Evaluation of Personalized Information Systems: Application in Intelligent Transport System

Makram Soui, Christophe Kolski, Mourad Abed, Guillaume Uster

Evaluation of Personalized Information Systems: Application in Intelligent Transport System

M.Soui[1], C.Kolski[1], M.Abed[1], G. Uster[2]

[1] LAMIH-UMR CNRS 8530, University of Valenciennes and Hainaut-Cambrésis, Le Mont Houy,
59313 Valenciennes cedex 9, France
{Makram.Soui, Mourad.Abed, Christophe.Kolski, emmanuelle.grislin}@univ-valenciennes.fr

[2] The French national institute of research for transport and security, INRETS, Villeneuve d'Ascq, France
guillaume.uster@inrets.fr

Abstract— **Thanks to the emergence of the Personalized Information System (PIS), it becomes possible to supply the user only with the pertinent information that directly interests him/her and suits his/her preferences. However, we need beforehand to evaluate these systems in real situations. In this paper, we point out the insufficiencies in the evaluation of the PIS. Then, we propose a new evaluation method for PIS. Finally, we describe the evaluation results of a demonstrator developed during a project called MouverPerso. This demonstrator was tested among a group of subjects in the University of Valenciennes.**

Index Terms— **Evaluation, Human-Computer Interaction (HCI), Intelligent Transport System (ITS), Personalized Information System (PIS)**

I. INTRODUCTION

Nowadays, the information systems, addressed to the users and mainly in the field of transport, tend to be more and more personalized. In the midst of a universal project labelled Intelligent Transport System (ITS), researchers' primary preoccupation is to provide personalized information for the public transport users. In fact, the PIS is a system which has the capacity to be adapted to the user taking in consideration his/her preferences [1]–[2]–[3]. For example, in transport field, the traveler hopes to have at his/her disposal only some information, just what he/she is directly interested in [4]–[5]. In addition, the traveler may have access to a reliable, multi-modal and personalized information using various supports (PC, PDA, mobile phone, etc.) [6]. Though, there are different methods and approaches to conceive PIS systems, to our knowledge and at the present point of research, we notice a lack of methods to evaluate the personalization quality of PIS. This article is made of three principal parts: at first we focus on the insufficiencies concerning the evaluation of PIS. The second part is devoted to describe the basic principles of a proposed method that permits the evaluation of SIP. The third and last part is meant to describe the evaluation results of the demonstrator developed during a project called MouverPerso.

II. INSUFFICIENCY IN THE PIS EVALUATION

The evaluation of the interactive system has been a recurrent problem since the last three decades. To ameliorate the quality of human machine interaction many studies were oriented towards the evaluation of the interactive system from different angles and view points. Concerning this subject we should mention that several papers and works had defined the basic principals of the evaluation as well as the methods used in this evaluation [7]–[8]–[9]–[10]–[11]–[12]–[13]–[14]–[15]. We notice that the focal point in these works was on the utility and usability while the dimension of personalization was neglected. Nowadays, the PIS users are facing many difficulties to interact with the badly studied or evaluated personalized systems that do not always answer their needs. Due to the complexity of personalized system interface, new criteria and methods are needed to evaluate the human machine interaction. In this context, we propose a method which is based on the evaluation criterion of PIS detailed in [15] and on the explicit intervention of the user who fills up questionnaires.

III. PROPOSITION OF A METHOD FOR EVALUATING PIS

The process of this evaluation is illustrated in the fig. 1. To give a clear structure to this proposition we use the SADT formalism; a well known in software engineering and in human machine interaction. This method is made up of three phases. According to the SADT formalism, we find in the box A0 (the box in top of diagram) the general objective that consists in the evaluation of PIS. To reach this target we have decomposed it into three sub-targets presented in the boxes A1, A2 and A3 (preparation, evaluation and analysis).
- Phase A1: it represents the preparation stage wherein the evaluator chooses the representative tasks on which the evaluation will be based. The evaluator also prepares two types of documents necessary to the evaluation. The first is a general questionnaire including general information about the users and the second is an index-card that includes the definition of every criterion and the parts which the user fills up during the experimentation. We distinguish seven global criteria, detailed in [15].
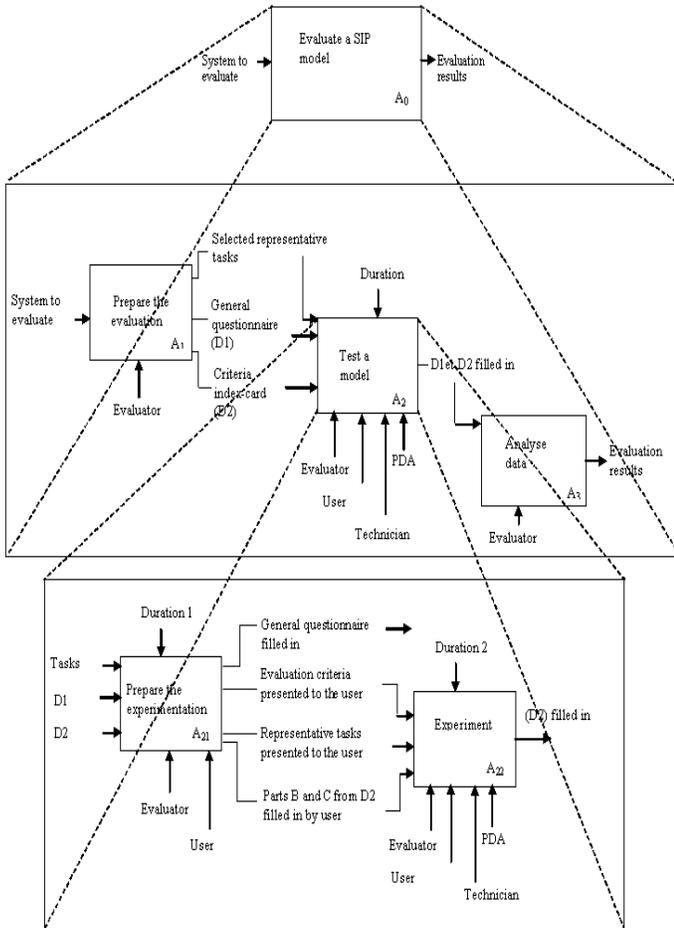
Fig. 1. Evaluation phases of the proposed method.

- Phase A2: it represents the stage of testing the model; it is made of two sub-phases (A21 and A22):

- Sub-phase A21: it represents the pre-experimental stage; the evaluator presents to the users the tasks already chosen in addition to the evaluation criteria and their definition. Then the evaluator asks them to select the most important criteria according to their needs and/or preferences. The users must attribute to every selected criterion a weight representing the importance accorded to this criterion.

- Sub-phase A22**:** it represents the experimental stage, in fact the users have to evaluate the system progressively while executing the tasks in terms of the already defined criteria. Then he/she allocates to every criterion a mark according to his/her level of satisfaction and according to a well predefined scale.

- Phase A3: it represents the analysis phase. In this level, the evaluator calculates to every user the level of satisfaction N_s that translates the level of personalisation N_p of the system applying the following formula:

$$N_s = N_p = \frac{\sum_{K=1}^{n} W_K \times I_K \times N_K}{\sum_{k=1}^{n} W_K \times I_K}$$

With

n: number of criteria the user is concerned with.

W_K: the weight of interest representing the importance the user gives to this criterion.

I_K: the criterion index of activation.

   1 if the user U is concerned with the criterion k

   0 if not

N_K: the mark attributed by the user to the criterion k

   0   if the user is not satisfied at all

   0.25   if the user is a bit satisfied

   0.5   if the user is fairly satisfied

   0.75   if the user is satisfied

   1   if the user is very satisfied

The result is a value restricted between 0 and 1.0 so that the more the satisfaction level is close to 1 the more the system is adapted to the user. To judge the system according to its degree of personalization, we calculate the average of the satisfaction level N_s of all the subject participated in the evaluation.

IV. CASE STUDY: IN ITS DOMAIN

● Context: Our work is a part of the project « MOUVER.PERSO» achieved with the collaboration of National Institute of Research on Transport and its Security (INRETS). This system aims to incite the usage of collective transport by ensuring the complementarity between different modes of transport and ameliorating the quality as well as the availability of the personalized information.

● Protocol test: The objective of this experience was to evaluate the demonstrator MouverPerso taking as basis the proposed approach and the criteria we have just defined for the evaluation of SIP.

- The participants: Twenty three people have participated in this evaluation, two expert evaluators, twenty subjects (twenty students in computer science) and a technician.

- The evaluated task: We focused on a representative task of the application, a research of itinerary after adding an appointment. This task needs the consultation of at least three interfaces of the application (an interface which permits to add an appointment, an interface which permits to consult the itinerary details and an interface which permits to consult the appointment list). The dynamic task is represented by Statechart diagram from UML (see fig. 2).

- Tools and techniques: In this experimentation, several tools were used namely:

▪ The general questionnaire: the evaluator invites the user to fill in this questionnaire which includes some general information such as last name, first name, email, age, gender. This questionnaire is distributed among the subjects in the pre-experimental step.

▪ The criterion index-card: the user makes use of his/her index-card to judge the system for example s/he could attribute weights and marks, mention problems and/or draw the

attention to them and may propose ideas to improve the system during the experience.

▪ A PDA (Portable Digital Assistant): is a personal assistant taking the form of a mobile digital appliance. It consists of a computer equipped with a tactile screen and a stylet (fig. 6).
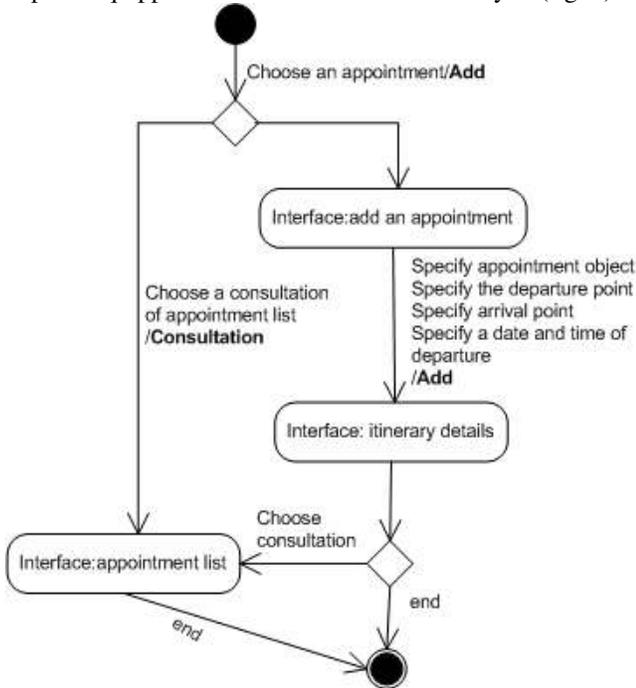

Fig. 2. The dynamic task of adding an appointment.

V. RESULTS AND ANALYSIS

This part represents the principal results issued from the test and it is concerned with the dimensions related to the content and container personalization.

A. Content Personalization

Three criteria were considered in the personalization of content: the preferences, point of focus and user experience as we see in fig.3.: 47% of the users are very satisfied, 41% are satisfied, 6% are less satisfied and finally 6% of them are not satisfied at all (see fig. 3).
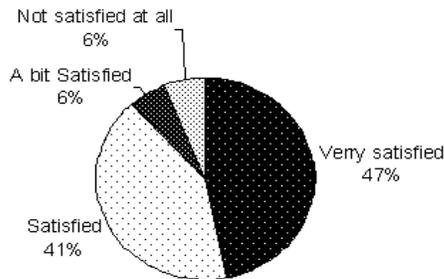

Fig. 3. The subjects' opinion about the personalization related to the content.

In order to give more details about the students' opinion concerning the personalization content, we have compared the average of their level of satisfaction according to the three criteria related to the quality of content personalization. For preferences and experiences, the averages of satisfaction in connection with these criteria exceed 0.6. For the interests the average is less satisfying (see fig. 4).
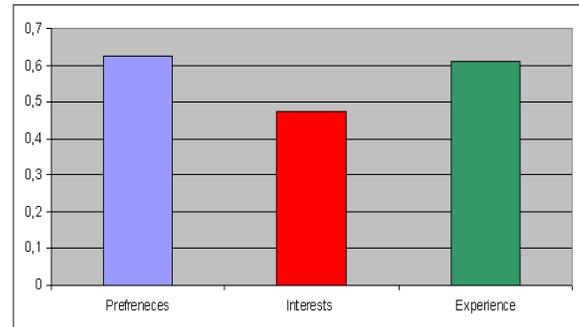

Fig.4. Average of the student's level of satisfaction per personalized criterion related to the content.

B. Container Personalization

The personalization related to the container groups four criteria: adaptation to interactive platform, adaptation to environment, adaptation to user's behaviour and finally adaptation to users' physical capacities (accessibility). The rate of satisfaction is 33% for the very satisfied subject, 17% for the satisfied and 50% for the bit satisfied ones (see fig. 5).
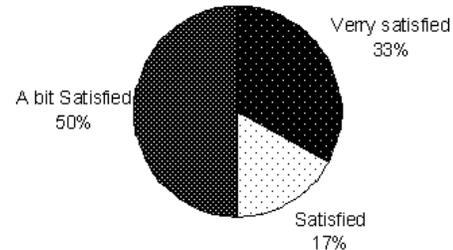

Fig. 5. The subjects' opinion about the personalization related to the container.

In order to give more details about the students' opinion concerning the personalization related to the container, we have compared the average of their level of satisfaction according to the four criteria related to these dimensions.

Fig.6. Test with PDA

After having tested the application in using two different interactive supports (PC, PDA) (see fig.6.) the subject gives his/her opinion about the adaptation of the system to the interactive platform. Concerning the adaptation criterion of the users behaviour, the students' opinion are centred on the

neutral response (the average = 5), while as showed in fig.7, the student have a disapproving opinion about the adaptation to the users' physical capacity and to the environment (the average <4).
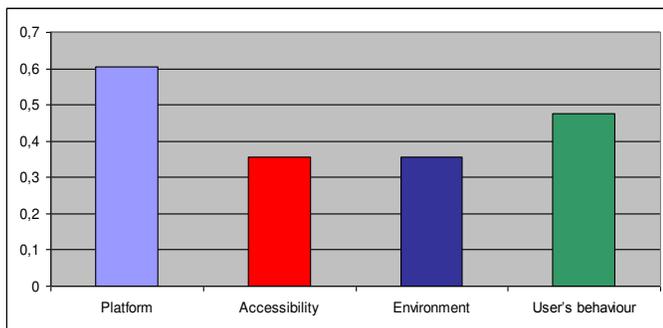

Fig.7. Average of the student's level of satisfaction per personalized orientation related to the container.

VI. PROBLEMS MENTIONED BY THE SUBJECTS

The fig. 8 shows that the majority of the subjects have mentioned some problems related to experience, accessibility and users behaviour. About accessibility a problem is cited several times which is the smallness of the characters that causes a bad legibility. Concerning the behaviour, the subjects notice the lack of warning signals during the validation task. The subjects had mentioned few problems related to the preferences, to interests and the interactive platform. This could be explained by the importance the designers of this system had given to the dimensions related to the personalization of the content and to the adaptation to the interactive platform when compared with the others criteria.
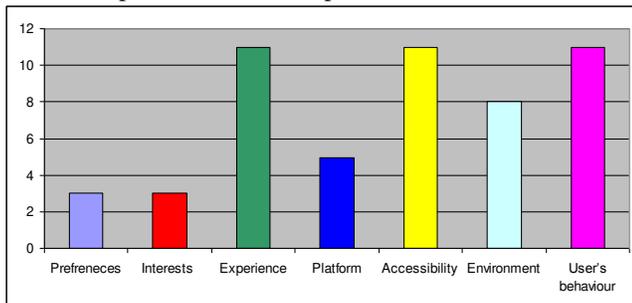

Fig.8. Problems repartition according the evaluation criteria evaluation.

VII. CONCLUSION AND PERSPECTIVES

In this article we have mentioned the insufficiencies in the evaluation of PIS and proposed a method to evaluate such systems. This method was tested among a group of subjects using the demonstrator MouverPerso. The information we have collected allowed us to know the subjects' level of satisfaction, the problems they have encountered and the ideas they proposed to ameliorate this system. This method is based on the explicit user intervention and on the existence of a real system (model, prototype or final system). Despite that it provides concise results, since they are explicitly obtained from users' opinions, the evaluator and user spend much time

to achieve PIS evaluation. It also disturbs the user in his/her main activity. Besides, this evaluation needs several iterations to find maximum of problems. These reasons make us think about another evaluation method that does not need the direct intervention of the user. The principle of this proposed method which makes the object of our future research will be based on the usage traces of the system.

REFERENCES

[1] R. Barrett, P. Maglio and D. Kellem, How to personalize the Web, in Proceedings of CHI'97, 1997.

[2] I. Cingil, A. Dogac and A. Azgin, A broader approach to personalization, Communications of the ACM 43 (8) 136–141, 2000.

[3] P. Maglio and R. Barrett, Intermediaries personalize information streams, Communications of the ACM 43(8) 96–101, 2000.

[4] C. Basu, H. Hirsh, and Cohen, W., Recommendation as Classification: Using Social and Content-Based Information in Recommendation. In C. Rich and J. Mostow (Eds.), Proceedings of the 15th National Conference on Artificial Intelligence, pp. 714-720. Madison, USA, AAAI Press/MIT Press, 1998.

[5] R. Hoyer, and O.Czogolla, Approach to personalised information services to public transport. In 9th world congress on intelligent transportation systems, Chicago Illinois, October-14-18, 2002.

[6] G.D Abowd, A.K.Dey, P.J. Brown, M. Smith, and P. Steggles, Towards a Better Understanding of Context and Context-Awareness. Lecture Note In Computer Science, Vol. 1707. Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing, Karlsruhe, Germany, pp. 304-307, September, 1999.

[7] R.Wilson., E.N. Corlett, Evaluation of human work, a practical ergonomics methodology.Taylor&Francis, 1990.

[8] B. Senach, Evaluation de l'ergonomie des IHM, Actes du Congrès ERGO-IA'90, ergonomie et informatique avancée, Biarritz, 1990

[9] A. Whitefield and al., A framework for human factors evaluation. Behaviour & Information Technology, vol 10, n°1, pp. 65-79, 1991.

[10] M. Sweeney and al., Evaluating user-computer interaction: a framework. Int. J. Man-Machine Studies, 38, 689-711, 1993.

[11] S. Balbo, Evaluation ergonomique des interfaces utilisateur, un pas vers l'automatisation. Thèse de doctorat, Université de Grenoble I, 1994.

[12] J-C. Tarby, H. Ezzedine, , C-D Tran,., P. Laporte, and C. Kolski, Traces using aspect oriented programming and interactive agent-based architecture for early usability evaluation: basic principles and comparison. J. Jacko (Ed.), Human-Computer Interaction, Part I, HCII 2007, LNCS, Springer-Verlag, pp. 632-641, 2007.

[13] J. Huart, C. Kolski, M. Sagar, Evaluation of multimedia applications using inspection methods: The Cognitive Walkthrough case. Interacting With Computers,16 ,pp. 183-215. 2004.

[14] H. Ezzedine, A.Trabelsi, C. Kolski Modelling of an interactive system with an agent-based architecture using Petri nets, application of the method to the supervision of a transport system. Mathematics and Computers in Simulation, 70, pp. 358-376. 2006.

[15] M. Soui, M. Abed, C. Kolski, K. Ghedira, Criteria devoted to evaluate personalized interactive systems. M. Abid, A. Hadj Kacem, M. Jmaiel, M. Lahiani (Ed.), Nouvelles tendances technologiques en génie électrique & informatique, GEI'2007, Edition CPU, Sfax, pp. 119-125, mars, ISBN 978-9973-61-631-9 , 2007.