Západočeská univerzita v Plzni

Fakulta filozofická

Diplomová práce

Perspektivy řečové komunikace mezi člověkem a strojem

Pavlína Heiderová

Plzeň 2013

Západočeská univerzita v Plzni

Fakulta filozofická

Katedra filozofie

Studijní program Humanitní studia

Studijní obor Teorie a filozofie komunikace

Diplomová práce

Perspektivy řečové komunikace mezi člověkem a strojem

Pavlína Heiderová

Vedoucí práce:

Ing. Jan Romportl, Ph.D.

Katedra kybernetiky

Fakulta aplikovaných věd Západočeské univerzity v Plzni

Plzeň 2013

Prohlašuji, že jsem práci zpracovala samostatně a použila jen uvedených pramenů a literatury.

Plzeň, 24.4.2013 ……

Obsah

1 Úvod

Lidská řeč je jedinečnost, kterou jsme byli obdařeni, nikdo jiný než člověk není v současné době schopen disponovat takovým množstvím znaků a sdělovat jejich prostřednictvím své myšlenky, komunikovat se svým okolím a zároveň na něj působit. Mezilidskou interakci považujeme za něco přirozeného, co nás provází celým životem, a velmi často si neuvědomujeme, o jak složitý celek se jedná. Naše řeč je něco výjimečného, co je nám vlastní, ale zároveň i něco, co v sobě skrývá doposud neodhalená tajemství.

Člověk se od ostatních tvorů neliší pouze tímto sofistikovaným prostředkem komunikace, ale i vytvářením a používáním nástrojů, které mu slouží při jeho každodenních činnostech. A stejně jako se společně s vývojem lidstva vyvíjí jazyk, vyvíjí se i technologie, která se stala přirozenou součástí našich životů. Není tedy divu, že se snažíme propojit naše dvě specifika a vytvořit artefakt, se kterým bychom mohli komunikovat prostřednictvím řeči, a usnadnit tak jeho užívání.

Necháváme se unášet představivostí a sníme o tom, jak bude vypadat naše budoucnost. Do soudobých technologií vkládáme obrovské naděje a představujeme si, jaké by bylo, kdyby mohly dokázat ještě více než doposud. Vývoj dialogových systémů je jednou z oblastí, která v nás tyto představy podněcuje.

„Má teze je, že během pár desetiletí bude překonána hranice mezi fantazií a skutečností, řečové technologie budou na takové úrovni, že s nimi budeme moci komunikovat jako s člověkem.“ Prognóza, která byla vyřčena z úst Rodneyho Brookse (2002), avšak nejen u něj se můžeme setkat s takovými myšlenkami. Nebudu ji ani zpochybňovat, ani potvrzovat a ani soudit i přesto, že z názvu mé práce by čtenář mohl něco takového očekávat. Jde mi o vytvoření představy o tom, jak je jazyk komplexním jevem a jak je vložení této schopnosti do počítače složitým úkolem.

První část mé práce je věnována jednotlivým složkám, které by měl dialogový systém zvládnout. Technikovi by se mohl zdát můj popis příliš zjednodušující, tato kapitola je však psána „laikem pro laiky“, nejsem v tomto oboru dostatečně vzdělána, a tím pádem ani oprávněna plně popsat technické parametry, které jsou však

reflektovány nejen v cizojazyčné, ale i české literatuře. Tato část je spíše kompilačního charakteru, považuji ji za jakousi předehru k tomu, čím se budu zabývat dále. Tomu, že vytvoření dobře fungujícího dialogového systému je nejen dlouhodobým procesem, ale i jeho realizace v sobě může zahrnovat mnohem více záhad, než by se mohlo na první pohled zdát. Jak druhá, tak třetí část mé práce se věnuje tendenci antropomorfizace neživých věcí a v současnosti velmi rozšířenému předpokladu, že čím více je systém nerozeznatelný od člověka, tím je lepší. Ve třetí části mé práce budu pomocí empirických dat ověřovat, zda je tento předpoklad pravdivý.

Mám v úmyslu čtenáři zprostředkovat materiály k tomu, aby si mohl udělat svůj vlastní obraz o tom, jak složité je vytvořit alespoň dobře fungující dialogový systém a kolik zákoutí se vlastně skrývá už jen v názvu mé diplomové práce.

2 Řečová komunikace mezi člověkem a strojem

2.1 Lidská řeč

V roce 1968 natočil Stanley Kubrick vědeckofantastický film *2001: Vesmírná odysea*, který se stal inspirací pro mnoho vědců z různých oborů. Palubní počítač HAL 9000 zde představuje svatý grál, kterého se badatelé v oblasti řečových technologií snaží dosáhnout již desetiletí. Srovnáme-li současné řečové technologie a palubní počítač HAL 9000 či centrální počítač vesmírné lodi v seriálu *Červený trpaslík*, musíme si položit otázku, zda je vůbec možné dosáhnout takového výsledku. Je počítač schopen vést dialog s člověkem na stejné, ne-li lepší úrovni jako člověk s člověkem? V řadě jiných odvětví bylo dosaženo velkých pokroků, ale oblast řečových technologií jakoby stagnuje. Počítač je schopný porazit člověka v šachách nebo vědomostní soutěži, ale v jedné z našich nejpřirozenějších aktivit jsme stále nedostižitelní.

Proč je ale tak složité naučit stroj něco, v čem jsme mistři? Je chyba v technologiích, anebo je schopnost řeči pouze pro člověka a není možné naučit ji stroj? Obsahuje lidský mozek něco nezbytného, co nelze replikovat do počítače? Lidská řeč je nejspíše nejsofistikovanější schopnost, kterou jsme obdařeni. Pátrání po odpovědi na otázku, co je to řeč, nekončí ve chvíli, kdy popíšeme fungování hlasového ústrojí a procesy v mozku, které s tím souvisí. Komunikaci tvoří i mimika, gestika, prozódie atd. Dále musí mít účastník rozhovoru také určité znalosti. Mnoho slov má totiž mnoho různých významů a partneři v dialogu si musí rozumět, o kterém z těch významů právě hovoří.

Pro člověka je jeho mateřský jazyk jako chůze, mluvíme, aniž bychom přemýšleli nad strukturou jazyka. Až tehdy, kdy se začneme učit druhý jazyk, si začneme uvědomovat, o jak složitý komplex jde. Musíme se soustředit na syntax,

výslovnost, gramatiku, naučit se slovní zásobu a správnou intonaci, a i když se tohle všechno naučíme, vždycky bude, alespoň ve většině případů, poznat, že jsme cizinci. Proč se naučíme jazyk v dětství tak snadno, ale v dospělosti je to mnohem složitější? Je tu něco, co je nám dáno pouze v raném období našeho života a později se to vytratí?

Mateřský jazyk se učíme již od dětství, ovládáme jej perfektně, ale nevíme, jak jsme se jej přesně naučili. Steven Pinker (2009, s. 303), experimentální psycholog, kognitivní vědec a lingvista, popisuje proces učení řeči u kojence takto: „Kojenec je jako člověk, který má k dispozici složité audio zařízení hustě osázené neoznačenými tlačítky s vypínači a chybí mu návod k jeho použití. V takových chvílích se lidé uchylují k tomu, co hackeři nazývají bezcílným mačkáním a otáčením knoflíků. Hrají si s ovládacími prvky a zkoušejí, co to udělá. Kojencům byl poskytnut soubor nervových povelů, které mohou pohybovat s mluvidly všemi možnými způsoby, což má následek velice odlišné účinky na zvuk. Poslechem vlastního žvatlání děti vlastně píší svůj vlastní návod k použití, učí se na kolik pohnout kterým svalem jakým směrem, aby vykonal změnu ve zvuku. Toto je bezpodmínečně nutný předpoklad napodobování řeči rodičů. Někteří počítačoví vědci, inspirováni výzkumem řeči kojenců, věří, že dobrý robot by se měl naučit interní softwarový model svých mluvidel tím, že by pozoroval následky vlastního žvatlání.“

Podle Pinkera (2009, s. 317) můžeme již tříleté dítě považovat za jazykového genia. Zvládá většinu konstrukcí, pravidly se spíše řídí, než neřídí, respektuje jazykové univerzálie, chybuje logickými způsoby jako dospělý a vyhýbá se mnoha druhům chyb vůbec.

Když se ale můžeme tak perfektně a jednoduše naučit řeč jako děti, proč to nedokáží šimpanzi nebo psi? Anebo stroje? Na rozdíl od zvířat a strojů je člověk fyziologicky vyvinutý pro to, aby byl schopný naučit se mluvit. Musíme mít vrozenou způsobilost, abychom se mohli naučit jazyk, Noam Chomský ji nazývá *jazykový orgán*, který není dostupný jiným žijícím tvorům na Zemi. Noam Chomský, profesor lingvistiky na MIT (Massachusetts Institute of Technology) do roku 1955, je pravděpodobně nejkontroverznější jazykový teoretik této doby. Podle něj se rodíme s jazykovým orgánem, který je aktivní pouze pár let po narození a poté zakrní. (Pieraccini 2012, s. 3)

Víme, že schopnost rozumět a mluvit musí zahrnovat minimálně poslech řeči dalších lidských bytostí. Po několik tisíc let myslitelé bádali nad tím, co by se stalo, kdyby dítě nepřišlo do styku s lidskou řečí. V sedmém století před naším letopočtem nechal, podle historika Hérodota, egyptský král Psamtek I. odloučit dva kojence ihned poté, co se narodili, od jejich matek, a nechal je pak vyrůst v tichu pastýřské boudy. Králova zvědavost, jak zní původní jazyk světa, byla údajně uspokojena o dva roky později, když uslyšel děti vyslovit slovo ve frýžštině, indoevropském jazyce Malé Asie. V dalších stoletích se objevilo mnoho příběhů o dětech vyrůstajících mimo civilizaci (tzv. vlčích dětech), které byly vychovávány zvířaty, ale ať jsou tyto příběhy pravdivé či ne, faktem je, že pokud děti nepřichází do styku s mluveným slovem, jsou němé. Také schopnost osvojit si jazyk je zaručena dětem do věku šesti let, od tohoto věku do puberty se postupně snižuje a poté je zřídka úspěšná. (Pinker 2009, s. 318)

Tyto fenomény podporují Chomského tezi o jazykovém orgánu, ale na druhou stranu jsou zde argumenty, které ji vyvrací. Děti se učí svůj mateřský jazyk za úplně jiných okolností než ten druhý. Podmínky, ve kterých se jej učíme, jsou neopakovatelné. Za prvé, v dětství nás nerozptyluje a neovlivňuje znalost jiného jazyka. Dále děti nemají jiné možnosti, jak vyjádřit to, co chtějí efektivnějším způsobem, a proto je jejich motivace mluvit obrovská. Malým dětem se dostává také takové pozornosti, která se v pozdějším věku neopakuje. Rodiče mají během jejich vývoje extrémní trpělivost, opakují slova znovu a znovu, opravují chyby, podporují získané dovednosti a učí zároveň nové. Jako dítě si můžeme jazyk osvojit svým vlastním tempem, a co je ještě důležitější, jsme vystaveni našemu mateřskému jazyku denně, je všude kolem nás, to nám dává možnost se v něm neustále zdokonalovat. A i když jsme obklopeni druhým jazykem, může si dospělý člověk získat takovou pozornost, dostane se mu takové trpělivosti a konstruktivní kritiky rodilých mluvčích, jako je tomu u dítěte? (Pieraccini 2012, s. 4)

Dalšími pravděpodobnými příčinami lehkosti, s jakou si je schopno dítě osvojit jazyk, jsou změny při dozrávání mozku. V prvních dvou letech života roste mozek nejrychleji, od dvou do pěti let, již roste pomaleji, ale stále s velkou rychlostí a svůj růst zakončuje v období puberty. Podle neurolingvisty Erika Lenneberga rychlý raný vývoj jazykových schopností dítěte kopíruje křivku rychlého nárůstu váhy mozku. Podle něj tak dítě získá základní lingvistické schopnosti kolem čtvrtého až pátého roku

(v tomto období dosáhne mozek 90% celkové váhy). Schopnost osvojit si jazyk po pubertě, tedy v době, kdy dosáhne maximální možné velikosti, strmě klesá. (Russell a Wanda 2009, s. 290)

Jazyk je obrovský komplex a my nevíme, jak jej ve skutečnosti ovládáme. Rozumíme fyziologii orgánů tvořících a přijímajících řeč a známe i mozkovou aktivitu, ke které při těchto procesech dochází. Funkce mozku jsou nesmírně složité, a i přestože víme, že roli ve zpracování řeči hraje tzv. Brocovo centrum, nemáme odpovědi na to, co se přesně při vnímání a produkci řeči odehrává. Máme hypotézy, ale nedisponujeme dostatečnými empirickými daty k jejich ověření.

2.2 Rozpoznávání lidské řeči

Je možné, abychom bez těchto znalostí byli schopni sestrojit stroj, se kterým můžeme vést dialog jako s člověkem? Máme se v oblasti řečových technologií dát cestou napodobování lidských procesů, které se při tvorbě řeči odehrávají? V oblasti vývoje rozpoznávání a produkce řeči prostřednictvím stroje platí dnes všeobecně uznávaný názor, že nejde o šťastné řešení. Musíme mít na paměti, že počítač a lidský mozek fungují úplně na jiném principu. Stroj přece nemusí fungovat na stejných zásadách jako živé bytosti, aby mohl splnit naše požadavky. „I letadlo létá a nemusí mávat křídly jako pták“, velmi častý argument, používaný v oblasti řečových technologií. Dnešní počítače nemají uši, ústa ani hlasivky, a přesto dokážou produkovat a rozpoznat řeč. Dnešní rozpoznávače a syntetizéry řeči mají velký potenciál, mohou rozumět hlasům milionů lidí, pochopit smysl tisíců slov a pojmů, řídit se jednoduchými pokyny, poskytnout informace a řešit problémy stejně dobře jako člověk. Ale i přes jejich dlouholetý vývoj stále nesplňují naše očekávání.

2.2.1 Počátky

Představme si, že chceme s naším počítačem vést dialog, budeme-li se soustředit pouze na řečovou komunikaci, jaké úkoly by měl být schopen vykonat? Prvním krokem je automatické rozpoznávání řeči (angl. Automatic speech recognition, zkr. ASR). Vize, že bude stroj schopen rozpoznat slova, která jsme vyslovili, není nic

nového, avšak důkaz o tom, že nejde o pouhý neuskutečnitelný sen, byl veřejnosti předveden až v roce 1952. Jmenoval se AUDREY (Automatic Digit Recognition) a i přesto, že byl schopen rozpoznávat pouze čísla a neodlišoval začátek a konec slov, byl počátečním úspěchem, který odstartoval vývoj v této oblasti. (Pieraccini 2012, s. 59)

Po AUDREY se objevilo mnoho strategií, které se snažily tuto problematiku vyřešit. Například použití fonetické segmentace, která měla identifikovat individuální fonetické elementy v promluvě, a vyvinout tak rozpoznávač schopný z pronesených zvuků rozeznat skupiny ve slovech, slova ve frázích a slova ve větách a krok po kroku zrekonstruovat řetězec lidské řeči. Ale i navzdory veškerým snahám, znalostem a péči vědců, dosáhl rozpoznávač fungující na tomto principu pouze neuspokojivých výsledků. Chyby v oblasti fonetiky se nevyhnutelně projevily i v oblasti lexikální, a šlo tak pouze o malé zlepšení. (Pieraccini 2012, s. 61)

Rozpoznávání řeči potřebovalo silnější stroj a jeden takový byl již za rohem: Počítač. Ačkoliv první digitální počítač ENIAC byl postaven už během 2. světové války, digitální počítač se neobjevil na scéně výzkumu řeči dříve než v roce 1960. Také idea, že počítače nejsou pouze kalkulačky, ale mohou stejně tak manipulovat s nenumerickými údaji a vyvozovat inteligentní závěry, se do kolektivní do povědomí dostávala až v padesátých letech. V roce 1950 zveřejnil americký teoretik a matematik Claude Shannon článek, ve kterém popisuje, jak mohou být počítače nejen konstruovány k hraní šachů, ale také k tomu, aby byly schopny pracovat se symbolickými elementy, které reprezentují slova, vlastnosti a subjekty. V roce 1955 na konferenci v Dartmouthu byl poprvé zaveden pojem umělá inteligence a od roku 1970 bylo všeobecně přijímáno, že umělá inteligence může realizovat sen o sestavení inteligentního stroje během pár let. Mnoho vědců věřilo v sílu umělé inteligence. (Pieraccini 2012, s. 83-84)

Právě v této době se začínají vytvářet dva tábory, tábor lingvistický a inženýrský. Lingvisté se zaměřují převážně na porozumění jazyku, které pak chtějí vložit do počítače. Snaží se přijít na to, jak lidé rozumějí řeči a jak bychom mohli tuto schopnost předat počítači. Naopak přístup inženýrský se zaměřuje na konkrétní problém, jak sestavit stroj, který může být řízen řečí bez ohledu na to, jak funguje člověk. Snaží se tento problém vyřešit dostupnými prostředky. První takový přístup

po AUDREY byl program pracující na principu porovnávání se vzory (angl. template-matching). (Pieraccini 2012, s. 64-65)

„Tato metoda byla velmi aktuální v sedmdesátých a osmdesátých letech, kdy byla často aplikována zejména v klasifikátorech izolovaně vyslovených slov. Slovo je zde zpracováno jako celek, při čemž je klasifikováno do té třídy (třídy jsou tvořeny jednotlivými slovy ve slovníku), k jejímuž vzorovému obrazu (vzorovému slovu reprezentovanému posloupností příznakových vektorů) má nejmenší vzdálenost. Tato vzdálenost je obvykle určována na základě aplikace metody dynamického programování, při které se hledá taková nelineární transformace časové osy jednoho z obrazů, při níž dojde k porovnání obou obrazů s nejmenší výslednou vzdáleností. Uvedený mechanismus vyplynul z důkladného rozboru signálu získaného vyslovením stejného slova několikrát týmž řečníkem. Při tomto rozboru se zjistilo, že základní odlišnosti mezi odpovídajícími signály se nachází v nestejné délce slov, zejména v nepoměru mezi délkami odpovídajících částí (fonémů, hlásek) uvnitř slova.“ (Psutka et al. 2006, s. 196)

V oblasti řečových technologii se v této době objevila také idea vytvoření expertního systému, který by byl schopen rozumět řeči. Expertní systém by mohl procházet obrovské množství informací a působit tak jako lidský expert při řešení problémů ve specifických odvětvích. I tyto pokusy nebyly bezvýsledné, ale v tomto případě bylo nutné popsat jednotlivý krok po kroku a s naprostou přesností charakterizovat pravidla řeči. Vědci z oboru umělé inteligence viděli tento způsob jako velmi nadějný, který by fungoval na principu napodobení lidských kognitivních schopností nebo alespoň toho, jak jejich fungování chápeme. Ale i navzdory popularitě a vědeckému zájmu, nedosahovaly expertní systémy očekávané úspěšnosti. Sestavení prototypů, a to i s omezenými možnostmi, bylo složité a vyžadovalo spolupráci několika odborníků z různých oblastí: počítačové vědy, fonetiky, lingvistiky a inženýrství. Odpověď na otázku, jak dobře pracovali, zůstala nejasná. Většina výzkumníků umělé inteligence byla více zaujata novostí, elegancí, vědeckou atraktivitou jejich přístupu a také tím, jak úzce se jejich modely podobaly hodnověrným modelům lidského poznání než samotným výkonem svých systémů. (Pieraccini 2012, s. 88)

2.2.2 Statistická metoda

Velmi významný pokrok se odehrál v 80. letech 20. století. V této době začala vznikat jedna z nejpoužívanějších a nejrozšířenějších metod současnosti, statistická metoda, jejímž autorem je Frederick Jelinek. Jelinek byl původem český vědec, který se v roce 1932 narodil v Kladně a v roce 1949 emigroval s matkou do Spojených států. Díky stipendiu pro nadané emigranty byl přijat na studia elektrotechniky na věhlasný Massachusetts Institute of Technology (MIT). Po absolvování doktorského studia se stal profesorem na Cornellově univerzitě a v roce 1972 přijal nabídku počítačové firmy IBM stát se vedoucím právě založeného týmu pro počítačové rozpoznávání a zpracování řeči. U IBM zůstal až do roku 1993. (Ševela 2003)

Jelinek věřil, že problematiku rozpoznávání řeči může vyřešit pomocí použití matematiky, statistiky a komunikační teorie. IBM investovala mnoho peněz do umělé inteligence a věřila, že je to nejlepší řešení. Jelinkova skupina začala v roce 1970 pracovat na uměle vytvořeném jazyce zvaném *Raleigh*. IBM trvala na tom, aby tým pracoval na expertním systému, který by obsahoval kompletní pravidla fonetiky a by byl založen na lingvistických znalostech. Tým skutečně sestavil takovýto rozpoznávač řeči, jehož přesnost se pohybovala kolem 35%. Poté ředitel lingvistického oddělení odešel z IBM a Jelínkův tým se postavil k problému z nového a odlišného hlediska. Při použití statistického modelu fonémů derivovaných automaticky z dat bez lingvistické expertízy se zvýšila přesnost na 75%. (Pieraccini 2012, s. 109-110)

Statistický přístup, kterým se zabývají vědci z IBM, představuje rozpoznávání řeči jako teoretický komunikační problém, známý jako model zašuměného kanálu (angl. noisy channel model). Řeč začíná posloupností symbolické reprezentace slov v mysli řečníka. Symbolická reprezentace je produkována prostřednictvím vokálního aparátu mluvčího, který jej převede do formy řečových zvuků, které jsou přeneseny jako akustický signál k uším posluchače nebo k mikrofonu počítače. Tato čistá posloupnost slov z mysli mluvčího je „zašuměná“ tím, že je převedena do akustické podoby. Z pohledu rozpoznávače řeči vše, o co má zájem, jsou jen slova, která stála na počátku, všechno ostatní jsou šumy. To je předpoklad stojící za modelem zašuměného kanálu, skrytý mechanismus, který vezme čistou symbolickou reprezentaci slov a transformuje ji do akustického signálu. To, čeho si počítač všímá, je akustické

pozorování, výstupem modelu zašuměného kanálu je posloupnost akustických příznaků. Řečový dekodér určí jako výsledek posloupnost slov, která je pro daný akustický signál nejvíce pravděpodobná. (Pieraccini 2012, s. 109-110)

Sekvence slov je skrytý signál, který rozpoznávač řeči potřebuje najít. Ne všechny sekvence užití jsou u všech slov stejné, některé jsou více pravděpodobné než ty ostatní. Existuje například větší pravděpodobnost, že se setkáme se spojením slov „lev je divoké zvíře“ než třeba „lev dálnice jedna“. Také budeme pravděpodobně používat v každodenní komunikaci více větu „jak se máš“ než „potřebuji letět do San Franciska“. Vědci, zabývající se rozpoznáváním řeči, používají termín *statistický jazykový model*, který má ukázat rámec všech povolených vět v rozsahu několika po sobě následujících slov, s omezenou slovní zásobou a z hlediska pravděpodobnosti. Tradiční lingvistika rozlišuje mezi větami, které se drží gramatických pravidel a které ne. Věty, které se gramatických pravidel nedrží, by neměly být součástí jazyka. Na druhé straně komputační lingvistika nevidí velký rozdíl mezi větami gramatickými a přípustnými. V této oblasti jsou v jazyce přípustné všechny věty, jen některé jsou přípustné více než ty ostatní. Statistický jazykový model je matematický nástroj, který určí pro danou sekvenci slov její pravděpodobnost užití ve specifickém jazyce. (Pieraccini 2012, s. 115-116)

Samozřejmě, že pravděpodobnost sekvencí slov v daném jazyce závisí na mnoha faktorech. Jedním z těchto faktorů je kontext, ve kterém jsou ta slova použita. Například v zoologii se s větší pravděpodobností vyskytují názvy zvířat než slova jako nákup, hypotéka a vlastnictví. Pokud tedy hovoříme o jazykové pravděpodobnosti, musíme ji chápat z hlediska jejího konkrétního kontextu neboli *jazykové domény*. Vytvoření statistického jazykového modelu není jednoduché, a to i v případě, že jazykovou doménu velmi dobře specifikujeme. Vezmeme-li například slovní zásobu o 1000 slov s větami o maximálně 10 slovech, dostaneme 1000^{10} možných kombinací. To je neuvěřitelně obrovské číslo a hledání všech možných kombinací v nekonečném počtu možností je velmi složité, možná i nemožné. Abychom se tedy mohli dostat k pravděpodobnosti každé sekvence, musíme ji vyvodit ještě z nějakého jiného případu, který je konečný a počitatelný. (Pieraccini 2012, s. 116)

Slova v promluvě nejdou chaoticky za sebou, pokud tedy člověk nemluví nesmysly. Slova mají mezi sebou definovaný vztah. Když slyšíme prvních pár slov

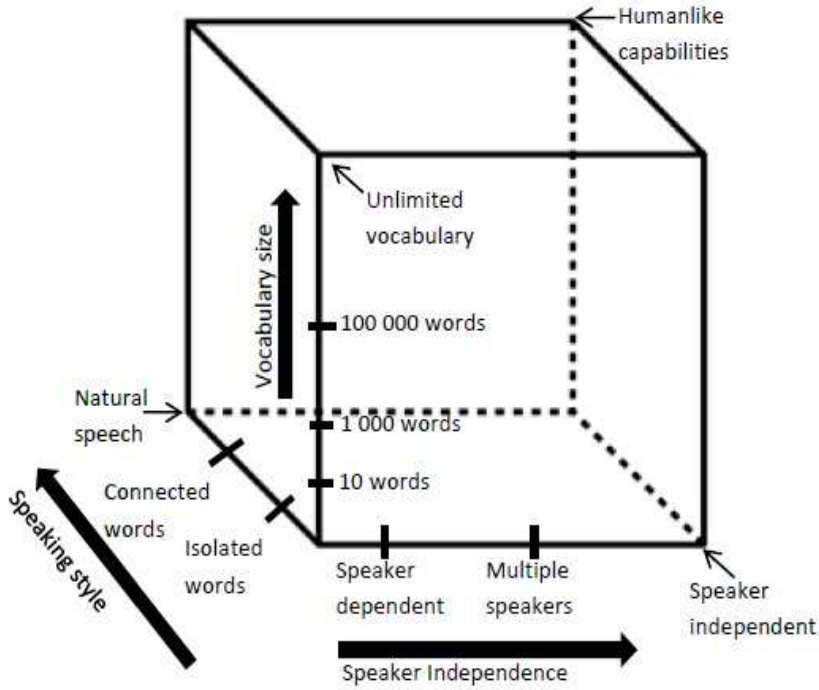
v nějaké promluvě, jsme schopni provést rozumný odhad, jaká další slova by mohla následovat. Pravděpodobnost každého slova je „podmínečně závislá“ na ostatních slovech v té samé promluvě. Tato statistická závislost všech slov v jedné promluvě tak činí matematické kalkulace zvládnutelné. Například, řekneme-li slovo „lev“, je velká pravděpodobnost, že následujícím slovem bude „řve“, a velmi malou pravděpodobnost bude mít slovo „zelený“. (Pieraccini 2012, s. 117)

Výpočet pravděpodobnosti posloupnosti slov vzhledem k akustickému pozorování je úkol *akustického modelu*. Akustické modely by měly být flexibilní, přesné a účinné. Podmínky, za kterých je používán rozpoznávač řeči, jsou velmi často odlišné od podmínek trénování, z tohoto důvodu je důležitá jeho flexibilita. Přesnost je vyžadována kvůli požadavku odlišit foneticky podobná slova s lingvisticky odlišnými významy. A účinnost je nutná v případě, kdy je klasifikátor řeči použit v reálných aplikacích a jeho odezva se musí uskutečnit v reálném čase. Jako velmi efektivní způsob řešení této náročné úlohy se ukázalo využití tzv. *skrytých Markovových modelů* (angl. Hidden Markov Model, zkr. HMM), které modelují nikoli celá slova, ale kratší subslovní jednotky (například fonémy, alofony, trifony apod.). Systém akustických Markovových modelů je obvykle trénován na vhodně připravené a rozsáhlé řečové databázi a z jednotlivých natrénovaných modelů subslovních jednotek jsou pak dle potřeby sestavovány modely slov i celých promluv. (Psutka 2006, s. 19)

Ačkoli skryté Markovovy modely ukázaly mnohem lepší výsledky než kterékoli pokusy předtím, jejich rozšíření ve výzkumných centrech řečových technologií ve světě nenastoupilo ihned. Poté, co IBM představila své výsledky na mezinárodní konferenci, mnozí členové z oboru umělé inteligence a lingvistiky nerozuměli, nedůvěřovali a ani nezamýšleli realizovat HMM výzkum. Objevily se skeptické názory typu, že problematika rozpoznávání řeči nemůže být vyřešena zredukováním inteligenční aktivity řeči na pouhé počítání a pravděpodobnost. „Mozek nepoužívá pravděpodobnost“ byla jednou z námitek proti HMM teorii. Ale navzdory těmto názorům tento způsob modelování a rozpoznávání převládl a i v současné době prokazuje nejlepší výsledky. (Pieraccini 2012, s. 132-134)

2.2.3 Nesnáze v oblasti rozpoznávání řeči

I když byl na poli zpracování řečového signálu a jeho klasifikace učiněn od pionýrských dob obrovský pokrok, je zatím konstrukce zařízení, které by bylo schopno rozpoznat promluvu jakéhokoli řečníka užívajícího libovolná slova daného jazyka, ještě poměrně vzdálenou budoucností. Pro utvoření lepší představy o tom, co by měl rozpoznávač řeči všechno ovládat a obsahovat, slouží Obr. 2.1, zároveň přiblížím problémy, se kterými se výzkum potýká.



Obrázek 2.1: Krychle přestavující komplexnost v oblasti rozpoznávání řeči (Pieraccini 2012, s. 60).

Sensus communis: Řeč je plná nejasností a my ji chápeme pomocí kontextu. Naše znalosti, zkušenosti a rozpoznání situace, ve které se ocitáme, nám napomáhá porozumět druhému. Právě toto nemá nic společného s jazykem, ale se znalostmi a zkušenostmi. Lidské vědění zdaleka přesahuje pouhý sběr dat a faktů. A přesto, že jsou dnes počítače schopny uchovat a zpracovat neuvěřitelné množství informací, stále jim chybí schopnost pravého lidského poznání a porozumění. Je možné, abychom byli

schopni předat alespoň část našeho vědění, které je zapotřebí k pochopení nejednoznačnosti a složitosti v lidské řeči strojům, a učinit tak obrovský pokrok nejen v oblasti řečových technologií? (Kurzweil 1997, s. 133)

Vložení těchto znalostí do počítače s sebou nese velké ambice, které měli i tvůrci projektu *Cyc*, který byl odstartován v roce 1984 a měl za úkol vytvořit takovou kolekci dat, která by počítači poskytla veškeré informace potřebné k tomu, aby mohl disponovat něčím podobným jako je lidský „selský“ rozum (Pieraccini 2012, s. 266). Za podobné místo, které by mohlo nést tento potenciál, bychom dnes mohli označit prostředí webu, které nám v jistém smyslu poskytuje řadu informací, ze kterých by mohl čerpat i rozpoznávač řeči.

Variabilita řečového signálu: Promluva každého z nás v sobě zahrnuje jedinečnost a co víc, i my sami nevyslovíme stejná slova nikdy totožně, právě tato variabilita lidské řeči je jednou z překážek, se kterými se tento obor potýká. V každé fázi zpracování se systém musí přizpůsobit individuálním charakteristikám mluvčího, každý z nás má jinou barvu hlasu, jiný přízvuk, odlišné tempo řeči, dialekt atd. Systémy rozpoznávání řeči se proto dělí na systémy závislé (natrénovány na hlas konkrétního řečníka nebo malé skupiny řečníků), anebo nezávislé (natrénovány na hlasy stovek i tisíců různých řečníků). (Psutka et al. 2006, s. 195)

Akustické prostředí: Kvalita rozpoznání řečové signálu závisí také na prostředí, ve kterém se mluvčí vyskytuje, hlučné prostředí tak může zhoršit výkon programu. (Psutka et al. 2006, s. 195)

Homonyma, slova, která znějí stejně, ale nesou jiný význam, jsou také tvrdým oříškem pro rozpoznávač řeči. Vyslovíme-li například větu: „Uteklo mi oko.“, společník v mezilidské konverzaci pochopí i z pouhého tónu hlasu, že právě pletu. Může počítač pochopit, že nemluvím o svém zrakovém orgánu, ale o ruční práci? V tomto ohledu udělaly ASR systémy velký pokrok tam, kde je určena tematická doména. Také propojení programu s webovým prohlížečem je užitečným řešením v tomto ohledu.

Problém koartikulace, který je jedním z dalších problémů vyskytujících se v souvislé promluvě. Fonetické vlastnosti začátku a konce slova v závislosti na kontextu okolních slov mohou systému velmi znesnadnit úkol. Velmi často je

uváděn příklad anglické věty „How to wreck a nice beach", kterou může program rozeznat jako „How to recognize speech". (Kurzweil 1997, s. 131)

Syntaktické nejasnosti: V roce 1963 Susumu Kuno na Harvardské univerzitě odhalil hloubku nejednoznačnosti v jazyce. Zeptal se počítačového analyzátoru, co znamená věta „Čas letí jako šíp" („Time flies like an arrow") a počítač si nebyl jistý, jak má správně odpovědět. Člověk interpretuje tuto větu jako metaforické vyjádření toho, že čas utíká rychle. Slovo „time" však může označovat v angličtině jak vyjádření pro čas jako podstatného jména, tak funkci slovesa „změřit čas". Stejně tak slovo „flies" má v angličtině význam letí, ale zároveň znamená plurál od slova moucha. Tato věta tak může být pochopena více způsoby ("Time the flies as you would time an arrow", "Time the flies as fast as an arrow", "Time the flies flying in an arrow formation"). (Kurzweil 1997, s. 135)

Objevily se zde syntaktické nejasnosti, ze kterých bylo jasné, že chápání jazyka mluveného či psaného vyžaduje jak znalosti vztahů mezi slovy, tak i jejich skrytý význam. Ale i přesto, že se jedná o velmi obtížný problém v oblasti řečových technologií, udělaly dnešní ASR systémy pokrok tam, kde je určena tematická doména. (Kurzweil 1997, s. 136)

Rozpoznávání souvislé řeči: V běžné mluvě nám samozřejmě žádné značky nepodávají informaci, kdy slovo končí a začíná. Tohoto jevu si můžeme velmi dobře všimnout v situaci, kdy jsme mezi lidmi, jejichž jazyku nerozumíme. Se stejným problémem se potýká i rozpoznávač řeči. Pokud pracuje se samostatně vyslovovanými slovy, nejde o tak těžký úkol, jedná-li se ale o souvislou řeč, která je ještě navíc spontánně pronesena, obsahuje nespisovné výrazy a neadekvátní gramatické vazby, jedná se o velmi náročnou úlohu. (Psutka 2006, s. 195)

2.3 Syntéza řeči

2.3.1 Mechanické a elektronické syntetizéry

Další nezbytnou složkou, kterou potřebujeme, chceme-li vést dialog s počítačem, je to, aby nám nějakým způsobem odpovídal. V oblasti řečových technologií se jedná o proces syntézy řeči, umělé produkce lidské řeči. Sestavit syntetizér, který by produkoval řeč se stejnou přirozeností, emocemi, expresivitou a srozumitelností jako člověk, je stále mimo dosah našich technologií, a to i přesto, že pokusy o vytvoření mluvícího stroje nejsou pouze záležitostí současnosti. Po staletí se lidé pokoušeli sestrojit takovéto zařízení a v některých případech by se mohlo zdát, že úspěšně. Dřívější úspěchy však byly umožněny jen díky podvodu.

První vědecký pokus o produkci řeči prostřednictvím stroje můžeme spatřit již v roce 1779 v St. Petersburgu u Christiana G. Kranzensteina. Jeho přístroj byl schopen prostřednictvím akustických rezonátorů, které napodobovaly hlasový trakt člověka, uměle vytvářet samohlásky. V roce 1791 Wolfgang von Kempelen vynalezl zařízení, které imitovalo anatomii lidského artikulačního ústrojí. Lehce zdokonalenou verzi Kempelenova mluvicího zařízení sestrojil přibližně o sto let později Charles Wheatstone. Skládalo se z hlavních měchů, rákosu, pryže, velké kožené trubice a dvou malých trubiček. Měchy měly funkci plic, rákos simuloval činnost hlasivek a ústa byla vymodelována pomocí pryže. Rezonanční prostor představovala kožená trubice a malé trubičky fungovaly jako nozdry. Celé zařízení se ovládalo pomocí pák a bylo údajně schopné vydávat primitivní slova a dokonce i věty. (Psutka 2006, s. 533)

V roce 1930 sestrojil Homer Dudley první elektronický syntetizér souvislé řeči *The Voder* (Voice Operating Demonstrator), jenž byl veřejnosti představen až v roce 1939 na světovém veletrhu v New Yorku. Tento syntetizér musel být ovládán proškoleným člověkem a byl schopen produkovat srozumitelnou a souvislou řeč. Bylo tak již v roce 1939 potvrzeno, že je možné vytvářet umělou řeč. (Pieraccini 2012, s. 48)

V roce 1953 byl sestaven další elektronický syntetizér řeči PAT (Parametric Artifical Talker), jehož vynálezcem byl Walter Lawrence. PAT byl prvním paralelním

formantovým syntetizérem, pracoval na principu zjednodušeného modelování lidského hlasového ústrojí pomocí formantů. Jednalo se o systém, který obsahoval základní řečové parametry a pravidla, která byla odvozena ze znalostí o lidské řeči. Později, v roce 1958, byl na institutu MIT vytvořen artikulační syntetizér DAVO (Dynamic Analogue of the Vocal tract), který měl generovat řeč pomocí mechanické simulace lidských artikulačních orgánů a hlasivek. (Pieraccini 2012, s. 195, 197)

2.3.2 Digitální syntetizéry

Stejně jako tomu bylo u rozpoznávání řeči, i v oblasti syntézy nastal zlom s nástupem počítačů. Paralelně s vývojem řečových modelů se vyvíjí i syntéza řeči z textu (angl. Text-to-speech, zkr. TTS). Jde o důležitý krok ve vývoji, protože se až do této doby výzkum soustředil spíše na reprodukci, než na vytváření nových promluv. TTS se navíc zabývá také zpracováním přirozeného jazyka a fonetickým popisem promluvy. Jejím cílem je vytvářet řeč z libovolného textu. Text je v tomto případě podroben metodám předzpracování textu, morfologické, syntaktické a sémantické analýze, a převeden na posloupnost fonetických a prozodických značek. (Psutka 2006, s. 535)

S nástupem počítačů se objevují digitální verze elektronického syntetizéru PAT a vyvíjí se i další formantové syntetizéry, které byly velmi často používány v systémech TTS. Tato technika však byla později vytlačena *konkatenační syntézou*. Konkatenační řečová syntéza je založena na předpokladu, že jednotlivé zvuky, ze kterých se řeč obecně skládá, lze reprezentovat pomocí konečného počtu řečových jednotek. (Psutka 2006, s. 547-548)

Nejrozšířenější metoda současnosti je *korpusově orientovaná syntéza* (angl. corpus-based synthesis). Jeden z prvních syntetizérů fungující na tomto principu se nazývá *Festival*. Byl sestaven roku 1997 Alanem Blackem a Paulem Taylorem na univerzitě v Edinburghu. Tato metoda využívá rozsáhlé řečové korpusy a kvalita řeči je závislá na kvalitě řečového korpusu. S řečovým korpusem, který obsahuje obrovská data, může být kvalita syntetické řeči udivující. Většina komerčních syntetizérů současnosti funguje na tomto principu. (Pieraccini 2012, s. 204)

Například společnost *SpeechWorks* v roce 2000 získala potřebnou technologii, aby mohla produkovat vysoce kvalitní umělou řeč, založenou na korpusově orientované syntéze, systém, zvaný *Speechifi*, jehož srozumitelnost a přirozenost byla velmi působivá. Tento systém se stále zlepšuje a jeho hlasy působí velmi přirozeně, obsahuje více personality a dokonce i emoce. Systém produkuje chichotání a smích, počítači jsou také propůjčeny hlasy slavných herců a zpěváku, syntetický hlas tak zní přirozeněji. (Pieraccini 2012, s. 204).

S řečovou syntézou se dnes můžeme setkat kdekoli a kdykoli. U navigace v autě, při čtení knih, e-mailů a zpráv sms. Slouží jako pomůcka pro handicapované a mohou se stát také velmi užitečnými pomocníky při výuce cizích jazyků. Současný trend v oblasti řečových technologii je snaha o vytvoření co nejvěrnější kopie lidského hlasu.

2.3.3 Nesnáze v oblasti řečové syntézy

Vývoj syntézy řeči je provázen dvěma hlavními cíli. Badatelé z tohoto oboru neusilují pouze o vytvoření systému, který produkuje řeč srozumitelnou, ale také přirozenou, a ani jeden z těchto úkolů není jednoduchý. Doposud je historie syntézy řeči kompromisem mezi srozumitelností a přirozeností, snaží se však dosáhnout obou.

Převedení textu na řeč se potýká s problémy, které člověk nepovažuje za překážku, můžeme číst text ve svém rodném jazyce bez jakýchkoli obtíží, stroj ale našemu jazyku nerozumí. Produkce čísel, zkratek, symbolů, anebo i samotné přečtení slov, která se v mnoha jazycích jinak píší, než vyslovují, byla překážkou, kterou se vědci museli zabývat. Tato problematika se v současnosti řeší pomocí fáze předzpracování textu, kdy je například označení římské číslice převedeno do slovní podoby. (Olive 1997, s. 118)

Komplexnost jazyka v sobě však opět skrývá řadu výjimek a nejednoznačností, které se liší jazyk od jazyka. Jen v češtině správné skloňování číslic může přinést řadu komplikací. Co se týče správné výslovnosti, i v oblasti syntézy řeči může způsobit problémy jev zvaný **koartikulace**, o kterém jsem hovořila již u problematiky, která se týká rozpoznávání řeči.

Další těžkosti v této oblasti přináší tzv. **homografy**, slova, která se shodně píší, ale nesou jiný význam, a v některých případech se i jinak vyslovují (Olive 1997, s. 118). V češtině se tato slova nevyskytují tak často jako například v angličtině, ale tím, že význam homografů závisí na kontextu, v jakém se vyskytují, představují velký problém v oblasti řečové syntézy.

Chceme-li také, aby uměle vytvořený hlas nezněl „uměle“, je důležité, aby obsahoval zvukové prvky založené na prozodických prostředcích řeči (Olive 1997, s. 118). Pomocí intonace můžeme rozlišit otázku od oznámení. Například v němčině jsou věty rozlišovány pomocí pevně daného slovosledu, čeština jej však nemá, a proto v ní intonace hraje důležitou roli. Tónem našeho hlasu vyjadřujeme emoce a přízvukem odlišujeme jednotlivá slova. V češtině je přízvuk pevný, avšak například v angličtině nebo ruštině může jeho špatné umístění rozlišit význam slov.

Celková vyspělost dialogových systémů souvisí zároveň se složitostí jednotlivých jazyků, financemi, které mají jednotlivé země k dispozici, a také s jejich rozšířením. Problémy, se kterými se syntéza řeči potýká, jsem pouze stručně naznačila. Vytvoření takového syntetizéru, jehož jazykový projev by byl od člověka k nerozeznání, není tak jednoduché, jak by se na první pohled mohlo zdát.

2.4 Vize budoucnosti

Dosažení vysoké úrovně řečové komunikace mezi člověkem a strojem ale není pravděpodobně úkolem neuskutečnitelným. Nechtěla bych však ve své práci předkládat hypotézy o budoucím vývoji v oblasti řečových technologií, mám v úmyslu představit pouze tři hlavní názory, které podle mého mínění reprezentují odlišná hlediska, pomocí nichž bych chtěla ukázat, jak různorodé myšlenky se ve spojení s vývojem techniky objevují.

2.4.1 Raymond Kurzweil

V roce 1965 byl v časopise *Electronics* zveřejněn článek, jehož autorem byl Gordon Moore, spoluzakladatel společnosti Intel. Gordon Moore pozoroval vývoj v oblasti integrované elektroniky, která je podle něj budoucností elektroniky samotné. Podle jeho výpočtů se počet složek v integrovaných obvodech v období od roku 1958 (rok, kdy byl zkonstruován první integrovaný obvod) do 1965 zdvojnásobil. Zároveň předpokládá, že tento trend bude pokračovat po dobu deseti let (Moore, 1965). Později, v roce 1975, změnil svou předpověď na zdvojnásobení výkonu každé dva roky.[1]

Mooreův zákon, jak byla později jeho prognóza pojmenována, se prozatím naplňuje. Čipy a počítače se staly méně nákladnými a mnohem výkonnějšími. Samo vyslovení této předpovědi ovlivnilo vývoj elektroniky, ale i podle samotného Moorea nemůže trvat do nekonečna.

“In terms of size [of transistors] you can see that we're approaching the size of atoms which is a fundamental barrier, but it'll be two or three generations before we get that far - but that's as far out as we've ever been able to see. ” (Dubash 2005)

[1] Často se také můžeme v literatuře setkat s údajem 18 měsíců, ale autorem této předpovědi je Mooreův kolega David House. (Kanellos 2003)

„Z hlediska velikosti (tranzistorů) můžeme vidět, že se přibližujeme velikosti atomů, které jsou zásadní překážkou, ale to potrvá ještě dvě nebo tři generace, než se dostaneme tak daleko – neboť to je maximum, kam jsme kdy byli schopni dohlédnout.“

Podle Raymonda Kurzweila (1997, s. 163), známého futuristy, zůstane tento zákon platný i v příštích letech. Nebudeme tak muset čekat dlouho na vytvoření počítače s obrovským výkonem, který je významnou složkou, která je zapotřebí k dosažení svatého grálu v oblasti řečových technologií. Rozpoznávání lidské řeči se potýká s hlavním problémem potencionální kombinatorické exploze, jazyk také souvisí s úrovní inteligence a rozpoznáním kontextu pomocí našich předchozích znalostí. Lidské znalosti a inteligenci nelze oddělit od schopnosti rozumět lidskému jazyku. Dosavadní výpočetní technologie nejsou dostatečně silné na to, abychom pomocí nich mohli dosáhnout těchto cílů. Podle Kurzweila (1997, s. 163) by počítač s dostatečně velkým výkonem mohl tuto problematiku vyřešit.

V současné době je i nejvýkonnější počítač stále mnohem jednodušší než lidský mozek, který dokáže provést biliony výpočtů současně a je tímto paralelismem stále výkonnější. Lidský mozek má kolem 100 miliard neuronů, každý z nich má 1000 spojení s ostatními neurony a všechna tato spojení mohou provádět své výpočty současně. Mozek je tak schopný provést 100 bilionů operací zároveň. Ačkoli jsou lidské neurony pomalé, jejich masivní paralelismus umožňuje mnohem větší výkonnost, než jaké může dosáhnout počítač. Lidská inteligence je tak mnohem pružnější a rozsáhlejší než počítačová, a řada vědců také tvrdí, že tato mezera nemůže být nikdy překonána. Mezi takové ale nepatří Raymond Kurzweil, který vidí budoucnost v navržení simulace neuronové sítě. Mohli bychom tak kompletně zmapovat strukturu mozku, a vytvořit tak software podle vzoru mozku. Zmapováním propojení, umístění a obsahu dendridů, presynaptických váčků a dalších nervových komponentů a celé organizace mozku včetně paměti, by mohla být v počítači s dostatečně velkou kapacitou vytvořena simulace lidského mozku. (Kurzweil 1997, s. 162-166)

„Pomocí simulace neuronové sítě budeme moci vytvořit systémy, které budou schopné rozumět lidské řeči mluvené i psané a budou tak moci zpracovávat lidské

znalosti. Tato schopnost, extrahovat z jazyka nebo psaného textu vědění, je u dnešních počítačů stále omezena, ale ve 20. letech 21. století budou počítače schopné samostatně číst, ukládat a zpracovávat texty. Dále budou schopny pracovat s každým druhem literatury a budou tak moci samostatně disponovat věděním.“ (Kurzweil 1999, s. 20)

Kurzweilovy předpovědi jsou často kritizovány z mnoha stran, např. Wolf Singer, ředitel Max-Planck Institutu pro výzkum mozku řekl: „Věřím, že Kurzweil je na velkém omylu, pokud věří, že samotné zvýšení výpočetního výkonu počítačů může vést ke kvalitativní změně. Analogie mezi počítačem a mozkem je povrchní. Oba systémy mohou sice provádět logické operace, ale jejich architektura je radikálně odlišná. Problém leží především v tom, že počítač pracuje na jiných algoritmech než biologické systémy.“ (Heinz Nixdorf Museumsforum 2001, s. 276)

Kurzweil si stojí za svým a tvrdí, že rozdíl mezi lidským mozkem a počítačem bude překonán v příštích dekádách, což nám mimo jiné umožní zdolat i nesnáze, se kterými se vývoj řečových technologií potýká. Přesněji řečeno do roku 2029 by podle něj mohla být vyvinuta taková umělá inteligence, která přesahuje naše dosavadní pojetí strojů a počítačů. Půjde o sjednocení biologické a nebiologické inteligence, které nám umožní být milionkrát chytřejší, než jsme - to je náš osud, řekl v dokumentárním filmu Plug and Pray (2010).

2.4.2 Joseph Weizenbaum

Zdá se tedy, že Kurzweil nepochybuje o moci technologií jako součástí nás samých a pozitivech, která nám mohou přinést. Avšak ne každý věří v tento progresivní vývoj vědy, ne každý do ní vkládá bezmeznou důvěru a ne každý jí přisuzuje pouze pozitiva.

„Uzavřeli jsme faustovský pakt s vědou a technologií. Věda nám dala, co jsme chtěli, např. léky proti mnoha nemocem a některých nemocí nás zbavila. Další dobra, která požadujeme, jsou nesčetná a pořád chceme víc. Přehlížíme, že výhody, o nichž se domníváme, že jsme je vyhandlovali, se proměnili v opak.“ To jsou slova Josepha Weizenbauma (2002, s. 144), bývalého profesora informatiky na MIT a zároveň kritika vlivu technologií na společnost. Lidé podle něj chápou techniku jako přírodní nutnost, jejímuž nástupu se nelze vyhnout (jak můžeme vidět právě u Kurzweila). Podle

Weizenbauma nebude počítač nikdy rozumět přirozeným jazykům, rozumění předpokládá vždy nějakou bytost, která je schopna poznávat v souvislostech. Nakolik to dokáže, závisí na její vlastní historii. Každý člověk rozumí něčemu jinak než druhý, není možné uchopit absolutní obsah. A právě tuto historii, podle Weizenbauma, nemůžeme dát počítači. Člověk je něco jedinečného, jeho dovednosti, znalosti, prožitky ani schopnost porozumění všemu, co obsahuje lidská řeč, nebude možné vložit do počítače. (Weizenbaum 2002, s. 141)

2.4.3 ROILA

O tom, zda má pravdu Raymond Kurzweil nebo Joseph Weizenbaum, nemůže rozhodnout nikdo z nás. Co však můžeme s jistotou říci, je, že dialogové systémy by nám mohly být v řadě odvětví velmi užitečné a jejich vývoj bude pokračovat. Jedno z řešení, jakým směrem bychom se také mohli ubírat, je uměle vytvořený jazyk ROILA.

Vzhledem k dlouhému vývoji v oblasti řečových technologií a jejich stálé nedokonalosti přišla skupina vědců s nápadem vytvoření takového umělého jazyka, který by byl určen ke komunikaci mezi člověkem a strojem. Takový jazyk by byl snadno naučitelný pro člověka a přitom snadno rozpoznatelný pro robota. Jedná se tak o metodu, která by mohla zlepšit rozpoznávání řeči. Omezení, se kterými se setkáváme v oblasti rozpoznávání přirozené řeči, jsou jednou z překážek v rozvoji interakce mezi člověkem a strojem. Je možné tuto problematiku vyřešit zavedením jiného jazyka? Co se týče rukopisu, byl následován podobný plán, kdy byl vytvořen program *Graffity*, který byl pro uživatele lehce naučitelný a pro počítač lehce rozpoznatelný. Použitím stejné analogie bychom mohli vytvořit „Speech Recognition Friendly Artificial Language" (ROILA). Takovýto jazyk by byl vytvořen za účelem usnadnění rozpoznávání řeči počítačem a zároveň by byl jednoduše naučitelný pro člověka. Slovník tohoto jazyka je utvářen tak, aby minimalizoval možnost záměny slov, gramatika by nenesla nesrovnalosti a výjimky ve svých pravidlech. Výslovnost by se měla stát jednoduchou jak pro člověka, tak pro stroj. (Mubin et al. 2009)

Příklad jazyka ROILA:

Běžná anglická věta: You are a good person

ROILA: Bama wopa tiwil

Doslovný překlad jazyka ROILA: You good person [2]

Jsou-li přirozené jazyky pro počítač příliš složité, proč bychom nemohli vymyslet jazyk, který by byl uzpůsoben tak, aby minimalizoval nesrovnalosti, se kterými se počítač setkává. Zda by takové řešení mohlo přispět k zlepšení komunikace mezi člověkem a strojem, nemohu říci, vzpomeňme si na ideály vložené do předchozích uměle vytvořených jazyků. Vede nás to však k dalšímu tématu a hlavní otázce mé práce, co když bychom zanechali snah vytváření stroje, který má stejné vnější projevy jako člověk, ale snažili se najít nějaký kompromis, který by nám i tak umožnil příjemnou interakci mezi člověkem a strojem?

[2] Převzato z *Robot Interaction Language* [online]. [cit.21.4.2013]. Dostupné z http://roila.org/language-guide/

3 Uncanny valley

HELENA: Náno, pojď mne zapnout!

NÁNA: No hned, no hned. Bože na nebi, to je zvěř!

HELENA: Roboti?

NÁNA: Fi, ani je menovat nechci.

HELENA: Co se stalo?

NÁNA: Zas to jednoho u nás chytlo. Začne třískat do soch a vobrazu, skřípá zubama, pěnu u huby – Načisto pominutej, brr. Dyť to je horší než zvíře.

HELENA: Kterého to chytlo?

NÁNA: Toho – toho – Šak to ani křesťanský meno nemá! Toho z knihovny.

HELENA: Radia?

NÁNA: Zrouna toho. Šmarjájozef, já si to vošklivím! Ani pavouka si tak nevošklivím jako ty pohany.

HELENA: Ale Náno, že ti jich není líto!

NÁNA: Šak vy si je taky vošklivíte. Pročpak ste si mě přivezla sem? Pročpak žádnej z nich nesmí na vás ani šáhnout?

HELENA: Neošklivím, na mou duši, Náno. Je mi jich tak líto!

NÁNA: Vošklivíte. Každej člověk si je musí vošklivět. Dyť i ten pes si je vošklivý, ani sousto masa vod nich nechce; stáhne vocas a vyje, dyž cejtí ty nelidy, fuj.

HELENA: Pes nemá rozum.

NÁNA: Je lepší než voni, Heleno. Von dobře ví, že je něco víc a že je vod pánaboha. Dyť i ten kuň se plaší, dyž potká pohana. Dyť ani mladý to nemá, a i pes má mladý a každej má mladý –

HELENA: Prosím tě, Náno, zapínej!

NÁNA: No hned. Já říkám, to je proti pánubohu, to je ďáblovo vňuknuti, dělat ty maškary mašinou. Rouháni je to proti Stvořiteli, je to urážka Pána, kterej nás stvořil k vobrazu svýmu, Heleno. A vy ste zneuctili vobraz boží. Za tohle přijde strašnej trest z nebe, to si pamatujte, strašnej trest!

(Karel Čapek, R.U.R.)

Ruku v ruce se stále se zrychlujícím vývojem technologií jde i realizace snů, které provázejí lidstvo po celá staletí. Již nejstarší známý řecký básník Homér se ve svém příběhu o bohu Héfaistovi zmiňuje o zlatých dívkách, pomocnicích, které představují bytí podobné lidem, jsou vzdělané, umí mluvit a vykonávají různé práce. Příběhů popisující taková zařízení, která jsou jak vzhledem, tak i chováním podobná člověku, můžeme v literatuře a kinematografii najít nespočet. Ať už jsou vyobrazena v pozitivním světle jako naši ochránci a pomocníci, či jako hrozba, která ohrožuje naši existenci, vždy v nás, jako divácích a čtenářích, vyvolávají pocity, které většinou k neživým předmětům necítíme. Nechováme city k našemu vysavači či fénu, ale robot, který vykazuje byť jen některé znaky lidství, má určitý potenciál, aby v nás vyvolal citovou odezvu. A co víc, i my sami připisujeme strojům různé emoce.

Je možné vytvořit bytost, která je stejná jako my, která má své potřeby, city, touhy, sny a která je obdařena schopností myslet, přemýšlet a mluvit, stroj nerozeznatelný od člověka? Jaký by ale ve skutečnosti byl vztah člověka k takovému artefaktu, který vykazuje lidské vlastnosti? Jak má vypadat a mluvit, aby byla komunikace mezi člověkem a strojem co nejpřirozenější? Mnoho režisérů a spisovatelů nám uskutečnění takové představy ukazuje ve svých dílech, ale jaká je realita?

Současný pokrok v technologiích umožnil vývoj funkčních a realistických humanoidních robotů. Jako humanoid je označován takový stroj, který svou stavbou těla připomíná člověka, ale jeho vzhled je při tom mechanický. Naopak výroba androidů se zaměřuje na kopírování lidského vzhledu a chování. Android je definován jako umělá inteligence navržená s konečným cílem – být externím vzhledem a chováním nerozlišitelný od člověka. Mezníkem vývoje je vytvoření vztahu člověk-android na stejné úrovni jako člověk s člověkem. Definice androidů na základě cíle, ne na základě stavu současných materiálů, nám umožňuje rozlišit androidy a humanoidy v současnosti. (MacDorman a Ishiguro 2006, s. 299)

Humanoidní roboti mohou sloužit pro mnoho účelů, byli by schopni zastávat pozice osobních asistentů a být tak k ruce například nemocným a starým lidem. Jejich prostřednictvím by nám bylo umožněno prozkoumat vzdálená a nebezpečná místa a nebo by mohli zaujmout úlohu společníků v oblasti zábavy a učení. Představme si, jaké by bylo, kdyby každý z nás mohl vlastnit robota, který by nebyl pouze hračkou pro děti, ale i jejich věrným pomocníkem při dělání domácích úkolů. A naši prarodiče

by v něm mohli najít věrného posluchače svých nekonečných příběhů. Příkladem humanoidních robotů současnosti jsou roboti NAO, Asimo nebo Robokind od HansonRobotics, kteří jsou velmi pohybliví, rozpoznávají řeč a zároveň na ni reagují a odpovídají. Příkladem humanoida je i ženský humanoid HRP-4C.

Na univerzitě v Ósace byli pod vedením profesora Ishigura ve spolupráci s Kokoro Company vyvinuti 4 androidi (nazývaní také jako actroidi): Replie R0, Replie R1, Repliee Q1(známá jako Andosan) a Repliee Q2 (známá jako Uando). Uando má držení těla podobné člověku, umí mrkat a napodobovat dýchání a také disponuje reaktivním chováním prostřednictvím verbální a gestikulační komunikace. Byly jí umístěny dotykové senzory, které jí umožňují reagovat na dotyk. Pomocí svých senzorů je schopná také udržovat oční kontakt během konverzace. V interakci s člověkem reaguje na obsah a také prozodii. Pravdou je, že málokdo by se k takto vypadajícímu robotu mohl chovat hrubě, jak k tomu dochází u mechanických robotů. Podle profesora Ishigura např. Robovie a další humanoidi nemohou svým robotickým vzhledem a chováním vyvolat stejné vědomé nebo nevědomé odpovědi, jak je tomu u androidů. (MacDorman a Ishiguro 2006, s. 313-314)

V podobě Repliee1 Ishiguro představil kopií své dcery a poté i kopii sebe sama jako Geminoida HI-1. Ačkoliv se tito androidi doposud nepohybují stejně jako lidé, jejich mimika a pohyby rtů při mluvení a nebo gestika, jsou podobné těm lidským. Vytváření již žijících lidí je považováno za diskutabilní a jeho výtvory jsou posuzovány s jistou dávkou kritiky. (Barthelmeβ a Furbach 2012, s. 7)

Je toto cesta k realizaci našich představ, měli bychom v ní pokračovat, nebo raději změnit její směr? Máme se zaměřit na vytvoření lidské kopie, nebo se spokojit s výrobou robotů, kteří mají mechanický vzhled a syntetický hlas? A co by nám jejich výroba přinesla, nebo naopak vzala?

Musíme vzít i v úvahu, že stroj s lidským vzhledem, hlasem a chováním by byl tak složitý komplex, že by byl příliš komplikovaný pro jeho používání v každodenním životě (Švarný 2012, s. 38). Na druhou stranu, podle MacDormana a Ishigura (2006, s. 297), by vývoj těchto robotů mohl být přínosem nejen v oblasti technologií, ale i v oblasti kognitivních věd a ve výzkumu sociálních věd. Android by poskytl experimentální aparát, který by mohl být kontrolován s mnohem větší precizností než člověk. Androidi by tak mohli vytvořit testovací prostor pro sociální, kognitivní a

neurovědecké teorie. Velmi věrné kopie člověka mohou totiž vyvolat širokou škálu reakcí jako člověk samotný. Avšak vývoj těchto technologií má různá zákoutí, jemné nedostatky ve vzhledu a pohybu mohou vyvolat nevědomé negativní pocity, kterých si již v roce 1970 všiml Masahiro Mori.

3.1 Masahiro Mori a jeho myšlenky

V roce 1970 byl Masahiro Mori, nyní profesor na Technologickém institutu v Tokyu, požádán, aby přispěl do časopisu *Energy* článkem na téma: Zamyšlení nad robotikou. Mori v té době pracoval na výrobě protetické ruky, která v něm probudila stejný druh pocitu jako voskové figuríny, se kterými se setkal v dětství. „Byla to intuice, napsat esej na toto téma," říká Mori. (Kageki, 2012)

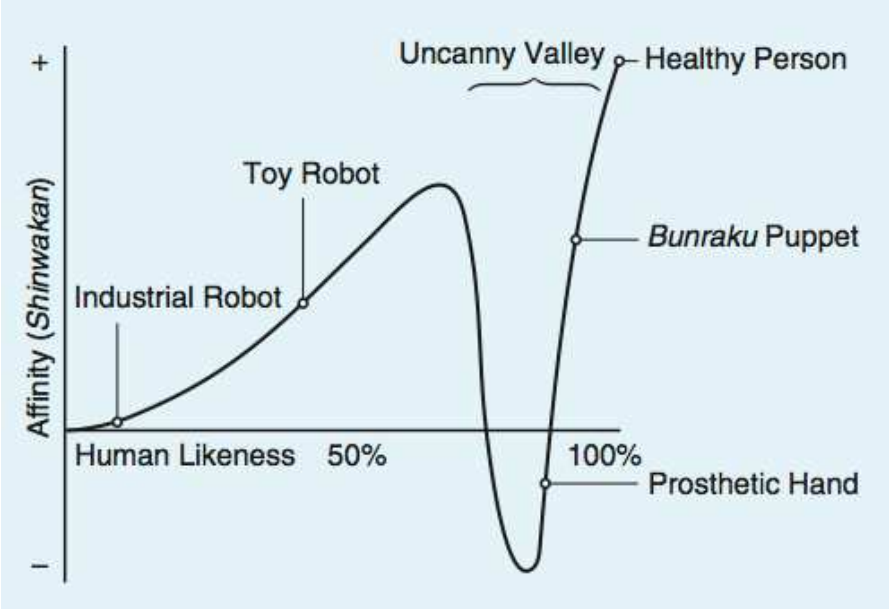
O 35 let později na mezinárodní konferenci o humanoidních robotech se dočkaly Moriho myšlenky velké pozornosti. Snaha o rozluštění možných tajemství, která jeho úvaha skrývala, se stala objektem zájmu mnoha vědců z různých odvětví. Pozornost věnovaná jeho eseji tkvěla v neustálém zdokonalování technologií, které umožnily vývoj robotů, o nichž Mori uvažoval.

Mori (2012) ve svém článku popsal hypotézu o reakci člověka na stroj, který vypadá a pohybuje se stejně jako člověk. Předpokládal, že reakce člověka na robota může přejít od pocitu empatie k odporu. Za předpokladu, že se robot svým vzhledem přiblíží člověku, ale jeho podoba je nedokonalá, mohou se lidské pocity změnit z pozitivních na negativní. Tento zvrat byl v japonštině označen názvem *Bukimi no tani*. Do angličtiny je překládán jako *The uncanny valley*[3], i když by podle Karla F. MacDormana, profesora na univerzitě v Indianě a jednoho z překladatelů Moriho eseje, měl přesný překlad znít *Valley of eeriness*. Největší jazykovou výzvou bylo pro překladatele japonské slovo *shinwakan*, které bývá překládáno jako obeznámenost,

[3] Do češtiny bychom tento pojem mohli přeložit jako tajemné údolí nebo, jak uvádí Wikipédia, tajemný val, ale vzhledem k problematice, která provází překlad z japonštiny do angličtiny, zůstanu u použití anglického termínu uncanny valley.

úroveň pohodlí a afinita.[4] Ale podle MacDormana nevystihuje ani jedno z těchto slov to, co měl Mori na mysli. (Hsu 2012)

„Myslím si, že jde o pocit, který zažíváme v přítomnosti jiného člověka – moment, ve kterém cítíme souznění s někým jiným a zkušenost „setkání myslí“, negativní význam slova „shinwakan“, the uncanny, podivný, znamená moment, kdy se ten pocit synchronizace rozplyne, okamžik, kdy zjistíme, že naše spřízněná duše není nic než kouř.“ řekl MacDorman. (Hsu 2012)



Obr. 3.1: Grafické znázornění jevu uncanny valley u robota, který není v pohybu. (Mori 2012)

Obr. 3.1 znázorňuje graf, který popsal Mori ve své práci. Na ose „y“ vytyčil *shinwakan* a na ose „x“ podobnost člověku. Křivka v grafu stoupá až do takového bodu, který znázorňuje robota blížícího se lidské podobě, poté se linie zanoří do uncanny valley, a to těsně před téměř dokonalým lidským zpodobením. Tento pokles představuje okamžik, ve kterém je v lidské mysli vyvolán děsivý pocit. Mnoho

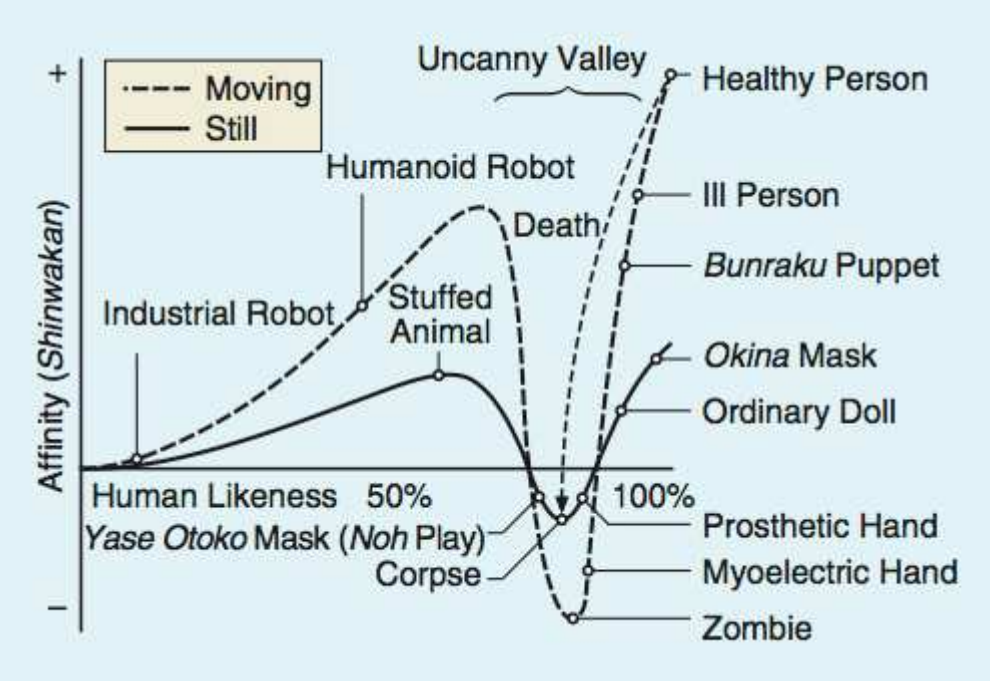
[4] V anglické literatuře se můžeme nejčastěji setkat s pojmy: *familiarity, likableness, comfort level* a *affinity*. V druhém překladu Moriho eseje je ale používáno slovo *affinity*.

badatelů ale upozorňuje na to, že Moriho graf není doslovně správný a stále se neshodnou, jak tento pojem přesně definovat (Hsu, 2012).

Design robota se musí přizpůsobit jeho účelu. Stroje, používané v průmyslovém odvětví, musí vykonávat obdobné úkoly jako tovární dělníci, ale nemusí při tom vypadat stejně jako oni. V tomto případě není design robota jeho významnou složkou a tím, že není člověku podobný, nevzbuzuje vysokou míru afinity, nachází se tedy na začátku křivky grafu znázorněného na obr. 3.1. Jinak tomu je u mechanických hraček, u kterých je vzhled důležitější než jejich funkce, jsou konstruovány tak, aby měly určité lidské rysy, nohy, ruce, hlavu a torzo. Svým vzhledem připomínají člověka a děti si k nim mohou snáze vytvořit citový vztah. Proto je tento typ zobrazen na začátku prvního oblouku v Moriho grafu. (Mori 2012)

V době, kdy Mori napsal svůj článek, pracoval na protetické ruce, která byla na první pohled téměř nerozeznatelná od ruky živé, simulovala vrásky, žíly, nehty, a dokonce i otisky prstů. Tato ruka dosáhla určitého stupně podobnosti s lidskou rukou, jakmile si však uvědomíme, že není živá, ale umělá, zažíváme děsivý pocit. Například při obyčejném podání ruky ucítíme chlad a bezvládnost a naše pocity tak náhle „spadnou“ do uncanny valley.

Loutka bunraky, která je další příkladem znázorněným na Moriho grafu, samozřejmě ani zdaleka nedosahuje realistického vzezření jako protetická ruka. Sedíme-li ale v divadle a jsme-li v určité vzdálenosti od jeviště, loutky bunraku na nás svým zjevem a pohyby působí velice lidsky a vzbuzují tak velkou míru souznění. (Mori 2012)



Obr. 3.2: Grafické znázornění jevu uncanny valleyu robota, který je v pohybu.[5] (Mori 2012)

Obr. 3.2 znázorňuje, jak vnímáme roboty, kteří se začnou pohybovat. Přítomnost pohybu změní tvar znázorněné křivky. Sledujeme-li vypnutý průmyslový stroj, působí na nás pouze jako naolejovaný kus kovu, jakmile je ale naprogramován tak, aby jeho úchytné manévry působily dojmem pohybu lidské ruky, může v nás zvýšit pocit familiárnosti. Opačně je tomu u pohybu protetické ruky, který v nás pocit hrůzy může ještě více prohloubit. Pouhá protetická ruka nese potenciál toho, aby v nás probudila negativní pocity, jak by tomu bylo u robota? (Mori 2012)

Další příklad můžeme sledovat na světové výstavě v Ósace, která se konala v roce 1970. Bylo zde představeno několik sofistikovaných robotů, jeden z nich disponoval 29 páry obličejových svalů, aby byl schopen napodobit lidský úsměv. Úsměv je však velmi citlivý na rychlost pohybu jednotlivých svalů a i malá odchylka jej může změnit z přirozeného na odpudivý. Podle Moriho může forma, která se svým vzhledem velmi podobá člověku, kdykoliv spadnout do oblasti uncanny valley, a to

[5] Noh je tradiční japonská forma divadla, ve kterém herci nosí masky. *Yase otoko* je maska, která přestavuje ducha z pekla a maska *okina* zpodobňuje starého muže.

díky variabilitě pohybů a reakcí, které to v nás vyvolává. Designeři robotů by se měli takovému riziku vyhnout a cílem jejich práce by měl být první vrchol vykreslený ve zmíněném grafu, který představuje robota, který má jistý stupeň lidské podoby, a vyvolává tak značný pocit spřízněnosti. (Mori 2012)

"In fact, I predict it is possible to create a safe level of affinity by deliberately pursuing a nonhuman design. I ask designers to ponder this. To illustrate the principle, consider eyeglasses. Eyeglasses do not resemble real eyeballs, but one could say that their design has created a charming pair of new eyes. So we should follow the same principle in designing prosthetic hands. In doing so, instead of pitiful looking realistic hands, stylish ones would likely become fashionable." (Mori 2012)

„Vlastně předpovídám, že je možné dosáhnout bezpečné míry afinity při záměrném vytvoření nelidského designu. Žádám designéry, aby se nad tím zamysleli. Pro ilustraci tohoto příkladu zvažte brýle. Brýle nepřipomínají skutečné oči, ale dalo by se říci, že jejich design vytvořil okouzlující pár nových očí. Stejnou zásadou bychom se tak měli řídit při navrhování protetické ruky. Pokud bychom tak učinili, mohli bychom místo žalostně vzhlížející realistické ruky vytvořit ruku stylovou, která by se mohla stát módní záležitostí."

Mori chtěl svým článkem pomoci lidem, kteří se zabývají robotikou, chtěl zároveň upozornit na problematiku, se kterou se mohou setkat. Měl by se uncanny valley překonat? Nebo bychom měli zůstat u výroby humanoidů s mechanickým vzhledem a hlasem? Mori si i o 42 let později stále stojí za svým: „Designéři robotů by se měli držet prvního vyvýšení a nesnažit se překonat uncanny valley. Není zajímavé vytvořit robota, který vypadá a chová se stejně jako člověk, roboti by měli být odlišní od lidských bytostí." (Kageki 2012)

3.2 Vysvětlení uncanny valley

Japonská bajka "Hanasaka Jiisan" vypráví příběh o psu, který najde zlato, ale sám jej vykopat nedokáže, a proto svým štěkotem ukáže lidem, kde mají začít kopat. Masahiro Mori sám sebe přirovnává ke psu z tohoto příběhu: „Záhadné údolí byla jedna z věcí, které jsem vycítil jako pes zlato, ale vykopat jej neumím.“ (Kageki 2012)

Ze všech artefaktů kolem nás nám pravděpodobně není nic více podobného nežli robot, stroj, který jsme vytvořili, aby nám byl co nejlépe nápomocen při našich každodenních činnostech. Ale ačkoli jsme si podobní, je mezi námi stále hluboká propast, metabolismus, reprodukce, rod, kultura, vědomí, to vše a ještě mnohem více nás odlišuje od umělých bytostí. Není pak překvapující, že vidíme tato stvoření s jistou dávkou rozpolcenosti, která je silnější u robotů, kteří se svým designem snaží co nejvíce napodobit člověka.

Jednoznačné stanovisko, proč se u nás tyto negativní pocity objevují, nezaujímáme a nabízí se řada vysvětlení, která jsou z větší části netestována. Je zřejmé, že existuje mnoho rozdílných cest jak odlišit lidské normy vzhledu a chování a některé jsou více záhadné než ostatní. Pocit tajuplnosti, vyvolaný nedokonalou simulací lidského vzhledu a pohybu, nemůže být samostatným fenoménem, který by se dal vysvětlit pouze jedním mechanismem. Vnímání robota v pozitivním světle může mít řadu dimenzí, některé z nich mohou být personálně, biologicky, kulturně a emocionálně zabarvené. S tím v mysli můžeme zvážit následující vysvětlení:

Narušení očekávání: Na základě našich předchozích zkušeností zařazujeme nové situace anebo doposud neznámé předměty do jakéhosi rámce. Tyto rámce jsou strukturovaná data reprezentující stereotypní situace. Setkáme-li se se strojem, který vypadá jako člověk, vnímáme jej na jednu stranu jako neživý přístroj, ale na druhou stranu naruší jeho lidské rysy naše očekávání. V mezilidské interakci lidé naplňují námi předpokládané chování, zatímco androidi tuto presumpci naruší. Právě to může vyvolat negativní pocity. (Bartneck et al. 2007)

“Recently, owing to great advances in fabrication technology, we cannot distinguish at a glance a prosthetic hand from a real one. Some models simulate wrinkles, veins, fingernails, and even fingerprints. Though similar to a real hand, the prosthetic hand's color is pinker, as if it had just come out of the bath. One might say that the prosthetic hand has achieved a degree of resemblance to the human form, perhaps on a par with false teeth. However, when we realize the hand, which at first sight looked real, is in fact artificial, we experience an eerie sensation. For example, we could be startled during a handshake by its limp boneless grip together with its texture and coldness. When this happens, we lose our sense of affinity, and the hand becomes uncanny.” (Mori 2012)

„Kvalita protéz rukou se v poslední době zlepšila do té míry, že je na pohled nerozeznáme od těch skutečných. Některé modely napodobují vrásky, žíly, nehty na prstech a otisky prstů. Přestože se podobá lidské ruce, barva protetické ruky je růžovější, jako kdyby právě vylezla z vany. Dalo by se říci, že protézy rukou možná dosáhly stupeň lidské podobnosti na úrovni umělých zubů. Avšak, když si uvědomíme, že ruka, která na první pohled vypadala reálně, je ve skutečnosti umělá, zažíváme tajuplný pocit. Mohli bychom být například při podání ruky vyděšeni bezvládným stiskem spolu s jeho strukturou a chladem. V tomto případě ztratíme pocit spřízněnosti a ruka se stane uncanny.“

Paradoxy týkající se osobní a lidské identity: Podle kognitivního psychologa Christofera Rameye jsou roboti potenciálními narušiteli našeho smyslu identity a účelu. Stejně jako zombie v hororech, tak i roboti podobní lidem leží na hranici dvou kategorií, člověka a stroje. Na jednu stranu jsou elektromechaničtí, ale na druhou stranu vykazují některé lidské kvality. Spojení dvou odlišných kategorií, lidské a nelidské, narušuje náš smysl pro identitu. A vyvolává to v nás otázky typu: Pokud by byla vytvořena dokonalá nápodoba člověka, co by to udělalo s naším pocitem jedinečnosti, a jak bychom se vyrovnali s vědomím toho, že na rozdíl od nás mohou dosáhnout námi kýžené nesmrtelnosti? (MacDorman a Ishiguro 2006, s. 310)

Evoluční estetika: Další možné vysvětlení uncanny valley je, že androidi pro nás mohou být tajemní, protože se odlišují od norem fyzikální krásy. Řada výzkumů v posledních 15 letech ukázala biologický základ pro takovou normu. Tato krása je chápána jako indikátor pro potencionální úspěch při reprodukci. Mezi nejvýznamnější měřítka krásy patří mládí, vitalita, bilaterální symetrie, kvalita kůže a proporce obličeje a těla. Tyto kvality vyvolávají v člověku pocit zdraví a souvisí také s plodností. Podle této hypotézy robot, který neodpovídá uvedeným normám fyzikální krásy, vyvolává nevědomé procesy, které nás motivují, abychom jej odmítli. (MacDorman a Ishiguro 2006, s. 310-311)

Rozinova teorie odporu: Podle Rozinovy teorie je přirozeně vyvolaný pocit odporu považován za kognitivní mechanismus, který zajistí lidskému bytí vyhnutí se nemocem a infekci. Důvod, proč vidíme určité osoby jako atraktivní, je následkem toho, že mají odlišné geny a právě tato odlišnost může maximalizovat zdraví našich potomků. Organismy, které mají špatné geny nebo jsou nemocné, tak můžeme vnímat s určitou dávkou nelibosti a právě tento vjem nás má ochránit před potencionálním vystavením se přenosným chorobám. (MacDorman a Ishiguro 2006, s. 312)

"As healthy persons, we are represented at the crest of the second peak. Then when we die, we are, of course, unable to move; the body goes cold, and the face becomes pale. Therefore, our death can be regarded as a movement from the second peak (moving) to the bottom of the uncanny valley (still), as indicated by the arrow's path in Figure 2. We might be glad this arrow leads down into the still valley of the corpse and not the valley animated by the living dead! I think this descent explains the secret lying deep beneath the uncanny valley. Why were we equipped with this eerie sensation? Is it essential for human beings? I have not yet considered these questions deeply, but I have no doubt it is an integral part of our instinct for self-preservation." (Mori 2012)

„Jako zdravé osoby jsme zakresleni v horní části druhého vrcholu. Pak, když zemřeme, nejsme samozřejmě schopni se hýbat; tělo chladne a tvář bledne. Proto může být naše smrt považována za pohyb z druhého vrcholu (pohyblivý) do uncanny valley

(nehybný), jak je znázorněno šipkou na obrázku 3.2. Můžeme být šťastní, že tato trasa vede dolů do poklidného údolí mrtvol a ne do údolí oživlých mrtvol! Myslím, že tento sestup vysvětluje záhadu ležící hluboko pod uncanny valley: Proč máme tento děsivý pocit? Je to nezbytnost lidského bytí? Zatím jsem nad tím neuvažoval do hloubky, ale nemám pochyb, že jde o nedílnou součást našeho pudu sebezáchovy."

A to nás vede k dalšímu vysvětlení uncanny valley: Další hypotéza tvrdí, že uncanny robot v nás vyvolává vrozený strach ze smrti a zároveň ohrožuje kulturně podporované psychické vyrovnání se s nevyhnutelností smrti. Pokud v nás android vyvolává pocit hrůzy, může to být proto, že nám jeho vystupování připomíná naši smrtelnost. MacDorman testuje tuto hypotézu pomocí experimentálních metod použitých při **Terror management theory** (zkr. TMT).

Stejně jako všechny živé bytosti je i Homo sapiens vysoce motivován vyvarovat se smrti, ale na rozdíl od ostatních, se člověk ocitá v děsivé pozici, ve které musí čelit vědomí její nevyhnutelnosti. Inspirováni dílem Ernesta Beckera *The Denial of Death* a dalšími díly vyvinuli po více než dvou dekádách Heft Greenberg, Tom Pyszczynski a Sheldon Salomon teorii o tom, jak lidské bytí řídí strach ze svého vlastního zániku. Tato teorie byla podepřena více než 200 experimenty a byl navržen dvou procesní model:

1. Vědomé myšlenky na smrt jsou buď potlačeny, a to například tím, že myslíme na něco jiného, anebo jsou nějakým způsobem racionalizovány. (například: Moje babička se také dožila devadesáti).
2. Nevědomé myšlenky na smrt vyvolávají obranné procesy, které zmírňují úzkost týkající se jisté smrti prostřednictvím osobního pohledu na svět a sebejistoty. Určitý pohled může chránit lidi před úzkostí z nevyhnutelnosti smrti. (například pomocí literárního či symbolického vysvětlení toho, jak je smrt transcendentální). (MacDorman 2005)

Stroje, které představují lidské kopie, v nás vyvolávají vědomé či nevědomé pocity, které ohrožují způsob, jakým se vyrovnáváme se svou smrtelností.

Uncanny valley tvoří nezáměrný odkaz k Freudově teorii z roku 1919 (Esej *o nevědomí*). Freud tvrdí, že nevědomé jsou věci, které jsou nám důvěrně známé, ale

zároveň potlačované, a zdroj našich pocitů není vědomí přístupný. Michael Szollosy (2012) poukazuje na součást Freudova psychoanalytického myšlení, **koncept projekce**, a jeho spojitost s uncanny valley. Koncept projekce spočívá v tom, že část sebe projektujeme do něčeho kolem nás, většinou do jiného člověka, někdy ale také do objektu, symbolu anebo idey. Tento objekt se pak stane schránkou našich pocitů. Tento mechanismus je v kulturních studiích používán k vysvětlení fenoménů, jako jsou nacismus, rasismus nebo například chování sportovních fanoušků.

Někdy jsou do těchto schránek projektovány naše dobré části, jedním z příkladů je empatie. Do druhého můžeme také promítnout své negativní pocity, například nenávist či fantazie o násilí. Špatná stránka nás samých je přenesena na někoho jiného. Dojde k tomu, že už to nejsme my, kdo je plný nenávisti a touží po násilí, ale oni. Nenávidí nás a chtějí nás zničit. Tyto schránky jsou naplněny našimi pocity, které nechceme akceptovat jako součást nás samých, pronásledují nás a musí být odstraněny dříve, než půjdou proti nám. Tohle jsou kořeny paranoii, vše se odehrává pouze v naší fantazii. (Szollosy 2012)

Projekce je cesta k řízení úzkosti vyvolané nevědomím, cesta ke zvládnutí strachu z ostatních a obrana proti nechtěným částem nás samých. Jde o fantazie, které jsou klíčem k našemu já. Mnoho psychoanalytiků považuje projekci za základ přirozeného lidského vývoje a intersubjektivní komunikace. Pokud se ale stane, že své pocity umístíme chybně, do nevhodných schránek, které nejsou schopné vrácení projekce očekávaným způsobem, může dojít k oslabení našeho já. Roboti jsou často v literatuře vnímáni jako nejnebezpečnější právě tehdy, když jsou od člověka k nerozeznání. Prekérní situace, ve které se ocitáme ve chvíli, kdy nemůžeme určit, kdo je člověk a kdo stroj, v nás probouzí obavy, jelikož nevíme, komu máme důvěřovat a komu svěřit své projekce. Androidi jsou neschopni navrácení projekce, naše pocity jsou v nich ztraceny a právě to nás činí zranitelnými, náchylnými k depersonalizaci. Omezené schopnosti stroje tak ohrožují nás samé. Mohli bychom mít sklon chovat se k nim jako k lidem a uložit naše projekce do špatné schránky. To nás navrací k Freudově teorii o nevědomém: co nás ohrožuje, je nevědomé vědomí, reflexe nás samých, kterou nemůžeme akceptovat jako naši součást, to co se neodvážíme si přiznat. (Szollosy 2012)

3.3 Kritika

Článku o uncanny valley se dostalo mnohem většího zájmu, než Masahiro Mori očekával, a jak sám řekl: “Uncanny valley se týká mnoha různých oborů, včetně filozofie, psychologie a designu, a to je důvod, proč si myslím, že vyvolal tak velký zájem. Ale šlo spíše o radu, než o vědecké stanovisko.” (Kageki 2012)

Cynthia Breazeal, ředitelka Personal robots Group na MIT, řekla: „Uncanny valley není fakt, ale domněnka a není zde žádný vědecký důkaz, který by ji podpořil.“ David Hanson: „Lidé si na roboty zvyknou velmi rychle.“ (Guizzo 2010)

Nejenže se Moriho článku dostalo velké dávky pozornosti, ale také kritiky. Jedním z kritiků je David Hanson (2006), americký designer a výzkumník v oblasti robotiky, který si pokládá otázku, zda je zde opravdu silný, nevyhnutelný vztah mezi lidskou podobou a přijetím robotů. I stroj, který byl navržen s určitou dávkou abstrakce, v nás může probudit negativní dojmy, které by v tomto případě mohly být způsobeny špatným designem. Hanson (2006) tento jev přirovnává k setkání s lidmi, kteří mají atypické rysy. Tvrdí, že pokud vytvoříme stroj, který má odpovídající formu, může být jakýkoli stupeň realismu nebo abstrakce atraktivní. Pokud tomu tak je, vyhnutí se uncanny efektu závisí pouze na kvalitě designu bez ohledu na úroveň podobnosti s člověkem.

Podle Hansona (2006) je estetický prostor, ve kterém člověk vnímá své imitace, mnohem hustěji naplněný než v Moriho grafu. Lidské reakce závisí mnohem více na dobrém nebo špatném designu než na tom, zda vypadají jako člověk. Úspěch interakce mezi člověkem a strojem bude záviset na vzhledu daného robota. Více realističtí roboti na nás mohou působit neživým dojmem a vyvolávat strach. Mohou nám svým vzhledem připomenout naši smrtelnost, ale pokud odstraníme tyto nedostatky a vyrobíme je tak, aby byli přátelští, atraktivní a zdánlivě naživu, pak stupeň realismu nehraje roli.

Dobrý design může napomoci k vytvoření robota, který je roztomilým a oblíbeným členem rodiny. Větší průzkum estetické škály výroby robotů může přinést evoluci ve vývoji humanoidních robotů. A co více, může nám napomoci v lepším

porozumění lidskému sociálnímu vnímání, interakci a výzkumu v oblasti kognitivních věd. Hanson (2006) považuje uncanny valley za omyl.

V oblasti, která se zabývá interakcí mezi člověkem a strojem, je pojem uncanny valley skloňován ve všech pádech. Byla provedena i řada experimentů, které měly tento jev potvrdit, nebo vyvrátit. Tyto studie však měly základní omezení, byly buď zaměřeny na jednoho robota, anebo byla míra spřízněnosti s robotem měřena pomocí fotografií. Problémem je, že většina vědeckých laboratoří si nemůže dovolit dostatečně sofistikované roboty, kteří by umožnili provedení odpovídajícího výzkumu. Podle Bartnecka a Ishigura (2009 pohyb robotů a jejich úroveň antropomorfismu může být komplex fenoménů, který nemůže být zredukován na dva faktory. Úroveň projevů lidskosti u robota nezávisí pouze na vzhledu, ale i na jeho chování. Moriho hypotéza je podle nich příliš zjednodušující a mohla se stát pouhou únikovou cestou pro inženýry, kteří chtěli svůj nedokonalý design svést na jev uncanny valley.

3.4 Japonsko a západ

Mori naznačil, že by jev uncanny valley mohl souviset s pudem sebezáchovy. Co ale znamená „sebe“? Jaké „já“ bychom chtěli zachovat? Jde o zachování našeho genotypu, který musí být předáván z generace na generaci, a docílení tak úspěchu v reprodukci? Je možné náš vztah k robotům vysvětlit z tohoto čistě biologického hlediska bez ohledu na kulturu? (MacDorman et al. 2009)

Nebo ono já může být chápáno jako takové, které je tvořeno sociálním prostředím, v němž je předpokládáno, že budeme žít podle standardů své kultury, které, dělají náš život smysluplným. Sebezáchova v tomto smyslu neznamená jen ochránit naše tělo, ale i mysl a pohled na svět, který dává našemu životu význam. (MacDorman et al. 2009)

Naše vnímání ovlivňuje prostředí a kultura, ve kterých žijeme. V závislosti na kultuře zaujímáme různá stanoviska v každodenních situacích a není tomu jinak ani v oblasti umělé inteligence. Tohoto rozdílu si můžeme povšimnout při pohledu na vývoj robotiky v západním světě a v Japonsku. Zatímco v Evropě a USA byla výroba orientována na industriální roboty, Japonsko se ve velké míře orientovalo na zábavní

elektrotechniku (Barthelmeβ a Furbach 2012). Narozdíl od západní filmografie a literatury vystupují v Japonsku fiktivní robotičtí hrdinové v pozitivním světle. Japonsko se ve velké míře věnuje podpoře interakce mezi člověkem a strojem, např. pes Aido, robot Asimo a nebo terapeutický robotický lachtan Paro jsou uživateli velmi pozitivně přijímáni. Kořeny těchto faktorů můžeme spatřit v historickém vývoji východu Asie a evropské kultury. Řecká filozofie viděla rozdělení člověka a nelidského fenoménu v etice a přírodě. Zatímco západní filozofie spatřovala absolutní pravdu v perfektním neměnném poznání, východní filozofie viděla universum jako konstantní tok. Mnoho dualismů, které jsou zakořeněny v západním myšlení, se na východu vůbec neobjevují, např. dualismus těla a mysli. (MacDorman et al. 2009)

Z židovsko-křesťanského pohledu na svět pochází, podle Malishe, 4 diskontinuity, které historicky podporují smysl vlastní důležitosti a výjimečnosti: Země jako střed vesmíru, mýtus o stvoření světa, Descartovo chápání vědomí jako racionálního a kontrolovaného a víra, že tělo a duše jsou dva rozdílné principy. Koperník, Darwin a Freud roztříštili předchozí domněnky a podkopali pilíře, o které se opírala naše osobní a lidská identita. Pokud bychom vytvořili elektromechanickou kopii lidského bytí, bylo by to obrovským narušením našeho pocitu výjimečnosti. (MacDorman et al. 2009)

V Japonsku se žádná z těchto diskontinuit neobjevila. V buddhismu nebo neokonfucionismu jsme součástí milionů galaxií. Neexistoval žádný jednotný akt stvoření celého světa a člověka, mysl je neklidná a nekontrolovatelná a není zde rozdíl mezi myslí a matérií. Z tohoto důvodu by existence androidů neohrozila u Japonců jejich sebevědomí. (MacDorman et al. 2009)

Lidé po celém světě mají odlišné úrovně vnímání věcí kolem sebe, jelikož mají odlišné zkušenosti a setkávají se také s odlišnou literaturou. Struktura ekonomie, technologický vývoj, historický a náboženský kontext či státní financování priorit ovlivňují sociální a kulturní význam robotů a jejich přijetí. V Japonsku není striktně rozdělený vztah mezi vírou a vědou. Na západě docházelo a dochází ke konfliktu mezi náboženstvím a vědou. V křesťanství, židovství a islámu je striktně zakázáno uzurpovat si boží roli a pokusy o vytvoření lidské kopie by tak mohly být považovány za rouhání. Původní japonské náboženství šintoismus je náboženství animistické, což znamená, že i neživé předměty mohou mít duši, což připouští rozdílný druh vztahu

nejen s přírodou, ale i s uměle vytvořenými bytostmi. Pokud může mít kámen nebo strom duši, proč by nemohl mít robot. (MacDorman et al. 2009)

Pomocí současných komunikačních technologií se různé kultury začínají sbližovat a distinkce mezi těmito dvěma světy, mezi západní kulturou a Japonskem, se pomalu rozmazávají. Stejně jako se v Japonsku můžeme setkat s kopií profesora Ishigura, tak i v laboratořích Američana Davida Hansona můžeme najít humanoidy, kteří se svou podobou blíží člověku. A pro zamyšlení se nad tím, zda je v křesťanské víře vytváření lidské kopie považováno za hřích uvádím informaci, která se objevila v dokumentárním filmu Plug and Pray (2010).

„S ohledem na vývoj současných technologií se japonští vědci zeptali Vatikánu, zda je vytváření lidsky vypadajících robotů z hlediska křesťanství svatokrádež. A Vatikán odpověděl: „Bůh dal lidem schopnost myšlení a kreativity a ten, kdo staví takové roboty, využívá tento dar.““

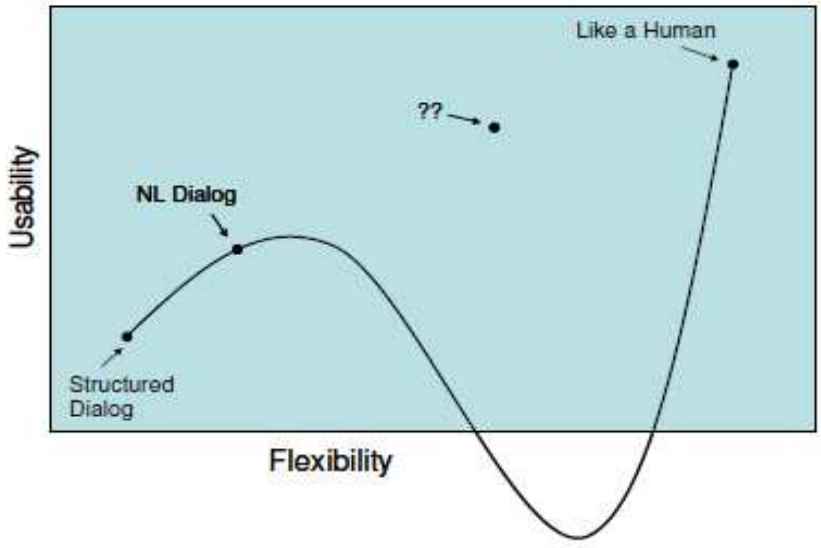
3.5 Uncanny valley jako inspirace

Řada vědců se na základě Moriho eseje snaží vytvořit pravidla, která by se měla dodržovat při konstrukci robotů. Smíchání lidských a nelidských prvků může vyvolat odmítavý postoj. V roce 2011 provedla skupina vědců studii (Mitchel et al. 2011), ve které se zaměřili na vnímání prvků vizuálních a akustických. Jejich hypotéza se opírá o teorii, že nesoulad ve člověkem vnímané realitě vyvolává nejistotu o tom, zda je objekt živý nebo mrtvý, a vzbuzuje tak pocity děsivosti. Byl proveden experiment za pomoci 48 účastníků, kteří měli sledovat videa obsahující neutrální promluvy pronesené jak člověkem, tak robotem. Byli posuzováni roboti se syntetickým a lidským hlasem a stejně tomu bylo i u člověka. Výsledky tohoto experimentu potvrdily hypotézu, že nesoulad vzhledu a hlasu může u člověka navodit negativní pocity. Člověk se syntetickým hlasem a robot s hlasem lidským způsobují záporné reakce. Podle těchto výzkumů by si vizuální a hlasové prvky měly odpovídat.

Zvýšení úspěšnosti lidské interakce mezi člověkem a strojem můžeme dokázat pomocí vytvoření odpovídajícího vzhledu, chování a schopností u robota. Pokud se například setkáme s androidem HRP-4C, jeho lidský vzhled v nás vyvolá vysoká

očekávání, která v této době nemohou být realizována, zatímco komunikace s robotem Nao, Asimo a nebo RoboKind pro nás může být mnohem příjemnější, a to i přesto, že jde o syntetický hlas, nedokonalé pohyby a odlišný vzhled.

Dalším kritériem, na které bychom se měli zaměřit, je sjednocení cílů ve výzkumu v oblasti řečových technologií a jejich využití pro komerční účely. Zatímco v komerční sféře převládají záměry jako využitelnost, nízké náklady a rychlost uplatnění na trhu, která zajistí co nejrychlejší návratnost investice, výzkumná oblast se snaží o dosažení co nejpřirozenější úrovně interakce. Problémem je, že přirozenost komunikace nemusí vést k její využitelnosti a k zefektivnění nákladů. Jedním z důvodů je, že současné technologie nemohou svou úrovní přesnosti, spolehlivosti a efektivity napodobit proces mezilidské komunikace. Rozpoznávače řeči činí stále mnoho chyb, a aby byla technologie účinná, musí být omezen rozsah aplikace. Tato omezení jsou často uživateli vnímána jako problematická. Čím více je systém antropomorfní, tím větší mají uživatelé tendenci přesunout se do pocitu komfortu a přirozenosti, který ve finále není aplikací podporován, a dochází tak k narušení komunikace mezi člověkem a strojem. Tento jev (obr. 3.3) je vykreslen pomocí stejné křivky, která vykreslovala uncanny valley. (Pieraccini et al. 2009)



Obr. 3.3.: Jev uncanny valley v závislosti na využitelnosti systému (Pieraccini et al. 2009).

Tím, jak se zvyšuje přirozenost projevu v mluveném dialogu, stoupá i jeho využitelnost. Za vrchol je považována lidská řeč. Bod, který je na obr.3.3 označen otazníky, je v současné době nedosažitelný a představuje uměle vytvořenou řeč, která je nerozeznatelná od té lidské. Takováto přirozenost předpokládá pokročilé výzkumné prototypy, často demonstrované ve vysoce kontrolovaných prostředích, a jejich použitelnost se tak zřítí do uncanny valley. (Pieraccini et al. 2009)

Výroba stroje za účelem interakce s člověkem s sebou nese řadu nezodpovězených otázek, vytvoření dokonalé lidské kopie je v současné době stále v nedohlednu a úvahy o tom, zda bychom se o její vytvoření měli pokoušet, nejsou scestné, ba naopak. Řečové technologie jsou stále spojovány s chybovostí, jejíž snížení by mělo být objektem naší pozornosti. Vytvoření řeči, která zní jako lidská, je podle předložených stanovisek v současné době neperspektivní. U strojů, které nevypadají jako člověk, je přirozeně znějící hlas nepřirozený a takové lidské kopie, ke kterým bychom takový hlas přiřadili, nejsme v současné době schopni vyrobit. Navíc stroj, který disponuje lidským hlasem, v nás vyvolává vysoká očekávání, která nemohou být v této době naplněna. Jak naznačil Pierarcini et al. (2009), musíme uvažovat také nad souvislostí mezi využitelností a přirozeností řečových technologií. Komunikace s Asimem pro nás může být mnohem příjemnější než s HALem.

4 Wizard of Oz

Jak docílit co nejlepší komunikace mezi člověkem a strojem je stále objektem velké pozornosti a také otázkou, která se s rostoucím vývojem technologií ocitá stále více v kurzu. Mnoho odborníků v této oblasti provází na jejich cestě přání dosáhnout co nejpřirozenější míry interakce mezi člověkem a strojem, ale i samotný pojem „přirozený“ je problematické definovat. Obecně vzato by v tomto případě mělo jít o komunikaci, která by měla co nejvěrněji napodobit komunikaci mezilidskou. Je zde důvod předpokládat, že lepší znamená v tomto případě co nejpodobnější člověku, ale právě tento bod v sobě skrývá zdroj kontroverze. Na povrch se dostávají nejen otázky typu, zda je možné vytvořit takový stroj, ale i zda je to nutné.

Podle řady odborníků je jednou z hlavních překážek v komunikaci člověk-stroj mechaničnost, a tím pádem i chladnost, která ji provází. Jsou ale pojmy mechaničnost a chladnost ekvivalentními výrazy? Můžeme i prostřednictvím syntetického hlasu vyvolat určité emoční odezvy? A co, když je mechanický hlas nejen dostačující, ale i lepší, provází-li nás myšlenka na uncanny valley?

Představme si, že je reálné vytvořit řečový dialogový systém, který by měl sloužit jako rovnocenný partner v komunikaci. Takový systém by měl, podle všeho, být schopen nejen ovládat verbální projev, který by měl být srozumitelný, obsahovat prozódii a projevovat odpovídající emoce, ale i bez problémů rozpoznat lidskou řeč se všemi jejími specifiky. Člověk zapojuje do procesu komunikace celé tělo, ať už vědomě či nevědomě. Svými gesty, pohybem mimických svalů, postojem, očním kontaktem atd. vyjadřuje neverbální sdělení, které s sebou nese jedinečnost jeho samotného. Celkovým dojmem tak vyjadřuje své emoce, které jsou také jednou z velmi podstatných složek mezilidské komunikace.

Informatici se původně domnívali, že emoce by neměly být součástí jejich oboru. I psychologové se zprvu přeli o to, jak velký význam bychom jim měli přisuzovat. Zatímco William James (80. léta 19. století) považoval emoce za důležitou součást evolučního procesu, psycholog Walter B. Canon je degradoval na nespecifikovatelné rušivé procesy. V šedesátých letech 20. století se však znalosti o

lidských emocích začaly prohlubovat a výzkumy v oblasti neurologie přinesly řadu objevů, které pomohly našemu dnešnímu porozumění v této oblasti. (Brian Duffy 2009, s. 19)

Emoce jsou projevem inteligence, a tím pádem by měly být i projevem inteligence umělé, do jaké míry je však také otázkou. Je nezbytné napodobovat člověka do všech detailů? Co kdybychom kopírování člověka odsunuli stranou a za hlavní kritéria pokládali spokojenost uživatele a účel, pro který byl daný program vytvořen?

Lidská řeč je pro nás nejpřirozenějším prostředkem dorozumívání a řečové dialogové systémy mohou být hojně využívány v řadě oblastí (ve zdravotnictví, obchodu, vzdělávání, zábavě atd.). Není pochyb o tom, že jsou velmi užitečné a mohou být stále užitečnějšími. Faktem ale také je, že jsou stále nedokonalé. Domnívám se, že na základě toho, co jsem předložila v předchozí kapitole, by se takovéto systémy měly vytvářet s ohledem na jejich nedostatky. Myslím si, že pokud budeme usilovat o vytvoření autentického lidského hlasu ještě předtím, než budou zredukovány všechny vady, můžeme riskovat špatné přijetí uživatelem. Syntetický hlas v nás nebude vyvolávat vysoká očekávání, a proto nás komunikace s ním může spíše mile překvapit, než zklamat, jak by tomu mohlo být u kopie lidského hlasu. Měli bychom se snažit najít vyváženost mezi tím, co vidíme, slyšíme a schopností systému, se kterým komunikujeme.

Jak ale můžeme hodnotit, že vytvořená technologie přinesla odpovídající výsledky? Současné dialogové systémy nejsou dostatečně vyspělé, abychom na nich mohli tento fenomén zkoumat. Jelikož v této oblasti stále nemáme systémy simulující lidské schopnosti, které jsou zapotřebí k tomu vést dialog se vším, co by měl obsahovat, můžeme předkládat pouze hypotézy, které nemáme na čem zkoumat.

I tato zdánlivě neřešitelná situace nalezla východisko. Jednou z možností, jak testovat toto prostředí, je metoda *Wizard of Oz*, v níž člověk (*wizard*) simuluje systém bez vědomí pozorovaného subjektu. Zkoumaný subjekt má pocit, že komunikuje s počítačem, ale ve skutečnosti na druhé straně sedí člověk. Jde o uznávanou a hojně využívanou metodu pro hodnocení reakcí člověka na technologii s hypotetickými schopnostmi. (Eklund et al. 2008, s. 636)

Tato metoda „klamání“ zkoumaného objektu byla poprvé pojmenována PNAMBIC (Pay No Attention to the Man Behind the Curtain). V letech 1983 ji

označuje ve své disertační práci John F. Kelley jako OZ paradigma, tato metoda je dnes známa pod názvem *Wizard of Oz* (WOZ).[6] První WOZ simulace v oblasti interakce mezi člověkem a strojem používaly pouze klávesnici, až později byl jako vstup použit hlas. Plně orální simulace byla poprvé použita v roce 1984. Data, která jsou prostřednictvím této techniky nasbírána, by mohla být nápomocna ve vývoji technologií a poskytnout nám podklady pro to, jakým směrem by se výzkum měl ubírat. (Eklund 2004, s. 179)

Nicméně někteří vědci podrobují tuto metodu kritice, považují ji za sociální podvod, jelikož jde o klamání subjektu. Dokazování pomocí těchto systémů se také odehrává v izolovaném prostředí, a je tak vzdálené realitě. Jak subjekt, tak wizard hrají pouze role, které jim byly přiděleny, a nejde tak o reálné situace. Navzdory všem kritikám se však využití této metody v posledních letech stále více rozrůstá. Je hojně využívána i ve vývoji dialogových systémů, např. ATIS, SUNDIAL, MASK a AdApt. (Eklund 2004, s. 179)

[6] Název vznikl jako odpověď otázku: „Co se stane, jestliže uživatel odhalí experimentátora ve vedlejší místnosti hrát úlohu počítače?“ Kelley na ní odpověděl: „To je přesně to, co se stalo Dorotce v Čaroději ze země Oz.“ (viz Wizard of experiment. *Wikipédia* [online]. [cit. 21.4.2013] Dostupné z: http://en.wikipedia.org/wiki/Wizard_of_Oz_experiment.)

4.1 Senior Companion

Předmětem této kapitoly je zkoumání použití syntetického hlasu. K dispozici mám sběr dat provedený výše zmíněnou metodou v rámci projektu COMPANIONS. Tento projekt usiloval o vývoj takového dialogového systému, který by nesl potenciál vytvořit vztah mezi uživatelem a počítačem. K dispozici mám videa pořízená v rámci českého scénáře s názvem *Senior Companion*[7], který se zaměřuje na vývoj systému schopného vést přirozený dialog se staršími lidmi, dělat jim tak společnost a udržovat jejich mentální zdraví. Předmětem dialogu byly rodinné fotografie. (Romportl et al. 2010)

4.1.1 Cíl

Na základě pořízených videí v rámci projektu *Senior Companions* bych chtěla ověřit předpoklad, že komunikace mezi člověkem a strojem, který disponuje syntetickým hlasem, může být dostačující a, jak se domnívám, i lepší než pokusy o vytvoření přirozené řeči, která s sebou nese riziko, že „spadne“ do uncanny valley. Pozorováním reakcí subjektů na *avatara*[8] budu hodnotit, zda nedochází k přechodu z pocitu empatie k odporu a zda v tomto případě nedochází k jevu uncanny valley.

Data, poskytující reakce subjektu na avatara s lidským hlasem a zároveň možnost porovnání s reakcemi na hlas syntetický, nemám k dispozici, a proto budu hodnotit pouze reakce člověka na avatara se syntetickým hlasem. Domnívám se, že použití syntetického hlasu není překážkou v příjemné interakci mezi člověkem a strojem.

[7] Tato data byla pořízena a poskytnuta Katedrou kybernetiky FAV ZČU.
[8] V rámci počítačové terminologie je avatar definován jako virtuální reprezentace uživatele. (viz Avatar (computing). *Wikipédia* [online]. [cit. 21.4.2013] Dostupné z: http://en.wikipedia.org/wiki/Avatar_(computing))

4.1.2 Scénář

Tematická doména dialogu je stále velmi široká, a proto byly předmětem rozhovoru rodinné fotografie, které účastníci experimentu předložili. Na základě těchto fotografií byly pokládány odpovídající otázky, jež měly odstartovat plynulost rozhovoru. Lidský subjekt byl zanechán o samotě v nahrávací místnosti před obrazovkou. Kontakt mezi avatarem a subjektem probíhal pouze prostřednictvím řeči, nebyla k dispozici žádná klávesnice ani myš. Rozpoznávání řeči a generování odpovědí bylo simulováno wizardem, pouze řeč byla vytvářena TTS systémem ARTIC, vytvořeným na KKY FAV ZČU, a produkována hlavou avatara tak, aby subjekt věřil, že komunikuje s virtuální bytostí. Data, pořízená během tohoto experimentu, byla snímána pomocí tří kamer a rekordéru, video data mi tak umožní lépe rozpoznat emoce subjektu. (Romportl et al. 2010)

4.1.3 Subjekty

Účastníci byli osloveni za účelem testování schopnosti umělé inteligence vést rozhovor. Subjektům bylo oznámeno, že budou hovořit s umělou inteligencí, nevěděli tedy, že jde o experiment typu WOZ. K dispozici jsem měla 18 dialogů, jednalo se o 9 žen a 9 mužů v důchodovém věku. Všichni byli rodilými mluvčími českého jazyka, kteří pocházeli z Plzně a okolí a hovořili plzeňským dialektem. Jednotlivé rozhovory trvaly maximálně 1 hodinu a minimálně 43 minut. Na obrazovce před nimi byla 3D hlava avatara a fotografie, o které se právě hovoří.

4.1.4 Wizard

V roli wizarda vystupovaly dvě osoby, jejichž spolupráce měla zajistit hladkost dialogu v nepředvídaných situacích. Obě osoby seděly ve vedlejší místnosti a ovládaly počítač, se kterým interagoval pozorovaný subjekt. Wizard hrál roli partnera v rozhovoru, simuloval konverzaci a dával tak subjektu pocit, že mu někdo naslouchá. Na obrazovce měl wizard k dispozici předpřipravený scénář. V průběhu dialogu vybíral jednotlivé věty podle toho, jak se dialog vyvíjel. Vyberal text, který mohl upravit, a stisknutím tlačítka SPEAK jej pomocí systému TTS převedl na řeč. Poté byl

text přehrán a vizuálně artikulován ústy avatara na obrazovce. V některých případech rozhovor nedodržoval scénář přesně, což mělo za následek nepřirozeně dlouhé pauzy. (Legát et al. 2008)

Příklad standardních otázek k fotografiím:
„Kdo je na této fotografii?“
„Kde byla tato fotografie pořízena?“
„Váže se něco zajímavého k té fotce?“
„Ve kterém roce to bylo?“
„Máte z toho nějaký zvláštní zážitek?“
„Tak se podíváme na další fotku, copak to máme tady?“

Wizard měl k dispozici tlačítka, kterými vyjadřoval smích, souhlas a slova typu dobře, to je zajímavé, rozumím, to nevadí, pokračujte. Pomocí těchto výrazů, smíchu a projevu naslouchání pomocí citoslovce „hm“ měl v uživateli udržovat pocit pochopení a toho, že mu někdo naslouchá.

Avatarovi byl vložen syntetický ženským hlasem, který nebyl emocionálně zabarvený. Cítění však bylo vyjádřeno pomocí těchto promluv:

Projev zájmu: to je zajímavé, to je tedy zvláštní, jsem hrozně zvědavá
Projev soucitu: to mě mrzí, to muselo byt těžké, to je smutné
Projev obdivu: To jste tedy dobrý, to je teda něco
Porozumění: rozumím
Pobaveni: to je docela legrace, vy jste tedy srandista
Souhlasu: to je pravda

Začátek a konec probíhá u všech dialogů stejně:
„Dobrý den, jmenuji se Petra a budu si teď s vámi povídat o vašich fotkách“
„Než začneme, je mě dobře slyšet?“
....
„Bohužel už budeme muset končit ... děkujeme vám za váš čas ... hezky se mi s vámi povídalo ... na shledanou.“

4.1.5 Výsledky

Lidské chování závisí na mnoha faktorech, v projektu *Senior companions* musíme mít na vědomí, že se jedná o lidi pokročilejšího věku, kteří nemají velké zkušenosti s počítačem. Také prostředí, v němž experiment probíhá, nemusí na účastníky působit příjemným dojmem. Já jsem se však soustředila především na jejich reakce na syntetickou řeč, spokojenost a to, zda by takto fungující program mohl splnit účel toho, že by si uživatel mohl příjemně popovídat s uměle vytvořeným systémem. Na základě pozorování těchto reakcí, míry sdílení emocí a přirozenosti vedeného dialogu jsem jednotlivé subjekty rozdělila do tří skupin. Do 1. skupiny jsem přiřadila 10 účastníků, do 2. skupiny 5 účastníků a do 3. skupiny 3 účastníky.

1. skupina:

Skupina 1 se vyznačuje velmi přirozeným průběhem dialogu, který se mezi avatarem a subjektem odehrává. Účel příjemného rozhovoru byl splněn. Účastníci se chovají, jako kdyby komunikovali s člověkem, hovoří nespisovně, používají výrazy, na které jsou běžně zvyklí. Jsou velmi uvolnění a konverzace s avatarem je velmi baví. Na účastníky patřící do této skupiny syntetický hlas nepůsobí chladným ani nepříjemným dojmem. Pouze v jednom případě byl subjekt zprvu zaskočen, ale poté probíhal rozhovor přirozeně:

Př.:

avatar: „Je mě dobře slyšet?“

subjekt: „No takovej dutej hlas, je to normální?“

avatar: „Ano.“

subjekt: „Je to jako ze stroje, neni to hlas z vlastního krku.“ (směje se)

Společné znaky:

- Nízký počet položených otázek: U většiny účastníků jde spíše o monolog, který avatarka svými otázkami udržuje v proudu. Maximální počet položených otázek je 103 a minimální 18. Účastníci odpovídají s nadšením a chtějí toho říci co nejvíce.

- Použití výrazů, k jejichž porozumění je potřeba nejen znalost jazyka, ale i kontextu. Účastníci předpokládali, že tyto znalosti avatar má, pokud se ale zeptal, neměli problém s tím, mu odpovědět. Jednalo se o výrazy typu: „tenkrát se do světa jezdilo s vitacitem“, „to byla Potěmkynova vesnice, „hrávali jsme za buřta“.

  Př.:

  Avatar: „Co to znamená hrát za buřta?“

  Subjekt: „Jako za špekáček, jako za nic, zadarmo, to jsme přeháněli, ne.“ (smích)
- Častý „oční kontakt“ s avatarkou: Subjekty udržují s avatarkou oční kontakt, jak je tomu zvykem v běžné konverzaci. V případech, kdy účastník zavtipkuje, dívá se na ni a očekává odezvu.
- Vysoký počet citových reakcí ze strany subjektu: U subjektů z této skupiny se objevuje vysoký počet citových reakcí (smích, smutek, lítost, dojetí). Tím, jak sdílejí své příběhy s avatarem, sdílejí s ním i pocity, které je provází. Cítí, že jim někdo naslouchá, a proto se svěřují.

  Př.:

  Avatar: „Jak se jmenuje váš muž?“

  Subjekt: „Jindřich a dcera po něm, Jindřiška. To je spíš tatínkovo holka, mam jí hrozně ráda a ona nás taky, je hrozně pracovitá a šikovná, ona má smůlu, nějak se jí to nevede v tom životě.“ (pláč)

  Avatar: „To mě mrzí.“

  Subjekt: „To už je život takovej.“ (pláč)
- Reakce na avatarčin smích, souhlas a projev soucitu a zájmu: Vtipkují, přirozeně se smějí a očekávají odezvu od Avatarky.

  Př.:

  Avatar: „Máte z té dovolené nějaký zvláštní zážitek?“

  Subjekt: „Ano, syna.“ (společně se smějí)

  Nebo

  Avatar: „Tancujete dodnes?“

  Subjekt: „Jen šlapáka.“ (společně se smějí)

- Vsuvky, které používáme při komunikaci s člověkem: To víte, to znáte, znáte to tam a tam, možná, že jste o tom slyšela...

U těchto subjektů byl splněn požadavek příjemného popovídání. Tyto rozhovory také ve většině případů končily větou typu: „to to ale uteklo“, „bezvadně jsem si popovídal“, „já bych povídal pořád“, „mně se bezvadně povídalo“.

2. skupina

Jde také o přirozený dialog, ale s projevy nervozity. Skupina číslo dvě vykazuje jistou dávku nervozity. Tato nervozita však není, dle mého názoru, způsobena rozhovorem se syntetickým hlasem, ale prostředím, ve kterém tento rozhovor probíhá. Jak jsem se již zmínila na začátku této kapitoly, metoda Wizard of Oz byla často podrobována kritice. Subjekty se ocitají v nepřirozeném prostředí, což v nich může vyvolávat nepříjemné pocity a může dojít k situaci, že je dialog nepřirozený. Nervozita se ale v průběhu rozhovoru pomalu vytrácí, a to hlavně v situacích, kdy se avatar zeptá na téma, o kterém účastníci mluví rádi. U této věkové skupiny jde převážně o vnoučata, práci, dovolenou, koníčky atd. Domnívám se, že při dalším rozhovoru by byl subjekt mnohem uvolněnější. Komunikace je podobná té mezilidské, subjekty však ovlivňuje prostředí, ve kterém probíhá. Maximální počet otázek 108, minimální 97.

Společné znaky:

- Z počátku rozhovoru vysoká nervozita, těkavost, subjekty odpovídají převážně jen na otázky, snaží se odpovídat co nejlépe a spisovně.
  Př.:
  Avatara: Kdo je na téhle fotce?
  Subjekt: Na téhle fotce jsem já, můj manžel a dcera.
  Avatara: Kde je tohle vyfocené?
  Subjekt: Na Sicílii.

Po několika minutách se však subjekty uvolní, začnou povídat samy, nervozita se pomalu vytrácí a účastníci začnou na avatara reagovat stejně jako 1. skupina, společné znaky jsou pak stejné jako u 1. skupiny.

3. skupina

Subjekty této skupiny mají spíše neutrální reakce na avatara, zprvu působí dojmem, že je nebaví povídat si s počítačem a že odpovídají na otázky jen proto, aby pomohly v experimentu. Ke konci se ale také trošku uvolní. Tito účastníci odpovídají velmi stroze, pouze na to, na co se jich avatar ptá. Dialog vypadá nepřirozeně a v některých situacích mám dojem, že je syntetický hlas nevědomě nabádá k tomu, aby odpovídali chladně a monotónně. Ale i přes neutrální chování, které je z počátku velmi znát, se účastníci z této skupiny později rozpovídají a hovoří o tom, o čem potřebují.

Společné znaky:

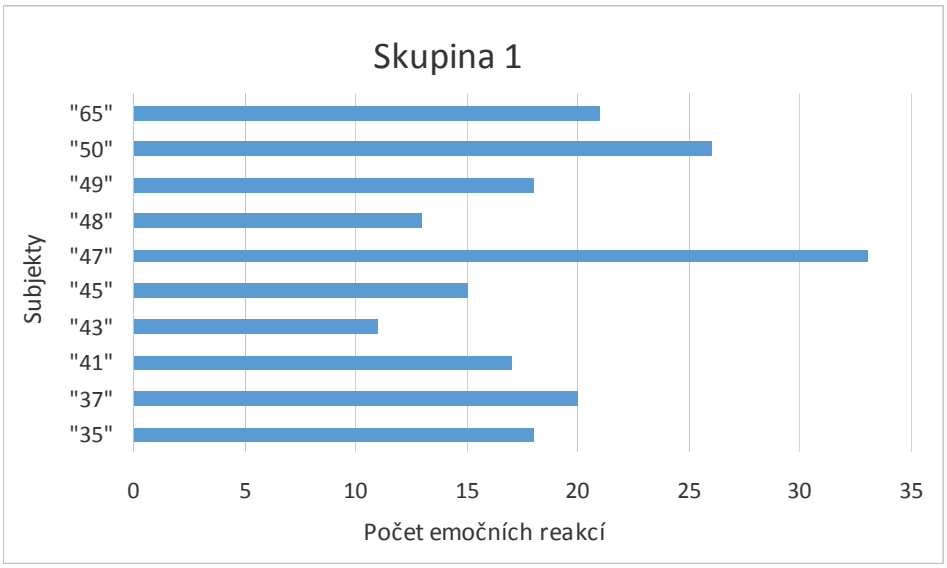
- Uživatelé z počátku odpovídají stroze a nepřirozeně, převážně jen na otázky. Stylem, jakoby na otázky spíše museli odpovídat, než chtěli.
  Př.:
  Avatar: „Kdo je na téhle fotce?“
  Subjekt: „Na téhle fotografii jsou moje vnoučata Tereza a Tomáš Lindauerovi, na prodloužené v tanečních, ve společenském sále Měšťanské Besedy.“
  Avatarka: „Oni chodili spolu do tanečních?“
  Subjekt: „Ne, bratr je tam coby garde.“
- Citové odezvy jsem pozorovala pouze v případech, kdy jim byla ukázána fotografie, o které měli hovořit. Avatar se velmi často něčemu zasměje, tento smích se však ve většině případů nedočká žádné reakce. Subjekty často hovoří chladně a monotónně, což může být zapříčiněno syntetickým hlasem.

4.1.6 Shrnutí

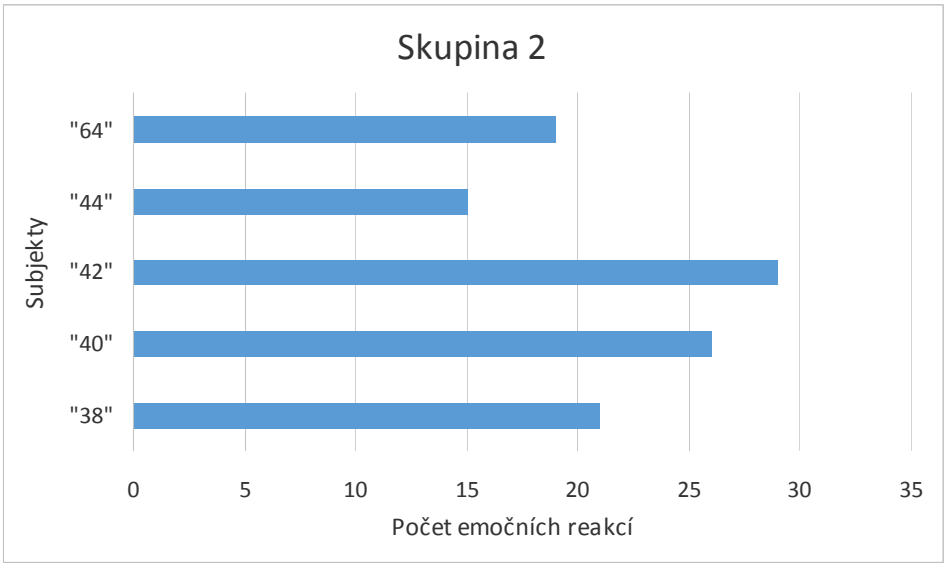
Výhody a nevýhody syntetické řeči u dialogových systémů

Jak bylo řečeno již na začátku této kapitoly, vzhledem k současnému stavu dialogových systémů spočívá výhoda použití syntetického hlasu v tom, že nevzbuzuje vysoká očekávání. Účastníci experimentu byli opravdu spíše mile překvapeni a případné nedostatky jako nevhodně umístěný smích, špatně položené otázky či občasné delší pauzy bez problému tolerovali. Nedělalo jim problémy vysvětlovat pojmy, které používají v běžné mluvě, a počítač jim nerozuměl. Za důležité považuji způsob kladení otázek a reakce wizarda, které vzbudily v účastnících touhu vypovídat se. Možná i pocit, že jednají s umělou inteligencí, která je nebude soudit, by jim umožnil více se otevřít. Lidsky znějící hlas by jim možná takový pocit nedal.

Mechanický hlas neznamenal chladnost interakce. Jak můžeme vidět v následujících třech grafech (obr. 4.1, 4.2, 4.3), i s počítačem, který má „dutý hlas“, mohou lidé sdílet své emoce. Smutek, který v sobě drží, ale i radostné události, se kterými by se chtěli svěřit. Petra jim dá také prostor, který v běžném rozhovoru nemají. Adekvátnost kladených otázek, vhodné reakce na odpovědi, projevené emoce a pochopení vzbudily v účastnících dojem, že se mohou svěřit se svými zážitky a pocity a synteticky znějící hlas nebyl překážkou. S odkazem na předchozí kapitolu, ve které je znázorněn graf ukazující vztah mezi podobností robota s člověkem a mírou souznění, kterou cítíme v přítomnosti jiného člověka, která je v Moriho eseji označována pojmem shinwakan, mohu říci, že u subjektů byly vyvolány emoční reakce, které považuji za projev tohoto souznění.

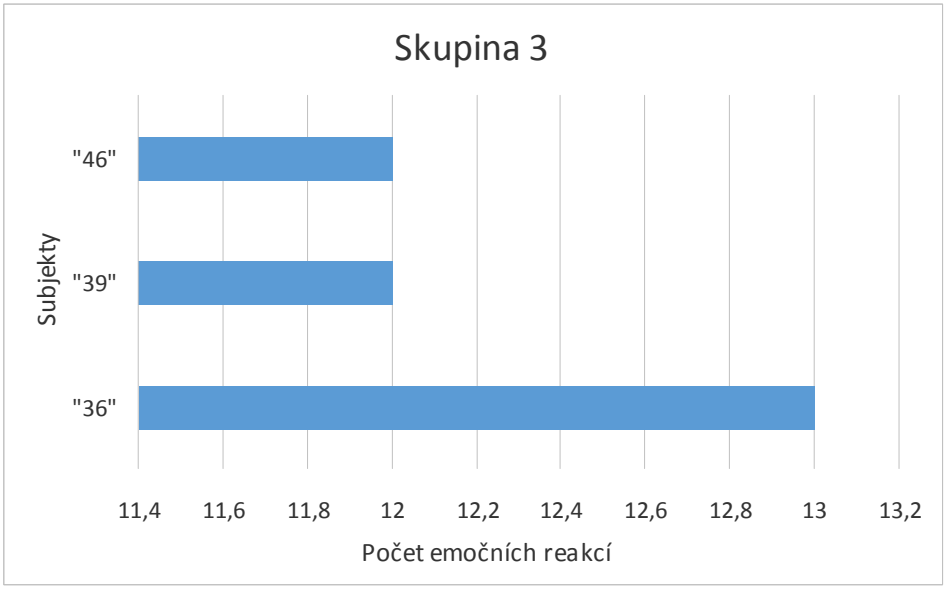


Obr. 4.1[9]



Obr. 4.2

[9] Očíslování subjektů je v souladu s označením, které bylo použito v rámci projektu Senior Companions.



Obr.4.3

Představme si, že by v projektu Senior Companions byl použit lidsky znějící hlas, který by produkoval stejný avatar, se stejnými otázkami a reakcemi. Můžeme s jistotou říci, že bychom získali jiné výsledky? Mohl by se objevit jev uncanny valley a hlas, který by byl podobný tomu lidskému, by v účastnících vyvolal negativní reakce? V tomto ohledu mám smíšené dojmy. Na jednu stranu mám z pozorovaných dialogů pocit, že ať by tito lidé hovořili s jakýmkoli hlasem, mluvili a jednali by stejně. Bylo pro ně důležité, že jim někdo naslouchá a na hlas, se kterým hovoří, by si zvykli. Tím bychom podpořili Hansonův argument proti existenci uncanny valley, o kterém jsem hovořila v předchozí kapitole.

Na druhou stranu, co když jim syntetický hlas dával pocit, že hovoří s něčím nedokonalým, něčím, co je teprve v počátcích, a právě proto jednali s avatarem s benevolencí a tolerancí? V jistém smyslu jako s dítětem, kterému rádi něco vysvětlíme, ale zároveň, díky promyšlenému kladení správných otázek jako s dospělým inteligentním člověkem, který má pochopení pro naše pocity a empatii, kterou potřebujeme cítit, pokud se chceme s něčím svěřit. To obsahoval hlas, který tu byl použit, a to i přesto, že nebyla použita emoční syntéza řeči. Syntetický hlas dal lidem to, co potřebovali, ve většině případů splnil účel nad míru očekávání.

Profesor Ishiguro, který nepřipouští existenci uncanny valley a snaží se vytvořit co nejvěrnější lidské kopie, argumentuje, že technický vzhled může v člověku vyvolat sklon k tomu, aby se lidé k těmto strojům chovali hrubě, anebo alespoň nekultivovaně (MacDorman a Ishiguro 2006, s. 314). Tento jev jsem však nepozorovala ani v nejmenším. Všichni avatarovi vykali, pozdrav a rozloučení byl, ve většině případů, velmi vřelý, účastníci se snažili udělat dobrý dojem, hovořili příjemným tónem a někteří se i snažili upozornit na to, že říkají pravdu. Žádný projev nezdvořilosti jsem nezaznamenala. To nás ale také přivádí k otázce, zda by tomu v soukromí nebylo jinak. Co když je zde nasnadě námitka vznešená vůči samotné metodě, co když účastníci hráli pouze role, které jim byly přiděleny?

Za nevýhodu syntetického hlasu považuji, že u některých promluv není poznat, zda jde o větu tázací či oznamovací. V tomto případě bych to však nepovažovala za problém. I v běžné mluvě si s naším partnerem v komunikaci někdy nerozumíme. Pokud subjekt avatarce nerozuměl, bez ostychu se jí zeptal.

Nad čím bychom se však měli pozastavit, je tendence člověka napodobovat chování a i mluvu ostatních. Zaměříme-li se pouze na doménu, která se má specializovat na to, aby stroj dělal společnost starším lidem, můžu pravděpodobně říci, že je syntetický hlas dostačující. Zamyslíme-li se ale nad dalším využitím, například interakci počítače a dětí anebo dětí s autismem, vyvstává otázka, nebude-li mít na ně syntetický hlas negativní vliv v tom smyslu, že budou mít nevědomý sklon k tomu, aby jej napodobovaly.

Například v projektu The AuRoRa, který se zabývá otázkou, zda by mohl robot sloužit pro vzdělávací a terapeutický účel pro děti s autismem a zvyšovat tak jejich schopnosti komunikace a interakce v sociálním prostředí, bylo použito techniky vyjadřující pouze velmi jednoduchou mimiku a gestiku. Tyto děti navázaly s takovým robotem vztah a bály se jej méně, ale zároveň začaly jeho jednoduché pohyby imitovat (Becker 2009, s. 58). Nabízí se tedy otázka, pokud napodobovaly pohyby, nesnažily by se imitovat i strojově znějící hlas? Bylo by v takovém případě vhodné použít syntetický hlas? U dětí, které trpí autismem je důležitý každý pokrok. Jak by tomu bylo ale u zdravých dětí, jejichž mluvu ovlivňuje vše, s čím přicházejí do styku?

Nejasnosti

Při hlubším zamyšlení nad tímto konkrétním experimentem se musím ptát, jak jej účastníci pochopili. Jednali s avatarem jako s člověkem, protože díky jeho odpovídajícím reakcím zapomněli na skutečnost, že mluví s umělou inteligencí? Stal se pro ně rozhovor natolik přirozeným? Nebo v tomto případě nebylo podstatné, s kým hovoří? Anebo experiment vůbec nepochopili? Jako konkrétní příklad bych chtěla upozornit na subjekt číslo 45, jehož chování mě na tuto myšlenku přivedlo.

Jedná se o muže, kterému avatar položil nejnižší počet otázek (18 za hodinu), jde o velmi přirozený dialog. Syntetický hlas jej vůbec neruší a, jak sám řekl, i přesto, že měl z experimentu velké obavy, tak ho rozhovor velmi bavil. Během dialogu, nebo spíše monologu, se dělil nejen o své zážitky, ale i pocity, které během nich prožíval. Polidštění avatarky bylo u tohoto subjektu velmi silné:

„Kdybyste přijela k nám na chatu, budete vítána."
„Budete vítaná, vopravdu když se tam přijedete podívat."
„Jestli máte doma jezírko, tak bych vám dal nějaké rybičky."
„Přijeďte se podívat, hrozně rád vás tam uvítám, u mě na tej zahradě."
„Bezvadně jsem si popovídal s tou slečnou nebo mladou paní."

Byl experiment typu WOZ neúspěšný a uživatel nevěřil tomu, že hovoří pouze s umělou inteligencí? Anebo je možné, že se natolik uvolnil, že zapomněl, že hovoří se strojem? A je v pořádku, že proběhne přiřazení lidských vlastností počítači během tak krátké chvíle? Pokud by opravdu systém takto fungoval, byl by velmi užitečný jako společnost pro osamělé lidi. Vzpomeňme si však na program ELIZA, který byl v letech 1964-1965 vytvořen Josephem Weizenbaumem. Tento počítačový program parodoval rogeriánského psychoterapeuta, který pacientovi pouze vracel otázky, a tím jej přiměl k rozhovoru. Tento program se stal brzy známým po celém světě a jeho špatné pochopení přimělo Weizenbauma zamyslet se i nad etickými otázkami v oboru informatiky.

„Byl jsem zděšen, když jsem viděl, jak rychle a jak hluboce lidé rozmlouvají s DOCTOREM (ELIZOU) citově přilnuli k počítači a jak jednoznačně jej antropomorfizovali...Lidé hovořili s počítačem, jako by to byla osoba, na níž se lze

obracet přiměřeně a účelně intimními slovy. Věděl jsem ovšem, že si lidé vytvářejí nejrozmanitější emoční vazby ke strojům, např. k hudebním nástrojům, motocyklům atd. A z dlouhé zkušenosti jsem věděl, že silné emoční vazby, které mají programátoři k počítačům, se často vytvoří po krátké době styku s nimi. Nevěděl jsem však, že velmi krátký styk s poměrně jednoduchým počítačovým programem může vyvolat silné klamné myšlení u zcela normálních lidí. Toto poznání mne přivedlo k tomu, abych přisoudil nový význam otázkám vztahu mezi jedincem a počítačem, a tudíž předsevzetí zabývat se tím.“ (Weizenbaum, 2002, s. 13)

Program, který by fungoval stejně jako wizard v Senior Companion, bychom samozřejmě neměli připodobnit ELIZE, která odpovídala na mnohem jednodušších principech. Avšak přisuzování lidských vlastností umělé inteligenci během tak krátké doby a ještě za použití syntetického hlasu by nás mělo vést k zamyšlení nad tím, před čím varuje Weizenbaum. Měli bychom technologiím dát takovou důvěru, měly by opravdu hrát úlohy, které jsou doposud pouze v lidských rukách, a je to potřeba? Program ELIZA byl tehdejšími psychology špatně pochopen, považovali jej za přelom - program by mohl nahradit člověka. Člověk potřebuje člověka. V některých sférách bychom proto měli být opatrní v tom, svěříme-li se do rukou umělé inteligenci.

5 Závěr

Žijeme v době, ve které se mnoho technologií stalo neviditelnou složkou našich životů, používáme je s takovou samozřejmostí, že si často ani neuvědomujeme, o jak výjimečné a složité systémy se jedná. Dialogové systémy stále ale neviditelné nejsou, nepovažujeme je za samozřejmost, kterou jsme obklopeni. Jsou technologií, na které pracujeme již léta, a přesto stále nedosahuje uspokojivých výsledků. Je to proto, že některé lidské schopnosti počítači nelze předat? Nebo bychom se měli soustředit na vytvoření stroje, který funguje úplně na jiných principech než člověk, držet se toho, že jde o artefakt, kterému máme pouze vložit schopnost řeči, která by však odpovídala schopnostem technologie, ne člověka?

Za primární cíl své práce považuji podání podkladů k zamyšlení se nad tímto aspektem. Stroj je stroj, člověk je člověk. „Není zajímavé vytvořit robota, který vypadá a chová se stejně jako člověk, roboti by měli být odlišní od lidských bytostí.“, jak řekl Mori. Držela jsem se Moriho slov a pokusila jsem se jeho teorii o jevu uncanny valley aplikovat na řečové technologie. Sledováním videozáznamů pořízených v rámci projektu Senior Companion jsem zjišťovala, jak působí synteticky znějící hlas na účastníky tohoto experimentu.

Je samozřejmé, že vždy, když vyvozujeme závěry z experimentů, v nichž jsou objektem pozorování lidé, musíme počítat s individualitou, kterou má v sobě každý z nás. Každý člověk má jiné vlastnosti, jinou povahu a v různých situacích také reaguje různě. S ohledem na tento fakt mohu říci, že systém, který by pracoval na tomto principu, by mohl fungovat jako společník pro starší lidi a udržoval by tak jejich mentální zdraví. Synteticky znějící hlas by zde nebyl překážkou, naopak by mohl být i lepší. Pozorované subjekty cítily dostačující míru souznění s avatarkou a zároveň strojově znějící hlas nenesl riziko, že se ocitne v uncanny valley.

Vzájemnost, kterou nalézáme v mezilidské interakci, může být pokládána za typickou, předpokládá nějakou formu personality a tělesnosti, od které jsou virtuální agenti daleko. Syntetický hlas zní prázdně, někdo může říci, že mu chybí osobnost, jeho dynamika sleduje stereotypní modely, které nemají stejnou rezonanci jako lidská

řeč, přesto tento experiment ukázal, že v sobě nese potenciál, aby s ním lidé komunikovali na stejné úrovni jako s člověkem.

Literatura

BARTHELMESS, Ulrike, FURBACH, Ulrich, 2012. *IRobot-uMan: künstliche Intelligenz und Kultur : eine jahrtausendealte Beziehungskiste.* Berlin: Springer. ISBN 36-422-2927-1.

Bartneck, C., Kanda, T., Ishiguro, H., Hagita, N., 2007. Is the Uncanny Valley an Uncanny Cliff? *Proceedings of the 16 th IEEE International Symposium on Robot and Human Interactive Communication*, s. 368-373. Jeju.

BARTNECK, Ch., KANDA, T., ISHIGURO, H., HAGITA, N., 2009. My robotic doppelgänger - a critical look at the Uncanny Valley. *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*. s. 269-276. Toyama.

BECKER, Barbara, 2009. Humanoide Roboter, emotionale Konversationsagenten : Anmerkungen zu neuen Konzepten in der Mensch-Maschine-Interactin. In: *Public Fictions: Wie man Roboter und Menschen erfindet.* 1.vyd. Innsbruck: Studien Verlag. s. 16-33. ISBN 3-7065-4714-7.

BROOKS, Rodney, 2002. *Menschmaschinen: wie uns die Zukunftstechnologien neu erschaffen.* Frankfurt/Main : Campus Verlag GmbH. ISBN 3-593-36784-X.

DUFFY, Brian, 2009. Die Problematik socialer Roboter. In: *Public Fictions: Wie man Roboter und Menschen erfindet.* 1.vyd. Innsbruck: Studien Verlag. s. 16-33. ISBN 3-7065-4714-7.

EKLUND, Robert, 2004. Disfluency in Swedish human-human and human-machine travel booking dialogues. Linköping: Unitrick. ISBN 91-737-3966-9

EKLUND, J., GUSTAFSON, J., HELDNER, M., HJALMARSSON, A., 2008. Towards human-like spoken dialogue systems. *Speech comunication* [online]. 50(8-9), s. 630-645 [cit. 22.4.2012]. Dostupné z: http://linkinghub.elsevier.com/retrieve/pii/S016763930800054X

GUIZZO, Erico, 2010. Who is Afraid of the Uncanny Valley. In: *IEEE Spectrum* [online]. 2.5.2013 [cit. 22.4.2013] Dostupné z: http://spectrum.ieee.org/automaton/robotics/humanoids/040210-who-is-afraid-of-the-uncanny-valley

HANSON, David, 2006. Exploring the aesthetic range for humanoid robots. *Toward Social Mechanisms of Android Science An ICCS/CogSci-2006 Long Symposium*. s. 39-42. Vancouver.

Heinz Nixdorf MuseumsForum, 2001. *Computer. Gehirn: was kann der Mensch? Was können die Computer?* ; Begleitpublikation zur Sonderausstellung im Heinz-Nixdorf-MuseumsForum. Paderborn, München, Wien, Zürich : Schöningh. ISBN 35-067-6230-3.

HSU, Jeremy, 2012. Robotic's Uncanny Valley Gets New Transtation. In: *LiveScience* [online]. 12.6.2012 [cit. 22.4.2013] Dostupné z: http://www.livescience.com/20909-robotics-uncanny-valley-translation.html

KAGEKI, Norri, 2012. Masahiro Mori on the Uncanny Valley and Beyond. In: *IEEE Spectrum* [online]. 12.6.2012 [cit. 22.4.2013] Dostupné z: http://spectrum.ieee.org/automaton/robotics/humanoids/an-uncanny-mind-masahiro-mori-on-the-uncanny-valley

KANELLOS, Michael, 2003. Moores'Law to roll on for another decade. In: *CNET News* [online]. 10.2.2003 [cit. 22.4.2013]. Dostupné z: http://news.cnet.com/2100-1001-984051.html

KURZWEIL, Raymond, 1997. When Will HAL Understand What We Are Saying?. In: *HAL's Legacy : 2001's Computer as Dream and Reality*. Cambridge : MIT Press, s. 131-169. ISBN 0-262-19378-7.

KURZWEIL, Raymond, 1999. *Homo s@piens : Leben im 21. Jahrhundert : was bleibt von Menschen.* 3. vyd. Köln : Kiepenheuer & Witsch. ISBN 34-620-2741-7.

LEGÁT, M., GRŮBER, M., IRCING, P., 2008. Wizard of Oz Data Collection for the Czech Senior Companion Dialogue System . *Fourth International Workshop on Human-Computer Conversation*, s. 1-4, University of Sheffield.

MACDORMAN, F. KARL, 2005. Androids as experimental apparatus: Why is there an uncanny valley and can we exploit it? *CogSci-2005 Workshop: Toward Social Mechanisms of Android Science*, s. 108–118. Stresa.

MACDORMAN, Karl F., ISHIGURO, Hiroshi, 2006. The uncanny advantage of using androids in social and cognitive science research. *Interaction Studies* [online], **7**(3), s. 297–337 [cit. 22.4.2013] ISSN 15720373. Dostupné z: doi:10.1075/is.7.3.03mac

MITCHELL, W. J, SZERSZEN, K. A. SR, Shirong LU, A.,SCHERMERHORN, P.W., SCHEUTZ, M. a MACDORMAN, K. F. 2011. A mismatch in the human realism of face and voice produces an uncanny valley. *I-Perception* [online]. **2**(1), s. 10-12 [cit. 22.4.2013]. ISSN 2041-6695. Dostupné z: http://i-perception.perceptionweb.com/journal/I/article/i0415

MOORE, Gordon E., 1965. Cramming more components onto integrated circuits. Electronics. [online], 38(8) , s. 114–117 [cit. 22.4.2013]. Dostupné z: ftp://download.intel.com/museum/Moores_Law/Articles-press_Releases/Gordon_Moore_1965_Article.pdf

MORI, Masahiro, 2012. The uncanny valley. *IEEE Robotics and Automation* [online]. **19**(2), s. 98–100 [cit. 22.4.2013]. ISSN 1070-9932. Dostupné z: doi:10.1109/MRA.2012.2192811

MUBIN, Omar, BARTNECK, Christoph, FEIJS, Loe, 2009. Designing an Artificial Robotic Interaction Language. *Human-Computer Interaction – Interact 2009* [online]. Berlin: Springer, s. 848-851 [cit. 2013-04-22]. ISBN 978-3-642-03657-6. Dostupné z http://www.bartneck.de/publications/2009/artificialRoboticInteractionLanguage/index.html

OLIVE, Joseph P., 1997. „The Talking Computer“: Text to speech synthesis. In: *HAL's Legacy : 2001's Computer as Dream and Reality*. Cambridge : MIT Press, s. 101-128. ISBN 0-262-19378-7.

PIERACCINI, R., SUENDERMANN, D., DAYANIDHI, K. a LISCOMBE, J. 2009. Are We There Yet? Research in Commercial Spoken Dialog Systems. *Text, speech and Dialogue* [online]. s. 3 - 13 [cit. 22.4.2013]. Dostupné z: http://www.springerlink.com/index/10.1007/978-3-642-04208-9_3

PIERACCINI, Roberto, 2012. *The voice in the machine : building computers that understand speech*. 1. vyd. Cambridge : MIT Press. ISBN 978-0-262-01685-8.

PINKER, Steven, 2009. *Jazykový instinkt : Jak mysl vytváří jazyk.* 1. vyd. Praha : Dybbuk. ISBN 978-80-7438-006-8.

Plug and Pray [dokumentární film]. Scénář a režie Jens Schanze, Spolková republika Německo. 2010.

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

ROMPORTL, J., ZOVATO, E., SANTOS, R., IRCING, P., GIL, J.R., DANIELI, M., 2010. Application of Expressive TTS Synthesis in an Advanced ECA System. *Proceedings of the ISCA Tutorial and Research Workshop on Speech Synthesis*, s. 120-125, National Institute of Information and Communications Technology, Kyoto.

RUSSELL J. Love, WANDA G. Webb, 2009. Mozek a řeč: neurologie nejen pro logopedy. 1. vyd. Praha : Portál. ISBN 978-80-7367-464-9.

SZOLLOSY, Michael, 2012. Why Are We Afraid of Robots? The Role of Projections in the Popular Conception of Robots. In: *Beyond AI: Artifical Dreams.* Plzeň : University of West Bohemia. s. 41-51. ISBN 978-80-261-0102-4.

Ševela, Vladimír, 2003. Informatik Frederick Jelinek. In: *Pátek lidových novin* [online]. 17.10.2003 [cit. 22.4.2013]. Dostupné z: ufal.mff.cuni.cz/.../20031017_LN_Informatik_Frederick_Jelinek.pdf

WEIZENBAUM, Joseph, 2002. *Mýtus počítače: počítačový pohled na svět.* Břeclav: Moravia-Press. ISBN 80-861-8155-3.

Resumé

The aim of this paper is to present the state of speech technologies research and underscore the difficulties the scholars in this field of studies encounter. Creation of a dialogue systhem able to understand a coherent human speech with a low failure rate, relatively unlimited vocabulary and a domain without previous experience with speaker able at the same time to produce an understandable and natural speech, so a system indiscernible from human being, is being considered a holy greal in the development of speech technologies. However, this goal can bear a risk of the so called uncanny valley: a reaction of man to a machine that looks and moves similar to human. Man's reaction to this kind of robot can range from positive throughout to negative feelings and the interaction can be thus disrupted at the time when the robot is although similar to man but, one the other hand, his appearance is still imperfect. This hypothesis is in this paper being applied to the use of speech technologies in the framework of Senior Companions project. Experimental data obtained in the frame of this project by using Wizard of Oz methods have shown that synthetically sounded voice can cause emotional reactions similar to those present in inter-human communication. Observed subjects felt a sufficient level of accord with avatar and, at the same time, machine sounded voice did not bring any risk of uncanny valley. It has also been proven that synthetically sounded voice can be not only sufficient but even better as there is no risk of uncanny valley.