ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta elektrotechnická
Katedra radioelektroniky

Využití strojového učení pro modelování binaurálního slyšení

Utilization of Machine Learning in Binaural Hearing Model

Diplomová práce

Studijní program:	Komunikace, Multimédia a Elektronika
Studijní obor:	Multimediální technika
Vedoucí práce:	Ing. František Rund, Ph.D., Ing. Jaroslav Bouše

Ekaterina Koshkina

Praha 2017

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

ZADÁNÍ DIPLOMOVÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Koshkina** Jméno: **Ekaterina** Osobní číslo: **412013**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra radioelektroniky**

Studijní program: **Komunikace, multimédia a elektronika**

Studijní obor: **Multimediální technika**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Využití strojového učení pro modelování binaurálního slyšení

Název diplomové práce anglicky:

Utilization of Machine Learning in Binaural Hearing Model

Pokyny pro vypracování:

Seznamte se s problematikou binaurálního slyšení se zaměřením na určování směru příchodu zvuku. Dále se seznamte s problematikou strojového učení. Navrhněte rozšíření existujícího modelu binaurálního slyšení [1] o část umožňující lokalizaci směru vnímaného zvuku na bázi strojového učení. Zaměřte se na lokalizaci v horizontální rovině, zvažte možnost lokalizace ve vertikální rovině. Implementujte navržené rozšíření v prostředí Matlab a výsledky porovnejte se subjektivními daty.

Seznam doporučené literatury:

[1] Bouše, J. Model of binaural interactions. 2015. Doktorandské minimum. Praha: ČVUT FEL, Katedra radioelektroniky.
[2] Hastie T., Tibshirani R., Friedman J., The Elements of Statistical Learning, Springer-Verlag New York, Second Edition, eBook ISBN 978-0-387-84858-7, 2009

Jméno a pracoviště vedoucí(ho) diplomové práce:

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

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

Ing. Jaroslav Bouše, katedra radioelektroniky FEL

Datum zadání diplomové práce: **13.02.2017** Termín odevzdání diplomové práce: **26.05.2017**

Platnost zadání diplomové práce: **31.08.2018**

Podpis vedoucí(ho) práce	Podpis vedoucí(ho) ústavu/katedry	Podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

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

.	
Datum převzetí zadání	Podpis studentky

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.

Dne 26. května 2017 v Praze

……..

Poděkování

Děkuji svým vedoucím diplomové práce Ing. Františku Rundovi, Ph.D. a Ing. Jaroslavu Boušemu za věcné připomínky, cenné rady, trpělivost a vstřícnost při konzultacích během zpracovávání této diplomové práce.

Tato diplomová práce vznikla za podpory Grantové agentury ČVUT v Praze, čísla grantů SGS14/204/OHK3/3T/13 a SGS17/190/OHK3/3T/13.

Abstrakt

Tato práce se zabývá problematikou binaurálního slyšení se zaměřením na lokalizaci zdroje zvuku pomocí strojového učení. Obsahem práce jsou čtyři experimenty s algoritmy detekce úhlu příchozího zvuku. V prvním experimentu je úhel statického zdroje zvuku v přední horizontální polorovině detekován pomocí klasifikátoru *k*-nejbližších sousedů s využitím binaurálních modelů (LSO a MSO). Ve druhém experimentu se určuje poloha statického a pohyblivého zdroje zvuku v přední horizontální polorovině pomocí klasifikátoru *k*-nejbližších sousedů a umělé neuronové sítě. Výsledky tohoto experimentu jsou na rozdíl od prvního experimentu získány pro spojení výstupů binaurálních modelů a jsou porovnané se subjektivními daty. Další částí práce je aplikace navrženého algoritmu pro klasifikaci úhlu v celé horizontální rovině. Posledním experimentem je lokalizace zvuku ve vertikální rovině. Úspěšnost algoritmů je ověřená na databázi zvukových signálů. Výstupem diplomové práce je systém, který je schopný lokalizovat zdroj zvuku v prostoru z výstupů binaurálních modelů. Systém dosahuje relativně vysoké úspěšnosti. V případě využití klasifikátoru *k*-nejbližších sousedů se průměrná úspěšnost pohybuje okolo 60 % a v případě umělé neuronové sítě okolo 90 %.

Klíčová slova: binaurální model slyšení, LSO, MSO, azimut, elevace, lokalizace, statický zdroj zvuku, pohyblivý zdroj zvuku, extrakce příznaků, RMS, klasifikace, *k*-NN, ANN, MATLAB

Abstract

This diploma thesis deals with the problem of binaural hearing with focus on the sound source localization utilizing machine learning. This work consists of four experimental algorithms of incoming sound detection. In the first experiment, the angle of the static sound source on the frontal horizontal plane is detected by the *k*-Nearest Neighbors classifier and binaural models (LSO and MSO). In the second experiment, the position of static and dynamic sound sources on the frontal horizontal plane is determined by the *k*-Nearest Neighbors classifier and Artificial Neural Network. In comparison to the first experiment, the results of the second experiment are obtained by combining the binaural models' outputs. These results are compared with the subjective data. The next part of this thesis applies the proposed algorithm to the angle localization on the entire horizontal plane. The last experiment addresses the problem of the sound localization on the vertical plane. The success rate of the algorithms is verified on the database of audio signals. The outcome of this diploma thesis is a system which can locate the sound source in a space from the outputs of the binaural models. The system achieves a high success rate. The average success rate is around 60 % for the *k*-Nearest Neighbors classifier and about 90 % for Artificial Neural Network.

Key words: binaural auditory model, LSO, MSO, azimuth, elevation, localization, static sound source, dynamic sound source, feature extraction, RMS, classification, *k*-NN, ANN, MATLAB

Obsah

Seznam obrázků a tabulek

Seznam obrázků

Seznam tabulek

Seznam použitých zkratek

ANN Artificial Neural Network
BPG Back-Propagation of Gradient
CNS Central Nervous System
CV Cross-Validation
DCN Dorsal Cochlear Nucleus
DFT Discrete Fourier Transform
DTF Directional Transfer Function
FIR Finite Impulse Response
HRIR Head-Related Inpulse Response
HRTF Head-Related Transfer Function
ILD Interaural Level Difference
ITD Interaural Time Difference
***k*-NN** *k*-Nearest Neighbors
LE Localization Error
LSO Lateral Superior Olive
MAA Minimal Audible Angle
MSO Medial Superior Olive
NN Nearest Neighbor
PHFB Patterson-Holdsworth Filter Bank
PRTF Pinna Related Transfer Function
RMS Root-Mean Square
SNN Spiking Neural Network
SOC Superior Olivary Complex

Úvod

Lokalizace zvuku je významnou vlastností lidského sluchu (*kapitola 1*) a jedná se o proces určování polohy zvukového zdroje v prostoru. Pozici lokalizovaného zvuku je možné popsat v polárních souřadnicích pomocí tří proměnných: úhlu v horizontální rovině (azimutu), úhlu ve vertikální rovině (elevace) a vzdálenosti. Zdroj zvuku lze lokalizovat za pomoci strojového učení. Strojovým učením se podrobněji zabývá *kapitola 2*.

Současným stavem na poli algoritmů pro lokalizaci zdroje zvuku na bázi strojového učení se zabývá *kapitola 3.* Lokalizace zvuku se v současné době využívá v hodně aplikacích. Praktický význam této problematiky souvisí s hlubším pochopením procesu lidského slyšení. Tyto poznatky lze uplatnit například v prostorové akustice, při vývoji asistivních pomůcek (naslouchadel), při vývoji a testování systémů prostorového zvuku, a také pro ověřování algoritmů virtuální reality.

Cílem této diplomové práce je především navrhnout rozšíření existujícího modelu binaurálního slyšení [9] o část umožňující lokalizaci směru vnímaného zvuku. Rozšíření má být implementováno na bázi strojového učení se zaměřením na lokalizaci v horizontální rovině. Výsledky lokalizace mají být srovnány se subjektivními daty. Nakonec má být zvážena možnost lokalizace ve vertikální rovině.

Tato diplomová práce je výsledkem dlouhodobé činnosti v rámci projektů ze studentských grantů SGS ČVUT. Součástí diplomové práce jsou čtyři experimenty s algoritmy pro lokalizaci zdroje zvuku. Prvním experimentem je lokalizace statického zdroje zvuku v přední horizontální polorovině pomocí klasifikátoru *k*-NN (*kapitola 5.1*). Výsledky tohoto experimentu byly publikované v [32]. Dalším experimentem je lokalizace statického a pohyblivého zdroje zvuku v přední horizontální polorovině pomocí *k*-NN a ANN (*kapitola 5.2*). Tato problematika je také součástí vlastní publikace [33]. Funkčnost tohoto algoritmu je ověřena třetím experimentem, který spočívá v lokalizaci statického zdroje zvuku v celé horizontální rovině (*kapitola 5.3*). Poslední částí práce je aplikace algoritmu lokalizace statického zdroje zvuku v horizontální rovině na lokalizaci statického zdroje zvuku ve vertikální rovině (*kapitola 5.4*).

Princip algoritmů především spočívá v několika dále popsaných krocích a je detailně popsán v *kapitole 4*. Vstupní signál je filtrován přenosovou funkcí hlavy, odpovídající určitému úhlu, a následně zpracováván binaurálním modelem slyšení. Ze získaných výstupů tohoto modelu se pak pomocí RMS extrahují příznaky, které jsou využity ke klasifikaci. Výstupem implementovaných algoritmů lokalizace je zejména závislost chyby lokalizace na referenčním úhlu.

1 Lidský sluch

Sluch je jedním z lidských smyslů a jedná se o schopnost vnímat zvukové podněty. Tato kapitola se zabývá popisem vnímání zvukových signálů.

1.1 Fyziologie sluchu

Princip lidského sluchu je založen na zachycení a zpracování zvukových signálů a jejich přeměně na nervové vzruchy. Lidské ucho se rozděluje na tři základní části – vnější, střední a vnitřní ucho (*Obr. 1.1*).

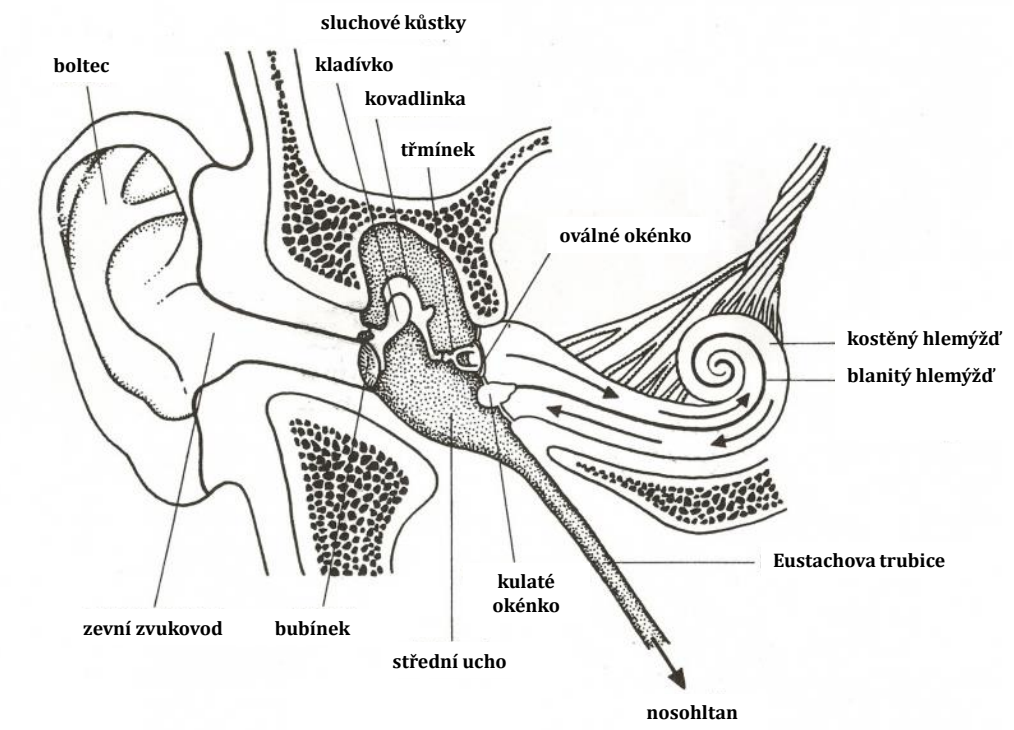
Tato kapitola čerpá z literatury [45], [56], [58]. **Vnější ucho** se skládá z boltce, zvukovodu a bubínku. Boltec umožňuje směrování akustických vln do zvukovodu. Geometrie boltce má za následek, že dochází ke frekvenční filtraci, která je závislá na směru příchodu zvuku. Tato prostorová závislost frekvenční filtrace napomáhá při rozlišování směru příchodu zvuku. Důležitou vlastností zvukovodu jsou jeho přenosové vlastnosti. Přenosová funkce zvukovodu ovlivňuje především citlivost lidského sluchu ve frekvenční oblasti lidské řeči.

Signály, které se dostaly do zvukovodu, procházejí dále k vazivové blance – bubínku, který se chvěje v důsledku příchozích zvukových vln a zároveň brání průchodu cizorodých těles do **středního ucha**. Zvukové vlny se dále přenáší přes bubínek do středního ucha, do bubínkové dutiny, ve které se nacházejí středoušní kůstky, Eustachova trubice, oválné a okrouhlé okénko. Dutina je vyplněná vzduchem a ohraničená lebkou. Středoušní kůstky (kladívko, kovadlinka, třmínek) tvoří pákový převod a spolu s poměrem velikostí bubínku a oválného okénka tvoří akustický impedanční transformátor mezi vzduchem a kapalinou uvnitř hlemýždě. Eustachova trubice, která spojuje bubínkovou dutinu a nosohltan, vyrovnává tlak ve středním uchu s tlakem okolního prostředí.

Oválné okénko je vstupem do **vnitřního ucha**, které obsahuje dva orgány (rovnovážný orgán a hlemýžď). Rovnovážný orgán dává mozku informace o vnímání polohy a pohybu hlavy. Hlemýžď je stočená trubička naplněná tekutinou. Uvnitř hlemýždě je bazilární membrána, na které se nachází tzv. Cortiho orgán. Tlakové vlny pohybují pružnou bazilární membránou, která deformuje vláskové buňky nacházející se v Cortiho orgánu, kde dochází k přeměně zvukového signálu na signál elektrický, který je pomocí sluchového nervu odváděn do mozku a tam dále zpracováván.

Dle [41] se nervové impulsy šíří sluchovým nervem a postupují ke komplexu horní olivy (SOC, Superior Olivary Complex), kde se schází a zpracovává informace z levého a pravého ucha. SOC se dělí především na laterální superior olivu (LSO, Lateral Superior Olive) a mediální superior olivu (MSO, Medial Superior Olive). LSO a MSO slouží k lokalizaci zdroje zvuku v horizontální rovině. Lokalizace zvuku v MSO probíhá na základě měření časového rozdílu příchodu signálu k uším. LSO naproti tomu pracuje na základě měření úrovňového rozdílu signálu.

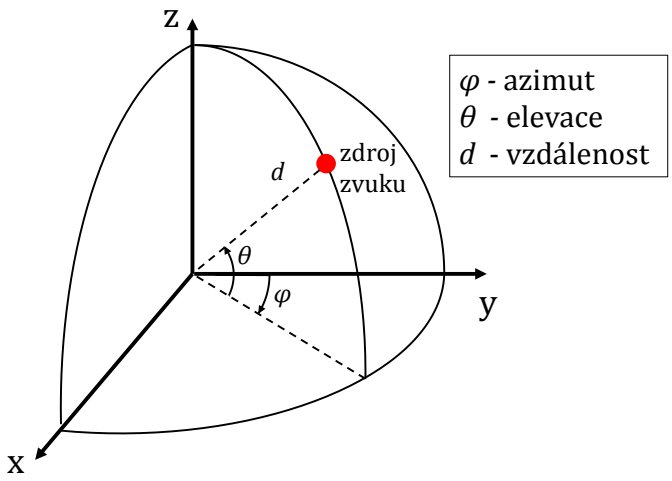
Informace o lokalizaci ve vertikální rovině se nezpracovává v SOC. Pro toto slouží tzv. dorsální kochleární jádro (DCN, Dorsal Cochlear Nucleus) [29].



Obr. 1.1 *Anatomie sluchového orgánu, převzato z* [48]

1.2 Binaurální slyšení

Binaurální slyšení znamená vnímání zvuku dvěma ušima. Jednou z hlavních vlastností binaurálního slyšení je to, že umožňuje určovat pozici a vzdálenost zdroje zvuku *d* (*Obr. 1.***2**). Jinými slovy umožňuje lokalizovat zdroj zvuku. Pozici zdroje zvuku lze určit na základě úhlu v horizontální rovině (azimutu, φ) a úhlu ve vertikální rovině (elevace, θ) [7]. Principy lokalizace zvuku vychází z fyziologických vlastností sluchu. Člověk není vždy schopen přesně odhadnout polohu zdroje zvuku, což vede k tzv. chybám lokalizace (LE, Localization Error). LE je rozdíl mezi odhadovanou a opravdovou pozicí zdroje zvuku v prostoru [19].



Obr. 1.2 *Pozice zdroje zvuku v polárních a prostorových souřadnicích, podle* [36]

Schopnost odhadu vzdálenosti zdroje zvuku vychází především z osobní zkušenosti. Interpretace spočívá ve frekvenční analýze. Díky fyzikální podstatě šíření zvuku lze předpokládat, že u vzdálenějšího zdroje zvuku nastane výraznější pokles vyšších frekvenčních složek [54].

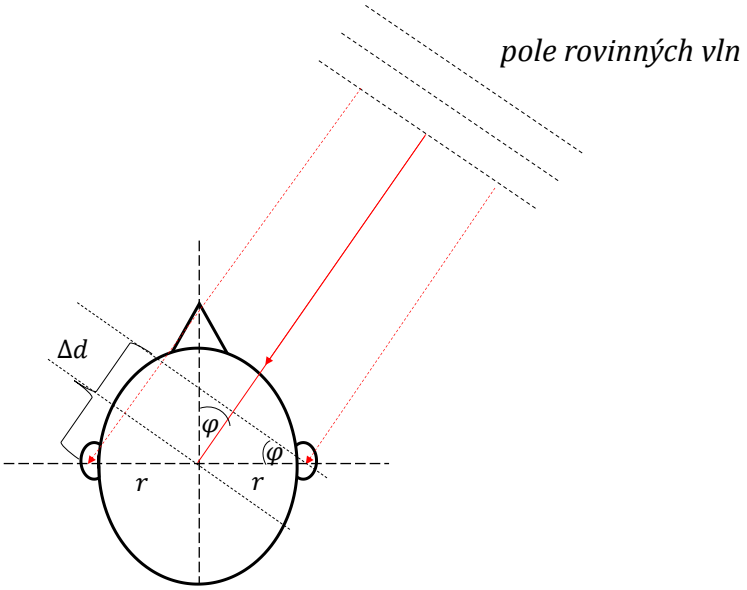
1.2.1 Horizontální rovina

Binaurální slyšení v horizontální rovině je především spojeno s rozdílem časů (ITD, Interaural Time Difference) příchodu signálu do pravého a levého ucha a také rozdílem jejich úrovní (ILD, Interaural Level Difference).

ITD vzniká kvůli odlišnému času příchodu signálu do levého a pravého ucha [7]. Tento časový rozdíl lze jednoduše spočítat při znalosti dráhového rozdílu vln (Δd), které přišly do levého a pravého ucha (*Obr. 1.3*). ITD v případě zdroje umístěného daleko od kulaté hlavy lze aproximovat následujícím vztahem [58]:

$$ITD = \frac{\Delta d}{c} = \frac{2 \cdot r \cdot \sin\varphi}{c} \text{ [s]}, \qquad \textbf{(1.1)}$$

kde c je rychlost šíření zvuku, r je poloměr hlavy a φ je azimutální úhel.

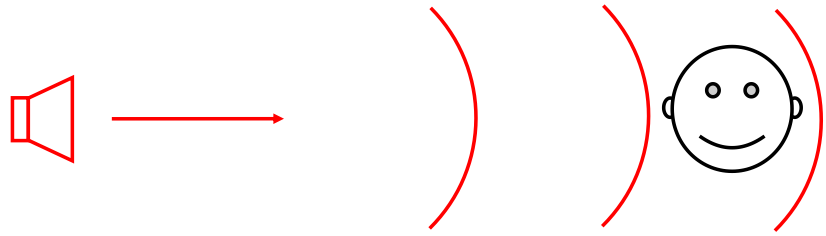


Obr. 1.3 *Ilustrační příklad výpočtu ITD*

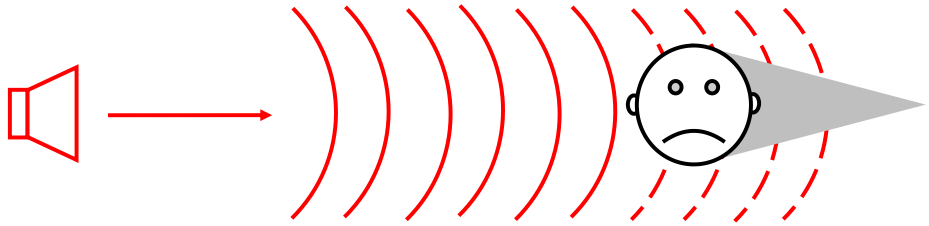
ILD je způsoben různou hodnotou úrovně zvuku [7]. Různá úroveň zvuku v uších nastane především v důsledku akustického stínu, který tvoří hlava. Akustický stín vzniká, když je vlnová délka příchozí vlny menší než poloměr hlavy, a také pokud je zdroj zvuku v blízkosti hlavy (*Obr. 1.4 a Obr. 1.5*) [17]. Hodnota ILD se udává v decibelech. Tento úrovňový rozdíl lze spočítat podle vztahu [12]:

$$ILD = 20 \log_{10} \frac{A_R}{A_L} \text{ [dB]}, \qquad \textbf{(1.2)}$$

kde A_R a A_L jsou úrovně akustického tlaku na bubínku pravého a levého ucha.



Obr. 1.4 *Ohyb vlny okolo hlavy*

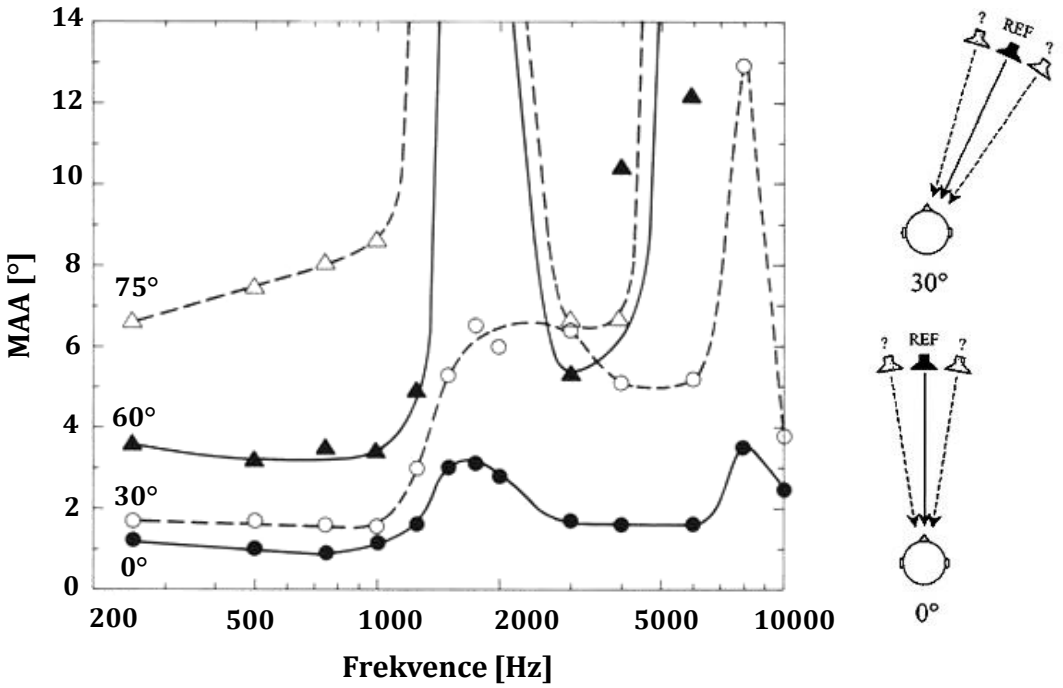


Obr. 1.5 *Vznik akustického stínu (vlnová délka je menší, než poloměr hlavy)*

Z tzv. duplexní teorie [57] vychází, že ITD se uplatňuje při lokalizaci nižších frekvencí (<1,5 kHz) a ILD se uplatňuje při lokalizaci vyšších frekvencí. V oblasti mezi 1,5 a 2 kHz, kde se částečně uplatňují obě metody, dochází k výraznému poklesu přesnosti lokalizace [44]. Hartmanova studie [22] oproti duplexní teorii naznačuje, že se ILD uplatňuje i při lokalizaci na nízkých kmitočtech, a to zejména v extrémních azimutálních úhlech. Spolu s tím, že většina reálných přírodních zvuků obsahuje jak vysoké, tak i nízké frekvenční složky, je tedy nutné pro lokalizaci metody kombinovat [55].

Hodnota ITD se pohybuje v rozsahu jednotek až stovek mikrosekund, maximální hodnota odpovídá přibližně 660 μs pro zdroj umístěný v úhlu $\varphi = \pm 90°$ a pro frekvenci menší než 1,5 kHz [7], [62]. Maximální hodnota ITD je úměrná velikosti hlavy člověka. Pro ILD je prahová hodnota rovna přibližně 0,4 dB až 17 dB [15], [64]. Tyto charakteristiky umožňují lokalizovat zvuky přicházející zepředu v horizontální rovině s přesností okolo 3° až 4° [54].

Na *Obr. 1.6* je znázorněna závislost parametru MAA (Minimum Audible Angle) na frekvenci. MAA je minimální detekovatelná změna směru zdroje zvuku v horizontální rovině. A. W. Mills [44] uvádí experiment, ve kterém má posluchač před sebou dva zdroje zvuku: referenční zdroj a druhý zdroj umístěný na jednu, nebo na druhou stranu od referenčního zdroje. Posluchač má určit, zda se druhý zdroj zvuku slyší zprava, nebo zleva vůči referenčnímu. Hodnota MAA se rovná úhlu mezi referenčním a druhým zdrojem, u kterého byl směr v 75 % měření lokalizován správně. Hodnota MAA se měří pro několik poloh referenčního zdroje. Z tohoto experimentu vychází, že přesnost lokalizace je závislá na frekvenci a také na směru příchodu zvuku.

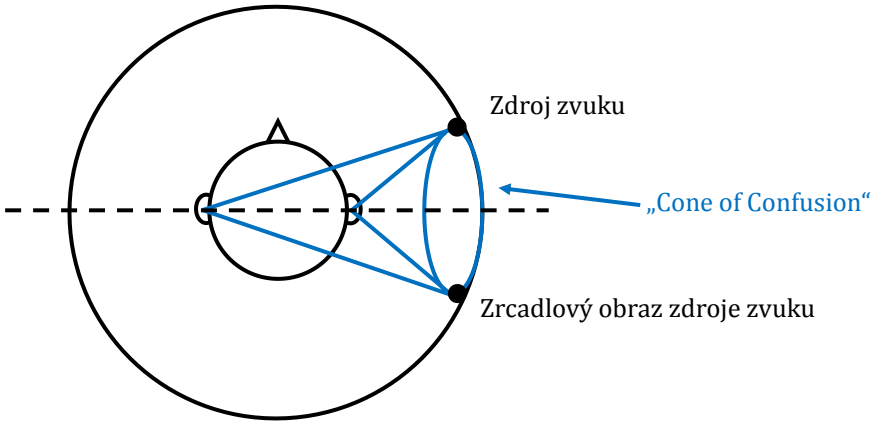


Obr. 1.6 *Závislost MAA na frekvenci, převzato z* [44]

1.2.2 Vertikální rovina

Existují situace, kdy se zdroj zvuku nachází ve stejné vzdálenosti od obou uší. Důsledkem toho je, že signály přicházející do obou uší jsou totožné (žádný časový rozdíl, shodné intenzity). Takový zdroj se musí nacházet v rovině kolmé ke spojnici obou uší uprostřed hlavy.

ITD a ILD jsou jenom nepatrně užitečné pro lokalizaci ve vertikální rovině na rozdíl od lokalizace v horizontální rovině. Za pomoci ITD a ILD není člověk schopen určit reálnou pozici zdroje zvuku, který je umístěn v oblasti tzv. *„Cone of Confusion"* (*Obr. 1.7*) [1], [44], [60]. Mezi jakýmikoli signály, které přicházejí z míst na povrchu kužele, budou stejné časové a úrovňové rozdíly (ITD a ILD). Pro spolehlivou lokalizaci zdroje zvuku je zapotřebí jiných charakteristik, které nejsou závislé na informacích z obou uší, jedná se o tzv. monaurální charakteristiky. Primární monaurální charakteristikou je funkce HRTF, která je probírána v následující kapitole.



Obr. 1.7 *„Cone of Confusion"*

Lokalizace ve vertikální rovině je především umožněna díky geometrickým vlastnostem ušního boltce, hlavy a těla, které fungují jako frekvenční filtry [3], [43], [49], [54]. Tyto filtry modifikují spektra zvukových signálů vstupujících do zvukovodu. Charakter filtrace záleží především na vzájemné pozici posluchače a zdroje zvuku ve vertikální rovině. Člověk rozpoznává elevaci porovnáváním spektra filtrovaného signálu se sadou přenosových funkcí hlavy, respektive se směrovými přenosovými funkcemi (viz kap. 1.3). Jako rozpoznaná hodnota elevace je zvolen směr, pro který přenosová funkce nejvíce korespondovala se spektrem.

Spektrální profil přenosových funkcí reprezentuje řadu širokých spektrálních maxim a ostrých minim. Tyto maxima a minima kolísají na frekvenční ose v závislosti na poloze zdroje zvuku. Podle [11] se první spektrální minimum nachází v rozsahu 6 až 12 kHz a má velký význam pro lokalizaci ve vertikální rovině.

Lokalizace je přesnější pro širokopásmové signály na vyšších frekvencích [50]. Přesnost lokalizace zvuku ve vertikální rovině je menší než v horizontální a odpovídá přibližně 10° – 15° [54].

Důležitou roli při lokalizaci zvuku ve vertikální rovině hraje osobní zkušenost, která je založená na dlouhodobém poslouchání a následné asociaci různých spektrálních charakteristik s určitými směry.

1.3 HRTF

Charakter zvuku se cestou od zdroje do bubínku změní [5], [7]. Modifikace mohou být způsobené například tvarem vnějšího ucha, tvarem hlavy, či těla. V charakteru změn je zakódovaná informace o poloze zdroje zvuku. Tyto změny mohou být popsány impulsní charakteristikou (HRIR, Head-Related Impulse Response). Impulsní charakteristika odpovídá konkrétní vzájemné poloze zdroje u levého, nebo pravého ucha. Po aplikaci Fourierové transformaci na HRIR vzniká HRTF (Head-Related Transfer Function, přenosová funkce hlavy). Přenos boltcem popisuje tzv. směrová přenosová funkce (DTF, Directional Transfer Function). DTF je v podstatě částí HRTF, která je zodpovědná za závislost na směru [42]. Přenosová funkce hlavy vychází ze speciálního měření a je individuální pro každého člověka.

1.4 Binaurální model

V této práci je použit binaurální model z [9]. Tento model se skládá ze dvou modelů, napodobujících chování mediální superior olivy (MSO) a laterální superior olivy (LSO) [8]. Princip tohoto modelu spočívá v tom, že se zprvu vstupní signál zpracovává v modelech lidských uší. Signál je aproximován FIR filtrem řádu 512, které simulují vnější a střední ucho. Aproximovaný signál je dále filtrován bankou filtrů, která simuluje frekvenční selektivitu hlemýždě (27 pásem v rozsahu 200 Hz až 7 kHz). Dále je signál filtrován dolní propustí, napodobující vláskové buňky. Filtrovaný signál je následně analyzován modely MSO a LSO, jejichž výstupem jsou tzv. lateralizační funkce. Lateralizací se myslí lokalizace zdroje zvuku uvnitř hlavy na spojnici mezi ušima.

2 Strojové učení

Strojové učení je oblast umělé inteligence, kde jsou počítačové algoritmy schopné rozhodovat nebo předpovídat budoucí vývoj na základě vstupních dat [53].

V oboru strojového učení se často uplatňuje **parametrizace.** Parametrizace je procesem určení nějaké číselné reprezentace, která může být použitá pro zlepšení charakterizace signálů pro účely rozpoznávání a je zásadním krokem v systémech klasifikace. Jako jednoduchý parametrizační prostředek je v této práci použit výpočet RMS (Root-Mean Square). RMS vyjadřuje efektivní hodnotu. Tato efektivní hodnota se dá jednoduše vypočítat pomocí kvadratického průměru. Parametr *RMS* pro signál *s* délky *N* (ve vzorcích) je dán vztahem [13]:

$$RMS = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^{N} s(n)^2}\,. \qquad \textbf{(2.1)}$$

Strojové učení se týká i oblasti klasifikace. **Klasifikace** je procesem rozdělování neznámých signálů do předem definovaných tříd. V této práci jsou třídami úhly příchozího zvuku.

Mezi způsoby strojového učení patří učení s učitelem a učení bez učitele. Následující popis vychází především z [23].

Metoda učení s učitelem je založená na učení funkce na základě tzv. trénovací množiny dat. Existuje tedy trénovací množina vzorů, která je složená z rysů vstupních objektů (vektorů příznaků) a požadovaných výstupů (označení tříd vstupních objektů). Cílem je najít neznámou závislost mezi vstupy a výstupy.

Učení bez učitele je případ, kdy není algoritmu poskytována označená trénovací množina. Při takovém učení se atributy vstupních objektů rozdělují na neprotínající se podmnožiny, tzv. klastry. Každý klastr se skládá ze shodných atributů objektů a rysy objektů různých klastrů se značně liší. Není tedy předem určené, jestli je vstupní objekt známý a patří do nějakého předem známého klastru, tj. třída vstupního objektu není předem známá.

Základní princip klasifikace spočívá v porovnání vstupního signálu s trénovací množinou vzorů.

Převážnou část kapitoly 2 jsem převzala ze své bakalářské práce, kde jsem se zabývala identifikací obsahu archivních zvukových záznamů [34].

2.1 Křížová validace

Přesnost klasifikačního modelu lze například určit pomocí tzv. metody křížové validace (CV, Cross-Validation). Přesnost klasifikace vyjadřuje míru spolehlivosti modelu správně klasifikovat data, na která model nebyl trénován, tedy data jemu neznámá. Křižová validace je založená na hodnocení přesnosti pomocí dat na základě

tzv. testovací množiny. Testovací množinou se myslí množina vzorů, pomocí které se hodnotí použitelnost modelu klasifikace [2], [16].

Jedním z typů křížové validace je tzv. *„v-fold Cross-Validation“* [30], kde se celá trénovací množina dat nejdříve rozdělí na *v* části. Algoritmus je následně natrénován na datech z (*v*-1) částí a testován na zbývající jedné části dat. Části se dále vyměňují *v*-krát, a ve výsledku je algoritmus trénován a testován na všech datech. Podle [30], [40] je odhad úspěšnosti spolehlivý při *v* = 10 (trénovací množině se přidělí 9/10 všech dat a testovací množině zbylá 1/10). V *Tab. 2.1* je znázorněn příklad křížové validace, kde je původní trénovací množina rozdělená na *v* = 5 podmnožin stejné velikosti (A-E). Zprvu je algoritmus trénován na částech B-E a testován na části A. Během 2. iterace se trénuje algoritmus na blocích A, C, D, E a testuje se na datech z části B. Bloky se vyměňují, dokud algoritmus nebude natrénován a otestován na všech částech.

	A	**B**	**C**	**D**	**E**
CV, 1. iterace	Testovací množina	Trénovací množina	Trénovací množina	Trénovací množina	Trénovací množina
CV, 2. iterace	Trénovací množina	Testovací množina	Trénovací množina	Trénovací množina	Trénovací množina
CV, 3. iterace	Trénovací množina	Trénovací množina	Testovací množina	Trénovací množina	Trénovací množina
CV, 4. iterace	Trénovací množina	Trénovací množina	Trénovací množina	Testovací množina	Trénovací množina
CV, 5. iterace	Trénovací množina	Trénovací množina	Trénovací množina	Trénovací množina	Testovací množina

Tab. 2.1 *Příklad metody „v-fold Cross-Validation“*

Křižová validace poskytuje přesnější hodnocení účinnosti algoritmu, než při testování jenom pomocí jedné části dat. Výsledné hodnocení je dáno průměrem úspěšností ze všech iterací.

2.2 Algoritmus *k*-nejbližších sousedů

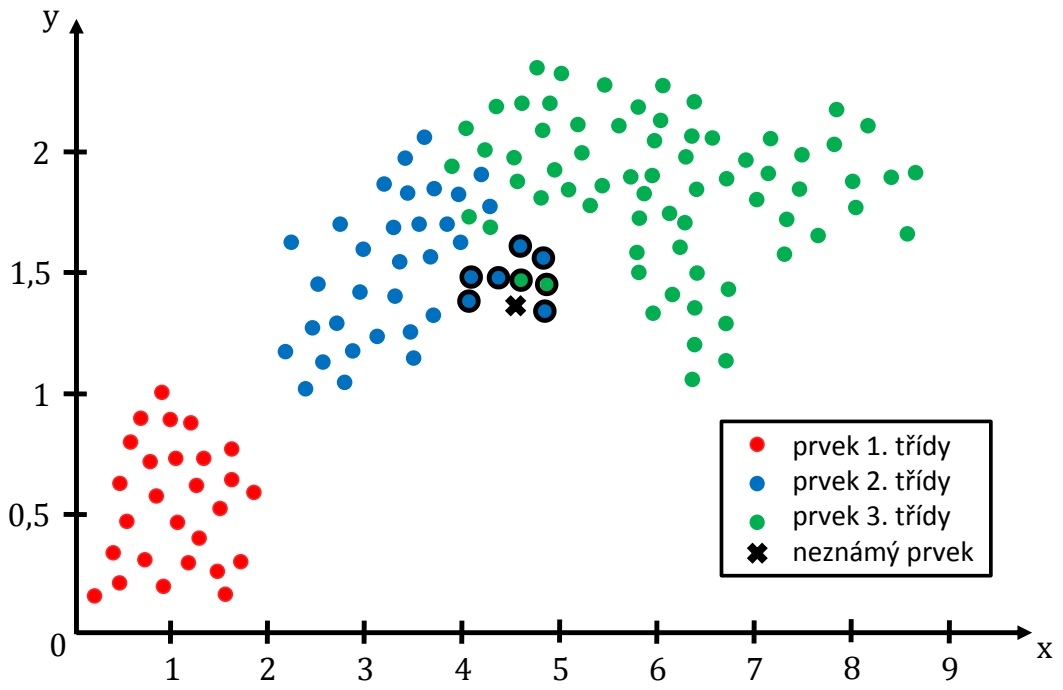
Metoda *k*-nejbližších sousedů (*k*-NN, *k*-Nearest Neighbors) je statistickou metodou klasifikace objektů [21], [25], [35]. Jedná se o metodu učení s učitelem, tj. metoda strojového učení pro učení funkce z trénovacích dat. Učení (trénování) je důležitým přípravným procesem pro klasifikaci. Trénovací sekvence obsahuje určité množství dat, patřících do konkrétních tříd, do kterých jsou potom testované části audio signálu ve fázi klasifikace zařazovány. Vytvořená trénovací množina je umístěná do některého místa *n*-rozměrného prostoru. Dále následuje proces klasifikace, kdy se testovaný prvek umísťuje do téhož prostoru a měří se *k*-nejbližších vzdáleností od testovaného prvku k trénovacím prvkům, tj. nalezení *k*-nejbližších sousedů. Objekt je pak klasifikován do té třídy, kam patří většina z těchto nejbližších sousedů.

Pro hledání nejbližší vzdálenosti v množině lze použít různé metriky. Dle literatury [10] je nejobvyklejší Euklidovská metrika, která je dána vztahem:

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}\,, \tag{2.2}$$

kde $x = x_1, x_2, \ldots, x_m$ jsou hodnoty parametrů testovacích dat a $y = y_1, y_2, \ldots, y_m$ jsou hodnoty parametrů trénovacích dat.

Ilustrační příklad klasifikace *k*-NN je znázorněn na *Obr. 2.1.*



Obr. 2.1 *Ilustrace klasifikace k-NN (Euklidovská metrika, k = 8)*

Hodnota *k* je celé číslo a může se pohybovat od jedné do počtu vzorků dat trénovací sekvence. Pokud se *k* = 1, jde o speciální zjednodušený případ, o tzv. metodu nejbližšího souseda (NN, Nearest Neighbor) [21].

Důležitým krokem je stanovení hodnoty *k*. Pro malé hodnoty není algoritmus stabilní vůči šumu. Při zvětšení této hodnoty se zvyšuje přesnost klasifikace, ale je potom těžší rozlišovat mezi jednotlivými třídami. Existují různé způsoby určení optimálního *k*. Jedním ze způsobů je stanovení *k* pomocí metody křížové validace (viz kap. 2.1), pomocí níž se zjišťuje úspěšnost pro různé hodnoty *k* a poté se vybere ta hodnota, pro kterou byla úspěšnost klasifikace nejvyšší [25].

Na *Obr. 2.2* je zobrazeno ilustrační blokové schéma algoritmu metody *k*-NN.

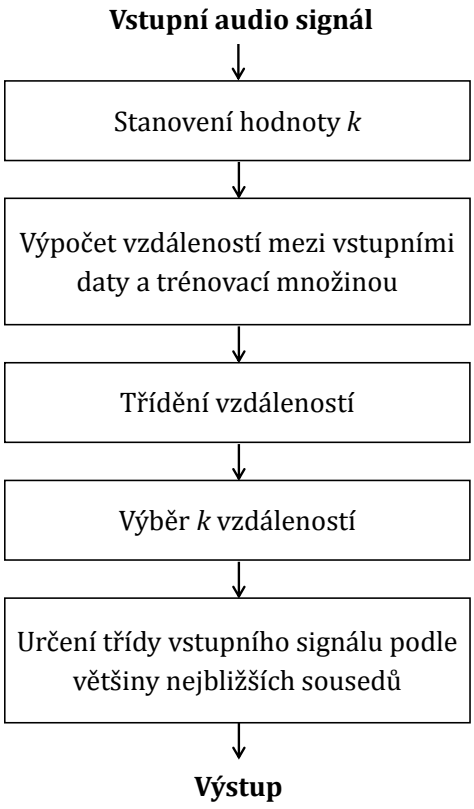
Metoda klasifikace *k*-NN má své výhody a nevýhody [25], [38].

Výhody:

- Velmi jednoduchý algoritmus
- Málo vstupních parametrů klasifikátoru (počet *k* a metrika)
- Učení je založeno na zapamatování trénovací množiny
- Efektivní a přesná

Nevýhody:

- Může být pomalým algoritmem (algoritmus musí vypočítat a roztřídit vzdálenosti mezi všemi prvky trénovací množiny, čím větší počet trénovacích dat, tím je menší rychlost klasifikace)
- Vyžaduje hodně paměti
- Výsledek zaleží na výběru metriky a počtu *k*



Obr. 2.2 *Algoritmus metody k-NN*

2.3 Umělá neuronová síť

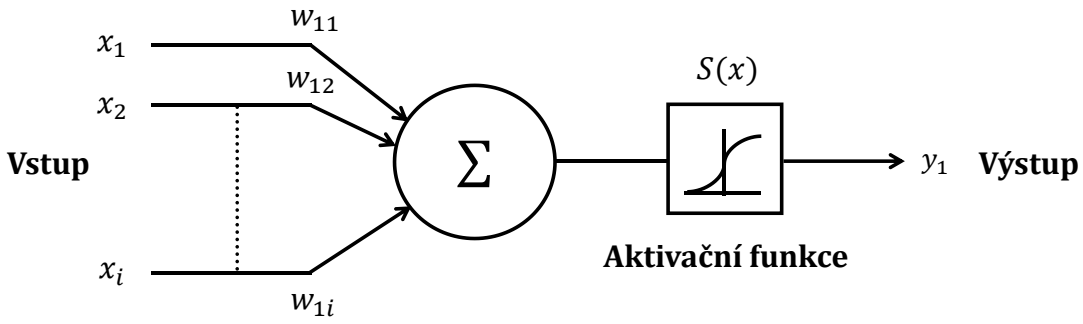
Umělá neuronová síť (ANN, Artificial Neural Network) je matematický model, jehož algoritmus napodobuje ve svém stylu chování biologické struktury (sítě nervových buněk živého organismu) [4].

ANN znamená systém spojených a vzájemně působících jednoduchých procesů (umělých neuronů) [6]. Hlavní funkcí umělého neuronu je generování výstupního signálu v závislosti na signálech přivedených na jeho vstupy. Neuron má libovolný počet vstupů, ale pouze jeden výstup. Vstupní data prochazejí jednosměrnými vazbami

tzv. synapsemi, kde každá synapse je chrakterizovaná vahovým koeficentem w_{ji}. Každá vstupní proměná x_i j-tého neuronu je násobena tímto synaptickým váhovým koeficientem w_{ji}. Dále se tyto násobky sčítají a výsledný součet je v neuronu transformován určitou nelineární přenosovou funkcí $S(x)$, tzv. aktivační funkcí, pomocí které se počítá výstupní signál umělého neuronu. Výstup základního modelu neuronu y_j je dán vztahem:

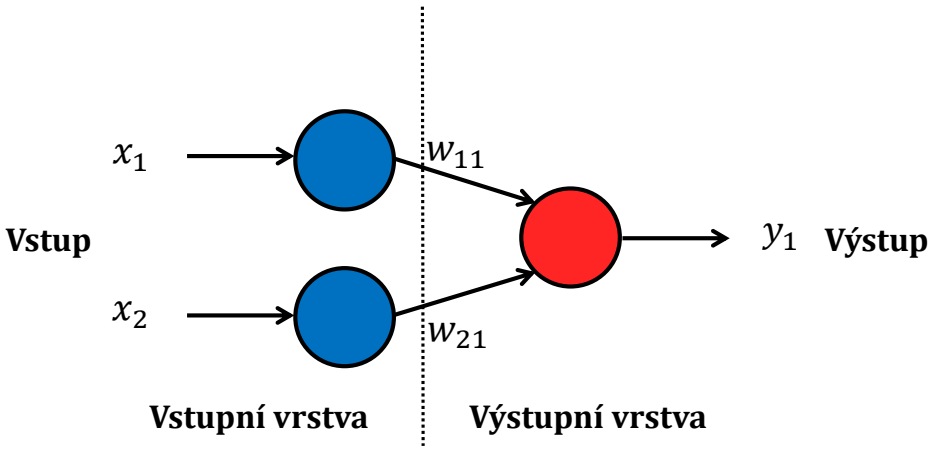
$$y_j = S\left(\sum_{i=1}^{n} x_i \cdot w_{ji}\right), \qquad \textbf{(2.3)}$$

kde n je počet vstupních proměných x_i [27], [51]. Grafické znázornění vztahu lze pozorovat na *Obr. 2.3*.



Obr. 2.3 *Základní model umělého neuronu, podle* [27]

Umělá neuronová síť může být jednovrstvá a vícevrstvá. Jednovrstvá neuronová síť (*Obr. 2.4*) má jenom jednu vstupní a jednu výstupní vrstvu, jedná se tedy o síť, ve které jsou všechny vstupní proměnné přímo spojené s výstupy [25].

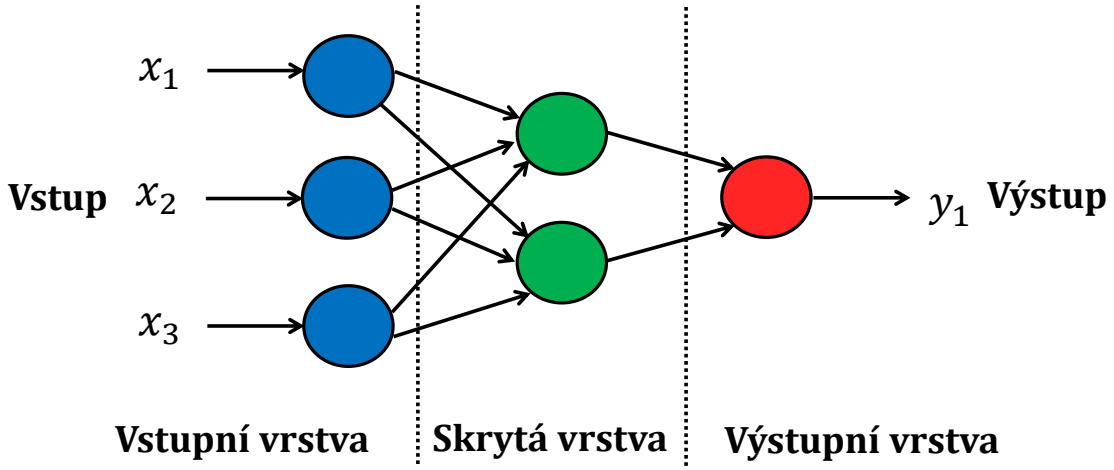


Obr. 2.4 *Jednovrstvá neuronová síť*

Vícevrstvá síť (*Obr. 2.5*) obsahuje minimálně tří vrstvy, jednu vstupní, minimálně jednu skrytou vrstvu a jednu výstupní vrstvu. Neurony se v jednotlivých vrstvách mezi sebou vzájemně neovlivňují, ale jsou závislé na neuronech z předchozí a následující vrstvy. Zároveň počet neuronů v jednotlivých vrstvách nezávisí na počtu neuronů ve vrstvách

jiných. Počet neuronů v jednotlivých vrstvách může být větší nebo rovný jedné. Data se v každé vrstvě zpracovávají paralelně.

Vstupní vrstva obsahuje neurony, které přijímají signál a posílají ho na vstupy neuronů následující skryté vrstvy [25]. Transferovaný signál postupně přechází do dalších skrytých vrstev, dokud nedojde do poslední výstupní vrstvy, jejíž výstup lze považovat za konečný pro celý model neuronové sítě.



Obr. 2.5 *Vícevrstvá neuronová síť*

Skryté vrstvy poskytují síti schopnost generalizovat (zobecňovat), což je schopnost rozhodovat o jiných datech, které nebyly součástí trénovacích dat. Literatura [28] uvádí, že velké úspěšnosti lze dosáhnout s neuronovou sítí, která má jednu nebo dvě skryté vrstvy. Velmi důležitý je výběr optimálního počtu neuronů ve skrytých vrstvách. Při příliš malém počtu neuronů se síť není schopna učit. Když je neuronů naopak hodně, zvyšuje se doba učení a může nastat přeučení sítě, tedy zhoršení schopnosti generalizace. T. Masters [37] v roce 1993 navrhl heuristické pravidlo *„Geometric Pyramid Rule"*, kde se počet neuronů ve skryté vrstvě (H) pro třívrstvovou síť (jedna skrytá vrstva) počítá podle následujícího vztahu:

$$H = \sqrt{N \cdot M}, \tag{2.4}$$

kde N je počet neuronů ve vstupní vrstvě a M je počet neuronů ve výstupní vrstvě. V případě čtyřvrstvé sítě (dvě skryté vrstvy) se počty neuronů v první a druhé skryté vrstvě (H_1, H_2) počítají podle vzahu:

$$H_1 = M \cdot \left(\frac{N}{M}\right)^{\frac{2}{3}}, \qquad H_2 = M \cdot \left(\frac{N}{M}\right)^{\frac{1}{3}}. \tag{2.5}$$

Pro účely klasifikace pomocí neuronových sítí se používá učení s učitelem (viz kap. 2). Trénovací množina je poslána na vstup neuronové sítě, po jejímž průchodu je poskytnut určitý výsledek. Tento výsledek se porovná s výsledkem požadovaným (třída dat

testovací množiny) a určí se chyba. Poté se upravují hodnoty synaptických váhových koeficientů tak, aby byla chyba mezi výstupní a požadovanou hodnotou pro všechny vzory učení minimální [18], [25]. Tento způsob učení se nazývá učení se zpětným šířením chyby (BPG, Back-Propagation of Gradient).

Literatura [25] a [31] uvádí popis kladů a záporů ANN.

Výhody:

- Algoritmus napodobuje funkci nervové soustavy živého organismu
- Umožňuje efektivně tvořit nelineární závislosti, které přesně popisují množinu dat
- Rychlý algoritmus (umožňuje paralelní zpracování informací)
- Odolný proti šumům vstupních signálů (model určí jejích škodlivost pro řešení klasifikace a automaticky zahodí tyto nepřínosné signály)

Nevýhody:

- Nelze zajistit opakovatelnost a jednoznačnost výsledků
- Není jasné, co se děje v rámci sítě (nelze prozkoumat kroky, jak byly vypočteny výstupní hodnoty, což komplikuje proces interpretace výsledků a modifikace sítě pro zlepšení přesnosti klasifikace)
- Mnoha krokové nastavení vnitřních prvků a vazeb mezi nimi
- Náročnost při vytvoření architektury ANN (pro různé příklady jsou různé architektury)

3 Současný stav metod lokalizace zvuku

Existuje řada algoritmů lokalizace binaurálního signálu. Většina z dohledaných algoritmů využívá pro klasifikaci úhlu umělou neuronovou síť.

V jedné z prací [46] se autoři zabývají lokalizací zdroje zvuku ve vertikální rovině. Navržený model se skládá ze dvou částí: spektrální analýzy pomocí umělého boltce a klasifikace elevace pomocí ANN. Algoritmus využívá toho, že savci lokalizují elevaci zvukového zdroje na základě monaurálních funkcí (v této práci se jedná o PRTF (Pinna Related Transfer Function)). Tyto funkce jsou závislé na elevaci zdroje zvuku a mají charakter úzkopásmových zádrží, které se liší v úrovni útlumu a ve frekvencích kmitočtových složek. Na vstupu neuronové sítě jsou použity frekvence a útlumy tří nejvíce významných složek v rozsahu 300 Hz až 4 kHz, které jsou detekované na základě frekvenční analýzy pomocí DFT (Discrete Fourier Transform). Autoři uvádějí chybu lokalizace do 5°.

V článku [14] autoři navrhli algoritmus lokalizace binaurálního signálu. Princip jejich algoritmu spočívá v několika krocích, ve kterých dochází k předzpracování vstupního signálu, výpočtu ITD a ILD charakteristik, clusterizaci a klasifikaci. Předzpracování je realizováno filtrací vstupního signálu bankou filtrů PHFB. ITD charakteristika je vypočtena vzájemnou korelací pravého a levého kanálu a také Jeffressovým modelem MSO. ILD je reprezentována modelem LSO. Výstupy těchto modelů jsou spojeny a použity pro clusterizaci pomocí tzv. metody „*k-Means*“ a tzv. samoorganizující mapy. Posledním krokem je detekce úhlu pomocí klasifikátoru *k*-NN a dvou různých typů neuronových sítí. Největší úspěšnost klasifikace je dosaženo v případě využití MSO modelu a při spojení výstupu z LSO a MSO. Tato úspěšnost se pohybuje v hodnotách vyšších než 90 %. Průměrná úspěšnost lokalizace s využitím modelu LSO je menší než 40 %.

V [59] je algoritmus lokalizace zdroje zvuku v horizontální rovině založen na použití třetí generace neuronových sítí (SNN, Spiking Neural Network). Na vstupu SNN se zpracovávají experimentálně získané HRTF pro každé ucho. Po několika dalších operacích se extrahují ILD charakteristiky, které dále slouží k detekci azimutu. Udávaná úspěšnost lokalizace pomocí SNN je přibližně 70 %.

Ben P. Yuhas se ve své práci [63] zabýval lokalizací řečových signálů v horizontální rovině. V algoritmu je vstupní signál zprvu předzpracováván modelem ucha a následně se z něj za pomoci dvou různých metod určují binaurální charakteristiky (ITD a ILD). Výsledný azimut je pak detekován pomocí ANN, která má vstupu získané binaurální charakteristiky. Autor ve svém algoritmu vyzkoušel také lokalizaci zdroje zvuku na surových datech, tj. pouze na výstupech z modelu ucha. Průměrná úspěšnost algoritmu se pohybuje okolo 80 %.

Ve zmíněných algoritmech jsou pro lokalizaci zdroje zvuku voleny různé strategie, využívající předzpracování signálů a následnou klasifikaci. Základní myšlenka algoritmu v této diplomové práci spočívá především ve využití jiného druhu modelace

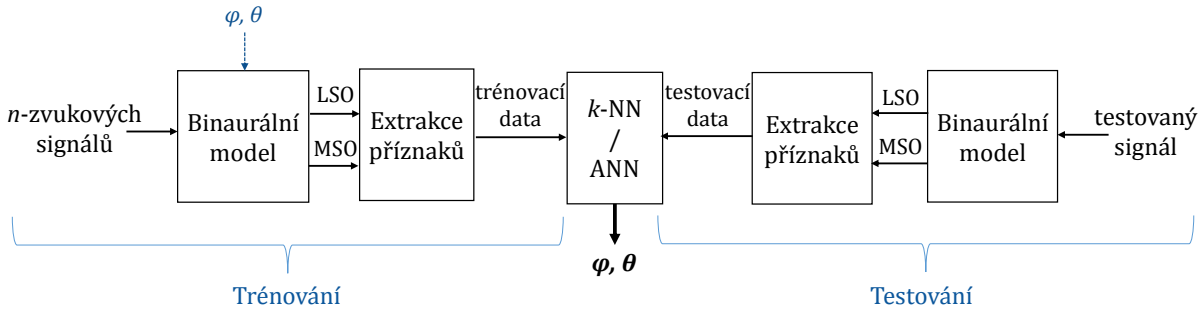
binaurálního slyšení. Použitý binaurální model je, stejně jako mnoho jiných, inspirován fyziologií lidského slyšení, ale zásadní rozdíl je v tom, že nepoužívá Jeffressovou zpožďovací linku. Nedávné experimentální výsledky z neurofyziologických dat a funkční magnetické rezonance naznačují, že Jeffressova zpožďovací linka není v savčím binaurálním systému přítomna [20], [39], [52]. Použité modely toto zohledňují a jsou těmito daty inspirovány.

4 Implementace algoritmu lokalizace

Implementovala jsem algoritmus lokalizace zdroje zvuku v horizontální a vertikální rovině s využitím klasifikátorů *k*-NN a ANN. Pro tyto klasifikátory jsem se rozhodla na základě prozkoumané literatury, existujících metod lokalizace zvuku (viz kap. 3) a vlastní zkušenosti s klasifikátory. Klasifikátor *k*-NN jsem již dříve implementovala ve své bakalářské práci [34], kde jsem se zabývala identifikací obsahu archivních zvukových záznamů. Kritériem výběru byly také výhody klasifikátorů popsané v kapitolách 2.2 a 2.3. Pro implementaci jsem zvolila programové prostředí MATLAB.

Základní princip algoritmu (*Obr. 4.1*) spočívá v předzpracování vstupního signálu pomocí binaurálního modelu slyšení (viz kap. 1.4), skládajícího se z modelů LSO a MSO. Výstupem tohoto binaurálního modelu jsou tzv. lateralizační funkce, ze kterých se následně extrahují příznaky, podle kterých probíhá klasifikace úhlu pomocí zmíněných klasifikátorů. Získané lateralizační funkce odpovídají určitým pásmům s centrálními frekvencemi f_c (viz kap. 1.4).

Ve vlastní publikaci [32] jsem jako příznaky pro klasifikaci testovala decimaci signálu a segmentální RMS. Význam segmentálního RMS a výsledky klasifikace s jeho použitím jsou uvedené v popisu prvního experimentu (viz kap. 5.1). V další vlastní publikaci [33] jsem jako příznaky použila už pouze RMS (viz kap. 2). Ve zbývajících experimentech (viz kap. 5.2 – kap. 5.4) jsem se rozhodla pro použití RMS na základě výsledků úspěšnosti klasifikace při použití těchto příznaků. Dalším důvodem pro výběr RMS byla také jednoduchost jeho výpočtu a to, že dokáže spolehlivě popsat obálku lateralizačních funkcí, které nesou informace o úhlech ve zvukových signálech.



Obr. 4.1 *Blokové schéma implementovaného algoritmu lokalizace zvuku*

5 Metody lokalizace a experimenty

V této kapitole jsou popsány implementované metody a experimenty pro lokalizaci zdroje zvuku v prostoru a je zde uvedeno porovnání výsledků se subjektivními daty.

5.1 Přední horizontální polorovina (*k*-NN)

V první části práce jsem se zabývala implementací algoritmu lokalizace statického zdroje zvuku v přední horizontální polorovině pomocí klasifikátoru *k*-NN (viz kap. 2.2). Parametr *k*, který charakterizuje počet nejpodobnějších prvků, jsem na základě experimentů stanovila na hodnotu 20. Klasifikačními třídami jsou azimuty příchozího zvuku $\varphi \in \langle -90°, 90° \rangle$ s krokem 5°, kde −90° odpovídá signálu umístěného u levého ucha, 0° odpovídá signálu před posluchačem v ose hlavy a 90° odpovídá signálu u pravého ucha.

Trénování klasifikátoru spočívá v generování širokopásmového šumu, který je následně filtrován HRTF funkcí (viz kap. 1.3), odpovídající určitému úhlu z rozsahu −90° až 90°. Sada HRTF byla naměřena na umělé hlavě KEMAR a je převzata z TU Berlin databáze [61].

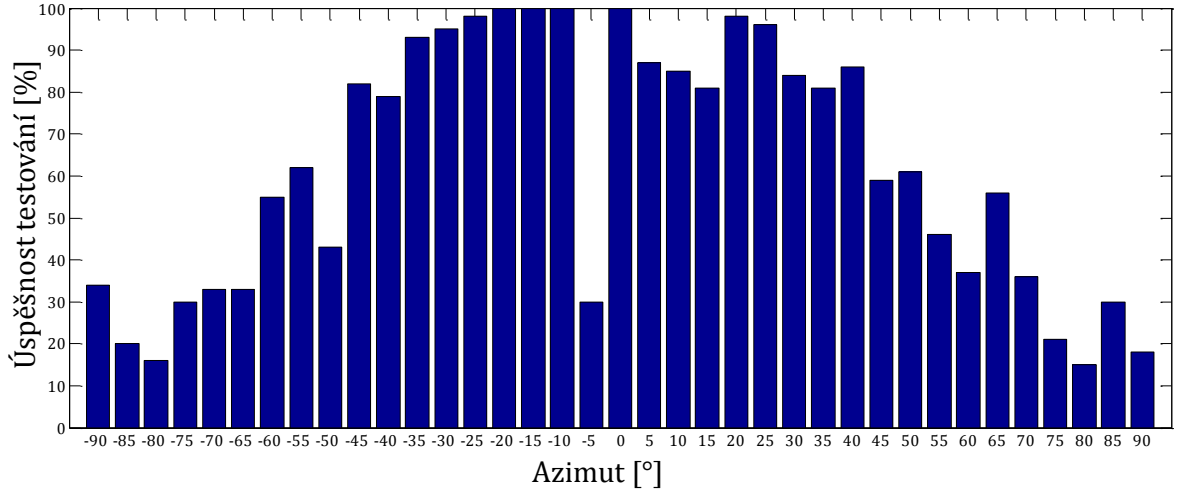
Na filtrovaný šum jsou dále aplikovány binaurální modely slyšení (LSO a MSO). Výstupy z těchto modelů jsou lateralizační funkce, které jsou dále pro každý model analyzovány zvlášť. Z těchto lateralizačních funkcí se pak počítají příznaky, které jsou získány pomocí segmentálního RMS. Segmentální RMS znamená, že je lateralizační funkce jednotlivých azimutů rozsegmentována a v každém segmentu délky 512 vzorků je vypočítaná hodnota RMS. Jinými slovy se získává obálka lateralizační funkce pro každý azimut. Trénovací množina obsahuje 20 vzorů pro každý azimut a je vytvořená pro modely LSO a MSO.

Dalším krokem je testování klasifikátoru. Testovací množina se tvoří stejně, jako trénovací, ale pro jiné realizace šumu. Hodnoty získaných příznaků testovací množiny pro jednotlivé azimuty jsou pak pomocí *k*-NN porovnávané se všemi hodnotami příznaků trénovací množiny a výstupem je třída odpovídající azimutu.

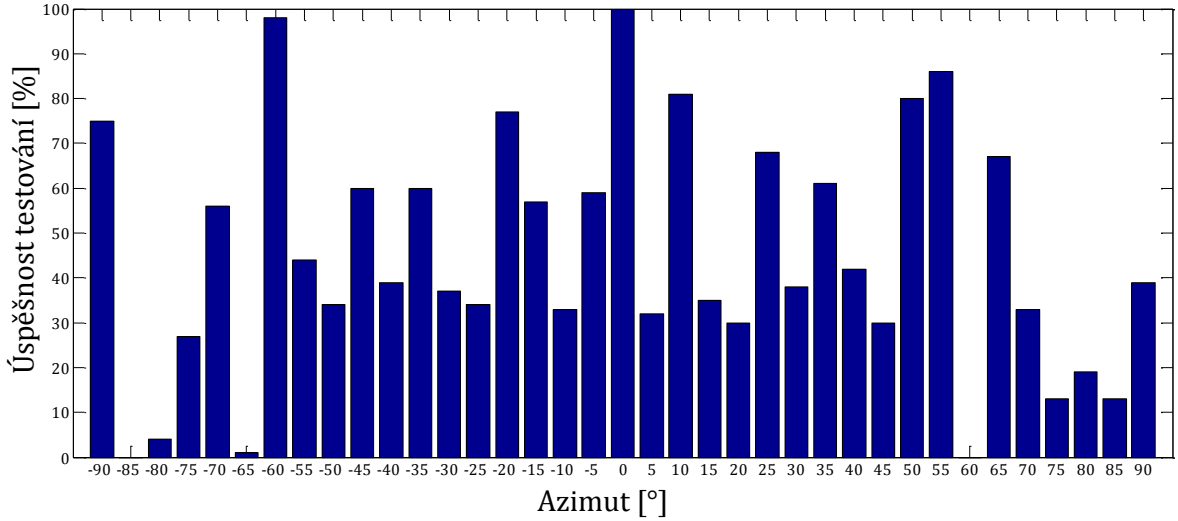
Výsledky této části práce byly publikované v [32] a jsou uvedeny na *Obr. 5.1* a *Obr. 5.2.* Úspěšnost je větší pro případ modelu LSO než MSO (*Tab. 5.1*). Úspěšnost algoritmu je vypočtena ze 100 vzorů z testovací množiny pro oba modely.

Model	Úspěšnost
LSO	60 %
MSO	46 %

Tab. 5.1 *Průměrná úspěšnost algoritmu pro modely LSO a MSO*



Obr. 5.1 *Úspěšnost algoritmu lokalizace statického zdroje zvuku pro model LSO pomocí k-NN (přední horizontální polorovina)*



Obr. 5.2 *Úspěšnost algoritmu lokalizace statického zdroje zvuku pro model MSO pomocí k-NN (přední horizontální polorovina)*

5.2 Přední horizontální polorovina (*k*-NN a ANN)

V dalším experimentu jsem se zabývala lokalizací statického a pohyblivého zdroje zvuku v přední horizontální polorovině pomocí *k*-NN a ANN (viz kap. 2), jejichž úspěšnost byla mezi sebou porovnávána. Výstupem celého algoritmu lokalizace je závislost chyby lokalizace (LE) na referenčním azimutu. Na grafech jsou znázorněny střední hodnoty a standardní odchylky vypočítané ze všech dat testovací množiny. Klasifikační třídy jsou stejné jako v předchozím experimentu (viz kap. 5.1).

Pro experimentální účely byly využity stimuly z databáze NOIZEUS [24], ITU Rec. P.501 [26], vlastní nahrávky a archivní nahrávky z filmů, které byly dodané vedoucím bakalářské práce [34]. Všechny tyto různě dlouhé nahrávky byly segmentovány na 950 sekundových signálů se vzorkovací frekvenci 44,1 kHz. Většina těchto nahrávek obsahuje řečový signál, ale nacházejí se zde také signály obsahující hluk, hudbu, nebo tóny hudebních nástrojů. Sada HRTF použitá v této části práci je převzata z TU Berlin databáze [61].

Sekundové signály ze sestavené databáze jsou filtrovány pomocí HRTF (viz kap. 1.3) odpovídající určitému azimutu. Filtrované signály jsou dále zpracovávané binaurálními modely (LSO a MSO). Z výstupů těchto modelů (lateralizačních funkcí) se pak počítají RMS příznaky (viz kap. 2). Na rozdíl od předchozího experimentu (viz kap. 5.1) jsou výstupy těchto modelů sloučeny dohromady už během trénování. Rozhodnutí o spojení výstupů vychází z pokusů studovaných v rámci vlastní publikace [33], kde jsem výstupy z modelů zkoušela kombinovat až po samostatné klasifikaci. Predikovaný azimut byl vypočítán podle vzorce:

$$\begin{cases} \dfrac{\varphi_{LSO} + \varphi_{MSO}}{2}, & |\varphi_{LSO} - \varphi_{MSO}| \leq 20 \\ \varphi_{LSO}, & |\varphi_{LSO} - \varphi_{MSO}| > 20. \end{cases} \qquad \textbf{(5.1)}$$

Takový způsob kombinace však přinášel nepřesnosti, a to především z toho důvodu, že správná klasifikace pomocí LSO modelu byla v mnoha případech ovlivněna chybnou klasifikací pomocí MSO modelu.

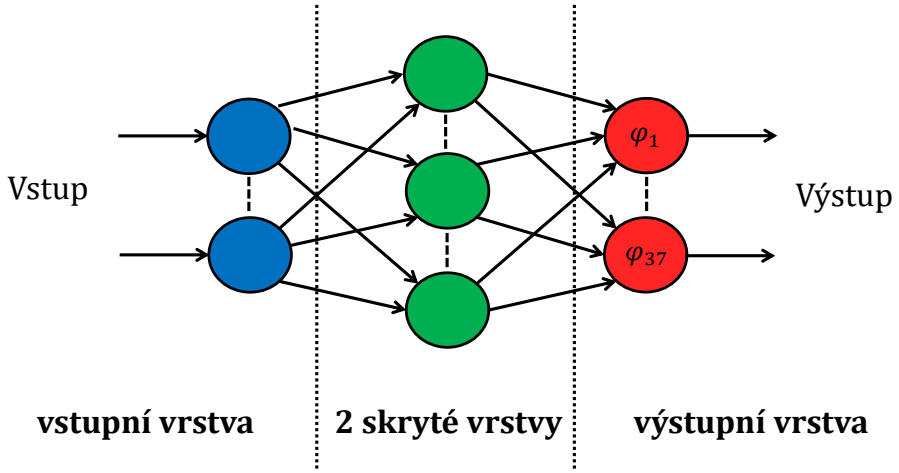
Z příznaků získaných ze zmiňovaných sekundových signálů je vytvořena **trénovací množina**, která obsahuje 950 příznakových vektorů (RMS hodnoty pro LSO a MSO modely) pro každý azimut a pásmo. Vygenerovaná trénovací množina je reprezentovaná maticí s rozměrem $950 \times 37 \times 27 \times 2$, kde 950 je počet signálů v množině, 37 je počet azimutů, 27 je počet centrálních frekvencí a 2 je dimenze příznakového vektoru. Na základě poznatků duplexní teorie (viz kap. 1.2.1) jsem omezila výstupy modelu MSO pouze na frekvenční pásma do 1 kHz. Výstupy LSO jsou počítány ve všech pásmech přibližně do 7,5 kHz.

Učení klasifikátoru ***k*-NN** spočívá pouze ve vytvoření této trénovací množiny. Pro učení **ANN** je třeba na datech z této množiny síť natrénovat. Pro implementaci jsem použila „Neural Network Toolbox" v MATLAB [47]. Byla vytvořena síť s topologií uvedenou na *Obr. 5.3.*

Topologie ANN (viz kap. 2.3):

- **Vstupní vrstva:** 37 neuronů (matice vypočítaných příznaků signálů, kde počet sloupců odpovídá počtu trénovacích dat pro všechny azimuty (950 × počet azimutů) a počet řádků je délka trénovacích signálů (počet centrálních frekvencí f_c pro LSO a MSO modely)
- **Skryté vrstvy** (2): 37, 37 neuronů
- **Výstupní vrstva**: 37 neuronů (každý neuron odpovídá určitému azimutu)

Matice vstupních parametrů je poslána do neuronové sítě, ze které vzejde určitý výsledek. Tento výsledek se porovná s požadovaným výsledkem a určí se chyba. Požadovaný výsledek odpovídá matici obsahující hodnoty v rozsahu 0 až 1, kde 0 znamená, že určité vstupní hodnoty vypočítaných příznaků neodpovídají určitému azimutu a 1 znamená, že odpovídají. Poté se automaticky upravují hodnoty synaptických váhových koeficientů tak, aby vznikla co nejmenší chyba (viz kap. 2.3). Neuronová síť je připravena ke klasifikaci po dokončení úpravy vah.



Obr. 5.3 *Topologie použité umělé neuronové sítě (ANN)*

5.2.1 Statický zdroj zvuku

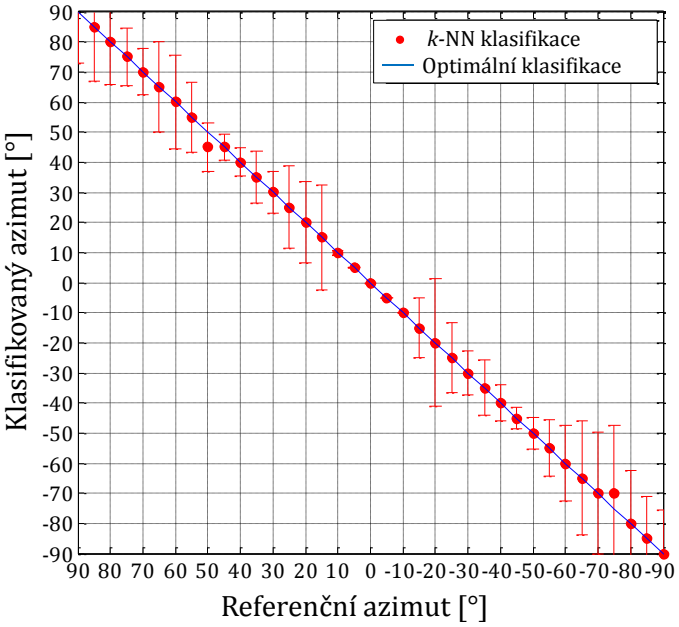
Pro testovaní klasifikátorů byla využita metoda *„v-fold Cross-Validation“* (viz kap. 2.1). Vytvořená trénovací množina dat, obsahující 950 signálů pro každý azimut, byla rozdělena na $v = 10$ podmnožin. Jedna podmnožina slouží jako testovací množina ($1/10 \times 950$ dat) a zbylé podmnožiny jako nová trénovací množina ($9/10 \times 950$ dat). Klasifikátory se natrénují na nové trénovací množině a pomocí testovací množiny se otestuje jejich úspěšnost. Tento proces se opakuje 10krát a pokaždé s prohozenými podmnožinami tvořícími novou trénovací a testovací množinu. Výsledné hodnoty úspěšnosti klasifikátorů k-NN a ANN jsou dány průměrnou hodnotou z hodnot úspěšnosti všech deseti iterací testování.

Pro klasifikátor ***k*-NN** je důležité stanovit parametr k na optimální hodnotu. Pro stanovení této hodnoty jsem využila také metodu *„v-fold Cross-Validation“* (viz kap. 2.1 a kap. 2.2), kde jsem určovala úspěšnost během všech 10 iterací pro různé hodnoty $k \in \langle 1, 35 \rangle$. Na základě výsledků této metody jsem pro k zvolila hodnotu 30.

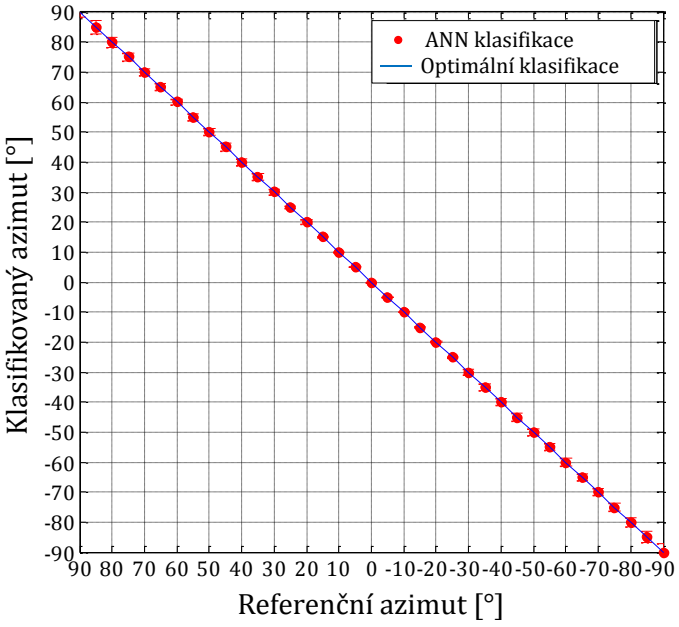
Pro testování k-NN jsou příznakové vektory testovací množiny porovnávané pomocí Euklidovské vzdálenosti se všemi příznakovými vektory trénovací množiny. Výstupem testování je třída odpovídající azimutu (viz kap. 2.2).

Testování klasifikátoru **ANN** spočívá v posílání testovacích dat na vstup sítě, kde se podle adaptovaných synaptických vah klasifikuje azimut testovaného signálu.

Na *Obr. 5.4* a *Obr. 5.5* jsou znázorněny výstupy algoritmu lokalizace statického zdroje zvuku pro klasifikátory k-NN a ANN. Červené body ukazují výsledky testování. V optimálním případě by měly body ležet na modré čáře.



Obr. 5.4 *Výsledky lokalizace v závislosti na referenčním azimutu pro k-NN pro statický zdroj zvuku (přední horizontální polorovina)*



Obr. 5.5 *Výsledky lokalizace v závislosti na referenčním azimutu pro ANN pro statický zdroj zvuku (přední horizontální polorovina)*

Tab. 5.2 znázorňuje úspěšnost klasifikátorů pro různé chyby lokalizace (LE) azimutu (viz kap. 1.2). ANN je schopna lepší generalizace, a proto dosahuje lepších výsledků než *k*-NN.

LE azimutu / **Klasifikátor**	**0°**	**5°**	**10°**	**> 10°**
***k*-NN**	67 %	18 %	4 %	11 %
ANN	97 %	3 %	0 %	0 %

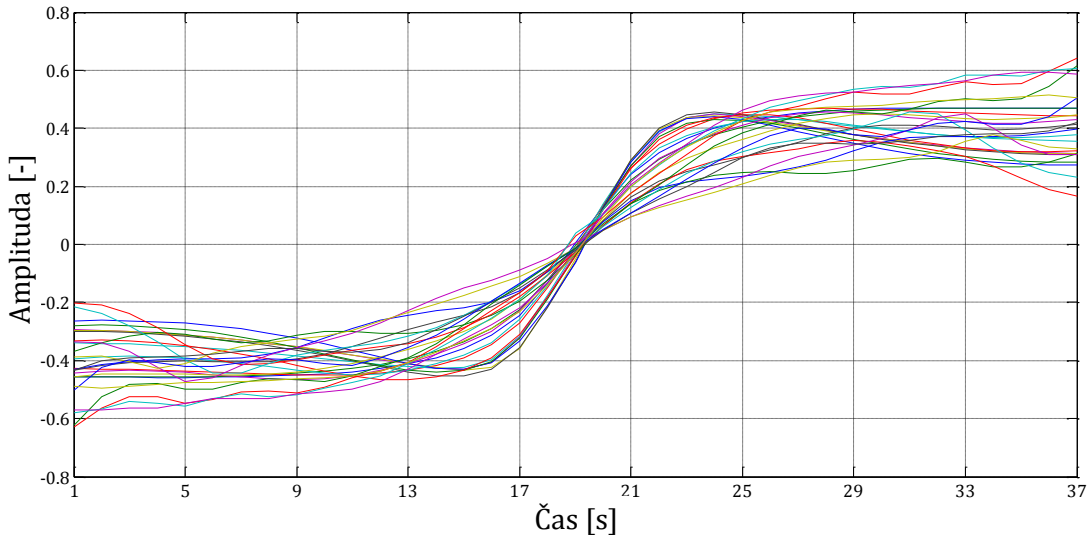
Tab. 5.2 *Průměrná úspěšnost lokalizace azimutu pro k-NN a ANN (přední horizontální polorovina)*

Porovnání se subjektivními daty

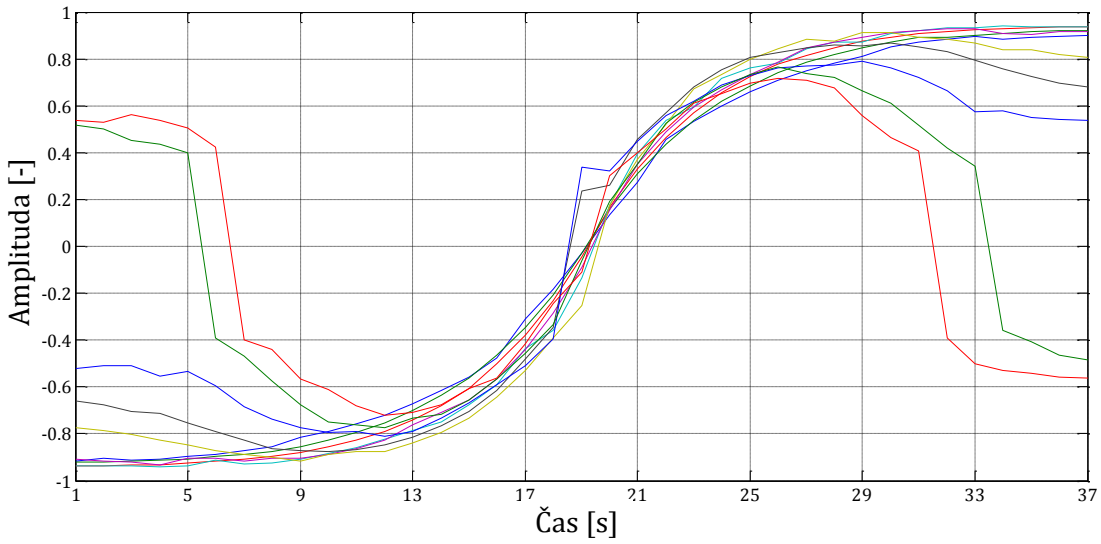
Klasifikátory mají větší úspěšnost při detekci azimutů okolo 0°, než při detekci vyšších hodnot azimutů (*Obr. 5.4 a Obr. 5.5*). Získané výsledky korespondují se subjektivními daty z experimentu A.W. Mills (viz kap. 1.2.1), kde je hodnota MAA nejmenší pro $\varphi = 0°$.

5.2.2 Pohyblivý zdroj zvuku

Testovací polohovaný signál představuje signál, u kterého se lineárně mění hodnota azimutu v čase. Hodnoty azimutů náleží rozsahu hodnot od $-90°$ do $90°$ s krokem 5°. Testovací signál je také generován pomocí HRTF, binaurálního modelu (respektive pomocí modelů LSO a MSO), a extrahovaní RMS příznaků z lateralizačních funkcí. Polohovaný parametrizovaný signál je 37 sekund dlouhý signál, kde každá sekunda odpovídá různému azimutu z definovaného rozsahu. Na *Obr. 5.6* a *Obr. 5.7* jsou znázorněny polohované parametrizované lateralizační funkce pro modely LSO a MSO. Průběh tvaru křivek v čase znázorňuje změnu azimutu v čase.

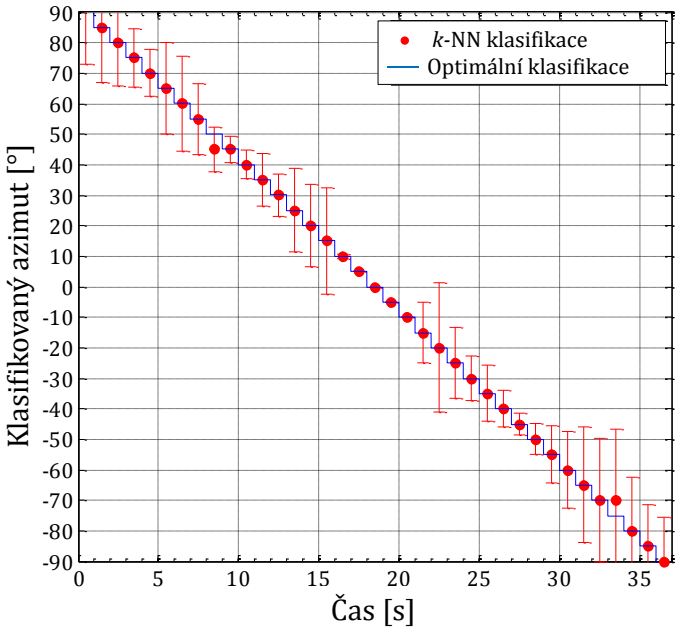


Obr. 5.6 *Polohované parametrizované lateralizační funkce pro model LSO (pro 27 frekvenčních pásem)*

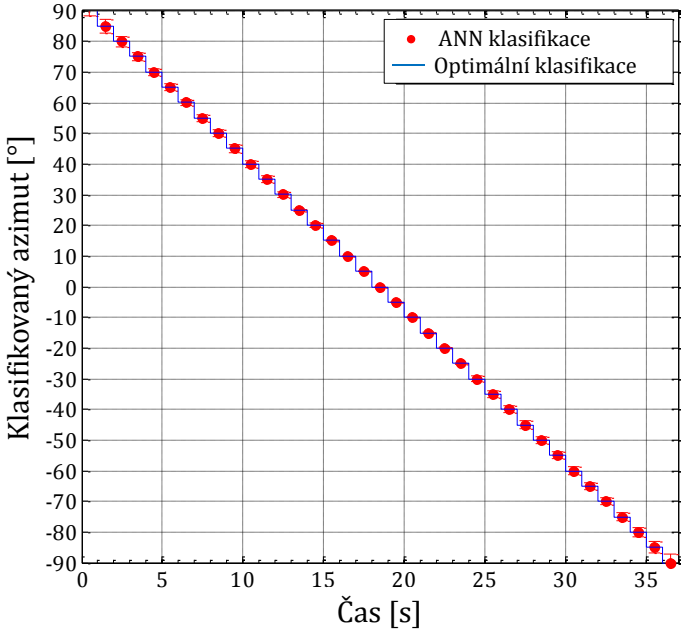


Obr. 5.7 *Polohované parametrizované lateralizační funkcí pro model MSO (pro 10 frekvenčních pásem)*

Při testování s pohyblivým zdrojem zvuku se také využívá metody *„v-fold Cross-Validation“* (viz kap. 2.1). Rozdíl je v obsahu testovací množiny, kterou tvoří signály polohované v horizontální rovině. Během testování je polohovaný parametrizovaný signál rovnoměrně segmentován na 37 segmentů, což odpovídá počtu klasifikačních tříd (počtu azimutů). Délka segmentu se rovná délce trénovacího signálu, tj. jedné sekundě. Po segmentaci je každý segment zpracováván samostatně, což znamená, že klasifikace každého segmentu odpovídá procesu lokalizace statického zdroje zvuku s vyžitím *k*-NN a ANN (viz kap. 5.2.1). Výsledkem je závislost klasifikovaného azimutu v čase (*Obr. 5.8* a *Obr. 5.9*).



Obr. 5.8 *Výsledky lokalizace v závislosti na referenčním azimutu pro k-NN pro dynamický zdroj zvuku (přední horizontální polorovina)*

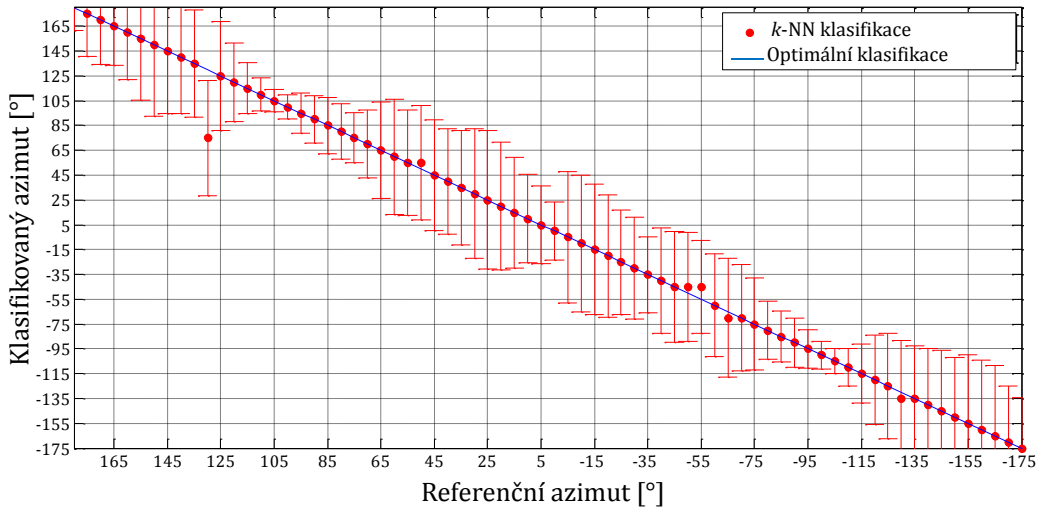


Obr. 5.9 *Výsledky lokalizace v závislosti na referenčním azimutu pro ANN pro dynamický zdroj zvuku (přední horizontální polorovina)*

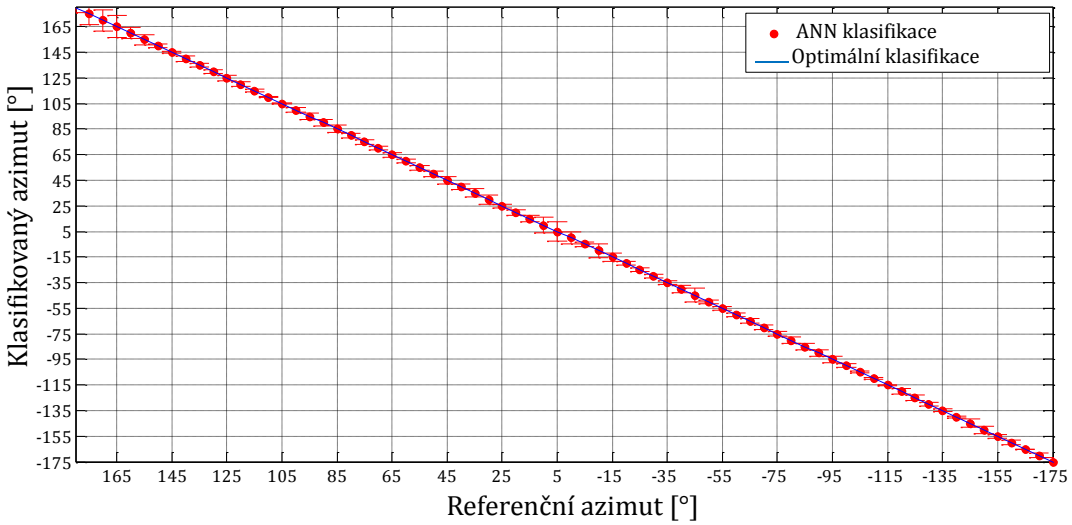
5.3 Celá horizontální rovina (*k*-NN a ANN)

Zabývala jsem se také experimentem, který spočíval v testování statického zdroje zvuku pro celou horizontální rovinu $\varphi \in \langle -180°, 180° \rangle$. Princip algoritmu je stejný jako v předchozím experimentu (viz kap. 5.2) až na to, že je algoritmus natrénován pro celou horizontální rovinu.

Výsledky testování s využitím ANN (*Obr. 5.11*) jsou lepší, než výsledky s využitím *k*-NN (*Obr. 5.10).* Tímto experimentem jsem ověřila, že implementovaný algoritmus pro přední polorovinu lze aplikovat i pro celou horizontální rovinu s relativně vysokou úspěšností (*Tab. 5.3*).



Obr. 5.10 *Výsledky lokalizace v závislosti na referenčním azimutu pro k-NN pro statický zdroj zvuku (celá horizontální rovina)*



Obr. 5.11 *Výsledky lokalizace v závislosti na referenčním azimutu pro ANN pro statický zdroj zvuku (celá horizontální rovina)*

LE azimutu / Klasifikátor	0°	5°	10°	> 10°
***k*-NN**	56 %	14 %	4 %	26 %
ANN	98 %	2 %	0 %	0 %

Tab. 5.3 *Průměrná úspěšnost lokalizace azimutu pro k-NN a ANN (celá horizontální rovina)*

5.4 Vertikální a horizontální rovina (ANN)

Pokusila jsem se aplikovat algoritmus lokalizace statického zdroje zvuku v horizontální rovině (viz kap. 5.2) na lokalizaci statického zdroje zvuku ve vertikální rovině.

V tomto algoritmu jsou využity stejné databáze signálů jako v experimentu v kapitole 5.2, ale bylo využito pouze 60 sekundových nahrávek namísto 950 nahrávek. Sekundové signály jsou zprvu filtrovány přenosovou funkcí hlavy odpovídající určitým elevacím v rozsahu $\theta \in \langle -30°, 60° \rangle$ s krokem 10° a azimutům v rozsahu $\varphi \in \langle -90°, 90° \rangle$ s krokem 5°. Záporné hodnoty úhlů odpovídají umístění zdroje dole a na levé straně vůči ose hlavy. Kladné hodnoty odpovídají umístění zdroje zvuku nahoře a na pravé straně vůči ose. Polohovaný signál je pak předzpracován binaurálními modely a z výstupů modelů se pro každé frekvenční pásmo z určitého rozsahu extrahují RMS příznaky. Pro model LSO je tento frekvenční rozsah omezen na 200 Hz až 7 kHz a pro model MSO je omezen do 1 kHz.

Pro filtraci vstupního signálu jsem použila HRTF a také jsem vyzkoušela filtrovat vstupní signál pomocí DTF (viz kap. 1.3). Sada těchto přenosových funkcí je převzata z ARI databáze [4], kde byla naměřena na hlavě člověka s označením „nh2“. Všechny možné kombinace azimutů a elevací jsou klasifikačními třídami pro lokalizaci zvuku. Počet těchto kombinací, respektive tříd, je 370 (37 hodnot azimutů × 10 hodnot elevace).

Jako klasifikátor jsem použila pouze ANN, kde jsou pro klasifikaci uvažovány frekvence důležité pro lokalizaci ve vertikální rovině (viz kap. 1.2.2). Tento experiment využívá následující topologii sítě, založenou na heuristickém pravidlu *„Geometric Pyramid Rule“* (viz kap. 2.3).

Topologie ANN:

- **Vstupní vrstva:** 37 neuronů
- **Skryté vrstvy** (2): 80, 172 neuronů
- **Výstupní vrstva**: 370 neuronů

Pro testování klasifikátorů je využita metoda *„v-fold Cross-Validation“*. Vytvořená množina dat, obsahující 60 signálů pro každou kombinaci azimutu a elevace, je rozdělena na 10 podmnožin ($v = 10$). Jedna podmnožina odpovídá testovacím datům (6 signálů pro každou kombinaci úhlů) a zbylé podmnožiny odpovídají trénovacím datům (54 signálů pro každou kombinaci úhlů). Klasifikátor se pak testuje pro všechna možná prohození těchto podmnožin a výsledky úspešnosti každého testování se zprůměrují.

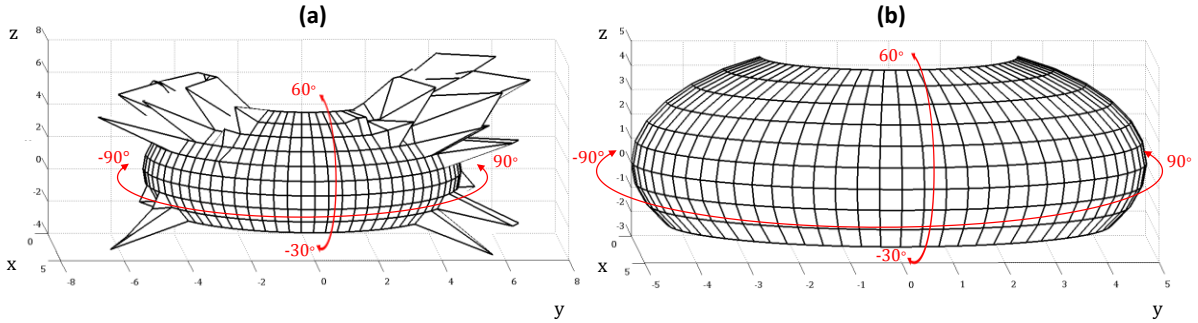
Výstupy tohoto experimentu jsou grafy ve tvaru polosféry znázorňující velikost chyb lokalizace (LE) azimutu a elevace. Na *Obr. 5.12* jsou výstupy algoritmu využívajícího DTF a na *Obr. 5.13* jsou výstupy algoritmu využívajícího HRTF. Lokalizační chyby jsou na grafech znázorněny výstupky na povrchu polosféry. Tyto výstupy vznikly převodem polárních souřadnic do kartézských:

$$x = r \cdot \cos\theta \cdot \cos\varphi,$$
$$y = r \cdot \cos\theta \cdot \sin\varphi, \qquad \textbf{(5.2)}$$
$$z = r \cdot \sin\theta.$$

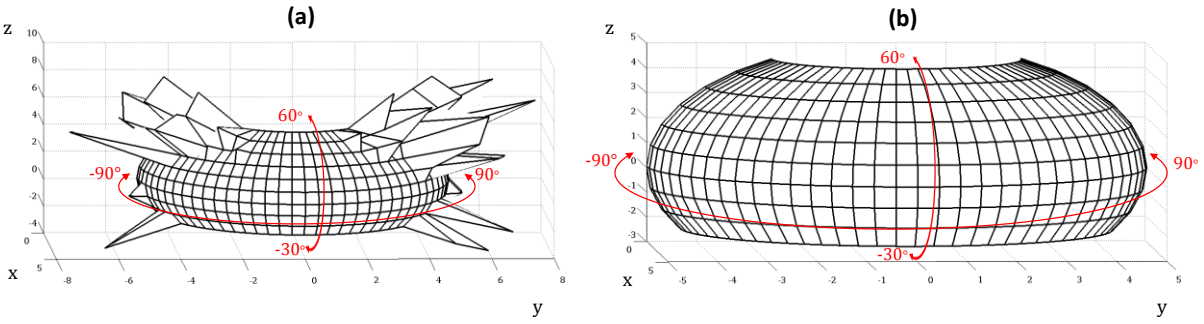
Rozdíl úspěšností algoritmů (*Tab. 5.4*) s využitím DTF a HTRF je nepatrný, což ukazuje, že pro implementovaný algoritmus nehraje výběr charakteristiky roli. Z *Obr. 5.12(a)* a *Obr. 5.13(a)* lze pozorovat, že pro hodnoty azimutů a elevací vzdálenějších od 0° je chyba lokalizace větší, než pro úhly okolo 0°. Chyba lokalizace elevace vyšla v tomto experimentu pro případ využití DTF i pro případ využití HRTF nulová (*Obr. 5.12(b)* a *Obr. 5.13(b)*).

LE / Klasifikátor	0°	5°	10°	> 10°
ANN (DTF)	93 %	6 %	1 %	0 %
ANN (HRTF)	92 %	7 %	1 %	0 %

Tab. 5.4 *Průměrná úspěšnost lokalizace azimutu a elevace pro ANN (DTF a HRTF, vertikální a horizontální rovina)*



Obr. 5.12 *Grafické znázornění velikosti chyby lokalizace pro DTF (vertikální a horizontální rovina): (a) – azimut, (b) – elevace; lokalizační chyby jsou znázorněny výstupky na povrchu polosféry*



Obr. 5.13 *Grafické znázornění velikosti chyby lokalizace pro HRTF (vertikální a horizontální rovina): (a) – azimut, (b) – elevace; lokalizační chyby jsou znázorněny výstupky na povrchu polosféry*

Nulová hodnota LE elevace může být způsobena efektivitou ANN (tím, že je sít dobře natrénovaná). Dále může být nulová chyba zapříčiněna modely LSO a MSO. Vliv modelů jsem ověřila na experimentu lokalizace elevace při nulových hodnotách azimutů pomocí ANN s využitím DTF a také na lokalizaci azimutu při nulových hodnotách elevace.

V těchto případech byl na vstupní signál aplikován pouze model ucha, bez aplikace modelů LSO a MSO. Jinými slovy jsou na vstupu ANN signály z levého a pravého ucha. Výsledná úspěšnost tohoto ověření je uvedena v *Tab. 5.5* Nulová chyba lokalizace v elevaci může být také zapříčiněna malým počtem testovacích dat a velkým krokem mezi možnými hodnotami elevace (10°).

LE	**0°**		**5°**		**10°**		**> 10°**	
	bez modelů	s modely	bez modelů	s modely	bez modelů	s modely	bez modelů	s modely
Elevace	96,6 %	100 %	0 %	0 %	1,2 %	0 %	2,2 %	0 %
Azimut	86 %	93 %	14 %	6 %	0 %	1 %	0 %	0 %

Tab. 5.5 *Srovnání úspěšnosti lokalizace s využitím binaurálních modelů a bez využití modelů pomocí ANN (DTF)*

Závěr

Cílem této diplomové práce bylo implementovat metody lokalizace v horizontální rovině za pomoci binaurálního modelu [9] a strojového učení. Získané výsledky jsem měla porovnat se subjektivními daty, a zvážit také možnost lokalizace ve vertikální rovině.

V teoretické části práce popisuji zkoumanou problematiku a v praktické části se zabývám popisem implementovaných metod.

Základní princip implementovaných algoritmů spočívá ve filtrování vstupního signálu přenosovou funkcí hlavy, následném předzpracování filtrovaného signálu binaurálními modely (LSO a MSO), dále v extrahování RMS příznaků z výstupů LSO a MSO modelů a v klasifikaci pomocí *k*-NN a ANN.

Zprvu jsem se zabývala lokalizací statického zdroje zvuku v přední horizontální polorovině pomocí *k*-NN. Výsledky klasifikace byly získány zvlášť pro LSO a MSO model. Z tohoto experimentu jsem zjistila, že lokalizace je úspěšnější pro LSO model (60 %).

Dále jsem rozšířila tento algoritmus o lokalizaci statického a pohyblivého zdroje zvuku pomocí *k*-NN a ANN, kde je lokalizovaný azimut detekován pro spojené výstupy z LSO a MSO. Tento algoritmus jsem aplikovala i pro celou horizontální rovinu, čímž jsem ověřila funkčnost algoritmu. Úspěšnost vychází v případě těchto dvou experimentů větší pro ANN (97 % a 98 %) než pro *k*-NN (67 % a 56 %). Klasifikátory také mají větší úspěšnost při určování azimutů kolem 0°, než při detekci vyšších hodnot azimutů, což koresponduje s daty ze subjektivních měření [44].

V posledním experimentu jsem využila algoritmus lokalizace statického zdroje zvuku v horizontální rovině na lokalizaci statického zdroje zvuku ve vertikální rovině pomocí ANN s využitím přenosových funkcí HRTF a také DTF. Průměrná úspěšnost detekce elevace vyšla 100 %, což je nejspíš způsobeno vlivem klasifikátoru a binaurálních modelů. Vliv binaurálních modelů jsem ověřila na experimentu, kde se úhel detekuje bez jejich využití. Nulová chyba také může být zapříčiněna malým počtem testovacích dat a větším krokem mezi hodnotami elevace.

Implementované algoritmy jsou spolehlivější v případě využití ANN než *k*-NN. ANN má na rozdíl od *k*-NN schopnost generalizace, což umožňuje neuronové síti dosahovat lepších výsledků. Testování algoritmu je zhruba 2krát rychlejší pro *k*-NN než ANN. Doba klasifikace jedné sekundové nahrávky na běžném osobním počítači vyšla pro *k*-NN okolo 6 ms a pro ANN okolo 12 ms. Klasifikátor *k*-NN má oproti ANN výhodu především v tom, že je jasné, co se v rámci klasifikátoru děje. U ANN nelze úplně prozkoumat kroky, jak byly vypočteny výstupní hodnoty.

Navržený systém dosahuje vysoké úspěšnosti pro lokalizaci zdroje zvuku. Dosažená průměrná úspěšnost je ve většině případech srovnatelná s úspěšností jiných zkoumaných algoritmů. Na rozdíl od jiných metod uplatňuje tento systém binaurální

model, který nepoužívá Jeffressovou zpožďovací linku. Algoritmus je připraven pro lokalizaci libovolného neznámého signálu.

Stanovené cíle byly dosaženy a vytvořené algoritmy jsou svou strukturou připraveny na možné vylepšení a na další rozšíření existujícího binaurálního modelu slyšení. Mezi možné vylepšení systému lokalizace by šlo například zařadit zvýšení adaptabilnosti vůči měnicímu se prostředí nebo použití jiných příznaků.

Seznam použité literatury a zdrojů

[1] ALGAZI, V. Ralph, et al. 2002. Approximating the head-related transfer function using simple geometric models of the head and torso. *The Journal of the Acoustical Society of America*, 112.5: 2053-2064.

[2] ALPAYDIN, Ethem. 2010. *Introduction to Machine Learning: Second Edition*. The MIT Press Cambridge, Massachusetts London, England. ISBN 978-0-262-01243-0.

[3] AL'TMAN, YA. A. 1990. *Slukhovaya sistema*. Nauka, Leningr. Otd-niye.

[4] ARI HRTF Database, *The Acoustics Research Institute of the Austrian Academy of Sciences* [online]. [cit. 2017-05-07]. Dostupné z: http://www.kfs.oeaw.ac.at/hrtf

[5] BEGAULT, Durand R.; TREJO, Leonard J. 2000. 3-D sound for virtual reality and multimedia.

[6] BISHOP, Christopher M. 1995. *Neural networks for pattern recognition*. Oxford university press.

[7] BLAUERT, Jens. 1997. *Spatial hearing: the psychophysics of human sound localization*. MIT press. ISBN 978-0-262-02413.

[8] BOUSE, Jaroslav; VENCOVSKY, Vaclav. 2015. Two-channel models of medial and superior olive based on psychoacoustics. *BMC Neuroscience*, 16.Suppl 1: P276.

[9] BOUŠE, Jaroslav. 2015. Model of binaural interactions. Doktorandské minimum. Praha: ČVUT FEL, Katedra radioelektroniky.

[10] BURRED, Juan José. 2003. An objective approach to content-based audio signal classification. *Masterthesis. Technische Universiteat Berlin, Berlin.*

[11] BUTLER, Robert A.; HUMANSKI, Richard A.; MUSICANT, Alan D. 1990. Binaural and monaural localization of sound in two-dimensional space. *Perception*, 19.2: 241-256.

[12] CALMES, Laurent. *Biologically inspired binaural sound source localization and tracking for mobile robots*. 2009. PhD Thesis. RWTH Aachen University.

[13] DAINTITH, John. 2009. *A Dictionary of Physics (6 ed.)*. Oxford University Press. ISBN 9780199233991.

[14] DAVILA-CHACON, Jorge; MAGG, Sven; LIU, Jindong; WERMTER, Stefan. 2013. Neural and statistical processing of spatial cues for sound source localisation. In: *Neural Networks (IJCNN), The 2013 International Joint Conference on*. IEEE. p. 1-8.

[15] GARDNER, William G. *3-D audio using loudspeakers*. Springer Science & Business Media, 1998.

[16] GEISSER, Seymour. 1993. *Predictive inference*. CRC press.

[17] GELFAND, Stanley A.; LEVITT, Harry. 1998. *Hearing: An introduction to psychological and physiological acoustics*. New York: Marcel Dekker.

[18] GERSHENSON, Carlos. 2003. Artificial neural networks for beginners. *arXiv preprint cs/0308031*.

[19] GRANTHAM, D. Wesley; HORNSBY, Benjamin WY; ERPENBECK, Eric A. 2003. Auditory spatial resolution in horizontal, vertical, and diagonal planes. *The Journal of the Acoustical Society of America*, 114.2: 1009-1022.

[20] GROTHE, Benedikt. 2003. New roles for synaptic inhibition in sound localization. *Nature Reviews Neuroscience*, 4.7: 540-550.

[21] GYÖRFI, László.; DEVROYE, Luc.; LUGOSI, Gábor. 1996. A probabilistic theory of pattern recognition. ISBN 0-387-94618-7.

[22] HARTMANN, William M.; RAKERD, Brad; CRAWFORD, Zane D. 2016. Transaural experiments and a revised duplex theory for the localization of low-frequency tones. *The Journal of the Acoustical Society of America*, 139.2: 968-985.

[23] HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. 2001. *The elements of statistical learning*. Springer, Berlin: Springer series in statistics.

[24] HU, Yi; LOIZOU, Philipos C. 2007. Subjective comparison and evaluation of speech enhancement algorithms. *Speech communication*, 49.7: 588-601.

[25] CHUBUKOVA, I.A. 2008. Data Mining: ucheb. Posobiye Osnovy informatsionnykh tekhnologiy. *Internet-un-t inform. Tekhnologiy.*

[26] ITU. 2012. Test signals for use in telephonometry, Int. Telecomm. Union, Geneva, Switzerland, ITU-T Rec. P.501.

[27] JAIN, Anil K.; MAO, Jianchang; MOHIUDDIN, K. Moidin. 1996. Artificial neural networks: A tutorial. *Computer*, 29.3: 31-44.

[28] JHA, Girish Kumar. 2007. Artificial neural networks and its applications. *IARI, New Delhi, girish_iasri@ rediffmail.com.*

[29] KANDEL, Eric R.; SCHWARTZ, James H; JESSEL, Thomas M. 2000. *Principles of neural science*. New York: McGraw-hill, pp. 591-624.

[30] KOHAVI, Ron. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Ijcai*. p. 1137-1145.

[31] KORNEYEV, D.S. 2007. *Ispol'zovaniye apparata neyronnykh setey dlya sozdaniya modeli otsenki i upravleniya riskami predpriyatiyami*. Upravleniye bol'shimi sistemami: sbornik trudov, №17, s. 81-102

[32] KOSHKINA, Ekaterina; BOUSE, Jaroslav. 2016. Lazy learning sound localization algorithm utilizing binaural auditory model. In: *Proc. of 20th International Student Conference on Electrical Engineering POSTER 2016*. 4 pp.

[33] KOSHKINA, Ekaterina; BOUSE, Jaroslav. 2017. Localization in Static and Dynamic Hearing Scenarios: Utilization of Machine Learning and Binaural Auditory model. In: *Proc. of 21th International Student Conference on Electrical Engineering POSTER 2017*. 5 pp.

[34] KOSHKINA, Ekaterina. 2015. *Identifikace obsahu archivních zvukových záznamů* Bakalářská práce. ČVUT v Praze. Vedoucí práce František Rund. 53 s.

[35] LAROSE, Daniel T. 2014. *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.

[36] LETOWSKI, Tomasz R.; LETOWSKI, Szymon T. 2012. *Auditory spatial perception: Auditory localization*. ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD.

[37] MASTERS, Timothy. 1993. *Practical neural network recipes in C++*. Morgan Kaufmann. ISBN: 0-12-479040-2.

[38] MATYASKO, A. A., KHAUSTOV V. A. 2012. Klassifikatsiya dokumentov v vektornom prostranstve. Sravneniye metodov Rokkio i metoda k-blizhayshikh sosedey. *BGUIR*. ISBN 978-985-488-926-9.

[39] MCALPINE, David; GROTHE, Benedikt. 2003. Sound localization and delay lines–do mammals fit the model?. *Trends in neurosciences*, 26.7: 347-350.

[40] MCLACHLAN, Geoffrey; DO, Kim-Anh; AMBROISE, Christophe. 2005. *Analyzing microarray gene expression data*. John Wiley & Sons.

[41] MEDDIS, Ray. 2010. *Computational Models of the Auditory System*. Springer US. ISBN 978-1-4419-5934-8.

[42] MIDDLEBROOKS, John C.; GREEN, David M. 1990. Directional dependence of interaural envelope delays. *The Journal of the Acoustical Society of America*, 87.5: 2149-2162.

[43] MIDDLEBROOKS, John C.; GREEN, David M. 1991. Sound localization by human listeners. *Annual review of psychology*. 42.1: 135-159.

[44] MILLS, Allen William. 1958. On the minimum audible angle. *The Journal of the Acoustical Society of America*, 30.4: 237-246.

[45] MOORE, Keith. 2013. Clinically Oriented Anatomy. 7th ed. Lippincott Williams & Wilkins. ISBN 978-1-4511-8447-1. pp. 848-849.

[46] MURRAY, John C.; ERWIN, Harry R. 2011. A neural network classifier for notch filter classification of sound-source elevation in a mobile robot. In: *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE. p. 763-769.

[47] Neural Network Toolbox, *The MathWorks* [online]. [cit. 2017-05-07]. Dostupné z: https://www.mathworks.com/help/nnet/

[48] NOVOTNÝ, Ivan; HRUŠKA, Michal; NEJTKOVÁ, Jana; VÁŇA, Michal. 1995. *Biologie člověka: pro gymnázia*. Fortuna.

[49] PAVANI, Francesco, et al. 2002. A common cortical substrate activated by horizontal and vertical sound movement in the human brain. *Current Biology*, 12.18: 1584-1590.

[50] ROFFLER, Suzanne K.; BUTLER, Robert A. 1968. Localization of tonal stimuli in the vertical plane. *The Journal of the Acoustical Society of America*, 43.6: 1260-1266.

[51] ROJAS, Raul. 1996. *Neural Networks: A Systematic Introduction*. Springer Science & Business Media.

[52] SALMINEN, Nelli H.; TIITINEN, Hannu; YRTTIAHO, Santeri; MAY Patrick J. C. 2010. The neural code for interaural time difference in human auditory cortex. *The Journal of the Acoustical Society of America*, 127.2: EL60-EL65.

[53] SAMUEL, Arthur L. 1959. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3.3: 210-229.

[54] SAPOZHNIKOV, M.A. 1978. *Elektroakustika*. Moskva: Svyaz'.

[55] SCHNUPP, Jan; NELKEN, Israel; KING, Andrew. 2011. *Auditory neuroscience: Making sense of sound*. MIT press.

[56] STANDRING, Susan (ed.). 2015. *Gray's anatomy: the anatomical basis of clinical practice*. Elsevier Health Sciences.

[57] STRUTT, John William. 1907. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13.74: 214-232.

[58] TBBITS, Adam; RICHARDSON Paul. 2005. *Gray's Anatomy for students*. Philadelphia: Elsevier /Churchill Livingstone.

[59] WALL, Julie A.; MCDAID, Liam J.; MAGUIRE, Liam P.; MCGINNITY, Thomas M. 2012. Spiking neural network model of sound localization using the interaural intensity difference. *IEEE transactions on neural networks and learning systems*, 23.4: 574-586.

[60] WALLACH, Hans. 1939. On sound localization. *The Journal of the Acoustical Society of America*, 10.4: 270-274.

[61] WIERSTORF, Hagen; GEIER, Matthias; SPORS, Sascha. 2011. A free database of head related impulse response measurements in the horizontal plane with multiple distances. In: *Audio Engineering Society Convention 130*. Audio Engineering Society.

[62] WOODWORTH, R. S., 1938. *Experimental Psychology*. New York: Holt, Rinehart, Winston.

[63] YUHAS, B. P. 1992. Automated sound localization through adaptation. In: *Neural Networks, 1992. IJCNN., International Joint Conference on*. IEEE. p. 907-912.

[64] ZHANG, Jie; MAO, DongXing. 2010. Dependence of binaural loudness summation on interaural level difference and frequency for pure tones. *Science China Physics, Mechanics and Astronomy*, 53.5: 834-841.

Přílohy

Příloha A (DVD)

Příloha B (dodatečná tabulka výsledků)

Příloha A

Přiložené DVD obsahuje složky s implementovanými algoritmy pro všechny 4 experimenty: ***01_experiment***, ***02_experiment***, ***03_experiment***, ***04_experiment***.

Popis těchto složek a jak postupovat při práci s programem je podrobně popsán v souboru **INFO.pdf**.

Složka ***BinauralModel*** je dodána vedoucími práce a obsahuje AMTtoolbox, SOFA toolbox, LTfat toolbox.

Příloha B

V příloze B je srovnávací tabulka obsahující průměrné úspěšnosti implementovaných algoritmů pro 2. až 4. experiment.

LE / **Klasifikátor**	**0°**	**5°**	**10°**	**> 10°**
Lokalizace zdroje zvuku v přední horizontální polorovině				
***k*-NN**	67 %	18 %	4 %	11 %
ANN	97 %	3 %	0 %	0 %
Lokalizace zdroje zvuku v celé horizontální rovině				
***k*-NN**	56 %	14 %	4 %	26 %
ANN	98 %	2 %	0 %	0 %
Lokalizace zdroje zvuku ve vertikální rovině				
ANN (HRTF)	92 %	7 %	1 %	0 %
ANN (DTF)	93 %	6 %	1 %	0 %

Tab. 1 *Průměrná úspěšnost implementovaných algoritmů*