A Wizard of Oz Interface with Qtrobot for Facilitating the Handwriting Learning in Children with Dysgraphia and its Usability Evaluation*

Jianling Zou[1][0000-0002-3074-8090], Soizic Gauthier[2,3][0000-0002-1699-8544], Salvatore M. Anzalone[1][0000-0001-5163-970X], David Cohen[2,3], and Dominique Archambault[1]

[1] CHArt laboratory, Paris 8 University, Paris, France
jianlingzou93@gmail.com
[2] Department of Child and Adolescent Psychiatry, Pitié Salpêtrière University Hospital, Paris, France
soizic.gauthier@aphp.fr
[3] CNRS UMR 7222, Institut des Systèmes Intelligents et Robotiques, Sorbonne University, Paris, France

Abstract. This paper presents the design of a Wizard of Oz interface with a library of robot behaviors developed for caregivers on handwriting learning in children with dysgraphia. We assess the final version of the interface with 15 care participants during a single simulated session. All care participants fulfilled the System Usability Scale and Attrakdiff 2 scale. In sum, the results confirm the system's usability, but there is still room for improvement.

Keywords: Wizard of Oz interface · Social robot · Dysgraphia · Humain-robot interaction.

1 Introduction

Handwriting is an essential skill to describe thoughts or objects, and express emotions. This ability seems natural for most people. However, some children demonstrate that their writing abilities are inconsistent with their ages. We call this disturbance in handwriting dysgraphia[5]. Children with dysgraphia are often viewed by teachers or parents as neglectful or lazy rather than being perceived as children suffering from a learning disability they cannot control. This situation directly impacts the children's self-esteem, self-confidence, and social relationships[4][6]. When handwriting difficulties are detected, children are usually addressed by occupational therapists with the approach focus on attention deficit that impacts handwriting automatization or on handwriting itself[6]. The pen-and-paper exercises which are very close to the tasks carried out in school

* Supported by National Research Agency (ANR) of France.

will be proposed for remediation to automatize the writing process[3]. Some children with dysgraphia express frustration, even refusal , regarding treatment sessions. The review of Cifuentes et al. highlight social robots are efficiently used to assist with therapies for children with NDD. And in the case of autism spectrum disorders, social robots can support clinical effectiveness of therapeutic session[13]. With the aim to facilitate the learning process of handwriting in children with dysgraphia and to break this vicious circle, we present the design of a tool based on a social robot. It's a Wizard of Oz interface which supports the caregivers in selecting the robot's behavior to apply the learning-by-teaching method to ask the child to teach the robot learning handwriting.

2 Methodology

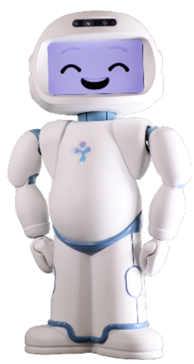


Fig. 1. QTRobot's apperance. [Public domain] via symbioïde.

Fig. 2. Learning-by-teaching setup

Luxai's QTRobot[4] (**Fig 1**) was chosen as the preferred social robot platform due to its expressive abilities in terms of social cues and, above all, broader facial expressions and good stability, with an appropriate size for interaction with children, of 64 cm . This social robot has been designed to increase therapy efficiency by encouraging active and engaged interaction[9]. Meanwhile, a serious game platform dedicated to practicing handwriting called Dynamilis[5] was used as a complementary tool to the robot. Among the various games provided by Dynamilis, the game that was used to train multi-features is co-writer. It was the best-suited game for our learning-by-teaching scenario. In this game, the child becomes the robot's teacher and teaches it how to write. In our previous work[7], Gargot et al. used the co-writer game between a Nao robot and a 10-year-old

[4] Luxai's offical website : https://luxai.com/

[5] The serious game "Dynamilis": https://dynamilis.com/en/

boy with a complex neurodevelopmental disorder combining developmental coordination disorder and severe dysgraphia. The structure of the study consisted of 20 consecutive weekly sessions (i.e., 500 minutes).

As shown in **Fig 2**, the robot is near the table and faces the child. The child has an iPad to play the serious game, and the caregiver has a tablet to control the robot. Meanwhile, the robot has a fake tablet to show the child that he is learning writing. This article focuses on evaluating the usability of the Wizard-of-Oz interface that caregivers will use during rehabilitation sessions. Another successive article will present the study of rehabilitation sessions and its result.

Following a literature review and an observatory study, we created the first set of behaviors based on the robot's role in our learning-by-teaching scenario of all of the significant robot-child interactions in the work indicated above[4]. Robot's complex behaviors involve gestures, speeches, and emotions. Gestures are predefined fluid movements created by physically moving the robot's joint while recording the dynamic of their different positions and angles, exploiting the gesture API built-in with the QTRobot. In addition, we have developed a standardized file format for these behaviors, keeping the name and the speed of the gesture, a sequence of facial expressions, and similar speeches. The program randomly selected one of these similar speeches when the behaviors were played to reduce repeatability and increase the robot's vitality.

There was a certain complexity in developing multi-platform interfaces because of different operating systems (such as Android, iOS, Windows) for various electronic devices that might be used in the future. In the end, we decided to build a web interface to adapt to multiple platforms. HTML 5, JavaScript, and CSS have been used for the web's front-end. Then, we used the Flask framework in Python as a server to ensure communication between the web and the robot. After every button click, the robot played the corresponding behavior through Robot Operating System (ROS). Furthermore, the user interface design followed Jakob Nielsen's ten usability heuristics[12]. After an iterative design with six children and six caregivers who participated in nine rehabilitation sessions (this study will be presented in another article), we invited fifteen caregivers (five teachers, five psychomotor therapists, five speech therapists) to evaluate the usability of the final version of our system. **Table 1** describes each participant.

The evaluations were conducted in the form of single simulated sessions, each session was about an hour. After a tutorial about how to use the interface, every caregiver can test the interface in thier own way during half an hour. Then they were given some tasks and had to control the robot as if they were with a child. In the end, we asked the caregiver to complete two well-validated scales of user's experience evaluation: the French version of SUS (System Usability Scale)[8] and the Attrakdiff 2 scales[11]. In more detail, SUS is a Likert scale comprising ten items with a total score that varies between 0 and 100. AttrakDiff questionnaire is also a Likert scale containing twenty-eight items divided into four subscales, each comprising seven questions. The AttrakDiff subscales are as follows[10]:

Table 1. Description of caregivers participants who carried on a final evaluation of system

Id.	Age*	Sex	Occupation	Years of experience*
ProFinal1	24	F	Psychomotor therapist	1
ProFinal2	27	F	Speech therapist	4
ProFinal3	22	M	Psychomotor therapist	1
ProFinal4	45	F	Psychomotor therapist	0 (2 years of internship)
ProFinal5	22	F	Psychomotor therapist	0 (2 years of internship)
ProFinal6	51	F	Speech therapist	26
ProFinal7	29	F	Speech therapist	3
ProFinal8	58	F	Teacher	36
ProFinal9	31	F	Speech therapist	4
ProFinal10	54	F	Speech therapist	21
ProFinal11	48	F	Teacher	25
ProFinal12	23	F	Psychomotor therapist	0 (2 years of internship)
ProFinal13	54	F	Teacher	25
ProFinal14	60	M	Teacher	37
ProFinal15	62	F	Teacher	32
**Age and years of experience at the research beginning*				

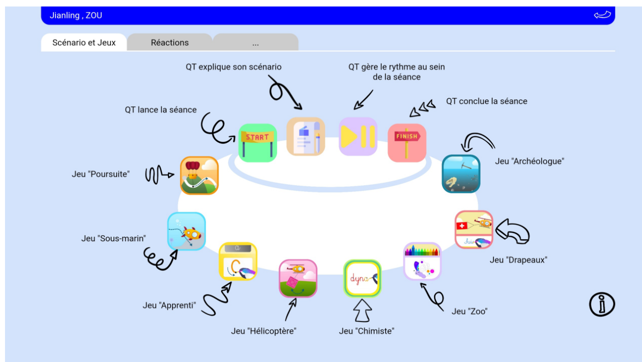
- **Pragmatic quality (PQ)**: describes the usability of the product and indicates how well it enables users to achieve their goals (in the sense of realizing a task);
- **Hedonic-stimulation quality (HQ-S)**: indicates the extent to which the product can support the need for stimulation;
- **Hedonic-identity quality (HQ-I)**: indicates to what extent the product allows the user to identify with it;
- **Overall attractiveness (ATT)**: describes the overall value of the product based on the perception of pragmatic and hedonic qualities.

We used the SUS to evaluate the WoZ interface and the AttrakDiff 2 to assess the robot's behaviors.

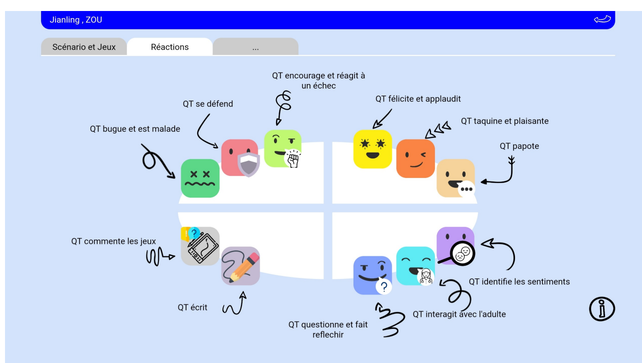
3 Result

At the final version of our system, we created 120 different behaviors classified into 33 menus according to their usage. In the interface presentation, we designed two web pages, one is the login page, and the other is the control page. The login

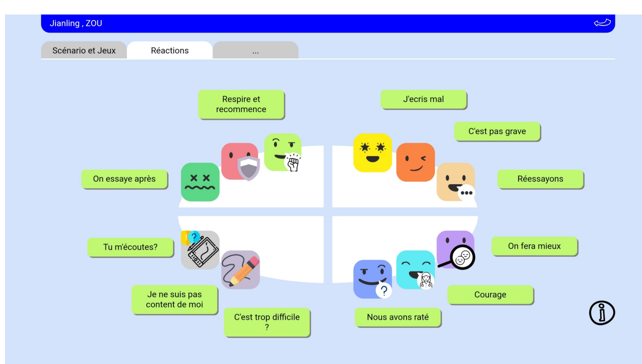
(a) Login page



(b) Control page – scenario and game tab



(c) Control page – reaction tab



(d) Buttons of a menu

Fig. 3. The final version of the Wizard of Oz interface

page can collect the names of the participants, which can be included in some pre-programmed behaviors, where the robot calls their names for personalization. The control page includes the 33 menus into two tabs: scenario and game tab, reaction tab. As shown in the **Figure 3**, we also used dividers to sub-categorize menus to make it easier for users to find the buttons they need quickly. When the user double-clicked the menu icon, all allocated buttons expanded in an ellipse, starting from the menu's position. Every button was in the same color as the icon of the menu it belonged to. Once the button was clicked on, the robot played the corresponding predefined behavior.

As illustrated in **Figure 4**, even though the scores given by participants are spread out, most of the scores of our interface are above 70 at the SUS, which is in the "Acceptable" interval of acceptability[1][2]. We can therefore say our interface is good and acceptable. Compared to speech therapists or teachers, psychomotor therapists gave more positive results for our system. Furthermore, in **Figure 5**, we represent the box-plot of the four sub-scales scores of Attrakdiff 2. Every sub-scale score corresponds to the average score of 7 items belonging to this sub-scale which is between -3 to 3. The average values of fifteen participants for these four sub-scales are more than one, which shows our robot's behaviors are positive in all these dimensions but with possible improvements.And the average of overall attractiveness score is higher than other dimension and has less dispersion, which shows in general our system is of good quality. However, one caregiver gave a SUS score of 45 to the WoZ interface, which corresponds to a "poor" system. And one caregiver gave a negative score in both AttrakDiff hedonic sub-scales. This caregiver explained he/she thought it was not easy to find the best-suited behaviors among so many choices and added he/she generally is not comfortable with new technology. This indicates that, on the one hand, the robot's behaviors could be improved to make it more easily usable and enjoyable. For example, a semi-autonomous system could be added to give choice suggestions. On the other hand, a tutorial session of the interface can be proposed to caregivers before practical therapist sessions with children.

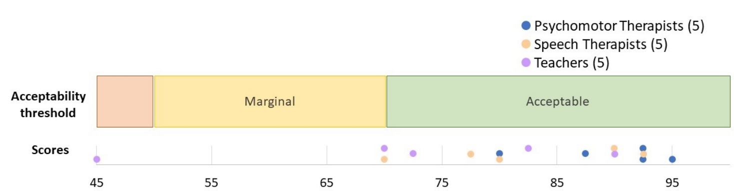


Fig. 4. Interpretation of the Total SUS Test Score (adapted from [1][2])

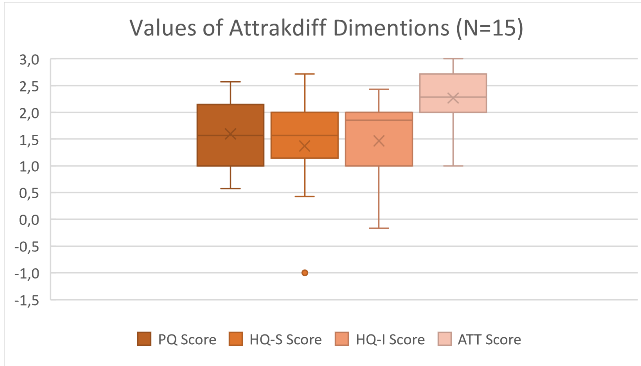


Fig. 5. Values of Attrakdiff Dimensions (N=15)

4 Discussion and Conclusion

This work focused on developing a usable WoZ interface for caregivers adapted to multi-platforms to facilitate handwriting learning in children with dysgraphia. The WOZ interface presented was developed and updated based on our iteration development experience with six children and six caregivers, which will be presented in our successive article. We found our system to be positive based on the usability testing results we got for the final version. However, since all the participants did not test the system with real children in an actual situation, the score of the usability evaluation might be superior. Even though we mentioned before the evaluation that each scale should be used to assess the corresponding dimension (WoZ interface or robot's behavior), many participants indicated having difficulty evaluating them separately, which can be a bias of user test results. In the future, we will conduct practical experiments with more caregivers and children based on the original iteration development experience set up. This will provide the database for our next step in automating the robot and getting more user testing results.

References

1. Bangor, A., Kortum, P., Miller, J.: Determining what individual sus scores mean: Adding an adjective rating scale. Journal of usability studies **4**(3), 114–123 (2009)
2. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. Intl. Journal of Human–Computer Interaction **24**(6), 574–594 (2008)
3. Berninger, V.W., Vaughan, K.B., Abbott, R.D., Abbott, S.P., Rogan, L.W., Brooks, A., Reed, E., Graham, S.: Treatment of handwriting problems in beginning writers: Transfer from handwriting to composition. Journal of Educational Psychology **89**(4), 652 (1997)
4. Chung, P., Patel, D.R.: Dysgraphia. International Journal of Child and Adolescent Health **8**(1), 27 (2015)

5. Chung, P.J., Patel, D.R., Nizami, I.: Disorder of written expression and dysgraphia: definition, diagnosis, and management. Translational pediatrics **9**(Suppl 1), S46 (2020)
6. Feder, K.P., Majnemer, A.: Handwriting development, competency, and intervention. Developmental Medicine & Child Neurology **49**(4), 312–317 (2007)
7. Gargot, T., Asselborn, T., Zammouri, I., Brunelle, J., Johal, W., Dillenbourg, P., Archambault, D., Chetouani, M., Cohen, D., Anzalone, S.M.: "it is not the robot who learns, it is me." treating severe dysgraphia using child–robot interaction. Frontiers in Psychiatry p. 5 (2021)
8. Gronier, G., Baudet, A.: Psychometric evaluation of the f-sus: Creation and validation of the french version of the system usability scale. International Journal of Human–Computer Interaction **37**(16), 1571–1582 (2021)
9. Grossard, C., Palestra, G., Xavier, J., Chetouani, M., Grynszpan, O., Cohen, D.: Ict and autism care: state of the art. Current opinion in psychiatry **31**(6), 474–483 (2018)
10. Hassenzahl, M., Burmester, M., Koller, F.: Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In: Mensch & computer 2003, pp. 187–196. Springer (2003)
11. Lallemand, C., Koenig, V., Gronier, G., Martin, R.: Création et validation d'une version française du questionnaire attrakdiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. European Review of Applied Psychology **65**(5), 239–252 (2015)
12. Nielsen, J.: Ten usability heuristics (2005)
13. Valentine, A.Z., Brown, B.J., Groom, M.J., Young, E., Hollis, C., Hall, C.L.: A systematic review evaluating the implementation of technologies to assess, monitor and treat neurodevelopmental disorders: A map of the current evidence. Clinical psychology review **80**, 101870 (2020)