



Topic of dissertation thesis

Academic year 2026/2027

Title (Engl./Slovak)	Interpretable Speech Signal Modelling for Speaker Characterization /Interpretovateľné modelovanie rečového signálu na charakterizáciu hovoriaceho		
Institute	Faculty of Electrical Engineering and Information Technology		
Place	University of Žilina		
PhD. programme	telecommunications		
Supervisor	doc. Ing. Roman Jarina, PhD		
Co-supervisor	-		
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:	E-mail:	Web-page
	+421 41 513 2213	roman.jarina@uniza.sk	https://kmikt.uniza.sk/

Dissertation topic abstract

This PhD study focuses on interpretable and explainable feature extraction and analysis, with emphasis on speaker representations and/or emotion modelling. The research aims to develop transparent modelling approaches that explicitly link acoustic and prosodic features to tasks such as voice biometrics and the estimation of a speaker's emotional or physiological state.

Although deep learning has significantly advanced speech processing, modern speaker characterization systems often lack transparency. By integrating representation learning with explainability techniques, the project seeks to better understand how non-linguistic information is encoded in speech while improving model robustness and reliability.

The expected outcomes include novel interpretable feature extraction frameworks and deeper insights into speaker and affect representation, contributing to more transparent and trustworthy speech-based AI systems.

Extended information, research responsibilities and tasks of PhD. candidate

Although deep neural networks (DNN) have significantly improved speaker characterization performance, they often operate as black-box models, and their decision-making processes remain difficult to interpret. Interpretable modelling refers to approaches in which model structures and representations are inherently understandable and directly linked to meaningful acoustic or prosodic properties, while explainable approaches provide insight into complex models by showing which features influenced a prediction and why a particular decision was made.

The PhD candidate will conduct research in the area of interpretable and explainable modelling of speech signals for speaker characterization. The work will focus on the development of transparent feature extraction and modelling approaches that explicitly relate acoustic and prosodic characteristics of speech to speaker identity, biometric traits, and emotional or physiological states.

The candidate will:

- design and implement interpretable feature extraction frameworks for speech analysis, analyse how non-linguistic information is encoded in learned representations,
- develop and evaluate models for voice biometrics and/or emotion/physiological state recognition,
- validate proposed methods on benchmark datasets and real-world scenarios.

The project provides access to speech datasets, experimental infrastructure, and opportunities for collaboration, scientific publications, and applied research activities.

The PhD candidate is expected to publish results in international conferences and journals in speech and multimedia processing and artificial intelligence, 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

- Basic to intermediate programming skills in Python,
- Experience with data analysis, signal processing, and machine learning methods,
- Interest in working with speech and audio data.

Research and Personal Competences

- Strong interest in speech/audio technologies, artificial intelligence, and audio data-driven methods,
- 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.