

Oponentní posudek disertační práce

Autor práce: Tomáš Oramus

Název práce: Vliv zvuku na vnímání audiovizuálního díla

Postup řešení výzkumného problému

Předložená disertační práce se zabývá třemi vybranými současnými zvukovými technologiemi a jejich působení na diváka/posluchače/hráče (uživatele). Práce se dělí do dvou částí.

První, teoretická, část práce se zabývá koncepty prožitku a způsoby jejich zkoumání. První kapitola této části přehledně představuje jednotlivé koncepty – imerzi, prezenci, flow, kognitivní absorpci a narativní transportaci. Druhá kapitola pak podává přehled metod objektivního a subjektivního měření prožitku. Třetí kapitola diskutuje jednotlivé koncepty ve vztahu k tématu práce a dochází k závěru, že vhodný koncept je nutné vybrat vždy pro konkrétní výzkumnou otázku, ale obecně je prezence vhodnější pro zkoumání vlivu technologie a imerze pro zkoumání obsahu, který daná technologie nabízí.

Druhá, praktická část práce, se zabývá samotným vlivem zvuku na vnímání AV díla pro tři vybrané konkrétní technologie. Čtvrtá kapitola uvádí do praktické části a pátá, šestá a sedmá kapitola se věnují jednotlivým vybraným technologiím – binaurálnímu audiu, virtuální realitě a Dolby Atmosu. Možná trochu nezvykle jsou uvedené kapitoly pojaty jako v podstatě nezávislé celky, každá obsahuje přehled současného stavu a vývoje dané technologie (často psaný poměrně populární formou), dosavadního výzkumu v oblasti, formulaci hypotéz, popis metod, prezentaci a diskuzi výsledků.

Pátá kapitola je věnována binaurálnímu zvuku. Úvod do problematiky je stručný a zjednodušený, což je pochopitelné vzhledem k rozsáhlosti dané problematiky a omezenému rozsahu práce. Výzkum v oblasti HRTF je velmi rozsáhlý, příklady uvedené v práci jsou opravdu jen příklady. Vlastní experiment byl věnován nikoli přímým binaurálním nahrávkám, ale ukázkám vytvořeným prostřednictvím binaurálních pannerů, což je v pořádku, protože tímto způsobem je binaurální audio často vytvářeno. Samozřejmě, získané výsledky mohou být ovlivněny volbou panneru a zejména danou ukázkou. Pro test byly použity dvě ukázky rozdílného charakteru, bohužel z jejich popisu není možné posoudit jejich přesný obsah. Respondenti byli rozděleni do dvou skupin, jedné byly prezentovány dvě ukázky pouze v binaurálním formátu, druhé pouze ve stereo formátu. Respondenti vyplňovali dotazník (IPQ) po poslechu každé ukázky. Rozdělení respondentů do dvou skupin je pochopitelné vzhledem k organizaci testu, nicméně srovnání obou skupin pak může být problematické. Respondenti hodnotí ukázky absolutně, z výsledků je zřejmé, že na škále 0–6 se pohybují někde kolem středních hodnot – ale je otázkou, zda by stejně odpovídali, kdyby měli přímé srovnání obou technologií. Bylo by zajímavé zjistit, zda by se na výsledcích nějak projevilo, kdyby jedna z ukázek byla prezentovaná ve stereu a druhá binaurální. Výsledky experimentu nepotvrdily předpokládanou hypotézu. Autor pro tento fakt nabízí několik důvodů, které jsou samozřejmě relevantní. Zvážil bych ale ještě vliv samotné metodiky testu – dvě oddělené skupiny, každá zkoumala jinou technologii, vliv složení skupin, vliv použitých sluchátek (případně i vliv ukázky, použitých

pluginů. . .).

Šestá kapitola byla věnována vlivu zvuku ve virtuální realitě. Po úvodu a přehledu stavu problematiky se autor opět věnuje formulaci hypotéz a návrhu vlastního experimentu. Experiment byl velmi náročný, časově, technologicky i organizačně. Opět byli posluchači rozděleni do dvou skupin, jedna se zvukem, druhá bez zvuku. Obě skupiny hrály stejné hry ve VR i na monitoru. Z výsledků je zřejmé, že se prokázal vliv platformy (VR vs monitor), vliv přítomnosti zvuku sice průměrné hodnoty také naznačují, ale není průkazný. Nicméně toto bych opět považoval za vliv zvolené metodiky. Každý respondent měl možnost porovnat obě platformy, nicméně hru hrál jen v jedné variantě (bez zvuku/se zvukem), a proto patrně jeho odpovědi byly ovlivněné hlavně platformou. Je zvláštní, že by nepřítomnost zvuku u hudební hry neměla vliv na pocit flow. Součástí experimentu byl i sběr objektivních dat, tepové a dechové frekvence. Ohledně zpracování naměřených objektivních dat bych uvítal více informací. Pokud jsou tyto informace dostupné, tak by bylo zajímavé sledovat jednotlivé průběhy v čase. Způsob výpočtu finálních parametrů by bylo vhodné popsat matematickým vztahem. U relativních odchylek není zřejmý přesný algoritmus výpočtu a to, zda nejdříve byla vypočtena odchylka (nebo poměr klidové a sledované hodnoty) jednotlivých respondentů a pak to bylo zprůměrováno, nebo naopak. Za výsledek hodný pozornosti pokládám, že významný rozdíl v tepové frekvenci mezi variantou se zvukem a bez zvuku byl přítomen jen při hraní na monitoru.

Sedmá kapitola práce je věnována systémům prostorového zvuku v kině, konkrétně rozdílu mezi formátem 5.1 a Dolby Atmos. Po nezbytném úvodním přehledu se věnuje návrhu a popisu vlastního experimentu. Popisovaný experiment byl extrémně náročný na organizaci a přípravu. Experimentu se zúčastnil vysoký počet respondentů (127). V tomto případě byl test uspořádán jako slepý A-B test, tj. přímé porovnání dvou variant. Bohužel autor nespécifikuje délku ukázek a časový odstup mezi prezentací jednotlivých variant. Jak autor prezentuje v jiné části práce, v případě delších ukázek a větších odstupů experiment zkoumá spíše sluchovou paměť respondentů, což může být dle mého názoru také důvod v rozdílech mezi skupinou zkušených a nezkušených posluchačů. Experiment zkoumal celkem pět atributů. Největších rozdílů mezi formáty bylo dosaženo v atributu lokalizace a experiment prokázal, že celkově posluchači jsou schopni formáty rozpoznat a (mírně) preferují Dolby Atmos. Zajímavý je rozdíl mezi AV a audio ukázkami. Autor nabízí dvě vysvětlení, jednak že u AV ukázek je pozornost posluchačů více strhávána obrazem než zvukem, jednak, že audio ukázky byly vytvořeny přímo pro potřeby výzkumu a proto lépe reprezentují dané rozdíly. Osobně považuji druhou možnost za velmi pravděpodobnou, zvláště pokud, jak autor píše, se tvůrci staví k využívání možností objektového zvuku spíše zdrženlivě. Proto rozdíly nemusí být dané samotnou technologií, ale způsobem, jakým s ní tvůrci pracují. V tomto kontextu dává smysl i výsledek, že největší rozdíl byl zaznamenán v lokalizaci, a např. v obklopení je rozdíl méně výrazný. Možná by bylo zajímavé do srovnání zařadit i oficiální propagační ukázky Dolby Atmos, které se snaží využívat možností formátu.

Použité metody a splnění stanoveného cíle

Pro zhodnocení metod použitých v práci je potřebné připomenout, že obecně hodnocení kvality zvuku je velmi komplexní a náročný úkol. Pro určité aplikace (např. hodnocení kvality algoritmů ztrátové komprese) existují standardizované metody, ale

i v takovém případě je vytvoření a správné vyhodnocení subjektivních testů velmi náročné. Základním problémem je, že lidské vnímání je velmi adaptabilní a schopnost absolutního hodnocení je omezená. Proto je potřeba zajistit takové podmínky, aby respondent skutečně hodnotil jen zkoumaný atribut a vliv dalších atributů byl eliminován. A klíčovou bývá tréninková fáze experimentu, která umožňuje respondentům seznámit se s metodikou testu. Nicméně v případech zkoumaných autorem na sebe narážejí často protichůdné požadavky. To se týká např. délky ukázek, kdy by bylo ideální použít krátké ukázky a porovnávat jednotlivé technologie, nicméně to pro zkoumání konceptů prožitku není možné, protože prožitek vyžaduje určitý čas. Dalším problémem je výběr ukázek, kdy pro potřeby testování je ideální velké množství ukázek pokud možno emočně neutrálních (aby respondent nehodnotil jak se mu ukázka „líbí“ po obsahové stránce), a respondentovi neznámých – což opět zde nebylo možné, protože prožitek nemůže být emočně neutrální, navíc výběr ukázek byl limitován dostupností ukázek pro dané platformy a formáty.

Autor se tedy musel často rozhodovat pro řešení, která sice nebyla ideální, ale byla nejlepší z dostupných. Tím např. bylo rozdělení respondentů pro první a druhý experiment do dvou skupin, které prováděly hodnocení za odlišných podmínek (v prvním případě stereo/binaurální, ve druhé hraní her se zvukem/bez zvuku), aniž by měly referenci k druhé možnosti. Toto podle mého názoru mohlo ovlivnit výsledky a zapříčinit, že se nepotvrdily některé hypotézy. Pro zpracování získaných dat autor používá vhodných a obvyklých statistických metod a správně interpretuje jejich výsledky.

Stanovený cíl, tj. prozkoumat tři současné technologické fenomény spjaté s rozdílnými způsoby zvukové reprodukce a jejich vliv na divácké vnímání byl nepochybně splněn.

Použití odborné literatury

Zvolené téma je velice široké, resp. existuje k němu literatura z rozličných oblastí. Autor musel shromáždit a prostudovat informace ve velmi širokém rozsahu. Kladně hodnotím práci s literaturou i velké množství odkazované literatury – cca 170 pramenů.

Formální stránka

Práce je psaná velmi čtivě a srozumitelně. Občas dochází ke směšování odborného a téměř hovorového jazyka, ale není to na úkor srozumitelnosti práce. Drobným stylistickým nedostatkem je, že autor se občas nedrží terminologie kterou zavedl (např. na str. 23 mluví o ludickém zapojení, ale v textu níže používá synonymum „herní“). Občas (např. v 1.1.1) považuje nadpis oddílu za část textu, což může být nepřehledné pro čtenáře kteří nadpisy nečtou. Pro jednotku času používá nesprávný název vteřina. V části vyhodnocení jednotlivých hypotéz se autor odvolává na hypotézu jen číslem, bylo by vhodné hypotézu zopakovat, aby čtenář nemusel listovat zpět a hledat čeho se hypotéza týká. Určitě by bylo možné také vylepšit prezentaci výsledků, výsledky subjektivních experimentů 1 a 2 jsou prezentovány především pomocí tabulek, grafická reprezentace by byla přehlednější. U grafů 1–5 prezentujících objektivní výsledky je poněkud matoucí spojování nesouvisejících dat pevnou

čarou. U grafů 6 a 8 by bylo dobré, kromě průměrných hodnot, vyznačit i jejich rozptyl, jako je tomu u grafů z objektivních dat.

Z hlediska sazby jsou poněkud nevýrazné nadpisy oddílů, zvláště u druhé a třetí úrovně (menší písmo než vlastní text). Dále jsou přítomny drobné typografické prohřešky, například nevhodný řádkový zlom (často i po jednopísmenných předložkách a spojkách), popisky tabulek pod tabulkami místo nad nimi apod. Formátování seznamu literatury odpovídá normě ISO 690. Významnou část tvoří online zdroje (včetně Wikipedie), ale to považuji vzhledem k tématu práce za akceptovatelné. Na literaturu je odkazováno prostřednictvím poznámek pod čarou, zdroj je vždy jasně zřejmý.

Otázky k obhajobě

1. Na str. 65 je uvedeno, že ideální pro binaurální audio jsou vložná sluchátka (in-ear). Nicméně pro experiment s binaurálním audiem byla použita uzavřená sluchátka Bose 700 – mohlo toto mít vliv na výsledek testu? Mohla nějakým způsobem ovlivnit výsledek funkce ANC? Jsou dostupné informace o funkci ANC, tj. zda dochází např. k potlačení externího hluku v každém kanále nezávisle?
2. Diskutujte možný vliv zvolené metodiky na výsledky – metoda bez reference a oddělené skupiny pro hodnocení různých technologií. To se týká prvního a druhého experimentu. Jak mnoho je možné zobecnit výsledky, získané pro konkrétní technologii (např. konkrétní binaurální panner), konkrétní ukázky a konkrétní skupinu posluchačů?
3. Skupiny byly u prvního a druhého experimentu rozděleny náhodně, podle času příchodu. – Podařilo se při náhodném dělení udržet vyrovnané skupiny (věk, pohlaví, zkušenost...)? Je nějaký rozdíl ve výsledcích u druhého experimentu pokud respondent začínal hrát ve VR a pokud začínal hrát na monitoru? (Odpovídají si průměry a rozptyly těchto podskupin?) I u třetího experimentu byla shromažďována data umožňující jemnější dělení do skupin, je možné např. nalézt nějaké rozdíly mezi studenty a profesionálními zvukaři? Nebo podle věku? (S věkem většinou dochází k nárůstu práhu sluchu na vyšších frekvencích i zhoršení kognitivních schopností, což může být kompenzováno větší zkušeností.)
4. Konkretizujte prosím podmínky objektivního měření a zpracování naměřených dat u druhého experimentu. Pokud jsou dostupné údaje o změnách dat v průběhu hry, je z nich možné něco usoudit? Při sledování jednotlivých/vybraných respondentů, je i zde vidět pokles hodnot vůči klidovému?
5. U třetího experimentu prosím konkretizujte délky ukázek a časový rozestup mezi dvojicí ukázek – pokud byly ukázky na různých DCP tak pravděpodobně nebylo možné je reprodukovat těsně za sebou. Byl stejný rozestup i u zvukových ukázek, které byly na stejném DCP?
6. Lze dát nějak do souvislosti rozdíly v LEQ (tabulka 16) a výsledky pro atribut dynamiky? Zajímavý fakt je, že pro AV ukázky jsou vyšší hodnoty LEQ pro formát 5.1. a u audio ukázek u Dolby Atmos.

Původnost a přínos práce

Autor předložil komplexní a nepochybně původní práci, shrnuje současný stav a přináší nové výsledky. Nezpochybitelný přínos práce je v uspořádání velmi náročných experimentů prezentovaných v druhé části práce. Vzhledem k množství shromážděných dat by určitě bylo přínosem, kdyby byla dostupná odborné komunitě pro další výzkum. Za zvlášť cenné pokládám poznatky ohledně porovnání systému prostorového zvuku v kinech. Kromě samotných výsledků je práce také důležitým příspěvkem k samotné problematice subjektivního testování.

Závěr

Vzhledem k výše uvedenému, i přes zmíněné drobné nedostatky, prezentovaná práce jednoznačně prokázala autorovu schopnost samostatné tvůrčí činnosti v dané oblasti výzkumu. Tuto schopnost dokládá i celková publikační aktivita autora. Prezentovaná práce splňuje požadavky standardně kladené na disertační práce. Práci doporučuji k obhajobě.

V Praze, 13. ledna 2022

Ing. František Rund, Ph.D.
katedra radioelektroniky
FEL ČVUT v Praze
Technická 2
166 27 Praha 6