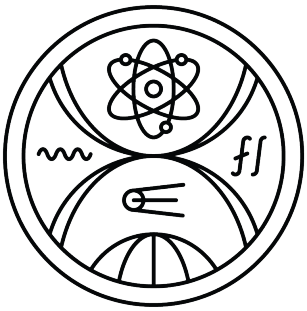
Univerzita Komenského v Bratislave
Faktulta matematiky, fyziky a informatiky

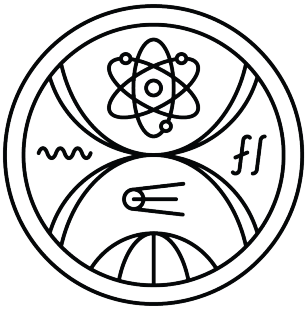


ANALÝZA MODELOV PRE ROZPOZNÁVANIE HOVORENEJ REČI

Diplomová práca

2026 Bc. Antónia Amosová

Univerzita Komenského v Bratislave
Faktulta matematiky, fyziky a informatiky



ANALÝZA MODELOV PRE ROZPOZNÁVANIE HOVORENEJ REČI

Diplomová práca

Study program:	Aplikovaná informatika
Branch of study:	Aplikovaná informatika
Department:	Katedra aplikovanej informatiky
Supervisor:	prof. RNDr. Roman Ďurikovič, PhD.
Consultant:	Mgr. Miroslav Hirjak

Bratislava, 2026 Bc. Antónia Amosová

Týmto vyhlasujem, že som túto tézu napísal sám, iba s pomocou referenčnej literatúry, pod starostlivým dohľadom môjho poradcu pre diplomovú prácu.

.....

Bratislava, 2026 Bc. Antónia Amosová

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Antónia Amosová
Študijný program: aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Analýza modelov pre rozpoznávanie hovorenej reči
Analysis of Spoken Speech Recognition Models

Anotácia: Práca sa sústreďuje na rozdelenie modelov do dvoch kategórií – serverové riešenia (cloudové, využívajúce výpočtové zdroje vzdialených serverov) a offline riešenia (lokálne, fungujúce bez pripojenia na internet). Na základe experimentálnej analýzy budú jednotlivé modely porovnané z hľadiska presnosti rozpoznávania (WER – Word Error Rate), rýchlosti spracovania, výkonu, jednoduchosti integrácie a modifikovateľnosti, ako aj robustnosti voči rôznym formám hovoreného prejavu – vrátane regionálnych dialektov a emocionálnych odtieňov hlasu.

Cieľ: Cieľom diplomovej práce je preskúmať, analyzovať a porovnať existujúce modely pre automatické rozpoznávanie reči (ASR – Automatic Speech Recognition) so zameraním na ich aplikáciu pre slovenský jazyk. Výstupom práce bude odporúčanie a implementácia najvhodnejšieho modelu, ktorý dosiahne cieľovú presnosť nad 75% pri spracovaní slovenskej reči v reálnych podmienkach. Súčasťou riešenia bude aj návrh metodiky trénovania a doladenia modelu na špecifické slovenské jazykové a akustické charakteristiky. Práca má potenciál prispieť k rozvoju technológií spracovania reči pre slovenské prostredie a k praktickému využitiu v aplikáciách vyžadujúcich spoľahlivú konverziu hovoreného slova do textu.

Literatúra: https://www.sciencedirect.com/org/science/article/pii/S1526149222002880
https://kjset.kiu.ac.ug/assets/articles/1718789093_review-of-techniques-used-in-speech-signal-processing.pdf
https://arxiv.org/abs/2509.19270

Vedúci: prof. RNDr. Roman Ďurikovič, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. RNDr. Tatiana Jajcayová, PhD.

Spôsob sprístupnenia elektronickej verzie práce:
bez obmedzenia

Dátum zadania: 24.11.2025

Dátum schválenia: 24.11.2025 prof. RNDr. Roman Ďurikovič, PhD.
garant študijného programu

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

.....
študent

.....
vedúci práce

Poďakovanie

TODO

Abstract

TODO

Keywords:

Abstrakt

TODO

Kľúčové slová:

Obsah

Terminológia

Motivácia

TODO

Kapitola 1

Úvod

Táto práca sa zameriava na riešenie problému nedostatočnej presnosti súčasných modelov pre rozpoznávanie slovenskej hovorenej reči v reálnom čase. V rámci prvej časti analyzujeme existujúce modely automatického rozpoznávania reči (ASR) pre slovenský jazyk, pričom sa zameriame na riešenia podporujúce spracovanie audia v reálnom čase, ako aj na modely určené na prepis statických nahrávok.

Na základe analytického porovnania a kritéria najnižšej chybovosti (Word Error Rate – WER) vyberieme optimálny základný model. Tento model bude následne jemne doladený (fine-tuned) s využitím relevantných, verejne dostupných rečových datasetov pre slovenčinu.

Návrhová a implementačná fáza práce sa opiera o dva kľúčové metodologické piliere. Prvým je prístup predstavený v práci [2], ktorá demonštrovala úspešnú adaptáciu architektúry Whisper na spracovanie v reálnom čase pomocou dvojpriechodového (two-pass) dekódovania. Táto metodika umožní transformovať robustný offline model na systém schopný generovať okamžitý rečový prepis s minimálnou latenciou. Druhým pilierom je zabezpečenie hardvérovej udržateľnosti a efektívnosti na mobilných zariadeniach, kde sú výpočtové zdroje limitované. V tejto oblasti sa práca opiera o závery autorov [1], ktorí detailne analyzujú vplyv konfigurácie parametrov modelu na spotrebu energie. Integrácia týchto dvoch prístupov umožní vytvoriť model, ktorý dosahuje nielen vysokú presnosť a nízku latenciu prepisu, ale je zároveň optimalizovaný pre beh na energeticky obmedzenom hardvéri.

Kapitola 2

Analýza a porovnanie modelov

2.1 Typy modelov pre rozpoznávanie reči

TODO

2.2 Rozpoznávanie hovorenej reči

TODO

Kapitola 3

Model pre rozpoznávanie hovorenej reči (ASR)

3.1 Skúmané modely

TODO

3.2 Úprava modelu na streamovací

TODO

3.3 Zmenšenie modelu

TODO

3.4 Optimalizácia výkonu modelu pre mobilné zariadenia

TODO

Kapitola 4

Výskum

TODO

4.1 Navrhovaný model

TODO

4.1.1 Príprava datasetu pre trénovanie

TODO

4.2 Dolaďovanie navrhovaného modelu

TODO

Kapitola 5

Výsledky

TODO

Záver

TODO

Literatúra

[1] Yang Li, Yuan Shangguan, Yuhao Wang, Liangzhen Lai, Ernie Chang, Changsheng Zhao, Yangyang Shi, and Vikas Chandra. Breaking down power barriers in on-device streaming ASR: Insights and solutions. In Weizhu Chen, Yi Yang, Mohammad Kachuee, and Xue-Yong Fu, editors, *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 3: Industry Track)*, pages 616–626, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics.

[2] Haoran Zhou, Xingchen Song, Brendan Fahy, Qiaochu Song, Binbin Zhang, Zhendong Peng, Anshul Wadhawan, Denglin Jiang, Apurv Verma, Vinay Ramesh, Srivas Prasad, and Michele M. Franceschini. Adapting Whisper for Streaming Speech Recognition via Two-Pass Decoding. In *Interspeech 2025*, pages 4428–4432, 2025.