

Automatická kontrola výslovnosti

Jan Tupý¹

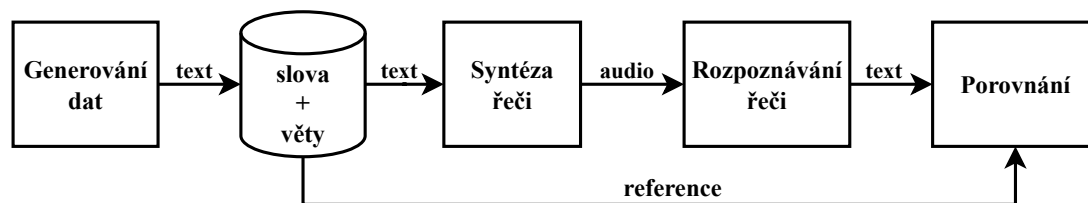
1 Úvod

Řeč je zásadním nástrojem lidské komunikace, který umožňuje lidem sdílet myšlenky, emoce a záměry. Ve světě neustále se rozvíjejících technologií se hlasové dialogové systémy stávají nezbytnými nástroji v mnoha oblastech, včetně zákaznické podpory, asistivních technologií, vzdělávacích aplikací a lékařské rehabilitace. Cílem této práce je vytvořit webovou aplikaci, která automaticky kontroluje řeč uživatele. Aplikace je určena pro osoby, které se rekonvalescencí snaží znovu získat schopnost mluvit, jako jsou například pacienti po mozkové příhodě.

Tato práce se nejprve zaměřuje na testování různých rozpoznávačů řeči, které jsou jádrem pro správné fungování navržené webové aplikace sloužící k automatické kontrole výslovnosti. Dále se zabývá návrhem aplikace a její realizací. Aplikace využívá grafémovou analýzu pro porovnání vyslovených slov s jejich referenčním textem, čímž poskytuje uživatelům okamžitou zpětnou vazbu na jejich výslovnost. Hlavním cílem je, aby výslovnost uživatelů byla srozumitelná, což se měří pomocí Levensteinovy vzdálenosti, která kvantifikuje shodu na procentuální bázi.

2 Testování rozpoznávačů řeči

Testování se soustředí na tři rozpoznávače řeči, z nichž každý využívá různé technologie a přístupy k rozpoznávání řeči. Jedním z nich je tradiční rozpoznávač založený na architektuře **TDNN** (*Time Delay Neural Network*) s akustickým modelem, kde dekódování probíhá nad fonémovým výstupem s využitím slovníku a jazykového modelu. Dva další modely jsou založeny na modernější architektuře **wav2vec 2.0**, konkrétně model (Lehečka et al., 2022), z nichž jeden využívá jazykový model a druhý grafémový přístup bez přidaného jazykového modelu. Na obrázku 1 je zobrazené schéma testování ASR, kde se nejprve vygeneruje text, který se následně syntetizuje (použito 7 různých hlasů - *Unit selection* (Tihelka et al., 2018)). Nahrávky se pomocí jednotlivých rozpoznávačů převedou zpět na text a porovnají s referencí.

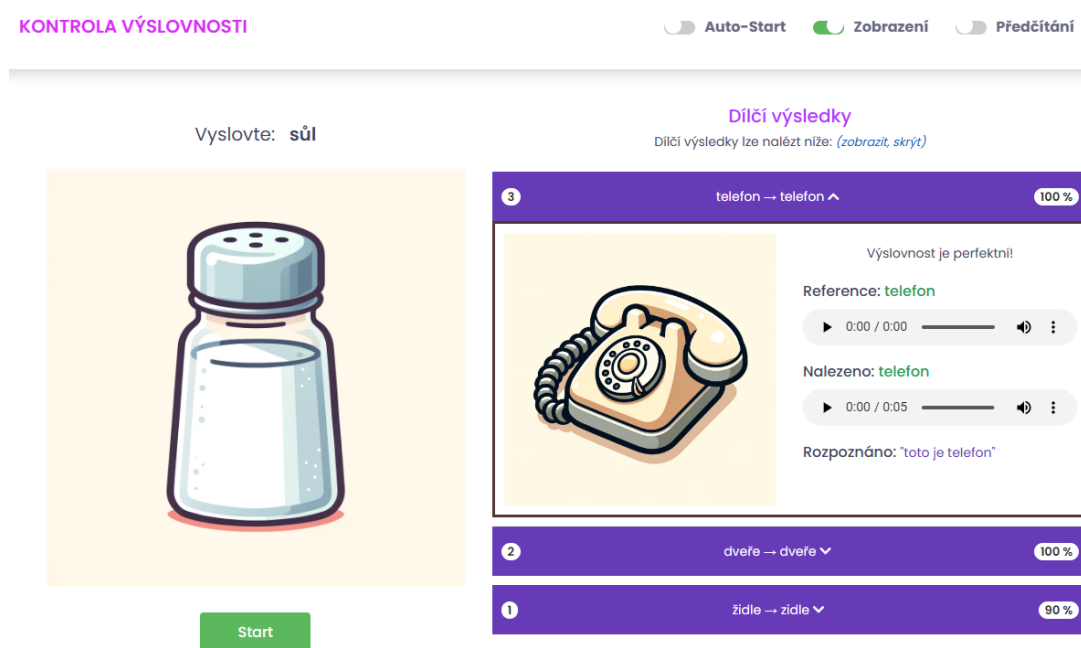


Obrázek 1: Schéma průběhu testování rozpoznávání řeči (ASR)

¹ student navazujícího studijního programu Kybernetika a řídicí technika, specializace Umělá inteligence a automatizace, e-mail: jtupy@students.zcu.cz

3 Webová aplikace na kontrolu výslovnosti

Aplikace je realizována jako webová stránka s důrazem na responzivitu, aby byla správně zobrazena na všech možných zařízeních. Přidávání dalších slov je navrženo tak, aby bylo co nejjednodušší: stačí vložit obrázky do příslušné složky a přidat textový název. Při spuštění aplikace se objeví postupně vyskakovací okna pro nastavení aplikace, upozornění, doporučení a instrukce. Poté se zobrazí hlavní část aplikace (Obrázek 2), kde lze začít se samotnou výslovností. Slova lze vyslovovat samostatně nebo ve větě, ale vždy v prvním pádě. Vyhodnocení probíhá kontinuálně a na konci je celkové vyhodnocení. Hodnocení je založeno na procentuální shodě vysloveného slova s referenčním textem pomocí Levensteinovy vzdálenosti (LD). U LD je možné nastavit libovolné ceny změn v rámci tzv. *loss matrix*, která umožňuje různé ceny pro různé znaky (například cena za záměnu znaků *b* za *h* bude vyšší než za záměnu *a* za *á*).



Obrázek 2: Ukázka webové aplikace na kontrolu výslovnosti

4 Závěr

V rámci testování vyšel jako nejvhodnější rozpoznávač **wav2vec 2.0** bez přidaného jazykového modelu na výstupu, který využívá grafémový přístup. Výsledkem je webová aplikace, která umožňuje automatickou kontrolu výslovnosti.

Literatura

- Lehečka, J., Švec, J., Pražák, A., Psutka, J. (2022) *Exploring Capabilities of Monolingual Audio Transformers using Large Datasets in Automatic Speech Recognition of Czech*. In: *Proc. Interspeech 2022*, pp. 1831–1835.
- Tihelka, D., Hanzlíček, Z., Jůzová, M., Vít, J., Matoušek, J., Grüber, M. (2018) *Current State of Text-to-Speech System ARTIC: A Decade of Research on the Field of Speech Technologies*. In: *Text, Speech, and Dialogue*, Springer International Publishing, Cham, pp. 369–378. ISBN 978-3-030-00794-2.