

W ostatnich latach pojawiło się mnóstwo aplikacji, urządzeń i usług głosowych, których podstawą jest automatyczne rozpoznawanie mowy. Interfejsy głosowe człowieka z maszyną stanowią nieodzowną technologię urządzeń takich jak inteligentne głośniki, asystenci głosowi, boty głosowe, jak również urządzenia kontrolowane głosem takie jak roboty. Choć temat automatycznego rozpoznawania mowy (ang. automatic speech recognition - ASR) jest tematem badań naukowców od kilku dekad, wiele nowych możliwości ale i wyzwań pojawiło się wraz z gwałtownym rozwojem głębokiego uczenia oraz sztucznej inteligencji. Wydajność najnowszych systemów ASR opartych na głębokim uczeniu dla mowy pochodzącej od jednego mówcy w dobrych warunkach akustycznych jest bliska wydajności ludzkiej. Jednakże w momencie gdy mowa jest nagrywana w rzeczywistych warunkach akustycznych z sygnałami zakłócającymi, innymi źródłami dźwięku i hałasem otoczenia, zwłaszcza jeśli w tym samym czasie aktywnych jest więcej mówców, jej wydajność zazwyczaj dramatycznie spada, nawet uniemożliwiając korzystanie z tej technologii w tak trudnych warunkach akustycznych.

W ramach projektu zespół badawczy skupi się na problemie automatycznego rozpoznawania mowy w trudnych warunkach akustycznych z wieloma mówcami, innymi dźwiękami zakłócającymi i hałasem otoczenia. W tym celu zbadane zostaną metody głębokiej ekstrakcji, które zapewnią najbardziej odpowiednią reprezentację cech związanych z mową z sygnałów mikrofonowych, które można skutecznie wykorzystać do celów rozpoznawania mowy. Celem badań jest sprawdzenie skuteczności różnych rodzajów modeli głębokiego uczenia, zarówno modeli ASR opartych na głębokich sieciach neuronowych (DNN) oraz ukrytych modelach Markowa (HMM), ale również nowatorskie podejście end-to-end, w którym modele akustyczny, leksykonowy i językowy są złożone w jedną całość - pojedynczą sieć neuronową - która jest wspólnie optymalizowana. Wprowadzimy również dodatkowe informacje o mówcy i jego lokalizacji w celu poprawy wyników rozpoznawania mowy z jedno- i wielokanałowych nagrań mowy w trudnych warunkach akustycznych. Wyniki projektu doprowadzą do stworzenia bardziej odpornych modeli opartych na ASR DNN, które będą dostosowane do trudniejszych, ale także bardziej naturalnych scenariuszy zastosowań. Projekt ma również na celu rozwój i tworzenie modeli ASR dla języka polskiego, które mogą być integrowane i wykorzystywane przez pracowników naukowych z różnych dyscyplin. Dobrym przykładem możliwego zastosowania jest diagnostyka i monitorowanie kilku rodzajów chorób poprzez semantyczną analizę mowy ludzkiej.