

Západočeská univerzita v Plzni

Fakulta filozofická

Diplomová práce

Hypotéza „uncanny valley“ ve vztahu k syntetické řeči

Daniela Tisarová

Plzeň 2014

Západočeská univerzita v Plzni

Fakulta filozofická

Katedra filozofie

Studijní program Humanitní studia

Studijní obor Teorie a filozofie komunikace

Diplomová práce

Hypotéza „uncanny valley“ ve vztahu k syntetické řeči

Daniela Tisarová

Vedoucí práce:

Ing. Mgr. Jan Romportl, Ph.D.

Katedra kybernetiky

Fakulta aplikovaných věd Západočeské univerzity v Plzni

Plzeň 2014

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci zpracovala samostatně s použitím odborné literatury a pramenů, jež jsou uvedeny v seznamu použitých zdrojů.

V Plzni 10. 4. 2014

Daniela Tisarová

Poděkování

Chtěla bych poděkovat především vedoucímu své diplomové práce, Ing. Mgr. Janu Romportlovi, Ph.D., za pomoc a rady, které umožnily realizaci této práce a výzkumu, který je s ní spojený. Děkuji také za nepostradatelné vedení v průběhu vyhodnocování výsledků výzkumu. Dále bych ráda poděkovala Ing. Alešovi Pražákovi, Ph.D. za technickou podporu a doporučení ohledně provedení experimentální části a za korektury v části teoretické. V neposlední řadě děkuji všem účastníkům experimentu, kteří měli odvahu stát se objektem zkoumání.

Obsah

1	Úvod.....	1
2	Teoretická část.....	2
2.1	Syntetická řeč	2
2.2	Syntéza řeči.....	3
2.3	Kvalita syntetické řeči	4
2.4	Využití TTS	5
2.5	Automatické rozpoznávání řeči	7
2.6	Dialogové systémy	8
2.7	Wizard of Oz	10
2.8	The Uncanny Valley	11
2.8.1	Podoba člověka a stroje.....	14
2.8.2	Vztah člověka a stroje	15
2.8.3	Hlas ve vztahu k „uncanny valley“	15
2.9	Shrnutí	16
3	Praktická část	17
3.1	Uvedení do problematiky experimentu	17
3.2	Experiment.....	18
3.3	Nepostradatelné součásti experimentu.....	19
3.4	Experimentální protokol	21
3.5	Metodologie.....	24
3.5.1	Sledování jedinci	26
3.5.2	Rozhovor	27
3.5.3	Dotazník	27
3.5.4	Prostředí	29

3.6	Etické aspekty výzkumu	29
3.7	Hodnocení experimentu	30
3.8	Důvod zrealizování experimentu	34
4	Analytická část.....	35
4.1	Popis výsledků experimentu	35
4.2	Nástroje vyhodnocování	35
4.3	Vyhodnocení výsledků	37
4.3.1	Rozhovory	43
4.3.2	Data z dotazníků	44
4.4	Shrnutí experimentu	45
4.4.1	Problémy	46
4.4.2	Návrhy řešení nedostatků	46
4.4.3	Příčiny nepotvrzení hypotézy	48
5	Závěr.....	50
6	Přílohy	51
7	Seznam použité literatury a pramenů	61
8	Resumé	63

1 Úvod

Na následujících stránkách se pokusím představit stručný přehled teoretických pojmů, jako je syntetická řeč, syntéza řeči, systém převodu textu na řeč, systém automatického rozpoznávání řeči a dialogový systém. Budu usilovat o nastínění současného stavu ve vývoji systémů a o výklad těchto pojmů v kontextu lidského vnímání. Dále se budu snažit objasnit význam hypotézy „uncanny valley“, jak v jejím původním pojetí v souvislosti s neživými entitami, tak ve vztahu k syntetické řeči.

Právě druhá možnost pojetí hypotézy „uncanny valley“ by měla tvořit hlavní konstrukci k výzkumu, který zamýšlím uspořádat pro využití této diplomové práce. Experiment by měl potvrdit nebo vyvrátit, zda se hypotéza „uncanny valley“ dá aplikovat i na syntetickou řeč. V případě, že by se neprokázal vztah hypotézy a syntetické řeči, mohlo by to mít pozitivní vliv na další rozvoj řečových technologií.

Motivací k naplánování výzkumu, který by prokázal vztah teoretické hypotézy a uměle vytvořené řeči, byla snaha nezabývat se jen teorií, nýbrž se věnovat také určité praxi. Předpokladem k provedení experimentů pak bylo, že by výsledky mohly být uplatnitelné v dalším výzkumu, který by se mohl týkat zlepšení přijímání syntetické řeči lidmi, popřípadě obecně pozitivnějšího akceptování nových technologií.

Mým cílem je zjistit, zda lze hypotézu uplatnit omezeně na pouhé zvukové podněty. Zajímá mě, jestli by to mohlo vyvolat podobné účinky, jako jsou popisovány v rámci „uncanny valley“ při spojení vizuálních a auditivních stimulů. Pokud ne, mohly by být syntetické hlasy využívány v daleko širším rozpětí, a to bez nebezpečí negativních dopadů „uncanny valley“.

2 Teoretická část

2.1 Syntetická řeč

Pod pojmem „syntetická řeč“ si většina lidí představí uměle vyrobenou řeč. V 18. století se snaha o výrobu řeči jiného původu než z lidských mluvidel ještě promítala do výroby nejrůznějších mechanických mluvících strojů, akustických rezonátorů a přístrojů, které lidskou řeč jen vzdáleně připomínaly. Od 2. poloviny 20. století prodělaly řečové technologie překotný vývoj a aktuálně se řeč uměle vyrábí výhradně pomocí počítače.¹

Řečová syntéza prodělala výrazný pokrok především proto, že se mnoho lidí snaží umožnit strojům produkovat řeč, popřípadě rozpoznávat lidskou řeč a nějakým způsobem na ni reagovat. Proč se o to ale vlastně snažíme? Nestačí, že si rozumíme my lidé? Za pokrokem řečové syntézy stojí touha vědců přiblížit lidem stroje pomocí nejpřirozenějšího nástroje komunikace mezi lidmi, tj. mluveným slovem. Vize vědců je pak taková, že díky tomuto přiblížení lidí a strojů by se komunikace mezi nimi mohla stát přirozenější a zároveň by mohla být bezprostřední a naprosto běžnou součástí našich životů.²

Jediným cílem syntézy řeči tedy není získání umělé řeči, ale jde nám o takovou řeč, která bude znít téměř jako lidská, která dokáže vyjádřit citové zaujetí mluvčího a emoce a zároveň bude stále srozumitelná a jasná. Tento cíl je stále ještě nedosažený, navzdory obrovským pokrokům, kterých řečové technologie dosáhly.³ Stejně tak se můžeme domnívat, že dokud řeč umělá nedosáhne srovnatelné úrovně přirozenosti s řečí lidskou, bude tento problém stále v oblasti vědeckého zájmu.⁴

¹ PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*, s. 529.

² Tamtéž, s. 529.

³ PIERACCINI, R. *The Voice in the Machine*, s. 191.

⁴ ROMPORTL, J. *Zvyšování přirozenosti strojově vytvářené řeči v oblasti suprasegmentálních zvukových jevů*, s. 19.

2.2 Syntéza řeči

Syntéza řeči, se kterou se nejčastěji setkáváme, je charakteristická tím, že na vstupu do systému máme k dispozici pouze text, který je převáděn na řeč (TTS – Text-To-Speech synthesis). Tento typ řečové syntézy je tím nejdostupnějším způsobem získání umělé řeči, ale zároveň je také nejsložitější počítačovou úlohou při umělé výrobě řeči. Syntetická řeč je vyráběna prostřednictvím syntetizérů řeči, které na základě fonetických a prozodických charakteristik vytváří výsledný řečový výstup.⁵ Nárůst složitosti převodu řeči na text přináší především to, že při konverzi z textu nemáme k dispozici tak detailní charakteristické rysy řeči v podobě přesných fonetických informací (tj. sled fonémů) a prozodických vlastností řeči (jako je například melodie, intenzita a trvání hlásek, výška hlasu apod.). V textu totiž máme k dispozici pouze souslednost slov doplněnou o interpunkční znaménka.⁶

Systém TTS se skládá ze dvou hlavních částí. První část se zabývá zpracováním přirozeného jazyka (NLP – Natural Language Processing). Modul NLP se zabývá analýzou a zpracováním textu před vstupem do systému a poskytuje informace, co přesně se má vytvořit v řečové syntéze. Po zpracování modulem NLP má druhá ze základních částí systému TTS k dispozici sled fonémů s prozodickými značkami. Druhý modul je tvořen syntetizérem řeči, který zpracovává řečový signál, tj. provádí vlastní výrobu řeči.⁷

Ještě předtím, než je syntetizér řeči schopen generovat řeč, musí se naučit jak vůbec číst text. K tomu potřebuje ovládat fonetickou transkripci (převod z grafémů na fonémy). To znamená, že syntetizér musí umět přeložit textové sekvence symbolů do sledu fonetických symbolů za užití odpovídajících fonémů. V tu samou chvíli musí přístroj identifikovat textové značky intonace (jako například interpunkci) a musí je převést do správného lexikálního přízvuku, který náleží daným slovům. Těmto všem činnostem však předchází jedna aktivita, která je v celém procesu převodu textu na řeč stěžejní. Zmíněnou činností je jakési předzpracování textu, které mluvčí provádí přímo v mysli, aniž by o tom musel přemýšlet. Například když máme v textu uvedeno: „Schůzka proběhla 1. 12.“, přečteme numerické hodnoty jako „prvního dvanáctý“, popřípadě „prvního prosince“, nikoli „jedna tečka dvanáct tečka“. Takových typů rozhodování nalezneme v jakýchkoli textech nespočet. Naše rozhodnutí pak většinou spočívá na znalosti kontextu a na interpretaci významu slov. A stejný typ aktivity musí udělat i řečový syntetizér právě ve fázi předzpracování textu před samotným provedením řečové syntézy. Stejně potíže přináší i samotná interpunkce, číslice, zkratky aj.

⁵ PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*, s. 529.

⁶ Tamtéž, s. 614.

⁷ Tamtéž, s. 614.

I kvůli nutnosti textového předzpracování je pak samotný úkol řečové syntézy komplexní a závisí i na typu jazyka; každý jazyk má speciální pravidla, která se musí nějakým způsobem formalizovat.⁸

Syntéza řeči z textu si klade za cíl uskutečnit převod víceméně jakéhokoli textu do řeči takovým způsobem, kterého by byl schopen každý rodilý mluvčí daného jazyka. V ideálním případě by syntéza měla být takové kvality, aby nebyla rozlišitelná od lidské mluvy. Jinými slovy, účelem syntézy řeči je nějakým způsobem učinit z řeči umělé a synteticky vytvořené řeč přirozeně znějící.⁹ Problémem však také zůstává definice spojení „přirozeně znějící řeč“, popřípadě „přirozená řeč“ vůbec. Existuje mnoho způsobů, jak vyslovit hlásku, slovo, větu v lidské řeči a žádný z mluvčích nebo z posluchačů nemá v mysli nějaký ideální vzor toho, jak by promluva měla být vyslovena, aby byla správně srozuměna. Syntetický výrok nemusí mít naprosto precizní výslovnost, a stále může být srozumitelný. Dokonce může syntetická promluva znít nevšedně či zvláštně, ale bude jí rozuměno až do té úrovně, dokud bude spadat do jakési podmnožiny toho, co jsou mluvčí zvyklí běžně slýchat.¹⁰

2.3 Kvalita syntetické řeči

To nejdůležitější, čeho je třeba dosáhnout v oblasti systémů TTS, je produkovat kvalitní řeč. Z toho vyplývají i další nutné náležitosti takové řeči, například její srozumitelnost a přirozenost. Syntetická řeč by neměla být náročná ani na poslech, ani na udržení pozornosti, tudíž by neměla být monotónní. Jakmile se podaří dosáhnout těchto požadavků, může dojít k daleko většímu rozšíření syntetické řeči a zároveň k jejímu lepšímu přijímání i od neoborné veřejnosti.¹¹

Při hodnocení kvality převodu textu na řeč se setkáváme s problémy, které se týkají především subjektivního posuzování výsledků. Možnosti hodnocení jsou poměrně omezené, neboť objektivní metody hodnocení nemáme k dispozici. Často se tedy musíme spolehnout na názory jednotlivců, kteří ohodnotili syntetizované promluvy z poslechových testů. Názory testovaných se pak mohou v mnohém lišit a není snadné nalézt jednotný názor či pohledy mnoha lidí shrnout do společného závěru.¹²

Dalším problémem je také to, že vlastně nemáme jednotný vzor, jak vnímat kvalitní promluvu a jak kvalitní řeč vlastně posoudit. Každý jedinec může mít jiné představy o tom,

⁸ PIERACCINI, R. *The Voice in the Machine*, s. 200 – 203.

⁹ PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*, s. 614.

¹⁰ PIERACCINI, R. *The Voice in the Machine*, s. 196.

¹¹ PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*, s. 633.

¹² Tamtéž, s. 627.

jak zní kvalitní řeč. Obecně však lze říci, že můžeme hodnotit především srozumitelnost umělé řeči a také to, jestli je promluva přirozená. Postupem času se ustálil soubor testů, díky kterému můžeme parametry syntetické řeči navzájem porovnávat. To nás pak může upozornit na chyby, které se v umělé řeči vyskytují, popřípadě na rozdíly či pokroky, kterých syntéza řeči z textu dosáhla. Při hodnocení srozumitelnosti syntetické mluvy je posluchač vyzván, aby určil, jak dobře promluvě rozumí, zda v ní vnímá nepřirozené přechody, pomlky apod. Při testování přirozenosti je pak nutné posuzovat řeč v globálním měřítku, tedy nezaměřovat se na detaily výslovnosti, pauz apod.¹³

Jiná nejasnost, která je svým způsobem nadřazená předcházejícím problémům, je otázka výběru syntetického hlasu. Tento výběr velkým podílem přispívá buď k nadšenému přijímání, nebo k nenávratnému odmítání syntetické řeči. Problémem však je, že nemáme žádný formální, objektivní způsob jak určit, jaký syntetický hlas je lepší než ostatní. Tento aspekt může být posuzován pouze subjektivně. Vědci, kteří se zabývají řečovou syntézou, mají k dispozici statisticky významný počet zkoumaných jedinců, kterým bylo k poslechu poskytnuto dostačující množství zvukových ukázek syntetické řeči. Subjekty tyto ukázky ohodnotily podle jejich srozumitelnosti a podle míry podobnosti s lidským hlasem. Jedině takovým způsobem se lze dostat na obstojně spolehlivé vyhodnocení kvality syntetické řeči.¹⁴

2.4 Využití TTS

Možností, jak využít systémy převodu textu na řeč, je celá řada. Nejmarkantnější význam mají tyto systémy především pro lidi s nejrůznějšími hendikepy. Nejprve nás samozřejmě napadne využití pro jedince nevidomé, kteří díky převodu textu na řeč mohou mít přístup k informacím v textové podobě. Jednou z možností využití je čtení textových zpráv a nejrůznějších textových příkazů na mobilních telefonech, další z variant využití je taktéž automatické předčítání textových zdrojů, jako jsou například knihy, noviny apod. Systémy TTS mohou být potřebné také pro lidi postižené například poruchami hlasu nebo pro jedince, kteří jsou zcela němí. Systémy řečové syntézy mohou v praxi nahradit hlas hendikepovaných, kteří pomocí speciálních klávesnic na nejrůznějších přístrojích napíší text, který se má syntetizovat, a přístroj poté promluvu pronese za ně. Rovněž je možné speciálně upravit

¹³ PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*, s. 627.

¹⁴ PIERACCINI, R. *The Voice in the Machine*, s. 196.

mobilní telefony, skrze které mohou hlasově postižení s využitím řečové syntézy komunikovat i s lidmi, kteří se nenachází v jejich blízkosti.¹⁵

Využívání funkce automatického čtení je nepostradatelné například pro nevidomé, ale může přinášet užitek i pro mnohé jiné účely v oblasti každodenního života běžných lidí. Systémy TTS je možné použít tam, kde je větší výhodou disponovat zvukovou nahrávkou než textem. Může se jednat o informace o odjezdech dopravních prostředků, o již zmíněné čtení zpráv, pošty, knih, což je využitelné nejen pro postižené, ale například i pro starší část populace, pro lidi upoutané na lůžko, pacienty v nemocnicích apod. Tato funkce také může být využita lidmi, kteří potřebují v zaměstnání vykonávat více aktivit současně, a čtení textu by je zdržovalo, nebo dokonce ohrožovalo. Takových profesí je celá řada, od řidičů přes zaměstnance obsluhující složité přístroje nebo manuálně pracující až po dispečery či piloty, kteří využívají nejrůznější varovná hlášení v kombinaci s vizuálními zobrazeními.¹⁶

Systémy převodu textu mají své nezastupitelné využití také v telekomunikačních službách, ve kterých usnadňují přístup k nejrůznějším druhům informací. Už jen při volání na mobilního operátora zakoušíme interakci s hlasovým automatem, což přináší úsporu času i snížení nároků na využití lidských zdrojů v tomto oboru. Dnes už takto můžeme získávat téměř libovolné informace, od zpráv o počasí, dopravě, sportu či kultuře až po osobní informace, například o výsledcích výběrového řízení nebo o svém bankovním účtu apod. Mnohá z těchto technických vylepšení využívají asistenci dialogových systémů, ve kterých se využívá i automatického rozpoznávání řeči; systém obdrží dotaz, rozpozná jej a v dostupných textech systému nalezne odpověď, kterou převede z textu do promluvy.¹⁷

Velice pokrokovým využitím syntézy z textu na řeč je výuka cizích jazyků, která se může stát nepostradatelným pomocníkem hlavně lidem, kteří se učí cizímu jazyku v domácím prostředí. Tento způsob vzdělávací opory je nepostradatelný především pro lidi, kteří nemají možnost komunikace s ostatními mluvčími cizího jazyka. Na využití systémů TTS k výuce jazyka není třeba disponovat syntézou nejvyšší kvality, jelikož ucho mluvčího je citlivé na rodný jazyk, nicméně je shovívavější k ostatním jazykům.¹⁸

V neposlední řadě lze systémy TTS hojně využít i v oblasti zábavy a multimediální komunikace, ať už se jedná o komunikaci člověka a stroje, nebo o touhu pobavit a rozptýlit se nejrůznějšími mluvčími aplikacemi, počítačovými hrami, přístroji či hračkami pro děti.¹⁹

¹⁵ PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*, s. 631.

¹⁶ Tamtéž, s. 632.

¹⁷ Tamtéž, s. 631.

¹⁸ Tamtéž, s. 632.

¹⁹ Tamtéž.

2.5 Automatické rozpoznávání řeči

Automatické rozpoznávání řeči (ASR – Automatic Speech Recognition) je způsob převodu řeči na text pouze na základě mluveného slova a systému, který umí řeč rozpoznávat. Stěžejními částmi systému ASR jsou akustická analýza, akustický model a jazykový model. Dalším důležitým nástrojem je řečový dekodér, který je zodpovědný za rozpoznávání akustických signálů. Systém rozpoznávání identifikuje slova, která byla vyslovena, a zobrazí je formou textu.²⁰ Automatické rozpoznávání je nyní založeno převážně na použití statistických metod, přičemž řečový dekodér určuje výsledné pořadí slov, které je nejvíce pravděpodobné pro daný akustický signál.²¹ K určení největší pravděpodobnosti posloupnosti slov slouží tzv. statistický jazykový model, který je specifický pro každý jazyk. Pravděpodobnost sledu slov v daném jazyce souvisí především s kontextem, ve kterém jsou slova užitá. Pravděpodobný výskyt slov může být také ovlivněn množstvím ustálených výrazů, které se v jazyce vyskytují. Určování pravděpodobnosti zjednodušuje úkol rozpoznávání řeči, jelikož slova po sobě většinou nenásledují chaoticky, ale mají mezi sebou definovaný vztah. Tím je pak možné odhadnout, jak velké jsou vyhlídky, že určitá slova budou následovat za sebou.²²

ASR může být negativně ovlivňováno vnějšími vlivy, které na systém mohou často působit. Mezi vlivy, kvůli kterým je ohroženo rozpoznávání řeči, patří především působení okolí (např. šumy nebo zvuk hovoru v bezprostřední blízkosti systému, popřípadě zhoršená akustika místnosti), dále i nejrůznější charakteristiky řečníka (kvalita řeči, vady řeči, přízvuk, nářečí, způsob a charakter mluvy, rychlost řeči apod.) a nakonec i ovlivnění přenosovým kanálem (problémy mohou způsobovat odlišné vlastnosti různých druhů mikrofonů nebo jejich vzdálenost od řečníka, nepříznivě působí také výpadky signálu a šumy samotného přenosu).²³

Při automatickém rozpoznávání řeči můžeme pozorovat velký rozdíl při zpracování spontánních a předem připravených promluv. Předem připravená řeč, popřípadě řeč čtená má korektní formu a gramatickou stavbu, projev řečníka bývá plynulý. U spontánní řeči se mnohdy setkáváme s nižší mírou plynulosti, spontánní promluva bývá obohacena o prvky, které nelze charakterizovat jako řečové (např. přeréknutí, nedokončená slova, hlasité nádechy,

²⁰ PIERACCINI, R. *The Voice in the Machine*, s. 58.

²¹ Tamtéž, s. 109 – 110.

²² Tamtéž, s. 116 – 117.

²³ PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*, s. 400.

odkašlávání, nesprávná větná stavba apod.), a tím samozřejmě způsobujeme potíže systému rozpoznávání.²⁴

Aktuálně jsou systémy rozpoznávání řeči konstruovány na základě jazykových modelů, které čerpají především z psaných textů, tzn. že systémy nejsou reálně uzpůsobeny na mluvenou řeč. Mimoto akustické modely jsou trénovány na čteném projevu a v tichém prostředí, proto může v kontaktu se spontánním projevem v autentickém prostředí docházet k častějšímu výskytu chyb.²⁵

2.6 Dialogové systémy

Hlasové dialogové systémy umožňují uživatelům komunikovat s počítačovými systémy nebo aplikacemi prostřednictvím hlasových povelů, nebo dokonce přirozenou řečí. Vytvoření určitého rozhraní mezi člověkem a strojem je základním účelem hlasových dialogových systémů. Na jedné straně tedy máme uživatele, který udává příkazy prostřednictvím hlasu, a na druhé straně je k dispozici aplikace, jež je řízena počítačem a je schopná nějakým způsobem reagovat na lidský hlas. Způsob reakcí dialogového systému se liší podle toho, k čemu je systém určený. Existují různé druhy dialogových systémů – od poměrně jednoduchých, které jsou schopny interagovat pouze prostřednictvím omezeného množství odděleně pronesených povelů, až po systémy komunikující s uživatelem prostřednictvím souvislé přirozené řeči.²⁶

Současný state-of-the-art hlasových dialogových systémů však předpokládá, že dialog mezi člověkem a strojem probíhá pouze v určité doméně, tedy že rozmluva se odvíjí na omezené téma (např. rezervační systém vstupenek, jízdenek, či nákup zboží apod.), nelze si zatím představovat plynulý rozhovor na libovolný námět. Aby mohla být komunikace člověka s dialogovým systémem považována za úspěšnou, je třeba zabezpečit „porozumění“, kterého by dosáhl strojově řízený systém. Tohoto „porozumění“ je dosaženo pomocí modulu, který zajišťuje řízení dialogu, včetně modulů rozpoznávání a syntézy řeči. Jeden z nejdůležitějších modulů v tomto systému je ten, který obstarává porozumění posloupnosti slov, přičemž se předpokládá, že by měl spojovat výpovědi komunikačního protějšku v kontextu s dalšími známými informacemi. Tento předpoklad samozřejmě podněcuje další nesnáze v kontinuálním komunikačním řetězci člověka a systému; problémy vyvstávají především

²⁴ PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*, s. 401.

²⁵ Tamtéž, s. 401.

²⁶ Tamtéž, s. 644.

v získávání zdrojů znalostí, kterými by měl systém disponovat, aby byl schopen souvislé rozmluvy.²⁷

Existuje několik přístupů, které se využívají při řízení dialogu. Jedním z nich je dialogový systém s konečným počtem stavů, který je typický tím, že systém předem „zná“ veškeré povolené dialogy a vede uživatele k jejich použití. Systém není náročný na výkon a funkčnost, ale je schopen reagovat pouze na omezené množství promluv, a pro uživatele tak může být interakce s tímto typem systému poněkud zdoluhavá a nepříjemná.²⁸ Druhý z přístupů, který je možno využívat při řízení dialogů, je systém, který využívá strukturu rámců, tzn. že systém se snaží uživatele dovést k poskytnutí veškerých dat, které systém potřebuje. Na základě toho poté může systém podat uživateli požadované informace. Rámcový přístup je pružnější než systémy s konečným počtem stavů, protože jeho struktura umožňuje vkládat i více informací najednou. Tento způsob se využívá například v dopravních informačních systémech, kdy uživatel může uvést odkud, kam a kdy chce jet i v jedné promluvě a systém je schopen to zaznamenat.²⁹ Poslední z možností, jak řídit dialog, je využití dialogového systému založeného na agentech. Tento systém zajišťuje nejkomplexnější a nejvariabilnější pokrytí hlasových požadavků na základě spolupráce mezi agenty, kteří si vyměňují data a znalosti, přičemž informace musí být vzájemně kompatibilní na několika úrovních. Dialogový systém založený na agentech je schopný pokrýt velké množství požadavků, ale je z těchto tří systémů zdaleka nejsložitější, přičemž programátoři se musí potýkat s požadavky na velmi obsáhlý slovník, na pokročilost jazykového modelu a konečně na komplexnost scénářů na generování promluv.³⁰

Dialogový systém je charakteristický tím, že musí být schopen odpovídat na dotazy, ale zároveň je nutné, aby uměl i klást otázky. Kladení otázek je vlastně základní nutností pro jakýkoliv typ konverzace, bez níž by nebylo možné komunikaci rozvinout, popřípadě by rozmluva byla rozvíjena pouze jednostranně. Interakce mezi dvěma lidmi je charakteristická svou smíšenou iniciativou, protože obě strany můžou kdykoli převzít iniciativu, můžou zasáhnout do tématu hovoru apod. Každá ze stran by měla odpovídat na položené dotazy a tím poskytovat dodatečné informace a zároveň pokládat další otázky, aby se konverzace přirozeným způsobem rozvíjela.³¹

²⁷ PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*, s. 644.

²⁸ Tamtéž, s. 676.

²⁹ Tamtéž, s. 678.

³⁰ Tamtéž, s. 681.

³¹ PIERACCINI, R. *The Voice in the Machine*, s. 167.

V dialogových systémech je jistý typ iniciativy také nutný. Míra iniciativy se však může u různých typů dialogového systému lišit. Existuje i extrémní případ systému s iniciativou pocházející zcela od uživatele, ve kterém stroj nemá žádnou možnost intenzivně zasáhnout do vývoje děje či promluvy. V tomto případě stroj nepokládá otázky, takže zkrátka nemá jinou možnost než interpretovat všechny výroky uživatele jako požadavky, na které okamžitě odpovídá. Existuje i druhý extrém na opačné straně spektra, kdy má veškerou iniciativu pouze systém a uživatel nemá žádnou. Uživatel může pouze zodpovídat položené dotazy, ovšem nemůže nic upřesnit, na nic se doptat apod. Postupem času se ukázalo, že systémy založené pouze na iniciativě uživatele nejsou příliš funkční v reálném světě, oproti tomu systémy, kde iniciativu přebírá pouze stroj (tj. systémy řízených dialogů), se ukázaly jako velice užitečné v mnoha aplikacích. Mezi těmito dvěma extrémními případy je velké množství stupňů systémů se smíšenou iniciativou, přičemž ideální pozici pro optimální stupeň iniciativy v komunikaci člověka a stroje ještě neznáme.³²

2.7 Wizard of Oz

Při výzkumech interakce mezi člověkem a strojem je často potřeba velká sbírka dat a poznatků ohledně reakcí lidí na komunikaci se strojem. Pro tyto účely však výzkumníci potřebují data získaná z interakce mezi člověkem a strojem dokonce dříve, než mají aktuálně k dispozici takový typ systému, který by byl schopen v interakci s lidmi obstát. Z toho důvodu výzkumníci často využívají data z experimentálních situací, ve kterých si zkoumané subjekty pouze mysleli, že komunikují se strojem. Díky tomu pak výzkumní pracovníci získají cenné informace o tom, jak systém sestrojít, nebo o tom, jak by člověk reagoval na určité výrobní možnosti. Tento způsob nahrazení stroje či jeho části člověkem je známý jako metoda Wizard of Oz (WoZ).³³

Prostřednictvím WoZ lze simulovat počítačový systém a vzbudit v testovaném subjektu dojem, že je doopravdy v kontaktu s jakousi vyzpělou technologií. Díky tomuto způsobu řešení je možné vyhodnotit reakce lidí na technologii, která ještě není vyvinuta. Tato metoda má jisté podobné znaky s tzv. Turingovým testem, ve kterém má účastník testu odhalit, zda komunikuje se strojem, nebo s člověkem. Rozdílem v metodě WoZ je to, že člověk při testování ani není na možnost komunikace s lidskou bytostí upozorňován.

³² PIERACCINI, R. *The Voice in the Machine*, s. 168.

³³ Tamtéž, s. 154.

2.8 The Uncanny Valley

Termín „uncanny“ se poprvé snažil vysvětlit německý psychiatr Ernst Anton Jentsch ve svém díle „Zur Psychologie des Unheimlichen“, které pochází z počátku 20. století. Termín „uncanny“, popřípadě německý výraz „Unheimlich“, Jentsch vysvětluje jako pocit intelektuální nejistoty, který vlastně nedokážeme definovat, ani když se nacházíme přímo uprostřed tohoto pocitu. Anglická synonyma k pojmu „uncanny“ jsou například uncomfortable, gloomy, uneasy apod. Tento koncept o třináct let později rozpracovává Sigmund Freud, který jej definuje jako opak k pojmu familiární. Navzdory tomu, že Freud popisuje význam slova „uncanny“ v jiném oboru, nežli je robotika, přibližuje jej jako představu, která může být najednou dobře známá a zároveň i cizí a odlišná. Podle Freuda to pak vyúsťuje právě v pocit něčeho nepopsatelně podivného.³⁴

Pokud je termín „uncanny valley“ překládán do českého jazyka, většinou se užívá spojení „prapodivné údolí“, popřípadě „tajemný val“. Nicméně překlad se užívá jen pro ujasnění pojmu. Pojmenování hypotézy většinou bývá zachováno v anglické verzi.

Před více než čtyřiceti lety japonský robotik Masahiro Mori, který učil na Tokijském technologickém institutu, napsal esej o tom, jak si představoval reakce lidí na roboty, kteří vypadají a chovají se téměř jako člověk. Konkrétně ve své práci vyslovil hypotézu, která tvrdí, že lidská odezva na robota podobného člověku se může přesunout od pocitů empatie k pocitům odporu v případě, že podobnost robota s člověkem je nějakým způsobem nedokonalá. Tento posun vnímání by se měl udát tím, že se vzhled robota velmi přibližuje k úrovni, na které vypadá skoro „jako živý“, avšak té živosti či opravdovosti nakonec nedosáhne. Tento spád na hladinu děsivosti je znám jako „uncanny valley“.³⁵

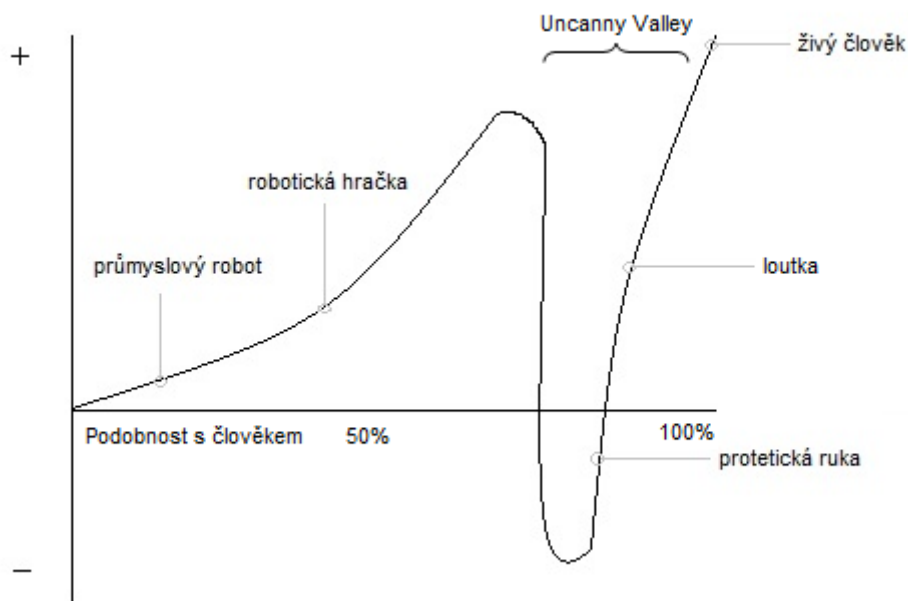
Moriho esej se objevila v tu dobu neznámém japonském časopise už roku 1970 a spoustu let zůstala nepovšimnuta. Avšak v poměrně nedávné době začal koncept „uncanny valley“ masivně poutat pozornost v robotice, v nejrůznějších vědeckých kruzích, a navíc i v populární vědě. Někteří výzkumníci se dokonce zabývali možnými dopady Moriho hypotézy na interakci člověka a stroje a na animace počítačové grafiky, zatímco jiní vědci zkoumali její biologické a sociální základy. Předpokládá se, že současný zájem v oblasti „uncanny valley“ se bude jen zintenzivňovat i vzhledem k tomu, jak se technologie neustále vyvíjejí a jak se výzkumníci snaží budovat roboty, kteří vypadají jako člověk.³⁶

³⁴ FREUD, S. *Das Unheimliche*, s. 297 – 300.

³⁵ MORI, M. *The Uncanny Valley* [online]. [30. 3. 2014] s. 98. Dostupné z: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6213238>>.

³⁶ Tamtéž, s. 98.

Mori ve svém článku přímo popisuje, že si povšiml, že čím více se přibližujeme na úroveň toho, že robot vypadá jako člověk, tím více náš pocit blízkosti k objektu stoupá. V jistém bodě se však křivka stoupání náklonnosti zastaví a jsme svědky náhlého propadu do „uncanny valley“.



Obr. 2.1: Grafické znázornění hypotézy „uncanny valley“.³⁷

Na obrázku lze vidět všechny body, které Mori zmínil ve známém grafu zobrazujícím křivku naší náklonnosti k neživým objektům. Křivka začíná u průmyslových robotů, kteří nevypadají jako lidé a u nichž je prvotní především jejich funkčnost. Lidé k nim tedy necítí žádnou vazbu právě díky tomu, že si nejsou vůbec podobní s lidskými bytostmi. Autor poté postupuje k popisu robotických hraček, u kterých se designéři zaměřují spíše na vzhled než na funkce. Robotické hračky se už začínají člověku podobat o něco více, i proto jsou jimi děti naprosto fascinovány. Poté autor zmiňuje protetické končetiny, které už se nachází v propadu do „uncanny valley“. Dnes již většinou nedokážeme rozpoznat umělou ruku od skutečné lidské ruky jen letmým pohledem. Až při bližším zkoumání můžeme zjistit, že „neživá“ ruka má zvláštní narůžovělou barvu nebo že její pohyby jsou nezvykle strnulé. Jakmile dojdeme

³⁷ MORI, M. *The Uncanny Valley* [online]. [30. 3. 2014] s. 99. Dostupné z: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&number=6213238>>.

k tomuto zjištění, zakoušíme právě onen pocit děsivosti a nepřírozenosti. Ovšem Mori v tomto ohledu vyzdvihuje další aspekt, který může být signifikantní ve vztahu k „uncanny valley“, a tou je rozdíl statických a pohyblivých objektů. Například umělá ruka ležící před námi na stole v nás nebude vyvolávat stejné pocity, jako když se nás ta ruka, nasazená na něčí končetině, bude snažit pohladit po tváři. Uvědomění si rozdílu objektu nehybného a objektu v pohybu samozřejmě ještě vzrůstá při interakci s něčím větším, než je jen umělá ruka, například s robotem. Moriho křivka nadále postupuje k loutkám, přičemž on přímo popisoval loutky japonského tradičního divadla Bunraku. Loutky sice nejsou přesnou kopií člověka, ale jisté podobnosti tam jsou, navíc při představení, když se loutky hýbou, můžou člověku připadat „jako živé“. Na konci křivky je poté člověk jako vrchol, se kterým jsou všechny ostatní objekty srovnávány.³⁸

V posuzování toho, co je pro lidskou bytost běžné a co už je zkrátka podivné, hraje roli velmi mnoho faktorů. Například když se robot podobný člověku snaží usmát, může to zintenzivnit náš pocit blízkosti k němu, jelikož úsměv je intuitivně vnímán jako pozitivní projev emocí. Avšak když si představíme, že robotí úsměv se formuje pomaleji než ten lidský, o mnoho pomaleji, přičemž ani výsledný výraz není naprosto shodný s člověkem, naše vnímání celé situace se propadá do té děsivosti v oblasti „uncanny valley“.³⁹

Mori uzavírá své myšlenky názorem, že jediným spolehlivým způsobem, jak se vyhnout „uncanny valley“ a jak zajistit bezpečnou úroveň pro pohodlnou interakci lidí a strojů, je snažit se objekty sestrojít tak, aby nepřipomínaly lidské bytosti. Nakonec se autor ještě táže, proč vlastně lidstvo bylo vybaveno tímto prapodivným pocitem a zda je to pro nás nějakým způsobem potřebné či užitečné. Mori se domnívá, že tento pocit je jakousi integrální částí našeho pudu sebezáchovy. Těmito závěrečnými řečnickými otázkami dává podněty právě k hojnému rozpracovávání teorie „uncanny valley“ a snaze k jejímu objasnění, potvrzení či vyvrácení.⁴⁰

³⁸ MORI, M. *The Uncanny Valley* [online]. [30. 3. 2014] s. 98 – 99. Dostupné z: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6213238>>.

³⁹ Tamtéž, s. 100.

⁴⁰ Tamtéž.

2.8.1 Podoba člověka a stroje

Hlavní otázkou tak zůstává, zdali se vůbec věda má pouštět do sestrojování robotů, kteří jsou podobní člověku. Spousta strojů nemá lidské rysy a je pohodlné s nimi komunikovat a to bez negativních pocitů. Podle jiných autorů je naopak nutné snažit se sestavit roboty, kteří se svým vzhledem přibližují člověku. Například David Hanson, americký vědecký pracovník a designér robotů, který je zodpovědný za sestrojení celé série realistických humanoidních robotů, tvrdí, že snaha robotiky realisticky zobrazovat objekty tak, aby byly podobné člověku, nám nabídne jedinečný nástroj pro zkoumání lidského sociálního vnímání a poznávání. Sestrojení „společenské bytosti“ ve všech jejích detailech může pomoci lépe pochopit sociální inteligenci jak ve vědeckém, tak v uměleckém pojetí.⁴¹ David Hanson od roku 2001 do roku 2008 sestrojil se svým týmem dvacet robotů, mezi nimiž byl i robot v podobě Alberta Einsteina, který dokáže vyjadřovat emoce. Podle Hansona je důležité soustředit se na dvě věci. Jednou z nich je přijímání robotů lidmi. Druhým důležitým aspektem je pak přirozený vzhled robota. Vzhled robota podobný lidskému vzezření by měl zajistit příjemnější a intuitivní interakci člověka s robotem. Člověk by měl uvěřit, že je robot živý.⁴² Hanson tvrdí, že přirozenost a lidskost robotů jsou jedny z klíčových požadavků k navázání úspěšné interakce mezi člověkem a strojem.



Obr. 2.2: Humanoidní robot „Albert Hubo“.⁴³

⁴¹ ŠVARNÝ, P. *A Visit on the Uncanny Hill*, s. 36.

⁴² HANSON, D. *David Hanson: Robots that "show emotion"* (přednáška na konferenci TED2009) [online]. [31. 3. 2014] Dostupné z: <http://www.ted.com/talks/david_hanson_robots_that_relate_to_you#t-171981>.

⁴³ Autor neuveden. *The 50 Best Robots Ever* [online]. [4. 4. 2014] Dostupné z: <http://archive.wired.com/wired/archive/14.01/images/FF_120_bots34_f.jpg>.

2.8.2 Vztah člověka a stroje

Ve spojení s hypotézou „uncanny valley“ vyvstává problém s tím, že teorie nezohledňuje myšlenku dlouhodobé spolupráce s umělými objekty. Nutnost zvyknout si na nová vylepšení v oblasti vědy hraje důležitou roli v utváření sociálního života člověka a provází lidstvo v mnoha sférách lidské interakce. Je tedy možné, že jakýkoli pocit děsivosti nebo nepohodlnosti při setkání s robotem podobným člověku by zkrátka vymizel po několika dnech interakce.⁴⁴

Vědu a výzkum v robotice by mělo spíše zajímat zlepšení stavu interakce člověka a stroje, nikoli co nejlidštější vzhled robotů. Naším největším zájmem by stále měla být snaha zlepšit „vztahy“ mezi člověkem a strojem za každou cenu. Kdybychom se zaměřili například pouze na obličejové rysy robotů, možná bychom sestrojili stroj, který má skvělou mimiku, který odráží ve výrazech své „pocity“, a třeba bychom dokázali z fyziognomie jeho tváře předvídat jeho reakce, ale jeho složitost a nároky na energii by možná byly příliš vysoké na používání takového objektu v každodenním životě. Na druhou stranu pokud zcela opomeneme důležitost obličejových rysů, můžeme výrazně ztížit interakci lidí a strojů.⁴⁵

2.8.3 Hlas ve vztahu k „uncanny valley“

Některé interakce s roboty dokonce ani nepotřebují vizuální přenos na to, aby správně fungovaly. Příkladem toho je počítačový program Eliza nebo inteligentní osobní asistent Siri, který funguje ve verzi 5 v Apple iOS a reaguje pouze prostřednictvím hlasu. Takové aplikace zřejmě nejsou v ohrožení z propadu do „uncanny valley“, ale pouze ukazují, jaký problém může hypotéza představovat. Emoční úprava umělého hlasu může být pro lidi dostatečnou zpětnou vazbou, a interakce mezi člověkem a strojem tak může být dokonce podobná mezilidské komunikaci. Zásadním bodem v tomto působení jsou rozdíly, kterým lidé přiřkládají důležitost vizuálních a auditivních stimulů. Tyto rozdíly jsou velmi individuální a nutné náležitosti, které každý jedinec preferuje v komunikaci, se mohou lišit.⁴⁶

Výzkum v této práci se zaměřoval právě především na auditivní stimuly. Ačkoli tyto podněty nebyly detailně zkoumány přímo v kontextu „uncanny valley“, existují některé studie, které dokazují jasné preference uživatelů k určitým aspektům hlasu. Například výzkumy americké doktorky Kathryn Dragerové a jejího týmu se snažily odhalit u posluchačů jisté korelace v názorech na porozumění syntetické řeči. Většina z posluchačů pak preferovala

⁴⁴ ŠVARNÝ, P. *A Visit on the Uncanny Hill*, s. 37.

⁴⁵ Tamtéž, s. 38.

⁴⁶ Tamtéž.

a pozitivněji ohodnotila přirozenou řeč před syntetickou řečí. I toto může být jedním z důkazů, že lidé jsou sluchově „náročnější“ a vyžadují perfektní kvalitu syntetického hlasu. Podle teorie o vztahu komunikace a prostoru (Communication Accommodation Theory) se lidé během komunikace přizpůsobují svému protějšku v dialogu. Podle této teorie by to samé mělo platit i v interakci mezi člověkem a strojem. Používání nepřirozeného hlasu je pak aspekt, kterému se člověk neumí přizpůsobit. To způsobuje nesoulad, který se projevuje v tom, že lidé očekávali jakousi lidskost, která by měla být hlasu vlastní, ale té se jim nedostává. Nepřirozený či až robotický hlas tak může být nepřekonatelnou překážkou k tomu, aby člověk a stroj vedli dialog přibližující se mezilidskému stylu interakce. Překážky mohou také být podpořeny tím, že představivost člověka je nezastavitelná a člověk intuitivně spojuje přirozeněji znějící hlas s lidskou představou mluvčího, od kterého by měl ten hlas pocházet.⁴⁷

2.9 Shrnutí

Uvedené teoretické informace o řečových technologiích jsou základním uvedením do stěžejní části práce, která se týká praktického provedení experimentálního výzkumu. Důvodem popisu veškerých výše uvedených náležitostí je, že většina z nich hrála roli v provedeném experimentu. Experiment byl uspořádán za účelem posouzení, zda hypotéza „uncanny valley“ může nějakým způsobem souviset se syntetickou řečí. Ačkoli použití některých popisovaných aspektů bylo v experimentálním výzkumu spíše teoretické či imaginární, bylo nutné uvést jejich funkce a účel k pochopení teoretických požadavků k uspořádání výzkumu.

Účel, požadavky a přesný průběh zkoumání je podrobně popsán v následujících kapitolách. Dále zde jsou popsány některé konkrétní případy, celkové výsledky testování spolu se závěrem celého experimentu.

⁴⁷ SZERSZEN, K. A. *The Audio/Visual Mismatch and the Uncanny Valley*, s. 6 – 8.

3 Praktická část

3.1 Uvedení do problematiky experimentu

Vzhledem k tématu své diplomové práce jsem se rozhodla ve svém bádání zaměřit více prakticky. Cílem mé práce je experimentálně posoudit platnost hypotézy „uncanny valley“ na základě poslechových testů syntetické řeči. Původní vizí bylo zaměřit se na již provedený pokus Senior Companion; k dispozici jsem totiž měla nahrávky a videa pořízené při pokusech, jež přinesly poměrně překvapivé výsledky.⁴⁸ Pokus Senior Companion je částí evropského projektu COMPANIONS, na kterém se podíleli partneři ze čtrnácti států z Evropy a Spojených států. Tento projekt vytvořil virtuální komunikační partnery ve snaze změnit způsob, jakým lidé smýšlí o svých vztazích k počítačům a internetu. Systém, jenž byl vytvořen v rámci tohoto projektu, přesahuje hranice tradičního způsobu interakce mezi člověkem a počítačem, který je založený na klasickém zadávání a následném plnění úkolů. Oproti tomuto běžnému přístupu se projekt snažil rozšířit existující oblast sociální interakce mezi lidmi o sféru komunikace člověk – stroj.⁴⁹

České pojetí projektu COMPANIONS se zabývalo snahou navázat přirozenou komunikaci dialogového systému a účastníka dialogu. Tato snaha probíhala prostřednictvím rozhovorů se seniory. Existoval předpoklad, že účastníci dialogu budou systém vnímat spíše negativně – předpokládaly se strohé, vlažné či až negativní reakce, i v důsledku použití nepřiliš kvalitního stupně syntetické řeči. Uvedená očekávání se však nenaplnila, většina účastníků vnímala rozhovor se systémem velmi pozitivně, přátelsky a s emočním zapojením reagovala na podněty systému a nevnímala v interakci, až na výjimky, žádný problém. To však mohlo být způsobeno věkem respondentů, kteří neměli předchozí zkušenosti s dialogovými systémy, ani nebyli obzvláště obeznámeni s řečovými technologiemi atp. Tyto závěry však byly poměrně detailně zpracovány již jinými diplomanty.⁵⁰

Navíc v pokusu Senior Companion se nijak neposuzovala hypotéza „uncanny valley“, zkoumaní neměli popisovat, jaké pocity měli z rozhovoru, zda pro ně nebylo divné povídat si se strojem atd. V pokusu bylo stěžejní navázat a udržet komunikaci člověka a stroje. Jelikož si

⁴⁸ HEIDEROVÁ, P. *Perspektivy řečové komunikace mezi člověkem a strojem*, s. 53.

⁴⁹ European Commission. *COMPANIONS* [online]. [28. 1. 2014] Dostupné z [www: <http://www.companions-project.org/?p=home>](http://www.companions-project.org/?p=home).

⁵⁰ HEIDEROVÁ, Pavlína. *Perspektivy řečové komunikace mezi člověkem a strojem*. Plzeň, 2013. Diplomová práce (Mgr.). Západočeská univerzita v Plzni, Fakulta filozofická. Vedoucí práce Jan Romportl. OTRADOVCOVÁ, Markéta. *Multimodální komunikace člověk-stroj*. Plzeň, 2010. Diplomová práce (Mgr.). Západočeská univerzita v Plzni, Fakulta filozofická. Vedoucí práce Jan Romportl.

tato práce určila za cíl posoudit hypotézu „uncanny valley“, bylo nutné provést jiný experiment, který by se zajímal právě o ty pocity a názory zkoumaných. Nový experiment tedy není pokračováním pokusu se seniory, pouze byl v mnohém inspirován některými postupy či metodami zmíněného pokusu. Zvolili jsme podobný přístup jak k účastníkům, tak k vedení rozhovoru (stavba rozhovoru, typy otázek apod.). Jednou z konkrétních inspirací byla například metoda Wizard of Oz, která byla stěžejním nástrojem.

3.2 Experiment

Vzhledem k již podrobně popsanému pokusu Senior Companion a jeho cílové skupině, kterou tvořila především starší generace, byla pozornost zkoumání přesunuta na mladší generaci. Jednalo se především o skupinu současné mladé populace, jež je tvořena studenty a absolventy, kteří se ve sféře moderní techniky orientují více, pracují s technickými přístroji téměř nepřetržitě a nečiní jim to nijaké potíže. Toto tvrzení je opřeno o dotazníkové šetření⁵¹, které proběhlo od 25. do 27. října 2013 ve skupině 64 respondentů, jejichž věkový průměr se pohyboval kolem 24 let. Necelých 70 % z dotazovaných vlastní minimálně dva či více prostředků komunikačních technologií, ať už se jedná o smartphone, tablet, notebook atp. Stejný počet respondentů (celkem 44 z dotazovaných) používá uvedené přístroje i několik hodin denně a nedovede (nebo nechce) si představit život bez používání nových technologií. V experimentu bylo nutné zaměřit se na posouzení, jestli dostatečná obeznámenost s novými technologiemi může ovlivnit vnímání jedinců, tzn. zda jsou k novým technologiím přátelštější, nebo právě naopak nedůvěřivější. Existoval předpoklad, že díky tomu, že většina mladé populace neustále používá moderní technologie a je zvyklá žít v technice přátelsky nakloněném světě, vnímala by pokročilejší stádia techniky s povděkem, a nikoli jako přítěž či překážku.

Na samotném provedení experimentu se velkou měrou podíleli vědecko-výzkumní pracovníci Západočeské univerzity Ing. Aleš Pražák, Ph.D. a Ing. Mgr. Jan Romportl, Ph.D. Zmínění přispěli radami ohledně způsobu provedení pokusu, poradili nejvhodnější metody a nástroje k provedení a varovali před možnými riziky v průběhu zkoumání. Pokus byl založen na provedení krátkých rozhovorů s dialogovým systémem prostřednictvím programu Skype umožňujícím volání přes internet a videohovory, stejně jako posílání zpráv a souborů mezi

⁵¹ Dotazníkové šetření bylo provedeno ve spolupráci se dvěma kolegyněmi ze studijního oboru TFK na ZČU v souladu s náležitostmi platnými pro kvantitativní výzkum. Viz HENDL, Jan. *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. Vyd. 2., opr. Praha: Portál, 2006. 583 s. ISBN 8073671239.

komunikanty. Experimenty probíhaly od 5. 11. do 25. 11. 2013 a zúčastnilo se jich celkem 30 respondentů, 15 žen a 15 mužů. Věkový průměr respondentů byl 23,5 let.

3.3 Nepostradatelné součásti experimentu

Ideálním nástrojem experimentu by byl hlasový dialogový systém. Podrobnější informace ohledně funkce dialogových systémů jsou uvedeny v kapitole 2.6 „Dialogové systémy“.⁵² K provedení experimentu bylo potřeba, aby systém nejen rozpoznával, co účastník pokusu říká, ale především bylo zapotřebí, aby systém adekvátně reagoval na řečené a následně poskytoval subjektu očekávanou a částečně předvídatelnou reakci, zkrátka takovou, jakou očekáváme v běžné konverzaci. Technologie dialogových systémů je na úrovni, která se pomalu přibližuje srovnání s interakcí mezi lidmi, ale prozatím stále komunikace mezi člověkem a strojem nedosahuje plynulosti, míry porozumění a možností poskytovat adekvátní reakce v takové míře jako mezilidská komunikace. Existuje poměrně velké množství obstojně reagujících dialogových systémů, které jsou schopny detekovat, co subjekt říká, a poskytnout přiměřenou odezvu. Hlasový dialogový systém, který to dokáže, byl vyvinut na katedře kybernetiky Západočeské univerzity v Plzni. Dialogový systém je schopen poskytnout informace o odjezdech a příjezdech vlaků a je možné s ním komunikovat prostřednictvím přirozeného jazyka s libovolnou stavbou věty.⁵³ Na druhou stranu možnost reakcí systému je poměrně omezená, protože systém není schopen aktivně reagovat na neočekávané podněty. Typ dialogového systému, který byl použit k informování ohledně dopravní situace, by se hodil k využití právě v experimentu. Musel by však být naprogramován k poskytování reakcí, které byly v pokusu stěžejní (tzn. umět zareagovat i na neočekávané situace při rozhovorech s účastníky, snažit se podnítit další konverzaci, umět vyhodnotit, kdy disponujeme dostatečným množstvím dat, a navíc posoudit, zda i účastník obdržel dostatečnou informaci, které by mu následně umožnily vyplnit dotazník apod.). Takový systém by byl velice náročný na naprogramování, což nebylo možné jak z technických, tak z časových důvodů. Proto jsme se pokusili napodobit dialogy podle vzoru nádražního systému a zvolili jsme postup prostřednictvím již zmiňovaného přístupu Wizard of Oz. Tato metoda nabízí možnost, jak simulovat dialogový systém, který by dokázal rozumět, adekvátně zareagovat a zároveň by jeho rozhodování netrvalo příliš dlouhou dobu. Postup, jakým bylo možno dosáhnout takových náležitostí, probíhal prostřednictvím osoby, jež

⁵² Strana 8 v této práci.

⁵³ ŠVEC, Jan. *Nádraží – hlasový dialogový systém* [online]. [22. 10. 2013] Dostupné z [www: <http://www.youtube.com/watch?v=VZ8I5M_mu2M>](http://www.youtube.com/watch?v=VZ8I5M_mu2M).

předstírala, že je dialogovým systémem, aniž by o tom zkoumaný subjekt věděl. Účastník experimentu měl mít pocit, že vede dialog s počítačem, ačkoli na druhé straně seděl člověk (tzv. wizard).⁵⁴

V konkrétním případě mých experimentů jsem úlohu wizarďa zaujímala já sama, samozřejmě bez vědomí účastníka. V rámci simulace rozhovoru člověka a počítače wizard musel dodržovat určité typy reakcí charakteristické pro nedokonalou strojovou komunikaci, tak aby se přibližoval strojovému „chápání“ světa. Mezi tyto typy reakcí patřila především nízká rychlost odezev, jelikož se předpokládalo, že systému by zabralo více času rozpoznat řeč účastníka, zpracovat ji, vybrat adekvátní odpověď a pronést ji. Dalším z charakteristických rysů systému, který byl simulován, byla snaha o jeho určitou „nebystrost“, tedy nutnost parodovat neznalost určitých pojmů, nespisovných výrazů, spojení z obecné češtiny a nepochopení kontextuálních souvislostí. Systém byl nucen doptávat se účastníků, aby správně „porozuměl“ komunikaci. Podobným stylem probíhaly všechny rozhovory, jež byly pořizeny. Materiál rozhovorů má rozsah celkem 7,9 hodin.

Nejdůležitější součástí experimentu byla aplikace SpeechTech TTS se sadou vlastních umělých hlasů. Při rozhovoru byly používány dva typy syntetických hlasů s rozdílnou kvalitou. Verze syntetického hlasu je rozhodujícím faktorem kvality hlasu a zároveň určuje i nároky na výpočetní výkon. Jako vysoce kvalitní syntetický hlas byl použit hlas „Iva“ ve verzi 2.10 a jako druhý syntetický hlas byl použit hlas „Alena“ ve verzi 2.6. Hlas Aleny měl záměrně znít roboticky, neuměle.⁵⁵ Během rozhovoru z aplikace zaznívaly věty, které do ní byly zapsány (formou TTS – převod textu na řeč) a účastník rozhovoru je slyšel, jako kdyby k němu ze strany komunikujícího protějšku mluvil systém (robot). Přenos zvuku z aplikace SpeechTech TTS k účastníkům byl možný díky tomu, že zvuk byl produkován ze záznamového přehrávání, nikoliv z pole mikrofónu. V praxi to tedy znamenalo, že cokoli se přehrálo na počítači, slyšel subjekt na druhé straně Skypu. Další neodmyslitelnou součástí pokusu pak bylo nahrávání celého rozhovoru, k čemuž sloužil program GoldWave.

⁵⁴ EKLUND, J., GUSTAFSON, J., HELDNER, M., HJALMARSSON, A. *Towards human-like spoken dialogue systems. Speech communication* [online]. [25. 3. 2014] s. 630 – 645. Dostupné z: <<http://linkinghub.elsevier.com/retrieve/pii/S016763930800054X>>.

⁵⁵ Autor neuveden. *SpeechTech TTS – syntéza řeči* [online]. [30. 11. 2013] Dostupné z [www: <http://www.speechtech.cz/produkty/synteza-rci.html>](http://www.speechtech.cz/produkty/synteza-rci.html).

3.4 Experimentální protokol

Během každého experimentu proběhlo následující: zkontaktování účastníka na Skypu, zaslání instrukcí a případná konzultace nejasností či dotazů, které vznesl zúčastněný, a poté probíhal samotný hovor. Před rozhovorem obdržel pozorovaný subjekt instrukce s vesměs obdobným obsahem: „Experiment spočívá v provedení krátkých hovorů s hlasovým dialogovým systémem, který bude probíhat prostřednictvím programu Skype. Účastník provede krátký rozhovor s řečovým dialogovým systémem jako rovnocenným partnerem v komunikaci. Konverzace bude řízena systémem, který je schopen nejen ovládat verbální projev, ale zároveň je schopen rozpoznávat lidskou řeč. Předmětem dialogu bude obrázek, který obdržíte před samotným hovorem a budete jej se systémem probírat. Po dokončení rozhovoru bude vaším úkolem zhodnotit, jaký dojem jste měli z hovoru, jaké pocity ve vás dialog vzbuzoval (především budete hodnotit interakci se systémem, nikoli jak pěkný se vám zdá určitý hlas, naopak důležitější je hodnocení vašeho celkového dojmu z rozhovoru, vašeho komfortu při komunikaci a vaší představy, kdy a jak by pro vás bylo pohodlné a výhodné komunikovat s dialogovým systémem).“

Účastníci rozhovoru byli dále informováni o předpokládané délce hovoru (maximálně hodina včetně podání instrukcí, samotného rozhovoru s dialogovým systémem i vyplnění dotazníku), o předpokládaném „scénáři“ hovoru (tedy dva kratší rozhovory se dvěma systémy s různým hlasem) a také byli varováni, že reakce protistrany může být někdy zpomalená, nepřiměřená či nečekaná. Zároveň při jakýchkoli nejasnostech se dobrovolníci mohli ptát přímo daného systému, či vznést dotaz prostřednictvím Skype chatu. Na závěr instrukcí bylo znovu zdůrazněno, že se nemají hodnotit estetické vlastnosti jednotlivých hlasů. Nechtěli jsme vědět, který z hlasů je „hezčí“, ale se kterým z hlasů by si účastníci byli schopni představit reálnou interakci, zkrátka z kterého z rozhovorů budou mít lepší pocit.

V průběhu rozhovoru bylo využito polostrukturovaného dotazování, což umožnilo naplnit definovaný účel dotazníku a proces získávání informací byl tím pádem mnohem přizpůsobivější. Každý hovor byl nahráván a probíhal podle předem částečně připraveného scénáře, ve kterém se přes pozdrav, uvítání a souhlas s nahráváním volně přešlo k zaslání obrázku a následné debatě o něm.⁵⁶

⁵⁶ HENDL, J., *Kvalitativní výzkum*, s. 168.

Ke každému pokusu bylo výhodné mít připravený seznam vět, které bylo nutno využít téměř v každém rozhovoru. Mezi tyto věty patřily například promluvy typu:

„Slyšíme se?“

„Dobrý den, jmenuji se Iva a budu si teď s vámi chvíli povídat.“

nebo: „Dobrý den, jmenuji se Alena a budu si teď s vámi chvíli povídat.“

„Pošlu vám obrázek, o kterém si budeme povídat.“

„Ještě než začneme, slyšíte mě dobře?“

„Ještě se musím zeptat, zda vám nevadí, že bude hovor nahráván?“

„Po skončení hovoru vám pošlu dotazník, poprosím vás o jeho vyplnění.“

„Máte nějaké otázky, než začneme?“

„Obdržel jste obrázek?“

„Tak můžeme začít.“

„Co mi můžete říct k obrázku?“

„Můžete mi jej více popsat?“

„Chodíte často na autobusové nádraží?“

„Chodíte často na vlakové nádraží?“

„Jezdíte často autobusem?“

„Jezdíte často vlakem?“

„Proč?“

„Aha.“

„Odkud jste?“

„Nemusíte dojíždět za studiem?“

„Jaké preferujete dopravní prostředky?“

„Už jsem se toho od vás hodně dozvěděla.“

„Jsme tak v půli našeho rozhovoru.“

„Předám slovo kolegyni.“

„Vydržte chvíli.“

Existovala snaha nechat mluvit účastníka co nejvíce a správně mířenými dotazy ho směřovat k tématu, o kterém bude moci tázaný povídat delší dobu. Přírozenou konverzaci a důkaz zájmu o respondenta se wizard snažil podpořit doplňujícími otázkami a údivem („Aha, to je velmi zajímavé, napadá vás k tomu ještě něco?“). Nicméně u některých jedinců nebyla přítomna touha debatovat, takže systém (tj. wizard, tj. já) musel vyvinout více aktivity pro zachování diskuze, která by alespoň částečně naplnila požadavky mezilidské (respektive mezi-strojově-lidské) komunikace. Přibližně v polovině rozhovoru došlo k vystřídání „mluvčích“ a jeden ze syntetických hlasů byl nahrazen tím druhým. Důležité také bylo poskytnutí informací účastníkovi o průběhu dialogu, proto účastníci často slyšali věty typu: „Jsme přibližně v půli našeho rozhovoru.“, „Už jsem se toho od vás hodně dozvěděla.“ apod.⁵⁷

Pořadí syntetických hlasů bylo záměrně střídáno u každého dobrovolníka a zároveň i pořadí zasílání dvou obrázků, které účastníci obdrželi, bylo úmyslně měněno, aby pevně nekorespondovalo ani s jedním z hlasů, ani s pořadím zaslání. Na prvním z obrázků bylo zachyceno Hlavní nádraží Českých drah v Plzni a na druhém bylo zobrazeno Centrální autobusové nádraží v Plzni. Účelně byly vybrány fotky, u nichž bylo nejpravděpodobnější, že vyvolají minimální negativní reakce, a zároveň to byly fotky, u kterých se předpokládalo, že zachovají udržitelné téma hovoru, tedy téma dopravy, cestování apod. Vzhledem k tomu, že cílová skupina studentů většinou pobývala v Plzni, bylo téměř jisté, že zmíněná nádraží budou znát, a zároveň bylo pravděpodobné, že dojíždí za studiem, a tím pádem jsou nuceni používat nějaký dopravní prostředek.

Zakončení rozhovoru je také velmi důležité, především kvůli poslední možnosti předání informací, které mohou ovlivnit účastníka a jeho odpovědi v dotazníku.⁵⁸ Konec konverzace opět proběhl dle předem připraveného scénáře. V rámci rozloučení se účastník dozvěděl, se kterým z hlasů právě domluvil, aby byla maximalizována možnost, že ve formuláři dokáže opětovně určit, který hlas patřil Aleně a který Ivě. Poslední věcí, kterou účastník v průběhu slyšel, bylo poděkování za účast na výzkumu a výzva k vyplnění dotazníku.

⁵⁷ HENDL, J., *Kvalitativní výzkum*, s. 170.

⁵⁸ Tamtéž, s. 167.

Otázky na konci rozhovoru byly opět u většiny rozhovorů stejné:

„Můžete mi k tomu říct víc?“

„Napadá vás k tomu ještě něco?“

„Dobře.“

„Tak to mi asi bude stačit.“

„Děkuji za váš čas.“

„Nyní vám pošlu dotazník a poprosím vás o jeho vyplnění.“

„Loučí se s vámi Iva.“ nebo „Loučí se s vámi Alena.“

„Děkuji.“

„Na slyšenou.“

Samotný rozhovor trval v průměru 16 minut a 15 vteřin. Vyplnění dotazníku bezprostředně po rozhovoru mělo především eliminovat případné nepřesnosti, které by způsobil delší časový odstup, sdílení dojmů se třetí stranou, rešerše informací o stavu řečových technologií apod. Bylo potřeba, aby účastník bezprostřední dojmy věnoval právě dotazníku. Avšak to nevylučuje další možné nejasnosti a nepřesnosti, které mohly být do výzkumu a statického vyhodnocení vneseny. Těmto rizikům se budu věnovat v následujících kapitolách.

3.5 Metodologie⁵⁹

V experimentu převažoval výzkum převážně kvantitativního charakteru. V něm dominují hlavně nahodilé výběry, pokusy a striktně uspořádané sběry dat, např. pomocí dotazníků. V kvantitativním typu výzkumu existuje předpoklad, že chování lidského jedince může být měřeno a do určité míry i předpověděno. Právě to je stěžejním přístupem především v části vyhodnocování výsledků experimentu, kde se vyskytuje statistická analýza dat.⁶⁰ Kvantitativní výzkum dále může mít experimentální či neexperimentální podobu. V konkrétním případě, který je popisován v této práci, se jednalo o experimentální přístup. Ten je charakteristický tím, že výzkumník se snaží svým přičiněním způsobit určitou změnu odlišnou od běžného stavu situací či zkušeností, které člověk zažívá v bezprostředním kontaktu s realitou. Tuto úlohu plně naplňovala funkce wizarda v experimentu. Výzkumníka

⁵⁹ Vypracováno dle struktury postupu uvedené v: HENDL, J., *Kvalitativní výzkum*, s. 321.

⁶⁰ HENDL, J., *Kvalitativní výzkum*, s. 44.

posléze zajímá způsob reakcí jedinců, schopnost jejich přizpůsobení se nové situaci a zároveň i subjektivní zhodnocení situace od každého jedince. Mezi příznačné rysy experimentálního přístupu patří především to, že zkoumané subjekty pochází ze známého vzorku populace (např. studenti), dále nutnost obměny výchozích podmínek a snaha otestovat určitou hypotézu.⁶¹

Ne všechny charakteristiky kvantitativního výzkumu se však dají aplikovat na postup experimentu popisovaného v této práci. Jisté podobnosti můžeme nalézt i v typických znacích kvalitativního výzkumu. Paralely mohou být spatřovány především v tom, že se pracuje s limitovaným množstvím účastníků, sběr dat se odehrává na určitém místě (konkrétně u účastníka doma) a z důvodu relativně nízkého počtu zúčastněných (30 účastníků) zde vyvstává problém se vztažením výsledků na širší část populace. Výsledky kvalitativních výzkumů zpravidla ztělesňují spíše soubor subjektivních hodnocení, jež nelze snadno zobecnit, a to především z důvodu různorodosti dat a volné struktury metod.⁶² Právě to bylo problémem i v dotaznících experimentu, v nichž bylo občas obtížné nalézt společné atributy v odpovědích zúčastněných. Názory účastníků se často lišily a jejich počet byl poměrně nízký na to, aby se typy odpovědí signifikantně opakovaly. Mimoto problém s poměrně volnou metodou se vyskytoval také v rozhovorech, u nichž byly naplánované jen hrubé rysy průběhu, avšak každý rozhovor byl jedinečný a určitým způsobem odlišný. Na druhou stranu velkou výhodou kvalitativního přístupu je získání důkladného popisu jednotlivých případů, jejichž porovnáváním můžeme zjistit například to, jakou roli hrál ve zkoumání kontext a podmínky jednotlivých případů.⁶³

Metodicky by bylo možné experiment rovněž zařadit pod tzv. kolektivní případovou studii, která detailněji zkoumá větší množství případů obecného jevu. Podle tohoto přístupu lze dosáhnout rozsáhlejšího pochopení dané problematiky prostřednictvím monitorování podobných situací. Zmíněný postup je možné využít při snaze otestovat nějakou teorii nebo také v komparativních výzkumech.⁶⁴

Stále více tak převažuje názor, že kvalitativní a kvantitativní výzkumy nutně nejsou protikladnými přístupy, ale že se vzájemně doplňují. Jan Hendl uvádí některé rozdíly mezi kvalitativním a kvantitativním výzkumem, jež uspořádal podle autorů Alana Brymana a Lawrence B. Mohra. Pokud bychom chtěli posoudit převažující zaměření experimentu, nedostaneme jasné výsledky. Například volný vztah výzkumníka k subjektu (nebyl nutný

⁶¹ HENDL, J., *Kvalitativní výzkum*, s. 44 – 45.

⁶² Tamtéž, s. 50.

⁶³ Tamtéž, s. 51.

⁶⁴ Tamtéž, s. 105.

žádný osobní vztah mezi výzkumníkem a účastníky pokusu) je charakteristický pro kvantitativní výzkum, postavení výzkumníka spíše uvnitř situace naopak spadá pod kvalitativní výzkum. V rámci kvantitativního přístupu šlo o snahu potvrdit či falzifikovat hypotézu, avšak výzkumná teorie byla spíše slabě strukturovaná (tedy neměla jasně vymezený postup), takže by pravděpodobněji náležela k výzkumu kvalitativnímu. Pokus se snažil zobecnit získaná data a výsledky, což je jedním ze znaků kvantitativního modelu, ale charakter dat, jež jsou bohatá a hloubková, na druhou stranu dotvívá strukturu kvalitativního přístupu.⁶⁵

Závěrem k metodice lze tedy říci, že přístup experimentu nebyl ani kvalitativní, ani kvantitativní, nýbrž smíšený. Smíšený výzkum slučuje metody a techniky obou přístupů v rámci jedné studie. Díky tomu bylo možné sledovat širší rozpětí zkoumaných hodnot (jak subjektivní reakce účastníků a jejich individuální hodnocení, tak srovnatelné, zejména demografické údaje v dotaznících, jež bylo možné kvantifikovat a porovnat) a zároveň se podařilo eliminovat nevýhody jednotlivých přístupů (například bylo možné díky dotazníkům zobecnit rozdrobená data získaná z rozhovorů).⁶⁶ Během experimentu tedy byly použity metody smíšeného výzkumu, konkrétně rozhovor a posléze dotazník, který náleží kvantitativnímu výzkumu a jehož výsledky mají vyšší relevanci a možnost zobecnění.

3.5.1 Sledování jedinci

Výběr jedinců, kteří se mohli zúčastnit experimentu, probíhal v období od října do listopadu 2013. Cílení jedinci byli především studenti, či přesněji řečeno mladá generace, tedy osoby ve věkovém rozsahu od 20 do 30 let. Avšak omezovat něčí účast nebylo nutné a koneckonců ani potřebné, jelikož lidé, kteří projeví zájem o účast ve výzkumu, se pohybovali v těchto věkových hranicích.

„Nábor“ účastníků probíhal především na akademické půdě Západočeské univerzity, zejména v 1. a 2. ročníku navazujícího magisterského studia oboru Teorie a filozofie komunikace na Katedře filozofie a také v hodinách předmětů ‚Umělá inteligence‘ a ‚Historie kybernetiky a umělé inteligence‘ v rámci Katedry kybernetiky. Nábor byl možný především díky vyučujícím jednotlivých předmětů, kteří mi umožnili v jejich vyučovací hodině prezentovat zamýšlenou vizi mého experimentu. Ve snaze přesvědčit studenty k účasti na

⁶⁵ HENDL, J., *Kvalitativní výzkum*, s. 54 – 55.

⁶⁶ Tamtéž, s. 61.

výzkumu byly potenciálním účastníkům sděleny určité informace o něm (viz kapitola 3.4 „Experimentální protokol“)⁶⁷ společně se stručnou prezentací účelu pokusu.

Na základě prezentací ve vyučovacích hodinách na ZČU se přihlásilo celkem 37 studentů, avšak ve skutečnosti se k samotnému experimentu odhodlalo jen 13 z nich. Zbytek studentů buď vůbec nezareagoval na dodatečný e-mail, ve kterém byli účastníci tázáni na datum a čas, ve kterém by se mohli pokusu zúčastnit, nebo se někteří ozvali v době, kdy byl výzkum již vyhodnocován apod. Zbýlých 17 účastníků bylo tvořeno převážně lidmi z ostatních fakult, kteří byli ochotni participovat na výzkumu, popřípadě jedinci, kteří se doslechli o experimentu a na základě vlastního zájmu se také chtěli zúčastnit.

3.5.2 Rozhovor

Rozhovor s účastníky lze charakterizovat jako polostrukturovaný. Probíhal podle částečně předpřipraveného scénáře, přesto však ne vždy se stejným výsledkem. Během interview byla nutná improvizace pro zachování „plynulé“ konverzace a zároveň při výskytu technických potíží (problémy se spojením, špatný signál Skypu, druhá strana neslyšela syntetickou řeč atd.) se rozhovor musel přizpůsobit ztíženým podmínkám.

Informace získané z rozhovorů byly velmi podrobné a poměrně subjektivní vzhledem ke každému individuálnímu účastníkovi. Z tohoto důvodu bylo po uskutečnění rozhovoru využito ještě statisticky relevantnějšího tázání – tedy dotazníku.

3.5.3 Dotazník

Dotazník⁶⁸ byl vytvořený na základě pravidel platných pro kvantitativní výzkum, který je složený především z nejrůznějších metod získávání odpovědí od respondentů. Zvláštní pozornost byla věnována formám otázek, které měly být skutečně otevřené, neutrální, ale přitom z nich mělo být jasně a na první přečtení pochopitelné, na co je respondent tázán.⁶⁹ V dotazníku byla použita jistá forma sondáže (neboli probing), jež slouží k prohlubování odpovědí v konkrétním směru⁷⁰ (viz dotazník *Experiment k diplomové práci* otázka č. 3: Která z konverzací vám byla bližší (méně nepříjemná)? Byli byste si schopni představit další interakci se systémem? – u této otázky byl respondent tázán na to, která z konverzací pro něj

⁶⁷ Strana 21 v této práci.

⁶⁸ TISAROVÁ, D. *Experiment k diplomové práci* [online]. [25. 11. 2013] Dostupné z [www: <https://docs.google.com/forms/d/12gKaSyMwYvOTAwgcUoCwQa_aecgS4Bey25mE1gJLOGw/viewform>](https://docs.google.com/forms/d/12gKaSyMwYvOTAwgcUoCwQa_aecgS4Bey25mE1gJLOGw/viewform).

⁶⁹ HENDL, J., *Kvalitativní výzkum*, s. 164, 169.

⁷⁰ Tamtéž, s. 170.

znamenal menší „utrpení“ a zároveň jsme se chtěli dostat hlouběji a zjistit, zda by dokázal pokračovat v podobném typu rozhovoru se systémem).⁷¹

S tím souvisí největší problém tohoto typu dotazování. Kdybychom se ptali přímo, který z hlasů byl účastníkovi méně nepříjemný, každý uchazeč by vybral pro něj ten „hezčí“ (tzn. ve většině případů ten se syntézou vyšší kvality), nicméně vzhledem k tomu, že jsme se ptali na celou konverzaci, která jim byla méně nepříjemná, vystavili jsme se tak riziku, že účastníci budou odpovídat například na téma hovoru té určité konverzace. Například dotazovaný muž č. 4 s iniciály PB se během druhé části dialogu s Alenou zmínil o zdravotních problémech otce způsobených vážným úrazem, což v něm vzbudilo negativní pocity, o nichž neměl zájem dále mluvit. Naproti tomu v první části rozhovoru s Ivou mluvil o cestování s přítelkyní a podobných pozitivních tématech. V dotazníku poté uvedl, že méně příjemným hlasem pro něj byla právě Alena, se kterou mluvil o zdravotních problémech otce. Můžeme spekulovat, zda negativnější ohodnocení hlasu Aleny bylo způsobeno prostým estetickým cítěním, či tématem hovoru.

I v dotazníku docházelo k míšení různých typů sběru dat, přičemž pro tuto formu bylo typické zahrnout do obsahu otevřené otázky s volnou odpovědí i uzavřené otázky s pouze jednou možností výběru.⁷² Struktura otázek v dotazníku byla členěna tak, aby touha respondenta sdělovat více informací postupem dotazníku vzrůstala. Toho bylo dosaženo tím, že probing byl umístěn na začátek dotazníku, abychom se dozvěděli bezprostřední reakce dotazovaného, jež podporovaly sdílnost účastníka. Začátek dotazníku byl složený z uzavřených otázek s pouze jednou možnou odpovědí, jelikož jsme potřebovali jednoznačnou reakci dotazovaného. Následovaly otázky otevřené, v nichž měl respondent možnost vyjádřit se v neomezeném rozsahu. Tázání pokračovalo polouzavřenými otázkami, ve kterých účastník mohl vybrat jednu z možností, nebo uvést vlastní odpověď. Závěrem na tázaného čekaly nutné otázky demografické a kontextové, jež se zajímaly o identifikační charakteristiky jedince. Na úplném konci dotazníku měl ještě respondent možnost vyjádřit se k čemukoli ohledně dotazníku, rozhovoru či samotného experimentu.⁷³

Při získávání vyplněných dotazníků nebyl zaznamenán s účastníky žádný problém. V některých případech se v dotazníku vyskytoval až přespříliš podrobný popis situací, které vyvstaly během rozhovoru, což mohlo komplikovat objektivní zhodnocení a shrnutí

⁷¹ TISAROVÁ, D. *Experiment k diplomové práci* [online]. [25. 11. 2013] Dostupné z [www: <https://docs.google.com/forms/d/12gKaSyMwYvOTAwgcUoCwQa_aecgS4Bey25mE1gJLOGw/viewform>](https://docs.google.com/forms/d/12gKaSyMwYvOTAwgcUoCwQa_aecgS4Bey25mE1gJLOGw/viewform).

⁷² HENDL, J., *Kvalitativní výzkum*, s. 59.

⁷³ Tamtéž, s. 168 – 169.

dotazníků, protože názory nebyly jednotné, často si protirečily, nebylo možné je snadno kvantifikovat apod. Nicméně i to se dá považovat za hodnotnou zpětnou vazbu.

3.5.4 Prostředí

Prostředí, ve kterém experiment probíhal, se taktéž lišilo případ od případu, jelikož každý účastník měl možnost provádět jak rozhovor, tak vyplnění dotazníku z domova. To samozřejmě přinášelo mnohé výhody – jako větší pohodlí účastníků, minimalizace nervozity i z důvodu uzpůsobení data a času podle individuálních možností zúčastněného. Na druhou stranu z domácího prostředí mohly vyvstávat i určité nedostatky výzkumu, jmenovitě kupříkladu již zmiňované technické problémy a nemožnost jejich rychlého vyřešení kompetentními odborníky. Mimoto experiment mohl být narušen přítomností rodinných příslušníků či spolubydlících a koneckonců mohlo dojít i k záměrnému sabotování průběhu výzkumu vzhledem k nepřítomnosti „autorit“.

3.6 Etické aspekty výzkumu

V souvislosti s výzkumem se objevují i nejrůznější etické problémy běžně spojené s takovými typy zkoumání. Je samozřejmé, že hned v úvodu rozhovoru byl účastník tázán, zda souhlasí s nahráváním. Všichni dotazovaní souhlasili, avšak každý z nich měl (nebo minimálně měl mít) neúplnou představu o tom, s kým hodlá konverzovat. Ta představa měla být následující, a sice: „Budu si povídat s robotem, který nemá pocity, nemůže o mně nikde a nikomu vypovídat, mohu se cítit bezpečně“. Ačkoliv je evidentní, že zúčastnění věděli o nahrávání, u většiny z nich interview neprobíhalo jako běžná mezilidská komunikace. Pocit komunikace se strojem (třebaže domnělým) v nich musel vzbuzovat zvláštní pocity. Účastníci byli přátelští, shovívaví, jelikož si mysleli, že systém plně nevnímá to, co říkají, nebo že to minimálně není schopen interpretovat z lidského hlediska. Existuje velká pravděpodobnost, že kdyby rozhovor probíhal s živým člověkem a měl by charakter stejného scénáře jako v případě experimentu, průběh i výsledek konverzace by byl zcela odlišný. Předně by žádný z účastníků neměl tolik trpělivosti při komunikaci s lidskou bytostí, nevysvětloval by pojmy, nikdo by se neostýchal říct, že ho rozhovor nudí a že už by jej chtěl ukončit, možná by někteří z participujících byli ironičtější, sarkastičtější nebo by odmítali spolupracovat atp. A je naprosto jisté, že by nikdo z účastníků neuváděl tak detailní či osobní informace. Navíc kdyby zúčastnění měli podezření, že osobu výzkumníka znají, přestože jen povrchně, byli by

v kontaktu s touto osobou ještě opatrnější. Pokud ani v osobní interakci nesdílí s cizími lidmi podrobnosti ze svého života, virtuální komunikace by nejspíše nebyla výjimkou.

Tím pak vyvstává onen etický problém výzkumu. Navzdory tomu, že jsme obdrželi souhlas s nahráváním, je otázka, zda je etické nahrávat někoho, kdo vlastně netuší, s kým onen rozhovor vede. Jedinci souhlasili se záznamem rozhovoru člověka a stroje, ale jeden z účastníků té diskuze se proměnil. Ze stroje se bez upozornění „stal“ člověk, takže by se dalo polemizovat nad tím, zda zaniká původní souhlas, nebo jestli se běžně a beztržně může jeden z komunikujících transformovat. Brali bychom to i nadále jako samozřejmost, kdyby tomu bylo naopak a náš partner v komunikaci by se běžně mohl změnit v robota? Účastníci byli oklamáni – výzkumem, výzkumníkem, celým experimentem – aniž by posléze měli možnost své rozhodnutí změnit, koneckonců jejich souhlas byl již zaznamenán a uložen.

Navzdory tomu ti zúčastnění, kteří byli s různě dlouhým časovým odstupem informováni o (ne)existenci Aleny a Ivy, neprojeví žádné známky nevole či pobouření. Většina z nich projevila mírné zklamání nad tím, že takový dialogový systém ještě není funkční, jiní tvrdili, že to celou dobu tušili, a někteří vyjádřili jen pouhé překvapení.

3.7 Hodnocení experimentu

Všichni respondenti reagovali na rozhovor se „systémem“ velice příznivě. Konverzace byla udržována někdy s větším nadšením a chutí, jindy nadšení nebylo tak patrné, komunikace v takových případech byla spíše vlažná či znuděná, ale nikdo experiment prvoplánově nebojkotoval. Polovina dotazovaných uvedla, že je rozhovor s dialogovým systémem velmi zaujal (uvedlo to 5 žen a 10 mužů; zajímavé bylo, že čím více byly ženy informovány v této oblasti, tím více je pokus zaujal; u mužů tento podíl nebyl tak markantní, experiment zaujal i laiky či muže jen zběžně informované v této oblasti). Druhá polovina zhodnotila, s výjimkou jediného respondenta, rozhovor jako „celkem zajímavý“.

Velká část účastníků se přitom věnovala studiu humanitního oboru (56,6 %), pod technický obor spadalo 43,3 % respondentů. Celkem 16,6 % z dotazovaných již byly výdělečně činné osoby. Studenti z humanitního oboru vesměs pocházeli z navazujícího magisterského oboru Teorie a filosofie komunikace na ZČU, jež spolupracuje kromě jiného také s Katedrou kybernetiky ZČU, a studenti mají taktéž přehled o dostupných technologiích, ať už se jedná o ty řečové či o dialogové systémy apod. Obecně řečeno byla pozornost zaměřena na studenty, kteří mohli k tématu uvést vlastní zkušenosti z oboru svého studia, přičemž zejména u techniků existovalo pravděpodobné riziko, že mohou odhalit úlohu

wizarda. O to větší bylo překvapení, že i jedinci s technickým vzděláním byli přesvědčeni o pravosti systému.

Jen dva respondenti během rozhovorů vyjádřili pochybnost, zdali za počítačem nesedí člověk, nicméně když jim tato pochybnost byla vyvrácena, rozhovor pokračoval dál. Uvedu zde příklad jednoho z rozhovorů, ve kterém se účastník chtěl přesvědčit o tom, zda je robot skutečný.

Subjekt: „Chtěla byste mít také telefon, kdybyste nebyla program v počítači?“

Systém: „Ano, asi chtěla. Je to asi pohodlné jej používat.“

Subjekt: „Ano, to je pravda. A jaké to je být program v počítači?“

Systém: „Je to docela nudné.“

Subjekt: „Tomu rozumím. A neodpovídá za vás teďka někdo skutečný?“

Systém: „Ne. Měl jste s telefonem někdy nějaký zvláštní zážitek?“

Subjekt: „Ano měl.“

Systém: „Povíte mi jaký?“

Subjekt: „Ano, povím. Jednou jsem jel na lodi s telefonem v kapse, kde jsem se vyklopil a telefon byl asi minutu až dvě úplně pod vodou, zcela utopený. A takže jsem ho doma rozebral, nechal jsem ho vyschnout a za tři dny, asi, znova fungoval, ale od té doby funguje mnohem pomaleji.“

Systém: „To je velmi zajímavé. Muselo to být nebezpečné.“

Samozřejmě je možné, že ani po zamítnutí dotazu účastníci už systému nevěřili, ale nebylo to znát ani na probíhajícím hovoru, ani na vyplněných dotaznících.

Účastníci rozhovoru se vesměs chovali k systému velmi přátelsky, se zájmem či alespoň se slušností odpovídali na zvědavé a neutuchající dotazy systému, který se snažil rozvinout konverzaci. Většina dobrovolníků se během rozhovoru snažila mluvit spisovně, zřetelně, pomalu a stručně, což se ne vždy projevovalo i v běžné mluvě těchto osob. Někteří ze zúčastněných dokonce měli snahu zrcadlit strojové reakce systému (např. když se systém zeptal: „Co vidíte na obrázku?“, účastník se snažil odpovědět větou, ve které by řekl všechno důležité: „Je tam vyobrazeno Centrální autobusové nádraží v Plzni.“). Jeden z účastníků dokonce tuto snahu zjednodušovat svou řeč vyjádřil v dotazníku jako skutečnost, která ho překvapila: „Překvapilo mě, že se automaticky nutím zjednodušit vlastní řeč, podobně jako kdybych mluvil na malé děti. Snažil jsem se, aby mi systém co možná nejlépe rozuměl.“ Jakmile se systém dotázal na nějaké účastníkovi známější téma, respondenti byli schopni

odpovídat i rozvitou větou, případně se zasmáli, což byl systém schopen doplnit svým umělým „smíchem“. Mnozí z účastníků tuto odchylku od strojové komunikace posléze uvedli v dotaznících jako zábavný aspekt hovoru. Všichni z nich uvedli, že byli mile překvapeni kvalitou, schopnostmi a pokročilostí systému, stejně tak jako úrovní řečové syntézy. K dokreslení strojové nedokonalosti se v každém rozhovoru vyskytoval dotaz na nějaký výraz, dvojsmysl či nejasnou formulaci. Příklad diskuze s jednou účastnicí při debatě o dojíždění z Plzně do Kraslic:

Systém: „Jak dlouho trvá cesta domů?“

Subjekt: „Většinou asi tři hodiny.“

Systém: „Jezdíte sama?“

Subjekt: „Jezdím obvykle s přítelem.“

Systém: „Přítel je také z Kraslice?“

Subjekt: „Říká se z Kraslic. A přítel není z Kraslic, je z Ostrova.“

Systém: „Aha, promiňte....(pauza) Na jakém ostrově bydlí?“

Subjekt: „Ne, Ostrov je město.“ (smích)

Dalších respondentů se systém ptal např. na to, co je to metal, cestář, mše; zkratka jakýkoli termín, u kterého byl předpoklad, že by ho systém nemusel znát. Tím však vzrůstala doba prodlev v komunikaci, protože tyto bezprostřední reakce bylo nutné rychle vymyslet, zformulovat, zapsat do systému a provést syntézu. Takové odezvy tudíž byly proneseny s výraznějším časovým odstupem, než měly předem připravené odpovědi. Pro 30 % respondentů byly právě dlouhé prodlevy tím nejnepříjemnějším faktorem na celém rozhovoru.

Ačkoli prodlevy v komunikaci byly záměrné, zůstává otázkou, zda nebyly spíše na škodu. Vždyť v případě, že bude v blízké budoucnosti sestaven takový dialogový systém, který bude schopný komunikovat s člověkem v podstatě nezávazně na jakékoli téma, existuje opodstatněný předpoklad, že takové prostoje v komunikaci již nebudou aktuální. Někteří respondenti byli prodlevami tak znechuceni, že absolutně neměli chuť si povídat se systémem dál, což bylo samozřejmě znát na celém rozhovoru. Jeden z nich poté uvedl, že důvodem negativního pocitu z rozhovoru byla až ze tří čtvrtin „nekonečná“ doba čekání na odpověď systému. V takových typech konverzace bylo obzvláště těžké nalézt společné téma hovoru, jelikož účastníci odpovídali na dotaz co nejkratší odpovědí, případně prostřednictvím odpovědi ano/ne.

Jedním z příkladů budiž následující rozhovor, u něhož byl subjekt zřetelně znuděný, na nahrávce je slyšet, že během rozhovoru píše na klávesnici, pracuje s dalšími technickými prostředky a tón jeho hlasu je ironický až sarkastický. Tento rozhovor patřil mezi kratší nahrávky, trval 12 minut, 9 sekund, přičemž ten úplně nejkratší rozhovor jen o málo překonal hranici devíti minut. Naopak nejdéle trávající nahrávka dosáhla téměř dvaceti šesti minut.

System: „Co vidíte na obrázku?“

Subjekt: „Nádraží.“

System: „Můžete mi jej více popsat?“

Subjekt: „Jeden, dva, tři, čtyři, pět, šest, sedm, osm vlaků a budova nádraží.

A mnoho kolejí.“

System: „Chodíte často na vlakové nádraží?“

Subjekt: „Ano...“

System: „Jezdíte rád vlakem?“

Subjekt: „Ano...“

System: „Kam nejčastěji jezdíte?“

Subjekt: „Do centra Prahy.“

System: „Odkud jste?“

Subjekt: „Z Prahy.“

System: „Aha a v Praze jezdíte vlakem?“

Subjekt: „Ano.“

K vyhodnocení dotazníkové šetření vzhledem k hypotéze „uncanny valley“ je třeba uvést, že navzdory očekávání 73,3 % respondentů vnímalo jako méně příjemný hlas v komunikaci hlas Aleny (zřetelně umělý hlas ve verzi 2.6), nehledě na to, zda byl Alenin hlas v rozhovoru v prvním pořadí, či až na druhém místě. Negativní reakce se nelišily ani vzhledem k pohlaví respondenta – hlas Aleny vnímalo záporně dvanáct mužů a jedenáct žen. Tento výsledek jsem se snažila potvrdit zrcadlově položenou otázkou „Který hlas byste si dokázali představit jako partnera ke komunikaci (k interakci)?“, v odpovědi dominovala ve čtrnácti případech Iva a ve dvanácti případech byly uvedeny obě. Otázka, která se vztahovala spíše k dojmu z celkové konverzace, také spíše potvrdila převládající typ odpovědí, jelikož v ní osm mužů a dvanáct žen označilo konverzaci s Ivou jako tu příjemnější.

Tato data však budou v průběhu práce popsána zevrubněji pomocí statistických metod a tabulek, spolu s formulováním celkových výsledků a předpokládaného závěru.

3.8 Důvod zrealizování experimentu

Důvod uspořádání experimentu byl prostý, chtěli jsme ověřit teoretickou hypotézu o vztahu syntetické řeči a teorie „uncanny valley“ v praxi. Existují mnohá zkoumání o interakci robotů a lidí, eventuálně o komunikaci lidí a strojů, které ani neměly působit lidsky. Proto nám přišlo zajímavé snažit se zjistit, zda by kontakt se syntetickou řečí mohl přinášet stejné či podobné výsledky a pocity, nebo zda bude lidské vnímání ke strojově vytvořené řeči shovívavější. Zároveň byl hlavní inspirací k provedení experimentu pokus Senior Companion. Zpočátku bylo v plánu nasimulovat téměř shodný experiment, který probíhal právě v pokusu Senior Companion. Jedním z původně plánovaných rozdílů bylo zapojení mladších účastníků do zmíněného projektu. Takový projekt se však ukázal časově velmi náročným jak na provedení, tak na organizaci. Proto jsme přistoupili k jednodušší variantě experimentu prostřednictvím skypového rozhovoru s účastníky, která také nabízela poměrně velké množství možností uskutečnění a vyhodnocení.

Jedna z věcí, která byla na myšlence zrealizování výzkumu nejvíce odlišná od klasických témat diplomových prací, byla právě možnost převést teoretické poznatky hypotézy do praxe a ověřit je v jejich reálném provedení. Bylo třeba otestovat, zdali je hypotéza japonského robotika platná i dnes, zhruba 44 let od jejího prvotního formulování. Pokud by se prokázalo, že hypotéza je aktuální i v současné době, otázkou by stále zůstávala redukovatelnost robotů a strojů na pouhou schopnost řeči, ve které využíváme pouze auditivního vnímání, nikoliv i vizuálního, jako je tomu například u robotů nebo jako tomu bylo i v případě pokusu Senior Companion. Je pravdou, že platnost původní hypotézy směřovala za vztahem a interakcí lidí s roboty či s konkrétními fyzickými věcmi. Nicméně v tomto experimentu jsme se snažili posoudit spíše hranice verbální komunikace lidí a strojů. Snažili jsme se zhodnotit přirozenost komunikace a univerzálnost lidských úsudků – jestli by pro většinu lidí byla táž komunikace srovnatelná, jestli, popřípadě jak velké rozdíly by se daly vysledovat v hodnocení různými jedinci a zda by bylo vůbec možné najít paralely mezi jednotlivými názory. Celý experiment tak primárně směřoval ke snaze prokázat, že hypotéza „uncanny valley“ je platná i ve vztahu k syntetické řeči.

Cílem experimentu bylo ověřit, zda preference syntetického hlasu nějakým způsobem souvisí s kvalitou jeho syntézy. Jinými slovy, chtěli jsme zjistit, zda je možné, že pro většinu účastníků měla úroveň hlasu zásadní vliv na ohodnocení tohoto hlasu. Tuto tezi jsme se snažili potvrdit nebo vyvrátit tím, že jsme hodnotili různá kritéria v rozhovorech a dotaznících a hledali jsme mezi nimi souvislosti.

4 Analytická část

4.1 Popis výsledků experimentu

Výsledky, ke kterým jsme dospěli v experimentech či v dotaznících, mohly být ovlivněny velkým množstvím různých faktorů. Jedním z nich je například rozlišení pohlaví jednotlivých účastníků, které mohlo mít vliv na výsledné rozdílné vnímání hlasu u mužů a žen. Následující možnou příčinou mohlo být i pořadí hlasů, ve kterém byly hlasy předvedeny účastníkovi, potažmo se původcem rozdílných výsledků mohlo stát i pořadí obrázku. Ostatními hodnotami, na kterých záleželo, byla i obeznamenost účastníků s řečovými technologiemi, případně vzdělání v humanitním či technickém oboru.

Co bylo tím důležitým činitelem, který měl za následek preferování jednoho či druhého hlasu? Popíšeme zde všechny signifikantní aspekty, které mohly ovlivnit výsledný závěr.

4.2 Nástroje vyhodnocování

Vzhledem k tomu, že primární roli v kvantitativním výzkumu zastává testování hypotéz, zaměřili jsme svou pozornost na nejrůznější hypotézy, které bylo možné vytěžit z dat, jež jsme měli k dispozici. Vytvořili jsme si řadu „pracovních“ domněnek – tzv. nulových hypotéz, které sloužily k odhalení závislosti jedné veličiny na druhé. Nulová hypotéza pak předpovídala, jak velká by mohla být souvislost mezi proměnnými, kdyby se ukázalo, že zmíněné předpoklady jsou založené na pravdě.⁷⁴

Klíčová data, se kterými jsme pracovali, pocházela buď z rozhovorů s účastníky, z vyplněných dotazníků, nebo ze samotných charakteristických vlastností rozhovorů (např. trvání rozhovoru, délka vstupu jednotlivých hlasů v rozhovoru apod.). V nulových hypotézách poté dominovaly nominální proměnné, jako je pohlaví, technická či humanitní orientace jedince atd., spolu se zásadními proměnnými intervalovými, jakými byly například věk účastníka a délka trvání individuálních rozhovorů.⁷⁵

⁷⁴ DISMAN, M. *Jak se vyrábí sociologická znalost*, s. 180.

⁷⁵ Tamtéž, s. 188 – 189.

Hlavními nástroji testování hypotéz byl aritmetický průměr (jenž lze vyjádřit jako „intervalový popis střední hodnoty“⁷⁶), směrodatná odchylka a tzv. chí-kvadrát test.

Směrodatná odchylka sleduje zejména to, jestli se vzorky v souboru vyznačují stejnorodostí, a definuje tak například hodnotu, kterou se liší nízká veličina od té průměrné. Pearsonův chí-kvadrát test je statistický nástroj, který umožňuje odhadnout míru pravděpodobnosti dvou proměnných. Díky této metodě lze odhalit potenciální vztah mezi dvěma veličinami.

Zmíněným testem docházelo k ověřování hypotéz v prostoru kontingenčních tabulek, v nichž bylo primárním úkolem porovnat skutečné četnosti hodnot vzhledem k očekávaným četnostem. Skutečné četnosti byly ty reálné hodnoty, ke kterým jsme došli prostřednictvím výzkumu, a k hodnotám očekávaných četností jsme se dostali jednoduchým výpočtem. Tabulka očekávaných četností ukazuje, jak by to vypadalo, kdyby hodnoty v tabulce byly nezávislé, přičemž primární byl předpoklad, že platí určitá nulová (pracovní) hypotéza. V našem případě docházelo k ověřování přítomnosti vztahu mezi veličinami, používali jsme tedy test nezávislosti, který nám pomohl odhalit (ne)závislost mezi jednotlivými členy tabulky. Nulová hypotéza tak vždy předpokládala, že jednotlivé prvky v tabulce jsou na sobě navzájem nezávislé. Výpočet nám tuto skutečnost posléze buď potvrdil, nebo vyvrátil.

K potvrzení, či vyvrácení došlo pomocí tzv. p-hodnoty, která nám odhalila velikost rozdílů mezi skutečnými a očekávanými četnostmi a zároveň ukázala, jak moc je pravděpodobné, že by se v jedné hypotéze mohly společně vyskytovat konkrétní skutečné a očekávané hodnoty. P-hodnoty se mohly pohybovat v rozmezí od nuly do jedné. Pokud by byla p-hodnota menší než 0,05, musela by být nulová hypotéza zamítnuta, tzn. že porovnávané hodnoty by na sobě byly nezávislé. V případě takového výsledku máme totiž poměrně vysokou míru pravděpodobnosti, že jsme hypotézu zamítli oprávněně, máme 95% jistotu, že je v takovém případě nulová hypotéza neplatná.⁷⁷

⁷⁶ DISMAN, M. *Jak se vyrábí sociologická znalost*, s. 190.

⁷⁷ Autor neuveden. *III. Cvičení ze statistiky* [online]. [21. 3. 2013] Dostupné z [www: <http://ulb.upol.cz/praktikum/statistika3.pdf>](http://ulb.upol.cz/praktikum/statistika3.pdf).

4.3 Vyhodnocení výsledků

Pomocí kontingenčních tabulek a Pearsonovu chí-kvadrát testu jsme hodnotili celkem třináct hypotéz. Sedm z nich bylo vyhodnoceno pouze na základě výsledků dotazníkového šetření a pro zbylých šest bylo využito také informací z nahrávek rozhovorů. Většina z těchto třinácti hypotéz přinesla jen poměrně nejasné výsledky, a to především kvůli relativně nízkému počtu zúčastněných. Účastníků bylo celkem třicet, avšak pokud jsme chtěli zhodnotit například zvlášť mužské a zvlášť ženské pohlaví, dostali jsme se už jen na polovinu z původních třiceti a s přihlédnutím k široké nabídce možných odpovědí v dotaznících se počty možných skupin ještě dále drobily.

První hypotéza, která byla podrobena zkoumání, se snažila vyloučit vliv pořadí přehrání hlasu ve vztahu k následným preferencím jednoho z hlasů. Nešlo o posouzení pokročilosti či kvality syntézy, ale jednalo se o víceméně subjektivní zhodnocení toho, jak se účastníkovi hlas líbil, jak mu byl příjemný na poslech. Tomu, aby dotazovaný vnímal a soudil pouze kvalitu syntetického hlasu, jsme se snažili vyhnout již v instrukcích k experimentu, které byly účastníkům zasílány před provedeným výzkumem.⁷⁸ Nejprve jsme chtěli posoudit, zda pořadí hlasů bylo tím rozhodujícím faktorem při odlišném hodnocení hlasů. Pořadí hlasů bylo náhodně měněno u každého rozhovoru, abychom vyloučili to, že by všem účastníkům vyhovoval jen jeden typ hlasu. Bylo totiž možné, že by někteří účastníci mohli preferovat například první hlas, protože by si ho po zkušenosti s druhým hlasem v mysli určitým způsobem zidealizovali, popřípadě by to samozřejmě mohlo být i naopak – druhý hlas by mohli mít čerstvě v paměti, a tak by se jim první z nich zdál nutně horší. Z tohoto důvodu bylo tedy během hovorů střídáno pořadí hlasů, abychom posléze mohli odhalit, zda to v názorech účastníků bylo signifikantním jevem, či nikoliv. Nulová hypotéza k otázce „Který z hlasů byl pro vás méně příjemný v komunikaci?“ zněla: „Preference hlasu není závislá na pořadí přehrání.“ Jinými slovy jsme chtěli prokázat, že není signifikantní rozdíl mezi znakem „první Iva“ a znakem „první Alena“ a že tyto charakteristiky nemají souvislost se znakem „horší Iva“ a „horší Alena“. P-hodnota chí-kvadrát testu byla 0,7419, tím pádem jsme hypotézu museli vyvrátit se závěrem, že je možné, že na pořadí nezáleží. Museli jsme to však vyjádřit jako pouhou možnost, jelikož p-hodnota se nepřibližuje hodnotě 1 příliš výrazně. Na druhou stranu je ovšem p-hodnota výrazně vyšší než 0,05, takže bychom mohli vyjádřit vysokou pravděpodobnost tohoto výsledku, tj. že na pořadí přehrání nezáleží. Pro představu

⁷⁸ Strana 21 této práce.

přikládám zmiňovanou tabulku skutečných četností a očekávaných četností pro právě popisovanou hypotézu.

Otázka: **Který z hlasů byl pro vás méně příjemný v komunikaci?**

nulová hypotéza (H₀): Preference hlasu nezávisí na pořadí přehrání.

skutečné četnosti

	horší Iva	horší Alena	Celkem
první Iva	4	11	15
první Alena	3	11	14
Celkem	7	22	29

očekávané četnosti

	horší Iva	horší Alena	
první Iva	3,6207	11,3793	15
první Alena	3,3793	10,6207	14
Celkem	7	22	29

p-hodnota chí-kvadrát testu: 0,741854468

H₀ je s vysokou pravděpodobností platná, na pořadí přehrání nezáleží.

Pořadí přehrání jsme považovali za jeden z velmi důležitých faktorů, proto jsme ověřovali hypotézu: „Preference hlasu nezávisí na pořadí přehrání.“ také zvlášť pro muže a zvlášť pro ženy. Jedenáct mužů i jedenáct žen uvedlo jako méně příjemný hlas Aleny. U obou pohlaví se potvrdila nulová hypotéza, tedy že není významná souvislost mezi pořadím přehrání a následnou preferencí jednoho či druhého syntetického hlasu. U mužů byla p-hodnota chí-kvadrát testu podstatně vyšší než u žen. U mužů jsme se dostali až na hodnotu 0,9227, což znamená, že pravděpodobnost, že je hypotéza platná, přesahuje 92 %. U žen je tato pravděpodobnost vyšší než 64 %, což je samozřejmě také signifikantní výsledek. Ani

ženy, ani muži tedy nebyli ovlivněni pořadím přehrání hlasu. Pořadí přehrání bylo zcela nezávislé na celkových preferencích účastníků, stejně tak pořadí a preference nijak výrazně nevybočovaly ani odděleně při posuzování pohlaví účastníků. Hodnocení „horší Iva“ a „horší Alena“ je tedy absolutně nenáhodné od pořadí přehrání, jinak řečeno, účastník nebyl vůbec ovlivněn tím, který z hlasů slyšel jako první.

Z toho důvodu, že jsme se chtěli naprosto přesvědčit o nezávislosti pohlaví na preferencích účastníků, jsme posuzovali vliv pohlaví ještě u další z otázek, která zněla: „Který hlas byste si dokázali představit jako partnera v komunikaci?“. Snažili jsme se zjistit, který z hlasů byl akceptován lépe pro případnou další konverzaci, přičemž zde bylo možno vybírat z odpovědí „Alena“, „Iva“, „Obě“, „Ani jedna“. Nulová hypotéza pak zněla: „Preference hlasu nezávisí na pohlaví.“. P-hodnota testu byla velmi vysoká, dokonce přesáhla hodnotu 0,9, takže se nám již ve druhém případě potvrdilo, že preference syntetického hlasu byla nezávislá na pohlaví účastníka.

Další hypotézou, kterou jsme chtěli ověřit, byla otázka obeznámenosti účastníků s řečovými technologiemi. Pokud by se nulová hypotéza nepotvrdila, existoval by zde vztah mezi preferencí syntetického hlasu a mírou pokročilosti účastníka v oblasti řečových technologií. Zúčastnění mohli své schopnosti ohodnotit jako naprosto laické nebo v podobě znalosti povrchní, běžné či pokročilé. Hodnota chí-kvadrát testu však dosáhla velikosti 0,3431, takže předpoklad, že preference hlasu nebyly ovlivněny mírou obeznámenosti účastníka s řečovými technologiemi, se naplnil.

Zajímal nás také vliv typu vzdělání účastníků na zaujetí rozhovorem. Většina z účastníků uvedla, že je rozhovor zaujal „velmi“, případně „celkem“. Nulová hypotéza v tomto případě tvrdila, že míra zaujetí rozhovorem nezávisí na typu vzdělání. Po provedení chí-kvadrát testu jsme dospěli k p-hodnotě 0,0890, což by potvrdilo nulovou hypotézu, tzn. že míra zaujetí rozhovorem nebyla udávána oborem vzdělání. Avšak p-hodnota je jen o tři setiny vyšší než hraniční hodnota 0,05, takže tuto situaci bychom měli zařadit spíše mezi hraniční případy. Nejsme schopni s jistotou určit, kam by se hodnoty posunuly, kdybychom měli větší vzorek testovaných subjektů. Hypotézu, že zaujatost nezávisí na vzdělání účastníka, nemůžeme tedy stoprocentně ani přijmout, ani odmítnout.

Ve snaze interpretovat získaná data co nejpodrobněji, jsme zkoumali, zda by obor, ve kterém je účastník vzdělaný, mohl mít vliv na to, který z hlasů nakonec byl preferován. Předpokládali jsme, že například vzdělání jedince v technické sféře by mohlo nějakým způsobem ovlivnit odpovědi, a tudíž i preference jednotlivých hlasů. Dalo by se očekávat, že technické vzdělání účastníky předurčuje k tomu, že jsou otevřenější k technologickému

pokroku, mají více zkušeností v tomto oboru, a proto by mohli lépe přijímat pokročilejší typ hlasové syntézy. Testovali jsme tedy následující nulovou hypotézu: „Preference hlasu je nezávislá na oboru studia.“, přičemž jsme se snažili odhalit případný vztah mezi znaky „horší Iva“, „horší Alena“ a znaky „technický obor“ a „humanitní obor“. P-hodnota chí-kvadrát testu byla nižší než 0,05, dosáhla pouze hodnoty 0,0107, takže nulovou hypotézu bylo třeba zamítnout. Obor studia tedy nějakým způsobem ovlivnil preference hlasu. Z tohoto typu testu se už bohužel nedozvíme, jakým způsobem může studium konkrétního oboru ovlivňovat určité preference, ale důležitým závěrem je fakt, že hodnocení hlasů v experimentu souviselo s faktorem humanitního či technického zaměření jedince. Tato hypotéza spolu s tabulkami skutečných a očekávaných četností si zaslouží zvláštní pozornost, z toho důvodu je přiložena níže k nahlédnutí.

Otázka: Vliv technického či humanitního vzdělání účastníků na preference

H₀: Preference hlasu je nezávislá na oboru studia.

skutečné četnosti

	technický obor	humanitní obor	Celkem
horší Alena	12	10	22
horší Iva	0	7	7
Celkem	12	17	29

očekávané četnosti

	technický obor	humanitní obor	Celkem
horší Alena	9,103448276	12,89655172	22
horší Iva	2,896551724	4,103448276	7
Celkem	12	17	29

p-hodnota chí-kvadrát testu: 0,010706648

H₀ je neplatná, tzn. že preference hlasu závisí na oboru studia.

V případě nutnosti zamítnutí nulové hypotézy jsme se snažili potvrdit data tabulkou, která kvantitativně vystihuje nulovou hypotézu.

	technický obor	humanitní obor	Celkem
horší Alena	100%	59%	1,588235294
horší Iva	0%	41%	0,411764706
Celkem	1	1	

100 % účastníků z technického oboru preferovalo Ivu, z humanitního oboru jich bylo 59 %. Tato tabulka tedy potvrzuje výsledky chí-kvadrát testu. Technicky zaměřeným účastníkům pokročilejší syntéza (hlas Ivy) ani v jednom z případů nepřipadala horší než méně dokonalá syntéza (hlas Aleny). Ve všech případech uvedli, že horší pro ně byl hlas Aleny. Zdá se, že nějakým způsobem pro ně byl problémem větší podíl robotičnosti v hlase Aleny. Humanitně zaměření jedinci naopak nebyli tímto faktorem ovlivněni natolik, aby se to významně odrazilo v jejich hodnocení. Ať už to bylo z toho důvodu, že jim na tom tolik nezáleželo nebo si nebyli vědomi rozdílu v hlasech, případně pro ně byly důležité jiné charakteristiky (celkový pocit z rozhovoru, téma, forma otázek apod.). „Technici“ byli výrazně negativně naladěni vůči Aleně, u „humanistů“ se naladění proti Aleně neprojevilo tak markantně. To však neznamená, že by humanitně zaměření jedinci přijímali hlas Aleny lépe, zkrátka jej jen tak často neodmítali jako technicky zaměření účastníci. Kdyby se experimentu zúčastnily jen technicky vzdělané osoby, mohlo by dojít k výraznému zkreslení výsledků. Důležitým závěrem této hypotézy tedy zůstává, že vzdělání účastníků ovlivňuje preference hlasu.

Další vyhodnocování dat se uskutečňovalo prostřednictvím informací získaných ze samotných nahrávek rozhovorů. Mezi typy těchto dat patřila jak celková délka rozhovoru, tak zvláště délka rozhovoru s Ivou a délka rozhovoru s Alenou, přičemž trvání rozhovoru s jedním hlasem se měřilo od prvního vstupu hlasu až do rozloučení hlasu a účastníka. Proto ne vždy dá součet trvání rozhovoru s Ivou a rozhovoru s Alenou dohromady čas trvání celé nahrávky. Proto jsme také zhodnotili délku a počet vstupů jednotlivých syntetických hlasů, aby bylo možné objektivněji porovnávat jednotlivé rozhovory.

Celkově se experimentu zúčastnilo třicet jedinců, avšak dvě nahrávky se nezachovaly, takže jsme hodnotili pouze dvacet osm nahrávek, tedy dvacet osm zúčastněných, z toho

čtrnáct mužů a čtrnáct žen. Z tohoto počtu zúčastněných třináct bylo technicky zaměřených a patnáct humanitně zaměřených. Většinu charakteristik rozhovorů jsme porovnávali vzhledem k preferencím hlasu, přičemž jeden účastník ohodnotil oba hlasy jako rovnocenné, byl tedy z většiny vyhodnocování vyjmut, aby se zachovalo hodnocení distinkce horší Iva – horší Alena.

Průměrná délka celého rozhovoru byla šestnáct minut, patnáct sekund. Pomocí směrodatné odchylky jsme z průměru určili délku krátkého rozhovoru (dvanáct minut, třicet osm sekund) a délku dlouhého rozhovoru (devatenáct minut, padesát dva sekund). Krátký rozhovor měli čtyři účastníci, dlouhý rozhovor pouze tři účastníci a zbytek, tj. dvacet jedna účastníků, se vešel do průměrné délky rozhovoru. Průměrná délka rozhovoru s Ivou byla osm minut a dvě sekundy. Trvání rozhovoru s Alenou zabralo v průměru sedm minut, třináct sekund.

Z těchto údajů jsme se snažili vyhodnotit, zda trvání celého rozhovoru mohlo ovlivnit, jaký hlas následně účastníci ve větší míře preferovali. Nulovou hypotézou byla teze, že preference hlasu nezávisí na délce rozhovoru. P-hodnota tabulky přesáhla 0,5394, takže tato hypotéza se nepotvrdila. Zkoumali jsme tedy odděleně délku rozhovoru s Ivou a zvlášť délku rozhovoru s Alenou. Bylo možné předpokládat, že pokud byl účastník delší dobu vystaven jednomu z hlasů, mohlo to následně ovlivnit jeho preference, ať už pozitivně, nebo negativně. V pozitivním hodnocení by se to mohlo odrazit tak, že si účastník s hlasem lépe „popovídal“, zatímco v negativním posouzení by to mohlo reflektovat jakousi přímou úměrnost – čím delší dobu byl účastník vystaven určitému hlasu, tím rostl jeho nepříjemný pocit z hlasu. Tyto předpoklady se však nepotvrdily. Hodnotili jsme obě hypotézy – tedy závislost délky rozhovoru s Ivou ve vztahu k preferencím hlasu, stejně tak jako závislost délky rozhovoru s Alenou vzhledem k preferencím hlasu – a v obou případech se nulová hypotéza potvrdila, tzn. potvrdilo se, že délka rozhovoru jak s Ivou, tak s Alenou neměla souvislost s preferencemi hlasů. Jediným rozdílem mezi těmito dvěma hypotézami bylo, že p-hodnota při posuzování délky rozhovoru s Ivou byla vyšší (0,5764) než p-hodnota vypovídající o vlivu délky rozhovoru s Alenou (0,25). To by svědčilo o tom, že preference určitého hlasu závisely méně na délce především v rozhovorech s Ivou. Vše je posuzováno vzhledem k výsledkům, které jsme měli k dispozici, je tedy možné, že kdybychom disponovali více vzorky, dostali bychom výsledky odlišné.

Další z faktorů, který jsme posuzovali vzhledem k délce rozhovoru, bylo opět humanitní či technické zaměření účastníků. Zde však bylo problémem, že většina rozhovorů

se bez potíží vešla do průměrné délky hovoru. Nulová hypotéza, která měla za úkol zjistit, zda je délka rozhovoru nezávislá na zaměření účastníka, se tedy opět potvrdila.

Poslední hypotézy, které jsme testovali, hodnotily počet vstupů jednotlivých hlasů vzhledem k následným preferencím. Hlas Ivy měl v rozhovorech průměrně 30,8 vstupů. Rozlišili jsme tak rozhovor s nízkým, středním a vysokým počtem vstupů, přičemž většina rozhovorů (celkem osmnáct) se opět držela v oblasti průměru. P-hodnota nulové hypotézy „Preference hlasu nezáležela na počtu vstupů, které měla Iva v rozhovoru.“ se výrazně přiblížila hodnotě 1 (p-hodnota byla 0,9776). Tato hypotéza je tedy s vysokou pravděpodobností platná a mohli jsme vyloučit vliv počtu vstupu hlasů vzhledem k preferencím.

Totéž jsme ověřili i vzhledem k počtu vstupů Aleny, jichž bylo průměrně 28,7 na jeden rozhovor. Nulová hypotéza se také potvrdila, tedy preference hlasu nezáleží na počtu vstupů, které měla Alena v rozhovoru, avšak p-hodnota nebyla tak vysoká jako u výsledků Ivy. P-hodnota dosáhla přibližně velikosti 0,2522.

4.3.1 Rozhovory

Jedním z možných problémů porovnatelnosti výsledků experimentu mohla být individuálnost rozhovorů. Tento fenomén lze sledovat i vzhledem k technickým náležitostem hovorů, jako je jejich trvání, nebo z podstatně subjektivnějších hodnocení rozhovorů, jako je například individuální přístup účastníků (zvědavost, touha vést rozhovor a odpovídat, popřípadě stud, nemluvnost apod.). Každý hovor se lišil nejen počtem minut, ale i počtem vstupů jak účastníka, tak syntetického hlasu. Počet vstupů také záležel na různých odchylkách od běžného průběhu rozhovoru. V případě, že se během rozhovoru vyskytly technické problémy, počet vstupů samozřejmě stoupl úměrně tomu, jak dlouho potíže přetrvávaly, eventuálně jakým způsobem a zda vůbec se je podařilo eliminovat.

Z průměrného počtu vstupů jak Aleny, tak Ivy lze vypočítat počet průměrných vstupů na jeden celý rozhovor, což je zhruba 29,7 vstupů. Přitom počet otázek, které se nutně v každém rozhovoru přesně opakovaly, byl 23. Mezi tyto otázky patřily vstupy, ve kterých docházelo k představení umělé mluvčí, k podání instrukcí účastníkům především s informacemi o tom, co se bude dít v následujících minutách, co se od zúčastněného bude požadovat na závěr rozhovoru, následovaly obecné otázky ohledně obrázku a nakonec se do počtu univerzálních vstupů počítalo i rozloučení hlasu s účastníkem. Typy těchto dotazů jsou

uvedeny již v kapitole 3.4 „Experimentální protokol“.⁷⁹ Shodnost vstupů u všech rozhovorů se tedy pohybovala přibližně kolem 77 %, což by bylo možné pokládat za dostačující shodný faktor k tomu, aby byly rozhovory nějakým způsobem srovnávány.

4.3.2 Data z dotazníků

Pokud jde o dotazníky, vyskytovalo se v nich celkem osmnáct otázek. Uvedené dotazy by bylo možné rozdělit do tří skupin. V první skupině byly dotazy, které se snažily získat informace o účastníkovi – ptaly se tedy na iniciály, věk, vzdělání, datum, kdy se jedinec zúčastnil dotazníku apod. Informace z této skupiny dotazů jsme již analyzovali společně se souhrnem otázek spadajících do druhé skupiny. Do ní patřily především otázky, u kterých bylo nutné vybrat si z fixně nabízených odpovědí. To bylo potřeba především kvůli pozdějšímu kvantitativnímu vyjádření odpovědí. Třetí skupinu otázek tvořily dotazy s možností volné odpovědi. V nich měl účastník možnost vyjádřit svůj názor na syntetické hlasy, na právě ukončený rozhovor, popřípadě na celý průběh experimentu.

Zúčastnění měli za úkol shrnout celý rozhovor jedním slovem, přičemž převažovaly hlavně pozitivní odpovědi. 23 % popsalo rozhovor jako zajímavý, přibližně 32 % jej zhodnotilo jako zvláštní či překvapivý, ostatní kladné reakce (jako např. pokrokový, obdivuhodný, plnohodnotný atd.) zaujímaly přes 26 %. Zhodnocení nevelké rychlosti systému, eventuálně přítomnost pauz apod. vyjádřilo jedním slovem 10 % zúčastněných a téměř 7 % účastníků považovalo rozhovor spíše za komický.

Další otázka zjišťovala, zda účastníky během rozhovoru něco překvapilo. Odpovědi dotazovaných se typově poměrně odlišovaly, nebylo možné všechny odpovědi sloučit do několika variant odezev. Avšak nejmarkantnější podíl podobných odpovědí vyjadřoval údiv nad vhodnými, rychlými a přirozenými reakcemi, které ocenilo celkem 26,67 % respondentů, a překvapení nad „smyslem pro humor stroje“, které vyjádřilo 16,67 % účastníků.

Následoval také dotaz, co bylo pro účastníky na rozhovoru nejpříjemnější. Odpovědi byly taktéž velmi různorodé, nicméně 13 % účastníků uvedlo jako příjemný faktor rozhovorů schopnost empatie „robotů“ spolu s příjemností a slušností, se kterou systém vystupoval. Jiná větší skupina společných odpovědí označila schopnost naslouchat a rozvést hovor. Touto pozitivní charakteristikou ocenilo rozhovor přes 43 % účastníků.

Dále nás zajímalo, co bylo pro účastníky naopak na rozhovoru nejméně příjemné. 30 % dotazovaných uvedlo, že to, co jim na rozhovorech nejvíce vadilo, bylo právě dlouhé

⁷⁹ Strana 22 v této práci.

čekání na odezvu systému a časté prodlevy v komunikaci. Pro 27 % byla překážkou hlavně umělost hlasu, nezvyklá intonace, nedokonalá návaznost slov, přičemž většina lidí v tomto ohledu negativně hodnotila spíše hlas s méně pokročilou syntézou (hlas Aleny).

V otázce „Při jaké situaci byste si dovedli představit interakci s takovým typem hlasu?“ se většina respondentů shodla na tom, že by se takový systém dal použít prakticky v jakýchkoliv informačních systémech, při rozmluvách s operátory, pro zrychlení vyhledávání nejrůznějších informací, na přepážkách úřadů, při telefonických objednávkách, případně i v situacích, kdy se člověk potřebuje někomu svěřit a vypovídat se.

Většina účastníků nevyužila posledního dotazu, kde měli respondenti možnost vyjádřit cokoli, co by chtěli dodat k rozhovoru, syntetickým hlasům či k experimentu jako takovému. Pokud se vyjádřili, tak vesměs obdobně, jako již v předchozích odpovědích. Převažovaly názory, že experiment byl zajímavý, že si zatím nedokážou představit, že by se s podobnými systémy běžně setkávali v každodenním životě, nicméně zároveň vyjádřili přesvědčení, že by si postupně zvykli i na nedostatky, které systém vykazoval.

4.4 Shrnutí experimentu

Celá práce se vlastně snažila jistým způsobem potvrdit, nebo vyvrátit nulovou hypotézu, která by zněla: „Preference hlasu nezáleží na kvalitě syntézy.“. K této nulové hypotéze však není možné udělat kontingenční tabulku, jelikož nelze porovnávat lepší a horší syntézu hlasu z toho důvodu, že všichni účastníci měli v rozhovoru obě syntézy.

Celkově se tedy k posouzení hlasů jasně vyjádřilo dvacet devět účastníků, sedmi z nich se jako horší hlas jevil hlas Ivy, dvaceti dvěma zúčastněným připadal horší hlas Aleny. Třicátý respondent ohodnotil oba hlasy jako rovnocenné, přičemž těžko soudit, zda by je oba ocenil jako dobré, nebo oba jako špatné. Proto hodnotíme pouze dvacet devět názorů, které byly vyhraněné jedním, či druhým směrem.

Poměr sedm ku dvaceti dvěma je v zásadě signifikantní, ale nemůžeme jím bez zaváhání teorii vztahující se k řečové syntéze a jejímu vztahu k hypotéze „uncanny valley“ jednoznačně potvrdit. Zároveň však tento poměr není ani dostatečným důvodem k tomu, abychom hypotézu zcela vyvrátili. Závěry tedy nejsou zcela jasné ani jedním směrem, a to především kvůli počtu zúčastněných. Třicet respondentů není zanedbatelný počet, avšak kdyby byl počet účastníku vyšší, statistické výsledky by mohly být považovány za podstatně relevantnější. Další možností, proč nešlo hypotézu výslovně přijmout ani odmítnout, by mohlo být to, že ji nelze objektivně zhodnotit, protože člověk takové věci posuzuje pouze

subjektivně a není schopen odhlédnout od dosavadních zkušeností, zážitků, a především se nedokáže zřeknout estetického hodnocení faktů. Jistý je ovšem fakt, že přes 73 % zúčastněných preferovalo Ivu, ať už jejich důvody byly jakékoliv.

4.4.1 Problémy

V průběhu experimentu jsme se setkávali s nejrůznějšími potížemi, které znesnadňovaly či zdržovaly jeho průběh i samotné vyhodnocování výsledků. Tyto problémy se zprvu týkaly například i zajištění potřebného množství účastníků. Ačkoli se pro účastníka jednalo maximálně o hodinový úsek, ve kterém bylo nutné zaobírat se rozhovorem a dotazníkem, mnoho z původních zájemců účast odřeklo, někteří z nich nereagovali na e-maily, ačkoli předtím účast přislíbili apod. I z tohoto důvodu se samotná realizace poměrně zdržela, rozhovory tak probíhaly celý měsíc.

Další potíže představovaly především technické problémy, které se ukázaly jako velice naléhavé. Pokud měla jedna ze zúčastněných stran během rozhovoru problémy s připojením, průběh rozhovoru byl závažně narušen a docházelo k nutnosti často opakovat promluvy apod. To se svým způsobem mohlo projevit i při zpracování výsledků, kdy například jeden z posledních rozhovorů utrpěl v jeho průběhu několik zvukových výpadků, dokonce muselo dojít k obnovení skypového spojení a délka rozhovoru se prodloužila minimálně o sedm minut. Rozhovor, který mohl mít průměrnou délku, se tak vlivem technických problémů prodloužil ze středního trvání na trvání dlouhé, nemluvě o počtu vstupů, který nutností opakovat promluvy vzrostl nejméně dvojnásobně.

Zaznamenali jsme také četné problémy s dotazníky, jelikož někteří jedinci, možná vlivem špatného internetového spojení, odeslali více stejných dotazníků, které se již nepodařilo odstranit z některých souhrnných vyhodnocení dotazníků. Mnoho respondentů si také nepozorně přečetlo instrukce k vyplňování dotazníků či zadání otázek a odpovídalo na něco jiného, popřípadě si někteří z nich spletli jména syntetických hlasů a poté bylo poměrně náročné jejich chyby odhalit a napravit.

4.4.2 Návrhy řešení nedostatků

Pro eliminaci zmiňovaných problémů by bylo zapotřebí zrealizovat obdobný typ zkoumání, ve kterém by se však podařilo nějakým způsobem vyloučit popisované potíže a ve kterém by se zároveň povedlo potlačit subjektivní hodnocení účastníků. Větší míry objektivity

by bylo možné docílit například navýšením počtu syntetických hlasů v rozhovorech, podrobnějšími dotazníky, a především vyšším počtem zúčastněných.

Statisticky relevantnější by také bylo, kdyby účastníci experimentu pocházeli z různých věkových skupin a disponovali různorodějším vzděláním, protože jak se ukázalo, faktor oboru vzdělání byl důležitou proměnnou v posuzování syntetických hlasů. Hypotéza, která posuzovala vliv technického nebo humanitního vzdělání na určování preferencí hlasů, se ukázala jako klíčová. Prokázal se určitý vliv typu vzdělání účastníka na preference v oblasti řečové syntézy. Bylo by tedy přínosné podrobit tuto zásadní hypotézu podrobnějšímu zkoumání s vidinou získání univerzálnějších výsledků.

Dalším z možných vylepšení by mohlo být rovnoměrnější střídání syntetických hlasů v rozhovorech, čímž by se mohlo docílit vyšší možnosti srovnání odpovědí účastníků. Ve výzkumu se totiž nepodařilo zachovat stejný počet mužů, kteří měli Ivu jako první, a těch, kteří měli Alenu jako první, stejně tak to dopadlo i u žen. Podařilo se rovnoměrně vystřídat pořadí hlasů v celé skupině třiceti testovaných, nicméně zvláště ve skupině mužů a zvláště ve skupině žen se muselo sledovat více faktorů najednou a nakonec se tyto nesrovnalosti projevíly při vyhodnocování výsledků.

K eliminaci technických problémů při experimentech se nabízí řešení v podobě zajištění prostoru, kde by bylo možné realizovat rozhovory. Tím by se zajistilo srovnatelné prostředí pro všechny účastníky, nicméně by to mohlo odradit ty z nich, kterým možnost konání v domácím prostředí vyhovovala.

V oblasti dotazníků se nabízí hned několik možných zlepšení. Jedním z nich by bylo snížit počet odpovědí, které mohli účastníci volit, aby se tak usnadnilo kvantitativní vyhodnocování dat. Místo otevřených odpovědí by bylo lepší umístit do dotazníků pouze možnost výběru z konkrétních druhů odpovědi, eventuálně omezit otevřené odpovědi množstvím povolených znaků. Dále by bylo třeba zúčastněné „donutit“ k volbě jen jedné možnosti, tudíž eliminovat nerozhodné odpovědi typu „oba“, „ani jeden“ apod.

V neposlední řadě by bylo ideální účastníky nějakým způsobem motivovat k vyšší bdělosti a pozornosti při vyplňování dotazníků. Nejlepší možností by byla finanční kompenzace za svědomitou účast v experimentech. Další z možností by bylo projít znovu s každým účastníkem vyplněný dotazník, aby respondent zkontroloval správnost uvedených odpovědí. K tomu by ovšem opět bylo zapotřebí změnit místo konání z domácího prostředí a souviselo by s tím i zvýšení časové náročnosti každého experimentu.

Samotné vyhodnocení pokusu nabídlo některé významné závěry, které by bylo zajímavé dále rozpracovávat. V průběhu experimentu by samozřejmě bylo možné mnoho věcí

vylepšit, proto zde byly uvedeny problémy, se kterými jsme se museli potýkat a návrhy na jejich řešení či eliminaci. Zároveň se experiment může stát základním modelem pro budoucí podrobnější výzkum.

4.4.3 Příčiny nepotvrzení hypotézy

K potvrzení hypotézy „uncanny valley“ ve vztahu k syntetické řeči bylo zapotřebí potvrdit, že lidé budou hlas, který je podobnější lidskému znění, vnímat spíše negativně, což se v závěrech nepotvrdilo. V porovnání například s projektem Senior Companion měl experiment zřetelné výhody v podobě provedení pokusu v domácím prostředí, takže se mohla vyloučit příčina nervozity či nejistoty subjektu z neznámého místa či z přítomnosti cizích osob. Dalším rozdílem od zmíněného projektu může být i to, že důchodci v projektu Senior Companion komunikovali s avatarkou, kterou viděli na obrazovce, oproti tomu účastníci tohoto Experimentu měli na obrazovce pouze fotky jako předmět diskuze, a mohlo tak dojít k většímu odcizení od hlasu a zároveň si nemuseli připouštět možnou neexistenci původce hlasu. Pokud bychom však posuzovali výsledky z projektu Senior Companion, hypotéza „uncanny valley“ by na něj byla neuplatnitelná, jelikož projekt vůbec nebyl zaměřen na posuzování Moriho hypotézy. Senioři komunikovali s avatarem, který disponoval řečovou syntézou poměrně nízké kvality, a ze závěrů vyplývá, že v těchto kontextech robotický hlas účastníkům vyhovuje spíše, než kdyby byl hlas podobný lidskému.⁸⁰

Mohlo být důvodem nepotvrzení hypotézy „uncanny valley“ to, že jsme v interakcích „dokonalý“ systém záměrně zpomalovali, abychom mu dodali punc nutné strojovosti, ale zároveň jsme mu schopností aktivně a plynule reagovat na komunikačního partnera odebrali ty strojové reakce? Z toho by vyplývalo, že jsme se snažili uměle napasovat pokus na hypotézu z roku 1970, která může být již dávno překonána nejen v technologiích, ale i v myšlení lidí. Podle mého názoru existuje vysoká pravděpodobnost, že současná mladá generace je již připravená na takový technologický pokrok a dá se říci, že už očekává „víc“. Přestože zkoumané subjekty byly zpočátku zaskočeny schopností systému správně reagovat, poměrně rychle si na jeho možnosti zvykly a v té chvíli zjistily, že jim systém zkrátka nestačí. Na syntézu řeči jsme zvyklí již z nejrůznějších nejmodernějších technologií, od mobilních telefonů začínající, po chladničky a pračky konče. K tomu, abychom pocíťovali jistý zvláštní pocit, který by se mohl blížit tomu, co popisoval Mori, bychom potřebovali rozvinutější, kvalitnější a rychlejší systém. Dalším důvodem nepotvrzení teorie může být také to, že bez té

⁸⁰ HEIDEROVÁ, P. *Perspektivy řečové komunikace mezi člověkem a strojem*, s. 59.

konkrétní vizuální podoby byl hlas přenášený skrz počítač vnímán dostatečně izolovaně a bez spojitosti s konkrétní entitou, ze které by hlas pocházel, v účastníkovi nevyvolával tak silnou reakci. Jiná možná příčina existuje ve hrozbě toho, že průběh pokusu nebyl uchopen správně, úloha wizarda nebyla představována tak, jak by měla být, a v dialozích byla patrná přílišná „lidskost“ či jiné charakteristiky, kterých by dialogový systém v dohledné době nemohl být schopen. Tou nejobávanější příčinou nepotvrzení hypotézy je to, že účastníci nepochopili dotazy a reagovali spíše na líbivost hlasu než na hypotetickou představu budoucí interakce s dialogovým systémem tohoto typu. S tím souvisí i důležitý fakt, že několik účastníků uvedlo s jistým časovým odstupem od pokusu, že hodnotili konverzaci s jednotlivými řečovými syntézami jako celek – tedy spíše hodnotili, jak si v danou chvíli se systémem popovídali, jak se jim líbila témata a reakce systému, popřípadě hodnotili i samotnou propracovanost systému apod.

5 Závěr

V diplomové práci byly nastíněny souhrnné informace o současném stavu systémů převodu textu na řeč a rozpoznávání řeči. Tyto informace byly základem pro uspořádání vlastního experimentu, který měl ověřit hypotézu „uncanny valley“ ve vztahu k syntetické řeči. Cílem jeho provedení bylo zjistit, zda kvalita syntetického hlasu souvisí s příznivějším přijímáním uměle vytvořeného hlasu posluchači.

Uměle vytvořená řeč čerpá a vždy bude čerpat inspiraci z té lidské. Řeč jako komunikační nástroj osvojený přirozeným způsobem pouze člověkem bude vždy konkurentem pro syntetickou řeč. Paradoxně proto mají lidé na hodnocení syntetické řeči vyšší nároky a jsou méně tolerantní k nepřirozenostem či chybám v uměle vytvořené řeči.⁸¹ Toto tvrzení je uplatnitelné i na tuto práci. Z dostupných výsledků však nelze tuto tezi potvrdit stoprocentně. Můžeme však zhodnotit, že po provedení experimentů se rýsuje vize, že ten hlas, který je méně přirozený pro lidské ucho (tedy více nepodobný lidskému hlasu), je právě kvůli vyšším nárokům lidí vnímán negativně, a naopak ten, který se přibližuje lidskému hlasu, je vnímán jako přirozenější.

Otázkou však zůstává, jak dalece by tento trend pokračoval. Je možné, že jakmile by tento vzestup přirozenosti hlasu dosáhl určité hranice, vnímání „přespříliš“ přirozeného hlasu by se náhle „propadlo“ do „uncanny valley“. Další možností je, že v interakci s konkrétní materiální entitou (např. s přístrojem na nákup jízdenek, který by mluvil lidsky znějícím syntetickým hlasem) by se propad do „uncanny valley“ stal náhle markantnější a příkřejší, než jen v pouhé komunikaci na základě telefonního rozhraní, kdy pracujeme více se svojí představivostí a nedoléhá na nás fakt neexistence komunikačního protějšku tak naléhavě.

Nejdůležitějším závěrem experimentu tedy zůstává fakt, že z dostupných výsledků nelze hypotézu „uncanny valley“ ve vztahu k syntetické řeči přesvědčivě ani vyvrátit, ani potvrdit. K tomu, abychom disponovali obecně uplatnitelnými závěry, by bylo třeba zapojit do výzkumu větší počet účastníků, než který jsme měli k dispozici. Počet třiceti zúčastněných se zpočátku jevil jako dostačující, avšak v průběhu vyhodnocování výsledků, například při rozlišování účastníků na muže a ženy ve snaze odhalit signifikantní rozdíly v hodnocení, se prokázalo, že k získání relevantních hodnot by počet zúčastněných musel být v ideálním případě minimálně dvakrát vyšší. Proto by bylo záhodno vytvořit obdobný experiment s eliminací problémů, které doprovázely tento výzkum, jelikož pak by výsledky mohly být jednoznačné.

⁸¹ PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V., *Mluvíme s počítačem česky*, s. 645.

6 Přílohy

V příloze jsou k dispozici všechny tabulky, které byly využity ke statistické analýze výsledků experimentu.

Otázka: **Který z hlasů byl pro vás méně příjemný v komunikaci?**

nulová hypotéza (H₀): není signifikantní rozdíl mezi znakem "první Iva" a znakem "první Alena"

skutečné četnosti

	horší Iva	horší Alena	celkem
první Iva	4	11	15
první Alena	3	11	14
Celkem	7	22	29

očekávané četnosti (kdyby na sobě znaky (tj. pořadí přehráví) byly nezávislé)

	horší Iva	horší Alena	
první Iva	3,6207	11,3793	15
první Alena	3,3793	10,6207	14
Celkem	7	22	29

p-hodnota chi square testu

0,741854468

kdyby p-hodnota byla menší než 0,05, pak by bylo nutné H₀ zamítnout, tj. na pořadí přehrávání by záleželo
když je p-hodnota větší než 0,05, nelze H₀ zamítnout, tj. v takovém případě JE MOŽNÉ, že na pořadí nezáleží

Otázka: **Který z hlasů byl pro vás méně příjemný v komunikaci?** – zvlášť pro muže a zvlášť pro ženy

skutečné četnosti

Muži

	horší Iva	horší Alena	celkem
první Iva	2	7	9
první Alena	1	4	5
Celkem	3	11	14

očekávané četnosti

Muži

	horší Iva	horší Alena	Celkem
první Iva	1,928571429	7,071428571	9
první Alena	1,071428571	3,928571429	5
Celkem	3	11	14

nulová hypotéza (H₀): u mužů preference hlasu nezávisí na pořadí přehrání

p-hodnota chí testu: 0,922650255

H₀ je s vysokou pravděpodobností platná

skutečné četnosti

Ženy

	horší Iva	horší Alena	Celkem
první Iva	2	4	6
první Alena	2	7	9
Celkem	4	11	15

očekávané četnosti

Ženy

	horší Iva	horší Alena	Celkem
první Iva	1,6	4,4	6
první Alena	2,4	6,6	9
Celkem	4	11	15

nulová hypotéza (H₀): u žen preference hlasu nezávisí na pořadí přehrání

p-hodnota chí testu: 0,633553476

p-hodnota je výrazně vyšší než 0,05, H₀ je s vysokou pravděpodobností platná, tedy pořadí přehrání nebylo signifikantní vzhledem k preferencím u ženského pohlaví

Ani muži, ani ženy nebyli ovlivněni pořadím přehrání hlasu – pořadí hlasu bylo nezávislé jak vzhledem k preferencím, tak vzhledem k pohlaví.

Otázka: **Který hlas byste si dokázali představit jako partnera v komunikaci?**

H0: Preference hlasu nezávisí na pohlaví.

skutečné četnosti

	muži	ženy	celkem
Alena	1	0	1
Iva	8	6	14
Obě	5	7	12
Celkem	14	13	27

očekávané četnost

	muži	ženy	celkem
Alena	0,518518519	0,481481481	1
Iva	7,259259259	6,740740741	14
Obě	6,222222222	5,777777778	12
Celkem	14	13	27

p-hodnota chí testu: 0,905792863

p-hodnota je vyšší než 0,05, H0 je s vysokou pravděpodobností platná, preference hlasu nezávisí na pohlaví

Otázka: **Obeznamenost účastníků s řečovými technologiemi**

H0: Preference syntetického hlasu nezávisí na obeznamenosti účastníka s řečovými technologiemi.

skutečné četnosti

	preference Alena	preference Iva	celkem
naprostý laik	1	5	6
povrchní znalost	1	4	5
běžná znalost	3	12	15
pokročilá znalost	2	1	3
Celkem	7	22	29

očekávané četnosti

	preferance Alena	preferance Iva	celkem
naprostý laik	1,448275862	4,551724138	6
povrchní znalost	1,206896552	3,793103448	5
běžná znalost	3,620689655	11,37931034	15
pokročilá znalost	0,724137931	2,275862069	3
Celkem	7	22	29

p-hodnota chí testu: 0,34305987

p-hodnota je vyšší než 0,05, H₀ je s vysokou pravděpodobností platná, tedy preference hlasu nejsou ovlivněny mírou obeznámenosti účastníka s řečovými technologiemi

Otázka: **Vliv technického či humanitního vzdělání účastníků na zaujetí rozhovorem**

H₀: Míra zaujetí rozhovorem nezávisí na typu vzdělání.

skutečné četnosti

	technické vzdělání	humanitní vzdělání	celkem
zaujetí velmi	9	6	15
zaujetí celkem ano	4	10	14
celkem	13	16	29

očekávané četnosti

	technické vzdělání	humanitní vzdělání	celkem
zaujetí velmi	6,724137931	8,275862069	15
zaujetí celkem ano	6,275862069	7,724137931	14
celkem	13	16	29

p-hodnota chí testu: 0,08901855

p-hodnota je vyšší než 0,05, H₀ je s vysokou pravděpodobností platná, tedy míra zaujetí rozhovorem nebyla udávána oborem vzdělání

hypotéza není jasná – hypotézu, že zaujatost nezávisí na vzdělání nemůžeme ani odmítnout, ani přijmout

H0: Preference hlasu je nezávislá na oboru studia.

skutečné četnosti

	technický obor	humanitní obor	celkem
horší Alena	12	10	22
horší Iva	0	7	7
celkem	12	17	29

očekávané četnosti

	technický obor	humanitní obor	celkem
horší Alena	9,103448276	12,89655172	22
horší Iva	2,896551724	4,103448276	7
Celkem	12	17	29

p-hodnota chí testu: 0,010706648

p-hodnota je nižší než 0,05, H0 je tedy nutné zamítnout – tím pádem **obor studia nějakým způsobem ovlivňuje preference hlasu**

potvrzení podle tabulky, která kvantitativně vystihuje nulovou hypotézu – 92 % technického oboru preferovalo IVU, 59 % humanitního oboru preferovalo IVU – tato distinkce je podle chí testu pravdivá

	technický obor	humanitní obor	celkem
horší Alena	100%	59%	1,588235294
horší Iva	0%	41%	0,411764706
celkem	1	1	

VZDĚLÁNÍ OVLIVŇUJE PREFERENCE

Zde je k nahlédnutí tabulka zaznamenaných hodnot rozhovorů. Barevně je označena délka trvání rozhovorů, která se odlišovala od průměrné délky. Zeleně jsou označeny dlouhé rozhovory, oranžově jsou označeny krátké rozhovory. Barevně nevyznačené rozhovory měly průměrnou délku trvání.

	Obor	Pořadí	Délka roz.	Délka roz. I	Dél. vstupu I	Vstupy I	Délka roz. A	Dél. Vst. A	Vstupy A
AK	H	1. A, 2. I	21:54	15:40	129	44	05:40	50	18
ZL	H	1. I, 2. A	14:53	06:00	81	31	05:30	57	24
PH	H	1. A, 2. I	17:50	07:00	81	32	10:00	121	43
PB	T	1. I, 2. A	17:09	06:30	74	22	10:00	102	35
RF	H	1. A, 2. I	19:46	08:00	89	25	10:30	125	39
JK	H	1. I, 2. A	17:45	10:10	113	41	06:35	65	23
JK	T	1. A, 2. I	17:09	06:15	71	26	09:40	116	41
IH	H	1. I, 2. A	19:25	10:30	115	41	07:15	72	26
PV	T	1. A, 2. I	17:40	06:00	69	25	10:20	126	44
DH	T	1. I, 2. A	16:09	07:30	87	32	08:10	99	32
AR	H	1. A, 2. I	10:46	04:10	65	24	06:00	98	34
VR	T	1. I, 2. A	13:02	07:40	106	34	04:20	51	17
LJ	H	1. A, 2. I	14:41	05:45	64	20	08:00	91	35
SS	H	1. A, 2. I	18:50	10:00	88	32	08:00	106	34
JŠ	H	1. I, 2. A	14:57	07:47	118	37	06:30	85	29
AI	H	1. A, 2. I	14:04	04:25	58	21	08:40	105	35
MŠ	H	1. I, 2. A	12:47	08:15	96	32	04:00	50	18
BM	T	1. A, 2. I	15:17	08:00	94	34	06:15	84	33
KK	H	1. I, 2. A	16:54	09:35	103	35	06:25	68	23
DP	T	1. I, 2. A	16:24	09:20	91	27	06:15	61	18
KČ	T	1. I, 2. A	15:38	08:30	101	33	06:20	66	19
JH	T	1. A, 2. I	25:56	13:25	151	40	12:00	121	37
JŠ	T	1. I, 2. A	13:39	07:20	92	30	05:15	67	20
PK	H	1. A, 2. I	12:09	04:30	59	21	06:15	101	36
JK	T	1. I, 2. A	16:33	10:00	96	31	06:00	62	18
VF	H	1. I, 2. A	09:35	05:35	80	27	03:20	57	19
MT	T	1. I, 2. A	11:19	06:00	88	29	05:00	70	25
BČ	T	1. I, 2. A	23:01	11:00	126	35	09:40	109	27
MŠ X	H								
SŠ X	H								
PRŮMĚR			16:15	08:02	92,32142857	30,75	07:13	85,178571	28,64286

průměrná délka rozhovoru	16:15
směrodatná odchylka	03:37
krátký rozhovor	12:38
střední rozhovor	16:15
dlouhý rozhovor	19:52

průměrná délka rozhovoru s Ivou		08:02
směrodatná odchylka		02:39
Krátký		05:22
Střední		08:02
Dlouhý		10:41

průměrná délka rozhovoru s Alenou		07:13
směrodatná odchylka		02:12
Krátký		05:00
Střední		07:13
Dlouhý		09:25

Délka celého rozhovoru vs. Preference hlasu

H0: Preference hlasu nezávisí na délce rozhovoru

skutečné četnosti

	horší Iva	horší Alena	Oba	celkem
krátký rozhovor	0	4	0	4
střední rozhovor	6	14	1	21
dlouhý rozhovor	0	3	0	3
Celkem	6	21	1	28

očekávané četnosti

	horší Iva	horší Alena	Oba	celkem
krátký rozhovor	0,857142857	3	0,142857143	4
střední rozhovor	4,5	15,75	0,75	21
dlouhý rozhovor	0,642857143	2,25	0,107142857	3
Celkem	6	21	1	28

p-hodnota 0,539406458

p-hodnota je vyšší než 0,05, takže preference hlasu nezávisí na délce rozhovoru

Délka rozhovoru s Ivou vs. Preference

H₀: Délka rozhovoru s Ivou je nezávislá na preferencích

skutečné četnost

	horší Iva	horší Alena	celkem
krátký	1	2	3
Střední	5	16	21
Dlouhý	0	3	3
Celkem	6	21	27

očekávané četnosti

	horší Iva	horší Alena	celkem
krátký	0,666666667	2,333333333	3
Střední	4,666666667	16,333333333	21
Dlouhý	0,666666667	2,333333333	3
Celkem	6	21	27

p-hodnota 0,576361394

p-hodnota je vyšší než 0,05, takže preference hlasu nezávisí na délce rozhovoru s Ivou

Délka rozhovoru s Alenou vs. Preference

H₀: Délka rozhovoru s Alenou je nezávislá na preferencích

skutečné četnost

	horší Iva	horší Alena	celkem
krátký	1	3	4
Střední	5	11	16
Dlouhý	0	7	7
Celkem	6	21	27

očekávané četnosti

	horší Iva	horší Alena	celkem
krátký	0,888888889	3,111111111	4
Střední	3,555555556	12,444444444	16
Dlouhý	1,555555556	5,444444444	7
Celkem	6	21	27

p-hodnota 0,250033414

p-hodnota je vyšší než 0,05, takže preference hlasu nezávisí na délce rozhovoru s Alenou

H0: Délka rozhovoru je nezávislá na humanitním či technickém zaměření

skutečné četnosti

	humanitní zaměření	technické zaměření	celkem
krátký rozhovor	3	1	4
střední rozhovor	11	10	21
dlouhý rozhovor	1	2	3
celkem	15	13	28

očekávané četnosti

	humanitní zaměření	technické zaměření	celkem
krátký rozhovor	2,142857143	1,857142857	4
střední rozhovor	11,25	9,75	21
dlouhý rozhovor	1,607142857	1,392857143	3
celkem	15	13	28

p-hodnota 0,536750334

p-hodnota je vyšší než 0,05, takže délka rozhovoru je nezávislá na zaměření účastníka

Počet vstupů hlasu vs. Preference

H0: Preference hlasu nezáležely na počtu vstupů, které měla Iva v rozhovoru

průměrný počet vstupů Iva	30,75
směrodatná odchylka	6,461423992

nízký počet vstupů	24,28857601
střední počet vstupů	30,75
vysoký počet vstupů	37,21142399

skutečné četnosti

	horší Iva	horší Alena	celkem
nízký počet vstupů	1	4	5
střední počet vstupů	4	13	17
vysoký počet vstupů	1	4	5
celkem	6	21	27

očekávané četnosti

	horší Iva	horší Alena	celkem
nízký počet vstupů	1,111111111	3,888888889	5
střední počet vstupů	3,777777778	13,22222222	17
vysoký počet vstupů	1,111111111	3,888888889	5
Celkem	6	21	27

p-hodnota 0,977566386

p-hodnota je vyšší než 0,05, hodně se blíží k hodnotě 1, takže je s velkou pravděpodobností platná – preference hlasu nezáleží na počtu vstupů, které měla Iva v rozhovoru

H0: Preference hlasu nezáležely na počtu vstupů, které měla Alena v rozhovoru

průměrný počet vstupů Alena	28,64285714
směrodatná odchylka	8,59032309

nízký počet vstupů	20,05253405
střední počet vstupů	28,64285714
vysoký počet vstupů	37,23318023

skutečné četnosti

	horší Iva	horší Alena	celkem
nízký počet vstupů	1	6	7
střední počet vstupů	5	10	15
vysoký počet vstupů	0	5	5
Celkem	6	21	27

očekávané četnosti

	horší Iva	horší Alena	celkem
nízký počet vstupů	1,555555556	5,444444444	7
střední počet vstupů	3,333333333	11,66666667	15
vysoký počet vstupů	1,111111111	3,888888889	5
Celkem	6	21	27

p-hodnota 0,252195419

p-hodnota je vyšší než 0,05, preference hlasu nezáleží na počtu vstupů, které měla Alena v rozhovoru

7 Seznam použité literatury a pramenů

Autor neuveden. *III. Cvičení ze statistiky* [online]. [21. 3. 2013] Dostupné z www: <<http://ulb.upol.cz/praktikum/statistika3.pdf>>.

Autor neuveden. *SpeechTech TTS – syntéza řeči* [online]. [30. 11. 2013] Dostupné z www: <<http://www.speechtech.cz/produkty/synteza-rci.html>>.

Autor neuveden. *The 50 Best Robots Ever* [online]. [4. 4. 2014] Dostupné z: <http://archive.wired.com/wired/archive/14.01/images/FF_120_bots34_f.jpg>.

European Commission. *COMPANIONS* [online]. [28. 1. 2014] Dostupné z www: <<http://www.companions-project.org/?p=home>>.

DISMAN, M. *Jak se vyrábí sociologická znalost: příručka pro uživatele*. 3. vyd. Praha : Karolinum, 2000. 374 s. ISBN 80-246-0139-7.

EKLUND, J., GUSTAFSON, J., HELDNER, M., HJALMARSSON, A. *Towards human-like spoken dialogue systems. Speech communication* [online]. [25. 3. 2014] s. 630 – 645. Dostupné z: <<http://linkinghub.elsevier.com/retrieve/pii/S016763930800054X>>.

FREUD, S. *Das Unheimliche* [online]. [1. 4. 2014] Dostupné z www: <<http://www.gutenberg.org/files/34222/34222-h/34222-h.htm>>.

HANSON, D. *David Hanson: Robots that "show emotion"* (přednáška na konferenci TED2009) [online]. [31. 3. 2014] Dostupné z: <http://www.ted.com/talks/david_hanson_robots_that_relate_to_you#t-171981>.

HEIDEROVÁ, Pavlína. *Perspektivy řečové komunikace mezi člověkem a strojem*. Plzeň, 2013. Diplomová práce (Mgr.). Západočeská univerzita v Plzni, Fakulta filozofická.

HENDL, J. *Kvalitativní výzkum: základní teorie, metody a aplikace*. 2., aktualiz. vyd. Praha : Portál, 2008. 407 s. ISBN 978-80-7367-485-4.

MORI, M. *The Uncanny Valley* [online]. [30. 3. 2014] S. 98 – 100. Dostupné z:
<<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6213238>>.

PIERACCINI, R. *The Voice in the Machine: Building Computers That Understand Speech*. Cambridge : MIT Press, 2012. 325 s. ISBN 978-0-262-01685-8.

PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V. *Mluvíme s počítačem česky*. Vyd. 1. Praha : Academia, 2006. 746 s. ISBN 80-200-1309-1.

ROMPORTL, J. *Zvyšování přirozenosti strojově vytvářené řeči v oblasti suprasegmentálních zvukových jevů*. Plzeň, 2008. Disertační práce. Západočeská univerzita, Fakulta aplikovaných věd.

SZERSZEN, K. A. *The Audio/Visual Mismatch and the Uncanny Valley: An Investigation Using a Mismatch in the Human Realism of Facial and Vocal Aspects of Stimuli*. Indiana, 2010. Diplomová práce. Indiana University, Faculty of the School of Informatics.

ŠVARNÝ, P. *A Visit on the Uncanny Hill*. In: *Beyond AI: Artificial Dreams*. Plzeň : University of West Bohemia. s. 35 – 40. ISBN 978-80-261-0102-4.

ŠVEC, Jan. *Nádraží – hlasový dialogový systém* [online]. [22. 10. 2013] Dostupné z www:
<http://www.youtube.com/watch?v=VZ8I5M_mu2M>.

8 **Resumé**

The aim of this paper is to evaluate if there is a relation between the Uncanny Valley hypothesis and the speech which is produced artificially. For this reason there is a mention of the brief introduction to the actual state of speech technologies. Specifically technologies such as Text-To-Speech synthesis and Automatic Speech Recognition are presented, which are described in the context of the use in dialogue systems.

Another concept which is discussed in this paper is the Uncanny Valley hypothesis which describes the relation of the human to the inanimate object; in the first place to robots. The hypothesis is considered primarily in the relation to the synthetic speech. Experimental research was organized, which focused mainly to the accepting and evaluation of the synthetic speech.