**Návrh jazykového modelu a modulu extrakce klíčových slov pro systém převodu hlasových příkazů v přirozené češtině do akcí robotického zařízení**

**(Language model and keyword extraction design for system infering robotic device actions from spoken commands in natural czech language)**

**Navrhovatel:**

*Ing. Martin Vejvar*

Ústav počítačové a řídicí techniky

Fakulta chemicko-inženýrská

**Další členové řešitelského týmu:**

*Ing. Nikita Dovzhenko* (spoluřešitel)

Ústav počítačové a řídicí techniky

Fakulta chemicko-inženýrská

*prof. Ing. Jan Náhlík, CSc.* (školitel)

Ústav počítačové a řídicí techniky

Fakulta chemicko-inženýrská

# Anotace

Projekt se zabývá dokončením návrhu univerzálního systému hlasového ovládání robotického či jiného zařízení, který využívá moderní metody strojového učení. Systém bude schopen zpracovávat hlasové příkazy v českém jazyce a na jejich základě určovat série akcí, které mají být zařízením provedeny. Hlasové příkazy přitom mohou být formulovány přirozeným způsobem ve větách a nejsou explicitně vázány na pevně danou sadu klíčových slov. Dokončení návrhu tohoto systému spočívá ve **vytvoření jazykového modelu** pro zpřesňování textových přepisů hlasových příkazů vytvářených již hotovým akustickým modelem **a na něj navazujícího modulu pro extrakci klíčových slov** z těchto textových přepisů.

# Aktuální stav problematiky

Využívání hlasových příkazů pro bezdotykové a dálkové ovládání programů či zařízení prochází v posledních letech velkým rozvojem. Zásadním podnětem k tomuto vývoji se stalo rozšíření levné výpočetní síly a vytvoření velkého množství datových korpusů obsahujících zvukové záznamy mluveného slova s jejich textovými přepisy. Souběh těchto jevů umožnil aplikaci metod strojového učení (Machine Learning) v oblasti zpracování a rozpoznání přirozené řeči (Natural Language Processing, NLP) a daly vzniknout zejména systémům založeným na Skrytých Markovových Modelech jako je např. CMU Sphinx [1]. Tyto systémy sice nabízejí vysokou přesnost rozpoznávání řeči, přitom však vyžadují značnou lingvistickou expertízu a v současnosti jsou již překonány modernějšími principy, zejména hlubokými neuronovými sítěmi (Deep Neural Networks, DNN) [2].

NLP systémy založené na hlubokých neuronových sítích jsou schopny mnohem lépe využít stále rostoucího množství dat v datových korpusech pro dosažení vysoké kvality rozpoznávání řeči. Typickými příklady tohoto přístupu v oblasti NLP jsou systémy DeepSpeech [3, 4, 5] společnosti Baidu, systém Wav2Letter od výzkumného týmu umělé inteligence společnosti Facebook [6] nebo systém pro rozpoznávání mluvené konverzace Watson (Large Vocabulary Conversational Speech Recognition, LVCSR) od IBM.

Všechny systémy i veškeré publikace v této oblasti jsou však omezeny převážně na anglický jazyk. Výzkum moderních metod rozpoznávání řeči a hlasového ovládání s využitím českého jazyka zatím zaostává. Spojení moderních metod jako jsou DNN s českými datovými korpusy nebylo, vyjma navrhovatelovy diplomové práce, dosud probádáno. Zásadním nedostatkem současných systémů je také jejich malá univerzálnost. Pokud zařízení (resp. program) disponuje hlasovým ovládáním, jedná se o unikátní a nepřenositelnou aplikaci optimalizovanou výhradně pro dané zařízení (resp. program). **V tomto projektu je proto položen důraz na návrh univerzálního systému převádějícího česky mluvené věty obsahující příkazy na požadované akce volitelného zařízení, a tím odstranění obou výše zmíněných nedostatků.**

[1] Lamere, Paul et al.: The CMU SPHINX-4 speech recognition system. In: IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2003), Hong Kong. Vol. 1. 2003, pp. 2-5.

[2] Maas, Andrew L et al.: Building DNN acoustic models for large vocabulary speech recognition. In: Computer Speech & Language 41 (2017), pp. 195-213.

[3] Hannun, Awni Y. et al.: Deep Speech: Scaling up end-to-end speech recognition. In: CoRR abs/1412.5567 (2014). arXiv: 1412.5567. url: http://arxiv.org/abs/1412.5567.

[4] Amodei, Dario et al.: „Deep Speech 2: End-to-End Speech Recognition in English and Mandarin. In: CoRR abs/1512.02595 (2015). arXiv: 1512.02595. url: http://arxiv.org/abs/1512.02595.

[5] Battenberg, Eric et al.: Exploring neural transducers for end-to-end speech recognition. In: 2017 IEEE Automatic Speech Recognition and understanding Workshop (ASRU). IEEE. 2017, pp. 206-213.

[6] Collobert, Ronan, Puhrsch, Christian, and Synnaeve, Gabriel.: Wav2Letter: an End-to-End ConvNet-based Speech Recognition System. In: CoRR abs/1609.03193 (2016). arXiv: 1609.03193. url: http://arxiv.org/abs/1609.03193.

# Vlastní plán projektu

Projekt si dává za cíl dokončit systém převodu hlasových příkazů v přirozené češtině do akcí robotického zařízení. Půjde o část navazující na již připravený akustický model, který byl vytvořen v rámci diplomové práce navrhovatele [[1]](#footnote-1). Budou vytvořeny zejména modul jazykového modelu a modul extrakce klíčových slov.

## Rozpis dílčích cílů projektu:

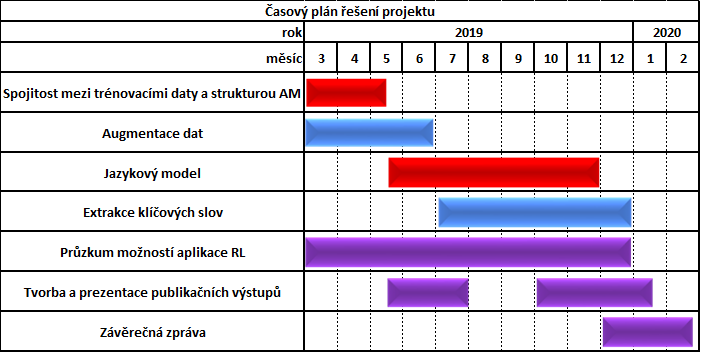
1. Ověření možností zpřesnění akustického modelu při převodu hlasových příkazů do textové podoby
   1. Určení spojitosti mezi množstvím trénovacích dat a strukturou akustického modelu (Ing. M. Vejvar)
   2. Augmentace dat pro umělé rozšíření trénovací sady dat (Ing. N. Dovzhenko)
2. Vývoj jazykového modelu (LM) pro zvýšení gramatické přesnosti přepisů z akustického modelu (Ing. M. Vejvar)
3. Návrh modulu pro extrakci klíčových slov z textových přepisů pro konečné určení akcí robota (Ing. N. Dovzhenko)
4. Průzkum možností metod posilovaného učení (Reinforcement Learning) k naučení robotického zařízení provádět akce nejen podle určených klíčových slov, ale i s ohledem na informace o okolním prostředí (oba)
5. Publikace dosažených výsledků na konferencích a v časopisech (oba)

## Postup řešení

1. První fáze projektu bude přímo navazovat na výsledky diplomové práce a bude se zabývat možnými způsoby zpřesnění akustického modelu. Základem akustického modelu je hluboká neuronová síť se strukturou skládající se z 9 vrstev, která byla natrénována pro zpracování a přepis vět v přirozené mluvené češtině do textové podoby na dvou rozsáhlých datových korpusech neformální mluvené češtiny PDTSC 1.0 a ORAL2013.
   1. V tomto kroku bude blíže prozkoumán vliv velikosti množiny trénovacích dat (množství unikátních zvukových nahrávek použitých k trénování akustického modelu) a struktury akustického modelu (počet a typy vrstev hluboké neuronové sítě, která je jeho základem).
   2. Zvyšování množství unikátních trénovacích dat je možno dosáhnout několika způsoby. Zřejmé je opatření více nahrávek z dosud nevyužitých korpusů mluvené češtiny. Jejich kvalitu však nelze předem zaručit a proces návrhu vhodného předzpracování je pro každý korpus unikátní a zdlouhavý. Alternativou je právě augmentace dat, kdy jsou již existující nahrávky upraveny tak, aby simulovaly nahrávky unikátní. K nahrávkám z již využitých korpusů proto budou přidány různé typy šumových signálů, čímž se jednak zvětší počet nahrávek a jednak bude docíleno vyšší přesnosti a robustnosti systému v rušných prostředích.
2. Jelikož pro správné vyvození akcí zařízení je zásadní dosažení co nejvyšší přesnosti přepisů mluveného slova, bude pro zvýšení gramatické přesnosti výstupů z akustického modelu vytvořen a natrénován jazykový model založený, obdobně jako akustický model, na principech učení hlubokých neuronových sítí. Pro natrénování LM bude opatřen dostatečně velký korpus česky psaného textu.
3. Ve třetí fázi projektu, která může být realizována paralelně s fází druhou, bude navržen systém pro extrakci volitelných klíčových slov z textových přepisů mluveného slova. Klíčová slova budou sloužit jakožto zástupné pojmy pro mapování akcí, které má dle hlasového příkazu zařízení realizovat.
4. Pohyb robotických či jiných mobilních zařízení v prostředí je v praxi obtížný, zejména pokud se jedná o prostředí ne úplně známé či dokonce proměnlivé. Aby zařízení mohlo plnit akce požadované hlasovými příkazy v takovém prostředí, bude nutno zařízení naučit, jak klíčová slova a informace o prostředí zkombinovat, aby akce byly fyzikálně proveditelné. Jedním z možných způsobů, jak tohoto chování docílit je metoda využití posilovaného učení (Reinforcement Learning) ke zmíněným účelům. V rámci čtvrté fáze projektu bude proto proveden průzkum možností aplikace RL k tomuto účelu.

Pozn.: Stejně jako v diplomové práci bude v programové části projektu využíván programovací jazyk Python a doplňující knihovny Numpy, Scipy, Tensorflow 2.0, aj.

# Časový plán řešení projektu



Barevné označení odpovídá hlavním řešitelům příslušné části:

* **Ing. Martin Vejvar**
* **Ing. Nikita Dovzhenko**
* **všichni řešitelé**

# Předpokládané výstupy

Mezivýsledky projektu budou přihlášeny na konferenci *Computational Methods in Systems and Software* konající se mezi 14. a 16. říjnem 2020 a v rámci konference předloženy k publikaci ve čtvrtém vydání sborníku *Advances in Intelligent Systems and Computing (Springer)*[[2]](#footnote-2). Závěrečné výsledky projektu budou navrženy k publikaci v impaktovaných časopisech *IEEE Computational Intelligence Magazine,* *Human-Computer interaction* nebo *Cybernetics and systems*.

Dále lze předpokládat, že tento vytvořený systém hlasového ovládání bude mít i značný aplikační potenciál pro řízení průmyslových i jiných zařízení.

# Údaje o připravenosti navrhovatele, řešitelského týmu a řešitelského pracoviště

**Navrhovatel Ing. Martin Vejvar** je absolventem bakalářského studijního oboru Inženýrská informatika (s vyznamenáním) a magisterského programu Senzorika a kybernetika v chemii na VŠCHT Praha (cena rektora za vynikající studijní výsledky).

V současnosti studuje první ročník doktorského studia v oboru Technická kybernetika. Projekt přímo navazuje na diplomovou práci navrhovatele, jejíž náplň a návaznost již byla vysvětlena výše v kap. 2.

Doktorand se již 3. rokem zabývá strojovým učením a je zručný v jazyku Python 3 (ve výzkumné sféře ML hojně využívaném) a modulech Pandas, NumPy, SciPy a Tensorflow 1.0 i 2.0.

Jeho práce byly třikrát oceněny v ústavních kolech SVK 2016-2018 (1., 3. a 2. místo).

V rámci studia strojového učení se stal držitelem certifikátu úspěšného dokončení Nanodegree programu “Natural Language Processing” od vzdělávací organizace Udacity, Inc.

V minulém roce se již podílel na účasti VIGA jako spolunavrhovatel projektu “Zvýšení selektivity polovodičových plynových senzorů za pomoci modulace pracovních podmínek” pod vedením navrhovatele Ing. Fary.

**Spoluřešitel Ing. Nikita Dovzhenko** je absolventem bakalářského studijního oboru Forenzní analýza a magisterského programu Forenzní chemie na VŠCHT Praha.

Bakalářskou práci *Digitální forenzní analýza* vypracoval pod vedením prof. Ing. Jana Náhlíka, CSc. a diplomovou práci *Modifikované elektrody pro elektrochemickou detekci psychoaktivních látek* se školitelem doc. Mgr. Taťjanou Šiškanovou, CSc.

Aktuálně studuje první ročník doktorského studia v oboru Technická kybernetika s tématem disertační práce *Metody posilovaného učení v modelování a řízení biotechnologických procesů* pod vedením prof. Ing. Jana Náhlíka, CSc.

**Školitel prof. Ing. Jan Náhlík, CSc.** je odborníkem v oboru technické kybernetiky.

V současné době působí jako profesor v Ústavu počítačové a řídicí techniky.

Dlouhodobě se zabývá moderními metodami řízení biotechnologických procesů, včetně metod umělé inteligence od expertních řídicích systémů[[3]](#footnote-3) až po současné metody automatického učení a neuronové sítě[[4]](#footnote-4).

Působil jako odpovědný řešitel řady národních i mezinárodních projektů s touto tématikou v programech GAČR, TAČR, COPERNICUS, 5. RP EU a 6. RP EU.

V tomto projektu odpovídá za organizaci práce a odborné konzultace.

# Finanční nároky projektu

Celkové nároky na projekt jsou odhadnuty na 118 335 Kč. Z celkové částky bude 59 216 Kč vynaloženo na osobní náklady a 25 600 Kč na provozní náklady. Osobní náklady se skládají ze stipendií navrhovatele a spoluřešitele ve výši 2x 28 000 Kč, tj. celkem 56 000 Kč, a odměny pro školitele ve výši 2 400 Kč, ke které se vztahuje odvod na sociální a zdravotní pojištění za školitele ve výši 816 Kč. Provozní náklady 25 600 Kč pokrývají jednak konferenční poplatek 9 400 Kč na mezinárodní konferenci 4th Computational Methods in Systems and Software 2020 a jednak zápisné ve výši 16 200 Kč na kurz Deep reinforcement Learning, což je tzv. Nanodegree program (MOOC kurz) vzdělávací organizace Udacity, Inc. Přehledný rozpis nákladů je v následující tabulce.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Náklady** | |  | **Výše** |  | | *I. osobní* | |  |  |  | |  | stipendium | Ing. Martin Vejvar | 28000 | Kč | |  | stipendium | Ing. Nikita Dovzhenko | 28000 | Kč | |  | mzdové náklady školitele (5%) | prof. Ing. Jan Náhlík, CSc. | 2400 | Kč | |  | soc. a zdrav. pojištění (34% mzdy) | | 816 | Kč | |  |  | součet | 59216 | Kč | | *II. provozní* | |  |  |  | |  | konferenční poplatek | CoMeSySo 2020 | 9400 | Kč | |  | nanodegree kurz | Deep Reinforcement  Learning | 16200 | Kč | |  |  | součet | 25600 | Kč | | ***přímé náklady*** | | součet | **84816** | Kč | | *III. režijní (39,52%)* | |  | 33519 | Kč | | **Celkové náklady projektu** | | | **118335** | **Kč** | |

# Publikace řešitelského týmu

Náhlík, J., Vovsík, J.: Laboratoř řízení bioprocesů. In: Kadlec, K., Kmínek, M., Kadlec, P. (@) *Měření a řízení v potravinářských a biotechnologických výrobách*. Ostrava: Key Publishing s.r.o., 2015, s. 12. ISBN 978-80-7418-232-7.

Kohout, J., Cigler, J., Široký, J., Hrnčiřík, P., Náhlík, J., Mengler, J.: Application of Knowledge-based Control on Antibiotics Production Fermentation Process. Proceedings of the International Workshop on Computational Intelligence for Multimedia Understanding IWCIM 2015. 2015.

Mareš, J., Kukal, J., Hrnčiřík, P., Náhlík, J.: Process state classification of fed-batch fermentation based on process variables analysis. *Biochemical Engineering Journal*. 2016, 112(August), 178-185. ISSN 1369-703X.

Náhlík, J., Hrnčiřík, P., Mareš, J., Rychtera, M., Kent, CA.: Towards the design of an optimal strategy for the production of ergosterol from Saccharomyces cerevisiae yeasts. *Biotechnology Progress*. 2017, 33(3), 838-848. ISSN 1520-6033.

Hrnčiřík, P., Náhlík J., Mareš, J.: Strategies for Automated Control of the Bioproduction of Mcl-PHA Biopolymers. *Chemical and Biochemical Engineering Quarterly*. 2017, 31(3), 241-250. ISSN 0352-9568.

1. Vejvar, M. *Návrh systému ovládání robotických i jiných zařízení hlasovými příkazy v přirozené řeči*. Diplomová práce, VŠCHT, Praha, 2019, GitHub: *https://github.com/vejvarm/speech-recognition-at-UCT-Prague*. [↑](#footnote-ref-1)
2. https://www.springer.com/series/11156 [↑](#footnote-ref-2)
3. Hrnčiřík, P., Náhlík, J., Vovsík, J.: The BIOGENES System for Knowledge-based Bioprocess Control. *Expert Systems with Applications* 23, 2, 145-153 (2002). ISSN: 0957-4174 [↑](#footnote-ref-3)
4. Vaněk, M., Hrnčiřík, P., Vovsík, J., Náhlík, J.: On-line estimation of biomass concentration by neural network using information about metabolic state. *Bioprocess and Biosystems Engineering* ….. doplnit…. ISSN: 1615-7591 [↑](#footnote-ref-4)