F3

Fakulta elektrotechnická
Katedra kybernetiky

Bakalářská práce

Využití hlasového ovládání ve VR

Petr Staňo

Leden 2024

Vedoucí práce: Ing. David Sedláček, Ph.D.

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Staňo** Jméno: **Petr** Osobní číslo: **499194**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra kybernetiky**

Studijní program: **Otevřená informatika**

Specializace: **Základy umělé inteligence a počítačových věd**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Využití hlasového ovládání ve VR

Název bakalářské práce anglicky:

Voice Control in VR

Pokyny pro vypracování:

1. Seznamte se se způsoby a možnostmi rozpoznávání přirozené řeči využitelné pro virtuální realitu (VR).
2. Navrhněte komponenty pro herní engine Unity, které umožní realizaci nabídky ovládané tradičními VR technikami (např. ukazování, přímá interakce) a současně hlasem (např. tradiční 2D nabídky nebo 3D menu).
3. Navrhněte a implementujte testovou VR scénu, na které demonstrujete použitelnost implementovaných komponent.
4. Navrhněte testové scénáře, kdy bude práce uživateli nějak ztížena, např. časová tíseň, nepřehlednost, omezený stupeň interakce rukama.
5. Metodami uživatelského testování porovnejte tradiční techniky a ovládání hlasem.
6. Zhodnoťte preferované způsoby ovládání. Pro realizaci hlasového ovládání a VR aplikace se omezte na VR platformu Meta Quest, pro rozpoznávání řeči použijte knihovnu Mama-AI (https://themama.ai/).

Seznam doporučené literatury:

[1] Jason Jerald, The VR Book: Human-Centered Design for Virtual Reality. 2015. Association for Computing Machinery and Morgan & Claypool, New York, NY, USA.
[2] Joseph J. LaViola, Jr. et all. 3D User Interfaces: Theory and Practice, second edition. 2017. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA.
[3] Dan Jurafsky and James H. Martin, Speech and Language Processing, 3rd ed. draft. 2023. dostupné online: https://web.stanford.edu/~jurafsky/slp3/

Jméno a pracoviště vedoucí(ho) bakalářské práce:

Ing. David Sedláček, Ph.D. katedra počítačové grafiky a interakce FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **08.02.2023** Termín odevzdání bakalářské práce: **09.01.2024**

Platnost zadání bakalářské práce: **22.09.2024**

Ing. David Sedláček, Ph.D.	prof. Ing. Tomáš Svoboda, Ph.D.	prof. Mgr. Petr Páta, Ph.D.
podpis vedoucí(ho) práce	podpis vedoucí(ho) ústavu/katedry	podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Student bere na vědomí, že je povinen vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací.
Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

. _____
Datum převzetí zadání

_____
Podpis studenta

Poděkování / Prohlášení

Rád bych poděkoval vedoucímu Ing. David Sedláček, Ph. D. a Michalu Mýlkovi za pomoc a za jejich rady při implementaci této bakalářské práce.

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze, 8.ledna 2024

Abstrakt / Abstract

Tento dokument ukazuje implementaci komponenty pro engine Unity 3D. Komponenta umožňuje hlasovou interakci uživatele s menu umístěném ve scéně. V rámci bakalářské prace proběhlo i uživatelské testování k porovnání jednotlivých implementací uživatelských rozhraní.

Klíčová slova: menu, NLP, hlasová interakce, Unity 3D.

This document shows an implementation of a component for Unity 3D engine utilizing user voice interaction for menu controll and navigation.

Keywords: menu, NLP, voice interaction, Unity 3D.

Title translation: Voice control in VR

Obsah /

1 Úvod 1

2 Existující implementace hlasového ovládání 2

2.1 Hlasoví asistenti 2
2.2 Radio General 2
2.3 Starship Commander 3

3 Hardware 4

4 Uživatelská rozhraní 5

4.1 2D menu 5
4.2 3D menu 6
4.3 Hlasové menu 7

5 Použité technologie 8

5.1 Oculus Interaction SDK 8
5.1.1 Interakce pomocí paprsku . . 8
5.1.2 Interakce uchopení 8
5.2 Natural language processing . . . 8
5.3 Plugin MAMA AI 9
5.3.1 Nexus 10
5.3.2 ResponseNLP 10
5.3.3 SampleDialog 10

6 Návrh 12

6.1 Komponenta hlasové interakce . 12
6.2 Ukázková scéna 12
6.3 Jazykový model 13

7 Implementace 14

7.1 Implementace komponent uživatelského rozhraní 14
7.1.1 MenuController 14
7.1.2 Choice Dictionary 14
7.1.3 Value Setter 14
7.2 Implementace komponent hlasového zadávání 14
7.2.1 NLP Validator, Entity Validator 15
7.2.2 NLP Input, NLP Submit . 15
7.2.3 NLP Navigator 15
7.3 Implementace ukázkové scény . 15

8 Testování 18

8.1 Uživatel 1 18
8.2 Uživatel 2 19
8.3 Uživatel 3 19
8.4 Uživatel 4 20

9 Závěr 21

Literatura 22

A Instalační manuál 23

A.1 Použití mé NLP komopnenty: . 23

B Trénovací soubor NLU 24

Tabulky / Obrázky

5.1 Ukázková implementace detekce záměru PIZZA 10
6.1 Tabulka záměrů a jejich příkladů 13
8.1 Výsledky z testování s uživatelem 1 18
8.2 Výsledky z testování s uživatelem 2 19
8.3 Výsledky z testování s uživatelem 3 19
8.4 Výsledky z testování s uživatelem 4 20

2.1 Uživatelské rozhraní hry Radio General. Zdroj: `https://store.steampowered.com/app/1011610/Radio_General/` 3
2.2 Uživatelské rozhraní hry Starship Commander. Zdroj: `https://store.steampowered.com/app/598400/Starship_Commander_Arcade/` 3
3.1 Rozvržení tlačítek ovladače Quest 4
4.1 2D menu v prostředí Oculus 6
4.2 Ukázka 3D výběru zbraně ve hře Until You Fall 7
5.1 Ukázka výsledku rozpoznávání záměru 11
6.1 Diagram pro komponenty 12
7.1 Pohled na prvky v testovací scéně 16
7.2 Interakce uživatele mopocí rukou 17

Kapitola 1
Úvod

Virtuální realita (VR) umožňuje vytvářet aplikace schopné umístit uživatele do 3D prostředí. Tato prostředí poskytují uživatelům, oproti zobrazení na obrazovkách počítačů, nové metody interakce a vnímání okolního světa. Cílem VR není modifikovat uživatelovo vnímání okolního světa, ale úplně nahradit okolní prostor virtuálním. K zobrazení světa a interakce s ním je použit headset a nějaké vstupní zařízení podle volby uživatele nebo dostupnosti, více níže 3. Součástí interakcí ve VR jsou i uživatelská rozhraní. Různé aplikace mohou mít odlišné implementace uživatelských rozhraní, což ovlivňuje jejich podobu a způsob interakce. V rámci této práce jsou zmiňovány 3 typy rozhraní blíže popsány níže: 2D, 3D a hlasové. Využití hlasové interakce se v posledních letech rozšiřuje. Hlasové zadávání je využito v mobilních zařízeních, k vyhledávaní na internetu, pro uživatelskou podporu, chatovací aplikace, a i v herním průmyslu. Často může docházet k problémům s rozpoznáním slov při interakci hlasem, což může způsobovat frustraci uživatele. Na druhou stranu může být zadávání hlasem intuitivní a snazší formou interakce, hlavně pro lidi, kteří nejsou s tradičním způsobem zadávání seznámeni.

Cílem projektu je vytvořit sadu komponent pro herní engine Unity3D, umožňující snadnou tvorbu menu reagující na hlasové příkazy od uživatele a použít tyto komponenty v demonstrační aplikaci, ve které bude možné porovnat hlasové ovládání s tradičním způsobem ovládání.

Kapitola 2
Existující implementace hlasového ovládání

V následující sekci jsou zmíněny některé existující aplikace využívající hlasovou interakci s uživatelem.

2.1 Hlasoví asistenti

Hlasoví asistenti využívají hlasové rozpoznávání a natural language processing (NLP), zmíněný níže 5.2, pro rozpoznání záměru uživatele. Vyslovená věta se převede na sadu příkazů, které je zařízení schopné provést a následně se vykonají. Většina hlasových asistentů je také schopná příkaz nebo vykonanou akci přečíst zpět uživateli jako potvrzení vykonané operace. Aktivace naslouchání asistenta probíhá buď stiskem příslušného tlačítka nebo za pomoci Automated Speech Recognition (ASR), který aktivuje asistenta, když uživatel vysloví příslušnou frázi jako např. „Ok, google“, „Hey Siri“ nebo „Hey Facebook“. Tito asistenti se často používají v běžných zařízeních jako jsou mobilní telefony, chytrá zařízení v domácnostech, ovládání světel nebo teploty, apod. Uživatelské rozhraní hlasových asistentů se skládá pouze z textového pole ukazující doposud vyrozuměnou větu a prostoru pro zobrazení výsledku příkazu. V roce 2011 byl uveden Siri od firmy Apple, který se stal prvním moderním hlasovým asistentem. Siri je mobilní asistent a její funkce zahrnují obsluhu telefonních hovorů, posílání textových zpráv, spouštění funkcí telefonních aplikací jako kalendář, poznámky nebo budík nebo vyhledávání na internetu. Amazon Alexa (2014) od firmy Amazon je domácí asistent, který se často používá k automatizaci domácnosti a poskytuje možnosti vývojářům pro vytváření vlastních aplikací přes Alexa Skills Kit [1]. Asistent Hey Facebook (2020) běžící na zařízeních Oculus Quest/Quest 2 umožňuje provádět některé úkony ve virtuální realitě pomocí hlasových příkazů[2]. Dalšími hlasovými asistenty jsou také Cortana (2013) od firmy Microsoft, Google Assistant (2016) od firmy Google. [1]

2.2 Radio General

Radio General je počítačová strategická hra odehrávající se v období druhé světové války. Cílem hráče je předávat příkazy jednotkám na bojovém poli pomocí svého rádia. Odezva jednotek je pouze rádiová a hráč si musí jejich pozice sám značit na mapě pomocí figurek. Hráčovi se průběžně zobrazuje vyslovený příkaz a jeho možná pokračování podle toho, co hráč říká2.1. Hráči mají také možnost zadávat příkazy pomocí myši prostřednictvím 2D menu. Tato možnost usnadňuje hru hráčům, kteří nejsou schopni určitou část příkazu vyslovit nebo jim detekce hlasu nerozumí.

[1] https://developer.amazon.com/en-US/alexa/alexa-skills-kit
[2] https://www.meta.com/help/quest/articles/in-vr-experiences/oculus-features/what-you-can-say-with-voice-commands/

Obrázek 2.1. Uživatelské rozhraní hry Radio General zobrazující použití hlasových příkazů (menu vlevo dole).

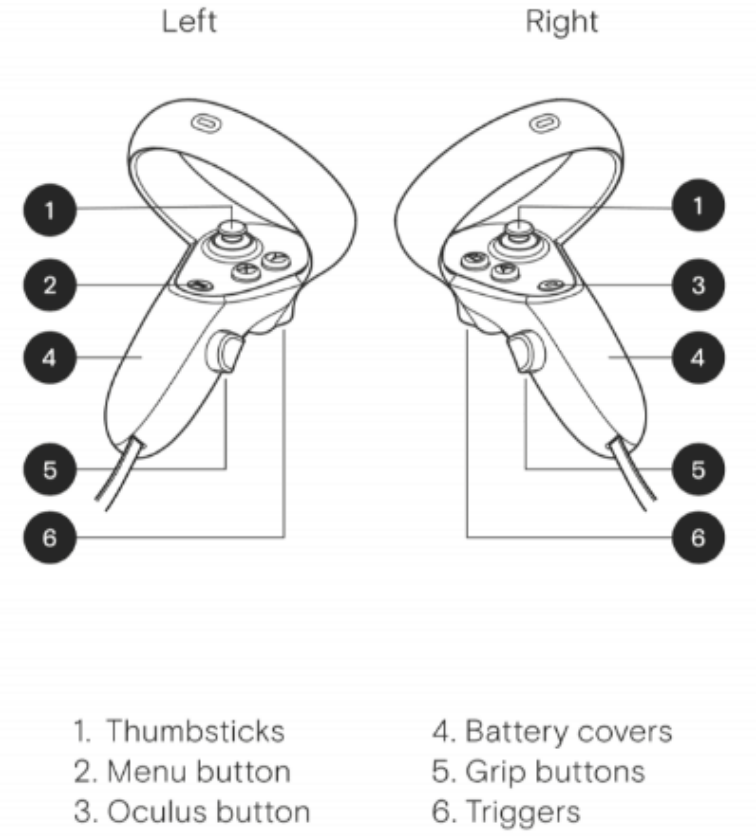
2.3 Starship Commander

Starship Commander je krátká zkušenost pro virtuální realitu trvající asi 15 minut. Hráče je posazen do bitevní vesmírné lodi XR71 a za pomoci palubního počítače se může vydat do bitvy. Hra obsahuje větvící se příběh, který se mění každou hru podle příkazů, které hráč zadá. Loď se ovládá pomocí hlasu a při bojových sekcích příběhu je míření řešeno pohledem hráče, kam se dívá tam je zaměřeno. Na obrázku 2.2 je zobrazena sekce hry, kde počítač čeká na reakci hráče jestli zůstat ve formaci nebo ne. Hra neobsahuje žádné standardní uživatelské rozhraní pro alternativní kontrolu lodi nebo zobrazení aktuálně vyrozuměné věty.

Obrázek 2.2. Uživatelské rozhraní hry Starship Commander v bojové sekci hry.

Kapitola 3
Hardware

Virtuální prostředí je uživateli promítáno skrze headsety obsahující 2 displeje, na kterých je zobrazen obraz pro korespondující oko uživatele, které displej vidí. Obraz prochází čočkami, které slouží k převodu světla z displejů na 3D obraz, který je následně viditelný pro uživatele. Pohyb uživatelé je sledován buď speciálními zařízeními, vysílající světelné informace do headsetu, který podle nich upravuje svou pozici, nebo kamerami na headsetu (inside-out tracking), které snímají okolní svět a podle klíčových bodů zjišťují svou polohu. Vstup od uživatele je získán z ovladačů. Většina základních ovladačů používá jednotný popis a umístění tlačítek viz.3.1 Rozmanitost ovladačů a vstupních zařízení pro virtuální realitu nabízí širokou škálu možností pro uživatele. Existují specializované ovladače nebo nástavce pro specializované aplikace např.: ovladače ve tvaru hudebních nástrojů, rukavice, vesty se zpětnou vazbou a nástavce pro ovladače ve tvaru zbraní pro simulátory. Platforma Oculus poskytuje možnost sledovat a detekovat i fyzické ruce uživatele. V této bakalářské práci byl použit headset Oculus Quest/Quest 2 s funkčností inside-out tracking a Quest ovladači. Inside-out tracking umožňuje zařízení orientovat se v prostoru pomocí kamer umístěných přímo na headsetu.



Obrázek 3.1. Rozvržení tlačítek ovladače Quest

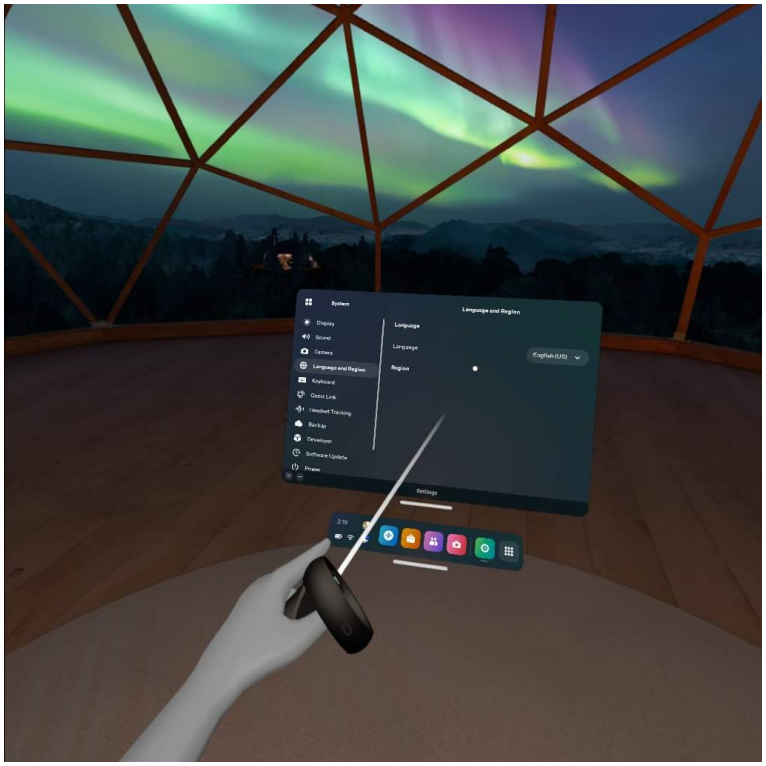
Kapitola 4
Uživatelská rozhraní

Standardní implementace menu ve virtuální realitě je občas nevhodná nebo těžko navigovatelná. Ovladače ve virtuální realitě poskytují přesnou detekci jejich polohy a orientace, toto může způsobit obtížné ovládání virtuálních menu, které jsou umístěny dál od uživatele, protože i mírný třes rukou způsobí, že se požadovaná akce nevykoná. Dalším problémem jsou menu, která mají vysoký počet vnořených prvků. S vyšším počtem prvků roste prostor, který menu zabírá a pro uživatele může být obtížné se v něm orientovat. Prostorná menu jsou nepřehledná a poskytují příliš velké množství informací pro uživatele najednou. Navzdory jeho nevýhodám je standardní způsob interakce stále běžný ve většině aplikací. [2]

4.1 2D menu

2D menu se skládá z jednoho či více 2D panelů umístěných ve virtuálním prostředí, které obsahují text nebo aktivní prvky pro uživatelskou interakci. Interakce s těmito menu obvykle probíhá pomocí dotyku nebo vysílání paprsku. Na obrázku je 2D menu přítomné v domovském prostředí Oculus 4.1. Výhodou 2D menu je jejich podobnost s tradičním uživatelským rozhraním z 2D obrazovek. Pokud menu obsahuje mnoho prvků nebo nestandardní ovládací prvky může být interakce s nimi nepřesná nebo nepříjemná. Velkým problémem se stává zadávání textu, protože virtuální klávesnice musí mít dostatečně velká tlačítka, aby na ně šlo ukázat. Pokud chceme obsáhnout celou standardní klávesnici je její zobrazení velké a přesun virtuálního kurzoru trvá dlouhou dobu. [3]

2D menu existují v různých formách a plní různé účely. Menu může být umístěno u nějakého objektu a pohybuje se ve světě s ním. Některé menu jsou umístěny volně ve světě buď ve fixní poloze nebo s nimi může uživatel pohybovat viz. prostředí Oculus 4.1. Některé menu sledují uživatele a pohybují se společně s ním, podle toho kam se zrovna dívá. Toto chování je vhodné když chceme získat pozornost uživatele a využívá ho například úvodní menu pro nastavení Guardian systému pro Oculus Quest.



Obrázek 4.1. 2D menu přítomné v domovském prostředí Oculus.

4.2 3D menu

3D menu přímo reprezentuje předmět který chce uživatel vybrat. Interakce s prvky rozhraní poté probíhá různými metodami. Od uživatele může být vyžadováno ať předmět uchopí, dotkne se jej nebo na něj ukáže pomocí paprsku. Příkladem tohoto menu je způsob výběru zbraně ve VR hře `Until You Fall`, kde hráč vybranou zbraň uchopí a tím si ji i uloží do inventáře 4.2. Výhodou tohoto rozhraní je, že jednotlivé prvky přímo reprezentují požadavek uživatele. 3D menu mají vyšší prostorovou náročnost oproti 2D a jsou nevhodné pro použití s předměty, které nelze od sebe dobře rozlišit pohledem. Pro lepší vysvětlení akce vyvolané interakcí je vhodné kombinovat 2D a 3D rozhraní. 3D rozhraní reprezentuje předmět, a pokud se uživatel dostatečně přiblíží nebo předmět uchopí, zobrazí se 2D menu obsahující bližší informace nebo popisek výběru. [3]

Obrázek 4.2. Ukázka 3D výběru zbraně ve hře Until You Fall.

4.3 Hlasové menu

Hlasová interakce uživatele s aplikací nevyžaduje zobrazení žádných funkčních prvků. Aplikace naslouchá pro hlasový vstup od uživatele a reaguje na určité podněty. Pro úspěšnou implementaci hlasových rozhraní je nutná znalost cílů aplikace a možných vstupů od uživatelů. Aplikace nemusí reagovat na všechny vstupy ale měla by mít několik možných vstupů pro dosažení stejné interakce. Možnost hlasové interakce je často doprovázena specifickým hlasovým rozhraním. Toto rozhraní poskytuje informace nebo možnosti interakce pro uživatele. Rozeznaná slova mohou být také zobrazována uživateli v textové podobě k poskytnutí zpětné vazby.[3]

Kapitola 5
Použité technologie

Níže popsané technologie jsou využity pro vytvoření hlasem ovládaného menu v enginu Unity 3D. Pro práci se vstupem z mikrofonu uživatele byl použit plugin od firmy MAMA AI, popsán níže.

5.1 Oculus Interaction SDK

Pro implementaci ovládání virtuálního menu je využívána knihovna pro Unity 3D od společnosti Meta: Oculus Integration SDK. Tato knihovna poskytuje několik možností, jak manipulovat s předměty a s menu. Pro manipulaci s menu je použita metoda Ray Interaction popsána níže. Knihovna poskytuje také možnost manipulace s objekty uchopením a zmáčknutím. Všechny typy interakce využívají k zajištění funkcionality stejnou komponentu. Pro umožnění interakce s objekty je nutné k ovladači přiřadit komponent Interactor. Objekt, se kterým pak lze manipulovat musí využívat komponent Interactable [1].

5.1.1 Interakce pomocí paprsku

Interakce pomocí paprsku (Ray Interaction) je metoda manipulace s objekty využívající metody vyslání paprsku z ovladače uživatele. Při kolizi paprsku s objektem, který má přiřazen Interactable a Collider, se zobrazí na povrchu objektu virtuální kurzor a akce výběru je spuštěna při stisku triggeru. Menu využívající tuto metodu musí využívat komponentu PointableCanvas a pro správnou detekci kolizí musí být ve scéně objekt, obsahující PointableCanvasModule.[2]

5.1.2 Interakce uchopení

Interakce uchopení (Grab Interaction) využívá kolize s detekční oblastí ovladače k uchopení a manipulací s objektem. Další komponenty umožňují interakci s fyzikálně aktivními Rigidbody nebo schopnost výpočtu rychlosti a trajektorie, umožňující hození předmětu.[3]

5.2 Natural language processing

Natural language processing (NLP) je způsob analýzy textu za pomoci výpočetních prostředků. Definice NLP je podle [4] soubor metod pro analýzu a reprezentaci lidmi srozumitelného textu za účelu umožnění zpracování daného textu na úrovni lidského posluchače/čtenáře. NLP systém může řešit několik z následujících úkonů, čím více dokáže řešit, tím více se přibližuje ke schopnosti porozumění na lidské úrovni.

[1] https://developer.oculus.com/documentation/unity/unity-isdk-architectural-overview/
[2] https://developer.oculus.com/documentation/unity/unity-isdk-canvas-integration/
[3] https://developer.oculus.com/documentation/unity/unity-isdk-using-with-physics/

Fonologie zpracovává výslovnost slov, ať už samostatně nebo ve skupině s jinými slovy. V NLP systému využívající hlas jako vstup je zvuk analyzován a zakódován a tento signál je předán využitému jazykovému modelu.

Tvarosloví se zabývá skládáním slov pomocí předpon, přípon a kořene, časováním a skloňováním. Význam jednotlivých částí se nemění, a tedy pokud je obdrženo neznámé slovo může se jeho význam odhadnout podle jeho částí.

Větosloví (Syntax) zkoumá závislost jednotlivých slov v textu na ostatních. Větosloví zahrnuje popis vztahů větných členů nebo rozdělení na souvětí a věty jednoduché.

Semantic je vrstva na které se zpracovává smysl slov, který nelze rozlišit na slovníkové úrovni. Tato vrstva zkoumá závislosti z předchozích vrstev, aby ke každému slovu přiřadila právě jeden význam.

Discourse vrstva umožňuje analýzu celého textu. Věty nebere jak samostatnou část textu, ale kouká na ně jako na celek. Na této vrstvě dochází ke změně zájmen na objekty, které reprezentují.

Pragmatika se stará o porozumění slovům, které potřebují pro jejich rozlišení extra kontext nedostupný z textu samotného. Vrstva bere v potaz i záměry a strategie mluvčího a důsledky textu. Porozumění na této úrovní lze reprezentovat větou: „Fotka se nevešla na polici, protože byla příliš velká.“. V druhé větě člověk z analýzy textu nejspíš pochopí že „příliš velká“ byla fotka, ale počítač, který si nedokáže uvědomit souvislost, že větší police by nijak neovlivnila schopnost fotky se do té původní vejít by tuto větu nevyhodnotil správně. [4]

5.3 Plugin MAMA AI

Pro vývoj klasifikace vět byl poskytnut plugin od firmy MAMA AI pro interakci s jazykovým modelem. Jazykový model je založen na frameworku RASA[4] pro vytváření hlasových asistentů. Plugin poskytuje možnost vytvoření vlastního modelu pro klasifikaci záměru a entit z textu. Pro trénování modelu je nutné poskytnout soubor ve formátu YAML obsahující názvy záměrů, ukázkové věty a vyhledávací strategii pro extrakci entit, níže je příklad implementace pro záměr objednání pizzy s možnostmi výběru mezi sýrovou, salámovou a olivovou 5.1. Plugin obsahuje několik tříd pro obsluhu jazykového modelu.

```
- intent: BURGER
    examples: |
      - Dal bych si [dvojitý](type) burger.
      - Chtěl bych hamburger.
      - Chtěl bych [slaninový](type) hamburger.
      - Hamburger prosím.
      - Mám chuť na [sýrový](type) burger.
      - Dám si hamburger, prosím.
      - [dvojitý](type) burger.
      - [sýrový](type) burger.
      - [slaninový](burger-type) burger.
    - lookup: type
      examples:
      - dvojitý
      - sýrový
      - slaninový
```

[4] https://rasa.com/docs/

Tabulka 5.1. Ukázková implementace jazykového modelu pro detekci záměru PIZZA.

5.3.1 Nexus

`Nexus` slouží k interakci s NLP. Při načtení scény se z konfiguračního souboru vytvoří a nastaví jednotlivé služby potřebné k rozpoznávání. Nexus využívá službu speech-to-text (STT) k převodu audia zachyceného od uživatele na text, který je dál průběžně rozpoznáván jazykovým modelem. Při obdržení výsledku od NLP dojde ke spuštění událostí, které této službě naslouchají. Přidání a odebrání posluchače se zajistí funkcí `Nexus.RegisterOnResult()`. Výsledek NLP je zabalen do třídy `ResponseNLP`. Při povolení hlasové syntéze v konfiguračním souboru je možné využít funkce převodu textu na mluvená slova. Tato funkcionalita může být dobrou zpětnou vazbou pro uživatele nebo nástrojem pro ladění programu ve virtuální realitě. Pro provedení hlasové syntéze slouží funkce `Nexus.syntesize(string)` přijímající jako parametr větu, která má být vyslovena.

5.3.2 ResponseNLP

`ResponseNLP` je datová třída obsahující detekovaný záměr `IntentNLP`. Informace o záměru obsahují jeho název a hodnotu jistoty modelu. Další pole třídy `ResponseNLP` je list `EntityNLP` pro uložení detekovaných entit. Každá entita obsahuje svůj název, text, který je k entitě přiřazen z rozpoznávané věty, a hodnotu jistoty modelu.

5.3.3 SampleDialog

`SimpleDialog` a objekt ve kterém je umístěn jsou poskytnuty jako demonstrace použití pluginu. V projektu je využíván jako vizuální feedback pro uživatele při používání hlasového zadávání. Objekt se skládá z několika textových polí ukazující výsledek ze služby STT, finální otázku odeslanou ke zpracování modelem, výsledný záměr a jeho entity 5.1. Při běhu aplikace je v jednotlivých polích zobrazen příslušný výsledek.

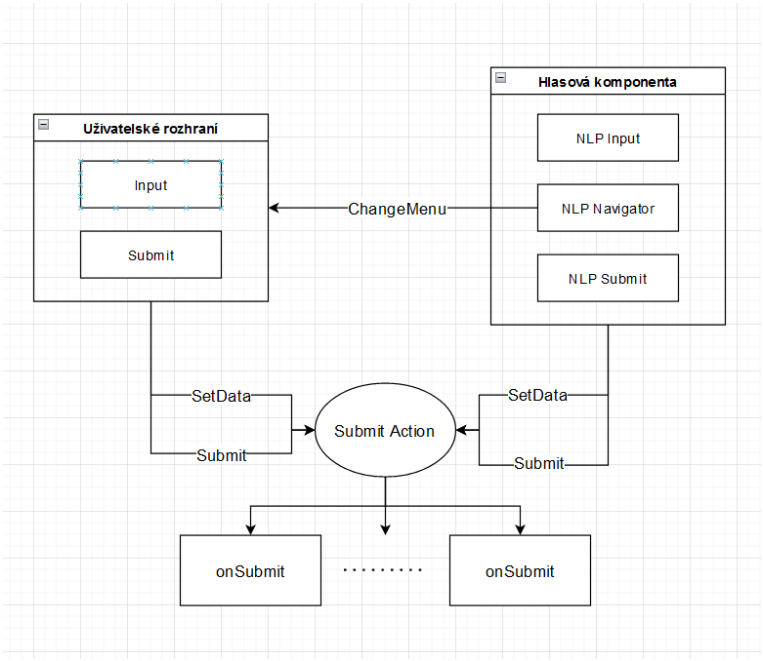
Obrázek 5.1. Ukázka výsledku rozpoznávání záměru.

Kapitola 6
Návrh

V následujících sekcí je popsán návrh jednotlivých částí bakalářské práce.

6.1 Komponenta hlasové interakce

Komponenta musí být schopna reagovat na události přicházející od NLP rozhraní, zpracovávat příchozí záměry uživatele a entity a poskytovat možnost interakce s těmito daty v dalších objektech scény. Zároveň by rozpoznání záměru mělo poskytovat interface pro vizuální indikaci uživateli. V projektu je uvažována práce s hlasovým rozhraním jako součástí menu připomínající webový formulář. Menu by si mělo udržovat hodnoty svých vstupních polí v podobném formátu jako komponenta NLP a po odesílací akci poskytne zadaná data na výstupu. Hlasové rozhraní bude možné využít samostatně nebo napojené na existující uživatelské rozhraní. Diagram 6.1 naznačuje schopnosti komponent manipulovat s menu umístěnými ve scéně.



Obrázek 6.1. Diagram funkcionality komponent pro práci s menu.

6.2 Ukázková scéna

Pro ukázku schopnosti komponent reagovat na hlasový vstup uživatele bylo vybráno prostředí restaurace. Uživatel bude mít možnost objednat si z několika nabídnutých předmětů jídla. Scéna bude obsahovat výše zmíněné typy uživatelských rozhraní, které budou reagovat na vstup od uživatele. Pro otestování rozdílů v jednotlivých rozhraní

bude scéna obsahovat hluboké menu, které vyžaduje větší počet interakcí od uživatele a mělké menu, které je rozsáhlejší a obsahuje všechny nabízené objednávky. uživatel bude moc mezi těmito menu přepínat.

Uživateli bude také ztížena interakce s těmito menu v několika scénářích. První dva scénáře představí podoby uživatelského rozhraní (hluboké a mělké). V dalším scénáři využije uživatel hlasového ovládání k objednání daných položek. V dalším scénáři bude po uživateli vyžadováno použití virtuální klávesnice pro zvolení položek menu. V aplikaci bude také implementována možnost využití interakce rukami. Pomocí rukou bude moci uživatel vybírat předměty i jednotlivé objednané předměty uchopit.

6.3 Jazykový model

Pro detekci záměrů uživatele byl vypracován jednoduchý model rozeznávající jednotlivá jídla a popřípadě jejich druhy v podobě entit. Tabulka ukazuje použité záměry, jejich entity a i ukázkové věty, jak si je uživatel může vyžádat.

Záměr	Entity	Ukázkové věty
Káva	Ledová, Macchiato, Cappuccino	Dám si kávu. Macchiato. Ledovou kávu, prosím.
Pivo		Pivo. Chtěl bych pivo.
Burger	Sýrový, Dvojitý, Slaninový	Mám chuť na burger. Dal bych si sýrový burger.
Hranolky	Malé, Střední, Velké	Ještě si dám hranolky. Malé hranolky.
Pití		Mám žízeň. Chtěl bych něco k pití.
Jídlo		Mám hlad.

Tabulka 6.1. Tabulka záměrů a jejich příkladů ve větě.

Kapitola 7
Implementace

V této kapitole je popsána implementace komponent řešící interakci s menu ve scéně a implementace ukázkové scény demonstrující všechny výše zmíněné typy uživatelských rozhraní. Jsou zde také popsány pomocné třídy pro usnadnění manipulace s datovými třídami vrácené jazykovým modelem.

7.1 Implementace komponent uživatelského rozhraní

V této sekci jsou popsány komponenty, které jsou součástí implementace uživatelských rozhraních. Pro funkčnost menu je vyžadována pouze komponenta `MenuController` a `ValueSetter`. Ostatní komponenty jsou využity při interakci s prvky menu nebo při hlasovém zadávání.

7.1.1 MenuController

Základem implementace uživatelského rozhraní je komponenta `MenuController`. Komponenta obsahuje rozhraní pro uložení hodnot do slovníku přijímací dva `string`y. Tento slovník má stejný formát dat jako hodnoty přicházející od NLU. Slovník je při odeslání menu předán posluchačům této komponenty k dalšímu zpracování. Pro přidání funkcí naslouchající odeslání využívá komponenta třídu `UnityEvent`, která umožňuje přidat události už v editoru a při zavolání definuje parametr `Dictionary<string, string>`. Komponenta má schopnost přepínat aktivní zobrazené položky, které lze přiřadit v editoru společně se jménem pro adresaci menu v navigačních komponentách. Při zavolání `ChangeMenu(string)` dojde k vypnutí aktivního menu a zobrazení menu, které má stejný název jako parametr funkce. Pokud dojde k události odeslání, `MenuController` provede odeslání až v průběhu dalšího snímku, aby nenastal problém odeslání menu před tím, než se všechny potřebné hodnoty nastaví.

7.1.2 Choice Dictionary

`ChoiceDictionary` je využit spolu s aktivním prvkem `TypeSelect` pro ukládání dat z výběrových prvků. Pro správnou funkci standardních rozhraní, když uživatel žádnou interakci neprovede, se při aktivaci menu obsahující tuto komponentu uloží do dat menu aktuální hodnota prvku.

7.1.3 Value Setter

`ValueSetter` je pomocná komponenta ukládající vybrané hodnoty do datové proměnné menu. Komponenta je použita například při stisku tlačítka, aby se uložil záměr uživatele před odesláním.

7.2 Implementace komponent hlasového zadávání

Komponenty, zpracovávající hlasový vstup od uživatele jsou ve scéně umístěny pouze jednou a jejich výsledné akce jsou distribuovány mezi jejich posluchače. Pro správnou

funkci komponent je ve scéně vyžadováno umístění komponenty `NLP Main`. Ta ovládá zda-li je hlasová interakce aktivní nebo ne.

7.2.1 NLP Validator, Entity Validator

Tyto objekty jsou datovými kontejnery pro pravidla, které využívají ostatní komponenty k ověření správnosti/úplnosti příchozích dat. Každý objekt má funkci `Validate()`, která přijímá výsledek od NLP. Validace probíhá kontrolou záměru a pokud je to i dále specifikováno tak dojde i ke kontrole správné entity jinak na entitě nezáleží.

7.2.2 NLP Input, NLP Submit

Třída `NLPInput` po přijetí výsledku hlasového vstupu uloží záměr a entity do slovníku ke zpracování. Pokud příchozí příkaz je kompletní (je rozpoznán záměr i entita, kde je to vyžadováno) tak se příchozí data odešlou všem posluchačům `NLPSubmit`. Odeslání se provede při splnění podmínek jednoho z validačních pravidel.

7.2.3 NLP Navigator

Tato komponenta řídí změnu zobrazeného menu v rámci jednoho uživatelského rozhraní. Při přijetí výsledku NLP je ověřena sada pravidel `NLPValidator`. Pokud některé z pravidel splní podmínky je zobrazené menu změněno podle klíče, odpovídající názvu menu v komponentě `MenuController`.

7.3 Implementace ukázkové scény

Při načtení do aplikace má uživatel možnost seznámit se s implementovaným prostředím. Za uživatelem je zobrazené dodatečné menu sloužící jako administrátorský panel. Tento panel poskytuje možnost změny typu jednotlivých menu a aktivace omezení pro uživatele. 7.1

Ukázková scéna obsahuje dvě uživatelská rozhraní na levé a na pravé straně od uživatele, pomocí kterých si uživatel může objednat některé z nabízených jídel. Při objednání se na stole objeví vybrané jídlo, které reaguje na uchopení od uživatele (uchopení funguje i při interakci rukama 7.2). Jídlem je možné manipulovat, hodit s ním nebo ho sníst. Akce snědení probíhá přiložením jídla k oblasti hlavy uživatele. Při snědení jídlo zmizí a objeví se několik částic indikující provedení akce.

Nalevo od uživatele je standardní 2D menu, které je možné navigovat pomocí ovladačů. Menu obsahuje tlačítka pro navigaci nebo provedení objednávky a vstupní prvky ve formě výběru možnosti.

Napravo od uživatele je zobrazena implementace 3D menu. Jednotlivé prvky reprezentují jídlo, které se při zvolení objedná. Pokud uživatel na předmět ukáže, zobrazí se pomocný text, který obsahuje celý název jídla. Toto menu může být využito pro rychlejší orientaci v názvech jídel i při využití hlasového rozhraní.

Hlasové rozhraní lze aktivovat pomocí admin menu zobrazené za uživatelem. Spuštěné naslouchání je indikované otevřením robotického asistenta. Poté co ze vstupu dojde k rozpoznání záměru provede se příslušná akce.

Při zapnutém omezení zadávání klávesnice je nad každém menu zobrazeno vstupní pole. Při výběru tohoto pole se zobrazí systémová klávesnice. Zadaný text se musí shodovat s položkou na vybraném menu bez diakritiky.

Menu zobrazené před uživatelem spouští testovací scénáře. Při spuštění scénáře je na kulatý stůl u stěny scény položen předmět sloužící jako předloha. Tento předmět je cíl,

který má uživatel objednat ke splnění úkolu. Jednotlivé scénáře jsou popsány výše6.2 a vyžadují správné objednání 6 předmětů. Po skončení scénáře jsou zobrazeny některé statistiky uživatele. Zobrazené informace jsou počet dokončených úkolů, počet chyb při objednávání a průměrný čas ke splnění objednávky.

Jednotlivé předměty mají možné kolize a je možné je na sebe skládat. Menu s popisem 'nejvyšší věž' změří pozici nejvýše umístěného objednaného předmětu ve scéně. Tlačítko 'start' pod menu s popisem 'Zbývající čas' spustí testovací scénář, kde má uživatel 30 vteřin k tomu, aby si objednal co nejvíce předmětů podle předlohy.

Obrázek 7.1. Pohled na prvky v testovací scéně

Obrázek 7.2. Interakce uživatele mopocí rukou

Kapitola 8
Testování

Testování s uživateli probíhalo za mé přítomnosti a během průběhu testů jsem si dělal poznámky na postup uživatelů. Při spuštění aplikace jsem vysvětlil uživatelům ovládání aplikace a nechal je ať si zkusí několik interakcí s prvky testovacího prostředí. Poté, co se seznámili s ovládáním byli uživatelé instruováni ať pomocí menu 'Začít test' spustí testovací scénář. Testovací scénáře požadovali od uživatele 4 úspěšné interakce v každé sekci. První dva testy využívali mělké a hluboké uživatelské rozhraní. V tomto úkolu byli uživatelé požádání o vyzkoušení obou metod interakce s rozhraními; pomocí ovladačů i za pomoci sledování rukou. V dalším úkolu byla dovolena pouze hlasová interakce. Čtvrtý úkol ztížil uživatelům interakci v podobě zadávání pomocí klávesnice, do které je zapotřebí vložit celý název požadované položky(bez diakritiky). Poslední scénář uživatelům umožnil interakci jak pomocí obou menu, tak i hlasovém ovládáním. Po provedení testovacích scénářů jsem se uživatelů dotázal na dodatečné informace a jejich pocity z testování. Uživatelů jsem se ptal na jejich předchozí zkušenosti s VR a počítačem a jak by ohodnotily jednotlivé typy rozhraní.

8.1 Uživatel 1

Po načtení do aplikace se uživatel rychle seznámil s ovládáním. Před začátkem scénáře se chvíli bavil stavěním věže z objednaných jídel. V průběhu testů uživatel poznamenal, že možnost výběru předmětů pomocí seznamu ve formě `drop-down` ve 2D menu je těžko ovladatelná z důvodu velikosti tohoto prvku. Při ovládání hlasem neměl uživatel ve většině objednávek žádné problémy, pouze při objednávání položky 'střední hranolky' musel požadavek několikrát opakovat. Následující testy s klávesnicí a ovládání rukama uživateli nedělaly žádné potíže.

Po testování uživatel preferoval jako metodu interakce mělké 3D menu. Podle uživatele je toto menu přehledné a rychle se v něm orientuje. Další varianty menu (jak 2D tak 3D) byly pro uživatele také vhodné metody interakce. Hlasové zadávání se uživateli nelíbilo, protože i když fungovalo dobře, tak potřebného výsledku dosáhl rychleji pomocí klasického menu.

Předchozí zkušenosti s VR: méně než 1 hodina používání

Práce s počítačem: denně

Úkol	počet chyb	průměrný čas
Mělké menu	0	4,2s
Hluboké menu	2	7,8s
Hlasové zadávání	12	16,7s
Zadávání klávesnicí	0	25,0s
Práce s rukami	0	6,3s

Tabulka 8.1. Výsledky z testování s uživatelem 1

8.2 Uživatel 2

Uživatel 2 měl již s VR zkušenosti a úkoly s interakcí pomocí ovladačů i rukou mu nedělali žádný problém. Uživatelovi se více líbila varianta 3D menu se zobrazenými předměty. Při hlasovém zadávání se mu nelíbila doba odezvy a nepřítomnost zpětné vazby, jestli jsou jeho příkazy rozeznány.

Po testování uživatel také preferoval jako metodu interakce mělké 3D menu. Hlasové zadávání se uživatel popsal jako obtížné.

Předchozí zkušenosti s VR: občasně využívá pro zábavu

Práce s počítačem: denně

Úkol	počet chyb	průmerný čas
Mělké menu	0	3,8s
Hluboké menu	0	5,4s
Hlasové zadávání	8	21,4s
Zadávání klávesnicí	2	17,3s
Práce s rukami	0	5,7s

Tabulka 8.2. Výsledky z testování s uživatelem 2

8.3 Uživatel 3

Uživatel 3 byl nezkušený jak s používáním VR a počítač používal pouze příležitostně. Tomuto uživateli trvalo déle se s ovládáním seznámit. Během testování se často stávalo, že se požadovaná položka na 2D menu neobjednala, protože při stisku triggeru ovladače, uživatel pohnul rukou a sjel tím z tlačítka a interakce se nezaregistrovala. Vzhledem k této obtížnosti s ovládáním trvala interakce s klávesnicí delší dobu než s ostatními uživateli. Při hlasovém zadávání měl uživatel méně potíží, ale nelíbilo se mu, že nevidí název požadované položky během testovacího scénáře. Interakce s rukami byla pro uživatele více intuitivní, ale uživatel měl problém pochopit, jak správně provést gesto pro uchopení předmětu k interakci.

I přes to, že uživatel měl problémy s 2D menu a s ovládáním ve virtuální realitě preferoval uživatel tradiční uživatelské rozhraní před hlasovém. Nejvhodnější metoda objednání bylo opět 3D mělké menu. Pro hlasové zadávání se uživateli nelíbilo, že se musel občas opakovat. Při požadavku o zhodnocení menu na stupnici 1-10 zvolil uživatel 8 pro standardní ovládání a 5 pro hlasové.

Předchozí zkušenosti s VR: žádné

Práce s počítačem: málo, asi 1x za měsíc

Úkol	počet chyb	průměrný čas
Mělké menu	4	44,9s
Hluboké menu	6	53,4s
Hlasové zadávání	5	42,2s
Zadávání klávesnicí	3	72,1s
Práce s rukami	4	47,2s

Tabulka 8.3. Výsledky z testování s uživatelem 3

8.4 Uživatel 4

Uživatel 4 měl stejně jako uživatel 1 předchozí zkušenosti s VR a rychle si osvojil ovládání. Během testovacích scénářů poznamenal, že obrázky (3D menu) jsou jednodušší. Když uživatel pracoval s hlasovým ovládáním řekl, že je to pomalé obtížnější než standardní menu. Při práci s klávesnicí se uživateli nelíbil zdlouhavý proces zadávání.

Podobně jako předchozí uživatelé preferoval k zadávání 3D menu a hlasové příkazy využil pouze když bylo potřeba k testovacímu scénáři.

Předchozí zkušenosti s VR: mírné používání 2x za měsíc

Práce s počítačem: denně

Úkol	počet chyb	průměrný čas
Mělké menu	1	4,3s
Hluboké menu	2	6,9s
Hlasové zadávání	8	16,7s
Zadávání klávesnicí	0	33,0s
Práce s rukami	0	23,7s

Tabulka 8.4. Výsledky z testování s uživatelem 4

Kapitola 9
Závěr

V rámci práce jsem vytvořil komponenty umožňující integraci hlasového ovládání ve virtuální realitě. Komponenty mi při vývoji ulehčili vývoj uživatelských rozhraní, velkou výhodou je možnost interakce s objekty v přímo v editoru Unity. V rámci práce jsem provedl testování uživatelských rozhraní s uživateli. I když jsem si myslel, že méně zkušení uživatelé budou preferovat snazší hlasové ovládání oproti jiným metodám, které vyžadují více interakcí, většina uživatelů preferovala vybírání předmětů z nabídky. Při testování uživatelé preferovali pohybovat se ve 3D menu se všemi možnostmi pro objednaní zobrazenými. V ostatních typech rozhraní nebyl pro uživatele větší rozdíl, kromě výběru z možností typu 'dropdown', který byl pro těžší ovládat. Hlasové ovládání nebylo uživateli preferováno z důvodu delší doby odezvy a pohodlnější možnosti zadávání. Pro vyšší četnost využití hlasového zadávání by na základě zpětné vazby od uživatelů bylo zapotřebí buď ztížit práci se standardním menu nebo zobrazení názvů předmětů nebo příkazů, které jsou uživateli dostupné.

Literatura

[1] *Historie hlasových asistentů.* https://voicebot.ai/2017/07/14/timeline-voice-assistants-short-history-voice-revolution/.

[2] Jason Jerald. *The VR Book: Human-Centered Design for Virtual Reality.* Association for Computing Machinery and Morgan Claypool, 2015. ISBN 9781970001129.

[3] Yannick Weiss, Daniel Hepperle, Andreas Sieß a Matthias Wölfel. *What User Interface to Use for Virtual Reality? 2D, 3D or Speech-A User Study.* In: 2018.

[4] *Liddy, E.D. 2001. Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc.* https://surface.syr.edu/cgi/viewcontent.cgi?article=1043&context=istpub.

Příloha A
Instalační manuál

Pro zprovozněni projektu je zapotřebí zip soubor rozbalit do zvoleného adresáře. Konfigurační soubor pluginu MAMA AI se nachází na adrese /Assets/Resources/Text/Player/config.txt. Soubor již obsahuje připravené hodnoty pro běh aplikace, ale pokud je zapotřebí konfiguraci nebo přístupové údaje změnit, tak se všechny nacházejí v tomto souboru.

A.1 Použití mé NLP komopnenty:

Ke správné funkci komponent je potřeba ve scéně singleton komponenty `GameManager` s odkazem na objekt s komponentou `Nexus`. Tyto komponenty umožní funkcionalitu NLP rozhraní. Pro interakci s NLP stačí dále zaregistrovat události `GameManager.instance.OnNexusResult` a `GameManager.instance.OnNexusInterrim`. Poskytnuté NLP třídy ve složce /Assets/NLPComponents poskytují snazší zprovoznění interakce s NLP za použití validačních pravidel pro záměry a entity.

Příloha B
Trénovací soubor NLU

```
- intent: BURGER
  examples: |
    - Dal bych si [dvojitý](burger-type) burger.
    - Chtěl bych hamburger.
    - Chtěl bych [slaninový](burger-type) hamburger.
    - Hamburger prosím.
    - Mám chuť na [sýrový](burger-type) burger.
    - Dám si hamburger, prosím.
    - [dvojitý](burger-type) burger.
    - [sýrový](burger-type) burger.
    - [slaninový](burger-type) burger.
  - lookup: burger-type
    examples:
    - dvojitý
    - sýrový
    - slaninový
- intent: SIDE_FRIES
  examples: |
    - Chci [malé](fries-size) hranolky.
    - Chci [střední](fries-size) hranolky.
    - Chci [velké](fries-size) hranolky.
    - Ještě si k tomu dám hranolky.
    - [malé](fries-size) hranolky.
    - [střední](fries-size) hranolky.
    - [velké](fries-size) hranolky.
    - hranolky.
    - Ještě jedny hranolky.
    - S hranolky, prosím.
  - lookup: fries-size
    examples:
    - malé
    - střední
    - velké
- intent: DRINK
  examples: |
    - Chtěl bych něco k pití.
    - Ještě něco k pití.
    - Mám žízeň.
- intent: FOOD
  examples: |
    - Mám hlad.
```

```
    - Dal bych si něco k jídlu.
    - Jídlo.
    - Jídlo, prosím.

- intent: COFFEE
  examples: |
    - Chtěl bych kávu.
    - Chtěl bych [cappuccino](coffee-type)
    - Chtěl bych [macchiato](coffee-type)
    - Chtěl bych [ledovou kávu](coffee-type).
    - Kafe.
    - [cappuccino](coffe-type)
    - [macchiato](coffee-type)
    - [ledovou kávu](coffee-type)
    - Kávu.
    - Kafe, prosím.
    - Dám si kávu.
    - Dám si [cappuccino](coffee-type).
    - Dám si [ledovou kávu](coffee-type).
  - lookup: coffee-type
    examples:
    - macchiato
    - ledovou kávu
    - cappuccino

- intent: BEER
  examples: |
    - Pivo.
    - Chtěl bych pivo.
    - Dám si pivo.
    - Pivo, prosím.
```