

Rečové prispôsobovanie sa medzi človekom a automatickým systémom

Štefan Beňuš

Univerzita Konštantína Filozofa v Nitre, Štefánikova 67 94974 Nitra
Ústav Informatiky SAV, Dúbravská cesta 9, 845 07 Bratislava
sbenus@ukf.sk

Abstrakt

Základnou otázkou tohto výskumu je to, akým hlasom by mal rozprávať robot alebo komunikačný systém, aby jeho príkazy a rady boli dôveryhodné. Pri medziľudskej komunikácii dochádza k prispôbovaniu sa medzi rečníkmi prirodzene a je spojené s kladným hodnotením spolurečníka, úspešnosti komunikácie, a ďalšími pozitívami. Vychádzajúc z tohto poznania sme vyvinuli komunikačnú hru, v ktorej človek hrá kartovú hru a počas nej dostáva rady od avatarov. Skúmame, či hráč bude viac dôverovať radám od toho avatara, ktorého syntetický hlas sa prozodicky prispôbuje hlasu hráča. Doterajšie výsledky naznačujú, že rečové prispôbovanie automatického systému človeku môže ovplyvňovať dôveryhodnosť systému a aj rozhodovanie človeka, ktorý so systémom komunikuje.

1 Úvod

Komunikácia medzi strojom a človekom v súčasnosti už bežne prebieha v rečovej modalite, čo je možné doložiť úspešnosťou a obľúbenosťou komunikačných asistentov ako Siri od Apple, Cortana od Microsoftu, Alexy od Amazonu, alebo Google Assistant. Tento boom bol umožnený veľkým progresom v počítačovom spracovaní hlasového signálu pri rozpoznávaní ľudského hlasu a generovaní syntetického hlasu. Charakteristiky dialógov medzi človekom a strojom však ešte ani zďaleka nedosahujú komplexnosť medziľudskej komunikácie. Komplexnosť vzťahov, ktoré sa dialógmi vytvárajú a upevňujú, je teda omnoho väčšia medzi ľuďmi ako pri komunikácii človek-stroj.

Jednou z oblastí, kde sa tento rozdiel prejavuje je komunikatívne prispôbovanie sa, ktoré sa v literatúre rôznych vedeckých oblastí označuje rôzne, ako napríklad entrainment, accommodation, adaptation, alignment, synchrony, a inak, pozri napr. prehľadové štúdie Beňuš (2014) alebo Brannigan a kol. (2011). V rečovej medziľudskej interakcii bolo takéto prispôbovanie preukázané v mnohých lingvistických ako aj paralingvistických rovinách dialógu, napríklad vo fonetickej realizácii jednotlivých slov (Pardo 2006), výbere lexikálnych jednotiek a syntaktických štruktúr (Brennan 2000, Reitter a kol. 2007, Brannigan a kol. 2000), časových charakteristikách preberania slova (ten

Bosch a kol. 2005, Beňuš a kol. 2011), alebo akusticko-prozodických vlastnostiach ako výška a intenzita hlasu alebo tempo reči (napr. Levitan a Hirschberg 2011).

Komunikatívne prispôbovanie medzi spolurečníkmi je prepojené na sociálnu rovinu medziľudskej komunikácie, keďže existuje silná korelácia medzi mierou prispôbovania a pozitívneho vnímania spolurečníka a aj celkovej interakcie. Táto vnímanka je formalizovaná v Communication Accommodation Theory (CAT, Giles a kol. 1991), ktorá hovorí, že miera prispôbovania je jednou z možností ako dosiahnuť (alebo ovplyvniť) mieru vzájomnej blízkosti medzi spolurečníkmi. V princípe, väčšia miera prispôbovania vedie k pozitívnejšiemu výsledku komunikácie a rečové prispôbovanie tak odzrkadľuje snahu o sociálnu integráciu.

V poslednej dobe v snahe využiť tieto pozitívne črty prispôbovania a vylepšiť automatické dialógové systémy sa formy komunikatívneho prispôbovania začali skúmať aj pri interakcii človek-stroj (napr. Bell a kol. 2003, Brannigan a kol. 2010). Implementácia prispôbovania sa automatického systému k svojmu užívateľovi má niekoľko výhod. Na jednej strane je dobré ak sa automatický systém snaží implementovať črty medziľudského dialógu aby sa tak zvyšovala prirodzenosť dialógových systémov. Pri aplikáciách zameraných na poskytovanie služieb alebo plnenie úloh môže takéto prispôbovanie pozitívne ovplyvňovať vnímanie užívateľa smerom k takémuto systému ako aj vnímanie úspešnosti celej interakcie podobne, ako to bolo pozorované v medziľudských interakciách. Pri aplikáciách, kde automatický systém slúži ako konverzačný spoločník, asistent pri doučovaní, alebo spoločník seniorov s medicínskymi a spoločenskými úlohami, môže takéto prispôbovanie pozitívne ovplyvniť spriaznenosť užívateľa s takýmto systémom, čo následne môže viesť k väčšej efektívnosti pri plnení úloh. V nedávnej štúdii napríklad Lubold a kol. (2016) zistili, že akusticko-prozodické prispôbovanie sa robotického pomocníka pri doučovaní signifikantne zvyšuje "social presence", teda vnímanie interakcie s takýmto robotickým spoločníkom ako prirodzené, príjemné, a motivujúce.

Z kognitívneho hľadiska je samozrejme výskum vzájomného rečového prispôbovania sa dôležitý pri formulácii modelov kognitívneho systému reči, ktoré donedávna boli založené na poznatkoch hlavne z

monologického jazykového prejavu, ale modely kognitívnych jazykových schopností založené na podstate dialógu sa v poslednej dobe dostávajú do popredia (napr. Pickering a Garrod 2013). Popri výskume kognitívneho rečového systému vstupuje do popredia aj výskum rozhodovacích procesov človeka pri jeho komunikácii so strojom a ako toto rozhodovanie môže byť ovplyvnené.

Je potrebné však spomenúť, že prispôsobovanie a jeho potenciálne pozitívne črty sa nemusia v interakcii medzi človekom a strojom objavovať automaticky. Strupka a kol. (2016) napríklad zistili, že pri komunikácii s robotmi, ktorí hovorili mužskými a ženskými hlasmi, užívatelia prispôbili intenzitu a intonáciu hlasu pri rozdielnom pohlaví robotického hlasu, ale efekt bol opačný: hlasové charakteristiky užívateľov sa nepribližovali robotovým, ale sa od nich vzdŕaľovali.

Smer výskumu popísaný vyššie by sme radi posunuli preskúmaním možnosti, že akusticko-prozodické prispôsobovania automatických syntetických hlasov ovplyvňuje dôveryhodnosť užívateľa voči týmto hlasom a môže na základe zmeny dôveryhodnosti aj ovplyvniť ich rozhodovanie.

2 Metódy

2.1 Prispôsobovanie

Spôsobov, ako je možné rečové prispôsobovanie v dialógoch uchopiť, je niekoľko. Levitan a Hirschberg (2011) popísali tri základné: podobnosť, konvergencia, a synchronnosť. Pri podobnosti ide o porovnanie priemernej hodnoty určitej charakteristiky, napríklad tempa reči, počas celého dialógu medzi rečníkmi. Môže sa to využiť napríklad na porovnanie takto formulovanej miery podobnosti medzi rečníkom a jeho/jej spolurečníkom na jednej strane a na druhej strane medzi rečníkom a "imaginárnym spolurečníkom", teda napríklad priemernou hodnotou u ostatných hovoriacich v databáze, ktorí však s rečníkom nekomunikovali. Ak je rozdiel rečník-spolurečník menší, ako rečník-imaginárny spolurečník, došlo k rečovému prispôsobovaniu sa medzi rečníkom a spolurečníkom.

Pri konvergencii ide o mieru približovania sa hodnôt určitej rečovej charakteristiky medzi spolurečníkmi počas celého dialógu. Je to teda čiastočne dynamická veličina (aj keď vyjadrená statickou jednou hodnotou z celého rozhovoru) miery časovej zmeny podobnosti.

Poslednou formou uchopenie prispôsobovania je synchronnosť. Toto je najviac dynamický koncept, ktorý vyjadruje relatívnu koordináciu spolurečníkov pri každej zmene hovoriaceho, teda pri každom prevzatí slova (turn exchange). Nejde tu teda o konvergenciu, keďže "vzdialenosť" medzi spolurečníkmi sa nemusí zmenšovať, ale o prispôsobovanie sa smeru zmeny. Napríklad, ak Rečník 1 hovoril rýchlejšie ako Rečník 2

a v určitom momente Rečník 1 spomalí, Rečník 2 sa bude správať synchronne, ak aj on spomalí a nepriblíži sa tak Rečníkovi 1, ale jeho správanie bude konvergentné, ak trochu zrýchli, aby sa tak "vzdialenosť" medzi rečníkmi zmenšila.

Ako forma prispôsobovania v našej komunikatívnej hre opísanej nižšie bola zvolená práve synchronnosť. Hlavnými dôvodmi bolo to, že synchronnosť sa hojne vyskytuje v medziľudskej komunikácii (napr. Edlund a kol. 2009, Levitan, a kol. 2015), a umožňuje aktívnejší a variabilnejší prístup k manipulácii syntetických hlasov počas hrania hry v porovnaní s konvergenciou alebo podobnosťou.

2.2 Komunikatívna hra Kvarteto s Pomocníkmi

2.2.1 Myšlienka

Hra Kvarteto je prispôbenuou verziou hry *GoFishwithHelpers*, ktorú navrhla vo svojej dizertačnej práci R. Levitan (2014). Hra testuje základnú otázku tohto výskumu, či hráč dôveruje viac tomu avatarovi, ktorého hlas sa mu prispôsobuje ako avatarovi, ktorého hlas ostáva statický, alebo sa aktívne neprispôsobuje tým, že sa pohybuje asynchronne voči hlasu hráča.

Použitá bola jednoduchá hra Kvarteto, kde si hráč pýta od svojho súpera kartu určitej hodnoty s cieľom zozbierať všetky štyri karty tejto hodnoty a vytvoriť tak Kvarteto. Úlohou hráča je nazbierať čo najviac bodov ktoré sa udeľujú pri získaní každej karty od súpera (10 bodov), pri získaní kvarteta (100 bodov), a 50 bodov sa stráca v situácii ak si hráč vyžiada kartu a súper žiadnu kartu tejto hodnoty nemá. Súperom hráča je počítač, ktorého karty hráč nevidí. Myšlienka využitia kartovej hry ovládanej hlasom na rôzne účely v slovenčine nie je nová (napr. Nagy 2006 využíva kvarteto na zlepšenie povedomia o čísliciach pre žiakov základných škôl). V tejto hre ale pri každom ťahu hráč musí hlasovo požiadať o radu od jedného z dvoch avatarov, akú kartu by si mal od súpera vypýtať. Hráč vidí na obrazovke obrázky dvoch jasne rozdielnych ženských tvári (Monika a Tereza) a príklady otázok, ktoré môže použiť, napr. "Monika, ktorú kartu by si mi odporučila?". Rozpoznávač zistí z tejto otázky identitu avatara, ktorý bol o radu požiadaný, a pri následnej rade sa ukáže obrázok požadovaného avatara a syntetizátor prehrá danú radu, napr. "Ja by som odporučila šestku". Hlas jedného z avatarov sa počas celej hry aktívne dynamicky prispôsobuje hráčovi, t.j. každá jeho rada je podobná prozodickým charakteristikám otázky, ktorou si hráč radu vypýtal, a hlas druhého z avatarov sa hlasu hráča neprispôsobuje.

Dôležitým predpokladom testovania základnej výskumnej otázky je to, aby kvalita rady, ktorú avatar poskytne, neslúžila hráčovi na rozhodovanie, komu bude viac dôverovať. To znamená, že chceme predísť tomu, aby sa hráč rozhodoval podľa kvality rád. Toto je zabezpečené dvomi súčasťami hry. Prvou je

algoritmická implementácia, ktorá zabezpečuje, že rozdiel výsledných bodových ziskov hráča plynúcich z rád avatarov je počas hry udržiavaný na minime, čo znamená, že obidva avatari dávajú v priemere celej hry rovnako dobré rady. Druhou je to, že hráč nevidí karty protihráča (počítača), ktorý pri každom ťahu dostane novú sadu kariet. Tým je možné, že pri každej rade existuje v princípe "lepšia" rada, ktorá by hráčovi priniesla viac bodov, ako aj "horšia" rada, ktorá by priniesla bodov menej. Tieto vlastnosti hry umožňujú, aby hráč, aj napriek tomu, že algoritmicky oba avatari poskytujú rady približne rovnakej kvality, mohol (mylne) predpokladať, že jeden z avatarov predsa len poskytuje lepšie rady.



Obr. 1: Grafické užívateľské rozhranie pre hru Kvarteto s pomocníkmi. Vpravo hore je tlačidlo na spustenie a zastavenie rozpoznávania (pri stlačení signalizované ikonou mikrofónu), pod nimi obrázky avatarov a návrhy otázok, vľavo hore karty súpera (počítača) a vľavo dole karty hráča s aktuálnym skóre.

2.2.2 Implementácia

Naša implementácia je pozmenená oproti tej, ktorú použila Levitan (2014) vo viacerých ohľadoch, ktoré popíšeme nižšie. Využíva sa programovacie prostredie Python a http komunikácia so serverom zabezpečujúcim rozpoznávanie reči hráča a syntézu reči avatarov. Grafické užívateľské prostredie je ilustrované na Obr. 1 spolu s popisom jednotlivých súčastí.

Základom implementácie je extrakcia prozodických charakteristík reči s hlasu hráča a následná reprodukcia týchto charakteristík v hlase prispôbujúceho sa avatara. V tomto príspevku prezentujeme výsledky dvoch experimentov: v prvom sa prispôbovalo len tempo reči a v druhom priemerná intenzita, výška hlasu (f_0) a tempo reči.

Tempo reči je možné získať buď priamo z akustického signálu odhadom počtu slabík (napríklad použitím rozšíreného softvéru na fonetickú analýzu Praat (Boersma a Weenink 2017) a dostupných skriptov na takéto určenie tempa reči v Praat-e (Mertens 2014, de Jong a Wempe 2009), alebo

použitím výstupu z rozpoznávača, kde je počet slabík v rozpoznanom texte možné algoritmicky odvodiť podľa počtu slabičných elementov. Pri experimentálnom testovaní bola metóda použitia výstupu rozpoznávača viac spoľahlivá a bola teda implementovaná v hre Kvarteto. Priemerné hodnoty výšky hlasu a intenzity sa použili z výstupu Praat skriptu (de Jong a Wempe 2009), ktorý sme mierne prispôbobi našim potrebám.

Podstatným rozdielom medzi súčasnou implementáciou a tou, ktorú použila Levitan (2014) je spôsob prispôbovania, teda reprodukcie prozodických vlastností reči hráča v syntetickej reči prispôbujúceho sa avatara. Levitan zvolila reprodukciu absolútnych nameraných hodnôt z hlasu hráča v hlase avatara. V našich experimentoch sme zvolili relatívne prispôbovanie synchronnosti popísanej v podkapitole 2.1. Pri každom ťahu si hráč vypýta radu, vypočíta sa tempo reči, výška a intenzitu hlasu, tieto hodnoty sa porovnávajú s hodnotami v otázke v predchádzajúcom ťahu, a v syntéze následnej rady prispôbujúceho sa avatara sa posunú o tento rozdiel. Ak hráč zrýchli alebo hovoril hlasnejšie oproti svojmu predchádzajúceho ťahu, prispôbivý avatar tiež zrýchli alebo bude hovoriť hlasnejšie.

Pri implementácii hlasu avatara, ktorý sa neprispôbovuje sme testovali dva spôsoby. V prvom (*Adapt-Stable*) sa hlas avatara vôbec nehybal a počas celej hry produkoval predvolené hodnoty daného syntetického hlasu. V druhom (*Adapt-Antiadapt*) sa hodnoty v syntéze neprispôbivého hlasu posúvali v opačnom smere oproti zmene nameranej v hlase hráča: ak si hráč vypýtal radu vyšším hlasom ako v minulom ťahu o napr. 5 Hz, neprispôbivý avatar poskytol radu hlasom o 5Hz zníženou f_0 .

Syntéza a rozpoznávanie bolo zabezpečené systémom vytvoreným na Ústave Informatiky SAV. Rozpoznávač využíva voľne dostupný dekodér reči Julius (<http://julius.osdn.jp/en-index.php>) obohatený o trifónový akustický model a jazykový model slovenčiny. Syntetizátor slovenčiny je založený na štatistickej parametrickej metóde voľne dostupného softvéru HTS (<http://hts.sp.nitech.ac.jp>) vytvorenej z jedného ženského hlasu hovorenej slovenčiny na základe databázy čítaných textov.

V tejto implementácii je dialóg zjednodušený tak, že rozpoznávač len zistí identitu cieľového avatara (Monika-Tereza), a nápoveda sa vyberie náhodne z viacerých prednastavených viet. Hráč nemá možnosť sám si vyberať karty a radu avatara musí poslúchnuť. Jeho jedinou voľbou a stratégiou je to, ktorého avatara sa opýta.

3 Experiment 1

3.1 Probandi

V experimente sa zúčastnilo 20 probantov. Všetci boli študenti nitrianskych univerzít, z toho 14 žien a 6 mužov.

3.2 Procedúra

Probandi sa zúčastnili experimentu po vyslovení informovaného súhlasu a boli finančne odmenení za svoj čas. Na začiatku vyplnili sociometrický dotazník a skrátený osobnostný test TIPI (Ten Item Personality Inventory, Gosling a kol. 2003). Potom sa oboznámili s pravidlami hry a inštrukciou, že pri hre budú komunikovať s dvomi avatarami, ktorí väčšinou dávajú výhodné rady, ale niekedy môžu robiť aj chyby a ich úlohou v experimente je nahráť čo najviac bodov a tak zísťovať, ktorý z avatarov poskytuje lepšie rady.

Po úvodnom oboznámení sa s hrou a grafickým užívateľským rozhraním hráči pristúpili k samotnej hre s dvomi avatarami (Monikou a Terezou) v celkovej dĺžke 45 ťahov. Na záver probanti vyplnili krátky dotazník, v ktorom vyberali, ktorý avatar podľa nich dával lepšie rady a mal príjemnejší hlas a hodnotili hlasové charakteristiky avatarov.

Hlasy avatarov sa líšili iba v tempe reči. U prispôsobujúceho sa avatara sa tempo v radách synchronizovalo s tempom hráčovských žiadostí o radu. U neprispôsobivého avatara sa buď tempo nemenilo (*Stable*) alebo sa prispôbovalo asynchrónne, teda v opačnom smere (*Antiadapt*). Polovica probantov hrala hru v spôsobe *Adapt-Stable* a druhá polovica v móde *Adapt-Antiadapt*. V oboch podskupinách sa polovici probantov prispôbovala Monika a polovici Tereza.

3.3 Analýza dát

Primárne závislé premenné, pomocou ktorých sa testovala základná otázka tohto výskumu, boli dve formy pomeru rád, ktoré si hráč vypýtal od prispôsobujúceho a neprispôsobujúceho avatara. Pri pôvodnom pomere (*Skóre*) sa spočítal počet rád vyžiadaných od jednotlivých avatarov. Pri váhovanom pomere (*Váhované Skóre*) sme sa snažili podchytiť možnosť, že vplyv prispôsobovania na dôveru hráča k avatarovi sa vyvíja počas hrania a očakávali sme, že vzťah medzi prispôsobovaním a dôveryhodnosťou bude silnejší ku koncu hry ako na začiatku hry. Použili sme jednoduché váhovanie, pri ktorom sa porovnáva súčet poradových čísel ťahov, v ktorých si hráč pýtal radu o jedného alebo druhého avatara. Žiadosti o radu ku koncu hry majú väčšiu váhu ako tie pri začiatku hry.

Štatistická analýza využila párovaný t-test v relevantných podskupinách probantov a tiež aj lineárnu regresiu (*Váhovaného Skóre* ako závislej premennej a prispôsobovania (*Adapt-Stable* a *Adapt-Antiadapt*)

spolu s identitou avatara (*Monika-Tereza*) ako nezávislých premenných.

3.4 Výsledky

Výsledky experimentu ukazujú na vyvrátenie našej základnej hypotézy, že probanti budú viac dôverovať hlasu avatara, ktorý sa im prispôsobuje v jedinej akusticko-prozodickej vlastnosti tempa reči. Párové t-testy v skupine, ktorá hrala hru v móde *Adapt-Antiadapt* preukázali síce nesignifikantnú, ale konzistentnú, tendenciu, že hráči si častejšie pýtali radu od toho avatara, ktorý sa im neprispôboval (*Antiadapt*), pri neváhovanom skóre to bolo $t[9] = -1,75$, $p = 0,11$ a pri váhovanom $t[9] = -1,68$, $p = 0,13$. Lineárna regresia v tejto skupine ukázala signifikantne negatívny vzťah medzi prispôsobovaním sa a obidvomi verziami *Skóre* v tejto podskupine: $t(17) = -2,41$, $p = 0,028$ pre neváhované a $t(17) = -2,3$, $p = 0,034$ pre váhované. Žiadny signifikantný efekt sme nepozorovali v druhej skupine, ktorá hrala v móde *Adapt-Stable*. Identita avatara (Monika/Tereza) ani pohlavie probanta nemali vplyv na výsledky uvedené vyššie.

3.5 Diskusia

Výsledok, že ľudia viac dôverujú syntetickému hlasu, ktorý sa im neprispôsobuje (t.j. prispôsobuje asymetricky) je v rozpore s predchádzajúcimi výsledkami Levitan (2014), ktoré sú najviac porovnateľné s našou procedúrou, ako aj s očakávaniami, ktoré boli podporené kladným vzťahom medzi rečovým prispôsobovaním a pozitívnymi aspektami ako úspešnosť komunikácie alebo sympatickosť spolurečníka, ktorý sa prispôsobuje.

Veľmi podobné výsledky z neváhovaného a váhovaného skóre napovedajú, že v tomto type scenára komunikatívnej hry nedochádza k postupnému upevňovaniu vzťahu medzi rečovým prispôsobovaním a dôveryhodnosťou.

Je možné, že prispôsobovanie iba na základe tempa reči je vnímané inak, ako prispôsobovanie iných prozodických charakteristík ako napríklad výšky hlasu alebo intenzity. V nasledujúcom experimente sme chceli overiť, či súbor charakteristík, podľa ktorých sa syntetický hlas prispôsobuje užívateľovi má vplyv na rozvoj vzťahu medzi rečovým prispôsobovaním a pridali sme k tempu reči aj prispôsobovanie priemernej výšky a intenzity hlasu.

4 Experiment 2

4.1 Probandi

V experimente 2 sa zúčastnilo 24 probantov, z toho 17 žien a 7 mužov opäť väčšinou študentov nitrianskych univerzít.

4.2 Procedúra

Procedúra bola identická ako pri experimente 1 (Podkapitola 3.2), iba spôsob prispôsobovania teraz neobsahoval len tempo reči, ale aj priemernú výšku hlasu (f0) a intenzitu.

4.3 Analýza dát

Analýza dát bola identická ako pri experimente 1 (Podkapitola 3.3).

4.4 Výsledky

V porovnaní s experimentom 1 sa výsledky Experimentu 2 uberajú opačne: probanti dôverovali viac hlasu avatara, ktorý sa im prispôboval ako tomu, ktorý sa im neprispôboval. Párové t-testy v skupine, ktorá hrala hru v móde *Adapt-Antiadapt* preukázali signifikantný efekt, že hráči si častejšie pýtali radu od toho avatara, ktorý sa im prispôboval (*Adapt*), pri neváhanom skóre to bolo $t[11] = 1,87, p = 0,08$ a pri váhanom $t[11] = 2,49, p = 0,03$. Lineárna regresia v tejto skupine tiež ukázala signifikantne pozitívny vzťah medzi prispôbovaním sa a obidvomi verziami *Skóre* v tejto podskupine: $t(21) = 2,58, p = 0,017$ pre neváhané a $t(21) = 3,43, p = 0,025$ pre váhané. Žiadny signifikantný efekt sme nepozorovali v druhej skupine, ktorá hrala v móde *Adapt-Stable*. Identita avatara (Monika/Tereza) ani pohlavie probanta opäť nemali vplyv na výsledky uvedené vyššie.

4.5 Diskusia

Výsledok tohto experimentu je opačný, ako v Experiment 1. Probandi teraz viac dôverovali tomu avatarovi, ktorého syntetický hlas sa im symetricky prispôboval.

Podobne ako v Experimente 1 sa výsledky v neváhanom a váhanom skóre líšia len minimálne a ukazujú na robustnosť pozorovania, že vzťah medzi prispôbovaním sa a dôveryhodnosťou je relatívne stabilný.

	Prispôsobovanie syntetického hlasu v:	Dôveryhodnejší hlas
Exp1	Tempe reči	neprispôsobujúci
Exp2	Tempe, intenzite a výške hlasu	prispôsobujúci

Tab. 1: Sumár výsledkov oboch experimentov

5 Všeobecná diskusia

Dlhodobým cieľom tohto výskumu je zistiť, či prispôbovanie akusticko-prozodických vlastností reči avatarov alebo robotických systémov svojmu používateľovi môže pozitívne ovplyvniť dôveru tohto

používateľa voči samotnému automatickému systému. Úvodné výsledky sumarizované v Tabuľke 1 poskytujú niekoľko zaujímavých zistení.

Za najdôležitejší výsledok považujeme to, že výsledky oboch experimentov ukazujú, že prispôbovanie v akusticko-prozodických charakteristikách syntetickej reči avatara, alebo iného automatického dialógového systému, má potenciál ovplyvňovať dôveryhodnosť tohto systému v očiach jeho užívateľa.

Zaujímavým a neočakávaným výsledkom je rozdiel v spôsobe, akým je vzťah medzi prispôbovaním a dôveryhodnosťou ovplyvnený výberom akusticko-prozodických vlastností, ktoré sa na rečovom prispôbovaním zúčastňujú. Pri prispôbovaní na v tempe reči sme pozorovali negatívny vzťah - teda že probanti dôverovali prispôsobujúcemu sa avatarovi **menej** ako neprispôsobujúcemu sa - a pri kombinácii tempa reči, výšky a intenzity hlasu sme pozorovali pozitívny vzťah - probanti **viac** dôverovali prispôsobujúcemu sa avatarovi. Tento výsledok poskytuje niekoľko zaujímavých hypotéz na ďalšie testovanie.

Je možné, že tempo reči je v princípe odlišná charakteristika ako kombinácia výšky tónu a intenzity v tom, ako ovplyvňujú vnímanie aj realizáciu vzťahu užívateľa k automatickému systému. Táto hypotéza je podporená v rámci embodimentu kognitívnej vedy v tom zmysle, že výška a intenzita hlasu sú fyziologicky silne prepojené, kde napríklad väčší sub-glotálny tlak má za následok vyššiu intenzitu a výšku základného tónu, a naproti tomu tempo reči je vo veľkej miere nezávislé od ostatných dvoch charakteristík.

S tým súvisí možnosť, že pozitívne vnemy, ktoré sa môžu spájať so zvýšenou mierou dôveryhodnosti v dialógoch, sa viažu u ľudí na asymetrické charakteristiky tempa reči ale na symetrické charakteristiky výšky a intenzity hlasu. V experimente 1, kde sa prispôbovalo len tempo reči, tak prevládol negatívny vzťah a v experiment 2 hlasové charakteristiky prevážili nad tempom reči, čo vyústilo do pozitívneho vzťahu.

Keďže sme v oboch experimentoch pozorovali len minimálne rozdiely medzi váhaným a neváhaným skóre, domnievame sa, že to naznačuje stabilitu vzťahu medzi prispôbovaním a dôveryhodnosťou v tomto scenári komunikatívnej hry.

Je tiež dôležité spomenúť, že preferencia voči prispôsobujúcemu (alebo neprispôsobujúcemu) avatarovi nemusí automaticky znamenať väčšiu dôveryhodnosť a môže byť spojená s (iba) príjemnejším hlasom toho ktorého avatara. V hre je však hráč motivovaný dosiahnuť čo najlepší výsledok a musí sa spoliehať na radu jedného z avatarov a tak výber avatara (s príjemnejším hlasom) je úzko prepojený na dôveryhodnosť, a my teda skúmame či prispôsobujúci hlas je viac dôveryhodnejší.

V budúcnosti chceme skúmať aj to, či vzťahu medzi rečovým prispôbovaním a dôveryhodnosťou je

ovplyvnený osobnostnými charakteristikami probanta. Ako bolo popísané v Podkapitole 3.2, máme pre každého probanta dvoch experimentov informácie zo zjednodušeného osobnostného testu a plánujeme preskúmať, či pozitívny a negatívny vzťah medzi prispôbovaním a dôveryhodnosťou je posilnený alebo oslabený u ľudí s rôznymi osobnostnými charakteristikami.

PodĎakovanie

Tento príspevok vznikol s podporou grantovej agentúry Air Force Office for Scientific Research (Air Force Materiel Command, USAF) v rámci grantu číslo FA9550-15-1-0055 a sčasti aj VEGA v rámci grantu 2/0197/15. Moje poďakovanie patrí aj spolupracovníkom Ramirovi Galvezovi a Agustinovi Gravanovi z University of Buenos Aires, Rivke Levitan z CUNY Brooklyn College v New Yorku, a Marianovi Trnkovi a Milanovi Ruskovi z UI SAV Bratislava.

Literatúra

- Bell, L., Gustafson, J., a Heldner, M. (2003). Prosodic adaptation in human-computer interaction. V zborníku ICPhS, str. 2453–2456.
- Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognitive Computation* 6(4): 802-813.
- Beňuš, Š., Gravano, A. a Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, 43(12): 3001-3027.
- Boersma, P. a Weenink, D. (2017). Praat: doing phonetics by computer [computer program]. <http://www.praat.org>.
- Branigan, H., Pickering, M., a Cleland, A. (2000). Syntactic co-ordination in dialogue. *Cognition* 75: B13–B25.
- Branigan, H.P., Pickering, M.J., Pearson, J. a McLean, J.F. (2010). Linguistic alignment between humans and computers. *Journal of Pragmatics*, 42: 2355–2368.
- Brennan, S. (1996). Lexical entrainment in spontaneous dialog. V zborníku *International Symposium on Spoken Dialog (ISSD)*.
- Giles, H., Coupland, N. a Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. V H. Giles, N. Coupland a J. Coupland *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge: Cambridge University Press, 1–68.
- de Jong, N.H. a Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41: 385.
- Edlund, J., Heldner, M., a Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. V zborníku *Interspeech*.
- Gosling, S., Rentfrow, P. a Swann, W. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality* 37(6): 504–528.
- Levitan, R. (2014). Acoustic-prosodic entrainment in human-human and human-computer dialogue. Ph.D. dissertation, Columbia University.
- Levitan, R., Beňuš, Š., Gravano, A. a Hirschberg, J. (2015). Entrainment in Slovak, Spanish, English, and Chinese: A cross-linguistic comparison. V zborníku *SIGdial*, str. 325-334.
- Levitan, R. a Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. V zborníku *Interspeech*, str. 3081–3084.
- Lubold, N., Walker, E. a H. Pon-Barry. (2016). Effects of Voice-Adaptation and Social Dialogue on Perceptions of a Robotic Learning Companion. V zborníku *ACM/IEEE International Conference on Human Robot Interaction*, str. 255-262.
- Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. V zborníku *Speech Prosody*.
- Nagy, M. (2006). Penguin Quart-Slovak digit speech recognition game based on HMM Marek Nagy. V zborníku *Artificial Intelligence Applications and Innovations*, str. 179-186.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119 (4): 2382–2393.
- Pickering, M. a Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36 (4) :329-347.
- Reitter, D., Keller, F. a Moore, J. 2006. Computational modelling of structural priming in dialogue. V zborníku *HLT/NAACL*.
- Strupka, E., Nieburgh, O. a Fischer, K. (2016). Influence of Robot Gender and Speaker Gender on Prosodic Entrainment in HRI. V zborníku *IEEE International Symposium on Robot and Human Interactive Communication*.
- ten Bosch, L., Oostdijk, N. a Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Communication* 47: 80–86.