THE BOOK OF ELLIE: AN INTERACTIVE BOOK FOR TEACHING THE ALPHABET TO CHILDREN

Eleni Papadaki[1], Xenophon Zabulis[1], Stavroula Ntoa[1], George Margetis[1], Panagiotis Koutlemanis[1], Polykarpos Karamaounas[1], Constantine Stephanidis[1,2]

[1]Foundation for Research and Technology – Hellas, Institute for Computer Science (FORTH – ICS)
[2]Computer Science Department, University of Crete
{elpap, zabulis, stant, gmarget, koutle, karam, cs}@ics.forth.gr

ABSTRACT

Augmented Reality environments have shown to be relevant and valuable in many instances of the educational process. Accounting for the advantages and conventional gains of learning through physical books and printed matter in general, this paper presents an approach towards augmenting both such media. This work has elementary school as a context and presents an approach towards augmenting a physical book and associated learning cards, with the purpose of providing a playful approach to learning the alphabet. The two principal activities involved in studying from an elementary school book are augmented: learning, during which the student receives information about letters, phonemes, and words, and practicing where questions are asked to the young student in order to consolidate the recently acquired knowledge. The proposed implementation is evaluated initially as to its performance and accuracy and then as to its usability and suitability for efficient and intuitive interaction.

Index Terms— Augmented reality book, natural interaction, educational software, tangible interface

1. INTRODUCTION

Augmented Reality is an emerging technological trend that brings technology enhancement in daily activities of people, while AR is gradually gaining focus both in the research community and the industry.

Using printed books versus e-books is an active topic of research with several pertinent studies and research efforts, that reveal the beneficial aspects of each medium (e.g., [1], [2], [3], [4], [5], [6], [7]). This work, aiming to provide an approach that contributes to the learning process, proposes the combination of the benefits of paper-based learning with those of electronic editions and Augmented Reality. Furthermore, the motivation for playfulness in the interaction targets the deep engagement of children in the learning process and, thereby, the efficiency of learning.

The contributions of this paper are as follows. First, a method for the context-based augmentation of a school book and associated learning cards is proposed, focusing on the task of learning the alphabet. Then, a method for natural hand interaction with printed books and cards is proposed, eliminating the need of input devices; this transparent interaction facilitates student focus on the learning process. Finally, an evaluation of these methods in the task of alphabet learning is provided.

The remainder of this paper is organized as follows. Related work is reviewed in Section 2. In Section 3, interaction with the proposed system is overviewed, while the technical components that enable this interaction are presented in Section 4. Evaluation experiments are reported in Section 5. Finally, Section 6 concludes this work and provides directions for future research.

2. RELATED WORK

Paper augmentation was introduced in the 90's by HP's "DigitalDesk" [8] and its successor "EnhancedDesk" [9]. Both these systems offered users the ability to interact with physical paper via touch, while related information was presented near the paper. More recently, pertinent research efforts were intensified, leading to more sophisticated systems that augment physical books in more realistic approaches [10].

In the field of interacting with physical paper the majority of approaches make use of special artefacts like digital pens, haptic devices etc. For example "Paper++" [11], uses a special paper printed with non-obtrusive patterns to determine the page and the position in the page that the pen touches, while "PapierCraft" [12] uses pen gestures on the paper in order to recognize users' commands, such as active reading, or copy and paste information. Another example is "SESIL" [13], which offers recognition of book pages, stylus-based indication, as well as simple handwriting and gesture recognition. The "Mixed Reality Book" [14] is another example of augmented book content, page recognition and interaction via pen, based on markers printed on each page. Finally, Intelligent Paper [15] refers to augmenting foldable prints such as newspapers, maps and pocket books, using a customised pen. This paper proposes a novel hand-based interaction approach, eliminating the need for any special equipment, fostering the natural process of pointing-while-reading, which is usually followed by young readers.

Over the past few years, teaching methods have adopted serious games as a means of increasing the efficiency of learning methods. Educational games based on physical cards (no electronic media or computer involved) have been studied in this work in the context of assessing the educational effects of the approach. Such examples include the "Voyager: Satellites" card game [16] and the "Trading Card Games" [17], whose prominent characteristic is that of using illustrated cards to captivate the learner interest and reinforce his/her motivation. Purely electronic card games (played on the computer screen) have also been studied in the same context, such as "Binary Numbers" [18], where students learn the binary system by playing a variation of the "Blackjack" card game against the computer. In the domain of Augmented Reality, the "Educational tabletop mini games" [19] are played with physical cards, which are visually recognized, and have been developed in the context of an Ambient Intelligence classroom [20], [21]. In this work, physical cards are visually recognized but, also, their spatial arrangement as produced by the learner becomes a means of user interaction.

3. SYSTEM OVERVIEW

The "Book of Ellie" is the augmented version of a classic schoolbook for teaching the Greek alphabet to primary school children. The book introduces the alphabet letters and combinations of them by increasing the difficulty level. For each letter or letter combination, relevant images and text involving the specific letter(s) are provided. The short stories for each letter are structured around dialogues and activities of a Greek family, with the protagonist being Ellie, one of the four children. In the augmented version of the book, Ellie has become an animated character, constantly available to assist the young learner by reading phrases from the book, asking questions or providing advice.

The Book of Ellie consists of two main modes, the *Book* and the *Game mode,* which users can interchange at any time during interaction. When in Book mode, the system recognizes the currently open book page and displays its electronic version. The child can interact with the physical book by pointing with its finger at any letter or text passage of interest and listen to its pronunciation. In Game mode, the child is prompted to answer a series of test questions on the learning material, by placing cards on the table.

The system setup, as shown in Fig. 1a, consists of a television screen (32'') for visual and audio output, an "Asus Xtion Pro" RGBD camera, and a PC running the software. The RGBD camera is used to recognize and localize book pages and cards, as well as detect and localize fingertip contacts on the book and table. The physical book and paper cards (e.g., depicting letters, simple objects, or animals) are interactive components of the system.

4. COMPONENTS

In this section, a detailed overview of the two modes of interaction and their technical implementation is provided.

4.1 Book mode interaction

In this mode, the child is engaged in learning letters and their combinations by reading and listening appropriate text passages and viewing related images for each letter. The young learner interacts with the physical book by turning its pages and, also, pointing at the printed letters and texts in order to listen to them. Each text snippet of the physical book is associated with a native speaker audio recording, played each time the learner places its finger upon it.

In more details, a vision process (see 4.2) first detects the book and then displays the currently open pages on the nearby screen. Every time a page is turned in the physical book, the electronic page on the screen changes accordingly. Thereby, an electronic representation of the appearance of each page is required, and is typically acquired from the PDF version of the book. For each page of the book, particular areas of interest (usually letters or phrases) have been specified, triggering a specific event when the child's finger points at them. Touch recognition is achieved using the depth image of the RGBD camera (see 4.2).

Initially, Ellie is waiting for the user to touch a phrase on the physical book page. When this occurs, the phrase is highlighted on the screen and Ellie reads it while pointing towards it (see Fig. 1b). Each phrase is transformed into a series of lip-synched visual mouth phonemes, providing Ellie's speaking animation with a realistic appearance. In this way, phonetic pronunciation is understood, casting the application ideal for alphabet learning.

4.2. Interaction with books

A book recognition module detects the presence of known books on the desktop, recognizes them, and estimates their location and orientation on the desktop coordinate frame. This is achieved by using the recognition system described in [13]. Recognition is based on a priori available digital images of each book page. This system additionally provides the location and orientation of the book.

Using these location and orientation estimates, the boundaries of the book are predicted. The corresponding portion of the depth map is then extracted and transformed to an orthocanonical view with respect to the book's coordinate frame. Using this information, a height map of the book's surface is generated. Each pixel of this map shows the distance of the corresponding book surface location from the workspace surface.

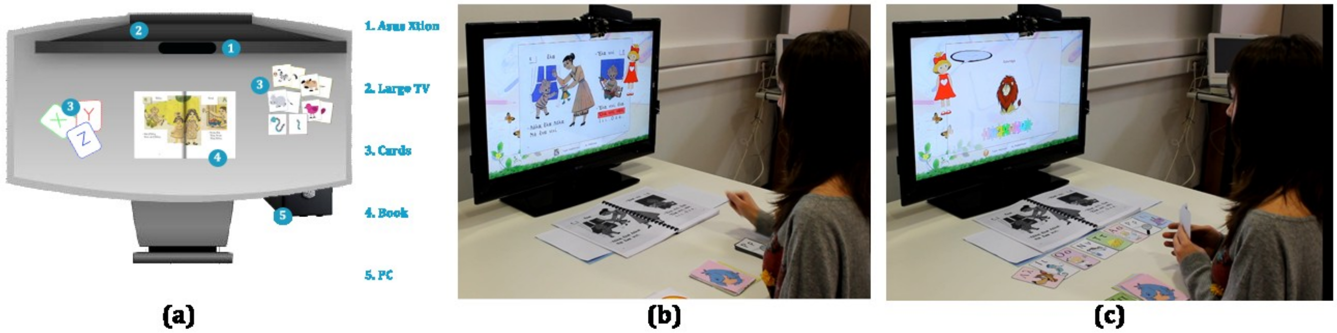

Fig. 1. (a) System setup (see text); (b) Book mode interaction: the user has just pointed at a phrase in the physical book, which is highlighted in the screen as Ellie reads it aloud; (c) Spell-the-word card game

When the user points at a location on the book, his hand creates an increment in the height map. The contact point is detected by finding the location where this increment starts to occur, that is, where a discontinuity in the, otherwise smooth, book surface is detected. This location is found by scanning the columns of the height map iteratively, looking for the topmost height discontinuity. In iteration i, the median height for the topmost k_i pixels of each scanline is computed. For each scanline, the height at the k_i–th pixel from the top is compared with the median value for this scanline. If a deviation larger than threshold q_L (10mm) from the median is observed, a contact point is assumed to be found. In order not to spuriously detect as in contact a finger that hovers above the page, this deviation is required to be less than q_H (17mm). If no contact is found, k_i is increased by 1 and the method proceeds to the next iteration. The contact point is tracked through Kalman filtering to reduce jitter and recover from tracking failures.

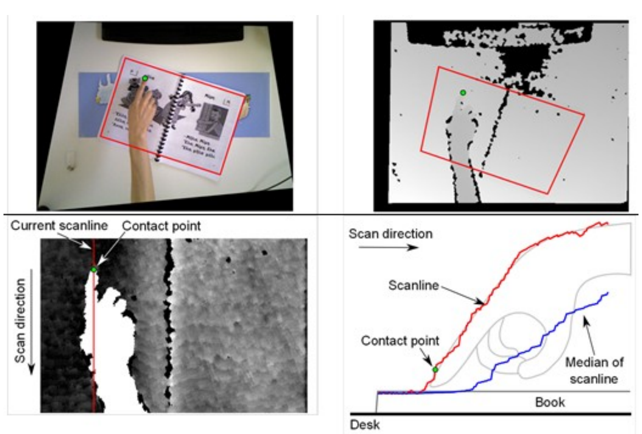

Fig. 2. Book recognition and contact detection. Top left: Estimation of the book boundaries in camera image. Top right: The portion of the depth map used to generate the height map. Bottom left: The generated height map, displaying a top view of the book. Dark pixels are closer to the workspace surface than lighter ones. Red line indicates the scanline used to generate the bottom right image. Bottom right: Schematic overview of the contact detection method, from a side view. The red curve corresponds to the red scanline from the bottom-left image and represents distance from the workspace surface. The blue curve corresponds to the median of the scanline. As the scan progresses from left to right, more samples from the scanline are included in the median computation.

The book locations estimated in camera coordinates are transformed into the desk's coordinate frame, through a homography that maps locations on the planar surface of the desk to plane of the acquired image. This homography is estimated at system setup through a calibration process that employs a checkerboard as the calibration object. The checkerboard is placed on the desk and aligned with it constituting a physical reference to the desk's coordinate plane. Camera lens distortion is also estimated during this calibration and the acquired images are, correspondingly, rectified.

The fingertip contact locations are provided directly in the, intrinsic, coordinate frame of book page. This is achieved through the recognition process [13] that establishes point correspondences between the acquired and the a priori stored image of the recognized book page. These correspondences are employed to estimate another homography. This homography maps fingertip contact locations to the stored representation of the recognized book page.

4.3 Game mode interaction

Swapping to Game mode, the child is introduced to an educational card game which acts as a recapitulation of the letters that have been taught and a teaching tool for the spelling of some basic words. Mode swapping is achieved by placing special purpose (utility) cards on the desk, one for each mode (see Fig. 3a and 3c). The child can activate a mode any time and resume its interaction from where it was left, i.e., resume reading from the last page that was open or continue answering the last question that was not successfully completed.

Fig. 3 Utility cards (a) Switch to Book mode, (b) Next question (c) Switch to Game mode.

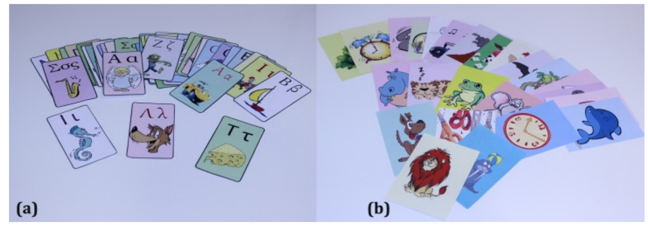

Fig. 4 Learning cards (a) letter (b) picture (objects & animals).

During the Game mode, Ellie asks the child 24 questions while visual feedback is provided on the screen. In more details, there are two types of questions which are randomly chosen: (a) find the picture that begins with a given letter and (b) spell the word of a given picture, using letters from a pack of letter-cards.

There are 44 letter-cards (some letters exist more than once) and 24 picture-cards (see Fig. 4) at the user's disposal, allowing children to answer questions by placing them on the table. For the cards' recognition, their digital representation is stored in a database, along with their dimensions and a unique id. When a card is placed on the table, a recognition algorithm matches it with its digital form. The visual process not only recognizes a card and returns its unique identification number, but also estimates its orientation and position in relation to other cards on the table.

In the first question category, feedback is provided by Ellie both through speech and text, to assist the learner by suggesting to place another card on the table, or to place just one card (in case the child leaves two or more cards on the table). When the depicted object or animal is successfully identified, Ellie provides confirmation feedback, while a characteristic sound is also heard (e.g., a telephone ringing, a lion roaring, etc.), and the system advances to the next question.

Feedback is more intricate in the case of "spell the word" questions. Visual feedback is provided by displaying the letters in the order in which the corresponding letter cards have been placed on the table, while letters are represented as individual puzzle pieces. Additional feedback is provided by Ellie regarding the correctness of the answer. In more details, the following checks are performed: accuracy of the number of cards placed on the table (i.e., if there are missing or extraneous cards), correctness of the number of the letters placed, and cards' order appropriateness. Once the requested word is correctly spelled, Ellie congratulates the player, an animation showing the individual puzzle pieces joint is displayed (see Fig. 1c), and the player proceeds to the next question.

After all questions have been asked (one per alphabet letter), the game ends and the player's score in a letter grade scale (A, B, C) appears. If, during the game, the player has difficulty to answer a specific question or does not wish to, the question can be skipped by placing a special purpose card on the table (see Fig. 3b).

4.4 Interaction with cards

This module detects and recognizes the cards that appear on the table at any given time instant. In addition, it characterizes the spatial arrangement by which the cards are laid out on the table, in order to detect whether they are in a linear spatial arrangement and properly oriented.

Card recognition and localization works in a very similar manner as the detection and localization of the book on the table. The system described in [13] is again employed, but now for the a priori available images of the cards. As in Section 4.2, the location and 2D orientation of the recognized cards is provided on the table's coordinate frame.

To reason whether the detected cards are (approximately) arranged on a straight line on the table, the following approach is adopted. Let the set of the detected cards, which is required to contain at least two elements. The elements of this set are sorted based on their distance from a corner of the table, into the ordered set C_S. Let the jagged line that connects the centers of the cards in C_S. The length, let L_T, of this line is the sum of distances of each element's center from the center of its succeeding element. Let the line segment that connects the centers of the first and last element of C_S and L_S the length of this segment. The arrangement is considered as linear if L_T is approximately equal to the L_S ($L_S / L_T < 0.8$) and, also, if the distances of all centers from L_S are below threshold (2cm).

Finally, the orientation of cards is required to be compatible with that of the line. The cards are considered to be properly oriented if their orientation is not more than 20 degrees different than that of the line. In this way, arrangements that include misaligned cards (e.g., upside-down) are not recognized as valid.

5. SYSTEM EVALUATION

In the context of system assessment, a performance and accuracy evaluation, as well as a preliminary usability evaluation have been conducted, as reported in the next sections.

5.1. Performance and accuracy

In [13], the book page and card recognition systems were tested to recognize book pages with a 100% confidence in books of approximately 200 pages, in a variety of orientation and occlusion conditions. The worst case of occlusions covered ~50% of a page. It was observed that approximately a third of each page must be visible to be recognized.

To assess the system's suitability as an interactive touch display, the accuracy of fingertip contact detection was measured. The experiment was performed by 5 users that were naive to the experimental hypotheses. A checkerboard that was printed upon a planar sheet of paper was placed upon the desk and aligned with it. A checkerboard detector and a method for extrinsic calibration of the camera were employed, using the OpenCV library, to detect the checker corners and estimate the relative pose of the camera to the checkerboard. The 3D locations of the checker corners were the estimated and considered as ground truth. Users were instructed to touch with their finger the inner corners at a particular order, while the system was acquiring images. The error was quantified as the distance between the 3D touch estimate and a 3D corner. Each user touched a total of 70 dots upon the entire desk. The average error was ~0.8cm in the center of the desk and ~1.2cm in the periphery. The reason for this variability is that peripheral regions of the desk are imaged more obliquely (and, thus, by less pixels per unit area) by the depth camera. It was concluded that the accuracy of fingertip detection is sufficient as long as the smallest contact areas are double the average detection error.

5.2. Usability

Four usability experts participated in a heuristic evaluation [22] of the system, aiming to identify and eliminate as many problems as possible, before proceeding to testing the system with children. Three of the evaluators had experience regarding the usability of interfaces for children, one due to her engagement with designing and evaluating such interfaces, and two of them as parents of young children. The number of evaluators employed is appropriate in the context of a heuristic evaluation, since between three and five evaluators have been found to identify a reasonably high percentage of the usability problems (between 74% and 87%) [23]. Furthermore, if double specialists (i.e., usability experts with knowledge regarding the specific type of interfaces) are employed [23] their error finding performance is even better.

The evaluators were explained the purpose of the system, as well as the main interaction concepts and were left to explore the system by themselves. During each evaluation session two more individuals were present: one facilitator noting down the evaluator's comments, and the chief developer aiming to provide assistance in case of a system error or to explain the functionality of the system if asked.

The evaluators found in total 28 issues that were classified in three categories: General, Book mode, Game mode. The issues identified by each evaluator were aggregated in one report, having duplicates removed. Then, they were rated by each evaluator according to their severity on a scale from 1 (aesthetic problem only) to 4 (usability catastrophe) and the average score for each problem was calculated in order to prioritize issues that should be immediately eliminated. The collected data are shown in Fig. 5.

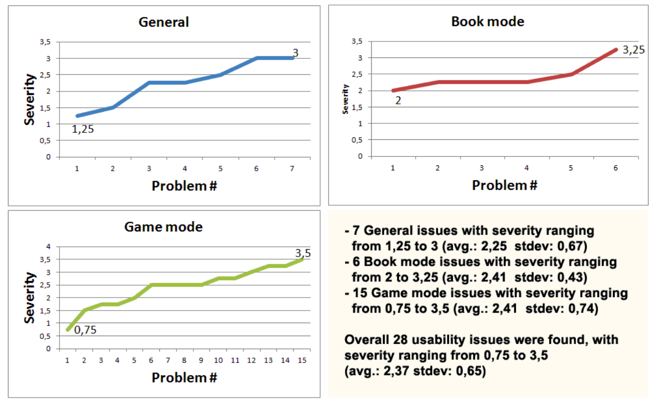

Fig. 5 Results of heuristic evaluation, overviewed per category.

The overall comment of the evaluators was that the system is very interesting and in general well designed for the target audience. The evaluators have identified some issues that need to be addressed and suggested improvements in order to enhance the overall user experience. One of the most important findings was that a help functionality guiding the child on how to interact with the system is required, since the interaction techniques may be novel to the child. Such guidance could include highlighting interactive book areas at long idle times and providing assistance when actions beyond the expected ones are carried out (e.g., the child closes the book, the child throws on the table a different card than expected, such as a letter card during the find the picture game, etc).

Furthermore, most of the evaluators agreed that the feedback provided by the system needs to be enhanced by adding audio messages and visual prompts in certain specific occasions, e.g., for changes of the currently active mode, during the game play to indicate the number of remaining questions, as well as in correct and erroneous answers during the cards game. In relation to the last concern, the feedback providing method was suggested to be reconsidered so as to provide messages more promptly, and be enhanced with hints in case the learner has serious difficulties in answering the question.

Finally, one concern raised was that the total number of cards may be too large for a young child to manage. This, however, is a concern that remains to be answered during user testing with primary school students.

6. CONCLUSIONS AND FUTURE WORK

This paper has presented Ellie's book, combining both physical (book and cards) and electronic media in the context of an educational application, aiming to teach elementary school students the alphabet. The current work has been evaluated in terms of performance, accuracy and usability. The evaluation indicated that the system's recognition algorithms perform sufficiently and that its design is appropriate for the target user group. Useful conclusions were derived about features of the system that could be improved, before proceeding to test with children.

Future work includes user testing with young students, initially with a small yet sufficient number (e.g., 15 – 20 students) in order to assess the usability of the improved system. Next, a large-scale experiment involving a large number of children with various learning profiles will be carried out, aiming to assess not only the overall user experience, but also the learning outcomes of the proposed system. To this end, the system will be supplemented with additional curricula, such as numbers and foreign language learning.

ACKNOWLEGEMENTS. This work is supported by the FORTH-ICS internal 'Ambient Intelligence and Smart Environments' RTD Programme. Authors thank Eftychia Varoucha for recording Ellie's voice.

7. REFERENCES

[1] W. Sutherland-Smith, "Weaving the literacy Web: Changes in reading from page to screen", The reading teacher (2002): 662-669

[2] K. O'Hara, and S. Abigail, "A comparison of reading paper and on-line documents." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 1997

[3] C. Spencer, "Research on learners' preferences for reading from a printed text or from a computer screen.", The Journal of Distance Education/Revue de l'Éducation à Distance 21.1 (2006): 33-50

[4] M. C. Murray, and P. Jorge, "E-Textbooks are Coming: Are we Ready?", Issues in Informing Science and Information Technology 8 (2011): 49-60

[5] O. Korat, and S. Adina, "Electronic books versus adult readers: Effects on children's emergent literacy as a function of social class", Journal of Computer Assisted Learning 23.3 (2007): 248-259.

[6] M. T. de Jong, and A. G. Bus, "How well suited are electronic books to supporting literacy?", Journal of Early Childhood Literacy 3.2 (2003): 147-164

[7] S. Davis, and C. Pearman, "Early readers and electronic texts: CD-ROM storybook features that influence reading behaviors". Reading Teacher 58: 446–454, 2005.

[8] P. Wellner, "Interacting with paper on the DigitalDesk", Commun, ACM 36, 7 July 1993, pp. 87–96

[9] M. Kobayashi, and H. Koike, "EnhancedDesk: integrating paper documents and digital documents", 1998, Asia Pacific Computer Human Interaction (APCHI'98). IEEE CS, pp. 57–62

[10] M. Billinghurst, H. Kato, I. Poupyrev, "The MagicBook: Moving Seamlessly between Reality and Virtuality", IEEE Computer Graphics, pp. 6-8, 2001

[11] P. Luff, C. Heath, M. Norrie, B. Signer, and P. Herdman, "Only touching the surface: creating affinities between digital content and paper". Proceedings of the 2004 ACM conference on Computer Supported Cooperative work, ACM, New York, NY, USA, 523–532

[12] C. Liao, F. Guimbreti, K. Hinckley, and J. Hollan, "Papiercraft: a gesture-based command system for interactive paper", ACM Transactions on Computer-Human Interaction (TOCHI) 14, no. 4 (2008): 18

[13] G. Margetis, X. Zabulis, P. Koutlemanis, M. Antona, C. Stephanidis, "Augmented interaction with physical books in an Ambient Intelligence learning environment", Multimedia Tools and Applications, Jan. 2012

[14] R. Grasset, M. Billinghurst, A. Dünser, H. Seichter, "The Mixed Reality Book: A New Multimedia Reading Experience", April 28-May 3, 2007 San Jose, CA, USA.

[15] M. Dymetman, M. Copperman, "Intelligent Paper", Int. Conference on Electronic Publishing, Document Manipulation, and Typography, 392-406, 1998.

[16] D. R. Smith and E. Munro, "Educational card games", Physics Education 44(5):479, 2009.

[17] P. Chen, R. Kuo, M. Chang, J.-S. Heh, "Designing a Trading Card Game as Educational Reward System to Improve Students' Learning Motivations", Transactions on Edutainment III (2009): 116-128.

[18] M. Kordaki, "A computer card game for the learning of basic aspects of the binary system in primary education: Design and pilot evaluation", Education & Information Technologies, 16(4):395-421, 2011.

[19] M. Korozi, A. Leonidis, G. Margetis, P. Koutlemanis, X. Zabulis, M. Antona, and C. Stephanidis, "Ambient educational mini-games", In Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 802-803. ACM, 2012.

[20] A. Leonidis, G. Margetis, M. Antona, and C. Stephanidis, "ClassMATE: enabling ambient intelligence in the classroom", World Academy of Science, Engineering and Technology, 66: 594 – 598.

[21] Leonidis, A., Korozi, M., Margetis, G., Ntoa, S., Papagiannakis, H., Antona, M., Stephanidis, C. A Glimpse into the Ambient Classroom. Bulletin of the IEEE Technical Committee on Learning Technology, 14(4): 3-6, 2012.

[22] J. Nielsen, and R. Molich, "Heuristic evaluation of user interfaces", SIGCHI conference on Human factors in computing systems: Empowering people, 249-256, 1990

[23] J. Nielsen, "Finding usability problems through heuristic evaluation", SIGCHI conference on Human factors in computing systems, ACM, 1992.