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): HMM-based speech recognition, Deep Neural Networks recognition;
- Text to Speech Synthesis (TTS): Voice Cloning, Speaking Style Transplantation;
- Speaker diarization, Emotion and speaking style automatic recognition;
- Online audio digital repositories: Web based interfaces, Metadata descriptors;
- Speech Coding at low bit rate for applications in telecommunications: RPE-LTP, CELP, MPEG, SPEEX;
- Secure authentication using voice, Interactive Voice Response Systems.

Text Processing:
- Sentiment analysis using dimensional and categorical models;
- Unsupervised sentence polarity prediction;
- Natural Language Processing using machine learning techniques;

Team

Prof. Dr. Eng. Mircea Giurgiu, Dr. Eng. Adriana Stan, Drd. Eng. Victor Rachita, Drd. Eng. Andrei Homodi.
External collaborators: University of Tirgu-Mures: Dr. Eng. Jozsef Domokos; SC. Fortech srl: Eng. Remus Pop, Dr.eng. Reka Hints; SC. Cloud Troopers: Eng. Ciprian Costea.

Representative projects

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. 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
2. Adriana Stan, Cassia Valentini-Botinhao, Mircea Giurgiu, Simon King, "Phonetic Segmentation of Speech using STEP and t-SNE", In *Proc. of the 8th International Conference on Speech Technology and Human-Computer Dialogue (SpeD),* Bucuresti, Romania, 2015.
3. Jószef Domokos, Adriana Stan, Mircea Giurgiu, "An Approach to Lexical Stress Detection from Transcribed Continuous Speech Using Acoustic Features", In *Proc. 22nd Telecommunications Forum,* Belgrade, Serbia, 2014.
4. O. Watts, S. Gangireddy, J. Yamagishi, S. King, S. Renals, A. Stan, M. Giurgiu, "Neural Net Word Representations for Phrase-Break Prediction Without a Part of Speech Tagger", *In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* Florence, Italy, 2014.
5. R. San-Segundo, J.M. Montero, M. Giurgiu, I. Muresan, S. King, “Multilingual Number Transcription for Text-to-Speech Conversion”, *In 8th ISCA Workshop on Speech Synthesis,* pages 85-89, Barcelona, Spain, August 2013.
6. A. Stan, O. Watts, Y. Mamiya, M. Giurgiu, R. A. J. Clark, J. Yamagishi, S. King, "TUNDRA: A Multilingual Corpus of Found Data for TTS Research Created with Light Supervision", *In Proc. Interspeech*, Lyon, 2013.
7. O. Watts, A. Stan, R. Clark, Y. Mamiya, M. Giurgiu, J. Yamagishi, S. King, "Unsupervised and lightly-supervised learning for rapid construction of TTS systems in multiple languages from ‘found’ data: evaluation and analysis*", In Proc. SSW8,* Barcelona, 2013.
8. A. Stan, O. Watts, Y. Mamiya, M. Giurgiu, R. A. J. Clark, J. Yamagishi, S. King, "TUNDRA: A Multilingual Corpus of Found Data for TTS Research Created with Light Supervision", *In Proc. Interspeech*, Lyon, 2013.
9. M. Giurgiu, A. Kabir, “Automatic Transcription and Speech Recognition of Romanian Corpus RO-GRID”, *Proceedings of 35th International Conference on Telecommunications and Signal Processing (TSP 2012),* ISBN: 978-1-4673-1118-2, 3-4 July, Prague, 2012, pp. 465-468.
10. A. Kabir, M. Giurgiu - "Modelling Human Speech Perception in Noise", - *International Journal of Circuits, Systems and Signal Processing, ISSN: 1998-4464, Issue 4, Volume 5,* 2011, pp. 415-422.
11. M. Giurgiu, A. Kabir, - “Improving Automatic Speech Recognition in Noise by Energy Normalization and Signal Resynthesis”, *Proceedings of 2011 IEEE International Conference on Intelligent Computer Communications and Processing (ICCP2011),* 25-27 August, Cluj-Napoca, ISBN: 978-1-4577-1479-5, pp. 311-314.
12. Stan, A., Pop, F.-C., Cremene, M., Giurgiu, M., and Pallez, D. Interactive Intonation Optimisation Using CMA-ES and DCT Parametrisation of the F0 Contour for Speech Synthesis. In Proceedings of the 5th Workshop on Nature Inspired Cooperative Strategies for Optimisation, volume 387 of *Studies in Computational Intelligence*. Springer.
13. Stan, A., Yamagishi, J., King, S., and Aylett, M. The Romanian speech synthesis (RSS) corpus: Building a high quality HMM-based speech synthesis system using a high sampling rate. *Speech Communication*, 53(3):442-450.

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 modeling 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. ALISA - A lightly supervised speech segmentation and alignment tool;
2. TUNDRA - A corpus of 14 European languages collected from audiobooks;
3. NORMA - Statistical machine translation-based text NORMAlization tool;
4. DEXTER - Speaker recognition and diarization in audio-video talk shows;
5. AUDIOOR - AUDIO Online Repository, a web based repository of audio and text resources;
6. DeREVERB - Blind speech enhancement and DeREVERBeration tool;
7. SENTIMENT - Sentence polarity predictor for SENTIMENT analysis;
8. TextPREDICT - Fast Text input PREDICTionon 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.