

Predstavitev tem za zaključne naloge

doc. dr. Janez Žibert

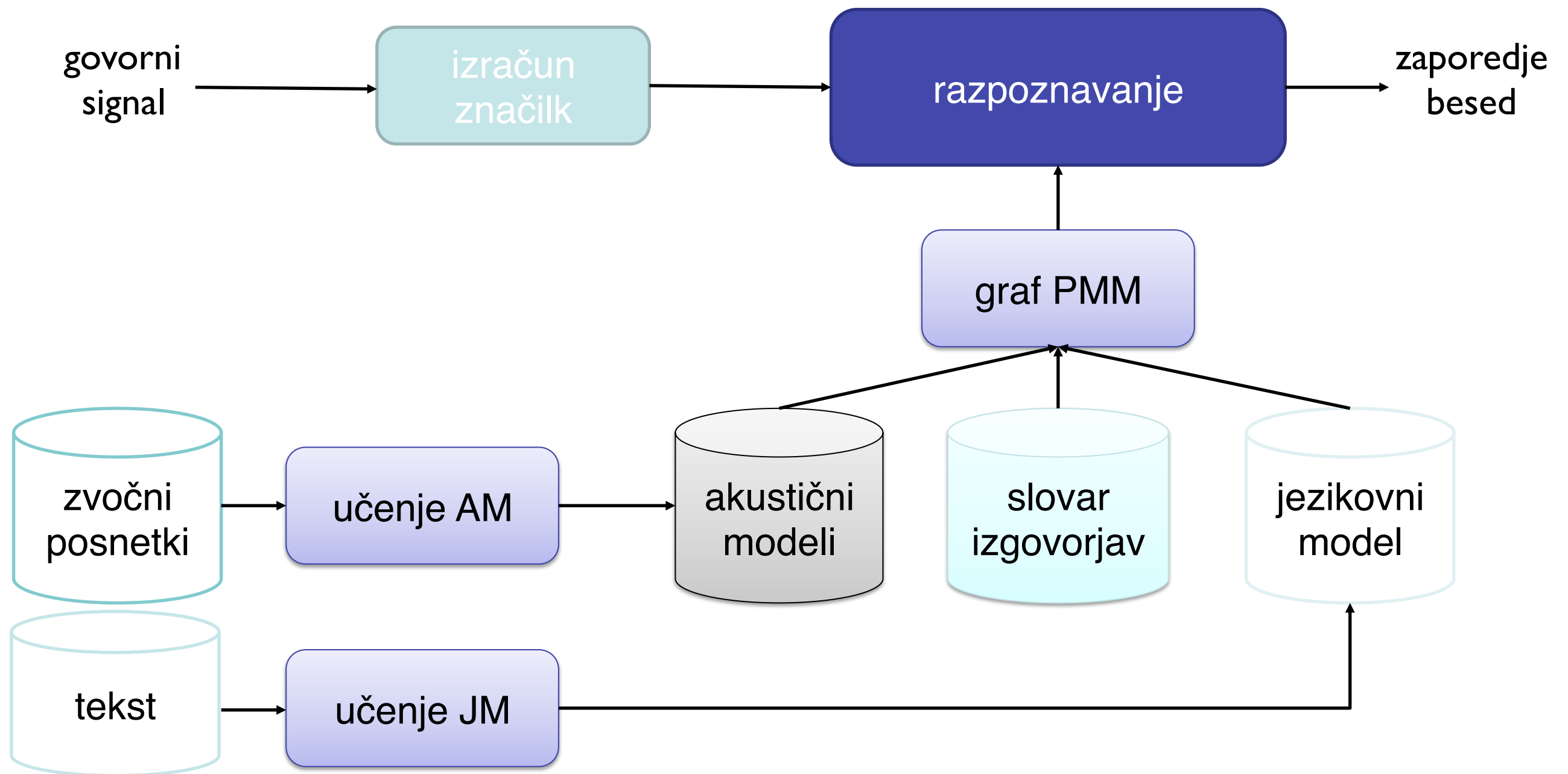
Predstavitev tem

- Govorne tehnologije
 - razpoznavanje govora z orodjem Sphinx,
 - razpoznavanje čustvenih stanj v govoru,
 - indeksiranje posnetkov po govorcih/razpoznavanje govorcev,
 - tvorba slovenskega govora v sistemu Mary TTS.
 - jezikovno modeliranje
- Obdelava slik in videa
 - aplikacije s senzorjem XBox Kinect: aplikacije prepoznavanja gibov za izgradnjo t.i. pametnih vmesnikov
 - analiza satelitskih slik: GUI vmesniki za označevanje in analizo sat. slik
- Statistično modeliranje
 - analiza okoljskih podatkov,

Govorne tehnologije

- govorni signali:
 - parametrizacija signala: spektrogram, melodični spekter, koeficienti melodičnega kepstra.
- razpoznavanje govora:
 - osnovna zgradba sistemov za razpoznavanje govora,
 - predstavitev praktičnih aplikacij razpoznavanja govora.
- tvorjenje govora – sinteza govora:
 - izvedba HMM sinteze slovenskega govora v sistemu Marry TTS.
- jezikovno modeliranje:
 - izvedba n-gramskega modeliranja za ugotavljanje podobnosti med dokumenti

Osnovna shema sistema za razpoznavanje govora



Govorne tehnologije

- Razpoznavanje govora z orodjem [Sphinx](http://cmusphinx.sourceforge.net/)
<http://cmusphinx.sourceforge.net/>

CMU Sphinx Downloads

Software

CMU Sphinx toolkit has a number of packages for different tasks and applications. It's sometimes confusing what to choose. To cleanup, here is the list

- Pocketsphinx — recognizer library written in C.
- Sphinxbase — support library required by Pocketsphinx
- Sphinx4 — adjustable, modifiable recognizer written in Java
- CMUclmtk — language model tools
- Sphinxtrain — acoustic model training tools

We recommend you to use the latest available releases:

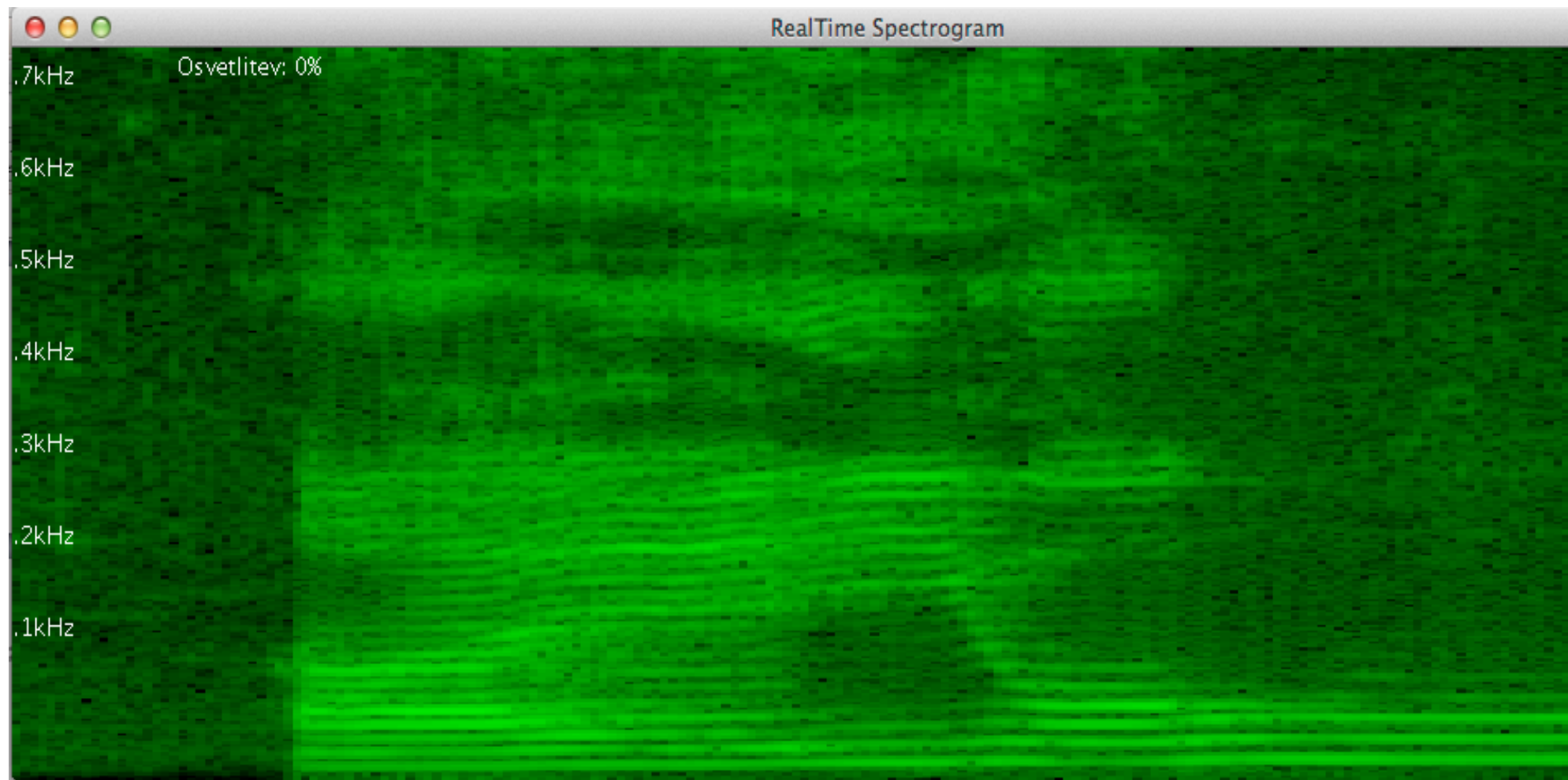
- [sphinxbase-0.7](#)
- [pocketsphinx-0.7](#)
- [sphinx4-1.0beta6](#)
- [sphinxtrain-1.0.7](#)
- [cmuclmtk-0.7](#)

If you want to try bleeding edge version, download subversion snapshots or checkout from subversion. Then compile packages from the source code, but they will be stable:

- [pocketsphinx-snapshot](#)
- [sphinxbase-snapshot](#)
- [sphinx3-snapshot](#)
- [sphinx4-snapshot](#)
- [cmuclmtk-snapshot](#)
- [sphinxtrain-snapshot](#)

Govorne tehnologije

- Parametrizacija zvočnih signalov:



Govorne tehnologije

- Razpoznavanje govora z orodjem [Sphinx](#)

```
janez@vox-mac: ~/Work/Demos/sphin
[java] Used: This: 145.15 Mb Avg: 144.99 Mb Max: 145.15 Mb
[java] # ----- linguist stats -----
[java] # Total states: 2251
[java] # class edu.cmu.sphinx.linguist.flat.HMMStateState: 1191
[java] # class edu.cmu.sphinx.linguist.flat.BranchState: 132
[java] # class edu.cmu.sphinx.linguist.flat.PronunciationState:
[java] # class edu.cmu.sphinx.linguist.flat.NonEmittingHMMState:
[java] # class edu.cmu.sphinx.linguist.flat.GrammarState: 1
[java] # class edu.cmu.sphinx.linguist.flat.ExtendedUnitState: 3
[java] ... done changing
[java] set to null
[java]
[java] REF:      one four seven three five three three
[java] HYP:      oh
[java]
[java] Accuracy: 0.000%  Errors: 7 (Sub: 1 Ins: 0 Del: 6)
[java] Words: 7  Matches: 0  WER: 100.000%
[java] Sentences: 1  Matches: 0  SentenceAcc: 0.000%
[java] This Time Audio: 3.10s Proc: 3.04s Speed: 0.98 X re
[java] Total Time Audio: 3.10s Proc: 3.04s Speed: 0.98 X re
[java] Mem Total: 211.65 Mb Free: 54.99 Mb
[java] Used: This: 156.66 Mb Avg: 156.66 Mb Max: 156.66 Mb
[java] <sil> SIL SIL SIL SIL SIL SIL oh OW_oh
```

Live Decoder!

Decoder:

Test file:

Say:

Recognized:

Statistics:

Word Accuracy: 0.0%

Sentence Accuracy: 0.0%

Speed: 0.98 X RT

Cumulative Speed: 0.98 X RT

Stop speaking

Govorne tehnologije

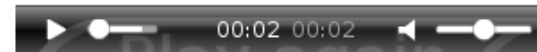
- Razpoznavanje čustvenih stanj v govoru



| | | | | | | |
|-------|------|--------|-------|-------------------|---------------|-------|
| Domov | News | Events | Users | Aktivni prispevki | sokrat-track1 | šefka |
|-------|------|--------|-------|-------------------|---------------|-------|

Nahajate se tu: [Domov](#) > [šefka](#) > [šefka-Track-1-sp1-528](#)

Anotate the media



1608.054 1609.398 **spk1** : manj kot drek na cesti
1609.398 1610.017 **spk1** : [diha]
1610.017 1611.322 **spk1** : mene še črvi ne bi hoteli
1611.322 1613.028 **spk1** : [vdih]
1613.028 1615.143 spk1 : Toplečka pa mi je dala župo
1615.143 1615.648 **spk1** : [vdih]
1615.648 1616.901 **spk1** : pa spalno srajco
1616.901 1617.417 **spk1** : [vdih]
1617.417 1618.339 **spk1** : pa na kolnkišti
1618.339 1619.426 **spk1** : sem lahko sedela

Please anotate media.

Mark appropriate ■

Please mark with your best ...

emocija 1 ▾

Mark appropriate ■

Please mark with your best ...

Toplečka
pa
je
dala

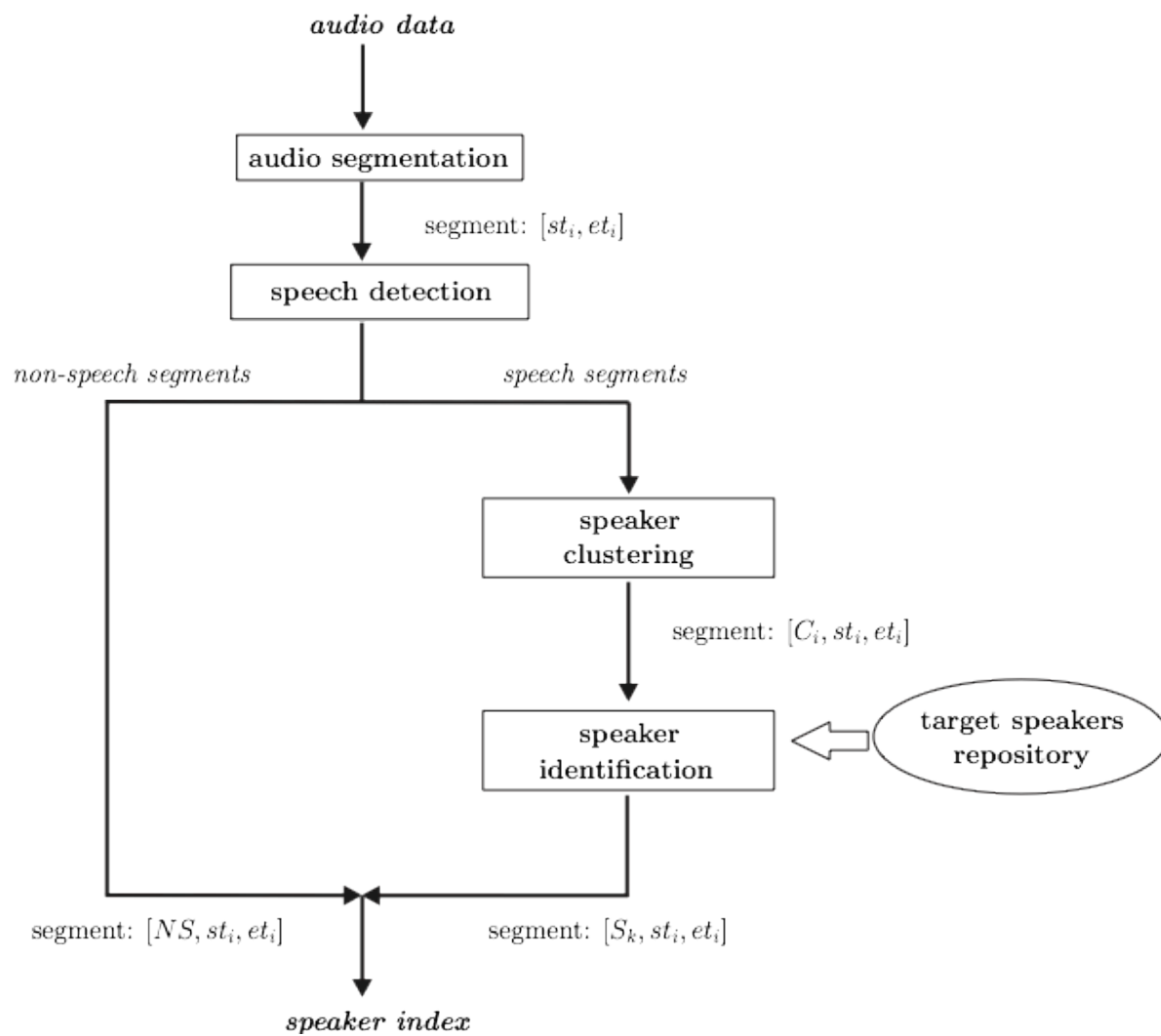
mi
župo

→ ← ↑ ↓

Anotate

Govorne tehnologije

- Indeksiranje posnetkov po govornih / razpoznavanje govorcev



The screenshot shows the 'Audio speaker search tool 2' interface. It features a menu bar (File, Tools, Help) and a 'Database statistics' section with the following data:

| | | | |
|----------------------------|-----|-----------------------------|----------|
| Loaded audio files: | 13 | Total audio time: | 12:09:59 |
| Loaded segmentation files: | 13 | Total speaker time: | 10:41:25 |
| Loaded score file: | yes | Number of tracked speakers: | 41 |

There are checkboxes for 'Add audio name to segments' (checked) and 'Remove speaker names from segments' (unchecked). A search field contains 'Manica J Ambrozic' and a 'Find speaker' button. A 'Score threshold' is set to 0.5. Below this is a table of search results:

| Speaker ID | Audio Name | # Segm... | Conf. Score |
|-------------------|---------------------|-----------|-------------|
| Manica J Ambrozic | dnevnik-110603-1900 | 6 | 1.128 |
| Manica J Ambrozic | dnevnik-150703-1900 | 26 | 0.995 |
| Manica J Ambrozic | dnevnik-100603-1900 | 24 | 0.890 |
| Manica J Ambrozic | dnevnik-090603-1900 | 26 | 0.844 |
| Manica J Ambrozic | dnevnik-130603-1900 | 33 | 0.841 |
| Manica J Ambrozic | dnevnik-110603-1900 | 1 | 0.806 |
| Manica J Ambrozic | dnevnik-110603-1900 | 14 | 0.744 |

A histogram titled 'Score distribution of speaker: Manica J Ambrozic' shows the probability distribution of speaker scores. On the right, a list of search results for 'Manica J Ambrozic' is displayed, showing time intervals and scores for various audio files. At the bottom, there are 'Play WAV' and 'Save WAV' buttons.

Govorne tehnologije

- Sinteza slovenskega govora:
 - z uporabo HMM v sistemu Mary TTS: <http://mary.dfki.de/>



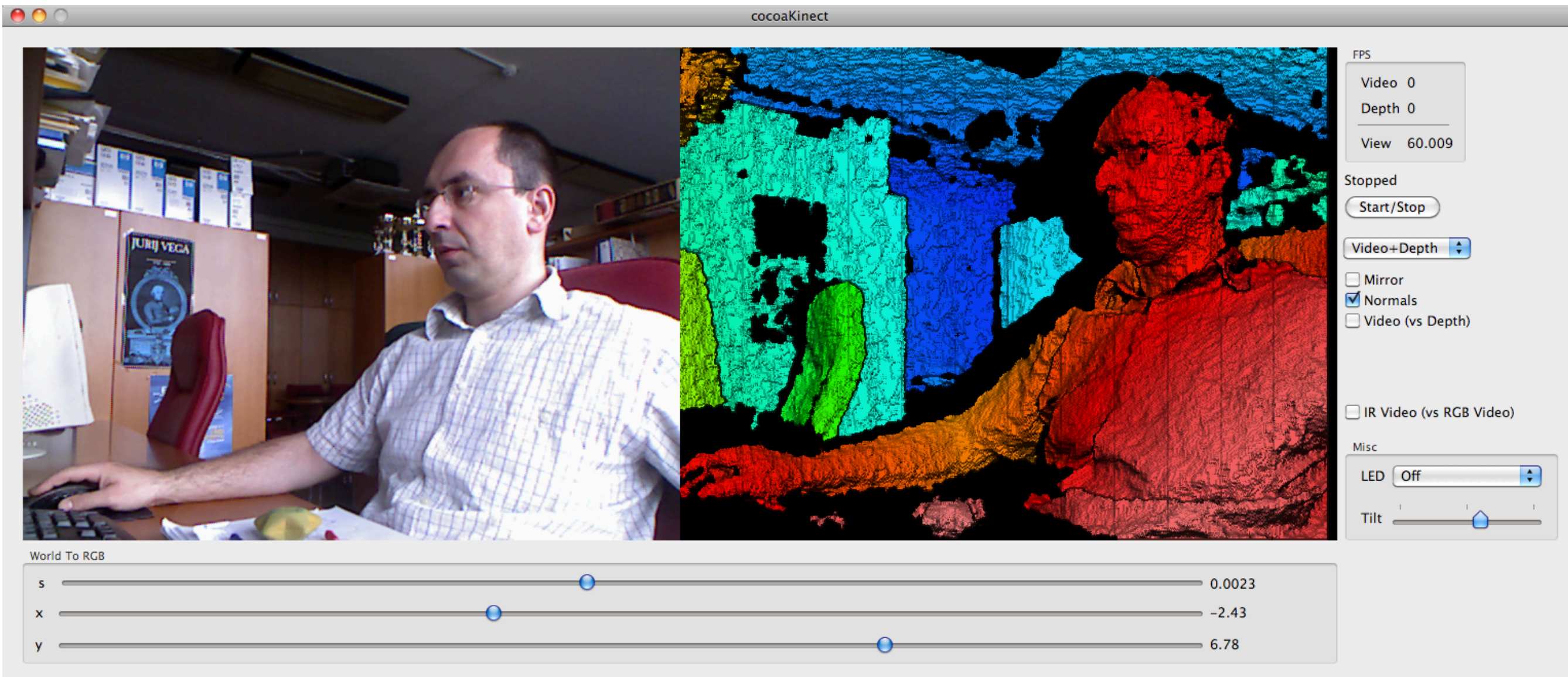
- Dodajanje novega jezika v sistem:
<https://github.com/marytts/marytts/wiki/New-Language-Support>

Jezikovno modeliranje

- Izgradnja n-gramskih jezikovnih modelov:
 - HTK Toolkit,
<http://htk.eng.cam.ac.uk/>
 - CMU Statistical Language Modeling (SLM) Toolkit:
<http://www.speech.cs.cmu.edu/SLM/toolkit.html>
 - Primerjava med modeliranj.
- Uporaba modelov za preverjanje podobnosti med dokumenti

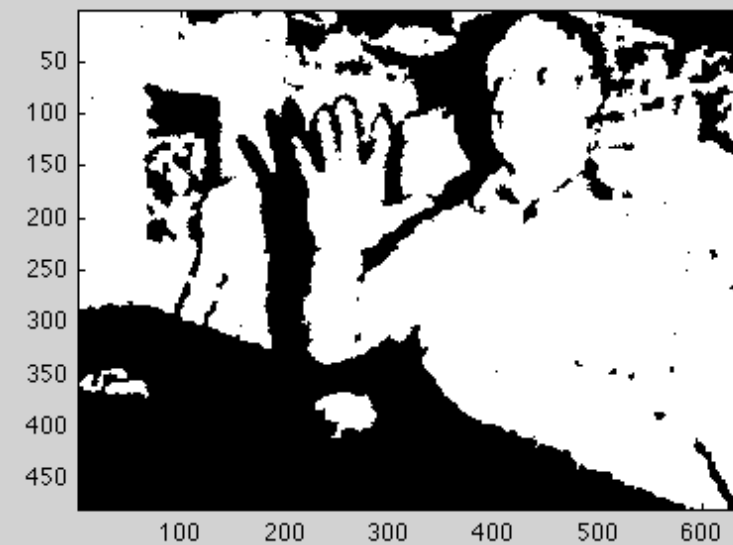
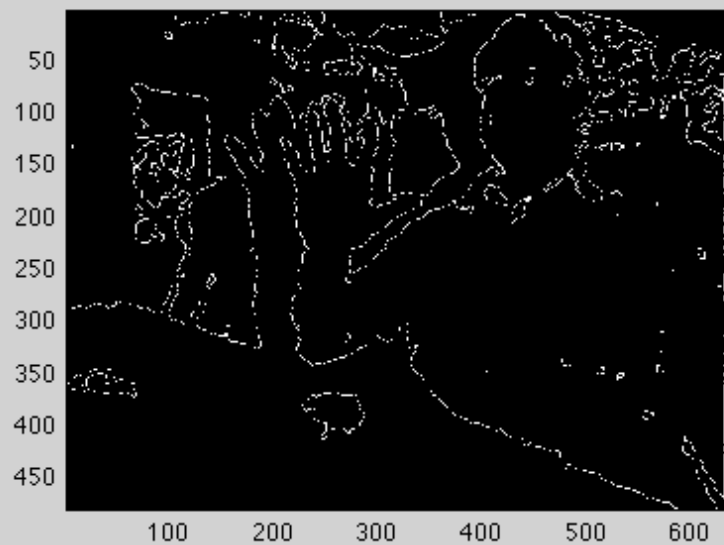
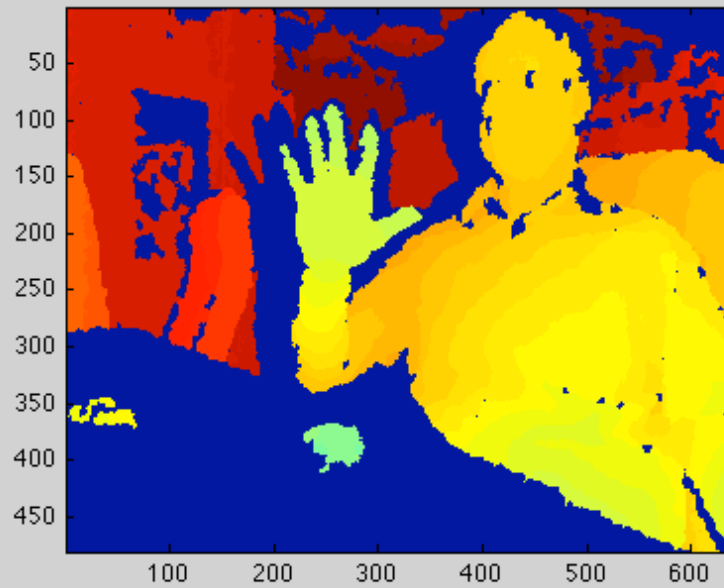
Obdelava videa in slik

- Aplikacije s senzorjem Kinect



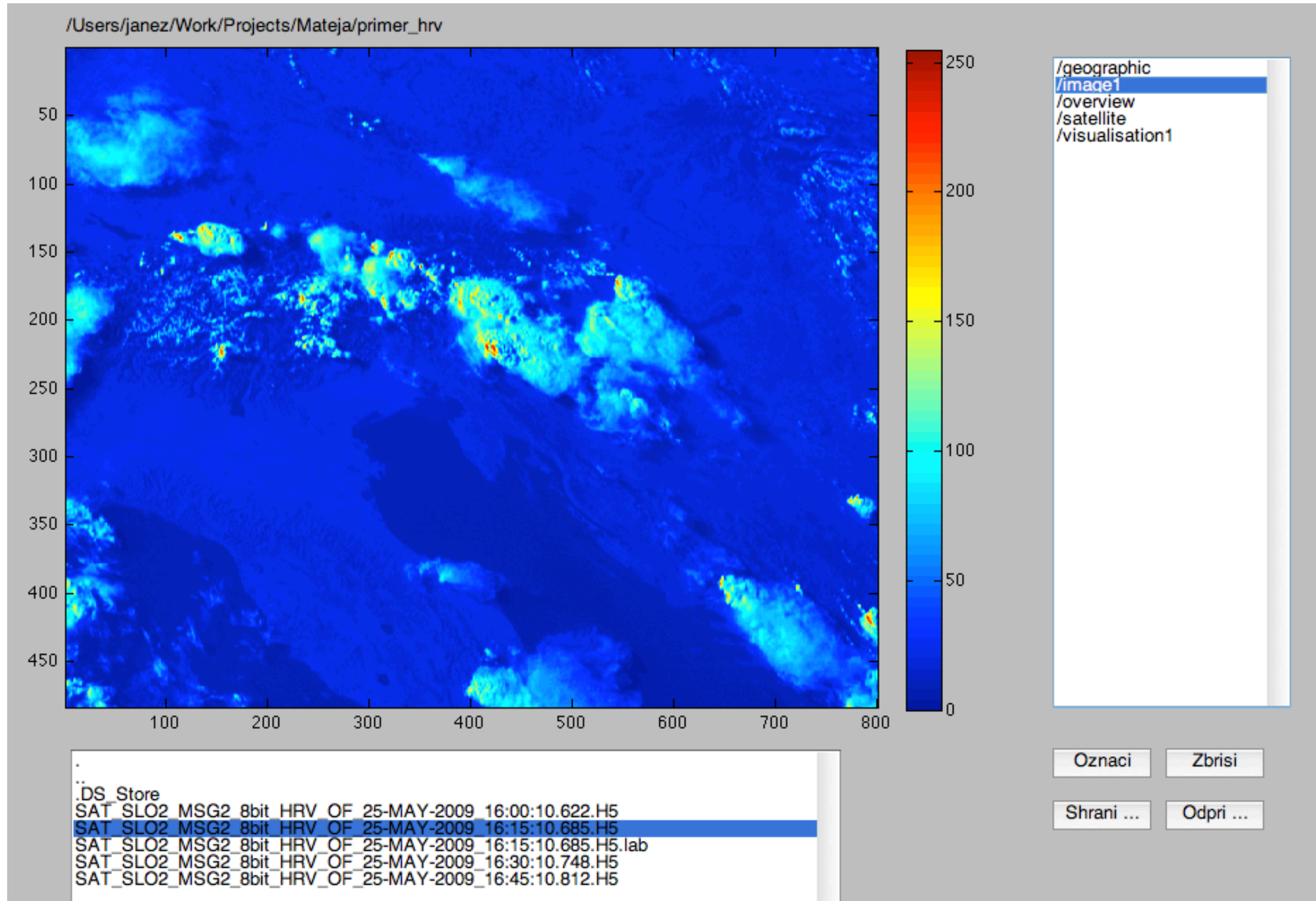
Obdelava videa in slik

- Aplikacije s senzorjem Kinect: Matlab



Obdelava slik

- Analiza satelitskih slik: detekcija premikanja oblačnosti



Statistično modeliranje

- Geopedia, <http://www.geopedia.si/>
iOS, Android:
 - Prikaz okoljskih in meteoroloških meritev v Kopru in okolici
 - Prikaz podatkov pridobljenih s pomočjo LIDAR meritev.
 - Prikaz podatkov o suši, pridobljenih iz različnih virov.

