



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta elektrotechnická

Katedra radioelektroniky

Aproximace výstupů modelu binaurálního slyšení

Approximation of Outputs of the Binaural Hearing Model

Individuální projekt č.2

Studijní program: Komunikace, Multimédia a Elektronika

Studijní obor: Multimediální technika

Vedoucí práce: Ing. František Rund, Ph.D.

Aleksandr Kartavykh

Praha 2018

České vysoké učení technické v Praze
Fakulta elektrotechnická
Katedra radioelektroniky

Zadání individuálního projektu č.2

Student: Aleksandr Kartavykh

Studijní program: Komunikace, Multimédia a Elektronika

Obor: Multimediální technika

Název tématu: Aproximace výstupů modelu binaurálního slyšení

Pokyny pro vypracování:

Seznamte se s problematikou binaurálního slyšení se zaměřením na určování směru příchodu zvuku. Navrhněte rozšíření existujícího modelu binaurálního slyšení o funkci umožňující popsat chování výstupních funkcí modelu binaurálního slyšení. Výstupem bude procentuální poměr shody výsledků aproximace se skutečnými výstupy modelu. Implementujte navržené rozšíření v prostředí Matlab a výsledky porovnejte s úspěšností přístupu pomocí strojového učení.

Vedoucí práce: Ing. František Rund, Ph.D.

Platnost zadání: do konce zimního semestru 2017/18

Z.S.

V Praze dne 11.10.2017

Čestné prohlášení

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

Dne 11.10.2017 v Praze

.....

Podpis studenta

Abstrakt

Tato práce se zabývá popisem chování výstupních funkcí modelu binaurálního slyšení. Zde je popsána základní problematika a aplikované principy, je zkoumána úspěšnost lokalizace pomocí aproximace ve srovnání s klasifikací pomocí strojového učení. Na základě výsledků bylo vytvořeno experimentální rozšíření v prostředí Matlab. Výstupem práce je procentuální poměr shody výsledků aproximace se dvěma skutečnými výstupy modelu, který tvoří 98,81% resp. 73,61%.

Klíčová slova: binaurální model slyšení, LSO, MSO, azimut, elevace, lokalizace, statický zdroj zvuku, aproximace, MATLAB

Abstract

This thesis focuses on the description of the behavior of output functions of the binaural hearing model. It describes basic issues and applied principles, examines the success of the localisation by approximation in comparison with the classification by machine learning. Based on the results, the experimental extension was created in the Matlab environment. The output of the work is the compliance ratio of the approximation results with two real outputs of the model, which are 98.81% and 73.61% resp.

Key words: binaural auditory model, LSO, MSO, azimuth, elevation, localization, static sound source, approximation, MATLAB

Seznam obrázků

Obrázek č.1 Ilustrační příklad výpočtu ITD [1]	7
Obrázek č. 2 Ohyb vlny okolo hlavy [1]	8
Obrázek č. 3 Vznik akustického stínu [1]	8
Obrázek č. 4 Etapy iteračního výběru koncových bodů	11
Obrázek č. 5 Blokové schéma algoritmu lokalizace zvuku	12
Obrázek č. 6 Průběh hodnot ITD pro přední horizontální polorovinu	12
Obrázek č. 7 Průběh hodnot ILD pro přední horizontální polorovinu	13
Obrázek č. 8 aproximované lateralizační funkce z MSO	14
Obrázek č. 9 Průběh aproximované lateralizační funkce z LSO	14

Seznam tabulek

Tabulka č.1 Koeficienty aproximačních polynomů	13
--	----

Seznam použitých zkrátek

DFT Discrete Fourier Transform

DTF Directional Transfer Function

FIR Finite Impulse Response

HRIR Head-Related Impulse Response

HRTF Head-Related Transfer Function

ILD Interaural Level Difference

ITD Interaural Time Difference

LSO Lateral Superior Olive

MSO Medial Superior Olive

Obsah

1. Úvod.....	6
2. Binaurální slyšení	7
2.1. Horizontální rovina	7
2.2. HRTF.....	8
2.3. Binaurální model	9
3. Aproximace.....	10
3.1. Polynomiální aproximace po částech.....	10
3.2. Iterační výběr koncových bodů.....	11
4. Realizace experimentu	12
4.1. Výstup binaurálního modelu.....	12
4.2. Aproximace polynomem 5-ého stupně	13
4.3. Shrnutí.....	15
5. Závěr.....	16
Citována literatura.....	17
Přílohy.....	18

1. Úvod

Binaurální slyšení je jednou z nejdůležitějších lidských schopností, protože mj. nám umožňuje lokalizovat v prostoru zdroj přicházejícího zvuku. Pozici lokalizovaného zvuku je možné popsat v polárních souřadnicích pomocí tří proměnných: úhlu v horizontální rovině (azimutu), úhlu ve vertikální rovině (elevace) a vzdálenosti [1].

Výsledkem zkoumání dané oblasti je model binaurálního slyšení z [3]. Ve druhé kapitole je popsán jak sám model, tak i okolnosti, za kterých pracuje. V diplomové práci [1] zmíněný model byl použit pro experiment s využitím strojového učení pro lokalizaci směru přicházejícího zvuku, kde výstupní funkce modelu byly parametrizovány a sloužily jako trénovací a testovací množiny pro strojové klasifikátory. Cílem této práce je zkusit jiný přístup k lokalizaci a navrhnout rozšíření modelu o funkci, která by dokázala popsat chování výstupních funkcí modelu pomocí aproximace. Následné porovnání s výsledky experimentu se strojovým učením dokáže říct, který postup je pro tuhle úlohu vhodnější.

Ve třetí kapitole je stručně popsána aproximace jako taková a také jsou uvedeny nejčastěji používané metody aproximace s ohledem na charakter aproximovaných dat (v daném případě jde o soubor experimentálních bodů, mezi kterými je předpokládána závislost).

Ve čtvrté kapitole je rozepsána praktická část dané práce: trochu je dotčen experiment se strojovým učením, protože začátek práce s modelem je v obou případech stejný, dále je implementována jedna ze zmíněných ve třetí kapitole metoda aproximace (polynomiální), potom jsou uvedeny číselné výsledky a na konci kapitoly je vyhodnocení výsledků a srovnání dvou přístupů k dané problematice.

2. Binaurální slyšení

Binaurální slyšení znamená, že vnímáme zvuk dvěma ušima. Lidské ucho je vysoce komplexní systém a neuvěřitelně precizní měřicí přístroj. Ať přichází hlas, který slyšíme, zepředu, zezadu, nebo ze strany – naše uši jednoznačně vnímají směr, a dokonce i vzdálenost.

Tahle kapitola je inspirována diplomovou prací [1], proto zde na něj dál nebude odkazováno.

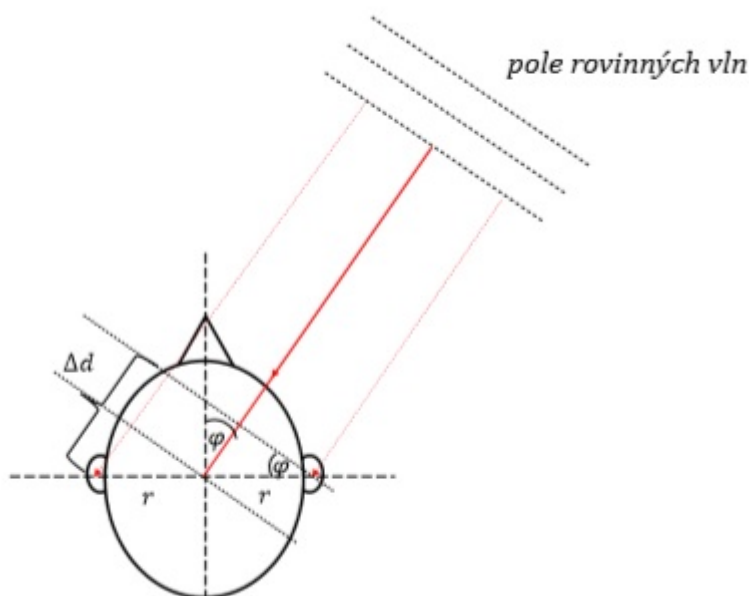
2.1. Horizontální rovina

Rozhodující jsou zde dvě podstatné schopnosti binaurálního (obouušního) slyšení: časové rozdíly mezi oběma ušima (Interaural Time Difference, ITD) a rozdíly jejich úrovní (Interaural Level Difference, ILD).

ITD vzniká kvůli odlišnému času příchodu signálu do levého a pravého ucha [6]. Tento časový interval lze jednoduše spočítat při znalosti dráhového rozdílu vln (Δd), které přišly do levého a pravého ucha (Obr. 1). ITD v případě zdroje umístěného daleko od kulaté hlavy lze aproximovat následujícím vztahem:

$$ITD = \frac{\Delta d}{c} = \frac{2 \cdot r \cdot \sin \varphi}{c} [s] \quad (1)$$

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



Obrázek č.1 Ilustrační příklad výpočtu ITD

ILD je způsoben různou hodnotou úrovně zvuku [6]. Odlišnosti zvukové úrovně způsobuje naše hlava, která tvoří přirozenou překážku (a tvoří tzv. "akustický stín"), od níž se zvuk odráží.. Akustický stín vzniká, když je vlnová délka přichozí vlny menší než poloměr hlavy, a také

pokud je zdroj zvuku v blízkosti hlavy (Obrázek č. 2 a Obrázek č. 3) [2]. Hodnota ILD se udává v decibelech. Tento úrovnový rozdíl lze spočítat podle vztahu

$$ILD = 20 \log_{10} \frac{A_R}{A_L} \quad (2)$$

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



Obrázek č.2 Ohyb vlny okolo hlavy



Obrázek č.3 Vznik akustického stínu

Hodnota ITD se pohybuje v rozsahu jednotek až stovek mikrosekund, maximální hodnota odpovídá přibližně 660 μ s pro zdroj umístěný v úhlu $\varphi = \pm 90^\circ$ a pro frekvenci menší než 1,5 kHz [6]. Maximální hodnota ITD je úměrná velikosti hlavy člověka. Pro ILD je prahová hodnota rovna přibližně 0,4 dB až 17 dB. Tyto charakteristiky umožňují lokalizovat zvuky přicházející zepředu v horizontální rovině s přesností okolo 3° až 4° .

2.2 HRTF

Charakter zvuku se cestou od zdroje do bubínku změní [6]. 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 [7]. Přenosová funkce hlavy vychází ze speciálního měření a je individuální pro každého člověka.

2.3. Binaurální model

V této práci je použit binaurální model z [3]. 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.

3. Aproximace

Aproximací rozumíme nepřesné, ale věrné vyjádření nějaké entity, nejčastěji čísla nebo funkce, ale také fyzikálního zákona (přírodního jevu) [4]. K využití aproximace se přistupuje tehdy, když pro analytické řešení není dostatek informací, nebo toto je příliš náročné na provedení. I v případě, kdy je přesné vyjádření známé, může aproximace být užitečná — poskytnutím dostatečně přesného řešení a zároveň podstatným snížením složitosti problému.

Pro souhrn experimentálních bodů (x_i, y_i) kde $i = 0, 1, 2, \dots, M$ a kde platí, že body (x_i, y_i) a (x_{i+1}, y_{i+1}) jsou nejbližší sousedy, úlohou aproximace je najít takovou funkci $\hat{y} = g(x)$, že chyba aproximace, tj. velikost odchylky původních bodů (x_i, y_i) od bodů $(x_i, g(x_i))$, bude minimální [4].

Současně mezi nejpoužitelnějšími metodami aproximace funkčně spojených experimentálních hodnot patří [5]:

- Polynomiální aproximace po částech;
- Iterativní výběr koncových bodů.

3.1. Polynomiální aproximace po částech

V případě použití tyhle metody aproximační funkce má tvar:

$$\hat{y} = a_0 + a_1x + a_2x^2 + \dots + a_Nx^N \quad (3)$$

kde a_n jsou koeficienty polynomu.

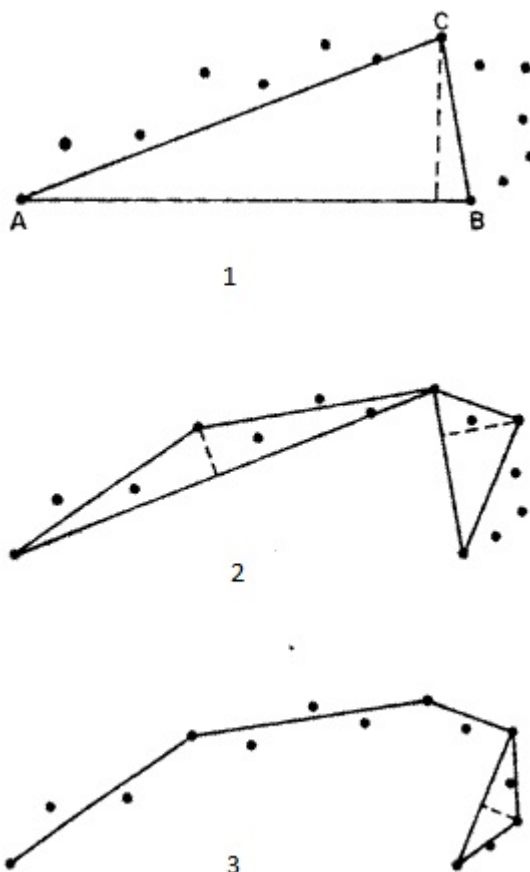
Po dosažení experimentálních hodnot do vztahu (3) dostáváme soustavu rovnic

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^N \\ 1 & x_1 & x_1^2 & \dots & x_1^N \\ 1 & x_2 & x_2^2 & \dots & x_2^N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_M & x_M^2 & \dots & x_M^N \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_M \end{bmatrix} \quad (4)$$

V případě lineární aproximace je třeba určit jen koeficienty a_0 a a_1 . V opačném případě, když funkce $\hat{y} = g(x)$ je polynom n -tého řádu, vztah (4) platí pro každý experimentální bod, tj. aproximační polynom prochází každým experimentálním bodem. Takový polynom je jenom jeden, ale dá se k němu dospět i jinou cestou [5]. Také podle [5] polynomem 5-ého stupně se dá aproximovat téměř jakoukoli křivku, proto pro konečnou implementaci byla zvolena tato metoda.

3.2. Iterační výběr koncových bodů

Tato metoda spočívá v lineární aproximaci po částech. Algoritmus aproximace je znázorněn na obrázku č.4. Na začátku jsou koncové body A a B spojeny přímkou. Pak se hledá bod C nejvzdálenější vůči dané přímce, který se stává propojením dvou přímek AC a BC. Tento postup se opakuje pro každou přímkou až všechny body budou “dobře” aproximovány.



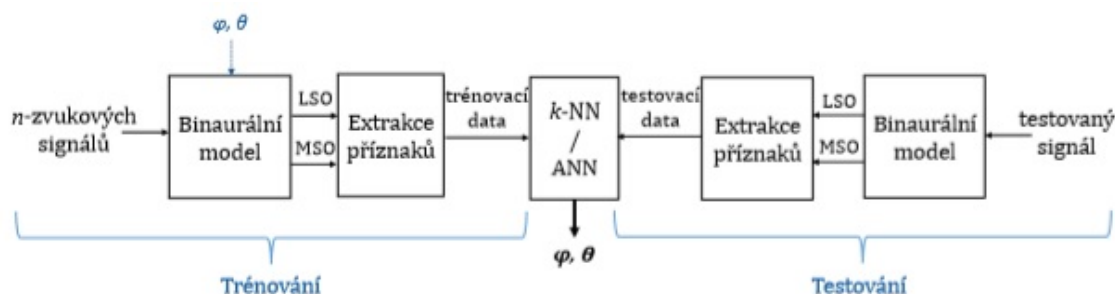
Obrázek č. 4 Etapy iteračního výběru koncových bodů

Jednoduchost metody je na druhé straně vyvážena vznikem chyb zapříčiněných chybami v původních experimentálních hodnotách [5].

4. Realizace experimentu

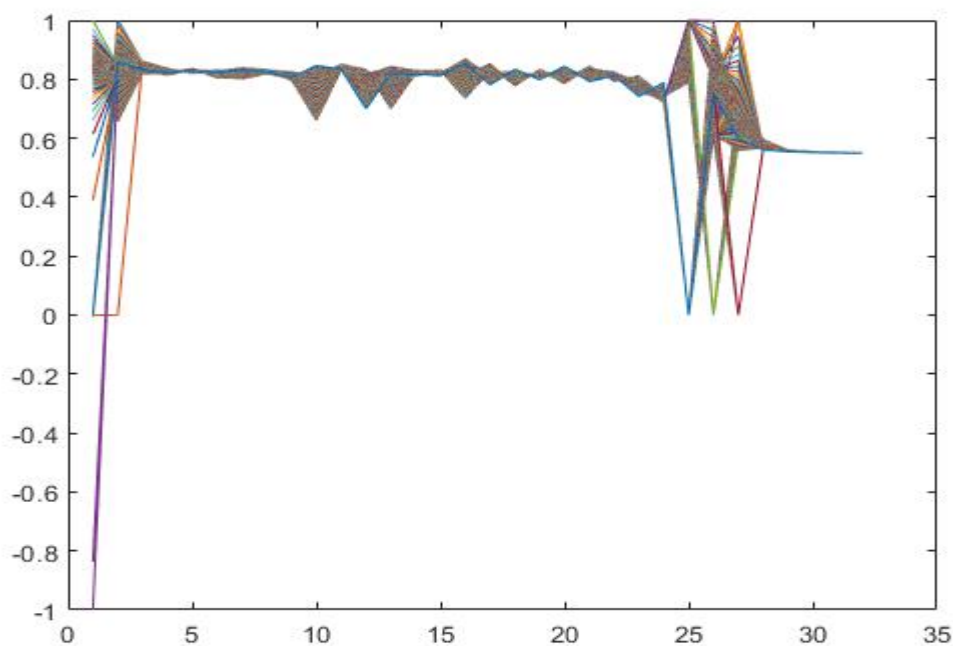
4.1. Výstup binaurálního modelu

Ve své práci [1] se E. Koshkina zabývala parametrizací lateralizačních funkcí na výstupu MSO a LSO pro výslednou analýzu směru příchodu zvuku za pomoci strojového učení. Na obrázku č. 5 je znázorněn algoritmus tohoto experimentu.

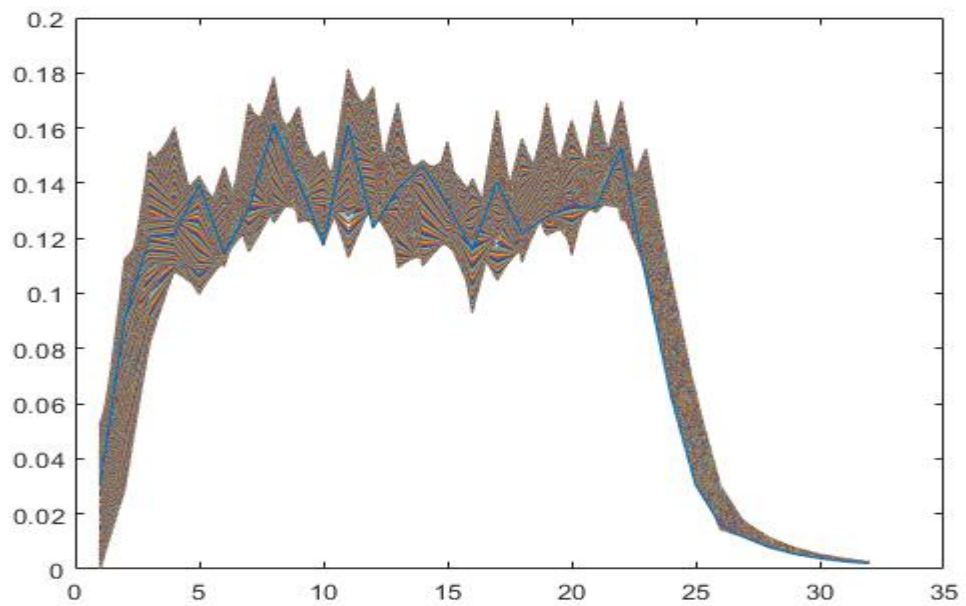


Obrázek č.5 Blokové schéma algoritmu lokalizace zvuku

V první realizaci experimentu [1] se generuje 20 šumových signálů, které pak analyzuje a filtruje binaurální model. Vypočtené hodnoty ITD a ILD jsou vztaženy k frekvenčním pásmům. Před extrakcí parametrů jsou lateralizační funkce segmentovány do 32 částí podle azimutálního úhlu pro přední horizontální polovinu ($< -90^\circ ; 90^\circ >$ s krokem 5°). Výslednou závislost hodnot ITD a ILD na azimutálních segmentech jsem ve svém projektu aproximoval (Obrázek č. 6, Obrázek č.7).



Obrázek č.6 Průběh hodnot ITD pro přední horizontální polovinu



Obrázek č.7 Průběh hodnot ILD pro přední horizontální polorovinu

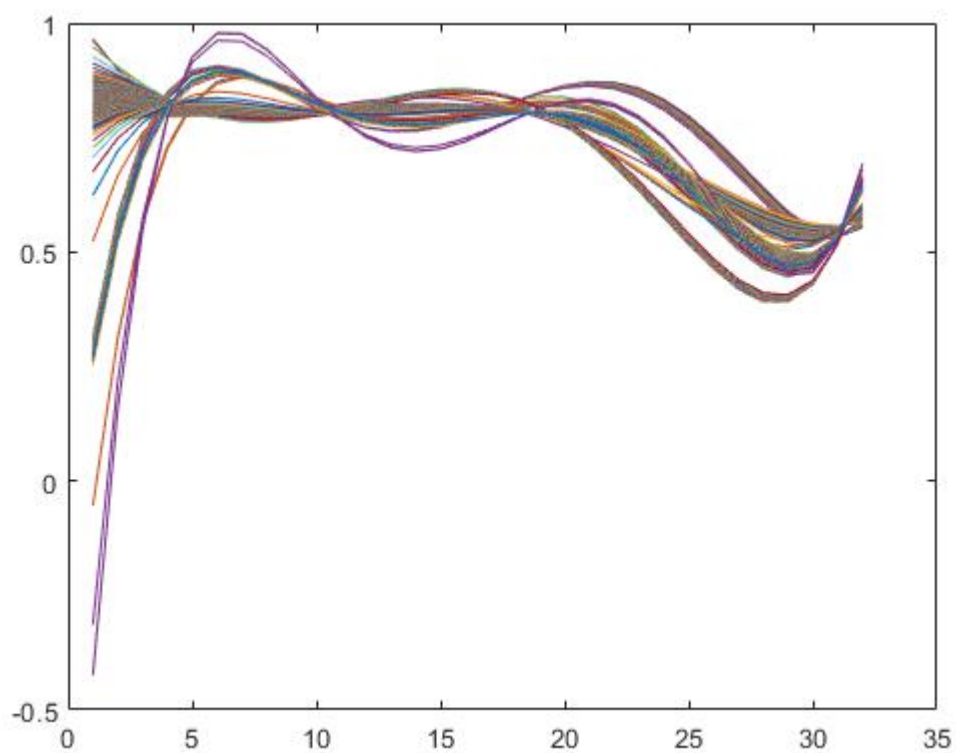
4.2. Aproximace polynomem 5-ého stupně

V dané práci byla použita metoda polynomiální aproximace po částech. V prostředí Matlab jsem vypracoval funkci `aproximace.m`, kde vstupní hodnotou je lateralizační funkce z modelu LSO, resp. MSO, která představuje matici 32 x 512, a výstupními hodnoty jsou: 1) procentuální poměr skutečných hodnot k funkčním hodnotám polynomu, 2) součet čtverců odchylek skutečných hodnot od funkčních hodnot polynomu.

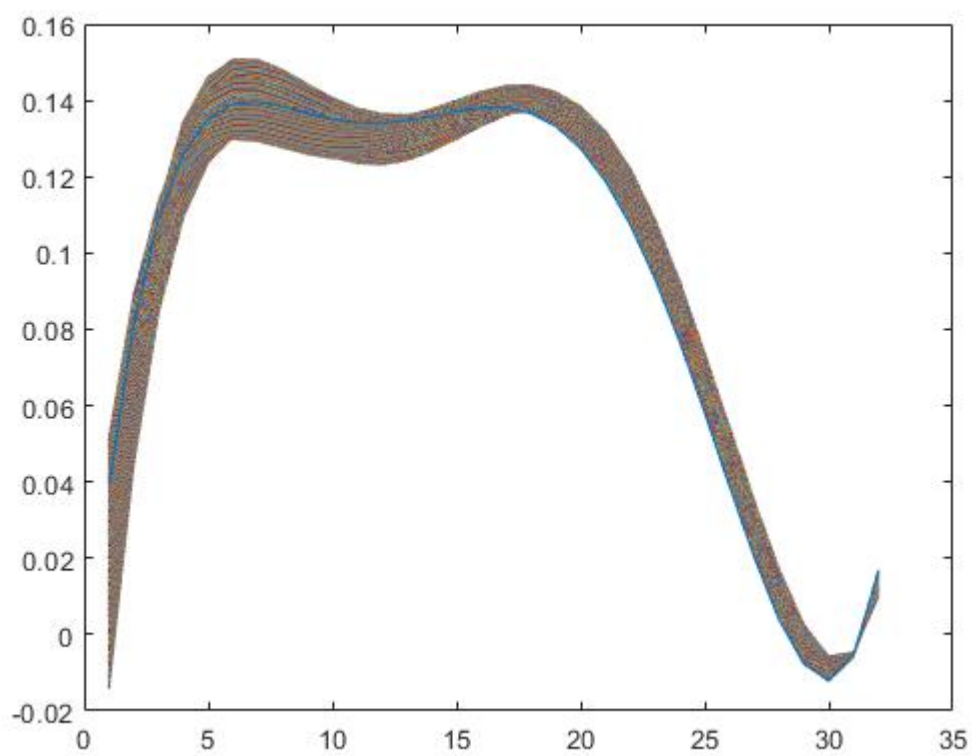
Pro model MSO bylo dosaženo shody 98,81% a součet čtverců odchylek je 0,6795; pro model LSO bylo dosaženo shody 73,61% a součet čtverců odchylek je 0,0097. Tabulka č.1 uvádí jednotlivé koeficienty aproximačních polynomů podle vztahu (3) pro oba modely, a obrázky č. 8 a 9 graficky znázorňují průběh lateralizačních funkcí po aproximaci.

Model\Koeficient	a_0	a_1	a_2	a_3	a_4	a_5
MSO	-0,1107	0,4486	-0,0716	0,0050	-0,0002	0,0000
LSO	-0,0221	0,0721	-0,0118	0,0009	0,0000	0,0000

Tabulka č.1 Koeficienty aproximačních polynomů



Obrázek č.8 Průběh aproximované lateralizační funkce z MSO



Obrázek č.9 Průběh aproximované lateralizační funkce z LSO

4.3. Shrnutí

Dosažený poměr shody tvoří 73,61% pro model LSO a 98,81% pro model MSO. Z porovnání průběhu hodnot do a po aproximaci je vidět, že u aproximací obou modelu po 30 kroku (to odpovídá úhlu 80°-90°) dochází k chybnému nárůstu hodnot ILD a ITD, zatímco skutečné hodnoty při daném úhlu buď jsou konstantní anebo klesají. Přičemž pro model LSO chybný nárůst hodnot na konci je mnohem větší než pro model MSO, což vlastně zčásti mohlo způsobit výsledný rozdíl poměrů shody (25,20%).

Z pohledu na koeficienty aproximačních polynomů je možné předpokládat, že pro model LSO (kde jsou koeficienty aproximačního polynomu 4. a 5. řádu nulové) by byla vhodnější a úspěšnější aproximace polynomem menšího stupně.

V práci [1] strojový klasifikátor na základě parametrizovaných lateralizačních funkcí určoval směr přicházejícího zvuku tím, že jednotlivému úhlu azimutu přiřazoval odpovídající hodnoty. Dosažená úspěšnost přiřazení tvoří 60% pro model LSO a 46% pro model MSO. Když uvedené výsledky porovnáme s aproximačními výsledky, pak můžeme považovat aproximaci za přesnější a proto vhodnější postup v dané problematice.

5. Závěr

Cílem této práce bylo popsat chování lateralizačních funkcí na výstupu modelu binaurálního slyšení pomocí aproximace. Získané výsledky jsem pak měl porovnat s výsledky experimentu lokalizace zdroje zvuku pomocí strojového učení.

Teoretická část práce popisuje problematiku binaurálního slyšení a aproximace, praktická část popisuje implementaci zmíněných teoretických poznatků a zhodnocení výsledků.

Ukázalo se, že aproximace polynomem 5-ého stupně dokáže důvěrohodně popsat závislost hodnot ITD a ILD na azimutálním úhlu (s přesností 98,81% pro výstup MSO, resp. 76,61% pro výstup LSO). Také se ukázalo, že přístup s aproximací udává lepší výsledky, než přístup se strojovým učním (o 13,61% přesnější pro LSO, o 52,81% přesnější pro MSO).

Stanovené cíle byly dosaženy, ale zůstává velký prostor pro prohloubení do daného tématu a další zkoumání, např. zkoušení jiných metod aproximace, nebo aplikace použité metody na jiné zdroje zvuku.

Citovaná literatura

- [1] KOSHKINA, Ekaterina. 2017. Využití strojového učení pro modelování binaurálního slyšení. Diplomová práce. ČVUT v Praze. Vedoucí práce František Rund. 51 s.
- [2] GRAY, Henry. 2000 . Anatomy of the Human Body. New York: Bartleby.com. [online] <http://www.bartleby.com/107/228.html>
- [3] BOUŠE, Jaroslav. 2015. Model of binaural interactions. Doktorandské minimum. Praha: ČVUT FEL, Katedra radioelektroniky.
- [4] LAURENS, Jean-Paul. 1975. Aproksimacija i optimizacija. Monografie. Moskva: Mir
- [5] PRETT, U. 1982. Cifrovaja obrabotka izobraženij. Monografie. Moskva: Mir [online] http://sernam.ru/book_prett2.php?id=81
- [6] BLAUERT, Jens. 1997. Spatial hearing: the psychophysics of human sound localization. MIT press. ISBN 978-0-262-02413.
- [7] MIDDLEBROOKS, John C.; GREEN, David M. 1990. Directional dependence of interaural envelope delays. The Journal of the Acoustical Society of America, 87.5: 21492162
- [8] BOUSE, Jaroslav; VENCOVSKY, Vaclav. 2015. Two-channel models of medial and superior olive based on psychoacoustics. BMC Neuroscience, 16.Suppl 1: P276.

Přílohy

- `aproximace.m`, `spousteni.m` – skripty Matlab připravené během práce
- `segmenty_lso.mat`, `segmenty_mso.mat` – výstupní hodnoty MSO a LSO pro 1. realizaci v [1]