

Intelligent Assistive System Using Real-Time Action Recognition for Stroke Survivors

Jean-Baptiste, Emilie; Nabiei, Roozbeh; Parekh, Manish; Rotshtein, Pia; Russell, Martin; Jancovic, Peter; Fringi, Evangelia; Drozdowska, B; Baber, Chris

DOI:
[10.1109/ICHI.2014.13](https://doi.org/10.1109/ICHI.2014.13)

License:
Other (please specify with Rights Statement)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Jean-Baptiste, E, Nabiei, R, Parekh, M, Rotshtein, P, Russell, M, Jancovic, P, Fringi, E, Drozdowska, B & Baber, C 2014, Intelligent Assistive System Using Real-Time Action Recognition for Stroke Survivors. in *2014 IEEE International Conference on Healthcare Informatics (ICHI)*. Institute of Electrical and Electronics Engineers (IEEE), pp. 39-44, IEEE International Conference on Healthcare Informatics 2014 (ICHI 2014), Verona, Italy, 15/09/14. <https://doi.org/10.1109/ICHI.2014.13>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

(c) 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

Checked June 2015

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

UNIVERSITY OF BIRMINGHAM

Research at Birmingham

Intelligent Assistive System Using Real-Time Action Recognition for Stroke Survivors

Jean-Baptiste, Emilie; Nabiei, Roozbeh; Fringi, Evangelia; Parekh, Manish; Drozdowska, Bogna; Baber, Christopher; Jancovic, Peter; Rotshtein, Pia; Russell, Martin

Citation for published version (Harvard):

Jean-Baptiste, E, Nabiei, R, Fringi, E, Parekh, M, Drozdowska, B, Baber, C, Jancovic, P, Rotshtein, P & Russell, M 2014, 'Intelligent Assistive System Using Real-Time Action Recognition for Stroke Survivors' Paper presented at IEEE International Conference on Healthcare Informatics 2014 (ICHI 2014), Verona, Italy, 15/09/14 - 17/09/14, .

[Link to publication on Research at Birmingham portal](#)

General rights

When referring to this publication, please cite the published version. Copyright and associated moral rights for publications accessible in the public portal are retained by the authors and/or other copyright owners. It is a condition of accessing this publication that users abide by the legal requirements associated with these rights.

- You may freely distribute the URL that is used to identify this publication.
- Users may download and print one copy of the publication from the public portal for the purpose of private study or non-commercial research.
- If a Creative Commons licence is associated with this publication, please consult the terms and conditions cited therein.
- Unless otherwise stated, you may not further distribute the material nor use it for the purposes of commercial gain.

Take down policy

If you believe that this document infringes copyright please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Intelligent Assistive System Using Real-Time Action Recognition for Stroke Survivors

Emilie M. D. Jean-Baptiste*, Roozbeh Nabiei*, Manish Parekh*, Evangelia Fringi†, Bogna Drozdowska†, Chris Baber*, Peter Jancovic*, Pia Rotshein†, and Martin Russell*

*School of Electronic Electrical and Computer Engineering

†School of Psychology

University of Birmingham, Birmingham, UK,

Email: [emj198, rxn946, mxp534,b.a.drozdowska, c.baber,p.jancovic,p.rotshein,m.j.russell]@bham.ac.uk

Abstract—CogWatch is an EU project developing technologies for cognitive rehabilitation of stroke patients. The CogWatch prototype is an automatic system to re-train patients with Apraxia or Action Disorganization Syndrome (AADS) to complete activities of daily living (ADLs). This paper describes the approach to automatic planning based on a Markov Decision Process, and real-time action recognition (AR) based on instrumented objects using Hidden Markov Models. The experimental results demonstrate the ability of a psychologically plausible planning system integrated in a Task Model (TM) to improve task performance via user simulation, and the viability of the approach to AR.

I. INTRODUCTION

In the UK alone it is estimated that over 150,000 people have a stroke each year [16]. It is estimated that 68% of stroke survivors suffer from Apraxia or Action Disorganization Syndrome (AADS), which leads to the impairments of cognitive abilities to complete activities of daily living (ADLs) [26], [27], [28]. AADS can have a serious impact on patients' ability to live independent lives in their own homes. For example, when making a cup of tea or preparing a snack, patients might perform a wrong sequence of actions, skip steps, or misuse objects with possible safety implications. Caregivers can provide assistance, but patients who aspire to independent living may be unwilling to accept this as a long-term solution. Hence, the objective of the CogWatch project [17], [5] is to develop an advanced and intelligent assistive system to re-train patients on how to carry out ADLs. To achieve this the system must be able to monitor the patient's progress through the ADL and provide appropriate guiding cues or feedback when an error is detected. Thus recognition of the individual actions that make up the ADL, and planning of the optimal strategy the patient should follow during the task are critical.

Several assistive systems have been designed to increase independent ADLs completion of cognitively impaired patients. These can range from alarms that remind people of their task, to interactive ubiquitous computing systems that provide step-by-step guidance on how to perform the activity. An example is the electronic system developed by Levinson [4] that uses artificial intelligence (AI) to automatically generate the best plan to complete the required steps during the task. Boger et al. and Kautz et al. developed an AI based approach to determine when and how to deliver cues to cognitively impaired patients during everyday tasks such as cooking [29] and handwashing [19]. The Autominder System developed by Pollack et al. [25] uses dynamic Bayesian networks as an

underlying domain model to coordinate prompts delivering through activities. Capturing ADL through sensors is not new [1], [2], [14], [12], [7], [8], [13], [3], and when combined with AI, Barucha showed that sensing capabilities improve the reability and viability of assistive systems [20]. Nevertheless, there has been less attention devoted to the definition and recognition of subtasks in these activities. Moreover, despite the growing interest in incorporating smart techniques, most of the devices required manual input from the user and were not able to sense the context in which they were being used or the users' preferences. Many devices were not designed to indicate when the task is completed, or when an error from the user has occurred, what type was this error, and if assistance was needed. The implication of involving clinical specialists in the system development have also received little attention in the studies. Thus, we aim to demonstrate that a context-aware planning system (referred as Task Model (TM)) integrating AI technics, which takes into account clinical specialists' knowledge and users' preferences makes an assistive system more efficient, when combined with a complex action recognition system (AR). Independently we provided evidences that those two components are valid: the AR was tested with data from healthy and patients, and the TM using a simulated user based on performances of healthy and patients. We are currently collecting data on the efficacy of the real system.

II. SYSTEM ARCHITECTURE OVERVIEW

CogWatch's current goal is to re-train stroke survivors providing them autonomous guidance during tea making. The complete system (figure 1) comprises sensorized objects (mug, kettle, jug), an AR, a Markov Decision Process based TM and a Prompting System. It works as follows: First, the patient chooses the type of tea out of four options (e.g. black tea with sugar). This information is passed to the TM that uses it to provide correct strategies. Patient's behaviour is detected via sensors attached to the objects used during the task. This data is communicated wirelessly to the AR which aim is to recognise what action the patient has performed. The AR outputs are passed to the TM, which is in charge of the planning and monitoring of the patient's progress through the task. In other words, each time the patient makes an action, the AR outputs an observation, and the TM records it in order to determine the patient's state (i.e. its understanding of what the patient has achieved so far). The state s is passed to the Action Policy module that plans what should be done next (what "optimal strategy" a should be suggested) in order to assist

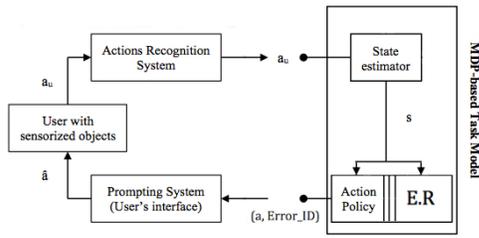


Fig. 1. Structure of CogWatch: a_u and a denote the user’s action and the TM’s strategy, s is the user’s state, ER - the error recognition module, and the circumflex indicates an estimate. *Error – ID* is the user’s error type.



Fig. 2. A jug fitted with a CogWatch Instrumented Coaster (CIC) (left) and an ‘open’ CIC (right)

the patient. Moreover, in contrast to most previous AI planning systems, here we also implemented an Error Recognition (E.R) module. This module analyzes the state s in order to identify potential errors in the patient’s plan. Finally, the TM outputs a recommendation for the next best action, and if needed alerts the Prompting system that an error has occurred. The Prompting System uses a table designed by clinicians to map the output from the TM to the type of cue that should be retrieved. We next describe each core components in more details: the AR and the TM.

A. HMM-based Action Recognition based on Instrumented Objects

For the system to assist the patient properly, it must be able to recognize the patient’s actions (sub-goals). The AR is in charge of this matter and uses HMMs based on instrumented objects. Indeed, sensors are integrated into the kettle, mug and milk jug. To avoid patient confusion, the chosen solution is to package the sensors and circuitry into an instrumented coaster (figure 2), the ‘CogWatch Instrumented Coaster (CIC)’, that is fitted to the underside of the object. This is inspired by the MediaCup concept [3]. The sensors in the CIC are a 3-axis accelerometer and 3 force sensitive resistors (FSRs), together with a Bluetooth transmitter to send data to a host computer. The CIC was designed according to 3 criteria: (i) Its sensors should capture relevant data to identify the sub-goals of tea-making. (ii) The CIC needs to fit under an object, so that the FSRs can record the object’s weight. The size of the CIC was defined by the circumferences of the bases of the coffee mug and milk jug. Finally, (iii) The selection of sensors was based on power efficiency relative to cost of the sensors and their processing capability. When the objects are moved the sensors communicate relevant data to the HMMs based AR.

Although HMMs are a generic framework for statistical sequential pattern processing, they have been developed most intensely for speech recognition (ASR). An HMM-based ASR system has four parts: a feature extraction component that converts speech into a sequence of acoustic feature vectors

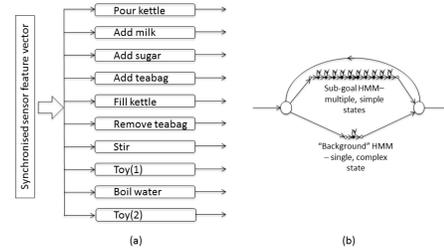


Fig. 3. Structure of the CogWatch AR: Parallel sub-goal detector set (a), and a single detector (b).

Y , a statistical grammar, typically based on n-grams, which gives the vocabulary and the probability of each vocabulary word given previous words, a pronunciation dictionary, which specifies one or more phone-level transcriptions of each vocabulary word, and a set of phone-level acoustic HMMs. Typically these are context sensitive to account for co-articulation effects. Viterbi decoding finds the sequence of words W such that an approximation to the probability $P(W|Y)$ is maximised. There are a number of important differences between AR and ASR which determine the design of our HMM-based AR system: (i) in ASR words occur sequentially, whereas in AR actions can occur in overlapping time (e.g. because the subject uses both hands). Therefore finding the most probable sequence of actions is inappropriate in AR, (ii) in ASR the same features are used by all HMMs, whereas in AR it is clear that different subsets of features are appropriate for recognising different sub-goals (e.g., sensors attached to the kettle are not directly relevant to ‘Pour Milk’). Finally, (iii) in AR there is no universally accepted equivalent to a ‘phone set’. The CogWatch AR is a parallel array of ‘sub-goal detectors’, each dedicated to a different sub-goal (figure 3(a)). The input to a detector is the sub-vector of sensor outputs that are relevant to that sub-goal. A detector consists of a multiple state sub-goal HMM and a single state ‘background’ HMM, with each HMM state associated with a Gaussian mixture model (GMM). These HMMs are configured in parallel, so that they compete to explain the input data (figure 3(b)). The data is processed separately for each parallel detector using a Viterbi decoder [10]. A detector’s output up to time t is generated as soon as its classification of the data up to t is unambiguous, using partial trace-back [11], and the memory used to store alternative explanations of the data up to t is freed. In this way the decoders can run indefinitely. The real-time CogWatch AR uses HMM file formats from the hidden Markov model toolkit (HTK) [22]. Thus HMM parameters can be optimised off-line using HTK and then transferred to the CogWatch AR.

B. MDP-based Task Model

The TM is based on a Markov Decision Process (MDP). The MDP is a mathematical tool for planning, learning and describing decision-making in probabilistic environments. It is defined as a four-tuple (S, A, P, C) [6], where: S is a finite set of states, A is a finite set of actions, P is the transition function ($P(s, a, s')$ denotes the probability of reaching state s' from s given that action a was taken), and $C(s, a)$ the cost of taking a in state s . Given an MDP, the problem is to find the optimal strategy π^* which is a mapping from states to actions, where $\pi^*(s)$ is the best action to perform in state s . Its computation

is based on the policy value $V^*(s)$, which is the expected sum of costs incurred by a session starting in state s at time $t = 0$, and following π^* , until the final state is reached at T_F . Note that angle brackets indicate expected values.

$$V^*(s) = \langle \sum_{t=0}^{T_F} c(s_t, a_t) \rangle, \quad (1)$$

where $s_0 = s$ and $a_t = \pi^*(s_t)$. It can also be expressed as:

$$V^*(s_t) = \min_a [\langle c(s_t, a) \rangle + \sum_s P(s, a, s_t) V^*(s)]. \quad (2)$$

π^* is then defined by:

$$\pi^*(s_t) = \operatorname{argmin}_a [\langle c(s_t, a) \rangle + \sum_s P(s, a, s_t) V^*(s)] \quad (3)$$

Those strategies are computed using a Monte Carlo Algorithm [9]. Prior to implement it, the MDP's theory has to take into account CogWatch's context:

1) *Action Space*: Using the principles of task analysis [15], each type of tea is decomposed into a hierarchy of sub-goals, tasks and sub-tasks. CogWatch currently focuses on the first level, where eight sub-goals have been identified, plus one common error (9), and one potentially hazardous activity (10). These are:

- 1) "Fill Kettle" (using water from a pre-filled jug),
- 2) "Boil Water",
- 3) "Pour Kettle" (i.e. pour boiling water into the mug),
- 4) "Add Teabag",
- 5) "Add Sugar",
- 6) "Add Milk",
- 7) "Stir",
- 8) "Remove Teabag",
- 9) "Pour Cold Water from Jug into Mug", and
- 10) "Toy with the Kettle".

2) *State Space*: States of the MDP-based TM directly correspond to user's states. Specifically, a state is a sequence of actions performed. For example, a state can be represent as a list: $s_k = [a_1, a_2, a_4, \dots]$, where each a_n is an action in the Action Space. In other words, the state space is a list containing all the states the user can reach while completing the task.

3) *Transition Function*: CogWatch is currently used with real participants under the supervision of a clinician. Thus, we make the assumption that the transition function $P(s, a, s')$ is binary. Suppose:

- The current state s corresponds to the sub-goal sequence a_1, a_2, \dots, a_n ,
- a is the sub-goal g , and
- s' is a_1, a_2, \dots, a_n, g

Then $P(s, a, s') = 1$ and $P(s, a, t) = 0$ for all states $t \neq s'$.

4) *Cost Function*: The cost function $C(s, a)$ is a mechanism to incorporate human judgment about the importance of different types of behaviour into the MDP. In our case, we combined two types of functions: one that allows the MDP to find the fastest strategy (i), and another one that takes into account the way participants successfully perform the task

TABLE I. DATA USED IN AR DEVELOPMENT. DURATIONS ARE IN HOURS. 'TOY(1)' AND 'TOY(2)' CORRESPOND TO THE PATIENT 'TOYING' WITH THE WATER JUG AND KETTLE, RESPECTIVELY.

Sub-goal	Trials	Dur.	Sub-goal	Trials	Dur.
Pour kettle	138	6.43	Stir	138	5.55
Add milk	123	4.06	Toy(1)	26	0.73
Add sugar	120	3.42	Boil water	125	2.21
Add teabag	144	3.11	Toy(2)	30	1.14
Fill kettle	146	7.12			
Rem. teabag	134	6.33	Full trial	99	25.11

and clinicians' preferences (ii). This decision has been taken because the fastest strategy may be valid, but is not necessarily psychologically plausible. Later we will demonstrate that when the cost function (i) is combined with relevant knowledge from users and clinicians (i.e., cost function (ii)), it allows the TM to generate more meaningful and efficient strategies during the task. When using the cost function (i) only, the TM's strategies will be referred as Non-Psychologically Plausible; when using the combination (i) and (ii), the TM's strategies will be referred as Psychologically Plausible.

III. EXPERIMENTS AND RESULTS

A. AR Evaluation

Twenty-six participants, aged between 18 and 80, completed multiple individual sub-goals and multiple full tea-making trials. In all cases synchronized sensor outputs were recorded. The recordings are summarised in table I. In total there are 1124 recordings of isolated actions (40.09 hours) and 134 recordings of complete tea-making (25.11 hours). Full trials took place under five different conditions. Subjects were asked to: (i) Make a cup of tea as they would normally make it for themselves, (ii) make a cup of tea in a different way, and follow cues presented on a screen, which gave step-by-step instructions on how to make the tea, where (iii) the cues followed the participant's normal way of making tea, (iv) the cues differed from the participant's routine way of making tea, and (v) the cues were in random order and did not lead to successful tea-making. HMM optimisation was performed off-line with HTK. The optimised parameters are the number of states N in the sub-goal HMMs ($N = 5, 10, 20, 30, 40, 50, 60$), with each state associated with a single Gaussian PDF, and the number of GMM components M in the single state background' HMM ($M = 1, 2, 8, 32, 64, 128, 256, 512$). Results are presented for the two sub-goals 'Pour Kettle' and 'Add Milk'. Five-fold cross-validation was used for parameter estimation, evaluation and testing. For each pair of values N and M , 60% of the 'target' sub-goal data and 60% of the other sub-goal data, were used to estimate the parameters of the sub-goal and background HMMs, respectively. For initialisation of the sub-goal model, each recording of the sub-goal in the training set was divided into N equal segments, and the data in the n th segments was used to estimate the mean and diagonal covariance matrix of the n th HMM state. The HMM parameters were then optimised using the standard Baum-Welch algorithm [22], [10]. For the background model, all of the recordings of the non-target sub-goal in the training set were used to estimate the mean and (diagonal) covariance matrix of a single Gaussian PDF. This was repeatedly divided and then optimised using the E-M algorithm and standard tools in HTK [22]. An evaluation set, comprising 20% of the 'target'

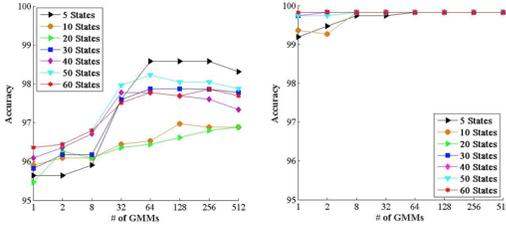


Fig. 4. Performance of detectors ‘Pour kettle’ (left) and ‘Add milk’ (right). Horizontal axes are number of ‘background’ model GMM components. Different graphs correspond to different numbers of states in the ‘sub-goal’ HMM.

sub-goal data and 20% of the recordings of other sub-goals that were not in the training set, was used to test each sub-goal detector. The results, averaged over all 5 folds, are shown in figure 4. The ‘sub-goal accuracy’ A is defined by

$$A = 1 - (D + I)/K \cdot 100 \quad (4)$$

where K , D and I are the number of test examples of the sub-goal, the number of times that the sub-goal is not detected (the number of deletions), and the number of times that the sub-goal is falsely detected (the number of insertions), respectively. For each fold, the optimal numbers of sub-goal HMM states and background model GMM components were chosen based on these results. Next, a test set comprising the remaining 20% of the ‘target’ sub-goal data and 20% of the other sub-goal data, was used to test the optimal configurations of each of the sub-goal detectors. The accuracy, averaged over the 5 folds, is 98.13% for ‘Pour Kettle’ and 99.82% for ‘Add Milk’. Finally, both detectors were tested together on the 99 full trials, using HMM parameters chosen according to figure 4 and trained using all of the ‘isolated sub-goal’ recordings. This resulted in an overall accuracy of 93.83%, with 5 state sub-goal models and 128 component background models. Because the AR comprises separate detectors for each sub-goal, and the final test is on complete trials, this results is not affected by the addition of further sub-goal detectors.

B. TM Evaluation

The initial evaluation of the Task Model consists of measuring its ability to suggest valid strategies only. This was done because we were interested in the successful completion of a single trial. So here we did not report results on the efficacy of the integration of the E.R module, as it was assumed that its impact would only be seen across trials. To reliably evaluate the TM’s action policy, a large number of repetitive interactions between participants and the system are necessary. To achieve this, we created a simulated impaired user *SimU* based on data from fifty-two control and cognitively impaired participants, aged between 21 and 82, who completed four types of tea (black tea, black tea with sugar, white tea, white tea with sugar) 100 times. Integrating this data into its mechanism, the *SimU* generates plausible sequences of actions while interacting with a virtualization of CogWatch. The experiments we ran showed the validity of the TM’s strategies, and that psychologically plausible strategies are more effective than non-psychologically plausible valid strategies.

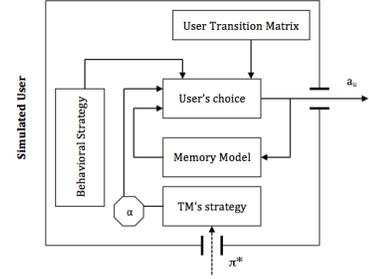


Fig. 5. Structure of the Simulated User.

1) *Simulated User*: Figure 5 shows the architecture of the *SimU*. The core of the *SimU* is the module *User’s choice*. It takes as inputs four parameters: *User’s transition Matrix*, *Behavioural strategy*, *Memory model*, and *TM’s strategy*. The latter collects the optimal strategies π^* sent by the TM during the task, which the *SimU* complies with a probability α . The *User’s transition Matrix* is based on action bigram probabilities calculated from examples of action sequences performed by real participants. In an attempt to make this mechanism plausible and compensate the limitations induced by the use of bigram probabilities, the *SimU* has different *behavioural strategies*, which permit it to overcome its potential lack of knowledge of what action to output next. The *Memory model* module gives the possibility to decide how the *SimU* remembers or forget what it performed in the past. Taking into account those four parameters, the *User’s choice* then chooses which action a_u to output.

2) *Virtualization of CogWatch*: The *SimU* is integrated into a virtualization of CogWatch, which follows the same structure of the real system as described in figure 1. The *SimU* replaces the module called *User with sensorized objects* and is directly connected to a virtual AR and Prompting System, the MDP-based TM remains the same. The virtual AR is implemented as a simple $N \times N$ confusion matrix C (N is the number of sub-goals) whose i, j^{th} entry is the probability that the AR system outputs sub-goal j when the user executes sub-goal i . As the AR output is verified by a clinician in the current system, we assume that the virtual AR is 100% accurate, so that C is the identity matrix.

3) *Experiments*: During the first experiment, the *SimU* tried to make each type of tea 3000 times. Table II shows its success rate when executing the strategies suggested by the TM and when ignoring it. For black tea and black tea with sugar, the *SimU*’s success rate is 100% when it complies with the TM’s strategies. This means that the TM’s strategies are 100% accurate for those two tasks. When the *SimU* ignores the TM and performs the tasks following its own plan, its success rate significantly decreases: 76% for black tea, 54.1% for black tea with sugar. In the case of white tea and white tea with sugar, the assistance of the TM’s strategies also permits to increase the *SimU*’s success rate, but the latter is no longer 100%. This is not due to the TM’s strategies, but linked to the fact that CogWatch is an after-effect system, where an action has to be made by the user for the TM to plan what to should be done next. If the first action is an error that cannot be corrected, the system cannot help the user. Apart from those cases, the results highlight that the TM’s psychologically plausible strategies always permit a compliant *SimU* to increase

TABLE II. SIMULATED USER’S SUCCESS RATES WHEN INTERACTING WITH THE COGWATCH SIMULATOR. PSYCHOLOGICALLY PLAUSIBLE STRATEGIES RETRIEVED BY THE TM.

Types of tea	Assisted by TM	Ignoring TM
Black Tea	100.0%	76.0%
Black Tea With Sugar	100.0%	54.1%
White Tea	97.1%	48.0%
White Tea With Sugar	96.0%	24.2%

its ability to succeed each task.

In the second experiment, we compare the *SimU*’s success rate at varying compliance when the TM outputs Psychologically Plausible (P.P) strategies, or Non-Psychologically Plausible (N.P-P) ones. As explained in section II-4, this is due to the cost function used in the MDP. In figure 6, we see that when the *SimU* is 100% compliant to the TM’s strategies, whether the latter outputs a N.P-P or P.P strategy has no impact on the *SimU*’s performance. This is an indicator that both N.P-P and P.P strategies are valid. Nevertheless, as soon as the *SimU* decreases its compliance to the TM’s outputs, we can see (figures 6(a-b-d-e)) that its success rate is higher when the strategies are psychologically plausible. In (figures 6(a-b)), when the *SimU* follows a N.P-P strategy with a compliance at 20% during the task, its success rate is the same as if it was trying to perform the task by itself (0% compliance). We can then conclude that if both strategies are valid, the P.P one is optimal compared to N-P.P. To make a parallel with a realistic situation, the P.P strategy can be seen as a familiar one; a strategy able to take into account the ways a clinician would perform the tasks or the optimal ways the patients are used to perform when they succeed. So, with P.P strategies, when the user completes the task and accepts to comply, the TM succeeds to redirect the user on the most efficient ways of succeeding the task. On the other hand, even if a N-P.P strategy is always correct, it does not take into account the patient’s habits, which then leads to more users’ failures. Indeed, in [23], [24] De Klein and Graybiel highlighted the impact of familiar and unfamiliar sequences on success rate. We learn that familiar sequences are easier to execute and requires less effort and energy, as this is done through a sub-cortical structure where the sequence is reduced to a single unit. In contrast for novel sequences or sequences that diverge from the familiar one, one needs to use cortical mechanisms (more effort, higher demands on resources) to re-create them.

IV. DISCUSSION

Our experiments have shown that the MDP-based Task Model integrated in a simulation of the current CogWatch System permits to correctly assist a virtually impaired simulated user during each task. We believe that similar results will be observed when experiments being run with real participants are completed. From an architectural and computational point of view, to have implemented this virtual simulation of CogWatch allowed us to validate the Task Model’s capability to fulfil the requirements needed for the system to be a context-aware intelligent assistive device. In the future, the TM will be extended to other types of tasks, such as teeth brushing. The flexibility of its current structure already allows such extension, but specific errors definitions will have to be defined and integrated. Another issue that will be tackled is the granularity of the actions the TM analyses. Indeed, we would like to

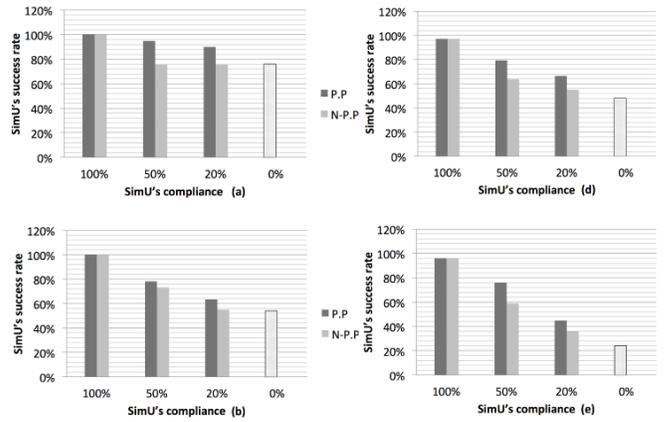


Fig. 6. *SimU*’s success ratio at varying compliance to the Task Model’s strategy. P.P and N-P.P stand for Psychologically Plausible and Non-Psychologically Plausible strategy; a, b, d and e correspond to black tea, black tea with sugar, white tea, white tea with sugar.

deal with actions that are at a lower level of hierarchy. For example, we would like the TM to understand if the patient fails adding an ingredient it is because an object is hold with a wrong grip, or because he/she fails other complex dexterous movements. As far as the AR is concerned, we showed that the error rates for unseen isolated sub-goals (1.87% and 0.18% for ‘Pour Kettle’ and ‘Add Milk’) are extremely low and validate combining instrumented objects with HMM-based action recognition. The error rate for complete trials, 6.17% overall error rate for two detectors, is higher. However, it represents insertion or deletion of just 7 instances of ‘Pour Kettle’ or ‘Add Milk’ in 2.5 hours of trials and thus represents extremely good performance. Since the real-time CogWatch AR gives identical results to HTK, this is also the performance of the real-time system. In addition, the CogWatch AR with two detectors runs in real-time on live data. Beyond these examples, numerous challenges remain such as the amount of human supervision needed by the AR. Currently, when used with real participants, the AR is supervised by a clinician, so its outputs always perfectly correspond to the user’s actions. As our goal is to have as little human supervision as possible, in the next prototypes we will have to take into account the uncertainties associated to the AR. Indeed, the AR may misrecognize some actions performed by the user during the task, which means that the state the TM believes the user to be in might not be true all the time. Such uncertainties are inevitable, and a MDP-based Task Model cannot cope with them well. Thus, a solution will be to replace the MDP with a Partially Observable MDP (POMDP) [18] which can accommodate uncertainties in the state space. So, when the AR will output an observation, instead of choosing a strategy based on the most probable state, a POMDP-based TM will base it on a probabilistic distribution over all states. In the context of a fully automatic system, we believe that such enhancements will permit obtaining a more accurate representation of the environment, and a robust guidance under uncertainties.

ACKNOWLEDGMENT

This work has been supported by the European Commission under the grant FP7-ICT-2011-288912.

REFERENCES

- [1] Amft, O. and Tröster, G. "Recognition of dietary activity events using on-body sensors", *Artificial Intelligence in Medicine*, 42, 121-136, 2008.
- [2] Hasan, M. K., Rubaiyat, H. A., Lee, Y. K., and Lee, S. "A reconfigurable HMM for activity recognition". In *Advanced Communication Technology 20 ICAC 2008. 10th IEEE Int. Conf on* , Vol. 1, pp. 843-846, 2008.
- [3] Gellersen, H-W., Beigl, M. and Krull, H. "The MediaCup: awareness technology embedded in an everyday object", In H-W. Gellersen (ed.) *Handheld and Ubiquitous Computing 1st International Symposium HUC99*, Berlin: Springer, 308-310, 1999.
- [4] Levinson, R., "PEAT : The planning and execution assistant and training system", *Journal of Head Trauma Rehabilitation*, 12, 2, 769-775, 1997.
- [5] Hermsdörfer, J., Bienkiewicz, M., Cogollor, J.M., Russell, M., Jean-Baptiste, E., Parekh, M., Wing, A.M., Ferre, M., Hugues, C., "CogWatch Automated Assistance and Rehabilitation of Stroke-induced Action Disorders in the Home Environment", *HCI Aspects of Optimal Healing Environments*, 2013.
- [6] Puterman, M., "Markov Decision Processes: Discrete Stochastic Dynamic Programming". In: John Wiley & Sons, Inc., 1994.
- [7] Junker, H., Lukowicz, P. and Troster, G., "Continuous recognition of arm activities with body-worn inertial sensors", In *Proc. 8th IEEE International Symposium on Wearable Computers*, Vol. 1, 188-189, 2004.
- [8] Maekawa, T., and Watanabe, S., "Unsupervised Activity Recognition with User's Physical Characteristics Data", In *Proc. 15th IEEE International Symposium on Wearable Computers* , 89-96, 2011.
- [9] Levin E., Pieraccini R., Eckert W., "A stochastic model of human machine interaction for learning dialogue strategies." *IEEE Transactions on Speech and Audio Processing*: Vol.8, No. 1., January, 11-23, 2000.
- [10] Rabiner, L.R., "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. IEEE*, vol. 77, no. 2, 1989.
- [11] Spohrer, J.C., Brown, P.F., Hochschild, P.H. and Baker, J.K., "Partial traceback in continuous speech recognition", *Proc. IEEE Int. Cong. Cybernetics and Society*, 1980.
- [12] G Hong, Y.J., Kim, I.J., Ahn, S. C., and Kim, H.G., "Activity Recognition Using Wearable Sensors for Elder Care". *Second IEEE Int. Conf. on Future Generation Communication and Networking*, 302305, 2008.
- [13] Wagner, J., Ploetz, T., Halteren, A. V., Hoonhout, J., Moynihan, P., Jackson, D., Ladha, C., "Towards a Pervasive Kitchen Infrastructure for Measuring Cooking Competence". *Proc Int. Conf Pervasive Computing Technologies for Healthcare*, 107-114, 2011.
- [14] Hondori, H.M., Khademi, M. and Lopes, C.V., "Monitoring intake gestures using sensor fusion (Microsoft Kinect and inertial sensors) for smart home telerehab setting", *Proceedings of the 1st Annual Healthcare Innovation Conference*, 2012
- [15] Hughes, C.M., Baber, C., Bienkiewicz, M., Hermsdorfer, J., "Application of human error identification (HEI) techniques to cognitive rehabilitation in stroke patients with limb apraxia. Access in Human-Computer Interaction. Applications and Services for Quality of Life. Springer Berlin Heidelberg:463-471, 2013
- [16] The Stroke Association website, <http://www.stroke.org.uk/about-stroke>
- [17] CogWatch: Cognitive rehabilitation of apraxia and action disorganization syndrome, <http://www.cogwatch.eu/>
- [18] Williams, J.D., Poupart, P., Young, S., "Factored partially observable Markov Decision Processes for dialogue management". *Computer Speech and Language*: p. 393-422, 2005.
- [19] Boger, J., Hoey, J., Poupart, P., Boutillier, C., Fernie, G., Mihailidis, A., "A Planning System Based on Markov Decision Processes to Guide People With Dementia Through Activities of Daily Living". *IEEE Transactions on Information Technology in Biomedicine*:323-333, 2006
- [20] Bharucha, A., Anand, V., Forlizzi, J., Dew, M. A., Reynolds III, C.F., Stevens, S., Howard Wactlar, M.S., "Intelligent Assistive Technology Applications to Dementia Care: Current Capabilities, Limitations and Future Challenges." *American journal of Geriatric Psychiatry*:17(2), 88-104, 2009.
- [21] LoPresti, E.F., Mihailidis, A., Kirsch, N.L., "Assistive technology for cognitive rehabilitation: State of the art." *Neuropsychological Rehabilitation*: 14, 5-39, 2004.
- [22] Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V. and Wood-land, P., "The HTK Book (version 3.2)", Cambridge University Engineering Department, 2002.
- [23] De Kleine, E., Van der Lubbe, R.H., "Decreased load on general motor preparation and visual-working memory while preparing familiar as compared to unfamiliar movement sequences.", *Brain Cogn.* 2011, 75(2):126-34, 2011.
- [24] Graybiel, A.M., "The basal ganglia and chunking of action repertoires." *Neurobiol Learn Mem.* Jul-Sep;70(1-2):119-36, 1998.
- [25] Pollack, M.A, Brown, L., Colbry, D., McCarthy, C.E., Orosz, C., Peintner, B., Ramakrishnan, S., and Tsamardinos, I., "Autominder: An Intelligent Cognitive Orthotic System for People with Memory Impairment." *Robotics and Autonomous System (RAS)*, 44(3-4): 273-282, 2003.
- [26] Zwinkels, A., Geusgens, C., van de Sande, P., van Heugten, C., "Assessment of apraxia: inter-rater reliability of a new apraxia test, association between apraxia and other cognitive deficits and prevalence of apraxia in a rehabilitation setting." *Clinical Rehabilitation* 18: 819-827, 2004.
- [27] Barbieri, C., De Renzi, E., "The executive and ideational components of apraxia." *Cortex.* 24:535-544, 1988.
- [28] Bickerton, W., Riddoch, M.J., Samson, D., Balani, A., Mistry, B., Humphreys, G.W., "Systematic assessment of apraxia and functional predictions from the Birmingham Cognitive Screen." *Journal of Neurology, Neurosurgery, and Psychiatry*, 83, 513-521, 2012.
- [29] Kautz, H., Etzioni, O., Fox, D., Wel, D., "Foundations of Assisted Cognition Systems", 2003.