

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra kybernetiky



► **FAKULTA
APLIKOVANÝCH VĚD
ZÁPADOČESKÉ
UNIVERZITY
V PLZNI**

KKY/ARŘ

Yauheni Petrachenka
14. listopadu 2025

1 Experiments

Cílem této práce bylo experimentálně otestovat a vzájemně porovnat dva modely automatického rozpoznávání řeči – Whisper a Wav2Vec2. Pro vyhodnocení přesnosti rozpoznávání byla použita metrika WER (Word Error Rate), definovaná standardním způsobem jako poměr součtu substitucí, vložení a delecí k celkovému počtu slov v referenční větě.

V experimentech byl použit model Whisper (verze Turbo) a model Wav2Vec2 (Large-XLSR-53), který byl dále fine-tuningován pro češtinu. Počet parametrů obou modelů je uveden v přiložené tabulce 2.

Jako testovací vstup byla zvolena česká věta „Modrý stín pomalu přeplouvá přes tichou hladinu jezera.“ Tato věta byla následně uměle zašuměna čtyřmi úrovněmi šumu s hodnotami SNR = 0 dB, 10 dB, 20 dB a 30 dB. Oba modely poté rozpoznávaly jednotlivé varianty signálu i jeho nešuměnou verzi. Výsledky jsou shrnutы v tabulce 1.

Z provedených experimentů vyplývá, že model Whisper dosahuje vyšší přesnosti rozpoznávání než Wav2Vec2. Whisper vykazuje znatelnou chybovost pouze v nejvíce zašuměných případech (SNR = 0 dB a 10 dB), zatímco Wav2Vec2 vykazuje chyby ve všech testovaných podmírkách. Na druhé straně Wav2Vec2 dosáhl výrazně nižší doby zpracování, a tedy vyšší rychlosti inference.

Dále bylo otestováno rozpoznávání běloruštiny pomocí modelu Whisper. Experimenty ukázaly, že model často nesprávně identifikuje jazyk a místo běloruštiny volí ruštinu (ověřeno na více příkladech, než je uvedeno v kódu). Po explicitním zadání jazyka sice již nedochází k chybné detekci jazyka, avšak model stále nedokáže větu kvalitně rozpoznat. Rovněž není schopen spolehlivě zachytit některé foneticky specifické běloruské hlásky, které nahrazuje zvuky podobnými ruštině.

Model	WER	Noise	Time
Wav2vec2	0.750	-	8.173
Wav2vec2	0.875	0	3.779
Wav2vec2	0.875	10	3.751
Wav2vec2	0.875	20	4.740
Wav2vec2	0.875	30	3.797
Whisper	0.0	-	16.951
Whisper	1.0	0	63.134
Whisper	0.375	10	16.469
Whisper	0.0	20	116.562
Whisper	0.0	30	16.194

Tabulka 1: Výsledky experimentů

Model	počet parametrů
Whisper	809M
Wav2vec2	317M

Tabulka 2: Počet parametrů