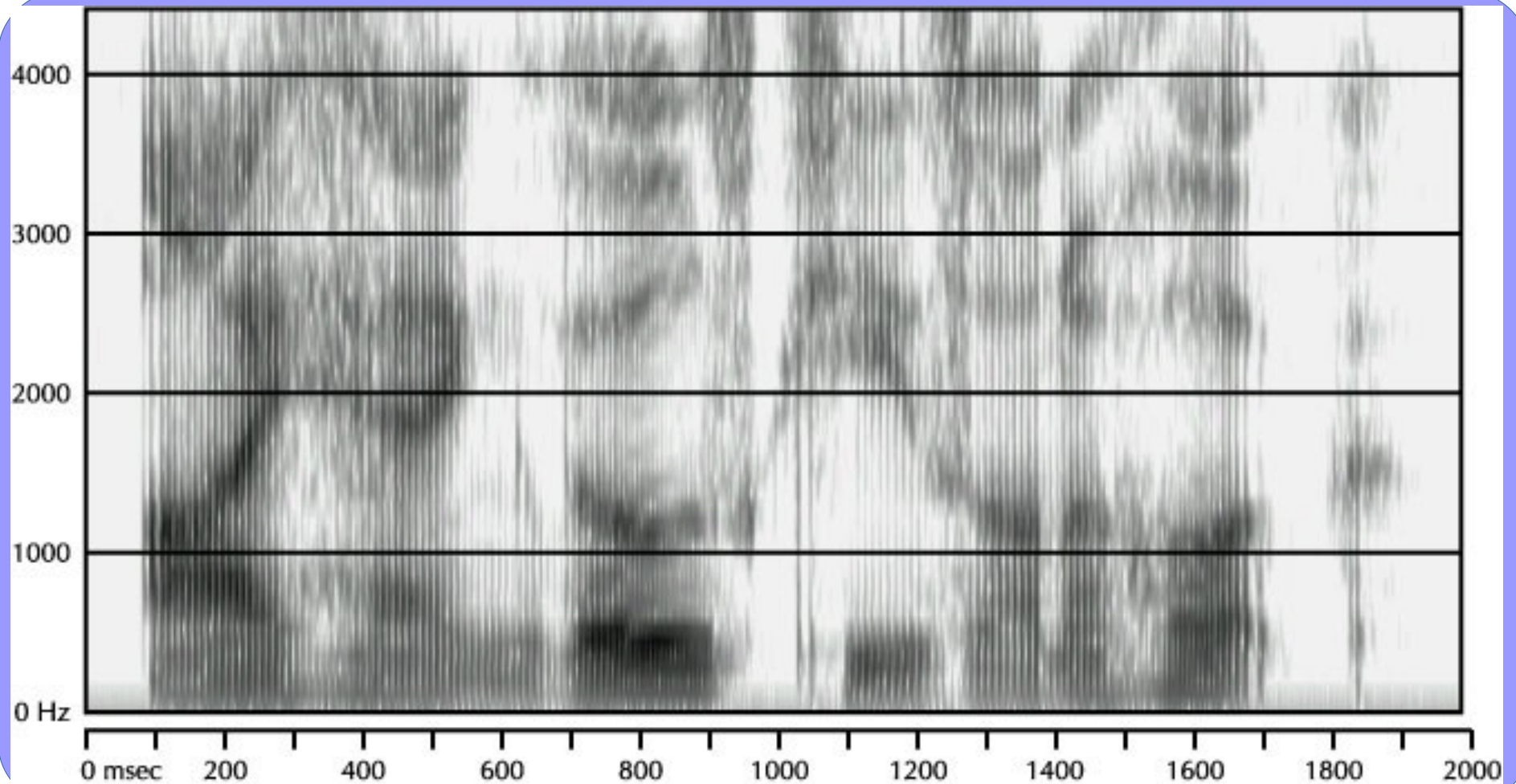
Priroda zvuka

- Ljudsko uho prirodno može da čuje zvukove na frekvencijama od 20Hz do oko 40kHz.
- Kada hoćemo da analiziramo prirodu zvučnog signala, jedan način da to uradimo je pomoću spektrograma. On nam omogućava da vidimo zvučni signal u tri dimenzije:
 - Na horizontalnoj osi: vreme
 - Na vertikalnoj osi: frekvencija
 - Intenzitet tačke na grafiku: amplituda

Priroda zvuka

- Ljudsko uho prirodno može da čuje zvukove na frekvencijama od 20Hz do oko 20kHz.
- Kada hoćemo da analiziramo prirodu zvučnog signala, jedan način da to uradimo je pomoću spektrograma. On nam omogućava da vidimo zvučni signal u tri dimenzije:
 - Na horizontalnoj osi: vreme
 - Na vertikalnoj osi: frekvencija
 - Intenzitet tačke na grafiku: amplituda

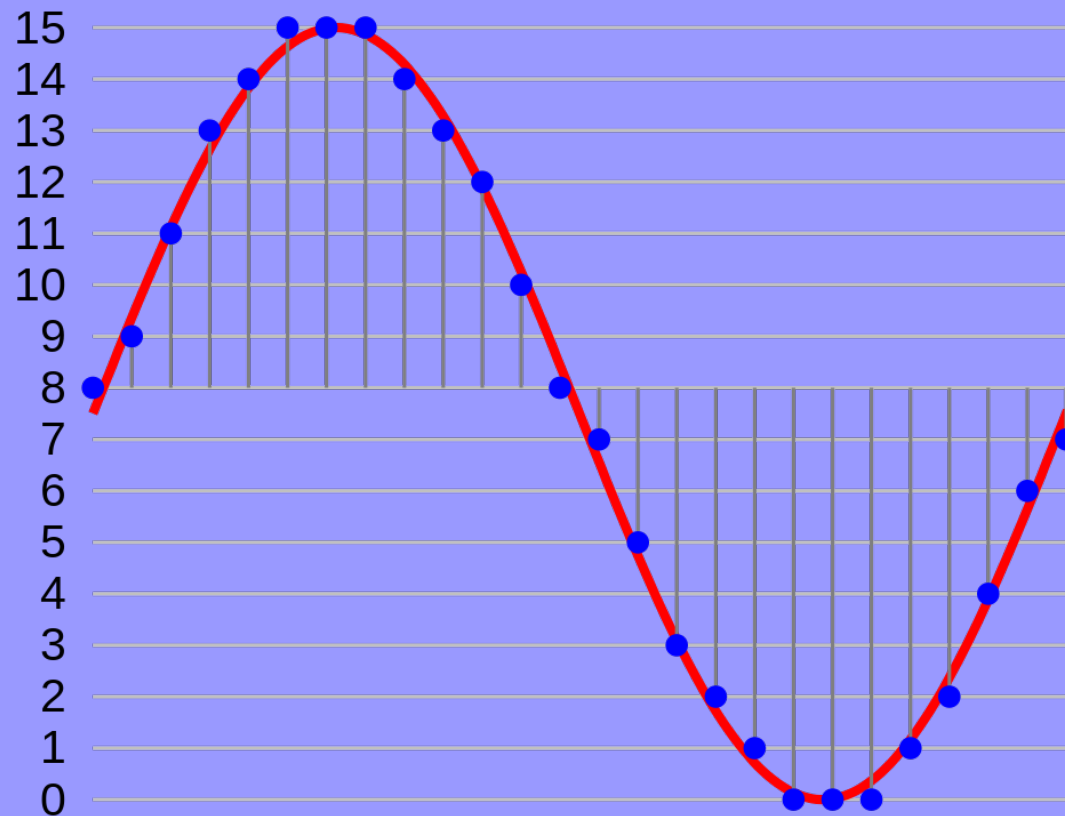
Primer spektrograma



Semplovanje

- Odabiranje vrednosti analognog signala u određenim intervalima. Semplovanje se često radi u jednakim intervalima, mada može da bude i promenljivo.
- Vrednost koja se dobija se kvantizuje, tj. svodi se na diskretnu vrednost koja pripada domenu koji smo unapred odredili.
- Kvantizacija može da se radi na više načina.

Primer kvantizacije



Niquist-ova teorema

- Da bi se pravilno semplovao signal frekvencije f , neophodno je da frekvencija semplovanja bude najmanje $2 \cdot f$.
- Pošto ljudski sluh razaznaje zvukove do frekvencije $\sim 22\text{kHz}$, za potrebe prepoznavanja govora sa najvećim mogućim kvalitetom nam je adekvatna frekvencija semplovanja od 44kHz .

Vežba

- Kao analogni signal imamo sinusoidu frekvencije 1KHz. Amplituda je 2 (opseg -1, 1)
- Definirati kvantizaciju i konstruisati tabelu dobijenih sempla za jedan period, uz sledeća ograničenja:
 - Frekvencija semplanja 4KHz; 2 bita po semplu
 - Frekvencija semplanja 4KHz; 4 bita po semplu
 - Frekvencija semplanja 8KHz; 4 bita po semplu
 - Frekvencija semplanja 16KHz; 8 bita po semplu
- Na osnovu dobijenih sempla, plotovati dobijeni digitalni signal.

WAV format

- WAV koristi RIFF specifikaciju multimedijalnih fajlova.
- Signal je zapisan kao čist PCM (Pulse-code modulation), tj. direktno prenesena semplovana vrednost signala u digitalni oblik.
- Fajl se sastoji iz tri glavna odeljka:
 - RIFF zaglavlje
 - WAVE deskriptor
 - Data odeljak

RIFF zaglavlje

Ime	Opis	Širina	Offset	Endian
ChunkID	ASCII zapis slova „RIFF“	4	0	big
ChunkSize	Veličina fajla bez prvih 8 bajtova	4	4	little
Format	ASCII naziv formata fajla (u našem slučaju - „WAVE“)	4	8	big

WAVE format

- Kada RIFF polje „Format“ sadrži vrednost „WAVE“, to znači da se očekuju još dva odeljka – jedan za parametre wave formata i drugi za same podatke.

WAVE parametri

Ime	Opis	Širina	Offset	Endian
Subchunk1ID	ASCII zapis slova „fmt “	4	12	big
Subchunk1Size	16 kod PCM – veličina ovog odeljka bez prvih 8 bajtova	4	16	little
AudioFormat	1=PCM, druge vrednosti naznačavaju kompresiju	2	20	little
NumChannels	Broj kanala	2	22	little
SampleRate	Frekvencija semplovanja (8000, 44100, ...)	4	24	little
ByteRate	$\text{SampleRate} * \text{NumChannels} * \text{BitsPerSample} / 8$	4	28	little
BlockAlign	$\text{NumChannels} * \text{BitsPerSample} / 8$	2	32	little
BitsPerSample	8, 16, ...	2	34	little

Data odeljak

Ime	Opis	Širina	Offset	Endian
Subchunk2ID	ASCII zapis slova „data“	4	36	big
Subchunk2Size	NumSamples * NumChannels * BitsPerSample / 8 (veličina samih podataka koji slede, u bajtovima)	4	40	little
Data	Podaci	*	44	little

Vežba

- Analizirati WAV fajl pomoću HEX editora.
- Napisati program koji otvara WAV fajl i ispisuje kompletno zaglavlje.