Resource-Based Modality Selection in Robot-Assisted Cognitive Training

Aleksandar Taranović,
Aleksandar Jevtić
Institut de Robòtica i Informàtica
Industrial, CSIC-UPC
Barcelona, Spain
{ataranovic,ajevtic}@iri.upc.edu

Joan Hernández-Farigola,
Natalia Tantinyà,
Carla Abdelnour
Fundació ACE. Alzheimer Treatment
& Research Center, Barcelona, Spain
{jhernandez,ntantinya,cabdelnour}@
fundacioace.com

Carme Torras
Institut de Robòtica i Informàtica
Industrial, CSIC-UPC
Barcelona, Spain
torras@iri.upc.edu

ABSTRACT

The majority of socially assistive robots interact with their users using multiple modalities. Multimodality is an important feature that can enable them to adapt to the user behavior and the environment. In this work, we propose a resource-based modality-selection algorithm that adjusts the use of the robot interaction modalities taking into account the available resources to keep the interaction with the user comfortable and safe. For example, the robot should not enter the board space while the user is occupying it, or speak while the user is speaking. We performed a pilot study in which the robot acted as a caregiver in cognitive training. We compared a system with the proposed algorithm to a baseline system that uses all modalities for all actions unconditionally. Results of the study suggest that a reduced complexity of interaction does not significantly affect the user experience, and may improve task performance.

KEYWORDS

Cognitive Training, Multimodal Interaction, Socially Assistive Robotics

1 INTRODUCTION

With the increase in average life expectancy, the size of the elderly population is rising, which results in the augmentation of some chronic diseases like dementia [4]. Dementia is usually treated with Cognitive Stimulation Therapy (CST) that is provided in daycare centers and specialized institutions as a complementary technique to other treatments. In this work, we focus on cognitive training that deals with Sequential Memory, the ability to recall the information presented in order (both the sequence and the content). Each user has specific preferences and requirements, and the robot
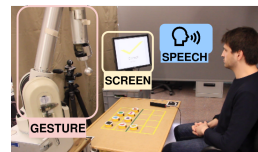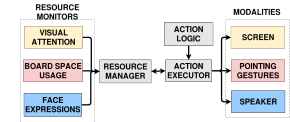
(a) Experimental setup   (b) Resource-based modality selection

Figure 1: Resource-based modality-selection system

should act according to them. We focus on personalizing the robot assistant regarding the usage of its modalities; hence we developed a resource-based modality selection (RBMS) algorithm. We use the proposed algorithm on an assistive-robotic system (Fig. 1(a)) that helps users to sort shapes printed on tokens of the same size. Our robotic system uses three different modalities (gesture, speech, and screen) to interact with its users. It can perform multiple actions using individual or combinations of modalities, transferring between them depending on the circumstances.

People often simultaneously use several modalities to perform an action, even when one is sufficient; e.g., a greeting can be done by speech, but also by waving. In most situations, using multiple modalities simultaneously can increase the favorable outcome of the action [3]. However, often it is impossible or impractical to use all modalities for every action. Additionally, the robot has four modalities for observing the behavior of the user (speech recognition in combination with facial expression recognition, face orientation detection, and board space usage).

2 RESOURCE-BASED MODALITY SELECTION

The same goal can be accomplished in various ways; e.g., Expressing an idea can be done verbally or by writing it down. In this paper, we define actions from the standpoint of the goal they achieve, not the manner they are executed. In the mentioned example, expressing an idea is an action, while speech and writing are modalities used to execute it. In other words, an action is a semantic element defined by the goal, and a modality is a manner of achieving the goal of an action. The proposed algorithm, shown in Fig. 1(b), executes actions using only modalities for which it has the appropriate resources. Action logic decides which action will be executed and sends that decision to the action executor. In our scenario, action logic depends on the exercise rules. The action executor communicates with the resource manager to obtain the resource state. Resource monitors track the usage of an assigned resource and inform the resource manager about any changes.

As stated in [2], resources in HRI can be divided into two categories: physical and cognitive. While the former includes resources

that characterize the environment and elements in it, the latter defines the cognitive state of the user. Prominent cognitive resources are visual attention and speaking floor, both of which are observed by our robotic system. In our resource-based algorithm, each modality requires specific resources. E.g., the robot should speak when the speaking floor is unoccupied so that its interlocutors can hear it. Similarly, showing an image on the screen is useful only when the user is looking at it, and thus the robot has the visual attention of the user. Losing a resource makes some modalities pointless, but their continued execution does not seriously affect the user. In case of the screen, if the user diverts their gaze, that action will not transmit the message to the user, but it is safe. However, it can be dangerous if the user moves their hand into the workspace of a robot arm in motion. Therefore, the system must track resources during action execution, and if a necessary resource is lost, it must perform an adequate termination function. For the robot arm example, a termination function can be a movement to a default position, at a safe distance from the user.

3 EXPERIMENTAL SETUP

The system consists of a Barrett's WAM robotic arm, screen, speakers, Kinect camera, web camera, and microphone. The robot can perform actions using any of the three modalities independently, without any conflict in their resource demands. The robot arm is used for gesture interaction with a user. The screen displays the shapes and robot's utterances. The microphone and speakers allow interaction through speech. A Kinect camera placed above the board is used to track the positions of the shapes and presence of the user's hands. The distribution of the tokens is evaluated each time the user retracts the hands from the board space. The front web camera is used to recognize user head orientation and facial expression: head orientation indicates the user's visual attention to different elements of the system or environment, and open mouth suggests the user's intention to speak.

Before starting the exercise,[1] the users are informed about the rules. Initially, they see the shapes one by one, with each picture being displayed for two seconds. The goal of the exercise is to sort the shapes in the initially shown order in the shortest possible time. After the last shape is displayed, the user is informed to begin sorting. If the token is placed in the correct position, the robot will provide feedback, and the user can make the next move. The user can verbally ask the robot for assistance, which is penalized by adding time (15 s) to the overall exercise completion time. However, if the user makes a mistake, penalization is even higher (30 s). This way, the users are encouraged to ask for help when they are not confident about the next shape. Moreover, if the user takes longer than a specific predefined time (15 s) per one shape move, the system will automatically help them, but they will be penalized with additional time (15 s). The role of penalization is to increase the user's motivation and focus on the exercise. Furthermore, after each correct move, the system provides confirmatory feedback.

4 RESULTS AND CONCLUSIONS

Experiments were performed with six participants with an engineering background, ages between 24 and 32 (M = 28.17, SD = 3.87).
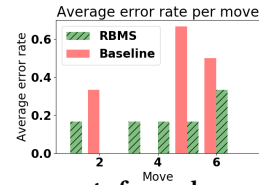

Figure 2: Average error rate for each move for the RBMS and the baseline systems

We compared our RBMS system with the baseline system that uses all modalities for all actions unconditionally. The experiments were within-subject and counter-balanced. After each exercise, participants filled a five-point Likert-scale questionnaire about how likable and useful was the completed exercise. After the second exercise, they also compared the two exercises. They could provide their overall opinion on the experiments in an open-ended question.

Numerous methods for evaluation of human-robot interaction have been proposed in literature, but there is no consensus among researchers how a comprehensive evaluation of an interactive robotic system should be performed [1]. However, obtained experimental results provide several insights into the user behavior. In the questionnaire, participants provided similar ratings of the RBMS and the baseline systems. In terms of usefulness: RBMS (M=3.5, SD=1.26), baseline (M=3.67, SD=1.11) and in terms of likability: RBMS (M=3.67, SD=0.75), baseline (M=3.67, SD=0.94). In the comparison question, the RBMS system was rated slightly better (M=3.17, SD=1.07). This shows that the system's complexity can be reduced by removing some of its features without affecting the user experience. Moreover, users made fewer errors when interacting with the RBMS system (M=0.86, SD=0.64), than with the baseline (M=1.29, SD=1.58). The average error rate for each move is shown as a bar graph in Fig. 2. Most of the participants made their moves in the first several seconds (M=2.08, SD=2.42), and none reached the time limit. In the open answer question, some participants stated that the RBMS system is better because it was faster. We noticed that some participants did not wait for the confirmatory gesture to finish. The RBMS system stops the confirmatory gesture when users put their hands into the board space to make the next move. This allows the RBMS system to respond faster than the baseline system. Obtained results suggest that users of the RBMS system tend to make fewer errors and perceive this system to be faster than the baseline system, with similar usefulness and likability. In future work, we will investigate these trends with a larger group of participants performing exercises of different levels of difficulty.

REFERENCES

[1] Suna Bensch, Aleksandar Jevtić, and Thomas Hellström. 2017. On Interaction Quality in Human-Robot Interaction. *Proceedings of the 9th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART* 1, 182–189.

[2] Crystal Chao and Andrea Thomaz. 2016. Timed Petri nets for fluent turn-taking over multimodal interaction resources in human-robot collaboration. *The International Journal of Robotics Research* 35, 11 (2016), 1330–1353.

[3] Matthew Turk. 2014. Multimodal interaction: A review. *Pattern Recognition Letters* 36, 1 (2014), 189–195.

[4] Johannes Wancata, M. Musalek, R. Alexandrowicz, and M. Krautgartner. 2003. Number of dementia sufferers in Europe between the years 2000 and 2050. *European Psychiatry* 18, 6 (2003), 306–313.

---

[1] A video of the exercise can be seen at: http://www.iri.upc.edu/groups/perception/hri2018_rbms/.