St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences

HAVRUS Corpus: High-speed Recordings of Audio-Visual Russian Speech

Vasilisa Verkhodanova, Alexander Ronzhin,
Irina Kipyatkova, Denis Ivanko, Alexey Karpov,
and Miloš Železný

25 August 2016, Budapest, Hungary

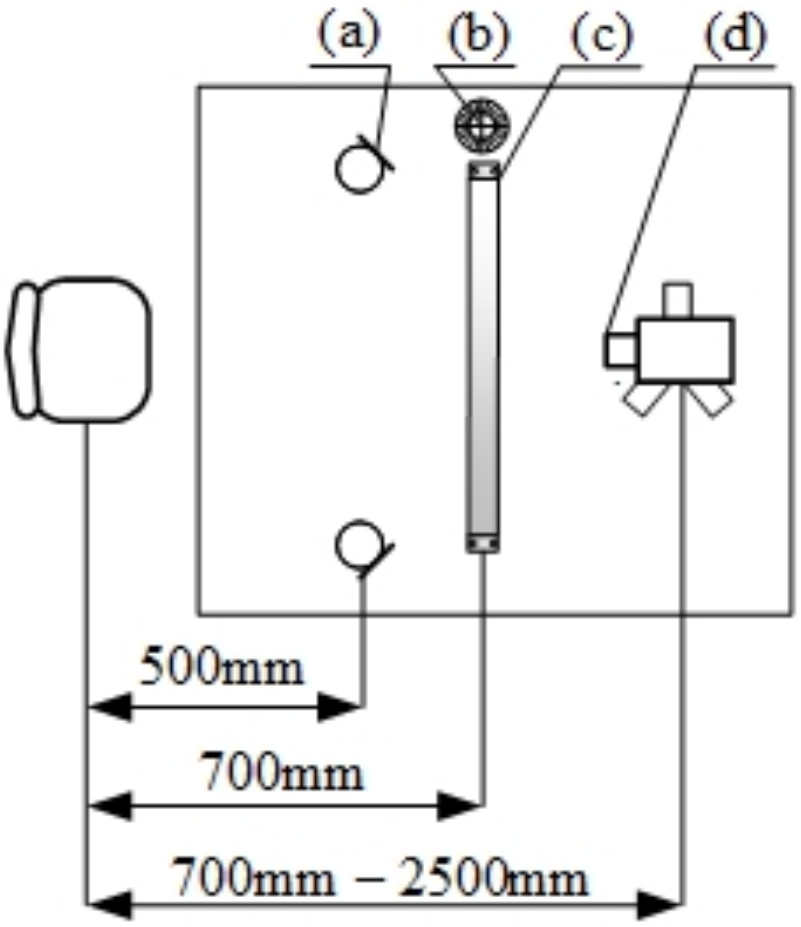
Presentation outline

- Approaches and aspects for development of audiovisual database
- Software architecture for recording audio-visual speech databases
- HAVRUS corpus description
- Conclusions

Examples of AV databases

Audio-visual databases	Format of recording scenario	Size	Audio annotation	Video annotation
CHIL	Lectures/meetings	~60 hours	orthographic transcripts, environmental events	2D & 3D head location
AusTalk	Read/ spontaneous speech	~3000 hours	word-level and orthographic transcripts	-
UWB-05-HSCAVC	Read speech	40 hours	phrase boundaries	mouth position and size
AVICAR	Read speech in car	- (86 speakers)	orthographically transcribed	-

Setup for audio-visual speech recording



- Video setup
  - JAI Pulnix RMC-6740.
  - Lenses:
    - Navitar NMV-25M23;
    - KOWA LM3NCM;
    - KOWA LM6NCM.
- Audio setup
  - M-audio QUAD
  - M-Audio EIGHT
  - Oktava – MK012

Independent Audio and Video Sensors

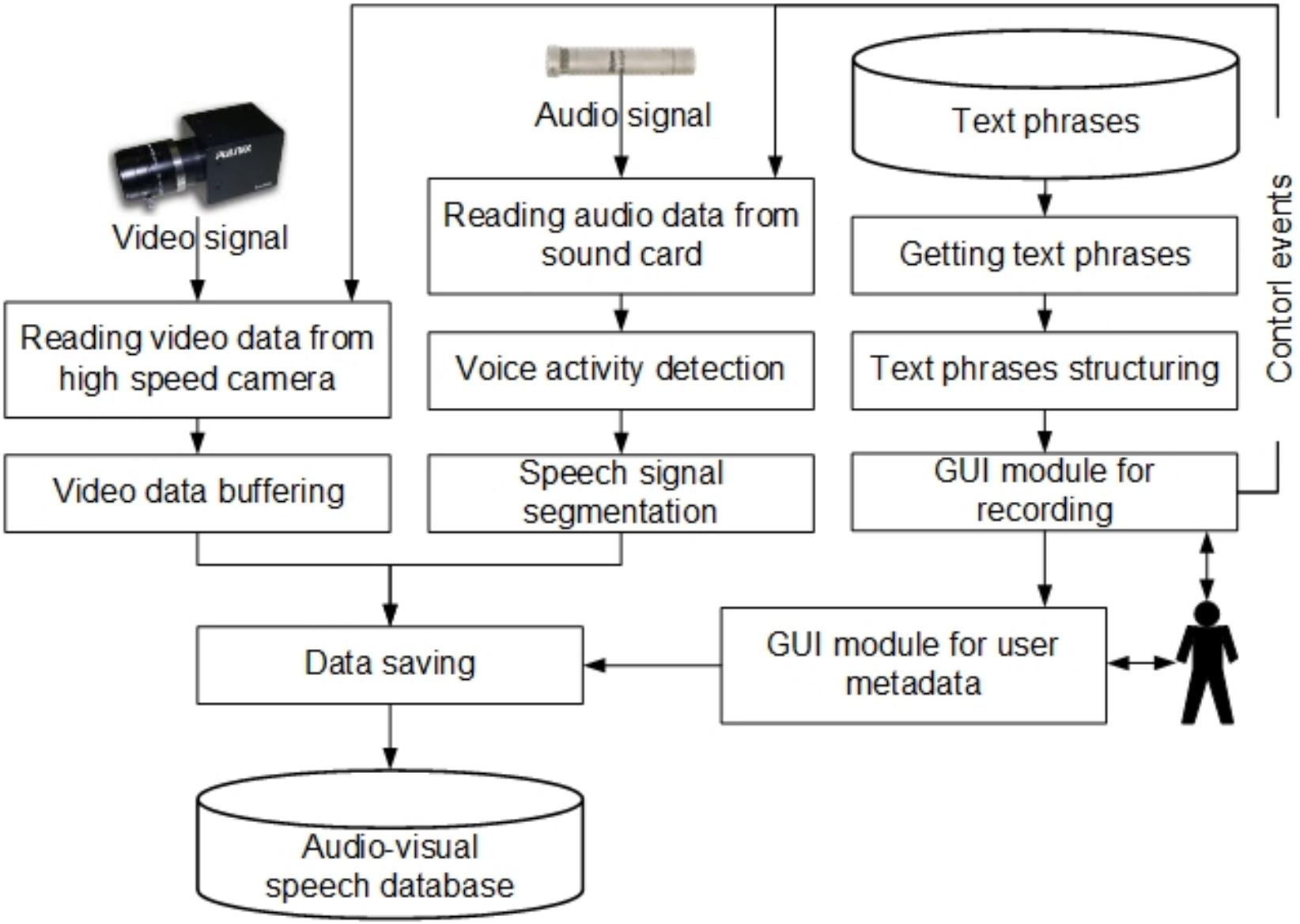
❖ **Oktava MK-012 condenser microphone:**

- medium-sized condenser microphone
- cardioid direction diagram (various capsules)
- captures acoustic sounds in range of 20-20kHz
- XLR interface with 48V phantom power

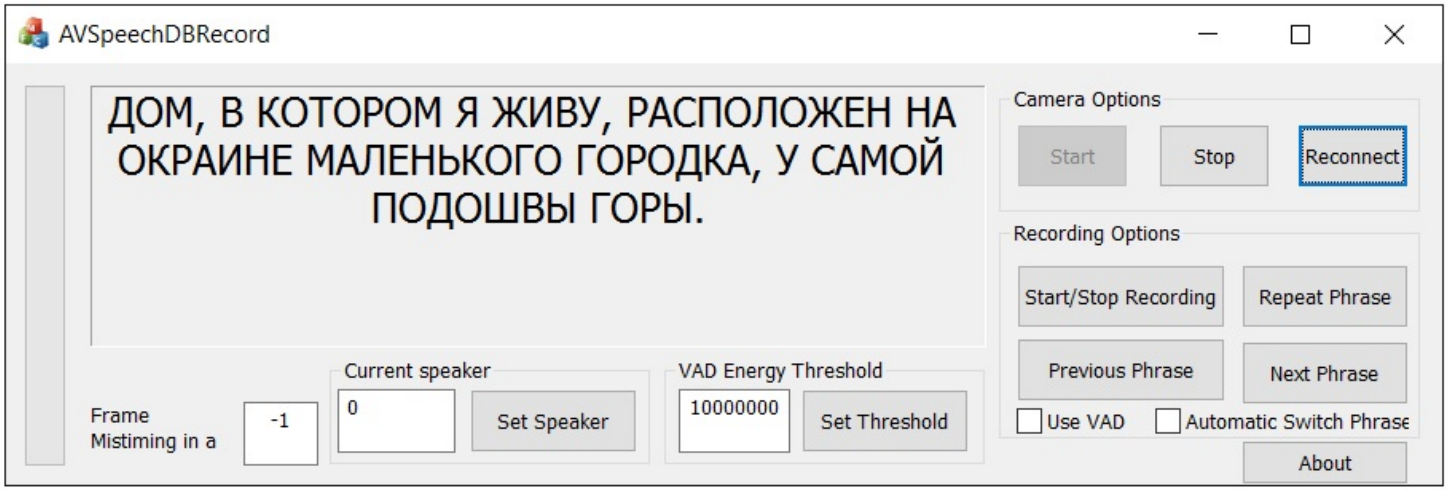
❖ **JAI Pulnix RMC-6740GE high-speed camera:**

- high-speed progressive scan camera
- full image resolution: 640x480 pixels at 200 fps
- partial scan mode of up to 3200 fps
- 24-bit color images
- 4:3 (3:4 if rotated) image format
- Gigabit Ethernet interface

The software for audio-visual speech database recording

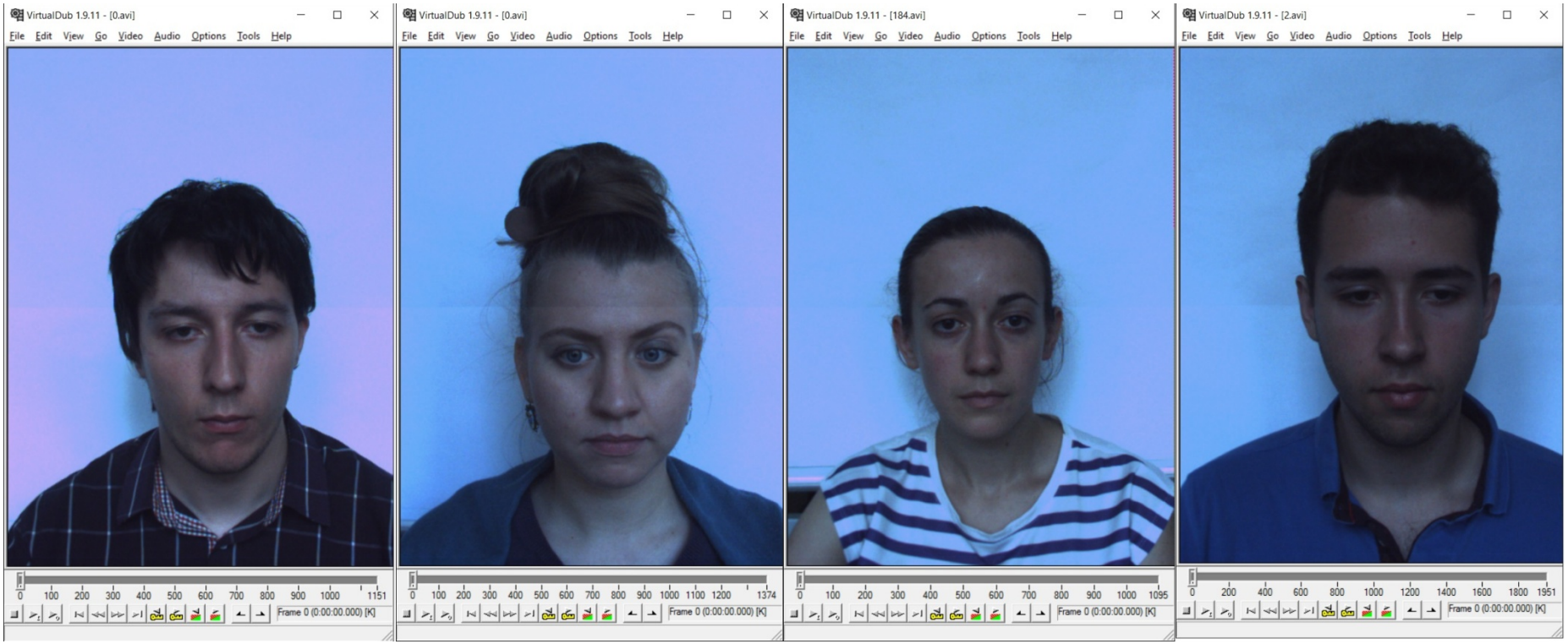


Example of the software GUI



- Video parameters:
  - Optical resolution - 640x460;
  - 200 frames per second;
  - Distortion from -0.2 to 0.46, based on installed lenses;
- Audio parameters:
  - 16 or 44kHz sampling rate;
  - SNR more than 30dB
  - PCM WAV format

Screenshots from HAVRUS



Conclusions

- The collected corpus HAVRUS comprises recordings of 20 native monolingual speakers of Russian with no language or hearing problems. Each speaker pronounced 200 Russian sentences. HAVRUS is meant for further research and experiments on audiovisual Russian speech recognition.
- This research is financially supported by the Ministry of Education and Science of the Russian Federation, agreement No 14.616.21.0056 (reference RFMEFI61615X0056), project "Research and development of audio-visual speech recognition system based on a microphone and a high-speed camera", as well as by the Czech Ministry of Education, Youth and Sports, project No LO1506.

Thank you!

- Speech and Multimodal Interfaces Laboratory
- Address: 39, 14 Line,

  St. Petersburg, Russia, 199178
- Phone/Fax : +7 (812) 3280421
- Web: www.hci.nw.ru