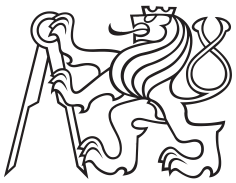


České
vysoké
učení technické
v Praze

F3

Fakulta elektrotechnická
Katedra radioelektroniky

Porovnání modelů lokalizace zdroje zvuku člověkem

Aneta Drhová

Školitel: Ing. František Rund, Ph.D.

2022

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Drhová** Jméno: **Aneta** Osobní číslo: **483456**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra radioelektroniky**

Studijní program: **Elektronika a komunikace**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Porovnání modelů lokalizace zdroje zvuku člověkem

Název bakalářské práce anglicky:

Comparison of Models of Human Sound Source Localization

Pokyny pro vypracování:

Seznamte se s problematikou lokalizace zdroje zvuku u člověka a s dostupnými modely této lokalizace. Porovnejte minimálně dva dostupné modely z hlediska přesnosti lokalizace, výpočetní náročnosti, popř. dalších kritérií. Výsledky zkoumaných modelů vyhodnoťte vzhledem k subjektivním datům.

Seznam doporučené literatury:

[1] Koshkina, E: Využití strojového učení pro modelování binaurálního slyšení. Diplomová práce FEL ČVUT 2017, dostupné z http://hdl.handle.net/10467/68455

[2] Bouse, J. – Vencovský, V. – Rund, F. – Marsalek, P. Functional Rate-Code Models of the Auditory Brainstem for Predicting Lateralization and Discrimination Data of Human Binaural Perception. The Journal of the Acoustical Society of America. 2019, 145(1), 1-15. ISSN 0001-4966. DOI https://dx.doi.org/10.1121/1.5084264

Jméno a pracoviště vedoucí(ho) bakalářské práce:

Ing. František Rund, Ph.D. katedra radioelektroniky FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **15.09.2021** Termín odevzdání bakalářské práce: **20.05.2022**

Platnost zadání bakalářské práce: **19.02.2023**

Ing. František Rund, Ph.D.	doc. Ing. Stanislav Vítek, Ph.D.	prof. Mgr. Petr Páta, Ph.D.
podpis vedoucí(ho) práce	podpis vedoucí(ho) ústavu/katedry	podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Studentka bere na vědomí, že je povinna vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

. Datum převzetí zadání

Podpis studentky

Poděkování

Tímto bych chtěla poděkovat Ing. Františku Rundovi, Ph.D. za odborné vedení práce, rady a připomínky k formě i obsahu a za čas, který mi věnoval. Dále děkuji své rodině a přátelům za podporu během celého studia.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze, 20. 5. 2022

Abstrakt

Tato bakalářská práce se zabývá problematikou lokalizace zdroje zvuku člověkem. Byly vytvořeny dva modely napodobující chování lidského slyšení se zaměřením na určování polohy zdroje umístěného v prostoru. Zkoumán je zejména vliv přítomnosti binaurálního modelu implementovaného v jednom z vytvořených algoritmů. K vyhodnocení zpracovaných dat je v rámci obou modelů provedeno za pomoci strojového učení. Testy provedené na obou implementovaných modelech neprokázaly výrazné rozdíly v přesnosti lokalizace. Pro tvorbu modelů a testování jejich přesnosti lokalizace bylo použito počítačové prostředí Matlab.

Klíčová slova: MATLAB, lokalizace, binaurální model slyšení, HRTF, strojové učení, umělá neuronová síť

Školitel: Ing. František Rund, Ph.D.

Abstract

This thesis focuses on the topic of human sound source localization. Two models were created to simulate the behaviour of human hearing in respect to its ability to localise sound source in space. Important subject of the study is the impact of binaural model being implemented in one of the algorithms. Both models include machine learning as a means of evaluation of the processed data. Tests performed on both implemented models did not show significant differences in localization accuracy. Matlab environment is used for modelling and accuracy testing.

Keywords: MATLAB, localization, binaural auditory model, HRTF, machine learning, artificial neural network

Title translation: Comparison of Models of Human Sound Source Localization

Obsah

Obrázky

Tabulky

Kapitola 1
Úvod

Jednou z hlavních funkcí lidského ucha je lokalizace zdroje zvuku. Bakalářská práce se zabývá touto problematikou, tedy procesem určování polohy zdroje zvuku umístěného v prostoru.

Člověk vnímá zvuk dvěma ušima, využívá tedy tzv. binaurálního slyšení. Fyziologie lidského sluchu a způsoby lokalizace zdroje za pomoci binaurálního vnímání zvuku jsou popsány v kapitole 2. Vyhodnocení lokalizace zdroje lze provést například pomocí algoritmů strojového učení. Jeden z algoritmů je také použit v této práci. Teoretické poznatky týkající se strojového učení a umělých neuronových sítí jsou stručně vysvětlené v části práce 3.

Existuje celá řada binaurálních modelů navržených pro různé účely. Své uplatnění nachází například při vývoji sluchových pomůcek, pro měření kvality prostorových audio systémů a mnoho dalších. Kapitola 4 seznamuje stručně s různým pohledem na využití principu binaurálního slyšení, kterým se zmíněné vybrané práce zabývají.

Cílem této práce je především seznámení se s problematikou lokalizace zdroje zvuku člověkem. Za tímto účelem jsou v rámci práce implementovány dva modely lidského slyšení, vytvořené na základě prací [1] a [2]. Návrh a implementace modelů se primárně zaměřují na vliv přítomnosti binaurálního modelu lidského slyšení ve vytvořeném algoritmu lokalizace. Za účelem prokázání tohoto vlivu byl vytvořen jeden model, který nezahrnuje binaurální část a jeden model, který v sobě implementovanou binaurální část má. Práce se zabývá tím, jaký vliv tento rozdíl v implementaci může způsobit.

V části 5 je podrobně popsán proces tvorby obou modelů, včetně popisů jednotlivých částí algoritmu a tvorby vstupních dat. Kapitola 5.5 se zabývá trénováním použité neuronové sítě a procesem jejího testování na vytvořené množině vstupních dat. Následné výsledky přesnosti lokalizace obou implementovaných modelů jsou popsané v kapitole 6.

Tvorba modelů lidského slyšení a jejich následné trénování a testování je prováděno v počítačovém prostředí Matlab.

Kapitola 2
Teorie

2.1 Fyziologie sluchu

Sluch, jako schopnost vnímat zvukové podněty, je důležitým lidským smyslem. Fyziologie lidského sluchu je velmi obsáhlá a komplikovaná oblast a je mnoho věcí, které stále nevíme. Pro potřeby této práce jsou poznatky zjednodušeně vysvětleny v následující kapitole, která čerpá z literatury [3], [4].

Zvuk, jakožto mechanické vibrace, je přiváděn vnějším uchem do ucha středního, dále do ucha vnitřního, kde je transformován na bioelektrický impuls, který je dále zvukovou drahou veden ke zpracování do sluchového centra.

Vnější ucho např. podle [4] sestává z boltce a ze zvukovodu. Boltec díky svému tvaru umožňuje směrování akustických vln do zvukovodu a mj. spektrální filtraci zvuku v závislosti na směru dopadu, což napomáhá rozpoznat umístění zdroje zvuku. Akustická vlna takto dále postupuje vnějším uchem až k bubínku.

Bubínek je součást **středního ucha**, jedná se o tenkou membránu, která zvuk ve formě akustické vlny transformuje na mechanické vibrace kůstek, nacházejících se ve středním uchu. Díky pákovému mechanismu těchto středoušních kůstek, společně s poměrem plochy bubínku a oválného okénka dochází podle [3] na rozhraní mezi středním a vnitřním uchem k vysoké koncentraci akustického tlaku.

Zvuk se šíří dále do **ucha vnitřního**, ve kterém se nachází hlemýžď. Jedná se o stočenou trubičku, naplněnou kapalinou, která se při mechanických vibracích rozvlní a tím rozvibruje membránu, na které se nachází tzv. Cortiho orgán. Vnitřní ucho slouží k transformaci mechanického vlnění na bioelektrickou energii, k čemuž dochází dle [2] právě v Cortiho orgánu.

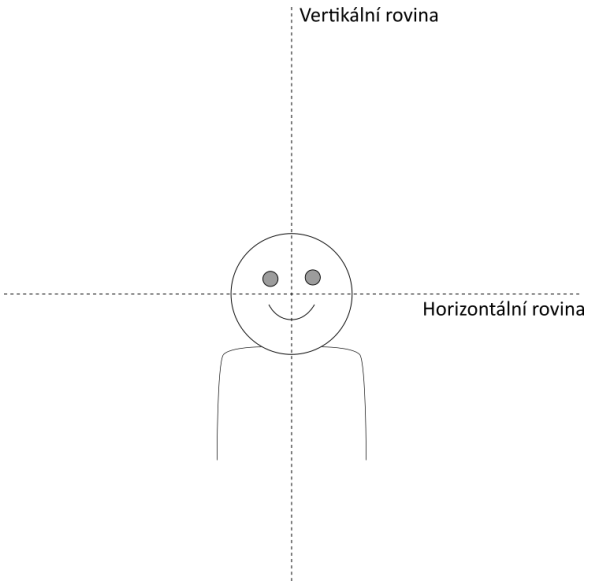
Bioelektrický signál je dále pomocí zvukového nervu poslán ke zpracování do mozku. Důležitou součástí drah sluchového systému je tzv. komplex horní olivy, **SOC** (Superior Olivary Complex), jedná se o hlavní místo zpracování sluchových informací. SOC se primárně rozděluje na laterární superior olivu, LSO (Lateral Superior Olive) a na mediální superior olivu, MSO (Medial Superior Olive).

2.2 Binaurální vnímání zvuku

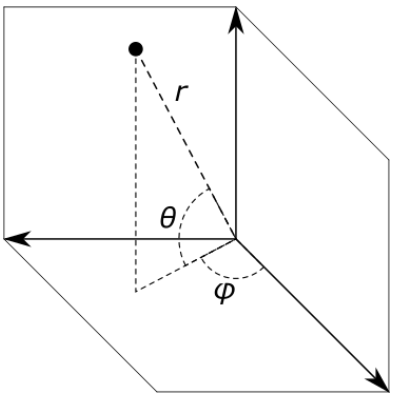
Člověk používá k vnímání zvuku v prostoru dvě uši, umístěné v horizontální rovině na opačných stranách hlavy, [1]. Schopnosti slyšet zvuk dvěma ušima se říká tzv. binaurální vnímání zvuku. Oproti poslechu zvuku jen jedním uchem, tzv. monaurální vnímání zvuku, přináší binaurální slyšení mnoho výhod. Jednou z nich je schopnost lokalizace zdroje zvuku v prostoru. Existuje řada způsobů, kterými člověk analyzuje příchozí zvuky, v následujících kapitolách je jeden z nich popsán.

2.2.1 Lokalizace zvuku

K lepšímu popisu prostoru je vhodné si jej rozdělit na dvě roviny, na rovinu horizontální a na rovinu vertikální, viz obr. 2.1. Polohu zdroje zvuku v takto rozděleném prostoru určujeme, stejně jako ve sférických souřadnicích, pomocí tří parametrů, pomocí vzdálenosti bodu od středu hlavy (r), úhlu v horizontální rovině a úhlu ve vertikální rovině, viz obr. 2.2 Lidskou hlavu můžeme aproximovat jako kouli, na jejichž opačných stranách jsou umístěny dvě uši, mezi kterými tím vzniká určitá vzdálenost, která hraje významnou roli při lokalizaci zdroje v horizontální rovině.



Obrázek 2.1: Horizontální a vertikální rovina vzhledem k modelu lidské hlavy



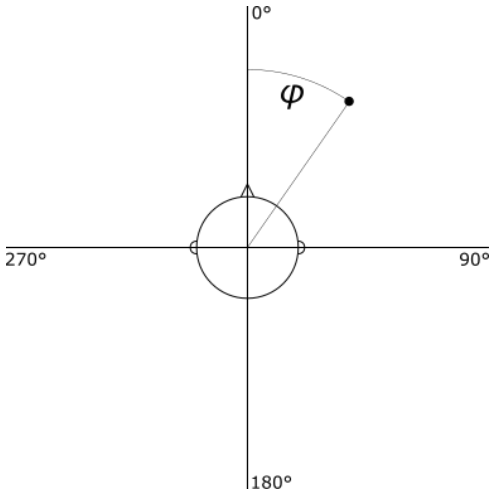
Obrázek 2.2: Umístění zdroje zvuku v prostorových souřadnicích

2.2.2 Horizontální rovina

Hlavními prvky používanými při určení zdroje zvuku v prostoru jsou dle [1] rozdíly intenzity a času příchozích zvuků do levého a pravého ucha.

Horizontální rovinu si můžeme rozdělit na tzv. azimutální úhly φ, viz obr. 2.3. V závislosti na tomto úhlu příchozího zvuku se dostane zvukový signál k jednomu uchu dříve než k uchu druhému. Výjimkou jsou případy pro 0° a 180°, kdy je zdroj umístěn v ose souměrnosti hlavy a akustické signály dorazí k oběma uším zároveň a časový rozdíl, tzv. ITD (Interaural Time Difference) je nulový. Oproti tomu nejvyšší časový rozdíl nastává v případech pro 90° a 270°, jedná se o situaci, kdy akustický signál přichází k hlavě z boku, kolmo na osu souměrnosti hlavy. [3] zmiňuje, že kromě směru příchozího zvuku má na hodnotu ITD nezanedbatelný vliv kmitočet příchozího zvuku. Lidské ucho je schopno tento časový rozdíl reagistrovat jen při dostatečně nízkých frekvencích.

Jak již bylo zmíněno, povrch hlavy je z fyzikálního hlediska překážkou v cestě akustického signálu, což způsobuje, že vzdálenější ucho se může dostat do tzv. akustického stínu. Šířící se vlny o vyšších frekvencích se na povrchu hlavy ohýbají a odráží a následně přichází ke vzdálenějšímu uchu s nižší hladinou intenzity zvuku, než k uchu umístěnému blíže ke zdroji. Tak vzniká úrovňový rozdíl v signálech příchozích k jednotlivým uším, tzv. ILD (Interaural Level Difference). Vzhledem k tomu, že tento jev je závislý na kmitočtu vlny, v určitých případech s rostoucí frekvencí roste ILD, tak při dostatečně nízkých frekvencích akustický stín nevzniká a hodnota ILD je nulová. V reálném prostředí tento jev nebývá jednoznačný, protože jak již bylo zmíněno, překážkou v cestě vln není jen povrch lidského těla, ale i okolní prostor, a tak intenzitní rozdíl signálů může být menší, žádný nebo naopak hladina odražených vln, příchozích ke vzdálenějšímu uchu, může být vyšší než k uchu bližšímu.



Obrázek 2.3: Úhel azimutu v horizontální rovině

2.2.3 Vertikální rovina

Ke zjištění polohy zdroje v prostoru však nestačí pouze zmíněný popis v horizontální rovině. Zejména v situacích, kdy se zdroj nachází ve stejné vzdálenosti od obou uší a ITD a ILD jsou nulové, je potřeba popsat polohu zdroje i v rovině vertikální. Lokalizace zvuku ve vertikální rovině je již obtížnější a na rozdíl od lokalizace v horizontální rovině v tomto případě člověk využívá k rozpoznání polohy zdroje zejména zrakových vjemů a zkušenosti. Vertikální rovinu si, podobně jako rovinu horizontální, můžeme rozdělit na tzv. elevační úhly θ.

Častým problémem při lokalizaci zvuku je tzv. předozadní zmatení, tedy stav, kdy je obtížné rozlišit, zda je zdroj zvuku umístěn vepředu či vzadu. Dochází tak dle [5] při poslechu nižších frekvencí, zhruba pod 300 Hz, což jsou vlnové délky, při kterých není hlava dostatečnou překážkou v cestě akustické vlny. Lokalizace ve vertikální rovině je uskutečněna především na základě spektrálních změn akustického signálu, které jsou dle [2] způsobeny geometrickými vlastnostmi ušního boltce, hlavy a těla.

Takový způsob určování polohy lze využít i při lokalizaci pouze jedním uchem, tzv. monaurální slyšení.

Kombinací obou způsobů lokalizace zvuku, monaurálního a binaurálního, lze dosáhnout možnosti detekce zvuku jak ve vertikální, tak v horizontální rovině prostoru.

2.3 HRTF

Zvuk je z fyzikálního hlediska mechanické vlnění částic prostředí. Zdroj zvuku rozkmitá okolní prostředí, čímž vznikají zvukové vlny sférického charakteru. V případě, že v cestě zvukových vln nestojí žádná překážka, vzniká dle [3] tzv. volné zvukové pole. Pokud však vlna narazí na překážku ve své cestě šíření, dochází buď k ohybu, tzv. difrakci nebo odrazu tzv. reflexi vlny.

Takovou překážkou v cestě vln mohou být na příklad stěny měřeného

prostoru, objekty v něm umístěné a neméně důležité lidské tělo. Během cesty zvukových vln, od zdroje zvuku k bubínku, dochází ke kontaktu s částmi lidského těla, na jehož povrchu dochází ke zmíněným difrakcím a reflexím zvukových vln, což má za následek změnu charakteru přijímaného zvuku. Vliv na tyto změny může mít například tvar trupu člověka, hlavy nebo samotného vnějšího ucha a na základě těchto změn lze získat i informaci o poloze zdroje zvuku.

K popisu změn charakteru zvuku šířícího se od zdroje, umístěného v prostoru, do lidského ucha se, viz [6], používá přenosová funkce hlavy, HRTF (Head-Related Transfer Function), jejíž součástí jsou zmíněné monaurální a binaurální informace. Vliv na tvar této funkce má, jak název napovídá, interakce šířící se vlny s povrchem lidské hlavy, případně i trupu. Přenosová funkce hlavy se získává pomocí Fourierovy transformace naměřených impulzních odezev hlavy, HRIR (Head-Related Impulse Response). Existují různé způsoby, jak získat jednotlivé sady HRTF, je například možné měřit je přímo. I metod měření HRIR existuje mnoho, jednou z možností je reproduktor, který vysílá z daného směru testovací zvuk do prostoru, ve kterém je umístěn model lidské hlavy. Příchozí zvukový signál je měřen na obou uších modelu a tím jsou započítané tzv. meziušní rozdíly signálů, podrobněji popsané v kapitolách 2.3 a 2.2.3. Měření se opakuje s mnoha různými polohami reproduktoru, čímž se získává řada impulzních odezev. Konkrétní získané přenosové funkce závisí na konkrétním tvaru těla, a proto jsou individuální pro každého člověka.

Kapitola 3
Strojové účení

Strojové učení je soubor algoritmů, které analyzují vzorová data a učí se z nich. Na základě takto získaných parametrů si vytváří datové modely, které následně využívá ke klasifikaci neznámých dat a formulovat předpovědi. Nejčastěji se využívají dva algoritmy strojového učení, tzv. učení s učitelem a učení bez učitele.

Strojové učení s učitelem spočívá v tom, že algoritmus má předem k dispozici tzv. trénovací množinu dat, jedná se o vstupní objekty, ke kterým jsou přiřazené požadované výstupy. V průběhu trénování se mění vnitřní stav systému podle toho, zda učinil správný závěr nebo ne. Algoritmus hledá neznámý vztah mezi vstupními a výstupním daty.

Pokud se algoritmus učí bez učitele, nemá poskytnutou trénovací množinu dat a učí se nezávisle na nějakém učiteli. Algoritmus hledá v neoznačených datech určité vzorové parametry (vlastnosti), které si rozdělí do vlastních tříd. Každá třída obsahuje skupinu podobných parametrů, které se naopak mezi třídami výrazně liší, vlastnosti tříd se tedy neprotínají. Když jsou algoritmu podána vstupní data neznámé třídy, posuzuje, v jaké míře obsahují dané parametry a podle toho zařadí data do příslušné třídy.

3.1 Umělá neuronová síť

Umělá neuronová síť, ANN (Artificial Neural Network), je dle [7] výpočetní model, který svým chováním napodobuje neuronové sítě v mozku živého organismu. Základní výpočetní jednotkou ANN jsou umělé neurony, které představují biologické neuronové buňky u organismů. Problematika fungování biologických nervových struktur je velmi obsáhlá a stále o ní nevíme mnoho. V dnešní době je možné simulovat skutečně komplikované a rozsáhlé struktury, ale od simulace sítí blízkých složitosti lidského mozku jsme daleko

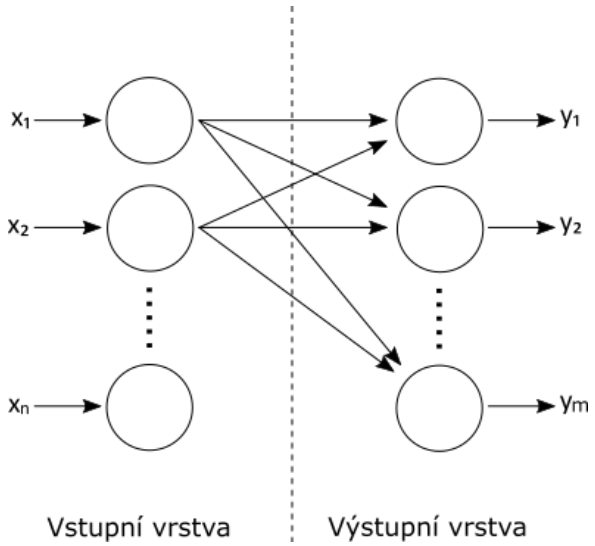
Umělé neurony jsou jednoduché matematické funkce, které zpracovávají vstupní signály, vytváří jim odpovídající výstupní signály a komunikují mezi sebou v rámci sítě. Každá takováto jednotka má jeden vstupní kanál a libovolný počet výstupních kanálů, tzv. synapsí, po kterých probíhá jednosměrná komunikace. Každý vstup x_i, který přichází do neuronové buňky je násoben odpovídajícím váhovacím koeficientem w_i, který je při procesu trénování neuronové sítě individuálně upravován. Z matematického hlediska váhovaný

výstup y_i neuronu je jednoduchým součtem všech vstupů násobených váhovacími koeficienty. Takováta hodnota je následně v neuronu upravena zkreslením b a tzv. aktivační funkcí $\phi(x)$, dle vztahu

$$y_i = \phi(b + \sum_{i=1}^{N} x_i \cdot w_i). \tag{3.1}$$

Neurony jsou v síti uspořádány do tzv. vrstev, které jsou systematicky propojeny tak, aby spolu neurony jednotlivých vrstev mohly komunikovat. Informace je přenášena mezi neurony dvou sousedních vrstev, a to zpravidla směrem od vrstvy vstupní k vrstvě výstupní, k přenosu informace mezi neurony v rámci jedné vrstvy nedochází. Používají se také sítě, které ve své sktruktuře zahrnují zpětné vazby a cykly, které představují schopnost přijímat již zpracované vstupy, mají tedy schopnost paměti. Existuje mnoho typů architektury neuronových sítí, z pohledu počtu vrstev je můžeme rozdělit na dva druhy.

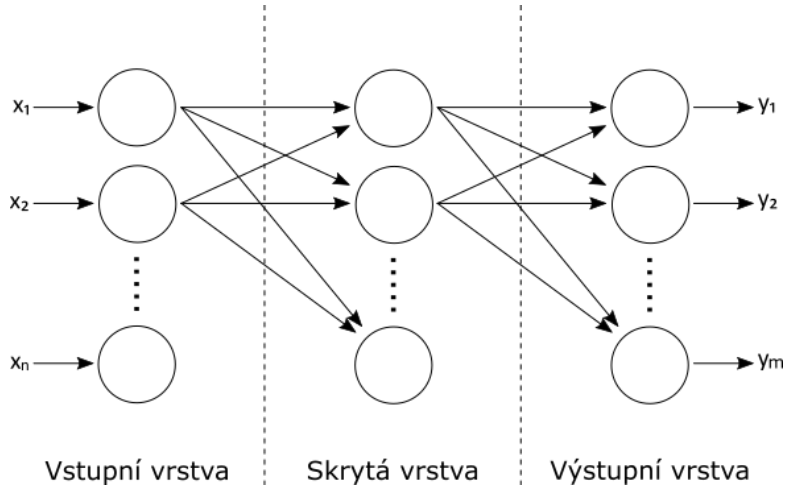
Základním a nejjednodušším uspořádáním je síť jednovrstvá. Skládá se pouze z vrstvy vstupní a výstupní, viz obr. 3.1.



Obrázek 3.1: Jednovrstvá síť

Oproti síti jednovrstvé obsahuje vícevrstvá síť alespoň jednu skrytou vrstvu, umístěnou mezi výstupní a vstupní vrstvu, viz obr. 3.2. Jednou z nejpoužívanějších architektur neuronových sítí je vícevrstvá dopředná síť, tzv. Feed-forward net, která umožňuje pouze přenos informace od vstupní vrstvy k vrstvě výstupní.

Díky přítomnosti skrytých vrstev nastává stav, kdy je každá další vrstva trénována na odlišné funkci závislé na výstupu předchozí vrstvy a síť je schopna rozpoznávat složitější funkce. Takové zpracování informací vytváří schopnost sítě lépe zobecňovat parametry vstupních dat, což zlepšuje účinnost při analyzování neznámých vstupních dat, na kterých nebyla síť natrénována.



Obrázek 3.2: Vícevrstvá síť

Kapitola 4

Dostupné binaurální modely

Existuje celá řada binaurálních modelů, které jsou navržené pro různé účely. Některé modely simulují fyziologii binaurálního slyšení a jiné využívají binaurálních principů za účelem zpracování signálů. Tato kapitola se některými z nich stručně zabývá.

4.1 Binaural analysis/synthesis of interior aircraft sounds

Autoři [8] se věnují analýze a syntéze zvuků v interiéru letadel. Práce se nejprve zabývá spektrálními a prostorovými parametry, určenými k popisu charakteristiky binaurálního zvuku uvnitř letadla. Následně použili model signál+šum, který rozšířili o principy interaurální koherence a interaurálního fázového rozdílu. Byl tedy navržen binaurální model tak, aby reprodukoval nejen spektrální, ale i prostorové charakteristiky zvuků. Následně byl tento algoritmus otestován a vyhodnocen na nahraném zvuku letadla.

4.2 Binaural Model Based Adaptive Binaural Noise Reduction

Vzhledem k tomu, že díky parametrům ITD a ILD je binaurální sluch schopen lokalizovat signály přicházejících různých směrů a také snižovat hladinu nežádoucí šumové signály, nachází v dnešní době binaurální technika řadu uplatnění. V práci [9] byla navržena metoda určena k redukci binaurálního šumu pro uživatele naslouchátek a tím napomáhá ve zlepšení kvality zvuku. Použitá metoda efektivně odděluje požadovaný zvuk od okolního šumu takovým způsobem, který velice věrně napodobuje lidský sluchový systém. K tomu autor použil známý model ke zpracování signálu, tzv. Boddenův procesor „koktejlové party“, toto schéma je však výpočetně velice náročné a je obtížné využívat ji v aplikacích v reálném čase. Proto navrhl zjednodušený binaurální model, fungující na základě detektoru zvukové aktivity, který pomáhá rozlišovat segmenty ticha od segmentů řeči.

Pravý a levý vstupní signál binaurálního modelu byl rozložen pomocí

filtrů gamatonové analýzy na určitý počet kritickým pásem. Každý signál kritického pásma byl zpracován modelem napodobujícím vláskové buňky, ten byl v tomto případě vytvořen filtrem dolní propusti. Následně byla provedena cross-korelace signálů pravého a levého kanálu v rámci každého z kritických pásem a výsledný signál byl navzorkován. Takto zpracované signály byly dále analyzovány, snímky s velkým výskytem maxim signálu byly označeny za segmenty řeči a naopak snímky s téměř nulovým signálem byly označeny za segmenty ticha. Následně na segmenty řeči bylo aplikováno adaptivní potlačení šumu. Tato navrhovaná metoda je díky binaurálnímu modelu mnohem jednodušší a také vhodnější pro aplikace v reálném čase, kterými právě naslouchátka jsou.

4.3 Deep neural network models of sound localization reveal how perception is adapted to real-world environments

V práci [10] byl vytvořen model, určený k lokalizaci zvuku za použití informací dostupných skutečnému lidskému posluchači.

Model obsahuje lidskou hlavu a trup, vnější ucho (zvukovod) a hlemýždě. Hlava, trup a vnější ucho byly simulovány pomocí HRIR, nahrané na standardním lidském modelu. Hlemýžď byl vytvořen pomocí banky filtrů (pásmová zádrž), napodobující frekvenční selektivitu lidského ucha. Výstup byl následně usměrněn a navzorkován na 4 kHz. Takto zpracované výstupy z obou uší tvoří vstup pro neuronovou síť. Model byl trénován tak, aby popsal umístění zdroje akustického signálu pomocí azimutu a elevace, přičemž horizontální rovina byla rozdělena s 5° rozestupy a vertikální rovina 10° rozestupy.

Trénovací data byla tvořena sadou stereo audio signálů, kterým byla přiřazena pozice v prostoru vzhledem k hlavě. Signály byly získány z nahrávek zvuku a přirozeného šumu, jejichž zdroj byl umístěn v prostoru a součástí signálů jsou dále odrazy od stěn místnosti. Tyto odrazy byly následně filtrované pomocí HRIR.

4.4 Functional rate-code models of the auditory brainstem for predicting lateralization and discrimination data of human binaural perception

Model navržený v práci [4] se skládá ze dvou hlavních částí, z části periferní a části binaurální. Periferní část sestává z bloků simulujících filtraci vnějšího a střední ucha, filtraci hlemýždě a filtraci vnitřních vláskových buněk. Vnější a střední ucho bylo vytvořeno pomocí FIR filtru řádu 512 a frekvenční selektivita hlemýždě byla modelována pomocí banky filtrů. Takto filtrovaný signál je dále zpracováván binaurální částí, která se skládá ze dvou modelů napodobujících chování MSO a LSO. Pro pravou a levou periferii mozku jsou vytvořený samostatné modely MSO a LSO, ale každý z nich získává vstupní

informace z obou částí. Výstupem těchto dvou modelů jsou tzv. lateralizační funkce.

Tento model byl autorkou práce [2] ještě doplněn o zpracování vstupních signálů pomocí funkcí HRIR a o algoritmy strojového učení, o umělou neuronovou síť a KNN klasifikátor.

Kapitola 5

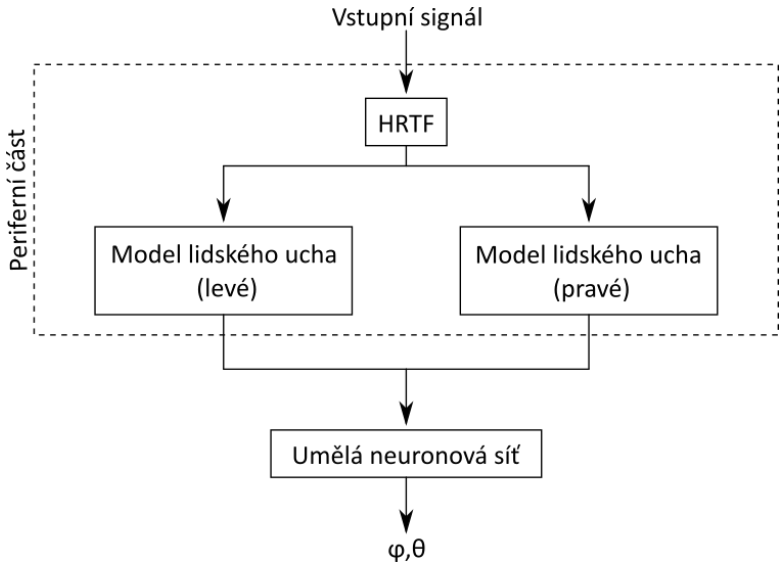
Implementace modelů

V této práci byly navrženy dva modely napodobující chování lidského sluchu, určené k lokalizace zdroje zvuku v prostoru. Návrh a implementace modelů se primárně zaměřovaly na vliv přítomnosti binaurálního modelu lidského slyšení v celém algoritmu lokalizace.

Na základě dostupné literatury popisující dnes používané metody lokalizace byla k výslednému vyhodnocení zvolena neuronová síť. Implementace navržených modelů byla provedena v počítačovém prostředí Matlab.

První implementovaný model byl inspirován článkem [10]. Signál přivedený na vstup modelu je nejprve zpracován periferní částí, která simuluje filtraci lidské hlavy a vnějšího ucha (HRTF) a ucha vnitřního. V této části je jednoduchý zvukový signál rozdělen na dva kanály, které jsou zmíněným způsobem filtrovány každý zvlášť v odpovídajícím modelu lidského ucha, viz obr. 5.1.

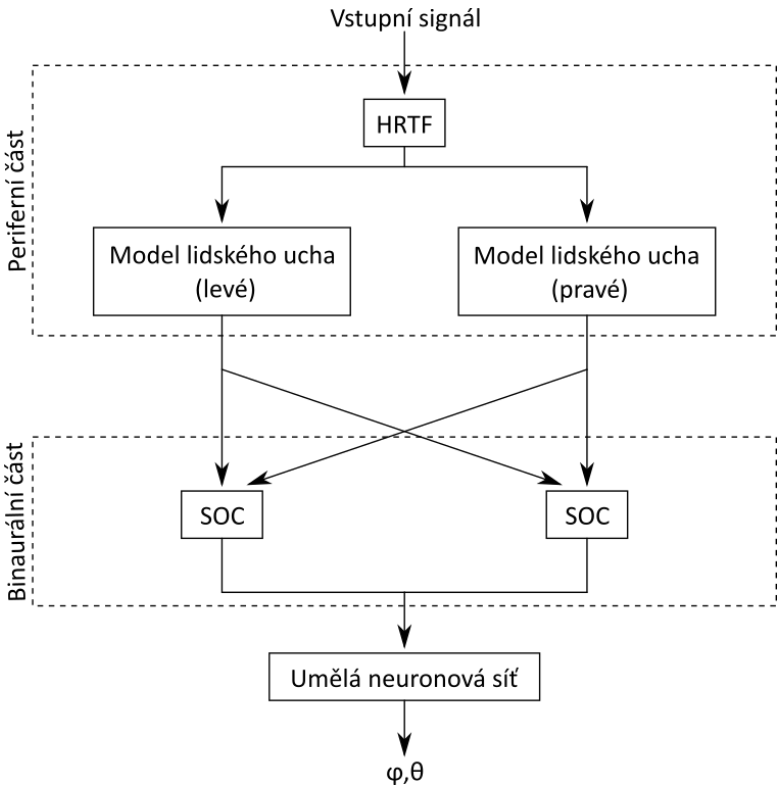
Takto upravené signály z obou kanálu jsou sloučeny a společně zpracovány neuronovou sítí. Výstupem neuronové sítě jsou zanalyzované parametry předzpracovaného zvuku, úhly azimutu a elevace odpovídající přiřazené poloze zvuku, podrobněji v 5.3.



Obrázek 5.1: Model bez binaurální části

Druhý implementovaný model, který byl navržen na základě práce [4], viz obr. 5.2, zahrnuje, na rozdíl od předešlého modelu, binaurální část, vloženou mezi periferní část a neuronovou síť. Periferní část, stejně jako v prvním modelu napodobuje filtraci lidské hlavy, vnějšího a vnitřního ucha.

Na rozdíl od předešlého modelu, oba vytvořené a poté filtrované signály nejsou po výstupu z periferní části sloučeny, jsou každý zvlášť přivedeny do zmíněné binaurální části. Signál z pravého i levého kanálu jsou přivedeny na dva modely simulující komplex horní olivy, viz obr. 5.2. Výstupem binaurální části modelu jsou tzv. lateralizační funkce, které obsahují informaci o lokalizaci zdroje zvuku. Teprve takto zpracované signály jsou sloučeny a analyzovány dohromady v umělé neuronové síti, která je implementována stejným způsobem jako v prvním modelu.



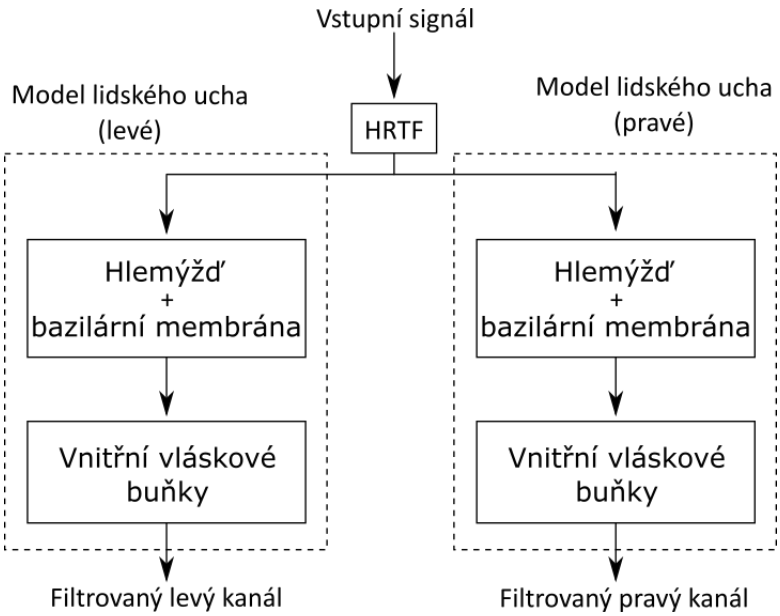
Obrázek 5.2: Model s binaurální částí

5.1 Periferní část

Na vstup celého modelu byl přiveden jednokanálový vstup. Prvním krokem, před samostatným zpracováním modelem lidských uší, bylo vytvořit prostorový zvuk, který bude možné později lokalizovat. Toho bylo dosaženo konvolucí vstupního signálu s funkcí HRIR. Sada funkcí HRIR byla převzata z databáze The HUTUBS head-related transfer function (HRTF) database [11]. Takto byl vstupní signál rozdělen na dva kanály, na signál pro pravé ucho a na signál pro levé ucho.

Každý z kanálů byl přiveden na vstup příslušného modelu lidského ucha, viz obr. 5.3. Oba signály byly zpracovány ve dvou blocích filtrů. Selektivita hlemýždě a nelinearita bazilární membrány byly simulovány pomocí banky filtrů DRNL (Dual Resonance Non-Linear filterbank) [12]. Přenos signálu pomocí vnitřních vláskových buněk je napodoben prostřednictvím filtru dolní propusti.

Na vstupu celého modelu byl vygenerován širokopásmový šum o 500 vzorcích. Ten byl následně pomocí HRIR rozdělen na dva vektory stejné délky, které dále byly zmíněným způsobem filtrovány. Na výstupu celé periferní části byly připravené dva vektory, levý kanál a pravý kanál, každý o délce 20385 prvků. V případě modelu s binaurální částí byly takto zpracované vektory poslány na její vstup a v případě modelu bez binaurální části byly následujícím způsobem upraveny. Tyto dva signály, byly nejprve sloučeny do jednoho vektoru o délce 40770 prvků, to znamená, že byly poskládány za sebe. Tento výsledný vektor byl poslán ke zpracování na vstup neuronové sítě.



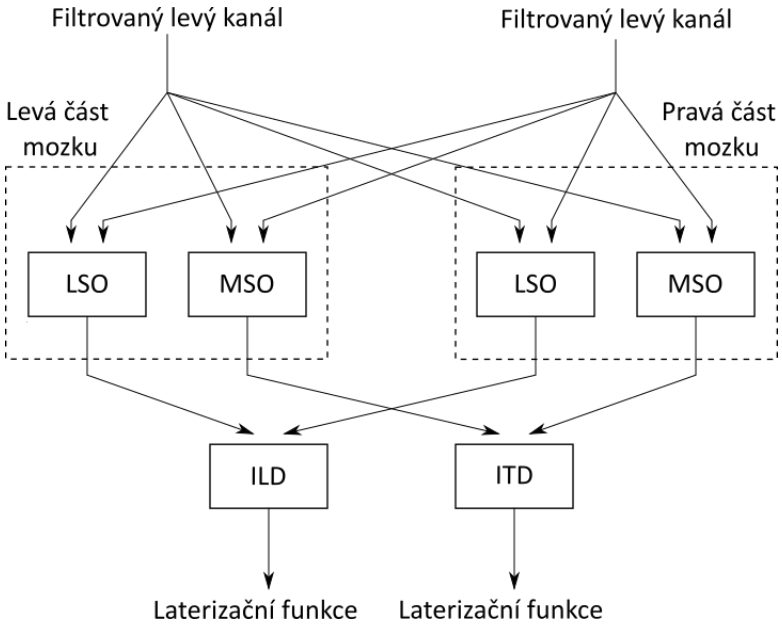
Obrázek 5.3: Periferní část

5.2 Binaurální část

Implementovaný binaurální model je převzat z práce [1].

Dvoukanálový signál zpracovaný periferní částí je přiveden na vstup binaurální části. Tato část se skládá ze dvou modelů, LSO a MSO, napodobujících chování komplexu horní olivy, viz 2.1. Oba modely jsou implementované samostatně pro každou polovinu mozku a každý z modelů přijímá informaci jak z levého kanálu, tak z pravého kanálu. Signály filtrované modelem LSO jsou, viz obr. 5.4, dále zpracované v centru mozku zodpovědném za zpracování parametrů ILD a signály filtrované modelem MSO jsou dále zpracované v centru, které vyhodnocuje ILD.

Na vstup binaurálního modelu byly přivedeny z periferní části dva vektory, každý z nich o délce 20385 prvků, odpovídající levému a pravému kanálu. Zpracování binaurálním modelem nemělo žádný vliv na rozměry obou signálů. Výstupem binaurálního modelu jsou vektory tzv. lateralizačních funkcí. Ty jsou před vstupem do neuronové sítě sloučeny do jednoho vektoru o délce 40770 vzorků, který odpovídá za sebe umístěným vektorům levého a pravého kanálu.



Obrázek 5.4: Binaurální část

5.3 Umělá neuronová síť

Ke zpracování filtrovaných vstupních dat byla v prostředí Matlab zvolena čtyřvrstvá umělá neuronová síť feedforwardnet. Jedná se o dopřednou síť, která má vstup přiveden na první vrstvu, následují dvě skryté vrstvy a poslední čtvrtá vrstva vytváří výstup sítě.

V každé vrstvě jsou data zpracovávána určitým počtem umělých neuronů, který je zvolen na základě velikosti vstupního datasetu. E. Koshkina ve své práci [2] uvádí pravidlo, tzv. Geometric Pyramid Rule, podle kterého se dá určit vhodný počet neuronů ve skrytých vrstvách. Pro použitou čtyřvrstvou neuronovou síť se určuje počet neuronů v první skryté vrstvě H_1 a ve druhé skryté vrstvě H_2 podle vztahů

$$H_1 = M \cdot \left(\frac{N}{M}\right)^{\frac{2}{3}}, \tag{5.1}$$

$$H_2 = M \cdot \left(\frac{N}{M}\right)^{\frac{1}{3}}, \tag{5.2}$$

kde N je počet neuronů ve vstupní vrstvě a M je počet neuronů ve výstupní vrstvě.

Počet neuronů ve vstupní vrstvě je přímo diktován vstupními daty. Vstupní data jsou ve formátu vektoru obsahujícího postupně všechny kanály získané modelem lidského slyšení. Velikost vstupní vrstvy trénované sítě je 40770 neuronů.

Výstupní vrstva je obdobně diktována formátem výstupu. Pro testovaný model to znamená vektor pravděpodobnosti směru zvuku z diskrétního počtu azimutů a elevací. Počet neuronů výstupní vrstvy trénované sítě je 17.

Na základě vztahů 5.1, 5.2 a velikosti vstupní a výstupní vrstvy byly použity skryté vrstvy o zvolených rozměrech $H_1 = 3050$ neuronů a $H_2 = 227$ neuronů.

5.4 Vytváření vstupního datasetu

Pro vytvoření prostorového signálu bylo nutné stanovit si rozsah zkoumaného prostoru a rozčlenit si ho na diskrétní hodnoty úhlů, pomocí kterých se dají popsat konkrétní body v prostoru. Byla zvolena celá horizontální rovina prostoru (360°) a rozdělena na azimutální úhly s krokem 30°. Vertikální rovina byla zvolena v rozsahu -60°až 60°, kdy 0° odpovídá vodorovné rovině ve výšce spojnice obou uší. Úhly elevace byly ve vertikální rovině rozmístěny rovněž s krokem 30°.

Aby došlo k pokrytí celého takto zvoleného prostoru, ve vytvořené síti úhlů, byly v každém úseku vygenerovány dva směry dané náhodně vygenerovanou dvojicí azimutu a elevace, spadající do příslušného úseku mezi dvěma diskrétními hodnotami. Konkrétní poloha těchto meziúhlů byla dána pravděpodobností, se kterou se vyskytují mezi dvěma nejbližšími hodnotami. To znamená, že každému meziúhlu azimutu a elevace byla přiřazena hodnota mezi 0 a 1, poměrově odpovídající vzdálenosti od příslušných krajních úhlů úseku. Na příklad při náhodném vygenerování úhlu 15°, který spadá doprostřed úseku mezi hodnotami 0° a 30°, je popsán pravděpodobností 0.5. Pro každou zvolenou dvojici meziúhlů byl vygenerován vstupní signál.

5.5 Trénování a testování sítě

Pro oba modely byla vygenerována stejná neuronová síť. Pro oba případy byl stejným způsobem vygenerován trénovací dataset s rozdílem v použití binaurálního modelu. Pro trénování pak v obou případech byla použita vygenerovaná neuronová síť bez pokusu o přetrénování na novém datasetu nebo z jednoho modelu na druhý. Výsledkem byly dvě neuronové sítě natrénované na příslušná vstupní data.

Po natrénování neuronové sítě byl dvacetkrát vygenerován dataset stejného typu, jako ten, na kterém byla síť trénována, s jedním meziúhlem. Tento testovací dataset byl nově vytvořený, tedy během trénování sítě na něm nebylo provedeno trénování, validace, ani testování. Dále byly na modelech testovány reálné zašuměné signály. Nahrávky zvuků byly získány z databáze NOIZEUS [13]. Za účelem testování vytvořených modelů bylo použito 20 zvukových stop ve formátu .wav, jejichž obsahem jsou řečové signály, na které byly aplikovány šumové signály stejné délky se SNR 0 dB, 5 dB, 10 dB a 15 dB.

Kromě testování vstupních dat, zpracováním přímo odpovídajících zvolenému modelu, byla každá z obou neuronových sítích otestována na datech určených pro trénování druhé sítě. Tedy nejen, že síť, natrénována na datech zpracovaných modelem bez binaurální části, byla testována na množině dat zpracovaných modelem bez binaurální části, ale byla rovněž testována na množině zpracované modelem s binaurální částí. Stejný způsob testování byl aplikován na neuronovou síť natrénovanou na datech zpracovaných modelem s binaurální částí. Její úspěšnost byla ověřena jak na množině dat zpracovaných modelem s binaurální částí, tak na množině dat zpracovaných modelem bez binaurální části.

V rámci tohoto testování byla spočítána úspěšnost natrénovaných sítí na zmíněných množinách dat a následně vykreslena prostřednictvím box plot grafů, viz kapitola 6. Výsledná úspěšnost natrénovaných modelů je spočítána pomocí hodnoty MSE (Mean Squared Error). Nejprve je stanovena chyba odhadu tak, že je zjištěn rozdíl mezi hodnotami vektoru očekávaného výstupu neuronové sítě a získaného výstupního vektoru. Na hodnoty takto získaného chybového vektoru je následně aplikovaná druhá mocnina a výsledná hodnota MSE se získá aritmetickým průměrem těchto hodnot.

Současně byla určena přibližná hodnota chybovosti lokalizace zvuku člověkem. Při výpočtu této hodnoty MSE se postupovalo stejným způsobem jako při výpočtu MSE neuronové sítě. Chybový vektor byl získán rozdílem očekávaného výstupu a přibližného výstupu, který by určil lidský sluch. Do tohoto přibližného výstupu byla započítána rozlišovací schopnost člověka 5° pro úhly azimutu a 10° pro úhly elevace. Získaná hodnota MSE pro lidský sluch je 0.0146.

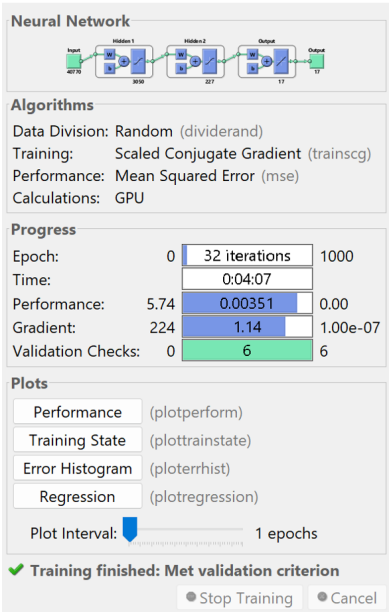
Kapitola 6

Výsledky trénování a testování

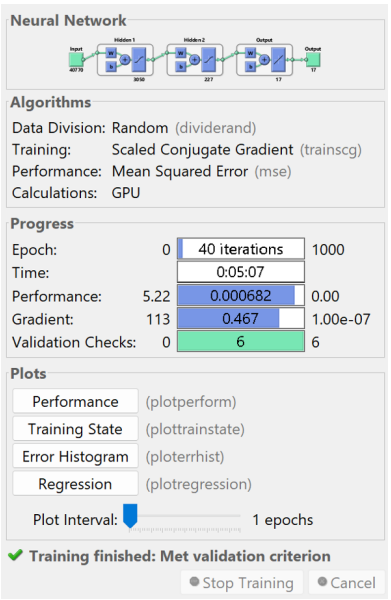
V této kapitole jsou popsány výsledky procesu trénování a následmě výsledky testů již natrénovaných modelů.

Obrázek 6.1 zobrazuje informace o dokončeném trénování obou neuronových sítí. V první části okna je znázorněna architektura implementované ANN, tedy vstupní a výstupní vrstva, jedna skrytá vrstva a počet neuronů nacházejících se v jednotlivých vrstvách. V sekci Progress je vypsán celkový počet iterací, které proběhly během trénování, doba, po kterou se neuronová síť trénovala a úspěšnost natrénování sítě (Performance). Pomocí tlačítek v sekci označené Plots je možné vykreslit případné grafy zobrazující vlastnosti sítě a jejího trénování.

a.



b.

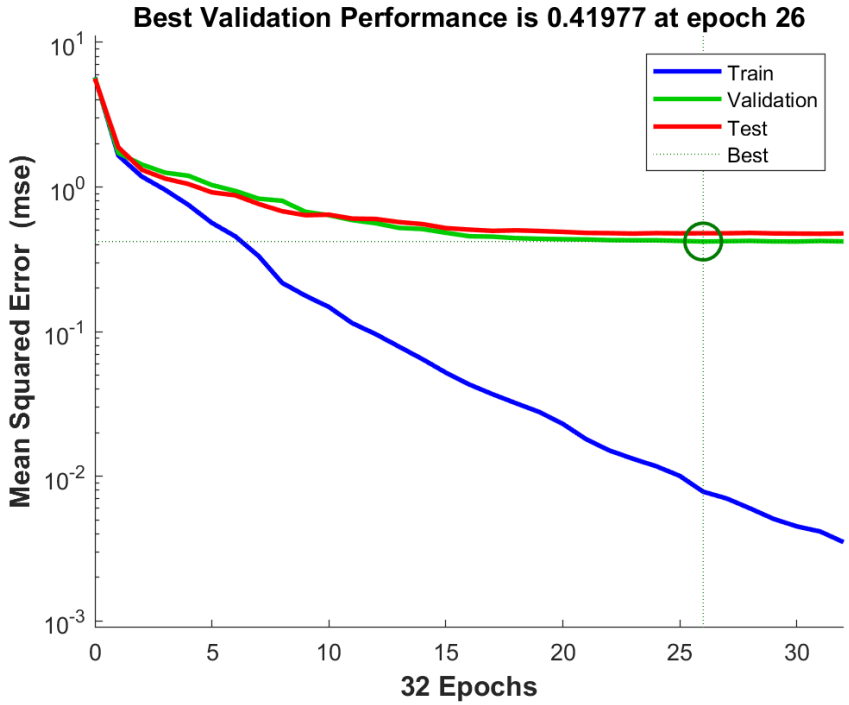


Obrázek 6.1: Trénování neuronové sítě bez binaurální části a. modelu s binaurální částí b.

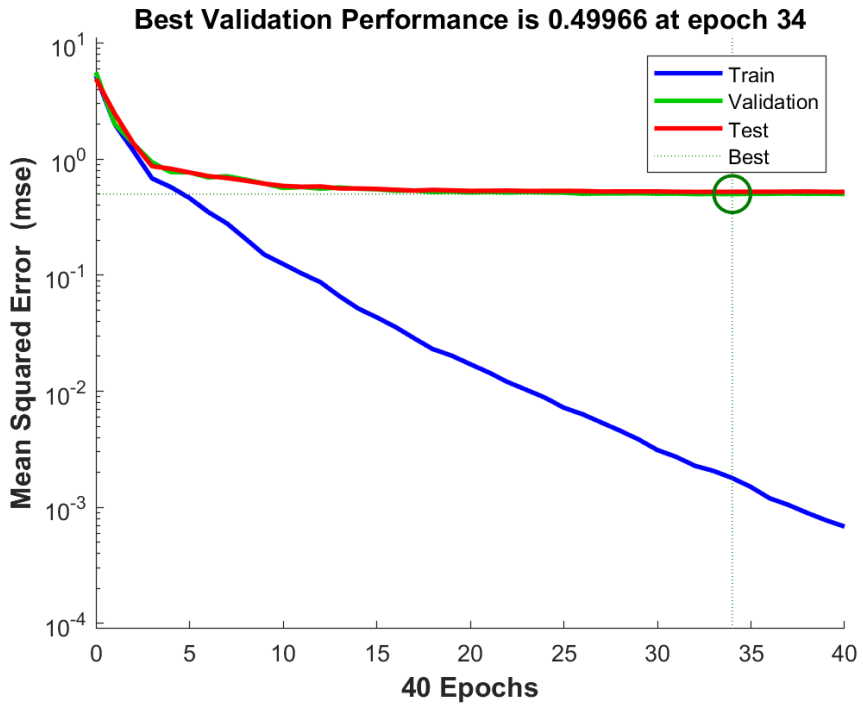
Performance trénování ANN implementované v modelu bez binaurální části je vykreslena na obrázku 6.2. Uvedený počet iterací je, na rozdíl od počtu iterací v trénovacím okně, hodnota, při které dosáhlo trénování neuronové sítě nejlepší hodnoty MSE. Na obrázku 6.3 je stejným způsobem znázorněna performance modelu s binaurální částí.

Na obrázcích grafů performance je vidět porovnání počtu iterací, které skutečně proběhly a hodnotu iterace, při které dosáhl model sítě nejlepšího výsledku MSE. Je tomu tak proto, že po dosažení nejlepšího výsledku model ověřuje, zda se tato hodnota bude dále zlepšovat či dojde k jejímu ustálení. V těchto případech dosáhlz modelz optimální hodnoty chybovosti o 6 iterací dřve, než bylo celkové trénování ukončeno.

V případě modelu bez binaurální části dosáhl algoritmus nejnižší hodnoty MSE 0.41977 již po 26 iteracích. Což je v porovnání s modelem obsahujícím binaurální část, který opritámlní hodnoty 0.49966 dosáhl po 34 iteracích, dříve.



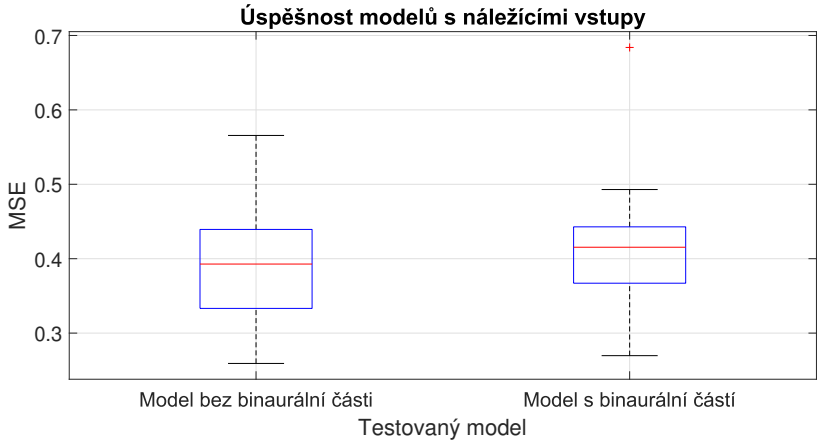
Obrázek 6.2: Ověření úspěšnosti trénování modelu bez binaurální části



Obrázek 6.3: Ověření úspěšnosti trénování modelu s binaurální částí

6.1 Testování širokopásmového šumu

Na obrázku 6.4 je vykreslena dosažená přesnost lokalizace oběma implementovanými modely. Jedná se o případ, kdy byla neuronová síť, implementovaná v modelu bez binaurální části, testována množinou dat zpracovaných modelem bez binaurální části. Rovněž neuronová síť, implementovaná v modelu s binaurální částí, byla testovaná jí odpovídající množině dat. Z grafu je patrné, že model bez implementované binaurální části má nepatrně lepší úspěšnost lokalizace a hodnota chybovosti má, na rozdíl od druhého modelu, větší rozptyl. V porovnání s hodnotou MSE lidského slyšení jsou hodnoty těchto modelů o jeden řád vyšší.



Obrázek 6.4: Úspěšnost obou implementovaných modelů testovaných na přímých vstupních datech

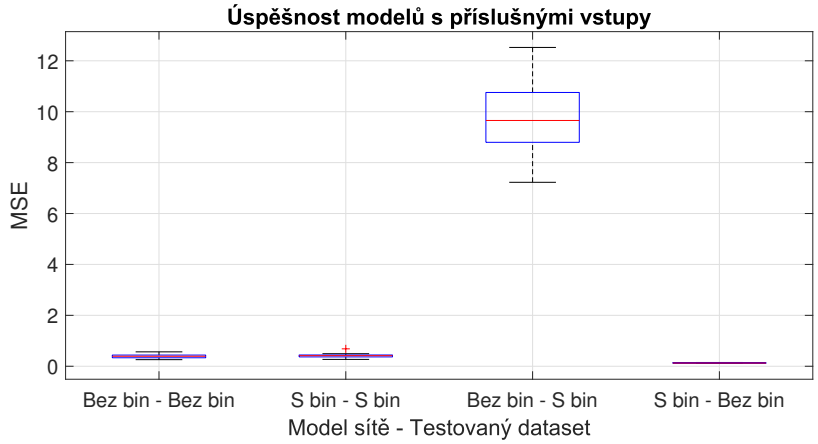
Testování úspěšnosti lokalizace pro oba implementované modely, společně s oběma variantami testovacích dat, je vykresleno na obrázku 6.5.

První dva grafy se týkají přesnosti neuronových sítí testovaných na množině dat odpovídající trénovací množině. Boxplot označený „Bez bin – Bez bin" zobrazuje testování ANN, která byla trénována i testována na datech zpracovaných modelem bez implementované binaurální části. Druhý boxplot, označený „S bin – S bin" vykresluje přesnost ANN, trénované i testované na datech, která byla předem zpracována binaurální částí modelu.

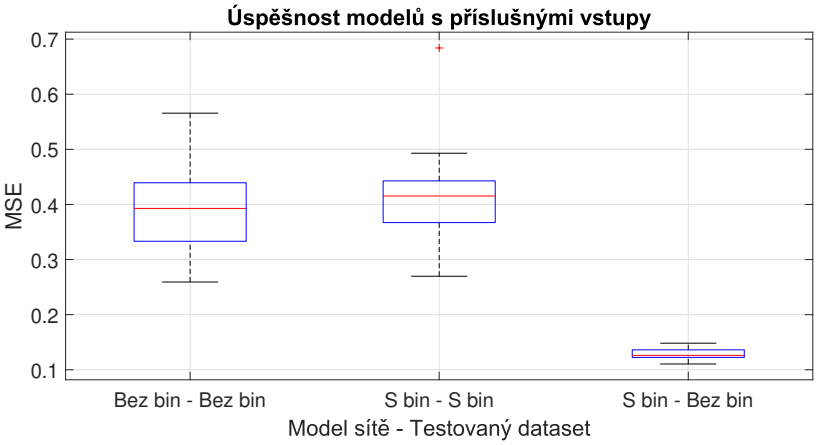
Druhé dva grafy zobrazují přesnost lokalizace těchto neuronových sítí, které byly testovány na prohozených datech. Boxplot označený „Bez bin – S bin" odpovídá úspěšnosti testování ANN implementované v modelu bez binaurální části, avšak testované pomocí datasetu, který byl zpracován binaurálním modelem. Poslední boxplot „S bin – Bez bin" zobrazuje přesnost ANN, která byla trénována na datech zpracovaných binaurální částí modelu a testována pomocí množiny dat, zpracované modelem bez implementované binaurální části.

Z grafů je patrné, že model, který nemá implementovanou binaurální část a byl testovaný na datech, které byly zpracované binaurální částí modelu, má výrazně vyšší hodnotu chybovosti, to zhruba desetkrát a má větší rozptyl.

Pro lepší viditelnost ostatních hodnot je první, druhý a čtvrtý boxplot z grafu 6.5 znovu vykreslen v obrázku 6.6. Zde je vidět, že oba modely, které byly testované na stejných datech, na kterých byly také trénovány, mají téměř stejnou hodnotu MSE i rozptyl. V porovnání s ostatními boxploty má zcela jednoznačně nejnižší hodnotu chybovosti model, jehož součástí je implementovaný binaurální model, ale byla na něm testována data, která předem nebyla zpracována binaurální částí.



Obrázek 6.5: Úspěšnost obou implementovaných modelů na přímých i prohozených vstupních datech

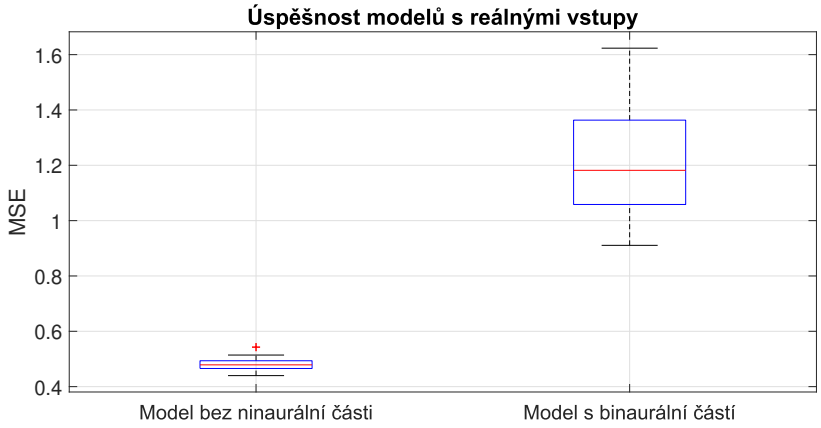


Obrázek 6.6: Úspěšnost obou implementovaných modelů na přímých i prohozených vstupních datech 2

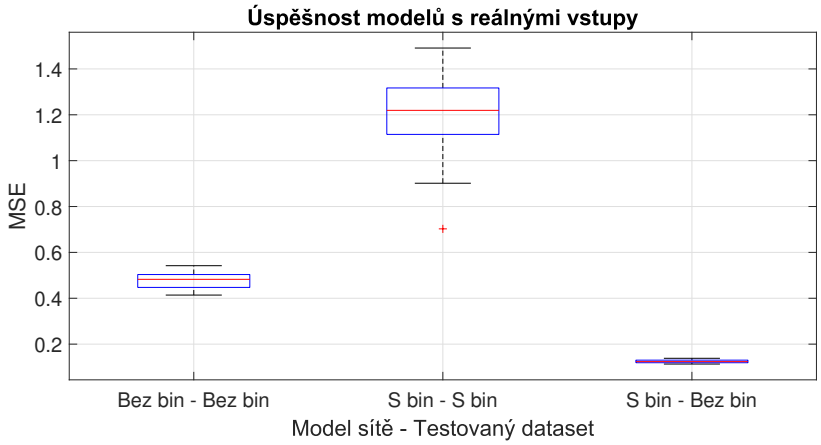
6.2 Testování reálných signálů

Na obrázku 6.7 je vykreslena dosažená přesnost lokalizace oběma implementovanými modely, které byly testovány pomocí nahraných zvukových signálů. Na modelu bez binaurální části byla testovaná množina dat zpracovaných modelem bez binaurální části a na modelu s binaurální částí, byla testovaná množina, která byla zpracovaná binaurální částí. Z grafu je patrné, že model bez implementované binaurální části má výrazně lepší úspěšnost lokalizace a hodnota chybovosti má, na rozdíl od druhého modelu, menší rozptyl.

Úspěšnost testování zvukových nahrávek na obou modelech přímo i křížem lze vidět na obrázku 6.8. Případ, kdy byl testován model bez binaurální části pomocí dat zpracovnaých binaurální částí měl při testování širokopásmovým šuměm velmi vysokou chybovosti, a proto byl při testování reálných signálů vynechán. Z grafu je patrné, že nejvyšší hodnotu MSE má neuronová síť implementovaná v modelu s binaurální částí a testována na datech zpracovaných binaurální částí. Tato hodnota je o dva řády vyšší než hodnota MSE lidského sluchu.



Obrázek 6.7: Úspěšnost obou implementovaných modelů na vstupech skutečných nahrávek zvuku



Obrázek 6.8: Úspěšnost obou implementovaných modelů na přímých i prohozených vstupech skutečných nahrávek zvuku

Kapitola 7

Závěr

Teoretické poznatky a metody popsané v první části práce byly použity při návrhu umělé neuronové sítě, která byla natrénována na dvou trénovacích množinách. Jedna z těchto množin byla předem zpracována modelem zahrnujícím binaurální část a druhá množina modelem, který neměl implementovanou binaurální část.

Po natrénování sítí se zdálo, že využití binaurálního modelu zvyšuje náročnost trénování bez pozitivního vlivu na přesnost lokalizace. Tento výsledek byl demonstrován v obrázcích 6.1, ze kterých je možné vyčíst, že počet iterací potřebných pro natrénování neuronové sítě je při použití binaurálního modelu o zhruba 23.5 procent vyšší. Z obrázků 6.3 a 6.2 je patrné, že při potřebném počtu iterací trénovaní pro každý model je výsledná přesnost odhadu nepatrně vyšší pro model bez implementované binaurální části.

Pro otestování potenciálního jiného významu binaurálního modelu byla na několika menších data setech zjištěna úspěšnost trénování sítě pro případ, kdy je lokalizace signálu, zpracovaného jedním modelem, provedena pomocí sítě natrénované na druhém modelu.

Očekávaným výsledkem bylo, že zkombinování vstupů zpracovaných binaurální částí a sítě trénované na vstupech zpracovaných modelem bez binaurální části, má výrazně nižší přesnost než při použití vstupních dat, na kterých byla příslušná neuronová síť natrénována. Oproti tomu zkombinování vstupů zpracovaných modelem bez binaurální části se sítí natrénované na datech zpracovaných binaurálním modelem překvapivě vykázalo výrazně vyšší přesnost odhadu lokalizace než nejen opačná kombinace, ale i testování na přímých datech. To naznačuje, že význam binaurálního modelu by mohl být, spíš než v praktickém použití při lokalizace zdroje zvuku, v použití pro trénování modelů strojového učení pro účel lokalizace.

Testování obou modelů pomocí signálů reálných zvukových nahrávek vykazuje podobné výsledky jako testování pomocí širokopásmového šumu, na kterém byly modely natrénovány. Je však patrné, že model s implementovanou binaurální částí je citlivější na výrazně odlišná testovací data od dat trénovacích. V případě testování dat zpracovaných modelem s binaurální částí i bez binaurální části se výrazně zvýšila hodnota chybovosti lokalizace. U modelu bez implementované binaurální části došlo jen k mírnému zhoršení lokalizace.

Je však důležité zmínit, že v rámci této práce byla umělá neuronová síť trénována na relativně malé množině vstupních dat. Se zlepšením přesnosti natrénování neuronové sítě za účelem lepší lokalizace zdroje zvuku v prostoru by mohlo pomoci trénování sítě na větším vstupním data setu. V moderním strojovém učení se používají umělé neuronové sítě, které mají, co se počtu skrytých vrstev a počtu použitých neuronů ve vrstvách týče, podstatně větší rozměry a komplikovanější struktury než neuronová síť použitá v rámci této práce. Také by zajisté mělo význam zjistit, jak se účinnost trénování změní v případě, kdy je neuronová síť trénována na data setu zpracovaném binaurálním modelem a následně dotrénována ještě na menším data setu, zpracovaném modelem bez binaurální části. Tímto způsobem trénovaná síť by byla implementována v modelu, který nezahrnuje binaurální část. Takovýto způsob trénování sítí se v praxi běžně používá a nazývá se "fine-tuning".

Cílem této práce však nebylo získání modelu s co nejvyšší přesností lokalizace, ale modelu, který se snaží co nejvíce přiblížit přesnosti lokalizace člověka. Při porovnání chybovosti lokalizace obou modelů s chybovostí lidského sluchu, vyjádřené pomocí hodnoty MSE, je patrné, že nejblíže k požadované hodnotě MSE, rovné 0.0146, se nachází model s binaurální částí, který zpracovává data filtrovaná pouze periferní částí, o hodnotě MSE 0.1246. Naopak největší rozdíl vykazuje model bez implementované binaurální částí, který zpracovává signály filtrované binaurální částí, o hodnotě MSE 9.5657.

Tato práce ukázala, že i v době strojového učení a neuronových sítí má s největší pravděpodobností používání binaurálního modelu v úlohách lokalizace zdroje zvuku v prostoru význam.

Bibliografie

[1] J. Bouse, V. Vencovský, F. Rund a P. Marsalek, "Functional Rate-Code Models of the Auditory Brainstem for Predicting Lateralization and Discrimination Data of Human Binaural Perception", *The Journal of the Acoustical Society of America*, 2019.

[2] E. Koshkina, "Využití strojového učení pro modelování binaurálního slyšení", Diplomová práce, FEL, 2017.

[3] R. Havlík, "Vliv binaurálního slyšení na srozumitelnost řeči při použití kompetitivního šumového signálu", Disertační práce, LÉKAŘSKÁ FAKULTA MASARYKOVY UNIVERZITY, 2010.

[4] J. Bouše, "Models and experiments of binaural interactions", Doctoral Thesis, FEL, 2020.

[5] O. Novotný, "Psychoakustická měření binaurálních vlastností lidského sluchu", Diplomová práce, VUT, 2010.

[6] V. R. Algazi a R. O. Duda, "Approximating the head-related transfer function using simple geometric models of the head and torso", *The Journal of the Acoustical Society of America*, 2002.

[7] G. K. Jha, *Artificial Neural Networks and its applications - apps.iasri.res.in*, 2007. URL: `http://apps.iasri.res.in/ebook/EBADAT/5-Modeling%20and%20Forecasting%20Techniques%20in%20Agriculture/5-ANN_GKJHA_2007.pdf`.

[8] C. Verron, P.-A. Gauthier, J. Langlois a C. Guastavino, "Binaural analysis/synthesis of interior aircraft sounds", in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, s. 177–180. DOI: `10.1109/ASPAA.2011.6082313`.

[9] H. Huang a C. Kyriakakis, "Binaural Model Based Adaptive Binaural Noise Reduction", in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, 2006, s. 1110–1113. DOI: `10.1109/ACSSC.2006.354926`.

[10] A. Francl a J. H. McDermott, "Deep neural network models of sound localization reveal how perception is adapted to real-world environments", *The Journal of the Acoustical Society of America*, 2020.

[11] B. Fabian, D. Manoj, P. Robert et al., *The HUTUBS head-related transfer function (HRTF) database*, 2019. DOI: 10.14279/depositonce-8487. URL: http://dx.doi.org/10.14279/depositonce-8487.

[12] E. Lopez, *A physical model of sound diffraction and reflections in the human Concha*, pros. 2015. URL: https://www.academia.edu/10114556/A_physical_model_of_sound_diffraction_and_reflections_in_the_human_concha.

[13] Y. Hu a P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms", *Speech Communication*, 2007.

Příloha A

Implementace

Přílohou je soubor binauralni_modely.zip obsahující:

- inicializace.m

  startování potřebných toolboxů

- generateNNinput.m

  filtrování signálů a úprava formátu vstupu do ANN

- NNcreateandtrain_bin.m

  generování datasetu, vytvoření a trénování neuronové sítě modelu s binaurální částí

- NNcreateandtrain_nobin.m

  generování datasetu, vytvoření a trénování neuronové sítě modelu bez binaurální části

- test.m

  testování modelů

- mse_human.m

  výpočet hodnoty mse lidského sluchu

- readme.txt

  soupis potřebných toolboxů a souborů