



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta elektrotechnická

## **Syntéza virtuálního akustického prostoru**

### **Synthesis of virtual acoustic space**

Bakalářská práce

Studijní program: Komunikace, multimédia a elektronika

Studijní obor: Multimediální technika

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

**Ondřej Lahoda**

Praha 2017

## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Lahoda** Jméno: **Ondřej** Osobní číslo: **434981**  
Fakulta/ústav: **Fakulta elektrotechnická**  
Zadávající katedra/ústav: **Katedra radioelektroniky**  
Studijní program: **Komunikace, multimédia a elektronika**  
Studijní obor: **Multimediální technika**

## II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

**Syntéza virtuálního akustického prostoru**

Název bakalářské práce anglicky:

**Synthesis of Virtual Acoustic Space**

Pokyny pro vypracování:

Seznamte se s problematikou syntézy virtuálního akustického prostoru. Implementujte vybranou metodu (např. Matlab) vytvoření virtuálního akustického prostoru pro prezentaci prostřednictvím sluchátek. Navrhněte experiment pro ověření věrnosti modelovaného virtuálního prostoru. Experiment realizujte a výsledky vyhodnoťte na základě poslechového testu.

Seznam doporučené literatury:

- [1] Carliele, S. Virtual Auditory Space: Generation and Applications. R. G. Landes Company, 1996
- [2] Zölzer, U. ed. DAFX - Digital Audio Effects (Second Edition), John Wiley & Sons, 2011

Jméno a pracoviště vedoucí(ho) bakalářské práce:

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

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

Datum zadání bakalářské práce: **17.02.2017**

Termín odevzdání bakalářské práce: **26.05.2017**

Platnost zadání bakalářské práce: **31.08.2018**

\_\_\_\_\_  
Podpis vedoucí(ho) práce

\_\_\_\_\_  
Podpis vedoucí(ho) ústavu/katedry

\_\_\_\_\_  
Podpis děkana(ky)

## III. PŘEVZETÍ ZADÁNÍ

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

\_\_\_\_\_  
Datum převzetí zadání

\_\_\_\_\_  
Podpis studenta

# Poděkování

Rád bych poděkoval Ing. Františku Rundovi, Ph.D. za cenné rady, vzácný čas a věcné poznámky během psaní této práce. Za možnost získání vlastních HRTF na měřicí sestavě děkuji Ing. Ondřeji Šupkovi za její zhotovení. Dále děkuji spolužákům a pedagogům Konzervatoře Jaroslava Ježka za odbornou účast v poslechovém testu. Nakonec bych chtěl poděkovat svým rodičům za podporu a motivaci během studií.

## Prohlášení

Prohlašuji, že jsem předloženou bakalářskou práci vypracoval samostatně a že jsem uvedl veškeré informační zdroje v seznamu literatury. Práce je napsána v souladu s Metodickým pokynem o etické přípravě vysokoškolských závěrečných prací.

V Praze dne 26. května 2017

Ondřej Lahoda .....

# Abstrakt

Tato bakalářská práce se zabývá virtuálním akustickým prostorem a jeho syntézou. Nejprve jsou vysvětleny principy lokalizace zvukového zdroje v prostoru. Dále je popsána Head Related Transfer Function, její využití pro syntézu virtuálního akustického prostoru a měření HRTF binaurálními mikrofony. Syntetizovaným modelem akustického prostoru je bicí souprava z pohledu hráče. Syntéza tohoto modelu je implementována v prostředí Matlab. Výsledky jsou vyhodnoceny subjektivními poslechovými testy.

## Klíčová slova

HRTF, binaurální mikrofony, audio syntéza, bicí souprava, subjektivní poslechové testy

# Abstract

The topic of this thesis is virtual acoustic space synthesis. There is a brief introduction to methods of sound localization principles in human ear. The Head Related Transfer Function is explained and used for own method of space synthesis in Matlab. Binaural microphones are used for reference recordings and measurements of HRTF. The practical model of virtual acoustic environment is a drumkit recorded and synthesised from the player's own perspective. The results of experiments are evaluated by subjective listening test.

## Key words

HRTF, binaural microphones, audio synthesis, drumkit, subjective listening test

# Obsah

1	Úvod.....	4
2	Akustický prostor.....	5
2.1	Lokalizace.....	5
2.1.1	Horizontální rovina .....	5
2.1.2	Vertikální rovina .....	6
2.1.3	Vjem vzdáleností.....	6
2.2	Head-Related Transfer Function .....	6
3	Syntéza.....	7
3.1	Binaurální syntéza .....	7
3.2	Konvoluce .....	7
3.3	Prostorová akustika .....	8
3.3.1	Model místnosti.....	8
3.3.2	Odražená intenzita.....	8
3.3.3	Prvotní odrazy .....	9
4	Psychoakustika.....	11
4.1	Vizuální vjem .....	11
4.2	Statistické zpracování .....	11
5	Binaurální mikrofony.....	12
5.1	Binaurální nahrávky .....	12
5.2	Použité binaurální mikrofony.....	12
5.2.1	Roland CS-10EM.....	12
5.3	Pozice mikrofonů v uchu.....	14
6	Bicí souprava .....	15
6.1	Hi-hat.....	16
6.2	Snare drum .....	16
6.3	Bass drum .....	17
6.4	Floor tom .....	17

6.5	China .....	17
7	Vlastní virtuální akustický prostor.....	18
7.1	Nahrání zvukových vzorků.....	18
7.2	HRTF měření a databáze .....	18
7.3	Využití systému HRTFmeas .....	19
7.4	Model místnosti.....	20
7.5	Zpracování signálu .....	20
7.6	Syntéza místnosti.....	22
7.7	Grafické rozhraní.....	22
8	Poslechový test.....	24
8.1	Cíl poslechového testu.....	24
8.2	Testovaná skupina .....	25
8.3	Technické parametry poslechového testu.....	25
8.4	Příprava a instrukce testu.....	26
8.5	Průběh testu .....	26
8.5.1	Určení azimutu .....	26
8.5.2	Určení elevace.....	27
8.5.3	Určení vzdálenosti.....	27
8.6	Výsledky testu .....	28
8.7	Reprezentace výsledků.....	30
9	Závěr .....	31
10	Zdroje.....	32
	Použité zkratky.....	34
	Příloha.....	34

# 1 Úvod

Práce se zabývá virtuálním akustickým prostorem a jeho syntézou. Pod tímto pojmem je myšleno vytvoření zvukových nahrávek se zaznamenaným nebo modelovaným prostorovým vjemem. Různé formy virtuální reality jsou velmi aktuální téma, které využívá technologických novinek a rozšiřuje možnosti lidského vnímání různých přístrojů. Tato práce je tedy jedním z mnoha přístupů k virtuální realitě, konkrétně v oblasti hudby a jejího záznamu a reprodukce.

V úvodní části jsou popsány základní principy lokalizace zdrojů zvuku a funkce, která tyto principy reprezentuje, přenosová funkce hlavy Head-Related Transfer Function. Tato funkce je základním stavebním kamenem celého experimentu. Využití HRTF pro syntézu virtuálního akustického prostoru je vysvětleno podrobněji v samostatné kapitole, ve které je také zpracován vliv prvotních odrazů v uzavřené místnosti na tvar a využití této funkce. Vzhledem k vyhodnocení praktického experimentu subjektivními poslechovými testy je v teoretickém úvodu krátký úvod do oboru psychoakustiky.

V praktické části je popsán způsob záznamu binaurálními mikrofony a jejich technické parametry. Binaurálními mikrofony jsou také měřené vlastní HRTF pomocí sestavy Ing. Ondřeje Šupky [1].

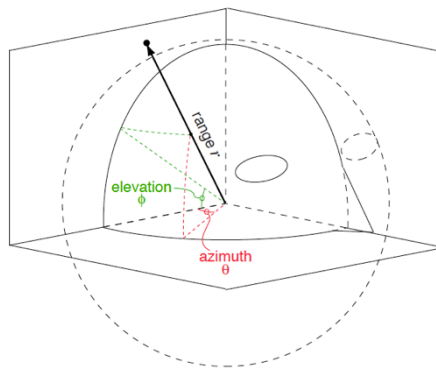
Pro model virtuálního akustického prostoru je zvolena bicí souprava z pohledu hráče. Její výhodou je různorodé umístění jednotlivých prvků v prostoru. Reálná bicí souprava je zaznamenána binaurálními mikrofony a klasickou metodou, která neobsahuje informaci o pozici v prostoru. Tyto nahrávky jsou pak použity pro syntézu prostorového vnímání, jejíž přehrávání je implementováno intuitivním grafickým rozhraním. V poslední kapitole jsou hodnoceny výsledné nahrávky metodou subjektivního poslechového testu.



## 2 Akustický prostor

### 2.1 Lokalizace

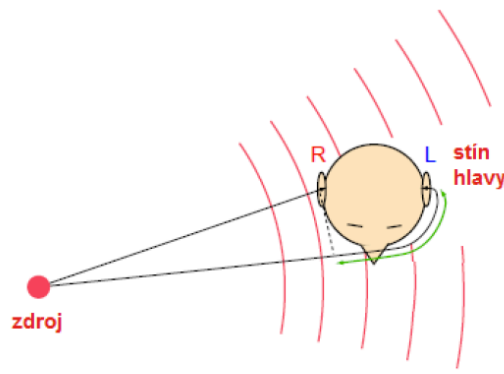
Při popsání lokalizace zdroje zvuku v prostoru si zvolíme souřadný systém, který ukazuje Obrázek 1. Rozlišíme tedy horizontální rovinu s azimutem  $\theta$ , vertikální rovinu s elevací  $\phi$  a vzdálenost  $r$ .



Obrázek 1: Souřadný systém [5]

#### 2.1.1 Horizontální rovina

V horizontální rovině kolem hlavy posluchače lze polohu zdroje zvuku určit z rozdílů signálů levého a pravého ucha. Při výchylce zdroje v horizontální rovině dochází k časovému zpoždění mezi totožnými signály a tím pádem k pozorovanému fázovému rozdílu. Časové zpoždění se nazývá Interaural Time Difference (ITD). Vlivem rozdílného útlumu mají signály také jinou amplitudu tedy parametr Interaural Level Difference (ILD). Cestu signálů popisuje Obrázek 2.



Obrázek 2: ITD a ILD [5], překlad [17]

### 2.1.2 Vertikální rovina

Ve vertikální rovině provádí lidské ucho frekvenční filtraci. Před zpracováním signálu ve vnitřním prostoru lidského sluchového ústrojí dochází k několikanásobným odrazům zvuku. Signál se odráží od lidského torza, hlavy a především od boltce vnějšího ucha. Odražené a přímé signály způsobují hřebenový frekvenční filtr, díky kterému umí člověk lokalizovat zdroje ve vertikální rovině. Kvalita lokalizace zvuku také záleží na spektrálním složení původního signálu. Pokud je signál širokopásmový s dostatečnou energií vyšších frekvencí, je lokalizace snadnější. Z toho plyne, že lokalizace pro některé reálné zdroje zvuku s úzkým frekvenčním pásmem může být obtížnější. Tvar částí lidského těla značně personalizuje prostorové vnímání každého jedince, to je jeden z největších problémů následného zpracování a experimentů.

### 2.1.3 Vjem vzdálenosti

Vzdálenost zdroje zvuku lze zjednodušeně stanovit z amplitudy signálu, musíme ale vědět nebo umět odhadnout původní intenzitu v místě zdroje, tento princip lze tedy spíše aplikovat v simulacích, než v reálných situacích. U známých zvuků a prostorů určuje člověk vzdálenost opět z kmitočtového spektra, kdy s rostoucí vzdáleností ubývají vysoké frekvence.

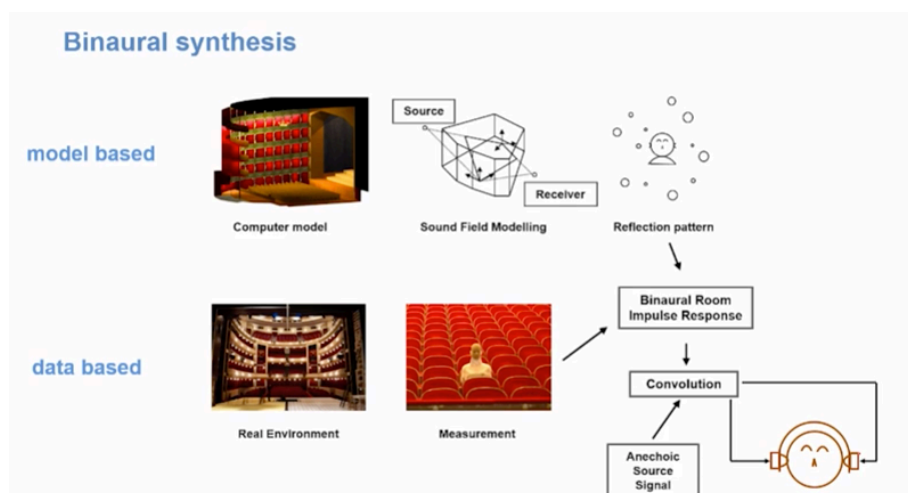
## 2.2 Head-Related Transfer Function

Výše popsané jevy ovlivňující lokalizaci zvuku lze dle [2] popsat přenosovou funkcí lidské hlavy (Head-Related Transfer Function) HRTF. HRTF je obrazem Fourierovy transformace impulsní odezvy lidského ucha na Diracův impuls (Head-Related Impulse Response) HRIR. Impulsní odezvy musí být ideálně měřené v bezodrazové místnosti. Pokud jsou měřené v prostoru s odrazy, zahrnuje HRTF i informaci o místnosti a dle [3] se nazývá BRTF (Binaural Room Transfer Function). Tyto pojmy se v praxi běžně zaměňují, nebo obecně skrývají pod pojmem HRTF.

## 3 Syntéza

### 3.1 Binaurální syntéza

Při modelování akustického prostoru pro následnou reprodukci můžeme zvolit několik přístupů. Schéma, které ukazuje Obrázek 3, zobrazuje různé fáze binaurální syntézy. Rozlišuje dva zásadní rozdíly přístupu: modelový a datový. Modelovaná syntéza nezahrnuje žádná naměřená data a impulsní odezvy je potřeba matematicky namodelovat. Datový přístup využívá změřených impulsních odezvy v reálném prostoru. Poslední krok před reprodukcí syntetizovaného binaurálního signálu je konvoluce zvukového vzorku s HRTF respektive BRTF.



Obrázek 3: Schéma binaurální syntézy [3]

### 3.2 Konvoluce

Konvoluce dvou signálů [4] je matematická operace zpracovávající dvě funkce. Pro diskrétní čas (digitální zpracování) je dána sumou:

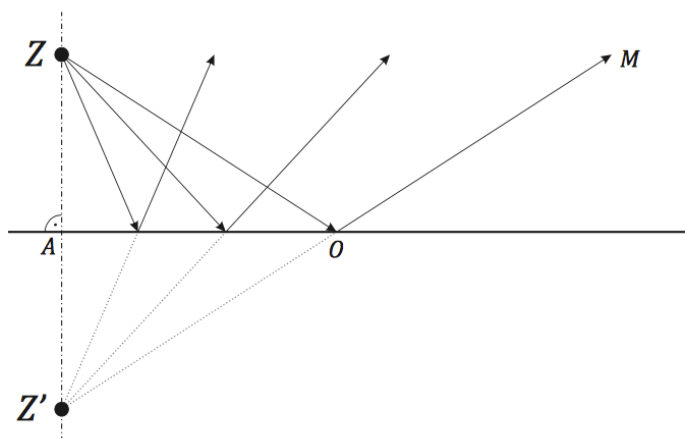
$$y[n] = \sum_{k=-\infty}^{+\infty} x[k]h[n-k] \quad (3.1)$$

Podle [5] konvolucí signálu zdroje s levým a pravým kanálem HRIR umístíme zvuk do přesného místa v prostoru pro následný poslech na sluchátkách. Tento krok také znázorňuje Obrázek 3 jako závěrečný krok binaurální syntézy. Výsledný vjem samozřejmě záleží na použité HRIR.

## 3.3 Prostorová akustika

### 3.3.1 Model místnosti

K šíření zvuku v místnosti a k prostorové akustice lze přistupovat mnoha způsoby. Jedním z běžných a přehledných způsobů modelování odrazů zvuku je princip zrcadlení zdrojů [6]. Metoda zrcadlení zdrojů vychází z geometrické akustiky. Zvukové pole můžeme popisovat pomocí paprsků. Při zanedbatelné vlnové délce vzhledem k rozměrům překážky můžeme využít jednoduchého zákona odrazu: úhel dopadu se rovná úhlu odrazu. Za překážku, od které dojde k odrazení zvukového paprsku, umístíme zdánlivý zdroj v přesně osově souměrné poloze. Metodu zrcadlení zdrojů zobrazuje Obrázek 4. Geometrický přístup je náročný pro řešení komplexnějších problémů. Pro vícenásobné odrazy musíme zrcadlit jak původní zdroj, tak jeho první zdánlivý zdroj. Geometricky lze přistupovat i k odrazům od kulových stěn, veškerá tato problematika je zpracována v [6]. V některých případech lze členitější prostory zjednodušit na rovinné stěny a také lze zanedbat vícenásobné odrazy a pracovat pouze s prvotními. Pokud takové zanedbání nelze provést, je lepší prostor popisovat z pohledu statistické akustiky.



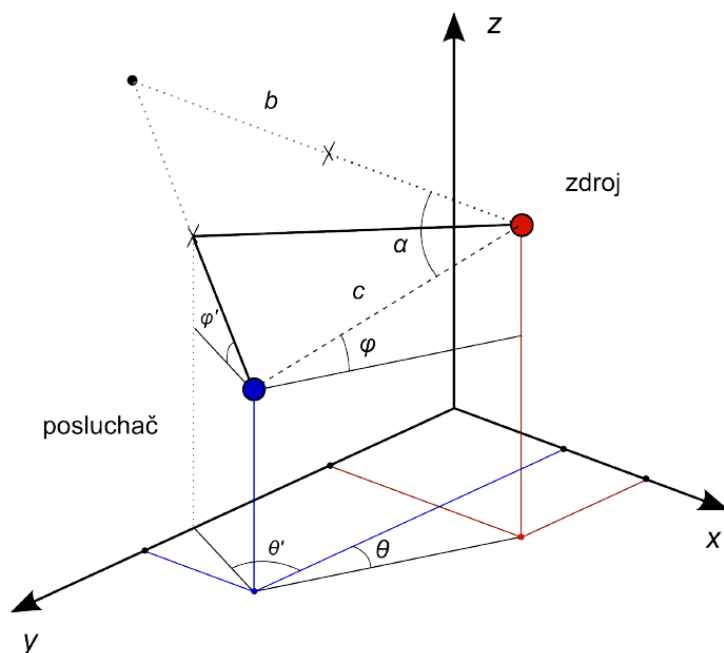
Obrázek 4: Metoda zrcadlení zdrojů [21]

### 3.3.2 Odražená intenzita

Dopadem zvukové vlny na překážku se může část její energie pohltit. Paprsek tak není zcela odražen a pohlcená energie je charakterizována činitelem zvukové pohltivosti  $\alpha$ . Zrcadlový zdroj je pak  $1 - \alpha$  krát slabší [7]. Činitel zvukové pohltivosti je závislý především na materiálu překážky. Při určitém zjednodušení mohou mít některé materiály činitel pohltivosti nulový a odrážet tedy celou energii dopadajícího paprsku. Metody geometrické akustiky se používají k předběžným odhadům zvukového pole, ale i v numerických výpočtech akustiky uzavřeného prostoru.

### 3.3.3 Prvotní odrazy

Při popisu prostoru souřadnou soustavou z kapitoly 2 musíme najít azimut a elevaci pro prvotní odrazy metody zrcadlení zdrojů. Popis prostoru ukazuje Obrázek 5.



Obrázek 5: Prvotní odrazy [8]

Dle [8] můžeme vzdálenost uraženou paprskem prvotního odrazu vyjádřit kosinovou větou.

$$d_n = \sqrt{b^2 + c^2 - 2bc \cos \alpha_n} \quad (3.2)$$

Kde  $b$  je vzdálenost reálného a zdánlivého zdroje,  $c$  je vzdálenost reálného zdroje od posluchače a  $\alpha_n$  je úhel, který svírají tyto strany dle obrázku (Obrázek 5). Pro odraz od přední stěny můžeme tento úhel získat z azimutu zdroje vůči posluchači.

$$\alpha = \begin{cases} 180^\circ - \theta & \text{pro } \theta < 180^\circ \\ \theta - 180^\circ & \text{pro } \theta \geq 180^\circ \end{cases} \quad (3.3)$$

Ze sinové věty poté určíme azimut prvotního odrazu.

$$\theta_p = \begin{cases} \sin^{-1}\left(\frac{\sin(\alpha)c}{d_n}\right) & \text{pro } \theta < 180^\circ \\ 360^\circ - \sin^{-1}\left(\frac{\sin(\alpha)c}{d_n}\right) & \text{pro } \theta \geq 180^\circ \end{cases} \quad (3.4)$$

Výpočet elevace prvotního odrazu získáme z pravoúhlých trojúhelníků se stejnou absolutní výškou reálného i zdánlivého zdroje od poloroviny nulové elevace.

$$\phi_p = \sin^{-1}\left(\frac{\sin(\phi)c}{d_n}\right) \quad (3.5)$$

Odražený paprsek dorazí do místa posluchače zpožděný o čas  $t$ . Čas zpoždění lze jednoduše určit z rozdílu dráhy přímého paprsku a paprsku prvotního odrazu a z rychlosti šíření zvuku ve vzduchu.

$$t = \frac{d_n - c}{v} \quad (3.6)$$

Kde  $d_n$  je vzdálenost uražená odraženým paprskem,  $c$  je vzdálenost zdroje a posluchače a  $v$  je rychlost šíření zvuku ve vzduchu s hodnotou přibližně  $330 \text{ ms}^{-1}$ .

## 4 Psychoakustika

Psychoakustika je věda zabývající se vnímáním a reakcemi člověka na akustické podněty. Má velmi rozšířenou působnost a z důvodu často abstraktních pojmů a definic obsahuje mnoho přístupů k řešení problému. Hodnocení v rámci psychoakustiky může být jak subjektivní, tak objektivní. Objektivní hodnocení vjemů člověka je pro některé případy ale neřešitelný problém. Subjektivní testy jsou naopak zatíženy individuálními úsudky jednotlivých respondentů. Cílem poslechového testu je hodnocení kvality zvuku a nalezení vazby mezi psychoakustickými veličinami a kroky syntézy, které lze ovlivnit [7]. Hodnocení slovním popisem je velmi náročné zpracovat, ale poskytuje vhodné termíny pro charakteristiku veličiny. Doplnění o hodnocení na definované škále umožňuje vhodnější zpracování. Při vyhodnocení testů je třeba lidské vjemy převést na definované škály, které je možné dále statisticky zpracovat. Škály mohou být definované různým způsobem a jejich podrobné dělení je v [9]. Rozlišují se především z hlediska dalšího zpracování. Pokud převádí psychologické vjemy na jasně definovanou stupnici s daným intervalem, můžeme například vyhodnotit celkový průměr se směrodatnou odchylkou.

### 4.1 Vizualní vjem

Pro lokalizaci zdroje zvuku není pro lidský mozek akustický signál prioritou. U zdrojů, které lze lokalizovat vizuálně, převládá tato informace nad informací akustickou [10]. Při některých experimentech je snaha oba vjemy oddělit. Ve virtuální realitě se naopak nedostatky simulovaného vizuálního vjemu snažíme doplnit sluchovým vjemem správně reprezentujícím zvolený prostor.

### 4.2 Statistické zpracování

Podle zvolené stupnice hodnocení subjektivních testů jsou různé možnosti statistického zpracování výsledků poslechového testu. Pokud má stupnice jasně definované vyjádření hodnocení číslem, můžeme z individuálních hodnot vypočítat aritmetický průměr. Pro ten lze určit směrodatnou odchylku.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (4.1)$$

## 5 Binaurální mikrofony

### 5.1 Binaurální nahrávky

Prostorový záznam zvuku můžeme pořídít přímo bez syntézy za použití binaurálních mikrofonů. Pojem binaurálních nahrávek nelze zaměňovat s klasickým stereofonním záznamem. Ačkoli se také jedná o dvoukanalový záznam zvuku, binaurální nahrávka je pořízena mikrofony umístěnými stejně jako lidské uši. V některých případech se jedná o umělou hlavu s vestavěnými mikrofony přímo v ušních kanálcích (např. Neumann KU-81). Jednodušší variantou jsou miniaturní mikrofony připomínající sluchátka, umístěné na figuríně nebo přímo na lidské hlavě. Výsledky všech binaurálních nahrávek značně ovlivňuje tvar uší a hlavy, na které byly pořízeny. Průměrné rozměry umělých hlav mohou zajistit přibližnou funkčnost pro více uživatelů. Nahrávky přímo na konkrétní osobě zaručí nejpřesnější vjem pro daného posluchače. V takovém případě hovoříme o personalizované HRTF. K personalizované HRTF můžeme dojít i modelováním, vždy jsou ale potřeba parametry konkrétní lidské hlavy.

### 5.2 Použité binaurální mikrofony

V rámci semestrálního projektu jsem testoval dostupné binaurální mikrofony katedry radioelektroniky FEL ČVUT. Díky testu jsem znal parametry různých párů a možnost jejich využití. Pro pořízení referenčních nahrávek binaurálními mikrofony jsem vybral mikrofony Roland CS10EM. Při výběru jsem zohledňoval i další vlastnosti mikrofonů, které z výsledku testování nebyly patrné. Jako důležitý parametr jsem nakonec zvolil možnosti praktického využití mikrofonů. Mikrofony Roland CS-10EM jako jediné nabízí zároveň funkci sluchátek, což je pro praktické využití velmi vhodné. Ostatními parametry mikrofonů zároveň nijak negativně neovlivňují výsledky měření.

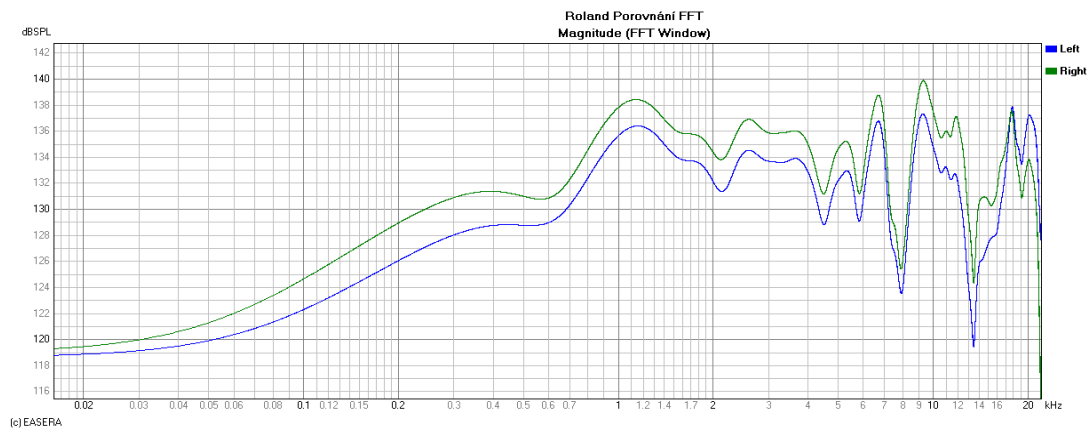
#### 5.2.1 Roland CS-10EM



Obrázek 6: Binaurální mikrofony Roland CS-10EM [roland.com]



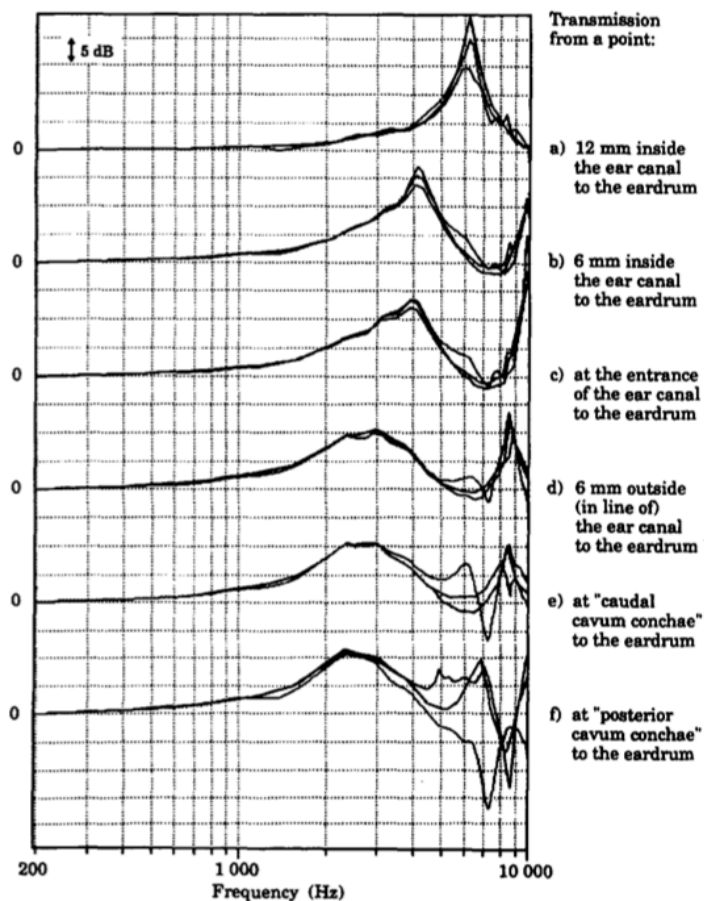
Roland CS-10EM jsou mikrofony spojené s funkcí sluchátek. Jedná se o elektretové kondenzátorové mikrofony, které ke své funkci potřebují fantomové napájení 48 V. Jelikož byly mikrofony využity i pro měření HRTF, byl na ně kladen požadavek vyrovnané frekvenční charakteristiky. Orientačně změřenou frekvenční charakteristiku představuje Obrázek 7. Zvlnění grafu zobrazuje především charakteristiku měřicího reproduktoru, z grafu ale můžeme pozorovat vyrovnanou závislost obou mikrofonů v páru. Drobná odchylka je viditelná v okolí 14 kHz. Konstrukce mikrofonů bohužel neumožňuje ideální umístění membrány na okraj ušního kanálu, tento problém ovlivňuje právě vyšší kmitočty, proto rozdíl charakteristik mikrofonů v okolí 14 kHz neovlivní výsledky více, než ostatní chyby.



Obrázek 7: Frekvenční charakteristika mikrofonů Roland CS-10EM

### 5.3 Pozice mikrofonů v uchu

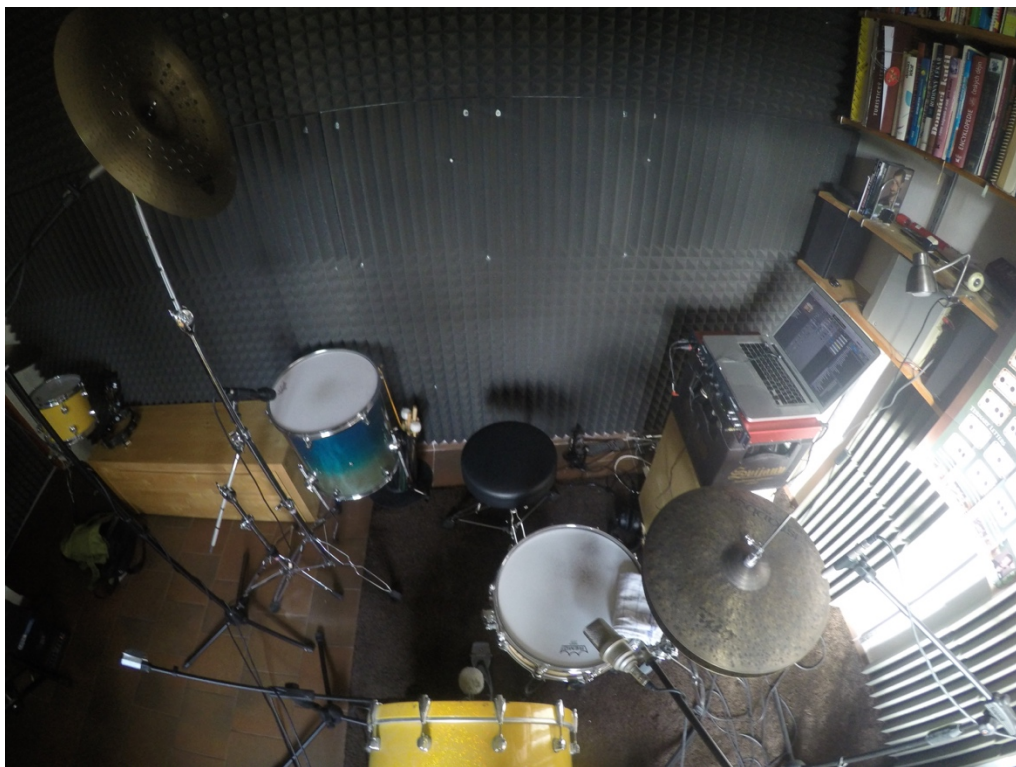
Dle frekvenčních závislostí různých umístění mikrofonů v uchu (Obrázek 8), může umístění mikrofonů vně ušního kanálu vykazovat změny pro frekvence nad 5 kHz. Pro zvukové zdroje obsahující vyšší kmitočty tedy může dojít ke zkreslení výsledného vjemu.



Obrázek 8: Různé pozice mikrofonů v ušním kanálu [20]

## 6 Bicí souprava

Pro reálné ověření experimentů a teorie jsem jako model akustického prostoru zvolil bicí soupravu. Bicí souprava se skládá z mnoha rozličných zdrojů zvuku umístěných do tradici daných poloh. Pro experiment prostorového vnímání je velmi výhodná vzhledem k extrémním polohám zdrojů zvuku vůči hlavě hráče. Pro zpracování byly vybrány hlavní části bicí soupravy a prvky v krajních polohách pro rozdílné lokalizace zdroje zvuku. Bicí souprava, která byla použita při nahrávání je na fotografii (Obrázek 9). Tabulka 1 obsahuje pozici jednotlivých prvků v souřadném systému zvoleném dle odstavce 2.1 se středem hlavy hráče, na které byly umístěny binaurální mikrofony. Azimut nabývá hodnot v intervalu ( $0^\circ$ ;  $360^\circ$ ). Nulová a  $360^\circ$  hodnota popisuje polohu přímo před hráčem. Hodnota  $90^\circ$  reprezentuje zvuk vpravo od hráče, hodnota  $180^\circ$  přímo za zády a  $270^\circ$  je zvuk vlevo od hráče. Elevace teoreticky nabývá hodnot ( $-90^\circ$ ,  $90^\circ$ ). Nulová hodnota elevace je pro horizontální rovinu, kladný směr je vzhůru a záporný směrem dolů. Dále budou popsány vybrané nahrané prvky bicí soupravy.



Obrázek 9: Bicí souprava při experimentu

Tabulka 1: Rozmístění prvků v souřadné soustavě.

	Azimut $\theta$ [°]	Elevace $\phi$ [°]	Vzdálenost $c$ [cm]
Hi hat	315	-45	85
Snare drum	345	-45	85
Bass drum	15	-60	135
Floor tom	120	-45	102
China	90	30	75

## 6.1 Hi-hat

Činely hi-hat jsou tvořeny párem činelů ve vodorovné poloze. Tradičně jsou umístěny v levé části soupravy z pohledu hráče. V souřadné soustavě zvolené dle kapitoly 0 byla naměřena poloha činelů ve vzdálenosti 85 cm v azimutu 315° a s elevací -45° (Tabulka 1). Hi-hat má ostrou barvu zvuku s velmi krátkým dozvukem. Frekvenční charakteristika je výrazná hlavně pro vyšší kmitočty. Pro záznam zvukového vzorku nablízko byl použit mikrofon Rode NT5 umístěný v těsné blízkosti horního činelu. Mikrofon Rode NT5 je kondenzátorový tužkový mikrofon s pŕpalcovou membránou. Má kardioidní směrovou charakteristiku a je vhodný především pro záznam činelů nebo celkového zvuku nástroje jako takzvaný overhead mikrofon.

## 6.2 Snare drum

Pochodový buben, hovorově virbl, je centrálním prvkem soupravy. Často je využíván i jako samostatný nástroj. Umístěn je výhradně přímo před hráčem ve výšce pasu. Toto umístění je ve zvolené souřadné soustavě pro hodnoty azimutu 345° (mírná výchylka vlevo), elevace -45° a ve vzdálenosti 85 cm. Na spodní bláně má připevněné strunění, které mu dává jeho charakteristický bzučivý zvuk. Frekvenčně se nachází ve středním pásmu. Pochodový buben byl nahrán mikrofonem Joemeek JM37dp, což je velkomembránový kondenzátorový mikrofon s kardioidní směrovou charakteristikou a disponuje přepínatelným útlumem -10 a -20 dB. Mikrofon byl umístěn u horní úderové blány bubnu. Kvůli vysoké intenzitě hladiny akustického tlaku bicích nástrojů obecně, zvláště pochodového bubínku, bylo vhodné využití toho mikrofonu s útlumem 20 dB.

## 6.3 Bass drum

Basový buben je největším a zároveň nejhlouběji znějícím bubnem soupravy. S pochodovým bubnem tvoří centrální pár soupravy. Obvykle je položen na zemi ve vodorovné pozici, úderové blány bubnu jsou tedy kolmo k podlaze a buben lze rozeznávat pedálem ovládaným nohou hráče. Kvůli technice hry pedálem je umístěn přímo před hráče. Toto umístění ve zvolené souřadné soustavě reprezentují hodnoty azimutu  $15^\circ$  (mírná výchylka vpravo), elevace  $-60^\circ$  (ze spodu) a vzdálenost 135 cm. Frekvenčně se buben nachází v nízkém pásmu s vrcholem okolo 300 Hz. Zvukový vzorek nablízko byl zaznamenán mikrofonem Akg D112, což je dynamický mikrofon s robustní membránou. Mikrofon je klasicky používán na nástroje s nízkou frekvencí, jelikož má lehce zvýrazněnou frekvenční charakteristiku právě pro nízké kmitočty.

## 6.4 Floor tom

Floor tom, hovorově kotel, je buben postavený na nožičkách v pravé části bicí soupravy. V závislosti na rozměru se pohybuje v nízké až střední části kmitočtové charakteristiky. Pro experiment byl buben umístěn do vzdálenější polohy než je obvyklé, dostal se tak jako jediný prvek soupravy do zadní poloviny souřadného systému, což je pro porovnání prostorové lokalizace žádoucí. Toto umístění představují hodnoty azimutu  $120^\circ$  (více než  $90^\circ$  - zadní polovina), elevace  $-45^\circ$  (stejně jako pochodový buben) a vzdálenost 102 cm. Floor tom byl snímán mikrofonem Akg P5, což je standardní dynamický mikrofon užívaný pro širokou škálu zvuků. Floor tom svým kmitočtem spadá do tohoto průměru pro zvolený mikrofon.

## 6.5 China

Činel typu china je efektním prvkem bicí soupravy. Má tříštivý zvukový charakter s dlouhým dozvukem. Pro experiment jsem jeho polohu zvolil vysoko nad pravým ramenem hráče, tím je jako jediný prvek v rovině kladných hodnot elevace. Jeho umístění je popsáno hodnotami azimutu  $90^\circ$  (přímo vpravo), elevace  $30^\circ$  a vzdálenosti 75 cm. Stejně jako činely hi-hat byl zaznamenán mikrofonem Rode NT5 vhodným pro záznam činelů.

## 7 Vlastní virtuální akustický prostor

### 7.1 Nahrání zvukových vzorků

Pro ideální porovnání při poslechovém testu byly pořízeny ve stejný čas totožné nahrávky binaurálními mikrofony a blízkými mikrofony. Nahráváno bylo přes zvukovou kartu Focusrite Scarlett 18i20 do software Logic Pro X. Zvuková karta disponuje 8-mi vstupními kanály. Ty byly tedy obsazeny binaurálním párem mikrofonů Roland, párem mikrofonů Rode NT5 pro činely, mikrofonem Joemeek pro virbl a mikrofony Akg D112 a P5 pro basový buben respektive floor tom. Pro binaurální pár byla pečlivě nastavena stejná úroveň zesílení předzesilovače, ostatní vstupy byly nastaveny pro dostatečně silný signál. Použité vzorky byly exportovány do formátu .wav vzorkovací frekvencí 44,1 kHz při bitové hloubce 24 bitů pro následné zpracování v Matlabu. V software Logic nebyly jednotlivé stopy nijak upravovány žádnými efekty. Stopy nahrané napřímo jsou dále zpracovávány při syntéze. Binaurální nahrávky slouží jako reference pro syntézu. Jejich zpracování záleží na požadavcích porovnávání. V některých případech nereprezentují ideální výsledek, kterého chceme dosáhnout. Zpracování binaurálních nahrávek pomocí běžných efektů je ale náročné. Dle [11] jsou binaurální nahrávky zabarvené kmitočtovou závislostí HRTF. Některé mikrofony s umělou hlavou (Neumann KU-100) toto zabarvení kompenzují vlastní ekvalizací.

### 7.2 HRTF měření a databáze

Pro zpracování signálu a následnou binaurální syntézu je potřeba získat HRTF. Dle teorie je možné HRTF modelovat, tato cesta je ale velmi náročná, protože je potřeba zahrnout všechny drobné odrazy a vlastnosti lidského ucha respektive celého těla. Jednodušší, ale neméně náročná cesta, je měření reálných HRTF. Pro obecné využití k syntéze libovolného místa v prostoru je zapotřebí HRTF pro každý možný směr v prostoru. Měření se obvykle soustřeďuje na dostatečné rozlišení v horizontální rovině s ideálním krokem alespoň  $5^\circ$ . Ve vertikální rovině je dostačující krok  $10^\circ$  [1]. Z těchto požadavků vyplývá velká časová náročnost získání HRTF pro různé směry. Z tohoto důvodu vznikají databáze již naměřených HRTF, například [12]. V databázích obvykle nemusíme najít danou HRTF pro požadovaný směr. Pro získání požadovaného směru můžeme přistoupit k metodám interpolace nebo extrapolace pro chybějící hodnoty. Při dopočítávání nenaměřených hodnot je ale potřeba postupovat velmi opatrně a sofistikovaně. Ve většině případů je výhodnější se spokojit s rozlišením dostupné databáze a dopustit se chyby pro polohu původního signálu. Z těchto okolností a z personalizované závislosti HRTF vyplývá značné omezení dostupných databází.

### 7.3 Využití systému HRTFmeas

Pro získání vlastních HRTF jsem použil systém [1], který je součástí diplomové práce Ing. Ondřeje Šupky. Soustava se skládá z otočné židle a šesti reproduktorů s různou elevací v rozsahu  $-40^\circ$  až  $60^\circ$  s krokem  $10^\circ$ . Soustava je zhotovena pro obecné měření HRTF a je ovládána z prostředí Matlab. Ovládání soustavy je uživatelsky velmi přívětivé a obsahuje i program pro analýzu a zpracování naměřených dat. Výstupem jsou tedy naměřené a vhodně zpracované HRTF ze zvolených směrů v struktuře zpracovatelné Matlabem. Díky dostatečné drobnosti kroků systému [1] lze získat HRTF z přesně požadovaných míst s odchylkou menší, než je schopnost popisu lokalizace zvuku člověkem. Při měření HRTF se obvykle kompenzuje frekvenční charakteristika měřicích mikrofonů. Pro získání impulsních odezev na soustavě HRTFmeas byly použity stejné binaurální mikrofony Roland CS-10EM jako pro referenční nahrávky. Kompenzace frekvenční charakteristiky, která je často problémová, není tedy nutná a naopak je stejné frekvenční zkreslení požadované. Na soustavě byla provedena dvě měření. Poprvé byly změřeny HRTF ze směrů reálných pozic zdrojů zvuku uložené ve struktuře `zmereneHRTF.mat`. Dodatečně byly změřeny HRTF pro směry prvotních odrazů uložené ve struktuře `odrazy.mat`. Systém nabízí několik metod měření. Vzhledem k malému počtu měřených dat nebylo nutné využít pokročilou metodu měření překrývaným sweepem [1]. Měření tedy probíhalo metodou Sine sweep s nepozměněným základním nastavením systému, tedy s délkou sekvence  $\frac{2^{14}}{f_s}$ , dobou trvání impulsu 0,3 s a dobou trvání druhé harmonické impulsu 0,03 s. Díky možnosti spouštění měření s dostatečným zpožděním pro nasazení mikrofonů a nastavení do měřicí polohy je možné samostatně měřit osobní HRTF bez další výpomoci. Použité HRTF jsou tedy personalizované na mou vlastní hlavu (autor: Ondřej Lahoda).

## 7.4 Model místnosti

Místnost, ve které byly pořízeny nahrávky bicí soupravy, má obdélníkový půdorys o rozměrech 3 x 5 metrů. Je částečně uzpůsobena pro nahrávání, především bicí soupravy. Většina stěn, podlaha a strop jsou ošetřeny akusticky pohltivým materiálem. Od těchto překážek dochází k prvotním odrazům zvukových paprsků s výrazně menší energií, jak je popsáno v odstavci 3.3.2. Vzhledem ke geometrickému popisu prostoru tyto odrazy zanedbáme. Jediným výrazným odrazivým prvkem místnosti je okno v přední stěně, které bylo úmyslně nezakryté. Sklo je materiál s velmi malým činitelem akustické pohltivosti o hodnotě  $\alpha = 0,027$  pro frekvenci 512 Hz [13]. Odraz energie od něj tedy můžeme považovat roven 1. Stěna je vzhledem k souřadnému systému z kapitoly 0 kolmá na přímkou nulového azimutu. Ze vztahů uvedených v odstavci 3.3.3 získáme postupným dosazením původních hodnot azimutu  $\theta$ , elevace  $\phi$  a vzdálenosti  $c$  každého prvku bicí soupravy nové hodnoty prvotních odrazů azimut  $\theta_p$ , elevace  $\phi_p$  a vzdálenosti  $d_n$ . Ze vztahu 3.6 poté jednoduše zjistíme zpoždění odraženého paprsku vůči přímému signálu. Takto vypočtené hodnoty pro každý prvek soupravy nalezneme v Tabulka 2.

Tabulka 2: Parametry paprsků prvotních odrazů

	Azimut $\theta_p$ [°]	Elevace $\phi_p$ [°]	Vzdálenost $d_n$ [cm]	Zpoždění $t_n$ [ms]
Hi hat	350	-10	345	7
Snare drum	356	-11	319	7
Bass drum	4	-13	532	12
Floor tom	51	-7	556	14
China	8	4	605	16

Odrazy od překážek prostoru zaručují charakteristický dozvuk. Člověk očekává od různých prostorů konkrétní dozvuk získaný postupnými zkušenostmi [7]. Zahnutí dozvuku prvotních odrazů do syntézy virtuálního prostoru přidá výslednému vjemu přirozenost. Z Tabulka 2 je patrné, že odrazy od přední stěny místnosti soustředí paprsky blíže k počátku souřadné soustavy jak z pohledu azimutu, tak z pohledu elevace. Signály odrazů vždy urazí větší dráhu než přímé signály, což zohledňuje vypočtené zpoždění  $t_n$ .

## 7.5 Zpracování signálu

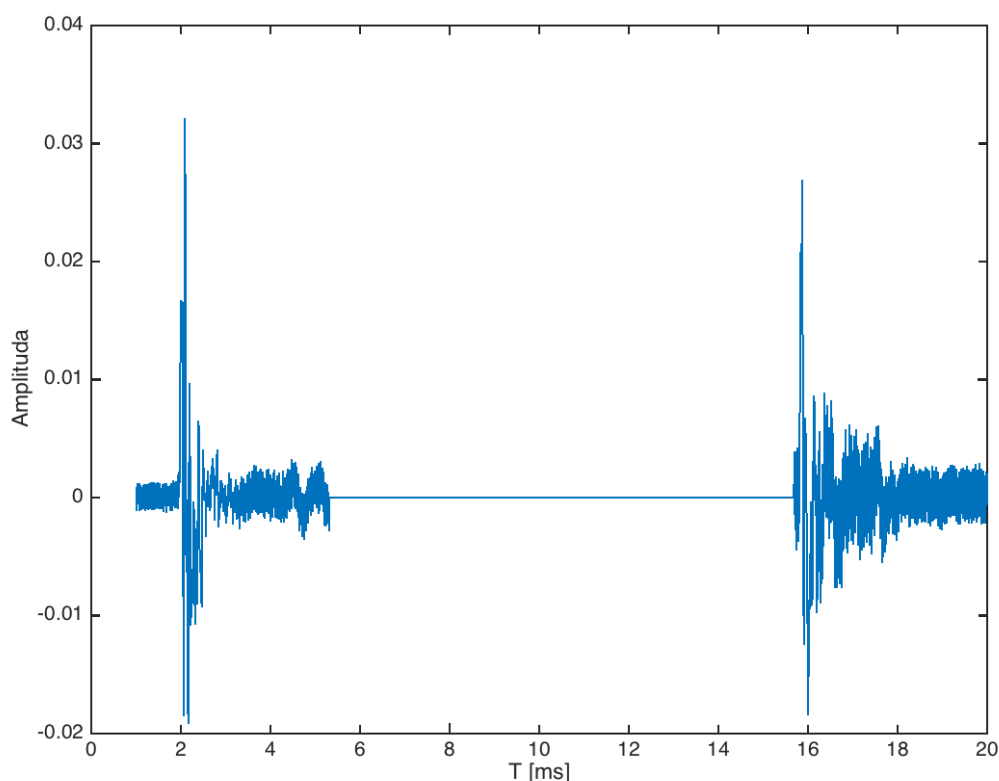
Zpracování nahraných signálů a následná syntéza virtuálního akustického prostoru jsou implementovány v prostředí Matlab. Zvukové vzorky nahrané binaurálními mikrofony slouží jako reference pro dosažení stejného vjemu syntézou. Zvuky nahrané mikrofony nablízko jsou postupně zpracovány k následnému porovnání výsledku. Syntézu prvků získáme z vlastní funkce `SynthHRTF`. Základní prvek zpracování je konvoluce HRTF s pořízeným zvukem. Tato samotná operace by měla zajistit umístění zvuku do daného místa HRTF. Zpracování je však velmi citlivé a potřebuje ošetřit základní problémy. Prvotní testování pořízených vzorků



probíhalo na konvoluci s HRTF z databáze univerzity MIT [12]. Tato databáze ale neobsahuje požadované elevace. V blízkých elevacích neobsahuje ani některé z azimutů, pro prvotní testování jsem tedy zvolil nepřesné hodnoty HRTF pro orientační zpracování. HRTF z databáze byly také pořízeny na jiné binaurální mikrofony než referenční nahrávky. Pro databázi MIT byl použit model lidské hlavy z průměrovaných rozměrů populace tzv. „dummy head“. Nahrávat takovýmto mikrofonem není v principu experimentu možné, překáží totiž reálná hlava hráče v jehož pozici chceme syntézu provádět. Rozdíly v rozměru hlavy způsobily obtížnější lokalizaci pro některé prvky. Pravděpodobně odlišné frekvenční charakteristiky mikrofonů silně změnilly barvu syntetizovaného zvuku oproti referenční nahrávce. Tato změna je však způsobena i dalšími principy syntézy a je těžké ji zcela potlačit. Další zpracování signálů jsem tedy provedl po získání vlastních HRTF pro stejné mikrofony referenční nahrávky. Měření vlastních HRTF je popsáno v kapitole 7.3. Na soustavě Ing. Ondřeje Šupky [1] bylo možné získat HRTF z přesně požadovaných azimutů. Systém je omezen spodní hranicí elevací  $-40^\circ$ . Při uvážení kroku  $10^\circ$  pro elevaci tato hodnota vyhovovala i pro požadované elevace  $-45^\circ$ . Jediná elevace basového bubnu  $-60^\circ$  tedy nebyla možná naměřit správně. HRTF basového bubnu by ideálně bylo možné extrapolovat vhodnou metodou, pro účel této práce to ale není nutné. Získané HRTF byly ze systému zpracovány se vzorkovací frekvencí 96 kHz. Pro konvoluci se zvuky o vzorkovací frekvenci 44,1 kHz je byla potřeba převzorkovat. Po vypočtení poměru vzorkovacích frekvencí šla použít jednoduchá funkce Matlabu `resample`. Dle měření [14] funkce `resample` přidává nežádoucí spektrální čáry jako důsledek aliasingu. V poměru k celkové metodě syntézy tento jev není slyšitelný a funkce `resample` je dostačující.

## 7.6 Syntéza místnosti

Pro zahrnutí prvotních odrazů místnosti popsaných v odstavci 0 je vytvořena funkce `RoomHRTF`. Tato funkce vychází z postupu funkce `SynthHRTF`, ale upravuje původní odezvu přímého signálu. Ve funkci jsou načtena data změřená pro parametry prvotních odrazů z tabulky 2. Tyto HRIR z pozic prvotních odrazů jsou s definovaným zpožděním přičteny k HRIR přímých signálů. Vzhledem k poměrně délce impulsních odezví řádově desetin milisekund a vypočteném zpoždění jednotek milisekund se průběhy nepřekrývají a výsledná HRIR použitá ke konvoluci se skládá z dvou následujících impulsů. První nese informaci přímého signálu a druhý odraženého (Obrázek 10).



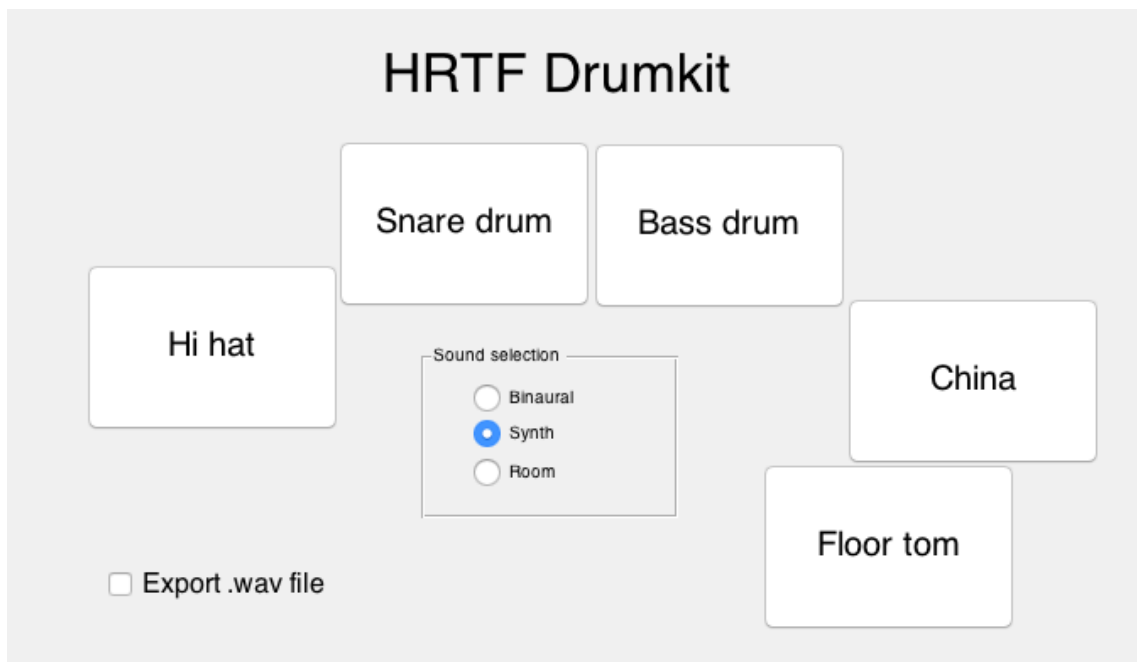
Obrázek 10: HRIR původního signálu i prvotního odrazu, floor tom, pravý kanál.

## 7.7 Grafické rozhraní

Za účelem vypracování poslechových testů vzniklo grafické rozhraní GUI pro přehrávání různých možností zvukových vzorků (Obrázek 11). Tlačítka pro spuštění jednotlivých nástrojů bicí sady jsou intuitivně rozmístěna dle přibližné horizontální pozice nástrojů. Přepínání módu přehrávání má možnosti:

- a) Binaurální nahrávky (Binaural)
- b) Syntetizované zvuky pomocí HRTF (Synth)
- c) Syntetizované zvuky se zahrnutými prvotními odrazy v místnosti (Room)

Uživatel tak může interaktivně přehrávat jednotlivé zvuky simulace a volit daný zdroj zvuku a princip syntézy. Zaškrťovací políčko *Export.wav file* označuje, zda dochází k uložení přehrávaného zvuku. Uložení probíhá pouze pro oba módy syntézy, v případě referenčních nahrávek by došlo k uložení stejné verze jako je načtena. Nahrávky se ukládají pod označením *Prvek + mód + formát*. Grafické rozhraní se spustí pomocí skriptu `drumkit.m`.



Obrázek 11: Grafické rozhraní programu `drumkit.m`

## 8 Poslechový test

### 8.1 Cíl poslechového testu

Z poslechového testu lze sledovat více závěrů. Primárně test vyhodnocuje schopnost lokalizace ve zvoleném virtuálním akustickém prostoru bez vizuálního vjemu, avšak s určitým očekáváním správné pozice původního zdroje zvuku. Další krok poslechového testu může vést k porovnání referenčních nahrávek binaurálními mikrofony vůči syntetizovanému zvuku. U syntetizovaného zvuku lze také pozorovat vliv přidání prvotních odrazů v místnosti. Určení lokalizace klade důraz především na hodnocení azimutu a elevace, vjem vzdálenosti a místnosti je vyhodnocen pouze relativním porovnáním.

## 8.2 Testovaná skupina

K vyřazení individuálního prvku každého psychoakustického měření je potřeba experiment opakovat na rozsáhlejší skupině posluchačů. Dle požadavků konkrétní formy psychoakustického experimentu je zapotřebí zvolit vhodnou skupinu respondentů. Pro obecné výsledky měření je vhodná náhodná skupina o co největším počtu jedinců. U experimentů, které nejdou provádět hromadně ale s počtem respondentů roste časová náročnost poslechové části. Náhodné složení posluchačů zajišťuje určité chyby hodnocení pro respondenty bez zájmu a vědomostí v oboru akustiky. Například při testování poslechových vlastností hudebních sálů jsou mezi posluchači běžní návštěvníci koncertů, případně hudebníci samotní. S ohledem na povahu experimentu je vlastní testovaná skupina složena především ze spolužáků a pedagogů Konzervatoře Jaroslava Ježka oboru hry na bicí nástroje. Respondenti jsou tedy denně zvyklí vnímat modelované akustické prostředí bicí soupravy právě z pohledu hráče. Pro hodnocení jim tedy není nutné pečlivě popisovat ideální výsledky simulace a nahrávek. Hodnocení lokalizace zdrojů zvuku může být tedy více podrobné. Nevýhodou zvolené skupiny je zvýšený nárok na uměleckou kvalitu nahrávek, která není v technickém provedení zcela dodržena.

Tabulka 3: Složení respondentů poslechového testu

pořadí	jméno respondenta	nástroj	praxe
1	Lukáš Doksanský	Bicí	Pedagog KJJ
2	Jakub Nývlt	Bicí	Student KJJ
3	Roman Vícha	Bicí	Pedagog KJJ
4	Adam Albrecht	Klavír	Student KJJ
5	Jakub Kaprál	Bicí	Student KJJ
6	Jan Zlatuška	Klavír	Student KJJ
7	Camilo Caller	Perkuse	Pedagog KJJ
8	Kryštof Danczi	Saxofon	Student KJJ
9	Vladimír Lahoda	Kytara	Ve volném čase
10	Vojtěch Hrubý	Syntezátory	Ve volném čase

## 8.3 Technické parametry poslechového testu

Ačkoli existují metody pro poslech binaurálních nahrávek přes reproduktory, nejběžnější forma poslechu přes sluchátka vychází z principu binaurálního záznamu a dodržuje reprodukci pro každé ucho odděleně. Volba sluchátek je nejkritičtější technický parametr reprodukce binaurální nahrávky. Dle [15] je potřeba odstranit frekvenční závislost sluchátek zvolených pro reprodukci. Metoda kompenzace inverzním digitálním filtrem není vhodná, jelikož se frekvenční charakteristika sluchátek mění s každým opakovaným měřením v závislosti na drobné změně pozice sluchátek vůči uchu. Tato závislost je nejmenší pro extraaurální sluchátka, vzhledem

k jejich otevřené konstrukci nezakrývající lidské ucho. Tato sluchátka se však svým principem blíží poslechu přes reproduktory, který přináší mnohem složitější problémy poslechu. Zdroj [16] naopak doporučuje uzavřená sluchátka blokující ušní kanál s vhodnou kompenzací frekvenční odezvy. Kombinací obou přístupů je možné využít uzavřená sluchátka s dostatečně plochou frekvenční charakteristikou bez použití kompenzace, která při změně frekvenční charakteristiky působí proti svému smyslu a signál poškozuje. K poslechovému testu byla zvolena uzavřená sluchátka Beyerdynamic DT770M. Tato sluchátka jsou standardem v nahrávacích studiích a svou frekvenční charakteristikou splňují uvedené požadavky.

## 8.4 Příprava a instrukce testu

Vzhledem k typu virtuálního akustického prostoru formou typické bicí soupravy se test nesnaží oddělit vizuální či paměťovou složku vjemu prostoru od akustické. Respondentům je předem vysvětlena povaha testu a potvrzeno, že se syntéza snaží dosáhnout reálného vjemu při hře na bicí nástroje. Vzhledem ke složení dotazované skupiny z řad hráčů na daný nástroj, není potřeba seznámení s očekávanými výsledky. V případě nejasností je respondentům potvrzena původní poloha nástrojů dle fotografie (Obrázek 9). Součástí testu je i hodnocení referenčních nahrávek zaznamenaných binaurálními mikrofony. Toto hodnocení slouží k určení celkové chybovosti záznamu akustického prostoru. Ani v ideálním případě poslechu pro vlastní personalizovanou HRTF nedojde ve všech případech ke správnému prostorovému vnímání. Dotazovaní tedy mají možnost preferovat syntetizovanou nahrávku před referenční.

## 8.5 Průběh testu

Respondenti mohou volně interagovat s grafickým rozhraním programu `drumkit.fig`. Postupně si u každého prvku bicí soupravy mohou poslechnout referenční binaurální nahrávku, prvotní syntetizovaný zvuk bez zavedených prvotních odrazů v místnosti a výsledný zvuk syntézy včetně odrazů zvukových paprsků v místnosti. Díky pevně definované pozici původního zdroje zvuku mohou určit odchylku od očekávané pozice.

### 8.5.1 Určení azimutu

V rovině azimutu je zvolena škála hodnot od 1 do 5. Jednotlivé stupně reprezentují následující odchylky lokalizace.

- 1) Znamka 1 hodnotí správnou, očekávanou polohu lokalizace.
- 2) Znamka 2 znamená odchylku v očekávaném kvadrantu, tedy v rozsahu  $\pm 45^\circ$
- 3) Znamka 3 reprezentuje časté zmatení v některé z os souřadného systému, tedy hodnoty azimutu  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$  a  $270^\circ$ .
- 4) Znamka 4 je určení směru z jiného než očekávaného kvadrantu.
- 5) Znamka 5 poté reprezentuje přesně opačný směr azimutu než je očekáván.

### 8.5.2 Určení elevace

V závislosti elevace jsou definované pouze tři pojmy. Posluchač rozlišuje elevace se zápornou hodnotou, tedy směr zdola, nulovou hodnotou, tedy horizontální polorovinu a kladnou hodnotou, tedy směr shora. Vzhledem k pozici reálného zdroje, který není nikdy v horizontální nulové rovině elevace, může odchylka od očekávané hodnoty být v nulové rovině, nebo v opačné. Elevace se tedy hodnotí na následující škále:

- 1) Znamka 1 reprezentuje očekávaný směr elevace
- 2) Znamka 2 reprezentuje vjem z nulové, horizontální roviny elevace
- 3) Znamka 3 reprezentuje opačnou polorovinu elevace, než je očekávána

### 8.5.3 Určení vzdálenosti

Pojem vnímání vzdálenosti je těžké definovat a proto je spojen s vjemem rozměru místnosti, který je zachycen na binaurální referenční nahrávce a snaží se ho modelovat syntéza v módu *Room*. Respondenti jsou tedy dotazováni, zda se vnímání rozměru místnosti modelováním odrazů zlepší, nebo zůstane nezměněné. Znamka 1 reprezentuje zlepšení a znamka 0 reprezentuje stejný vjem místnosti po přidání prvotních odrazů. Pro vjem místnosti bylo zásadní porovnávání s referenční nahrávkou binaurálními mikrofony.

## 8.6 Výsledky testu

Výsledky jednotlivých respondentů a následné zpracování jsou dostupné v příloze *SubjektivniTesty.xlsx*. Z jednotlivých individuálních hodnocení je určen aritmetický průměr a vypočtena směrodatná odchylka. Následující tabulky 4, 5 a 6 obsahují výsledky celkového zpracování pro jednotlivé módy přehrávání. Výsledky jsou dále zpracovány formou sloupcových grafů pro lokalizaci azimutu a elevace.

Tabulka 4: Subjektivní testy, binaurální nahrávky

Binaurální nahrávky				
	Azimut $\bar{\theta}$	$\sigma$	Elevace $\bar{\phi}$	$\sigma$
Hi hat	1,6	0,8	1,8	0,6
Snare	1,6	0,7	1,0	0,0
Bass drum	2,1	0,7	1,6	0,7
China	1,3	0,5	1,3	0,5
Floor tom	1,1	0,3	1,2	0,4

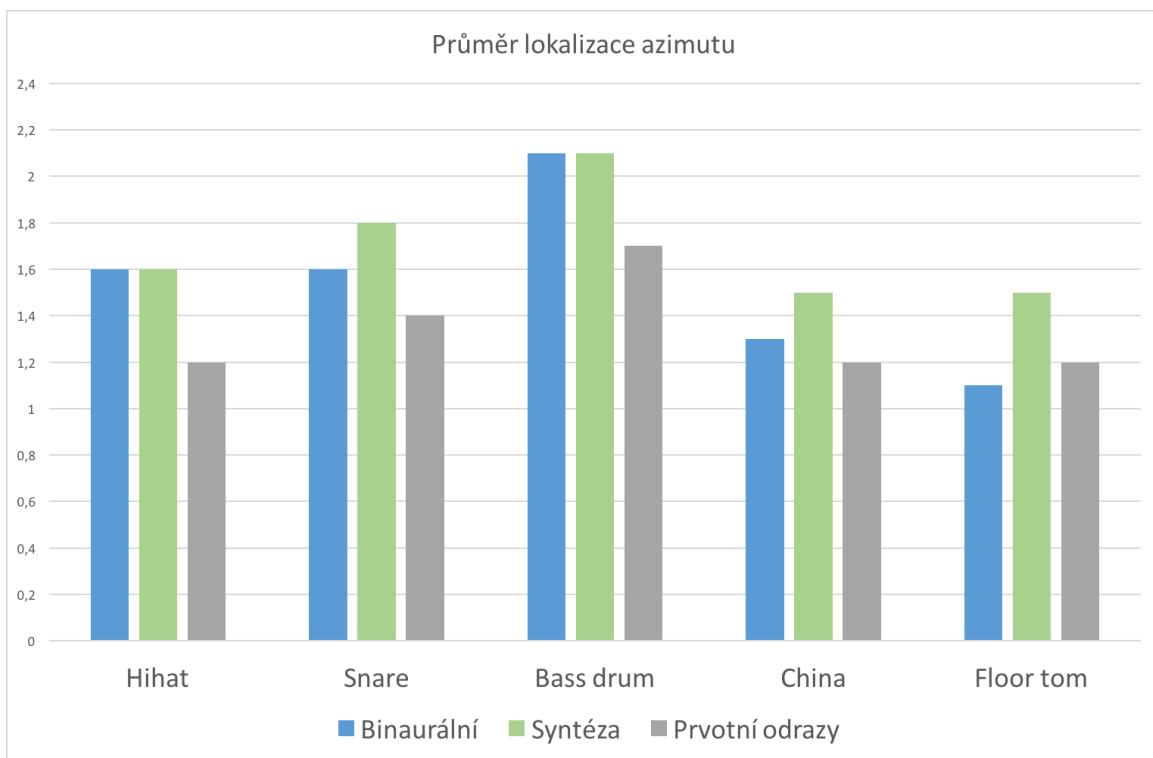
Tabulka 5: Subjektivní testy, syntetizované nahrávky

Syntetizované nahrávky SynthHRTF (bez prvotních odrazů)				
	Azimut $\bar{\theta}$	$\sigma$	Elevace $\bar{\phi}$	$\sigma$
Hi hat	1,6	0,7	1,7	0,6
Snare	1,8	0,9	1,4	0,5
Bass drum	2,1	0,7	1,2	0,4
China	1,5	0,8	1,0	0,0
Floor tom	1,5	0,7	1,6	0,7

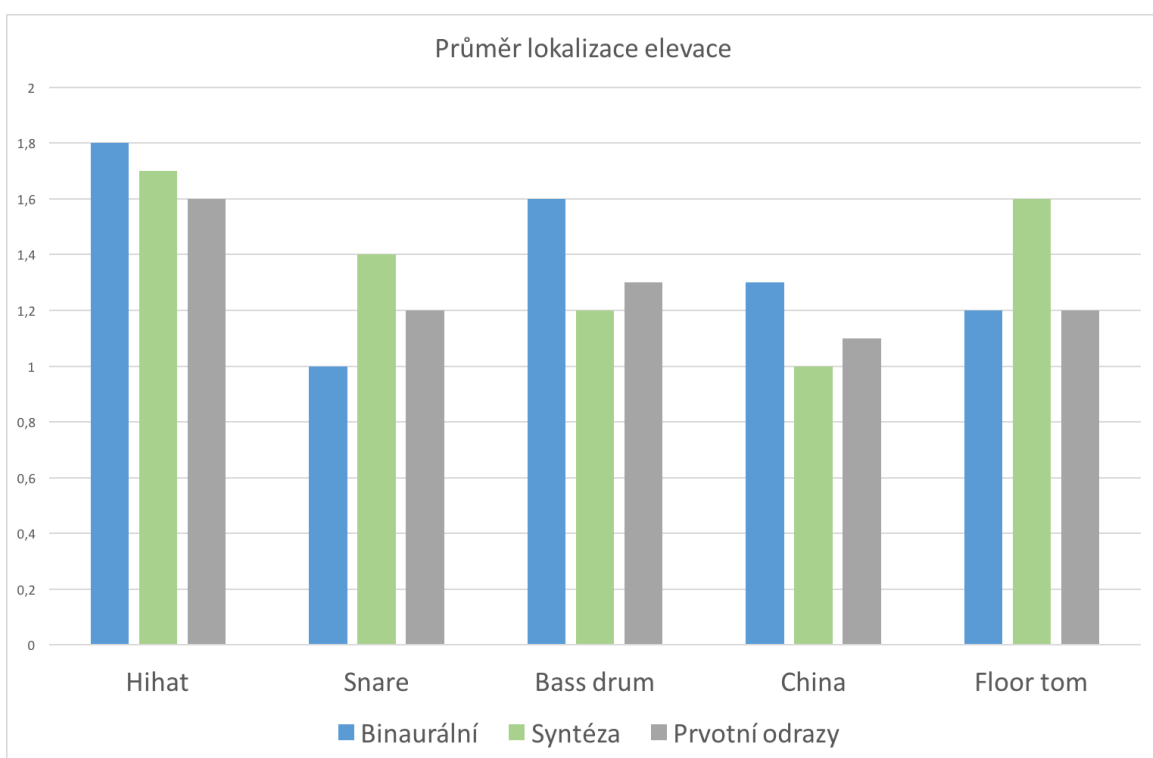
Tabulka 6: Subjektivní testy, model prvotních odrazů, včetně vlivu na vjem vzdálenosti

Syntetizované nahrávky RoomHRTF (včetně prvotních odrazů)					
	Azimut $\bar{\theta}$	$\sigma$	Elevace $\bar{\phi}$	$\sigma$	Vzdálenost [%]
Hi hat	1,2	0,4	1,6	0,5	90
Snare	1,4	0,5	1,2	0,4	80
Bass drum	1,7	0,5	1,3	0,5	100
China	1,2	0,4	1,1	0,3	10
Floor tom	1,2	0,4	1,2	0,4	70





Obrázek 12: Graf hodnocení pro azimut



Obrázek 13: Graf hodnocení pro elevaci

## 8.7 Reprezentace výsledků

Z celkového hodnocení lokalizace a kvality zvuku ve virtuálním akustickém prostoru lze vyvodit několik závěrů. Vzhledem ke kombinaci znalosti očekávané polohy zdroje zvuku a lokalizaci hodnoceného zvuku bylo možné test provést bez personalizované HRTF. Důležitá část porovnání výsledků je mezi syntézou bez prvotních odrazů a syntézou s prvotními odrazy v místnosti. Pro všechny prvky se zlepšil výsledek při určování azimutu. Při určování elevace se hodnocení zhoršilo pro basový buben a chinu. Hodnota se však změnila méně, než je střední odchylka. Pouze u basového bubnu překročilo průměrné hodnocení azimutu hodnotu 2, pro model s prvotními odrazy hodnota klesne, lokalizaci tedy poslední model pomůže i vůči referenčním nahrávkám. Vjemu vzdálenosti z naprosté většiny pomohl model místnosti, díky svému přirozenějšímu dozvuku. Vyjimku s opačnou hodnotou tvoří činel china, pro který model odrazů vjem místnosti nezlepšil. Pouze 10 % posluchačů určilo zlepšení, tedy pouze 1 posluchač. Důvodem je především dlouhý dozvuk samotného zvukového vzorku, který zamezí vjemu krátkého zpoždění prvotního odrazu. Při lokalizaci azimutu ale model místnosti pomohl i pro tento vzorek. Testy jsou vyhodnoceny pouze v závislosti lokalizace zdroje zvuku, respondenti nehodnotili ostatní rozdíly referenčních a syntetizovaných nahrávek, které jsou především v zvukové barvě.

## 9 Závěr

Cílem práce bylo vytvořit virtuální akustický prostor. Pojetí experimentu formou modelu bicí soupravy je velmi konkrétní případ. Výsledky experimentu jsou tedy vztaženy pouze pro danou situaci a nelze je příliš zobecnit. Výhodou je ale jasně definovaná podoba syntetizovaného prostoru a jeho parametry.

Z teoretického hlediska bylo potřeba prostudovat metody lokalizace zdrojů zvuku a naučit se zpracovávat signál HRTF. V části nahrávání jsem využil jak novou znalost záznamu binaurálními mikrofony, tak klasický postup nahrávání bicí soupravy. V neposlední řadě bylo vhodné ovládat i samotnou hru na nástroj a chápání vjemu prostoru z pohledu hráče.

Práce ukazuje praktické využití soustavy pro měření HRTF. Pro budoucí subjektivní testy je možné každému posluchači nahrát personalizovanou HRTF z požadovaných směrů a porovnat výsledky vůči obecné verzi. Toto porovnání by mohlo vést k obhájení nebo vyvrácení nutnosti personalizovat získané HRTF různých dostupných databází.

Samotný kód programu pro syntézu a přehrávání v Matlabu je málo tvárný. Tato skutečnost vychází především z pevně daných parametrů experimentu, které nemají nutnost flexibility. Dalším omezením je naměření pouze potřebných HRTF pro daný experiment. Při změření celkové databáze může být experiment rozšířen o další prvky soupravy, o jiný nástroj nebo může být rozšířen model prvotních odrazů v místnosti. Každý krok by bylo ideální důkladně vyhodnotit poslechovými testy, protože je prostorové vnímání velmi citlivé i na drobné změny.

Praktický vývoj tohoto experimentu by mohl vést k implementaci pro přehrávání vzorků elektronické bicí soupravy, u které se výrobci předhánějí s inovativními metodami zvukové syntézy. Prostorový vjem při samotné hře by mohl být technologickým krokem vpřed.

## 10 Zdroje

- [1] O. Šupka, Sestava pro měření HRTF, Praha: ČVUT FEL katedra radioelektroniky, 2016.
- [2] G. Enzner, C. Antweiler a S. Spors, Trends in Acquisition of Individual Head-Related Transfer Functions, Springer Berlin Heidelberg, 2013.
- [3] S. Weinzierl, Communication Acoustics, Berlin: edX.org, 2016.
- [4] „Wikipedie: Konvoluce,“ [Online]. Available: <https://cs.wikipedia.org/wiki/Konvoluce>.
- [5] M. Mandel, Spatial sound, Columbia University Dept. of Electrical Engineering, 2008.
- [6] F. Kolmer a J. Kyncl, Prostorová akustika, Bratislava: SNTL, 1980.
- [7] O. Jiříček, Úvod do akustiky, Praha: ČVUT, 2002.
- [8] A. Havel, VST implementace strukturálního modelu HRTF, Praha: ČVUT, 2011.
- [9] J. Mikšátko, Metodika Psychoakustických experimentů pro testování subjektů poslechové skupiny, Brno, 2016.
- [10] E. Zwicker a H. Fastl, Psychoacoustics - Facts and Models, Berlin: Springer-Verlag, 1999.
- [11] D. H. Cooper a J. L. Bauck, Prospects for transaural recording, Journal of the AES, 1989.
- [12] B. Gardner a K. Martin, „HRTF Measurements,“ MIT, 1994. [Online]. Available: <http://sound.media.mit.edu/resources/KEMAR.html>.
- [13] D. Steiner, „Akustika,“ [Online]. Available: <http://www.steiner.cz/david/akustika/>.
- [14] „Infinite Wave,“ [Online]. Available: <http://src.infinitewave.ca>. [Přístup získán 2017].
- [15] Z. Sharer a A. Lindau, Evaluation of equalisation methods for binaural signals, AES, 2009.
- [16] D. Griesinger. [Online]. Available: <http://www.davidgriesinger.com>. [Přístup získán 2017].
- [17] J. Stuchlík, Externalizace virtuálních zdrojů zvuku, Praha: ČVUT FEL katedra Radioelektroniky, 2015.
- [18] S. Carliele, Virtual Auditory Space: Generation and Applications, R. G. Landes Company, 1996.

- [19] U. Zölzer, DAFX - Digital Audio Effects (Second Edition), John Wiley & Sons, 2011.
- [20] H. Moller, Fundamentals of Binaural Technology, Aalborg: Aalborg University, 1992.
- [21] M. Ryšavý, Modelování akustiky uzavřených prostorů paprskovou metodou, Brno: VUT Brno, 2013.

## Použité zkratky

- HRTF – Head-Related Transfer Function
- HRIR – Head-Related Impulse Response
- BRTF – Binaural Room Transfer Function
- ITD – Interaural Time Difference
- ILD – Interaural Level Difference

## Příloha

- *Syntéza virtuálního akustického prostoru.pdf* – písemná práce
- *SubjektivniTesty.xlsx* – zpracování subjektivních testů a grafické výsledky
- *Tracks*
  - *Binaural* – Binaurální referenční nahrávky
    - *BinauralBassDrum.wav*
    - *BinauralChina.wav*
    - *BinauralFloortom.wav*
    - *BinauralHihat.wav*
    - *BinauralSnare.wav*
  - *Source* – Nahrávky určené pro následné zpracování v matlabu
    - *BassDrum.wav*
    - *China.wav*
    - *FloorTom.wav*
    - *Hihat.wav*
    - *Snare.wav*
  - *Export* – Nahrávky exportované programem `drumkit.m`
- *Matlab* – Složka s jednotlivými funkcemi
  - `drumkit.m` – hlavní program, který volá ostatní funkce a spouští GUI
  - `drumkit.fig`
  - `roomHRTF.m`
  - `synthHRTF.m`
  - `nastroj.m`
  - `zpozdeni.m`
  - `convolute.m`
  - `odrazy.mat`
  - `zmereneHRTF.mat`