Over the recent years we have seen an emergence a various voice-based applications, devices, and services in which human speech is recognized automatically by a machine. Voice-based human-machine interfaces are the core of the underlying technology for smart speakers, digital home assistants, voice bots, as well as a breadth of voice-controlled devices including humanoid robots. Although automatic speech recognition (ASR) has been an active research field for several decades, a real irruption of new research opportunities but also challenges has been associated with the emergence of deep learning and artificial intelligence. The performance of recent deep learning based ASR systems for speech of a single speaker in good acoustic conditions is reaching nearly that of humans. However, when speech is recorded in real acoustic environments with interfering sound sources and background noise, especially when more speakers are active at the same time, its performance typically dramatically drops, often even disabling the technology from use in such difficult acoustic conditions.

In this project the project team will aim to tackle the problem of robust automatic speech recognition in difficult acoustically challenging conditions with multiple speakers, other interfering sounds and background noise. To this end, we will investigate deep extraction methods that provide the most suitable feature representations from the microphone signals that can be effectively used for speech recognition purposes. We will study different kinds of deep learning models, ASR models based on deep neural networks (DNNs) and hidden Markov model (HMM) as well as novel end-to-end approaches in which acoustic, lexicon and language models are all folded into a single neural network that is jointly optimized. We will also incorporate additional speaker and localization information in order to achieve improved speech recognition results from single- and multichannel recordings of speech in acoustically difficult conditions. The project results will lead to more robust ASR DNN based models that are tailored to more difficult but also more natural application scenarios. The project is also aimed at advancing and creating ASR models for Polish language that can be integrated and used by researchers from various disciplines. A good example of a possible application includes diagnostics and monitoring of several kinds of disease by a semantic analysis of human speech.