

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

DATABÁZE NAHRÁVEK PRO DETEKCI HLASOVÉ AKTIVITY

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

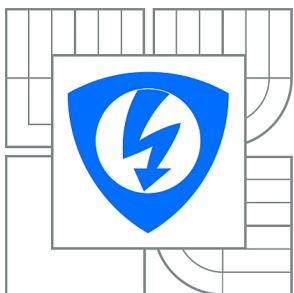
PAVEL PELIKÁN

BRNO 2011



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ**

ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

DATABÁZE NAHRÁVEK PRO DETEKCI HLASOVÉ AKTIVITY

DATABASE OF RECORDINGS FOR DETECTION OF VOICE ACTIVITY

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

PAVEL PELIKÁN

VEDOUcí PRÁCE

SUPERVISOR

Ing. IVAN MÍČA

BRNO 2011



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Pavel Pelikán

ID: 119570

Ročník: 3

Akademický rok: 2010/2011

NÁZEV TÉMATU:

Databáze nahrávek pro detekci hlasové aktivity

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s problematikou detekce hlasové aktivity (VAD) obecně a podrobněji se zabývejte požadavky na databázi nahrávek pro testování výkonnosti VAD algoritmů. Vyberte z existujících nástrojů pro označování nahrávek. Na základě získaných znalostí navrhnete a vytvořte databázi akusticky problematických izolovaných slov, čteného slova a spontánních promluv. Součástí databáze má být stručný popis zvoleného formátu značek a možnost konverze do maticového formátu vhodného pro načtení v Matlabu.

DOPORUČENÁ LITERATURA:

- [1] PSUTKA, J., et al. Mluvíme s počítačem česky. Praha: Academia, 2006. 752 s. ISBN 80-200-1209-1.
- [2] KACUR, J. The Concept of Task Specific Speech Database for VAD Systems. In 48th International Symposium ELMAR-2006., s. 155-158. ISBN 953-7044-03-3.

Termín zadání: 7.2.2011

Termín odevzdání: 2.6.2011

Vedoucí práce: Ing. Ivan Míča

prof. Ing. Kamil Vrba, CSc.

Předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Tato práce se zabývá detekcí řeči (VAD) a požadavky na vytvoření databáze. Byl vybrán jeden z existujících nástrojů na označování nahrávek. Na základě získaných znalostí byla vytvořena databáze s izolovanými slovy, větami, čteným textem a spontánní řečí. Praktická část se skládá z podrobného popisu vytvoření databáze a vytvoření značek. Dále se práce zabývá konverzí značek do Matlabu. Součástí jsou také některé pomocné skripty pro práci se značkami. Vypracovaná databáze byla vytvořena v bezodrazové komoře a obsahuje nahrávky od 16 řečníků.

KLÍČOVÁ SLOVA

VAD, databáze nahrávek, označování nahrávek, konverze značek, mluvčí, spontánní řeč, izolovaná slova, izolované věty

ABSTRACT

This thesis deals with voice activity detection (VAD) and requirements for creating a speech database. One of the existing tools for marking recordings was chosen. On the basis of the gained knowledge, database of isolated words, sentences, text and spontaneous speech was created. The practical part consists of a detailed description of database and mark creation. Furthermore, the thesis deals with the conversion of marks into Matlab. There are also some auxiliary scripts for operations with marks. Prepared and a database was created in an anechoic chamber and includes recordings from 16 speakers.

KEYWORDS

VAD, database of recordings, conversion marks, speakers, spontaneous speech, reading texts, isolated words, isolated sentences

PELIKÁN, Pavel *Databáze nahrávek pro detekci hlasové aktivity*: bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2011. 38 s. Vedoucí práce byl Ing. Ivan Míča

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Databáze nahrávek pro detekci hlasové aktivity“ jsem vypracoval samostatně pod vedením vedoucího bakalářské 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 bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

Brno

.....

(podpis autora)

Rád bych touto cestou poděkoval svému vedoucímu bakalářské práce Ing. Ivanu Míčovi za výborné vedení, prospěšné konzultace a značnou pomoc při nahrávání databáze. Moje poděkování si zaslouží též všichni mluvčí, kteří se podíleli na vzniku nahrávek. Rád bych poděkoval také Mgr. Pavlu Rajmicovi, Ph.D. za přípravu šablony pro vysázení bakalářské práce v programu \LaTeX .

OBSAH

Úvod	10
1 Teoretické základy	11
1.1 Řeč	11
1.1.1 Vlastnosti mluvené řeči	11
1.1.2 Pauzy	11
1.1.3 Vymezení řeči	11
1.2 Detekce řeči	13
1.2.1 Metoda výpočtu krátkodobé energie	13
2 Databáze nahrávek	14
2.1 Přípravy	14
2.1.1 Velikost databáze	14
2.2 Průběh nahrávání	16
2.3 Závěrečná fáze	18
2.3.1 Značkování	18
2.3.2 Struktura databáze	21
2.4 Konverze formátu značek do Matlabu	23
2.5 Poznámky k databázi	24
2.5.1 Možnosti rozšíření	25
2.5.2 Zkušenosti nabyté vytvářením k databáze	25
2.5.3 Pomocné skripty	26
2.5.4 Velikost databáze	29
3 Závěr	31
Literatura	32
Seznam příloh	33
A Přepis nahrávek	34
A.1 Izolovaná slova	34
A.2 Izolované věty	35
A.3 Čtený text	36
B Náhled značek	37
B.1 Ukázka značkování spontánní řeči	37
B.2 Ukázka značkování izolovaných slov	38

SEZNAM OBRÁZKŮ

1.1	Ukázka problematického označení hranic slova	12
2.1	Graf srovnání značek ze dvou detektorů a správného označování . . .	19
2.2	Formát značkovacího souboru	21
2.3	Informace k databázi	22
2.4	Informace o řečnících	22
2.5	Struktura databáze	22
2.6	Matice vzniklá konverzí z TextGridu u čteného textu	24
2.7	Matice vzniklá konverzí z TextGridu u izolovaných slov	24
2.8	Graf podílu nahrávek dle obsahu	29
B.1	Náhled značek	37
B.2	Náhled značek izolovaných slov	38

SEZNAM TABULEK

2.1	Parametry databáze	15
2.2	Nahrávací zařízení	16
2.3	Velikost databáze	30
A.1	Izolovaná slova	34
A.2	Izolované věty	35
A.3	Čtený text	36

ÚVOD

Ve své práci se zabírám problematikou detekce hlasové aktivity a dále zejména požadavky na vytvoření databáze nahrávek k účelu testování VAD algoritmů.

Na základě získaných zkušeností s vytvořením prototypu databáze v semestrálním projektu a také na základě konzultací, jsem vytvořil databázi, která obsahuje různé typy promluv. Konkrétně se jedná o nahrávky izolovaných slov, vět, čteného textu a spontánní promluvy. Nahrávky této databáze vznikly v bezodrazové komoře VUT. K nahrávkám je připojen soubor značek, jeho formát je popsán v této práci.

Tento dokument je rozdělen do dvou hlavních kapitol. V první jsou uvedeny základní poznatky ohledně řeči a její detekce, je nastíněna jedna metoda detekce hlasové aktivity. V druhé kapitole je rozebrána samotná databáze nahrávek, její vznik od příprav přes nahrávání v bezodrazové komoře až po zpracování nahraných dat. Podstatnou součástí práce jsou značky, popisující hlasovou aktivitu, respektive ticho.

Obsahem práce je také popis a funkce zdrojového kódu ke konverzi do programu Matlab. V příloze jsou uvedeny texty sloužící jako materiál pro řečníky. Výstup této práce, databáze nahrávek, která je obsahem přiloženého média, by měla sloužit k testování efektivity jednotlivých VAD algoritmů. Její výhodou je „čistota“ nahrávek, neboť nahráváním v bezodrazové komoře bylo zamezeno vzniku rušivých elementů a šumu.

1 TEORETICKÉ ZÁKLADY

1.1 Řeč

Na úvod je vhodné seznámit se s některými základními znalostmi a pojmy v oblasti řeči. Není cílem této práce seznámit čtenáře s tvorbou řeči, hlasovým ústrojím člověka. Řeč je nejpřirozenější způsob přenosu informací mezi lidmi. Bohužel, řeč každého člověka je odlišná a dvě stejná slova, z úst dvou lidí, mohou obsahovat mnoho rozdílů.

1.1.1 Vlastnosti mluvené řeči

Komunikace pomocí mluvené řeči je nejpřirozenější způsob přenosu informací mezi lidmi. Je přenášena komunikačním kanálem ve formě akustických vln, tzv. akustického signálu. Podstatou tohoto akustického signálu je vlnění elastického prostředí v oboru slyšitelných frekvencí. Skrývá v sobě dvě důležité složky, první je složka akustická, která představuje amplitudově-frekvenční časové spektrum a druhou složkou je lingvistická, která v sobě nese informace, které jsou pro komunikaci stěžejní (lze studovat z hlediska fonetiky, morfologie apod.). Mimo těchto dvou složek lze pozorovat další vlastnosti, které jsou vlastní mluvcímu, např. intonaci, hlasitost, barvu hlasu, artikulaci, rytmus a rychlost řeči. Součástí mluveného projevu, zejména spontánního, jsou také chyby a přerušky.

1.1.2 Pauzy

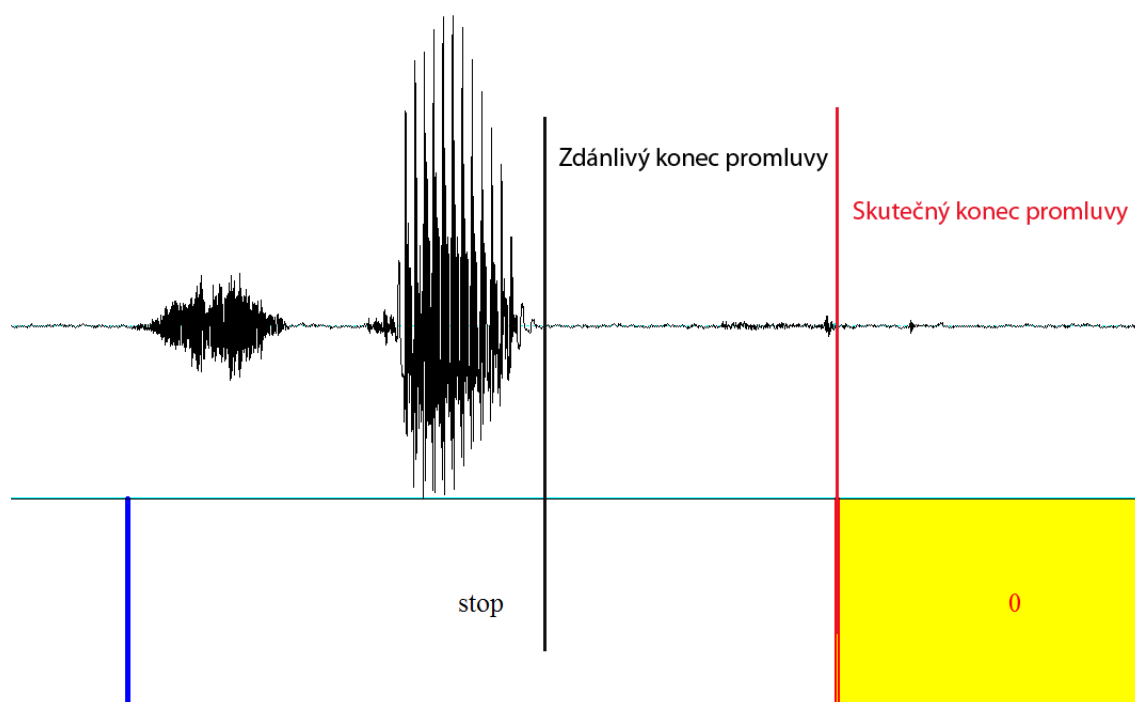
Velmi důležitou součástí samotné řeči jsou pasáže, které jsou indikátory hraničních bodů promluvy. Pauzy ovlivňují nejen časování, ale také srozumitelnost. Délky pauz jsou různé, může se jednat o mezery, kterými se oddělují jednotlivé větné úseky nebo celé věty. Umístění, délka, počet pauz a typ pauzy velmi ovlivňuje plynulost a přirozenost řeči. Velkou roli hrají pauzy při detekci hlasové aktivity. Pokud je totiž pauza doplněna či dokonce vyplněna např. vzdechnutím, mlasknutím, zakašláním apod., může detektor označit takovou pauzu jako řeč. Jasně vymezení pauzy a řeči může být v mnoha případech značně problematické.

1.1.3 Vymezení řeči

Vymezením řeči je myšlena technika ručního značkování nahrávek nejčastěji pomocí popisových os. Stanovení hraničních bodů promluvy je důležité zejména v případech rozpoznávání izolovaných slov a krátkých frází, nicméně i v případech souvislé řeči

je třeba rozhodnout, kdy má detektor řeči překládat výsledky klasifikace. Existují jevy, které mají vliv na kvalitu určení řeči [1]. Jsou to zejména tyto:

- Akustické pozadí, ve kterém je řeč nahrána. Patrné je to zejména na nahrávkách, kdy je v pozadí obsažen zvuk jako např. jiná řeč, větrák, automobil apod. Takovým jevům se tato práce vyhnula volbou bezodrazové komory.
- Rušení způsobené přenosovým kanálem.
- Nedokonalá artikulace, která způsobuje již zmiňované doprovodné zvuky, které řečník produkuje během promluvy (mlaskání, cvakání a klikání různými částmi artikulačních orgánů, hlasité vzdychání apod.), přičemž energetická úroveň těchto zvuků může být srovnatelná s úrovní řeči.
- Speciálním případem jsou slova, u kterých je obtížné definovat začátek a konec. Konkrétní jevy jsou uvedeny v kapitole 2.1.1. Pro ilustraci takového těžko definovatelného případu je přiložen obrázek 1.1. V tomto případě se jedná o průběh slova „stop“ a jeho měkkou okluzivu „p“ na konci slova. Průběh takové hlásky je velmi nevýrazný, stejně tak jeho energie.



Obr. 1.1: Ukázka problematického označení hranic slova

1.2 Detekce řeči

Jelikož se tato práce zabývá databází nahrávek pro detekci hlasové aktivity, bylo vhodné se seznámit se samotnou detekcí hlasové aktivity VAD. Konkrétně byl vybrán jeden z několika existujících algoritmů, který je založen na metodě výpočtu krátkodobé energie. Detekce řeči je známá pod označením VAD (voice activity detection).

Algoritmy detekce lze nalézt v mnoha různých aplikacích pro zpracování signálu. Nejznámější detekční metody jsou založeny na zmiňovaném výpočtu krátkodobé energie, dále na průchodu nulou, spektrálních změnách, detekce základního tónu apod.

1.2.1 Metoda výpočtu krátkodobé energie

Algoritmus této metody je poměrně jednoduchý. Nejprve dojde k segmentaci signálu na malé části, segmenty. Délka jednotlivých segmentů musí být volena dostatečně dlouhá, aby bylo zaručeno, že budou bezchybně změřeny jednotlivé parametry. Na druhé straně musí být ale dostatečně krátká, aby bylo možno jednotlivé segmenty aproximovat konstantními hodnotami. Vzhledem k vlastnostem lidského ucha je volena délka segmentu nejčastěji 10 až 30 ms. Pro každý segment je vypočítána jeho krátkodobá energie. Funkci krátkodobé energie signálu [1] lze definovat vztahem:

$$E_n = \sum_{k=-\infty}^{\infty} [s(k)w(n-k)]^2, \quad (1.1)$$

kde $s(k)$ je vzorek signálu v čase k a $w(n)$ reprezentuje příslušný typ okénka. Vypočtená hodnota signálu je následně porovnávána s prahem a na základě tohoto porovnání se rozhodne, zda je obsažena řeč či nikoli. Práh je tedy vlastně funkce s výstupem pravda/nepravda. Pravda představuje řeč, nepravda pauzu. Pro vyhodnocení metody je potřeba provést tzv. označování nahrávky. Jedná se o označování nahrávky na základě poslechu a průběhu signálu. Jsou označeny části, které obsahují řeč a části, které jsou považovány za pauzy. Existuje celá řada nástrojů pro práci se zvukem, které takovouto techniku zvládnou. Například jsou to volně dostupné programy jako Audacity či Praat. Praat má dokonce implementován svůj detekční algoritmus, u kterého lze nastavit parametry detekované řeči.

Na závěr se porovnává detekce spočítaná s ručně označovaným signálem a zhodnotí se výsledky. Problém může nastat nahrávkami s natolik šumivým prostředím, kde úroveň šumu dosahuje hodnot prahu, je tedy nutno uvážit nastavení velikosti prahu.

2 DATABÁZE NAHRÁVEK

Na základě zkušeností získaných v semestrálním projektu byla navržena databáze, jejíž rozměry byly částečně upraveny v průběhu příprav. Tyto úpravy se týkaly zejména velikosti databáze. Cílem bylo najít kompromis mezi velikostí databáze a únosností času a práce, které bylo nutné věnovat zpracování a zejména značkování nahrávek. Vznik kompletní databáze lze rozdělit na tři hlavní části. Záležitostí té první je promyslet kolik řečníků bude zapotřebí, v jakém prostředí se bude nahrávat, na jaké zařízení, v jakém formátu apod. Druhá fáze je samotné nahrávání a třetí je finální zpracování nahrávek a vytvoření potřebné dokumentace.

2.1 Přípravy

Ve fázi příprav bylo zapotřebí promyslet materiál ke vzniku nahrávek, promyslet formát materiálu, vytisknout podklady přijatelné formy pro řečníky. Součástí příprav bylo také zprovoznění techniky v učebně a provést několik zkušebních nahrávek. Jako první v této fázi padlo rozhodnutí o prostředí, v němž se bude nahrávat. Byla zvolena bezodrazová komora laboratoři PA-327 na Ústavu telekomunikací. Ač nahrávky mohou znít nepřírozně z důvodu absorpce akustických vln na stěnách bezodrazové komory, tak velkou výhodou pro další zpracování je „čistota“ nahrávek. Tím je myšleno minimální šumové pozadí.

2.1.1 Velikost databáze

Dále bylo rozhodnuto o konečném počtu mluvčích. Z původně zamýšleného počtu 20 mluvčích, se změnil počet na konečných 16. I tak se jedná o databázi střední velikosti. Těchto vybraných 16 řečníků bylo osloveno a byl domluven termín nahrávání. Důležitou součástí přípravy bylo vytvoření samotného obsahu databáze. S ohledem na zadání práce obsahuje databáze těchto 5 celků:

Izolovaná slova

Bylo vybráno 100 slov. Nešlo o náhodný výběr, nýbrž o výběr slov, u kterých je obtížné definovat začátek a konec [1]. Jedná o následující případy:

- jsou na začátku či na konci promluvy měkké okluzivy, tj. /p/, /t/, /k/;
- jsou na konci promluvy znělé okluzivy /b/, /d/, /g/;
- jsou na začátku či na konci neznělé frikativy, např. /f/, /s/, /š/, /ch/;
- jsou na konci promluvy znělé frikativy /v/, /z/, /ž/, /h/;
- jsou na konci promluvy nosní souhlásky /m/, /n/, /ň/;

Každý tento případ je obsažen ve zmiňovaných 100 slovech minimálně pětkrát, což je dostačující pokrytí všech jevů. Těchto 100 slov bylo seřazeno dle abecedy a uspořádáno do sloupců, z důvodu přehlednosti, viz příloha A.1.

Izolované věty

V této sekci bylo zvoleno náhodným výběrem 12 různých vět. Většinou jde o věty z tiskovin či vymyšlené věty viz příloha A.2.

Izolované věty s hlukovým pozadím

Další fází nahrávání byly opět tytéž izolované věty, které ovšem byly čteny se sluchátky, do kterých byl puštěn záznam hlukového prostředí. Takto byl simulován Lombardův efekt, který je spojen se změnami řeči za účelem lepší srozumitelnosti v hlučném prostředí. Toto hlučné prostředí však není součástí nahraných dat.

Čtený text

V této sekci byla vybrána poměrně krátká povídka, jejíž čtení trvalo v průměru 2 minuty a 20 sekund, viz příloha A.3.

Spontánní promluva

K této fázi spontánní řeči byla připravena krátká pohádka, která byla puštěna řečníkovi těsně před nahráváním. Po jejím zhlédnutí byl řečník připraven ji převyprávět vlastními slovy bez jakýkoliv podkladů. Cílem bylo získat od každého řečníka 2-3 minuty spontánního monologu.

Všechny podklady byly vytištěny a připraveny ve více kopiích. Parametry databáze jsou zaznamenány v přehledné tabulce 2.1.

Počet řečníků	16 (10 mužů, 6 žen)
Počet izolovaných slov	100
Počet vět	12
Počet vět s hlukem	12
Čtený text	299 slov (Ø 2:20)
Spontánní řeč	Ø 2:31

Tab. 2.1: Parametry databáze

2.2 Průběh nahrávání

Po dokončení materiálu a také smluvení konkrétního času nahrávání se přistoupilo k přípravě nahrávacích prostor. Přibližně hodinu před příchodem prvního řečníka či skupiny se začaly připravovat mikrofony, zesilovač, nahrávací software a občerstvení pro mluvčí. Konkrétní informace o nahrávacím zařízení jsou uvedeny v tab. 2.2.

Mikrofon	RØDE NTG-1, UTKO inv. c 1000134824
Zesilovač	Behringer ULTRAGAIN PRO, model MIC2200
Zvuková karta	LAYLA by echo, 20-bit multitrace recording system
Nahrávací software	Cubase 4
dále	držák na mikrofon, spojovací kabely, sluchátka

Tab. 2.2: Nahrávací zařízení

Obvykle přistupovala k jednomu nahrávacímu cyklu skupina 4 lidí. Po příchodu takové skupiny následovalo:

1. Uvítání a organizace

V rámci uvítání byli mluvčí přivedeni na pracoviště a usazeni. Pracovištěm je myšlena učebna PA-327 na Ústavu telekomunikací, jejíž součástí je bezodrazová komora. Bylo jim nabídnuto občerstvení. Dále byli seznámeni se zevrubným plánem nahrávání, v případě dotazů a nejasností vedoucí nahrávání vysvětlil, co nebylo zřejmé a mohlo se přejít ke konkrétním nahrávacím fázím. V bezodrazové komoře dostal každý mluvčí pokyny, jakou pozici zaujmout. Vzdálenost mluvčího od mikrofonu se pohybovala okolo 0,5 m. Mluvčí byl otočen k mikrofonu čelem a stál. Když byl přítomen i vedoucí nahrávání, držel se stranou od mikrofonu a mluvčího, ale byl připraven zasáhnout v případě nesrovnalostí či nespokojenosti s průběhem nahrávání. Každá konkrétní nahrávka byla ukládána do samostatné stopy, přehledně uložena do správného adresáře a průběžně zálohována. Před vstupem dalšího mluvčího byla vždy načtena nová stopa.

2. Izolované věty

V této části přistupovali jednotliví mluvčí do bezodrazové komory. Každý dostal dokument s jednotlivými větami. Do komory přistupoval za asistence vedoucího nahrávání, který dotyčnému dával pokyny, v jakém pořadí věty číst. Byl tak navazován kontakt s mluvčím, který byl podstatný ve fázi izolovaných vět s hlukem. Po bezchybném přečtení všech vět se mluvčí prostřídali.

3. Izolovaná slova

Opět dostal každý podklady s izolovanými slovy a po jednom přistupovali do bezdrazové komory. Byli požádáni, aby všechna slova četli přibližně se stejnou pauzou. Přítomnost vedoucího nahrávání nebyla nezbytná, ovšem i přesto byl přítomný. Nakonec se to ukázalo jako prozíravé, protože vedoucí ještě v komoře korigoval velikost pauz mezi jednotlivými slovy. V případě chyby bylo slovo přečteno znovu a při zpracování byla chyba vystřížena.

4. Poslech povídky

V této části byla nejprve puštěna nahrávka od profesionálního řečníka. Mluvčí měli tudíž lepší možnost seznámit se s daným textem a inspirovat se ve svém vlastním přednesu.

5. Čtený text

Po poslechu se řečníci po jednom, tentokrát bez asistence vedoucího, odebrali do komory, kde přečetli daný text.

6. Izolované věty s hlukem

Po nahrávání čteného textu celé skupiny byl v programu Cubase puštěn na výstup do komory (do sluchátek) záznam z hlukového prostředí (konkrétně to byl záznam z konference či debaty více lidí). V komoře byl přítomen samozřejmě mluvčí a také vedoucí nahrávání. Ten dal pokyn k nasazení sluchátek (nasadil si je řečník i vedoucí) a podobně jako v nahrávání vět bez hluku, nechal číst jednotlivé věty v náhodném pořadí. Byl tak záměrně vyvoláván kontakt s mluvčím. Někteří mluvčí se snažili komunikaci zjednodušit gestikulací, vedoucí však na tuto snahu nerefletoval a verbálně opakoval své požadavky. V případě, že vedoucí nerozuměl promluvě, dal pokyn k zopakování položky. Tato taktika motivovala mluvčí k větší srozumitelnosti promluv.

7. Zhlédnutí pohádky

Předposlední fází bylo zhlédnutí krátké pohádky. Záměrně byla vybrána nepříliš známá pohádka. Projekce trvala asi 9 minut. Na jejím základě se mohlo přistoupit k finální fázi nahrávání.

8. Spontánní řeč

V komoře byly opět přítomny dvě osoby, mluvčí i vedoucí nahrávání. Vedoucí uvedl nahrávku a předal slovo mluvčímu, z čehož vyplývá, že vedoucí je součástí nahrávek spontánní řeči. Dále byl přítomen pro případné korigování vyprávění. Pokud mluvčí zabíhal příliš do podrobností, byl šetrně usměrněn a naopak pokud byl příliš strohý, vedoucí měl snahu příběh rozvést. Cílem bylo získat dvě až tři minuty spontánní řeči.

9. Poděkování

Po nahrání posledního mluvčího proběhlo poděkování a rozloučení se všemi mluvčími.

Nahrávání bylo rozděleno do několika dnů, přičemž v jeden den bylo možné získat materiál od více lidí, pokud šlo o počet větší než 4, byli mluvčí rozděleni do více skupin. Při počtu 5 a více by nastaly poměrně dlouhé prodlevy při čekání, protože pokud byl jeden mluvčí v komoře, ostatní čekali, než přijdou na řadu. Volnou chvíli mohli vyplnit studováním dalšího nahrávacího cyklu či volnou zábavou.

Po skončení nahrávání bylo samozřejmě nutné uklidit všechny použité přístroje a pomůcky a navrátit pracoviště do výchozího stavu.

Vedoucím nahrávání je v této kapitole označen autor práce.

2.3 Závěrečná fáze

Závěrečnou fází můžeme označit celou proměnu a zpracování všech nahraných dat do konečné podoby, což je výstup této práce, databáze nahrávek. Součástí je tedy přenesení všech dat z učebny a její následné zpracování, třídění a stříhání, značkování a vytváření dokumentace. Všechny nahrávky jsou uloženy ve formátu PCM s příponou `.wav`. Jedná se o bezztrátový formát s vysokou kvalitou zvuku, nevýhodou může být větší velikost souborů, což se samozřejmě projevilo větší velikostí celé databáze. Všechny nahrávky jsou navzorkovány frekvencí 44 100 Hz.

V souvislosti se stříháním je vhodné uvést:

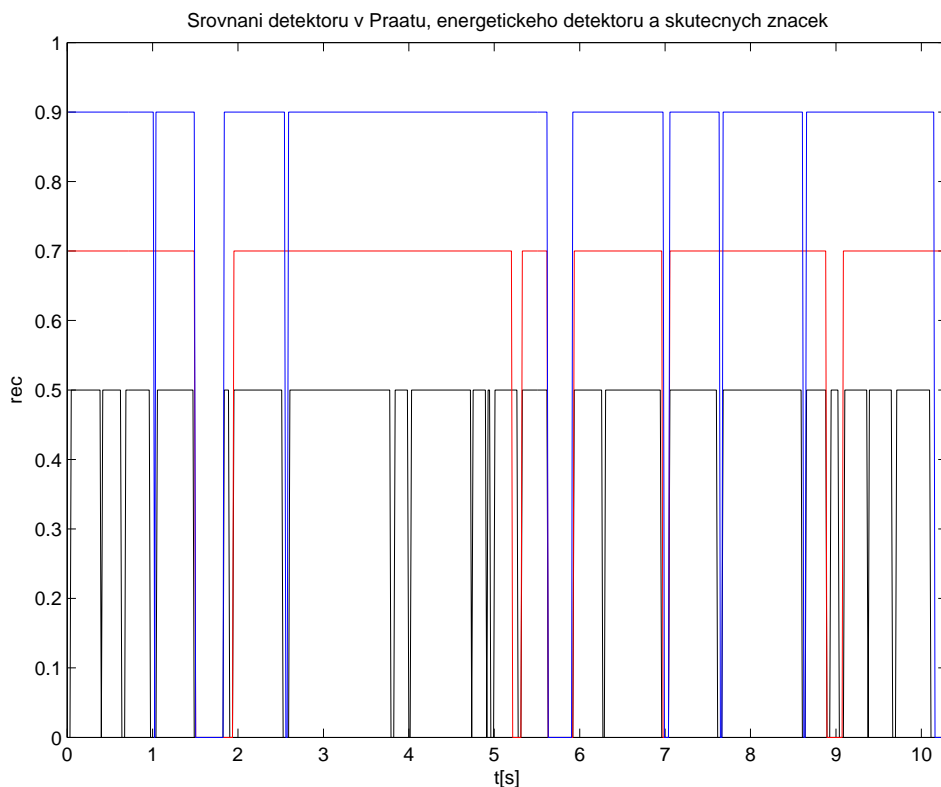
- Izolované věty – vybraných 12 vět bylo rozstříháno a jsou v databázi uloženy jako soubory `1veta.wav` až `12veta.wav`. Stejně tak u izolovaných vět s hlukem.
- Izolovaná slova – protože 100 souborů nahrávek a k tomu navíc 100 textových souborů se značkami by se jevilo jako velmi komplikované řešení, byla zvolena varianta 100 slov v rámci jednoho souboru. V případě potřeby je možné jednotlivá slova vystříhat, viz pomocný skript v kapitole 2.5.3. Z originální nahrávky byly vystříhány přeřky, případně některé markantní doprovodné nežádoucí zvuky (například zakašlání).
- Čtený text – u čteného textu k žádnému vystřihávání nedocházelo. Tato nahrávka je uvedena jako celek od začátku čtení až ke konci.
- Spontánní řeč – přirozeně, ani zde nedocházelo k žádnému vystřihávání.

Po rozřídění nahrávek do složek s jednotlivými řečníky následovala velmi podstatná součást této práce, značkování nahrávek. Pro její důležitost je zařazena do samostatné podkapitoly.

2.3.1 Značkování

Pro značkování byl vybrán softwarový nástroj Praat verze 5.1.05. Jelikož by značkování bylo samo o sobě velmi náročné, byl využit nástroj programu Praat, který dovede

na základě několika parametrů detekci zvuku, resp. ticha provést. Nalezneme ho pod Annotate → to TextGrid (silences). Vstupními parametry jsou například: práh ticha v dB, minimální intervaly trvání ticha a zvuku, popisky ticha a popisky zvuku. Tímto nástrojem bylo provedeno jakési předznačkování. Alternativou za takovýto nástroj v Praatu bylo předznačkování pomocí energetického detektoru. Tato alternativa se však jevila komplikovaná a časově náročnější než využití detektoru v Praatu. Pro srovnání obou těchto detektorů slouží obrázek 2.1, na kterém je graf s průběhy obou detektorů a ručního označování. V Praatu byly použity výchozí parametry detektoru a práh v energetickém detektoru byl záměrně vybrán tak, aby jeho výstup byl co nejpřesnější (ve srovnání s označovanou nahrávkou). Detekována byla náhodně vybraná věta z databáze. Černou barvou je vyznačen výstup energetického detektoru, červenou označování z detektoru Praatu a modrou ruční označování. Hodnoty na ose x 0,5; 0,7 a 0,9 představují řeč dle jednotlivých značek.



Obr. 2.1: Graf srovnání značek ze dvou detektorů a správného označování

Jak vidno z grafu, Praat dovedl vymezit řeč a ticho mnohem efektivněji a přesněji než se to podařilo energetickému detektoru. Ten se jevil velmi citlivě i při použití nízkého práhu a zobrazoval o mnoho víc pasáží s „tichem“, než v nahrávce ve skutečnosti bylo. Čili pro shrnutí, předznačkování energetickým detektorem bylo realizováno

pouze jako pokusný vzorek. Používat ho na celou databázi by nebylo vhodným řešením, už z důvodu přítomnosti efektivnějšího nástroje programu Praat, díky němuž bylo ušetřeno velké množství času a úsilí. Stanovení zmiňovaných parametrů v detektoru Praatu se lišilo u různých řečníků i typů nahrávek. Snahou bylo přiblížit se co nejvíce ideálnímu označování, toho se docílit nedalo a všechny značky musely být ručně přesunuty do správné polohy. Některé značky musely být vymazány, některé bylo nutné přidat. Jistým zrychlením značkování je osvojení si a využití klávesových zkratk, díky nimž bylo ušetřeno značné procento celkového času.

Počet os

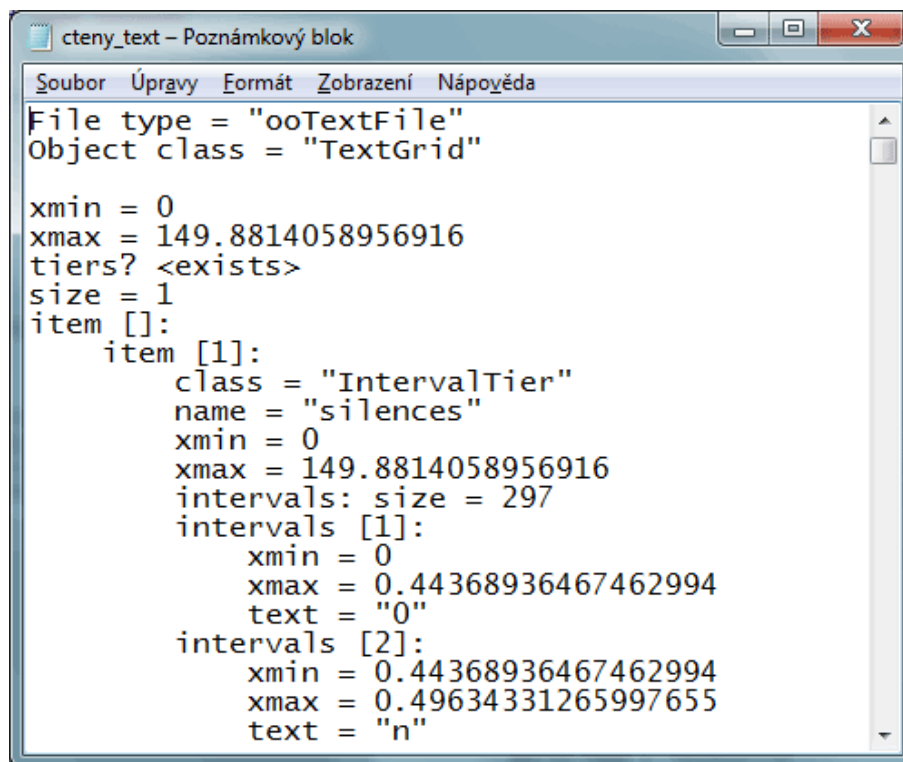
Ke každé nahrávce byla přidána jedna popisová osa. V úvaze byla také druhá osa pro druhého řečníka, a to v případě nahrávek spontánní řeči, kde se ve velmi malé míře vyskytuje vedoucí nahrávání, který uvádí hlavního mluvčího, případně vstupuje do promluvy. Ovšem těchto vstupů je naprosté minimum a druhá popisová osa by měla svůj účel, pokud by se promluvy mluvčích prolínaly, což v této databázi najdeme jen v ojedinělých případech. Proto bylo z úvahy o dvou osách odstoupeno. Jednalo by se o zbytečnou komplikaci například při konverzi značek do Matlabu.

Značky

Byly vybrány 2 stěžejní značky pro řeč a pro ticho. Řeč je vyznačena popisem „1“, naopak ticho je označeno popisem „0“. Ve spontánní řeči a čteném textu se ale vyskytují i jiné jevy, které nelze označit jako ticho, ani jako řeč. Je to zejména mlaskání, kašlán, jisté klapání, hlasité nádechy či výdechy nebo například smích. Tyto jevy jsou považovány jako neverbální. Takovéto pasáže jsou označeny popisem „n“. Pro ilustraci takovýchto značek je uveden obrázek B.1. Ve spontánní řeči je obsažen také druhý mluvčí, který je označen popisem „2“. Jiné značky jsou ovšem použity u nahrávek izolovaných slov. U nich je nejen označena řeč, ale v popisu je zároveň uvedeno, o jaké slovo se jedná. Z toho vyplývá, že součástí popisové osy je 100 odlišných značek. Pro lepší představu je uveden obrázek B.2. Samotné značkování probíhalo ve dvou etapách. Nejprve byly vytvořeny značky popisující ticho a řeč. V druhé fázi proběhla kontrola značek a v případě spontánní řeči a čteného textu byla přidána značka „n“, označující neverbální projevy.

Program Praat automaticky vytváří textový dokument s koncovkou `.TextGrid`, která má pevně daný formát. V záhlaví obsahuje položky, které pro potřeby této databáze nejsou nutné, důležitá data se vyskytují až v položkách `xmin` a `xmax`, což označuje začátek a konec nahrávky v sekundách. Další podstatná data se vyskytují v

jednotlivých „itemech“, což jsou samostatné úseky, oddělené dvěma značkami, označujícími začátek a konec (opět položkami `xmin` a `xmax`). Součástí takového itemu je položka `text`, která obsahuje popis značky, například „0“ nebo „1“. Pro ilustraci je uveden obrázek 2.2



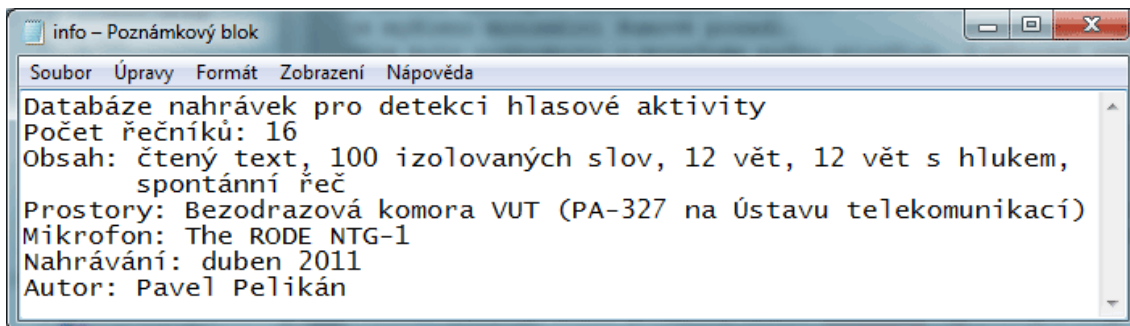
```
cteny_text - Poznámkový blok
Soubor Úpravy Formát Zobrazení Nápověda
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 149.8814058956916
tiers? <exists>
size = 1
item []:
  item [1]:
    class = "IntervalTier"
    name = "silences"
    xmin = 0
    xmax = 149.8814058956916
    intervals: size = 297
    intervals [1]:
      xmin = 0
      xmax = 0.44368936467462994
      text = "0"
    intervals [2]:
      xmin = 0.44368936467462994
      xmax = 0.49634331265997655
      text = "n"
```

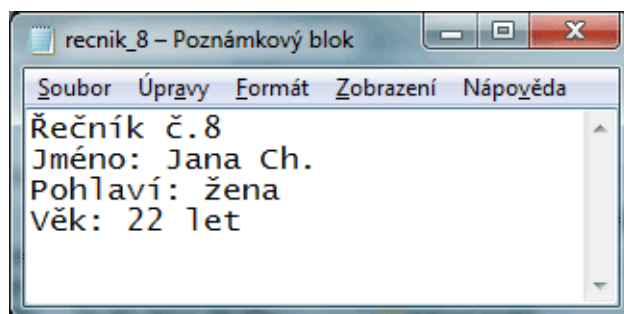
Obr. 2.2: Formát značkovacího souboru

2.3.2 Struktura databáze

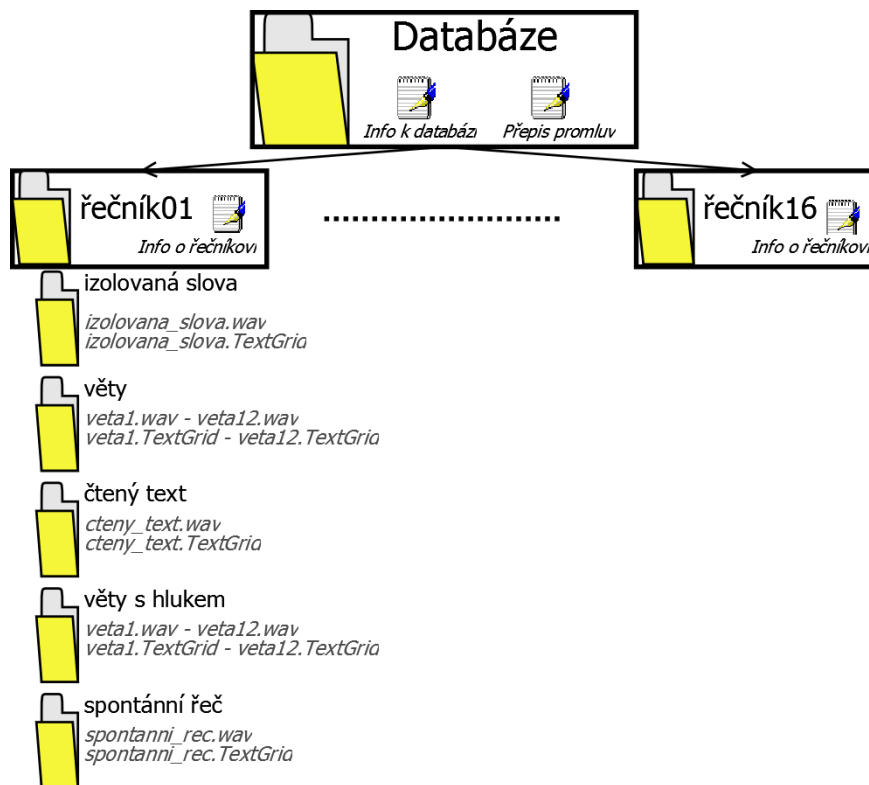
Snahou bylo udělat databázi co nejpřehlednější. A proto se složka s databází skládá z 16 složek (pojmenovaných `recnik01` až `recnik16`), dále je součástí textový soubor se základními informacemi o databázi (obrázek 2.3) a ze složky přepis, která obsahuje přepisy izolovaných slov, vět a čteného textu, viz příloha A. Každá složka s řečníkem obsahuje 5 celků nahrávek společně se značkami a mimo to je součástí také textový soubor s informacemi o řečníkovi, konkrétně o pohlaví, věku a jméně, viz obrázek 2.4. Pro lepší představu struktury celé databáze je vytvořen diagram, který je znázorněn na obrázku 2.5.



Obr. 2.3: Informace k databázi



Obr. 2.4: Informace o řečníkovi



Obr. 2.5: Struktura databáze

2.4 Konverze formátu značek do Matlabu

Jelikož formát značek z výstupu programu Praat není nejvhodnější formou pro další zpracování, zabývá se tato práce také konverzí do maticového formátu vhodného pro načtení v Matlabu. K tomuto účelu byl v programu Matlab vytvořen jednoduchý skript, který načte daný soubor `.TextGrid` (je nutné k němu uvést cestu), zahodí zbytečná data, načte začátky a konce jednotlivých značek a jejich význam. Na závěr zkontroluje, zda velikost celé nahrávky koresponduje s velikostí popisové řady. Vytvořený skript je možné si prohlédnout ve výpisu kódu:

Konverze souboru TextGrid do matice v Matlabu, konverze.m

```
1 fid = fopen(['D:\cteny_text.TextGrid']); %Doplnit cestu!
2 for i=1:11 % Nevýznamná data
3     fgetl(fid);
4 end
5 %hlavička
6 xmin=sscanf(fgetl(fid), '          xmin = %f');
7 xmax=sscanf(fgetl(fid), '          xmax = %f');
8 velikost=sscanf(fgetl(fid), '          intervals: size = %f');
9 final=cell(3, velikost);
10 for i=1:(velikost);
11     fgetl(fid);
12     final{1,i}=sscanf(fgetl(fid), '          xmin = %f');
13     final{2,i}=sscanf(fgetl(fid), '          xmax = %f');
14     final{3,i}=sscanf(fgetl(fid), '          text = "%s"');
15 end
16 if xmin==final{1,1} & xmax==final{2, velikost}; %kontrola
17     fprintf('KONVERZE PROBĚHLA V POŘÁDKU.\n');
18 else error('Chyba v TextGridu');
19 end
20 fprintf('1. řádek - začátek intervalu, 2. řádek - konec');
21 fprintf('3. řádek: popis značky');
22 out = final
```

Na základě této konverze vznikne matice o 3 řádcích a x sloupcích, x je počet itemů (značek). Jelikož Matlab standartně nedovoluje, aby matice obsahovala znaky i číslice, je použita matice `cell`. Příklad matice, která vznikne konverzí souboru se značkami ke čtenému textu, je obrázku 2.6.

První řádek označuje začátek, druhý řádek označuje konec itemu a v třetím je

	1	2	3	4	5	6	7	
1	0	0.6845	1.2917	1.3225	1.8287	2.5464	3.4501	3.5
2	0.6845	1.2917	1.3225	1.8287	2.5464	3.4501	3.5106	4.0
3	'0''	'1''	'0''	'1''	'0''	'1''	'1''	'0''
4								

Obr. 2.6: Matice vzniklá konverzí z TextGridu u čteného textu

řetězec, který popisuje, co se v daném itemu nachází, např. „1“, „0“, „n“. Problém se bohužel objevil při konverzi izolovaných slov. Důvodem byla diakritika v souboru. Ta v databázi zůstala záměrně, neboť by v některých případech nebylo jasné o jaké slovo se vlastně jedná, např. slovo „bázeň“ může bez diakritiky získat zcela jiný význam. Proto je nutné před konverzí v Matlabu odstranit diakritiku. Poté už konverze proběhne bez komplikací, viz obrázek 2.7.

	1	2	3	4	5	6	7	
1	0	0.4875	1.4069	1.9803	2.5288	3.2782	3.9946	4.6
2	0.4875	1.4069	1.9803	2.5288	3.2782	3.9946	4.6944	5.5
3	'0''	'antropolog''	'0''	'atom''	'0''	'bazen''	'0''	'bc
4								

Obr. 2.7: Matice vzniklá konverzí z TextGridu u izolovaných slov

Soubor konverze.m je součástí přiloženého média. Je uložen v adresáři „Skripty“.

2.5 Poznámky k databázi

Všichni mluvčí jsou Češi, až na výjimky studenti ve věku 21-23 let. Databáze byla nahrávána v období na přelomu března a dubna roku 2011.

Z důvodů lepší kompatibility s jiným kódováním, jsou názvy adresářů a souborů bez diakritiky. Obsah souborů týkajících se informací k celé databázi a jednotlivým řečníkům byl zachován v plnohodnotné češtině.

2.5.1 Možnosti rozšíření

Ač rozměry této databáze jsou poměrně velké, nabízí se možnost ji zvětšit o další mluvčí, například o ženy, kterých je menší množství než mužů. Nebo ji rozšířit o mluvčí jiných věkových kategorií. V rámci zachování kompatibility by bylo příhodné:

- Zachovat prostředí, čili zvolit také bezodrazovou komoru, jinak by nahrávky měly odlišnou podobu (zejména úroveň šumového pozadí), a ta by se projevila při použití různých detektorů VAD. Zachovat prostředí je v podstatě podmínka rozšíření, nikoliv doporučení. V případě nedodržení by rozšířená databáze obsahovala dva různorodé celky nahrávek.
- Použít podobné přístroje. Není to nutné, ale zejména v otázce mikrofону by bylo vhodné zvolit alespoň model s podobnými parametry jako byl užit při nahrávání této databáze.
- Zachovat formát dat. Pro zvukové nahrávky formát PCM a pro značky textový dokument. Praat používá výstup textového souboru `nazev.TextGrid`, nicméně i značky ve formátu `nazev.txt` by byly akceptovatelné. V případě použití jiného nástroje pro značkování by se změnil formát značek, což by ovlivnilo například konverzi do Matlabu.

Prostory pro rozšíření:

- Globální rozšíření (stejný obsah nahrávek).
- Rozšíření o mluvčí ženského pohlaví, kterých je menšina v této databázi.
- Rozšíření o mluvčí jiných věkových kategorií, než jsou použity v této databázi.
- Rozšíření o nahrávky v jiném jazyce, než je čeština.
- Rozšíření o nahrávky jiného obsahu, což by však dost ovlivnilo faktor kompatibility.

2.5.2 Zkušenosti nabyté vytvářením k databáze

Na základě zkušeností s procesem vytváření kompletní databáze vzniklo několik doporučení, které mohou být nápomocné například při rozšiřování databáze nebo při vytváření databáze naprosto nové. Daná doporučení jsou uvedena v následujícím textu:

K přípravě

- Pečlivě vybrat text k nahrávkám, odstranit chyby a vytisknout ho raději větším písmem.
- Vyhnout se náročným textům, raději volit text s jednoduchými větami, než s dlouhými souvětími. Pokud je vybrán obtížnější text záměrně, nechat ho mluvčím nastudovat.
- Uvědomit si, že čím budou obsáhlejší nahrávky, tím více práce bude se stříháním, značkováním.

K nahrávání

Jelikož mluvčí přišli dobrovolně je doporučováno:

- Při domluvě termínu nahrávání vyjít mluvčímu maximálně vstříc.
- Neměnit stanovený čas nahrávání tentýž den.
- Chovat se vstřícně, mluvit věcně a srozumitelně, zachovat trpělivost.
- Počítat s časovou rezervou před i po nahrávání, nespěchat.
- Při nahrávání stop a jejich ukládání se řídit dle českého přísloví „Dvakrát měř a jednou řež“. Každá nahrávka byla pro jistotu nahrávána zvlášť do jednotlivých stop. Po nahrání každé stopy se kontrolovalo, zda uložení do stopy proběhlo v pořádku.
- Navodit přátelskou atmosféru a nestresovat mluvčí, někteří mohou být před mikrofonem mírně nervozní a nepřírození.

Ke značkování

- Použít nějaký detektor na předznačkování k usnadnění práce.
- Opět se nedoporučuje spěchat, vznikají tak chyby.
- Vybrat vhodný nástroj, zjistit si jeho klávesové zkratky, které značně urychlují práci.

A na závěr doporučení – zálohovat data.

2.5.3 Pomocné skripty

V rámci práce vznikly některé pomocné skripty v Matlabu, které sloužily k urychlení práce nebo měly jinou funkci. Tyto skripty jsou součástí přiloženého média a jsou uloženy společně se skriptem ke konverzi v adresáři „Skripty“ v matlabovském formátu s příponou .m

Kontrola chyb

Během práce mohly vzniknout chyby ve značkování. Nejčastější chybou bylo opomenutí popisu značky. Značka tedy byla vytvořena, ovšem její popisek neobsahoval žádný znak. Procházet celou databází z tohoto důvodu by bylo zbytečně náročné. K účelu detekce takové chyby byl použit jednoduchý skript:

Kontrola chyb, kontrola.m

```
1 fid = fopen(['D:\cteny_text.TextGrid']);
2 for i=1:13 % Nevýznamná data
3     fgetl(fid);
4 end
5 velikost=sscanf(fgetl(fid),'          intervals: size = %f');
6 for i=1:(velikost);
7     fgetl(fid);fgetl(fid);fgetl(fid);
8     kontrola=sscanf(fgetl(fid),'          text = "%s"');
9     if kontrola == ('')
10        fprintf('Byla nalezena chyba v itemu číslo %d\n' ,i);
11    return
12    end
13 end
14 fprintf('Zadne chyby');
```

Pokud byla nalezena chyba, tj. prázdná značka, na výstupu bylo vypsáno v jakém itemu. Po opravě chyby byl soubor kontrolován znovu dokud program nevypsál hlášku „Zadne chyby“.

Práce s izolovanými slovy

Protože jsou izolovaná slova uložena v jednom souboru, není možné s nimi pracovat samostatně. Aby to bylo možné, je potřeba slova rozstříhat. Z tohoto důvodu vznikl skript, který dovede dané slovo na základě jeho popisku (a současně obsahu) uložit zvlášť do pracovního adresáře. Práce s takovým skriptem je jednoduchá, nejprve je zapotřebí na druhém až čtvrtém řádku doplnit správné cesty k oběma souborům (ke zvukovému i popisovému). Další položkou, kterou je třeba vyplnit je na 14. řádku hledané slovo (v příkladu je uvedeno slovo „epidiaskop“). Naposledy je zapotřebí v řádku číslo 21 zadat název nově vzniklého souboru. V případě potřeby je možné slovo vystříhnout s určitou rezervu před ním a po něm. To by se provedlo přičtením, resp. odečtením jistého počtu vzorků v řádku číslo 20, viz příklad „Korekce délky vystřiženého slova“.

Rozstřihání izolovaných slov, strih.m

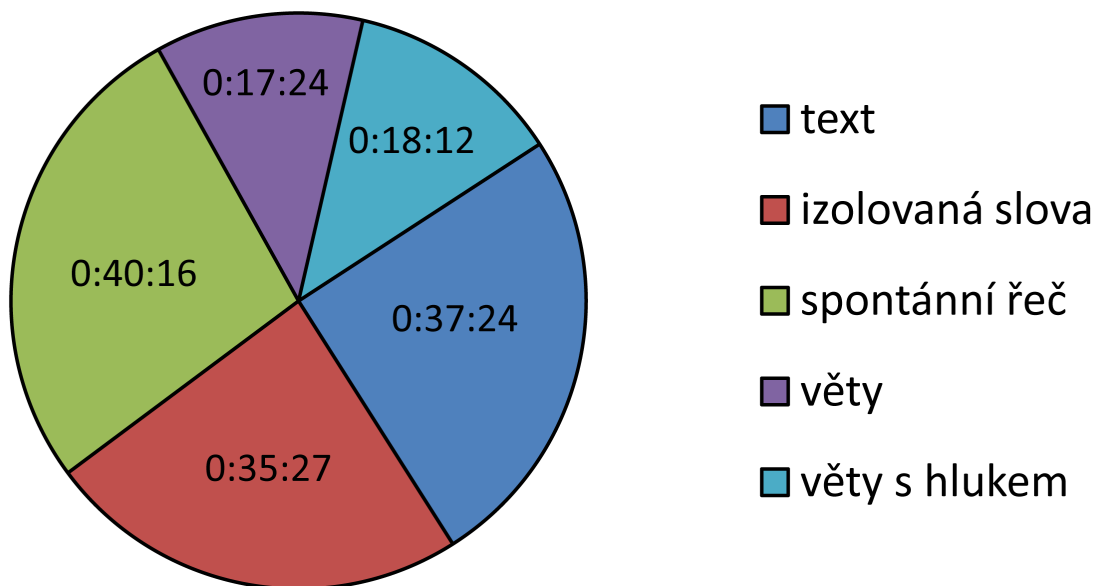
```
1 %Nacteni potrebných dat, doplnit cesty!
2 sw=wavread('D:\izolovana_slova.wav');
3 [sw,Fs,Nbits]=wavread('D:\izolovana_slova.wav');Ts=1/Fs;
4 fid = fopen(['D:\slovabez.TextGrid']);
5 for i=1:13 % Nevýznamná data
6     fgetl(fid);
7 end
8 velikost = sscanf(fgetl(fid),'           intervals: size = %f'
9 );
10 for i=1:velikost;
11     fgetl(fid);
12     xmin=sscanf(fgetl(fid),'           xmin = %f');
13     xmax=sscanf(fgetl(fid),'           xmax = %f');
14     slovo=sscanf(fgetl(fid),'           text = "%s"');
15     a=strcmp (slovo, 'epidiaskop');%Doplnit hledané slovo!
16     if a==1
17         fprintf('Slovo nalezeno v itemu číslo %d\n' ,i);
18         break
19     end
20 end
21 x=sw((xmin/Ts):(xmax/Ts));
22 wavwrite(x,44100,16,'epidiaskop.wav');%Zadat název souboru!
```

Korekce délky vystřiženého slova

```
1 x=sw(( (xmin/Ts)-0.5/Ts) : ((xmax/Ts)+0.5/Ts));%+-0,5 sekundy
```

2.5.4 Velikost databáze

Vzhledem k množství nahrávek a použití formátu PCM, je celkový objem dat poměrně veliký. Celá databáze má velikost 756 MB. Každý z mluvčích namluvil pro potřeby databáze v průměru 9 minut a 18 sekund sestříhaných dat. Objem nesestříhaných nahrávek byl samozřejmě ještě větší. Celkový čas všech nahrávek činí 2 hodiny 28 minut a 43 sekund. Největší objem má spontánní řeč, následovaná čteným textem, viz obrázek 2.8 s grafem podílu jednotlivých položek k celé databázi.



Obr. 2.8: Graf podílu nahrávek dle obsahu

Pro představu velikosti databáze slouží tabulka 2.3. Lze z ní vyčíst jednotlivé časy nahrávek v minutách a sekundách, jejich součty i celková doba trvání všech nahrávek (údaj vpravo dole).

Řečník	text	slova	spontánní řeč	věty	věty s hlukem	Součet
č. 1	02:29	02:01	02:16	01:04	01:08	0:08:58
č. 2	02:31	02:18	03:28	01:06	01:13	0:10:36
č. 3	02:18	02:01	02:11	01:11	01:13	0:08:54
č. 4	02:32	02:02	02:16	01:10	01:12	0:09:12
č. 5	02:14	02:37	02:11	01:01	01:08	0:09:11
č. 6	02:23	02:22	03:33	01:06	01:15	0:10:39
č. 7	02:02	01:47	02:45	01:07	01:07	0:08:48
č. 8	02:08	01:40	02:29	01:05	01:01	0:08:23
č. 9	02:05	02:11	02:55	00:59	01:03	0:09:13
č. 10	02:53	02:15	02:48	01:07	01:12	0:10:15
č. 11	02:25	02:17	02:13	01:05	01:10	0:09:10
č. 12	02:40	03:24	02:45	01:19	01:12	0:11:20
č. 13	02:08	02:14	02:06	01:02	00:55	0:08:25
č. 14	02:03	01:50	02:04	00:58	01:00	0:07:55
č. 15	02:24	02:42	02:13	01:06	01:19	0:09:44
č. 16	02:09	01:46	02:03	00:58	01:04	0:08:00
Součet	0:37:24	0:35:27	0:40:16	0:17:24	0:18:12	2:28:43

Tab. 2.3: Velikost databáze

3 ZÁVĚR

Cílem této bakalářské práce bylo vytvořit databázi nahrávek pro detekci hlasové aktivity. Podílelo se na ní 16 různých osob, řečníků. Jedná se tedy o databázi s velkým počtem řečníků (5-50 řečníků) [1]. Vzniklá databáze obsahuje nahrávky o celkovém objemu téměř dvě a půl hodiny, které se dělí do 5 celků dle obsahu (izolovaná slova, spontánní řeč, věty, věty s hlukem a čtený text). Jedná se o objem dat 756 MB. Databáze vznikala v prostorách bezodrazové komory VUT, což se projevilo dobrou kvalitou nahrávek. V práci jsou označeny nejen řečové sekvence a ticho, ale také neverbální prvky řeči nebo doprovodné nežádoucí zvuky. Pro další práci s databází je možné formát značek převést do výhodnějšího formátu (matice v Matlabu) pomocí přiloženého konvertujícího skriptu. Pro práci s jednotlivými izolovanými slovy je přiložen skript, který dovede daná slova vyhledat a rozstříhat do samostatných souborů.

LITERATURA

- [1] PSUTKA, J. et al. *Mluvíme s počítačem česky* Praha : Academia, 2006. 752 s. ISBN 80-200-1209-1.
- [2] KACUR, J. *The Concept of Task Specific Speech Database for VAD Systems* In 48th International Symposium ELMAR-2006., s. 155-158. ISBN 953-7044-03-3.

SEZNAM PŘÍLOH

A Přepis nahrávek	34
A.1 Izolovaná slova	34
A.2 Izolované věty	35
A.3 Čtený text	36
B Náhled značek	37
B.1 Ukázka značkování spontánní řeči	37
B.2 Ukázka značkování izolovaných slov	38

A PŘEPIS NAHRÁVEK

A.1 Izolovaná slova

Antropolog	Mikrob	Hašíš	Svatebčan
Atom	Nářez	Hloh	Šašek
Bázeň	Nekov	Chatrč	Šašlik
Bontón	Odlet	Chirurg	Šedesát
Cestopis	Onkolog	Chlap	Škrob
Cypřiš	Pařez	Chlup	Švihadlo
Čeněk	Pavilón	Chobot	Šleh
Čenich	Pláň	Chov	Špendlík
Černoch	Pláštík	Chrup	Špunt
Čich	Poslech	Chřestýš	Štěp
Děkan	Postup	Ihned	Teorém
Dohad	Přítrž	Infinitiv	Těsnopis
Epidiaskop	Přízeň	Jasmín	Tis
Extrém	Refýž	Jestřáb	Tíseň
Falzum	Rozpad	Kamarád	Tovaryš
Fantóm	Řev	Kauf	Tragéd
Fermež	Říz	Kebab	Trh
Feudalismus	Sejf	Kloub	Trýzeň
Filozof	Sjezd	Koloběh	Úraz
Fotograf	Sluch	Koncert	Úvoz
Fotosyntéza	Sněm	Kontrašpionáž	Všelicos
Fyziolog	Souš	Kozoroh	Výstřih
Gong	Stávkokaz	Lázeň	Western
Graf	Stop	Lov	Zub
Gynekolog	Stručnost	Masáž	Žlutek

Tab. A.1: Izolovaná slova

A.2 Izolované věty

1. Maminka byla strašně rozčilená, tatínek jako vždy zůstal klidný a rozvážený.
2. Tento fakt se budeš muset znovu naučit, Petře.
3. Stál tam veliký zasklený pavilón.
4. A jak se vůbec dokázali svatebčané tak rychle přemístit do radnice?
5. Platy vedoucích pracovníků se stále více přibližují průměrné mzdě v zemi.
6. Šašlik je jistý druh kebabu, které je oblíbené zejména ve státech bývalého Sovětského svazu, konkrétně v Gruzii, Arménii, Ázerbájdžánu, Uzbekistánu a na Ukrajině.
7. Stůj, nebo strelím !
8. Krásné počasí, které poslední dny panuje v Česku, přineslo zvýšenou viditelnost, tedy meteorologický jev, který je možné si užít jen párkrát za rok.
9. Se správně lepidovou japonskou rýží připravíte opravdu chutné a poctivé japonské sushi i doma.
10. Rokytnice nad Jizerou usiluje o povolení letního provozu lanovky na Lysou horu.
11. V prosinci 2010 se konalo školní kolo matematické olympiády kategorie A.
12. Jedním z cílů projektu je navázat co nejširší partnerství mezi Ústavem matematiky a středními školami ve vybraných krajích ČR.

Tab. A.2: Izolované věty

A.3 Čtený text

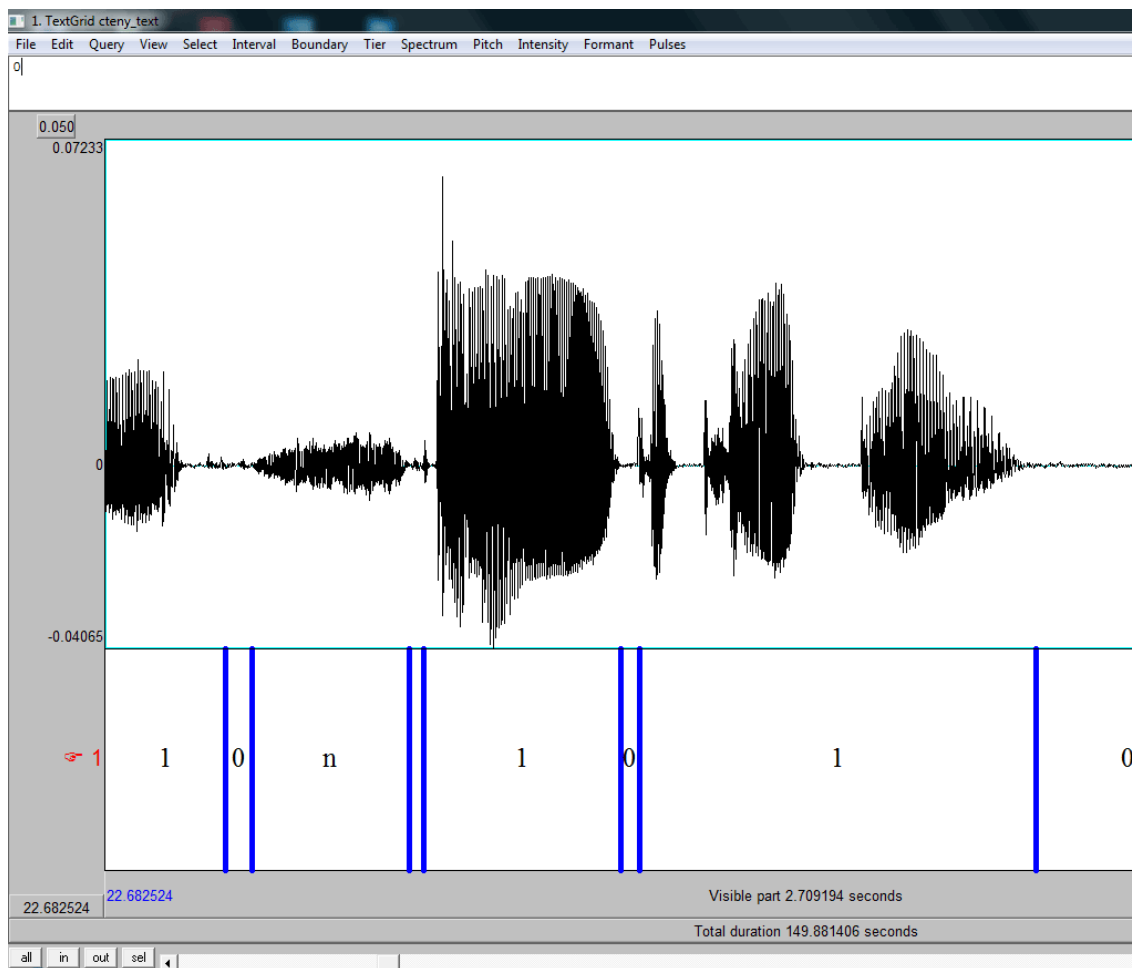
Pokojová květina (Jiří Suchý)

Mám doma malou pokojovou květinu a jsem do ní, jak se říká, blázen. Původně jich bylo víc, ale nemohly beze mne žít, a tak, když jsem se jednou vrátil z dovolené, měly nezdravě žlutou barvu a to byla taky poslední barva, kterou v životě měly. Jenom tuhle jednu květinu se mi podařilo zachránit. A na ní teď lpím. A stává se, že se do mých úvah připele někdy i tato květina - nevím ani, jak se jmenuje, nebyla mi představena. Ale to ještě zvyšuje její tajuplnost a přitažlivost. Znova a znova si uvědomuju, že je živá. A naplňuje mě smíšenými pocity, že tato živá bytost bývá někdy svědkyní i těch nejintimnějších okamžiků, které se koneckonců vyskytují v životě každého člověka. Třeba stříhání nehtů u nohou. Jednou, když jsem byl se svou květinou sám doma, dostal jsem bláznivý nápad: oslovil jsem ji. „Květino,“ pravil jsem, „víš, že tě ošetřuju, zalévám, listy rosím, poskytuju hnojivo OBM a vystavuju tě na slunci. Občas přihnojuji i Hortusem nebo Herbaponem - máš se jako prase v žitě. A mně jde vlastně jen o to, abys lístečkem pohnula, nebo pestíkem zavrtěla, případně bliznou mrskla a tím mi dala najevo, že o mně víš, že stojíš o mou péči, že mě máš ráda. Jde to, nekecej, jsi přece živá!“ Takhle přesně jsem k ní hovořil a pak jsem se velice soustředěně zahleděl na vyjmenované partie. A tu se stala neuvěřitelná věc: květina se ani nepohnula. Zůstala tajemnou. Jsem přesvědčen, že jí to dalo mnoho přemáhání, ale neprojevila se. Protože jenom my, lidé, dovedeme plýtvat sympatiemi a za čas zraňovat tím, že je odpíráme. A tak mi kytky dala lekci, o kterou jsem původně nestál - lekci v ovládnutí citů. A pochopil jsem, že kytky jsou na světě nejen pro ozdobu, ale i pro poučení.

Tab. A.3: Čtený text

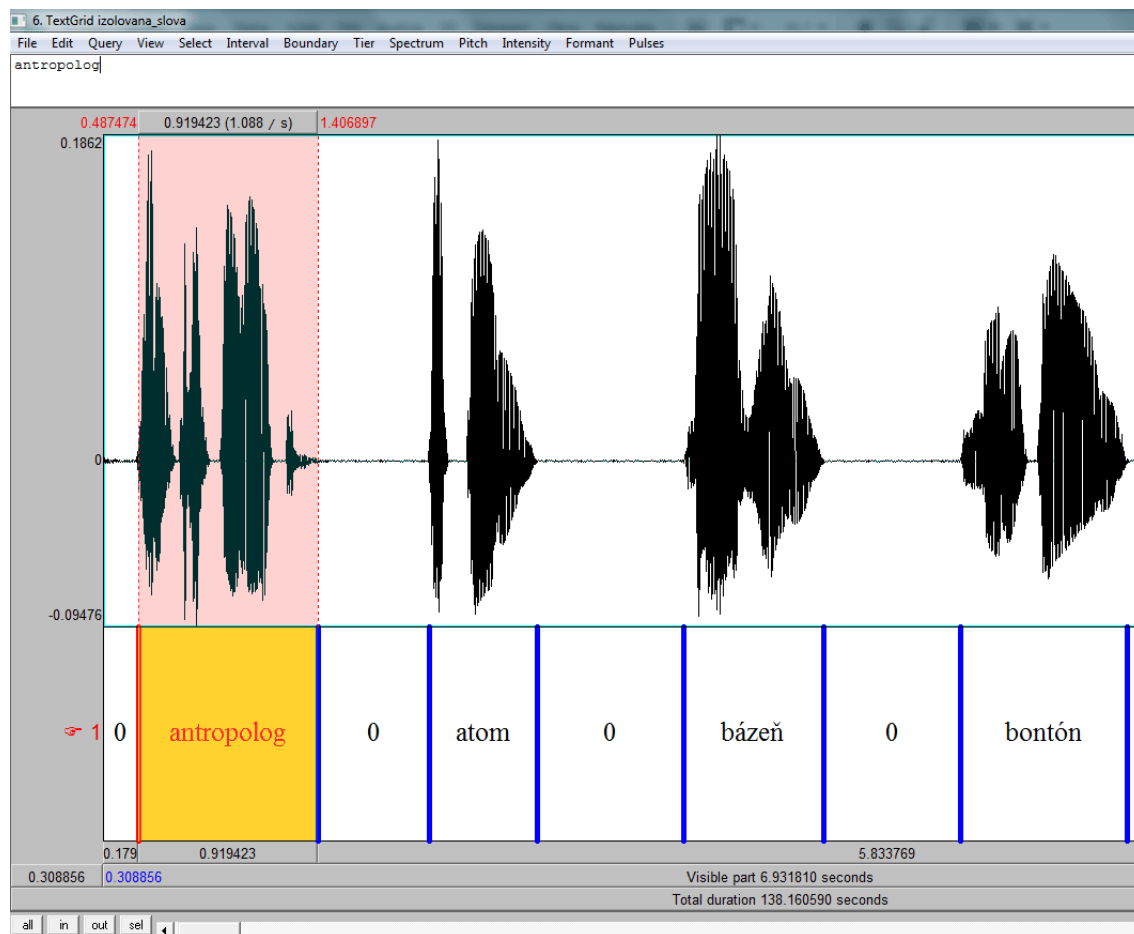
B NÁHLED ZNAČEK

B.1 Ukázka značkování spontánní řeči



Obr. B.1: Náhled značek

B.2 Ukázka značkování izolovaných slov



Obr. B.2: Náhled značek izolovaných slov