VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ
FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

NÁVRH UŽIVATELSKÝCH TESTŮ V ROBOTICE

DESIGN OF USER STUDIES IN ROBOTICS

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — MARTIN JUŘÍK
AUTHOR

VEDOUCÍ PRÁCE — Ing. ZDENĚK MATERNA
SUPERVISOR

BRNO 2016

Vysoké učení technické v Brně - Fakulta informačních technologií

Ústav počítačové grafiky a multimédií Akademický rok 2015/2016

Zadání bakalářské práce

Řešitel: **Juřík Martin**
Obor: Informační technologie
Téma: **Návrh uživatelských testů v robotice**
Design of User Studies in Robotics
Kategorie: Uživatelská rozhraní

Pokyny:

1. Seznamte se s principy návrhu uživatelských testů pro ověřování grafických uživatelských rozhraní.
2. Seznamte se se specifiky návrhu uživatelských testů pro potřeby robotiky.
3. Vyhledejte a analyzujte několik publikovaných článků zahrnujících uživatelské testování.
4. Popište metodiky a nástroje používané pro návrh a realizaci uživatelských testů.
5. Proveďte návrh jednoho nebo více uživatelských testů.

Literatura:

- Steinfeld, Aaron, et al. "Common metrics for human-robot interaction."*Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 2006.
- Chen, Jessie YC, Ellen C. Haas, and Michael J. Barnes. "Human performance issues and user interface design for teleoperated robots." *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 37.6 (2007): 1231-1245.

Pro udělení zápočtu za první semestr je požadováno:

- Body 1 až 3.

Podrobné závazné pokyny pro vypracování bakalářské práce naleznete na adrese http://www.fit.vutbr.cz/info/szz/

Technická zpráva bakalářské práce musí obsahovat formulaci cíle, charakteristiku současného stavu, teoretická a odborná východiska řešených problémů a specifikaci etap (20 až 30% celkového rozsahu technické zprávy).

Student odevzdá v jednom výtisku technickou zprávu a v elektronické podobě zdrojový text technické zprávy, úplnou programovou dokumentaci a zdrojové texty programů. Informace v elektronické podobě budou uloženy na standardním nepřepisovatelném paměťovém médiu (CD-R, DVD-R, apod.), které bude vloženo do písemné zprávy tak, aby nemohlo dojít k jeho ztrátě při běžné manipulaci.

Vedoucí: **Materna Zdeněk, Ing.**, UPGM FIT VUT
Datum zadání: 1. listopadu 2015
Datum odevzdání: 18. května 2016

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
Fakulta informačních technologií
Ústav počítačové grafiky a multimédií
612 66 Brno, Božetěchova 2

doc. Dr. Ing. Jan Černocký
vedoucí ústavu

Abstrakt

Vytvořit kvalitní uživatelské testování pro potřeby robotiky může být výzva. Tato bakalářská práce se zabývá celkovým procesem návrhu testů pro uživatele. Součástí práce je i detailní popis metod pro testování, které se v robotice využívají. Cílem této práce je pomoci budoucím studentům s uživatelským testováním.

Abstract

It may be challenging to design and perform a valid user study in the field of robotics. This bachelor's thesis gives a brief overview of commonly suggested guidelines and used methods. Its aim is to help future students dealing with user testing to carry out their studies successfully.

Klíčová slova

Metriky, Objektivní metriky, Subjektivní metriky, Metody ohodnocení, Uživatelské rozhraní, Wizard of Oz, Návrh testování, Experiment

Keywords

Metrics, Objective metric, Subjective metrics, Methods of evaluation, User interface, Wizard of Oz, User study, Experiment

Citace

JUŘÍK, Martin. *Návrh uživatelských testů v robotice.* Brno, 2016. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Materna Zdeněk.

Návrh uživatelských testů v robotice

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Zdeňka Materny. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

. . . . . .
Martin Juřík
18. května 2016

Poděkování

Touto cestou bych chtěl poděkovat vedoucímu mé práce panu Ing. Zdeňku Maternovi za jeho ochotu, pomoc a odborné rady, které mi při zpracování této práce poskytl.

© Martin Juřík, 2016.
Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Obsah

1 Úvod **3**

2 Metriky a metody ohodnocení **4**
2.1 Metody ohodnocení 4
2.2 Metody ohodnocení a subjektivní metriky 5
2.2.1 Vlastní ohodnocení 5
2.2.2 Rozhovor 8
2.2.3 Pracovní zátěž účastníků 8
2.3 Metody ohodnocení a objektivní metriky 9
2.3.1 Sledování chování 9
2.3.2 Psychofyzické měření 9
2.3.3 Výkonnost při úkolech 9

3 Důležité faktory pro testování **13**
3.1 Jak správně navrhnout experiment pro testování 14
3.1.1 Obecné informace při vytvoření testování 15

4 Analýza vybraných experimentů **17**
4.1 Objektivní vs. Subjektivní metriky 17
4.1.1 Objektivní metriky 17
4.1.2 Subjektivní metriky 17
4.1.3 Strategie experimentu 18
4.1.4 Závěr 19
4.2 Najdi a zachraň 19
4.2.1 Metody ohodnocení 19
4.2.2 Závěr 20
4.2.3 Doporučení 21

5 Wizard of Oz **23**
5.0.1 Obecné informace 23
5.0.2 Analýza experimentu a SUXES 23

6 Návrh testování **28**
6.1 Protokol 29

7 Závěr **35**

Literatura **36**

Přílohy **38**
Seznam příloh 39

A Obsah CD **40**

B Informovaný souhlas **41**

C NASA-TLX **42**

Kapitola 1

Úvod

Roboti nám v dnešní době dávají možnosti, které jsme si před několika lety nedovedli ani představit. Proto je najdeme ve spoustě oblastech jako je průmysl [10], lékařství, vojenství, ale i záchranné mise [2]. Lidé se snaží nahradit svou účast při situacích, které jsou životu nebezpečné, v situacích, kam se lidská ruka nedostane, nebo jen pro usnadnění si práce.

Než je robot nasazen do plného provozu, musí podstoupit dlouhý proces ladění a testování. Existují různé studie na toto téma, kde se výzkumníci snaží simulovat různé podmínky a následně v těchto podmínkách testovat specifické typy robotů. Tyto roboty testují uživatelé, jejich přínos je tedy značný. Jsou to právě oni, kteří nepřímo rozhodnou o tom, které grafické rozhraní nebo které zařízení je nejvhodnější pro daného robota. Když totiž výzkumnící tvoří závěr, tak většinou právě z jejich výsledků. Aby výzkumníci tyto testy mohli vytvořit, musí si jasně stanovit, jak test bude probíhat, jaké metody využíjí a co bude výstupem daného testování, např. zda uživatel splní úkol rychleji s rozhraním A nebo B a kolik při každém udělá chyb.

Tato práce se zabývá studiem různých metod testování. Dále jaké metody výzkumníci využívají k otestování daného rozhraní a jak takové uživatelské testování vypadá. Práci jsem rozdělil do pěti navazujících kapitol, přičemž první kapitola shrnuje různé nástroje a testování v robotice. V druhé kapitole jsou obecné rady, jak správně navrhnout experiment a s ním i spojené testování. Třetí kapitola obsahuje dva experimenty, které využívají různé metody ohodnocení. Ve čtvrté kapitole je pak ukázán jiný typ testování, konkrétně pro průmyslové robotické rameno s ukázkou experimentu. Práce je zakončena pátou kapitolou, ve které se snažím využít všechny získané informace pro návrh testu uživatelského rozhraní pro robota PR2. Tento test je založen na bakalářské práci z roku 2015.

Kapitola 2

Metriky a metody ohodnocení

Existují různé metody ohodnocení experimentu. Tyto metody využívají tzv. metriky. Ty v robotice dělíme na objektivní a subjektivní. Jejich kombinace [9] jsou velmi používané. Jedná se o měření podle stanovených pravidel při testování robotů a uživatelů.

Objektivní metriky jsou takové, u kterých výzkumník nepotřebuje názor účastníků. Ti jsou měřeni různými metodami pro zjištění výkonnosti při daném úkolu. Jedním z příkladů v Human Robot Interaction (interakce člověka s robotem, dále jen HRI), kde jsou tyto metriky zastoupeny, je „NIST Aréna – najdi a zachraň“ [17]. Tato aréna často slouží jako inspirace pro různé experimenty a testování (viz Kapitola 4.2). Testování spočívá v tom, že je vytvořeno bludiště s různými okolními vlivy a robot v něm musí najit člověka resp. figurínu. Při tomto testování jsou použity metriky na celkový výkon robota s člověkem a nebere se zde v potaz kým byla mise splněna, ale zda byla splněna. To, jestli úkol splnil robot, operátor nebo spolupráce člověka s robotem, není cílem tohoto testu. Metriky, které se zde používají jsou např. čas, počet kolizí a počet překážek, kterým se robot vyhnul [12].

Subjektivní metriky mají za úkol zjistit co nejvíce informací od účastníků testování.

Vytvořit metriku, která by byla natolik všeobecná a šla použít na každý typ robota, však není možné, protože rozmanitost použití robotů je příliš velká. Byly vytvořeny základní typy metrik, které jsou aplikovatelné na různé roboty. Metriky mohou být systémové (systém/robot sám měří hodnoty) a nebo tištěné/elektronické (osoba, nejčastěji výzkumník si dělá poznámky písemně nebo elektronicky – vše měří člověk) [12].

2.1 Metody ohodnocení

Díky těmto metodám lze posoudit, jak lidé reagují na ovládaní určitých robotů nebo které rozhraní je lepší.

Existuje 5 základních metod, které se využívají ve studiích HRI a pracují s metrikami [1]:

1. vlastní ohodnocení
2. rozhovor – nebo také „interview“
3. sledování chování testujícího člověka – video

4. psychofyzické měření
5. výkonnost při úkolech

Nejvíce využívané metody jsou 1. a 3. Obecně platí, že čím více metod je zahrnuto, tím lépe. Základ pro vyhodnocení HRI je metoda „vlastní ohodnocení“. Ve většině případů se jedná o subjektivní metriku ve formě dotazníku [1].

Využít alespoň tři metody ohodnocení je předpoklad dobré studie, hlavně proto, že poskytnou více informací. Důležité je nespokojit se s výsledky z jedné metody ohodnocení (např. pouze z výsledků dotazníku) a to ani tehdy, pokud jsou jasné a bez chyb. Stejné výsledky by se měly potvrdit i v další metodě ohodnocení. Pokud však jsou tyto dva výsledky v rozporu, třetí metoda ohodnocení by měla podpořit jeden z těchto výsledků nebo vysvětlit, proč předchozí dva vůbec v rozporu jsou. Nejhorší je pak možnost, pokud jsou v rozporu všechny tři výsledky. To by znamenalo, že výzkumné otázky mohly být špatně pochopeny, špatně vytvořeny, případně celý plán testování byl špatně navrhnut [1].

Při testování je také důležité mít dobrou měřící techniku pro účastníky, např. pokud se snaží výzkumník zjistit úroveň stresu či úzkost účastníků při interakci s robotem, měl by využít metody vlastní ohodnocení, nahrávaní na video a psychofyzické měření, jako je rychlost dýchaní, vodivost kůže nebo EKG[1] [1].

2.2 Metody ohodnocení a subjektivní metriky

Následující část popisuje metody ohodnocení a jaké konkrétní metriky se v nich využívají.

2.2.1 Vlastní ohodnocení

Jedná se o nejvíce využívanou metodu. Tato metoda zahrnuje psychometriku, dotazníky, nebo průzkumy. Díky této metodě účastníci poskytnou vlastní názor na to, jak se během testování cítili, jakou měli motivaci a všeobecně co si o testu myslí. Tato metoda může poskytnout hodnotné informace, ale často zde bývá problém s validitou či potvrzením dat. Účastníci totiž nemusí odpovědět na otázku podle sebe, ale podle toho, jak by ostatní teoreticky odpověděli nebo také podle toho, jakou odpověď výzkumníci očekávají. Výsledky též mohou být ovlivněny náladou účastníků v daný den. I proto je nutné využít více typů měření, jako je rozhovor, psychofyzické měření či výkonnost v daném úkolu pro pochopení výsledků účastníka [1].

Dotazníky

Nejvíce využívaným vlastním ohodnocením je dotazník. Tento fakt potvrzuje i to, že se objevují téměř v každém testování. Dotazník je metoda kvantitativního výzkumu pro sběr dat [18]. Jak se takové dotazníky dělí shrnuji v následující části i s konkrétními příklady. Jaké otázky využívají výzkumníci v experimentech lze vidět v kapitole 5.0.2.

Otevřené otázky jsou především takové, které nabízejí pouze vlastní textové odpovědi. Jsou určeny ke zkoumání specifických odpovědí respondentů, které není možné vyjádřit jinak než slovním popisem.

[1] ElektroKardioGrafie – snímání elektrické srdeční aktivity

Příklad: Co byste udělal v experimentu jinak? – respondent by pak odpovídal celou větou dle libosti.

Výhody	**Nevýhody**
+ Respondent může poskytnout svůj pohled na věc	− Nelze zanést výsledky do grafu
	− Časově náročné vyhodnocení

Polo-uzavřené otázky je vhodné použít tam, kde je třeba získat konkrétní i textovou odpověď. Textová odpověď je obvykle definovaná jako samostatná odpověď na posledním místě. Respondenti ji vybírají, pokud si nemohou vybrat z nabídky definovaných odpovědí.

Příklad: Jak jste se dozvěděl o tomto experimentu?

a) plakát

b) internet

c) z doslechu

d) jinak *(respondent vyplní jinou možnost)*

Výhody	**Nevýhody**
+ Přehlednost a snadnost vyplnění	− Do výsledkových grafů nelze zpracovat textové odpovědi
+ Možnost textové odpovědi pro přesné výsledky	

Uzavřené otázky jsou nejvíce využívané v dotaznících, v nichž respondent vybírá jednu nebo více z předpřipravených odpovědí. Díky jejich jednoznačnosti a jednoduché konstrukci je velmi snadné je rychle a efektivně zpracovat.

Příklad dichometické otázky: Jakého jste pohlaví?

a) muž

b) žena

Výhody	**Nevýhody**
+ Snadnost a rychlost vyplnění	− Náročnější formulace otázky a možnosti odpovědi
+ Snadné zpracování a vyhodnocení	− Není vhodná pro otázky s velkým množstvím odpovědí

U trichotomické otázky přibývá třetí možnost, většinou neutrální. Tento typ otázky se volí tehdy, pokud by respondent mohl mít problém z určení jasné odpovědi. Dalším typem jsou otázky polytomické (Likertova škála). Ty nabízí respondentům více jak tři odpovědi a mohou být vytvořeny jako výběrové nebo výčtové.

U výběrové otázky respondent vybírá pouze jednu odpověď za účelem získání jasné odpovědi. Výčtová otázka nabízí respondentovi možnost vybrat jednu nebo více odpovědí.

Likertova škála je oblíbený typ otázek, kdy je respondent dotázán, aby vyjádřil svůj názor. Často se objevuje v dotaznících po skončení experimentu [5]. Jedná se o speciální polytomickou otázku, která se snaží zjistit postoj, spokojenost nebo zkušenost respondenta. Tvoří se podobně jako uzavřená otázka s výběrem možností, avšak odpovědi jsou pevně dané a většinou jich bývá pět (také se vyskytuje se sedmi a více odpověďmi) [15].

Příklad:

	Velmi spokojen		Nejsem ani spokojen ani nespokojen		Velmi neposkojen
Jak jste byl spokojen s ovládáním robota?	○	○	○	○	○

Výhody
+ Snadné vyhodnocení

+ Snadno pochopitelné
+ Nenutí se rozhodnout mezi stanovenými odpovědmi

Nevýhody
− Odpověď dané otázky může být ovlivněna předchozí odpovědí

Matice je druh uzavřené otázky, která využívá Likertovu škálu. Na každém řádku obsahuje buď vlastní otázku nebo body předotázky. Pro každou položenou otázku jsou ve sloupcích připraveny stejné odpovědi. Předotázkou rozumíme blíže nespecifikovanou otázku nad samotnou maticí otázek. Ta se vztahuje k jednotlivým bodům.

Příklad:

Do jaké míry souhlasíte s těmito výroky?	Souhlasím	Spíše souhlasím	Nevím	Spíše nesouhlasím	Nesouhlasím
Nevadilo mi strávit experiment v těsném prostředí.	○	○	○	○	○
Ve stísněném prostoru jsem se cítil nepříjemně.	○	○	○	○	○

Výhody
+ Umožňuje sjednotit více podobných otázek

Nevýhody
− Příliš velká matice může působit nepřehledně
− Respondent se může lehce v matici ztratit

Řazení je další specifickou otázkou. Respondent řadí jednotlivé odpovědi podle sebe, kdy na vrchol z pravidla umisťuje pro sebe nejvhodnější možnost.

Příklad: Seřaďte následující zařízení podle toho, jak Vám nejvíce vyhovovalo.

- Volant
- Joystick
- Tablet
- Gamepad

Výhody	**Nevýhody**
+ Respondent vytvoří pomyslný osobní žebříček odpovědí	− Složitější zpracování dat

2.2.2 Rozhovor

Je nepřímo spojený s vlastním ohodnocením. Rozhovor může být strukturovaný nebo nestrukturovaný. U prvního zmíněného je položeno několik otázek, kdy všechny tyto otázky jdou ve stejném pořadí každému účastníkovi testování. Nestrukturovaný rozhovor funguje tak, že jsou otázky měněny na základě odpovědí účastníka. Například se může více rozvinout předešlá otázka. Rozhovor nám může pomoci rozšířit informace získané již při vlastním ohodnocení. Na druhou stranu se mohou objevit problémy – jsou 3 typy odpovědí [1]:

- Pouze kladné – účastník na vše přikyvuje
- Pouze záporné – účastník vše odmítá
- Vhodná odpověď – účastník odpoví to, co si myslí, že je správné

Pokládat otázky při rozhovoru, které by přinesly kvalitní odpovědi, může být těžší, než se zdá. Určitě je dobré nahrávat rozhovor na video, případně nahrávat jen zvuk pro pozdější vyhodnocení.

2.2.3 Pracovní zátěž účastníků

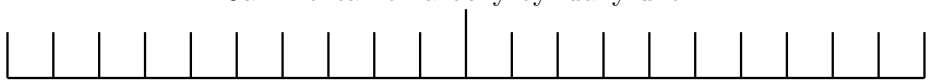
Posouzení pracovní zátěže se dá získat díky nástroji NASA-TLX [16]. Ten obsahuje šest jednoduchých kategorizovaných otázek a lze jej využít jak v tištěné, tak i elektronické podobě. Jedná se o vysoce aplikovatelnou metodu, kterou lze použít na spoustu systémů založených na spolupráci stroj-člověk. NASA-TLX byl použit např. při testování semi-autonomních robotů, kde se ukázalo, že více autonomie přináší menší zátěž [12]. Tento výsledek však nemusí potvrdit výsledky jiných metod [9].

Skládá se z 6 otázek, které mají danou škálu odpovědí a má následující kategorie:

- mentální nároky
- fyzické nároky
- časové nároky
- výkon
- námaha
- frustrace

Příklad otázky na mentální nároky:

Jak mentálně náročný byl daný úkol?



Velmi málo — Velmi náročný

Celý dotazník je přiložen jako příloha C.

2.3 Metody ohodnocení a objektivní metriky

Následující část popisuje často využívané metody ohodnocení v robotice a také jak vypadají konkrétní objektivní metriky.

2.3.1 Sledování chování

Druhá nejvíce využívaná metoda ohodnocení, která se občas vyskytuje současně s psychofyzickým měřením nebo s vlastním ohodnocením. Jedná se o metodu, která využívá nahrávání účastníků na video. Výhodou této metody je, že můžete nahrát chování účastníka a nemusíte se tak spoléhat na jeho názor (proč se zachoval tak, jak se zachoval u některého úkolu apod.). Nahrávky jsou poté často hodnoceny dvěma nebo více hodnotiteli. Nicméně nutno podotknout, že účastníci ve většině případech ví, že jsou nahráváni a u některých to může způsobovat nervozitu nebo nesoustředění se na daný úkol [1]. Výhodou je, že z této metody se dá po skončení experimentu extrahovat hned několik metrik na celkový výkon při plnění úkolu.

2.3.2 Psychofyzické měření

Výhodou je, že člověk nedokáže ovlivnit autonomní nervový systém a proto by výsledky měly být přesné. Navíc toto měření není tak stresující jako například nahrávání na kameru. Měření však nemusí mít očekávaný výsledek a to může vést až k chybnému vysvětlení dat. Proto je nutné před začátkem testu položit účastníkům otázku jak se cítí, jestli jsou zdraví, stav mysli apod., aby se pak provedly případné úpravy výsledků [1]. Zahrnuje především měření tepové frekvence nebo také kožní odpor.

2.3.3 Výkonnost při úkolech

Kategorie úkolů, které se používají při uživatelském testování. Každá tato kategorie obsahuje specifické metriky [12].

Navigace Úkol dostat robota z bodu A do bodu B. Důležité je vědět kde robot je, kam má robot jít, jak se tam dostane a jak se vypořádá s prostředím a překážkami.

Vnímání Cílem úkolu je porozumět vzdálenému prostředí. Využití je důležité pro hledání, dohled, identifikace cílu. Tento úkol vyžaduje interpretování senzorových dat.

Řízení Cílem úkolu je zkoordinovat akce a chování člověka a robota, ať už je to chování nezávislé nebo skupinové.

Manipulace Úkol zaměřený na interakci robota s prostředím. Nejčastěji se jedná o robota, který má uchopovací rameno. Využití najde v geologii, konstrukci nebo při likvidaci munice. Pro provedení tohoto úkolu je nutné přesně definovat, co bude ovlivněno – jaká věc. Dále se musí specifikovat, co bude provedeno, např. toto polož sem.

Socializace Úkol, který by měl znatelnou sociální interakci. Využití zahrnuje např. „tour guiding“ (rady při vycházkách), starání se o zdraví (terapie, pomoc při pohybu), zábavu a testování lidské inteligence. Vytvořit takový úkol vyžaduje rozpoznávání uživatele, chápání sociální komunikace a chování.

Úkolové metriky

Následující podkapitola shrnuje důležité informace o objektivních metrikách, které se využívají ve zmíněných kategoriích. Vzhledem k tématu bakalářské práce nebudou některé úkoly a jejich metriky podrobně rozebírány. Také se nebudu dál věnovat oblasti sociální robotiky.

Tyto metriky lze využít při uživatelském testování.

Navigační úkoly a metriky

Pro testování navigace jsou vytvářeny různé úkoly na přesun, např. dostat se na určitou destinaci a přitom se vyhnout určitým překážkám. Pro takové úkoly jsou doporučovány následující objektivní metriky, které poskytnou informace, jak si operátor nebo robot vedl. Tyto metriky mohou být aplikované na autonomního i na semi-autonomního robota [12]:

- procenta navigačních úkolů, která byla splněna
- odchylky od plánované trasy
- překážky, kterým se robot vyhnul
- překážky, kterým se robot nevyhnul

Nejvíce používaná metrika spojená s měřením účinnosti při navigačních úkolech je čas. Metriky spojené s časem:

- potřebný čas ke splnění úkolu
- operátorův čas na úkol
- průměrný čas, než se robot dostane přes překážky

Další možné metriky použité v navigačních úkolech:

- počet zasáhnutí operátora za jednotku času (toto zasáhnutí může být plánované nebo neplánované)
- zprůměrovaný počet zásahů operátora za jednotku času
- průměrný čas potřebný pro zásah
- poměr časů operátor : robot – pokud operátor 10 minut nastavuje navigační plán, který by robotu umožňoval úspěšnou navigaci na jednu hodinu, tak je poměr 1 : 6

Úkoly na vnímání a metriky

Při úkolech mohou být závěry vykonány robotem (lokalizace, detekce překážek, atd.), člověkem (např. identifikace oběti na video záznamu), nebo jejich kombinací např. robot směrující operátorovu pozornost na cílovou oblast. Závěr je však vytvořen člověkem.

HRI metrika pro vnímání se dá rozdělit na pasivní vnímání (interpretace získaných sonzorových dat) a aktivní vnímání (pohybuje se od relativně pasivních úkolů jako posuv a náklon kamery až k ovládání robotova pohybu při hledání) [12].

Shrnutí důležitých metrik

Pasivní vnímání Metrika kolik % objektů bylo detekováno při vykonávání úkolu.

Aktivní vnímání Zahrnuje aktivní identifikaci a stacionární hledání.

Aktivní identifikace – měří výkon při rozpoznávacích úkolech zahrnujících mobilitu.

Objektivní metriky aktivní identifikace:

- výkonnost: čas nebo úsilí k potvrzení identifikace
- úsilí: počet pohybů kamery

Stacionární vyhledávání – měří výkon vyhledávacích úkolů, které nezahrnují mobilitu. Může obsahovat ovládání kamery nebo spojování dat.

Objektivní metriky pro stacionární hledání:

- výkonnost jako čas k vyhledání

Aktivní vyhledávání – měří výkon vyhledávacích úkolů zahrnujících mobilitu.

Objektivní metriky pro aktivní vyhledávaní:

- výkonnost: vynaložený čas a úsilí (např. pro identifikaci cíle)
- identifikační chyby: počet nesprávných cílů, počet vynechaných cílů, atd.

Řízení a využívané metriky

- Fan Out – typ meření, který definovali ve studii Goodrich a Olsen [3]. Snaží se zjistit, kolik robotů se stejnými dovednostmi může být efektivně kontrolováno člověkem.
- Čas než dojde k zasáhnutí – pokud operátor nevěnuje pozornost robotovi, může se objevit časové zpoždění. To se vytvoří v případě, že robot zjistí problém a operátor na něj ihned nezareaguje, právě kvůli nepozornosti. Zásah operátora může být fyzický (mechanická manipulace) nebo kognitivní (je potřeba učinit nějaké rozhodnutí, pomoc při rozeznávání nějaké situace). Tato metrika se objevuje v případech, že operátor ovládá více robotů najednou a přepíná mezi nimi [12].

Manipulace a využívané metriky

- Náraz – klíčová metrika v téměř všech manipulačních úkolech. Počet neúmyslných kolizí mezi manipulátorem a prostředím (včetně úkolových objektů). Typ nárazu se dělí na tvrdý (silný náraz, který by mohl poškodit objekt, či robota) nebo měkký (jemný naraz s objektem, bez újmy) [12].
- Čas – další klíčová metrika, např. jak dlouho trvá přesunout objekt z místa A do místa B [9].

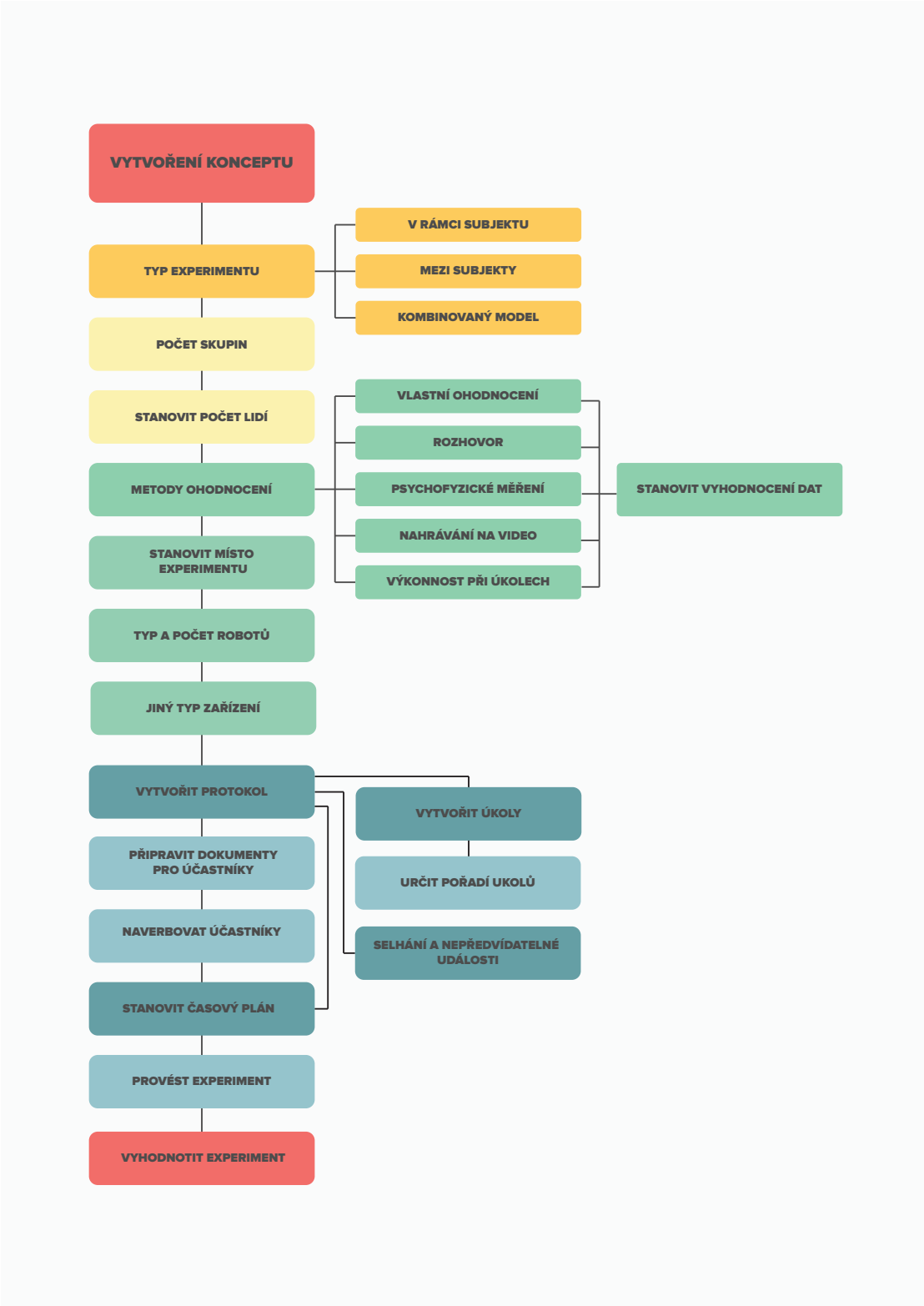
Kapitola 3

Důležité faktory pro testování

Úspěšné testování vyžaduje zvážení následujících faktorů [1].

- Co je cílem testování neboli, co se chce daným testováním dokázat.
- Kolik skupin bude zahrnuto pro testování.
- Kolik bude potřeba účastníků na jednu skupinu.
- Jaká metoda ohodnocení bude použita. (vlastní ohodnocení, atd.)
- Jaký typ prostředí nebo prostory budou potřebné.
- Jaký typ robota se použije pro testování.
- Kolik různých robotů se použije.
- Jaké vybavení bude potřeba. (počítač, přístroje na měření chování, nahrávací zařízení apod.)
- Jakým způsobem by se vyřešily/zvládly chybové situace. (něco v testu selže, kamera, robot apod.)
- Jaké úkoly budou účastníci plnit.
- Jakým způsobem budou účastníci hledáni.
- Jaký typ asistence bude potřeba.

3.1 Jak správně navrhnout experiment pro testování



Obrázek 3.1: Schéma procesu vytvoření experimentu [1]

Celý proces lze vidět na obrázku 3.1. Pro začátek je nutné určit, o jaké vyhodnocení testů půjde. To se většinou určuje na základě hypotézy. Výzkumník by si měl být vědom toho, co chce daným testem dokázat. Existují 3 možnosti, jak navrhnout experiment:

V-rámci-subjektu Nejvíce rozsáhlá možnost je v-rámci-subjektu [1], ve které má každý účastník stejné podmínky. Jinými slovy, každý účastník projde stejně navrženým testem. Tato metoda není tak časově náročná, protože jsou vytvořeny jedny podmínky pro testování. Dalším důležitým faktem je to, že je potřeba méně účastníků. Nicméně má to i svá proti v tom, že u některých testů by se u účastníků mohl projevovat tzv. učící efekt. A to v úkolech, které jsou si podobné nebo stejné. Tomu se dá vyhnout např. tím, že testové úkoly jsou prováděny v náhodném pořadí [14]. Člověk se učí a to, co člověku poprvé přijde složité, se mu na podruhé může zdát mnohem lehčí. Obzvláště, když dostane stejný nebo podobný úkol, který má vykonat. Tento způsob byl použitý například v článku [10], kdy byla použita právě jedna skupina pro testování různých vstupních modalit.

Mezi-subjekty Při metodě mezi-subjekty jsou vytvořeny nejméně dvě skupiny. Počet skupin závisí na počtu testovacích podmínek a účastníci ze skupin zkouší pouze jednu testovací podmínku (jedna podmínka v rámci skupiny). Výsledky mezi skupinami jsou poté srovnávány. Jelikož jsou účastníci vystaveni pouze jedné z podmínek testování, neměl by se u nich projevovat učící efekt a ani možnost čerpání znalostí z předešlého úkolu. Limity této metody jsou v tom, že porovnání skupin nemusí přinést očekávané výsledky a to hlavně proto, že některé výsledky skupiny mohou být ovlivněny některým z účastníků. Ti mohou mít více zkušeností s danou problematiku apod. U této metody je základem dobře rozdělit skupiny. Nejlepší volbou by byl vstupní test, který by zjistil, jaké zkušenosti mají účastníci s danou problematikou a na základě těchto výsledků by pak účastníci byli rovnoměrně rozdělení do skupin. Tím se předejde problémům silná skupina/slabá skupina. Další nevýhodou může být to, že je potřeba dvojnásobný počet účastníků [1].

Kombinovaný model Kombinace dvou výše zmiňovaných metod. Tato metoda se například používá při testování dvou a více robotů, nebo pokud existuje více testovacích podmínek [1].

3.1.1 Obecné informace při vytvoření testování

Protokol postupu

Protokol, který bude obsahovat kompletní informace o průběhu testování od začátku až do konce. Časový harmonogram, počet úkolů, v jakém pořadí se úkoly budou provádět, jak a kde budou ukládána/zpracovávána data z testování atd. Je dobré tento protokol dát přečíst účastníkům nebo je alespoň obeznámit s nadcházejícími částmi protokolu a celkovým testováním.

Lokace a prostředí

Základem je určit, v jakém prostředí se bude testování provádět. Zda to bude venkovní prostor, laboratoř, virtuální prostředí nebo online. Prostředí by mělo reprezentovat co nejvíce reálnou situaci. Účastníci se pak chovají přirozeně. Při testování nemusí být špatné vytvořit specifické světelné podmínky či pouštět různé zvukové nahrávky pro věrohodnější situaci.

Typ a počet robotů

Je důležité vědět, jaký typ a v jaké situaci robota chceme testovat a hlavně kolik robotů při testování využijeme.

Vybavení

Stanovit si, jaké vybavení bude pro testování potřeba. Je nutné počítat s tím, že vybavení při testování může selhat (mechanická závada apod.). Je proto nutné mít náhradní variantu, např. více kamer, rezervní přístroj na měření tepu atd.

Jak správně zvolit účastníky?

Podle toho, co se daným testováním snažíme dokázat. Asi nejlepší způsob je hledat účastníky na některé z fakult. Složit skupinu/skupiny jen z technických fakult případně pracovat s lidmi, kteří jsou obeznámeni s touto problematikou, by však nemuselo přinést výsledky, které potřebujeme. Stejně tak není dobré říct svým kamarádům a známým (experimentátor by se mohl chovat při testování jinak ke kamarádům než u neznámých lidí – může mít potřebu radit a pomáhat). Proto zkusit publikovat plakáty nebo inzerce na veřejných portálech. Zajímavou možností mohou být i děti. Zde se naskytuje možnost zeptat se žáků základní školy (alespoň 8. – 9. třída). Tím získáme velké zastoupení různých lidí pro testování. Je vhodné do testování zahrnout také ženy. Celkový počet se odvíjí od toho, jaký typ vyhodnocení se bude provádět (v-rámci-subjektu, mezi-subjekty atd.). Také je dobré nechat si některé účastníky jako rezervu v případě, že by někdo nepřišel.

Kapitola 4

Analýza vybraných experimentů

Následující kapitola obsahuje shrnutí dvou konkrétních experimentů. Při prvním uživatelé testovali pozemního robota opatřeného ovládacím ramenem. Druhý experiment, který byl inspirován NIST arénou, testoval dva různé roboty při misi najdi a zachraň.

4.1 Objektivní vs. Subjektivní metriky

Následující experiment byl proveden v článku [9], ve kterém výzkumníci využívají objektivní i subjektivní metriky. Cílem experimentu je zjistit výkonnost (viz použité metriky níže) třech různých strategií a také, zda objektivní a subjektivní metriky budou vykazovat stejné výsledky.

4.1.1 Objektivní metriky

- čas ke splnění úkolu – nahrávaní na video (pokud byl úkol proveden úspěšně, byla nahrávka uchována)
- počet chyb – každé selhání bylo poznamenáno
- manipulace s ramenem – během úkolu byla nahrávána, čas se spustil ve chvíli, kdy byl předmět uchopen a ukončil se, když byl předmět upuštěn

4.1.2 Subjektivní metriky

- NASA-TLX – účastníci po provedení každého úkolu vyplnili NASA-TLX
- rozhovor – účastníci byli po skončení experimentu dotázaní na popsání každé jednotlivé strategie

Pro tento experiment byl využit pozemní robot s ovládacím ramenem, který fungoval ve dvou operačních módech. Pohybový mód (přesun robota) a manipulační mód (ovládání ramene).

Dvacet čtyři účastníků si vyzkoušelo ovládání tohoto robota přes vzdálené prostředí pro tři nastavené strategie. Plnili jednoduchý úkol, který zahrnoval pohybový i manipulační mód. Úkolem bylo cestovat s robotem přes místnost k určitému bodu a tam přesunout objekt. Aby účastníci dokončili úkol, museli se nejprve dostat skrze překážky. Jakmile se dostali na destinaci, vzali daný objekt a přesunuli jej do koše.

4.1.3 Strategie experimentu

Strategie 1: Hlavní pracovní prostor

S touto strategií se robot automaticky přepíná mezi módy podle toho, kde se nachází. Pokud je robot příliš daleko od svého cíle, tak operátor robota přesune blíže k pracovnímu místu. Pokud je robot v pracovním prostoru, tak se robot sám přepne na manipulační mód.

Strategie 2: Dosah ramene určí prostor

Dosah ramene určuje kontrolní mód. Systém se automaticky přepne z manipulačního módu na pohybový, když se rameno natáhne na svou maximální délku a nedosáhne na potřebné místo v pracovním prostoru.

Strategie 3: Přidělení kontroly

Pokud je robot mimo pracovní prostor, je aplikovaná regulace rychlosti. V tomto případě zůstane rameno zafixováno v maximální poloze dosahu. Vozidlo pak následuje konec ramene. (robotické kleště)

Metodou testování byl zvolen v-rámci-subjektu. Výzkumníci jednotlivé strategie střídali v následujícím pořadí, aby eliminovali učící efekt.

- 1/3 účastníků testovala módy v pořadí: S1-S2-S3
- 1/3 účastníků testovala módy v pořadí: S2-S3-S1
- 1/3 účastníků testovala módy v pořadí: S3-S1-S2

Zpětná vazba U strategie 1 byli téměř všichni účastníci zmateni, když nevěděli, zda používají v danou chvíli vozidlo nebo rameno. Poznamenali, že tento systém se ovládal velmi složitě.

U strategie 2 účastníci věděli, kdy se vozidlo bude pohybovat, protože by rameno museli nastavit do jeho limitu, aby se přepnulo do pohybového módu. Většinu času proto používali jen pohybový mód. Využili toho, že rameno bylo vysunuté na svou maximální délku. Poté se s koncovou rukou dostali za objekt a rameno vrátili tak, aby objekt uchopili.

Všichni účastníci konstatovali, že strategie 3 byla nejvíce intuitivní. Účastník zajel s robotem dostatečně blízko a poté vykonával manipulační úkol.

Kvantitativní metriky Aby získali kvantitativní ohodnocení různých přístupů (strategií) tak z naměřených hodnot spočítali medián časů (potřebný čas ke splnění úkolu), počet selhání a medián manipulovatelnosti pro každý úkol účastníka.

	Strategie 1	Strategie 2	Strategie 3
Čas ke splnění	206s	202s	174s
Počet splnění	61	48	23
Manipulovatelnost	0.727	0.829	1.00
NASA-TLX	60.67	56.83	53.67

Tabulka 4.1: Tabulka dat

4.1.4 Závěr

Všechny naměřené hodnoty zanesli do bodových a krabicových grafů (boxplot) a vytvořili závěr. I přesto, že subjektivní metriky NASA-TLX a rozhovor poskytly představu o výkonnosti jednotlivých strategií, tak objektivní metriky podaly pro výzkumníky kvalitnější informace a to hlavně nameřeným údajům. Závěrem bylo, že výsledky objektivních metrik se více shodovaly se zpětnou vazbou získanou z rozhovorů, než s NASA-TLX.

4.2 Najdi a zachraň

Při tomto experimentu [1, 4] se výzkumníci nechali inspirovat NIST arénou.

Celkem 128 účastníků testovalo dva rozdílné roboty používané na úkoly najdi a zachraň. Každý robot pracoval ve dvou módech - standardní a emotivní. Celý experiment byl vykonáván v zatemnělé místnosti se zvukovým doprovodem, který měl simulovat katastrofu. Cílem bylo zjistit, jak účastníci budou reagovat na různé módy robotů.

Výzkumníci zvolili metodu kombinovaný model s tím, že metoda mezi-subjekty se použila na operační módy robota (standardní x emotivní) a v-rámci-subjektu se používal na jednotlivé roboty (Inuktun Extreme-VGTV a iRobot Packbot). Standardní mód fungoval tak, že robot okamžitě přistupoval k účastníkům, měl rychlé a nepravidelné pohyby, zvýšil se do své maximální polohy a poté se otočil od účastníků a sledoval okolní prostředí. Toto chování se ukazovalo jako agresivní a mělo za následek zvýšení vzrušení u účastníků. Při emotivním módu přistupoval robot k účastníkům pomalu a projevoval záje. Veškeré pohyby byly pomalé a robot zůstal ve sníženém postoji. Poté se díval směrem k účastníkovi, aby ukazoval zdvořilost. Kombinovaný model použili z toho důvodu, protože měli čtyři testující podmínky a usoudili, že je jich mnoho pouze pro metodu v-rámci-subjektu. Pokud tedy experiment má příliš moc podmínek, je moudré zvážit variantu kombinovaný model.

Účastníci byli náhodně rozdělení do dvou skupin (ovládán standardně nebo emočně). Každý účastník si vyzkoušel oba roboty.

4.2.1 Metody ohodnocení

Tato studie použila čtyři metody ohodnocení (vlastní ohodnocení – dotazník, nahrávání na video, psychofyzické měření a strukturovaný rozhovor nahrávaný na audio). Bližší informace nebyly publikovány. Výzkumníci využili matematické statistiky, konkrétně testování hypotéz. Výpočty prováděli skrze software G*Power3.

Nahrávání na video bylo prováděno ze čtyř různých kamer za použití nočního vidění a infračervených zařízení:

- pohled na obličej včetně horní části těla
- horní kamera poskytující pohled na scénu
- pohled od účastníka
- pohled od robota

Synchronizovat všechny kamery může být obtížné, proto výzkumníci doporučují synchronizovat kamery například pomocí rozsvěcování světel v místnosti, kde se experiment provádí. Za předpokladu, že kamery mají zabudovaný mikrofon, je další možností použití zvuku.

Při psychofyzickém měření pak využili pět různých signálů – EKG, vodivost kůže, hrudní dýchání, břišní dýchání, krevní tlak a při tom využili Thought Technology ProComp5 Infiniti system[1]. Díky pěti měřeným signálům dokázali lépe pochopit chování účastníků a jejich pocity.

Po provedení experimentu každý účastník absolvoval strukturované interview. Toto interview nebylo publikováno. Z experimentu lze pochopit, že se jednalo o klasické otázky typu, jak se účastník cítil, co se mu nejméně líbilo, zda byl experiment dostatečně reálný apod. Účastníkům byla předem nabídnuta možnost odmítnutí následného zveřejnění videí a audio nahrávek, ve kterých se objevili. Tato možnost by se měla nabízet při každém testování.

Jak bylo uvedeno, výzkumníci se inspirovali NIST arénou a vytvořili podmínky zničené budovy (viz Obrázek 4.1). Účastník testu byl umístěn v menší kabince s potrhanou látkou, která měla za úkol vytvořit pocit stísněného prostoru kvůli autentičnosti. Pro věrohodnost byla přinesena do laboratoře i reálná suť, pro pocit zříceniny. Roboti byli v obou módech předprogramováni, aby pohyb byl konzistentní. Cesta, kterou roboti prošli, byla navržena z video nahrávek z různých experimentů [7, 8]. I proto asi výzkumnici nepoužili žádné metriky specifické pro NIST arénu.

Obrázek 4.1: Připravené prostředí pro experiment [4]

4.2.2 Závěr

Výsledky ukázaly, že účastníci se cítili mnohem klidněji, když robot pracoval v emotivním módu. Celkově působil více přátelsky. Výzkumníci došli k závěru, že využití takového robota při záchranné misi může pozitivně ovlivnit psychický stav oběti nehody, což by mohlo snížit možné riziko šoku.

[1] http://www.thoughttechnology.com/pro5.htm

4.2.3 Doporučení

Autoři článku [1] vytvořili seznam následujících doporučení.

Rady pro provedení experimentu Tyto rady by měli usnadnit celkové provedení experimentu a pomohou zjistit potenciální chyby v návrhu. Nicméně ani nejlepší návrh se neobejde bez problému. Je nutné počítat s náhradním zařízením, které je potřebné pro experiment, případně počítat s časovou rezervou – např. když některý z účastníků přijde pozdě nebo vůbec. Klíčem je mít náhradní plán a očekávat i ty nejhorší scénáře, protože mohou nastat.

1. Vytvořit protokol studie se všemi instrukcemi. Tento protokol je pak možné použít pro vytvoření obecných informací pro účastníka testu (tento dokument taktéž zahrnuje podpisový souhlas). Protokol obsahuje:
   - způsoby hledání účastníků
   - připravený text pro účastníky – představení, co se snažíte dokázat, jak dlouho to zhruba zabere apod.
   - jednotlivé úkoly v pořadí v jakém se budou provádět
   - časový rozvrh
   - informace jak a kam se budou ukládat data
   - jakékoliv další aktivity spojené s experimentem
2. Provést několik zkoušek plánovaného protokolu, dokud všechny problémy nejsou zjištěny a vyřešeny.
3. Ujistit se, že je dostatek všech potřebných zařízení a že jsou v záloze připraveny ihned na výměnu.
4. Snažit se připravit na nečekané události náhradními plány jako je vyřešení selhání zařízení, případně že nedorazí účastníci nebo váš asistent, který může být pro experiment nezbytný.
5. Snažit se počítat s časovým zpožděním např. účastníci přijdou pozdě, selže některé ze zařízení, je nutné ho vyměnit. Nejhorší scénář je pak selhání zařízení, které může vést až ke zrušení experimentu a případné pozdržení celé studie.

Další rada

- Pokud je potřebný asistent pro váš experiment, snažit se hledat někoho, kdo dané problematice rozumí.
- Pokud se jedná o experiment, který zahrnuje psychofyzické měření, nebylo by špatné zvolit například studenty/odborníky na psychologii.

Dopad validních studií na HRI a použití robotů ve společnosti HRI je oblast, která jde stále dopředu a základem kvalitní studie je použít dobré výzkumné metody. Určit vhodné množství účastníků a tři či více metod na vyhodnocení může poskytnout validitu a důvěryhodnost. To může mít dobrý vliv i na veřejnost. Ta mnohem lépe příjme roboty do svých domovů, škol a zaměstnání, pokud bude vědět, že roboti jsou dobře testovaní na základě kvalitních experimentů pro bezpečnost a celkovou účinnost. Nebo mohou budoucí výzkumníci čerpat z kvalitní studie pro jejich experiment a ještě jej vylepšit [1].

Kapitola 5

Wizard of Oz

V předešlých kapitolách se počítalo s tím, že robot je již naprogramován a reaguje na ovládaní uživatele. Jak může vypadat uživatelské testování pro robota, ktery není naprogramován lze vidět v následující kapitole včetně konkrétního experimentu.

5.0.1 Obecné informace

Běžnou metodou používanou výzkumníky pro HRI experimenty je Wizard of Oz [11]. Díky této metodě lze použít při experimentu jakoukoliv úroveň autonomie, protože systém robota (mozek) nahrazuje výzkumník. Základní myšlenka této metody je, že účastník by si měl myslet, že používá reálný systém. Nicméně někteří výzkumníci jsou trochu znepokojeni z této metody, protože má jisté limity a hlavně se nejedná ani tak o HRI, ale spíše o Human-Human Interaction (interakce člověka s člověkem) za pomocí robota. I tak je tato metoda velkým přínosem a dovoluje výzkumníkům vyzkoušet uživatelské rozhraní bez toho, aby potřebovali naprogramovaného robota. Největší výhoda pak spočívá v tom, že umožňuje otestovat princip bez vlivu konkrétní implementace nebo technologie.

5.0.2 Analýza experimentu a SUXES

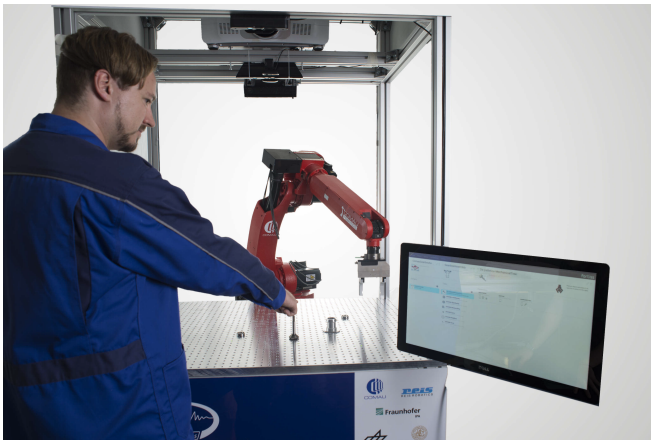
Následující experiment byl proveden v článku Analysis and Semantic Modeling of Modality Preferences in Industrial Human-Robot Interaction [10]. Výzkumníci se v tomto experimentu pokoušeli určit, která z potenciálních modalit by byla nejlepší pro robota, kterého by uživatel dokázal naprogramovat tím, že by ho naváděl při plnění prvotního úkolu.

Pro tento experiment bylo pozváno třicet účastníků, kteří si vyzkoušeli jednotlivé modality.

Modality pro naprogramování robota:

- dotyková obrazovka
- gesta
- řeč
- 3D pero

Jejich hypotéza byla postavena na tom, že využít multimodální přístup usnadňuje naprogramování robotů a že používání gest/3D pera bude preferováno více než dotek a řeč. Pro věrohodnost použili robotické rameno umístěné v pracovním prostoru (viz Obrázek 5.1).



Obrázek 5.1: Pracovní prostor s nenaprogramovaným robotickým ramenem [10]

Návrh

Studie zahrnovala tři časté úkony pro průmyslové roboty používané v malých a středně velkých společnostech:

1. seber a polož
2. skládání součástek
3. svařování

Vyhodnocení bylo založeno na SUXES metodice [13]. Ta umožňuje porovnání, co účastník očekává a jaká je realita. Ohodnocování účastníků při SUXES se skládá ze čtyř fází a osmi kroků (viz Obrázek 5.2). Tento experiment jednotlivé fáze detailně popisoval i s možností nahlédnutí do dotazníků.

Obrázek 5.2: SUXES fáze a kroky [13]

První fáze sestává z úvodu a dotazníku na obecné informace: poskytnout krátký úvod o jednotlivých fázích, co je cílem experimentu a jaké otázky obsahuje dotazník. Jak vypadá takový dotazník viz níže.

Jednoduchý dotazník pro získání základního přehledu o účastnících experimentu:

1. Váš věk?
2. Vaše pohlaví?
3. Jaké jsou vaše schopnosti při využívání počítačů?
   (a) Začátečník: Umím zapnout PC a používat internetový prohlížeč
   (b) Běžný uživatel: Umím využívat většinu programů nainstalovaných na svém PC a zvládám správu souborů
   (c) Pokročilý uživatel: Umím si vyměnit hardwarové komponenty a instalovat nové softwarové vybavení
   (d) Expert: Umím naprogramovat nový software a vím, jak počítačový hardware funguje

Tyto informace slouží k pozdějšímu vyhodnocení. Výzkumníci chtějí vědět, jak si při experimentu vedli ženy a muži nebo například jaké jsou jejich znalosti z oblasti informatiky

Druhá fáze sestává z dotazníku na použití a očekávání.

Následující fáze obsahovala otázky na použití, co účastník od jednotlivých modalit a úkolů očekává. Tato část dotazníku obsahovala otázky na seřazení.

Ukázka dotazníku:

Seřaďte modality podle toho, jak by Vám nejvíce vyhovovaly v jednotlivých úkolech?
Úkol: Vyber objekt

- dotyk
- 3D pero
- ukázat rukou
- řeč

Jedna z otázek na to, kterou možnost by účastník preferoval nejvíce, a kterou méně.

Třetí fáze sestává z experimentu a dotazníku na získané zkušenosti.

Účastníci plnili jednotlivé úkoly, které byly stejné pro všechny modality. Nejdříve si vyzkoušeli dotykové rozhraní, poté gesta následovány řečí a nakonec 3D pero.

Po vykonání všech úkolů následoval další dotazník. S výjimkou úvodní otázky byl dotazník stejný jako u očekávání.

Úvodní otázka dotazníku:

Seřadit modality podle kognitivního zatížení. (Která z modalit potřebovala nejvíce koncentrace, kde jste museli přemýšlet nejvíce). Nejvyšší - velká kognitivní zátěž, Nejnižší - nízká kognitivní zátěž).

- dotyk
- 3D pero
- ukázat rukou
- řeč

Otázka velice podobná té z NASA-TLX. Která se zaměřuje na pracovní zátěž. Výzkumníci se však neptají na náročnost úkolů, ale na konkrétní modality.

Čtvrtá fáze sestává z dotazníku na vlastní názor účastníků. Dotazník se skládal z otázek, které měly pět možností (Likertova škála) a následně pak šest (matice).

Příklad otázky se škálou 1 - 5:

Jaký je váš názor při následujících možnostech?

	O dost lehčí, než jsem očekával		jak jsem očekával		Mnohem složitější, než jsem si myslel
Dotyk.	○	○	○	○	○
Řeč.	○	○	○	○	○
Gesta.	○	○	○	○	○
3D pero.	○	○	○	○	○
Celkově, využívání systému bylo.	○	○	○	○	○

Příklad otázky se škálou 1 - 6:

	Silně nesouhlasím	Nesouhlasím	Víceméně nesouhlasím	Víceméně souhlasím	Souhlasím	Silně souhlasím
Je zábavné programovat robota využíváním několika modalitami.	○	○	○	○	○	○
Systém by neměl dávat na výběr v modalitách, ale vybrat tu, která je nejvíce vhodná pro daný úkol.	○	○	○	○	○	○
Citím se lépe, když je na výběr méně modalit.	○	○	○	○	○	○

Celý dotazník, který byl využit pro toto testování lze nalézt na https://github.com/fortiss/robotics-mmio-preferences.

Závěř

Pro výpočty a grafy výzkumníci využívali metody z matematické statistiky o testování hypotéz. Konkrétně aritmetický průměr, směrodatnou odchylku, p-hodnotu párového t-testu, p-hodnotu nepárového t-testu, p-hodnotu Kruskal-Wallis testu, Kruskal-Wallis test/Dunnův test. Všechny údaje a porovnání zanášeli do grafů.

Cílem dotazníků bylo získat základní informace o tom, co si uživatelé myslí o daném systému a o možnosti různých modalit. Výsledky ukázaly, že uživatelé byli celkově spokojeni s nastavenou obtížností úkolů a s časem, který na ně potřebovali.

Po vyhodnocení výsledků třiceti účastníků se zjistilo, že účastníci nejvíce preferují gesta před 3D perem, dotykem a řečí. Jedním ze závěrů byl snížit počet dostupných modalit pro specifické úkoly. Experti tvrdí, že by se intuitivnost zvětšila, pokud by systém před-vybral nejvhodnější modalitu, zatímco ne-experti by chtěli zkoušet systém a s ním i různé modality. Jak ukazuje tento experiment, dobrý návrh a následné provedení mohou poskytnout od účastníků spoustu přínosných informací.

Kapitola 6

Návrh testování

Následující testování je založeno na bakalářské práci Uživatelské rozhraní pro řízení robota při uchopování objektů od Bc. Romana Manďáka [6]. Cílem jeho experimentu bylo zjistit, které zařízení – 2D myš, 3D myš nebo gamepad, je vhodné pro nejrychlejší naučení ovládaní a jakých výsledků dosahují jednotlivé zařízení. Byl využit robot PR2 (viz Obrázek 6.1), který pracoval ve dvou módech: 3DOF – pohyb po osách x, y, z a 6DOF (přidána i rotace kolem těchto os). Nutno konstatovat, že testování nebylo tématem zmíněné práce a proto má určité nedostatky. Jako typ experimentu zvolil v-rámci-subjektu, ale při počtu 5 lidí nemusel získat věrohodná data. Je možné, že výsledky mohly být ovlivněny náhodou. Abych neprováděl úplně stejný typ experimentu, rozhodl jsem se trochu změnit jeho původní cíl (učící křivka) a zjistit, které zařízení si vede nejlépe při plnění úkolů (čas pro dokončení úkolu, nárazy, celková spokojenost účastníků). Při tomto návrhu jsem se snažil zohlednit všechny rady a metody, na které jsem narazil při studiu.

Obrázek 6.1: PR2 [6]

Jak je uvedeno v kapitole 3.1.1 a 4.2.3, základem je vytvořit protokol pro testování, který bude obsahovat veškeré náležitosti o experimentu a testování. Následující body ukazují protokol pro návrh experimentu doplněné komentářem.

6.1 Protokol

- Testování z BP Návrh testování z Uživatelské rozhraní pro řízení robota.
- Počet robotů – jeden PR2.
- Místo – RoboLab VUT FIT
- Určit, které ze tří zařízení a v jakém módu podává nejlepší výsledky (čas při plnění úkolů, chybovost).
- Úkoly převzaty a mírně rozšířeny pro získání více dat.
- Účastník se snaží splnit jednoduché úkoly, když přemísťuje objekty z jednoho místa na druhé.
- Metriky byly převzaty a mírně rozšířeny (použity subjektivní i objektivní).

Tuto metodu jsem zvolil, abych měl přesnější výsledky neovlivněné postupným zlepšováním účastníků (eliminování učícího efektu). Ti by mohli v poslední fázi testování (při používání třetího zařízení) vykazovat mnohem lepší výsledky, protože se naučili s robotem zacházet z předchozích úkolů s jiným vstupním zařízením. Tyto skupiny budou vytvořeny na základě jednoduchého dotazníku, který vytvoří představu o základních dovednostech účastníků. Toto rozdělení musí být provedeno před samotným testováním, aby se předešlo nesrovnalostem (počet lidí nevyhovuje, skupiny jsou nerovnoměrně rozděleny atd.)

- Dotazník bude online přístupný a bude rozeslán potenciálním účastníkům. Klasický dotazník s uzavřenými a otevřenými otázkami:
    1. Váš e-mail?
    2. Vaše pohlaví?[1]
        (a) muž
        (b) žena
    3. Váš věk?
    4. Jaké jsou vaše schopnosti při využívání počítačů?
        (a) Začátečník: Umím zapnout PC a používat internetový prohlížeč
        (b) Běžný uživatel: Umím využívat většinu programů nainstalovaných na svém PC a zvládám správu souborů
        (c) Pokročilý uživatel: Umím si vyměnit hardwarové komponenty a instalovat nové softwarové vybavení
        (d) Expert: Umím naprogramovat nový software a vím, jak počítačový hardware funguje

[1] Chci vědět, jak si při úkolech vedly i ženy. Také by bylo dobré využít tuto informací pro rovnoměrné rozdělení skupin.

5. Už jste se někdy setkal s 3D myší?
   (a) Nevím, co je 3D myš
   (b) Vím, co to je, ale nikdy jsem ji nepoužíval.
   (c) Ano, už jsem ji používal.
6. Pokud jste se setkali s 3D myší, tak kde?
7. Už jste se někdy setkal s gamepadem?
   (a) Nevím, co je gamepad.
   (b) Vím, co to je, ale nikdy jsem ho nepoužíval.
   (c) Ano, už jsem ho používal.
8. Používali jste gamepad někde jinde, než u počítačových/konzolových her?
   (a) ano
   (b) ne

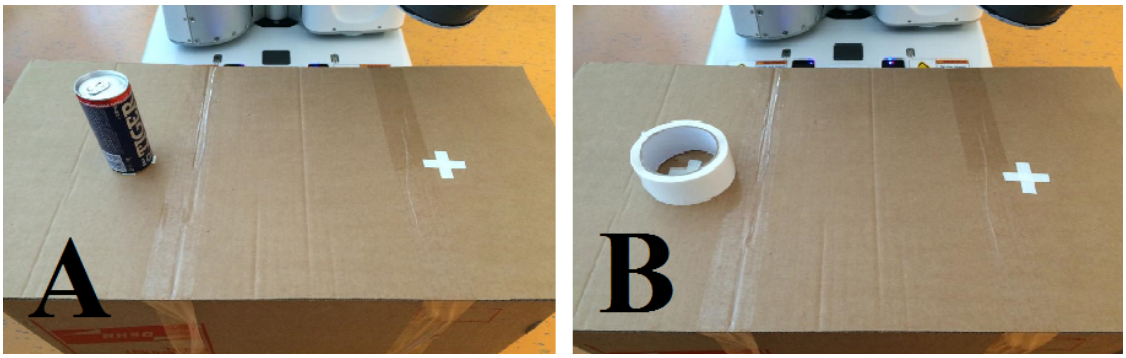
2D myš do dotazníku nezahrnuji, protože lze počítat s tím, že každý běžný uživatel se s tímto zařízením již setkal. Těmito otázkami zjistím základní informace o účastnících. Ty rozdělím rovnoměrně podle toho, jak znají jednotlivá vstupní zařízení. Tím předcházím například tomu, že by se v jedné skupině mohla setkat spousta „hráčů“, kteří jsou s gamepadem hodně sžiti. Po rovnoměrném rozdělení vytvořím časový harmonogram a pošlu mail účastníkům, kdy se mají dostavit a zda jim čas vyhovuje. E-mail v dotazníku slouží pouze jako ID účastníka a jeho zkontaktování.

- Pro testování nebude třeba asistent
- Metriky budou prováděny ručně.
- Objektivní metriky:
  - čas ke splnění úkolu – stopkami se bude měřit čas ke splnění úkolu; časomíra se pouští ve chvíli, kdy účastník provede první tah a končí ve chvíli, kdy je úkol splněn; čas se poté zapisuje do poznámek
  - kolize – ke každému času si budu značit počet kolizí s jiným předmětem nebo s pracovní deskou; nahrávaní posledních úkolů v obou módech na video (ověření výše zmiňovaného měření + metrika počet otočení na pracovní prostor); kameru by bylo dobré umístit vysoko, aby snímala pracovní prostor robota, ale i účastníka – horní část těla (ruky a hlava)

Úkolů a účastníků bude mnoho, proto by se nahrávaly pouze poslední úkoly (v posledním úkolu by měla být výkonnost nejvyšší). Tyto nahrávky budou sloužit k dalšímu vyhodnocení a srovnání průběhu testů jednotlivých zařízení. Chci z videa ověřit výsledky, které získám při ručním měření – čas, počet nárazů. Také mě zajímá viditelné vynaložené úsilí pří úkolech. Při testování si totiž nelze všímat všeho – např. výrazy, pohyby hlavou a pohyby rukou účastníka. Bylo by dobré nahrávku dát posoudit i jiné osobě, která by řekla svůj názor na výkon dotyčného. Další metrikou, kterou bych chtěl aplikovat, je počet pohledů (otočení hlavy na pracovní prostor). Tuto informaci bych také mohl zjistit z videa.

- Subjektivní metriky:
  - vytvořený online dotazník (skrze nějaký portál), který účastník po skončení testu vyplní přes tablet – snazší zpracování dat
  - způsob výběru a hledání účastníků: získat vzorek širší populace → zeptat se na fakultě, navštívit základní školu a zkusit oslovit; dnes poměrně oblíbené sociální sítě by mohly být nápomocné (různé skupiny); tím se dá zajistit široké spektrum účastníků
  - náhradní zařízení/plán: technika může selhat → mít v zásobě od každého zařízení alespoň jeden kus; počítat i s náhradní kamerou; pokud selže tablet s dotazníkem, mít raději i tištěnou verzi dotazníku
  - účastníci mohou dorazit pozdě, nechat 10 minut rezervu – raději ať přijdou dříve
- Úkoly
  - Nejdříve úkoly na 3DOF a poté na 6DOF.
  - Než úkol začne, bude účastníkovi demonstrováno ovládání robota. Poté dostane zhruba dvě minuty na zkoušení ovládání (bez objektů). Následně budou probíhat testovací úkoly.

Originálně byly vytvořeny čtyři úkoly a dva na vyzkoušení 6.2. Zvolené objekty jsem se rozhodl volit od největšího po nejmenší – tím zajistím mírné zvyšování obtížnosti. Pro jednotlivé úlohy střídám jiná místa přesunů. Originálně všechny úkoly byly přesunuty ze stejného místa na stejné místo. Zvýšením počtu testovacích úkolů chci, aby se účastník více sžil s ovládáním pro snazší průběh testu.



Obrázek 6.2: Ukázka testovacích úkolů ze zmíněné BP [6]

Úkoly v módu 3DOF v následujícím pořadí:

1. Test 1. Přemístit válec z bodu A do bodu B.
2. Test 2. Přemístit kostku z bodu C do bodu D.
3. Test 3. Přemístit kvádr z bodu E do bodu F.
4. Přemístit půl litrovou láhev z bodu G do bodu H.

5. Přemístit plechovku z bodu G do bodu H
6. Přemístit puk z bodu I do bodu J
7. Přemístit krabičku z bodu I do bodu J.

Úkoly v módu 6DOF v následujícím pořadí:

1. Test 1. Přemístit kousek trubky z bodu A do bodu B.
2. Test 2. Přemístit návin lepicí pásky z bodu C do bodu D.
3. Test 3. Přemístit misku z bodu E do bodu F.
4. Přemístit půl litrovou láhev z bodu G do bodu H.
5. Přemístit plechovku z bodu G do bodu H
6. Přemístit puk z bodu I do bodu J
7. Přemístit návin lepicí pásky z bodu I do bodu J.
8. Přemístit misku z bodu A do bodu B.
9. Přemístit misku z bodu A do bodu B. Přemístit puk z bodu C do bodu B. (Umístit puk do misky)

- Odhadovaný čas pro jednu osobu je cca 25 minut. Pro testování budou vyhrazeny určité dny. Nelze počítat s tím, že by se v jeden den mohli sejít všichni účastníci.
  - Fáze 1: Úvod do experimentu

    Vyplnění obecného dotazníku proběhlo před testováním, jednotliví účastnici jsou očekáváni. Časový harmonogram byl řešen individuálně a je nastavena 10 minutová časová rezerva, pro případné zpoždění účastníků. Úvodní proslov: Děkuji, že jste přišel/přišla.V průběhu testování si vyzkoušíte ovládat robota za použití vstupního zařízení (2D myš, 3D myš nebo gamepad). Nejdříve Vám bude robot a zařízení představeno a poté začne Váš test, ten bude trvat asi 25 minut. Dále bych Vás chtěl poprosit, abyste si přečetli obecné informace a poté podepsali informovaný souhlas (viz Příloha B).
  - Fáze 2: Představení robota a úkolů

    Před sebou máte robota PR2. Prostřednictvím vstupního zařízení budete plnit šestnáct úkolů, kdy budete přesouvat objekty z jednoho místa na druhé. Těmto úkolům se postupně zvyšuje obtížnost. Nejdříve budete ovládat robota skrze rozhraní, kde budete pracovat s osami x, y, z. Ve druhé části si pak úkoly zopakujete v jiném módu, který přidává možnost otáčení kolem těchto os. Než začnete s testem jednotlivých módů, tak Vám vždy předvedu testovací úkol a vy si poté budete moci na dvě minuty vyzkoušet volné ovládání robota.
  - Fáze 3: Vykonání experimentu a dotazník

    Každá skupina bude vyhodnocena zvlášť a výsledky budou srovnávány. Předběžný plán vyhodnocení:
    * počet nárazů v každém úkolu (každý mód zvlášť)

* počet nárazů celkově
* průměr časů jednotlivých úkolů (každý mód zvlášť)
* celkový počet otočení hlavou v posledním úkolu
* nejčastěji zvolené možnosti v dotazníku
* průměrná známka ovládaného zařízení

Na dotazník byla využita Likertova škála o pěti možnostech.

	Špatné		Nic moc		Skvělé
Zařízení, které jsem ovládal bylo.	○	○	○	○	○
Obtížnost úkolů byla lehká.	○	○	○	○	○
Zvládnul bych těžší úkol.	○	○	○	○	○
Ztrácel jsem při úkolu trpělivost.	○	○	○	○	○

Chci vědět, jestli se jejich tvrzení o obtížnosti shoduje s výsledky.

Poslední zmíněnou otázkou se chci zeptat účastníků na celkový pocit, protože by se během testování mohlo objevit spoustu chyb (robot udělal něco, co jsme nechtěli), někteří účastníci mohou být více labilní. Tyto výsledky by mohly poukázat i na to, zda bylo rozhraní dobře navrhnuto.

Jelikož bylo náplní práce i GUI (Graphic User Interface), tak jsou převzaty a doplněny následující otázky:

	Nesouhlasím	Spíše nesouhlasím	Neutrální	Spíše souhlasím	Souhlasím
Virtuální kurzor byl přesný.	○	○	○	○	○
Vizualizace scény byla přesná	○	○	○	○	○
Vizualizace kurzoru (orientace a viditelnost) byla vhodná	○	○	○	○	○

Následující otázky jsem zařadil proto, abych zjistil, zda si účastníci byli vědomi toho, že se otáčeli na pracovní desku nebo to dělali automaticky pro ověření.

Před-otázka: Co si myslíte o následujícím tvrzení?

	Nesouhlasím	Spíše nesouhlasím	Neutrální	Spíše souhlasím	Souhlasím
Musel jsem si kontrolovat pracovní desku kvůli pozici ramene	○	○	○	○	○
Kontroloval jsem desku častokrát	○	○	○	○	○

Kapitola 7

Závěr

Cílem této práce bylo poskytnout přehled pro budoucí studenty, který by popisoval metody uživatelského testování a s tím i spojené informace o celkovém průběhu testů. Vytvořit kvalitní návrh může byt obtížnější než se zdá, ale stanovený postup a pečlivý přístup a mohou celkový proces vytváření testu zjednodušit.

Tuto práci jsem vytvořil na základě článků, obsahujících informace o experimentech, ve kterých uživatelé testovali různé přístupy a roboty. Také jsem se zaměřil na ukázky konkrétních metrik a při jakých metodách se využívají. Ukázalo se, že nejvíce využívané metriky jsou ty, které využívají ve svém měření čas. Tyto metriky se objevují velice často a lze je označit za nezbytné. Práci zakončuji návrhem testu, který je založený na testování prováděné v bakalářské práci pana Manďáka [6]. Nejprve jsem si stanovil koncept testování a poté si vytvořil protokol o celém postupu. V návrhu jsem se snažil využít všechny dostupné informace, abych vytvořil co nejlepší testování, které by poskytlo kvalitní informace.

Literatura

[1] Bethel, C. L.; Murphy, R. R.: *Review of Human Studies Methods in HRI and Recommendations 2010.* Online, 2010-07-14.

[2] Chen, J. Y.; Haas, E. C.; Barnes, M. J.: *Human Performance Issues and User Interface Design for Teleoperated Robots. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, ročník 37, 2007: s. 1231–1245.

[3] Goodrich, M. A.; Olsen, D. R.: *Seven principles of efficient human robot interaction. Systems, Man and Cybernetics, 2003. IEEE International Conference on*, ročník 4, 2003: s. 3942–3948.

[4] L.Bethel, C.; Murphy, R. R.: *Emotive Non-Anthropomorphic Robots Perceived as More Calming, Friendly, and Attentive for Victim Management. AAAI Fall Symposium Series*, 2010.

[5] Lewis, J. R.: *Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. ACM SIGCHI Bulletin*, ročník 23, 1991: s. 78–81.

[6] Manďák, R.: *Uživatelské rozhraní pro řízení robota při uchopování objektů.* Bakalářská práce, Vysoké učení technické v Brně, 2015.

[7] Murphy, R. R.; Riddle, D.; Rasmussen, E.: *Robot-assisted medical reachback: a survey of how medical personnel expect to interact with rescue robots. Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, 2004: s. 301–306.

[8] Murphy, R. R.; Riddle, D. R.; Burke, J. L.: *Robot-assisted medical reachback: using shared visual information. ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, 2005: s. 635–642.

[9] Pham, C. D.; Phan, H. N. T.; From, P. J.: *Evaluation of Subjective and Objective Performance Metrics for Haptically Controlled Robotic Systems. Norwegian University of Life Scienes*, ročník 35, 2014: s. 147–157.

[10] Profanter, S.; Perzylo, A.; Somani, N.; aj.: *Analysis and Semantic Modeling of Modality Preferences in Industrial Human-Robot Interaction.* IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

[11] Riek, L. D.: *Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. Journal of Human-Robot Interaction*, ročník 1, 2012: s. 119–136.

[12] Steinfeld, A.; Fong, T.; Kaber, D.; aj.: *Common Metrics for Human-Robot Interaction*. HRI 2006 Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, 2006.

[13] Turunen, M.; Hakulinen, J.; Melto, A.; aj.: *SUXES - User experience evaluation method for spoken and multimodal interaction*. INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association.

[14] WWW stránky: *36-309/749 Experimental Design for Behavioral and Social Sciences*. [Online].
URL http://www.stat.cmu.edu/~hseltman/309/

[15] WWW stránky: *Field Research in organizational psychology*. [Online].
URL https://psyc450.wordpress.com/2011/12/05/the-likert-scale-advantages-and-disadvantages/

[16] WWW stránky: *NASA TLX: Task Load Index*. [Online].
URL http://humansystems.arc.nasa.gov/groups/tlx/index.html

[17] WWW stránky: *NIST*. [Online].
URL http://www.nist.gov/el/isd/testarenas.cfm

[18] WWW stránky: *Survio*. [Online].
URL http://www.survio.com/cs/blog/rubrika/serialy

Přílohy

Seznam příloh

A Obsah CD **40**

B Informovaný souhlas **41**

C NASA-TLX **42**

Příloha A

Obsah CD

- Tato práce ve formátu pdf
- Zdrojové kódy tohoto dokumentu v adresáři /bptex
- Plakát

Příloha B

Informovaný souhlas

Testování uživatelského rozhraní robota PR2 za pomoci vstupního zařízení
Obecné informace
Budete pracovat s robotem PR2, který bude na základě vašeho ovládání plnit jednotlivé úkoly. Všechny úkoly jsou tvořeny přesouváním objektu z jednoho místa na druhé. Budete využívat jedno z přidělených zařízení(2D myš, 3D myš, gamepad). Celý test zabere cca 25 minut a bude se skládat ze tří fází.

1. fáze - Úvod do experimentu
2. fáze - Představení robota a úkolů
3. fáze - Vykonání experimentu a dotazník

Souhlasím s účastí na tomto experimentu Fakulty informačních technologií v Brně. Svým podpisem dávám souhlas k tomu, že mohou být pořizovány audio či video nahrávky, které zachycují průběh experimentu. Také souhlasím s tím, že tyto nahrávky mohou být publikované a použity pro další práci. Potvrzuji, má že účast je dobrovolná bez nároku na jakoukoliv finanční odměnu.

Jméno a Příjmení:

V Brně dne:

Podpis:

Příloha C

NASA-TLX

Měření pracovní zátěže. Prosím, vyplňte: Jméno/Příjmení/Úkol/Datum

Mentální nároky: Jak mentálně náročný byl daný úkol?

Velmi málo — Velmi náročný

Fyzické nároky: Jak fyzicky náročný byl daný úkol?

Velmi málo — Velmi náročný

Časové nároky: Jaké bylo tempo úkolu?

Velmi pomalé — Velmi rychlé

Výkon: Jak jste si vedl/vedla při plnění úkolu?

Výborně — Špatně

Námaha: Jak moc jste se musel/musela snažit?

Velmi málo — Hodně

Frustrace: Jak moc jste se citil/citila nevhodně, stresováně, otráveně?

Velmi málo — Hodně