

Výzkumný tým na Technické univerzitě v Liberci – ITE TUL

Vedoucí týmu: Prof. Ing. Jan Nouza, CSc.

Označení týmu: Ústav informačních technologií (ITE) - SpeechLab a ASAP Group
Technická universita v Liberci (TUL)

1. Obsah výzkumu – state-of-the art

Výzkumný tým působí na TUL od roku 1994 a zabývá se zejména zpracováním mluvené řeči (SpeechLab, <https://www.ite.tul.cz/speechlab/>) a dále obecnějším zpracováním především akustických signálů (ASAP Group, <https://asap.ite.tul.cz/>). V roce 2019 tým tvoří 1 profesor, 2 docenti, 4 pracovníci s titulem PhD, 6 doktorandů a 1 technický pracovník.

V oblasti zpracování řeči je hlavní pozornost zaměřena zejména na výzkum a vývoj metod automatického přepisu řeči v multilingválním prostředí (efektivní adaptace akustických a jazykových modelů na nové jazyky), vývoj a implementace on-line a off-line systémů pro automatický přepis na individuálních i distribuovaných platformách, využití v aplikacích typu člověk – stroj (diktování, hlasové ovládání, monitoring a analýza telefonních hovorů apod.), vývoj a aplikace metod rychlé identifikace klíčových slov v proudu řeči nebo v rozsáhlých zvukových archivech.

Skupina ASAP se zabývá metodami zpracování signálů z více senzorů a jejich aplikacím ve zlepšování řečového signálu, redukce šumu, separace signálů a zpracování biosignálů. Hlavním předmětem zájmu jsou metody slepé separace a slepé extrakce signálů na základě pokročilých matematických modelů, výzkumem teoretických vlastností a mezí těchto modelů, vývojem rychlých algoritmů pro odhad parametrů a různými modifikacemi, které efektivně kombinují matematický model s nejrůznějšími formami apriorní informace včetně částečného využití natrénovaných modelů. Cílovými aplikacemi jsou mimo jiné i robustní rozpoznávání řeči a pasivní sledování (tracking) cíle v reálných podmínkách hluku a interferencí.

2. Klíčoví výzkumníci

Prof. Jan Nouza – zaměření na počítačové zpracování řeči, v od roku 1995 se jako řešitel nebo spoluřešitel podílel na výzkumných projektech GAČR (6), TAČR (4), MŠMT (2), MK (1), MV (1) a aktivně se účastnil několika evropských projektů COST. V databázi WoS je uveden jako (spolu)autor u 75 článků (232 citací/ 119 bez autocitací, h-index 7), v databázi SCOPUS 80 článků (494/224 citací, h-index 11). Podílel se významnou měrou na realizaci několika komerčně nasazených produktů výzkumu a vývoje (např. NewtonDictate, Multilingual Transcription Platform, programů pro motoricky hendikepované uživatele PC, apod.)

Doc. Zbyněk Koldovský – zaměření na zpracování digitálních signálů, na pravděpodobnostní a algebraické modely vícerozměrných dat za účelem separace/extrakce latentních zdrojů. Od roku 2007 řešitel a spoluřešitel GAČR (4) projektů, zahraničních (2) grantů a jednoho projektu smluvního výzkumu. (Spolu)autor 14 časopisových článků (většina v IEEE), 70+ článků recenzovaných konferencí (většina IEEE), 3 kapitol v knize a dvou patentů. Ve WoS má h-index 12 (626/498 citací) a ve SCOPUSu h-index 15 (939/707 citací).

3. Klíčové metody a technologie

Tým vyvinul již v roce 2004 vlastní modulární platformu pro automatické rozpoznávání řeči pracující v reálném čase s rozsáhlými slovníky (nad 500 tisíc slov). Platforma je neustále rozvíjena a doplňována o nové moduly. Využívá se jak na běžných PC, tak i ve vlastním distribuovaném prostředí (NanoGrid), nebo např. v cloudu Azure. V současné době pro ni existují moduly založené na metodách strojového učení umožňující nasazení ve 20 evropských jazycích. Komerčně je nasazena zejména v diktovacích programech (kromě češtiny též ve slovenštině, polštině, chorvatštině, slovinštině a srbštině) a v systémech automatického on-line přepisu a monitoringu TV a R vysílání ve více než 20 zemích. (Posledním přidaným jazykem, je finská varianta švédštiny).

Skupina ASAP vyvíjí od roku 2006 algoritmy pro analýzu nezávislých komponent (EFICA, BGSEP, BARBI, MULTICOMBI), slepou separaci audiosignálů (T-ABCD), řídkou rekonstrukci dat (SPARIR) a dále pro tenzorové rozklady CP a INDSCAL. V poslední době se orientuje na vývoj nových modelů pro slepou extrakci signálů, výpočet Rao-Cramérových mezí a nové rychlé algoritmy.

4. Top 3 výsledky v posledních 5 letech

Publikace:

Z. Koldovský and P. Tichavský, "Gradient Algorithms for Complex Non-Gaussian Independent Component/Vector Extraction, Question of Convergence, " IEEE Transactions on Signal Processing, vol. 67, no. 4, pp. 1050-1064, Feb. 2019.

Z. Koldovský and F. Nesta, "Performance Analysis of Source Image Estimators in Blind Source Separation," IEEE Transactions on Signal Processing, Vol. 65, No. 16, pp. 4166-4176, ISSN:1053-587X, Aug 2017.

P. Cerva, J. Silovsky, J. Zdansky, J. Nouza, L. Seps, "Speaker-adaptive speech recognition using speaker diarization for improved transcription of large spoken archives", Speech Communication, vol. 55, no. 10, pp. 1033-1046, 2013

5. Top 5 projektů v posledních 5 letech

TAČR (TA04010199)- MULTILINMEDIA - Multilingual platform for automatic monitoring and analysis of media, (2015-2017)

TAČR (TH03010018) – DeepSpot - Multilingual technology for spotting and instant alerting, 2018-2021

N62909-18-1-2040, Office of Naval Research Global, NICOP – Adaptive Algorithms for Independent Component/Vector Extraction (2018)

GAČR 17-00902S, The Czech Science Foundation, Advanced Joint Blind Source Separation Methods (2017-2019)

DA-15-114599, California Community Foundation, Noise reduction of far field speech recordings using two or more microphones (2014-2017)

6. Spolupráce s průmyslovou a klinickou praxí:

Výzkumný tým spolupracuje už od roku 2007 s významnou českou firmou v oblasti monitoringu, společností Newton Media, a.s., od roku 2009 též s firmou Newton Technologies, a.s. Praktickým výsledkem spolupráce jsou komerčně nasazené programy pro diktování (Newton Dictate, prodávány nyní v 6 jazykových verzích a

nasazený v tisících kusech ve státních institucích i soukromých firmách). Firma Newton Media také provozuje službu automatického přepisu a monitoringu několika set zdrojů (zejména televizních a rozhlasových stanic, provozovaných ve 20 evropských jazycích a odebíraných klienty z více než 10 zemí). Ve spolupráci s Newton Technologies byly dále realizovány zakázky, jejichž součástí je nasazení řečových technologií např. při jednání zastupitelstev a podobných orgánů, v monitoringu telefonních hovorů v call centrech, apod. K největším realizovaným zahraničním zakázkám patřilo např. vybavení všech soudů v Chorvatsku diktovacími programy, či nasazení stejného programu na polských prokuraturách.