



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

**Fakulta elektrotechnická
katedra radioelektroniky**

Sonifikace polohy objektu

Sonification of an Object Position

Bakalářská práce

Studijní program: Komunikace, multimédia a elektronika

Studijní obor: Multimediální technika

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

Petr Erbes

Praha 2014

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.

Datum: 20. 5. 2014

.....
podpis bakalanta

Poděkování

Rád bych tímto poděkoval vedoucímu práce Ing. Františku Rundovi, Ph.D. za přínosné rady a připomínky k mé bakalářské práci. Dále bych rád poděkoval svým rodičům za finanční podporu při studiu a rovněž všem dobrovolníkům, kteří se účastnili experimentů.

Zadání

Seznamte se s problematikou sonifikace polohy objektu.

V prostředí Pure Data navrhňte a realizujte algoritmus pro sonifikaci polohy pohybujícího se objektu z videosignálu. Parametry implementovaného algoritmu ověřte poslechovým testem.

Abstrakt:

Tato závěrečná práce se zabývá problémem sonifikace videosignálu. V práci jsou shrnuty základy lokalizace zdroje zvuku lidským sluchem, dále je vysvětlen pojem sonifikace včetně popisu jednotlivých typů a několika sonifikačních metod sloužících ke zpracování videosignálu či pohybu. Součástí práce jsou testy zkoumající subjektivní spojení mezi páry: „elevace – výška tónu“ a „jas obrazu – jasnost tónu“. Následně je v prostředí Pure Data navržen algoritmus sloužící k sonifikaci polohy pohybujícího se objektu z videosignálu. Elevace objektu je mapována na výšku tónu, azimut je prostřednictvím ITD a ILD převeden na stereofonii signálu, dále je sonifikován jas objektu a rychlost pohybu. V závěru práce jsou provedeny testy, které ukazují, že rozlišení jasu na základě navržené struktury je poměrně obtížné, zatímco rozpoznání sonifikované trajektorie a její rychlosti funguje velice přesně.

Klíčová slova: *sonifikace, Pure Data, virtuální polohování zvuku*

Abstract:

This final project deals with problem of sonification of video. In this project are summarized fundamentals of sound localization by human hearing, further the meaning of sonification is explained including description of different sonification types. Also several techniques for video processing and move sonification are described. Further there are tests investigating subjective connection between couples: 'elevation – pitch' and 'brightness of video – brightness of sound'. Then an algorithm for video sonification is devised using platform Pure Data. Elevation is mapped on to pitch and azimuth is transformed to stereophony of sound using ITD and ILD, further brightness and speed of movement are sonified. In conclusion there are tests, which show that the recognition of brightness based on the proposed structure is relatively difficult, while recognizing sonified trajectory and its speed works very accurately.

Key words: *sonification, Pure Data, virtual sound positioning*

Obsah

1	Úvod	1
2	Teorie slyšení a lokalizace zvuku	2
2.1	Souřadný systém	2
2.2	Lokalizace zvuku v horizontální rovině	3
2.2.1	ITD – Interaurální časová diference	3
2.2.2	ILD – Interaurální intenzitní diference	4
2.3	Lokalizace zvuku v horizontální i mediální rovině	5
2.4	Omezení spojená s vnímáním směrovosti	5
2.5	Směrovost zvuku a vnímání výšky tónu v prostoru	7
2.5.1	Výška tónu a další obrazové parametry	7
2.6	Spojení sluchových a zrakových vjemů	8
2.6.1	„Jasnost“	8
2.6.2	„Jas“	9
3	Sonifikace	10
3.1	Základní charakteristika a definice pojmu sonifikace	10
3.1.1	Definice sonifikace	10
3.1.2	Využití sonifikace	10
3.1.3	Rozdělení sonifikace	11
3.2	Různé způsoby sonifikace videosignálu a pohybu	13
3.2.1	Theremin jako předchůdce sonifikace	13
3.2.2	Cesta vlakem a Skrjabinův kód	13
3.2.3	Převod grafu pohybu na spektrogram	14
3.2.4	Spojení nervových center a sonifikace pohybu paže	15
3.2.5	Rehabilitace po CMP a sonifikace pohybu	15
4	Praktická část	17
4.1	Návrh algoritmu	17
4.2	Motivace použití PD a zpracování videosignálu	19
4.2.1	Získání souřadnic pohybujícího se objektu	19
4.3	Spojení jasu a jasnosti (experiment a implementace)	20
4.3.1	Subjektivní testy spojování <i>jasu</i> a <i>jasnosti</i>	20
4.3.2	Implementace mapování jasu	23
4.4	Mapování elevace na výšku tónu (experiment a implementace)	25
4.4.1	Subjektivní vnímání výšky tónu jako funkce elevace	25
4.4.2	Implementace mapování elevace	30
4.5	Implementace mapování azimutu	35
4.5.1	Interaurální časový rozdíl	35

Implementace ILD.....	36
5 Testování algoritmu.....	39
5.1 Testování vnímání rychlosti.....	39
5.1.1 Výsledky.....	39
5.2 Souhrnný test algoritmu.....	39
5.2.1 Výsledky.....	40
6 Závěr.....	42
7 Zdroje:	43
8 Soubory na přiloženém CD.....	46
9 Přílohy	47

Seznam použitých zkratek

PD – Pure Data

ITD – Interaurální časová diference (*interaural time difference*)

ILD – Interaurální intenzitní diference (*interaural level difference*)

JND – Právě postřehnutelný rozdíl (*just noticeable difference*)

MAA – Minimální slyšitelný úhel (*minimal audible angle*)

BR – Jasnost (*brightness*)

HRTF – Hlavová přenosová funkce (*head related transfer function*)

GEM – Graphics Enviroment for Multimedia

IIR – Nekonečná impulsní odezva (*infinite impulse respnose*)

DP – Dolnoproputní filtr

CMP – Cévní mozková příhoda

1 ÚVOD

Tato práce se zabývá tématem sonifikace obrazu. V úvodu teoretické části je popsán základ binaurálního slyšení, následuje část o vztahu výšky tónu a umístění v prostoru. Tato část je zakončena kapitolou o vztahu mezi různými smyslovými oblastmi.

Dále práce pojednává o sonifikaci, která je zde definována a zároveň jsou uvedena možná využití sonifikace jako takové. Poté jsou popsány základní sonifikační metody a v závěru teoretické části je vylíčeno několik existujících sonifikačních algoritmů, které slouží k převodu videa, či pohybu do zvukové podoby.

V praktické části práce je popsán návrh sonifikačního algoritmu v prostředí Pure Data, který slouží k převodu polohy pohybujícího se objektu z videosignálu. K převodu elevace je použito kódování pomocí výšky tónu, azimut je převeden prostřednictvím stereofonie, velikost pohybu určuje amplitudu případně rytmus zvuku, a konečně jas objektu je mapován na jasnost tónu.

V průběhu implementací jednotlivých částí jsou provedeny dva testy, které zkoumají subjektivní spojení jasu obrazu a jasnosti tónu a spojení výšky tónu a umístění v prostoru.

V závěru je popsáno testování algoritmu a výsledky jsou vyhodnoceny.

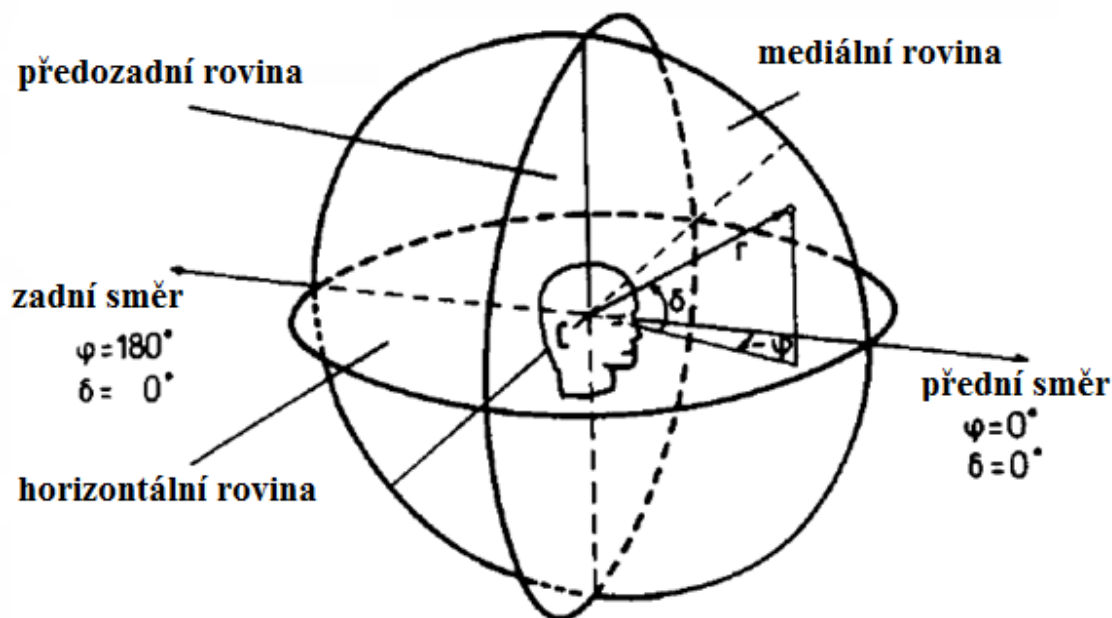
Hlavní motivací pro vznik této práce bylo hledání alternativ k virtuálnímu polohování zvuku pomocí HRTE, protože přesnost této metody i přes relativně vysokou náročnost na její implementaci, nedosahuje uspokojivých výsledků. Metodu navrženou v této práci by bylo možné rozvinout například pro navigační, rehabilitační, či signalizační účely.

2 TEORIE SLYŠENÍ A LOKALIZACE ZVUKU

2.1 Souřadný systém

Abychom se mohli zabývat otázkou polohování virtuálního zvuku v prostoru, musíme nejdříve definovat souřadný systém, se kterým budeme pracovat. Výhodný pro tento účel je sférický systém, kde se posluchač nachází ve středu pomyslné koule definované tímto systémem. Souřadný systém je ilustrován na obr. 1.

Poloha objektu je definována pomocí dvou úhlů (φ ; δ) a vzdálenosti r od posluchače. Polohu v horizontální rovině definuje velikost úhlu φ . Pokud mluvíme o poloze v horizontální rovině, můžeme rovněž mluvit o azimutu objektu. Umístění objektu v mediální rovině charakterizuje velikost úhlu δ a bývá nazýváno elevací objektu. Popisovaný souřadný systém je možné vidět na obrázku níže.



Obrázek 1: Souřadný systém použitý pro popis slyšení [1]

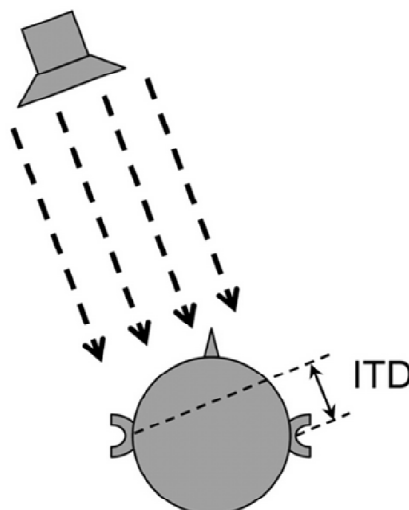
2.2 Lokalizace zvuku v horizontální rovině

Abychom dále mohli pracovat se směrovostí zvukového signálu, zmíníme v této části způsoby, jakými člověk vnímá směrovost zvuku.

Na počátku 20. století přišel Rayleigh s tzv. duplexní teorií, která říká, že mozek pro rozlišení směru používá dvě odlišné metody založené na časové, nebo intenzitní diferencii zvukového signálu přijatého levým a pravým uchem [2]. Která z metod je v danou chvíli dominantní, závisí především na frekvenci zvukových signálů a s ní spojenými fyzikálními a biologickými omezení pro každý ze způsobů vnímání směrovosti.

2.2.1 ITD – Interaurální časová diference

Duplexní teorie říká, že mozek může rozlišovat, z jakého směru zvuk přichází podle zpoždění v příjmu zvuku pravým a levým uchem, což je znázorněno na obr. 2.



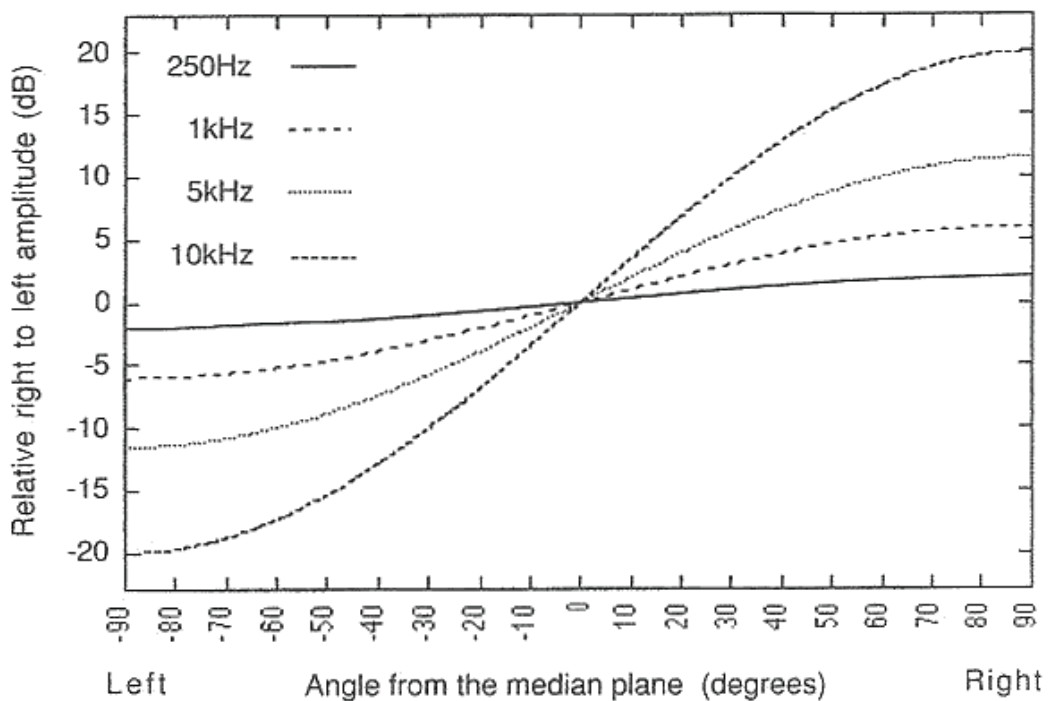
Obrázek 2: Schematické zobrazení interaurální časové diference [3]

Abychom byli schopni rozlišit posun fáze levého vůči pravému uchu, musí být vlnová délka přijímaného zvuku větší, než je vzdálenost obou uší [4]. Z experimentů prováděných na modelu hlavy tvaru koule (tedy určité aproximace tvaru hlavy) vyplynulo, že největší takto dosažitelné zpoždění je 0,66 ms, což odpovídá frekvenci 1500 Hz a vlnové délce 21 cm, tedy vzdálenosti uší průměrného člověka. Rozlišení směru pomocí ITD se tedy omezuje pouze na frekvence zhruba do 1500 Hz.

Další omezení spojená s touto metodou vyplývají ze schopnosti nervových buněk předávat mozku informaci o fázi přijímaného zvuku, ta se totiž začíná ztrácet kolem frekvence 1 kHz [5]. Mluvit proto při vyšších frekvencích o ITD by bylo i z tohoto důvodu bylo nepřesné, protože mozek u frekvencí vyšších 1 kHz nerozlišuje fázi signálu, ale pouze jeho přítomnost.

2.2.2 ILD – Interaurální intenzitní diference

Druhý způsob, jakým ucho rozpoznává směr dopadajícího zvuku je rozdíl v intenzitě dopadajícího zvuku. Tento rozdíl je způsoben tzv. akustickým stínem hlavy, případně také směrovostí ušního boltce. Vzhledem ke způsobu šíření zvukových vln prostředím, je tento způsob prostorového vnímání zvuku možné uvažovat až od vyšších frekvencí, protože pouze tyto vyšší frekvence hlavu „neobejdou,“ a dojde tedy k jejich částečné filtraci. Spektrum přijímaného zvuku levým a pravým uchem bude rozdílné právě o chybějící (nebo částečně potlačenou) složku frekvencí vyšších 1,5 kHz [8].

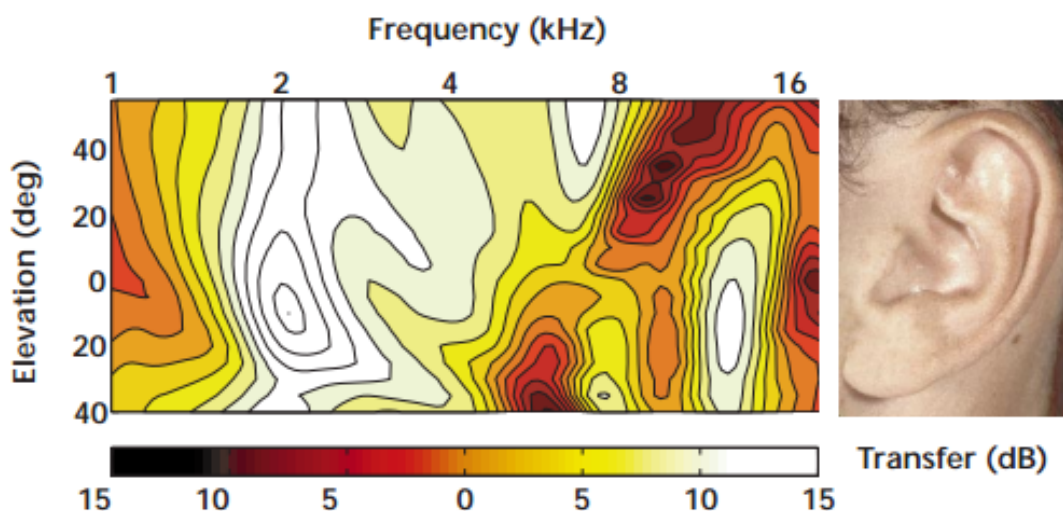


Obrázek 3: Závislost relativní amplitudy na azimutu [4]

Na obr. 3 je vynesena závislost relativní amplitudy signálu přijatého levým a pravým uchem na azimutu. Můžeme vidět, že závislost není lineární, ale blíží se spíše sinusovému průběhu [4]. Dále můžeme pozorovat, že rozdíl hlasitostí signálu o frekvenci 1 kHz při azimutu 90° je přibližně roven 6 dB, což je vzhledem k tomu, že za minimální slyšitelný rozdíl při vnímání hlasitosti bývá označována hodnota 1 – 2 dB velice málo, uvědomíme-li si, že se jedná o největší možný úhel a tedy i největší možný rozdíl amplitud. Rozlišitelnost směrovosti při 1 kHz je tedy velice omezená a jedná se o jakousi hranici použitelnosti ILD. Na druhé straně je nutno poznamenat, že lidský mozek je schopen řídit se pomocí ILD i na nižších frekvencích, pouze se jedná o jev, který přirozeně nenastává, ale je ho možné simulovat například pomocí amplitudové panoramy.

2.3 Lokalizace zvuku v horizontální i mediální rovině

Vzhledem k tomu, že vnitřní ucho není uzpůsobeno tak, aby samo o sobě rozlišovalo vertikální směr příchozího zvuku, využívá mozek k rozlišení směrovosti dalších vlastností zvukového signálu [5]. Hlavní roli zde hraje vliv boltce, který funguje jako filtr, jehož přenos závisí na frekvenci a na elevaci příchozího zvuku. Mozek tedy rozpoznává směr příchozího zvuku ze změn ve frekvenčním spektru signálu. Pro správné rozlišení směru je proto zapotřebí zvukový podnět o co největší šířce pásma a zároveň významnou roli hraje znalost nefiltrovaného - původního zvuku.



Obrázek 4: Graf přenosu boltce v závislosti na frekvenci a elevaci [4]

Zajímavé experimenty s frekvenčními vlastnostmi ušního boltce byly provedeny v roce 1998 Hofmanem, Riswickem a Opstalem na Universitě v Nijimegenu [5]. Během nich byl uživatelům do ucha vložen malý nástavec, který měnil frekvenční vlastnosti boltce, což v prvních dnech vedlo k značnému zhoršení orientace, avšak zhruba po třiceti dnech se vnímání směrovosti přiblížilo původnímu stavu – mozek se přizpůsobil novým frekvenčním vlastnostem sluchového traktu, což potvrzuje, jak flexibilní lidský mozek dokáže být.

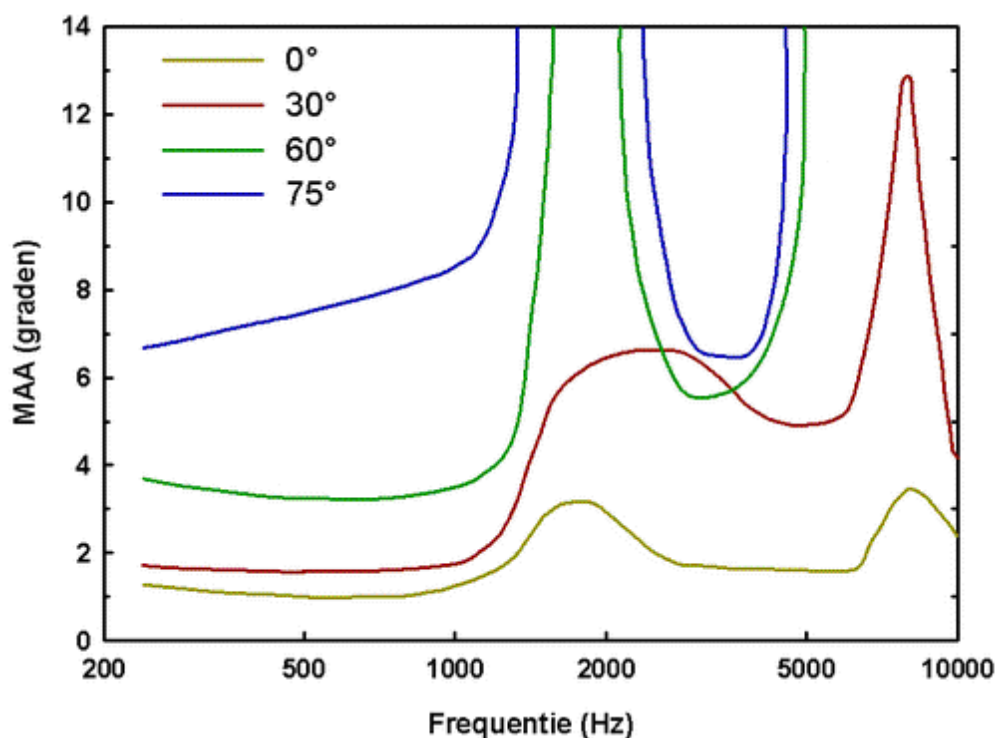
Kromě boltce se při rozpoznávání směrovosti uplatňují také odrazy od ramen a hrudníku. Ty fungují jako hřebenové filtry a mozek díky způsobeným změnám ve spektru může přesněji rozlišovat směr zvuku.

2.4 Omezení spojená s vnímáním směrovosti

Rozlišení směrovosti zvukového signálu je ovlivněno několika faktory. Z geometrických zákonitostí ITD a ILD vzniká tzv. kužel neurčitosti [6], který určuje prostor, pro který není pouhým využitím ITD a ILD možné určit polohu objektu, protože hodnoty ITD a ILD jsou na různých bodech kuželu stejné. Tento problém je do určité míry spíše teoretického rázu, protože při poslechu v reálném prostředí rozlišuje mozek směrovost pomocí filtrace boltcem, případně si posluchač může napomocnat natáčením hlavou. Odtud

může částečně pramenit instinkt otáčet se za zvukem, který slyšíme pouze jedním uchem – kromě potřeby identifikovat původ zvuku zde může hrát roli i snaha o přesnější určení polohy.

Další omezení spojená s rozlišením směrovosti závisí na frekvenci, hlasitosti i poloze objektu [7]. Rozlišitelnost směrovosti se nazývá minimální slyšitelný úhel (MAA z anglického *minimal audible angle*), ten se výrazně mění v závislosti na azimutu i v závislosti na frekvenci [8], jak je možné vidět na obrázku níže.



Obrázek 5: Závislost MAA na frekvenci signálu pro různé hodnoty azimutu, kde na vodorovné ose je vynesena frekvence v Hz a na svislé ose MAA ve stupních [8]

Je velice zajímavé, že při frekvencích do 1500Hz závisí MAA převážně na azimutu. To je způsobeno tím, že ILD na takto nízkých frekvencích nehraje téměř žádnou roli. Naopak veliký růst MAA v oblasti kolem 1500Hz lze vysvětlit tak, že zde končí frekvenční pásmo, ve kterém může fungovat ITD.

2.5 Směřovost zvuku a vnímání výšky tónu v prostoru

Studiem vztahu výšky tónu a umístění v prostoru se zabývá práce E. Rusconiové [9], přičemž následující text popisuje nejen její práci, ale zmiňuje i ostatní experimenty, o kterých Rusconiová píše, nebo z kterých vychází. Výška tónů bývá velice často spojována s umístěním zdroje zvuku v prostoru. Tento jev je na první pohled zřejmý již ze slovního označení pro „hluboké“ a „vysoké“ tóny, kde kromě češtiny můžeme stejnou analogii nalézt například v angličtině („low – high“), němčině („tief – hoch“) ale i v jiných jazycích (např. v čínštině, polštině, španělštině a dalších). Spojování výšky s umístěním v prostoru můžeme rovněž vidět v notovém zápisu, kde vyšší tón zaujímá vyšší místo v notové osnově.

Otázkou vnímání výšky tónu ve spojitosti s prostorem se jako první v třicátých letech 20. st. zabývala C. Prattová, která ze skrytého reproduktoru přehrávala posluchačům čtyři různé tóny, vzdálené od sebe vždy jednu oktávu, přičemž posluchači měli za úkol postavit se na schody mezi pokojem a sklepem tam, kde se by se podle nich tón měl nacházet. Prattová tímto experimentem potvrdila, že člověk podvědomě spojuje vyšší tóny s vyšším umístěním v prostoru, protože všichni dotyční spojili nejnižší místo s nejhlubším tónem, nevyšší s nejvyšším atd. Podobné pokusy provedl v roce 1934 Trimble. Roku 1968 Roffler a Buttler opakovali pokus C. Prattové. Všichni dospěli ke stejným závěrům – tedy že vyšším tónům náleží vyšší místo v prostoru.

Roku 1963 provedl Mudd experiment zkoumající vztah dvou tónů a jejich relativního umístění [8]. Posluchačům byly postupně přehrávány dvojice tónů. Ti před sebou měli dvojrozměrnou plochu, do které mohli umístit kolíky reprezentující dva tóny, přičemž první tón se nacházel vždy uprostřed této plochy. Z pokusů vyplynulo, že člověk nevnímá výšku tónu pouze vertikálně, ale rovněž horizontálně. Vyšším tónům náležel prostor nahoře vpravo, nižším dole vlevo.

Dvojrozměrným umístěním zvuků dle výšky tónu se zabývá experiment Eleny Rusconiové a kol. [9]. Experiment, který byl proveden na lidech hudby a hry na nástroj neznalých i na profesionálních hudebnících, spočíval v zjišťování rychlosti reakce na výšku tónu v závislosti na umístění tlačítek sloužících k odpovídání. Z testů vyplynulo, že reakce byla vždy rychlejší, když pro vyšší tón bylo tlačítko umístěno výše, rovněž se potvrdilo, že reakce je rychlejší, když pro nižší tón bylo tlačítko umístěno vlevo. Zatímco vertikální spojování výšky tónu s umístěním prostoru funguje vždy (tedy pro profesionální hudebníky i pro lidi nehudební), horizontální vnímání se značně liší. Profesionální hudebníci měli výrazně větší potřebu spojovat hluboké tóny s levou stranou, vysoké s pravou, z čehož vyplývá, že hra na klavír (nebo jiné klávesové nástroje) výrazně ovlivňuje vnímání výšky tónu v souvislosti s jeho směrovostí.

2.5.1 Výška tónu a další obrazové parametry

Collier a Hubbard ve své studii Hudební stupnice a hodnocení jasu [10] mluví o souvislosti výšky tónu s jasnou obrazu. V pokusu z roku 1996 Hubbard předkládá účastníkům experimentu stejně velké obdélníky v dvanácti různých odstínech šedi (od

černé po bílou) a cílem experimentu je zjistit, jak souvisí hudební intervaly s vnímáním jasu. Z experimentu vyplynulo, že klesajícím intervalům byly přiřazovány tmavější odstíny, stoupajícím intervalům odstíny světlejší. Rovněž bylo zjištěno, že větší intervaly zaujímaly krajní polohy jasu: to znamená, že veliký stoupající interval byl označen velice světlým obdélníkem, veliký klesající interval obdélníkem, jehož barva se blížila černé.

Z dalších experimentů, které Collier a Hubbard provedli, vyplývá, že posluchači si spojují hudební stupnici zahrnou od vyššího tónu s jasnější barvou, stupnici zahrnou níže s temnější barvou. Studie rovněž mluví o změnách ve frekvenčním spektru klarinetu a jejich spojení s vnímáním jasu, poslední kategorií zmiňovanou v souvislosti s jasnem je tempo, kdy rychlejší tempo podle výsledků experimentu bývá spojováno se světlejší barvou.

2.6 Spojení sluchových a zrakových vjemů

„Barva zvuku byla už ve svých prvních definicích vztahována k subjektivní představě kvality zvuku a pro její verbální popis se převzala adjektiva z jiných smyslových oblastí, především pak ze zraku a hmatu.“ Záměrně uvádím tuto citaci Syrového publikace Hudební akustika [11], protože dále v práci se budeme zabývat právě převodem zrakových (či obrazových) kvalit na zvukové vlastnosti.

Pro popis barvy zvuku často slouží protikladné pojmové páry, které by měly vystihovat vjem barvy zvuku, protože popis zvuku jedinou matematickou veličinou je (pokud ne nemožný) obtížný.

Toto jsou některé na sobě nezávislé pojmové páry, které slouží k popisu barvy zvuku:

světlý ↔ tmavý

měkký ↔ ostrý

slabý ↔ silný

Takovýchto pojmových párů samozřejmě existuje mnoho a to nejen pro popis zvuku, ale v přeneseném smyslu i pro popis nástroje, o jehož kvalitách prakticky barva zvuku (tónu) vypovídá.

2.6.1 „Jasnost“

Syrový kromě podvědomého spojování vjemů z různých smyslových oblastí uvádí studii Beauchampa, který zavedl pojem „jasnost“ zvuku (BR ; *brightness*). Tato veličiny vyjadřuje poměr vyšších harmonických ve spektru.

$$BR = \frac{\sum_{n=1}^k k * c_k}{c_0 + \sum_{n=1}^k c_k} \quad (1)$$

Kde c_k vyjadřuje velikost k-té harmonické složky v lineárním měřítku a c_0 prahovou velikost užitečné harmonické složky.

2.6.2 „Jas“

Jak bylo zmíněno výše, zvukové parametry bývají často pojmenovávány ekvivalenty převzatými z jiných smyslových vjemů. Jedním konkrétním příkladem může být následující dvojice:

brightness ↔ brightness

Záměrně uvádíme anglickou podobu výrazů, kde na jedné straně *brightness* znamená *jasnost* zvuku (tj. vlastnost popsanou v předchozí kapitole), druhým významem tohoto slova je *jas* obrazu. Vzhledem k tomu, že v praktické části budeme vztahovat *jas* videa na *jasnost* zvuku, uvádíme zde rovněž charakteristiku *jasu*.

Jas obrazu je definován jako měrná veličina svítivosti. V barevném prostoru RGB může být jas μ spočítán jako aritmetický průměr hodnot RGB [12].

$$\mu = \frac{R+G+B}{3} \quad (2)$$

Vzhledem k rozdílné citlivosti lidského oka na různé barvy (respektive na části barevného spektra) by ovšem bylo v rozporu s vlastnostmi lidského zraku uvažovat jas jako aritmetický průměr složek RGB [13]. Abychom při implementaci obsáhli rozdílnou citlivost lidského oka na různé barvy, budeme hodnoty RGB váhovat podle vztahu mezi barevnými prostory XYZ ↔ RGB [13].

$$\begin{aligned} X &= 0,607 * R + 0,173 * G + 0,200 * B \\ Y &= 0,299 * R + 0,587 * G + 0,144 * B \\ Z &= 0,000 * R + 0,066 * G + 1,116 * B \end{aligned} \quad (3)$$

Místo hodnoty jasu μ budeme za hodnotu jasu považovat jasovou složku *Y*.

3 SONIFIKACE

3.1 Základní charakteristika a definice pojmu sonifikace

Pod pojmem sonifikace rozumíme proces, při kterém dochází k převodu libovolných dat na zvukový signál. Jedná o způsob transformace hojně využívaný v mnoha odvětvích lidské činnosti.

Sonifikace nejčastěji slouží tam, kde zvuková podoba dat je pro lidské vnímání přirozenější, či intuitivnější než původní podoba signálu. Sonifikace se rovněž využívá jako doplněk dalších informačních zdrojů. Podrobné rozdělení sonifikace do skupin dle využití naleznete v následující části.

3.1.1 Definice sonifikace

Abychom mohli mluvit o sonifikaci, musí být podle Thomase Hermanna [14] splněny následující podmínky:

- 1) Proces musí objektivně reflektovat vlastnosti vstupních dat, nebo jejich vztahy.
- 2) Transformace dat do zvukové podoby musí být systematická.
- 3) Sonifikace musí být reprodukovatelná.
- 4) Systém musí být použitelný s jinými daty, zároveň opakovatelně použitelný s daty původními.

3.1.2 Využití sonifikace

V článku Theory of Sonification dělí Walker a Nees [14] sonifikaci do těchto čtyř základních skupin a uvádějí některé ze zmíněných příkladů:

1) Alarmy, výstrahy, varování

Do této skupiny spadají veškeré zvukové zprávy, které indikují, že se něco stalo, nebo stane. Mohli bychom zmínit například ohlášení příjmu elektronické pošty, domovní zvonek, či pípání mikrovlnné trouby.

Vývoj v této oblasti se v poslední době zaměřuje na předání většího množství informací obsažených v samotném alarmu. Pokud bychom zmínili například požární alarm, namísto běžného významu „*hoří, evakuujte budovu a volejte 150*“ by alarm mohl obsahovat rovněž informaci o rozsahu či umístění požáru.

2) Zvukové složky indikující stav a vývoj funkce

Do této skupiny spadají veškeré sonifikační metody, které uživateli předávají informace o stavu určité veličiny v určitou chvíli. Příkladem může být signalizace vzdálenosti předmětů za automobilem při parkování. Uživatel (řidič) má na jedné straně vizuální kontrolu vzdálenosti a pro upřesnění využívá i zvukové vyjádření této vzdálenosti.

V jiných případech jsou takto sdělovány informace, které uživatel nemůže kontrolovat vizuálně například proto, že zrak potřebuje k něčemu jinému. Dobrým případem je audifikace činnosti srdce během operace, kdy lékař nemůže průběh elektrokardiogramu kontrolovat vizuálně, zatímco zvuková složka ho nijak nezatěžuje.

3) Výzkum dat

Do této skupiny spadají takové techniky, které bychom mohli označit za „pravou sonifikaci.“ Je to takový převod dat, který primárně neslouží k signalizačním účelům, ale jehož cílem je transformace určitého souboru dat tak, že jeho zvuková podoba přinese snadnější, přesnější, intuitivnější či jakkoli kvalitativně jiné zkoumání těchto dat.

4) Umění, zábava, sport a cvičení

Do této skupiny spadají veškeré počítačové hry, které využívají zvukovou složku. Z her založených vyloženě na zvukové stránce bychom mohli zmínit *Audio ping-pong*,¹ či *Vlimbo*². Do této skupiny bychom rovněž mohli zařadit veškeré metody pro sonifikaci polohy využívané ve zvukové složce počítačových her (například *Far Cry 3 HRTF 3D Sound*³).

Významnou skupinou v této oblasti je audifikace svalové činnosti sloužící k rehabilitaci atrofovaných svalů, poúrazové rehabilitaci apod.

Poslední skupinu tvoří využití sonifikace při uměleckých projektech. Můžeme zmínit například inscenaci *Antikódy na Nové Scéně Národního divadla*, ve které je pohyb tanečnicků přenášen na podobu hudební složky.

3.1.3 Rozdělení sonifikace

Sonifikaci můžeme rozdělit do tří základních skupin podle způsobu transformace dat na zvukový signál. Walker a Ness zmiňují tyto tři způsoby a jejich příklady [14]:

1) Audifikace

Audifikace je nejpřirozenějším způsobem sonifikace, kdy data jsou přímo přenesena do zvukové podoby. Příkladem může být sonifikace seizmických vln, nebo elektrických signálů snímaných ze svalů. Tento postup může zahrnovat změnu ve frekvenci, či rychlosti přehrávání.

2) Sonifikace mapováním parametrů

Pod tímto pojmem rozumíme takovou sonifikaci, jejímž cílem je nalezení mapovací funkce (kódu) pro převod datové struktury do zvukové podoby. Je to takový způsob sonifikace, který se snaží najít takové mapovací funkce, které akceptují uživatelskou rovinu (tj. srozumitelnost, pohodlnost) a zároveň stále splňují čtyři podmínky, které uvádí

¹ Audio ping-pong je hra využívající pouze zvukovou stránku; navrženo pro Android.

Dostupné z: <<https://play.google.com/store/apps/details?id=SteveAudio.PingPong>>.

² VLimbo je audio hra založená na sonifikaci polohy telefonu. AVISSAR (a kol.) *An audio game app using interactive movement sonification for targeted posture control*. ICAD. 2013.

³ Far Cry 3 je počítačová hra využívající polohování zvuku založené na simulaci HRTF. Zvukovou ukázkou je možné nalézt na: <<http://www.youtube.com/watch?v=dvRqJZEfIuM>>.

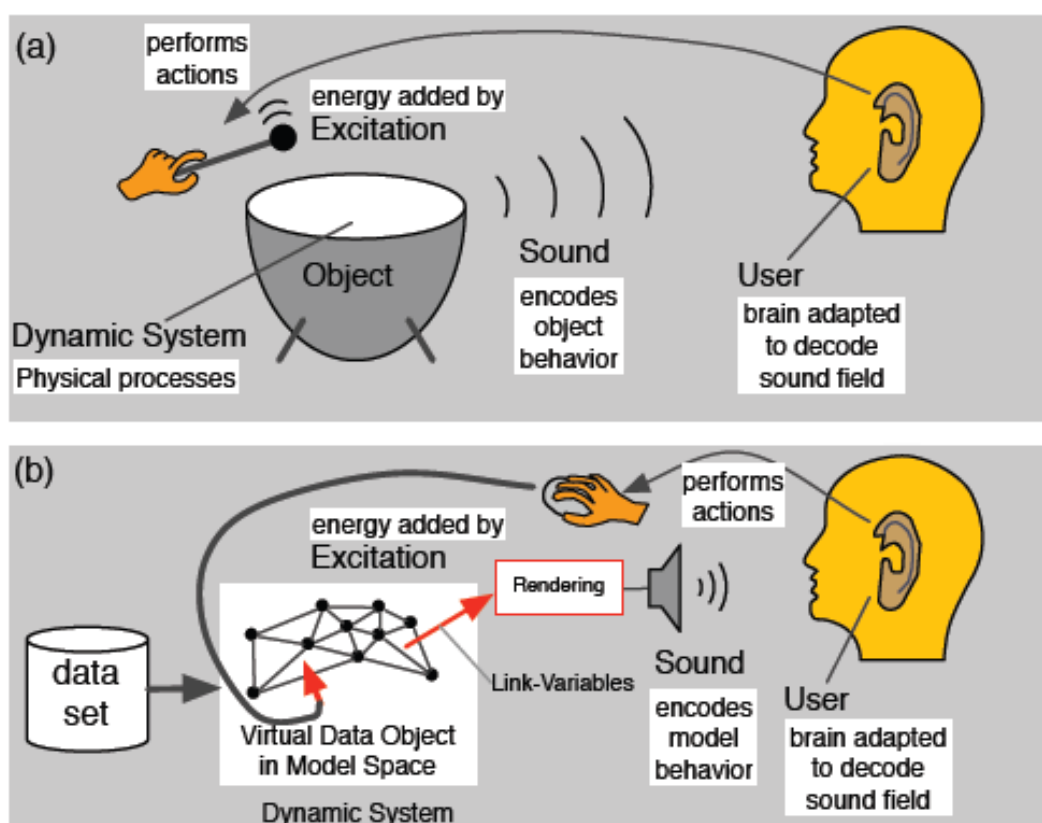
Thomas Hermann ve své definici sonifikace. Do určité míry lze souloví „*mapování parametrů*“ chápat jako akcentování faktu, že sonifikace je především hledáním určitého *kódu* pro převedení libovolné informace do zvukové podoby.

Možné topologie mapování parametrů:

- Jedna informace – jeden parametr zvukového signálu.
- Jedna informace – více parametrů zvukového signálu.
- Více informací – jeden parametr zvukového signálu.

3) Modelově založená sonifikace

Tento způsob sonifikace vychází z faktu, že člověk přirozeně vnímá zvuk jako odezvu okolí (či objektu) na určitou aktivitu. Podstatou modelově založené sonifikace je vytvoření modelu (objektu) podle určité datové struktury. Model tedy sám o sobě zvuk netvoří, ale vytváří odezvu na aktivitu uživatele.



Obrázek 6: Schéma modelově založené sonifikace, kde v části a) je znázorněn vznik zvuku jako odezvy reálného systému na lidskou činnost, na obrázku b) je znázorněn vznik zvuku jako odezvy sonifikačního algoritmu na lidskou činnost [14]

Pro lepší popis modelově založené sonifikace nám poslouží obrázek 6. V reálném světě bychom ze zvuku, který je odezvou bubnu na úder paličkou, analyzovali vlastnosti bubnu (obrázek 6a). V případě modelově založené sonifikace analyzujeme vlastnosti dat ze zvukové odezvy virtuálního modelu (obrázek 6b).

Vzhledem k zaměření sonifikace na člověka – uživatele, je realizace vždy průsečíkem objektivního přístupu s uživatelsky nejpřirozenějším a nejpříjemnějším (tedy subjektivním) přístupem.

3.2 Různé způsoby sonifikace videosignálu a pohybu

V této části následuje popis několika sonifikačních metod určených k transformaci obrazu do zvukové podoby, na konci této kapitoly popíšeme některé způsoby sonifikace pohybu.

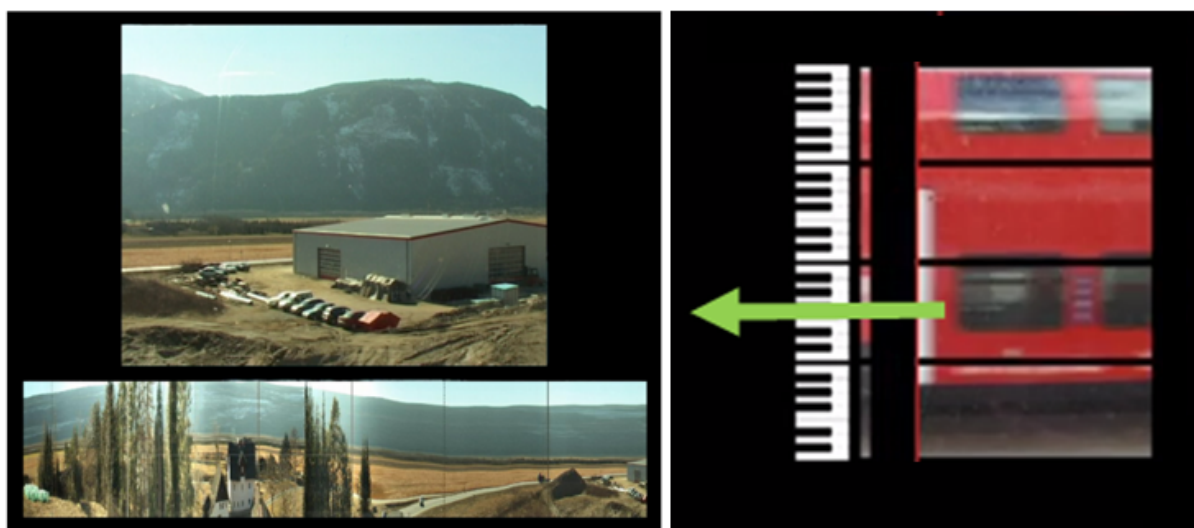
3.2.1 Theremin jako předchůdce sonifikace

Za jeden z prvních pokusů o sonifikaci polohy a pohybu objektu bychom mohli označit nástroj jménem „Theremin.“ Jedná se o bezdotykový hudební nástroj, který v první polovině dvacátého století navrhl Ruský vynálezce Těrmen.

Princip nástroje je založen na přeladování zánějů dvou oscilátorů vlivem kapacity lidské ruky. Výška tónu je řízena vzdáleností od jedné antény, hlasitost vzdáleností druhé ruky od druhé antény [15].

3.2.2 Cesta vlakem a Skrjabinův kód

Skupina studentů University Johannesse Keplera ve složení Knees, Pohle a Wildmer navrhla algoritmus, umožňující real-time transformaci obrazu snímaného videokamerou z okna vlaku do zvukové podoby [16]. Snahou bylo vytvořit takový systém, který bude reflektovat státnost cesty, ale zároveň při každé nové cestě vznikne unikátní hudební kompozice.

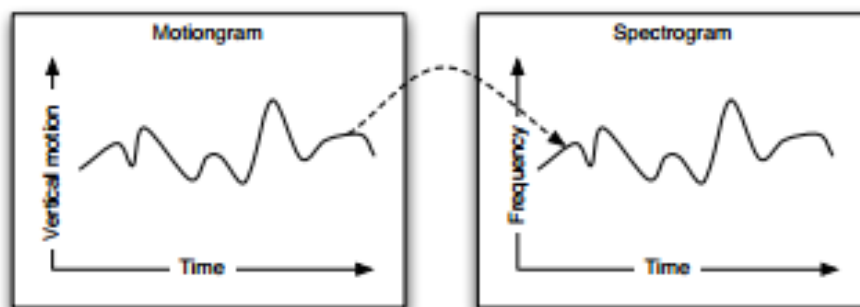


Obrázek 7: Snímek algoritmu sloužícího pro sonifikaci cesty vlakem: v levé horní části vidíme současný snímek, pod ním zhuštěný záznam cesty a vpravo znázornění přiřazení oktáv k jednotlivým skupinám pixelů [16]

Princip systému je zřejmý z obr. 7. Nejprve je z videosignálu vybrán každý sedmý snímek, čímž vzniká zhuštěná kopie cesty i základní rytmická struktura skladby. Z tohoto snímku je dále vybrán prostřední sloupec pixelů, který je následně rozdělen tak, aby každý díl bylo možné přiřadit jednomu tónu z celkových čtyř oktáv. Mapování barvy na příslušné tóny odpovídá kódu Alexandra Skrjabina [17], což byl ruský skladatel a filosof, který se snažil o propojení hudby s filosofií a vědou, kde inspirován Newtonovou Optikou navrhl systém přiřazující barvám jednotlivé tóny⁴.

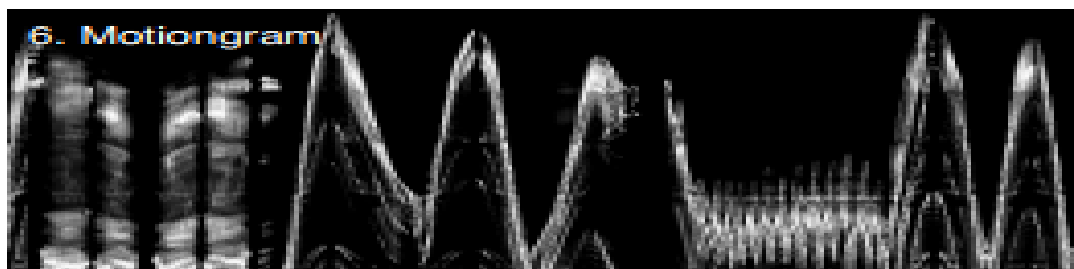
3.2.3 Převod grafu pohybu na spektrogram

Výzkum A. R. Jensenia z univerzity v Oslu se zabývá převodem informace o pohybu z videosignálu do zvukové podoby [18]. K této práci je využíváno prostředí MAX a rovněž jako v předchozím případě se jedná o real-time transformaci.



Obrázek 8: Znázornění základní myšlenky sonifikačního algoritmu: převod motiongramu na spektrogram, kde na vodorovných osách je vynesena čas a dochází k převodu vertikální polohy na frekvenci [18]

Základem algoritmu je převod *motiongramu* (grafu polohy objektu) na spektrogram zvukového signálu, kde vertikální poloha objektu určuje centrální frekvenci pásmové propusti. Bodu, který se pohybuje u spodního okraje videa, náleží hluboké frekvenční složky, bodu pohybujícímu se u horního okraje vyšší frekvence. Rozsah frekvencí pro plný rozsah hodnot na vertikální ose není v práci zmíněn a horizontální pohyb není vůbec uvažován.



Obrázek 9: Motiongram (graf pohybu) [18]

⁴ Záznamy z různých cest je možné nalézt na: <http://www.cp.jku.at/projects/soundtracks/>.

Algoritmus pracuje ve čtyřech krocích. Nejprve je videosignál převeden do černobílé, poté je vypočítán rozdíl dvou po sobě následujících snímků. Následně je zjištěna poloha pohybujícího bodu a ta je pomocí inverzní Fourierovy transformace převedena na zvukový signál.

3.2.4 Spojení nervových center a sonifikace pohybu paže

Práce Vinkena a kol. [19] se zabývá otázkou, do jaké míry jsou nervové oblasti zodpovědné za vnímání pohybu a jeho zvukové podoby propojené. Je člověk schopen rekonstruovat pohyb pouze ze zvukového záznamu?

Pro účely experimentu bylo vybráno šest „základních“ pohybů paže člověka sedícího u stolu a ty byly následně rozděleny do tří skupin dle svého charakteru:

- Kategorie I: pomalé cyklické pohyby (*malování kruhu, míchání v hrnci*).
- Kategorie II: pomalé lineární pohyby (*zvedání sklenice a pití, lití vody do hrnce*).
- Kategorie III: rychlé cyklické pohyby (*čištění zubů, pilování nehtů*).

Vlastní sonifikace nabývala sedmi možných podob, přičemž se jednalo o různé kombinace čtyř základních zvukových parametrů, kterými byl pohyb mapován. Pomocí amplitudy byla mapována dynamika pohybu, výška tónu reprezentovala elevaci, frekvenční složení vzdálenost od uživatele a pohyb v horizontální rovině byl mapován stereofonií zvuku.

Závěry, které vyplynuly z experimentu, bychom mohli shrnout do tří hlavních bodů. Na počátku byla správnost rozpoznání pohybu pouze z jeho zvukové reprodukce rovna 42%, zatímco na konci experimentu (zhruba po čtyřiceti minutách průběžných testů) se správnost zvýšila na 62%. Je důležité připomenout, že testovaní během experimentu nebyli nijak informováni o správnosti odpovědí, což poukazuje na rychlou schopnost učení se v této oblasti. Další závěr, který z testu vyplynul, je fakt, že takřka nezáleží na tom, jaké parametry sonifikace využívá (tedy o jakou kombinaci *amplitudy, frekvence, stereofonie a spektrálního složení* se jedná), protože správnost odpovědí nabývala pro všechny kombinace téměř stejných hodnot. Experiment Vinkena a kol. potvrdil, že je pro člověka přirozené vnímat zvukovou podobu pohybu a pohyb samotný jako dvě spojené části.

Z výsledků experimentu vyplývá, že nervová centra zodpovědná za pohybový a sluchový vjem jsou propojená, což otevírá dveře novým rehabilitačním metodám, kdy například vhodnou stimulací sluchových center můžeme urychlit rekonvalescenci po cévní mozkové příhodě, či zpomalit šíření onemocnění u lidí trpících Alzheimerovou chorobou.

3.2.5 Rehabilitace po CMP a sonifikace pohybu

Zatímco výše zmíněná studie zkoumá, do jaké míry je možné mluvit o propojení sluchových vjemů s řízením pohybu, práce *Real-Time Sonification of Movement for an Immersive Stroke Rehabilitation Environment* Willise a kol. [20] se již zabývá konkrétní implementací sonifikace pro účely rehabilitace po cévní mozkové příhodě a to konkrétně

stimulací pohybů paže pomocí zvukově obrazového virtuálního prostředí.

Pro toto využití byl navržen systém, kde pohyb paže pacienta je sledován šesti kamerami a následně mapován na několik hudebních nástrojů. Virtuální prostředí předává pacientovi zvukovou i obrazovou zpětnou vazbu vztaženou k pohybu paže.

Rehabilitace zde nabývá podobu hry, jejímž cílem je virtuální paží (která přesně sleduje pohyb reálné paže) „uchopit“ virtuální kelímek. Poté, co je pacient seznámen s touto formou vizuální podoby prostředí, je ruka s kelímkem nahrazena abstraktním obrazem, jehož pohyb a rotace odpovídají přesnosti, s jakou se pacient blíží, či vzdaluje „zóně uchopení kelímku“ – tzn. místu, které je cílem pohybové akce.

Vzhledem k tomu, že rehabilitace je obvykle dlouhodobý proces, vyžadující aktivní účast a motivaci pacienta, je důležité, aby algoritmus byl z uživatelského hlediska co nejpohodlnější, a zároveň pacientovi v reálném čase předával zpětnou vazbu, která podporuje jeho výkonnost a snahu o další užívání vedoucí k urychlení rehabilitace.

4 PRAKTICKÁ ČÁST

4.1 Návrh algoritmu

Pro návrh vlastní struktury algoritmu jsme vyšli z článků a studií uvedených v předchozí kapitole, kde byl kladen důraz především na sonifikační řešení založená na principu převodu elevace na výšku tónu, což je způsob, který budeme využívat i v naší práci. Pro převod elevace na výšku tónu jsme se rozhodli z těchto důvodů:

- **Ve srovnání s algoritmy simulujícími HRTF je převod elevace na výšku tónu přesnější.**

- *více o tomto problému v kapitole 3.5 o převodu elevace na výšku tónu a v práci A. Štolce [26]*

- **Jedná se o velice intuitivní převod elevace na jinou veličinu.**

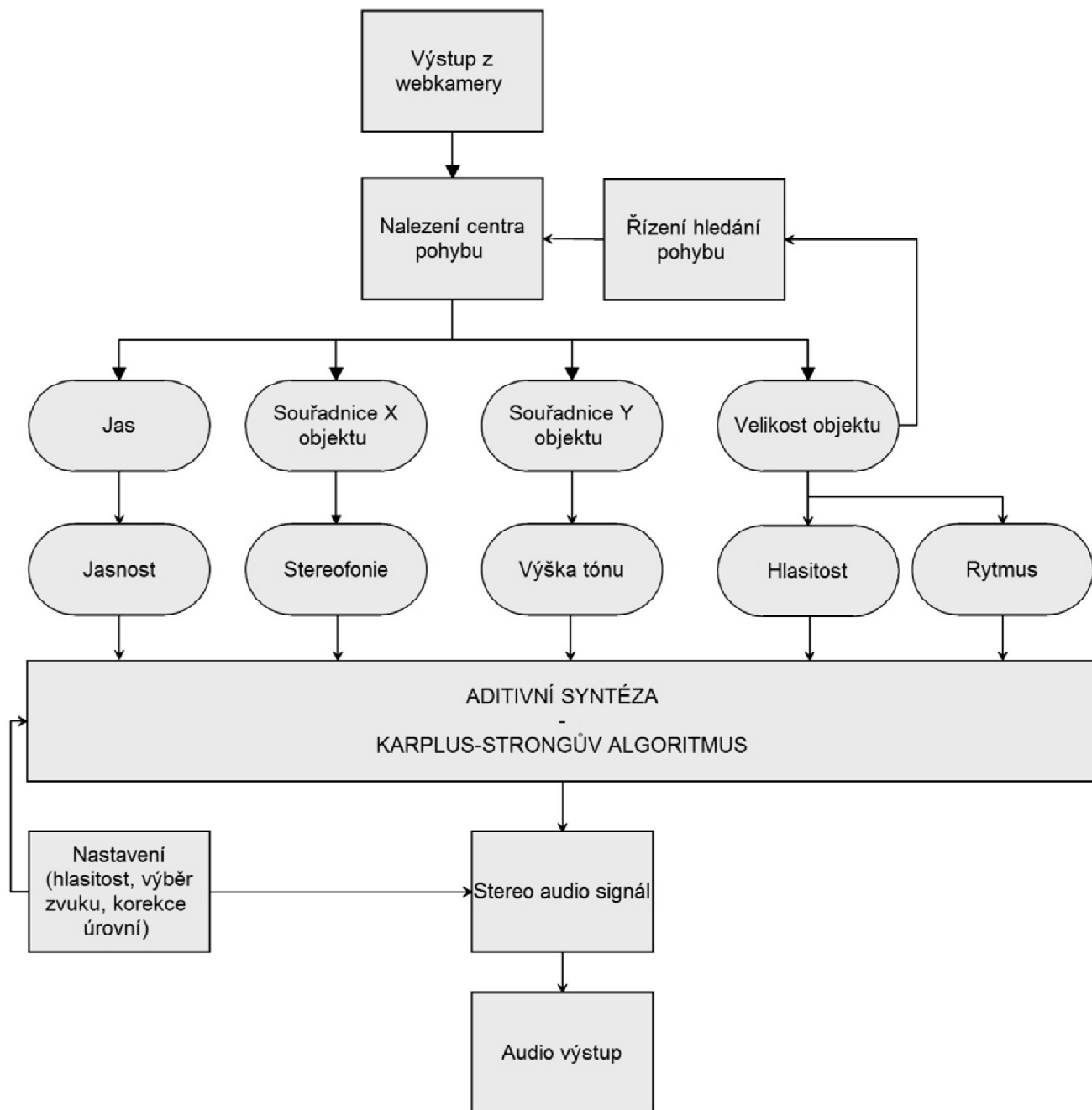
- *více v kapitole 1.5 o práci C. Prattové a E. Rusconiové a kol. [9]*

Velikost azimutu je mapována na stereofonii signálu, což je způsob, který využívá práce Vinnkena a kol. [19]. V našem případě se jedná o implementaci ITD a ILD. Hlavním důvodem, proč jsme se rozhodli nepřevádět informaci o azimutu na jinou zvukovou veličinu, byla snaha zachovat algoritmus uživatelsky co nejsrozumitelnější a nejintuitivnější.

V další části algoritmu jsme vyšli z kapitoly 1.6 a především ze zmiňované práce Colliera a Hubbarda [10] o vztahu mezi jasem obrazu a zvukovými parametry, načež jsme přistoupili k převodu jasů objektu na jasnost zvuku, tak jak ji popisuje V. Syrový v publikaci Hudební akustika [11].

Rychlost pohybu je převedena na amplitudu vytvářeného zvukového signálu a na rytmus zvuku tvořeného pomocí Karplus-Strongova algoritmu.

Jednotlivé části algoritmu včetně detailnějších motivací budou blíže rozebrány v následujících kapitolách. Pro lepší orientaci v textu je na následující straně přiložen obrázek znázorňující strukturu algoritmu.



Obrázek 10: diagram znázorňující strukturu sonifikačního algoritmu, kde po nalezení polohy pohybujícího se objektu jsou souřadnice použity pro mapování výšky tónu a stereofonie, jas objektu určuje jasnost tónu a velikost pohybu rytmus tvoření Karplus-Strongova algoritmu a amplitudu signálu, dále následuje vlastní syntéza a řízení algoritmu uživatelem

4.2 Motivace použití PD a zpracování videosignálu

Práce A. Jensenia [18] popisuje převod videa do zvukové podoby v prostředí MAX, což je komerční verze PD, přičemž obě prostředí vznikla především pro usnadnění práce s datovými toky v reálném čase. Abychom mohli snadno pracovat s obrazem snímaným webkamerou počítače, využíváme PD i v naší práci.

4.2.1 Získání souřadnic pohybujícího se objektu

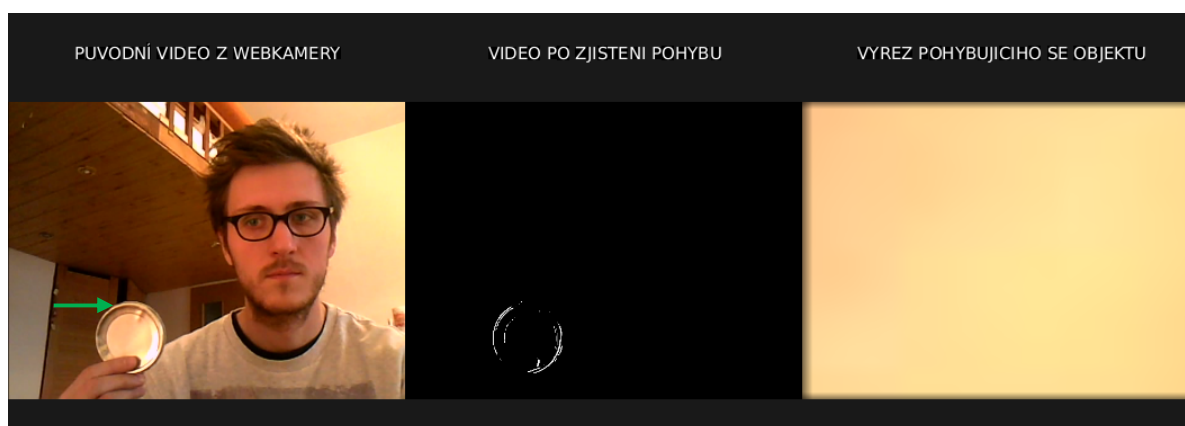
Pro práci s obrazovými daty slouží v PD knihovna GEM [21], ze které je pro naši aplikaci klíčový objekt [pix_video], sloužící k propojení webkamery počítače s PD. *Pokoušeli jsme se připojit k programu více webkamer, abychom rozšířili zpracovávané zorné pole, jak jsme však zjistili, je to možné pouze v komerčním prostředí MAX.*

Objekt, který následuje za výstupem webkamery se nazývá [pix_movement] a slouží k nalezení pohybu ve videu. Objekt [pix_movement] počítá diferenci dvou po sobě následujících snímků, přičemž se jedná o časově založený IIR filtr. Informace o pohybu je uložena do alfa kanálu videa, a ten je odeslán na obrazovku monitoru a pro určení polohy pohybujícího se objektu je rovněž odeslán na vstup objektu [pix_blob].

Objekt [pix_blob] počítá souřadnice s největší koncentrací aktivních pixelů, a protože je na vstup objektu přiveden videosignál, který nese pouze informaci o pohybu (viz prostřední rámeček obrázku 11), je výstupem objektu [pix_blob] informace o poloze pohybujícího se objektu a o velikosti pohybu (kterou je rovněž možné vnímat jako velikost objektu).

Řízení vyhledávání pohybu

Pokud se hodnota velikosti pohybu blíží maximu, může docházet k výraznému zhoršení určení polohy, proto je v algoritmu zařazena zpětná vazba, která zajišťuje, že když se hodnota velikosti pohybu blíží maximu, dojde k zvětšení prahové hodnoty, která určuje, kolik sousedních pixelů v po sobě následujících snímcích se musí změnit, aby tato změna mohla být považována za „pohyb.“ Stejným způsobem jsou odstraněny obrazové šumy.



Obrázek 11: Snímek z běhu algoritmu při zpracování videa; v levé části vidíme obraz před zpracováním s naznačeným směrem pohybu víčka, uprostřed zpracované video po nalezení pohybu a vpravo vidíme výřez pohybujícího se objektu

4.3 Spojení jasu a jasnosti (experiment a implementace)

Jak bylo zmíněno v kapitole 1.6, Beauchamp zavedl veličinu zvanou *jasnost* zvuku. Naším cílem bude mapovat *jas* pohybujícího se objektu na *jasnost* zvuku. Hlavní motivací pro nás bylo hledání analogií mezi obrazovou a zvukovou smyslovou oblastí.

4.3.1 Subjektivní testy spojování *jasu* a *jasnosti*

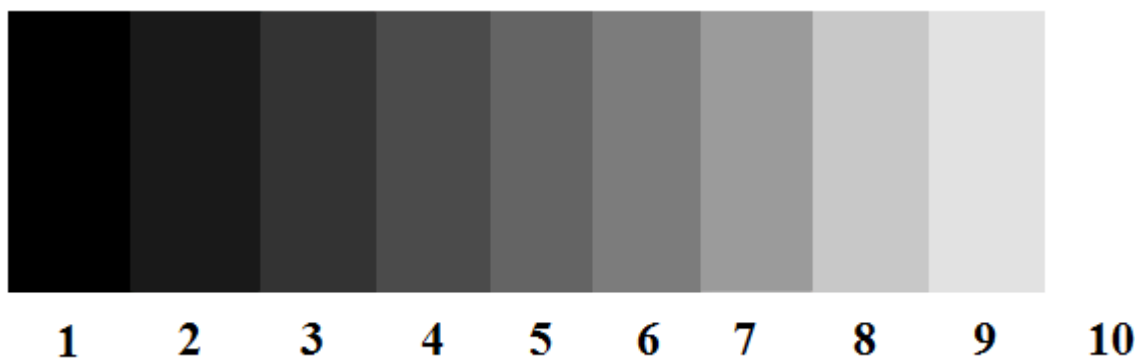
Před tím, než jsme přistoupili k implementaci, provedli jsme test, který měl ověřit, či vyvrátit předpoklad, že člověk podvědomě spojuje tóny vyšší jasnosti s vyšším jasnem obrazu. K tomuto testu jsme přistoupili proto, že studií na toto téma zatím mnoho nevzniklo.

Průběh testu

Na začátku byl účastníkům experimentu předán obrazový vzor skládající se z deseti obdélníků různých stupňů šedi od černé ($R=0$, $G=0$, $B=0$) do bílé ($R=255$, $G=255$, $B=255$), přičemž barva každého z deseti obdélníků byla vypočtena podle následujícího vzorce.

$$R_{k+1} = G_{k+1} = B_{k+1} = \frac{255}{9} * k; k \in \langle 0; 9 \rangle ; \quad (4)$$

Součástí vzoru bylo číslování od jedné do deseti, kde černá měla hodnotu jedna, bílá deset, jak je možno vidět na následujícím obrázku.



Obrázek 12: vzor deseti stupňů šedi použitý pro testování

Stimul délky 3s s náběžnou a sestupnou hranou délky 100ms měl po celou dobu experimentu konstantní frekvenci a amplitudě základní harmonické 440Hz. Jasnost jednotlivých stimulů byla náhodně měněna. Mezi jednotlivými stimuly byla časová prodleva v délce 10s.

Před začátkem testu bylo účastníkům přehráno deset stimulů náhodných jasností, aby tak získali představu, v jakém spektru zvuků se bude test pohybovat. Každý účastník analyzoval dvacet různých stimulů. Úkolem testované osoby bylo přiřadit každému tónu jeden stupeň šedi. Posluchačům nebyl před začátkem sdělen cíl, ani předpokládaný výsledek testu. Experiment byl proveden na šesti osobách ve věku 15 – 50 let. Všichni

uvedli, že nemají žádné sluchové vady a že se nikdy neúčastnili podobného testu.

Cíle experimentu

Cílem experimentu nebylo nalezení konkrétních hodnot jasu, které odpovídají konkrétním hodnotám jasnosti, ale snažili jsme se zjistit, zda existuje nějaká závislost ve spojování jasnosti a jasu, případně zda je tato závislost obdobná pro více testovaných osob. Zjišťovali jsme závislost změny jasu na změně jasnosti.

Výsledky experimentu a komentáře testovaných osob

Výsledky ukázaly, že *pět z šesti osob pro vyšší jasnost tónu volilo vyšší jas obrazu*. Test rovněž ukázal, že všechny naměřené průběhy mají charakter lineární závislosti. Testování si během krátké chvíle dokázali vytvořit svůj vlastní klíč použitý pro vztahování jasnosti k jasu, přičemž tento klíč měl vždy charakter lineární závislosti mezi změnou jasnosti a vnímáním stupňů šedi.

Pět z šesti testovaných osob se v komentářích k průběhům testů vyjadřovalo velice obdobně ve smyslu, že jasnějším tónům přiřazovali světlejší odstíny z toho důvodu, že zvuk obsahující vyšší podíl vyšších harmonických složek považovali za analogický k bílému světlu, které je rovněž složeno z širokého spektra frekvenčních složek.

Jedna testovaná osoba spojovala jasnější tóny s nižšími hodnotami jasu, avšak i v tomto případě měl zjištěný průběh charakter lineární závislosti. Při výkladu této výjimky vycházíme z komentáře uvedeného k průběhu měření. Zmíněná osoba uvedla, že tón, který bychom mohli označit za „čistší“ co do spektrálního složení (v extrémním případě pouze sinusový průběh) odpovídá bílé barvě proto, že „právě bílá barva je symbolem lehkosti a čistoty“ (citace komentáře k průběhu testu).

Jedna z testovaných osob uvedla, že nejpřirozenějším klíčem pro spojení jasu obrazu a jasnosti zvuku je takové mapování, kde tmavším barvám odpovídají hluboké tóny a světlejším barvám tóny vyšší. Tento způsob mapování ovšem neodpovídá experimentům popsaným v kapitole 1.5, které mluví o výšce tónu ve vztahu s elevací, což je zároveň spojení, kterého využíváme i v naší práci.

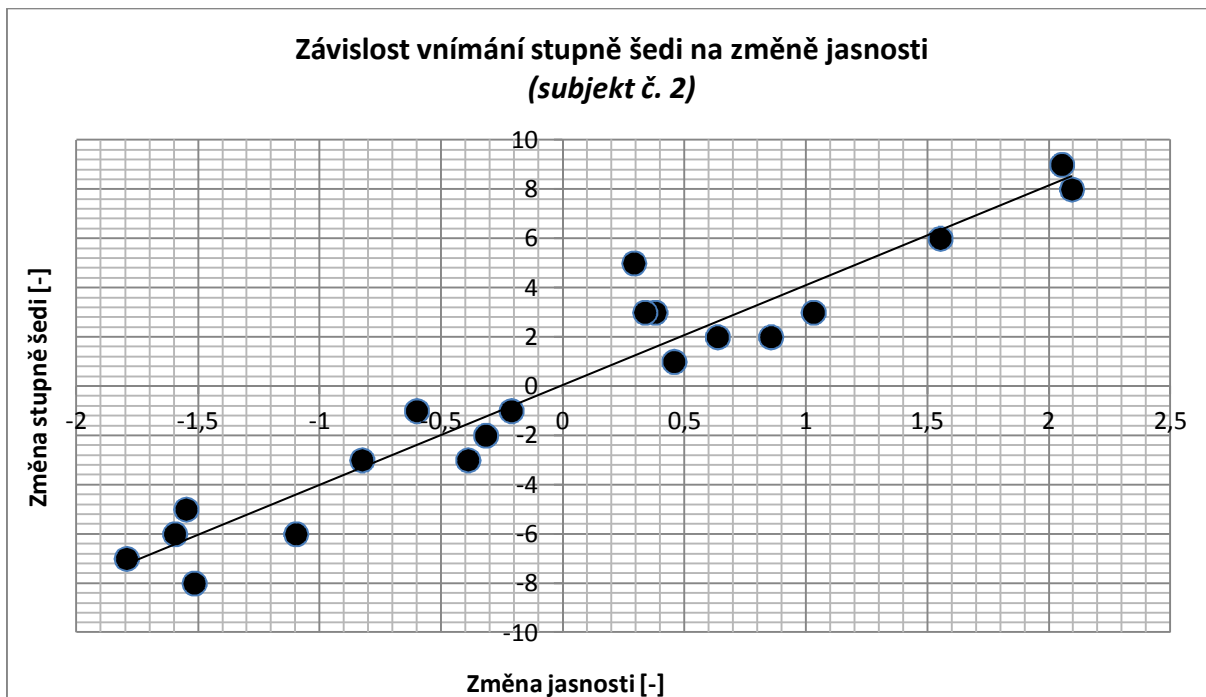
Změřené závislosti vnímání změny stupně šedi na změně jasnosti tónu

Na obrázku 13 můžeme pozorovat graf závislosti změny stupně šedi na změně jasnosti. Ve všech grafech záměrně neuvádíme hodnotu jasu, ale vycházíme z námi zvolené stupnice, kterou jsme přiřadili jednotlivým stupňům šedi (1 až 10).

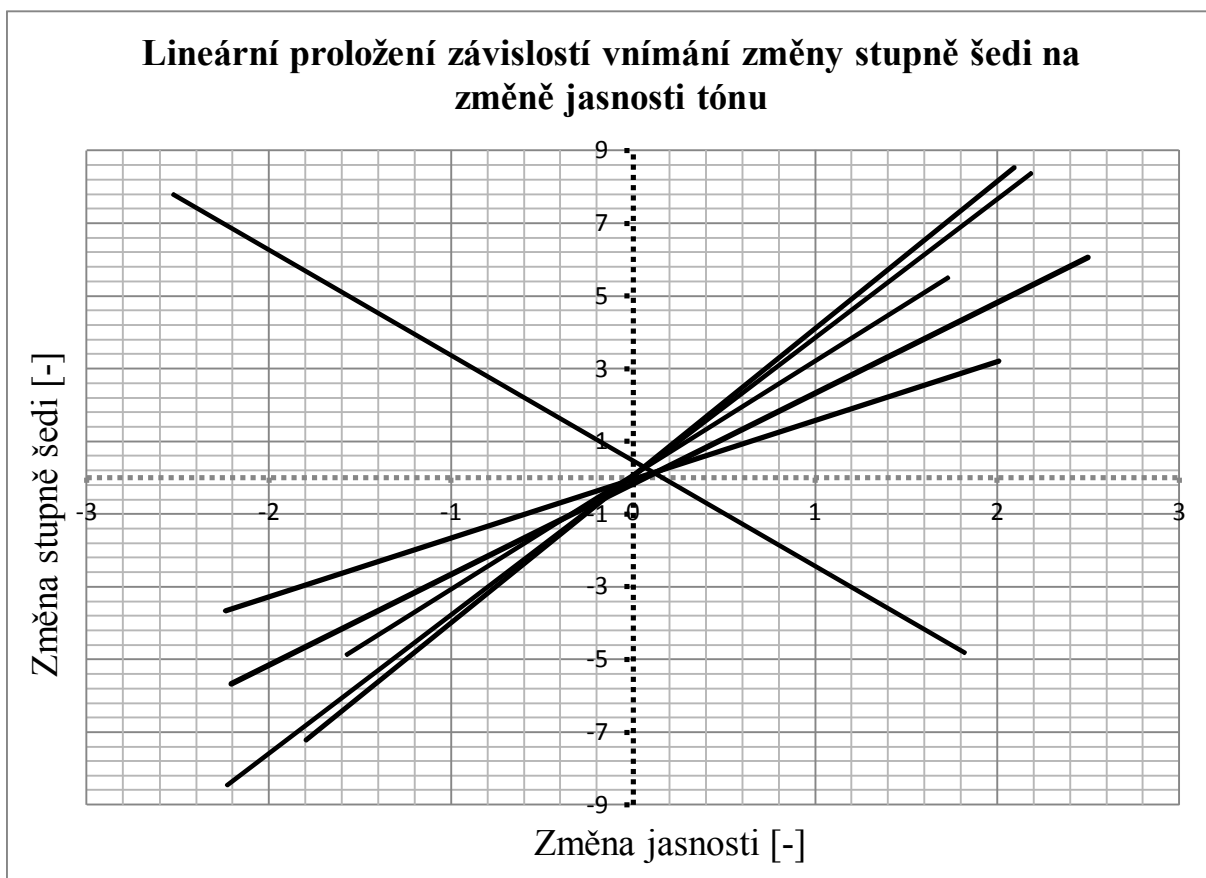
Vztah mezi jasem a námi zvolenou stupnicí šedi je lineární.

Na obrázku 14 můžeme vidět graf, ve kterém jsou vynesena veškerá provedená měření, s tím rozdílem, že pro větší přehlednost již nejsou zobrazeny konkrétní hodnoty, ale v grafu jsou vyneseny pouze lineární proložení naměřených hodnot.

Tabulku všech naměřených hodnot včetně jednotlivých grafů a patch sloužící k tomuto testu naleznete na přiloženém CD pod názvy „TEST_jasnost_vysledky.xls“ a „poslechovy_test_jas.pd“. Grafy naleznete rovněž v tištěné příloze práce.



Obrázek 13: graf měření subjektu č. 2 závislosti vnímání změny stupně šedi v závislosti na změně jasnosti tónu, v grafu je rovněž vyneseno lineární proložení naměřených hodnot



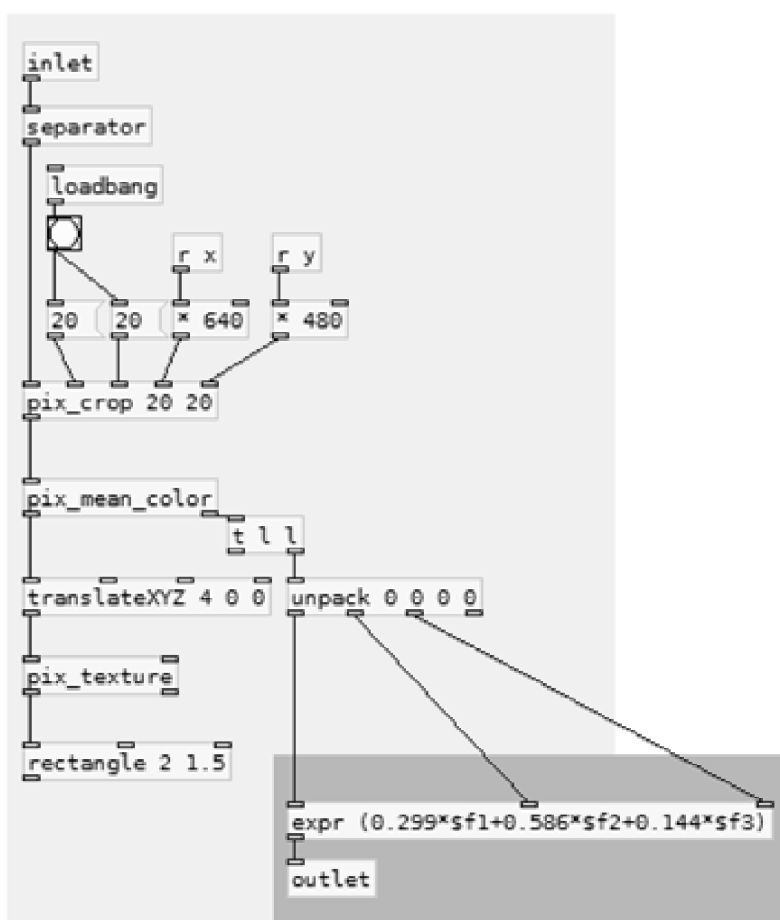
Obrázek 14: graf zobrazující vnímání změny jasu jako funkce jasnosti, v grafu jsou pro lepší přehlednost vyneseny lineární proložení změřených hodnot

4.3.2 Implementace mapování jasu

Vzhledem k výsledkům experimentu jsme přistoupili k mapování *jasové složky Y* na *jasnost tónu*.

Zjištění barvy a jasu pohybujícího se objektu

Polohu pohybujícího se objektu jsme našli v kapitole 3.2, jeho souřadnice použijeme k další analýze. Barvu a jas objektu zjistíme tak, že z původního videa vyjmeme 20x20 pixelů z místa, kde se objekt nachází. K tomu slouží objekt [pix_crop], jehož vstupními parametry jsou velikost výřezu a poloha objektu vztažená ke konkrétnímu rozlišení videa. Průměrné barevné složky tohoto výřezu zjistíme pomocí objektu [pix_mean_color], odtud poté vypočteme hodnotu jasové složky *Y* podle vztahu (3) uvedeného v kapitole 1.6.1.



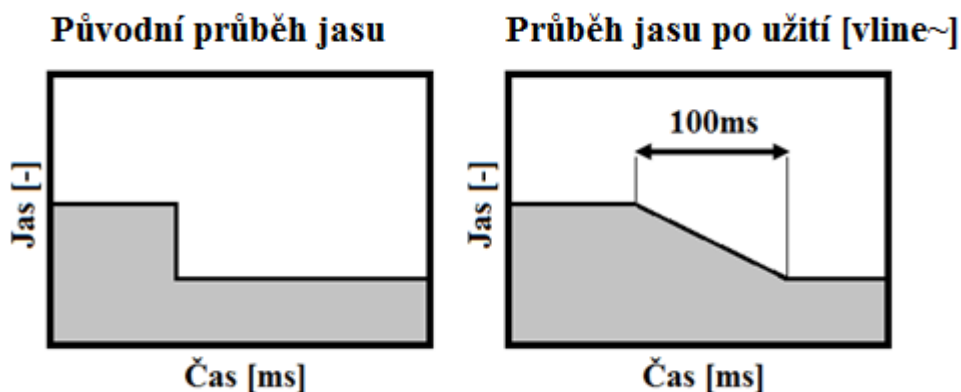
Obrázek 15: Část patche sloužící ke zjištění jasové složky pohybujícího se objektu; objekt [pix_crop] má úlohu získání výřezu videa, objekt [pix_mean_color] počítá průměrné barevné složky výřezu, z nich je poté pomocí objektu [expr] vypočtena jasová složka

3.2.2.1 Lineární aproximace řídicí hodnoty jasu

Abychom předešli skokovým změnám hodnoty jasu, ze kterých by mohly pramenit nežádoucí zvukové artefakty, nastavíme systému pevně danou hodnotu času, během které

ke změně jasů dochází. Změna nikdy není skoková, nýbrž je aproximována lineárním průběhem trvajícím 100ms.

Pro tento účel využíváme objekt [vline~], který při každé inicializaci vytvoří posloupnost hodnot počínaje současnou hodnotou po inicializační hodnotu v námi zvoleném čase (tedy během již zmíněných 100ms). Stejněho principu využíváme v práci vždy, když potřebujeme eliminovat skokové změny hodnot.



Obrázek 16: Ilustrace uměle vloženého přechodu mezi dvě hodnoty jasů sloužící k odstranění skokových přechodů dvou hodnot

Konkrétní způsob, jakým je hodnota jasové složky následně mapována na jasnost tónů, je uveden v následující kapitole, kde jsou popsány veškeré formy zvukové syntézy, kterých v této práci využíváme. Pro lepší přehlednost uvádíme mapování jasové složky v rámci těchto částí.

4.4 Mapování elevace na výšku tónu (experiment a implementace)

4.4.1 Subjektivní vnímání výšky tónu jako funkce elevace

Před tím, než přistoupíme k implementaci převodu elevace na výšku tónu, provedeme několik experimentů vycházejících z práce C. Prattové [9] a dalších, a pokusíme se zjistit přesnost, s jakou je možné použít výšky tónu pro orientaci v mediální rovině. Dále se budeme zabývat otázkou vlivu kalibrační posloupnosti na orientaci.

Průběh testu

Testovaným osobám byly postupně přehrávány jednotlivé stimuly a jejich úkolem bylo posvítit na místo v prostoru, které podle nich nejlépe reprezentovalo daný tón. Pro tento účel jsme připevnili baterku na kamerový stativ, jehož pohyb byl omezen pouze na změnu vertikálního směřování světla. Šířka světelného svazku byla omezena tak, že v místě dopadu vznikal obrazec ve tvaru kruhu o průměru 20cm. Celý test probíhal ve tmě, nejnižší možná poloha se nacházela na úrovni podlahy, nejvyšší na úrovni stropu, hodnota výšky mohla nabývat 0 – 270cm. Stimuly neměly charakter harmonického průběhu, ale jednalo se o tóny stejného frekvenčního složení, jaké bylo následně použito při tvorbě algoritmu. K tomuto řešení jsme přistoupili proto, abychom tímto testem zároveň zjistili přesnost vnímání elevace navrženého algoritmu. Zvukovou ukázkou jednoho stimulu i celé stupnice je možno nalézt na přiloženém CD pod názvem „stimul_elevace.mp3.“ a „stupnice_elevace.mp3.“ Zvuk byl přehráván ze sluchátek.

Modifikace testu č. 1

Stimul délky 5s s náběžnou a sestupnou hranou délky 100ms nabýval pěti možných tónů vzdálených od sebe jednu oktávu. Jednalo se o tóny C, c, c1, c2, c3, jejichž pořadí bylo náhodně měněno. Konkrétní frekvence a čísla MIDI značek naleznete v tabulce níže. Mezi stimuly byla časová prodleva v délce 10s. Závislost byla měřena pro dva bloky po pěti stimulech.

Tón	MIDI	Frekvence [Hz]
C	36	65,4063913251
c	48	130,8127826503
c1	60	261,6255653006
c2	72	523,2511306012
c3	84	1 046,5022612024

Tabulka 1: Frekvence a MIDI hodnoty jednotlivých tónů [22]

Modifikace testu č. 2

Délka stimulu i princip testu byl totožný jako u modifikace č. 1, ovšem namísto pěti tónů bylo frekvenční pásmo mezi C a c3 rozděleno tak, že výšky tónů náležely do diatonické stupnice C dur. Takovéto rozložení tónů si můžeme představit jako bílé klapky klavíru v rozsahu čtyř oktáv od C po c3. Závislost byla měřena pro deset stimulů.

Modifikace testu č. 3

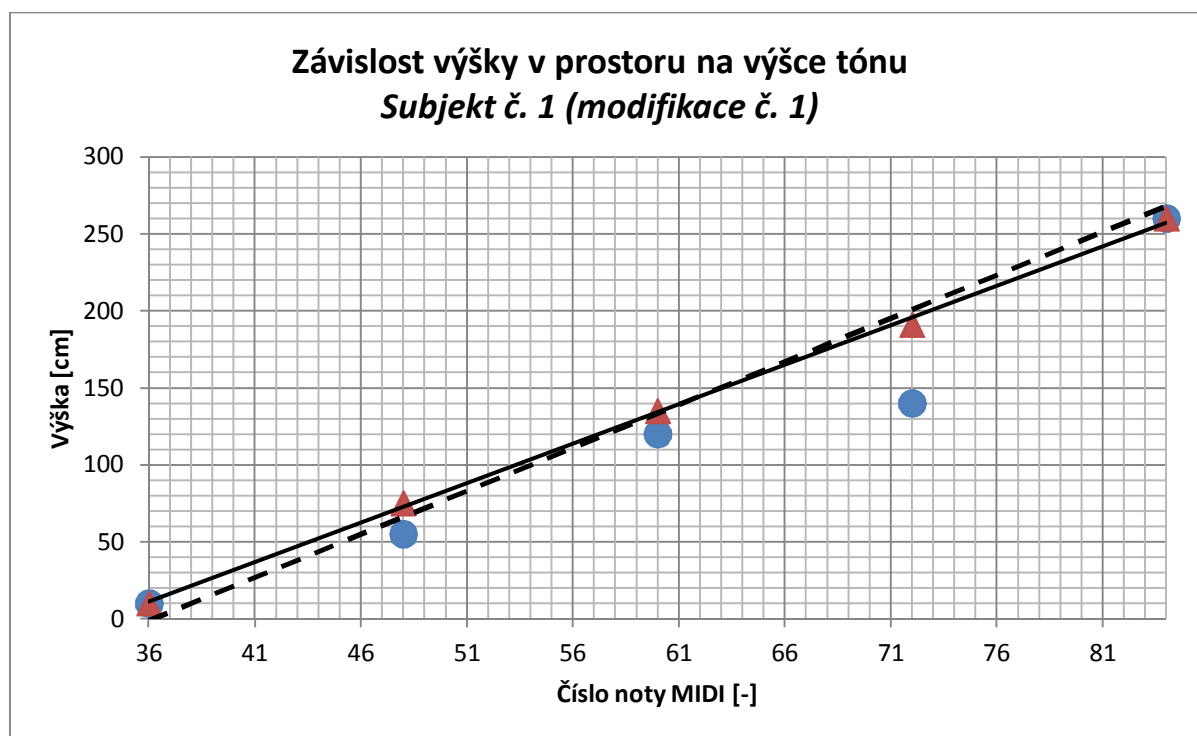
Poslední modifikace testu vychází z té předchozí, ovšem po každém třetím stimulu byla testovaným osobám přehrána „kalibrační posloupnost“ tónů $C - c - c1 - c2 - c3$, kde velké C určovalo nejnížší možnou polohu v prostoru, tříčárkované $c3$ určovalo polohu nejvyšší. Každý tón z této posloupnosti trval 0,5s. Cílem této modifikace bylo zjistit, do jaké míry je člověk schopný „kalibrovat“ svůj sluch a zpřesnit výsledky na základě přehrávané posloupnosti tónů, která dělí prostor do stejných intervalů. „Kalibrační posloupnost“ se nachází na přiloženém CD pod názvem „kalibrační_posloupnost_elevace.mp3.“

Posluchačům nebyl před začátkem sdělen cíl, ani předpokládaný výsledek testu. Experiment byl proveden na pěti osobách ve věku 22 – 50 let. Všichni uvedli, že nemají žádné sluchové vady, nikdo z nich se dříve neúčastnil obdobného testu. Všichni se testu účastnili dobrovolně.

Výsledky experimentu

V kapitole 1.5 bylo zmíněno několik experimentů zabývajících se vztahem výšky tónu a umístěním v prostoru. Všechny tyto práce došly k závěru, že člověk podvědomě spojuje nižší umístění v prostoru s nižšími hodnotami frekvence.

Pro ověření tohoto předpokladu byla klíčová hned první modifikace testu, protože právě v jejím průběhu si testovaní vytvářeli „klíč“ pro převod výšky tónu na hodnotu elevace. Na obrázku níže můžeme vidět výsledek měření prvního subjektu první modifikace testu.

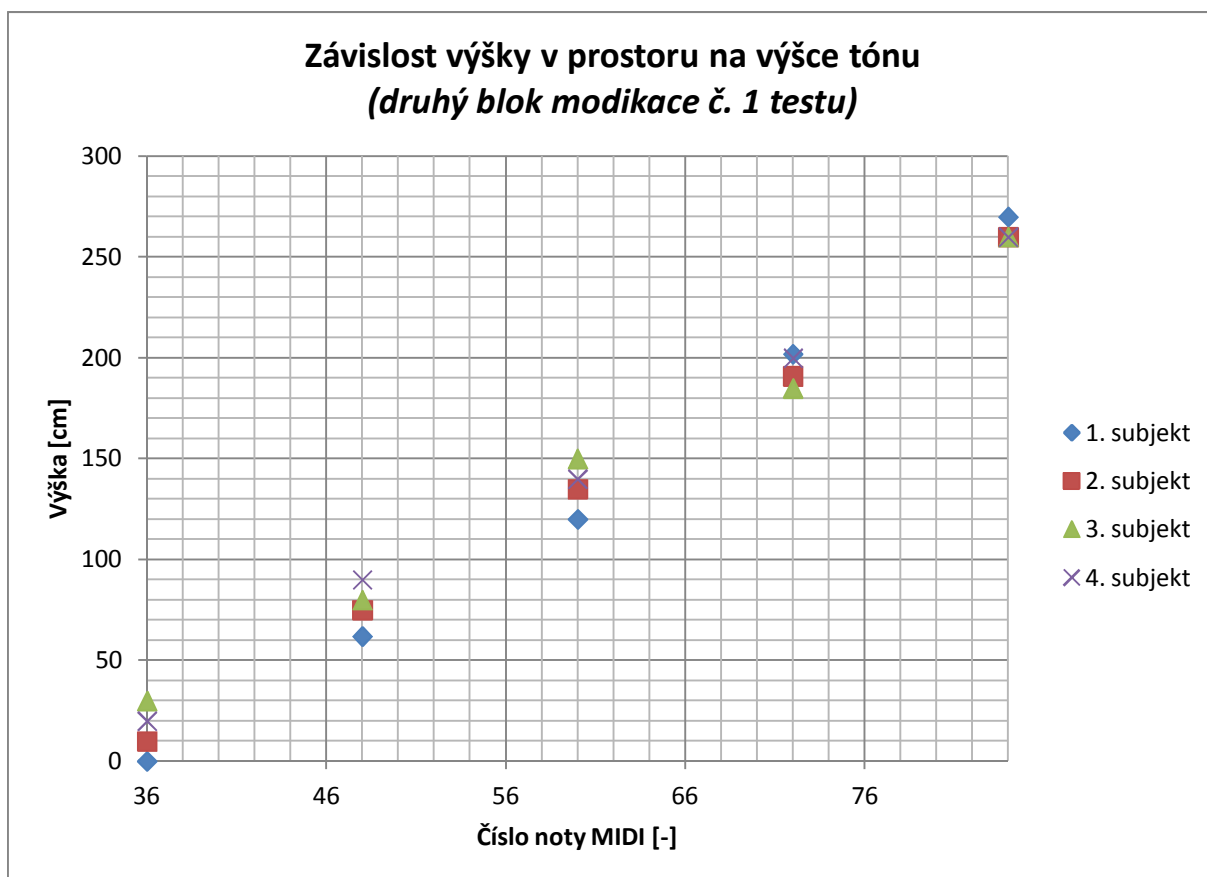


Obrázek 17: Graf závislosti umístění v prostoru na výšce tónu, na vertikální ose je vynesena výška tónu ve formátu MIDI, na horizontální ose výška v prostoru, modré body znázorňují první polovinu měření, červené druhou polovinu, přerušovaná čára je lineárním proložením ideálního případu a plná čára je lineárním proložením druhé poloviny měření

Modré body znázorňují prvních pět odměřených hodnot, červené body znázorňují druhou polovinu měření; přerušovaná čára je lineárním proložením ideálního případu, kde durová stupnice v intervalu C-c3 je rovnoměrně rozložena mezi 270 cm výšky a plná čára je lineárním proložením druhé poloviny naměřených bodů.

Z obrázku můžeme vidět vývoj ve vnímání zvukově-prostorové škály. Když člověk zjistí, že rozložení, které původně zvolil, zcela neodpovídá celé stupnici, dokáže další výsledky upravit a zpřesnit. To můžeme pozorovat z toho, že druhá polovina měření (červené body) se více blíží ideálnímu případu. Nejvýrazněji je možné tuto změnu škály pozorovat na tónu c2 (MIDI hodnota=72), kde při prvním přiřazení byla hodnota výšky rovna 140 cm, při druhém 192 cm, přičemž ideální hodnota výšky pro tón c2 je rovna 202,5 cm. Rozdíl prvního a druhého zaznamenání výšky pro tón c2 dobře reprezentuje flexibilitu a schopnost učení se, která je pro funkčnost většiny sonifikačních metod nutností i předpokladem.

Pro názornost jsme tato zjištění popsali na měření jednoho ze subjektů, ale možné je stejné zákonitosti pozorovat u všech testovaných osob. Naměřené hodnoty a grafy se nachází na přiloženém CD v souboru pod názvem „TEST_elevace_vysledky.xls“, dále se zde nachází prázdný dokument s instrukcemi názvem „TEST_elevace_zadani.xls“ a patche sloužící k testování jednotlivých modifikací pod názvy „poslechovy_test_elevace_1.pd“, „poslechovy_test_elevace_2.pd“ a „poslechovy_test_elevace_3.pd“.



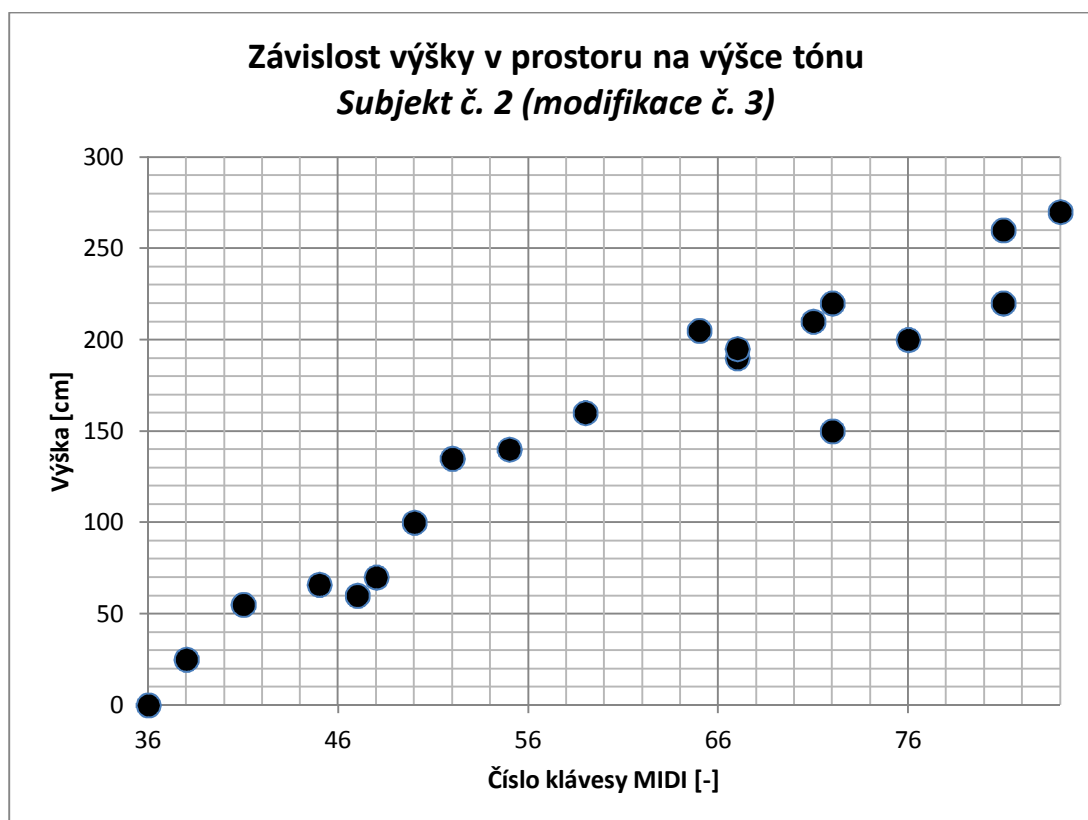
Obrázek 18: Graf naměřených hodnot výšky pro jednotlivé subjekty modifikace č. 1 testu, na vertikální ose je vynesena výška tónu ve formátu MIDI, na horizontální ose výška v prostoru

Z obrázku výše můžeme vidět, že naměřené hodnoty se pro všechny subjekty blíží lineárnímu průběhu. Hodnoty mediánu absolutní odchylky elevace pro jednotlivé subjekty jsou vyneseny v následující tabulce.

	Subjekt č. 1	Subjekt č. 2	Subjekt č. 3	Subjekt č. 4
Medián absolutní hodnoty odchylky elevace [°]	5,17	6,67	10,83	6,67

Tabulka 2: Medián absolutní hodnoty odchylky elevace jednotlivých subjektů modifikace č. 1

Z hodnot mediánu odchylek můžeme vidět, že orientace v prostoru podle modifikace č. 1 testu – tedy podle pěti tónů vzdálených od sebe jednu oktávu je lepší, než JND lidského sluchu, které se pro orientaci v mediální rovině pohybuje kolem 10° [26]. Problémem ovšem je velice nízký počet hodnot, které lze takovýmto způsobem lokalizovat. Z tohoto důvodu jsme přistoupili k větší diferencovanosti frekvenčního spektra – tedy k modifikaci č. 2 a 3. Na následujícím obrázku můžeme vidět rozložení bodů při modifikaci č. 3.

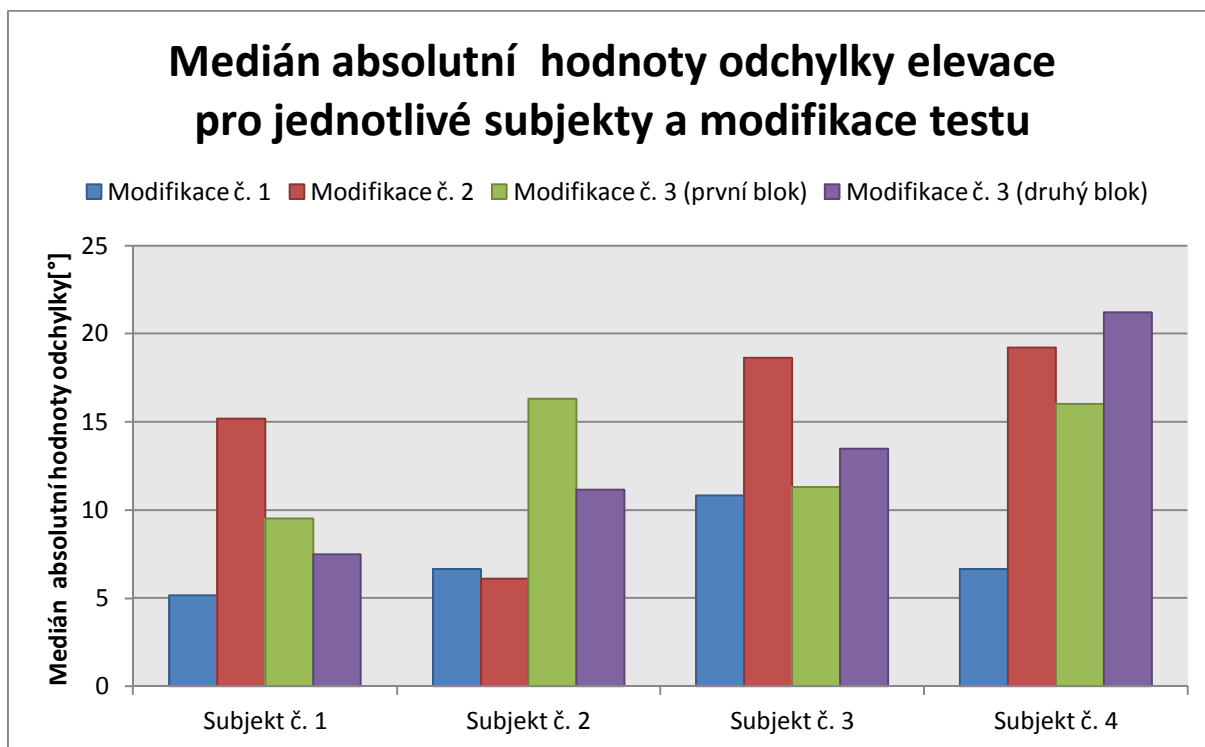


Obrázek 19: Graf závislosti umístění v prostoru na výšce tónu modifikace č. 3 testu, na vertikální ose je vynesena výška tónu ve formátu MIDI, na horizontální ose výška v prostoru

Na obrázku výše můžeme vidět rozložení hodnot pro jeden subjekt měření modifikace č. 3 testu, kde na vodorovné ose je vynesena výška tónu a na svislé ose výška v prostoru. Stejně jako u první modifikace i v tomto případě jsou hodnoty rozmístěny tak, že

hlubokým tónům náleží umístění v prostoru níže a vysokým umístění výše, přičemž mezi výškou tónu a umístěním v prostoru lze nalézt lineární vztah.

A. Štolc v práci Využití sonifikace pro navigační účely [26] zmiňuje, že JND lidského sluchu pro orientaci v prostoru je zhruba 10° a JND nejkvalitnějších sonifikačních algoritmů založených na simulaci HRTF dosahuje hodnot kolem 15° . Na následujícím obrázku můžeme vidět medián absolutních hodnot odchylek všech modifikací testu.



Obrázek 21: Graf zobrazující medián absolutní hodnoty odchylky elevace pro jednotlivé subjekty a jednotlivé modifikace testu. Jednotlivé sloupce zobrazují postupně veškeré modifikace testu elevace a jednotlivé čtveřice sloupců náleží vždy k jednomu subjektu.

Zdánlivě nejlepších výsledků dosahovala modifikace č. 1, kde ovšem znělo pouze pět tónů. Dále však můžeme vidět, že medián odchylek naměřených hodnot od ideálního průběhu se pohybuje v rozmezí 6° - 21° , což jsou výsledky blíží se 15° , která Štolc [26] zmiňuje jako nejlepší možné výsledky dosažitelné pomocí implementace HRTF. Dále můžeme předpokládat, že dlouhodobým tréninkem by došlo k zpřesnění odpovědí.

Posledním problémem, který jsme těmito testy zkoumali, byl vliv tzv. „kalibrační posloupnosti.“ Náš předpoklad byl takový, že s přidáním „kalibrační posloupnosti“ dojde k zpřesnění odpovědí, což nastalo u první a třetí testované osoby. Z komentářů, které testovaní uvedli po skončení experimentu, však vyplývá, že „kalibrační posloupnost“ byla pro všechny přítěží a orientaci v tónové řadě z větší části ztěžovala. Všichni testovaní rovněž uvedli, že zatímco bez kalibrační posloupnosti byl způsob jejich odpovídání spíše intuitivní, po jejím zařazení do experimentu se vyhodnocování změnilo ve více racionální.

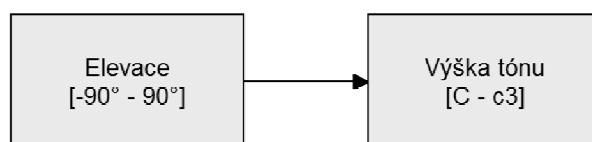
Přesnost odpovědí jistě souvisí s hudebním sluchem každého respondenta, avšak i přes veškeré nepřesnosti můžeme přistoupit k implementaci vztahu mezi výškou tónu a elevací, protože se ukázalo, že klíč, který dělí elevaci do stejných intervalů a ty poté přiřazuje tónům durové stupnice, byl pro všechny testované osoby intuitivním a pochopitelným.

4.4.2 Implementace mapování elevace

Po zapnutí programu je možné zvolit ze dvou druhů zvuku včetně toho, že jejich úrovně lze libovolně kombinovat. První možností je zvuk vzniklý aditivní syntézou, v druhém případě se jedná zvuk založený na fyzikálním modelování dle Karplus-Strongova algoritmu.

Rozdělení vertikální osy

Jak bylo zjištěno v experimentech popsanych v předchozí kapitole, rozdělení vertikální osy do intervalů náležících diatonické durové stupnici je pro uživatele intuitivní, proto jsme tento způsob mapování elevace využili i v naší práci.



Obrázek 22: Výška tónu jako funkce elevace včetně příslušných rozsahů

Interval $C - c3$ vymezuje celkový počet dvaceti osmi tónů, které spadají do stupnice C dur, hodnota elevace je proto rovněž kvantizována na dvacet osm hodnot. Odtud vyplývá, že rozdíl dvou sousedních tónů odpovídá změně elevace přibližně o $6^\circ 25'$, což je maximální rozlišení elevace navrženého algoritmu. Jak jsme zjistili experimenty výše, již po prvním seznámení s algoritmem je uživatel schopen se takovéto hodnotě přesnosti přiblížit, což však výrazně záleží na jeho hudebním sluchu.

Konkrétní provedení převodu elevace na výšku tónů v PD probíhá tak, že jednotlivým hodnotám elevace jsou přiřazeny indexy v tabulce, ze které se vyčítá MIDI hodnota výšky tónu a ta je následně převedena na frekvenci. Tabulku k nahlédnutí naleznete v příloze pod názvem *C_dur.txt*.

Aditivní syntéza

Libovolný periodický signál je možné vyjádřit jako řadu kosinových průběhů. Výstupní signál aditivní syntézy lze popsat následujícím výrazem [23].

$$y(t) = \sum_{k=1}^K r_k * \cos(2\pi k f_0 t + \varphi_0) \quad (5)$$

K účelu tvorby zvuku využíváme v PD jedenáct objektů [osc~], které slouží jako generátory harmonického signálu. Vstupními parametry jsou frekvence, případně fáze signálu. Pro tento počet oscilátorů jsme se rozhodli na základě přiložené tabulky č. 3 (viz

níže), která popisuje amplitudy jedenácti vyšších harmonických pro některé charaktery zvuku a jednak jsme při tvorbě algoritmu ověřili, že harmonické vyšší jedenácté při jejich nízké intenzitě nehrají výraznou roli při tvorbě barvy zvuku. Syrový [11] k problému složení spektra uvádí, že intervaly, které zásadně určují subjektivní vjem barvy tónu, jsou vyčerpány do šesté harmonické. V naší aplikaci jsme využili charakteru „klarinet.“

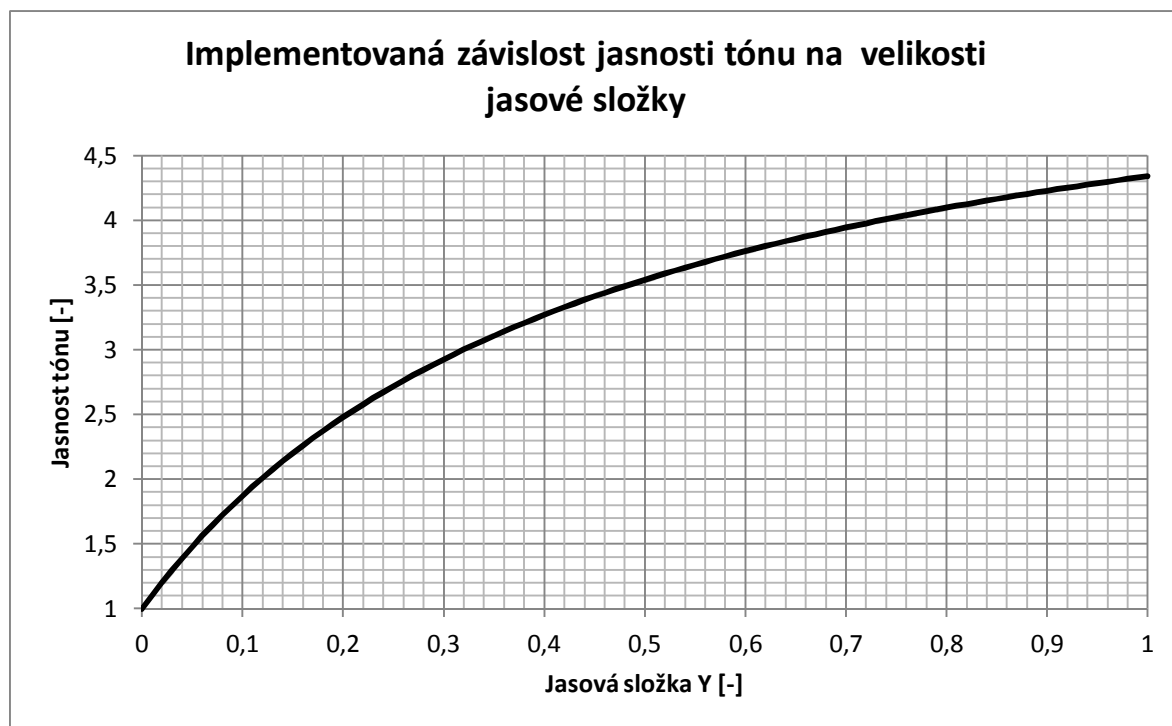
Poř. harmonické	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
trubka	0,17	0,63	0,57	0,98	0,56	0,68	0,02	0,05	-	-	-
harmonika	8,60	0,45	3,40	0,50	0,42	0,13	0,13	0,16	0,04	0,35	0,02
klarinet	1,00	0,00	0,75	0,00	0,50	0,00	0,14	0,50	0,00	0,12	0,17

Tabulka 3: Tabulka amplitud vyšších harmonických tří různých nástrojů [24]

Jak bylo uvedeno v předchozí kapitole, budeme mapovat hodnotu jasu na hodnotu jasnosti. Provedenými experimenty jsme zjistili, že implementace by měla jasnějším objektům přiřazovat tóny vyšší jasnosti (více v kap. 3.3). Jasová složka může nabývat hodnot v rozmezí 0-1, amplitudu vyšších harmonických poté počítáme podle následujícího vzorce.

$$c_k(t) = n_k * Y(t); k > 2; c_1 = 1 \quad (6)$$

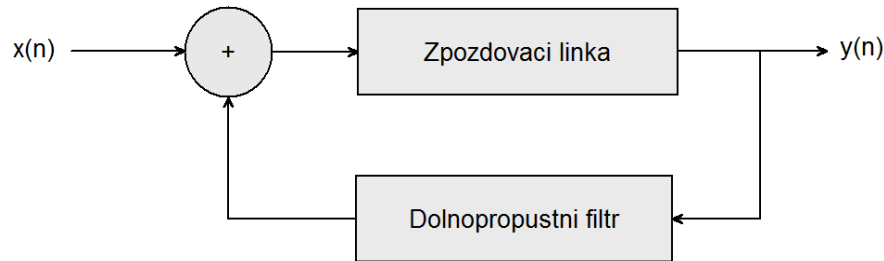
Kde $c_k(t)$ je hodnota amplitudy k-té harmonické, n_k je amplituda k-té harmonické z tabulky výše, $Y(t)$ je jasová složka pohybujícího se objektu. Amplituda první harmonické je na jasu nezávislá, rovná jedné. V průběhu této operace je rovněž počítána hodnota jasnosti podle vzorce (1) popsáno v kapitole 1.6. Konkrétní závislost jasnosti na velikosti jasové složky můžeme pozorovat na následujícím obrázku.



Obrázek 23: Implementovaná závislost jasnosti tónu na jasové složce pro aditivní syntézu; na vodorovné ose je vynesena hodnota jasové složky Y, na svislé ose hodnota jasnosti syntetizovaného tónu

Karplus-Strongův algoritmus

Druhým zvukem, který je možno využít pro sonifikaci objektu je zvuk vzniklý na základě Karplus-Strongova algoritmu, což je metoda tzv. fyzikálního modelování strunných nástrojů, která byla poprvé prezentována v 80. letech, pro naši práci vycházíme převážně z rozšíření, které popisuje Jeffe a Smith [25].



Obrázek 24: Schematické znázornění Karplus-Strongova algoritmu [25]

Princip algoritmu je založen na průchodu krátkého impulsu šumu zpoždovací linkou a dolnoproustním filtrem (DP) ve smyčce. Zpoždovací linka jako digitální filtr vyššího řádu reprezentuje vlastnosti struny, šumový impuls zastupuje drnknutí na strunu.

Řízení výšky tónu

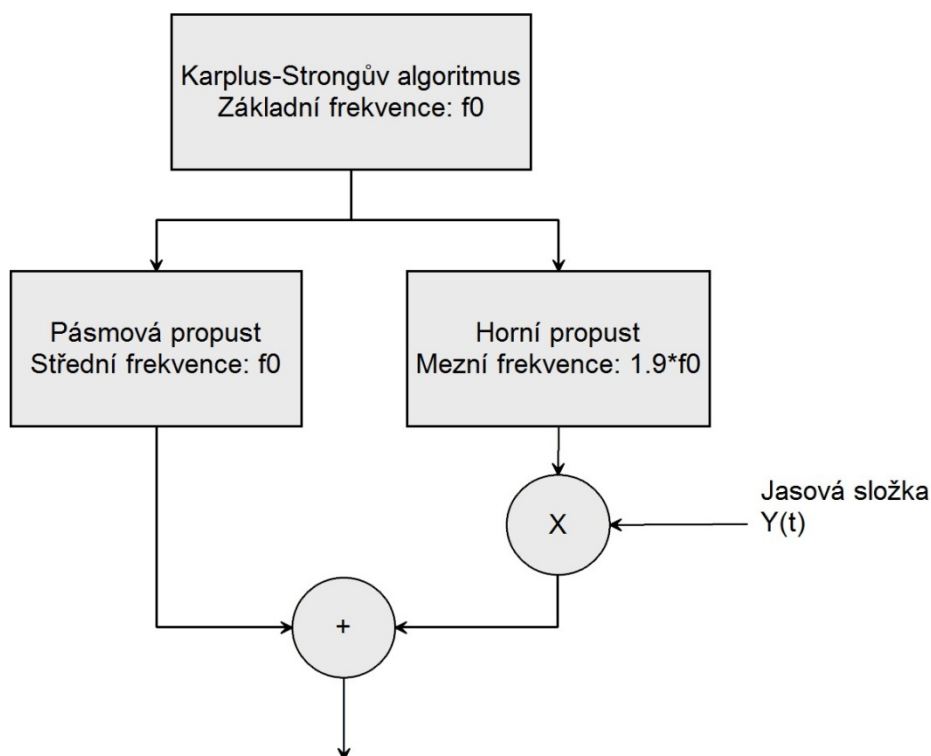
Základní rezonanční frekvence soustavy udává výšku tónu a je dána rezonanční frekvencí filtru, která závisí na délce zpoždovací linky [25]. Pro počet vzorků N a vzorkovací frekvenci f_s můžeme základní rezonanční frekvenci f_1 vypočítat podle následujícího vzorce.

$$f_1 = \frac{f_s}{N + \frac{1}{2}} \quad (7)$$

Řízení jasnosti tónu v závislosti na jasové složce

Stejně jako v případě aditivní syntézy i u Karplus-Strongova algoritmu provádíme převod jasové složky na jasnost tónu. Docíleno je toho modifikací vztahu pro aditivní syntézu. Zvuk je nejdříve filtrací rozdělen do dvou kanálů, přičemž první kanál vzniká průchodem signálu pásmovou propustí naladěnou na hodnotu první harmonické, druhý kanál vzniká průchodem signálu horní propustí s mezním kmitočtem pod frekvencí druhé harmonické. Pro tento účel využíváme objekty [biquad~], což jsou filtry se dvěma nulami a dvěma póly. K nalezení koeficientů pro tento typ filtrů slouží objekty [bandpass] a [highpass], jejichž vstupními parametry jsou hodnoty mezních kmitočtů.

Po rozdělení signálu na základní harmonickou a vyšší harmonické, je řešení obdobné jako u aditivní syntézy. Kanál s první harmonickou zůstává na jasu nezávislý, zatímco amplituda kanálu, který obsahuje vyšší harmonické je řízena hodnotou jasové složky. S rostoucí jasovou složkou roste jasnost tónu. Pro názornější představu implementace mapování jasu slouží následující obrázek.



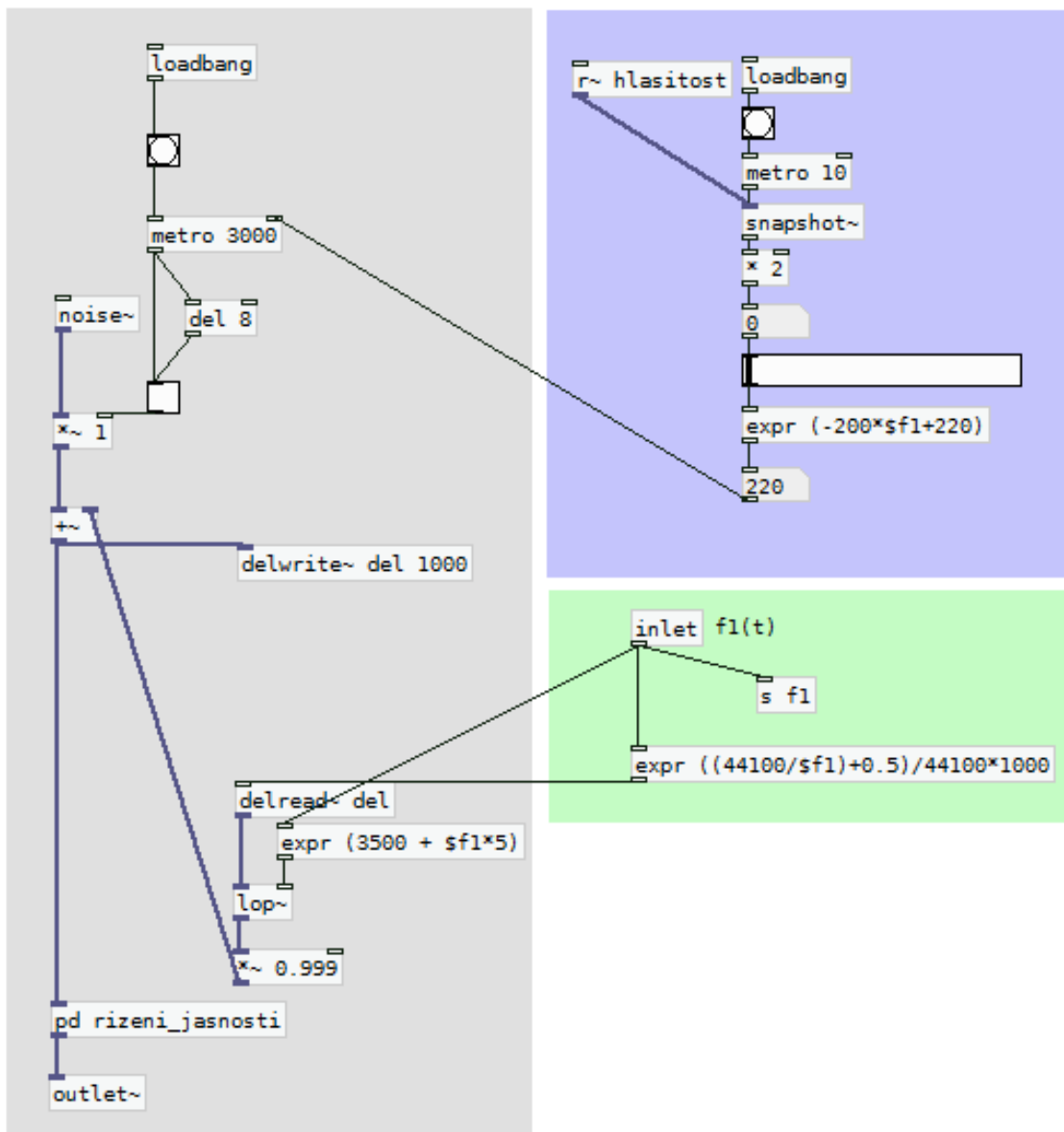
Obrázek 25: Diagram znázorňující řízení jasnosti tónu vzniklého dle Karplus-Strongova algoritmu, kde signál je pomocí filtrace rozdělen do dvou kanálů: na základní harmonickou a vyšší harmonické, jejichž amplituda je poté řízena velikostí jasové složky $Y(t)$

Řízení rytmu

Nejvýraznější část tónu vzniklého pomocí Karplus-Strongova algoritmu trvá přibližně 400ms, poté dochází k výraznému omezení vyšších harmonických složek, což je problém, který by mohl výrazně poznamenat především rozlišitelnost jasu pohybujícího se objektu. Z tohoto důvodu je nutné tvořit tóny v určitém rytmu. Vzhledem k době trvání výrazné části jsme se rozhodli pro základní periodu $T_{k-s}=220\text{ms}$, a abychom zvýšili sdělnost sonifikační metody, perioda závislá na velikosti pohybu. Hodnota velikosti pohybu $\beta(t)$ je normována do intervalu 0 – 1. Periodu T_{k-s} s jakou je tvořen nový tón potom spočítáme podle následujícího vzorce.

$$T_{k-s}(t) = 220 - \beta(t) * 200 \quad (8)$$

Minimální perioda, s jakou je tvořen nový tón je 20ms, což je krajní mez, při které již dochází k přechodu mezi zvukem, který má charakter drnknutí na strunu a zvukem, jehož charakter se blíží šumu.



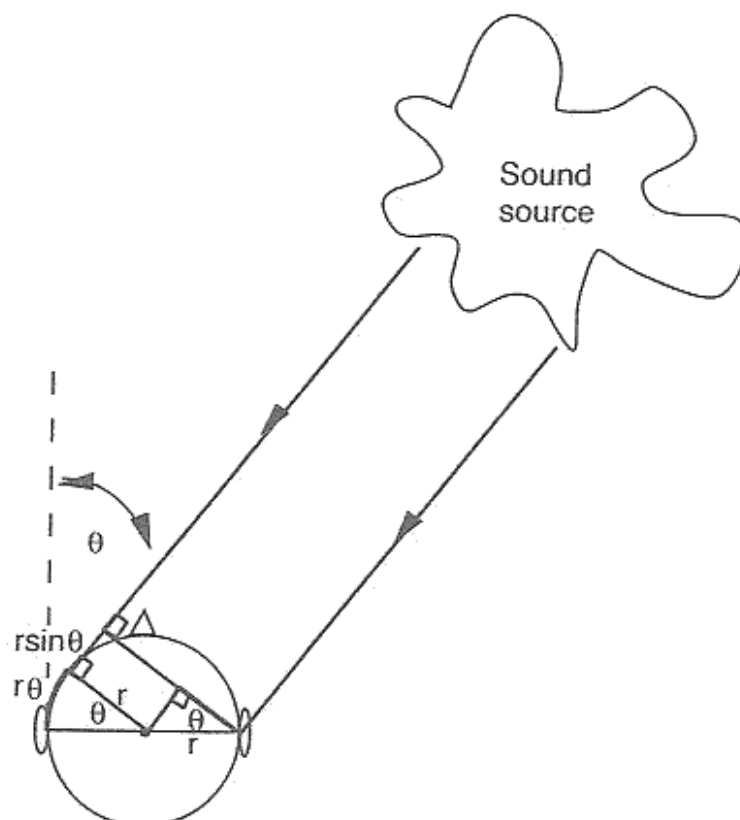
Obrázek 26: Realizace Karplus-Strongova algoritmu v PD, kde na počátku stojí šumový impuls o délce 8ms, ten poté prochází zpoždovací linkou [delwrite~], dolní propustí [lop~], poté dochází k řízení jasnosti tónu; výpočet rytmu v závislosti na velikosti pohybu probíhá v modrém rámečku, výpočet základní frekvence systému v rámečku zeleném

4.5 Implementace mapování azimutu

Pro sonifikaci polohy v horizontální rovině využíváme modifikace ITD a ILD popsané v kapitole 1.2.

4.5.1 Interaurální časový rozdíl

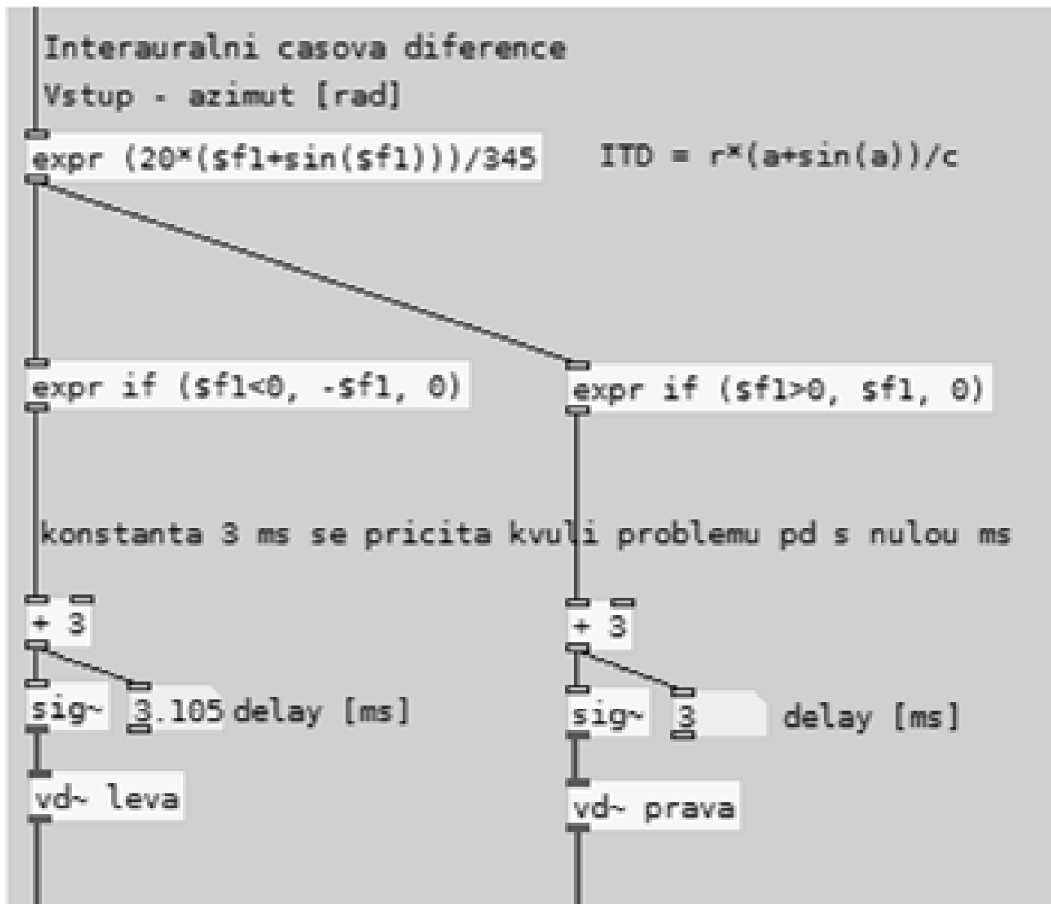
Interaurální časový rozdíl vypočítáme podle geometrických zákonitostí šíření zvukové vlny.



Obrázek 27: Geometrie lokalizace zdroje zvuku na základě rozdílu vzdáleností levého a pravého ucha od zdroje zvuku

Při výpočtu uvažujeme vzdálenost, kterou musí zvuk urazit kolem hlavy, přičemž hlavu považujeme za kouli poloměru r , kde vzdálenost uší je 20cm. Rychlost zvuku c při výpočtu je $345 \text{ m}\cdot\text{s}^{-1}$ [3].

$$ITD = \frac{r * (\theta + \sin(\theta))}{c} \quad (9)$$



Obrázek 28: Implementace ITD v PD na základě vztahu (9), kde po výpočtu ITD je touto hodnotou řízen objekt [vd~] levého a pravého kanálu sloužící jako zpožďovací člen s proměnnou hodnotou zpoždění, k hodnotě zpoždění ITD se pro oba kanály přičítají 3ms kvůli problému PD s nulou

Implementace ILD

Jak bylo uvedeno výše, interaurální intenzitní diference je závislá nejen na úhlu, ale rovněž na frekvenci signálu. Způsob, který využijeme, bývá nazýván amplitudovým panorámováním. Ville Pullki ve studii *Virtual Sound Source Positioning Using Vector Base Amplitude Panning* [27] popisuje několik možných řešení. Jedním z nich je lineární panorama, jak ji popisuje následující vzorec, kde g_{Leva} a g_{Prava} jsou hodnoty amplitud jednotlivých kanálů.

$$g_{Leva} + g_{Prava} = konst. \quad (10)$$

Dalším možným způsobem je obdobné řešení, které ovšem uvažuje závislost amplitudy levého a pravého ucha kvadraturně.

$$g_{Leva}^2 + g_{Prava}^2 = konst. \quad (11)$$

Pro oba způsoby by krajní případ vypadal tak, že pro azimut $\varphi=90^\circ$ je $g_{Leva}=0$ a $g_{Prava}=konst.=1$. Další možný způsob bývá označován jako trigonometrické vyjádření panoramy.

$$\frac{tg(\theta)}{tg(\theta^0)} = \frac{g_{Leva}-g_{Prava}}{g_{Leva}+g_{Prava}} \quad (12)$$

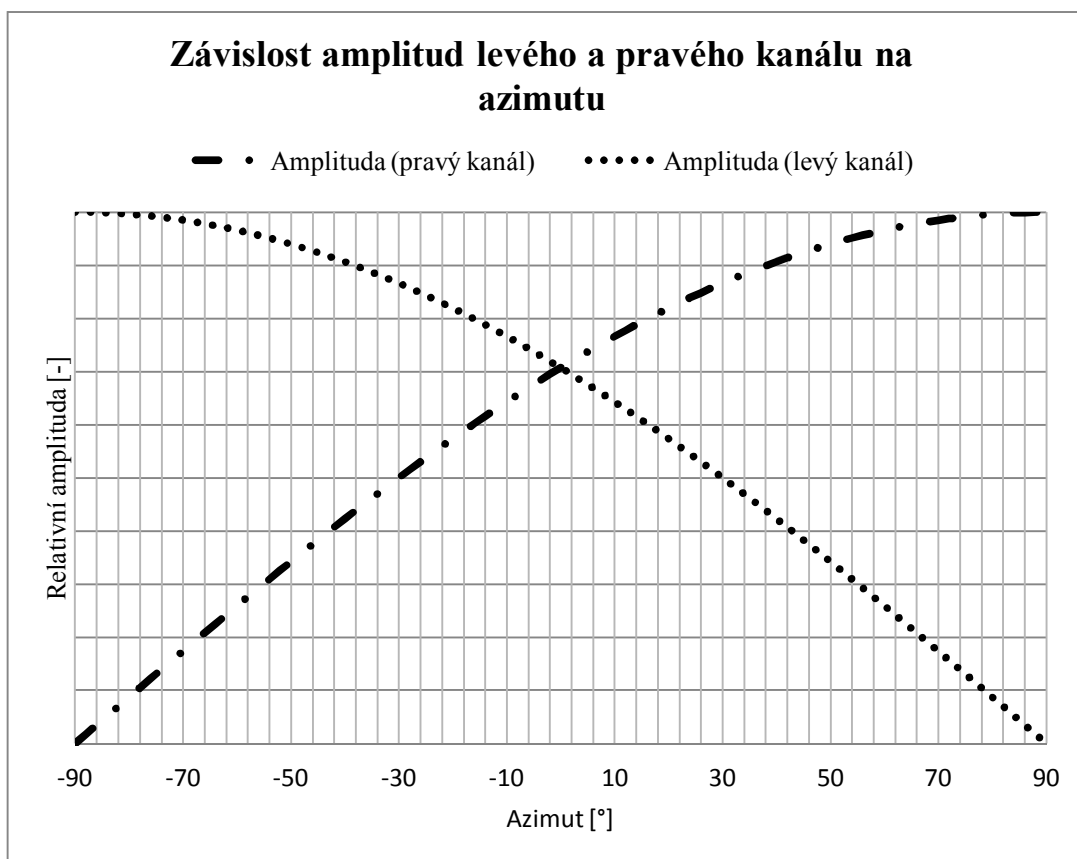
Kde θ^0 označuje úhel, který svírá posluchač s reproduktory. My využijeme vzorec, který zmiňuje David Griesinger v práci *Stereo and Surround Panning in Practice* [14]. Jedná se o sinový tvar amplitudového panorámování.

$$g_{Leva} = \cos(\theta) * input; g_{Prava} = \sin(\theta) * input \quad (13)$$

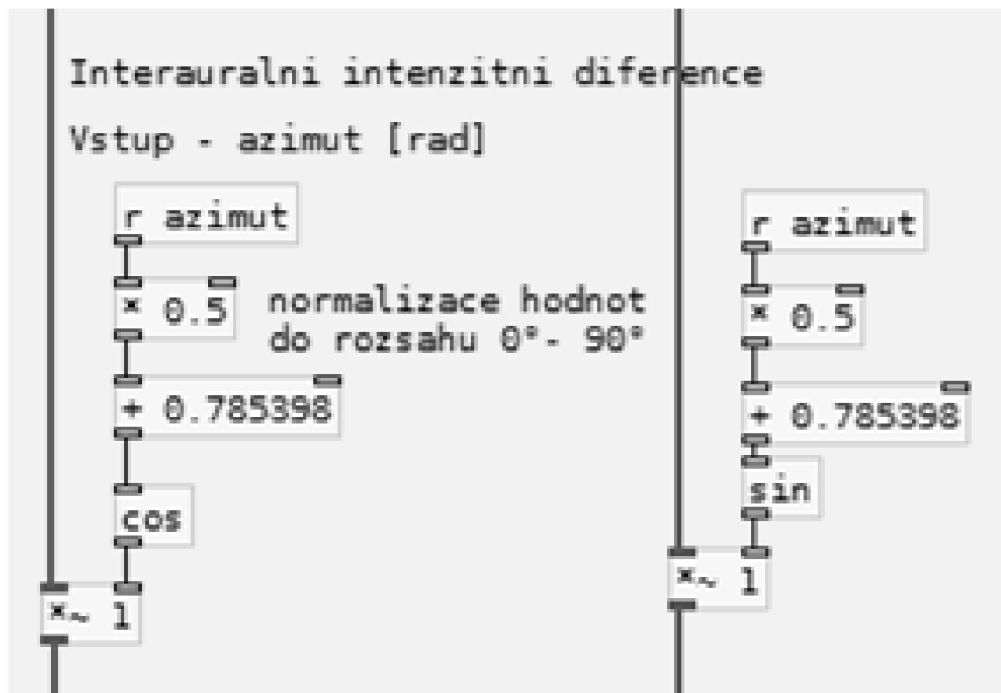
Důležité je, že vysílaný výkon je po celou dobu, pro všechny úhly konstantní, což plyne ze vztahu (7) resp. (8).

$$\sin^2(\theta) + \cos^2(\theta) = 1 \quad (14)$$

$$g_{Leva_output}^2 + g_{Prava_output}^2 = input^2 \quad (15)$$



Obrázek 29: Závislost amplitudy levého a pravého kanálu v závislosti na azimutu použité implementace intenzitní diference



Obrázek 30: Implementace ILD v PD podle vztahu (13), kde hodnoty azimutu jsou normalizovány do mezí $0^\circ - 90^\circ$ a amplitudy levého a pravého kanálu závisí na kosinu případně na sinu normalizovaného azimutu, průběh těchto amplitud v závislosti na azimutu je možno vidět na obr. 29

5 TESTOVÁNÍ ALGORITMU

5.1 Testování vnímání rychlosti

Test implementace rychlosti pohybujícího se objektu byl rozdělen do dvou bloků, přičemž v obou blocích bylo testovaným osobám přehráno deset dvojic stimulů a jejich úkolem bylo určit, který z dvojice stimulů reprezentuje rychleji pohybující se objekt. První modifikace testu vznikla při běhu všech částí algoritmu. Následně byl proveden modifikovaný test, kde tvorba stimulu nepodléhala změnám azimutu a elevace: testovaná osoba tedy musela určit rychlost pohybu na základě rytmu tvoření Karplus-Strongova algoritmu a amplitudy. Test byl proveden celkem na pěti osobách. Zvuk byl přehráván ze sluchátek.

5.1.1 Výsledky

Z celkem padesáti odpovědí všech respondentů byly při první modifikaci testu pouze dvě chybné, což svědčí o tom, že tato implementace funguje velice dobře. Hlavní důvod tak nízké chybovosti odpovědí zřejmě spočívá v tom, že informace o pohybu je předávána hned čtyřmi způsoby:

- Amplitudou signálu
- Rytmem tvoření Karplus-Strongova algoritmu
- Délkou trvání stejného pohybu provedeného za odlišnou dobu
- Rychlostí změny výšky tónu případně stereofonie pro různé rychlosti pohybu

Průměrná chybovost při vyřazení změny elevace a azimutu byla 18%, což znamená, že informace o rychlosti je výrazně nesena také pomocí periody tvoření Karplus-Strongova algoritmu a amplitudy.

Na přiloženém CD naleznete zvukové soubory obou modifikací testu pod názvy „TEST_rychlost1.wav“ a „TEST_rychlost2.wav“. V příloze se rovněž nachází soubor obsahující správné řešení i výsledky jednotlivých subjektů pro obě modifikace pod názvem „TEST_rychlost_reseni.xls“.

5.2 Souhrnný test algoritmu

Úkolem testované osoby bylo přiřadit přehrávanému stimulu jeden ze tří načrtnutých průběhů a zároveň zvolit jednu ze tří hodnot jasu. Každý stimul byl přehrán pouze jednou, mezi jednotlivými stimuly byla časová prodleva v délce 5s. Celkem bylo testováno patnáct osob, jednalo se jak o hudebníky, tak nehudebníky ve věku 22 – 50 let. Každá testovaná osoba analyzovala tři bloky po deseti stimulech, dohromady tedy bylo provedeno 450 měření. Před začátkem testu byl popsán princip, na kterém algoritmus pracuje, zároveň bylo přehráno video obsahující více sonifikovaných sekvencí.

5.2.1 Výsledky

Testem jsme ověřili, že na základě navrženého systému je možné správně rozpoznat trajektorii objektu. Správnost odpovědí se s průběhem testu výrazně zlepšovala, což je možné vidět na následující tabulce.

	Průměrná správnost odpovědí při určování trajektorie [%]	Průměrná správnost odpovědí při určování jasu [%]
Blok č. 1	73,3	34,7
Blok č. 2	82,7	40,7
Blok č. 2	93,3	38,0

Tabulka 4: správnost odpovědí při určování trajektorie a jasu pro jednotlivé bloky testu

V průběhu testu se ukázalo, že navržený algoritmus je snáze pochopitelný hudebně založeným uživatelům, stejně tak u mladého dospělého člověka existuje předpoklad, že pochopí navrženou strukturu rychleji, než uživatel starší čtyřicet let. Tento fakt můžeme vidět z tabulky níže, kde je vynesena správnost odpovědí v jednotlivých blocích pro hudebníky, nehudebníky a pro skupinu lidí starších čtyřiceti let.

Měřená skupina:	Průměrná správnost odpovědí při určování trajektorie v jednotlivých blocích [%]		
	Hudebníci	Nehudebníci	Starší čtyřiceti let
Blok č. 1	78,3	55	50
Blok č. 2	88,3	65	75
Blok č. 3	100	82,5	95

Tabulka 5: správnost určování trajektorií v závislosti na vztahu k hudbě a věku

Z testů provedených v kapitole 4.4 vyplývá, že průměrná chyba elevace je rovna 15,2°, což je hodnota, která se blíží nejkvalitnějším systémům pracujícím s aplikací HRTF. Konkrétní hodnoty se nachází v tabulce níže.

Průměrná minimální chyba elevace [°]	Průměrná chyba elevace [°]	Průměrná maximální chyba elevace [°]
0,0°	15,2°	43,9°

Tabulka 5: odchylky elevace zjištěné pomocí testu za použití kamerového stativu s možností směřování světelného toku (viz kap. 4.4)

Na rozdíl od rozpoznání trajektorie činilo určování jasu uživatelům znatelně větší problémy. Z provedených testů nemůžeme prohlásit, že by někdo dokázal s jistotou jas rozpoznat. Problém určování hodnoty jasu všichni testovaní shodně komentovali tak, že se soustředili především na rozpoznání směru objektu, přičemž jas považovali za vedlejší a většinou se jím ani nestihli zabývat. Možným důvodem nízké úspěšnosti může být rovněž princip testu, kdy testovaní určovali hodnotu jasu bez dané reference, což je po jednom zhlédnutí videa promítnutého před začátkem testu bez delšího tréninku velice obtížné, ačkoli medián správnosti odpovědí hudebníků dosáhl v druhém a třetím bloku testu 45%, což je hodnota, kterou lze považovat za určitý náznak toho, že navržená struktura je částečně funkční i při těchto podmínkách. Při konkrétním použití by problém rozlišení jasu mohl být částečně potlačen tím, že při poslechu zvuku různých objektů by si uživatel vytvořil vlastní referenční hodnotu jasu. Tento předpoklad se opírá o testy uvedené v kapitole 4.3, ze kterých je zřejmé, že člověk je schopen frekvenční složení zvuku rozpoznat a spojit ho s jasem.

Souhrnným testem se dále ukázalo, že navržená struktura, kdy relativně malá část zorného pole je mapována na celé pole sluchové, nečiní uživateli problémy, což může být způsobeno například analogií s mobilními telefony a tablety, kde je taktéž malý obraz spojován s celým sluchovým polem.

V průběhu dílčích testů, jejichž výsledky do práce nakonec nebyly zahrnuty, se ukázalo, že použití algoritmu činí výrazně větší problémy mladistvým, nehledě na jejich hudební sluch. V další práci by bylo zajímavé zaměřit se na zkoumání navržené struktury ve spojitosti s věkem uživatele.

Na přiloženém CD naleznete vlastní zvukovou stupu testu „TEST_souhrny.wav“, dále pak video „VIDEO_popis_k_testu.avi“, které bylo testovaným osobám přehráno před začátkem testu, dále dokument „TEST_souhrnny.pdf“, který obsahuje instruktáž a nevyplněný arch testu s načrtnutými průběhy trajektorie pohybu, nakonec pak soubor „TEST_souhrny_vysledky.xls“, ve kterém jsou vyneseny počty chybných odpovědí jednotlivých subjektů, a ty dále zpracovány a vyneseny do krabicových grafů.

6 ZÁVĚR

Tato práce se zabývá sonifikací videosignálu. V úvodu teoretické části byl popsán princip binaurálního slyšení, poté bylo zmíněno několik studií pojednávajících o vztahu mezi elevací a výškou tónu, následovala část o vztahu mezi různými smyslovými oblastmi. Ve třetí kapitole byl vysvětlen pojem sonifikace, včetně její definice a popisu základních sonifikačních metod. Zároveň bylo popsáno několik algoritmů, které pracují s převodem videa, či pohybu do zvukové podoby.

Pro účel sonifikace obrazu byl v prostředí Pure Data navržen program, který zpracovává videosignál snímaný webkamerou počítače, jeho návrh je popsán v kap. 4.1, popis implementace dílčích částí je popsán v kap. 4.3 a 4.4. Algoritmus pracuje tak, že nejprve ve videu nalezne pohybující se objekt, následně informaci o jeho elevaci mapuje na výšku tónu a azimut na stereofonii vzniklého signálu. Jas objektu je mapován na frekvenční složení zvuku tvořeného pomocí Karplus-Strongova algoritmu a aditivní syntézy. Před implementací navržené struktury byly provedeny dva testy zkoumající subjektivní vztah mezi elevací a výškou tónu a mezi jasností zvuku a jasnou obrazu, průběhy testů jsou popsány v kap. 4.3 a 4.4. Potvrdilo se, že člověk přirozeně spojuje vyšší hodnoty elevace s vyššími tóny a druhý test ukázal, že pět z šesti testovaných osob přiřazovalo zvyšující se hodnotě jasnosti zvuku vyšší hodnoty jasu.

V páté kapitole je popsán souhrnný test celého algoritmu, kde bylo úkolem testovaných osob přiřadit každému stimulu jeden z načrtnutých průběhů trajektorie a jednu hodnotu jasu, přičemž většina testovaných osob navrženou strukturu pochopila již během analyzování první desítky stimulů. Ke konci testu byla průměrná správnost rozpoznání trajektorie objektu rovna 93,3%. Dále se ukázalo, že rozpoznání jasu při současném sledování trajektorie činí uživatelům veliké problémy. Z výsledků testu vyplývá, že navržený algoritmus nejvíce vyhovuje uživateli, který je mladý dospělý hudebník.

Úkolem této práce bylo seznámit se s prostředím PD a v něm navrhnout algoritmus sloužící k sonifikaci polohy objektu z videosignálu, což také bylo splněno. Přínosem této práce je snaha o *hledání souvislostí* mezi zvukovou a obrazovou smyslovou oblastí prostřednictvím experimentů se vztahy „výška-výška“ a „jas-jasnost“. Hlavní výhodou navržené struktury je její *intuitivnost* a současně vysoký objem informací, které zvukový signál přenáší. Konkrétně se jedná o mapování jasu a rychlosti, které zároveň nijak nezatěžuje základní strukturu polohování objektu, ale je k ní pouze přidanou hodnotou, což je směr kterým se dnes ubírá více sonifikačních metod.

Vytvořený algoritmus by bylo možné rozšířit tak, aby při současném pohybu více objektů docházelo k jejich paralelní sonifikaci. V další práci by bylo zajímavé zkoumat vlastní propojení zvukové a obrazové smyslové oblasti v lidském mozku. Popsanou strukturu by bylo možné využít pro navigaci nevidomých, signalizační, či rehabilitační účely.

7 ZDROJE:

- [1] SAMIDH (a kol.). *Localization of Sound in Micro Gravity*. Poslední revize 1999. [online]. Obrázek ve formátu JPG. [cit. 2013-4-11]. Upraveno. Dostupné z: <<http://web.media.mit.edu/~raffik/zero-g/aup/linked-final.html>>.
- [2] DIAMOND, Luis Scott. *Rayleigh Duplex Theory*. [online]. Poslední revize 27.5.2010. [cit. 2014-2-1]. Dostupné z: <<http://diamonddissertation.blogspot.cz/2010/05/rayleigh-duplex-theory.html>>.
- [3] PNAS. *Interaural time difference (ITD)*. Obrázek ve formátu JPG. [cit. 2014-2-15]. Dostupné z: <<http://www.pnas.org/content/105/2/401/F4.expansion.html>>.
- [4] *Spatial audio*. 2014. [online]. Obrázek ve formátu JPG. [cit. 2014-2-18]. Dostupné z: <<http://www-users.york.ac.uk/~th632/interauralIntensityDifference.html>>.
- [5] HOFMAN, RISWICK, OPSTAL. *Relearning sound localization with new ears*. Nijmegen, 1998. University of Nijmegen, Department of Medical Physics and Biophysics. Dostupné z: <<http://www.mbfys.ru.nl/~johnvo/papers/nm98.pdf>>.
- [6] WILSON, Tracy. *How Virtual Surround Sound Works*. [online]. Poslední revize 31.5.2007. [cit. 2014-24-2]. Dostupné z: <<http://electronics.howstuffworks.com/virtual-surround-sound2.htm>>.
- [7] KEITH, FRANKLIN, ROBERTS. *Pie Chart Sonification*. Canterbury, 2003. University of Kent, Computing laboratory, Canterbury, England, UK. ISBN: 0-7695-1988-1.
- [8] BILSEN, LAMORÉ. *Richtighoren*. [online]. Poslední revize květen 2011. [cit. 2014-3-1]. Dostupné z: <<http://www.audiologieboek.nl/hfm/hfd2/2-7-1.htm>>.
- [9] RUSCONIA, KWANA, GIORBANOB. *Spatial representation of pitch height*. Elsevier [online sborník]. 2005. University College London, Università degli Studi di Padova. ISSN: 0010-0277. Dostupné z: <<http://www.mathematicalbrain.com/pdf/2005ERBKBGBB.PDF>>.
- [10] COLLIER, HUBBARD. *Musical scales and brightness evaluations: Effects of pitch, direction, and scale mode*. University of North Carolina at Pembroke, Texas Christian University. Dostupné z: <http://timothyhubbard.net/colhub_MS2004.pdf>.
- [11] SYROVÝ, Václav. *Hudební akustika*. Praha: Akademie múzických umění; 2003. ISBN 978-80-7331-161-2

- [12] WIKIPEDIA. *Brightness*. [online]. Poslední revize 12.2.2013. [cit. 2014-3-10]. Dostupné z: <<http://en.wikipedia.org/wiki/Brightness>>.
- [13] POYNTON, Charles. *Color technology. Color FAQ*. [online]. c2006. [cit. 2014-3-5]. Dostupné z: <http://www.poynton.com/notes/colour_and_gamma/ColorFAQ.html>.
- [14] HERMANN, HUNT, NEUHOFF (a kol.). *The Sonification Handbook. 1 st. ed.*, 2011. Berlin: Logos Publishing House. ISBN 978-3-8325-2819-5. Dostupné z: <<http://sonification.de/handbook/>>.
- [15] STANĚK, Jan. *Theremin – konstrukce hudebního nástroje*. STRETECH 2009, ČVUT v PRAZE. ISBN 978-80-01-04367-7. Dostupné z: <<http://www1.fs.cvut.cz/stretech/2009/pdf/1096.pdf>>.
- [16] KNEES, POHLE, WILDMER. *Real-Time Synaesthetic Sonification of Train Journeys*. [online]. Linz, 2008. MM '08 Proceedings of the 16th ACM international conference on Multimedia [sborník konference]. s. 1117-1118. ISBN: 978-1-60558-303-7
- [17] WIKIPEDIA. *Alexander Scriabin*. [online]. Poslední revize 21.2.2014. [cit. 2014-26-2]. Dostupné z: <http://en.wikipedia.org/wiki/Alexander_Scriabin>.
- [18] JENSENIUS. *Motion-sound Interaction Using Sonification based on Motiongrams*. [online]. Oslo, 2012. University of Oslo, Department of Musicology. ISBN: 978-1-61208-177-9. Dostupné z: <https://www.duo.uio.no/bitstream/handle/10852/26900/Jensenius_2012.pdf>.
- [19] VINKEN (a kol.). *Auditory Coding of Human Movement Kinematics*. Hanover, 2013. Leibniz University Hanover, Institute of Sports Science, Hanover, Germany. Dostupné z: <http://sonification-online.com/wp-content/uploads/2014/02/Vinken-et-al_Auditory-Coding-of-Human-Movement-Kinematics.pdf>.
- [20] WALLIS, RIKAKIS, OLSEN (a kol.). *Real-Time Sonification of Movement for an Immersive Stroke Rehabilitation Environment*. [online]. Arizona State University. Arts and Media Engineering Department. Dostupné z: <<http://ame2.asu.edu/projects/mrrehab/papers/ICADpaper-16.pdf>>
- [21] PUCKETTE, M. S. Pure Data - Community site. Poslední revize 11.1.2014. [cit. 2014-01-13]. Dostupné z: <<http://puredata.info/>>.
- [22] TIŠNOVSKÝ, Pavel. *Rozhraní MIDI na stolních počítačích*. [online]. Poslední revize 19.2.2009. [cit. 2014-15-4]. Dostupné z: <<http://www.root.cz/clanky/rozhrani-midi-na-osobnich-pocitacich/>>.

[23] WIKIPEDIA. *Additive synthesis*. [online]. Poslední revize 21.4.2014. [cit. 2014-21-4]. Dostupné z: <http://en.wikipedia.org/wiki/Additive_synthesis>.

[24] ČMEJLA, SOVKA. *Úvod do číslicového zpracování signálů - cvičení*. ČVUT Praha. 2005.

[25] JAFFE, SMITH. *Extensions of the Karplus-Strong Plucked-String Algorithm*. *Computer Music Journal*, vol. 7, no. 2, s. 56-69, 1983.
ISSN 0148-9267

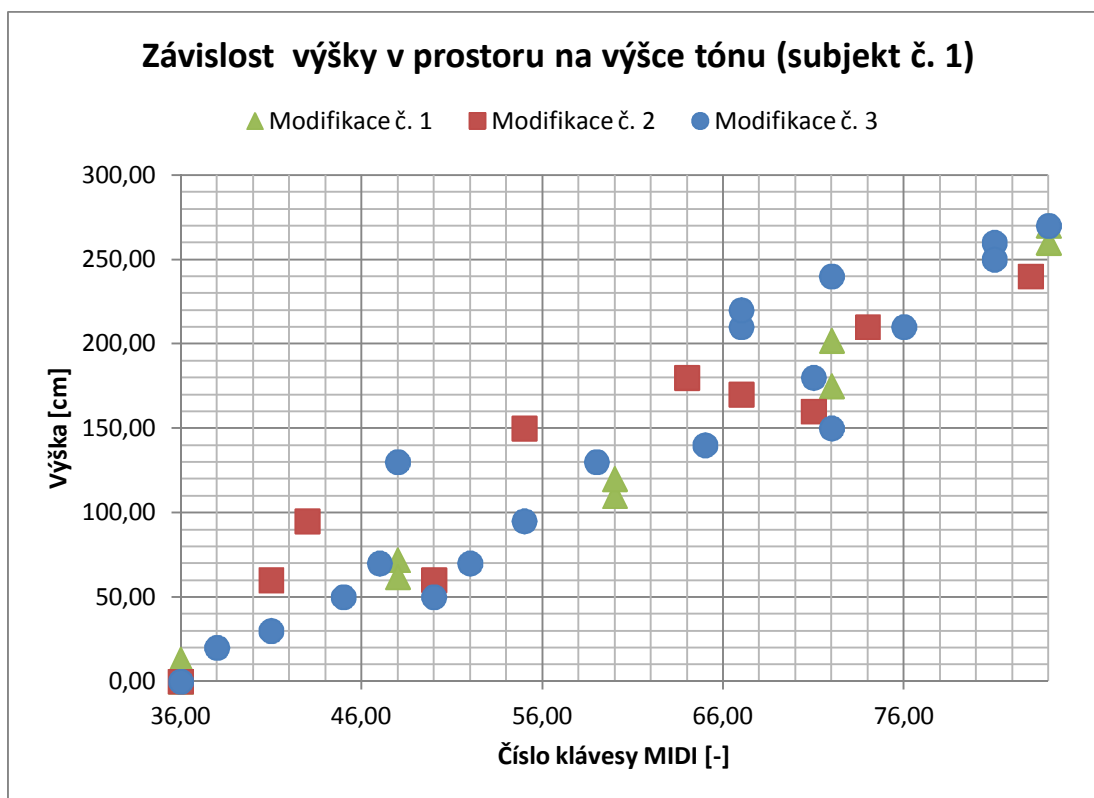
[26] ŠTOLC, A. *Využití sonifikace pro navigační účely*. Praha: České vysoké učení technické. Fakulta elektrotechnická. Katedra radioelektroniky, 2013. Vedoucí diplomové práce Ing. František Rund, Ph.D.

[27] PULKKI, *Virtual Sound Source Positioning Using Vector Base Amplitude Panning*. 1997. *Journal of the Audio Engineering Society*, vol. 45, no. 6, s 456-466.
ISSN: 15494950.

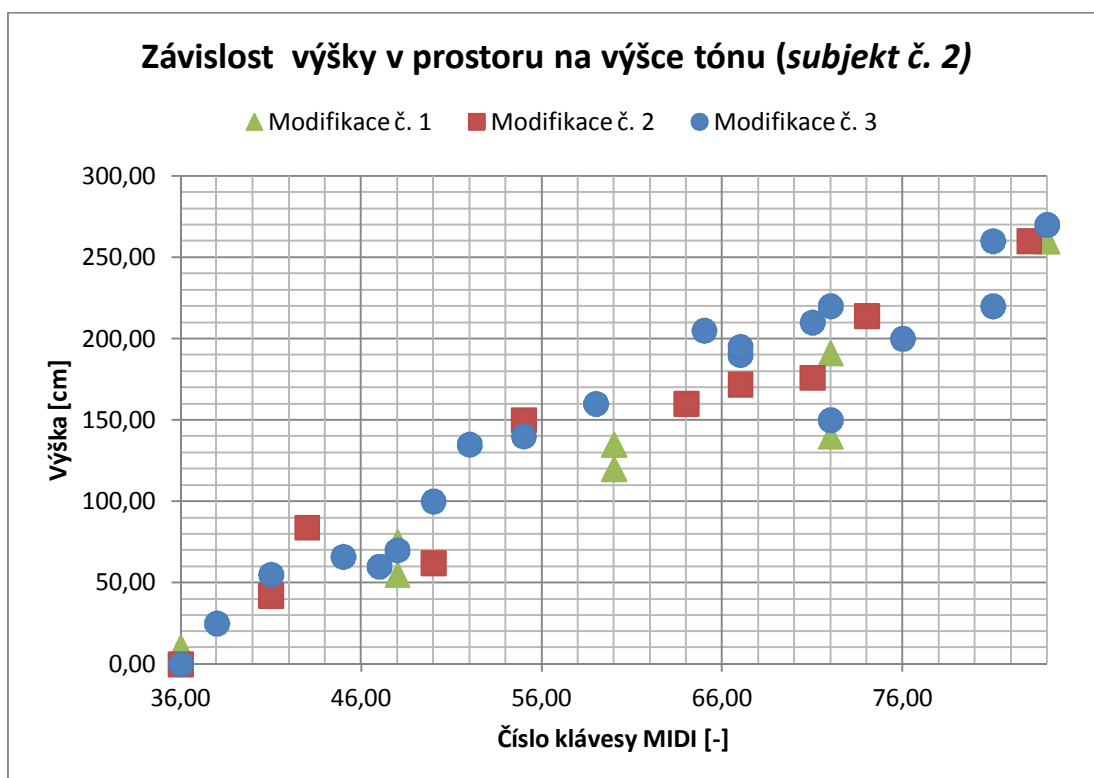
8 SOUBORY NA PŘILOŽENÉM CD

sonifikace.pd	- hlavní program
readme.txt	- návod k ovládní programu
C_dur.txt	- dokument sloužící k tvorbě stupnice C dur
TEST_elevace_zadani.xls	- zadání pro testy elevace
TEST_elevace_vysledky.xls	- naměřené hodnoty a grafy testů elevace
Poslechovy_test_elevace_1.pd	- test modifikace č. 1 elevace a výšky tónu
Poslechovy_test_elevace_2.pd	- test modifikace č. 2 elevace a výšky tónu
Poslechovy_test_elevace_3.pd	- test modifikace č. 3 elevace a výšky tónu
stupnice_elevace.mp3	- ukázka rozsahu výšek testu elevace
stimul_elevace.mp3	- ukázka stimulu pro test elevace
kalibracni_posloupnost.mp3	- ukázka kalibrační posloupnosti testu elevace
Poslechovy_test_jas.pd	- test vjemu jasnosti tónu
TEST_jasnost_vysledky.xls	- naměřené hodnoty a grafy testů vjemu jasnosti
TEST_rychlost_vysledky.xls	- naměřené hodnoty a grafy testů vjemu rychlosti
TEST_rychlost1.wav	- test pro zjištění rychlosti
TEST_rychlost2.wav	- test pro zjištění rychlosti bez elevace a azimutu
TEST_souhrnny_zadani.pdf	- zadání testů rozpoznání trajektorie a jasu
TEST_souhrnny.wav	- test pro rozpoznání trajektorie a jasu
TEST_souhrnny_vysledky.xls	- výsledky testů rozpoznání trajektorie a jasu
Video_popis_k_testu.avi	- video znázorňující strukturu programu pro test
Video_ukazka_cinnosti.avi	- video obsahující několik sonifikovaných pohybů
Prilohy.txt	- textový dokument s popisem jednotlivých příloh

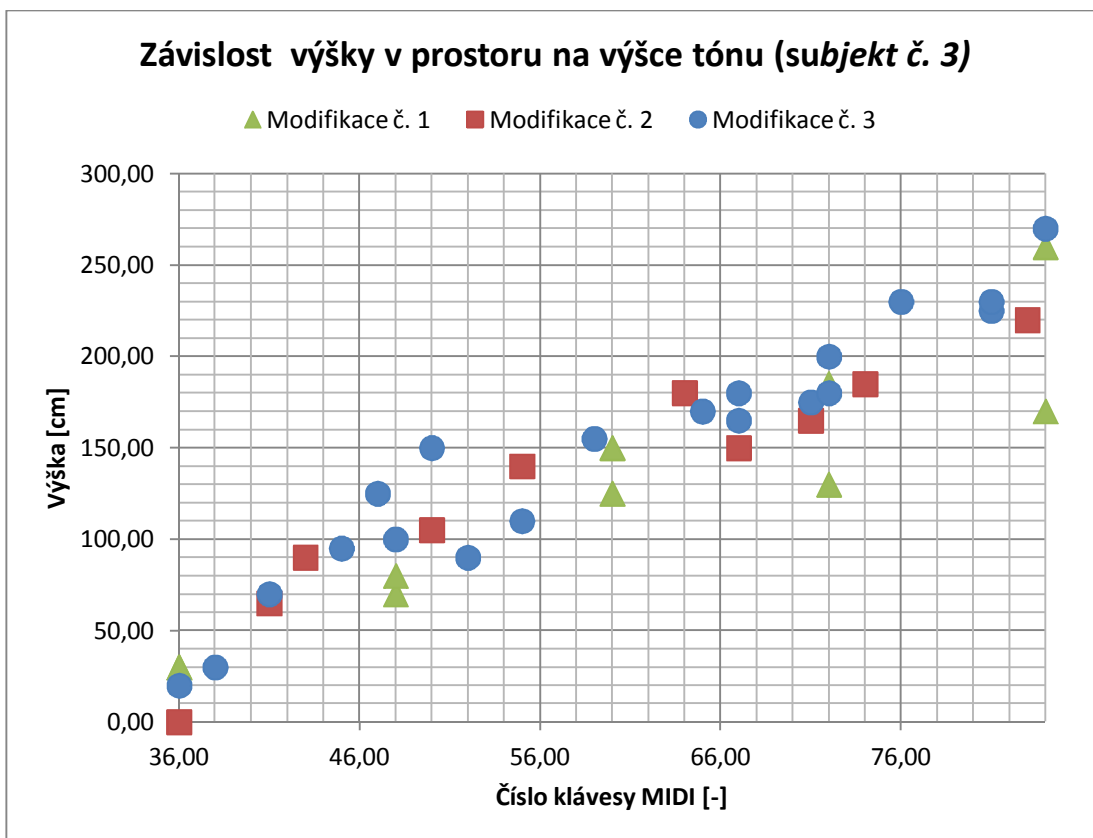
9 PŘÍLOHY



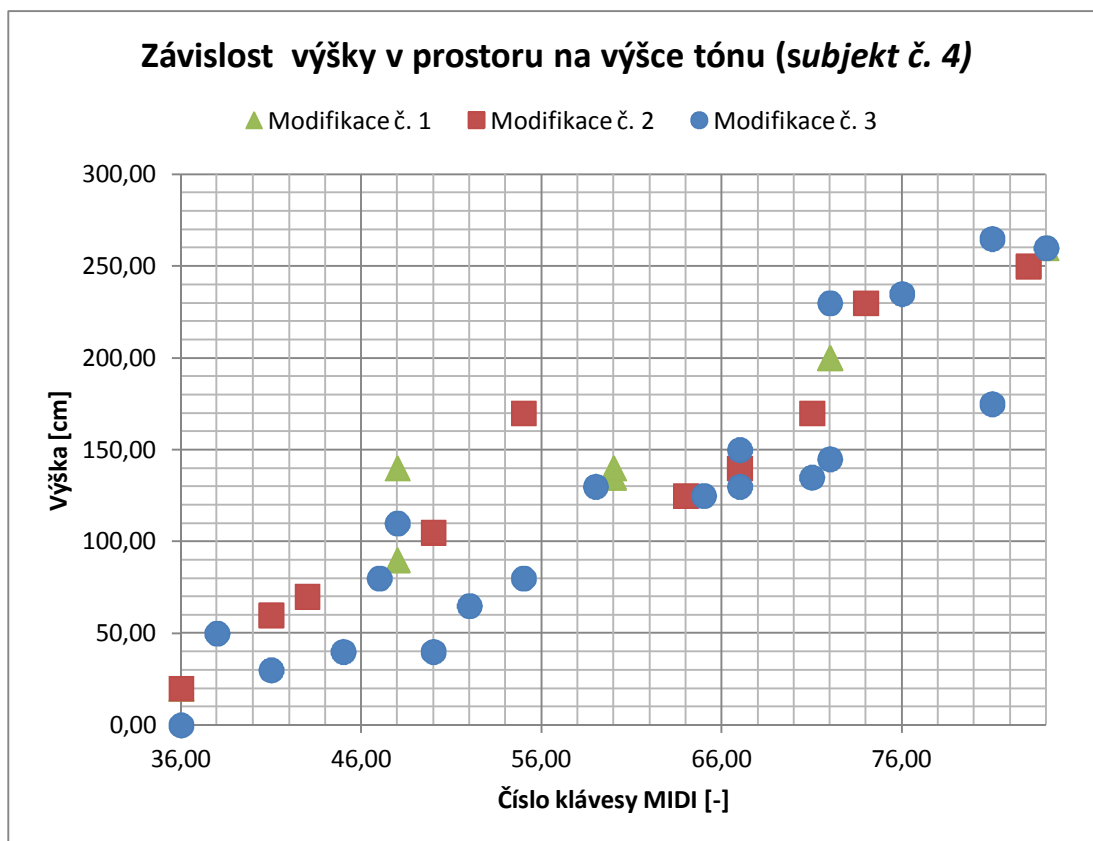
Obrázek 31: Závislost výšky v prostoru na výšce tónu subjektu č. 1 všech modifikací



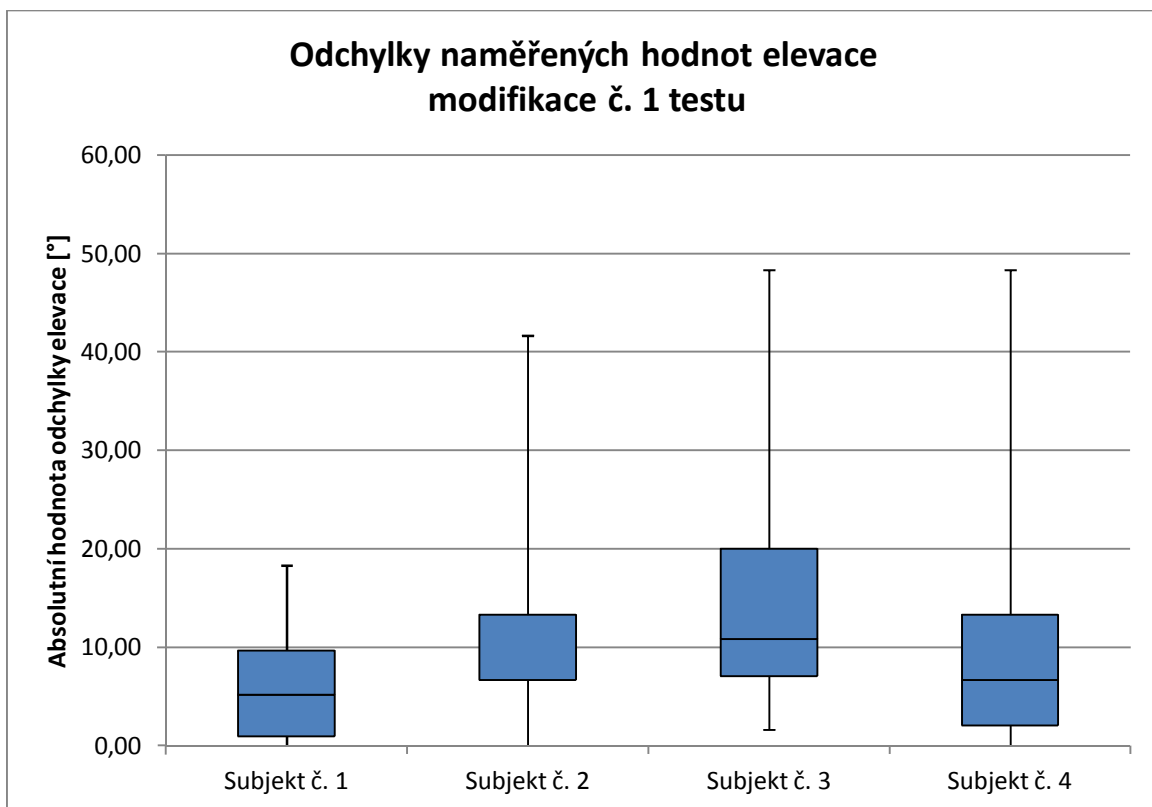
Obrázek 32: Závislost výšky v prostoru na výšce tónu subjektu č. 2 všech modifikací



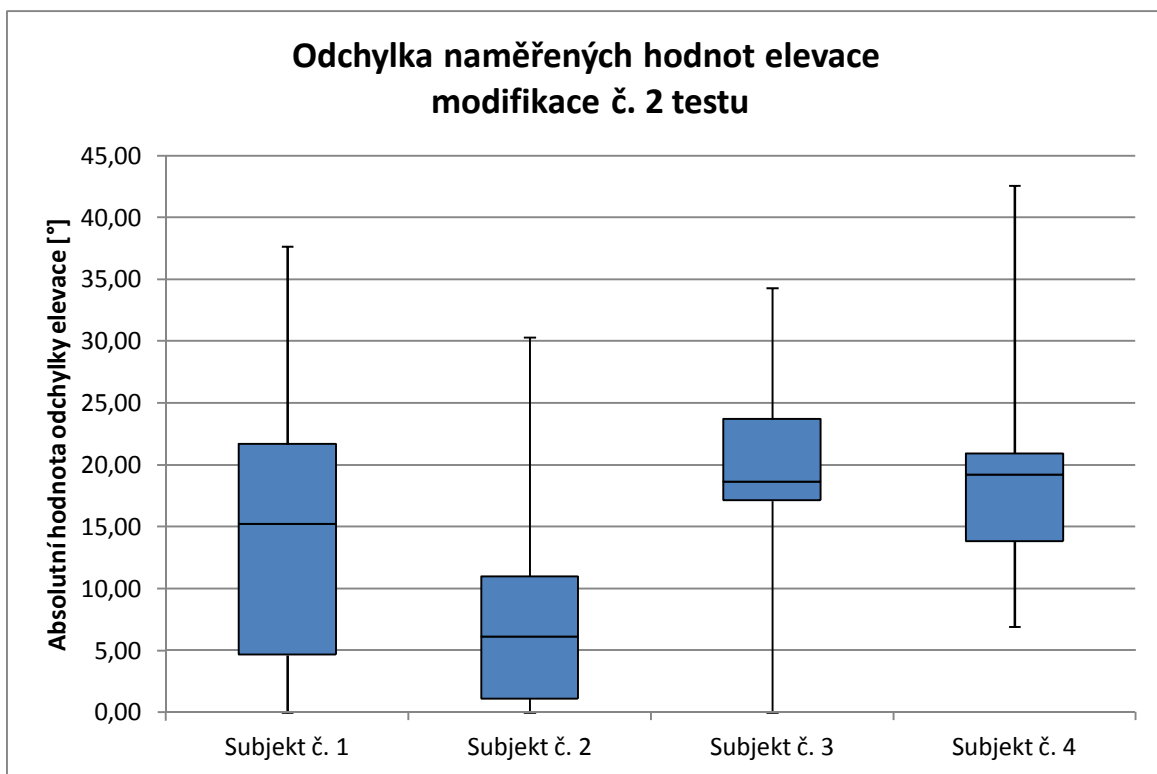
Obrázek 34: Závislost výšky v prostoru na výšce tónu subjektu č. 3 všech modifikací



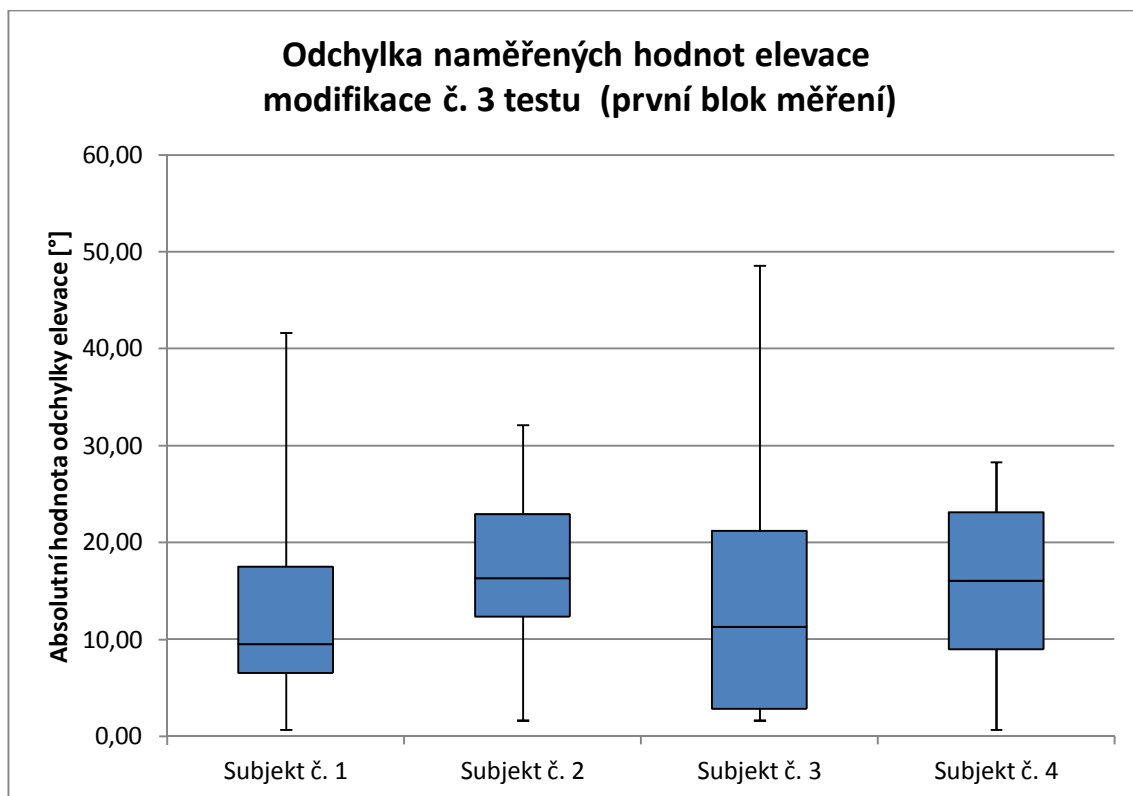
Obrázek 35: Závislost výšky v prostoru na výšce tónu subjektu č. 4 všech modifikací



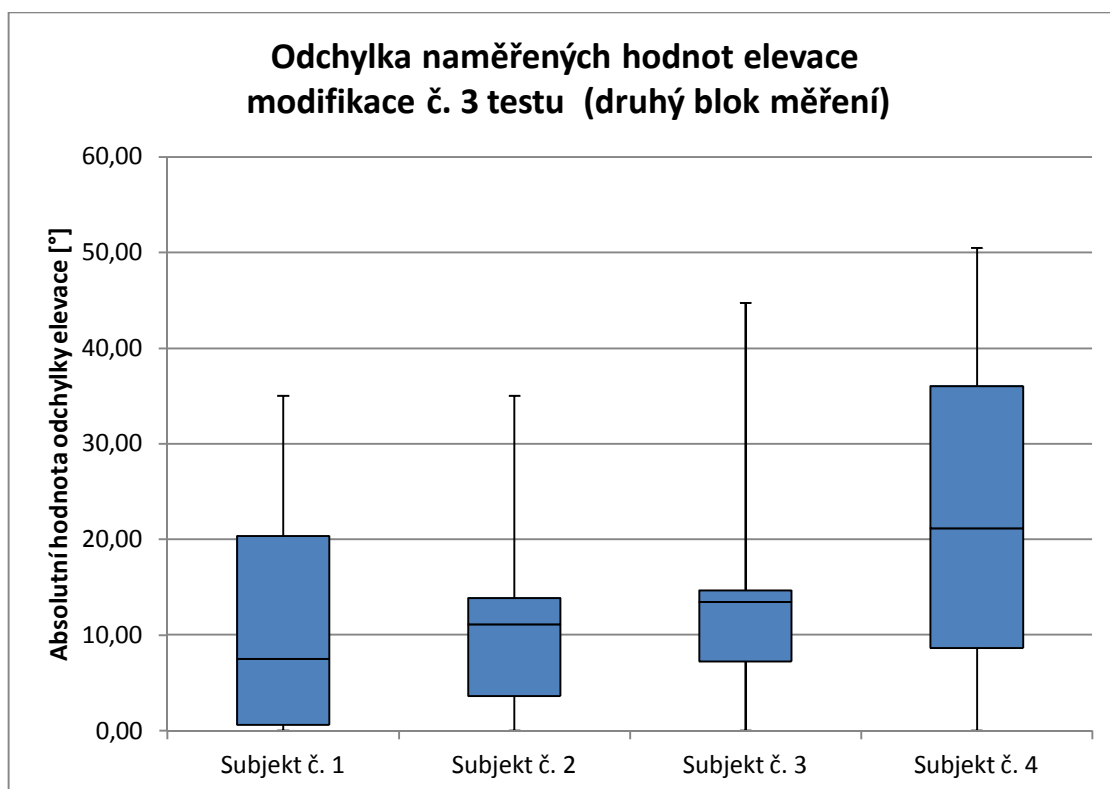
Obrázek 36: Krabicový graf odchytek naměřených hodnot elevace modifikace č. 1 testu



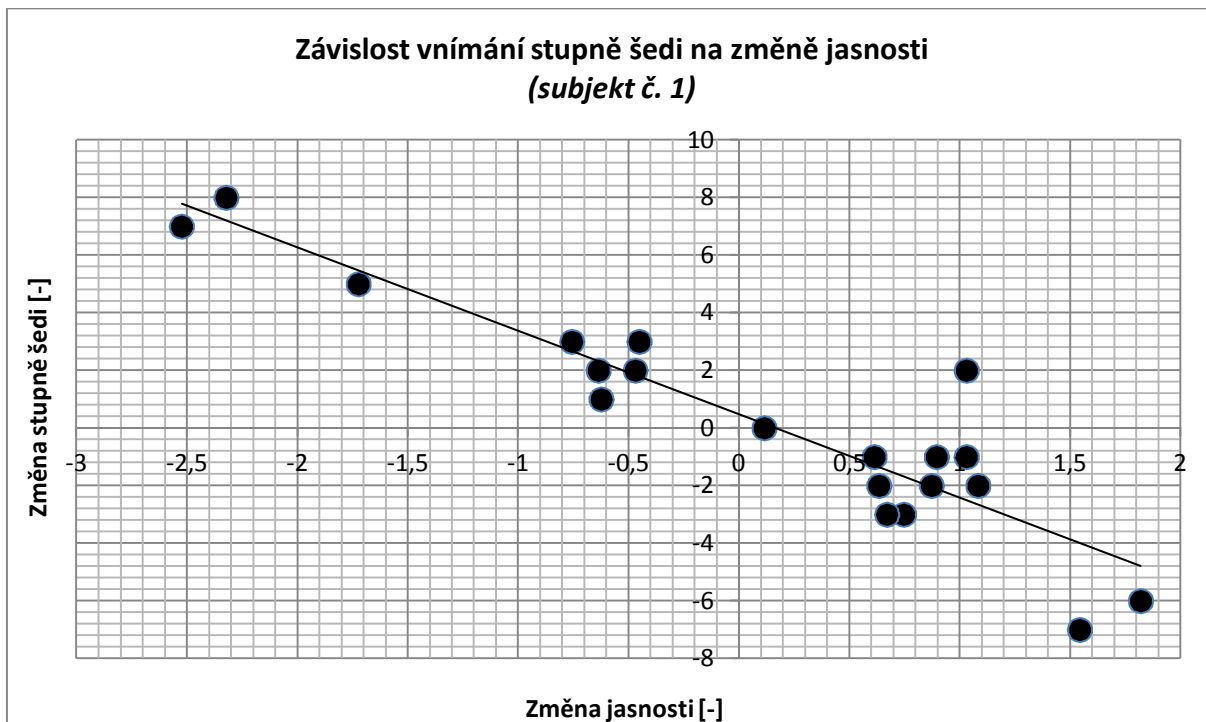
Obrázek 37: Krabicový graf odchytek hodnot elevace modifikace č. 2 testu



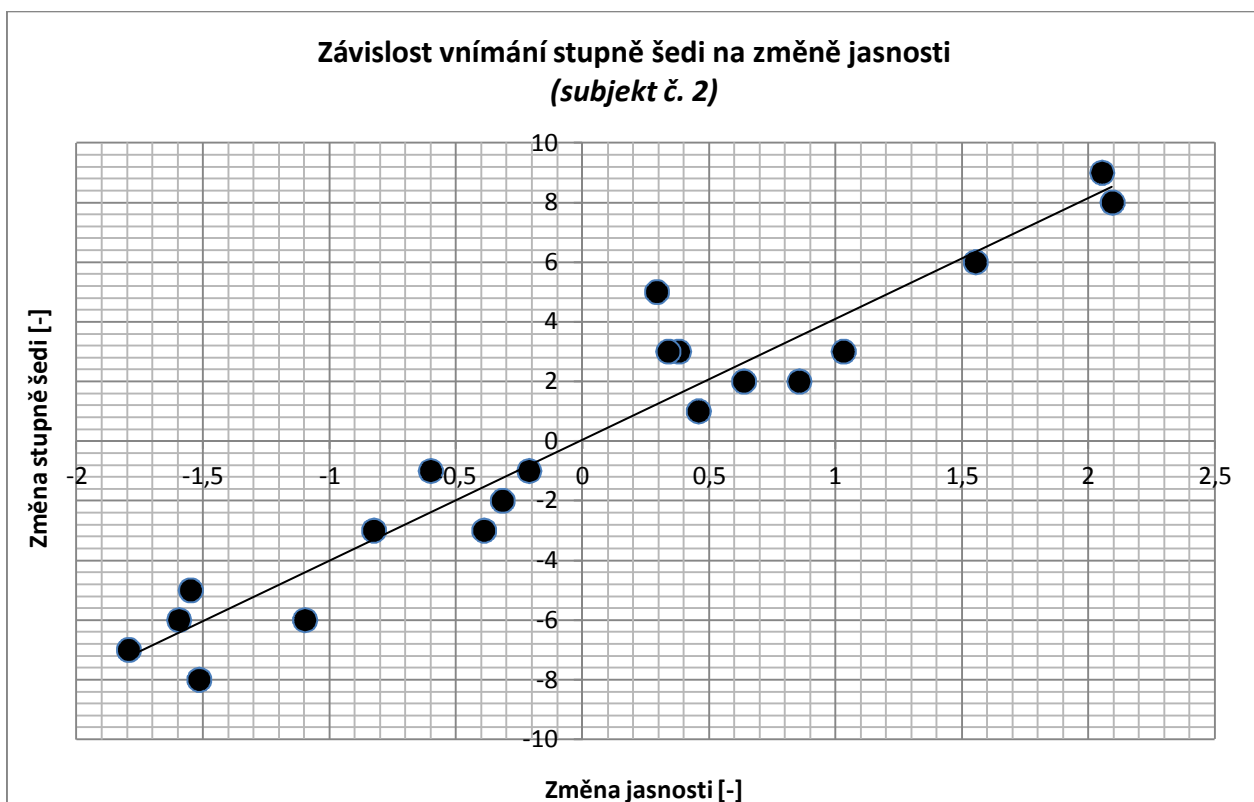
Obrázek 38: Krabicový graf odchylek naměřených hodnot elevace modifikace č. 3 testu pro první blok experimentu



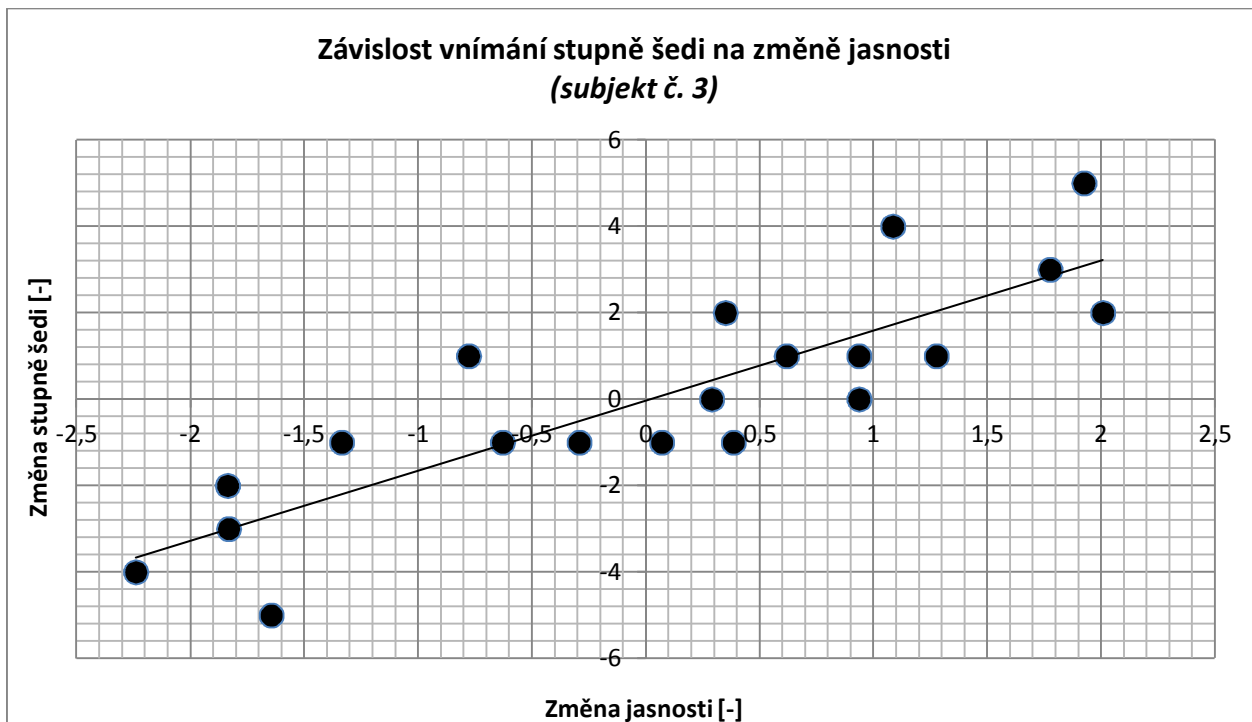
Obrázek 39: Krabicový graf odchylek naměřených hodnot elevace modifikace č. 3 testu pro druhý blok experimentu



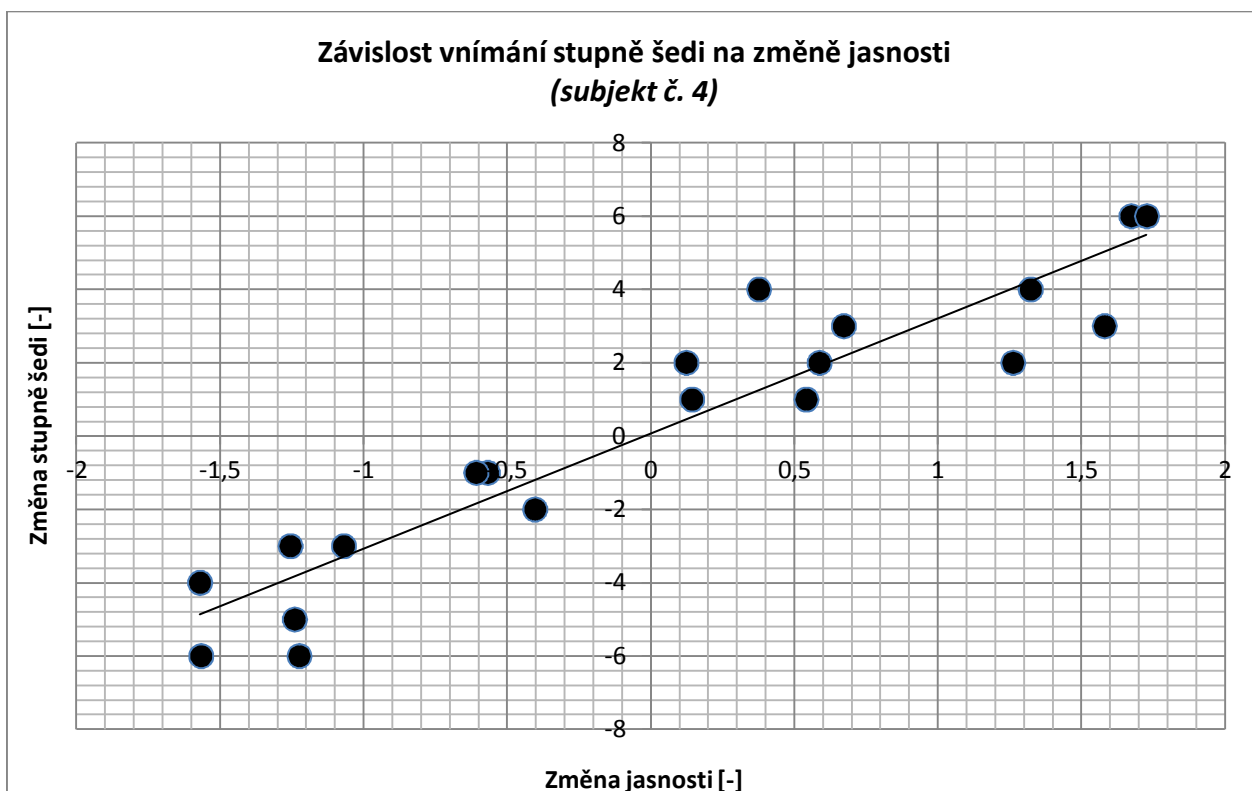
Obrázek 40: Závislost subjektu č. 1 vnímání změny stupně šedi v závislosti na změně jasnosti tónu včetně lineárního proložení naměřených hodnot



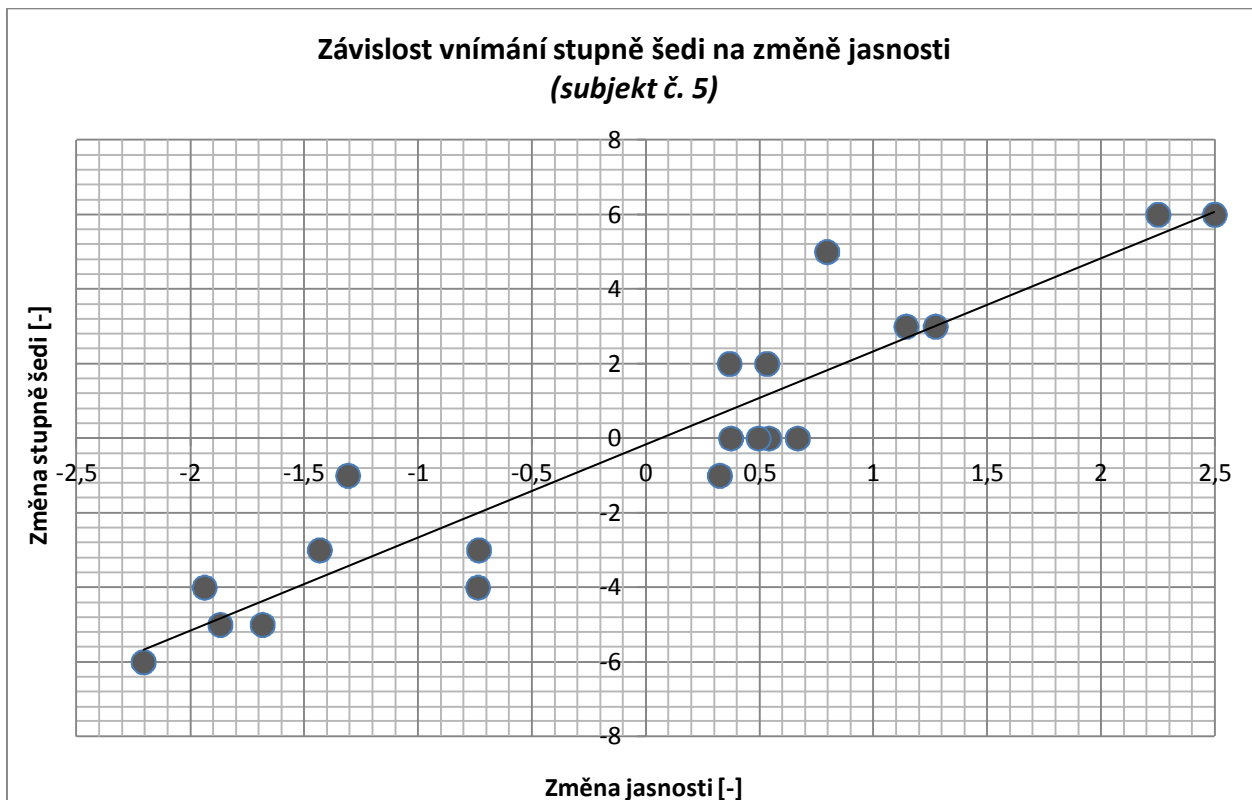
Obrázek 41: Závislost subjektu č. 2 vnímání změny stupně šedi v závislosti na změně jasnosti tónu včetně lineárního proložení naměřených hodnot



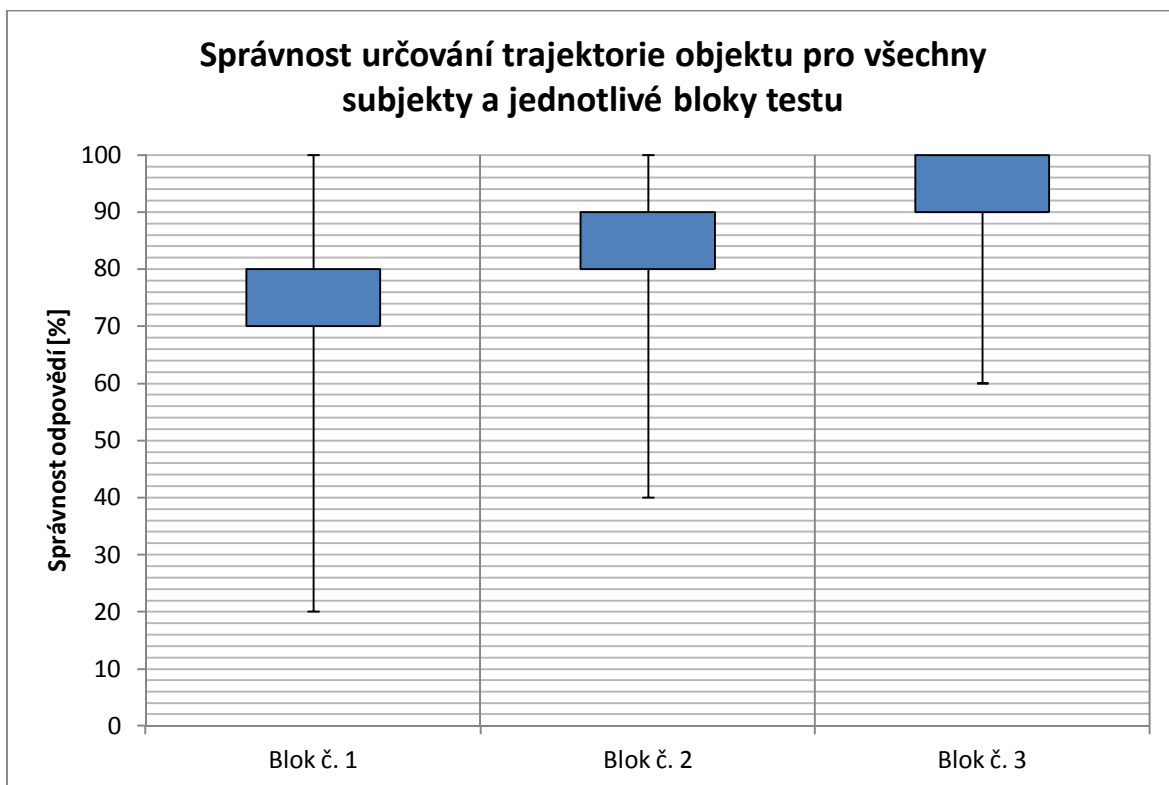
Obrázek 42: Závislost subjektu č. 3 vnímání změny stupně šedi v závislosti na změně jasnosti tónu včetně lineárního proložení naměřených hodnot



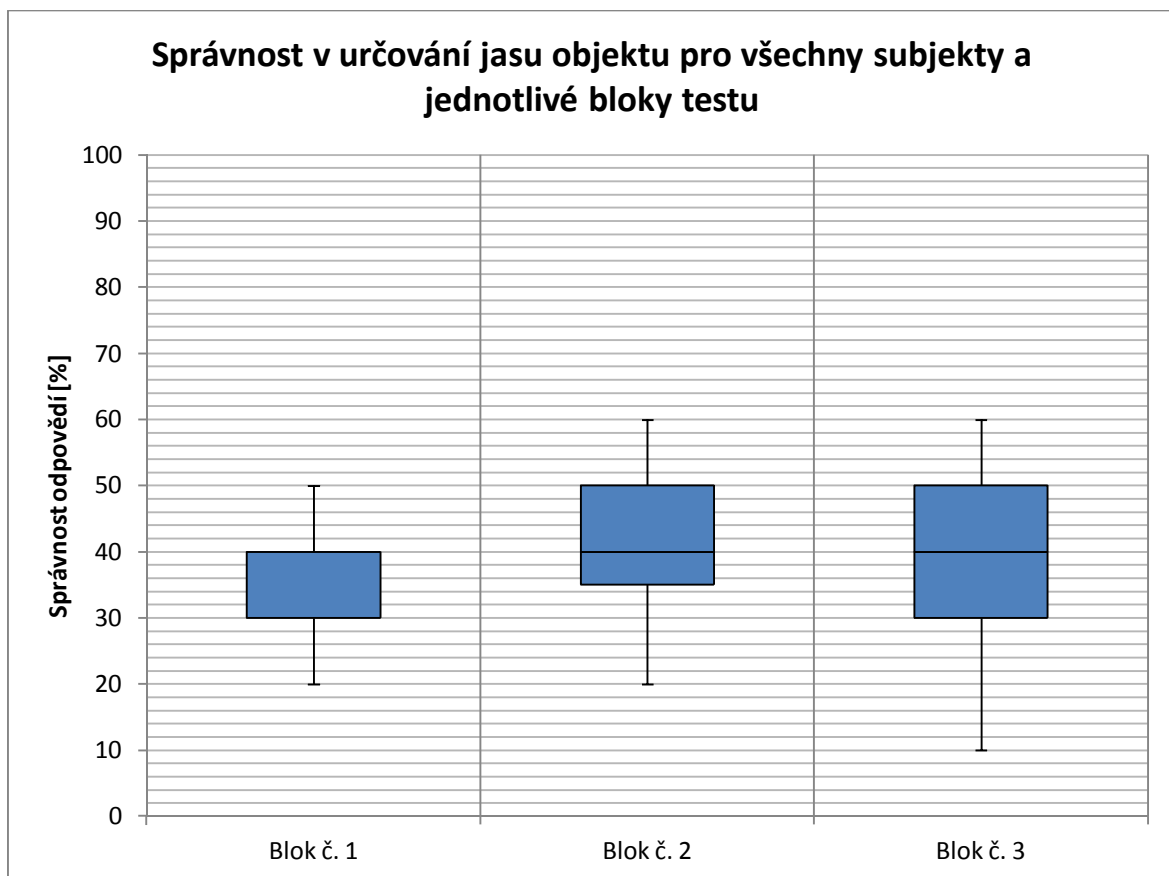
Obrázek 43: Závislost subjektu č. 4 vnímání změny stupně šedi v závislosti na změně jasnosti tónu včetně lineárního proložení naměřených hodnot



Obrázek 43: Závislost subjektu č. 5 vnímání změny stupně šedi v závislosti na změně jasnosti tónu včetně lineárního proložení naměřených hodnot



Obrázek 44: Krabicový graf správnosti určování trajektorie objektu pro všechny subjekty; jedná se o souhrnný test celého algoritmu; medián a dolní kvartil nabývají stejné hodnoty



Obrázek 45: Krabicový graf správnosti určování jasu objektu pro všechny subjekty; jedná se o souhrnný test celého algoritmu; medián v bloku č. 1 splývá s dolním kvartilem