UNIVERSITY OF ŽILINA
Faculty of Electrical Engineering and Information Technology

Topic of dissertation thesis

Academic year *2026/2027*

Title (Engl./Slovak)	Compact and Generalizable Neural Models for Speech and Audio Processing / Kompaktné a generalizovateľné modely neurónových sietí pre spracovanie reči a audia
Institute	Faculty of Electrical Engineering and Information Technology University of Žilina
Place	Žilina, Slovakia
PhD. programme	telecommunications
Supervisor	doc. Ing. Roman Jarina, PhD Department of multimedia and information-communication technology
Co-supervisor	Ing. Maroš Jakubec, PhD.
Study form	Internal
Study duration	3 years (internal form)
Study language	Slovak, English
Start date	1.9.2026
Research domain	Speech processing, Speaker recognition, Machine learning, audio
Contact person	Phone: +421 41 513 2213 E-mail: roman.jarina@uniza.sk Web-page https://kmikt.uniza.sk/

Dissertation topic abstract

While large deep neural networks (DNN) have achieved remarkable performance across various speech tasks, including emotion recognition, speaker characterization, and audio scene analysis, they are often computationally expensive and require extensive labelled data.

The PhD study aims to investigate strategies such as knowledge distillation and zero-shot learning to enable lightweight, generalizable models that can operate efficiently in resource-constrained environments and adapt to new tasks without task-specific training data. By combining modern representation learning approaches with these techniques, the research seeks to improve the accessibility, scalability, and applicability of deep learning models for speech and audio processing.

Extended information, research responsibilities and tasks of PhD. candidate

Modern deep learning architectures have transformed speech and audio processing, enabling breakthroughs in emotion recognition, speaker identification, and audio understanding. However, state-of-the-art models are typically extremely large, computationally intensive, and require extensive labelled training datasets. This project addresses these challenges by developing efficient and generalizable deep learning frameworks using techniques such as knowledge distillation, model compression, and zero-shot learning.

The PhD candidate will:
- design and implement DNN architectures for speech and audio tasks,
- explore knowledge distillation techniques to transfer knowledge from large “teacher” models to lightweight “student” models,
- investigate zero-shot and cross-task learning methods to enable generalization to unseen audio domains,
- evaluate developed methods on benchmark datasets across multiple speech/audio tasks.

The project provides access to large-scale speech and audio datasets, GPU infrastructure, and opportunities for collaboration, publications, and applied research activities.

The PhD candidate is expected to publish results in leading conferences and journals in AI, speech, and audio processing, actively participate in research collaborations, and contribute to teaching and dissemination activities.

Candidate profile

Education and Professional Background

- Master's degree (MSc, MEng, or equivalent) in Telecommunications, Computer Science, or a closely related technical field.

Technical Skills

- Intermediate to advanced programming skills in Python, with experience in relevant libraries (e.g., PyTorch, TensorFlow, scikit-learn).
- Experience with data analysis, and machine learning methods, familiarity with deep learning techniques, model training, and evaluation.

Research and Personal Competences

- Strong interest in artificial intelligence, and data-driven methods,
- Interest in working with speech and audio data.
- Strong analytical skills and solid mathematical foundation,
- Ability to communicate scientific results effectively (e.g., through publications, conference presentations, or patents),
- Proficiency in written and spoken English.

The dissertation topic is supported by the projects:

- “Understanding speaker voice changes due to aging for robust speaker voice representation” APVV-MVP-24-038, funded by the Slovak Research and Development Agency, and
- "Speech signal processing for speaker authentication and emotion recognition using deep learning", 1/0684/25, funded by the Slovak Grant Agency VEGA.