

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Využití hlasového ovládání ve VR
Jméno autora:	Vít Gardoň
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačové grafiky a interakce
Oponent práce:	Ing. Michal Mýlek
Pracoviště oponenta práce:	MAMA AI (~mimo ČVUT)

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	průměrně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zvažoval jsem, zda přeci jen nepovažovat zadání za lehčí, ale přesvědčila mne nutnost zvládnout VR plugin pro sledování rukou i ovladačů.	

Splnění zadání	splněno s menšími výhradami
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Relativní mělkost stavového stromu aplikace spolu s velmi nízkým počtem položek v každém okamžitém stavu (max. 5) podle mého nedovolují demonstrovat užitečnost hlasového ovládání a aplikace je tak designem předurčena spíše k ovládání rukama. Pokud by se podařilo ukázat „one shot entry“, tedy například rovnou vytvořit auto s červenou metalízou jedinou větou, byl bych spokojen. Hlasové ovládání pro tento druh mělkých množin není uživateli preferováno a také se to plně ukazuje v testovací části.	

Zvolený postup řešení	částečně vhodný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Soudím, že na mnohé nedostatky výsledné aplikace by se přišlo konzultací s testery / uživateli již při designu nebo v průběhu vývoje. Rozhodnutí oddělit ovládání rukou a ovladačem do dvou různých scén považuji za nešťastné. Celkový dojem kazí řada nedokonalostí. Scéna působí znepokojivě, bez podlahy. Spawn hráče je přesně uprostřed auta. Texty na zdech nejsou při pohledu z boku v jedné rovině. Výchozí scéna je ta ovládaná rukama, ale pokud si člověk připraví ovladače, dějí se smutné věci. Dotykovému ovládání by slušelo vylepšení detekce „prstu v tabletu“ vs. „prst se zase vzdálil“. Časový test je příliš nepružný, ale chválím jakoukoliv snahu o zapracování zpětné vazby uživatelů. Volba angličtiny jako jazyka interakce mohla také dost ovlivnit závěry. Text nezmiňuje, zda byli testeři rodilí mluvčí, ale STT modely jsou tradičně velmi citlivé na nedokonalosti výslovnosti. Na druhé straně v kódu vyzdvihují systém vlastních tříd jako abstrakci nad poskytnutým výsledkem NLU. Autor se nespokojil s jedním velkým switch multiplexorem, místo toho (dle mého elegantně) navrhl a implementoval hierarchické třídni řešení.	

Odborná úroveň	B - velmi dobře
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Textová část odborností vesměs obstojí. Občasné faktické diskrepance (např. poskytnuté NLU není postaveno na RASA již od druhé iterace). Také patrně omylem došlo k záměně pojmů text-to-speech (TTS) a speech-to-text (STT) v preludiu odstavce 2.1. Sekce 1.1 shrnující použité technologie je notně zjednodušená, ale na správnou úroveň detailů. Kód scriptingu scény by bylo dobré refactorovat konzistentněji. Hlavičky autora někde chybí, není občas jasné, zda je předkládán kód převzatý nebo autorský. Komentáře kódu existují, ale jsou to spíše návěští než vysvětlení.	

Formální a jazyková úroveň, rozsah práce

D - uspokojivě

Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku.

Text práce je psán česky, abstrakt též v AJ. Hůře se čte, dalo mi práci sledovat autorovu myšlenku v množství stylistických a formulačních chyb, gramatických nedokonalostí a častých doslovných zdvojení stejných frází (např. 2.1, AudioSaveManager vs. Nexus). Moc by pomohlo ještě jedno kolo korektury.

Přál bych si více obrázků, zvláště v Kapitole 5 věnované výsledkům testování, kde je jediná tabulka a žádný graf. Těžko pak sledovat či prezentovat závěry výzkumu. Jako podklad pro diskusi preferuji kombinaci tabulek, schémat a grafů. Při pracovní prezentaci je často prostor nejvýše desítek sekund a obrázek pak zachrání den.

Typografické potíže se nevyhnuly ani Příloze E.

Výběr zdrojů, korektnost citací

C - dobře

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posudte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Kdybych neměl znalost použité domény, zřejmě by mi předkládané portfolio zdrojů nestačilo. Citacemi se šetřilo na nesprávných místech. Kapitola o testování dle mého necituje žádnou konkrétní metodiku a naměřená data mohla být přehledově uvedena aspoň v příloze.

Kondenzovaná rešerše o STT, NLU a TTS je dostatečně detailní pro bakalářskou práci, i zde bych si ale přál více obrázků, schémat, odkazů do nich a citací.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Výsledná aplikace by podle mých kritérií potřebovala doladit, aby ji bylo možné použít k prezentaci dialogu ve VR. Importované detailní modely aut jsou v kontrastu s billboardovou strohostí zdí a podlahy, texty rozvěšené po stěnách jsou patrně převzaté z naší demonstrační scény a působí rozpačitě. Uživatelův spawn je uprostřed budoucího auta. Bylo by třeba detekovat přítomnost ovladačů namísto přepínání scén. Někdy se v reakci na vstup nic nestane, přestože podle textové zpětné vazby byl výsledek NLU doručen, pravděpodobně občasný bug v reakční logice aplikace.

Aby bylo pro uživatele aspoň teoreticky zajímavé použít hlasový vstup, bylo by třeba zásadně rozkošatit položky menu, nebo umožnit „one shot entry“. Překlopil bych aplikaci do češtiny, v případě, že očekávání uživatelé nejsou rodilí mluvčí, případně poptal specializovaný STT model pro anglicky mluvící české rodáky. Každé procento chybovosti STT se tranzitivně projevuje ve schopnostech dialogu správně reagovat.

Není to nic, co by jedno intenzivní integrační sezení nevyřešilo, ale takto celek působí bohužel nedotaženě a neodzkoušeně.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uvedte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **C - dobře**

Datum: 30.5.2023

Podpis: Michal Mýlek