

AUTOMATIC TRANSCRIPTION AND SUBTITLING OF SLOVAK BROADCAST NEWS DATA

¹J. Juhár, ¹J. Staš, ¹D. Hládek, ¹M. Lojka, ¹P. Vizslay, ²M. Rusko, ²M. Trnka, ²S. Darjaa, ²R. Sabo, ²M. Ritomský

¹Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovakia

²Institute of Informatics, Slovak Academy of Science in Bratislava, Slovakia

¹jozef.juhar@tuke.sk, ²milan.rusko@savba.sk

Abstract

We have developed a pilot system for automatic transcription and subtitling of broadcast news for the Slovak language, which can further develop into a simple application for daily use by TV news editors, publishing on various news web portals or other users who create e.g. custom video blogs. The pilot system consists of three independent modules designed for offline and online speech transcription and time synchronization of speech with text transcription. The proposed technology allows to streamline or completely replace manual transcription, create fully automated transcriptions and can also be used to index audiovisual content in large television archives. We intend to protect the individual modules of the presented research activity in the near future by utility models.

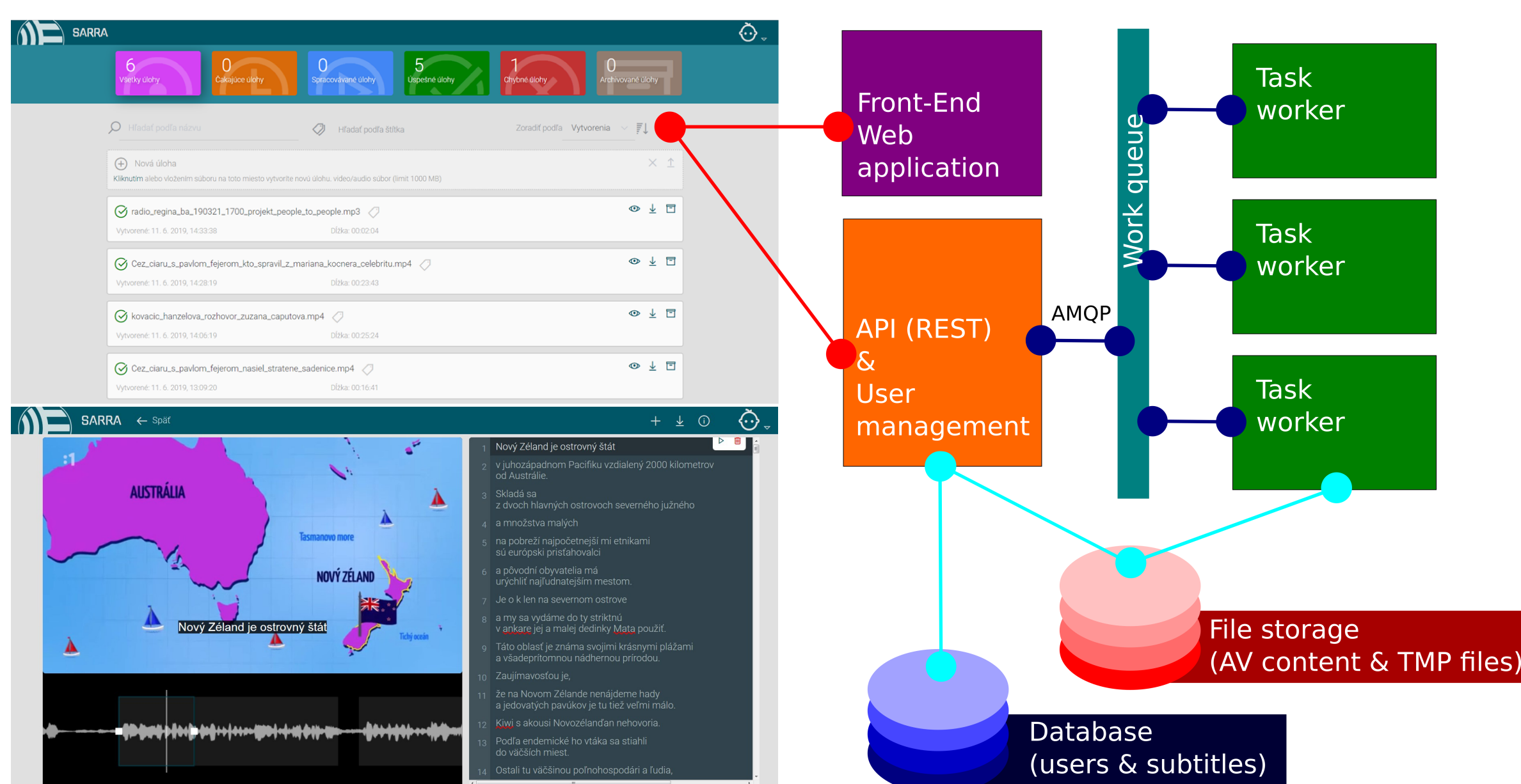


Fig. 1: Block diagram of the proposed offline subtitling system accessible through a web application

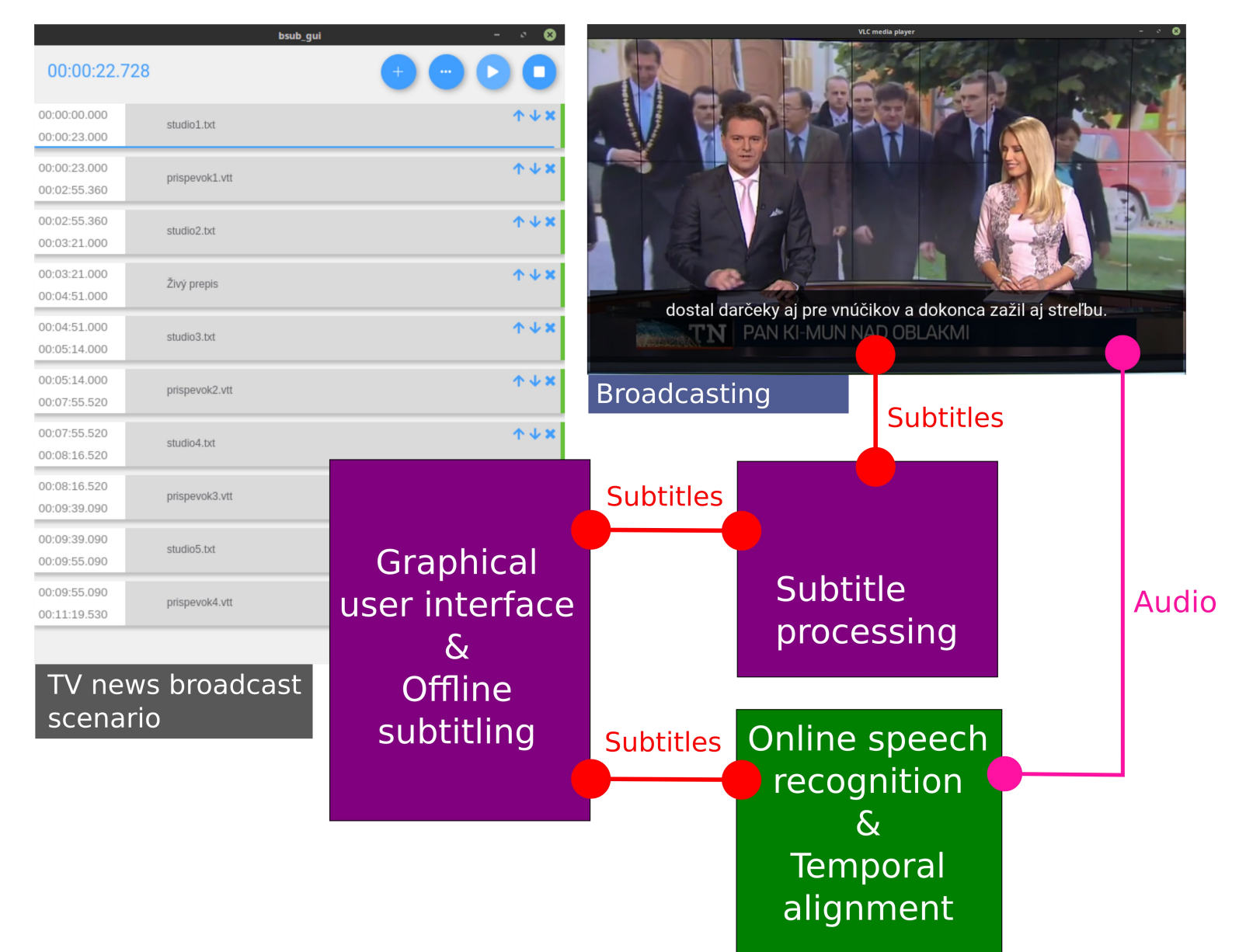


Fig. 2: Block diagram of interactive user interface for creation of TV news broadcasting scenario

Offline subtitling of audiovisual content

a. Task server

- interaction between user and task processing (workers) using private database and file storage [1]
- allows the user to create a new task, upload audiovisual content, query the status of the task and result of the processing
- assigns selected tasks to groups and then to the queue to process them

b. Task workers

- automatic speech segmentation based on principal component analysis
- speaker diarization based on Gaussian mixture models and i-Vectors
- acoustic models trained on almost 1,000 hours of manually annotated speech recordings [4]
- language models trained on 2,25 billion tokens/words
- advanced adaptation of acoustic and language models [3]
- spontaneous speech recognition based on deep learning (Kaldi) [6]
- the text post-processing and creation of subtitles

c. Web application

- for administrators provides access to user management
- for users provides access to individual processing tasks
- the interactive graphical user interface includes a simple subtitles editor and viewer (see Fig. 1)

Online subtitling of audiovisual content

- the input speech is recorded by the microphone
- the decoding is performed simultaneously with the incoming content [2]

Principle

- simultaneous automatic speech segmentation and spontaneous speech recognition with voice activity detection over a predefined period a time
- adaptation of acoustic models based on i-Vectors
- the text post-processing and creation of subtitles

Temporal alignment of subtitles with audiovisual content

- a principle similar to that of online subtitling
- the recognition network is limited by the content of the text in subtitles [5]
- reduced computational costs, improved accuracy, faster response

Principle

- recognition network is created from preprocessed subtitles [5]
- automatic speech segmentation with hypothesis tracking
- adaptation of acoustic models based on i-Vectors

User interface for creation of TV news broadcast scenario

- interactive graphical user interface (see Fig. 2)
- automatic switching of subtitles created using the previous three systems
- scenarios can be moved by dragging (the timing changes automatically)

References

- [1] Lojka, M., et al.: Slovak broadcast news speech recognition and transcription system, in Advanced in Network-Based Inform. Systems (NBIS 2018), LNDECT 22, Springer, 2019, 385–394
- [2] Staš, J., Lojka, M., Vizslay, P., Hládek, D., Juhár, J.: Automatické titulkovanie spravodajských relácií v slovenskom jazyku, in Data a znalosti & WIKT, Košice, Slovensko, 2019, 72–77
- [3] Staš, J., Hládek, D., and Juhár, J.: Dual-space re-ranking model for efficient document retrieval, user modeling and adaptation, in Proc. of ELMAR, Zadar, Croatia, 2018, 203–206
- [4] Vizslay, P., et al.: An extension of the Slovak broadcast news corpus based on semi-automatic annotation, in Proc. of LREC 2016, Portorož, Slovenia, 2016, 4684–4687
- [5] Stan, A., Bell, P., and King, S.: A grapheme-based method for automatic alignment of speech and text data, in Proc. of SLT 2012, Miami, Florida, USA, 2012, 286–290
- [6] Povey, D., et al.: The Kaldi speech recognition toolkit, in Proc. of Automatic Speech Recognition and Understanding, (ASRU), Waikoloa, Hawaii, USA, 2011, 1–4

Acknowledgement

The poster describes the results of R&D of speech technologies for the Slovak language supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the project VEGA 1/0511/17 and the Slovak Research and Development Agency under the project of applied research APVV-15-0517 and bilateral cooperation project SK-TW-2017-0005.