Perceived differences between on-line and real robotic failures

Alexander Mois Aroyo
University of Waterloo

Dario Pasquali
Istituto Italiano di Tecnologia

Francesco Rea
Istituto Italiano di Tecnologia

Giulio Sandini
Istituto Italiano di Tecnologia

Perceived differences between on-line and real robotic failures*

Aroyo A.M.[1], Pasquali D.[2], Kothig A.[1], Rea F.[2], Sandini G.[2], Sciutti A.[3]

Abstract — **Robotic failures are an inevitable occurrence. This study tries to shed light on how people perceived failures and how much they affect the interaction. Continuing the work on a previous reliable study, this research gathers information on how people perceive a failure in an online validation study. After the failures have been selected, they were applied in a real-world game-like scenario where participants played a Treasure Hunt game with iCub. Initial results show that failure perception is more severe in the online study rather than the actual game.**

***Index Terms* — Human Robot Interaction; Trust; Faultiness.**

I. Introduction

Trust is important in human nature, it helps to delegate tasks to others, build relationships and cooperate in society. In a similar way, many robotic researchers believe that trust is crucial in human robot interaction [1]–[5]. Technology, however, is not fully reliable and can trigger unwanted responses in people when overtrusted [2], [3]. Research has also shown that the perceived trust people have towards robots decreases if the robot fails [4], [5].

This research is an extension of a previous study named the Treasure Hunt (TH) where participants played an interactive game with iCub in which they had to find 5 hidden eggs in a room using hints provided by the robot in order to win a monetary prize. If participants managed to find the eggs, iCub suggested them to gamble their money; if they manage to find another egg, they will double the prize, or lose everything otherwise. For more information consult the TH paper [1].

This new version, aptly named Unreliable Treasure Hunt, introduces a series of failures with the purpose to study how they will affect people's perception of the robot.

II. Methodology

A. Design phase

Robotic unreliability and failures are quite common, however there is still a low number of studies that analyze the perception and effect of those failures on humans. Based on a literature review from Honig and Oron-Gilad [6] and on the context of this game scenario, mechanical failures were chosen over other types of failures as the latter ones could be associated to the hints' difficulty instead of the failures per se.

The following four technical unreliable behaviors* were designed to affect the perceived reliability of iCub. Two were audio related, and the others, control based. All of them occurred in almost identical timing across participants.

Distorted good luck (audio): Just before the game starts, iCub wishes them good luck with a distorted voice.

Abrupt pointing (control): iCub repeatedly fails to point the location of one of the eggs.

Noised hint (audio): One of the spoken hints is replaced with a meaningless distorted speech.

Fake crash (control): iCub simulates a full crash and shut down bending over, afterwards it recovers. See Figure 1.

The failures were designed so to produce a sense of malfunction form the participants' point of view, however, they did not heavily affect the game timewise: eventually the pointing was performed, while a transcript of the speech was available on the TV. The reason behind this design is to allow the comparison between the initial Treasure Hunt game in [1].

TABLE I. Online validation of failures

Failures	Credibility	Artificiality	Severity
Distorted good luck			
Turntable (TT)	M=4.64, SD=1.65	M=4.32, SD=4.53	22%
Noised TT	**M=4.92, SD=1.57**	**M=3.98, SD=1.75**	**48%**
Repeated word	M=4.44, SD=1.87	M=4.2, SD=1.61	30%
Abrupt pointing			
Random	M=3.76, SD=1.74	M=4.62, SD=1.59	10%
Abrupt	M=5.48, SD=1.37	M=3.2, SD=1.47	16%
Abrupt w. jerk	**M=5.66, SD=1.44**	**M=3.14, SD=1.72**	**74%**
Noised hint			
Interference	M=5.72, SD=1.08	M=3.20, SD=1.51	12%
Shortcut	M=5.86, SD=1.28	M=2.88, SD=1.66	26%
Shortcut noise	**M=5.78, SD=1.3**	**M=3.38, SD=1.89**	**62%**
Fake crash			
Fast Slow, NA[a]	M=3.6, SD=1.71	M=4.98, SD=1.87	20%
Slow Slow, NA[a]	M=3.6, SD=1.85	M=4.78, SD=1.84	20%
Fast Slow, A[b]	**M=4.36, SD=1.82**	**M=4.26, SD=1.86**	**34%**
Slow Slow, A[b]	M=4.34, SD=1.89	M=4.56, SD=1.83	26%

a. No audio; b. audio.

B. Online validation study

The failures were modified with different nuances so to validate the most effective one; in the audio ones the distortion was varied, while in the control ones the movements and speed were changed. In order to understand whether the design is effective, all different behaviors were shown to 50 participants (56% females, average age of 32 ± 8.3 years) in an online survey. On a 7-point Likert scale, participants were asked to evaluate the credibility, artificiality and severity of the failures. The results are shown in TABLE I. where items in bold correspond to high degree of severity combined with high degree of credibility and relatively lower degree of artificiality.

C. Game scenario study

The failures selected by the online validation study were introduced in the original TH game explained in the first

[1]A.M. Aroyo is corresponding author and A. Kothing are affiliated with the University of Waterloo {alexander.aroyo, austin.kothig@uwaterloo.ca}.

[2]D. Pasquali is affiliated with DIBRIS, Università di Genova, ICT, Italian Institute of Technology (IIT). Together with F. Rea and G. Sandini are affiliated with RBCS, IIT.

[3]A. Sciutti is affiliated with CONTACT, IIT.

The emails are the following: {dario.pasquali, francesco.rea, giulio.sandini, alessandra.sciutti@iit.it}

*youtu.be/playlist?list=PLm3z-RU1nzLOQnU92j6GpboajxEGUD2dp section and in [1]. 68 healthy Italian participants took part in this new variation of the experiment with an average age of 37.96±13.81 years; 46% males; no previous robot experience.

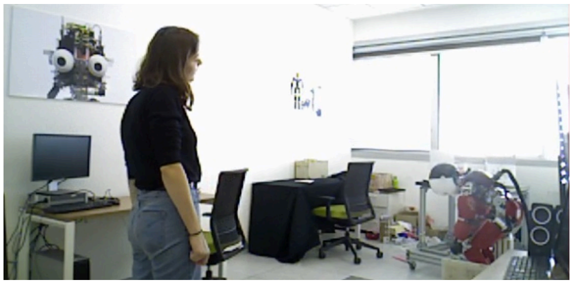

Figure 1. Participant reacting to iCub's failure.

After the experiment, participants were asked to fill in a post questionnaire, then they were debriefed about the failures of the robot, and were asked a small set of additional questions. Some of these questions were the same asked in the online validation study as to understand whether the effects of an online video observation were the same as in a real-game scenario. The questions analyzed in this extended abstract are explained in the next section.

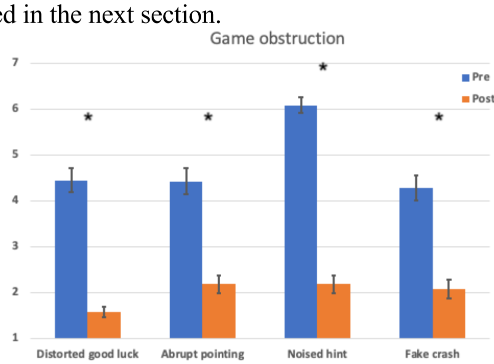

Figure 2. Pre/Post values of game obstruction: 0 = not obstructive at all; 7 = very obstructive. Statistical significance with p<0.01 marked by *.

III. Results

The following questions are shown in the *pre* and *post* studies. Pre refers to the 50 participants of the online validation study; 49% were participants of the original TH game, and all were fully informed about the game dynamics. In the post, 68 different participants played the game.

The first question was to evaluate in a 7-point Likert scale how much the failures would obstruct their game. In Figure 2. it can be seen a drastic and statistically significant decrease (two sample *t*-test) in the perception of the failure (pre) and the actual game (post) for each one of the failures. Another set of questions asked to both groups was whether the trust they had in the robot would increase, decrease or stay the same after iCub's failure. In Figure 3. it can be seen that in pre, people's perceived trust towards the robot would decrease after the failure; while in the actual game, people's perception of their trust towards the robot remained the same. In the pre phase, participants were hypothetically asked what they would do if iCub offered them to gamble after all the failures; on a 7-point Likert scale (1-not probable at all; 7-very probable) the mean was 4.12 (SD=2.3). In the real game scenario (post), 25% of the participants reached the gambling phase, and 100% of them decided to gamble. As a verification check, all participants (pre and post) ordered the severity of the failures in the same way; and both replied that after seeing all four failures, they believe it's possible that there would be a new failure: on 7-point Likert scale (1-not probable at all; 7-very probable) pre: 5.38(SD=1.32); post: 4.08(SD=1.61).

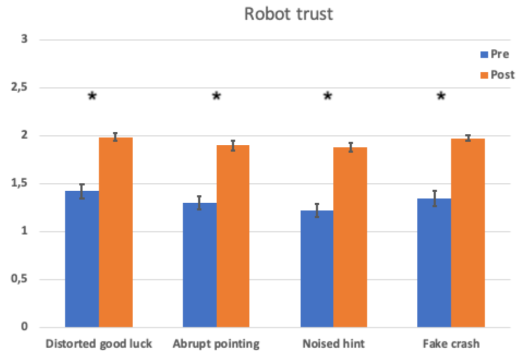

Figure 3. Pre/Post values robot trust: 0 = I do not trust robots; 1 = I trust it less than before; 2 = I trust it the same way as before; 3 = I trust it more than before. Statistical significance with p<0.01 marked by *.

IV. Discussion

There are big differences between physical embodiment of the robot and its virtual representation, as seen in literature [7]. The significant decrease seen in Figure 2. could be explained because in the videos participants are consciously imagining how the failure will affect their game; while people playing the game were so engaged that they may not even realize the robot is failing, or they just perceived it as normal behavior as they have not experienced previously a situation in which iCub did not fail. Participants who experienced the real failure, later on and retrospectively thinking may have realized that it was not such of a big issue. This idea could also explain why the trust perception seen in Figure 3. did not decrease with time as suggested by [4], [5]. Surprisingly but similarly to the original TH, all participants who reached the gambling phase decided to accept the challenge. A possible explanation of the last two effects could be that, nevertheless iCub was faulty, it kept working and providing hints. A more in-depth analysis is needed to understand the various reasons participants behaved that way. This study also prompts the need to research in more details how (non)failure would affect the same participants over a long period of time.

References

[1] A. M. Aroyo, F. Rea, G. Sandini, and A. Sciutti, "Trust and Social Engineering in Human Robot Interaction: Will a Robot Make You Disclose Sensitive Information, Conform to its Recommendations or Gamble?," *IEEE RAL.*, vol. 3, no. 4, pp. 3701–3708, 2018.

[2] A. M. Aroyo *et al.*, "Will People Morally Crack Under the Authority of a Famous Wicked Robot?," in *27th ROMAN*, 2018.

[3] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner, "Overtrust of Robots in Emergency Evacuation Scenarios," *Elev. ACM/IEEE Int. Conf. Hum. Robot Interact.*, pp. 101–108, 2016.

[4] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust," *Proc. Tenth Annu. ACM/IEEE Int. Conf. HRI.*, pp. 141–148, 2015.

[5] A. Rossi, K. Dautenhahn, K. L. Koay, and M. L. Walters, "The impact of peoples' personal dispositions and personalities on their trust of robots in an emergency scenario," *Paladyn*, 2018.

[6] S. Honig and T. Oron-Gilad, "Understanding and resolving failures in human-robot interaction: Literature review and model development," *Frontiers in Psychology*. 2018.

[7] W. A. Bainbridge, J. W. Hart, E. S. Kim, and B. Scassellati, "The benefits of interactions with physically present robots over video-displayed agents," *Int. J. Soc. Robot.*, vol. 3, no. 1, pp. 41–52, 2011.