This is the author's accepted manuscript. The final published version of this work (the version of record) is published in Companion of the 2020 ACM/IEEE International Conference on Human Robot Interaction (HRI '20 Companion), March 23-26, 2020, Cambridge, United Kingdom, ACM, New York, NY, USA, available at DOI: 10.1145/3371382.3378278. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Challenges of a Real-World HRI Study with Non-Native English Speakers: Can Personalisation Save the Day?

Bahar Irfan*
CRNS, University of Plymouth, UK
bahar.irfan@plymouth.ac.uk

Mehdi Hellou*†
Polytech Sorbonne, France
mehdi.hellou@etu.sorbonne-universite.fr

Alexandre Mazel
SoftBank Robotics Europe, France
amazel@softbankrobotics.com

Tony Belpaeme
ID Lab, Ghent University, Belgium
CRNS, University of Plymouth, UK
tony.belpaeme@plymouth.ac.uk

ABSTRACT

Real-world studies allow for testing the limits of HRI systems and observing how people react to failures. We developed a fully autonomous personalised barista robot and deployed the robot on an international student campus for five days. We experienced several challenges, the most important one being speech recognition failures due to foreign accents. Nonetheless, these failures showed a different perspective on HRI, and we demonstrate how personalisation can overcome a negative user experience.

KEYWORDS

Personalisation; Real-World Study; Natural Language Interaction

ACM Reference Format:
Bahar Irfan, Mehdi Hellou, Alexandre Mazel, and Tony Belpaeme. 2020. Challenges of a Real-World HRI Study with Non-Native English Speakers: Can Personalisation Save the Day?. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20 Companion), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3371382.3378278

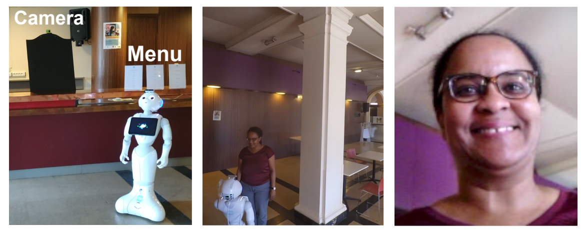

Figure 1: (a) Experiment setup with Adapted Pepper, (b) external camera image, (c) internal camera image.

1 INTRODUCTION

Deploying autonomous robots in the real-world adds several challenges, such as incomplete data and dropouts, decreasing the success rate of the interaction [3]. Nonetheless, such studies are necessary to create reliable systems for long-term interactions. Moreover, user experience can decrease over time if the robot uses a fixed set of behaviours, which can be overcome by personalisation [5].

There are only a few studies [2, 7, 9] that explored fully autonomous personalisation in dialogue for long-term HRI. However, none of these studies were conducted in the real-world.

We designed a study to analyse the effects of personalisation in a real-world application using a barista robot which recalls the user's previous orders in subsequent interactions, similar to a barista in a local coffee shop. Hence, we developed a fully autonomous robot with user recognition, automated speech recognition (ASR) and a rule-based dialogue management system (RBDMS). During the real-world study, we faced several failures that greatly affected the user experience. However, personalisation mitigated interaction failures and the *negative* user experience. In this paper we describe our system and study, and focus on the challenges suggesting solutions to overcome them in the future.

*Both authors contributed equally to this research.
†This work was conducted during the author's internship at SoftBank Robotics Europe.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom
© 2020 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-7057-8/20/03.
https://doi.org/10.1145/3371382.3378278

2 METHODOLOGY

2.1 Rule-Based Dialogue Management System

The RBDMS is modelled on a real-world barista who: (1) requests the drink order, (2) size, (3) snacks, (4) confirms the order, (5) changes the order if necessary, (6) takes the customer's name, (7) notes the order pick up location, (8) says goodbye. Typically, a customer can ask for the order in one sentence, removing the need of (2) and (3), however, we separated the steps to reduce the errors and aid speech recognition. Template-matching and dialogue state tracking are used to match the user responses to the phrases in the RBDMS.

2.2 Personalisation

Multi-modal incremental Bayesian Network [4] (MMIBN) is used for online user recognition. MMIBN combines face recognition with soft biometrics (age, gender, height and time of interaction)[1] for reliable identification.

The interaction is personalised by recalling the most frequent or recent order in the database, which is suggested to the user, e.g.,

[1]Obtained from NAOqi: http://doc.aldebaran.com/2-5/

"Hello, Jane! Would you like to have a large coffee and a chocolate cookie again?". The user can accept or change the order. If the user is incorrectly identified, the name is requested, and the corresponding most frequent order is suggested. If the user is new, the interaction is pursued as described in Sec. 2.1. At the end of the interaction, the robot updates the database with the order and adds the new user.

2.3 Speech Recognition

We used *Adapted Pepper* robot[2] (shown in Figure 1), which has an improved microphone system with lower noise compared to an off-the-shelf robot. We used NAOqi voice activity detection and Google Cloud Speech-to-Text for online speech recognition.

Speech recognition was optimised with a band-pass filter based on 5 monologues from personalised barista phrases with 12 non-native English speakers, providing an exact match accuracy of 47% and an error (1-BLEU score) of 0.34.

2.4 Real-World Experiment

We conducted a 5 day study in the coffee bar of an international student campus, Cité Internationale Universitaire de Paris, with 18 non-native English speakers (11 M, 7 F) within the age range of 22-47 (M=28.2, SD=7.0). The study had three conditions: enrolment (EC), non-personalisation (NPC), personalisation (PC). EC and NPC have the same structure described in Sec. 2.1; EC is the first interaction with users, whereas NPC depicts the second and third interactions. PC is the corresponding (as in Sec. 2.2) condition to NPC.

The perceived performance of the robot and the experience of the users are evaluated through a questionnaire. The speech recognition performance is analysed, and the reactions of the participants are evaluated through the robot camera and an external camera[3].

3 RESULTS AND DISCUSSION

Due to the technical difficulties further outlined in this section, only 5 out of 18 EC, 4 out of 6 NPC, 3 out of 9 PC interactions were successful (i.e., completed and correct order was delivered). We did not interfere with the experiment unless the robot was stuck at a phrase for a prolonged time or had an apparent connection failure, in which case, we asked if the participant would like to repeat the interaction. Due to the low number of subsequent encounters, the resulting Bayes factors are between 0.3-3, suggesting inconclusive statistical significance between conditions [6], thus, we interpret the implications of the trends in the results.

The results in Figure 2 support that a higher percentage of users received the correct order and had more complete interactions in NPC. However, in PC, a higher percentage of users enjoyed the interaction, looked forward to the next one, and preferred to interact with the robot as a barista in a coffee shop. These findings suggest that personalisation can improve the negative experience of users, which is a key result of conducting a real-world study.

The primary cause of failure was speech recognition, necessitating users to repeat phrases. The underlying reasons are foreign accents of non-native speakers, latency due to connection problems, quietly speaking users, user's distance from the robot, and

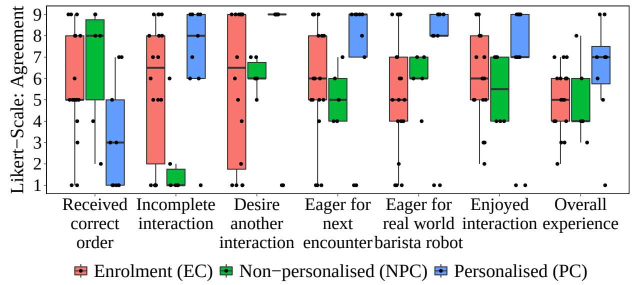

Figure 2: User questionnaire results.

the accuracy of the ASR on the audio obtained from the robot's microphones.

The second major problem was the fixed order structure of RBDMS. The system failed to understand when the user ordered the items in a combined sentence, switched the order of items, or confirmed the order first and then tried to change the order later.

The users did not realise when the robot incorrectly identified them, thus, online learning in MMIBN and RBDMS updated the wrong user, thereby, causing PC to have a worse success rate.

4 SUGGESTIONS FOR REAL-WORLD STUDIES

These technical difficulties caused the participants to repeat their phrases several times, change their wording, and even accept wrong orders, but these are unlikely to happen in the real-world when the customers are in a hurry. For deploying robots to the real-world, we need solutions that are reliable and can recover from failures.

Our results showed that ASR is not accurate enough for real-world applications, hence, a touchscreen interface for text or image-based interaction can be used. However, such methods decrease the naturalness of the interaction. Thus, it is preferable to improve the accuracy of ASR, by constraining grammar [8], ensuring a reliable WiFi or using an onboard ASR, and using high-quality microphones. Low ASR accuracy in foreign accents can be overcome by personalising the interaction with the user's native language.

We should also account for user errors by designing systems that are flexible and robust. For example, confirming the identity before the order can overcome errors in online learning for MMIBN. Moreover, a neural network with a long-term memory would be more suitable than a rule-based dialogue management system for reverting changes in the state of the dialogue [1]. We are developing such a system for personalised interactions.

Nevertheless, failures enabled us to observe the high positive impact of personalisation on the negative user experience, which showed the importance of evaluating technologies outside of controlled environments and studying how people respond to failures.

ACKNOWLEDGMENTS

This work has been supported by the EU H2020 Marie Skłodowska-Curie Actions ITN project APRIL (grant 674868) and SoftBank Robotics Europe. The authors would like to thank Cité Internationale Universitaire de Paris for hosting the study, and the participants for their time and patience.

[2]Created for MuMMER project: http://mummer-project.eu.
[3]Participants signed consent forms under the University of Plymouth ethical approval for audio and video recording and image sharing.

REFERENCES

[1] Antoine Bordes, Y-Lan Boureau, and Jason Weston. 2016. Learning End-to-End Goal-Oriented Dialog. arXiv:cs.CL/1605.07683

[2] Nikhil Churamani, Paul Anton, Marc Brügger, Erik Flieundefinedwasser, Thomas Hummel, Julius Mayer, Waleed Mustafa, Hwei Geok Ng, Thi Linh Chi Nguyen, Quan Nguyen, and et al. 2017. The Impact of Personalisation on Human-Robot Interaction in Learning Scenarios. In *Proceedings of the 5th International Conference on Human Agent Interaction*. Association for Computing Machinery, New York, NY, USA, 171âĂŞ–180. https://doi.org/10.1145/3125739.3125756

[3] Christian Dondrup, Lynne Baillie, Frank Broz, and Katrin Lohan. 2018. How can we transition from lab to the real world with our HCI and HRI setups? In *2018 ACM/IEEE International Conference on Human-Robot Interaction, 4th Workshop on Public Space Human-Robot Interaction: MobileHRI: Mobile human-robot interaction*. Chicago, IL, USA.

[4] Bahar Irfan, Natalia Lyubova, Michael Garcia Ortiz, and Tony Belpaeme. 2018. Multi-modal Open-Set Person Identification in HRI. In *2018 ACM/IEEE International Conference on Human-Robot Interaction Social Robots in the Wild workshop*. Chicago, IL, USA.

[5] Bahar Irfan, Aditi Ramachandran, Samuel Spaulding, Dylan F. Glas, Iolanda Leite, and Kheng Lee Koay. 2019. Personalization in Long-Term Human-Robot Interaction. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Daegu, South Korea, 685–686. https://doi.org/10.1109/HRI.2019.8673076

[6] Harold Jeffreys. 1939. *Theory of probability* (1 ed.). The Clarendon Press, Oxford.

[7] Zerrin Kasap and Nadia Magnenat-Thalmann. 2012. Building Long-Term Relationships with Virtual and Robotic Characters: The Role of Remembering. *Vis. Comput.* 28, 1 (2012), 87–97. https://doi.org/10.1007/s00371-011-0630-7

[8] James Kennedy, Séverin Lemaignan, Caroline Montassier, Pauline Lavalade, Bahar Irfan, Fotios Papadopoulos, Emmanuel Senft, and Tony Belpaeme. 2017. Child Speech Recognition in Human-Robot Interaction: Evaluations and Recommendations. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Vienna, Austria, 82–90. https://doi.org/10.1145/2909824.3020229

[9] Xiqian Zheng, Dylan F. Glas, Takashi Minato, and Hiroshi Ishiguro. 2019. Four memory categories to support socially- appropriate conversations in long-term HRI. In *2019 ACM/IEEE International Conference on Human-Robot Interaction Personalization in Long-Term Human-Robot Interaction (PLOT-HRI19) workshop*. Daegu, South Korea.