VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2022 Bc. Jan Malucha

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV RADIOELEKTRONIKY

DEPARTMENT OF RADIO ELECTRONICS

PROGRAM NA PODPORU SPRÁVNÉ VÝSLOVNOSTI CIZÍHO JAZYKA

PROGRAM TO SUPPORT CORRECT PRONUNCIATION OF FOREIGN LANGUAGES

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Jan Malucha

VEDOUCÍ PRÁCE

SUPERVISOR

prof. Ing. Milan Sigmund, CSc.

BRNO 2022

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Diplomová práce

magisterský navazující studijní program **Elektronika a komunikační technologie**

Ústav radioelektroniky

Student: Bc. Jan Malucha ***ID:*** 203286

Ročník: 2 ***Akademický rok:*** 2021/22

NÁZEV TÉMATU:

Program na podporu správné výslovnosti cizího jazyka

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s problematikou zpracování řečových signálů. Vypracujte rešerši významných publikací a dostupných softwarových prostředků na podporu a kontrolu správné výslovnosti. Vyřešte vstup řečového signálu do PC v reálném čase. Naprogramujte v Matlabu nebo Pythonu a ověřte na řečovém signálu algoritmy na výpočet spektrogramů, znělosti, základního tónu, formantových kmitočtů a na určování hranic mezi fonetickými úseky řeči. Jednotlivé parametry by se měly registrovat v časové posloupnosti a pak celkově statisticky. Vytvořte autonomní program přizpůsobený na mluvenou angličtinu, který bude vyhodnocovat správnou výslovnost uživatele na základě porovnání se vzorovou výslovností. Pomocí testů vyberte vhodné porovnávací kritérium. Hodnocení by mělo vycházet z relativní podoby se vzorem nikoliv sledovat absolutní shodu, tj. imitaci. Do programu implementujte vzorovou výslovnost odborných výrazů ze dvou oblastí: teorie obvodů a zpracování signálů. Napište stručný návod na tvorbu vlastního slovníku. Na základě typického akcentu v mluvené angličtině u českých mluvčích se pokuste vytvořit přídavný modul, který je schopen detekovat rodilého českého mluvčího.

DOPORUČENÁ LITERATURA:

[1] PSUTKA, J., MÜLLER, Z., MATOUŠEK, J., RADOVÁ, V. Mluvíme s počítačem česky. Praha: Academia, 2006.

[2] KRČMOVÁ, M. Fonetika a fonologie. Brno: Masarykova univerzita Eplortál, 2008.

[3] GIANNAKOPOULOS, T., PIKRAKIS, A. Introduction to Audio Analysis: A MATLAB Approach. New York: Academic Press, 2014.

Termín zadání: 11.2.2022 ***Termín odevzdání:*** 25.5.2022

Vedoucí práce: prof. Ing. Milan Sigmund, CSc.

prof. Dr. Ing. Zbyněk Raida
předseda rady studijního programu

ABSTRAKT

Cílem diplomové práce je vývoj autonomního programu pro kontrolu a vyhodnocování výslovnosti cizího jazyka. Teoretická příprava zahrnuje pojednání o historii a významu počítačové podpory výuky jazyků; zároveň obsahuje stručný úvod do fonetiky, fonologie a problematiky výslovnosti, dále také úvod do zpracování řečových signálů se zaměřením na výpočetní metody k analýze výslovnosti. V rámci praktické části byla zpracována rešerše v současné době rozšířených a dostupných nástrojů ke kontrole výslovnosti a byla naprogramována řada algoritmů využitelných pro její analýzu. Některé z algoritmů byly uzavřeny do grafického uživatelského rozhraní jako základní prototyp funkčního programu na podporu výslovnosti. Dosažené výsledky a směr dalšího vývoje jsou diskutovány v závěru.

KLÍČOVÁ SLOVA

výslovnost, CALL, anglický jazyk, přízvuk, melodie, intonace

ABSTRACT

The aim of the diploma thesis is the development of an autonomous program for checking and evaluating the pronunciation of a foreign language. Theoretical preparation includes a treatise on the history and importance of computer-assisted language teaching; it also contains a brief introduction to phonetics, phonology and pronunciation issues, as well as an introduction to speech signal processing with a focus on computational methods for pronunciation analysis. In the practical part, a search of currently widespread and available tools for pronunciation control was processed and a number of algorithms usable for its analysis were programmed. Some of the algorithms have been enclosed in a graphical user interface as a basic prototype of a functional program to support pronunciation. The achieved results and the direction of further development are discussed in the conclusion.

KEYWORDS

pronunciation, CALL, English language, stress, melody, intonation

MALUCHA, Jan. *Program na podporu správné výslovnosti cizího jazyka*. Brno, 2021, 66 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky. Vedoucí práce: Prof.Ing. Milan Sigmund, CSc.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Program na podporu správné výslovnosti cizího jazyka“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno ....

podpis autora

PODĚKOVÁNÍ

Tímto velice děkuji vedoucímu této práce, p. prof. Ing. Milanu Sigmundovi, CSc., za nepostradatelnou asistenci při jejím vypracování.

Obsah

Seznam obrázků

Seznam tabulek

Úvod

Snaha o využití počítačů pro podporu vzdělávání se projevuje již několik desetiletí. Tento trend kontinuálně roste společně s neustále se zvětšujícím významem informačních technologií ve všech oblastech lidské činnosti, a to zejména od doby masivního rozšíření osobních počítačů v 80. letech. Současný pandemický stav a široce diskutovaný fenomén distančního vzdělávání dále akceleruje aplikaci počítačů a technologií ve vzdělávání. Konkrétní podoba a koncepce podpůrných vzdělávacích technologií je však stále kontroverzním tématem, kdy řada odborníků před tímto trendem varuje v souvislosti s narušováním pozornosti či údajnou neefektivitou. Zároveň se s časem mění obecné přístupy k metodice vzdělávání a s tím se mění i nároky na tyto nástroje. Lze tedy jednoznačně říci, že vývoj technologických prostředků k podpoře výuky nedosáhl svého maxima a má smysl mu věnovat značnou pozornost s cílem maximálního zvýšení efektivity výuky a snížení negativních dopadů.

Počítačová podpora výuky cizích jazyků je jedním z mnoha okruhů aplikujících technologie ve vzdělávání. Ačkoli klasické podpůrné prostředky ve formě např. internetových gramatických testů a doplňovacích cvičení jsou velmi rozšířené a studenty často využívané, stále je značný nedostatek těchto prostředků v oblasti zlepšování výslovnosti jazyka, případně pokulhává jejich variabilita. Může to souviset také s nízkým důrazem na výuku výslovnosti, kdy jsou na úkor ovládnutí mluvené formy jazyka prioritizovány spíše okruhy pro získání schopnosti práce se psanou formou, tzn. gramatika a slovní zásoba. Vzhledem k množství lidí využívajících mluvenou angličtinu v profesním životě však má smysl výuce výslovnosti věnovat podobnou pozornost. Technologie zde mohou zmíněný nepoměr vyvážit a nabídnout učitelům možnost zlepšovat výslovnost studentů bez nutnosti upozadění jiných oblastí výuky.

Vývoj prostředků pro podporu výslovnosti jazyků prolíná některé široce se rozvíjející obory, jako jsou strojové učení a zpracování přirozeného jazyka, zpracování signálů a elektroakustika, počítačová lingvistika či fonetika a cílem je vytvoření takového nástroje, který by studentovi co nejvíce usnadnil osvojování druhého jazyka s důrazem na mluvenou formu. Cílem této práce je takový nástroj navrhnout.

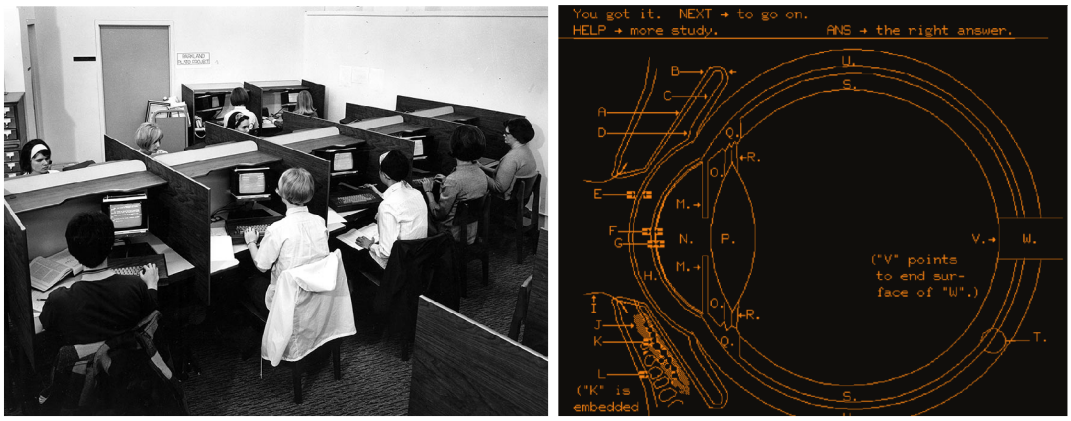
Ve své první části se práce zaměřuje na teoretické aspekty počítačové podpory výuky jazyků, přičemž je zde nastíněna historie těchto prostředků a perspektiva jejich využití (Kap. 1). Zároveň je zde zpracována rešerše současně dostupných prostředků pro kontrolu výslovnosti jazyka. Průzkum proběhl primárně formou diskuze s množstvím odborníků z oblastí lingvistiky, počítačové lingvistiky, vzdělávání a logopedie. Z větší části se jedná o odborníky z Masarykovy univerzity v Brně. Nástroje doporučené jimi či referovanými publikacemi byly v rešerši stručně popsány (Kap. 2). Dále je teoreticky popsána problematika jazyka jakožto komunikačního prostředku, jeho základní členění na vrstvy fonetickou a fonologickou a základní

vhled do aspektů výslovnosti (Kap. 3). Další kapitoly spadají pod praktickou část práce a jsou zaměřeny na zpracování signálu z hlediska výslovnosti. Jsou zde představeny a stručně popsány naprogramované funkce v prostředí MATLAB, určené pro analýzu výslovnosti v řečovém signálu; mimo algoritmy použité ve výsledném programu jsou zde také nastíněny rozpracované funkcionality pro budoucí vývoj, které z různých důvodů nebyly dokončeny nebo odladěny (Kap. 4). Představené algoritmy jsou shrnuty do kompaktního autonomního programu názorně popsaného blokovým schématem (Kap. 5). Nakonec je demonstrováno použití programu pomocí GUI (Kap. 6). Závěrem je práce shrnuta a jsou diskutovány dosažené výsledky (Kap. 7).

1 Počítačová podpora výuky jazyka

1.1 Historie CALL

Koncepce využívání počítačů pro zefektivnění výuky jazyků má poměrně dlouhou historii. Prvopočátky pokusů o reálné zavedení výpočetní techniky do vzdělávání sahají do 60. – 70. let minulého století, kdy univerzity vyvíjely software pro velké halové počítače, tzv. *mainframes*. Myšlenkou bylo ušetřit učitelům rutinní práci a drilování ve prospěch komunikačních cvičení či uvolnění více prostoru na opravování studentových chyb. Příkladem může být projekt PLATO (Programmed Logic for Automatic Teaching Operations), provozovaný Univerzitou Illinois – jednalo se o výukový počítačový systém určený k asistenci při studiu řady kurzů včetně francouzského, německého, čínského nebo latinského jazyka. Přistupovalo se k němu prostřednictvím řady terminálů, schopných na svou dobu pokročilého vykreslování grafiky. Obsahoval jednoduchá drilová cvičení nebo *text-to-speech technologii* [1], pro čínský jazyk nabízel i funkci rozpoznávání tónu [2].



Obr. 1.1: Projekt PLATO [24]

Na počátku 80. let došlo k rozšíření osobních počítačů. V této dekádě proběhla transformace prostředků počítačové výuky jazyků z drilování na formu více založenou na komunikaci a interaktivitě [3]. Tyto transformace zároveň probíhaly (a stále probíhají) ruku v ruce s novými koncepcemi a náhledy na vzdělávání obecně. Se vzrůstajícím zájmem o vývoj specializovaného software byl ustaven pojem CALL (Computer Assisted Language Learning); tento pojem byl v průběhu času definován několika způsoby, v [4] je popsán jako „jakýkoliv proces, při kterém student používá počítač a jehož výsledkem je zlepšení jazykových schopností“. Je však třeba říci, že termín CALL je nyní především chápán jako zastřešující pro celou řadu dalších okruhů, do kterých se s rozvojem technologií rozvětvil. Příklady těchto

okruhů jsou Mobile Assisted Language Learning (MALL) nebo Web Enhanced Language Learning (WELL). CALL stojí na rovnocenné úrovni s někdy používaným pojmem TELL (Technology Enhanced Language Learning). Terminologie každopádně zkratku CALL ustanovuje do popředí při reprezentaci využívání technologií pro výuku jazyků, ať už se jedná o stolní počítače, notebooky, tablety či mobilní zařízení.

Zájem o počítačem podporovanou výuku jazyků s příchodem 90. let dále stoupal. Stejně tak rostla produkce technologií určených k tomuto účelu; ta povětšinou byla v režii vývojařů v akademické sféře, šlo tedy o jednotlivce, malé týmy a skupiny nadšenců. Pro navýšení objemu financování prostřednictvím grantů se formovaly odborné asociace [5], z nichž nejvýznamnější jsou IALLT (International Association for Language Learning Technology), evropská EUROCALL (European Association for Compute Assisted Language Learning), americká CALICO (Computer Assisted Language Instruction Consortium), a zmíněnými založená organizace WorldCALL. Mimo financování projektů jsou jimi pořádány vědecké konference, některé z organizací také vydávají odborné časopisy, např. magazín FLTMAG org. IALLT nebo CALICO Journal.

Po roce 2000 přišel prudký rozmach World Wide Webu a fenoménu Web 2.0, charakterizovaného přechodem internetu z konceptu statického zdroje informací na platformu, jejímž spolutvůrcem se stal samotný uživatel (do protikladu se zde staví např. experty napsaná encyklopedie Britannica vůči Wikipedii, dynamicky vytvářené uživateli internetu). Objevily se platformy pro bloggování, sdílení multimediálního obsahu apod. Vývoj do dnešních dnů jednoznačně směřoval cestou SW aplikací do mobilních zařízení a i v České republice již ve školství zazněly názory typu „Notebooky a tablety do každé třídy“. V [2] je naznačen další vývoj CALL, který by měl jít směrem integrace technologií do vyúky jazyků tak, že technologie přejde z role „hračky“ do role podobné učebnici, tabuli, tužce atd.

Diskuze na téma technologií ve vzdělávání nejsou ojedinělé. Někteří přítomnost počítačů ve školách považují za nežádoucí a dokonce škodlivý fenomén bez reálného přínosu pro učení [6], který narušuje studentovo soustředění a neposkytuje mu plnohodnotnou možnost nad problémem přemýšlet. Mnoho autorů naopak přirovnalo vzestup informačních technologií k vynálezu knihtisku [7] a považují za nezbytné ve školství posouvat počítačovou gramotnost do popředí, v korelaci s pronikáním technologií do všech aspektů lidské činnosti, od zdravotnictví či průmyslu po zcela běžný život. Řešení problémů, které vyvstávají v tomto názorovém střetu, je možné navrhnout pouze na základě relevantního výzkumu reálné efektivity technologií v procesu vzdělávání, a to např. srovnáním studijních výsledků studentů v závislosti na používání či nepoužívání těchto technologií.

1.2 Motivace k vývoji CALL

Současná doba byla pádem železné opony v roce 1991 naprosto jednoznačně nasměrována k dominujícímu anglocentrismu a anglický jazyk se stal novou lingua franca s úlohou, kterou ve starší historii plnila latina či řečtina. Na různé úrovni jím hovoří kolem 1.5 miliardy lidí a je dnes běžně vyučován na všech stupních škol. Verbální komunikace je neoddělitelnou součástí sociálního života člověka a vzájemné porozumění je zde stěžejní. Protože se z angličtiny stal globální komunikační prostředek, má smysl směrovat úsilí do zvyšování obecné schopnosti se s její pomocí efektivně dorozumívat, a to jak ve formě psané, tak i mluvené.

Osvojování cizího jazyka je podobor aplikované lingvistiky, v zahraniční literatuře někdy označovaný jako SLA (Second Language Acquisition). Jedná se o interidisciplinární vědní obor, který zkoumá proces učení se jinému jazyku, než je jazyk mateřský. Kromě lingvistiky kombinuje také psychologii nebo pedagogiku.

Výslovnost je jedním z klíčových aspektů při osvojování cizího jazyka. Skládá se z více elementů, mezi něž patří mj. artikulace, přízvuk, intonace či plynulost řeči (popsáno dále). Podle některých zdrojů je na nácvik výslovnosti při výuce cizích jazyků obecně kladen menší důraz z důvodu působení mnoha faktorů včetně neucelených osnov výuky v této oblasti [8], kdy se nácvik omezuje na behaviorální model *listen and repeat*.

Akvizice prvního jazyka je specifický proces, velmi odlišný od osvojování druhého jazyka v pozdějším věku. V této oblasti probíhá z různých důvodů nepříliš extenzivní výzkum. Některé zdroje tvrdí, že dítě si svůj jazyk osvojuje již během těhotenství. Bylo např. zjištěno, že novorozenec pláče melodií svého mateřského jazyka – německý klesavou intonaci, francouzský naopak stoupavou, a dokáže také rozeznat samohlásky [9]. Výslovnost druhého jazyka, který se člověk učí, je silně ovlivněna návyky z mateřského jazyka. Dochází zde k interferenci prvního a druhého jazyka [10], což ústí v převádění vzorců mateřštiny, jako je výslovnost hlásek, rytmus či melodie (není zde řeč o současné akvizici dvou mateřských jazyků najednou, tzv. bilingvismus). Toto převádění je výraznější s rostoucím věkem začínajícího studenta; podle [11] je věk dokonce nejdůležitějším prediktorem v kvalitě osvojení si cizího přízvuku. Typickým problémem je výskyt nových fonémů. Výslovnost je dána souhrou všech částí mluvního ústrojí, na toto ovšem mluvčí při spontánní řeči soustředěn není a bez podvědomé znalosti tvorby těchto zvuků dochází k výraznému zkreslení řeči. Jiným problémem může být odlišnost intonace. Zatímco např. ruský, ale i český jazyk je velmi monotónní, anglický nebo čínský jazyk má velmi výraznou melodii a její variace hrají nesmírně důležitou roli při rozlišování významu slova či emocionálního ladění promluvy.

Špatná výslovnost, angl. *misarticulation*, se může lišit podle jazykové úrovně

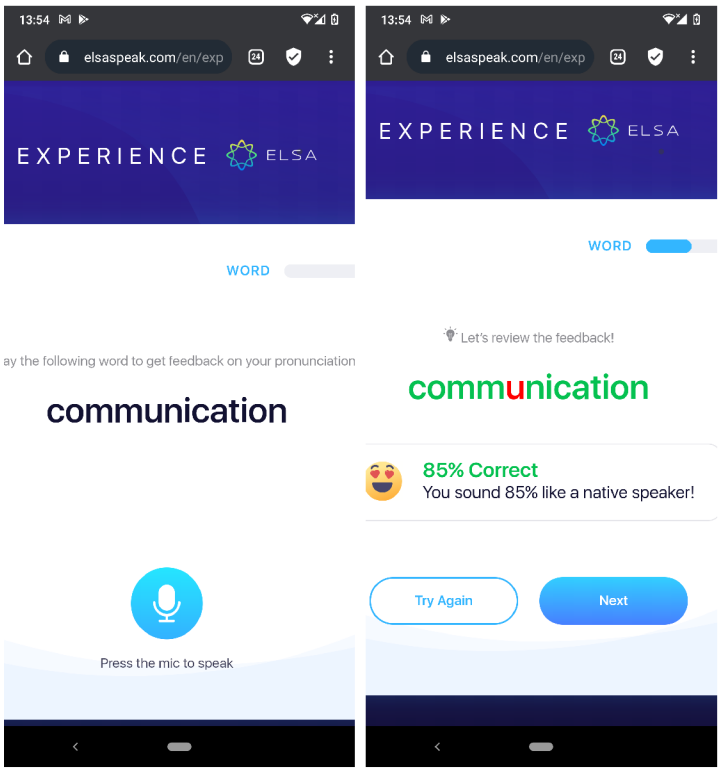
daného mluvčího, od schopnosti se dorozumět „lámaným jazykem“ až po naprostou neschopnost vyjádřit myšlenku tak, aby ji protějšek pochopil. Vyvstává zde otázka, zda má smysl se zabývat zlepšováním výslovnosti nerodilého mluvčího, který, ač neplynulou mluvou bez intonace, správné artikulace a se špatným přízvukem, je ve výsledku také schopen se dorozumět, a předat alespoň primitivním způsobem stejnou informaci, jakou by sofistikovanějším způsobem předal zkušený mluvčí. Lze odpovědět, že zde záleží na více faktorech, jako pole působnosti mluvčího (např. amatérské, profesionální), ambice a geografická lokace. Jistě nemá smysl nabízet pilování výslovnosti jazyka někomu, kdo žije mimo anglicky mluvící zemi, ve svém poli působnosti využívá maximálně čtenou angličtinu pro omezenější vyhledávání informací a pro koho je mluvená forma pouze ojedinělou záležitostí v rámci např. krátkých přátelských konverzací, a to ani ne nutně s rodilým mluvčím.

Stále však existuje určité procento nerodilých mluvčích s polem působnosti na odbornější úrovni, ať už se jedná o akademiky, lektory či experty v odborných profesích, kteří se vlivem své činnosti nevyhnutelně musí pohybovat v anglickém prostředí (komunikace se zákazníky, prezentování výsledků, výuka zahraničních studentů atd.), a přesto jejich jazyková úroveň není vysoká. Takový člověk při komunikaci s rodilým mluvčím (ale i s velmi pokročilým uživatelem) radikálně ztrácí na důvěryhodnosti. Je charakteristickým rysem rodilého mluvčího přisuzovat protějšku sofistikovanost spíše na základě formy sdělení než samotného obsahu, a dobrá výslovnost evokuje vyšší sociální prestiž. Výslovnost je totiž první zjevná věc, které si posluchač na komunikaci všímá, a to již po několika prvních slovech. Gramatika a slovní zásoba zde překvapivě nehraje tak vysokou roli - není příliš obvyklé užívat komplikovanou a méně frekventovanou slovní zásobu, a její limitace nezabraňuje mluvčímu vyjádřit myšlenku ne jedním neznámým slovem, ale několika jednoduššími. Z toho důvodu je výslovnost prvotním kritériem k označení mluvčího za dobrého či špatného uživatele angličtiny [12]. Výhodu v tomto mají mluvčí, jejichž mateřský jazyk pochází ze stejné jazykové rodiny jako jazyk studovaný, resp. jazyky jsou příbuzné. Holanďan tak bude daleko snadněji imitovat anglický přízvuk než Čech. Lze prohlásit, že výuka výslovnosti má obzvláště pro Čechy velký smysl a vzhledem k nepříliš širokému spektru příležitostí k jejímu pravidelnému procvičování má vývoj podpůrných nástrojů budoucnost, a to i z případného komerčního hlediska.

2 Rešerše dostupných CALL nástrojů

2.1 ELSA

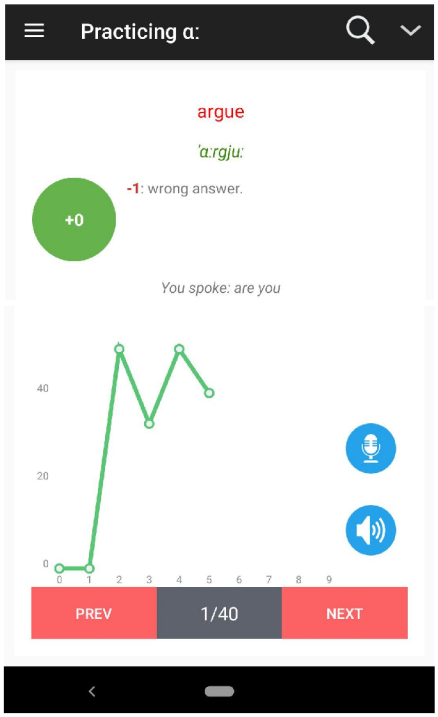
ELSA [19] je v současné době zřejmě nejpokročilejší technologie od počátku zaměřená čistě na výuku a kontrolu výslovnosti anglického jazyka. Jedná se o vietnamsko-americký projekt vyvíjený od roce 2015, založený na pokročilých algoritmech umělé inteligence. Kontrola výslovnosti probíhá na segmentální úrovni. Nástroj je určen pro co nejširší okruh uživatelů včetně naprostých začátečníků, je tedy koncipován jednoduchou formou – na obrazovce je vypsáno slovo, student jej přečte v módu aktivního nahrávání a nahrávka je následně zpracována. Výstupem, resp. zpětnou vazbou, je vyznačení nesprávně vyslovených hlásek ve slovu, jakož i možnost přehrání studentovy a vzorové výslovnosti pro vzájemné porovnání [20]. ELSA obsahuje obrovské množství výslovnostních cvičení, jakož i rozsáhlý úvod do anglických hlásek. Nevýhody nástroje ELSA podle [20] tkví jednak v přetrvávající tendenci identifikovat nesprávně vyslovené hlásky jako správné, druhak ve faktu, že hláskový inventář nástroje je omezen pouze na zvuky typické pro anglický jazyk a není tak možné lépe rozeznat zvuky produkované studentem.



Obr. 2.1: ELSA

2.2 English Phonetic Pronounciation, Listening Practice

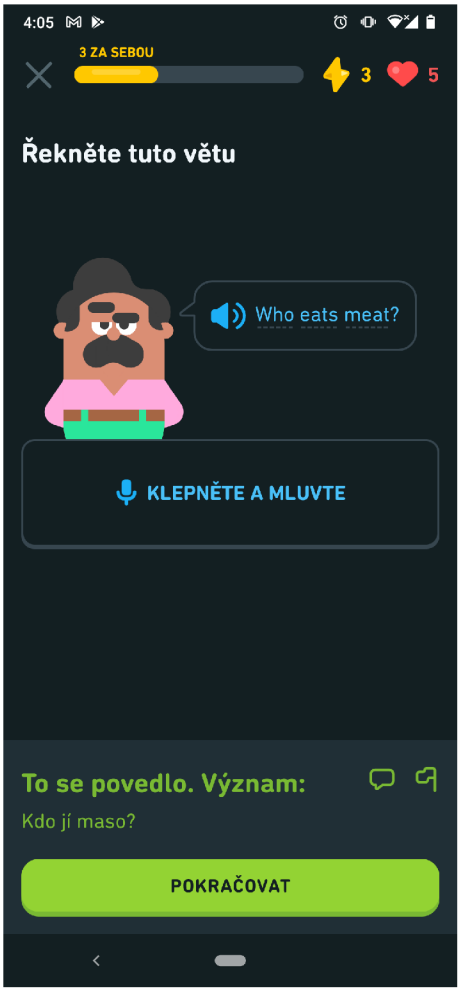
Tato aplikace byla vyvinuta firmou Stavira Vietnam a je přímo určena ke kontrole výslovnosti. Obsahuje jak úvod do výslovnosti anglických hlásek s videotutoriály, tak procvičovací mód. Koncept je zde podobný jako u nástroje ELSA – je zobrazeno anglické slovo, jež má uživatel se zapnutým mikrofonem přečíst. Po pořízení nahrávky proběhne zpracování vstupu algoritmem. Výstupem je potvrzení / zamítnutí správnosti výslovnosti, společně s grafem, kde osa x značí pořadí pokusu a hodnota na ose y roste či klesá podle výsledku pokusu. Zásadní rozdíl oproti ELSA je zde způsob zpracování dat. Zatímco ELSA je založena na proprietárním algoritmu přímo určeném ke kontrole správné výslovnosti, EPP využívá *speech-to-text* technologii spol. Google na bázi neuronových sítí; jde o stejnou technologii, jíž využívá známý Google Translate. *Speech-to-text* technologie je však navržena především pro porozumění [20] – i když je výslovnost špatná, technologie použije speciální jazykový model pro zjištění, co konkrétně mohlo být řečeno. Z toho důvodu je zpětná vazba velmi často nerelevantní. Další nevýhodou je zde naprostá absence zpětné vazby na výslovnost dílčích hlásek či instrukce ke zlepšení výslovnosti. Aplikace tak zjednodušeně řečeno pouze kontroluje, zda je uživatelův projev srozumitelný samotnému *speech-to-text* algoritmu, nikoli potenciálnímu lidskému protějšku.



Obr. 2.2: Stavira EPP

2.3 Duolingo

Duolingo je aplikace pro komplexní výuku velkého množství jazyků, a to včetně češtiny. Používá proprietární algoritmy tzv. umělé inteligence a v modulu zaměřeném na *speaking* kontroluje výslovnost technologií *speech-to-text*. Úspěšnost, resp. efektivitu výukového programu v dílčích okruzích včetně výslovnosti tvůrci zaštiťují množstvím průzkumů úspěšnosti uživatelů studujících jazyk pouze za použití aplikace Duolingo, a to pomocí jazykových testů třetích stran, např. Versant [21].
Vzhledem k použití speech-to-text technologie je kontrola výslovnosti omezena podobným způsobem, jako u předchozího nástroje. Rovněž zde chybí jakákoliv dílčí kontrola jednotlivých hlásek slova a veškerá zpětná vazba se v tomto případě omezuje pouze na přijetí / zamítnutí pokusu.



Obr. 2.3: Duolingo

2.4 Versant / Versant Pro

Versant je technologií vyvíjenou americkou společností Pearson od konce 90. let. Jde o automatický jazykový test zahrnující okruhy *Speaking*, *Listening*, *Reading* a *Writing* s funkcí automatického hodnocení, tzv. *auto-marking*. Tato technologie má školám a firmám poskytnout komplexní počítačový test anglického, španělského či arabského jazyka s odborným ohodnocením. Modul kontroly výslovnosti je obsažen ve cvičeních typu *Read aloud*, kdy je vyžadováno přečtení odstavce textu, a *Repeat*, kde se požaduje zopakovat přehranou větu. Hodnocení výslovnosti je založeno na algoritmech umělé inteligence. Zpětná vazba na studentovu výslovnost je pak zahrnuta v konečném ohodnocení celého testu, a to formou verbálního hodnocení, např. *„Candidate produces a range of meaningful sentences. Candidate speaks with adequate rhythm but with some inappropriate phrasing and pausing, and produces many vowels and consonants in a clear manner, although some sounds are non-native.“*, a bodování.
Tato technologie je určena pro testování, nejedná se tedy o výukový nástroj / aplikaci s možností souvislého procvičování výslovnosti podobným stylem, jaký nabízí např. ELSA.

Obr. 2.4: Pearson Versant

SCORE REPORT

VERSANT

Versant English Placement Test

Test Identification Number: 12345678
Test Completion Date: January 1, 2015
Test Completion Time: 1:23 PM (UTC)

OVERALL SCORE
53

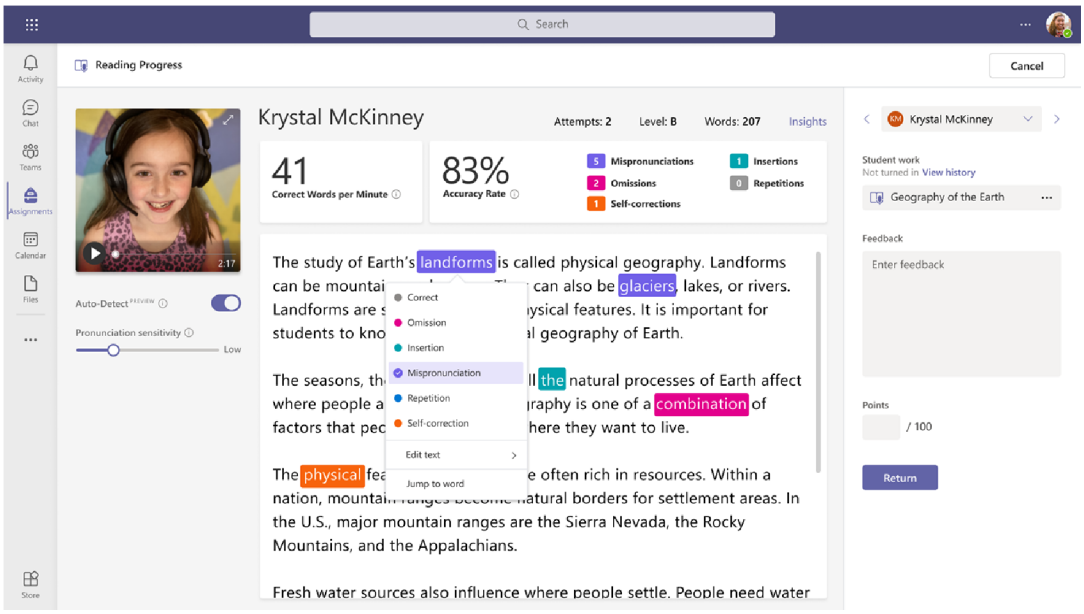
SKILL AREA	SCORE
Overall	53
Speaking	57
Listening	43
Writing	57
Reading	52

SKILL AREA	UNDERSTANDING CANDIDATE'S CAPABILITIES
Overall	Candidate can handle many utterances using a variety of words and structures, and can follow and sometimes participate in a native-paced conversation. Pronunciation is mostly intelligible; candidate can express some composite information on familiar topics to a cooperative listener. Candidate understands texts using a variety of words and structures, and given enough time can produce written texts for general purposes. Writing contains errors or inappropriate word choice, but the message is clear to a sympathetic reader.
Speaking	Candidate produces a range of meaningful sentences. Candidate speaks with adequate rhythm but with some inappropriate phrasing and pausing, and produces many vowels and consonants in a clear manner, although some sounds are non-native.
Listening	Candidate understands simple everyday conversational speech when it is spoken clearly and directed at him/her.

Obr. 2.5: Pearson Versant - vyhodnocení

2.5 MS Teams Reading Progress

Reading Progress je nástroj vyvinutý firmou Microsoft. Jde o e-learningový plugin platformy MS Teams určený pro distanční výuku, jehož účelem je asistence učiteli při kontrole plynulého čtení pomocí funkce *Auto-detect*, která detekuje chyby či anomálie ve výslovnosti. Nástroj je založen na algoritmech umělé inteligence a je určen pro analýzu celistvého bloku čteného textu, ve kterém je schopen najít jak výslovnostní chyby, tak přeřeky, vynechané části atd. Typické použití je následující – učitel do systému nahraje odstavec textu a nastaví parametr *Pronunciation Sensitivity*, který udává citlivost autodetekčního algoritmu na výslovnostní chyby. Text je zobrazen studentovi, který pořídí nahrávku čtení daného textu a uploaduje ji do systému. Nástroj poté automaticky porovná nahrávku s textem, vyhodnotí plynulost projevu, dílčí chyby ve slovech a provede klasifikaci chyb / anomálií v projevu do kategorií *Ommision*, *Insertion*, *Mispronounciation*, *Repetition* a *Self-correction*. Zpětnou vazbou je zde text s barevně vyznačenými slovy, kdy barva reprezentuje vyhodnocenou kategorii chyby. Učitel poté sám rozhoduje, co ze zpětné vazby se zobrazí studentovi.



Obr. 2.6: Reading Progress

Nástroj Reading Progress je v současné době dostupný zdarma pro všechny uživatele platformy MS Teams. Jeho nevýhodou je však netransparentní vyhodnocování výslovnostních chyb, kdy jedinou informací o chybě je informace o její přítomnosti. Funkce *Auto-detect* je tak koncipována spíše jako časoúsporná pomůcka učiteli, která jej má upozornit na možná problematická místa v projevu; učitel tato místa musí sám analyzovat a odpovídajícím způsobem vyhodnotit.

3 Jazyk a komunikace

Jazyk jakožto prostředek komunikace je detailně popsán v [32]. Pojem komunikace (lat. communicare - spojení) je z elektrotechnického hlediska povětšinou chápán jako zastřešující koncept v oborech, jejichž cílem je navázání spojení prostřednictvím komunikačního kanálu a přenos informace, přičemž snaha odborníků je primárně směřována k co největší optimalizaci tohoto procesu, a to za využití aplikace nových technologií. Prakticky se tímto zabývá celá řada odvětví, mj. radiokomunikace, telekomunikace. Základní myšlenka komunikace však zůstává zcela stejná od prvopočátků jejího vývoje – jedná se o prostou snahu o předání informace, emoce či ideje. Budeme-li se zabývat pouze komunikací jakožto formou mezilidské interakce (a pomineme např. komunikaci mezi stroji), lze zcela jednoznačně prohlásit, že v průběhu vývoje člověka sehrává nejdůležitější roli komunikace verbální. Ta se v různých formách vyvinula ve všech společnostech nezávisle na sobě a de facto umožnila vznik moderní lidské civilizace.

Verbální komunikace umožňuje sdílení informací prostřednictvím komplexního systému, jemuž říkáme jazyk. Jedná se o soubor prostředků či pravidel, které umožňují kódování a dekódování zpráv komunikujících subjektů, a to za předpokladu, že oba subjekty jsou stejného systému, resp. jazyka, znalé. Věda zabývající se jazykem se nazývá lingvistika. Praktickou realizací verbální komunikace je řeč; ta má dvě formy – akustickou a grafickou. Starší a přirozenější formou je realizace akustická, která jest realizována prostřednictvím řečového ústrojí člověka.

Akustická verbální komunikace může být rovněž doprovázena neverbálními prostředky. Těmito prostředky se zabývá vědní obor paralingvistika a jedná se o gestikulaci (pohyby těla), mimiku (výraz tváře), haptiku (doteky) či proxemiku (prostorové chování).

Jazyk, jakožto systém ke zprostředkování komunikace, využívá pro vyjádření popisovaného objektu reprezentantů zvaných znaky - ty umožňují formulaci konkrétního sdělení. Pojem znak může být chápán ve více rovinách. Příkladem nechť jsou systémy indoevropských jazyků, které reprezentují nejnižší významonosné znaky (fonémy) uspořádanými subsystémy (abeceda); zároveň lze však za znak označit i celé slovo. Znaky se zabývá vědní obor sémiotika, úzce související s lingvistikou. Uvedené obory jsou velice široké a komplexní popis jazykového systému dalece přesahuje možnosti této práce, jejímž předmětem je navíc pouze akustická stránka řeči. Pro její účely tedy stačí jazyk popsat jednoduchým vrstvovým modelem:

Pro akustickou formu verbální komunikace mají význam dvě nejnižší vrstvy – fonetická a fonologická.

lexikologie	resistance
morfologie	resist-ance
fonologie	r-e-s-i-s-t-a-n-ce
fonetika	ɹ-ɪ-z-ɪ-s-t-ə-n-s

Tab. 3.1: Nejnižší vrstvy jazyka

3.1 Fonetika

Fonetika je mezivědní obor kombinující lingvistiku, akustiku a anatomii. Její poznatky jsou pro samotnou lingvistiku výchozím bodem, jde tedy o její základní disciplínu. Předmětem zkoumání fonetiky je čistě zvuková stránka jazyka, tzn. nejsou brány v potaz žádné sémiologické aspekty. Fonetický řetězec tvoří artikulace (tvorba zvuku), akustika (přenos zvuku) a percepce (vnímání zvuku). Fundamentální fonetickou jednotkou jest fón neboli hláska. Fonetiku lze studovat univerzálně i odděleně, tzn. úzce zaměřenou na jeden konkrétní jazyk.

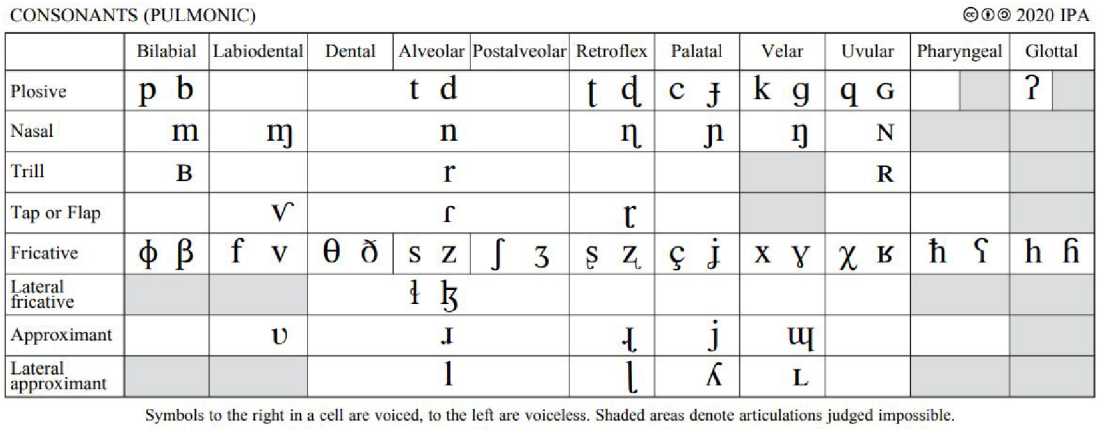
Pro vyjádření jazyka psanou formou je nezbytné osvojit si soubor znaků daného jazyka – tato množina je obecně zvána abeceda. Paralelně k psané formě, pro vyjádření jazyka formou mluvenou je nutné znát soubor zvuků daného jazyka. Prvky této množiny se nazývají hlásky, angl. phonemes. Tento překlad může do terminologie vnášet zmatky – v české terminologii je hláska pojmem fonetiky (akustický pohled na řeč), zatímco foném spadá do oblasti fonologie (sémantický pohled na řeč).

Na přelomu 19. a 20. století byl Mezinárodní fonetickou asociací ustaven systém IPA (International Phonetic Alphabet), definující množinu symbolů, kterou lingvisté užívají pro popis zvuků mluveného jazyka. Cílem bylo vytvoření zcela univerzálního systému, který by postihoval veškeré mluvené jazyky; z toho důvodu byla IPA tabulka pravidelně aktualizována a doplňována o nové zvuky. Každému konkrétnímu zvuku je zde přiřazen jediný exaktní symbol. Tento lingvistický nástroj je narozdíl od standardní abecedy zcela nezávislý dialektu jazyka.

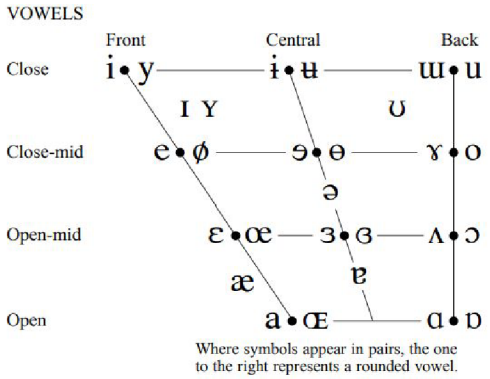
IPA byla vytvořena pouze pro popis lidské řeči. Neobsahuje tedy veškeré člověkem produkovatelné zvuky. Zároveň je třeba říci, že zdaleka ne všechny země tento systém adaptovaly; v českých bilingvních slovnících jsou používány pouze ty IPA symboly, které reprezentují zvuky nevyskytující se v češtině. Význam IPA pro český jazyk oproti např. angličtině není tak velký, vlastností českého jazyka je, že slova se vyslovují stejně, jako se píší.

Hlásky se klasifikují několika způsoby. Jedním z nejznámějších je rozdělení fonémů na samohlásky (vokály) a souhlásky (konsonanty) podle způsobu jejich tvorby ve vokálním traktu – samohlásky jsou buzeny proudem vzduchu z plic a modulovány dutinami v traktu, charakterizuje je přítomnost rezonančních kmitočtů zvaných for-

mantové kmitočty. Vznikají rezonancemi v dutinách vokálního traktu, každý formant odpovídá právě jedné rezonanci. Jejich vzájemným poměrem je možné jednotlivé vokalické hlásky do jisté míry rozlišovat. Samohlásky mají oproti tomu charakter šumu vznikající kladením překážek proudícímu vzduchu v traktu, např. jazyku.



Obr. 3.1: Konsonanty podle IPA



Obr. 3.2: Vokály podle IPA

3.2 Fonologie

Fonologie je lingvistická disciplína, která již kromě akustické stránky bere v potaz i sémiologický aspekt jednotlivých zvuků. Na znaky nahlíží a zkoumá je z hlediska jejich významonosné funkce. Základní jednotkou je zde foném. Ve vztahu k fonetice lze prohlásit, že hláska jest akustickou realizací fonému.

Ve psané formě jazyka je základní jednotkou grafém (písmeno), přičemž každý foném je v textu označen svým odpovídajícím grafémem, případně tzv. spřežkou. Fonémy a grafémy tedy do jisté míry existují paralelně (u jazyků bez výskytu spřežek

ve psané formě jsou pak zcela paralelní). Samotný foném ovšem není jednoznačný – pro každý foném totiž může existovat podmnožina jeho hláskových variant, tzv. alofonů. Např. foném /n/ s příslušným grafémem „n“ se vyslovuje jinak ve slovech lano a banka. Ve druhém případě dochází k mírné změně ve tvarování hlásky vlivem podvědomého zjednodušování výslovnosti a proti /n/ vzniká varianta (alofon) /ŋ/. Jednou z výhod systému IPA je, že s alofony počítá a přiřazuje jim odpovídající znaky (viz zmíněný /ŋ/).

Hranice mezi fonetikou a fonologií, resp. hláskou a fonémem, může působit poněkud nejednoznačně. Pro názorné vysvětlení rozdílu mezi hláskou a fonémem lze tedy za příklad vzít uvedená slova „lano“ a „banka“ – z fonologického hlediska obě slova obsahují foném /n/; z hlediska fonetického první slovo obsahuje hlásku /n/ a druhé hlásku /ŋ/ a tím pádem se slova liší. Laicky řečeno - sepíšeme-li všechny grafémy a spřežky daného jazyka, za fonémy můžeme označit množiny možných akustických realizací jednotlivých grafémů a spřežek. Foném se tedy jeví jako poměrně abstraktní pojem.

Jak bylo řečeno, lze hlásky dělit na konsonanty a vokály. Prvky množiny vokálů v českém jazyce jsou alofony fonémů a, e, i, o, u, jejich dlouhé varianty a dvojhlásky au, ou, eu. České konsonanty jsou alofony fonémů p, b, t, d, ť, ď, k, g, f, v, s, z, c, š, ž, č, dž, ř, x, h, m, n, ň, l, r, j.

3.3 Výslovnost

Pojmu správná výslovnost se věnují např. [31] a [32]. Podobně jako existují pravidla pravopisu, existují také kodifikace norem výlučně pro akustickou realizaci řeči, resp. pro správnou výslovnost. Správné výslovnosti lze dosáhnout aplikací poznatků fonetiky. Je předmětem zkoumání vědního oboru ortoepie. Výslovnostní normy se začaly formovat až se vznikem sféry kultivovaného projevu, kde byla správnost výslovnosti udávána novými komunitami jejích nositelů. Toto formování bylo dále akcelerováno s rozšiřováním sdělovací techniky (do té doby byly běžné výrazné regionální rozdíly ve stylu řeči). Normy jsou udány ve specializovaných dokumentech věnující se výslovnosti spisovné češtiny.

Při analýze výslovnosti češtiny je někdy nutné postupovat s přihlédnutím ke psané formě řeči. Vztah těchto dvou forem se může různit pro jiné jazyky. Český jazyk je v řadě případů čten stejně jako psán; jeho psaná forma (latinské písmo) ovšem obsahuje velké množství diakritických znaků, oproti např. cyrilici, která byla vytvořena přímo pro zvuky typické slovanským jazykům.

Výslovnost, resp. chyby ve výslovnosti, zpravidla kontrolujeme ze dvou hledisek. Buďto sledujeme výslovnost fonémů či jejich spojení, nebo zkoumáme zvukové vlastnosti souvislé řeči; v prvním případě mluvíme o hledisku fonémickém, ve druhém

případě o prozódickém.

Fonémická stránka sleduje výslovnost jednotlivých hlásek, případně jejich spojení. Český jazyk má celou řadu pravidel určujících chování hlásek v různých spojeních. Příkladem může být změna místa tvoření hlásky v řečovém ústrojí, čímž vznikne alofon (výše zmíněná výslovnost [baŋka] místo [banka] u slova „banka“), dále např. artikulační splývání (splynutí /ts/ slova „studentský“ - [studentskí] na [studenckí]), pravidlo pro vypouštění a vkládání hlásek ([jsem] má častou výslovnost [sem], případně vkládání hlásky /u/ u slova [osm]) a především fenomén spodoby znělosti, kdy párové znělé a neznělé hlásky ovlivňují znělost některých hláskových spojení (např. tužka [tuška], sbor [zbor]).

Prozódie označuje vlastnosti zvukové realizace řeči na suprasegmentální úrovni. Předmětem zkoumání zde nejsou jednotlivé hlásky, jako je tomu u fonémického pohledu; řeč vnímáme jako kontinuální proud zvuku, resp. řetězec vokalizovaných a šumových segmentů, a kontrolují se modulace tohoto proudu. Jako nejnižší suprasegmentální jednotka se zde chápe slabika. Členení slova na slabiky není pro všechny jazyky univerzální. Zatímco český jazyk člení slovo podle slabikotvorných jader (slabikotvorné znělé souhlásky či samohlásky), anglický jazyk i v jednoslabičném slově připouští i několik hlásek, které by v češtině byly označeny za slabikotvorné. Příkladem může být jméno Ch**a**rl**e**s, které angličané vnímají jako jednoslabičné. Českému mluvčímu se díky zvýrazněným slabikotvorným samohláskám ovšem jeví jako dvouslabičné [32].

Modulace řečového proudu v časovém průběhu se provádí prostřednictvím tzv. modulačních faktorů; např. faktor energie udává přízvuk a rytmus promluvy, melodie řeči je dána průběhem základního tónu atd. Tyto atributy bývají pro jazyky typické – jejich zkoumáním pak můžeme analyzovat správnost výslovnosti z průběhu řečového signálu. V praxi jsou modulační faktory i fonémické vlastnosti charakterizující řeč často označovány pojmem příznak.

Cílem této práce bude analýza správnosti výslovnosti z hledisek melodie, přízvuku a trvání suprasegmentálních jednotek.

3.3.1 Přízvuk

Přízvuk je z lingvistického hlediska víceznačný pojem. Nejčastěji chápán jako zvýraznění určité slabiky zvýšením její síly; tato definice však de facto nemusí být úplná, protože zvýšení síly je zpravidla doprovázeno také dalšími změnami, jako je zvýšení hlasu či změna artikulace, a ty dále přízvuk dotvářejí. Pojem přízvuk pak může být chápán jako komplexní parametr řeči; pro účely této však bude dále chápán jako pouhé zvýšení síly slabiky, přičemž uvedený komplexní parametr budiž označen pojmem intonace.

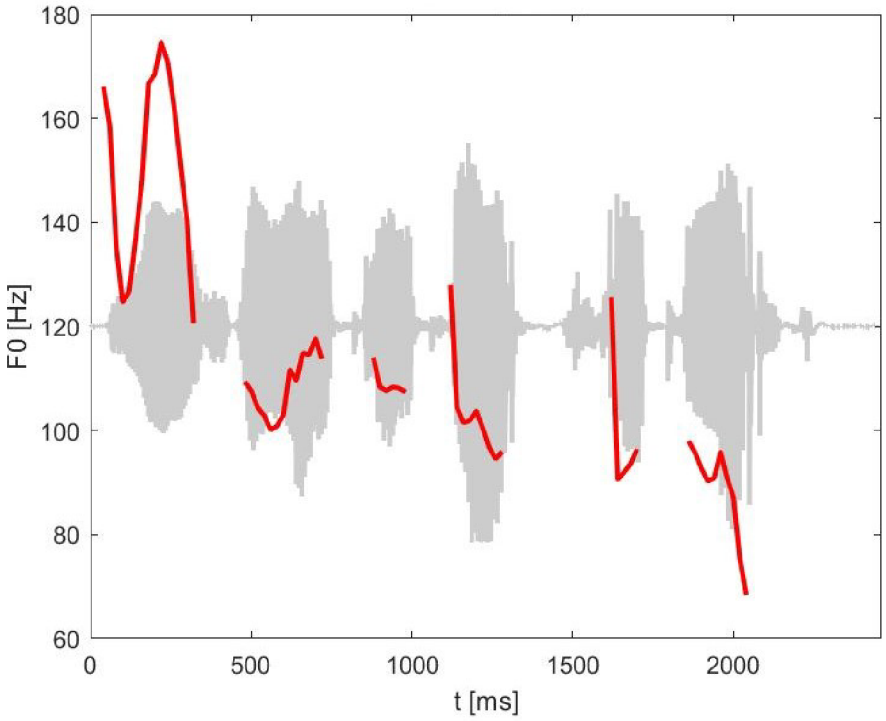
Přízvuk je zpravidla možné klást nejen na pouhou slabiku. Na suprasegmentální úrovni hraje významnou roli také přízvuk větný, jenž může měnit význam či podtext celého sdělení. Pro různé jazyky má přízvuk obecně také odlišná pravidla. Zatímco čeština patří společně s např. polštinou či italštinou mezi jazyky s pevným přízvukem (český jazyk klade přízvuk na první slabiku, polština a italština na předposlední), anglický jazyk je typický přízvukem pohyblivým. Vyskytuje se v něm velké množství dvojic slov, jejichž význam se s položením přízvuku na odlišnou slabiku zcela změní; toto se často projevuje u podstatných jmen a sloves, např. anglické slovo "*record*"bude s přízvukem na první slabice vyjadřovat podstatné jméno „nahrávka“, zatímco přízvuk na druhé slabice význam změní na sloveso „nahrát“. Lze tedy prohlásit, že správnost výslovnosti z hlediska přízvuku je daleko významnější, než např. v českém jazyce. V češtině lze přízvuk často zcela zanedbat bez velkého rizika neporozumění mezi mluvčími. Zajímavostí je kladení přízvuku na předposlední slabiku místo první u obyvatelů Ostravska, jejichž styl mluvy je silně ovlivněn geografickou blízkostí a kulturní provázaností s mluvčími polského jazyka.

3.3.2 Melodie

Jak bylo řečeno dříve, hláska může podle způsobu svého tvoření ve vokálním traktu mít šumový nebo vokalizovaný charakter. Zapojením hlasivek do produkce řeči se objevuje základní tón, vyjádřitelný také jako fundamentální frekvence F0 [Hz] nebo základní perioda řeči T0 [ms]. Měřením základního tónu je možné získat řadu informací o promluvě a řečníkovi, včetně jeho emocionálního stavu či regionálního dialektu. Tón v průběhu řeči není konstantní – zákadní perioda T0 se s časem prodlužuje a zkracuje. Důsledkem je melodie, tzn. změna tónu řeči s časem. Měřítkem tohoto jevu je parametr jitter, popisující vzájemnou časovou odlišnost jednotlivých krátkodobých period. Melodie se pro jednotlivé jazyky velmi liší. Bylo již zmíněno, že ruský jazyk nevykazuje časté kolísání melodie a jeví se méně emocionální, než např. anglický jazyk, kde monotónnost může do jisté míry měnit význam promluvy – např. výraz “*Oh, really!*” s výrazně kolísavou melodií je jednoznačně vnímán jako projev překvapení, stejný výraz vyslovený monotónně však jasně evokuje znuděnost. Nevýrazná melodie může dokonce vyvolávat dojem jisté arogance [12]. Melodie a přízvuk (tzn. intonace) jsou tedy zcela zásadním kritériem pro rozeznání významu promluvy z hlediska emocionálního podtextu (sarkasmus, hněv či rozhořčení atd.). Existují také jazyky, kde je intonace přímým prostředkem pro určení samotného významu slova, podobně jako fonetická stavba. Typické to je např. pro čínský jazyk, jehož výslovnostní gramatika se opírá o soubor tří tónů a jejich vzájemné kombinování pro rozlišení významu stejně fonematicky znějících slov.

Standardní hlasový rozsah můžů bývá vymezován mezi 50 a cca 250 Hz. U žen

je tento rozsah vyšší, relativně vysokých hodnot horní meze mohou dosahovat např. operní pěvkyně.

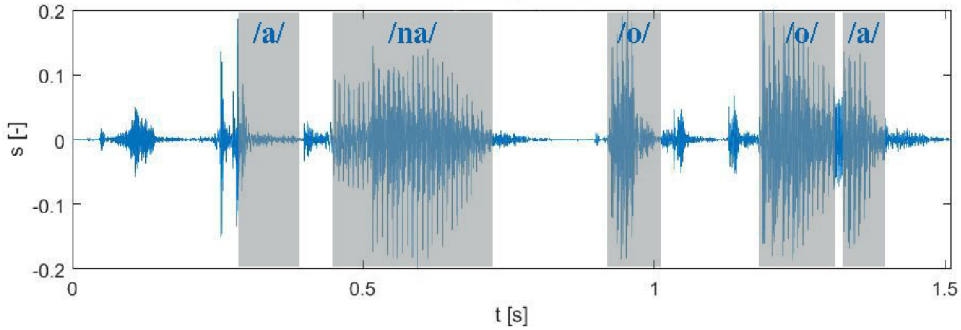


Obr. 3.3: Vizualizace melodie

3.3.3 Znělost

Tento parametr dává informaci čistě o samotné přítomnosti či nepřítomnosti základního tónu hlásky, generovaného kvaziperiodickým kmitáním hlasivek. Fonémy lze obecně klasifikovat jako znělé či neznělé. Na řečový signál reprezentující slovo potom pohlížíme jako na posloupnost znělých či neznělých fonémů. Pauzy se také mohou vyskytovat mezi fonémy, tzn. uprostřed slov. Za příklad můžeme vzít průběh českého sousloví "*špatná podkova*", který zobrazuje následující graf. Znělé fonémy jsou v grafu vyznačeny šedou barvou, neznělé fonémy a tiché pauzy jsou neoznačeny. Lze zde mimo jiné pozorovat typickou krátkou pauzu mezi fonémy /d/ a /k/ ve slově podkova. Tato pauza je zapříčiněna přechodem mezi pozicemi mluvního ústrojí – pozice se samozřejmě mění spojitě, což způsobuje přítomnost přechodových jevů.

Je třeba říci, že anglický jazyk je z hlediska znělosti obzvláště citlivý na správnou výslovnost fonémů; při špatné výslovnosti se totiž může zcela změnit význam slova, viz např. slova "*bat*"(netopýr) a *bad* (špatný). Český jazyk na tento jev naopak náchylný není, např. slova let a led mohou být beze ztráty porozumění vyslovena



Obr. 3.4: Vizualizace znělosti

stejně. Z tohoto důvodu má u českých studentů velký význam hlídat čistotu výslovnosti každého anglického fonému a rozlišení jeho znělosti či neznělosti; to platí obzvláště u těch fonémů a jejich alofonů, které se v češtině nevyskytují, a jenž tudíž vyžadují zvýšenou pozornost při nácviku správné výslovnosti [26]. Za nejproblematičtější fonémy se u českých mluvčí jeví foném /th/ s alofony /ð/ a /θ/ [27].

Citlivost anglického jazyka na rozlišení znělosti fonémů lze demonstrovat na tzv. minimálních párech. Jedná se o fonologický fenomén vyskytující se v jazycích mluvených i znakových [13]. Popisuje páry slov odlišného významu, které se liší pouze jediným fonémem. Rozdíl může spočívat v délce fonému, může jít o kontrast jeho znělé / neznělé varianty, různé alofony či zcela jinou hlásku. Pozice této hlásky ve slově může být na začátku, uprostřed i na konci. Následující tabulka uvádí příklady minimálních párů pro rozdíl fonému (jeho znělá / neznělá varianta):

Neznělé	**Znělé**
pin /pɪn/	bin /bɪn/
rot /ɾot/	lot /lot/
seal /sil/	zeal /zil/
hat /hæt/	head /hæd/

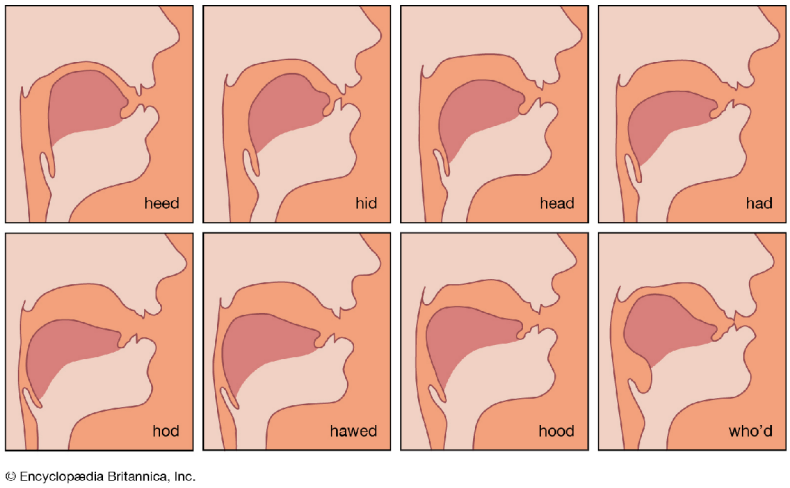
Tab. 3.2: Minimální páry

3.3.4 Trvání

Trvání jednotlivých segmentů (slabik či fonémů) v daném slově má zásadní vliv na celkový rytmus promluvy. Zároveň se velmi úzce váže k celkové intonaci slova. Proudloužení, ale i zkrácení slabiky ji může přízvučně zdůrazňovat.

3.3.5 Formantové kmitočty

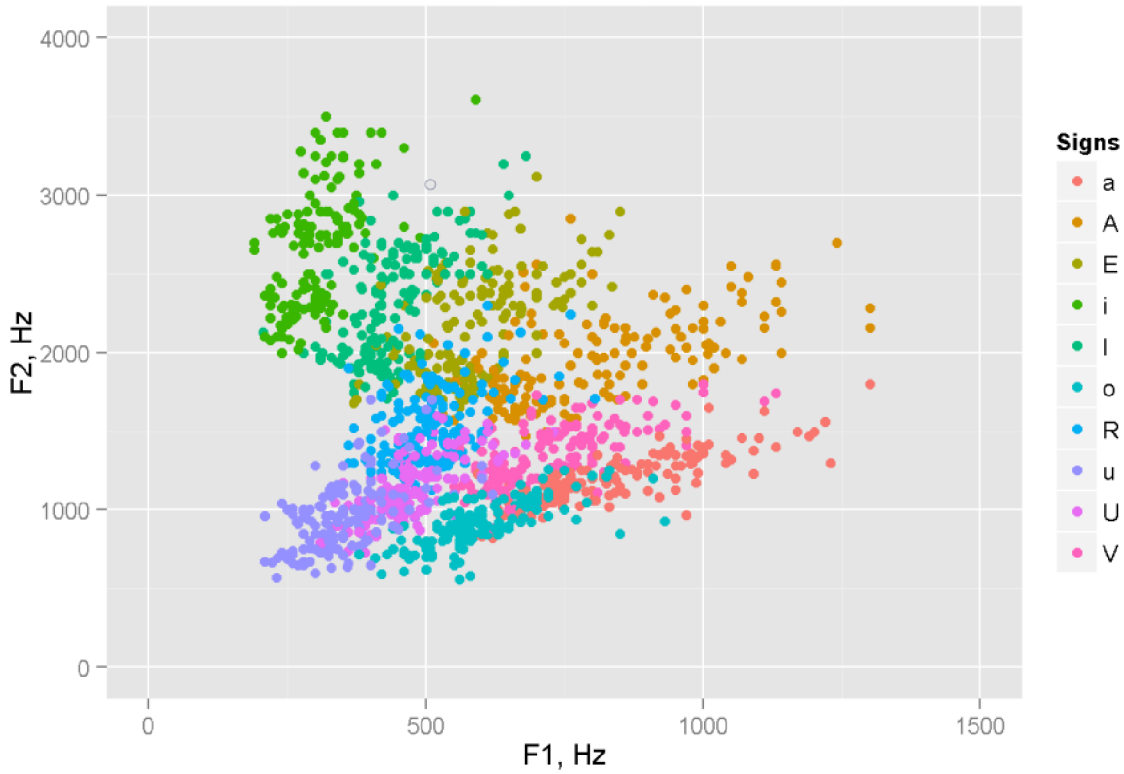
Řečový signál je frekvenčně bohatý. Základní frekvence F0, tvořená hlasivkami, je při tvorbě řeči filtrována řečovým ústrojím. Tento filtr je unikátní pro každého člověka a jeho charakteristika je dána fyziologickou podobou řečového ústrojí. Nachází se zde celá řada rezonančních dutin, např. dutina ústní, dutina nosní, hltan či samotná lebka. Stav některých dutin se dynamicky mění podle vyslovované hlásky, např. změnou pozice jazyka, otevření rtů atd. (pro každý generovaný foném je možné změřit konkrétní přenosovou funkci vokálního traktu [14]). Společně s oscilací hlasivek (produkující F0) tak v dutinách dochází ke vzniku dalších významných rezonančních kmitočtů v podobě příčných, podélných i vázaných módů, jež se nazývají formantové kmitočty, označované jako F1, F2, F3... Fn. Při vzniku formantů hrají nejdůležitější roli dutina ústní a hltan, jež jsou asociovány s F1 (hltan) a F2 (dutina ústní) [15].



Obr. 3.5: Postavení artikulátorů pro různé hlásky [30]

Každá hláska může být popsána pomocí pozic formantů na kmitočtové ose. Měřením pozic formantů a ustavením jejich referenčních hodnot pro české hlásky se v roce 2011 zabývali [16], pro anglický jazyk např. [17]. Velmi výhodné je, že fonémy mohou být rozlišeny pouze na základě pozic F1 a F2 (pro zpřesnění výsledků i F3), což zároveň umožňuje zobrazení fonémů ve dvourozměrném prostoru o jedné veličině, kterou je frekvence [Hz]:

Nevýhodou popisu hlásek formantovými kmitočty je značná změna pozice formantů podle mluvčího. Podle [18] se frekvence vokálu /i/ mohou pro jednotlivé osoby lišit v rozmezí od 190 Hz do 590 Hz pro F1 a od 2000 Hz do 3610 Hz pro F2. Zároveň však bylo autory dokázáno, že poměr F1 a F2 vždy zůstává stejný (u vokálu /i/ v hodnotě 1:9), a navíc že poměry formantů jsou nezávislé na pohlaví a věku



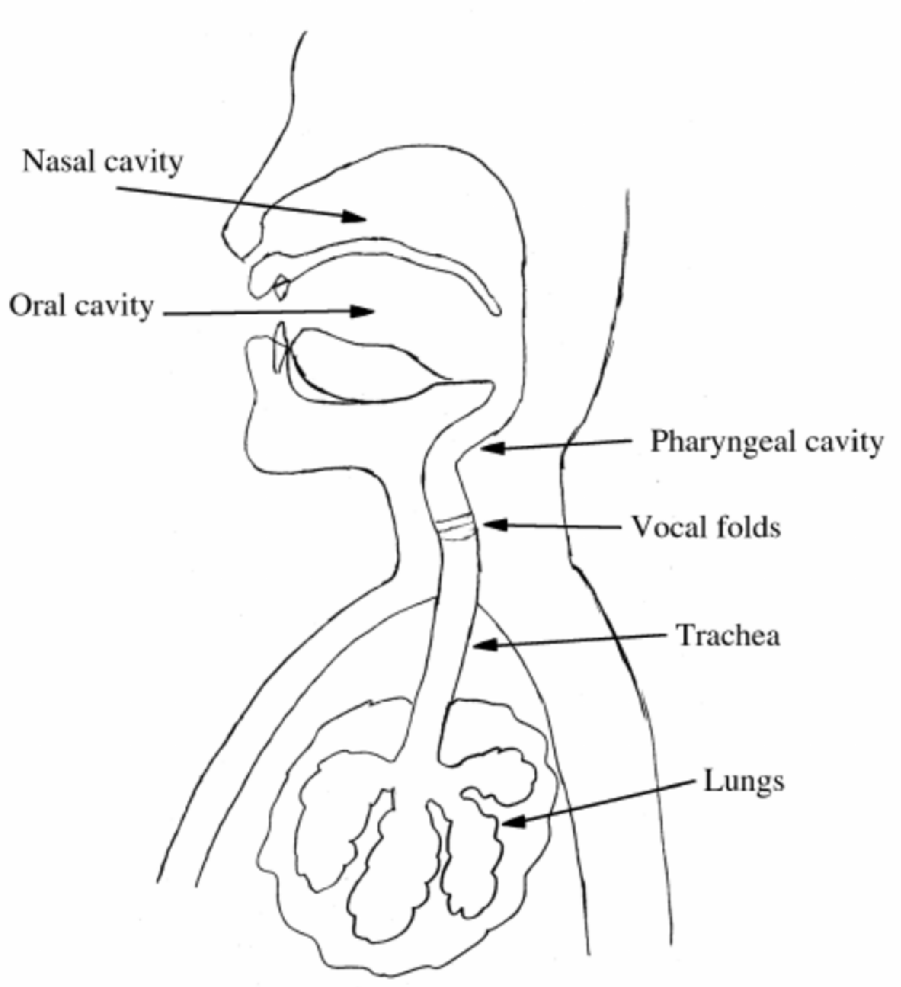
Obr. 3.6: Anglické fonémy (muži i ženy) [17]

mluvčích, což jsou dva parametry silně ovlivňující délku vokálního traktu mluvčího a jeho F0.

Je třeba říci, že analýza formantových kmitočtů postrádá smysl u neznělých hlásek, jež mají charakter šumu a jejichž spektrum nevykazuje rezonance.

4 Řečový signál a jeho zpracování

Řečový signál je předmětem oboru zpracování řečových signálů. Generujícím systémem je lidský vokální trakt neboli řečové / mluvní ústrojí, sestávající z respiračního a fonačního ústrojí. Výstupem systému je kmitání vzduchu, jenž lze zachytit kmitáním membrány ucha, případně sejmout membránou mikrofonu a převést na elektrický signál. Ten bývá zpravidla před zpracováním navzorkován a tedy převeden na signál diskrétní.

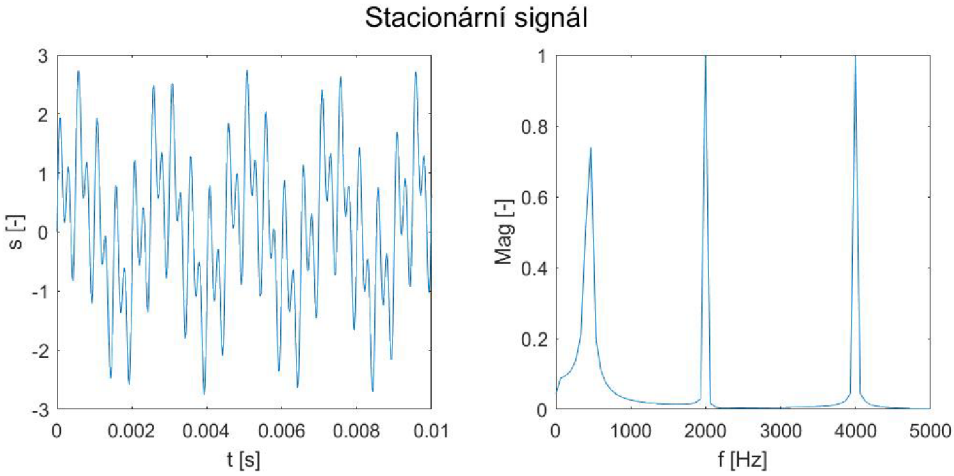


Obr. 4.1: Mluvní ústrojí člověka [29]

Základními zkoumanými vlastnostmi řečového signálu z hlediska teorie signálů jsou stacionarita, ergodicita a časová invariance. Řečový signál je považován za ergodický a časově invariantní, v závislosti na neseném obsahu signálů však může být stacionární či nestacionární. To je spojeno s nutností segmentace signálu.

4.1 Stacionarita a segmentace

Stacionarita je pro zpracování signálu nesmírně významným kritériem. Signál je stacionární jen a pouze tehdy, když se jeho frekvenční složky nemění s časem. Zatímco stacionární signál je možné popsat jeho spektrem, pro popis nestacionárních signálů je nezbytné sledovat změnu spektra v závislosti na čase, tzn. relevantní informaci nenese spektrum získané diskrétní Fourierovou transformací signálu jako celku, ale spektrogram, tzn. spektrum v závislosti na čase. Typickým příkladem stacionárního signálu může být funkce o třech harmonických komponentech, kdy $f_1 = 440$ Hz, $f_2 = 2000$ Hz, $f_3 = 4000$ Hz:



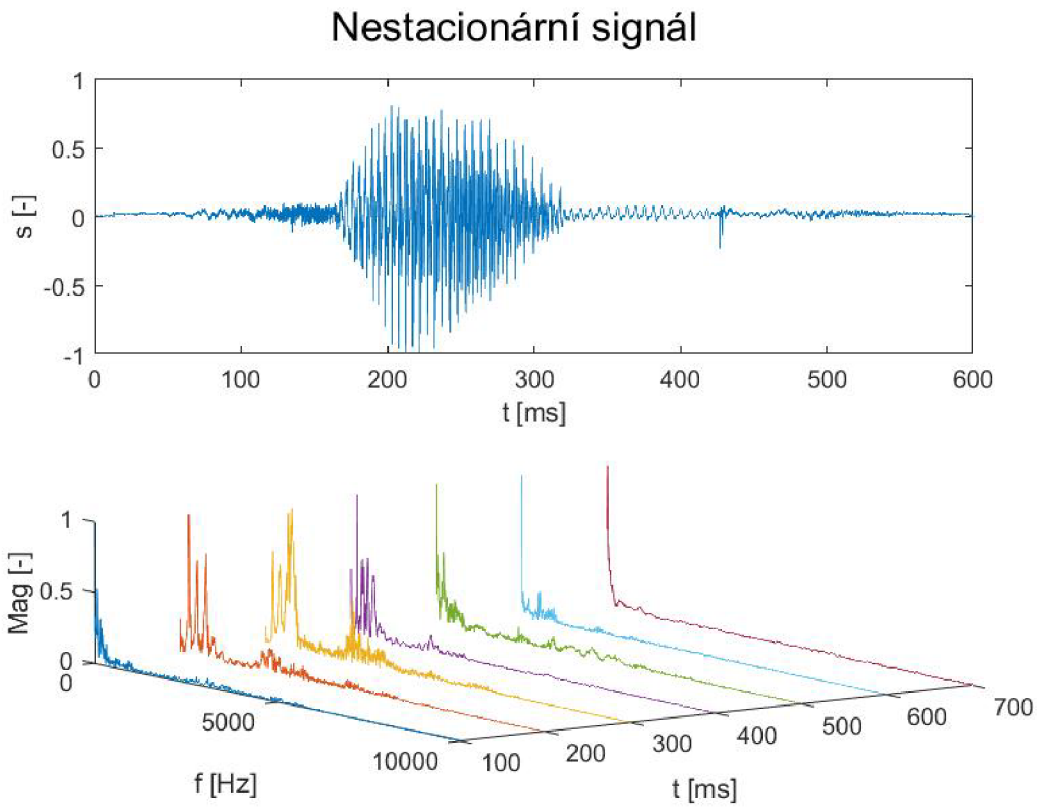
Obr. 4.2: Příklad stacionárního signálu - součet funkcí sin

Za stacionární signál konvenčně považujeme také řečový signál nesoucí např. pouze průběh hlásky /a/. Ačkoli při tvorbě hlásky dochází v řečovém ústrojí k přechodovým jevům a na signálu se to projeví fází náběhu a doznívání (signál narůstá z minima nebo do něj naopak konverguje), tyto části mohou být zanedbány a za užitečnou brána jen ustálená část mezi nimi. Hlásku /a/ potom můžeme popsat exaktním spektrálním průběhem.

Pokud ovšem pracujeme s řečovým signálem nesoucí slovo o několika fonémech, případně celou větu, je signál jednoznačně nestacionární. Spektrální komponenty se v jednotlivých úsecích signálu mění a tudíž spektrum takového signálu nenese relevantní informaci o signálu. Je potom nutné jej popsat funkcí času a kmitočtu pomocí spektrogramu. Příkladem [1] nechť jest slovo "*head*"na obr.4.3.

Zpracování řečových signálů nestacionaritu signálu typicky řeší pomocí oken. Za stacionární signál je obecně brán krátkodobý úsek signálu nestacionárního (tzv. krátkodobý segment), trvající v rozmezí <10, 30> ms. Při zpracování je tak ze

[1] Segmentace na 100 ms je stále nedostačující pro zajištění stacionarity, obrázek je ilustrační.



Obr. 4.3: Příklad nestacionárního signálu - slovo

signálu extrahována taková část násobením určitým oknem. Existuje více typů oken, pro účely práce je využito obdélníkové okno.

4.2 Řečový signál z hlediska výslovnosti

Jak bylo řečeno v kap.3.3, výslovnost můžeme sledovat z hlediska fonémického a prozódického. Vycházíme z nahrávky řeči ve formě řečového signálu – reprezentace vlnění prostředí v závislosti na průběhu času. Z fonémického hlediska je pro nás zajímavá jakost jednotlivých hlásek, tedy kvalita jejich vyslovení či jejich trvání. V tomto ohledu tedy je nutné řečový signál rozčlenit na základní fonematické segmenty a jejich parametry dále jednotlivě analyzovat nástroji obecného zpracování signálu. Z hlediska prozódického toto členění není kritické – zde jsou důležité zejména modulační faktory v průběhu času, udávající lingvistické vlastnosti promluvy; tyto faktory také úzce souvisí s obecnými vlastnostmi elektrických signálů.

Vlastnosti, které jsou pro řečový signál stěžejní, jsou melodie, přízvuk, trvání a artikulace zvuku. Tyto čtyři parametry zcela zásadně ovlivňují celkový zvuk jazyka. Z prozódického hlediska platí, že parametr melodie je v průběhu zvukového proudu

tvořen modulačním faktorem změny základní frekvence a přízvuk je závislý na modulačním faktoru energie či síly zvuku. Trvání jednotlivých segmentů (fonémů) je řazen do fonémické kategorie vlastností řeči a je dán čistě relativním časovým trváním segmentu vůči ostatním segmentům ve slově; do této kategorie spadá rovněž artikulace, která závisí na momentálním spektrálním složení daného segmentu.

Publikace [25] bere v potaz poměrně širokou řadu signálových faktorů / příznaků, které se váží k reálným vlastnostem řeči a jimiž tedy lze hodnotit její kvalitu. Signál reprezentující foném, slabiku, slovo nebo větu pak zobrazuje jako cluster v N-dimenzionálním prostoru, kde každý rozměr N symbolizuje jeden z příznaků. Je-li cílem zhodnocení kvality řeči ve smyslu „rodilosti“, zjišťujeme vzájemnou vzdálenost clusteru referenčního a analyzovaného signálu. Příklady zmíněných parametrů zobrazuje následující tabulka, převzatá z [25]. Parametry se zde klasicky kategorizují jako fonémické a prozódické:

Feature Category	Feature Name
Phonemic	Phone-level log-likelihood scores, GOP
	Vowel durations, duration trigrams
	Phoneme pair classifiers
	spectral features (formants)
	Articulatory-acoustic features
Prosodic (Intonation, Stress, fluency)	distances between stressed and unstressed syllables
	Mean, max, min power per word (energy)
	F0 contours (slope and maximum)
	rate of speech (words per second/minute)
	Trigram models to model phoneme duration in context
	Phonation/time ratio, mean phoneme duration
	Articulation Rate (phonemes/sec)
	Mean and standard deviation of long silence duration
	Silences per second
	Frequency of disfluencies (pauses, fillers etc)_
	Total and mean pause time (i.e. duration of interword pauses)

Obr. 4.4: Příklady parametrů k celkovému hodnocení výslovnosti [25]

Ve fonémické kategorii zde lze najít jak trvání segmentů (vowel durations), tak i příznaky artikulace (formants, articulatory-acoustic features). Prozódická kategorie zde zahrnuje průběh základní frekvence F0 i energii signálu. Následující podkapitoly popisují výpočetní metody k předzpracování signálu nebo k získání určitých příznaků, přičemž je uvedena jejich souvislost s výslovností.

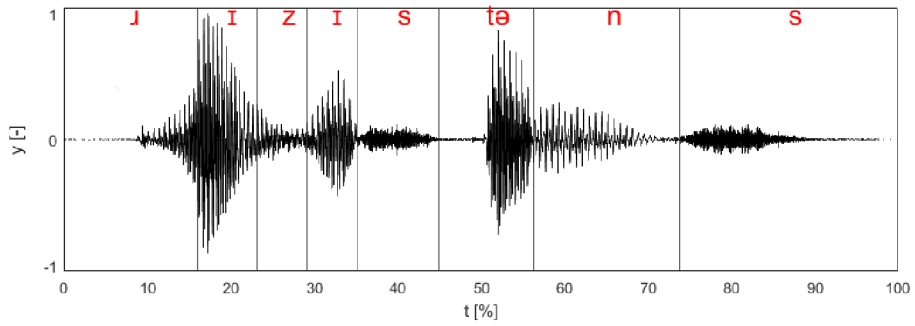
4.3 Výpočetní metody a naprogramované algoritmy

4.3.1 Vstup řečového signálu a relativizace os

Pro sejmutí řečového signálu je vhodné využít integrovaný mikrofon používaného zařízení (notebooku, smartphonu), případně headsetu. Vzhledem k záměru vytvoření efektivního programu ke kontrole výslovnosti má také velký smysl snažit se veškeré zpracování signálů, stejně jako odezvu, implementovat co nejvíce do reálného času. Pomocí integrovaných funkcí v MATLAB byl naprogramován algoritmus **record**, která po spuštění v reálném čase sejme vstupní zvukový signál a převede ho na vektor vzorků. **record** byl naprogramován s velmi jednoduchou konceptualitou - po jeho spuštění se zapne mikrofon snímající zvuk a zároveň se spustí odpočet 5 vteřin. Po uběhnutí tohoto odpočtu je mikrofon vypnut a nasnímaný zvukový signál uložen do proměnné y. Součástí této funkce je také algoritmus na identifikaci řeči mezi úseky ticha či mírného šumu. Díky tomu lze průběh slova v nahrávce izolovat.

Je nutné také vzít v potaz fakt, že mluvčí může při tréninku výslovnosti konkrétního výrazu každý pokus vyslovit různě rychle. Celková délka slova v jednotkách milisekund může tedy mít značný rozptyl, a to se negativně projeví na další analýze fonémických parametrů řeči. Počítá se tedy s relativizací časové osy vstupního signálu, kdy je osa v jednotkách milisekund převedena na procentuální osu. Za tímto účelem byla naprogramována jednoduchá funkce **time_ProcentualniPrizpusobeni**.

Vstupní signál je také vhodné normovat nejvyšší hodnotou - osa y potom má rozpětí [-1; 1] kolem střední hodnoty signálu. Střední hodnota je velmi důležitým indikátorem, k němuž bude vztahována řada spojitých modulačních faktorů, resp. příznaků. Při analýze řeči často není relevantní absolutní hodnota příznaku, místo toho je analyzována poloha příznaku vůči jeho střední hodnotě. Z tohoto důvodu budou také příznaky energie i melodie normovány a nevyhodnocovány v absolutní hodnotě, nýbrž relativně vůči své střední hodnotě v průběhu celého signálu. Příklad zpracovaného signálu slova "*resistance*"je na obr. 4.5.



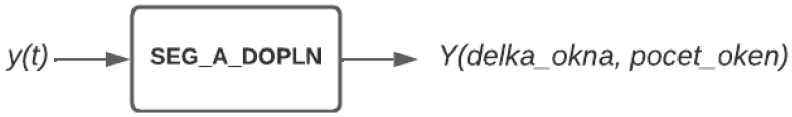
Obr. 4.5: Vstupní signál "resistance"s manuálním naznačením poloh hlásek

4.3.2 Segmentace

Protože cílem je zpracování signálu celých slov složených z několika hlásek, očekává se na vstupu nestacionární řečový signál a proto je nutné při zpracování zajistit jeho kvazistacionaritu pomocí oken, resp. segmentace.

MATLAB Algoritmus

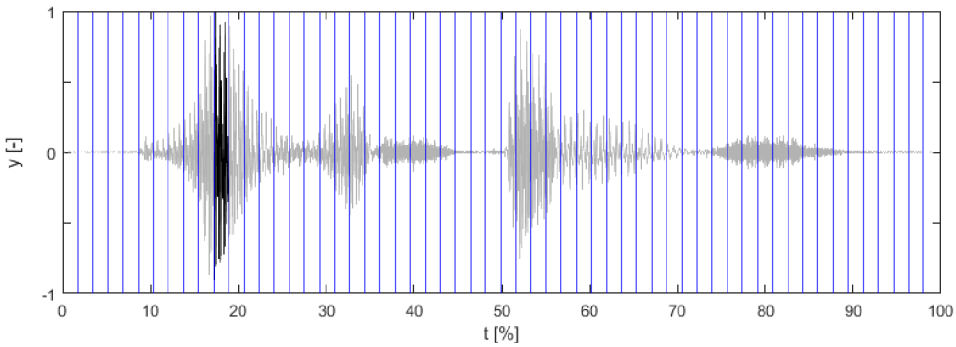
Pro zachování stacionarity signálu byla naprogramována funkce **SEG_A_DOPLN**:



Obr. 4.6: Funkce SEG A DOPLN

Nejdůležitějším vstupem této funkce je řečový signál y, společně s doplňkovými informacemi o jeho vzorkovací frekvenci Fs a délce kýženého stacionárního okna *delka_okna*. Tato funkce vstupní signál doplní nulami tak, aby podíl celkového počtu vzorků signálu a počtu vzorků okna bylo celé číslo; díky tomu může být signál dále rozdělen do matice Y, jejíž sloupce jsou jednotlivá okna signálu y. Takto je signál rozčleněn na krátkodobé segmenty, jež považujeme na stacionární; v řadě dalších algoritmů pak zpracování probíhá postupně po těchto oknech.

Výsledné členění vstupního signálu slova "*resistance*"pomocí naprogramované funkce **SEG_A_DOPLN** je ilustrováno na následujícím grafu; jednotlivé krátkodobé stacionární segmenty jsou vyznačeny modrými vertikálními čarami. Další zpracování pak probíhá postupně, např. zvlášť pro černě vyznačený 11. segment, tzn. *Y(:,11)*.



Obr. 4.7: Segmentace vstupního signálu

4.3.3 Energie

Přízvuk je modulován zesílením vzduchového proudu z plic a projevuje se zesílením produkované řeči. Analýza síly, resp. energie signálu, je tedy fundamentálním způsobem určování přízvuku, ale bude využita i pro zjišťování délky fonémických segmentů. Výpočetní metodou pro tento účel může být algoritmus STE (Short Time Energy). Spojení „short time“ zde upozorňuje na fakt, že objektem zkoumání je pouze krátkodobý úsek, resp. dílčí stacionární segment, analyzovaného signálu. Jedná se o velmi jednoduchou a robustní výpočetní metodu, jež pracuje přímo se vzorky signálu:

$$STE = \sum_{n=0}^{N-1} x(n)^2 \tag{4.1}$$

kde:

N = počet vzorků segmentu (okna)
x = vzorek signálu
n = index pořadí vzorku

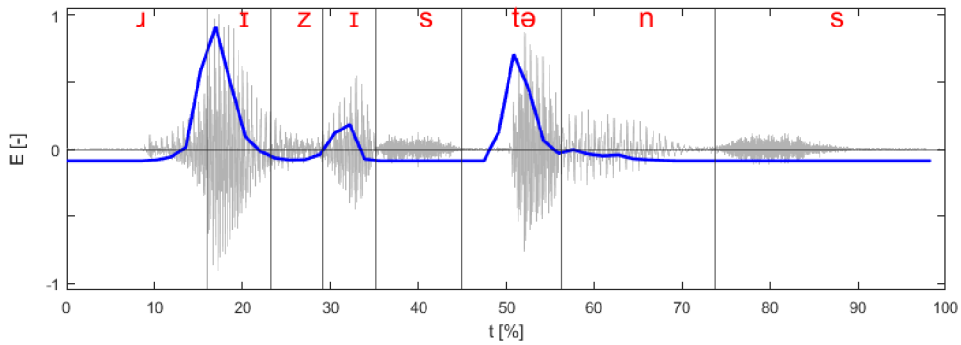
MATLAB Algoritmus

Pro určení krátkodobé energie byl naprogramován algoritmus **energie**:

Obr. 4.8: Funkce energie

Vstupem této funkce je matice stacionárních segmentů Y. Jádrem funkce je výpočet sum a jejich umocnění pro každý jednotlivý krátkodobý segment, resp. sloupec matice; výsledný vektor umocněných sum je dále normalizován max. hodnotou a centrován kolem své střední hodnoty. Výstupem je tedy vektor energie E, viz obr. 4.9.

Následující graf ilustruje průběh krátkodobé energie podél vstupního signálu; díky manuálnímu naznačení jednotlivých hlásek lze názorně sledovat zvýšení energie na polohách některých z nich. Podle definice přízvuku v kapitole 3.3.1 odtud lze získat intuici o přízvučnosti slova.



Obr. 4.9: Průběh energie podél vstupního signálu

4.3.4 Znělost, neznělost, ticho

Znělost se v řečovém signálu projevuje přítomností základního tónu generovaného hlasivkami. Neznělé hlásky oproti tomu mají charakter šumu. Lze předpokládat, že ve znělých úsecích bude energie signálu vyšší než v místech, kde převažuje šumový charakter. Díky tomu lze znělost v průběhu signálu určovat z křivky krátkodobé energie, kdy porovnáváním hodnot s příslušnými prahy bude krátkodobý segment o dané energii klasifikován jako znělý či neznělý.

MATLAB Algoritmus

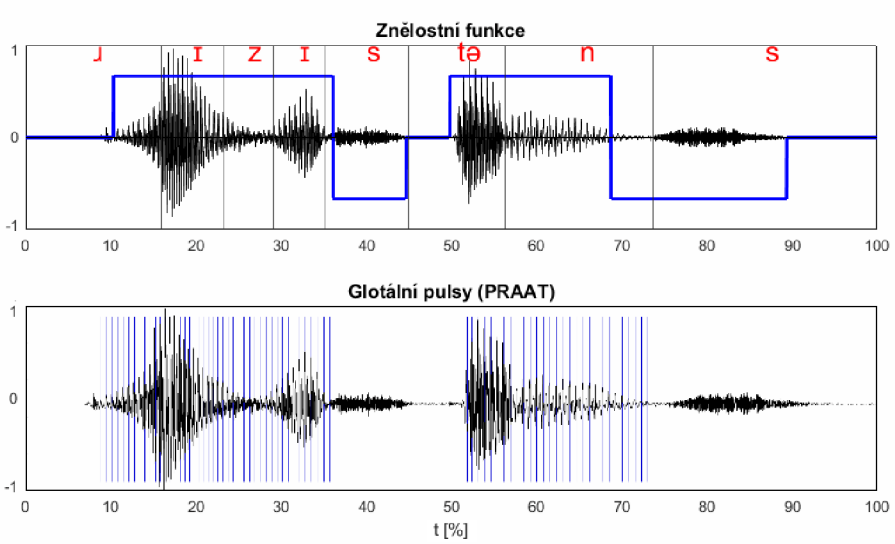
Pro výpočet znělosti byla naprogramována funkce **znelost_STE**:

Obr. 4.10: Funkce znelost STE

Nejdůležitějším vstupem funkce je matice *Y*, společně s doplňkovými informacemi jako experimentálně zvolené hodnoty prahů *prah_znelosti*, *prah_ticha*; protože v ní jsou integrovány další funkce pro pomocné výpočty vč. *SEG_A_DOPLN*, je doplňkovým vstupem také *delka_okna* a *max_pauza*. Jádro funkce spočívá ve výpočtu vektoru energie *E* a následné srovnávání hodnot s prahy v cyklu (i = 1:1:*pocet_oken)*. Výsledky se ukládají do výstupního vektoru *znelost* - pro hodnoty vyšší než *prah_znelosti* je momentální hodnota vyhodnocena jako znělá a *znelost(i)* = 1; hodnota v rozmezí *prah_znelosti* a *prah_ticha* je vyhodnocena jako neznělá a *znelost(i)* = -1; vše pod hodnotou *prah_ticha* je určeno jako ticho a *znelost(i)*

= 0. Nejdůležitějším výstupem funkce **znelost_STE** je tříhodnotový vektor *znelost(pocet_oken)*, jenž lze vykreslit a reprezentovat jako znělostní funkci, viz obr. 4.11.

Následující grafy ilustrují průběh znělosti u vstupního řečového signálu (jde opět o angl. slovo "*resistance*"). První z grafů zobrazuje třístavovou znělostní funkci (vektor *znelost*) podél signálu. Lze vidět, že podle očekávání je signál znělý v místech, kde se nacházejí znělé hlásky, zároveň je patrná krátká tichá pauza před /t/ způsobená specifickou spojitou změnou rezonančních dutin vokálního traktu. Druhý z grafů je výstupem softwarového nástroje PRAAT, který má integrovanou funkci určování znělosti metodou detekce glotálních pulsů; jak lze vidět, oba grafy se liší jen mírně. Odlišnost je patrná zejména u dokmitávání znělých hlásek - zatímco metoda STE tuto oblast již považuje za neznělou či tichou vlivem nízké energie pod prahem znělosti, metoda glotálních pulsů stále pulsy detekuje.



Obr. 4.11: Znělost

4.3.5 Základní tón

Melodie je modulována změnou periody kmitání hlasivek při tvoření vokalizovaných, tzn. znělých, hlásek. V řečovém signálu se projevuje přítomností stálého či kolísajícího tónu. Zřejmě nejrozšířenějšími metodami pro určování základního tónu signálu jsou ty, které pracují ve frekvenční oblasti; nejtypičtějším příkladem může být aglo-ritmus FFT. Základní tón je zde určen jako globální maximum v průběhu funkce spektra po určitém pre-processingu. Metoda použitá v této práci však pracuje v časové doméně a je založena na autokorelaci signálu. Autokorelační funkce ACF je ve zpracování signálu široce využívána k detekci periodických struktur. Přídomek

auto zde upozorňuje na fakt, že se jedná o verzi algoritmu klasické korelace, jehož vstupy jsou nyní dvě kopie totožného signálu. ACF je definována následovně:

$$R(\tau) = \sum_{n=0}^{N-1} x(n)x(n+\tau) \tag{4.2}$$

kde:

N = počet vzorků segmentu (okna)
x = vzorek signálu
n = index pořadí vzorku
τ = posun [n]

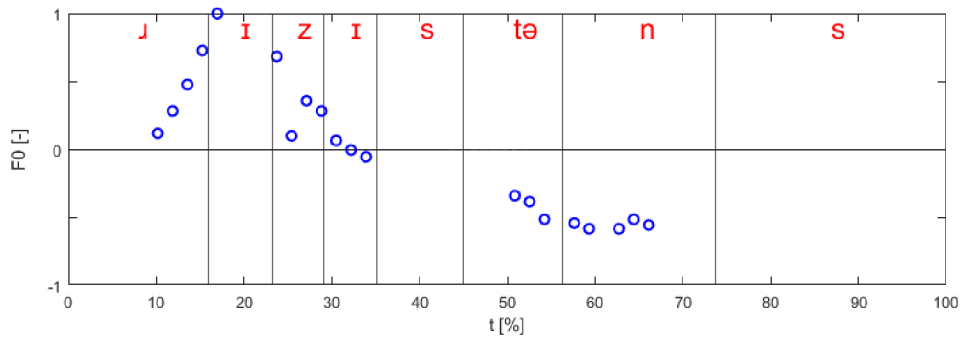
Pro vektor posunu tau jsou počítány korelační koeficienty, které mají typickou vlastnost měřítka podobnosti. Pokud je tedy v analyzovaném signálu přítomna periodická struktura, pro násobky určité hodnoty posuvu se překrývající kopie nacházejí ve fázi a vzájemná podobnost je tak vysoká. Vlastností výstupní autokorelační funkce tedy je, že za přítomnosti periodické struktury se v jejím průběhu objeví lokální maxima, a to na násobcích základní periody T0 zmíněné periodické struktury. Informaci o T0, potažmo i základním kmitočtu F0, tak lze velmi jednoduše získat změřením vzdálenosti sousedních maxim.

Pro určení melodie byla naprogramována funkce **pitch_ACF**:

Obr. 4.12: Funkce pitch STE

Hlavním vstupem této funkce je segmentovaný vektor *Y*. Dalšími doplňkovými vstupy jsou *Fs*, *delka_okna* a vektor *znelost_y*. Jádrem funkce je výpočet vektoru základního kmitočtu *F_Y*. Vstupní signál je ovšem nejprve filtrován dolní propustí o mezní frekvenci 400 Hz pro odstranění vyšších harmonických složek nad dosažitelným hlasovým rozsahem, čímž je získána matice *Y_filtered*. Protože má dále smysl hledat základni frekvenci F0 pouze u znělých segmentů, jsou neznělé a tiché segmenty ignorovány. Pro znělé segmenty je vypočtena krátkodobá autokorelační funkce ACF a z jejího průběhu je extrahována poloha prvního maxima v jednotkách vzorků. Tato hodnota je dále převedena na časový údaj a frekvenci F0. Ta je postupně pro každý segment ukládána do výsledného vektoru *F_Y*; neznělé a tiché segmenty jsou zaznamenávány jako prázdná hodnota. Tento vektor je dále normován kolem své střední hodnoty, výstupem tedy není hodnota frekvence v [Hz], ale relativní průběh.

Je třeba říci, že funkce **pitch_ACF** je velmi citlivá na délku krátkodobého segmentu. Musí v něm být přítomny alespoň dvě periody vstupního signálu, v opačném případě nemohou být nalezena maxima autokorelační funkce ani výsledná základní frekvence. Délka okna 20 ms tuto podmínku zpravidla vždy splňuje. Příliš dlouhé okno by naopak zmenšovalo rozlišení výsledku a významně by se snižovala jeho přesnost.



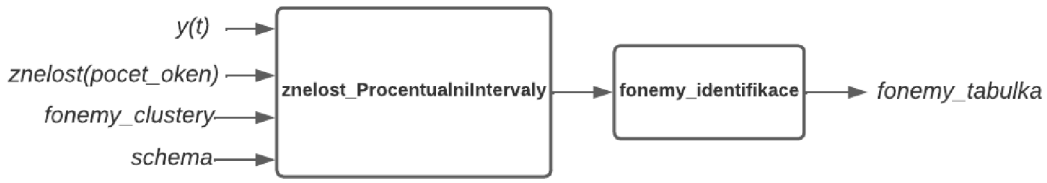
Obr. 4.13: Vizualizace růstu a poklesu melodie vůči střední hodnotě melodie

4.3.6 Trvání fonematických úseků

Pro získání hodnoty délky konkrétní fonematické jednotky je nutné signál nejprve rozčlenit. Rozčlenění nahrávky konkrétního slova na jednotlivé hlásky či slabiky je složitá (a v případě slabik i nejednoznačná) úloha, jenž je stále řešena v rámci vývoje speech-to-text technologie. Protože vývoj uspokojivě přesného algoritmu by byl časově příliš náročný, bylo rozčlenění značně zjednodušeno. Zde implementovaná metoda je založena na analýze znělosti.

Pro členění signálu na fonematické segmenty vycházíme z předpokladu, že známe fonetický přepis analyzovaného slova. Příkladem opět budiž angl. slovo „*resistance*“ ([rɪzɪstəns], v překladu „odpor“). Z jeho fonetické transkripce systémem IPA předem víme, které hlásky či clustery hlásek jsou znělé (červenou barvou zvýrazněné) a které naopak neznělé, případně lze odvodit, kde jsou úseky ticha – díky tomu můžeme slovo chápat nejen jako posloupnost fonémů či slabik, ale také jako posloupnost nově definovaných pseudoslabičných fonematických segmentů – clusterů znělých či neznělých hlásek. V případě slova "*resistance*“ je tedy charakteristické členění /ɹɪzɪ/ + /s/ + /-/ + /tən/ + /s/ podle schématu znělosti [1; -1; 0; 1; -1]; prostřední prázdný segment značí krátkou pauzu před plozivem /t/. U uvedených fonematických segmentů je dále možné určovat jejich časové trvání - cílem je tedy přiřazení znělým, neznělým a tichým úsekům řečového signálu příslušných fonémických segmentů.

Pro identifikaci fonematických segmentů v průběhu signálu bylo naprogramováno několik dílčích funkcí.



Obr. 4.14: Funkce pro výpočet trvání

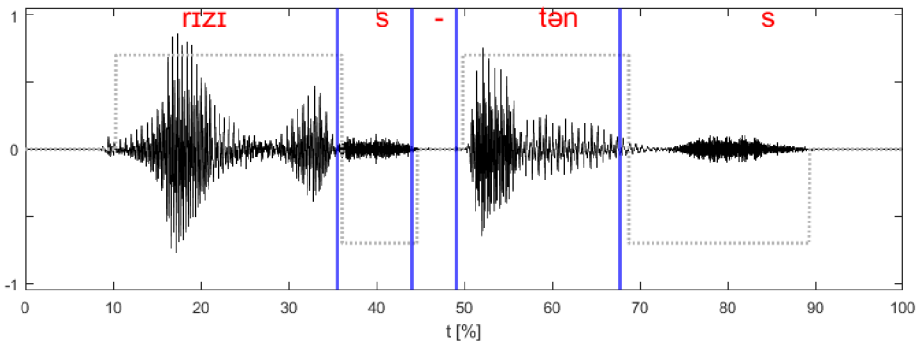
Pomocí funkce **znelost_STE** je nejprve vypočten vektor znělosti *znelost* daného řečového signálu *y*. Následně jsou s použitím naprogramované pomocné funkce **znelost_ProcentualniIntervaly** vyhledány a označeny přechody mezi znělými, neznělými a tichými úseky v průběhu signálu *y*. Tímto jsou přesně nalezeny polohy hranic mezi clustery znělých hlásek, neznělých hlásek a úseky ticha.

Následně jsou další pomocnou funkcí **fonemy_identifikace** ohraničeným úsekům signálu *y* přiřazeny clustery konkrétních fonémů podle znělostního schématu - předpokladem je, že po výpočtu průběhu znělosti podél řečového signálu je nalezeno stejné schéma, jaké je předpokládáno pro očekávané slovo, tzn. pro slovo „*resistance*“ je očekáváno zmíněné schéma znělosti [1; -1; 0; 1; -1] při fonematické segmentaci na clustery /ɹɪzɪ/ + /s/ + /-/ + /tən/ + /s/; nepřítomnost tohoto schématu, která může vzniknout např. eliminováním pauzy před /t/ nebo nahrazením znělého /z/ za neznělé /s/, způsobí chybovou hlášku programu. Důvodem je, že program je navržen na kontrolování parametrů již správně artikulovaného slova, nikoliv na kontrolu samotné artikulace. Ošetření této varianty bude předmětem dalšího vývoje programu.

Výsledkem zpracování signálu algoritmy může být následující matice:

Rozmezí [%]		**Znělost**	**Clustery**
10	36	*1*	ɹɪzɪ
36	44	*-1*	s
44	49	*0*	-
49	68	*1*	tən
68	88	*-1*	s

Z matice je nyní velmi jednoduché získat procentuální hodnoty délek clusterů hlásek. Následující graf vizualizuje automatické rozčlenění řečového signálu na fonematické úseky - clustery hlásek.



Obr. 4.15: Vizualizace automatického členění na fonematické úseky

4.3.7 Interpretace příznaků ve vztahu k řeči

Protože program, jímž se tato práce zabývá, má být primárně určen běžnému uživateli bez lingvistických znalostí, které by mu umožňovaly vyvozovat závěry o své výslovnosti přímo z průběhů vypočtených příznaků, je nezbytné řešit také uživatelsky přívětivou interpretaci výsledků. Tato interpretace by měla být velmi jednoduchá a nabízející jasnou odezvu k výslovnosti. Zcela optimální formou odezvy se zde jeví verbální oznámení či doporučení.

Díky funkci **fonemy_identifikace** je možné v signálu identifikovat jednotlivé clustery hlásek a přesně vymezit hranicemi jejich polohu, nabízí se tudíž možnost v průběhu každého clusteru vyhodnotit jeho trvání, energii a melodii. Bylo naprogramováno několik pomocných funkcí, které srovnají hodnoty energie a melodie v jednotlivých úsecích, a to vůči jejich střední hodnotě v průběhu signálu; jak bylo řečeno v kapitole 4.3.1, při analýze výslovnosti je důraz kladen na relativní hodnoty parametru vůči jeho střední hodnotě, spíše než na abs. hodnoty. **V případě, že energie daného clusteru překračuje střední hodnotu energie řečového signálu, je cluster označen jako přízvučný. Podobně je postupováno v případě melodie - pokud melodie v úseku daného clusteru překračuje střední hodnotu melodie průběhu celého řečového signálu, je na tomto clusteru detekován vzrůst hlasu. V opačném případě je identifikován pokles hlasu.** Tímto postupem by bylo možné uživateli formovat odezvu např. následující formou:

The section no. 1 - ɹɪzɪ (voiced) is 26 percent long.
The section no. 2 - s (unvoiced) is 8 percent long.
The section no. 2 - -(silence) is 5 percent long.
The section no. 3 - tən (voiced) is 19 percent long.
The section no. 4 - s (unvoiced) is 20 percent long.

Your word has stressed /ɹɪzɪ/; stressed /tən/.

Your word has voice rise at ɹɪzɪ; voice fall at tən.

Uvedené odezvy by byly použitelné v případě jednoduché analýzy jediné nahrávky hlasu uživatele. Protože však je nutné přesně znát správnou výslovnost daného slova pro vytvoření relevantní zpětné vazby, je nezbytné do programu implementovat nahrávku vzorové výslovnosti lektora, s níž je výslovnost uživatele srovnávána (a to pro všechny příznaky - trvání, energii i přízvuk). Počítá se tedy s tím, že program bude mít interně uložen i etalon výslovnosti - délku každého clusteru i míru energie a přízvuku v každém clusteru. Vypočtené příznaky uživatelova signálu pak s nimi bude porovnávat pro získání představy o odchylkách od referenční "etalonové"výslovnosti. Bylo naprogramováno několik pomocných algorimů, které tyto dvojice příznaků vyhodnocují. Díky tomu je finálně možné podat uživateli verbální zpětnou vazbu ve formě doporučení na zlepšení výslovnosti. V případě špatné výslovnosti může odezva mít například následující podobu:

The section no. 1 - ɹɪzɪ (voiced) is 3 percent longer.
The section no. 2 - s (unvoiced) is 1 percent shorter.
The section no. 2 - -(silence) is 3 percent longer.
The section no. 3 - tən (voiced) is 2 percent shorter.
The section no. 4 - s (unvoiced) is 6 percent longer.

Your word has stressed /ɹɪzɪ/; stressed /tən/. Model word has stressed /ɹɪzɪ/.

Your word has voice rise at /ɹɪzɪ/; voice fall at /tən/. Model word has voice rise at /ɹɪzɪ/; voice fall at /tən/.

4.4 Metody a algoritmy pro budoucí vývoj

V průběhu vývoje programu bylo navrženo několik prototypových funkcionalit pro potenciální budoucí využití. Vývoj těchto funkcionalit vyžaduje větší množství času a bude dále pokračovat. V následujících podkapitolách jsou tedy pouze stručně zmíněny základní principy jejich zamýšleného fungování.

4.4.1 Formantové kmitočty

Jak bylo popsáno v kapitole 3.3.3, znělé hlásky jsou tvořeny proudem zvuku filtrovaného vokálním ústrojím v určitém postavení. Hypoteticky tedy je možné konkrétní znělou hlásku odhadnout z jejího izolovaného časového nebo frekvenčního průběhu. Může toho být dosaženo zpracováním signálu několika naprogramovanými algoritmy, popsanými v následujících odstavcích:

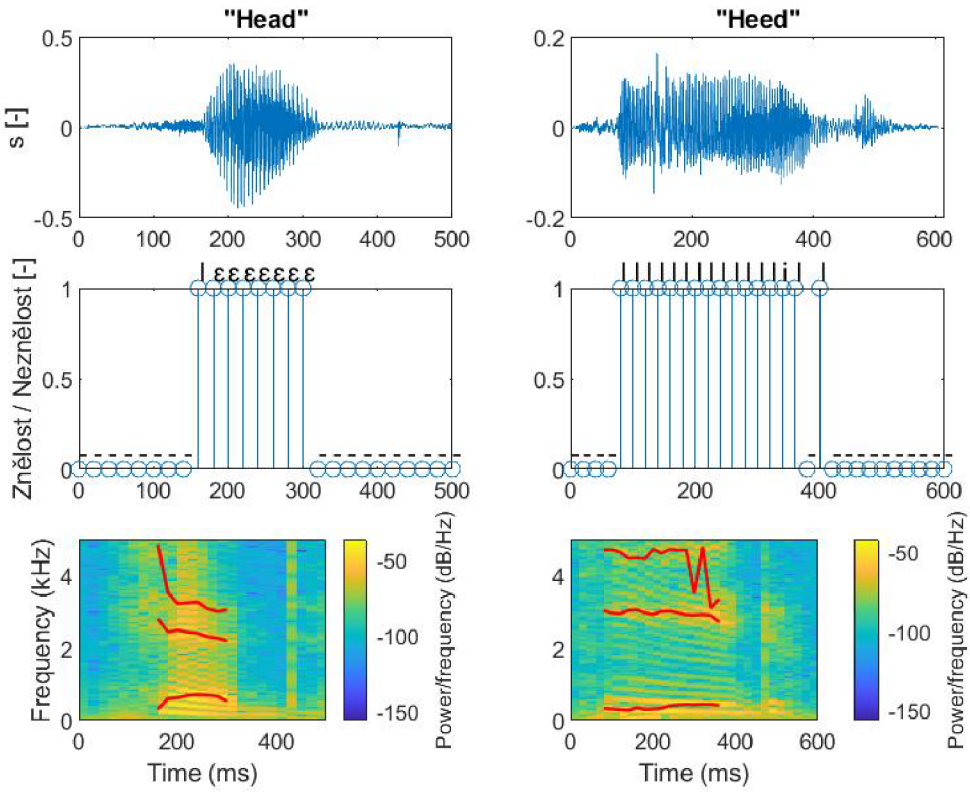
Obr. 4.16: Funkce pro identifikaci hlásek

Ve většině případů se formantové kmitočty určijí na základě analýzy spektrální obálky za použití metody LPC Linear Predictive Coding [23]. Jde o výpočetně nenáročnou metodu založenou na odhadu parametrů modelu tvorby řeči podle stacionárního segmentu řečového signálu, resp. odhadu budoucího vzorku signálu na základě lineární kombinace několika předchozích vzorků. Modelem se zde rozumí simulace hlasového ústrojí pomocí celopólového IIR filtru [22]. Implementace algoritmu byla zjednodušena využitím předpřipraveného algoritmu, dostupného na oficiální nápovedě MATLAB. Byla vytvořena funkce **dip_Formanty**, jejímž cílem je rozčlenění vstupního řečového signálu na krátkodobé segmenty, klasifikování segmentů na znělé a neznělé a výpočet formantových kmitočtů pro každý znělý segment. Tato funkce opět používá algoritmus **energie**, jež vstupní signál rozčlení na krátkodobé segmenty a klasifikuje jako znělé a neznělé. Hlavním vstupem funkce je analyzovaný signál *y* společně s dalšími doplňkovými parametry - vzorkovací frekvence *Fs*, parametr *delka_okna* udávající informaci o délkce segmentu v [ms] pro algoritmus **energy** a parametr *pocet_koeficientu*. Poslední uvedený parametr *pocet_koeficientu* je určen k výpočtu LPC uvnitř funkce; určuje rozsah modelu filtru a má přímý vliv na správnost určení formantů. Ačkoli je doporučeno pro řečové signály volit hodnoty 10 – 15, experimentálně byla hodnota nastavena do defaultní hodnoty 30, protože vykazovala nejlepší výsledky. Výstupem funkce je matice formantů o rozměru [pocet_formantu, pocet_oken]. Tato funkce byla doplněna další funkcí **dip_PomeryFormantu**, určenou k extrakci matice [3, pocet_oken] obsahující první 3 formanty pro každý krátkodobý segment. Dále jsou vypočteny jejich poměry jako návratová hodnota funkce. Výsledné poměry jsou porovnávány s referenčními hodnotami formantových kmitočtů pro jednotlivé hlásky pomocí třetí funkce **dip_UrceniFonemu**. To se provádí srovnáním euklidovských vzdáleností poměrů formantů pro daný segment vůči referenčním poměrům. Reference byly získány z [17]. Následující tabulka znázorňuje přiřazení dvojic F1 a F2 hláskám (ženy):

F1 [Hz]	310	430	610	860	850	590	470	370	760	500
F2 [Hz]	2790	2480	2330	2050	1220	920	1160	950	1400	1640
F3 [Hz]	3310	3070	2990	2850	2810	2710	2680	2670	2780	1960
F1/F2	0.111	0.173	0.262	0.420	0.697	0.641	0.405	0.389	0.543	0.305
F2/F3	0.843	0.808	0.779	0.719	0.434	0.339	0.434	0.356	0.504	0.837
hlásky	i	I	e	æ	a	c	U	u	v	d

Tab. 4.1: Formanty

Díky tomu je možné automaticky přepisovat řečový signál na některé fonetické symboly IPA:



Obr. 4.17: Příklad Určení fonému

4.4.2 Fonematické členění na hlásky

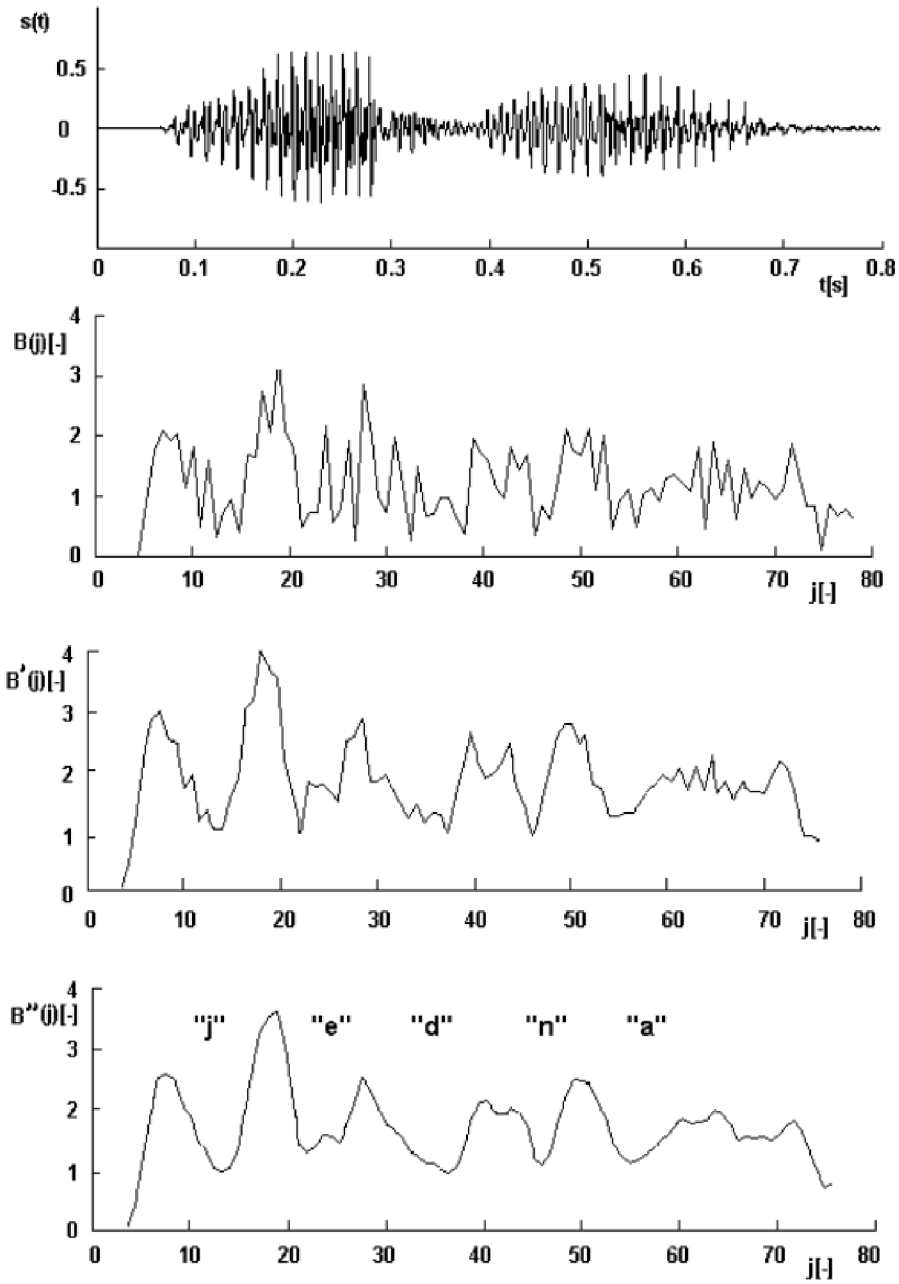
Současná verze programu při členění nahrávky slova počítá s nově definovanou suprasegmentální fonematickou jednotkou v podobě clusterů znělých či neznělých hlásek, viz kap. 4.3.6. Protože jde o velmi základní členění, v budoucích verzích programu je žádoucí slovo členit na slabiky či hlásky. Členění na slabiky je nejednoznačné, protože vymezení slabiky může být velmi subjektivní jak v různých jazycích, tak pro několik mluvčích jednoho jazyka. Z toho důvodu v úvahu připadá automatické členění na hlásky.

$$B(j) = b\frac{|R(0)_{j+l_1} - R(0)_{j-l_2}|}{R(0)_{j+l_1} + R(0)_{j+l_1}} + \sum_{k=1}^{K} |R(0)_{k+l_1} - R(0)_{k-l_2}| \tag{4.3}$$

kde:

$R(0)$ = energie signálu
k = řád autokorelačních koeficientů
b = multiplikativní konstanta
j = pořadí analyzovaného krátkodobého segmentu
l_1 = počet kroků časového posuvu dopředu [segment]
l_2 = počet kroků časového posuvu dozadu [segment]

Podle [33] je možné pomocí autokorelační funkce získat křivku *B(j)* a její vyhlazenou verzi *B'(j)*, jejichž lokální maxima určují přechody mezi hláskami. Obr. 4.18 znázorňuje tento proces graficky. Byla naprogramována funkce **krivka_B**, jejímž výstupem jsou přechody mezi hláskami; lazení této funkce je však velmi náročné a výsledek není spolehlivý bez dalších podpůrných funkcí. Členění na fonematické úseky proto v programu zůstává na úrovni znělých a neznělých clusterů.



Obr. 4.18: Určení přechodu mezi hláskami [23]

4.4.3 Detekce českého mluvčího

Pro detekci českého mluvčího je nutné znát charakteristické chyby českých mluvčích při výslovnosti angličtiny. Styl řeči zatížený těmito chybami je někdy neformálně nazýván jako "Czenglish". Pokus o získání informací o těchto chybách byl proveden testem pomocí vytvořených algoritmů na určování trvání, energie a melodie. Test byl proveden na 11 mluvčích - českých studentech ústavu UREL na VUT FEKT, tzn. všichni prošli zkouškou z angličtiny a jejich jazyková úroveň je přinejmenším středně pokročilá. Výslovnost byla měřena na 10 slovech bez předchozího nácviku. Tato slova byla vybrána z tematických okruhů zpracování signálu a teorie obvodů. Každý z jedenácti subjektů nahrál svou výslovnost pro každé slovo, nahrávky byly následně analyzovány naprogramovanými algoritmy. Takto byly získány příznaky - informace o trvání a vektory energie a zákl. frekvence.

Pro identifikaci parametru, který nejtypičtěji určuje odchylku českého mluvčího, byly dále vypočteny hodnoty procentuální podobnosti mezi příznaky vzorové výslovnosti a výslovností subjektů:

Pro **trvání** byly porovnávány součty délek jednotlivých znělých clusterů uživatele a lektora.

$$V_p = 100(\sum_{v_{user}} / \sum_{v_{lecturer}})[\%] \tag{4.4}$$

kde:

V_p = procentuální podobnost trvání
v_{user} = znělé procento signálu uživatele
$v_{lecturer}$ = znělé procento signálu lektora

Pro **přízvuk** je za míru podobnosti vzato maximum korelační funkce mezi vektory E lektora a uživatele, převedené na procenta.

$$E_p = 100(max(R_{E_{user},E_{lecturer}}(t1,t2)))[\%] \tag{4.5}$$

kde:

E_p = procentuální podobnost přízvuku
E_{user} = vektor energie uživatele
$E_{lecturer}$ = vektor energie lektora
$t1, t2$ = časový posun korelační funkce (použito 0 - 10 ms)

Pro **melodii** je za míru podobnosti vzato maximum korelační funkce mezi vektory F lektora a uživatele, převedené na procenta.

$$F_p = 100(max(R_{F_{user},F_{lecturer}}(t1,t2)))[\%] \tag{4.6}$$

kde:

F_p = procentuální podobnost přízvuku
F_{user} = vektor energie uživatele
$F_{lecturer}$ = vektor energie lektora
$t1, t2$ = časový posun korelační funkce (použito 0 - 10 ms)

Pro vyhodnocení celkové výslovnosti jsou výsledky všech tří parametrů rovnoměrně váhovány a sečteny do parametru SR. Následující tabulka ilustruje výsledek jednoho ze subjektů; ve spodním řádku jsou vypočteny průměrné hodnoty výslovnosti pro všechna kritéria - E_p, F_p, V_p a SR.

Slovo	**Vyhodnocená kritéria**			
	E_p [%]	F_p [%]	V_p [%]	**SR** [%]
magnitude	89	84	99	91
envelope	65	89	100	85
harmonicity	92	87	53	78
distortion	89	64	96	83
transformation	74	65	82	74
spectrogram	88	77	91	85
optimized	73	74	92	80
amplitude	62	73	92	76
resolution	69	79	92	80
component	58	84	82	75
	mean	mean	mean	mean
	75	77	87	80

Následující tabulka statisticky shrnuje výsledky všech 11 subjektů. Lze z ní vyvodit jasný závěr, že čeští mluvčí u anglického jazyka nejvíce chybují v přízvuku. Parametr E_p popisující přízvuk má průměrnou podobnost se vzorem rovnu 64% při velkém rozptylu 23 - 97 %. Pro detekci českého mluvčího by tedy mělo smysl se zaměřit právě na tento parametr.

Kritérium	Vyhodnocení	
	Rozptyl [%]	Průměr [%]
E_p	23 – 97	64
F_p	43 – 89	71
V_p	36 – 100	81
SR	44 – 91	72

Modul na detekci českého mluvčího by mohl pracovat na bázi porovnávání přízvučnosti u jednotlivých slabik. V případě, že by byly známy typické odchylky polohy přízvuku, např. záměna [rɪzɪ'stəns] a [rɪ'zɪstəns], bylo by snadné mluvčího odhadnout jako Čecha. Vytvoření modulu by však vyžadovalo velmi rozsáhlý výzkum odchylek přízvuku českých mluvčích pro velký soubor slov (přízvuk v anglickém jazyce je pohyblivý a tedy těžko generalizovatelný) a algoritmus členění slova na slabiky. Z toho důvodu byla předložena pouze koncepce modulu.

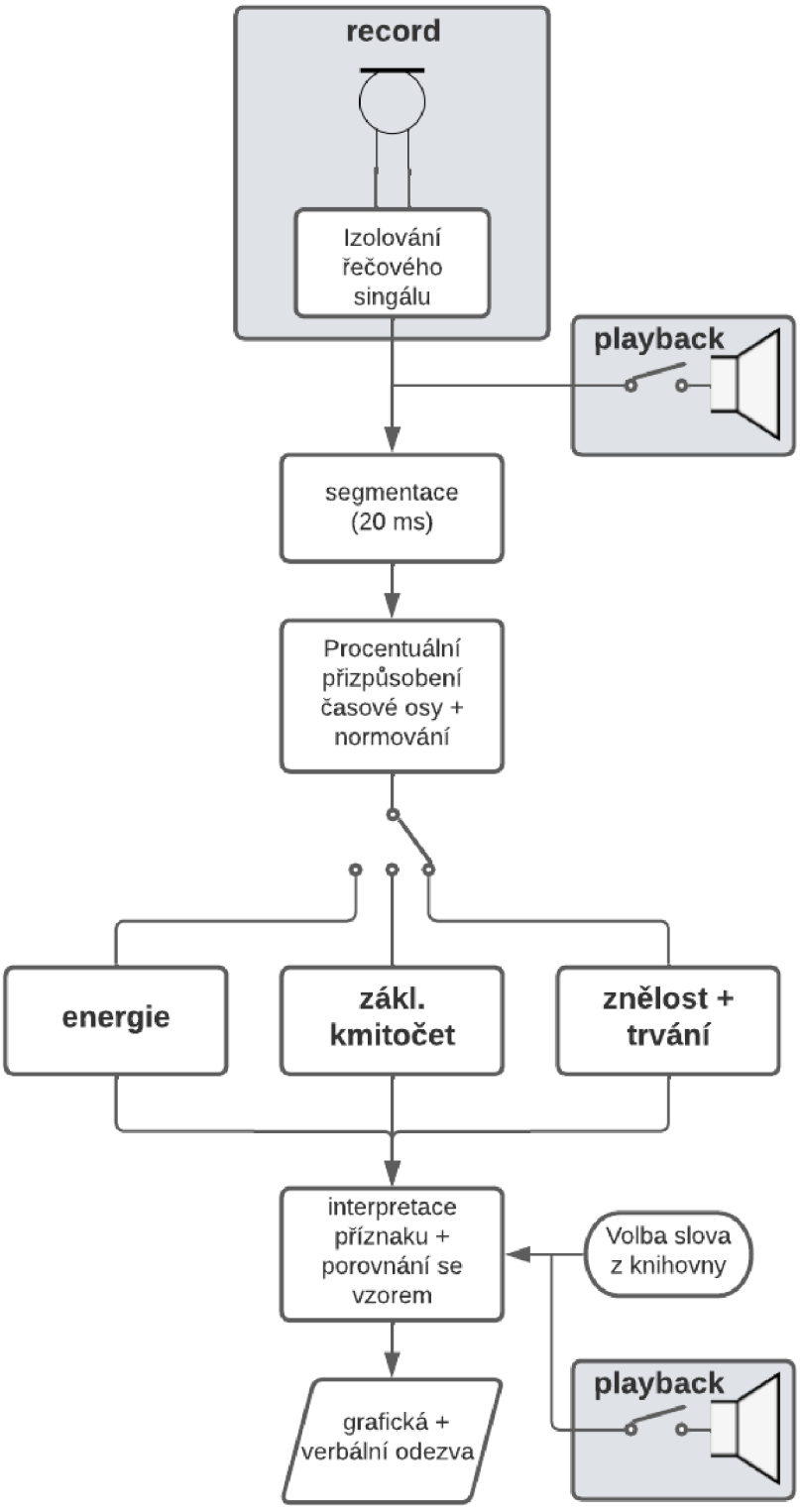
5 Popis programu

Na základě provedené rešerše již existujících CALL nástrojů lze prohlásit, že tyto nástroje mají vesměs společné nevýhody - jsou netransparentní a zpravidla nedávají uživateli detailní zpětnou vazbu na více parametrů výslovnosti zvlášť. Z toho důvodu byl navržen koncept nového programu, který tyto nedostatky vyváží a bude se soustředit na výpočty příznaků řečového signálu, které bude separovaně interpretovat jako dílčí parametry výslovnosti. Zpětná vazba bude podávána zvlášť podle výběru konkrétního parametru.

Program bude disponovat integrovanou knihovnou / slovníkem, z něhož si uživatel vybere slovo k procvičení. Technická koncepce vychází z toho, že správnost výslovnosti, resp. konkrétního parametru výslovnosti (tedy přízvuk, melodie, trvání) bude měřena vůči referenčním hodnotám, vypočteným z nahrávky lektora či rodilého mluvčího. Tato nahrávka, společně s referenčními hodnotami, bude uložena v paměti programu pro každé konkrétní slovo ve slovníku.

Tvorba takového slovníku tedy pro každé obsažené slovo vyžaduje nahrávku namluvenou lektorem či rodilým mluvčím; následně je nutné nahrávku zpracovat, tzn. rozčlenit ji na fonémické segmenty a definovat znělostní schéma. Zároveň by součástí datasetu slova měly být i hodnoty délky fonematických segmentů a hodnoty příznaků podél nich. Díky tomu bude umožněno vzájemné porovnání výslovnosti uživatele s výslovností lektora.

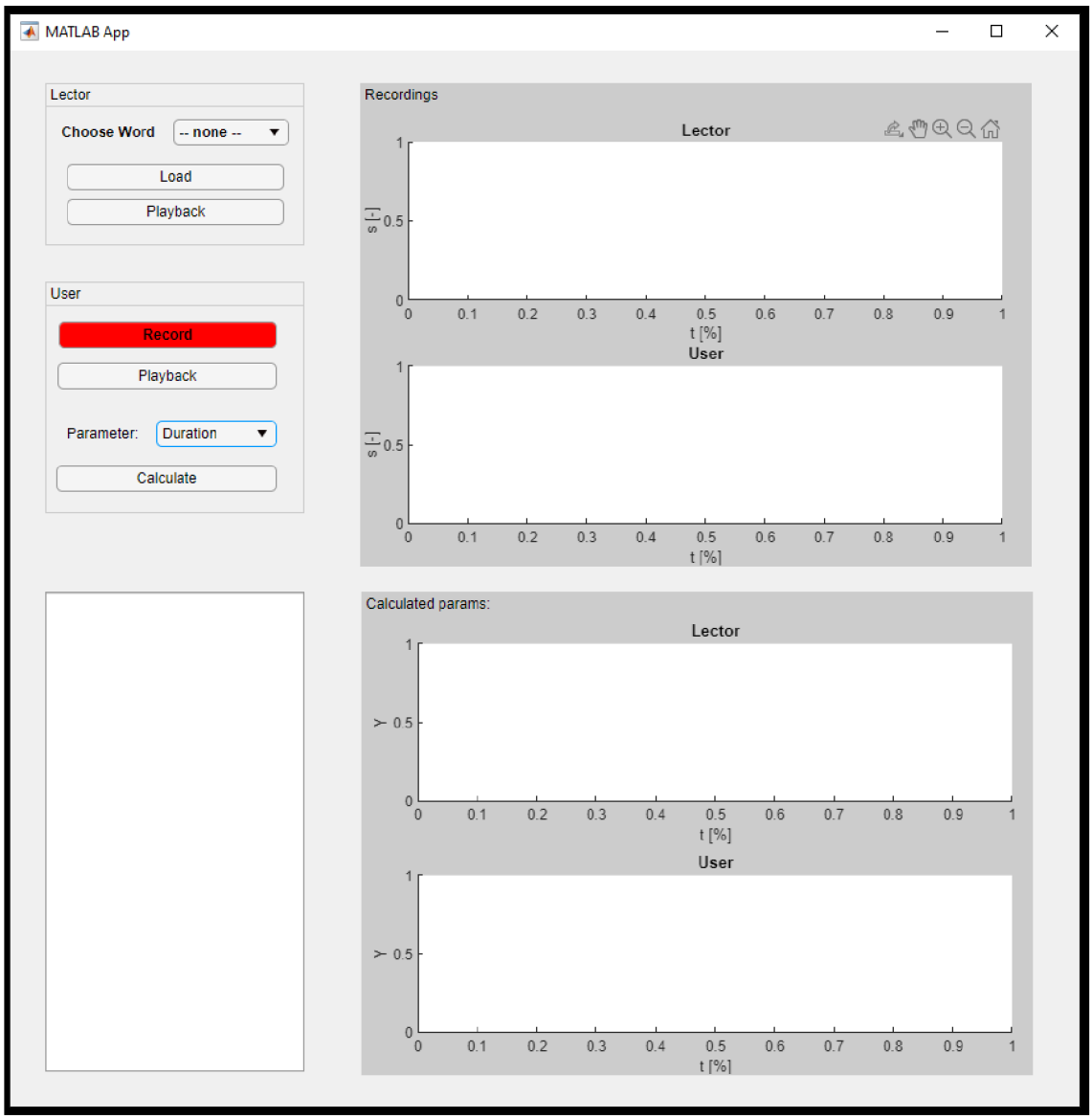
Celková koncepce programu je nastíněna na schématu 5.1. Uživatel nejprve ze slovníku vybere slovo k procvičení. Tímto se načtou parametry daného slova a jsou k dispozici pro porovnání. Následně uživatel spustí nahrávání, namluví dané slovo a vybere parametr výslovnosti - přízvuk, melodii či délku. Tím je zároveň vybrána metoda k výpočtu příznaku uživatelovy nahrávky. Příznak je vypočten, interpretován jako parametr výslovnosti a porovnán s lektorovými daty. Uživateli je pak zobrazena grafická i verbální zpětná vazba k výslovnosti.



Obr. 5.1: Schéma programu

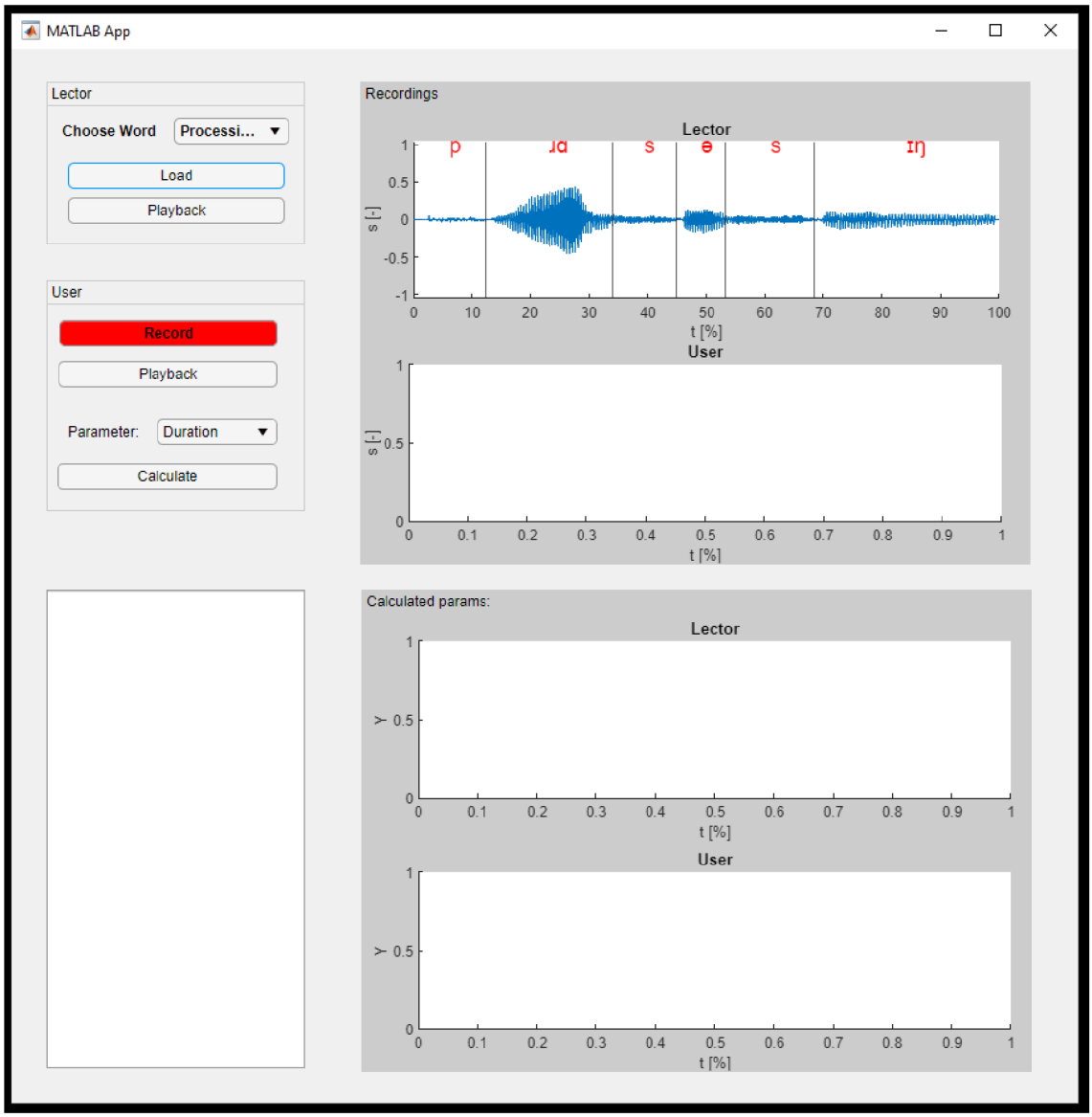
6 GUI a testování

Program disponuje grafickým uživatelským rozhraním pro jednoduché použití. V horní části se nachází menu pro výběr slova ze slovníku a možnosti playbacku. V dolní části je menu pro nahrání pokusu, také s možností playbacku. Dále se v rolovacím menu nabízí možnost výběru parametru. Vlevo dole je pak okno pro zobrazování verbální odezvy. Okna pro grafy v horní části slouží ke zobrazení samotných řečových signálu lektora a uživatele, ve spodní části pak ke grafickému zobrazení příznaku.



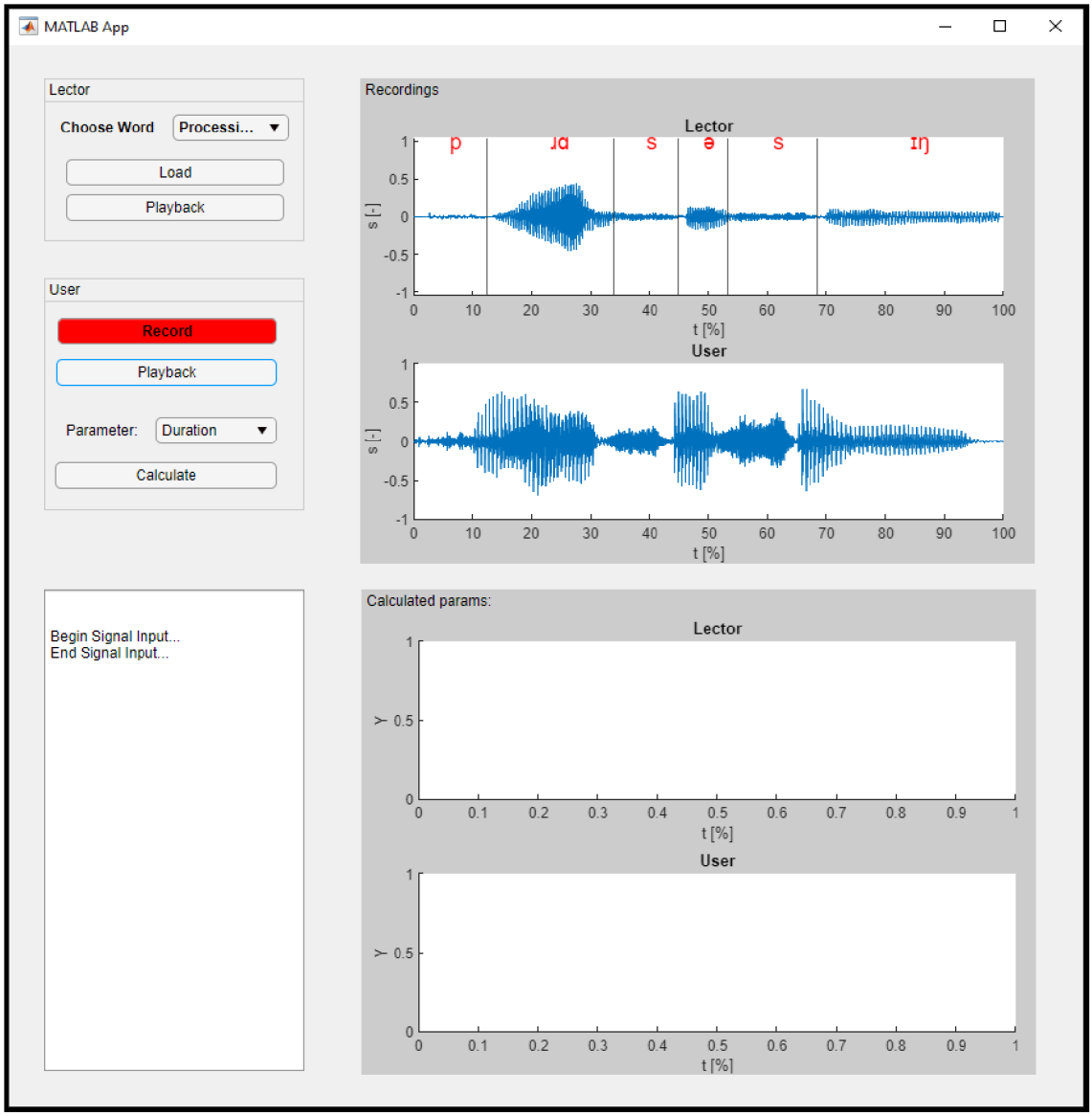
Obr. 6.1: GUI programu

Při použití je prvník krokem výběr slova k procvičení a jeho načtení. Výběr se provede pomocí rolovacího menu, načtení je provedeno tlačítkem *Load*. V prvním z oken se objeví průběh řeči, rozčleněné na fonematické úseky - clustery hlásek.



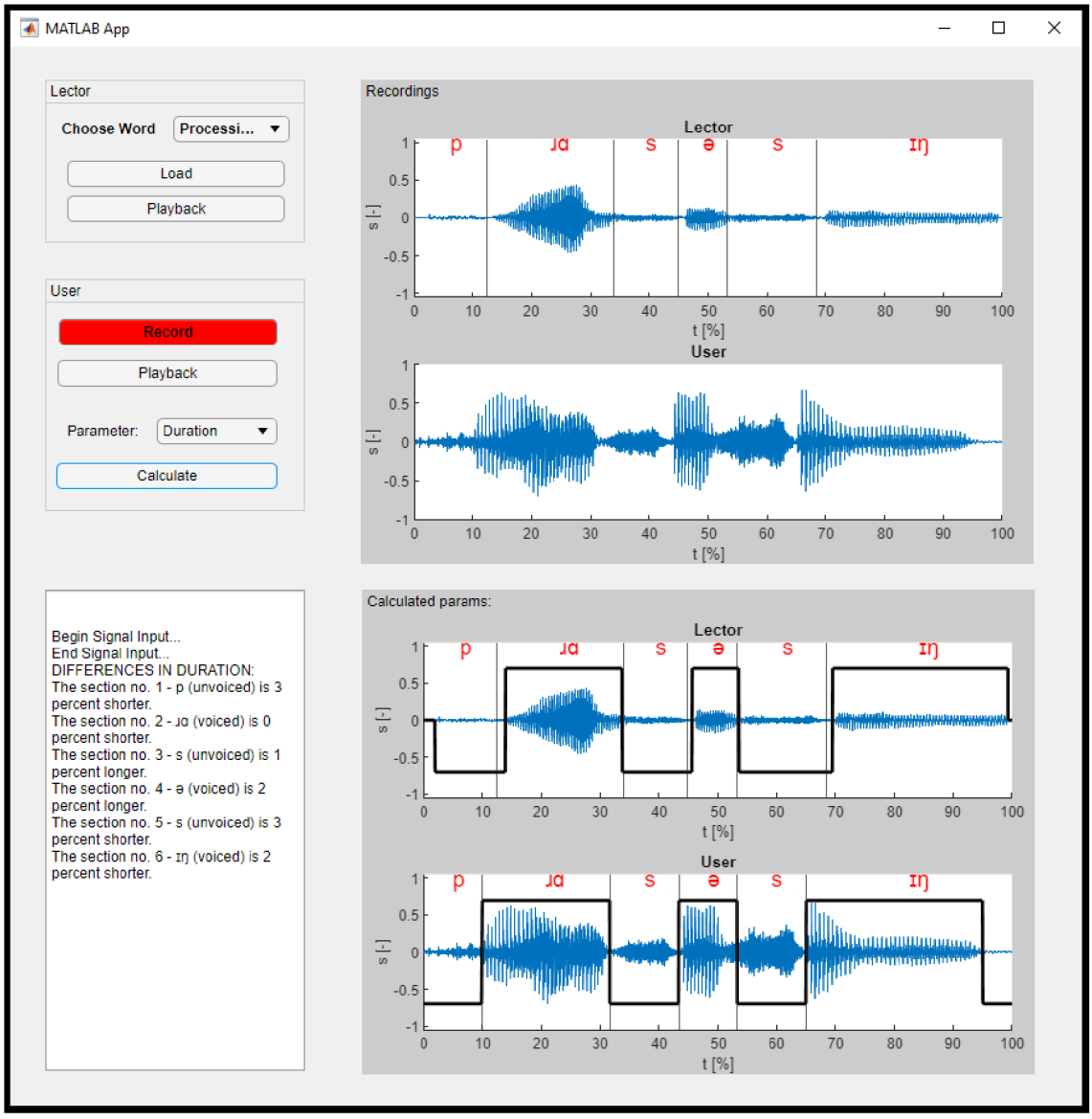
Obr. 6.2: Load

Následně je možné spustit nahrávání a pokusit se o vyslovení vybraného slova. Čas pro nahrávání je standardně nastaven na 5 vteřin, start je signalizován hláškou *"Begin Signal Input"* a stop hláškou *"Stop Signal Input"*. V druhém okně se objeví izolovaný a uložený průběh řeči, zatím nezpracovaný.



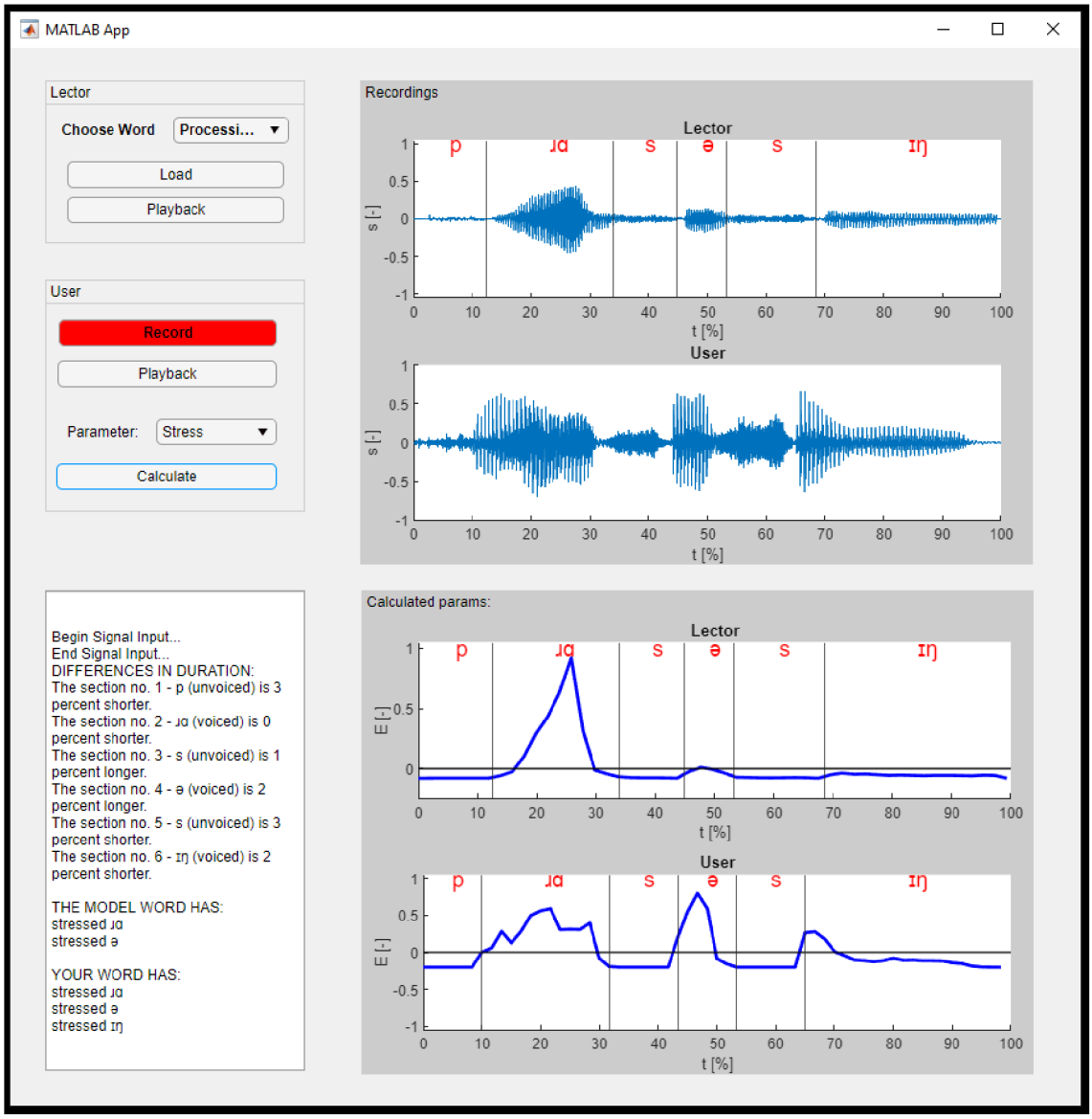
Obr. 6.3: Record

Následně je možné vybírat mezi kontrolovanými parametry pomocí příslušného rolovacího menu. Jako první analyzovaný parametr byla v testu vybrána délka fonematických úseků *Duration*. Po výběru je proveden výpočet. Výsledky jsou názorně zobrazeny v oknech - grafická reprezentace vykresluje oba řečové signály společně se znělostní funkcí a automatickým rozčleněním uživatelova signálu na fonematické úseky. Informace o odlišnostech ve výslovnosti je vypsána v textovém okně.



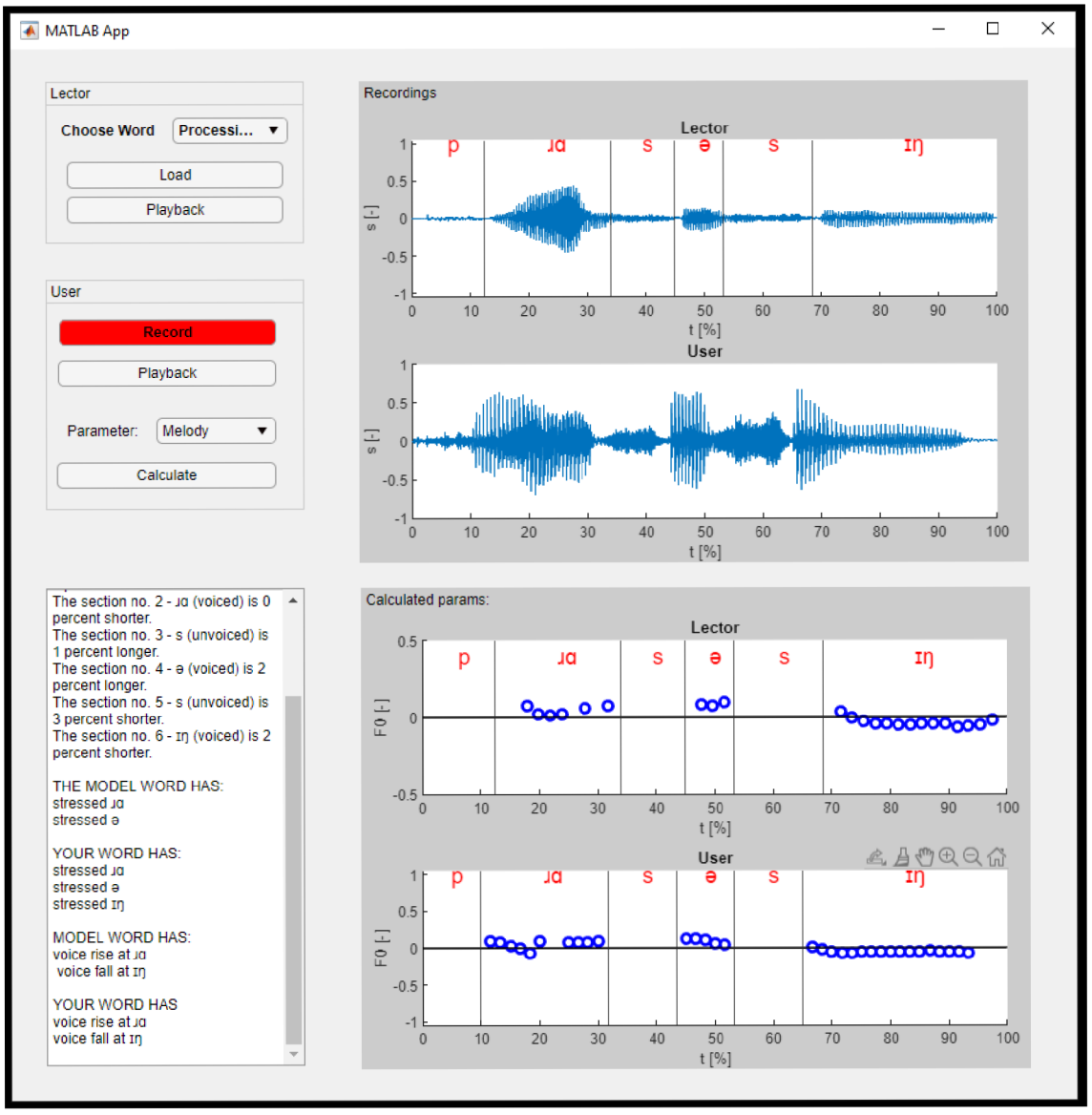
Obr. 6.4: Duration

Následujícím analyzovaným parametrem je přízvuk - *Stress*. Po stisku tlačítka *Calculate* se opět vykreslí grafická a verbální zpětná vazba. Na grafech lze pozorovat odlišnosti v přízvuku na již rozčleněném signálu; pokud v průběhu dílčího fonematického úseku funkce E překročí zobrazený práh střední hodnoty, je úsek klasifikován jako přízvučný. Verbální zpětná vazba v textovém okně srovnává přízvučnost fonematických úseků řečových signálů lektora a uživatele.



Obr. 6.5: Stress

Posledním z možných parametrů k analýze je melodie - *Melody*. Je vykreslen průběh melodie společně s prahem střední hodnoty tohoto parametru; z jeho překročení jsou identifikovány vzrůsty a poklesy hlasu. Verbální zpětná vazba je pak srovnává ve stejném duchu jako u přízvuku.



Obr. 6.6: Melody

7 Závěr

Závěrem lze prohlásit, že řady cílů vytyčených v zadání práce bylo dosaženo. Při vypracování teoretické části byla vynaložena snaha o dosažení kompromisu mezi přijatelným rozsahem výkladu a jeho podrobností, a to obzvláště u lingvistických pojmů, jejichž znalost je nezbytná pro obory zpracování řečových signálů a počítačové lingvistiky. Značný čas zabralo kompletování rešerše současně dostupných CALL nástrojů; průzkum byl prováděn za pomoci velmi rozsáhlé skupiny až 90 odborníků z MUNI (nebo jiných organizací), kteří byli kontaktováni skrze odborné předměty studijních programů představených na webových stránkách jejich univerzity. Díky tomu bylo navázáno množství velmi zajímavých kontaktů, jichž může být do budoucna využito.

V kapitolách věnujících se zpracování řečových signálů byly ustaveny vztahy mezi parametry reálné řeči, jako jsou přízvuk, melodie atd., a příznaky řečových signálů. Na základě těchto vztahů byly v prostředí MATLAB ver. 2020 naprogramovány a představeny algoritmy pro určování těchto příznaků a jejich interpretaci. Vzhledem k zaměření MATLABu (jakožto pokročilého kalkulátoru spíše než programovacího jazyka) byl před formalitami programování a optimalizací primárně kladen důraz na přehledné a jasně interpretovatelné výsledky, které jsou ve formě grafů postupně zobrazovány. Ty dále otevírají cestu k široké praktické využitelnosti koncipovaného programu. Úspěch v této oblasti dokládá umístění na studentské odborné soutěži EEICT 2022 v Brně a účast na konferenci TELFOR 2021 v Bělehradě; při obou příležitostech byly prezentovány dílčí části programu v různých vývojových fázích. V práci byly rovněž představeny koncepty metod a funkcionalit pro budoucí vývoj programu, jejichž cílem je zejména radikální zvýšení jeho spolehlivosti.

V souvislosti se spolehlivostí a robustností je třeba říci, že představený program je stále na úrovni prototypu. Hlavním cílem i přínosem mělo být otestování metod zpracování signálu pro získání jasné představy o jejich reálném využití v budoucím plně funkčním nástroji. Celý program je velmi náchylný na řadu chyb a eventualit; místy nepředvídatelné chování je zcela zřejmě zapříčiněno absencí příslušných mechanismů pro ošetření situací, jako je zvýšená úroveň šumu na vstupu, nesprávná izolace řečového signálu v pětisekundové nahrávce, ale také nepřítomností fungujícího algoritmu pro kontrolu artikulace alofonů atd. Program jako takový tedy zdaleka není připraven pro skutečný provoz a bude vyžadovat nespecifikovatelné množství času na další vývoj.

Budoucí vývoj se bude ubírat směrem naznačeným v kapitole 4.4. Kromě toho bude prioritou program co nejvíce přiblížit *realtime processingu* - velmi zajímavou perspektivou tedy je přesun od MATLABu k nízkoúrovňovým jazykům, např. C. Snahou bude zpracování krátkodobých segmentů v reálném čase současně s pro-

bíhajícím nahráváním a plynulé vykreslování dílčích parametrů výslovnosti. Tento koncept souvisí zejména se snahou o řešení nedostatků současně dostupných CALL nástrojů, které jsou velmi netransparentní. K tvorbě rozsáhlejšího slovníku také pravděpodobně bude nutná spolupráce s jazykovými experty. Tato nutnost se projevila zejména u řešení problematiky pohyblivého přízvuku angličtiny v kapitole 4.4.3.

Literatura

[1] DAVIES, G. *CALL (computer assisted language learning).* [online]. 2016 Dostupné z URL: <http://www.llas.ac.uk/resources/gpg/61>

[2] THOMAS, M.; REINDERS, H.; WARSCHAUER, M. *Contemporary Computer-Assisted Language Learning.* Contemporary Studies in Linguistics, A&C Black, 2012. ISBN 1441134506, 9781441134509

[3] RICHTER, R. *Multimedia im Phonetikunterricht. Programmangebote und Anwendungsperspektiven.* Informationen Deutsch als Fremdsprache, ročník 25, č. 5, 1998: 577-589.

[4] FATHALI, S.; EMADI, A. *CALL Research in Iran: An Integrative Review of the Studies between 2007 and 2019.* Computer Assisted Language Learning, ročník 22, č. 3, 2021: 33-51.

[5] DELCLOQUE, P. *The History of Computer Assisted Language Learning: Web Exhibition.* [online]. ICT for Language Teachers. ICT4LT. October 3. Dostupné z URL: <https://eurocall.webs.upv.es/textos/history_of_call.pdf>

[6] *Neurológ Stránský: Ľuďom klesá IQ, nevedia riešiť problémy a stávajú sa otrokmi technológií.* EDUWORLD [online]. Dostupné z URL: <https://eduworld.sk/cd/jaroslava-konickova/4666/neurolog-stransky-technologie-rozhovor>

[7] BARNETT, M. *Literacy, technology and 'technological literacy'.* Int J Technol Des Educ 5, 1994: 119-137

[8] JAMES, A. R. *Review of The Teaching of Pronunciation, by P. MacCarthy.* TESOL Quarterly, ročník 14, č. 2, 1980: s. 246-250.

[9] *Učte se cizí jazyky tak jako miminka.* Magazín Univerzity Karlovy [online]. Dostupné z URL: <https://www.ukforum.cz/rubriky/veda/511-ucte-se-cizi-jazyky-tak-jako-miminka>

[10] KREJČOVÁ, E. *Problematika osvojování druhého jazyka – teorie a praxe (na základě výuky bulharštiny jako cizího jazyka v slovanském kontextu).* Brno: Masarykova univerzita, Filosofická fakulta, 2018. Habilitační práce

[11] ÇAKIR, İ.; BAYTAR, B. *Foreign language learners' views on the importance of learning the target language pronunciation.* Journal of Language and Linguistic Studies, ročník 10, č. 2, 2014: 99-110.

[12] NURULLAYEVNA, S. N. *THE KEY OF EFFECTIVE COMMUNICATION IS PRONUNCIATION.* European Journal of Humanities and Educational Advancements, ročník 1, č. 4, 2020: 5-7.

[13] STRACHOŇOVÁ, H. *Český znakový jazyk jako samostatný komunikační systém.* Celostátní foniatrický seminář: Následná rehabilitační péče o dítě se sluchovým postižením, 2019.

[14] FREIWALD, M. *Vytvoření výpočtového modelu lidského vokálního traktu.* Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2018. 61 s. Vedoucí bakalářské práce Ing. Pavel Švancara, Ph.D.

[15] *PHO_013 - Linguistic Micro-Lectures: Formants - YouTube.* [online]. Dostupné z URL: <`https://www.youtube.com/watch?v=sqfhA9mwAuA&ab_channel=TheVirtualLinguisticsCampus`>

[16] SKARNITZL, R.; VOLÍN, J. *Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny.* Akustické listy, ročník 18, č. 1, 2012: 7-11.

[17] PETERSON, G. E.; BARNEY, H. L. *Control methods used in a study of the vowels.* The Journal of the acoustical society of America, ročník 24, č. 2, 1952: 175-184.

[18] GOMULKIN, D. *The variance of F1/F2 ratio within and across different vowel types.* 2017 10.13140/RG.2.2.35919.07840.

[19] *The world's best way to improve your English pronunciation.* ESLASPEAK [online]. Dostupné z URL: <`https://elsaspeak.com/en/`>

[20] HOLAJ, R.; POŘÍZKA, P. *L2 Czech annotation for automatic feedback on pronunciation.* Journal of Linguistics, ročník 72, č. 2, 1952: 510-519.

[21] *Duolingo - Nejlepší způsob na světě, jak se naučit jazyk.* DUOLINGO [online]. Dostupné z URL: <`https://cs.duolingo.com/efficacy`>

[22] HODAŇ, D. *Konverze hlasu.* Brno: Vysoké učení technické v Brně, Fakulta informačních technologií, 2016. Vedoucí bakalářské práce Černocký Jan.

[23] *MathWorks - Makers of MATLAB and Simulink - MATLAB & Simulink* [online]. Dostupné z URL: <`https://www.mathworks.com/help/signal/ug/formant-estimation-with-lpc-coefficients.html`>

[24] *PLATO – Illinois Distributed Museum. Illinois Distributed Museum* [online]. Dostupné z URL: <https://distributedmuseum.illinois.edu/exhibit/plato/>

[25] WITT, S. M. *Automatic error detection in pronunciation training: Where we are and where we need to go.* International Symposium on automatic detection on errors in pronunciation training, 2012.

[26] REED, M.; LEVIS, J. *The Handbook of English Pronunciation.* Hoboken: John Wiley & Sons, 2019.

[27] TAUBER, M. *Czenglish: Common Mistakes in English Pronunciation Made by Czech People.* Plzeň: Západočesná univerzita v Plzni, Bakalářská práce, 2017.

[28] MALUCHA, J. *Computer Based Evaluation of Speech Voicing for Training English Pronunciation.* 29th Telecommunications forum TELFOR, 2021.

[29] BÄCKSTRÖM, T. *Linear Predictive Modelling of Speech – Constraints and Line Spectrum Pair Decomposition.* 951-22-6946-5.

[30] SCOTT, I. *phonetics - Vowels | Britannica. Encyclopedia Britannica | Britannica.* [online]. Dostupné z URL: <https://www.britannica.com/science/phonetics/Vowels>

[31] *ORTOEPIE | Nový encyklopedický slovník češtiny* [online]. Copyright © Masarykova univerzita, Brno 2012 Dostupné z URL: <https://www.czechency.org/slovnik/ORTOEPIE>

[32] *Internetová jazyková příručka* [online]. Copyright © Dostupné z URL: <https://prirucka.ujc.cas.cz/>

[33] SIGMUND, M. *Analýza řečových signálů.* Skripta FEKT VUT v Brně. Brno: MJ servis, 2000