Adaptive Modality Selection Algorithm in Robot-Assisted Cognitive Training

Aleksandar Taranović[1], Aleksandar Jevtić[1] and Carme Torras[1]

Abstract— Interaction of socially assistive robots with users is based on social cues coming from different interaction modalities, such as speech or gestures. However, using all modalities at all times may be inefficient as it can overload the user with redundant information and increase the task completion time. Additionally, users may favor certain modalities over the other as a result of their disability or personal preference. In this paper, we propose an Adaptive Modality Selection (AMS) algorithm that chooses modalities depending on the state of the user and the environment, as well as user preferences. The variables that describe the environment and the user state are defined as resources, and we posit that modalities are successful if certain resources possess specific values during their use. Besides the resources, the proposed algorithm takes into account user preferences which it learns while interacting with users. We tested our algorithm in simulations, and we implemented it on a robotic system that provides cognitive training, specifically Sequential memory exercises. Experimental results show that it is possible to use only a subset of available modalities without compromising the interaction. Moreover, we see a trend for users to perform better when interacting with a system with implemented AMS algorithm.

I. INTRODUCTION

The number of people suffering from dementia is rising [1]. Consequently, demands for additional care from trained individuals are also growing. However, the number of caregivers is not increasing sufficiently to support the elderly population. Therefore, it is necessary to develop technological solutions to overcome this problem [2]. Socially Assistive Robots (SAR) provide a viable solution. Several papers examined different scenarios and the use of SAR [3], [4] as well as their application for cognitive training [5]. Moreover, research has shown that SAR yields better results than other technologies, like tablets [6].

While many applications in robotics have clearly-defined metrics related to goal attainment, SAR should also consider social features like trust, engagement, persuasiveness, etc., [7], [8] that are difficult to measure. This is why an assistive robot needs to develop a user profile and adjust its actions according to it. Moreover, a disability of the user should significantly modify the behavior of the robot. For example, the robot should not use speech, or other audio signals, when interacting with users with hearing problems.

Natural human-human interaction is multimodal [9], and therefore social robots should also be capable of multimodal

[1]Institut de Robotica i Informatica Industrial, CSIC-UPC, Barcelona, Spain. E-mail: {ataranovic, ajevtic, torras}@iri.upc.edu
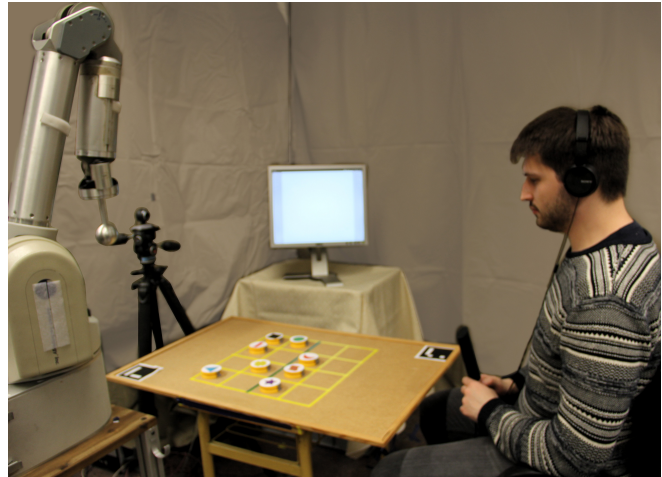
Fig. 1: Experimental setup for the sequential memory exercise.

interaction [10]. Multiple modalities enable robots to perform the same action in different ways. However, in certain situations the choice of modalities is unambiguous. If the user is not looking at the robot, speech could be used to obtain the attention of the user. However, often the same action can be performed in multiple manners. If the user is looking at the robot, it could do a pointing gesture to direct user's attention. Moreover, it could combine speech and gesture and use both modalities simultaneously. The exact choice of modalities should depend on the user profile, which we describe with modality preferences. Using gestures may seem intrusive to some users, while speech can confuse or fail when interacting with users with hearing impairment.

In this paper, we propose an Adaptive Modality Selection (AMS) algorithm for choosing a subset of modalities for performing actions. The goal of the algorithm is to execute actions only using modalities that satisfy certain conditions and that prove to be successful with the particular user. The success probability of a modality for a certain user defines the preference for that modality. When choosing preferred modalities, our algorithm relies on a Bayesian method for solving Multi-Armed Bandit (MAB), with Thompson sampling [11]. MAB has been used to personalize content on the web, like news article recommendation [12], it has also been used for selecting a robot policy [13]. Moreover, MAB has proven useful for grasp planning in scenarios with high uncertainty [14].

We describe the environment and the user using variables defined as resources [15], [16]. Depending on the values of resources during the use of modalities, we categorize modalities as successful or unsuccessful. This information

(a) Diagram of a system with the AMS algorithm.

(b) Executing action A_2 that has modality M_2 conflicted with modalities M_1 and M_j over resources R_1 and R_k, respectively.

(c) After resolving conflicts for resources between modalities, the AMS algorithm chose to use modalities M_1 and M_j.
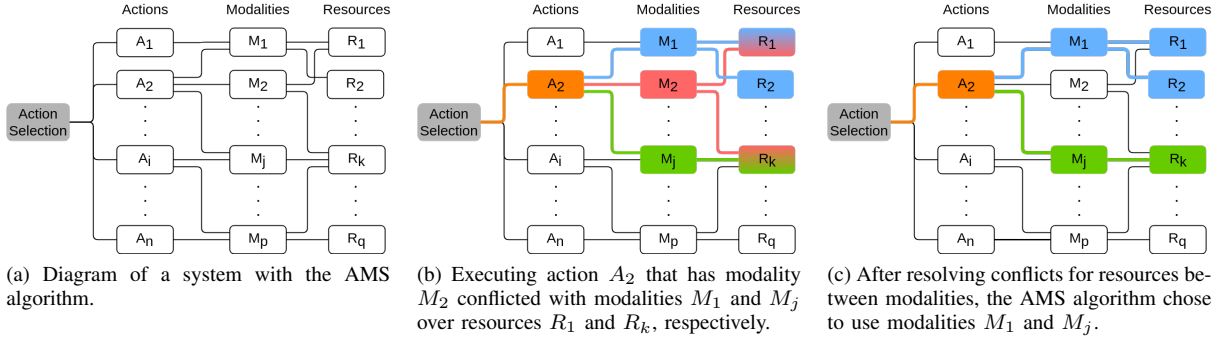
Fig. 2: Example diagram of a system implementing the AMS algorithm while selecting modalities.

is later used to determine the user preference towards that modality.

We applied the proposed algorithm in a robotic system for assisting users performing cognitive training, specifically a Sequential memory exercise (Fig. 1), which can be used as part of the therapy for people with mild cognitive impairment or Alzheimer's disease. The exercise has simple rules and goals but can be a challenge even for users without cognitive problems. Users are initially informed about a sequence of elements, and afterward, they should sort the elements in the presented order. In our application, the elements are shapes printed on tokens of the same size. Our robotic system can perform actions using three modalities: screen, speech, and gestures. To observe the user and track the progress of the exercise, the system uses two cameras, a Kinect and a webcam, and a microphone.

The proposed AMS algorithm enables the system to select the most successful modalities and remove unwanted redundancies. Initial experimental results show that the users rate the system with the AMS algorithm similarly to using all modalities unconditionally. Moreover, we observed a trend for users to make fewer errors when interacting with the AMS system.

II. ADAPTIVE MODALITY SELECTION ALGORITHM

The proposed AMS algorithm is defined by sets of actions $\mathbf{A} = \{A_1, A_2, ..., A_n\}$, modalities $\mathbf{M} = \{M_1, M_2, ..., M_p\}$, resources $\mathbf{R} = \{R_1, R_2, ..., R_q\}$ and their connections. Each action can be executed using one or more modalities, while each modality depends on the value of one or more resources. An example of the relationship between actions, modalities, and resources can be seen in Fig. 2a. The goal of the AMS algorithm is to choose a set of modalities to be used for executing an action. Robot actions are task-specific, and they are provided to the AMS algorithm, which defines a set of eligible modalities to execute such actions. For the sequential memory exercise, the selection of the next robot action is a result of either user performance or user request (e.g., error or request for help).

Resources $R_k, k \in \{1, ..., q\}$, represent variables that describe the environment and the user state and they are divided into physical and cognitive. A physical resource is the state of any object relevant to the robot. Moreover, the occupancy of the space can be modeled as a physical resource. The second

group of resources describes the cognitive state of the user. A common use of cognitive resources is to describe the focus of user attention; for example, when the user is looking at the screen can be used as an indicator to show certain content on the screen.

An action A_i is a semantic element specified by the goal it realizes, and it can be performed with different modalities. However, depending on the user's preferences and available resources, only a subset of modalities is used. One example of an action in our sequential memory exercise scenario is the *assist* action that informs the user of the correct shape for the current position. The robot can show the correct shape on its screen, point towards it using its arm, or say the name of the correct shape. In this case, the action is also characterized by the name of the correct shape, which is treated as an action parameter that is passed to the selected modalities.

A modality M_j is one way of instantiating an action, and each modality depends on certain resources. A modality in use can change the value of a resource; however, if the resource gets an undesired value, the modality is regarded as unsuccessful. Moreover, if a resource value changes from a desired to an unwanted value, the system should stop the modality and mark it as unsuccessful. For certain modalities, it is critical to check the resource state before applying them. For example, gestures performed by the robotic arm require the workspace to be unoccupied before the movement starts. If this is not the case, it can be dangerous to use it. Hence, the initial values of resources are the eliminating criteria when selecting some modalities. The time in which a resource should obtain a relevant value, and the duration it should hold that value are modality specific.

An important part of the AMS algorithm is the Thompson method for solving the MAB problem, which the algorithm uses in two steps. The MAB problem describes situations when an agent has to choose between different options, called arms, with each option providing an unknown reward. One way of solving the MAB problem is Thompson sampling. In our application, an arm is a modality described by its preference, which is also the probability it is successful. Hence, it can be described with a Bernoulli distribution, and our MAB is Bernoulli bandit problem [11]. Therefore, the Thompson sampling algorithm will use the Bernoulli conjugate, Beta distribution, to draw samples from. The Beta distributions for each arm is defined by the number

of successful and unsuccessful trials. Finally, Thompson sampling selects the arm with the largest sampled value.

Besides the conditional dependency between modalities and the values of resources, an important factor in modality selection is user preference towards a certain modality. Our algorithm associates and tracks the number of successful n_s^{ij} and unsuccessful n_u^{ij} uses for all pairs of actions A_i and modalities M_j. We define the preference towards modality M_j for action A_i as the ratio of the number of successful uses and the total number of uses of that modality for that action:

$$p_{ij} = \frac{n_s^{ij}}{n_s^{ij} + n_u^{ij}} \tag{1}$$

Firstly, the proposed AMS algorithm checks if all modalities are capable of executing the chosen action A_i by having their resources in the required initial state. If that condition is satisfied, the modality is added to the set of potential modalities $\mathbf{M}_p$. If no modalities pass this step, the system stops and informs the user.

Secondly, for each potential modality M_j, the system takes two samples x_{ij} and y_{ij} with Beta distributions defined by the number of times the modality was successfully, $n_{s,t}^{ij}$, or unsuccessfully, $n_{u,t}^{ij}$, used to execute the action A_i at the time t:

$$x_{ij} \sim Beta(n_{s,t}^{ij}, n_{u,t}^{ij}), \tag{2}$$

$$y_{ij} \sim Beta(n_{u,t}^{ij}, n_{s,t}^{ij}) \tag{3}$$

If $x_{ij} > y_{ij}$, modality M_j is left in the set of potential modalities $\mathbf{M}_p$, and all its resources are added to the set of relevant resources $\mathbf{R}_r$. Otherwise, the modality M_j is eliminated from set $\mathbf{M}_p$. This step is regarded as if the system is choosing between two arms. If the set of potential modalities $\mathbf{M}_p$ is empty after this step, the system creates random variables x_{ij} and y_{ij} again, until at least one modality is selected as potential. Afterward, all relevant resources R_r are checked for conflicts.

If multiple modalities $\mathbf{M}_c = \{M_{c_1}, M_{c_2}, ..., M_{c_N}\}$, require the same resource, the algorithm selects the modality M_j:

$$j = \underset{j \in \{c_1, c_2, ..., c_N\}}{\operatorname{argmax}} x_{ij} \tag{4}$$

and removes other modalities that are in conflict with the selected modality M_j from the set of potential modalities $\mathbf{M}_p$. This conflicts-resolving procedure is equivalent to the Thompson sampling process for the best arm. Furthermore, this step is repeated until all the conflicts for resources are resolved. In the end, modalities that remain in the set $\mathbf{M}_p$ are used.

An example diagram of the AMS algorithm is shown in Fig. 2. In this example, action A_2 is executed, and modalities M_1, M_2, and M_j are initially added to the set of potential modalities. However, resources R_1 and R_k are required by multiple modalities (Fig. 2b). After a possible conflict resolution, the choice of modalities is M_1 and M_j as shown in Fig. 2c.

TABLE I: Randomly generated connections between modalities and resources for the general system simulations.

| modality	resources	modality	resources
m_0	r_7	m_5	r_2, r_3
m_1	r_0, r_7, r_8	m_6	r_4, r_5, r_7
m_2	r_1, r_6, r_7	m_7	r_1, r_2, r_8
m_3	r_0, r_1, r_2, r_3	m_8	r_3, r_5, r_7, r_9
m_4	r_1, r_4, r_7	m_9	r_7, r_9

III. SIMULATION OF AMS SYSTEMS

To examine the properties of the proposed AMS algorithm, we tested the system in two different types of simulations. Firstly, we analyzed a general system with ten modalities and ten resources (Table I). We generated random connections between modalities and resources, with a constraint that one modality can depend on the state of at most four resources. This constraint was put in place to enable the selection of multiple modalities simultaneously. Secondly, we simulated users that interact with the robotic system for supervising sequential memory exercises (Fig. 1). This system can perform two actions with three different modalities and three different resources (Fig. 4). The first set of experiments was performed to evaluate the overall performance of the system, while the second set was used to evaluate our use case.

A. Simulation of the General system

The General system (Table I) was tested in three different conditions for one action that was executed 500 times. Moreover, we repeated the simulations ten times, and all figures in Fig. 3 represent the mean value of all simulations. Firstly, we tested the system instances where all modalities have a high success probability (p=0.9). The goal was to test how different modalities are represented, especially the modalities that have numerous conflicts. Fig. 3a and 3d show that all modality preferences converge towards the high values (0.9). However, some modalities with multiple conflicts require more time, whereas modality m_5 is used more often than others due to a small number of conflicts. Secondly, we wanted to test how long it takes for the system to eliminate one unsuccessful modality m_0 (p=0.1) while others have high success probability (p=0.9). Fig. 3b and 3e show that the unsuccessful modality did not converge to the value 0.1, because the system learned that this modality has a low probability of success, and therefore was rarely chosen. Finally, we wanted to see how the system behaves when half of the modalities has a high success probability (p=0.9) and the other half has a low success probability (p=0.1). This simulation shows how fast the system learns different success probabilities for multiple modalities. A similar situation occurred as in the previous condition, and we can see that unsuccessful modalities are filtered out after a relatively small number of interactions, and are rarely used afterward.

### B. Simulation of the sequential memory exercise

In the simulation of our use case, two actions are executed: confirm and assist. The former informs users that they correctly positioned the last token, while the latter advises
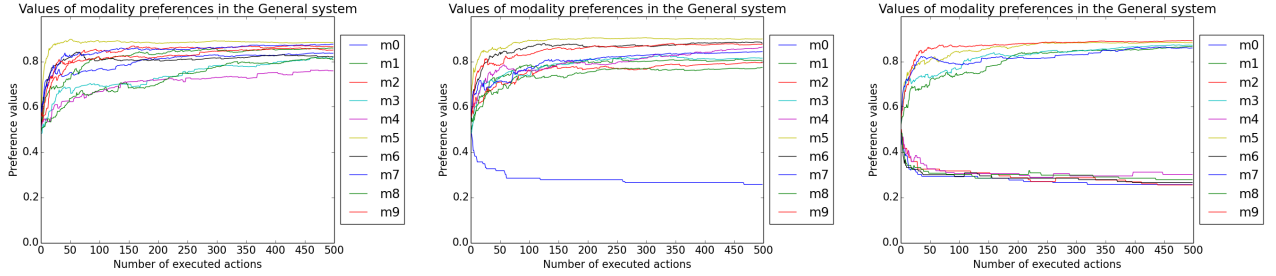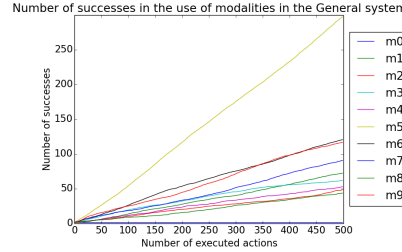
(a) Modality preferences in the simulation of the General system when all modalities have success probability p=0.9.

(b) Modality preferences in the simulation of the General system when all modalities have success rate p=0.9, except modality m_0 (p=0.1).

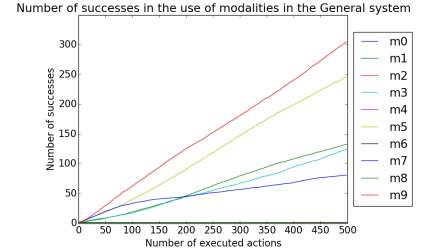(c) Modality preferences in the simulation of the General system when modalities with odd indexes have success probability p=0.9, and modalities with even indexes have success rate p=0.1.
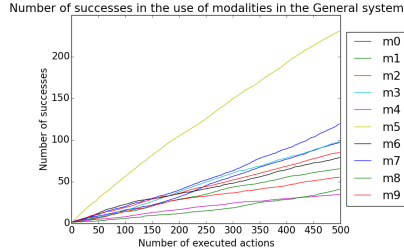
(d) Number of successes in the use of given modalities in the simulation of the General system when all modalities have success probability p=0.9.

(e) Number of successes in the use of given modalities in the simulation of the General system when all modalities have success rate p=0.9, except modality m_0 (p=0.1).

(f) Number of successes in the use of given modalities in the simulation of the General system when modalities with odd indexes have success probability p=0.9, and modalities with even indexes have success probability p=0.1.

Fig. 3: The simulation results for the General system.

users about the correct shape. Both actions can be performed using three different modalities: speech, screen, and gesture. The screen requires the visual attention of the user, while gesture requires that the visual attention is on the robotic arm and that the board space is free. The speech modality depends on the occupancy of the speaking floor. The diagram of the simulated system for the sequential memory exercise is shown in Fig. 4.

The probability of guessing the i-th shape in sequence p_{mi} is calculated using the assumption that the probability of correctly selecting a shape linearly deteriorates as the shape number increases. We model this assumption with

Fig. 4: The diagram of the system for sequential memory exercises.

Fig. 5: Sequential memory exercise simulation.

a linear function defined by the initial probability p_{Minit} and deterioration rate for one step p_{Mstep}. However, the probability of correctly guessing p_{gi} must be higher than the probability that a randomly selected shape is the correct one. Therefore, if there are n shapes in the sequence, the probability p_{gi} is calculated as:

$$p_{gi} = max(p_{Minit} - (i-1)p_{Mstep}, \frac{1}{n_s - i + 1}) \quad (5)$$

Parameters p_{Minit} and p_{Mstep} influence how often the assist action will be used.

We tested the system with users with different profiles (Table. II) which did ten exercises. The users that have success probability for all modalities less than 0.5 were not considered. This condition was put in place because adapting to users that do not want or cannot interact with the system is not possible and is not the goal of the proposed algorithm. Moreover, users that dislike interacting with the system can lead to a large number of failed actions, thus significantly increasing the entire simulation process. Therefore, out of 128 possible user profile we eliminated 16, and in the end, we simulated the exercise with 112 different users. The goal was to observe how many exercises it would take for the values of modality preferences to converge. Results show that the values change insignificantly after only two exercises.
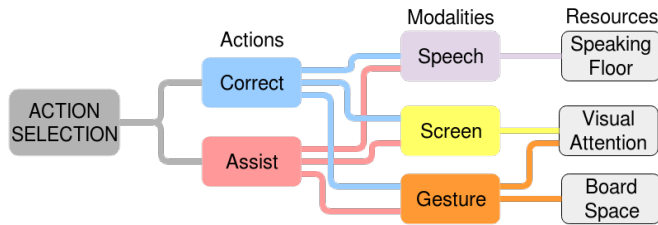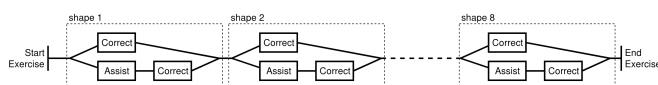
IV. EXPERIMENTS

The main components of our robotic system are a screen, a WAM robotic arm, headphones, a microphone, a webcam

TABLE II: The possible values of relevant parameters for different user profiles.

Variable	Possible values
screen success probability	0.1, 0.3, 0.7, 0.9
gesture success probability	0.1, 0.3, 0.7, 0.9
speech success probability	0.1, 0.3, 0.7, 0.9
memory profile	$(p_{Minit} = 0.6, p_{Mstep} = 0.05)$, $(p_{Minit} = 0.9, p_{Mstep} = 0.05)$

and a Kinect camera (Fig. 1). The screen shows shapes and textual information, while the WAM robotic arm can perform various gestures. The headphones can generate utterances, while the microphone is used for voice recognition. Above the table is a Kinect camera with dual use. First, it provides RGB images used for tracking the positions of the tokens, and secondly, it is used for detecting if the user's hand is intruding the board space. The webcam is positioned on the opposite side of the table, directed towards the user's face. The images that the webcam provides are used to detect the user's head orientation, needed for detecting visual attention of the user. Moreover, those images are used to detect if the user's mouth is open. If it is open, then the speaking floor resource should be marked as occupied by the user.

The exercise[1] starts by showing shapes on the screen and speaking their names in a specific order. The goal of the exercise is to arrange the shapes on the board in the initially shown order in the shortest time. The user is offered the possibility to ask for assistance; however, this action increases the overall exercise time by 15 s. This duration of time addition was determined by our initial test, as the maximum before a user makes a guess. To motivate users to ask for assistance when they are not sure what is the correct shape, they are penalized with an additional 30 s if they make a mistake. In case the user doubts what the system tried to do, he or she can verbally ask the system to repeat the action. In this case, the system considers the action, and all used modalities, as unsuccessful.

As in the simulation, the developed system for sequential memory exercises can execute two actions: confirm and assist, and it has three modalities and three resources (Fig. 4). Unoccupied board space is an initial condition for the gesture. In case of the speech, the system demands that the speaking floor is initially free since we do not want the robot to speak over the users. Both the gesture and speech modalities have predefined time periods that need to elapse for them to be successful. Those periods represent the time it takes for the modality to successfully fulfill the goal of the action. In other words, they represent the time a relevant resource needs to have the desired values so that the system can successfully finish the action. In case of the speech, this modality is successful if the speaking floor remains unoccupied by the user until it finishes speaking the required information. The gesture modality is successful if the board space is unoccupied, and the user's attention is directed towards the robotic arm so that its gestures are observed.

[1]A demonstration of an exercise can be seen at: http://www.iri.upc.edu/groups/perception/AMSAlgorithm.

The screen modality depends on the visual attention of the user, which does not need to initially have a desired value. However, it has a predefined time period in which it needs to obtain its relevant resource (visual attention should be on the screen). Hence, if the user never looks at the screen, this modality is considered as a failure.

A. Participants

A total of 12 participants, ages between 22 and 33 (M=26.92, SD=3.94), including students of different degree levels, and the administrative staff were recruited for the experiments. The experiments were within-subject with participants randomly divided into two groups. The first group initially performed the first five exercises with the AMS system followed by five exercises with the baseline system, while the other group performed in the opposite order. In the baseline exercises, the system used all modalities for all actions. The number of exercises was chosen because simulations showed that the AMS system learns user preferences after a few exercises. Moreover, it enables us to finish the experiments in a reasonable amount of time for each user (30 min), which is an important factor since we do not want to make our participants tired. These experiments were intended to be a "proof of concept" before testing the system in a daycare facility with patients with mild dementia.

Before doing the exercise, users were informed about the rules of the exercise, signed a consent form and they did an initial exercise to become familiar with the system. During the initial exercise, we showed them how different modalities work. Special focus was put on the gesture for the correct action, since it may not be intuitive for all users. Finally, they were asked to fill out a questionnaire where they rated the interaction with the system after the fifth and after the last exercise. Additionally, we compared the performance in the experiment regarding the number of errors and the duration of exercises. At the end of the experiment, the participants were offered refreshments for their involvement.

B. Experimental results

The main part of the questionnaires are the ten questions with a seven-point Likert-scale (1 - strongly disagree, 7 - strongly agree). The first five questions were about the first five exercise, while the other five about the last five exercises. The questions were formulated in the same way for both parts and all participants filled the same questionnaire. The participants did not know if they were interacting with the AMS system or the baseline system, they only knew that the way the system interacts with them is different in the first five exercises in comparison to the following five. The results of the questionnaires are presented in Table III. The difference in responses between the AMS and the baseline system was insignificant, with the AMS being rated slightly better in most of the questions. This fact indicates that it is possible to not use some of the modalities and hence simplify the complexity of the system. The obtained results show that all users for all actions have the highest preference towards speech, followed by the screen and, in the last place the

TABLE III: The mean values of answers in the questionnaire with a seven-point Likert-scale.

Question	AMS	Baseline
I liked interacting with the robotic system.	4.92	4.67
The behavior of the robotic system distracted me.	2.33	2.66
I felt comfortable interacting with the robotic system.	4.50	4.58
The feedback and/or assistance of the robotic system was provided in a useful way for doing the exercise.	4.92	5.16
I preferred interacting with the robotic system in the fifth rather than in the first exercise.	4.08	3.75

gesture. Overall, the least preferred modality was the robot gesture executing the "correct" action.

Since the total number of errors differs between users, we calculated the ratio between the total number of errors for the AMS and the baseline and obtained mean values for the whole set of users M=0.85, SD=0.38. Because of the high deviation, it is not possible to make a statistically significant conclusion that users make fewer errors when interacting with the AMS system, but the results show a trend in this direction that we plan to further investigate in our future work. Additionally, for each user, we computed the ratio between the exercise completion times using the AMS and baseline systems. The obtained values (M=0.93, SD=0.17) show that the systems had similar performance and that the users were able to finish the exercise slightly faster when interacting with the AMS system.

V. Conclusion and Future Work

In this paper, we propose the Adaptive Modality Selection algorithm for human-robot interaction and its application to cognitive training. The algorithm enables robotic systems to use only modalities that satisfy certain conditions, that we described as resource states. Through interaction, the algorithm learns the user's preferred modalities for executing actions. To validate the AMS system, we simulated its behavior in a general system, with random connections between modalities and resources. The results show the system can successfully adapt to users by learning their modality preferences.

Afterward, we simulated our use case, a sequential memory exercise performed by a carefully designed repertoire of users. The simulated results indicated that a small number of exercise performances would be enough for the system to learn the preferences of the users. The experimental results suggest that it is possible to reduce the number of used modalities without compromising the interaction. Furthermore, we saw a trend for users to perform better when interacting with the AMS system, which, although promising, needs to be validated with a thorough experimental design.

Now that we have a reasonably working prototype, we plan to evaluate it on patients with mild cognitive impairment in a daycare facility with which we collaborate in the SOCRATES project. Moreover, we would like to test the system with users with hearing and vision difficulties. Additionally, we plan to implement the AMS algorithm on robotic systems for other forms of cognitive training. Therefore, we will test the AMS algorithm in different scenarios with a higher number of actions and a diverse set of modalities.

References

[1] World Health Organization, "Dementia a public health priority," *Dementia*, p. 112, 2012.

[2] A. Di Nuovo, F. Broz, N. Wang, T. Belpaeme, A. Cangelosi, R. Jones, R. Esposito, F. Cavallo, and P. Dario, "The multi-modal interface of robot-era multi-robot services tailored for the elderly," *Intelligent Service Robotics*, vol. 11, no. 1, pp. 109–126, Jan 2018.

[3] S. M. Rabbitt, A. E. Kazdin, and B. Scassellati, "Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use," *Clinical Psychology Review*, vol. 35, pp. 35 – 46, 2015.

[4] M. A. Salichs, I. P. Encinar, E. Salichs, Á. Castro-González, and M. Malfaz, "Study of scenarios and technical requirements of a social assistive robot for alzheimer's disease patients and their caregivers," *International Journal of Social Robotics*, vol. 8, no. 1, pp. 85–102, Jan 2016.

[5] A. Tapus, C. Tapus, and M. J. Mataric, "The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia," in *2009 IEEE International Conference on Rehabilitation Robotics*, June 2009, pp. 924–929.

[6] J. A. Mann, B. A. MacDonald, I.-H. Kuo, X. Li, and E. Broadbent, "People respond better to robots than computer tablets delivering healthcare instructions," *Computers in Human Behavior*, vol. 43, pp. 112 – 117, 2015.

[7] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, ser. HRI '06. New York, NY, USA: ACM, 2006, pp. 33–40.

[8] S. M. Anzalone, S. Boucenna, S. Ivaldi, and M. Chetouani, "Evaluating the engagement with social robots," *International Journal of Social Robotics*, vol. 7, no. 4, pp. 465–478, Aug 2015.

[9] M. Turk, "Multimodal interaction: A review," *Pattern Recognition Letters*, vol. 36, pp. 189 – 195, 2014.

[10] T. Kollar, A. Vedantham, C. Sobel, C. Chang, V. Perera, and M. Veloso, "A multi-modal approach for natural human-robot interaction," in *Social Robotics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 458–467.

[11] S. Agrawal and N. Goyal, "Analysis of thompson sampling for the multi-armed bandit problem," in *Proceedings of the 25th Annual Conference on Learning Theory*, S. Mannor, N. Srebro, and R. C. Williamson, Eds., vol. 23. PMLR, 25–27 Jun 2012, pp. 39.1–39.26.

[12] O. Chapelle and L. Li, "An empirical evaluation of thompson sampling," in *Advances in Neural Information Processing Systems 24*. Curran Associates, Inc., 2011, pp. 2249–2257.

[13] P. Matikainen, P. M. Furlong, R. Sukthankar, and M. Hebert, "Multi-armed recommendation bandits for selecting state machine policies for robotic systems," in *2013 IEEE International Conference on Robotics and Automation*, May 2013, pp. 4545–4551.

[14] M. Laskey, J. Mahler, Z. McCarthy, F. T. Pokorny, S. Patil, J. van den Berg, D. Kragic, P. Abbeel, and K. Goldberg, "Multi-armed bandit models for 2d grasp planning with uncertainty," in *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, Aug 2015, pp. 572–579.

[15] C. Chao and A. Thomaz, "Timed petri nets for fluent turn-taking over multimodal interaction resources in human-robot collaboration," *The International Journal of Robotics Research*, vol. 35, no. 11, pp. 1330–1353, 2016.

[16] A. Taranović, A. Jevtić, J. Hernández-Farigola, N. Tantinyà, C. Abdelnour, and C. Torras, "Resource-based modality selection in robot-assisted cognitive training," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '18. New York, NY, USA: ACM, 2018, pp. 255–256.