Log Data Analysis of Long-Term Household Trials: Lessons Learned and Pitfalls

Markus Bajones, Astrid Weiss, Markus Vincze
Vienna University of Technology
firstname.lastname@tuwien.ac.at

***Abstract*—In this paper, we present preliminary results from the participants in a long-term household field trial with the service robot HOBBIT. HOBBIT is a robot developed in order to enable independent living at home for older adults. The field trial involved 18 seniors in as many household in total, distributed over Austria, Sweden, and Greece. Users were left alone with the robot for 3 weeks and the field trials were accompanied by a method triangulation consisting of questionnaires, interviews, cultural probing material, and logging data. The focus of this paper is put on a reflection of the log data. What happened when we left the user together with the robot at his/her home and is the log data really the right source to answer this question? Descriptive results show the challenges and difficulties of interpreting and only using log data to gain insights in HRI. Our results show that it is possible to rely on log-data, if detailed understanding of the robot's inner workings exists, and the user's impression of the interaction is taken into account and that it will benefit the understanding during autonomous HRI.**

I. Introduction

Long-term studies with service robots in homes are still rare. There are prominent studies on Roomba vacuum cleaning robots [1] and also first studies with more complex robotic systems, such as Kompai[1]. Studies like these involve numerous methodological challenges that need to be solved for beneficial long-term evaluation of robots with end users. As stated by Lohse et al. [2], in order to improve the interaction with a robotic system, it is crucial to analyze at the same time what happens on the system level (informatics perspective) and on the interaction level (interaction studies and task analysis perspective). Therefore, in order to understand what humans do in the interaction with an autonomous robot and why they do it, it is also crucial to log what is happening on the system side. However, in long-term field trials an analysis at the same time is not possible, as a researcher cannot always be present. In this paper we present a reflection on logging data gathered in three-weeks field trials with 18 households in total, distributed over Austria, Sweden, and Greece. We logged all activities users were performing with the service robot HOBBIT (see section 2 for details) and also all autonomously started events by HOBBIT. We will present the challenges we had to face with this huge amount of data and reflect on the advantages and disadvantages of using logging data in long-term trials.

II. HOBBIT robot

HOBBIT was developed for a project within the European FP7 framework. Its main objective is to enable older people to feel safe in their homes. The risk of falling can be considered a "trigger mechanism" for institutionalization into an elderly care facility [3]. An overview of all robot functionalities is given by Fischinger et.al. [4]; next we only present the most common functionalities during the field trials with the second generation prototype of HOBBIT as shown in figure 1.

Fig. 1. Hobbit - Prototype 2

Call Robot To command the robot to come to the user from any place in the apartment, so called call buttons were installed at different places. If a call button is pressed HOBBIT moves to this position, detects and moves toward the user (0.3 m safety distance). Users can bring the robot closer with a gesture.

Pick-up Objects can be picked up by the robot after the user starts the pick-up activity and pointing at the object on the floor. The robot will move into position to detect a suitable object ($0.03\,\mathrm{m} \leq x \leq 0.1\,\mathrm{m}$, $\leq 0.5\,\mathrm{kg}$), and re-position and grasp it if a possible grasp location has been found. After the pick-up the floor is checked again to confirm that the object has been picked up.

Learn and Bring Object: To teach the robot a new object, the user can put it onto a turntable that is rotated while observed with the robot's upper camera. A 3-D model is generated out of the collected data, which the user is then able to name and use in the bring object task. HOBBIT is moving to search locations, performs recognition and moves back to the user and guides her or him to the location where the object was found.

Follow Me: Uses leg tracking to calculate navigation goals behind the user. If the user is in the field of view and moving away from the robot, HOBBIT is moving to those locations. Otherwise, the robot stops the action and informs the user about this.

Go To: Provides the user the option to send HOBBIT to one out of up to six learned positions inside the home as well as to the docking station for autonomous recharge. HOBBIT offers gestures, speech, and a touchscreen as input modalities. Recognizable gestures and speech commands are yes, no, help, reward (a functionality to praise the robot after a task), and come closer. Speech can additionally be used for go to, learn/bring object and pick-up. Touch input includes all mentioned commands as well as the option to name an object during the learning procedure.

III. Logging

We wanted to assess the following criteria with logging data. (1) Flexibility: Logging the number of times different modalities are successfully used. (2) Utility: Logging the number of robot commands issued by the user. The logging was implemented as a ROS node that collected data from a set of different sources and stored them in a database for analysis and log files were kept for debugging purposes. Each robot stored the database locally on its internal main computer for later retrieval to compensate for unreliable Internet connections. However, an uncaught bug in the database node led to the issue that under certain circumstances the logging stopped completely and without notification to the responsible developers. To compensate for some of the lost data, (119 out of 371 days) of the user-trials at the point of noticing the issue, we started extracting additional and confirmatory data from the debug log files by incorporating them into the database for further analysis. A complete recovery was not possible, but we were able to retrieve data for 36 otherwise lost days. Unfortunately we were not able to verify how complete the recovered data is, as we could only assess the amount of data we were able to recover, but not how much was lost.

IV. Results

We assumed that one input modality will be preferred more than others and that specific commands will suit better for one input modality than the other. Tables I, II and III show interesting tendencies with respect to preferred interaction modalities. On the first glance it seems that gestures were the most preferred input modality, even more preferred than the touchscreen which we assumed to have the highest usage rate. This counter-intuitive result however, was caused by many false/positive gesture recognitions that happened when the user was moving the arm towards the touchscreen. This was above all true for the over sensitive *come closer*-gesture. This also explains the high numbers for the *come closer*-gesture during most trials. However, content-wise it is convincing that *come closer* and *yes* are the most frequently used gestures as both were used when the robot asked for confirmation to move closer to the user (the most often used action). Similarly, we can argue that *yes* is the most often used speech command, while for touch input the *go to* command and the *recharge* command were most commonly started. Taking into account the many false positives in gesture recognition for the first few participants of the trials, it is rather the case that the touchscreen was used most, followed by speech and gesture.

TABLE III
USAGE FREQUENCIES OF MODALITIES - GREEK USERS

Modality	GRC1	GRC2	GRC3	GRC4
Gestures	247	237	20	12
Most often	Come	Come	Cancel	No
used	229	191	10	5
Speech	34	47	19	6
Most often	Yes	Yes	Stop	Yes
used	13	13	13	5
Touchscreen	162	155	58	10
Most often	Go to	Recharge	Recharge	Recharge
used	27	47	36	8

TABLE VI
USAGE FREQUENCIES OF ROBOT ACTIONS - GREEK USERS STARTED/CANCELED

Action	GRC1	GRC2	GRC3	GRC4
Call HOBBIT	143/49	91/40	50/12	21/14
Pick-up	8/6	10/10	10/9	0/0
Teach object	5/3	1/0	1/1	0/0
Bring object	0/0	0/0	0/0	0/0
Follow me	0/0	4/2	0/0	0/0
Go to	25/13	23/7	9/2	3/0
Recharge[2]	41	78	36	8

Tables IV, V and VI show that for the almost every user the most often executed command was *Call HOBBIT*. As the charging station of the robot was placed outside of the living room (that the users saw as the room in which they would interact the most with the robot), the *Call HOBBIT* command can be seen as the initial phase of most interactions (i.e. executed before any other user-initiated interaction). However, a closer look also shows us that almost all commands were half of the time canceled. How can this high cancellation rate be explained? One reason is that starting one command before

TABLE I
USAGE FREQUENCIES OF MODALITIES - AUSTRIAN USERS

Modality	AUT1	AUT2	AUT3	AUT4	AUT5	AUT6	AUT7
Gestures	744	293	182	434	31	111	199
Most often	Come	Come	Come	Come	Yes	Yes	Yes
used	434	232	127	323	17	44	88
Speech	476	353	523	78	478	279	344
Most often	Yes	Yes	Yes	Yes	Yes	Yes	Yes
used	307	205	432	34	373	197	166
Touchscreen	473	428	341	216	526	885	443
Most often	Go to	Recharge	Recharge	Go to	Go to/Reward	Go to	Stop
used	104	97	86	51	126	213	119

TABLE II
USAGE FREQUENCIES OF MODALITIES - SWEDISH USERS

Modality	SWE1	SWE2	SWE3	SWE4	SWE5	SWE6	SWE7
Gestures	456	227	401	4	337	23	73
Most often	Come	Come	Come	Fall	Come	Come	Come
used	354	131	336	3	287	17	40
Speech	228	87	167	4	10	40	56
Most often	Yes	Yes	Stop	Yes	Yes/Stop	Stop	Stop
used	119	49	53	3	3	19	20
Touchscreen	689	286	487	7	122	91	492
Most often	Go to	Go to	Recharge	Social Role[3]	Go to	Recharge	Go to
used	194	68	114	4	44	24	233

TABLE IV
USAGE FREQUENCIES OF ROBOT ACTIONS - AUSTRIAN USERS STARTED/CANCELED

Action	AUT1	AUT2	AUT3	AUT4	AUT5	AUT6	AUT7
Call HOBBIT	340/190	172/81	74/21	97/75	71/42	146/60	349/263
Pick-up	26/24	78/67	18/18	4/2	18/15	272/250	39/35
Teach object	14/1	21/10	8/0	2/1	1/0	0/0	3/2
Bring object	0/0	0/0	0/0	0/0	15/3	36/1	5/4
Follow me	7/4	25/15	3/3	5/5	26/20	18/6	14/7
Go to	89/33	29/8	18/5	46/26	117/24	386/85	41/17
Recharge[2]	49	186	126	16	88	312	168

TABLE V
USAGE FREQUENCIES OF ROBOT ACTIONS - SWEDISH USERS STARTED/CANCELED

Action	SWE1	SWE2	SWE3	SWE4	SWE5	SWE6	SWE7
Call HOBBIT	242/123	97/57	340/131	6/4	40/26	39/12	140/61
Pick-up	30/30	19/19	42/40	0/0	9/9	0/0	9/9
Teach object	14/4	4/2	1/0	0/0	1/0	0/0	0/0
Bring object	1/1	2/1	0/0	0/0	0/0	0/0	0/0
Follow me	4/1	7/7	4/2	0/0	0/0	0/0	0/0
Go to	193/109	66/31	84/39	0/0	43/25	18/11	224/70
Recharge[2]	135	46	207	1	0	34	320

another running one is finished is counted as cancellation. Secondly, in cases where the robot autonomously started an action this canceled a user action, something that caused dissatisfaction on the user side, (e.g. when the robot started moving to the docking station for autonomous recharge) while the user was still playing a game. Regarding the numbers for *pick-up*, it has to be mentioned that these only tell us that the command was issued by the user and fully executed by the user; it does not tell us if an object was successfully grasped.

Regarding the numbers for *teach object*, it has to be mentioned that only the user procedure of teaching an object was executed, but the object was not actually learned or stored in the database, which is the reason why *Bring Object* was not performed at all in many trials. The ranking of frequencies therefore lies within the expected range. Overall, in most cases *Call HOBBIT* was issued most often, followed by *Go to*. All other robot commands were not used that often.

V. REFLECTION

In HRI evaluation research we are aware that self-reporting data alone is not sufficient, as well as log-data alone is not sufficient to understand the reasons behind user preferences with a robotic system [2]. However, it is not possible to rely on the log-data to understand the actions and intentions of the users, but only the discrete events that the robot was able to observe. The feedback from the users during the field trial showed that the logged data was not always consistent with the actual events. Situations in which the logging proved to be most helpful, were those when the possibility to directly assess the outcome of changes and improvements, that were introduced during the running trials, was needed. In such situations the immediate effect on the robotic behavior could be observed without the need to request feedback from the user. Those improvements were introduced mainly to reduce false positives in gesture or speech recognition (reported by users and verified in our lab) to improve the reactive behavior of the robot towards the user. Thus, we conclude that log-data can be used to improve our understanding of long-term Human-Robot interactions, but it is not possible to interpret the logged data without a detailed knowledge of the underlying robotic system and that log-data can only be seen as an addition to self-reporting data and questionnaires to assess long-term trials.

VI. ACKNOWLEDGMENTS

This work has been partially funded by the European Commission under grant agreements FP7-IST-288146 HOBBIT.

REFERENCES

[1] J. Fink, V. Bauwens, F. Kaplan, and P. Dillenbourg, "Living with a vacuum cleaning robot," *International Journal of Social Robotics*, vol. 5, no. 3, pp. 389–408, 2013.
[2] M. Lohse, M. Hanheide, K. J. Rohlfing, and G. Sagerer, "Systemic interaction analysis (sina) in hri," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, 2009, pp. 93–100.
[3] M. Dramé, F. Fierobe, P.-O. Lang, D. Jolly, F. Boyer, R. Mahmoudi, D. Somme, I. Lanièce, D. Heitz, J.-B. Gauvain *et al.*, "Predictors of institution admission in the year following acute hospitalisation of elderly people," *The journal of nutrition, health & aging*, vol. 15, no. 5, pp. 399–403, 2011.
[4] D. Fischinger, P. Einramhof, K. Papoutsakis, W. Wohlkinger, P. Mayer, P. Panek, S. Hofmann, T. Koertner, A. Weiss, A. Argyros *et al.*, "Hobbit, a care robot supporting independent living at home: First prototype and lessons learned," *Robotics and Autonomous Systems*, 2014.

[1] http://www.aal-domeo.org
[2] The number of cancellations for Recharge was not yet evaluated.
[3] The *Social role* command triggered the change of certain parts in the robot's behavior.