

Usability evaluation of handheld devices: A case study for a museum application

Adrian Stoica¹, Georgios Fiotakis¹, Jorge Simarro Cabrera², Henar Muñoz Frutos², Nikolaos Avouris¹, and Yannis Dimitriadis²

¹ University of Patras, Human-Computer Interaction Group, GR-26 500 Rio Patras, Greece
{Stoica, fiotakis }@ee.upatras.gr; avouris@upatras.gr
<http://www.hci.ee.upatras.gr>

² University of Valladolid, Intelligent and Cooperative Systems, 47011 Valladolid, Spain
yannis@yllera.tel.uva.es

Abstract. In this paper, we describe a usability evaluation study of a system involving PDAs, designed to be used in a traditional historical/ cultural museum. The system permits collaboration of small groups of museum visitors through mobile handheld devices. The key characteristics of the system are described first, which include a server and a client component and a tool that permits authoring of new activities. The usability evaluation study that involved typical users revealed some of the limitations of the design. The reported findings can be of use to practitioners interested in following similar approaches relating to evaluation of mobile technology.

1 Introduction

This paper discusses our experience with designing and evaluating collaborative learning activities in a Museum, using handheld devices (PDAs).

Usability evaluation of mobile systems is a new area of research (Kjeldskov and Stage, 2004). There are not yet clearly defined techniques and methodologies to conduct usability evaluation studies for such systems. In mobile systems the context and the surroundings as well as the other people around play an important role making this way the evaluation process extremely difficult. Taking these considerations into account for usability evaluation of mobile systems we need to use extensively field based studies. There are three fundamental difficulties reported in literature (Kjeldskov and Stage, 2004) regarding field studies: (a) it is complicated to establish realistic studies that capture key situations, (b) it is very difficult to apply evaluation techniques such as observation and think aloud, (c) the data collection is far more complicated and the control over the environment is very limited.

On the other hand the laboratory based evaluation studies significantly reduce these problems. But in this case there is the disadvantage of the lack of realism. There are approaches that try to recreate or to simulate the real context of use in the laboratory, though some times it is very difficult or quite impossible to recreate realistically the context and the surroundings.

Kjeldskov and Stage (2004) proposed a set of new techniques for evaluating the usability of mobile systems that are focused on (a) the different ways the user is moving and the attention needed to navigate as well as the notion of divided attention.

Our case study concerned usability evaluation of a mobile application designed for a traditional historical and cultural museum. The usability evaluation technique is laboratory based of activities of typical users with a considerable degree of realism.

The application prototype under evaluation is also described in this paper. Its objective is to augment interaction with the museum through a mystery play that stimulates children's' imagination. The plot involves a number of puzzles that relate to the exhibits of the Museum and their solution brings rewards to the players. These puzzles, the most typical examples of which involve images of certain exhibits and scrambled verses from manuscripts of the Museum, necessitate collaboration for their solution, as the necessary pieces are spread in the mobile devices of the members of the group. The aim of the activity is to mix the real and the virtual world and to make children work together in a collaborative way in this setting. The application shares many characteristics with other Museum based mobile systems (Raptis et al. 2005).

In the rest of the paper we first present an overview of usability evaluation technology and then we discuss the design and the architecture of the developed system. We describe next the usability evaluation study of the system and finally outline conclusions we have drawn from both the development and the evaluation phases.

2 Usability Evaluation of mobile applications

Software usability has been the subject of many international standards, directives and theoretical and empirical research during the last years. At the same time many practical techniques for measuring usability have been proposed to be introduced in the interactive software development life-cycle (Dix et al, 2004, Schneiderman 2003, Avouris 2001, etc.).

Usability was originally related with making systems easy to use and easy to learn, as well as supporting the users during their interaction with computer equipment. There have been however many attempts to relate the term to more attributes and metrics. According to ISO 9241-11 standard Usability is defined as the "extent to which a product can be used with effectiveness, efficiency and satisfaction in a specified context of use" (ISO 9241). The attributes which a product requires for usability depend on the nature of the *user*, *task* and *environment*. A product has therefore no intrinsic usability, only a capability to be used in a particular context. So usability cannot be assessed by studying a product in isolation.

There are three potential ways in which the usability of a software product could be measured, according to (ISO 9241):

(a) By analysis of the *features of the product*, required for a particular context of use. Usability could be measured by assessing the product features required in a particular context. However ISO 9241 only gives partial guidance. Of the many potential design solutions compatible with ISO 9241, some will be more usable than others.

(b) By analysis of the *process of interaction*. Usability could be measured by modeling the interaction between a user carrying out a task with a product. However,

current analytic approaches do not give very precise estimates of usability. As the interaction is a dynamic process in the human brain, it cannot be studied directly.

(c) By analyzing the *effectiveness and efficiency*, which results from use of the product in a particular context, and measuring the satisfaction of the users of the product. These are direct measures of the attributes of usability.

So, in general, techniques to measure usability-related factors include (i) inspection methods, (ii) testing methods and (iii) inquiry methods.

For systems including mobile devices, a combination of these techniques is often used. In a survey of evaluation studies involving mobile technology, 71% of the studies have been performed through laboratory experiments, which revealed a tendency towards building systems based on trial and error and evaluating systems in controlled environments at the expense of studying real use of systems. As a result the question of what is useful and what is perceived problematic from a user perspective has not been adequately addressed (Kjeldskov and Graham, 2003). Traditional usability evaluation techniques can hardly be applied in this context as heuristics have not been defined and there is not enough collective background experience from user studies, as the technology is in continuous evolution. In the study discussed in this paper, we apply a process testing method with participation of typical users. Analysis of the collected observations of the evaluation study is performed using an ethnographic tool that permits mixing of multiple sources of observational data, necessary requirement in evaluation studies involving mobile technology, when the users move in physical space and are difficult to track. In the next section we provide a brief description of the system before discussing the evaluation study.

```
<event>
  <time>19 : 43 : 05</time>
  <time2>00 : 15 : 24</time2>
  <user>PDA 2</user>
  <action>15.-G2.-Send image</action>
  <attribute>PDA receiver (number 1).
    Message :Xen_9\Xenopoulos6.jpg </attribute>
  <typology />
  <comments />
  <added_by_user>false</added_by_user>
  <id_event>48</id_event>
</event>
```

Figure 1 – An event of the logfile

3 The Museum Mobile Application

The Museum system under evaluation is based on a client – server architecture. A description of the environment is included in (Simarro Cabrera et al. 2005).

An important characteristic of the application is that the server produces a centralized log file (in XML format) of the actions that take place during the visit and optionally this log file can be combined with video recording of the visit allowing evaluation of the activity during the visit. The format used in this logfile is that defined in the context of the *Collaboration Analysis Tool (ColAT)*, see Avouris et al.

(2004). This feature has permitted using this tool in the analysis of the collected data, as discussed in the following section. In figure 1 an event as collected by the logfile is shown. This event concerns request for exchange of a piece of a puzzle by one user (PDA2) to another user (PDA1).

The client front-end (see fig.2) includes an interface through which the user can obtain information about the rooms in the museum, synchronization functionality for the code of the puzzles as well as for the content used by them. Synchronizing the games automatically allows easy management of the system having to produce the desired update only in one place and the system propagating it automatically on all clients upon first connection.

The clients are very light, in order to save the scarce resources of the PDA. The database access is done via the server that acts as a proxy for cases that the client itself could have been fetching information from the database – this approach offloads some more the client and also has some management benefits with respect to the fact that there is no need to transport over the wireless network the login credentials for the database server and there is no overhead in the settings of the client.

The server supports virtually any number of connected clients being limited only by the hardware and the operating system resources. At every new client connection it spawns a new thread that is in charge with communicating with the specific client connected.

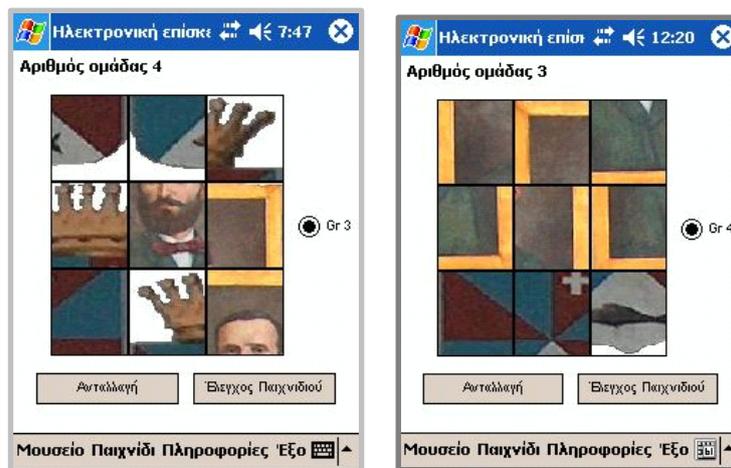


Figure 2 – PDA screens of two of the group members during the image puzzle game

Next we discuss more specifically the characteristics of the two puzzles implemented: the text and image descramble games. Typical client environment for these games is shown in figure 2.

The text Game's goal is to compose correctly a manuscript out of fragments of text. Each group member receives different verses of a poem. Afterwards they should explore the room they are in order to find the manuscript that fits the verses they have received. So, they should attempt first to put in order the sentences they have re-

ceived. The problem is that they will only receive a subset of the verses. So they have to exchange them with the other group members in order to form correctly the poem. Once they receive all the needed verses, they will have to place them in the proper order in order to complete successfully the particular game.

In the Image Game each student has to solve a puzzle. Each one receives some pieces of the puzzle but not all of them, while there are also some pieces of others' images. Due to this, they will have to exchange the pieces between them in order to have all the parts of the image, a necessary step for solving the puzzle correctly. When the users receive the pieces of the puzzle they should look for the item they think it could be the solution of its puzzle. The fact of seeing the solution of the puzzle – that for example can be found in the physical environment like a painting, a picture, statue etc – supports them in solving the puzzle.

Both these collaborative games are initiated in such a way that the pieces are allocated in group members that have mutual interest to exchange them, so that deadlocks are avoided (i.e. a deadlock may occur when a member of the group has pieces of interest to somebody else who has nothing to offer in exchange). However in future more arbitrary distribution of items may be allowed, especially when elder players are involved. In such case more complex problem solving strategies have to be defined by the players if they wish to avoid possible deadlocks for the sake of the overall group performance.



Figure 3 – View of the evaluation study

4 Setting of the Usability Evaluation study

The developed prototype was evaluated by a controlled experiment in a setting that resamples a typical context of use of the system (A Historical/Cultural Museum). A school party of children in their first year of senior high-school (aged 15 to 16, the expected typical user age) participated in the study. The party was made of twelve

(12) children (3 boys and 9 girls) divided in three groups of 2 teams having two members each that were randomly formed. Two of these groups were made only of girls and one group was composed of 3 boys and one girl.

The teams gathered the clues and then each group had to debate and to discover collaboratively what the combined clues were in order to solve the problem.

The experiment was recorded by 3 video and 2 audio recorders for further analysis using the ColAT analysis tool (Avouris et al 2004, Fiotakis et al 2004) which interrelate activity logs video and observers notes in the same environment. So through ColAT the actions that the users performed during the use of the PDAs, that were logged by the server, where synchronized with the videos. During the experiment at least one evaluator observed the behavior of the children. The observers also played the role of the museum guides and they explained to the participants how to use the application and what is the goal to follow in playing the games.

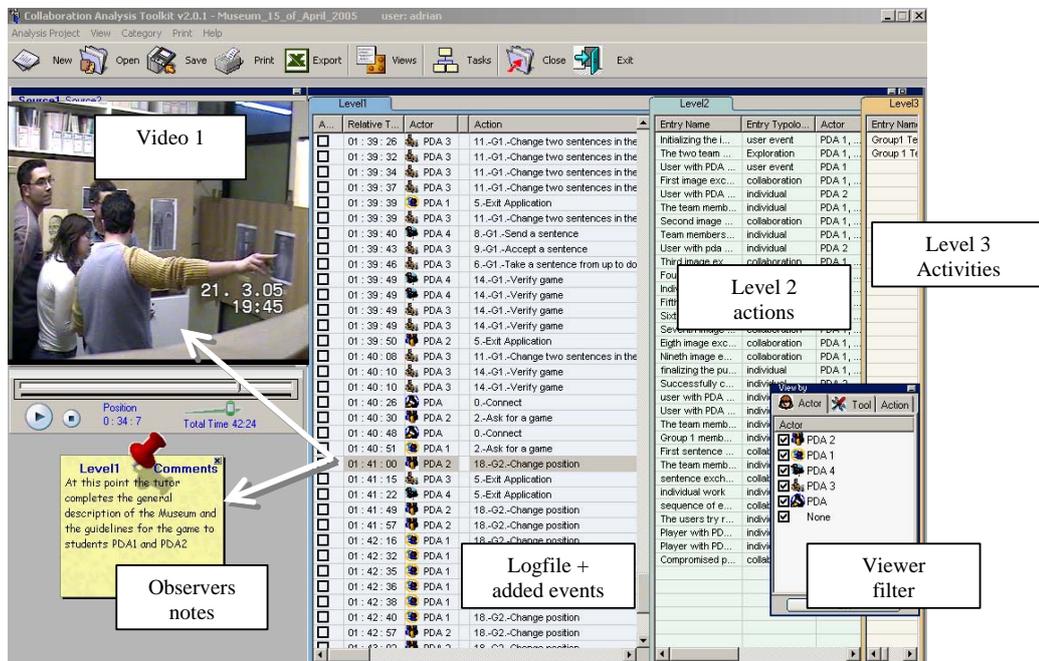


Figure 4 – Analysis environment for the evaluation study: ColAT

5 Findings of the study

There are several types of observations that have been made. In all the teams for every game the log has a similar pattern that shows that the participants were engaged first in exploring at first the interface of the games. After the period of exploration they start the real game play to achieve the desired goal. The interface is very intuitive.

tive considering that none of the participants have used before a handheld, though all of them have mobile phones (so they are used to mobile, small display constrained devices) and over 30% of them have a PC at home.

Observers have noted that the teams that identified the exhibits in the Museum walls and used them as reference completed the games in a much shorter time. Overall it was observed that the PDAs drew most of the attention, as the participants at most used the surroundings just as a means to solve the games and to get the reward in the form of the clue.

The TextGame generally took much longer to complete than the ImageGame as there were no fragments of the text in the physical space, though the poems chosen were well known to the children. As opposed to the ImageGame the TextGame involves some scrolling in order to place the verses in the right spots and it does not support an overview of the text. The children that have not succeeded to solve the text puzzle expressed their wish to have the solution presented in the game – feature that is not present. This feature is not present in the ImageGame either, however the fact that the images can be found in the environment fulfills this requirement.

Also the increased time required to do the TextGame could be related to the fact that higher cognitive load is required, as the users need to remember the sequence of the verses that was even harder due to the necessary splitting of the poem in short phrases.

Table 1. ImageGame Analysis results

G*	P**	Member	ImageGame				
			Time	Events	Exchanges	Finish	
1	1	PDA 1	6'35"	101	9	3	Yes
		PDA 2	6'18"			6	Yes
	2	PDA 3	3'05"	47	3	0	Yes
		PDA 4	3'31"			3	Yes
2	1	PDA 1	4'22"	46 ~	3	2	Yes
		PDA 2	6'26"			1	Yes
	2	PDA 3	2'49"	41	5	4	Yes
		PDA 4	2'56"			1	Yes
3	1	PDA 1	2'29"	47	4	3	Yes
		PDA 2	3'32"			1	Yes
	2	PDA 3	4'59"	71	7	4	Yes
		PDA 4	4'14"			3	Yes
Average:			4'16"	~59			
Average by profile:		1	4'57"	~65	~5		
		2	3'36"	53	5		

* Group, ** Profile

The optimal strategy for the ImageGame needed only three (3) exchanges in order to have both players win status. The optimal strategy for the TextGame was different from one profile to another and needed 4 phrases to be changed over, 2 from one member to the other and 2 in the opposite way for profile 2, and respectively 6 phrases (3 and 3) for profile 1 in order to have both players win status.

Table 2. TextGame Analysis results

G*	P**	Member	TextGame				
			Time	Events	Sentences sent	Finish	
1	1	PDA 1	16'15"	76	6	2	Yes
		PDA 2	14'19"			4	No
	2	PDA 3	3'15"	24	4	2	Yes
		PDA 4	3'17"			2	Yes
2	1	PDA 1	12'27"	57	11	4	No
		PDA 2	11'32"			7	No
	2	PDA 3	3'46"	41	6 ***	3	Yes
		PDA 4	3'29"			3	Yes
3	1	PDA 1	6'08"	29	4	0	Yes
		PDA 2	5'59"			4	No
	2	PDA 3	3'24"	31	4	2	Yes
		PDA 4	3'20"			2	Yes
Average:			7'16"	43			
Average by profile:		1	11'07"	54	7		
		2	3'25"	32	~5		

* Group, ** Profile, *** 5 successful + 1 unsuccessful sentences sent

The results presented in Tables 1 and 2 show us that there is a significant difference between profiles 1 and 2 especially for the TextGame. We can notice a very big difference in the average times by profile required to play the TextGame (profile 1 with 11'07" compared with profile 2 with 3'25") as well as in terms of successful finalization of the game (profile 1 33,33 % compared to profile 2 with 100%).

The explanation for these results lies in the differences in the contents of the two profiles where profile 1 had a larger quantity of text than 2. Also the poem of profile 2 is well known, as it is the first verse of the Greek National Anthem, see Table 4. The larger poem of profile, also resulted in noticeable differences in the interaction. As already mentioned the children were familiar with small screen devices but they were not used with the stylus. From the observations that were made during the experiment turned out that in the TextGame they had difficulties to scroll the text with the stylus and often they tended to miss a phrase when they moved from one scroll page to another. The lesser quantity of text in profile 2 allowed the players to overcome easier these problems while the text in profile 1 made the game harder for them. Some times during the game children felt frustrated and often remarks like "I can not stand it any more!" or "This is impossible!" were heard.

Also we can note that since the verses in the task for profile 1 were less recognizable by the participants they had even more difficulties in completing the puzzle. So none of these teams had the patience to complete in both PDAs the text. In the teams that finished the game they stopped immediately after one of the team members got the corresponding clue. Opposite to this behavior was that of profile 2 where all participants wanted to complete the task on their PDA.

Table 3. Profile differences in the TextGame

TextGame	Well known?	Total lines	Common lines	No of sentences to be send in order for both to win	Words in the text	Characters in the text without spaces
Profile 1	Should be*	13	7	6	28	142
Profile 2	Yes	11	6	4	16	82

*Though profile 1 verses are in the children curriculum the verses of profile 2 have a much higher probability to be known as they belong to the national hymn.

Table 4. Content of the TextGame

<i>Profile 1 text</i>	<i>Profile 2 text</i>
Στων Ψαρών την ολόμαυρη ράχη περπατώντας η δόξα μονάχη, Μελετά τα λαμπρά παλικάρια και στεφάνι στην κόμη φορεί Καμωμένο από τα λίγα χορτάρια που απόμειναν στην έρημη γη.	Απ' τα κόκαλα βγαλμένη Των Ελλήνων τα ιερά, Και σαν πρώτα ανδρειωμένη, Χαίρε, ω χαιρε, Ελευθεριά!

Profile 2 results are very interesting with regard to the times needed for solving the ImageGame and the TextGame. The average time for the TextGame is lower than the average time for the image game, an observation against our intuition. The explanation is that the users got accustomed to the devices during the ImageGame, that was in all teams the first game they have chosen to play. So they had already learned the basic elements of interaction and they got the idea of the games. Repeating the ImageGame resulted in reduced completion times as expected. This behavior is clear in the profile 1 where the times required to complete the TextGame were high compared with the times required to complete the ImageGame for just 2 extra lines of additional text, as the increase of required time seems to be exponentially function of text quantity.

Comparing the average times required by profile 1 for ImageGame with the one required by profile 2 we can see a slight difference that is explained through the fact that the images in profile 2 contained slightly more edges and shapes that eased the completion of the game.

The strategies that the teams adopted for solving the games differed from team to team. The most common was to stick together and sometimes move around to find the relevant picture on the walls of the museum. One exception was in the 3rd group 1st team have exchanged the pieces of the desired image to form and then they split, each partner moving close to the relevant exhibit in order to finalize the puzzle. For the TextGame, profile 1 players only tried to find information in the physical environment that might help them to complete the task with the verses. It is of interest to note that there were teams that adopted an optimal strategy in terms of exchanges. In spite of the fact that other teams had less optimal strategies the times for completion were close. Collaboration between team members was in various forms: the collaboration

patterns expected and provided by the system – exchange pieces of images or text sentences through the game interface; verbal collaboration; helping the team mate to find the relevant information in the physical space as well as some times selecting on PDA the desired piece of information.

6 Conclusions

Design and usability evaluation of the Museum mobile system has been a hard task. Despite the many technical difficulties related to the selected development environment, the application entered a beta testing phase after six months of development and recently the prototype went into the evaluation phase that involves a series of usability evaluation studies like the one described in section 5. While the system goes through testing the approach is also under consideration through different perspectives, among which its educational value, the impact on the museum when deployed, the impact on future visitors and applicability in other similar settings.

Some of the findings concern the activity characteristics: A possible drawback of the proposed activity is related to the fact that we may be building a tour oriented to the PDA instead of the museum, so that it could be possible that visitors interact more with the handheld devices than with the exhibits. In our evaluation study, the visitors indeed used the exhibits rather as auxiliary material towards the main objective, i.e. solving the given problem, however the conditions were simulated and the emphasis of the particular study was clearly on examining interaction during the activity. An additional concern, as discussed in the evaluation section, is related to the fact that it is possible that the games are too difficult for children of the target age group. If children find the games difficult to play they will lose interest in them. However the developed authoring environment has the capability of adjusting the level of difficulty in various ways, so thorough testing of future activities with children of the target age group should be done before introducing the activity in a museum.

Overall the activity is not invasive of the environment, as no visible intervention is required in a museum, while the scenario can be adjusted for many different kinds and sizes of Museums.

Finally the developed prototype is expected to be used for studying various aspects of group interaction through mobile devices in the near future and validate the developed usability evaluation methodology, especially since it is combined with logging mechanism and powerful analysis tools like the ColAT environment.

Acknowledgements

The research reported here has been made possible through funding by the Socrates/Erasmus Mobility Program between the University of Patras and University of Valladolid. Special Thanks are also due to K. Demeti of the Museum of Solomos and Eminent Zakynthians for providing access to the collection in the context of long-term collaboration.

References

1. Avouris N., Komis V., Margaritis M., Fiotakis G., (2004), An environment for studying collaborative learning activities, *Journal of International Forum of Educational Technology & Society* , 7 (2), pp. 34-41, April 2004.
2. Avouris N., (2001), Introduction in *Human-Computer Interaction*, Diavlos Publication, Athens, 2001 (in Greek).
3. Dix A., Finley J., Abowd G., Beale A., (2004), *Human-Computer Interaction*, Prentice Hall,
4. Fiotakis G., N. Avouris, V. Komis, N. Tselios, (2004), Qualitative Data Analysis Tools in the frame of Activity Theory: The case of CoLAT, *Proceedings 4th ETPE Conf.*, Athens.
5. Hsi, S., *The Electronic Guidebook: A Study of User Experiences using Mobile Web Content in a Museum Setting*, IEEE Int. Workshop on Wireless and Mobile Technologies in Education (WMTE'02), 2002
6. ISO 9241-11 International Standard on Ergonomic Requirements for office work with visual display terminals (VDT), Part 11: Guidance on Usability, ISO, 1997.
7. Kjeldskov J., Graham C. (2003), A Review of Mobile HCI Research Methods, *Proc. Mobile HCI 2003*.
8. Kjeldskov J., Stage J., (2004) New Techniques for Usability Evaluation of Mobile Systems, *International Journal of Human-Computer Studies*, volume 60, pp. 599-620.
9. Kusunoki F., M. Sugimoto, H. Hashizume, (2002), Toward an Interactive Museum Guide System with Sensing and Wireless Network Technologies, *IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'02)*
10. Raptis D., N. Tselios, N. Avouris, (2005), Context-based design of mobile applications for museums: A Survey of existing practices, *Proc. MobileHCI 2005, Salzburg, 2005..*
11. Seppälä P., H. Alamäki, (2002), Mobile Learning and Mobility in Teacher Training, *IEEE International Workshop WMTE'02*
12. Simarro Cabrera J., Muñoz Frutos H., Stoica A.G., Avouris N., Dimitriadis Y., Fiotakis G., Demeti Liveri K., (2005), Mystery in the Museum: Collaborative Learning Activities using Handheld Devices, *Proc. MobileHCI 2005, Salzburg, September 2005*.
13. Schneiderman B. , (2003), *Designing the User Interface*, Addison Wesley.
14. Tzu-Chien Liu, Hsue-Yie Wang, Jen- Kai Liang, Tak-Wai Chan, Jie- Chi Yang, (2002), Applying Wireless Technologies to Build a Highly Interactive Learning Environment, *IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'02)*
15. Yi-Chan Deng, Ming-Zhang Do, Li-Jie Chang, Tak-Wai Chan, (2004), PuzzleView: Enhanced Workspace Displaying for Group Interaction with Tablet PCs, *2nd IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'04)*.
16. Yatani K., M. Sugimoto, F. Kusunoki, (2004), Musex: A System for Supporting Children's Collaborative Learning in a Museum with PDAs, *2nd IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'04)*