



Univerzitet u Beogradu-Elektrotehnički fakultet  
Katedra za signale i sisteme



# PREPOZNAVANJE GOVORA POMOĆU KEPSTRALNIH KOEFICIJENATA



Kandidat: Marija Rakonjac, br. indeksa 2020/0222

Mentor: Željko Đurović

Beograd, avgust 2024. godine

# UVOD

- Prepoznavanje govora - tehnologija koja omogućava računarima i drugim uređajima da razumeju i interpretiraju ljudski govor
- Cilj diplomskog rada - dizajnirati sistem za prepoznavanje govora primenom kepstralnih koeficijenata i klasifikatora k najbližih suseda koji koristi ograničen rečnik za klasifikaciju reči
- Glavni zadaci su ispitivanje efikasnosti kepstralne analize za ovakav problem, kao i uticaj podele reči na segmente na tačnost klasifikacije

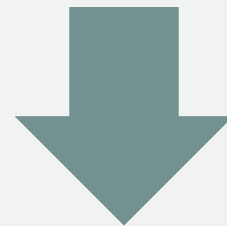


# MODELIRANJE I NASTANAK GOVORNOG SIGNALA

- Govorni signal – akustički talas koji nastaje kao rezultat vibracija glasnih žica
- U nastanku govornog signala učestvuju: pluća, dijafragma, glasne žice, glotis i farinks (ždrelo), usna i nosna šupljina, jezik, usne, zubi, velum (resica)
- Model uniformne tube – pojednostavljeni model za analizu akustičkih karakteristika govornog signala
- Pretpostavka: poprečni presek vokalnog trakta je isti celom dužinom

# ELEMENTI KEPSTRALNE ANALIZE

**KEPSTAR**



**SPEKTAR**

# ELEMENTI KEPSTRALNE ANALIZE

**KEPSTAR**



**SPEKTAR**

**KVEFRENCIJA**



**FREKVENCIJA**

**LIFTEROVANJE**



**FILTEROVANJE**

**RHAMONIK**



**HARMONIK**

# IZRAČUNAVANJE KEPSTRA

vremenski domen

$$C(x(t))$$

=

kepstar

$$F^{-1}[\log(F[x(t)])]$$

spektar

log spektar

**KEPSTAR**

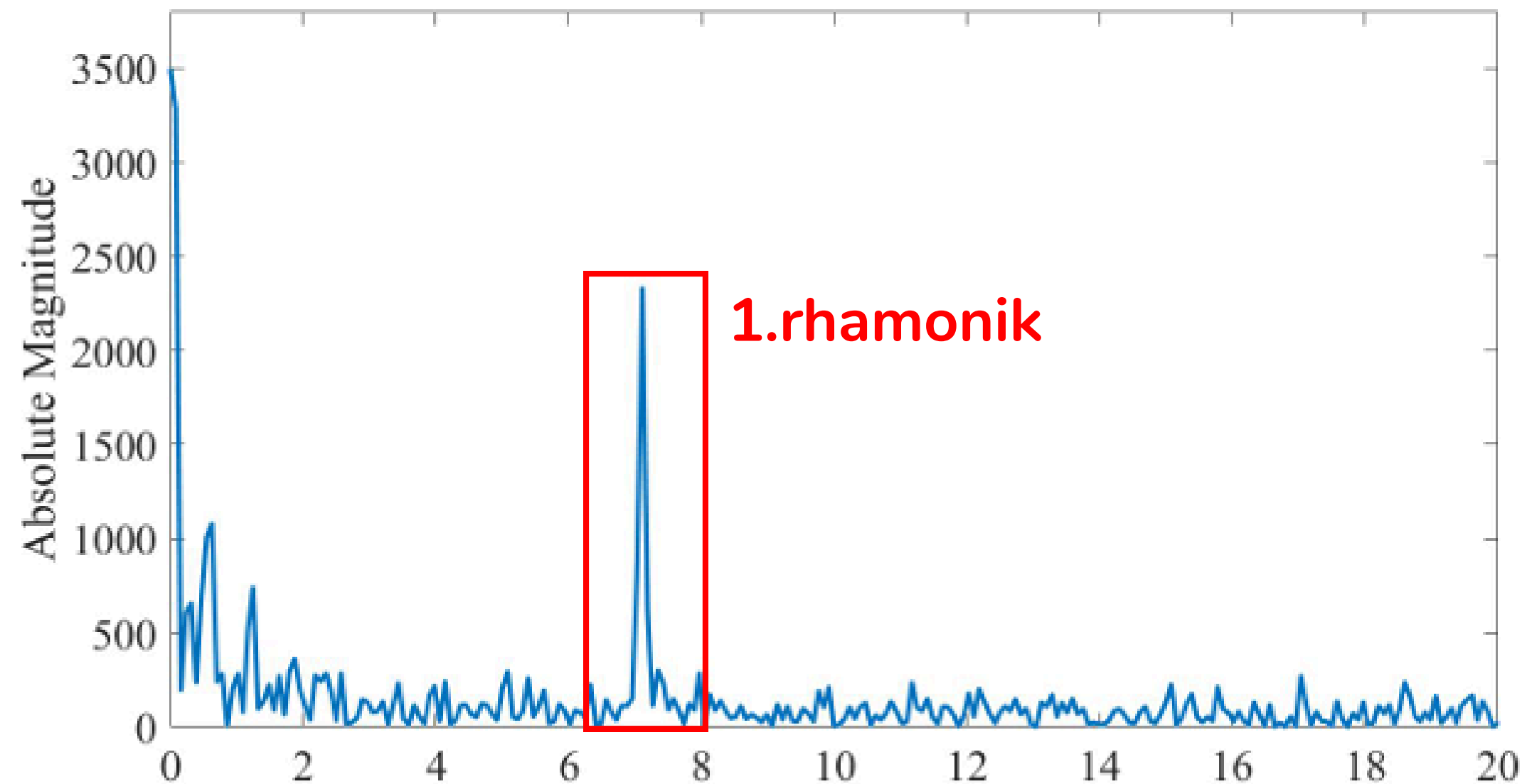


**SPEKTAR SPEKTRA**



**KEPSTAR**

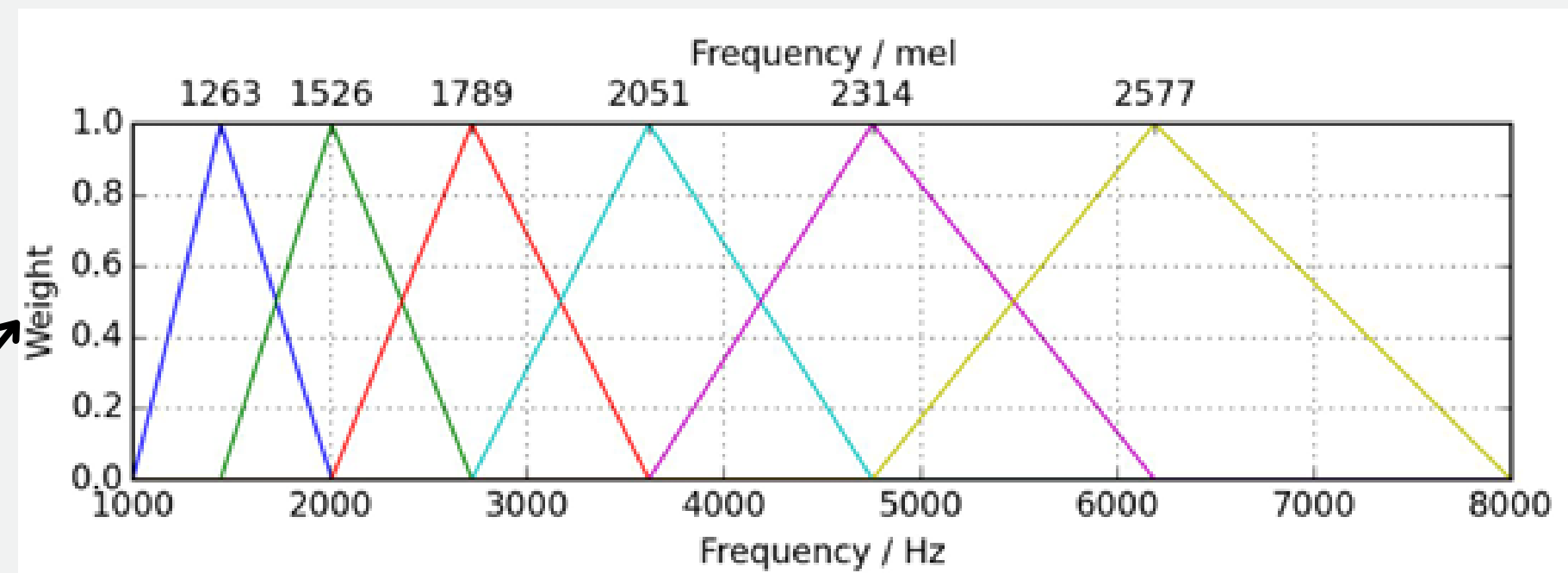
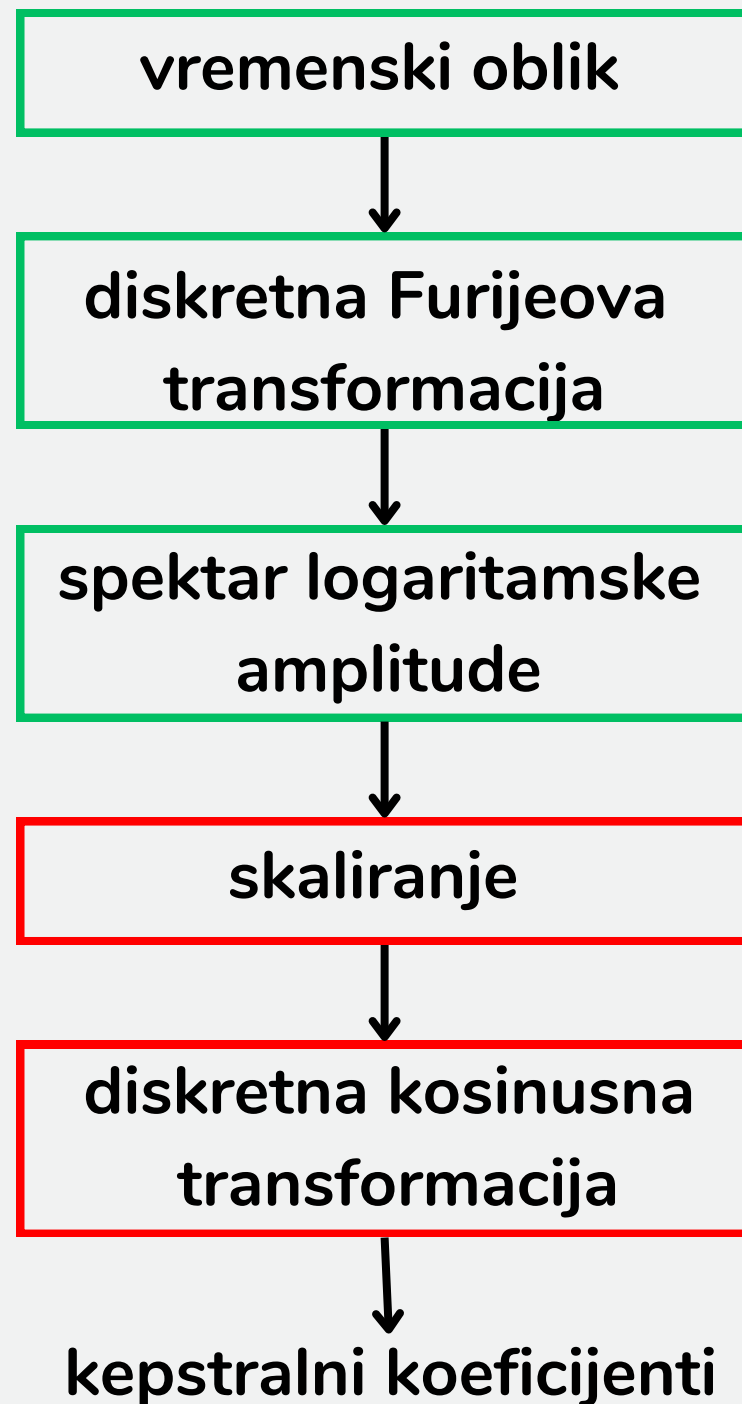
# KEPSTAR



kvefrencija (ms)



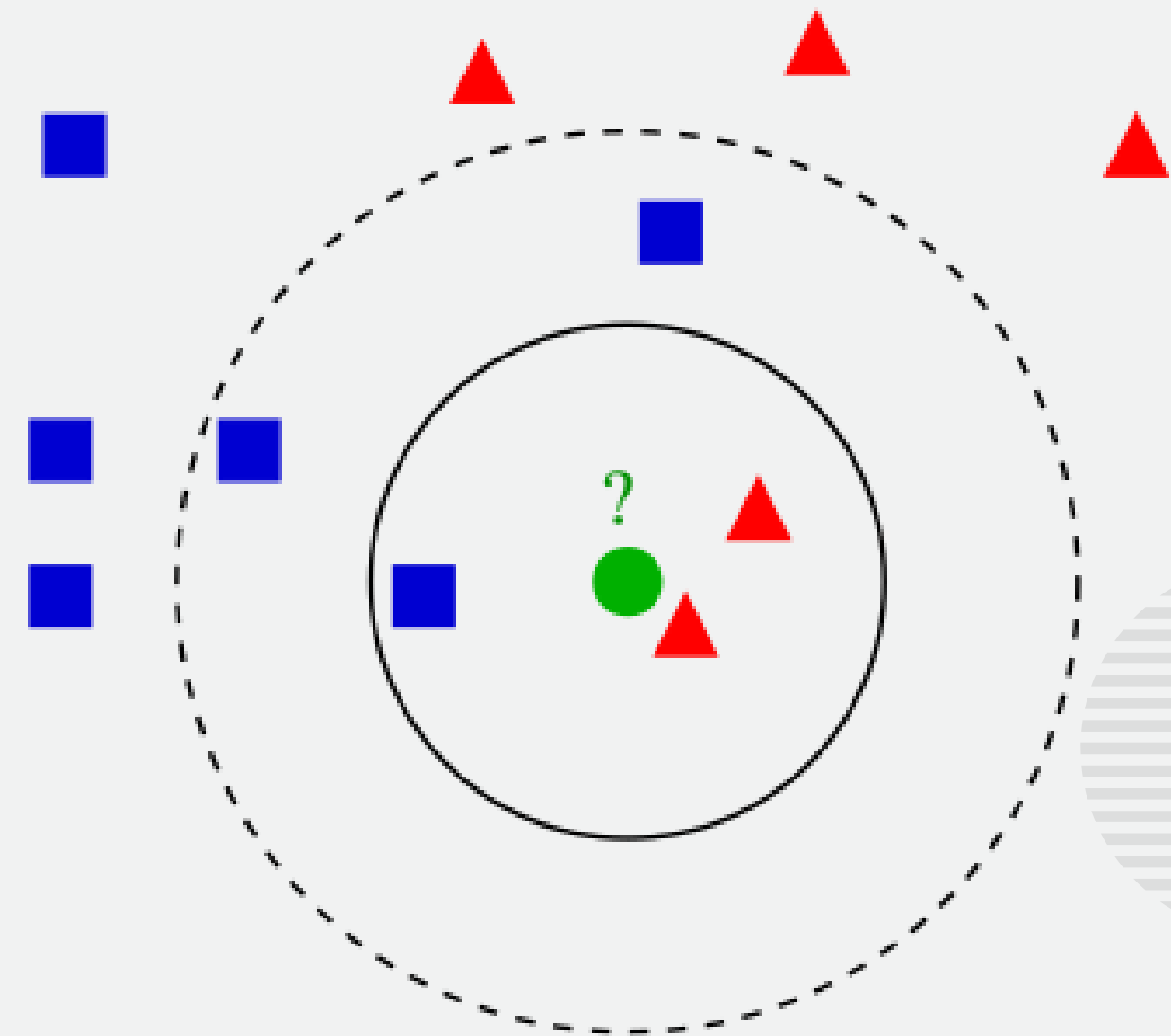
# KEPSTRALNI KOEFICIJENTI



- jednostavnija verzija Furijeove transformacije
- daje realne koeficijente

# METODA K NAJBЛИŽIH SUSEDA

- K najbližih suseda (eng. *k Nearest Neighbours-kNN*) je jednostavna neparametarska metoda koja se koristi u problemima klasifikacije oblika i regresije
- Odabir odgovarajuće vrednosti  $k$  je ključan za performanse algoritma
- Prednosti: Jednostavan za implementaciju i prilagodljiv novim podacima
- Ograničenja: Računski zahtevan i podložan prokletstvu dimenzionalnosti



# OPIS SISTEMA ZA PREPOZNAVANJE GOVORA

Formiranje baze  
podataka



Predprocesiranje



Izdvajanje obeležja



Klasifikacija oblika

# FORMIRANJE BAZE PODATAKA

Formiranje baze  
podataka



Predprocesiranje



Izdvajanje obeležja



Klasifikacija oblika

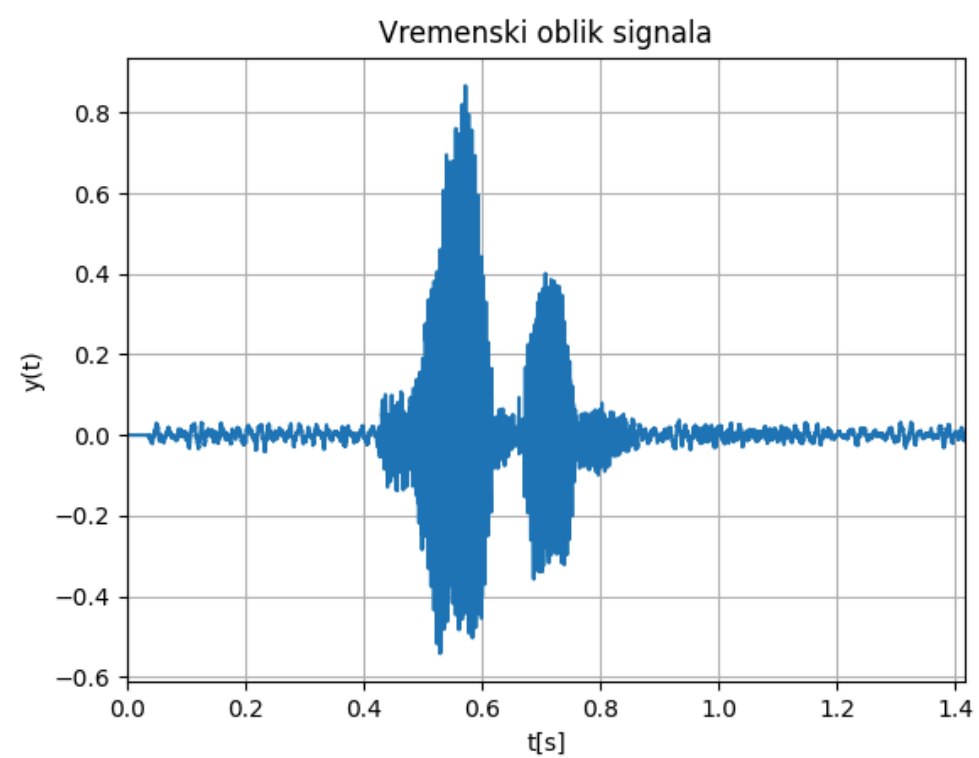
# FORMIRANJE BAZE PODATAKA

- 5 reči – **JEDAN, SEDAM, SIGNAL, GROŽĐE, KUĆA**
- 30 govornika – 14 osoba ženskog pola, 16 osoba muškog pola
- Većinom starosti između 20 i 25 godina
- 5 deteta
- 2 osobe starosti 45+
- Svaka reč izgovorena 15 puta od strane jednog govornika – ukupno 2250 govornih snimaka

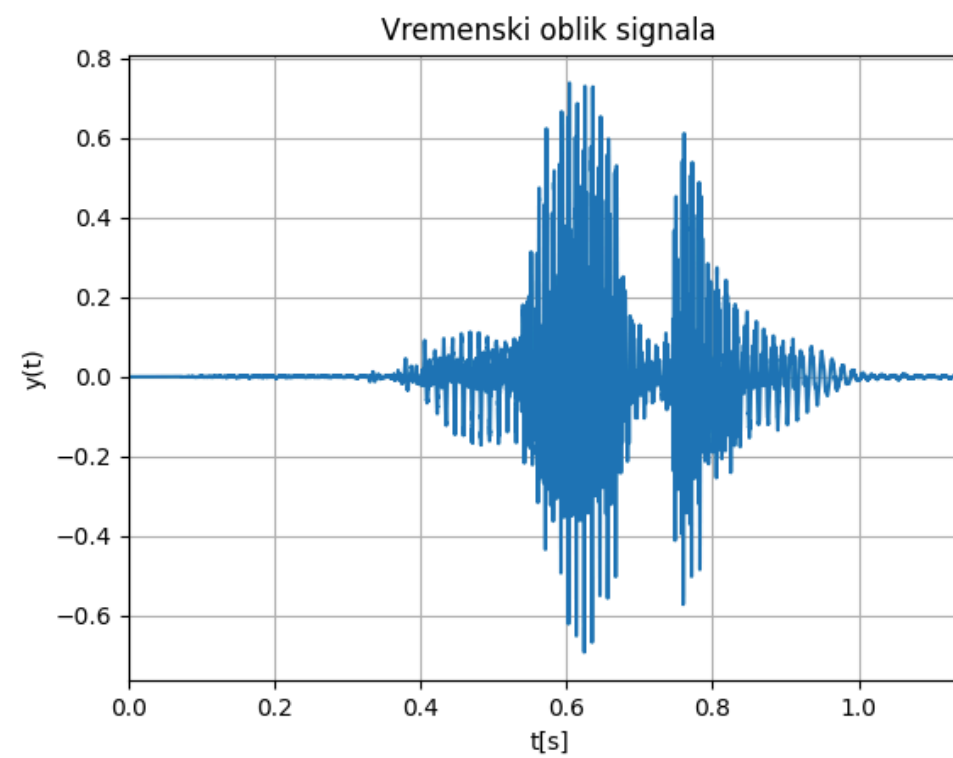
# PRIKAZ SIGNALA

JEDAN

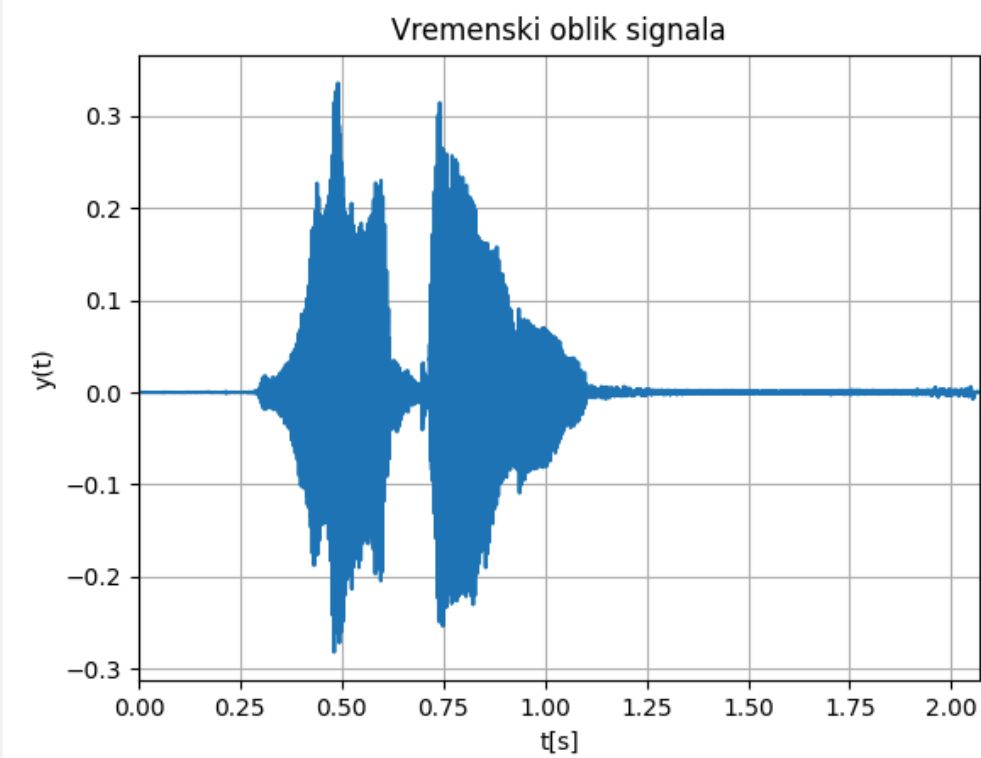
ŽENA



MUŠKARAC



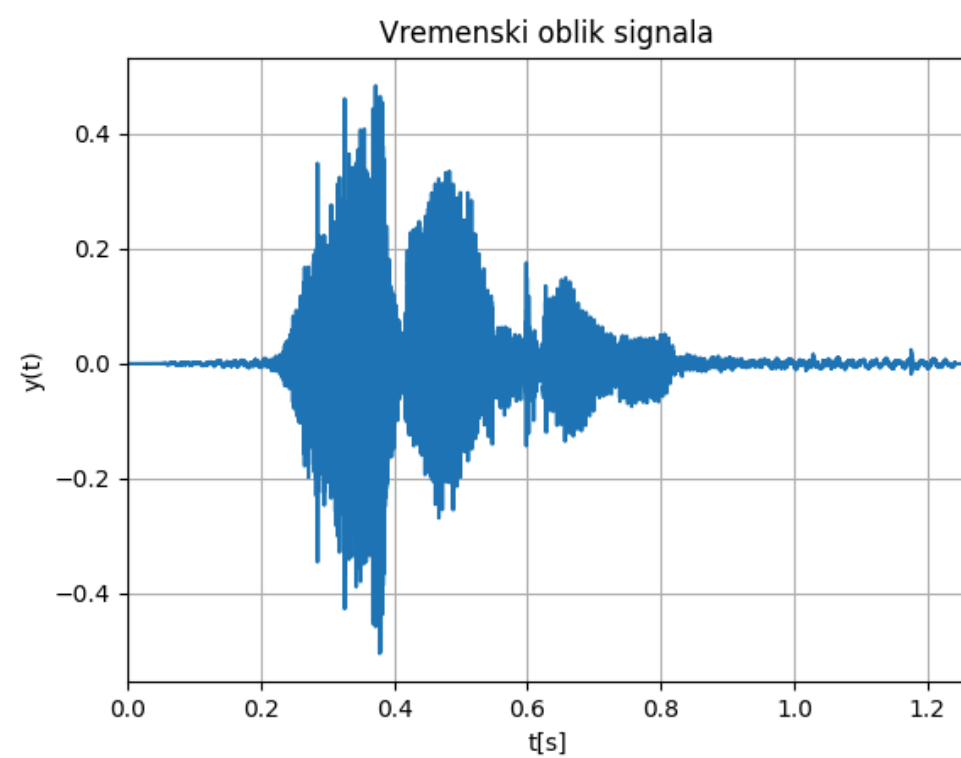
DETE



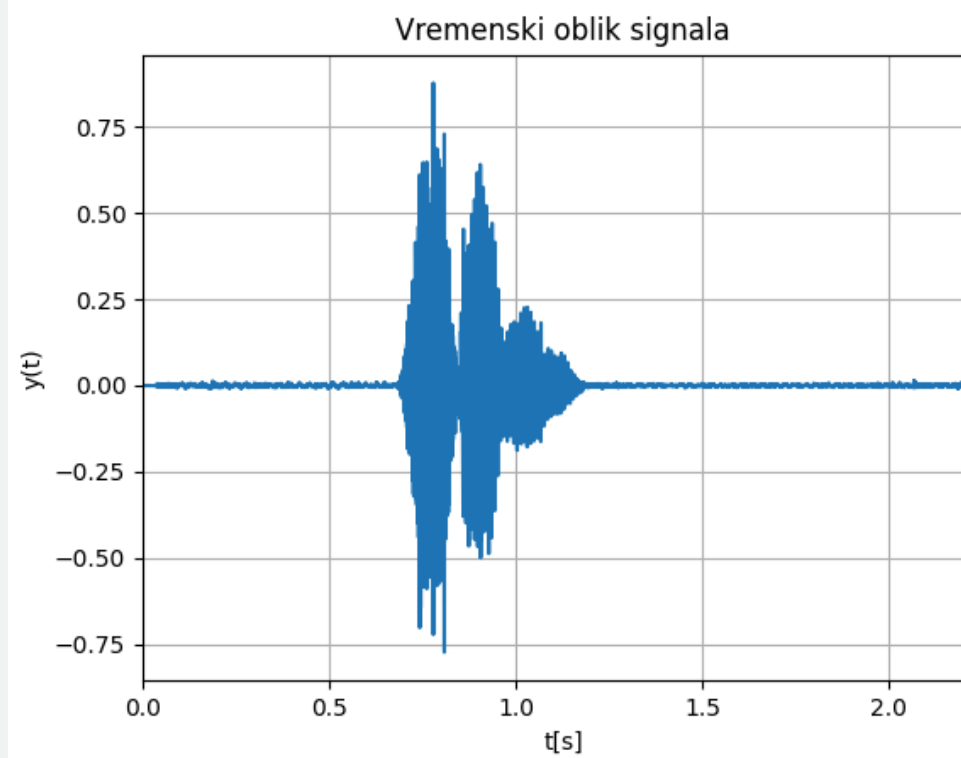
# PRIKAZ SIGNALA

SEDAM

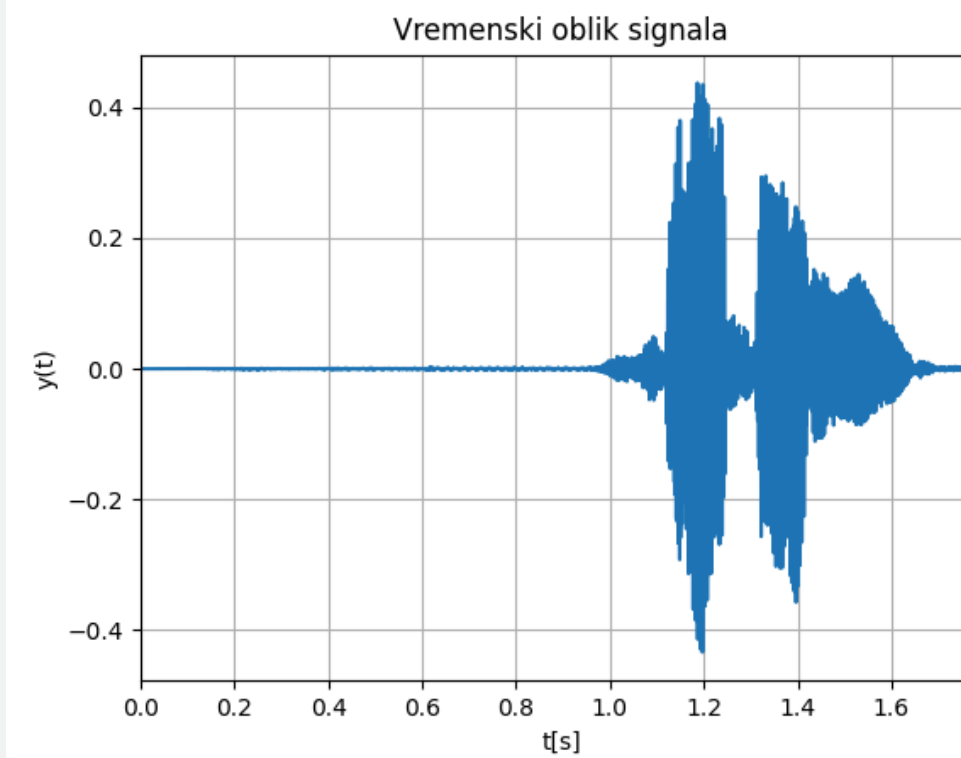
ŽENA



MUŠKARAC



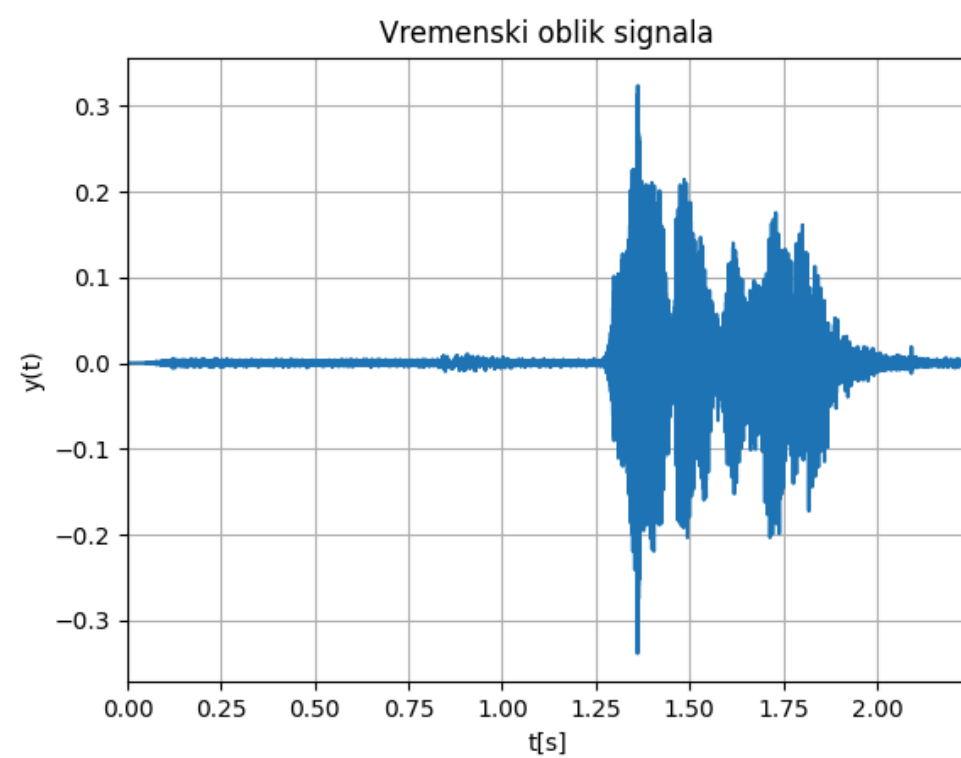
DETE



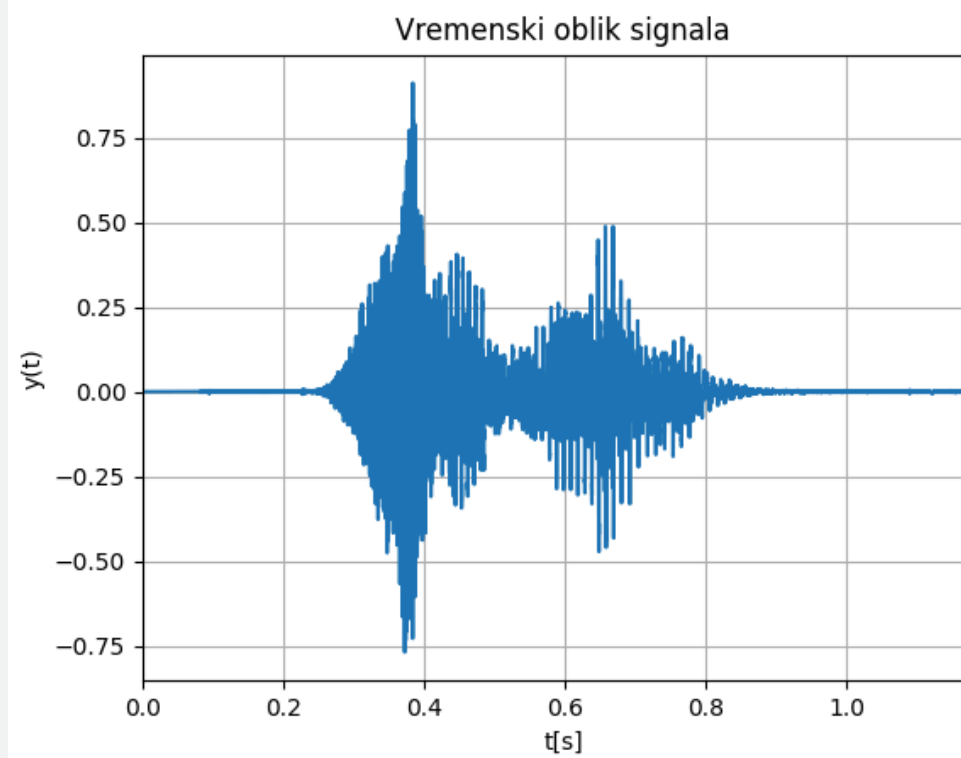
# PRIKAZ SIGNALA

## SIGNAL

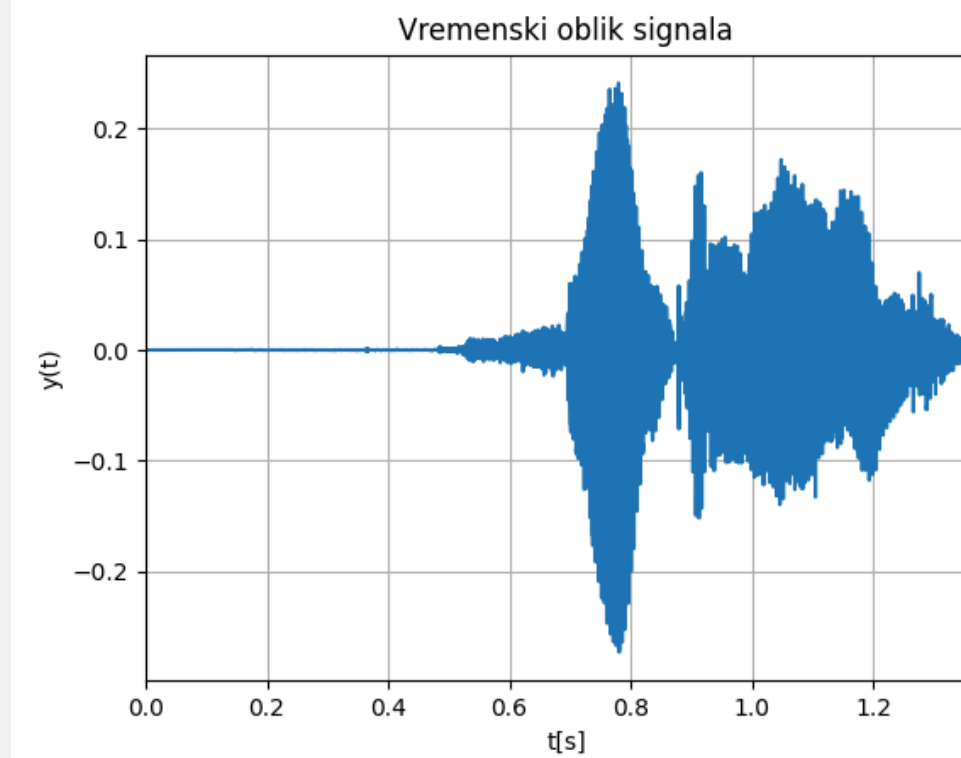
### ŽENA



### MUŠKARAC



### DETE

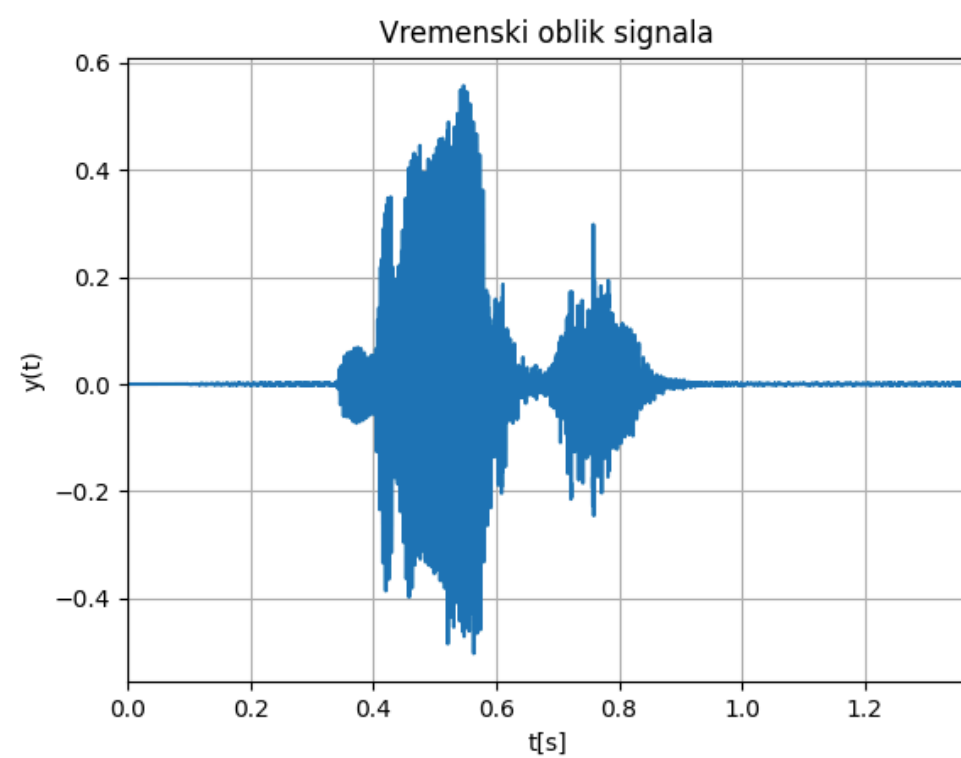




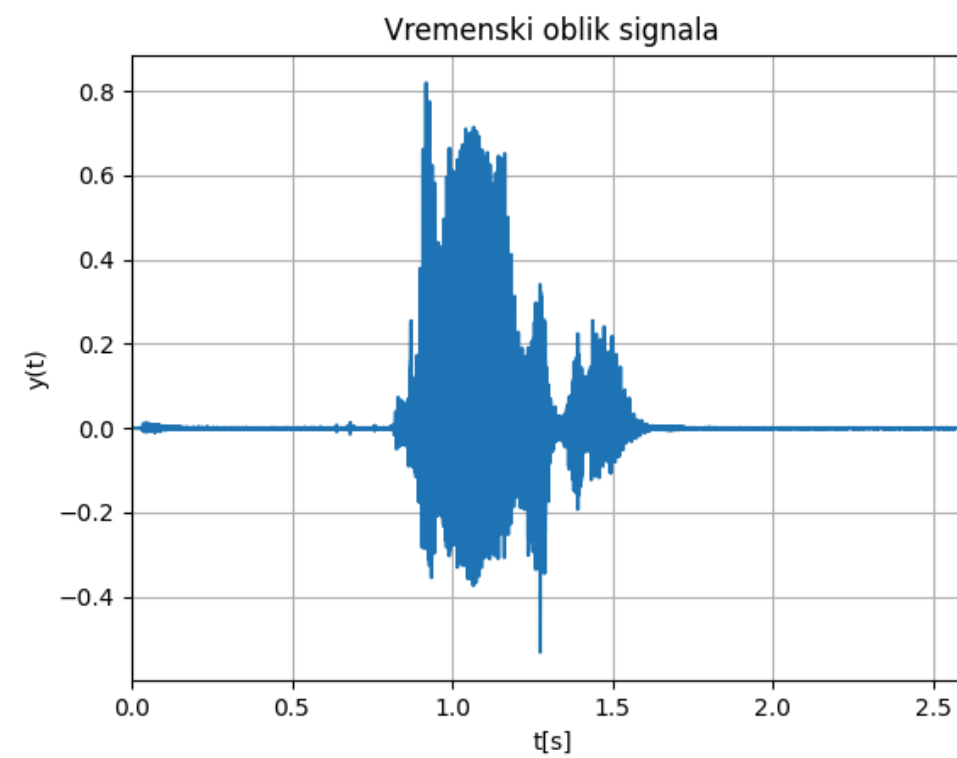
# PRIKAZ SIGNALA

GROŽĐE

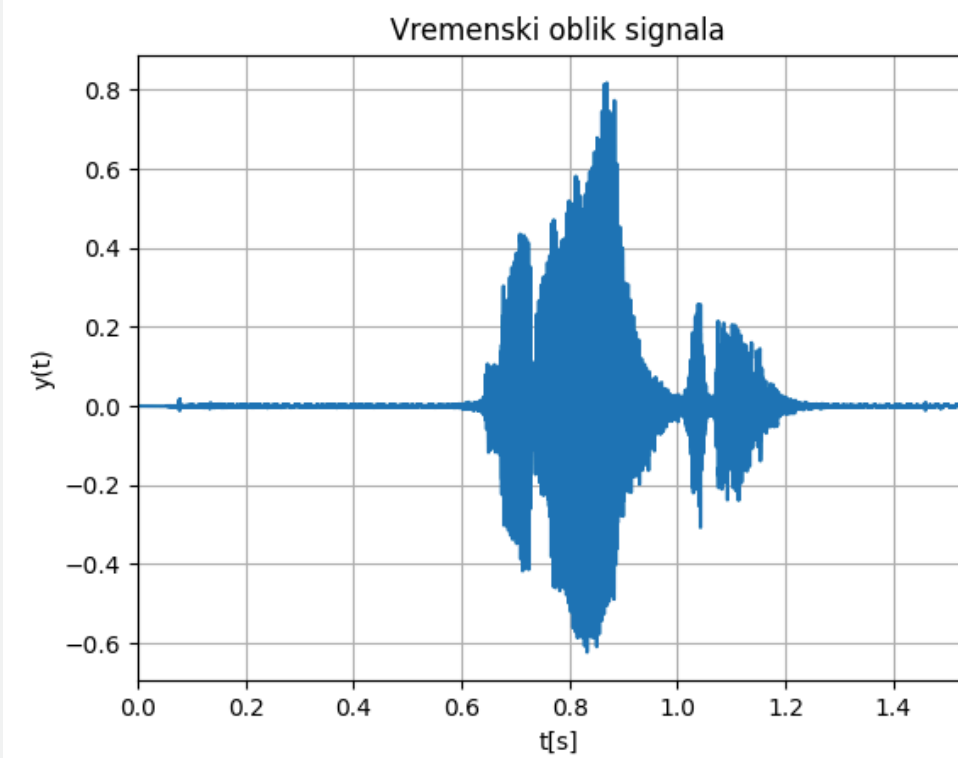
ŽENA



MUŠKARAC



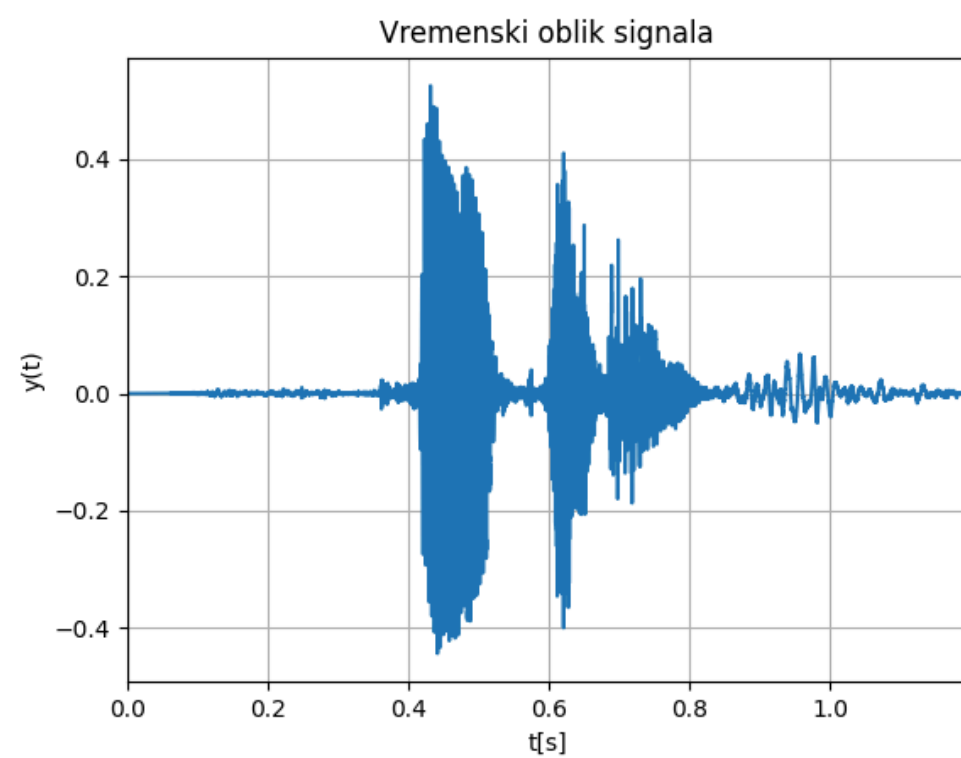
DETE



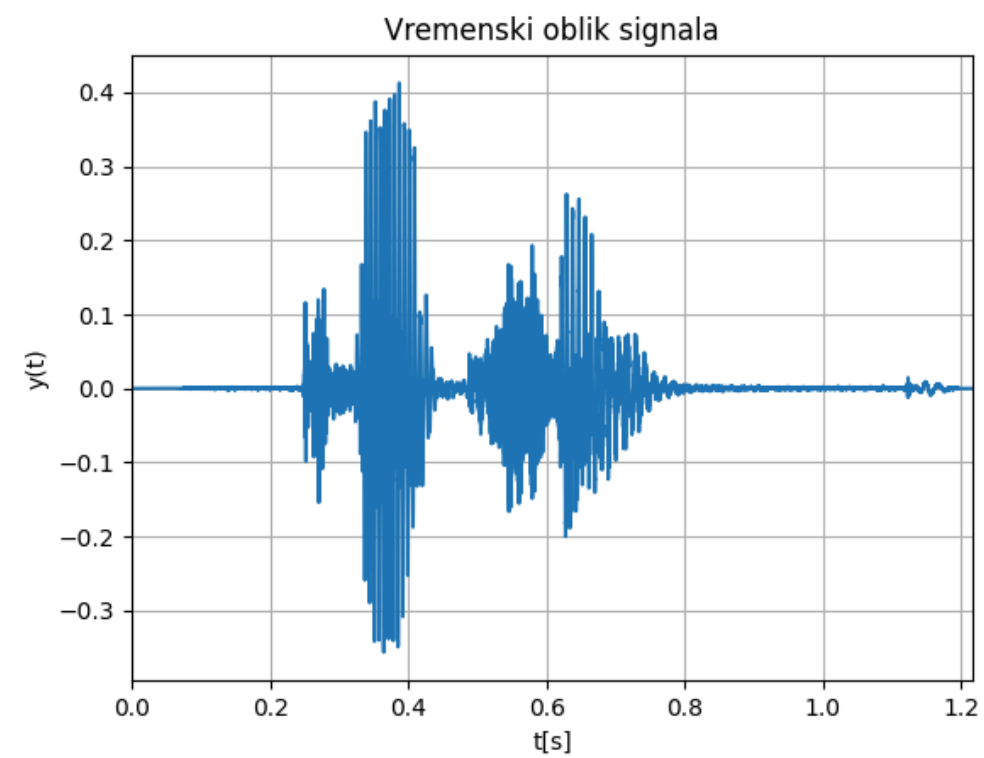
# PRIKAZ SIGNALA

KUĆA

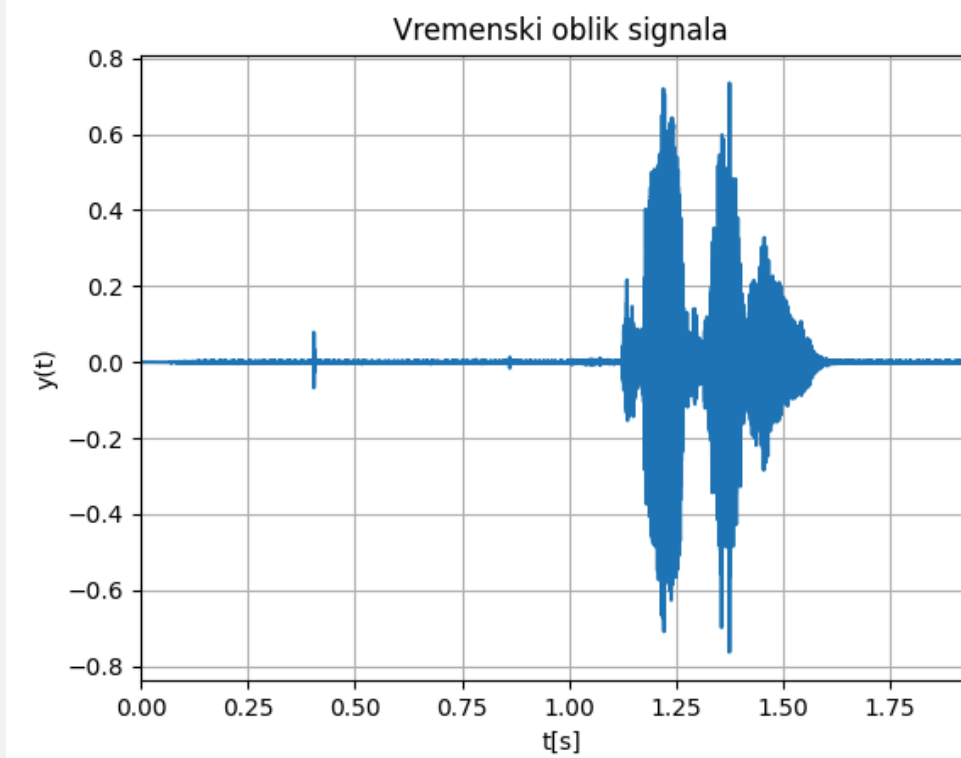
ŽENA



MUŠKARAC



DETE



# PREDPROCESIRANJE

Formiranje baze  
podataka



Predprocesiranje



Izdvajanje obeležja

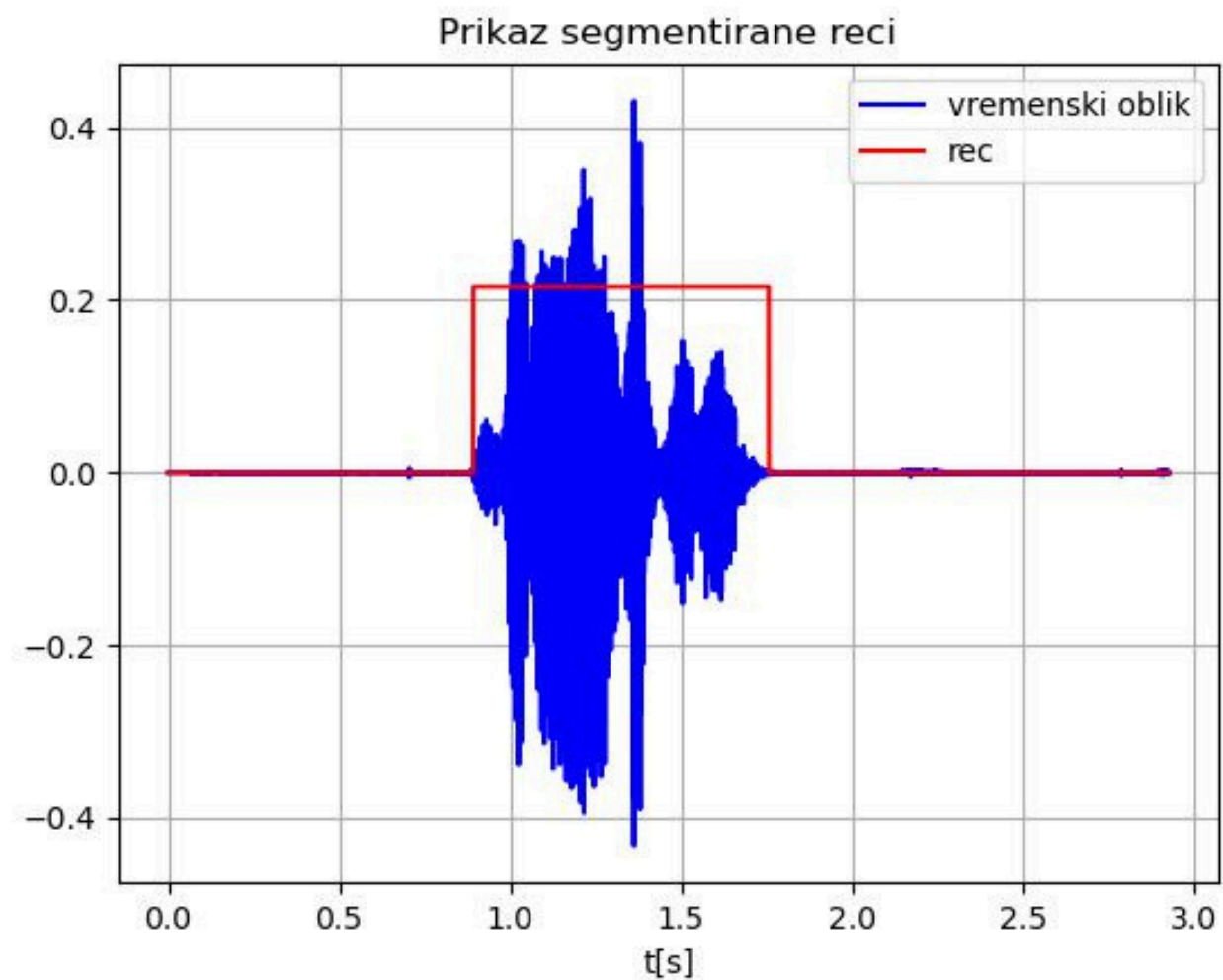


Klasifikacija oblika

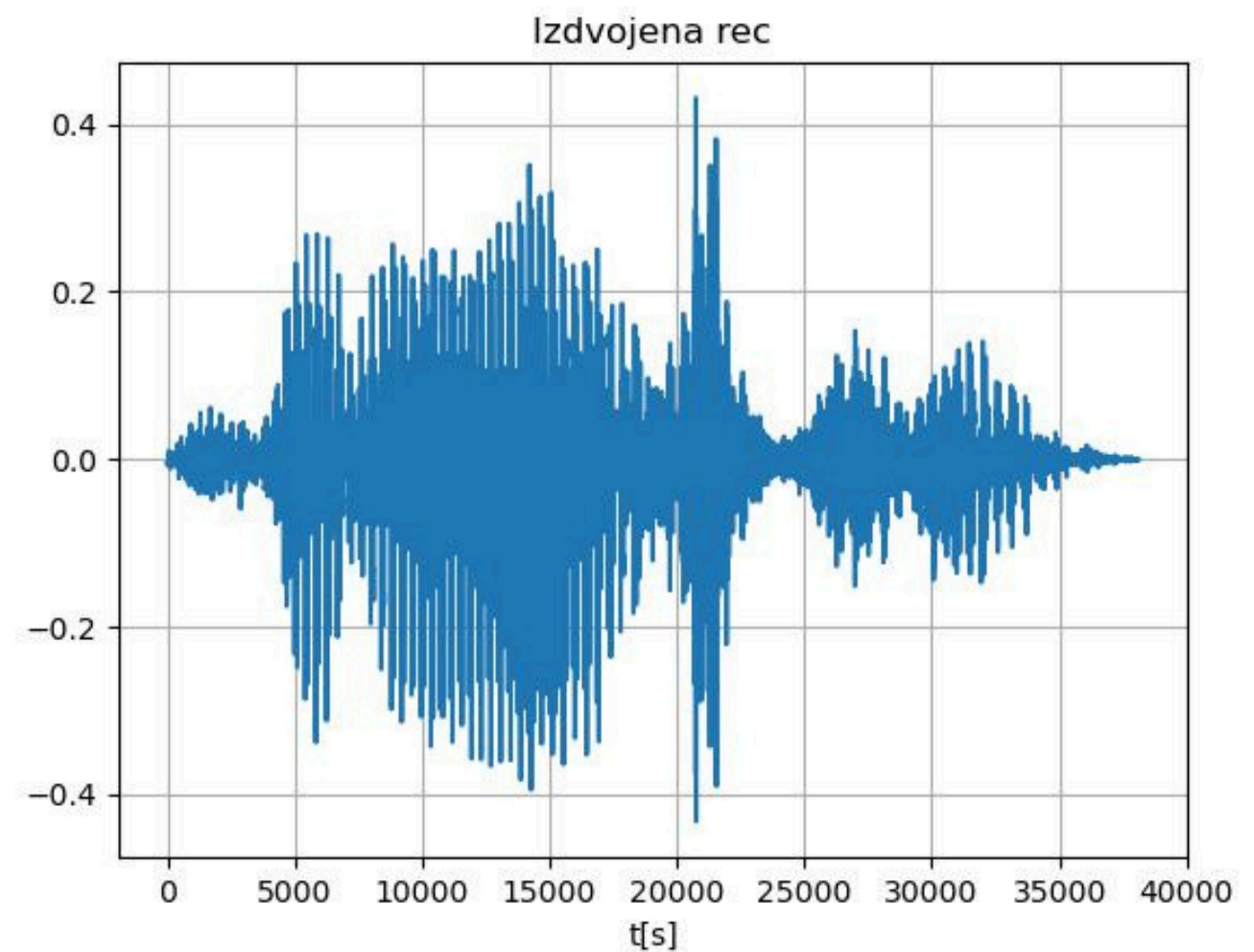
# PREDPROCESIRANJE

- Postupak predprocesiranja:
  1. **Uklanjanje šuma** – primena Butterwoth-ovog filtera (opseg 300-4000Hz)
  2. **Segmentacija reči** – određivanje granica reči; uklanjanje tišine
  3. **Normalizacija podataka** – po varijansi
  4. **Pre-emphasis filter** – pojačava više frekvencije signala u odnosu na niže frekvencije
- Korišćene teorijske karakteristike govornog signala:
  - **Kratkovremenska energija E** – mera snage signala u određenom vremenskom intervalu
  - **Brzina prolaska kroz nulu ZCR** – mera frekvencijskog sadržaja signala

# PREDPROCESIRANJE



**ORIGINALNI SIGNAL**



**NAKON SEGMENTACIJE**

# IZDVAJANJE OBELEŽJA

Formiranje baze  
podataka



Predprocesiranje



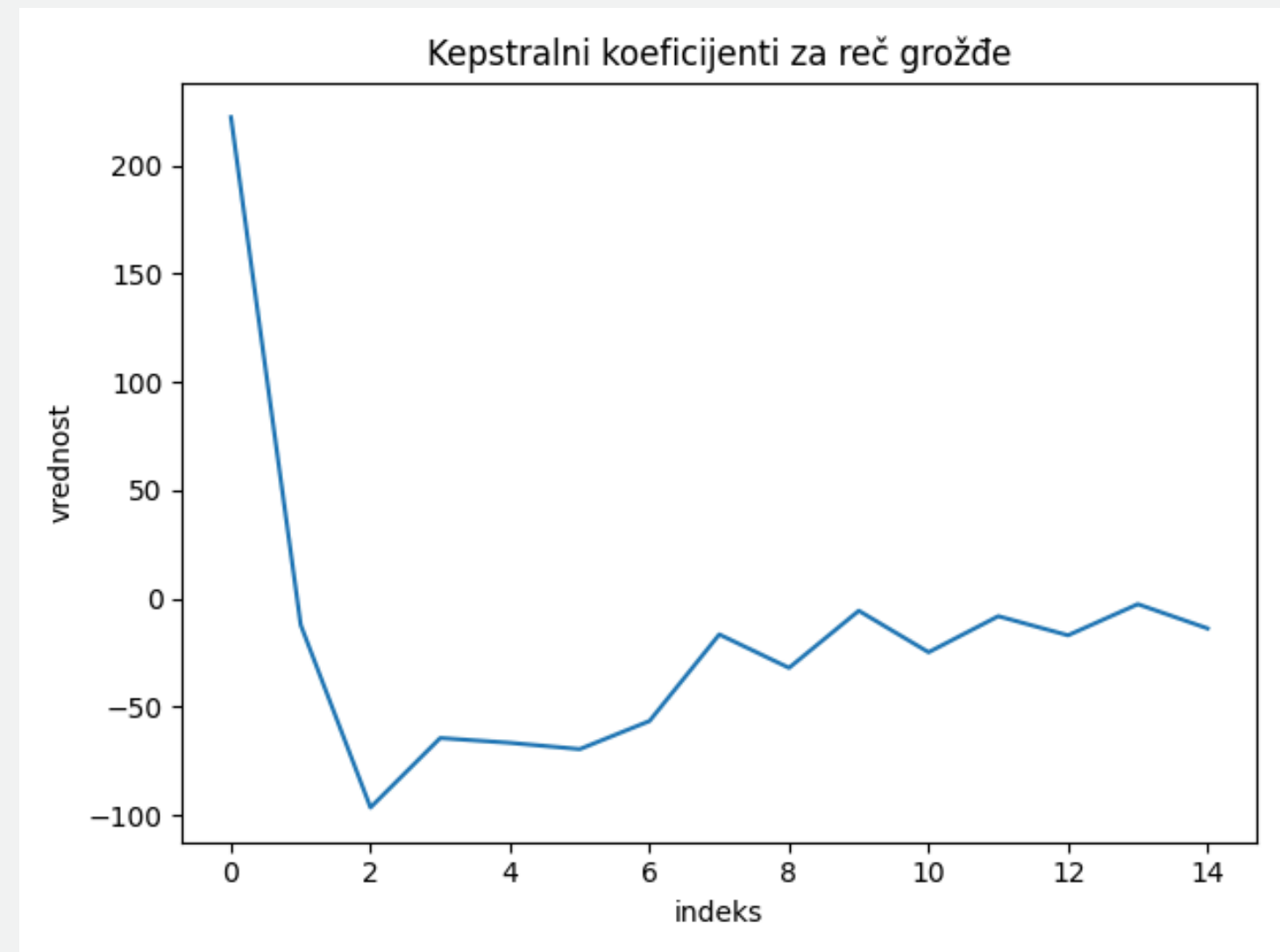
Izdvajanje obeležja



Klasifikacija oblika

# IZDVAJANJE OBELEŽJA

- Red kepstralne analize (broj koeficijenata) -određuje koliko dobro model može da predstavi spektralne karakteristike govornog signala
- Usvojeno 15 za broj kepstralnih koeficijenata
- Pretpostavka da je signal stacionaran, pa se ne koristi prozorovanje



# KLASIFIKACIJA OBLIKA

Formiranje baze  
podataka



Predprocesiranje



Izdvajanje obeležja



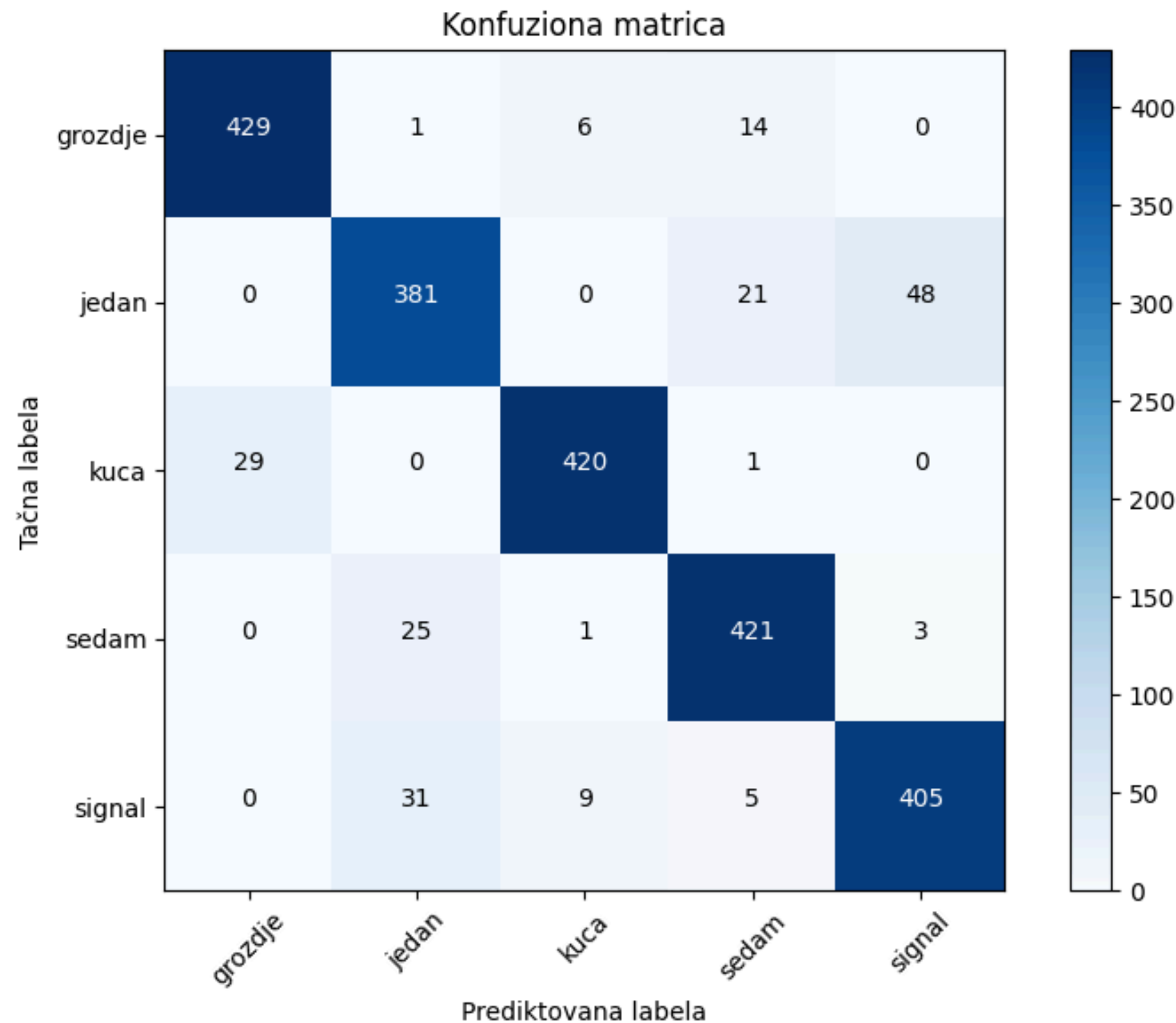
Klasifikacija oblika



# KLASIFIKACIJA OBLIKA

- Baza podataka se deli na trening i test skup
- Za treniranje klasifikatora k najbližih suseda korišćena je pretraga hiperparametara sa petostrukom kros-validacijom (eng. **Grid Search**)
- Postavlja se mreža koja pretražuje sledeće parametre:
  - broj suseda: testira vrednosti 3, 5, 7 i 9
  - težinsku funkciju: testira slučaj kada svi susedi imaju istu težinu i kada se daje prednost susedima koji su bliži trenutnoj tački
  - metriku: testira Euklidsku i Menhetn distancu
- Za slučaj kada je signal podeljen na 2 dela, za svaki segment se izračunava 15 koeficijenata, tako da ukupno postoji 30 obeležja; analogno za 3 dela

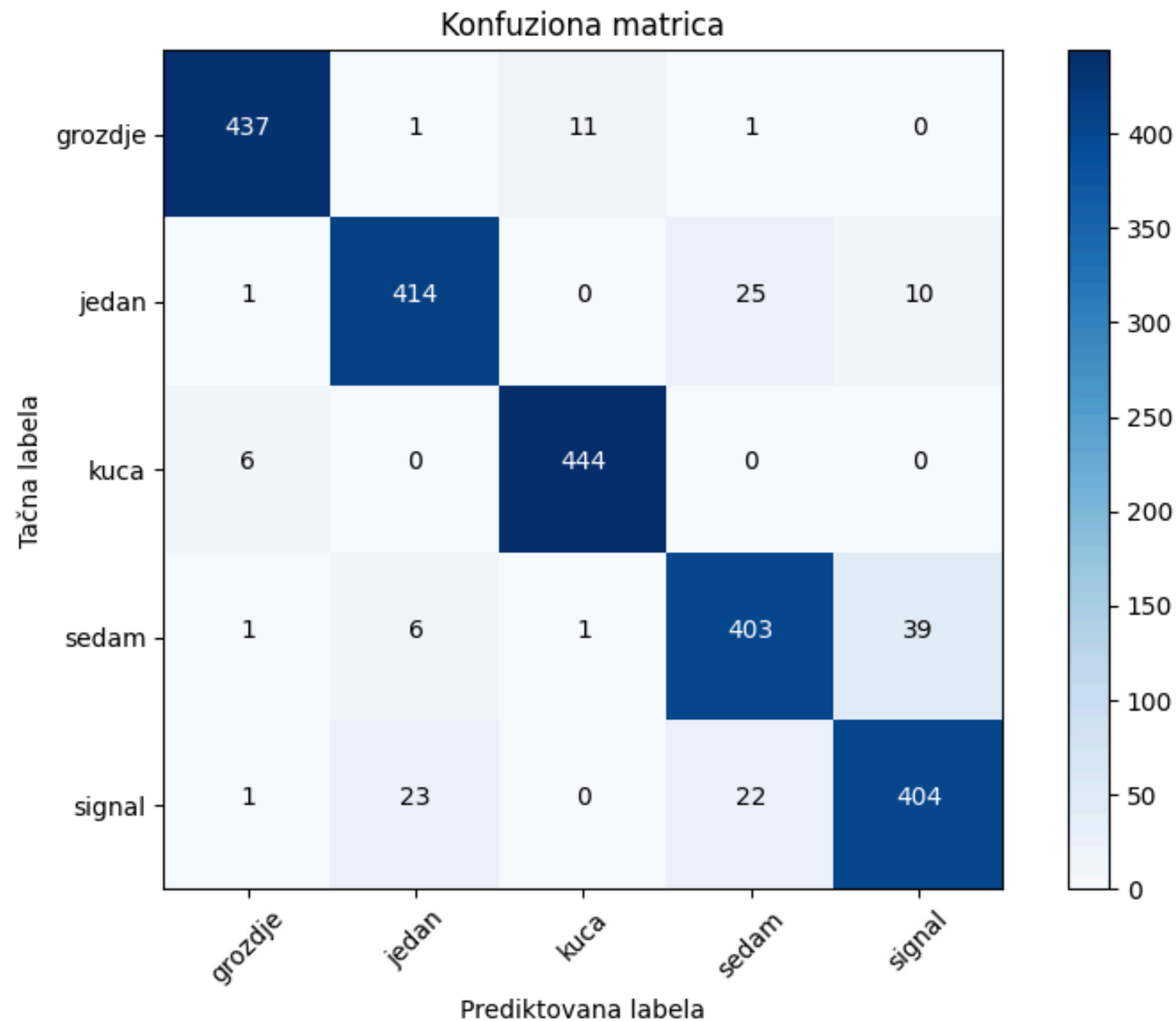
# REZULTATI



## NEPODELJENI SIGNALI

- Optimalan broj suseda: 3
- Optimalna težinska funkcija: funkcija težina distance
- Optimalna metrika: Euklidska distanca
- Tačnost klasifikacije: **91%**

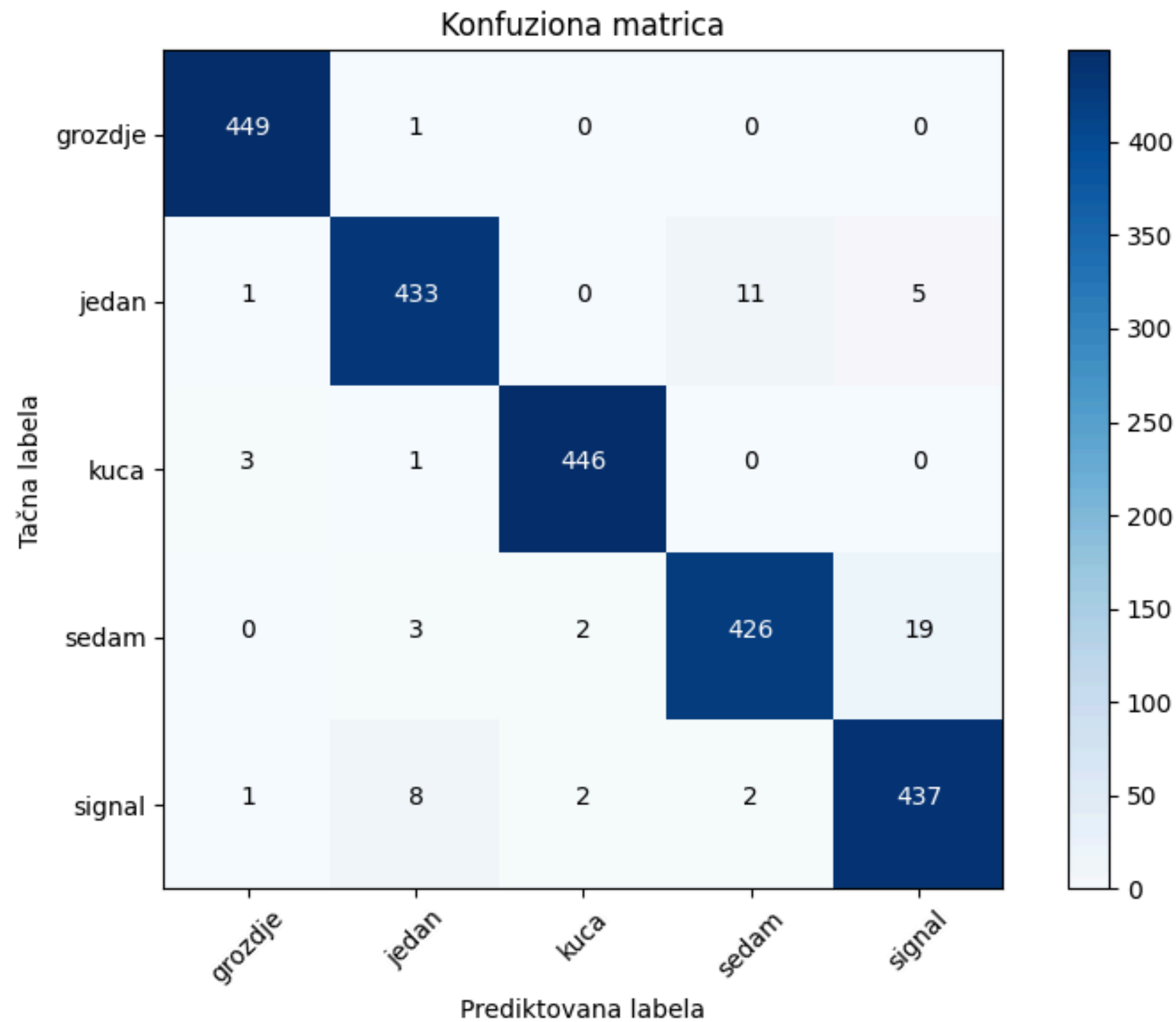
# REZULTATI



## SIGNALI PODELJENI NA 2 DELA

- Optimalan broj suseda: 9
- Optimalna težinska funkcija: funkcija težina distance
- Optimalna metrika: Euklidska distanca
- Tačnost klasifikacije: **93%**

# REZULTATI




## SIGNALI PODELJENI NA 3 DELA

- Optimalan broj suseda: 7
- Optimalna težinska funkcija: funkcija težina distance
- Optimalna metrika: Menhetn distanca
- Tačnost klasifikacije: **97%**



# ZAKLJUČAK

- Kepstralni koeficijenti i klasifikator k najbližih suseda su se pokazali kao efikasni za prepoznavanje govora
  - Demonstracija nestacionarnosti signala - prozorovanje znatno unapređuje performanse sistema, čineći ga otpornijim na varijabilnosti unutar govornog signala i boljim za detekciju lokalnih karakteristika signala
  - Razlike u govornim stilovima i individualnim karakteristikama govornika dovode do varijacija u signalima – izazov za tačnu klasifikaciju
  - Fonetska sličnost reči takođe unosi konfuziju unutar modela
  - Dalje istraživanje uključuje primenu alternativnih tehnika ekstrakcije obeležja, primenu drugih metoda klasifikacije, kao i implementaciju dubokih neuralnih mreža
- 



**HVALA NA PAŽNJI!**

**PITANJA?**

