SIMULATION OF A HUMAN MACHINE INTERACTION: LOCATE OBJECTS USING A CONTEXTUAL ASSISTANT

Chikhaoui Belkacem
Pigot Hélène
Domus Laboratory, Computer Science Department, Faculty of Science,
University of Sherbrooke, Sherbrooke (QC), J1K 2R1 Canada
email: {belkacem.chikhaoui, helene.pigot}@usherbrooke.ca

ABSTRACT

The standard development of human machine interfaces needs the respect of ergonomic norms and rigorous approaches, which constitutes a major concern for computer system designers. The increased need on easily accessible and usable interfaces leads researchers in this domain to create methods and models that make it possible to evaluate these interfaces in terms of utility and usability. This paper presents a study about the simulation of a human machine interaction with an interface of a contextual assistant, using the cognitive architecture ACT-R emphasizing on the time execution of tasks. The results of our model were consistent with those obtained by the Fitts Law model which is a powerful analytical method for evaluating human machine interfaces, developed in this study mainly to support our results.

INTRODUCTION

The evaluation of Human Machine Interfaces (HMI) is becoming increasingly important and constitutes an integral part in the development cycle of computer systems. While the development of interfaces presents some challenges, their evaluation needs rigorous methods to ensure they fulfill the initial specifications and the quality of accessibility, usability and usefulness (Nielsen and Phillips, 1993; Eugenio et al., 2003). Two main approaches for evaluation are currently used, empirical approaches and analytical approaches. Empirical approaches are essentially based on performances or opinions of users gathered in laboratories or other experimental situations. These approaches are user-focused. Unlike the empirical approaches, analytical approaches are not based directly on the user performance, but rather, on the automated examination of interfaces using well-defined structures and rigorous analysis techniques.
The HMI should be resumed by the actions of pushing buttons displayed on a screen. According to this approach the Fitts law estimates the time needed to reach the targets displayed on the interface. Nevertheless, the HMI implies three human components, which must be taken in account. The first component is perceptual. In our case the human perceives the signal in a visual manner. The second one is cognitive. Here the human retrieves in his memory the object required and reasons to satisfy specific goals. The third one is motor and necessitates pressing on the selected button.
In this study, we aim to evaluate the interaction with an interface of a contextual assistant developed for cognitively impaired people. The aim of this application is to assist people while preparing meals in their kitchen by using cognitive assistance (Pigot et al., 2005). Due to the related population and the kind of errors they commit we need to take in account the cognitive part involved in the HMI. We then use a powerful analytical method based on cognitive models, emphasizing the cognitive analysis of the tasks and the time execution. We choose to base our analytical method on the cognitive architecture ACT-R (Anderson et al., 2004). Thanks to ACT-R the interaction is decomposed in rules simulating the cognitive behavior of a human using the contextual assistant. We first present an overview of the cognitive architecture ACT-R and of the contextual assistant. Once the task simulated is defined, the model, we developed, is introduced and the results of the simulation are compared to the time estimated by the Fitts law to interact with the contextual assistant.

BACKGROUND

In this section we present an overview of the cognitive architecture ACT-R, and then we introduce the contextual assistant application and the interface to be modeled.

Cognitive architecture ACT-R

The cognitive architecture ACT-R is built to simulate and understand human cognition (Anderson et al., 2004, 2005). It consists of a set of modules integrated through a central production system. ACT-R is an hybrid architecture that combines two subsystems: symbolic system including semantic and procedural knowledge, and subsymbolic system evaluating knowledge activations. The subsymbolic system assigns activations to chunks (semantic knowledge) and rules (procedural knowledge).

The activation level is one of the criteria to choose the more predominant knowledge available at a specific time. In ACT-R the perceptual and motor modules are used to simulate interfaces between the cognitive modules and the real world (Byrne, 2001; Bothell, 2004).

Visual and Motor Modules of ACT-R

The visual module that is part of the perceptual modules, has two subsystems, the positional system (where) and the identification system (what) that work together in order to send the specified chunk to the visual module. The positional system is used to find objects. When a new object is detected, the chunk representing the location of that object is placed in the visual-location buffer according to some constraints provided by the production rule (Bothell, 2004). The identification system is used to attend to locations which have been found by the positional system. The chunk representing a visual location will cause the identification system to shift visual attention to that location. The result of an attention operation is a chunk, which will be placed in the visual buffer (Byrne, 2001; Bothell, 2004). The motor module contains only one buffer through which it accepts requests (Bothell, 2004). Two actions are available in ACT-R, to click with the mouse or press a key on the keyboard.

Contextual Assistant

The Contextual assistant application is developed to assist persons with cognitive disabilities (Pigot et al., 2007a; Lussier-Desrochers et al., 2007). The aim is to foster autonomy in the daily living tasks and particularly during complex cooking tasks such as preparing pancakes, or spaghetti (Pigot et al., 2007b). The cooking task is decomposed of steps displayed on a touch screen. The two first steps consist of gathering the utensils and ingredients necessary to the recipe (Figure 1). The other steps explicit the recipe using photo and video on the screen as well as information dispatched all around the kitchen. The contextual assistant is specifically designed to help people remembering the places where the objects are stored. To do so, the contextual assistant contains an interface called the locate application displaying the objects to search. When an object is pushed in the main interface, the contextual assistant looks for the location of that object in the environment using techniques of pervasive computing and indicates the location by highlighting the appropriate locker containing that object as shown in Figure 2. In this study we simulate the first two steps of the spaghetti recipe. They consist of first knowing the list of objects to gather, either utensils or ingredients, and then to use the locate application in order to find each object.
The contextual assistant interface is displayed on a 1725L 17" LCD Touchscreen, with 13.3" (338 mm) horizontal and 10.6" (270 mm) vertical useful screen area. It is configured to 1024 x 768 optimal native resolution running Macintosh. The screen is fixed under a closet nearby the oven in order to be easily accessible and also protected against the cooking splashes.

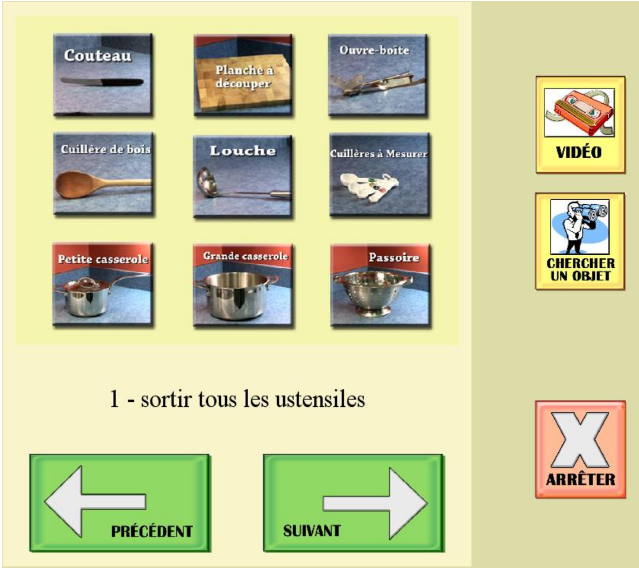

Figure 1: Main interface of the contextual assistant

Figure 2: Locker state when an object is pushed

MODELING THE INTERACTION WITH THE CONTEXTUAL ASSISTANT USING ACT-R

In this section, we present the modeling process of the tasks involved in our study, which are gathering utensils and gathering ingredients, emphasizing on the perceptual and cognitive parts, using the perceptual motor modules of ACT-R.

Task analysis: gathering utensils and ingredients

We model the first two steps of the recipe, gathering utensils and gathering ingredients. The interactions with the touch screen are simulated without taking in account the time taken by the subject to pick up the objects in the environment. The two first steps require three subtasks (Figure 3). The first subtask consists of activating the locate application in order to locate each object required by the recipe. This is done by pushing the button "LOOK-FOR-OBJECT" (in French, "CHERCHER-UN-OBJET"), which is displayed on the main interface of the contextual assistant (Figure 1). The second subtask is to locate each object, either utensils or ingredients, needed in the current step by pushing the button corresponding to the object in the locate application. The third task consists of coming back to the main application in order to know the next step of the recipe. The tree decomposition is presented in figure 3, where the translation in English is available to compare the tasks tree from the interface of Figure 1. The nodes in capital indicate the action to click on the named button, while the other nodes represent tasks to be decomposed.

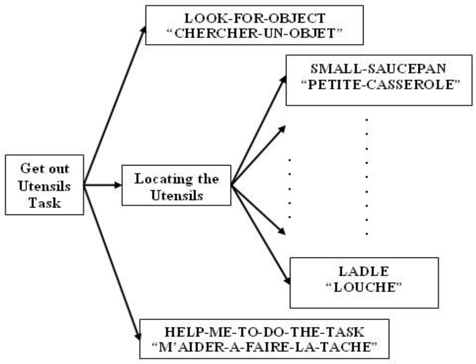

Figure 3: Tree representing the gathering utensils task

Gathering ingredients and utensils model

The model developed aims to simulate the HMI during the two first steps of the recipe. In that task, three different interfaces are involved, the interface of the locate application and the two of the contextual application displaying the utensils and ingredients needed in the recipe. The model uses ACT-R to emphasize the cognitive processes involved when looking for an object and choosing the button to push. It is decomposed of three phases, the visual phase, the recognizing phase and the motor phase. The visual phase consists of localizing the object to perceive and then identifying it. We consider that all buttons displayed on the screen are objects, either the button used to locate a utensil or ingredient, or the buttons to navigate in the interface. The first one is the button "LOOK-FOR-OBJECT" as described in Figure 3. Then, all the utensils needed in the recipe are presented in the visual interface of ACT-R. Finally, to complete the first step of the recipe, the button "HELP-ME-TO-DO-THE-TASK" is presented in order to come back to the main interface of the contextual assistant and pursue the second step of the recipe. Each object of the interface is displayed at defined coordinates (x, y) on the screen. These coordinates specify the request made to the visual-location buffer of ACT-R, which creates a chunk representing the location of the specified object. After that, the identification system identifies the name of the object and creates a chunk placed in the visual buffer. The location and identification phases last 185 ms (Bothell, 2004; Byrne, 2001). The objects are presented to the visual module of ACT-R by the mean of a list of all the objects (buttons of the interface) to be pushed on. Figure 4 shows some ACT-R productions responsible of the visual encoding phase.

```
(P start-application
  =goal> ISA     begin ; Initializing the model
==>
  -imaginal>
  +visual-location>
  ;Making request of the visual-location buffer
      ISA        visual-location
      :attended     nil
  +goal> ISA          get-object
         state        find-location )
(P attend-utensil
   =goal> ISA     get-object
          state   find-location
  ; Move attention to the location
  ; screen-x 122 and screen-y 250
   =visual-location>
      ISA        visual-location
      screen-x     122
      screen-y     250
   ?visual> state free
==>
  +visual> ISA    move-attention
           screen-pos  =visual-location
 =goal>  state     attend  )
```

Figure 4: Example of some ACT-R productions responsible for the visual encoding phase

The recognizing phase begins when the chunk of the object is placed in the visual module. This phase implies to recover that specific chunk from the declarative memory. The result of this phase is a chunk that represents the object with some characteristics as color, localization on the screen, name, and kind of object. The motor phase consists of activating the motor actions via a request to the motor buffer in order to click on the object. The three phases process is applied for each object displayed in the interface for the two steps of the recipe. The gathering utensils and ingredients model finishes when the last object of the ingredient list is reached.

The ACT-R model is developed using the ACT-R 6 environment. No noise is introduced in the perceptual motor modules. no retrieval error is modeled in the recognizing phase. These restrictions lead to a deterministic model. Figure 5 shows an example of execution traces of the ACT-R model for the visual encoding and the shift attention actions respectively. The visual-location request takes place at time 0.050 seconds and the request to move-attention is made at time 0.100 seconds. The encoding needs still 0.085 seconds to be completed and store the chunk into the visual buffer.

```
0.000   PROCEDURAL    PRODUCTION-SELECTED START-APPLICATION
0.000   PROCEDURAL    BUFFER-READ-ACTION GOAL
0.050   PROCEDURAL    PRODUCTION-FIRED START-APPLICATION
THE SUBJECT STARTS TO LOOK FOR NEW OBJECT
0.050   PROCEDURAL    MODULE-REQUEST VISUAL-LOCATION
0.050   PROCEDURAL    MODULE-REQUEST GOAL
0.050   PROCEDURAL    CLEAR-BUFFER IMAGINAL
0.050   PROCEDURAL    CLEAR-BUFFER VISUAL-LOCATION
0.050   PROCEDURAL    CLEAR-BUFFER GOAL
0.050   VISION        Find-location
0.050   VISION        SET-BUFFER-CHUNK VISUAL-LOCATION LOC1
0.050   GOAL          CREATE-NEW-BUFFER-CHUNK GOAL ISA GET-OBJECT
0.050   GOAL          SET-BUFFER-CHUNK GOAL GET-OBJECT0
0.050   PROCEDURAL    CONFLICT-RESOLUTION
0.050   PROCEDURAL    PRODUCTION-SELECTED ATTEND-UTENSIL
0.050   PROCEDURAL    BUFFER-READ-ACTION GOAL
0.050   PROCEDURAL    BUFFER-READ-ACTION VISUAL-LOCATION
0.050   PROCEDURAL    QUERY-BUFFER-ACTION VISUAL
0.100   PROCEDURAL    PRODUCTION-FIRED ATTEND-UTENSIL
SHIFT ATTENTION TO A SPECIFIED LOCATION ON THE SCREEN
0.100   PROCEDURAL    MOD-BUFFER-CHUNK GOAL
0.100   PROCEDURAL    MODULE-REQUEST VISUAL
0.100   PROCEDURAL    CLEAR-BUFFER VISUAL-LOCATION
0.100   PROCEDURAL    CLEAR-BUFFER VISUAL
0.100   VISION        Move-attention LOC1-0 NIL
0.100   PROCEDURAL    CONFLICT-RESOLUTION
0.185   VISION        Encoding-complete LOC1-0 NIL
0.185   VISION        SET-BUFFER-CHUNK VISUAL TEXT1
```

Figure 5: Example of execution trace of the ACT-R model for the visual encoding action

Results of the ACT-R model

Figure 6 shows the progress of time depending on progress in the task of get out utensils and get out ingredients respectively. The first task (get out utensils) lasted 6510 ms and the second task (get out ingredients) lasted 8101 ms, the overall time to complete the whole task equal to the sum of the two previous times: 7107 + 8101 = 15208 ms. The time taken to gather the utensils and ingredients is linear depending on the number of objects to search. No differences are observed between the object locations on the screen.

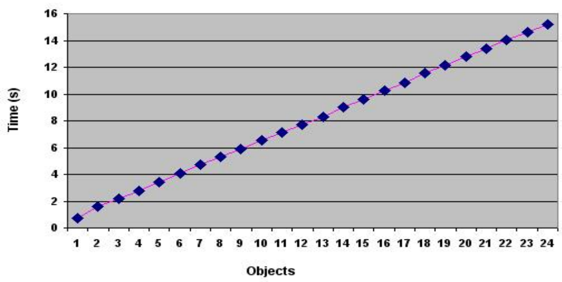

Figure 6: Progress of time depending on progress in the tasks of get out utensils and get out ingredients

MODELING THE INTERACTION WITH THE CONTEXTUAL ASSISTANT USING FITTS LAW

In order to support and validate our results, we used the Fitts Law model, widely used in the evaluation of human machine interfaces. In the Fitts Law, the movement time is proportional to the target amplitude and inversely proportional to the target width.

Fitts Law model

In human machine interfaces, the formulation of Fitts Law (Fitts, 1954) states that the movement time (MT) is function of target amplitude (A) and target width (W). Our model is based on the Mackenzie's [1995] version of Fitts Law in which the movement time (MT) follows the equation:

$$MT = a + b * \log_2(\frac{A}{W} + 1) \quad (1)$$

The second term of the equation (1): $\log_2(\frac{A}{W} + 1)$ is known as the index of difficulty ID, where a and b are constants derived empirically. They can be interpreted by the y-intercept and the slope of a predictive linear regression equation (MacKenzie, 1995) (MacKenzie et al., 1991). In our study, the user- interface interaction is based on the use of a touchscreen, assuming that users remain standing at a distance of 30 cm from the touchscreen, and point directly on the displayed objects by touching them using their index finger. The index finger is held down before starting the interaction, which constitutes the start position. After each pointing action, users returned their index finger to the start position, and the procedure continued like that.

Results of the Fitts Law model

Table 1 shows the index of difficulty values obtained when applying the formulate $\log_2(\frac{A}{W} + 1)$ on some objects displayed in the interface, and the corresponding predicted movement time (MT) obtained by applying the equation (1). The target amplitude (A) remains constant while the button width (W) varies as seen in figure 1.

Object-Name	A (cm)	W (cm)	ID (bits)	MT (ms)
BIG-SAUCEPAN	30	5.8	2.625	614.125
NEXT-BUTTON	30	7.6	2.306	553.834
LOOK-FOR-OBJECT	30	3.8	3.152	713.728
MUSHROOMS	30	5.8	2.625	614.125

Table 1: Index of Difficulty values for some Objects in the Interface and the corresponding movement time

The total time of the whole task applying the Fitts Law is estimated using the following equation:

$$MT_{Total} = \sum_{i=1}^{n} MT_i \quad (2)$$

Where n represents the number of objects used by the user in the interface, and MT_i the corresponding movement time of each object. The total movement time of the whole task applying the equation (2) is: 14977 ms (14.977 s).

COMPARISON OF RESULTS

The results of the predicted time of the task gathering utensils, gathering ingredients and the predicted time of the whole task in both models ACT-R model and Fitts Law model are shown in Table 2.

Tasks	ACT-R	Fitts Law
Predicted time of getting out Utensils Task (ms)	7107	6954
Predicted time of getting out Ingredients Task (ms)	8101	8023
Predicted time of the whole Task (ms)	15208	14977

Table 2: Time estimation of gathering utensils task, gathering ingredients task and the whole task in both models ACT-R and Fitts Law

The ACT-R results are consistent with the Fitts Law model as shown in Figure 7. The predicted time to point each object is very close in both models ACT-R and Fitts Law.

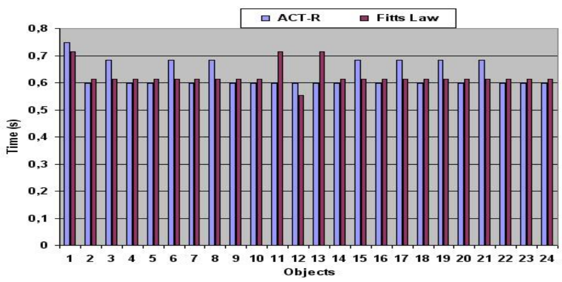

Figure 7: Comparison between the predicted time of each object using ACT-R model and Fitts Law model

GENERAL DISCUSSION

The ACT-R model we developed is proved robust and efficient in our analysis. In fact, the results obtained by the ACT-R model were consistent with those obtained by the Fitts Law model in terms of the predicted time execution of tasks as mentioned previously; this demonstrates that cognitive models and particularly ACT-R can give good predictions in the evaluation of HMI. The results of the ACT-R model show that, the size of objects in the interface is not taken into consideration, and our model does not make difference in the predicted time of the pushing "HELP-ME-TO-DO-THE-TASK" button for example, and the pushing "WOODEN-SPOON" object; these two actions have the same predicted time which equals to 597 ms. Unlike ACT-R model, the Fitts Law model takes in account the object's size in the interface. The predicted time for the pushing "HELP-ME-TO-DO-THE-TASK" button using the Fitts Law is 713 ms and the predicted time for the pushing "WOODEN-SPOON" object is 614 ms. However, some differences are noted as presented in Figure 7. The simulation of the HMI with the object number 11 and 13 takes more time with the Fitts Law. It corresponds to buttons representing the action "HELP-ME-TO-DO-THE-TASK" and "LOOK-FOR-OBJECT" respectively. This is due to the smaller size of these buttons (width = 3.8 cm), compared with the size of the other objects. On the other side, the object number 12 necessitates less time to be pushed. It corresponds to the button representing the action "NEXT", which has the largest size (width = 5.8 cm) in the interface. The simulation of the HMI with the object number 1 as shown in Figure 7, takes more times with the ACT-R model, it corresponds to the button representing the action "LOOK-FOR-OBJECT". This is due to the initialization of the model such as the goal buffer, the retrieval buffer and the visual buffers. In the ACT-R model, the focus is essentially on the visual encoding and recognizing of objects and how to interact with the interface using motor actions. This is supported by some scientific literature such as the use of cognitive models in the evaluation of expert cell phone menu interaction (Amant et al., 2007). The results of the ACT-R model are considered suitable and correct comparing them to those obtained by the Fitts Law model. In fact, as shown in Table 2 the estimated time of the whole task in the ACT-R model (15208 ms) is very close to the Fitts Law model time estimated to 14977 ms. We believe nevertheless, that our study lays out new perspectives of research in this domain particularly how to use perceptual motor modules of the ACT-R architecture to simulate the HMI.

CONCLUSION

The main goal of our study is to evaluate the HMI of a contextual assistant by simulating the HMI, focusing on the time execution of tasks. We used the cognitive architecture ACT-R as a powerful tool to develop our model. Our ACT-R model consists of two parts, the model of the interface of the contextual assistant which represents the environment to interact with, and the model of the cognitive processes required to interact with the interface. The perceptual part of the cognitive processes constitutes the difficult part in our ACT-R model, due to the scarcity in the documentation about the perceptual module in the literature. The results of the ACT-R model were compared with those obtained by the Fitts Law model, developed in this study in order to argue and support our results. The results of our model were consistent with the results of the Fitts Law model. Our model gives a good prediction of user performance, which makes it powerful and realistic.

FUTURE IMPROVEMENTS

The model we developed constitutes the first step of the evaluation of HMI using a contextual assistant. Three futur improvements will add scientific validity to our model. First, the results of our model were compared with those obtained by the Fitts Law model. The results of the Fitts Law model are not always good and exact, but have a certain percentage of errors. It would be interesting to do some experiments with real persons to collect real data and compare them with our results. Second, our model is deterministic and does not make errors. It should be extended to allow errors in the pointing actions. These errors are essentially related to memory problems that may occur in the task modeling (Serna et al., 2005; Dion and Pigot, 2007) and during the interaction with the interface of the contextual assistant. Finally, the action of searching an object is resumed to the HMI with the touch screen. The contextual assistant offers an interaction with the environment to help people recovering utensils and ingredients dispatched in the kitchen. It would be interesting in the future to model this part and simulate the movement of users picking up the objects in the kitchen. Therefore, the extended model should simulate people making a task with contextual assistant and the errors committed by people with cognitive impairments.

REFERENCES

Amant R.S.; Horton T.E.; and Ritter F.E., 2007. *Model-based Evaluation of Expert Cell Phone Menu Interaction. ACM Transactions on Computer-Human Interaction*, 14(1), 1–24.

Anderson J.R.; Bothell D.; Byrne M.D.; Douglass S.; Lebiere C.; and Qin Y., 2004. *An Integrated Theory of the Mind. Psychological Review*, 111, 136-1060.

Anderson J.R.; Taatgen N.A.; and Byrne M.D., 2005. *Learning to Achieve Perfect Time Sharing: Architectural Implications of Hazeltine, Teague, Ivry (2002). Journal of Experimental Psychology: Human Perception and Performance*, 31, No. 4, 749–761.

Bothell D., 2004. *ACT-R 6.0 Reference Manual.* Working draft.

Byrne M.D., 2001. *ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. International Journal of Human-Computer Studies*, 55, 41–84.

Dion A. and Pigot H., 2007. *Modeling cognitive errors in the realization of an activity of the everyday life.* In *Cognitio 2007.*

Eugenio B.D.; Haller S.; and Glass M., 2003. *Development and Evaluation of NL interfaces in a Small Shop. AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, 1–8.

Fitts P.M., 1954. *The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology*, 47, NO. 6, 381–391.

Lussier-Desrochers D.; Lachapelle Y.; Pigot H.; and Bauchet J., 2007. *Apartments for People with Intellectual Disability: Promoting Innovative Community Living Services.* In *2nd International Conference on Intellectual Disabilities/Mental Retardation.*

MacKenzie I.S., 1995. *Movement Time Prediction in Human Computer Interfaces.* In *Readings in human-computer interaction*. 2nd, 483–493.

MacKenzie I.S.; Sellen A.; and Buxton W., 1991. *A comparison of input devices in elemental pointing and dragging tasks.* In *Proceedings of the CHI '91 Conference on Human Factors in Computing Systems.* New York: ACM, 161–166.

Nielsen J. and Phillips V.L., 1993. *Estimating the relative usability of two interfaces: heuristic, formal, and empirical methods compared. Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*, 214–221.

Pigot H.; Bauchet J.; and Giroux S., 2007a. *Assistive devices for people with cognitive impairments*, John Wiley and Sons, chap. 12. The Engineering Handbook on Smart Technology for Aging, Disability and Independence.

Pigot H.; Lussier-Desrochers D.; Bauchet J.; Lachapelle Y.; and Giroux S., 2007b. *A smart home to assist recipes completion.* In *Festival of International Conferences on Caregiving, Disability, Aging and Technology (FICCDAT), 2nd International Conference on Technology and Aging (ICTA).*

Pigot H.; Savary J.P.; Metzger J.L.; Rochon A.; and Beaulieu M., 2005. *Advanced technology guidelines to fulfill the needs of the cognitively impaired population.* In *3rd International Conference On Smart Homes and health Telematic (ICOST).* IOS Press, Assistive Technology Research Series, 25–32.

Serna A.; Pigot H.; and Rialle V., 2005. *Modeling the performances of persons suffering Alzheimer's disease on an activity of the daily living.* In *18th Congress of the International Association of Gerontology.*