



# 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

## ANALÝZA PROSODICKÝCH A SPEKTRÁLNÍCH VLASTNOSTÍ HLASOVÉ KOMUNIKACE V ŘÍZENÍ LETOVÉHO PROVOZU

ANALYSIS OF PROSODIC AND SPECTRAL PROPERTIES OF VOICE COMMUNICATION IN  
AIR TRAFFIC CONTROL

### BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

#### AUTOR PRÁCE

AUTHOR

Jakub Simonides

#### VEDOUcí PRÁCE

SUPERVISOR

prof. Ing. Zdeněk Smékal, CSc.

BRNO 2022

# Bakalářská práce

bakalářský studijní program **Audio inženýrství**  
specializace Zvuková produkce a nahrávání  
Ústav telekomunikací

**Student:** Jakub Simonides

**ID:** 212580

**Ročník:** 3

**Akademický rok:** 2021/22

## NÁZEV TÉMATU:

### Analýza prosodických a spektrálních vlastností hlasové komunikace v řízení letového provozu

#### POKyny PRO VYPRACOVÁNÍ:

Řízení letového provozu (ATC) je služba poskytovaná ze země letadlům v řízeném vzdušném prostoru a poskytuje poradenské služby letadlům v neřízeném vzdušném prostoru. Hlavním účelem ATC je předcházet kolizím, organizovat a urychlovat tok letového provozu a poskytovat informace a další podporu pilotům.

Cílem bakalářské práce je analýza prosodických a spektrálních charakteristik hlasové komunikace mezi pilotem a řídicí věží za účelem rozpoznání směru komunikace, tedy určení, zda daný úsek řeči přichází od pilota či věže. Pro tento účel lze použít spektrálních charakteristik řeči, které již obsahují informace o přenosovém kanálu, ale i unikátních prosodických vlastností (intonace a tempo), které jsou rozdílné mezi pilotem a podzemním dispečerem ATC komunikace. V rámci navazující bakalářské práce realizujte aplikaci pro analýzu.

#### DOPORUČENÁ LITERATURA:

[1] Mary, Leena. Extraction of Prosody for Automatic Speaker, Language, Emotion and Speech Recognition. Springer, 2018.

[2] Hofbauer, Konrad, Stefan Petrik, and Horst Hering. "The ATCOSIM Corpus of Non-Prompted Clean Air Traffic Control Speech." LREC. 2008.

**Termín zadání:** 7.2.2022

**Termín odevzdání:** 31.5.2022

**Vedoucí práce:** prof. Ing. Zdeněk Smékal, CSc.

**doc. Ing. Jiří Schimmel, Ph.D.**  
předseda rady studijního programu

#### UPOZORNĚNÍ:

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

## **Abstrakt**

Práca analyzuje prozodické a spektrálne vlastnosti oboch smerov komunikácie riadenia letovej prevádzky, opisuje rozdelenie komunikácie na segmenty v smere od pilota k veži a od veže k pilotovi na základe prepisu. Následne sú jednotlivé segmenty podrobené hĺbkovej analýze ich vlastností. Analýza sa zameriava na spektrum reči v obidvoch smeroch, konkrétne na intenzitu, základnú frekvenciu F0, spektrálnu strmosť a centroid. Navyše ku tomu je vykonaná aj analýza rýchlosti reči a detekovania reči. Na základe rozdielnosti týchto parametrov, pre každý smer komunikácie, sú určené faktory, ktoré dopomáhajú k automatickému rozpoznaniu smeru komunikácie.

## **Klíčová slova**

ATC komunikácia, spektrum, reč, rozpoznávanie hlasovej aktivity, spektrálne vlastnosti, prozodické vlastnosti, tempo, intenzita, sklon, centroid, základná frekvencia

## **Abstract**

This thesis analyses the prosodic and spectral features of bi-directional air traffic control communication, describes how the communication was split to segments, according to the source, via transcription. After the splitting, the segments are deeply analyzed for their spectral and prosodic features. The analysis itself, focuses on the spectral aspects of intensity, fundamental frequency F0, slope and centroid. Additionally, tempo and voice activity detection data were measured, to support the spectral aspects as well. Because of the differences between the ATC controller's and pilots' spectral aspects, the direction of the communication can be automatically determined, with relatively high success percentage.

## **Keywords**

ATC communication, spectrum, speech, voice activity detection, spectral aspects, prosodic aspects, tempo, intensity, slope, centroid, fundamental frequency

## **Poděkování**

Rád by som poďakoval pánovi Ing. Hicham Atassimu, Ph.D., za odbornú metodologickú a pedagogickú pomoc pri vypracovaní tejto bakalárskej práce.

## **Bibliografická citace**

SIMONIDES, Jakub. *Analýza prosodických a spektrálních vlastností hlasové komunikace v řízení letového provozu* [online]. Brno, 2022 [cit. 2022-05-21]. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/141278>. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce Zdeněk Smékal.

## Prehlásenie autora o pôvodnosti diela

<b>Meno a priezvisko študenta:</b>	<i>Jakub Simonides</i>
<b>VUT ID študenta:</b>	<i>212580</i>
<b>Typ práce:</b>	<i>Bakalárska práca</i>
<b>Akademický rok:</b>	<i>2021/22</i>
<b>Téma záverečnej práce:</b>	<i>Analýza prosodických a spektrálnych vlastností hlasové komunikace v řízení letového provozu</i>

Prehlasujem, že svoju záverečnú prácu som vypracoval samostatne pod vedením vedúcej/ho záverečnej práce a s použitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú všetky citované v práci a uvedené v zozname literatúry na konci práce.

Ako autor uvedenej záverečnej práce ďalej prehlasujem, že v súvislosti s vytvorením tejto záverečnej práce som neporušil autorské práva tretích osôb, najmä som nezasiahol nedovoleným spôsobom do cudzích autorských práv osobnostných a som si plne vedomý následkov porušenia ustanovení § 11 a nasledujúcich autorského zákona č. 121/2000 Sb., vrátane možných trestnoprávných dôsledkov vyplývajúcich z ustanovení časti druhej, hlavy VI. diel 4 Trestného zákonníku č. 40/2009 Sb.

V Brne dňa: 13. decembra 2021

-----  
podpis autora

# Obsah

<b>ZOZNAM OBRÁZKOV</b> .....	<b>9</b>
<b>ZOZNAM TABULIEK</b> .....	<b>10</b>
<b>ÚVOD</b> .....	<b>11</b>
<b>1. ATC</b> .....	<b>12</b>
1.1 HISTÓRIA LETECTVA.....	12
1.2 HISTÓRIA ATC KOMUNIKÁCIE.....	12
1.3 RÁDIOVÁ KOMUNIKÁCIA.....	13
1.4 FRAZEOLÓGICKÝ SLOVNÍK VYUŽITÝ V ATC KOMUNIKÁCIÍ.....	14
1.4.1 Označovanie výšok.....	16
1.4.2 Letová hladina (Flight level).....	16
1.4.3 Minimálny pokles a rozhodujúca výška.....	16
1.4.4 Čas.....	16
1.4.5 Výškomer.....	16
1.4.6 Rýchlosť a smer vetra.....	16
1.4.7 Smer letu.....	17
1.4.8 Čísla prístávacích dráh.....	17
1.4.9 Rádiové frekvencie.....	17
1.5 ATCOSIM KORPUS.....	17
<b>2. PROZÓDIA</b> .....	<b>20</b>
2.1 REČ.....	20
2.2 ROZPOZNÁVANIE REČI.....	21
2.3 ROZPOZNÁVANIE HLASOVEJ AKTIVITY.....	24
2.3.1 Vylepšenie signálu reči.....	24
2.4 ŠPECIFICKÉ VLASTNOSTI REČI REČNÍKA.....	26
2.5 SPEKTRÁLNE DESKRIPTORY REČNÍKA.....	26
2.5.1 Spektrálny centroid.....	27
2.5.2 Spektrálna entropia.....	27
2.5.3 Spektrálny sklon.....	27
2.5.4 Spektrálna špicatosť.....	28
2.5.5 Spektrálna plochosť.....	28
2.6 PRAVDEPODOBNOŠŤ ROZPOZNANIA.....	28
2.7 TEMPO.....	28
2.8 JAZYKOVÉ ŠPECIFICKÉ ASPEKTY REČI.....	30
2.9 MEL KEPSTRUM.....	31
<b>3. VYUŽITÝ SOFTVÉR</b> .....	<b>32</b>
3.1 PYTHON.....	32
3.1.1 Parselmouth.....	32
3.1.2 NLTK.....	33
3.1.3 NumPy.....	33
3.1.4 Pandas.....	33
3.2 PRAAT.....	33
<b>4. PREVEDENIE</b> .....	<b>34</b>
4.1 FORMÁTOVANIE DÁT.....	34
4.2 DELENIE PODĽA SMERU KOMUNIKÁCIE.....	35
4.3 EXTRAKCIA INFORMÁCIÍ O TEMPE.....	35
4.4 ZÍSKANIE VÝSLEDNÝCH DÁT.....	35
4.5 GENEROVANIE GRAFOV.....	37
4.6 AUTOMATICKÉ ROZPOZNÁVANIE SMERU KOMUNIKÁCIE POMOCOU ROZHODOVACÍCH STROMOV	38
<b>5. ZÁVER</b> .....	<b>40</b>

<b>LITERATÚRA .....</b>	<b>41</b>
<b>ZOZNAM SYMBOLOV A SKRATIEK.....</b>	<b>43</b>
<b>ZOZNAM PRÍLOH .....</b>	<b>44</b>



# ZOZNAM OBRÁZKOV

Obrázok 1.1 – Číslovanie dráh, prevzaté z [2] .....	17
Obrázok 1.2 Krátka pasáž segmentu ( <i>THANKS BIZEX THREE TWENTY NINE TURN LEFT HEADING ONE CORRECTION ZERO NINER ZERO</i> ), signál časového priebehu .....	18
Obrázok 1.3 Krátka pasáž segmentu ( <i>THANKS BIZEX THREE TWENTY NINE TURN LEFT HEADING ONE CORRECTION ZERO NINER ZERO</i> ), spektrogram pre signál reči .....	19
Obrázok 2.1 Priebežná zmena kontúr F0 pre text "The MAN in the STREET", kde veľkými písmenami sú označené slová s dôrazom, prevzaté z [4].....	20
Obrázok 2.2 Euklidovská vzdialenosť LPCC vektorov, prevzaté z [4] .....	22
Obrázok 2.3 Kontúry F <sub>0</sub> pre (a) TIMIT, (b) NTIMIT a (c) CTIMIT rovnakého rečníka pre tú istú vetu, prevzaté z [4] .....	22
Obrázok 2.4 Zmena kontúry F0 pri neutrálnej emócií [4] .....	23
Obrázok 2.5 Extrahovanie vlastností reči pomocou redukcie spektrálneho šumu a frame-droppingu, prevzaté z [7] .....	25
Obrázok 2.6 Graf energie signálu reči poškodeného šumom pri znižujúcom sa SNR, prevzaté z [7] .....	25
Obrázok 2.7 Rozdiel v dynamike kontúr F0 pri rozprávaní čísel od 1 do 10 dvomi rozdielnymi ženami. [4] .....	26
Obrázok 2.8 Dĺžka správy, doba trvania a rýchlosť reči – Vysvetlivky: # - číslo prenosu AT – počet letových inštrukcií v správe W – počet slov v správe Syl – počet slabík v správe D – dĺžka správy v sekundách W/Sec – rýchlosť reči v slovách za sekundu T/Sec – rýchlosť reči v inštrukciách za sekundu R – správnosť odpovede pilota R – nesprávne pochopenie informácie pilotom P – čiastočne správne pochopenie informácie pilotom C – správne pochopenie informácie pilotom, prevzaté z [1] .....	30
Obrázok 4.1 Prepis pred formátovaním .....	35
Obrázok 4.2 Prepis po formátovaní.....	35
Obrázok 4.3 Graf časového priebehu segmentu komunikácie s detekciou reči.....	36
Obrázok 4.4 Graf zobrazujúci intenzitu hlasu v čase, pre jeden segment komunikácie v smere veža pilotovi.....	37
Obrázok 4.5 Boxplot pre intenzitu v čase pre jeden segment komunikácie v smere veža pilotovi .....	38
Obrázok 4.6 Graf podstatnosti príznakov pre správne určenie smeru komunikácie.....	39

## ZOZNAM TABULIEK

Tabuľka 1.1 - Svetelné signály, prevzaté z [2].....	12
Tabuľka 1.2 – Využitie frekvencie v ATC, prevzaté z [2].....	14
Tabuľka 1.3 Štandardná frazeológia pre čísla a písmená, prevzaté z [2].....	16
Tabuľka 1.4 Príklad prepisu reči korpusu ATC0 .....	19

# ÚVOD

Riadenie letovej prevádzky, ATC, sa zaoberá poskytovaním pokynov a informácií o letovej prevádzke tak, aby sa zabezpečila bezpečná a plynulá prevádzka leteckej dopravy. Motiváciou práce je, aby sa dalo s istotou identifikovať, akým smerom komunikácia prebieha. Kvôli veľkému širokospektrálnemu šumu v pozadí a nie práve prirodzeným slovníkom používaným v reči, v porovnaní s bežnou komunikáciou, totiž nie je vôbec jednoduché určiť smer komunikácie. Podľa publikácie *Misunderstandings in ATC Communication*, častokrát až spätný pohľad na komunikáciu vyjasní, že si ju interpretovali nesprávne. [1] V reálnom čase by to znamenalo, že korekciu zle podanej informácie, je treba opraviť hneď a predísť tým prípadným nehodám ktoré by v rámci leteckej prepravy mohli mať katastrofálne následky.

Cieľom tejto práce a zároveň aj jej dôvodom výberu je zjednodušenie identifikácie zdroja ATC komunikácie a tým aj jednoduchšie rozpoznanie zdroja komunikácie pre osobu tretej strany. Keďže sa letecká doprava stále rozširuje a je stále viac a viac populárnejšia aj v rámci vnútroštátnych letov, je treba nesmierne dbať na bezpečnosť pasažierov všetkými možnými bezpečnostnými prvkami. Jedným z nich je práve aj ATC komunikácia, keďže práve tá dohliada na bezpečnú nepretržitú prevádzku, nielen na letiskách. Je nevyhnutnou súčasťou celej modernej leteckej infraštruktúry a letecká doprava je od tejto komunikácie nesmierne závislá. Ak by komunikácia medzi riadiacou vežou a pilotom náhle zlyhala na celom svete, malo by to dopad na celú ľudskú spoločnosť, nielen kvôli osobnej, ale aj nákladnej preprave. Celý svet by sa jednoducho zastavil.

# 1. ATC

## 1.1 História letectva

Riadenie letovej prevádzky sa vzťahuje na „predchádzanie zrážok lietadiel a podporu usporiadaného toku letovej prevádzky“. Okrem toho sa vzdušný priestor, kde sa vykonáva riadenie letovej prevádzky, nazýva riadený vzdušný priestor (anglicky: Controlled airspace). Amplitúdová modulácia sa celosvetovo používa ako metóda modulácie rádiových vln pre rádiotelefony používané na riadenie letovej prevádzky. Spočiatku, pri skorých začiatkoch letectva, nebola žiadna, verbálna ani neverbálna, komunikácia medzi pilotom a letiskom. Spočiatku preto, lebo letiská ani neexistovali. Každý človek si mohol skonštruovať nejaký letecký stroj. Tieto prvé lietadlá teda často havarovali, najmä kvôli ich technickej nedokonalosti. Časom sa ale aj napriek prvotným obavám, začalo brať v úvahu užitočnosť leteckej dopravy. V roku 1911 bola letecky doručená prvá pošta. Pri zapojení Spojených Štátov Amerických do Prvej Svetovej Vojny, sa letectvo ukázalo byť užitočné aj za účelmi skúmania a prepravy zbraní. [2]

## 1.2 História ATC komunikácie

Pred skorými 30. rokmi minulého storočia bol jediný bezpečnostný prvok let cez deň a za priaznivých podmienok. Postupne sa ale bezpečnosť vyvíjala. Najskorší spôsob ATC komunikácie, bol človek, ktorý stál na viditeľnej pozícii letiska a mával farebnými vlajkami na letisku St. Louis, Missouri v 1929. Kvôli viacerým problémom, ako napríklad lety so zlou viditeľnosťou, sa postupne vlajky prestali používať a nahradili ich svetelné pistole, ktoré vedeli zasvietiť lúč svetla priamo na konkrétne lietadlo. Táto metóda sa využíva v takmer všetkých radiaciách vežiach dodnes, ak náhodou nie je možné využiť rádiovú komunikáciu. Svetelné pokyny sa líšili podľa toho, či bolo lietadlo práve vo vzdušnom priestore alebo na letisku. [2]

Farba a typ signálu	Význam z pohľadu pilota na zemi	Význam z pohľadu pilota vo vzduchu
Stála zelená	Povolenie k odletu	Povolenie k pristátiu
Blikajúca zelená	Povolenie k pojazdu	Vrátenie sa k pristávaniu
Stála červená	Stop	Pokračovať v krúžení a dať prednosť inému lietadlu
Blikajúca červená	Pojazd mimo práve požitú pristávaciu dráhu	Letisko je nebezpečné na pristátie - nepristávať
Blikajúca biela	Návrat na štartovací bod z letiska	-
Meniaca sa červená a zelená	Dbajte na mimoriadnu opatrnosť	Dbajte na mimoriadnu opatrnosť

Tabuľka 1.1 - Svetelné signály, prevzaté z [2]

Od 30. rokov sa značne tieto signály nezmenili. Keďže dopyt po leteckej preprave bol veľký, tak ani tento spôsob už nebol dostatočný na pomaly sa zhusťujúcu leteckú prepravu. Kvôli tomu bolo treba vyvinúť nejaký obojsmerný typ komunikácie, ktorým bola rádiová komunikácia. Počiatky rádiovkej ATC komunikácie sú z letiska v Clevelande, Ohio. Práve tu bola postavená veža na starom hangári a boli v nej používané 15W vysielačky s dosahom až 15 míľ pri rôznych poveternostných podmienkach. Problém však bol s ich počiatočnou veľkosťou a hmotnosťou alebo, pri malých lietadlách, s nedostatkom energie na ich napájanie. Keďže veža bola najvyšším bodom letiska, pracovníci veže mali dokonalý prehľad o celom priebehu dopravy na letisku. Postupne sa, po vzore toho letiska, transformovali aj iné rušné letiská k tomuto typu udržiavania kontroly nad plynulosťou letovej premávky. Problémom bola stále hmotnosť a veľkosť celého systému rádiovkej komunikácie v porovnaní s tým o koľko ziskov by sa letecké spoločnosti pripravili, keby tento priestor nevyužijú na miesto pre pasažierov. Rovnako ešte neboli zriadené žiadne pravidlá v rámci rádiovkej komunikácie, čo niekedy vedelo byť viac na príťaž ako na úžitok. Niektorí piloti kontaktovali ATC až pri pristávaní, pričom niektorí už keď boli 5 až 10 míľ ďaleko. V konečnom dôsledku sa ale rádiová ATC komunikácia potvrdila byť dobrou cestou pre budúcnosť leteckej prepravy.

Ako sa rádiové veže stávali bežnou súčasťou letísk, leteckí inžinieri už boli možní vyrábať nové lietadlá, ktoré boli veľkokapacitné, rýchlejšie ale hlavne, ktoré boli osadené rádiovými systémami pre umožnenie komunikácie. Umožnil sa tým let v podmienkach, kedy by sa 10 rokov dozadu nemohol uskutočniť. Piloti boli možní pristávať pri viditeľnosti menej ako 2 míle. Ako sa ale letectvo stávalo rozšírenejším, ľudia začínali mať strach pre prípadné zrážky lietadiel alebo poškodenie majetku spôsobené lietadlom a tak žiadali, o zákaz letovej dopravy ponad isté územia. Toto by znamenalo veľmi rýchly koniec komerčnej leteckej dopravy. V 1934, pre tieto opodstatnené obavy, Kongres schválil Predsedníctvo Obchodného Letectva, ktoré zodpovedalo za reguláciu dopravy národných liniek. Tieto všetky pokroky v letectve sa ukázali ako prínosné a postupne sa schválil tzv. The Air Commerce Act, ktorý mal dopomôcť rýchlo sa rozširujúcej leteckej doprave prostredníctvom letísk a rôznych navigačných prvkov. Postupne sa ku komunikácii pridávali aj ďalšie technológie pre leteckú bezpečnosť, napríklad radar. [2]

### 1.3 Rádiová komunikácia

Aj keď sa rádiová technika za ten čas rôzne menila, od počiatkov z Clevelandu zostala rádiová komunikácia hlavným spôsobom ATC komunikácie. Prvým typom komunikácie, bola jednostranná, kedy sa z veže dal kontaktovať pilot no nie naopak, keďže spočiatku bol problém s veľkosťou inštalácie aj vysielača aj prijmača. Riešením v rámci bezpečnosti ale muselo byť použitie prijmačej technológie vo veži a vysielačej v lietadle. Aby sa eliminovala interferencia, vysielače v lietadlách používali inú frekvenciu ako tie, ktoré boli vo vežiach. Tento systém sa nazýva ako systém duplex komunikácií. Systém má isté nevýhody, avšak to bol prvý systém, ktorý umožňoval obojsmernú komunikáciu pilota s ATC. Časom sa z neho vyvinul systém simplex ktorý pracuje na jednej frekvencii obojsmerne. Keďže letecká doprava nepozostáva len z komerčných letov, bolo treba prideliť frekvencie k reálnym využitiam. Väčšina rádiových komunikácií, momentálne

využívaných v letectve, vie komunikovať takmer s každým komunikačným kanálom, ktorý je k dispozícii, či už to je s lietadlom, alebo v rámci veži. [2]

<b>Frekvencie</b>	<b>Použitie</b>
108.000–117.950	Navigačné pomôcky
118.000–121.400	ATC
121.500	Núdzové hľadanie a záchrana
121.600–121.925	Služby letiska a ELT test
121.950	Letecká inštrukcia a podpora
121.975	FSS poradenstvo pre súkromné lety
122.000–122.050	Pomocná služba pre prebiehajúce lety
122.075–122.675	FSS poradenstvo pre súkromné lety
122.700–122.725	UNICOM
122.750	Medzilietadlová vo vzduchu
122.775	Letecká inštrukcia a podpora
122.800	UNICOM
122.825	Vnútroštátne VHF
122.850	Multicom
122.875	Vnútroštátne VHF
122.900	Multicom
122.925	Multicom
122.950	Unicom
122.975–123.000	Unicom
123.050–123.075	Unicom
123.100	Letecké hľadanie a pomoc
123.125–123.275	Letecké skúšobné stanice
123.300	Letecká podpora
123.325–123.475	Letecké skúšobné stanice
123.500	Letecká podpora
123.525–123.575	Letecké skúšobné stanice
123.600–123.650	FSS poradenstvo leteckého dopravcu
123.675–128.800	ATC
126.200	vojenské ATC
128.825–132.000	Vnútroštátne VHF
132.025–136.975	ATC

Tabuľka 1.2 – Využitie frekvencie v ATC, prevzaté z [2]

## 1.4 Frazeologický slovník využitý v ATC komunikácii

Medzinárodný jazykový štandard, využitý v ATC komunikácii, je Angličtina. Využitie Francúzštiny, Španielčiny alebo Ruštiny je tiež povolené, pokiaľ je to materinským jazykom pilota a zároveň aj operátora. Aby sa predišlo zle pochopeným správam, bol na to vyvinutý a prispôsobený aj samotný slovník a forma skladby informácie pri ATC komunikácii. Štandardne sa používa formát kde ako prvá časť správy je identifikácia

cieľového príjmateľa správy. To slúži na upozornenie príjmateľa na prichádzajúcu správu. Druhou informáciou v poradí je identifikácia odosielateľa správy, pre určenie toho, kto je zdroj. Tretím v poradí je obsah správy ktorý je bežne vo formáte štandardom FAA. Nakoniec je uzatvorenie komunikácie, ktoré prebieha spôsobom, že strana ktorá uzatvára komunikáciu sa identifikuje, pre jednoduchosť spätnej identifikácie. Viacero čísel aj písmen môže znieť podobne a vzhľadom aj na viaceré prízvuky pri rôznych jazykoch bolo potrebné vytvoriť určitý štandard ktorý by sa využíval v ATC komunikácii, aby sa predošlo možným chybám. Práve kvôli tomuto, bol vytvorený frazeologický slovník špeciálne pre ATC. Pre čísla sa využíva hovorenie po číslovkách. Číslo 10 by sa teda povedalo ako one zero ak sa nenachádza v skupinovej forme. Vyslovenie skupinovej formy byt bolo napríklad pre číslo 132 ako one thirty-two a samostatne one three two. [2]

Znak	Slovo	Výslovnosť
0	Zero	Zee-ro
1	One	Wun
2	Two	Too
3	Three	Tree
4	Four	Fow-er
5	Five	Fife
6	Six	Six
7	Seven	Sev-en
8	Eight	Ait
9	Nine	Nin-er
A	Alpha	Al-fah
B	Bravo	Brah-voh
C	Charlie	Char-lee
D	Delta	Del-ta
E	Echo	Eck-oh
F	Foxtrot	Foks-trot
G	Golf	Golf
H	Hotel	Hoh-tell
I	India	In-dee-ah
J	Juliett	Jewlee-ett
K	Kilo	Key-loh
L	Lima	Lee-mah
M	Mike	Mike
N	November	Nov-em-ber
O	Oscar	Oss-cah
P	Papa	Pah-pah
Q	Quebec	Key-beck
R	Romeo	Row-me-oh
S	Sierra	See-air-ah
T	Tango	Tang-go
U	Uniform	You-nee-form
V	Victor	Vik-tah

W	Whiskey	Wiss-key
X	X-ray	Ecks-ray
Y	Yankee	Yang-key
Z	Zulu	Zoo-loo

Tabuľka 1.3 Štandardná frazeológia pre čísla a písmená, prevzaté z [2]

#### 1.4.1 Označovanie výšok

Označenie výšok sa v ATC bežne udáva v nadmorskej výške (MSL), s výnimkou oblakov, kedy sa využíva nadzemská výška. Pri nadzemskej výške sa na to však dopredu upozorňuje. Pri vyslovovaní sa rozdeľujú Tisícky a stovky zo skupín. Napríklad výška 12 500 nad morom by sa vyslovila ako one two thousand five hundred. [2]

#### 1.4.2 Letová hladina (Flight level)

Letovej hladine vždy predchádzajú slová Flight level a každá číslovka by sa mala vyslovovať samostatne. Pri letovej hladine 180 by to teda bolo Flight level one eight zero. [2]

#### 1.4.3 Minimálny pokles a rozhodujúca výška

Vždy sa udávajú podľa tabuliek momentálneho stavu. Pri vyslovovaní MDA 1950 sa znovu číslovky uvádzajú samostatne ako minimum descent altitude one niner five zero. [2]

#### 1.4.4 Čas

Aby sa predišlo nedorozumeniam, všade na svete v ATC sa používa štandardný čas UTC, ktorý je rovnaký ako Greenwich mean time (GMT). Odstraňuje sa ním nutnosť vyjadrovania A.M. a P.M., kvôli využívaniu 24 hodinového formátu. Frazeologicky sa vždy vyslovuje ako len štvorčíslicie a následne slovom zulu, kvôli ujasneniu, že sa jedná o čas. Čas 7:20 AM by sa teda zmenil na 0720 ako time zero seven two zero zulu a 7:20 PM by bol 1920. one niner two zero zulu. [2]

#### 1.4.5 Výškomer

Každý pilot musí mať dáta o aktuálnom barometrickom tlaku aby výškomer mohol byť správne nakalibrovaný. Pri vyslovovaní sa vyslovujú číslovky samostatne a desatinná čiarka sa vynecháva. Predchádza tomu slovo altimeter. Pri hodnote výškomera 29.92 by sa teda vyslovilo altimeter two niner niner two. Piloti, ktorí majú výškomer v metrických jednotkách musia svojvoľne konvertovať hodnotu, alebo požiadať o konverziu jednotiek. [2]

#### 1.4.6 Rýchlosť a smer vetra

Smer vetra na letisku sa vždy určuje vo vzťahu k magnetickému severu a opisuje smer odkiaľ vietor fúka. Smer sa vždy zaokrúhľuje k najbližším desiatim stupňom. Vietor ktorý prúdi zo severu na juh bude teda 360° vietor a vietor z východu je 90° vietor. Medzinárodný štandard na určenie rýchlosti vetra sa udáva v uzloch. Frazeologicky sa vyslovuje započatím so slovom wind, nasleduje údaj smeru vetra, potom prichádza slovo "at" a za ním nasleduje rýchlosť vetra. Ak sa smer vetra mení tak prichádza slovo variable, ak prístroje na meranie vetra nefungujú, používa sa slovo estimated a ak sa mení rýchlosť tak sa pred rýchlosť pridáva slovo gust. Pri premenlivom vetre z juhovýchodu



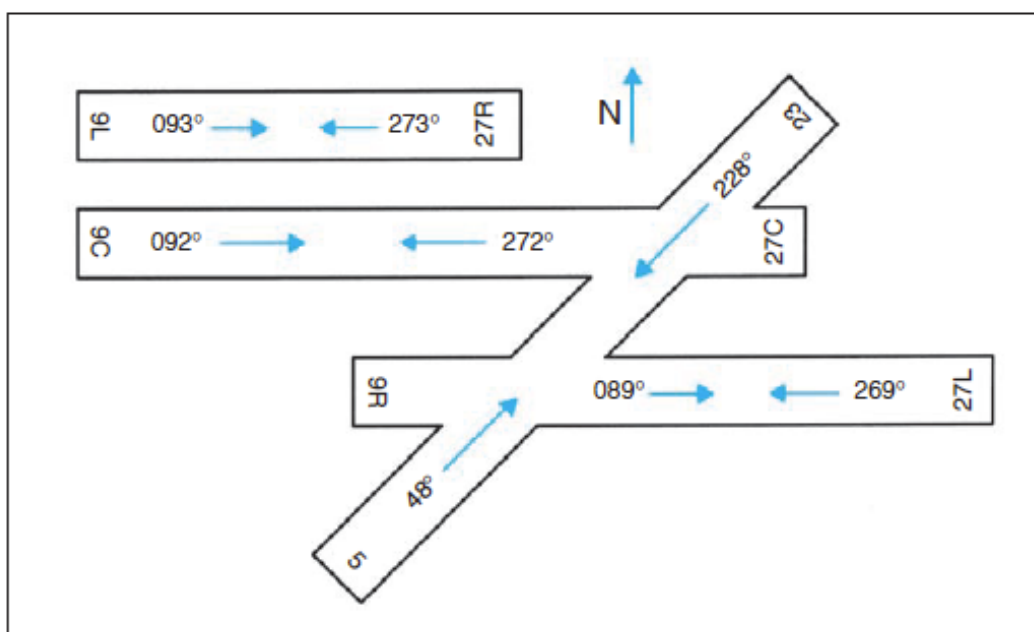
pri rýchlosti 12 uzlov miestami 35 uzlov by sa vyslovilo wind one five zero variable at one two gusts to three five. [2]

#### 1.4.7 Smer letu

Smer letu sa rovnako uvádza vo vzťahu k magnetickému severu. Ak obsahuje menej ako tri číslice, pridávajú sa na začiatok nuly, aby sa z toho stalo trojčiferné číslo. Predchádza im slovo heading. Pri smere letu 225° je výslovnosť heading two five five. [2]

#### 1.4.8 Číslo pristávacích dráh

Pristávacie dráhy sú očíslované tiež podľa magnetického severu. Pristávacia dráha smerom na sever by teda bola označená 36 z vynechaním poslednej nuly. Vyslovovaniu vždy predchádza slovo runway, nasleduje číslo po číslovkách a ak ich je viac na rovnaký smer, za číslom nasleduje right center alebo left. Ak teda je dráha smerom 138° a jej číslo je 14R, tak výslovnosť je Runway one four right. [2]



Obrázok 1.1 – Číslovanie dráh, prevzaté z [2]

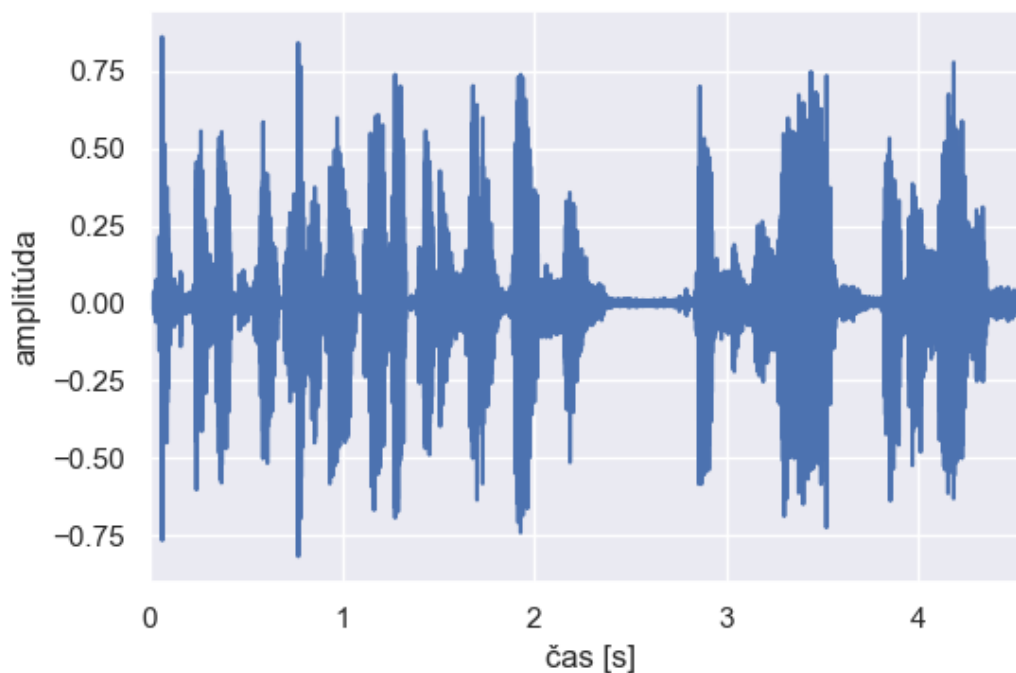
#### 1.4.9 Rádiové frekvencie

Pri frekvenciách sa číslovky znovu udávajú samostatne a frekvencie sú od seba vždy vzdialené o 25kHz. Prvé desatinné číslo sa vždy vyslovuje, aj keby bolo nulové. Tretie desatinné sa nevyslovuje nikdy, keďže je vždy 0 alebo 5 a dá sa predpokladať. Frekvencia sa explicitne vyslovuje len ak je v kHz. Pri frekvencii 131,725 MHz by sa vyslovovalo one three one point seven two. [2]

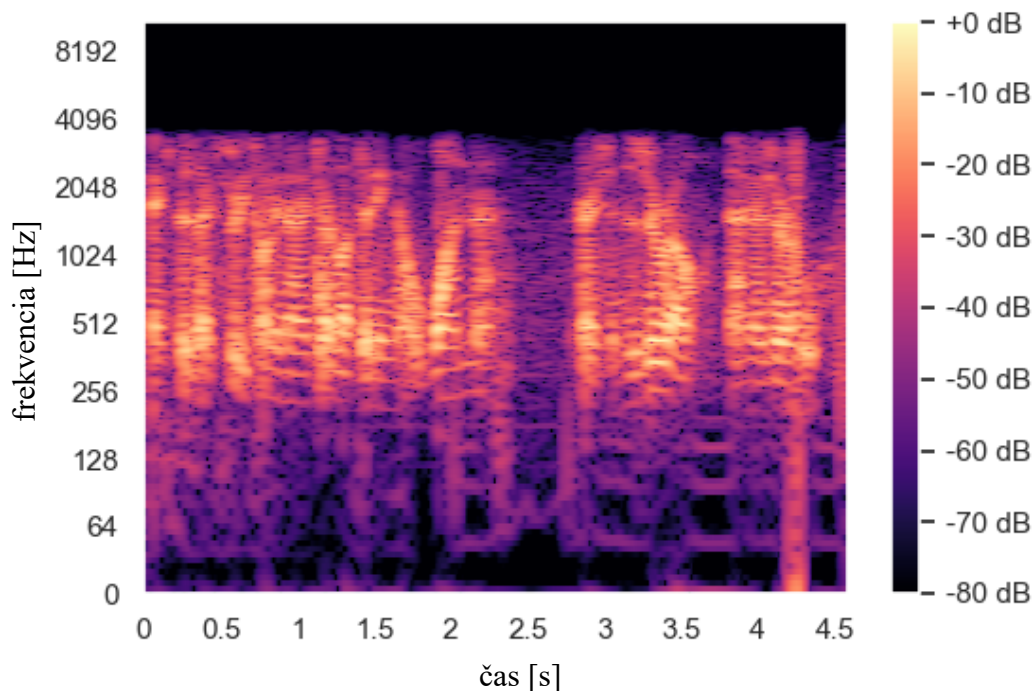
### 1.5 ATCOSIM korpus

ATCOSIM korpus je plynulá a čistá databáza rečí operátorov ATC komunikácie pre civilné lety. Skladá sa z 10 hodín reči, nahratých v prirodzených podmienkach počas simulácií v reálnom čase, zahŕňajúc všetky pravidlá, ktoré ATC komunikácia musí

dodržiavať a riadiť sa nimi. Do ešte nedávnej minulosti, bolo využitie rozpoznávania reči v rámci ATC komunikácie v ešte veľmi skorých začiatkoch, práve kvôli vysokému šumu, neznalosti rečníka a obmedzenej slovnej zásoby, no postupne je rozpoznávanie reči stále väčšou súčasťou takmer každého odvetvia kde figuruje hovorené slovo aj pri nepriaznivých podmienkach. Cieľom vytvorenia korpusu, bolo ponúknuť nejakú vzorku ATC komunikácie na ďalšiu analýzu. Aj napriek tomu, že reálna komunikácia sa archivuje, nastáva problém s jej získaním, distribúciou a faktom, že prebieha pomocou rádiového a nie širokopásmového spojenia. Simulácie prebiehali, v takmer rovnakom prostredí, na aké sú ATC operátori zvyknutí. V ďalšej miestnosti, sa nachádzajú pseudo-piloti, ovládajúci simulované lietadlo. Obsahom korpusu je však len reč operátorov, kvôli nereálnosti podmienok pilotov. Čo sa týka operátorov, tak všetci účastníci simulácie, boli aktuálne zamestnaní aktívni operátori. Skladali sa z piatich mužov a štyroch žien, s rodnými jazykmi Nemčiny, Švajčiarskej Nemčiny a Švajčiarskej Francúzštiny. Mikrofón použitý na zber záznamov, bol slúchadiel typu Sennheiser HME 45-KA. Signál z mikrofónu aj so stavom push-to-talk bol zaznamenaný na digitálnu audiokazetu so vzorkovacou frekvenciou 32 kHz a rozlíšením 12 bitov. [3]



Obrázok 1.2 Krátka pasáž segmentu (*THANKS BIZEX THREE TWENTY NINE TURN LEFT HEADING ONE CORRECTION ZERO NINER ZERO*), signál časového priebehu



Obrázok 1.3 Krátka pasáž segmentu (*THANKS BIZEX THREE TWENTY NINE TURN LEFT HEADING ONE CORRECTION ZERO NINER ZERO*), spektrogram pre signál reči

Na analýzu signálu reči bol použitý program Praat. Prepis z tejto komunikácie bol následne preformulovaný do britskej Angličtiny, bez interpunkcie a veľkých písmen. Apostrofy boli použité len pri privlastňovaniach, číslovky, písmená, navigačné prvky a signály boli prepísané podľa leteckého frazeologického slovníka, boli použité špeciálne znaky, pre bližšie špecifikácie a nejasnosti boli popísané v hranatých zátvorkách. Prepis bol zhotovený jednou osobou, pre zachovanie konzistentnosti. [3]

*((FROM NERA3788) (NUM L02F1-0001)  
 (TO F1-1)  
 (TEXT THOUSAND ONE NINETY WE (QUOTE LL) GIVE YOU THAT ON THE SPEED  
 AND WE  
 (QUOTE RE) CLEARED FOR THE APPROACH AH NERA THIRTY SEVEN EIGHTY  
 EIGHT WE  
 (QUOTE LL) HOLD SHORT OF TWO SEVEN)  
 (TIMES 1.49 6.57))*

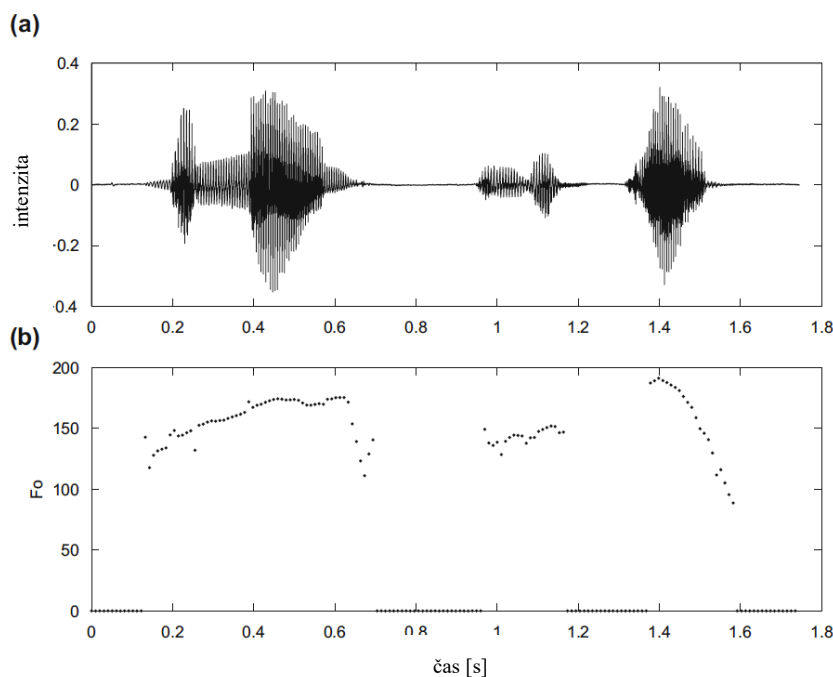
*((FROM F1-1) (NUM L02F1-0002)  
 (TO NERA3788 GAA329)  
 (TEXT THANKS BIZEX THREE TWENTY NINE TURN LEFT HEADING ONE  
 CORRECTION ZERO  
 NINER ZERO)  
 (TIMES 6.59 11.17)  
 (COMMENT "CONTROLLER TALKED TO TWO AIRCRAFT IN SAME TRANSMISSION"))*

Tabuľka 1.4 Príklad prepisu reči korpusu ATC0

## 2. PROZÓDIA

### 2.1 Reč

Reč je skladbou viacerých foném, ktoré sa následne spájajú do slabík a vytvoria slovo. Reč, ako ľudstvu najprirodzenejší komunikačný prostriedok, si teda rýchlo zabrala prvenstvo v ATC komunikácii. Zároveň je aj najefektívnejšia. Priemerne dokáže človek povedať 120 až 250 slov za minútu a tak niet pochyb, že v rámci hustej leteckej dopravy, je reč tým najlepším komunikačným kanálom. Prozódia sa vzťahuje na všetky fonetické vlastnosti, ktoré sa vyskytujú v reči. Zahŕňa intonáciu alebo výšku tónu, prízvuk, dôraz, relatívnu dĺžku tónu, rytmus, energiu, pauzy, kvalitu hlasu a intonáciu. Vzťahuje sa na tie prvky, ktoré sa môžu líšiť v závislosti od kontextu a dôrazu. Sú jazyky, kde je dôraz v určitom slove permanentne určený, no keďže v ATC komunikácii sa používa prevažne angličtina, musíme brať ohľad na to, že v angličtine má viacero rovnakých slov iný význam práve podľa toho, na ktorú slabiku dáme dôraz. Preto je aj v anglickej ATC komunikácii nesmierne dôležité, ktorým slovám dáme dôraz na aké miesto. Intonácia je dôležitým aspektom prozódie. Intonáciu vieme vyjadriť ako zmenu výšky hlasu v čase. Z toho sa dá v konečnom dôsledku odvodiť aj základná frekvencia tónu hlasu  $F_0$  v čase. Pre mužov je to zväčša medzi 80 až 200 Hz, pre ženy 180 až 400 Hz. Medzi týmito hodnotami  $F_0$  buď stúpa alebo klesá. [4] [5]



Obrázok 2.1 Priebežná zmena kontúr  $F_0$  pre text “The MAN in the STREET“, kde veľkými písmenami sú označené slová s dôrazom, prevzaté z [4]

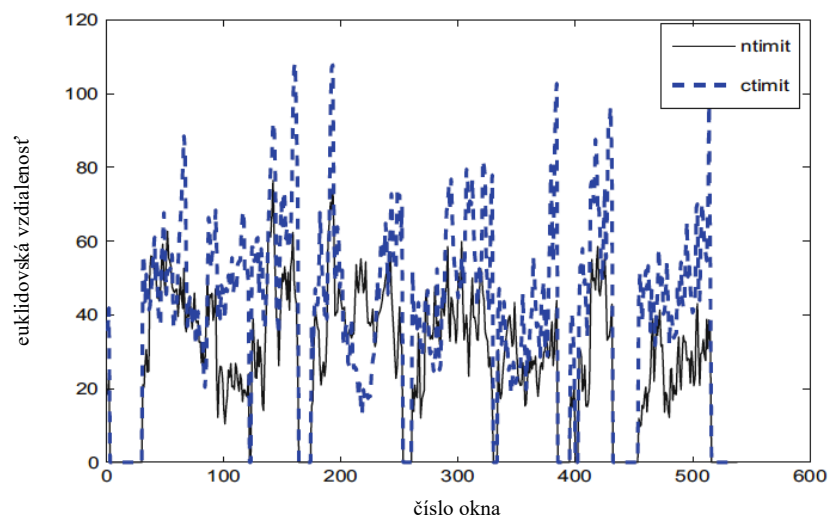
Ďalšou prozodickou vlastnosťou je rytmická zmena. Rytmickou zmenou sa rozumejú rôzne dĺžky zvukov v slove. Slabika v slove je základnou jednotkou časovania v reči. Čo sa týka analýzy všetkých týchto vlastností reči, sú jedinečné u každého jedinca a tak sa stávajú akýmsi unikátnym identifikátorom človeka. Práca sa zameriava na konkrétnu analýzu týchto vlastností pochádzajúcich z viacerých zdrojov a ich následné porovnanie, za účelom analýzy podobných a rozličných aspektov komunikácie. [4]

## 2.2 Rozpoznávanie reči

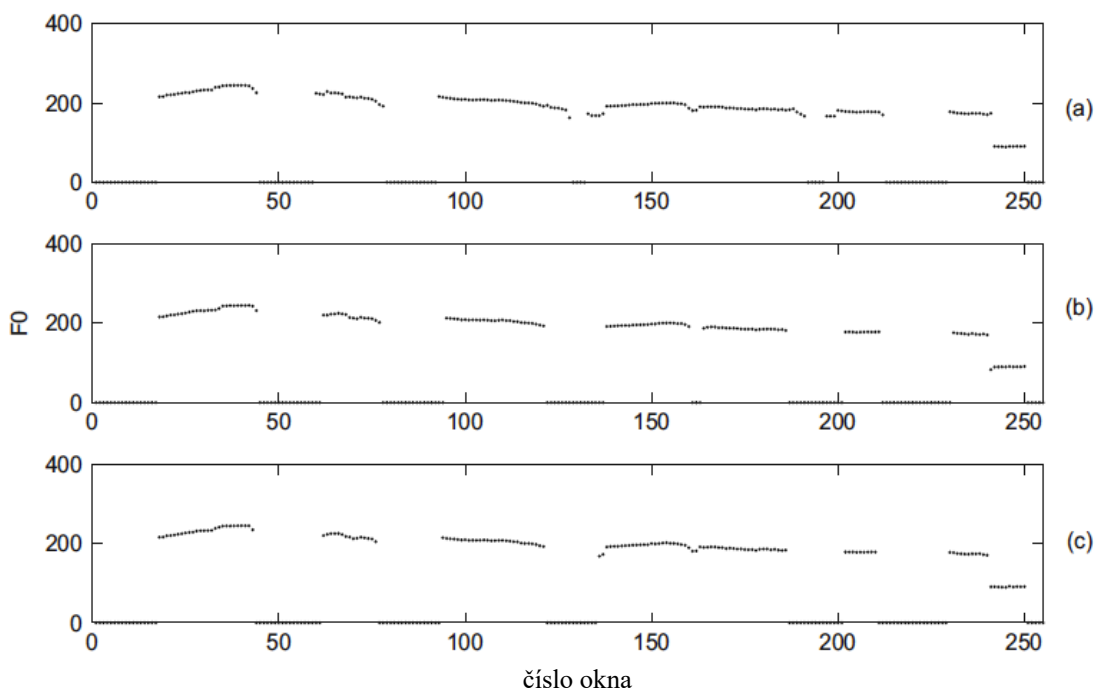
Prozódia je taktiež podstatným faktorom v rámci rozpoznávania reči. Navzájom sa poznajúce osoby, sa medzi sebou dokážu jednoducho identifikovať priradením hlasu k prejavu alebo výzoru. Rozdiely v hlase sú určené podľa rozličných rozmerov a tvarov hlasových orgánov a artikulátorov. Takisto ako faktor hrá úlohu aj rozdielny prízvuk, slovná zásoba tempo a iné subjektívne prejavy reči. Moderné programy na rozpoznávanie reči, berú v úvahu práve tieto vlastnosti rečníka a na základe nich priradujú hlas k osobe. Na základe vedeckých zistení je už teraz známe, že počítače dokážu identifikovať rečníka lepšie, ako samotní ľudia. Identifikácia prebieha postupne na základe niekoľkých krokov.

Predspracovanie – Získanie signálu reči, jeho zdigitalizovanie a správne navzorkovanie.

Extrahovanie vlastností reči – Rozpoznávanie je hrubo založené na spektrálnych/kepstrálnych vlastnostiach získaných pomocou spektrálnej analýzy. V ideálnom prípade je potreba mať k dispozícii viacero podobných zdrojov signálu, viacero rozdielnych zdrojov signálu, istý počet čo najčistejších a aj čo najšumivejších vzoriek signálu a malo by v nich byť obsiahnutých viacero emócií. Hlavným faktorom k dosiahnutiu očakávaných výsledkov je získanie hlasového záznamu v dobrých akustických podmienkach. Pri zlej akustike sa totiž v spektrálnej analýze reči vyskytuje aj chyba, ktorá vzniká zlými akustickými podmienkami, čo znamená že v analýze nebude len čistá reč ale aj napríklad ruch, ktorý vzniká akustikou miestnosti. Samotné spektrum hlasu hovoriacej osoby je teda značne rozdielne v akusticky ideálnej miestnosti, vo vonkajšom prostredí a taktiež v kokpite pilota alebo v samotnej riadiacej veži. Rovnako tento problém môžeme pozorovať pri ATC komunikácii, kde nám spektrum ovplyvňuje šum prenosového kanálu. Pri analýze ATC komunikácie teda musíme brať v úvahu všetky faktory ktoré ovplyvňujú spektrum a vo výsledkoch ich aj zohľadniť. Pôsobenie týchto faktorov si je možné všimnúť na ukážke nižšie pri vyslovení frázy “Don’t carry an oily rig like that“. Kanály pomenované TIMIT, NTIMIT a CTIMIT sú ukážkou variácie F0 pri nahrávaní blízko pri mikrofóne, cez kanál s vysokým obsahom šumu a v prostredí mobilnej siete. Pri charakteristikách kontúr F0 sa dá potvrdiť, že robustnosť prozodických vlastností je natoľko silná, že zmeny kontúr F0 sú minimálne, no stále postrehnutelné. [4] [6]

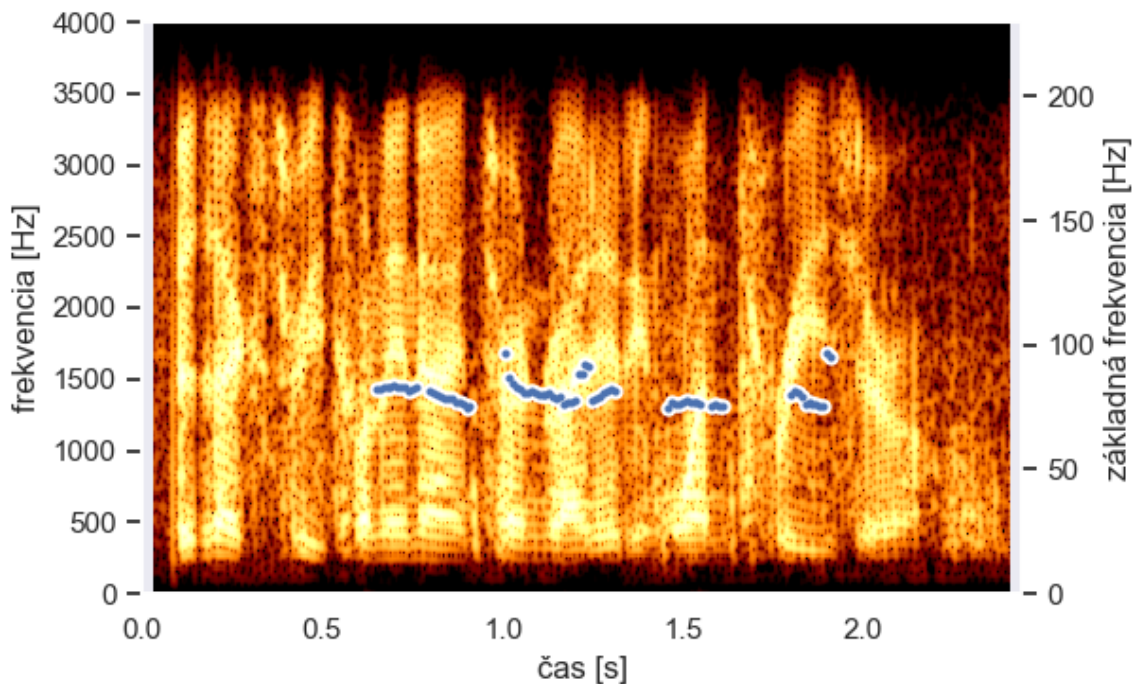


Obrázok 2.2 Euklidovská vzdialenosť LPCC vektorov, prevzaté z [4]



Obrázok 2.3 Kontúry  $F_0$  pre (a) TIMIT, (b) NTIMIT a (c) CTIMIT rovnakého rečníka pre tú istú vetu, prevzaté z [4]

Pri reči netreba zabúdať ani na emóciu. Rozdielne kontúry  $F_0$  môžeme získať z toho istého zdroja len pri emočnom rozdieli v reči. Takisto emočný faktor vplýva aj na rozdiely tempa a páuz. [4]



Obrázok 2.4 Zmena kontúry F0 pri neutrálnej emócií [4]

Pre priaznivé rozpoznávanie reči sú často preferované krátkodobé spektrálne vlastnosti (napr. MFCC), práve pre ich vysokú presnosť a dáta v reálnom čase.

Priradenie výsledkov k vzorom – Ako posledný krok je potrebné výsledky inej vzorky, priradiť ku dopredu pripraveným výsledkom z kalibračnej vzorky. Na základe porovnania, sa zistia rozdiely a podobnosti a na základe nich sa určí a rozpozná, o ktorého rečníka sa jedná. [6]

Treba samozrejme rátať aj s možnými chybami. Pri prostrediach, kde sa kalibračná vzorka získala s vysokým množstvom šumu, treba mať na vedomí, že ak vzorka na porovnanie má rovnako vysoké množstvo ale iného šumu, nemusí byť rečník identifikovaný správne. Takisto treba dať do úvahy rôzne faktory typu okolie, zdravotný stav rečníka atď. kedy bola vzorka získaná. Pri zmene ostatných vplyvov, sa porovnávací vzorka taktiež nemusí zhodovať so vzorkou kalibračnou. Preto, ak kalibračná vzorka bola získaná v ideálnom prostredí, sa môže pred porovnávaním vzoriek použiť ešte nejaký filter nežiadúcich vplyvov. [6]

V prípade, že je známe aké nežiadúce vplyvy boli prítomné pri získavaní kalibračného materiálu, je možné ich odfiltrovať ešte pred samotným porovnaním, no stále v rámci kalibračných vzoriek. Tento spôsob je však možné využiť jedine v prípade, vedomosti o týchto vplyvoch. Tento spôsob je veľmi presný, ale v praxi je voľba čo najideálnejšieho prostredia presnejšia. [6]

## 2.3 Rozpoznávanie hlasovej aktivity

Rozdeliť rozprávanie od nerozprávania, vôbec nie je tak jednoduché, ako sa na prvý pohľad môže zdať. Najväčším faktorom ovplyvňujúcim správne detekovanie reči, je šum. Za posledné roky, sa mnoho ľudí pokúsilo o vytvorenie čo najpresnejšieho systému VAD. VAD je skratka z anglického prekladu Voice Activity Detection. Najviac pokusov bolo práve pomocou vývoja rôznych algoritmov, ktoré skúmali, ako čo najefektívnejšie odstrániť nežiadúci šum a zároveň zachovať čo najlepšiu kvalitu signálu. Spôsoby prevedenia boli napríklad pomocou zistenia prahov energie, detekciou vysokých frekvencií signálu, hĺbkovej analýzy spektra, zero-crossing rate, merania periodicity, alebo kombináciou viacerých. [7]

### 2.3.1 Vylepšenie signálu reči

Cieľom spôsobu vylepšenia signálu, je odfiltrovať šumovú zložku signálu. Prvotné pokusy dosiahnutia cieľa však nebrali v úvahu, že reč aj šum nie sú statické procesy, preto prvotný spôsob spektrálneho odčítania nebol nijako zložitý ani efektívny. Bol opísaný rovnicou:

$$S(f) = |X(f)| - |N(f)|, \quad (2.1)$$

kde  $S(f)$  je spektrum čistej reči,  $N(f)$  je spektrum šumu, získané z častí signálu, kde reč nie je prítomná a je odčítané zo spektra aktuálneho okna signálu reči so šumom  $X(f)$ . Neskôr sa však tento spôsob zdokonaľoval. Upravený spôsob spektrálneho odčítania, bol vyjadrený rovnicou:

$$S(f) = \max\{|X(f)| - \alpha|N(f)|, \beta|X(f)|\}, \quad (2.2)$$

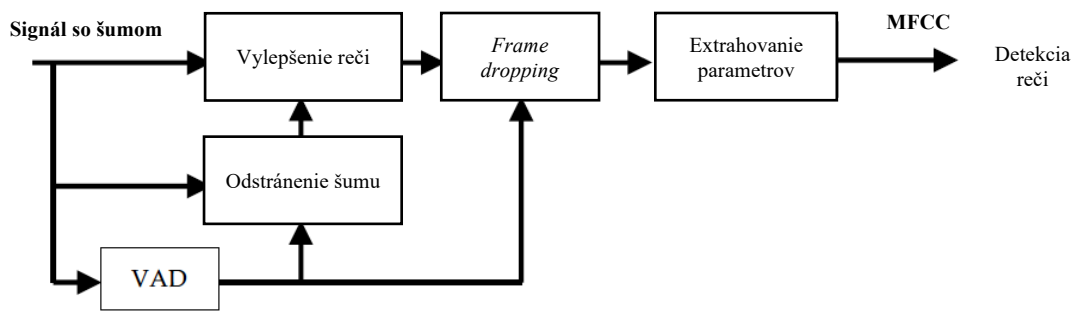
ktorá zahŕňa aj faktor nadmerného odčítania  $\alpha$  a maximálny útlm signálu  $\beta$ . Ďalšou metódou tohto spôsobu bol Wienerov filter vyjadrený ako:

$$W(f) = \frac{\Phi_{ss}(f)}{\Phi_{ss}(f) + \Phi_{nn}(f)}, \quad (2.3)$$

s potrebou vyjadrenia spektrálneho výkonu čistej reči  $\Phi_{ss}(f)$  a spektrálneho výkonu samotného šumu  $\Phi_{nn}(f)$ . [7]

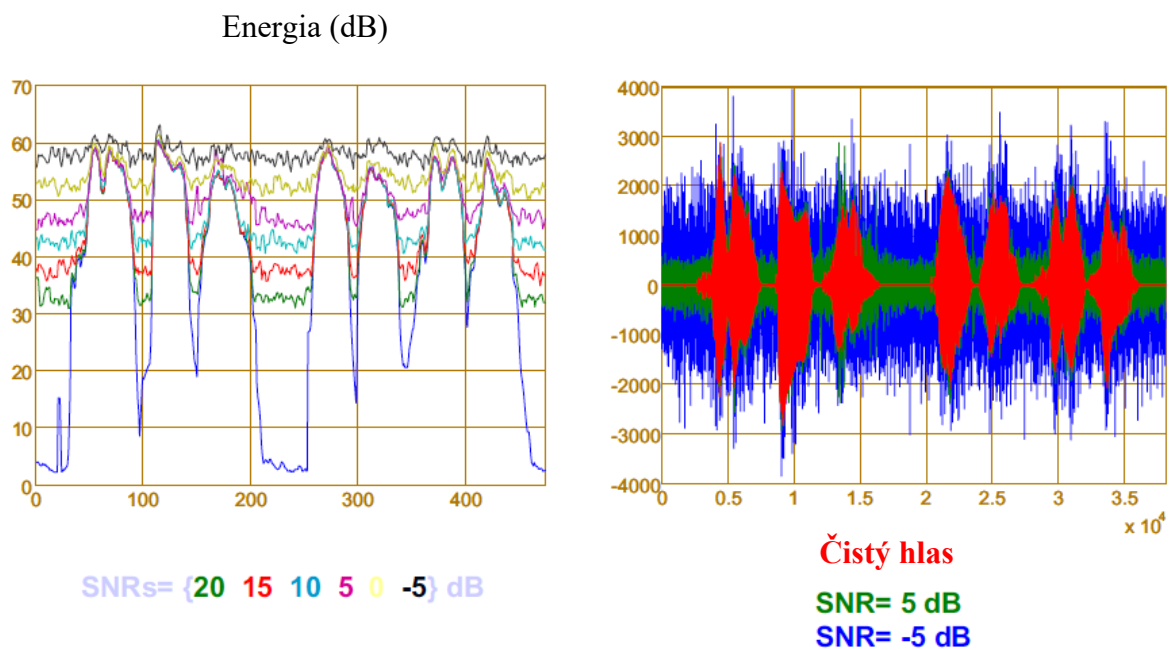
VAD však pri týchto metódach je veľmi pomocná technika docielenia adekvátneho výsledku. V pasážach, kde nie je detekovaná reč, sa získavajú informácie o šume, aby mohol byť využitý niektorý z algoritmov na vylepšenie signálu reči. Ďalšou možnosťou je využiť metódu vynechania okien signálu, kde reč nie je detekovaná (frame-dropping). Kombináciou týchto dvoch možností je možné docieľiť, že po využití spektrálneho odčítania alebo Wienerovho filtru na konkrétny signál, sa aplikujú MFFC a ich deriváty prvého a druhého stupňa postupne pre každé okno signálu a vytvoria vektor príznaku vhodný na ďalšie rozpoznanie.





Obrázok 2.5 Extrahovanie vlastností reči pomocou redukcie spektrálneho šumu a frame-droppingu, prevzaté z [7]

Problém s VAD pri signáloch s vysokým šumom je, že takýto signál degraduje vektor príznaku. Pri vysokom šume, šum kompletne prekryje signál reči, čo je problémom pre všetky iterácie systému VAD pri nízkych SNR hodnotách pri dynamickom šume. SNR je skratka pre pomer signálu k šumu.



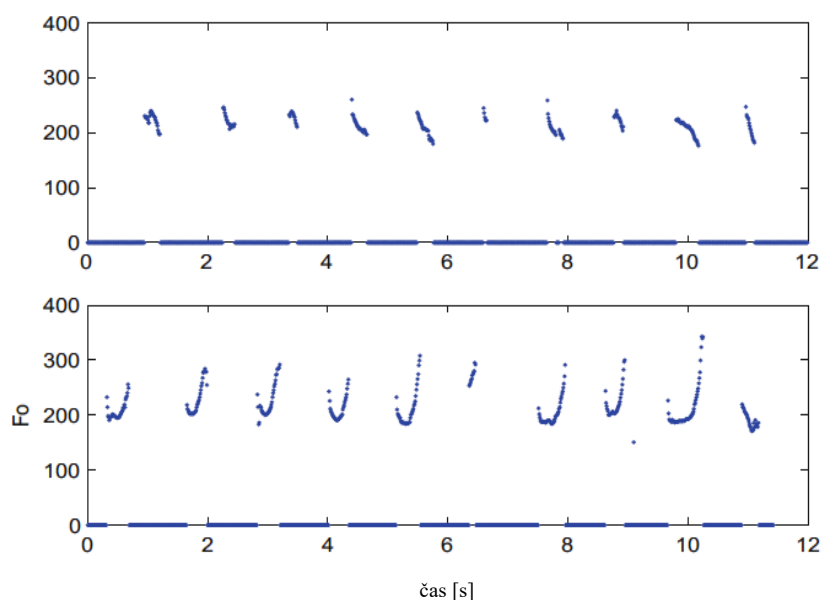
Obrázok 2.6 Graf energie signálu reči poškodeného šumom pri znižujúcom sa SNR, prevzaté z [7]

## 2.4 Špecifické vlastnosti reči rečníka

Každý rečník má svoje unikátne charakteristické črty, buď podľa rečníkových fyziologických charakteristík hlasových orgánov alebo nejakých zvykov. Rozdiely môžu byť zhrnuté nasledovne:

1. Veľkosť a tvar vokálneho traktu
2. Excitačné charakteristiky
3. Prozódia
4. Idiolekt
5. Sémantika

Prvé tri z vyššie uvedených črt vieme vyzistiť zo signálu a zvyšné dve sú prítomné v samotnom obsahu textu. [4]



Obrázok 2.7 Rozdiel v dynamike kontúr F0 pri rozprávaní čísloviek od 1 do 10 dvomi rozdielnymi ženami. [4]

## 2.5 Spektrálne deskriptory rečníka

Spektrálne deskriptory rečníka pomáhajú určiť rozdiely medzi rečníkmi na základe ich odlišností v hlasivkách. Deskriptory akými sú napríklad spektrálny centroid, sklon, entropia, špicatosť a plochosť majú isté rozdiely medzi viacerými rečníkmi. Tieto rozdiely sú značné najmä medzi pohlaviami a preto sa nimi dá jednoduchšie identifikovať napríklad pohlavie rečníka, kvôli prirodzeným rozdielom tvaru hlasiviek medzi mužmi a ženami. Ďalšie aspekty ktoré sa z daných deskriptorov dajú určiť, sú rozpoznanie emócie a emočná valencia. Na správne určenie deskriptora, potrebujeme mať kompletnú a celú nerušenú dĺžku signálu. [5]

### 2.5.1 Spektrálny centroid

Spektrálny centroid predstavuje, kde sa nachádza ohnisko hmoty rozsahu rečníka. Podobne sa vzťahuje k ohnisku rýchlostnej mapy a väčšinou sa používa v parametroch farby hlasu. Centroid sa tak stáva dôležitým indikátorom určenia farby hlasu. Farba vyzobrazuje mieru vysokej frekvencie a odhaduje sa porovnaním energie cez vopred charakterizovanú medznú frekvenciu s frekvenciou v jej blízkosti. V digitálnej hudobnej produkcii sa centroid využíva tiež na automatické určenie farby. Ďalej sa využíva na meranie vibrácie ostroti. Počíta sa ako vážený priemer frekvencií prítomných v signáli, stanovený pomocou Fourierovej transformácie s ich veľkosťou ako hmotnosťou, ako je znázornené v rovnici.

$$\mu_1 = \frac{\sum_{k=b_1}^{b_2} f_k s_k}{\sum_{k=b_1}^{b_2} s_k}, \quad (2.4)$$

Kde  $f_k$  je frekvencia v Hz,  $s_k$  je spektrálna hodnota a  $b_1$  a  $b_2$  sú hranice pásma.

### 2.5.2 Spektrálna entropia

Entropia predstavuje mnohostrannú povahu systému. Entropia je proporciou neusporiadanosti a zvykne sa využívať na kvantifikáciu vrcholnosti privlastnenia. Prvky entropie sú určené z krátkodobej Fourierovej transformácie. Dokazujú, že znelé miesta reči majú nižšiu entropiu, pretože sa tam nachádzajú jasné formanty. Entropia sa pravidelne využíva na lokalizáciu ticha a prítomnosti reči pri rozpoznávaní reči tak, že spektrum reči je viac komplexné počas fragmentov reči ako počas šumových častí. Spektrálna entropia a spektrálny **jitter** sú jedni z najlepších vlastností na rozlíšenie modálneho, dýchavičného a stlačeného hlasu. Entropia sa počíta nasledujúcou rovnicou.

$$Entropia = \frac{-\sum_{k=b_1}^{b_2} s_k \log(s_k)}{\log(b_2 - b_1)}, \quad (2.2)$$

Kde  $f_k$  je frekvencia v Hz,  $s_k$  je spektrálna hodnota a  $b_1$  a  $b_2$  sú hranice pásma. [5]

### 2.5.3 Spektrálny sklon

Amplitúda harmonických frekvencií, ktoré vznikajú v dôsledku prekrývajúcich sa vokálnych vibrácií, klesá o 12 dB na oktávu, čo znamená, že vždy, keď sa frekvencia spáruje, následná zníži svoju intenzitu o 12 dB. Spektrálny sklon určuje hodnotu práve tohoto zníženia kde

$$\text{sklon} = \frac{\sum_{k=b_1}^{b_2} (f_k - \mu_f)(s_k - \mu_s)}{\sum_{k=b_1}^{b_2} (f_k - \mu_f)^2}, \quad (2.3)$$

$f_k$  je frekvencia v Hz,  $\mu_f$  je priemerná frekvencia,  $s_k$  je spektrálna hodnota,  $\mu_s$  je priemerná spektrálna hodnota a  $b_1$  a  $b_2$  sú hranice pásma. Podľa sklonu, sa dá identifikovať dôraz rečníka. K segregácii spektrálneho sklonu dochádza v skorej mladosti. Spektrálny sklon je najvýraznejší, keď je energia v nižších formantoch oveľa dominantnejšia ako energia vo vyšších formantoch. [5]

### 2.5.4 Spektrálna špicatost'

Spektrálna špicatost' je faktor, ktorý ukazuje blízkosť usporiadania tranzientov a ich oblastí vo frekvenčnej oblasti. Je to časť vrcholovosti a je primeraným ukazovateľom impulzivnosti signálu.

$$\mu_4 = \frac{\sum_{k=b_1}^{b_2} (f_k - \mu_1)^4 s_k}{\mu_2^4 \sum_{k=b_1}^{b_2} s_k}, \quad (2.4)$$

Pri tejto rovnici:  $f_k$  je frekvencia v Hz,  $s_k$  je spektrálna hodnota  $\mu_1$  je centroid,  $\mu_2$  je rozloženie spektra a  $b_1$  a  $b_2$  sú hranice pásma. [5]

### 2.5.5 Spektrálna plochosť

Spektrálna plochosť je miera využívaná pri digitálnom spracovaní signálu na popisanie zvukového spektra. Zvyčajne sa odhaduje v decibeloch a určuje, či je zvuk ako tón, alebo je skôr rušivého charakteru. Kvantifikuje podiel geometrického priemeru spektra k aritmetickému priemeru spektra:

$$plochosť = \frac{\left(\prod_{k=b_1}^{b_2} s_k\right)^{\frac{1}{b_2-b_1}}}{\frac{1}{b_2-b_1} \sum_{k=b_1}^{b_2} s_k}, \quad (2.5)$$

kde  $f_k$  je frekvencia v Hz,  $s_k$  je spektrálna hodnota a  $b_1$  a  $b_2$  sú hranice pásma. Spektrálna plochosť je súčasťou akustických signálov, ktorá sa využila v mnohých aplikáciách spracovania signálov. Je možno ju určiť v rámci vopred definovaného podpásma, namiesto celého pásma. Toto určenie je jedným z mnohých zvukových deskriptorov používaných v štandarde MPEG-7. [5]

## 2.6 Pravdepodobnosť rozpoznania

Rozpoznanie sa dá nazvať ako priradenie najpravdepodobnejšieho rozprávača, jazyka, emócie alebo nejakej zvukovej jednotky  $C^*$  vstupnej reči z množiny existujúcich rozprávačov, jazykov, emócií ale zvukových jednotiek. Ak je určené že  $\{C_i\}$ ,  $1 \leq i \leq M$  označuje množinu existujúcich rozprávačov, jazykov, emócií ale zvukových jednotiek a  $O$  označuje pozorovania zistené zo vstupnej reči, tak rovnica rozpoznávanie v rámci pravdepodobnosti bude znieť nasledovne:

$$C^* = \arg \max P(C_i|O), \quad (2.6)$$

kde  $P(C_i|O)$  je aposteriórna pravdepodobnosť člena  $C_i$ , pre daný výskyt reči vyjadrený pre  $O$ . [4]

## 2.7 Tempo

Tempo je ďalším podstatným aspektom v rámci analýzy prozodických vlastností reči. Podľa rýchlosti reči, vieme určiť nie len náladu rečníka, ale aj podmienky v ktorých rozpráva. Tempo je neoficiálne definované ako rýchlosť reči, čo znamená, že koľko

jednotiek reči rečník vysloví za určitý čas. Oficiálna definícia však neexistuje, keďže sa na tom viacerí vedci nevedia zhodnúť. V publikáciách často nie je jasné, či sa ako rýchlosť reči nemyslí, celé pásmo, vrátane páuz. Jednotkou reči sa najčastejšie poníma jedna slabika. Vieme to vyjadriť ako rovnicu:

$$T = \frac{N_s}{T_d}, \quad (2.5)$$

kde  $N_s$  predstavuje počet slabík a  $T_d$  celkový čas rozprávania. Rýchle tempo, napríklad, môže vystihovať nahnevaného rečníka ale muselo by sa zahrnúť viacero ďalších podmienok.

Pri hneve býva zväčša  $T_d$  nižšie ako pri smútku. Priemerná dĺžka slabiky teda môže byť vyjadrená ako:

$$T_{avg} = \frac{T_{sd}}{N_s}, \quad (2.6)$$

kde  $T_{sd}$  predstavuje celkovú dobu trvania slabík a  $N_s$  počet slabík.

Každá prestávka, dlhšia ako 100ms, je považovaná ako pauza. Priemerná doba pauzy teda môže byť vyjadrená ako:

$$T_{pavg} = \frac{T_{pd}}{N_p}, \quad (2.7)$$

kde  $T_{pd}$  predstavuje celkovú dobu trvania pauzy a  $N_p$  počet páuz.

Najdlhšia doba pauzy, má tiež významný význam v rámci vyjadrenia nálady rečníka. Pri dlhších pauzách, sa jedná zväčša o opatrnosť, smútok v porovnaní s kratšími pauzami pri hneve alebo radosť.

Cruttenden [9] tvrdí, že priemerné tempo Britsky hovoriaceho človeka je okolo 6 slabík za sekundu, no pri nadšení môže byť vyššie, resp. sa môže priebežne zvyšovať. To znamená, že rýchlosť reči sa môže meniť dynamicky. [4] [10]

V ATC komunikácií, sa považuje príliš rýchle tempo reči, za príčinu chybné podaných informácií. Podľa štúdie McCandlessa v Salt Lake City termináli z roku 1991, sa preukázalo zvýšené tempo reči v komunikácií pri hustej leteckej premávke, čo spôsobilo častejšie opakovanie a aj opravovanie chybných informácií od operátorov. Naopak, podľa štúdií z [1], sa údaje o chybovosti značne líšia. Môže to byť spôsobené časom zberu dát (11 am), keďže v tomto čase letisko nie je úplne zaneprázdnené a zároveň, letisko Portland, nie je úplne rušným letiskom. [1]

#	AT	W	Syl	D	W/Sec	T/Sec	R
131	6	34	54	8.5	4.0	.71	P
35	6	38	61	8.7	4.37	.69	P
148	6	36	58	8.9	4.05	.67	E
207	5	22	35	5.55	3.96	.90	P
230	5	26	41	6.63	3.92	.75	P
33	5	30	48	6.8	4.41	.74	E
121	5	31	52	7.2	4.31	.69	P
58	5	32	54	8.1	3.95	.62	P
44	5	33	55	8.15	4.05	.61	C
96	5	35	58	8.6	4.07	.58	P
249	4	12	24	3.8	3.16	1.05	P
68	4	20	33	4.15	4.82	.96	P
51	4	16	29	4.55	3.52	.88	C
214	4	22	35	5.05	4.36	.79	P
7	4	22	37	5.6	3.93	.71	C
243	4	26	43	6.15	4.23	.65	C
264	4	30	47	6.4	4.69	.63	P
195	4	31	46	6.5	4.77	.62	P
19	4	24	37	6.5	3.69	.62	P
177	4	28	45	8.15	3.44	.49	C -> NR
256	2	14	22	4.0	3.5	.69	E
21	2	13	20	2.9	4.48	.50	(E)

Obrázok 2.8 Dĺžka správy, doba trvania a rýchlosť reči – Vysvetlivky: # - číslo prenosu AT – počet letových inštrukcií v správe W – počet slov v správe Syl – počet slabík v správe D – dĺžka správy v sekundách W/Sec – rýchlosť reči v slovách za sekundu T/Sec – rýchlosť reči v inštrukciách za sekundu R – správnosť odpovede pilota R – nesprávne pochopenie informácie pilotom P – čiastočne správne pochopenie informácie pilotom C – správne pochopenie informácie pilotom, prevzaté z [1]

Z výsledkov sa dá usúdiť, že vyššia rýchlosť reči, nutne nemusí znamenať vyššiu chybovosť. Faktorom chybovosti však môže byť rýchle podanie viacerých informácií, v porovnaní s rýchlym podaním menšieho počtu informácií, ešte k tomu vo vôbec nie optimálnych podmienkach. Táto štatistika tak nedokazuje, že pri rýchlejšej reči, je pravidlom, že chybovosť bude väčšia. [1]

## 2.8 Jazykovo špecifické aspekty reči

Je viacero rečových aspektov, ktoré sú rozdielne pri rozličných jazykoch ako napríklad akustická fonetika, prozódia, fonotaktika, slovník používaných slov a slovná skladba. Keďže ATC komunikácia pri lokálnych letoch neprebíha len v anglickom jazyku, tieto aspekty treba brať v úvahu aj pri analýze v rozličných jazykoch kde sa napríklad intonácia, rytmus alebo dôraz pri vyslovovaní rôznych slov mení v závislosti od hovoreného jazyka. [4]

## 2.9 Mel Kepstrum

Mel frekvencia, bola zavedená do fonetiky pre presnejší opis tónov podľa subjektívnych vnemov človeka. Po viacerých sluchových vedeckých experimentoch bolo zistené, že tóny vnímané ľudským uchom majú nelineárny vzťah s frekvenciou určenou parametrami sínusoidy alebo Fourierovou transformáciou, kvôli fyziológii slimáka a bazilárnej membrány. Experimenty, popisujúce fyziológiu ucha dokazujú, že frekvencia Mel má lineárny vzťah s frekvenciou pod 1000 Hz a logaritmický keď je frekvencia nad 1000 Hz. MFCC, sú kepstrálne koeficienty, odrážajúce ľudské sluchové vnímanie tónov. Na výpočet MFCC daného signálu sa spektrum na lineárnej frekvencii najprv transformuje na spektrum frekvencie Mel (spektrum Mel) a potom sa vypočítajú koeficienty kepstra na spektre Mel. [11]

Vzťah medzi frekvenciou Mel a skutočnou frekvenciou možno vyjadriť vzorcom:

$$Mel(f) = 2595 \ln(1 + f/700), \quad (2.5)$$

kde,  $f$  je frekvencia. Frekvenciu hlasu, je možné rozdeliť do série trojuholníkových filtračných sekvencií, ktoré sa nazývajú Mel filtre. Existujú dve výhody používania trojuholníkového pásmového filtra. Spektrum môže byť vyhladené, aby sa eliminoval harmonický efekt a zobrazili sa charakteristiky formantu hlasu (preto tón alebo výška hlasu nie je prezentovaná v MFCC). Ďalšou výhodou je, že je možné získať vyššiu koncentráciu charakteristických parametrov. [12]

## 3. VYUŽITÝ SOFTVÉR

### 3.1 Python

Python je objektovo orientovaný programovací jazyk, značne ovplyvnený jazykmi Java a C++, ktorého počiatky sú v neskorých osemdesiatych rokoch minulého storočia. Python bol vyvinutý ako skriptovací jazyk a dodnes sa nachádza v jeho neustálom vývoji. V momentálnom štádiu jeho rozvoja, je neuveriteľne flexibilným jazykom, pre dátovú analýzu a má nespočetne mnoho využití v inžinierskej praxi. Má mimoriadne jednoduchú základnú syntax a je jedným z najprehľadnejších súčasne využívaných programovacích jazykov. Programy využívajúce jazyk Python, nie sú kompilované do strojového kódu, ale sú spúšťané pomocou interpreteru, ktorý zaručuje rýchly test a následný debugging programu. Nutnosťou je preto mať v počítači nainštalovaný Python interpreter.

Výhodami programovacieho jazyku Python sú napríklad, že jeho distribúcia je open-source, čo znamená, že jeho využívanie je zadarmo a zároveň je jeho zdrojový kód verejne prístupný. Ďalej, je to jazyk, ktorý je dostupný pre všetky väčšie platformy, takže programy vytvorené pomocou Pythonu, sa dajú spúšťať na takmer všetkých operačných systémoch. Zároveň je veľmi jednoduché nainštalovať nielen samotný interpreter, ale aj dodatky a knižnice, ktoré bude jazyk neskôr využívať. Využíva taktiež základné prvky programovacích jazykov ako napríklad triedy, metódy, funkcie a dedičnosť, podobne ako iné objektovo orientované programovacie jazyky. Najviac podobností, však má s jazykom, ktorý má rovnako vedecké zameranie, MATLAB, kde tiež vidno podobnosť v tom, že je veľmi prehľadný, s potrebou odsadzovať jednotlivé riadky kódu a takisto rozlišuje veľkosti písmen. Python, je však oproti Matlabu viac výkonný, kompletný modul s mnoho funkciami, môže byť zahrnutý v jedno súbore, oproti mnoho .m súborom v Matlabe, viac verejne dostupných knižníc a má napríklad aj jednoduchšiu prácu s funkciami ako argumentami. Pre tieto dôvody, bol pre túto konkrétnu prácu preferovaný Python namiesto Matlabu. Nevýhodou však je, že viacero funkcií sa nachádza v Matlabe už natívne, avšak do Pythonu ich treba importovať pomocou knižníc. V prípade analýzy signálu reči a následnému spracovaniu dát, boli využité knižnice parselmouth, nltk, scipy, numpy, noisereduce, matplotlib, librosa, pyACA, pandas, re, webrtcvad a pydub. [13] [14]

#### 3.1.1 Parselmouth

Parselmouth je Python knižnica, využívajúca software Praat. Jej výhodou oproti ostatným Praat knižniciam je, že namiesto toho, aby znovu implementovali časti Praatu v Pythone, má Parselmouth priamy prístup ku natívnemu C/C++ kódu v Praate. Tým sa dosiahnu výsledky totožné, ako by sa dosiahli cez samotný program Praat, avšak v rámci Python skriptu. Práve toto umožnilo, pracovať s množstvom dát, získaných z programu Praat a zahrnúť to do komplexného Python skriptu, spolu s kombináciou iných knižníc, ktoré boli využívané, pre dosiahnutie a následné porovnanie výsledných hodnôt. [15]



### 3.1.2 NLTK

Ďalšou využitou knižnicou, je knižnica NLTK. NLTK, znamená Natural Language Toolkit (súbor nástrojov prirodzeného jazyka). Jej využitie je práve na prácu s dátami ľudského jazyka, konkrétne na extrahovanie údajov dát tempa reči na základe prepisu. Špecificky na to bola využitá časť NLTK knižnice, zvaná corpus, ktorá číta prepisy komunikácie a z nich vie určiť počet slabík. [16]

### 3.1.3 NumPy

Numpy, odvodené z Numerical Python, je knižnica, ktorá je základom práce s numerickými dátami v Pythone. Jej využitie je takmer vo všetkých vedeckých a inžinierskych programoch, využívajúcich jazyk Python a jej API je využívané aj v knižniciach Pandas, SciPy, pre import súborov s príponou .wav a Matplotlib, pre vyzobrazovanie grafov, zo zistených hodnôt, ktoré sú taktiež využité na získanie výsledkov analýzy. NumPy pracuje s viacdimenziálnymi poľami hodnôt a taktiež s maticami, ktoré sú nevyhnutnosťou pre porovnanie výsledkov viacerých aspektov reči. [17]

### 3.1.4 Pandas

Knižnica Pandas, je určená na vytváranie dátových štruktúr a následnej práce s nimi. Jej cieľom, je byť najvýkonnejšou open-source knižnicou svojho druhu, pre prácu v ľubovoľnom jazyku. Jej dve primárne štruktúry dát sú, Series (1-dimenziálna) a DataFrame (2-dimenziálna). Práve DataFrame sa využíva na tvorbu matic, ktoré slúžia na dátové porovnania, spolu s možnosťou pomenovania riadkov a stĺpcov. [18]

## 3.2 Praat

Praat je program, ktorý je určený na analýzu hovoreného slova a dokáže zo vzorky audio súboru analyzovať reč a následne vygeneruje konkrétne dáta, ktoré sa viažu s danou vzorkou reči. Medzi jeho pokročilé funkciami patrí však aj napríklad syntéza reči. [19]

## 4. PREVEDENIE

Cieľom bakalárskej práce, bol zber dát na samotnú analýzu prozodických a spektrálnych vlastností rečníkov z nahrávok zvukových súborov nepretržitej ATC komunikácie medzi pilotmi a vežami z viacerých letísk, ich následné porovnanie a zhotovenie štatistík, na čo najpresnejšie automatické rozpoznávanie smeru ATC komunikácie. Zvukové súbory, spolu s prepismi komunikácie, boli poskytnuté vedúcim práce, z korpusu ATC0, pozostávajúceho zo zaznamenatej reči komunikácie na použitie pri podpore výskumných a vývojových aktivít v oblasti robustného rozpoznávania reči v doménach podobných riadeniu leteckej dopravy (niekoľko reproduktorov, hlučné kanály, relatívne malá slovná zásoba, obmedzený jazyk atď.). Zvukové súbory boli vo formáte .sph, čo je audio súbor, vytvorený vo formáte NIST SPHERE. Daný formát sa využíva práve v hlasovom rozpoznávaní a vlny v tomto formáte, majú bežne 16-bitovú pulzne kódovú moduláciu a vzorkovaciu frekvenciu o výške 16 kHz. Formát .sph však python neprečíta a tak bolo nutné danú nahrávku prekonvertovať do formátu, s ktorým python problém mať nebude. To bolo docielené za použitia programu SoX, ktorý audio súbor prekonvertoval na Waveform Audio File Format. Pre každý zvukový súbor sú poskytnuté úplné prepisy, vrátane časov začiatku a konca každého prenosu. Každý let je identifikovaný svojím číslom letu. ATC0 pozostáva z troch subkorpusov, jeden pre každé letisko, na ktorom boli zhromaždené prenosy - Dallas Fort Worth (DFW), Logan International (BOS) a Washington National (DCA). Kompletná sada obsahuje približne 70 hodín vysielania radiaciach a pilotných vysielaní prostredníctvom antén a rádiových prijímačov, ktoré boli umiestnené v blízkosti príslušných letísk. Korpus ATC0 zhromaždila spoločnosť Texas Instruments na základe zmluvy s agentúrou DARPA. Vyrobil ho Národný inštitút pre štandardy a technológie na distribúciu konzorciom Linguistic Data Consortium [20]. Komunikáciu bolo najskôr potrebné rozdeliť na segmenty a rozdeliť ich podľa zdroja komunikácie. Následne sa každý segment analyzoval, a z analýzy boli extrahované informácie o intenzite, spektre a základnej frekvencii reči (hodnoty pre priemer, maximum, minimum a štandardnú odchýlku). Z údajov o spektre, boli taktiež navyše extrahované údaje o spektrálnom sklone a centroide. Poslednými dátami potrebnými pre analýzu, boli údaje o tempe a najväčšej dĺžke pauzy resp. kontinuálnej reči a ich pomer.

### 4.1 Formátovanie dát

Analýze samotných dát predchádzalo rozčlenenie nepretržitej, cca. dvojhodinovej komunikácie na segmenty. Jeden segment komunikácie sa rozumie ako jeden súbor s príponou .wav kde je extrahovaná práve jedna pasáž komunikačného kanála. Rozdelenie bolo dosiahnuté pomocou skriptu v programovacom jazyku python, kde v prvom kroku, je modifikovaný textový dokument s prepisom komunikácie, za účelom dosiahnutia formátu, ktorý obsahuje jedine údaj smeru komunikácie a počiatkový a konečný čas segmentu s vynechaním každej inej informácie, ktorá sa v pôvodnom počiatkovom prepise nachádzala. Formátovanie dát, prebiehalo pomocou čítania textového dokumentu riadok po riadku. Ak sa v riadku nachádzala informácia o smere komunikácie pomocou využitia regexu (regulárne výrazy), skript našiel nasledujúci riadok, popisujúci časový údaj a vymazal všetko nepotrebné medzi tým. Na záver pre jednoduchšiu čitateľnosť, bolo prepísané kompletne pomenovanie lietadla a ATC pracoviska na PA prípadne AP.

```
((FROM NERA3788) (NUM L02F1-0001)
(TO F1-1)
(TEXT THOUSAND ONE NINETY WE (QUOTE LL) GIVE YOU THAT ON THE SPEED AND WE
(QUOTE RE) CLEARED FOR THE APPROACH AH NERA THIRTY SEVEN EIGHTY EIGHT WE
(QUOTE LL) HOLD SHORT OF TWO SEVEN)
(TIMES 1.49 6.57))
```

Obrázok 4.1 Prepis pred formátovaním

```
PA1
1.49 6.57
```

Obrázok 4.2 Prepis po formátovaní

## 4.2 Delenie podľa smeru komunikácie

V druhom kroku, za použitia knižnice pydub, sa z modifikovaného textového dokumentu načítajú dáta a pomocou cyklu “for“, program počiatočne vyhodnocuje, aký je údaj o smere komunikácie. Ak riadok začína na PA (pilot veži) resp. AP (veža pilotovi), tak čítanie prejde do ďalšieho riadku. V ďalšom riadku sa časový údaj od – do, rozdelí na dva stringy, ktoré sa uložia do premenných, nazvaných var1 a var2. Stringy je potrebné konvertovať na dátový typ float, keďže sa pracuje s desatinnými číslami a následne ich vynásobiť tisícmi. Násobenie bolo nutné vykonať, kvôli prevodu zo sekúnd na milisekundy. Audio súbor sa nakoniec segmentuje podľa časových údajov a po segmentácii, je uložený do príslušného priečinku, s názvom podľa smeru prebiehajúcej komunikácie, rovnako do súboru s príponou .wav. Súbor sú uložené číselne vzostupne, pre obe strany komunikácie.

## 4.3 Extrakcia informácií o tempe

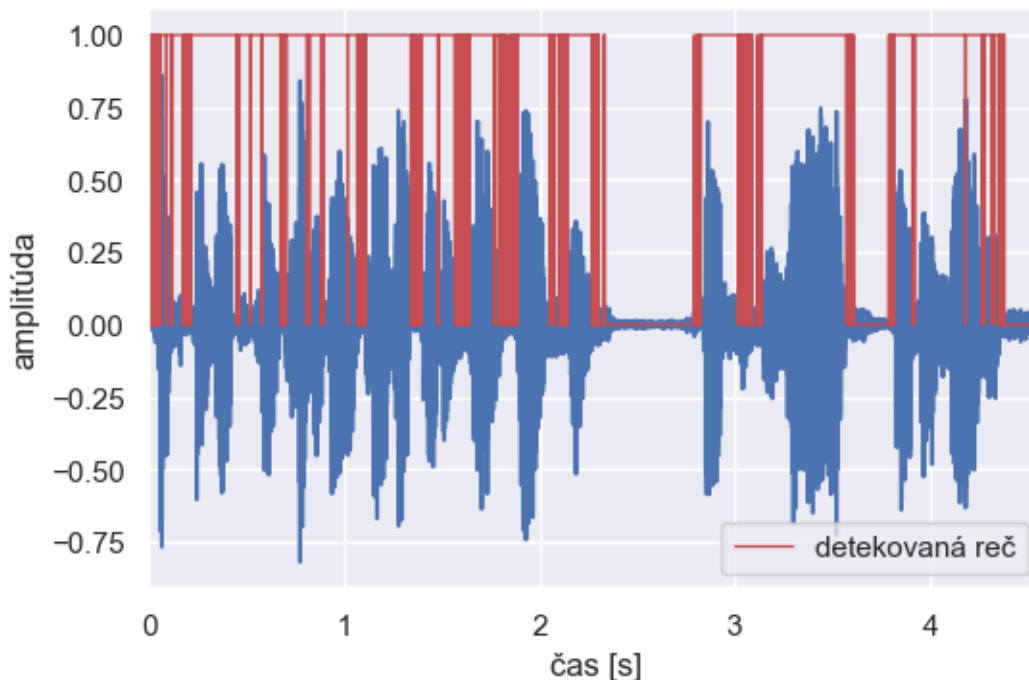
Tempo je podľa nezáväznej dohody, definované ako počet slabík za určitý čas. Údaje tempa boli zistené na základe cyklu, kde sa znova čítal každý riadok pôvodného prepisu komunikácie postupne, vyhodnocovali sa jedine údaje o smere a konkrétny prepis informácie a postupne sa pridávali dáta do premenných. Potom, za použitia cyklu for, ktorý postupne čítal každý segment konverzácie v určitom smere a následne, za použitia knižnice NLTK, vyhodnotil počet slabík. Ako posledný krok, program vydělil počet slabík dĺžkou nahrávky. Z toho bol extrahovaný údaj o tempe reči, ktorému bol pridelený stĺpec hodnôt v matici príznaku.

## 4.4 Získanie výsledných dát

Pre získanie dát, bolo využitých viacero knižníc a programov. Po rozdelení segmentov do patričných priečinkov, bolo získaných vyše 1000 segmentov z jednej nepretržitej komunikácie. 1000 segmentov, je dostatočná vzorka na zistenie bližších detailov. Tie boli extrahované znovu pomocou jazyku python v kombinácii s aplikáciou Praat. Pre analýzu ATC komunikácie sú obzvlášť potrebné dáta o intenzite a základnej frekvencii hlasu  $F_0$ , ktoré sa aj pomocou pythonu s Praatom podarilo extrahovať. Analýza samotných vzoriek bola obsiahnutá pomocou Praat knižnice pre python, parselmouth, tak, že pomocou cyklu for, boli postupne analyzované všetky segmenty komunikácie v oboch smeroch. Každý segment bol individuálne postupne načítaný a preskúmaný.

Po zistení týchto parametrov, sa pomocou knižnice pyACA, akumulovali údaje a následne z nich boli vytvorené grafy spektra, centroidu a skolnu.

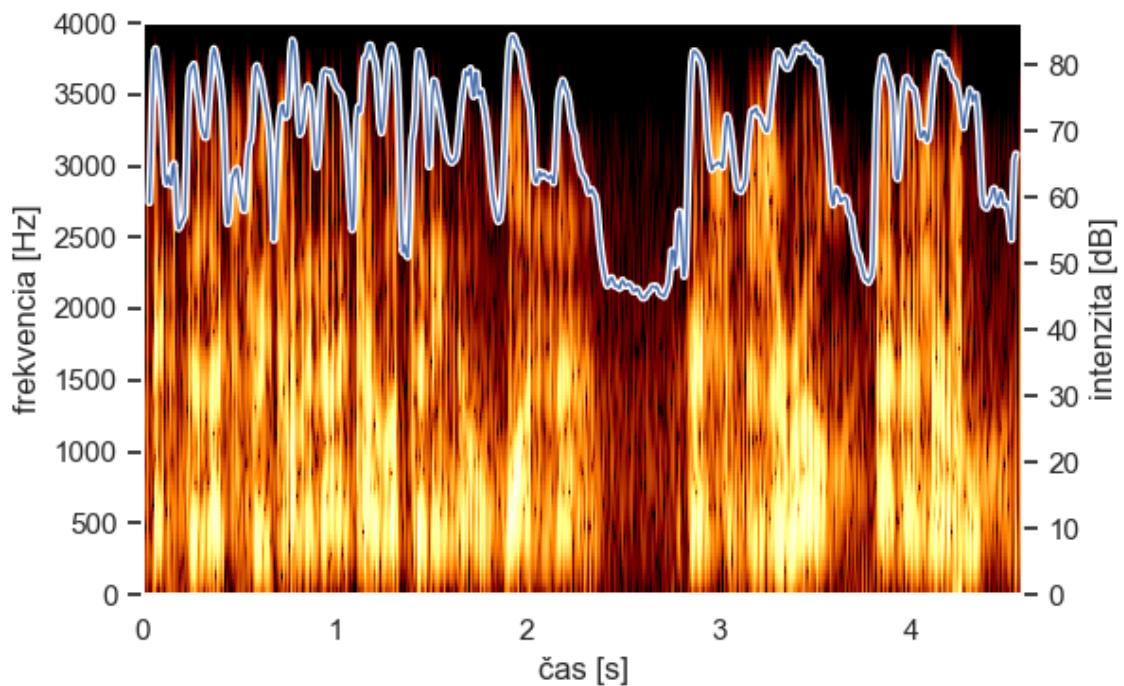
Pomocou knižnice Pandas, bol vytvorený DataFrame, ktorý slúži ako matica pre každý smer komunikácie, kde do každého riadku, boli vypísané číselné hodnoty analýz pre intenzitu, spektrum a základnú frekvenciu reči (hodnoty pre priemer, maximum, minimum a štandardnú smerodajnú odchýlku). Z údajov o spektre, boli taktiež navyše extrahované údaje o spektrálnom sklone a centroide. Do tejto matice bolo pri každom kroku pridané aj tempo, ktoré bolo už skôr obsiahnuté. Ako posledný krok pre získanie kompletných údajov pre dôslednú analýzu reči a jej porovnanie, bolo potrebné zahrnúť detekciu ľudského hlasu. Ku tomu bola využitá knižnica webrtcvad. Pri Voice Activity Detection bolo potrebné, správne si určiť vzorkovaciu frekvenciu. Maximálna dĺžka doby jedného okna reči, s ktorou webrtcvad vie pracovať bola 30ms. Pre čo najpresnejšie výsledky, bola zvolená práve táto doba, keďže čím dlhšie okno reči, tým spoľahlivejšie vie knižnica určiť, či je v danom okne reč prítomná alebo nie. Znovu boli načítané všetky segmenty konverzácie postupne. Každý segment sa jednotlivito čítal jednotlivito po oknách, analyzoval ich a vracal jedine boolean hodnoty 0 alebo 1 podľa toho, či v danom okne bola reč prítomná alebo nie. V prípade, že reč prítomná bola resp. nebola, si program ukladal počet hodnôt 1 alebo 0 do premenných true\_frames a false\_frames. Z nich boli následne extrahované detaily o najdlhšej kontinuálnej reči, o najdlhšej konštantnej pauze a nakoniec aj o pomere detekovanému hlasu ku nedetekovanému. Práve tieto tri údaje boli nakoniec doplnené ako posledné tri stĺpce a tak vytvorili kompletnú maticu údajov v jednom smere komunikácie. Tento istý postup sa zopakoval aj pre druhý smer aby sa následne mohli výsledky porovnať. Matice boli nakoniec zlúčené a pridaný bol fiktívny stĺpec, ktorý určoval smer prebiehajúcej konverzácie.



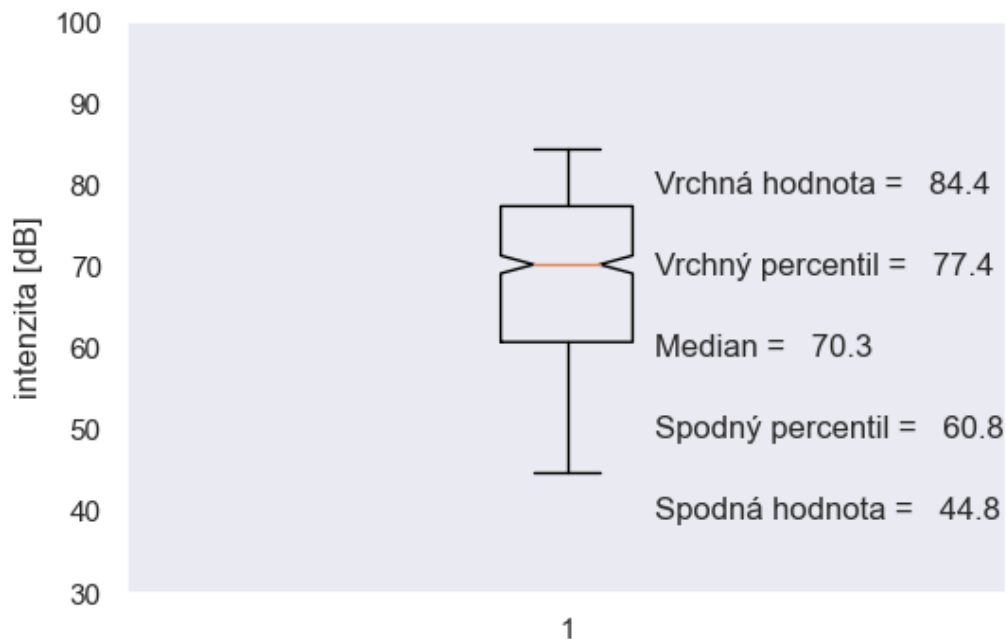
Obrázok 4.3 Graf časového priebehu segmentu komunikácie s detekciou reči

## 4.5 Generovanie grafov

Ku každému segmentu, boli vygenerované grafy analýz pomocou knižnice matplotlib. Grafy boli vygenerované pre amplitúdu hlasu v čase, intenzitu hlasu v čase a základnú frekvenciu  $F_0$ , grafy spektra, centroidu a strmosti a v neposlednom rade aj graf detekovanej reči v rámci časového priebehu signálu pre každý konkrétny segment komunikácie. Pre porovnanie a určenie podobností a rozličností dát, boli taktiež vytvorené boxploty, kvôli porovnaniu a priblíženiu všetkých údajov jednej strany komunikácie. Grafy boli uložené vo formáte .png a pomenované podľa príslušných segmentov komunikácie a sú uložené v prílohách. V prílohe A sú pripnuté boxploty, rozdelené podľa príslušného smeru komunikácie, kde v ľavom stĺpci sa nachádzajú hodnoty pre kontúry  $F_0$  v čase a v pravom hodnoty intenzity pre ten istý segment komunikácie. Na prvý pohľad je zrejmé, že pri vybraných segmentoch sú kontúry  $F_0$  v smere pilot veži vyššie, ako pri smere opačnom. Pri tých istých segmentoch sa však hodnoty intenzity značne nelíšia. Nie len preto bolo potrebné, pre hĺbkovú analýzu príznakov, využiť strojové učenie, ktoré dokáže presne určiť podstatu jednotlivých príznakov, pri veľkom počte vzorkov



Obrázok 4.4 Graf zobrazujúci intenzitu hlasu v čase, pre jeden segment komunikácie v smere veža pilotovi

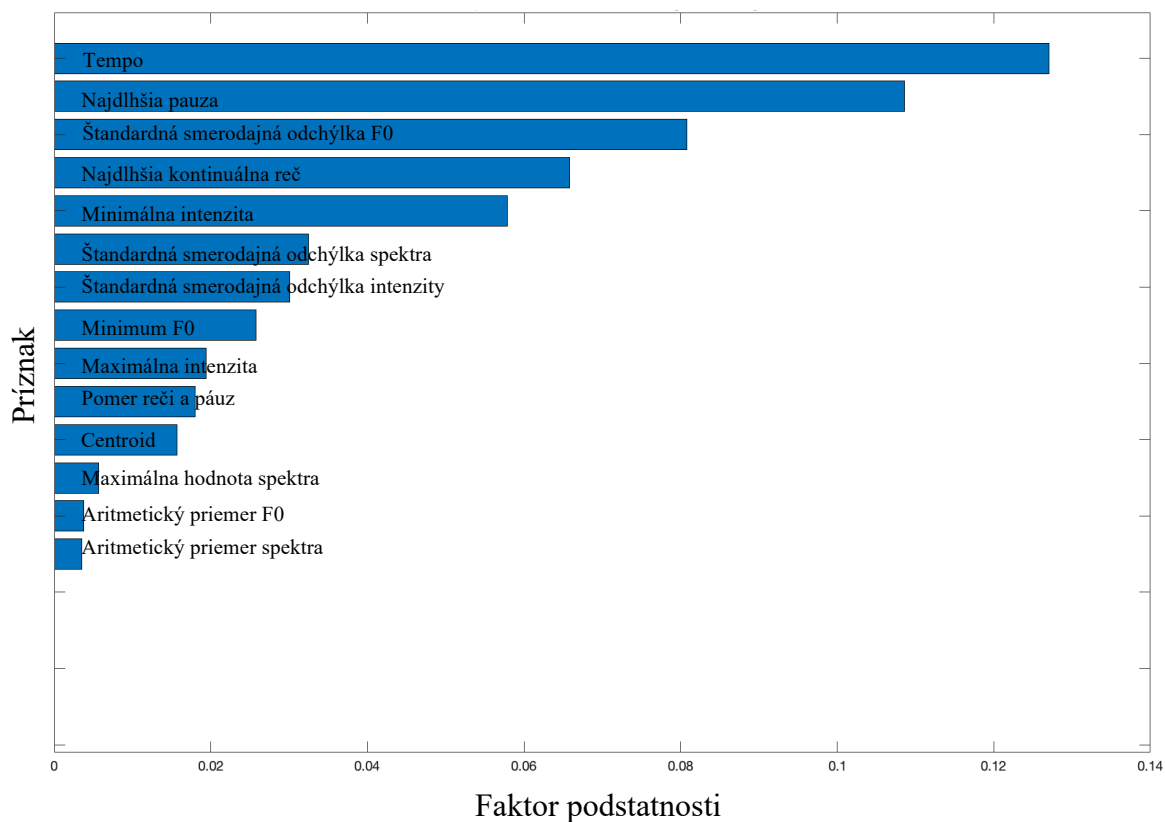


Obrázok 4.5 Boxplot pre intenzitu v čase pre jeden segment komunikácie v smere veža pilotovi

## 4.6 Automatické rozpoznávanie smeru komunikácie pomocou rozhodovacích stromov

Pre porovnanie výsledkov, bola vytvorená komplexná matica, kde v riadkoch boli uložené výsledky jednotlivých parametrov analýzy. Ako faktor rozlišovania hodnôt, bol smer prebiehajúcej ATC komunikácie jednotlivého segmentu komunikácie, nachádzajúci sa, ako posledný stĺpec matice. Parametrami pre porovnávanie hodnôt boli práve zvolené prozodické vlastnosti. Výsledná matica o rozmeroch 1024x18 sa importovala do prostredia Matlab a pomocou aplikácie classificationLearner sa porovnali parametre podľa smeru komunikácie. Inštancii bolo práve 1024 pri osemnástich príznakoch. V matici boli porovnávané parametre tempa, najdlhšej a najkratšej pauzy, ich pomeru, spektrálneho centroidu, spektrálneho sklonu, samotných spektier a nakoniec maximálnych a minimálnych hodnôt intenzity a kontúr F0 vrátane ich smerodajnej odchýlky. Dáta z matice sa porovnávali krížovou validáciou s piatimi iteráciami. Klasifikácia výsledkov prebehla podľa rozhodovacieho stromu Fine Tree. Po trénovaní za pomoci strojového učenia, bol získaný výsledok presnosti určenia smeru komunikácie, na základe zistených parametrov z matice 80,2%, čo je dostatočnou úspešnosťou pre správne určenie väčšiny segmentov komunikácie. Algoritmom minimálnej redundantnosti, maximálnej relevantnosti boli zistené najpodstatnejšie parametre pre rozlíšenie smeru. Najpodstatnejším faktorom, pre správne určenie smeru komunikácie, bolo z výsledkov analýz tempo, ktoré bolo značne vyššie pri komunikácii, kde zdrojom bola riadiaca veža. Hodnoty sa pri tempe pohybovali pri viacerých segmentoch až nad úroveň 7 – 9. Druhým najpodstatnejším faktorom, bola taktiež prozodická vlastnosť páuz. Pri aspekte najdlhšej pauzy, bol znovu dominantnejší smer komunikácie zo smeru

riadiacej veže. Ako tretí najpodstatnejší faktor, bola štandardná odchýlka kontúr základnej frekvencie F0.



Obrázok 4.6 Graf podstatnosti príznakov pre správne určenie smeru komunikácie

## 5. ZÁVER

Ako prvý krok bolo obsiahnutie dát s ktorými s následne bude pracovať. K dispozícii bolo viacero nahrávok nepretržitej komunikácie z rôznych letísk a aj ich prepis do textového dokumentu, z korpusu ATC0.

Aby mohlo dôjsť k automatickému rozpoznávaniu smeru komunikácie, bolo najprv nutné rozdeliť danú nahrávku na samostatné súbory a takisto ich následne roztriediť a uložiť podľa príslušnému smeru komunikácie spolu s vzostupným číslovaním pre prehľadné dohľadanie podľa prepisu. Následne už rozdelené nahrávky boli analyzované pomocou skriptu v programovacom jazyku python a takisto boli extrahované dáta o amplitúde v čase, intenzite hlasu v čase, samotného spektra, spektrálneho sklonu, centroidu, tempa, detekcii reči a v neposlednom rade aj údaje o základnej frekvencii F0 v čase daného segmentu. Z grafov boli vytvorené boxploty, ktoré boli následne porovnané, pre získanie obrazu o spektrálnych a prozodických vlastnostiach oboch strán komunikácie. Po ich získaní boli dáta porovnané podľa toho, čo majú spoločné a čo odlišné v rámci rozdelenia podľa smeru komunikácie a na základe toho, bolo možné rozlíšiť, aké hodnoty údajov sprevádzajú komunikáciu v určitom smere. Okrem boxplotov boli údaje overené aj priamo z matice pomocou Matlabu, kde sa určila presnosť určenia zdroja komunikácie na základe analyzovaných parametrov na 80,2%. Zároveň boli zistené aj parametre, zoradené podľa relevantnosti v určovaní zdroja komunikácie. Najdominantnejšími faktormi pre určenie smeru, bolo práve tempo, dĺžka páuz a štandardná odchýlka pri kontúrach základnej frekvencie F0.

Výsledkom práce bolo zistenie, že práve kombináciou prozodických a spektrálnych vlastností rečníka sa môže s vyššou pravdepodobnosťou určiť smer prebiehajúcej ATC komunikácie. Obidva typy týchto vlastností idú so sebou ruka v ruke, a spoliehať sa len na jeden z nich, by značne ovplyvnilo všetky výsledky meraní.



## LITERATÚRA

- [1] BARSHI, Immanuel a Candace FARRIS. *Misunderstandings in ATC communication : language, cognition, and experimental methodology*. 2016. Abingdon: Routledge, 2016. ISBN 978-0-7546-9933-0.
- [2] NOLAN, Michael S. *Fundamentals of Air Traffic Control*. 5th ed. Clifton Park (New York): Cengage, c2011. ISBN 978-1-4354-8272-2.
- [3] HOFBAUER, Konrad, Stefan PETRIK a Horst HERING. *The ATCOSIM Corpus of Non-Prompted Clean Air Traffic Control Speech*.
- [4] MARY, Leena. *Extraction of Prosody for Automatic Speaker, Language, Emotion and Speech Recognition*. 2nd ed. Cham: Springer, 2019. ISBN 978-3-319-91171-7.
- [5] S. A. A. Qadri, T. S. Gunawan, T. Wani, M. F. Alghifari, H. Mansor and M. Kartiwi, "Comparative Analysis of Gender Identification using Speech Analysis and Higher Order Statistics," *2019 IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, 2019, pp. 1-6, doi: 10.1109/ICSIMA47653.2019.9057296.
- [6] RAO, K. Sreenivasa a Sourjya SARKAR. *Robust Speaker Recognition in Noisy Environments*. 1. Springer, 2014. ISBN 978-3-319-07129-9.
- [7] GRIMM, Michael a Kristian KORSCHEL, ed. *Robust Speech Recognition and Understanding* [online]. 1. Londýn: I-Tech Education and Publishing, 2007 [cit. 2022-05-10]. Dostupné z: <https://www.intechopen.com/books/3371> doi:10.5772/35
- [8] Online Measuring of Available Resources - Scientific Figure on ResearchGate. Available from: [https://www.researchgate.net/figure/The-two-dimensions-of-emotions-Valence-negative-positive-and-arousal-low-high-Every\\_fig1\\_318031044](https://www.researchgate.net/figure/The-two-dimensions-of-emotions-Valence-negative-positive-and-arousal-low-high-Every_fig1_318031044) [accessed 17 Apr, 2022]
- [9] Cruttenden, A. *Intonation*, Cambridge University Press, 1986.
- [10] ARNFIELD, Simon, Peter ROACH, Jane SETTER, Peter GREASLEY a Dave HORTON. *Emotional Stress and Speech Tempo Variation* [online]. Lisbon, 1995, 13-15 [cit. 2022-04-26]. Dostupné z: [https://www.isca-speech.org/archive/open/sus\\_95/sus5\\_013.html](https://www.isca-speech.org/archive/open/sus_95/sus5_013.html)
- [11] J. Zhuo, K. Xu and J. Wan, "Target classification for passive sonar based on Mel cepstrum and image texture," *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, 2016, pp. 1682-1686, doi: 10.1109/ICSP.2016.7878114.
- [12] H. Xu, X. Zhang and L. Jia, "The extraction and simulation of Mel frequency cepstrum speech parameters," *2012 International Conference on Systems and Informatics (ICSAI2012)*, 2012, pp. 1765-1768, doi: 10.1109/ICSAI.2012.6223385.
- [13] KIUSALAAS, Jaan. *Numerical Methods in Engineering with Python*. Cambridge: Cambridge University Press, 2005. ISBN 978-0-511-85287-6.

- [14] GRIMM, Michael a Kristian KORSCHEL, ed. *Robust Speech Recognition and Understanding* [online]. 1. Londýn: I-Tech Education and Publishing, 2007 [cit. 2022-05-10]. Dostupné z: <https://www.intechopen.com/books/3371>  
doi:10.5772/35
- [15] JADOUL, Yannick. *Parselmouth - Praat in Python, the Pythonic way* [online]. [cit. 2022-05-23]. Dostupné z: <https://parselmouth.readthedocs.io/en/stable/>
- [16] AARSEN, Tom. *NLTK :: Natural Language Toolkit* [online]. [cit. 2022-05-23]. Dostupné z: <https://www.nltk.org/>
- [17] Harris, C.R., Millman, K.J., van der Walt, S.J. et al. *Array programming with NumPy*. *Nature* 585, 357–362 (2020). DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2).  
Dostupné z: <https://www.nature.com/articles/s41586-020-2649-2>
- [18] Jeff Reback, Wes McKinney, jbrockmendel, Joris Van den Bossche, Tom Augspurger, Phillip Cloud, gyoung, Sinhrks, Adam Klein, Matthew Roeschke, Simon Hawkins, Jeff Tratner, Chang She, William Ayd, Terji Petersen, Marc Garcia, Jeremy Schendel, Andy Hayden, MomIsBestFriend, ... Mortada Mehyar. (2020). pandas-dev/pandas: Pandas 1.0.3 (v1.0.3). Zenodo.  
<https://doi.org/10.5281/zenodo.3715232>
- [19] GOUSKOVA, Maria. *Praat Tutorial* [online]. [cit. 2022-05-23]. Dostupné z: <https://www.gouskova.com/2016/09/03/praat-tutorial/>
- [20] Godfrey, John J.. *Air Traffic Control Complete LDC94S14A*. Web Download. Philadelphia: Linguistic Data Consortium, 1994.

## ZOZNAM SYMBOLOV A SKRATIEK

### Skratky:

LPCC	Linear Prediction Cepstrum Coefficient
WAV	Waveform Audio File Format
ATC	Air Traffic Control
MFCC	Mel Frequency Cepstral Coefficient
API	Application Programming Interface
SNR	Signal-to-noise ratio

### Symboly:

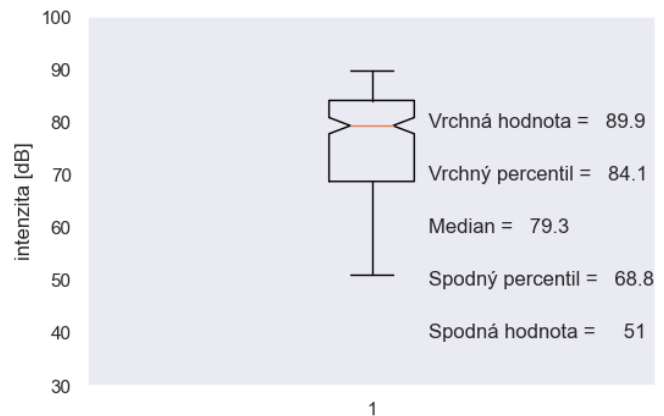
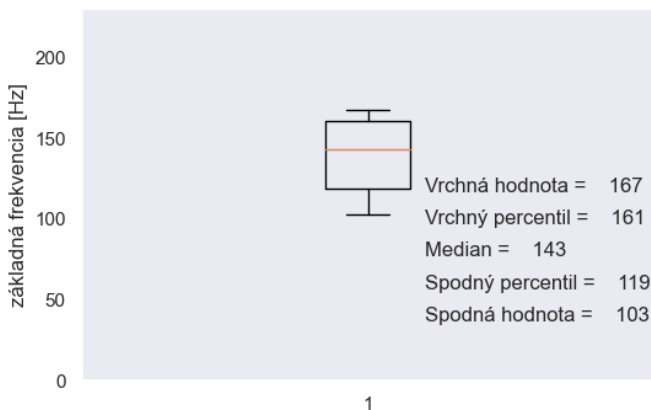
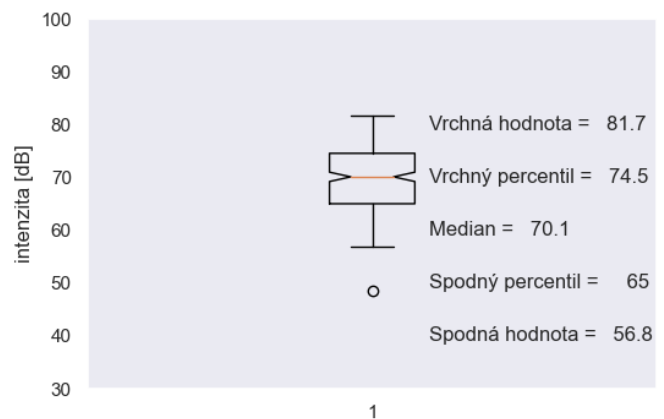
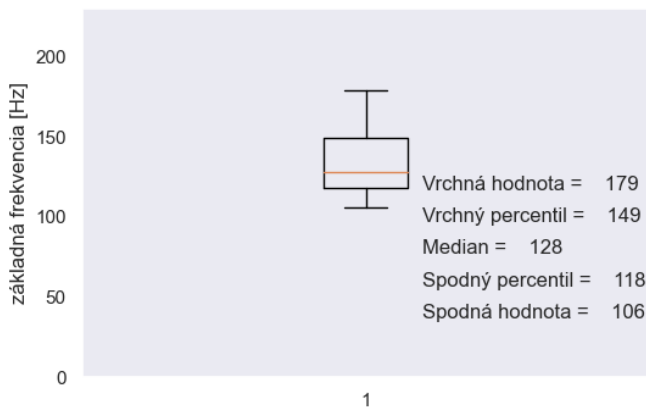
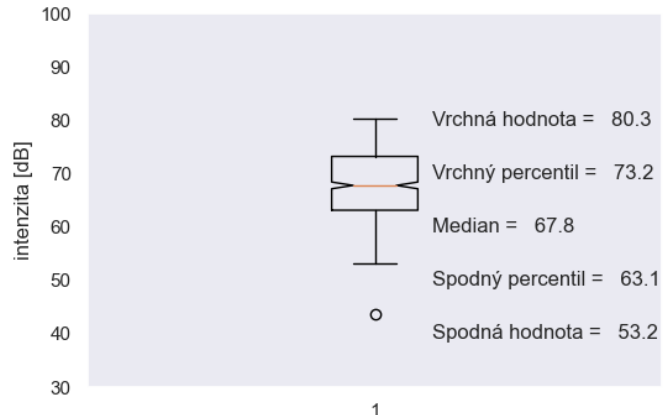
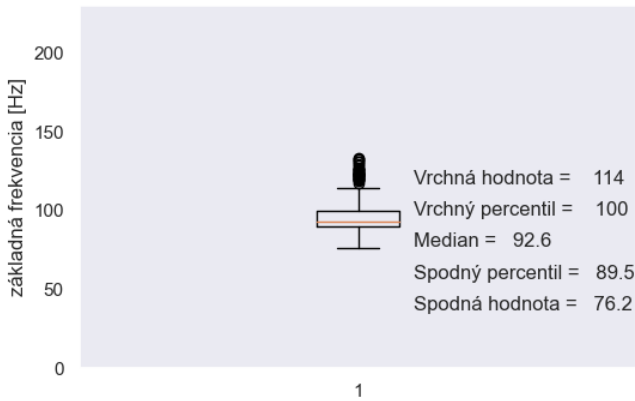
$S(f)$	spektrum čistej reči	(-)
$X(f)$	reč so šumom	(-)
$N(f)$	spektrum šumu	(-)
$\alpha$	faktor nadmerného odčítania	(-)
$\beta$	maximálny útlm	(-)
$W(f)$	Wienerov filter	(-)
$\Phi_{ss}(f)$	spektrálny výkon čistej reči	(-)
$\Phi_{nn}(f)$	spektrálny výkon samotného šumu	(-)
$f_k$	frekvencia	(Hz)
$s_k$	spektrálna hodnota	(-)
$b.$	hranice pásma	(-)
$\mu_f$	priemerná frekvencia	(Hz)
$\mu_s$	priemerná spektrálna hodnota	(-)
$\mu_1$	centroid	(-)
$\mu_2$	rozloženie spektra	(-)
$T$	čas rozprávania	(s)
$N_s$	počet slabík	(-)
$N_p$	počet páuz	(-)
$T_{savg}$	priemerná dĺžka slabiky	(s)
$T_{pavg}$	priemerná doba pauzy	(s)

# ZOZNAM PRÍLOH

PRÍLOHA A -	NAMERANÉ HODNOTY .....	45
A.1	NAMERANÉ HODNOTY – BOXPLOTY V SMERE PILOT VEŽI.....	45
A.2	NAMERANÉ HODNOTY – BOXPLOTY V SMERE VEŽA PILOTOVI.....	46

# Príloha A - Namerané hodnoty

## A.1 Namerané hodnoty – boxploty v smere pilot veži



## A.2 Namerané hodnoty – boxploty v smere veža pilotovi

