

## SPEECH PROCESSING RESEARCH GROUP

### Contact details

Name	<b>Speech Processing Research Group</b>
Acronym	<b>SPEECH</b>
Logo	
Site	<a href="http://speech.utcluj.ro">http://speech.utcluj.ro</a>
Address	26 Baritiu Str., 400027 Cluj-Napoca, Romania
Faculty Department	<b>Faculty of Electronics, Telecommunications and Information Technology, Telecommunications Department.</b>
Telephone	+40 264 202452
Fax	+40 264 591689
Director	Prof. dr. eng. Mircea.Giurgiu
e-mail	<a href="mailto:Mircea.Giurgiu@com.utcluj.ro">Mircea.Giurgiu@com.utcluj.ro</a>



### 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;
- Audio Deep Fake detection and countermeasures to combat disinformation
- 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.

#### Natural Language Processing:

- Sentiment analysis using dimensional and categorical models;
- Automatic question and answer systems in natural language using LLMs (BERT-based architectures);
- Natural Language Processing using machine learning techniques;

### Team

**Prof. dr. eng. Mircea Giurgiu**, drd.ing. Gal Oscar, drd.ing. Cărbunescu Maeva, drd.ing. Istrate Ciprian. MSc. Axente Andrei, MSc. Popescu Valentin, MSc. Melinte Cosmin, MSc. Zinveliu Ioana, MSc. Dolca Florica, MSc. Bucur Andrei. External collaborators: Prof. Jozsef Domokos (Univ. Sapiientia), dr.ing. Alin Cordos (PixelData).

### Representative projects

**HRIA** – “Romanian Hub for Artificial Intelligence”, 2025-2030, SMIS 2021-351416, <https://hria.upb.ro/>  
**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 – 2021, <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. Axente and M. Giurgiu, "Audio Fakes in the Noise: Hybrid DNN-Based Detection with SE-U-Net, ResNet-18, and Wav2Vec-XLSR Embeddings," Proc. of 33rd Telecommunications Forum, 25-26 Nov 2025, Belgrade, Serbia, doi: 10.1109/TELFOR67910.2025.11314280.
2. V.-D. Popescu and M. Giurgiu, "Optimizing Speech Emotion Recognition with Attention-Driven Deep CRNN," Proc of 33rd Telecommunications Forum, 25-26 Nov 2025, Belgrade, Serbia, doi: 10.1109/TELFOR67910.2025.11314307.
3. O. Gal, M. Giurgiu, "Empowering Accessibility: Application for Individuals with Visual Impairment," in Lecture Notes in Networks and Systems, 2025, vol 1660. Springer, Cham. [https://doi.org/10.1007/978-3-032-07109-5\\_6](https://doi.org/10.1007/978-3-032-07109-5_6).
4. I. Zinveliu, F. Dolca, M. Giurgiu, "AudioDepfake Detection using MFCC Features and CNNs", Proc. of SYNASC-WSAI, 23-25 Sept 2025, Timisoara,
5. O. Gal, M. Giurgiu, "Objective Evaluation of Romanian TTS Systems: Glow-TTS and VITS", Proc. of SYNASC-WSAI, 23-25 Sept 2025, Timisoara,
6. D. Melinte, M. Giurgiu, "Speaker Anonymization via GAN-driven Artificial Embeddings and Prosody Transfer", Proc. of IEEE 21st ICCP Conference, 16-18 Oct 2025, Cluj-Napoca,
7. M. Carbunescu, M. Giurgiu, "Gaze TRacking via L2CS-Net Pipeline with Ridge Regression and Kalman POst-filtering", Proc. of IEEE 21st ICCP Conference, 16-18 Oct 2025, Cluj-Napoca,
8. 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.
9. 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.

### Significant solutions:

Automatic Speech Recognition for speech to text transcription of medical reports; HiFi text to speech synthesis and 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 categorial and dimensional models; Polarity prediction using Vector Space Models; IoT solutions in Cloud.

### Products and technologies:

1. S2T-MEDICAL – Automatic Speech to Text transcription of medical reports;
2. HiFi-GAN-TTS – High Fidelity Generative Adversarial Network for Text To Speech synthesis;
3. Q&A-BERT – Question and Answer system in Romanian using BERT architecture of Large Language Models.
4. VoiceDrive – Voiced based control to support autonomous Driving
5. GAZE2SPEECH - Automatic conversion tool from gaze tracking to speech;
6. TUNDRA - A corpus of 14 European languages collected from audiobooks;
7. NORMA - Statistical machine translation-based text NORMALization tool;
8. DEXTER - Speaker recognition and diarization in audio-video talk shows;
9. AUDIOOR - AUDIO Online Repository, a web based repository of audio and text resources;
10. VoiA – Voice Assistance tool using the conversational AI tool Nvidia NeMo;
11. SENTIMENT - Sentence polarity predictor for SENTIMENT analysis;
12. TextPREDICT - Fast Text input PREDICTion on mobile devices.
13. „DeepFake-Hybrid” - Software for audiodepfake detection in noisy environments
14. „ANONIMIZE” – Software tool for speaker anonymization in Romanian language

## 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 2026