VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ
DEPARTMENT OF TELECOMMUNICATIONS

VYTVOŘENÍ DATABÁZE UMĚLE ZAŠUMENÝCH AUDIONAHRÁVEK V AKUSTICKY KONTROLOVANÉ MÍSTNOSTI

CREATING A DATABASE OF AUDIO RECORDINGS WITH ARTIFICIAL NOISE IN AN ANECHOIC CHAMBER

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — **Vojtěch Hájek**
AUTHOR

VEDOUCÍ PRÁCE — **Ing. Pavol Harár**
SUPERVISOR

BRNO 2017

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Bakalářská práce

bakalářský studijní obor **Audio inženýrství**
Ústav telekomunikací

Student: Vojtěch Hájek ***ID:*** 174302
Ročník: 3 ***Akademický rok:*** 2016/17

NÁZEV TÉMATU:

Vytvoření databáze uměle zašumených audionahrávek v akusticky kontrolované místnosti

POKYNY PRO VYPRACOVÁNÍ:

Cílem práce je teoreticky popsat metodologii nahrávání lidského hlasu v bezodrazové komoře a metodologii nahrávání exteriérových šumů. Dále student nahraje alespoň 300 minutových mono audio nahrávek lidského hlasu v bezodrazové komoře a alespoň 4 různé stereo nahrávky exteriérových šumů, z nichž každá bude dlouhá alespoň 10 minut. Zastoupení nahrávek mužského a ženského hlasu musí být vyvážené v poměru 50/50, tzn. že ve výsledku bude nahraných minimálně 150 minut záznamu mužského a 150 minut záznamu ženského hlasu. Jedna osoba může nahrát maximálně 25 nahrávek. Obsah textové předlohy pro nahrávání není důležitý, čtený text však na sebe musí smysluplně navazovat. Může se tedy jednat o úryvek z novinového článku, nebo knihy. Hlavním cílem práce je za použití výše zmíněného materiálu vytvořit databázi nahrávek čistého lidského hlasu a jejich kombinací s externími šumy, který bude moct sloužit jako trénovací dataset pro model hluboké neuronové sítě.

DOPORUČENÁ LITERATURA:

[1] SCHIMMEL, J. Electroacoustics - Laboratroy Practices. Brno, VUT v Brně. 2015. p. 1 - 59.

[2] Rumsey, F. Spatial Audio. Focal Press, 2001. ISBN 0-240-51623-0 (EN)

Termín zadání: 1.2.2017 ***Termín odevzdání:*** 8.6.2017

Vedoucí práce: Ing. Pavol Harár
Konzultant:

doc. Ing. Jiří Mišurec, CSc.
předseda oborové rady

Abstrakt

V této bakalářské práci se zabývám tvorbou databáze zvukových nahrávek a následným vytvoření databáze nahrávek hlasu, které byly nahrány v bezodrazové komoře. Databáze byla vytvořena tak, aby mohla být použita pro učení neuronové sítě s cílem oddělit řeč od hluku pozadí. Z tohoto důvodu jsou součástí databáze i záznamy hluků, které slouží pro umělé zašumění nahrávek hlasu. Dataset zahrnuje nahrávky 18 řečníků ve věku od 16 do 76 let. Polovina řečníků byli muži, polovina ženy. Dataset obsahuje 405 nahrávek hlasu o průměrné délce 46,7 vteřin a celkové délce 315 minut. Kombinací každé nahrávky hlasu s každou nahrávkou šumu ve třech úrovních odstupu signálu od šumu vzniklo 7290 uměle zašumených nahrávek hlasu.

Klíčová slova

Dataset nahrávek hlasu, dataset nahrávek hluku, neuronové sítě, deep learning, bezodrazová komora

Abstract

This bachelor thesis deals with theory of creating the database of sound records and subsequent creating the database of speech records in the anechoic chamber. Database was created as training dataset for learning process of the artificial neural network, which will be able to separate the speech from background noise. Therefore as the part of the database there are also the recordings of various types of noise that will be used as background noise for the voice recordings. The dataset contains records taken from 18 speakers aged from 16 to 76 years. Half of the speakers were men, half women. Database contains 405 records of speach of average length 46,7 secons and total length 315 minutes. By combining each speech record with each noise record at three levels of signal-to-noise ratio was created 7290 mixed records.

Keywords

Database of speech, Database of noise, Neural network, Deep learning, Anechoic chamber

Bibliografická citace

HÁJEK, V. *Vytvoření databáze uměle zašumených audionahrávek v akusticky kontrolované místnosti.* Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2017. 48 s. Vedoucí bakalářské práce Ing. Pavol Harár.

Prohlášení

Prohlašuji, že svou bakalářskou práci na téma „Vytvoření databáze uměle zašumených audionahrávek v akusticky kontrolované místnosti“ 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/nebo majetkových a jsem si plně vědom následku porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonu (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.

V Brně ....

podpis autora

Poděkování

Děkuji vedoucímu bakalářské práce ing. Pavlovi Harárovi za výborné vedení, odbornou pomoc a za vždy rychlou odezvu při vzájemné komunikaci. Dále bych chtěl poděkovat doc. Ing. Jiřímu Schimmelovi, Ph.D. za pomoc při nahrávání v bezodrazové komoře a také všem řečníkům za ochotu při nahrávání.

Faculty of Electrical Engineering
and Communication

Brno University of Technology
Technicka 12, CZ-61600 Brno, Czechia

http://www.six.feec.vutbr.cz

Výzkum popsaný v této diplomové práci byl realizovaný v laboratořích podpořených projektem Centrum senzorických, informačních a komunikačních systémů (SIX); registrační číslo CZ.1.05/2.1.00/03.0072, operačního programu Výzkum a vývoj pro inovace.

EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI

Obsah

1. Úvod

Cílem této bakalářské práce je vytvoření vědeckého datasetu, který bude obsahovat audio nahrávky lidského hlasu. Dataset bude sloužit jako trénovací a testovací množina pro umělé neuronové sítě, které oddělí hluk na pozadí nahrávky od řeči. Součástí datasetu jsou i nahrávky hluku, které budou složit pro umělé zašumení nahrávek řeči.

V teoretické části bakalářské práce se zabývám strukturou a vlastnostmi neuronových sítí a také způsobem učení neuronových sítí. Úvodní část obsahuje také rešerši existujících obdobných datasetů. Poté se v této práci zabývám teorií nahrávání. Poznatky z této části umožnily vytvořit nahrávky v maximální možné kvalitě. Tato kapitola obsahuje i přehled stereofonních mikrofonních technik pro vytvoření nahrávek externích šumů. Na základě teoretické části bakalářské práce byl dataset zašumených nahrávek řeči navržen.

Ve druhé části práce se zabývám tvorbou datasetu, jeho rozsahem, hardwarem použitým při nahrávání, postupem nahrávání, postprodukčními úpravami nahrávek a také strukturou, ve které bude dataset uchován. Součástí práce je také seznam řečníků, kteří se na nahrávání hlasových ukázek podíleli a textové ukázky, které řečníci četli.

Hlavním přínosem mé práce je vytvoření datasetu, který je kompletně v českém jazyce a díky použitému vybavení a pořízení nahrávek v bezodrazové komoře je navíc zaručena jeho vysoká kvalita. Dataset bude použitelný jako trénovací dataset pro neuronové sítě s různým zaměřením.

2. Teoretické základy

2.1. Neuronové sítě

Umělé neuronové sítě využívají strukturu biologických sítí jako svou předlohu. Neuronové sítě jsou základem lidského mozku. Umělé sítě jsou aplikovány při řešení složitých úloh, které lze jen velmi obtížně popsat matematicky. Podobně jako biologické sítě vynikají ty umělé schopností učit se a na základě získaných zkušeností odhadnout řešení problému.

Neuron je základním prvkem neuronových sítí. Jeho funkcí je zpracování, uchování a přenos informací v podobě signálu. Z těla neuronu vystupuje velké množství výběžků. Výstupy neuronu se nazývají axony. Na svém konci se dělí na další výběžky, které se nazývají synapse. Vstupy neuronu jsou dendrity. Z neuronu se přenášená informace šíří axonem přes synapse a dendrity do dalších neuronů, které s informací dále nakládají. Spojením mnoha neuronů pomocí synapsí a dendritů vzniká neuronová síť. Například mozková kůra člověka je tvořena 13 až 15 miliardami neuronů.

Biologické funkce neuronu jsou u umělého neuronu nahrazeny funkcemi matematickými. Umělý neuron má obecně *n vstupů* $x_1,\ldots,x_n$, které modelují dendrity. Dohromady tvoří vstupy neuronu vektor $\boldsymbol{X}=[x_1,\ldots,x_n,]$. Podle umístění neuronu v síti mohou být vstupy buď výstupy jiných neuronů, nebo jsou to vstupy z vnějšího okolí. Vstupy jsou ohodnoceny synaptickými váhami $w_{1i},\ldots w_{ni}$, které určují jejich propustnost, tedy vliv vstupu neuronu na jeho výstup. Nastavováním vah se zabývá proces učení. Vstupy a jejich váhy společně tvoří agregační funkci y_a, kterou lze vyjádřit matematicky následovně:

$$y_a = \sum_{i=1}^{\infty} x_i \cdot w_i \qquad (2.1)$$

Pokud hodnota agregační funkce překročí prahovou hodnotu neuronu, převede aktivační funkce $f(y_a)$ vstupní potenciál y_a na výstupní hodnotu neuronu.

Způsob, jakým jsou neurony v síti propojeny, se nazývá typologie sítě. Nejčastěji využívanou typologií je tzv. vrstvená architektura. Neurony jsou v ní uspořádány do vrstev. Podle pozice jsou vrstvy označeny jako vstupní vrstva, výstupní vrstva a mezi těmito dvěma se nachází minimálně jedna skrytá vrstva. Počet vrstev společně s počtem neuronů ve vrstvách se označuje jako konfigurace sítě. [1]

2.2. Deep learning

Výstup neuronu je přiveden na vstup neuronů sousední vrstvy. Proces nastavování vah těchto spojení se nazývá proces učení. V roce 1949 se Donald Hebb ve své knize „The Organization of Behaviour“ zabývá návrhem učícího pravidla pro mezineuronové rozhraní (synapse). Toto pravidlo mění během učení váhu mezi dvěma neurony na základě jejich aktivity. Pokud jsou oba neurony aktivní, váha je zvětšena, pokud neaktivní, je zmenšena. [1]

Pokud existuje trénovací množina vstupních a výstupních dat, provádí se nastavování vah neuronové sítě jako tzv. učení s učitelem. Vstupní vektor je zpracován sítí a její výstup je porovnán s očekávaným výstupem. Na základě chyby jsou následně váhy modifikovány. Proces se opakuje tak

dlouho, dokud se výsledky sítě stále zlepšují, nebo je chyba mezi výstupem sítě a požadovaným výstupem sítě dostatečně malá.

Aby bylo dosaženo potřebného chování sítě, musí být učící množina dostatečně reprezentativní. Testovací množina je část datasetu, na které se učení neprovádí, ale slouží pouze k ověření správné funkčnosti neuronové sítě. [2]

Neuronové sítě mají schopnost sledovat trendy a pravidla i z komplikovaných průběhů dat. Při správné aplikaci dokážou díky zobecňování poměrně přesně předpovědět údaje, které nebyly součástí trénovacích dat. Díky tomu jsou využitelné v mnoha oblastech. Neuronové sítě například dokážou předvídat vývoj měnových kurzů, nebo rozpoznávat texty. Jejich dalším využitím je také řízení samoobslužných systému, jako například řízení dopravní signalizace, nebo autopilot, který je právě vyvíjen pro automobily.

V oblasti zpracování zvuku slouží neuronové sítě například k diagnostikování Parkinsonovi choroby. Tato diagnóza probíhá z nahrávek řeči pacienta, tedy neinvazivní metodou, na rozdíl od běžného invazivního vyšetření. Dalším využitím je například rozpoznání jednotlivých řečených slov a jejich převod na text. [3]

2.3. Rešerše existujících datasetů s nahrávkami řeči

V této kapitole se zabývám již existujícími datasety řeči. Nejsou zde samozřejmě zahrnuty všechny, ale pouze ty významnější nebo něčím zajímavé. Rešerše slouží k vytvoření představy o rozsahu a obsahu jednotlivých databází.

2.3.1. TIMIT

Dataset obsahuje nahrávky 630 řečníků, kteří hovoří 8 hlavními dialekty americké angličtiny. Skupinu řečníků tvoří ze 70% muži, ze 30% ženy. Je nahráno 10 vět od každého řečníka. Věty obsahují foneticky bohatý text. Dataset vznikl jako trénovací a testovací množina pro vývoj systému pro automatické rozpoznání řeči. Dataset není volně dostupný, pouze pro členy LDC (Linguistic Data Consortium), nebo za peněžní poplatek.[4]

2.3.2. SPINE2

SPINE je zkratkou pro Speech in noisy environments. Tento dataset slouží k rozpoznávání řeči, která je maskována různými hluky z vojenského prostředí. Dataset tvoří rozhovory několika párů mluvčích. V první části vedou dva páry mluvčích dohromady 64 rozhovorů o délce od 1 do 4 minut. Celková délka nahrávek první části je asi 207 minut (1,6 Gb). Ve druhé části je zaznamenáno celkem 32 rozhovorů dvou párů mluvčích o celkové délce 207 minut (811 Mb). Poslední část tvoří nahrávky celkem 16 párů. Každý pár vede 4 rozhovory (64 rozhovorů celkem) o délce 423 minut (1,6 Gb). Stejně jako TIMIT je tento dataset dostupný pouze pro členy LDC a nebo za poplatek. [5]

2.3.3. TED

Dataset obsahuje záznamy prezentací v Angličtině od rodilých mluvčích a cizinců z konference EUROSPEECH 1993 v Berlíně. Dataset tvoří nahrávky 224 řečníků o průměrné délce prezentace

15 minut + 5 minut diskuze. Celková délka nahrávek činí něco kolem 75 hodin. Také tato databáze je dostupná za poplatek nebo členům LDC. [6]

2.3.4. Santa Barbara Corpus of Spoken American English

Zkráceně také SBCSAE. Tento dataset tvoří velké množství nahrávek lidí z různých regionů USA, různého věku, etnika, vzdělání a sociální příslušnosti. Součástí jsou záznamy pořízené při různých situací, jako například: konverzace, pomluvy, pracovní konverzace, politické projevy, pohádky, školní výuka a další. Dataset je rozdělen na čtyři části, každá asi o 6 hodinách nahrávek. I tento dataset je zdarma dostupný pro členy LDC, pro ostatní pouze za poplatek. [7]

2.3.5. TIDIGITS

Dataset se skládá ze sekvencí čísel, čtených od 326 mluvčích, z toho 111 mužů, 114 žen, 50 chlapců a 51 dívek. 77 číselných sekvencí od každého řečníka je rozděleno do tréninkové a testovací množiny. Sekvence obsahují různý počet číslic. Dataset je také součástí archivu LDC a je zdarma pouze pro členy. [8]

2.3.6. Aurora

Databáze je navržena jako tréninková množina pro algoritmus, který rozpoznává řeč v hlučném prostředí. Základ tohoto datasetu je dataset TIDIGITS. K číselným sekvencím datasetu TIDIGITS bylo přidáno 8 nahraných šumů a hluků. Nahrávky obsahují pouze frekvenční pásmo od 0 do 4 kHz. Pro získání tohoto datasetu je nutné právo užívat dataset TIDIGITS. [9]

2.3.7. Berlin Database of Emotional Speech

Tento dataset obsahuje nahrávky emoční řeči. 10 herců (5 žen a 5 mužů) simuluje 7 různých emocí na 10 větách v Německém jazyce. Nahrávky byly vytvořeny v bezodrazové komoře. Databáze celkem obsahuje téměř 800 vět. Projevy emocí byly posuzovány z hlediska kvality a z více než 60% byly označeny za přirozené. Databáze je volně dostupná. [10]

2.3.8. Parkinson Speech Dataset

Dataset se skládá z nahrávek hlasu 20 zdravých lidí (10 mužů a 10 žen) a 20 lidí s Parkinsonovou chorobou (6, žen a 14 mužů). Od každého mluvčího je nahráno 26 ukázek, které zahrnují hlásky „a“ a „o“, čísla, slova nebo krátké věty. Tento dataset je volně dostupný. [11]

2.3.9. ISOLET

Dataset obsahuje nahrávky 150 řečníků hláskujících dvakrát každé písmeno abecedy, čímž vzniklo 52 vzorků na jednoho řečníka. Řečnící jsou rozděleni po 30 do pěti skupin. Každou skupinu tvoří nahrávky 15 mužů a 15 žen. Skupiny isolet1 až isolet4 slouží jako tréninkové množiny, skupina isolet5 je testovací množinou. Celková velikost datasetu je 150Mb. Databáze je dostupná za malý poplatek a může být zdarma kopírována. [12]

Tab. 1: Přehled dostupných datasetů

Název	Obsah	Počet řečníků	Použití	Dostupnost
TIMIT	6300 nahrávek	630	rozpoznání řeči	Placený, pro LDC zdarma
SPINE2	14 hodin	40	oddělení řeči od šumu	Placený, pro LDC zdarma
TED	75 hodin	224	rezpoznání řeči	Placený, pro LDC zdarma
SBCSAE	24 hodin	216	rozpoznání řeči	Placený, pro LDC zdarma
TIDIGITS	25102 nahrávek	326	rozpoznání číslic	Placený, pro LDC zdarma
Aurora	28028 nahrávek	104	oddělení řeči od šumu	Nutné právo pro užívání TIDIGITS
Berlin Database of Emotional Speech	800 nahrávek	10	rozpoznání emocí	Zdarma
Parkinson Speech Dataset	1040 nahrávek	40	odhalení Parkinsonovy choroby	Zdarma
Isolet	7797 nahrávek	150	rozpoznání jednotlivých písmen	Placený, poplatek zahrnuje pouze náklady na vytvoření kopie

3. Teorie nahrávání

Snahou každého autora jakýchkoliv zvukových záznamů by mělo být pořízení nahrávek v nejvyšší kvalitě, jaká je za daných okolností možná. Pořízený materiál může být využitelný i pro jiné účely, než pro jaké byl původně plánován a případná nízká kvalita záznamu nemusí být dostačující. Při vytváření záznamů zvuku je často volen kompromis mezi kvalitou zvukového záznamu a objemem dat. Při pořizování nahrávek řeči je pro správnou volbu parametrů nahrávacího řetězce dobré znát základní vlastnosti lidského hlasu, jako je například jeho frekvenční a dynamický rozsah. [13]

3.1. Vlastnosti lidského hlasu

Lidské hlasové ústrojí se jako každý akustický systém skládá z excitátoru (budící element - proud vzduchu), oscilátoru (kmitající element - hlasivky) a z rezonátoru (zesilující a vyzařující element - soustava rezonančních dutin vokálního traktu). Vlivem pohybu dýchacích svalů dochází ke změně objemu plic a k vydechování vzduchu. Ten vzduch proudí do hrtanu, kde jsou uloženy hlasivky. Ty jsou při tvorbě hlásek (vokálů) přitisknuty těsně k sobě. Proudící vzduch je periodicky rozkmitává a kvaziperiodické změny tlaku vytváří primární akustický signál. Ten postupuje přes hltan do ústní a nosní dutiny, odkud je vyzářen do okolního prostoru. Hlasivkový tón nemá během řeči stálou frekvenci. Ta kolísá v závislosti na emocionálním stavu řečníka. Tato frekvence se může pohybovat mezi 70 Hz a 680 Hz.

Všechny hlásky obsahují své charakteristické formanty (lokální maxima ve spektrálním složení tónu). Ty jsou dány konkrétními rozměry vokálního traktu. Formantové oblasti určují kvalitativní a kvantitativní vlastnosti vokálů. První tři formanty jsou rozhodující pro rozlišení jednotlivých vokálů. Čtvrtý až sedmý formant jsou pro všechny hlásky stejné a mají vliv pouze na barvu hlasu řečníka.

Kromě kvaziperiodického signálu hlásek vzniká v hlasovém ústrojí i stochastický signál souhlásek (konsonantů). Takový šumový signál je vytvářen vlivem turbulentního proudění vzduchu, ke kterému dochází díky překážkám v hlasovém ústrojí. Analýza formantových oblasti a jejich přesné určení je u souhlásek mnohem obtížnější než u samohlásek. Je to dáno především jejich spektrálním složením.

Řečový signál je v podstatě řada vokálů, která je vhodně narušována konsonanty (souhláskami). Díky časové koordinaci procesů vytvářejících vokály a konsonanty vzniká souvislá řeč. Při řeči se každá slabika pohybuje v jiné tónové výšce, což zapříčiňuje melodický spád řeči, který je typickým znakem některých jazyků a dialektů.

Výškový rozsah lidského hlasu se s věkem mění. U kojence se pohybuje kolem 440 Hz. S přibývajícím věkem se nejprve rozšiřuje směrem dolů, později nahoru. V pubertě klesá hlas chlapců až o oktávu. Průměrná výška mužského hlasu kolísá při běžném hovoru mezi 100 Hz a 200 Hz. U ženského hlasu o oktávu výše.

Dynamický rozsah hovorové řeči se pohybuje od 20 dB při šepotu až k 80 dB při křiku. Hladina akustického tlaku při běžném hovoru je asi 65 dB ve vzdálenosti jednoho metru od řečníka. Průběh dynamiky má u řečového signálu impulzní charakter, průměrná hladina se pohybuje asi 12 dB pod maximální hodnotou.

Frekvenční rozsah řeči pokrývá celé pásmo slyšitelnosti. K dobré srozumitelnosti je však třeba pásmo mnohem užší. Toto pásmo musí ale obsahovat všechny formanty. Například u telefonů a některých intercomů je přenášen signál s frekvencí pouze do 4 kHz. Řeč je srozumitelná, pouze v ní nejsou zachyceny sykavky (/s/,/z/). [14]

3.2. Vzorkovací frekvence

Lidské ucho je schopné vnímat zvuky v kmitočtovém pásmu od 20Hz do 20kHz. Pro volbu vzorkovací frekvence je rozhodující horní hranice tohoto spektra. Pravidlo pro určení vzorkovacího kmitočtu se nazývá Shannon-Nyquistův teorém:

„Přesná rekonstrukce spojitého, frekvenčně omezeného signálu z jeho vzorků je možná tehdy, pokud byla vzorkovací frekvence vyšší než dvojnásobek nejvyšší harmonické složky vzorkovaného signálu.“ [15]

Minimální vzorkovací frekvence je tedy 40kHz. Nesplnění poučky má za následek překrývání sousedních spekter vzorkovaného signálu. Tento jev se nazývá aliasing. Aliasingu je možné zabránit pomocí antialiasingového filtru, což je kmitočtový filtr typu dolní propust. Pokud chceme signál zaznamenat s nižší vzorkovací frekvencí, dojde k ořezání vyšších harmonických složek. Nejčastěji používané vzorkovací frekvence v oblasti zpracování audio signálu jsou 44,1 kHz, 48 kHz a jejich celočíselné násobky.

3.3. Bitová hloubka

Zatímco vzorkovací frekvence ovlivňuje spektrum výsledného signálu, bitová hloubka ovlivňuje jeho amplitudu. Kvantování je v podstatě zaokrouhlování aktuální výchylky na předem dané kvantizační úrovně. Kvantizační zkreslení je rozdíl mezi skutečnou hodnotou signálu a hodnotou přiřazenou během kvantování. Bitovou hloubku je třeba volit tak, aby toto zkreslení bylo dostatečně malé. Čím je však kvantizační krok menší, tím více místa zabírá výsledný soubor v úložišti. Počet bitů udává počet těchto úrovní. Při 16 bitech může být hodnotě napětí vstupního signálu teoreticky přiřazeno až 65 536 úrovní (2^{16}). To pokrývá dynamický rozsah 94dB. Při bitové hloubce 24 bitů pokrývá dynamický rozsah celou oblast slyšitelnosti (od prahu slyšitelnosti po práh bolesti). Dynamický rozsah řeči nabývá hodnot od 20 dB při šepotu po 80 dB při křiku. Z tohoto důvodu je tedy bitová hloubka 16 bitů pro záznam řeči dostačující. Rozdíl mezi 16 a 24 bitovým záznamem není lidské ucho schopno postřehnout.

3.4. Vstupní hlasitost

Správné nastavení vstupní hlasitosti je při nahrávání audio signálů naprosto zásadní. Většina zařízení většinou neumožňuje nastavit příliš nízké hodnoty vzorkovací frekvence nebo bitové hloubky, takže jejich znalost pro laika není tak důležitá. Při špatném nastavení vstupní hlasitosti může vzniknout nahrávka buď „přebuzená“ nebo téměř neslyšitelná. Pokud má použité nahrávací zařízení vysoký odstup signálu od šumu (signal-to-noise ratio, SNR), dá se špatně slyšitelná nahrávka zesílit na požadovanou úroveň, ale SNR se tím výrazně zhorší a nahrávka nemusí dosahovat potřebných kvalit. Pokud je vstupní signál hlasitější než maximální úroveň, kterou je přístroj schopen zachytit, upozorní na tento jev dioda s označením „peak“ (případně overload, clip).

V tomto případě dojde k nekompletnímu zaznamenání zvukové vlny. Okamžitá výchylka vlny je mimo rozsah přístroje, proto dojde k ořezání vlny a vzniku harmonického zkreslení.

Vstupní hlasitost je vhodné nastavovat o 6 dB až 12 dB nižší, než je maximální úroveň, kterou je přistroj schopen přijmout. Tímto krokem zajistíme dostatečný odstup signálu od šumu a zároveň je ponechána značná rezerva, kdyby došlo k neočekávanému nárůstu hlasitosti zdroje zvuku.

3.5. Formát

Při nahrávání zvuku je nutné pro uchovávání nahrávek používat formáty bez komprese nebo formáty s bezztrátovou kompresí. V současnosti nejpoužívanějším takovým formátem je Waveform Audio File Fromat (WAV). Tento formát podporuje různé nastavení bitové hloubky, vzorkovací frekvenci i různý počet přenášených kanálů.

3.6. Mikrofony

Mikrofony dělíme především podle principu převodu zvuku na elektrický signál na kondenzátorové a dynamické. Oba tyto systémy mají řadu modifikací. Mikrofony dále dělíme podle směrové charakteristiky, což je závislost citlivosti mikrofonu na úhlu mezi zdrojem zvuku a akustickou osou mikrofonu v konstantní vzdálenosti mikrofonu od zdroje zvuku. Znázorňuje se zpravidla grafy pro různé frekvence v polárních souřadnicích. Za základní směrové charakteristiky mikrofonů považujeme kulovou (všesměrovou), osmičkovou a kardioidní (ledvinovou) charakteristiku. [16]

3.6.1. Dynamický mikrofon

Principem konstrukce se dynamické mikrofony podobají reproduktorům. Jejich základ tvoří membrána, která je mechanicky spojena s cívkou. Cívka se pohybuje v magnetickém poli permanentního magnetu. Membrána kmitá podle změn akustického tlaku a rozkmitává tím cívku. Pohyb vodiče v magnetickém poli vyvolává v závitech cívky slabý elektrický proud, jež je nutné zesílit mikrofonním předzesilovačem.

Parametry dynamického mikrofonu ovlivňuje především rozměr a hmotnost membrány a hmotnost cívky. Větší membrána zaručuje lepší odstup od šumu, ale větší hmotnost se projevuje větší setrvačností. To má dopad v podobě poklesu přenosu na vyšších kmitočtech. Navíc zvuky, které přicházejí ze směru mimo osu, dopadají na jednu stranu membrány dříve než na druhou. Některé vyšší frekvence mohou dopadnout na membránu s jinou fází, čímž se dále zhoršují vlastnosti mikrofonu na vyšších kmitočtech.

Počet závitů a typ vodiče ovlivňují hmotnost cívky. Vyšší hmotnost cívky ovlivňuje vyšší kmitočty stejně jako těžší membrána. Z tohoto důvodu mají dynamické mikrofony nižší frekvenční rozsah než například kondenzátorové mikrofony. Počet závitů můžeme snížit použitím silnějšího magnetického pole. Proto používají moderní dynamické mikrofony magnet z neodymia, což umožňuje zkonstruovat dynamický mikrofon s vyrovnaným přenosem až do 20kHz.

Dynamické mikrofony produkují poměrně slabý výstupní signál. Z tohoto důvodu se hodí spíše pro snímání hlasitějších zdrojů zvuku, nebo musí být umístěny v těsné blízkosti zdroje zvuku.

3.6.2. Kondenzátorový mikrofon

Membrána kondenzátorového mikrofonu tvoří jednu elektrodu kondenzátoru. Kmitáním membrány se mění vzdálenost mezi elektrodami a tím i kapacita kondenzátoru. Na odporu mezi napájecím napětím a kondenzátorem můžeme pozorovat změny napětí, které odpovídají změnám kapacity. Celý systém funguje pouze za přítomnosti elektrického náboje na deskách kondenzátoru. Proto musí být systém napájen tzv. Phantomovým napájením (48V).

Ve srovnání s dynamickými mikrofony jsou kondenzátorové konstrukčně složitější. Díky současným technologiím je možné vyrobit velmi tenkou pokovenou membránu, díky čemuž má minimální setrvačnost a dokáže lépe reagovat na vyšší frekvence než membrána s cívkou dynamického mikrofonu. Kapacitní mikrofony jsou velmi citlivé a mají nízký šum. Další výhodou je vyrovnaný přenos v celém pásmu slyšitelných kmitočtů. Kondenzátorové mikrofony mohou trpět ve vlhkém prostředí snižováním citlivosti. Příčinou poklesu citlivosti je vodivé spojení desek kondenzátoru vlivem vlhkosti vzduchu.

3.6.3. Kulová (všesměrová) charakteristika

Mikrofon s kulovou charakteristikou se také nazývá „tlakový“ mikrofon. Reaguje pouze na změny tlaku vzduchu. Membrána je upevněna na okraji vzduchotěsné dutiny. Tlak na zadní část membrány je konstantní, na přední straně se závisí na zvukových vlnách. Nezáleží na směru zvuku dopadajícího na membránu, pouze na jeho amplitudě.

Na vyšších kmitočtech je nemožné dosáhnout všesměrové charakteristiky, protože samotný mikrofon vytváří akustický stín. Mikrofon je ze zadní strany a z boku méně citlivý na vysoké kmitočty než zepředu. Z tohoto důvodu je vhodné tento mikrofon zkonstruovat co nejmenší, aby směrová charakteristika byla kulová pro co nejvyšší frekvence.

Výhodou mikrofonů s kulovou charakteristikou je především absence tzv. proximity efektu, což je narůstání basů se snižující se vzdáleností. U kardioidních i osmičkových systémů se s tímto efektem setkáme. Všesměrové mikrofony jsou schopné zachytit poměrně věrně i zvuky přicházející ze směru mimo osu. Proto se hodí například na nahrávání orchestru, pokud chceme zachytit i vliv koncertního sálu. Naopak je tento mikrofon velmi nevhodný pro nahrávání jednoho zdroje zvuku v hlučném prostředí.

3.6.4. Osmičková charakteristika (Figure eight, Bi-directional)

Tento systém umožňuje zvukovým vlnám přístup na membránu ze dvou stran. Zvuk je rovnoměrně snímán zepředu i zezadu, ale z boků je necitlivý, protože zvuková vlna přichází na přední i zadní stranu membrány současně, takže nenastává žádný rozdíl tlaku vzduchu.

Tento systém je ovlivněn proximity efektem. Tento jev nemusí být nutně nevýhodou. Lze jej využít jako umělecký záměr. Například snížení vzdálenosti mezi ústy hlasatele a mikrofonem přidá jeho hlasu zesílení na spodních frekvencích. Tím hlas získá sytost a příjemné zabarvení. Někteří výrobci udávají vliv proximity efektu v kmitočtové charakteristice.

Mikrofony s bi-directional charakteristikou se dnes používají spíše výjimečně, a to ve speciálních aplikacích, například při snímání systémem MS (stereofonní technika, 2 mikrofony: jeden s kardioidní charakteristikou pokrývá zdroje signálu ve středu stereofonního obrazu, mikrofon s osmičkovou charakteristikou složí ke snímání zvuku přicházejícího ze stran)

3.6.5. Kardioidní (ledvinová) charakteristika

Zvuk, který přichází zepředu na membránu, způsobuje rozdíl tlaku před a za membránou. Zvuky přicházející zezadu a ze stran rozdíl tlaku, alespoň teoreticky, nezpůsobují, díky speciálně zkonstruovanému zvukovodu. Ve skutečnosti jsou snímány i zvuky ze stran, ale ne tak účinně jako zepředu.

Od kardioidních systémů jsou odvozeny mikrofony superkardioidní nebo hyperkardioidní. Ty potlačují více signál ze stran než kardioida, ale na rozdíl od mikrofonu s ledvinovou charakteristikou jsou více citlivé ke zvukům, které přicházejí zezadu.

Směrové mikrofony jsou vhodné především při aplikacích, kdy chceme potlačit zvuky přicházející ze směru mimo osu mikrofonu (hlučné prostředí nebo potlačení odrazů prostoru). Kardioidní systémy jsou jednoznačně nejpoužívanějším typem mikrofonu při živých hudebních vystoupeních, kdy je potřeba minimalizovat tzv. přeslechy. Přeslechem rozumíme zachycení zvuku z jiného zdroje, než který mikrofon snímá. Přeslech navíc bývá díky proximity efektu často nepříjemně frekvenčně zabarven.

3.7. Stereofonní nahrávání a mikrofonní techniky

Stereofonní techniky nahrávání do své práce zahrnuji kvůli vytváření nahrávek šumů, které budou dostupné v monofonní i stereofonní verzi. Stereofonní nahrávání je snaha o zachycení zvuku v daném prostoru co nejpřesvědčivěji. Dosud neexistuje žádný systém, který by dokázal věrně sejmout a reprodukovat akustické pole. Existuje však několik technik stereofonního snímání, které se ideálu přibližují a vytváří celkem přesvědčivý prostorový vjem.

Při snímání zvuku dvěma a více mikrofony dochází ke kombinaci přímého zvuku se zvukem zpožděným a ke vzniku efektu hřebenového filtru. Vlivem různých fází obou signálů dochází ke zvýraznění některých frekvencí a naopak k zeslabení nebo dokonce k úplnému potlačení jiných. [16][17]

3.7.1. Systém XY

Jedním z nejpoužívanějších způsobů stereofonního snímání zvuku je použití koincidenčního páru, označovaného XY. Tento systém se skládá ze dvou kordioidních mikrofonů se shodnou charakteristikou a mezi sebou svírají úhel od 60° do 120°. Na zvoleném úhlu závisí šířka stereofonního obrazu. Typicky se volí úhel 90°. Obě mikrofonní kapsle jsou umístěny co nejblíže k sobě. Tím se zamezí vzniku hřebenového filtru při mono poslechu nahrávky. Oba mikrofony jsou směrové, takže jeden snímá více levou a druhý pravou část prostoru. Tento systém zaznamenává pouze změny intenzity v prostoru, ale nezachycuje žádné změny fáze nebo frekvenční změny, které vznikají při binaurálním slyšení.

Nespornou výhodou tohoto systému je především mono kompatibilita, která je zajištěna zanedbatelnou vzdáleností obou mikrofonů. Tato skutečnost je však zároveň i nevýhodou, protože absence fázových rozdílů oslabuje výsledný stereofonní efekt. Díky nedokonalostem ve směrové charakteristice kardioidních mikrofonů mimo jejich osu může dojít k ochuzení středu stereofonního obrazu o vyšší frekvence. Systém XY i přes tyto nevýhody dosahuje slušných výsledků, proto je hojně používán.

3.7.2. Systém MS

U tohoto systému lze vytvořit kvalitnější střed stereofonního obrazu než u systému XY. Navíc je monaurálně kompatibilní. Systém MS tvoří jeden kardioidní (nebo všesměrový) mikrofon (M), který pokrývá především střed stereofonního obrazu. Druhý mikrofon s osmičkovou charakteristikou (S) pokrývá levou a pravou stranu. Signál ze stranového mikrofonu je pro jeden kanál k signálu středového mikrofonu přičten (M+S) a pro druhý je od něj odečten (M-S). Při mono poslechu získáme pouze kanál středového mikrofonu.

Výhodou systému je možnost změnit šířku stereofonního obrazu. Lze toho docílit změnou poměru úrovně středového mikrofonu ke stranovému. Pokud zcela potlačíme signál ze stranového, zůstane pouze monofonní signál středového mikrofonu. Zvýšením úrovně stranového mikrofonu vznikne rozšířený stereo obraz.

Výhodou této techniky snímání je věrné zachycení středu zásluhou středového mikrofonu. Stranové signály se díky opačné polaritě při sloučení do jednoho kanálu zcela odečtou. Na druhou stranu osmičková charakteristika stranového mikrofonu může vést k ne zcela přesvědčivě sejmutému signálu přicházejícímu z prostoru mimo jeho osu. Protože je vzdálenost mezi mikrofony, stejně jako u techniky XY, minimální, nepřináší tato metoda snímání žádné fázové informace.

3.7.3. Systém AB

Dva mikrofony s kulovou charakteristikou umístěné ve větší vzdálenosti od sebe zachycují kromě rozdílů intenzit i časové a fázové rozdíly v prostoru. Vzdálenost mezi mikrofony je vhodné volit tak, aby mezi nimi nevznikala hluchá místa. Při dostatečné vzdálenosti mezi mikrofony je hřebenový filtr méně znatelný. I přesto však není tento systém zcela mono kompatibilní. Tento systém poskytuje dobrou lokalizaci zdroje zvuku a dobrý vjem hloubky prostoru. Nedoporučuje se tento systém umisťovat do blízkosti zdroje zvuku, protože jeho případný drobný pohyb může ve stereofonním obrazu vyvolat dojem velkého posunu. Správné rozmístěná je u tohoto systému zásadní, proto je vhodné pořídit několik testovacích nahrávek pro různá umístění mikrofonů a následně je poslechem sluchátky porovnat. Nahrávky pořízené systémem AB jsou vhodné spíše pro poslech na reproduktorových soustavách.

4. Nahrávky hlasu

4.1. Využití nahrávek

Tyto nahrávky byly vytvořeny jako součást trénovací množiny pro model hluboké neuronové sítě, která oddělí řeč od šumové složky. Nahrávky hlasu vznikly v bezodrazové komoře Ústavu telekomunikací na VUT v Brně a poté do nich byl uměle přidán šum. Nahrávky lidského hlasu jsou pořízeny od 18 řečníků, 9 mužů a 9 žen, v co nejširším věkovém rozpětí. Díky tomu by nahrávky mohly sloužit i pro neuronové sítě, které odhadnou věk nebo rozpoznají pohlaví řečníka. Pro tento účel však databáze obsahuje vzorky příliš malého počtu řečníků a pro využití v této oblasti by bylo vhodné tuto databázi rozšířit o nahrávky získané od většího počtu mluvčích. Databáze je unikátním především tím, že všechny nahrávky jsou v českém jazyce. Při vytváření datasetu byl kladen důraz na vysokou kvalitu nahrávek, čehož je dosaženo použitím kvalitního nahrávacího řetězce a především pořízením nahrávek v bezodrazové komoře.

4.2. Rozsah nahrávek

Požadovaný výstup mé bakalářské práce je 300 minutových mono audio nahrávek hlasu (150 mužského hlasu a 150 ženského hlasu), přičemž jeden řečník smí nahrát maximálně 25 nahrávek. Po dohodě s vedoucím práce bylo rozhodnuto, že vytvořené nahrávky nebudou minutové, ale kratší, avšak celková délka nahrávek datasetu (300 minut) bude přinejmenším dodržena. Pokud to bude možné, bude však překročena. Poměr mužů a žen zůstane zachován 1:1 a maximální počet nahrávek na jednu osobu také.

Původně bylo naplánováno, že dataset budou tvořit nahrávky od 24 řečníků- 12 mužů a 12 žen. Po komplikovaném organizování nahrávacích frekvencí a obtížnému hledání termínu, který by vyhovoval řečníkům a kdy bezodrazová komora byla k dispozici, byl počet řečníků snížen na 18 (9 mužů a 9 žen). Nejmladšímu řečníkovi bylo v době nahrávání 16 let a 10 měsíců, nejstaršímu 76 let a 7 měsíců. Dataset tedy zahrnuje nahrávky řečníků různých věkových skupin a v tomto věkovém rozmezí jsou řečníci rozprostřeni v rámci možností rovnoměrně.

Od všech řečníků bylo celkem pořízeno 450 nahrávek, z nichž bylo 45 nahrávek vyřazeno. Soubor nahrávek řeči tedy tvoří 405 nahrávek lidského hlasu o celkové délce 315 minut, které byly sestříhány přibližně z 9 hodin záznamu.

4.3. Specifikace použitých mikrofonů a ostatního hardwaru

K nahrávání byl použit free-field kondenzátorový mikrofon Brüel&Kjær 4189, který je určen pro použití v bezodrazové komoře nebo ve volném poli. Dále byl použit mikrofonní předzesilovač Brüel&Kjær Type 2669, A/D-D/A převodník RME ADI-2, který byl pomocí rozhraní ADAT propojen s PCI zvukovým rozhraním počítače RME Hammerfall. Pro nahrávání a editaci byl použit DAW software Cubase 7, pro postprodukční úpravy taktéž Cubase a Wavelab 7, určený pro mastering. Vzorkovací frekvence byla zvolena 48 kHz, bitová hloubka pro nahrávání 24 bitů.

4.4. Průběh nahrávání

Nahrávky hlasu byly pořízeny na základě získaných zkušeností nabytých při pořizování testovacích nahrávek, které vznikly na podzim roku 2016 a sloužily k ověření kvality nahrávek. Bylo pořízeno celkem 12 nahrávek od dvou řečníků (muže a ženy). Celková délka těchto nahrávek je 7 minut a 47 vteřin. Díky těmto nahrávkám byly upraveny texty ukázek, které řečníci četli. Novinové články se ukázaly jako nevhodné, protože se v nich vyskytuje velké množství jmen. I když řečníci nečtou ukázku poprvé, dělají jim tato jména problémy. Z tohoto důvodu tvoří nové ukázky převážně úryvky knižních textů. Pokud se v ukázkách vyskytují jména, tak všeobecně známá a lehce vyslovitelná. Soubor textových ukázek tvoří celkem 35 ukázek dlouhých přibližně 100 slov (délka ukázek není striktně dodržována). Každý mluvčí nahrál pouze 25 ukázek, tudíž 10 ukázek, které se mu četly nejhůře, mohl vyřadit.

Vždy nejpozději týden před termínem nahráváním byly řečníkům předány texty, aby byl řečníkům poskytnut dostatečný čas pro jejich pročtení. Tím pádem proběhlo nahrávání plynuleji a s menším počtem přeřeknutí, než kdyby řečníci viděli texty poprvé až při nahrávání. Před příchodem mluvčích byl připraven nahrávací řetězec a realizována komunikace mezi kontrolní místností a bezodrazovou komorou pro udělování pokynů řečníkům. V bezodrazové komoře byla vyznačena vzdálenost 50 cm od mikrofonu, ve které řečníci stáli.

Obr. 4.1: Řečník během nahrávání

Po příchodu byl řečníkům vysvětlen postup nahrávání a proběhla případná individuální domluva, která se týkala přestávek a střídání řečníků. Pokud bylo přítomných více řečníků zároveň, osvědčil se postup střídat řečníky po určitém počtu ukázek (nejčastěji po deseti). Každý řečník byl na konci tohoto bloku ukázek náchylný na přeřeknutí, tudíž byl vystřídán jiným řečníkem a mohl si

odpočinout, zatímco byly pořizovány nahrávky jiného mluvčího. Nedocházelo tudíž k dlouhým časovým prodlevám a minimalizoval se počet opětovného nahrávání stejné ukázky kvůli přeřeknutí. Pokud přicházeli řečníci postupně a během nahrávání byl přítomen vždy jen jeden řečník, dodržovaly se domluvené bloky ukázek také, ale přestávky byly kratší, řečníci neměli tolik času na regeneraci a čas na pořízení ukázek od jednoho řečníka byl delší a řečníci dělali více chyb. Pro průběh nahrávání je tedy rozhodně lepší první postup s dvěma nebo třemi řečníky, kteří se střídají. Pokud však řečníci spěchali a nemohli při nahrávání trávit tolik času, byl preferován druhý způsob.

Během nahrávání nedocházelo k zastavování záznamu po ukázkách, ale nahrávalo se vše, opět kvůli úspoře času. Před každou ukázkou řekl mluvčí číslo ukázky. Pokud došlo k přeřeknutí nebo zadrhnutí řečníka během čtení ukázky, začal ve většině případů řečník číst od začátku věty, ve které došlo k chybě. Nepovedená část věty byla následně postprodukčně odstraněna. Pokud bylo v jedné ukázce větší množství chyb a bylo by nutné provádět mnoho střihů, byl řečník požádán, aby danou ukázku (nebo její část) přečetl znovu. Díky velkému množství střihů v jedné ukázce by se mohla vytratit autentičnost projevu řečníka. Do nahrávání jsem zasahoval pouze výjimečně. Pokud řečník udělal chybu, většinou sám zastavil, a začal číst od začátku dané věty. Pokud četl dál, byl požádán o zopakování pouze chybné věty po dokončení ukázky.

Nahrávky od všech 18 řečníků byly pořízeny v průběhu první poloviny roku 2017 během šesti nahrávacích frekvencí, které proběhly v bezodrazové komoře v učebně SC5.50 v budově T12 na VUT v Brně. První nahrávací frekvence se konala 3. 3. 2017, poslední 10. 5. 2017.

4.5. Problémy při vytváření nahrávek

Před nahráváním datasetu řeči byly vytvořeny testovací nahrávky, aby odhalili případné nedostatky v nahrávacím řetězci nebo nedokonalosti v průběhu nahrávání. Bohužel se nepodařilo všechny nedostatky odhalit a byly zjištěny až v průběhu vytváření datasetu.

Hlavním problémem byla bezesporu manipulace řečníků s papíry s texty. Kvůli lepšímu pocitu při čtení a také s ohledem na vyšší věk některých řečníků jsem se rozhodl, že papíry ponechám řečníkům v ruce, aby se jim snáze četly, než kdyby byly umístěny na stojanu na noty. Tuto možnost jsem v testovacích ukázkách ověřoval a žádné hluky papíru se v nich nevyskytují. Bohužel se tyto hluky projevily až během vytváření datasetu a kvůli tomu musely být některé ukázky zkráceny nebo vyřazeny.

Dalším problémem bylo pronikání nízkofrekvenčních hluků do bezodrazové komory vlivem nedostatečného útlumu jejich stěn. V některých nahrávkách se vyskytuje dupání lidí pohybujících se v okolí komory. Tyto zvuky se ve frekvenčním spektru nahrávek objevují pod hranicí 100 Hz, takže nezasahují do frekvenční oblasti lidského hlasu.

U řečníků, kteří přečetli větší množství ukázek, se objevovala únava, která se projevovala větším množství chyb a přeřeknutí při čtení textů. Aby nedocházelo k zbytečnému prodlužování nahrávací frekvence, byla znovu nahrána pouze věta s chybou nebo část textu, kde se chyby vyskytovali. Tato nepovedená část byla poté vystřižena z ukázky a byla nahrazena nově nahraným úryvkem.

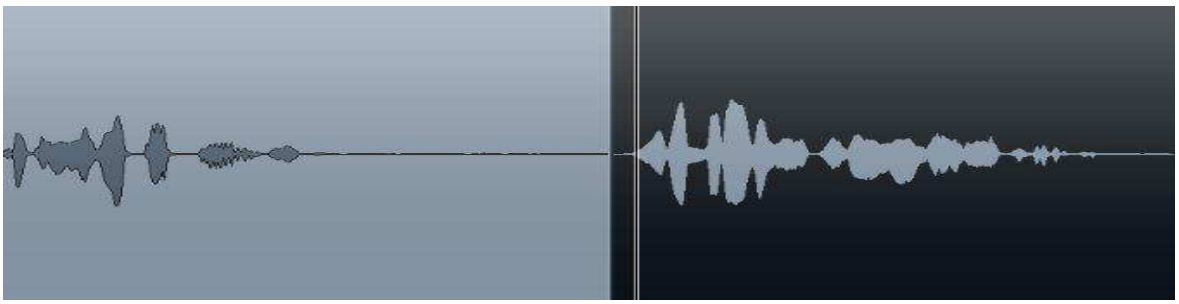
V některých nahrávkách se objevuje šum, který vznikl vlivem nastaveného většího zesílení předzesilovače. Větší zesílení bylo v některých případech nastaveno kvůli nižší úrovni vstupního signálu.

Většinu těchto nedostatků bylo možné odstranit pomocí postprodukčních úprav, které byly prováděny v editačním programu Cubase a v masteringovém programu Wavelab. U některých nahrávek nebylo možné dosáhnout dostatečné kvality, proto byly z výsledného datasetu vyřazeny, nebo byla ponechána pouze jejich část.

4.6. Postprodukční úpravy

4.6.1. Střih

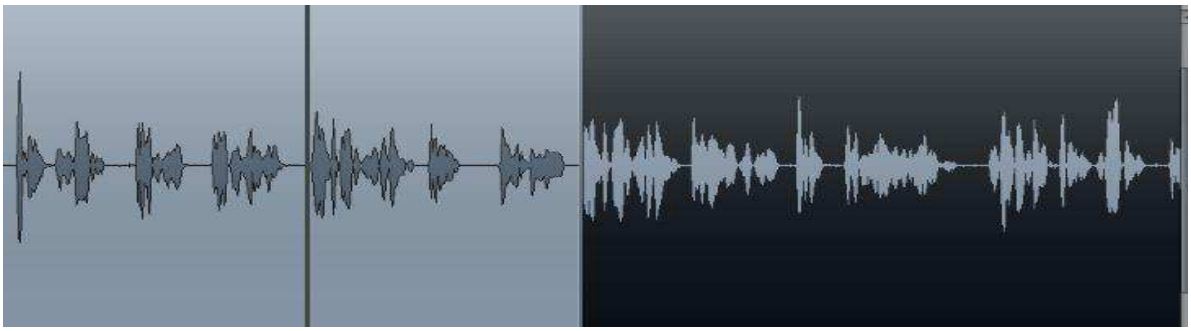
Všechny střihy byly prováděny v programu Cubase 7. Jelikož při nahrávání nedocházelo k zastavení záznamu, musel být vždy proveden střih na začátku a na konci každé přečtené ukázky. Střih na začátku ukázky je zobrazen na obrázku č. 4.2. Na obrázku úplně vlevo je záznam čísla ukázky a těsně po střihu v tmavé části začíná samotná ukázka.



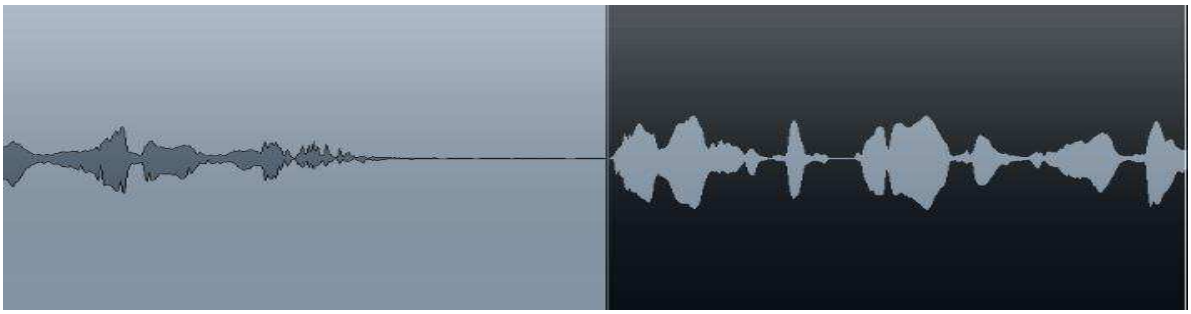
Obr. 4.2: Střih na začátku ukázky

Střih na konci byl proveden bezprostředně po posledním slovu, aby se na konci ukázky nevyskytoval nějaký šum, který pochází od řečníků (například otáčení papírů, šustění oblečení a podobně). Při střihu na konci ukázky bylo třeba dávat pozor na některé souhlásky (především na „t“), protože mohlo snadno dojít k jejich odstřižení, jelikož ve vizualizovaném záznamu připomínají právě záznam nějakého lupnutí nebo prasknutí.

Pokud při čtení ukázky došlo k chybě, musel být střih proveden i uprostřed ukázky. V ideálním případě se místo střihu vyskytovalo na začátku věty, řečník se po chybě zastavil a začal číst znovu od začátku věty a dočetl ukázku až do konce. Potom bylo místo střihu voleno po nádechu a před začátkem věty s chybou. Od tohoto místa dále byl zbytek nepovedené části ukázky vymazán a nahrazen znovu přečtenou částí. Tento případ je zachycen na obrázku č. 4.3, první střih označuje svislá černá čára, za ní se nachází věta s chybou. Na hranici světlé a tmavé oblasti se nachází druhý střih, podobný střihu na začátku ukázky. Na obrázku č. 4.4 níže, je stejná část ukázky zobrazena detailněji po vymazání prostřední části a po spojení obou zbývajících oblastí. V místě překrývání byl použit nástroj Crossfade, který zaručí hladký přechod mezi jednotlivými částmi.



Obr. 4.3: Střih věty s chybou



Obr. 4.4: Navázání na původní část nahrávky po vystřižení chybné věty

Pokud řečník po chybě dočetl ukázku do konce a opakoval pouze nepovedenou větu, nacházela se v ukázce dvě místa střihu. První střih se nacházel mezi nádechem a začátkem věty. Druhý střih, který navazoval na původní ukázku, se lišil případ od případu: v některých případech byl proveden ještě před koncem věty, která byla opakována. Velmi záleželo na pomlkách, které řečník dělal mezi větami v souvětí a na pozici chyby.

Některá drobná přeřeknutí, kterých jsem si během nahrávání nevšiml nebo která nebyla příliš závažná, jsou v ukázkách ponechána. Pokud řečník drobné zaváhání ihned napravil a neponechal mezeru mezi chybným slovem a zopakovaným slovem, byl střih v tomto místě obtížně proveditelný. Střih byl přesto proveden a poté bylo posouzeno, zda zní tato část ukázky věrohodně. Pokud ne, byla ponechána původní verze s drobnou chybou.

Pro střih byl v editačním programu Cubase použit nástroj „Split“ se zapnutým módem „Snap to Zero Crossing“. Tento mód zaručuje střih v bodě s nulovou amplitudou signálu. Díky tomu při střihu nedojde k náhlému poklesu amplitudy signálu, který je během přehrávání doprovázen charakteristickým lupnutím nebo prasknutím. Po kontrole místa střihu byly obě části nahrávky k sobě pevně spojeny pomocí nástroje „Glue“, aby nedošlo k posunutí některé části nahrávky.

4.6.2. Potlačení šumu

Všechny ostatní úpravy nahrávek byly prováděny v masteringovém programu Wavelab 7. Nejprve byl potlačen šum předzesilovače, který předzesilovač produkoval při větším zesílení. Předzesilovač umožňoval pouze skokové změny zesílení. Po zjištění nízké úrovně odstupu signálu od šumu (SNR) při větším zesílení předzesilovače byla používána pouze nižší hodnota zesílení. Nicméně část nahrávek byla tímto chybným nastavením poškozena.

Pro zvětšení odstupu signálu od šumu byl použit VST zásuvný modul DeNoiser od firmy Steinberg s maximální úrovní redukce šumu. Tento modul byl aplikován na všechny nahrávky, i když žádný šum na pozadí nahrávky nebyl slyšitelný.

4.6.3. Filtrace spektra signálu

Po zvětšení úrovně SNR byly u nahrávek potlačeny nízké kmitočty, aby bylo z nahrávek odstraněno dupání. To bylo provedeno filtrem typu horní propust s mezním kmitočtem 120 Hz a se strmostí 48 dB/okt. Po použití tohoto filtru bylo dupání dostatečně potlačeno a není pří poslechu nahrávek patrné.

4.6.4. Normalizace úrovně

Jelikož se hlasitost projevu jednotlivých řečníků velmi lišila, bylo nutné postprodukčně nastavit hlasitost jednotlivých nahrávek na stejnou úroveň. Tento proces byl poslední úpravou nahrávek před jejich závěrečným exportem.

Program Wavelab umožňuje normalizovat úroveň nahrávek na přesnou efektivní hodnotu hlasitosti. Všechny nahrávky byly normalizovány na úroveň *L = -25 dB*. Efektivní hodnota hlasitosti nahrávek řeči mohla být zvolena i o 5 dB vyšší, ale při normalizaci úrovně hlasitosti nahrávek hluků by při nastavení stejné efektivní hodnoty jako u nahrávek hlasu mohlo docházet k ořezání špiček signálu, proto byla raději zvolena nižší hodnota. Při závěrečném exportu byla snížena bitová hloubka z 24 bitů na 16. Pro záznam mluveného slova je tato hodnota dostačující a dojde k úspoře místa na disku.

4.7. Struktura databáze

Pro snadnější orientaci v databázi jsou nahrávky jednotlivých řečníků roztřízeny do složek. Název složky tvoří unikátní kódové označení řečníka, ze kterého lze určit pohlaví a věk řečníka. Jednotlivé údaje jsou odděleny pomlčkami. Prvním znakem kódu je písmeno M nebo F, pro označení pohlaví řečníka (M pro muže). Další dvě číslice udávají věk řečníka a poslední dvě jeho pořadové číslo v průběhu vytváření datasetu.

V každé složce se nachází maximálně 25 nahrávek řečníka, ve většině případů však méně, kvůli vyřazeným nahrávkám. Jejich název tvoří opět kódové označení řečníka a číslo ukázky. Obě části jsou také odděleny pomlčkou. Texty jednotlivých ukázek jsou uvedeny v příloze.

Například název nahrávky M-61-17-020 označuje nahrávku ukázky číslo 020, kterou čte řečník muž, kterému bylo v době nahrávání 61 let a který nahrával jako 17. v průběhu vytváření datasetu.

Z celkových 315 minut nahrávek řeči tvoří 165 minut záznam mužského hlasu a 150 minut ženského hlasu. Z délky nahrávek jednotlivých řečníků můžeme pozorovat, že s rostoucím věkem průměrná délka nahrávek vzrůstá. Výjimkou je řečník F-22-02, který i přes nízký věk dosahuje průměrné délky nahrávky téměř 55 vteřin a také řečník F-76-15, u kterého je velké množství nahrávek vyřazeno nebo zkráceno, proto průměrná nahrávka dosahuje délky jen 42 vteřin. Až na tyto výjimky platí tento trend jak u mužů, tak u žen. K hlubšímu zkoumání této zajímavosti by bylo nutné získat nahrávky většího počtu řečníků.

4.8. Návrhy na rozšíření databáze

Databáze sice obsahuje poměrně velké množství nahrávek, ale pro některé účely by nemusela být dostačující. Je možné databázi rozšířit podle potřeby, ale v rámci zachování kompatibility nahrávek je nutné dodržet tyto podmínky:

- Zachovat prostředí. Je bezpodmínečně nutné vytvořit nahrávky v bezodrazové komoře, v ideálním případě ve stejné komoře. V případě nedodržení této podmínky by databáze obsahovala v podstatě dva různé celky nahrávek.
- Zvolit podobné přístroje. Při rozšiřování databáze je nutné použít zejména mikrofon s podobnými parametry, jaké měl mikrofon použitý pro pořízení původních nahrávek.

Dataset by mohl například po rozšíření o nahrávky většího počtu řečníků sloužit jako trénovací množina pro neuronové sítě, které rozpoznají pohlaví nebo odhadnou věk řečníka. Pro tento účel dataset v současné podobě obsahuje nahrávky příliš malého počtu řečníků.

V případě pořizování nahrávek v bezodrazové komoře na VUT doporučuji komoru využívat ve dnech, kdy se po budově pohybuje co nejmenší množství lidí. To znamená v pátek odpoledne a o víkendu. Díky tomu lze minimalizovat přítomnost dupání v nahrávkách.

Dále doporučuji umístit texty, které budou čtenáři číst na notový stojan, aby nedrželi řečníci papír v ruce a nedocházelo k záznamu nežádoucích zvuků, který v případě původního datasetu vedl k vyřazení téměř 50 nahrávek.

5. Nahrávky šumu

5.1. Využití nahrávek

Nahrávky šumů a ruchů primárně slouží k umělému zašumění nahrávek hlasu. Uměle zašuměné nahrávky budou využity jako trénovací a testovací množina pro umělé neuronové sítě, které budou schopny oddělit řeč od šumové složky. Pro tento účel využití nahrávek jsou dostačující monofonní nahrávky. Nicméně nahrávky této databáze jsou nahrány stereofonním systémem XY. Díky tomu je možné pro zašumení použít signál z levého, nebo pravého kanálu. Stereofonní nahrávky je možné použít také za jiným, například uměleckým účelem. Některé nahrávky jsem použil například pro svůj umělecký projekt. Nahrávky pořízené tímto systémem jsou plně mono kompatibilní, takže sloučení obou kanálů do jednoho nahrávku nijak nepoškodí. Všechny nahrávky byly pořízeny v Brně nebo v jeho blízkém okolí.

5.2. Specifikace použitých mikrofonů a ostatního hardwaru

K vytvoření databáze šumů jsem si vybral příruční rekordér Zoom H4nSP, především pro jeho kompaktnost. Jeho součástí jsou dva všesměrové kondenzátorové mikrofony v konfiguraci XY s možností volby úhlu, který svírají (90° nebo 120°). Tento rekordér umožňuje A/D převod do formátu WAV se vzorkovací frekvencí 44.1 kHz, 48 kHz nebo 96 kHz a bitovou hloubkou 16 nebo 24 bitů. Díky jeho velikosti a malé hmotnosti je pro terénní nahrávání ideální a díky systému XY je navíc i monaurálně kompatibilní. Pro nahrávání byla použita vzorkovací frekvence 48 kHz, bitová hloubka 24 bitů a úhel mezi XY mikrofony 90°.

5.3. Údaje o jednotlivých nahrávkách

Hluk H01

- **místo:** křižovatka ulic Česká a Joštova, zastávka tramvaje 4,5 a 6, naproti přes ulici knihkupectví Dobrovský.
- **datum a čas:** 23. 11. 2016, začátek nahrávání 16:32, konec nahrávání 16:42
- **délka nahrávky:** 8:56
- **zaznamenané zvuky:** příjezdy a odjezdy tramvají, projíždějící automobily, procházející dav lidí

Hluk H02

- **místo:** roh ulic Veveří a Kotlářská, zastávka tramvaje 3, 9, 11 a 12.
- **datum a čas:** 20. 4. 2017, začátek nahrávání 16:42, konec 16:54
- **délka nahrávky:** 9:04
- **zaznamenané zvuky:** příjezdy a odjezdy tramvají, projíždějící auta a auta stojící na křižovatce, procházející lidé

Hluk H03

- **místo:** Technologický park Brno, poblíž lávky přes silnici 640 (Hrádecká, Královo Pole)
- **datum a čas:** 21. 4. 2017, začátek nahrávání 11:30, konec nahrávání 11:42
- **délka nahrávky:** 11:37
- **zaznamenané zvuky:** především projíždějící auta, na začátku nahrávky také šustění suchého listí.

Hluk H04

- **místo:** Vlaková linka S41, část trasy mezi Brnem a Moravskými Bránicemi
- **datum a čas:** 23. 4. 2017, začátek nahrávání 18:10, konec nahrávání 18:20
- **délka nahrávky:** 15:20
- **zaznamenané zvuky:** zvuky vlakové soupravy, lidé procházející uličkou

Hluk H05

- **místo:** Komenského náměstí, u budovy HF JAMU
- **datum a čas:** 25. 4. 2017, začátek nahrávání 13:39, konec nahrávání 13:53
- **délka nahrávky:** 12:35
- **zaznamenané zvuky:** projíždějící tramvaje a automobily

Hluk H06

- **místo:** Ulice Husova, naproti hotelu International
- **datum a čas:** 25. 4. 2017, začátek nahrávání 14:10, konec nahrávání 14:23
- **délka nahrávky:** 12:20
- **zaznamenané zvuky:** projíždějící tramvaje a automobily

5.4. Postprodukční úpravy

5.4.1. Střih

Z hlediska primárního využití nahrávek šumu je nežádoucí, aby se v nich vyskytoval zřetelně lidský hlas. Přítomnost lidského hlasu by mohla mít vliv na proces učení neuronové sítě a negativně ovlivnit funkčnost sítě. Při pořizování nahrávek v městském prostředí však bylo velmi obtížně dosažitelné pořídit nahrávku bohatou na ruchy a bez přítomnosti zřetelných frází. Proto byly z pořízených nahrávek části, ve kterých se hlas vyskytoval příliš zřetelně vystřiženy.

Vzhledem k povaze nahraného signálu, tedy hluku, u kterého se mění intenzita i směr, odkud přichází, téměř náhodně, byl střih při odstraňování řeči proveden téměř na libovolném místě a po vymazání nežádoucí části byly oddělené části nahrávky spojeny pomocí nástroje Crossfade, který zaručí hladký přechod mezi jednotlivými částmi nahrávky v místě střihu. V některých případech (například v nahrávce H04) bylo zapotřebí použít funkci „Equal Power“, protože v místě prolínání obou částí nahrávky docházelo k poklesu hlasitosti vlivem nekoherentnosti obou signálů.

Střih byl proveden, stejně jako v případě nahrávek z bezodrazové komory, v editačním programu Cubase.

5.4.2. Normalizace

Po sestříhání nahrávek byla nastavena jejích hlasitost, stejně jako u nahrávek hlasu, pomocí masteringového softwaru Wavelab. Každá nahrávka byla uložena ve třech úrovních hlasitosti: L_0 = -25 dB, L_{-5} = -30 dB a L_{-10} = -35 dB. Po zašumení nahrávek hlasu bude tedy odstup signálu od šumu 0 dB, 5 dB a 10 dB. Tato hodnota je však pouze orientační, protože zatímco u nahrávek hlasu je hodnota dynamického rozsahu malá, u nahrávek hluku je extrémně vysoká.

Přesná hodnota odstupu signálu od šumu bude tedy záviset na pozici v nahrávce hluku, kam bude nahrávka hlasu vložena. Pro učení neuronové sítě však není přesná hodnota odstupu signálu od šumu podstatná. Pro smíchání nahrávek šumu s nahrávkami hlasu byla opět snížena bitová hloubka nahrávek šumu z 24 bitů na 16 bitů. Součástí datasetu jsou však i 24 bitové stereofonní nahrávky, které však nejsou normalizovány.

5.5. Návrhy na rozšíření databáze

Většina mnou pořízených nahrávek v rámci bakalářské práce pochází z městského prostředí, bylo by tedy vhodné v případě potřeby databázi rozšířit o nahrávky z jiných míst, kde se budou vyskytovat rozdílné hluky. Přínosné by mohly například být nahrávky z továren, z okolí vodních toků a podobně.

6. Databáze uměle zašumených nahrávek hlasu

Po všech postprodukčních úpravách byly nahrávky hlasu uměle zašumeny nahranými hluky. Každá nahrávka hlasu byla smíchána s každým z šesti šumů ve třech různých úrovních odstupu signálu od šumu. Výsledkem je tedy 7290 zašumených nahrávek hlasu s odstupem signálu od šumu 0 dB, 5 dB a 10 dB.

6.1. Skript pro kombinaci audionahrávek

Kombinování nahrávek bylo provedeno za pomocí skriptu v jazyce Python, který je uveden v příloze. Tento skript byl vytvořen ve spolupráci s Ondřejem Talárem[1]. Skript využívá command line utilitu SoX [18], díky které lze v jazyce Python zpracovávat audio signály. Skript vybere náhodnou pozici v nahrávce šumu, na kterou vloží nahrávku hlasu a upraví délku nahrávky šumu podle délky nahrávky hlasu. Díky tomu nebudou žádné dvě nahrávky hlasu zkombinovány se dvěma totožnými nahrávkami šumu. Navíc jsou použity signály z obou kanálů stereofonního záznamu a jejich součet.

Pro použití skriptu je nutné zadat vlastní cestu k nahrávkám hlasu, k nahrávkám šumu a do cílové složky, do které budou nahrávky uloženy. Výsledný počet nahrávek by měl být násobkem nahrávek hlasu a nahrávek šumu. Na konci procesu je počet vytvořených nahrávek porovnán s předpokládaným počtem. Pokud množství souborů neodpovídá, je vypsána chybová hláška, která zobrazí počet předpokládaných a skutečně vytvořených nahrávek. Během testování skriptu došlo k vytvoření menšího počtu nahrávek jen ve dvou případech: při různých vzorkovacích frekvencích kombinovaných signálů a při různých formátech, ve kterých byly signály uloženy (WAV a MP3). Při rozdílné bitové hloubce nebo při různém počtu kanálů kombinovaných nahrávek byly všechny očekávané soubory vytvořeny.

6.2. Struktura databáze

Označení zašumené nahrávky se skládá ze dvou částí- z označení nahrávky hlasu a z názvu nahrávky šumu. Obě tyto části jsou odděleny podtrhávací čarou („_“). Označení nahrávky hluku je navíc doplněno o informaci o použitém kanálu stereofonního záznamu a o odstupu signálu od šumu.

Například nahrávka s označením M-51-14-023_H02-L-0dB označuje nahrávku ukázky číslo 023, kterou nahrál muž ve věku 51 let, která je zkombinována se signálem z levého kanálu hluku H02 se stejnou hlasitostí signálu a hluku.

Kvůli velkému objemu dat jsou nahrávky uložena na 6 DVD discích o kapacitě 8,5 Gb, které jsou součástí přílohy. První disk obsahuje nahrávky hlasu a nahrávky šumu a to jak v rozlišení 24 bitů, tak v 16 bitové verzi, která byla použita na kombinaci s nahrávkami hlasu. Nahrávky šumu s bitovou hloubkou 24 bitů jsou stereofonní a není normalizována jejich hlasitost. Na prvním disku je také uložena veškerá dokumentace k datasetu. Dalších 5 disků potom obsahuje celkem 7290 zašumených nahrávek lidského hlasu. Přesný obsah jednotlivých disků je uveden v příloze.

[1] TALÁR, Ondřej Redukce šumu audionahrávek pomocí hlubokých neuronových sítí: diplomová práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2017. 29 s. Vedoucí práce byl Ing. Pavol Harár

6.3. Vlastnosti databáze

Databáze vytvořená v rámci této semestrální práce se svými vlastnostmi nejvíce podobá databázím Aurora a SPINE, které byly popsány v této práci výše (2.3.2 a 2.3.6). Hlavním rozdílem mezi databázemi je způsob vzniku nahrávek. Tato databáze vznikla nahráváním čteného textu, zatímco ve SPINE jsou zachyceny nahrávky rozhovorů a Aurora vychází z datasetu TIDIGITS, což jsou nahrávky sekvencí čísel. Hluky, které obsahují nahrávky databáze SPINE, jsou z vojenského prostředí. Tato databáze obsahuje nahrávky hluků z městského prostředí. Ostatní databáze převyšují tuto databázi několikanásobně počtem řečníků i délkou a počtem nahrávek. Rozsahem nahrávek řeči se tato databáze řadí k těm menším, například k Berlin Database of Emotional Speech (2.3.7), která se však zaměřuje na nahrávky emoční řeči. Tato databáze i Berlin Database of Emotional Speech vznikaly v bezodrazové komoře. Celkovým počtem nahrávek (7290) se však řadí mezi středně velké databáze. Největší odlišností proti ostatním databázím je jazyk nahrávek. Až na několik výjimek jsou všechny ostatní v anglickém jazyce, tato databáze je v českém jazyce. Databáze se také vyznačuje vyrovnaností délky nahrávek mužů a žen (9 mužů a 9 žen) a také širokým věkovým rozpětím řečníků, kteří jsou v daném věkovém rozpětí zastoupeni rovnoměrně a v rámci věkových skupin vyrovnaně.

7. Závěr

V rámci této bakalářské práce byly nastudovány postupy pro vytvoření vědeckého datasetu nahrávek lidského hlasu, který bude sloužit jako trénovací množina pro hluboké neuronové sítě. Na základě těchto znalostí a na základě zkušeností získaných při pořizování testovacích nahrávek byl takový dataset navržen a následně vytvořen.

Dataset se skládá z 405 nahrávek řeči, 6 nahrávek hluku a z 7290 zašumených nahrávek řeči. Nahrávky řeči byly pořízeny od 18 řečníků (9 mužů a 9 žen) a jejich celková délka je 315 minut. Nejmladšímu řečníkovi bylo v době nahrávání 16 let a 10 měsíců, nejstaršímu 76 let a 7 měsíců. V tomto věkovém rozmezí byli řečníci v rámci možností rovnoměrně rozprostřeni. Nahrávky šumu pochází z městského prostředí a jejich celková délka je téměř 70 minut. Každá nahrávka řeči byla zašumena každou nahrávkou šumu se třemi rozdílnými hodnotami odstupu řeči od šumu. Takto vzniklo 7290 zašumených nahrávek. Dataset se vyznačuje poměrně vysokou kvalitou nahrávek řeči, které vznikly v bezodrazové komoře, a je unikátní především tím, že jsou všechny nahrávky v českém jazyce. Součástí datasetu je také tabulka s informacemi o každé nahrávce, skript v jazyce Python, pomocí kterého byly nahrávky hluku a hlasu kombinovány a průvodní dokument prezentující souhrnné informace o datasetu.

Seznam použité literatury

[1] PUTNA, L.: Predikce vývoje kurzu pomocí umělých neuronových sítí. Brno: Vysoké Učení Technické v Brně, Fakulta Informačních Technologií, 2011.

[2] BARBER, S.: AI : Neural Network for beginners – CodeProject.

[3] NIELSEN, Michael A. Neural networks and deep learning. URL: http://neuralnetworksanddeeplearning.com/, 2015.

[4] GAROFOLO, John S., et al. DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1. *NASA STI/Recon technical report n*, 1993, 93.

[5] GONG, Yifan. Speech recognition in noisy environments: A survey. *Speech communication*, 1995, 16.3: 261-291.

[6] LAMEL, Lori, et al. The translanguage English database (TED). In: *ICSLP*. 1994.Santa Barbara Corpus of Spoken American English

[7] DU BOIS, John W., et al. Santa Barbara Corpus of Spoken American English. *CD-ROM. Philadelphia: Linguistic Data Consortium*, 2000.

[8] LEONARD, R. Gary; DODDINGTON, George. Tidigits. *Linguistic Data Consortium, Philadelphia*, 1993.

[9] PEARCE, David, et al. The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In: *INTERSPEECH*. 2000. p. 29-32.

[10] BURKHARDT, Felix, et al. A database of German emotional speech. In: *Interspeech*. 2005. p. 1517-1520.

[11] SAKAR, Betul Erdogdu, et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 2013, 17.4: 828-834.

[12] COLE, Ron. The ISOLET spoken letter database. 1990.

[13] ZÍKOVÁ, Magdaléna; KŘIVAN, Jan. Nahrávání v terénním lingvistickém výzkumu: jak získat kvalitní záznam řeči?. *Studie z aplikované lingvistiky-Studies in Applied Linguistics*, 2014, 5.1: 65-82.

[14] SYROVÝ, Václav. *Hudební akustika*. Akademie múzických umění, 2003.

[15] LAVRY, Dan. Sampling Theory For Digital Audio. Lavry Engineering, Inc. Available online: *http://www*.lavryengineering.com/documents/Sampling_Theory. pdf (checked 24.5. 2010), 2004.

[16] VLACHÝ, Václav. *Praxe zvukové techniky*. Muzikus, 2008.

[17] SCHIMMEL, J. *Electroacoustics.* Electroacoustics. 1. Brno: Vysoké učení technické v Brně, 2015. s. 1-130. ISBN: 978-80-214-5170- 4

[18] BAGWELL, Chris; KLAUER, U. Sox-sound exchange. *Online Website*, 2010.

[19] ORWELL, George; GÖSSEL, Gabriel. *Farma zvířat*. Radioservis, 2009.

[20] SAINT-EXUPÉRY, A. de. Malý princ. 1998.

[21] JIROTKA, Zdeněk. *Saturnin*. Karolinum, 2006.

[22] ŠEDIVÝ, Petr. Anglie a Skotsko se vzepřely vrchnosti, za vlčí máky mohou dostat trest. *Idnes.cz* [online]. [cit. 2017-05-18]. Dostupné z: http://fotbal.idnes.cz/anglie-a-skotsko-nastoupily-se-symboly-vlciho-maku-mohou-za-to-dostat-trest-1v0-/fot_reprez.aspx?c=A161112_010559_fot_reprez_pes

[23] STEINBECK, John; VENDYŠ, Vladimír. *O myších a lidech*. Československý spisovatel, 1960.

[24] HEMINGWAY, Ernest. *Stařec a moře*. Praha: Odeon, 2015. ISBN 978-80-207-1621-7.

[25] ADAMS, Douglas; HOLLANOVÁ, Jana. *Stopařův průvodce po galaxii*. Hynek, 1998.

Seznam obrázků

Seznam příloh

A Seznam řečníků

B Textový materiál

C Skript na kombinaci audionahrávek

D Přílohy na DVD

A Seznam řečníků

Řečník M-21-01

- **Pohlaví:** muž
- **Věk:** 21 let, 10 měsíců
- **Datum nahrávání:** 3. 3. 2017
- **Počet nahrávek:** 24
- **Délka nahrávek:** 17:02
- **Poznámky:** na výběr pouze ukázky 001 až 032

Řečník F-22-02

- **Pohlaví:** žena
- **Věk:** 22 let, 2 měsíce
- **Datum nahrávání:** 3. 3. 2017
- **Počet nahrávek:** 22
- **Délka nahrávek:** 20:19
- **Poznámky:** na výběr pouze ukázky 001 až 032

Řečník F-16-03

- **Pohlaví:** žena
- **Věk:** 16let, 10 měsíců
- **Datum nahrávání:** 3. 3. 2017
- **Počet nahrávek:** 22
- **Délka nahrávek:** 14:49
- **Poznámky:** na výběr pouze ukázky 001 až 032

Řečník F-19-04

- **Pohlaví:** žena
- **Věk:** 19 let, 5 měsíců
- **Datum nahrávání:** 3. 3. 2017
- **Počet nahrávek:** 19
- **Délka nahrávek:** 12:25
- **Poznámky:** na výběr pouze ukázky 001 až 032

Řečník M-70-05

- **Pohlaví:** muž
- **Věk:** 70 let, 1 měsíc
- **Datum nahrávání:** 31. 3. 2017
- **Počet nahrávek:** 25
- **Délka nahrávek:** 23:40

Řečník M-65-06

- **Pohlaví:** muž
- **Věk:** 65 let, 8 měsíců
- **Datum nahrávání:** 31. 3. 2017
- **Počet nahrávek:** 21
- **Délka nahrávek:** 20:20
- **Poznámky:** dlouholetý kuřák

Řečník F-64-07

- **Pohlaví:** žena
- **Věk:** 64 let, 5 měsíců
- **Datum nahrávání:** 31. 3. 2017
- **Počet nahrávek:** 24
- **Délka nahrávek:** 20:08
- **Poznámky:** vada řeči

Řečník M-18-08

- **Pohlaví:** muž
- **Věk:** 18 let, 9 měsíců
- **D Datum nahrávání:** 07. 04. 2017
- **Počet nahrávek:** 21
- **Délka nahrávek:** 12:33

Řečník F-44-09

- **Pohlaví:** žena
- **Věk:** 44 let
- **Datum nahrávání:** 07. 04. 2017
- **Počet nahrávek:** 22
- **Délka nahrávek:** 18:39

Řečník M-39-10

- **Pohlaví:** muž
- **Věk:** 39 let, 10 měsíců
- **Datum nahrávání:** 07. 04. 2017
- **Počet nahrávek:** 25
- **Délka nahrávek:** 17:03
- **Poznámky:** dlouholetý kuřák, operace hlasivek

Řečník M-22-11

- **Pohlaví:** muž
- **Věk:** 22 let, 4 měsíce
- **Datum nahrávání:** 07. 04. 2017
- **Počet nahrávek:** 25
- **Délka nahrávek:** 18:01

Řečník M-34-12

- **Pohlaví:** muž
- **Věk:** 34 let, 9 měsíců
- **Datum nahrávání:** 10. 04. 2017
- **Počet nahrávek:** 24
- **Délka nahrávek:** 17:37

Řečník F-34-13

- **Pohlaví:** žena
- **Věk:** 34 let, 6 měsíců
- **Datum nahrávání:** 10. 04. 2017
- **Počet nahrávek:** 22
- **Délka nahrávek:** 15:52

Řečník M-51-14

- **Pohlaví:** muž
- **Věk:** 51 let, 8 měsíců
- **Datum nahrávání:** 21. 4. 2017
- **Počet nahrávek:** 22
- **Délka nahrávek:** 17:52

Řečník F-76-15

- **Pohlaví:** žena
- **Věk:** 76 let, 7 měsíců
- **Datum nahrávání:** 21. 4. 2017
- **Počet nahrávek:** 15
- **Délka nahrávek:** 10:33

Řečník F-70-16

- **Pohlaví:** žena
- **Věk:** 70 let, 2 měsíce
- **Datum nahrávání:** 21. 4. 2017
- **Počet nahrávek:** 22
- **Délka nahrávek:** 21:17

Řečník M-61-17

- **Pohlaví:** muž
- **Věk:** 61 let, 11 měsíců
- **Datum nahrávání:** 10. 5. 2017
- **Počet nahrávek:** 25
- **Délka nahrávek:** 21:05

Řečník F-26-18

- **Pohlaví:** žena
- **Věk:** 26 let, 10 měsíců
- **Datum nahrávání:** 10. 5. 2017
- **Počet nahrávek:** 25
- **Délka nahrávek:** 16:11

B Textový materiál

Textová ukázka 001 až 003 pochází z knihy Farma zvířat [19] od George Orwella. Ukázka 004 je z knihy Malý Princ [20], jejímž autorem je Antoine de Saint-Exupéry. Ukázky 005 až 008 čerpají z knihy Saturnin [21] od Zdeňka Jirotky. Ukázka 009 je částí novinového článku [22], který vyšel na serveru idnes.cz a autorem je Petr Šedivý. Ukázky 010 až 012 pocházejí z knihy O myších a lidech [23] od Johna Steinbecka, ukázky 013 až 032 z knihy Stařec a moře [24] od autora Ernesta Hemingway**e.** Poslední 3 ukázky (033-035) pochází ze Stopařova průvodce po galaxii [25] od Douglase Adamse.

001

Přišel listopad a s ním nárazy jihozápadních vichrů. Stavba se musela přerušit, protože na míchání cementu bylo příliš vlhko. Jedné noci přišla tak silná vichřice, že se budovy kymácely v základech a se střechy stodoly odnesl vítr několik tašek. Slepice se probudily, kvokaly hrůzou, protože se jim zdálo, že někde v dálce slyší ránu jako z děla. Když ráno zvířata vyšla ze stájí, zjistila, že stožár vlajky je zlomený a jeden jilm na konci sadu vyvrácený z kořenů jak ředkvička. A vzápětí zazněl z hrdel všech zvířat zoufalý sten. Jejich zrakům se naskytl hrůzný pohled: větrný mlýn se zbortil.

002

Celé léto klapala práce na farmě jako hodinky. Všechna zvířata byla šťastnější, než by si kdo mohl představit. Každé sousto píce bylo pro ně opravdovým potěšením, neboť vše pocházelo z jejich vlastní práce, nebyla to již almužna od nenáviděného pána. Když odešli neužiteční lidští příživníci, každý se mohl dosyta najíst. A ačkoliv neměla žádné zkušenosti, získala mnohem více volného času. Objevila se samozřejmě spousta potíží: při sklizni obilí musela zvířata například sešlapávat klasy jako za dávných časů a vyfukovat plevy vlastním dechem, protože na farmě nebyla mlátička.

003

Jedno nedělní odpoledne, když se zvířata shromáždila, aby vyslechla příkazy, oznámil Napoleon, že se rozhodl pro novou hospodářskou politiku. Od nynějška bude Farma zvířat obchodovat se sousedními farmami. Pochopitelně, že ne za komerčním účelem, ale jen proto, aby mohla získat nejpotřebnější materiál. Větrný mlýn musí stát nad vším ostatním, řekl. V současné době proto dojednává prodej fůry sena a části letošní úrody ječmene, a později, bude-li nutné získat další peníze, domluví se prodej vajec. Jak řekl Napoleon, slepice by tuto oběť měly uvítat jako svůj zvláštní přínos pro stavbu větrného mlýna.

004

Pátá planeta byla velmi zvláštní. A ze všech nejmenší. Bylo tu právě tak dost místa, aby se sem vešla pouliční svítilna a lampář. Malý princ si nedovedl vysvětlit, k čemu může být někde na planetě bez domů a bez lidí svítilna a lampář. Přesto si v duchu řekl:

Možná že ten člověk je zbytečný. A přece je méně zbytečný než král, než domýšlivec, než byznysmen a než pijan. Jeho práce má aspoň smysl. Když rozsvítí svítilnu, jako by se zrodilo o hvězdu nebo o květinu víc. Když zhasne svítilnu, jako by květina nebo hvězda šly spát. Je to moc hezké zaměstnání a opravdu užitečné, protože je hezké.

005

Příštího dne bylo krásné nedělní jitro. Ležel jsem v poduškách a vdechoval svěží vzduch, který proudil otevřeným oknem ze zahrady. Po bydlení v obývací lodi vychutnával jsem až do dna pobyt v solidním domě svého dědečka. Pociťoval jsem příjemné vědomí bezpečnosti, jistotu, že se neozve z paluby poplašný Saturninův křik, abych se držel, že jede kolem parník "Praha". Přitom jsem uvažoval, že musím promluvit se Saturninem vážné slovo. Doktor Vlach se totiž na naší lodi občas vyjadřoval o dědečkovi s notnou dávkou obvyklého sarkasmu a Saturnin těm hovorům naslouchal s neskrývaným potěšením. Obávám se, že nabyl o dědečkovi mínění poněkud neuctivého, a mrzí mne, že jsem ty hovory trpěl.

006

Těžké černé mraky byly stále nižší a stále více se stmívalo. Všiml jsem si, že jsme se všichni instinktivně přikrčili. Slečna Barbora zaplatila benzín, stiskla startér a v desíti metrech už jela na trojku. Prolétli jsme městem a Barbora rozsvítila reflektory. Reflektory v poledne! Díval jsem se před sebe v napjatém očekávání, že se v nejbližší chvíli rozpoutá peklo. Saturninovi podivně svítily oči a já jsem měl pocit, že má radost jednak z toho, že se děje něco neobvyklého, jednak ze slečny Barbory. Řídila vůz s obdivuhodnou bravurou a přidávala plyn s lehkomyslností lidí, kteří nikdy automobil nerozbili.

007

Nakonec měl doktor Vlach docela rozumný plán. Navrhoval, abychom počkali ještě dva dny, nebude-li započato se stavbou mostu. Potom jsme měli vzít potravu na poslední den s sebou a vydat se na cestu k jeho srubu. Je to krásná a pohodlná túra, a vyjdeme-li kolem osmé hodiny ráno, dorazíme tam brzo po poledni. Je tam dost místa, abychom tam všichni přenocovali. Samozřejmě, nebude to tak pohodlné jako zde, ale nic jiného nám nezbývá. Příštího dne vyrazíme časně ráno k Bílému sedlu, obejdeme prameny řeky, přes Vřesové studánky sestoupíme do údolí a pokusíme se ještě téhož dne dorazit do městečka. Bude to úkol poněkud tvrdý, ale lze to dokázat.

008

Příští den uplynul, aniž by se stalo cokoliv pozoruhodného. Dělníci pana Novotného se neobjevili. Saturnin, jak jsem předpokládal, byl nadšen vyhlídkou na náš podnik a říkal tomu "pochod hladu". Pravil, že se takové věci stávají v Číně, jenže v daleko větším rozsahu. Tam se na takový pochod dá třeba milión lidí. Dědeček říkal, že je to nesmysl, protože kdyby se milión lidí postavil do řady, mohli by si podávat potraviny z městečka až sem. Saturnin namítal, že každý člověk z toho miliónu musí jíst, a že kdyby si každý v té řadě ukousl, zemřeli bychom tu hladem stejně.

009

Angličtí a skotští fotbalisté měli ve vzájemném zápase na rukou černé pásky s obrázky vlčích máků, který je symbolem Dne válečných veteránů. Oběma reprezentacím za to teď od Světové fotbalové federace hrozí trest.

FIFA před pár dny zamítla žádost Anglie a Skotska, aby jejich hráči mohli navléct pásky s vlčími máky. Federace později své vyjádření korigovala, že nic nezakázala, ale jen upozornila na pravidla.

Anglie a Skotsko se odradit nenechaly. Fotbalisté pásky ve vzájemném zápase navlékli. Teď za to mohou dostat trest. Oběma reprezentacím hrozí pokuta či dokonce odečet bodů v kvalifikaci o postup na mistrovství světa.

010

Všechen život tu na chvíli vymřel, když se na pěšině, tam kde ústila ve volné prostranství u zelené tůňky, vynořili dva lidé. Šli po ní husím pochodem, a i když se teď ocitli na volném prostranství, jeden z nich zůstával vzadu. Na hlavě měli oba černé beztvaré klobouky a přes ramena měli oba přehozeny pevně svinuté rance z houní. První z nich byl malý a pohyblivý, se snědým obličejem, nepokojnýma očima a ostrými, pevnými rysy. Všechno bylo na něm docela určité: drobné silné ruce, útlé paže, tenký a kostnatý nos. Za ním kráčel jeho opak, obr s beztvarým obličejem, s velkýma bledýma očima, s širokými rameny; a kráčel trochu neohrabaně, nohy trochu vláčel, asi tak jako medvěd vláčí tlapy.

011

Barák byl dlouhé obdélníkové stavení. Uvnitř vybílený, podlaha nenatřená. Ve třech stěnách byla malá čtvercová okénka, ve čtvrté bytelné dveře s dřevěnou závorou. U stěn stálo osm lůžek: pět pokrytých houněmi, na ostatních ležely jutové slamníky. Nad každým lůžkem byla přibita bednička od jablek. Její přední strana byla odtržena, takže ten, kdo na tom lůžku spal, měl na své věci dvoupatrovou poličku. A ty poličky nesly náklad všelijakých drobností, mýdla, pudru, holicích strojků a časopisů specializovaných na Divoký západ, které lidé z rančů rádi čtou a rádi se jim posmívají, ale v skrytu srdce jim věří.

012

Na jednom konci ohromné stáje bylo vysoko narovnáno čerstvé seno a nad tou kupou, zavěšen na kladce, visel čtyřramenný drapák na balíky sena. K druhému konci stáje se seno svažovalo jako nějaké horské úbočí a pak následovala rovinka novým senem ještě nenaplněná. U bočních stěn bylo vidět jesle a mezi příčkami prokukovaly hlavy koní.

Bylo nedělní odpoledne. Hovící koně okusovali zbylé hromádky sena, hryzali do dřeva žlabů a chřestili ohlávkovými řetízky. Škvírami ve stěnách prořezávalo se do stáje odpolední slunce a jasnými čarami lehalo na seno. Vzduch byl rozbzučen mouchami, líným odpoledním bzukotem.

013

Stařec byl hubený, vyzáblý a zátylek měl zrytý hlubokými rýhami. Na lících mu vyvstaly hnědé skvrny kůže, zrohovatělé na ochranu před odrazem slunce v tropickém moři. Ty skvrny mu sahaly po stranách obličeje až dolů a ruce měl zjizvené hlubokými zářezy od toho, jak se lopotil s těžkými rybami na šňůrách. Žádná z těch jizev však nebyla čerstvá. Byly tak staré jako výmoly v poušti. Všechno na něm bylo staré, až na jeho oči. Ty měly stejnou barvu jako moře a hleděly vesele a nezkroušeně.

014

Seděli na terase a mnoho rybářů si dělalo ze starce legraci a on se nezlobil. Jiní, z těch starších, se na něho dívali a bylo jim smutno. Ale nedávali to najevo a hovořili zdvořile o proudu, o tom, v jaké hloubce za sebou dnes vláčeli šňůry, o stále pěkném počasí a o tom, co viděli. Rybáři, kteří měli toho dne štěstí, byli už zpátky, vyvrhli své úlovky a odnesli je naložené v celé délce napříč přes dvě prkna do rybárny, kde čekali na nákladní auto s ledem, aby je odvezlo na trh do Havany. Ti, kdo chytili žraloky, dopravili je do továrny na jejich zpracování naproti přes zátoku.

015
Sebrali z člunu výstroj. Stařec nesl přes rameno stěžeň a chlapec nesl bednu se stočenými tvrdými pletenci hnědých šňůr, bodec s hákem a harpunu s násadou. Bednička s návnadou ležela na zádi loďky vedle kyje, jehož se užívalo ke zdolání velkých ryb, když byly přitaženy po bok člunu. Nikdo by byl starému nic neukradl, ale bylo lépe vzít plachtu a tlusté šňůry domů, protože jim škodila rosa a protože, i když by mu místní lidé docela jistě nic neukradli, stařec měl za to, že bodec a harpuna ponechané ve člunu by jen zbytečně někoho pokoušely.

016
Když se chlapec vrátil, stařec spal na židli a slunce už zapadlo. Chlapec stáhl s postele starou vojenskou pokrývku a přehodil ji přes opěradlo židle a starci přes ramena. Byla to zvláštní ramena, dosud mocná, třebaže velice stará. I jeho šíje byla dosud statná a rýhy nebyly tak vidět, když stařec spal a hlava mu padala dopředu. Košili měl tolikrát záplatovanou, že byla jako jeho plachta a záplaty vyrudly sluncem do mnoha různých odstínů. Starcova hlava však byla velice stará, a když měl zavřené oči, byla jako tvář bez života. Noviny mu ležely na kolenou a tíha jeho paže je tam přidržovala ve večerním větříku. Byl bos.

017
Usnul ve chvilce a zdálo se mu o Africe, kde byl jako chlapec. O dlouhých zlatých plážích a o bílých pobřežích, tak bílých, až z nich bolely oči, o vysokých mysech a mohutných hnědých horách. Plavil se teď podél těch břehů každou noc a slyšel ve snách řev příboje a viděl, jak se tím příbojem blíží čluny domorodců. Cítil ve spánku dehet a koudel ve spárách paluby a cítil vůně Afriky, které z jitra přinášel větřík z pevniny.

Když ucítil větřík z pevniny, probouzel se zpravidla a oblékal se, aby šel vzbudit chlapce. Ale dnes přišla vůně větříku z pevniny velice záhy a stařec věděl, že je ještě brzy v jeho snu a pokračoval ve snění, aby viděl, jak se bílé vrcholky ostrovů vynořují z moře.

018
Stařec pomalu srkal kávu. To bylo všechno, co za celý den pozře a on věděl, že ji musí vypít. Dávno ho už teď omrzelo jíst a nikdy si s sebou nebral nic k obědu. Měl na přídi loďky láhev vody a to bylo všechno, co na celý den potřeboval.

Chlapec se mezitím vrátil se sardinkami a s dvěma návnadami zabalenými do novin a oba pak sešli po stezce k loďce. Cítili pod nohama omletý písek. Nadzdvihli člun a sešoupli jej do vody.

019
Oblaka nad pevninou se teď nakupila jako hory a pobřeží se změnilo v dlouhou zelenou čáru, za kterou se zvedaly šedomodré kopce. Voda teď měla barvu temně modrou, tak temnou, že byla skoro fialová. Když se podíval do hloubky, viděl v tmavé vodě červený plankton a zvláštní světlo, vrhané teď sluncem. Pozoroval své šňůry a hleděl, jak splývají přímo dolů, až ve vodě zmizely z dohledu a byl rád, že vidí tolik planktonu, protože to znamenalo ryby. Zvláštní světlo, kterým slunce prozařovalo vodu, znamenalo pěkné počasí, právě tak jako tvar oblak nad pevninou.

020

Nic se nestalo. Ryba prostě pomalu odplouvala a stařec ji nemohl přitáhnout ani o coul blíž k hladině. Šňůru měl pevnou, dělanou na těžké ryby, a teď si ji přehodil přes záda a táhl, až se napjala tak, že od ní odstřikovaly kapičky vody. Pak začala šňůra ve vodě pomalu syčet a on stále táhl, vzpíraje se o veslařské sedátko a zakláněje se proti směru tahu. Člun se pomalu hnul k severozápadu.

Ryba plula vytrvale a tak pomalu cestovali po klidném moři. Ostatní vnadidla byla dosud ve vodě, ale nedalo se nic dělat.

021

Pokud mohl stařec určit z pozorování hvězd, nezměnila ryba za celou tu noc směr ani rychlost. Po západu slunce se ochladilo a starci uschl na zádech, na pažích a na starých nohou studený pot. Dříve během dne vzal pytel, který přikrýval bedničku s návnadou a rozprostřel ho, aby se na slunci usušil. Po západu slunce si jej uvázal kolem krku tak, že mu splýval na záda a opatrně jej podstrkal pod šňůru, kterou měl teď přes ramena. Podložil pytlem šňůru jako polštářkem a našel si způsob, jak se opřít dopředu o příď, takže měl skoro pohodlí. Ve skutečnosti ta pozice byla jenom o něco méně nesnesitelná, jemu však připadala skoro pohodlná.

022

Pak začal litovat velikou rybu, kterou chytil na hák. Je báječná a tak podivná a kdo ví, jak je stará, myslel si. Ještě nikdy jsem neměl co dělat s tak silnou rybou, nebo s rybou, která by si vedla tak divně. Snad je příliš moudrá na to, aby skákala. Strhala by mě, kdyby začala skákat nebo se najednou divoce utrhla. Ale možná že se už mockrát chytila a ví, že musí bojovat právě takhle. Nemůže vědět, že proti ní stojí jen jeden člověk a že je starý. Ale jaký to je obr a kolik vynese na trhu, jestli má dobré maso! Zabral na vnadidlo jako samec a táhne jako samec a bojuje beze vší paniky. Rád bych věděl, jestli má v hlavě nějaký plán, nebo jestli je právě tak zoufalý, jako jsem já?

023

Snad jsem se neměl stát rybářem, pomyslel si. Ale k tomu jsem se narodil. Určitě nesmím zapomenout sníst toho tuňáka, až začne svítat.

Někdy před svítáním zabralo něco na jedno z vnadidel, které vláčel za sebou. Slyšel, jak praskl klacek a šňůra se začala rychle odvíjet přes obrubeň loďky. Vytáhl v temnotě z pochvy nůž, nechal si všechen tah ryby spočinout na levém rameni, zaklonil se dozadu a uřízl šňůru.

024

Snažil se zvýšit napětí, ale šňůra byla už od chvíle, kdy ryba zabrala, napnutá až k prasknutí a stařec cítil ten ukrutný odpor, když se zaklonil dozadu, aby táhl a věděl přitom, že už do toho nemůže dát víc síly. Nesmím za žádnou cenu škubnout, připomínal si. Každé škubnutí rozšiřuje ránu, kudy vnikl hák, a když ryba začne skákat, mohla by jej vyplivnout. Ostatně se sluncem je mi líp a jednou se aspoň do něho nemusím dívat.

Na šňůru se nachytaly žluté chaluhy, ale stařec věděl, že to jenom přispívá brzdění a měl z toho radost. Byly to žluté chaluhy z golfského proudu, které v noci tolik světélkují.

025
Pomyslel si, jak se někteří lidé bojí octnout se v malém člunu z dohledu země, a věděl, že mají pravdu v měsících, kdy přichází náhlá nepohoda. Ale teď bylo období hurikánů. Když se zrovna žádný hurikán nestrhne, jsou měsíce hurikánů nejlepší v celém roce.

Když má přijít hurikán, pozná to člověk podle známek na obloze celé dny předem - jestliže je na moři. Na souši to nepoznají, myslel si stařec, protože nevědí, po čem se mají dívat. Nad zemí se taky nejspíš mění tvar mraků. Ale teď se žádný hurikán nechystá.

026
Je to veliká ryba a já ji musím udolat, myslel si. Nesmím ji ani na chvíli nechat, aby si uvědomila svoji sílu a co by dokázala, kdyby vyrazila k útěku. Kdybych byl na jejím místě, vynaložil bych teď všechnu sílu a vrhl bych se vpřed, až by něco prasklo. Ale ryby bohudíky nemají tolik rozumu jako my, kdo je zabíjíme, i když jsou ušlechtilejší a silnější.

Stařec už viděl hodně velkých ryb. Viděl hodně takových, které vážily přes pět metráků a dvě nebo tři ryby těch rozměrů v životě ulovil, ale nikdy ne sám. A teď byl sám a z dohledu země a byl připoután k největší rybě, jakou kdy spatřil.

027
Teď, když už rybu uviděl, dovedl si ji představit, jak pluje ve vodě s purpurovými prsními ploutvemi, roztaženými jako křídla, a jak svým velkým vztyčeným ocasem krájí temnotu. Rád bych věděl, jak asi v té hloubce vidí, pomyslel si. Má obrovské oči a kůň, který má oči daleko menší, vidí ve tmě. Já jsem taky kdysi viděl ve tmě. Ne v naprosté tmě. Ale skoro tak jako kočka.

Slunce a vytrvalý pohyb prstů vyhnaly teď křeč z levičky úplně a on na ni začal přenášet více námahy z tahu a nahrbil svaly na zádech, aby pošinul bolestivý zářez provazce o kousek dál.

028
Mám v hlavě dost jasno, odpovídal si v duchu. Až moc jasno. Tak jasno, jako jsou jasné hvězdy, moje sestry. Ale přece jen musím spát. I hvězdy spí a měsíc a slunce spí. A dokonce i oceán někdy spí, v jistých dnech, kdy není žádný proud a hladina je plochá a klidná.

Nezapomeň se vyspat, přikazoval si. Přinuť se k tomu a vymysli si něco jednoduchého a jistého, co podniknout se šňůrami. Obešel bych se bez spánku, odporoval sám sobě. Ale bylo by to příliš nebezpečné. Vydal se po rukou a po kolenou zpátky na záď, dávaje při tom pozor, aby nepoplašil rybu škubnutím. Možná že sama napůl spí, napadlo ho.

029
Přivázal rybu k přídi a zádi a k veslařské lavičce uprostřed. Byla tak veliká, že mu to připadalo, jako by si k boku připoutával mnohem větší člun. Uřízl kus šňůry a přivázal rybě spodní čelist kolem jejího meče, aby se jí neotvírala huba a aby pluli hladce, jak to jen půjde. Pak vztyčil stěžeň, narovnal ráhno, záplatovaná plachta se vzedmula, člun se dal do pohybu a stařec vyplul k jihozápadu.

Nepotřeboval kompas, aby poznal, kde je jihozápad. Stačilo mu cítit vítr a pozorovat dmutí plachty. Měl v láhvi ještě dva doušky vody a jeden z poloviny vypil, jakmile snědl garnáty. Na tak veliké zatížení plul člun docela dobře a stařec jej řídil s rukojetí kormidla v podpaží.

030
Pluli dobře a stařec si máčel ruce ve slané vodě a snažil se udržet si jasnou hlavu. Díval se v jednom kuse na rybu, aby se ujistil, že je všechno v pořádku. To bylo hodinu předtím, než ho napadl první žralok.

Ten žralok se neobjevil náhodou. Vyplaval už dříve z vodních hlubin, jakmile se temné mračno krve rozptýlilo v moři. Vyřítil se nahoru tak rychle a tak naprosto bezhlavě, že prorazil hladinu modré vody a octl se na slunci. Pak padl zpátky do moře, zachytil pach a vyrazil ve sledu člunu a ryby.

031
Teď věděl, že byl dočista poražen a že se nedá už vůbec nic dělat. Zabalil si ramena do pytle a vrátil člun do kurzu. Plulo se mu teď lehce a byl prost jakýchkoliv myšlenek a pocitů. Všechno už bylo za ním. Řídil tedy člun, aby se dostal do svého domovského přístavu tak hladce a rozvážně, jak to jen půjde. V noci napadli žraloci ohlodaný trup, jako kdyby někdo sbíral drobečky se stolu. Stařec si jich ani nevšiml a nevšímal si vůbec ničeho než kormidlování. Uvědomoval si jen jak lehce a bystře teď loďka pluje, když nemá po boku žádnou velkou zátěž.

032
V chatrči opřel stěžeň o stěnu. Nahmatal potmě láhev s vodou a napil se. Pak si lehl na lůžko. Přetáhl si pokrývku přes ramena a potom přes záda a přes nohy a usnul s tváří dolů na novinách a s rukama nataženýma dlaněmi vzhůru.

Když chlapec ráno nahlédl do dveří, zastihl jej ve spánku. Vítr dul tak silně, že čluny dnes nevypluly a chlapec dlouho spal a pak přišel ke starcově chatrči, jako přicházel každé ráno. Viděl, že stařec oddychuje a pak spatřil starcovi ruce a rozplakal se. Vyšel velice tiše ven, aby přinesl trochu kávy a celou cestu po silnici plakal.

033
Dálnice jsou zařízení, které umožňují jistým lidem řítit se z bodu A do bodu B značnou rychlostí, zatímco jiní lidé se značnou rychlostí řítí z bodu B do bodu A. Lidé, co bydlí v bodě C, který leží přesně uprostřed, se občas musí divit, co je na bodě A tak úžasného, že se taková spousta lidí z bodu B jen třese na to, aby se tam dostali. A co je tak zajímavého na bodě B, že taková spousta lidí z bodu A stojí o to se tam dostat. A často si přejí, aby se lidi už jednou ksakru rozhodli, kde vlastně chtějí být.

034
„Opakuji. Mluví k vám váš kapitán, tak přestaňte dělat, co právě děláte a dávejte pozor. Za prvé, přístroje ukazují, že máme na palubě dva stopaře. Tak nazdar, ať jste, kdo jste. Rád bych, aby bylo jasno. Nemám z vás ani trochu radost. Dalo mi hodně práce, než jsem se dostal tam, kde jsem teď a nestal jsem se kapitánem stavební lodi, jen abych dělal taxíka. Poslal jsem hlídku, aby po vás pátrala, a až vás najdou, vyrazím vás z lodi. Když budete mít moc velkou kliku, tak vám možná předtím přečtu pár svých básní.“

035

Vzpomínky na Zemi se mu míhaly hlavou. Bylo mu z toho špatně. Ať se snažil sebevíc, nedokázal si představit, že celá Země je pryč. Bylo to na něho trochu moc. Pokoušel se v sobě vyvolat nějaké pocity. Myslel na rodiče a na sestru. Už nejsou. Žádná reakce. Myslel na všechny blízké lidi. Žádná reakce. Pak si vzpomněl na jakéhosi neznámého člověka, který před dvěma dny stál před ním ve frontě v samoobsluze a náhle ho píchlo u srdce. Samoobsluha je pryč, se vším co v ní bylo. Celá Anglie existuje jen v jeho vzpomínkách. Vzpomínky vězely spolu s ním v téhle smradlavé zatuchlé kosmické lodi. Převalila se přes něj vlna klaustrofobie.

C Skript na kombinaci audionahrávek

```
import os
import subprocess
from os.path import basename
from random import randint

noiseDir = r"C:\Path\To\NoiseFolder"
signalDir = r"C:\Path\To\SignalFolder"
outputDir = r"C:\Path\To\OutputFolder"
tmpFile = r"C:\Path\To\OutputFolder\C.wav"

pathsToNoise = [os.path.join(noiseDir,fn) for fn in next(os.walk(noiseDir))[2]]
pathsToSignal = [os.path.join(signalDir,fn) for fn in next(os.walk(signalDir))[2]]

print(pathsToNoise)
print(len(pathsToNoise))
print(pathsToSignal)
print(len(pathsToSignal))
nofiles = len(pathsToSignal) * len(pathsToNoise)
progress = nofiles + 1

for noisePath in pathsToNoise:
  for signalPath in pathsToSignal:
    progress -=1
    subprocess.call(["sox", noisePath, tmpFile, "trim", str(randint(0,294))], shell=True)
    try:
      signalLength = str(float(subprocess.check_output(["sox", "--i", "-D", signalPath], shell=True)))
      outFileName = "{0}_{1}.wav".format(os.path.splitext(basename(signalPath))[0],
                                         os.path.splitext(basename(noisePath))[0])
      outFilePath = os.path.join(outputDir, outFileName)
      subprocess.call(["sox", "-m", tmpFile, signalPath, outFilePath, "trim", "0", signalLength],    shell=True)

    except subprocess.CalledProcessError as e:
        print(e.output)
    print(progress)

os.remove(tmpFile)

        sanitycheck = len(os.listdir(outputDir))
        if sanitycheck == nofiles:
            print( 'Everything seems to be fine, all {0} files were created.' .format(nofiles))
        else:
             print('There should be {0} files created, but there is only {1} files.'.format(nofiles, sanitycheck))
        print('DONE')
```

D Přílohy na DVD

Součástí práce je 5 DVD disků, které obsahují elektronickou verzi práce, skript v jazyce Python, který slouží ke kombinaci audionahrávek, nahrávky, průvodní dokument v angličtině, prezentující souhrnné informace o datasetu a tabulku, která obsahuje informace o jednotlivých nahrávkách. Níže je uveden podrobný obsah jednotlivých disků:

DISK 1

- Elektronická verze bakalářské práce
- Průvodní dokument v anglickém jazyce (Datasheet.pdf)
- Tabulka obsahující informace o jednotlivých nahrávkách (List_of_records.ods)
- Skript pro kombinaci audionahrávek (Script.py)
- Složka s nahrávkami hluku (Noise records)
- Složka s nahrávkami hlasu (Speech records)
- Složka obsahující textové předlohy (Texts)

DISK 2

- Zašumené nahrávky řečníků:
  F-16-03
  F-19-04
  F-22-02
  F-26-18
  F-34-13

DISK 3

- Zašumené nahrávky řečníků:
  F-44-09
  F-64-07
  F-70-16
  F-76-15

DISK 4

- Zašumené nahrávky řečníků:
  M-18-08
  M-21-01
  M-22-11
  M-34-12

DISK 5

- Zašumené nahrávky řečníků:
  M-39-10
  M-51-14
  M-61-17

DISK 6

- Zašumené nahrávky řečníků:
  M-65-06
  M-70-05