



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

3D AUDIO STANDARDS V PROSTŘEDÍ DOMÁCÍHO KINA

3D AUDIO STANDARDS IN HOME THEATRE ENVIRONMENT

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Daniel Kováč

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Miroslav Balík, Ph.D.

BRNO 2017

Bakalářská práce

bakalářský studijní obor **Audio inženýrství**
Ústav telekomunikací

Student: Daniel Kováč

ID: 170614

Ročník: 3

Akademický rok: 2016/17

NÁZEV TÉMATU:

3D Audio standardy v prostředí domácího kina

POKYNY PRO VYPRACOVÁNÍ:

Na základě již provedené analýzy reprodukce zvukových objektů dle standardu MPEG-H realizujte algoritmus, který bude schopen interpretovat vstupní metadata dle standardu MPEG-H a následně provede vektorové panoramování zvukového objektu, který metadata definuje. Algoritmus bude podporovat reproduktorové konfigurace 5.1.2, 5.1.4 a 7.1.4. Algoritmus realizujte v Matlabu s grafickým uživatelským rozhraním. Nejdříve bude provedena grafická interpretace vstupních metadat a následně dojde k vygenerování zvukových signálů pro všechny reproduktory dle zvolené reproduktorové konfigurace. Proveďte poslechové testy, které porovnejí subjektivní vnímání polohy zvukového objektu v prostoru při použití konfigurací 5.1.2, 5.1.4 a 7.1.4.

DOPORUČENÁ LITERATURA:

[1] PULKKI, Ville. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. Journal of the Audio Engineering Society, 45(6), 456-466, 1997.

[2] FUEG, Simone, et al. Design, coding and processing of metadata for object-based interactive audio. In: Audio Engineering Society Convention 137. Audio Engineering Society, 2014.

Termín zadání: 1.2.2017

Termín odevzdání: 8.6.2017

Vedoucí práce: Ing. Miroslav Balík, Ph.D.

Konzultant:

doc. Ing. Jiří Mišurec, CSc.
předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské 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 části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

Abstrakt

Tato práce se zabývá objektově založeným zvukem. Je zde popsán rozdíl mezi kanálově a objektově založeným zvukem a vývoj tohoto zvuku do dnešní podoby. Následně jsou popsány formáty, které pracují s objekty, a ty jsou porovnány mezi sebou. Je zde specifikovaný formát MPEG-3D Audio, který je dnes standardem, a uveden způsob polohování virtuálního zdroje zvuku technikou vektorově amplitudového panorámování založeného na vektorech. Dále je zde popsáno, jakým způsobem byl sestaven kodér a dekodér, podporující reproduktorové konfigurace 5.1.2, 5.1.4 a 7.1.4, podle specifikace standardu MPEG-H 3D Audio. V poslední řadě se práce zabývá problémem při panorámování objektového zvuku na standardní konfiguraci a je zde také zmínka o poslechovém testu.

Klíčová slova

Objektově založený zvuk, kanálově založený zvuk, MPEG-H 3D Audio, metadata, VBAP, panorámování, kodér, dekodér, reproduktorová konfigurace

Abstract

This work deals with object-based sound. There is a description of the difference between channel-based and object-based sound and the development of this sound to today's form. Subsequently, formats that work with objects are described, and they are compared with each other. MPEG-H-3D Audio format is specified here, which is a standard today, and also the method of positioning the virtual audio source by vector-base amplitude panning. It also describes how an encoder and decoder supporting speaker configurations 5.1.2, 5.1.4 and 7.1.4 has been developed according to the MPEG-H 3D Audio specification. Lastly, the work deals with the problem of panning the object sound to the standard configuration and there is also a mention of the listening test.

Keywords

Object-based audio, channel-based audio, MPEG-H 3D audio, metadata, VBAP, panning, encoder, decoder, speaker configuration

Bibliografická citace

KOVÁČ, D. *3D Audio standardy v prostředí domácího kina*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2017. 53 s. Vedoucí bakalářské práce Ing. Miroslav Balík, Ph.D..

Prohlášení autora o původnosti

Prohlašuji, že svou bakalářskou práci na téma 3D Audio standardy v prostředí domácího kina jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a 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/nebo majetkových a jsem si plně vědom následku porušení ustanovení § 11 a následujících autorského zákona č.121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonu (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne 7.6.2016

.....
podpis autora

Poděkování

Mé poděkování patří v první řadě mému vedoucímu práce panu docentu Miroslavu Balíkovi za cenné rady, trpělivost a ochotu. Dále bych chtěl poděkovat panu doktoru Jiřímu Schimmelovi rovněž za ochotu a pomoc při chystání poslechového testu v laboratoři. Nakonec bych rád vyjádřil vděk všem lidem, kteří se tématu věnovali přede mnou, takže jsem mohl čerpat z jejich zveřejněných prací.

V Brně dne 7.6.2016

.....
podpis autora

Obsah

Seznam obrázků	9
Seznam tabulek.....	10
Úvod.....	11
1 Historie a vývoj reprodukce prostorového zvuku.....	12
1.1 Prostorový zvuk založený na kanálech.....	12
1.1.1 Analogový záznam zvuku	12
1.1.2 Digitální záznam zvuku	13
1.2 Prostorový zvuk založený na objektech	13
2 Formáty a standardy založené na objektech	14
2.1 Formát AURO 3D	14
2.2 Formát Dolby Atmos.....	15
2.2.1 Nástroje pro editaci zvuku formátu Dolby Atmos.....	17
2.3 Formát DTS-X.....	17
2.4 Standard MPEG-H 3D Audio.....	18
2.4.1 Otevřený standard Audio Definition Model.....	19
2.5 Porovnání současných formátů a standardů	19
3 Specifikace standardu MPEG-H 3D Audio	20
3.1 Princip dekódování prostorového zvuku	20
3.1.1 Signály typu Kanály	20
3.1.2 Signály typu Objekty.....	21
3.1.3 Signály typu Spatial Audio Object Coding	21
3.1.4 Signály typu Higher Order Ambisonics	22
3.2 Struktura metadat v bitovém proudu	22
3.3 Přenos a dekódování objektových metadat	23
3.3.1 Dekódování metadat komprimovaných lineární interpolací.....	23
3.3.2 Dekódování metadat komprimovaných s malým zpožděním.....	24
4 Virtuální polohování zvukového zdroje	25
4.1 Vektorově založené amplitudové panoramování	25
4.1.1 Dvourozměrné panoramování	25
4.1.2 Trojrozměrné panoramování	27
5 Algoritmus interpretující vstupní metadata standardu MPEG-H sestavený v Matlabu.....	28
5.1 Kodér standardu MPEG-H	28
5.1.1 Vytvoření souboru .OAM a zápis do něj.....	29

5.1.2	Hlavička souboru .OAM	29
5.1.3	Čtení a úprava vstupních zvukových dat.....	29
5.1.4	Vytvoření a zápis metadat statického a dynamického objektu.....	30
5.2	Dekodér standardu MPEG-H	33
5.2.1	Načítání souboru .OAM a výpis z něj	34
5.2.2	Rozmístění reproduktorů v prostoru.....	34
5.2.3	Výpočet umístění stropních reproduktorů s korekcí hlasitosti	35
5.2.4	Úprava kanálových signálů	37
5.2.5	Panoramování objektového signálu.....	37
5.2.6	Vytvoření a úprava výstupních signálů	42
6	Poslechové testy	43
6.1	Příprava pracoviště	44
6.2	Ověření správnosti elektroakustického řetězce	45
	Závěr.....	46
	Použitá literatura.....	48
	Seznam použitých zkratk	50
	Seznam příloh.....	52

Seznam obrázků

Obrázek 2.1: Pracovní postup Dolby Atmos A/V Receiveru pro domácí kino	15
Obrázek 2.2: Dolby Atmos reproduktorová konfigurace 5.1.2 pro domácí kino.....	16
Obrázek 2.3: Dolby Atmos reproduktorová konfigurace 5.1.4 pro domácí kino.....	16
Obrázek 2.4: Dolby Atmos reproduktorová konfigurace 7.1.4 pro domácí kino.....	17
Obrázek 3.1: Blokové schéma dekodéru MPEG-H 3D Audio.....	20
Obrázek 3.2: Proces dekódování metadat lineární interpolací (na obrázku jeden rámeček)	24
Obrázek 4.1: Vektorově založené amplitudové panoramování pro 2 reproduktory	25
Obrázek 5.1: Graficky zpracovaný kodér standardu MPEG-H.....	28
Obrázek 5.2: Graficky zpracovaný dekodér standardu MPEG-H.....	33
Obrázek 5.3: Umístění stropních reproduktorů.....	35
Obrázek 5.4: Modifikovaná technika VBAP	38
Obrázek 5.5: Trajektorie objektu zepředu dozadu (po oblouku) při konfiguraci 7.1.4.....	41

Seznam tabulek

Tabulka 2.1: Přehled konfigurací formátu Auro 3D	14
Tabulka 2.2: Porovnání formátů objektově založeného zvuku	19
Tabulka 3.1: Přehled základních informací zakódovaných v objektových metadatech	23
Tabulka 5.1: Azimuty reproduktorů pro konfiguraci 5.1.2	34
Tabulka 5.2: Azimuty reproduktorů pro konfiguraci 5.1.4	34
Tabulka 5.3: Azimuty reproduktorů pro konfiguraci 7.1.4	34
Tabulka 5.4: Pořadí kanálů v matici a jejich značení.....	43
Tabulka 6.1: Polohy statických objektů při poslechovém testu	43

Úvod

V dnešní době se při ozvučování především kinosálů, ale také domácích kin nebo konferenčních místností, stále více rozmáhá koncept 3D prostorového zvuku. Existuje několik způsobů, jakými lze docílit jeho reprodukci. Nejpoužívanější z nich je založený na tzv. objektech a jeví se dnes jako nejefektivnější. Většina formátů 3D prostorového zvuku, které se dnes vyskytují, pracují právě s tímto objektově založeným zvukem.

Mou snahou v nadcházejícím textu je přiblížit čtenáři, jak objektově založený zvuk funguje, jakým principem se objekty přenáší v datovém toku, dále pak přiblížit způsob jejich kódování a následného rendrování do prostoru. Nejdříve popíši, jak se reprodukce prostorového zvuku vyvíjela až do nynější podoby zvuku objektově založeného. Poté porovnáím formáty, které se v dnešní době vyskytují, především na základě jejich reproduktorových konfigurací a na základě jejich reprodukce objektů. Mojí snahou bude popsat, jaké kódování a dekódování formáty používají a jaká je charakteristika jejich bitových toků. Budu se snažit zjistit, jestli jsou formáty mezi sebou navzájem kompatibilní a jestli existuje určitý způsob převodu jejich objektů. Po té, co formáty popíši, provedu jejich srovnání. Specifikuji standard MPEG-H 3D Audio tak, aby bylo možné z této specifikace sestavit dekodér tohoto standardu, a následně se budu věnovat polohování virtuálního objektu v prostoru využitím panoramovacích technik

Sestavím dekodér na základně specifikovaného standardu MPEG-H 3D audio, který bude podporovat konfigurace 5.1.2, 5.1.4 a 7.1.4. Algoritmus bude interpretovat vstupní metadata daného standardu a následně z těchto informací panoramovat signál zvukového souboru na zvolenou reproduktorovou konfiguraci. Protože reproduktorová konfigurace je do jisté míry závislá na velikosti místnosti, bude dekodér také obsahovat algoritmus, který dá uživateli potřebné informace o rozmístění reproduktorů v prostoru. Vše bude přehledně sestaveno v grafickém uživatelském prostředí, kde se budou také při dekódování souboru zobrazovat informace o daném souboru a objektu, jehož trajektorie se navíc bude vykreslovat do grafu. Mým cílem bude provést poslechové testy, které porovnají subjektivní vnímání polohy objektu v prostoru při různých konfiguracích. Z těchto testů by mělo být jisté, zda mají stropní reproduktory vliv na subjektivní vnímání polohy objektu a do jaké míry je vnímán rozdíl při poslechu s dvěma stropními reproduktory a čtyřmi.

1 Historie a vývoj reprodukce prostorového zvuku

Za prostorový zvuk můžeme považovat ten zvuk, který je rozšířený o druhý nebo třetí rozměr. V případě druhého rozměru mluvíme o horizontální rovině šíření zvuku, čehož docílíme vícekanálovým zvukovým systémem. Z mono reprodukce tedy rozšiřujeme systém na stereo reprodukci, což je rozšíření z jednoho kanálu na dva. Tento systém se v dnešní době používá nejvíce, protože je nejdostupnější a to zejména při poslechu hudby nebo hraní počítačových her. Pro stereo poslech sluchátky nebo také dvěma reproduktory (s posluchačem musí tvořit rovnostranný trojúhelník) je možné docílit prostorového vjemu zvuku za pomoci tzv. HRTF (Head-Related Transfer Function - přenosová funkce hlavy) **Chyba! Nenalezen zdroj odkazů.** Zde se při mixu signálu využívá simulace přenosové funkce lidské hlavy, torza a uší. Za pomoci rozdílných vlastností zvuku v obou kanálech (intenzita, barva, časové zpoždění mezi kanály) pak člověk dokáže lokalizovat zdroj zvuku v prostoru. Přidáním dalších kanálů do horizontální roviny docílíme tzv. surround systému. Zvětší se tak plocha, ve které lze prostorové informace lokalizovat. Pro posluchače se také stává zvukové prostředí věrnější, protože signál k němu může dorazit z více směrů. Nemusí se tedy simulovat odražený zvuk, neboť zvuk sám o sobě je již modulován přenosovou soustavou lidského těla (akustický stín hlavy, odraz od uší,...) a na základě toho člověk dostává informaci o poloze zdroje. Oproti mono nebo stereo poslechu tak může člověk přijímat signál i ze stran nebo ze zadu. Toho se využívá hlavně ve filmovém průmyslu. Pokud chceme zvuk vnímat i v třetím rozměru, je třeba přidat další zvukové kanály do vertikální roviny a využít tak odrazu zvuku od našich ramen, obvykle se jedná o reproduktory připevněné na strop.

1.1 Prostorový zvuk založený na kanálech

1.1.1 Analogový záznam zvuku

Jako první významná firma zabývající se prostorovým zvukem byla Dolby Laboratories, kterou založil roku 1965 americký vědec Ray Dolby. Prvotním zaměřením společnosti bylo potlačení statického šumu ve zvukových nahrávkách. Později se začala také zabývat vícekanálovým zvukem a v polovině 70. let uvedla na svět formát Dolby Stereo. Do této doby byla kina ozvučována pouze monofonně nebo stereofonně, popřípadě vícekanálově avšak neefektivně za pomoci drahého magnetického systému Cinema Scope od společnosti FOX. Dolby Stereo se snažilo o zaznamenání čtyř zvukových stop na filmový pás. Tři reproduktory jsou za plátnem a jsou označovány jako levý, střední a pravý kanál. Čtvrtý kanál je tzv. prostorový (Surround kanál) a je reprodukován před plátnem pomocí reproduktorů umístěných za diváky a po stranách sálu. Protože stopy nebyly v té době ještě komprimované a na filmový pás se vlezly pouze dvě, využilo se maticového kódování čtyř stop do dvou (Left Total a Right Total). Před reprodukcí je signál opět možné rozdělit zpět nebo snadno reprodukovat pouze dvěma kanály. Po čase se formát rozrostl na tzv. Dolby stereo 70mm Six Track, byla přidána basová složka, která není na filmovém pásu přenášena samostatně, ale až

daný přehrávač odfiltruje od obou stop frekvence vyšší 120 Hz a vzniká tak kanál dnes známý jako výstup pro subwoofery. Společnost Dolby Laboratories se stále posouvala dopředu a vyvíjela další formáty. Dolby Surround, který byl v roce 1982 představen veřejnosti, ocenili zejména příznivci domácího kina s přehráváním pomocí kazet VHS nebo Betamax. Finálním analogovým formátem se stal Dolby Pro Logic, který využívá již dnes klasický 5.1¹ zvuk, ale také 6.1, 7.1 nebo 9.1 kanálový prostorový zvuk, který se však pro domácí kina příliš neuchytil.

1.1.2 Digitální záznam zvuku

S příchodem digitalizace se adaptovaly i formáty týkající se přenosu a reprodukce prostorového zvuku. Vědci z Dolby Laboratories představili formát Dolby Digital, který se nejčastěji vyskytuje v konfiguraci 5.1 (levý L, centrální C, pravý R, levý prostorový kanál Ls a pravý prostorový kanál Rs + kanál LFE (low-frequency effects), který obsahuje basovou složku). Dolby Digital však také podporuje mono nebo dřívější konfigurace známé pro Dolby Stereo nebo Dolby Surround. Nejdříve se používal pro poskytování digitálního zvuku zaznamenaného na 35 mm filmový pásek a to především v kinosálech a později se začal používat i v digitálním televizním vysílání a stal se standardem pro záznam digitálního zvuku na disky DVD a jako základní formát pro Blue-Ray disky. Jeho zvukový formát obsahuje ztrátovou kompresi, dnes je standardem a nese označení AC-3. Může mít maximální datový tok 640kbit/s. Dolby Digital tak začal konkurovat společnosti Digital Theatre System (DTS), vytvořené Beardem a Stevenem Spielbergem, která vydala svůj formát v roce 1993. Přestože systém využívá ztrátovou kompresi, jeho datový tok může dosahovat až 1536 kbit/s. DTS využívá všechny známé reproduktorové konfigurace, ale oproti Dolby Digital nebo SDDS neobsahuje zálohu v podobě analogového záznamu. SDDS je zkratka pro další systém digitálního ozvučování kin a celý název je Sony Dynamic Digital Sound. Formát je vyvíjen společností Sony a vznikal paralelně s ostatními známými formáty, je velice kvalitní co se týče zvuku, ale jeho reproduktorová konfigurace je atypická a taktéž postrádá více efektových kanálů. Pro ozvučení domácího kina systém příliš nevyhovuje.

1.2 Prostorový zvuk založený na objektech

V dnešní době se do podvědomí dostává nový efektivní způsob reprodukce prostorového zvuku. Jedná se o systém, kde zvuk není uchován a přenášen několika kanály, ale pouze jako jedna mono zvuková stopa (popřípadě stereo) s přidanými informacemi o pozici, kde by měl v prostoru zvuk zaznít. To má obrovskou výhodu v objemu datového toku. Zvukový signál je v něm nejčastěji přenášen nekomprimovaný a čistý tj. pouze klasicky převáděn z analogového na digitální za pomoci lineární pulzní kódové modulace (LPCM signál). Se zvukovým signálem se v datovém toku přenáší také zakódovaná metadata popisující daný signál a společně tvoří tzv. jeden zvukový objekt. Tento objekt je posílán do dekodéru, kde se metadata dekodují na potřebné informace a finální směšovač následně rendruje zvuk do reproduktorů. Zakódovaná metadata nesou informace o zvuku a zajišťují, že zvukový signál bude v celém řetězci správně přenášen, zpracován a rendrován podle našich představ [1][3].

¹ Jedná se o číslo popisující počet kanálů. Číslice za tečkou je označení pro kanál přenášející basovou složku.

2 Formáty a standardy založené na objektech

2.1 Formát AURO 3D

Auro 3D je první komerční formát ovládající zvuk nejen založený na kanálech ale také na objektech (ten později dostal název AuroMax [4], ale nenachází využití v ozvučení domácích kin). Jedná se v podstatě o hybrid, který umí pracovat v obou režimech, podle toho, jaký je náš záměr. Byl vyvinut Belgičskou společností Auro Technologies a jako poprvé jej představil vynálezce Wilfried Van Baelen roku 2006 v Paříži. Ještě téhož roku Baelen sestavil tým inženýrů a sestrojil tzv. Auro-Codec a Auro-Matic technologii, jejíž algoritmus dokáže kterýkoliv druh PCM zvukového signálu (mono, stereo nebo prostorový zvuk) namixovat na přirozený zvuk, který je reprodukován jakoukoli Auro 3D reproduktorovou konfigurací nebo pouze jako stereo signál do sluchátek. Při své práci jako první zavedl termín Immersive Sound („pohlcující zvuk“), ten se dnes standardně používá v kontextu s 3D zvukem. Cílem Auro Technologies je docílení co nejlepšího zvuku použitím co nejmenšího počtu kanálů, které jsou uspořádané ve třech vrstvách. První spodní vrstva je pro nás známa oblast, kde se klasicky umísťují reproduktory pro reprodukci prostorového zvuku, a nazývá se Ear-level. Nad ní leží vrstva druhá a nese název Height, ta slouží pro kanály, které jsou umísťovány s elevací 30° (Tím je myšlen úhel svírající zem a pomyslnou přímkou, která protíná reproduktor a posluchače vprostřed místnosti). Poslední vrstva Top je nejvýše a vymezuje prostor, kde jsou reproduktory umísťovány s elevací 90°, většinou jsou montovány na strop sálu. Přehled všech konfigurací je zobrazen v tabulce 2.1. Objektově založený formát zvuku Auro Max však obvykle pracuje s větším množstvím kanálů (20 až 32) a používá rendrovací technologii od Barco Audio Technologies. Baelenova myšlenka spočívá v tom, že při nahrávání zvuku reálného prostředí, kde se nachází několik elementů, by bylo zbytečně složité každý element nahrávat zvlášť jako objekt, proto je lepší využít osvědčené stereofonní nebo jiné techniky, které vedou k danému počtu kanálů, představující celé zvukové pole. Tyto kanály lze pak jednoduše použít jako kanálově založený zvuk nebo také statické objekty. Další nevýhodu vidí v hudebním průmyslu, kde není možné s objekty pracovat tak jako s běžnými stopami a mohlo by to tak tvořit komplikace při finálním masteringu nahrávky. Celá vize Auro Max pohlcujícího zvuku je tedy postavená především na brilantním základu v podobě klasického prostorového zvuku založeném na nahrávce, mixu a distribuci zvukových stop, a ten je dále doplněn statickými nebo dynamickými objekty pro doladění celkového zvuku [5][6].

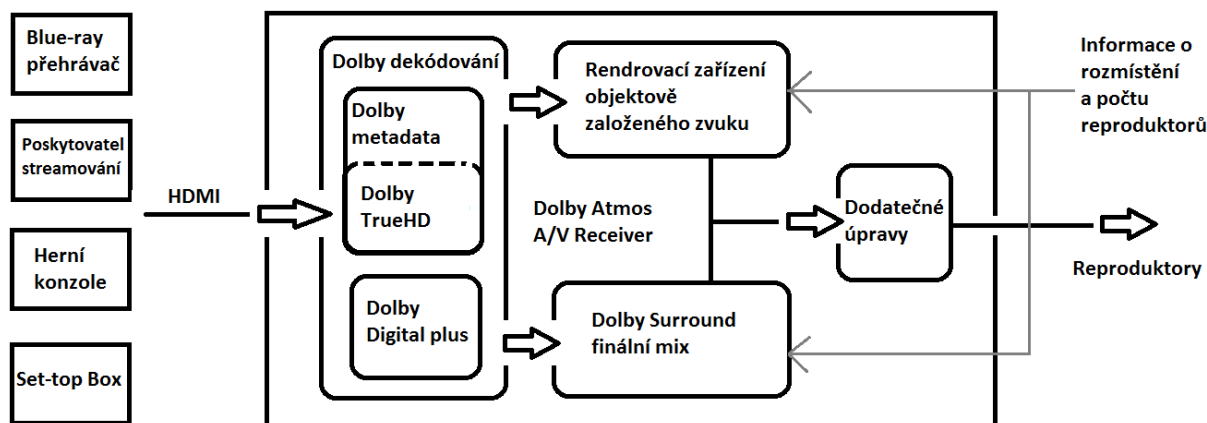
Tabulka 2.1: Přehled konfigurací formátu Auro 3D

Konfigurace	L	R	C	LFE	Ls	Rs	Lb	Rb	HL	HR	HC	T	HLs	HRs
Auro 8.0	x	x			x	x			x	x			x	x
Auro 9.1	x	x	x	x	x	x			x	x			x	x
Auro 10.1	x	x	x	x	x	x			x	x		x	x	x
Auro 11.1	x	x	x	x	x	x			x	x	x	x	x	x
Auro 11.1 (7+4)	x	x	x	x	x	x	x	x	x	x			x	x
Auro 18.1	x	x	x	x	x	x	x	x	x	x	x	x	x	x

2.2 Formát Dolby Atmos

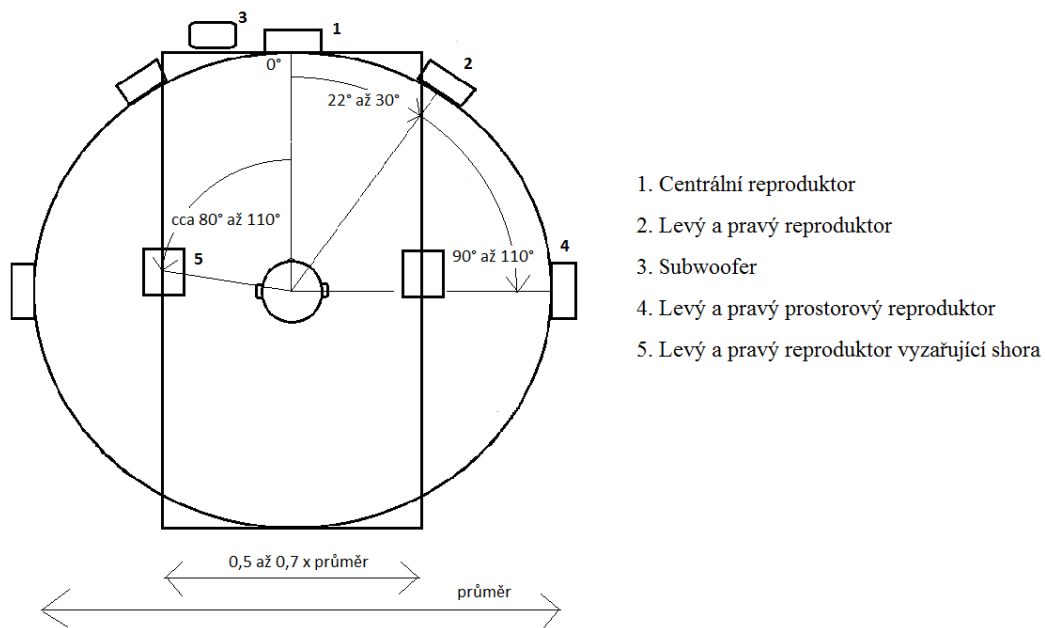
Společnost Dolby se nejprve snažila řešit různé problémy a úskalí kanálově založeného zvuku v kinosálech. Cílem bylo osvobodit filmové tvůrce od limitací, které kanálový zvuk představuje, vyvinutím systému umožňující přesně determinovat, jak a kde by se měl zvuk v prostoru promítat. Vědci z Dolby Laboratories se usnesli názoru, že tak lze docílit pouze pomocí zvuku objektivě orientovaného a vytvořili sadu nástrojů, pomocí které dokážou vytvářet opravdově věrné objektivé stopy. O tom se veřejnost přesvědčila v tu chvíli, když byl roku 2012 formát Dolby Atmos, jak jej vědci nazvali, poprvé použit ve filmu Rebelka. V nadcházejících dvou letech filmové společnosti po celém světě vydali více než 230 filmů, jejichž zvuk byl v ateliérech formován pomocí tohoto formátu.

V kinech se Dolby Atmos spoléhá především na konfiguraci kanálů 9.1 společně s až 118 současně přehrávanými zvuky objektivě založenými, přičemž každý objekt (DCI PCM zvuková stopa mono nebo stereo + příslušné metadata) má svou vlastní stálou lokalizaci v prostoru nebo může být dynamický a svou pozici měnit. To je docíleno spolehlivými metadaty, které přesně definují daný zvukový objekt. Procesor Dolby Atmos používaný v kinech podporuje rendrování objektů až pro 64 reproduktorů. Co se týče domácích kin, tam stačí do A/V receiveru zadat informace o reproduktorech (počet, typ a rozmístění v prostoru) a algoritmus si už poradí s výsledným zvukem. Čím více reproduktorů se v místnosti nachází, tím lépe receiver výsledný zvuk rendruje. Je jen potřeba umisťovat reproduktory symetricky v párech, popřípadě po jednom jako centr. Konfigurace, které Dolby Atmos doporučuje, jsou standardní reproduktorové uspořádání (stereo, 5.1 nebo 7.1), které ještě rozšiřuje o reproduktory vyzařující od stropu. Vznikly tak známé konfigurace 5.1.2, 5.1.4 nebo 7.1.4. Limitem receiveru je uspořádání 24.1.10, to znamená 24 reproduktorů na stěnách, 10 na stropu a 1 subwoofer. Zvuk, který slyšíme shora, nemusí být nutně vyzařován reproduktory montovanými na strop, Dolby Atmos totiž přichází i s jiným řešením v podobě nahoru vyzařujících reproduktorů. Ty jsou umisťovány ve stejné reproduktorové skříně jako reproduktory postranní, jen jsou nasměrovány na strop. Využívá se tedy odrazů od stropu a člověk tak vnímá, že přicházející zvuk slyší shora. A/V receiver dokáže zpracovat jakýkoli kanálově založený formát zvuku a vyrobit z něj 3D zvuk optimální pro uživatelskou konfiguraci [7][8].

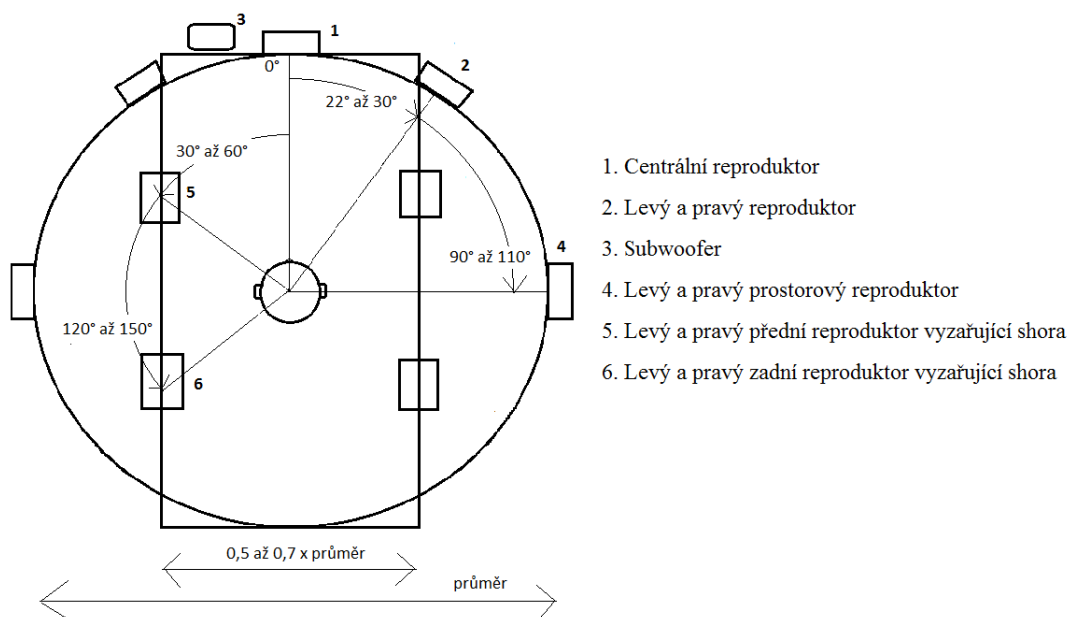


Obrázek 2.1: Pracovní postup Dolby Atmos A/V Receiveru pro domácí kino

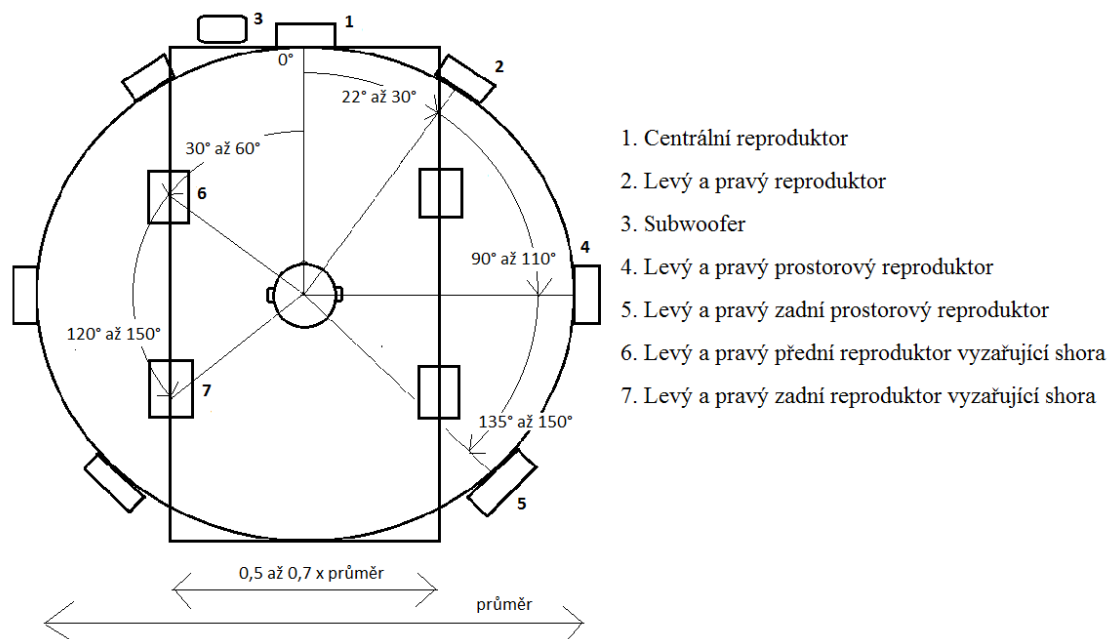
Umístění reproduktorů v prostoru, které Dolby Atmos používá, je založené na doporučení *ITU-R BS.775-3* [9]. Podle něj se reproduktory umístěné vepředu místnosti považují za referenční a všechny reproduktory v místnosti by měly být ekvidistantní od pozice posluchače. Pokud to není umožněno, je třeba zajistit přehrávání zvuku s časovým zpožděním u reproduktorů postavených blíže posluchači a vykompenzovat tak vzdálenost. Všechny reproduktory by dále měly být ve stejné výšce nad zemí a to v oblasti uší, doporučení ITU-R tuto vzdálenost od země k uším specifikuje na 1,2 metru. Zadní prostorové reproduktory mohou být o něco výše než přední, neměly by být však vyšší než 1,5 metru od země. Dolby Atmos dále uvádí, že reproduktory vyzářující shora by bylo vhodné umístit v rozmezí 2,3 až 4,26 metrů vertikálně od posluchače [10].



Obrázek 2.2: Dolby Atmos reproduktorová konfigurace 5.1.2 pro domácí kino



Obrázek 2.3: Dolby Atmos reproduktorová konfigurace 5.1.4 pro domácí kino



Obrázek 2.4: Dolby Atmos reproduktorová konfigurace 7.1.4 pro domácí kino

2.2.1 Nástroje pro editaci zvuku formátu Dolby Atmos

System, který pracuje se zvukem a metadaty formátu Dolby Atmos, využívá DAW editační zařízení Pro Tools a zahrnuje monitorovací aplikaci a panoramovací plug-in od Dolby Atmos. V monitorovací aplikaci se editují data pro kanálově nebo objektově založený zvuk (pouze však zvukové stopy) a posílají se skrz rozhraní HD MADI do zařízení Dolby RMU, což je komponent, který pracuje jako rendrovací a masteringová jednotka. Metadata pro objektově založený zvuk vznikají v panoramovacím plug-inu a posílají se do Dolby RMU přes Ethernet. Existují 4 způsoby jak metadata doručit:

- Jeden nebo více zařízení Pro Tools posílají metadata do Dolby RMU = základní způsob
- Základní způsob, jen je do systému přidán mixážní pult (Avid D-command nebo Avid D-control)
- Základní způsob rozšířený o mixážní pult nebo jiné zařízení s podporujícím softwarem (Avid System 5 nebo iPad obsahující Dolby iPad Panner-Controller). Metadata jdou z RMU do externího zařízení, z tamá upravené zpět do RMU, pak do Pro Tools a z tamá všechny metadata zpět do RMU
- Mixážní pult jako master s možností automatizace, odtud jsou posílány metadata do Dolby RMU a z tamá do Pro Tools v režimu nahrávání [11]

2.3 Formát DTS-X

V Dubnu roku 2015 představila společnost Digital Theatre System formát objektově založeného zvuku, který výrazně konkuruje formátu Dolby Atmos. Nový DTS-X, jak formát vývojáři nazvali, dokáže při reprodukci v kinosálech rendrovat neomezený počet objektů

a nezáleží přitom ani na počtu nebo umístění reproduktorů v sále. Pokud jsou reproduktory v prostoru nesymetricky rozloženy a nelze je umístit jinak, DTS-X za pomoci hodnot změřených kalibračním mikrofonem upraví výsledný zvuk tak, aby byl slyšet jako při správném rozložení. Formát nepodporuje žádný kanálově založený zvuk a je založen pouze na objektech, které jsou virtuálně polohovány každý zvlášť na platformě MDA Creator, jejíž licence je zdarma k dostání a funguje jako instance v editačním prostředí Pro Tools. MDA, což je zkratka pro Multi-Dimensional Audio, je otevřený standard, který převádí klasický prostorový zvuk založený na kanálech na zvuk objektově založený. Takový proces společnost pojmenovala DTS Neural X. Ten dokáže na objektově založený zvuk převést prostorový zvuk, který byl mixován na konfiguraci stereo, 5.1 nebo 7.1. Platforma umožňuje panoramovat objekt až pro 32 reproduktorů, což je limitem pro A/Vreceiver. Pro lepší poslech společnost Digital Theatre System doporučuje použít stropní reproduktory a nezavrhuje ani myšlenku, že by se nahoru vyzařující reproduktory od Dolby Atmos staly s formátem DTS-X kompatibilní. V současné chvíli neexistuje jakýkoli typ reproduktorové konfigurace, kterou by formát podporoval nebo doporučoval. V tomto ohledu je velice otevřený a nechává to na uživateli. DTS-X představuje také funkci, s kterou je možné s dialogy ve filmu pracovat, jako s jedním objektem, který uživatelé mohou řídit nezávisle na ostatním zvuku (nastavení jazyka, hlasitosti, ...). To záleží jen na zvukovém mistru, který dělal mix daného filmu, jestli ponechá objekt uzamčený nebo otevřený a umožní tak uživateli nastavovat jeho parametry [12][13].

2.4 Standard MPEG-H 3D Audio

MPEG-H 3D Audio byl představen skupinou Moving Picture Experts Group roku 2014. Tato skupina byla založena Mezinárodní organizací pro normalizaci (ISO) spolu s Mezinárodní elektrotechnickou komisí (IEC). Celý název pro standard, který je součástí skupiny MPEG-H, obsahující další standardy, je MPEG-H Part 3: 3D Audio. Standard nabízí univerzální a efektivní kódování a rendrování prostorového pohlcujícího zvuku na jakoukoli známou konfiguraci od stereo, přes 5.1 až po 22.2 nezávisle na vstupním formátu kódování. Dokáže pracovat jak s kanálově založeným zvukem tak objektově založeným zvukem a navíc také se zvukem založeným na ambisonii vyššího řádu (HOA - Higher-order Ambisonics), který dnes ještě není příliš rozšířený. Jako základní funkci umožňuje normalizovat hlasitost podle standardizovaného doporučení ITU-R BS.1770-3, které vychází z EBU R128. Pro MPEG-H 3D Audio bylo navrženo několik technologií kódování, které byly testovány, načež ta nejlepší z nich se uplatňuje, dosahuje datového toku 512kbit/s pro konfiguraci 5.1 a nejvýše 1.2 Mbit/s pro 22.2 konfiguraci. Objektově založený zvuk MPEG-H 3D Audio je velmi interaktivní a umožňuje uživateli npř. měnit pozici zvukového zdroje v prostoru nebo měnit jazyk daného programu. Dále při přenosu sportovního utkání umožňuje výběr dialogu (komentátora nebo týmu), zesílení či ztlumení různých objektů (zvuk z tribuny na té straně týmu, které fandím) a další funkce. Objektová metadata jsou navržena tak, aby podporovala funkce modelu ADM (pojednávám kapitola 2.4.1), což se očekává také pro jiné formáty 3D zvuku, aby fungovala jistá kompatibilita. V únoru 2015 na 111. zasedání skupiny MPEG proběhla konečná etapa procesu, při kterém se formát MPEG-H 3D Audio ustanovil jako mezinárodní standard. Bližší specifikace standardu především se zaměřením na strukturu

datového toku a proces dekódování prostorového zvuku uvádí kapitola Specifikace standardu MPEG-H 3D Audio.

2.4.1 Otevřený standard Audio Definition Model

Delší dobu se postrádal jednotný model metadat, který by dostatečně popsal jejich funkci a uvedl celistvý zvukový formát. V roce 2014 Evropská vysílací unie vyvinula Audio Definition Model (ADM) [14], což je model, který nám dává kompletní technický popis zvuku v rámci souboru a popis umožňující jeho správné rendrování. Hlavním požadavkem je aby metadata plně popisovala zvukový signál. To umožní distribuci signálu, ať už je jakéhokoliv druhu. ADM využívá souborový formát Broadcast Wave Format (BWF), jedná se v podstatě o stejný formát jako WAVE (WAV), jen rozšířený o příslušená metadata, a dále pak využívá značkovací jazyk XML jako soubor pravidel pro kódování metadat ve formátu. Tento jazyk je čitelný jak člověkem, tak počítačem. Model je rozdělen do dvou sekcí, obsahová část popisuje samotný zvukový signál, část formální zase metadata a jejich kódování. Většina z formátových elementů je definována ještě před příchodem vzorků zvukového signálu, zatímco obsahové elementy hrají svou úlohu až po generování zvukového signálu. Přestože je model popsán pro formát BMF, lze jej použít jako obecný, aby jej mohli využívat i jiné formáty. Pro zajištění kompatibility napříč všech systémů se stal ADM otevřeným standardem [15].

2.5 Porovnání současných formátů a standardů

V následující tabulce je zobrazeno porovnání aktuálních formátů z hlediska jejich stáří, podpory konfigurací 5.1.2, 5.1.4 nebo 7.1.4 a maximálního počtu kanálů, se kterým A/V receiver dokáže pracovat. Dále pak porovnává kompatibilitu formátů s otevřeným standardem ADM pro převod na jiné formáty a také parametry² popisující vlastní bitový proud.

Tabulka 2.2: Porovnání formátů objektově založeného zvuku

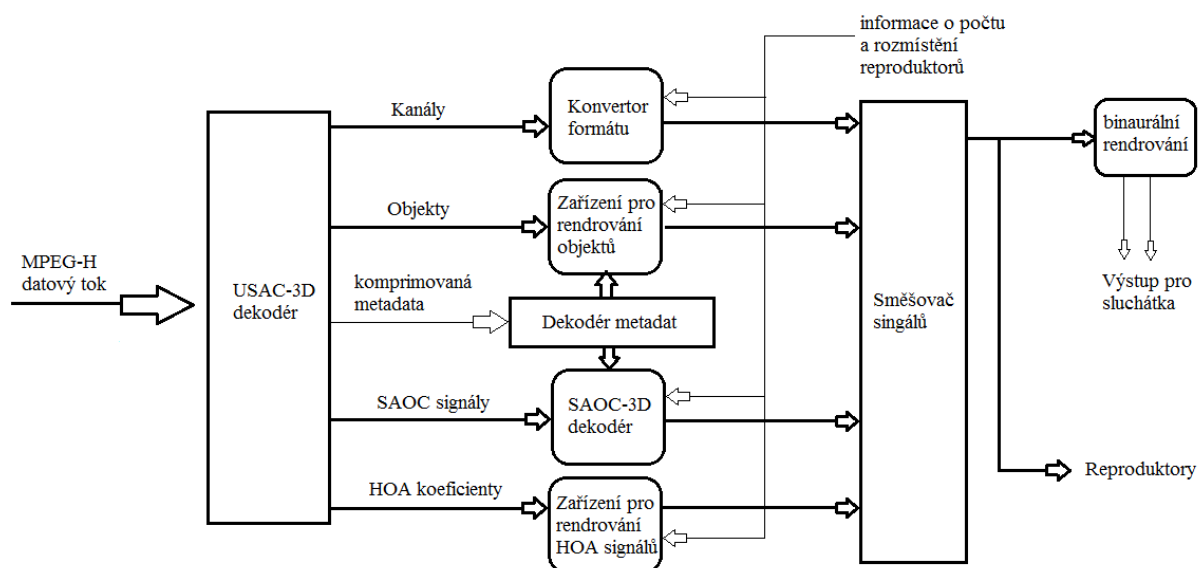
	Auro 3D	Dolby Atmos	DTS-X	MPEG-H
představení	květen 2006	duben 2012	duben 2015	červenec 2014
vydání	leden 2012	červen 2012	červenec 2015	březen 2015
konfigurace 5.1.2	ne	ano	ano	ano
konfigurace 5.1.4	ne	ano	ano	ano
konfigurace 7.1.4	ne	ano	ano	ano
max. počet kanálů	18.1	24.1.10	32	22.2
kompat. s ADM	ano	nezveřejněno	nezveřejněno	ano
bitový proud	1509 kbit/s (6)	6612 kbit/s (8)	1509 kbit/s (6)	512 kbit/s (6)
délka zv. rámce	512 vzorků	40 vzorků	512 vzorků	1024 vzorků
vz. frekvence	44,1 až 384 kHz	48, 96 a 192 kHz	48 kHz	48 kHz
bitová hloubka	16 a 24 bitů	16,20 a 24 bitů	24 bitů	24 bitů
komprese	ztrátová	bezztrátová	ztrátová	nezveřejněno

² Zjištěno při analýze zvuk. vzorků v přehrávacím programu Daum Potplayer. Pro standard MPEG-H 3D Audio vzorek nenalezen, informace byly vyčteny z [16]. Závorka u parametru bitový proud značí počet kanálů.

3 Specifikace standardu MPEG-H 3D Audio

3.1 Princip dekódování prostorového zvuku

Celý princip dekódování je popsán v mezinárodním standardu ISO/IEC DIS 23008-3 [17] vydaného roku 2014. MPEG-H datový tok, který nese všechny potřebné informace, musí nejdříve projít celým řetězcem v dekodéru, abychom mohli slyšet výsledný zvuk. Jako první krok se za použití nástrojů kódovacího systému USAC-3D zjistí jaký formát je v datovém toku obsažen a jakou kompresi používá. Následně se dekóduje zpět na čistý PCM signál (soubor .WAV) popřípadě na další doprovázející data a roztrídí se podle typu signálu. Signál může být typu Kanály, Objekty, SAOC nebo HOA. Každý z nich je odlišný a s každým se následně jinak pracuje, všechny však mají stejný cíl a to projít směšovačem signálů na konci řetězce aby mohli být následně reprodukovány. USAC je zkratka pro Unified Speech and Audio Coding [18], tento formát byl vyvinut skupinou MPEG a používá se pro kódování hudby nebo mluveného slova s využitím velmi nízké přenosové rychlosti (do 64 kbit/s). Ve svém vlastnictví má mimo jiné i známý standard AAC, který pracuje se ztrátovou kompresí digitálního zvuku a je nástupce formátu mp3. Pro nové požadavky v kontextu s 3D zvukem byl USAC formát pouze rozšířen o nové nástroje, které jsou založeny na vjemových účincích 3D reprodukce, tím se zvýšila účinnost kódování. Dále obsahuje mechanismus, který dokáže v komprimovaném datovém toku přečíst 3D obsah a následně jej roztrdit na jednotlivé typy signálu, nebo také nové aspekty, které umožňují výhodné adaptivní streamování přes IP síť. Vznikl tak rozšířený formát, který nese název USAC-3D.



Obrázek 3.1: Blokové schéma dekodéru MPEG-H 3D Audio

3.1.1 Signály typu Kanály

Nejjednodušší způsob, jak docílit prostorového zvuku, je pomocí několika zvukových stop, nazývajících se Kanály, přičemž každý signál je designován na jeden reproduktor, který je umístěn v předem známé pozici od posluchače. Jedná se tedy o klasický kanálově založený

zvuk. Konfigurace reproduktorů v domácím prostředí pro standard MPEG-H může být v podstatě jakákoliv, která se dnes používá (2.0, 5.1, 7.1, 9.1, 22.2). Pro doporučení *ITU-R BS.775-3* se stala konfigurace 5.1 nejběžnější verze, která se používá. Pro 3D zvuk, který využívá navíc shora vyznačující reproduktory, však žádné doporučení neexistuje. Při tvorbě tzv. mixu zvukovým mistrem je kanálově založený zvuk vázán na předem danou konfiguraci. Pokud přehráváme zvuk na konfiguraci, která obsahuje menší počet reproduktorů, než na jaký se prováděl daný mix je potřeba provést tzv. downmixing, pokud konfigurace obsahuje více reproduktorů, provádí se tzv. upmixing. Ani jedna úprava se však neobejde bez ztráty na kvalitě výsledného zvuku. Signály typu Kanály, které dekodér USAC-3D rozpozná, putují do konvertoru formátů, který má za úkol provést upmix nebo downmix, na požadovaný počet reproduktorů, které jsme do dekodéru zadali. Takto konvertované kanály jdou do směšovače, kde pro každý kanál je přiřazen svůj reproduktor umístěný v prostoru.

3.1.2 Signály typu Objekty

Objekty obsahují pouze jednu stopu a při mixování zvukovým mistrem jsou nezávislé na reproduktorové konfiguraci, proto se objektově založený zvuk stává mnohem víc kompatibilní než kanálový a umožňuje také uživateli interaktivou se zvukem. Příslušná metadata popisují, kde a jak by signál měl v prostoru zaznít. Objekt může být statický a mít stálou pozici v prostoru nebo dynamický a svou pozici měnit. To popisují objektová metadata, která putují komprimovaná do dekodéru metadat. Komprese je prováděna buď s využitím lineární interpolace, nebo s tzv. malým zpožděním. V dekodéru se metadata dekodují na potřebné informace a zařízení pro rendrování objektů dále použije tyto informace, aby správně panoramoval příslušnou zvukovou stopu. Rendrovací zařízení vytvoří virtuální polohu objektu pomocí techniky VBAP (Vector-based amplitude panning – Vektorově založené amplitudové panoramování) a namíchá signál na daný počet reproduktorů tak, aby při jeho přehrávání posluchač slyšel zvuk přicházet z toho místa, které bylo popsáno v zakódovaných metadatech. Jak jsou metadata přenášena a kódována pojednává kapitola 3.3.

3.1.3 Signály typu Spatial Audio Object Coding

Jedná se taktéž o signály objektového typu, ovšem s jinou technikou kódování a následného rendrování. Technologie SAOC [23] je založená na přenosu objektových signálů které byly v kodéru převedeny na downmix jednoho mono nebo stereu signálu a na další příslušná metadata, která se přenáší separátně do dekodéru. V dekodéru se downmix opět dekóduje na původní objekty a za pomoci metadat probíhá jejich rendrování na potřebný počet kanálů, o kterém jsme rendrovací zařízení nechali informovat. Rendrovací zařízení může být součástí dekodéru, ale taky nemusí. Metadata separovaná od objektových signálů se přenáší dedikovaně v odlišném bitovém proudu. Jejich množství roste lineárně s počtem objektů v signálu, přesto je množství bitů potřebné pro dekódování těchto dat zanedbatelné ve srovnání s datovým tokem kódovaného downmix signálu. Dekodér SAOC-3D obsahuje vylepšení a nové nástroje pro kódování objektového signálu, jako jsou: možnost provádět downmixing na větší počet kanálů než stereo nebo nový systém rendrování na místo MPEG Surround rendrovací techniky, upravený pro 3D poslech. Systém SAOC-3D je přínosem pro telekonferenční místnosti, kde většinou hovoří více lidí najednou. Nalézá zde uplatnění, když

se každý hlas přenáší jako objekt a je rendrován tak, aby každý z nich posluchač slyšel přicházet z jiného směru. Dále se využívá v herním nebo hudebním průmyslu.

3.1.4 Signály typu Higher Order Ambisonics

Alternativní formou jsou signály, které jsou kódovány systémem Higher Order Ambisonics. Ten staví na rozsáhlé matematické myšlence, která považuje zvukovou scénu jako celé zvukové pole. Provádí expanzi vlnového pole do sférických harmonických vln, které dané akustické vlnové pole determinují, využívá tak znalosti zvukové pole, které se v okolí posluchače vyskytuje. Původní ambisonie je konfigurována pro 4 kanály a nazývá se prvního řádu. Pomocí těchto čtyř kanálů by mělo být možné matematicky simulovat zvukové pole, které představuje jakýkoli počet vyzařujících reproduktorů v prostoru. Při nahrávání zvukového pole je jeden kanál použit pro signál braný z všesměrového mikrofonu, další 3 mikrofony jsou osmičkového charakteru a každý z nich nahrává signál, se kterým se pracuje jako s jedním rozměrem 3D prostoru. Pro rozšíření sweet spotu se přidávají další kanály do systému a vzniká tak ambisonie vyššího řádu (HOA). Matematický model vypočítá časově proměnné koeficienty, které popisují vlnové pole. Ty jsou potřebné pro finální rendrování zvuku poté, co jsou podrobeny syntéze. Systém Hogher Order Ambisonics má potenciál pro velice kvalitní reprodukci prostorového zvuku, ale dnes ještě není příliš znám a podporován.

3.2 Struktura metadat v bitovém proudu

Transportní mechanismus v bitovém proud MPEG-H využívá tzv. pakety. Veškerá data, ať už kódovaný audio signál nebo konfigurační metadata, jsou uloženy do jednotlivých paketů. Důležité je zajistit, aby byly oba typy dat dobře synchronizované. Metadata, která nesou veškeré informace o obsahu v bitovém proudu, mají svůj pevný řád a popisují obsah, který je strukturován v následujících paketech:

Element: Můžeme jej označit jako nejmenší článek. Jedná se o zvukové stopy nebo zvukové události (letadlo, řeč, kytara), které jsou obsaženy ve zvukové scéně. Mohou být typu Kanál, Objekt, SAOC nebo HOA. Metadata přiřazují každému elementu svoje ID a ‚vlajku‘, která definuje, jestli je element tzv. screen-related. Pokud se jedná o element typu Objekt nebo SAOC, objektová metadata dále definují jejich pozici a dynamický charakter v prostoru (azimut, elevace, ...).

Skupina: Všechny elementy patří do určité skupiny, ta může obsahovat pouze elementy stejného typu. Nemůže obsahovat jinou skupinu. Metadata popíší, kolik skupin se ve zvukové scéně bude nacházet a každé skupině přiřadí ID. Dále definují prioritu skupiny, její typ (odvozený od typu jejich elementů) a seznam ID elementů ve skupině.

Přepínací skupina: Obsahuje pouze nějaké skupiny, což má za následek to, že se v ní nachází seskupení elementů, které se vzájemně vylučují. Používají se pro zajištění toho, že právě jenom jedna skupina, která se v ní nachází, je aktivní. To

uživateli umožňuje přepínání mezi npř. dialogy, které jsou v různých jazycích, přičemž by nebylo rozumné přehrávat několik dialogů najednou. Metadata přiřazují přepínací skupině ID, dále popisují počet skupin, který se v ní nachází současně s jejich jednotlivými ID a jako poslední informaci popisují ID skupiny, která je v přepínací skupině jako defaultní.

Zvuková scéna: Jedná se o největší článek zahrnující všechny data. Obsahuje přepínací skupiny, skupiny a elementy. Metadata popisují počet skupin a přepínacích skupin, které se ve zvukové scéně nachází.

3.3 Přenos a dekódování objektových metadat

Objektová metadata nesou informace potřebné pro správné rendrování elementů ve zvukové scéně. Azimut, elevace a poloměr udávají polohu elementu v prostoru a gain udává energii, se kterou je element reprodukován.

Tabulka 3.1: Přehled základních informací zakódovaných v objektových metadatech

informace	jednotka	rozsah hodnot
azimut	° (stupeň)	- 180 až 180
elevace	° (stupeň)	- 90 až 90
poloměr	<i>m</i> (metr)	0 až 1000
gain	- (lineární nezáporné číslo)	0 až 1000

Poté, co SAOC-3D dekodér rozpozná v bitovém proudu signál typu Objekt. Pošle příslušná objektová metadata do dekodéru metadat, kde se rozpozná, jak jsou metadata komprimovaná a podle toho se volí jejich dekódování. Statická objektová metadata se do dekodéru přenesou pouze jednou a to na začátku každého elementu. Dynamická metadata se přenáší několikrát během přenosu elementu a zapříčiňují tak, že výsledný zdroj zvuku mění polohu v čase. Kvůli dynamickým změnám signálu je třeba metadata posílat velmi často tj. každých 1024 zvukových vzorků. Protože objektová metadata se přenáší ve zvukové scéně před každým novým elementem a pokud jde o dynamická, tak také mnohokrát během jednoho elementu, zavádí se jejich komprese, která zaručí menší objem datového toku.

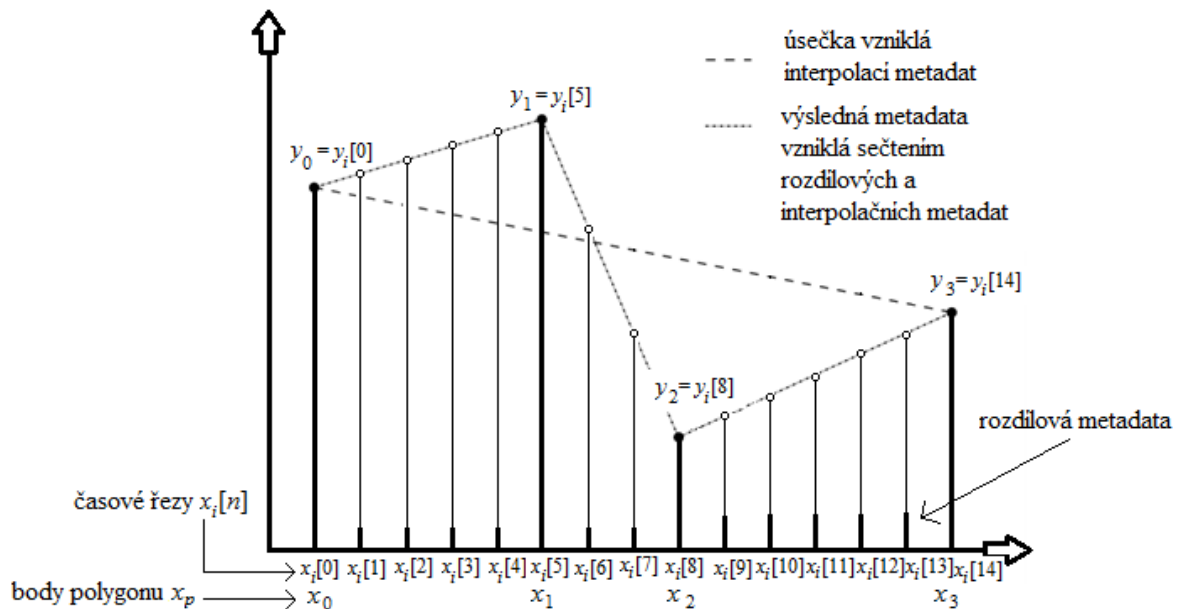
3.3.1 Dekódování metadat komprimovaných lineární interpolací

Informace pro polohu objektu jsou běžně popsány v signále s diskretním časem $y[n]$. Indexy n diskretního času si však nesmíme plést se vzorky zvukového signálu, které jsou v toku obsaženy s mnohem vyšší frekvencí. Proto se indexy n pro hodnoty metadat nazývají časové řezy. Myšlenka spočívá v rozdělení signálu $y[n]$ na segmenty určité délky (s určitým počtem časových řezů), které jsou aproximovány pomocí nejméně dvou polygonů (každý polygon je zároveň časový řez a jeho pořadí se značí indexem p). Tyto segmenty nazýváme okna. Každé přenesené okno obsahuje jeden nový polygon a jeden koncový. Pro další následující okno je vždy koncový polygon předcházejícího okna novým počátečním polygonem stávajícího okna. Každé okno obsahuje několik časových řezů, jejichž perioda

odpovídá délce R jednoho zvukového rámce obsahujícího PCM vzorky. Délka zvukového rámce je obvykle určována podle velikosti jednoho snímku ve video signálu. Všechny hodnoty časových řezů mezi mezními polygony obsažené v okně jsou pro každou informaci (azimut, elevace, poloměr, gain) počítány pomocí lineární interpolace. Kompresi spočívá v přenosu tzv. přenosových metadat, která jsou výsledkem rozdílu normálních (žádaných) hodnot a interpolačních hodnot, což jsou okamžité hodnoty ležící na přímkce protínající první a poslední polygon jednoho okna. Pro výpočet lineární interpolace, potažmo její aktuální hodnoty $y_i[n]$ v příslušném časovém oknu $x_i[n]$, použijeme rovnici:

$$y_i[n] = \left(1 - \frac{x_i[n] - x_p}{x_{p+1} - x_p}\right)y_p + \left(\frac{x_i[n] - x_p}{x_{p+1} - x_p}\right)y_{p+1}, \quad (3.1)$$

kde index p značí pořadí polygonu a index i pořadí okna. Pokud chceme zjistit hodnoty diferenčních metadat v daných časových řezech, musíme odečíst hodnoty normálních zakódovaných metadat od hodnot na interpolační křivce.



Obrázek 3.2: Proces dekódování metadat lineární interpolací (na obrázku jeden rámeček)

3.3.2 Dekódování metadat komprimovaných s malým zpožděním

Jedná se o kompresy založené na tzv. rozdílové pulsní kódové modulaci (DPCM), která je navíc modifikovaná. Dekodér pracuje s malým zpožděním. Nejprve zjistí, jestli jsou metadata v bitovém proudu obsažena absolutní nebo diferenční. Z absolutních metadat čte dekodér informace přímo, v případě diferenčních musí nejprve data převést na absolutní pomocí rovnice:

$$y[n] = y[n - 1] + d[n], \quad (3.2)$$

kdy y je výstupní hodnota, d je hodnota rozdílových metadat a n značí příslušný časový řez.

4 Virtuální polohování zvukového zdroje

V současné chvíli existuje několik způsobů virtuálního polohování zvukového zdroje, mezi nejpoužívanější patří panoramování technikou VBAP, DBAP (Distance-based amplitude panning - Amplitudové panoramování založené na vzdálenosti) nebo pomocí ambisonie.

4.1 Vektorově založené amplitudové panoramování

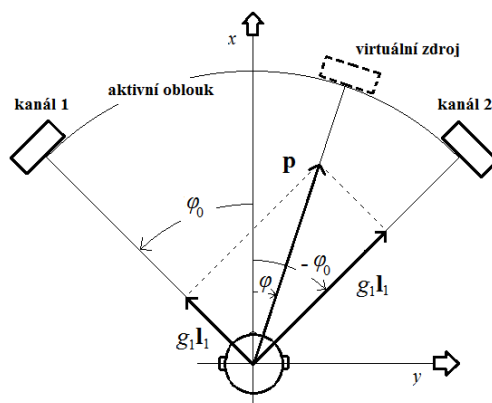
Vektorově založené amplitudové panoramování (VBAP) [24] je metoda polohování virtuálního zdroje pomocí znalosti vektorů směřujících od posluchače k reproduktorům. Tyto vektory spolu interagují (sčítají se a odčítají) a výsledkem je jeden výsledný vektor, který směřuje k virtuálnímu zdroji zvuku. Takový přístup umožňuje použití jakéhokoli konečného počtu reproduktorů nacházejícího se v prostoru kolem posluchače. Tyto reproduktory mohou být polohovány ve vertikální i horizontální rovině. Podmínkou je, aby směřovali k posluchači a byly od něj ekvidistantní. S větším počtem reproduktorů roste přesnost polohy výsledného panoramovaného zvuku.

4.1.1 Dvourozměrné panoramování

Pro 2 reproduktory, které vysílají koherentní akustický signál s rozdílnou amplitudou, posluchač slyší výsledný signál přicházející pouze z jednoho (virtuálního) zdroje zvuku. Pozice virtuálního zdroje je určena polohou reproduktorů v prostoru a rozdílnou intenzitou (velikostí amplitudy) jejich signálů, přicházejících k posluchači. Amplitudy signálů jsou ovládány zesilovacími faktory g_1 a g_2 . Aby nedocházelo ke změně hlasitosti virtuálního zdroje při různých změnách jeho polohy, je vyžadováno, aby platila rovnice:

$$g_1^2 + g_2^2 = C \quad (4.1)$$

přičemž $C > 0$ značí konstantní hodnotu a můžeme ji označit jako hlasitost. Čím větší je parametr C , tím blíže k posluchači se bude virtuální zdroj přibližovat. Pokud nebudeme hlasitost nijak regulovat, bude virtuální zdroj ležet někde mezi oběma reproduktory na kružnici je protínající. Tato křivka se nazývá aktivní oblouk.



Obrázek 4.1: Vektorově založené amplitudové panoramování pro 2 reproduktory

Směr přicházejícího zvuku virtuální zdroje se řídí stereofonními zákony popsané reformulovanou sinovou větou:

$$\frac{\sin \varphi}{\sin \varphi_0} = \frac{g_1 - g_2}{g_1 + g_2} \quad (4.2)$$

přičemž $0^\circ < \varphi_0 < 90^\circ$, $-\varphi_0 \leq \varphi \leq \varphi_0$ a $g_1, g_2 \in [0,1]$. Veličina φ_0 udává úhel mezi osou x v horizontální rovině kolmou na posluchače a přímkou protínající reproduktor a posluchače zatímco veličina φ udává úhel mezi jistou osou x a přímkou protínající virtuální zdroj a posluchače. Rovnice (4.2) platí jen v případě, že je posluchač otočený čelem k reproduktorům. Pokud otáčí hlavou, aby sledoval dynamický virtuální zdroj, je třeba použít tangentskou větu:

$$\frac{\tan \varphi}{\tan \varphi_0} = \frac{g_1 - g_2}{g_1 + g_2} \quad (4.3)$$

Za konstantní hlasitosti C můžou být zesilovací faktory g_1 a g_2 vypočítány použitím rovnice (4.1) spolu s rovnicí (4.2) nebo (4.3). Ty lze však použít pouze pro signály s kmitočtem nižším než 600 Hz.

V dvojrozměrném panoramování můžeme brát stereofonní rozmístění reproduktorů, jako dvourozměrnou vektorovou základnu. Ta je popsána jednotkovými vektory

$$\mathbf{l}_1 = \{l_{11} \quad l_{12}\}^T, \quad (4.4)$$

$$\mathbf{l}_2 = \{l_{21} \quad l_{22}\}^T, \quad (4.5)$$

kteří směřují od posluchače k reproduktorům. Index T značí, že matice je transponovaná. Jednotkový vektor \mathbf{p} směřující k virtuálnímu zdroji je lineární kombinací vektorů \mathbf{l}_1 a \mathbf{l}_2 :

$$\mathbf{p} = \{p_1 \quad p_2\}^T, \quad (4.6)$$

$$\mathbf{p} = g_1 \mathbf{l}_1 + g_2 \mathbf{l}_2. \quad (4.7)$$

Protože víme, že zesilovací činitele g_1 a g_2 jsou skalární veličiny, můžeme rovnici upravit do maticového tvaru:

$$\mathbf{p}^T = \mathbf{g} \mathbf{L}_{12}, \quad (4.8)$$

$$\mathbf{g} = \{g_1 \quad g_2\}, \quad (4.9)$$

$$\mathbf{L}_{12} = \{\mathbf{l}_1 \quad \mathbf{l}_2\}^T. \quad (4.10)$$

Rovnici (4.8) můžeme upravit a dále do ní dosadit a získáme tak vektor \mathbf{g} , který obsahuje potřebné zesilovací činitele g_1 a g_2 :

$$\mathbf{g} = \mathbf{p}^T \mathbf{L}_{12}^{-1} = \{p_1 \quad p_2\} \begin{Bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{Bmatrix}^{-1}. \quad (4.11)$$

Rovnice má řešení, pokud existuje \mathbf{L}_{12}^{-1} . To nastane, pokud $\varphi_0 \neq 0^\circ \wedge \varphi_0 \neq 90^\circ$, pro takové krajní hodnoty bychom prováděli pouze jednorozměrné panoramování. Zesilovací činitele g_1 a g_2 vypočítané rovnicí (4.11) korespondují s tangentskou rovnicí (4.3). Pokud je základna ortogonální, to znamená, že $\varphi_0 = 45^\circ$, zesilovací činitele g_1 a g_2 jsou shodné s těmi vypočítanými systémem ambisonie. Pokud vektory \mathbf{l}_1 a \mathbf{l}_2 nespírají 45° , je třeba zesilovací činitele normalizovat, aby korespondovaly s rovnicí 25, pomocí následující rovnice:

$$\mathbf{g}^{\text{norm}} = \frac{\sqrt{C}\mathbf{g}}{\sqrt{g_1^2 + g_2^2}}. \quad (4.12)$$

C značí hlasitost, na kterou chceme normalizovat. Pro vícekanálovou reprodukci v horizontální rovině systém panoramuje v jednu chvíli pouze pro jednu stereofonní vektorovou základnu. Každá z nich tvoří svůj vlastní aktivní oblouk, na kterém se promítá virtuální poloha zdroje. Dohromady tvoří kružnici kolem posluchače a při příchodu zvukového signálu systém rozhodne, která stereofonní základna bude použita. Zesilovací faktory ostatních základen se snaží udržet si nulovou hodnotu. Při dynamickém zvuku může virtuální poloha zdroje přecházet z jednoho oblouku na druhý (většinou sousední). Je důležité, aby zesilovací činitel té základny, která právě ukončila svoji působnost, okamžitě klesl na nulu. To proto, aby již dále neovlivnila polohu zdroje na novém aktivním oblouku. Aktuální reproduktorový pár je vybírán podle nenormalizovaných zesilovacích faktorů z rovnice (4.11), podle toho, jestli neobsahují záporné hodnoty. Pokud ano, systém je srovná na nulu a tím se stává základna neaktivní.

4.1.2 Trojrozměrné panoramování

Tři reproduktory nyní tvoří vektorovou základnu a v matici tím tedy přibývá další jednotkový vektor \mathbf{l}_3 . Virtuální zdroj zvuku se zobrazuje na tzv. aktivním trojúhelníku. Jedná se v podstatě o stejný systém jako v dvojrozměrném panoramování, v každém jednotkovém vektoru je pouze obsažen nový parametr značící třetí rozměr. Zesilovací faktory reproduktorů musí při sečtení opět udržovat stejnou hlasitost:

$$g_1^2 + g_2^2 + g_3^2 = C. \quad (4.13)$$

Pro jejich vypočtení použijeme rovnici:

$$\mathbf{g} = \mathbf{p}^T \mathbf{L}_{123}^{-1} = \{p_1 \quad p_2 \quad p_3\} \begin{Bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{Bmatrix}^{-1}. \quad (4.14)$$

Normalizaci vektorů provedeme následovně:

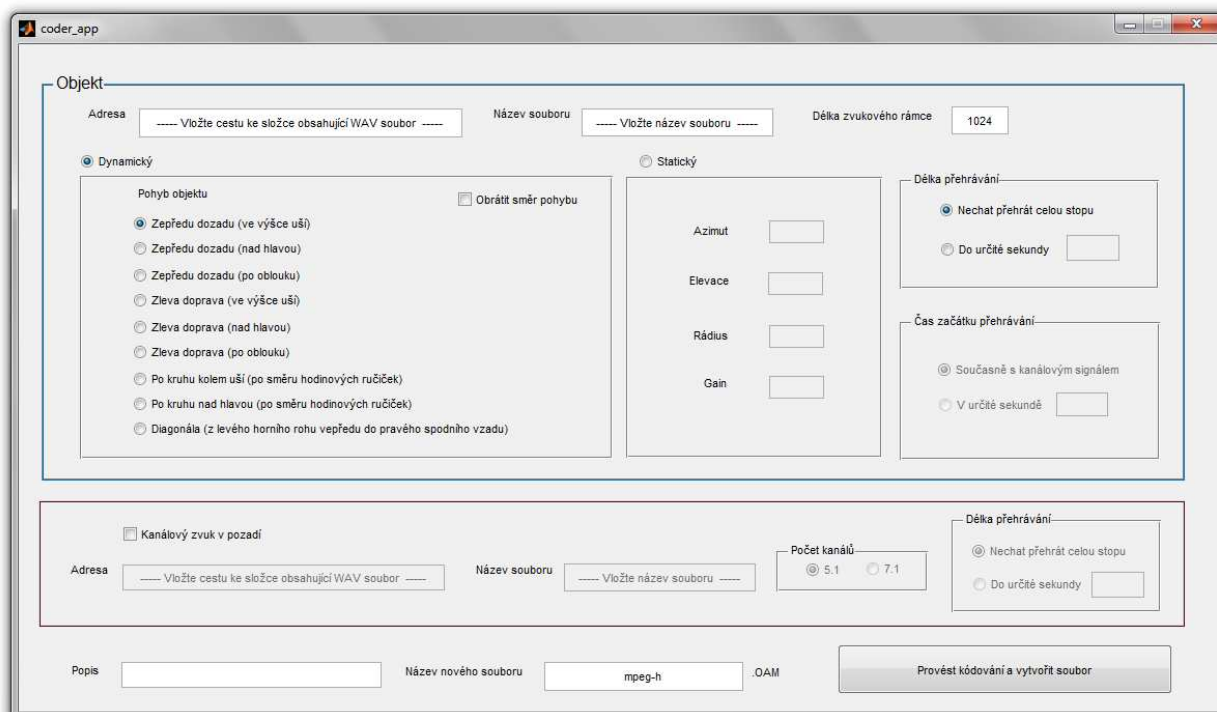
$$\mathbf{g}^{\text{norm}} = \frac{\sqrt{C}\mathbf{g}}{\sqrt{g_1^2 + g_2^2 + g_3^2}}. \quad (4.15)$$

5 Algoritmus interpretující vstupní metadata standardu MPEG-H sestavený v Matlabu

Algoritmus byl programován v grafickém uživatelském prostředí programovacího jazyku Matlab. Podle nastudovaných informací o daném standardu byl sestaven řetězec, na jehož začátku uživatel na straně kodéru nahraje audio soubor a dodá příslušné informace. Vytvoří se soubor, ve kterém jsou informace zakódovány formou metadat. Po nahrání souboru do dekodéru se informace získají zpět a na konci řetězce se vytvoří zvukový soubor připravený k přehrání.

5.1 Kodér standardu MPEG-H

Kodér je graficky zpracován tak, aby byl variabilní a uživatel si tak mohl vybrat, zda chce objektový zvuk přehrávat staticky nebo dynamicky (měnit polohu objektu v prostoru), celý nebo do určité sekundy a s jakou délkou zvukového rámce. Dále je možnost zakódovat kanálový zvuk do pozadí a nastavit stejně jako u objektu délku přehrávání. Zde je nutné zvolit počet kanálů, pro který byl zvuk mixován, aby s danou konfigurací mohl program dále pracovat. Pokud zvolíme kanálový zvuk v pozadí, máme možnost vybrat, jestli chceme přehrávat objekt současně s kanálovým zvukem nebo až od určité sekundy.



Obrázek 5.1: Graficky zpracovaný kodér standardu MPEG-H

5.1.1 Vytvoření souboru .OAM a zápis do něj

Vytvoření souboru je provedeno funkcí `fopen`. Následným příkazem `fwrite` se provádí zápis proměnných, kterým se přiřazuje text jako znaky ze sady ASCII s kódem 10 a číslice nebo vektor v Little-endian. Každý zápis probíhá postupně pod příslušným ID, proto záleží na jeho pořadí. Jakmile je vše potřebné zapsáno do souboru, zavře se příkazem `fclose`.

5.1.2 Hlavička souboru .OAM

Hlavička nese obecné informace o souboru a nachází se na jeho začátku. Řídí se podle standardu MPEG-H a do souboru se zapisuje v daném pořadí následovně:

1. *format_id_string*
 - ID formátu, text, pevně nastaveno na `testDanielKováč`
2. *format_version*
 - verze formátu, číslice, pevně nastaveno na 1
3. *number_of_channel_signals*
 - počet kanálových signálů, číslice, při kanálovém zvuku v pozadí rovna 1
4. *number_of_object_signals*
 - počet objektových signálů, číslice, pevně nastaveno na 1
5. *description_string*
 - popis formátu, text, zadáno v grafickém uživatelském rozhraní
6. *channel_file_name*
 - název kanálového souboru, text, zadáno v grafickém uživatelském rozhraní
7. *object_description*
 - název objektového souboru, text, zadáno v grafickém uživatelském rozhraní.

Informace vyskytující se v textu s tímto fontem kurzívou znázorňují zároveň také proměnné, s kterými se pracuje v programu.

5.1.3 Čtení a úprava vstupních zvukových dat

Aby kodér a následně i dekodér pracoval správně, je potřeba načítat zvukové soubory se vzorkovací frekvencí 48 kHz.

Data kanálového zvuku

Pokud uživatel zadal kanálový zvuk v pozadí, nejprve se do souboru zapíše příslušný počet kanálů. Poté se načítají vzorky zvukového souboru pomocí funkce `audioread`. V případě, že chceme, aby zvuk zazněl jen do určité sekundy, načítá se pouze určitý počet zvukových vzorků. Počet vzorků n je roven součinu času t , ve kterém skončí přehrávání, a vzorkovací frekvence f_{vz} .

$$n = t f_{vz} . \quad (5.1)$$

Do souboru se zapíše počet vzorků, následně data (zvukové vzorky) pro každý kanál zvlášť a nakonec vzorkovací frekvence.

Data objektového zvuku

Postup čtení a zápisu je stejný jako u dat kanálového zvuku. Navíc zde však probíhají výpočty týkající se zvukového rámce. Jeden zvukový rámeček zahrnuje počet vzorků, po kterém se do vektoru ukládají metadata. V případě dynamického objektu budou tato metadata vždy s rozdílnými hodnotami. Panoramovací jednotka v dekodéru pak pro každý rámeček vypočítá jinou polohu v prostoru. Čím kratší zvukový rámeček bude, tím plynuleji by se měl posluchači při reprodukci jevit výsledný pohyb objektu. Příliš krátký rámeček ale bude zbytečně zatěžovat dekodér a prodlužovat celkovou dobu panoramování. Pokud uživatel nezadá jinak, délka zvukového rámce je nastavena na standardních 1024 vzorků. Protože kodér a následně pak i dekodér pracuje pokaždé se stejnou délkou zvukového rámce, je třeba zajistit, aby poslední rámeček obsahoval stejný počet vzorků, jako rámce ostatní. Proto se po rozdělení všech vzorků do rámečků poslední z nich doplní nulami o chybějící počet vzorků tak, aby byl naplněn celý zvukový rámeček. Počet nul p je dán rovnicí:

$$p = R - \text{rem}\left(\frac{n}{R}\right), \quad (5.2)$$

příčemž R je délka jednoho zvukového rámce a rem je funkce, která vrací zbytek po celočíselném dělení n/R . Do souboru se pak zapisuje celkový počet zvukových vzorků i s nulami:

$$n = n + p \quad (5.3)$$

a nakonec délka zvukového rámce. Samotné vzorky se zapisují až po vytvoření metadat objektu.

5.1.4 Vytvoření a zápis metadat statického a dynamického objektu

Každý zvukový rámeček je provázen metadatami. Ta se ukládají společně s daty postupně do jednoho vektoru, který je po naplnění zapsán do souboru. Metadata nesou dohromady 6 informací v podobě čísel a ukládají se do vektoru v následujícím pořadí:

1. *sample_index*
2. *object_index*
3. *position_azimuth*
4. *position_elevation*
5. *position_radius*
6. *gain_factor*.

Informace *sample_index* značí začátek přehrávání aktuálního objektu v celé zvukové scéně a v případě přehrávání pouze objektového zvuku samostatně je nastavena na hodnotu 0. Pokud zadáme kanálový zvuk v pozadí, bude nést hodnotu rovnou počtu kanálových zvukových vzorků, které se přehrají, než se začne přehrávat objekt. Uživatel si může v grafickém prostředí vybrat, v jakém čase t se přehrávání objektu spustí:

$$\text{sample_index} = (tf_{vz}) - f_{vz}. \quad (5.4)$$

Informace *object_index* udává pořadí objektu ve zvukové scéně. V našem případě je její hodnota pevně nastavena na 0. Další čtyři informace definují polohu a hlasitost objektu v prostoru a jsou popsány v tabulce 3.1.

Aby mohl kodér do vektoru postupně přidávat střídavě metadata a data, je nejdříve potřeba zjistit celkový počet zvukových rámců N :

$$N = \frac{n}{R}. \quad (5.5)$$

Statický objekt

V případě statického objektu jsou metadata stejná pro každý rámeček a opakují se tolikrát, jaký je počet rámců. Až se do vektoru přidají zvukové vzorky z posledního rámečku, vektor se zapíše do souboru a po něm se jako poslední ze všeho zapíše vzorkovací frekvence.

Dynamický objekt

Zadáme-li v grafickém prostředí, že objekt bude dynamický, informace o poloze se budou s každým novým rámečkem měnit. Uživatel si může vybrat z devíti trajektorií objektu a u každé z nich navíc obrátit směr pohybu.

Systém funguje tak, že postupně se pro aktuální rámeček počítají odlišná metadata. Nejprve je nutno nadefinovat vzdálenost V mezi počátečním a koncovým bodem trajektorie. Z této vzdálenosti získáme několik malých úseků s délkou v pomocí následující rovnice:

$$u = \frac{V}{N - 1}. \quad (5.6)$$

Tím zajistíme, že každý rámeček bude mít svou vlastní odlišnou polohu v prostoru. Dále se s úseky pracuje podle toho, jaká trajektorie byla zvolena:

- Zepředu dozadu (ve výšce uší)
 - $V = 8 \text{ m}$,
 - $x = \frac{V}{2} - (i - 1)u$,
 - $y = 0$,
 - $z = 0$,
- Zepředu dozadu (nad hlavou)
 - $V = 8 \text{ m}$,
 - $x = \frac{V}{2} - (i - 1)u$,
 - $y = 0$,
 - $z = 2$,

- Zepředu dozadu (po oblouku)
 - $V = 180^\circ$,
 - $position_azimuth = 0$,
 - $position_azimuth = 180$ pro $elevation > 90$,
 - $elevation = (i - 1)u$,
 - $position_elevation = V - position_elevation$
pro $position_elevation > 90$,
 - $position_radius = 2$,

- Zleva doprava (ve výšce uší)
 - $V = 8$ m,
 - $x = 0$,
 - $y = -\frac{V}{2} + (i - 1)u$,
 - $z = 0$,

- Zleva doprava (nad hlavou)
 - $V = 8$ m,
 - $x = 0$,
 - $y = -\frac{V}{2} + (i - 1)u$,
 - $z = 2$,

- Zleva doprava (po oblouku)
 - $V = 180^\circ$,
 - $position_azimuth = -90$,
 - $position_azimuth = 90$ pro $elevation > 90$,
 - $position_elevation = (i - 1) \cdot u$,
 - $position_elevation = V - position_elevation$
pro $position_elevation > 90$,
 - $position_radius = 2$,

- Po kruhu kolem uší (po směru hodinových ručiček)
 - $V = 360^\circ$,
 - $position_azimuth = (i - 1)u$,
 - $position_azimuth = -V + position_azimuth$
pro $position_azimuth > 180$,
 - $position_elevation = 0$,
 - $position_radius = 2$,

- Po kruhu nad hlavou (po směru hodinových ručiček)
 - $V = 360^\circ$,
 - $position_azimuth = (i - 1) \cdot u$,
 - $position_azimuth = -V + position_azimuth$
pro $position_azimuth > 180$,
 - $position_elevation = 50$,
 - $position_radius = 2$,

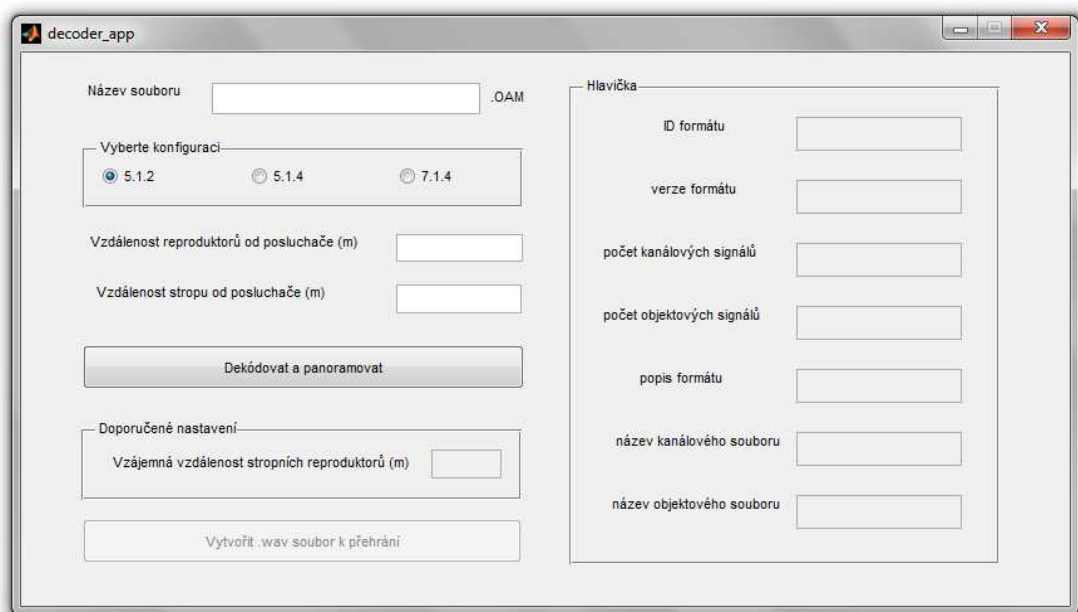
- Diagonála (z levého horního rohu vepředu do pravého spodního vzadu)

- $V = 6 \text{ m}, V_z = 2 \text{ m},$
 - $x = \frac{V}{2} - (i - 1) \cdot u,$
 - $y = -\frac{V}{2} + (i - 1) \cdot u,$
 - $z = V_z - (i - 1) \cdot u_z,$

příčemž index i značí pořadí aktuálního rámce. Do souboru se zapisují metadata ve sférické soustavě souřadnic, pro usnadnění se však u některých trajektorií počítalo se souřadnicemi kartézskými. Pokud je poloha definována v kartézské soustavě souřadnic, je třeba provést převod do soustavy souřadnic sférické funkcí `cart2sph`. Následnou funkcí `rad2deg` se převedou radiány na stupně. V případě zadaného obrácení směru pohybu se pouze úseky přičítají s jinou polaritou. Zápis metadat, dat a vzorkovací frekvence je na stejném principu, jako u objektu statického.

5.2 Dekodér standardu MPEG-H

Grafické uživatelské prostřední dekodéru je sestaveno tak, aby uživateli byly přehledně podány informace o souboru a také informace o správném přehrávání. Poté, co uživatel zadá název souboru, může si vybrat, na jakou konfiguraci si přeje soubor panoramovat. Dále je třeba zadat, s jakou vzdáleností reproduktorů od posluchače a vzdáleností stropu od posluchače, bude dekodér pracovat. Jakmile se soubor dekóduje, vypíše se hlavička souboru, která o něm nese obecné informace. Program vypočítá doporučené nastavení vzájemné vzdálenosti stropních reproduktorů a zobrazí jež uživateli spolu s trajektorií zvukového objektu. V tuhle chvíli je zvukový soubor připraven k vytvoření.



Obrázek 5.2: Graficky zpracovaný dekodér standardu MPEG-H

5.2.1 Načítání souboru .OAM a výpis z něj

Podobně, jak se v kodéru vytvářel soubor a zapisovala se do něj data a metadata, dekodér zase zpět ze souboru data a metadata vypisuje. Soubor se nejdříve načte funkcí `fopen` a následně používá funkci `fgetl`, pokud čte text, a funkci `fread`, pokud čte číslice. Jestliže používá druhou funkci a čte čísla, potřebuje vědět, jestli se jedná o jednu číslici nebo vektor. Vždy je tedy nutné na vstupu funkce dodat, kolik číslic má dekodér číst. Nejprve se přečte hlavička souboru a dále pak data a metadata přesně ve stejném pořadí, v jakém byla do souboru zapsána. Nakonec se soubor zavře funkcí `fclose`.

5.2.2 Rozmístění reproduktorů v prostoru

Reproduktorové konfigurace se řídí doporučením *ITU-R BS.775-3* [9]. Standard MPEG-H však neuvádí bližší informace o tom, s jakým rozmístěním v prostoru dekodér pracuje, proto byly použity konfigurace, které jsou popsány v kapitole 2.2. Elevace horizontálních reproduktorů je rovna 0° a elevace stropních je vždy 60° . Azimuty se liší podle konfigurace. Pro větší přehled a pozdější jednodušší výpočty jsou azimuty reproduktorů řazeny do vektoru, s kterým dekodér následně pracuje. Na prvním místě je centrální kanál a pak pokračují další azimuty po směru hodinových ručiček. Azimuty stropních reproduktorů začínají od jedenácté pozice vektoru. Některé konfigurace nemají určité kanály, přesto jsou kanály naplněny nulami nebo takovou hodnotou, aby odpovídala kanálu na začátku pomyslné kružnice. To umožní panoramovací jednotce pracovat efektivně.

Tabulka 5.1: Azimuty reproduktorů pro konfiguraci 5.1.2

pozice	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
značení	C	R	Rs	Ls	R	C	x	x	x	x	Tl	Tr	x	x	x
Azimut [°]	0	30	110	-110	-30	0	0	0	0	0	-90	90	0	0	0

Tabulka 5.2: Azimuty reproduktorů pro konfiguraci 5.1.4

pozice	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
značení	C	R	Rs	Ls	R	x	x	x	x	x	Tl	Tr	Tls	Trs	x
Azimut [°]	0	30	110	-110	-30	0	0	0	0	0	-135	-45	45	135	-135

Tabulka 5.3: Azimuty reproduktorů pro konfiguraci 7.1.4

pozice	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
značení	C	R	Rs	Rb	Lb	Ls	L	x	x	x	Tl	Tr	Tls	Trs	x
Azimut [°]	0	30	90	150	-150	-90	-30	0	0	0	-135	-45	45	135	-135

5.2.3 Výpočet umístění stropních reproduktorů s korekcí hlasitosti

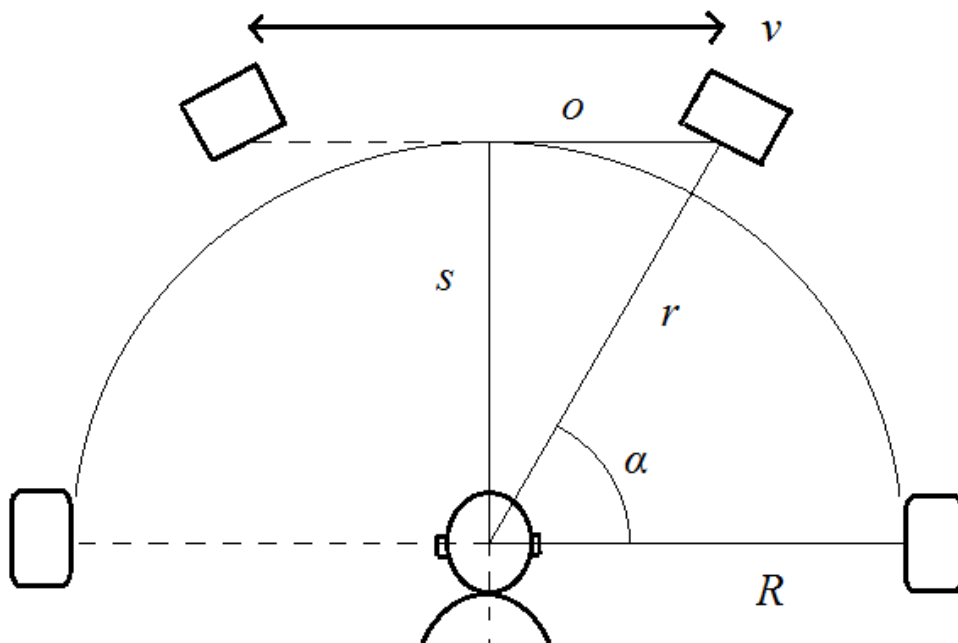
Pro zajištění ekvidistantního umístění všech reproduktorů, je třeba stropní reproduktory umístit pod elevací $\alpha = 60^\circ$. To proto, že levý a pravý reproduktor jsou navzájem také pod tímto úhlem a při konfiguraci 7.1.4 tento úhel svírají všechny reproduktory v horizontální rovině krom centrálního. Každá místnost má ovšem jinak vysoký strop a proto je třeba vypočítat vzdálenost v , v jaké se mají reproduktory nacházet, aby dodržovaly daný úhel. Výpočet je následovný:

$$o = \frac{s}{\tan(\alpha)}, \quad (5.7)$$

$$v = 2o, \quad (5.8)$$

přičemž s je vzdálenost stropu od posluchače a o je vzdálenost reproduktoru k ose kolmé k posluchači. To lze však použít pouze při dvou stropních reproduktorech. Při čtyřech musíme rovnici upravit:

$$v = \sqrt{2o^2}. \quad (5.9)$$



Obrázek 5.3: Umístění stropních reproduktorů

Korekce hlasitosti

Pokud nám strop nedovolí umístit stropní reproduktory do stejné vzdálenosti od posluchače, jako reproduktory v horizontální rovině, je třeba provést korekci hlasitosti. Pokud jsou stropní reproduktory blíže, je potřeba, aby zněli tišeji a při větší vzdálenosti více nahlas. Tím se srovná intenzita zvuku I v místě posluchače:

$$I = \frac{P}{S}, \quad (5.10)$$

přičemž P je výkon zvukového vlnění dopadající na plochu S . Protože vlnění bodového zdroje zvuku se šíří kulově, použijeme výpočet pro povrch koule:

$$S = 4\pi r^2. \quad (5.11)$$

Obecně můžeme říct, že intenzita zvuku vzroste přibližně čtyřikrát s dvounásobnou vzdáleností r zdroje od posluchače. Kodér pracuje tak, že vzdálenost horizontálních reproduktorů od posluchače R odpovídá intenzitě zvuku rovné 1. Pro tuhle vzdálenost a intenzitu spočítá výkon zvukového vlnění:

$$P = I4\pi R^2. \quad (5.12)$$

Vzdálenost r stropních reproduktorů od posluchače se vypočítá pomocí Pythagorovy věty:

$$r = \sqrt{s^2 + o^2}. \quad (5.13)$$

Při stejném výkonu jako u horizontálních reproduktorů bude nová intenzita zvuku větší, pokud bude $r < R$ a naopak. Snahou je tedy reproduktor virtuálně posunout blíže, pokud je ve vzdálenosti větší než D , aby se kompenzovala hlasitost pro intenzitu zvuku odpovídající rozdílu vzdáleností. Proto je třeba počítat s takovou virtuální vzdáleností r_v , která vznikne po posunutí reproduktoru o vzdálenost přesně opačnou, než se doopravdy posunula:

$$r_v = R + R - r. \quad (5.14)$$

Virtuální intenzita zvuku I_v pro stropní reproduktory se tedy bude počítat následovně:

$$I_v = \frac{P}{4\pi r_v^2}. \quad (5.15)$$

Data \mathbf{d} (zvukové vzorky) kanálů pro stropní reproduktory se pak násobí touto intenzitou:

$$\mathbf{d} = \mathbf{d}I_v. \quad (5.16)$$

5.2.4 Úprava kanálových signálů

Pokud soubor obsahuje kanálový zvuk, může se stát, že počet těchto kanálů neodpovídá počtu reproduktorů v horizontální rovině při aktuální konfiguraci. Zde je třeba provést konverzi kanálů na vyšší počet reproduktorů (upmix) nebo případně na nižší (downix).

Konverze počtu kanálů z 5 na 7

Při konverzi na vyšší počet reproduktorů bylo využito dvourozměrného panoramování technikou VBAP, která je popsána v kapitole 4.1.1. Prostorové kanály v konfiguraci 5.1 se uvažují jako virtuální objekty, jejichž azimuty jsou rovny -110° a 110° . Tyto objekty jsou panoramovány na kanály prostorové s azimuty -90° a 90° a kanály zadní s azimuty -150° a 150° pro konfiguraci 7.1. Po získání zesilovacích činitelů \mathbf{g}^{norm} se těmito hodnotami násobí data \mathbf{d} (zvukové vzorky) příslušného kanálu:

$$\mathbf{d} = \mathbf{d}\mathbf{g}^{\text{norm}}. \quad (5.17)$$

Konverze počtu kanálů ze 7 na 5

V případě konverze na nižší počet reproduktorů se posupuje stejně jako v případě konverze na vyšší počet, pouze se data \mathbf{d} hodnotami zesilovacích činitelů \mathbf{g}^{norm} nenásobí, ale dělí:

$$\mathbf{d} = \frac{\mathbf{d}}{\mathbf{g}^{\text{norm}}}. \quad (5.18)$$

5.2.5 Panoramování objektového signálu

Výčet metadat

Z vektoru, který dekodér přečetl ze souboru, je třeba získat zpět informace o objektu v podobě metadat. Jak jsou ve vektoru metadata zakódována pojednává kapitola 5.1.4. Každá informace má ve vektoru svoji pozici před prvním rámcem a pak před každým dalším. Informace *sample_index*, *object_index* a *gain_factor* se opakují se stejnou hodnotou, proto je stačí číst z vektoru pouze jednou na začátku a to na pozicích 1, 2 a 6. Informace o pozici objektu se musí z vektoru číst tolikrát, z kolika rámců je objekt tvořen. Pozice p ve vektoru pro informaci *position_azimuth* se spočítá následovně:

$$p_{\text{position_azimuth}} = 3 + (w - 1)(R + 6), \quad (5.19)$$

pro informaci *position_elevation*:

$$p_{\text{position_elevation}} = 4 + (w - 1)(R + 6), \quad (5.20)$$

a pro informaci *position_radius*:

$$p_{\text{position_radius}} = 5 + (w - 1)(R + 6), \quad (5.21)$$

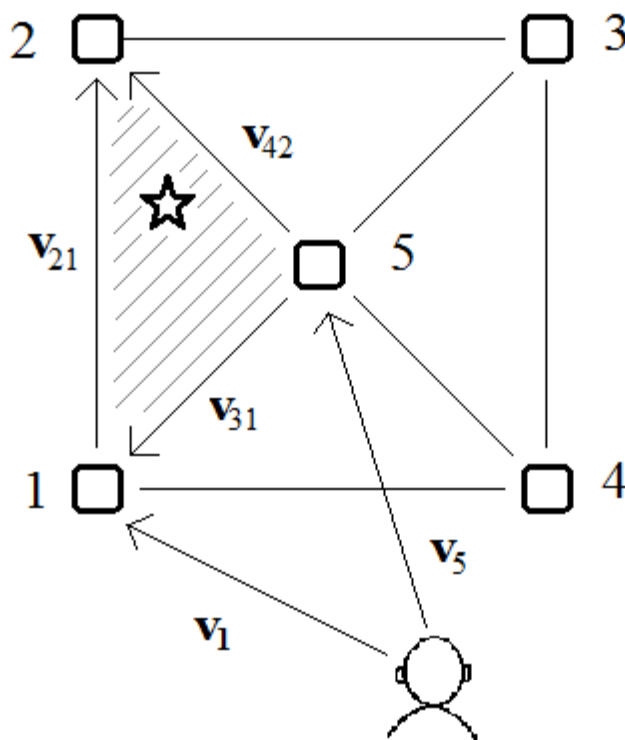
přičemž w značí pořadí následujícího zvukového rámce a R je délka jednoho rámce.

Panoramování využitím rozšířené techniky VBAP

Pro virtuální polohování objektu je využita technika VBAP pro trojrozměrné panoramování, která je popsána v kapitole 4.1.2. Informace pro polohu objektu jsou dány metadaty příslušného rámece a z nich se vypočítá jednotkový vektor p . Informace o poloze reproduktorů jsou přednastaveny v dekodéru, přičemž vzdálenost reproduktorů je zadána uživatelem a úhly se berou z vektorů, které jsou popsány v kapitole 5.2.2. Tím je dána matice L a po jejím vynásobení vektorem p dostáváme zesilovací činitele g . Pro každý rámeček se počítají nové zesilovací činitele. Nejdříve je třeba rozdělit reproduktorovou konfiguraci do několika trojúhelníků tak, aby pokryly celý prostor, přičemž jejich vrcholy jsou tvořeny reproduktory. Vždy dva reproduktory jsou společné pro sousední trojúhelník a trojúhelníky se navzájem nesmí překrývat, jinak mohou při přechodu objektu z jednoho trojúhelníku do druhého vznikat lupance nebo nežádoucí hlasitostní změny. Pro každý trojúhelník se počítají zesilovací faktory. Pokud jsou všechny tři nezáporné, přiřadí se příslušným kanálům.

Řešení problémů při nejednoznačnosti trojúhelníku

Někdy může vzniknout situace, při které není jednoznačné, jaký trojúhelník je třeba použít. Jedná se především o prostory, kde jsou dva možné sousední trojúhelníky osově souměrné podle osy x nebo y . V našem případě jde o prostor tvořený stropními reproduktory nebo reproduktory L_s, R_s, T_l, T_r při konfiguraci 5.1.2, reproduktory L_s, R_s, T_l, T_r při konfiguraci 5.1.4 a reproduktory L_b, R_b, T_l, T_r při konfiguraci 7.1.4. V takových případech je nutné použít rozšíření techniky VBAP, které popisuje patent WO2013181272A2 [25]. V tomhle rozšíření se provádí modifikace trojúhelníků menšími trojúhelníky tvořící čtverec popřípadě jiný čtyřúhelník.



Obrázek 5.4: Modifikovaná technika VBAP

Nejprve je nutné zvolit polohu virtuálního reproduktoru, který bude uprostřed čtverce. To se provede vypočítáním vektoru \mathbf{v}_5 směřujícího k reproduktoru:

$$\mathbf{v}_5 = \mathbf{v}_1 + \frac{(\mathbf{v}_{21} \times \mathbf{v}_{42}) \cdot (\mathbf{v}_{31} \times \mathbf{v}_{42})}{|\mathbf{v}_{31} \times \mathbf{v}_{42}|^2} \mathbf{v}_{31}, \quad (5.22)$$

přičemž symbol \times značí vektorový součin a symbol \cdot součin skalární. Rovné závorky znamenají velikost vektoru. Tento vektor je třeba přepočítat na jednotkový vektor \mathbf{L}_5 :

$$\mathbf{L}_5 = \frac{\mathbf{v}_5}{\sqrt{(v_{51}^2 + v_{52}^2 + v_{53}^2)}}. \quad (5.23)$$

Při výpočtu trojice zesilovacích činitelů se dva počítají pro fyzické reproduktory a třetí pro reproduktor virtuální. Dvojice fyzických reproduktorů se vybírá podle polohy objektu. Objekt leží v malém trojúhelníku vzniklém překrýváním dvou původních trojúhelníků. Zesilovací činitele \mathbf{g}^{125} pro reproduktory 1,2 a 5 se počítají klasicky podle rovnice (4.14). Zesilovací činitele fyzických reproduktorů se pak určí následovně:

$$\begin{aligned} g_1 &= 2g_1^{125} + g_5^{125}, \\ g_2 &= 2g_2^{125} + g_5^{125}, \\ g_3 &= g_5^{125}, \\ g_4 &= g_5^{125}. \end{aligned} \quad (5.24)$$

Umístění virtuálního reproduktoru pro reproduktory tvořící lichoběžník

V případě rovnoramenného lichoběžníku, který je tvořen dvěma zadními reproduktory a dvěma stropními, se při testování stalo, že vektor \mathbf{v}_5 podle rovnice směřoval pro každý malý trojúhelník jiným směrem a zesilovací činitele \mathbf{g}^{125} později v některých případech vycházely všechny záporně, přestože vektor virtuálního objektu \mathbf{p} směřoval do tohoto prostoru. Proto se rovnice (5.22) musí nahradit jinou, která zajistí, že vektor bude směřovat vždy do těžiště lichoběžníku a tam se bude nacházet i virtuální reproduktor. Vzdálenost z těžiště od dolní základny se spočítá následovně:

$$z = \frac{(S_1 z_1 + S_2 z_2 + S_3 z_3)}{S}, \quad (5.25)$$

kde $S_1 = S_3$ je obsah trojúhelníku po rozdělení lichoběžníku s obsahem S na čtverec s obsahem S_2 a dva stejné trojúhelníky:

$$S_1 = \frac{d-v}{4} h, \quad (5.26)$$

$$S_2 = v h, \quad (5.27)$$

$$S = \frac{d+v}{2} h, \quad (5.28)$$

přičemž v je vzdálenost stropních reproduktorů a spočítá se podle rovnice (5.8) nebo (5.9) a d značí vzdálenost reproduktorů \mathbf{L}_4 a \mathbf{L}_1 umístěných v horizontální rovině:

$$d = |\mathbf{l}_4 - \mathbf{l}_1|. \quad (5.29)$$

Proměnná h je výška lichoběžníku a v našem případě velikost vektoru jdoucího od středu jedné základny ke středu druhé. Pro její výpočet se použije tato rovnice:

$$h = |\mathbf{d}_{1/2} - \mathbf{v}_{1/2}|, \quad (5.30)$$

$$h = \left| \frac{(\mathbf{l}_4 + \mathbf{l}_1)}{2} - \frac{(\mathbf{l}_3 + \mathbf{l}_2)}{2} \right|, \quad (5.31)$$

kde $\mathbf{d}_{1/2}$ je vektor směřující ke středu spodní základny a $\mathbf{v}_{1/2}$ vektor směřující ke středu horní základny. Proměnná $z_1 = z_3$ v rovnici (5.25) značí vzdálenost těžiště trojúhelníku od spodní základny a proměnná z_2 vzdálenost těžiště obdélníku od základny:

$$z_2 = \frac{1}{2}h. \quad (5.32)$$

Vzdálenost těžiště trojúhelníků od spodní základny se bude s konfigurací měnit. Pro konfigurace 5.1.2 a 5.1.4 se vypočítá následovně:

$$z_1 = \frac{1}{3}h. \quad (5.33)$$

Pro konfiguraci 7.1.4 bude platit rovnice:

$$z_1 = \frac{2}{3}h. \quad (5.34)$$

Dále je nutné zjistit, jaký úhel α svírá lichoběžník s horizontální rovinou. To provedeme tak, že vektor $\mathbf{v}_{1/2}$, posuneme o stejnou vzdálenost směrem k posluchači, jakou nese hodnota souřadnice x pro vektor $\mathbf{d}_{1/2}$, a pro tento bod spočítáme elevaci pomocí funkce `cart2sph`. Pro souřadnici z vektoru \mathbf{v}_5 směřujícímu k těžišti lichoběžníku se použije tento vztah:

$$v_5(z) = \sin(\alpha)z, \quad (5.35)$$

Pro souřadnici x :

$$v_5(x) = -R + \cos(\alpha)z, \quad (5.36)$$

kde R značí vzdálenost horizontálních reproduktorů od posluchače. Souřadnice y bude vždy nulová, protože těžiště rovnoramenného lichoběžníku se bude nacházet v prostředí šířky lichoběžníku, který se nachází na ose x soustavy souřadnic pro posluchače a reproduktorovou konfiguraci.

Korekce hlasitosti objektových signálů

Zesilovací činitele se po vypočítání normalizují podle rovnice (4.15), aby celková hlasitost po panoramování nevzrostla nebo neklesla. Dále se hlasitost objektu mění v závislosti na vzdálenosti od posluchače. Informace o této vzdálenosti se nazývá *position_radius* a je zakódována v metadatech. Aby hlasitost při stále menší vzdálenosti od posluchače nevzrůstala až do nekonečna, program všechen *radius* menší než je hodnota 0,5 ponechá právě této hodnotě. Všechny objekty blíže, než 0,5 m se jeví stále ve stejné vzdálenosti. Dekodér pracuje právě tak, že poloměru 0,5 přiřadí virtuální intenzitu zvuku rovné hodnotě 1. Pro tuhle vzdálenost a intenzitu spočítá výkon zvukového vlnění pomocí rovnice (5.12). Výsledná virtuální intenzita s novým poloměrem se spočítá podle rovnice (5.15). Nakonec se touto intenzitou I_v vynásobí normalizované zesilovací činitele aktuálního rámece:

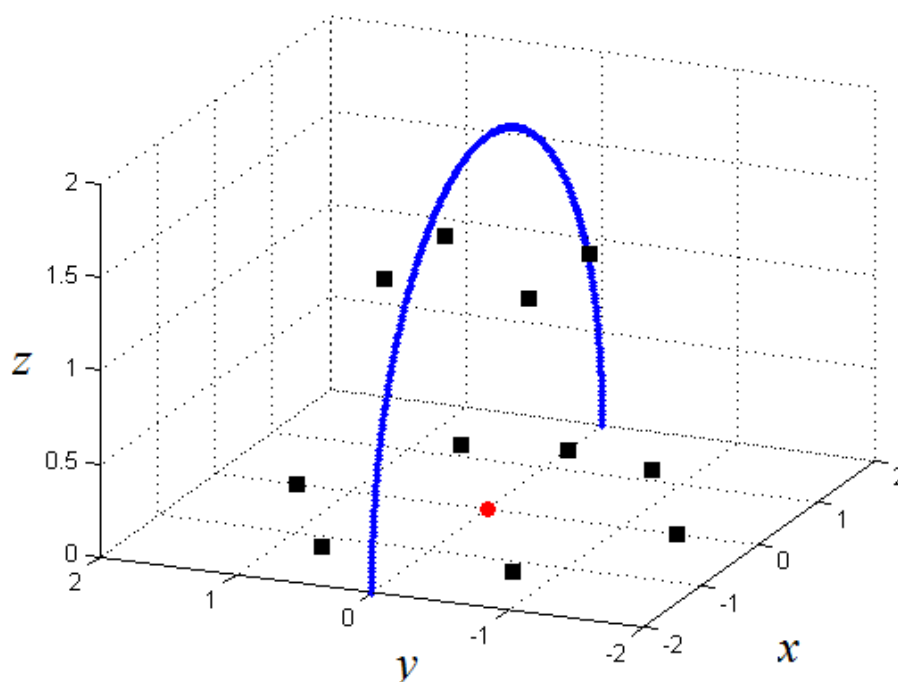
$$\mathbf{g}_{\text{rad}} = \mathbf{g}_{\text{norm}} I_v. \quad (5.37)$$

V metadatech je ještě zakódována informace *gain_factor*. Hodnotou této informace se znova násobí zesilovací činitelé:

$$\mathbf{g}_{\text{gain}} = \mathbf{g}_{\text{norm}} \mathbf{gain_factor}. \quad (5.38)$$

Grafická prezentace reproduktorů a trajektorie objektu

Po vyčtení informací o poloze objektu se pro každý rámeček aktuální poloha zobrazuje v jednom grafu, stejně jako umístění reproduktorů při zvolené konfiguraci.



Obrázek 5.5: Trajektorie objektu zepředu dozadu (po oblouku) při konfiguraci 7.1.4

5.2.6 Vytvoření a úprava výstupních signálů

Nejprve je třeba odebrat z vektoru aktuální zvukový rámec, který se bude násobit zesilovacími činiteli. Pozice prvního (začátečního) vzorku rámce p_z se vypočítá následovně:

$$p_z = 7 + (w - 1)(R + 6), \quad (5.39)$$

přičemž w značí pořadí aktuálního rámce a R značí délku jednoho rámce. Pozice posledního (konečného) vzorku rámce p_k se vypočítá následovně:

$$p_k = (7 + (w - 1)(R + 6)) + (R - 1). \quad (5.40)$$

Data \mathbf{D} jednoho zvukového rámce určené pozicemi p_z a p_k jsou násobeny vypočítanými zesilovacími činiteli \mathbf{g}_{gain} :

$$\mathbf{D} = \mathbf{D}\mathbf{g}_{\text{gain}} \quad (5.41)$$

Po vynásobení vznikají vektory délky R , které se ukládají do objektové matice na pozicích příslušných daným kanálům. Ostatní kanály jsou zastoupeny vektory, které jsou naplněny nulami. Pro každý nový rámec se nové vektory přičítají k vektorům v objektové matici, dokud není matice naplněna.

Pokud je zakódován kanálový zvuk v pozadí, matice kanálových dat a matice objektových dat se sčítají. K tomu je však potřeba mít matice o stejné velikosti a to vznikne pouze v případě, že bylo uživatelem nastaveno stejně dlouhé přehrávání pro oba zvuky. Počet kanálů je stejný, ale délka matice je potřeba doplnit nulami. Informace *sample_index*, která byla dekodována, nese informaci o tom, po jaké době od přehrávání kanálového zvuku se spustí zvuk objektový. To znamená, jaký počet nul z_z je třeba doplnit na začátku objektové matice. Pokud po doplnění nulami do objektové matice je součet počtu nul z_p a počtu objektových zvukových vzorků n_o menší než počet kanálových vzorků n_{ch} , je třeba doplnit v objektové matici nuly také za objektovými vzorky. Počet těchto nul z_z se vypočítá následovně:

$$z_z = n_{\text{ch}} - n_o - z_p. \quad (5.42)$$

V případě, že je objektová matice po přidání nul před objektové vzorky větší než matice kanálová, je třeba do kanálové matice přidat za zvukové vzorky nuly s počtem z_{ch} :

$$z_{\text{ch}} = n_o + z_p + n_{\text{ch}}. \quad (5.1)$$

V tuhle chvíli jsou matice připravené k sečtení. Každý sloupec ve výsledné matici představuje jeden zvukový kanál a jejich pořadí v matici je nezávislé na reproduktorové konfiguraci. Pokud konfigurace nepodporuje některé kanály, jsou tyto sloupce vyplněny nulami.

Tabulka 5.4: Pořadí kanálů v matici a jejich značení

pořadí	1	2	3	4	5	6	7	8	9	10	11	12
kanál	L	R	C	LFE	Ls	Rs	Lb	Rb	Tl	Tr	Tls	Trs

Výsledná matice se zapíše do souboru .wav za použití funkce audiowrite.

6 Poslechové testy

Poslechový test porovnávající subjektivní vnímání polohy zvukového objektu v prostoru při použitých konfiguracích 5.1.2, 5.1.4 a 7.1.4 je rozdělen do dvou částí.

Statická část

Pro první část bylo vygenerováno šest zvukových souborů objektu statického pro každý typ reproduktorové konfigurace. Vždy tři objekty jsou umístěny do prostoru tak, aby byly blízko u sebe. První trojice se nachází před posluchačem mírně vlevo a druhá za posluchačem mírně vpravo. Polohy objektů jsou zaznamenány v následující tabulce s tím, že rádius je vždy roven vzdálenosti reproduktorů (v našem případě 1,4 m) a gain byl nastaven na 1.

Tabulka 6.1: Polohy statických objektů při poslechovém testu

Název zvukového souboru s objektem	azimut (°)	elevace (°)
stat.p.1	-30	15
stat.p.2	-45	30
stat.p.3	-15	45
stat.z.1	180	50
stat.z.2	165	35
stat.z.3	140	65

Trojice souborů se bude spouštět postupně za sebou jak je uvedeno v tabulce 6.1. Posluchač určí, pořadí zazněných objektů. S rostoucím počtem kanálů při různých reproduktorových konfiguracích by měla růst správnost určení pořadí.

Dynamická část

Pro druhou část bylo vygenerováno osm zvukových souborů objektu dynamického pro každý typ reproduktorové konfigurace. Posluchač v této části musí určit pořadí přehrávaných objektů s různou trajektorií v jedné skupině objektů. Tři skupiny přehrávaných objektů jsou tvořeny objekty s odlišnými trajektoriemi. Výsledky testu by měly zhodnotit, do jaké míry mají reproduktory umístěné ve vertikální rovině vliv na subjektivní vnímání polohy objektu a jaká reproduktorová konfigurace je pro trajektorie tohoto typu dostačující.

1. Skupina:
 - Zepředu dozadu (ve výšce uší)
 - Zepředu dozadu (nad hlavou)
 - Zepředu dozadu (po oblouku)

2. Skupina:
 - Zleva doprava (ve výšce uší)
 - Zleva doprava (nad hlavou)
 - Zleva doprava (po oblouku)

3. Skupina:
 - Po kruhu kolem uší (po směru hodinových ručiček)
 - Po kruhu nad hlavou (po směru hodinových ručiček)

6.1 Příprava pracoviště

Testy byly připraveny v laboratoři SC1.22 na VUT. Počet kanálů u konfigurace 7.1.4 je vysoký a bylo nutné dohromady použít jedenáct reproduktorů. Kanál pro LFE nebyl použit, protože nemá vliv na prostorový vjem posluchače. Bylo použito šest monitorů typu Genelec 1032A, jeden centrální monitor typu Event TR a čtyři stropní monitory typu KRK VXT 4, ty byly namontovány na speciálně sestavené kovové konzole a vyzdviženy do potřebné výšky. Protože každý typ reprodukuje zvuk s jinou citlivostí, bylo třeba nastavit, aby při stejném signálu reproduktory zněly v místě posluchače stejně hlasitě. To se provedlo za použití zvukoměru, který byl postaven na stojan do místa posluchače ve výšce uší. Do každého kanálu se postupně pouštěl bílý šum o stejné hlasitosti a zvukoměr měřil průměrnou hladinu akustického tlaku. Všechny monitory jsou aktivní a jsou na zadních stranách vybaveny potenciometry pro regulaci citlivosti. Pro každý reproduktor se upravila postupně citlivost tak, aby zvukoměr ukazoval stejnou hodnotu. Monitory typu KRK namontované na konstrukci reprodukovaly zvuk celkově s menším akustickým tlakem, proto bylo nutné ještě pro ostatní typy snížit citlivost na zvukové kartě M-Audio ProFire. Zvukové karty byly použity dvě. V první hlavní kartě byly na výstupu připojeny reproduktory v horizontální rovině (monitory Genelec a Event) a ve druhé kartě reproduktory stropní (monitory typu KRK). Tyto dvě karty byly navzájem připojeny přes rozhraní ADAT a do notebooku pomocí sběrnice FireWire. Bílý šum se generoval z digitálního zvukového editoru Sound Forge. Po té, co byly nastaveny citlivosti reproduktorů, bylo možné spouštět zvukové soubory .wav vygenerované vytvořeným dekodérem.

6.2 Ověření správnosti elektroakustického řetězce

Po té, co se podařilo zprovoznit kodér a dekodér podle standardu MPEG-H 3D audio, bylo nutné poslechem ověřit správnost celého programu a především panoramovací jednotky. Při prvním poslechu bylo znatelné, že panoramovací jednotka, která využívá trojrozměrné panoramování techniky VBAP nefunguje zcela správně. Problém vykazovala při trajektorii objektu v úrovni uší, ať už zepředu dozadu nebo zleva doprava, kdykoli objekt procházel hlavou posluchače. Bylo zaznamenáno přeskočení zvukového signálu do reproduktoru, který se nacházel na druhé straně pomyslné kružnice při průchodu objektu posluchačem. Dále v místech, kde je prostor tvořen čtveřicí reproduktorů, byly po průchodu objektu tímto místem slyšet výrazné hlasitostní změny. Zvuk v reproduktorech přeskakoval, podle toho, jaký trojúhelník byl zrovna aktivní. Podařilo se nalézt patent, který by tento problém s přeskakováním v prostoru měl odstranit. Po přeprogramování algoritmu podle informací z patentu se podařilo vyřešit přeskakování v prostoru tvořeném čtveřicí stropních reproduktorů. Prostor za zády posluchače však stále trpěl hlasitostními změnami. Bylo zjištěno, že zde program počítá stále pro trojici reproduktorů na místo čtveřice. To bylo zapříčiněné tím, že patent dokázal panoramovat zvuk do čtyř reproduktorů pouze v případě, jednalo-li se o čtverec. V čase, který zbýval do odevzdání bakalářské práce, se pracovalo na ladění panoramovací jednotky, aby fungovala správně. Nakonec se algoritmus podařilo zprovoznit tak, aby byl přechod dynamického objektu hladký. To však bylo ověřeno pouze individuálně. Při poslechovém testu bylo alespoň ověřeno, že při konfiguraci 7.1.4 objekt, jehož trajektorie se nachází nad hlavou posluchače, je subjektivně rozeznatelný od objektu, jehož trajektorie je ve výšce uší. Poslechové testy jsou připravené a zvukové soubory vygenerované a uložené v příloze na CD. Problém techniky VBAP při trajektorii procházející hlavou posluchače je nevyřešen a nebyl nalezen patent, který by to řešil. Proto na CD nejsou mezi vygenerovanými soubory objekty s trajektorií Zepředu dozadu (ve výšce uší) a Zleva doprava (ve výšce uší).

Závěr

Tato práce se zabývá objektivě založeným zvukem a formáty, které jej používají. V první kapitole jsem popsal, jak se vyvíjel prostorově založený zvuk až do současného stavu. Dále jsem představil současné 3D formáty a provedl jejich srovnání, které je realizováno v tabulce. Analyzoval jsem dostupné zvukové vzorky formátů a porovnal parametry jejich bitových proudů. Pro formát MPEG-H-3D Audio, který je dnes standardem, se mi nepodařilo nalézt zvukový vzorek nebo alespoň část jeho datového toku. S největší pravděpodobností tento datový tok hledaného standardu není skupinou MPEG zveřejněn. Pro tentýž formát jsem provedl v kapitole 3 jeho specifikaci se zaměřením na strukturu jeho datového toku a dekódování objektivě založených metadat, která nesou potřebné informace pro správné panoramování objektu do prostoru. V kapitole 4 jsem uvedl způsoby virtuálního polohování zdroje. Ze všech metod jsem si vybral a detailně popsal způsob panoramování technikou VBAP, jelikož je využíván dekodérem MPEG-H 3D Audio.

Jelikož se mi nepodařilo sehnat zvukový vzorek nebo alespoň část datového toku formátu MPEG-H 3D Audio, nebylo možné provést dekódování metadat a následné panoramování objektivě založeného zvuku do prostoru. Proto jsem musel nejprve vytvořit kodér, který pracuje podle specifikace standardu MPEG-H 3D Audio. V kapitole 5 je popsáno, jakým způsobem jsem prováděl sestavení algoritmu v Matlabu pro celý elektroakustický řetězec od zakódování metadat po jejich dekódování a rendrování objektu do prostoru. V kodéru je nejprve vytvořen soubor s koncovkou .OAM, do kterého se zapisují obecné informace o souboru v podobě hlavičky a dále také zvukové vzorky dat, které jsou do kodéru načítány uživatelem. Do souboru se dále zapisují metadata, která kodér vytváří z informací zadaných uživatelem v grafickém uživatelském prostředí. Algoritmus jsem navrhl tak, aby si uživatel mohl vybrat, zda chce společně s objektem přehrávat kanálový zvuk v pozadí, jak dlouho si přeje oba zvuky přehrávat a v jaké sekundě po přehrávání kanálového zvuku se spustí zvuk objektu. Uživatel si také může vybrat, zda chce přehrávat objekt staticky nebo dynamicky. Pro případ, že vybere dynamický objekt, jsem připravil celkově devět trajektorií, které jsou programovány tak, aby při jejich přehrávání bylo jasné, zda je slyšet rozdíl mezi objektem pohybujícím se v úrovni uší a objektem pohybujícím se nad hlavou posluchače.

Následně jsem sestavil dekodér, ve kterém se načítá soubor s koncovkou .OAM a dekódují se z něj obecné informace o souboru v podobě hlavičky, které se uživateli zobrazí v grafickém uživatelském prostředí. Aby uživatel věděl, jaká trajektorie byla v souboru zakódována, zobrazí se její průběh v grafu i s reproduktorovou konfigurací, na kterou si uživatel přál objekt panoramovat. Sestavil jsem algoritmus, který po zadání vzdálenosti reproduktorů od posluchače a vzdálenosti stropu od posluchače dá uživateli informace o tom, do jaké vzdálenosti od sebe by měl stropní reproduktory umístit. V případě, že se výška stropu neshoduje se vzdáleností reproduktorů od posluchače, algoritmus srovná hlasitosti stropních reproduktorů tak, aby je posluchač vnímal stejně reproduktory v horizontální rovině. V případě, že uživatel zadal konfiguraci, jejíž počet reproduktorů v horizontální rovině neodpovídá počtu kanálů pro kanálový zvuk v pozadí, provede se konverze těchto kanálů na zvolenou konfiguraci. Algoritmus panoramování objektu jsem sestavil tak, aby využíval

techniku VBAP pro trojrozměrné panoramování tak, jak specifikuje standard MPEG-H 3D Audio.

Připravil jsem poslechové testy, před kterými jsem si chtěl individuálně ověřit funkčnost programu. Při ověřování jsem zjistil, že panoramovací jednotka nefunguje správně a poslechové testy porovnávající subjektivní vnímání polohy objektu při různých reproduktorových konfiguracích v tu chvíli nemělo význam provádět. Algoritmus jsem opravil po nalezení patentu, který rozšiřuje techniku VBAP o modifikaci trojúhelníků tvořící reproduktory na čtyřúhelník. Po dalším ověřování funkčnosti jsem zjistil, že tento patent funguje pouze v případě, že čtyřúhelník tvořený reproduktory je pouze čtverec. Protože reproduktory stropní se zadními tvoří lichoběžník, zde patent nefungoval a to se dalo poznat i poslechem v podobě hlasitostních změn při průchodu objektu tímto prostorem. Navrhl jsem tedy nový algoritmus, založený na patentu, který nyní počítá zesilovací činitele správně i pro rovnoramenný lichoběžník. Ten jsem ověřil při poslechu.

Provedl jsem individuální poslechový test na konfiguraci 7.1.4 pro různé trajektorie objektu a slyšel jsem znatelný rozdíl mezi trajektorií Zepředu dozadu (nad hlavou) a Zepředu dozadu (po oblouku). Rozdíl byl znatelný také mezi trajektoriemi Zleva doprava (nad hlavou) a Zleva doprava (po oblouku) stejně tak jako mezi trajektoriemi Po kruhu nad hlavou a Po kruhu kolem uší. Z toho usuzuji, že stropní reproduktory mají vliv na subjektivní vnímání polohy objektu ve vertikálním směru. Bohužel jsem se potýkal v práci s různými překážkami a nezbyl čas, abych to ověřil pro více lidí a při různých konfiguracích. Program je ale funkční a připravený pro testování a další používání. V příloze na CD jsem zanechal všechny vygenerované zvukové vzorky připravené pro testování. Poznatek z práce je takový, že trojrozměrné panoramování technikou VBAP není dostačující pro standardní reproduktorové konfigurace, pokud se neprovede jeho rozšíření a modifikaci trojúhelníků na čtverec nebo lichoběžník. Dále si také zákonitě nedokáže poradit při průchodu objektu přímo skrze posluchače.

Použitá literatura

- [1] POTISK, Tilen. *Head-Related Transfer Function* [online]. Faculty of Mathematics and Physics: University of Ljubljana, 2015
- [2] NÁBĚLĚK, Lukáš. *Prostorový zvuk a jeho využití* [online]. Brno : Janáčkova akademie muzických umění v Brně, divadelní fakulta, Ateliér audiovizuální tvorby a divadla, 2015.
- [3] REICHL, Jaroslav. *Encyklopedie fyziky* [Online]. Dostupné z: <http://fyzika.jreichl.com/main.article/view/1359-digitalni-zvuk>.
- [4] AURO, Technologies. *Next generation Immersive Sound systém* [online]. BARCO NV, 2015.
- [5] AURO, Technologies. *AURO-3D HOME THEATER SETUP Installation Guidelines* [online]. Belgie : Auro Technologies, 2015
- [6] DAELE, Bert Van a Wilfried VAN BAELEN. *Principles behind a revolutionary codec* [online]. AURO-3D OCTOPUS CODEC, 2011.
- [7] LABORATORIES, Dolby. *Dolby Atmos. Next-Generation Audio for Cinema* [online]. San Francisco: Dolby Laboratories, 2014.
- [8] LABORATORIES, Dolby. *Dolby Atmos for the Home Theater* [online]. San Francisco: Dolby Laboratories, 2015.
- [9] *Recommendation ITU-R BS.775-3: Multichannel stereophonic sound system*. International telecommunication union, 2012.
- [10] LABORATORIES, Dolby. *Dolby Atmos Home Theater Installation Guidelines – Červenec 2016*. [Online] Dostupné z: <https://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-home-theater-installation-guidelines.pdf>
- [11] LABORATORIES, Dolby. *Authoring for Dolby Atmos Cinema* [Online]. San Francisco: 2013
- [12] PNEDLEBURY, Ty. *DTS:X: The immersive audio standard, explained* [Online]. 2016 Dostupné z: <https://www.cnet.com/news/dts-x-the-dolby-atmos-alternative-explained/>
- [13] COHEN, Steven. *High-Def Digest's DTS:X Home Theater Guide* [Online]. 2015 Dostupné z: http://www.highdefdigest.com/news/show/DTS/dtsx/Surround_Sound/objectbased-mixing/Home_Audio/Receivers/Dolby_Atmos/Denon/Marantz/The_Avengers/age-of-ultron/Marvel/highdef-digests-dtsx-home-theater-guide/22341
- [14] *AUDIO DEFINITION MODEL: METADATA SPECIFICATION*. EBU, Tech 3364, Status: Version 1.0: Geneva: 2014

- [15] MORSTON, David. *The Audio Definition Model* [Online]. Position paper for the Fourth W3C Web and TV Workshop: London
- [16] *Encoder Input Format for MPEG-H 3D Audio: CODING OF MOVING PICTURES AND AUDIO*. Geneva, Swiss: INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013.
- [17] Information technology- High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D ISO/IEC DIS 23008-3, 2014
- [18] The ISO/MPEG Unified Speech and Audio Coding Standard – Consistent High Quality for all Content Types and at all Bit Rates. MAX NEUENDORF and AES Members. Fraunhofer Institute for Integrated Circuits IIS, Erlangen, 91058 Germany, et al.
- [19] FUEG, Simone, a další. *Design, coding and processing of metadata for object-based interactive audio* [Online]. In: Audio Engineering Society Convention 137. Audio Engineering Society, 2014
- [20] Jürgen Herre, Johannes Hilpert, Achim Kuntz, and Jan Plogsties, *MPEG-H 3D Audio — The New Standard for Coding of Immersive Spatial Audio* [Online], IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 9, NO. 5, AUGUST 2015
- [21] Jürgen Herre, Johannes Hilpert, Achim Kuntz, and Jan Plogsties. *MPEG-H Audio - The New Standard for Universal Spatial / 3D Audio Coding. International Audio Laboratories Erlangen* [Online]. A Joint Institution of Universität Erlangen-Nürnberg and Fraunhofer: Německo
- [22] Seungkwon Beack, Jongmo Sung, Jeongil Seo, and Taejin Lee, *MPEG Surround Extension Technique for MPEG-H 3D Audio*[Online]. Etri Journal 2016
- [23] Jonas Engdeg, et all. *Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding* [Online], 1Dolby Sweden AB, Gåvlegatan 12A, 11330, Stockholm, Sweden, et all
- [24] VILLE PULKKI, AES Member. *Virtual Sound Source Positioning Using Vector Base Amplitude Panning** [Online]. Helsinki University of Technology: Laboratory of Acoustics and Audio Signal Processing, FIN-02015 HUT, Finland
- [25] World intellectual property organization. Object-based audio system using vector base amplitude panning. Pierre-Anthony Stivell Lemieux, Roger Wallace Dressler, Jean-Marc Jot. Patent cooperation treaty. WO2013181272A2. 2013

Seznam použitých zkratek

ADAT	Standard pro přenos digitálního zvuku mezi vybavením
ADM	Audio definitiv model - Specifikace popisující metadata
ASCII	American Standard Code for Information Interchange - Americký standardní kód pro výměnu informací
DAW	Digital audio workstation - Digitální zvuková stanice
DBAP	Distance-based amplitude panning - Amplitudové panoramování založené na vzdálenosti
DPCM	Differential linear pulse-code modulation - Rozdílová pulzně kódová modulace
DTS	Digital Theatre System - Formát prostorového zvuku
DVD	Digital Versatile Disc - Formát digitálního nosiče
HOA	Higher-order Ambisonics - Ambisonie vyššího řádu
HRTF	Head-Related Transfer Function - Přenosová funkce hlavy
ID	IDentification - Identifikace
IEC	International Electrotechnical Commission - Mezinárodní elektrotechnická komise
ISO	International Organization for Standardization - Mezinárodní organizace pro normalizaci
LPCM	Linear pulse-code modulation - Lineární pulzně kódová modulace
MADI	Multichannel Audio Digital Interface - Standard komunikačního protokolu
PCM	Pulse-code modulation - Pulzně kódová modulace
SAOC	Spatial Audio Object Coding - Metoda kódování zvukového objektu
SDDS	Sony Dynamic Digital Sound - Formt prostorového zvuku
USAC	Unified Speech and Audio Coding - Kodér komprimující zvuková data
VBAP	Vector-based amplitude panning - Amplitudové panoramování založené na vektoru
VHS	Video Home System - Standard pro videokazety
WAV	Waveform audio file format - Zvukový formát ukládání zvuku

Značení kanálů:

C	central	centrální
Hl	high left	vrchní levý
Hls	high left surround	vrchní levý prostorový
Hr	high right	vrchní pravý
Hrs	high right surround	vrchní pravý prostorový
L	left	levý
Lb	left back	levý zadní
LFE	low frequency effect	efektový
Ls	left surround	levý prostorový
R	right	pravý
Rb	right back	pravý zadní
Rs	right surround	pravý prostorový
T	top	stropní
Tl	top left	stropní levý
Tls	top left surround	stropní levý prostorový
Tr	top right	stropní pravý
Trs	top right surround	stropní pravý prostorový

Seznam příloh

A	Obsah CD	53
A.1	Bakalářská práce.....	53
A.2	Zdrojový kód	53
A.3	Zvukové soubory k zakódování.....	53
A.4	Zvukové soubory k testování.....	53
A.5	Příručka uživatele	53
A.6	Help	53

A Obsah CD

A.1 Bakalářská práce

Na CD je uložena bakalářská práce ve formátu PDF totožná s tou vytištěnou, její název je `3D Audio standardy v prostředí domácího kina.pdf`.

A.2 Zdrojový kód

CD obsahuje 4 soubory programového prostředí Matlab, z toho dva soubory jsou funkce psané v jeho programovacím jazyce. První z nich nese název `coder_app.m` a je mu přiřazen soubor pro grafické uživatelské prostředí zvaný `coder_app.fig`. Druhá funkce se nazývá `decoder_app.m` a jí přísluší grafické uživatelské prostředí s názvem `decoder_app.fig`.

A.3 Zvukové soubory k zakódování

Na disk byly vypáleny 2 zvukové soubory. Jeden z nich je `příroda.wav` a je tvořen šesti kanály. Druhý soubor nese název `moucha.wav` a jedná se o mono signál v hodný pro objekt. Tyto soubory byly získány z webu <http://freesound.org/> a byly použity v programu pro generování výsledných zvukových souborů z dekodéru. Uživatel může tyto soubory v programu použít, pokud nemá žádné jiné vyhovující soubory po ruce.

A.4 Zvukové soubory k testování

Ve složce `/poslechový test/` jsou zvukové soubory generované dekodérem připravené pro poslechový test. Jsou rozříděné do jednotlivých složek podle reproduktorové konfigurace. Pojmenovány jsou buďto podle trajektorie nebo v případě statických objektů podle tabulky 6.1 s tím, že za pomlčkou je navíc uveden typ konfigurace. Ve složce je také textový dokument `poslechový test.txt` nesoucí patřičné informace.

A.5 Příručka uživatele

Aby se uživatel mohl snadno orientovat v grafickém uživatelském prostředí kodéru a dekodéru, vznikl textový dokument `příručka uživatele.txt`. Tento soubor obsahuje informace, jak správně aplikace používat.

A.6 Help

Textový dokument s názvem `help.txt` by měl pomoci uživateli při potížích. Nese informace o správném spouštění programu a zacházení s funkcemi.