Računalnik, ki sliši in govori slovensko

Janez Žibert
janez.zibert@upr.si

Fakulteta za matematiko, naravoslovje in informacijske tehnologije
Univerza na Primorskem

Fakulteta za elektrotehniko, Univerza v Ljubljani

NAMESTO UVODA: SIRI

- iPhone: SIRI

Pregled predavanja

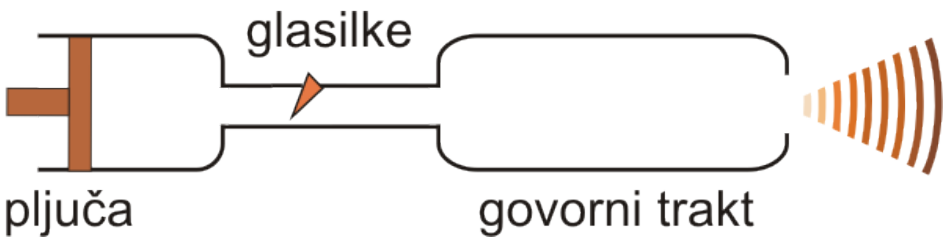
- govorni signali:
  - osnovne lastnosti govornih signalov in zaznavne zmožnosti človeka,
  - spektralna analiza govornih signalov: spektrogram, melodični spekter, koeficienti melodičnega kepstra.

- razpoznavanje govora:
  - osnovna zgradba sistemov za razpoznavanje govora,
  - predstavitev praktičnih aplikacij razpoznavanja govora.

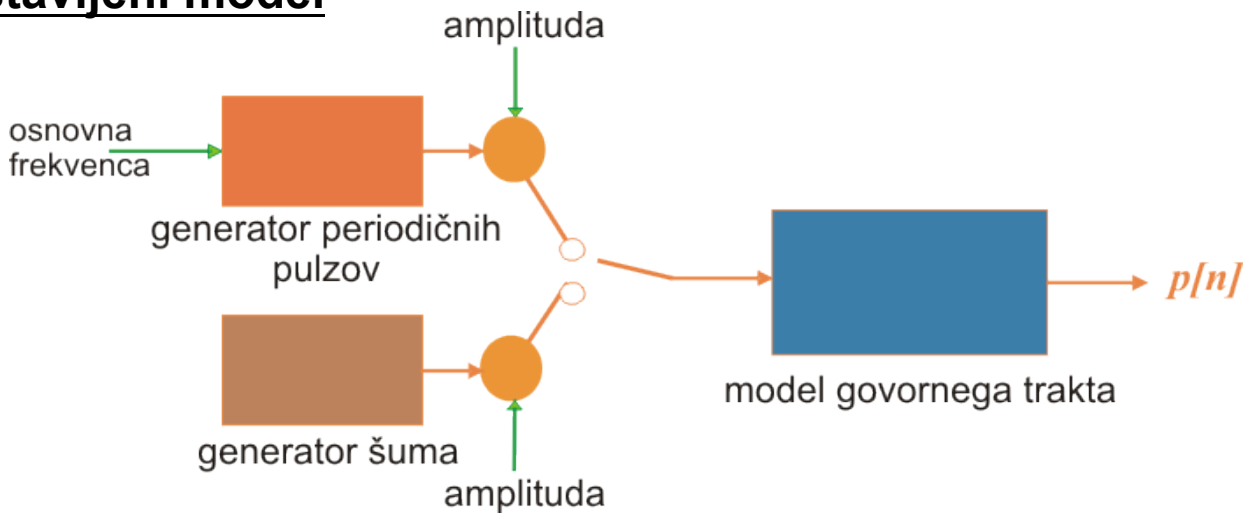
- tvorjenje govora – sinteza govora:
  - osnovna zgradba sistemov za sintezo govora,
  - predstavitev praktičnih aplikacij sinteze govora.

Tvorba govora pri človeku

Tvorjenje govora

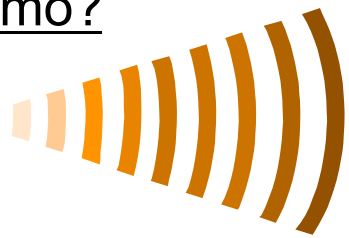


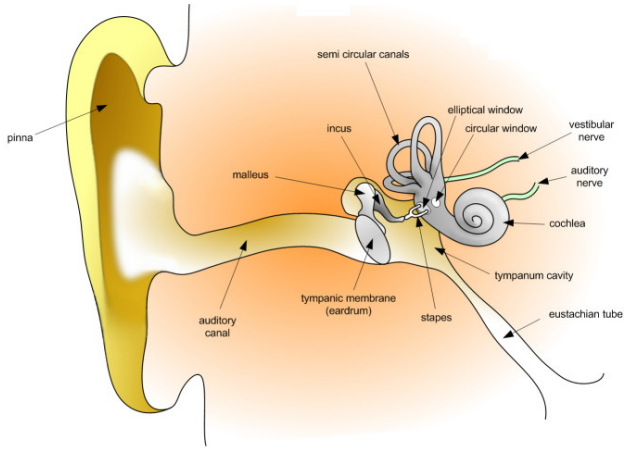
Poenostavljeni model



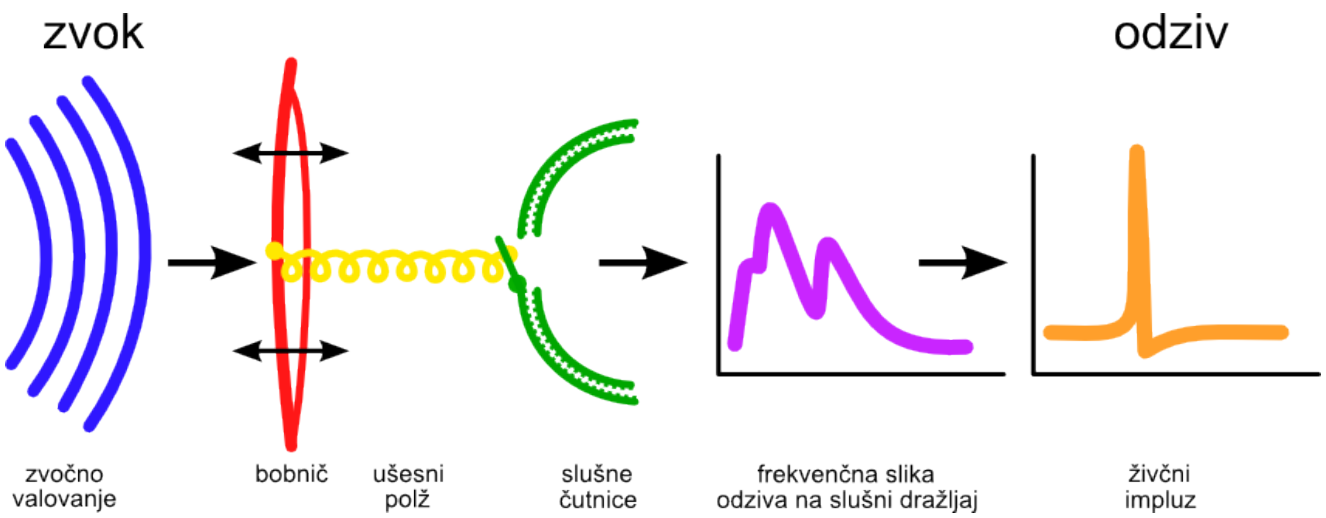
Zaznavanje zvoka pri človeku

Kako slišimo?





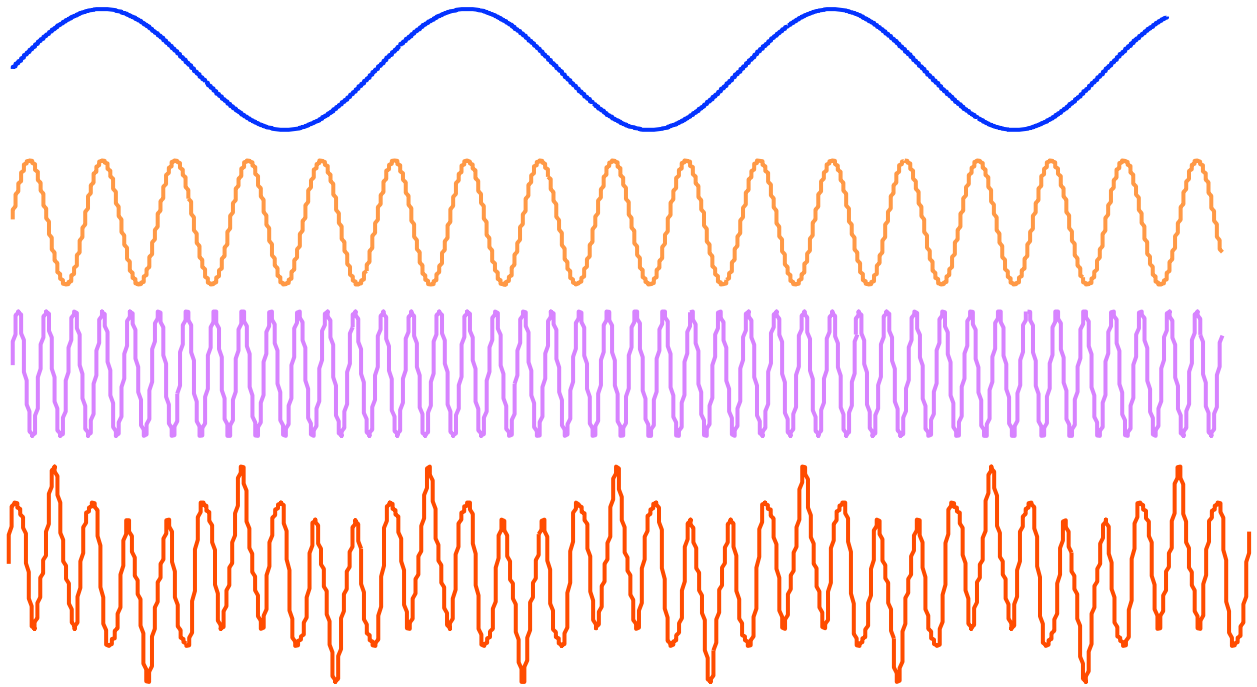
Poenostavljen model



Kaj slišimo: primeri periodičnih nihanj

$$y(t) = 8\sin(\omega_0 t) + 20\sin(5\omega_0 t) + 3\sin(13\omega_0 t)$$

	ω_0	$5\omega_0$	$13\omega_0$
Amplitudni spekter : $\lvert F(\omega)\rvert$	8	20	3
Fazni spekter : $\phi(\omega)$	0	0	0



$\sin(\omega_0 t)$

$\sin(5\omega_0 t)$

$\sin(13\omega_0 t)$

$y(t)$

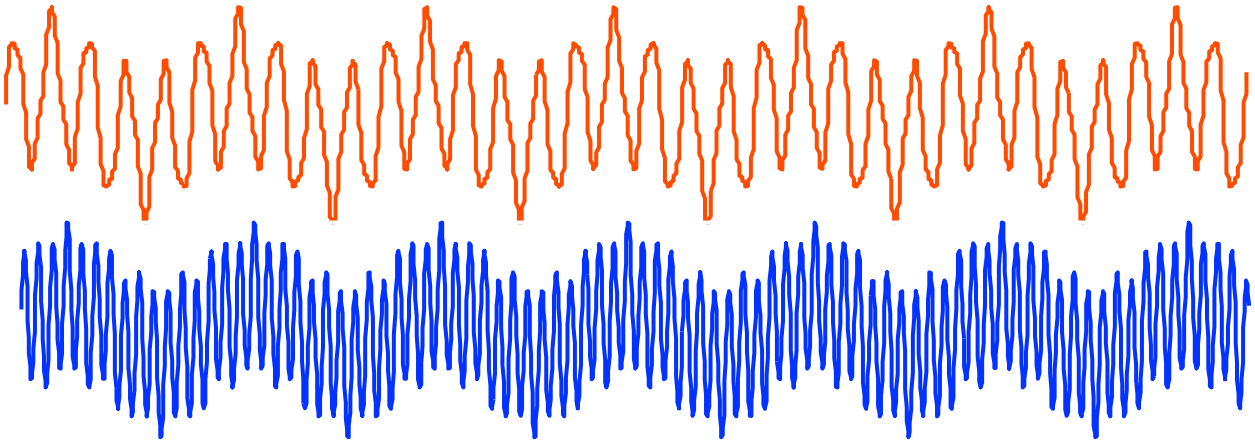
Kaj slišimo: razlike v amplitudnem spektru

$$y_1(t) = 8\sin(\omega_0 t) + 20\sin(5\omega_0 t) + 3\sin(13\omega_0 t)$$
$$y_2(t) = 8\sin(\omega_0 t) + 3\sin(5\omega_0 t) + 20\sin(13\omega_0 t)$$

	ω_0	$5\omega_0$	$13\omega_0$
$\lvert F_1(\omega)\rvert$	8	20	3
$\lvert F_2(\omega)\rvert$	8	3	20

$$\phi_1(n\omega_0) = \phi_2(n\omega_0)$$

Človek **ZAZNAVA** razlike v amplitudnem spektru.



$y_1(t)$

$y_2(t)$

Kaj slišimo: razlike v faznem spektru

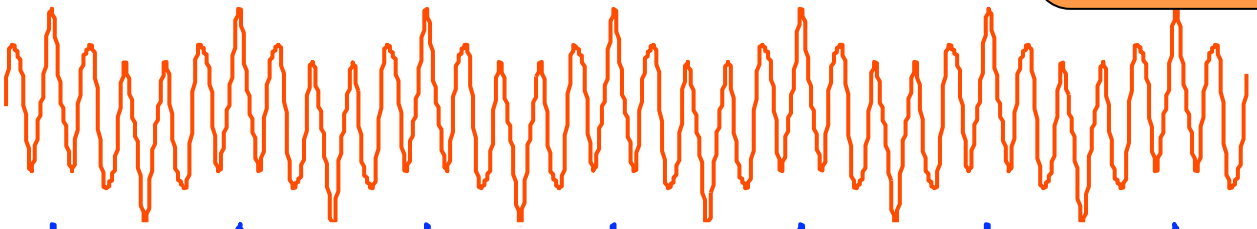
$$y_1(t) = 8\sin(\omega_0 t) + 20\sin(5\omega_0 t) + 3\sin(13\omega_0 t)$$
$$y_2(t) = 8\sin(\omega_0 t) + 20\sin(5\omega_0 (t-5)) + 3\sin(13\omega_0 t)$$

$$|F_1(n\omega_0)| = |F_2(n\omega_0)|$$

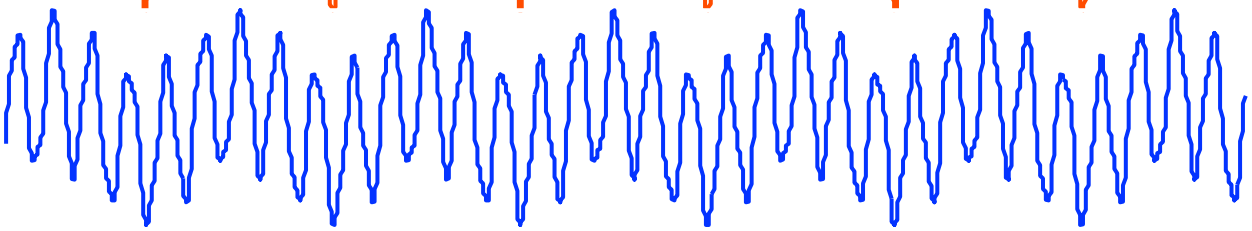
	ω_0	$5\omega_0$	$13\omega_0$
$\phi_1(\omega)$	0	0	0
$\phi_2(\omega)$	0	5	0

Človek **NE ZAZNAVA** razlik v faznem spektru.

Informacijo v zvoku 'nosi' le **amplitudni spekter**.



$y_1(t)$



$y_2(t)$

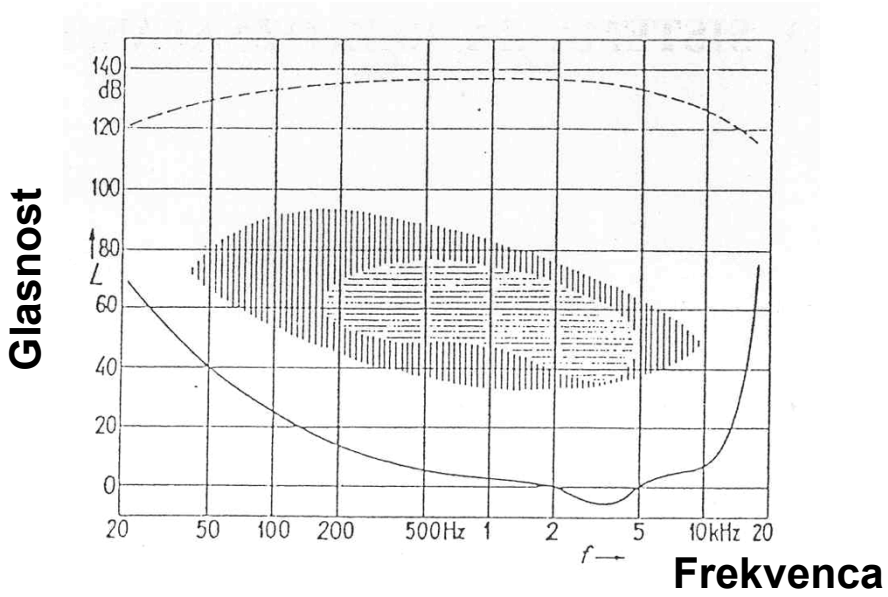
Lastnosti govornega signala

- Govorni zvočni signal je sestavljen iz periodičnih in neperiodičnih delov.

- Zveneče govorne signale na dovolj kratkem časovnem intervalu lahko opredelimo kot "*skoraj*" periodične (kvaziperiodične signale).
  - Perioda (osnovni ton, ki ga določa frekvenca nihanja glasilk) se s časom spreminja (intonacija).

- Nezveneči deli govora so <u>povsem neperiodični</u>.

Zaznavne zmožnosti človeka



- Območje slišnosti je odvisno od **glasnosti** in **frekvence**:
  - Vodoravno črtkano – območje govornega signala
  - Navpično črtkano – območje glasbe
- **Uho ni fazno občutljivo.**
- Sposobnost zaznavanja časovnih sprememb med **5 in 20 msec.**
- **Sposobnost zaznavanja razlik v signalu se s frekvenco spreminja.**
- Minimalen čas za opredelitev signala med **50 in 200 msec.**

PRETVORBA V FREKVENČNI PROSTOR

- Fourierjeva vrsta:

  vsak periodičen signal lahko opišemo s kombinacijo sinusnih nihanj različnih period (frekvenc).

- Fourierjeva transformacija:

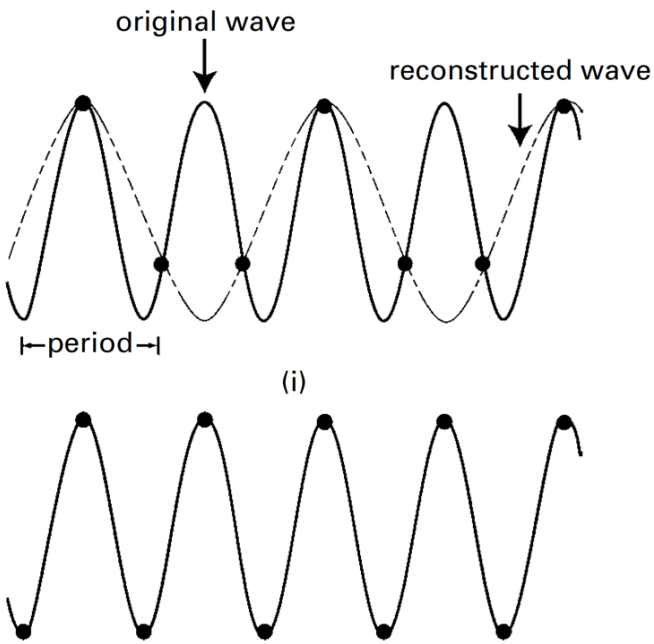
  Fourierjevo analizo lahko uporabimo tudi za neperiodične signale, kjer predpostavimo, da je perioda neskončna.

Demo.

IZREK O VZORČENJU

- Nyquist-Shannonov izrek:
  Da lahko rekonstruiramo signal brez napake, mora biti frekvenca vzorčenja signala, f_s, dvakrat večja od največje frekvence, ki je vsebovana v signalu.

$$f_s \geq 2 \times f_{\max}$$ ali $$f_N \geq f_{max}$$



$$f_s/2 = f_N$$

Nyquistova frekvenca

Zajemanje govornega signala

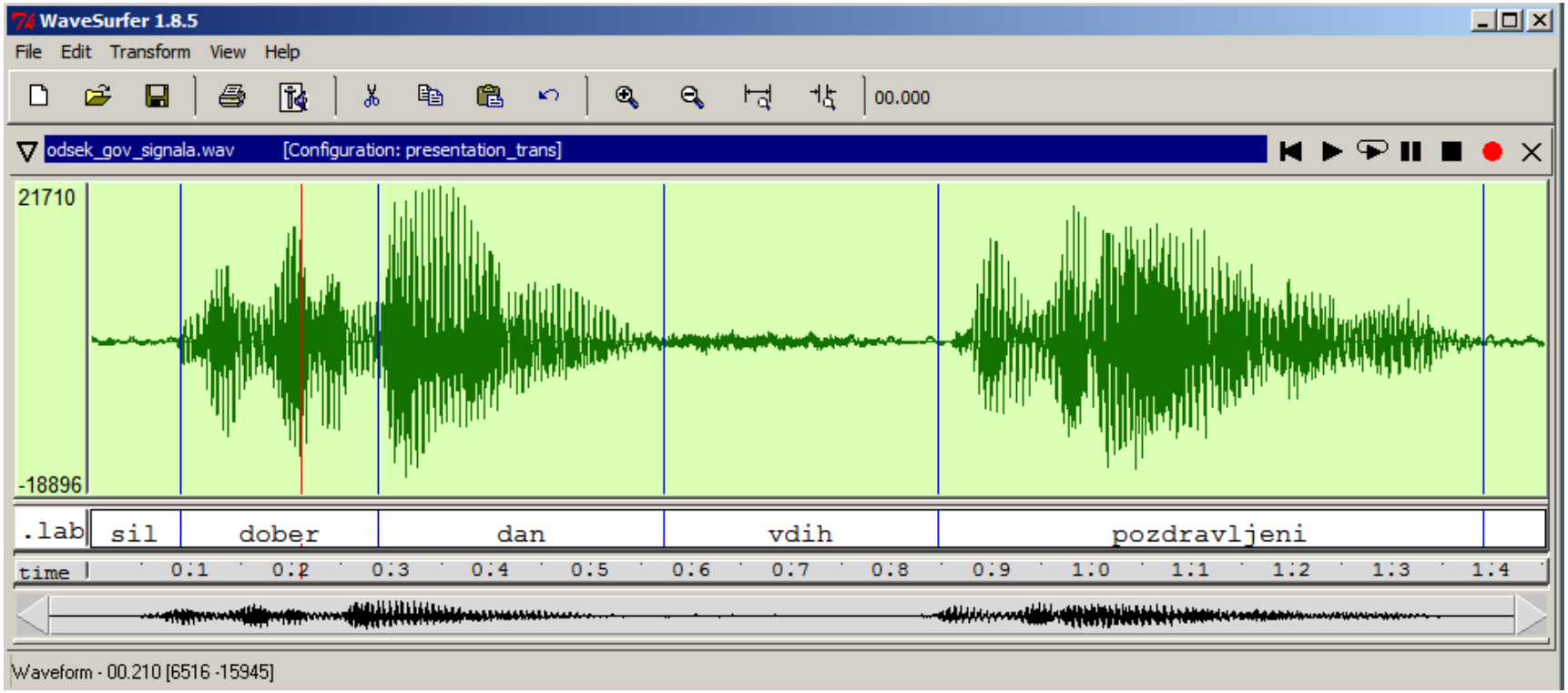
- Mikrofon: pretvorba zvočnega valovanja v analogen električni signal (osnovni princip: zvočno valovanje povzroči nihanje membrane v elektro-magnetnem polju - več vrst izvedb, obrat: zvočnik)

- **Digitalizacija** (A/D pretvornik: zvočna kartica):
  - vzorčenje – diskretizacija po času
  - kvantizacija – diskretizacija po amplitudi

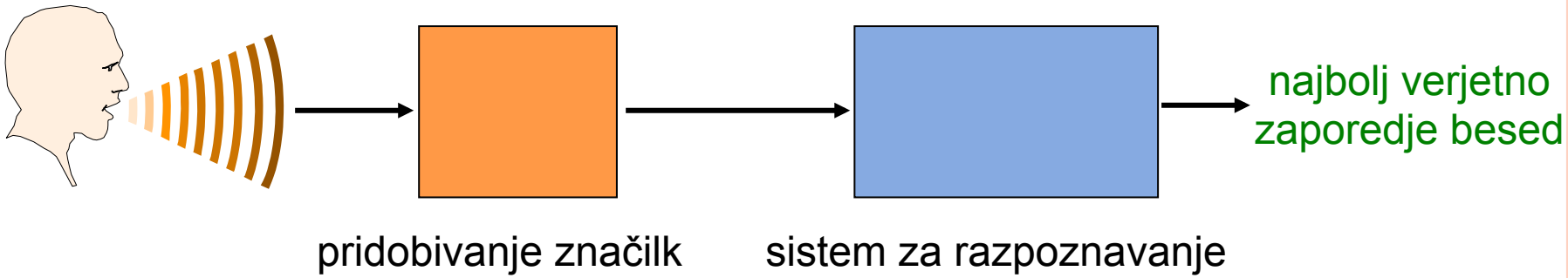
  - V praksi uporabljamo naslednje vrednosti:
    - frekvenca vzorčenja: 8 kHz (telefon) – 16 kHz
    - natančnost kvantizacije: 10 – 16 bit-ov (običajno od - 2^{15} do $+2^{15}$)

- *Pri frekvenci vzorčenja 16 kHz in 16-bitni kvantizaciji, človek ne zazna več razlik pri reprodukciji digitalnega signala v primerjavi z originalom.*

Govorni signal



Razpoznavanje govora z računalnikom

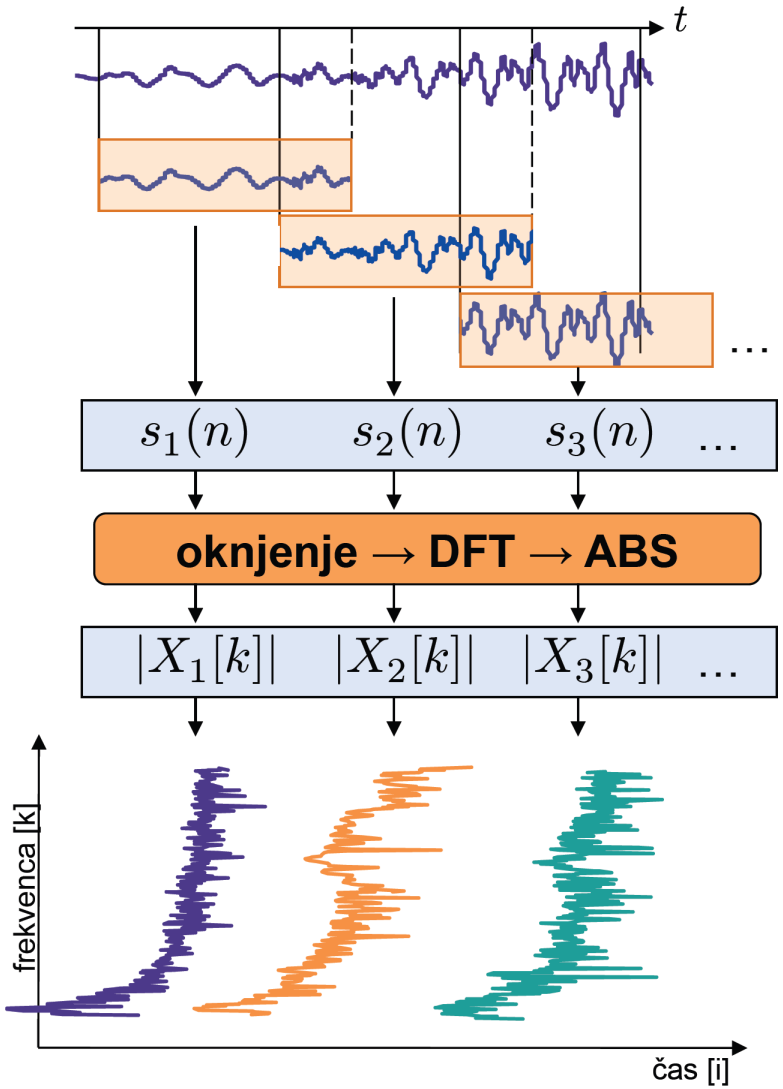


- Osnovni princip delovanja sistema:
  - <u>zaporedje vektorjev značilk:</u>
    - govorni signali pretvorimo v zaporedje vektorjev značilk,
  - <u>razpoznavanje:</u>
    - na podlagi statističnega modela določamo najbolj verjetno zaporedje besed glede na vhodno zaporedje vektorjev značilk,
    - statistične modele ‘učimo’ iz predhodno označenih posnetkov.

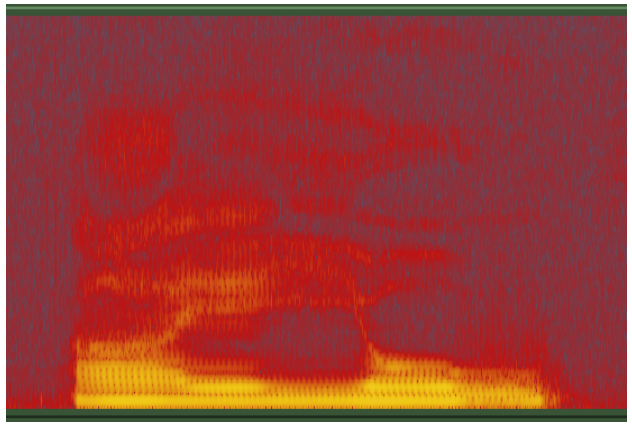
Obdelava govornih signalov za razpoznavanje govora

- Vektorji značilk:
  - Izvajati je potrebno frekvenčno analizo govora.
  - Ker se "narava" zvočnega signala od glasu do glasu spreminja, je potrebno opraviti frekvenčno analizo govornega signala na dovolj ozkih časovnih intervalih, ki hkrati ne zajamejo več kot en glas (5 – 200 msec).
  - Ker s sluhom ne zaznamo sprememb v faznem spektru, je pri frekvenčni analizi dovolj, da opazujemo le amplitudni spekter.
  - Značilke naj nosijo informacije o "stanju" govornega trakta pri tvorbi posameznih glasov – modeliranje govornega trakta.

Spektrogram = časovno-frekvenčna predstavitev



SPEKTROGRAM

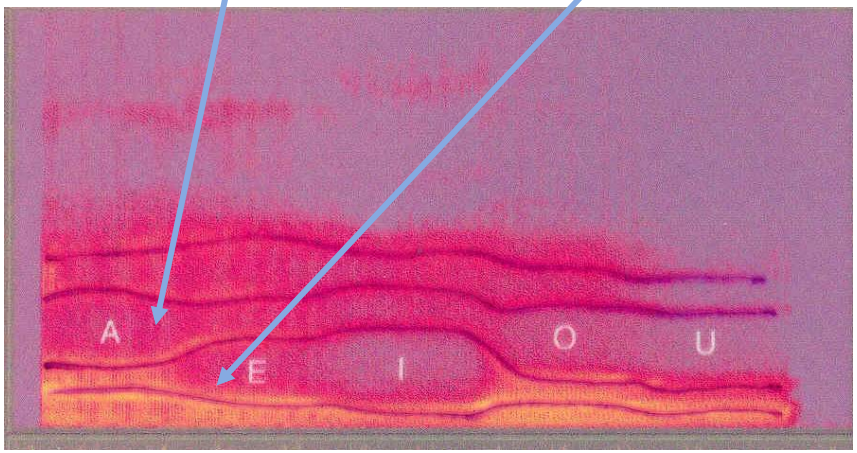


Kaj vidimo v spektrogramu? (demo)

Formanti - resonančne frekvence govornega trakta pri izreku določenega glasu.

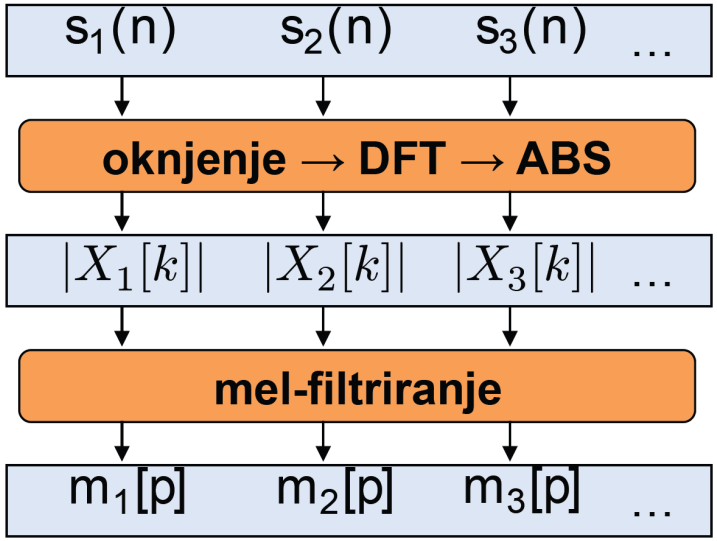
formantna frekvenca **F2**

formantna frekvenca **F1**

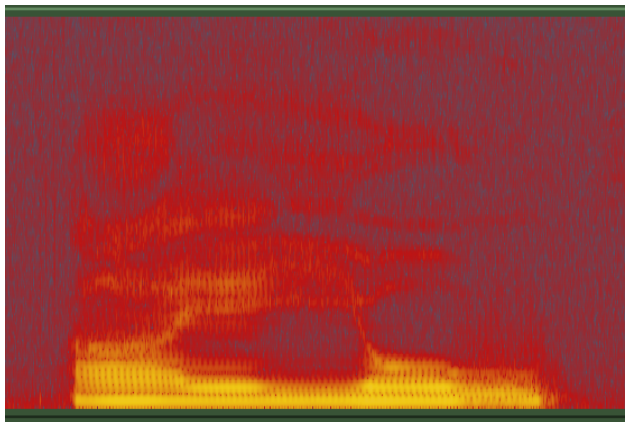


Samo-glasniki	**F1** (Hz)	glasnost	**F2** (Hz)	glasnost
u	**300**	močna	**900**	šibka
o	500	močna	1000	šibka
a	**950**	močna	**1500**	šibka
e	700	šibka	2000	močna
i	300	šibka	3000	močna

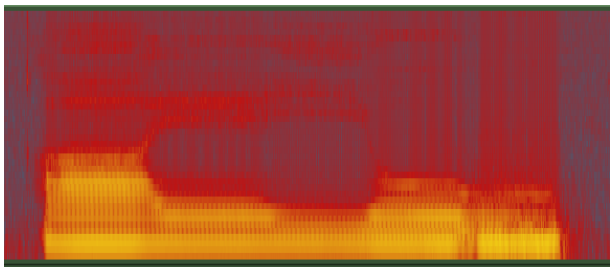
Melodični spekter



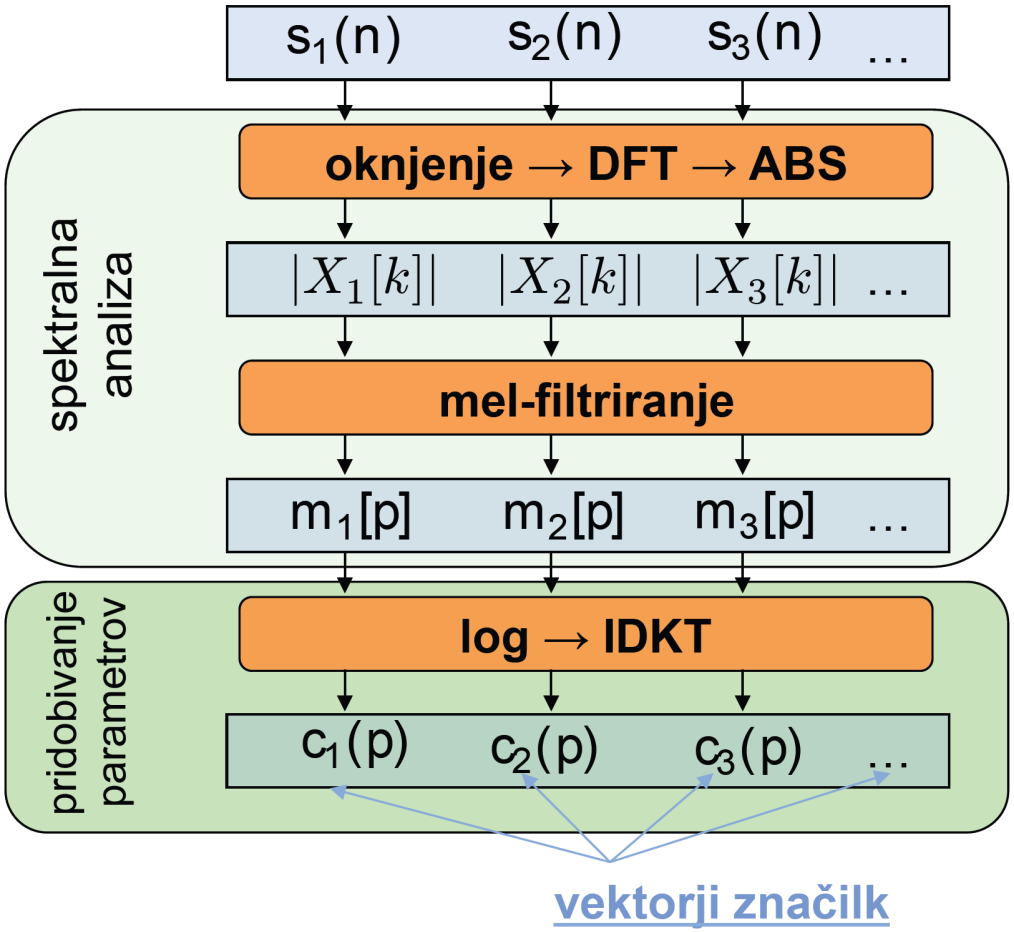
spekter

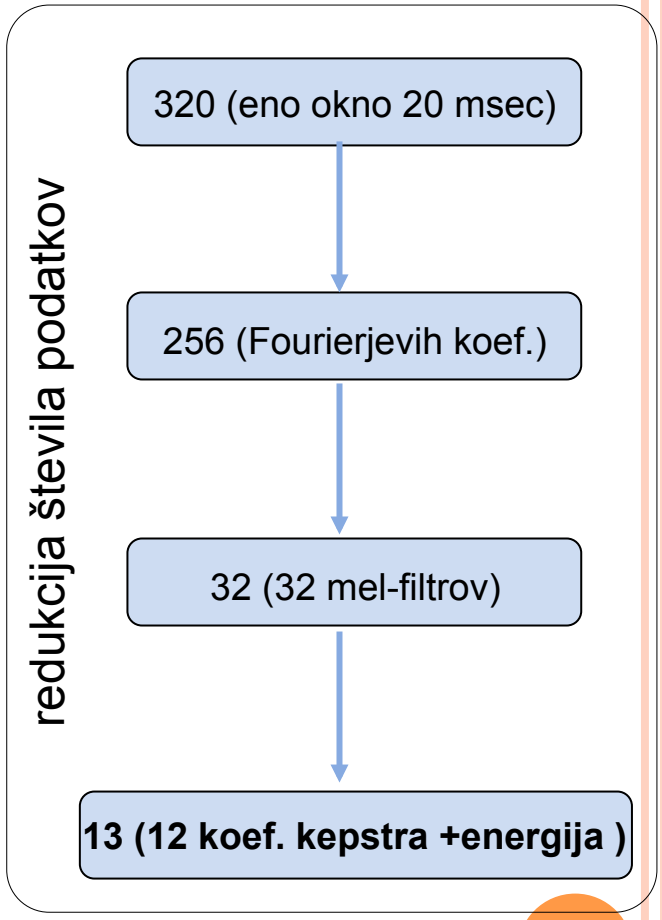


melodični spekter

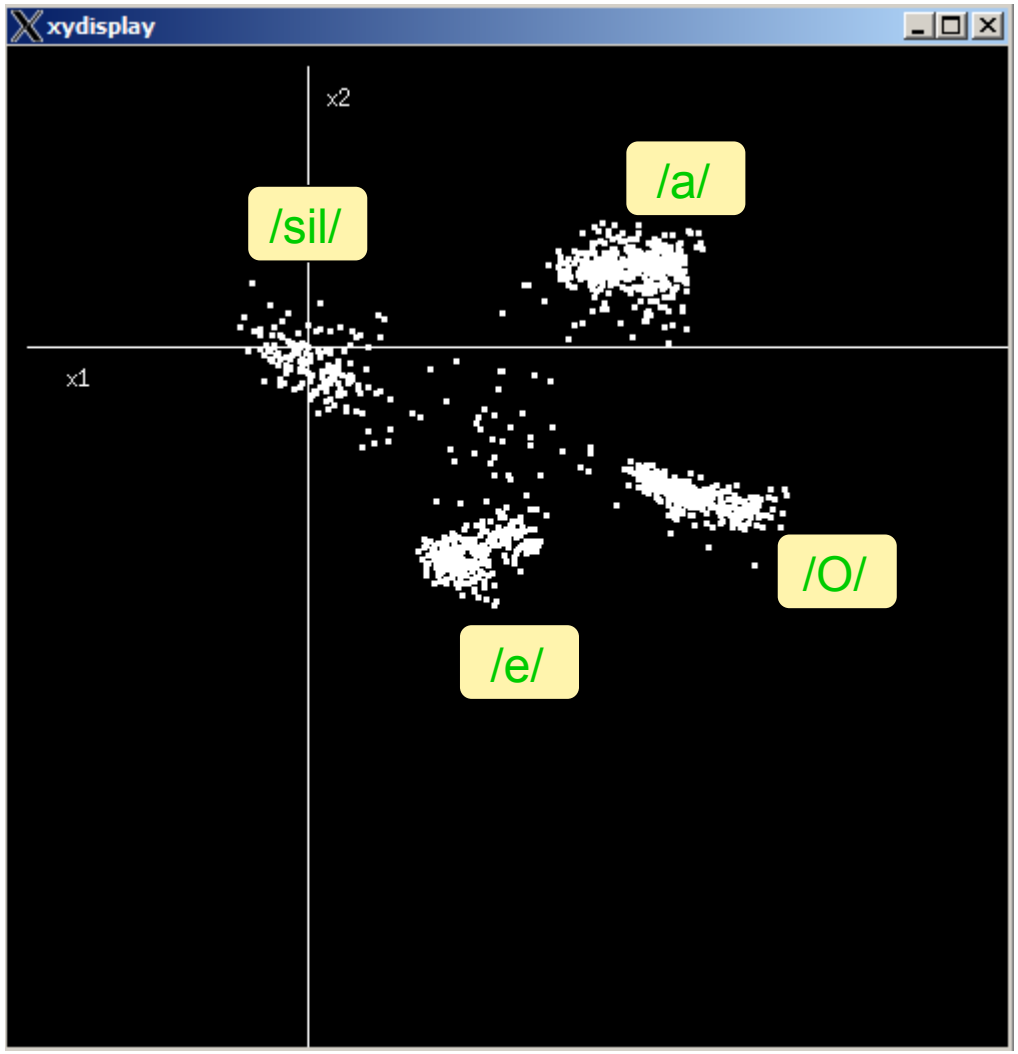


Koeficienti melodičnega kepstra





Razlikovanje glasov s koeficienti melodičnega kepstra

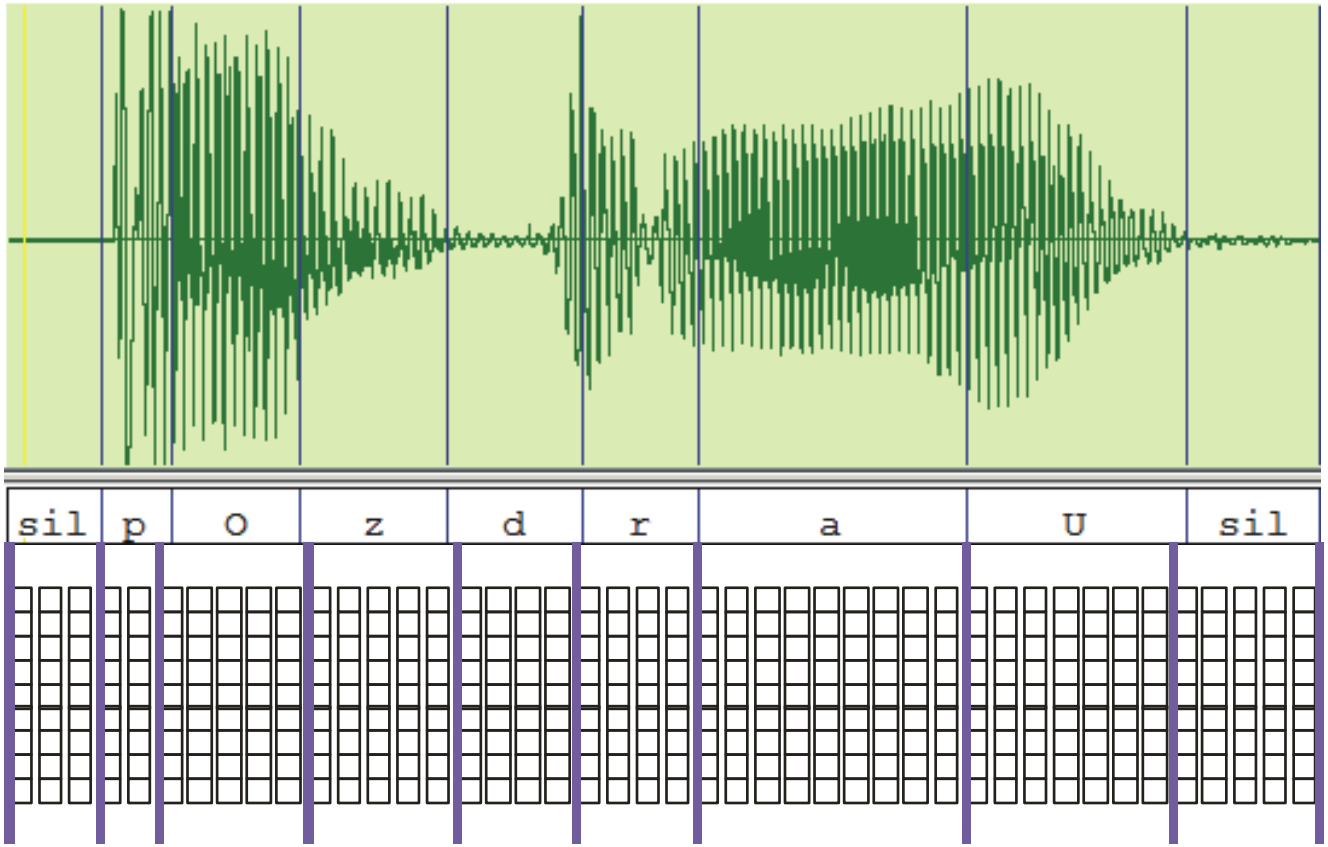


Kaj smo dobili?

signal

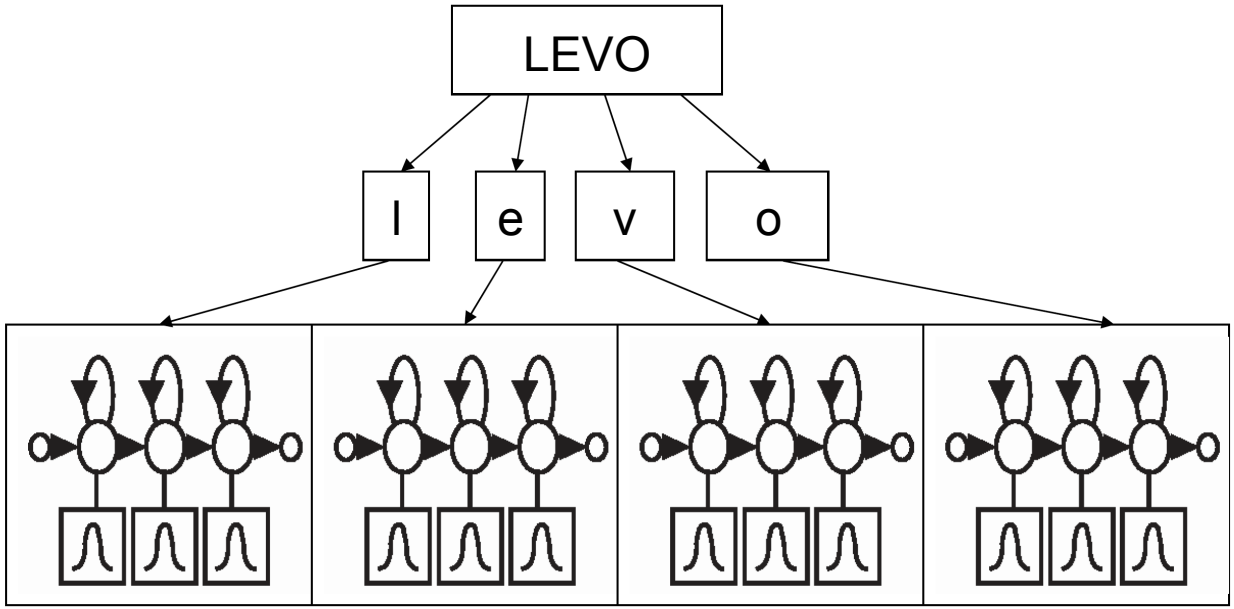
labele glasov

vektorji značilk



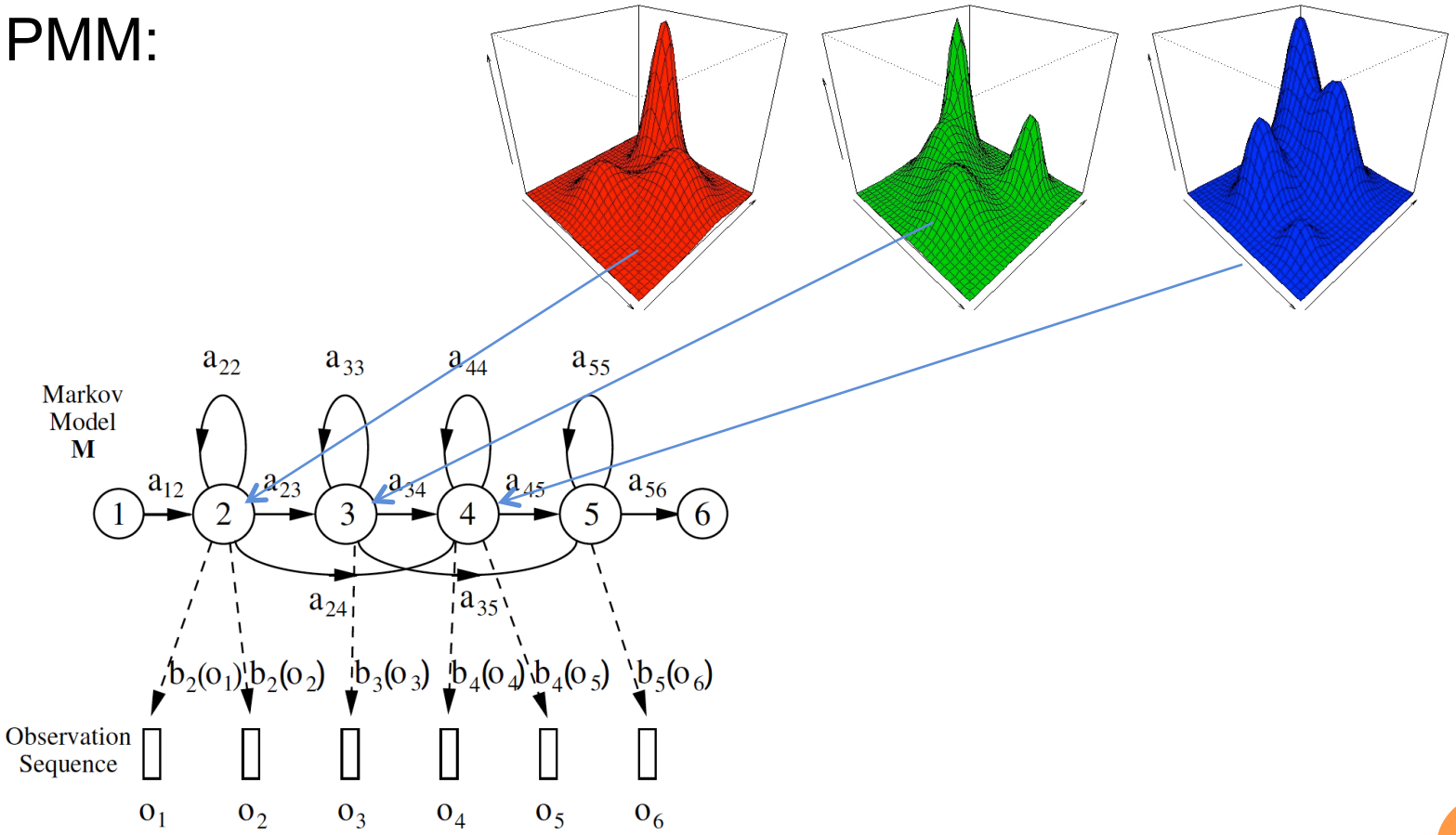
RAZPOZNAVANJE GOVORA: AKUSTIČNI MODELI

- Besede so sestavljene iz glasov.
- Vsak glas modeliramo z akustičnimi modeli:
  - PRIKRITI MARKOVOVI MODELI - HMM



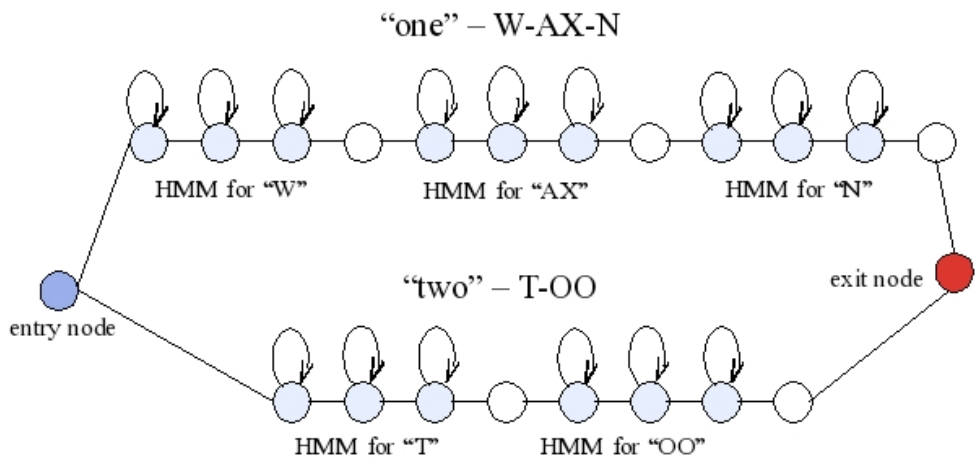
RAZPOZNAVANJE GOVORA: AKUSTIČNI MODELI

- PMM:



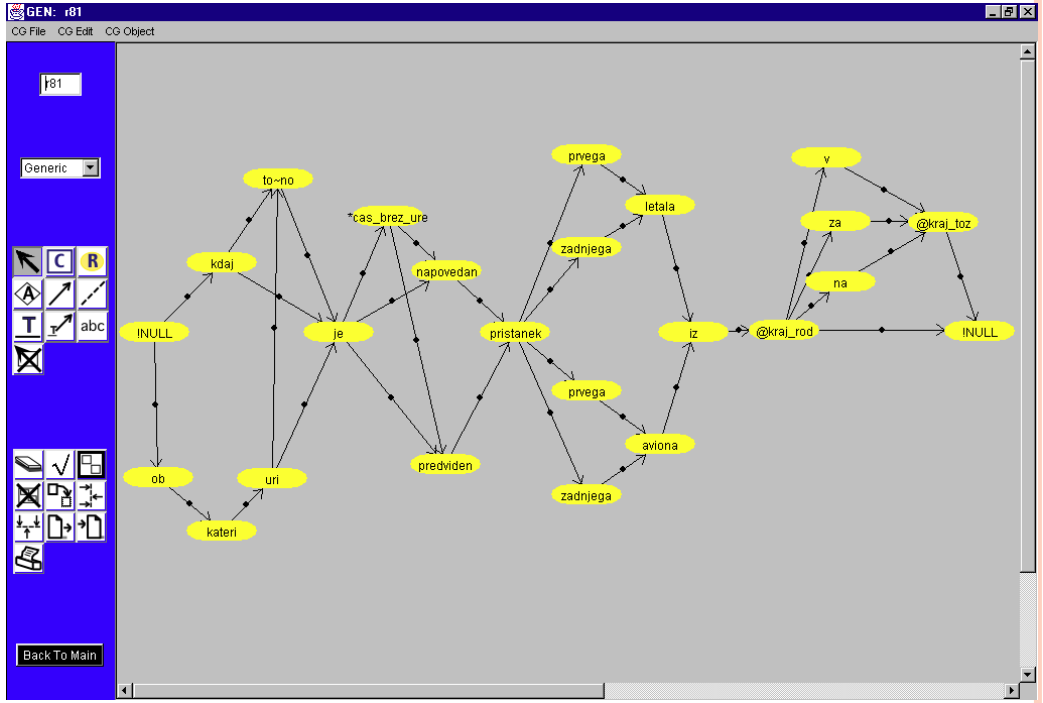
RAZPOZNAVANJE GOVORA: GRAF PMM

- Akustične modele posameznih glasov združimo skupaj glede na besede, ki jih želimo razpoznavati.
- Tako dobimo velik graf PMM.

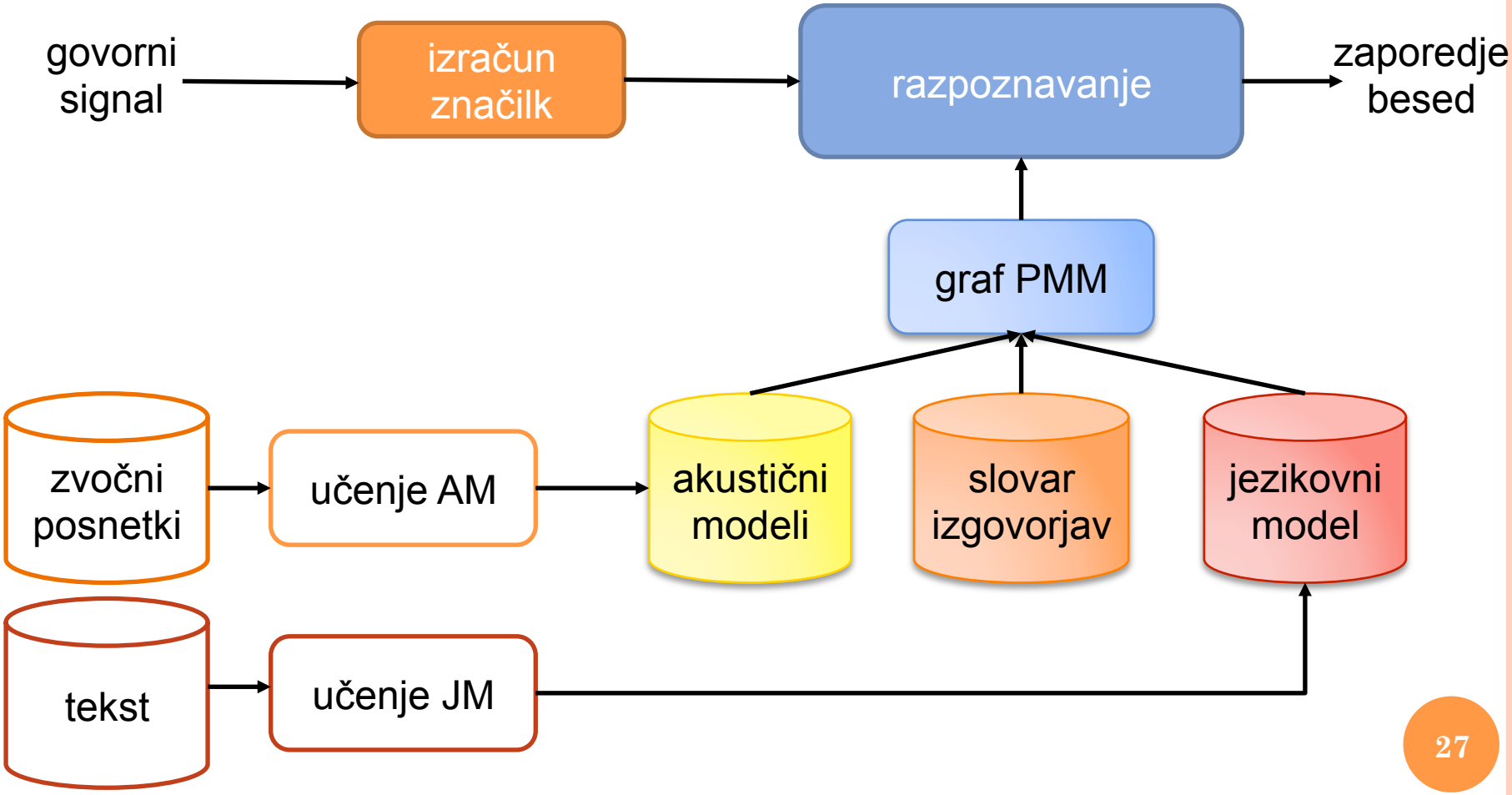


KAKO TVORIMO GRAF PMM: JEZIKOVNI MODEL

- Jezikovni model:
  - definiran s pravili:
    - potrebno jih je določiti,
    - odvisen od področja uporabe,
    - razpoznavanje govora boljše
  - **verjetnostni model**:
    - vsako zaporedje besed je možno, vendar različno verjetno,
    - splošen model, primeren za širšo uporabo,
    - razpoznavanje govora manj zanesljivo.



OSNOVNA SHEMA SISTEMA ZA RAZPOZNAVANJE GOVORA



Razpoznavanje govora: postopek

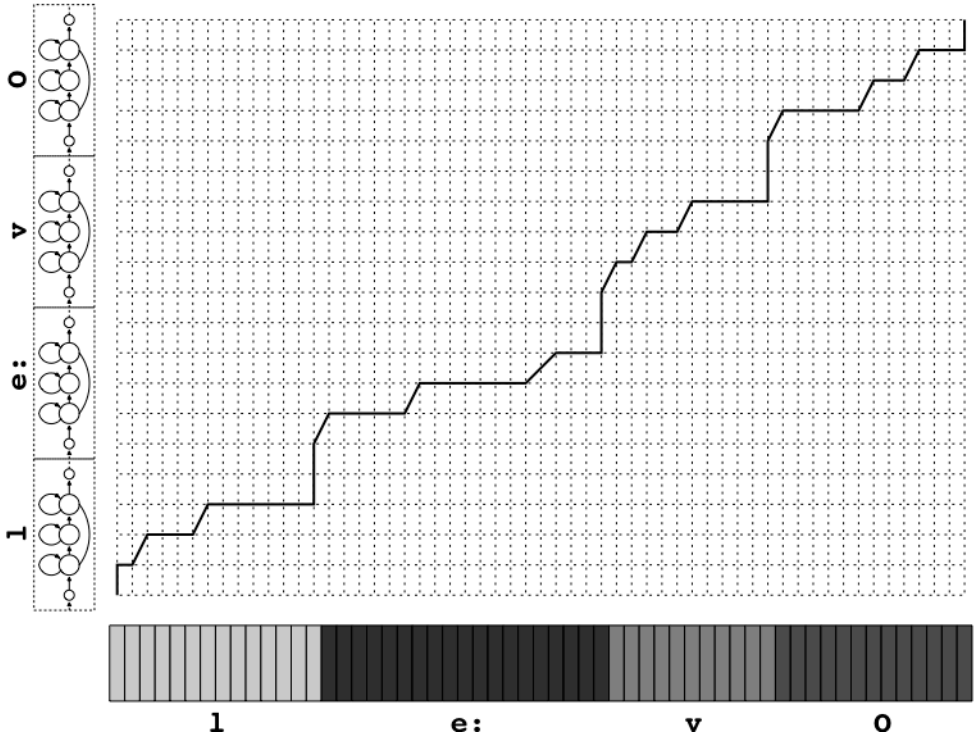
- **<u>Razpoznavanje:</u>**

  - <u>Osnovna naloga</u>:

    zaporedje vektorjev značilk, ki jih izračunamo iz govornega signala mora dobiti oznake posameznih glasov (besed).

  - <u>Postopek</u>:
    Viterbijevo dekodiranje

    Iščemo najbolj verjetno kombinacijo besed (sestavljeno iz glasov), ki ustreza vhodnemu zaporedju vektorjev značilk.



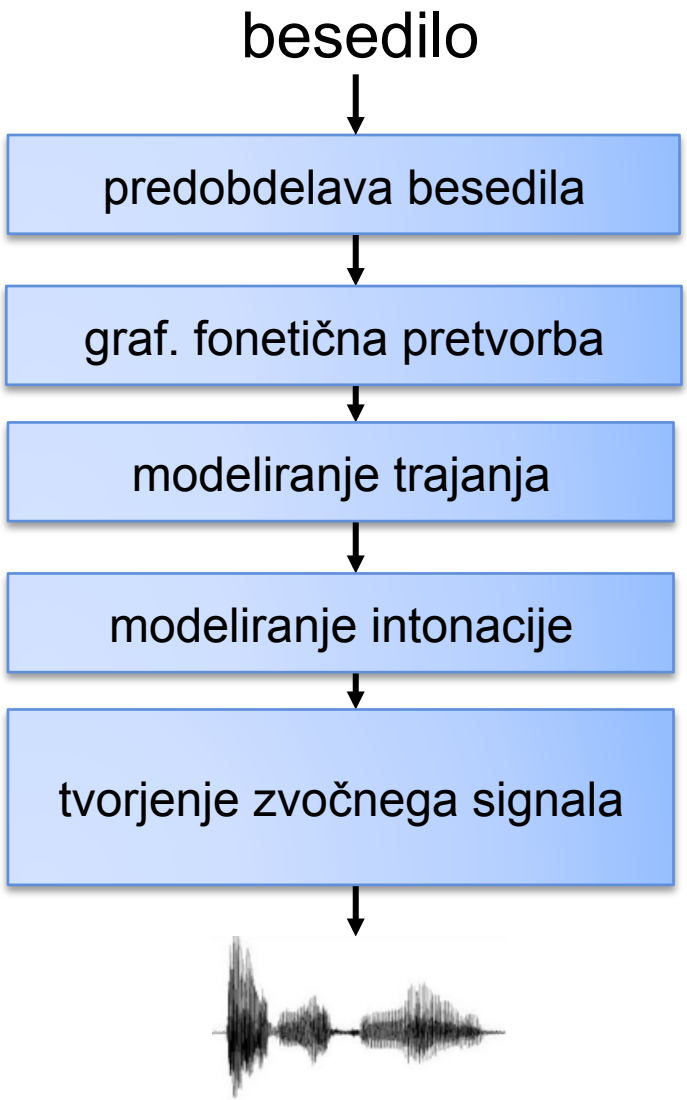
Razpoznavanje govora - primeri

- razpoznavanje števk:
  - število besed, ki jih je potrebno razpoznati, je 10.
  - zato je iskanje najbolj verjetnih zaporedji besed enostavno: iskalni prostor je majhen.
  - napak pri razpoznavanju je malo.

- razpoznavanje vremenskih napovedi:
  - število besed, ki jih je potrebno razpoznati, je okoli 2300.
  - iskalni prostor je večji: večja mreža besed.
  - napak pri razpoznavanju je več.

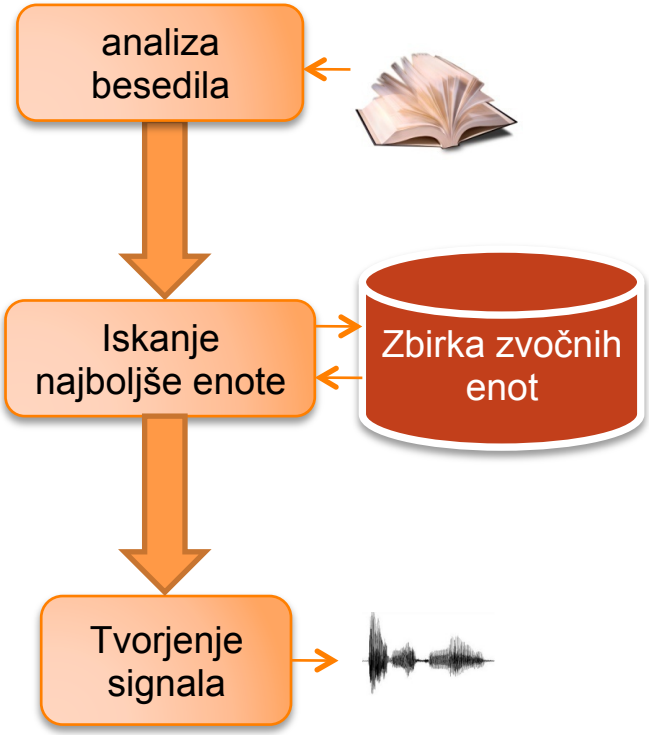
UMETNO TVORJENJE GOVORA: SINTEZA GOVORA

- Gre za obraten proces kot pri razpoznavanju govora.
- Računalnik zna izgovoriti poljuben tekst, ki mu ga zapišemo.



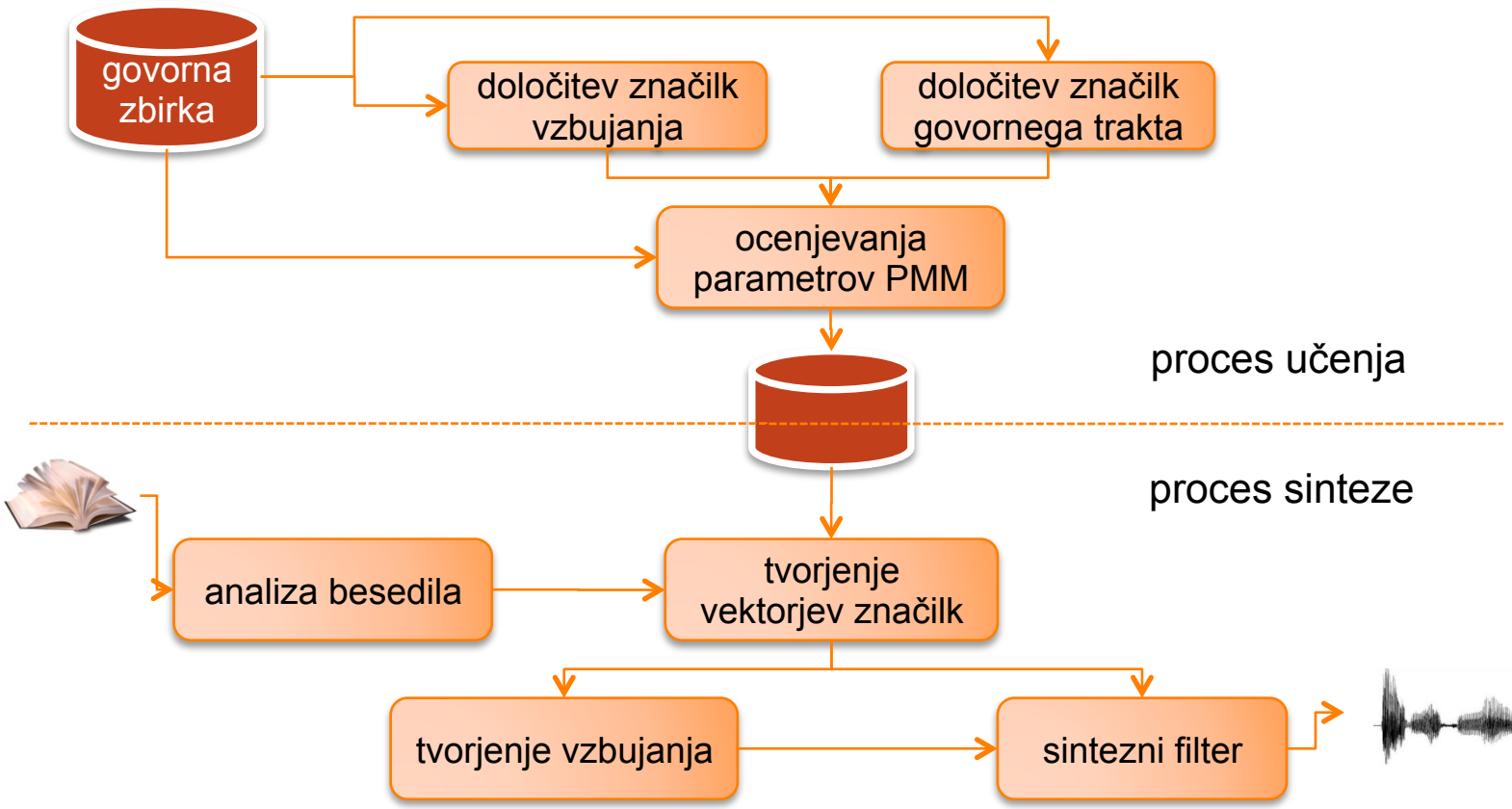
UMETNO TVORJENJE GOVORA: SINTEZA GOVORA

- Korpusna sinteza (*ang .unit-selection synthesis*)
  - visokokavosten govor dosežemo z združevanjem vnaprej posnetih govornih enot
  - tovrsten pristop potrebuje veliko število govornih posnetkov
  - naloga: poskušamo najti najbolj primerne enote govora – problem iskanja



UMETNO TVORJENJE GOVORA: SINTEZA GOVORA

- Statistični pristop: uporaba PMM
  - lahko enostavno vplivamo na kvaliteto sinteze govora s spremembo parametrov PMM



Umetno tvorjenje govora: Sinteza govora

- Primeri sinteze govora v slovenščini:
  - Difonska sinteza.
  - PMM sinteze: različni glasovi.
  - Povečevanje naravnosti govora.

Zaključek

- Govorne tehnologije:
  - sistemi za razpoznavanje govora,
  - sistemi za umetno tvorjenje govora,
  - sistemi za dialog med človekom in računalnikom,
  - sistemi za samodejno prevajanje med jeziki,
  - sistemi za identifikacijo govorcev,
  - sistemi za detekijo in analizo čustvenega stanja govorca.

- Razvoj govornih tehnologij izjemno odvisen od jezika:
  - potrebno je pridobiti veliko označenih zvočnih posnetkov,
  - potrebno je upoštevati značilnosti in posebnosti jezika,
  - drugi nam ne bodo razvili takšnih sistemov.