Attentive User Interface for Interaction within Virtual Reality Environments Based on Gaze Analysis

Florin Barbuceanu[1], Csaba Antonya[1], Mihai Duguleana[1], and Zoltan Rusak[2]

[1] Transilvania University of Brasov, Department of Product Design and Robotics, 29 Eroilor Bldv., Brasov, Romania
[2] Technical University of Delft, Department of Design Engineering Landbergstraat 15, 2628CE, Delft, The Netherlands
{Florin.Barbuceanu,Antonya,Mihai.Duguleana}@unitbv.ro, z.rusak@tudelft.nl

Abstract. Eye movements can carry a rich set of information about someone's intentions. In the case of physically impaired people gaze can be the only communication channel they can use. People with severe disabilities are usually assisted by helpers during everyday life activity, which in time can lead to a development of an effective visual communication protocol between helper and disabled. This protocol allows them to communicate at some extent only by glancing one towards the other. Starting from this premise, we propose a new model of attentive user interface featured with some of the visual comprehension abilities of a human helper. The purpose of this user interface is to be able to identify user's intentions, and so to assist him/her in the process of achieving simple interaction goals (i.e. object selection, task selection). Implementation of this attentive interface is accomplished by way of statistical analysis of user's gaze data, based on a hidden Markov model.

Keywords: gaze tracking, eye tracking, attentive user interface, hidden Markov model, disabled people.

1 Introduction

Special eye movements (i.e. blinking, rough saccadic shifts or fixations) can be easily recognized and associated with particular intentions or emotional states [1]. Fast blinking could signal the presence of stress, while looking around randomly might indicate a lack of interest. Meaningful information can be extracted from eye movements and interpreted through special attentive user interfaces that are designed to detect the relation between eye movements and user's intentions. When, for example, an interlocutor gazes longer towards another, it is supposed he/she is expecting an answer [2]. Also, previous work shows that by using a virtual assistant a social natural communication protocol based on eye movements can be induced to the user [3]. In this case users can gaze directly towards assistant's face when they want to capture its attention, then towards a certain object to indicate the interest on it, and eventually back to the agent, waiting for a response. This small set of procedures is similar to a simple human to human visual communication protocol. Based on these

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2011, LNCS 6762, pp. 204–213, 2011.

results, a new generation of virtual assistants has been implemented [4], which follows the gaze of the user and adapts the content displayed on a screen depending on subject's interests. This type of user interfaces constitutes a powerful tool for the disabled, reducing the workload of the interaction procedures.

1.1 Previous Work on Attentive User Interfaces

In a general context, the state of uncertainty is characterized by random or irregular eye movements. When all the other communication means are altered, eye movements and gaze gestures can successfully be used to exchange valuable data with the disabled. An attentive interaction interface for disabled should be able to copy some of the comprehension abilities of a human helper and assist him/her when the helper is not around. In order to achieve this, the interface must track the gaze of the users and infer their intensions based on the contextual information of the environment and the relation between gazed objects [5]. A comparative study with the aim of evaluating user's contextual preferences and states of uncertainty based on the analysis of eye movements is presented in [6]. iDict is an application designed to ease the translation of unknown words from a text, through an attentive user interface able to determine when the users are in a state of uncertainty. It compares the measured data about eye movements to existing data on normal dynamics of eye movements measured during text reading. This approach has been extensively studied in the literature. In iDict the process of eye movement analysis is triggered, if at any given time eye movements are abnormal from the way they are expected to be, a process of eye movement analysis is triggered. The results of the eye movement's analysis could for instance state that the process of text comprehension is discontinued and the attentive interface can or cannot take several alternative actions to help the reader getting back on track. One possible action in this context is to translate the words where the discontinuity occurred, and check whether this helps the users going further with their reading. This feature enables users to speed up the process of text comprehension, since normally they would have been constrained to open a dictionary to search for the unknown word. Using an algorithm based on a comprehension difficulty factor [7], an amazing 91% success rate in detecting unknown words, and a very small 2.4% false alarm rate were achieved.

A second study evaluates the uncertainty during the process of reviewing answers given previously to a set of questions. Based on eyes movements' analysis, a "strength-of-belief" (SOB) factor has been defined to characterize the level of uncertainty. In this study scanning eye movements are considered correspondent to high rates of SOB factor, while transition movements indicate low rates. The users have also given a subjective estimation of SOB for each answered reviewed, which has been compared with the real value of SOB. The results indicate that user's estimations of answer correctness are closely related to the values of the SOB factor determined from eye movements' analysis. Their technique can thus successfully be applied for detection of user's uncertainty.

AutoSelect is a user preference detection system based on eye movements' analysis, exploiting the so called gaze "cascade effect" discovered in [8]. This effect consists in the gradually gaze shifting between two similar objects, with the tendency to lean more towards the preferred objects, by the time the preference clears up in

user's mind. The success rate achieved by the AutoSelect system in an experiment based on 8 subjects, 4 males and 4 females is 81%,

The iTourist system provides interactive information about the location of interest on a city map, based on eye movements' analysis [9]. It was designed with the purpose of testing how much extra information can be provided to the user based only on gaze tracking. Every time the users gaze longer over a certain location on the map the system displays guiding information about it. The results indicate a clear advantage over a human guide, because it is hard for humans to constantly follow users' gaze and figure out what they are interested in, while iTourist can easily do this, constantly providing guiding information about the locations of interest.

2 Conscious vs. Unconscious Action Triggering

Face-to-face communication between humans is a complex process of information transfer both verbally and visually. Not only the words count for the data exchanges, but also the tone of the voice, speech rhythm, face expression, all these important clues about the intrinsic emotional and cognitive states. The results in [10] shows that eyes and mouth are the dominant areas gazed during a dialog.

Simple eye gestures drive the communication within the Eye Bed interface research project [11]. A selection is made through a prolonged blink. Gazing around randomly signifies lack of interest, while staring indicates attention. If eyes get closed after lying in a bed, it means the user wants to sleep. Eyes opened after a long time sleep means the user will wake up, and so the light in the room can be opened, the radio may be started up etc. This kind of simple eye movements can be easily interpreted and implemented in effective attentive user interfaces.

Predefined conscious eye's gestures can carry cognitive data about user's intentions, significantly improving the communication between humans. They simplify the process of message extraction from eye movements data, but they also constrain the user to perform unnatural eye movements, overloading thus with motor control tasks the perceptual function of the visual channel [12]. In this case eyes play a role of an output communication channel, used to express user's intentions through consciously controlled eye movements. But because the eye is a perceptual organ not meant for motor controlling tasks [7], interaction actions based on gaze should be triggered without any tedious eye gestures, but rather they should be detected from the context. Specific sequences of eye movements can be sufficient to decide when an action needs to be triggered. An assistive interface able to detect the correct significance of a specific sequence of eye movements and take the corresponding action at the right time can free the user from performing unnecessary burden eye gestures. This means that the cognitive load at the user is transferred to the attentive interface, which needs to come with more sophisticated and more precise intention detection algorithms. When designing an attentive interface based on unconscious action triggering, there are several issues to confront with. For instance, the "Midas touch" problem appears when a false alarm dwell time selection is triggered [13]. The consequence of improper intention detection could reside in ineffective communication leading to frustration and eventually rejection. To compare, conscious

action triggering can improve the accuracy of the intention recognition through predefined eye gestures, at the cost of overloading the natural functions of the eyes, while unconscious action triggering is more natural, does not disturb the normal functions of eyes, with a possible cost of lower accuracy and higher complexity attentive systems implementations. If the accuracy problems can be minimized through highly robust attentive systems, then unconscious action triggering interaction would ultimately be the right interaction interface to be implemented. Just imagine how difficult the usability of the iDict application [6] would be if instead of automatic detection of uncertainty, the users should blink their eyes every time a translation of a word is needed.

3 The Experimental Setup

Our experiments were conducted in a virtual reality environment to test the usability of the human-computer interaction interfaces developed for people with severe locomotion disabilities. The virtual environment is a kitchen with several objects placed on the tabletop. The user can interact with these objects by looking towards them using a simple command based gaze interaction protocol, or a more complex attentive user interface based on a hidden Markov model implementation. Images are projected on a 3D stereoscopic visualization screen and the subject is immersed in the virtual environment through a pair of polarized 3D glasses. The eye tracking device used was the head-mounted model ASL H6-HS-BN [14].

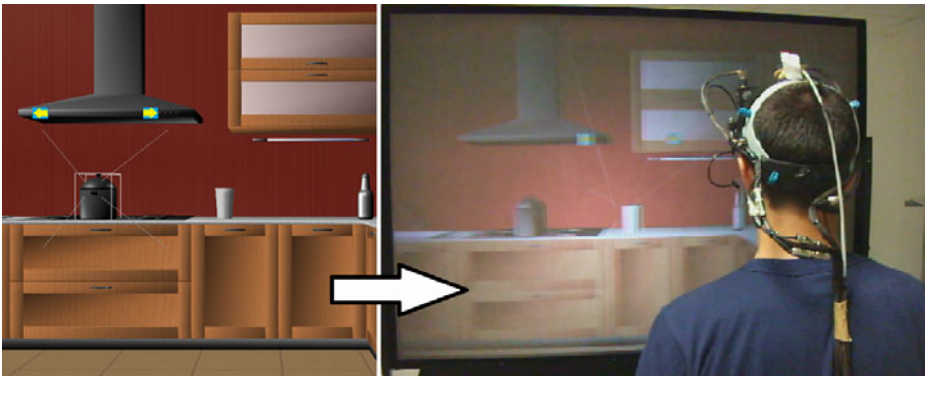

Fig. 1. A virtual kitchen containing objects the user can interact with: a bottle, a glass, and a pot. The virtual scene is projected on a 3D visualization screen in front of the user.

The graphics of the virtual environment is rendered by the XVR (Extreme Virtual Reality) software platform. The 3d model of the kitchen (Fig. 1) is composed by 3dsmax objects available at [15]. Selection of virtual objects available in the virtual kitchen is made by intersecting the line of sight, detected by the eye tracking device, and the objects, through the IsColliding() function, available whithin XVR SDK. Provided that the visual angle accuracy of the eye tracking device is 0.5^{o} [14], using a

bar object with a given thickness, instead of a thin line, the chances of successful selection of objects through gaze is increased [16]. The distance between neighboring virtual objects is about 50 cm, and the distance to the user is 3 m. This corresponds to a spatial separation between two virtual objects of approximately 9^0 from the user's point of view (Fig.2).

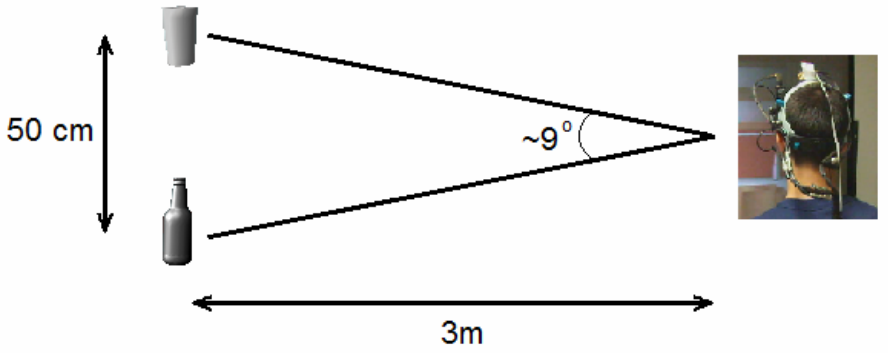

Fig. 2. Spatial separation of neighboring objects, from user's point of view

With the given eye tracking measurement accuracy of 0.5^0, at 3m distance from the user, the accuracy of selection is about ±2.6 cm around the real gaze point. The glass and the bottle at this distance have dimensions that can negatively influence the precision of eye tracking, so the selection can easily fail. In order to increase the success rate of selections, the thickness of the bar was chosen to be 10 cm. After some tests this thickness proved to generate most successful selections and it was adopted for further experiments.

4 Attentive User Interface Based on Hidden Markov Chains

Thoughts, intentions or preferences of a person are hidden deep inside the mind. Physically impaired persons, who cannot use conventional communication channels, face a major drama of not being able to express their thoughts or needs when necessary. Because of the hidden nature of these cognitive states, one solution for guessing them is to use a hidden Markov chains approach in the implementation of an attentive user interface based on gaze tracking.

A hidden Markov model is a hidden stochastic process which cannot be directly observable, but it generates dependent non-hidden stochastic emissions [17]. By analyzing a sequence of input data and knowing the transition and emission matrixes, it is possible to estimate the state they were generated from. In this implementation, the sequence of emissions consists in the recorded identification (ID) number of the objects gazed by the user over time (Tab. 1). Several objects have been displayed in a virtual kitchen, a bottle, a glass, a pot and a tap. Each object has an ID of its own, as follows:

Table 1. IDs of the virtual objects the user wants to interacts with

	Glass	Bottle	Tap	Wheelchair	Closet door	Pot
Object ID	**1**	**2**	**3**	**4**	**5**	**6**

Within the hidden Markov model five states were considered, each one corresponding to a different intention the user may have. These states are described below:

S1 = "I want to pour water from bottle into glass";
S2 = "I want to drink water from the glass";
S3 = "I want to wash the glass";
S4 = "I want to open the upper closet";
S5 = "I want to poor water from glass to pot".

The transition matrix contains the probability of system transitions from one state to another. The values in this matrix influence the correct evolution of the statistical system. For example in Tab. 2, the system can easily evolve from S1 to S2, S2 to S3, S3 to S4, and S4 to S5, but it is unlikely that it will evolve from S5 to S1, S2 or S3, since the probabilities of these transitions were set to 0.1, in order to prevent the system from evolving in those states.

Table 2. System transition matrix of the hidden Markov model considered

	S1	S2	S3	S4	S5
S1	**0.5**	**0.2**	**0.05**	**0.05**	**0.2**
S2	**0.3**	**0.4**	**0.2**	**0.05**	**0.05**
S3	**0.2**	**0.05**	**0.4**	**0.3**	**0.05**
S4	**0.1**	**0.1**	**0.1**	**0.3**	**0.4**
S5	**0.1**	**0.1**	**0.1**	**0.4**	**0.3**

The sequence of input data is compared with the values of the observations/ emission matrix, which has a unique set of data for every state. Thus, if a sequence of input data matches the emission matrix of a specific state, the probability that the system might evolve in that state is evaluated, and if it is high enough, the system will evolve.

The same data from Tab. 2 and 3 are reproduced in a more intuitive representation in Fig. 4 and 5. If counted, a total of 13 transitions have real chances to take place. The other transitions represented with thin lines are common situations in which the system cannot go, either because it is not possible or it is not desirable.

Table 3. Observations/emissions matrix of each state of the hidden Markov model considered

	O(t)	O(t+1)
S1 (O1)	**ID1**	**ID2**
S2 (O2)	**ID1**	**ID4**
S3 (O3)	**ID1**	**ID3**
S4 (O4)	**ID4**	**ID5**
S5 (O5)	**ID1**	**ID6**

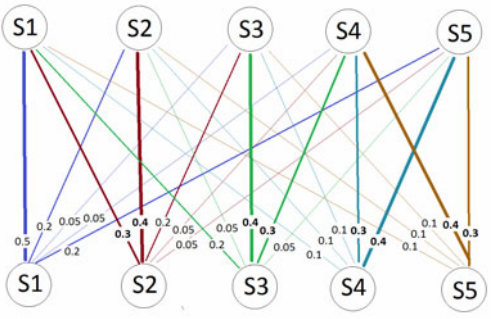

Fig. 4. Visual representation of the chosen states transitions probabilities of the hidden Markov model considered. Thicker lines and the associated values indicate higher probabilities of state transition. The representation is unidirectional, upwards.

In Fig. 5, each observation/emission has a highest correspondence connection with one state, which means it is most probably that the specific emission is produced by that particular state. These emissions are compared with the sequences of incoming data.

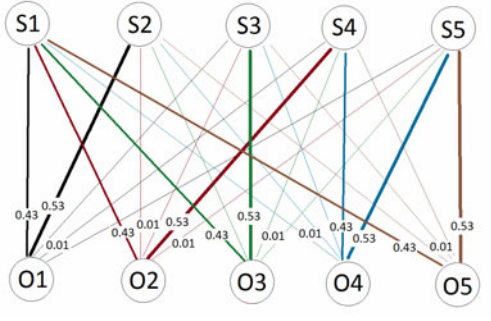

Fig. 5. Visual representation of the correspondence between observations/emissions and system states. Each observation has a most significant correspondent state.

Given a particular sequence seq=[1,2,1,4,1,3,4,5,1,6], the system will evolve as expected from one state to the other. The estimatedStates vector demonstrates the reaction of the system to a particular user input sequence:

```
>> seq=[1,2,1,4,1,3,4,5,1,6]
seq =
1 2 1 4 1 3 4 5 1
estimatedStates=hmmviterbi(seq,trans,emis)
>> estimatedStates =
1 1 1 2 2 3 4 4 5
```

This example corresponds to the case in which the users gaze towards the glass, than towards the bottle, glass, wheelchair, glass, tap, wheelchair, closet door, glass and eventually pot. The estimated states show that when the users gaze sequentially

over the glass and bottle, the attentive interface determines that they want to pour some water from bottle to glass, which is correct. As it can be seen, detection of intentions in this way is simple and precise.

To go further with the analysis of system's response, given a sequence of incoming data which match two not directly transmittable states, it is noticeable that the system goes in an intermediate state until it reaches the final matched state:

```
>> seq=[1,2,1,3]
seq =
1 2 1 3
estimatedStates=hmmviterbi(seq,trans,emis)
>> estimatedStates =
1 1 2 3
```

If we consider the description of states S1, S2 and S3 involved in this example, it is a matter of logic that is not a common procedure to wash the glass if it was clean and some water was just poured into it. It is more likely that first the user should drink that water, and afterwards it might be necessary to have it cleaned. If direct transition is not possible, the system can evolve in intermediate states from which it can go further. As the number of states implemented is higher, more appropriate actions can be triggered in case of such direct transitions, which are either not recommended either impossible to be performed. Thus, the attentive interface will be able not only to understand what the user wants, but also to suggest alternatives when certain tasks are difficult to complete. To ensure a more accurate response, the system can wait for a longer sequence of inputs from the user (five user inputs, or more). In this way, if erroneous selections are made, the system can reject them very efficiently and infer the correct user intention.

5 Design of Experiments

The goal of the conducted experiments was to asses the effectiveness of the attentive user interface implementation based on hidden Markov model. The final purpose was to evaluate the accuracy of intention detection algorithm and also whether this interface is more natural compared to a command based interaction interface. 10 subjects have tested the two interfaces mentioned and at the end each one filled in a questionnaire about their experience of interacting with the virtual objects. Half of them were informed about the possible operation they could perform with virtual objects through the attentive interface. A description of all the five possible states was presented to these subjects, so they became aware about the possibilities and limitations of this interface. The other subjects were left to discover these features all by themselves during the experiments. They were all informed about the functionality of the command based interaction interface. This interface features instant selection of virtual objects, if the user gazes towards them. At the end they all answered a common set of questions, and in addition the five subjects who were not informed about the possibilities of the attentive interface, were asked whether their first impression of discovering it's features by themselves was positive or negative. The common requirement for all users were: 1) to count for the overall number of failed intention detections; 2) to compare which interface is more natural and less obtrusive.

6 Discussion of Results

The task of the users was to perform selections of virtual objects with the two interaction interfaces developed. After repeating the tests three times, those five subjects who were previously informed about the features of the attentive interface, managed to reach an 88% success rate, with an average of 1.8 failed intention detections per subject. The numbers are based on the answers they filled in the questionnaire after they completed the tests. The selection procedures made through the command based user interface reached a 100% precision, mostly due to the sufficiently distant arrangement of objects in the virtual environment. The answers to the second question indicates that the attentive interface based on the hidden Markov model implementation reacts more naturally and does not divert their attention from the surrounding environment. Some interesting user's comments state that it is fun to connect two objects with the eyes and see the system detecting what they can do with them. Those five users not aware about the features of the attentive interface reported a higher rate of failure, as some of them were expecting different possible operations between the gazed objects than the ones implemented in the system. Thus, the average failed detections were 5.6, with a corresponding 62% success rate. Some of them were not sure about the exact number of failed detections, as they were diverted from counting by the surprising discovery of unexpected connections between different objects. The answers to the second question state that for some users, the attentive interface is more natural, while for some others the difference is insignificant. They all confirmed that the attentive interaction interface is more intuitive than the command based interface.

Although this current implementation of an attentive user interface based on hidden Markov model is a rough one, the users of the system still managed to score an amazing success rate of 88%. This is encouraging as the possibilities of improving the accuracy are numerous.

Acknowledgment. This work was supported by the Romanian National University Research Council (CNCSIS-UEFISCDI), under the Grant INCOGNITO: Cognitive interaction between human and virtual environment for engineering applications, Exploratory Research Project PNII – IDEI 608/2008.

References

1. Majaranta, P., Räihä, K.J.: Twenty years of eye typing: systems and design issues. In: Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA 2002), pp. 15–22. ACM Press, New York (2002)
2. Selker, T.: Visual attentive interfaces. BT. Technol. J. 22(4), 146–150 (2004)
3. Prendinger, H., Ma, C., Yingzi, J., Nakasone, A., Ishizuka, M.: Understanding the effect of life-like interface agents through eye users' eye movements. In: Proceedings of Seventh International Conference on Multimodal Interfaces (ICMI 2005), pp. 108–115. ACM Press, New York (2005)
4. Eichner, T., Prendinger, H., André, E., Ishizuka, M.: Attentive presentation agents. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 283–295. Springer, Heidelberg (2007)

5. Vertegaal, R.: Designing attentive interfaces. In: Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA 2002), pp. 22–30. ACM Press, New York (2002)
6. Prendinger, H., Hyrskykari, A., et al.: Attentive interfaces for users with disabilities: eye gaze for intention and uncertainty estimation. Univ. Access Inf. Soc. 8, 339–354 (2009)
7. Hyrskykari, A., Majaranta, P., Räihä, K.-J.: From gaze control to attentive interfaces. In: Proceedings of HCII 2005. Erlbaum, Mahwah (2005)
8. Shimojo, S., Simion, C., Shimojo, E., Scheier, C.: Gaze bias both reflects and influences preference. J. Nat. Neurosci. 6(12), 1317–1322 (2003)
9. Qvarfordt, P., Zhai, S.: Conversing with the user based on eyegaze patterns. In: Proceedings of the ACM CHI 2005 Conference on Human factors in Computing Systems, pp. 221–230. ACM Press, New York (2005)
10. Bailly, G., Raidt, S., Elisei, F.: Gaze, conversational agents and face-to-face communication. J. Speech Communication 52, 598–612 (2010)
11. Selker, T., Burleson, W., Scott, J., Li, M.: Eye-Bed. In: Workshop on Multimodal Resources and Evaluation in conjunction with the Third International Conference on Language Resources and Evaluation, LREC (2002)
12. Toet, A.: Gaze directed displays as an enabling technology for attention aware systems. J. Comp. in Human. Beh. 22, 615–647 (2006)
13. Jacob, R.J.K.: What you look at is what you get: eye movement-based interaction techniques. In: Proceedings of the SIGCHI Conference on Human factors in Computing Systems: Empowering People, New York, pp. 11–18 (1990)
14. Applied Sciens Laboratories, `http://www.asleyetracking.com`
15. 3d models on turbosquid, `http://www.turbosquid.com`
16. Barbuceanu, F., et al.: Evaluation of the average selection speed ratio between an eye tracking and a head tracking interaction interface. In: 2nd Doctoral Conference on Computing Electrical and Industrial Systems, Costa de Caparica, Portugal, pp. 181–186 (2011)
17. Cappe, O., Moulines, E., Ryden, T.: Inference in Hidden Markov Models. Springer, Heidelberg (2009)