

## **Vaizdo, garso, natūralios kalbos apdorojimo metodų bei didžiaisiais ir vizualiais kalbos modeliais grįstų sprendimų kūrimas ir tyrimas**

Dr. Jolita Bernatavičienė, prof. dr. Gražina Korvel, prof. dr. Virginijus Marcinkevičius, prof. dr. Darius Plikynas, doc. dr. Gintautas Tamulevičius, prof. dr. Povilas Treigys

Nuoseklus ir visaapimantis garso ir/ar vaizdo signalų apdorojimas turėtų būti atliekamas keliais aspektais. Pirmiausia apdorojamas fizinis signalas. Tuo tikslu signalas skaitmenizuojamas, o jam apdoroti taikomi įvairūs formalizuoti diskretiniai metodai, leidžiantys analizuoti sudedamąsias dalis, išskirti charakteringus duomenis. Signalams generuoti, koduoti bei aprašyti naudojami įvairūs diskretieji modeliai. Taikydami šiuos metodus ir modelius mes galime identifikuoti asmenį, atpažinti kalbos arba vaizdo turinį, įvertinti asmens emocinę būseną, atlikti ankstyvąją balso ir /ar vaizdo kokybės vertinimą, kalbos analizėje – transformuoti žmogaus balsą, atstatyti triukšmingų įrašų kokybę ar kitaip apdoroti signalą. Toliau atliekamas aukštesnio lygio signalo apdorojimas, taikant dirbtinio intelekto metodus, skirtus automatinei kalbos, garso ir vaizdo analizei ir generavimui. Pavyzdžiui, natūralios kalbos apdorojimo procesas apima gramatinę kalbos nagrinėjimą, sintaksinę analizę, semantinę interpretaciją bei kalbos pragmatiką. Pagrindinės teksto apdorojimo užduotys yra teksto santraukos sudarymas, informacijos paieška, skirtingų kalbų teksto vertimas, atsakymai į klausimus, temų modeliavimas ir teksto klasifikavimas. Šnekos apdorojimo atveju – sintezavimas ir atpažinimas, vaizdų analizėje – scenos segmentavimo, aprašymo, objektų atpažinimo, sekimo, klasifikavimo ar kitos užduotys pasitelkiant didžiuosius ir vizualiuosius kalbos modelius ar kitus pažangiausius metodus.

## **Developing and researching image, sound, natural language processing techniques and solutions based on large and visual language models**

Consistent and comprehensive audio and image signal processing should be carried out in several aspects. First, the physical signal is processed. To this end, the signal shall be digitised, and various formalised discrete methods shall be applied to process it, allowing for the analysis of the components and the extraction of characteristic data. Various discrete models are used to generate, encode and describe the signals. These methods and models allow us to identify a person, recognise the content of speech or video, assess a person's emotional state, perform an early assessment of voice and/or video quality, or, in speech analysis, transform the human voice, restore the quality of noisy recordings, or otherwise process the signal. Secondly, the higher-level signal processing follows after the physical signal is processed. Artificial intelligence techniques are typically used for automatic speech, audio, and video analysis and generation. For example, natural language processing includes grammatical analysis, syntactic analysis, semantic interpretation and language pragmatics. The main text-processing tasks are summarising text, information retrieval, translating cross-lingual text, answering questions, modelling topics and text classification. In speech processing: synthesis and recognition; image analysis: scene segmentation, description, object recognition, tracking, classification, and other tasks, use large and visual language models or other state-of-the-art methods.