

PROGRAM FOR ASSISTANCE IN LEARNING ENGLISH PRONUNCIATION

Jan Malucha

Master Degree Programme (1), FEEC BUT

E-mail: xmaluc00@vutbr.cz

Supervised by: Milan Sigmund

E-mail: sigmund@feec.vutbr.cz

Abstract: This study presents a compact MATLAB tool to assist students in learning the correct pronunciation of English language, with use of three known methods for estimation of speech signal voicing. A short introduction into speech voicing analysis is given and the developed program is briefly described. The practical use of the program was verified by test with non-native speaker.

Keywords: speech signal, pronunciation, voicing

1 ÚVOD

Vlivem pandemické situace se pojem „distanční výuka“ v současnosti stává velmi často skloňovaným tématem. S rozvojem komunikačních technologií se toto téma začíná objevovat i v diskuzích o budoucnosti vzdělávání a vyvstávají otázky, zda vzdělávání na dálku zůstane nadále pevnou součástí běžného studia. Obzvláště ve sféře středoškolského vzdělávání toto světové trendy skutečně naznačují, a má tudíž smysl zaměřit pozornost na vývoj komplexních nástrojů k podpoře distanční formy výuky. Zejména v oblasti výuky cizích jazyků lze s výhodou aplikovat poznatky teorie zpracování řečových signálů, jedné z disciplín obecného zpracování a analýzy signálů, široce využívané v oblastech od rádiových komunikací až po zdravotnické aplikace. Výuka cizích jazyků klade velký důraz na mluvenou a poslechovou formu výuky, a je zde tedy možné zefektivnit zdokonalování se ve správné výslovnosti např. anglického jazyka za podpory komplexních nástrojů analyzující řečový signál získaný skrze mikrofon.

2 ZNĚLOST A NEZNĚLOST ŘEČI

Samotný řečový signál je časovou reprezentací zachycující mluvenou řeč. Z fyzikálního hlediska jde o reprezentaci chvění přenosového média, zpravidla vzduchu, vyvolané budičem, jímž je hlasové ústrojí člověka. Jeho podoba se více či méně může lišit pro různé jazyky. V případě většiny indoevropských jazyků však můžeme dané řečové signály lingvisticky považovat za časové průběhy reprezentující slova (tedy spojení hlásek) oddělená úseky ticha. Fonologicky jednotlivé hlásky dále dělíme na znělé a neznělé, a to podle podstaty jejich tvorby [1] – znělé hlásky jsou tvořeny kvazi-periodickým kmitáním hlasivek a mohou tudíž mít základní frekvenci, resp. melodii (např. hláska /a/), oproti tomu hlásky neznělé nejsou tvořeny hlasivkami a typicky mají charakter šumu (např. hláska /š/). Úseky ticha jsou kromě pauz mezi slovy či větami typické pro rozhraní mezi určitými hláskami (např. /dk/ ve slově ředkev), což je způsobeno přechodem mezi různými stavy hlasového ústrojí.

Budeme-li se zabývat anglickým jazykem, je třeba říci, že na rozdíl od českého jazyka je angličtina až velmi citlivá na správnou výslovnost hlásek v rámci daného slova. Při špatné výslovnosti či jejich záměně totiž může velmi často dojít k naprosté změně jeho významu (např. rozdíl mezi slovy *ship* [lod'] a *sheep* [ovce]). Zároveň je třeba věnovat zvýšenou pozornost hláskám, které se v českém jazyce vůbec nevyskytují [2], a jež při řeči pro správné vyznění zpravidla vyžadují jistý cvik. Znělost/neznělost je tedy jedním z hlavních atributů správné výslovnosti.

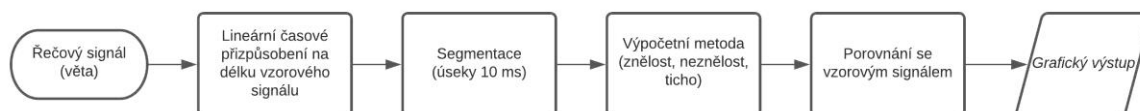
3 CÍLE PRÁCE

Předmětem tohoto příspěvku je vytvoření programu k analýze řečových signálů, jehož praktickou funkcí má být vyhodnocení a evaluace parametru znělosti a neznělosti v průběhu vstupního řečového signálu a srovnání jeho vlastností s parametry ideálního protějšku. Výstupem má být grafické znázornění výsledků, přizpůsobených k interpretaci ve vztahu k výuce cizího jazyka.

4 VYTVOŘENÝ PROGRAM

Metodami vhodnými a často používanými k určování znělosti a neznělosti jsou: STE (short time energy), ZCR (zero crossing rate) a HNR (harmonic-to-noise ratio). Všechny pracují na základě určování míry šumu v daném segmentu řečového signálu - metoda STE využívá výpočet krátkodobé energie signálu [1], ZCR určuje četnost změny polarit signálu [3], resp. průchodů rovnovážnou polohou, a HNR analyzuje poměr periodické složky signálu ke složce šumové [4]. Výpočetní algoritmy jsou rozšířeny o schopnost určení úseků ticha, což omezí vliv pauz v různých variantách daného signálu, které by jinak byly určeny jako neznělé a mohly by značně zkreslovat výsledky. Program byl vytvořen v prostředí MATLAB.

Typickou vlastností neznělé hlásky je její šumový charakter, což způsobuje vyšší položení na frekvenční ose oproti znělým hláskám periodického charakteru, zároveň vykazuje nižší energii a náhodnost průběhu, zatímco znělá hláska vykazuje velmi málo šumu a dominuje periodická složka. Úsek ticha se svou podobou nejvíce blíží nulové hodnotě. Analýza probíhá po velmi krátkých úsecích signálu, aby byla zajištěna stacionarita šumové složky. Program tedy nejprve rozčlení vstupní signál na krátkodobé úseky o délce 10 ms a ty jednotlivě zpracuje vybranou výpočetní metodou, čímž je porovná se šumem a následným prahováním je označí za znělé/neznělé či ticho. Souborem výstupních dat analýzy signálu je četnost výskytu úseků znělosti, neznělosti a ticha. Tato četnost je porovnána s četností vzorového signálu pomocí Euklidovské vzdálenosti a výsledkem celé procedury pak je hodnota podobnosti mezi analyzovaným a vzorovým signálem.

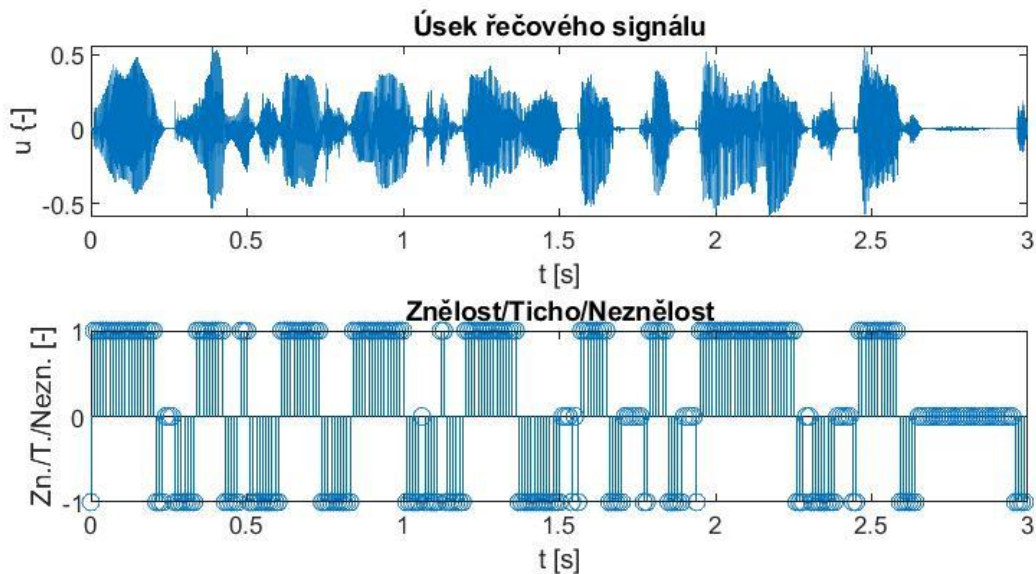


Obrázek 1: Blokové schéma programu

5 PRAKTICKÉ VYUŽITÍ A TESTOVÁNÍ PROGRAMU

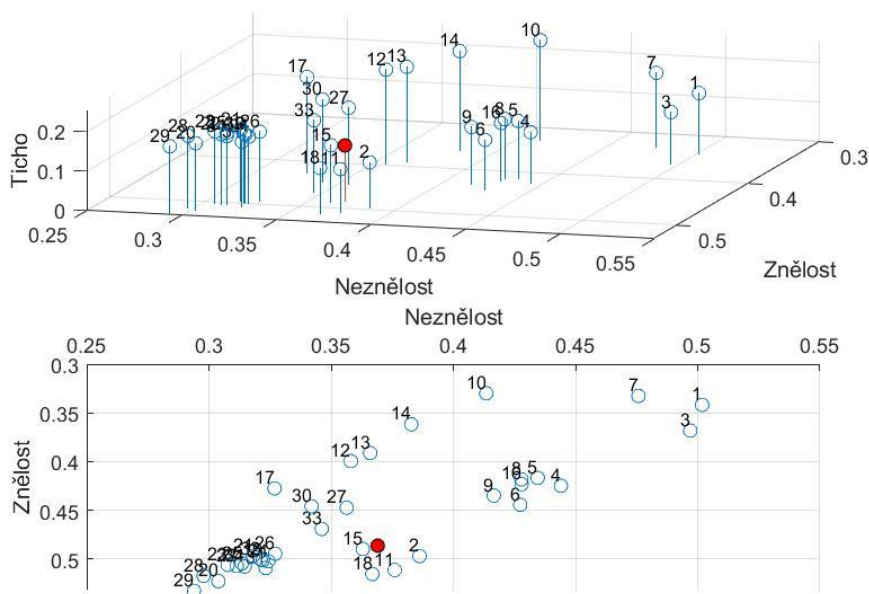
Praktický význam programu spočívá v podpoře distanční výuky cizích jazyků, a to i bez účasti vyučujícího. Příkladem může být kontrola správné výslovnosti anglické věty. Anglický jazyk je velmi zvučný a citlivý na správnou srozumitelnost, do čehož negativně vstupuje neschopnost správné artikulace znělých a neznělých pasáží jednotlivých slov. To se pro mluvčího uvyklého na slovanské jazyky může projevat např. zkracováním znělých samohlásek, přílišným zvýrazňováním některých neznělých souhlásek či zadržáváním při přechodech mezi jednotlivými hláskami.

Pro analýzu těchto vlivů byla vybrána následující anglická věta: „*Wall cavities need to be inspected by an expert to ensure walls are secure and any damaged wall-cavity insulation will also need to be removed.*“ Její záznam, namluvený rodilým mluvčím, má délku trvání cca 9 sekund. Obr. 2 zobrazuje krátkou část časového průběhu této věty společně s grafickým znázorněním pasáží vyhodnocených jako znělé/tiché/neznělé – kladná hodnota potvrzuje znělost, nulová hodnota značí ticho a záporná neznělost. Analýza byla provedena za použití výpočetní metody STE, která se ukázala být optimální pro daný účel díky efektivní schopnosti rozlišit krátkodobé úseky neznělosti a ticha. Metoda ZCR se v tomto ohledu projevila jako méně spolehlivá vlivem relativně silného šumu při porovnání nahrávek, metoda HNR se pro detekci ticha ukázala být takřka nevhodná.



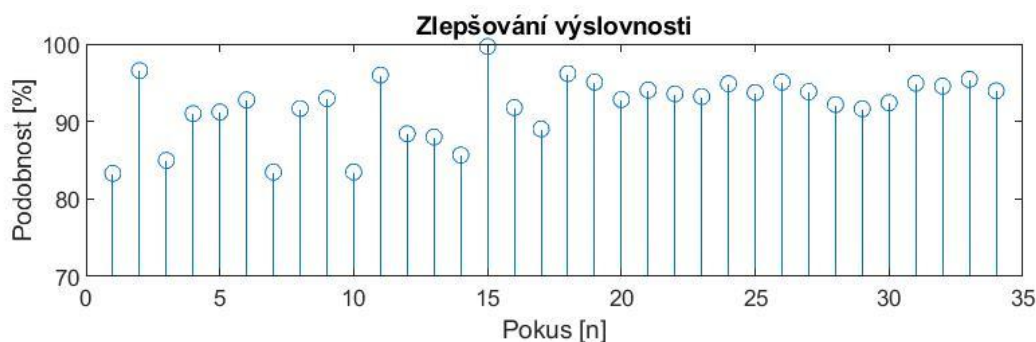
Obrázek 2: Řečový signál rodilého mluvčího

Výše uvedená nahrávka byla přehrávána českému studentovi, jenž je profesionálním učitelem angličtiny hodnocen jako průměrný až lehce nadprůměrný, za účelem zdokonalování jeho výslovnosti. Student následně nahrál 34 pokusů o správné vyslovení dané věty. Na Obr. 3 je trojrozměrný graf s osami interpretujícími procentuální četnost znělosti, neznělosti a ticha. Červený bod značí poměr četnosti pro vzorový signál namluvený rodilým mluvčím, modré body značí signály jednotlivých 34 pokusů českého studenta a jsou číslovány podle pořadí pokusu.



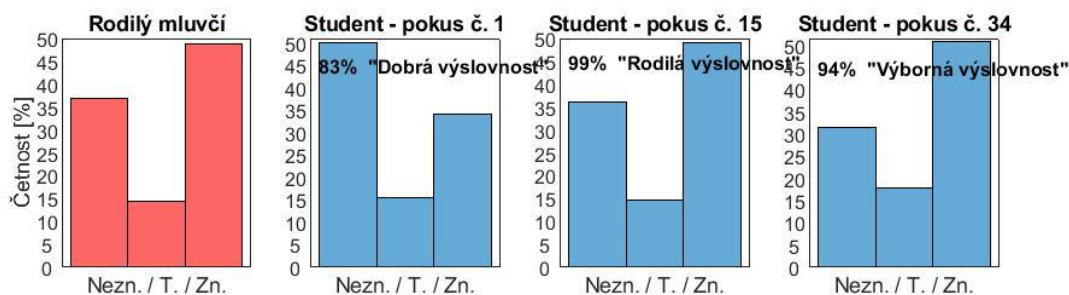
Obrázek 3: Postupné zlepšování výslovnosti anglické věty

Jak lze vidět, v prvních pokusech kvalita výslovnosti studenta výrazně kolísá, přiblížení se správné výslovnosti se děje spíše náhodně. Se zvyšujícím se počtem opakování, a postupným zdokonalováním, se poloha pokusů soustřeďuje kolem bodu nedaleko polohy vzorového signálu a výslovnost se stabilizuje, jak ukazuje Obr. 4. To, že se poslední pokusy nesoustřeďují přímo u vzorové věty, je dáno stálou několikaprocentní odlišností přízvuků. K dosažení dokonalého britského přízvuku by bylo zapotřebí např. strávit určitý čas v zemi typické tímto přízvukem či dlouhodobější nácvik.



Obrázek 4: Zlepšování výslovnosti s opakováním (procentuální podobnost s roditěm mluvčím)

Program je možné rozšířit o funkci slovního hodnocení výslovnosti na základě shody pokusu se vzorem; procentuální hranice a dané hodnocení určuje učitel. Zpětná vazba pro studenta pak bude srovnání histogramu vzoru a pokusu, procentuální podobnost a slovní hodnocení, viz Obr. 5.



Obrázek 5: Vybrané pokusy správné výslovnosti

6 ZÁVĚR

Oproti mnoha jiným oblastem studia klade studium jazyků velký důraz na mluvenou a poslechovou stránku výuky. Z tohoto důvodu má jeho distanční forma studia za podpory výpočetní techniky velký potenciál. Předložený program analyzuje řečové signály pomocí již existujících matematických metod, přidanou hodnotou je zde zejména uživateli srozumitelné grafické znázornění přiblížování se správné výslovnosti z hlediska znělosti/neznělosti. Nástroj tak nabízí možnost atraktivního zefektivnění výuky cizích jazyků. Dalším krokem ve vývoji programu by pro zpřesnění analýzy mohlo být rozšíření o schopnost identifikace jednotlivých hlásek a vyhodnocení podobnosti s hláskami vzorového signálu.

PODĚKOVÁNÍ

Tento příspěvek vznikl za podpory projektu interního grantu VUT v Brně FEKT-S-20-6361.

REFERENCE

- [1] J. Psutka, L. Müller, J. Matoušek, and V. Radová, *Mluvíme s počítačem česky*. Academia, Praha, 2006.
- [2] M. Reed and J. Levis, *The Handbook of English Pronunciation*. John Wiley & Sons, Hoboken, 2019.
- [3] R. G. Bachu, S. Kopparthi, B. Adapa, and B. D. Barkana, Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy. In *Proc. Advanced Techniques in Computing Sciences*. Springer, Dordrecht, 2010, pp. 279-282.
- [4] J. P. Teixeira, C. Oliveira, and C. Lopes, Vocal acoustic analysis-jitter, shimmer and HNR parameters, *Procedia Technology*, 2013, vol. 9, pp. 1112-1122.