# A POMDP Framework for Modelling Human Interaction with Assistive Robots

Tarek Taha, Jaime Valls Miró and Gamini Dissanayake

*Abstract*— **This paper presents a framework for modelling the interaction between a human operator and a robotic device, that enables the robot to collaborate with the human to jointly accomplish tasks. States of the system are captured in a model based on a partially observable Markov decision process (POMDP). States representing the human operator are motivated by behaviours from the psychology of the human action cycle. Hierarchical nature of these states allows the exploitation of data structures based on algebraic decision diagrams (ADD) to efficiently solve the resulting POMDP. The proposed framework is illustrated using two examples from assistive robotics; a robotic wheel chair and an intelligent walking device. Experimental results from trials conducted in an office environment with the wheelchair is used to demonstrate the proposed technique.**

## I. Background

Human-robot interaction (HRI) is a branch of the robotics science that focuses on modelling, implementing and evaluating the collaboration between robotic systems and human partners to produce "human helper" systems that are practical, efficient and accepted as well as enjoyable. HRI has evolved rapidly from systems with only a limited teleoperation capabilities and sometimes with simple video feedback. The increasing number of application domains in which robots can/will be deployed in the future and the presence of humans in many of these domains are the main motivations driving further developments in HRI.

Modelling human behaviour is a challenging and a complex task and it is perhaps unreasonable to expect artificial systems that are able to model humans in great detail using the probabilistic models that are currently available. However, a better understanding of the underlying interaction and the psychological states of humans during that interaction can be obtained by examining studies in related fields such as psychology. These can then be used to motivate the development of models that are adequate and tractable to realise the objective of building intelligent devices that can effectively collaborate with humans. Remainder of this section is devoted to providing a background to the human action cycle, and the literature on the use of POMDPs in modelling human robot interaction. This is followed by a brief outline of the organisation of this paper.

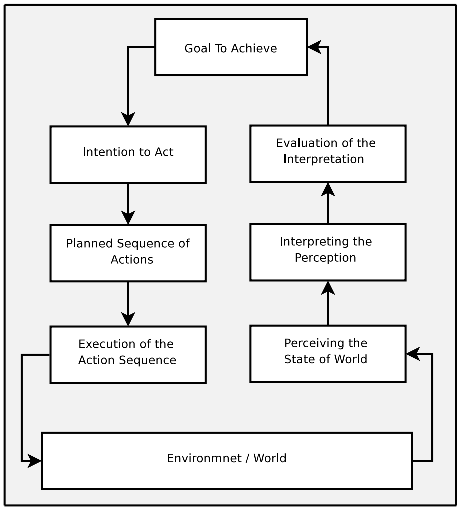
Goal To Achieve
Intention to Act
Planned Sequence of Actions
Execution of the Action Sequence
Evaluation of the Interpretation
Interpreting the Perception
Perceiving the State of World
Environmnet / World

Fig. 1. The Seven Stages of Action [1].

### A. Human Action Cycle

The "Human Action Cycle" described by Norman in "The Design of Everyday Things" [1] provides the framework humans go through to plan, execute and evaluate activities or tasks. The "Stages of Execution and Evaluation" captures the human behaviour in this context where the human usually starts by defining what needs to be achieved, then plan a sequence of intended actions to be performed. During plan execution the consequences of the actions taken are examined and evaluated. The evaluation stage starts with the human perception of the world which is first interpreted according to expectations and then compared/evaluated with respect to both the intentions and goals. The seven stages of action that captures these ideas is described in Figure 1. Norman used this concept to develop guidelines on how to engineer a device and how to evaluate it's performance and design an intuitive interaction/interface layer.

Relationship between intention and planning also needs to be examined to further exploit the work of Norman in HRI. Bratman, in his books "Faces of Intentions" [2] and "Intention, Plans, and Practical Reason" [3] develops and explains the theory of intention. According to Bartman, intentions are treated as elements of partial plans of action. These plans are considered a core component of practical reasoning, and can be thought of as roles needed to support the structure of human activities over time. Bratman presents impact of these ideas on a wide range of issues, including the relationship between intention and intentional action,

This work is supported by the Australian Research Council (ARC) through its Centre of Excellence programme, and by the New South Wales State Government. The ARC Centre of Excellence for Autonomous Systems (CAS) is a partnership between the University of Technology Sydney, the University of Sydney and the University of New South Wales.

All authors are with the ARC Centre of Excellence for Autonomous Systems (CAS), Faculty of Engineering, University of Technology Sydney (UTS), NSW2007, Australia {t.taha,j.vallsmiro,g.dissanayake}@cas.edu.au

and the distinction between intended and expected effects of what one intends. Bratman also elaborates the commitment involved in intending, and explores its implications on the fundamental understanding of temptation and self-control, shared intention and shared cooperative activity, and moral responsibility. Selection of the states for representing human status and the hierarchical relationships between these states presented in this paper are motivated by the work of Norman and Bratman described above.

### B. *POMDPs in HRI*

POMDPs have been especially effective in health care and assistance applications. For instance in [4], a real-time system to assist persons with dementia during handwashing was presented. The assistance was given in the form of verbal and visual prompts, or through the call for a human caregiver's help. The system used only video inputs, and combined a Bayesian sequential estimation framework to track hands and towels. The system was successful in estimating user states, such as awareness, responsiveness and overall dementia level.

This paper exploits the conditional independence between some of the states that represent the human behaviour, making it possible to use techniques based on algebraic decision diagrams (ADDs) to simplify and solve the POMDPs generated. ADDs [5] are a general case of BDD (binary decision diagrams).

The remainder of the paper is organised as follows. The proposed strategy for modelling the human interaction layer is presented in section II. Section III illustrates how this framework can be used to efficiently solve human robot interaction associated with two assistive robotic applications. This is followed by results from experiments with a robotic wheelchair in section IV, and the conclusions are presented in section V.

## II. A Human Aware POMDP Model for HRI

It is proposed to describe the interaction between a human and a robot in the following way. The human starts the interaction with a robotic agent by indicating a **task** that the robotic agent is capable of performing. The human then plans a sequence of **intentions** (these can be voice commands, hand gestures, joystick indication, touch screen input...) to refer to direct actions that the robot should perform to achieve the task; or provides a global goal encapsulating a sequence of actions (for example issuing a command such as ”get me a cup of water”). During the action execution, the human will evaluate the interaction and reveal a feedback in the form of **satisfaction** with the ingoing interaction, thus allowing the robot to enhance or re-evaluate the actions taken to better comply with the human's needs. Moreover, the robot evaluates the situation through access to the human's **status** (such as the competence level of the user) in order to adapt its behaviour and preserve the natural social aspects of this collaboration. Conceptually, it is proposed to extend the classical two gulf model of seven stages of action in the literature [1] by incorporating a decision making block based on a POMDP as depicted in Figure 2, to capture the communication layer that facilitates the flow and interpretation of information between the human and the robot. This information flow can then be translated into an action plan executed by the robot, and supervised/guided by the human.

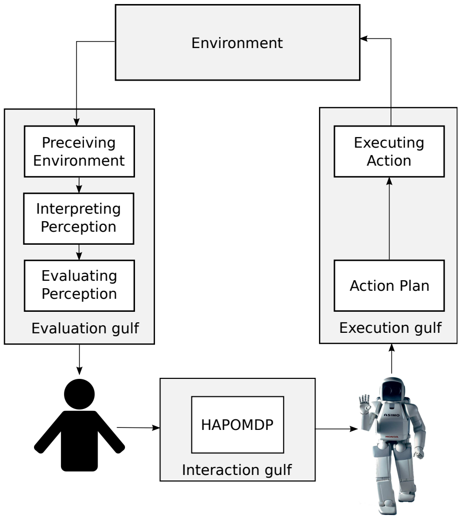

Fig. 2. An extension to the seven stages of action theory to include a third gulf (Interaction gulf) acting as a communication bridge between the human and the robot.

It is proposed that variables “Tasks”, “Intentions”, “Satisfaction” and “Status” be used to model the interaction within the POMDP. Subsection II-A discusses interaction variables in detail.

### A. *HRI Variables*

As discussed variables such as user's satisfaction and intention are important to accurately recognise the user's needs and measure the success of the interaction. The proposed model contains four HRI relevant state variables that are described below. The states can be factored into a set of variables that represent the natural layers of the HRI structure, making it possible to simplify the resulting POMDP. It is important to note that the state space can also include other variables relevant to the application at hand as will be discussed in section III.

*1) Intention, $In$:* In much of the previous work on HRI, intention and plan recognition are used as if they represent the same thing. However, in practice, intentions and plans should be treated separately for a better understanding of the user's intention. From a philosophical point of view [6], [7] intentions is the attitude that directs future long term planning. In this work intentions are used to direct an intermediate action and a set of these actions determine a task. Therefore, intention represents a variable that helps in selecting an immediate action to achieve a long term task or plan.

*2) Satisfaction, $Sa$:* This variable represents the user's satisfaction in the outcome of the collaboration with the robot. This indicates the level of success in the interaction. The combination of the intention variable and the satisfaction variable provides a powerful mix that can adapt and perfect the interaction. The satisfaction variable can also be thought of as a switch that when modelled properly enables the human interacting with the robot to dictate a change of plan or stop an action without having to physically reset the system or restart the algorithm. It can also work as a reward function that awards the robot for executing the correct action, and penalises it when the wrong action has been executed.

*3) Status, $St$:* This variable defines the status of the users, for example the awareness level, the load/stress level, response level and competence level. Identifying this variable is essential in determining the level of confidence in the user's contribution to the planning process and/or the amount of assistance the human requires from the robotic system. This variable is very useful when modelling assistive robotic applications where one can not always rely on the human to give an informative and clear indication of his/her desires. For instance, people with dementia are not capable of recognising or remembering their requirements or current state, and wheelchair drivers with a hand tremor are usually unable to give a noise-free joystick input to indicate their destination. Status could be observed using external sensors, predicted based on prior domain knowledge or manually specified as inputs to the system.

*4) Tasks, $Ta$:* This variable represents the set of missions to be executed during the collaboration. This, for example, could be a navigation task, a walking assistance task, a human following task, assisting with door opening or even a calendar notifications reminding the user to take medicine or watch his favourite TV show. The tasks variable can be also be used as a switching parameter that facilitates the modelling of a multi-task sequential decision making process.

### *B. Modelling and Solving the POMDP*

Typically, three steps are necessary to produce a solution for a problem formulated within the POMDP framework. The first step involves specifying the model. In the proposed POMDP structure, this includes:

- Determining the set of $tasks$ the system can perform
- Determining the set of $intentions$, $status$ and $satisfaction$
- Defining the observations and with what kind of sensors should be used to obtain them
- Determining the $actions$ that the system can execute
- Defining the $reward$ function
- Specifying the conditional dependencies between the variables

The second step is to learn the transition and observation functions from a set of training data. The third step involves solving the resulting POMDP model to obtain the optimal policy, which translates observations to the most rewarding actions.

Conditional independence between some of the state variables makes it possible to develop a concise ADD representation [5] of the resulting POMDP. Before building the ADD, the temporal effect of actions on variables should be presented properly. Actions usually have a direct effect on variables under certain conditions and these implicitly determine state transition. A DBN (Dynamic Bayesian Network) can be used to represent the effect of each action $a \in A$ on the variables probabilistically. If two set of variables $X = \{X_1, ..., X_n\}$ and $X^{'} = \{X_1^{'}, ..., X_n^{'}\}$ are used to present the pre and post action state of the system, directed arcs from a variable $X$ to $X^{'}$ indicate the probability of variable $X^{'}$ given it's parent $X$ after the execution of action $a$. Typically, a CPT (conditional probability table) is needed for each post-action variable $X_1^{'}$ to specify the probability of the variable after action $a$ given it's parents. However, ADDs can be used instead to represent the structure by exploiting the regularities in the CPTs [8]. Reward function can also be represented similarly. Once the model is defined as a collection of ADDs the dynamics of the system in the factored form will become:

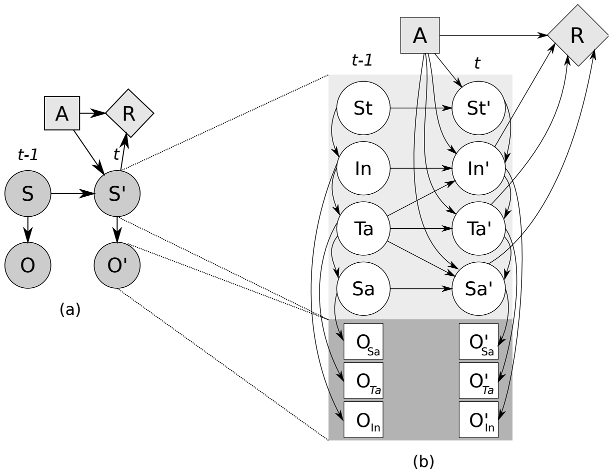

Fig. 3. Progression from (a) the classical POMDP structure, to (b) a 2-slice DBN with the interaction layer/gulf.

$$\begin{aligned}
T(s,a,s^{'}) &= T(<x_1,x_2,...,x_n>,a,<x_1^{'},...,x_n^{'}>)\\
O(a,s^{'},o) &= O(a,<x_1^{'},...,x_n^{'}>,o)\\
P_{a,o} &= T(<x_1,...,x_n>,a,<x_1^{'},...,x_n^{'}>)\\
&\quad O(a,<x_1^{'},...,x_n^{'}>,o)
\end{aligned}$$

The generic interaction model described above assumes that the $status$ of a user determines her/his ability to provide an $intention$, while $intentions$ define the $task$ in hand, and the $task$ is evaluated by the $satisfaction$. The $tasks$, $intentions$, and $satisfactions$ can be observed by external non-intrusive sensors (cameras, temperature sensors ...) or by intrusive sensors (like joysticks, buttons, touch screens ...). This model is illustrated in Fig. 3, where the state space is factored into a set of the four variables: $status$ ($St$), $intention$ ($In$), $tasks$ ($Ta$) and $satisfaction$ ($Sa$). The arcs in this figure reflect the following concept: intention ($In$), task ($Ta$) and satisfaction ($Sa$) are not directly observable, but inferred from observations. Changes in satisfaction ($Sa$)can be caused by the task ($Ta$) independent of the user"s

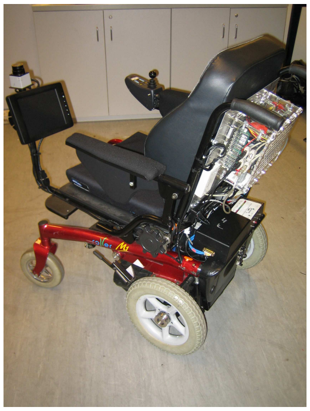

Fig. 4. The robotic wheelchair platform used in the experiments.

direct status/ability ($St$) or intention ($In$). The task ($Ta$) is determined from the user's intention($In$) which is in turn affected by the user's ability to give that indication ($St$). Intra slice arc connections show that the Task ($Ta$) at time $t-1$ can have effect on the intention ($In$), task ($Ta$) and satisfaction ($Sa$) in time slice $t$. Reward is given for intentions ($In$) that lead to the correct task execution($Ta$) and result in user satisfaction($Sa$). This model structure acts as a guideline and is not strict. It can be adapted to different applications if needed. The variable sets, however, are assumed constant. With this factored representation, the transition function of the model in Fig. 3 can be represented as:

$$\begin{aligned} Pr(S'|S,A) &= Pr(St',In',Ta',Sa'|St,In,Ta,Sa) \\ &= Pr(St'|St,A)Pr(In'|St',In,Ta,A) \\ &\quad Pr(Ta'|In',In,Ta,A) \\ &\quad Pr(Sa'|Ta',Ta,Sa,A) \end{aligned}$$

Once the POMDP model variables are specified and the transition links between them are defined, the model states are translated into the appropriate ADD representation using the format developed by Hoey [8]. The model is then solved offline using the symbolic Perseus method [9] to obtain an optimal policy. When operating on-line, observations are used to update the state beliefs which are then used with the optimal policy to generate the most rewarding action.

## III. Selected Applications

Two robotic systems have been selected to demonstrate the viability of POMDP modelling for HRI interaction with the proposed framework.

### A. *Instrumented Wheelchair*

The robotic wheelchair used for this example is illustrated in Fig. 4. The aim of this application is to ensure that the interaction appears transparent to the user, demanding a minimal input from her/him to automatically perform the required fine motion control to take the user to an intended final destination. This is to be achieved by observing the joystick input to determine the user's immediate intended action, and combine this with knowledge of the current location to predict the final destination. At any time, the user is allowed to expresses his/her dissatisfaction of the interaction by giving a joystick input that is different from the direction of motion, to trigger an automatic action correction followed by a new action selection.

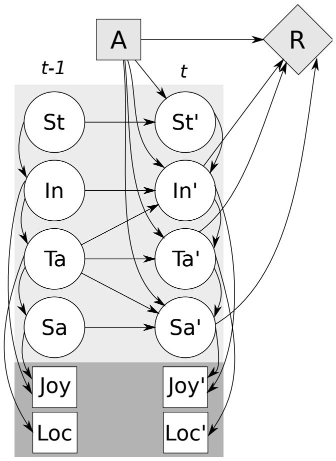

Fig. 5. Two slice DBN where: ($Ta$) is the variable set, ($Sa$) is the user's satisfaction, ($Int$) is the user's intention, ($St$) is the user's status, ($Joy$) is the user's joystick observation, and ($Loc$) is the location.

TABLE I
Wheelchair Model Variables

| State Variables | Values |
|---|---|
| Intention | Right, Left, Up, Down, Nothing |
| Satisfaction | Satisfied, Unsatisfied |
| Status | Competent, Struggling, Reliant |
| Task | Navigation |
| Joystick | Up, Down, Right, Left, Nothing |
| Actions | North, South, East, West, Stop |

The proposed framework results in the 2-slice DBN shown in Fig. 5. The user input is detected through the joystick observation ($Joy$). The intention ($Int$) and the location ($loc$) are then used to determine the user's destination in the ($tasks$) set. As described above, the joystick observation is also used to indicate user dissatisfaction.

The set of variables defining the POMDP model, listed in Table I,is described below.

- Task ($Ta$): The task variable here represent the navigation assistance task described using two state variable sets, location $L = \{S0, S1, \ldots, S49\}$ representing all nodes in the topological map, and destination $D = \{d0, d1, d2, d3, d4, d5\}$ representing the list of all possible destination nodes. The combination of these two variables determine the path the user needs to follow reach the destination.
- Intention ($In$): The intention here is the indication of the immediate action the user want the wheelchair to perform to get from one node to the next. It can be $In = \{Up, Down, Right, Left, Nothing\}$ and is observed using a joystick.
- Satisfaction $Sa$: Satisfaction set consists of only two values $Sa = \{Satisfied, Unsatisfied\}$, representing

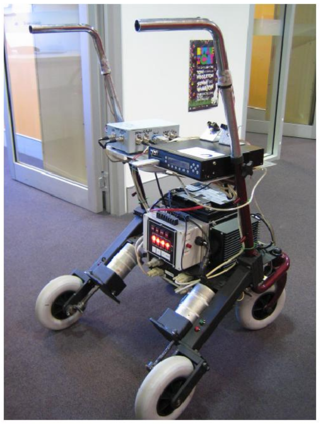

Fig. 6. The instrumented walker platform.

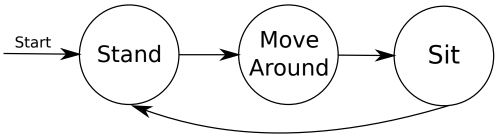

Fig. 7. State diagram of the walker transitions between tasks.

TABLE II
WALKER MODEL VARIABLES

| State variables | Values |
|---|---|
| Intention | Right, Left, Up, Down , Nothing, Stand, Sit |
| Satisfaction | Engaged, Cautious, Frustrated |
| Status | Competent, Struggling, Reliant, Distracted |
| Task | Stand-up, Sit down, Move-around |
| Strain gauges | High, Medium, None, Negative |
| Infra-red | Close, Far |
| Actions | North, South, East, West, Nothing, Lock Motors |

the satisfaction or the dissatisfaction with the current interaction. Absence/presence of a joystick input between topological nodes indicate satisfaction/dissatisfaction.

- Status ($St$): The status defines how capable the wheelchair user is in diving the wheelchair. This input helps to define the confidence to be placed upon the user's joystick input. A competent user for instance is capable of providing a stable joystick input. The set of status is defined by $St = \{Competent, Struggling, Reliant\}$. This can be obtained be asking the user to perform some calibration tests to measure her/his ability to give the correct joystick input. In this scenario, the joystick signal could be analysed to observe the status.
- Actions, $A$: The actions that the system can perform. In this case, these are the navigation actions defined by the set $A = \{North, South, East, West, Stop\}$.
- Observations, $O$: The observation set consists of the $location$ and the $intention$. The $location$ is obtained from the localisation system, while the $intention$ is observed using the signal from the $joystick$ when the wheelchair is sufficiently close to a node.
- Transition, $T$: This represents the dynamics of the system. In this example, the transition encodes information about the topological map states and the preferred routes of the user. Data collected during a training period can be processed to automatically generate $T$.
- Rewards, $R$: The reward function need to be constructed so as to penalise the system with a negative reward for actions that result in a dissatisfied user, and reward the system for driving the user correctly to destinations through the preferred route.

### B. Intelligent Walker

This example models a therapeutic training task that allows patients to train by themselves with minimal supervision. The patient start from a sitting down position at a certain location and then uses the walker to stand up. Once in a standing position, the patient then moves around in the designated area before returning back to the start location to sit down. Fig. 6 shows the instrumented walker developed for this purpose. It is equipped with two infra-red proximity sensors to determine if the user is standing or sitting; two strain gauges mounted on the handles to measure the amount of stress the user is exerting on the handles to determine the intended direction of the user; and a laser sensor mounted on the front of the walker is used for localisation and obstacle avoidance. The state diagram for the walker in a therapeutic training task is depicted in Fig. 7. The model variables for the walker platform example are summarised in Table II.

## IV. EXPERIMENTAL RESULTS

The proposed framework was experimentally evaluated using the robotic wheelchair described in [10] and illustrated in Fig. 4. The transition model was learned using simulated training data and was then solved to obtain an optimal policy, this policy was then tested in a navigation experiment. The learned model embeds knowledge about the most visited destinations and the preferred routes. The wheelchair system can predict the most probable destination the user is trying to reach at any node. If the user gives no joystick input at a node, then the system selects the most probable action from the policy based on the current belief and initiates the navigation. If the user is not happy with direction of travel, then she/he can give a joystick signal different from the direction of travel to indicate a dissatisfaction with the ongoing action, the system will then re-evaluate the belief and select a new action.

In this navigation experiment the user was attempting to navigate from the topological node 30, to destination $d3$ representing node 26 in the map. The user starts by giving a down joystick input, the belief is updated and a the wheelchair start going to node 29, the user doesn't give an input at this stage, so the wheelchair continues by selecting the next best action which in this case drive the wheelchair to node 31. The user realises that the wheelchair is not going to where she/he desired, so expresses dissatisfaction by giving a joystick input during this motion. This input

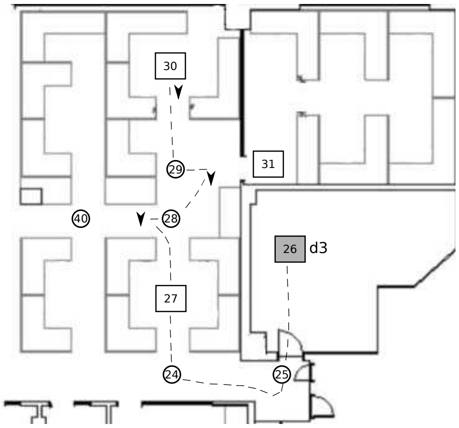

Fig. 8. Path traversed during a navigation experiment. Arrows represent joystick observations, dashed line represents the traversed path.

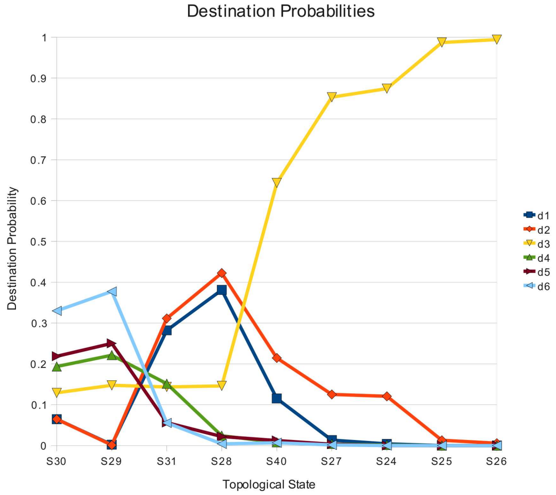

Fig. 9. Change of destination probabilities over time. $Sxx$ represent the topological states that the wheelchair had to pass through.

triggers a belief update that produces a new action directing the wheelchair to node 28. At this node, the user doesn't give an input again, so the wheelchair again selects the best action to comply with the policy that has already been generated and starts navigating to node 40. When the user quickly realises that the action being performed is incorrect and expresses his dissatisfaction by giving a joystick input indicating the correct direction of travel. The wheelchair changes direction and moves to node 27 then correctly chooses the rest of action to navigate to destination $d3$. Fig. 8 shows the path traversed during this experiment, where arrows indicate a joystick input. Fig. 9 illustrates the progression of the destination belief at the topological nodes visited while navigating, while Fig. 10 shows the values of the state variables at each node during the navigation (only the most probable state is shown).

## V. Summary

This paper presented a natural interaction gulf/layer modelled using POMDP. The proposed framework also highlighted the importance of predicting the user's intention and getting feedback from the user in terms of "satisfaction" to better enhance the quality of the interaction. The HRI variables used in the interaction layer were carefully selected to represent the natural human action cycle. This proposed strategy was successful in modelling a wheelchair assistive navigation system as to allow the wheelchair user to have maximum control over the wheelchair with a minimum amount of input, allowing at any time a change of plan or corrective actions to be communicated.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Destinations | d1 | | | | | | | | | |
| | d2 | | | ■ | ■ | | | | | |
| | d3 | | | | | ■ | ■ | ■ | ■ | ■ |
| | d4 | | | | | | | | | |
| | d5 | | | | | | | | | |
| | d6 | ■ | ■ | | | | | | | |
| Joystick | Up | | | | | | | | | |
| | Down | | ■ | | ■ | | | | | |
| | Right | | | | | | | | | |
| | Left | ■ | | ■ | | ■ | ■ | ■ | ■ | |
| | Nothing | | | | | | | | | ■ |
| Time | Struggling | | | | | | | | | |
| | Competent | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| | Reliant | | | | | | | | | |
| Action | North | | | | | | | | ■ | |
| | South | ■ | | ■ | | ■ | ■ | | | |
| | East | | ■ | | | | | ■ | | |
| | West | | | | ■ | | | | | |
| | Stop | | | | | | | | | ■ |
| Sat | Satisfied | ■ | ■ | | ■ | | ■ | ■ | ■ | ■ |
| | Unsatisfied | | | ■ | | ■ | | | | |
| | Location | S30 | S29 | S31 | S28 | S40 | S27 | S24 | S25 | S26 |

Fig. 10. State of the system variables at each topological state. $Sat$ refers to the satisfaction variable, locations denoted by $Sxx$ represent the topological states.

## References

[1] D. Norman. *The design of everyday things*. New York: Doubleday, 1990.

[2] Michael Bratman and Ernest Sosa. *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge UniversityPress, 1999.

[3] M. Bratman. *Intention, plans, and practical reason*. Harvard University Press, 1999.

[4] J. Hoey, A. V. Bertoldi, P. Poupart, and A. Mihailidis. Assisting persons with dementia during handwashing using a partially observable markov decision process. In *Proceedings of the International Conference on Vision Systems (ICVS)*, Biefeld, Germany, 2007.

[5] R. I. Bahar, E. A. Frohm, C. M. Gaona, G. D. Hachtel, E. Macii, A. Pardo, and F. Somenzi. Algebric decision diagrams and their applications. *Formal methods in system design*, 10(2):171–206, 1997.

[6] M. E. Bratman. *Intention, Plans, and Practical Reason*. Cambridge: Harvard University Press, 1990. Re-issued 1999.

[7] S. Ossowski. *Co-ordination in Artificial Agent Societies*. Springer Berlin / Heidelberg, 1999.

[8] J. Hoey, R. St-Aubin, A. Hu, and C. Boutilier. Spudd: Stochastic planning using decision diagrams. In *Proc. of the Conference on Uncertainty in Artificial Intelligence*, pages 279–288, 1999.

[9] M. T. J Spaan and N. Vlassis. Perseus: Randomized point-based value iteration for pomdps. *Journal of Artificial Intelligence Research*, 24:195–220, 2005.

[10] T. Taha, J. V. Miro, and G. Dissanayake. Pomdp-based long-term user intention prediction for wheelchair navigation. In *Proc. of IEEE International Conference on Robotics and Automation*, pages 3920–3925, 2008.