# Využití funkce rozpoznání řeči za účelem zlepšení výslovnosti anglických slov

Bc. Martin Chaloupka
Faculty of Informatics and Management
University of Hradec Kralove,
Hradec Kralove, Czech Republic
chaloma1@uhk.cz

***Abstract*—Projekt se zabývá vytvořením mobilní aplikace, která by uživateli zlepšila výslovnost anglických slov. V článku je tedy rozebráno, jaké funkce aplikace obsahuje a jak by to šlo udělat jinak. V aplikaci se kladl důraz na offline použití a správnou funkcionalitu. Po chvíli přemýšlení byly využity základní funkcionality systému Android. Třídy „TextToSpeech" a „SpeechRecognizer" s využitím „Room" databáze, která funguje nad SQLite (lokální databází systému android), které mají hlavní roli v této aplikaci a fungují i v nejzákladnějších verzích operačních systémů android. Hlavní výhodou je jednoduchost a funkcionalita i přesto, že offline verze má nakonec velkou chybovost z důvodu omezené lokální knihovny.**

***Keywords-SpeechRecognizer; Android; Exercise; Pronunciation; Room database; Text-To-Speech; Speech-To-Text;***

## I. INTRODUCTION/ÚVOD

**OBSAH**

V moderní době, kdy není výkon ani úložný prostor velkým problémem se hodně rozvíjejí technologie jako je například automatické rozpoznání řeči pro hlasové ovládání zařízení nebo převod řeči na text k usnadnění práce s psaním dokumentů. Velké společnosti jako Google, IBM, Amazon, Microsoft a Apple již mají své algoritmy pro rozpoznání řeči velmi vyvinuté s více než dostačující přesností pro většinu světových jazyků. (6)

Hlasové ovládání se využívá v produktech jako je Alexa, Google Home, Siri a další…. Jde hlavně o usnadnění přístupu k informacím pro uživatele. Hodně se také vyvíjí tzv. inteligentní domácnost, kde má uživatel hlasovou kontrolu nad teplotou v pokoji nebo svítivostí žárovek.

Naopak překlad řeči na text neboli STR (Speech-to-text recognition) má spíše využití ve vzdělávací sféře. Například při přednáškách, kdy STR zaznamená proslov přednášejícího a poté poskytne studentům psanou formu dané lekce. V civilní sféře se pak používá například k hlasovému psaní mobilních zpráv. (2)

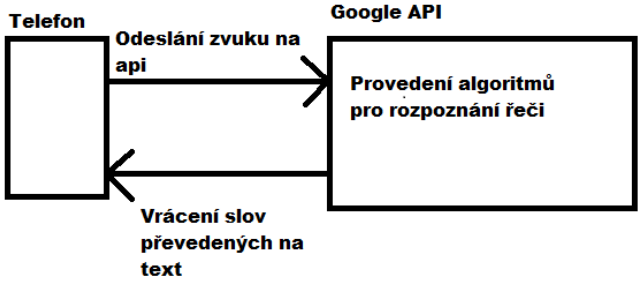
Většina těchto firem své algoritmy integrovala do moderních inteligentních telefonů nebo chytrých reproduktorů. V těchto zařízení pak dále pomáhají uživateli s různými každodenními úkoly. (3)

V dnešní době se lidé učí anglicky už od dětských let. Učí se ve škole, z her nebo i filmových titulků. Z těchto typů učení podle osobní zkušenosti vzniká jeden velký problém a tím je správné vyslovování slov. Většinou je špatná výslovnost upevněna prostředím, kde daný jazyk není mateřský, jelikož si člověk ani sám neuvědomí, že něco vyslovuje špatně a tak nemá šanci se zlepšit.

Mobilní aplikace se stávají čím dál důležitější v každodenním použití mladých uživatelů. V současné době je trh s mobilními aplikacemi velmi žádaný a právě tyto aplikace jsou vytvářeny za účelem zjednodušení života klientů. (8)

Technologie STR využívá strojového učení. To znamená, že se s každými dodanými daty zlepšuje v přesnosti překladu. Aby byla taková technologie úspěšná, bylo by zapotřebí předhodit stovky hodin hlasových dat. Z tohoto důvodu je jednoduší vybrat STR jedné z velkých firem, která nemá omezující limit pro případné testování aplikace. (9)

Z těchto důvodů je hlavním cílem v tomto projektu vytvoření mobilní aplikace (cvičně nazvané „Pronounce corrector"), která bude použitelná v každodenní situaci třeba i v offline modu. S touhle aplikací dokáže každý člověk procvičovat svou anglickou výslovnost ať už je kdekoliv a stačit by mu k tomu měl i starší mobil s operačním systémem Android. Další funkcí této aplikace bude poslech a generování slov z lokální databáze. Uživatel si bude moct zobrazit slova, ve kterých nejvíce chyboval včetně minulých procvičování.




**Obrázek 1 - Funkce rozpoznání řeči. Zdroj: Vlastní**

## II. PROBLEM DEFINITION/ DEFINICE PROBLÉMU

**OBSAH**

Pokud chce v dnešní době člověk vycestovat do ciziny nebo i rozšířit své podnikání za hranice státu, je nutné znát cizí jazyk. Pokud se však člověk učí další jazyky než svůj mateřský, vzniká problém ve správné výslovnosti. Takové chyby vznikají například při učení z psaných slov, třeba při prohlížení sociálních sítích nebo hraní počítačových her. A protože anglický jazyk je v moderním věku nejrozšířenější, tato práce se bude zabývat právě využitím hlasového rozpoznání tohoto jazyka za účelem zlepšení výslovnosti daných slov.

Nejúčinnějším způsobem ke vzdělání je bezpochyby přes zařízení, které člověk nosí stále u sebe. Uživatel se tedy může vzdělávat kdykoliv se mu zachce. Takovým zařízením je myšlen mobilní telefon. Mobilní aplikace jsou v dnešní době velmi populární u mladých jedinců. Z tohoto důvodu je tedy vhodné tyto aplikace využití pro zlepšení vzdělávacích technik. (7)

Ačkoliv se mobilní zařízení čím dál více zrychlují, operace potřebné k úspěšnému zpracování zvukových dat jsou stále celkem náročné. Z toho důvodu je lepší vybrat serverové aplikační rozhraní (dále jen API) s dostatečnou přesností překladu, na které se tato data budou zasílat ke zpracování. (5) Je také vždycky lepší najít takový server, který nemá zbytečně nízké limity přenosu dat.

Tedy dalším problémem, na který lze narazit je potřeba rychlého a stabilního připojení k internetu. (4) Když pomineme velký síťový provoz u serverové části, která je většinou spravována velkými společnostmi, je třeba řešit připojení klienta. Ačkoliv mobilní sítě nabírají na rychlostech a poslední dobou i na obsahovém limitu (většina operátorů již nabízí 10 GB měsíčně ve svých tarifech), tak jako stabilní připojení se dnes považuje pouze 4G / LTE síť. Toto připojení sice umí většina nových zařízení, ale problém je v pokrytí signálu například v údolích nebo vesnicích. V dnešní době je také všude wifi připojení, ať už v kavárnách nebo v autobusech. V připojení k neznámému poskytovali je zase bezpečnostní riziko. To se však v této práci řešit nebude, jelikož jde pouze o testový projekt, který by pouze poukázal na možnost využití STR technologie k procvičení výslovnosti jazyka. Dále v tomto projektu budeme tedy předpokládat bezchybné připojení klienta například na zabezpečenou školní wifi, protože datovou náročnost a stabilní připojení neovlivníme.

Jedním z implementačních problémů je několika hodinová limitace překladů řeči na text od vybraných API Google Cloud Speech-to-text nebo IBM Watson. Avšak Google také nabízí překlady pro Android systém bez časového omezení přes Java třídu nazvanou „SpeechRecognizer“, která zajišťuje přístup k dané překladatelské službě. A jelikož to je třída dělaná právě pro systém Android, byla proto vybrána pro tento projekt.

## III. NEW SOLUTION / NOVÉ ŘEŠENÍ

**OBSAH**

Když už se ví jaké jiné má STR využití, probereme v tomto projektu nový směr, kterým by tato technologie šla dále využít. Čím dál více lidí využívá špatnou výslovnost anglického jazyka z mnoha různých důvodů. Takto špatně vyslovená slova pak vedou ke zbytečnému nedorozumění mezi lidmi s různým mateřským jazykem. Proto se tato práce zabývá, zda by šla STR technologie použít ke zlepšení výslovnosti jednotlivých anglických slov. Teorie je tedy taková, že pokud mechanismus dokáže rozpoznat vyslovené slovo, pak by dané slovo dokázal rozpoznat i jiný člověk mluvící anglicky.

Tato aplikace bude tedy postavená na operačním systému Android, kvůli jednoduchosti implementace a vnořené podpory od firmy Google. Aby byla možnost vyzkoušení funkce v offline modu, bude zapotřebí využít lokální databázi. Aby byla opět udržena jednoduchost projektu, bude využitá SQLite databáze. Jde o databázi určenou pro lokální použití v operačních systémech Android. Z toho důvodu je SQLite databáze velmi ořezaná a o hodně jednodušší oproti klasickým databázím. To však v tomhle projektu nevadí. Implementována bude v podobě „Room persistence library“. Zkráceně Room je knihovna, která tvoří vrstvu abstrakce nad SQLite databází a zajišťuje nám tak novější a lehčí přístup k databázi. (1)


**Obrázek 2 - Databázový model aplikace**

Na obrázku č. 2 je popsán databázový model aplikace. U slov je hlavní název v textové podobě vůči kterému se porovnávají přeložené textové slova. Dále se zaznamenává „rating“, důležitý u možnosti vyřazení slov ze cvičícího programu (například při špatné funkčnosti). Poslední atribut je „ranking“, sloužící ke zjištění poměru správného nebo špatného vyslovení. Hodnotící tabulka zkoumá vše důležité pro vytvoření grafů a statistik pro uživatele sloužící k lepšímu přehledu jeho pokroku. Statistiky by pak dále mohli pomoct vývojářům, až by se odeslaly s jedinečným identifikátorem aplikace nebo podle uživatelského loginu na privátní server.

Poslechová aktivita neboli převod textu na řeč je další věc, která je bezlimitně podporovaná firmou Google v operačním systému Android. Implementace je tedy vskutku jednoduchá. Při nastartování vhodné aktivity předáme objektu „TextToSpeech“ kontext a jazyk ve kterém má aktivita fungovat. V tomto případě anglický jazyk. Poté lze už jen zavolat metodu „speak“ s textovým parametrem, který se má vyslovit.

Aktivita k procvičení výslovnosti je delegována na objekt „SpeechRecognizer“. K této třídě je již potřeba bezpečnostní povolení k přístupu mikrofonu zajištěný příkazem v manifestu. Tento objekt zajišťuje přístup k offline i online Google API, kde se zpracovává zvuk složitými algoritmy. Vstupem je tedy základní nastavení jazyka a preferencí. Výstupem se poté posílá list slov seřazených podle procentuální shody zaslaného zvuku a skutečného slova ze slovníku. Zvuk a nastavení jazyka je do objektu přenášeno třídou „RecognizerIntent“, která začne nahrávat zvuk po stisknutí tlačítka.

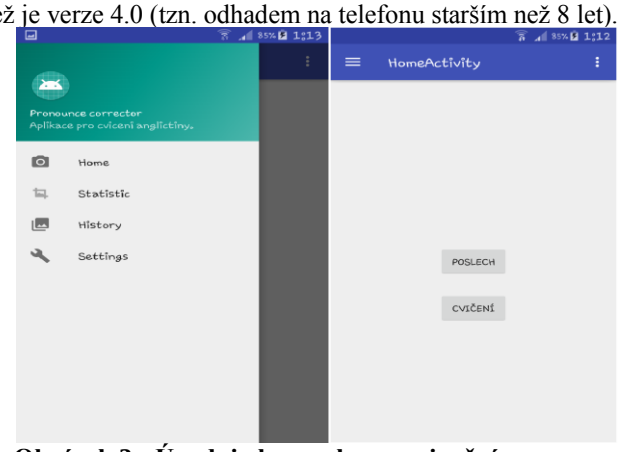
Slovo, které nám vrátí Google API v textové podobě je potom porovnáno s vygenerovaným slovem z databáze. Zde je několik možností jak dané slovo porovnat. Nejjednodušší způsob je využití metody „equals“ datové typu „string“. Jde o metodu, která porovná všechny znaky obou textových polí na stejné pozici. Pokud se jeden z nich nerovná, pak tato metoda skončí se záporným výsledkem. Složitější metoda by byla v možnosti rozebrat hlasové slovo převedené na text na jednotlivé znaky. Pokud by se všechny znaky nebo aspoň většina nacházela v generovaném slovu, pak by výslovnost byla uznána a zapsána ve výsledném hodnocení. Jelikož však Google API vrací rozpoznané slovo, bude v tom projektu stačit základní porovnávací metoda.

Aby byla aplikace úspěšná, uživatel by měl mít možnost základního vlivu na generovaná slova. V této aplikaci bude proto možnost vyřadit neoblíbená nebo nefunkční slova.

## IV. Implementation / Implementace řešení

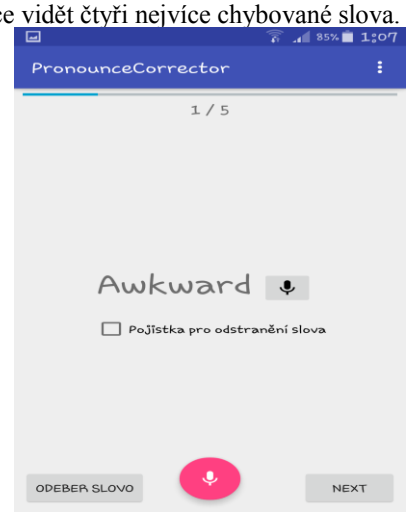
### OBSAH

Aplikace je napsaná v jazyce Java s frameworkem pro operační systém Android. Minimální SDK aplikace je 14. To znamená, že aplikace nebude fungovat na nižším androidu než je verze 4.0 (tzn. odhadem na telefonu starším než 8 let).



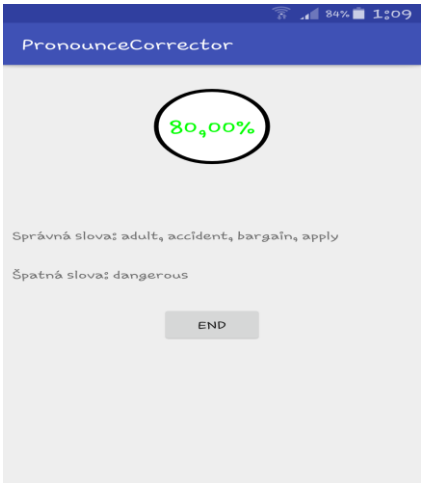

**Obrázek 3 - Úvodni obrazovka s navigačním menu**

Na obrázku č. 3 lze vidět úvodní aktivitu aplikace s možným navigačním menu. Menu může být otevřenou kliknutím na tři svislé čárky v levém horním rohu. Z této aktivity může uživatel začít lekci v poslechovém cvičení nebo klasické cvičení ve výslovnosti náhodně generovaných slov. Z navigačního menu si poté může uživatel prohlédnout své minulé výsledky v „History“ záložce. Ve „Statistic“ si lze poté prohlednout graf se špatnými a správnými slovy nebo dokonce vidět čtyři nejvíce chybované slova.




**Obrázek 4 - Aktivita pro cvičení výslovnosti slov**

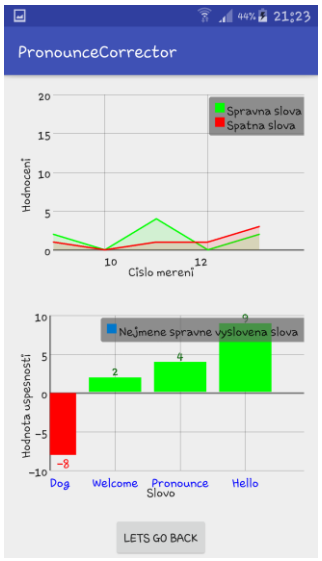
Obrázek č. 4 zaznamenává hlavní aktivitu v aplikaci. V horní části je vidět pokrok v jednotlivém cvičení. Každé cvičení je na různý počet slov, který lze změnit v nastavení aplikace. V tomhle případě je v dané lekci pět slov, které by měl uživatel vyslovit a tak se cvičení započítalo do celkového hodnocení. Po každém stisknutí „Next“ tlačítka se vygeneruje nové slovo. Uživatel si dané slovo může nechat vyslovit kliknutím na ikonu černého mikrofonu. Aby sám vyslovil slovo, stačí zmáčknout růžový tlačítko s bílým mikrofonem. Po stisknutí má uživatel zhruba 3s na vyslovení daného slova. Po vyslovení slova může uživatel znovu zkoušet slovo vyslovit, jediné co se potom počítá je hodnota daného slova, nicméně v hodnocení bude pouze zapsáno jedno špatné slovo. Poslechová aktivita vypadá stejně, akorát funguje na automatizovaném principu, kdy každé pár sekund se generuje nové slovo a samostatně se dvakrát po sobě vysloví. Rychlost generování a vyslovování lze nastavit v aplikaci.




**Obrázek 5 - Vyhodnocení cvičící aktivity**

Po dokončení cvičící lekce se zobrazí aktivita s procentuální úspěšností a výpisem správných a špatných slov podle obrázku č. 5. Poté je uživatel odkázán na úvodní obrazovku a lekce je zapsána do statistik. Uživatel si může prohlédnout všechny minulé lekce v záložce „History“ v navigačním menu (viz obrázek č. 6).


**Obrázek 6 - Historie lekcí uživatele**




**Obrázek 7 - Statistická aktivita uživatele**

V posledním obrázku č. 7 si uživatel může prohlédnout, jak moc se zlepšil nebo jaká čtyři slova si spletl nejvícekrát. Horní graf správných a špatných slov lze dotykem prstu horizontálně posouvat.

## V. TESTING OF DEVELOPED APPLICATION / TESTOVÁNÍ VYVINUTÉ APLIKACE - ŘEŠENÍ

**OBSAH**

Z důvodu neschopnosti virtuálního stroje se vstupním zvukem je zapotřebí testovat danou aplikaci na fyzickém zařízení. Z tohoto důvodu je funkcionalita aplikace testována pouze autorem projektu.

Při prvním testování je do aplikace vloženo pět testových slov: „hello, goodbye, welcome, pronounce, dog“. Aplikace je připojena k internetu a při správné výslovnosti aplikace správně zaznamenává všechny slova až na slovo „dog“. Problémem je, že serverové API vracelo pouze jedno slovo s největším procentem rozpoznání. V tomto případě se jednalo o slovo „doc“, což je zkratka pro doktora a vyslovuje se stejně s původním slovem. Oprava je v celku jednoduchá. Od serveru se teď bere až 5 rozpoznaných slov, které se poté porovnávají se zadaným slovem.

Databáze nepotřebuje zvukovou formu slov, a proto jde přidávání slov celkem bez problému, jelikož stačí pouze jejich textová verze. Aplikace nyní testuje zhruba padesát různých slov. Bohužel v offline použití se člověk musí hodně soustředit, aby dané slovo bylo správně zaznamenáno. Tento problém bohužel nelze vyřešit z důvodu lokální ořezané knihovny pro rozpoznání slov od společnosti Google. V budoucnu by bylo možné tento problém obejít s využitím jiného API pro rozpoznání hlasu.

## VI. CONCLUSIONS / ZÁVĚRY

Závěrem bych shrnul, že aplikace je plně funkční a to i v offline modu. Tudíž uživatel může procvičovat svoji výslovnost se svým mobilem, ať je kdekoliv. Bohužel bych však doporučoval tuto aplikaci jen začátečním studentům angličtiny, protože ty samé algoritmy, které se snaží rozlišit, co přesně uživatel řekl, také přispívají k malé nepřesnosti ve výslovnosti. Pokud tedy uživatel jemně zkomolí cvičící slovo, serverový algoritmus dané slovo stejně rozpozná. Takovou nepřesnost lze ovšem prominout.

Avšak využití v offline modu nedopadlo moc dobře. Zdá se, že jazyková knihovna od Googlu je pro limity velikosti mobilní paměti velice osekaná. Tudíž bych tuto technologii pro využití bez internetu nedoporučoval. V budoucnu by však šlo vyzkoušet knihovnu CMUSphinx, která je také zdarma a mezi komunitou celkem doporučovaná.

### Seznam obrázků:



### References / Reference


1. **Schramm, Daniel.** *Mobilní aplikace pro inventarizaci majetku MU.* [Online] 2019. [Cited: 8 19, 2019.] https://is.muni.cz/th/v2psm/Mobiln__aplikace_pro_inventarizaci_majetku_MU.pdf.
2. **Shadiev, Rustam, et al.** *Applications of speech-to-text recognition and computer-aided translation for facilitating*

*cross-cultural learning through a learning activity: issues and their solutions.* [Online] 12 12, 2017. [Cited: 8 19, 2019.] https://link.springer.com/article/10.1007%2Fs11423-017-9556-8.

3. **Palanica, Adam, et al.** *Do you understand the words that are comin outta my mouth?* [Online] 6 20, 2019. [Cited: 8 19, 2019.] https://www.nature.com/articles/s41746-019-0133-x.pdf.

4. **McGraw, Ian, et al.** *Personalized speech recognition on mobile devices.* [Online] 3 11, 2016. [Cited: 8 19, 2019.] https://arxiv.org/pdf/1603.03185.pdf.

5. **Chin-ChenChang, Hsiao-LingWu and Chin-YuSun.** *Notes on "Secure authentication scheme for IoT and cloud servers".* [Online] 7 2017. [Cited: 8 19, 2019.] https://www.sciencedirect.com/science/article/pii/S1574119215002151.

6. **Dharmale, Gulbakshee, Patil, Dipti D. and Thakare, V. M.** *Implementation of Efficient Speech Recognition System on Mobile Device for Hindi and English Language.* [Online] 2019. [Cited: 8 19, 2019.] https://thesai.org/Downloads/Volume10No2/Paper_12-Implementation_of_Efficient_Speech_Recognition_System.pdf.

7. **Bryndová, Vlasta.** *Využití mobilních aplikací jako podpory pro výuku.* [Online] 9 18, 2018. [Cited: 8 19, 2019.] https://theses.cz/id/9blddn/?isshlret=Vlasta%3BBryndov%C3%A1%3B;zpet=%2Fvyhledavani%2F%3Fsearch%3DbryndovC3%A1%20vlasta%26start%3D1.

8. **Bc. Manček, Lubor.** *Využití mobilních aplikací ve volnočasových pohybových aktivitách studentů Masarykovy univerzity.* [Online] 2016. [Cited: 8 19, 2019.] https://is.muni.cz/th/k76ql/Diplomova_prace_Lubor_Manek.pdf.

9. **Bansal, Sameer, et al.** *Low-Resource Speech-to-Text Translation.* [Online] 6 18, 2018. [Cited: 8 19, 2019.] https://arxiv.org/pdf/1803.09164.pdf.