VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

KLASIFIKACE PŘÍKAZŮ Z EMG POMOCÍ NEURONOVÉ SÍTĚ

COMMAND CLASSIFICATION FROM EMG USING NEURAL NETWORK

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE **JÁN ZAUŠKA**
AUTHOR

VEDOUCÍ PRÁCE **Ing. IGOR SZŐKE, Ph.D.**
SUPERVISOR

BRNO 2020

Vysoké učení technické v Brně
Fakulta informačních technologií

Ústav počítačové grafiky a multimédií (UPGM) Akademický rok 2019/2020

Zadání bakalářské práce

Student: **Zauška Ján**
Program: Informační technologie
Název: **Klasifikace příkazů z EMG pomocí neuronové sítě**
Command Classification from EMG Using Neural Network
Kategorie: Zpracování řeči a přirozeného jazyka

Zadání:

1. Nastudujte základy snímání aktivity svalů (EMG). Nastudujte dodaná data (záznamy EMG pro cca 16 hlasitě/šeptem vyslovených příkazů a vět). Nastudujte dostupné knihovny a literaturu o EMG. Nastudujte základy NN.
2. Definujte trénovací a testovací sadu. Definujte metriku pro měření úspěšnosti. Vyberte nebo navrhněte metody pro klasifikaci či verifikaci příkazů z EMG za použití NN.
3. Implementujte zvolené metody, natrénujte a vyhodnoťte NN. Experimentujte a zlepšete navržené metody.
4. Zhodnoťte výsledky a navrhněte směry dalšího vývoje.
5. Vyrobte A2 plakátek a cca 30 vteřinové video prezentující výsledky vaší práce.

Literatura:

- Dle pokynů vedoucího

Pro udělení zápočtu za první semestr je požadováno:

- Body 1, 2 a část bodu 3 ze zadání.

Podrobné závazné pokyny pro vypracování práce viz https://www.fit.vut.cz/study/theses/

Vedoucí práce: **Szőke Igor, Ing., Ph.D.**
Vedoucí ústavu: Černocký Jan, doc. Dr. Ing.
Datum zadání: 1. listopadu 2019
Datum odevzdání: 28. května 2020
Datum schválení: 29. dubna 2020

Abstrakt

Táto práca sa zaoberá klasifikáciou 15 príkazov (krátkych slov) z malej dátovej sady nahranej pomocou sEMG elektród umiestnených na tvári a krku rečníka. V nahrávkach sú rozlíšené dva typy reči – audible speech, čo je klasická reč, a silent speech, teda reč, pri ktorej je potlačené vydávanie zvuku. Práca popisuje spracovanie EMG signálu, extrakciu príznakov, návrh klasifikátoru a výsledky klasifikácie. Ako klasifikátor bola použitá vlastná architektúra konvolučnej neurónovej siete. V práci sa tiež nachádza mnoho experimentov porovnávajúcich presnosť klasifikácie silent a audible speech.

Abstract

This work deals with classification of 15 commands (short words), from small dataset recorded by sEMG electrodes placed on face and neck of speaker. Two types of speech are differentiated in recordings – audible speech, what is classic speech and silent speech, hence speech, in which sound output is suppressed. This work describes EMG signal processing, feature extraction, classifier design and classification results. The convolutional neural network architecture was used as a classifier. There are a lot of experiments in this work that compare the classification accuracy of silent and audible speech.

Kľúčové slová

EMG, AI, CNN, Elektromyografia, Silent speech, Neurónové siete, Konvolučné neurónové siete

Keywords

EMG, AI, CNN, Electromyography, Silent speech, Neural Networks, Convolutional Neural networks

Citácia

ZAUŠKA, Ján. *Klasifikace příkazů z EMG pomocí neuronové sítě*. Brno, 2020. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Igor Szőke, Ph.D.

Klasifikace příkazů z EMG pomocí neuronové sítě

Prehlásenie

Prehlasujem, že som túto bakalársku prácu vypracoval samostatne pod vedením pána Ing. Igora Szőkeho, Ph.D. Uviedol som všetky literárne pramene, publikácie a dalšie zdroje, z ktorých som čerpal.

.....
Ján Zauška
27. mája 2020

Poďakovanie

Ďakujem svojmu vedúcemu Ing. Igorovi Szőkemu, Ph.D. za jeho čas, cenné rady a usmerňovanie pri tvorbe tejto práce.

Obsah

Kapitola 1

Úvod

Komunikácia je každodennou súčasťou nášho života. Za posledné roky sa výrazne zmenila a na svedomí to má hlavne rýchly vývoj technológií. V dnešnej dobe je možné komunikovať s ľuďmi po celom svete v priebehu sekúnd, čo bolo niekedy nepredstaviteľné. Tam ale vývoj zďaleka nekončí. Naopak, stále sa objavujú nové spôsoby komunikácie.

V posledných rokoch bolo uskutočnených niekoľko výskumov zameraných na rozpoznávanie reči z elektromyografického (EMG) signálu meraného pomocou EMG elektród umiestnených na tvári a krku rečníka. Ich motiváciou sú potenciálne široké možnosti použitia vo viacerých oblastiach. V medicíne ide o pomoc ľuďom s rečovým postihnutím, ktorým by mohla komunikácia pomocou EMG výrazne uľahčiť život. Významné uplatnenie sa naskytá aj v oblasti komunikácie v hlučnom prostredí a pod vodou. EMG signál je meraný priamo z rečníkovho tela, teda zvuky z okolia naň nemajú vplyv. Na druhej strane, existujú aj situácie, kde je potrebné zachovať ticho, čo je pri bežnej reči veľmi obtiažne. Okrem toho občas nie je vhodné používať hlasitú komunikáciu z dôvodu ochrany súkromia. Ďalším z potenciálnych využití je ovládanie inteligentných zariadení. Práve využitie pri ovládaní inteligentných zariadení sa zdá byť veľmi sľubné, keďže k tomu stačí oproti bežnej komunikácii použitie malej slovnej zásoby.

Práve klasifikácia malej slovnej zásoby, presnejšie 15 príkazov s aritmetickým významom, je témou tejto práce. Všetky potrebné dáta mi boli poskytnuté z predošlého výskumu *Towards Continuous Silent Speech Synthesis from Non-Invasive Bio-Physiological Activity* (ďalej predchádzajúci výskum). Jeho technická správa sa nachádza na priloženom pamäťovom médiu, keďže ešte nebola publikovaná. Síce sa výskum primárne zameriaval na syntézu reči, no pokúšali sa v ňom aj o klasifikáciu.

V kapitole 2 budú zhrnuté informácie o rozpoznávaní reči z EMG signálu a existujúci výskum. Kapitola 3 sa bude zaoberať priblížením neurónových sietí. Viac sa zameriam na konvolučné neurónové siete, ktoré budú použité pre klasifikáciu. Po viac teoreticky zameranej časti nasleduje kapitola 4, popisujúca dostupnú dátovú sadu a spôsoby jej rozdelenia pre trénovanie a testovanie neurónovej siete. Ešte predtým je potrebné si dodané dáta spracovať. Kapitola 5 sa preto zaoberá spracovaním EMG signálu a následne extrakciou príznakov z neho. Tie budú slúžiť ako vstup klasifikátoru navrhnutého v kapitole 6. Na záver sú v kapitole 7 zhodnotené výsledky. Tiež sa tu nachádza pár zaujímavých experimentov a návrh smerov ďalšieho vývoja.

Kapitola 2

Rozpoznávanie reči z EMG signálu

V tejto kapitole budú na začiatku stručne popísané rozdielne typy reči a metóda EMG použitá pre ich meranie. Neskôr bude zhrnutý existujúci výskum v oblasti rozpoznávania reči z EMG signálu a najčastejšie používané metódy. Tiež sa tu nachádza popis predchádzajúceho výskumu spomenutého v úvode, ktorý používal rovnakú dátovú sadu.

2.1 Typy reči

Pod pojmom reč si väčšina ľudí predstaví typ, ktorý je počuť. To je len jeden typ reči, nazývaný *audible speech*. Reč je možné rozdeliť podľa úrovne zapojenia rečových orgánov[1] až na štyri typy. Okrem *audible speech* ide o *silent speech*, *motor imagery speech* a *inner speech*. V tejto sekcii sú stručne popísané. Pre označenie typov reči sú použité anglické názvy, keďže v slovenčine ešte neexistujú ich zaužívané preklady. Táto práca sa zaoberá iba rečou typu *silent* a *audible*, ale pre zaujímavosť krátko spomeniem aj zvyšné dva typy.

2.1.1 Audible speech

Audible speech je klasická reč. Ide o jediný zo spomenutých typov reči, ktorý je počuť. K vytvoreniu zvuku pri artikulácii dochádza prúdom vzduchu vychádzajúcim z pľúc, prechádzajúcim pomedzi hlasivky. Artikuláciou sa myslí použitie jazyka, pier, sánky a ostatných rečových orgánov (artikulátorov). [8]

2.1.2 Silent speech

Pri silent speech rečník artikuluje rovnako ako pri audible speech, teda aj aktivita svalov použitých pri reči by mala byť totožná. Rozdiel je v tom, že rečník potlačí prúd vzduchu vychádzajúci z pľúc, čím zamedzí vydávaniu zvuku. Pri tomto type reči často dochádza k problému, že rečník nedokáže vysloviť daný text bez vydania zvuku, ale je prítomný tichý šepot. Keď sa to aj rečníkom podarilo, mnoho z nich uviedlo, že ich artikulácia sa oproti audible speech výrazne zmenila. [23]

2.1.3 Motor imagery speech

Motor imagery speech je typ reči, pri ktorom nedochádza k pohybu artikulátorov. Sú však stimulované mentálne (rečník myslí na ich používanie). [15]

[1] Orgány, ktoré sa podieľajú na tvorbe reči

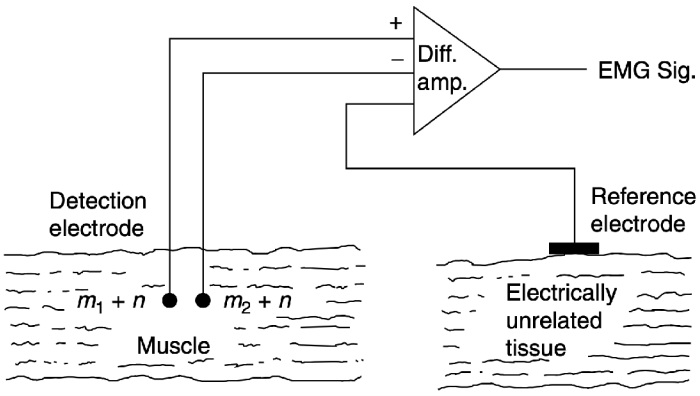
2.1.4 Inner speech

Inner speech sa taktiež označuje aj ako vnútorný monológ. Človek ho používa štvrtinu času, kedy je pri vedomí. Dochádza k nemu napríklad vtedy, keď je potrebné si udržať myšlienku v krátkodobej pamäti. Taktiež hrá dôležitú úlohu pri plánovaní, riešení problémov, ale aj pri písaní či čítaní. [19]

2.2 Elektromyografia

Elektromyografia (EMG) [5] sa zaoberá detekciou, analýzou a použitím elektrického signálu, ktorý vzniká pri kontrakcii svalov. Tento signál sa nazýva EMG signál. Reprezentuje elektrický prúd generovaný prúdom iónov prechádzajúcim cez membrány svalových vlákien, ktorý sa propaguje cez priľahlé tkanivá až k povrchu detegujúcej elektródy. Je dôležité si pamätať, že výsledný signál je funkciou ako membrány svalových vlákien, tak aj prístroja použitého na jeho získanie. Elektrická aktivita vo vnútri svalu alebo na povrchu kože môže byť ľahko získaná umiestnením elektródy na toto miesto. Pre meranie je potrebné umiestniť druhú, referenčnú elektródu, do prostredia, ktoré je elektricky tiché alebo obsahuje elektrický signál nezávislý na meranom signáli. Umiestnenie elektród pri meraní je znázornené na obrázku 2.1.

Ako bolo spomenuté, svalovú aktivitu je možné merať buď v jeho vnútri alebo na povrchu. Prvá z metód, kedy je elektróda pripevnená k ihle, ktorá je aplikovaná priamo do svalu, sa nazýva ihlová elektromyografia. Druhou metódou je povrchová elektromyografia, označovaná aj sEMG (surface EMG). Pri tejto metóde je elektróda pripevnená k povrchu kože blízko meraného svalu. Na rozdiel od ihlovej elektromyografie ide o neinvazívnu metódu.



Obr. 2.1: Meranie EMG signálu pomocou bipolárnej elektródy (prevzaté z [5]).

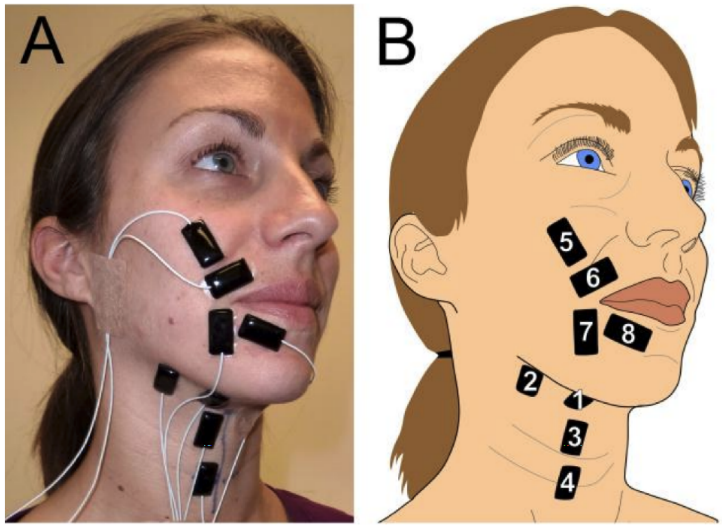
Pri EMG je kvalita výsledného signálu ovplyvnená vonkajšími a vnútornými faktormi. Vonkajším faktorom sa venuje zvýšená pozornosť z dôvodu, že je možné ich ľahko ovplyvniť. Medzi ne patrí napríklad poloha elektród vzhľadom na nameraný sval, vzdialenosť a veľkosť elektród, kontakt elektródy s kožou a externý šum. K vnútorným faktorom patrí svalová aktivita meraného svalu, aktivita okolitých svalov a elektrická aktivita iných tkanív (napríklad EKG signály[2]).

[2]Elektrokardiografické signály

2.3 Existujúci výskum

V oblasti rozpoznávania reči z EMG bolo uskutočnených viacero výskumov. Tejto téme sa výskumníci venujú približne od 80. rokov, kedy prebehol jeden z prvých výskumov v tejto oblasti [20]. Z množstva existujúcich výskumov som sa zameral hlavne na výskumy používajúce dáta namerané podobným spôsobom.

Vo výskume [6] bolo na meranie použitých osem EMG elektród strategicky umiestnených na povrchu tváre a krku. Toto rozloženie je zobrazené na obrázku 2.2.



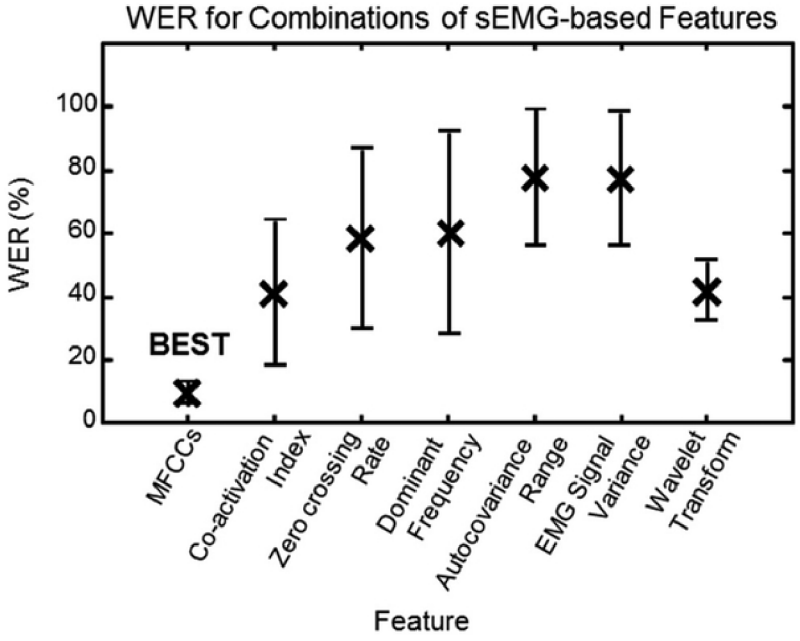
Obr. 2.2: Rozmiestnenie EMG elektród použité vo výskume [6].

Z EMG signálu boli extrahované MFCC[3] použité pre klasifikáciu vyslovených slov. Ako klasifikátor bol použitý GMM[4]. Nachádza sa tu porovnanie výsledkov pri rozdelení signálu na rámce rôznej dĺžky v kombinácii s rôznou dĺžkou, o ktorú sa po sebe idúce rámce prekrývajú (overlap). Z testovaných kombinácií v rozsahu od 20 ms do 40 ms najlepšie výsledky dosiahla u prvého rečníka dĺžka rámca 40 ms a overlap 20 ms, v druhom prípade dĺžka rámca 40 ms a overlap 25 ms. Z toho bol vyvodený záver, že optimálne hodnoty pre rôznych rečníkov sa môžu líšiť. Okrem toho konštatujú, že pri rozpoznávaní reči z EMG signálu je vhodné použiť väčšiu dĺžku rámca oproti rozpoznávaniu reči z audia. Vo výskume sa tiež uvádza, že pri použití štyroch EMG elektród namiesto ôsmich došlo len k minimálnemu poklesu v úspešnosti klasifikácie.

Výskum [17] sa taktiež zaoberal rôznym počtom EMG elektród. Najlepšie výsledky boli dosiahnuté použitím ôsmich elektród. V tomto výskume sa bohužiaľ nenachádza použité rozmiestnenie elektród, takže ich nie je možné porovnať s výskumom [6]. Je možné, že pri rozmiestnení elektród na obrázku 2.2 boli jednotlivé elektródy umiestnené blízko seba a zachytávali signály z rovnakých svalov, preto zníženie ich počtu výrazne neovplyvnilo presnosť klasifikácie. Výskum [17] porovnáva tiež použitie rôznych typov príznakov získaných z EMG signálu. Z porovnania kombinácií siedmich druhov príznakov zobrazených na obrázku 2.3

[3] Mel frekvenčné kepstrálne koeficienty
[4] Gaussian Mixture Model

je vidieť, že MFCC príznaky dosahujú výrazne lepšie výsledky ako ostatné typy. Výskum na klasifikáciu používa GMM a HMM[5].



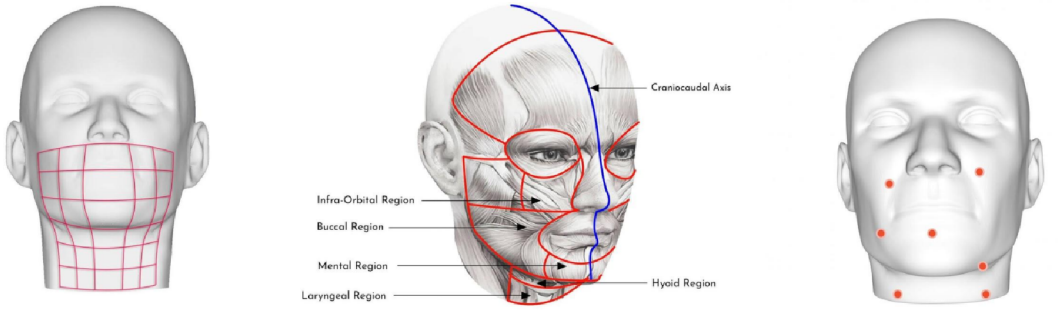
Obr. 2.3: WER[6]pri použití rôznych kombinácií príznakov. X reprezentuje priemernú hodnotu pri kombinácií rôznych typov príznakov (prevzaté z [17]).

Vo výskume [16] bola skúmaná podobnosť dát nahraných v rôznych sedeniach. Vzhľadom na pokles presnosti z 97,3 % pri použití dát z jedného nahrávacieho sedenia až na 76,2 % pri použití dát z rozdielnych sedení usúdili, že umiestnenie EMG elektród na presne rovnaké miesta v rôznych sedeniach je nesmierne dôležité. Pri použití dát z omnoho viac sedení došlo k zlepšeniu presnosti na 84,1 %. Nachádza sa tu tiež experiment s vplyvom typu (silent a audible) reči na výslednú presnosť. Z testovaných troch rečníkov dvaja dosahovali pri audible reči lepšie výsledky. Pripisujú to faktu, že tretí rečník mal najviac skúseností so silent speech. Taktiež si všimli, že s pribúdajúcimi skúsenosťami rečníkov sa presnosť na silent speech dátach zvyšovala.

A. Kapur vo svojej práci [11] používa na klasifikáciu silent speech konvolučnú neurónovú sieť. Ako jej vstup sú použité MFCC príznaky. Použité rozmiestnenie siedmich EMG elektród znázornené na obrázku 2.4 je takmer totožné s rozložením v predchádzajúcom výskume, keďže z neho vychádzalo.

V rámci práce bolo navrhnuté a zostrojené aj zariadenie *AlterEgo* so zabudovanými EMG elektródami, ktoré sa nachádza na obrázku 2.5. Tým, že EMG elektródy sú v ňom pevne umiestnené, by sa pri opätovnom používaní mala zachovať ich vzájomná poloha a tiež poloha vzhľadom k rečníkovi používajúcemu toto zariadenie. Motiváciou pre návrh tohto zariadenia bolo podľa autora priblížiť sa prepojeniu človeka a technológií. Konktrétne ho označuje ako IA (Intelligence-Augumentation) zariadenie.

[5]Hidden Markov Model
[6]Word Error Rate



Obr. 2.4: Rozloženie EMG elektród použité vo výskume [11].

Obr. 2.5: Zariadenie AlterEgo navrhnuté vo výskume [11].

Ďalšou výhodou oproti konkurenčným zariadeniam je podľa autora fakt, že AlterEgo dosahuje vysokú presnosť aj v prípade, že rečník pri rozprávaní neotvára ústa. Uvádzaná presnosť rozpoznávania číslic dosahuje až 92 %. Autor medzičasom vydal aj aktualizovanú verziu zariadenia zobrazenú na obrázku 2.6. Aktualizovaná verzia už používa namiesto pôvodných siedmich EMG elektród len štyri, čo pomohlo hlavne z praktického hľadiska.

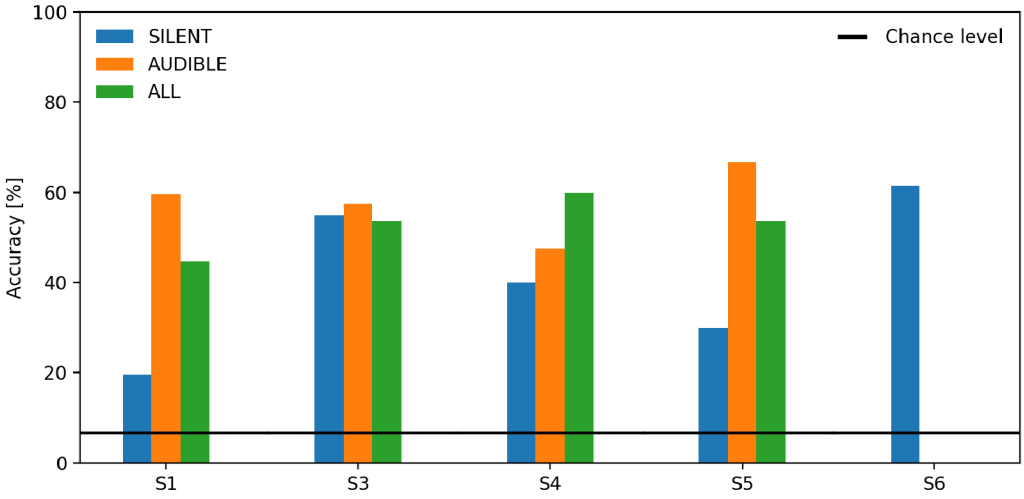
Obr. 2.6: Novšia verzia zariadenia AlterEgo.[7]

[7] http://news.mit.edu/2018/computer-system-transcribes-words-users-speak-silently-0404

2.3.1 Predchádzajúci výskum

Ako som spomenul v úvode, v tejto oblasti bol uskutočnený výskum nazvaný *Towards Continuous Silent Speech Synthesis fromNon-Invasive Bio-Physiological Activity*. Podieľali sa na ňom výskumníci z organizácií Vysoké Učení Technické v Brně, Honeywell Labs a NorthEastern University.

Primárnym cieľom výskumu nebola klasifikácia príkazov ale syntéza reči. Aj napriek tomu sa tu nachádza pokus o klasifikáciu príkazov použitím metódy najbližšieho suseda na základe korelácie. Výsledky klasifikácie sú zobrazené na obrázku 2.7. Vzhľadom na to, že v tejto práci používam rovnaké dáta, budem sa snažiť tieto výsledky zlepšiť použitím neurónovej siete pre klasifikáciu.



Obr. 2.7: Porovnanie dosiahnutej presnosti klasifikácie pre jednotlivé sedenia S1 – S6 (prevzaté z predchádzajúceho výskumu).

Ďalšie pokusy, ktorých výsledky by mohli byť použiteľné v tejto práci, sa nachádzajú v prílohe *Command Classification Task*. V týchto pokusoch boli použité MFCC príznaky extrahované z EMG signálu. Na klasifikáciu tu bolo použité skóre podobnosti nahrávok získané pomocou DTW[8] s kosínusovou vzdialenosťou na meranie podobnosti medzi MFCC príznakmi. Po vypočítaní podobnosti pre všetky páry nahrávok bola zvolená prahová hodnota, zostrojená ROC[9] krivka a vyrátaná hodnota AUC[10].

Prvým z pokusov bola snaha o redukciu počtu MFCC príznakov extrahovaných z EMG. K tomu bola použitá neurónová sieť natrénovaná s MFCC príznakmi z EMG na vstupe a MFCC príznakmi z audia na výstupe. Pri redukcii na 5 príznakov došlo oproti použitiu pôvodných 160 príznakov k zhoršeniu výsledkov – presnejšie došlo k poklesu hodnoty AUC z 0,69 na 0,60.

Posledným pokusom bolo porovnanie rôznej dĺžky rámca použitej pri rozdelení EMG signálu. V porovnaní dĺžky rámca o veľkosti 20 ms a 100 ms dosiahli obe veľkosti približne rovnaké výsledky. Hodnota AUC pre 100 ms bola 0,60 a pre 20 ms 0,59.

[8] Dynamic Time Warping
[9] Receiver operating characteristic
[10] Area Under Curve

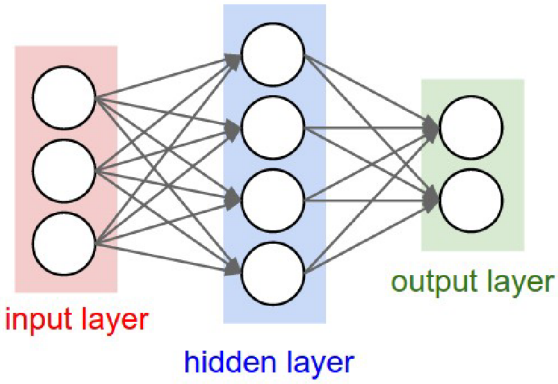
Kapitola 3

Umelé neurónové siete

V tejto kapitole sa nachádza krátky popis umelých neurónových sietí (ďalej neurónové siete) a ich základného stavebného prvku – umelého neurónu (ďalej neurón). Podrobnejšie sa venuje konvolučným neurónovým sieťam, ktoré budú použité ako klasifikátor. Tiež je tu spomenutý autoenkóder, pomocou ktorého bola v rámci jedného experimentu zredukovaná dimenzionalita vstupných dát klasifikátoru.

3.1 Štruktúra neurónovej siete

Neurónová sieť je zložená z veľkého množstva umelých neurónov. Tie sú usporiadané do vrstiev a vzájomne poprepájané. Jednotlivé vrstvy sa môžu skladať z ľubovoľného počtu neurónov. Neurónová sieť môže obsahovať rôzny počet vrstiev. Každá neurónová sieť obsahuje vstupnú a výstupnú vrstvu. Väčšina medzi nimi navyše obsahuje minimálne jednu skrytú vrstvu. Označenie skrytá vrstva znamená iba to, že nejde o vstupnú ani výstupnú vrstvu. Počet vrstiev neurónovej siete je nazývaný *hĺbka siete*. Do hĺbky siete sa nepočíta vstupná vrstva, takže neurónová sieť so vstupnou vrstvou, jednou skrytou a výstupnou vrstvou znázornená na obrázku 3.1 má hĺbku dva. Na obrázku je konkrétne zobrazená neurónová sieť zložená z plne prepojených vrstiev. Práve plne prepojená vrstva je najčastejšie používaným typom v klasických neurónových sieťach. Každý z neurónov v takejto vrstve je prepojený s každým neurónom v nasledujúcej vrstve. [13]

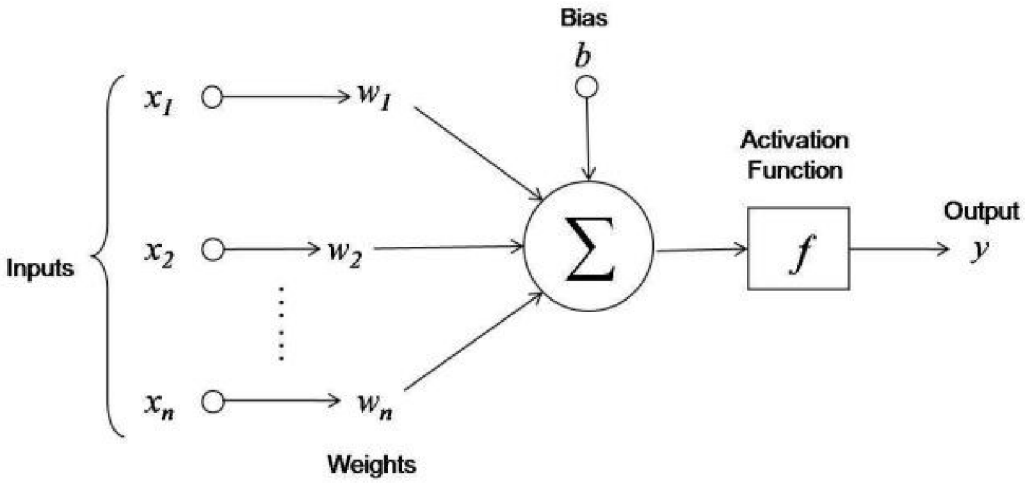


Obr. 3.1: Schéma neurónovej siete s jednou skrytou plne prepojenou vrstvou (prevzaté z [13]).

3.2 Neurón

Ako bolo spomenuté v predchádzajúcej sekcii, jednotlivé neuróny sú vzájomne prepojené. Neurón na obrázku 3.2 obsahuje n vstupov $x_1, \ldots, x_n$, ktoré sú prepojené s výstupmi iných neurónov. Každý z týchto vstupov je ohodnotený váhou w_n, ktorá udáva dôležitosť konkrétneho vstupu. Neurón navyše obsahuje extra hodnotu nazývanú *bias*. Presnejšie ide o váhu pre vstup, ktorého hodnota je vždy 1. Sumou vstupov vynásobených s im prislúchajúcimi váhami a prirátaním biasu b dostaneme vnútorný potenciál neurónu. Z vnútorného potenciálu je takzvanou aktivačnou funkciou f vyrátaná výstupná hodnota. Tá predstavuje výstup neurónu y, ktorý sa prenáša na vstupy iných neurónov. Výstup neurónu je teda možné vyjadriť ako:

$$y = f\left(b + \sum_{i=1}^{n} x_i w_i\right). \tag{3.1}$$



Obr. 3.2: Schéma neurónu[1].

3.3 ReLU

Aktivačné funkcie sú dôležitou súčasťou neurónových sietí. Používajú sa hlavne z dôvodu zavedenia nelinearity. ReLU (Rectified Linear Unit) sa v posledných rokoch stala veľmi populárnou aktivačnou funkciou. Medzi jej výhodami sú výrazné urýchlenie procesu trénovania oproti iným aktivačným funkciám a fakt, že môže byť implementovaná pomocou jednoduchých operácií [13]. Funkcia ReLU je definovaná vzťahom:

$$ReLU(x) = max(0, x). \tag{3.2}$$

[1] https://www.datacamp.com/community/tutorials/neural-network-models-r

3.4 SoftMax

Funkcia SoftMax[2] sa používa najmä pri klasifikácii viacerých tried. Presnejšie keď jedna vzorka dát patrí len do jednej triedy. Funkcia má na vstupe vektor K reálnych čísel, ktoré normalizuje na vektor K hodnôt pravdepodobnostného rozloženia. Tieto hodnoty sú v intervale (0, 1) a ich súčet je rovný 1. Výstupných K hodnôt $x_1, \ldots, x_K$ reprezentuje pravdepodobnosti pre každú z K tried. Zvykne sa používať po výstupnej vrstve neurónovej siete, preto sa výstupná vrstva s použitím SoftMax často označuje ako SoftMax vrstva. Funkcia SoftMax je definovaná vzťahom:

$$SoftMax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}. \tag{3.3}$$

3.5 Generalizácia

Neurónová sieť je sama o sebe nepoužiteľná, kým nie je natrénovaná na dátach. Trénovaním sa myslí iteratívny proces, kedy sa upravujú jednotlivé váhy neurónov s cieľom nájsť ich najoptimálnejšie hodnoty. Konkrétne pri použití neurónovej siete pre klasifikáciu je cieľom, aby na základe vstupných dát správne predikovala, do ktorej triedy patria.

Po natrénovaní neurónovej siete sa môže stať, že sa síce naučí správne predikovať triedy pre vstupné dáta, na ktorých bola trénovaná, ale nie pre podobné dáta, na ktorých trénovaná nebola. Je to spôsobené tým, že sa neurónová sieť naučí až príliš veľa informácií špecifických pre trénovacie dáta – pretrénuje sa. V takom prípade je možné povedať, že zle generalizuje. Generalizácia sa dá zlepšiť mnohými spôsobmi, napríklad zvýšením počtu trénovacích dát alebo použitím neurónovej siete s vhodným počtom parametrov. Počet parametrov je vlastne počet váh, ktoré sa neurónová sieť počas trénovania dokáže naučiť. Pri väčšom počte parametrov je šanca na pretrénovanie vyššia. Okrem vhodného počtu parametrov sa na zlepšenie generalizácie používajú aj iné techniky, napríklad dropout. [3]

3.5.1 Dropout

Dropout je metóda, ktorá aproximuje paralelné trénovanie veľkého množstva neurónových sietí. Ide o často používanú regularizačnú metódu, ktorá tiež zlepšuje generalizáciu. Hlavnou myšlienkou dropoutu je, že sa počas trénovania každému neurónu pridelí pravdepodobnosť p, s ktorou bude v aktuálnej iterácii deaktivovaný – vylúčený (dropped out). Vylúčením rozdielnych neurónov dočasne vznikne nová architektúra siete. Je dôležité poznamenať, že dropout sa používa len počas trénovania. Počas testovania sa namiesto neho používa jednoduchá aproximačná metóda. Ak bol neurón pri trénovaní vylúčený s pravdepodobnosťou p, pri testovaní sú jeho váhy vynásobené hodnotou $1 - p$. [22]

3.5.2 Batch normalizácia

Ďalšou metódou, ktorá môže pomôcť so zlepšením generalizácie je batch normalizácia. Batch normalizácia sa používa na normalizáciu vstupu jednotlivých vrstiev neurónovej siete za účelom optimalizácie trénovania. Ukázalo sa, že táto technika prináša niekoľko benefitov. Okrem urýchlenia trénovania medzi ne patrí zvýšenie stability neurónovej siete. Batch normalizácia je obzvlášť užitočná v neurónových sieťach s väčšou hĺbkou.[14]

[2] https://victorzhou.com/blog/softmax/

3.6 Konvolučné neurónové siete

Konvolučné neurónové siete (CNN) sú typom neurónových sietí najčastejšie spájaným s analýzou obrazu. Sú použiteľné aj pri iných problémoch, kde majú dáta známu mriežkovú štruktúru, napríklad pri spracovaní signálov. Konvolučné neurónové siete sú nazvané podľa matematickej operácie konvolúcie. Ako CNN sa označujú neurónové siete, ktoré používajú konvolúciu v aspoň jednej z vrstiev. [9, s.330–345]

Na rozdiel od klasických neurónových sietí zložených z plne prepojených vrstiev, kde je každý z neurónov prepojený s každým neurónom z predchádzajúcej vrstvy, sa v CNN používa *lokálna konektivita*. Lokálna konektivita znamená, že je každý neurón spojený iba s malou časťou predchádzajúcej vrstvy. Toto pomáha redukovať celkový počet parametrov. CNN sa rovnako ako klasické neurónové siete skladajú zo sekvencie vrstiev. Používajú tri základné typy vrstiev: konvolučné vrstvy, pooling vrstvy a plne prepojené vrstvy.

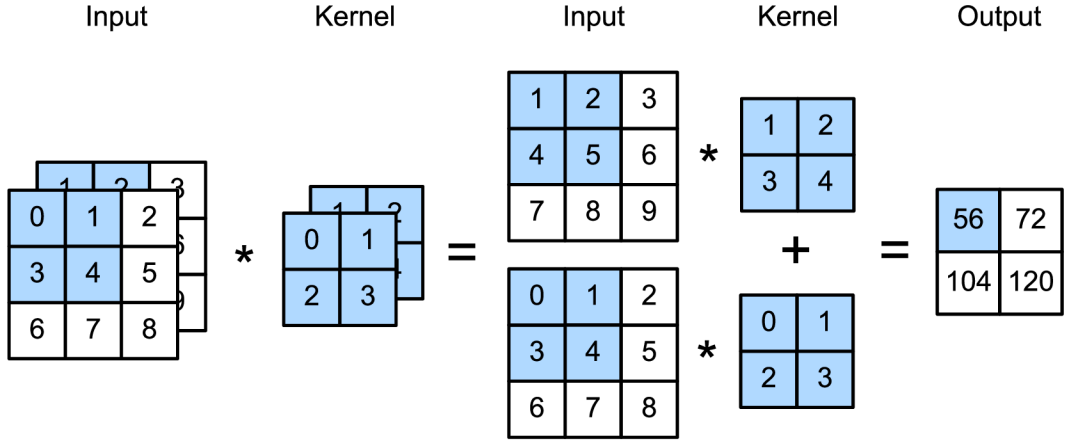
3.6.1 Konvolučná vrstva

Konvolučná vrstva je základným stavebným prvkom CNN. Používa operáciu konvolúciu, ktorá je definovaná ako:

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n), \tag{3.4}$$

kde K predstavuje pole váh nazývané kernel a I predstavuje vstup.

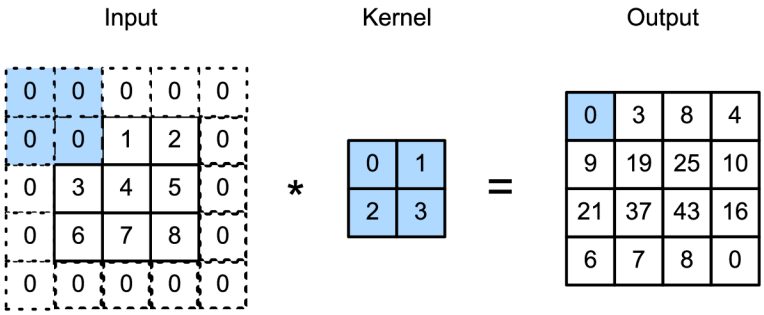
Konvolúcia pre dvojrozmerný vstup s dvomi kanálmi je znázornená na obrázku 3.3. Ako z neho vyplýva, prebehne konvolúcia jednotlivých kanálov s jednotlivými kernelmi a výsledky sa sčítajú. Počet kernelov je vždy rovný počtu kanálov vstupu. Táto skupina kernelov sa nazýva *filter*. Výstup konvolúcie sa zvykne nazývať mapa príznakov. V konvolučných vrstvách sa často používa viacero filtrov, čím pre každý z nich vznikne jedna mapa príznakov.



Obr. 3.3: Znázornenie konvolúcie 2D vstupu s dvomi kanálmi pri veľkosti kernelu 2×2 (prevzaté z [24]).

Hyperparametre konvolučnej vrstvy sú:

- **Veľkosť kernelu** (K) je pri 2D konvolúcii reprezentovaná výškou K_h a šírkou K_w. Počet kernelov vo filtri sa nedefinuje, pretože ako bolo spomenuté, musí byť rovnaký ako počet kanálov vstupu.
- **Strieda** (S) reprezentuje krok, o ktorý sa filter posúva po vstupe.
- **Padding** (P) pomáha pri zachovaní informácií z okrajov vstupných dát. Tiež sa používa na zachovanie rozmerov vstupu. Dosiahne sa to pridaním P riadkov a stĺpcov s hodnotou 0 (zvyčajne) okolo vstupných dát. Padding s hodnotou 1 je znázornený na obrázku 3.4.
- **Počet filtrov** (F_n) sa označuje aj ako hĺbka konvolučnej vrstvy, pretože hĺbka výstupu je rovná použitému počtu filtrov. Znamená to teda, že pri použití F_n filtrov bude výstupom konvolučnej vrstvy F_n máp príznakov.



Obr. 3.4: Konvolúcia s použitím paddingu (prevzaté z [24]).

Výšku O_h a šírku O_w výstupnej mapy príznakov je možné pre vstup s rozmermi I_h a I_w vyrátať podľa vzťahov:

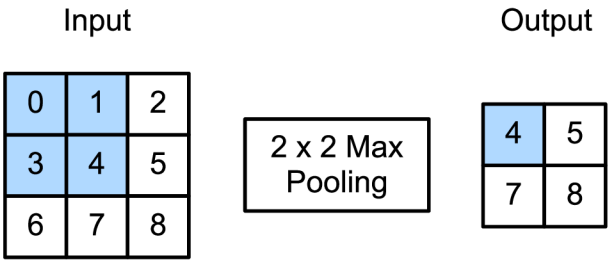
$$O_h = floor\left(\frac{I_h + 2P - K_h}{S} + 1\right), \tag{3.5}$$

$$O_w = floor\left(\frac{I_w + 2P - K_w}{S} + 1\right). \tag{3.6}$$

3.6.2 Pooling vrstva

Pooling vrstva sa používa na zmenšenie rozmerov mapy príznakov. Zvyčajne sa vkladá medzi dve po sebe nasledujúce konvolučné vrstvy. Podobne ako v konvolučnej vrstve, pooling pozostáva z *okna* o pevnej veľkosti, ktoré je postupne posúvané po vstupe o hodnotu striedy. Na rozdiel od konvolučnej vrstvy neobsahuje žiadne parametre. Pooling operácie namiesto toho rátajú priemernú alebo maximálnu hodnotu v danom okne, podľa toho, či ide o max pooling alebo average pooling. [24]

Veľkosť okna používaného pre 2D pooling je zvyčajne 2×2. Špeciálnym prípadom je globálny pooling, čo je pooling s veľkosťou okna rovnakou ako veľkosť vstupu. Výstupom je len jedna hodnota pre jednu mapu príznakov. Pooling je aplikovaný na každú vstupnú mapu príznakov samostatne, čiže ich počet na výstupe je rovnaký ako na vstupe.



Obr. 3.5: Znázornenie max poolingu s oknom o veľkosti 2x2 a striedou 2 (prevzaté z [24]).

Medzi hyperparametre pooling vrstvy, rovnako ako u konvolučnej, patria strieda a padding. Padding sa zvyčajne nepoužíva, keďže cieľom je zmenšenie rozmerov vstupu. Dôležitými hyperparametrami sú veľkosť okna W a typ pooling operácie (max alebo average). Výpočet výstupnej výšky O_h a šírky O_w je vyjadrený vzťahmi:

$$O_h = floor\left(\frac{I_h + 2P - W_h}{S} + 1\right), \tag{3.7}$$

$$O_w = floor\left(\frac{I_w + 2P - W_w}{S} + 1\right). \tag{3.8}$$

3.7 Autoenkóder

Autoenkóder je špeciálny typ neurónovej siete používaný pri redukcii dimenzionality dát. Ako autoenkóder môže byť použitá klasická neurónová sieť aj CNN. Presnejšie ide o neurónovú sieť, ktorá je trénovaná, aby skopírovala vstup na výstup. Medzi vstupnou a výstupnou vrstvou obsahuje aspoň jednu skrytú vrstvu s menším počtom neurónov ako má vstupná a výstupná vrstva. Táto vrstva sa zvykne nazývať *bottleneck*, keďže je to najužšia časť autoenkóderu. Bottleneck obmedzuje počet informácií, ktoré prechádzajú celou sieťou a tým ju núti naučiť sa reprezentovať vstupné dáta pomocou menšej dimenzionality. [9, s.502–525]

Architektúru autoenkóderu je možné rozdeliť na dve hlavné časti:

- **Enkóder** - Redukuje dimenzionalitu dát na veľkosť bottlenecku. Cieľom je zachovať podstatné informácie, nepotrebné informácie a šum naopak odstrániť.
- **Dekóder** - Vstupom dekóderu je výstup z enkóderu - kompresované vstupné dáta. Dekóder sa z nich snaží zrekonštruovať pôvodné dáta.

Pri použití autoenkóderu na redukciu dimenzionality dát sa po natrénovaní odstráni dekódovacia časť, teda ostane iba enkóder a výstupnou vrstvou sa stane bottleneck.

Kapitola 4

Dátová sada

V tejto kapitole sa nachádza popis dátovej sady použitej v tejto práci a jej analýza. Potrebné dáta boli nahrané v predchádzajúcom výskume, takže táto práca sa nezaoberá ich získavaním.

4.1 Obsah dátovej sady

EMG nahrávky v dátovej sade je možné podľa obsahu rozdeliť na dve skupiny:

- nahrávky krátkych viet z korpusu TIMIT [7],
- nahrávky slov s aritmetickým významom (ďalej príkazy).

Nahrávky viet sú použité v jednom z experimentov. Okrem toho boli pre prácu použité nahrávky 15 príkazov. Konkrétne ide o číslice 0 – 9 a matematické operácie násobenie, delenie, sčítanie, odčítanie a percento. Obsah nahrávok je v anglickom jazyku, takže presnejšie ide o príkazy `zero`, `one`, `two`, `three`, `four`, `five`, `six`, `seven`, `eight`, `nine`, `add`, `subtract`, `divide`, `multiple` a `percent`. Práve tieto príkazy budú pri klasifikácii reprezentovať jednotlivé triedy.

Vo všetkých nahrávkach rozpráva rovnaký rečník. Dokopy bolo nahraných šesť sedení označených S1 - S6. Prvých päť sedení je rozdelených na bloky obsahujúce dvanásť párov nahrávok. Každý pár nahrávok je zložený zo silent a audible nahrávky, teda nahrávky silent speech a audible speech. Z dvanástich párov osem obsahuje vety z korpusu TIMIT a štyri páry obsahujú príkazy. Počet blokov v sedeniach sa líši, pretože dĺžka sedenia závisela od subjektívnej únavy rečníka. Sedenie S6 obsahuje len silent nahrávky príkazov, ktorých tu je ale viac ako v ostatných sedeniach.

4.1.1 Priebeh nahrávania

Pár audible a silent nahrávky obsahuje rovnaký vyslovený text a je nahraný ihneď po sebe. Na začiatku každej nahrávky bola ukázaná vizuálna pomôcka – text, ktorý má rečník vysloviť. Sekundu po nej bola pustená audio nahrávka obsahujúca daný text nahraný syntetickým hlasom. Ďalej nasledovalo pípnutie, ktoré slúžilo ako pokyn k začatiu rozprávania. Po nahratí zaznelo ďalšie pípnutie signalizujúce ukončenie nahrávania. Poradie silent a audible nahrávky v rámci jedného bloku je rovnaké. V nasledujúcom bloku sa poradie vždy vymenilo, aby sa predišlo skresleniu, ktoré by mohlo byť spôsobené poradím.

4.1.2 Organizácia dát

Každé zo sedení obsahuje poznámky k sedeniu, tabuľku s informáciami o jednotlivých nahrávkach, EMG dáta a maticu triggerov pre rozdelenie EMG.

Poznámky k sedeniu

Poznámky k sedeniu obsahujú informácie o prípadných chybách a iných okolnostiach, ktoré by mohli mať vplyv na namerané hodnoty. Ak sa v sedení poznámky nenachádzajú, nedošlo počas nahrávania k žiadnym komplikáciám.

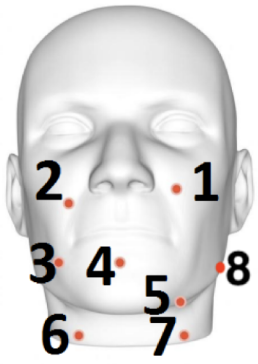
Tabuľka s informáciami

Tabuľka s informáciami obsahuje popis jednotlivých nahrávok. Tieto informácie slúžia k následnému spracovaniu. Konkrétne sa tu nachádzajú nasledujúce položky:

- **Target:** text vyslovený v nahrávke.
- **PromptType:** obsahuje hodnoty `Timit` a `Command` slúžiace na rozlíšenie, či ide o vetu z korpusu TIMIT alebo príkaz.
- **StimuliOrder:** informácia o tom, či bola skôr nahraná silent alebo audible nahrávka.
- **StimuliType:** obsahuje hodnotu `loud` pre audible a `silent` pre silent nahrávku.
- **Rejected:** obsahuje hodnotu 0 ak je nahrávka v poriadku, 1 ak je nahrávka z nejakého dôvodu nepoužiteľná.

EMG signál

EMG signál bol meraný pomocou ôsmich sEMG elektród. Sedem z nich bolo umiestnených rovnako ako popisuje výskum [11]. Navyše bola použitá jedna EMG elektróda umiestnená na ľavom žuvacom svale. Pozície všetkých ôsmich elektród sú znázornené na obrázku 4.1. Vzorkovacia frekvencia nahraného EMG signálu je 2048 Hz. EMG signál bol nahrávaný súvisle počas celého sedenia. Pre rozdelenie na jednotlivé nahrávky slúžia triggery označujúce začiatok a koniec nahrávky.



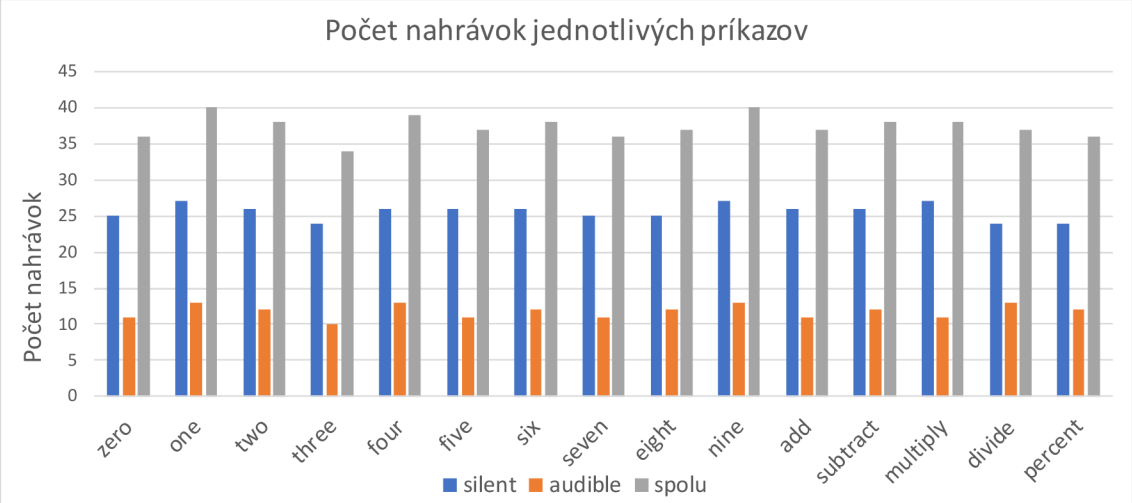
Obr. 4.1: Umiestnenie EMG elektród (prevzaté z predchádzajúceho výskumu).

4.1.3 Analýza dátovej sady

Celkovo dátová sada obsahuje **561** nahrávok príkazov, z toho **178** audible a **383** silent. Silent nahrávok je kvôli S6, kde bol nahraný iba tento typ, približne dvakrát viac. Triedy sú relatívne vyvážené, ako vyplýva aj zo sumarizujúcej tabuľky 4.1 a grafu početnosti jednotlivých nahrávok na obrázku 4.2.

Tabuľka 4.1: Sumarizácia počtu nahrávok jednotlivých slov v sedeniach.

	Audible						**Silent**						
	S1	**S2**	**S3**	**S4**	**S5**	**Spolu**	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**Spolu**
zero	0	1	3	3	4	**11**	0	1	3	3	4	14	**25**
one	3	1	3	2	4	**13**	3	1	3	2	4	14	**27**
two	2	1	3	3	3	**12**	2	1	3	3	3	14	**26**
three	2	1	2	2	3	**10**	2	1	2	2	3	14	**24**
four	3	1	3	3	3	**13**	3	0	3	3	3	14	**26**
five	2	1	3	2	3	**11**	2	1	3	2	3	15	**26**
six	3	1	3	2	3	**12**	3	1	3	2	3	14	**26**
seven	2	1	2	3	3	**11**	2	1	2	3	3	14	**25**
eight	2	1	3	3	3	**12**	2	0	3	3	3	14	**25**
nine	2	2	2	3	4	**13**	2	2	2	3	4	14	**27**
add	3	1	3	2	3	**12**	3	1	3	2	3	14	**25**
subtract	2	1	3	3	3	**12**	2	1	3	3	2	15	**26**
divide	2	1	2	3	3	**11**	1	1	2	3	3	14	**24**
multiply	3	1	3	3	3	**13**	3	1	3	3	3	14	**27**
percent	3	1	2	3	3	**12**	3	1	2	3	3	12	**24**
Spolu	**33**	**16**	**40**	**40**	**48**	**178**	**33**	**14**	**40**	**40**	**47**	**210**	**383**



Obr. 4.2: Graf početnosti nahrávok jednotlivých príkazov.

4.2 Rozdelenie dát pre trénovanie a testovanie

Presnosť klasifikátoru z veľkej časti závisí na veľkosti a kvalite dátovej sady. Dôležité je taktiež rozdelenie dátovej sady na sady pre učenie a testovanie. Testovaním modelu na iných dátach ako na tých, na ktorých bol trénovaný sa overí jeho schopnosť generalizovať. Často používaný spôsob rozdelenia je použiť 60 % z celej dátovej sady na trénovanie, 20% na validáciu a zvyšných 20% na testovanie. Validačná dátová sada sa používa pri ladení hyperparametrov, preto môžu byť výsledky na tejto sade skreslené. Z toho dôvodu sa na finálne vyhodnotenie používa rozdielna dátová sada – testovacia.

4.2.1 K-násobná krížová validácia

K-násobná krížová validácia (k-fold cross validation) je technika často používaná pri malých dátových sadách. Dátová sada sa najskôr náhodne premieša a následne sa, ako naznačuje názov, rozdelí na k častí s približne rovnakou veľkosťou. Potom prebehne k iterácii, teda jedna iterácia pre každú z častí. Táto časť sa použije ako testovacia sada a spojením ostatných $k - 1$ častí sa vytvorí trénovacia sada. V každej iterácii sa model natrénuje na trénovacej sade a vyhodnotí na testovacej sade. Vznikne teda celkovo k výsledkov, ktoré sa nakoniec spriemerujú. Ako k sa zvyčajne volí hodnota 5 alebo 10. Na rozdiel od klasického rozdelenia na trénovaciu, testovaciu a validačnú sadu, pri krížovej validácii sú postupne pre testovanie použité všetky vzorky z dátovej sady. Krížová validácia sa často používa pri ladení hyperparametrov modelu a tiež pri výbere z viacerých modelov. [21]

4.2.2 Rozdelenie podľa nahrávacích sedení

Ďalším spôsobom rozdelenia dát pre trénovanie a testovanie je rozdeliť ich podľa sedení, v ktorých boli nahrané. V tomto konkrétnom prípade je to obzvlášť vhodné z dôvodu, že sa tým overí funkčnosť na nevidených sedeniach. Dá sa predpokladať, že výsledky dosiahnuté týmto spôsobom sa budú viac približovať výsledkom dosiahnutým pri klasifikácii nahrávok z nového sedenia. Pri tomto rozdelení je vždy jedno zo sedení vybrané ako testovacie a ostatné sú použité na trénovanie. Ako testovacie sedenia sú použité sedenia S1, S3, S4 a S5. Sedenie S2 ako testovacie použité nebolo, pretože obsahuje málo nahrávok a sedenie S6 preto, že obsahuje len silent nahrávky. Navyše obsahuje takmer rovnaké množstvo nahrávok ako ostatné sedenia spolu. To by prudko znížilo veľkosť trénovacej sady, čím by mohli byť ovplyvnené výsledky.

Kapitola 5

Spracovanie signálu a extrakcia príznakov

V tejto kapitole bude popísaný celý postup spracovania EMG signálu. Prvým krokom je filtrovanie EMG signálu za účelom odstránenia šumu. Z takto filtrovaného signálu sú následne extrahované príznaky, ktoré budú neskôr použité ako vstup klasifikátoru.

5.1 Spracovanie signálu

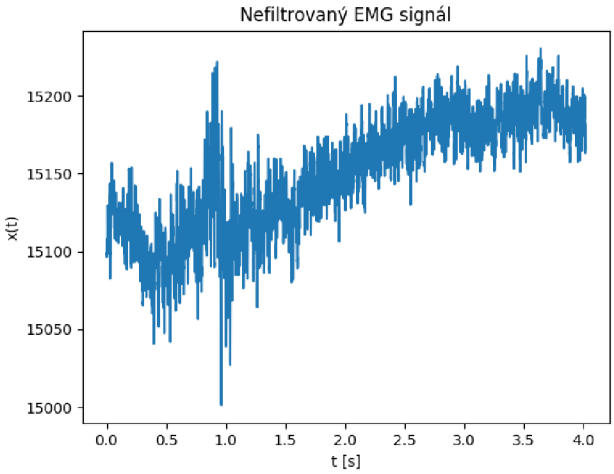
Použitie signálu priamo nameraného pomocou EMG je síce možné, ale takýto signál obsahuje okrem svalovej aktivity, ktorá nás zaujíma, taktiež šum. Ľudské telo sa správa ako anténa – povrch tela je neustále zaplavovaný elektrickým a magnetickým žiarením, ktoré je zdrojom elektromagnetického šumu. Elektromagnetické zdroje z okolia spôsobujú nechcený signál a zatieňujú signál nahrávaný zo svalu. Amplitúda šumu z okolia je niekedy až trikrát väčšia ako amplitúda EMG signálu, ktorý nás zaujíma (podľa [4]).

5.1.1 Odfiltrovanie šumu

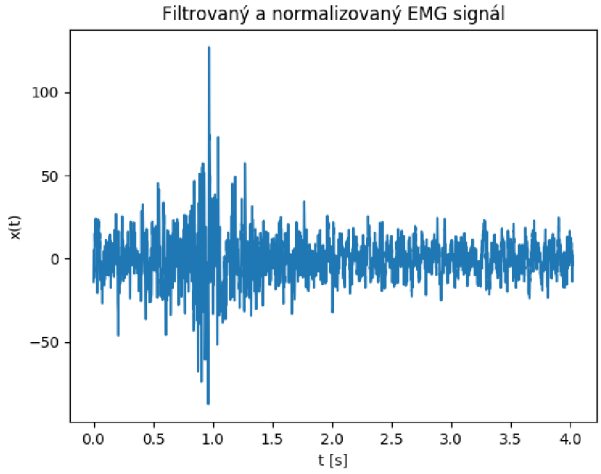
Medzi dominantné zdroje šumu patrí šum s frekvenciou 50/60 Hz a jeho harmonické frekvencie[1] od spotrebičov v elektrickej sieti, nazývaný tiež Power-Line Interference (PLI). Frekvenčné spektrum signálu nameraného pomocou sEMG je zvyčajne z rozsahu od 0 Hz do 450 Hz. Radí sa preto odfiltrovať frekvencie vyššie ako 400–450 Hz (podľa [2]). Ďalšia nechcená zložka signálu je spôsobená pohybom káblov pripevneným k elektródam a pohybom samotných elektród. Pri aktivácii sa sval skráti, čím sa mierne zmení poloha elektródy vzhľadom na sval. Týmto vzniká šum s frekvenciou 1–10 Hz. Tieto frekvencie sú ale produkované aj svalovou aktivitou, takže čo sa týka odfiltrovania nízkych frekvencií, odporúčania pre spodnú hranicu sú rôzne. Väčšinou sa spodná hranica pohybuje v rozsahu 5–28 Hz (podľa [4]).

Z EMG signálu bola odfiltrovaná frekvencia 60 Hz a jej harmonické frekvencie, keďže EMG signál bol nahraný v USA, kde je sieťová frekvencia 60 Hz. Následne boli odfiltrované frekvencie nižšie ako 5 Hz a vyššie ako 450 Hz, čím boli dosiahnuté najlepšie výsledky. Nakoniec bola od vyfiltrovaného signálu odčítaná priemerná hodnota (mean normalization). EMG signál pred spracovaním sa nachádza na obrázku 5.1 a signál po spracovaní je zobrazený na obrázku 5.2.

[1] Harmonické frekvencie sú celočíselnými násobkami základnej frekvencie



Obr. 5.1: Pôvodný EMG signál.



Obr. 5.2: EMG signál po spracovaní.

5.1.2 Rozdelenie na nahrávky

Ďalším krokom bolo rozdelenie signálu na jednotlivé nahrávky, keďže, ako bolo spomenuté v sekcii 4.1, celé sedenie bolo nahrané ako súvislý EMG signál. Nahrávky dlhé štyri sekundy obsahovali pomerne veľkú časť bez reči. Experimentálne bolo zistené, že reč začína približne 0,15 sekundy po signalizácii začiatku nahrávania a celá sa nachádza v nasledujúcich dvoch sekundách. Z tohto dôvodu bolo odrezaných 0,15 sekundy zo začiatku nahrávky a 1,8 sekundy z jej konca, čím sa pôvodná dĺžka skrátila takmer o polovicu.

Takto spracovaný EMG signál je už pripravený na extrakciu príznakov. Extrakcia príznakov prebieha samostatne pre každý z ôsmich kanálov EMG signálu.

5.2 Extrakcia príznakov

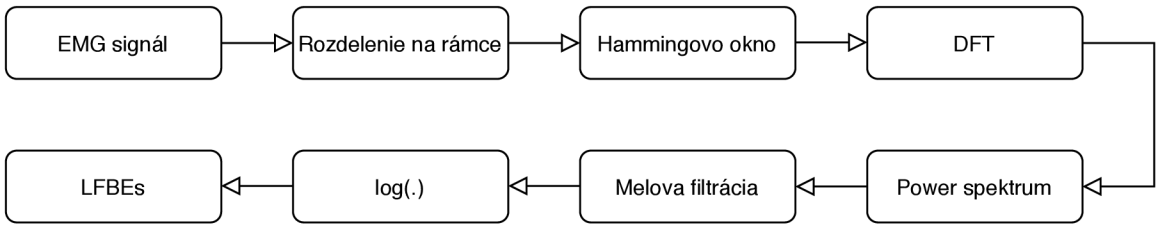
Extrakcia príznakov sa používa za účelom získať z pôvodných dát iba najvýznamnejšie charakteristiky. Na základe viacerých existujúcich výskumov spomenutých v sekcii 2.3, ktoré sa zhodovali v tom, že pre klasifikáciu reči z EMG dát sú vhodné MFCC príznaky, boli v tejto práci použité taktiež. MFCC príznaky sa používajú hlavne pri práci so zvukom, no ako sa ukázalo, sú vhodnou voľbou aj pri rozpoznávaní reči z EMG signálu.

Bolo zistené, že pre použitie s neurónovými sieťami môže byť lepším riešením použiť logaritmus energií banky Melových filtrov (Log-Mel Filter Bank Energy, LFBE)[2]. Tie sú získané rovnakým postupom ako MFCC, jediným rozdielom je vynechanie posledného kroku pri výpočte MFCC príznakov – diskrétnej kosínusovej transformácie (DCT). DCT sa používa na dekoreláciu LFBE. MFCC boli veľmi populárne hlavne v časoch, keď boli ako modely pre rozpoznávanie reči často používané HMM a GMM. Neurónové siete oproti nim dokážu lepšie pracovať s vysoko korelovaným vstupom, preto môžu byť LFBE príznaky lepšou voľbou (podľa [18]). V tejto práci boli testované obe varianty a lepšie výsledky boli dosiahnuté s použitím LFBE. Pre ich extrakciu bola použitá implementáciu z predchádzajúceho výskumu.

Extrakcia príznakov prebieha osobite pre každú nahrávku. Keďže pri nahrávaní bolo použitých 8 EMG elektród, nahraný signál má 8 kanálov. Príznaky sú extrahované samostatne pre každý z týchto kanálov. Skupina príznakov extrahovaných z jednej nahrávky reprezentuje jednu vzorku vstupných dát klasifikátoru.

5.2.1 LFBE príznaky

Extrakcia LFBE príznakov z EMG signálu pozostáva z viacerých krokov znázornených na obrázku 5.3. Informácie použité v tejto časti som čerpal zo zdrojov [1, 18].



Obr. 5.3: Bloková schéma extrakcie príznakov z EMG signálu.

Signál sa najskôr rozdelí na krátke rámce o dĺžke N vzoriek. Tento krok sa robí kvôli tomu, aby bolo možné zachytiť charakteristiky v jednotlivých častiach signálu. Nasledujúce rámce sa navyše prekrývajú o určitú časť (overlap), čo pomáha pri zachytení kontextu.

Jednotlivé rámce sú následne vynásobené Hammingovým oknom. Hammingovo okno vyjadrené vzťahom 5.1 slúži na potlačenie vzoriek signálu pri krajoch okna.

$$w_n = 0.54 - 0.46\ cos\left(\frac{2\pi n}{N-1}\right) \tag{5.1}$$

To pomáha zmenšiť rozdiel medzi hodnotou prvej a poslednej vzorky v rámci kvôli ich návaznosti, čo je nápomocné pre nasledujúci krok, ktorým je diskrétna Fourierova Transformácia

[2]Často sa používa aj označenie Mel–frekvenčné spektrálne koeficienty (MFSC)

(DFT). DFT slúži pre výpočet frekvenčného spektra signálu v danom rámci. Presnejšie povedané, N vzoriek rámca $x_0, \ldots, x_{N-1}$ sa prevedie na N hodnôt frekvenčného spektra $X_0, \ldots, X_{N-1}$ podľa vzťahu 5.2.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \tag{5.2}$$

Pri DFT sa predpokladá, že signál je periodický a spojitý. Práve toto je dôvodom pre použitie Hammingovho okna. Z výstupu DFT sa vyráta power spektrum podľa vzťahu 5.3, čím sa získa hodnota energie pre jednotlivé frekvencie.

$$S_k = |X_k|^2 \tag{5.3}$$

Tieto energie sú následne vynásobené bankou trojuholníkových Melových filtrov. Tie sú rovnomerne rozložené po frekvenčnej osi podľa Melovej stupnice vyjadrenej vzťahom 5.4, kde f je frekvencia prevádzaná do Melovej stupnice. Susedné filtre sa prekrývajú o polovicu.

$$Mel(f) = 1125 \; ln\left(1 + \frac{f}{700}\right) \tag{5.4}$$

Tento krok, teda násobenie banky Melových filtrov a power spektra, sa nazýva Melova filtrácia. Výsledkom je hodnota energie pre každý z banky filtrov. Nakoniec sa z výsledných energií pre jednotlivé filtre vypočíta logaritmus, čím sa zmenší dynamický rozsah hodnôt. Tým vznikne logaritmus energie banky Melových filtrov – LFBE. Ako posledný krok je od príznakov z celej nahrávky odčítaná ich priemerná hodnota (zvlášť pre každý kanál).

5.2.2 Parametre zvolené pri extrakcii príznakov

Pri extrakcii LFBE príznakov je dôležité použiť správne hodnoty pre ich počet, dĺžku rámca a časť, o ktorú sa nasledujúce rámce prekrývajú. Najlepšie hodnoty boli zistené experimentálne. Podrobnejšie informácie budú popísané v časti 7.3.1. Tu len spomeniem, že na základe experimentov bola zvolená kombinácia 10 LFBE príznakov extrahovaných z rámcov o dĺžke 100 ms, prekrývajúcich sa o 75 ms. Pri týchto hodnotách z jednej nahrávky dlhej 2,05 sekundy vznikne 80 rámcov, teda $80 \times 10 \times 8$ LFBE príznakov.

Kapitola 6

Návrh a implementácia klasifikátoru

V tejto kapitole sú popísané navrhnuté architektúry CNN, ktoré slúžia na klasifikáciu príkazov z EMG nahrávok. Tiež sa tu nachádza krátky popis implementácie.

6.1 Návrh klasifikátoru

Dôležitou časťou tejto práce je návrh neurónovej siete slúžiacej na klasifikáciu príkazov z EMG nahrávok. Konkrétne sú použité CNN popísané v kapitole 3.6, ktoré už boli úspešne použité v podobnej úlohe. Práve CNN bola použitá ako klasifikátor vo výskume [11], ktorý sa zaoberá klasifikáciou slov zo silent speech a dosahuje v nej výborné výsledky. Popis architektúry z tohto výskumu bol použitý ako základ pre architektúru prvej CNN vytvorenej v tejto práci, nazvanú podľa daného výskumu – AlteregoNet. Ďalej, v snahe zlepšiť dosiahnuté výsledky, bola navrhnutá vlastná architektúra nazvaná SilentNet.

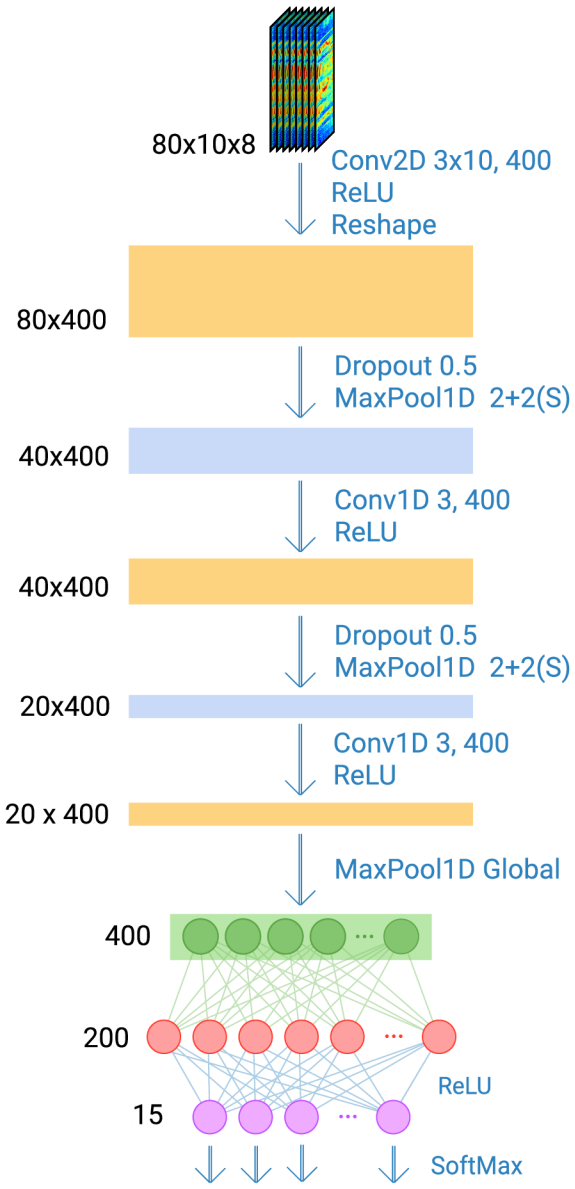
6.1.1 AlteregoNet

V spomenutom výskume [11] sa nachádza popis použitej CNN, ale je relatívne stručný, takže architektúra implementovaná v tejto práci pravdepodobne nebude úplne totožná. Okrem toho bolo upravených pár drobností, s ktorými boli dosiahnuté lepšie výsledky – napríklad zredukovanie použitia dropoutu. Napriek tomu sa pôvodnej architektúre snažila priblížiť. Zaujímavé je, že autor uvádza, že použil 1D konvolučnú neurónovú sieť aj napriek faktu, že ako jej vstup používa 2D vstupné dáta (MFCC príznaky). V tejto práci bolo ako riešenie zvolené použitie 2D konvolučnej vrsty so šírkou kernelu rovnou šírke vstupných dát, čím vznikol výstup so šírkou 1. S ním sa ďalej pracovalo ako s 1D dátami.

Navrhnutá architektúra, ktorú som nazval AlteregoNet, je znázornená na obrázku 6.1. $80 \times 10 \times 8$ na vstupe reprezentuje počet použitých rámcov, počet príznakov pre jeden rámec a počet kanálov. Počet rámcov závisí na dĺžke vstupného signálu a parametroch zvolených pri extrakcii príznakov. Pre väčšiu názornosť som použil konkrétne hodnoty zvolené v podsekcii 5.2.2. Je dôležité zdôrazniť, že sa ráta aj s použitím odlišného počtu rámcov a príznakov.

Časť označená ako `Conv2D` symbolizuje 2D konvolučnú vrstvu. Príslušné hodnoty 3×10 a 400 definujú použité rozmery kernelu a počet filtrov v danej konvolučnej vrstve. Rovnako sú popísané 1D konvolučné vrstvy s jediným rozdielom, že kernel je jednorozmerný. Každá konvolučná vrstva je nasledovanú aktiváciou ReLU. `MaxPool1D` s hodnotou `2 + 2(S)` sym-

bolizuje max pooling vrstvu s oknom o veľkosti 2 a striedou 2. Označenie `Global` pri max pooling vrstve znamená, že ide o globálnu max pooling vrstvu. Výstup z tejto vrstvy je vstupom plne prepojenej vrstvy s 200 neurónmi, nasledovanej aktiváciou ReLU. Za ňou sa nachádza už len výstupná vrstva s 15 neurónmi, teda jedným pre každú z klasifikovaných tried, nasledovaná funkciou SoftMax.



Obr. 6.1: Architektúra CNN AlteregoNet.

6.1.2 SilentNet

Okrem architektúry AlteregoNet bola navrhnutá aj úplne vlastná architektúra CNN, nazvaná *SilentNet*. Na rozdiel od predchádzajúcej architektúry bolo použitých viac 2D vrstiev vzhľadom k tomu, že vstupné dáta sú dvojrozmerné. Pri návrhu bol zvolený viac experimentálny prístup, aby sa zistilo, čo najlepšie funguje v tomto konkrétnom prípade. Bolo to síce veľmi časovo náročné, ale po mnohých experimentoch s je výsledkom architektúra na obrázku 6.2. Systém použitý pre popis jednotlivých vrstiev je rovnaký ako v kapitole 6.1.1. Upresnené by malo byť jedine označenie `MaxPool 2×2 + 2(S)`, čo symbolizuje 2D max pooling vrstvu s oknom o veľkosti 2×2 a striedou 2 v oboch smeroch. Navyše sa tu nachádza `BatchNorm`, čo symbolizuje batch normalizáciu popísanú v časti 3.5.2.

Architektúra obsahuje celkovo štyri konvolučné vrstvy a tri max pooling vrstvy, z toho posledná je globálna. Pri vstupe s ôsmimi kanálmi má 1 278 939 naučiteľných parametrov, čo je mierne viac ako AlteregoNet s 1 092 415 parametrami. Zaujímavosťou je prvá konvolučná vrstva s 300 filtrami a veľkosťou kernelu 1×1. Konvolučná vrstva s kernelom o veľkosti 1×1 sa používa na zmenu počtu kanálov vstupu pri zachovaní výšky a šírky. Často sa používa práve na redukciu počtu kanálov, ale je možné ju použiť taktiež na zvýšenie počtu kanálov. Práve týmto spôsobom je použitá v tejto práci. Je to mierne neštandardné riešenie, ale viedlo k lepším výsledkom.

6.2 Implementácia

Pre implementáciu modelov CNN bol použitý framework Keras[1]. Keras je open-source framework napísaný v jazyku Python. Presnejšie ide o vysokoúrovňové API podporujúce viaceré backend platformy, z ktorých bola konktrétne zlovená platforma TensorFlow[2]. Hlavnou výhodou frameworku Keras je jednoduchosť prototypovania modelov, keďže pre pridanie jednej vrstvy neurónovej siete typicky stačí jeden riadok kódu.

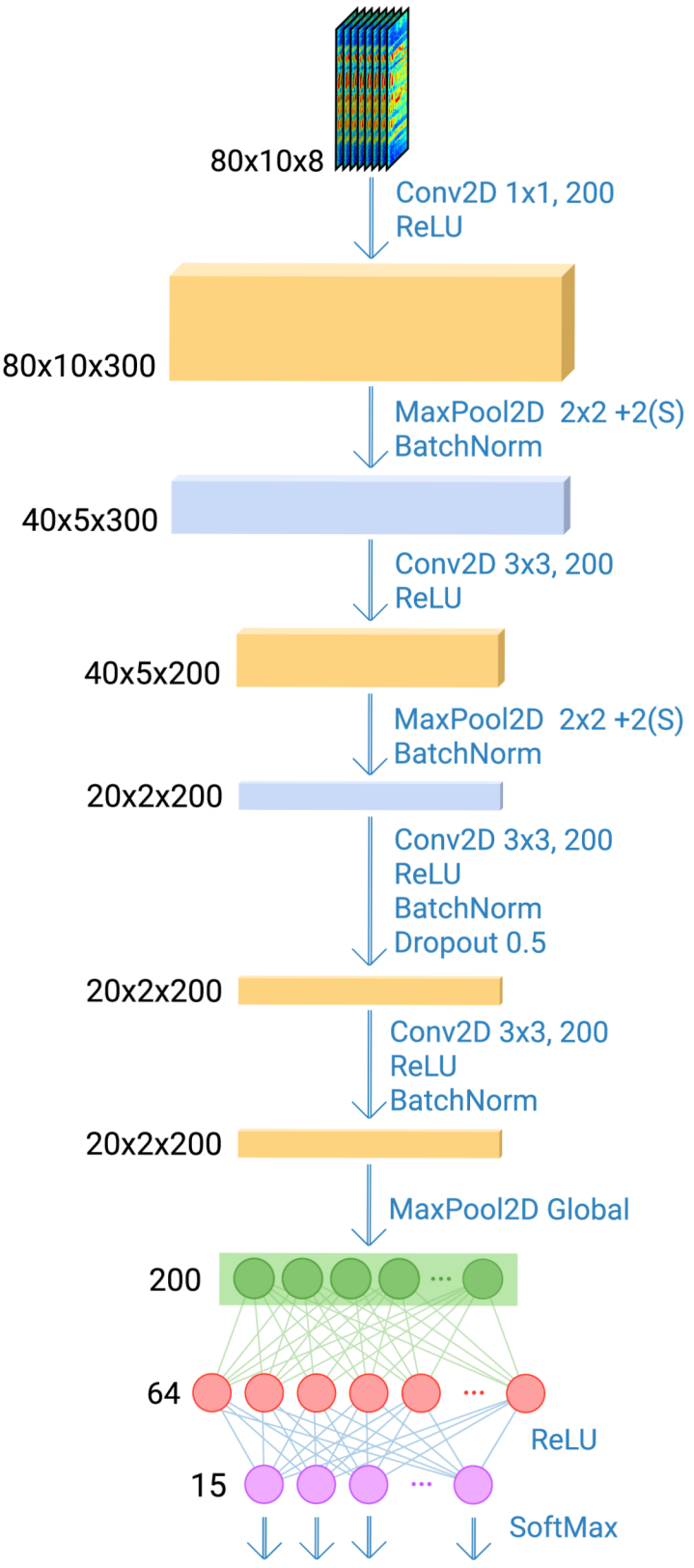
Konkrétne bola na prototypovanie modelov použitá trieda `Sequential`, určenú na tvorbu modelov zložených zo sekvencie vrstiev s jedným vstupom a výstupom. Vytvorenie modelu je veľmi jednoduché, keďže Keras obsahuje aj implementáciu všetkých typov vrstiev, ktoré sa nachádzajú v navrhnutých architektúrach. Stačí vytvoriť inštancie potrebných vrstiev, inštanciu triedy `Sequential` a pridať do nej vytvorené vrstvy v príslušnom poradí.

6.3 Trénovanie

Pred použitím boli navrhnuté modely natrénované na dátach. Jedna vzorka vstupných dát predstavuje jednu nahrávku rozdelenú na rámce, z ktorých boli extrahované LFBE príznaky postupom popísaným v časti 5.2.1.

Pri trénovaní je dôležitý optimalizátor pre úpravu váh a taktiež spôsob ich inicializácie. Ako optimalizátor bol použitý Adam [12]. Pre inicializáciu váh bola zvolené metóda nazývaná *Kaiming He inicializácia*, podľa jej autora, ktorý vo výskume [10] vysvetľuje, prečo je pri použití aktivačnej funkcie ReLU táto metóda inicializácie váh vhodnejšia.

[1] https://keras.io/
[2] https://www.tensorflow.org/



Obr. 6.2: Architektúra SilentNet.

Kapitola 7

Dosiahnuté výsledky a experimenty

V tejto kapitole sa nachádza vyhodnotenie úspešnosti klasifikátorov navrhnutých v predchádzajúcej kapitole a ich porovnanie s predchádzajúcim výskumom. Ďalej sa tu nachádza množstvo experimentov s použitím príznakov získaných rôznymi spôsobmi a porovnanie klasifikácie silent a audible speech.

7.1 Meranie úspešnosti klasifikácie

Aj napriek tomu, že existuje mnoho rozdielnych typov klasifikátorov, meranie úspešnosti je pre všetky z nich založené na podobných princípoch. Pri klasifikácii, kedy sú známe skutočné triedy, do ktorých patria jednotlivé vzorky dát, je jednoduché určiť, či sa predikovaná trieda zhoduje so skutočnou. Presnejšie pri klasifikácii každej vzorky dát môžu nastať štyri situácie, ktoré sú základom pre väčšinu metrík používaných pri klasifikácii[1]:

- **True positive (TP)** - klasifikátor predikoval, že vzorka dát patrí do danej triedy a skutočne do nej patrí.
- **True negative (TN)** - klasifikátor predikoval, že vzorka dát nepatrí do danej triedy a skutočne do nej nepatrí.
- **False positive (FP)** - klasifikátor predikoval, že vzorka dát patrí do danej triedy aj napriek tomu, že v skutočnosti do nej nepatrí.
- **False negative (FN)** - klasifikátor predikoval, že vzorka dát do danej triedy nepatrí aj napriek tomu, že v skutočnosti do nej patrí.

7.1.1 Presnosť

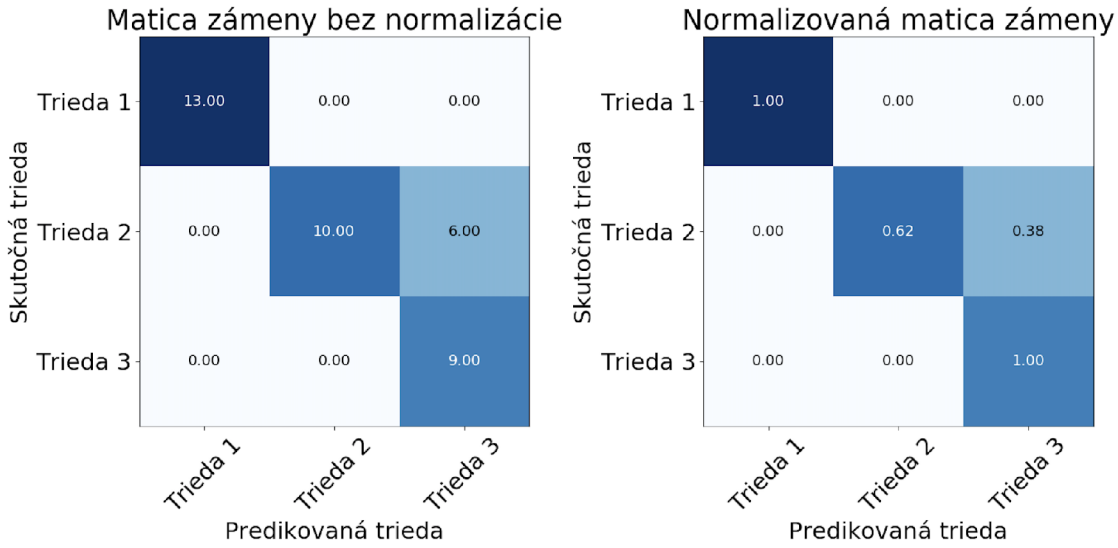
Pri klasifikácii viacerých tried je pravdepodobne najpoužívanejšou metrikou *presnosť* (accuracy). Presnosť je veľmi intuitívna metrika. Označuje pomer medzi počtom správnych predikcií a celkovým počtom predikcií. Presnosť je tiež možné definovať vzťahom[2]:

$$accuracy = \frac{TP}{TP + FP}. \tag{7.1}$$

[1] https://spark.apache.org/docs/latest/mllib-evaluation-metrics.html
[2] https://medium.com/apprentice-journal/evaluating-multi-class-classifiers-12b2946e755b

7.1.2 Matica zámeny

Matica zámeny (confusion matrix) zachytáva celkový prehľad výsledkov klasifikačného modelu. Používa sa pri klasifikácii dvoch alebo viacerých tried. Jedná sa o dvojrozmernú maticu. Jedným rozmerom sú skutočné triedy a druhým triedy predikované klasifikátorom. Pre každú kombináciu skutočnej a predikovanej triedy obsahuje odpovedajúci počet prípadov. Často sa tiež používajú normalizované hodnoty. Obe verzie sú zobrazené na obrázku **??**. V ideálnom prípade by matica obsahovala nuly všade okrem diagonály, teda všetky predikované triedy by boli rovnaké ako skutočné. Z matice zámeny sa dá tiež ľahko vyčítať, ktoré triedy sa často zamieňajú.



Obr. 7.1: Matica zámeny bez normalizácie a normalizovaná matica zámeny.

7.2 Vyhodnotenie úspešnosti

Úspešnosť vytvorených modelov bola nakoniec vyhodnotená na rozdielnych trénovacích a testovacích dátových sadách pre lepšiu predstavu o ich fungovaní za rôznych podmienok. Najskôr boli dáta rozdelené náhodne pomocou metódy 5-násobnej krížovej validácie. Tento postup bol použitý aj pri ladení hyperparametrov a architektúry výsledných modelov. Následne bolo použité rozdelenie podľa sedení, v ktorých boli nahrané, z ktorých bolo vždy jedno vybrané ako testovacie.

7.2.1 Priebeh merania úspešnosti

V rámci jedného merania bol model natrénovaný a výsledok vyhodnotený celkovo desaťkrát. Finálny výsledok je vyrátaný ako ich priemer. Tento postup je síce časovo náročný, ale bol zvolený za účelom dosiahnutia stabilnejších výsledkov.

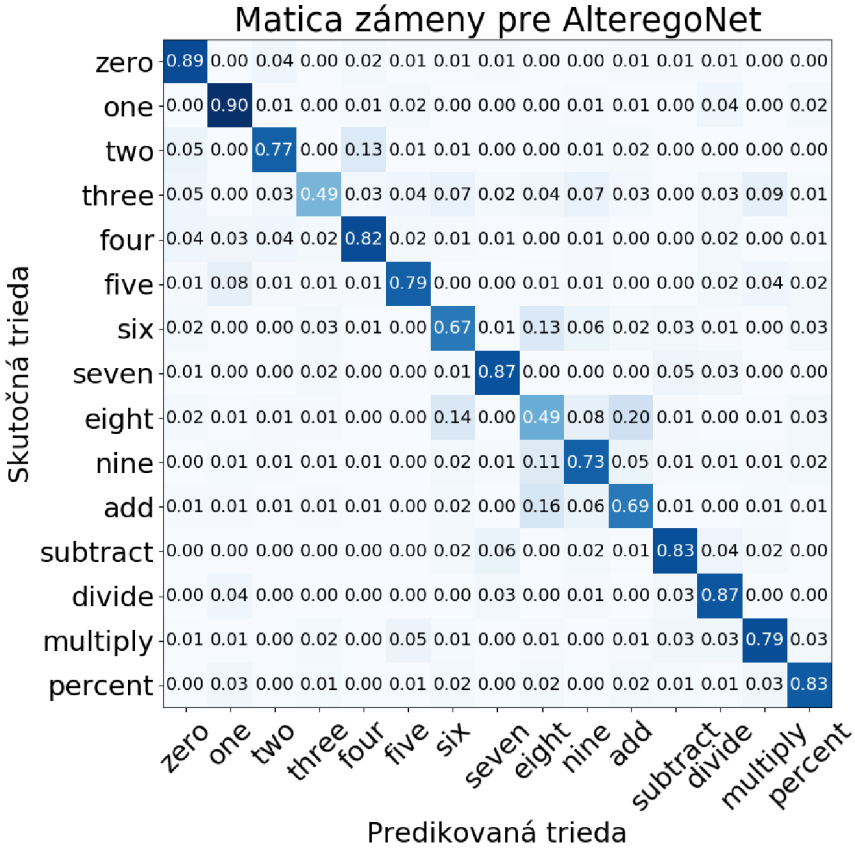
7.2.2 Výsledky krížovej validácie

Pre vyhodnotenie boli použité dáta zo všetkých šiestich sedení. Tie boli spojené dohromady, premiešané a následne rozdelené pomocou 5-násobnej krížovej validácie. Tabuľka 7.1 obsahuje presnosť oboch navrhnutých modelov pre každú z piatich skupín rozdelenia a taktiež priemernú presnosť. Vyššiu priemernú presnosť dosiahla architektúra SilentNet – 82,01 %.

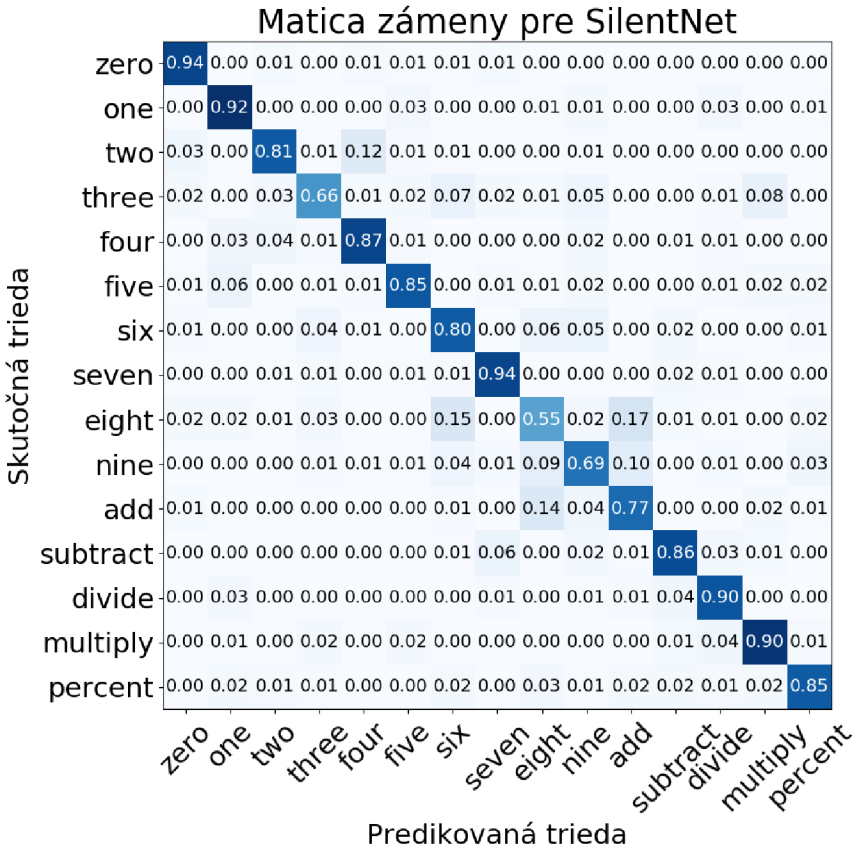
Tabuľka 7.1: Presnosť [%] klasifikácie v jednotlivých iteráciách (I1-I5) krížovej validácie.

	AlteregoNet	**SilentNet**
I1	76.19	84.07
I2	76.07	82.23
I3	74.82	82.23
I4	77.95	82.50
I5	76.52	79.02
Priemer	**76.32**	**82.01**

Ešte detailnejšie informácie je možné vyčítať z matice zámeny pre AlteregoNet na obrázku 7.2 a pre SilentNet na obrázku 7.3.



Obr. 7.2: Matica zámeny sumarizujúca výsledky krížovej validácie pre AlteregoNet.



Obr. 7.3: Normalizovaná matica zámeny sumarizujúca výsledky klasifikácie pri použití krížovej validácie a architektúry SilentNet. Použité boli nahrávky oboch typov reči (silent a audible) zo všetkých sedení (S1-S6).

Krásne z nich vidno metriky pre jednotlivé triedy a taktiež, ktoré triedy sa najčastejšie zamieňajú. Výrazne najhoršie výsledky dosiahla trieda `eight`. Najčastejšie bola zamieňaná s triedou `add`, čo sa dá pochopiť vzhľadom na veľmi podobnú výslovnosť. Zaujímavé je tiež, že zlé výsledky dosiahla trieda `three`. Toto prisudzujem faktu, že jej zastúpenie v dátovej sade je zo všetkých tried najmenšie. Rozdiel oproti triede `zero`, ktorá je druhou najmenej zastúpenou triedou a obsahuje len o dve vzorky dát viac je napriek tomu veľký, takže nie je vylúčené, že je to spôsobené aj inými faktormi.

7.2.3 Výsledky v rôznych sedeniach

Tentokrát bola meraná presnosť v jednotlivých sedeniach. Jedno sedenie bolo vždy použité ako testovacie a ostatné boli použité na trénovanie. Tým, že trénovacia sada neobsahuje žiadne dáta zo sedenia použitého pre testovanie sa podstatne sťažili podmienky pre klasifikáciu. Pre upresnenie treba dodať, že pri klasifikácii len jedného typu reči (silent alebo loud) nebol na trénovanie použitý ani opačný typ reči z testovacieho sedenia. Je to z dôvodu, aby bolo možné zistiť, akú presnosť bude dosahovať klasifikácia na nevidených sedeniach. Práve takto dosiahnuté výsledky by mohli vytvoriť predstavu o fungovaní za predpokladu,

že polohy EMG elektród sa v každom sedení mierne zmenia. Výsledky sú zhrnuté v tabuľke 7.2. Výrazne najlepšie výsledky boli vo všetkých prípadoch dosiahnuté v sedení S3, kde bola dosiahnutá presnosť klasifikácie až 87,87 % pre audible speech a 79,62 %. Predpokladám, že to bolo spôsobené práve dobrým umiestnením EMG elektród. Taktiež je možné spozorovať, že rozdiely medzi presnosťou klasifikácie audible a silent speech sa postupne zmenšovali. V sedení S1 bola presnosť klasifikácie audible speech až o 12,5 % vyššia ako presnosť klasifikácie silent speech. Tento rozdiel sa postupne zmenšoval a v sedení S5 dokonca dosiahla vyššiu presnosť klasifikácia silent speech. Toto len povtrdzuje zistenie z výskumu [16], kde sa uvádza, že presnosť klasifikácie silent speech rástla s pribúdajúcimi skúsenosťami rečníka.

Tabuľka 7.2: Presnosť [%] klasifikácie daného typu reči (Silent, Audible, Oba) v jednotlivých sedeniach. Sedenia S1, S3, S4 a S5 označujú testovacie sedenie. Pre trénovanie boli použité všetky ostatné sedenia.

	AlteregoNet			**SilentNet**		
	Audible	**Silent**	**Oba**	**Audible**	**Silent**	**Oba**
S1	75.76	55.00	65.38	82.58	60.00	71.29
S3	82.62	77.38	80.00	87.87	79.62	83.75
S4	69.12	76.00	72.56	77.88	74.62	76.25
S5	68.54	68.83	68.68	73.44	77.87	76.63
Priemer	**74.01**	**64.53**	**71.66**	**80.44**	**73.03**	**76.98**

7.3 Experimenty

V tejto sekcii sa nachádza mnoho experimentov, ktoré boli vykonané či už za účelom pokusov o zlepšenia výsledkov, alebo za účelom overiť presnosť klasifikácie za odlišných podmienok. Okrem vyhodnotenia výsledkov v predchádzajúcej sekcii, ktoré dávajú predstavu o celkovej úspešnosti, bolo experimentované s viac špecifickým rozdelením dát na trénovanie a testovanie. Nachádzajú sa tu experimenty s použitím len jedného typu reči na trénovanie, s použitím opačného typu reči z testovaného sedenia a tiež s použitím menšieho počtu kanálov. Pre experimenty v tejto sekcii bola použitá architektúra SilentNet.

7.3.1 Dĺžka rámca a počet príznakov

Veľký vplyv na presnosť klasifikácie má voľba parametrov pre extrakciu príznakov. Konkrétne je potrebné zvoliť dĺžku rámca, overlap a v neposlednom rade počet príznakov extrahovaných z každého rámca. V existujúcich výskumoch sa uvádza použitie rôznych kombinácií týchto parametrov, preto je ťažké na základe nich jednu kombináciu vybrať. Okrem toho, ostatné výskumy používajú na klasifikáciu odlišné metódy, čo tiež môže mať vplyv na výber parametrov. Boli vyskúšané rôzne kombinácie počtu príznakov v rozsahu 5–20 a dĺžky rámca od 25 do 100 ms s rozdielnou hodnotou pre overlap. Ako prvé boli testované rôzne počty príznakov. Samozrejme pri tomto pokuse musela byť použiť nejaká dĺžka rámca a overlap. Zvolená bola dĺžka rámca 50 ms a overlap 25 ms. Po nájdení optimálneho počtu príznakov, konkrétne 10, boli skúšané rôzne kombinácie dĺžky rámca a hodnoty overlap. Výsledky pre jednotlivé kombinácie sa nachádzajú v tabuľke 7.3. Po nájdení vhodných hodnôt boli znova otestované rôzne počty príznakov. Výsledky sú v tabuľke 7.4.

Tabuľka 7.3: Presnosť [%] klasifikácie s použitím rôznych kombinácií hodnôt pre dĺžku okna a overlap. Použitých bolo 10 LFBE príznakov.

		Overlap [%]			
		0	**25**	**50**	**75**
Dĺžka rámca [ms]	**25**	74.97	76.15	76.40	74.25
	50	74.52	79.45	80.87	78.43
	100	65.93	72.87	80.50	**82.01**

Tabuľka 7.4: Presnosť [%] klasifikácie s použitím rôzneho počtu LFBE príznakov extrahovaného pre jeden rámec. Použitá dĺžka rámca bola 100 ms a overlap 75 ms.

Počet príznakov	**5**	**10**	**15**	**20**
Presnosť [%]	76.90	**82.01**	81.23	81.25

7.3.2 Kompresia príznakov pomocou autoenkóderu

Po extrakcii príznakov z EMG signálu ešte prebehol experiment so snahou o ich kompresiu pomocou autoenkóderu. Ako autoenkóder bola použitá neurónovú sieť s dvomi plne prepojenými skrytými vrstvami, podobne ako to bolo spravené v predchádzajúcom výskume. Ako vstup a výstup bolo použitých 80 LFBE získaných z EMG (10 LFBE×8 kanálov). Na trénovanie boli použité TIMIT dáta rozdelené na rámce o veľkosti 100 ms prekrývajúce sa o 90 ms. Takéto veľké prekrytie bolo použité z dôvodu, že šlo o to získať čo najviac rámcov na trénovanie, keďže vstupom autoenkóderu bol jeden rámec. Po natrénovaní bol autoenkóder použitý na redukciu 80 LFBE príznakov extrahovaných z každého rámca na 10 príznakov. Týmto spôsobom ale došlo k zhoršenie výslednej presnosti. Lepším riešením mohlo byť použiť ako autoenkóder CNN, prípadne prispôsobenie architektúry klasifikátoru. V tomto prípade ale šlo len o experiment, ktorého cieľom bolo zistiť, či je možné dosiahnuť vyššiu presnosť klasifikácie aj s použitím jednoduchej architektúry autoenkóderu.

7.3.3 Použitie opačného typu reči z testovacieho sedenia

Pri vyhodnocovaní výsledkov v kapitole 7.2.3 neboli na trénovanie použité dáta opačného typu zo sedenia, ktoré bolo použité na testovanie. To znamená, že ak boli na testovanie použité napríklad silent dáta zo sedenia S1, audible dáta zo sedenia S1 vôbec neboli použité. Preto bol uskutočnený experiment, v ktorom boli počas trénovania použité aj tieto dáta. Predpokladalo sa, že to bude viesť k lepším výsledkom vzhľadom na to, že počas jedného sedenia sa poloha EMG elektród nemenila, teda by silent a audible dáta z rovnakého sedenia mali byt dosť podobné. Tento predpoklad sa aj naplnil, keďže sa presnosť klasifikácie pre silent zlepšila o 5,86 % a 4,42 % pre audible. Výsledky pre jednotlivé sedenia sú uvedené v tabuľke 7.5.

7.3.4 Použitie len jedného typu reči na trénovanie

Ďalším z experimentov bolo vyhodnotenie s použitím len jedného typu reči. Ako sa dalo očakávať, presnosť pri audible speech bola vyššia. Výsledky boli získané ako priemer jednotlivých presností získaných pomocou 5-násobnej krížovej validácie. Oproti priemernej

Tabuľka 7.5: Presnosť [%] s použitím opačného typu reči z testovacieho sedenia (S1-S5).

	Loud	**Silent**
S1	84.85	65.45
S3	93.50	86.00
S4	84.00	81.75
S5	77.08	82.34
Priemer	**84.86**	**78.89**

presnosti 82,01 % pri použití oboch typov reči, bola priemerná presnosť pre audible speech 78.77 % a 74.24 % pre silent speech. Presnosť pre silent speech je nižšia aj napriek tomu, že silent dát je podstatne viac. Tiež bola porovnaná presnosť klasifikácie v jednotlivých sedeniach, ktorá sa nachádza v tabuľke 7.6. Tu bol pre zaujímavosť model natrénovaný na jednom type reči vyhodnotený aj na opačnom type reči. Dopadlo to podľa očakávaní, teda presnosť klasifikácie opačného typu reči bola výrazne nižšia.

Tabuľka 7.6: Presnosť [%] klasifikácie s použitím len jedného typu reči (Audible alebo Silent) na trénovanie. Trénovanie označuje typ reči použitý pri trénovaní a testovanie typ reči použitý pri testovaní.

Trénovanie	**Audible**			**Silent**		
Testovanie	**Audible**	**Silent**	**Oba**	**Audible**	**Silent**	**Oba**
S1	76.36	38.79	57.58	51.82	47.58	49.70
S3	71.75	41.75	56.75	64.75	79.25	72.00
S4	75.25	54.25	64.75	47.25	73.25	60.25
S5	64.38	44.47	54.43	37.08	71.70	54.21
Priemer	**71.93**	**44.81**	**58.20**	**50.23**	**67.94**	**59.04**

7.3.5 Vplyv jednotlivých elektród

V tabuľke sa nachádza presnosť pre rôzne kombinácie použitých elektród, teda v praxi kombinácie kanálov s tým, že vždy jeden bol vynechaný. Ak by bez niektorého z kanálov boli dosiahnuté lepšie výsledky, znamenalo by to nielen že daný kanál neprispieva k zlepšeniu výsledkov, ale dokonca ich kazí. Jednotlivé presnosti sú veľmi podobné až na kanál CH4, bez ktorého došlo k najvyššiemu poklesu presnosti. Z toho sa dá usúdiť, že práve tento kanál má najväčší vplyv na výsledok. Na obrázku 4.1 je možne vidieť, že kanál CH4 sa nachádza na brade. Ďalej nepoužitie kanálu CH8 prinieslo najlepšie výsledky. Práve EMG elektróda použitá pre nahranie kanálu CH8 bola pridaná navyše oproti siedmim EMG elektródam vo výskume [11]. Z tohto experimentu vyplýva, že jej použitie nezlepšuje presnosť klasifikácie.

Tabuľka 7.7: Presnosť [%] pri vynechaní rozdielnych kanálov. CH1 - CH8 označuje vynechaný kanál.

	CH1	**CH2**	**CH3**	**CH4**	**CH5**	**CH6**	**CH7**	**CH8**
Presnosť [%]	82.06	81.59	82.10	78.17	80.85	81.10	80.82	82.74

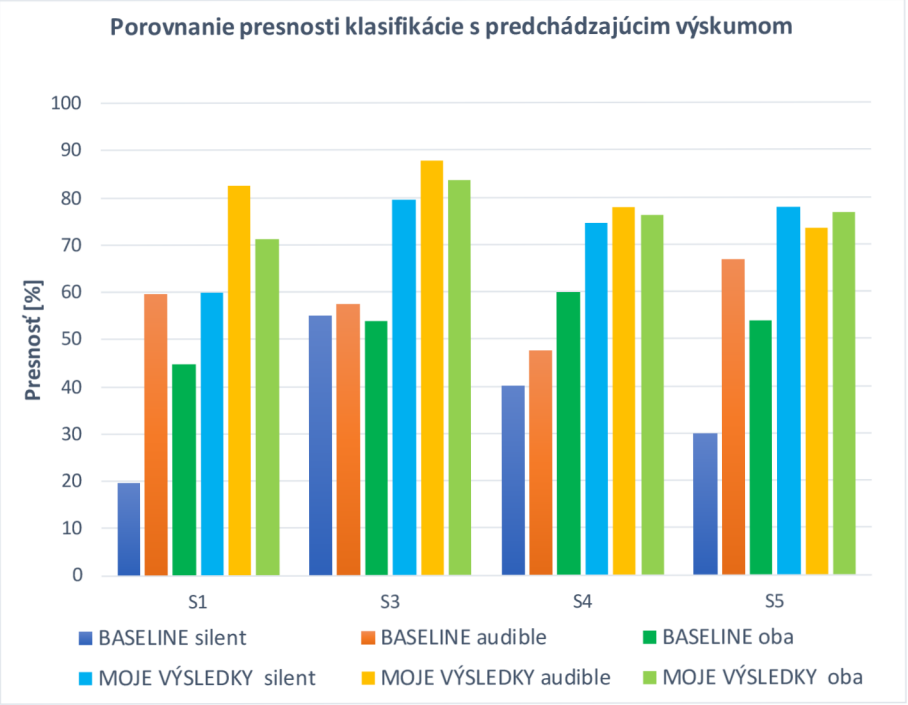
7.4 Zhrnutie výsledkov a porovnanie s predchádzajúcim výskumom

Dátová sada obsahuje len 561 nahrávok dlhých približne dve sekundy, čo je výrazne menej ako vo výskume AlterEgo [11], kde bolo spolu nahraných približne 31 hodín reči. Aj napriek malému množstvu dostupných dát bola dosiahnutá vysoká presnosť klasifikácie. Pri náhodnom predikovaní 15 tried je šanca, že sa predikovaná trieda bude zhodovať so skutočnou len 6,67%. S touto hodnotou sa porovnával predchádzajúci výskum, keďže cieľom bolo hlavne ukázať, že dosiahnuť výrazne vyššiu presnosť klasifikácie aj s takto malou dátovou sadou je možné. Priemerná presnosť klasifikácie s použitím oboch typov reči tu dosiahla 53,03 % [3] a to aj pri použití dát z rovnakého sedenia. V tejto práci bola úloha mierne sťažená tým, že pri vyhodnocovaní presnosti klasifikácie pre dané sedenie pri trénovaní neboli použité žiadne nahrávky z tohto sedenia. Aj napriek tomu bola dosiahnutá priemerná presnosť pre oba typy reči až 76,98 %. Pri klasifikácii jedného typu reči bola priemerná presnosť v sedeniach 80,44 % pre audible a 73,03 % pre silent speech, pri čom najvyššia presnosť dosiahnutá v sedení S3 je až 87,87 % pre audible a 79,62 % pre silent speech. Tieto hodnoty sa pri použití nahrávok opačného typu reči z rovnakého sedenia ešte zvýšili, čím sa v sedení S3 podarilo dosiahnuť až 93,50 % presnosť pre audible speech a 86,00 % pre silent speech. Pri trénovaní s rovnakým typom reči z rovnakého sedenia by sa presnosť pravdepodobne ešte zvýšila. Tieto výsledky môžu slúžiť ako predstava o tom, ako by sa presnosť zvýšila, keby bola zaručená rovnaká pozícia EMG elektród v rozdielnych sedeniach. K nejakým rozdielom v umiestnení elektród pravdepodobne dôjde, preto považujem výsledky z tabuľky 7.2 za viac približujúce sa výsledkom pri reálnom použití. Preto sú práve tieto hodnoty použité pri porovnaní s predchádzajúcim výskumom, ktoré je zhrnuté na obrázku 7.4.

7.5 Smery ďalšieho vývoja

Dosiahnuté výsledky sú dobré, no je možné ich ďalej zlepšovať. K tomu by mohlo pomôcť nahranie ďalších dát. Zaujímavé by bolo overiť aj ako sa zmení úspešnosť klasifikácie pri použití dát od viacerých rečníkov. Považujem to za obzvlášť dôležité vzhľadom k tomu, že schopnosť naučiť sa klasifikovať príkazy nezávisle od rečníka je podľa mňa pre praktické využitie kľúčová. Ďalej by bolo určite zaujímavé rozšíriť slovnú zásobu a overiť, ako sa zmení presnosť klasifikácie. Čo sa týka spracovania EMG signálu, tiež je tu priestor na experimentovanie vzhľadom na to, že existuje mnoho rôznych používaných techník. Čo sa týka samotného klasifikátoru, po výskume [11] sa tiež potvrdilo, že CNN sú vhodnou voľbou pre klasifikáciu príkazov z EMG signálu. To nemusí znamenať, že sú jedinou vhodnou voľbou a stále ostáva veľa priestoru pre výskumy s použitím iných typov neurónových sietí.

[3] Vyrátaná ako priemer presnosti klasifikácie v sedeniach S1, S3, S4 a S5.



Obr. 7.4: Porovnanie presnosti klasifikácie rôznych typov reči (silent, audible a oba) v jednotlivých sedeniach. BASELINE sú hodnoty dosiahnuté v predchádzajúcom výskume a MOJE VÝSLEDKY hodnoty dosiahnuté v tejto práci.

Kapitola 8

Záver

V tejto práci bol na úvod predstavený koncept rozpoznávania reči z EMG signálu. Pre lepšiu predstavu sa v kapitole 2 nachádza popis EMG a popis jednotlivých typov reči, ktoré boli pomocou EMG nahrávané. V tejto kapitole sa tiež nachádza zhrnutie existujúceho výskumu v tejto oblasti a používaných techník. Väčšina z týchto výskumov používala pre klasifikáciu iné metódy ako neurónové siete, takže táto oblasť je ešte relatívne nová a nepreskúmaná. Práve CNN sa javili ako vhodný typ pre klasifikáciu príkazov z EMG, preto sa práca detailnejšie zameriava na ich popis.

Neurónové siete ale reálne nie sú použiteľné kým nie sú natrénované na dátach. Popisu dostupných dát sa venuje kapitola 4. Nachádza sa v nej zoznam 15 príkazov, ktoré boli klasifikované a počet ich nahrávok v jednotlivých sedeniach. Celkovo bolo k dispozícii len 561 nahrávok príkazov, z čoho 178 bolo nahraných pri reči typu audible a 383 pri reči typu silent. Okrem popisu obsahu dátovej sady a jej štruktúry boli v tejto kapitole navrhnuté aj spôsoby ich rozdelenia pre účely trénovania a testovania neurónovej siete. Pre účely ladenia architektúry a hyperparametrov modelu bolo použité prevažne rozdelenie pomocou metódy 5-násobnej krížovej validácie. Pri vyhodnocovaní úspešnosti boli dáta rozdelené aj s ohľadom na sedenia, v ktorých boli nahrané.

Neurónová sieť by teoreticky mohla byť natrénovaná aj priamo s použitím EMG signálu, ale vzhľadom k malej veľkosti dátovej sady to nebolo vhodné. Z EMG signálu boli extrahované príznaky, ktoré ho charakterizujú menším počtom hodnôt (v sekcii 5.2). Existuje mnoho spôsobov extrakcie príznakov z EMG signálu, takže bolo dôležité nájsť vhodný typ príznakov. Väčšina existujúcich výskumov dosahujúcich najlepšie výsledky používala MFCC príznaky, preto boli v tejto práci použité tiež. Presnejšie boli zvolené im podobné LFBE príznaky, vhodnejšie pre použitie s neurónovými sieťami, ktorých extrakcia prebieha rovnako ako extrakcia MFCC s vynechaním niektorých krokov. Po extrakcii príznakov nasledoval pokus o ich kompresiu použitím autoenkóderu, ale neviedlo to k zlepšeniu výsledkov. Pre zvýšenie kvality výsledných príznakov bol pred ich extrakciou EMG signál spracovaný. Konkrétne z neho bol odstránený šum použitím rôznych filtrov. Najlepšie výsledky boli dosiahnuté pri odfiltrovaní PLI s frekvenciou 60 Hz a jej harmonických frekvencií. Okrem toho boli z EMG signálu odfiltrované frekvencie nižšie ako 5 Hz a frekvencie vyššie ako 450 Hz.

Extrahované príznaky boli použité ako vstup klasifikátoru. V kapitole 6 boli celkovo navrhnuté dve architektúry CNN pre klasifikáciu príkazov. Prvá architektúra nazvaná AlteregoNet je založená na popise existujúceho riešenia, ktorý je veľmi stručný. Ďalej bola navrhnutá vlastná architektúru nazvaná SilentNet. Návrh architektúry spočíval z veľkej časti v experimentovaní, keďže návrh architektúry neurónových sietí je náročný proces.

Navrhnuté architektúry CNN boli nakoniec implementované a natrénované. Výsledky sú zhrnuté v kapitole 7. Použitím navrhnutých CNN pre klasifikáciu sa podarilo podstatne zlepšiť výsledky oproti predchádzajúceho výskumu. Z použitých architektúr dosiahla lepšie výsledky architektúra SilentNet. Pri klasifikácii audible speech bola v jednom zo sedení dosiahnutá presnosť až 93,50 % a 86,00 % pre silent speech. Oproti tomu najvyššia dosiahnuté presnosť v predchádzajúcom výskume bola 66,7 % pre audible speech a 61,4 % pre silent speech. V rámci viacerých experimentov, ktoré porovnávali presnosť silent a audible speech, bolo zistené, že presnosť klasifikácie audible speech je celkovo vyššia, ale oproti presnosti dosiahnutej pri silent speech nie je až tak rozdielna. Ďalším zaujímavým zistením je, že presnosť klasifikácie silent speech sa v priebehu sedení približovala presnosti klasifikácie audible speech a v sedení S5 ho dokonca prekonala.

Pre ďalší vývoj by bolo vhodné nahrať viacej dát s použitím viacerých rečníkov a taktiež by mohlo byť zaujímavé vyskúšať rozšíriť slovnú zásobu. Okrem toho vidím priestor na zlepšenie v oblasti spracovania EMG signálu a určite stojí za pokus vyskúšať aj iné typy neurónových sietí. Stále ide o pomerne mladú a nepreskúmanú oblasť výskumu, no predpokladám, že sa v nej v blízkej dobe uskutoční mnoho výskumov, vzhľadom na široké možnosti využitia.

Literatúra

[1] *Mel Frequency Cepstral Coefficient (MFCC) tutorial* [online]. [cit. 2020-05-14]. Dostupné z: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/.

[2] Filtering the surface EMG signal: Movement artifact and baseline noise contamination. *Journal of Biomechanics*. 2010, roč. 43, č. 8, s. 1573 – 1579. Dostupné z: https://doi.org/10.1016/j.jbiomech.2010.01.027. ISSN 0021-9290.

[3] *Generalization in Neural Networks* [online]. 2018 [cit. 2020-05-14]. Dostupné z: https://www.deeplearningdemystified.com/article/fdl-5.

[4] Chowdhury, R., Reaz, M. B. I., Mohd Ali, M., A Bakar, A. A., Chellappan, K. et al. Surface Electromyography Signal Processing and Classification Techniques. *Sensors (Basel, Switzerland)*. September 2013, roč. 13, s. 12431–12466. Dostupné z: https://doi.org/10.3390/s130912431. ISSN 424-8220.

[5] De Luca, C. Electromyography. In: *Encyclopedia of Medical Devices and Instrumentation*. American Cancer Society, 2006, s. 99–105. Dostupné z: https://doi.org/10.1002/0471732877.emd097. ISBN 9780471732877.

[6] Deng, Y., Heaton, J. a Meltzner, G. Towards a practical silent speech recognition system. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. 2014, s. 1164–1168.

[7] Garofolo, J. S. TIMIT acoustic phonetic continuous speech corpus. *Linguistic Data Consortium, 1993*.

[8] Godfrey, A. E. Speech. Encyclopædia Britannica, inc. [online]. 2019, [cit. 2020-01-04]. Dostupné z: https://www.britannica.com/topic/speech-language.

[9] Goodfellow, I., Bengio, Y. a Courville, A. *Deep Learning*. MIT Press, 2016. Dostupné z: http://www.deeplearningbook.org. ISBN 9780262337373.

[10] He, K., Zhang, X., Ren, S. a Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *IEEE International Conference on Computer Vision (ICCV 2015)*. Február 2015, roč. 1502. Dostupné z: https://doi.org/10.1109/ICCV.2015.123.

[11] Kapur, A., Kapur, S. a Maes, P. Alterego: A personalized wearable silent speech interface. In: ACM. *23rd International Conference on Intelligent User Interfaces*. 2018, s. 43–53. Dostupné z: 10.1145/3172944.3172977.

[12] Kingma, D. P. a Ba, J. *Adam: A Method for Stochastic Optimization.* 2014. Dostupné z: http://arxiv.org/abs/1412.6980.

[13] Li, F.-F., Karpathy, A. a Johnson, J. CS231n: Convolutional Neural Networks for Visual Recognition. [online]. 2016. Dostupné z: https://cs231n.github.io/neural-networks-1.

[14] Luo, P., Wang, X., Shao, W. a Peng, Z. *Towards Understanding Regularization in Batch Normalization.* 2018.

[15] Maegherman, G., Nuttall, H. E., Devlin, J. T. a Adank, P. Motor Imagery of Speech: The Involvement of Primary Motor Cortex in Manual and Articulatory Motor Imagery. *Frontiers in Human Neuroscience.* 2019, roč. 13, s. 195. Dostupné z: https://doi.org/10.3389/fnhum.2019.00195. ISSN 1662-5161.

[16] Maier-Hein, L., Metze, F., Schultz, T. a Waibel, A. Session independent non-audible speech recognition using surface electromyography. In: *IEEE Workshop on Automatic Speech Recognition and Understanding, 2005.* 2005, s. 331–336. Dostupné z: https://doi.org/10.1109/ASRU.2005.1566521.

[17] Meltzner, G., Heaton, J., Deng, Y., Luca, G., Roy, S. et al. Development of sEMG sensors and algorithms for silent speech recognition. *Journal of Neural Engineering.* 2018, roč. 15. Dostupné z: https://doi.org/10.1088/1741-2552/aac965.

[18] Mohamed, A. rahman. *Deep Neural Network Acoustic Models for ASR.* 2014. 43 – 55 s. Dizertačná práca. University of Toronto.

[19] Morin, A. Inner Speech. In:. 2012. Dostupné z: http://doi.org/10.1016/B978-0-12-375000-6.00206-8.

[20] Morse, M. S. a O'Brien, E. M. Research summary of a scheme to ascertain the availability of speech information in the myoelectric signals of neck and head muscles using surface electrodes. *Computers in Biology and Medicine.* 1986, roč. 16, č. 6, s. 399 – 410. Dostupné z: https://doi.org/10.1016/0010-4825(86)90064-8. ISSN 0010-4825.

[21] Refaeilzadeh, P., Tang, L. a Liu, H. Cross-Validation. *Encyclopedia of Database Systems.* Január 2009, 532–538, s. 532–538. Dostupné z: https://doi.org/10.1007/978-0-387-39940-9_565.

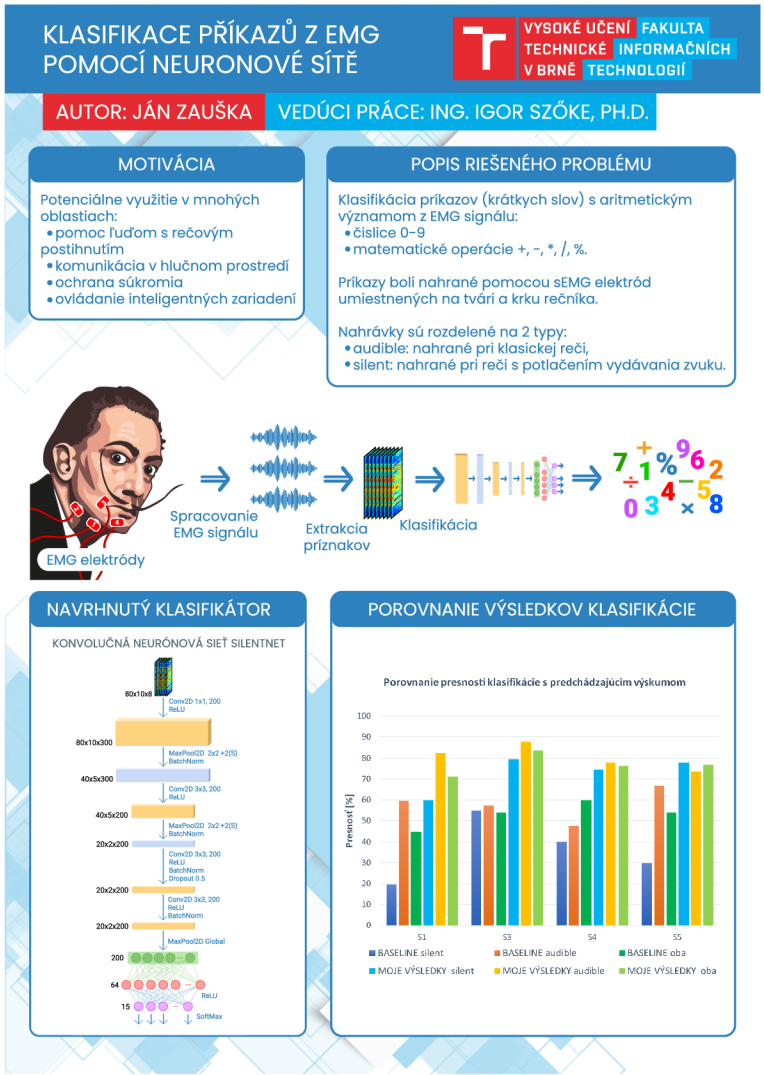
[22] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. a Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research.* 2014, roč. 15, č. 56, s. 1929–1941. Dostupné z: http://jmlr.org/papers/v15/srivastava14a.html.

[23] Wand, M. *Advancing Electromyographic Continuous Speech Recognition: Signal Preprocessing and Modeling.* 1. vyd. KIT Scientific Publishing, 2015. 30–32 s. ISBN 9783731502111.

[24] Zhang, A., Lipton, Z. C., Li, M. a Smola, A. J. *Dive into Deep Learning.* 2019. 231–257 s. Dostupné z: http://www.d2l.ai.

Príloha A

Plagát



Obr. A.1: Plagát prezentujúci výsledky práce.

Príloha B

Obsah priloženého pamäťového média

- **src**: zdrojové súbory pre spracovanie dát a klasifikáciu
- **sample_data**: ukážka použitej dátovej sady
- **trained_models**: natrénované modely
- **bp.pdf**: dokumentácia k bakalárskej práci
- **src_latex**: zdrojové súbory dokumentácie
- **poster.pdf**: plagát prezentujúci výsledky práce
- **video.mp4**: video prezentujúce výsledky práce
- **predchadzajuci_vyskum_ts.pdf**: technická správa z predchádzajúceho výskumu
- **manual.md**: popis spustenia spracovania dát a klasifikátoru