

Vedoucí týmu: Doc. Dr. Ing. Jan Černocký

Označení týmu: Skupina zpracování řeči na Fakultě informačních technologií VUT v Brně – BUT Speech@FIT

## 1. Obsah výzkumu – state-of-the art

BUT Speech@FIT byla založena v roce 1997 a od roku 2002 pracuje na Fakultě informačních technologií VUT v Brně. Poradcem skupiny je Prof. Hynek Heřmanský, výkonným ředitelem je Doc. Jan Černocký a jejím ředitelem výzkumu je Doc. Lukáš Burget. Skupina se věnuje výzkumu a vývoji technologií pro dolování informací z řeči - identifikace řečníka a jazyka, rozpoznávání jazyka a detekce klíčových slov. Na přelomu tisíciletí (před boomem hlubokého učení) byla jedna z mála, které v dolování z řeči využívaly neurální přístupy a dodnes přispívá k definici jejich state-of-the-art v oblasti řečových technologií. Skupina se proslavila fonémovým rozpoznávačem, neurálními přístupy k extrakci parametrů z řeči a efektivním multi-lingválním trénováním DNN pro rozpoznávání řeči. V oblasti rozpoznávání jazyka a řečníka se podílela na definici i-vektorů a end-to-end systémů a má dlouhodobě vynikající výsledky v mezinárodních evaluacích pořádaných americkým Národním institutem pro standardizaci a technologie NIST. Výzkumníci skupiny jsou pravidelně zváni na prestižní akce, jako např. Frederick Jelinek summer research workshops. Skupina získala od mezinárodní organizace ISCA pořadatelství největší světové konference o zpracování řeči Interspeech v roce 2021 v Brně.

BUT Speech@FIT se zúčastnila řady významných projektů financovaných z prostředků Evropské komise, sahajících od sběru řečových databází (SpeechDat-E - FP4), přes rozpoznávání meetingů (M4 - FP5, AMI - FP6, AMIDA - FP6) až k mobilní biometrii (MOBIO - FP7), detekci audiovizuálních událostí (DIRAC – FP6 a CareTaker - FP6) a využití řečových technologií v letectví (A-PiMod - FP7). Zajišťovala vědecké vedení projektu H2020 BISON zaměřujícího se na dovolování řeči z komerčních kontaktních center. Skupina je jediným českým pracovištěm s historií projektů financovaných vládou USA (IARPA programy BEST, Babel a Lorelei, DARPA programy RATS a Material, projekt financovaný Air Force). Je rovněž financována na národní úrovni (MŠMT, MPO, MO a MV) a je součástí IT4I Centra excellence.

BUT Speech@FIT intenzivně spolupracuje s průmyslem v ČR i v zahraničí. Stála u zrodu tří spin-off firem: Phonexia (založena 2006) dodává software pro řečovou analytiku pro komerční a bezpečnostní oblast, ReplayWell (2011) vyvíjí a komercializuje cloudové technologie pro indexování a dolování řečových záznamů a Voice Dimension (2017) poskytuje konsultační služby v oblasti bezpečnosti hlasových technologií pro bankovní sektor.

BUT Speech@FIT je aktivní ve vývoji open-source softwaru: spoluzaložila a aktivně se podílí na vývoji nejpoužívanějšího toolkitu pro dolování informací z řeči KALDI. a její fonémový rozpoznávač PHNREC získal mezinárodní renomé. Skupina disponuje zařízením pro masivní experimenty v oblasti dolování dat: >1000 CPU jednotkami na systému Linux, datovými servery s celkovou kapacitou >370 TB, desítkami GP-GPU akceleratorů a řečovými a jazykovými databázemi.

## 2. Klíčoví výzkumníci

Doc. Lukáš Burget, Ph.D. (SCOPUS h-index 35) – vědecký vedoucí skupiny, Gaussovské a neurální techniky ve strojovém učení pro dolování v řeči.

Martin Karafiát, Ph.D. (SCOPUS h-index 25) – rozpoznávání řeči, multilingvální přístupy, učení na omezených a heterogenních zdrojích.

Pavel Matějka, Ph.D. (SCOPUS h-index 22) – identifikace řečníka a jazyka – NIST evaluace, end-to-end systémy, adaptace na nové domény a přenosové kanály.

## 3. Klíčové metody a technologie

fonémové rozpoznávání, neurální přístupy k extrakci parametrů z řeči, multilingvální trénování DNN pro rozpoznávání řeči, i-vektorové a end-to-end systémy pro rozpoznávání jazyka a řečníka, neurální techniky pro jazykové modelování.

## 4. Top 3 výsledky

V letech 2005-2010 nejlepší fonémový rozpoznávač na světě, využitý mezinárodní komunitou pro úlohy indexace, vyhledávání klíčových slov a identifikace jazyka.

První funkční systém pro end-to-end rozpoznávání jazyka, vyvinutý ve spolupráci s Google Inc.

Efektivní přístup pro multilingvální trénování DNN pro rozpoznávání řeči, umožňující transfer z dobře reprezentovaného jazyka na jazyk s omezenými zdroji.

## 5. Top 5 projektů

EU H2020 Big Speech data analytics for cONtact centers - BISON (2015-2017).

DARPA RATS (2010-2017) – tým koordinovaný Raytheon BBN, USA.

IARPA BABEL (2012-2016) – tým koordinovaný Raytheon BBN, USA.

DARPA Lorelei (2015-2019) - tým koordinovaný USC-ISI, USA.

GAČR NEUREM3 - Neuronové reprezentace v multimodálním a mnohojazyčném modelování (2019-2023) - projekt excelence v základním výzkumu (hlavní řešitel).

## 6. Významný aplikační výstup či výsledek spolupráce s praxí nebo v praxi nasazen

Sada modelů pro automatické rozpoznávání řeči nasazená firmou Phonexia v produkčních systémech.

Systém pro indexování a vizualizaci nasazený firmou Replaywell v systémech <https://superlectures.com/> a <https://spokendata.com>

Řada absolventů ve významných firmách: Tomáš Mikolov (Facebook), Mirko Hannemann (Apple), Marcel Kockmann (Voice Trust), Petr Motlíček (IDIAP), atd.