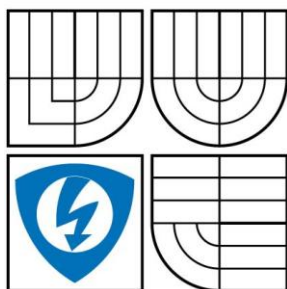


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKACNÍCH
TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

POPIS A REPREZENTACE DVOUROZMĚRNÝCH ZVUKOVÝCH SCÉN VE VÍCEKANÁLOVÝCH SYSTÉMECH REPRODUKCE ZVUKU

DESCRIPTION AND REPRESENTATION OF TWO-DIMENSIONAL SCENES IN
MULTICHANNEL SYSTEMS OF SOUND REPRODUCTION

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

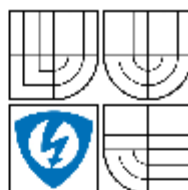
AUTOR PRÁCE
AUTHOR

Bc. MICHAL TRZOS

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. JIŘÍ SCHIMMEL Ph.D.

BRNO 2009



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Diplomová práce

magisterský navazující studijní obor
Telekomunikační a informační technika

Student: Bc. Michal Trzos
Ročník: 2

ID: 80477
Akademický rok: 2008/2009

NÁZEV TÉMATU:

**Popis a reprezentace dvourozměrných zvukových scén
ve vícekanálových systémech reprodukce zvuku**

POKYNY PRO VYPRACOVÁNÍ:

Prostudujte metody popisu a reprezentace dvourozměrných zvukových scén ve vícekanálových systémech reprodukce zvuku. Zaměřte se zejména na metody vektorově-bázového a amplitudově-bázového panoramování a využití ambisonie. V prostředí Matlab otestujte algoritmy různých metod reprezentace dvourozměrných zvukových scén a vzájemně je porovnejte. Vybrané algoritmy implementujte v programovacím jazyce C++ jako plug-in modul technologie VST pro zpracování signálu v reálném čase.

DOPORUČENÁ LITERATURA:

- [1] GARDNER, W., G. 3-D Audio Using Loudspeakers. Springer, 1998. ISBN 978-0792381563
- [2] BHARITKAR, S., KYRIAKAKIS, C. Immersive Audio Signal Processing. Springer, 2006. ISBN 978-0387284538
- [3] WIGGINS, B. An Investigation into the Real-Time Manipulation and Control of Three-Dimensional Sound Fields. PhD Thesis, University of Derby, 2004.

Termín zadání: 9.2.2009

Termín odevzdání: 26.5.2009

Vedoucí práce: Ing. Jiří Schimmel, Ph.D.

prof. Ing. Kamil Vrba, CSc.
Předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

ABSTRAKT

Tato práce se zabývá podněty, které ve sluchové soustavě člověka způsobují prostorové slyšení a polohovacími metodami, které těchto podnětů využívají. Podrobněji se zabývá metodou vektorově bázového amplitudového polohování a metodou ambisonie. Na základě popisu těchto metod byl realizován návrh a implementace ve formě plug-in modulu systému VST. Tato práce také obsahuje poslechový test ambisonie druhého řádu a analýzu získaných dat.

ABSTRACT

This thesis deals with cues used by the human auditory system to identify the location of sound and methods for sound localisation based these cues, namely, vector based amplitude panning and ambisonics, which are described in detail. These methods have been implemented as a VST plug-in module. This thesis also contains listening tests of second order ambisonics along with acquired data analysis.

KLÍČOVÁ SLOVA

ambisonie, ambisonie vyšších řádů, polohovací metody, amplitudové polohování, VST, vektorově bázové amplitudové polohování

KEYWORDS

ambisonics, high order ambisonics, HOA, VBAP, vector based amplitude panning, maxre, energy decoder, inphase decoder, velocity decoder

TRZOS, M. *Popis a reprezentace dvourozměrných zvukových scén ve vícekanálových systémech reprodukce zvuku*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. 62 s. Vedoucí diplomové práce Ing. Jiří Schimmel, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Popis a reprezentace dvourozměrných zvukových scén ve vícekanálových systémech reprodukce zvuku“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny uvedeny v seznamu literatury na konci práce. Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne 25. 5. 2009

.....
podpis autora

PODĚKOVÁNÍ

Děkuji vedoucímu diplomové práce Ing. Jiřímu Schimmelovi Ph.D. za metodickou pomoc a zorganizování poslechového testu.

V Brně dne 25. 5. 2009

.....

podpis autora

SEZNAM POUŽITÝCH SYMBOLŮ

$x_i(t), x(t)$ signály se spojitým časem

g_i zesilovací faktory

$\theta, \theta_0, \theta_S, \theta_T$ azimut, úhel mezi dvěma body v jedné rovině

α elevace, úhel mezi dvěma rovinami, z nichž jedna je referenční a horizontální

$\mathbf{p}, \mathbf{l}, \mathbf{L}, \mathbf{g}$ matice

$\mathbf{A}^T$ transponovaná matice $\mathbf{A}$

$\mathbf{A}^{-1}$ inverzní matice $\mathbf{A}$

$\mathbf{A}^+$ pseudoinverzní matice $\mathbf{A}$

$\mathbf{A}^*$ hermitovsky sdružená matice

$W, X, Y, Z, B_{mn}^\sigma$ ambisonické signály

Y_{mn}^σ sférické harmonické funkce řádu m , stupně n , kde $\sigma = \pm 1$ určuje pozitivní
a negativní stupně sférické harmonické funkce

$j_m(kr)$ sférická Besselova funkce řádu m

$h_m^-(kr)$ sférická Hankelova funkce řádu m

SEZNAM POUŽITÝCH ZKRATEK

IID Interaural intensity difference, interaurální rozdíl intenzity

ITD Interaural time difference, interaurální časový rozdíl

VBAP Vector base amplitude panning, vektorově bázové amplitudové polohování

IQR Interquartile range, interkvartilová vzdálenost

SVD Singular value decomposition, dekompozice singulárních hodnot

JAMA Balík funkcí pro práci s maticemi

TNT..... Template numerical toolkit, nástroj pro práci s maticemi

OBSAH

1	Úvod.....	11
2	Psychoakustika a směrové slyšení.....	12
2.1	Interaurální časový rozdíl.....	12
2.2	Interaurální rozdíl intenzity.....	15
3	Metody prostorové reprodukce zvuku.....	16
3.1	Amplitudové polohování.....	16
3.2	Stereofonie.....	17
3.3	Vektorově bazové amplitudové polohování.....	18
3.3.1	Formulace vektorové báze.....	19
3.3.2	Dvojměrné vektorově bazové amplitudové polohování pro dva a více reproduktorů.....	21
3.4	Ambisonie.....	22
3.4.1	Matematické základy.....	23
3.4.2	Ambisonické kódování.....	24
3.4.3	Ambisonické dekódování.....	27
3.4.4	Psychoakustická optimalizace ambisonického dekódu.....	29
3.4.5	Vektor energie a rychlosti.....	30
3.4.6	Optimalizovaný dekódér.....	31
3.4.7	Filtry pro optimalizovaný dekódér.....	34
4	Návrh a implementace algoritmu.....	37
4.1	Úvod.....	37
4.2	Virtual Studio Technology.....	38
4.3	Implementace metody vektorově bazového amplitudového polohování.....	38
4.3.1	Třídící algoritmus.....	40
4.3.2	Výpočet a normalizace zesilovacích faktorů.....	40
4.4	Implementace metody ambisonie.....	42
4.4.1	Ambisonický kódér.....	44
4.4.2	Ambisonický dekódér.....	44
5	Poslechový test.....	48
5.1	Metoda vyhodnocení lokalizace.....	48
5.2	Testovaný systém.....	48
5.3	Testovací signál.....	49
5.4	Průběh testu.....	49
5.5	Vyhodnocení testu a analýza hodnot.....	50

6	Závěr	52
6.1		52
7	Literatura.....	53
A	Přílohy k textu.....	57
A.1	Vývojový diagram bubble sortu pro metodu vektorově bázového amplitudového polohování	57
A.2	Konfigurace použita v poslechovém testu	58
A.3	Tabulka naměřených hodnot z poslechového testu.....	59
B	Příloha k CD	60
B.1	Adresářová struktura.....	60
B.2	Skripty pro prostředí Matlab	61

SEZNAM OBRÁZKŮ

Obr. 2.1: Znázornění dvou cest, které musí urazit zvuk ze zdroje do obou uší.....	13
Obr. 2.2: Znázornění binaurálního rozdílu vzdáleností u diskriminačního práhu.	14
Obr. 3.1: Zobrazení konfigurace pro stereofonní poslech.....	17
Obr. 3.2: Vektorový popis stereofonní konfigurace.....	20
Obr. 3.3: Dvojdímenzionální VBAP s pěti reproduktory. Vektory $\mathbf{l}_n$ ukazují na reproduktory; reproduktorové vektorové báze $\mathbf{L}_{mn}$ jsou formovány přilehlými reproduktory.	21
Obr. 3.4: Horizontální ambisonické komponenty prvního (nahore) a druhého řádu (dole), kde $\pm$ značí fázi.	25
Obr. 3.5: Kódování rovinné vlny přicházející ze směru šipky do ambisonických signálů prvního řádu.....	26
Obr. 3.6: Velikost vektoru energie r_E jako funkce azimutu virtuálního zdroje pro různé reproduktorové konfigurace; vektor rychlosti r_V je roven jedné.	31
Obr. 3.7: Grafická reprezentace směrových charakteristik dosažitelných ambisonií prvního řádu.	34
Obr. 3.8: Charakteristika fázově kompenzovaných filtrů pro $f_c = 370 \text{ Hz}$	36
Obr. 4.1: Blokové schéma implementace metody vektorově bazového amplitudového polohování.	39
Obr. 4.2: Zobrazení normalizovaných zesilovacích faktorů pro metodu VBAP.	42
Obr. 4.3: Kolísání velikosti vektoru energie při nedostatečném počtu reproduktorů; Ambisonie třetího řádu, šest reproduktorů.	43
Obr. 4.4: Blokové schéma ambisonického dekodéru.	45
Obr. 4.5: Směrové charakteristiky virtuálního mikrofону pro ambisonický dekodér prvního řádu bez optimalizace, s optimalizací $\max r_E$ a „ve fázi“.	47
Obr. 5.1: Signál použitý v poslechovém testu.....	49
Obr. 5.2: Cílové a naměřené azimuty pro ambisonický systém druhého řádu; Symboly reprezentují 25. kvartil, medián a 75. kvartil.....	50
Obr. 5.3: Interkvartilové rozsahy pro jednotlivé měřené azimuty poslechového testu	51

1 ÚVOD

V posledních letech stoupá obliba vícekanálových systémů pro reprodukci zvuku, které se díky nízké ceně dostávají i do domácností. Zajímavým paradoxem je, že vícekanálové systémy původně nevznikly za účelem obohacení hudebního poslechu, nýbrž k obohacení kinematografického zážitku. Motivace zavedení vícekanálových systémů je však společná, a to vytvoření zvukového pole obklopujícího posluchače, který vytváří iluzi jiného prostředí. Toto prostředí může být reálné, reprodukováné ze záznamu, či syntetické, uměle vytvořené.

Proto, abychom dokázali zaznamenávat a reprodukovat v našem poslechovém prostředí zvuková pole jiných prostředí, musíme nejdříve prozkoumat jakým způsobem lidský sluch tato zvuková pole vnímá. Podněty prostorového vnímání se zabývá kapitola 2.

Jsou-li podněty prostorového vnímání dostatečně popsány, můžeme na jejich základě definovat metody, které nám umožní manipulovat se záznamem, reprodukcí a syntézou zvukových scén. Veškeré popsané metody budou zjednodušeny do dvojrozměrného prostoru. Věnovat se budeme hlavně vektorově bázovému amplitudovému polohování a ambisonii vyšších řádů. Těmito metodami se bude zabývat kapitola 3.

Kapitola 4 řeší návrh a implementaci popsaných metod jako plug-in modulu technologie VST v programovacím jazyce C++. Jsou zde popsány implementační detaily, které bylo potřeba vyřešit v jednotlivých etapách implementace.

Kapitola 5 se zabývá poslechovou zkouškou. Je zde popsána metodika poslechové zkoušky, analýza a interpretace výsledků.

2 PSYCHOAKUSTIKA A SMĚROVÉ SLYŠENÍ

Dříve, než lze začít definovat metody pro prostorovou reprodukci zvuků, je třeba prozkoumat způsob, jakým lidská sluchová soustava vnímá zdroje zvuku v prostoru. Tyto průzkumy jsou prováděny zejména empirickou korelací působení fyzikálních parametrů zvuku a subjektivního vnímání zvuku.

Lidské ucho je orgán nedokonalý a obzvláště díky jeho nedokonalostem vzniká prostor pro optimalizaci systémů pro reprodukci zvuku z hlediska počtu reprodukčních kanálů. V této kapitole budou probrány některé z podnětů uplatňujících se při prostorovém slyšení a v následující kapitole budou definovány zákony pro kvalitativní posouzení prostorových reprodukčních systémů.

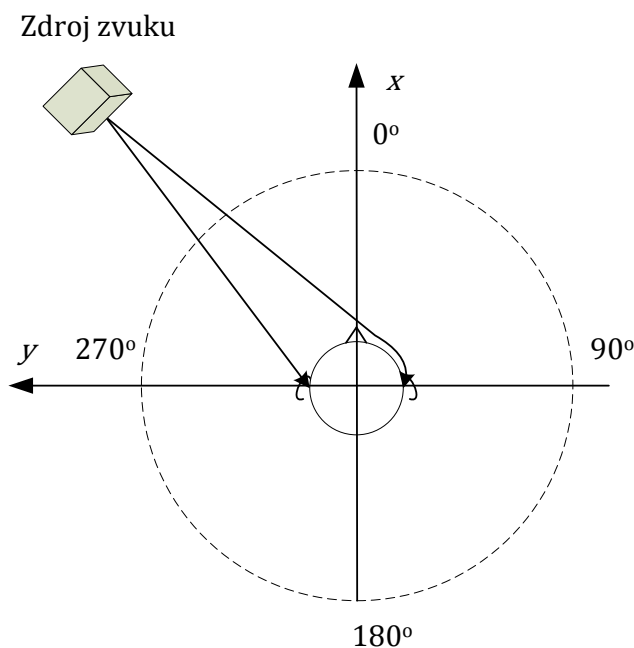
Podněty, kterých sluchová soustava využívá pro vnímání směrovosti zvuku, se z pohledu získávání podnětů pro analýzu dělí na dvě základní skupiny, na binaurální a monaurální. Binaurální jevy se zakládají na porovnávání zvukového vstupu z dvou odlišných detektorů. Monaurální jevy se zakládají spíše na sluchové paměti a směrové charakteristice ušního boltce.

Je pravidlem, že se lokalizace v prostoru daří lépe u složených zvuků a hluků než u čistých tónů [4]. Tento fakt je dán zkušenostmi získanými během života, spojováním zrakových a sluchových vjemů a závisí na několika faktorech, z nichž žádný sám o sobě nestačí k přesnému umístění zdroje.

2.1 Interaurální časový rozdíl

Interaurální časový rozdíl – ITD (Interaural time difference) je rozdíl doby, za kterou zvuk urazí vzdálenost mezi oběma ušima. Máme-li situaci, ve které je v horizontální rovině reprodukován zvuk, tak úhel, který svírá hlava a zdroj zvuku nazýváme *azimut*, kde 0° je azimut směřující přímo před posluchače, 90° napravo od posluchače, 270° nalevo od posluchače a 180° přímo za posluchačem (viz obr. 2.1).

Je zřejmé, že největší časový rozdíl bude pro zdroje zvuku umístěné na azimutu 90° a 270° . Naopak nulový časový rozdíl bude pro zdroje zvuku umístěné na azimutu 0° a 180° . To by ovšem znamenalo, že zdroje zvuku před posluchačem a za posluchačem budou nerozlišitelné.



Obr. 2.1: Znázornění dvou cest, které musí urazit zvuk ze zdroje do obou uší.

Není tomu tak díky směrové filtraci boltce a dynamické lokalizaci, které v takovýchto případech poskytují přídatné podněty pro lokalizaci zvukového zdroje.

Uvažujeme-li, že rychlost šíření zvuku v suchém vzduchu při teplotě 20 °C rovna $343 \text{ m} \cdot \text{s}^{-1}$ a průměr průměrné lidské hlavy (založený na modelu koule s ušima na azimutu 90° a 270°) je roven 0,15 m, můžeme vypočítat maximální vzdálenost, kterou zvuk urazí od jednoho ucha k druhému jako polovinu obvodu kružnice

$$d = \pi r = 0,075\pi \doteq 0,2355 \text{ m}, \quad (2.1)$$

kde d je polovina obvodu kružnice a r je poloměr hlavy. Nyní můžeme vypočítat čas, za který zvuk tuto vzdálenost urazí jako

$$ITD_{\max} = \frac{d}{v} = \frac{0,2355}{343} \doteq 0,687 \text{ ms}, \quad (2.2)$$

kde $ITD_{\max}$ je maximální časový rozdíl mezi dvěma ušima, d je vzdálenost, kterou urazí zvuk od jednoho ucha ke druhému a v je rychlost zvuku ve vzduchu.

Z tohoto vyplývá, že pro hlavu o průměru 0,15 m bude maximální časový rozdíl $ITD_{\max} \doteq 0,687 \text{ ms}$.

Lze definovat i minimální časový rozdíl, který je lidským uchem rozeznatelný. Tento minimální rozdíl se nazývá *diskriminační práh*, budeme jej označovat $ITD_{\min}$. Tento práh byl empiricky určen na úhel 4° [1], jak je vidět na obr. 2.2, můžeme díky němu získat binaurální rozdíl vzdáleností

$$D = d_1 + d_2, \quad (2.3)$$

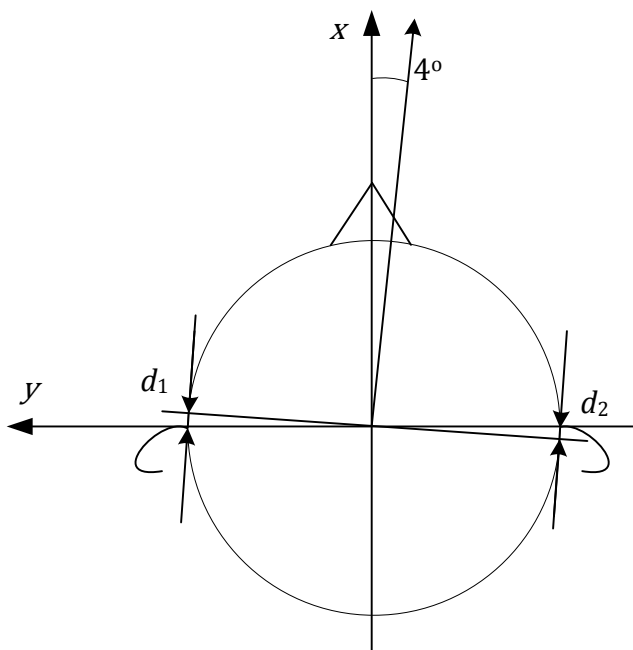
kde D je binaurální rozdíl vzdáleností, d_1 a d_2 jsou odchylky od mediální roviny. Protože d_1 a d_2 jsou totožné, můžeme vztah (2.3) napsat jako

$$D = d \sin \alpha, \quad (2.4)$$

kde d je průměr hlavy a α je diskriminační práh. Můžeme tedy použít znovu rovnici (2.2) pro výpočet časové hodnoty diskriminačního práhu

$$ITD_{\min} = \frac{D}{v} = \frac{d \sin \alpha}{v} = \frac{0,15 \sin 4^\circ}{343} = 30,5 \mu\text{s}. \quad (2.5)$$

Lidské ucho tedy není za daných podmínek schopno určit směrovost zvuku, u kterého je interaurální časový rozdíl menší než $30,5 \mu\text{s}$.



Obr. 2.2: Znázornění binaurálního rozdílu vzdáleností u diskriminačního práhu.

2.2 Interaurální rozdíl intenzity

Zatímco interaurální časový rozdíl – IID (Interaural intensity difference) je nezávislý na kmitočtu, interaurální rozdíl intenzity závislý je vlivem *akustického stínu* hlavy. U nízkých kmitočtů, s vlnovou délkou větší než je průměr hlavy, dochází při kontaktu s hlavou k difrakci okolo hlavy, zatímco u vysokých kmitočtů, s vlnovou délkou menší, než je průměr hlavy, bude zvuk hlavou utlumen. Hlava takto působí jako dolní propust. Kmitočet odpovídající vlnové délce průměru hlavy můžeme vypočítat podle rovnice

$$f = vd^{-1} = 343 \cdot 0,15^{-1} = 2,28 \text{ kHz}, \quad (2.6)$$

kde v je rychlost zvuku ve vzduchu a d průměr hlavy. Za daných podmínek vychází tento kmitočet jako 2,28 kHz. Ve skutečnosti se začíná sluchová soustava řídit IID dříve. Dle různých zdrojů [2], [3], [4] je to 1500 Hz. Do 1500 Hz tedy sluchová soustava určuje směrovost podle ITD, od 1500 Hz ji určuje podle IID. Tento přechod je pozvolný [4].

3 METODY PROSTOROVÉ REPRODUKCE ZVUKU

Pojem *prostorová reprodukce zvuku* označuje metody používané k reprodukci nebo syntéze prostorových atributů zvuku. Reprodukci myslíme přehrávání zvuků, které byly předem nahrány. Reprodukce může probíhat přímo (stereofonie), nebo může nejprve podstoupit dekódovací proces (DTS, Ambisonie). Syntézou pak máme namysli vytváření nebo simulaci zvukových scén.

Tyto metody se pak snaží buď o reprodukci zvukového pole, nebo berou v úvahu sluchovou soustavu člověka a snaží se o simulaci jevů, kterými působí zvuk na člověka v prostoru. Tyto jevy byly popsány v kapitole 2.

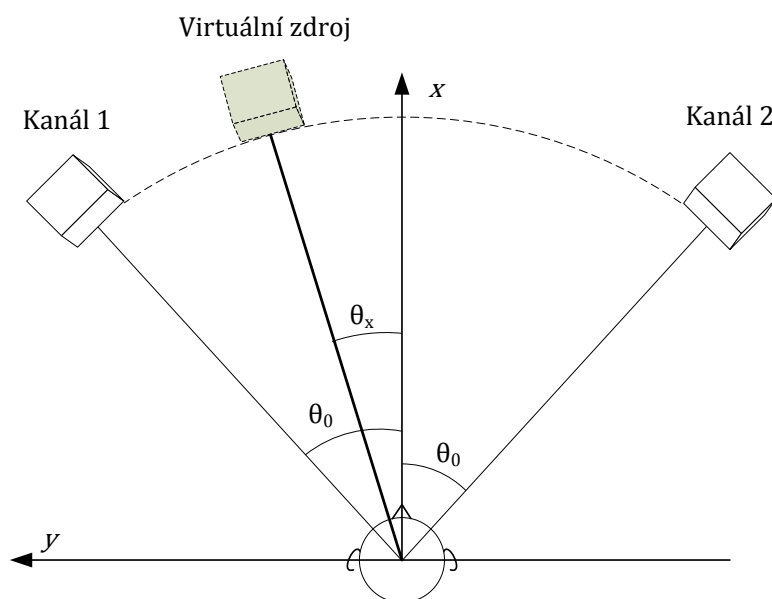
3.1 Amplitudové polohování

Amplitudové polohování je nejčastěji používanou polohovací metodou. Spočívá v aplikaci zesilovacího faktoru na amplitudy signálů jdoucích do reproduktorů. Tato technika je použitelná pro dva a více reproduktorů, které musí být umístěny ekvidistantně kolem posluchače. Obecná rovnice pro amplitudové polohování může být formulována jako

$$x_i(t) = g_i x(t), \quad i = 1, 2, \dots, N,$$

kde $x_i(t)$ je signál jdoucí do reproduktoru i , g_i je zesilovací faktor kanálu i , N je počet reproduktorů a $x(t)$ je vstupní signál.

U jednoduché metody amplitudového polohování dva reproduktory vyzařují koherentní signály, které mohou mít rozdílnou amplitudu. Posluchač vnímá iluzi jediné poslechové události, které říkáme *virtuální zdroj zvuku*. Tu můžeme umístit do horizontálního sektoru, který je tvořen posluchačem a reproduktory, změnou amplitud signálů jdoucích do reproduktorů. Typické rozložení reproduktorů je znázorněno na obr. 3.1.



Obr. 3.1: Zobrazení konfigurace pro stereofonní poslech.

Reproduktory jsou rozmístěny symetricky vůči mediální rovině. Amplitudy signálů jsou ovládány zesilovacími faktory g_1 , respektive g_2 . Reprodukory jsou obvykle umístěny pod úhlem $\theta_0 = 30^\circ$. Pokud se virtuální zdroj pohybuje, je třeba zajistit, aby měl konstantní hlasitost. To je možné provést normalizací zesilovacích faktorů. Zvukový výkon může být nastaven na konstantní úroveň C a můžeme formulovat aproximaci

$$g_1^2 + g_2^2 = C. \quad (3.1)$$

Parametr $C > 0$ v rovnici (3.1) může být považován za řízení hlasitosti virtuálního zdroje. Vnímání vzdálenosti zvuku je do jisté míry řízeno hlasitostí virtuálního zdroje, ale pro přesné řízení vzdálenosti by bylo třeba použít více psychoakustických elementů, jako jsou například odrazy a dozvuky.

3.2 Stereofonie

Stereofonie je speciální případ amplitudového polohování pro dva reproduktory umístěné v horizontální rovině před posluchačem.

Existuje několik zákonů, dle kterých se určuje azimut virtuálního zdroje θ z reproduktorových zesilovacích faktorů. Bauer [5] odvodil sinový polohovací zákon pro situaci, ve které je posluchač umístěn symetricky mezi reproduktory. Tato konfigurace je vidět na obr. 3.1.

Sinový zákon je definován jako

$$\frac{\sin \theta_S}{\sin \theta_0} = \frac{g_1 - g_2}{g_1 + g_2}, \quad (3.2)$$

kde g_1 a g_2 jsou zesilovací faktory reproduktorů, θ_S je vnímaný úhel virtuálního zdroje a θ_0 je úhel, který svírají reproduktory s osou x .

Bernfeld [6] odvodil polohovací zákon vylepšením modelu hlavy použitého u sinového zákona. Tomuto zákonu se říká tangentový zákon

$$\frac{\tan \theta_T}{\tan \theta_0} = \frac{g_1 - g_2}{g_1 + g_2}, \quad (3.3)$$

kde θ_T je vnímaný úhel virtuálního zdroje. Ostatní proměnné jsou stejné jako u sinového zákona. Tomuto zákonu rovněž odpovídá konfigurace na obr. 3.1, kde se za θ dosadí θ_S pro sinový zákon a θ_T pro zákon tangentový.

3.3 Vektorově bázové amplitudové polohování

Vektorově bázové amplitudové polohování – VBAP (Vector base amplitude panning) je polohovací technika umožňující použití neomezeného počtu reproduktorů ve dvourozměrném nebo trojrozměrném prostoru. Reproduktory musí být umístěny ekvidistantně kolem posluchače a poslechová místnost by neměla mít příliš dozvuků. Touto metodou lze vytvářet stacionární i pohyblivé virtuální zdroje zvuku v jakémkoli prostoru osazeném reproduktory [3].

3.3.1 Formulace vektorové báze

U metody vektorově bazového amplitudového polohování je dvoukanálová stereofonní konfigurace přeformulována jako dvojrozměrná vektorová báze. Tato báze je definována jednotkovými vektory $\mathbf{l}_1 = [l_{11} \ l_{12}]^T$ a $\mathbf{l}_2 = [l_{21} \ l_{22}]^T$, které směřují na reproduktory 1, respektive 2, jak je vidět na obr. 3.2. Horní index T představuje transpozici matice. S jednotkovým vektorem $\mathbf{p} = [p_1 \ p_2]^T$, který směřuje k virtuálnímu zdroji, můžeme zacházet jako s lineární kombinací reproduktorových vektorů,

$$\mathbf{p} = g_1 \mathbf{l}_1 + g_2 \mathbf{l}_2. \quad (3.4)$$

V rovnici (3.4) jsou g_1 a g_2 zesilovací faktory, se kterými můžeme zacházet jako s nezápornými skalárními proměnnými. Rovnice (3.4) pak může být zapsána v maticovém tvaru jako

$$\mathbf{p}^T = \mathbf{g} \mathbf{L}_{12}, \quad (3.5)$$

kde $\mathbf{g} = [g_1 \ g_2]$ a $\mathbf{L}_{12} = [\mathbf{l}_1 \ \mathbf{l}_2]^T$. Pokud existuje inverzní matice $\mathbf{L}_{12}^{-1}$, můžeme vypočítat přímo vektor zesilovacích faktorů

$$\mathbf{g} = \mathbf{p}^T \mathbf{L}_{12}^{-1} = [p_1 \ p_2] \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}^{-1}. \quad (3.6)$$

Matice L_{12}^{-1} existuje, pokud je $\theta_0 \neq 0^\circ$ a zároveň $\theta_0 \neq 90^\circ$, v opačném případě by se celá metoda zjednodušila pouze na jeden rozměr.

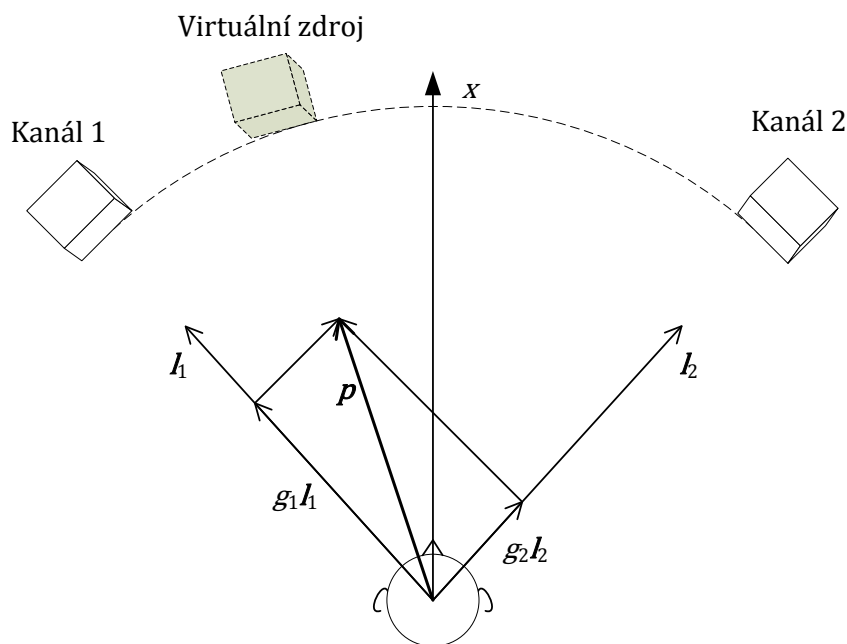
Zesilovací faktory g_1 a g_2 vypočítané rovnicí (3.1) splňují tangentový zákon, což je dokázáno v [3].

Pokud je reproduktorová báze ortogonální, tedy $\theta_0 = 45^\circ$, jsou zesilovací faktory totožné s těmi získanými pro ambisonický kódér [8] s tím rozdílem, že v ambisonickém kódéru mohou být záporné zesilovací faktory. V takovém případě je shodná absolutní hodnota zesilovacích faktorů.

Pokud je $\theta_0 \neq 45^\circ$, musí být zesilovací faktory normalizovány použitím rovnice

$$\mathbf{g}_{\text{vážený}} = \frac{\sqrt{C} \mathbf{g}}{\sqrt{g_1^2 + g_2^2}}. \quad (3.7)$$

Nyní zesilovací faktor $\mathbf{g}_{\text{vážený}}$ splňuje podmínku (3.1).

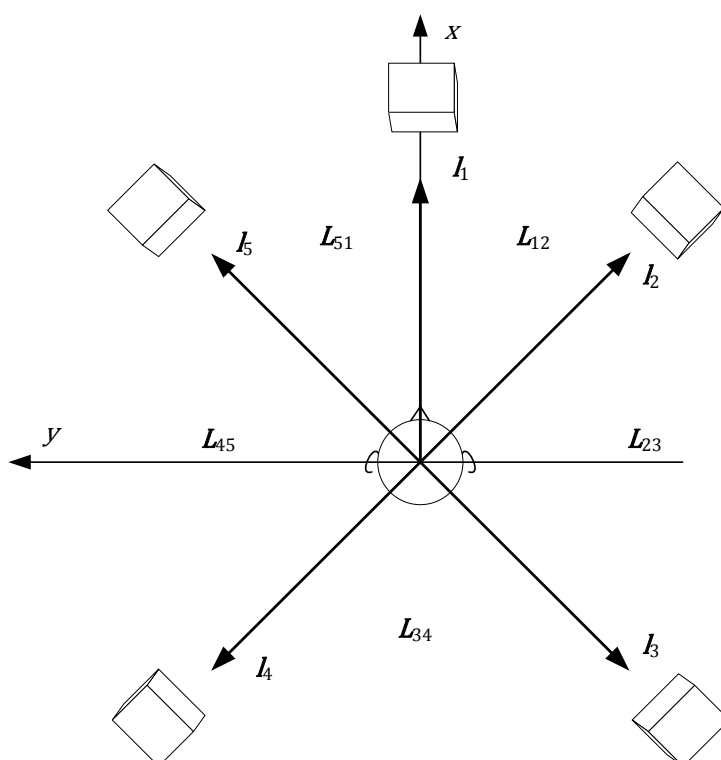


Obr. 3.2: Vektorový popis stereofonní konfigurace.

3.3.2 Dvojmrozměrné vektorově bázové amplitudové polohování pro dva a více reproduktorů

Metoda vektorově bázového amplitudového polohování může být použita pro téměř jakoukoli konfiguraci reproduktorů v horizontální rovině, protože takový systém může být také formulován vektorovými bázemi. Z celého systému jsou vždy vybrány dva reproduktory a signál jde jen do těchto dvou reproduktorů. Každý reproduktor pak patří do dvou párů.

Na obr. 3.3 je znázorněna aplikace dvojmrozměrného vektorově bázového amplitudového polohování na reproduktorový systém s pěti reproduktory. Virtuální zdroj zvuku může být vytvořen na spojnici mezi těmi reproduktory, na jejichž vektorové bázi se nachází. Zvukové pole, které může být vytvořeno metodou VBAP, je tedy spojením aktivních oblouků jednotlivých vektorových bází. U dvojmrozměrného VBAP je nejvhodnější utvořit vektorové báze přilehlými reproduktory. Na obr. 3.3 jsou vektorové báze reproduktorů zvoleny jako L_{12} , L_{23} , L_{34} , L_{45} a L_{51} . Jednotlivé vektorové báze se nepřekrývají. Při přechodu z jedné vektorové báze do druhé se zesilovací faktor vzdálenějšího reproduktoru snižuje na nulu, což umožňuje vytváření plynule navazujících pohyblivých virtuálních zdrojů zvuku. Tím, že se pro každý virtuální zdroj využívají



Obr. 3.3: Dvojdímenzionální VBAP s pěti reproduktory. Vektory l_n ukazují na reproduktory; reproduktorové vektorové báze L_{mn} jsou formovány přilehlými reproduktory.

pouze dva reproduktory a ostatní zůstávají nepoužité, vzniká jistá redundance, která má za následek zlepšení lokalizace. [3]

3.4 Ambisonie

Ambisonie je systém vyvinutý zejména Michaellem Gerzonem [9] a je založen na dekompozici sférických harmonických funkcí zvukového pole. Hlavním rozdílem, kterým se ambisonie liší od ostatních technik reprodukce prostorového zvuku, je separace počtu kanálů potřebných pro přenos a počtu reproduktorů potřebných pro reprodukci.

Typicky se používá větší množství reproduktorů než přenosových kanálů. Čím více reproduktorů, tím stabilnější je výsledné zvukové pole. Reprodukory mohou být uspořádány do různých formací, mezi nejpoužívanější patří regulární polygony, např. krychle. Neregulární rozmístění reproduktorů, například dle standartu ITU-R BS.775, je také běžně používáno, ale neexistuje pro ně analytické řešení dekodovací rovnice. Je tak třeba hledat řešení numerickými metodami [4].

Gerzon vypracoval meta-teorii nazvanou *směrová psychoakustika*, která obsahuje několik teorií lokalizace ve zvukovém poli.

Základními předpoklady směrové psychoakustiky jsou [7]:

- Na nízkých kmitočtech (asi do 500 Hz) mají signály přicházející do uší stabilní fázový rozdíl, který je menší než polovina vlnové délky. Tato fázová synchronizace znamená, že můžeme provést vektorový součet příspěvků ze všech reproduktorů k získání tzv. *zesílení vektoru rychlosti*

$$\mathbf{u} = \sum_i g_i \mathbf{l}_i, \quad (3.8)$$

což je představitel vnímaného směru v nízkých kmitočtech. V rovnici (3.8) je g_i zesilovací faktor i -tého reproduktoru a $\mathbf{l}_i$ je jednotkový vektor směřující od reproduktoru k posluchači.

- Na vysokých kmitočtech (asi od 500 Hz do asi 3500 Hz) je se signály přicházejícími do uší zacházeno jako s nekoherentními a sluchová soustava je citlivá převážně na *zesílení vektoru energie*

$$\mathbf{e} = \sum_i |g_i|^2 \mathbf{l}_i. \quad (3.9)$$

Těchto předpokladů poskytnutých směrovou psychoakustikou je využito pro vylepšení kvality vícekanálové reprodukce a je jich využíváno pro posouzení, je-li dekodér ambisonický. [9]

3.4.1 Matematické základy

Ambisonie je založena na sférické harmonické dekompozici zvukového pole, která vychází z vlnové rovnice převedené do sférické soustavy souřadnic, kde je bod $\vec{r}$ popsán poloměrem r , azimutem θ a elevací α . Tlakové pole pak může být vyjádřeno Fourier-Besselovou řadou, jejíž koeficienty jsou váhované násobky směrových funkcí $Y_{mn}^\sigma(\theta, \alpha)$ nazývaných *sférické harmonické funkce* a radiálních funkcí

$$\begin{aligned} p(\vec{r}) = & \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{0 \leq n \leq m, \sigma \pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \alpha) + \\ & + \sum_{m=0}^{\infty} j^m h_m^-(kr) \sum_{0 \leq n \leq m, \sigma \pm 1} A_{mn}^\sigma Y_{mn}^\sigma(\theta, \alpha) \end{aligned} \quad (3.10)$$

s vlnovým číslem k . [14]

Rovnice (3.10) je obecná rovnice pro případ kulové vrstvy, ve které se nevyskytuje žádný zdroj zvuku. Váhovací koeficienty B_{mn}^σ spojené se sférickými Besselovými funkcemi $j_m(kr)$ popisují průchozí pole, zatímco koeficienty A_{mn}^σ spojené se sférickými Henkelovými funkcemi $h_m^-(kr) = j_m(kr) - j n_m(kr)$, popisují odchozí pole (způsobené zdroji uvnitř koule). [14]

Ambisonie předpokládá centrovaný poslechový prostor, který neobsahuje virtuální zdroje zvuku. Proto se bere v úvahu pouze průchozí pole reprezentované koeficienty B_{mn}^σ , odchozí pole se pokládá rovno nule ($A_{mn}^\sigma = 0$). Koeficientům B_{mn}^σ ve frekvenční oblasti se říká *ambisonické signály*. [14]

Sférické harmonické komponenty B_{mn}^σ jsou úzce spojeny s tlakovým polem a jeho deriváty. První čtyři komponenty jsou dobře známy, $B_{00}^{+1} = W$ souvisí s tlakem, $B_{11}^{+1} = X$, $B_{11}^{-1} = Y$ a $B_{10}^{+1} = Z$ s gradientem tlaku, nebo také s akustickou rychlostí. Každá další skupina komponentů vyšších řádů poskytuje aproximaci zvukového pole pro větší okolí. [14]

V praxi, pouze omezený počet komponentů lze přesnět a využít. Proto se reprezentované pole typicky aproximuje zkrácenou řadou M prvků

$$p(\vec{r}) = \sum_{m=0}^M j^m j_m(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \alpha), \quad (3.11)$$

která obsahuje $K_{3D} = (M + 1)^2$ komponentů.

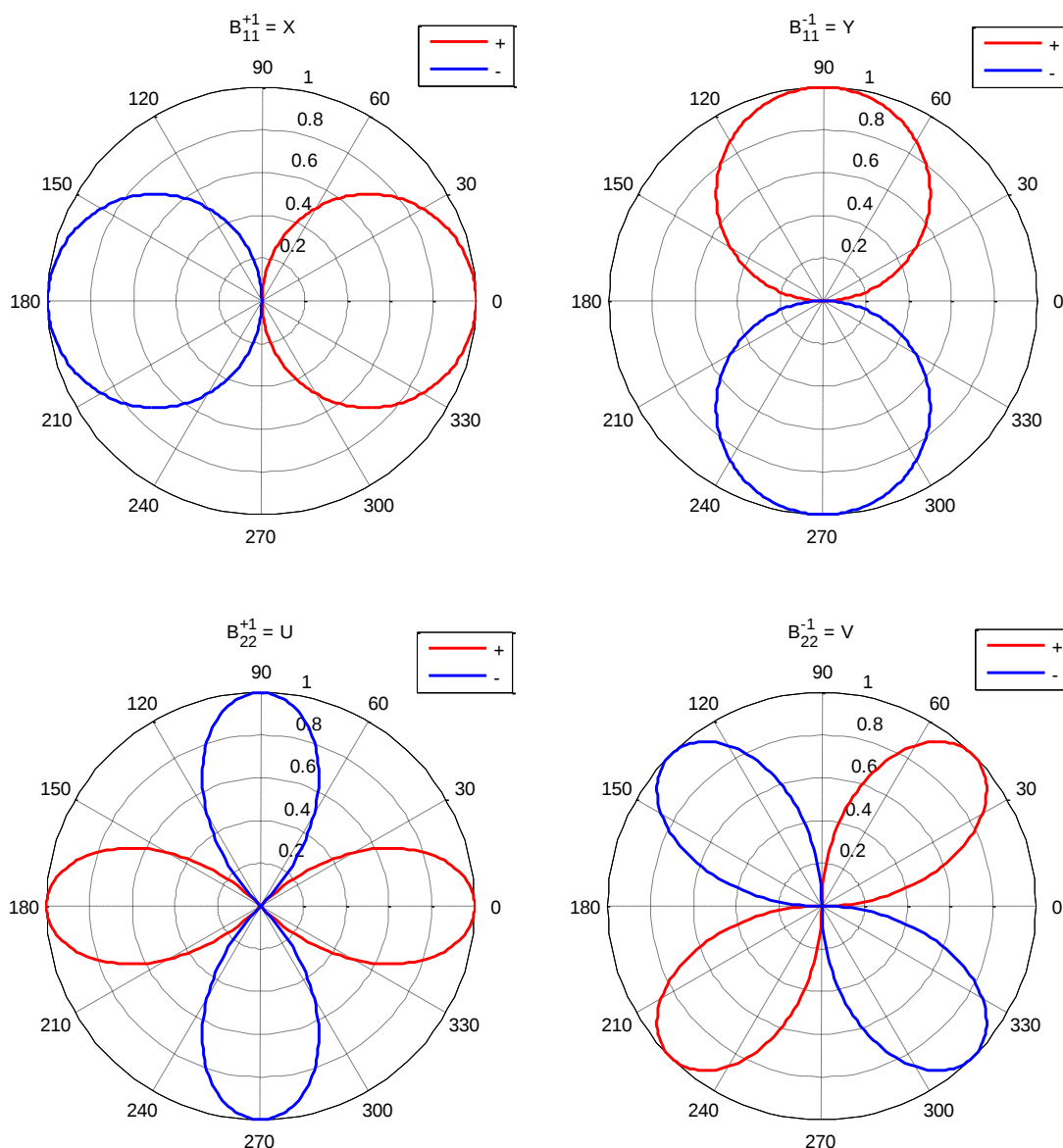
3.4.2 Ambisonické kódování

Kódování virtuálního zdroje do ambisonických signálů je založeno na předpokladu, že zdroj zvuku je dostatečně vzdálený, v ideálním případě v nekonečnu, aby jeho příspěvek byl aproximovatelný rovinnou vlnou [15].

Sférická harmonická dekompozice rovinné vlny přicházející ze směru (θ_S, α_S) , nesoucí signál S vede k jednoduchému tvaru vyjadřujícímu ambisonické komponenty [15]

$$B_{mn}^{\sigma} = S \cdot Y_{mn}^{\sigma}(\theta_S, \alpha_S), \quad (3.12)$$

kde B_{mn}^{σ} vyjadřuje ambisonické komponenty řádu m , stupně n s $\sigma = \pm 1$, kde komponenty s $\sigma = -1$ jsou otočeny o $90^\circ/m$ kolem osy z oproti komponentům s $\sigma = 1$. Signál S ze vzdáleného pole je tedy kódován aplikací reálných kódovacích koeficientů, které jsou vyjádřeny sférickými harmonickými funkcemi. Reálnost kódovacích koeficientů má význam v tom, že vlastnosti ambisonických komponentů jsou neměnné s kmitočtem [15].



Obr. 3.4: Horizontální ambisonické komponenty prvního (nahore) a druhého řádu (dole), kde $\pm$ značí fázi.

Rovnice ambisonického kódování ve tvaru (3.12) platí pro obecné trojrozměrné zvukové pole. Ne vždy je potřeba záznamu, nebo reprezentace trojrozměrného pole, proto lze definovat zjednodušený vztah pro dvojrozměrné horizontální zvukové pole.

Prvním zjednodušením je eliminace sférických harmonických složek, které jsou definovány podél osy z (vertikální osa). Použijí se tedy pouze ty sférické harmonické funkce Y_{mn}^σ , u kterých je $m = n$. Tyto funkce pro první a druhý řád jsou zobrazeny na obr. 3.5, kde je dobře patrná rotace u složek s $\sigma = -1$. Se stoupajícím řádem sférické harmonické funkce se zvyšuje její úhlové rozlišení.

Sférické harmonické funkce bez vertikální informace lze zjednodušit na cylindrické harmonické funkce, které jsou popsány v kartézském systému souřadnic [11]

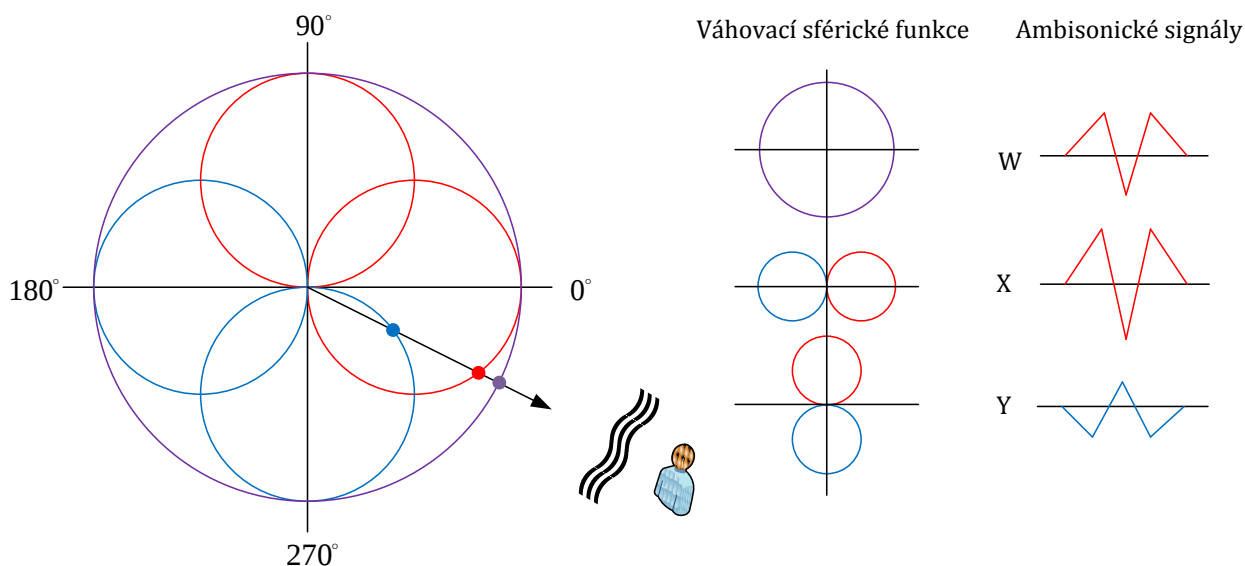
$$\begin{aligned}x &= \rho \cos \theta, \\y &= \rho \sin \theta, \\z &= z,\end{aligned}\tag{3.13}$$

kde x a y jsou pozice bodu v kartézském systému souřadnic, θ je azimut v cylindrickém systému souřadnic. Protože vertikální souřadnice z ani poloměr válce ρ nenesou pro horizontální ambisonii žádnou informaci, lze dosadit za $z = 0$ a $\rho = 1$.

Z (3.13) lze již odvodit zjednodušenou rovnici horizontálního ambisonického kódování prvního řádu [10]

$$\begin{bmatrix} W \\ X \\ Y \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ \cos \theta \\ \sin \theta \end{bmatrix} S,\tag{3.14}$$

kde W , X a Y jsou ambisonické signály, θ je azimut a S je vstupní signál. W není rovno jedné, jak by se dalo očekávat z definice sférických harmonických funkcí, ale $1/\sqrt{2}$. Je to dáno Furse-Malhamovou notací a je standartní pro ambisonii prvního řádu [19].



Obr. 3.5: Kódování rovinné vlny přicházející ze směru šipky do ambisonických signálů prvního řádu.

Obecně lze vyjádřit vztah pro dvojrozměrné ambisonické kódování řádu M jako dekompozici rozložení delta do cirkulárních Fourierových koeficientů [11]

$$\mathbf{B}(\theta) = S \left[\frac{1}{\sqrt{2}}, \cos(\theta), \sin(\theta), \cos(2\theta), \sin(2\theta), \dots, \cos(M\theta), \sin(M\theta) \right]^T, \quad (3.15)$$

kde $\mathbf{B}(\theta) = [B_{00}^{+1}(\theta), B_{11}^{+1}(\theta), B_{11}^{-1}(\theta), \dots, B_{MM}^{\sigma}(\theta)]^T$ jsou horizontální ambisonické komponenty a θ je směr, ze kterého rovinná vlna přichází.

U trojrozměrného ambisonického kódování prvního řádu lze kódovací rovnici získat převodem ze sférické soustavy souřadnic do kartézské soustavy souřadnic pomocí rovnice [11]

$$\begin{bmatrix} W \\ X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ \cos \theta \sin \alpha \\ \sin \theta \sin \alpha \\ \cos \alpha \end{bmatrix} S, \quad (3.16)$$

kde W, X, Y a Z jsou ambisonické signály, θ je azimut virtuálního zdroje a α elevace virtuálního zdroje.

3.4.3 Ambisonické dekódování

Ambisonické dekódování je postup, který vhodnou lineární kombinací ambisonických signálů ($W, X, Y, Z, \dots$) pro každý reproduktor v systému reprodukuje tlak a částicovou rychlost ve středové pozici systému. Takové lineární kombinaci se říká *exaktní dekódovací matice*, v literatuře také často *rychlostní dekódování*. Zařízení, které uplatňuje tento postup se nazývá *ambisonický dekodér*.

V kapitole 3.4.1 byly objasněny podmínky, které musí být splněny, aby bylo možné dekodér prohlásit za ambisonický. Tyto podmínky mají přímý vztah ke kvalitě lokalizace reprodukováného zdroje zvuku.

Každá reproduktorová konfigurace má unikátní dekódovací matici. Obecně lze rozlišit tři typy reproduktorových konfigurací [10]:

1. regulární polygony – např. čtverec, šestiúhelník, kostka, dvanáctiúhelník,

2. neregulární polygony s protilehlými páry – např. obdélník s poměrem stran $\sqrt{2}:1$, $\sqrt{3}:1$,
3. obecné neregulární polygony – např. ITU-R BS.775.

Pro první a druhý typ reproduktorových konfigurací lze odvodit exaktní řešení, pro třetí typ exaktní řešení neexistuje a je třeba jej získat numerickými metodami [10]. Ve všech případech musí počet reproduktorů přesáhnout počet ambisonických kanálů.

Základním cílem ambisonického dekodéru je znovu sestavit zvukové pole na pozici centrálního posluchače ambisonickými komponenty, které zvukové pole reprezentují. Tomuto principu se říká *princip znovu zakódování*. Za předpokladu, že jsou reproduktory dostatečně vzdáleny od poslechového bodu, jejich signály S_i jsou kódovány jako rovinné vlny s vektorem koeficientů $\mathbf{c}_i$ [14]:

$$\mathbf{c}_i = \begin{bmatrix} Y_{00}^{+1}(\theta_i, \alpha_i) \\ Y_{11}^{+1}(\theta_i, \alpha_i) \\ Y_{11}^{-1}(\theta_i, \alpha_i) \\ \dots \\ Y_{mn}^{\sigma}(\theta_i, \alpha_i) \end{bmatrix}, \quad \tilde{\mathbf{B}} = \begin{bmatrix} \tilde{B}_{00}^{+1} \\ \tilde{B}_{11}^{+1} \\ \tilde{B}_{11}^{-1} \\ \dots \\ \tilde{B}_{mn}^{\sigma} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} S_1 \\ \dots \\ S_N \end{bmatrix}, \quad (3.17)$$

kde $\mathbf{S}$ je vektor signálů vysílaných reproduktory a $\tilde{\mathbf{B}}$ je vektor ambisonických komponentů, které jsou znovu skládány (principem znovu zakódování).

Princip znovu zakódování lze napsat v maticové formě jako [14]

$$\tilde{\mathbf{B}} = \mathbf{C} \cdot \mathbf{S}, \quad (3.18)$$

kde $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_N]$ je znovu kódovací matice. Význam rovnice (3.18) je, že vysílané signály S_i přispívají k rekonstrukci ambisonických komponentů v poslechové bodě. Dekódovací operace tedy míří k získání signálů $\mathbf{S}$ z původních ambisonických signálů $\mathbf{B}$ [14]:

$$\mathbf{S} = \mathbf{D} \cdot \mathbf{B}. \quad (3.19)$$

Aby bylo možné zajistit $\tilde{\mathbf{B}} = \mathbf{B}$, systém podle rovnice (3.19) je potřeba invertovat. Dekódovací matice $\mathbf{D}$ je tedy typicky definována jako [14]

$$\mathbf{D} = \mathbf{C}^+ = \mathbf{C}^T \cdot (\mathbf{C} \cdot \mathbf{C}^T)^{-1}, \quad (3.20)$$

kde $+$ označuje pseudoinverzi. Rovnice (3.20) platí pouze v případě, že reproduktorů je stejně nebo více než ambisonických signálů, které jsou dekodovány. Pro regulární konfigurace reproduktorů může být vyjádření dekodovací matice $\mathbf{D}$ zjednodušeno jak pro trojrozměrnou, tak i pro dvojrozměrnou reprodukci [14]

$$\mathbf{D} = \frac{1}{N} \mathbf{C}^T, \quad (3.21)$$

kde N je počet reproduktorů dané reproduktorové konfigurace. Dekodovací matici ve tvaru rovnice (3.20) nebo (3.21) se říká *exaktní (rychlostní) dekodování*. Vyznačuje se tím, že reprodukuje vektor r_V roven jedné po všech azimutech [10].

3.4.4 Psychoakustická optimalizace ambisonického dekodéru

Psychoakustická optimalizace vychází z Gerzonovy metapie sluchové lokalizace [16], ve které prohlásil, že člověk používá různé mechanismy pro sluchovou lokalizaci a že kromě případů, kdy jsou jednotlivé sluchové podněty silně v konfliktu (podávají různé informace o směrovosti zvuku), je celkový sluchový dojem dán většinou souhlasných podnětů.

Gerzon [16] dále popisuje hierarchii modelů, z nichž pro každý derivuje lokalizační vektor, jehož směr určuje předpokládaný směr zvuku a jeho velikost popisuje stabilitu lokalizace. Pro reálný jednobodový zdroj je velikost lokalizačního vektoru 1. Pokud je velikost lokalizačního vektoru menší nebo větší, vnímaný směr se pohybuje, pokud posluchač otočí hlavou.

Dva nejjednodušší a možná i nejdůležitější modely jsou model *akustické částicové rychlosti*, který koresponduje s Makitovým modelem [17] a model *proudění energie*, který koresponduje s De Boerovým modelem [18]. Gerzon poznamenává, že prakticky všechny modely sluchové lokalizace, kromě boltcového zabarvení (spektrální) a impulzního interaurálního časového rozdílu, jsou speciální případy těchto dvou modelů. Tyto modely se často vyskytují v literatuře zabývající se ambisonií pod názvem *model rychlosti* a *model energie* s přidruženými lokalizačními vektory nazývanými *vektor rychlosti* a *vektor energie*. Tyto vektory korelují s měřeními interaurálního fázového rozdílu a interaurálního rozdílu intenzity.

Aplikací těchto psychoakustických modelů na návrh systémů pro prostorovou reprodukci Gerzon uvedl, že dekodér, nebo reprodukční systém pro obklopující zvuk je definován jako ambisonický, pokud je navržen tak, aby [16]

- směr rychlostního vektoru a vektoru energie byl souhlasný alespoň do kmitočtu okolo 4 kHz, s tím, že se reprodukováný azimut $\theta_V = \theta_E$ výrazně nemění s kmitočtem,
- na nízkých kmitočtech okolo 400 Hz byla velikost rychlostního vektoru blízko jedné pro všechny reprodukováné azimuty,
- na středních a vysokých kmitočtech, okolo 700 Hz až 4 kHz, byla velikost vektoru energie r_E maximalizována na co největší části poslechového prostoru.

Dle Gerzona [13, 16] nastává nejlepší možná lokalizace pro danou reproduktorovou konfiguraci, pokud je velikost rychlostního lokalizačního vektoru na nízkých kmitočtech rovna jedné a pokud je velikost energiového lokalizačního vektoru maximalizována na středních kmitočtech s tím, že přechod mezi těmito dvěma pásmy je mezi 300 Hz a 700 Hz. Tato teorie se zakládá na předpokladu, že pokud je rychlostní lokalizační vektor a energiový lokalizační vektor reprodukováného zvukového zdroje shodný s vektorem rychlosti a vektorem energie reálného zdroje, jejich vjem je stejný. Reprodukovaný zvukový zdroj bude tedy znít jako reálný.

3.4.5 Vektor energie a rychlosti

Vektor rychlosti ve středu reproduktorové soustavy je vypočítán jako suma příspěvků vektorů rychlosti každého z reproduktorů. Protože jsou ambisonické dekodovací rovnice odvozeny tak, aby rekonstruovaly rychlostní komponenty přesně, výsledkem je to, že vektor rychlosti má vždy jednotkovou velikost ve směru předpokládaného zdroje. Dle Gerzona je velikost vektoru rychlosti r_V pro střed reproduktorové soustavy vypočítána jako [9]

$$r_V \hat{r} = \text{Re} \left\{ \frac{\sum_{i=1}^n G_i \hat{u}_i}{\sum_{i=1}^n G_i} \right\}, \quad (3.22)$$

zatímco velikost vektoru energie je vypočítána r_E jako [9]

$$r_E \hat{r} = \frac{\sum_{i=1}^n |G_i|^2 \hat{u}_i}{\sum_{i=1}^n |G_i|^2}, \quad (3.23)$$

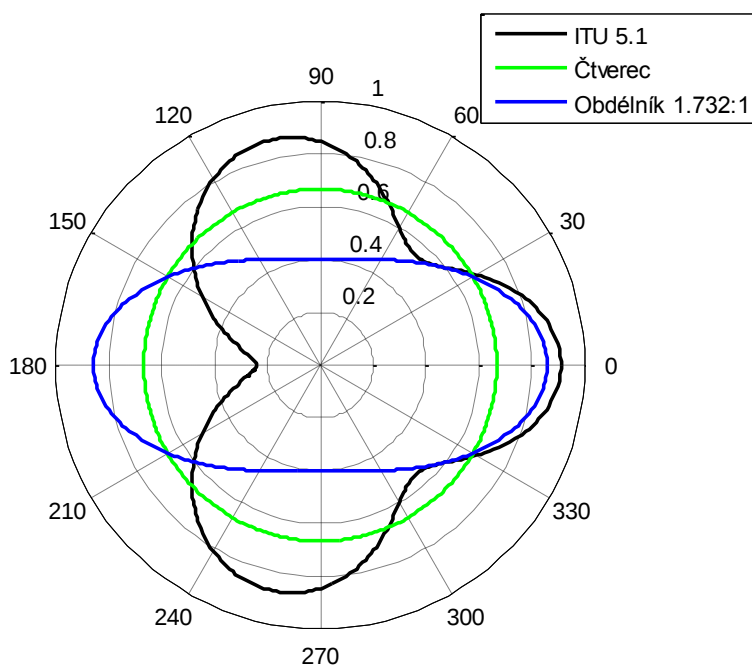
kde G_i je zesílení (v některých případech komplexní) ze zdroje do i . reproduktoru a $\hat{u}$ je jednotkový vektor směřující na i . reproduktor.

Na obrázku 3.6 je zobrazena velikost vektoru energie r_E pro exaktní (rychlostní) dekodér pro konfiguraci ITU-R BS.775, která má reproduktory na azimutech 0° , $\pm 30^\circ$ a $\pm 110^\circ$, čtvercovou konfiguraci, která má reproduktory na azimutech $\pm 45^\circ$ a $\pm 135^\circ$ a obdélníkovou konfiguraci s poměrem stran $\sqrt{3}:1$ s reproduktory na azimutech $\pm 30^\circ$ a $\pm 150^\circ$. [9]

Pro čtvercový dekodér má velikost r_E konstantní hodnotu pro všechny azimuty, stejně jako bude mít pro jakoukoliv jinou regulární konfiguraci. To mimo jiné znamená, že kvalita lokalizace bude konstantní ve všech směrech. U obdélníkové konfigurace má r_E větší velikost vpředu a vzadu, zatímco po bocích dosahuje jen malých hodnot. To má za následek lepší lokalizaci vpředu a vzadu, ale horší lokalizaci po stranách. Konfigurace ITU-R BS.775 poskytuje dobrou lokalizaci vpředu a po bocích ve směru zadních reproduktorů. Lokalizace vzadu bude však poněkud nekvalitní.

3.4.6 Optimalizovaný dekodér

V kapitole 3.4.5 byly uvedeny kritéria pro psychoakustickou optimalizaci a byly zavedeny



Obr. 3.6: Velikost vektoru energie r_E jako funkce azimutu virtuálního zdroje pro různé reproduktorové konfigurace; vektor rychlosti r_V je roven jedné.

lokalizační vektory r_E a r_V , které určují kvalitu reprodukce směrové informace v daném dekodéru. Protože v kapitole 2 byl popsán exaktní (rychlostní) dekodér, který reprodukuje velikost vektoru r_V rovnou jedné pro všechny azimuty, tato kapitola se bude zabývat přeměnou tohoto řešení, vhodného pro nízké kmitočty, pro splnění dalších psychoakustických kritérií, jmenovitě energiový dekodér $\max r_E$ a dekodér „ve fázi“. Jediným způsobem, jak provést tuto přeměnu, aniž by se změnil směr vektoru rychlosti r_V , je upravit poměr mezi tlakem (W) a rychlostí ($X, Y, Z \dots$) [10].

Dekodér $\max r_E$ soustřeďuje příspěvky jednotlivých reproduktorů na vyšších kmitočtech směrem k azimutu virtuálního zdroje [22]. Pro dekodér $\max r_E$ lze velikost vektoru r_E vyjádřit jako funkci poměru tlaku a rychlosti k , která má za předpokladu, že jde o horizontální dekodér prvního řádu s regulární topologií tvar [10]

$$r_E(k) = \frac{2k}{2k^2 + 1}, \quad (3.24)$$

který dostahuje maxima v hodnotě $k = \sqrt{2}/2 \approx -3,01$ dB [10]. Nyní je potřeba zajistit takové zesílení tlaku a snížení rychlosti, aby celková hlasitost mezi nízkými kmitočty a vysokými kmitočty zůstala nezměněná. Jedním ze způsobů je zachování střední kvadratické hodnoty [10]

$$W^2 + X^2 + Y^2 = 3 \quad (3.25)$$

na nízkých i vysokých kmitočtech a

$$\frac{X}{W} = \frac{Y}{W} = \frac{\sqrt{2}}{2} \quad (3.26)$$

na vysokých kmitočtech. Řešením rovnic (3.25) a (3.26) lze získat zesilovací faktory pro $\max r_E$ dekodování vysokých kmitočtů [10]

$$W = \sqrt{\frac{3}{2}} \approx 1,76 \text{ dB}, \quad (3.27)$$

$$X = Y = \sqrt{\frac{3}{4}} \approx -1,25 \text{ dB}. \quad (3.28)$$

Rovnice (3.27) a (3.28) platí pro horizontální ambisonický dekodér prvního řádu. Tyto zesilovací faktory by měly být aplikovány před dekódováním ambisonických signálů, nejlépe pomocí filtrů typu shelving [8], jejichž návrhu bude věnována další kapitola.

V [12] jsou odvozeny zobecněné vztahy pro ambisonický dekodér vyšších řádů. Jsou ve formě váhovacích koeficientů, kterými se váhují ambisonické signály. V tab. 3.1 jsou uvedeny vztahy pro výpočet těchto vah. M značí nejvyšší řád ambisonických komponentů použitých při dekódování a m řád aktuální dekódované ambisonické komponenty.

Tab 3.1: Vztahy pro výpočet vah pro různé druhy ambisonických dekodérů

dekodér	základní	$\max r_E$	ve fázi
váhy $w(m)$	1	$\cos\left(\frac{m\pi}{2M+2}\right)$	$\frac{M!^2}{(M+m)!(M-m)!}$

Použitím váhovacích koeficientů v tab. 3.1 lze sestavit dekódovací rovnici pro uvedené typy dekodérů [14]

$$\mathbf{D} = \frac{1}{N} \mathbf{C}^T \text{diag}(\mathbf{w}), \quad (3.29)$$

kde $\mathbf{w} = [w(0), w(1), \dots, w(M)]^T$.

Dekodér „ve fázi“ dekóduje ambisonické komponenty se směrovou charakteristikou ve tvaru kardioidu. Tento druh dekódování byl označen za nejlepší pro velké prostory [19]. Vyznačuje se tím, že jeho směrová charakteristika neobsahuje části se zápornou polaritou. Maximalizuje se tak poslechový prostor a zároveň by tento typ dekodéru mohl splňovat psychoakustické kritérium pro kmitočty nad 5 kHz [12]. Dosud však nebyly provedeny experimenty, které by tuto hypotézu potvrzovaly.

Jiným způsobem, definovaným v [4], [23] lze z ambisonie prvního řádu získat jakoukoli mikrofónovou odezvu prvního řádu (od všesměrové po tvar osmičky) v jakémkoli směru. To je možné použitím rovnice [4], [23]

$$\begin{aligned} g_w &= \sqrt{2}, \\ g_x &= \cos(\theta) \cos(\alpha), \\ g_y &= \sin(\theta) \cos(\alpha), \\ g_z &= \sin(\alpha), \\ S &= 0,5[(2-d)g_w W + d(g_x X + g_y Y + g_z Z)], \end{aligned} \quad (3.30)$$

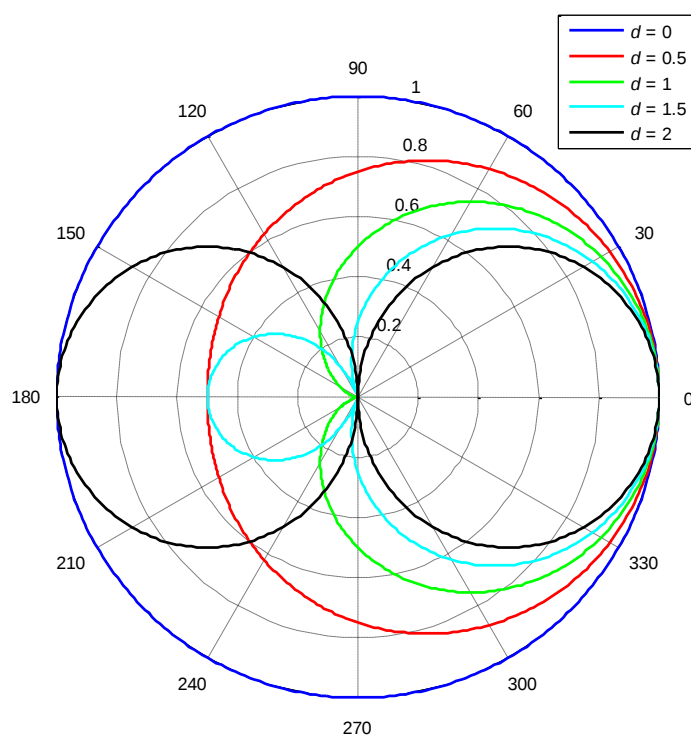
kde W, X, Y a Z jsou ambisonické signály prvního řádu, S je výstup reproduktoru, θ je azimut reproduktoru, α je elevace reproduktoru a $d \in (0; 2)$ je činitel směrovosti.

Činitel směrovosti d , jehož vliv na směrovou charakteristiku je vidět na obr. 3.7, může být zvolen na základě osobní preference, tedy podle nějaké formy empiricky ověřeného nastavení, nebo teoretickou kalkulací, která počítá optimální dekodovací schéma. [4]

3.4.7 Filtry pro optimalizovaný dekodér

Změnu z rychlostního ambisonického dekodování dle rovnice (3.20) na dekodování, které je optimalizováno pro nízké i vysoké kmitočty lze realizovat filtry typu shelving. Pro případ ambisonického dekodéru prvního řádu budou použity různé zesílení filtrů pro signál W a signály X a Y tak, aby se změnil relativní poměr úrovní těchto signálů, ale jejich fáze zůstala souhlasná. Pro horizontální dekodér, signál W je zesílen o 1,76 dB na vysokých kmitočtech a signály X a Y jsou zeslabeny o 1,25 dB na vyšších kmitočtech oproti rychlostnímu dekodování, jak vyplývá z rovnic (3.29) a (3.30) [9].

Pro ambisonický dekodér je důležité, aby filtry v něm použité byly ve fázi v celém



Obr. 3.7: Grafická reprezentace směrových charakteristik dosažitelných ambisonií prvního řádu.

kmitočtovém rozsahu. Tato podmínka je lehce splněna filtry s konečnou impulzní odezvou získané aproximací přenosové charakteristiky shelving filtru s nekonečnou impulzní odezvou. [9]

Filtry s konečnou impulzní odezvou jsou podstatně náročnější na výpočetní výkon, než filtry s nekonečnou impulzní odezvou. V [10, 20] jsou popsány metody návrhu filtrů s nekonečnou impulzní odezvou. Nejzajímavější z nich jsou fázově kompenzované filtry, které při součtu dvou filtrem oddělených pásem mají plochou modulovou kmitočtovou charakteristiku. Jejich přenosová funkce je dána rovnicí

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{a_0 + a_1 z^{-1} + a_2 z^{-2}}, \quad (3.31)$$

kde jsou koeficienty b_n přenosové funkce pro filtr typu dolní propust dány jako

$$\begin{aligned} b_0 &= \frac{k^2}{k^2 + 2k + 1}, \\ b_1 &= 2b_0, \\ b_2 &= b_0, \end{aligned} \quad (3.32)$$

koeficienty b_n filtru typu horní propust jsou dány jako

$$\begin{aligned} b_0 &= \frac{1}{k^2 + 2k + 1}, \\ b_1 &= -2b_0, \\ b_2 &= b_0, \end{aligned} \quad (3.33)$$

a nulové body a_m pro oba typy filtrů jsou dány jako

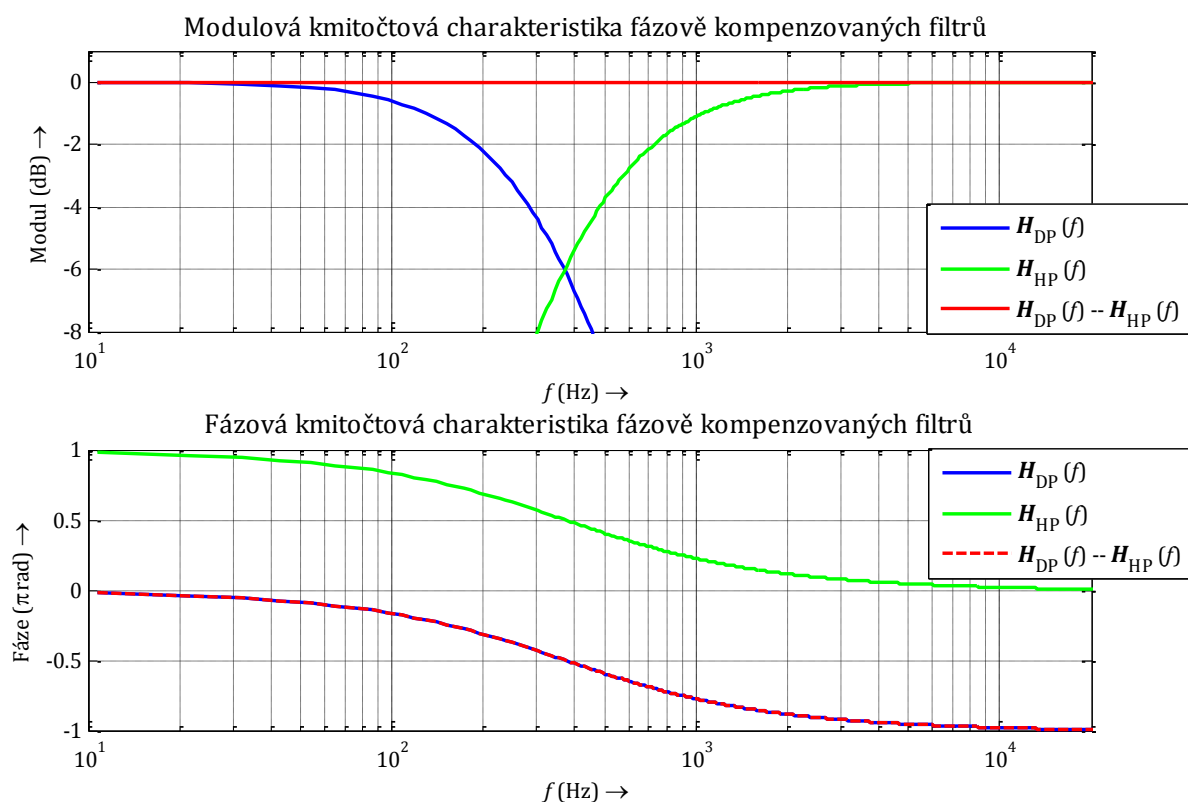
$$\begin{aligned} a_0 &= 1, \\ a_1 &= \frac{2(k^2 - 1)}{k^2 + 2k + 1}, \\ a_2 &= \frac{k^2 - 2k + 1}{k^2 + 2k + 1}, \end{aligned} \quad (3.34)$$

kde k je dáno vztahem

$$k = \tan \frac{\pi f_c}{f_{vz}}, \quad (3.35)$$

kde f_c je mezní kmitočet filtru a f_{vz} vzorkovací kmitočet.

Na obr. 3.8 je zobrazena charakteristika fázově kompenzovaných filtrů pro mezní kmitočet $f_c = 370$ Hz. Při součinu filtrovaných signálů dolní propustí H_{DP} a horní propustí H_{HP} je třeba obrátit fázi jednoho ze signálů. Výsledná modulová kmitočtová charakteristika je plochá a fázová kmitočtová charakteristika je totožná s fázovou kmitočtovou charakteristikou dolní propusti.



Obr. 3.8: Charakteristika fázově kompenzovaných filtrů pro $f_c = 370$ Hz.

4 NÁVRH A IMPLEMENTACE ALGORITMU

Součástí této práce je i návrh algoritmů vybraných metod v prostředí Matlab. Při jejich návrhu bylo počítáno s následující implementací v jazyce C++ jako plug-in modul technologie VST. Nebyly proto v zájmu usnadnění implementace používány funkce Matlabu, které by nebyly snadno převoditelné do jazyka C++. Při implementaci bylo třeba vyřešit několik dílčích problémů, jejichž popisem a řešením se bude tato kapitola zabývat.

4.1 Úvod

Z popsaných metod byly k implementaci vybrány metody VBAP a ambisonie. VBAP je zobecněná technika amplitudového polohování pro dvojrozměrný prostor. Jde o párovou polohovací techniku, tzn. v každém případě jsou aplikovány nenulové zesilovací faktory maximálně na dva kanály. Ambisonie je založená na mikrofonní technice, kde se zvukové pole zakóduje do ambisonických signálů, které je možno přenášet či uchovávat, a ze kterých lze následně dekodovat zpět původní směrovou zvukovou informaci.

Výhody metody VBAP spočívají zejména v tom, že je libovolně škálovatelná, poměrně jednoduchá na realizaci a přesná ve smyslu vnímané směrovosti zvuků. Mezi nevýhody však patří, že není příliš transparentní (ve výsledném zvukovém obrazu jsou příliš slyšet reproduktory, protože vždy využívá pár reproduktorů) a zatím neexistuje žádný vhodný způsob, jakým lze záznam VBAP ukládat a přenášet. [4]

Výhody ambisonie jsou zejména malý počet kanálů pro přenos zaznamenaného zvuku, nezávislost reproduktorové konfigurace na způsobu nahrávání, vysoká transparentnost, možnost maticových operací jako je zoomování a rotace kolem os. Nevýhodou je však neexistence analytického řešení pro neregulární reproduktorové konfigurace.

Algoritmy, stejně jako metody, nejsou závislé na reproduktorové konfiguraci, avšak pro znázornění jsou vztaženy k reproduktorové konfiguraci ITU-R BS.775 (umístění reproduktorů: centrální reproduktor na azimutu 0° , levý přední reproduktor na 30° , levý zadní na 110° , pravý zadní na 250° a levý přední na 330°).

4.2 Virtual Studio Technology

Virtual Studio Technology je systém pro zpracování zvukových signálů na osobním počítači. Skládá se z hostitelské aplikace a z VST plug-in modulu. Z pohledu aplikace je plug-in modul černá skříňka, se kterou komunikuje řízením parametrů a které předává prostřednictvím rozhraní VST data určená ke zpracování. VST plug-in modul v podstatě zapouzdřuje kompletní proces zpracování audia. Od hostitelské aplikace dostává již připravená audio data, je tedy odstíněna od procesu řízení a získávání datových toků. Samotná audio data jsou v plovoucí řádové čárce v rozsahu ± 1 . Přes rozhraní VST jsou dále přenášeny parametry, události a MIDI data. Systém VST také umožňuje vytvořit vlastní grafické rozhraní, které je odděleno od vlastního zpracování signálů a komunikuje s ním za pomoci rozhraní VST. To umožňuje řízení efektu i při absenci vlastního uživatelského rozhraní, kdy řízení obstarává hostitelská aplikace pomocí univerzálního uživatelského rozhraní. [24]

Pro zpracování signálů v reálném čase je potřeba zajistit i odpovídající správu vstupně-výstupních audio toků. Použití běžných ovladačů zvukových karet není pro tento způsob zpracování příliš vhodný, zavádí totiž značné zpoždění. Proto firma Steinberg vyvinula systém ASIO, který spolupracuje se systémem VST. Systém ASIO je přímo navržen pro práci v reálném čase, v podstatě to je nový ovladač zvukového hardwaru, pracuje na principu double-bufferingu, umožňuje také nastavit velikost vyrovnávacích pamětí. [24]

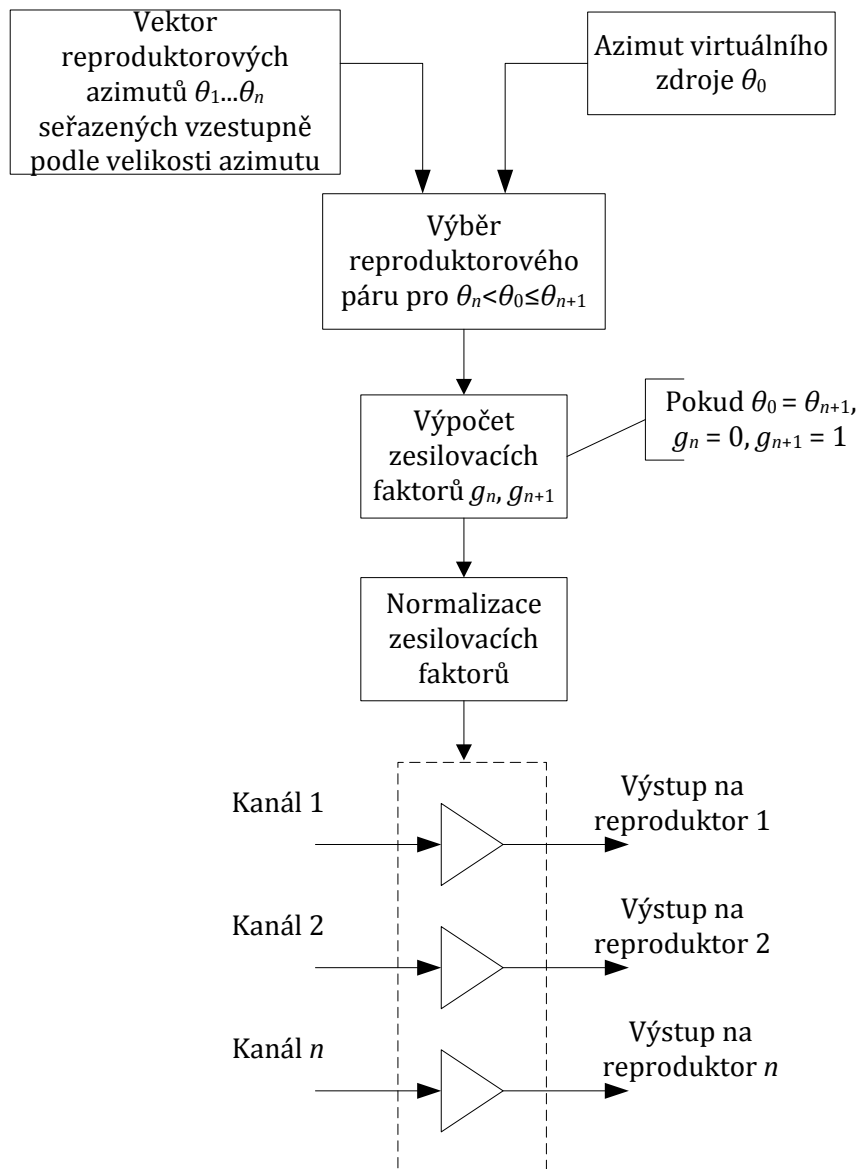
4.3 Implementace metody vektorově bázového amplitudového polohování

Implementace této metody je obecně popsána blokovým schématem (viz obr. 4.1). Nastavitelné parametry jsou:

- počet reproduktorů,
- azimut virtuálního zdroje,
- azimut jednotlivých reproduktorů.

Algoritmus se skládá z dílčích částí:

- výpočet a normalizace zesilovacích faktorů,
- třídící algoritmus.



Obr. 4.1: Blokové schéma implementace metody vektorově bazového amplitudového polohování.

Před výpočtem algoritmu je potřeba nastavit počet reproduktorů a jejich azimuty θ_n . Z těchto azimutů se celý systém rozdělí na vektorové báze $L_{n,n+1}$. Poté je zvolen azimut virtuálního zdroje θ_0 , který je umístěn do vektorové báze tak, aby $\theta_n < \theta_0 \leq \theta_{n+1}$. Následně jsou vypočítány a normalizovány zesilovací faktory pro jednotlivé kanály a směřují dále na reproduktorový výstup.

4.3.1 Třídící algoritmus

Pro výběr reproduktorového páru musí pro každý reproduktorový pár platit podmínka $\theta_n < \theta_0 \leq \theta_{n+1}$, která zaručuje, že se jednotlivé vektorové báze nebudou překrývat. Je tedy nutné, aby reproduktory byly seřazeny vzestupně podle jejich azimutu. K tomuto účelu byl implementován třídící algoritmus bubble sort, který sice má v nejhorším případě složitost $\mathcal{O}(n^2)$, ale protože je implementace omezena na dvanáct reproduktorů, algoritmus bude ukončen nejhůře po jedenácti krocích. K tomuto účelu byla použita struktura `Speaker`.

```
struct Speaker
{
    int index;      // pořadí reproduktoru
    float gain;     // zesilovací faktor
    float azimuth;  // azimut reproduktoru
};
```

Jednotlivé reproduktory mění pořadí, ve kterém jsou adresovány, nemění se však jejich `index`, který odpovídá pořadí reproduktoru. Vývojový diagram algoritmu třídění je v příloze A.1.

4.3.2 Výpočet a normalizace zesilovacích faktorů

Výpočet a normalizace zesilovacích faktorů vychází z rovnice (3.6). Tato rovnice obsahuje inverzní matici $\mathbf{L}_{12}^{-1}$, která splňuje podmínku $\mathbf{L}_{12}^{-1}\mathbf{L}_{12} = \mathbf{I}$, kde $\mathbf{I}$ je jednotková matice. Inverzní matici $\mathbf{L}_{12}^{-1}$ lze upravit použitím kofaktorové rovnice [3]

$$\begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}^{-1} = \frac{1}{l_{11}l_{22} - l_{21}l_{12}} \begin{bmatrix} l_{22} & -l_{12} \\ -l_{21} & l_{11} \end{bmatrix}. \quad (4.1)$$

Dosazením do rovnice (3.6) pak získáme

$$\mathbf{g} = \frac{1}{l_{11}l_{22} - l_{21}l_{12}} [p_1l_{22} - p_2l_{21} \quad p_2l_{11} - p_1l_{12}], \quad (4.2)$$

což při $\mathbf{g} = [g_1 \quad g_2]$ dává rovnice pro jednotlivé zesilovací faktory

$$g_1 = \frac{1}{l_{11}l_{22} - l_{21}l_{12}}(p_1l_{22} - p_2l_{21}),$$

(4.3)

$$g_2 = \frac{1}{l_{11}l_{22} - l_{21}l_{12}}(p_2l_{21} - p_1l_{12}).$$

Je vidět, že v rovnicích (4.3) se opakuje zlomek $1/(l_{11}l_{22} - l_{21}l_{12})$, ten je možno substituovat proměnnou Z a vypočítat pouze jednou. Po substituci a dosazení $l_{11} = \cos \theta_1$, $l_{12} = \sin \theta_1$, $l_{21} = \cos \theta_2$, $l_{22} = \sin \theta_2$, $p_1 = \cos \theta_0$ a $p_2 = \sin \theta_0$ dostaneme

$$Z = \frac{1}{\cos \theta_1 \sin \theta_2 - \cos \theta_2 \sin \theta_1},$$

$$g_1 = Z(\cos \theta_0 \sin \theta_2 - \sin \theta_0 \cos \theta_2),$$

(4.4)

$$g_2 = Z(\sin \theta_0 \cos \theta_1 - \cos \theta_0 \sin \theta_1),$$

kde θ_0 je azimut virtuálního zdroje, θ_1, θ_2 jsou azimuty reproduktorů a g_1 a g_2 jsou jejich zesilovací faktory. Zesilovací faktory je třeba normalizovat použitím rovnice (3.7) a dosazením $C = 1$ dostaneme

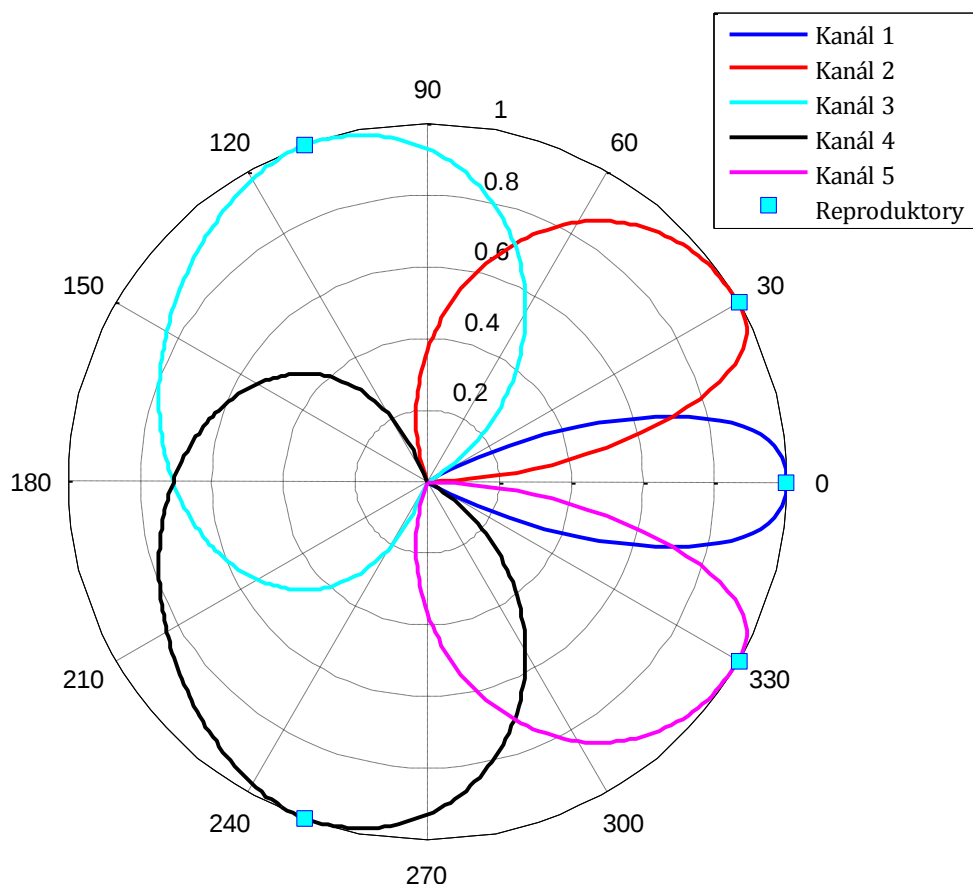
$$g_{1,\text{vážený}} = \frac{g_1}{\sqrt{g_1^2 + g_2^2}},$$

(4.5)

$$g_{2,\text{vážený}} = \frac{g_2}{\sqrt{g_1^2 + g_2^2}},$$

kde $g_{1,\text{vážený}}$ a $g_{2,\text{vážený}}$ jsou zesilovací faktory aplikované na reproduktory.

Na obr. 4.1 je pak zobrazena směrová charakteristika navrženého systému metodou VBAP, který vychází z rovnice (4.5).



Obr. 4.2: Zobrazení normalizovaných zesilovacích faktorů pro metodu VBAP.

4.4 Implementace metody ambisonie

Metoda ambisonie se principiálně skládá ze dvou základních částí, z ambisonického kodéru a ambisonického dekodéru. Ambisonický kodér zakóduje zvuk spolu s informací o požadovaném umístění v prostoru do ambisonických signálů. Tyto signály pak lze přenášet, ukládat, či dekódovat. Ambisonický dekodér dekóduje ambisonické signály tak, aby reprezentovaly co nejlépe původní zvuk i jeho umístění v prostoru pro libovolnou reproduktorovou konfiguraci.

K implementaci byl zvolen horizontální ambisonický systém prvního až třetího řádu, včetně psychoakustických optimalizací. Od řádu ambisonie se odvíjí počet kanálů nutných pro přenos ambisonických signálů. Pro počet ambisonických signálů horizontální ambisonie N platí [22]

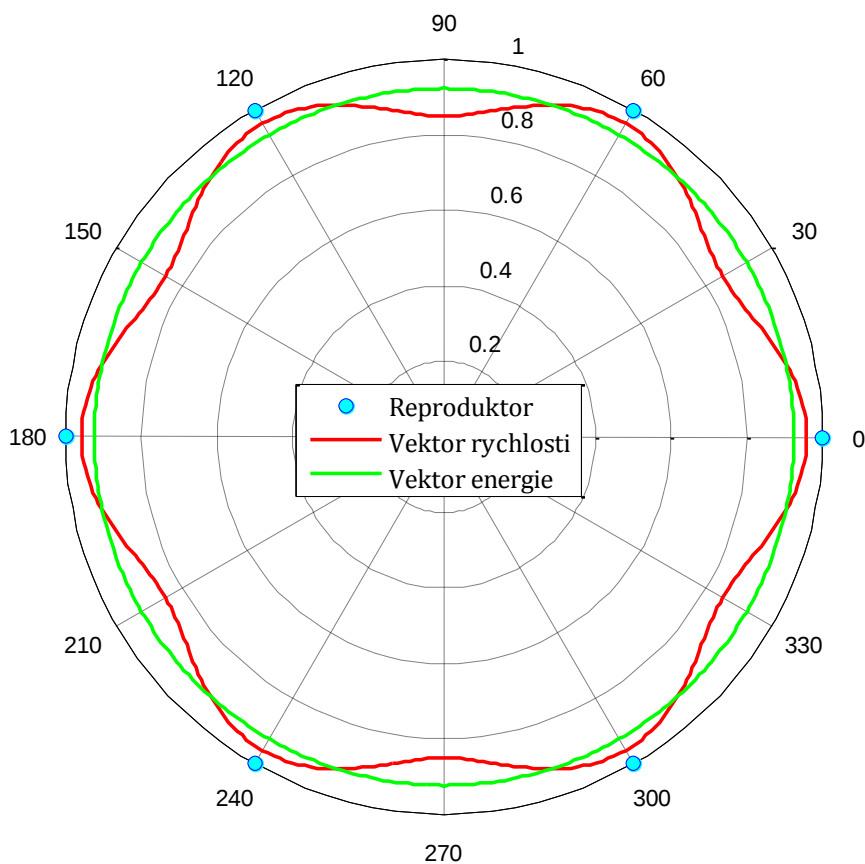
$$N = 2M + 1, \quad (4.6)$$

kde M je řád ambisonie. Pro přenos ambisonie prvního řádu jsou tedy potřeba tři, pro druhý pět a pro třetí sedm přenosových kanálů. Z toho vyplývá, že pro druhý a třetí řád ambisonie je vhodný přenos na médiích, které se v současnosti používají pro přenos šestikanálového, resp. osmikanálového zvuku.

Od řádu ambisonie se také odvíjí počet reproduktorů potřebných pro stabilní reprodukci dvojrozměrného zvukového pole, ten je dán vztahem [22]

$$S = 2M + 2, \quad (4.7)$$

kde S je počet reproduktorů a M je řád ambisonie. Pokud tato podmínka není splněna, reprodukováné zvukové pole není stabilní a dochází ke kolísání kvality lokalizace, která je vidět na obr. 4.3. Ambisonie je však hierarchální metoda a pouhým nepoužitím dvou kanálů lze snížit řád ambisonie a získat tak stabilní zvukové pole.



Obr. 4.3: Kolísání velikosti vektoru energie při nedostatečném počtu reproduktorů; Ambisonie třetího řádu, šest reproduktorů.

Implementaci lze rozdělit do několika dílčích částí:

- ambisonický kodér,
- ambisonický dekodér,
- ambisonické polohování (sériová kombinace ambisonického kodéru a dekodéru).

4.4.1 Ambisonický kodér

Implementace vychází z rovnice (3.15), která pro ambisonii prvního až třetího řádu má tvar

$$\begin{aligned} W &= \frac{1}{\sqrt{2}}, \\ X &= \cos \theta, \\ Y &= \sin \theta, \\ U &= \cos 2\theta, \\ V &= \sin 2\theta, \\ P &= \cos 3\theta, \\ Q &= \sin 3\theta, \end{aligned} \tag{4.8}$$

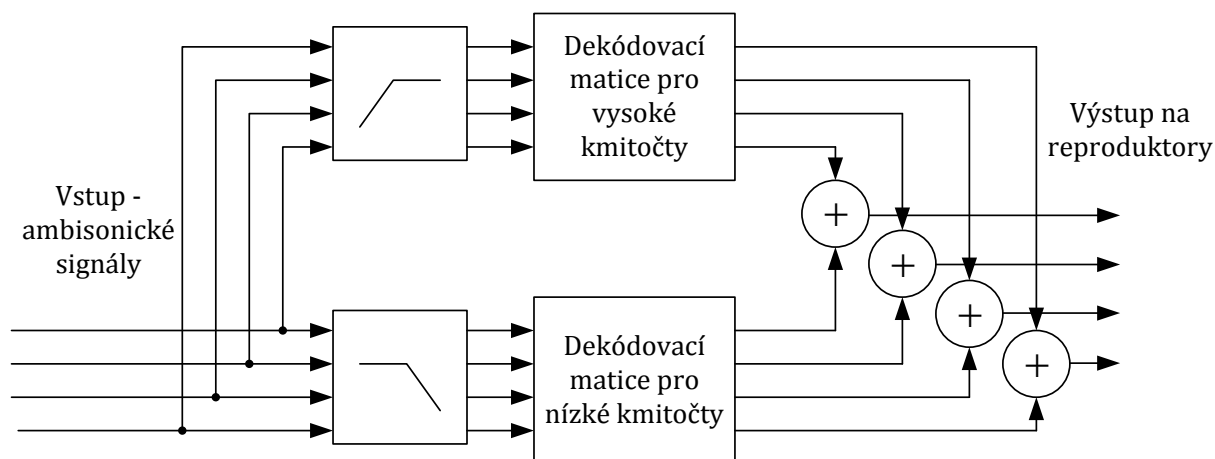
kde W , X a Y jsou ambisonické signály prvního řádu, U a V druhého řádu, P a Q třetího řádu a θ je azimut virtuálního zdroje.

U ambisonického dekodéru je tedy třeba nastavit pouze řád ambisonie a azimut virtuálního zdroje.

4.4.2 Ambisonický dekodér

Implementace ambisonického dekodéru vychází z rovnice (3.20). Jednodušší vztah (3.21) nelze použít, protože se předpokládá obecná reproduktorová konfigurace, která nemusí plnit podmínku regularity.

Blokové schéma ambisonického dekodéru s psychoakustickou optimalizací je zobrazeno na obr. 4.4. Skládá se z filtrů pro dvoupásmové zpracování a dvojice dekódovacích matic pro nízké a vysoké kmitočty. Protože návrh fázově kompenzovaných filtrů je popsán v kapitole 3.4.7, tato kapitola se bude zabývat pouze implementací výpočtu dekódovací matice a optimalizací této matice pro splnění psychoakustických kritérií popsaných v kapitole 3.4.4.



Obr. 4.4: Blokové schéma ambisonického dekodéru.

Hlavním implementačním problémem plynoucím z rovnice (3.20) je získání pseudoinverzní matice. K tomuto účelu je použita metoda dekompozice singulárních hodnot SVD, ve které je dána dekompozice obdélníkové matice A jako [25]

$$A = U \Sigma V^*, \quad (4.9)$$

kde U a V jsou unitární matice, Σ je diagonální matice singulárních hodnot matice A a V^* značí hermitovsky sdruženou matici V . Z takovéto dekompozice je možné získat inverzní matici A^+ z matice A pomocí vztahu [25]

$$A^+ = V \Sigma' U, \quad (4.10)$$

kde Σ' se získá převrácením hodnoty prvků s kladnou hodnotou na diagonále matice Σ .

Pro výpočet metody SVD byl použit balíček JAMA, který využívá knihovny TNT. Knihovna TNT obsahuje datový typ `Array2D< datový_typ >` pro počítání s maticemi. Umožňuje definici matice jako statické proměnné libovolného typu a složkám matice lze přistupovat jako k dvojrozměrnému poli.

```
Array2D< double > A; // vytvoří prázdnou matici typu double
Array2D< double > B(m,n); // vytvoří matici typu double s m řádky a n
sloupci
Array2D< double > C(m,n) = 0.0; // vytvoří matici typu double s m řádky, n
sloupci s prvky inicializovanými na hodnotu 0
```

Metodu SVD lze použít s jakýmkoliv reálným datovým typem. Výpočet SVD probíhá ihned po deklaraci.

```
SVD< double > matrix(A); // inicializace proměnnéou matrix, která je typu
SVD maticí A
matrix.getS(matrixS); // vrátí n x n matici singulárních hodnot
matrix.getU(matrixU); // vrátí m x n unitární matici
matrix.getV(matrixV); // vrátí n x n unitární matici
```

Po získání matic Σ' a V výpočtem převrácené hodnoty na diagonále, resp. Hermitovským sdružením lze získat pseudoinverzní matici A^+ funkcí `matmult`, která slouží k násobení matic.

```
inverseA = matmult(matrixV, matrixStransposed);
inverseA = matmult(inverseA, matrixUtransposed);
```

Matice `inverseA` je výslednou maticí ambisonického dekodéru.

Dále je třeba pro splnění psychoakustických kritérií váhovat dekodovací ambisonickou matici váhovacími koeficienty. Tato operace je založena na kombinaci rovnic (3.29) a (3.20)

$$\mathbf{D} = \mathbf{C}^+ \text{diag}(\mathbf{w}), \quad (4.11)$$

kde $\mathbf{D}$ je dekodovací matice, $\mathbf{C}^+$ je pseudoinvertovaný vektor ambisonických koeficientů vycházející z rovnice (3.18). $\mathbf{C}^+$ lze vyložit jako přepočet cylindrických váhovacích koeficientů, kterými bylo zachyceno zvukové pole na odezvu mikrofону, kterou by musel být snímán zdroj zvuku proto, aby daný reproduktor v soustavě reprodukoval zvukové pole v místě počátku.

Dekodovací matice $\mathbf{D}$ je též nazývána rychlostní dekodér. Je vhodná pro rekonstrukci zvukového pole na nízkých kmitočtech. Pro vyšší kmitočty a pro reprodukci ve větších prostorech je třeba tuto matici váhovat diagonálními koeficienty $\mathbf{w}$. Vztahy pro výpočet těchto koeficientů jsou zobrazeny v tabulce 3.1.

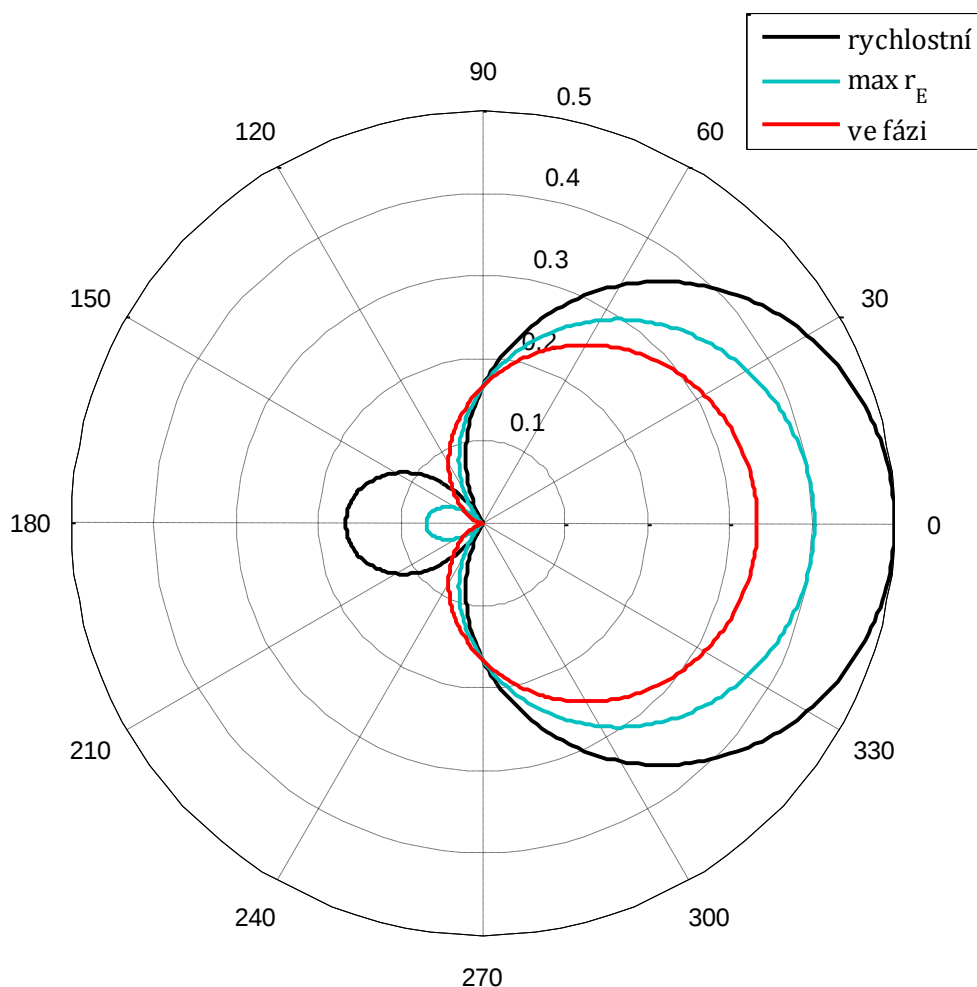
Metoda pro implementaci váhovacích koeficientů pro $\max r_E$ dekodér optimalizovaný na vyšší kmitočty dle tabulky 3.1 vypadá následovně.

```
float CAmbisonics::getMaxre(int m, int M)
{
    // m - současný řád ambisonie, M - maximální řád ambisonie
    return cosf(m*PI_/(2.f*M+2.f));
}
```

Podobně také pro dekodér „ve fázi“.

```
float CAmbisonics::getInphase(int m, int M)
{
    // m - současný řád ambisonie, M - maximální řád ambisonie
    return (factorial(M)*factorial(M))/(float)(factorial(M+m)*(factorial(M-
m)));
}
```

Na obr 4.5 jsou zobrazeny virtuální směrové mikrofonní odezvy pro všechny tři typy dekodérů. Je vidět, že se liší zejména velikostí a velikostí protifázové složky, což má mimo působení na různé podněty směrového slyšení přímý vliv na velikost oblasti kolem středu reproduktorového kruhu, kde bude zvukové pole kvalitně rekonstruováno. Rychlostní dekodér je optimalizován na centrálně umístěného posluchače, zatímco dekodér „ve fázi“ je vhodný pro posluchače vzdáleného od středu reproduktorového kruhu.



Obr. 4.5: Směrové charakteristiky virtuálního mikrofonu pro ambisonický dekodér prvního řádu bez optimalizace, s optimalizací $\max r_E$ a „ve fázi“.

5 POSLECHOVÝ TEST

Pro ověření funkčnosti implementace metody ambisonie byl proveden poslechový test. U poslechového testu je zásadní zvolit vhodnou metodu vyhodnocení lokalizace zvuku. V [22] jsou shrnuty některé používané metody.

5.1 Metoda vyhodnocení lokalizace

První používanou metodou bylo přímé vyhodnocování azimutu a elevace odhadem. Tomuto měření předcházela desetihodinová trénovací lekce, takže měření bylo časově náročné, navíc se musela po takové době projevit únava sluchu. Další metoda využívá zařízení, které sleduje směr natočení hlavy. U takového zařízení je ovšem problém s vyhodnocením azimutů, které jsou za posluchačem. Nejvhodnější metodou [22], která eliminuje mnohé zdroje chyb je metoda, ve které se po stimulaci vyhodnocuje poloha virtuálního zdroje pomocí reproduktoru umístěného na pohyblivém rameni. Posluchač nastaví toto rameno do pozice, odkud vnímá zdroj zvuku a tato pozice se zaznamená.

V tomto poslechovém testu byla použita jednoduchá metoda ukazování rukou a vyhodnocení azimutu bylo provedeno pomocí značek umístěných na podlaze. Těchto značek bylo 36 a byly umístěny rovnoměrně do kruhu kolem posluchače. Jednotlivé značky mezi sebou svíraly úhel 10° .

5.2 Testovaný systém

Systém použitý pro poslechový test se skládal ze šesti reproduktorů. Jako polohovací metoda byla zvolena ambisonie druhého řádu bez psychoakustických optimalizací. Rozložení reproduktorů, virtuální mikrofonní směrové odezvy, vektor rychlosti a vektor energie pro tento systém jsou zobrazeny v příloze A.2.

Místnost, ve které probíhalo měření byla částečně ztlumená. Dozvuk místnosti však měřen nebyl.

5.3 Testovací signál

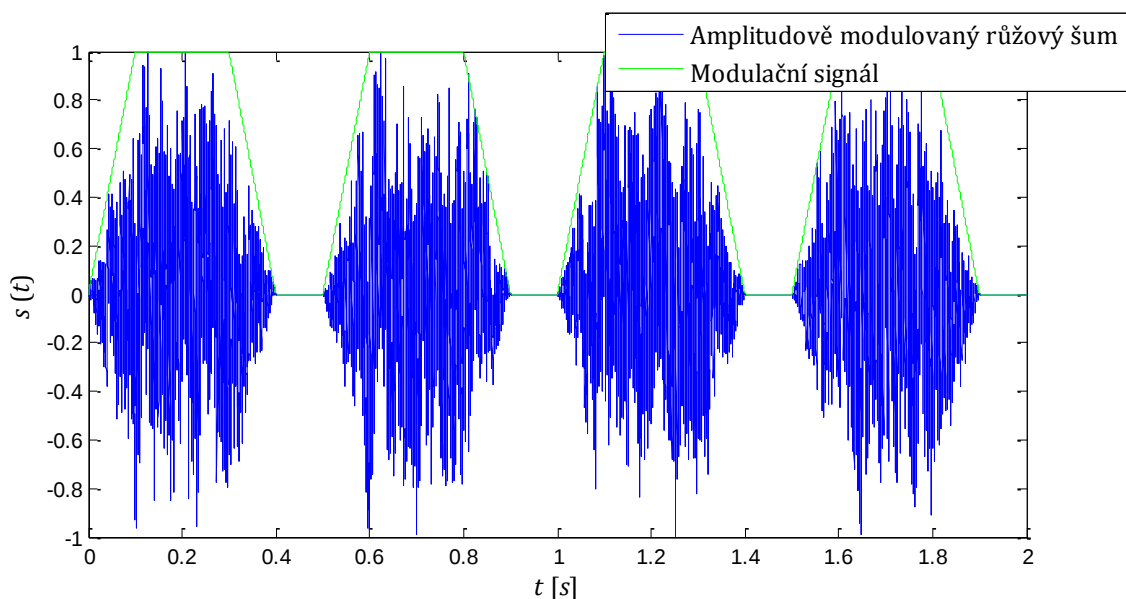
Jako testovací signál byl použit amplitudově modulovaný růžový šum, protože lépe aproximuje kmitočtovou odezvu sluchové soustavy než bílý šum a tím, že obsahuje všechny kmitočty je vhodný pro většinu podnětů používaných sluchovou soustavou k určení směru zvuku. Růžový šum je takový signál, jehož výkonová spektrální hustota je úměrná obrácené hodnotě kmitočtu. Jeho hlavní vlastností je, že v každé oktávě obsahuje stejný výkon.

Testovací signál má délku dvě sekundy. Je vytvořen čtyřmi pulzy růžového šumu, které jsou výsledkem modulace růžového šumu modulačním signálem (viz obr. 5.1). Náběžná hrana modulačního signálu je 100 ms, maximální amplitudou má 200 ms, sestupná hrana je 100 ms, následovaná 100 ms nulové úrovně.

5.4 Průběh testu

Poslechového testu se zúčastnilo 16 mužů ve věku 25 až 40 let, přičemž žádný nevykazoval známky poruchy sluchu. Práh slyšení však u žádného měřen nebyl. Každý z posluchačů vyhodnotil polovinu testu, tedy 18 hodnot.

Po příchodu posluchače do místnosti byl posluchač usazen do křesla, které bylo umístěno ve středu kruhu reproduktorů. Byl požádán, aby zavřel oči a byl otočen na židli, aby nevěděl,



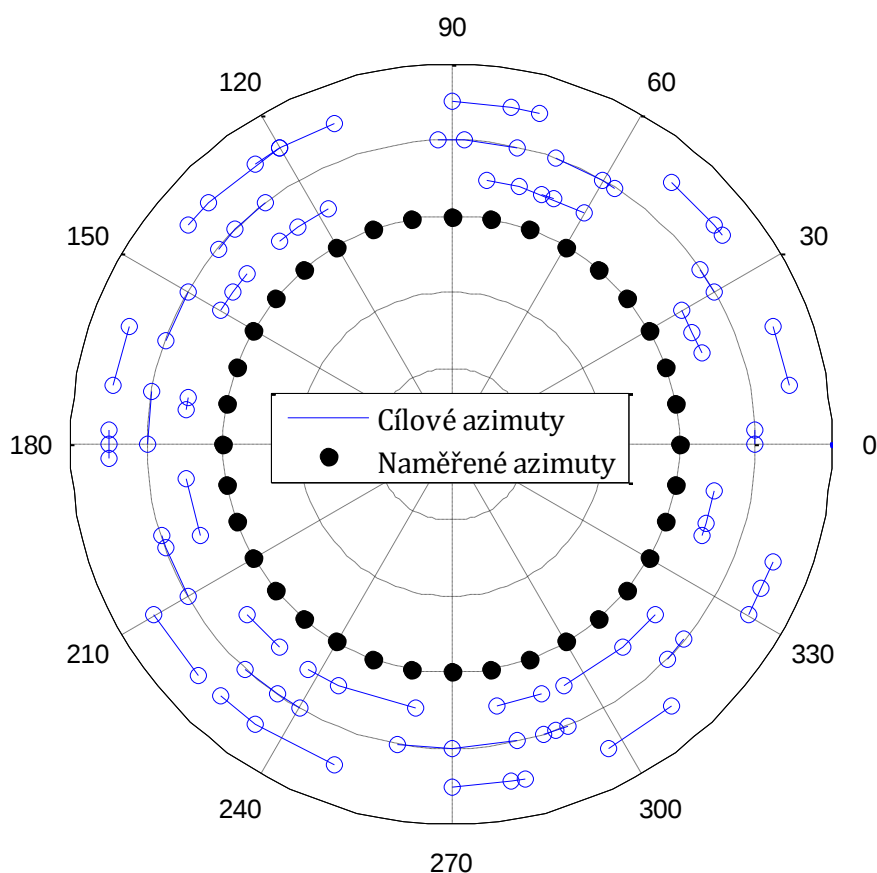
Obr. 5.1: Signál použitý v poslechovém testu.

kterým směrem jsou umístěny reproduktory. Je nutno podotknout, že po příchodu do místnosti posluchači věděli, kde jsou umístěny reproduktory, nevěděli ovšem, kterým směrem je umístěn azimut 0° , na který byli po otočení na židli namířeni.

Poté byla spuštěna série měření, kdy byl posluchači puštěn testovací signál a posluchač ukázal rukou, ze kterého směru si myslí, že zvuk přichází. Tento směr byl poté odečten od značek na podlaze a zaznamenán na připravený formulář. Pokud ukazovaný směr neodpovídal přesně značce, byl odečten azimut značky nejbližší ukázanému směru. Odchylka této metody je $\pm 5^\circ$. Pokud posluchač chtěl, byl mu testovací signál puštěn znovu.

5.5 Vyhodnocení testu a analýza hodnot

Hodnoty zaznamenané během poslechového testu jsou zobrazeny v příloze A.3. Oranžově jsou zvýrazněny hodnoty, u kterých došlo u posluchače k záměně směru před a za posluchačem a nejsou k analýze použity. Na obr. 5.2 jsou zobrazeny cílové azimuty, které byly podrobeny

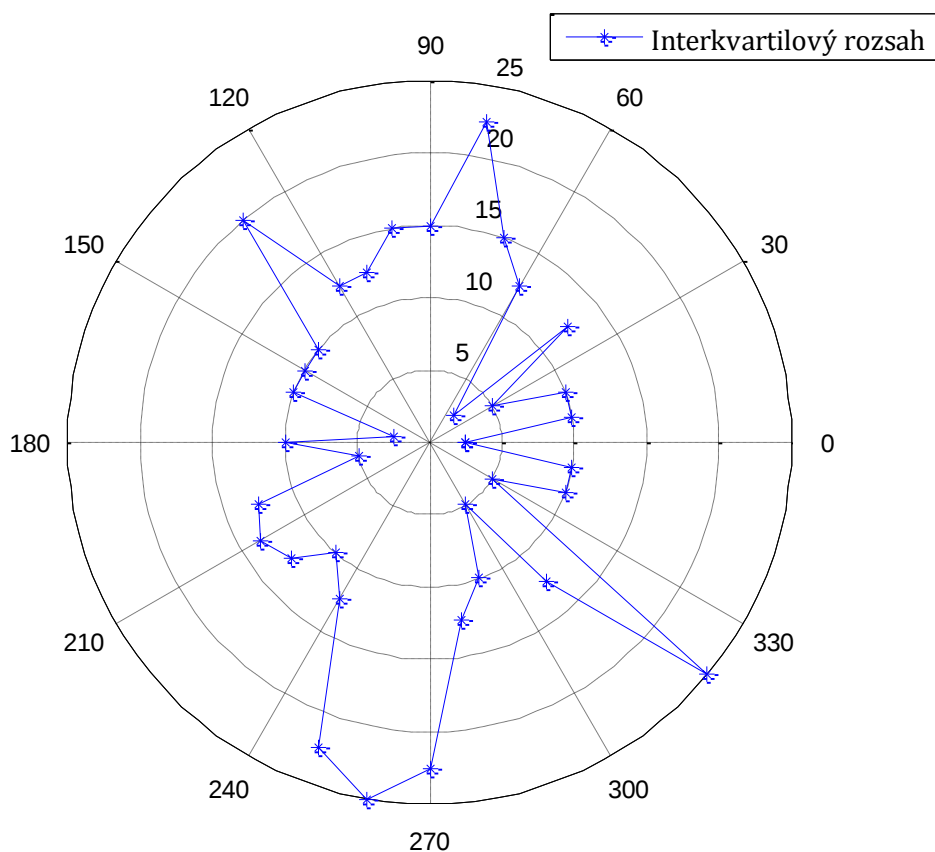


Obr. 5.2: Cílové a naměřené azimuty pro ambisonický systém druhého řádu; Symboly reprezentují 25. kvartil, medián a 75. kvartil

testu a 25. kvartil, medián a 75. kvartil, které byly získány z naměřených hodnot. 25. kvartil odděluje spodních 25 % hodnot od zbytku hodnot. Podobně 75. kvartil odděluje horních 75 % hodnot od zbytku hodnot. Medián, nebo také 50. kvartil rozděluje hodnoty na polovinu.

Dle obr. 5.2 lze tedy určit, že lokalizace je nejlepší vpředu mezi azimuty $\pm 45^\circ$, kde jsou všechny tři kvartily velmi blízko u sebe. Naopak nejhorší lokalizace je po bocích mírně vzadu, mezi 90° – 120° a 240° – 270° . V některých případech se intervaly mírně překrývají.

Na obr. 5.3 jsou zobrazeny interkvartilové rozsahy. Ty je možno interpretovat jako lokalizační mlhu. Pro ambisonický systém druhého řádu je predikována lokalizační mlha na 30° [12], této hodnotě se blíží lokalizace převážně po stranách, naopak vpředu a vzadu je lokalizace mnohem kvalitnější.



Obr. 5.3: Interkvartilové rozsahy pro jednotlivé měřené azimuty poslechového testu

6 ZÁVĚR

V této práci byly popsány vlastnosti lidského sluchu, zejména vnímání směrovosti zvuku a byly předvedeny dvě metody pro prostorovou reprodukci zvuku: ambisonie a vektorově bázové amplitudové polohování.

Na vnímání směrovosti sluchu se nejvíce podílí interaurální časový rozdíl a interaurální rozdíl intenzit. Interaurální časový rozdíl rozhoduje více u nižších kmitočtů a interaurální rozdíl intenzit u vyšších kmitočtů. Hranice mezi nižšími a vyššími kmitočty je v tomto případě zhruba 1500 Hz a přechod mezi nimi je pozvolný.

Vektorově bázové amplitudové polohování je polohovací metoda pro neomezený počet reproduktorů. Je založená na zobecnění principu stereofonie na dvojrozměrný systém pomocí vektorových bází. Umožňuje však i trojrozměrné polohování. Jde o relativně jednoduchou a účinnou metodu. Její hlavní nevýhodou je nemožnost záznamu a nepřenositelnost mezi různými reproduktorovými konfiguracemi.

Ambisonie umožňuje záznam, reprodukci a syntézů trojrozměrných zvukových scén. Kvalita lokalizace a počet kanálů potřebných pro reprodukci se odvíjí od řádu ambisonie. V této práci je použita pouze horizontální ambisonie. První řád horizontální ambisonie využívá tři kanály pro přenos či záznam a každý další řád přidává dva kanály. Ambisonie není závislá na reproduktorové konfiguraci, avšak pro neregulární konfigurace poskytuje horší výsledky.

Obě metody byly navrženy v prostředí Matlab. Při tvorbě algoritmů byl brán ohled na co nejsnadnější implementaci do jazyka C++. Nevyužívají proto interních funkcí prostředí Matlab, ani možnosti maticových výpočtů a využívají místo nich cykly.

Metody VBAP a ambisonie byly implementovány jako plug-in modul systému VST v jazyce C++. K plug-in modulu bylo vytvořeno GUI generované šablonou a dvojrozměrný polohovací kruhový prvek, který slouží k intuitivnějšímu nastavení azimutu.

Funkce implementace metody ambisonie byla ověřena poslechovým testem. Pro ambisonii druhého řádu je predikovaná hodnota lokalizační mlhy 30° , výsledky poslechového testu pak nasvědčují o lepší lokalizaci vpředu a vzadu ($2,5^\circ$ – 10°), zatímco po stranách je lokalizace podstatně horší ($12,5^\circ$ – 25°) a blíží se predikované hodnotě lokalizační mlhy.

7 LITERATURA

- [1] MARŠÁLEK, P. *Kódování zvuku ve sluchové dráze a jeho aplikace na prostorové slyšení*. [Elektronický dokument] Brno : ÚTKO FEKT, VUT, 29. 10 2008.
<http://fanous.utko.feec.vutbr.cz/marsalek_oct08/marsalek.pdf>.
- [2] GARDNER, W. *3-D Audio Using Loudspeakers*. Massachusetts : Springer, 1998. ISBN 978-0792381563.
- [3] PULKKI, V. *Virtual source positioning using vector base amplitude panning*. Espoo : Helsinki University of Technology, 2001. ISSN 1456-6303.
- [4] WIGGINS, B. *An Investigation into the Real-time Manipulation and Control of Three-dimensional Sound Fields. PhD Thesis*. Derby : University of Derby, 2004.
- [5] BAUER, B. B. Phasor Analysis of Some Stereophonic Phenomena. *J. Acoust. Soc. Am.*, 1961, Sv. 33, stránky 1536-1539.
- [6] BERNFELD, B. Attempts for Better Understanding of the Direction Stereophonic Listening Mechanism. *J. Audio Eng. Soc.*, 1973, Sv. 21.
- [7] ROCHESSO, D. Spatial Effects. Udo Zolzer. *Digital Audio Effects*. New York : John Wiley and Sons, Ltd.
- [8] ZÖLZER, U. *DAFX – Digital Audio Effects*, 1st ed. New York: John Wiley & Sons, Ltd, 2002, 533 p. ISBN 0-471-49078-4.
- [9] BENJAMIN, E. M., LEE, R., HELLER, A. J. Localization in Horizontal-Only Ambisonic Systems. In *Preprints of the 121st Convention of the Audio Engineering Society*. San Francisco, 2007. s. 12.
- [10] HELLER, A. J., LEE, R., BENJAMIN, E. M. Is My Decoder Ambisonic?. In *Preprints of the 125th Convention of the Audio Engineering Society*. San Francisco, 2008. s. 21.
- [11] MOREAU, S. *Étude et réalisation d'outils avancés d'encodage spatial pour la technique de spatialisation sonore Higher Order Ambisonics : microphone 3D et contrôle de distance*. Le Mans, Francie : University of Maine, 2006. 189 s. Dizertační práce.
- [12] DANIEL, J. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Paris, Francie : University of Paris, 2000. 319 s. Dizertační práce.

- [13] GERZON, M. Surround Sound Psychoacoustics. *Wireless World*. Leden 1974, vol. 80, s. 483-486.
- [14] DANIEL, J., NICOL, R., MOREAU, S. Further Investigations of High Order Ambisonics and Wave Field Synthesis for Holophonic Sound Imaging. In *Preprints of the 114th Convention of the Audio Engineering Society*. Amsterdam, The Netherlands, 2003. s. 136-153.
- [15] DANIEL, J. Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format. In *AES 23rd International Conference*. Copenhagen, Denmark, 2003.
- [16] GERZON, M. A. General Metatheory of Auditory Localisation. In *Preprints from the 92nd Audio Engineering Society Convention*. Vienna, Austria, 1992.
- [17] MAKITA, Y. On the directional localisation of sound in the stereophonic sound field. In *European Broadcasting Union journal : Part A - Technical*. 1962. s. 102-108.
- [18] DEBOER, K. Stereophonic sound Production. *Phillips Technical Review*. 1940. S. 107-144.
- [19] MALHAM, D. G. Experience with large area 3-D ambisonic sound systems. *Proceedings of the Institute of Acoustics*. 1992, vol. 5, no. 14, s. 209-215.
- [20] LEE, R. *Shelf filters for ambisonic decoders* [online]. 2008 [cit. 2009-04-12]. Dostupný z WWW: <<http://www.ambisonia.com/Members/ricardo/shelfs.zip>>.
- [21] WEISSTEIN, E. W. *Polar Coordinates* [online]. 1999-2009 , 9.3.2009 [cit. 2009-05-11]. Dostupný z WWW: <<http://mathworld.wolfram.com/PolarCoordinates.html>>.
- [22] BERTET, S., DANIEL, J., PARIZET, E. LAETITA, G. WARUSFEL, O. Investigation of the perceived spatial resolution of higher order ambisonics sound fields: A subjective evaluation involving virtual and real 3D microphones. At *30th International Conference of the Audio Engineering Society*. Finland, Saariselkä, 2007.
- [23] WIGGINS, B. The generation of panning laws for irregular speaker arrays using heuristic methods. At *31st International Conference of the Audio Engineering Society*. UK, London, 2007.
- [24] MAČÁK, J. *Návrh algoritmů číslicového zpracování signálů pro simulaci kytarových zesilovačů založených na obvodové analýze analogových prototypů*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008. Počet stran 76. Vedoucí diplomové práce Ing. Jiří Schimmel, Ph.D.

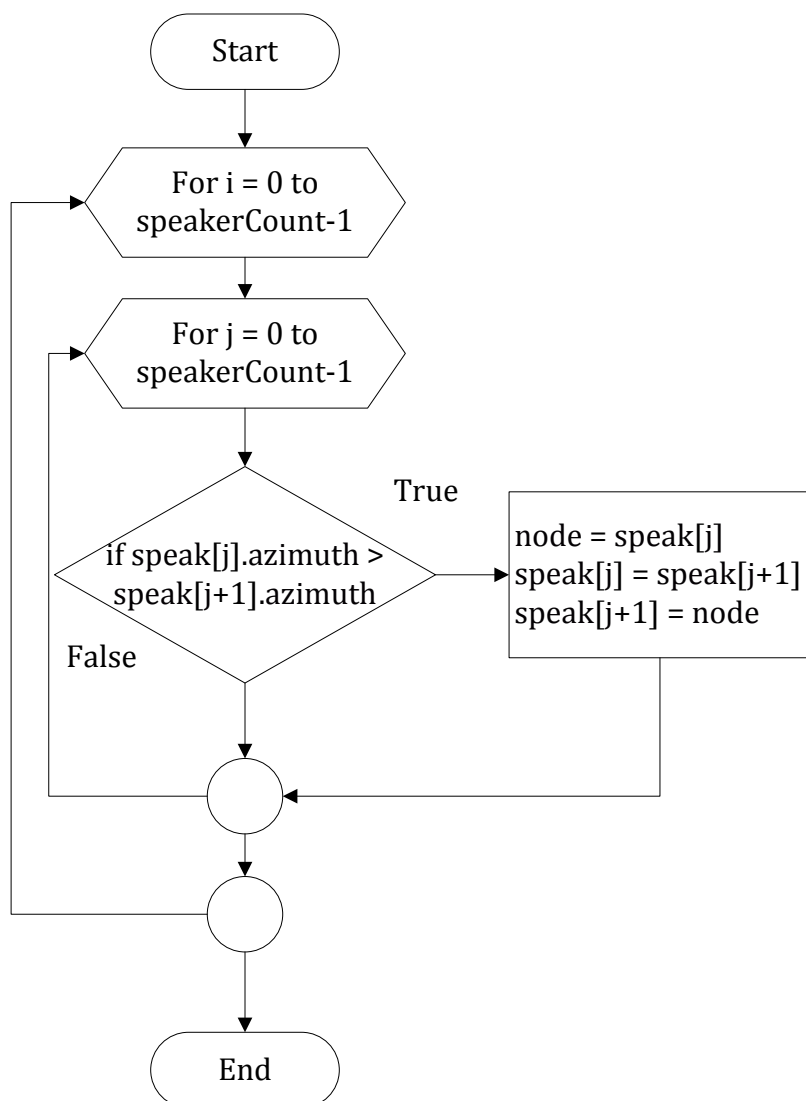
[25] GOLUB, G., KAHAN, W. Calculating the Singular Values and Pseudo-Inverse of a Matrix. In *Journal of the Society for Industrial and Applied Mathematics : Series B, Numerical Analysis*. 2nd edition. 1965. s. 205-224.

SEZNAM PŘÍLOH

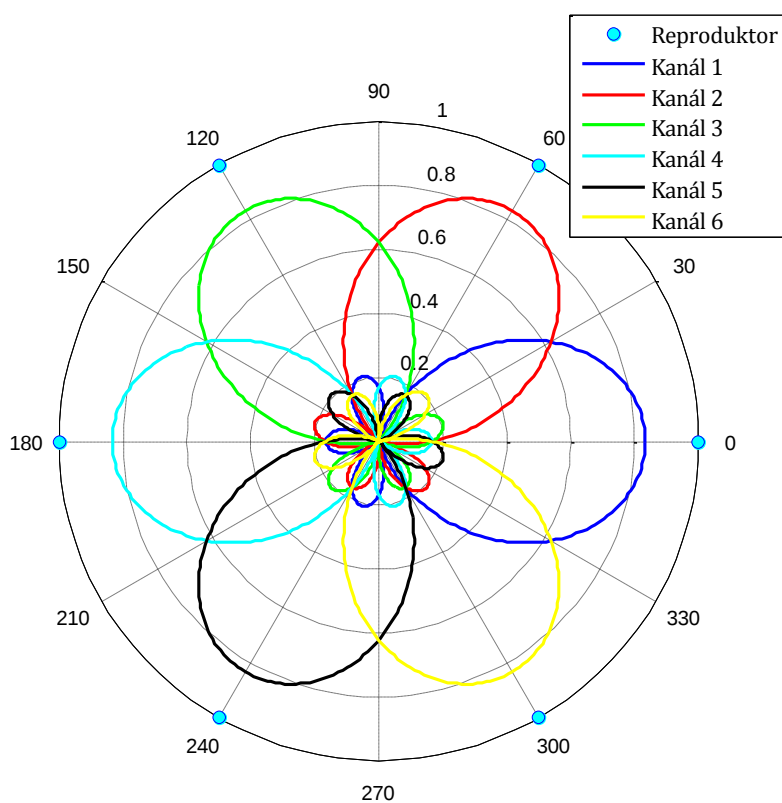
A	Přílohy k textu.....	57
A.1	Vývojový diagram bubble sortu pro metodu vektorově bazového amplitudového polohování	57
A.2	Konfigurace použita v poslechovém testu	58
A.3	Tabulka naměřených hodnot z poslechového testu.....	59
B	Příloha k CD	60
B.1	Adresářová struktura.....	60

A Přílohy k textu

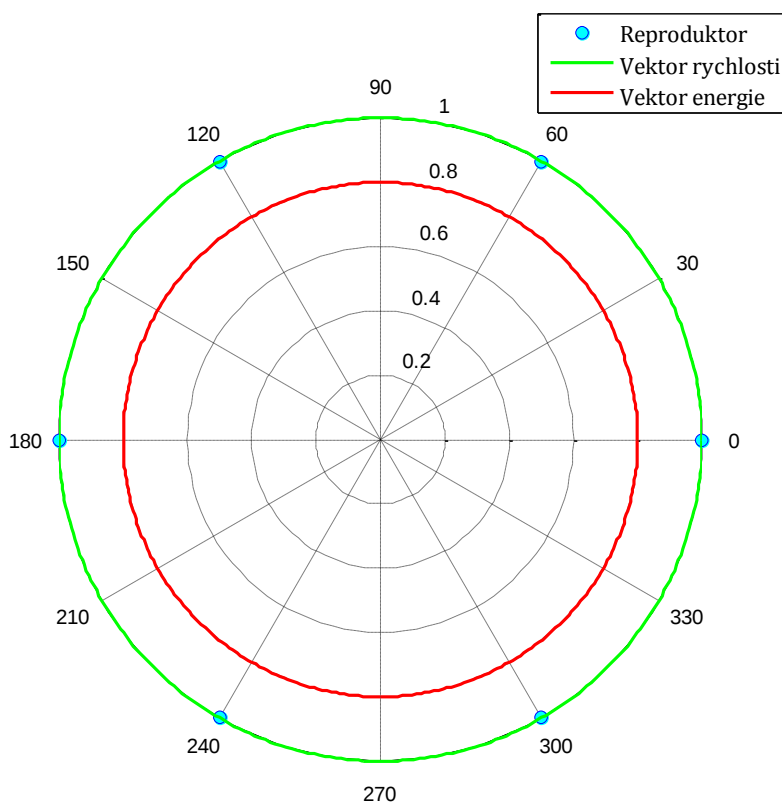
A.1 Vývojový diagram bubble sortu pro metodu vektorově bazového amplitudového polohování



A.2 Konfigurace použita v poslechovém testu



Směrové charakteristiky virtuálních mikrofونů použitých při poslechovém testu



Velikost vektoru rychlosti a vektoru energie pro ambisonický systém použitý pro poslechový test

A.3 Tabulka naměřených hodnot z poslechového testu

Úhel	Měření								První kvartil	Medián	Třetí kvartil	IQR
	1	2	3	4	5	6	7	8				
0	0	0	10	0	0	10	0	-10	0	0	2,5	2,5
10	10	10	20	10	20	0	20	10	10	10	20	10
20	30	30	20	30	20	20	20	40	20	25	30	10
30	30	30	60	30	50	20	30	30	30	30	35	5
40	30	50	40	50	50	30	40	40	37,5	40	50	12,5
50	70	70	70	70	70	50	60	90	67,5	70	70	2,5
60	50	60	70	70	70	40	60	60	57,5	60	70	12,5
70	80	80	60	90	60	90	80	90	75	80	90	15
80	50	60	80	90	90	60	70	80	60	75	82,5	22,5
90	90	80	70	90	85	110	60	100	77,5	87,5	92,5	15
100	80	120	140	110	110	140	120	120	110	120	125	15
110	120	120	110	130	130	130	110	140	117,5	125	130	12,5
120	120	150	130	140	120	140	130	140	127,5	135	140	12,5
130	120	120	120	140	140	140	160	130	120	135	140	20
140	150	140	140	140	140	150	150	150	140	145	150	10
150	150	160	150	160	140	30	150	170	150	150	160	10
160	160	170	160	160	150	20	170	190	160	160	170	10
170	170	170	170	170	190	170	140	180	170	170	172,5	2,5
180	180	170	170	170	180	0	180	180	170	180	180	10
190	170	180	170	180	180	190	190	180	177,5	180	182,5	5
200	200	180	190	180	200	200	210	200	187,5	200	200	12,5
210	200	200	200	210	210	180	230	190	197,5	200	210	12,5
220	210	210	210	230	200	230	220	210	210	210	222,5	12,5
230	220	220	210	230	230	230	250	230	220	230	230	10
240	210	240	220	240	230	240	240	230	227,5	235	240	12,5
250	240	230	230	250	220	250	250	220	227,5	235	250	22,5
260	280	230	240	260	240	250	270	230	237,5	245	262,5	25
270	290	260	270	270	250	300	260	280	260	270	282,5	22,5
280	290	270	270	280	270	290	280	280	270	280	282,5	12,5
290	290	280	290	280	290	280	280	270	280	280	290	10
300	290	290	290	290	280	300	310	280	287,5	290	292,5	5
310	310	310	300	290	310	310	310	280	297,5	310	310	12,5
320	320	300	320	310	310	280	320	280	295	310	320	25
330	320	320	300	320	320	320	320	280	315	320	320	5
340	340	340	330	340	330	330	350	270	330	335	340	10
350	350	340	340	350	345	340	360	340	340	342,5	350	10

B Příloha k CD

B.1 Adresářová struktura

Adresář „skripty“	Tento adresář obsahuje skripty v prostředí Matlab, které byly použity k návrhu metod vektorově bazového amplitudového polohování, ambisonie, frekvenčně kompenzovaných filtrů.
Adresář „plug-in“	Tento adresář obsahuje implementaci metod VBAP a ambisonie jako plug-in technologie VST v prostředí C++.
Adresář „obrázky“	Tento adresář obsahuje originální soubory k obrázkům použitým v této práci.

B.2 Skripty pro prostředí Matlab

<code>AMBI</code>	Vypočítá dekodovací matici pro ambisonické signály prvního až třetího řádu, zobrazí jejich virtuální mikforonní směrové charakteristiky a vypočítá a zobrazí vektor rychlosti a vektor energie.
<code>maxrE</code>	Pro zadaný řád ambisonie m a maximální řád ambisonie M vrátí váhovací koeficient pro dekodovací matici dekodéru energie.
<code>inph</code>	Pro zadaný řád ambisonie m a maximální řád ambisonie M vrátí váhovací koeficient pro dekodovací matici dekodéru „ve fázi“.
<code>VBAP2</code>	Vypočítá a zobrazí zesilovací faktory pro jednotlivé reproduktory při zadané reproduktorové konfiguraci.
<code>vypoctizisk</code>	Vrací normovaný zisk reproduktorových signálů.
<code>pmsplit</code>	Vrací koeficienty horní a dolní propusti fázově kompenzovaných filtrů pro zadaný mezní kmitočet a zobrazí jejich modulové a fázové frekvenční charakteristiky.