

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

AUTOMATICKÁ KLASIFIKACE VÝSLOVNOSTI HLÁSKY R

AUTOMATIC CLASSIFICATION OF PRONUNCIATION OF THE LETTER „R“

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Enrik Hrušovský

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Vratislav Harabiš, Ph.D.

BRNO 2018

Diplomová práce

magisterský navazující studijní obor **Biomedicínské a ekologické inženýrství**

Ústav biomedicínského inženýrství

Student: Bc. Enrik Hrušovský

ID: 183581

Ročník: 2

Akademický rok: 2017/18

NÁZEV TÉMATU:

Automatická klasifikace výslovnosti hlásky R

POKYNY PRO VYPRACOVÁNÍ:

- 1) Seznamte se se základními postupy odstraňování vad řeči u dětí předškolního věku. 2) Seznamte se s metodami pro klasifikaci a rozpoznávání řeči, zaměřte se především na výslovnost hlásky „R“ a proveďte literární rešerši. 3) Definujte podmínky a ve spolupráci s logopedy kliniky LOGO se zaměřte na určitý rozsah hlásky „R“. 4) Navrhněte metodu pro automatickou klasifikaci hlásky „R“ při výslovnosti určitých tréninkových slov. 5) Ve zvoleném programovém prostředí realizujte navrženou metodu. 6) Navrženou metodu otestujte na dostatečném počtu nahrávek a proveďte vyhodnocení úspěšnosti navrženého řešení. 7) Navrhněte vhodný postup pro nácvik a kontrolu výslovnosti hlásky „R“.

DOPORUČENÁ LITERATURA:

[1] KEJKLÍČKOVÁ, Ilona. Vady řeči u dětí: návody pro praxi. Praha: Grada, 2016, 224 s. ISBN 978-80-247-3-41-0.

[2] DONG, Yu, DENG, Li. Automatic speech recognition. Springer London, 2015, ISBN 9781447157786.

Termín zadání: 5.2.2018

Termín odevzdání: 18.5.2018

Vedoucí práce: Ing. Vratislav Harabiš, Ph.D.

Konzultant:

**prof. Ing. Ivo Provozník,
Ph.D.**

předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové 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.

Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně / Technická 3058/10 / 616 00 / Brno

Abstrakt

Táto diplomová práca sa zaoberá automatickou klasifikáciou hlásky R. Dáva si za cieľ vytvoriť program pre detekciu logopedickej vady výslovnosti hlásky „R” u detí. V práci sú spracované témy ako tvorba reči, logopédia, dyslália a následne spracovanie rečových signálov a ich metódy analýzy. V poslednej časti je navrhnutý program pre automatickú klasifikáciu výslovnosti hlásky „R”. K rozpoznaniu výslovnosti je použitý algoritmus MFCC pre extrakciu príznakov. Tieto príznaky sú následne klasifikované neurónovou sieťou do triedy so správnou alebo nesprávnou výslovnosťou a je vyhodnotená úspešnosť klasifikácie.

Kľúčové slová

tvorba reči, logopédia, dyslália, segmentácia, spektrálna analýza, lineárne prediktívna analýza, percepčne lineárne prediktívna analýza, mel-frekvenčné keprálne koeficienty, neurónová sieť, automatická klasifikácia výslovnosti

Abstract

This diploma thesis deals with automatic classification of vowel R. Purpose of this thesis is to make program for detection of pronunciation of speech defects at vowel R in children. In thesis are processed parts as speech creation, speech therapy, dyslalia and subsequently speech signal processing and analysis methods. In the last part is designed software for automatic detection of pronunciation of vowel R. For recognition of pronunciation is used algorithm MFCC for extracting features. These features are subsequently classified by neural network to the group of correct or incorrect pronunciation and is evaluated classification success.

Key words

speech creation, speech therapy, dyslalia, segmentation, spectral analysis, linear predictive coding, perceptual linear prediction, mel-frequency cepstral coefficients, neural network, automatic classification of pronunciation

Bibliografická citácia

HRUŠOVSKÝ, E. *Automatická klasifikace výslovnosti hlásky R*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2018. 67 s. Vedoucí diplomové práce Ing. Vratislav Harabiš, Ph.D..

Prehlásenie

Prehlasujem, že svoju diplomovú prácu s názvom "Automatická klasifikace výslovnosti hlásky R" som vypracoval samostatne pod vedením vedúceho diplomovej práce a s použitím odbornej literatúry a iných informačných zdrojov, ktoré sú všetky v práci citované a uvedené v zozname použitej literatúry na konci práce.

Ako autor diplomovej práce ďalej prehlasujem, že v súvislosti s vytvorením tejto diplomovej práce som neporušil autorské práva tretích osob, predovšetkým som nezasiahol nedovoleným spôsobom do cudzích autorských práv osobnostných a som si plne vedomý následkov porušenia ustanovenia §11 a následujúcich autorského zákona č. 121/2000 Sb., vrátane možných trestnoprávných dôsledkov vyplývajúcich z ustanovenia §152 trestného zákona č. 140/1961 Sb.

V Brne dňa

.....
(podpis autora)

Pod'akovanie

Ďakujem vedúcemu práce Ing. Vratislavovi Harabišovi, Ph.D. za užitočnú metodickú pomoc a cenné rady při spracovaní diplomovej práce. Ďalšie pod'akovanie patrí PHDr. Petrovi Staníčkovi za poskytnuté logopedické nahrávky z kliniky LOGO.s.r.o.

V Brne dňa

.....
(podpis autora)

Obsah

Úvod.....	12
1. Tvorba reči	13
1.1 Matematický model tvorby reči	15
1.2 Rečové jednotky.....	15
2. Logopédia	17
2.1 Narušená komunikačná schopnosť	17
2.2 Logopedická intervencia.....	17
3. Dyslália	19
3.1 Etiológia dyslálie	20
3.2 Klasifikácia dyslálie.....	20
3.2.1 Delenie z vývojového hľadiska.....	20
3.2.2 Delenie podľa rozsahu	21
3.2.3 Delenie z hľadiska kontextu	21
3.3 Diagnostika dyslálie.....	21
3.4 Terapia dyslálie	22
4. Spracovanie signálov	24
4.1 Detekcia hraníc slov.....	24
4.1.1 Krátkodobá energia signálu	24
4.1.2 Počet priechodov nulou.....	24
4.2 Segmentácia	25
4.2.1 Oknové funkcie.....	25
4.3 Preemfáza.....	25
5. Metódy pre analýzu rečových signálov	26
5.1 Spektrálna analýza	26
5.1.1 Furierova analýza.....	26
5.1.2 Vlnková analýza.....	27
5.2 Kepstrálna analýza	28
5.3 Lineárne prediktívna analýza	29
5.4 Perceptívne lineárne predikčné koeficienty	30
5.5 MEL-frekvenčné kepstrálne koeficienty.....	31
6. Klasifikátory	34
6.1 Algoritmus k-najbližších susedov	34
6.2 Gaussov zmiešaný model.....	35
6.3 Markovove modely	37
6.3.1 Markovove modely	37
6.4 Neurónové siete	39
6.4.1 Matematický model neurónu	39
6.4.2 Činnosť neurónovej siete	40
6.4.3 Učenie neurónových sietí.....	40
6.4.4 Dopredné neurónové siete.....	41
6.4.5 Rekurentné neurónové siete.....	41
7. Klasifikácia výslovnosti hlásky „R“	43
7.1 Naviazanie na semestrálnu prácu.....	43
7.2 Nahrávky tréningových slov	43
7.3 Súčasť programu pre klasifikáciu výslovnosti slov.....	44
7.4 Predspracovanie a výpočet MFCC.....	45
7.5 Extrahovanie MFCC pre natréňovanie neurónovej siete	48
7.6 Klasifikátor – neurónová sieť	48
7.6.1 Tréňovanie neurónovej siete pre klasifikáciu slov.....	49
7.6.2 Tréňovanie neurónovej siete pre klasifikáciu hlások.....	52
7.6.3 Tréňovanie a testovanie neurónovej siete pri použití všetkých slov.....	54

7.6.4 Trénovanie a testovanie neurónovej siete pri použití všetkých hlások.....	56
7.7 Automatická klasifikácia – načítanie nahrávky	57
7.7.1 Výsledky automatickej klasifikácie slov	57
7.7.2 Výsledky automatickej klasifikácie ručne nasegmentovaných hlások „R“	59
7.8 Zhodnotenie výsledkov	60
8. Postup nácviku a kontroly výslovnosti hlásky „R“	62
Záver	63
Použitá literatúra	64
Zoznam skratiek	66
Obsah priloženého CD	67

Zoznam obrázkov

Obr. 1.1: Model tvorby reči	15
Obr. 5.1: Blokové schéma výpočtu reálneho kepstra	29
Obr. 5.2: Trojuholníkové filtre pozdĺž frekvenčnej osy v Hz.....	32
Obr. 5.3: Blokové schéma výpočtu MFCC.....	33
Obr. 6.1: Klasifikácia K-NN do dvoch tried.....	35
Obr. 6.2: Výsledné zhluky MFCC pomocou GMM	37
Obr. 6.3: Uzly Markovovho modelu s prechodmi	37
Obr. 6.4: Matematický model neurónu	39
Obr. 7.1: Nahrávka slova ryba s logopedickou vadou (lyba)	45
Obr. 7.2: Nahrávka po odstránení jednosmernej zložky.....	45
Obr. 7.3: Nahrávka po preemfáze	46
Obr. 7.4: Nahrávka po normovaní	46
Obr. 7.5: Nahrávka po orezaní začiatku a konca	46
Obr. 7.6: Nahrávka po fft.....	47
Obr. 7.7: Výsledné MFCC nahrávky	48
Obr. 7.8: Bloková schéma predspracovania a výpočtu MFCC.....	48
Obr. 7.9: Výber dát (input a target) pre neurónovú sieť	49
Obr. 7.10: Rozdelenie počtu nahrávok pre tréning a testovanie	50
Obr. 7.11: Natréningovanie neurónovej siete	50
Obr. 7.12: Percentuálna úspešnosť tréningovania, validácie a testovania slov	51
Obr. 7.13: Krivka ROC pre slová	52
Obr. 7.14: Percentuálna úspešnosť tréningovania, validácie a testovania hlások.....	53
Obr. 7.15: Krivka ROC pre hlásky	53
Obr. 7.16: Počet slov pre tréningovanie, validáciu a testovanie	54
Obr. 7.17: Výsledok samootestovania sa neurónovej siete.....	55
Obr. 7.18: ROC krivka.....	55
Obr. 7.19: Počet hlások pre testovanie, validáciu a tréningovanie	56
Obr. 7.20: Výsledok samootestovania sa neurónovej siete.....	56
Obr. 7.21: Krivka ROC	57
Obr. 7.22: Zobrazenie znaku pre nesprávnu výslovnosť	58
Obr. 7.23: Zobrazenie znaku pre správnu výslovnosť	58
Obr. 7.24: Bloková schéma automatickej klasifikácie slov	58
Obr. 7.25: Bloková schéma automatickej klasifikácie hlások	59

Zoznam tabuliek

Tab. 7.1: Výsledok automatickej klasifikácie slov	58
Tab. 7.2: Výsledok automatickej klasifikácie slov spolu s testom neurónovej siete	59
Tab. 7.3: Výsledok automatickej klasifikácie slov	59
Tab. 7.4: Výsledok automatickej klasifikácie slov spolu s testom neurónovej siete	60

Úvod

Reč je prirodzená forma komunikácie medzi ľuďmi. Počas vývoja dieťaťa dochádza k osvojovaniu si dôležitých skúseností v rámci správnej výslovnosti. Tieto skúsenosti zdokonaľujeme a spoliehame sa na komunikáciu počas celého nášho života. Často sa vyskytujú rôzne poruchy komunikačnej schopnosti u detí. Výslovnosť hlásky „R“ je práve jednou z najčastejších porúch výslovnosti u detí ale aj následne u dospelých osôb. Náprava poruchy komunikačnej schopnosti je hlavnou náplňou práce logopédov, predovšetkým u detí. Cieľom tejto diplomovej práce je vytvoriť program, ktorý by dokázal klasifikovať výslovnosť hlásky „R“, a zhodnotil či hláska bola vyslovená správne alebo nesprávne.

V rámci práce je spracovaná teória v ktorej som sa venoval vývoju reči, logopédií a dyslálií, ktorá je jednou z najčastejšie sa vyskytujúcich porúch komunikačnej schopnosti. V teoretickej časti sú tiež zhrnuté metódy spracovania rečových signálov, s ich následnou analýzou a klasifikáciou. Analýza rečového signálu je popisovaná pomocou metód, ktoré extrahujú príznaky zo signálu. Tieto metódy sú LPC, PLP a MFCC. V poslednej kapitole teoretickej časti sú popísané princípy klasifikátorov, medzi ktoré patrí K-NN, GMM, HMM a neurónové siete.

Praktická časť sa venuje popisu zhotoveného programu a následne štatistickým údajom o jeho natrénovaní, testovaní a vyhodnotení. Taktiež je popísaná metóda nácviku a kontroly výslovnosti hlásky „R“.

1. Tvorba reči

Na tvorbe reči sa podieľajú CNS (centrálne nervová sústava), obzvlášť rečové centrá mozgovej kôry, čo sú výkonné oblasti širokého neurokognitívneho systému mozgu, ktoré zabezpečujú proces rečovej komunikácie. Ďalej je to realizácia nervových vzruchov činnosťou orofarciálnej sústavy. Orofariálna sústava pozostáva z vstupnej a výstupnej časti dýchacej sústavy, fonačné a artikulačné orgány, mimické svalstvo a systém hlavových dutín.

Dýchacia sústava predstavuje hlavný zdroj energie pre vznik reči. Podstatná časť je umiestnená v hrudnom koši, a sú to pľúca a dýchacie svaly. Medzi dýchacie svaly sa radia medzirebrové svaly a bránica. V procese nádychu dochádza k pohybu vzduchu, ktorý poskytuje zdroj energie pre reč. Následne pri výdychu vzniká výdychový prúd vzduchu v pľúcach, ktorý je základom pre tvorbu reči. Vzduch pri vydýchnutí sa z pľúc dostáva priedušnice, čo je trubicový orgán a následne prechádza hrtanom a nadhrtanovými dutinami. V tejto časti sa vydychovaný vzduch, ktorý je pri výdychu približne konštantný, modifikuje na rečový signál. Podľa dĺžky produkovaného rečového signálu sa predlžuje trvanie výdychu, a sila vydychovaného vzduchu má vplyv na silu hlasu ale aj jeho výšku.

Dôležitou časťou je fonačné ústrojenstvo, nazývané taktiež hlasové, ktoré je uložené v hrtane. Tento úsek dýchacej sústavy je ohraničený napojením na priedušnicu zdola a jazykovou kosťou zhora. Hrtan je orgán tvorený viacerými chrupavkami, väzmi a kĺbmi. Taktiež sa v ňom nachádza pár hlasivkových chrupaviek a ostré slizničné riasy, nazývané ako hlasivky, ktoré sú najdôležitejšou časťou hlasového ústrojenstva. Ostré slizničné riasy sú uložené naprieč hrtanom v jeho najužšom mieste, prichytené na hlasivkové chrupavky z jednej strany a na štítnu chrupavku z druhej strany. Taktiež ich dĺžka je typická, pre mužov 15mm a pre ženy o niečo menej a to 13mm. Medzi hlasivkami sa nachádza štrbina trojuholníkového tvaru nazývaná ako hlasivková štrbina. Hlasivky držia hlasivkovú štrbinu otvorenú pri kľudovej polohe hlasiviek, a teda pri nevdôle človeka produkovať reč, aby cez ňu mohol prechádzať vzduch pri nádychu a výdychu bez odporu. Pri tvorbe reči, fonačnej polohe hlasiviek, hlasivky uzatvoria cestu výdychovému prúdu vzduchu. V takejto pozícii sa hlasivky pod tlakom vzduchu stanú pružnejšími a začnú kmitať, a následne sa rýchlo zatvárajú a otvárajú, čo má za následok vznik vzduchovej vlny. Vzduchová vlna je periodický prúd hustejšieho a redšieho vzduchu, nazývaný tiež základný tón, ktorý predstavuje nosný zvuk reči. S ním je spojená aj frekvencia základného tónu, čo je frekvencia kmitania hlasiviek. Frekvencia kmitania sa líši u detí a dospelých, rovnako aj u mužov a žien. V prípade fonačnej polohy hlasiviek sa vytvárajú znelé zvuky reči. Ak sú hlasivky v kľudovej polohe, a teda neje produkovaný základný tón, vytvárajú sa neznelé zvuky reči v nadhrtanových dutinách.

Artikulačné ústrojenstvo je poslednou časťou, ktorá sa podieľa na tvorbe reči. Artikulačným ústrojenstvom sa vytvára veľké množstvo rôznych zvukov, ktoré charakterizujú reč. Pozostáva z artikulačných orgánov a nadhrtanových dutín, ktoré sú tvorené hrdlovou, ústnou a nosnou dutinou. Nadhrtanové dutiny sa pri tvorbe reči nepohybujú, takže sa podieľajú pasívne, oproti pohybujúcim sa, teda aktívne sa podieľajúcich artikulačných orgánov. Artikulačné orgány zahŕňajú orgány ústnej dutiny, a teda pery, jazyk, mäkké podnebie, sánka a mimické svaly. Najdôležitejším z artikulačných orgánov je jazyk, ktorý je veľmi pružný a prispôsobivý. Jazyk je schopný tvoriť veľmi veľa tvarov a rýchlo prechádzať z pozície do pozície, čo umožňuje produkciu najrôznejších zvukov reči. Ďalším artikulačným mechanizmom je uzatváranie a otváranie nosnej dutiny, ktoré sa zásadne účastní modifikácie sily a zafarbenia hlasu. Hlavové dutiny zosilňujú individuálne rôzne frekvencie zvuku, v závislosti na sile a spôsobe fonácie, a tak vzniká rôzne zafarbenie hlasu u konkrétnych ľudí.

Pomocou pohybu všetkých predchádzajúcich orgánov vznikajú rôzne hlásky. Samohlásky ako „A,E,I,O,U“ vznikajú usmerňovaním tónu pomocou tvarov orgánov ústnej dutiny. Hláska „H“ vzniká šumom priamo v hlasivkovej štrbine. Spoluhlásky ako „P,B,M,T,D,N,Ť,Ď,Ň,K,G“ vznikajú v mieste otvorenia predchádzajúceho záveru, ostatné ako sú „S,Z,Š,Ž,CH,L,V,F“ vznikajú v mieste úžiny šumom výdychového prúdu, „C a Č“ vznikajú kombináciou oboch predchádzajúcich prípadov. Zvláštnym prípadom sú „R a Ř“ , ktoré sú vibrantami a vznikajú v úžine s kmitavým pohybom jazyka. Toto delenie je podľa spôsobu tvorby, iným delením je podľa miesta vzniku. Podľa miesta vzniku sa hlásky rozdeľujú do 5 takzvaných artikulačných okrsokov, ktoré sa čísľujú rímskymi číslicami:

I. artikulačný okrsok – patria sem hlásky obojperové (bilabiálne) a retrozubné (labiodentálne).

Obojperové sa tvoria oboma perami, a patria medzi ne „P,B,M“.

Retrozubné sú tvorené prekážkou, ktorú vytvorila spodná pera a horné rezáky, a sú to „F,V“.

II. artikulačný okrsok – hlásky d'asnové (alveolárne), vznikajú prekážkou, ktorú tvorí hrot jazyka

na hornom d'asne výdychovému prúdu. Delia sa na zubod'asnové

(alveodentálne) - „T,D,N,C,S,Z,L,R,Ř“ a zadod'asnové (postalveolárne) -

„Č,Š,Ž“.

III. artikulačný okrsok – hlásky tvrdod'asnové (palatálne), tvorené priblížením jazyka k tvrdému

d'asnu, sú to „Ť,Ď,Ň,J“.

IV. artikulačný okrsok – hlásky mäkkod'asnové (velárne), tvorené na mäkkom d'asne. Sú to

„K,G,CH“.

V. artikulačný okrsok – hlásky hrtanové (laryngeálne) – tvorené priamo v hrtane, je to jediná hláska

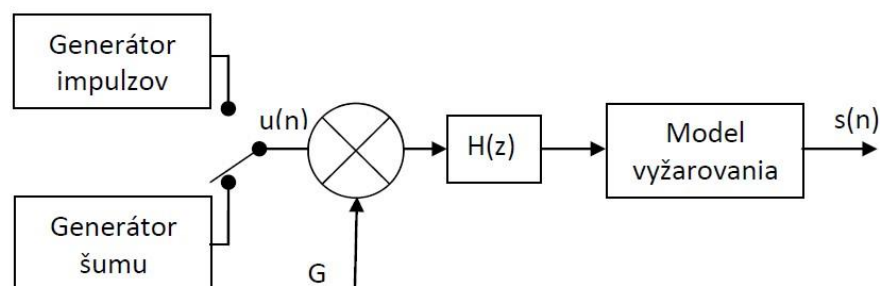
„H“. [13,18]

1.1 Matematický model tvorby reči

V procese výskumu vytvárania reči vznikla snaha matematicky popísať jeho proces a vytvoriť základný matematický model. Cieľom bol matematický popis takého zjednodušenia, ktoré bude rešpektovať tvorbu reči s minimálnymi požiadavkami a zložitou. Boli navrhnuté samostatné matematické modely pre jednotlivé časti produkcie reči, ako model hlasiviek, model hlasivkového traktu a model vyžarovania zvuku. Zjednotením týchto modelov vznikol výsledný celkový model tvorby reči, popísaný rovnicou

$$H(z) = \frac{1}{1 + \sum_{i=1}^Q a_i z^{-i}}, \quad (1.1)$$

kde $Q = 2K + 1$ a K je počet formantov zahrnutých do modelu. Tento model je následne budený G , ktorý je tvorený sledom impulzov pre znelé a neznelé zvuky. Normované vlastnosti budenia pre rôzne typy signálov predstavuje $u(n)$. Podľa obrázku 1.1 je signál tvorený dvoma generátormi, prvý je generátor pulzov pre znelé zvuky a druhý generátor šumu pre zvuky neznelé. Generátor pulzov má periódu generovania T_0 základnej periódy a generátor šumu generuje náhodný šum.



Obr. 1.1: Model tvorby reči

1.2 Rečové jednotky

Rečové jednotky sú dôležité pre systém rozpoznávania reči. Podstatnou časťou je segmentovať reč na rečové jednotky ako fonémy, alofóny, slabiky, a slová, pre ich čo najefektívnejšie rozpoznanie.

Fonéma je najmenšia komplexná jednotka zvukového systému jazyka, ktorá zahŕňa najmenší počet odlišných fonetických tried. V procese segmentácie by bola najvhodnejšia, ale ich automatické vymedzenie a klasifikácia je problematická, čo je spôsobené veľkou variabilitou akustickej realizácie fonémy.

Alofóna je pozíčná varianta fonémy, ktorá môže poskytnúť informáciu o hranici slabiky alebo slova. Nevýhodou použitia alofón je ich obrovský počet.

Slabiky sú fonetické útvary obsahujúce začiatočnú spoluhlásku, ďalej samohlásku a koncovú spoluhlásku. V procese segmentácie by slabiky boli vhodné pre ich jednoduchú detekciu jadra slabiky, no hranice slabiky z fonetického hľadiska nie sú jednoznačne určiteľné. Nevýhodou je tiež veľký počet slabík.

Slová môžeme definovať ako spojenie konečného počtu foném alebo slabík. Výhodou použitia slov pri segmentácii je vyvarovanie sa použitia komplikovaných algoritmov segmentácie nižších jednotiek ako foném a slabík a tiež ich jednoduchá segmentácia pri vyslovovaní izolovaných slov s dostatočnými pauzami. V súvislej reči kde sa pauzy medzi slovami veľmi skracujú je ich segmentácia zložitejšia. [15]

2. Logopédia

Začiatky oboru logopédie sa formovali už začiatkom 20. storočia. V roku 1924 viedenský lekár a foniatr Froeschels definoval logopédiu ako lekársku vedu, čo následne zmenil český lekár a foniatr Sovák, ktorý ju definoval ako obor špeciálne pedagogický. V ďalších rokoch rôzni lekári, nemedicínsky odborníci ale aj jazykovedci, popísali obor logopédie odlišne, čo znamená, že obor sa neustále mení a rozvíja. V súčasnosti sa radí sa medzi obory špeciálnej pedagogiky (defektológia). Slovo logopédia vychádza z dvoch gréckych slov - „logos“ znamená slovo a „paidea“ znamená výchova. Je definovaná ako výchova reči, alebo náuka o výchove a vzdelávaní jedincov s rečovými vadami. Súčasná logopédia sa neorientuje iba na deti, ale zaoberá sa aj narušením komunikačnej schopnosti u dospelých osôb. Obor logopédia má tesný vzťah k ostatným oborom špeciálnej pedagogiky, ako sú surdopedia, somatopedia, psychopedia aj oftalmopedia. Logopédia sa tiež zaoberá patologickou stránkou komunikačného procesu, to určuje vzťah s medicínskymi obormi ako pediatria, foniatra, otolaryngológia, stomatológia, neurológia, psychiatria a iné. Taktiež súvisí aj s jazykovednými obormi fonetiky a fonológie.

2.1 Narušená komunikačná schopnosť

Predmetom vedného odboru logopédie je narušená komunikačná schopnosť, ktorá je jedným zo základných termínov súčasnej logopédie. Definovať narušenú komunikačnú schopnosť je samo o sebe náročná úloha. Veľmi komplikované je už vymedzenie normality, to znamená určenie, kedy sa jedná o normu a kedy už môžeme hovoriť o narušení. Komunikačná schopnosť jednotlivca je narušená vtedy, keď niektorá rovina alebo viacero rovín jeho jazykových prejavov pôsobia interferenčne vzhľadom k jeho komunikačnému zámeru.

2.2 Logopedická intervencia

Je činnosť, ktorú odborník teda logopéd vykonáva. Logopedická intervencia je aktivita, ktorá je špecifická pre logopéda vo všetkých jej oblastiach, a logopéd ju uskutočňuje s určitým cieľom. Tieto ciele sú – 1. identifikovať, 2. eliminovať, teda zmierniť alebo aspoň prekonať narušenú komunikačnú schopnosť, alebo 3. predchádzať tomuto narušeniu, zlepšiť komunikačnú schopnosť. Logopedickú intervenciu chápeme ako zložitý multifaktoriálne podmienený proces, ktorý sa realizuje v záujme dosiahnutia týchto cieľov na troch úrovniach, a to :

1. logopedická diagnostika
2. logopedická terapia

3. logopedická prevencia.

Cieľom logopedickej diagnostiky je stanovenie diagnózy, kde termín „diagnosis“ znamená hĺbkové poznanie. Logopedická diagnostika má za cieľ čo najpresnejšie pochopenie narušenej komunikačnej schopnosti, ako východisko pre správny výber a aplikáciu intervenčných metód. Logopedická diagnostika má špecifikovať druh poruchy, príčinu vzniku, priebeh postihnutia, stupeň a následky narušenej komunikačnej schopnosti. Na základe tejto diagnostiky sa vytvorí plán logopedickej intervencie, podľa ktorého sa postupuje. Logopedická diagnostika má tri úrovne : orientačnú, základnú a špeciálnu. Orientačná sa robí v rámci screeningu, základná má za úlohu zistiť konkrétny druh a špeciálna na čo najpresnejšiu identifikáciu narušenej komunikačnej schopnosti. Závety logopedickej diagnostiky slúžia k výberu postupov, metód, foriem a prostriedkov logopedickej terapie.

Logopedickú terapiu charakterizujeme ako špecifickú aktivitu, ktorá sa realizuje špecifickými metódami, v špecifickej situácii zámerného učenia. Zámerné učenie znamená usmerňovanie a kontrolu logopéda v logopedických zariadeniach, k osvojovaniu špecifických vedomostí, zručností a návykov v oblasti komunikácie. Terapia má tri metódy : stimulujúcu, korigujúcu a reedukujúcu. Stimulujúca metóda sa aplikuje v terapii narušeného vývoja reči, korigujúca pri terapii dyslalie a reedukujúca pri afázií, kedy sa reedukujú stratené rečové funkcie.

V súčasnej logopédii sú tri typy prevencie, a to primárna, sekundárna a terciárna. Primárnou prevenciou je predchádzanie situáciám, ktoré ohrozujú určité narušenie komunikačnej schopnosti. Napríklad situácie, ktoré môžu vyvolať poruchy hlasu alebo podpora správneho vývoja reči u dieťaťa. Metódy primárnej prevencie sa uplatňujú obzvlášť v medicínskej oblasti. Sekundárna prevencia sa zameriava na určité rizikové skupiny, u ktorých sa prejavuje ohrozenie. Terciárna prevencia sa zaoberá jedincami s narušenou komunikačnou schopnosťou, aby u nich v dôsledku narušenej komunikačnej schopnosti nedochádzalo k problémom v socializácii.

V logopedickej praxi je náročné odlíšiť jednotlivé úrovne, pretože sa často diagnostika, terapia a prevencia prelínajú a nie je možné ich presne odlíšiť. Terapia bezprostredne nadväzuje na diagnostiku, je možnosť vzájomného prelínania takzvanej terapeutickkej diagnostiky, kedy už prvé kontakty logopéda s človekom s narušenou komunikačnou schopnosťou, môžu mať pozitívny terapeutický efekt. Ďalším prípadom je diagnostická terapia, ktorá sa vyskytuje pri komplikovanejších prípadoch narušenia komunikačnej schopnosti, kde sa diagnostika spresňuje v priebehu terapie a prelína sa s ňou. Terapia sa môže čiastočne prelínať aj s prevenciou. [7,13,18]

3. Dyslália

Dyslália je najčastejšie sa vyskytujúcou poruchou komunikačných schopností, vznikajúcou najčastejšie v priebehu predškolského veku. Zahrňuje artikuláciu jednej alebo viacerých hlások spôsobom, ktorý je zvukovo nápadný a odchyľuje sa od kodifikovanej formy výslovnosti. Tieto odchýlky sa označujú gréckymi písmenami s koncovkou a označením miesta alebo spôsobu odchýlenej artikulácie. Termínom dyslália sa označuje niekoľko prepojených úrovní. Hlavnou úrovňou je fonetická úroveň, ktorá dysláliu charakterizuje, vzťahuje sa na používanie jednotlivých hlások a predstavuje analytickú stránku reči. Na tejto úrovni sa porucha reči prejavuje vynechávaním hlások, ich zámenou za iné alebo nepresným vyslovovaním. Ďalšou úrovňou je fonologická, ktorá sa zaoberá používaním elementárnych hovorových znakov spojených do slabík, slov a viet. Tu sa poruchy reči prejavujú v plynulosti reči, kedy sú jednotlivé hlásky ovplyvnené predchádzajúcimi alebo nasledujúcimi.

Dyslália je vývojová vada, ktorá vzniká behom vývoja výslovnosti, a teda je najčastejšou poruchou komunikačnej schopnosti u detí. Pri vývoji dieťaťa a jeho komunikačných schopností, dochádza k napodobňovaniu počutých slov, čo môže viesť k ich nesprávnej reprodukcii. K nesprávnej reprodukcii dochádza ak dieťa nemá vytvorené presné akusticko-artikulačné okruhy. Postupne, s pribúdajúcim vekom výskyt nesprávnej výslovnosti klesá, príčinou je výuka čítania a písania, alebo vplyv logopedickej intervencie.

Dyslália sa vyskytuje častejšie u chlapcov ako u dievčat, v percentuálnom pomere 60:40%. Vyskytuje sa zhruba rovnako u detí inteligentných, ako aj u detí, ktoré navštevujú špeciálne školy. Dyslália sa väčšinou vyskytuje u jedincov v detskom veku, no nachádza sa aj u dospelých jedincov. Častým prípadom je výslovnosť hlások „R a Ľ“ a sykaviek u dospelých osôb.

Najčastejšími odchýlkami od kodifikovanej výslovnosti sú sigmatizmy, čo je odchylna artikulácia sykaviek a rotacizmy, teda odchylna artikulácia vibrantov, teda hlások „R a Ľ“. Frekventovaným závažným mechanizmom je velárny rotacizmus, kedy sa hláska tvorí pomocou vibrácie v hrdlovej oblasti, najčastejšie pomocou okraja mäkkého d'asna. Táto tvorba hlásky má za následok takzvané račkovanie. Iným mechanizmom je interdentalný sigmatizmus alebo addentalný sigmatizmus. V prípade prvého je pri tvorbe hlásky medzizubná poloha hrotu jazyka pri artikulácii sykavky, u druhého sa hrot jazyka dotýka alveol zubov, čo má za následok šušlanie. Medzi dyslalie patrí aj mogilalie, čo je porucha pri ktorej dieťa danú hlásku vynecháva. Ďalšou poruchou je paralélie, kedy dieťa danú hlásku zamieňa za inú. Vytvorené odchýlky v artikulácii hlások pretrvávajú po celý život jedinca, ak sa ich výslovnosť neupraví pomocou špecializovanej logopedickej starostlivosti.

3.1 Etiológia dyslálie

Etiológia, príčina vzniku dyslálie, sa rozdeľuje na funkčnú a organicky podmienenú.

U funkčnej dyslálie sú orgány, ktoré sa podieľajú na tvorbe reči bez poruchy. Porucha reči sa prejavuje motorickým typom, kedy sa vyskytuje neobratnosť rečových orgánov, alebo senzorickým typom. Senzorický typ sa vyskytuje v prípade nesprávneho vnímania a diferenciácie hovorových zvukov. Je to vývojový nedostatok pohybovej a sluchovej diferenciácie.

Organicky podmienená dyslália je spôsobená nedostatkami a zmenami na rečových orgánoch, taktiež ako následok porušenia sluchových dráh a CNS. Podľa lokalizácie ju delíme na akustickú (vady sluchu), labiálnu (defekty pier), dentálnu (defekty zubov), palatálnu (anomálie d'asna), lingválnu (anomálie jazyka), nasálnu (narušenie nasality). Jej príčiny rozdeľujeme na vnútorné a vonkajšie. Medzi vnútorné príčiny patrí porucha sluchu, nedostatočná diskriminácia zvukov, anatomické vady rečových orgánov, neuromotorické poruchy alebo kognitívno-lingvistické nedostatky. K vonkajším príčinám patria psychosociálne vplyvy.

3.2 Klasifikácia dyslálie

3.2.1 Delenie z vývojového hľadiska

U detí sa výslovnosť jednotlivých hlások zdokonaluje s pribúdajúcim vekom. V prvom rade sa dieťa naučí rozlišovať a pohybovo napodobňovať počuté zvuky, neskôr sa naučí presne zvládnuť zvukovú stránku reči, čo môže trvať u každého dieťaťa inú dobu. Fyziologická dyslália je nesprávna výslovnosť u detí do zhruba 5 roku, a považuje sa za fyziologickú, teda prirodzenú. Nastáva tiež predĺžená fyziologická dyslália, do zhruba 7 roku dieťaťa. Fyziologická dyslália sa upravuje vývojom diferenciačného útlmu, pomocou ktorého sa počuté slová analyzujú a vytvárajú správne akusticko-artikulačné spoje. Ak sa nesprávna výslovnosť hlások neupraví ani po 7 roku života dieťaťa, ide o chybnú výslovnosť, čo znamená, že je diagnostikovaná pravá dyslália.

Mogilália je chybou výslovnosti, kedy dieťa určitú hlásku v slovách vynecháva. Často namiesto hlásky používa neurčitý alebo nehláskový zvuk, čo môže pretrvávajúť až do 4 roku života dieťaťa. Ak táto nesprávna výslovnosť pokračuje aj po tomto roku života, jedná sa o vadu výslovnosti, ktorú označujeme predponou mogi- a príponou -izmus. Napríklad vynechávanie hlásky „R“ v reči sa nazýva mogirotacizmus. Inou chybou výslovnosti je paralália, ku ktorej dochádza, ak dieťa hlásku, ktorú ešte nedokáže vysloviť zamieňa za inú. Ak sa vada vyskytuje aj po určitom roku dieťaťa, kedy sa považuje vývoj výslovnosti danej hlásky za ukončený, označujeme ju predponou para- a príponou -izmus. Príkladom je zamieňanie hlásky „R“ za „L“ alebo „U“, takúto vadu nazývame pararotacizmus. Ak je hlásku tvorená iným spôsobom a na inom

mieste než odpovedá norme spisovného jazyka, jedná sa o vadu výslovnosti a označujeme ju príponou -izmus. Napríklad nesprávnej tvorby hlásky „K“ je kappacizmus, „L“ lambdacizmus, „R“ rotacizmus, „Ř“ rotacizmus bohemicus. Pri nahradzovaní správnej hlásky inou hláskou, existuje možnosť spontánneho osvojenia správnej výslovnosti, no pri chybné zafixovanom mechanizme tvorenia hlásky je nutná špeciálna starostlivosť logopéda.

3.2.2 Delenie podľa rozsahu

V tomto delení sa vyskytujú 3 typy dyslálie a to dyslália universalis, dyslália multiplex a parciálna dyslália.

Dyslália univerzalis (mnohopočetná) nastáva pri chybné výslovnosti väčšiny hlások. Najčastejšie dochádza k nahradzovaniu spoluhlások za hlásku „T“, čím sa reč stáva nezrozumiteľnou. U dyslálie multiplex je rozsah chybné vyslovovaných hlások menší v porovnaní s mnohopočetnou dysláliou, tým pádom je lepšia aj zrozumiteľnosť reči. U parciálnej dyslálie nastáva chyba výslovnosti u jednej alebo niekoľkých hlások. Delíme ju na monomorfnú, kedy chybná výslovnosť hlásky je z jednej artikulačnej oblasti a polymorfnú, chybná výslovnosť hlások z viacerých artikulačných oblastí.

3.2.3 Delenie z hľadiska kontextu

Z hľadiska kontextu sa vyskytuje hlásková a kontextová dyslália. U hláskovej sa týka jednotlivých hlások, u kontextovej jednotlivých slabík alebo slov. Pri kontextovej dyslálii sú izolované hlásky tvorené správne, no ich chybná výslovnosť nastáva v slabikách alebo slovách. Kontextová dyslália sa prejavuje nasledujúcimi symptómami : elízia (vynechávanie hlások), metatézia (prešmykovanie hlások), kontaminácia (zmiešavanie hlások), anaptixia (vkladanie hlások) a asimilácia (prispôsobovanie hlások).

3.3 Diagnostika dyslálie

Logopedické vyšetrenie je základom diagnostiky dyslálie. Pri diagnostike postupujeme tak, ako pri ostatných poruchách komunikačných schopností, a to nasledovne. Na začiatku je dôležité zhromaždenie anamnestických údajov, následne vyšetrenie sluchovej percie, fonematickej diferenciácie, celkovej motoriky ale aj jemnej motoriky hovorových orgánov. Dôležité je tiež vyšetrenie impresívnej a expresívnej zložky reči.

Diagnózu stanovuje zásadne logopéd, často v spolupráci s foniatrom a psychológom. Foniator určuje, či narušená komunikačná schopnosť nie je symptómom inej choroby. Diagnóza foniatra je často využívaná pri zložitejších poruchách artikulácie. Prácou psychológa je diagnostikovať a následne vyšetriť psychický stav dieťaťa.

Pri diagnostike dyslálie sa využíva základnej vyšetrovacej metódy, ktorou je rečová komunikácia, teda rozhovor s pacientom. Touto metódou sa získa predstava o súčasnom stave reči, a zistí sa rozsah a stupeň chybných výslovností. U detí je dôležitá prítomnosť rodičov poprípade zákonných zástupcov, a tiež prostredie logopedickej ambulancie musí byť príjemne zariadené. Rozhovor je zaznamenávaný mikrofónom, pre čisto hlasový záznam, poprípade videokamerou. Logopéd zaznamenáva nedostatky vo výslovnosti hlások do špeciálne vytvorených tabuliek, typických pre tieto vyšetrenia. Pri vyšetrení výslovnosti sa využívajú artikulačné testy a obrázkové testy. Môžu sa využiť aj prístrojové metódy ako sonograf a spektrograf, ktoré automaticky vyhodnocujú frekvenciu hlások pomocou frekvenčnej analýzy. Využívajú sa tiež špeciálne indikátory, S-indikátor a N-indikátor, ako aj počítačové metódy.

3.4 Terapia dyslálie

Terapia by mala začať u detí už v predškolskom veku, aby dieťa pred príchodom na základnú školu vedelo artikulovať všetky hlásky. Najzákladnejším predpokladom úspešnej terapie je vhodná motivácia dieťaťa, častou pomocou hry.

Platia štyri zásady priebehu terapie dyslálie podľa Seemana. Prvou je zásada krátkodobého cvičenia, kedy by sa výslovnosť mala precvičovať krátko ale čo najčastejšie. Druhou je zásada využitia sluchovej kontroly, ktorá kladie dôraz na sluchové vnímanie novo vytvorenej hlásky. Treťou je zásada používania pomocných hlások, sú to také ktoré už dieťa vie správne vyslovovať. Od nacvičovanej hlásky sa odlišujú zvukovo, ale sú si artikulačne podobné, napríklad pri vyvodzovaní hlásky „R“, sa používa ako pomocná hláska „D“. Poslednou zásadou je minimálna akcia, kedy by sa cvičenie malo robiť bez prehnaného úsilia a napätia artikulačných orgánov. Často sa artikulačné cvičenia začínajú šepotom, a až následne hlasitejšou rečou.

Vlastný priebeh terapie dyslálie sa delí na štyri základné etapy:

1. príprava cvičenia

Vykonávame ju k upresneniu artikulačných pohybov, k rozvoju motoriky artikulačných orgánov a taktiež k rozvoju schopnosti fonematickej diferenciácie, ktorá je dôležitá k schopnosti rozlišovať sluchom správnu a nesprávnu výslovnosť nacvičovanej hlásky. K prípravným cvičeniam patria cvičenia dychové, fonačné, cvičenia na rozvoj motoriky pier, sánky, jazyka, mäkkeho d'asna atď., podľa artikulačného miesta na ktorom sa precvičovaná hláska vytvára.

2. vyvodzovanie hlásky

Hlásky vyvodzujeme pomocou nepriamych a priamych metód. Nepriame sú cvičenia, pri ktorých sa napodobňujú prírodné alebo technické zvuky. Využívajú sa hlavne u menších detí. Pri priamych metódach navodzujeme tu hlásku, ktorú chceme precvičovať hlavne u starších detí.

3. fixácia výslovnosti hlásky

Ak sa podarí vyvodiť správnu výslovnosť hlásky, nasleduje fáza fixácie, teda upevnenia artikulačných stereotypov danej hlásky. Pri fixácii sa používa metódy nácviku spojovania vyvodenej hlásky s inými hláskami v najrôznejších spojeniach, obvykle v slabikách a slovách.

4. automatizácia hlásky

Táto záverečná etapa sa nesmie podceňovať alebo vynechať, pretože jej výsledkom je správna výslovnosť v spontánnej reči. Etapa sa začína jednoduchým opakovaním slov za pomoci zraku, teda pred zrkadlom a sluchu, napodobňovaním logopédovho vzoru. Následne pomenovanie rôznych obrázkov, predmetov, reprodukcia básničiek alebo čítanie. Nácvik sa ukončuje pri správnej výslovnosti hlások v monológu ale aj dialógu. [6,7,13,18]

4. Spracovanie signálov

4.1 Detekcia hraníc slov

Významným článkom v procese rozpoznávania reči je detekcia hraníc slov, ktorej cieľom je čo najpresnejšie rozdeliť vstupný signál na reč a šum. Dôležitou súčasťou je určenie hraníc, čo najrýchlejšie a najpresnejšie. Z toho dôvodu nemôže byť vstupný signál príliš dlhý, pretože bude výpočet dlho trvať, no ani príliš krátky. Čím kratší je signál tým sú okrajové časti vzoriek reči tesnejšie orezané, čo spôsobí nepresnejšiu detekciu práve týchto okrajových častí slov. Dôležité je zvoliť kompromis medzi rýchlosťou a účinnosťou detekcie.

Medzi základné metódy detekcie hraníc slov sa využívajú funkcie krátkodobej energie signálu, poprípade krátkodobé funkcie počtu prechodov nulou. Ich hlavnými výhodami je vysoká a výpočtovo nenáročná spoľahlivosť. Metóda spočíva v spracovávaní vstupného signálu, ktorý je rozdelený na malé časti, pričom pre každú časť sa vypočíta hodnota, ktorá závisí na type použitej funkcie. Ak hodnota prekročí stanovený prah, je príslušná časť vyhodnotená ako začiatok slova.

4.1.1 Krátkodobá energia signálu

Krátkodobá energia signálu sa využíva pre detekciu hraníc slov, ako bolo spomenuté v predchádzajúcej kapitole. Môže sa tiež využiť pre detekciu znelých hlások, ktoré sa vyznačujú väčšou energiou ako hlásky neznelé. Parameter krátkodobej energie sa často kombinuje s diferenciálnymi a akceleračnými koeficientmi, ktoré znižujú vplyv šumu prostredia, a je vyjadrený rovnicou

$$E = \sum_{n=0}^{N-1} s^2(n). \quad (4.1)$$

Krátkodobá energia signálu zvyrazňuje veľkú dynamiku reči, preto sa uplatňuje parameter intenzity daný rovnicou

$$E = \sum_{n=0}^{N-1} |s(n)|. \quad (4.2)$$

4.1.2 Počet priechodov nulou

Počet priechodov nulou rastie so zvyšujúcou sa frekvenciou. Tento parameter nie je výpočtovo náročný. Ak nastane zmena znamienka v časovej oblasti vo vzťahu daným rovnicou,

$$ZC = \frac{1}{2} \sum_{n=0}^{N-1} |sgn(x(n)) - sgn(x(n-1))|, \quad (4.3)$$

tak sa prechod zachytí pomocou funkcie

$$\text{sgn} = \begin{cases} 1 & \text{pre } x(n) \geq 0 \\ -1 & \text{pre } x(n) < 0 \end{cases} \quad (4.4)$$

4.2 Segmentácia

Dôležitou časťou predspracovania je segmentácia signálu reči. Keďže reč je nestacionárny náhodný signál, musíme ju pre následné spracovanie stacionarizovať. To je možné spraviť rozdelením rečového signálu na mikrosegmenty o určitej dĺžke trvania. Najčastejšie ju rozdelíme na segmenty o dĺžke 20 ms, pričom 10 ms časť sa využíva na prekryv, ktorý určuje periódu snímania segmentov a sme schopný zachytiť prechod medzi segmentami. Je možné ale tiež rozdelenie na 10 ms segmenty bez prekryvu, čo sa používa pri rýchlom oboznámení so signálom ako napr. zistenie šumu, znelosti fonémy. [4,16,17]

4.2.1 Oknové funkcie

Jednoduchou oknovou funkciou je pravouhlé okno, ktoré ma vo frekvenčnej reprezentácii veľké postranné laloky, čo je nevýhodou pri spracovaní rečového signálu. Jeho matematický zápis je

$$w(n) = \begin{cases} 1 & \text{pre } 0 \leq n \leq N - 1 \\ 0 & \text{pre } 0 > n > N - 1 \end{cases} \quad (4.5)$$

V procese segmentácie sme ako keby vstupný signál prenasobili pravouhlým oknom, čím sme do segmentu vniesli aj jeho frekvenčné skreslenie. Pre odstránenie tohto skreslenia použijeme inú oknovú funkciu, ktorá sa využíva v spracovaní rečového signálu, a to najčastejšie Hammingove, Hannove, Kaiserove alebo Bartlettovo okno. Matematický zápis Hammingovho okna je

$$w(n) = 0,54 + 0,46 \cos\left(\frac{2\pi n}{N-1}\right) \quad \text{pre } 0 \leq n \leq N - 1. \quad (4.6)$$

[16,17]

4.3 Preemfáza

Rečové ústrojenstvo vo všeobecnosti utlmuje vysoké frekvencie, čo na rečový signál pôsobí ako DP filter. Použijeme teda operáciu, ktorá vyrovná kmitočtovú charakteristiku rečového signálu, a teda HP FIR filter s konečnou impulzovou odozvou. Jeho prenosová funkcia je

$$H(z) = 1 - \lambda z^{-1} \quad , \quad (4.7)$$

kde $\lambda \sim 1$. [9]

5. Metódy pre analýzu rečových signálov

5.1 Spektrálna analýza

Táto kapitola popisuje súčasne používané prístupy pre analýzu akustických signálov, a teda aj ľudskej reči. Digitálne spracovanie a analýza reči je široko rozvinutý obor, ktorý má veľké množstvo nástrojov a algoritmov. Preto uvedieme len dva najrozšírenejšie používané prístupy, a to Fourierovu analýzu a Vlnkovú analýzu. Na základe tejto analýzy a jej výsledkov, môžeme signál klasifikovať a následne predspracovať pre efektívnejšie rozpoznávanie.

5.1.1 Furierova analýza

Fourierova analýza patrí medzi najbežnejšie nástroje pre spracovanie signálov. Jej počiatky siahajú do 19. storočia, do prác francúzskeho matematika a fyzika Josepha Fouriera. Základnou myšlienkou Fourierových radov je, že každá periodická funkcia sa dá rozložiť do nekonečného radu sínusov a kosínusov, teda harmonických priebehov. Takéto rozloženie poskytuje informáciu o frekvenčných zložkách skúmaného signálu. Z frekvenčného spektra vstupného zvukového signálu vieme pomocou spektrálnej analýzy získať súbor frekvenčných zložiek, z ktorých je daný signál zložený.

Spojité Fourierova transformácia

Spojité Fourierova transformácia (CFT) bola vytvorená pre využitie Fourierových radov v praxi, keďže Fourierove rady sú definované pre periodické a nekonečné signály, v praxi sú signály neperiodické a konečné. Fourierova transformácia transformuje signál z časovej oblasti do frekvenčnej. Matematicky môžeme CFT zapísať ako integrál

$$F(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt. \quad (5.1)$$

Výsledok transformácie $F(\omega)$ označujeme ako spektrálnu funkciu, teda spektrum vstupného signálu $x(t)$, kde ω je uhlová rýchlosť, t je čas a j imaginárna jednotka. Výsledkom je teda komplexná funkcia, z ktorej možno získať amplitúdové spektrum ako $|F(\omega)|$ a fázové spektrum ako $\arg(F(\omega))$.

Diskrétna Fourierova transformácia

Diskrétna Fourierova transformácia (DFT) je Fourierova transformácia aplikovaná na postupnosť komplexných čísel. DFT teda slúži na prevod z diskkrétnej časovej oblasti do diskkrétnej frekvenčnej oblasti. Matematicky DFT zapisujeme ako

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}nk}, \quad (5.2)$$

kde $X(k)$ je aktuálne frekvenčné spektrum, $x(n)$ je navzorkovaný vstupný signál, teda postupnosť komplexných čísel s N prvkami a $k = 0, 1 \dots N-1$.

Keďže DFT má vysokú výpočtovú náročnosť sa v praxi využíva rýchla Fourierova transformácia FFT, ktorá má nízku výpočtovú náročnosť. Výpočtová náročnosť FFT je $O(N \log N)$ oproti výpočtovej náročnosti $O(2N)$ pri DFT. Najpoužívanejším variantom je FFT je algoritmus Cooley-Turkey, kde dochádza k rekurzívnemu deleniu transformácie v každom kroku na dve menšie transformácie N_1 a N_2 s rozmerom $\frac{N}{2}$.

Diskrétna kosínusová transformácia

Obdobou DFT je diskrétna kosínusová transformácia (DCT), ktorej výstupy sú len reálne koeficienty. Existuje 8 variantov DCT, no bežne sa používajú iba 4. Najpoužívanejší je druhý variant, teda DCT-II, matematicky zapísaný ako

$$X(k) = \sum_{n=0}^{N-1} h(n)x(n)\cos\left(\frac{(2n+1)k\pi}{2N}\right), \quad (5.3)$$

kde $k = 0, 1 \dots N-1$. Jej výhodou je menšia redundancia oproti DFT. [4,9,16,17,24]

5.1.2 Vlnková analýza

Hlavnou nevýhodou Fourierovej transformácie je šírka okna, ktoré sa zvolí na začiatku a počas celej transformácie sa nemení. Preto bola v osemdesiatych rokoch minulého storočia vyvinutá vlnková transformácia, ktorá je schopná meniť veľkosť okna dynamicky počas transformácie, čím lokalizuje frekvenčné spektrum v čase s väčšou presnosťou. U vlnkovej transformácie je tiež rôzna rozlišovacia schopnosť v závislosti od frekvencie. Platí, že má lepšie frekvenčné rozlíšenie pre nízke frekvencie, teda vie presnejšie stanoviť hodnotu frekvencie ako polohu v čase. Naopak pri vysokých frekvenciách lepšie stanoví polohu v čase ako hodnotu frekvencie. Táto vlastnosť vyplynula z väčšiny reálnych signálov, ktoré obsahujú dlhotrvajúce nízkofrekvenčné zložky a krátkodobé zámity vyšších frekvencií.

Spojité vlnková transformácia

U spojitej vlnkovej transformácie (CWT) sa stretáme s označením materská vlnka, ktorá plní funkciu okna a označuje sa symbolom ψ . Dôležitou vlastnosťou materskej vlnky je jej konečné trvanie v čase a nulová stredná hodnota. Existuje mnoho druhov materských vlniek, ako Haarova a Morletova, ich použitie sa líši podľa typu aplikácie.

Princípom CWT je posun materskej vlnky cez celé trvanie signálu, a súčasne porovnáваме materskú vlnku so skúmaným signálom. V ďalších krokoch materskú vlnku dilatujeme a posúvame

v čase a celý proces opakujeme. Tieto dilatované a posunuté materské vlny sa nazývajú dcérske vlnky. Dilatáciu a posuv môžeme matematicky vyjadriť rovnicou

$$\psi^*(t, \tau, a) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right), \quad (5.4)$$

ψ^* je dcérska vlnka s posuvom τ a dilatáciou a , tzv. scaling factor. Následne je CWT definovaná ako

$$CWT(\tau, a) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) dt, \quad (5.5)$$

kde $x(t)$ je skúmaný signál.

Po priebehu transformácie sa výsledné koeficienty môžu zakresliť do grafu, v ktorom na jednej osi je čas a na druhej scaling factor a . Takýmto grafom je spektrogram.

Diskrétna vlnková transformácia

Nevýhodou CWT je jej veľká výpočtová náročnosť, keďže parametre τ aj a môžu nadobúdať všetky reálne hodnoty. To sa javí ako zbytočné, pretože reálne skúmané signály majú konečnú dĺžku trvania. Diskrétna vlnková transformácia (DWT) pracuje s radom filtrov. Tieto filtre sa rozdeľujú na filter typu dolnej priepusti (DP), ktorý prepúšťa len nízkofrekvenčné zložky a filter typu hornej priepusti (HP), ktorý naopak prepusti iba vysokofrekvenčné zložky. Filtre majú tiež v DWT špeciálne označenie, DP sa označuje ako aproximácia a HP ako detail.

Princípom DWT je filtrácia vstupného signálu na dve vetvy pomocou HP a DP. Získame teda vzorky aproximácie a vzorky detailov, ktoré sa označujú ako koeficienty DWT. Týmto spôsobom však pri každej filtrácii vzniká dvojnásobný počet vzoriek, pretože sa rozdelí iba frekvenčné spektrum a nie počet vzoriek. Pretože dvojnásobný počet vzoriek je nepotrebný, polovicu vzoriek z aproximácie aj detailu môžeme zahodiť a nestratíme žiadne frekvenčné informácie. Následne sa filtruje už len vetva DP, teda koeficienty aproximácie, tým istým spôsobom ako v prvom kroku. Tento proces môžeme opakovať, až kým nezostane jeden koeficient aproximácie a jeden koeficient detailu. Frekvenčné spektrum signálu vieme zostrojiť z jednotlivých detailov v každom stupni filtrácie. Pôvodný signál, naopak vieme zostrojiť z aproximácií v každom stupni. Tento zrekonštruovaný signál však bude obsahovať iba polovicu celkových vzoriek vstupného signálu, čo nám v konečnom dôsledku šetrí kapacitu diskovej jednotky. [9,16,17]

5.2 Kepstrálna analýza

Kepstrálna analýza vychádza z frekvenčnej analýzy a patrí medzi metódy nelineárneho spracovania rečových signálov. Pre zaznamenanie zmeny rečového signálu v spektrálnej oblasti, musíme analyzovať takú dĺžku segmentu, u ktorej predpokladáme, že signál bude stacionárny.

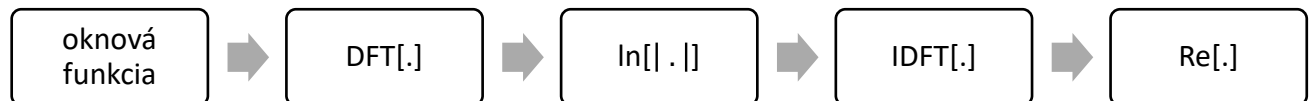
Diskrétno spektrum signálu $X(k)$ môžeme získať pomocou DFT

$$X(k) = DFT\{x(n)\} = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}kn}, \quad (5.6)$$

kde premenná k je frekvenčný index nadobúdajúci diskkrétne hodnoty $0, 1, \dots, N - 1$ kde N je počet bodov algoritmu DFT. Princípom keprálnej analýzy je získanie reálneho kepru, ktoré získame ako reálnu časť spätnej DFT prirodzeného logaritmu modulu spektrálnej funkcie $X(k)$

$$c(n) = \text{Re}\{DFT^{-1}[\ln|X(k)|]\}. \quad (5.7)$$

V keprálnej analýze sa z frekvenčnej oblasti vraciame opäť do časovej oblasti. [16]



Obr. 5.1: Blokové schéma výpočtu reálneho kepru

5.3 Lineárne prediktívna analýza

Lineárne prediktívna analýza, ktorá bola známa už v sedemdesiatych rokoch minulého storočia, je jednou z najpoužívanejších techník spracovania rečového signálu, označuje sa aj ako LPC koeficienty. Metóda je založená na modelovaní rečového ústrojenstva, a jej hlavnou výhodou je nízka výpočtová náročnosť a spätná rekonštrukcia spracovaného signálu.

Princíp LPC vychádza z predpokladu, že rečový signál je zložený zo stacionárnych úsekov určitej dĺžky, a môžeme predpovedať všetky nasledujúce úseky na základe lineárnej kombinácie predchádzajúcich stacionárnych úsekov. Túto definíciu matematicky zapíšeme ako

$$s(k) = \sum_{i=1}^M a_i s(k - i), \quad (5.8)$$

kde $s(k)$ je k -tá vzorka, a_i sú LPC koeficienty a M vyjadruje počet koeficientov. Neznáme koeficienty a_i sa vypočítajú pomocou minimalizácie chyby predikcie, ktorá je na úseku n rovná

$$E_n = \sum_n [s_n(k) - \tilde{s}_n(k)]^2, \quad (5.9)$$

kde

$$\tilde{s}_n(k) = -\sum_{i=1}^M a_i s(k - i). \quad (5.10)$$

Neznáme koeficienty a_i získame minimalizáciou tohto výrazu na stacionárnom úseku reči parciálnou deriváciou chybovej funkcie podľa koeficientov a_i , ktorú položíme rovnú nule

$$\frac{\delta}{\delta a_i} [E_n] = 0. \quad (5.11)$$

Touto parciálnou deriváciou dostaneme M lineárnych rovníc, ktorých výpočtom získame koeficienty lineárnej predikcie a_i . Výpočet sa realizuje najčastejšie pomocou korelačnej metódy, kedy na sústavu rovníc použijeme krátkodobú autokorelačnú funkciu a následne Durbin-

Levinsonov iteračný algoritmus. Korelačná metóda sa využíva pre svoju nižšiu výpočtovú náročnosť. Výpočet autokorelačných koeficientov pre kladné k a signál dĺžky N je popísaný rovnicou

$$R(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} s(n)s(n+k), \quad (5.12)$$

kde autokorelačná funkcia je $R(k)$ a $s(n)$ je vzorka signálu v čase n . Podobnosť signálu v čase n so signálom v čase $n+k$ je reprezentovaná výslednými korelačnými koeficientmi. Táto sústava rovníc sa môže prepísať do maticového tvaru a následne sa použije rýchly Durbin-Levinsonov algoritmus

$$E_n^{(0)} = R_n(0), \quad (5.13)$$

$$k_i = \frac{-[R_n(i) + \sum_{j=1}^{i-1} a_j^{(i-1)} R_n(i-j)]}{E_n^{(i-1)}}, \quad (5.14)$$

$$a_i^{(i)} = k_i, \quad (5.15)$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)}, \quad (5.16)$$

kde $j = 1, 2, \dots, i-1$.

$$E_n^i = (1 - k_i^2) E_n^{(i-1)}, \quad (5.17)$$

Pre získanie LPC koeficientov sa zredukovalo množstvo dát. Získané LPC koeficienty veľmi dobre opisujú spracovávaný signál. [14,16,24]

5.4 Perceptívne lineárne predikčné koeficienty

Percepčne lineárna prediktívna analýza, taktiež nazývaná ako PLP je ďalšia metóda pre spracovanie rečového signálu. Spektrálne parametre reprezentujú tie charakteristiky rečového signálu, ktoré sa najviac podobajú charakteristikám sluchového orgánu človeka. Percepčne významné charakteristiky rečového signálu získame metódou spektrálnej analýzy na nelineárnej frekvenčnej stupnici. PLP koeficienty sa najčastejšie využívajú u automatických systémoch rozpoznávania reči nezávislých od rečníka.

Výpočet PLP koeficientov pozostáva z postupu určitých základných operácií. Prvou operáciou je použitie DFT na segment reči, a teda pretransformovanie rečového signálu do frekvenčnej oblasti a vypočítanie jeho výkonového spektra. Túto operáciou popisuje rovnica

$$P(\omega) = \text{Re}[S(\omega)]^2 + \text{Im}[S(\omega)]^2 . \quad (5.18)$$

Nasledujúcou operáciou je pretransformovanie frekvenčnej stupnice na Barkovu stupnicu podľa vzťahu

$$\Omega(\omega) = 6 \ln \left(\frac{\omega}{1200\pi} + \sqrt{\left(\frac{\omega}{1200\pi}\right)^2 + 1} \right), \quad (5.19)$$

Ďalšou operáciou je konvolúcia výkonového spektra $P(\omega)$ a výkonového spektra $\psi(\omega)$ podľa vzťahu

$$\theta(\Omega_i) = \sum_{\Omega=-1,3}^{2,5} P(\Omega - \Omega_i) \psi(\Omega), \quad (5.20)$$

kde ψ je definovaná ako

$$\psi(\omega) = \begin{cases} 0 & \text{pre } \Omega < -1.3 \\ 10^{2.5(\Omega+0.5)} & \text{pre } -1.3 < \Omega < -0.5 \\ 1 & \text{pre } -0.5 < \Omega < 0.5 \\ 10^{-1(\Omega-0.5)} & \text{pre } 0.5 < \Omega < 2.5 \\ 0 & \text{pre } \Omega > 2.5 \end{cases} . \quad (5.21)$$

Nasleduje prevzorkovanie výkonového spektra $\theta(\Omega_i)$ na nižšiu frekvenciu, pri použití osemnástich spektrálnych vzorkov v pásme od 0 po 5 kHz s krokom 0,994 Barka. Prispôsobenie spektra $\theta(\Omega_i)$ citlivosti ucha, urobíme tak, že vynásobíme spektrum $\theta(\Omega_i)$ prenosovou funkciou filtra $E(\omega)$, ktorá ma tvar

$$E(\omega) = \frac{[(\omega^2 + 56,8 \times 10^6)\omega^4]}{[(\omega^2 + 6,3 \times 10^6)^2 \times (\omega^2 + 0,38 \times 10^9)]} . \quad (5.22)$$

Následne sa použije operácia pre kompresiu amplitúdy spektra podľa vzťahu

$$\theta(\Omega) = [\Xi(\Xi_b)]^{0,33} . \quad (5.23)$$

Poslednými operáciami sú inverzná DFT a LP analýza, po ktorých prevedení získame PLP koeficienty. [14,16,24]

5.5 MEL-frekvenčné kepstrálne koeficienty

V súčasnosti sa pre spracovanie rečového signálu používa okrem LPC aj Mel-frekvenčné kepstrálne koeficienty (MFCC). Výpočet MFCC je založený na využití zovšeobecneného princípu superpozície, a teda vychádza z homomorfného spracovania signálov. Snahou MFCC je upraviť signál tak, aby sa zvýraznili nižšie frekvencie, takže rešpektuje nelineárne vlastnosti vnímania zvuku ľudským uchom. Táto skutočnosť sa realizuje zavedením banky filtrov s trojuholníkovými priepustnými filterami rovnomerne rozloženými na melovskej frekvenčnej škále. Pre prepočet na melovskú frekvenčnú škálu platí vzťah

$$m = 2596 \log_{10} \left(1 + \frac{f}{700} \right), \quad (5.24)$$

kde f v [Hz] je kmitočet v lineárnej škále a m v [mel] je kmitočet v nelineárnej melovskej škále.

Pre prepočet z melovskej škály do frekvenčnej škály v [Hz] sa využíva vzťah

$$f = 700[\exp(0,887 \times 10^{-3}m) - 1]. \quad (5.25)$$

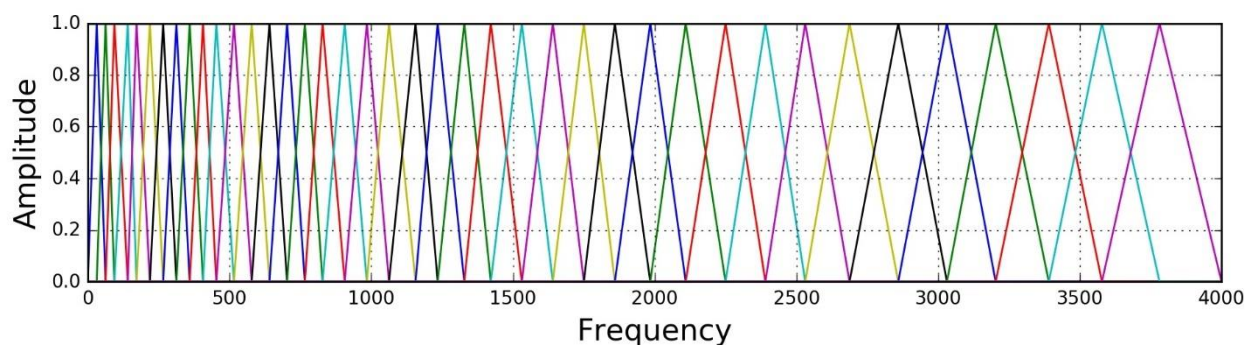
Melovskou škálou teda odlinearizujeme pôvodné frekvencie a zvýraznia sa frekvencie charakteristické pre ľudské ucho.

Princíp výpočtu MFCC koeficientov je založený na postupe využívajúcom viacerých metód. Na vstup systému sa privedú vzorky rečového signálu, ktoré vstupujú na filter preemfázy aby sa zvýraznili vyššie frekvenčné zložky signálu, a signál sa rozdelí na prekrývajúce sa časové úseky konštantnej dĺžky. Po takto upravenom signály sa naň aplikuje najčastejšie Hammingovo okno. Ďalším krokom je transformácia signálu pomocou DFT do spektrálnej oblasti a vypočíta sa výkonové spektrum.

Najdôležitejším krokom je melovská filtrácia, ktorá sa realizuje pomocou banky trojuholníkových pásmových filtrov pozdĺž melovskej frekvenčnej osy. Trojuholníkové filtre sú rozložené cez celé frekvenčné pásmo, teda od nuly až po Nyquistovu frekvenciu. Pre stredné frekvencie jednotlivých filtrov platí vzťah

$$b_{m,i} = b_{m,i-1} + \Delta_m, \quad (5.26)$$

kde $b_{m,i}, i = 1, 2, \dots, M^*$ a $\Delta_m = B_{mw}/(M^* + 1)$. Odozvy jednotlivých filtrov majú tvar rovnoramenných trojuholníkov a pre ich výpočet sa musia všetky koeficienty DFT prepočítať do melovskej frekvenčnej škály podľa vzťahu (5.19). Vhodnejším spôsobom je vyjadrenie trojuholníkových filtrov vo frekvenčnej škále v [Hz], pričom nie je potrebný prepočet koeficientov DFT. Trojuholníkové filtre môžeme vidieť na nasledujúcom obrázku 5.2.



Obr. 5.2: Trojuholníkové filtre pozdĺž frekvenčnej osy v [Hz]

Následne sa prepočítajú všetky stredné frekvencie $b_{m,i}$ s využitím vzťahu (5.20) na stredné frekvencie v [Hz] $b_i, i = 1, 2, \dots, M^* + 1$. Odozvy jednotlivých filtrov sú definované vzťahom

$$y_m(i) = \sum_{f=b_{i-1}}^{b_{i+1}} |S(f)| u(f, i), \quad i = 1, 2, \dots, M^*, \quad (5.27)$$

kde f sú frekvencie vyberané zo súboru frekvencií pri výpočte DFT a $u(f, i)$ je trojuholníkový filter. Pre vyjadrenie trojuholníkového filtra sa využíva vzťah

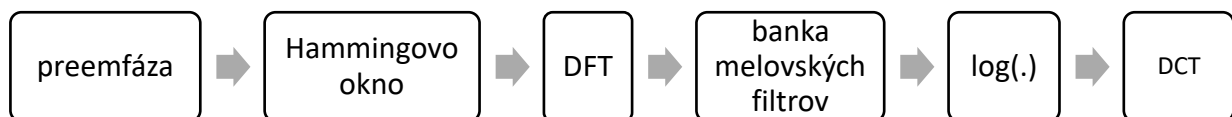
$$u(f, i) = \begin{cases} \frac{1}{b_i - b_{i-1}} (f - b_{i-1}) & \text{pre } b_{i-1} \leq f < b_i \\ \frac{1}{b_i - b_{i+1}} (f - b_{i+1}) & \text{pre } b_i \leq f < b_{i+1} \\ 0 & \text{pre ostatné} \end{cases} \quad (5.28)$$

Každý koeficient DFT sa vynásobí koeficientom pásmového filtra pre príslušný kanál a navzájom sa sčítajú. Následne sa výstupy $y_m(i)$ jednotlivých filtrov prevedú do logaritmickkej oblasti, čím sa do procesu spracovania vkladá dôležitý príznak kepstrálnej analýzy. Posledným krokom pre výpočet MFCC koeficientov $\{c_m(j)\}_{j=1}^M$ je výpočet pomocou DCT,

$$c_m(j) = \sum_{i=1}^{M^*} \log y_m(i) \cos\left(\frac{\pi j}{M^*} (i - 0,5)\right) \quad , \text{ pre } j = 0, 1, \dots, M, \quad (5.29)$$

kde M^* je počet pásiem filtra a M je počet melovských kepstrálnych koeficientov. [10,15,24]

Bloková schéma postupu výpočtu MFCC je na nasledujúcom obrázku.



Obr. 5.3: Blokové schéma výpočtu MFCC

6. Klasifikátory

Pojem klasifikácia znamená triedenie alebo hodnotenie. V prípade klasifikácie logopedických vád je to klasifikácia do dvoch tried, a to so správnou výslovnosťou alebo nesprávnou výslovnosťou na základe určitých koeficientov. Koeficienty môžu byť rôzne no najčastejšie sa využívajú PLP, PLC alebo MFCC popísané v predchádzajúcich kapitolách.

6.1 Algoritmus k-najbližších susedov

Algoritmus k-najbližších susedov (K-NN) pracuje na princípe, v ktorom sú vzory klasifikované podľa ich najbližších susedov do odpovedajúcich si tried. Jedná sa o jeden z najpopulárnejších klasických prístupov zhlukovej analýzy. Samotný model je v podstate tvorený tréningovou množinou a proces učenia spočíva v uschovaní vzorov v modeli. Vstupné atribúty musia byť numerické, aby bolo možné pre každé dva vzory vypočítať ich vzájomnú vzdialenosť. Vzorky z tréningovej množiny majú n číselných atribútov a každá jedna vzorka reprezentuje bod v N -rozmernom priestore. Ak chce klasifikátor určiť cieľový atribút neznámej vzorky, hľadá v priestore k vzoriek z tréningovej množiny tie, ktoré sú najbližšie danej neznámej vzorky pomocou miery. Mierou je vzdialenosť Euklidovská alebo Manhattanská.

Tréningovú množinu definujeme ako

$$\{(X_i, \varpi_i)\}_{i=1, \dots, K}, \quad (6.1)$$

kde X_i je vzorok ku ktorému je priradená trieda ϖ_i a K je veľkosť tréningovej množiny. Pre neznámy vzorok x hľadáme x_k také, aby platil vzťah

$$\|x_k - x\| = \min \|x_i - x\|_{i=1, \dots, K}. \quad (6.2)$$

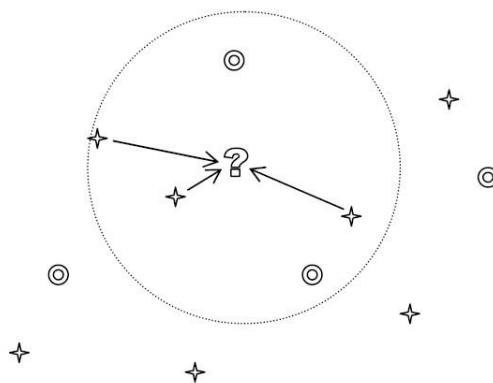
Následne vzorok x zaradíme do rovnakej triedy, do ktorej patrí x_k .

Pre výpočet vzdialenosti sa najčastejšie používa Euklidovská vzdialenosť definovaná vzťahom

$$d(x, y) = \sqrt{\sum_{i=1}^P (x_i - y_i)^2}, \quad (6.3)$$

kde d je hľadaná vzdialenosť a x_i sú vzorky tréningovej množiny, y_i je neznámy vzorok.

Postup klasifikácie je jednoduchý a definovaný v dvoch krokoch. Prvým krokom je vytvorenie hypergule okolo neznámej vzorky, ktorá obsahuje práve k vzorkov tréningovej množiny. Druhým krokom pozostáva z klasifikácie vzorky do tej triedy, ktorá je v hypergule zastúpená najväčším počtom vzorkov (obr.6.1).



Obr. 6.1: Klasifikácia K-NN do dvoch tried

Pri použití klasifikátora K-NN je nutné zvoliť vhodnú hodnotu pre k , kedy sa pri klasifikácii do dvoch tried zvolí k nepárne. Nevýhodou tohto klasifikátora je počítanie vzdialenosti pri každej klasifikácii, čo môže značne spomaliť proces a tiež kladie nemalé kapacitné nároky pre uschovanie modelu. [3]

6.2 Gaussov zmiešaný model

Klasifikátor na báze Gaussovského zmiešaného modelu (GMM) patrí medzi metódy využívajúce štatistické rozpoznávanie vzorov, ktorý vychádza z predpokladu, že modely rovnakých tried môžu mať podobné štatistické vlastnosti. Prebieha v ňom učenie bez učiteľa a dáta obsiahnuté v tréningovej množine nenesú informácie ohľadom príslušnosti k určitej triede.

Princíp GMM modelovania je založený na lineárnej kombinácii dielčích Gaussovských funkcií, kedy je každá funkcia definovaná svojou strednou hodnotou μ , smerodatnou odchýlkou σ a váhou π . Keďže sa obvykle pracuje s viacrozmernými vzájomne korelovanými dátami, je smerodatná odchýlka nahradená kovariančnou maticou Σ . Takúto viacrozmernú Gaussovskú funkciu zapisujeme ako

$$\eta(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} e^{-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}}, \quad (6.4)$$

kde P je rozmer funkcie rozdelenia pravdepodobnosti, x je vektor príznakov. Následne z Bayersovej teórie podmienenej pravdepodobnosti je možné, že lineárnou kombináciou takýchto funkcií vznikne GMM charakterizovaný vzťahom

$$f(x) = \sum_{i=1}^M a_i \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_i)}} e^{-\frac{(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)}{2}}, \quad (6.5)$$

kde M označuje počet Gaussovských funkcií rozdelenia pravdepodobnosti, a je váhovací parameter a d je počet príznakov. Rozptyl hodnôt použitých príznakov určuje kovariančná matica Σ o rozmere $d \times d$. Súčasne však musí platiť

$$\sum_{i=1}^M a_i = 1. \quad (6.6)$$

Hodnoty rozptylu pre jednotlivé x_i sú obsiahnuté na hlavnej diagonále a na vedľajšej diagonále sú prvky, ktoré vyjadrujú mieru korelácie medzi príznakmi.

V prípade, že je rečový signál spracovávaný súčasne vo viacerých prúdoch, a teda že rovnaký úsek signálu je popísaný viacerými príznakovými vektormi, môžeme rovnicu (6.5) upraviť do tvaru

$$p(x) = \prod_{r=1}^R \left[\sum_{i=1}^M a_i \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_i)}} e^{-\frac{(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)}{2}} \right]^{g_r}. \quad (6.7)$$

Exponent g_r vyjadruje váhu, ktorá odpovedá r -tému dátovému prúdu.

Klasifikačná funkcia GMM klasifikátoru má tvar

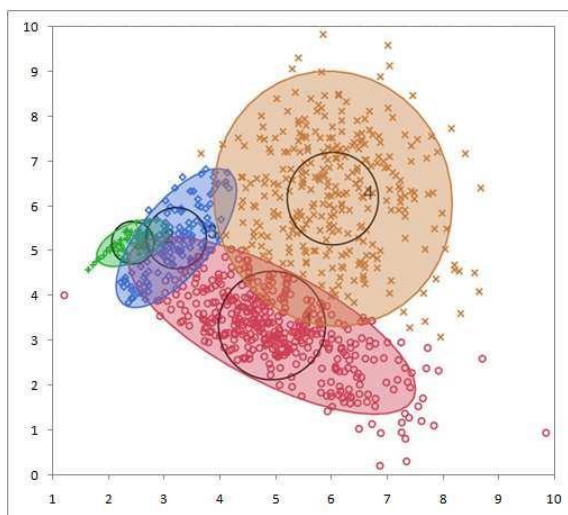
$$p(x) = \arg \max_C \sum_{i=1}^M a_i^C \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_i^C)}} e^{-\frac{(x-\mu_i^C)^T \Sigma_i^{C-1} (x-\mu_i^C)}{2}}, \quad (6.8)$$

kde hľadáme maximum zo všetkých C tried. V praktických aplikáciách sú klasifikátory často doplnené o prahovaciu konštantu b , ktorá určuje minimálnu potrebnú úroveň pre klasifikáciu.

Parametre zmiešaného modelu je možné z množiny pozorovaní odhadnúť pomocou EM algoritmu, čo v angličtine znamená Expectation Maximization. Tento algoritmus sa využíva pre tréovanie modelu na množine tréovacích dát a pozostáva z nasledujúcich krokov:

1. Inicializácia parametrov rozdelenia pravdepodobnosti a ohodnotenie inicializovaných hodnôt pomocou logaritmickej vierohodnosti.
2. Ohodnotenie zodpovednosti (váhové faktory pre každú vzorku) pomocou nastavených hodnôt parametrov.
3. Nový odhad parametrov s použitím zodpovedností nájdených v kroku 2.
4. Opakovanie krokov 2 a 3, kým sa hodnota logaritmickej vierohodnosti nemení, alebo je menšia ako zadaná prahová hodnota.

Výsledkom je pravdepodobnostné zaradenie jednotlivých pozorovaní do zhlukov. Algoritmus však naráža na problém určenia počtu zložiek GMM a štartovacieho bodu algoritmu, keďže pre odhad neexistuje jednoznačná metóda. [15,22]



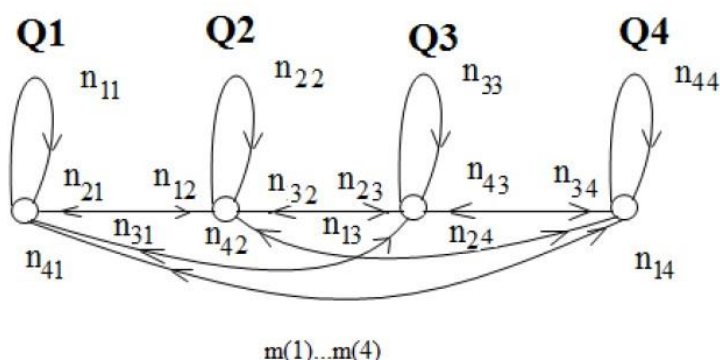
Obr. 6.2: Výsledné zhľuky MFCC pomocou GMM

6.3 Markovove modely

Najúspešnejšie metódy rozpoznávanie reči sú založené na štatistických metódach rozpoznávania, medzi ktoré patria aj skryté Markovové modely (HMM). V prípade použitia HMM predpokladáme, že sa charakter reči počas krátkeho časového úseku nemení, aj keď už vieme, že reč je nestacionárny proces. Metóda spočíva v predpoklade, že musí prísť k zmene rečového traktu pri prechode medzi dvoma rečovými jednotkami. Pozostáva z Markovových modelov, kde je každá rečová jednotka reprezentovaná jedným modelom. Najvhodnejším typom HMM pre modelovanie reči je ľavo-pravá HMM.

6.3.1 Markovove modely

Pravdepodobnostné modely, ktoré reprezentujú Markovove modely pozostávajú z uzlov Q_i a prechodmi medzi uzlami n_{ij} pre $i = 1, 2, \dots, N$ a $j = 1, 2, \dots, M$. Väzby medzi uzlami sú vyjadrené maticou pravdepodobnostných prechodov N a hodnoty uzlov maticou pravdepodobností generovaných vzorov M .



Obr. 6.3: Uzly Markovovho modelu s prechodmi

V prvej matici jednotlivé prvky popisujú pravdepodobnosť, že sa model z určitého uzla môže dostať do uzlov iných. Daný je aj vektor pravdepodobnosti počiatočných stavov π , ktorý vyjadruje počiatočný stav modelu. Takto definovaný Markovov model je daný vzťahom

$$\lambda = (N, M, \pi). \quad (6.9)$$

HMM teda spočíva vo vytvorení príslušných modelov, ktoré následne budeme porovnávať s neznámym slovom. Pre každé neznáme slovo sa vyčíslí pravdepodobnosť zhody tohto slova so všetkými vytvorenými modelmi, následne model s najväčšou dosiahnutou pravdepodobnosťou zhody s neznámym slovom určí akustickú reprezentáciu slova, a teda jeho rozpoznanie. V procese rozpoznania, či neznáme slovo bolo generované práve tým modelom sa využívajú dva algoritmy. Prvým je dopredný algoritmus a druhým je spätný algoritmus.

Dopredný algoritmus

Premennú $\alpha_t(i)$ definujeme ako pravdepodobnosť generovania čiastočnej postupnosti $\{o_1, \dots, o_t\}$ a stavu $q(t) = q_t$ na danom modeli λ , ktorý má N stavov a pozorovali sme T vektorov príznakov

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q(t) = q_i | \lambda), \quad (6.10)$$

kde hodnoty $\alpha_t(i)$ vypočítame rekurzívne:

Inicializácia

$$\alpha_1(i) = \pi_i b_i(o_1) \text{ pre } i = 1, 2, \dots, N. \quad (6.11)$$

Rekurzia pre $t = 1, 2, \dots, T - 1$ vyjadrená pomocou

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(\alpha_{ij}) \right] b_j(o_{t+1}) \text{ pre } i = 1, 2, \dots, N. \quad (6.12)$$

Hodnota výslednej pravdepodobnosti

$$P(O | \lambda) = \sum_{i=1}^N \alpha_t(i). \quad (6.13)$$

Spätný algoritmus

Pri spätnom algoritme je definovaná premenná

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T, q(t) = q_i | \lambda). \quad (6.14)$$

Inicializácia

$$\beta_T(i) = 1 \text{ pre } i = 1, 2, \dots, N. \quad (6.15)$$

Rekurzia pre $t = T - 1, T - 2, \dots, 1$. vyjadrená pomocou

$$\beta_t(i) = \sum_{i=1}^N \pi_i b_i(o_1) \beta_1(i) . \quad (6.16)$$

Hodnota výslednej pravdepodobnosti

$$P(O|\lambda) = \sum_{i=1}^N \pi_i b_i(o_1) \beta_1(i) . \quad (6.17)$$

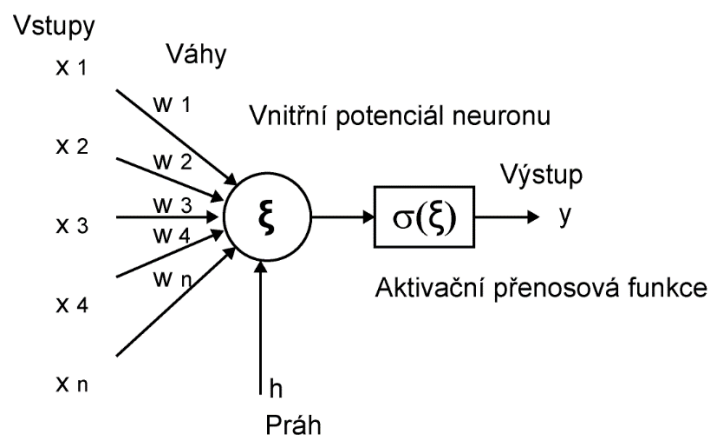
[2]

6.4 Neurónové siete

Neurónové siete patria medzi najmladšie so súčasných vedných disciplín. Vychádzajú z biologických neurónových sietí, z ktorých si prevzali veľmi zjednodušený matematický model. Využívajú sa pre klasifikáciu vstupných dát, princípom identifikácie známeho vzoru na vstupe a jeho následným rozpoznáním. Práve preto patria neurónové siete k veľmi efektívnej metóde využívanej pre rozpoznávanie reči.

6.4.1 Matematický model neurónu

Matematický model neurónu je odvodený od biologického neurónu, má viacero vstupov no len jeden výstup. Množina vstupov neurónu x_1, x_2, \dots, x_n , ktoré sú vynásobené odpovedajúcou hodnotou synaptickej váhy w_1, w_2, \dots, w_n , sú privedené na sumačný člen Σ , kde sa uskutoční súčet všetkých vážených vstupov a určí sa hodnota potenciálu neurónu. Na základe hodnoty potenciálu neurónu sa pomocou aktivačnej funkcie δ vypočíta výstup neurónu y . Model je zobrazený na obrázku.



Obr. 6.4: Matematický model neurónu

6.4.2 Činnosť neurónovej siete

Činnosť neurónovej siete môžeme rozdeliť na dve základné fázy. Prvou fázou je fáza učenia a druhou je fáza života. Vo fáze učenia neurónovej siete dochádza k ukladaniu znalostí do synaptických váh, ktoré sa v tejto fáze menia podľa vzťahu

$$\frac{\partial dW}{\partial dT} \neq 0 . \quad (6.18)$$

Samotné učenie neurónovej siete sa rozdeľuje na kontrolované a nekontrolované.

V druhej fáze, fáze života neurónovej siete, sa už samotné využitie naučených znalostí z fázy učenia. V tejto fáze sa synaptické váhy nemenia, podľa vzťahu

$$\frac{\partial dW}{\partial dT} = 0 . \quad (6.19)$$

6.4.3 Učenie neurónových sietí

Učenie neurónových sietí je pojem, ktorý znamená účelnú adaptáciu synaptických váh tak, aby celý systém mal na výstupe požadovanú odozvu. Princípom učenia je teda eliminácia alebo posilňovanie synaptických váh v závislosti na požadovanom výstupe neurónu. Učenie neurónových sietí sa rozdeľuje na dve techniky - kontrolované (učenie s učiteľom) a nekontrolované (učenie bez učiteľa). Technika učenie s učiteľom je charakteristická v tom, že požadovaný výstup neurónu je zadaný ručne užívateľom. Na rozdiel od techniky učenie bez učiteľa, v ktorom sa požadovaný výstup určí pomocou samoorganizujúceho mechanizmu.

V neurónových sieťach typu učenie s učiteľom, sa hodnoty synaptických váh nastavujú v závislosti na odchýlke aktuálneho výstupu od požadovaného. Takže, ak je odchýlka aktuálneho výstupu v tolerancii, považujeme sieť za naučenú. V opačnom prípade je dôležitá zmena synaptických váh, ktorá sa označuje aj ako delta pravidlo a vyjadríme ju:

$$\Delta w_{ij} = \eta x_i (d_i - y_i) , \quad (6.20)$$

kde η je koeficient rýchlosti učenia siete, x_i je aktuálny vstup, y_i je aktuálny výstup a d_i je výstup požadovaný. Následne sa výsledná hodnota zmeny synaptickej váhy Δw_{ij} pripočítaná k aktuálnej váhe podľa vzťahu

$$w_{ij} = w_{ij} - \Delta w_{ij} . \quad (6.21)$$

Najpoužívanejším typom učenia s učiteľom je algoritmus Back-Propagation, teda spätného šírenia chyby.

6.4.4 Dopredné neurónové siete

Sú najpoužívanejším typom architektúry neurónových sietí. V doprednej neurónovej sieti je signál privedený na vstup neurónu, ktorý ho spracuje na výstup a ten privedie na vstup ďalšieho neurónu.

Najjednoduchším typom doprednej siete je sieť perceptrónov, ktorá obsahuje iba vstupné a výstupné neuróny. V nej sú vstupné vzorky privádzané pomocou vstupných neurónov na výstupné, ktoré urobia výpočet. V sieti sú neuróny organizované do vrstiev, a takejto sieti hovoríme jednovrstvová neurónová sieť, pri ktorej sa na tréning používa delta pravidlo.

Rozšírením predchádzajúcej siete perceptrónov je pridanie ďalšej vrstvy, tzv. skrytej vrstvy medzi vstupnú a výstupnú vrstvu. Na tréning takejto siete sa využíva Back-Propagation algoritmus.

Okrem dopredných neurónových sietí sa ešte vyskytujú rekurentné, asociačné pamäte a samoorganizujúce mapy. Rekurentné siete majú aspoň jednu spätnú väzbu, ktorá umožňuje neurónu získať údaje z predchádzajúceho behu. Asociačné pamäte majú schopnosť zapamätať si obraz dát privedených na vstup pamäte počas tréningu a samoorganizujúce mapy obsahujú stovky až tisíce neurónov, a teda majú vysokú časovú náročnosť tréningu.

Alogoritmus Back-Propagation

Táto sieť je charakterizovaná svojim učebným procesom, v ktorom pracuje na princípe spätného šírenia chyby. Spočíva v postupnom vyhodnocovaní reakcie neurónovej siete na jednotlivé vstupy, v porovnávaní týchto výstupov s predlohami (správnymi odpoveďami) a v úprave synaptických váh tak, aby rozdiel medzi reakciou upravenej neurónovej siete a predlohy bol minimálny.

Algoritmus pozostáva z viacerých krokov. Prvým je inicializácia všetkých váh w_{ij} náhodne v rozsahu -1 až 1. V druhom kroku vezmeme ďalší vzorok z tréningovej sady dát x_1, x_2, \dots, x_n , u ktorých poznáme očakávané výstupy a necháme sieť tento vstup vyhodnotiť, čím získame hodnoty y_j . V treťom kroku sa vypočítajú nové hodnoty váh w_{ij} . V poslednom štvrtom kroku sa pokračuje krokom druhým pokiaľ sa neprejdú všetky vzorky, poprípade sa môžu vzorky prejsť aj viackrát.

6.4.5 Rekurentné neurónové siete

Rekurentná neurónová sieť predstavuje dynamicky zložitejší systém doprednej neurónovej siete. Táto sieť má podmnožinu neurónov (rekurentných neurónov), ktoré uchovávajú informáciu o svojich aktivitách z predošlých stavov. Pri výstupe z neurónu nerozhoduje len momentálny

výstup v čase t , ale aj vstupy z minulosti, a teda algoritmy učenia musia brať do úvahy aj túto skutočnosť. U dopredných neurónových sieťach je signál vedený iba jedným smerom, a teda dopredu, takže sieť nie je schopná rozoznať časovú závislosť dát. Naopak rekurentné siete sú schopné si vytvoriť stavovú reprezentáciu časového kontextu. Preto sa rekurentné siete uplatňujú pri riešení iných úloh, ktoré sa rozdeľujú do troch typov : klasifikačné, predikčné a generatívne úlohy. U klasifikačných úlohách sieť rozhoduje o tom, či daná postupnosť vstupov patrí do niektorej triedy. U predikčných úloh zisťuje neurónová sieť závislosť medzi dátami s využitím časového úseku, čím dokáže predikovať nasledujúcu hodnotu v čase. Posledné generatívne úlohy sú zložitejším typom predikčných úloh, pretože sieť musí pokračovať v časovom rade dát vzhľadom na tendenciu skrytú v danej postupnosti dát. [5,23]

7. Klasifikácia výslovnosti hlásky „R“

Nasledujúce kapitoly sa zaoberajú praktickou časťou diplomovej práce, a teda vytvorením a otestovaním programu pre automatickú klasifikáciu výslovnosti hlásky „R“ u určitých tréningových slov. K spracovaniu signálu reči a následnej klasifikácii výslovnosti bol použitý software Matlab verzia R2016a (9.0.0.341360). Matlab je programovacie prostredie s vlastným programovacím jazykom vychádzajúci z jazyka Fortran, ktorý sa špecializuje na vedeckotechnické numerické výpočty, návrhy algoritmov, meranie a spracovanie signálov, analýzu a iné. Názov Matlab vznikol skrátením slov Matrix Laboratory.

7.1 Naviazanie na semestrálnu prácu

V rámci semestrálnej práce boli naštudované rôzne metódy pre klasifikáciu a rozpoznávanie reči. Tieto metódy sú popísané v predchádzajúcich kapitolách. Semestrálna práca tiež pozostávala z návrhu metódy pre automatickú klasifikáciu hlásky „R“. Následne v tejto diplomovej práci som realizoval navrhnutú metódu, otestoval a vyhodnotil úspešnosť zvolenej automatickej klasifikácie. Popis programových častí ako aj vyhodnotením úspešnosti metódy sa budem venovať v nasledujúcich kapitolách.

7.2 Nahrávky tréningových slov

Nahrávky jednotlivých tréningových slov pre klasifikáciu hlásky „R“ boli z prevažnej časti získané od pacientov súkromnej kliniky LOGO s.r.o.. Nahrávky boli získané priamo z logopedickej ordinácie pomocou zakúpeného mikrofónu a nahrávané priamo PHDr. Petrom Staničkom. Aj keď ordinácia nie je z akustického hľadiska veľmi vyhovujúca, nahrávky neobsahujú šum z pozadia, čo je v konečnom dôsledku veľkou výhodou pre následné spracovanie zvukového signálu.

Poskytnutých bolo celkovo 15 nahrávok s nesprávnou výslovnosťou, pričom 5 nahrávok bolo od dievčat vo veku 5 až 6 rokov. Zvyšných 10 bolo od chlapcov vo veku od 5 do 11 rokov. Tieto nahrávky sa tiež líšili rôznym typom poruchy výslovnosti hlásky „R“. U dievčat sa vyskytli celkovo 4 poruchy. Po jednej nahrávke boli dievčatá s menej kmitným rotacizmom, príliš veľkým dôrazom na vibráciu, velárnym rotacizmom a dve nahrávky s paralalickým rotacizmom. U paralalického rotacizmu prvé dievča zamieňalo hlásku „R“ za „L“ a druhé za „G“. V prípade chlapcov to bola jedna nahrávka s bilabiálnym rotacizmom, velárnym rotacizmom a s prehnaným tlakom, štyri nahrávky s mogilalickým rotacizmom, tri nahrávky paralalického rotacizmu.

U paralalického rotacizmu zamienali všetci traja chlapci hlásku „R“ za „L“, buď čiastočne alebo úplne.

Nahrávok so správnou výslovnosťou bolo z logopedickej kliniky poskytnutých 5, jedna od dievčaťa a štyri od chlapcov. Nahrávky so správnou výslovnosťou som doplnil o ďalších 6, dve od ženy a štyri od chlapov. Je nutné podotknúť, že tieto nahrávky sú od ľudí vo veku od 18 do 24 rokov. Tieto nahrávky som nahral pomocou softwaru Audacity verzia 2.2.2, ktorý je slobodný open source, multiplatformový program na úpravu digitálneho zvuku, použil som pri tom externý mikrofón zo slúchadiel. Nahrávky boli značne zaťažené šumom z pozadia, ktorý som odstránil s využitím funkcie noise reduction v softwari Audacity. Táto funkcia si najskôr nahrála šum z pozadia a následne ho vyfiltrovala z nahrávky. Takto upravená nahrávka už sa dala subjektívne sluchom porovnať z nahrávkou z logopedickej kliniky. Celkovo nahrávok so správnou výslovnosťou je teda 11.

V každej nahrávke bolo 15 tréningových slov od jedného pacienta, takže nahrávku bolo nutné zostrihať na jednotlivé samostatné slová. K tomu som použil software Audacity. Celkovo bolo získaných 225 slov zlej výslovnosti, pričom 3 slová sa použiť nedali a 165 správnych slov.

Tieto tréningové slová sú rozdelené do troch skupín podľa pozície hlásky „R“. V prvej skupine sú slová s hláskou „R“ na začiatku slova, a sú to – ryba, rak, robot, rýč, ruka. V druhej skupine sú slová s hláskou „R“ v strede slova – strom, Brno, Praha, Franta, motorka. V poslednej skupine sú slová s hláskou „R“ na konci – vítr, motor, obr, mír, dar.

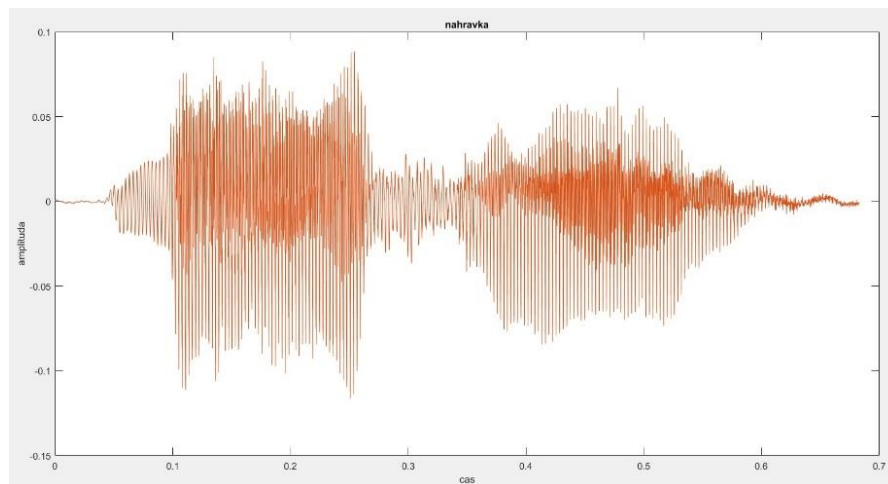
7.3 Súčasti programu pre klasifikáciu výslovnosti slov

Vytvorený klasifikátor pozostáva z dvoch zložiek, *klasifikácia slov* a *klasifikácia hlások*, v ktorých sú štyri matlabovské skripty. V zložke *klasifikácia slov* je vytvorená klasifikácia pre celé tréningové slová, a v zložke *klasifikácia hlások* je klasifikácia ručne nasegmentovaných hlások „R“. Skripty v zložkách sú inak zhodné. Prvý skript s názvom *mfcc_pre_nn.m*, ktorý má za úlohu extrahovať Mel-frekvenčné kepstrálne koeficienty pre natrénovanie neurónovej siete. Druhým skriptom je natrénovaná neurónová sieť s názvom *net.mat*, ktorá bola vytvorená pomocou NNtoolboxu v Matlabe. Nasledujúce dva sú pre automatickú klasifikáciu. Kód s názvom *nacitanie_nahravky.m*, v ktorej máme možnosť si vybrať predom nahrávku a klasifikovať jeho výslovnosť a skript *nahravanie_nahravky.m*, v ktorom slovo alebo hlásku, môžeme nahráť priamo a klasifikovať.

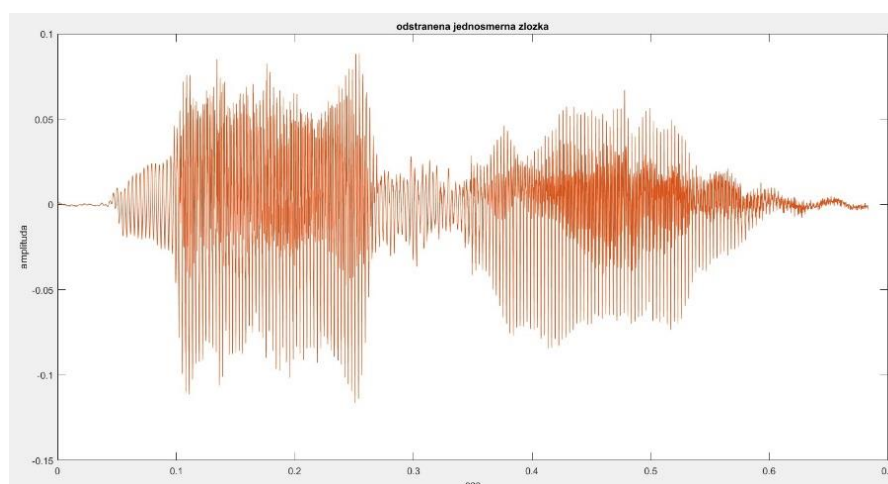
7.4 Predspracovanie a výpočet MFCC

Táto časť kódu sa vyskytuje v troch matlabovských skriptoch a to v *mfcc_pre_nn.m*, *nacitanie_nahravky.m* a *nahravanie_nahravky.m*. Je to najdôležitejšia časť programu, okrem neurónovej siete, skripty sa následne líšia iba vstupom a výstupom.

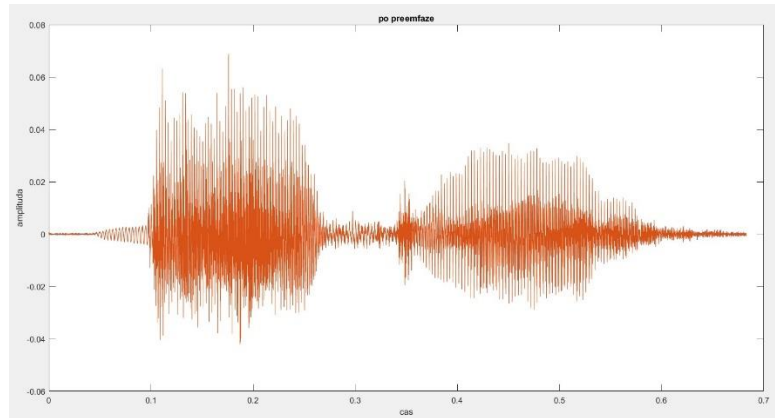
Súčasťou predspracovania je odstránenie jednosmernej zložky, ktorá nenesie žiadnu dôležitú informáciu a môže byť rušivá pri ďalšom spracovaní signálu, tým pádom je dôležité jej odstránenie. Program to realizuje pomocou *for clyklu*, kedy z každej vzorky v nahrávke odstráni jeho strednú hodnotu. Ďalšou časťou predspracovania je preemfáza, v ktorej sa vyrovnáva kmitočtová charakteristika rečového signálu, pretože vplyvom rečového ústrojenstva dochádza k potláčaniu vyšších frekvencií. V preemfáze je signál filtrovaný hornopriepustným FIR filtrom. Poslednou časťou je normalizácia signálu, kedy sa signál vydolí absolútnou maximálnou hodnotou daného signálu, čím sa normalizuje amplitúda do hodnôt $[1, -1]$.



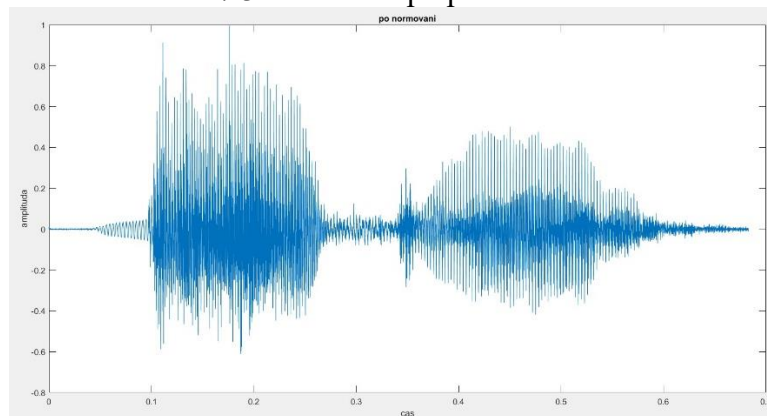
Obr. 7.1: Nahrávka slova ryba s logopedickou vadou (lyba)



Obr. 7.2: Nahrávka po odstránení jednosmernej zložky

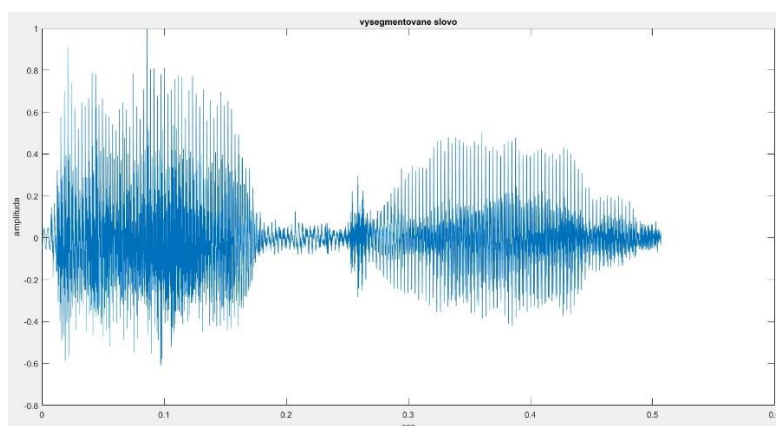


Obr. 7.3: Nahrávka po preemfáze



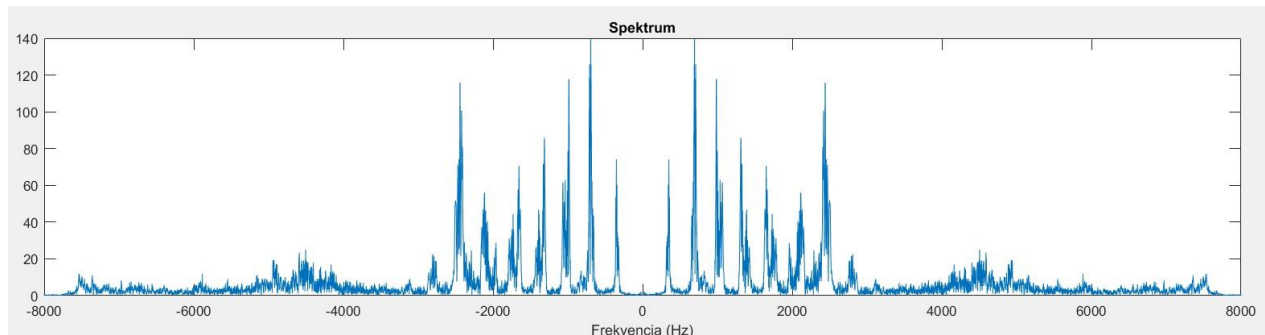
Obr. 7.4: Nahrávka po normovaní

Nasledujúcou časťou je orezanie začiatku a konca nahrávky, čo by sme mohli charakterizovať ako jednoduchú segmentáciu reči z nahrávky. Táto časť sa nenachádza v skriptoch *mfcc_pre_nn.m* a *nacitanie_nahravky.m*, pretože pri ručne nasegmentovaných hláskach nebola potrebná. V ostatných skriptoch je orezanie charakterizované veľkosťou amplitúdy, ktorá bola otestovaná na veľkom počte nahrávok, a jej hodnota je 0,05. Program teda pri orezaní začiatku nahrávky prechádza jednotlivé vzorky od začiatku, až kým narazí na hodnotu amplitúdy väčšiu ako 0,05, a následne odstráni celú časť nahrávky od začiatku až po danú vzorku presahujúcu túto hodnotu. Ten istý postup sa udeje aj pri orezaní konca, len program prechádza vzorky od konca nahrávky.



Obr. 7.5: Nahrávka po orezaní začiatku a konca

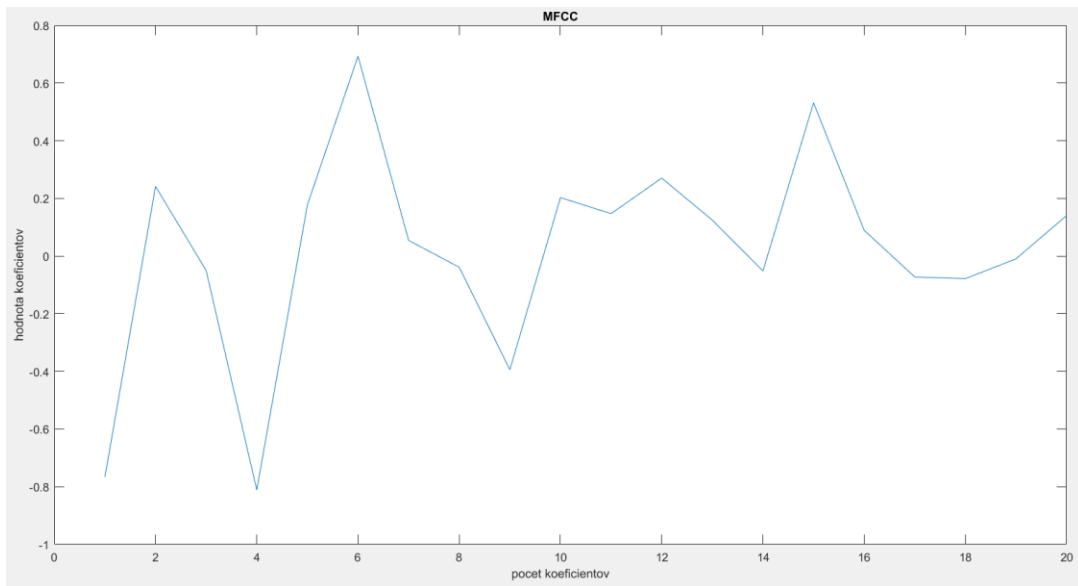
Následne prichádza výpočet Mel-frekvenčných keprálnych koeficientov z nahrávky. Podstatná časť tohto kódu bola použitá zo zdroja [11]. Na predspracovanú a orezanú nahrávku slova sa použije funkcia *enframe.m*, použitá zo zdroja [12], ktorá signál rozdelí na prekrývajúce sa úseky konštantnej dĺžky s určitým prekryvom. Úsek je nastavený premennou *frameSize* na hodnotu 256 a prekryv premennou *frameOverlap* na hodnotu 128. Na takto rozdelenú nahrávku sa aplikuje Hammingovo okno pomocou funkcie *hamming*. Ďalším krokom je transformácia nahrávky pomocou DFT do spektrálnej oblasti, funkciou *fft* a *abs*.



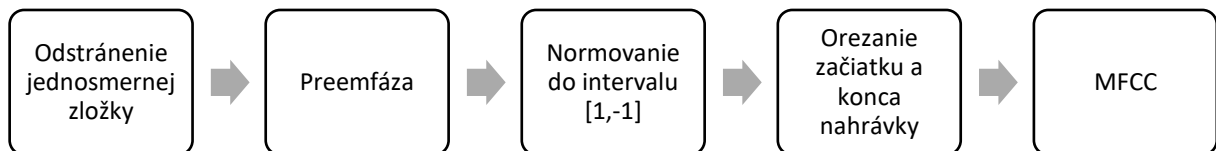
Obr. 7.6: Nahrávka po *fft*

Následne sa vyberie iba polovica spektra nahrávky, keďže spektrum je symetrické. V ďalšej časti je podstatná časť Melovskej filtrácie a to vytvorenie trojuholníkových filtrov charakteristických pre MFCC, kde môžeme pomocou premennej *Nofilters* nastaviť počet filtrov. Vytvorené filtre sa aplikujú na nahrávku v spektrálnej oblasti. Posledným krokom je zlogaritmovanie výstupu filtrov prirodzeným logaritmom a diskretná kosínusová transformácia.

Výstupom bola matica o rozmere $M \times S$, kde M je počet koeficientov a S je počet zložiek *fft*. Takto získaná matica MFCC z nahrávky sa ešte upravila spriemerovaním hodnôt S jednotlivých zložiek pre každý jeden koeficient. Zobrali sa teda všetky hodnoty prvého koeficientu MFCC, sčítali sa a vydělili množstvom hodnôt, a tak sme dostali jednu priemernú hodnotu pre koeficient MFCC. Následne sa táto operácia spravila pre druhý koeficient MFCC, až po posledný dvadsiaty. Výstupom sme teda dostali maticu o rozmere $M \times 1$, kde M je počet koeficientov teda 20.



Obr. 7.7: Výsledné MFCC nahrávky



Obr. 7.8: Bloková schéma predspracovania a výpočtu MFCC

7.5 Extrahovanie MFCC pre natrénovanie neurónovej siete

Ako už bolo spomenuté, skript s názvom *mfcc_pre_nn.m* má za úlohu extrahovať MFCC z jednotlivých nahrávok.

V tomto skripte si vyberieme zložku s jednotlivými nahrávkami, ktorými následne budeme trénovať neurónovú sieť. Následne celý postup výpočtu prebehne automaticky, vo *for* cykle si program postupne vyberá jednotlivé nahrávky, prebehne predspracovanie každej nahrávky a výpočet MFCC podľa postupu v kapitole 7.4. Každá matica MFCC z nahrávky o rozmere $M \times 1$ sa zapíše do výstupného textového súboru. Celý proces sa opakuje pre nasledujúce nahrávky zo zložky, ktorých matice MFCC sa následne zapíšu na ďalšie riadky vo výstupnom textovom súbore. Výstupom je teda matica MFCC nahrávok v textovom súbore, o rozmere $M \times N$. Kde M je počet MFCC koeficientov, ktorý je v našom prípade 20 a N reprezentuje počet nahrávok vo zvolenej zložke. MFCC koeficienty sú teda reprezentované stĺpcami, a jednotlivé nahrávky riadkami.

7.6 Klasifikátor – neurónová sieť

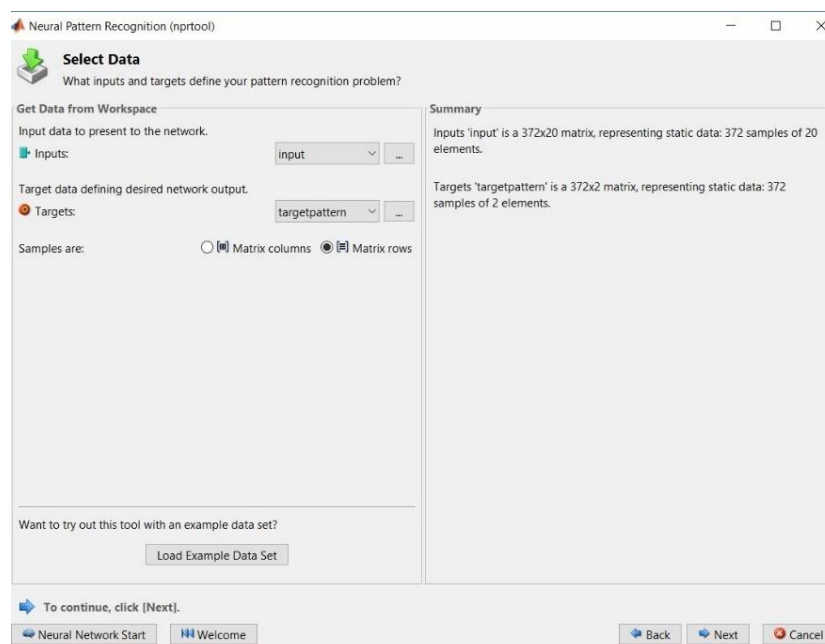
Pre klasifikáciu MFCC z nahrávky bola použitá neurónová sieť typu pattern recognition z Neural Network Toolboxu v Matlabe. Neural Pattern Recognition (nprtool) je neurónová sieť,

založená na rozpoznávaní vzorov, ktorá klasifikuje vstupy do súboru cieľových kategórií. Je to dvojvrstvová dopredná neurónová sieť, so skrytými sigmoidmi a výstupnými neurónmi typu softmax. Sieť bude natrénovaná funkciou *trainscg*, ktorá aktualizuje váhy podľa stupnice konjugovanej gradientovej metódy.

Klasifikátor vyhodnocuje MFCC nahrávok v skriptoch *nacitanie_nahravky.m* a *nahravanie_nahravky.m*. Po načítaní alebo nahratí nahrávky, prebehne predspracovanie, orezanie a výpočet matice MFCC nahrávky, o veľkosti $M \times 1$. Následne je matica MFCC vyhodnotená natrénovanou neurónovou sieťou, a nahrávka je klasifikovaná ako nesprávna pre výstup [1,0] z neurónovej siete a správna pre výstup [0,1].

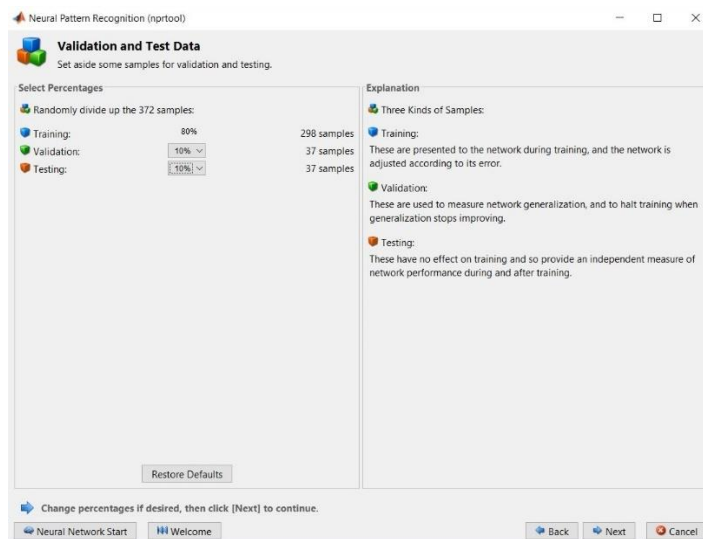
7.6.1 Trénovanie neurónovej siete pre klasifikáciu slov

Ako už bolo spomenuté v predchádzajúcej kapitole, využili sme neurónovú sieť typu pattern recognition z Neural Network Toolboxu. Pre natrénovanie siete bolo nutné mať maticu MFCC nahrávok slov, ktoré chceme natrénovať. Tieto nahrávky slov sú výstupom skriptu *mfcc_pre_nn.m* v podobe textového súboru s názvom *input.txt*. Taktiež bolo nutné vytvoriť target maticu, v ktorej je zadefinovaná trieda pre jednotlivé MFCC nahrávky slov v textovom súbore *input.txt*. Obe matice pozostávajú z rovnakého počtu riadkov, a teda počtu nahrávok slov. V target matici, teda textovom súbore *targetpattern.txt* je prvých 192 hodnôt zadefinovaných do triedy [1,0], ktorá odpovedá prvým 192 nahrávkam slov s nesprávnou výslovnosťou v *input.txt*, a nasledujúcich 180 hodnôt do triedy [0,1] odpovedajúcim 180 správne vysloveným nahrávkam slov v *input.txt*. Tieto dva textové súbory sú vybrané do neurónovej siete ako input a target, čo nám ukazuje obrázok 7.1.



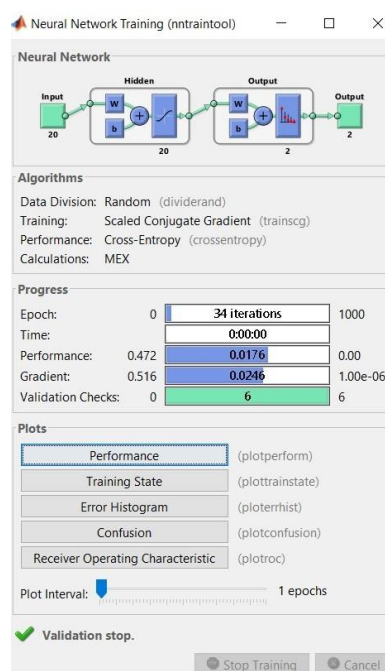
Obr. 7.9: Výber dát (input a target) pre neurónovú sieť

Následne sme si zvolili koľko nahrávok slov z celkových 372 sa využije pre natréovanie, validáciu a testovanie neurónovej siete. Zvolili sme si 298 (80%) nahrávok slov pre testovanie, 37 (10%) pre validáciu a taktiež 37 (10%) pre testovanie, ako môžeme vidieť na obrázku 7.2.



Obr. 7.10: Rozdelenie počtu nahrávok pre tréovanie a testovanie

V ďalšej časti bolo zvolenie počtu neurónov vo vrstve, ktorých sme zvolili 20. Následne sme neurónovú sieť natréovali s 298 slovami. Neurónová sieť sa natréovala po 34 iteráciach, čo možno vidieť na obrázku 7.3. Percentuálna úspešnosť tréovacej matice bola 97,3%, úspešnosť validácie, na ktorú bolo použitých 37 nahrávok bola 97,3%. Po natréovaní sa neurónová sieť otestovala s 37 nahrávkami, ktoré neboli použité na tréovanie a s úspešnosťou 97,3% ich priradila k správnej triede, kde 1 je trieda nesprávnej výslovnosti a 2 správnej. Tieto výsledky môžeme vidieť na obrázku 7.4.



Obr. 7.11: Natréovanie neurónovej siete



Obr. 7.12: Percentuálna úspešnosť tréovania, validácie a testovania slov

Na obrázku 7.5 môžeme vidieť krivku ROC (Receiver Operating Characteristic), čo je nástroj pre hodnotenie a optimalizáciu klasifikačného systému, ktorý ukazuje vzťah medzi senzitivitou a špecificitou daného systému. Na zvislej ose grafov je relatívna početnosť skutočne pozitívnych hodnôt (True Positive - TP) a na vodorovnej ose je relatívna početnosť falošne pozitívnych prípadov (False Positive - FP). Senzitivitu testovaných hodnôt môžeme vypočítať pomocou vzťahu

$$P(TP) = TP / (TP + FN) \quad (7.1)$$

kde FN je počet falošne negatívnych hodnôt. Po vyčíslení:

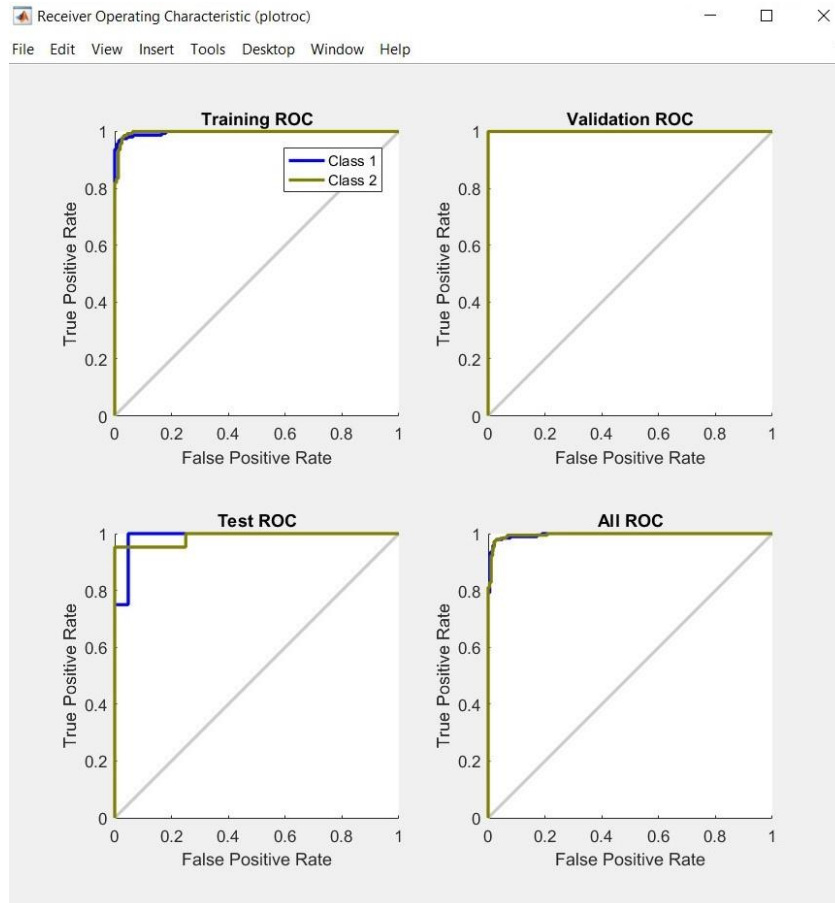
$$P(TP) = 20 / (20 + 1) = 0,952 = 95,2\%.$$

Špecificitu následne vypočítame podľa vzťahu

$$P(FP) = TN / (FP + TN) \quad (7.2)$$

a teda po vyčíslení dostávame

$$P(FP) = 16 / (0 + 16) = 1 = 100\%.$$



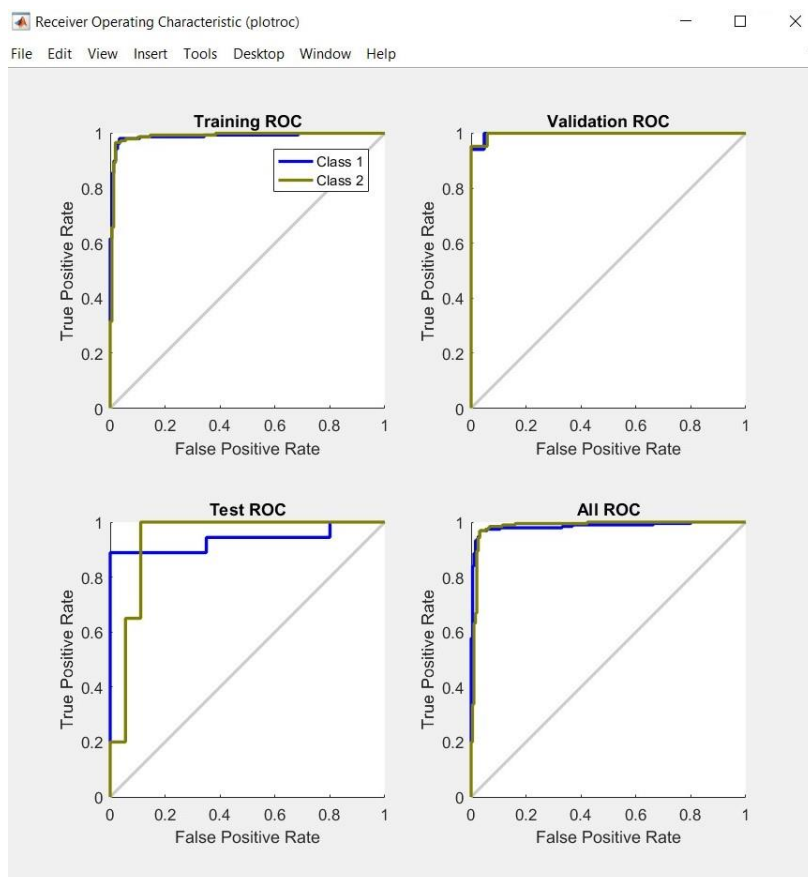
Obr. 7.13: Krivka ROC pre slová

7.6.2 Trénovanie neurónovej siete pre klasifikáciu hlások

Tak ako pri tréovaní neurónovej siete pre klasifikáciu slov, tak aj pre hlásky musíme najskôr pomocou skriptu *mfcc_pre_nn.m* získať maticu MFCC hlások v textovom súbore *input.txt* a vytvoriť si target maticu v textovom súbore *targetpattern.txt*. Matica *input.txt* obsahuje prvých 194 nesprávne vyslovených hlások, čo odpovedá triede [1,0] v matici *targetpattern.txt*, následných 190 správne vyslovených hlások odpovedá triede [0,1] v matici *targetpattern.txt*. Tieto matice sme tak isto ako opisujem v predchádzajúcej kapitole zadali pre natréovanie neurónovej siete, s rovnakým počtom neurónov vo vrstve, a tiež sme ich rozdelili. Pre natréovanie sa použilo 308 (80%) nahrávok hlásky, pre validáciu 38 (10%) tak ako aj pre testovanie. Po 58 iteráciach sa tréovanie zastavilo, a neurónová sieť sa otestovala na 38 nahrávkach hlásky s 92,1% úspešnosťou zaradenia do triedy s nesprávnou alebo správnou výslovnosťou, ako je možné vidieť na obrázku 7.6.



Obr. 7.14: Percentuálna úspešnosť tréovania, validácie a testovania hlások



Obr. 7.15: Krivka ROC pre hlások

Na obrázku 7.7 môžeme vidieť krivku ROC. Senzitivitu testovaných hodnôt môžeme vypočítať pomocou vzťahu (7.1), kde po vyčíslení dostávame:

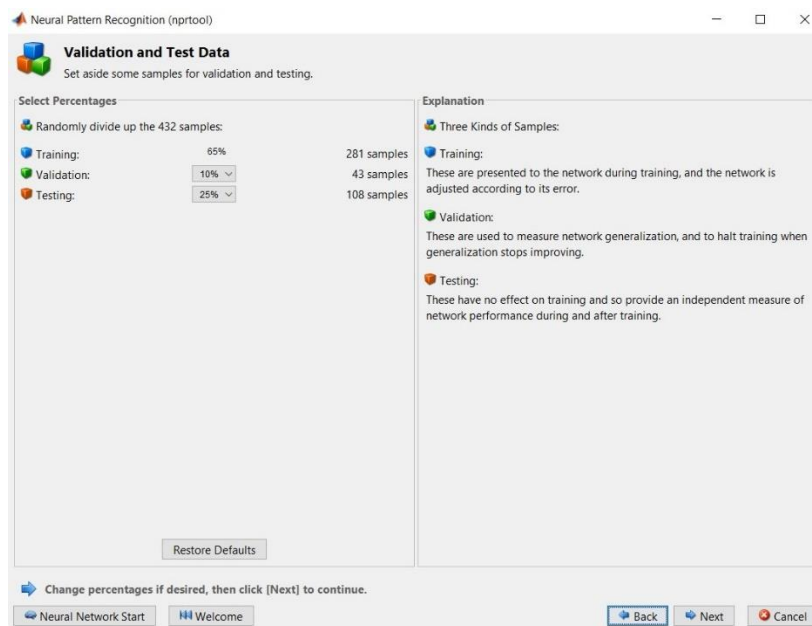
$$P(TP) = 20/(20 + 0) = 1 = 100\%.$$

Špecificitu následne vypočítame podľa vzťahu (7.2) a po vyčíslení dostávame:

$$P(FP) = 15/(3 + 15) = 0,833 = 83,3\%.$$

7.6.3 Trénovanie a testovanie neurónovej siete pri použití všetkých slov

V tejto kapitole si ukážeme ako sa neurónová sieť netrénovala a sama otestovala pri použití všetkých dostupných nahrávok slov. Počet všetkých dostupných nahrávok slov bol 432, z toho 222 s nesprávnou výslovnosťou a 210 so správnou. Opäť sme vytvorili matice v textových súboroch *input.txt* a *targetpattern.txt*. Pre natrénovanie neurónovej siete sme použili 281 (65%) slov so správnou a nesprávnou výslovnosťou, 43 (10%) slov sme využili pre validáciu. Na 108 (25%) slovách sa neurónová sieť sama otestovala, ako môžeme vidieť na obrázku 7.8.

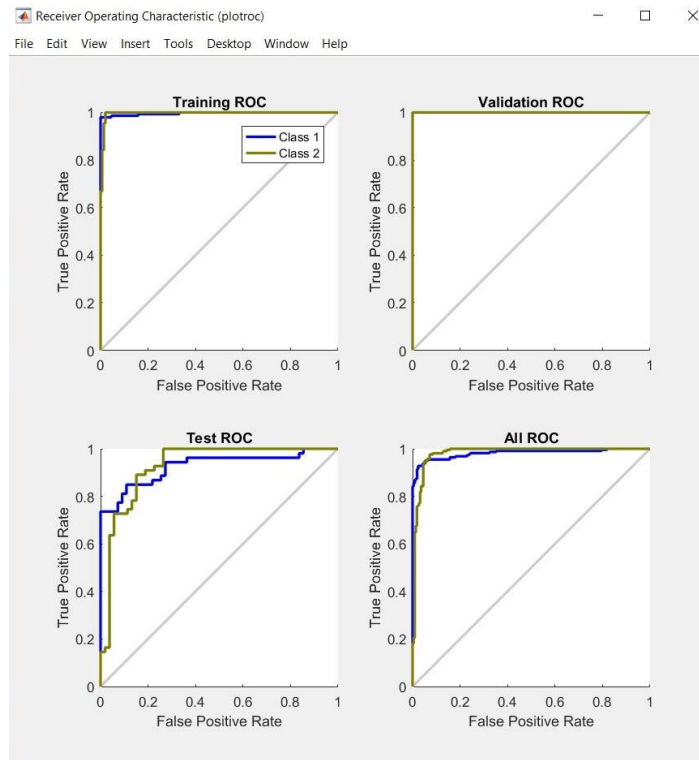


Obr. 7.16: Počet slov pre trénovanie, validáciu a testovanie

Počet neurónov v skrytej vrstve bol nastavený na hodnotu 20. Výsledok testovania neurónovej siete môžeme vidieť na obrázku 7.9, kde sa sama otestovala s 84,3% úspešnosťou.



Obr. 7.17: Výsledok samoostovania sa neurónovej siete



Obr. 7.18: ROC krivka

Senzitivitu testovania môžeme vypočítať pomocou vzťahu (7.1), kde po vyčíslení dostávame:

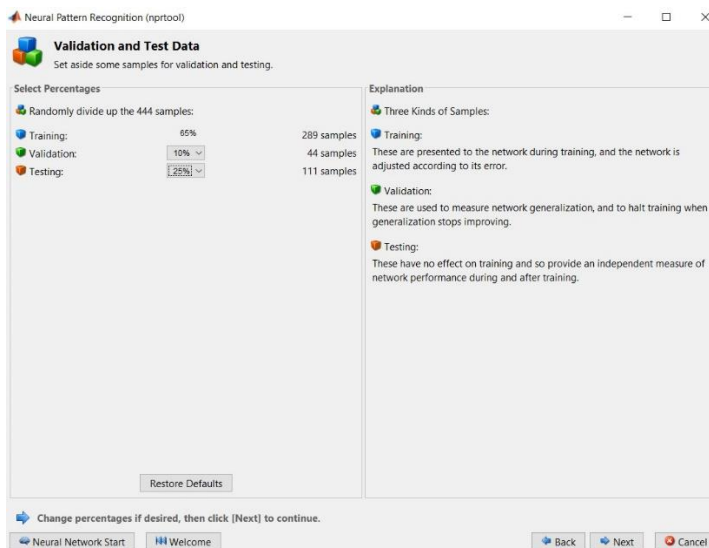
$$P(TP) = 50 / (50 + 5) = 0,91 = 91\%.$$

Špecificitu následne vypočítame podľa vzťahu (7.2) a po vyčíslení dostávame:

$$P(FP) = 41 / (12 + 41) = 0,773 = 77,3\%.$$

7.6.4 Trénovanie a testovanie neurónovej siete pri použití všetkých hlások

Tak ako v predchádzajúcej kapitole, neurónovú sieť natrénujeme a otestujeme s využitím všetkých dostupných ručne nasegmentovaných hlások „R“. Počet ručne nasegmentovaných hlások „R“ bol 444, z toho 224 s nesprávnou výslovnosťou a 220 s výslovnosťou správnu. Pre natrénovanie sme využili 289 (65%) hlások, validáciu 44 (10%) a pre samootestovanie sa 111 (25%) hlások s nesprávnou a správnu výslovnosťou.



Obr. 7.19: Počet hlások pre testovanie, validáciu a tréovanie

Ako môžeme vidieť na nasledujúcom obrázku, neurónová sieť sa dokázala sama otestovať s 93,7% úspešnosťou vyhodnotenia výslovnosti.



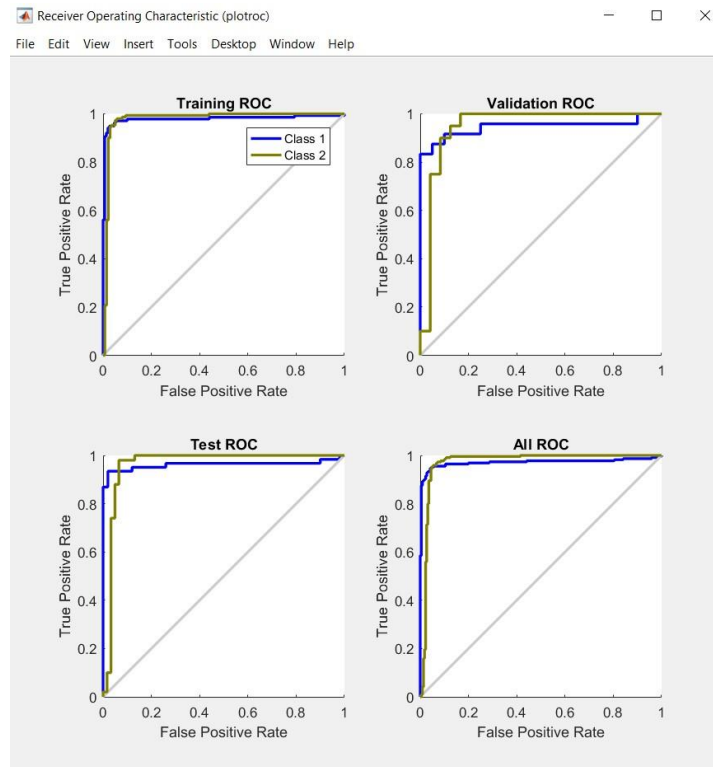
Obr. 7.20: Výsledok samootestovania sa neurónovej siete

Senzitivitu testovaných hodnôt môžeme vypočítať pomocou vzťahu (7.1), po vyčíslení:

$$P(TP) = 49/(49 + 1) = 0,98 = 98\%.$$

Špecificitu následne vypočítame podľa vzťahu (7.2) a teda po vyčíslení dostávame

$$P(FP) = 55/(6 + 55) = 0,902 = 90,2\%.$$



Obr. 7.21: Krivka ROC

7.7 Automatická klasifikácia – načítanie nahrávky

Skript *nacitanie_nahravky.m* sa využije pre automatickú klasifikáciu výslovnosti slov alebo hlások, s použitím natrénovanej neurónovej siete pre slová alebo pre hlásky, ktoré sú popísané v predchádzajúcich kapitolách.

7.7.1 Výsledky automatickej klasifikácie slov

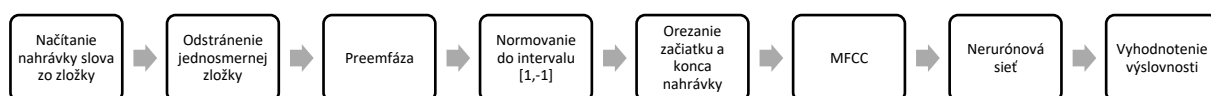
V skripte pre automatickú klasifikáciu slov s názvom *nacitanie_nahravky.m*, máme možnosť na začiatku si vybrať nahrávku slova zo zložky, ktorú chceme klasifikovať. Nahrávka sa predspracuje, oreže a vypočítajú sa MFCC, ako je popísané v kapitole 7.4. Následne vypočítaná matica MFCC nahrávky slova vstupuje na neurónovú sieť popísanú v kapitole 7.6.1 a prebehne klasifikácia. Nahrávku slova klasifikujeme do triedy s nesprávnou výslovnosťou alebo so správnou. Pri nesprávnej výslovnosti sa zobrazí obrázok



Obr. 7.22: Zobrazenie znaku pre nesprávnu výslovnosť



Obr. 7.23: Zobrazenie znaku pre správnu výslovnosť



Obr. 7.24: Bloková schéma automatickej klasifikácie slov

Pre vyhodnotenie klasifikácie som použil 30 nesprávnych nahrávok tréningových slov od 15 osôb a taktiež 30 správnych nahrávok tréningových slov od 15 osôb. Výsledok klasifikácie je môžeme vidieť v nasledujúcej tabuľke:

	Správne klasifikované	Nesprávne klasifikované
Slová s nesprávnou výslovnosťou	25	5
Slová so správnou výslovnosťou	11	19

Tab. 7.1: Výsledok automatickej klasifikácie slov

Z tabuľky 7.1 môžeme vidieť, že úspešnosť správnej klasifikácie bola 60%. Senzitivitu klasifikácie slov môžeme vypočítať pomocou vzťahu (7.1), kde po vyčíslení dostávame:

$$P(TP) = 11 / (11 + 5) = 0,687 = 68,7\%.$$

Špecificitu následne vypočítame podľa vzťahu (7.2) a po vyčíslení dostávame:

$$P(FP) = 25 / (19 + 25) = 0,568 = 56,8\%.$$

Ak by sme zarátali aj slová na ktoré sa otestovala neurónová sieť z kapitoly 7.6.1, dostali by sme iný výsledok klasifikácie:

	Správne klasifikované	Nesprávne klasifikované
Slová s nesprávnou výslovnosťou	25+16 = 41	5+1 = 6
Slová so správnou výslovnosťou	11+20 = 31	19+0 = 19

Tab. 7.2: Výsledok automatickej klasifikácie slov spolu s testom neurónovej siete

Podľa tabuľky 7.2 sa nám úspešnosť správnej klasifikácie zväčšila na hodnotu 74,22%. Hodnotu senzitivity vypočítanú podľa vzťahu (7.1), dostávame:

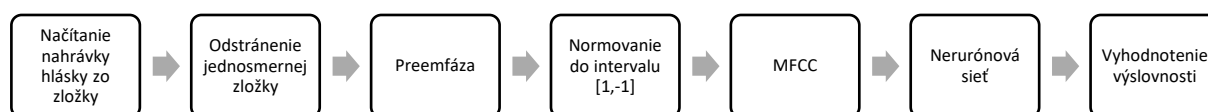
$$P(TP) = 31/(31 + 6) = 0,838 = 83,8\%.$$

Špecificitu vypočítame podľa vzťahu (7.2) a dostávame:

$$P(FP) = 41/(19 + 41) = 0,683 = 68,3\%.$$

7.7.2 Výsledky automatickej klasifikácie ručne nasegmentovaných hlások „R“

Tak ako pre automatickú klasifikáciu slov, tak aj pre automatickú klasifikáciu ručne nasegmentovaných hlások „R“ sa využije skript s názvom *nacitanie_nahravky.m*, v ktorom máme možnosť na začiatku si vybrať nahrávku hlásky zo zložky, ktorú chceme klasifikovať. Nahrávka sa predspracuje, avšak neoreže, pretože bola ručne segmentovaná, takže predpokladáme že bola segmentovaná správne. Následne sa vypočítajú MFCC, ako je popísané v kapitole 7.4. Takto vypočítaná matica MFCC nahrávky hlásky vstupuje na neurónovú sieť popísanú v kapitole 7.6.2 a prebehne klasifikácia. Nahrávku hlásky „R“ klasifikujeme do triedy s nesprávnou výslovnosťou (obr. 7.21) alebo so správnou (obr. 7.22).



Obr. 7.25: Blokovaná schéma automatickej klasifikácie hlások

Pre vyhodnotenie klasifikácie som použil 30 nesprávnych a taktiež 30 správnych nahrávok hlások „R“. Výsledok klasifikácie je môžeme vidieť v nasledujúcej tabuľke:

	Správne klasifikované	Nesprávne klasifikované
Hlásky s nesprávnou výslovnosťou	28	2
Hlásky so správnou výslovnosťou	5	25

Tab. 7.3: Výsledok automatickej klasifikácie slov

Z tabuľky 7.3 môžeme vidieť, že úspešnosť správnej klasifikácie bola 55%. Senzitivitu klasifikácie hlások vypočítame pomocou vzťahu (7.1), kde po vyčíslení dostávame:

$$P(TP) = 5/(5 + 2) = 0,714 = 71,4\%.$$

Špecificitu následne vypočítame podľa vzťahu (7.2) a po vyčíslení dostávame:

$$P(FP) = 28/(25 + 28) = 0,528 = 52,8\%.$$

Ak následne zarátame aj hlásky na ktoré sa otestovala neurónová sieť v kapitole 7.6.2, dostaneme:

	Správne klasifikované	Nesprávne klasifikované
Hlásky s nesprávnou výslovnosťou	28+15 = 43	2+0 = 2
Hlásky so správnou výslovnosťou	5+20 = 25	25+3 = 28

Tab. 7.4: Výsledok automatickej klasifikácie slov spolu s testom neurónovej siete

Podľa tabuľky 7.4 sa nám úspešnosť správnej klasifikácie zväčšila na hodnotu 69,38%. Hodnotu senzitivity vypočítanú podľa vzťahu (7.1), dostávame:

$$P(TP) = 25/(25 + 2) = 0,926 = 92,6\%.$$

Špecificitu vypočítame podľa vzťahu (7.2) a dostávame:

$$P(FP) = 43/(28 + 43) = 0,605 = 60,5\%.$$

7.8 Zhodnotenie výsledkov

Ako môžeme vidieť u výsledkov z predchádzajúcich kapitol, ak sa neurónová sieť sama testovala na nahrávkach slov alebo hlások, dokázala sa otestovať s 84,3% úspešnosťou pre slová a 93,7% pre hlásky. Teda pri samootestovaní sa so 108 slovami mala 84,3% úspešnosť klasifikovania, a pri samootestovaní sa so 111 hláskami mala 93,7% úspešnosť klasifikovania správnej alebo nesprávnej výslovnosti. Tieto výsledky dosahujú celkom vysokú úspešnosť a boli by vyhovujúce. Taktiež senzitivita dosahujúca 91% u slov a 98% u hlások je veľmi dobrá, špecificita 77,3% u slov je menej vyhovujúca no u hlások je opäť veľmi dobrá a to 90,2%.

Naopak pri klasifikácií, kedy sme neurónovú sieť natrénovali s použitím 298 slov správnej a nesprávnej výslovnosti, úspešnosť automatickej klasifikácie výslovnosti 60 slov klesla na 60%. 68,7% senzitivita a 56,8% špecificita nie je príliš vyhovujúca. Ak by sme zahrnuli do toho aj samootestovanie sa neurónovej siete, dostali by sme sa na úspešnosť 74,22%, čo je viac menej prijateľný výsledok. Senzitivita stúpla na dobrých 83,8% a špecificita na 68,3%. Ak sme neurónovú sieť natrénovali s 308 hláskami správnej a nesprávnej výslovnosti, úspešnosť automatickej klasifikácie 60 hlások klesla na 55%. Očakávali sme o veľa lepší výsledok, no ten ktorý sme dostali nie je vyhovujúci. Senzitivita testovania bola 71,4% a špecificita veľmi nízka 52,8%. So zahrnutím aj samootestovania neurónovej siete, úspešnosť klasifikácie stúpla na 69,38%, s vysokou senzitivitou 92,6% no nízkou špecificitou 60,5%.

S využitím všetkých nahrávok, z ktorých sme 65% využili na tréovanie a 25% na samootestovanie neurónovej siete sme získali veľmi uspokojivé výsledky. Následne pri automatickej klasifikácii sme dosahovali o veľa menšie výsledky úspešnosti. Tieto výsledky mohli byť spôsobené nekvalitnejšími nahrávkami správnej výslovnosti, ktoré som získal za pomoci priateľov, ktorý boli už dospelý. Moje nahrávky nedosahovali takú kvalitu ako nahrávky pacientov logopedickej kliniky, pretože som nemal k dispozícii kvalitnejší mikrofón. Tento problém by sa vyriešil ak by bolo poskytnutých viac správnych nahrávok od detí z logopedickej kliniky. U ručnej segmentácie hlások „R“ mohlo prísť tiež ku chybám, keďže som ich segmentoval väčšinou podľa sluchu. Taktiež rozdiel úspešnosti medzi samootestovaním a automatickou klasifikáciou, je rapídny a neviem si vysvetliť takto veľký rozdiel, keďže úprava aj výpočet MFCC z nahrávky je rovnaký v oboch skriptoch. Môže to byť spôsobené testovacou sadou nahrávok pri samootestovaní a pri automatickej klasifikácii. Tento problém by sa dal vyriešiť navýšením počtu nahrávok so správnou aj nesprávnou výslovnosťou.

8. Postup nácviku a kontroly výslovnosti hlásky „R“

Nácvik výslovnosti vyžaduje komplexnú, nielen logopedickú starostlivosť. Najvhodnejšou voľbou je začať s nápravou výslovnosti ešte pred nástupom dieťaťa do školy, kedy sa môže viac sústrediť na logopedický nácvik, ako počas školy keď začne mať viac povinností.

Dôležitou časťou pred samotným nácvikom správnej výslovnosti hlásky sú prípravné cvičenia. Cvičenia sú väčšinou aplikované formou hry so zapojením hlasu. Medzi prípravné cvičenia sa radia dychové cvičenia, ktorých cieľom je aby sa dieťa naučilo a dokázalo správne vdychovať a vydychovať pri komunikácii. Sú rôzne typy dychových cvičení, napríklad fúkanie do ľahkých papierových guľčiek, hovorenie básničky na jeden výdych a iné. Ďalším typom cvičenia je cvičenie pre rozvoj motoriky artikulačných orgánov, kde sa kladie dôraz na pohyblivosť pier, čeľuste, jazyka. Toto cvičenie tiež aplikujeme formou hry a využívame napodobňovanie rôznych zvukov. Rozvoj fonematického sluchu je ďalším typom cvičenia, umožňuje rozpoznávať jednotlivé slabiky a hlásky. U menších detí najskôr precvičujeme orientáciu sluchom v priestore a podľa zvuku poznávať predmety. U starších detí využívame rozpoznávanie rozdielov v trvaniach tónov, dĺžke samohlások a následne rozpoznávať jednotlivé hlásky v slovách.

Metódy nácviku správnej výslovnosti hlásky sú tri. Prvou metódou je substitučná, v ktorej k navodeniu správnej artikulácie sa využívajú hlásky iné. Aby jedinec dokázal správne vyslovovať hlásku „R“, musí zvládnuť výslovnosť alveolárnych hlások „T a D“. Tieto hlásky majú rovnaké miesto artikulácie ako hláska „R“, no nedochádza pri nich ku kmitom jazyka. Ďalšou takou hláskou je hláska „L“, ktorá svojou akustickou príbuznosťou a špecifickosťou tvorenia je príbuzná hláske „R“. Druhá metóda je mechanická, pri ktorej sa využívajú mechanické prostriedky. Základom tejto metódy je hláska „D“, ktorá sa vyslovuje predĺžene a tak sa rozkmitáva hrot jazyka. Kombinovaná metóda je poslednou a spojuje sa pri nej metóda substitučná aj mechanická.

Navrhnuť vlastnú metódu precvičovania je neľahká úloha aj pre logopéda. Využívajú sa metódy popísané vyššie, ktoré sú vyskúšané a prinášajú svoje výsledky správnej výslovnosti, poprípade si logopéd s dlhoročnou praxou vie tieto metódy vylepšiť. Pri zlej výslovnosti rotacizmov je logopedická náprava za pomoci logopéda nutná, pretože náprava bez pomoci logopéda nie je vo väčšine prípadov úspešná.

V tejto práci som vyskúšal vytvoriť program ako pomôcku ku klasifikácii správnej výslovnosti hlásky „R“. Výsledky však ukázali, že je to neľahká úloha. Kontrola výslovnosti teda aj naďalej spočíva v jej zhodnotení logopédom, poprípade osobou, ktorá pomáha dieťaťu s vyvedením správnej výslovnosti hlásky.

Záver

V dnešnej dobe nie sú 100% úspešné žiadne metódy, ktoré by dokázali rozpoznávať ľudskú reč, alebo detekovať logopedické vady výslovnosti. Stanoviť rozdiel medzi správnou a zlou výslovnosťou je neľahká úloha aj pre logopéda.

Nesprávna výslovnosť vibranty „R” sa nazýva dyslália a má rôzne príčiny vzniku. Najčastejším prípadom je velárny rotacizmus, čo má za následok račkovanie. Iným prípadom je mogirotacizmus kedy dieťa vibrantu vynecháva a pararotacizmus kedy ju zamieňa za inú hlásku. Všetky typy nesprávnej výslovnosti vibranty majú rušivý vplyv v prejave dieťaťa, a teda je ich nutné správne klasifikovať a následne výslovnosť upraviť v spolupráci s logopédom.

Cieľom tejto práce bolo navrhnúť metódu pre klasifikáciu výslovnosti hlásky „R“ a následne ju aj realizovať. Nahrávky boli poskytnuté z kliniky LOGO, ktoré museli byť spracované a doplnené o vlastné. Pre analýzu rečového signálu boli naštudované rôzne prístupy, ktoré sú popísané v tejto práci. Z týchto prístupov bola zvolená MFCC, pretože do určitej miery rešpektujú nelineárne vlastnosti vnímania zvukov ľudským uchom a teda sa snažia kompenzovať nelineárne vnímanie frekvencií. Ako klasifikátor bola zvolená neurónová sieť pre svoju robustnosť pri riešení podobných metód. Metódu sa mi podarilo zrealizovať, popis funkcie programu je popísaný v práci. Taktiež bola metóda otestovaná na dostatočnom počte nahrávok slov aj ručne nasegmentovaných hlások, keďže segmentácia slova na hlásky je náročná a nepodarilo sa mi ju úspešne zrealizovať. Výsledky testovania boli úspešné ak sa neurónová sieť otestovala sama, a to konkrétne správna klasifikácia v 84,3% nahrávok slov a 93,7% nahrávok hlások. Ak však boli predložené neurónovej sieti nahrávky, ktoré predtým neboli v tréningovej množine a mali sa klasifikovať, úspešnosť výrazne klesla na 60% a 55%.

S dosiahnutými výsledkami sa dá ďalej pracovať. Pre zlepšenie úspešnosti klasifikácie by bolo vhodné navýšenie počtu nahrávok slov, akusticky kvalitných, u ktorých by nebol problém so šumom, poprípade využiť inú metódu alebo kombináciu metód pre analýzu signálu.

Použitá literatura

- [1] ČERNOCKÝ, J. *Zpracování řečových signálů*. 2006.
- [2] ČERNOCKÝ, J. *Rozpoznávání řeči – HMM*. Dostupné z http://www.fit.vutbr.cz/study/courses/ZRE/public/pred/09_hmm/09_hmm.pdf .
- [3] CUNNINGHAM, P., DELANY, S. J. *K-Nearest Neighbour Classifiers*. 2007. Dostupné z <https://pdfs.semanticscholar.org/60f3/89f056ae250d896eb80b414a933537d51d6c.pdf> .
- [4] DELLER, J., HANSEN, J., PROAKIS J. *Discrete-Time Processing of Speech Signals (IEEE Press Classic Reissue)*. ISBN: 978-0-7803-5386-2.
- [5] HERTZ, J., KROGH, A., PALMER, R. G. *Introduction to the Theory of Neural Computation*, Addison-esley. 1991. ISBN:0-201-50395-6.
- [6] KEJKLÍČKOVÁ, I. *Vady řeči u dětí : návody pro praxi*. Praha : Grada, 2016. ISBN 978-80-247-3941-0.
- [7] KLENKOVÁ, J. *Logopedie: narušení komunikační schopnosti, logopedická prevence, logopedická intervence v ČR, příklady z praxe*. Praha: Grada, 2006. ISBN 978-80-247-1110-2.
- [8] KOTEK, Z. MAŘÍK, V. HLAVÁČ, V. PSUTKA, J. ZDRÁHAL, Z. *Metody rozpoznávání a jejich aplikace*. Praha : Academia. ISBN 80-200-0297-9.
- [9] KOZUMPLÍK, J., KOLÁŘ, R., JAN, J. *Číslicové zpracování signálů v prostředí Matlab*. Skripta FEKT VUT v Brně, 2001.
- [10] LOGAN, B. *Mel Frequency Cepstral Coefficients for Music Model*. Cambridge, 2000.
- [11] Mike Brookers (4. Máj 2007). *enframe.m (verzia 1.5) [Software]*. Dostupné z <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html> .
- [12] Mohit (27. septembra 2015). *Verifying MFCC Final Result*. Príspevok zaslaný na <https://dsp.stackexchange.com/questions/> .
- [13] NEUBAUER, K. *Logopedie učební text pro bakalářské stadium speciální pedagogiky*. Univerzita Hradec Králové : GAUDEAMUS, 2007. ISBN 978-80-7041-093-6.

- [14] NILSSON, M., EINARSSON, M. *IEEE Signal Processing Magazine*. Blekinge Institute of Technology, 2002.
- [15] PSUTKA, J., MULLER, L., MATOUŠEK, J. *Mluvíme s počítačem česky*. ACADEMIA, 2006. ISBN 80-200-1309-1
- [16] SIGMUND, M. *Analýza řečových signálů*. Skripta FEKT VUT v Brně, 2000.
- [17] SMÉKAL, Z. *Číslicové zpracování signálů*. Skripta FEKT VUT v Brně, 2006.
- [18] SOVÁK, M. *Logopedie předškolního věku*. Praha: Státní pedagogické nakladatelství, 1986. SPN 0-72-19 2.
- [19] ŠVENDA, Z. MICHÁLEK, L. *Automatické metody rozpoznávání řečníka ve forenzní praxi*. Praha : Policejní akademie České republiky v Praze, 2015. ISBN 978-80-7251-431-1.
- [20] TUČKOVÁ, J. *Úvod do teorie a aplikací umělých sítí*. Praha : ČVUT, 2003. ISBN 80-01-02800-3.
- [21] UHLÍŘ, J. SOVKA, P. POLLÁK, P. HANŽL, V. ČMEJLA, R. *TECHNOLOGIE HLASOVÝCH KOMUNIKACÍ*. Praha : České vysoké učení technické v Praze, 2007. ISBN 978-80-01-03888-8.
- [22] WATANABE K., WATANABE S. *Stochastic Complexities of Gaussian Mixtures in Variational Bayesian Approximation*. Journal of Machine Learning Research 7 625–644, 2006. ISBN 978-0-9719777-4-7.
- [23] WAIBEL, A. *Phoneme Recognition Using Time-Delayed Neural Networks*. ET AL, 1990. ISBN:1-55860-124-4.
- [24] ZIMMERMANN, J. *Spektografická a škálografická analýza akustického, řečového Signálu*. Prešov : Náuka, 2002. ISBN 80-89038-22-0.

Zoznam skratiek

- CNS – centrálna nervová sústava
- DP – dolnopriepustný
- HP – hornopriepustný
- FIR – filter s konečnou impulznou odozvou (Finite Impulse Response)
- CFT – spojitá Fourierova transformácia (Continuous Fourier Transform)
- DFT – diskretná Fourierová transformácia (Discrete Fourier Transform)
- FFT – rýchla Fourierova transformácia (Fast Fourier Transform)
- DCT – diskretná kosínusová transformácia (Discrete Cosine Transform)
- CWT – spojitá vlnková transformácia (Continuous Wavelet Transform)
- DWT – diskretná vlnková transformácia (Discrete Wavelet Transform)
- LPC – lineárne prediktívna analýza (Linear Predictive Coding)
- PLP – percepčne lineárne prediktívna analýza (Perceptual Linear Prediction)
- MFCC – Mel-frekvenčné keprálne koeficienty (Mel-frequency Cepstral Coefficients)
- K-NN – algoritmus k-najbližších susedov (K-nearest Neighbors)
- GMM – Gaussov zmiešaný model (Gaussian Mixture Models)
- EM – Expectation Maximization
- HMM – skryté Markovove modely (Hidden Markov Models)

Obsah priloženého CD

Priložené CD obsahuje 2 zložky : klasifikácia hlások R a klasifikácia slov.

V prvej zložke sa nachádzajú ručne nasegmentované hlásky, natrénované neurónové siete pre (net), skript pre vypočítanie MFCC z hlások pre neurónovú sieť (mfcc_pre_nn.mat) a skripty programu k automatickej klasifikácii hlások (nacitanie_nahravky.mat , nahravanie_nahravky.mat).

V druhej zložke sa nachádzajú slová, natrénované neurónové siete (net), skript pre vypočítanie MFCC zo slov pre neurónovú sieť (mfcc_pre_nn.mat) a skripty programu k automatickej klasifikácii slov (nacitanie_nahravky.mat , nahravanie_nahravky.mat).