SPEECH PROCESSING RESEARCH GROUP

Contact details

Name	**Speech Processing Research Group**	
Acronym	**SPEECH**	
Logo		
Site	http://speech.utcluj.ro	
Address	26 Baritiu Str., 400027 Cluj-Napoca, Romania	
Faculty Department	**Faculty of Electronics, Telecommunications and Information Technology, Telecommunications Department.**	
Telephone	+40 264 202452	
Fax	+40 264 591689	
Director	Prof. Dr. Eng. Mircea.Giurgiu	
e-mail	Mircea.Giurgiu@com.utcluj.ro	

Areas of expertise

Speech Processing:
- Automatic Speech Recognition (ASR): Deep Neural Networks (DNN) architectures (DeepSpeech, Transformer);
- Text to Speech Synthesis (TTS): systems based on Tacotron2, DCTTS, and FastSpeech DNN architectures;
- Speaker diarization, Emotion and speaking style automatic recognition;
- Speaker anonymization to ensure privacy and security;
- Neural network-based speech vocoders;
- Voice assistants using conversational AI tools.

Text Processing:
- Sentiment analysis using dimensional and categorical models;
- Automatic question and answer systems in natural language using Deep Neural Networks;
- Natural Language Processing using machine learning techniques;

Team

Prof. Dr. Eng. Mircea Giurgiu, Conf.dr.ing. Adriana Stan, drd.ing. Alexandra Drobut, drd.ing. Mihai Ciobanca. drd.ing. Gal Oscar. External collaborators: Prof. Jozsef Domokos (Univ. Sapientia), dr.ing. Alin Cordos (PixelData).

Representative projects

"**TDIH** – Transylvania Digital Innovation Hub", 2022-2025, H2020 Digital-2021-EDIH-01, Nr. 101083508, Coordinator: Transylvania IT Cluster, RO, https://transilvaniadih.ro/, WP2 (Test before Invest) leader.
ReTeRom – "Resources and technologies for developing human-machine interfaces in Romanian", PCCDI 2018 – 2020, http://speech.utcluj.ro/sintero
SWARA – "Mobile System for Rehabilitative Vocal Assistance of Surgical Aphonia" PN-II-PCCA, 2014-2017, http://speech.utcluj.ro/swara
Simple4All – "Speech synthesis that improves through adaptive learning" (EC-FP7, 2011-2014), http://simple4all.org
Sound2Sense – "Making sense of speech sounds" (EC-FP6, 2007-2011), http://www.sound2sense.eu
Text2Speech – "Development of software services for text to speech synthesis in Romanian language" (PN II INOVARE, 2008-2010);.
KeyToNature – (EC - eContent Plus, 2008-2010), http://www.key2nature.eu
EUROWEX – "Online platform using digital signature for the management of university activities" (EC – eTEN Trans European e-Services in the Public Interest, 2006-2008), http://www.eurowex.org
Pool2Business – "Project Organisation Online" (EC–EACEA-LLP, 2008-2010), http://www.pool2business.eu/

Significant results

The most representative publications of the past 5 years:

1. A. Stan, B. Lorincz, M. Nutu, M. Giurgiu, "The MARA Corpus: Expressivity in End-to-end TTS Systems using Synthesised Speech Data", The 11th Conf. on SPED 2021, Bucharest, 13-15 Oct. 2021.
2. B. Lorincz, A. Stan, M. Giurgiu, "An objective evaluation of the effects of recording conditions and speaker characteristics in multi-speaker deep neural speech synthesis", Procedia Computer Sciences, Vol. 192, pp. 756-765, 2021, Elsevier.
3. B. Lorincz, A. Stan, M. Giurgiu, "Speaker verification-derived loss and data augmentation for DNN-based multispeaker speech synthesis", Proc of EUSIPCO 2021.
4. Beata Lorincz, Maria Nutu, Adriana Stan, "Romanian Part of Speech Tagging using LSTM Networks", In Proceedings of the IEEE 15th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, 2019.
5. Maria Nutu, Beata Lorincz, Adriana Stan,"Deep Learning for Automatic Diacritics Restoration in Romanian", In Proceedings of the IEEE 15th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, 2019.
6. David A. Braude, Matthew P. Aylett, Caoimhin Laoide-Kemp, Simone Ashby, Kristen M. Scott, Brian O Raghallaigh, Anna Braudo, Alex Brouwer, Adriana Stan,"All Together Now: The Living Audio Dataset", In Proceedings of Interspeech, Graz, Austria, 2019
7. Adriana Stan, "Input Encoding for Sequence-to-Sequence Learning of Romanian Grapheme-to-Phoneme Conversion", In Proceedings of the 10th IEEE International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Timisoara, Romania, 2019.
8. Stan, Adriana; Dinescu, Florina; Tiple, Cristina; et al., The SWARA Speech Corpus: A Large Parallel Romanian Read Speech Dataset International Conference on Speech Technology and Human-Computer Dialogue (SpeD), published 2017
9. Stan, Adriana; Valentini-Botinhao, Cassia; Orza, Bogdan; et al., BLIND SPEECH SEGMENTATION USING SPECTROGRAM IMAGE-BASED FEATURES AND MEL CEPSTRAL COEFFICIENTS 2016 IEEE WORKSHOP ON SPOKEN LANGUAGE TECHNOLOGY (SLT 2016) Pages: 597-602 Published: 2016
10. Moldovan, Alexandru; Stan, Adriana; Giurgiu, Mircea, Improving Sentence-level Alignment of Speech with Imperfect Transcripts using Utterance Concatenation and VAD 2016 IEEE 12TH INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTER COMMUNICATION AND PROCESSING (ICCP)   Book Series: IEEE International Conference on Intelligent Computer Communication and Processing ICCP Pages: 171-174 Published: 2016
11. Adriana Stan, Yoshitaka Mamiya, Junichi Yamagishi, Peter Bell, Oliver Watts, Rob Clark, Simon King, "ALISA: "An automatic lightly supervised speech segmentation and alignment tool", In *Computer Speech and Language*, vol. 35, pp. 116-133, 2016

Significant solutions:

Voice cloning in TTS using small amount of speech data, Automatic alignment of speech and text data, Improve the speech synthesis by improved speaker similarity, Accent prediction in text using only speech data, Text processing using Finite State Transducers, Statistical language modelling for speech recognition and text to speech synthesis, Blind speech denoising and dereverberation, Automatic speech segmentation at syllable level, Unsupervised and language independent syllabification using statistical methods, Broadcast news speaker diarization and speech music discrimination, Emotion and speaking style recognition from audiobook data; Sentiment polarity prediction using categorigal and dimensional models, Polarity prediction using Vector Space Models,

Products and technologies:

1. RONNA – ROmanian Neural Network Api for Speech Synthesis (http://speech.utcluj.ro/ronna)
2. ALISA - A lightly supervised speech segmentation and alignment tool;
3. TUNDRA - A corpus of 14 European languages collected from audiobooks;
4. NORMA - Statistical machine translation-based text NORMAlization tool;
5. DEXTER - Speaker recognition and diarization in audio-video talk shows;
6. AUDIOOR - AUDIO Online Repository, a web based repository of audio and text resources;
7. VoiA – Voice Assistance tool using the conversational AI tool Nvidia NeMo;
8. SENTIMENT - Sentence polarity predictor for SENTIMENT analysis;
9. TextPREDICT - Fast Text input PREDICTion on mobile devices.

The offer addressed to the economic environment

Research & development	Text to speech synthesis integrated in specific solutions for telecommunications, Automatic speech recognition and assistive technologies for human computer interface, Interactive Voice Response Systems, Online multimedia repositories using intelligent indexing and content searching.
Consulting	Multimedia technologies, data modelling, data mining, advanced methods for signal processing, eLearning solutions, project management, data security.
Training	Speech Processing, Statistical methods for data processing, Microprocessor-based systems.

Last update on January 2024