Advances in Human-Computer Interaction

Interactive Play and Learning for Children

Guest Editors: Adrian Cheok, Hiroshi Ishii, Junichi Osada, and Owen Noel Newton Fernando

Interactive Play and Learning for Children

Advances in Human-Computer Interaction

Interactive Play and Learning for Children

Guest Editors: Adrian Cheok, Hiroshi Ishii, Junichi Osada, and Owen Noel Newton Fernando

Contents

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 954013, 3 pages
doi:10.1155/2008/954013

Editorial

Interactive Play and Learning for Children

**Adrian Cheok,[1,2] Hiroshi Ishii,[3] Junichi Osada,[4]
Owen Noel Newton Fernando,[1] and Tim Merritt[1]**

[1] *Mixed Reality Lab, Interactive and Digital Media Institute (IDMI), National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077*
[2] *Graduate School of Media Design, Keio University, Kanagawa-ken 223-8522, Japan*
[3] *MIT Media Lab, School of Architecture and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA*
[4] *NEC Design, Ltd., Shinagawa-Ku, Tokyo 141-0031, Japan*

Correspondence should be addressed to Adrian Cheok, adriancheok@mixedrealitylab.org

Received 24 November 2008; Accepted 24 November 2008

1. Introduction

One of the most socially and culturally beneficial uses of human computer interaction research is enhancing play and learning for children. It is very important to understand the needs of children and craft visionary interactive systems designed to enhance education and entertainment. In this special issue, we can study papers that provide visions and glimpses into the future of human computer interaction for children's play and learning.

Since the research findings of Piaget [1], educators have valued the activity of play as an essential stage for the development of the human mind. Through trial and error and a "learn by doing approach," children internalize concepts of complex physics, abstract logic, and social rules among others. Caillois [2] described man as essentially an organism that requires play. The positive psychology approach would also agree that this is especially critical in promoting the human to not only develop but truly thrive. It is very important to understand the needs of children and craft visionary interactive systems designed to enhance education and entertainment and provide those experiences and activities which allow the child to emerge as an empowered learner with confidence and inspiration.

2. Guide to the Special Issue

The focus of this special issue is to provide visions and glimpses into the future of human computer interaction for children's play and learning. It also discusses achievements on different technical and theoretical development of interactive play and learning for children. The special issue received an overwhelming number of submissions from researchers around the world focusing on various topics ranging from user-centered design to advanced technical developments. Considering originality, technical competence, user evaluation, and balance of topics, thirteen papers were accepted for the special issue. If we look at all of the papers as one body of works, it makes sense to divide them into three chapters according to the main focus. There were four focused on designing for children with disabilities, five focused on methodological issues of design for children, and four focused on specific technologies focused on supporting children in learning through play activities.

2.1. Design for Disabilities

Apart from being a most worthy application of research, it is a very well accepted strategy to focus on the design for people with disabilities as a way to improve designs for all users. By designing for users with specific needs, there are usability issues and needs that emerge which face the designer which are then transferrable to all users regardless of their abilities.

The paper by van Rijn and Stappers presents guidelines for designers of interactive systems for children with autism. These were developed during their LINKX project in order to design a language-learning toy for autistic children. Theoretical and practical implications for each guideline are presented through the experiences of the prototype development process so that other designers can use them to inform their design decisions.

The paper by Tuominen et al. presents research of their PROAGENTS multimodal learning environment, which supports blind children in explorative learning experiences with astronomy. The system utilizes haptic and auditive interaction techniques in the proactive agents that support and guide children, deepen their explorations, and discover the central concepts and other phenomena.

The paper by Hengeveld et al. presents research focused on children with multiple disabilities (e.g., both a cognitive and a motor disability) and stimulating the adult-child communication to decrease functional limitations. LinguaBytes is presented, which stimulates language development through interactive and adaptive play and learning in an environment incorporating tangible objects and multimedia content. The research is based on interactive storytelling and anchored instruction. An outline of the development of the LinguaBytes play and learning environment from the earliest studies, up to the current prototype, CLICK-IT, is provided.

The paper by Jacob and Mythili has developed a text-to-speech (TTS) system specifically targeting children of various abilities including the vocally disabled. Their system utilizes linguistic analysis instead of the signal processing techniques used in other existing concatenative TTS systems. The paper presents the results of a user study, informing their development and giving insights to the challenges and needs of such a system.

2.2. Methodology

The advancement of the methodological tools for the field of HCI research and for designers in general represents a fundamental way of sharing knowledge and the blueprints for successful outcomes. Papers in this section recognized the creativity of the child as a designer, user-centered design, and the development of new multimodal environments.

The paper by Nousiainen and Kankaanranta presents their research experiences in user-centered and participatory design while in collaboration with elementary school children in the design of learning environments, based on three projects and requirements gathering techniques. Through this research, issues related to design collaboration with children emerged, especially in terms of the children's feeling of ownership over the final outcome. The children's contribution yielded outcomes which were useful in informing the design of the user interface as well as the contents of the learning environments.

The paper by Abeele and Zaman presents theoretical framework and practical case for designing likeable interactive media applications for preschoolers. In this article, they have demonstrated how the uses and gratifications paradigm together with expectancy-value theory could fill in one of the blanks in designing likeable applications for preschoolers via the research project CuTI (Cuddle Toy Interface). The CuTI project aims at revealing those particular user gratifications and design attributes that are important to support playful behavior and fun activities of preschoolers in the home environment.

The paper entitled "Child-centered evaluation: broadening the child/designer dyad" by Pardo et al. discusses challenges in literal interpretation of user-centered design for children. They have explored the certain hard-to-reach or difficult-to-interpret situations such as evaluation of educational software intended to be used by children. The paper also argues that the added value that teachers can bring to the evaluation of educational software can improve the diagnostic power of the evaluation and its sensitivity to pedagogical issues.

The paper by Motoyoshi et al. proposes a novel strategy of designing play equipments and detailed explanation of the strategy through an example of a designing process of playing environment where the players' usage of the play equipments cannot be foreseen. This enables designers to interpret the playing environment as an information channel in terms of channel theory, which is also an appropriate loose constraint for designing a play environment.

The paper by Birchfield et al. introduces theoretical research from HCI and Education that reveals a convergence of trends focused on embodiment, multimodality, and composition. Authors argued that there is great potential for truly transformative work that aligns HCI and education research, and posit that there is an important opportunity to advance this effort through the full integration of the three themes into a theoretical and technological framework for learning.

2.3. Specific Applications and Technology

Four of the papers clearly introduced innovative uses of technology in the creation of interactive systems for learning and play. From the systems which inspire creativity to the systems which assist in the mastery of games of skill, there are lessons to be learned and inspiration to be drawn from these cutting edge systems.

The paper entitled "BlogWall: social and cultural interaction for children" by Cheok et al. is an extension of the existing text messaging to a new level of self-expression and public communication, combining visual art and poetry. By selecting proper elements, the application can be effectively used to build awareness among children about diverse cultural elements of humans as global beings while they interact and play with the system. The results of users' studies indicated that the application has a high level of social and engagement aspects.

The paper by Kerdvibulvech and Saito proposes a vision-based method for tracking guitar fingering position and actions by guitar players. A new framework for tracking colored finger markers is presented which integrates a Bayesian classifier into particle filters. The resulting application made from this system assists guitar learners by allowing them to automatically identify if they are using the correct chords required by the musical piece. This brings HCI research and applies engineering solutions to enhance the creativity and learning of guitar playing.

The paper by Uematsu and Saito presents visually enhanced sports entertainment applications: AR baseball presentation system and interactive AR bowling system.

Main advantage of these applications is that they do not require any special device such as positioning sensors or a high-performance PC because applications can be run on the tabletop in the real 3D world only with a web camera and a handheld monitor connected to a PC. Demonstration showed that children really enjoyed these applications because their actions in the real world affect the virtual world.

The paper by Uchiyama and Saito presents a pool supporting system with a camera-mounted handheld display based on augmented reality technology. Main objective of this research is to estimate and analyze the distribution of balls and to present visual aids without special equipment such as a magnetic sensor or artificial markers. Children are able to capture a pool table from an arbitrary viewpoint and see supporting information and ball behavior which are drawn on the captured images through the display. Results of the experiment showed that the accuracy of estimated ball positions is enough for analysis of ball arrangement.

3. Conclusion

We appreciate the opportunity to gather together a fine collection of academic papers which present a picture of the current landscape of the design of interactive and digital systems supporting learning and play for children. We hope that this special issue will inspire the researchers of today and tomorrow to create a better world promoting the learning and play activities of children of all ages.

Adrian David Cheok
Hiroshi Ishii
Junichi Osada
Owen Noel Newton Fernando
Tim Merritt

References

[1] J. Piaget, *The Construction of Reality in The Child*, Basic Books, New York, NY, USA, 1954, translated by Margaret Cook.
[2] R. Caillois, *Man, Play and Games*, Free Press of Glencoe, New York, NY, USA, 1961.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 639435, 8 pages
doi:10.1155/2008/639435

Research Article

The Puzzling Life of Autistic Toddlers: Design Guidelines from the LINKX Project

Helma van Rijn and Pieter Jan Stappers

ID-StudioLab, Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, 2628CE Delft, The Netherlands

Correspondence should be addressed to Helma van Rijn, h.vanrijn@tudelft.nl

Received 30 September 2007; Revised 17 April 2008; Accepted 24 July 2008

Recommended by Adrian Cheok

This paper presents guidelines for designers to help them consider what children with autism value in interactions with their environment. The guidelines were developed during the LINKX project in order to design a language learning toy for these children and are based on literature study, expert interviews, generative techniques, and prototype testing with users. We present both the theoretical or practical background of each guideline together with a discussion how the guideline was evident in the prototype of LINKX. Testing the prototype in the real world helped us to shape the prototype and the guidelines. This paper aims to share our guidelines with the design research community, so that others can use them as steppingstones in their work.

1. Introduction

Interactive toys are opening up new learning and play opportunities for children with autism. For example, the toys can keep track of the child's progress, evolve with a child, or give controlled feedback. But for this to succeed, the design of these toys must fit well within the experiential world of these children. Children with autism live in an experiential world that is different from that of other children or adults. Therefore, designers need to make effort in to understand these children's needs and preferences.

Although an admirable body of theory on autism is available in popular science (see, e.g., [1, 2]), much of it is not helpful for designers. Moreover, absorbing literature is often too time-consuming for design teams. We see two important ingredients that do work for designers: contact with the children to achieve empathic understanding, and guidelines on an operational level to help them translate that understanding into product ideas and concepts.

In the LINKX project, we developed guidelines ourselves in order to design a language-learning toy for children with autism. These guidelines, implemented in the design, were acquired through study and contextual research, such as expert interviews, generative techniques, and prototype testing with users. The prototype and the research techniques have been described [3–5]. In this paper, we want to share our guidelines with the design research community, so that others can use them as stepping-stones in their work.

2. Background of Autism

Autism is an inborn developmental disorder that affects around 91 people in every 10 000. Much variation exists between autistic children, and even with one child the diagnosis can change over time. Affected children may display a range of disabilities at many levels, such as impairment in social relationships, communication, and imagination [6]. Therefore, it is best to use the term "autism spectrum disorder" (ASD), although we use the term "children with autism" in this paper. The disorder distinguishes a number of classes, including pervasive developmental disorder-not otherwise specified (PDD-NOS), Aspergers Syndrome (sometimes referred to as "high-functioning autism"), and classic autism.

2.1. Puzzling Your Perceptions Together

The neurological setup of the autistic brain makes children with autism process information differently from children

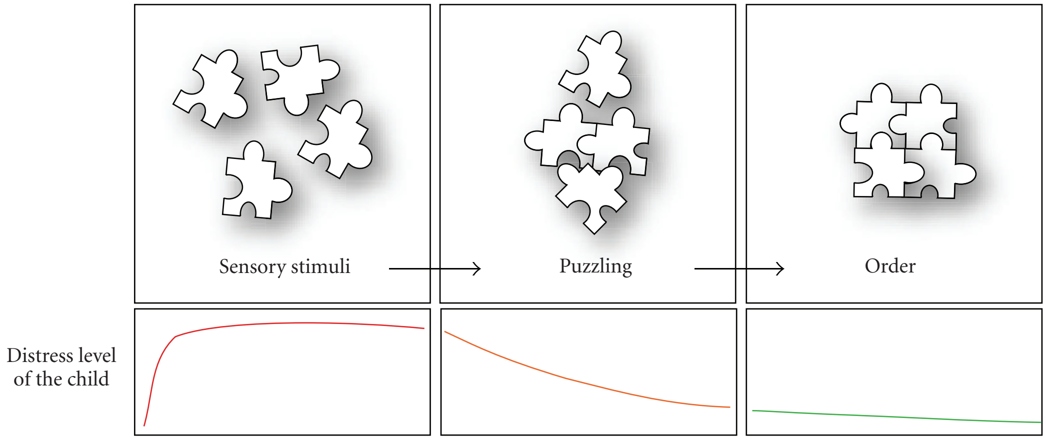


Figure 1: Children with autism need to puzzle their perceptions together to create order. Before puzzling, the child feels distress. After puzzling, the order makes the child feel at ease.

with typical development. Thus, children with autism experience sensations differently and also have difficulties with understanding these sensations.

Unlike most people, children with autism perceive sensations, such as sound, smell, and pain, as incoherent fragments. They do not automatically integrate different sensations into a meaningful whole as most people do. Therefore, they can be very sensitive to certain sensations, and insensitive to others. For example, a child can feel distress because of sticky materials, while it does not notice pain at all.

These children also struggle with sensemaking. Temple Grandin, an adult with autism who could not speak as a child, describes in her book her difficulty to come up with new categories [7]. When she was little, she used to categorize cats and dogs by their size. After her neighbours got a small dog, she had to come up with a new feature to make the distinction, which became the nose. Children with autism need to consciously make categories for everything they hear, see, smell, or feel. Each time something needs to be processed, the child has, in the words of de Bruin [8], to "puzzle it together." The time needed for this puzzle can vary from seconds to months. For example, one mother tried to teach her child to say the word "cookie." After trying it for weeks, she finally gave up. Surprisingly, the child said the word himself a month later. Apparently, it took the child that long to incorporate the word. This process of puzzling your perceptions together is visualized in Figure 1. For children with autism, it is important to eventually complete these puzzles in order to get some peace of mind.

This overall explanation about how children with autism process information is important for understanding the behaviour, needs, and feelings of these children. Most research on autism has brought forward classifications [9, 10], biological facts, or theories [11, 12]. Although this is valuable, it does not sufficiently inform and inspire designers. Designers mainly need to know and understand the behaviour and feelings of the children when they interact with "things" around them, such as toys and people. We deliberately categorised people as "things," because that is how many children with autism regard them. They have difficulty in relating to people. Most of us, including designers, have difficulty in imagining what this must be like, because we are used to conceive our environment in largely social terms.

2.2. The Three Boys Involved in the LINKX Project: Beer, Robbert, and Jakob

Three parents chose to cooperate in the LINKX project in response to a letter in which we asked for their expertise. The parents and their children (see Figure 2) were involved throughout the whole design process. These boys helped us to gain a general view on autism, because they differed in diagnosis, age, and language level. Although one child was easier to make contact with than the others, all of them could be triggered to try something new. From the children and their caregivers, we learned how creating order is a central activity in their lives. This is their only way to deal with the world, because of their difficulty processing information. The children will serve as examples for many of the guidelines.

3. A Toy Designed for Autistic Children: Learning Words with LINKX

Before going into depth about guidelines on how to design for autistic children, we first introduce the concept design LINKX, depicted in Figure 3 [3]. LINKX was designed as an interactive toy to help children learn their first 100 words, such as "door," "table," and "cabinet." The toy consists of speech-o-grams and interactive blocks. Before play starts, a caregiver prepares the toy by speaking words into the speech-o-grams. For example, a parent records the word "cabinet" in a speech-o-gram and attaches it to the cabinet. Next, children can start to play by linking blocks to speech-o-grams or to other blocks. Each time a block connects, it changes its lights and speaks the word. For example, when a child links a block with the speech-o-gram described above, the word "cabinet" moves into this block. When ready, this block plays the word

Figure 2: A short introduction to the three children in the LINKX project.

Figure 3: A presentation of LINKX, the concept design based on guidelines discussed in this paper. By playing with blocks, children cause the blocks to play the word they received from objects in their everyday environment. For example by connecting a block to the speech-o-gram attached to the cabinet, the block says "cabinet."

"cabinet." Each speech-o-gram has its own colour, which travels with the word, to help children to predict the result of making a new link. In this way, children can explore names of objects in their everyday environment by playful interaction. The LINKX project website contains a movie that shows the concept in use by the three involved children [5].

LINKX employed the guidelines, which are described in detail in the next section as follows. First, the fixed location of the speech-o-grams *structured* the child's environment. Explicit structure is important for children with autism. The *sensorial quality of feedback* such as the lights changing their position in the blocks and colour was fascinating for them. This is not only because the children enjoy sensory stimuli such as sound and light so much, but also they enjoy *the predictability* of when these stimuli appear each time after they made a link. For that reason, they liked to link the play elements together *again and again*. The blocks confirmed over and over that *the child was in control.* The children showed they had *an excellent memory for sensory detail*, because days after playing with the prototype, they still remembered the speech-o-gram's colours. Also during their play, the children remembered what word was inside a block with the help of the colour.

4. Guidelines on How to Design for Children with Autism

The attention to designing for children with autism is rather new. Earlier efforts have focused on the disorder itself, with an emphasis on curing a disease rather than improving the way the children can function with such an

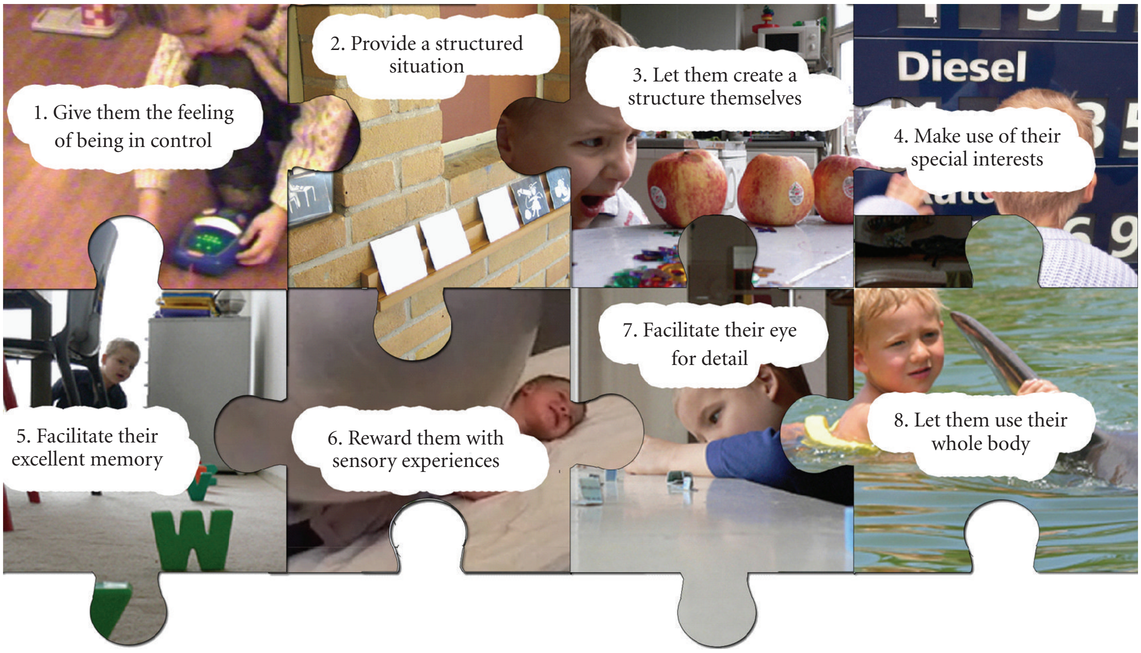


Figure 4: An overview of the design guidelines on how to design for children with autism. This puzzle is unfinished on purpose, to indicate that we expect more insights will follow in the coming years, and want to invite others to think along and help to extend it.

impairment. Approximately thirty-five years ago, researchers and designers started to realize the importance of using technology for their treatments (see, e.g., [13, 14]). Moreover, scientists and practitioners realize it cannot be cured at present. This indicates a need for guidelines on operational level to come to products for these children.

Figure 4 depicts an overview of our guidelines as an open-ended jigsaw puzzle, inviting others to think along and help to complete it.

In this section, we will explain each guideline, or puzzle piece, by giving theoretical background and examples from practise combined with how we implemented this in LINKX, the interactive toy we presented earlier. The concept of LINKX was designed on the basis of the guidelines. By means of building the prototype and testing it with the children we could evaluate our guidelines. Therefore, we present the guideline combined with LINKX examples. To illustrate this, the following figures show for each guideline a pair of puzzle pieces, in which the left piece was taken from Figure 4, the right piece from the design of LINKX.

4.1. Give Them a Feeling of Being in Control

Children with autism often have no idea how to make sense of their surroundings. Their way of processing information makes them feel they have little or no control over the situation. For that reason, the children enjoy interactions that make them feel in control. In the literature, this preference of children with autism is described in several cases (see, e.g., [10, 15]). Beer illustrates this preference with his alphabet toy computer. For example, he likes to press the letter "B" and hear it say "*bee*" over and over. On other days, he prefers the letter "A" or "C." He adores the lights and sounds that are triggered at his command. The same goes for his little number computer, which is depicted in Figure 5. This example shows how a product makes him feel in control by means of direct feedback, given at the position of the trigger and immediately afterwards. In that way, children can predict the effects of their actions. By triggering action over and over, the children reassure themselves that they are in control of the toy.

Figure 5: A sense of control. (a) Beer playing with his little number computer: by pushing the button he makes it count from one to ten. (b) Beer playing with LINKX: by linking a block to the speech-o-gram he makes it say the word "fishbowl."

In LINKX, the children felt in control when linking blocks. As a result of each link, a word and its associated colour travel from one block to the other. Direct feedback was given by light and sound. The little smile on their faces showed they were happy when the blocks turned on their lights and played a sound at their command. Beer and Robbert even broke out in laughter. The blocks clearly made

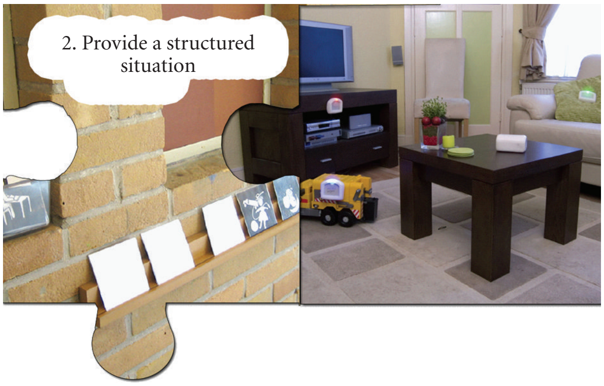


Figure 6: Structure: (a) a day-planner at a school for autistic children; activities are planned in time by means of pictograms. When an activity is completed, the pictogram is flipped over in order to provide structure. (b) An illustration of LINKX in a living room; the game is structured by means of speech-o-grams that have a fixed position, word, and colour.

Figure 7: Organising: (a) Beer loves to make sequences of objects; here you see Beer making a sequence with apples. (b) Robbert playing with LINKX; he loves to make blocks into the same colour and arrange them in groups.

them feel good about themselves. The importance of giving them the feeling of being in control became clear on the occasional moments the prototype gave a delayed or wrong reaction. For example at one time, the colours of two speech-o-grams interchanged. A red speech-o-gram attached to the kitchen cabinet became green, while the green one on the bucket became red. Jakob immediately sensed this loss of control and started to cry in distress. He refused to play again with the toy during that session. This example shows the importance of consistent and predictable product behaviour for these children.

4.2. Provide a Structured Situation

Children with autism have difficulty changing from one activity into another [16]. Their way of processing information makes it difficult for them to know the order of activities or what will happen in what situation. Activity schedules have proven to be successful (see, e.g., [17, 18]). Visual prompts, such as the day-planner depicted in Figure 6, stay visible at all times. The day-planner provides a structure, because the child is informed about the activities in time. When an activity is completed, the corresponding pictogram is flipped over, indicating it is time for the next activity. This structure helps them to both learn actions such as the sequence for going to the toilet, as well as providing a safe and structured environment in which a child can learn.

In LINKX, a structured situation is provided by means of the speech-o-grams (see Figure 6). These speech-o-grams are labels on which a parent can leave a visual prompt (e.g., pictogram or written word) and an audio prompt (spoken word). Both the presence of speech-o-grams on objects and the predictable result from each link structure the play situation. The fixed location, fixed content, and fixed colour of speech-o-grams give the children peace of mind. However, to challenge children to learn new words, parents can change the position, content, and colour of one or more speech-o-gram(s), whenever the child is ready for it.

4.3. Let Them Create a Structure Themselves

Not only do the children enjoy a structured situation, they enjoy creating a structured situation themselves as well. They like to play with toys they can organize, such as the game “memory” or jigsaw puzzles, and especially love to complete those games. A missing piece can cause panic. The children enjoy looking for similarities and differences as well. The children create structure by making differences explicit. Another thing they like to do is arranging objects in space. For example, Beer loves to arrange objects, such as apples in a sequence (see Figure 7), and looks for differences and similarities in that as well.

In LINKX, we tapped into their preference to create structure, because the blocks and speech-o-grams can be seen as a three-dimensional puzzle. By connecting the blocks to the speech-o-grams, the corresponding word travels from that speech-o-gram into the block. Moreover, the blocks can be organised nicely into a line or a group of blocks varying in colour. Robbert for example liked to make collections of blocks in one colour. Over and over, he made them all green, then blue, and red, and put them nicely together (see Figure 7). Beer on the other hand, preferred to reserve one block for the colour red, and another one for green. His structure lay in appointing one colour to one specific block.

4.4. Make Use of Their Special Interests

Some children develop special interests to create structure. Focusing on specific aspects of perception helps the children to make sense of the things they see, feel, and hear. For example, Jakob has special interest in colours, Robbert in sound (music in particular), and Beer in numbers (see Figure 8). These elements are everywhere and the children experience them as important regularities in daily life. These special interests can serve as a means to elicit them to interact with something and learn.

As expected, Robbert paid a lot of attention to the sounds (or words) the blocks produced. Although LINKX’ colours elicited Jakob to play, it distracted him from learning. He

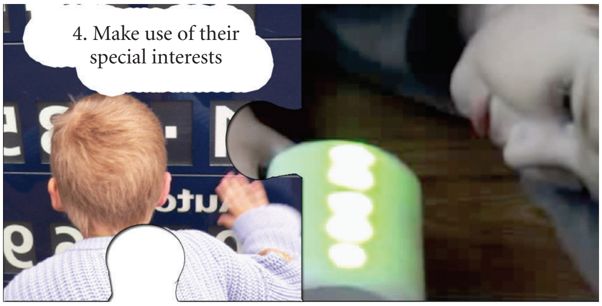


Figure 8: Special interests: (a) Beer's special interest is numbers. They are everywhere and provide him a safe feeling. (b) Jakob's special interest in colours encourages him to play with LINKX, although it hinders him from learning words.

Figure 9: An excellent memory: (a) Beer playing with his alphabet letters on the floor; he knows its sequence by heart, so he can lay the alphabet from "A" to "Z" and from "Z" to "A." (b) Beer plays with LINKX; he remembered the green lights belong to "fishbowl" and the red ones to "window."

was so focused on all the beautiful colours that he could not pay attention to the words LINKX aimed to teach him. His parents suggested it would be best for Jakob to use only one colour for all speech-o-grams, such as white. In that way, he could focus better on the sound. So, special interests can both help and hinder children in their play.

4.5. Facilitate Their Excellent Memory

Children with autism have a much better (visual) memory than children with typical development. An example that illustrates their excellent memory is the fact that Beer can arrange the alphabet from "Z" to "A" in a few seconds (see Figure 9). We have to "think backwards" but he just "sees" it. If our designs make use of their memory skills, we may help them understand more and learn better.

We could explore what Robbert remembered, because he speaks in small sentences and gives answers to questions. For the other two children in the LINKX project, this was impossible. After Robbert played with LINKX the second time, we removed the speech-o-grams and the blocks and put it away. Next, his father asked him whether he knew the colours of the speech-o-grams, which were attached to different objects, such as a cat litter box, kitchen door, carpet, and bike. Surprisingly, Robbert not only knew the colours of objects he had just played with, but he also knew the colours belonging to objects from the week before. For example, when his father asked, "what colour belongs to the bike?," Robbert answered, "the bike is red!." However, he not only remembered all of the colours, we discovered he could also name the gender of the voice that spoke each word. Both his mother and father had recorded words in LINKX. Although he did not recognize his parents from their voices, he did remember the gender of the person that had spoken the word. When asked, "who is the bike," he answered, "the bike is a mister."

Figure 10: Sensory experiences: (a) Beer gets rewarded with deep pressure during speech therapy. The speech therapist rolls the ball roughly over his body. This makes him feel relaxed and happy. (b) Beer playing with LINKX; he likes to look at the coloured lights from very close and laughs out loud.

4.6. Reward Them with Sensory Experiences

Children with autism often enjoy specific sensations. For example, Robbert is sensitive to sounds. The low pitch of a truck that drives through his street distresses him. However, he loves to put a little vibrating toy, which is meant to be looked at, against his ear to hear a low-pitched sound. Jakob's mother said, "stimuli are less disturbing when they are expected. It helps a lot when the child is in control over the stimuli." Although these children can be highly sensitive, or insensitive, to stimuli, they truly enjoy sensory rewards, such as sounds, music, and vibration, and deep pressure, as described by Grandin [7]. Figure 10 shows how Beer enjoys the deep pressure of a big ball being pushed onto his whole body.

The occasions when the prototype malfunctioned clearly exposed the children's sensory preferences and attention. For Robbert, LINKX' sounds were very important. When he linked a block to a speech-o-gram on the cat's litter box, the block's lights turned red, but the sound was not produced. After a moment, Robbert instructed the block to "say litter box." For Jakob, LINKX' colours were very important. Each time a block changed its colour, Jakob enthusiastically pronounced the colour "red," "green," or "blue." He did not mind the two times that the sound failed.

Figure 11: Eye for detail: (a) Beer investigates pieces of paper on the table. (b) Robbert counts five red lights on top of the blocks instead of the number of red blocks on the floor.

Although it seems that Jakob did not learn anything from playing with LINKX, since he focused on the colours, we are not sure. For a long time, it seemed Beer focused only on the lights as well, but later on, he surprised us, when we unexpectedly heard him mutter the word “fishbowl” as he walked towards the fishbowl. Maybe this could be the case for Jakob as well. Some children need more time to puzzle than others.

4.7. Facilitate Their Eye for Detail

Part of the children's excellent memory is their great eye for detail. When something changes, even a tiny detail, an autistic child will often notice immediately, while our blunt senses overlook it completely. For example, Robbert has a detailed view of books, CDs, and DVDs. When his father adds a new DVD to the row of DVDs, he notices immediately, only by looking at the sides of the covers. Also Beer likes to investigate things in detail (see Figure 11). He notices the smallest details around him.

While playing with the prototype of LINKX, Robbert's eye for detail was again clear when his mom asked him how many “red's” there were. Robbert immediately said, “five.” His mom answered, “no, I see one, two, three blocks.” Now Robbert counted, “one, two, three, four, five,” pointing at the red lights shining on one block instead of counting the three red blocks lying on the floor. Jakob also counted the lights on top of the blocks out loud.

4.8. Let Them Use Their Whole Body

Children with autism explore the world, just as other children, through play. By letting the children use their whole body, the learning activity becomes a multisensory unity of action, perception, cognition, and emotions. Hengeveld et al. [7] list more advantages to tangible interaction for children with multiple disabilities, such as more room for social interaction, a more personal interaction style, a slower pace, and a more active interaction. This multisensory unity is visualised in Figure 12 with a picture of Beer during his dolphin therapy in Florida. During the activity itself, he learned how to interact with the dolphin and got rewarded by the sensory experience of the warm water, involving action, perception, cognition, and emotions.

Figure 12: Whole body: (a) Beer swims with a dolphin in Florida. (b) Jakob plays with LINKX on the kitchen floor together with his sister.

LINKX provided a break from conventional computer-assisted language learning, which is mainly screen-based with words linked to drawn pictures of generic objects. LINKX shows how interactive toys can bring word learning to the real familiar environment of the child, involving the whole body, multiple senses, and motor actions. When Jakob played with LINKX, his sister naturally joined his play, because there were more blocks available. On the kitchen floor, she showed him there were more speech-o-grams around him, providing other important learning aspects, such as imitation and turn-taking.

5. Conclusions

These guidelines were developed on the basis of literature, our own and other people's experience and the development of the LINKX prototype in particular. Especially the real world helped us to shape the prototype and the guidelines. For example, the prototype tests showed us the importance of giving the opportunity to adjust the toy to a child's specific preferences or needs. The aim of this paper is to share these guidelines with the design research community, so that others can use them as stepping-stones in their work leading to more tangible toys for children with autism. We encourage others to use and improve our guidelines, and formulate new ones.

The interest of parents, pedagogues, and psychologists in LINKX shows the value of these guidelines, both to drive design and to open up discussion. This specific example shows how computers can help children to learn in other ways than in front of a computer screen, as is done in the CD-ROM language training games that work so well for children with typical development. By embedding computer intelligence into physical toys, several aspects can help the children enjoy their play in a predictable way. First, cause and effect are immediate at the place of action. Second, the toy refers to real objects in their physical familiar environment instead of (symbolic) representations in a generic drawing of an environment. Third, play takes place in a safe environment, their homes, involving their bodily actions and social interactions. These aspects help the children to learn in a way that makes sense to them and should be exploited more in interaction design for these children.

We hope this paper provides some insight into the experiential world of autistic children. In the long term, further research can give more certainty about the causes of autism, and how learning and play can be facilitated for these children. We learned that personal contact with the children was important for understanding and empathy. We hope the guidelines can provide a bridge between the theories and design research community. Moreover, we hope to inspire more designers with this story of Beer, Robbert, and Jakob, who often surprised us with their smart and inventive ways of dealing with the world.

Acknowledgment

The authors wish to thank the pedagogues, parents, and especially Beer, Robbert, and Jakob, who willingly and enthusiastically took part in the LINKX project.

References

[1] T. Grandin, *Thinking in Pictures*, Vintage, New York, NY, USA, 1996.

[2] O. Sacks, *An Anthropologist on Mars*, Vintage, New York, NY, USA, 1995.

[3] H. van Rijn, *LINKX a language toy for autistic toddles, developed in co-creation with parents and pedagogues*, M.S. thesis, Delft University of Technology, Delft, The Netherlands, 2007.

[4] H. van Rijn, "The LINKX project," 2007, http://studiolab.io.tudelft.nl/vanrijn/LINKX.

[5] H. van Rijn and P. J. Stappers, ""Codesigning LINKX": a case of gaining insight in a "difficult-to-reach user group"," in *Proceedings of International Association of Societies of Design Research (IASDR '07)*, The Hong Kong Polytechnic University, Kowloon, Hong Kong, November 2007.

[6] The National Autistic Society, 2000, http://www.nas.org.uk/.

[7] B. Hengeveld, R. Voort, H. van Balkom, C. Hummels, and J. de Moor, "Designing for diversity: developing complex adaptive tangible products," in *Proceedings of the 1st International Conference on Tangible and Embedded Interaction (TEI '07)*, pp. 155–158, Baton Rouge, La, USA, February 2007.

[8] C. de Bruin, *Geef me de vijf. Een praktisch houvast bij de opvoeding en begeleiding van kinderen met autisme*, Graviant, Doetinchem, The Netherlands, 2004.

[9] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Publishing, Arlington, Va, USA, 4th edition, 2000.

[10] N. Parés, A. Carreras, J. Durany, et al., "Promotion of creative activity in children with severe autism through visuals in an interactive multisensory environment," in *Proceedings of the 4th International Conference on Interaction Design and Children (IDC '05)*, pp. 110–116, Boulder, Colo, USA, June 2005.

[11] S. Baron-Cohen, *Mindblindness*, MIT Press, Cambridge, Mass, USA, 1995.

[12] L. M. Oberman and V. S. Ramachandran, "The simulating social mind: the role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders," *Psychological Bulletin*, vol. 133, no. 2, pp. 310–327, 2007.

[13] K. M. Colby, "The rationale for computer-based treatment of language difficulties in nonspeaking autistic children," *Journal of Autism and Developmental Disorders*, vol. 3, no. 3, pp. 254–260, 1973.

[14] T. R. Goldsmith and L. A. LeBlanc, "Use of technology in interventions for children with autism," *Journal of Early and Intensive Behavior Intervention*, vol. 1, no. 2, pp. 166–178, 2004.

[15] K. Dautenhahn, "Design issues on interactive environments for children with autism," in *Proceedings of the 3rd International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT '00)*, pp. 153–161, Sardinia, Italy, September 2000.

[16] D. J. Cohen and F. R. Volkmar, *Handbook of Autism and Pervasive Developmental Disorders*, John Wiley & Sons, New York, NY, USA, 1997.

[17] L. C. Bryan and D. L. Gast, "Teaching on-task and on-schedule behaviors to high-functioning children with autism via picture activity schedules," *Journal of Autism and Developmental Disorders*, vol. 30, no. 6, pp. 553–567, 2000.

[18] K. L. Pierce and L. Schreibman, "Teaching daily living skills to children with autism in unsupervised settings through pictorial self-management," *Journal of Applied Behavior Analysis*, vol. 27, no. 3, pp. 471–481, 1994.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 387076, 13 pages
doi:10.1155/2008/387076

Research Article

Proactive Agents to Assist Multimodal Explorative Learning of Astronomical Phenomena

Eva Tuominen,[1] Marjatta Kangassalo,[1] Pentti Hietala,[2] Roope Raisamo,[2] and Kari Peltola[3]

[1] *Unit for Early Childhood Education, Department of Teacher Education, University of Tampere, 33014 Tampere, Finland*
[2] *Tampere Unit for Computer-Human Interaction TAUCHI, Department of Computer Sciences, University of Tampere, 33014 Tampere, Finland*
[3] *Academic Development Unit, University of Tampere, 33014 Tampere, Finland*

Correspondence should be addressed to Eva Tuominen, eva.tuominen@uta.fi and Marjatta Kangassalo, marjatta.kangassalo@uta.fi

Received 1 October 2007; Revised 15 August 2008; Accepted 28 October 2008

Recommended by Adrian Cheok

This paper focuses on developing, testing, and examining the Proagents multimodal learning environment to support blind children's explorative learning in the area of astronomy. We utilize haptic, auditive, and visual interaction. Haptic and auditory feedbacks make the system accessible to blind children. The system is used as an exploration tool for children's spontaneous and question-driven explorations. High-level interaction and play are essential with environments for young children. Proactive agents support and guide children to deepen their explorations and discover the central concepts and relations in phenomena. It has been challenging to integrate together in a pedagogically relevant way the explorative learning approach, proactive agents' actions, haptic perception's possibilities, and the selected astronomical phenomena. Our tests have shown that children are very interested in using the system and the operations of the agents.

1. Introduction

Today's children are immersed in a world of play. Much of their everyday activities center around different kinds of digital facilities designed to bring them joy through playing. The dominating role of entertainment affects their relation to computers as well as their learning. In this paper we emphasize exploratory learning, in addition to the playful elements, in order to support children's conceptual learning and thinking. With this, we strive for a combination where entertainment is not so dominant, in line with the "hard fun" (instead of "soft fun") concept of Papert [1]. To accomplish this, we utilize multimodal interaction with novel devices and proactive software agents to provide just-in-time assistance for both sighted and blind children.

The exploratory learning approach can be examined by studying the phenomenon that will be the subject of investigation, the exploration activities themselves, and the learning environment where human and computer support is emphasized. The main aim of exploratory learning is to make it possible for children to understand more deeply the phenomenon in question so that they can predict and explain the phenomenon and enrich their knowledge, ultimately achieving conceptual change if needed. This means that children's exploratory activities should not be restricted only to hands-on activities, but rather aim at understanding the underlying theory and central abstract concepts as well as their relationships within the phenomenon. Attention should be directed to the way children explore the phenomenon and what kind of exploration strategies they use. Essential to a fruitful inquiry process is the ability to pose questions, construct explanations and hypotheses, and to search for new knowledge. The meaningfulness of the tasks and children's interest are central starting points for such an inquiry process. However, for the development of exploration strategies and in order to grasp the central ideas behind the observed phenomena, children often need support for their explorations. Other children and adults play important roles in supporting knowledge construction. Materials, tools, and technological equipment can also help to structure the inquiry process.

To support children's exploratory and conceptual learning of astronomical phenomena we constructed a multimodal learning environment that is based on dynamic haptic, auditive, and visual interaction. Haptic feedback provides the users with an additional presentation channel, making the system accessible for blind children, and increasing the level of engagement also for sighted children while they are studying the astronomical phenomena. The construction of the computer environment is based on the notion that children could use the system as an exploration tool for their spontaneous and question-driven explorations. Attention is paid to the support which the children might need in their explorations. Support in this computer environment is implemented through built-in proactive agents that aim to create situations that stimulate children's active and constructive inquiry processes. Agents provide personalized support for children, try to anticipate the problems they might have, and suggest pedagogically relevant new approaches.

In the following, the starting points and theoretical principles of designing and constructing the multimodal learning environment will be examined in more detail. Of specific interest is the introduction of pedagogical bases from the perspective of exploratory and inquiry learning in which human and computer support plays an important role.

2. Astronomical Phenomena as Content

When selecting the natural phenomena for the proactive computer simulation system of the research in question it was essential that the phenomena were important and significant with regard to life. The simulated phenomena have to awaken sufficient interest in the children and efficiently utilize possibilities offered by the new computer technology. Some of the most suitable phenomena are those that can be easily and illustratively presented in no other way, such as phenomena linked with space, the atmosphere, and many other phenomena in elementary physics. There has to be a coherent theory of the phenomenon, which in turn has to include important, central, and abstract concepts. A final selection criterion for the chosen natural phenomenon can be seen as its conformity to scientific principle which should form a clear, well-organized knowledge structure and theory. These aspects present a strong and defined foundation for the computer simulation (see, e.g., [2]).

In this research, the phenomena chosen for computer simulation were our solar system, the interrelations between the Earth and the Sun, the Earth, the atmosphere, and the interior layers of the Earth. Although the phenomenon as a whole is abstract, it is possible to concretize it by using the computer-based simulation. The core properties and features of the phenomenon can be built into the computer-based simulation, which then will assume a concrete form in the simulation run. The selected phenomena have also been modeled so that visually impaired and blind children can easily investigate the phenomena using a haptic device.

Earlier studies on children's conceptual learning in the area of astronomy have shown that children often face difficulties when they are learning these phenomena. This is because the early observations children make of their natural surroundings, and from which they construct their understanding, do not often correspond to the scientific view. The studies have shown that children's initial conceptions about the physical world are very deep rooted and difficult to change by means of instruction (e.g., [3, 4]). In the area of astronomy, children's naïve, everyday conceptions have been widely investigated and are already well documented. Children's knowledge about the Earth, especially, with regard to the Earth's shape, is one of the most studied phenomena in this area (e.g., [3, 5–7]). For example, Vosniadou and Brewer (e.g., [7]) have found that children may start with an initial model of the Earth, in which the earth is a flat object. This model is supposed to be based on children's everyday observations of their surroundings. Later, children may form "synthetic models." These models are formed when children try to combine the scientific knowledge, received, for example, from instruction, into their intuitive models. The synthetic models Vosniadou has found include the flattened sphere model (people live on flat parts of a spherical earth), the hollow sphere model (people live inside the spherical earth), and the dual earth model (a flat earth where we live and a spherical one in the sky). According to Vosniadou, to adopt the scientific view of the spherical Earth, children often need to revise their existing knowledge structure and to change some of the entrenched assumptions or beliefs in which the concept of Earth is embedded. The process of restructuring is slow and gradual, and since children try to retain as many of their assumptions as possible or to change them only partially, misconceptions are very likely [8, 9]. Ikospentaki and Vosniadou [10] have included blind children in their studies, and found that blind children also have initial or alternative models of the Earth although they lack empirical, optical data from their surroundings.

Earlier research on children's conceptual thinking has included phenomena such as the alternation of day and night, seasons, and phases of the moon (e.g., [5, 11, 12]). In all of these phenomena, a number of intermediated notions between the initial impressions and scientific notions have been discovered (e.g., [5]). Although the more naïve views showed reduction as the age increased, misconceptions still seemed to persist in many students up to 16 years of age.

In sum, these studies show that children may have various notions with regard to the shape of Earth and other astronomical phenomena, and that these notions are very persistent and difficult to change. They are also connected to one another. For example, Vosniadou et al. [9] have stated that the spherical shape of the Earth is not possible to understand without knowledge about gravity. Furthermore, to understand the scientific explanation of the day/night cycle, children must first understand that the Earth is a sphere. The spherical shape of the Earth and gravity are also prerequisites for understanding the spherical arrangements of layers inside the earth [9]. These findings have supported and provided background for the selecting of the microworlds for the simulation.

3. Explorative Learning and Play

Contemporary findings from cognitive science studies have demonstrated the importance of the individual's or learner's own activity and interest in learning. For example, based on the knowledge and results concerning the plasticity and self-organizing of the human brain, it can be said that the more complicated the phenomenon or matter in question, the more the learner's own active explorative action, like analyzing and organizing the material, is needed for understanding and comprehending. Up to certain limit, when the optimal level of attention and activation prevails, the self-organizing and shaping of brain takes place and models are constructed in the brain. Through perceptions, experiences, and active mental work, the memory traces corresponding to these models grow and strengthen, and knowledge structures are formed (see, e.g., [13–15]). Furthermore, when considering the acquisition of expert skills in some area, the organization of domain-specific knowledge structures and the learner's own active involvement and interest become even more important.

Playing has a significant role in young children's learning. Children are often naturally interested in various phenomena around them and ask questions and explore their surroundings. Many scientific concepts, knowledge, and skills develop throughout the early years through the child's active exploration and play. The explorations start with very concrete and small things, for example, when a small child drops things out of his/her pram. Soon this turns into a game where, after dropping, the child waits an adult to pick the toy up. Even very young children know where to look for the toy after dropping it; they look down. Through their experience they have learned a theory about the world around them; that is, when you release your hold of a thing it falls (on the ground) [16]. Later on, the children's explorations can take an even broader form as they take place in social interaction with other children and include ever more complicated matters. Besides constructing an understanding of the surrounding world, through their early explorations children also develop many scientific skills; they identify similarities and differences, and group things together. Early explorations are often seemingly unsystematic and unproductive, but as a child matures, exploratory skills also develop which then enables more skilled exploration and investigation to occur. The exploratory skills such as observation, raising questions, classification, and hypothesizing are important first steps in the development of other skills in the scientific process, especially planning, predicting, and investigating [16].

In science education, the role of children's interests and questions have become current themes through constructivist and socio-constructivist inquiry and project-based approaches that aim at fostering research-like processes of inquiry in education (see, e.g., [17–19]). In these approaches, learners are encouraged to initiate their own questions and hypotheses. These are then investigated under the teacher's guidance. The aims of inquiry learning at schools are related to learning more deeply about the phenomenon in question, namely, that the inquiry would concentrate on the key concepts in the phenomenon and understanding the theory behind the phenomenon. Through inquiry children can also learn the research skills, develop research strategies and metacognitive skills [20]. Often the inquiry process is described as a process that consists of posing problems, constructing explanations and hypotheses, searching for new information, and generating new questions as well as explanations. The progressively deepening nature of the inquiry process has been emphasized as well as the collaborative mode of learning. Moreover, information and communication technology has been used to facilitate the inquiry processes [21–24].

Inquiry or exploration activities that aim at understanding more deeply the phenomenon in question are deliberate processes, and in many cases some cognitive and socio-cultural support is a necessity. In schools, teachers have an important role to motivate and support students in their inquiry. In exploration-based science instruction, it has been emphasized that more effort is placed on students' thinking processes rather than on the need for correct answers, and that enough time must be given to the exploration of key concepts in one subject matter area [9]. It has also been considered important that students are provided with opportunities to work with phenomena instead of only watching teacher demonstrations, and that some cognitive scaffolding is available to help the students to find new and alternative ideas (e.g., [25, 26]). It has also been considered essential to pay attention to the order of acquisition of concepts in a given subject matter area. Teachers can take this into account when scaffolding the explorations. Meaningful and theoretically relevant experiences as well as providing models and external representations are important in clarifying the scientific explanations [9]. Models and external representations offer students opportunities to explore aspects of phenomena in other than linguistic form and they contribute for example the comprehension of complicated phenomena through providing visual presentations of interrelations in phenomena.

3.1. Computer Supported Exploration

Considering computer programs from the standpoint of the amount of support a child needs in exploration, we think that tool programs, as well as simulation programs, may leave the learner alone with his/her problems and the exploration process can be stuck for a long time. Of course, if a teacher or an adult is available, she/he can come to rescue, but often the teacher also has other learners to attend to. One possibility for the learner is to ask for help from more advanced peers in a classroom or in a networked environment. In fact, the web can be seen as an especially empowering environment for inquiry learning [21].

Knowledge-based programs (formerly called intelligent tutoring systems) are an interesting solution to the above-mentioned problem of support. These programs contain and maintain knowledge on the teaching domain, pedagogical approaches, as well as the individual learners, and present their support for the learner by combining these matters into guidance to assist the learner when she/he gets stuck or asks for help. The communication capabilities of modern tutoring systems are rather advanced. For example, some of these systems hold a conversation in natural language and in this way help students construct answers to deep-reasoning questions (e.g., AutoTutor system, [27]). AutoTutor is not merely an information delivery system, but it coaches students in actively constructing knowledge [27]. Some other tutoring systems, such as the Cognitive Tutors for mathematics from the Carnegie Learning Company, which is already in classroom use, excel in providing versatile interaction opportunities (in addition to diagnostic interventions) in problem-solving situations [28].

Lately, these tutoring systems are orchestrated from several coworking program entities, called intelligent agents. Agents can function “behind the scenes” as an implementation technique, or manifest themselves to the learner as animated characters on the screen. Much attention has been paid to animated pedagogical agents or characters presenting the support of the tutoring systems. Indeed, it is true that a vivid figure on the screen can make a difference to children, also in learning (see, e.g., [29]). We have constructed an agent-based learning environment [30] with several teacher agents and learning companion agents for the learner to select. Having the possibility to change your teacher or learning companion during the sessions helped the learners to find coworkers that suited their preferences [30]. However, we tend to agree with Graesser et al. [27] that it is the “message (the dialogue moves of the tutoring system)” that is more important than the “medium (the way it is presented to the learner).” This emphasizes the role of the “just in time” and informative dialogue moves, which are also in focus in proactive support for explorative learning. In addition, the intended users of our prototype are restricted to only a few of the typical input channels of computer programs, which set further challenges for the agent environment.

Visually impaired or blind children pose a tough problem for computer assistance. The program cannot rely on graphics, but the interface must depend on sound or the sense of touch (haptics). The degree of the disability, of course, dictates the environment design. However, it is clear that children with severe visual impairment also want to have the opportunity to access the learning environment, such as a game, alone and choose by themselves [31]. A design might include special tactile sheets for input and sound as feedback from the system, as in the case of Buaud et al. [31]. Baloian et al. [32] discuss general principles for educational software for visually disabled, and suggest an (intelligent) tutoring component in their architectural model. Their example system uses both the computer and LEGO blocks, both sound and touch.

3.2. The PICCO System as an Example of a System to Support Exploration in Astronomy

Kangassalo (e.g., [24, 33]) has examined the knowledge construction and conceptual understanding of phenomena in astronomy by pre and primary school children in learning environments where children have had possibilities to explore the astronomical phenomena using a computer-based multimedia simulation program PICCO [2, 34, 35].

The natural phenomena chosen for the PICCO program consist of the mutual relations of a number of phenomena, as well as the temporal and spatial relations of objects. The phenomenon chosen for the program is the changes in the sun’s light and warmth as experienced on Earth in relation to the mutual relations of the Sun and the Earth on the spatial level. The key concepts and phenomena in the program are time, night and day, the seasons, the Earth and the Sun, and space. The contents of the program are structured in such a way that the phenomenon can be first studied on the surface of the Earth, and children may proceed along various paths into the different areas of the program, according to their own interests. The areas include the mutual relations of the Earth and the Sun in space, the solar system and the planets, the mutual size and distance relations of the Earth and the Sun, and our solar system as part of the galaxy. The children may explore, for example, on the earth level what happens in the phenomenon in the environment at different time, in different months, and at different compass points. The phenomenon can be studied with the help of pictorial symbols of time and space, as well as with special research tools in the program, a dictionary, and a space shuttle. When studying the phenomenon from the surface of the Earth, for example, the seeming trajectory of the Sun and its position in the sky can be studied with the clock and the calendar, the changing of light and dark on different days of the year, the changes taking place in nature over months and seasons. At any moment a child can proceed up to the space level and explore the interrelationships of the Earth and the Sun at the space level. The program is a pictorial entity that proceeds in the form of events that are described realistically, as seen from the surface of the Earth, and on the spatial level based on analogy models. Reading or writing skills are not required for using the program. The child may proceed with studying the phenomenon according to his /her own interests and thoughts either alone, in pairs, or in small groups [2, 34, 35].

On the basis of the research experiments in both preschool and primary school teaching, it can be concluded that PICCO worked well as a tool for the children’s spontaneous and independent exploration. Exploring the phenomenon with PICCO supported the construction of the children’s knowledge of the phenomenon and directed learning towards established scientific knowledge. The children’s perceptions of the phenomenon as an entity and their outlinings of the mutual relations of the concepts and the link-ups developed and became more structured. As a synergy of PICCO and traditional teaching, considerable knowledge construction was detected. The children used a variety of research strategies while studying the phenomenon with the simulation program. The children’s conceptual models on the phenomenon acted as a basis through which they proceeded with exploring the phenomenon with PICCO. The research strategies children applied in their exploration were more advanced the more developed the children’s conceptual models of the phenomenon were (e.g., [24, 33, 36]).

4. Proagents Multimodal Learning Environment

Based on the theoretical and pedagogical background and research results with PICCO, a new Proagents system was designed. The leading design principles, in addition to the pedagogical principles, were the utilization of multimodal interaction (visual, auditory, and haptic) and proactive agents. The Proagents multimodal learning environment supports pre and primary school aged (6 to 8 years old) children's exploratory and conceptual learning in the domain of astronomy. Of special interest to us has been the construction of support for children's explorations and conceptual learning when they are using the system.

Multimodal interfaces [37] make use of the different senses and actions of the user. Multimodal systems have only recently been introduced in applications that are directly usable by a common computer user. Research on multimodal interfaces involves research on both human senses and capabilities, as well as how to construct such multimodal systems. We focused on the interaction design and content of the system, the highest level of which is visible for the user as the concrete multimodal user interface.

Disabled users may really benefit from multimodal interaction. There is a great amount of different disabilities and impairments, but in general adding new modalities, such as making use of a wider variety of senses, makes it possible for many users with special needs to make use of modern technology. This is why we have selected visually impaired children as the target users of our multimodal learning environment.

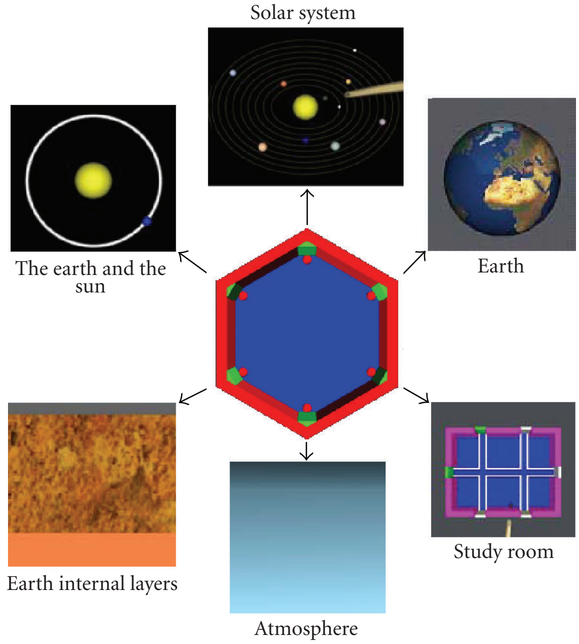


Figure 1: The structure of the simulation.

4.1. Pedagogical Approach of the System

The pedagogical approach of the system is based on exploratory learning. This means that a child can explore the selected astronomical phenomena independently, according to his/her own interests and questions. There are no rules or pathways on how to proceed in the program. The environment consists of six microworlds that a child can explore, see Figure 1. The user starts from the central station. From there she/he can move to one of the six microworlds. When the user is navigating from one microworld to another, she/he must travel through the central station. This lessens the likelihood of getting lost in the environment because the user is always only one step away from the central station. The microworlds are the solar system, the Earth and the Sun, the Earth, the atmosphere, and the interior layers of the Earth. The selected phenomena include central and abstract concepts related to the Earth's shape, to time and space, to the seasons, and to the alternation of day and night. The concepts and phenomena in microworlds are many ways connected to each other, and together they form a coherent theory.

The theoretical foundation and pedagogical basis for children's explorations are derived from the inquiry models that emphasize the role of questions as a starting point for the inquiry. One such model is the interrogative model of inquiry. This model was originally developed for the purposes of the philosophy of science (see [38–41]), but it has also been used to represent knowledge seeking in educational contexts (e.g., [18, 21]). In the interrogative model, scientific procedure is viewed as information seeking by questioning. More specifically, inquiry is defined as a series of questions the inquirer poses during his/her inquiry process, either to nature or to some other source of information. The inquirer tries to derive an answer to his/her initial question or problem by using his/her existing knowledge and by formulating and seeking answers to smaller questions. The acquisition of new knowledge raises new questions that have to be examined. By choosing the questions, the inquirer can direct the course of the inquiry according to his/her own plans [21, 38]. According to the interrogative model, an inquiry can be conceived as a dynamic, question-driven process of understanding [18].

In this research, applying the interrogative model means that a child's learning is viewed as an active process guided by his/her own questions and previous knowledge. The selection of the approach is largely based on Kangassalo's earlier research results [24, 33] with the PICCO computer environment, where a child has been seen to progress in his/her exploration process step by step on the basis of his/her earlier knowledge base concerning the phenomenon in question. In this project, we aim to continue the PICCO project by constructing support for children's explorations. The support aims at encouraging the formation of questions in child's mind as well as the process of seeking answers and explanations.

4.2. Proactive Agents as a Scaffolding Tool

It is possible to implement the scaffolding using software agents, independent program entities that can provide knowledgeable "just-in-time" support for learning. There are several existing agent applications that support the inquiry process and conceptual change in learners studying science. For example, Ting and Chong [42] have constructed an animated pedagogical agent that acts as a catalyst coaching learners through the scientific inquiry. White et al. have reported several studies on multiagent environments fostering scientific inquiry, conceptual change, and reflective learning; see for example, the Inquiry Island environment [43].

To support children's explorations in the system, we have utilized proactive agents as a scaffolding tool. Our program utilizes many kinds of agents. In this article, we describe the pedagogical agents that communicate with children. Each microworld in the program has its own assisting pedagogical agent. The agents are different imaginary characters, and they all have different names and voices. Since the system is developed also for visually impaired and blind children, the agents' operations are based mainly on auditory feedback.

The starting point for constructing the agents' support has been that the agents would support children's explorations in the program starting from the child's own activity, interest, and questions. For this reason the agents support the explorations mainly by making questions and suggestions to the child, and the child him/herself can choose whether to take the agents suggestions or to continue with his/her own exploration. With their suggestions and questions the agents aim at deepening and extending the child's explorations in the program, and to awaken questions in a child's mind. Decisions regarding the appropriate times for suggestions are based on, for example, the child's exploration time in a microworld and the child's exploration path in the program.

The agent's suggestions assist the child to find the central phenomena and concepts that are connected to the microworld the child is currently exploring. They also guide the explorations from one microworld to another, and in this way support the discovery of relations and explanations in the selected phenomena. From the viewpoint of conceptual learning, the agents guide the explorations from familiar everyday observations towards the causes and scientific explanations of phenomena. For instance, if the child is exploring the solar system and has already examined the different planets and their properties, an agent might suggest to the child that she/he could find the planet Earth. After that the agent may challenge the child to consider why only the Earth has people and animals on it, and suggest that the child might like to explore the Earth even closer. Furthermore, when exploring the Earth an agent may direct the child's attention to Earth's gravity with a comment like "did you notice when you traveled with your space shuttle you were pulled to the Earth's surface?" It can also challenge the child's thinking with arguments like "what happens to different objects when you throw them into air or drop them?" and offer explanations on gravity.

In summary, the agents try to guide the children to elaborate their previous knowledge through their questions and suggestions, and encourage them to examine the properties and relations in phenomena. The proactive agents also aim to help the children become conscious of their own exploration and thinking. The agents allow children to explore and proceed in different ways, and the child him/herself can continuously choose either to listen to what an agent wants to say or to ignore him.

Figure 2: Multimodal interface technology.

4.3. Multimodal Interaction

The learning environment is based on using a Reachin Display System [44] which integrates a PHANTOM force feedback device from SensAble Technologies [45] with a stereo monitor and supporting systems (see Figure 2). The PHANTOM device has a thick movable stick which provides haptic feedback. By using the stick, a child can feel three-dimensional entities. For visually impaired children, haptic and auditory modalities are the main presentation channels, and for children with normal eyesight, the system provides opportunities to receive and use the information obtained through multiple senses.

Haptics often involve active touch (intentional actions) that a person wants and chooses to do. In other words, when using the system the learner can control his or her actions and the speed of exploration. This is in contrast with passive learning, such as watching a video or a teacher demonstration, where the learner does not have to make any decisions or initiate actions. When comparing haptics, for example, to vision in the perception of objects, vision is usually considered superior although there are some important exceptions. Visual perceptions allow the learner to take in a lot of information at one time and it is more rapid and holistic than haptic perception. Haptics, on the other hand, help learners detect the properties of texture and some microspatial properties, like elasticity, compliance, and pattern [46]. In our system, it has been possible to represent the concept of gravity through haptics. Furthermore, features of the Earth's surface, such as oceans and the ground, have been possible to create as one object of exploration.

The proactive and adaptive functionality of the system is implemented by carefully modeling the astronomical phenomena as well as maintaining student profiles. Based on this long-term logging of user actions, rule-based agents initiate

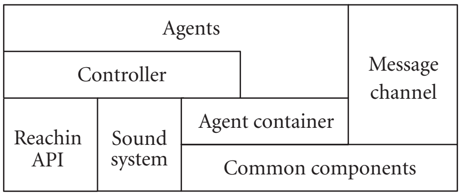


Figure 3: Layered view of the architecture.

proactive actions. The inference and learning capabilities of the agents will enable them to produce proactive behavior, such as to suggest autonomously interesting or important information at appropriate times. Since the targeted users of the system have either just started school or are in their first years of school, the most useful way to provide an agent's feedback is speech. This is also directly suitable for visually impaired children.

Haptic modality is also involved in agents' operations, for example, when an agent wishes to say something, the stylus in the child's hand shakes. The child can then decide whether he or she wants to hear the agent's message by pressing either a "Yes" or "No" button. In addition, agents may also refer in their messages to the sense of touch when the child is examining and feeling a microworld with the stylus.

Figure 3 shows the low-level technical structure of the agent-based system [47]. The controller is built on top of the Reachin API [44] which provides a means to interact with the 3D simulation environment and the haptic devices, and our own simple sound system. The basic agent architecture can be divided into three separate functional components. The system can be seen as the realization of a typical message dispatcher architecture where MessageChannel is the central dispatcher providing a centralized means for passing messages between agents around the network. The agent containers (and thus agents) are connected to the agent system via MessageChannel. The third component is the actual application that is handled by the controller. The functionality and structure of MessageChannel, agent container, and agents is loosely based on FIPA agent specification (http://www.fipa.org/).

The agent-based system architecture is used to construct higher-level agents that handle multimodal input and output [47]. The database agent stores application-dependent information and notifies the agents that have subscribed to be informed of a certain change in the database. The rule engine agent has been built on top of the CLIPS rule engine [48] and manages rule-based events in the user interface. Filter agents can be used to process originally low-level input in higher levels following the basic pipes and filters functionality. The controller agent handles the central control of the application and manages the logging agent that stores events in the system.

A detailed technical description of the system can be found in our earlier publications [47, 49]. The present paper adds to these publications especially in the areas of pedagogical design and testing of learning, usability, and interaction in the environment.

4.4. The Microworlds

In the following, the program's microworlds are described shortly. As mentioned in the previous sections, the system consists of six microworlds that the child can explore according to his/her own interests. In the *solar system microworld* the child can examine the planets in our solar system. The solar system is depicted as a whole on the screen; the planets, their orbits, and the sun. As the child touches one of the orbits, the program tells the child which of the planet's orbit is it. The orbit can be felt under the stylus as a groove, and a blind child can move the stylus along the groove. This design decision was made based on the need to provide some guidance for a blind child navigating in a three-dimensional space. We experimented with different alternative methods. In our pilot tests, the selected 2.5-dimensional representation helped the children to use the system and still get the benefit from three-dimensional haptic forms and textures. When the child touches the planet with the stylus, the program tells him/her what planet it is and where it is located in relation to the Sun. It is also possible to listen to more information on each of the planets by pressing the planet with the stylus.

In the *Earth and the Sun microworld*, the child can study the Earth's revolving around the Sun, as well as the changing of the seasons in Finland at the different stages of its revolution. As visual feedback, the screen shows the Sun, the Earth, and the Earth's orbit. When the child touches the various objects in the microworld, the program tells him/her where she/he is located. The orbit is presented as a groove, which enables the blind child to follow the orbit. The program also informs the child, when she/he finds the Earth, and the child can then move the Earth on its orbit around the Sun. At the different stages of the orbit, the program plays the sounds typical to the season in Finland. The program also tells about the Sun's light and warmth in each season.

In the *Earth microworld*, the child can study the round shape of the Earth as well as the Earth's surface. The Earth can be felt with the stylus as a three-dimensional round object. When touching the surface of the Earth, it is possible to feel the differences between solid ground and the oceans. Haptic feedback is supported by auditive feedback; when touching the surface of the Earth, one can hear the sounds of human habitation and nature; at the sea, one can hear sounds typical of the ocean (waves, seagulls). The Earth also has gravity that can be felt with the stylus as a light pull towards the Earth. Gravity is one of the things that were possible to represent very well through haptics. The user can also rotate the Earth around its axis by moving the stylus to the right of the Earth.

In the *Bowels of the Earth microworld*, the child can explore the insides of the Earth. The various layers of the Earth are represented as a cross-section of the northern hemisphere. The layers can be explored by touching them with the stylus.

In the *atmosphere microworld*, the child may study the Earth's atmosphere from the surface of the Earth to the upper layers of the atmosphere. Exploring the microworld is first and foremost based on auditive feedback. The haptic feedback is almost unnoticeable and light, and it aims to create a tangible "feeling of air." The child can freely move with the stylus through the different layers of the atmosphere. The program also tells the child, when the child him/herself so desires, about the characteristics and the importance of the atmosphere.

The study room is a room with six doors. When the child presses one of the doors with the stylus, the program gives the child a statement related to gravity, and the child can answer "yes" or "no" by pressing a button. There are grooves that lead to the doors, on the floor of the research laboratory; and they assist the blind child in finding his/her way to the doors with the help of haptic feedback.

5. User Studies: Testing and Evaluating the System with Children

Both blind and sighted children have participated in the planning and testing process of our system from the very beginning of the research project. The development and testing of the system continued throughout the project. There were four kinds of tests included in the system development: the testing of the system's manuscript, tests at the usability laboratory, children as expert evaluators, and a study of children's exploratory learning. In this section, these tests are presented as well as some results of the tests. The results and experiences we received from the pedagogical agents supporting children's explorations are presented in the next section.

5.1. Testing the Manuscript of the System

In the first phase, we tested the manuscript for the agents' operations with children. The testing included interviews of ten 5- to 9-year-old sighted children. The interviews were conducted mainly at a day care center, and the children were interviewed individually.

The interviews proceeded as follows. First, the children were asked about their interests and questions with regard to the selected phenomena. After that the interview concentrated on one or two microworlds. The child was presented with general idea(s) about the micro world(s) and some short narrations of the agents' operations were read aloud. Limiting the discussion only to one or two microworlds was considered necessary so that the interviews would not be too long for the child. However, since there were three interviewers, all of the microworlds were evaluated by children. When presenting and discussing the manuscript and the agents' operations, some assistive materials were used (such as a picture of the solar system, a globe, etc.). This material helped to concretize the phenomena and the manuscript. The child was then asked how he/she liked the manuscript, how the narration could be improved, and whether he/she thought that something could be added or removed from the manuscript. Some separate utterances and sentences that were designed for the agents were also read aloud, and children were asked to evaluate if there were any difficult words in these utterances.

One interview lasted about thirty minutes, and the interviews were recorded on tapes. This testing at the very beginning of our developing work helped us improve, for example, the lines we had designed for the agents, and at the same time, it gave us insight into children's interests and knowledge with regard to the phenomena chosen for the computer program.

5.2. Testing at the Usability Laboratory

We have also tested the usability of the system with children. These tests have been carried out in the usability laboratory and at the school for visually impaired children. The tests in the laboratory have been organized whenever some new solutions have needed testing, and both visually impaired and sighted children have participated in these tests at different phases of the system's development.

Two researchers participated in the testing situations. One of them conducted the actual test with the child and the other was responsible for observing and videotaping the situation. The researcher who conducted the testing situation had a kindergarten teacher's training and experience in working with children.

The children came to the laboratory with their parents. After the laboratory rooms were introduced to the child, the researcher talked with the child and the parent. The purpose of the discussion was to inform the child and the parent about the goal of the test, to tell about the system, and to have the child feel secure and comfortable.

The test began by familiarizing the child with the PHANTOM device. After that, the child was given a tangible model of the program's navigation tool. The navigation tool is called "Central Station," and it's a place where a child can choose which microworld he/she would like to explore. The researcher guided the child to explore the tangible model first with hands and then with a plastic stick. The plastic stick was modeled after the PHANTOM's stylus. At the same time the researcher explained to the child the function of the navigation tool ("Central Station"), and guided him or her to examine its shape. Familiarizing the child with the navigation tool using the tangible model helped the children comprehend the general structure of the program. This was important, as we wanted to support children's independent exploration.

After that, the child could start using the program. The goal was that the child would use the program as independently as possible. However, guidance and help was offered whenever the child needed it. To some extent, the researcher also tried to elicit children's ideas and comments during the use. The researcher, for example, sometimes asked the child to describe what he or she felt with the stylus, and made suggestions that encouraged the child to explore and examine the different objects in the microworlds.

After the child had used the system, he/she was briefly interviewed. The aim of this small interview was to hear the child's comments about the program: how easy or difficult it was to use, what the child had liked and disliked in the program, what could still be added to program, and so forth.

All the tests were recorded on video and afterwards the usability of the tested microworlds was evaluated. The evaluation of the usability in these tests meant that we mainly observed what kind of solutions supported children's independent exploration and in what kind of situations children most often needed researcher's help.

5.3. Children as "Expert Evaluators"

Two tests were carried out at the school for visually impaired children. The first of these tests included seven 12-year-old visually impaired children. The microworlds that were tested were the solar system, the Earth, and the study room. The aim of this testing was to use the system in natural surroundings and to receive as much feedback as possible from the usability of the system. The children in this test were older than the actual target group of our system, and the children's task in this test was to act as "child experts" to give comments and feedback about the system and to evaluate how they thought that smaller children than they would be able to use the system. In all test situations the children used the system one by one, and the researcher sat next to the child and assisted him/her when necessary.

Before the actual tests, the researchers introduced themselves to the children in their classroom, and told them about the test. The purpose of the test was emphasized: we wished the children to use the program, and to tell us what they thought about it, what they thought about the selected astronomical phenomena, and how they thought that the program could be improved. The children were also told that the program is designed for younger children than they were. The important thing was to create and maintain the children's motivation for the evaluation task, and to describe the device and the testing situation beforehand so that the children would feel secure and confident about the test.

Similar to the laboratory tests, two researchers participated in the testing situation. The children's use of the program was videotaped. The test took place in an empty room near the children's classroom. Each child was reserved a lesson period of 45 minutes for the test, but the time children used the system varied. This is because the use of the system was based on the children's own interest. At first, the child was familiarized with the system. Then the structure of the program was explained to the child with the help of the tangible model of the navigation tool (Central Station). After that, the child was allowed to explore the program freely. The researcher sat next to the child, and provided help whenever the child needed it. The researcher also tried to elicit the children's comments and thoughts during the use of the system, and asked the children such questions as "if you'd like to find some particular planet, how difficult would it be?" and "how did you understand what the agent just said?" The children were also briefly interviewed directly after the use. Below are some examples of the comments children gave regarding the agents and their operations after they had used the system.

> "*You could say, for example, how cold it is on Pluto…if possible?*" (solar system microworld)

> "*Maybe more information about the Sun and its structure—a small child may think that the Sun is solid, although it indeed is not.*" (solar system microworld)

> "*Earth Giant* (the agent) *could tell something about the people who live on the continents you are currently exploring.*" (The Earth microworld)

> "*I would add there also more challenging questions, really difficult ones, so that there would be questions of various levels of difficulty.*" (Research laboratory)

These were just some examples of children's comments. In sum, children were very motivated in evaluating the system, and their comments and suggestions seemed well thought and also realistic with regard to realizing them. The children's comments regarding the overall usability of the system were also very valuable to us, especially, when we were thinking that also smaller children should be able to use the system.

5.4. Study of Children's Exploratory Learning

The final test was carried out at the school for visually impaired children where two 7-8-year old blind children participated. Afterwards, a third child from this same age group participated, and this test was carried out in the usability laboratory a few months later. In this test we had three microworlds available: the Earth, the solar system, and the Earth and the Sun. One of the aims of this test was to examine blind children's exploratory and conceptual learning when using the system.

The two children who participated in the test at the school for visually impaired children used the system twice, for about a single lesson period of time (45 minutes) on successive days. The third child used the system at the university's usability laboratory once also for approximately 45 minutes. At first, a child was familiarized with the PHANTOM device, and a tangible model of the central station was also presented (see the description of the previous tests). After that, the child could use the program and explore the phenomena according to his/her own interest. In this test, we also had plastic tangible models of all the micro worlds. As the child explored a microworld, she/he was able to get a general picture of what he or she was exploring by touching the plastic model. As in the previous tests, the researcher helped the child whenever he/she needed help in using the system. In addition, the researcher encouraged and supported the child's explorations by posing questions and sometimes directing the child's attention to central objects or events in phenomena. However, the use of the program was always based on child's own interest and independent exploration.

The research data gathered included video recordings and log files of the children's use of the program. The log files provided information about the child's exploration pathways. The child's comments, questions, and other expressions as well as the researcher's guidance could be observed from the videotapes. To get an accurate picture of the child's exploration process, these video recordings were also transcribed next to the log files.

6. Results and Experiences on Pedagogical Agents Supporting Children's Explorations

In this section, the results and experiences we have had of the pedagogical agents that were designed to support children's exploration are discussed from the basis of the tests carried out. The discussion concentrates on the experiences we received from the final test carried out with three visually impaired 7-8-year-old children. The experiences and results are preliminary in nature because there were only three children who participated in the test, and also because during this test the environment was still under construction and the agents' action did not yet include all the messages that were designed to them.

To analyze the experiences we got from the pedagogical agents supporting children's exploration we transcribed the researcher's and a child's interaction from the video recordings next to the log files the computer had gathered. In this way it was possible to reliably identify the child's comments, questions, and other expressions and actions with regard to agents' messages from the data.

6.1. Children's Responses to Agents' Messages

During the children's exploration there came several messages from the agents (*child 1:* 8 messages at the first exploration time, 12 messages at the second exploration time; *child 2:* no log files saved from the first exploration time, 10 messages at the second exploration time; *child 3:* 9 messages). The children chose to receive almost all the messages (92% of all the messages were received). Most of the agents' messages were such that the agent told more information for the child about the object or phenomenon he/she was currently exploring (72% of all the messages the children chose to receive).

The children's reactions and responses to the agents' messages showed that the children were able to include the agents naturally in their exploration of the phenomenon, although the amount the child commented on or discussed the messages aloud varied. When examining the exploration process of child 1, who was the most talkative of the three children, from the point of view of his reactions and responses to agents and their messages, it seemed that the child became more interested in the agents and their messages during the exploration, and was able to include them better as a part of his inquiry as the exploration proceeded. During the first exploration time, this child received all of the agents' messages that became available and commented on one of them aloud. During the second exploration time, the child answered one of the agents' questions aloud, and afterwards, after receiving another message, he started to re-evaluate his earlier answer and thinking.

(In the Earth–the Sun microworld:)

09:43:18 **Sunny Anneli (an agent):** *You moved the Earth around the Sun quite fast with the stylus. How long do you think the real planet Earth takes to travel around the Sun? You can think about this while travelling around the Sun once again and I'll tell you what it is like here in Finland during the different parts of your journey.*

Child1: *and it took the two months.*

Researcher: *mm.*

Child1: *I think.*

Researcher: *well we can circle the orbit again and as we circle we can listen.*

(In the Earth micro world:)

09:52:41 **Earth Giant (an agent):** *You are now exploring the earth, the planet in which we live. The earth's circumference is approximately 40,000 kilometers. This means that if you could travel around the world by car, you would have to sit still for three weeks in a row.*

Child1: *It takes less time than the round of the sun*

09:53:03 Surface of earth

09:53:03 Ocean

Researcher: *mm.*

09:53:04 Ocean

Child1: *It.... I think I guessed it a bit wrong.*

09:53:09 Ocean

09:53:09 Ocean

Researcher: *well, what do you think about it now?*

09:53:09 Surface of earth

09:53:10 Surface of earth

Child1: *Now I think if you took an airplane it'd take only two weeks.*

09:53:11 Surface of earth.

There was also a phase where the child tested the agents' function through pressing the "Yes" and "No" buttons to different messages and looking what happens. Towards the end of the exploration the child seemed to learn to expect to hear more information about the phenomenon from the agents. For example, when there was a message coming from the agents, he told the researcher that *"I press "yes" so that I can hear information."* He also expressed a wish to hear an agent when he was in solar system and examined the Sun ("*If a message comes hear, I will press "yes""*). The last message from the agents elicited a question from a child, which could have lead to further exploration.

The two other children did not make many comments aloud. However, they listened to the agents' messages very carefully, and the one child (child 2) commented and wondered at the low voice of the Earth Giant many times. He was also very interested in the narrative elements of the system as he made many questions and comments regarding, for example, vehicles at the research station. He also noticed when there was a long break in the agents' messages ("*Now the program has not said anything for a long time.... This must be such an object that here it needs so much...*").

6.2. Future Challenges

When considering the phenomenon as a whole, how it was modeled in the program and how the agents managed to support the children's explorations from this perspective, it seems that the Earth and the Sun microworld (where the child can explore the Earth's revolution around the Sun) was the most challenging microworld to model for the blind users as well as to support their explorations there. In this microworld the movement of Earth in relation to the static Sun, and the connection of this to the changing of seasons, were the main objects of exploration. Making the movement possible for blind children to explore was very challenging. On the basis of this small test it seems that at least children whose conceptual model of the changing of seasons is at a very initial level and does not yet include features of the scientific model would need more support for this microworld to be outlined, or the support needs to be constructed differently in order to make the perspective from which the relations of the Earth and the Sun are examined clearer.

We consider the experiences we received from the pedagogical agents supporting the children's inquiry to be very promising, although there's still some development work and more extensive tests to be done. The technical solution for receiving and rejecting the agents' messages through pressing the buttons after the stylus in the child's hand had shaken seemed to work well because the children learned this very quickly. At the time of this research experiment, the agents' role as "information givers" became emphasized because of the nature of their messages at this developing phase. However, the children seemed very motivated towards the information given by the agents, and sometimes waited for the agents to tell them something. In future development work, it would be good to increase the amount of suggestions and questions in the agents' operations, and to carry out more extensive tests. In this research experiment, the researcher acted as the children's conversational partner and assisted the children in the use of the system. In the future it would also be interesting to examine how, for example, children discuss and explore the phenomenon together when using the system.

7. Conclusions

Information technology can be of great help in the understanding of astronomical phenomena. According to the earlier research it has also been found that children are very interested in these phenomena and they explore phenomena on the basis of their earlier knowledge. Children's conceptual understandings have seemed to develop in the direction of the currently accepted scientific knowledge when children have had possibilities to explore these phenomena using computer simulation system PICCO program [34, 35]. These earlier research results concern sighted children [24, 33].

We have constructed a system that supports blind and visually impaired children. For them, information technology provides greater opportunities to explore abstract phenomena and their spatial relationships. We used haptic devices for producing haptic sensations, as haptic perception is an important exploration means for blind and visually impaired children. Proactive agents were used to support children's explorations and conceptual learning in the domain of astronomy. It was challenging to construct and integrate in a pedagogically relevant way the explorative learning approach, proactive agents' action, haptic perception, and the selected astronomical phenomena for seeing, visually impaired and blind children's learning and exploration. Based on our results, the system is expected to support children's explorative action, the formation of questions, the construction of knowledge, and the enhancement of interest in the selected astronomical phenomena.

Acknowledgments

We express our gratitude to the members of our research group and the children and their parents for participating in our pilot experiments. Special thanks to Janne Järvi, Rami Saarinen, and Jouni Salo for their valuable work in the implementation of the system. This research (Grants no. 202179 and no. 202180) was funded by the Academy of Finland as part of its research program on Proactive Computing (PROACT) for the period 2002–2005 (http://www.uta.fi/proact/).

References

[1] S. Papert, "Hard Fun," Article in the Bangor Daily News (Bangor, Maine), 2002, http://www.papert.org/articles/HardFun.html.

[2] M. Kangassalo, "The pictorial computer-based simulation in natural sciences for children's use," in *Information Modelling and Knowledge Bases III: Foundations, Theory and Applications*, S. Ohsuga, H. Kangassalo, H. Jaakkola, K. Hori, and N. Yonezaki, Eds., pp. 511–524, IOS Press, Amsterdam, The Netherlands, 1992.

[3] I.-A. Diakidoy, S. Vosniadou, and J. D. Hawks, "Conceptual change in astronomy: models of the earth and of the day/night cycle in American-Indian children," *European Journal of Psychology of Education*, vol. 12, no. 2, pp. 159–184, 1997.

[4] W. Schnotz, S. Vosniadou, and M. Carretero, "Preface," in *New Perspectives on Conceptual Change*, W. Schnotz, S. Vosniadou, and M. Carretero, Eds., pp. 13–24, Elsevier Science, Oxford, UK, 1999.

[5] J. Baxter, "Children's understanding of astronomy and the earth sciences," in *Learning Science in the Schools: Research Reforming Practice*, S. M. Glynn and R. Duit, Eds., pp. 155–177, Lawrence Erlbaum, Mahwah, NJ, USA, 1995.

[6] J. Nussbaum, "The Earth as a cosmic body," in *Children's Ideas in Science*, R. Driver, E. Guesne, and A. Tiberghien, Eds., pp. 170–192, Open University Press, Milton Keynes, UK, 1985.

[7] S. Vosniadou and W. F. Brewer, "Mental models of the Earth: a study of conceptual change in childhood," in *Cognition and the Symbolic Processes: Applied and Ecological Perspectives*, R. R. Hoffman and D. S. Palermo, Eds., pp. 535–585, Lawrence Erlbaum, Hillsdale, NJ, USA, 1991.

[8] S. Vosniadou, "Capturing and modeling the process of conceptual change," *Learning and Instruction*, vol. 4, no. 1, pp. 45–69, 1994.

[9] S. Vosniadou, C. Ioannides, A. Dimitrakopoulou, and E. Papademetriou, "Designing learning environments to promote conceptual change in science," *Learning and Instruction*, vol. 11, no. 4-5, pp. 381–419, 2001.

[10] K. Ikospentaki and S. Vosniadou, "The development of observational astronomy in blind children," in *Proceedings of the 4th European Symposium on Conceptual Change: Philosophical, Historical, Psychological and Educational Approaches*, S. Vosniadou, C. Stathopouloum, X. Vamvakoussi, and N. Mamalougos, Eds., Delphi, Greece, May 2004.

[11] P. M. Sadler, "Misconceptions in astronomy," in *Proceedings of the 2nd International Seminar on Misconceptions and Educational Strategies in Science and Mathematics*, pp. 422–425, Ithaca, NY, USA, July 1987.

[12] S. Vosniadou and W. F. Brewer, "Mental models of the day/night cycle," *Cognitive Science*, vol. 18, no. 1, pp. 123–183, 1994.

[13] V. Virsu, "Aivot ja oppiminen," in *Suorin tie ei aina ole lyhin. Varhaiskasvatuksessa oppimisen alku*, pp. 14–18, Opetusalan Ammattijärjestö OAJ, Helsinki , Finland, 1997.

[14] T. Kohonen, *Self-Organization and Associative Memory*, Springer, Berlin, Germany, 3rd edition, 1984.

[15] OECD, *Understanding the Brain: The Birth of a New Learning Science*, Organization for Economic Co-operation and Development, 2007.

[16] J. Johnston, *Early Explorations in Science*, Open University Press, Berkshire, UK, 2nd edition, 2005.

[17] S. Carey and C. Smith, "On understanding the nature of scientific knowledge," in *Software Goes to School: Teaching for Understanding with New Technology*, D. N. Perkins, J. L. Schwartz, M. M. West, and M. S. Wiske, Eds., pp. 39–55, Oxford University Press, New York, NY, USA, 1995.

[18] K. Hakkarainen and M. Sintonen, "The interrogative model of inquiry and computer-supported collaborative learning," *Science and Education*, vol. 11, no. 1, pp. 25–43, 2002.

[19] M. Scardamalia and C. Bereiter, "Schools as knowledge building organizations," in *Today's Children, Tomorrow's Society: The Developmental Health and Wealth of Nations*, D. Keating and C. Hertzman, Eds., pp. 274–289, Guilford, New York, NY, USA, 1999.

[20] M. Kangassalo, "Tutkiva oppiminen," in *Kotipihasta maailmalle. Ympäristökasvatuksen karttakirja*, L. Niinikangas, Ed., pp. 28–42, BTJ Kirjastopalvelu, Helsinki, Finland, 2004.

[21] K. Hakkarainen, K. Lonka, and L. Lipponen, "Tutkiva oppiminen. Järki, tunteet ja kulttuuri oppimisen sytyttäjinä. 6. uudistettu painos," Porvoo: WSOY, 2004.

[22] K. Kumpulainen and D. Wray, *Classroom Interaction and Social Learning. From Theory to Practice*, Routledge, London, UK, 2002.

[23] K. Lonka, K. Hakkarainen, and M. Sintonen, "Progressive inquiry learning for children—experiences, possibilities, limitations," *European Early Childhood Education Research Journal*, vol. 8, no. 1, pp. 7–23, 2000.

[24] M. Kangassalo, *The formation of children's conceptual models concerning a particular natural phenomenon using PICCO, a pictorial computer simulation*, Ph.D. thesis, Acta Universitatis Tamperensis 559, University of Tampere, Tampere, Finland, 1997.

[25] T. Andre and M. Windshitl, "Interest, epistemological belief, and intentional conceptual change," in *Intentional Conceptual Change*, G. M. Sinatra and P. R. Pintrich, Eds., pp. 173–197, Lawrence Erlbaum, Mahwah, NJ, USA, 2003.

[26] G. Hatano and K. Inagaki, "When is conceptual change intended? A cognitive sociocultural view," in *Intentional Conceptual Change*, G. M. Sinatra and P. R. Pintrich, Eds., pp. 407–427, Lawrence Erlbaum, Mahwah, NJ, USA, 2003.

[27] A. C. Graesser, K. Moreno, J. Marineau, A. Adcock, A. Olney, and N. Person, "AutoTutor improves deep learning of computer literacy: is it the dialog or the talking head?" in *Proceedings of the 11th International Conference on Artificial Intelligence in Education (AIED '03)*, U. Hoppe, F. Verdejo, and J. Kay, Eds., pp. 47–54, IOS Press, Sydney, Australia, July 2003.

[28] K. R. Koedinger, J. R. Anderson, W. H. Hadley, and M. A. Mark, "Intelligent tutoring goes to school in the big city," *International Journal of Artificial Intelligence in Education*, vol. 8, pp. 30–43, 1997.

[29] J. Lester, S. Converse, B. Stone, S. Kahler, and T. Barlow, "Animated pedagogical agents and problem-solving effectiveness: a large-scale empirical evaluation," in *Proceedings of the 8th World Conference on Artificial Intelligence in Education*, pp. 23–30, Kobe, Japan, August 1997.

[30] P. Hietala and T. Niemirepo, "The competence of learning companion agents," *International Journal of Artificial Intelligence in Education*, vol. 9, no. 3-4, pp. 178–192, 1998.

[31] A. Buaud, H. Svensson, D. Archambault, and D. Burger, "Multimedia games for visually impaired children," in *Proceedings of the 8th International Conference on Computers Helping People with Special Needs (ICCHP '02)*, vol. 2398 of *Lecture Notes in Computer Science*, pp. 173–180, Linz, Austria, July 2002.

[32] N. Baloian, W. Luther, and J. Sánchez, "Modeling educational software for people with disabilities: theory and practice," in *Proceedings of the 5th International ACM SIGCAPH Conference on Assistive Technologies*, pp. 111–118, Edinburgh, UK, July 2002.

[33] M. Kangassalo, "Conceptual change in astronomical phenomena using PICCO," in *the 2nd European Symposium on Conceptual Change*, EARLI, Madrid, Spain, November 1998.

[34] M. Kangassalo, "PICCO—The pictorial computer simulation of a selected natural phenomenon for children's use," Computer program, The grant of the Academy of Finland, 1991.

[35] M. Kangassalo, *PICCO Day and Night, Seasons—the Earth and the Sun*, Cd-rom, Piccos Programs, 1998.

[36] M. Kangassalo, "PICCO as a cognitive tool," in *Information Modelling and Knowledge Bases VII*, Y. Tanaka, H. Kangassalo, H. Jaakkola, and A. Yamamoto, Eds., pp. 344–357, IOS Press, Amsterdam, The Netherlands, 1996.

[37] R. A. Bolt, ""Put-that-there": voice and gesture at the graphics interface," in *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 262–270, ACM Press, Seattle, Wash, USA, July 1980.

[38] J. Hintikka, "True and false logic of scientific discovery," in *Logic of Discovery and Logic of Discourse*, J. Hintikka, Ed., pp. 3–14, Plenum Press, New York, NY, USA, 1985.

[39] J. Hintikka, "What is the logic of experimental inquiry?" *Synthese*, vol. 74, no. 2, pp. 173–190, 1988.

[40] M. Sintonen, "How to put questions to nature," in *Explanation and Its Limits*, D. Knowles, Ed., vol. 27 of *Royal Institute of Philosophy Lecture Series*, pp. 267–284, Cambridge University Press, New York, NY, USA, 1990.

[41] M. Sintonen, "Why questions, and why just why-questions?" *Synthese*, vol. 120, no. 1, pp. 125–135, 1999.

[42] C.-Y. Ting and Y.-K. Chong, "Enhancing conceptual change through cognitive tools: an animated pedagogical agent approach," in *Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies (ICALT '03)*, pp. 314–315, Athens, Greece, July 2003.

[43] B. White, J. Frederiksen, T. Frederiksen, E. Eslinger, S. Loper, and A. Collins, "Inquiry Island: affordances of a multi-agent environment for scientific inquiry and reflective learning," in *Proceedings of the 5th International Conference of the Learning Sciences (ICLS '02)*, P. Bell, R. Stevens, and T. Satwicz, Eds., Erlbaum, Seattle, Wash, USA, October 2002.

[44] Reachin Technologies AB, http://www.reachin.se.

[45] SensAble Technologies Inc, http://www.sensable.com/.

[46] E. B. Goldstein, *Sensation and Perception*, Brooks/Cole, Pacific Grove, Calif, USA, 5th edition, 1999.

[47] R. Saarinen, J. Järvi, R. Raisamo, and J. Salo, "Agent-based architecture for implementing multimodal learning environments for visually impaired children," in *Proceedings of the 7th International Conference on Multimodal Interfaces (ICMI '05)*, pp. 309–316, ACM Press, Trento, Italy, October 2005.

[48] "CLIPS: A Tool for Building Expert Systems," http://clipsrules.sourceforge.net/.

[49] R. Saarinen, J. Järvi, R. Raisamo, et al., "Supporting visually impaired children with software agents in a multimodal learning environment," *Virtual Reality*, vol. 9, no. 2-3, pp. 108–117, 2006.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 381086, 13 pages
doi:10.1155/2008/381086

Research Article

The Development of LinguaBytes: An Interactive Tangible Play and Learning System to Stimulate the Language Development of Toddlers with Multiple Disabilities

Bart Hengeveld,[1] Riny Voort,[2] Caroline Hummels,[1] Jan de Moor,[3] Hans van Balkom,[2,3] Kees Overbeeke,[1] and Aadjan van der Helm[4]

[1] *Designing Quality in Interaction, Department of Industrial Design, Eindhoven University of Technology, Den Dolech 2, 5612 AZ Eindhoven, The Netherlands*
[2] *PonTeM, Wijchenseweg 122 D, 6538 SX Nijmegen, The Netherlands*
[3] *Department of Special Education, Behavioural Science Institute, Radboud University Nijmegen, Postbus 9104, 6500 HE Nijmegen, The Netherlands*
[4] *ID-StudioLab, Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands*

Correspondence should be addressed to Bart Hengeveld, b.j.hengeveld@tue.nl

Received 1 October 2007; Revised 12 August 2008; Accepted 19 October 2008

Recommended by Adrian Cheok

Young children with multiple disabilities (e.g., both cognitive and motor disabilities) are confronted with severe limitations in language development from birth and later on. Stimulating the adult-child communication can decrease these limitations. Within *LinguaBytes*, a three-year research program, we try to stimulate language development by developing an interactive and adaptive play and learning environment, incorporating tangible objects and multimedia content, based on interactive storytelling and anchored instruction. The development of a product for such a heterogeneous user group presents substantial challenges. We use a Research-through-Design method, that is, an iterative process of developing subsequent experiential prototypes and then testing them in real-life settings, for example, a center for rehabilitation medicine. This article gives an outline of the development of the *LinguaBytes* play and learning environment from the earliest studies up to the current prototype, *CLICK-IT*.

1. Introduction

Normal developing children acquire language skills seemingly effortless. However, this is not the case for non- or hardly speaking toddlers with multiple disabilities [1, 2]. These children are confronted with severe limitations in language, early literacy, and communication development from birth on. While *language* refers to the use of words in spoken, written, signed, or other symbolic forms in either the expressive or receptive modality [3], *early literacy* refers specifically to early behaviours, like storybook reading that precedes and develops into conventional literacy. These are evident in even very young children as part of informal daily experiences [4]. *Communication* refers to the transmission of meaning from one individual to another, whatever the means used (verbal, with and without speech, nonverbal, with and without vocal output) [5]. Communication implies a process of social interaction. In this article we focus on language development, with the notion that language development can be stimulated in adult-child communication and that early literacy activities are an important context to stimulate language development.

A major part of the non- or hardly speaking toddlers with multiple disabilities have the diagnosis *Cerebral Palsy* (CP). CP is an umbrella term encompassing a group of nonprogressive damage of the immature brain, before, during, or shortly after birth, with motor disabilities as a consequence. The fact that it is a nonprogressive disorder means that the brain damage does not worsen, but secondary deformities are common. The word *cerebral* means that the brain is injured; the word *palsy* refers to a weakness in the way a person moves or positions his or her body.

Subgroups of CP have been classified according to clinical signs: spastic (70–80 percent), which is characterized by muscles that are stiffly and permanently contracted; athetoid (10–20 percent), which is characterized by uncontrolled, slow movements; ataxic (5–10 percent), which affects depth perception and the sense of balance. Depending upon which muscle groups are affected, CP may also be classified as monoplegic, triplegic, or quadriplegic, for one, three, or four limbs, respectively. Diplegic usually refers to both legs being affected and hemiplegic for one side of the body. A toddler with CP has trouble controlling the muscles of the body, the child might not be able to walk, talk, eat, or play the way most children do.

If the part of the brain that controls speech is affected, a child with CP might have trouble talking clearly. Another child with CP might not be able to speak at all. So the limitations in language development can arise as a result of the brain injury. Other factors that contribute to the language limitations are as follows.

(i) *Motor Problems*. Because the arm and hand functions are retarded, these toddlers have restricted access to their environment and therefore an impoverished experiential base for language development [2]. Another result of motor problems is that the facial, gestural, and verbal expressions of toddlers with multiple disabilities can be hard to interpret by their caregivers, making it difficult to understand what the children are trying to communicate, especially since communication with non- or hardly speaking children is highly dependent on nonverbal expressions. As a consequence, these children receive less communicative reactions than normal developing children, or only reactions that are less rich in information. This leads to further impoverishment of the child's opportunities for language development.

(ii) *The Requirement of Much Physical Care*. Because a lot of time is needed for physical care, less time and attention is left for caregivers to spend on play and communication. The toddlers miss opportunities for learning from their caregivers and surroundings, which leads to a restricted environment [6].

The limitations in language development can also have serious repercussions on other developmental areas, such as the social, emotional, and personal development, since in this age the development of all skills is interdependent. Early intervention including augmentative and alternative communication (AAC) is essential to minimize these negative impacts and stimulate the development of language, emergent literacy, and communication [7]. The effectiveness of early intervention programs can be reinforced by the use of multimedia technology [4]. Although there are many contexts in which early intervention can take place, one receiving recent attention in the AAC literature is that of interactive storybook reading [8], since it plays an important role in the early development of language and literacy skills of young children [9].

Following a short inventory study of multimedia computer programs to stimulate the language development of toddlers with multiple disabilities in the Netherlands, we concluded that multimedia technology to stimulate the language development incorporating AAC was not available. It was decided to examine the need for multimedia technology that included AAC in a preliminary study and, if such a need would exist, what guidelines for such a multimedia program would be.

In this article, we describe how this preliminary study has led to *LinguaBytes*, a three-year research program, aimed at developing an interactive and adaptive play and learning environment for stimulating the language development of toddlers with multiple handicaps. We will describe (1) the abovementioned preliminary study; (2) the development and evaluation of a first prototype, the E-Scope; (3) the development and evaluation of the follow-up prototype, *KLEEd*; and (4) the current prototype, *CLICK-IT*, along with preliminary findings.

2. Preliminary Study

The aim of the preliminary study was twofold: (1) to execute a needs assessment and define initial guidelines and (2) to build and evaluate a preliminary program based on these guidelines [10]. The methods used were literature study and expert consultation using the Delphi method [11]. The Delphi method is a method for obtaining judgments from a panel of independent experts.

The literature study showed that 50% of the non- or hardly speaking toddlers with multiple disabilities have problems with their language development and early literacy [12, 13]. The period from the prelinguistic to the linguistic period is the most important phase in language development in which the foundation for further language development is laid. Another finding was that young children are more distracted by details on two-dimensional graphics than older children [14]. The results of the literature study were used as starting points for formulating the propositions used in the Delphi method.

The invited experts (2 linguists, 5 educational psychologists and speech therapists working with toddlers with multiple disabilities, 3 computer scientists, 3 teachers in special education, and 1 industrial designer) were asked to react on propositions in two subsequent phases (36 propositions in the first phase and 27 propositions in the second phase) via the Internet. The propositions were categorized in propositions about the target group (including the need for a program), the content, AAC, graphics, devices, and instructions and support. An example of a proposition in the category "graphics" is "*To show the symbolic function of a concept an animation of the concept is needed.*" On a five-point scale the experts had to fill in how much they (dis-) agreed with the proposition and argue why. In the second phase the propositions on which almost all experts (dis-) agreed were left out. The other propositions were reformulated or specified based on arguments given in the first phase. An example of a more specified proposition in the category "graphics" is "*For each concept it should be considered if an animation supports the meaning of the concept. In case of verbs and dynamic concepts animations are needed.*

In case of nouns like "*house*" *and* "*tree*" *animations are not needed*." Anonymous summaries of the experts' opinions from the previous phase as well as the rationale behind their judgments were given. Thus, experts were encouraged to revise earlier answers in light of the replies of other members of the group.

The results of the preliminary study confirmed the need for a multimedia program, and the results of the Delphi study led to a first global set of project guidelines.

2.1. Guidelines

The guidelines concerning the development of a preliminary *LinguaBytes* program were defined as follows.

(i) *Target Group*. The *LinguaBytes* multimedia computer program should aim at toddlers with a developmental age between 1 and 4 years.

(ii) *Language Development*. The content of the multimedia program should cover the transition from the prelinguistic to the linguistic period and the linguistic period with an accent on the early linguistic period. This outcome was also supported by literature [8, 15].

(iii) *Content*. The content should contain interactive story reading and story related exercises that would provide appropriate vocabulary for the child to explore, predict, and practice.

(iv) *Levels of Difficulty/Adaptivity*. The program should offer different levels of difficulty and grow along with the developing child.

(v) *AAC*. The program should make use of AAC and at least contain *picture communication symbols* (PCSs), since this is the most used form of supported communication for toddlers.

(vi) *Independency*. The child should, as much as possible, be able to use the program independently, which means that the child should be able to start, stop, and replay parts of the program. The replay is especially important because toddlers enjoy rereading stories, which has shown to be very powerful in supporting language, emergent literacy, and communication development [16]. Because children with disabilities often lack the possibility for independent exploration, it is important to give them these opportunities. This will have a positive impact on their social emotional development.

(vii) *Graphics*. The graphics should be simple without too much distracting details; animations should be used when needed, for example, in case of dynamic concepts like verbs.

2.2. Program

To verify the results of the Delphi study a prototype was built, a computer program that contained a nine-scene story about a boy who is going to sleep (Figure 1). This subject is close to the daily experiences of the toddler. The "core words" of the story were derived from Dutch word lists. Some core words in the story like "pyjama," "sleep," and "toothbrush" were highlighted on screen using PCS. Animations were only used to illustrate dynamic concepts like "undress" and "brushing teeth." The computer program supported the use of "traditional" PC input devices (mouse, trackball, etc.). By clicking one of the navigation icons at the bottom of the screen, the toddler could stop the story, go to the next or previous scene, or replay the current scene. The program was presented in a plenary meeting with the experts that participated in the preliminary study.

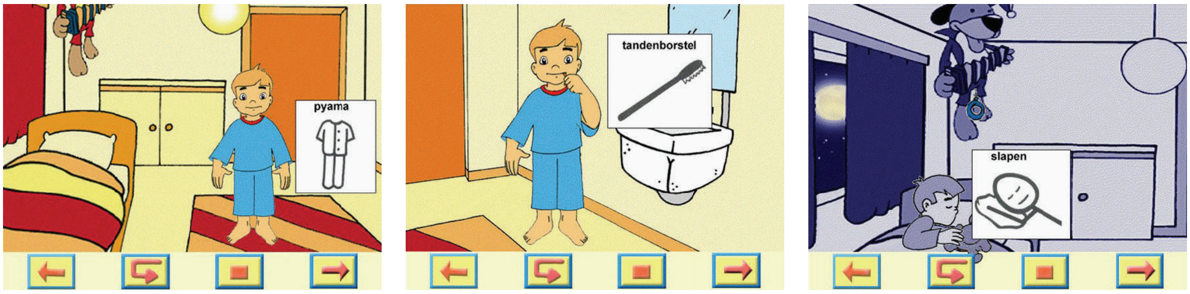
2.3. Evaluation

After demonstrating the prototype, the experts were divided into smaller groups and were asked to evaluate the prototype focusing on (1) the themes that should be incorporated in the content, (2) the graphical interface, and the (3) user interface in general. The experts were positive about the content of the story and proposed several other themes, for example, "eating and drinking," "animals," and "taking a bath." With regard to the graphical interface the experts mentioned that buttons like the stop and forward button should not be shown all the time. Concerning the design of the user interface and the hardware several more impacting aspects were indicated as follows.

(i) The program should be adjustable to the sensory-motor skills of the child to optimize the interaction for each individual child. If the designed interaction does not fit the child's skills, the child will be less motivated to engage in the program and eventually stop using it. This does not benefit the child's language development.

(ii) The program should appear to be more as a toy than PC-based computer program. This is for two main reasons. Firstly, one has to realize that practically none of the multimedia play and learning applications that have been developed for toddlers with multiple handicaps—mostly traditional, PC-based software—support the explorative natural interaction style of toddlers, making these programs less appealing than many toys. Most computer programs are assignment based, solitary and do not support the child's urge to explore. Interacting with a PC is simply not rich and social enough for toddlers. Our prototype was no exception. Secondly, the structure (menu-based decision making) and input (mostly button like) of most programs are not suitable for toddlers, due to the high cognitive load [17].

These aspects led us to conclude that, in order to stimulate the language skills of toddlers with multiple disabilities, we should design a different interaction that would be better tailored to their individual skills and needs. This should be a richer system that would facilitate active exploration and interaction with the environment, and would integrate interactive storytelling and AAC, capitalizing on new technology (embedded intelligence, sensor technology, tangible input



(a) (b) (c)

Figure 1: Screenshots from the preliminary study prototype.

systems). This could lead to improved and more effective play and learning systems for toddlers with multiple disabilities [18].

3. Follow-up Study

To research what such a product should look like, what its content should be, and how it should be used by toddlers with multiple handicaps, the *LinguaBytes* project started, a three-year research program. *LinguaBytes* aims to develop an adaptive and interactive play and learning system for stimulating the language competencies of toddlers, aged from 1 to 4 years with multiple handicaps.

3.1. Method

The development of a product for this highly heterogeneous user group is a complex process, in which numerous choices that are impacted by factors related to the child (e.g., motor, cognitive and linguistic skills, interests, and attractiveness), the therapist or parent (efficient to learn, maintain, and develop) as well as factors related to the product itself (technology, material, costs) have to be made. In order to keep this process structured and efficient, we use a constructive research method—in our field of design research more commonly known as "Research-through-Design." This is a process in which scientific knowledge is generated through iterations of designing, building, and testing experiential prototypes in real-life settings [19]. When designing interactive products like the *LinguaBytes* product, this process typically moves through several cycles of designing, building, testing, each cycle yielding refined guidelines for the content and the design of the product. This means that early iterations are often more diverging in character (focused on mapping out all aspects involved in the project) and later iterations more converging (refining within these aspects). As a consequence, research activities such as, for example, literature search, are often repeated throughout the process, on different levels of detail: from global to specific knowledge. Subsequent cycles are evaluated using process or formative evaluation [20, 21]. The aim of formative evaluation is to collect data with which the product can be improved.

Below we describe two cycles, in which two explorative prototypes of the *LinguaBytes* project were developed and tested: the *ExploraScope* and *KLEEd*. Concluding, we will outline the evolution towards a more definitive prototype, called *CLICK-IT*.

4. Explorascope

The preliminary study showed that there was a need for an early intervention multimedia program to stimulate the language competencies of toddlers with multiple disabilities in the Netherlands. It also showed that, in order for this program to be successful, it should appear to be more as a toy than computer program in the traditional sense and be highly flexible in order to create optimal learning settings for individual children. In the *LinguaBytes* project we have taken these guidelines as a starting point for actually designing such a system.

4.1. Guidelines

In order to get more insight in the scope of these guidelines we have conducted a broad literature search and consequently built and tested several cardboard models, mockups, and semifunctional 3D sketches. This resulted in an extension of our design guidelines as follows.

(i) *Playing*. Very young children learn mostly through play [22]. Play permits making mistakes and trying again. Therefore, the interaction with our system should be playful, in order to motivate the child and stimulate exploration [17, 23–25]. Within *LinguaBytes*, this could be done by taking the initiative away from the computer and giving it to the child, for example, by offering materials with which the child can control the content of the program.

(ii) *Social Interaction*. The new toy should focus on stimulating interpersonal interaction [26], because stimulating the communication between caregiver and child is essential [27–29]. This means, for example, that the *LinguaBytes* system should shift from solitary use to collaborate use, compared with PC-based programs [23, 30].

Figure 2: The E-Scope in its various configurations: (a) on the floor with pictures (top-left), (b) on a table with an integrated screen (bottom-left, mockup), (c) with a separate screen (middle), and (d) with an additional input device (right).

(iii) *Tangibility*. Especially for very young children, who naturally explore the world through play, interaction should be focused on all bodily skills. Tangible interfaces [31], for example, offer a number of advantages over the standard PC-interface, for example, stimulation of multiple senses and skills [32, 33], affording both actions and play, offering a slower pace [34] and thus more room for social interaction, a more personal interaction style, more involvement, and a more active interaction [34].

(iv) *Challenge*. The interaction should be challenging. Challenge is a key element of motivation [35]. It engages children by stimulating them to reach for the boundaries of their skills. We wish to challenge children by designing interactions that are appealing, rewarding, engaging, and fun. This also means that the content of our system should be tailored to the developmental level of the child.

(v) *Technology*. The *LinguaBytes* system should be highly adaptive to individual users to enable the diverse group of multihandicapped toddlers to use it independently. This optimizes the learning setting and avoids frustration. Supporting such adaptability requires advanced technologies, which are not capitalized on today. Embedded intelligence, wireless networking, and interactive, adaptive narratives offer possibilities for innovative designs. For example, small motors and sensors could be integrated in the interface to react to the child's behaviour during interaction, or even trigger behaviour.

(vi) *Appeal*. We wish to design products that are appealing to both disabled and able-bodied children by making products that resonate with them. Designs should be nonstigmatizing and can benefit from success formulas from the toy industry [7].

We have used these guidelines to design a new prototype called ExploraScope, or E-Scope.

4.2. Design

The E-Scope is a tangible controller that enables young children to learn simple concepts (e.g., sleep, clock, bear) through tangible interaction and play [18]. The E-Scope consists of a wooden ring-shaped toy with sensors and actuators, a computer with a wireless station, and a monitor. The E-Scope and the computer communicate through radio transceivers. All sensors, actuators, and batteries are built into the ringed layers of E-Scope.

E-Scope is adaptable to a child in the sense that it can be used in different configurations (Figure 2) to suit a child's preferred interaction style. A child can listen to stories or play educational games by rolling E-Scope over pictures that are lying on the floor. Each picture triggers a matching one-scene story. The buttons can be used for further deepening of the linguistic concepts within the scene. For example, within a scene about a goat at the farm, pushing a button can trigger auditory output (e.g., the sound the goat, the word "goat," a song of the goat) or visual output (e.g., the PCS of a goat, a different picture of a goat), or be used to highlight parts of the goat (legs, belly, tail, etc.).

Figure 3: The *KLEEd* prototype, with clockwise from the top left: the console, the navigation mat, the combination mat, tagged input material, and the hide and seek mat.

If another configuration is preferred, E-Scope can also be used on a table or other workspace. By rotating the upper part of the ring and pushing its buttons, which are connected, to individual scenes of the story, a child can interact with stories shown on an integrated or a separate screen, depending on the ergonomic and social requirements. If required, E-Scope can also be attached to alternative input devices, for example, single button or eye-movement interaction. In this last case, the upper ring is rotated by use of a motor.

4.3. Content: Story

A linear story about a girl called Jitte who is going to sleep, that is offered through the E-Scope, aims at being rich and engaging. Therefore, it uses a variety of visual and auditory outputs such as photos, drawings, symbols, sounds, and songs. The graphical style aims at being realistic for an optimal recognition of the concepts to be learned, but it lacks enough freedom to stimulate the imagination of the children.

4.4. Evaluation

The E-Scope was tested with configurations in Figures 2(a) and 2(c) with three children and three therapists in the center for rehabilitation medicine *St. Maartenskliniek* in Nijmegen, the Netherlands. Each session took half an hour and was conducted during a "regular" speech therapy session. The sessions were videotaped and the therapists were interviewed after the sessions.

The outcome was that the overall concept of the E-Scope—enabling young children to learn simple concepts (e.g., sleep, clock, bear) through tangible interaction and play [18]—was useful and promising. The children were exited by the stories and graphics and showed good concentration. The therapists were positive about the toy-like design and its playful sensorial character. They were enthusiastic about

Figure 4: The *KLEEd* prototypes in different setups.

the diversity in interaction styles but encouraged further adjustability for a more personal fit. The product should make more use of physical objects that could be adjusted to the skills (cognitive and motor) of the child. Also, the interaction style should be adaptable because for some children it would be too hard to push the (correct) buttons. Also, the graphics on the buttons were relatively small and could not be adjusted. Finally, one therapist indicated that she wanted an integrated screen to enhance social interaction by sitting opposite each other with E-Scope in the middle (configuration b). Unfortunately, integrating a circular screen in the E-Scope's ring would be a costly thing. We have not yet solved this problem but have built and tested a mockup version of this configuration. In this mockup, the E-Scope was placed in a fixed position over a circular tabletop projection. This proved to facilitate the desired eye contact with the child but of course heavily limited the freedom of moving and exploring the E-Scope.

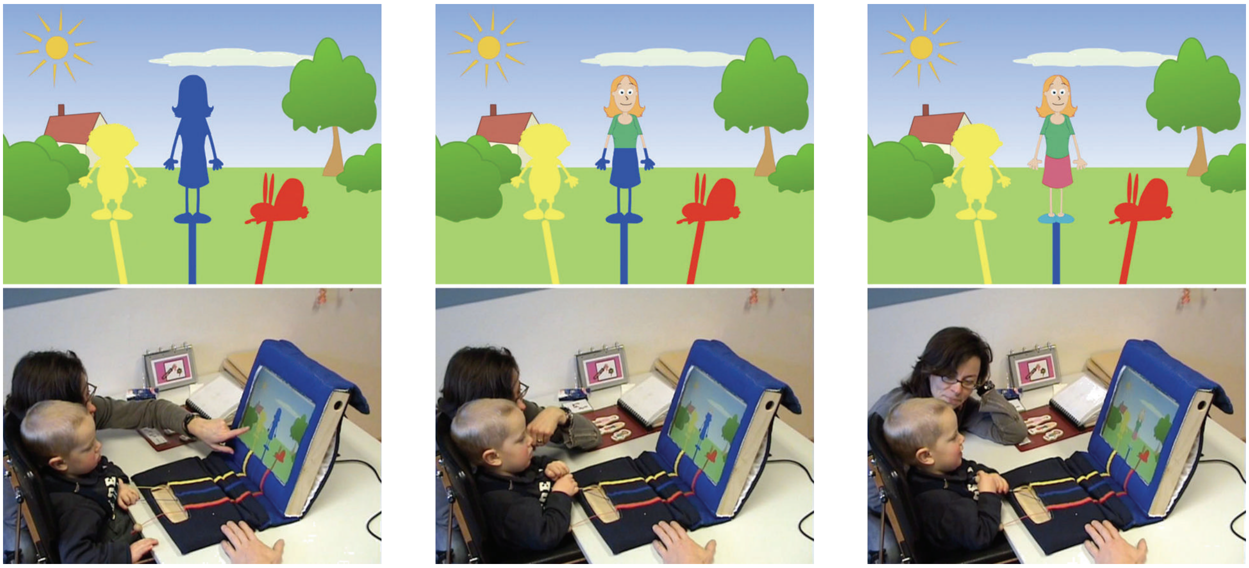
5. *KLEED*

The tests with E-Scope showed that a more playful, toy-like interface has great potential, provided it is tunable to the skills and needs of the individual child, not only cognitively, but also physically. The E-Scope already had some flexibility in terms of configuration and content, but the therapists indicated the necessity of further adjustability. Taking these results, we have elaborated our literature search for further deepening of our guidelines, and built and tested two semifunctional mockups, which led to the development of our second prototype, kids learn through engaging edutainment (*KLEEd*).

5.1. Guidelines

Taking the results of the evaluation of the E-Scope, additional guidelines were formulated as follows.

(i) *Physical objects*. The *LinguaBytes* system should allow for the use of a child's own preferred physical objects and AAC systems. This is important since not all toddlers are capable of symbolizing the world into abstract representations [36]. They should be able to use materials they know as a starting point.



(a) (b) (c)

Figure 5: The hiding exercise. From left to right: (a) the speech therapist asks the child who "the blue figure" could be, (b) the child pulls the blue string, and (c) thus reveals the blue figure.

(ii) *Adaptability and adaptivity.* This also means that the system should be flexible enough to support different input materials and levels of abstraction or difficulty. The *LinguaBytes* system could benefit from database technologies to set initial settings per child and monitor the child's development.

(iii) *Costs*. The *LinguaBytes* system should be affordable, despite being innovative. Otherwise it will not be feasible. To do so, the system could benefit from the advantages of modularity.

5.2. Design

Based on these and other guidelines, *KLEEd* was developed (Figure 3). *KLEEd* is a modular system consisting of exercise mats that can be connected to a central console, and upon these mats a standard set of tagged objects and additional tagged personal material can be used to hear and respond to interactive stories and exercises. Apart from exercise modules, a separate module for navigating through stories was developed. The modularity, tangibility, and adaptability of the system all add to its playfulness, appeal, and challenge, making it motivating for the child to use and learn.

The central console contains a 15-flatscreen monitor and electronics for connecting exercise modules to the system. The position of the monitor can be adjusted to the optimal learning settings of individual toddlers. This means that the screen can be placed in both a horizontal position, enabling the use of *KLEEd* on the floor or table, and a range of tilted positions. The central console is embedded in a sleeve of a soft and friendly material, that can be washed separately after screen and electronics are taken out.

Exercise modules can be easily attached to the console in different setups (Figure 4), enabling both individual and collaborative use, thus stimulating social interaction. Every exercise module has its own goal, for example, exercising phonological awareness (through rhymes or songs), semantics, syntax, or just free play. By giving every module its own goal, each can be designed optimally for the type of exercise, making the interaction more intuitive, engaging, and suitable for toddlers. Materials, textures, colours, sounds, and so forth will therefore vary between exercises, thus offering a wide range of sensory stimuli. Each module supports different difficulty levels, depending on the development of the child.

All parts of the prototype were made interactively using Phidgets sensors [37], Macromedia Flash, and MAX/MSP, a widely used graphical programming environment [38].

5.3. Content

5.3.1. Stories

Along with the *KLEEd* prototype two stories and two exercises have been developed within the semantic category "people." Both stories consisted of nine scenes. The first story concerns two children, Tom and Tess, playing with a ball and daddy who wants to join them, but when he does, he falls down and tears his trousers, so he has to go home to put on a new one. Some core words are "daddy," "cuddle," "play," "join," and "help." The second story also starts with Tom and Tess who are playing with a ball, but now a woman with a baby appears. The children want to see the baby and give the baby a kiss. They sing a song for the baby. Then the baby falls asleep and the woman with the baby goes home. Some core words in this story are "woman," "baby," "cuddle," "kiss," "sing," and "sleep." The core words for the stories and exercises have been chosen on the basis of the

(a) (b) (c)

Figure 6: The combination exercise. From left to right: (a) the wooden piece of "the woman" is selected, (b) placed on the wooden "subject" area, and (c) combined with the PCS verb-card "cuddle" into the sentence "the woman cuddles."

three Dutch word lists, *N-CDI* [39], the *Lexilijst* [40], and the list *Duizend-en-een-woorden* [41]. In each scene, up to three PCSs were shown to emphasize the core words in the scene.

5.3.2. Exercises

The aim of the first exercise is to stimulate vocabulary (words like "daddy," "woman," "baby," "boy"), turn-taking, cause and effect, and visual memory. In the exercise (Figure 5), three outlines of characters from the story were shown on screen. Pulling a cord on the accompanying exercise mat would "open" the corresponding character and trigger audio that said something about the character, for example, "this is daddy, daddy is a man." After this, the revealed character would be replaced by the outline of a different character.

The aim of the second exercise (Figure 6) was to construct two-word sentences. The child could choose a wooden character piece and combine it on the exercise mat with a PCS-verb card. The constructed sentence (e.g., "Tom" and "sleeping") would be animated on the screen and pronounced ("Tom is sleeping"). The aim of this exercise is to stimulate active syntax, turn-taking, and playing with elements of a sentence by letting the same character do something else or letting someone else do the same.

5.4. Evaluation

Seven children between the ages of 3 years 1 month and 6 years 1 month (with developmental ages between 1 year 5 months and 3 years 9 months) took part in the evaluation study. Four children visited a day care center for children with cognitive delays; these children had none or minor motor problems. The other three children visited a center for rehabilitation medicine. Two of these children had the diagnosis CP, the other had the diagnosis hydrocephalus. These children had moderate to severe problems with their arm and hand function: they could pick up something with their hands but were not able to pick up something with their thumb and forefinger. All seven children had a delay in language development.

For the evaluation, we used the following (1) a questionnaire concerning child data that was completed by the researcher on the basis of the children's files; (2) an observation list that was completed by the researcher to analyze the video material of the children; and (3) a questionnaire for the therapist of the child.

The most important results were as follows.

(i) The toddlers looked continuously at the screen and seemed interested in the story and the exercises. The therapists indicated that the concentration, speed of work, and motivation of the children were at least similar or better than in working with other, comparable materials. It was noted that the child's motivation could be even more enhanced by offering both child and therapist more control over the content and timing of the exercises. For this, a more extensive database should have been set up.

(ii) The children reacted well to the design, among which the use of material, colour, graphics, animations, and audio (type of voice, speed of the pronunciations). The separate modules with their own actions spaces and own goals were considered to make the interaction more intuitive for toddlers.

(iii) The use of AAC was considered satisfactory by the therapists, but from a design perspective it showed considerable drawbacks in terms of clarity and flexibility. We will address some here. Firstly, recent literature suggests (1) that it is best to offer language concepts within a visual scene [42] and (2) that it is best to show a communication symbol at the same location which the object refers to. In the case of onscreen animations this means that when, for example, the PCS of "daddy" is shown in the story, it should be placed as close to the onscreen daddy figure as possible. This however obscures part of the scene making it unclear. This effect increases with each symbol placed within the scene. Secondly, it is also preferred to reveal the communication symbol at the moment the corresponding audio is being pronounced. However, it could be seen that toddlers often looked away from the screen at this crucial moment, due to the fact that many toddlers with CP move around involuntary. This raises substantial timing problems for the animator. Thirdly, in order for the scene to be as clear as possible, it is necessary to keep the symbols small, making them harder to "read" for the toddlers, who often have visus problems. Finally, we used RFID-tagged cardboard PCS verb cards with the second exercise. However, these widespread communication symbols were often slightly customized or replaced by the caregivers or therapists in order to make them suitable for individual children. Some children did not understand the standard PCS but did understand a slightly altered version. Some children preferred using photograph representations. In other words, although the tangible symbol cards had their interaction advantages in the sense of exploration of, and control over content, it would be highly recommendable to allow therapists to customize the symbols.

(iv) The contents of the stories and exercises were considered suitable. All children liked the characters in the story; they spontaneously used the names of the children in the story. In spite of the fact that most children already mastered the core words, the vocabulary seemed to be chosen well as indicated by the therapists.

(v) The physical interaction needed for storytelling was not always suitable. One problem was that some children liked moving the story navigation handle so much that it disrupted the continuity and concentration. The tangible interaction in the combination exercise was clear for all children and provided no difficulties, not even for the multiply disabled children. The hiding exercise proved to be physically difficult for the multiple handicapped children.

(vi) The therapists indicated that they wanted to make choices in offering content to the child, so more (types of) stories and exercises should be implemented. The therapists did not give a high priority to integrating their own personalized pictures or audio in the program.

We are currently using these results for the development of *KLEEd*'s follow-up, *CLICK-IT*.

6. *CLICK-IT*

6.1. Guidelines

The results of the evaluation of *KLEEd* enabled us to refine our body of design guidelines. The most important alterations are as follows.

(i) *Control*. The *LinguaBytes* play and learning system should offer both child *and* therapist as much control over the content, exercises, and interaction itinerary as possible. This will benefit the child's comprehension of the content and the social interaction with therapist or parent.

(ii) *Adaptivity*. This means that the system should make both the software/content and hardware/interfaces highly adaptable and adaptive [43].

(iii) *Database Technology.* This means that the system should capitalize more on database technology and randomization (sounds, details, visual effects, etc.) to enhance the motivation of the child, within the constrictions of the individual learning settings.

(iv) *AAC*. The system should externalize the use of PCS or other symbols to both keep the visual content clear and give the initiative to use symbols to child and therapist/parent.

Based on this new body of design guidelines, we have recently started developing our current prototype, *CLICK-IT*.

Figure 7: Part of the *CLICK-IT* prototype.

6.2. Design

Like the previous prototype, *CLICK-IT* consists of a console, exercise modules, and a collection of input materials but the design shows the following significant changes.

(i) The exercise modules have been split into one general base unit, on top of which various interface modules can be placed.

(ii) The base unit contains most of the sensors, actuators, and processing, so that these can be used by any of the interface modules, thus reducing costs. Connecting an interface module automatically changes the setup of the base unit.

(iii) Due to various reasons, the fabrics used in the *KLEEd* prototype have been replaced by wood.

(iv) The base unit contains a slot in which the current user's identifying tag can be inserted. This will change the product's settings to fit the user optimally (level of difficulty of the content, sensitivity of sensors, etc.).

(v) To increase the flexibility of the system, more use has been made of tangible input materials. A major change in this respect is that the stories have become physical books again, that can be augmented by running them through the story reading module.

The *CLICK-IT* prototype consists of a console, a base unit, four different interface modules, a booklet of the story, and 15 input characters (Figure 7). Additionally, the 30 core words from the story are added on 7×7 cm (2.75×2.75 inch) cards, containing the word and an illustration of the word. The console contains a 17-flatscreen monitor, stereo speakers, and connectors for the base unit; the base unit itself contains various Phidgets sensors and connectors for the exercise modules and the exercise modules occasionally house additional electronics (speaker, slider, light sensor or DC motor). All parts are made of Ash wood using a 3D milling machine and plastics.

6.3. Content

The content of the *CLICK-IT* prototype was created in the same way as that of *KLEEd*: a body of core words was assembled based on the word lists mentioned earlier in this paper. This time another semantic category was chosen, "animals." In this context, a nine-scene story about a children's farm was created (Figure 8) along with seven exercises (four phonological exercises with incrementing levels of complexity, two syntax exercises, and a semantic exercise). All exercises and the stories were animated in Macromedia Flash and made interactive using MAX/MSP.

6.4. Evaluation

We are currently testing the *CLICK-IT* prototype at two centers for rehabilitation medicine in the Netherlands, with twelve children between the ages of 2 years 3 months and 3 years 10 months and a developmental age between 1 year 3 months and 3 years. At this point it is too early to draw definite conclusions, but the first signs are promising.

7. Discussion

In the introduction of this article, we have outlined some of the main factors that cause severe limitations in the language, emergent literacy, and communication development of very young children [1, 2, 6]. We have described the repercussions these limitations have on the total development of the child and identified the need for an interactive multimedia play and learning system to stimulate the linguistic development of the child and help diminish these repercussions [7–10]. A preliminary study resulted in the following two important conclusions that would be the foundation for the further development of this interactive multimedia play and learning system.

(1) The system should appear more as a toy than PC;

(2) the interface of the system should be not only adjustable to the cognitive and linguistic level of the child, but also to the child's other needs and skills (perceptual motor, social, and emotional).

These conclusions led to the iterative development of the *LinguaBytes* system, of which we have described two iterations. Throughout the paper we have highlighted the most important design guidelines that led up to our current work, *CLICK-IT*. We have summarized all guidelines in Table 1. Each column represents one iteration.

Focusing on the two, abovementioned foundation points for the development of our system, we have made some interesting observations.

Firstly, throughout our iterative process we have seen that the physical manipulability of our toy-like prototypes had the following major advantages for these children, compared to the familiar PC interface.

(i) By offering physical input materials (in the *KLEEd* prototype) and an interaction that is closer to their usual style of exploration, these children were offered more access to their environment, thus getting a richer base for language and emergent literacy development [2].

(ii) Consequently, our observations showed that the children generally had a longer attention span than usual and showed more initiative [32, 33]. These observations were confirmed in the therapists' questionnaires.

(iii) Additionally, using tangible input material slowed down the interaction, subsequently giving both children and caregivers more control over the timing of the interaction. For example, the *KLEEd* prototype clearly seemed to stimulate the communication between the therapist and the child; especially in working with the combination exercise children seem to communicate more than in other comparable situations, using the physical input material as an alternative communication means.

(iv) As a result of this, there are more opportunities for facial, gestural, and verbal expressions of the children, letting them evoke more communicative reactions of their surroundings.

Secondly, throughout our research we have seen that, in order *LinguaBytes* system to optimally fit all its different users, it is crucial that it is highly adaptable and adaptive. By the former we mean "adjustable by the user," by the latter "adjusting to the user." The *E-scope* already showed that the interactions in the different exercises should be more intuitive and suitable, that the system should allow for the use of a child's own preferred physical objects and AAC systems, and that it should offer the possibility to optimally fit each exercise to the individual child. And although these refinements were executed in the *KLEEd* prototype and the children and caregivers were positive about the content and the tangibility of the product, still they urged for further broadening of the content (more stories and exercises) and a more flexible and adaptive user interface with regard to the children's motor skills.

All this brings us to one of the major challenges of developing the *LinguaBytes* system: combining the two foundation points "more toy than PC" and "highly adaptive and adaptable." We clearly see the advantages of physical interfaces that adapt themselves to individual users, not only within our own target group but also for any other highly heterogeneous group of users. Actually developing such interfaces however becomes really complex, due to this high heterogeneity. We find incremental research or *Research through Design* [18] a helpful method for achieving this because it enables us to shift between different aspects of the design building on previously generated knowledge, slowly working toward a more or less complete body of guidelines.

Which brings us to the second challenge we wish to address with regard to interaction design: the complexity of mapping out guidelines when designing complex products such as the *LinguaBytes* system. Table 1 illustrates this quite well, since it shows the growth of our design guidelines as a rapidly expanding set. However, the guidelines it holds are still very general. The table does not incorporate detailed guidelines such as ergonomic dimensions. Of course, in order to fit Table 1 in this paper we had to cut it back to the bare minimum, but this illustrates an important thing:

Table 1: Summary of guidelines for the *LinguaBytes* product (time passes from left to right, guidelines accumulate).

Delphi study (p. 2)	Preliminary program (p. 3)	First prototype: Explorascope (p. 4, 5)	Second prototype: *KLEEd* (p. 6–9)	Towards third prototype: *CLICK-IT* (p. 9)
Target group				
Multi-handicapped toddlers with a developmental age between 1–4 years, w. language learning problems				
				Parents, therapists or caregivers of these toddlers
Language development				
Content should cover the prelinguistic/linguistic period transition and the linguistic period/early linguistic period transition				
Content should contain interactive story reading and story related exercises for the child to explore, predict, and practice			Support interactive storytelling and practicing and playing with phonology, semantics, and syntax.	
		Focus on stimulating interpersonal interaction: communication between caregiver and child		
			The interactions should support the linguistic meaning of the exercises	
AAC				
Make use of AAC (at least PCS)			Allow for the use of a child's own preferred AAC system	
				Externalize the use of PCS or other symbols
Adaptivity and adaptability				
Offer different levels of difficulty and grow along with the developing child		Interaction should be challenging, appealing, rewarding, engaging and fun.		
			Allow for the use of a child's own preferred material (e.g., personal toys)	
Independent use as much as possible (child can start, stop, and replay)	Be adjustable to the sensorimotor skills of the child	Enable different configurations (floor, table, etc.)		
		Capitalize on novel technologies, for example, embedded intelligence and wireless networking		
			Be affordable, despite being innovative	
			Benefit from modularity	
				Offer both child and therapist as much control as possible over the content, exercises, and interaction itinerary
				Externalize the use of PCS or other symbols
Graphic design				
Simple, nondistracting, animations when needed				
				Externalize the use of PCS or other symbols
	Physical design			
	Appear to be more toy than PC-based computer program			
		Appealing to both disabled and able-bodied children: nonstigmatizing		
		Benefit from success formulas from the toy industry	Offer a wide range of sensory stimuli (tactile, visual, auditory)	Use textiles only when needed
			Benefit from modularity	Redistribute modularity
		Interaction design		
		Interaction with our system should be playful in order to motivate the child and stimulate exploration		
		Taking the initiative away from the computer and giving it to the child, for example, by offering materials with which the child can control the content of the program		
		Shift from solitary use to collaborate use		
		Interaction should be focused on all bodily skills		
			Allow for the use of a child's own input material	

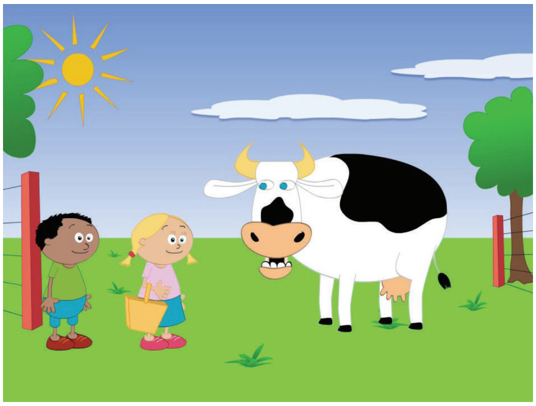


Figure 8: Screenshot from the story: Tom and Tess visit the cow.

mapping out the guidelines, requirements, and criteria for (developing) a complex product or system rapidly becomes complex itself. A table often does not suffice, due to the interdependencies of many guidelines: often, when one guideline changes it has repercussions on other guidelines, which in turn might lead to necessary changes in other guidelines. We feel that we need a new representation form, in order to keep track of all these changes. We are still investigating ways to tackle this problem, which we feel most developers of complex products or systems have encountered.

8. Conclusions

In this paper, we have outlined some of the main factors that cause severe limitations in the language, emergent literacy, and communication development of very young children, described the repercussions these limitations have on the total development of the child, and identified the need for an interactive multimedia play and learning system to stimulate the linguistic development of the child and help diminish these repercussions. We have described how we developed and tested two prototypes, ExploraScope and *KLEEd*. Concluding, we have given insights in current and future works.

The subsequent outcomes of the two prototypes indicate that the iterative process leading toward a definitive concept of *LinguaBytes* is promising. The current iteration with the *CLICK-IT* prototype as a result shows that the *LinguaBytes* system is gradually evolving into a more definitive concept and a valuable addition to the yet available early intervention products for non- or hardly speaking children with multiple disabilities.

We do feel however that in order for the system to be really effective, still more adaptivity and adaptability need to be designed and implemented. We hope the results from our current tests with CLICK-IT enable us to develop a final, fully adaptive prototype which we plan to build and test in early 2009.

Acknowledgments

The authors thank *Dr. W. M. Phelps-Stichting voor Spastici* and the consortium of additional sponsors for funding the *LinguaBytes* project. They also thank the therapists of the two centers for rehabilitation medicine, *Rijndam Revalidatiecentrum* in Rotterdam and *St. Maartenskliniek* in Nijmegen for their useful feedback and for enabling them to test the *LinguaBytes* prototypes.

References

[1] K. W. Falkman, A. Dahlgren Sandberg, and E. Hjelmquist, "Preferred communication modes: prelinguistic and linguistic communication in non-speaking preschool children with cerebral palsy," *International Journal of Language & Communication Disorders*, vol. 37, no. 1, pp. 59–68, 2002.

[2] J. Light, ""Let's go star fishing": reflections on the contexts of language learning for children who use aided AAC," *Augmentative and Alternative Communication*, vol. 13, no. 3, pp. 158–171, 1997.

[3] R. Paul, "Facilitating transitions in language development for children using AAC," *Augmentative and Alternative Communication*, vol. 13, no. 3, pp. 141–148, 1997.

[4] M. Smith, *Literacy and Augmentative and Alternative Communication*, Elsevier Academic Press, London, UK, 2005.

[5] L. Lloyd, "Comments on terminology," *Augmentative and Alternative Communication*, vol. 1, no. 3, pp. 95–97, 1985.

[6] C. Basil, "Social interaction and learned helplessness in severely disabled children," *Augmentative and Alternative Communication*, vol. 8, no. 3, pp. 188–199, 1992.

[7] J. C. Light and K. D. Drager, "Improving the design of augmentative and alternative technologies for young children," *Assistive Technology*, vol. 14, no. 1, pp. 17–32, 2002.

[8] J. L. Bedrosian, "Language acquisition in young AAC system users: issues and directions for future research," *Augmentative and Alternative Communication*, vol. 13, no. 3, pp. 179–185, 1997.

[9] C. E. Snow and A. Ninio, "The contracts of literacy: what children learn from learning to read books," in *Emergent Literacy: Writing and Reading*, W. H. Teale and E. Sulzby, Eds., pp. 116–138, Ablex, Norwood, NJ, USA, 1986.

[10] L. J. M. Van Balkom, J. M. H. De Moor, and R. Voort, *LinguaBytes. Een studie naar de ontwikkeling van een computerprogramma voor niet- of nauwelijks sprekende peuters met een motorische beperking*, EAC, Nijmegen, The Netherlands, 1st edition, 2002.

[11] M. Turoff and S. R. Hiltz, "Computer based Delphi processes," in *Gazing into the Oracle: The Delphi Method and Its Application to Social Policy and Public Health*, M. Adler and E. Ziglio, Eds., Kingsley, London, UK, 1996.

[12] W. L. Rush, *Journey out of Silence*, Media Productions and Marketing, Lincoln, Neb, USA, 1986.

[13] D. R. Beukelman and P. Miranda, *Augmentative and Alternative Communication: Management of Severe Communication Disorders in Children and Adults*, Paul H. Brookes, Baltimore, Md, USA, 2nd edition, 1998.

[14] S. C. Grover, "A field study of the use of cognitive-developmental principles in microcomputer design for young children," *Journal of Educational Research*, vol. 79, no. 6, pp. 325–332, 1986.

[15] J. E. Downing, *Teaching Communication Skills*, Paul H. Brookes, Baltimore, Md, USA, 2nd edition, 2005.

[16] P. King-DeBaun, *Early Literacy Engagement Progress Monitoring Checklist*, Creative Communicating, Park City, Utah, USA, 2006.

[17] L. Plowman and C. Stephen, "A "benign addition"? Research on ICT and pre-school children," *Journal of Computer Assisted Learning*, vol. 19, no. 2, pp. 149–164, 2003.

[18] C. Hummels, A. van der Helm, B. Hengeveld, et al., "Explorascope: an interactive, adaptive educational toy to stimulate the language and communicative skills of multiple-handicapped children," in *Proceedings of the 6th International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT '06)*, pp. 16–24, Esbjerg, Denmark, September 2006.

[19] B. Archer, "The nature of research," *Co-Design*, no. 2, pp. 6–13, 1995.

[20] B. N. Flagg, *Formative Evaluation for Educational Technologies*, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1990.

[21] J. R. Sanders and D. J. Cunningham, "A structure for formative evaluation in product development," *Review of Educational Research*, vol. 43, no. 2, pp. 217–236, 1973.

[22] J. Piaget, *Play, Dreams, and Imitation in Childhood*, W. W. Norton, New York, NY, USA, 1962.

[23] M. D. Brouwer-Janse, J. F. Suri, M. Yawitz, G. de Vries, J. L. Fozard, and R. Coleman, "User interfaces for young and old," *Interactions*, vol. 4, no. 2, pp. 34–46, 1997.

[24] J. P. Isenberg and N. Quisenberry, "Play: essential for all children. A position paper of the Association for Childhood Education International," *Childhood Education*, vol. 79, no. 1, pp. 33–39, 2002.

[25] D. R. Anderson and M. K. Evans, "Peril and potential of media for infants and toddlers," *Zero to Three*, vol. 22, no. 2, pp. 10–16, 2001.

[26] L. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press, Cambridge, Mass, USA, 1978.

[27] C. E. Snow and C. E. Ferguson, *Talking to Children*, Cambridge University Press, Cambridge, Mass, USA, 1977.

[28] E. Bates, *The Emergence of Symbols: Cognition and Communication in Infants*, Academic Press, New York, NY, USA, 1979.

[29] M. Tomasello, *Constructing a language. A Usage-Based Theory of Language Acquisition*, Harvard University Press, Cambridge, Mass, USA, 2003.

[30] D. Stanton and H. R. Neale, "The effects of multiple mice on children's talk and interaction," *Journal of Computer Assisted Learning*, vol. 19, no. 2, pp. 229–238, 2003.

[31] H. Ishii and B. Ullmer, "Tangible bits: towards seamless interfaces between people, bits and atoms," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*, pp. 234–241, Atlanta, Ga, USA, March 1997.

[32] V. Bayon, J. R. Wilson, D. Stanton, and A. Boltman, "Mixed reality storytelling environments," *Virtual Reality*, vol. 7, no. 1, pp. 54–63, 2003.

[33] D. Stanton, V. Bayon, H. Neale, et al., "Classroom collaboration in the design of tangible interfaces for storytelling," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*, pp. 482–489, Seattle, Wash, USA, March-April 2001.

[34] E. Strommen, "When the interface is a talking dinosaur: learning across media with ActiMates Barney," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)*, pp. 288–295, Los Angeles, Calif, USA, April 1998.

[35] T. W. Malone and M. R. Lepper, "Making learning fun: a taxonomy of intrinsic motivations for learning," in *Aptitude, Learning and Instruction*, R. E. Snow and M. J. Farr, Eds., vol. 3 of *Cognitive and Affective Process Analyses*, pp. 223–253, Lawrence Erlbaum, Hillsdale, NJ, USA, 1987.

[36] A. Bruckman and A. Bandlow, "HCI for kids," in *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, J. Jacko and A. Sears, Eds., Lawrence Erlbaum, Hillsdale, NJ, USA, 2002.

[37] Phidgets, March 2006, http://www.phidgets.com.

[38] Cycling '74, October 2006, http://www.cycling74.com/products/mmjoverview.

[39] I. Zink and M. Lejaegere, "N-CDI's: Lijsten voor communicatieve ontwikkeling," in *Aanpassing en Hernormering van de MacArthur CDI's van Fenson*, Acco, Leusden, The Netherlands, 2002.

[40] J. E. P. T. Schlichting, M. C. M. van Eldik, H. C. Lutje Spelberg, S. Van der Meulen, and B. F. Van der Meulen, *Schlichting Test voor Taalproductie*, Swets Zeitlinger, Lisse, The Netherlands, 1995.

[41] S. Bacchini, T. Boland, M. Hulsbeek, and M. Smits, *Duizend-en-een woorden. De allereerste Nederlandse woorden voor anderstalige peuters en kleuters*, SLO, Stichting Leerplanontwikkeling, Enschede, The Netherlands, 1st edition, 2005.

[42] J. Shook and W. Coker, "Increasing the appeal of AAC technologies using VSD's in preschool language intervention," in *Proceedings of the 22nd Annual International Technology and Persons with Disabilities Conference*, Los Angeles, Calif, USA, August 2006.

[43] B. Hengeveld, C. Hummels, K. Overbeeke, H. van Balkom, R. Voort, and J. de Moor, "Adapt or be adapted," in *Proceedings of the 2nd International Workshop on Physicality (Physicality '07)*, pp. 27–32, Lancaster, UK, September 2007.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 597971, 6 pages
doi:10.1155/2008/597971

Research Article

Developing a Child Friendly Text-to-Speech System

Agnes Jacob and P. Mythili

Division of Electronics, School of Engineering, Cochin University of Science and Technology, Kochi 682022, Kerala, India

Correspondence should be addressed to Agnes Jacob, nirag_2007@rediffmail.com

Received 11 November 2007; Revised 5 June 2008; Accepted 28 August 2008

Recommended by Owen Noel Newton Fernando

This paper discusses the implementation details of a child friendly, good quality, English text-to-speech (TTS) system that is phoneme-based, concatenative, easy to set up and use with little memory. Direct waveform concatenation and linear prediction coding (LPC) are used. Most existing TTS systems are unit-selection based, which use standard speech databases available in neutral adult voices. Here reduced memory is achieved by the concatenation of phonemes and by replacing phonetic wave files with their LPC coefficients. Linguistic analysis was used to reduce the algorithmic complexity instead of signal processing techniques. Sufficient degree of customization and generalization catering to the needs of the child user had been included through the provision for vocabulary and voice selection to suit the requisites of the child. Prosody had also been incorporated. This inexpensive TTS system was implemented in MATLAB, with the synthesis presented by means of a graphical user interface (GUI), thus making it child friendly. This can be used not only as an interesting language learning aid for the normal child but it also serves as a speech aid to the vocally disabled child. The quality of the synthesized speech was evaluated using the mean opinion score (MOS).

1. Introduction

There are various critical factors to be considered while designing a TTS system that will produce intelligible speech. Any TTS should be appealing to the child user whether it is used as a language learning aid or as a vocal aid. Children find learning more fun when their typed inputs are mapped to vocalized outputs. The first crucial step in the design of any concatenative TTS system is to select the most appropriate units or segments that result in smooth concatenation. This involves a tradeoff between longer and shorter units. In the case of shorter units, such as the phonemes, less memory is required. But, the sample collection and labeling procedures become more complex. The number of segments and the time required to cover the language increase steadily from word to the phoneme. In addition to being part of computationally manageable inventory of items, the synthesis segments chosen should capture all the transient and transitional information. The latter had been emphasized throughout this work, which in turn contributed to the smooth concatenation of speech segments in this TTS.

Even though speech is analog, phonemes are discrete. Inclusive of allophones, phonemes are less than hundred and are mainly the vowels, diphthongs, and consonants [1]. These allophones can be concatenated to produce smooth utterances without enormous computational effort compared to concatenation from the basic set of just 44 phonemes. Although phoneme appears to be an attractive linguistic unit for speech synthesis because of its limited number, most efforts [2] to string them together have failed. Pronunciation of phonemes depends on contextual effects, speaker's characteristics, and emotions. During continuous speech, the articulator movements depend on the preceding and the following phonemes. This causes some variations on how the individual phoneme is pronounced which lead to spontaneous variations in phoneme quality that is often known as coarticulation [3]. Thus, as per available facts [4], phoneme-sized building blocks were found to be unsatisfactory as synthesis segments because of the coarticulatory effects of the adjacent sounds. Further, one of the problems associated with segmenting words and storing the excised phonemes is the preservation of the characteristics of the sound, which is present at the transitions at the beginning

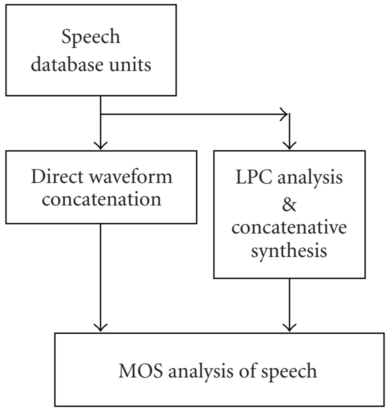


Figure 1: Overview of the TTS implementation.

and end of the segment. The characteristic sound of these transitions could be lost from both segments if they are smoothed together using signal processing techniques, resulting in a loss of naturalness in the utterance [5]. Hence, a major challenge in this method is that the boundaries between the phonemes correspond to areas that are acoustically volatile. Speech synthesized with phonemes as units is intelligible when each phoneme is represented by several allophones in the segment database. Different emotions and speaker characteristics could be implemented with such a database.

Figure 1 shows the major steps for implementing this work by the two methods. In the first method, a smooth, direct waveform concatenation of phoneme segments had been done, maximizing the speech quality in terms of naturalness and intelligibility. In the second method, LPC had been used to reduce memory requirements. A comparison of the above two methods, in terms of performance as well as the resources used, had been done using MOS. It was found that both methods gave a good quality TTS for children, which are not overloaded by too many analytical aspects and can be simulated in a short period compared to any other TTS. Sufficient care of the various linguistics aspects ensured natural sounding speech, inclusive of emotions.

2. Text/Vocabulary Selection

Selecting the vocabulary of the TTS is often referred to as fixing the target, since each of the utterances, for which this TTS system is designed, is called target and it is the speech output corresponding to a phonemic input. In the text selection phase, the target vocabulary as well as pitch, amplitude, and duration or speed of utterance was chosen. These varied styles correspond to different emotions. Being an experimental work with phonemes, the target was chosen to be a limited vocabulary of 635 different words covered by thirty-eight different phonemes and their allophones. Restriction of the targets words to a specific domain was seen to give better performance, since ultimately, the prosody is limited to the extent embedded in the recorded speech from which the segments are excised. Moreover, limiting the variety of utterances to one's anticipated needs can help one to provide sufficient samples of the segments without unnecessarily increasing the memory requirements of the database. This TTS had been initially designed to meet the minimal vocal requirements of a vocally impaired child. Alternately, it can also be modified to suit the language learning requirements of a normal child.

3. Text-To-Phoneme Conversion

The input text should be valid in the sense that it should belong to the set of target words. An error message is displayed in the graphical user interface (GUI), for cases falling outside the predetermined vocabulary (set of fixed inputs for which the speech database is designed). This project uses an event-driven front-end, which is simulated by means of a MATLAB program to transcribe input target words (the decided vocabulary) into their corresponding different phonemic forms. The international phonetic transcription (IPA) symbols are used. A dictionary-based approach has been used here for text-to-phoneme conversion. The pronunciation dictionary is initially designed in the context of the 635 different, typed input words only. The algorithmic complexity in the TTS design is considerably reduced due to this dictionary-based approach. The program stores words along with their correct pronunciation. This method is accurate and quick. The transcriptions are obtained from the sixteenth edition [6] of the *English Pronouncing Dictionary*. The pronouncing dictionary of this TTS provides essential information such as pronunciation of proper names and variant pronunciation than is usual in a general dictionary. Therefore, the primary aim of this TTS dictionary is to list pronunciation likely to be used by learners such as one that reflects the regional accent.

After considerable linguistics study and based on the findings from the literature survey, certain words are selected for the text corpus (to be recorded). These words are constituted by the phonemes of desired interest (present in the target words). Choosing specimen words for recording, followed by segmenting of these word recordings to extract the desired phonemes is the next crucial task and is related to linguistics.

4. Design of the Text Corpus

Careful design of the text corpus is an essential prerequisite to speech database design and involves the preparation of an inventory of words, which have the same phonemic constituents (along with contexts) of the target words. Care is taken to include three to five examples of words with the same allophone and these words are recorded. Even though the simplest approach is to add data to the speech database till the quality does not improve anymore, in all practical cases, there is a tradeoff between the quality and quantity. Each of the constituent phoneme segments of the target words is examined, and care is taken to select and record words having these very features. Examples of certain "design rules" formulated on the basis of the linguistics study are

Table 1: The design process.

Word (target)	Transcription	Words to record	Description
Eye	aI	Buy, why, bide B | aI, w | aI, b|aI|d	Diphthong. greater duration on first sound a: cases favored are word final or followed by voiced consonant.

given below [7]. These rules are also followed when selecting segments for concatenation as follows.

(1) A vowel preceding a voiced consonant in the same syllable tends to be longer than the same vowel preceding a voiceless consonant.

(2) The length of long vowels and diphthongs is very much reduced when they occur in syllables closed by consonants such as /p, t, k, s, h/.

(3) The consonant letter y can act as a vowel and as a consonant.

(4) The length of a phoneme is the least when it is in the middle of the word and maximum at the end.

The use of stressed words served to increase the duration of the phonemes by more than 20%. Since content words (words that are important for their meaning, e.g., nouns, adjectives, adverbs) are stressed, the text corpus to be recorded is constituted mostly by content words.

Based on the above criteria, a table was prepared with each row containing a word to be synthesized, its phonemic constituents, and examples of words with these constituent phonemes in varied positions in the word. A sample is shown in Table 1 above. The database is designed to offer sufficient coverage of the units to make sure that an arbitrary input sentence can be synthesized with more or less homogeneous quality.

5. Phoneme Database Development

A series of preliminary stages have to be fulfilled before the synthesizer can produce its first utterance. The database of WAV files is obtained by recording the natural voicing of the targets. A sampling frequency of 22.05 kHz was used to make the synthesized voice sound more pleasant. An amplitude resolution of 16 bits was used [8]. The recordings were done in male and female voices. Repeated segmentation of these speech files was done to excise phonemes/allophones from the speech database. As the output quality of any concatenative speech synthesizer relies heavily on the accuracy of segment boundaries in the speech database [9], manual method of segmentation was used.

In this work, the coarticulatory effect was put to good use by excising phonemes from different environments (surrounding phonemes), adding to the variability and naturalness of the database. These allophone segments were also stored as WAV files after appropriate labelling.

6. Concatenation

Relevant literature cites that concatenating and modifying the prosody of speech units without introducing audible artifacts are difficult [3]. In this work, this problem was overcame by appropriate linguistic design of the text corpus and careful preparation of the speech database. Moreover, the acoustic inventory used consists of a rich storage of needed allophones rather than phonemes. Several linguistic rules have been closely followed. For instance, we apply that vowels are key components [2] determining the synthesized voice quality. After manually editing the WAV files and trying out the direct waveform concatenation to identify the right constituent segments for any word in the predetermined vocabulary, the appropriate segments are concatenated programmatically to yield the synthesized speech. Sentences could also be synthesized with the prosody corresponding to those embedded in the segments [10]. Sentences made from segments of longer duration give rise to slow utterances and correspond to sad emotions, while sentences from short duration segments give rise to fast utterances corresponding to any happy, energetic person. These varied styles could be chosen using tags in the text entry. Allowing the choice between male and female voices provided an additional degree of customization [11]. A female voice, sampled at 22.05 kHz, was played back at a reduced sampling frequency of 17.5 kHz in order to produce a distinct male voice. Such voice conversion attempts proved to be amusing to children, instilled in them curiosity about the mechanism of speech production and improved their intellectual ability. This TTS system is a brilliant way to expose interested students to the basics of phonetics and motivates them to setup any TTS of their choice.

6.1. Parametric Representation of Speech

The second method uses the linear predictive coded (LPC) speech [3], wherein wave files are replaced by parametric models. In the LPC method of resynthesis, the voicing, pitch, gain, and LPC parameters were found for the down-sampled versions of the above constituent wave files, and speech was resynthesized. The source filter model [12] of speech production used here hypothesizes that an acoustic speech signal can be seen as a source signal (the glottal source, or noise generated at a constriction in the vocal tract), filtered with the resonances in the cavities of the vocal tract, downstream from the glottis, or the constriction. The LPC model was used for the prosody modification as it explicitly separates the pitch of a signal from its spectral envelope. Speech is parameterized by an amplitude control, voiced/voiceless flag, Fundamental Frequency (F0), and filter coefficients at a small interval. The F0 is the physical aspect of speech corresponding to perceived pitch. As a periodic signal, voiced speech has spectra consisting of harmonics of the fundamental frequency of vocal fold vibration. The loudness control is determined from the power of speech at the time frame of analysis. The concatenative approach to speech synthesis requires that speech samples should be stored in some parametric representation that will be suitable for connecting the segments and changing the signal characteristics like loudness and F0. F0 extraction algorithms determine the voicing of speech as well as the fundamental frequency.

The LPC analysis is done for the same candidate WAV files used in the direct waveform concatenation. The pitch detection function (program) developed in MATLAB takes the speech audio signal and divides it into 30 milliseconds frames, over which speech is quasistationary. These overlapping frames start every 15 milliseconds and were further Hamming windowed to avoid distortion [3]. The function returns the pitch value in hertz for voiced frames, whereas it returns a zero for unvoiced frames. Monotone speech was produced from the synthesizer by replacing the pitch signal calculated by the function with a vector of constant values. The constant values, that were selected, were 100 Hz and 380 Hz, which correspond to male and child voices, respectively.

The Synthesis function returns the reconstructed audio. Using the voicing, pitch, gain, and LPC coefficients, each frame was synthesized. These were then put together to form the synthesized speech signal. The initial 30 milliseconds signal was created based on the pitch information. If the pitch is zero, the frame is unvoiced. This means that the 30 milliseconds signal needs to be composed of white noise. White noise is noise with a flat spectrum (uniform power spectral density) over the entire frequency range of interest. The term "white" is used in analogy with white light, which is a superposition of all visible spectral components.

Unvoiced excitation is usually modelled as such a white noise limited to the bandwidth of speech [3]. In this implementation, MATLAB function "randn" was used for producing white noise. For nonzero pitch, a 30 milliseconds signal was created with pulses at the pitch frequency. These initial signals were filtered using the gain and filter coefficients and then connected together in overlapping frames, for smooth transition from one frame to the next.

6.2. MOS Evaluation

Evaluating synthetic speech formally is difficult as there are many complex factors, dealing with intelligibility, naturalness, and the flexibility to simulate different voices and speaking rates. Due to lack of suitable standards for comparison, objective methods could not be used in this work. Hence, evaluating synthetic speech output was almost exclusively a subjective process. Certain subjective tests such as the dynamic rhyme test (DRT) are not realistic for practical application [3]. Therefore, mean opinion score (MOS) [13] has been used to evaluate the quality of this TTS, mainly in terms of intelligibility and naturalness. We have used the five level scales given in Table 2 as they are easy and provide some instant, explicit information.

An MOS rating greater than 4 indicates good quality. Any rating between 3.5 and 4 indicates that the utterance possesses telephonic communication quality. Ten volunteers without any known hearing disabilities participated in the MOS evaluation of the outputs of both phases. The listeners were all nonnative speakers of English. As is required, none were experts in TTS. There were five teachers (one of them well versed in linguistics), two 8-year-old kids (who are used to synthetic voices), two doctors, and 1 person who had recovered from a voice loss recently. The participants were briefed about the project.

Table 2: Scales used in MOS.

Rating	MOS
1	Bad
2	Poor
3	Fair
4	Good
5	Excellent

Table 3: Sample MOS for sentences.

No.	Test sentences	MOS rating
1	Where are you going?	4.5
2	Please leave me alone.	4

Table 4: MOS for WAV concatenation.

No.	Words	MOS
1	Fine	5
2	Bill	4.9
3	Queue	4

7. Results and Discussion

Synthesis of polysyllabic words was done without any difficulty. Each participant randomly selected and listened to 25 words. The average MOS rating for each stimulus (TTS utterance) was later calculated and tabulated. As these tests were aimed at assessing the segmental as well as the suprasegmental characteristics, [14] the volunteers took further listening tests of a prepared list of 20 sentences and were asked to rate these using MOS. A sample is given in Table 3. The MOS tests were conducted individually for each listener. After hearing a test stimulus, the listener indicated his/her rating on a 5-point scale. The tests for both methods were administered at a stretch.

A sample of the MOS [13] evaluation of direct waveform concatenation of phonemes is as given below in Table 4. The average MOS ratings indicated that the speech output of the TTS is of good/"toll" quality.

However, as expected, results of the MOS evaluation of the LPC-based, parametric, concatenative TTS indicate that no words are rated as excellent. Listeners unanimously stated that the intelligibility of words increased considerably when these were embedded in sentence utterances rather than in isolation.

The memory comparison given in Table 5 highlights the advantage of incorporating the LPC model. There is considerable saving of memory though at the expense of the quality of the synthesized speech. The average memory gain factor was found to be 8.96, thus justifying the use of LPC as a means to achieve database compression. This is a significant

Table 5: Sample comparison of memory for LPC and WAV file storage.

Word	Constituent files	(X) WAV file storage (bytes)	(Y) LPC parametric storage (bytes)	(X)/(Y) gain factor
SO	S01SJ8 O1O8	24000	2760	8.7

achievement compared to the vast memory requirements of conventional unit selection-based methods.

Additionally, it has been found that the target vocabulary could be generalized in the sense that using the database of phonemes suitable to produce the predefined vocabulary, many other words could also be produced by suitable concatenation. Hence, this implementation is efficient.

As such, it can be implemented by any person with a basic knowledge in linguistics and programming. Since all can use the same basic program, this can, therefore, be self-implemented by students with minimal guidance from a tutor. Since children are often more receptive to certain voices like those of their teacher or parent, the database can be recorded in any of their preferred voices for better and enjoyable learning. The vocabulary also can be chosen as per the preferences of the child. This TTS in one's own voice or in any other preferred voice motivates the child learner to try out new words. The provision to record user's own voice and compare it with the TTS utterance provides sufficient motivation for the child learner to expand his vocabulary; if one keeps track of correct utterances of TTS and assigns appropriate scores to the user. Thus word building exercises can be made more interactive and amusing, as only correct words will be vocalized, whereby children can feel out the word and get immense satisfaction with each complete utterance.

Unlike unit selection concatenative systems [15] which make use of varied speech units and mostly engineering techniques like cost optimization and signal processing, this TTS implementation minimized its algorithmic complexity primarily by incorporating appropriate linguistic aspects like coarticulation and a carefully designed database to ensure smooth concatenation. Further reduction in algorithmic complexity was achieved by using table lookup methods for the grapheme to phoneme conversion. Though there are various speech aids for the vocally handicapped, any person who once possessed the ability to speak would normally prefer to use his/her own voice compared to any other robotic voice. Hence, this will also help children facing the risk of an impending vocal impairment due to some illness. Besides, this sort of speech synthesizer requires no additional expense for a person with a good computer along with speakers. With the proliferation of laptops and notebook computers, mobility is also not a problem.

8. Conclusion

In this paper, the implementation details of a child friendly phoneme-based concatenative TTS, with sufficient degree of customization and which uses linguistic analysis to circumvent most of the problems of existing concatenative systems, have been presented. The use of a dictionary-based approach for text-to-phoneme conversion along with a tailored speech database helps to avoid all algorithmic complexities and concatenation mismatches, characteristic of existing TTS systems. While conventional unit selection-based TTS requires hundreds of megabytes of memory, this TTS required only hundreds of kilobytes of memory.

Voice conversion feature has been incorporated in this TTS using the LPC method with provision for varying the voice quality over a wide range by varying the F0 values in the synthesis stage. Another feature of this work is that it was implemented using female voice, whereas most of the successful LPC-based TTSs have been implemented in male voice. This TTS further has add-on facility in that new words can be synthesized, after adding these words, their transcription, and constituent wave file names to their respective databases. The prosody of the utterance can be designed to vary depending on the nature of the recordings in the speech database from which the phoneme segments are excised. Thus, a simple, flexible, and efficient TTS that can be user defined has been setup with minimum resources to serve multiple purposes. Though this had been developed for English, it can be suitably modified for any other language. This TTS was found to be a successful vocal aid/language learning aid as the users were able to get a real feel of phonemes, the most basic speech units. The learning environment can be conditioned to any particular accent by using an appropriate combination of database and pronunciation dictionary. Alternately, content specific learning too can be encouraged implicitly. By suitable design of the TTS vocabulary and database, a child can be familiarized with all common terms associated with any specific topic. Thus, such a TTS helps the child user get acquainted with the regular as well as any other selective vocabulary.

References

[1] T. Parsons, *Voice and Speech Processing*, McGraw-Hill, New York, NY, USA, 1987.

[2] E. Keller, Ed., *Fundamentals of Speech Synthesis and Speech Recognition*, John Wiley & Sons, New York, NY, USA, 1994.

[3] D. O'Shaughnessy, *Speech Communications: Human and Machine*, Cambridge University Press, Cambridge, UK, 2001.

[4] C. Rowden, *Speech Processing*, McGraw-Hill, New York, NY, USA, 1992.

[5] T. Dutoit, *An Introduction to Text-to-Speech Synthesis*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1976.

[6] D. Jones, *Cambridge English Pronouncing Dictionary*, Cambridge University Press, Cambridge, UK, 2003.

[7] T. Balasubramanian, *A Textbook of English Phonetics for Indian Students*, Macmillan India, New Delhi, India, 2003.

[8] VoiceSynthesis, http://www.hitl.washington.edu/scivw/EVE/.

[9] M. Ostendorf and I. Bulyko, "The impact of speech recognition on speech synthesis," *IEEE Communications Magazine*, pp. 99–104, 2002.

[10] M. Tatham and E. Lewis, "Improving text-to-speech synthesis," *Proceedings of the Institute of Acoustics*, vol. 18, no. 9, pp. 35–42, 1996.

[11] A. S. Black, P. Taylor, and R. Caley, "The Festival Speech Synthesis System," http://www.festvox.org/festival/.

[12] A. M. Kondoz, *Digital Speech Coding for Low Bit Rate Communication*, John Willey & Sons, New York, NY, USA, 1994.

[13] C. Delogu, A. Paoloni, and P. Pocci, "New directions in the evaluation of voice input/output systems," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 4, pp. 566–573, 1991.

[14] Y. Sagisaka, "Speech synthesis from text," *IEEE Communications Magazine*, vol. 28, no. 1, pp. 35–41, 1990.

[15] T. Toda, H. Kawai, M. Tsuzaki, and K. Shikano, "Unit selection algorithm for Japanese speech synthesis based on both phoneme unit and diphone unit," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, vol. 1, pp. 465–468, Orlando, Fla, USA, May 2002.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 284056, 7 pages
doi:10.1155/2008/284056

Research Article

Exploring Children's Requirements for Game-Based Learning Environments

Tuula Nousiainen[1] and Marja Kankaanranta[1, 2]

[1] *Agora Center, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland*
[2] *Institute for Educational Research, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland*

Correspondence should be addressed to Tuula Nousiainen, tuula.nousiainen@jyu.fi

Received 8 October 2007; Revised 21 April 2008; Accepted 24 July 2008

Recommended by Adrian Cheok

End users' expertise in the development of new applications is acknowledged in user-centered and participatory design. Similarly, children's experience of what they find enjoyable and how they learn is a valuable source of inspiration for the design of products intended for them. In this paper, we explore experiences obtained from collaboration with elementary school children in the design of learning environments, based on three projects and three requirements gathering techniques. We also discuss how the children experienced the participation. The children's contribution yielded useful, both expected and unanticipated, outcomes in regard to the user interface and contents of the learning environments under development. Moreover, we present issues related to design collaboration with children, especially in terms of the children's feeling of ownership over the final outcome.

1. Introduction

The potential of involving children as active contributors in technology design is being increasingly acknowledged (e.g., [1–3]). Children are no longer seen only as research objects or as a passive target group for the development of new technologies. Instead, similarly as the end users' knowledge about, for example, their work practices is valued in participatory design with adults (e.g., [4]), also children's expertise about the issues they are familiar with is considered valuable. Several researchers (e.g., [5, 6]) have emphasized the need to better understand children's needs as technology users as their ways of interacting with technology often differ fundamentally from those of adults, for example, in terms of curiosity and tendency to explore, and their preference of working together. Moreover, children's wishes especially regarding the fun and motivating aspects may be difficult for adult designers to envision [3, 6]. A recently emerged perspective to designing applications for children is child-computer interaction (CCI), which applies HCI principles and participatory methods to this specific context (e.g., [6]). Moreover, the design of educational software has broadened the HCI perspective to bring in principles of pedagogical design: an approach referred to as learner-centered design (LCD) has been introduced as a way of bringing together HCI on the one hand and educational and developmental principles on the other [7, 8]. Additionally, it is worth looking into areas such as child-centered pedagogy and sociology for a multidisciplinary perspective on children's active citizenship and empowerment (e.g., [9, 10]).

Traditional user-centered design has been criticized for placing the users in the role of mere reactors to suggested solutions, not initiators of ideas (e.g., [11, 12]). In other words, the users' contribution is minimal—or left out altogether—in all other design phases except for the testing and evaluation of solutions in different iterations. To respond to this concern, researchers have developed and applied various requirements gathering techniques aiming to allow the users to initiate ideas early in the design process. As an alternative for traditional interviewing in the gathering of early requirements, for example, several studies have employed techniques specifically tailored to children. These include, for example, photo diaries [13], Kid Reporter in which children create "newspaper articles" related to the themes of the application [14], and "Mission from Mars" entailing an innovative interview technique and enhanced with photo collage creation [15, 16]. Moreover, more specific ideas for the functions and the appearance of the application being developed have been explored through different creative techniques based on drawings [3, 17], storyboarding [18, 19], and collaborative low-tech prototype creation [1, 20, 21].

Gathering ideas from children early in the design process has yielded useful insights into what children want in technology in general or in a specific type of application. Druin et al. [21], for example, discovered that children want especially control, variety, social interaction, and creative tools, and that they pay attention to the appearance, learnability, and "coolness" of an application, as well as to how rich it is in terms of the use of multimedia. Children's early involvement in requirements gathering has revealed clues also about, for example, gender differences in preferences related to technology, children's navigation skills, ways of presenting textual information, application-specific content-related preferences, the variety of elements to be included in user interfaces and their structures, and children's desire to personalize their applications [14, 15, 17–19].

The aims of the study presented in this paper were twofold. Firstly, the goal was to develop game-based learning environments that would respond to children's requirements, and secondly, the goal was to analyze the development process from the perspectives of both the developers and the children. This paper focuses on examining the significance of children's participation for the collaborative technology design process and for the attributes of the products. A specific feature was to carry out the design project in authentic school environments, which required the development of collaborative school-based methods for child-centered design. We address issues yielded by three idea gathering practices that were used in the projects, namely, user interface drawings, idea map creation, and the evaluation of existing learning environments.

2. Research Process

The principal research approach in our study was development research. It is employed in studies of educational interventions, addressing either the intervention itself, the process of developing it, or both [22, 23]. Using the principles of development research and case study research [24], we have studied the development processes of three game-based learning environments or websites for children. The three applications discussed in this paper are Talarius, Virtual Peatland, and Kids' Site (Child Ombudsman's website for children).

Talarius is a software tool with which children can create and play educational computer board games. Working in pairs or small groups, the children gather material relevant to the topic of their game (factual texts, images, sounds, videos). Based on this material, they create questions, design a game board, and play games created by their classmates. This paper deals with the development of the pilot version of the application, which took place in the academic year 2003-2004. A school class (23 children, ages 11-12) participated in the project from the requirements collection to the evaluation of the final outcome.

Virtual Peatland is a web-based learning environment about peatlands. It includes text-based information sections as well as interactive and game-like parts (Peatland Adventure game and various quizzes). In the years 2005 and 2006, an elementary school class (approximately 25 children at a time, ages 11-12) participated in the process through various workshop sessions, each of which dealt with a specific aspect related to the structure of the learning environment or presentation forms to be used in it.

Kids' Site is a part of the Finnish Child Ombudsman's website, aiming to make children aware of the UN Convention on the Rights of the Child, to enable them to have a say in issues concerning them, and to share stories about their everyday lives. A group of 7–9-year-old children from an after-school care facility took part in the project at different stages in the years 2006 and 2007. The group size varied from four to eight children between sessions. Additionally, a brief paper questionnaire about ideas related to the website was answered by 25 children.

In each case, the data consisted of development documents, outcomes of the design processes, interviews and questionnaires to the developers and the child participants, and observation of the design process documented in research journals. Experiences related to the development process were gathered both from the perspective of the children and that of the developers. The data was analyzed using a framework consisting of several research questions as a basis of analysis. In terms of this paper, from the children's point of view, the research questions were principally related to whether they felt that their ideas had had an effect on the final outcome, how they perceived their own expertise in the project, and how they experienced the participation activities in general. From the developers' perspective, it was examined how they felt the children's participation and ideas aided their work and affected their development solutions. Additionally, design session outcomes were compared to the final products in order to examine how the ideas provided by the children manifested in the final applications.

3. What Children Want in Game-Based Learning Environments

Several different idea gathering techniques and practices were used in the course of the projects. In this paper, we address three of them, namely, user interface (UI) drawings, idea maps, and evaluations of existing learning environments, presenting issues discovered through these activities and how they guided the development of the applications. They represent different approaches to idea generation: the drawings start from scratch and leave plenty of room for imagination, the idea maps entail collaborative and gradual development of ideas, and the evaluation of learning environments provides a collection of different existing solutions to draw upon. Additional ideas and feedback were obtained through prototype evaluation and field trials, but in this paper we focus on ideas obtained before the development of prototypes of any kind.

3.1. Ideas from UI Drawings

At the beginning of the Talarius project, the children expressed their ideas about the appearance and functionality of the application by drawing UI sketches during their art class. They could choose whether they wanted to draw the game-playing mode or the game editor. Each child created an individual drawing. The rationale for having children make drawings is a wish to allow them to design something new instead of just reacting to existing suggestions [17]. Drawing UI concepts suits the art class context well but there are certain challenges related to their use. Firstly, there is a risk that the children pay a great deal of attention to details while overlooking the bigger picture, that is, what the objects in their drawings do and how they behave [3]. Secondly, despite the assumption that expression by drawing is easy and natural for children, they might have problems, for example, with understanding the idea of drawing things that would be seen on the screen [19]. Some problems did surface in the Talarius project as well. Despite recognizing the gradual development of the Talarius prototype, the children had a hard time picturing a link between this change and their individual ideas. The developers, however, considered the UI drawings to be of great value for their work. They made use of the drawings by analyzing their main attributes and by placing them all on display on the walls of their office in order to be able to continually draw upon them for guidance in design decisions. In problematic situations related to, for example, the choice of colors or layout, it was easy to turn to the drawings for clues and inspiration.

The developers discovered that the UI elements included in the drawings and their locations reflected the children's familiarity with common Windows software. This suggested that the developers could adhere to familiar conventions in order to enhance the learnability and usability of the application, yet without making it too much of a tool and too little of a game. The importance of quick learning is highlighted especially in children's software, as they are often prone to lose their interest unless they get in control of the software rapidly (e.g., [21]). Another navigation-related issue evident in the pictures was that many tasks were performed using buttons (instead of menus), which suggested that the children wanted all the most essential options to be visible on the screen at all times in order to find them quickly. Figure 1 entails "Quit," "Continue," "Question," and "Information" buttons in the top right hand corner, and buttons for choosing the game piece in the bottom right hand corner.

In the drawings depicting the playing mode, it was observed that it was very important for the children to see quickly what the status of the game and standings of the players were at a given time. The children had very different ways of representing the standings in their pictures but the common element was that this information was clearly visible and multiple formats of presenting the status were often used. In Figure 1, the status of the player is represented

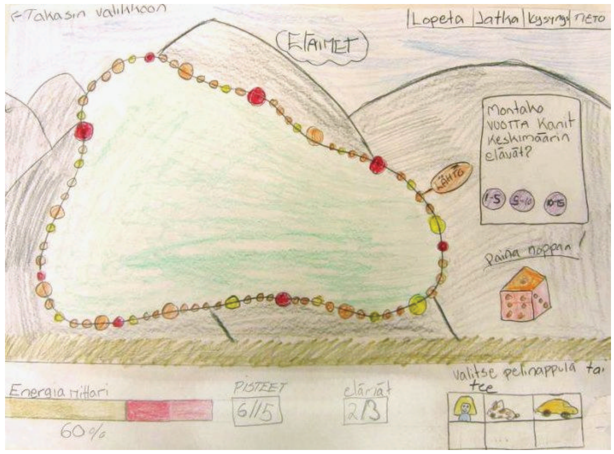


Figure 1: An example of the user interface drawings.

in terms of three different indicators: energy, points, and lives. In the game editor mode, the children's desire for as much freedom of choice as possible became apparent, which is in line with observations made by Druin et al. [21] about children wanting a large variety of functions and something to explore. According to the drawings, the children wanted the software to enable them, for example, to create backgrounds and characters, to choose freely the game path and the shape and color of the squares on it, to adjust the level of difficulty, and to add animations to the game.

3.2. Issues Discovered with Idea Maps

As brought up above, the children did not see the link between their individual drawings and the final outcome very well, which decreased their sense of ownership of the ideas. Guha et al. [2] have tackled this problem with a special technique for gradually merging ideas to make the process visible. In the Virtual Peatland project we took a somewhat related approach, attempting to make the different workshop sessions—as well as the activities within a session—build on each other more concretely. The workshops were, for a large part, based on gradual idea map creation.

The collaboration started with a session in which the children and the researchers created idea maps together to chart initial ideas about the content and structure of the site. In this paper, we leave the topic-specific content issues aside and only discuss children's opinions about more general questions, such as the presentation forms to be used in the learning environment. The children first made individual lists of their ideas and, based on them, suggested things to be included in a collective list of ideas. After this, the children suggested which issues on the list they considered the most important, and these were used as the basis for an idea map. Out of the individual lists and the collective list, the children highlighted games, pictures, quizzes, animations, and "learning by seeing and hearing" as the most important presentation forms, and these were taken into further discussion with the aid of the idea map. Similarly as the children in the Talarius project, they emphasized especially the importance of being able to create and explore—for example, to make a peatland, which became one of the two main tasks in the Peatland Adventure game—and having a variety of different activities to choose among. Later, in another session, the idea map was taken onto a more detailed and concrete level: the children worked in groups to select specific peatland species they considered the most interesting or important and to pick the most appealing presentation forms among those suggested previously. Another form of idea maps used in this project was a "game idea map" created by a group of children who were especially interested in game-related things. The main observations about their game idea were that their game had a boardgame-like structure and that they wanted various different kinds of subtasks, again a confirmation of the importance of variety and a large selection of things to do.

The suggestions obtained from these activities guided us to include a "from children to children" section in the learning environment, in addition to the Peatland Adventure game, to provide the users with varying activities. The children's section contains quizzes, crossword puzzles, and other types of tasks created by the participant children, video clips filmed during a peatland excursion with the children, and a Talarius-based board game about peatlands. In addition, through these content creation activities, we aimed to provide the children with a possibility to contribute concretely to the outcome and thereby promote their feeling of having an influence in the development of the website. The creation of contents did play an important role in terms of the children's feeling of ownership. Through these very concrete creations and their being straightforwardly present in the final outcome, the children saw more clearly that their participation had yielded something that furthered the development of the application. When comparing the children's experiences from the Talarius and Virtual Peatland projects, the latter succeeded with conveying the feeling of ownership better. When the children were asked whether they felt that their ideas showed in the final outcome, there were more negative than affirmative answers in the Talarius project, while in the Virtual Peatland project it was vice versa. Similarly when asked whether they considered that they had power in the development of the application, the difference between the two projects was noticeable. This was mainly due to the content creation activities which were in a considerably greater role in the Virtual Peatland project. The children felt empowered especially in terms of the quizzes they made. Moreover, in an interview conducted after the Virtual Peatland project, the content creation activities were also the best remembered aspects of the whole process.

3.3. Obtaining Ideas from Existing Learning Environments

Critiquing existing applications has proved a successful technique in design projects [19–21]. Having children point out pros and cons about different applications has helped the designers understand children's ways of interacting with technologies, uncover problems, and start developing solutions based on observations about children's preferences [19]. We asked the children to evaluate existing websites both in the Virtual Peatland project and in the Kids' Site project. Some of the websites were thematically related to the topics of the projects, whereas others were chosen merely because they included a variety of different activities and ways of presenting information. The children gave feedback on the websites, telling which features they did or did not like, and what kinds of ideas the websites gave them as regards the learning environment under development.

Consistently with observations made in previous research (e.g., [15, 21]) and confirming issues discovered with the other techniques in our projects, the children preferred sites which had a large variety of different activities or even a possibility to create something themselves, and it was very important to them what the site looked like. One interesting layout-related observation revealed through this technique in both projects was that having much empty space on a page was a negative thing in the children's opinion. Instead, the children's wish for a great variety of content existed also on single-page level. As the UI drawings also suggested, the desire to have plenty of visible elements concerned navigation as well: the children did not want to move back and forth much, they preferred having as direct access to all the parts of the site as possible.

In regards to the contents of the learning environment, the children wanted excitement and things that were related to real life. Their hope for real-life feel extended also to the visual appearance of the learning environment; the children liked illustrations that used real photographs. Additionally, as regards the illustrations, the children wanted to explore pictures with a great deal of detail, which is consistent with their preference for a variety and a great amount of content in general. The children also liked the idea of one main character that illustrates a learning environment—a game character or alternatively an animated "mascot" that appears on different pages throughout the website.

The children's opinions led us to introduce an animated cat character on Kids' Site to represent different sections of the site and to present the Rights of the Child. Moreover, the children created various different types of quizzes and puzzles for Virtual Peatland (as mentioned above) and Kids' Site to cater for the desire to have a great variety of activities. The real-life context is also represented in different ways in each of the sites: Virtual Peatland contains photographs taken and videos filmed by children, and Kids' Site has stories written by children about their everyday lives and how the Rights of the Child manifest in their lives. Children's possibility for creative participation is guaranteed especially on Kids' Site through several interactive activities involving their contribution to the website content.

4. Conclusions and Discussion

The significance of children's participation is divided into the issues related to carrying out collaborative design and research processes with children and those related to the outcomes of the design projects, namely, attributes of

Table 1: Issues discovered in the projects.

Category	Issue	Source
UI/Navigation	Using familiar UI conventions without risking game-likeness	UI drawings
	Controls clearly visible at all times	UI drawings
		Existing sites
	Game status and standings clearly visible at all times	UI drawings
UI/Appearance	Content-rich layouts, little empty space	Existing sites
	Realistic appearance, for example, photographs	Existing sites
Content/Theme	Themes related to real life	Existing sites
	Great variety in themes	Existing sites
		Idea maps
Content/Functionality	Much freedom of choice in functionality to allow exploration	UI drawings
		Idea maps
		Existing sites
	Possibility to create something	Idea maps
	A main character (a game character or a guiding "mascot")	Existing sites

children's learning environments. In this section, we will discuss these lessons separately.

4.1. Lessons Learned for Conducting Collaborative Projects with Children

One of the aims was to study the development processes and the activities they entailed from the perspective of the children and the developers alike. As we have seen above, involving children in the design process revealed and confirmed to the developers several issues that were useful to know when developing a game-based learning environment or a children's website. From the developers' point of view, the children's contribution was important in many aspects. According to the developers, children's ideas can guide the design of the appearance, functionality, and usability of the application. They especially stated that the children's participation helped them make the software more motivating for children. They also felt that the children's participation is valuable in enhancing the developers' understanding of children as technology users in general.

For the children themselves, participation in a real design project provided a chance to learn new things such as teamwork skills, shared decision-making, planning and designing, content-area and technology related knowledge, and activities which varied from their everyday schoolwork. Four areas of learning were identified in the analysis: (1) content area issues, (2) design skills ranging from general ones, such as more methodical ways of carrying out tasks, to more specific ones, such as different planning and design techniques, (3) social skills needed for collaboration in different compositions from pairs and small groups to speaking up before a larger group, and (4) learning skills, such as searching and evaluating information. In previous research, Druin [25] has also discovered learning outcomes in the same vein.

However, when working with school classes, supporting individual children's experience of contribution was a challenge. In the course of the study, we aimed to respond to this problem in two ways, firstly by improving the collaborative workshops through making them build on one another better, and secondly by incorporating concrete content-creation activities in the process. The content-creation perspective proved to be an important factor in enhancing the children's feeling of ownership: the project in which they were in a larger role was seen as more successful in terms of conveying this feeling, and the content creation activities were considered as the most enjoyable activities.

In summary, a good collaborative design process consists of varying activities in order to, firstly, provide the developers with rich data to draw upon in their implementation solutions and, secondly, maintain the children's interest in the participation process. Ideally, the activities build on each other, forming a logical and clear continuum which enables the children to follow the development of the application and the manifestation of their ideas. Moreover, the inclusion of concrete content creation activities enhances the children's feeling of ownership of the final outcome.

4.2. Lessons Learned for the Development of Children's Learning Environments

From the point of view of the development solutions, the issues discovered can be categorized into user interface and content related points. The UI issues comprise those having to do with navigation and those dealing with the appearance of the application, and the content-related issues concern the themes and the functionality of the applications. Table 1 summarizes these observations. We will discuss a few of these points in more detail.

As children want social interaction in technology [15, 21], being able to easily compare scores and standings is important to them in order to be able to monitor how they and others are doing. As regards the appearance, the children wanted each page or screen to be rich with content; not so abundant pages appeared to be boring to the children and to imply lack of content in the application altogether. Content-wise, our observations suggest similar issues as previous studies have brought up: the children want a great deal of content and functions to explore, and to create something new—or at least tweak and modify some aspects about the application [15, 21]. The children's preference for layouts using real photographs over those based on drawings might reflect their wish to avoid too "childish" appearances. The same might be true at the content level; themes dealing with real-life events and topics acknowledge children as part of the society, recognizing their interest in the world around them. On the other hand, the idea of a drawn (and perhaps animated) character as a guide on a website or as an avatar in a game is very appealing to them, most likely due to adding a more personal feel to the website or the game.

To summarize, children's desire for quick command of the application they are using manifests in several ways, such as by wanting to have UI elements (be they function buttons, navigation links, or content elements) constantly visible and readily available, and by calling for clear presentation of the status and standings in games. Content-wise, the most essential issue is versatility and richness of content; the children want creative activities, freedom of choice, and a large variety of things to explore and discover—both on the level of the whole application and, interestingly, also in terms of the structures of single pages or views.

Acknowledgments

Funding for the projects discussed in this paper has been provided by the National Technology Agency of Finland, the State Provincial Office of Western Finland, the European Social Fund, and the Ministry of Social Affairs and Health in Finland. We also wish to thank all the children, school teachers and after-school facility staff, and the developer teams for collaboration.

References

[1] A. Druin, "The role of children in the design of new technology," *Behaviour and Information Technology*, vol. 21, no. 2, pp. 1–25, 2002.

[2] M. L. Guha, A. Druin, G. Chipman, J. A. Fails, S. Simms, and A. Farber, "Working with young children as technology design partners," *Communications of the ACM*, vol. 48, no. 1, pp. 39–42, 2005.

[3] M. Scaife and Y. Rogers, "Kids as informants: telling us what we didn't know or confirming what we knew already?" in *The Design of Children's Technology*, A. Druin, Ed., pp. 27–50, Morgan Kaufmann, San Francisco, Calif, USA, 1999.

[4] D. Schuler and A. Namioka, "Preface," in *Participatory Design: Principles and Practices*, D. Schuler and A. Namioka, Eds., Lawrence Erlbaum, Hillsdale, NJ, USA, 1993.

[5] A. Bruckman and A. Bandlow, "HCI for kids," in *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, J. Jacko and A. Sears, Eds., pp. 428–440, Lawrence Erlbaum, Mahwah, NJ, USA, 2003.

[6] J. Read, "The ABC of CCI (Child Computer Interaction)," Interfaces 62, pp. 8–9, 2005.

[7] J. Good and J. Robertson, "CARSS: a framework for learner-centred design with children," *International Journal of Artificial Intelligence in Education*, vol. 16, no. 4, pp. 381–413, 2006.

[8] J. A. Rode, M. Stringer, E. F. Toye, A. R. Simpson, and A. F. Blackwell, "Curriculum-focused design," in *Proceeding of the 2nd International Conference on Interaction Design and Children (IDC '03)*, pp. 119–126, Preston, UK, July 2003.

[9] A. Clark and P. Moss, *Listening to Young Children: The Mosaic Approach*, National Children's Bureau, London, UK, 2001.

[10] R. A. Hart, *Children's Participation: The Theory and Practice of Involving Young Citizens in Community Development and Environmental Care*, Earthscan, London, UK, 1997.

[11] V. Nesset and A. Large, "Children in the information technology design process: a review of theories and their applications," *Library & Information Science Research*, vol. 26, no. 2, pp. 140–161, 2004.

[12] M. Scaife, Y. Rogers, F. Aldrich, and M. Davies, "Designing for or designing with? Informant design for interactive learning environments," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*, pp. 343–350, Atlanta, Ga, USA, March 1997.

[13] R. Oosterholt, M. Kusano, and G. de Vries, "Interaction design and human factors support in the development of personal communicator for children," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '96)*, pp. 450–457, Vancouver, Canada, April 1996.

[14] M. Bekker, J. Beusmans, D. Keyson, and P. Lloyd, "KidReporter: a user requirements gathering technique for designing with children," *Interacting with Computers*, vol. 15, no. 2, pp. 187–202, 2003.

[15] C. Dindler, E. Eriksson, O. S. Iversen, A. Lykke-Olesen, and M. Ludvigssen, "Mission from Mars—a method for exploring user requirements from children in narrative space," in *Proceedings of the 4th International Conference on Interaction Design and Children (IDC '05)*, pp. 40–47, Boulder, Colo, USA, June 2005.

[16] J. Verhaegh, I. Soute, A. Kessels, and P. Markopoulos, "On the design of Camelot, an outdoor game for children," in *Proceeding of the 5th International Conference on Interaction Design and Children (IDC '06)*, pp. 9–16, Tampere, Finland, June 2006.

[17] D. Bilal, "Draw and tell: children as designers of web interfaces," in *Proceedings of the 66th Annual Meeting of the American Society for Information Science and Technology (ASIST '03)*, pp. 135–141, Long Beach, Calif, USA, October 2003.

[18] L. Hall, S. Woods, K. Dautenhahn, and P. Sobreperez, "Using storyboards to guide virtual world design," in *Proceeding of the 3rd International Conference on Interaction Design and Children (IDC '04)*, pp. 125–126, College Park, Md, USA, June 2004.

[19] C. Jones, L. McIver, L. Gibson, and P. Gregor, "Experiences obtained from designing with children," in *Proceeding of the 2nd International Conference on Interaction Design and Children (IDC '03)*, pp. 69–74, Preston, UK, July 2003.

[20] A. Druin, "What children can teach us: developing digital libraries for children with children," *Library Quarterly*, vol. 75, no. 1, pp. 20–41, 2005.

[21] A. Druin, B. Bederson, A. Boltman, A. Miura, D. Knotts-Callahan, and M. Platt, "Children as our technology design partners," in *The Design of Children's Technology*, A. Druin, Ed., pp. 51–72, Morgan Kaufmann, San Francisco, Calif, USA, 1999.

[22] R. C. Richey, J. D. Klein, and W. A. Nelson, "Developmental research: studies of instructional design and development," in *Handbook of Research for Educational Communications and Technology*, D. Jonassen, Ed., pp. 1099–1130, Lawrence Erlbaum, Mahwah, NJ, USA, 2nd edition, 2004.

[23] J. van den Akker, "Principles and methods of development research," in *Design Approaches and Tools in Education and Training*, J. van den Akker, R. M. Branch, K. Gustafsson, N. Nieveen, and T. Plomp, Eds., pp. 1–14, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.

[24] R. K. Yin, *Case Study Research: Design and Methods*, Sage, Thousand Oaks, Calif, USA, 2nd edition, 1994.

[25] A. Druin, "Cooperative inquiry: developing new technologies for children with children," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*, pp. 592–599, Pittsburgh, Pa, USA, May 1999.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 719291, 11 pages
doi:10.1155/2008/719291

Research Article

The Extended Likeability Framework: A Theoretical Framework for and a Practical Case of Designing Likeable Media Applications for Preschoolers

Vero vanden Abeele[1] and Bieke Zaman[2]

[1] *e-Media Lab, Group T-Leuven Engineering College (Association K.U.Leuven), Vesaliusstraat 13, 3000 Leuven, Belgium*
[2] *IBBT, Centre for User Experience Research (CUO), Katholieke Universiteit Leuven Parkstraat 45 bus 3605, 3000 Leuven, Belgium*

Correspondence should be addressed to Vero vanden Abeele, vero.vanden.abeele@groept.be

Received 30 September 2007; Revised 30 April 2008; Accepted 18 August 2008

Recommended by Adrian Cheok

A theoretical framework and practical case for designing likeable interactive media applications for preschoolers in the home environment are introduced. First, we elaborate on the theoretical framework. We introduce the uses and gratifications paradigm (U&G). We argue that U&G is a good approach to researching likeability of media applications. Next, we complete the U&G framework with expectancy-value (EV) theory. EV theory helps us move from theoretical insights to concrete design guidelines. Together, the U&G framework and the EV model form the foundation of our extended likeability framework for the design and evaluation of interactive media applications, for preschoolers in the home environment. Finally, we demonstrate a practical case of our extended likeability framework via the research project CuTI. The CuTI project aims at revealing those particular user gratifications and design attributes that are important to support playful behaviour and fun activities of preschoolers in the home environment.

1. Introduction

Digital media are increasingly moving into everyday life. As a consequence, designers no longer exclusively focus on applications with productivity goals in a work-related context. Instead, the focus has also shifted towards design for entertainment, aiming at delivering a more social or emotional experience. Human-computer interaction (HCI) research has responded to this new trend and moved from the "second wave" to the "third wave" of HCI methods and theories. The first wave adapted methods and theories from cognitive psychology whereas the second wave was influenced by sociological principles and "situated" research methods. More particularly, second wave research still focused on technologies in a work-related context with explicit rational (i.e., instrumental) purposes. In contrast, third wave research is centred on emotional and cultural technologies in a more entertaining and everyday context [1]. Often, the third wave is equated with "user experience" research or the "new usability" of affective factors such as "fun"' and "trust" [2]. The recent shift within HCI from second to third wave research made the need for new or adapted research methods and theories more explicit, resulting in a move to ground human-centred research on user experience. Stelmaszewska et al. [3] also cite: "HCI lacks theories and methods to facilitate approaches to design products which allow for pleasurable, enjoyable, and entertaining Interaction."

In [4], we came to the same conclusion, namely, that HCI and child-computer interaction (CCI) lack a solid framework for likeability research with preschoolers. The need for theories to gain insight into the meanings that children give to new media has been acknowledged within HCI and CCI. For instance, Learning and Teaching Scotland [5, page 8] points out: "There is no detailed analysis of how children make meanings with interactive media and a conspicuous lack of evidence about what actually happens in social, cognitive, and affective domains." Hassenzahl [6]

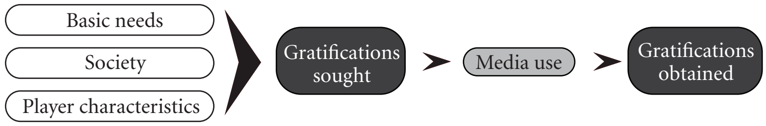


Figure 1: A simplified model of the uses and gratifications paradigm, by Lucas and Sherry [12] based on Rosengren's outline [10].

stresses that there is still a lack of successful methods to reveal experience-oriented product requirements: "...there are no specialized analysis techniques used in the industry. Therefore, it is important to develop appropriate analytical techniques to help product designers or usability engineers to gather hedonic requirements in a certain context of use."

In Section 2 of this article, we detail our theoretical foundations. We first elaborate on the uses and gratifications paradigm, well known in mass communications science, and promising for researching fun and likeability of media. Next, we introduce expectancy value theory, originating from social psychology. The EV theory helps us move from user gratifications to product attributes. The combination of the U&G framework and the EV model forms the blueprint for our extended likeability framework for the evaluation and design of interactive media applications.

In Section 3 of the article, we apply our extended likeability framework to fun media applications, for preschoolers in the home environment. Via the CuTI research project we demonstrate how high-level user gratifications can be translated into product attributes.

2. The Extended Likeability Framework

2.1. Uses and Gratifications, a Communication Science Perspective

2.1.1. The Uses and Gratifications Paradigm Outlined

The lack of available frameworks, theories, and models in HCI on likeability research for children explains why we turned our attention to communication sciences in order to find an appropriate framework that is "human-" centered. The uses and gratifications (U&G) paradigm, well known in mass communication science, is such a human-centered framework. The uses and gratifications paradigm accepts that users actively seek media, according to their needs and expectations, and give meaning to these media. The basic premise of the U&G paradigm is that people actively choose media to fulfil specific gratifications. U&G research is interested in what uses and gratifications the audience derives from media and what appeal media have for the audience. More particularly, the uses and gratifications (U&G) paradigm focuses on *how frequently* and especially *why* certain *types of (genres* within) media are used and *by whom* [7–9]. Rosengren [10, 11] presented a U&G model in which basic needs, individual differences, and contextual societal factors interact and result in a variety of gratifications that are sought from the media. People then use media in the hope that the gratifications they seek can actually be obtained.

Figure 1 shows a simplified model of the U&G paradigm, presented by Lucas and Sherry [12] and based on Rosengren [10]. The U&G model does not specify how exactly the mix of societal factors, individual characteristics, and basic needs leads to certain types of media behaviour; rather, it calls on researchers to fill in these details. However, it does acknowledge that no single factor drives media use; it is only the mix and the interaction among needs, contextual factors, and individual characteristics that can predict media use.

For our research, we find this all-encompassing view interesting because it is similar to the multiparadigmatic view in human-computer interaction. HCI researchers also stress the interplay between social/contextual, individual/psychological, and tool/media characteristics. Nevertheless, models or theories that are employed to research enjoyment or fun are often narrowed down to physio/psychological parameters [13, 14]. In contrast, the U&G framework allows for a global approach that does not only incorporate basic biological influences and psychological dispositions but also addresses the social context. The influence and importance of context for product experiences can be found with different HCI researchers and theorists. Goffman [15], for example, refers to "situated experience"; Overbeeke et al. [16], Hummels et al. [17], and Dourish [18] emphasize the importance of "context for experience"; Hassenzahl [6] mentions "context of use."

2.1.2. Three Assumptions for Adopting the U&G Framework

In order to be able to use the U&G framework as a foundation for our likeability framework, we need to accept the following assumptions [4].

(1) Assumption 1: Interactive Artefacts Can Be Seen as Media

The choice for the U&G framework implies that we abandon paradigms that are current in work on human computer interaction, such as the "tool perspective" or the "usability perspective." These perspectives are attuned to an adaptation of information and communication technologies to users' cognitive competences. Such adaptations result in a more efficient and effective performance of predefined tasks. Technologies are also often conceptualized as "tools" within the HCI community. These "tools" then function as an extension of the human body and/or mind, compensating for human shortcomings [20]. However, we no longer view artefacts as "tools," neither do we follow the "usability perspective"; instead we look upon applications as "media" [20]. The adaptation of the U&G framework in likeability research demands acceptance of the view that artefacts like interactive toys and games can be considered as media. The "media perspective" in HCI resembles the way communication sciences define the concept "media." Media are then defined as bridges that make communication and information possible between people [20]. This leads us to the first assumption: the U&G framework requires us to *accept that interactive toys, games, and other artefacts that support play activities can be thought of as media.*

(2) Assumption 2: Children are Active in Their Choice of Media

U&G starts from the premise that the audience is active; an important part of mass media use is assumed to be goal directed [7]. Instead of being passive bystanders, people choose to engage in a medium and their choice reflects their need for gratifications. Thus, need gratification and media choice lie within the audience member. The perspective of an active audience is illustrated by Schramm et al. [21, page 169] who state that "in order to understand television's impact and effect on children, we have first to get away from the unrealistic concept of what television does to children and substitute the concept of what children do with television."

Clearly, this idea, that one actively chooses to deal with a certain medium based on particular needs and the need for gratifications, is very similar to the user-centred approaches that prevail in HCI. For instance, we can refer to the goal-directed design approach of Cooper and Reimann [22], where design requirements are based on needs in order to fulfil goals, wishes, and dreams. Similarly to communication scholars that follow a U&G perspective, we *posit that children are active in their choice to use or not to use media/applications for certain reasons.*

(3) Assumption 3: Likeable Products Fulfil Gratifications

Finally, similarly to HCI researchers that approach users as purposeful and task oriented, U&G researchers make the assumption that people use media in order to serve their needs and fulfil their gratifications. This is an interesting perspective, especially in the context of designing for fun or for playful interaction, where the goal is not to produce output or accomplish productivity tasks but rather enjoying the process. With regard to the U&G paradigm, a distinction is made between *gratifications* that are *sought* (GS) and those that are *obtained* (GO) from media use. To meet the user needs, gratifications obtained should correspond with gratifications sought. In this respect, we formulate our final assumption: *in order to design likeable products for children, we need to fulfil those gratifications that children seek in specific media.*

2.1.3. Uses and Gratifications of Games and Play

Typically, research in the U&G tradition begins with the identification of the gratifications that are sought or the motivations for media use in relation to the traits of the audience. These studies tend to show no universal set of motivations or gratifications sought from media. Motivations vary across media, genres and user groups, and cultures [11]. However, several U&G studies (cf. [23, 24]) on traditional media with adults reveal the existence of four gratifications clusters that repeatedly occur:

(1) information/surveillance,
(2) personal identity,
(3) integration/social interaction/personal relationships,
(4) entertainment/division/escapism.

Recently, interesting studies have been conducted on video games and gratifications for young players. Sherry et al. [25], for example, conducted a survey (from elementary scholars to university students) in order to determine the uses and gratifications of playing video games. Sherry and Lucas [26] defined six gratifications that could explain why young people play video games:

(1) competition: to be the best player of the game;
(2) challenge: to push yourself to beat the game or to get to the next highest level;
(3) social interaction: to play as a social experience with friends;
(4) diversion: to pass time or alleviate boredom;
(5) fantasy: to do things that one cannot do in real life;
(6) arousal: to play for excitement.

Many more U&G studies exist and are carried out; in fact U&G is also seen as an excellent way to study new media [27]. New media credit users with even more control over their own activities. The active audience paradigm prevalent in U&G aligns well with new interactive media such as the internet or video games. In this respect, mass media scholars promote an adaptation of the U&G framework to new technologies: "A challenge is for researchers to adapt and mould the current conceptual framework to deal with new communications technologies" [28, page 241].

2.2. Expectancy Value Theory: From High Level Gratifications to Low Level Attributes

Typically, U&G studies result in a set of motives or gratifications that are still quite general or abstract. Let us recall the game gratifications that Sherry and Lucas defined: competition, challenge, social interaction, diversions, fantasy, and arousal. A designer might point out that these gratifications are interesting but do not provide useful operational guidelines for designing actual media products. Consequently, the U&G framework can be perceived as too abstract to offer concrete guidelines for designing interactive applications.

Therefore, to understand the relationship between the design of a media application and the gratifications of that media user, we rely on an extension/addition to the U&G framework. To make this transition, we rely on expectancy value theory, first introduced by Fishbein and Ajzen [29, 30] in the nineteen seventies and adopted and refined by many scholars and disciplines in later years.

2.2.1. Expectancy Value Theory

Expectancy value (EV) theory is based on the premise that all human behaviour (including media use) is shaped by behavioural intentions or attitudes. These attitudes are shaped by the expectations (beliefs) one holds about an object. In turn, these expectations are based upon specific attributes of the object. An expectation or belief then is the subjective (for the user) probability that an object has a given attribute. Furthermore, each attribute has either a positive or negative value associated with this expectation,

$$A_o = \sum_{i=1}^{n} b_i e_i. \tag{1}$$

As seen in the formal expression by Fishbein and Ajzen [29], the attitude towards an object (A_o) is a function of a belief about the object attributes (b_i) and the evaluation of an attribute (e_i). Expectancy value theory acknowledges that users learn to choose a product (or medium) because they expect (believe) that the product contains attributes that are instrumental to achieving desired consequences.

2.2.2. Expectancy Value Theory and Uses and Gratifications

EV theory was first linked to U&G by Palmgreen. "A fundamental assumption of U&G models, that of an active audience, is in fact founded upon the even more basic precept that audience members do have perceptions of the gratifications available from various alternatives, and that they act upon these perceptions." [8].

User gratifications are shaped by (1) the expectancy or the perceived probability that using a media application will have a certain consequence and by (2) the evaluation or the positive or negative affect towards the consequence.

U&G researchers Palmgreen and Rayburn [8] explain this relationship. They see gratifications as a function of the user evaluation of a media object. More particularly, Palmgreen and Rayburn base themselves on Fishbein and Ajzen's formal expression [29] and state that the gratifications sought (GS_o) in an object "o" are a function of (the weighted sum of) the beliefs (b_i) and evaluations (e_i) about the attributes "i" that object "o" possesses. More specifically, the gratifications that users seek (or avoid) from a media object "o" depend on the belief or the expectancy (=perceived probability) that "o" has a particular attribute "i" and the positive or negative evaluation of that attribute.

The expression results in the following formula:

$$GS_o = \sum_{i=1}^{n} b_i e_i. \tag{2}$$

The expectancies or beliefs about the product attributes can be learned and modified through direct experience, but might also be the result of communication and/or processes of induction and deduction [8].

One child might, for example ask his parents to buy a computer game about Peter Pan (i.e., object, "o") because he believes that he will like it (gratification sought or "GS_o"). Although the child has never played this computer game before, he nevertheless expects the game to be fun because of the central character (i.e., belief associated with a certain game attribute, "b_i"). Through previous experiences of watching the Peter Pan movie, the child learned that the adventures of Peter Pan are always fun and exciting (i.e., positive evaluation of game attribute, "e_i"). However, if the child does not expect positive attributes or if he is convinced that the object only induces negative attributes, then he would not ask to buy it. The child may also, for instance, choose to quit playing the computer game because it is too difficult and frustrating (beliefs with negative evaluations) which makes him loose control of the game ("GS").

2.2.3. Means-End Theory

In contrast to our simple Peter Pan example, in real life the gratifications sought are mostly determined by more than one belief about a positive or negative attribute. Each gratification is indeed determined by a set of beliefs and evaluations of attributes.

Unfortunately, Palmgreen's elaboration of U&G with EV theory did not generate many research projects within the U&G tradition that focus on how media attributes influence the audience behavioural intentions. Nor did Palmgreen elaborate on methods to extract these beliefs from the audience. However, EV theory has not only been linked to social psychology and communication sciences. In consumer research, EV theory parallels research based upon means-end theory [6, 31, 32].

The common premise of means-end theory and expectancy value theory is that users learn to choose a product (or medium) because it contains attributes that are instrumental to achieving desired consequences and fulfilling values. The common generic means-end chain consists of attributes (A), consequences or benefits (C), and values (V) [32]. The consequences of using a product in this case correspond to the "gratifications sought" within U&G research. In consumer research, means-end chains are used to explain what motivated consumers to desire certain products or product attributes. In human computer interaction, Subramony [33] and Zaman [34] successfully used means-end chains and consequently "laddering" as a qualitative research method to create design recommendations.

Means-end analysis is preferably done via laddering [31]. Laddering refers to an in-depth, one-on-one interviewing technique, used to develop an understanding of how users

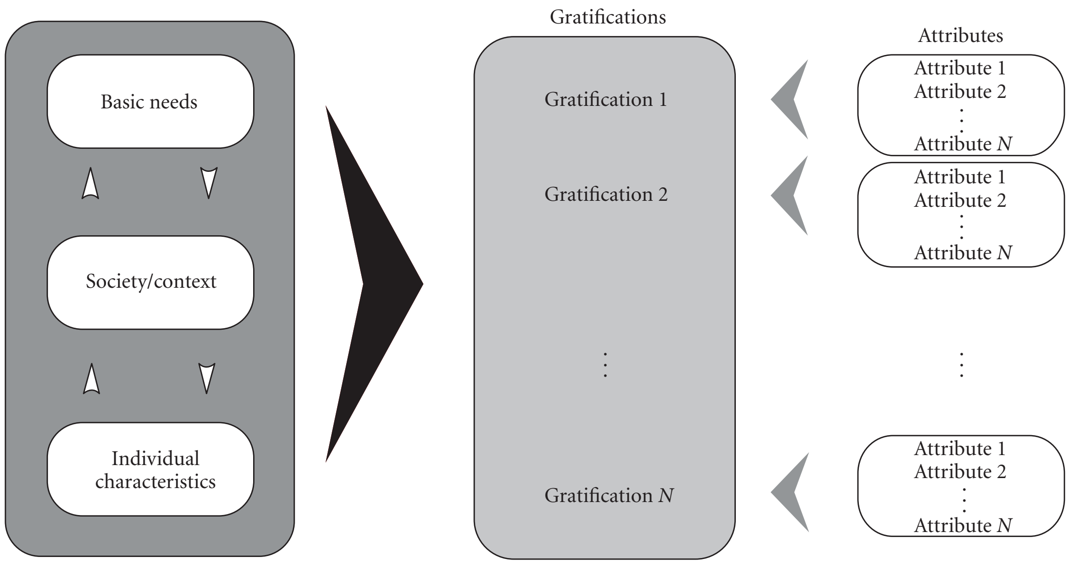


Figure 2: The extended likeability framework.

translate the attributes of a product into meaningful associations with the "self" [32]. Laddering involves a tailored interviewing format using primarily a series of directed probes, typified by the following question: "why is that important to you?" The goal is to identify linkages between key perceptual elements across the range of attributes, consequences, and values. These association networks or ladders, ranging from the concrete to the abstract, provide an understanding of the relations between a product's attributes and the user motivational perspective to acquire this product. Laddering allows discovering the underlying reasons why an attribute is important, in relation to the underlying gratification.

2.3. Combining Uses and Gratifications with Expectancy Value Theory

To conclude the first section of this article, we posit that we should combine U&G and EV theory in order to research and design likeable products for a given audience in a given context. In Figure 2 we illustrate how basic needs, society and context, and individual characteristics lead to gratifications. We should also understand how these gratifications are based upon expectancies and evaluations of attributes. In order to design likeable toys and interactive applications for preschoolers in the home environment, we need to understand how basic needs, society and context, and individual characteristics lead to preschoolers' gratifications that need to be fulfilled. Next, we need to understand which beliefs preschoolers hold with regard to the medium attributes and how these attributes are related to the gratifications of preschoolers. In the following section, we will discuss how we applied this underlying theoretical structure to a practical case for preschoolers.

3. CuTI: Applying the Extended Likeability Framework to Preschoolers in the Home Environment

Although many U&G studies are conducted on the gratifications for traditional media, and recently for digital media and games, no recent U&G or EV studies, to our knowledge, have been done on preschoolers and interactive media in the home environment. Therefore, we decided to take up a new research project, which we called CuTI [35]. CuTI is the acronym for "Cuddle Toy Interface," an on-going project aimed at revealing those particular user gratifications and design attributes that support playful behaviour and fun activities of preschoolers within the context of the home environment.

The gratifications and attributes we find can inform the design of interactive toys and games in later stages. We emphasize that CutTI is an on-going project. The case here is offered to provide insight into how to apply the theoretical framework. We wish to emphasize that the specific gratifications and attributes uncovered in this initial study are preliminary insights, needing further validation with confirmatory research modes.

3.1. CuTI Research Method

Researching five years old children in the home environment calls for appropriate research methods. A survey of what preschoolers prefer, and for what reasons, would not provide the in-depth information we were looking for, nor would this approach be effective with our young audience. Consequently, in [4, 34, 36, 37] we explain our child-centered methodology for researching the likeability framework.

Briefly stated, we first combined the existing literature on child-computer interaction and uses and gratifications research on games. Armed with this information and with a preliminary and theoretical classification of gratifications [4], we conducted qualitative user research in order to refine our findings. More specifically, we combined user diaries and cultural probe packages followed by participant observations and depth interviews [36, 37].

As research participants, we selected 8 households with five years old children. During a time span of one week, we researched these preschoolers and their families at home. At the beginning of the week, the researcher explained the purpose of the project and handed over the package consisting of user diaries and cultural probes to the preschoolers and their parents. A first observation of the child and his/her social and physical environment was made. At the end of the week, the researcher collected the diary and probes. These user diaries and probes allowed refining and validating the gratifications [4]. Next, the researcher conducted an interview with and observation of the child, guided by the diary and the results of the probing activities. The goal was to better understand "why" these activities were considered fun, and to attempt to link attributes to the gratifications. Hence, for each of the activities listed in the diary we asked the preschooler to explain "why it was fun." Because of the cognitive limitations of preschoolers, we converted the "laddering" method as listed in [33, 34] to "reflection in action." Basically, we asked children to elaborate on why something was fun, while demonstrating the activity or playing the fun activity.

At the end of the interview session, the researcher completed the data collection by interviewing the parent(s), asking additional information on the fun activities and also on the parental attitude and influence towards these fun activities. For more detailed information on the specific adapted research methodology for preschoolers in the home environment, we refer to [4, 34, 36, 37].

3.2. Applying the Extended Likeability Framework

To illustrate the practical use of the extended likeability framework (ELF), we now "fill it in" with the findings of the CuTI project. The aim is not only to discuss the exploratory gratifications and attributes found but also to demonstrate how the ELF can easily be filled in and used as a concrete design guide.

Figure 3 presents our extended likeability framework. It illustrates how an interaction between basic needs, contextual societal factors, and individual characteristics influences the gratifications children get or seek. In total, we discern five different gratification areas that make things fun and likeable. We hereby assume that the more a product fulfils these gratifications and possesses the related positively evaluated attributes, the more the product will be likeable for preschoolers.

In the following paragraphs, we will discuss each part of the extended likeability framework. First, preschoolers' basic needs, social and physical environment, and individual characteristics are dealt with (the left column of Figure 3). Then, we go into more detail on preschoolers' gratifications (the middle column). For every gratification, we will discuss how these gratifications are connected with the products' attributes (the right column).

3.3. Basic Needs, Contextual Influences and Individual Characteristics of Likeable Applications for Preschoolers in the Home Environment

3.3.1. Basic Needs

With "basic needs," universal biological and psychological motives are meant [9]. Often in this context, the needs pyramid of Maslow [28, 38] is referred to. In his needs pyramid, Maslow [38] mentions organic and physical needs, the need for security and safety, the need for social experiences, the need for recognition and appreciations as well as the need for self-actualization [28]. Besides basic needs mentioned by Maslow, there is a general consensus that play from a functional perspective is necessary for learning and development. Based upon the work of several developmental psychologists and ludologists [39–43] we add the need to play as a universal, basic need for children.

3.3.2. Society and Context

With "society," we refer to the prevailing social, political, cultural, economic, and media structures. Since young children are our main target group, insight in their social and cultural context is crucial. Adults inevitably forget what it is to be a child. The contextual landscape has changed, so that adults no longer understand what it is to grow up as a child today. Moreover, each child differs from eachother and may have grown up in a different social and physical environment. For our research in the context of the CuTI programme we limited ourselves to the nuclear family, consisting of two adults of both sexes, and one or more own or adopted children.

3.3.3. Individual Characteristics

Individual characteristics typically refer to the psychological nature and demographic characteristics of the audience. Rosengren [10] makes a distinction between intra- and extraindividual characteristics. The intraindividual characteristics encompass, for example, the personality traits and developmental stages, whereas someone's social position is an example of an extra individual characteristic.

3.4. Gratifications and Attributes of Likeable Applications for Preschoolers in the Home Environment

We now focus on the high level gratifications and low level attributes of likeable products. To arrive at the gratifications, we first combined the existing literature on child-computer interaction and uses and gratifications research to derive a

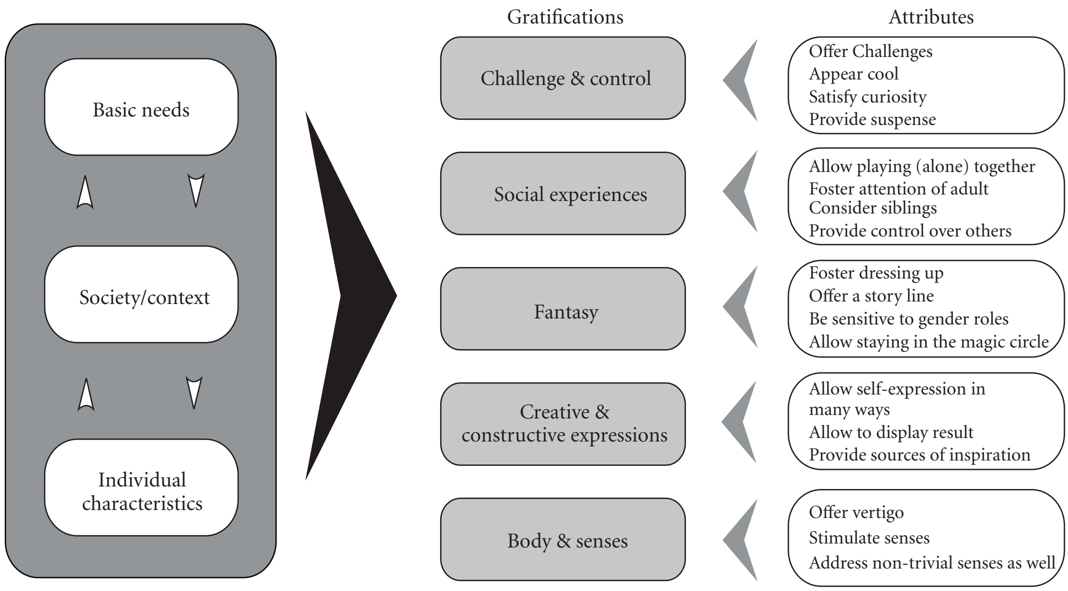


Figure 3: The extended likeability framework as applied to preschoolers in the home environment.

preliminary list of hypothetical uses and gratifications. Next, we conducted qualitative user research that allowed us to refine our list of gratifications and to come up with our definitive list of hypothetical gratifications. To arrive at the attributes we defined gratifications as a function of beliefs and evaluations of particular attributes.

We emphasize that we do not contend that our list of attributes is exhaustive (more research is needed to be able to claim reasonable exhaustiveness), nor do we pretend that one media application should try to incorporate all these attributes. Yet we do think that a list of attributes, such as the one derived from initial research such as ours, can inform the research community and help to understand the diverse gratifications and the ways to translate these into design solutions.

3.4.1. Gratification "Challenge and Control"

With "challenge and control," we adopt Csikszentmihalyi's theory of flow [44]. Flow theory deals with the fact that, in order to be truly challenging and "absorbing," an activity or task needs to address the right skill level. If the task is too easy it becomes boring; if it is too hard it becomes frustrating. If the challenge is tuned right, however, activities become so rewarding that they are done just for the sake of "doing it." Challenge is thus intertwined with development, with learning new information and mastering new skills. This gratification is also well researched and addressed in the research on child-computer interaction [45–48]. In our framework, "surprise" is conceptualized in the category of challenge. Surprise, control, and challenge are indeed important catalysts for likeability [49, 50].

3.4.2. Attributes of "Challenge and Control"

Addressing the right skill level, not too hard, not too easy, was very important for our preschoolers. Learning is part of this category and often mentioned as fun, for example, learning about dinosaurs. In our CuTI project, children often stressed that something was "very hard" in a proud way. To them, it was important that things were "not for babies." Being "cool" often involved a challenge or "danger." Children with younger siblings stressed that they could do things younger siblings could not. The other way around, children with older siblings really wanted to be able to do the same thing as their older brother or sister. Part of the challenge was also that things are unpredictable and satisfy curiosity. Fun activities that surprised or provided suspense (slimy worms, jack-in-the box, etc.) were popular. However, being scared could quickly turn aversive and become really scary. Especially because of the (sometimes) blurred boundary between the real and the magic world, we observed that some activities were on the edge of offering too much suspense at the expense of loosing "control."

We conclude that likeable media objects that focus on fulfilling the "challenge and control gratification" should consider the following attributes: (a) offering the right challenge, not too easy, not too hard; (b) appearing "cool" (as for older children or a little dangerous) and not for "babies"; (c) satisfying curiosity, offer something new; (d) Provide suspense while not being too scary.

3.4.3. Gratification "Social Experiences"

With this category, we refer to fun activities that occur in the presence of others. Rice offers the classification of play according to the degree of social involvement: unoccupied play, solitary play, onlooker, play, parallel play, associative play and finally cooperative play. All these different types of play indeed came up during our research.

Based on Schutz' interpretation [51], we distinguish between activities that point towards "inclusion" (being a member of a social group), "affection" (expression of being accepted), and "control of others" (being able to have an influence on others). Stelmaszewska et al. [3] and Jansz and Tanis [52] also refer to the positive relation between social experiences and likeability. Offering social experiences is important for the likeability of a media application.

3.4.4. Attributes of "Social Experiences"

With "social experiences," we denote activities that refer to being together with/or in the presence of others. Although preschoolers have an egocentric worldview, this certainly does not mean a lack of interest in social experiences. In our project, the activities that were mentioned as most fun by our subjects often implied the cooperation of a parent, sibling, or friend. However, playing together did not always literally mean playing "together" (associative or cooperative play). Often, playing together pointed to individual play, but still within the presence of a family member or friend (onlooker play, parallel play). Clearly, "playing alone together" was still preferred over "playing alone" (solitary play). We noticed that siblings were eagerly present to "disturb" or "facilitate" play, and again affected the fun activity profoundly. Furthermore, we noticed that the (undivided) attention of a parent or other adult was of particular value. Such activity often ended up as the most fun of the day. Also activities that involved no direct social experience but ensured inclusion in a social group indirectly were popular, for example, watching a television show in order to be part of the group. Finally, control over others was an important issue; being "the boss" or having a "dominant" role in playing was important, for example, being able to decide "who's next."

We conclude that likeable media objects that focus on fulfilling the "social experience gratification" should consider the following attributes: (a) allowing playing together or if not possible allowing playing "alone together"; (b) stimulating playing/attention of adults; (c) considering the siblings in the environment; (d) providing a sense of control over others.

3.4.5. Gratification "Fantasy"

With "fantasy," we include all activities that address role-playing, pretense play (also referred to as dramatic play or symbolic play), mimicry, and make-believe [39, 41]. Make-believe concerns the imagination of objects and events without any immediate link to the real world. According to developmental psychologists such as Berk [39, pages 217, 224], fantasy is a very important activity for preschoolers. With preschoolers, make-believe becomes more and more complex. From the age of three year, children can easily imagine objects or situations without an immediate link with real life. Through fantasy experiences, children can go beyond the boundaries of normal experience. This can make them feel more than what they really are. We found that fantasy and role-playing are extremely important and often precede the other components of fun activities, or coordinate the diverse other components of fun activities.

3.4.6. Attributes of "Fantasy"

With "fantasy," all activities that address role-playing, pretense play, mimicry, and make-believe are concerned. As for our project, we found that dressing up and objects that support fantasy roles (e.g., fairies or pirates) as well as everyday roles (e.g., construction worker, caregiver) were very popular. Offering a story that stimulated children to enter into a world of fantasy seemed to guarantee "fun." We noticed a strong gender disparity, where it was important that children could stay within consistent gender roles. Girls want to be princesses, boys want to be knights. Gender neutrality is accepted, but inverse gender roles inhibit play (e.g., a girl refused to play with a "male pirate" pawn).

Together with "fantasy," we encountered the magical thinking that is a characteristic of children of this age. While we observed role-playing, the boundaries between the real and the fictional world often disappeared,for example, snapping out of a role was often found difficult. Children placed a great weight on their imaginary lives and roles and liked to keep on their dresses and makeup when switching to another activity, they liked to stay in their "magic circle."

We conclude that likeable media objects that focus on fulfilling the "Fantasy gratification" should consider the following attributes: (a) fostering dressing up; (b) offering a story; (c) offering gender-sensitive or neutral roles; (d) allowing for staying in their magic circle.

3.4.7. Gratification "Creative and Constructive Expressions"

Important to preschoolers is the need for self-expression and self-disclosure, materialized or expressed in a noticeable way. This gratification concerns activities such as drawing, painting, modelling, constructing, storytelling, singing, and so on. Creating or constructing something is especially common between the ages of three and six years [39, page 251]. As preschoolers get older, their creative skills develop. Not only the process matters but also the end product. Five years old children are aware that they can create something and are proud of it. They have a need for appreciation and recognition of this. Artefacts of preschoolers become an alternative way to express feelings and thoughts.

3.4.8. Attributes of "Creative and Constructive Expressions"

In the CuTI project, we noticed that children could really be absorbed into expressive activities such as drawing, painting, handicrafts, but also constructing, carpentry, storytelling, and so on. Our preschoolers even liked simply sorting objects according to colour or shape. We found that activities that allow expression should contain attributes that both inspire and guide the activity. On the one hand, one should provide an example of what can be made (such as a mould or a picture), but on the other hand the child should be given enough flexibility to still "own" the creation. If the activity merely consists of putting an object in the right place, then it does not allow the child to demonstrate his/her creativity. Whether it is drawing, building, modelling, and so forth, it is very important that there is a clear result that can somehow be demonstrated. The prospect of having created something that can be shown or that can "last" is part of the pleasure of creation.

We conclude that likeable media objects that focus on fulfilling the "creative and constructive expressions gratification" should consider the following attributes: (a) allowing self-expression in a myriad of ways; (b) allowing a result that lasts and can be displayed; (c) providing sources of inspiration.

3.4.9. Gratification "Body and Senses"

Finally, "senses" refers to all behaviours in which the senses are involved such as eating, tickles, watching colour patterns, making strange noises, and so on [41]. Besides, these sensory stimulations, children are also keen on physical sensations ("body") such as jumping and running very fast. Roger Callois addressed this category of play based on bodily action as "vertigo" [53]. According to Berk [39], even simple repetitive motor movement can be considered the earliest form of play. We found that sensory stimulation in all its forms (not just motor movements) is important and considered fun for our preschoolers; it is not essential to distinguish between input or sensory perception (senses), and output via the motor system (body), as these blend into each other.

3.4.10. Attributes of "Body and Senses"

When focusing on current games, certainly on video games, the gratification "body and senses" is often not addressed, or narrowed down to impressive graphics, sound, and "button bashing." However, we found that "all" senses are important. Our preschoolers loved things that move around, make noises, have bright lights, and have a soft touch or nice smell. Further, games that involved tickling, massaging, cuddling, or touching were very popular. To our surprise, many fun activities also included eating; something we did not expect to find, but that appeared in the observations of most children. Finally, not surprisingly, children loved activities that involved running around, jumping, tumbling, which we summarize as "vertigo."

We conclude that likeable media objects that focus on fulfilling the "creative and constructive expressions gratification" should consider the following attributes: (a) addressing vertigo; (b) addressing senses, also nontrivial senses such as smelling or tasting (eating).

3.4.11. Combining Attributes

It is important to realize that most activities involved several gratifications and many attributes, for example, "taking a bath with mom" both involved "body and senses" and "social experiences." "sorting out pearls" was about "creative and constructive expression" but preschoolers also liked the activity because the colours or shapes were in attractive colours, which points to "body and senses" and is related to the "fantasy" gratification. "Walking in between sheep" involved "body and senses" but also "challenge and control." In fact, none of the activities that were listed could be narrowed down to one gratification, but oftentimes involved two or three of them.

3.5. Summarizing Gratifications and Attributes

To summarize, we believe that when designing likeable applications for preschoolers, we need to fulfil the gratifications mentioned in our likeability framework:challenge and control, social experiences, fantasy, creative and constructive expressions and body and senses. As for the CuTI case, the framework implies that an object will be likeable if preschoolers can obtain these gratifications from it. More particularly, we found that if the attributes mentioned in our extended likeability framework (see Figure 3) are provided, then the gratifications sought by preschoolers can be obtained. Because preschoolers will evaluate these attributes positively, they will also actively engage in using the media object and perceive the media object as "likeable." The more gratifications that can be obtained, the more likeable an object seems to be.

4. Discussion and Future Work

With the emergence of new technologies, the HCI community is fully aware of the need for theories and methods for likeability research. In this article, we aimed at demonstrating how the uses and gratifications paradigm together with expectancy value theory could fill in one of the blanks in designing likeable applications for preschoolers.

The paper remains exploratory, at best hypothesis-generating, in the methods used to elicit responses, in the list of attributes and gratifications derived and in the proposed structure linking these. At the same time, this outlines the further efforts are needed to achieve progress in this area.

As for the model structure and links, further inquiry into means-end analysis might help us to refine or to validate the ELF. As for the attributes and gratifications mentioned in our applied model, they should be considered as inputs for further confirmatory and structured research, provided that appropriate structured research tools can be found or developed to study and interview our young target audience. This brings us to our last endeavour. We are dealing with younger, not yet literate children, and we need to further investigate methods appropriate for eliciting responses from our young targets. Adapting research methods to preschoolers, such as the laddering technique, is a challenging but necessary condition in order to conduct further research.

Therefore, although we are convinced of the possibility of the ELF, we do not pretend that this framework is already fully established. However, we are convinced that the blueprint of the ELF can serve as a practical guide in many third-wave research projects dealing with broad issues. Furthermore, we hope that the list of gratifications and attributes, such as the one derived from our "initial" research, can inform the community and help to understand the diverse gratifications and the ways to translate these into design solutions.

References

[1] S. Bødker, "When second wave HCI meets third wave challenges," in *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles (NordiCHI '06)*, vol. 189, pp. 1–8, Oslo, Norway, October 2006.

[2] G. Cockton, "Designing worth is worth designing," in *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles (NordiCHI '06)*, vol. 189, pp. 165–174, Oslo, Norway, October 2006.

[3] H. Stelmaszewska, B. Fields, and A. Blandford, "Conceptualizing user hedonic experience," in *Proceedings of the 12th European Conference on Cognitive Ergonomics (ECCE '04)*, pp. 83–89, York, UK, September 2004.

[4] B. Zaman and V. vanden Abeele, "Towards a likeability framework that meets child-computer interaction & communication sciences," in *Proceedings of the 6th International Conference on Interaction Design and Children (IDC '07)*, pp. 1–8, Aalborg, Denmark, June 2007.

[5] C. Stephen and L. Plowman, *ICT in Pre-School: A 'Benign Addition'? A Review of the Literature on ICT in Pre-School Settings*, Learning and Teaching Scotland, Dundee, UK, 2002.

[6] M. Hassenzahl, "The effect of perceived hedonic quality on product appealingness," *International Journal of Human-Computer Interaction*, vol. 13, no. 4, pp. 481–499, 2001.

[7] E. Katz, J. G. Blumler, and M. Gurevitch, "Uses and gratifications research," *Public Opinion Quarterly*, vol. 37, no. 4, pp. 509–523, 1973.

[8] P. Palmgreen and J. D. Rayburn, "An expectancy-value approach to media gratifications," in *Media Gratification Research: Current Perspectives*, K. E. Rosengren, L. A. Wenner, and P. Palmgreen, Eds., Sage, Beverly Hills, Calif, USA, 1985.

[9] P. Palmgreen, "Uses and gratifications: a theoretical perspective," in *Communication Yearbook*, R. N. Bostrom, Ed., vol. 8, pp. 61–72, Sage, Beverly Hills, Calif, USA, 1984.

[10] K. E. Rosengren, "Uses and gratifications: a paradigm outlined," in *The Uses of Mass Communications: Current Perspectives of Gratifications Research*, J. G. Blumler and E. Katz, Eds., pp. 269–286, Sage, Beverly Hills, Calif, USA, 1974.

[11] K. E. Rosengren, L. A. Wenner, and P. Palmgreen, *Media Gratifications Research: Current Perspectives*, Sage, Beverly Hills, Calif, USA, 1985.

[12] K. Lucas and J. L. Sherry, "Sex differences in video game play: a communication-based explanation," *Communication Research*, vol. 31, no. 5, pp. 499–523, 2004.

[13] R. L. Mandryk, "Objectively evaluating entertainment technology," in *Proceedings of the CHI Extended Abstracts on Human Factors in Computing Systems (CHI '04)*, pp. 1057–1058, ACM, Vienna, Austria, April 2004.

[14] E. S. Tan and J. Jansz, "The game experience," in *Product Experience*, H. N. Schifferstein and P. P. Hekkert, Eds., pp. 531–556, Elsevier, San Diego, Calif, USA, 2008.

[15] E. Goffman, "Fun in games," in *Encounters: Two Studies in the Sociology of Interaction*, E. Goffman, Ed., pp. 15–81, Bobbs-Merrill, Indianapolis, Ind, USA, 1961.

[16] K. Overbeeke, C. Djajadiningrat, S. Hummels, S. Wensveen, and J. Frens, "Let's make things engaging," in *Funology: From Usability to Enjoyment*, M. Blythe, K. Overbeeke, A. Monk, and P. Wright, Eds., pp. 7–18, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[17] C. C. M. Hummels, K. Overbeeke, and A. van der Helm, "The interactive installation ISH: in search of resonant human product Interaction," in *Funology: From Usability to Enjoyment*, M. A. Blythe, K. Overbeeke, A. F. Monk, and P. C. Wright, Eds., pp. 265–274, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[18] P. Dourish, *Where the Action Is: The Foundations of Embodied Interaction*, MIT Press, Cambridge, Mass, USA, 2001.

[19] M. Blythe and M. Hassenzahl, "The semantics of fun: differentiating enjoyable experiences," in *Funology: From Usability to Enjoyment*, M. A. Blythe, K. Overbeeke, A. F. Monk, and P. Wright, Eds., pp. 91–42, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[20] L. Hult, M. Irestig, and J. Lundberg, "Design perspectives," *Human-Computer Interaction*, vol. 21, no. 1, pp. 5–48, 2006.

[21] W. Schramm, J. Lyle, and E. Parker, *Television in the Lives of Our Children*, Stanford University Press, Palo Alto, Calif, USA, 1961.

[22] A. Cooper and R. M. Reimann, *About Face 2.0: The Essentials of Interaction Design*, John Wiley & Sons, New York, NY, USA, 2003.

[23] J. G. Blumler and E. Katz, *The Uses of Mass Communications: Current Perspectives on Gratifications Research*, Sage, Beverly Hills, Calif, USA, 1974.

[24] D. McQuail, *McQuail's Mass Communication Theory*, Sage, London, UK, 2000.

[25] J. L. Sherry, K. Lucas, B. S. Greenberg, and K. Lachlan, "Video game uses and gratifications as predicators of use and game preference," in *Playing Video Games: Motives, Responses, and Consequences*, P. Vorderer and J. Bryant, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2006.

[26] J. L. Sherry and K. Lucas, "Video game uses and gratifications as predictors of use and game preference," in *Proceedings of the 53rd Annual Meeting of the International Communication Association (ICA '03)*, San Diego, Calif, USA, May 2003.

[27] T. E. Ruggiero, "Uses and gratifications theory in the 21st century," *Mass Communications and Society*, vol. 3, no. 1, pp. 3–37, 2000.

[28] A. H. Maslow, R. Frager, C. McReynolds, R. Cox, and J. Fadiman, *Motivation and Personality*, Addison Wesley-Longman, New York, NY, USA, 3rd edition, 1987.

[29] M. Fishbein and I. Ajzen, "Attitudes towards objects as predictors of single and multiple behavioral criteria," *Psychological Review*, vol. 81, no. 1, pp. 59–74, 1974.

[30] M. Fishbein and I. Ajzen, *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*, Addison-Wesley, Reading, Mass, USA, 1975.

[31] K. G. Grunert and T. Bech-Larsen, "Explaining choice option attractiveness by beliefs elicited by the laddering method," *Journal of Economic Psychology*, vol. 26, no. 2, pp. 223–241, 2005.

[32] T. J. Reynolds and J. Gutman, "Laddering theory, method, analysis, and interpretation," *Journal of Advertising Research*, vol. 28, no. 1, pp. 11–31, 1988.

[33] D. P. Subramony, Introducing a "Means-End" Approach to Human-Computer Interaction: Why Users Choose Particular Web Sites Over Others. AACE, Norfolk, 2002, http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/1a/15.pdf.

[34] B. Zaman, "Introducing contextual laddering to evaluate the likeability of games," *Cognition, Technology & Work*, vol. 10, no. 2, pp. 107–117, 2007.

[35] CuTI.Website for research on Cuddle Toy Interfaces, http://wiki.groept.be/confluence/display/Cu/Home.

[36] V. vanden Abeele and B. Zaman, "Reflection in action," 2008, http://wiki.groept.be/-confluence/download /attachments/4619/reflection+in+action_anonymous.pdf?version=1.

[37] V. vanden Abeele and B. Zaman, "Fun for preschoolers in the home environment," 2008, http://-wiki.groept.be/confluence/download/attachments/4619/Fun+for+preschoolers.pdf?version=1.

[38] A. H. Maslow, *Motivation and Personality*, Harper, New York, NY, USA, 1954.

[39] L. E. Berk, *Development through the Lifespan*, Allyn & Bacon, Boston, Mass, USA, 3rd edition, 1999.

[40] R. Oerter, "Spiel as Lebensbewältigung," in *Handbuch Medien: Computer Spiele*, J. Fritz and W. Fehr, Eds., pp. 59–65, Bundeszentrale für politische Bildung, Bonn, Germany, 2000.

[41] F. P. Rice, *Human Development*, Prentice-Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1992.

[42] B. Sutton-Smith, *The Ambiguity of Play*, Massachussets & Londen, Cambridge, UK, 1992.

[43] J. Van Looy, *The promise of Perfection. A Cultural analysis of games*, Ph.D. dissertation, Catholic University of Louvain, Louvain, Belgium, 2006.

[44] M. Csikszentmihalyi, *Flow: The Classic Work on How to Achieve Happiness*, Rider, London, UK, 2002.

[45] W. Barendregt and M. M. Bekker, "Extended guidelines for usability (and fun) testing with children," in *Proceedings of the 9th Annual SIGCHI.NL Conference "HCI Close to You"*, Den Haag, The Netherlands, October 2005.

[46] A. Druin, B. Bederson, A. Boltman, A. Miura, D. Knotts-Callahan, and M. Platt, "Children as our technology design partners," in *The Design of Children's Technology*, A. Druin, Ed., pp. 51–72, Morgan Kaufmann, San Francisco, Calif, USA, 1999.

[47] A. Gopnik, A. Meltzoff, and P. Kuhl, *How Babies Think*, Phoenix, London, UK, 2001.

[48] M. von Salisch, C. Oppl, and A. Kristen, "What attracts children?" in *Playing Video Games: Motives, Responses, and Consequences*, P. Vorderer and J. Bryant, Eds., Lawrence Erlbaum Associates, Mahwah, NY, USA, 2006.

[49] P. B. Brandtzaeg, A. Følstad, and J. Heim, "Enjoyment: lessons from Karasek," in *Funology: From Usability to Enjoyment*, M. Blythe, K. Overbeeke, A. F. Monk, and P. Wright, Eds., pp. 55–65, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[50] P. Vorderer, C. Klimmt, and U. Ritterfeld, "Enjoyment: at the heart of media entertainment," *Communication Theory*, vol. 14, no. 4, pp. 388–408, 2004.

[51] W. C. Schutz, *FIRO: A Three-Dimensional Theory of Interpersonal Behavior*, Rinehart, New York, NY, USA, 1958.

[52] J. Jansz and M. Tanis, "Appeal of playing online first person shooter games," *CyberPsychology & Behavior*, vol. 10, no. 1, pp. 133–136, 2007.

[53] R. Caillois, *Man, Play and Games*, Meyer Barash, Chicago, Ill, USA, 2001.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 597629, 9 pages
doi:10.1155/2008/597629

Research Article

Child-Centered Evaluation: Broadening the Child/Designer Dyad

Sofia Pardo, Steve Howard, and Frank Vetere

Interaction Design Group, Department of Information Systems, The University of Melbourne, Level 4 111 Barry Street, Carlton, Melbourne 3010, Australia

Correspondence should be addressed to Sofia Pardo, miriamp@pgrad.unimelb.edu.au

Received 1 October 2007; Accepted 6 July 2008

Recommended by Adrian Cheok

Some settings challenge a literal interpretation of user-centered design orthodoxy; that design is best done *for* a user, by designing *with* that user. We explore the value that a copresent proxy and interpreter can bring to certain hard-to-reach or difficult-to-interpret situations; in this case the evaluation of educational software intended to be used by children. We discuss the effect that introducing a teacher had on the results of the evaluation and conclude that adding an expert-based component to evaluations increased its diagnostic power.

1. Introduction

Certain hard-to-reach places make adopting a literal interpretation of user-centred design difficult. Designing with users can be challenging if those users have limited communication skills, restricted cognitive abilities [1], or if there are large power differentials present between designer and user. Here, we are concerned with bridging such distances, and our case is the evaluation of educational software intended for use by children.

We present two evaluations, firstly of current practice and secondly of a novel method called Kids and Teacher Integrated Evaluation (KaTIE), and we focus on the role and the effects of the teacher as an adjunct to the traditional designer/child dyad. KaTIE combines elements of both expert and user-based evaluation, and strongly embeds the teacher in the process, aiming to account for both the usability and pedagogical aspects of educational software in tandem.

The next section discusses child-centred design, and surfaces the challenges implied. Then, we describe and evaluate a widely adopted current approach to child-centred evaluation. Findings from this evaluation influenced the development of KaTIE, which we next describe in detail, and report the results of its evaluation. We conclude with a discussion of hybrid evaluation methods, and the effect that the teacher's presence had on the evaluation outputs.

2. Background

2.1. Child-centered design

A central tenet of child-centered design is the primacy of the child-designer conversation, wherein the voice of the child is heard first hand by the designer, without the mediating influence of teachers or parents [2, 3]. This view is however not without its difficulties. Some challenges implied by the child-designer dyad relate to the child's developing but still immature communication skills [4]. Other issues include potential power differentials between the designer and the child that may hamper the gathering of the child's genuine views and opinions [2, 5]. As a user group, children have particular characteristics that need to be taken into consideration as participants in technology evaluation. These considerations have driven the creation of methods specifically tailored for the inclusion of children [6].

Children have been involved as testers in the evaluation of technological products that serve different purposes, such as play, learning, entertainment, or functionality (e.g., enabling products such as word processors, searching engines) [7]. The input collected from children in the evaluation of these technologies has been considered as mostly regarding usability and the level of engagement or fun involved when interacting with the product [8]. Accounting for usability and fun is often sufficient for products whose goal is to entertain or facilitate particular tasks, but are of incomplete value for those products that aim to engage children in learning.

2.2. Challenges in the context of educational software

The evaluation of educational software places additional challenges to the child-designer dyad since a major aim of such evaluations is to gather feedback regarding children's understanding of the concepts/ideas conveyed in the software, as well as determining how the learning goals anticipated by the designer are being met. Learning is not an overt time-independent behavior that can be easily observed and easily measured [9]; gathering evidence of learning often requires the evaluator to make judgments of what is being observed or expressed, and to what learning unfolds over time. In this sense, learning is in sharp contrast with current approaches to the evaluation of usability or fun, where error rates and facial expressions [10] might be taken as indicators of these attributes, respectively.

In the context of educational software, the verbal exchange between designers and children is often the main communication channel by which designer can gauge whether children are gaining new understanding or refining their existing prior knowledge. In this sense, being able to engage with the children in a dialog that uncovers children's understanding is a major concern. In addition to the children's developing verbal skills, the new vocabulary associated with a particular knowledge domain can be problematic. Moreover, children are often unaware of the learning goals of the software and quite often these are not their own [7]. This lack of knowledge becomes a hurdle when trying to determine how well the software supports the children in attaining those goals, as they cannot provide feedback regarding the goals they have not attained yet [8]. Thus communication, domain, and pedagogical literacy each challenge the place of the child in child-centered evaluation.

3. Related work

The challenges associated with the dyad child-designer in the context of educational software have been addressed in different ways. However, the solutions given have mostly resorted to the employment of different evaluation methods and techniques that in conjunction with direct observation and dialog with the children may provide a more complete account of the educational effectiveness of the software. Pre- and postwritten tests have been used to determine any learning outcomes [11, 12]. The conduct of heuristic evaluations [13, 14] is also a common practice, as experts are able to provide informed views on the educational soundness of the instructional design implemented. Alternatively, the self-administration of questionnaires has been used to determine the pedagogical usability of educational software [15].

These methods are of peripheral nature to child-centered design practices, as they provide complementary views to those of children and designers. In this sense, results yielded by pre- and post-test, expert reviews or self-administered questionnaires need to be amalgamated in such a way that a single body of feedback is produced. This amalgamation process can be challenging, it has been found that children's views are not necessarily in synchrony with good educational practice [8]. What children may consider amusing and entertaining at the interface may hinder their engagement with a particular learning task.

The use of pre- and postwritten tests does not provide feedback regarding the process by which a correct or incorrect answer is given. Obtaining the expected answer on a written test is of little value for design purposes if account of how the children got to that answer is not provided. The predictions established through heuristic evaluations carried out by experts cannot anticipate the learning process that will take place when children interact with the software [9]. Therefore, determining the educational effectiveness of the software using this evaluation method has serious limitations. Lastly, self-administered questionnaires also share the limitations of pre- and postwritten tests as they mostly rely on children's perceptions of what they think they learned and little attention is given to the behaviors that took place during the interaction [16].

These limitations lead to the consideration of alternative evaluation practices that would preserve the benefits of directly involving children, include the views of experts, the designer's insights, and design agenda. As a consequence, these alternative practices are drawn from the child-based and expert-based evaluation methods [17] in order to overcome the limitations that the child-designer dyad poses in the context of educational software.

4. Understanding current practice

In many examples of current practice, the child-designer conversation is broadened by the inclusion of a third stakeholder, an "expert." The expert may bring literacy in technical issues, such as usability evaluator [16] pedagogical insights, such as a teacher or educational expert [18], domain knowledge such as a scientist if the software is designed to teach science concepts [9] or communication, or translation skills to the team [19].

Approaches that combine elements of child-based and expert-based evaluation methods can host different evaluation practices as different arrangements involving children, designers, and experts can take place. In this paper, we focus on arrangements wherein experts, more specifically teachers, are involved in order to determine the learning path followed by the children as they use the software and the attainment of the anticipated learning goals. This focus however is not indifferent to the overlapping nature of usability and learning when children interact with educational software. We adhere to the compartmentalized view of others [11] who consider that these two dimensions although interrelated can be focally evaluated.

Teacher involvement in evaluation, but also in other parts of the design process, has been seen as undesirable as the existing power relationship between teachers and students could lead to a situation wherein the children might feel tested or compelled to perform well [2]. This concern has kept teachers on the periphery of child-based evaluation. However, where computing is concerned these power differentials can be inverted as children, increasingly "digital natives," are often more technically literate than their "digital immigrant" teachers [20]. In this sense, the involvement of teachers does not seem to strengthen the existing power differentials between adults and children. Despite this concern, the following two evaluation methods involve teachers in child-based evaluations in two different ways. The subsequent findings present the influence teachers' participation had on the outcomes produced by the evaluations.

4.1. In-school evaluation (ISE)

ISE is an evaluation method created by one of the biggest educational software producer in Australia. This method has been used extensively across different educational contexts, and has involved over 300 schools in Australia and New Zealand. ISE emerged from the need to reach large numbers of children from different socioeconomic backgrounds in order to create software that is suitable for different types of learners in different contexts.

In ISE, both teachers and designers work in a parallel fashion with a pair of students each, often in the same room but apart from each other to minimize disruption. Teacher and designer use the same data collection protocol and this is provided by the designer as a form to be completed in real-time during the evaluation. The form guides the evaluation process with prescribed questions to ask the children and particular things to observe and record. This form contains the questions the designer has considered relevant in determining the educational effectiveness of the software. The evaluation form is often given to the teacher a week in advance along with the nearly completed software prototype so that familiarization can occur ahead of time.

The rationale behind ISE is partly economic, and partly about process standardization. Having the designer and the teacher play the same role halves the time needed for data collection, and helps facilitate comparisons across different child/school settings by standardizing the process.

4.2. Evaluating ISE

The ISE method was considered a suitable case study given its uniqueness in combining the Child-Based and Expert-Based evaluation methods. Similar hybrid evaluation practices had not been found to our knowledge in Child-Centred Design literature; much less an account of the type of outcomes produced by such practices.

4.2.1. Data collection

The data collection consisted of collection of artefacts, that is, the evaluation forms containing the hand-written notes of the evaluators (teacher and designer), video footage of the evaluation process and interviews with all participants (teacher, designer, and children). As this paper focuses on the feedback collected, the findings here presented are concerned exclusively with the analysis carried out on the hand-written notes.

The case study involved the participation of one designer from the software company, who was well experienced in the ISE method, two primary school teachers (from two different schools) who had used ISE a few times before, and a total of 16 children between 7–9 years old (8 children per school). The evaluations took place at two primary schools over a period of 2 hours each, thus each teacher performed two evaluations and the designer performed 4 evaluations in total (2 per school).

Four educational software prototypes were evaluated with the ISE method. Each prototype was evaluated twice, by the teacher and the designer in a simultaneous fashion. Two prototypes aimed to teach children about the relationship between sample space and likelihood of outcomes, the third one taught about classification of substances according to their behavior in water, and the fourth taught about energy chains.

4.2.2. Data analysis

The analysis of the feedback collected during the ISE evaluations was typed in into tables for coding. The coding scheme employed to analyze the feedback emerged from the data and it was refined through several iterations. As the ISE evaluation form contained prescribed questions, the feedback was broken down into instances that followed the same structure of the questions. In this way an instance could correspond to a prescribed question and its associated hand-written note, or to any notes that were not necessary linked to a question (e.g., notes written on the margin of the page).

The coding scheme consisted of four-high-level codes, which were broken down into subcodes for a total of 13 codes. The feedback was divided into reporting, descriptive, diagnostic, and advisory types. A single feedback instance could be coded as belonging to one of these types or a combination of them; hence they were further coded as single or combined. Feedback coded as reporting regarded to those wherein the evaluator was reporting the verbal response of the children to the questions asked. Within the reporting feedback, a distinction was made between those instances where the verbal answer/question referred to the content conveyed by the software, to the interface or to improvements children thought could be made to the software.

Descriptive feedback included observations made by the evaluators of what was taking place as the children interacted with the software. This feedback was broken down into those instances where the description referred to the interface/navigation, to technical problems, to the evaluator behavior, and to the learner behavior. The latter was further classified into those instances conveying learners' emotional responses, interaction between children and approach to learning task. Diagnostic feedback included those feedback instances wherein the evaluator made a judgment. This type of feedback was either diagnosing the interface/navigation or the learner's understanding or behaviors. Finally, advisory feedback included instances where the evaluator made suggestions to change or improve something in the software. These suggestions were classified into those addressing changes to the interface/navigation and those regarding the learning tasks or instructional design in order to better support children's learning.

The following is an example of a combined feedback instance collected through ISE method:

> Did the students enjoy the animations? Please record any comments. (Prescribed instruction) *"Is it finished"* [**Reporting**]
>
> *Would be good to see propeller working* [**Advisory**] *as these students don't know how a boat travel thru water* [**Diagnostic**]. *Would improve understanding* [**Diagnostic**].

As the evaluations were focused on determining the educational effectiveness of the prototypes, three expectations were anticipated regarding the amount and type of feedback collected. First, it was expected that the feedback coded as referring exclusively to the interface/navigation would be less in number. Second, it was expected that those feedback instances providing insights into the learner and the evaluator's behavior would have a significant percentage. Thirdly, as learning is not an overt behavior, it was expected that an increased number of diagnostic feedback would be found, either on its own or in conjunction with descriptive, reporting, and advisory feedback.

4.2.3. Findings

The results of the coding process are presented at two levels. First, differences and similarities between the teachers and the designer feedback are identified across all evaluations, and second overall tendencies on the type of feedback collected are also drawn. There were no major differences between the feedback collected by the teachers and the designer across the four-software prototypes evaluated. The amount of feedback collected by both was approximately the same, with teachers writing slightly more instances in three out of four evaluations. The amount of reporting, descriptive, and diagnostic feedback was similar, while advisory feedback was mostly reported by the designer.

The feedback collected by the teachers and the designer remained focused on the learning aspects of the interaction across all evaluations, as the number of single feedback instances regarding the interface/navigation of the software was less. Most reporting and descriptive feedback regarded learners' verbal and nonverbal behavior, respectively, hence providing a rich account of the children interaction with the learning tasks.

The similarities of the feedback collected by the teachers and the designer are perhaps related to the evaluation form and set up of the ISE method. The focus on learning aspects of the interaction can be associated to the overall flavor the evaluation form employed has, as most of the prescribed questions and issues to observe aim to explore children's understanding of the concepts introduced by the software. The highly structured evaluation form may have been beneficial in keeping both the teacher and the designer focused on the purpose of the evaluation, to determine the educational effectiveness of the software, and stop them from wandering into other aspects of the interaction, such as usability.

The lack of significant differences between the type of feedback collected by the teachers and designers' feedback can be a result of the prescriptive evaluation form, as this tended to homogenize the feedback collected. This standardization may have resulted in the reduction of the potential added value that involving teachers may have brought. Teachers' involvement in ISE method did not take advantage of their expertise as educator, as the setup resembled a form-filling exercise wherein the questions prescribed the flow and interaction between teacher and children. In this sense, the ISE method heavily relied on the evaluation form as opposed on the insights the teacher might have brought to the evaluation as an expert educator.

At an overall level, the type of feedback collected through the ISE method showed that a significant amount of the feedback instances was of descriptive and reporting nature with less number of diagnostic types. This tendency can also be the result of the structure of the evaluation form, as this one mostly encouraged the collection of these feedbacks. The diagnostic section on the form was located at the end and it consisted of a liker scale. The reduced amount of diagnostic feedback was not perceived as problematic by the designer (as commented in interview) as the main purpose was to be able to recreate what happened on the field (reporting and descriptive feedback) as opposed to emitting judgments. This is also in accordance with the nonexpert role played by the teachers during in the ISE method.

It can be assumed that the diagnosis regarding the educational effectiveness of the software with the ISE method is carried out outside the field based on what was reported and described in the evaluation form. This diagnostic practice takes as face value what is written down and requires the close examination of the hand-written notes to determine whether the prototype achieved what it was meant to. A detached examination of the feedback captured in the ISE evaluation form can be misleading as in some cases children's responses recorded on the form may require further interpretation as to: in which context and after how much and what sort of prompting were the children able to articulate a particular answer.

Our findings show that ISE method does not provide insight into educational effectiveness; rather this is determined by a third party based on what was described and reported on the evaluation forms. Moreover, teacher's involvement, although central to the evaluation, is not as a pedagogical expert, but of a form-filling aid. The collection of diagnostic feedback is fundamental in determining the educational effectiveness of the software, as it is in the field with all the necessary contextual clues that an informed and empirically based judgment can be attained. Although there is considerable value on reporting children's verbal responses/comments and describing their behaviors, these fall short in conveying a complete picture of the children's learning experience.

There are opportunities to provide an integrated commentary by merging child and expert-based evaluation methods in such a way that the teacher is allowed to play to his or her strengths. In doing so the teacher may be encouraged to provide diagnostic feedback regarding educational effectiveness. Encouraging the teacher and designer to share their insights and come to a collective view could facilitate this. Care would need to be taken so as not to overly constrain the teacher's commentary with design-centric questions. An open evaluation form could potentially encourage teachers to draw from their expertise and guide the dialog with the child.

5. Changing the place of the teacher

Although ISE is a hybrid child- and expert-based evaluation methods that includes a teacher, the centrality of the child-designer conversation is not diminished. Rather it is complemented by a child-teacher conversation, but one in which the teacher is playing the role of a designer. What might a teacher bring to the process if his or her pedagogical voice were strengthened? How might this be achieved?

KaTIE extends current practice, by facilitating a child-designer-teacher conversation, and aims to strengthen the account of the educational effectiveness of the software resulting from the evaluation. By including teachers as teachers, the evaluation's set up is changed radically. KaTIE is a collaborative [21] and lightweight [22] or discount method, wherein the inclusion of teachers, designers, and children is grounded in their respective areas of expertise in order to gather rapid insights into the educational effectiveness of software. KaTIE agrees with the view that the evaluation of the pedagogical design implemented in educational software remains an adults' matter [8] but also firmly believes that adults' judgment needs to be grounded not just on theoretical views but on direct observation and dialog with children.

The collaboration between the teacher and the designer in KaTIE takes place over a period of three consecutive hours distributed as follows.

(1) Preparation stage (1 hour). This stage aims to create rapport between teacher and designer since KaTIE is not based on a continuing relationship between them. There is a need to develop a shared understanding of the evaluation purpose as well as a sense of importance of their respective roles and perceptions. This stage also aims to familiarize the teacher with the software and the evaluation form that will be employed.

(2) Data collection stage (1 hour). The second stage involves observing and engaging in dialog with the children as they use the software, with a particular focus on children's understanding of the content conveyed by the software and the overall learning experience the software supports. In this stage the teacher leads the dialog with the children while the designer mostly observes in the background providing support if required or asking additional questions to the children. Both, teacher and designer also play the role of note takers.

(3) Reporting stage (1 hour). This last stage is of crucial importance in KaTIE as it consolidates the outcomes of the evaluation. This stage requires the teacher and the designer to engage in a conversation wherein their observations and written notes are shared with the purpose of creating a rich picture of the children's interaction with the software. This conversation is assisted with an open ended template. This stage also brings to the table the learning goals and objectives that have been anticipated by the designers and looks at them under the light of the teacher and designer observations. The reporting stage allows for the consideration of solutions that can address identified issues with the pedagogical design as well as more open ended brainstorming.

These three stages and the key tasks associated with them were summarized into three reference cards for the designers' guidance. These cards were given to the software designers along with two short documents containing the tenets on which the KaTIE method was based. Although KaTIE has its origins in ISE, it has distinctive characteristics. The number of children involved in KaTIE is less than the ISE method as a consequence of the three-hour duration of the evaluation. The teacher and designer focus their attention on the same pair of students, rather than a different pair each, in order to gather and merge their complementary views on the educational effectiveness of the software.

In addition, the evaluation form used in KaTIE does not contain prescribed questions to ask the children or specific things to observe as they interact with the software. The evaluation form provides general instructions to the facilitator (the teacher) to explore children's understanding of the content conveyed by the software. The openness of the evaluation form aims to encourage teachers to draw from their expertise and knowledge without being constrained by predetermined design-centric questions. This type of form was considered adequate as unanticipated issues the designer might have not thought of could be also identified by the teacher, hence adding extra value to the feedback otherwise collected by a prescribed form.

In terms of the evaluation outcomes, KaTIE's aim is to strengthen the commentary on the educational effectiveness of the software, without diminishing the usability and engagement components of the interaction. The ISE evaluation evidenced the predominant reporting and descriptive nature of the commentary produced by the teachers and the designers. KaTIE aims to increase diagnostic commentary, in the hope that these better account for educational effectiveness.

5.1. Comparing KaTIE and ISE

KaTIE's empirical evaluation consisted of four-field studies that aimed to contrast the commentary provided by the designer and the teacher in ISE with the combined commentary produced in KaTIE. All designers were given a brief introduction to the ISE method and the KaTIE method. Three out of the four designers who participated were new to both ISE and KaTIE, thus there were no bias or preferences towards either. One of the designers received instructions only on KaTIE as she mastered the ISE method.

5.1.1. Data collection

The data collection for the evaluation of KaTIE resembled the one employed in the evaluation of ISE: video footage, interviews, and collection of artefacts. As mentioned before the findings here presented regarding these artefacts exclusively.

Each field study involved the participation of one software designer, one primary school teacher, and 6 children between 7–10 years old. Four software companies and four primary schools were involved working in pairs; each pair evaluated different educational software, which was created by the associated software company. The four educational software evaluated had different learning purposes, one aimed to teach English as second languages (ESL) children about insects and their habitats; the second introduced hospital terminology and some relaxations techniques; the third software was an argument-mapping tool for which some online exercises involving the concepts of reasons and objections were tested, and the fourth consisted of a reading program that involved some reading comprehension activities.

Each primary school was visited twice, one visit for the implementation of ISE and the other for the implementation of KaTIE. The order in which ISE and KaTIE were implemented was alternated across the four-field studies to minimize order effects. The number of commentary sets collected per field study was three, two collected through the ISE method (one belongs to the teacher and the other to the designer) and the third through the KaTIE method (teacher and designer independent and combined notes).

5.1.2. Data analysis

The analysis carried out on the commentaries collected followed three steps. The first step was coding according to the previously described coding scheme. The second step involved an audience review and the third step involved an expert review panel. These reviews had a three-fold purpose; first, they aimed to provide some measure of the level of usefulness and comprehensiveness of the feedback collected through ISE and KaTIE methods. Second, they served as a way of testing the coding scheme, as the reviewers were not given any coding scheme to guide their reviews. And third to contextualize the findings in terms of what was desirable feedback for designers in general.

The coding of the feedback produced by ISE method followed the same procedure as in the first evaluation presented. The feedback produced by KaTIE on the other hand required a slightly different coding approach as there were no prescribed questions to guide the identification of feedback instances. As the feedback collected through KaTIE tended to be narrative, clues such as dashes and bullets that implied a different idea or change in topic.

After the feedback was typed in and coded, the audience review was undertaken. The audience was defined as a member of the design team that was/had designed the software under evaluation. Therefore, a second software designer for every software company that participated played the role of audience. The audience was given a typed in version of the feedback to facilitate comparisons, but they were also given the original written notes for reference if required. The feedback was deidentified, so the audience did not know which feedback belonged to which evaluator (teacher, designer, or both).

The expert review consisted of a panel of three experts, who had background in education and learning technologies. All experts had a good sense of designing technology for learning and teaching, just one of them had direct experience designing and evaluating software for primary school children. The expert panel gathered together four times, every time to review the three sets of feedback belonging to one of the softwares evaluated. Experts were also given a typed in version of the feedback and one of the researchers was present during all review sessions. As these experts had no knowledge of the software evaluated they were introduced to it by the researcher to ensure the feedback was contextualized within the particularities of each computer program. As with the audience review, experts were not given any coding or framework for reviewing besides some general instructions.

These instructions consisted of the identification of any differences across the three sets of feedback and the selection, when possible, of the feedback that was perceived as most useful and more comprehensive regarding children's learning processes as they interacted with the software.

5.1.3. Findings

The findings of the feedback collected through KaTIE method are presented in three sections, coding, audience review, and expert review.

Coding

The coding results of the feedback collected through the ISE method were consistent with those yielded in the previous evaluation described in Section 4.2. There were fewer instances of diagnostic feedback as compared to the reporting and descriptive types. Equally, advisory feedback remained very low. There were not also major differences between the teachers and the designer's feedback. This confirms the homogenizing effect, previously identified, that a prescribed evaluation form has on the type of feedback collected. The feedback also referred predominantly to the learner verbal and nonverbal behavior as opposed to the interface/navigation of the software.

The coding of the feedback gathered through KaTIE was collected in three sections associated with the notes taken independently by the teacher and the designer during the data collection stage and the concluding notes taken by the designer during the reporting stage. Although all these notes were considered part of the feedback, the dissection allowed a better account of where the different types of feedback were mostly being produced.

As it was expected, the notes taken by the designer and the teacher while observing the children interacting with the software tended to be of reporting and descriptive nature. One exception to this tendency was the notes taken by a teacher in one of the filed studies where the amount of diagnostic feedback was greater than the descriptive and reporting. This finding reflects the nature of the task teacher and designer were engaged in: observing and talking with the children; thus it can be reasonably expected that at this stage of the evaluation, less judgment would be recorded.

On the other hand, an increased number of diagnostic feedback was produced during the reporting stage wherein teacher and designer got to discuss their observations. The number of diagnostic feedback coded in the combined notes of the teacher and the designer resulted in an overall increase in the number of diagnostic feedback collected in KaTIE as opposed to ISE. This diagnostic feedback remained mostly focused on the learning aspects of the interaction despite the omission of prescribed questions.

An increased number of advisory feedback was also collected through KaTIE; however the number of those instances regarding children's interaction with the learning tasks or the overall instructional design was equal to those regarding the interface/navigation of the software. This finding suggests that teachers and designers also engaged in considering potential solutions to some of the issues found during the evaluation.

Audience review

All reviews collected from the audience identified similar types of feedback. This validated the coding scheme employed previously. The following interview extract evidences the similarity between the coding scheme and the audience views:

> "*... you really need to offer some view, some judgment as to what the implications of this observation is for improving the software or for improving the interaction between the kids and the software.... As a developer you often have to work with this and developers do not necessarily have an education background even though they are developing educational activities and so they need that interpretation there.*" Audience reviewer 4 [Diagnostic and descriptive feedback].

When the audience was asked to determine the usefulness and comprehensiveness of the feedback regarding the educational effectiveness of the software, the reviewers considered that the feedback collected through KaTIE was "too discursive" and that the format was difficult to read as the information followed a narrative style. This was considered cumbersome, as designers are often looking for the next steps they need to follow with the prototype. They also referred to the feedback in KaTIE as hard to interpret for a third party, as it lacked the structure the ISE feedback had in abundance.

As a consequence, the audience tended to select in most cases the feedback provided by ISE as their choice. However, some of the reviewers considered that the ISE and KaTIE feedback were complementary as the latter could provide a richer context for the answers and descriptions recorded in ISE. This finding shows that the feedback collected through KaTIE was overwhelming for most audience reviewers, hence it requires further filtering before a third party, who has not been in the field, could read it and make use of it.

In the context of formative evaluations that look into the educational effectiveness of the software, all types of feedback were considered desirable. Nonetheless, the diagnostic feedback as a "secondary source" type of feedback raised issues of reliability on the judgments made by the evaluators. The presence of advisory and diagnostic feedback was on the other hand seen as less useful if they were not complemented with other types of feedback that would provide context to particular suggestions or judgments. This shows that combined types of feedback are more useful to designers as compared to the single types. A reviewer affirmed that "things can always be done better," so if an advisory comment was not linked to a problematic issue identified with the software, this commentary would be considered less useful.

Expert review

The experts also identified very similar types of feedback across the four-group reviews undertaken. One drawback identified by the audience reviewers regarded the conciseness of some of the feedback found in ISE. Some of children's responses required the reader to "guess" how the children came about particular answers, and most importantly if the answers jotted down represented the first, second, or third attempt on the evaluator's behalf to uncover the children's understanding.

As the audience reviewers, the experts considered the presence of diagnostic feedback necessary when trying to determine the educational effectiveness of the software, as reporting and describing what the students are saying or doing is insufficient to convey the idea of the degree of engagement with the learning task. Nevertheless, the way in which the KATIE method supported the collection of diagnostic feedback was perceived, again, as not necessarily useful.

As it was mentioned earlier, most of the diagnostic feedback collected through KaTIE was produced during the reporting stage, hence these feedback tended to be detached of the reporting and descriptive feedback collected in the data collection stage. This was perceived by experts as problematic since for judgments to be informative a description of the context in which these are made is necessary. Some of the audience reviewers also expected these judgments to be grounded on evidence of children's behaviors or responses.

6. Discussion

Although both ISE and KaTIE can be situated in a hybrid space between child- and Expert-based evaluations, this is not so clearly reflected in the type of feedback collected by them. The feedback produced using ISE is predominantly child-based, that is, reporting and describing children's behaviors; while diagnostic comments are less accounted for reflecting the minor role of expert-based feedback. We argue that the feedback produced with KaTIE is more balanced in this sense, increasing expert-based feedback without diminishing child-based feedback. Due to this balance, KaTIE conveys a richer picture of the children's learning experience by providing hybrid and merged feedback. As it was suggested by the reviewers, in the case of diagnostic feedback, there is a need to provide contextual information, such as descriptive or reporting type of feedback, for a more complete account of the educational effectiveness of the software. In this context, combined feedback as opposed to single types may ensue in more useful and comprehensive results.

The inclusion of teachers as experts in KaTIE has added value to the feedback collected. However, this value was hard to see given the format in which the feedback was collected, its narrative style needs to be refined. The prescribed evaluation form used in ISE guided the reader and dissected the information in more manageable chunks, while the open-ended evaluation form used in KaTIE perhaps serves better a data collection purpose rather than a reporting purpose.

In the context of current practice, KaTIE has shown that it is possible to account for the educational effectiveness of educational software within the boundaries of child-based evaluations and that resorting to independent expert reviews is of limited value. Experts can be included in the child-designer dyad and have a favorable influence on the evaluation outcomes. The use of measuring instruments, such as pre- and postwritten tests, to account for learning outcomes [11, 12] is restricted in the context of formative evaluations as they cannot provide the type of diagnostic feedback obtained with KaTIE. Through observation and dialog with the children, the formulation of diagnostic feedback takes into account contextual information that is not available on a written test.

The use of self-reported questionnaires [16] is not in opposition to KaTIE or ISE methods, as children's perceptions on their learning experiences are also fundamental. However, these should be treated as *perceived* educational effectiveness, which may not conform to the perception of experts and designers. In determining the educational effectiveness of educational software, the involvement of experts brings theoretical and experiential views to the evaluation, the participation of designers accounts for the design rationale behind the software, and the contribution of children accounts for their views on what they find easy, difficult, or fun to do when using the software.

7. Conclusions and future directions

We have demonstrated the added value that teachers can bring to the evaluation of educational software designed for children. By entrenching the teacher meaningfully in the evaluation process *as an expert educator*, we have shown that the diagnostic power of the evaluation and its sensitivity to pedagogical issues can be improved. In contrast to the views of earlier work, we have shown that both designers and children welcome the teachers into the process, and if supported teachers have much to offer to child-centered evaluation.

However, the evaluation process exerts a powerful influence over its practice, and great care is needed in its design if the teachers' voice is to be heard. Teachers' multifarious acts support the child, translate for the designer, and provide the pedagogical critique missing from a typical usability evaluation. Future work should examine the multitude of methodological options for the conduct of child/designer/teacher-centric evaluations, and indeed design more generally.

KaTIE and our findings, in regard to the inclusion of other members of a user's community in the process, may also apply to other situations, where psychological or social factors compromise the user's role in a participative process, or the setting of use is out of reach of the design team.

Acknowledgments

We would like to acknowledge the valuable participation of the software designers, teachers who willingly embarked on the process of implementing a new evaluation method. Also, we would like to thank all the children that took part in the evaluations.

References

[1] P. Francis, L. Firth, and S. Balbo, "Approaching the sensitive user with care: an online Delphi study," in *Proceedings of the International Workshop on Appropriate Methods for Design in Complex and Sensitive Settings. Australasian Conference on Human Computer Interaction (OZCHI '05)*, Canberra, Australia, November 2005.

[2] A. Druin, *The Design of Children's Technology*, Morgan Kaufmann, London, UK, 1999.

[3] A. Druin, "The role of children in the design of new technology," *Behaviour & Information Technology*, vol. 21, no. 1, pp. 1–25, 2002.

[4] P. Markopoulos and M. Bekker, "Interaction design and children," *Interacting with Computers*, vol. 15, no. 2, pp. 141–149, 2003.

[5] C. Jones, L. McIver, L. Gibson, P. Gregor, et al., "Experiences obtained from designing with children," in *Proceedings of the 1st International Conference for Interaction Design and Children (IDC '03)*, Preston, UK, July 2003.

[6] L. Hanna, K. Risden, and K. J. Alexander, "Guidelines for usability testing with children," *Interactions*, vol. 4, no. 5, pp. 9–14, 1997.

[7] J. Read, "Unpublished IDC master class notes," in *Proceedings of the 5th International Conference for Interaction Design and Children (IDC '06)*, Tampere, Finland, June 2006.

[8] M. Scaife, Y. Rogers, F. Aldrich, and M. Davies, "Designing for or designing with? Informant design for interactive learning environments," in *Proceedings of Computer Human Interaction Conference (CHI '97)*, pp. 343–350, Atlanta, Ga, USA, March 1997.

[9] S. W. Draper, M. I. Brown, E. Edgerton, et al., "Observing and measuring the performance of educational technology," TILT Project report 1, University of Galsgow, Glasgow, UK, 1994.

[10] S. MacFarlane, G. Sim, and M. Horton, "Assessing usability and fun in educational software," in *Proceedings of 4th International Conference for Interaction Design and Children (IDC '05)*, pp. 103–109, Boulder, Colo, USA, June 2005.
[11] G. Sim, S. MacFarlane, and J. Read, "All work and no play: measuring fun, usability, and learning in software for children," *Computers & Education*, vol. 46, no. 3, pp. 235–248, 2006.
[12] M. Virvou and V. Tsiriga, "Involving effectively teachers and students in the life cycle of an intelligent tutoring system," *Educational Technology & Society*, vol. 3, no. 3, pp. 511–521, 2000.
[13] P. Albion, "Heuristic evaluation of educational multimedia: from theory to practice," in *Proceeding of the 16th Annual Conference of the Australasian Society for Computers in Learning for Tertiary Education (ASCILITE '99)*, Brisbane, Australia, December 1999.
[14] J. W. Robertson, "Usability and children's software: a user-centred design methodology," *Journal of Computing in Childhood Education*, vol. 5, no. 3-4, pp. 257–271, 1994.
[15] P. Nokelainen, "An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students," *Educational Technology & Society*, vol. 9, no. 2, pp. 178–197, 2006.
[16] K. Beattie, "How to avoid inadequate evaluation of software for learning," in *Interactive Multimedia in University Educational: Designing for Change in Teaching and Learning (A-59)*, K. Beattie, C. McNaught, and S. Wills, Eds., pp. 245–258, Elsevier Science, Amsterdam, The Netherlands, 1994.
[17] J. Preece, Y. Rogers, and H. Sharp, *Interaction Design, Beyond Human-Computer Interaction*, John Wiley & Sons, New York, NY, USA, 2002.
[18] R. Heller, "Evaluating software: a review of the options," *Computers & Education*, vol. 17, no. 4, pp. 285–291, 1991.
[19] L. van Leeuwen, "How children contribute to the design of technology for their own use," in *Developing New Technologies for Young Children*, J. Siraj-Blatchford, Ed., pp. 139–159, Trentham Books, London, UK, 2004.
[20] M. Prensky, "Digital natives, digital immigrants," *On The Horizon*, vol. 9, no. 5, pp. 1–6, 2001.
[21] V. John-Steiner, R. J. Weber, and M. Minnis, "The challenge of studying collaboration," *American Educational Research Journal*, vol. 35, no. 4, pp. 773–783, 1998.
[22] A. Monk, "Lightweight techniques to encourage innovative user interface design," in *User Interface Design, Bridging the Gap from User Requirements to Design*, L. E. Wood, Ed., pp. 109–129, CRC Press, Boca Raton, Fla, USA, 1998.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 258516, 7 pages
doi:10.1155/2008/258516

Research Article

A Mathematical Framework for Interpreting Playing Environments as Media for Information Flow

Tatsuo Motoyoshi, Takashi Hattori, Hiroshi Kawakami, Takayuki Shiose, and Osamu Katai

Department of Systems Science, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan

Correspondence should be addressed to Tatsuo Motoyoshi, motoyosi@sys.i.kyoto-u.ac.jp

Received 1 October 2007; Accepted 18 August 2008

Recommended by Adrian Cheok

This paper proposes a novel strategy of designing play equipments. The strategy introduces two loose constraints as a guideline for designers. The first constraint is "describing unit of play action chain" $\langle O, S, V \rangle$ based on Barthes' semiology, and the second is the infomorphism between designer, play equipment, and players based on channel theory. We provide detailed explanation of the strategy through an example of a designing process of playing environment where the players usage of the play equipment cannot be foreseen.

1. Introduction

When playing computer games, users sometimes tend to add their own original rules to make games more exciting; thus the games (or more generally speaking the play equipment) do not have a unique "right way" to be played but have the potential to be played in various ways. This potential is usually interpreted as a problem to be avoided in the field of conventional system design because it may lead to unexpected danger. In the field of designing play equipment, however, the potential is an important element for producing entertainment.

It is needless to say that designers of conventional systems, to, often assume users' various ways of using a system and handle them one by one. In designing play equipment in which the events caused by the equipment are prepared by the designers (e.g., computer games), a variety of decision branches are considered an essential element for making the play equipment "entertaining"; the knowledge regarding adequate size of the decision branches and how to create them is organized as gamenics [1].

On the other hand, users of such play equipment are not always satisfied with the given decision branches, and sometimes create a new and unpredicted way of playing. In this paper, we focus on this point and target outdoor play equipment that is often seen at playgrounds as equipment for which it is difficult to assume players' intentions.

We expect to gain broad knowledge that also covers "event-foreseeable play equipment" through analysis of such play equipment, for example, factors that cause entertainment or guiding principles of how to design "enjoyable-to-use tools". Unlike gamenics, this knowledge is not organized at present and is a challenging task to tackle.

It is impossible to design these types of targets by combining rigidly pre-constructed building blocks; "correspondence between playing actions and pleasure" and "correspondence between play equipment's characteristics and pleasure" are both context-dependent. Furthermore, we must not assume that the context can be fully clarified because we aim to construct a framework that can tolerate unpredictable ways of playing. This paper therefore first focuses on "correspondence between playing actions and pleasure" in Section 2 and proposes "describing units of the play action chain"; this unit introduces play equipment's components and its characteristics as a context that (not fully but to some extent) regulates the correspondence. This unit functions as a first "loose" constraint. Though play actions and pleasure cannot be uniquely correlated, allowing spontaneous description of the correspondence will lead to a denial of any systematic operation. The adoption of

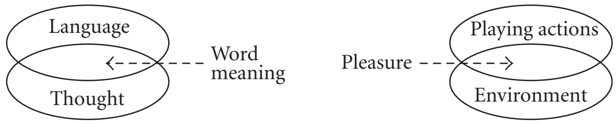


Figure 1: A unit of linguistic thought and playing action.

Table 1

Actions	Therblig
Climbing	Transport empty → grasp → transport loaded → use
Swinging	Transport empty → grasp → transport loaded → use

the new unit is namely an adoption of an intermediate (loose) level of constraint; we fix only the description format and vary the implication of the description itself, case by case. Next in Section 3, by using the format introduced in Section 2, we adopt a new strategy for designing targets whose influence cannot be foreseen. The main idea of the strategy is to individually describe the part that can be locally decided and describe what cannot be locally decided as a correspondence between local description. That is to say, give a correspondence between the 3 factors: assumption of the designer, assumption about the user, and an event that can be generally assumed to happen in a physical environment. Here, we give a second loose constraint. Though bijection, the most rigid correspondence, is too powerful, allowing spontaneous correspondence is meaningless. On loosening the powerful bijection constraint, we can adopt various strategies such as surjection and injection, but in this paper we adopt "infomorphism." Infomorphism is essentially a projection that is consolidated to formulate qualitative information flow, and formation of an infomorphism is assumed as formation of information flow.

2. Analysis of Perception-Action Chain Structure

In a playing environment, pleasures prompt new playing actions that in turn generate new pleasures. These perceptions of pleasures and playing actions are coupled dynamically. In other words, the playing environment involves many "perception-action chains" in terms of ecological psychology [2]. Vygotsky advocated in his psychological studies that the primitive unit of linguistic thought is not a single word or a syllable but "word meaning," which has both properties of language and thought [3]. These properties are coupled dynamically, and variations of these properties generate new "word meanings." Interpreting this notion of linguistic thought to playing environment, a unit of playing actions should have both properties of actions and environment, and should be able to describe relations between play equipment and players in order to grasp the generations of pleasures in the playing environment as shown in Figure 1. Furthermore, variations of playing actions or environment perceptions generating new pleasures correspond with the characteristic of the unit of linguistic thought. We describe analysis of the perception-action chain for outdoor play equipment in the next section, using one of four types of pleasure actions: "imitative action," "resting action," "vertiginous action," "challenging action," which are observed in using outdoor play equipment [4].

2.1. Perception-Action Chain Analysis Using Therblig

To describe processes of generation of pleasures from playing actions, we separate playing action using the "therblig" as a unit that is beyond an analyst's subjectivity. Therblig analysis [5] is the notation evolved from the observation of human manual operation and is made up of 17 types of basic motions called "therbligs." For example, describing two different actions "climbing" and "swinging" observed on the rope of outdoor play equipment using therbligs is performed as shown in Table 1.

This analysis gives the same result. This means that using the therblig cannot derive various pleasures from different playing actions.

Therblig has much less basic motions for analysis for playing actions, so we cannot describe all of the playing actions using therblig. Additionally, therblig does not have properties of environment perception. Generally each playing action does not correspond to a unique "pleasure," so we cannot derive pleasures generated in the relations between play equipment and playing environment even though we can assign one of the therbligs to each action. Furthermore, this method cannot allow us to describe generation of new pleasures from linked actions. Thus, the therblig, which is one of the reductionism units, is not suited for analysis of playing action. The "PTS" has more basic motions [6], but it is not suited for them in the same way.

2.2. A New Unit $\langle O, S, V \rangle$ Based on Barthes' Semiology

As noted in Section 2, reductionism units such as the therblig or PTS are not suited for analysis of playing action. Additionally, a unit of playing actions should have both properties of actions and environment and represents the "pleasure of a combination of actions" that is not just a integration of its constituent actions. In this section, we propose a new unit of playing action based on Barthes' semiology, "the fashion system," which has a syntax structure that is suited for representing linked playing actions.

The fashion system

Barthes proposed "a unit for describing fashion [7]," which is defined as an association of three elements: object (O), support (S), and variant (V) as follows:

S: the garment's technical existence. Support of signification (e.g., the collar),

V: the garment's signifying existence. The point of the matrix from which signification emerges (e.g., open/closed),

O: the material element of the garment. Object of the signification. (e.g., the cardigan).

For example, analysis of "a sporty cardigan" is as follows:

$$S(\text{collar}) + V(\text{open}) + O(\text{cardigan}) \equiv \textbf{sporty}. \quad (1)$$

A new unit $\langle O, S, V\rangle$

We propose "a unit of playing-actions $\langle O, S, V\rangle$" based on the fashion system in Barthes' semiology [7] as follows:

S: a technical factor that determines a position of a part of play equipment (e.g., dropped down vertically),

V: a playing action that is carried out in a part of play equipment (e.g., climbing),

O: a device to be a part of play equipment (e.g., a rope).

For example, describing two different actions "climbing" and "swinging" observed on the rope of outdoor play equipment is as follows, using one of four types of pleasure actions: "imitative action", "resting action", "vertiginous action", "challenging action", which are observed in using outdoor play equipment [4]:

$$\begin{aligned} &S_1(\text{dropped down vertically}) + V_1(\text{climbing}) + O_1(\text{rope}) \\ &\quad \equiv \text{pleasure of } \textbf{Challenging Action}, \\ &S_1(\text{dropped down vertically}) + V_2(\text{swinging}) + O_1(\text{rope}) \\ &\quad \equiv \text{pleasure of } \textbf{Vertiginous Action}. \end{aligned} \quad (2)$$

These units correspond to various pleasures. In particular, we can represent new actions and new pleasures involved in the same device. The substance of pleasures depends on choices of V (actions) in the same parts; also choices of V depend on S and O. Players recognize V and S as characteristic of O. These operations correspond to that of variant V in "the fashion system," in that V has the right of determination of fashion mode and becomes characteristic of O with S.

2.2.1. Binding of Units

In "the fashion system", multiple units link together and generate a new unit and describe fashion mode as follows:

$$\begin{array}{ccc} \langle\text{white} & \text{braid and white} & \text{buttons}\rangle \\ \langle S_1 V_1 \quad O_1\rangle & & \langle S_2 V_2 \quad O_2\rangle \\ \langle S_0 & V_0 & S_0'\rangle \\ & \mathbf{S}. & \end{array} \quad (3)$$

S_0 represents $\langle O_1, S_1, V_1\rangle$. Similarly, S_0' represents $\langle O_2, S_2, V_2\rangle$. Each meaning of units is transmitted to the final unit. We can apply this representation to the playing action as follows:

$$\begin{array}{cc} \text{dropped down vertical climbing rope} & \\ \langle S_1 \qquad V_1 O_1\rangle & \\ \text{drop the rope down} & \text{swinging bar} \\ \mathbf{S_2} & V_2 O_2. \end{array} \quad (4)$$

$\langle O_1, S_1, V_1\rangle$, which is represented by S_2, generates "challenging action" pleasure. S_2 makes a new unit with O_2 and V_2 and generates a new pleasure of "vertiginous action." This means that we can grasp the new pleasure on a larger scale when playing actions that are linked together, which is not described in the method using therbligs. We can adopt this method in analyzing the playing action in connected multiple parts of play equipment.

3. Information Channels in Playing Environment

We discussed the syntax structure of units of playing actions. In a playing environment, there are information flows involving generation processes of new playing actions that are observed on play equipment based on players' interpretations. We propose a mathematical framework for qualitative information flows of designing processes of the playing environment.

In the conventional way of designing play equipment, a designer recollects some devices (e.g., ladder) associated with a specific action (e.g., climb) and integrates the device as a part of play equipment. But in such a case, we cannot capture the emergence of new actions in the playing environment. We cannot adopt the strategy of deriving an action correlated to a combination of devices by consolidating actions correlated to its constituent devices. In this chapter, we interpret that individual descriptions of the part are not devices but a correspondence between three factors: assumption of the designer about "character of play equipment and pleasures," assumption of the user about "character of play equipment and actions afforded from," and equipment and actions that can be generally assumed to happen in a physical environment. Then, we adopt a strategy that can be locally decided on and describe what cannot be locally decided on as a correspondence between local descriptions.

Bijection, the most rigid correspondence, is too powerful, and allowing spontaneous correspondence is meaningless. On loosening the powerful bijection constraint, we can adopt various strategies such as surjection and injection, but in this paper we adopt "infomorphism" in channel theory [8, 9]. Infomorphism is essentially a projection that is consolidated to formulate qualitative information flow, and formation of an infomorphism is assumed as formation of information flow. So, we interpret the playing environment as a new medium that mediates play equipment and players, by focusing on interactions between artifacts and users, and define a model for representing designers—play equipment—players as information channels based on channel theory [8, 9]. Channel theory is the qualitative theory of information advocated by Barwise and Seligman.

Channel theory defines a classification (D) such as $\langle \text{typ}(D), \text{tok}(D), \models_D \rangle$. Given a set of tokens: tok(D) and a set of types: typ(D), $a \models_D \alpha$ means that the type of $a (\in \text{tok}(D))$ is $\alpha (\in \text{typ}(D))$. A unit of playing action $\langle O, S, V \rangle$ has both properties of action $\langle V \rangle$ and environment $\langle O, S \rangle$. It is appropriate to separate both properties to typ and tok in the classification of Channel theory. The separation of both properties can be done in an intensional or extensional way. In this paper, we adopt an extensional separation, in which recitation of properties of environments defines properties of actions that are attributes of playing environments. For playing environment, we define three classifications as follows.

Classification of designers assumption: D_{ds}

tok(D_{ds}): a set of "pleasures T_i".

typ(D_{ds}): a set of "features of playing environment E_i."

$T_i \models_{D_{\text{ds}}} E_i$: Pleasure T_i is observed in environment E_i.

Classification of equipment: D_{pl}

tok(D_{pl}): a set of "playing actions V_i."

typ(D_{pl}): a set of "parts of play equipment O_i."

$V_i \models_{D_{\text{pl}}} O_i$: a part of play equipment O_i enables an action V_i.

Classification of users: D_{ch}

tok(D_{ch}): a set of "playing actions of players V'_i."

typ(D_{ch}): a set of "combinations of parts and their feature $SO_i (= S_i + O_i)$."

$V'_i \models_{D_{\text{ch}}} SO_i$: SO_i enables an action V'_i.

In the classification of users (D_{ch}), types SO_i are defined by combining S_i in terms of units of playing actions $\langle O, S, V \rangle$ with parts of play equipment O_i of D_{pl}. S_i affects O_i in its ability to bring about some new play actions V_i.

The three classifications enable us to interpret designing play equipment as an establishment of an *infomorphism* in terms of channel theory.

3.1. Conventional Design of Play Equipment

An *infomorphism* $f : A \rightleftarrows B$ from A to B is defined as a contravariant pair of functions $f = \langle f^\wedge, f^\vee \rangle$ in channel theory. The function $f^\vee$ is a map from tok(B) to tok(A), and $f^\wedge$ is a map from typ(A) to typ(B). It satisfies the following property:

$$f^\vee(b) \models_A \alpha \Longleftrightarrow b \models_B f^\wedge(\alpha), \tag{5}$$

for each token $b \in \text{tok}(B)$ and each type $\alpha \in \text{typ}(A)$. Figure 2 shows an *infomorphism* $f : A \rightleftarrows B$ from A to B.

In the case of playing environment, the establishment of *infomorphism* from D_{ds} to D_{pl} can be seen as an accomplishment of designing play equipment in a conventional way. Namely, "a pleasure T_i in an environment $E_i (= f^\vee(O_i))$" in the mental world of a designer is implemented in the physical world as the action $V_i (= f^\wedge(T_i))$ by using equipment O_i.

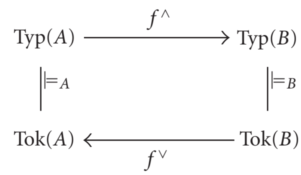


FIGURE 2: An infomorphism from classification A to B.

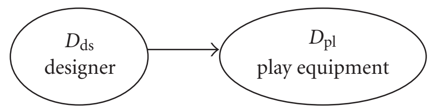


FIGURE 3: Conventional design process modeled by an infomorphism.

Figure 3 illustrates an image of establishment of an informorphism. This method describes only actions corresponding to parts of play equipment based on the designer's logic, and no actions generated by players' intentions. So, designers cannot foresee the emergence of new playing actions and pleasures.

3.2. Novel Design of Play Equipment

Channel theory defines a *channel* as a set of *infomorphisms* sharing the same codomain. In the case of a playing environment, establishment of an *information channel* between D_{ds} and D_{ch} via D_{pl} can be interpreted as an accomplishment of a novel design from the viewpoint of interpreting play environment as a medium of designers, play equipment, and children. Figure 4 illustrates an image of establishment of a channel.

3.2.1. A Concrete Example of Using Chu Map

Classification D_{ds}: first, a designer sets features of playing environments E_i and pleasures T_i in classification D_{ds} as follows. In this classification, correspondence relations of E_i and T_i are trivial, like $T_1 \models_{D_{\text{ds}}} E_2$, $T_2 \models_{D_{\text{ds}}} E_1$, $T_2 \models_{D_{\text{ds}}} E_4$, $T_3 \models_{D_{\text{ds}}} E_3$, and $T_3 \models_{D_{\text{ds}}} E_4$.

typ(D_{ds})

- E_1: environment that contains pleasure T_2.
- E_2: environment that contains pleasure T_1.
- E_3: environment that contains pleasure T_3.
- E_4: environment that contains pleasures T_2 and T_3.

tok(D_{ds})

- T_1: pleasure of resting action.
- T_2: pleasure of challenging action.
- T_3: pleasure of vertiginous action.

Classification D_{pl}: second, a designer sets parts of play equipment O_i and playing actions V_i in classification D_{pl} as

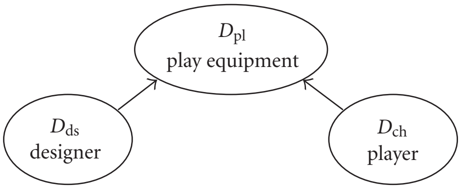


FIGURE 4: Novel design process modeled by a single channel.

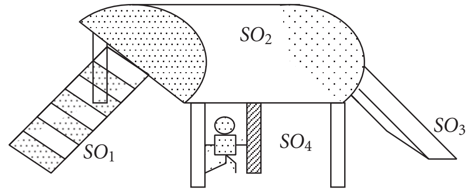


FIGURE 5: A slide derived from the design model using channel theory.

follows. To establish an *information channel* between D_{ds} and D_{ch} via D_{pl}, designers must set an informorphism D_{ds} to D_{pl}. In this case, a designer sets infomorphisms $f^{\wedge}(E_1) = O_1$, $f^{\wedge}(E_2) = O_2$, $f^{\wedge}(E_3) = O_3$, $f^{\wedge}(E_4) = O_4$, and $f^{\vee}(V_1) = T_2$, $f^{\vee}(V_2) = T_1$, $f^{\vee}(V_3) = T_3$.

typ(D_{pl})

- O_1: a ladder.
- O_2: a horizontal plate.
- O_3: a slope.
- O_4: a bar.

tok(D_{pl})

- V_1: climb vertically.
- V_2: walk.
- V_3: slide down.

Classification D_{ch}: finally, a designer creates units of playing actions $\langle O, S, V\rangle$, and sets SO_i, which are defined by combining S_i in terms of $\langle O, S, V\rangle$ with parts of play equipment O_i and V_i', which contain new playing actions emerged in the play environment in classification D_{ch} as follows. In this operation, designers must set S_i and V_i' to establish an information channel between D_{ds} and D_{ch} via D_{pl}. Figure 5 shows a slide derived from the design model using channel theory. Figure 6 shows an example of a channel represented by *Chu maps* [10].

typ(D_{ch})

- SO_1: a ladder that is put to SO_2 vertically.
- SO_2: a covered horizontal plate.
- SO_3: a steep slope that is put to SO_2.
- SO_4: a bar dropped down from SO_2.

tok(D_{ch})

- V_1': climb vertically.
- V_2': walk.
- V_3': slide down.
- V_4': walk into.
- V_5': walk outside.
- V_6': jump.
- V_7': scale up.

3.3. New Playing Actions Described in a Model of Information Channel

In the novel design, there are several alternatives for not only segmenting O_i but also S_i and V_i. Therefore, designers can foresee the emergence of new playing actions and pleasures.

At boundary between multiple parts

Designers have to select multiple elements when they select units at the boundary between multiple parts of play equipment. They become loose constraints that guide design processes. For example, setting S_2: "covered" generates new playing actions, V_4': "walk in," and V_5': "walk out":

(i) a playing action V_4' at the boundary between SO_1, SO_3, SO_4, and SO_2:

$$\begin{array}{c|cccc} & SO_1 & SO_2 & SO_3 & SO_4 \\ \hline V_4' & 1 & 0 & 1 & 1 \end{array}, \quad (6)$$

(ii) a playing action V_5' at the boundary between SO_2, SO_1, SO_3, and SO_4:

$$\begin{array}{c|cccc} & SO_1 & SO_2 & SO_3 & SO_4 \\ \hline V_5' & 0 & 1 & 0 & 0 \end{array}. \quad (7)$$

At one part

In the novel design of a play equipment, setting S_i makes arrangement O_i. Therefore, designers can set more playing actions in the same device, but selecting elements at one part of the play equipment depends on the designer's ability largely. For example, setting SO_1 and SO_4 generates new playing actions V_6': "jump." And setting SO_3 generates V_7': "scale up":

(i) a playing action V_6' in SO_1 and SO_4:

$$\begin{array}{c|cccc} & SO_1 & SO_2 & SO_3 & SO_4 \\ \hline V_6' & 1 & 0 & 0 & 1 \end{array}, \quad (8)$$

Typ(D_{ds}) / Tok(D_{ds})

	E_1	E_2	E_3	E_4
T_1	0	1	0	0
T_2	1	0	0	1
T_3	0	0	1	1

Typ(D_{ch}) / Tok(D_{ch})

	SO_1	SO_2	SO_3	SO_4
V_1'	1	0	0	1
V_2'	0	1	0	0
V_3'	0	0	1	1
V_4'	1	0	1	1
V_5'	0	1	0	0
V_6'	1	0	0	1
V_7'	0	0	1	0

Typ(D_{pl}) / Tok(D_{pl})

	E_1	E_2	E_3	E_4	O_1	O_2	O_3	O_4	SO_1	SO_2	SO_3	SO_4
V_1	1	0	0	1	1	0	0	1	1	0	0	1
V_2	0	1	0	0	0	1	0	0	0	1	0	0
V_3	0	0	1	1	0	0	1	1	0	0	1	1

Figure 6: A description of a "playing structure design" by Chu maps.

(ii) a playing action V_7' in SO_3:

	SO_1	SO_2	SO_3	SO_4
V_7'	0	0	1	0

. (9)

3.3.1. Constraint in Classification

Channel theory defines a constraint such as $\Gamma \vdash_A \Delta$. Given a classification A, a token $a \in \text{tok}(A)$ satisfies a sequent $\langle \Gamma, \Delta \rangle$ of typ(A) provided that if a is of every type in Γ, then it is of some type Δ. A constraint is a sequent $\langle \Gamma, \Delta \rangle$ for which, $\Gamma \vdash_A \Delta$. The following are part of constraints in classifications D_{ds}, D_{pl}, D_{ch}:

(1) constraint of D_{ds}

$$E_1 \vdash E_4, \qquad E_3 \vdash E_4, \tag{10}$$

(2) constraint of D_{pl}

$$O_1 \vdash O_4, \qquad O_3 \vdash O_4, \tag{11}$$

(3) constraint of D_{ch}

$$SO_1 \vdash SO_4. \tag{12}$$

Constraints $E_1 \vdash E_4$, $E_3 \vdash E_4$ in D_{ds} are directly translated in D_{pl}, but not in D_{ch}. This means that a designer's inference is described in D_{ds} and D_{pl}, because of D_{pl} being a step in which designers correspond parts of play equipment to playing actions, but which is not translated in D_{ch}. In other words, generating V_7' in tok(D_{ch}) disturbs establishment of $SO_3 \vdash SO_4$.

4. Discussion

In Channel theory, there are several ways of setting of *infomorphism* between same classifications. For example, we can set different *infomorphism* between 3 classifications D_{ds}, D_{pl}, D_{ch} shown in Figure 6. We can conclude that this case corresponds to players interpreting playing environments in different ways from designers' intentions, even though players seem to adopt the ways satisfying designers' intentions. We are now exploring the setting of classification in depth to describe this representation.

In this work, we discussed linked playing actions using the syntax structure of $\langle O, S, V \rangle$ binding them. Therefore, we are investigating a way of representing the linked playing actions by operation of classification in channel theory.

Furthermore, we discussed an environment where only a single play equipment and a single player exist. But a real playing environment can be seen as a medium for mediating plural equipment and players. Therefore, we are now investigating a way to develop our framework for multiactors.

5. Conclusions

We focused on a playing environment and analyzed the perception-action chains in order to argue playing environment on a theoretical ground. We showed that reductionism units such as therbligs are not suitable for analyzing perception-action chains. Then, we proposed representational units of playing actions $\langle O, S, V \rangle$, which have both properties of actions and environment.

There are high degrees of freedom for segmenting playing environments. $\langle O, S, V \rangle$ can be seen as one of the loose constraints for the segmentation. There are several alternatives for not only segmenting O_i but also selecting S_i and V_i. These constraints provide a guide for design. They are not excessively strict, therefore designers' ability can be utilized properly. But as long as designers follow the format $\langle O, S, V \rangle$, they are forced to pay attention to boundaries of $\langle O, S, V \rangle$, which enable designers to foresee the emergence of new playing actions.

Furthermore, $\langle O, S, V \rangle$ enables designers to interpret the playing environment as an *information channel* in terms of channel theory, which is also an appropriate loose constraint for designing a play environment. The establishment of an *information channel* between a designer and a player via play equipment supports in a mathematical way a good environment from the viewpoint of information flow.

References

[1] A. Saito and K. Ono, *Nintendo DS has Sold Reason—GEMUNIKUSU Interface Changes*, Shuwa System, Tokyo, Japan, 2007.
[2] J. J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, New York, NY, USA, 1979.
[3] L. S. Vygotsky, *Thinking and Speech*, Plenum Press, New York, NY, USA, 1987.
[4] M. Senda, *Play Structure—Design of Play Environment for Children*, Toppan Printing, New York, NY, USA, 1998.
[5] R. M. Barnes, *Motion and Time Study*, John Willey & Sons, New York, NY, USA, 1968.
[6] R. M. Barnes, *Industrial Engineering and Management: Problems and Policies*, Hive, Easton, Pa, USA, 1976.
[7] R. Barthes, *Systeme de la Mode*, Editions du Seuil, Paris, France, 1967.
[8] J. Barwise and J. Seligman, *Information Flow*, Cambridge University Press, Cambridge, UK, 1997.
[9] A. Shimojima, "What we can do with channel theory?" *Journal of Japan Society for Fuzzy Theory and Systems*, vol. 10, no. 5, pp. 13–22, 1998.
[10] V. R. Pratt, "Chu spaces from the representational viewpoint," *Annals of Pure and Applied Logic*, vol. 96, no. 1–3, pp. 319–333, 1999.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 874563, 19 pages
doi:10.1155/2008/874563

Research Article

Embodiment, Multimodality, and Composition: Convergent Themes across HCI and Education for Mixed-Reality Learning Environments

David Birchfield,[1] Harvey Thornburg,[1,2] M. Colleen Megowan-Romanowicz,[1,3] Sarah Hatton,[1,4] Brandon Mechtley,[1,5] Igor Dolgov,[1,6] and Winslow Burleson[1,5]

[1] *Arts, Media and Engineering, Arizona State University, Tempe, AZ 85281, USA*
[2] *Department of Electrical Engineering, Arizona State University, Tempe, AZ 85281, USA*
[3] *School of Educational Innovation and Teacher Preparation, Arizona State University, Tempe, AZ 85281, USA*
[4] *School of Art, Arizona State University, Tempe, AZ 85281, USA*
[5] *School of Computing and Informatics, Arizona State University, Tempe, AZ 85281, USA*
[6] *Department of Psychology, Arizona State University, Tempe, AZ 85281, USA*

Correspondence should be addressed to David Birchfield, dbirchfield@asu.edu

Received 6 October 2007; Revised 27 July 2008; Accepted 14 October 2008

Recommended by Adrian Cheok

We present concurrent theoretical work from HCI and Education that reveals a convergence of trends focused on the importance of three themes: embodiment, multimodality, and composition. We argue that there is great potential for truly transformative work that aligns HCI and Education research, and posit that there is an important opportunity to advance this effort through the full integration of the three themes into a theoretical and technological framework for learning. We present our own work in this regard, introducing the Situated Multimedia Arts Learning Lab (SMALLab). SMALLab is a mixed-reality environment where students collaborate and interact with sonic and visual media through full-body, 3D movements in an open physical space. SMALLab emphasizes human-to-human interaction within a multimodal, computational context. We present a recent case study that documents the development of a new SMALLab learning scenario, a collaborative student participation framework, a student-centered curriculum, and a three-day teaching experiment for seventy-two earth science students. Participating students demonstrated significant learning gains as a result of the treatment. We conclude that our theoretical and technological framework can be broadly applied in the realization of mixed reality, student-centered learning environments.

1. Introduction

Emerging research from Human Computer Interaction (HCI) offers exciting new possibilities for the creation of transformative approaches to learning. Current sensing, modeling, and feedback paradigms can enrich collaborative learning, bridge the physical/digital realms, and prepare all students for the dynamic world they face. When grounded in contemporary research from the learning sciences, HCI approaches have great promise to redefine the future of learning and instruction through paradigms that cultivate the students' sense of ownership and play in the learning process.

A convergence of recent trends across the Education and HCI research communities points to the promise of new learning environments that can realize this vision. In particular, many emerging technology-based learning systems are highly inquiry based, with the most effective being learner centered, knowledge centered, and assessment centered [1]. These systems are broadly termed as *student-centered learning environments* (SCLEs). Looking to the future of learning, we envision a new breed of SCLE that is rooted in contemporary Education and HCI research and is tightly coupled with appropriate curriculum and instruction design. Our research is focused on three concepts in particular: embodiment, multimodality, and composition which we define in Section 2.

We begin with a discussion of these key concepts and situate them in the context of both HCI and Education research. We present prior theoretical work and examples of the application of these three concepts in a variety of learning contexts. We then present our own work in the design and implementation of new platform for learning, the *Situated Multimedia Arts Learning Lab* (*SMALLab*). *SMALLab* (Figure 1) is a mixed-reality environment where students collaborate and interact with sonic and visual media through vocalization and full-body, 3D movements in an open, physical space. *SMALLab* emphasizes human-to-human interaction within a computational multimodal feedback framework that is situated within an open physical space. In collaboration with a network of school and community partners, we have deployed *SMALLab* in a variety of informal and formal educational settings and community-based contexts, impacting thousands of students, teachers, and community members, many from underserved populations. We have developed innovative curricula in collaboration with our partner institutions. We summarize past deployments along with their supporting pilot studies and present two recent examples as case studies of *SMALLab* learning. Finally, we present conclusions and describe our ongoing work and future plans.

Figure 1: *SMALLab* mixed-reality learning environment.

2. Prior Work

Recent research spanning Education and HCI has yielded three themes that inform our work across learning and play: *embodiment*, *multimodality*, and *composition*. Here, we define the scope of these terms in our research and discuss their theoretical basis before presenting examples of prior related applications.

2.1. Embodiment

2.1.1. Learning Sciences

By *embodiment* we mean that *SMALLab* interactions engage students both in mind and in body, encouraging them to physically explore concepts and systems by moving within and acting upon an environment.

A growing body of evidence supports the theory that cognition is "embodied"-grounded in the sensorimotor system [2–5]. This research reveals that the way we think is a function of our body, its physical and temporal location, and our interactions with the world around us. In particular, the metaphors that shape our thinking arise from the body's experiences in our world and are hence embodied [6].

A recent study of the development of reading comprehension in young children suggests that when children explicitly "index" or map words to the objects or activities that represent them, either physically or imaginatively, their comprehension improves dramatically [7]. This aligns well with the notion, advanced by Fauconnier and Turner [4], that words can be thought of as form-meaning pairs. For example, when a reader encounters the lexical form, "train" in a sentence, he can readily supply the sound form (trān). If he then maps it to the image of a train (a locomotive pulling cars situated on a track), we have a form-meaning pair that activates the student's mental model of trains, which he can then use to help him understand and interpret the sentence in which the word "train" appears [6].

SMALLab is a learning environment that supports and encourages students in this meaning-making activity by enabling them to make explicit connections between sounds, images, and movement. Abstract concepts can be represented, shared, and collaboratively experienced via physical interaction within a mixed-reality space.

2.1.2. HCI

Many emerging developments in HCI also emphasize the connections between physical activity and cognition [8–14], and the intimately embedded relationship between people and other entities and objects in the physical world [15–17]. The *embodied cognition* perspective [10, 14] argues based on strong empirical evidence from psychology and neurobiology [7, 18] that perception, cognition, and action, rather than being separate and sequential stages in human interaction with the physical world, in fact occur simultaneously and are closely intertwined. Dourish [8, 9] in particular emphasizes the importance of *context* in embodied interaction, which emerges from the interaction rather than being fixed by the system. As such, traditional HCI frameworks such as desktop computing (i.e., mouse/keyboard/screen) environments, which facilitate embodied interaction in a limited sense or not at all, risk binding the user to the system context, restricting many of his/her capacities for creative expression and free thought which have proven so essential in effective learning contexts. From cognitive, ecological, and design psychology, Shepard [17], Gibson [15], Norman [16], and Galperin [19] further emphasize the importance of the embedded relationship between people and things, and the role that manipulating physical objects has in cognition. Papert, Resnick, and Harel (see [20–23]) extend these approaches by explicitly stating their importance in educational settings. Design-based learning methodologies such as *Star Logo*, *Lego Mindstorms*, and *Scratch* [21, 24, 25] emphasize physical-digital simulation and thinking. These have proven quite popular and effective in fostering and orienting students' innate creativity toward specific learning goals.

In order for these tools to extend further into the physical world and to make use of the important connections provided by *embodiment*, they must include physical elements that afford embodied interactions. Ishii has championed the field of tangible media [26] and coined the term tangible user interfaces (TUIs versus GUI: graphical user interfaces). His Tangible Media group has developed an extensive array of applications that pertain to enhancing not only productivity (e.g., Urban Simulation, SandScape) but also artistic expression and playful engagement in the context of learning (e.g., I/O Brush, Topobo, and Curlybot) [27].

Some prior examples of HCI systems that facilitate elements of embodiment and interaction with immersive environments include the Cave Automated Visualization Environment (CAVE) [28]. CAVEs typically present an immersive environment through the use of 3D glasses or some other head-mounted display (HMD) that enables a user to engage through a remote control joystick. A related environment, described as a step toward the holodeck, was developed by Johnson at USC to teach topics ranging from submarine operation to Arabic language training [29]. In terms of extending physical activity through nontraditional interfaces and applying them to collaboration and social engagement, the Nintendo Wii's recent impact on entertainment is the most pronounced. The Wii amply demonstrates the power of the body as a computing interface. Some learning environments that have made strides in this area include *Musical Play Pen*, *KidsRoom,* and *RoBallet* [30–32]. These interfaces demonstrate that movement-based HCI can greatly impact instructional design, play, and creativity.

2.1.3. Example

A particularly successful example of a learning environment that leverages embodiment in the context of instructional design is *River City* [33–36]. *River City* is a multiuser, online desktop virtual environment that enables middle school children to learn about disease transmission. The virtual world in *River City* embeds a river in various types of terrain which influence water runoff and other environmental factors that in turn influence the transmission of disease through water, air, and/or insect populations. The factors affecting disease transmission are complex and have many causes, paralleling conditions in the physical world. Student participants are virtually embodied in the world, enabling exploration through avatars that interact with each other, with facilitators' avatars, and with the auditory and visual stimuli comprising the *River City* world. Participants can make complex decisions within this world by, for example, using virtual microscopes to examine water samples, and sharing and discussing their proposed solutions. In several pilot studies [33, 34], the level of motivation, the diversity and originality of participants' solutions, and their overall content knowledge were found to increase with *River City* as opposed to a similar paper-based environment. Hence, the *River City* experience provides at least one successful example of how social embodiment through avatars in a multisensory world can result in learning gains.

However, a critical aspect of embodiment not addressed by *River City* is the bodily-kinesthetic sense of the participant. Physically, participants interact with *River City* using a mouse and keyboard, and view 2D projections of the 3D world on a screen. The screen physically separates users' bodies from the environment, which implies that perception and bodily action are not as intimately connected as they are in the physical world, resulting in embodiment in a lesser sense [10]. In *SMALLab*, multiple participants interact with the system and with each other via expressive, full-body movement. In *SMALLab* there is no physical barrier between the participant and the audiovisual environment they manipulate. It has long been hypothesized [37] that bodily kinesthetic modes of representation and expression are an important dimension of learning and severely underutilized in traditional education. Thus, it is plausible that an environment that affords full-body interactions in the physical world can result in even greater learning gains.

2.2. Multimodality

2.2.1. Learning Sciences

By *multimodality* we mean interactions and knowledge representations that encompass students' full sensory and expressive capabilities including visual, sonic, haptic, and kinesthetic/proprioceptive. Multimodality includes both student activities in *SMALLab* and the knowledge representations it enables.

The research of Jackendoff in cognitive linguistics suggests that information that an individual assimilates is encoded either as spatial representations (images) or as conceptual structures (symbols, words or equations) [38]. Traditional didactic approaches to teaching strongly favor the transmission of conceptual structures, and there is evidence that many students struggle with the process of translating these into spatial representations [6]. By contrast, information gleaned from the *SMALLab* environment is both propositional and imagistic as described above.

Working in *SMALLab*, students create multimodal artifacts such as sound recordings, videos, and digital images. They interact with computation using innovative multimodal interfaces such as 3D physical movements, visual programming interfaces, and audio capture technologies. These interfaces encourage the use of multiple modes of representation, which facilitates learning in general, [39, 40] and are robust to individual differences in students' optimal learning styles [37, 41], and can serve to motivate learning [1].

2.2.2. HCI

Many recent developments in HCI have emphasized the role of immersive, multisensory interaction through multimodal (auditory, visual, and tactile) interface design. This work can be applied in the design of new mixed-reality spaces. For example, in *combining audio and video in perceptive spaces*, Wren et al. [42] describe their work in the development of environments utilizing unencumbered sensing technologies in situated environments. The authors present a variety of applications of this technology that span data visualization, interactive performance, and gaming. These technologies suggest powerful opportunities for the design of learning scenarios, but they have not yet been applied for this purpose.

Related work in arts and technology has influenced our approach to the design of mediated learning scenarios. Our work draws from extensive research in the creation of interactive sound environments [43–45]. While much of this work is focused on applications in interactive computer music performance, the core innovations for interactive sound can be directly applied in our work with students. In addition, we are drawing from the 3D visualization community [46] in considering how to best apply visual design elements (e.g., color, lighting, spatial composition) to render content in *SMALLab*.

There are many examples where HCI researchers are extending the multimodal tool set and applying it to novel technologically mediated experiences for learning and play. Ishii's *Music Bottles* offer a multimodal experience through sound, physical interaction, and light color changes as different bottles are uncorked by the user to release sounds. The underlying sensor mechanism is a resonant RF coil that is modulated by an element in the cork. Edmonds has chronicled the significant contribution physiological sensors have made to the interactive computational media arts [47]. *RoBallet* uses laser beam-break sensors, such as those found in some elevators and garage doors, along with video and sonic feedback to engage students in interactive choreography and composition. Cavallo argues that this system would enable new forms of teaching not only music but math and programming as well [32]. The work described in this paper builds upon this prior work and is similarly extending the tools and domains for multimodal HCI interfaces as they apply to learning and play.

2.2.3. Example—the MEDIATE Environment

One example of an immersive, multisensory learning environment which emphasizes multimodality is *MEDIATE*, an environment designed to foster a sense of agency and a capacity for creative expression in people on the autistic spectrum (PAS). Autism is a variable neuron-developmental disorder in which PAS are overwhelmed by the excessive stimuli, the noises and colors that characterize interaction in the physical world [48–50]. Perhaps as a result (although exact mechanisms and causes are unknown), PAS withdraw into their own world. They often find self-expression and even everyday social interaction difficult. *MEDIATE,* designed in collaboration with PAS, sets up an immersive 3D environment in which stimuli are quite focused and simplified, yet at the same time dynamic and engaging—capable of affording a wide range of creative expression. The *MEDIATE* infrastructure consists of a pair of planar screens alternating with a pair of tactile interface walls and completely surrounds the participant. On the screens are projected particle grids, a dynamic visual field which responds to the participant's visual silhouette, his/her vocalizations and other sounds, and his/her tactile interactions [49]. A specially designed loudspeaker system provides immersive audio feedback that includes the subsonic range, and interface walls provide vibrotactile feedback.

Multimodality in *MEDIATE* is achieved through the integration of sonic, visual, and tactile interfaces in both sensing and feedback. The environment is particularly impressive in that it can potentially supplant the traditional classroom space with one that is much more conducive to learning in the context of PAS. However, *MEDIATE* remains specialized as a platform for PAS rehabilitation and has not been generalized for use in everyday classroom instruction. By contrast, *SMALLab* emphasizes multimodality in the context of real-world classroom settings, where the immersive media coexists in the realm of everyday social interactions. *SMALLab* enables students and teachers to work together, physically interacting, face-to-face with one another and digital media elements. Thus, it facilitates the emergence of a natural zone of proximal development [51] where, on an informal basis, facilitators and student peer experts can interact with novices and increase what they are able to accomplish in the interaction.

Although *MEDIATE* was designed in collaboration with PAS [48], participants are not able to build in new modes of interaction or further customize the interface. This idea of *composition*, which comes from building, extending, and reconfiguring the interaction framework, is essential to engaging participants in more complex and targeted learning situations and has been integral to the design of *SMALLab*.

2.3. Composition

2.3.1. Learning Sciences

Composition refers to reconfigurability, extensibility, and programmability of interaction tools and experiences. Specifically, we mean *composition* in two senses. First, students compose new interaction scenarios in service of learning. Second, educators and mentors can extend the toolset to support new types of learning that is tailored to their students' needs.

In our design of the *SMALLab* learning experience, we proceed from the fundamentally constructivist view that knowledge must be actively constructed by the learner rather than passively received from the environment, and that the prior knowledge of the learner shapes the process of new construction [52]. Drawing on the views of social constructivists (i.e., Vygotsky, Bruner, Bandura, Cobb, Roth) we view the learning process as socially mediated, knowledge as socially and culturally constructed, and reality as not discovered but rather socially "invented" [40, 53, 54]. We venture beyond constructivism in subscribing to the notion that teaching and learning should be centered on the construction, validation, and application of models—flexible, reusable knowledge structures that scaffold thinking and reasoning [3, 55]. This constructive activity of modeling lies at the heart of student activity in *SMALLab*.

In their seminal work describing the situated nature of cognition, Brown et al. [56] observed that students in a classroom setting tend to acquire and use information in ways that are quite different from "just plain folks" (JPFs). They further revealed that the reasoning of experts and JPFs was far closer to one another than that of experts and students. They concluded that the culture of schooling, with its passive role for students and rule-based structure for social interactions, promotes decontextualization of information that leads to narrow procedural thinking and the inability to transfer lessons learned in one context to another. This finding highlights the importance of learning that is situated, both culturally and socially. *SMALLab* grounds students in a physical space that affords visual, haptic, and sonic feedback. The abstraction of conceptual information from this perceptual set is enabled through guided reflective practice of students as they engage in the modeling process.

Student engagement in *SMALLab* experience is motivated both by the novelty of a learning environment that affords them some measure of control [57] and by the opportunity to work collaboratively to achieve a specific goal, where the pathway they take to this goal *is not predetermined by the teacher or the curriculum*. Hence, *SMALLab* environment rewards originality and creativity with a unique digital-physical learning experience that affords new ways of exploring a problem space.

2.3.2. HCI

Compositional interfaces have a rich history in HCI, as evidenced by Papert and Minsky's Turtle Logo which fosters creative exploration and play in the context of a functional, lisp-based programming environment [24]. More recent examples of HCI systems that incorporate compositional interfaces include novice level programming tools such as *Star Logo*, *Scratch*, and *Lego Mindstorms*. Resnick extends these approaches through the *Playful Invention and Exploration* (PIE) Museum Network and the *Intel Computer Club Houses* [58], thus providing communities with tools for creative composition in rich, informal sociocultural contexts. Essentially, these interfaces create a "programming culture" at community technology centers, classrooms, and museums. There has been extensive research on the development of programming languages for creative practitioners, including graphical programming environments for musicians and multimedia artists such as Max/MSP/Jitter, Reaktor, and PD. This research has made significant contributions toward improving the impact and viability of programming tools as compositional interfaces.

Embedding physical interactions into objects for composition is a strategy for advancing *embodied multimodal composition*. Ryokai's I/O Brush [59] is an example of a technology that encourages composition, learning, and play. This system enables capture from the physical world through a camera in the end of a paint brush that allows individuals to capture colors and textures from the physical world and compose with them in the digital world. It can even take video sequences such as a blinking eye that can then become part of the user's digital painting. Composition is a profoundly empowering experience and one that many learning environments are also beginning to emphasize to a greater extent.

2.3.3. Example—Scratch

The Scratch programming environment [60] emphasizes the power of compositional paradigms for learning. Scratch enables students to create games, interactive stories, animations, music and art within a graphical programming environment. The interface extends the metaphor of LEGO bricks where programming functions snap together in a manner that prohibits programming errors and thus avoids the steep learning curve that can be a barrier to many students in traditional programming environments. The authors frame the goal of Scratch as providing "tinkerability" for learners that will allow them to experiment and redesign their creations in a manner that is analogous to physical elements, albeit with greater combinatorial sophistication.

Scratch has been deployed in a number of educational settings [25, 61]. In addition to focused research efforts to evaluate its impact, a growing Scratch community website, where authors can publish their work, provides mounting evidence that it is a powerful tool for fostering meaningful participation for a broad and diverse population.

Scratch incorporates multimodality through the integration of sound player modules within the primarily visual environment. However, it provides only a limited set of available tools for sound transformation (e.g., soundfile playback, speech synthesis) and as a consequence, authors are not able to achieve the multimodal sophistication that is possible within *SMALLab*. Similarly, Scratch addresses the theme of embodiment in the sense that authors and users can represent themselves as avatars within the digital realm. However, Scratch exists within the standard desktop computing paradigm and students cannot interact through other more physically embodied mechanisms.

2.4. Defining Play

With a focus on play in the context of games, Salen and Zimmerman [62] summarize a multitude of definitions. First they consider the diverse meanings and contexts of the very term "play." They further articulate multiple scopes for the term, proposing a hierarchy comprised of three broad types. The most open sense is "being playful," such as teasing or wordplay. Next is "ludic activity," such as playing with a ball, but without the formal structure of a game. The most focused type is "game play," where players adhere to rigid rules that define a particular game space.

Play and game play in particular have been shown to be an important motivational tool [63], and as Salen and Zimmerman note, play can be transformative as, "it can overflow and overwhelm the more rigid structure in which it is taking place, generating emergent, unpredictable results." Our work is informed by these broad conceptions of play that are applied to the implementation of game-like learning scenarios for K-12 content learning [62].

Jenkins offers an expansive definition of play as "the capacity to experiment with one's surroundings as a form of problem-solving" [64]. Students engaged in this type of play exhibit the same transformative effects as described by Salen and Zimmerman. We apply this definition of play as collaborative problem solving in our work with students in formal learning contexts.

2.5. Toward a Theoretical and Technological Integration

As described above, there has been extensive theoretical and practice-based research across Education and HCI that is aimed at improving learning through the use of embodiment, multimodality, and compositional frameworks. We have described examples of prior projects, each of which strongly emphasizes one or two of these concepts. This prior work has yielded significant results that demonstrate the powerful impact of educational research that is aligned with emerging HCI practices. However, while there are some prior examples of interactive platforms that integrate these principles [65], there are few prior efforts to-date that do so while leveraging the powerful affordances of mixed reality for content learning. As such there is an important opportunity to improve upon prior work.

In addition, many technologically driven efforts are limited by the use of leading edge technologies that are prohibitively expensive and/or too fragile for most real-world learning situations. As a consequence, many promising initiatives do not make a broad impact on students and cannot be properly evaluated owing to a failure to address the practical constraints of today's classrooms and informal learning contexts. Specifically, in order to see large-scale deployment on a two- to five-year horizon, learning environments must be inexpensive, mindful of typical site constraints (e.g., space, connectivity, infrastructure support), robust, and easily maintainable. It is essential to reach a balance between reliance upon leading-edge technologies and consideration of the real-world context in order to collect longitudinal data over a broad population of learners that will demonstrate the efficacy of these approaches.

Our own efforts are focused on advancing research at the intersection of HCI and Education. We next describe a new mixed-reality environment for learning, a series of formative pilot studies, and two recent in-school programs that illustrate the implementation and demonstrate the impact of our work.

3. *SMALLab:* Integrated HCI for Learning

SMALLab represents a new breed of student-centered learning environment (SCLE) that incorporates multimodal sensing, modeling, and feedback while addressing the constraints of real-world classrooms. Figure 2 diagrams the full system architecture, and here we detail select hardware and software components.

Physically, *SMALLab* consists of a $15' \times 15' \times 12'$ portable, freestanding media environment [66]. A cube-shaped trussing structure frames an open physical architecture and supports the following sensing and feedback equipment: a six-element array of *Point Grey Dragonfly* firewire cameras (three color, three infrared) for vision-based tracking, a top-mounted video projector providing real time visual feedback, four audio speakers for surround sound feedback, and an array of tracked physical objects (*glowballs*). A networked computing cluster with custom software drives the interactive system.

The open physical architecture of the space is designed to encourage human-to-human interaction, collaboration, and active learning within a computational framework. It can be housed in a large general-purpose classroom without the need for additional specialized equipment or installation procedures. The use of simple, unencumbered sensing technologies ensures that there is a minimal learning curve for interaction, yet it has been utilized in diverse educational contexts including schools and museums.

With the exception of the *glowballs*, all *SMALLab* hardware (e.g., audio speakers, cameras, multimedia computers, video projector, support structure) is readily available off-the-shelf. This ensures that *SMALLab* can be easily maintained throughout the life of a given installation as all components can be easily replaced. Furthermore, the use of readily available hardware contributes to the overall low cost of the system. We have custom developed all *SMALLab* software which is made freely available to our partner educational institutions.

SMALLab can be readily transported and installed in classrooms or community centers. We have previously disassembled, transported to a new site, reinstalled, and calibrated a functioning *SMALLab* system within one day's time.

3.1. Multimodal Sensing

Groups of students and educators interact in *SMALLab* together through the manipulation of up to five illuminated *glowball* objects and a set of standard HID devices including wireless gamepads, Wii Remotes [67, 68], and commercial wireless pointer/clicker devices. The vision-based tracking system senses the real-time 3D position of these *glowballs* at a rate of 50–60 frames per second using robust multiview techniques [69]. To address interference from visual projection, each object is partially coated with a tape that reflects infrared light. Reflections from this tape can be picked up by the infrared cameras, while the visual projection cannot. Object position data is routed to custom software modules (described below) that perform various real-time pattern analyses on this data, and in response, generate real-time interactive sound and visual transformations in the space. With this simple framework we have developed an extensible suite of interactive learning scenarios and curricula that integrate the arts, sciences, and engineering education.

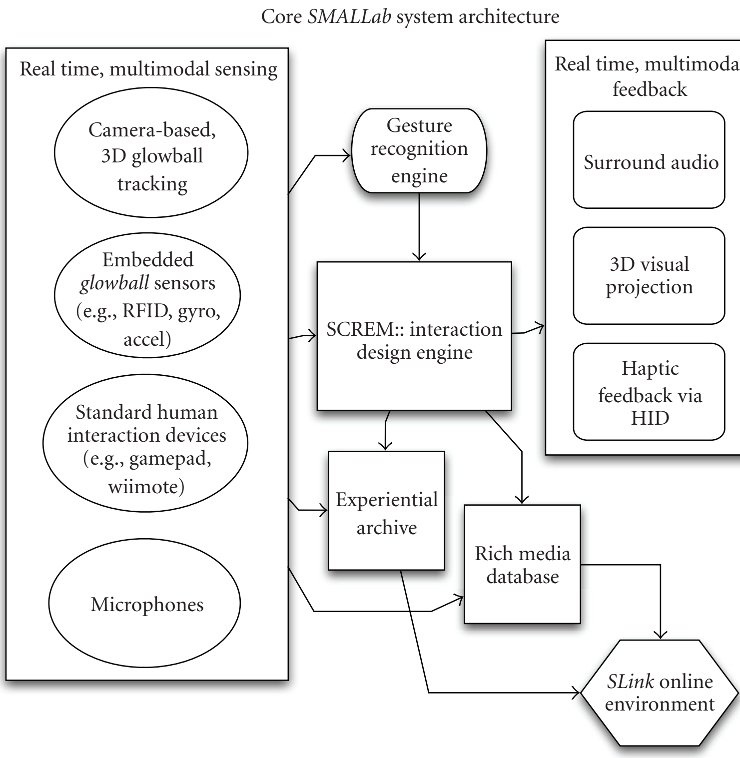


Figure 2: *SMALLab* software architecture.

3.2. Rich Media Database

SMALLab features an integrated and extensible rich media database that maintains multimodal content provided by students, teachers, and researchers. This is an important tool in support of multimodal knowledge representation in *SMALLab*. It manages audio, video, images, text, and 3D objects and enables users to annotate all media content with user-specific metadata and typed links between elements. The SCREM interface (described below) tightly integrates search and navigation tools so that scenario authors and students can readily access this media content.

3.3. SCREM

We apply the notion of composition at two levels. First, we have conceived of *SMALLab* as a modular framework to ensure that educators and administrators can continuously extend and improve it through the design and implementation of new scenarios. In this regard, *SMALLab* is not a one-size-fits-all solution, but rather, it enables an educator- and community-driven learning environment. Second, many *SMALLab* curricula emphasize learning through collaborative problem solving and open-ended design challenges. These approaches demand that students are able to readily design and deploy new interactive scenarios through the manipulation of powerful, yet easy to use interfaces—interfaces that provide both depth and breadth.

To this end we have developed an integrated authoring environment, the *SMALLab core for realizing experiential media* (*SCREM*). SCREM is a high-level object oriented framework that is at the center of interaction design and multimodal feedback in *SMALLab*. It provides a suite of graphical user interfaces to either create new learning scenarios or modify existing frameworks. It provides integrated tools for adding, annotating, and linking content in the *SMALLab* Media Content database. It facilitates rapid prototyping of learning scenarios, enables multiple entry points for the creation of scenarios, and provides age and ability appropriate authoring tools.

SCREM supports student and teacher composition at three levels. First, users can easily load and unload existing learning scenarios. These learning scenarios are stored in an XML format that specifies interactive mappings, visual and sonic rendering attributes, typed media objects, and metadata including the scenario name and date. Second, users can configure new scenarios through the reuse of software elements that are instantiated, destroyed, and modified via a graphical user interface. Third, developers can write new software code modules through a plug-in type architecture that are then made available through the high-level mechanisms described above. Depending on developer needs, low-level *SMALLab* code can be written in a number of languages and media frameworks including Max/MSP/Jitter, Javascript, Java, C++, Objective C, Open Scene Graph, and VR-Juggler.

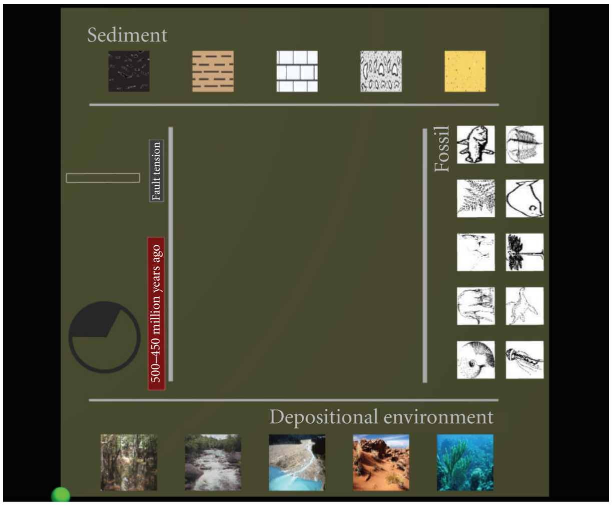


Figure 3: Screen capture of projected *layer-cake builder* scene.

3.4. *SLink* Web Portal

The *SMALLab Link* or, *SLink*, web portal [66] provides an online interface that enables teaching and learning to seamlessly span multiple *SMALLab* installations and to extend from the physical learning environment and into students' digital realms. It serves as three functions: (1) a supportive technology, (2) a research tool, and (3) an interface to augment *SMALLab* learning.

As a supportive technology, *SLink* acts as a central server for all *SMALLab* media content and user data. It provides functionality to sync media content that is created at a given *SMALLab* site to all other sites while preserving unique metadata. Similarly, *SLink* maintains dynamic student and educator profiles that can be accessed by teachers and researchers online or in *SMALLab*.

SLink is a research tool and an important component of the learning evaluation infrastructure. Through a browser-based interface, educational researchers can submit, search, view, and annotate video documentation of *SMALLab* learning. Multiple annotations and annotator metadata are maintained for each documentation element.

SLink serves as a tool for students where they can access or contribute media content from any location through the web interface. These media content and metadata will sync to all *SMALLab* installations. In ongoing work, we are expanding the *SLink* web interface to provide greater functionality for students. Specifically, we are developing tools to search and render 3D *SMALLab* movement data through a browser-based application. Student audio interactions can be published as podcasts, and present visual interactions presented as streaming movies. In these ways, *SLink* extends into the web our paradigms of multimodal interaction and learning through composition.

3.5. Experiential Activity Archive

All *glowball* position data, path shape quality information, SCREM interface actions, and projected media data are streamed in real time to a central archive application. Incoming data is timestamped and inserted into a MySQL database where it is made available in three ways. First, archived data can be replayed in real time such that it can be rerendered in *SMALLab* for the purpose of supporting reflection and discussion among students regarding their interactions. Second, archived data is made available to learners and researchers through the *SLink* web interface. Third, archived data can be later mined for the purposes of evaluation and assessment of *SMALLab* learning. We are currently developing a greatly expanded version of the activity archive that will include the archival of real-time video and audio streams, interfaces to create semantic links among entries, and tools to access the data from multiple perspectives.

4. Case Study: Earth Science Learning in *SMALLab*

Having presented a theoretical basis and described the development and integration of various HCI technologies into a new mixed-reality environment, we now focus on the application and evaluation of *SMALLab* for learning. This research is undertaken at multiple levels including focused user studies to validate subcomponents of the system [70, 71], and perception/action experiments to better understand the nature of embodied interaction in mixed-reality systems such as *SMALLab* [72]. Over the past several years we have reached over 25,000 students and educators through research and outreach in both formal and informal contexts that span the arts, humanities, and sciences [73–75]. This prior work serves as an empirical base that informs our theoretical framework. Here we present a recent case study to illustrate our methodology and results.

4.1. Research Context

In Summer 2007 we began a long-term partnership with a large urban high school in the greater Phoenix, AZ metropolitan area. We have permanently installed *SMALLab* in a classroom and are working closely with teachers and students across the campus to design and deploy new learning scenarios. This site is typical of public schools in our region. The student demographic is 50% white, 38% Hispanic, 6% Native American, 4% African American, and 2% other. 50% of students are on free or reduced lunch programs, indicating that many students are of low socioeconomic status. 11% are English language learners and 89% of these students speak Spanish at home. In this study, we are working with 9th grade students and teachers from the school's dedicated program for at-risk students. The program is a specialized "school within a school" with a dedicated faculty and administration. Students are identified for the program because they are reading at least two levels below their grade and have been recommended by their middle school teachers and counselors. After almost a year of classroom observation by our research team, it is evident that students are tracked into this type of program, not because they have low abilities, but because they are often underserved by traditional instructional approaches and exhibit low motivation for learning. Our work seeks to address the needs of this population of students and teachers.

Throughout the year, we collaborated with a cohort of high school teachers to design new *SMALLab* learning scenarios and curricula for language arts and science content learning. Embodiment, multimodality, and composition served as pillars to frame the formulation of new *SMALLab* learning scenarios, associated curricula, and the instructional design. In this context, we present one such teaching experiment. This case study illustrates the use of *SMALLab* for teaching and learning in a conventional K-12 classroom. It demonstrates the implementation of our theoretical framework around the integration of embodiment, multimodality, and composition in a single learning experience. Finally, we present empirical evidence of student learning gains as a result of the intervention.

4.2. Design and Teaching Experiment

The evolution of the earth's surface is a complex geologic process that is impacted by numerous interdependent processes. Geologic evolution is an important area of study for high school students because it provides a context for the exploration of systems thinking [76] that touches upon a wide array of earth science topics. Despite the nature of this complex, dynamic process, geologic evolution is typically studied in a very static manner in the classroom. In a typical learning activity, students are provided with an image of the cross-section of the earth's crust. Due to the layered structure of the rock formations, this is sometimes termed a geologic layer cake. Students are asked to annotate the image by labeling the rock layer names, ordering the layers according to which were deposited first, and identifying evidence of uplift and erosion [77]. Our partner teacher has numerous years of experiences with conventional teaching approaches in his classroom. Through preliminary design discussions with him, we identified a deficiency of this traditional instructional approach: *when students do not actively engage geologic evolution as a time-based, generative process, they often fail to conceptualize the artifacts (i.e., cross-sections of the earth's surface) as the products of a complex, dynamic process*. As a consequence, they struggle to develop robust conceptual models during the learning process.

For six weeks we collaborated with the classroom teacher, using the *SMALLab* authoring tools, to realize a new mixed-reality learning scenario to aid learning about geologic evolution in a new way. Our three-part theoretical framework guided this work: embodiment, multimodality, and composition. At the end of this process, the teacher led a three-day teaching experiment with seventy-two of his ninth-grade earth science students from the CORE program. The goals for the teaching experiment were twofold. First, we wanted to advance participating students' understanding of earth science concepts relating to geologic evolution. Second, we wanted to evaluate our theoretical framework and validate *SMALLab* as a platform for mixed-reality learning in a formal classroom-learning environment.

We identified four content learning goals for students: (1) understanding of the principle of *superposition*—that older structures typically exist below younger structures on the surface of the Earth; (2) understanding geologic evolution as a complex process with interdependent relationships between surface conditions, fault events, and erosion forces; (3) understanding that geologic evolution is a time-based process that unfolds over multiple scales; (4) understanding how the fossil record can provide clues regarding the age of geologic structures. These topics are central to high school earth science learning and are components of the state of Arizona earth and space science standards [78]. We further stipulate that from the theoretical perspective of modeling instruction [79, 80] students should be able to apply a conceptual model of geologic evolution that integrates both descriptive and explanatory elements of these principles.

Our collaborative design process yielded three parts: (1) a new mixed-reality learning scenario, (2) a student participation framework, and (3) an associated curriculum. We now describe each of these parts, discussing how each tenet of our theoretical framework is expressed. We follow this with a discussion of the outcomes with respect to our goals.

4.2.1. Interactive Scenario: Layer-Cake Builders

Figure 3 shows the visual scene that is projected onto the floor of *SMALLab*. Within the scene, the center portion is the layer-cake construction area where students deposit sediment layers and fossils. Along the edges, students see three sets of images. At the bottom they see depictions of depositional environments. At the top are images that represent sedimentary layers. To the right they see an array of plant and animal images that represent the fossil record. Each image is an interactive element that can be selected by students and inserted into the layer-cake structure. The images are iconic forms that students encounter in their studies outside of *SMALLab*. A standard wireless game pad controller is used to select the current depositional environments from the five options. When one student makes a selection, other students will *see* the image of the environment and *hear* a corresponding ambient sound file. One *SMALLab* glowball is used to grab a sediment layer—by hovering above it—from five options and drop it onto the layer-cake structure in the center of the space. This action will insert the layer into the layer-cake structure at the level that corresponds with the current time period. A second glowball is used to grab a fossil from ten options and drop it onto the structure. This action embeds the fossil in the current sediment layer. On the east side of the display, students see an interactive clock with geologic time advancing to increment each new period. Three buttons on a wireless pointer device are used to pause, play, and reset geologic time. A bar graph displays the current fault tension value in real time. Students use a Wii remote game controller [67, 68], with embedded accelerometers, to generate fault events. The more vigorously that a user shakes the device, the more the fault tension will increase. Holding the device still will decrease the fault tension. When a tension threshold is exceeded, a fault event (i.e., earthquake) will occur, resulting in uplift in the layer-cake structure. Fault events can be generated at any time during the building process. Subsequently erosion occurs on the uplifted portion of the structure.

Figure 4 illustrates that in addition to the visual feedback present in the scene, students hear sound feedback with each action they take. A variety of organic sound events including short clicks and ticks accompany the selection and deposit of sediment layers and fossils. These events were created from field recordings of natural materials such as stones. This feedback contributes to an overarching soundscape that is designed to enrich students' sense of immersion in the earth science model. In addition, key earth science concepts and compositional actions are communicated to the larger group through sound. For example, the selection of a depositional environment is represented visually through an image, and sonically through looping playback of a corresponding sound file. If a student selects the depositional environment of a fast moving stream, all students will see an image of the stream, and hear the sound of fast moving water. The multimodal display first alerts all students to be aware of important events in the compositional process. In addition, the dynamic nature of the fast moving water sound communicates important features of the environment itself that are not necessarily conveyed through image alone. Specifically, a fast moving stream is associated with the deposition of a conglomerate sediment layer that contains a mixture of large and small particles. The power of water to move large rocks and even boulders is conveyed to students through sound.

While students are engaged in the compositional process, sound is an important component of how they parse the activity and cue their own actions. Here we present a transcript from a typical layer-cake build episode, demonstrating how sound helps students to orient themselves in the process. In the transcription *T* is the teacher and *FS* indicates a member of the fossil selection team:

(The student holding the controller from the depositional environment group selects an environment and the sound of ocean waves can be heard. Responding to the sound cue without even looking up at the image of the depositional environment highlighted, the student controlling the glowball for sediment layer team moves to select limestone.)

FS1: Shallow ocean.

FS2: Wait, wait, wait.

(As the student holding the fossil glowball moves to make his selection. A fossil team member tells the boy with the glowball to wait because he could not see what sediment layer had been selected. After the sediment group and the fossil group made their selections, someone from the depositional environment team changes their selection. When the sound of a new environment is heard, the fossil team selector student (FS1) looks at the new environment and sees that the fossil he deposited is no longer appropriate for this environment. He picks up an image of a swimming reptile but then pauses uncertainly before depositing it.)

FS2: Just change it.

T: Just change it to the swimming reptile.

(The clock chimes the completion of one cycle at this point. The depositional environment team shifts their choice to desert and a whistling wind sound can be heard. Again, without even looking at the depositional environment image, the fossil group selector, FS2, quickly grabs a fossil and deposits it while the sediment layer girl runs back and forth above her 5 choices trying to decide which one to choose. She finally settles on one, picks and deposits it and then hands off the glowball and sits down. The next two selector students stand at the edge of the mat waiting for the clock to complete another cycle. The assessment team is diligently taking notes on what has been deposited. Another cycle proceeds as the sound of ocean waves can be heard. Students controlling the glowballs move quickly to make their selections without referring to the highlighted depositional environment.)

As shown in Figure 5, during the learning activities, all students are copresent in the space, and the scenario takes advantage of the embodied nature of *SMALLab*. For example, the concept of fault tension is embodied in the physical act of vigorously shaking the Wii Remote game controller. In addition this gesture clearly communicates the user's intent to the entire group. Similarly, the deliberate gesture of physically stooping to select a fossil and carrying it across the space before depositing it in the layer-cake structure allows all students to observe, consider and act upon this decision as it is unfolding. Students might intervene verbally to challenge or encourage such a decision. Or they might coach a student who is struggling to take action. Having described the components of the system, we now narrate and discuss the framework that enables a class of over twenty students to participate in the scenario.

4.2.2. Participation Framework

The process of constructing a layer cake involves four lead roles for students: (1) the depositional environment selector, (2) the sediment layer selector, (3) the fossil record selector, and (4) the fault event generator. In Figure 4, we diagram the relationship between each of these participant roles (top layer) and the physical interaction device (next layer down). The teacher typically assumes the role of geologic time controller.

In the classroom, approximately twenty to twenty-five students are divided into four teams of five or six students each. Three teams are in active rotation during the build process, such that they take turns serving as the action lead with each cycle of the geologic clock. These teams are the (1) depositional environment team and fault event team, (2) the sediment layer team, and (3) the fossil team. The remaining students constitute the evaluation team. These "evaluator" students are tasked to monitor the build process, record the

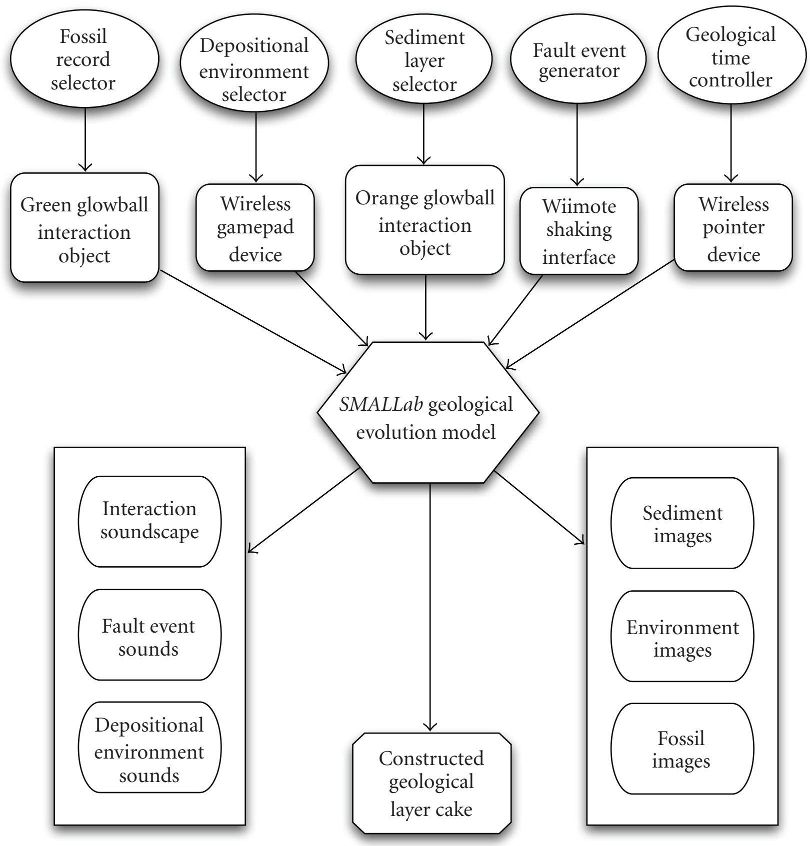


Figure 4: Layer-cake builder interaction architecture schematic.

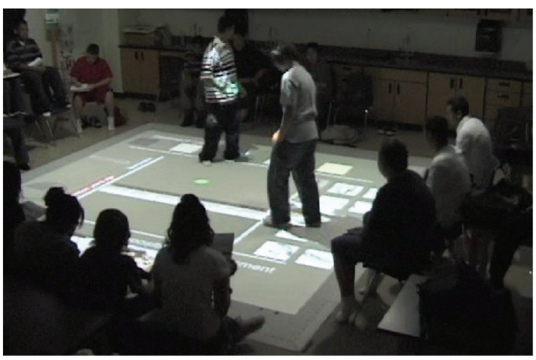


Figure 5: Students collaborating to compose a layer-cake structure in *SMALLab*.

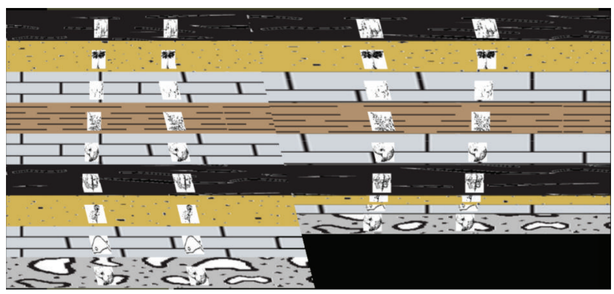


Figure 6: Layer-cake structure created in *SMALLab*.

activities of action leads, and to steer the discussion during the reflection process. Students are encouraged to verbally coach their teammates during the process.

There are at least two ways in which the build process can be structured. On the one hand, the process can be purely open ended, with the depositional environment student leading the process, experimenting with the outcomes, but without a specific constraint. This is an exploratory compositional process. Alternatively, the students can reference an existing layer-cake structure as a script such as the one pictured in Figure 6. This second scenario is a goal directed framing where only two students have access to the original script, but all participants must work together to reconstruct the original. At the end of the build cycle, students compare their structure against the original. In this discussion we narrate the goal-directed build process.

At the beginning of each geologic period, the lead "depositional environment" student examines the attributes of the source structure (e.g., Figure 6) and selects the appropriate depositional environment or surface condition on the earth. All students see an image and hear a sonic representation of the depositional environment. Based on that selected condition, another student grabs the appropriate sedimentary rock, and drops it onto the structure. While considering the current evolutionary time period and the current depositional environment, another student grabs a fossilized animal and lays it into the sedimentary layer. To address any potential student misconceptions, the teacher initially leads a discussion to clarify that fossilization is yet another example of a geologic process that students should be aware of, despite the fact that it is not a focus of this particular activity. If a student changes their mind, sediment and fossil layers can be replaced by another element within a given geologic time period. As the geologic clock finishes a cycle, the next period begins. The action lead passes their interaction device to the next teammate, and these students collaborate to construct the next layer. The rotation continues in like fashion until the layer cake is complete. In this manner, the layer-cake build process unfolds as a semistructured choreography of thought and action that is distributed across the four action leads and their teammates. The teams rotate their roles each time a new layer cake is to be constructed. The fossil students become evaluators, while the evaluators become the sediment layer team and so forth.

From a compositional perspective, this process is *open ended* and *improvisational*. By open ended we mean that any combination of depositional environments, sediment layers, fossils, and fault events can occur without constraint from the technology itself. By improvisational we mean that it unfolds in real time, and each participant acts with a clearly defined role, yet independently of the other students. The participation framework is analogous to a group of improvising jazz musicians. Students have individual agency to think and act freely. Yet they are bound by a constrained environment and driven by the shared goal of producing a highly structured outcome. Composition is distributed across multiple students where each has a clearly defined role to play and a distinct contribution to be made toward the collective goal. Collective success or failure depends on all participants. This process unfolds in real time with the expectation that there will be continuous face-to-face communication between participants.

This interaction model affords rich opportunities for whole group action and discussion about the relationship between in-the-moment events and the consequence of these decisions in the final outcome. For example, "fault event" students are free to generate earthquake after earthquake and explore the outcomes of this activity pattern including its impact on students who are depositing sediment layers and fossils. Through this experimentation, students come to understand that in the real world, just as in the model, periods of numerous fault events are often interspersed with periods of little activity. This is a system-level understanding of geologic evolution that must be negotiated by teams of students over the course of numerous cycles.

The learning activity is a form of structured play in two senses. Following Salen and Zimmerman's model, the layer-cake build process unfolds in a structured manner as defined by the interaction framework. However, the play activity can take different forms according to the metarules set by the teacher. For example, during the open-ended compositional process, play is akin to "ludic activity" where a clear game space is articulated in *SMALLab*, but there are not clearly defined start and end conditions. When the activity is structured with a reference layer-cake image and students are given the explicit goal to recreate that structure, the activity takes the form of goal oriented "game play." Jenkins' notion of play also frames the learning activity as he defines play to be "the capacity to experiment with one's surroundings as a form of problem-solving." Again, in both the open-ended and structured forms, the layer-cake build process is posed as a complex problem-solving activity that unfolds in real time. Importantly, individual participants must cooperatively integrate their thoughts and actions to achieve a shared success.

4.2.3. Curriculum

We collaborated with our partner teacher to design a curriculum that he implemented during a total of three, forty-five minute class periods across three consecutive days. The curriculum is informed by our overarching theoretical framework and is designed to foster student-centered learning. Student activity is structured around a repeating cycle of *composition* → *reflection*. From a modeling instruction perspective [79, 80], this activity cycle supports students' underlying cognitive process that we term as *knowledge construction* → *consolidation*. During the first phase of the cycle, (i.e., activity = *composition* and cognitive process = *knowledge construction*), students construct a simple conceptual model of the evolution of the earth's crust. Teams of students work together in real time to create a layer-cake representation of this model. By engaging in this hands-on, compositional activity, they continuously form, test, and revise their model. This phase is immediately followed by a second stage (i.e., activity = *reflection* and cognitive process = *knowledge consolidation*) in which students discuss their activities, analyze any flaws in decision making, make sense of the various aspects of the layer-cake structure, and challenge one another to justify their choices. This reflective activity leads to a consolidation of the conceptual model that was interactively explored during the first phase. With each iteration of this cycle, new elements are introduced and new knowledge is tested and consolidated, ultimately leading to a robust and coherent conceptual model of the process of geologic evolution.

As this was the first experience in *SMALLab* for most students, day one began with a brief introduction to the basic technology and an overview of the teacher's expectations. The teacher then introduced the technological components of the learning scenario itself and students were divided into teams to begin creating layer-cake structures in an open-ended, exploratory fashion. During this first day, the teacher structured the interactions, frequently pausing the scenario and prompting students to articulate their thinking before continuing the interaction. For example, he first started the geologic clock and asked the depositional environment team to select an environment, leading a discussion of the images and sounds, and what they represent. Once an environment was selected he would stop the geologic clock and ask the sediment layer team to discuss the sediment icons and why a particular selection would be appropriate or not. Restarting geologic time, the team selected their choice for the best sediment layer, placed it in the layer-cake structure. Similar discussions and actions unfolded for the selection of an appropriate fossil. Over the course of the class period, the teacher intervened less and less as the students improved in their ability to coordinate their activities and reason through the construction process on their own. Figure 6 shows an example of the outcome of a layer-cake build cycle. During each reflection stage, we captured screenshot of the layer-cake structure and uploaded and annotated it in the *SLink* database for later reference.

During day two, the teacher introduced the fault event interface and teams assumed this role in a similar manner as the exploration of day one. Discussions regarding the selection of the fossil record grew more detailed as students were challenged to consider both the environmental conditions and the sequence of geologic time in their selection process. For example, students reasoned through an understanding of why mammalian fossils should not appear early in the fossil record due to their understanding of the biological evolution of species. Midway through the class, the teacher moved students to the structured build process. He provided the “depositional environment” team with source images that show geologic cross-sections of the earth’s crust such as the one pictured in Figure 6. These students had to interpret the sequence of sediment layers and uplift/erosion evidence to properly initiate the environments and fault events that would cause the actions that followed to reproduce the source image. Only the few students on the “depositional environment” team had access to this source image. Thus all others’ actions were dependent on their decision making. For example, the “sediment” selection team could potentially add a rock layer that did not align with the source image for a particular geologic period. While this could stem from a misunderstanding by their action lead, this deviation might be due to the improper selection of a depositional environment. Or both the depositional environment and the sediment could be selected incorrectly, causing a chain of deviations that would have to be unraveled at the end of the build. Students continued iterating through the *composition → reflection* process, rotating roles with each cycle, structuring their successive interactions, and measuring their progress with the explicit goal of replicating the reference layer cake. The teacher at times guided this reflective process, but the student “evaluation” team members increasingly led these discussions.

On day three, the teacher led a summative assessment activity. Prior to the session, he worked in *SMALLab* to create a set of four layer-cake structures. He captured screenshots and printed images of these four structures. During class the students worked to recreate each of the structures in a similar manner as in day two. At the end of each build process, the “evaluation” team reported any deviations from the reference structure, and the build teams were given the opportunity to justify and defend their actions. The teacher assigned a grade to each student at the end of the class period. These grades were a measure of their mastery of the build process as indicated by their ability to effectively contribute to the replication of the source structure and/or justify any deviations. Similar to days one and two, team action leads rotated with each new geologic period, and teams rotated through the different roles each time a new script was introduced. During this class session the teacher made very few interventions as students were allowed to reason through the building and evaluation process on their own.

4.3. Outcomes

During the final in-class assessment activity on day three, all teams demonstrated an impressive ability to accurately reproduce the source structure. Collectively, the students composed fifteen layer cakes during day three. Eleven of the results were either a perfect match or within tolerable limits (e.g., only a slight deviation in the intensity of a fault event or no more than one incorrect sediment layer) of the source structure. Deviations typically stemmed from students’ selection of alternate fossils in circumstances where there was room for interpretation or minor deviations in the magnitude of fault events within a given geologic period. Students also exhibited improvement in their ability to justify their actions, developing arguments by the final day which suggest that they quickly developed robust conceptual models of the underlying content.

For example, below is a transcript of the teacher and students in a typical cycle of *composition → reflection* from day one of the treatment. The teacher is controlling geologic time during this episode. When the transcription begins, the students are in the middle of a layer-cake build process and they have just completed discussion about creating one layer in the process. After his first comment, he starts the geologic clock again, and the students commence constructing the next layer. In the transcriptions *T* is the teacher and students are identified by a first initial or *S* if the exact voice could not be identified.

T: *Alright, let’s go one more time.*

(Sound of rushing water. The students with the glowballs pick a sediment layer (sandstone) and a fossil (fish) and lay them into the scenario. This takes lass than 10 seconds. When they are done the teacher pauses the geologic clock to engage them in reflection.)

T: *Alright, depositional environment—what are we looking at?*

Ss: *A river.*

T: *A river. Sandstone. Is that a reasonable choice for a type of rock that forms in a river? (Shrugs) Could be…is there any other types of rock over there that form in a river. Chuck. What’s another rock over there that might form in a river?*

C: *In a river? I can’t find one…*

T: *In a river. (there is a pause of several seconds)*

S: *Conglomerate.*

T: *Alright. Conglomerate is also an acceptable answer. Sandstone's not a bad answer. Conglomerate is pretty good...big chunks of rock that wash down in the river. So, what kind of fossil did you put in?*

S: *A fish.*

T: *A fish, okay. A fish in a stream makes good sense. Let's think about the fossils that we have in here. First we have a trilobite and then we had a jellyfish, then we had a fern and then we had a fish, alright? Is there anything wrong with the order of these animals so far?*

S: *They're aging.*

T: *What do you mean, "they're aging"?*

S: *Evolution?*

T: *It's evolution so which ones should be the older fossils? (pause of several seconds)*

S: *...Trilobite?*

T: *Trilobite in this case...why the trilobite in this case? How do we know the trilobite's the oldest?*

S: *Because it's dead.*

T: *Just look at the picture. How do we know that the trilobite is oldest?*

S: *Because it's on the bottom?*

T: *We know that the oldest rocks are found...*

S: *On the bottom.*

T: *...on the bottom. So that's another thing that we want to make sure that we're keeping in check...we don't want to end up putting a whale on the bottom and a trilobite on top of a whale...because what kind of animal is a whale? (Pause) It's a mammal, alright? Mammals are relatively recently evolved. So let's pass off the spheres, guys. This next cycle I'm going to do a little different. I'm going to let two cycles go through without stopping you. Let's see how well we can do with the two cycles.*

Now we present a brief transcription of a typical episode from day 3. Here, students have just finished building a complete layer cake. One student team controlled the depositional environment and faulting events, another team controlled sediment layers, a third team, controlled fossils, and a fourth team acted as evaluators, determining the plausibility of various elements used in the construction.

T: *Alright, JR, What's the first rock supposed to be?*

JR: *They got them all right.*

T: *All the rocks are correct?*

JR: *Yeah.*

T: *Ok. How about depositional environments, and Walt you're going to have to help her with this...do all the depositional environments match up with the rocks that were chosen?*

W: *Yeah.*

T: *All the rocks match up...what about the fossils, A (student)?*

A: *They actually had some differences...*

T: *It doesn't have to exactly as it is on here. This is just a suggested order, right? What you need to do is figure out whether or not the ones they chose fit their environment.*

W: *Yeah. Well except for...*

S1: *Except for the fern in there...*

W: *Yeah, number 9 was supposed to be a fish, but it was a fern.*

T: *Ok, well, like I said, it doesn't necessarily have to be the fish that's there...is a fern possible as a formation of a fossil in a conglomerate, which is what type of depositional environment?*

S1: *Water...*

S2: *Stream...*

S3: *River...*

T: *A stream...is it possible for a fern to form a fossil in stream environment?*

Many voices: yeah...no, no...no...yeah...

T: *Alright. Bill says there is. Let's hear what you have to say Bill.*

B: *I just said that it can be.*

T: *Okay. How. How would that happen?*

D: *Cuz he thinks he knows everything.*

T: *David. Talk to Bill. I think you have a potential valid argument here but I want to hear it so we can make our...so we can judge.*

D: *That's cool.*

B: *Well...like, ferns grow everywhere, and if it lives near a river it could fall in...*

T: *Do ferns grow everywhere?*

S4: *No, not deserts.*

T: *Where do they typically grow? What do they need to grow?*

Ss: *Water.*

T: *Water. Would a fern growing next to a river make sense?*

Ss: *Yeah.*

T: *Do you think over the course of millions of years that one fern could end up preserved in a river environment?*

B: *Yeah. Fern plants could.*

T: *So since you guys over here are judging the fossils, Andy, do you accept his answer for why there's a fern there?*

A: *Yeah.*

T: *I would agree. I think that's an acceptable answer. It doesn't always have to be the way it pans out on the image here. Any other thing that you see? What about Allen, did they put the earthquake at the right point.*

A: *No. They're a little off.*

T: *How were they a little bit off?*

A: *They went, like, really long.*

T: *Could you be more specific. How did they go a little long?*

A: *She got excited. (Referring to the fact that she shook the Wii Remote hard for almost 10 seconds causing multiple faulting events.)*

S1: *I told you to stop.*

S2: *It's hard to do it right.*

S1: *Have an aneurism why don't you.*

T: *Okay. Allow me to just work it out. I'm mediating. I'm backing you guys up okay? So Allen, the important thing is, did it come at t he right time?*

A: *Yeah. It was just too long.*

T: *Okay. That's more important. Maybe when you use the Wii controller sometimes it's hard to know when to stop.*

S: *Yeah.*

T: *Do you think that this is acceptable the way that they did it.*

S: *Yeah.*

T: *I would agree with you as well. So were there any points taken off for any decisions that were made in creating this geologic cross section?*

S: *No. Not really. It was all good.*

T: *It was all good. . .alright. . .awesome*

These two transcripts demonstrate two important trends. First, there is a marked difference in the nature of the reflective discussion between the two days. The discussion in day 1 is exclusively led by the teacher as he prompts students to respond to direct questions. By day 3, while the teacher serves to moderate the discussion, he is able to steer the more free-flowing conversation in away that encourages students to directly engage one another. Second, owing to the open-ended nature of the build process, students are by day 3 considering alternative solutions and deviations in the outcomes. They discuss the viability of different solutions and consider allowable tolerances. This shows that they are thinking of the process of evolution as a complex process that can have multiple "acceptable" outcomes so long as those outcomes align with their underlying conceptual models.

To assess individual students' content learning gains, we collaborated with the classroom teacher, to create a ten-item pencil and paper test to assess students' knowledge of earth science topics relating to geologic evolution. Each test item included a multiple-choice concept question followed by an open-format question asking students to articulate an explanation for their answer. The content for this test was drawn from topics covered during a typical geologic evolution curriculum and aligning with state and federal science standards. All test concepts were covered in the teacher's classroom using traditional instructional methods in the weeks leading up to the experiment. As such, at the time of the pretest, students had studied (and learned) all of the test material to the full extent that would be typically expected. To be clear, the three-day teaching experiment did not introduce any new concepts but rather only reinforced and reviewed previously studied topics. This concept test was administered one day before and then one day after the *SMALLab* treatment. Every student in our partner teacher's earth science classes participated in the teaching experiment and thus we were not able to administer the test to a control group.

Table 1 shows the pre- and posttest scores for the seventy-two participating students. The summary is divided into two categories for the multiple-choice items and corresponding open-answer explanation items. Open-answer questions were rated on a 0–2 scale where a score of 0 indicates a blank response or nonsense response. A score of 1 indicates a meaningful explanation that is incorrect or only partially accurate. A score of 2 indicates a well-formed and accurate explanation. We computed a percentage increase and the Hake gain for each category. A Hake gain is the actual percent gain divided by the maximum possible gain [81]. Participating students achieved a 22.6% overall percent increase in their multiple-choice question scores, a 48% Hake gain ($P < .00002$, $r = 0.20$, $n = 72$, $std\ deviaton = 1.9$). They achieved a 40.4% overall percent increase in their explanation scores, a 23.5% Hake gain ($P < .000003$, $r = 0.60$, $n = 72$, $std\ deviation = 2.8$). These results reveal that nearly all students made significant conceptual gains as measured by their ability to accurately respond to standardized-type test items and articulate their reasoning.

We also observed that the student-centered, play-based nature of the learning experience had a positive impact on students. All participants were part of the school's CORE program for at-risk students. While many of these students are placed in the program due to low academic performance, after one year of observation, we see that this is often not due to a lack of ability, but rather to a lack of motivation to participate in the traditional culture of schooling. During our three-day treatment we observed high motivation from students. Many students who might otherwise disengage from or even disrupt the learning process emerged as vocal leaders in this context. These students appeared intrinsically motivated to participate in the learning activity and displayed a sense of ownership for the learning process that grew with each day of the treatment. As evidence of the motivating impact of play, we informally observed a group of students from outside the teacher's regular classes. These students previously spoke with their peers about their in-class experience and subsequently visited *SMALLab* during their lunch hour to "play" in the environment. For nearly a full class period these students composed layer-cake structures, working together, unsupervised by any teacher.

5. Conclusions

We have presented theoretical research from HCI and Education that reveals a convergence of trends focused

Table 1: Summary of pre- and posttreatment geologic evolution concept test results.

		Scores
Multiple-choice test items ($n = 72$)	Pretreatment multiple choice average score	6.82
	Posttreatment multiple choice average score	8.36
	Percent increase	22.6%
	Hake gain	48.5%
Free-response justifications ($n = 72$)	Pretreatment explanation average score	3.68
	Posttreatment explanation average score	5.17
	Percent increase	40.4%
	Hake gain	23.5%

on *embodiment*, *multimodality*, and *composition*. While we have presented several examples of prior research that demonstrates the efficacy of learning in environments that align work in HCI and Education, there are few examples of large-scale projects that synthesize all three of these elements. We have presented our own efforts in this regard, using the integration of these three themes as a theoretical and technological framework that is informed by broad definitions of play. Our work includes the development of a new mixed-reality platform for learning that has been pilot tested and evaluated through diverse pedagogical programs, focused user studies, and perception/action experiments. We presented a recent high school earth science program that illustrates the application of our three-part theoretical framework in our mixed-reality environment. This study was undertaken with two primary goals: (1) to advance students' knowledge of earth science content relating to geologic evolution, and (2) to evaluate our theoretical framework and validate *SMALLab* as a platform for mixed-reality learning in a formal classroom learning environment. Participating students demonstrated significant learning gains after only a three-day treatment and exhibited strong motivation for learning as a result of the integration of play in the scenario. This success demonstrates the feasibility of mixed-reality learning design and implementation in a mainstream formal school-based learning environment. Our preliminary conclusions suggest that there is great promise for the convergent themes of applied HCI and Educational research that are manifest in the *SMALLab* learning platform and our three-part theoretical base.

6. Future Work

We are currently working to increase the scope and scale of the *SMALLab* platform and learning programs. With regarding to the technological infrastructure, we are actively pursuing augmented sensing and feedback mechanisms to extend the system. This research includes an integrated framework for robotics, outfitting the tracked glowballs with sensors and wireless transmission capabilities, and integrating an active RFID system that will allow us to track participant locations in the space. We are extending the current multimodal archive to include real-time audio and video data that is interleaved with control data generated by the existing sensing and feedback structures.

With regard to learning programs, we continue our collaboration with faculty and students at a regional high school. We are currently collecting data that will allow us to evaluate the long-term impact of *SMALLab* learning that is correlated across multiple content areas, grades, and instructional paradigms. Concurrently we are developing a set of computationally based evaluation tools that will identify gains in terms successful *SMALLab* learning strategies and the attainment of specific performance objectives. These tools will be applied to inform the design of *SMALLab* programs, support student-centered reflection, and communicate to the larger HCI and Education communities our successes and failures in this research.

Acknowledgments

The authors gratefully acknowledge that this document is supported by the National Science Foundation CISE Infrastructure grant under Grant no. 0403428 and IGERT Grant no. 0504647. They extend their gratitude to the students, teachers, and staff of the Herberger College for Kids, Herrera Elementary School, Whittier Elementary School, Metropolitan Arts High School, Coronado High School, and ASU Art Museum for their commitment to exploration and learning.

References

[1] J. D. Bransford, A. L. Brown, and R. R. Cocking, Eds., *How People Learn: Brain, Mind, Experience, and School*, National Academy Press, Washington, DC, USA, 2000.

[2] E. Hutchins, "Material anchors for conceptual blends," in *Proceedings of the Symposium on Conceptual Blending*, Odense, Denmark, 2002.

[3] D. Hestenes, "Notes for a modeling theory of science cognition and instruction," in *Proceedings of the GIREP Conrerence on Modeling in Physics and Physics Education (GIREP '06)*, p. 27, Amsterdam, The Netherlands, August 2006.

[4] G. Fauconnier and M. Turner, *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*, Basic Books, New York, NY, USA, 2002.

[5] L. W. Barsalou, "Grounded cognition," *Annual Review of Psychology*, vol. 59, no. 1, pp. 617–645, 2008.

[6] M. C. Megowan-Romanowicz, *Framing discourse for optimal learning in science and mathematics*, Ph.D. dissertation, College of Education, Division of Curriculum and Instruction, Arizona State University, Tempe, Ariz, USA, 2007.

[7] A. M. Glenberg, T. Gutierrez, J. R. Levin, S. Japuntich, and M. P. Kaschak, "Activity and imagined activity can enhance young children's reading comprehension," *Journal of Educational Psychology*, vol. 96, no. 3, pp. 424–436, 2004.

[8] P. Dourish, *Where the Action is: The Foundations of Embodied Interaction*, MIT Press, Cambridge, Mass, USA, 2001.

[9] P. Dourish, "What we talk about when we talk about context," *Journal of Personal and Ubiquitous Computing*, vol. 8, no. 1, pp. 19–30, 2004.

[10] J. Rambusch and T. Ziemke, "The role of embodiment in situated learning," in *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, pp. 1803–1808, Stresa, Italy, July 2005.

[11] L. Suchman, *Plans and Situated Actions: The Problem of Human-Machine Communication*, Cambridge University Press, New York, NY, USA, 1987.

[12] F. Varela, E. Thompson, and E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press, Cambridge, Mass, USA, 1991.

[13] E. Hutchins, *Cognition in the Wild*, vol. 1, MIT Press, Cambridge, Mass, USA, 1995.

[14] M. Wilson, "Six views of embodied cognition," *Psychonomic Bulletin and Review*, vol. 9, no. 4, pp. 625–636, 2002.

[15] J. J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, Mass, USA, 1979.

[16] D. Norman, *The Psychology of Everyday Things*, Basic Books, New York, NY, USA, 1988.

[17] R. N. Shepard, "Perceptual-cognitive universals as reflections of the world," *Behavioral and Brain Sciences*, vol. 24, no. 4, pp. 581–601, 2001.

[18] A. M. Glenberg, D. Havas, R. Becker, and M. Rinck, "Grounding language in bodily states: the case for emotion," in *The Grounding of Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, R. Zwann and D. Pecher, Eds., Cambridge University Press, Cambridge, UK, 2005.

[19] P. I. Galperin, "Stage-by-stage formation as a method of psychological investigation," *Journal of Russian and East European Psychology*, vol. 30, no. 4, pp. 60–80, 1992.

[20] I. Harel, *Children Designers*, Ablex, Norwood, NJ, USA, 1991.

[21] S. Papert, *Mindstorms: Children, Computers and Powerful Ideas*, Basic Books, New York, NY, USA, 1980.

[22] S. Papert, *The Children's Machine*, Basic Books, New York, NY, USA, 1993.

[23] M. Resnick, "Technologies for lifelong kindergarten," *Educational Technology Research and Development*, vol. 46, no. 4, pp. 43–55, 1998.

[24] S. Papert, "Situating constructionism," in *Constructionism*, pp. 1–12, Ablex, Norwood, NJ, USA, 1991.

[25] K. Peppler and Y. B. Kafai, "From SuperGoo to Scratch: exploring creative digital media production in informal learning," *Learning, Media and Technology*, vol. 32, no. 2, pp. 149–166, 2007.

[26] H. Ishii and B. Ullmer, "Tangible bits: towards seamless interfaces between people, bits and atoms," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '97)*, pp. 234–241, Atlanta, Ga, USA, March 1997.

[27] H. Ishii, "Tangible User Interfaces," 2007, http://tangible.media.mit.edu/.

[28] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: the design and implementation of the CAVE," in *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 135–142, Anaheim, Calif, USA, August 1993.

[29] W. Swartout, R. Hill, J. Gratch, et al., "Toward the holodeck: integrating graphics, sound, character and story," in *Proceedings of the 5th International Conference on Autonomous Agents*, pp. 409–416, Montreal, Canada, May-June 2001.

[30] G. Weinberg, "The Musical Playpen: an immersive digital musical instrument," *Personal and Ubiquitous Computing*, vol. 3, no. 3, pp. 132–136, 1999.

[31] A. F. Bobick, S. S. Intille, J. W. Davis, et al., "The kidsRoom: a perceptually-based interactive and immersive story environment," *Presence: Teleoperators and Virtual Environments*, vol. 8, no. 4, pp. 369–393, 1999.

[32] D. Cavallo, A. Sipitakiat, A. Basu, et al., "RoBallet: exploring learning through expression in the arts through constructing in a technologically immersive environment," in *Proceedings of the 6th International Conference on Learning Sciences*, pp. 105–112, Santa Monica, Calif, USA, June 2004.

[33] C. Dede and D. J. Ketelhut, "Designing for motivation and usability in a museum-based multi-user virtual environment," in *Proceedings of the American Educational Research Association Conference*, Chicago, Ill, USA, April 2003.

[34] C. Dede, D. J. Ketelhut, and K. Ruess, "Motivation, usability and learning outcomes in a prototype museum-based multi-user virtual environment," in *Proceedings of the 5th International Conference of the Learning Sciences*, Seattle, Wash, USA, October 2002.

[35] D. J. Ketelhut, "The impact of student self-efficacy on scientific inquiry skills: an exploratory investigation in River City, a multi-user virtual environment," *Journal of Science Education and Technology*, vol. 16, no. 1, pp. 99–111, 2007.

[36] B. C. Nelson, D. J. Ketelhut, J. Clarke, E. Dieterle, C. Dede, and B. Erlandson, "Robust design strategies for scaling educational innovations: the River City MUVE case study," in *The Design and Use of Simulation Computer Games in Education*, pp. 209–231, Sense Press, Rotterdam, The Netherlands, 2007.

[37] H. Gardner, *Frames of Mind: The Theory of Multiple Intelligences*, Basic Books, New York, NY, USA, 1993.

[38] R. Jackendoff, "The architecture of the linguistic spatial interface," in *Language and Space*, P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garrett, Eds., pp. 1–30, MIT Press, Cambridge, Mass, USA, 1996.

[39] D. Cunningham, T. Duffy, and R. Knuth, "Textbook of the future," in *Hypertext: A Psychological Perspective*, C. McKnight, Ed., pp. 19–50, Ellis Horwood, London, UK, 1993.

[40] W.-M. Roth and M. K. McGinn, "Inscriptions: toward a theory of representing as social practice," *Review of Educational Research*, vol. 68, no. 1, pp. 35–59, 1998.

[41] G. Burt, "Media effectiveness, essentiality, and amount of study: a mathematical model," *British Journal of Educational Technology*, vol. 37, no. 1, pp. 121–130, 2006.

[42] C. R. Wren, S. Basu, F. Sparacino, and A. Pentland, "Combining audio and video in perceptive spaces," in *Proceedings of the 1st International Workshop on Managing Interactions in Smart Environments (MANSE '99)*, pp. 44–55, Dublin, Ireland, December 1999.

[43] C. Bahn, T. Hahn, and D. Trueman, "Physicality and feedback: a focus on the body in the performance of electronic music," in *Proceedings of the International Computer Music Conference (ICMC '01)*, San Francisco, Calif, USA, November 2001.

[44] A. Hunt, M. Wanderley, and M. Paradis, "The importance of parameter mapping in electronic instrument design," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME '02)*, Dublin, Ireland, May 2002.

[45] D. Wessel and M. Wright, "Problems and prospects for intimate musical control of computers," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME '01)*, pp. 1–4, Seattle, Wash, USA, April 2001.

[46] P. Rheingans and C. Landreth, "Perceptual principles for effective visualizations," in *Perceptual Issues in Visualization*, pp. 59–73, Springer, Berlin, Germany, 1995.

[47] L. Muller and E. Edmonds, "Living laboratories: making and curating interactive art," in *Proceedings of the ACM SIGGRAPH International Conference on Computer Graphics and Interactive Techniques*, Boston, Mass, USA, July-August 2006.

[48] EU Community Report, Information Society Technologies (IST) Programme, "A multisensory environment design for an interface between autistic and typical expressiveness," 2004.

[49] N. Parés, A. Carreras, J. Durany, et al., "MEDIATE: an interactive multisensory environment for children with severe autism and no verbal communication," in *Proceedings of the 3rd International Workshop on Virtual Rehabilitation (IWVR '04)*, Lausanne, Switzerland, Septembe 2004.

[50] N. Parés, P. Masri, G. van Wolferen, and C. Creed, "Achieving dialogue with children with severe autism in an adaptive multisensory interaction: the "MEDIATE" project," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 6, pp. 734–742, 2005.

[51] L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*, The Harvard University Press, Cambridge, Mass, USA, 1978.

[52] M. Dougiamas, "A journey into Constructivism," November 1998.

[53] P. Cobb, "Constructivism in math and science education," *Educational Researcher*, vol. 23, no. 7, p. 4, 1994.

[54] L. S. Vygotsky, *Thought and Language*, MIT Press, Cambridge, Mass, USA, 1962.

[55] R. Lesh and H. Doerr, *Beyond Constructivism: Models and Modeling Perspectives on Mathematics Problem Solving, Learning, and Teaching*, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2003.

[56] J. S. Brown, A. Collins, and P. Duguid, "Situated cognition and the culture of learning," *Educational Researcher*, vol. 18, no. 1, pp. 32–42, 1989.

[57] J. A. Middleton and Z. Toluk, "First steps in the development of an adaptive theory of motivation," *Educational Psychologist*, vol. 34, no. 2, pp. 99–112, 1999.

[58] M. Resnick, "Computer as paintbrush: technology, play, and the creative society," in *Play = Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth*, D. Singer, R. Golikoff, and K. Hirsh-Pasek, Eds., Oxford University Press, Cambridge, UK, 2006.

[59] K. Ryokai, S. Marti, and H. Ishii, "I/O brush: drawing with everyday objects as ink," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '04)*, pp. 303–310, Vienna, Austria, April 2004.

[60] M. Resnick, "All I really need to know (about creative thinking) I learned (by studying how chidren learn) in kindergarten," in *Proceedings of the 6th ACM SIGCHI Conference on Creativity and Cognition*, pp. 1–6, Washington, DC, USA, June 2007.

[61] Y. B. Kafai, K. Peppler, and G. Chin, "High tech programmers in low income communities: creating a computer culture in a community technology center," in *Proceedings of the 3rd International Conference on Communities and Technologies*, C. Steinfeld, B. Pentland, M. Achkermann, and N. Contractor, Eds., pp. 545–562, East Lansing, Mich, USA, June 2007.

[62] K. Salen and E. Zimmerman, *Rules of Play: Game Design Fundamentals*, MIT Press, Cambridge, Mass, USA, 2004.

[63] J. P. Gee, *What Video Games have to Teach Us about Learning and Literacy*, Palgrave Macmillan, New York, NY, USA, 2003.

[64] H. Jenkins, K. Clinton, R. Purushotma, A. Robison, and M. Weigel, Confronting the Challenges of Participatory Culture: Media Education for the 21st Century: MacArthur Foundation, 2006.

[65] SmartUs, "SmartUs: Games in Motion," 2008, http://www.smartus.com/.

[66] D. Birchfield, T. Ciufo, and G. Minyard, "SMALLab: a mediated platform for education," in *Proceedings of the ACM SIGGRAPH International Conference on Computer Graphics and Interactive Techniques*, p. 33, Boston, Mass, USA, July-August 2006.

[67] A. Shirai, E. Geslin, and S. Richir, "WiiMedia: motion analysis methods and applications using a consumer video game controller," in *Proceedings of the ACM SIGGRAPH Symposium on Video Games*, pp. 133–140, San Diego, Calif, USA, August 2007.

[68] M. Brain, "How the Wii Works," 2007, http://electronics.howstuffworks.com/wii.htm.

[69] H. Jin, G. Qian, and S. Rajko, "Real-time multi-view 3D object tracking in cluttered scenes," in *Proceedings of the 2nd International Symposium on Visual Computing (ISVC '06)*, pp. 647–656, Lake Tahoe, Nev, USA, November 2006.

[70] K. Tu, H. Thornburg, E. Campana, D. Birchfield, M. Fulmer, and A. Spanias, "Interaction and reflection via 3D path shape qualities in a mediated constructive learning environment," in *Proceedings of the International Workshop on Educational Multimedia and Multimedia Education (EMME '07)*, pp. 37–46, Augsburg, Germany, September 2007.

[71] K. Tu, H. Thornburg, M. Fulmer, and A. Spanias, "Tracking the path shape qualities of human motion," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 2, pp. 781–784, Honolulu, Hawaii, USA, April 2007.

[72] I. Dolgov, D. Birchfield, M. McBeath, H. Thornburg, and C. Todd, "Perception of approaching and retreating floor-projected shapes in a large, immersive, multimedia learning environment," *Journal of Perceptual and Motor Skills*. In press.

[73] A. Cuthbertson, S. Hatton, G. Minyard, H. Piver, C. Todd, and D. Birchfield, "Mediated education in a creative arts context: research and practice at Whittier Elementary School," in *Proceedings of the 6th International Conference on Interaction Design and Children (IDC '07)*, pp. 65–72, Aalborg, Danmark, June 2007.

[74] S. Hatton, D. Birchfield, and M. C. Megowan-Romanowicz, "Learning metaphor through mixed-reality game design and game play," in *Proceedings of the ACM SIGGRAPH Symposium on Video Games (Sandbox '08)*, pp. 67–74, Los Angeles, Calif, USA, August 2008.

[75] D. Birchfield, B. Mechtley, S. Hatton, and H. Thornburg, "Mixed-reality learning in the art museum context," in *Proceedings of ACM SIG Multimedia*, Vancouver, Canada, October-November 2008.

[76] D. Chen and W. Stroup, "General system theory: toward a conceptual framework for science and technology education for all," *Journal of Science Education and Technology*, vol. 2, no. 3, pp. 447–459, 1993.

[77] F. Lutgens, E. Tarbuck, and D. Tasa, *Foundations of Earth Science*, Prentice Hall, Upper Saddle River, NJ, USA, 4th edition, 2004.

[78] Arizona Department of Education, "Science Standard Articulated by Grade Level Strand 6: Earth and Space Science," 2005, http://www.ade.az.gov/standards/science/downloads/strand6.pdf.

[79] D. Hestenes, "Modeling methodology for physics teachers," in *Proceedings of the International Conference on Undergraduate Physics Education (ICUPE '96)*, College Park, Md, USA, July-August 1996.

[80] D. Hestenes, "Modeling games in the newtonian world," *American Journal of Physics*, vol. 60, no. 8, pp. 732–748, 1992.

[81] R. R. Hake, "Interactive-engagement vs traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses," *American Journal of Physics*, vol. 66, no. 1, pp. 64–74, 1998.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 341615, 8 pages
doi:10.1155/2008/341615

Research Article

BlogWall: Social and Cultural Interaction for Children

Adrian David Cheok,[1] Owen Noel Newton Fernando,[1] Janaka Prasad Wijesena,[1] Abd-ur-Rehman Mustafa,[1] Ramkumar Shankar,[1] Anne-Katrin Barthoff,[1] Naoko Tosa,[2] Yongsoon Choi,[1] and Mayank Agarwal[1]

[1] *Mixed Reality Lab, National University of Singapore, Singapore 119077*
[2] *Academic Center for Computing and Media Studies, Kyoto University, Kyoto 606-8501, Japan*

Correspondence should be addressed to Adrian David Cheok, adriancheok@mixedrealitylab.org

Received 11 October 2007; Accepted 11 August 2008

Recommended by Junichi Osada

Short message service (SMS) is extremely popular today. Currently, it is being mainly used for peer-to-peer communication. However, SMS could be used as public media platform to enhance social and public interactions in an intuitive way. We have developed BlogWall to extend the SMS to a new level of self-expression and public communication by combining art and poetry. Furthermore, it will provide a means of expression in the language that children can understand, and the forms of social communication. BlogWall can also be used to educate the children while they interact and play with the system. The most notable feature of the system is its ability to mix up and generate poetry in multiple languages such as English, Korean, Chinese poems, or Japanese "Haiku" all based on the SMS. This system facilitates a cultural experience to children unknowingly, thus it is a step into new forms of cultural computing.

1. Introduction

SMS or short message service is immensely popular among mobile phone users today [1]. The volume of short messages that Hong Kong people sent during the period December 2006 to November 2007 amounted to 3.28 billion, a record high since office of the telecommunications authority (OFTA) started collecting the figures in 2002, translating into an increase of 26% as compared to the same period in 2005 and 2006 [2]. But, it is primarily used for peer-to-peer communication. It has not widely being used for public or social communications.

BlogWall is an extension of the existing text messaging to a new level of self-expression and public communication, combining visual art and poetry. Furthermore, it provides a means of expression in the language that children can understand and the forms of social communication, which is an essential part of their lives. The application enables a person not only to express herself artistically, but also to entertain the masses in a form of digital graffiti. Figure 1 shows the concept design of BlogWall. The main feature of BlogWall is its ability to mix-up poetry from different poems and to generate new poetry based on users' SMS.

System such as "i.plot" [3] discovers hidden connections between unrelated words by tracing possible paths through a database, traversing many two-word connections built from content based on publicly available resources. BlogWall extends the concept of "i.plot" further, creating a new type of poetry by elevating SMS to new art form. The longest words in the SMS are selected to locate synonyms from the Internet (currently by using the "Free dictionary website" [4]). The synonyms are used to rate poem lines out of a database. The highest-rated lines will be shown to the user. The extension of this poetry mode is the formation of cultural poetry. The system uses databases containing Korean, Chinese, and Japanese poetry and their English meaning. Synonyms found from the Internet are used to select poetry from these databases. The original text message, poetry in original language, and its English meaning are displayed on the screen.

Other than poetry mode, there are several other modes available in BlogWall. It can directly display users SMS in an

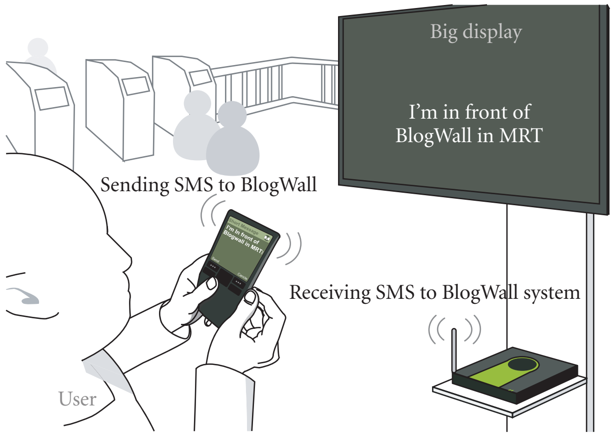


Figure 1: Concept design of BlogWall.

interesting way. The polling system available in BlogWall can be used to gather data from the public. Predefined keywords in the SMS can also be replaced by images when the SMS is displayed. Any of these modes can be activated based on users' requirements.

2. Why Generate Poetry by SMS?

People yearn to express themselves in public. Graffito is a one form of public expression that was very popular in mid-70s [5]. With the growth of modern technology, people find new avenues to express themselves. Blogs enable people to share their ideas with millions of others all over the world. Today, Internet websites such as "Youtube" and "Twango" are popular ways of sharing videos and pictures with others.

Nowadays, electronic displays can be seen far and wide. These public displays are mainly used for conveying information rather than providing artistic and social communication which is of fundamental importance in the dubious world order. Modern life is weighted down not only by the global dilemmas, but also by the pace and the stress associated with it. Ironically, people spend a substantial amount of their precious time at airports, train stations, and so forth. One of the devices that almost all people have is a mobile phone. It will be a groundbreaking concept with unlimited possibilities ahead if artistic and social communication can be promoted using mobile phones and public displays.

Technological development poses undeniable threats because it will detach people from their cultures to a certain degree. Children of today who are interacting with these new technologies are finding more and more things to get absorbed in the popular culture. This development is affecting some of the cultural aspects of great cultures. Their folk songs, cultural dances, events, traditional dresses, and so forth are fast becoming museum pieces. But, it is not an inherent limitation of the technological development that is causing this, but it is how we use this new technologies. New technologies have successfully been integrated to preserve various cultures aspects. But, true preservation will not happen in a piece of plastic or silicon. It should happen in the hearts and minds of people. Since children are finding lot less in common with their culture, an approach must be created to embed culture into applications that children can understand and adore.

Poetry is one aspect of culture that has a long history, which will be recaptured in our system. It is also a wonderful way of expressing our inner thoughts. Studies have shown that poetry can be used very effectively to improve mental health [6]. Every one, in some manner, has a certain level of imagination and creativity to express himself in poetry. But, it is very hard for most of the people to actually construct a poem. Hence, a methodology should be there to bridge this gap effectively. SMS provides ideal basis because it is extremely user friendly, and since most people of today have mobile phones, it is a readily available facility.

3. Related Research

Researchers around the world have been experimenting with different combinations of art, public displays, and mobile messaging. The mobile phone has already been used as a medium of self-expression [7]. Ballagas et al. [8] discuss enabling interactions with large public displays using mobile phone. They have used the embedded camera on mobile phones as an enabling technology. Ballagas et al.'s "Point & Shoot" technique allows users to select objects using visual codes to setup an absolute coordinate system on the display surface instead of tagging individual objects on the screen. Joe Blogg [9] is a public display, where users can contribute content by sending messages and images to it using their mobile phones. TexTales [10] is a large scale photographic installation to which people can send SMS text message captions. It can create technologically supported public discourse spheres in which they can both represent personal views and practice new ways of forming collective opinions.

One of the pioneering works in cultural computing was ZENetic computer [11]. It is an interface that evokes self-awakening through essential aspects of Zen Buddhist culture. It tries to offer users a chance to engage and understand Buddhist principles of "recreation" of the self. With stories portrayed in ink painting, haiku, and kimono, ZENetic conveys the rich allegorical interaction characteristic of Eastern philosophy. Another groundbreaking research was "Hitch Haiku" [12]. The system condenses an essence of a book into a Haiku, a Japanese minimal poem form. When a user chooses arbitrary phrases, system generates the haiku which includes the essence, then translates it into English. The cultural computing project ALICE [13] as an interactive, entertaining experience inspired from *Alice in Wonderland* [14]. In the scope of this project, interactive adventures are experiences provided by an augmented reality (AR) environment based on selected parts from Lewis Carroll's book *Alice's Adventures in Wonderland*. The user assumes the role of Alice and explores this interactive narrative. The project uses AR as a new medium for edutainment and entertainment as a particular carrier for cultural transformations.

The BlogWall consists of many of the features found on above-mentioned systems. SMS is the basis of BlogWall

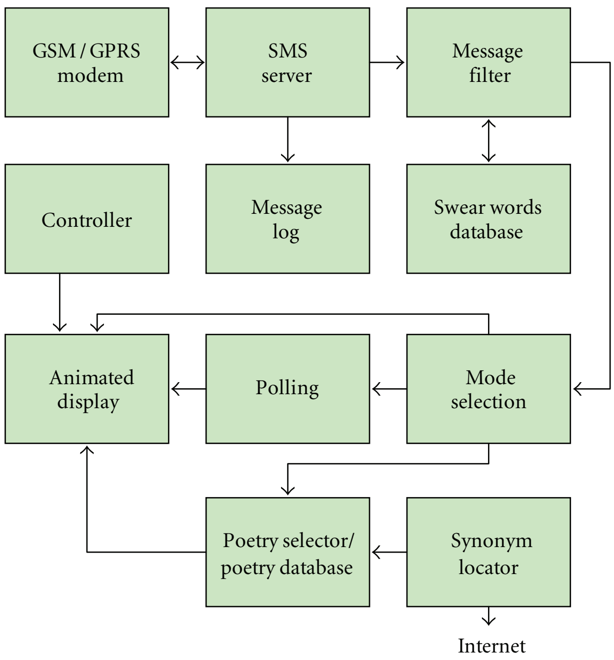


Figure 2: System overview.

(a)

(b)

Figure 3: Animated display.

operations like other systems. However, many of them directly display the SMS or use the mobile phone to control an action, but BlogWall uses the SMS to promote artistic and social communication through poetry. One of the main advantages of BlogWall over other cultural computing applications is its ease of use. Almost no domain knowledge is needed to use the BlogWall system. It has also been clearly designed to be used in public space by millions of users. Unlike other systems, using multilingual poetry, BlogWall can promote cross-cultural understanding.

4. System Architecture

The general setup of BlogWall requires a high-end computer with a good graphics card, projector, and a screen. Dedicated GSM/GPRS modem is used to receive SMS messages. The user stands in front of the projector screen and sends an SMS to a given number. The application issues AT commands to the modem to locate SMS. All the messages received by the server will be written to a log file along with originator phone number and the date/time. The application consists of several modes of operation. Based on the enabled modes, it offers different services to the user. The complete system architecture is shown in Figure 2.

5. Display Modes

Even though the main focus of BlogWall is poetry, there are several other ways it can be used. The additional modes of BlogWall can be activated based on user's requirements. These modes are mainly used as value addition to the overall system.

5.1. Standard Display Mode

If the application is in the standard display mode, it will immediately display the text message with some animation. When the system receives the SMS, it selects a random animation for the SMS as shown in Figure 3.

5.2. Polling Mode

The polling function is used to collect user opinions. The system displays polling question and available answers. The answers are indexed by single alphabetical letter within parenthesis. To vote, users send SMS with appropriate indexed letter of the answer to the system. BlogWall also has the capability to provide statistical data to system administrators.

5.3. Keyword Triggering Mode

Keyword triggering mode enables the application to trigger an internal function based on a word in the SMS. This feature is somewhat similar to the features found in popular chat programs like Windows Messenger. For example, if the SMS contains word "love" the application may replace the word "love" with an image of a heart. The keyword triggering mode can also display a small verse based on the words found in the SMS. The images as well as verses are selected from an internal database.

5.4. Poetry Mode

The most prominent feature of the application is its ability to mix poetry. In the poetry mode, user's SMS is analyzed to identify most prominent words. Those words are posted to a free dictionary website. The website will generate the synonyms related to the posted words, and the system finds them by analyzing the HTML response of the website. The synonyms are used to rank poems and poetry lines in an internal database. The database consists of lines of poetry and keywords related to each line and poem. When the system administrator adds a new poem to the system, all the keywords related to the poem and poem lines get generated and stored in separate table in the database.

The ranking of the poems is done based on the number of synonyms found in each poem. The poem that contains most number of synonyms gets the highest poem ranking. In the selected poem, the line that contains the highest number of synonyms gets the highest line ranking. Based on ranking, most suitable poetry lines will be selected to display.

For example, if the user sends an SMS, "I love thunder and rain" the selected words would be love, thunder, and rain. Some of the synonyms generated for word "love" may be passion, dearest, loved one. Based on the poem and line ranking, system might select the line "A heart whose love is innocent!" from the poem "She Walks in Beauty" by Lord Byron. Similarly, two more lines of poetry will be selected based on the synonyms generated for words thunder and rain. So, a possible output of this SMS might be the following:

> "A heart whose love is innocent!"
>
> "Boom, boom, boom,"
>
> "Rain from the clouds."

This unique ranking system enables the system to borrow lines of poetry from different poets. Therefore, the final outcome of the system could be unusual, surprising, or maybe amusing.

5.4.1. Multilingual Poetry

Due to the flexible nature of BlogWall, it can be extended to different cultures. The application has been extended to Korean, Japanese, Chinese, Malay, Tamil, and Sinhala poetry. The system consists of a very large poetry database. All the poetry in this database is translated to each of the displaying languages. When the system receives an SMS, it selects a line of poetry by ranking the English translation of it. The ranking system is similar to that of poetry mix-up method described above. Table 1 shows all the poetry generated for the SMS "I love thunder and rain" in multilingual poetry mode.

BlogWall system simultaneously displays the generated poetry in all different languages as shown in Figure 4. Each side of the rotating cube displays poetry in the respective language and its English meaning. Through the generation of multilingual poetry, we aim to create another step in the promotion of cross-cultural computing and cross-cultural understanding.

Table 1: Poetry in multilanguages.

Language	Poem
English	Rain of the cold season Wing of love The one of the thunder during the coldest season
Haiku	寒の雨 愛の羽根 寒雷の
Chinese	寒之雨 爱的翅膀 寒雷之
Malay	Hujan di musim sejuk Sayab percintaan Guruh di musim sejuk
Tamil	குளிர்கால மழை காதலின் சிறகு கடுங்குளிர்கால மின்னல்களில் ஒன்று
Sinhala	සීත කාලයේ වැස්ස ආලයේ අත්තටුව එකම අකුනක් සීත කාලයේ
Korean	추운 계절의 비 사랑의 날개 가장 추운 계절의 번개 중의 하나

6. BlogWall User Study

6.1. Overview of the User Study

The application has been demonstrated at several exhibitions and on permanent display at Singapore Science Centre (Figure 5). In those exhibitions, we have observed many people enjoying the interaction with BlogWall and we wanted to discover the interesting and exciting factors that attract the crowd towards this application. In this study, we focused on the social and engagement aspects of the user experience, while the cultural aspects will be a focus on future research.

To understand the users' engagement and social experience, we have used features described by Csikszentmihalyi's Flow theory [15]. The flow of the system is broken down as follows.

(1) *Concentration*. Since Concentration on the application is needed, the ability of the user to concentrate on the application is emphasized.

(2) *Challenge*. Does the system provide some level of challenge to the user.

(3) *User Skills*. User skill development and mastery should be supported in the system.

(4) *Control*. Users should be able to feel in control of their actions.

(5) *Clear Goals*. Clear defined goals should be given to the players at appropriate times.

(6) *Feedback*. Appropriate feedback from the system is given.

(a)

(b)

Figure 4: BlogWall poetry.

Figure 5: School children are interacting with BlogWall at Singapore Science Centre.

(7) *Connection*. Participants should feel deeply connected to the application and with little/no effort.

(8) *Social Interaction*. The system should support social interaction.

We came up with following set of questions to evaluate the system based on our hypotheses. Users were asked to interact with all the modes available in the system, including standard display, keyword triggering, and poetry. But, the main emphasis of the user study was focused on the poetry mode.

6.2. User Study Questions

User study questions and hypothesis pairs are as follows.

(1) User Concentration of the System

Question. Did the application managed to capture users' attention?

Hypothesis. The application should be able to capture users' attention.

Example Question. (1) Did the application grab your attention and maintain your focus?

Question. Were they able to concentrate on the task?

Hypothesis. User should be able to concentrate on the task at hand without any difficulties.

Example Question. (2) Were you able to concentrate on the tasks at hand?

(2) User Skills Required in the System

Question. Were the users able to interact with the system without much assistance?

Hypothesis. The only required skill expected in the system is the ability to send SMS. So the users should be able to proceed with little instructions.

Example Question. (3) Were you able to use the system without spending too much time at the instructions?

Question. Was learning to use the system fun?

Hypothesis. Time spent on learning to operate the system is small. Few simple instructions would be sufficient to start interacting with BlogWall. Still learning how to create poetry from SMS should be fun.

Example Question. (4) Was learning to use the system fun?

(3) Amount of Control Users Has of the System

Question. Did users feel they were in control instead of the system?

Hypothesis. The system behavior is directly connected to users' actions. So they should feel very much in control.

Example Question. (5) Did you feel in control of your actions in the system?

Question. Did they have control of the poetry output of the system?

Hypothesis. The poetry is generated using the SMS send by the user. BlogWall make use of the synonyms it found to generate poetry. Final output may be a surprising result to the users SMS. However, it is still derived from users SMS. Therefore, they are very much in control of the poetry output.

Example Question. (6) Did you feel you were in control of the poetry output of the system?

(4) Clear Goal Is Given to the Users

Question. Was the overall objective of the system is clear to the user?

Hypothesis. Just by looking at the system, users should be able to comprehend the goal of the system.

Example Question. (7) Was the objective of the system clear and presented in advance?

(5) Feedback Given to the Users by the System

Question. Did users feel they knew what is happening in the system all the time?

Hypothesis. Users should have a general notion of what is happening inside the system. But, they may not have a complete image of how the system is progressing.

Example Question. (8) Were you aware what is happening in the system at all times?

Question. Did the system provide fast feedback?

Hypothesis. The system is based on SMS. The time required to receive the SMS should be very small. But, it might take some time to generate poetry since it has to search the Internet for the synonyms. Still the feedback of the system should be very high.

Example Question. (9) Did the system give you immediate feedback of your actions?

(6) How Connected Are the Users to the System

Question. Were users deeply involved in interacting with the system?

Hypothesis. The main task for the user is sending an SMS. But to generate proper poetry, they have to think deeply about the SMS. Therefore, there should be a medium level of connection with the system.

Example Questions. (10) During participation, were you less aware of what is happening physically around you?

(11) Were you aware of the passing of time during participation?

Question. Were they emotionally involved with the system?

Hypothesis. Aesthetic creativeness is very much an emotional process. They may be emotionally involved with the creative process of poetry thus inadvertently absorbed with the system as their interaction is with the system. Users should be emotionally involved with the system.

Example Question. (12) Did you feel emotionally involved with the system?

(7) Social Interaction of the System

Question. Did it support social interaction?

Hypothesis. One of the main objectives of the system is to support social communication. Large number of people can interact with the system enabling social communication. Therefore, users should feel the application supporting social communication.

Example Question. (13) Do you suppose that the system support social communication?

Question. Does the system support social communities inside and outside the system?

Hypothesis. Generally, users are expected to assemble to where the BlogWall is setup in order to interact. However, users can also send SMS to the system and interact with it without physically being there. Therefore, they should feel BlogWall does support social communities.

Example Question. (14) Does the system support social communities inside and outside the system?

The original set of questions for the flow model has been created for a generic task (like surfing the Internet), therefore, some questions have been modified to adapt to the BlogWall system. For the user study, 13 subjects were randomly selected and their average age is 27.3 (60% male and 40% female).

6.3. Results of the User Study

Data collected from the survey after using the system are expressed as mean and standard deviation unless otherwise specified. Results of the survey are given in Table 2. Of all elements explored with this survey, most of them performed positively in the survey as more than 50% selected the favorable choice to the questions posed.

Unlike in computer games, users do not have to concentrate very hard on the task at hand. Users can take their time in interacting with the system. It attracted the users but does not demand total concentration. This may explain the answers received for questions 1 and 2.

One of the reasons we chose SMS as the basis for the poetry generation is that, it is very easy to use and according to the results of question 3, 69% users highly agree to that.

The overwhelming majority, 61%, responded positively to question 6. Thus, users clearly feel that they can control the poetry output of the system. However, they did not respond very positively to question 5. This means that users did not feel that they are in total control of their actions. The poetry generation does not depend totally on users SMS. BlogWall searches the Internet and finds synonyms for the selected words in the SMS. Therefore, the final poetry may

Table 2: Results of user study.

Qn	Options				
	Yes very	Yes	Fairly	Not really	No
1	8%	70%	22%	0%	0%
2	31%	46%	15%	8%	0%
3	69%	31%	0%	0%	0%
4	16%	47%	37%	0%	0%
5	16%	39%	45%	0%	0%
6	61%	28%	11%	0%	0%
7	0%	53%	31%	8%	8%
8	8%	45%	39%	8%	0%
9	31%	54%	15%	0%	0%
10	0%	24%	52%	24%	0%
11	0%	24%	45%	31%	0%
12	16%	31%	29%	24%	0%
13	47%	39%	14%	0%	0%
14	24%	52%	24%	0%	0%

not contain the exact words in the users SMS. In fact, this is a feature of BlogWall. But, it may explain the users' surprise of the outcome and the feeling of not totally being in control.

There is a clear indication that some users considered that the objective of the system is not clearly portrayed. 16% of the users responded negatively to question 7. One user commented that "short instruction information should be displayed along with the BlogWall." Based on these findings, we are now displaying clear instructions to the users wherever we install the BlogWall system. Short description is also presented on the BlogWall display.

Even though feedback time can be compromised by mobile phone network and Internet delays, 31% of users do agree that the system provides an immediate response. This means that the users feel that the system response time is satisfactory. There is a time gap between user sending the SMS and system receiving it. After receiving the SMS, the system searches the Internet to locate the synonyms. During those processes, there is no indication to the user that the system is processing her SMS, which may be the reason that 8% responded that they were unaware of the activities of the system at all times.

Only 24% were less aware of what is happening around them and 24% were aware of the passing of time during participation. This is not an unexpected result since BlogWall does not require the user to be fully mentally connected with it. The response to question 12 is disappointing, 24% users felt that there is no emotional connection with the system. This can be explained by the fact that the poetry selection algorithm selects poetry based on simple line ranking. The selected poem may not reflect the overall context of the users SMS. By developing a natural language processor, the system should be able to analyze the context or linguistic meaning of the SMS and generate the poetry accordingly.

None of the users responded negatively to questions 13 and 14. This indicated that people did feel that the system does support social communication. Also, they believed it supports social communities inside and outside the system. This is a very significant result since social interaction is one of the foremost aspirations of BlogWall. According to the user response, BlogWall has successfully achieved it.

7. Conclusions

BlogWall is a novel mobile artistic media application which promotes self-expression and public communication, combining visual art and poetry. By selecting proper elements, the application can be effectively used to build awareness among children about diverse cultural elements of humans as global beings. BlogWall can also be used for other purposes, such as learning applications. For instance, students and teachers could use the BlogWall to ask questions and provide answers in real-time. The users studies conducted indicate that the application has a high level of social and engagement aspects.

Acknowledgments

The initial version of the BlogWall is funded by Singapore Science Centre and it is installed in the iSpace at Singapore Science Centre.

References

[1] R. E. Grinter and M. A. Eldridge, "y do tngrs luv 2 txt msg?" in *Proceedings of the 7th European Conference on Computer Supported Cooperative Work (ECSCW '01)*, pp. 219–238, Kluwer Academic Publishers, Bonn, Germany, September 2001.

[2] Hong Kong SMS Usage Increased, 2007, http://www.chinatechnews.com/2008/02/01/6361-hong-kong-sms-usage-increased-in-2007.

[3] N. Tosa, A. R. Torroella, B. Ellis, S. Matsuoka, and R. Nakatsu, "Computing inspiration: i.plot," in *Proceedings of the 32nd International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '05)*, p. 3, ACM Press, Los Angeles, Calif, USA, July-August 2005.

[4] Free dictionary, http://www.thefreedictionary.com/.

[5] "Graffiti," http://en.wikipedia.org/wiki/Graffiti.

[6] S. McArdle and R. Byrt, "Fiction, poetry and mental health: expressive and therapeutic uses of literature," *Journal of Psychiatric & Mental Health Nursing*, vol. 8, no. 6, pp. 517–524, 2001.

[7] K. Cheverst, A. Dix, D. Fitton, et al., "Exploring mobile phone interaction with situated display," in *Proceedings of Workshop on Pervasive Mobile Interaction Devices (PERMID '05)*, pp. 43–47, Munich, Germany, May 2005.

[8] R. Ballagas, M. Rohs, and J. G. Sheridan, "Sweep and point and shoot: phonecam-based interactions for large public displays," in *Proceedings of Conference on Human Factors in Computing Systems (CHI '05)*, pp. 1200–1203, ACM Press, Portland, Ore, USA, April 2005.

[9] K. Martin, A. Penn, and L. Gavin, "Engaging with a situated display via picture messaging," in *Proceedings of Conference on Human Factors in Computing Systems (CHI '06)*, pp. 1079–1084, ACM Press, Montréal, Canada, April 2006.

[10] M. Ananny, C. Strohecker, and K. Biddick, "Shifting scales on common ground: developing personal expressions and public opinions," *International Journal of Continuing Engineering*

Education and Life-Long Learning, vol. 14, no. 6, pp. 484–505, 2004.

[11] N. Tosa, S. Matsuoka, and K. Miyazaki, "Interactive storytelling system using behavior-based non-verbal information: ZENetic computer," in *Proceedings of the 11th ACM International Conference on Multimedia (MULTIMEDIA '03)*, pp. 466–467, ACM Press, Berkeley, Calif, USA, November 2003.

[12] N. Tosa, "Hitch-Haiku," in *Proceedings of the 34th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '07)*, p. 250, ACM Press, San Diego, Calif, USA, August 2007.

[13] R. Nakatsu, M. Rauterberg, and P. Vorderer, "A new framework for entertainment computing: from passive to active experience," in *Proceedings of the 4th International Conference on Entertainment Computing (ICEC '05)*, vol. 3711 of *Lecture Notes in Computer Science*, pp. 1–12, Sanda, Japan, September 2005.

[14] L. Carroll, *Alice's Adventures in Wonderland*, Macmillan, London, UK, 1865.

[15] M. Csikszentmihalyi, *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play*, Jossey-Bass, San Francisco, Calif, USA, 1975.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 384749, 10 pages
doi:10.1155/2008/384749

Research Article

Guitarist Fingertip Tracking by Integrating a Bayesian Classifier into Particle Filters

Chutisant Kerdvibulvech and Hideo Saito

Department of Information and Computer Science, Keio University, 3-14-1 Hiyoshi, Kohoku-ku 223-8522, Japan

Correspondence should be addressed to Chutisant Kerdvibulvech, chutisant@ozawa.ics.keio.ac.jp

Received 3 October 2007; Revised 19 July 2008; Accepted 30 August 2008

Recommended by Adrian Cheok

We propose a vision-based method for tracking guitar fingerings made by guitar players. We present it as a new framework for tracking colored finger markers by integrating a Bayesian classifier into particle filters. This adds the useful abilities of automatic track initialization and recovery from tracking failures in a dynamic background. Furthermore, by using the online adaptation of color probabilities, this method is able to cope with illumination changes. Augmented Reality Tag (ARTag) is then utilized to calculate the projection matrix as an online process which allows the guitar to be moved while being played. Representative experimental results are also included. The method presented can be used to develop the application of human-computer interaction (HCI) to guitar playing by recognizing the chord being played by a guitarist in virtual spaces. The aforementioned application would assist guitar learners by allowing them to automatically identify if they are using the correct chords required by the musical piece.

1. Introduction

Due to the popularity of acoustic guitars, research about guitars is one of the most popular topics in the field of computer vision for musical applications.

Maki-Patola et al. [1] proposed a system called "Virtual Air Guitar" (VAG) using computer vision. Their aim was to create a virtual air guitar which does not require a real guitar (e.g., by using only a pair of colored gloves) but produces music similar to a player using a real guitar. Liarokapis [2] proposed an augmented reality system for guitar learners. The aim of this work is to show the augmentation (e.g., the positions where the learner should place the fingers to play the correct chords) on an electric guitar as a guide for the player. Motokawa and Saito [3] built a system called *Online Guitar Tracking* that supports a guitarist using augmented reality. This is done by showing a virtual model of the fingers on a stringed guitar as a teaching aid for anyone learning how to play the guitar.

These systems do not aim to track the fingering which a player is actually using (a pair of gloves are tracked in [1], and graphics information is overlaid on captured video in [2, 3]). We have different goal from most of these researches. In this paper, we propose a method for tracking the guitar fingering by using computer vision. Our research goal is to accurately determine and track the fingertip positions of a guitarist which are relative to guitar position in 3D space.

A challenge for tracking the fingers of a guitar player is that the guitar neck usually moves while the guitar is being played. It is therefore necessary to identify the guitar's position relative to the camera's position. Another important issue is recovery when the finger tracking fails. Our method for tracking the fingers of guitar player solves these problems.

At every frame, we first estimate the projection matrix of each camera by utilizing Augmented Reality Tag (ARTag) [4]. ARTag's marker is placed on the guitar neck. Therefore, the world coordinate system is defined on the guitar neck as the guitar coordinate system so the system allows the players to move guitar while playing.

We utilize a particle filter [5] to track the finger markers in 3D space. We propagate sample particles in 3D space and project them onto the 2D image planes of both cameras to get the probability of each particle to be on finger markers based on color in both images.

To determine the color probabilities being finger markers color, during the preprocessing we apply a Bayesian classifier that is bootstrapped with a small set of training data and refined through an offline iterative training procedure [6, 7]. Online adaptation of markers-color probabilities is then used to refine the classifier using additional training images. Hence, the classifier is able to deal with illumination changes, even when there is a dynamic background.

In this way, the 3D positions of finger markers can be obtained, so that we can recognize if the fingers of player are pressing the strings or not. As a result, our system can determine the complete positions of all fingers on the guitar fret. It can be used to develop instructive software such as a chord tracker for the guitar learner. One of the example applications [8] is to identify whether the finger positions are correct and in accord with the finger positions required for the piece of music that is being played. Because guitar players can automatically identify whether their fingers are in the correct position, it would be an invaluable teaching aid for people learning to play the guitar.

2. Related Works

Related approaches for finger detection and tracking of guitarists will be described in this section. Cakmakci and Berard [9] detected the finger position by placing a small Augmented Reality Toolkit's (ARToolKit) [10] marker on a fingertip of the player for tracking the forefinger position (only one fingertip). However, when we attempted to use the ARToolKit's markers to all four fingertips, some markers' planes were not simultaneously perpendicular to the optical axis of camera(s) in some angles, especially while the player was pressing their fingers on the strings. Therefore, it was quite difficult to accurately track the positions of four fingers concurrently by using the ARToolKit finger markers.

Burns and Wanderley [11] detected the positions of fingertips for the retrieval of the guitarist fingering without markers. They assumed that the fingertip shape can be approximated with a semicircular shape while the rest of the hand is roughly straight and uses the circular Hough transform to detect fingertips. However, utilizing Hough transform to detect the fingertips when playing the guitar is neither accurate nor robust enough. This is because a fingertip shape does not appear as a circular shape in some angles. Also, in real-life performance, the lack of contrast between fingertips and background skin adds a complication.

In addition, these two methods [9, 11] use only one camera on the 2D image processing. The problem with using one camera is that it is very difficult to classify whether the fingers are pressing the strings or not. Therefore, stereo cameras are needed (3D image processing). However, when using these methods it is sometimes difficult to employ stereo cameras because all fingertips may not be perpendicularly captured by the two cameras simultaneously.

We therefore propose a method to overcome this problem by utilizing four colored markers placed on the four fingertips to determine the positions of the fingertips. However, a well-known current problem of color detection is the control of the lighting. Changing the levels of light and limited contrasts prevents correct registration, especially when there is a cluttered background. The survey of detecting faces in images [12] provides an interesting overview of color detection. A major decision toward deriving a model of color relates to the selection of the color space to be employed. Once a suitable color space has been selected, one of the commonly used approaches for defining what constitutes color is to employ bounds on the coordinates of the selected space. However, by using the simple threshold, it is sometimes difficult to accurately classify the color when the illumination changes.

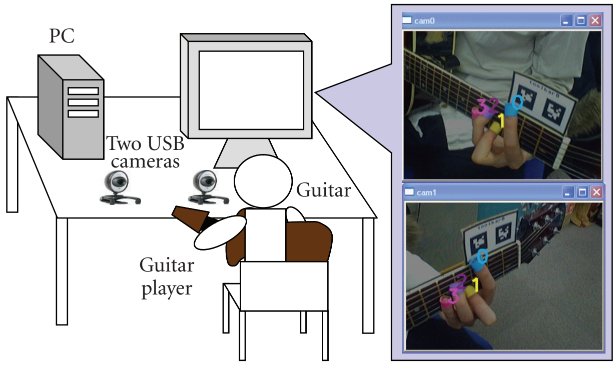


Figure 1: System configuration.

Therefore, we use a Bayesian classifier by learning color probabilities from a small training image set and then adaptively learn the color probabilities from online input images (proposed recently in [6, 7]). Applying this method, the first attractive property is that it can avoid the burden involved in the process of manually generating a lot of training data. From small amount of training data, it adapts the probability according to the current illumination and converges to a proper value. For this reason, the major advantage of using this method is its ability to cope with changing illumination because it can adaptively describe the distribution of the markers color.

3. System Configuration

The system configuration is shown in Figure 1. We use two USB cameras and a display connected to the PC for the guitar players. The two cameras capture the position of the left hand (assuming the guitarist is right-handed) and the guitar neck to obtain 3D information. We attach a 4.5 cm × 8 cm ARTag fiducial marker onto the top right corner of guitar neck to compute the position of the guitar (i.e., the poses of cameras relative to guitar position). The colored markers (with different color) are attached to the fingertips of the left hand.

4. Method

Figure 2 shows the schematic of the implementation. After capturing the images, we calculate the projection matrix in each frame by utilizing ARTag. We then utilize a Bayesian

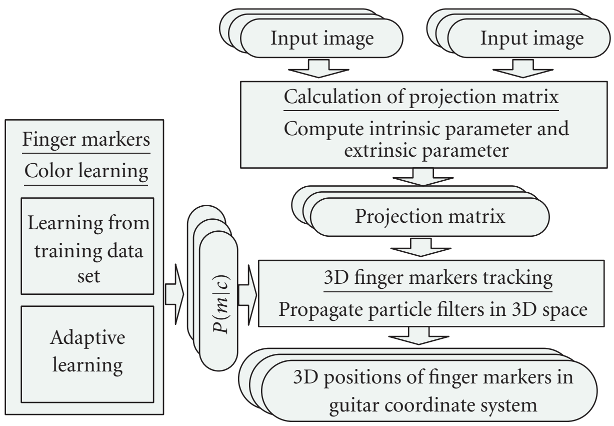


Figure 2: Method overview.

classifier to determine the color probabilities of the finger markers. Finally, we apply the particle filters to track the 3D positions of the finger markers.

4.1. Calculation of Projection Matrix

Detecting position of the fingers in captured images is the main point of our research, and the positions in images can give 3D positions based on stereo configuration of this system. Thus, it is necessary to calculate projection matrix (because it will be then used for projecting 3D particles to the image planes of both cameras in particle filtering step in Section 4.3). However, while the cameras are fixed, the guitar neck is not fixed to the ground, and therefore the projection matrix changes at every frame. Thus, we have to define the world coordinate system on the guitar neck as a guitar coordinate system (Figure 3). In the camera calibration process [13], the relation by projection matrix is generally employed as the method of describing the relation between the 3D space and the images. The important camera properties, namely, the intrinsic parameters that must be measured, include the center point of the camera image, the lens distortion, and the camera focal length. We first estimate intrinsic parameters during the offline step. As shown in (1), the matrices R and t in the camera calibration matrix describe the position and orientation of the camera with respect to world coordinate system. During online process, extrinsic parameters are then estimated in every frame by utilizing ARTag functions. Therefore, we can compute the projection matrix, P, by using

$$P = A[R,t] = \begin{bmatrix} \alpha_u & -\alpha_u \cot\theta & u_0 \\ 0 & \dfrac{\alpha_v}{\sin\theta} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{21} & R_{31} & t_x \\ R_{12} & R_{22} & R_{32} & t_y \\ R_{13} & R_{23} & R_{33} & t_z \end{bmatrix}, \quad (1)$$

where A is the intrinsic matrix, $[R,t]$ is the extrinsic matrix, u_0 and v_0 are the center point of the camera image, θ is the lens distortion, and α_u and α_v represent the focal lengths.

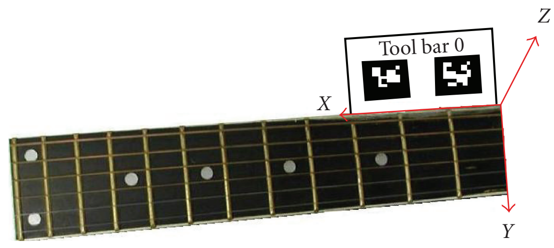


Figure 3: The world coordinate system on the top right corner of guitar neck.

4.2. Finger Markers Color Learning

This section will explain the method we used for calculating the color probabilities being finger markers color which will be then used in the particle filtering step (Section 4.3).

The learning process is composed of two phases. In the first phase, the color probability is learned from a small number of training images during an offline preprocess. In the second phase, we gradually update the probability from the additional training data images automatically and adaptively. The adapting process can be disabled as soon as the achieved training is deemed sufficient.

Therefore, this method will allow us to get accurate color probabilities of the finger markers from only a small set of manually prepared training images. This is because the additional marker regions do not need to be segmented manually. Also, because of the adaptive learning, it can be used robustly with changing illumination during the online operation.

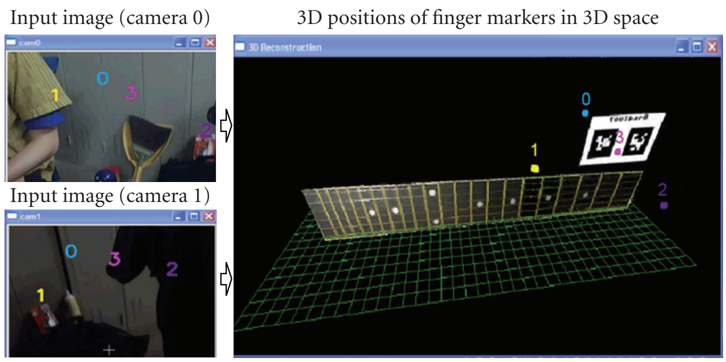
4.2.1. Learning from Training Data Set

During an offline phase, a small set of training input images (20 images) is selected on which a human operator manually segments markers-colored regions. The color representation used in this process is YUV 4:2:2 [14]. However, the Y-component of this representation is not employed for two reasons. Firstly, the Y-component corresponds to the illumination of an image pixel. By omitting this component, the developed classifier becomes less sensitive to illumination changes. Secondly, compared to a 3D color representation (YUV), a 2D color representation (UV) is lower in dimensions and, therefore, less demanding in terms of memory storage and processing costs.

Assuming that image pixels with coordinates (x, y) have color values $c = c(x, y)$, training data are used to calculate the following.

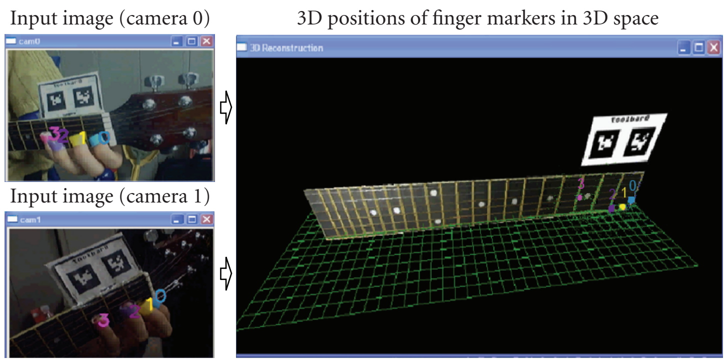
(i) The prior probability $P(m)$ of having marker m color in an image: this is the ratio of the marker-colored pixels in the training set to the total number of pixels of whole training images.

(ii) The prior probability $P(c)$ of the occurrence of each color in an image: this is computed as the ratio of the number of occurrences of each color c to the total number of image points in the training set.



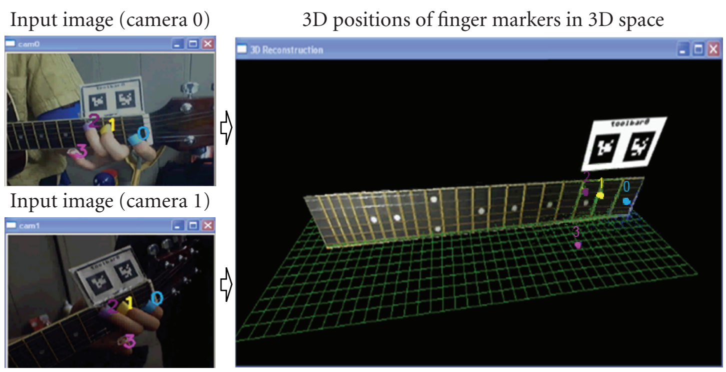
Initial stage: frame 10 (no guitar and no fingers in the scene)

(a)



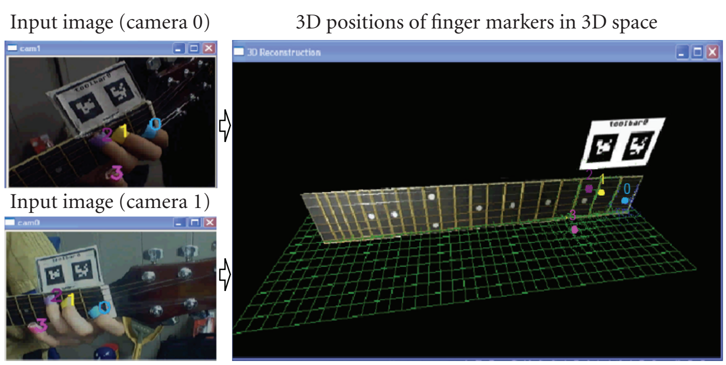
Playing stage: frame 50

(b)



Playing stage: frame 80

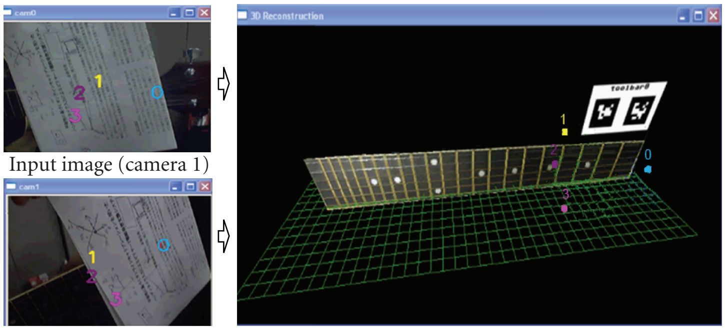
(c)



Playing stage: frame 110

(d)

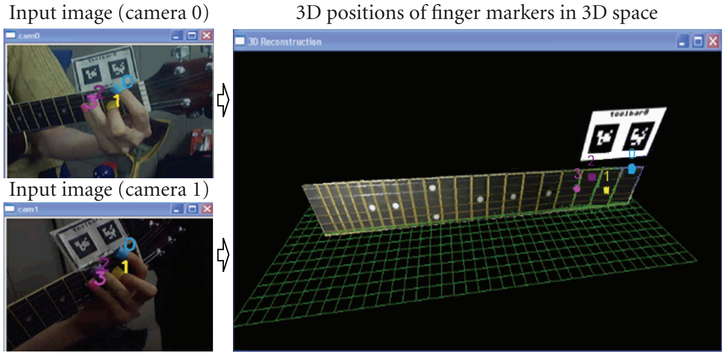
Input image (camera 0) 3D positions of finger markers in 3D space



Occlusion stage (tracking fails-back to initial stage): frame 150

(e)

Input image (camera 0) 3D positions of finger markers in 3D space



Recovering stage (playing stage again): frame 180

(f)

Figure 4: Representative snapshots from the online tracking experiment.

(iii) The conditional probability $P(c \mid m)$ of a marker being color c: this is defined as the ratio of the number of occurrences of a color c within the marker-colored areas to the number of marker-colored image points in the training set.

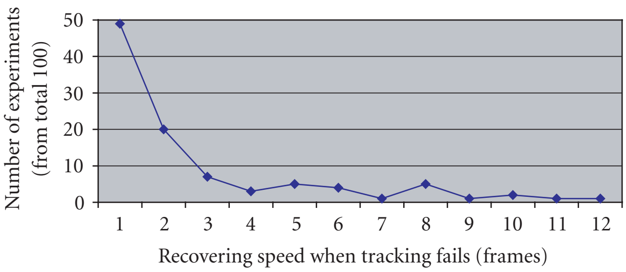
By employing Bayes' rule, the probability $P(m \mid c)$ of a color c being a marker color can be computed by using

$$P(m \mid c) = \frac{P(c \mid m)P(m)}{P(c)}. \quad (2)$$

This equation determines the probability of a certain image pixel being marker-colored using a lookup table indexed with the pixel's color. The resultant probability map thresholds are then set to be threshold $T_{\max}$ and threshold $T_{\min}$, where all pixels with probability $P(m \mid c) > T_{\max}$ are considered as being marker colored—these pixels constitute seeds of potential marker-colored blobs—and image pixels with probabilities $P(m \mid c) > T_{\min}$ where $T_{\min} < T_{\max}$ are the neighbors of marker-colored image pixels being recursively added to each color blob. The rationale behind this region growing operation is that an image pixel with relatively low probability of being marker colored should be considered as a neighbor of an image pixel with high probability of being marker colored. The values for $T_{\max}$ and $T_{\min}$ should be determined by test experiments (we use 0.5 and 0.15, resp., in the experiment in this paper). A standard connected component labelling algorithm is then responsible for assigning different labels to the image pixels of different blobs. Size filtering on the derived connected components is also performed to eliminate small isolated blobs that are attributed to noise and do not correspond to interesting marker-colored regions. Each of the remaining connected components corresponds to a marker-colored blob.

4.2.2. Adaptive Learning

The success of the marker-color detection depends crucially on whether or not the illumination conditions during the online operation of the detector are similar to those during the acquisition of the training data set. Despite the fact that using the UV color representation model has certain illumination independent characteristics, the marker-color detector may produce poor results if the illumination conditions during online operation are considerably different to those used in the training set. Thus, a means for adapting the representation of marker-colored image pixels according to the recent history of detected colored pixels is required. To solve this problem, marker-color detection maintains two sets of prior probabilities. The first set consists of $P(m)$, $P(c)$, $P(c \mid m)$ that have been computed offline from the training set. The second



Mean recovering speed (frames)	2.75
Standard derivation	2.64

FIGURE 5: Speed used for recovering from tracking failures.

is made up of $P_W(m)$, $P_W(c)$, $P_W(c \mid m)$ corresponding to $P(m)$, $P(c)$, $P(c \mid m)$ that the system gathers during the W most recent frames, respectively. Obviously, the second set better reflects the "recent" appearance of marker-colored objects and is therefore better adapted to the current illumination conditions. Marker-color detection is then performed based on the following weighted moving average formula:

$$P_A(m \mid c) = \gamma P(m \mid c) + (1-\gamma) P_W(m \mid c), \tag{3}$$

where γ is a sensitivity parameter that controls the influence of the training set in the detection process, $P_A(m \mid c)$ represents the adapted probability of a color c being a marker color, and $P(m \mid c)$ and $P_W(m \mid c)$ are both given by (2) but involve prior probabilities that have been computed from the whole training set [for $P(m \mid c)$] and from the detection results in the last W frames [for $P_W(m \mid c)$]. In our implementation, we set $\gamma = 0.8$ and $W = 5$.

Thus, the finger markers-color probabilities can be determined adaptively. By using online adaptation of finger markers-color probabilities, the classifier is able to easily cope with considerable illumination changes and also a dynamic background (e.g., moving guitar neck).

4.3. 3D Finger Markers Tracking

Particle filtering [5] is a useful tool to track objects in a clutter, with the advantages of performing automatic track initialization and recovering from tracking failures. In this paper, we apply particle filters to compute and track the 3D position of finger markers in the guitar coordinate system (the 3D information is used to help determine whether fingers are pressing a guitar string or not). The finger markers can then be automatically tracked, and the tracking can be recovered from the failures.

The particle filtering (system) uniformly distributes particles all over the area in 3D space and then projects the particles from 3D space onto the 2D image planes of the two cameras to obtain the probability of each particle to be finger markers. As new information arrives, these particles are continuously reallocated to update the position estimate. Furthermore, when the overall probability of particles to be finger markers is lower than the threshold we set, the new sample particles will be uniformly distributed all over the area in 3D space. Then the particles will converge to the areas of finger markers. For this reason, the system is able to recover the tracking. (The calculation is based on the following analysis.)

Given that the process at each time step is an iteration of factored sampling, the output of an iteration will be a weighted, time-stamped sample set, denoted by $\{s_t^{(n)}, n = 1, \ldots, N\}$ with weights $\pi_t^{(n)}$, representing approximately the probability-density function $p(X_t)$ at time t: where N is the size of sample sets, $s_t^{(n)}$ is defined as the position of the nth particle at time t, X_t represents the position in 3D of finger marker at time t, and $p(X_t)$ is the probability that a finger marker is at 3D position $X = (x, y, z)^T$ at time t.

The number of particles we used, N, is 900 particles in each marker. The dimensions of the 3D space we used in the particle filter are 20 cm, 15 cm, and 7 cm for x-axis, y-axis, and z-axis respectively, that is, $x \in [-3, 17]$, $y \in [-7, 8]$, and $z \in [-2, 5]$. Each axis and the origin of the world coordinate space are depicted in Figure 3. This size of 3D space is determined by measuring from the size of actual guitar we used. The size should cover the area of guitar neck in which the fingers of a player will hold a chord. If the size of 3D space is too small, the finger markers cannot be successfully tracked whenever the markers leave the scope of the defined 3D space. However, if the size is too large, the accuracy of tracking will decline. For this reason, the suitable size of 3D space should be determined.

The iterative process can be divided into three main stages:

(i) selection stage,

(ii) predictive state,

(iii) measurement stage.

In the first stage (the selection stage), a sample $s_t'^{(n)}$ is chosen from the sample set $\{s_{t-1}^{(n)}, \pi_{t-1}^{(n)}, c_{t-1}^{(n)}\}$ with probabilities $\pi_{t-1}^{(j)}$, where $c_{t-1}^{(n)}$ is the cumulative weight. This is done by generating a uniformly distributed random number $r \in [0, 1]$. We find the smallest j for which $c_{t-1}^{(j)} \geq r$ using binary search, and then $s_t'^{(n)}$ can be set as follows: $s_t'^{(n)} = s_{t-1}^{(j)}$.

Each element chosen from the new set is now subjected to the second stage (the predictive step). We propagate each sample from the set s'_{t-1} by a propagation function, $g(s_t'^{(n)})$, using

$$s_t^{(n)} = g\left(s_t'^{(n)}\right) + \text{noise}, \tag{4}$$

where noise is given as a Gaussian distribution with its mean $= (0, 0, 0)^T$. The value for the variance of the Gaussian distribution of noise should be determined by test experiments (we use the variance of Gaussian distribution $= 12\,\text{cm}^2$ in the experiment in this paper). Specifically, the particles should be distributed covering the area that the finger markers will move in the consecutive frame. Therefore, if the variance value is too low, the tracker will easily fail

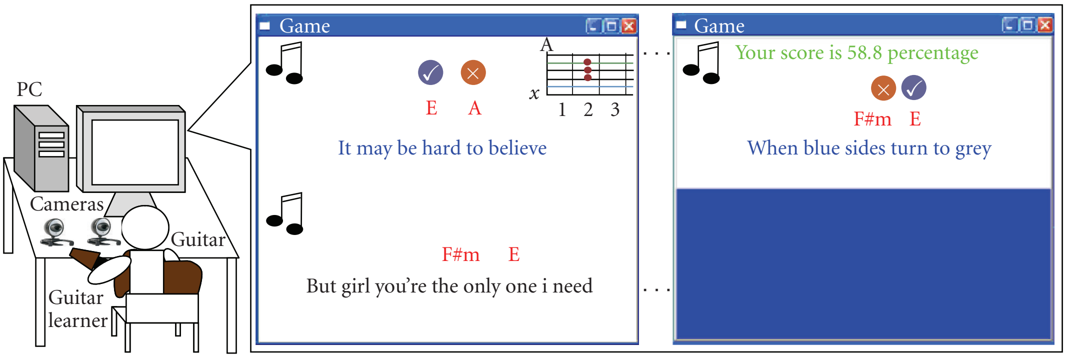


Figure 6: User interface of our guitar application.

to track and need to recover too frequently. On the other hand, if the variance is too high, the accuracy of tracking will decrease. Therefore, a suitable value of variance should be determined by test experiments.

Also, the accuracy of the particle filter depends on this propagation function. We have tried different propagation functions (e.g., constant velocity motion model and acceleration motion model), but our experimental results have revealed that using only noise information gives the best result. A possible reason is that the motions of finger markers are usually quite fast and constantly changing directions while playing the guitar. Therefore, the calculated velocities or accelerations in previous frame do not give accurate prediction of the next frame. In this way, we use only the noise information by defining $g(x) = x$ in (4).

In the last stage (the measurement stage), we project these sample particles from 3D space to two 2D image planes of cameras using the projection matrix results from (1). We then determine the probability whether the particle is on finger marker. In this way, we generate weights from the probability-density function $p(X_t)$ to obtain the sample-set representation $\{(s_t^{(n)}, \pi_t^{(n)})\}$ of the state density for time t using

$$\begin{aligned}\pi_t^{(n)} &= p\left(X_t = s_t^{(n)}\right) \\ &= P_A(m \mid c)_{\text{Camera}\,0} \times P_A(m \mid c)_{\text{Camera}\,1},\end{aligned} \tag{5}$$

where $p(X_t = s_t^{(n)})$ is the probability that a finger marker is at position $s_t^{(n)}$.

We assign the weights to be the product of $P_A(m \mid c)$ of two cameras which can be obtained by (3) from the finger markers color learning step (the adapted probabilities $P_A(m \mid c)_{\text{Camera}\,0}$ and $P_A(m \mid c)_{\text{Camera}\,1}$ represent a color c being a marker color in camera 0 and camera 1, resp.). Following this, we normalize the total weights using the condition

$$\Sigma_n \pi_t^{(n)} = 1. \tag{6}$$

(a) (b)

Figure 7: The system environment while users were testing our system.

Next, we update the cumulative probability, which can be calculated from normalized weights using

$$c_t^{(0)} = 0, \quad c_t^{(n)} = c_t^{(n-1)} + \pi_t^{(n)}{}_{\text{Total}} \quad (n = 1, \dots, N), \tag{7}$$

where $\pi_t^{(n)}{}_{\text{Total}}$ is the total weight.

Once the N samples have been constructed, we estimate moments of the tracked position at time step t as using

$$\varepsilon[f(X_t)] = \Sigma_{n=1}^{N} \pi_t^{(n)} s_t^{(n)}, \tag{8}$$

where $\varepsilon[f(X_t)]$ represents the centroid of each finger marker. The four finger markers can then be tracked in 3D space, enabling us to perform automatic track initialization and track recovering even in a dynamic background. The positions of four finger markers in the guitar coordinate system can be obtained.

5. Tracking Results

Representative results from our experiment are shown in this section. Figure 4 provides a few representative snapshots of the experiment. The reported experiment is based on a sequence that has been acquired. Two USB cameras with resolution 320×240 have been used.

The camera 0 and camera 1 windows depict the input images which are captured from two cameras. These cameras capture the player's fingers in the left hand positioning and the guitar neck from two different views. For visualization purposes, the 2D tracked result of each finger marker is also shown in camera 0 and camera 1 windows. The four colored numbers depict four 2D tracking results from the finger markers (forefinger [number0—light blue], middle finger [number1—yellow], ring finger [number2—violet] and little finger [number3—pink]).

The 3D reconstruction window, which is drawn using OpenGL, represents the tracked 3D positions of the four finger markers in guitar coordinate system. In this 3D space, we show the virtual guitar board to make it clearly understood that this is the guitar coordinate system. The four-color 3D small cubes show each 3D tracked result of the finger markers (these four 3D cubes correspond to the 2D four colored numbers in the camera 0 and the camera 1 windows).

In the initial stage (frame 10), when the experiment starts, there are no guitar and no fingers in the scene. The tracker attempts to find the color which is similar to the markers-colored region (i.e., the particle filter tries to find objects whose color is the most similar to the color of our markers in the background). For example, because the color of player's shirt (light yellow) is similar to a middle finger marker's color (yellow), the 2D tracking result of middle finger marker (number1) in the camera 0 window detects wrongly as if the player's shirt is the middle finger marker.

However, later during the playing stage (frame 50), the left hand of a player and the guitar enter the fields of cameras' views. The player is playing the guitar, and then the system can closely determine the accurate 3D fingertip positions which correspond to the 2D colored numbers in the camera 0 and the camera 1 windows. In this way, the system can perform automatic track initialization because it is using particle filtering.

Next, the player changes to next fingering positions in frame 80. The system can continue to correctly track and recognize the 3D fingering positions which correspond nearly to the positions of 2D colored numbers in the camera 0 and the camera 1 windows.

Following this, the player moves the guitar position (from the old position in frame 80) to the new position in frame 110 but still holds the same fingering positions on the guitar fret. It can be observed that the detected 3D positions of the four finger markers from different guitar positions (i.e., but the same input fingering on the guitar fret) are almost the same positions. This is because ARTag marker is used to track the guitar position.

Later on, in the occlusion stage (frame 150), the finger markers are totally occluded by the white paper. Therefore, the system is again back to find the similar colors of each marker (returning to the initial stage again). Also, when the ARTag marker cannot be found in the scene, the system cannot determine the accurate 3D tracking results in the guitar coordinate space. This is because the projection matrix cannot be obtained correctly.

However, following this in the recovering stage (frame 180), the occlusion of white paper is moved out, and then the cameras are capturing the fingers and guitar neck again. It can be seen that the tracker can return to track the correct fingerings (returning to the playing stage again). In other words, the system is able to recover from tracking failure because it is using particle filtering.

The reader is also encouraged to observe the illumination difference between camera 0 and camera 1 windows. Our experimental room has two main light sources which are located opposite to each other. We turned on the first light source of the room which is quite near for capturing images on camera 0, while we turned off the second light source (which is opposite to the first source) and is located near camera 1. Hence, the lighting used to test in each camera is different. However, it can be observed that the 2D tracked result of finger markers can be still determined without effects of different light sources in both camera 0 and camera 1 windows in each representative frame. This is because a Bayesian classifier and an online adaptation of color probabilities are utilized to deal with this.

We also evaluate the recovering speed whenever tracking the finger markers fails. Figure 5 shows the speeds used for recovering from lost tracks. In this graph, the recovering speeds are counted from initial frame where certainty of tracking is lower than threshold. At the initial frame, the particles will be uniformly distributed all over the 3D space as described in Section 4.3. Before normalized weights in particle filtering step, we determine the certainty of tracking from the sum of the weight probability of each distributed particle to be marker. Therefore, if the sum of weight probability is lower than the threshold, we assume that the tracker is failing. On the other hand, if the sum of weight probability is higher than threshold, we imply that the tracking has been recovered. Thus, the last counted frame will be decided at this frame (the particles have been already converged to the areas of finger markers). The mean recovering speed and the standard derivation are also shown in the table in Figure 5, in frames. We believe this recovering speed is fast enough for recovering of tracking in real-life guitar performance.

Finally, we will refer to a limitation of the proposed system. The constraint of our system is that, although the background we used can be cluttered, the background should not be composed of large objects which are the same color as the colors of finger markers. For instance, if the players wear their clothes which are of a very similar color to the markers' colors, the system cannot sometimes determine the output correctly.

6. Sample Application

Learning to play the guitar usually involves tedious lessons in fingering positions for the left hand. It is difficult for beginners to recognize by themselves whether they are accurately positioning their fingers on the string. For this reason, we developed the application of human-computer interaction

Table 1: Qualitative feedback from participants.

Criteria	Subjective feedback (number of respondents)				
	Excellent	Good	Fair	Poor	Total
Interest in application	14	1	—	—	15
System smoothness	8	5	2	—	15
User satisfaction	12	3	—	—	15
Ease of use	7	7	1	—	15
System naturalness (comfortable feeling)	3	5	6	1	15
Overall	10	4	1	—	15

(HCI) to guitar playing, named *InteractiveGuitarGame*, by applying the guitarist fingertip tracking method presented in this paper. This application aims to assist guitar learners by automatically identifying whether the fingertip positions are correct and in accord with the fingertip positions required for the piece of music that is being played.

6.1. Description of the Application

The *InteractiveGuitarGame* application recognizes the guitar chord being played by a guitar learner based on the 3D position of finger markers in the virtual guitar coordinate spaces. Then, the system gives real-time feedback to guitar learners telling them if they are using the correct chords required by the musical piece.

An example user interface of the application is shown in Figure 6. This application contains the lyrics, guitar chord charts, and vocal information. The lyrics are displayed on the screen in color which changes and is synchronized with the music. The guitar chord charts are shown in the top right of the user interface that can be used to guide student guitarists to play a song. Most importantly, this application recognizes if each successive chord contained in the music is being played correctly. It is considered that this would be of a great assistance to guitar learners because they are able to automatically identify if their own finger positions are correct and whether they are matching the correct chords required by the musical piece. This feedback is shown by small right/wrong symbols above the corresponding chords in real time. Finally, this application will show an overall evaluation score, as a percentage, indicating the user's accuracy when the performance has been completed which provides greater user friendliness. This application would be invaluable as a teaching aid for guitar learners.

6.2. Evaluation of the Application

We conducted a user study to evaluate the effectiveness of the aforementioned application. Fifteen users (nine inexperienced and six experienced guitar players) were asked to test our system. Each user took approximately 5–10 minutes to run the study. All users were able to use the system after a short explanation. After the individual tests, the users were asked to give qualitative feedback (Table 1). This included interest in the application, smoothness of the system, user satisfaction, ease of use of the interface, naturalness of the system, and overall impression. General comments on the test were also collected from the users. The system environment while users were testing our system is shown in Figure 7.

From this user study, we received mostly positive comments from users. Many users agreed that it is useful for learning to play the guitar. Also, they were satisfied with the smoothness and the speed of our system. They reasoned that this was because the system can run in real time and was synchronized with the real music. Moreover, many users indicated that they were impressed by the system, especially by the idea of developing this application. They gave the reason that this was their first time to see how a computer using cameras can give a feedback to players interactively for an actual stringed guitar. Interestingly, several users seemed to be more enthusiastic with the overall evaluation score when the song finished playing. They gave their reason since this made it more interesting as a real guitar game. Hence this evaluation score could attract learners to learn and enjoy the guitar lessons in music school.

However, the most common complaint was about finger markers. Some users indicated that, although they preferred the idea of this application, it was slightly difficult to use because colored markers were placed on the fingertips. In other words, since fingertip markers were required, some users felt less comfortable using our system. Thus the system seemed to be slightly unnatural. This was the reason why they gave lower scores to system naturalness (comfortable feeling), as presented in Table 1.

7. Conclusions

In this paper, we have developed a system that measures and tracks the positions of the fingertips of a guitar player accurately in the guitar's coordinate system. A framework for colored finger markers tracking has been proposed based on a Bayesian classifier and particle filters in 3D space. ARTag has also been utilized to calculate the projection matrix. This implementation can be used to develop instructive software such as a chord tracker for a guitar learner.

Although we believe that we can successfully produce an accurate system output, the current system has the limitation about the colored finger markers because finger markers are required in our current system. This sometimes makes it unnatural for playing guitar in real life. As future work, we intend to make technical improvements to further refine the problem of the finger markers by removing these markers which may result in even greater user friendliness.

Acknowledgments

This work is supported in part by a Grant-in-Aid for the Global Center of Excellence for High-Level Global Cooperation for Leading-Edge Platform on Access Spaces from the Ministry of Education, Culture, Sport, Science, and Technology in Japan.

References

[1] T. Maki-Patola, J. Laitinen, A. Kanerva, and T. Takala, "Experiments with virtual reality instruments," in *Proceedings of the 5th International Conference on New Interfaces for Musical Expression (NIME '05)*, pp. 11–16, Vancouver, Canada, May 2005.

[2] F. Liarokapis, "Augmented reality scenarios for guitar learning," in *Proceedings of the 3rd International Conference on Eurographics UK Theory and Practice of Computer Graphics*, pp. 163–170, Canterbury, UK, May 2005.

[3] Y. Motokawa and H. Saito, "Support system for guitar playing using augmented reality display," in *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '06)*, pp. 243–244, Santa Barbara, Calif, USA, October 2006.

[4] M. Fiala, "ARTag, a fiducial marker system using digital techniques," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 590–596, San Diego, Calif, USA, June 2005.

[5] M. Isard and A. Blake, "CONDENSATION—conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[6] A. A. Argyros and M. I. A. Lourakis, "Tracking skin-colored objects in real-time," in *Cutting Edge Robotics*, pp. 77–90, Advanced Robotic Systems International, Vienna, Austria, 2005.

[7] A. A. Argyros and M. I. A. Lourakis, "Tracking multiple colored blobs with a moving camera," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, p. 1178, San Diego, Calif, USA, June 2005.

[8] C. Kerdvibulvech and H. Saito, "Real-time guitar chord estimation by stereo cameras for supporting guitarists," in *Proceedings of the 10th International Workshop on Advanced Image Technology (IWAIT '07)*, pp. 256–261, Bangkok, Thailand, January 2007.

[9] O. Cakmakci and F. Berard, "An augmented reality based learning assistant for electric bass guitar," in *Proceedings of the 10th International Conference on Human-Computer Interaction (HCI '03)*, Crete, Greece, June 2003.

[10] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," in *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality*, pp. 85–94, San Francisco, Calif, USA, October 1999.

[11] A. M. Burns and M. M. Wanderley, "Visual methods for the retrieval of guitarist fingering," in *Proceedings of the 6th International Conference on New Interfaces for Musical Expression (NIME '06)*, pp. 196–199, Paris, France, June 2006.

[12] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[13] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ, USA, 2003.

[14] K. Jack, *Video Demystified*, Elsevier Science, Oxford, UK, 2004.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 145363, 14 pages
doi:10.1155/2008/145363

Research Article

Visual Enhancement for Sports Entertainment by Vision-Based Augmented Reality

Yuko Uematsu and Hideo Saito

School of Science for Open and Environmental Systems, Graduate School of Science and Technology, Keio University, 3-14-1 Hiyoshi, Koyoku-ku, Yokohama 223-8522, Japan

Correspondence should be addressed to Yuko Uematsu, yu-ko@ozawa.ics.keio.ac.jp

Received 1 October 2007; Revised 26 June 2008; Accepted 6 July 2008

Recommended by Adrian Cheok

This paper presents visually enhanced sports entertainment applications: AR Baseball Presentation System and Interactive AR Bowling System. We utilize vision-based augmented reality for getting immersive feeling. First application is an observation system of a virtual baseball game on the tabletop. 3D virtual players are playing a game on a real baseball field model, so that users can observe the game from favorite view points through a handheld monitor with a web camera. Second application is a bowling system which allows users to roll a real ball down a real bowling lane model on the tabletop and knock down virtual pins. The users watch the virtual pins through the monitor. The lane and the ball are also tracked by vision-based tracking. In those applications, we utilize multiple 2D markers distributed at arbitrary positions and directions. Even though the geometrical relationship among the markers is unknown, we can track the camera in very wide area.

1. Introduction

Augmented reality (AR) is a technique for overlaying virtual objects onto the real world. AR has recently been applied to many kinds of entertainment applications by using vision-based-tracking technique, such as [1–3]. AR can provide users with immersive feeling by allowing the interaction between the real and virtual world.

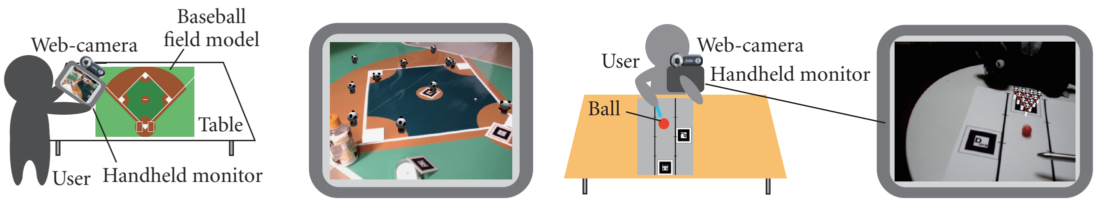
In these kinds of AR entertainment applications, virtual objects (world) generated with computer graphics are overlaid onto the real world. This means that the real 3D world is captured by a camera and the virtual objects are superimposed onto the captured images. By seeing the real world through some sort of displays, the users find that the virtual world is mixed with the real world and they can control the virtual world as well as the real world.

In such AR applications, the users carry a camera and move around the real world. Therefore, the pose and the position of the moving user's camera should be obtained so that the virtual objects can be appeared at correct position in the real world captured with the camera. Such camera tracking should also be performed in real time for interactive operations of the AR applications.

Vision-based camera tracking for such AR application is one of the popular research areas because the vision-based method does not require any special device except cameras, in contrast with sensor-based approach. For making the vision-based tracking robust and running in real time, marker-based approach is a reasonable solution, so we focus on marker-based approach. Especially, "AR-Toolkit" [4], which is using 2D square marker for the camera tracking, is a very popular tool for simple online AR applications that follow a marker-based approach. The camera's position and pose can be estimated in real time by using the 2D square markers.

This paper presents two AR applications: AR Baseball Presentation System and Interactive AR Bowling System. Users use these applications on the tabletop in the real world by using a web camera attached to a handheld monitor as shown in Figure 1.

AR Baseball Presentation System is an observation system of a virtual baseball game. Users place a real baseball field model on the tabletop and input a baseball game history (scorebook) that they want to watch into the system. Then, they can watch the game by replaying with virtual baseball



(a) AR Baseball Presentation System

(b) Interactive AR Bowling System

FIGURE 1: Our proposed AR applications.

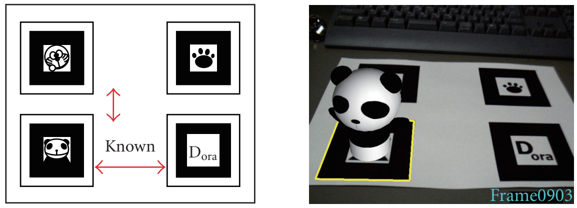
players on the field model in front of them. On the field model, 2D markers are placed for registration of the virtual players. Therefore, the users can watch the game from their favorite viewpoints around the field. This system focuses on visualizing the game by using a scorebook data, so that the user can understand the main game point. Thus, the detail part of the game such as player's gesture is not replayed. In this system, the virtual players are generated as a cartoon character instead of a human player. There is a limit even if a human player is made in great detail by CG. An increase in the number of polygon in CG is also big problem for real-time processing. Therefore, we decide to use a funny and friendly character.

With interactive AR Bowling System, users can enjoy the bowling game by rolling a real ball down a real bowling lane model placed on a tabletop in the real world. On the lane model, there are virtual pins generated with CG. They knock down the virtual pins by rolling the real ball. Touching and rolling the real ball provide a sort of tangible feeling in this system. It is well known that a tangible interface enhances the reality of communication [5–8]. Because of placing some markers on the lane model, the users can watch the lane and pins from free view points.

For registration of virtual objects such as the virtual players or the virtual pins, the motion of the user's camera is estimated by multiple 2D markers. In our applications, the multiple markers can be arranged at arbitrary positions and directions in the real world. Therefore, the users can start our applications only with free-arranging the markers. These applications are based on the multiple marker-based online AR system [9].

2. Related Work of Marker-Based AR Applications

There are some related works using 2D markers for AR applications. Henrysson et al. have proposed the AR tennis application which is used on the tabletop tennis court model [10]. On the tabletop tennis court, a few 2D markers are drawn for estimating the motion of user's camera. In their application, since a ball is a virtual object and user's position does not move a lot, the 2D markers are easily detected by the user's camera. On the other hand, in our baseball application, the users move around the baseball field to watch the baseball game from favorite view points. Therefore, a lot of markers should be arranged in the real world. Moreover, in our bowling application, since the ball is a real object, the markers may be occluded by the rolling ball. Therefore, the markers should be arranged not only on the table plane but also in various directions.



(a) Multiple markers

(b) Overlaid virtual object

FIGURE 2: Usual multiple marker-based registration of a virtual object.

Multiple markers are usually used by aligning themselves at measured intervals as shown in Figure 2 because the geometrical relationship of the multiple markers must be known [11–15]. In [15], they need the position and pose of a square marker and the position of a point marker in advance. In [14], they proposed marker-less registration method by settingup a learning process. In the learning process, however, the markers' geometrical information is required for learning the markers. In most cases, the task for the measurement of such information is implemented manually. However, this task is very time-consuming and not sufficiently accurate. Kotake et al. [16] proposed a marker-calibration method combining multiple planar markers with bundle adjustment. Although they do not require a precise measurement of the markers, they need a priori knowledge of the markers such as qualitative information to compute markers' geometrical information from a set of images by bundle adjustment, that is, multiple markers are coplanar.

In contrast, our registration method can freely distribute the multiple markers at arbitrary positions and directions. The geometrical relationship of the markers can be automatically estimated by constructing a 3D projective space which is defined by projective reconstruction of reference images. Through the projective space, then the geometrical arrangement of the marker planes is recovered in 3D. Therefore, we need not to manually measure the distance between the markers in advance. This algorithm is quite suitable for the AR applications when the users move around the real world and the markers may be occluded.

In this paper, we explain the algorithm of registration with multiple markers as described in Section 3. Then, AR Baseball Presentation System and AR Bowling System are introduced in Sections 4 and 5, respectively.

3. Registration Using Multiple Markers

In this section, we explain the algorithm of the registration method in the Multimarker-Based Online AR System [9]. This algorithm is based on [17].

Figure 3 shows a flowchart of the registration method. This registration method can be divided into two stages. At the first stage, the geometrical relationship of the markers is automatically estimated. For the estimation, a 3D projective space, which is a 3D virtual space, is defined by projective reconstruction of two reference images. The reference images are automatically selected from some candidate images. In our registration method, the geometrical relationship of the markers is represented as a transformation matrix called $\mathbf{T}_\mathbf{i}^{\mathbf{WP}}$ which relates each marker i and the projective space. These transformation matrices are computed once in advance.

At the second stage, a projection matrix $\mathbf{P}_\mathbf{i}^{\mathbf{WI}}$ from each marker i to the input image is computed. Those projection matrices and the transformation matrices, which are computed at the first stage, are integrated into projection matrices $\mathbf{P}_\mathbf{i}^{\mathbf{PI}}$ by (1), respectively,

$$\mathbf{P}_\mathbf{i}^{\mathbf{PI}} = \mathbf{P}_\mathbf{i}^{\mathbf{WI}}\left(\mathbf{T}_\mathbf{i}^{\mathbf{WP}}\right)^{-1}. \tag{1}$$

These projection matrices are based on the marker i and project the projective space onto the input image. Moreover, those $\mathbf{P}_\mathbf{i}^{\mathbf{PI}}$ are integrated into one projection matrix $\mathbf{P}^{\mathbf{PI}}$ by least-square method. Then virtual objects described in the projective space coordinate system are overlaid onto the input image by using the integrated projection matrix. These processes of the second stage are performed at every frame.

3.1. 3D Projective Space

A 3D projective space is constructed for estimating the geometrical arrangement of multiple planes placed at arbitrary positions and poses. The projective space is defined by projective reconstruction of two images which are captured from two different view points and called reference images. As shown in Figure 4, a 3D space P-Q-R is defined as a 3D projective space, which is projected onto the reference image A and B by following equations:

$$\begin{aligned} [u_A v_A 1]^\top &\simeq \mathbf{P_A}[\mathrm{PQR1}]^\top, \\ [u_B v_B 1]^\top &\simeq \mathbf{P_B}[\mathrm{PQR1}]^\top, \end{aligned} \tag{2}$$

$$\mathbf{P_A} = [\mathbf{I} \mid \mathbf{0}], \qquad \mathbf{P_B} = \left[-\frac{[\mathbf{e_B}]_\times \mathbf{F_{AB}}}{\|\mathbf{e_B}\|^2} \mid \mathbf{e_B}\right], \tag{3}$$

where, $[u_A, v_A, 1]^\top$ and $[u_B, v_B, 1]^\top$ are homogeneous coordinates of 2D points in the reference images, and $[P, Q, R, 1]^\top$ is also homogeneous coordinates of a 3D point in the projective space. $\mathbf{F_{AB}}$ is a fundamental matrix from the image A to B, $\mathbf{e_B}$ is an epipole on the image B, and $[\mathbf{e_B}]_\times$ is the skew-symmetric matrix [18].

Since the projective space is defined by projective reconstruction of the reference images, the accuracy of $\mathbf{F_{AB}}$ is important and is depending on the combination of the reference images. In this system, two reference images which have most accurate $\mathbf{F_{AB}}$ are automatically selected. The details will be described in next section.

3.2. Automatic Selection of Reference Images

The projective space is defined by the projective reconstruction of two reference images. Therefore, the fundamental matrix between the reference images is important to construct the accurate projective space. We introduce automatic selection algorithm of the reference images. The detail is shown in Figure 5.

First, the object scene is captured for a few seconds by a moving camera. This image sequence in which all the markers should be included becomes the candidate of the reference image. When two images are selected from the candidate images, projection matrices based on the markers in the selected reference images are computed by using the algorithm of [4], where $\mathbf{P_{A_i}}$ and $\mathbf{P_{B_i}}$ are the projection matrices which project marker i to the selected reference image A and B, respectively. Using each pair of the projection matrices, a fundamental matrix based on marker i is computed as in the following equation:

$$\mathbf{F_{AB_i}} = [\mathbf{e_{B_i}}]_\times \mathbf{P_{B_i}} \mathbf{P_{A_i}^-}, \tag{4}$$

where $\mathbf{P_{A_i}^-}$ represents the pseudoinverse matrix of $\mathbf{P_{A_i}}$ [18]. Then, one fundamental matrix is selected as $\mathbf{F_{AB}}$ which has the smallest projection error:

$$\text{error} = \mathbf{m_B}^\top \mathbf{F_{AB_i}} \mathbf{m_A}, \tag{5}$$

where $\mathbf{m_A}$ and $\mathbf{m_B}$ are corresponding points in the selected reference images.

When a projective space is temporarily constructed by the selected $\mathbf{F_{AB}}$ from (3), $\mathbf{T}_\mathbf{i}^{\mathbf{WP}}$ between each marker i, and the projective space is computed. Then, $\mathbf{P}_\mathbf{i}^{\mathbf{PI}}$ are computed and integrated into one projection matrix $\mathbf{P}^{\mathbf{PI}}$. Then, we compare these two projected coordinates $\mathbf{x_i}$, $\mathbf{x_i'}$:

$$\mathbf{x_i} = \mathbf{P_{B_i}}\mathbf{X_W}, \qquad \mathbf{x_i'} = \left(\mathbf{P^{PI}}\mathbf{T}_\mathbf{i}^{\mathbf{WP}}\right)\mathbf{X_W}. \tag{6}$$

Although these two coordinates should be equal, if the combination of the two reference images is not reasonable, they will be different. In such a case, we return to the phase of selecting a pair of temporary reference images. We iterate these processes until every difference of $\mathbf{x_i}$ and $\mathbf{x_i'}$ based on plane i becomes smaller than a few pixels. In the experiments, we decide the threshold as 3 pixels.

Even if the number of markers increases, only the time of computing the transformation matrices ($\mathbf{F_{AB_i}}, \mathbf{P^{PI}}, \mathbf{T}_\mathbf{i}^{\mathbf{WP}}$) is increased and the computation time is very short. Therefore, it is not a time-consuming process. The times of iteration is mainly decided by the number of the candidate reference images. When using 100 candidate reference images, the processing time of selecting reference images using 8 markers also took around 60 seconds as well as using 4 markers.

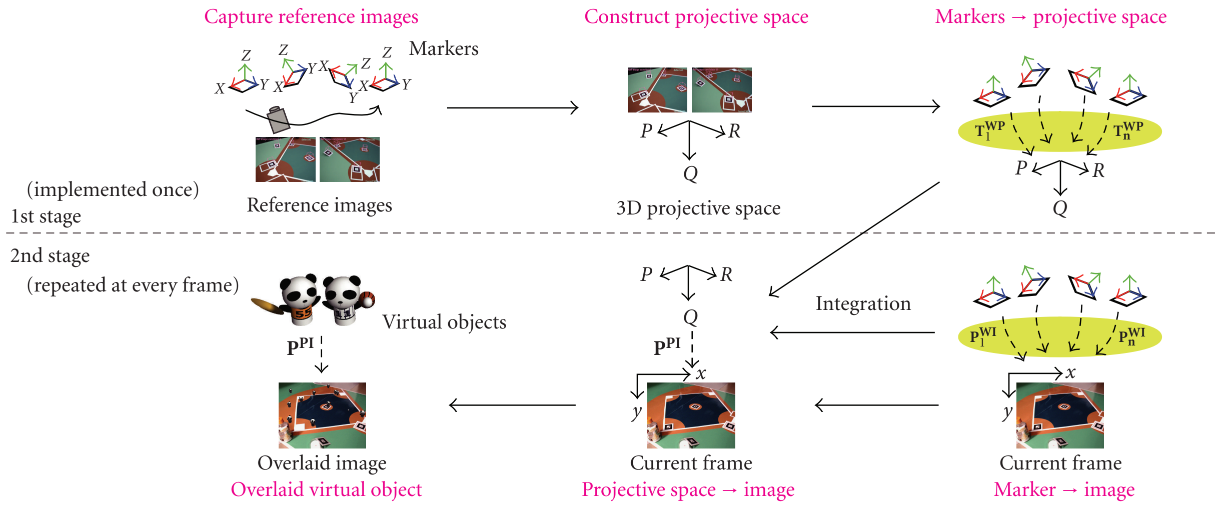


Figure 3: Overview of the registration method with 3D projective space.

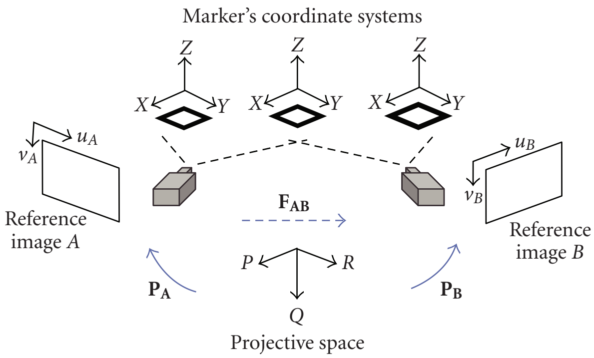


Figure 4: 3D projective space defined by projective reconstruction of the reference images.

4. AR Baseball Presentation System

AR Baseball Presentation System allows users to watch a virtual baseball game on the tabletop field model in the real world via a moving web camera attached to a handheld monitor. The virtual baseball game scene is synthesized with 3D CG players. These players are overlaid on the real field model. The users can interactively change their view points as their likes by applying the algorithm described in Section 3.

This system visually replays the baseball game which was previously played in the other place by a scorebook data, in which the game history they want to know is described. In contrast with the usual way to know the game history, such as watching the captured video or reading the recorded scorebook, our AR system can provide the users with much realistic sensation as an entertainment application.

4.1. Overview of Processing

Figure 6 shows overview of the system. Multiple 2D markers are distributed inside and outside of the baseball field model which is placed on the tabletop in the real world. The markers can be placed at arbitrary positions and poses without measuring the arrangement of them. The image of the tabletop field model is captured by a web camera attached and displayed on a handheld monitor.

This system can be divided into offline and online processes. At the offline process, first, a game history data file of a baseball game is prepared and loaded. In this file, history of game results are described play-by-play. Next, the field model is captured by the moving web camera for some seconds to automatically estimate the markers' arrangement. The detail of the algorithm is described in Section 3. These processes are executed once in advance.

At the online process, the three steps are repeated online: (1) synthesizing the baseball game scene while 1 play according the input data, (2) computing the camera's position and pose at the current frame, and (3) overlaying virtual players onto the field model. At the first step, when one line of the data file is read out, the positions of the players and the ball at every frame while 1 play are computed according to the data to render them on the field model. At the second step, the camera's rotation and translation are estimated using the markers in the current frame. At the final step, the virtual baseball scene, such as the players and the ball synthesized with CG, is overlaid onto the tabletop field model.

4.2. Input Scorebook Data File

The game played on the field model is the replayed game of the actual game which is synthesized according to input data file called "Scorebook Data File" (SDF). As shown in Figure 7, the game history of the actual game is described play-by-play in the SDF. "1 play" means the actions of the players and the ball from the moment that the pitcher throws the ball to the moment that the ball returns to the pitcher again. It is about for 15 to 30 seconds. The actions of the players and the ball in 1 play are described on one line in the SDF. The former part of the line represents the actions

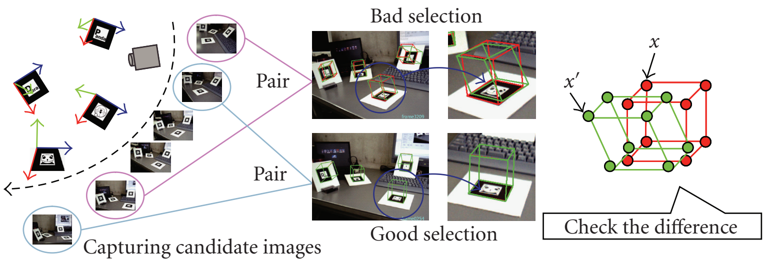


Figure 5: Automatic selection method of the reference images.

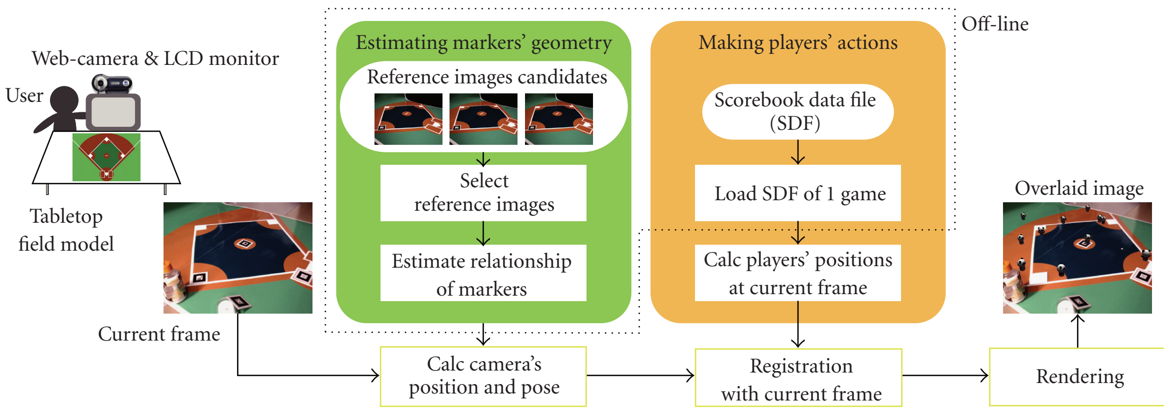


Figure 6: Overview of AR Baseball Presentation System.

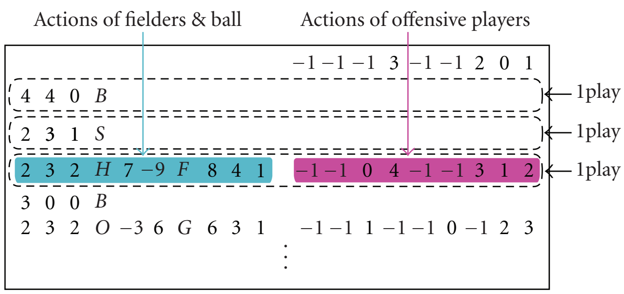


Figure 7: Scorebook Data File (SDF).

of the fielders and the ball, while the latter part describes the actions of the offensive players. This file is loaded in starting the system and is sequentially read out line-by-line at every 1 play. In this way, the actions of the baseball scene are described in the SDF.

4.3. Actions of Offensive Players

Offensive players indicate a batter, runners, and players who are waiting in the bench. Each player belongs to one of the six states as shown in Figure 8(a). The batter is in the batter's box, so its state is "0," third runner is "3," and the waiting players are "−1." In SDF, the destination state to which every player changes in each play is sequentially recorded. When one line of the file is read out, the destination of each player is decided according to the data as in Figure 8(b). Then, the game scene that 3D players are moving from the present state to the destination state while 1 play is created with CG.

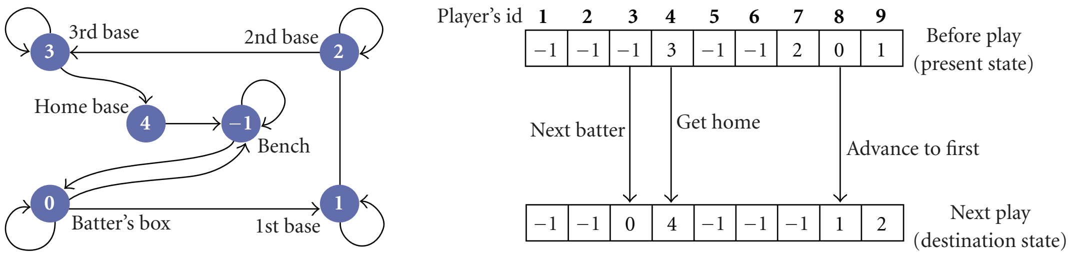
4.4. Actions of Fielders and Ball

In contrast to the offensive players who are just moving from present state to destination while 1 play, the fielders are doing some actions while 1 play, such as moving around the field and throwing and catching the ball, and so forth. Therefore, only the action of the ball is described in the SDF. Fielders move to catch the ball according to the action of the ball. The action of the ball while 1 play is described in Figure 9.

Fielders basically keep own positions. First, the ball is thrown by the pitcher and hit to the position which is described in part D of Figure 9. Then, the player whose position number is described in the first of part E moves to the position of part D to catch the ball. After catching the ball, the player throws the ball to the next player whose position number is described next. The next player moves to the nearest base and catches the ball. After the same iterations, finally, the ball is thrown to the pitcher.

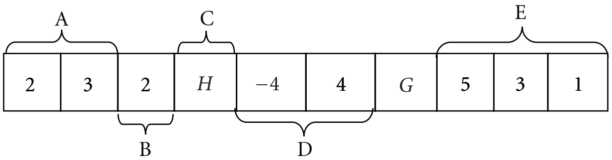
4.5. Demonstrations

We have implemented AR Baseball Presentation System with a web camera (ELECOM UCAM-E1D30MSV) attached to a handheld monitor connected a PC (OS:Windows



(a) State transition of offensive players (b) Example of Scorebook Data File for offensive players

Figure 8: Actions of the offensive players.



A: Pitching position of strike zone
B: Through/swing out/hit
C: Ball/strike/out/hit
D: Flying position on field model
E: Fielders' no. who catch the ball

Figure 9: Scorebook Data File of the fielders and the ball.

XP, CPU:Intel Pentium IV 3.2 GHz. The resolution of the captured image is 640 × 480 pixels. Multiple planar markers are distributed inside and outside the field model. In this case, one of the markers must be put on one of the bases in order to determine relationship between the field model and the markers. The other markers can be placed at arbitrary positions and poses. In these experiments, we use four markers and place one of them on the third base. A Scorebook Data File of a baseball game is manually prepared in accordance with Section 4.2. 3D models of virtual objects, such as players and a ball, are rendered with OpenGL (Algorithm 1).

First, the user places the baseball field model on the tabletop and distributes the markers. Next the object scene is captured with moving around the field model for 100 frames as candidates of the reference images. Inside of the system, the best pair of the reference images is automatically selected from the candidate images. Then, the projective space is constructed by the selected reference images. The geometrical relationship of the markers is also estimated. These automatic processes take about 60 seconds. After the automatic preparations, the user inputs a Scorebook Data File and starts the system. The virtual baseball game begins on the field model and the user can watch the game from favorite view point with moving around the real world.

Figure 10 presents a baseball game: team RED versus team WHITE. In this situation, team WHITE is in the field and team RED is at bat. The bases are loaded and 4th batter of team RED is in the batter's box (frame 0–15). The pitcher throws the ball (frame 15–29). The batter hits safely to left (frame 29–35), and then all runners move up a base (frame 50–89). In the result, team RED gets a score. In this experiment, frame rate of AR presentation is about 30 fps. Thus, user can see the baseball game at video rate.

Figure 11 shows some closeup views of the same scene. Since these images are captured from closeup view points, only a few markers can be captured in the field of view, and the captured markers are different in every frame. Even though particular markers are not continuously captured over the frames, the virtual players and the ball can be correctly registered onto the real tabletop field with the same world coordinate. This means that the consistency of the geometrical relationship between the camera and the virtual objects is kept properly although the geometrical arrangement of the markers is unknown.

In Figure 12, the angle of the camera with respect to the tabletop is too small to detect the markers lying on the tabletop plane. One marker is placed at different pose from the ground plane and the other markers are placed on the ground plane. In such a case, the markers which face to the same directions as the tabletop plane cannot be recognized because of the angle of the camera. If all the markers have to be on the same plane, it even fails recognition for most of the markers. In our registration method, however, the markers can face to various directions like Figure 12 because the markers can be placed at arbitrary positions and poses. The marker with the red cube is placed at different pose from the ground plane, so that this marker can be detected even in the case that the markers on the tabletop plane are not detected. Therefore, the registration can be stably continued even if the user moves the camera to any view point. This is a big advantage of the proposed system for applying to entertainment AR applications.

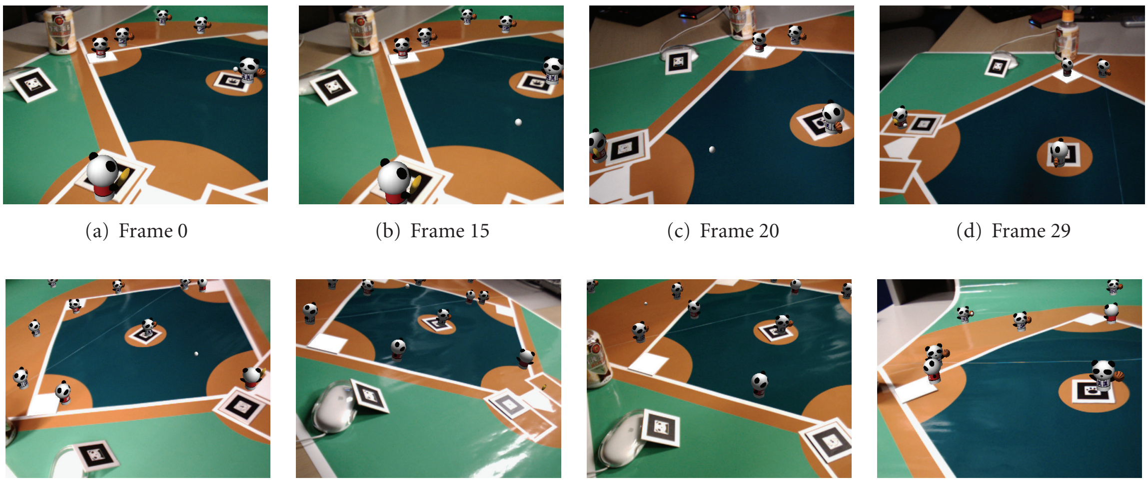
5. Interactive AR Bowling System

In the Interactive AR Bowling System, as shown in Figure 1(b), a real bowling lane model is placed on the tabletop. The users roll a real ball down the bowling lane model to knock down virtual pins generated with CG. Of course they can move around the lane model and can see the virtual bowling game scene from favorite view points by applying the algorithm presented in Section 3.

As for related work with bowling, Matysczok et al. have also proposed a bowling system using AR [19]. A user wears an HMD, in which a virtual ball, lane, and pins are

(1) *Arrangement*
Place field model and multiple markers at arbitrary positions and poses.
(2) *Capturing*
Capture the object scene as candidates for two reference images;
(3) *Input*
Input Scorebook Data File;
(4) *Observation*
Start system and observe game with moving around;

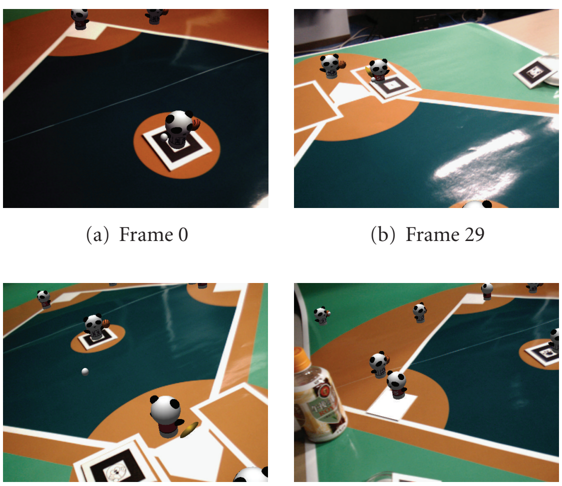
Algorithm 1: User's Operations.



(a) Frame 0 (b) Frame 15 (c) Frame 20 (d) Frame 29

(e) Frame 35 (f) Frame 50 (g) Frame 65 (h) Frame 89

Figure 10: Example of play: 4th batter of team RED sends a hit to left with the bases loaded and scores a goal.



(a) Frame 0 (b) Frame 29

(c) Frame 35 (d) Frame 89

Figure 11: Closeup views of the same scenes as Figure 10 seen from different view points.

(a) (b)

Figure 12: Most of the markers which face to the same directions as the tabletop cannot be detected. The marker which faces to different direction is detected successfully. (a) Marker detection: the red cube on the marker represents detected marker. (b) Augmented view: virtual objects are overlaid using the detected marker.

displayed, and interacts with the virtual ball by hand gesture. Since all objects, including the ball, are not real but virtual objects, this system just generates virtual bowling scenes with the input of hand gesture from the sensors. Thus, it is unnecessary to overlay the virtual scene onto the real scene like an AR system. Moreover, the user can hardly see the real world because the virtual lane is covering the real scene. Therefore, the meaning of AR that is mixing the real world and the virtual world is lost.

In our system, in contrast, since the ball is a real object, the virtual scenes are generated according to the ball's motion in the real scene. Therefore, it is effective to be an AR

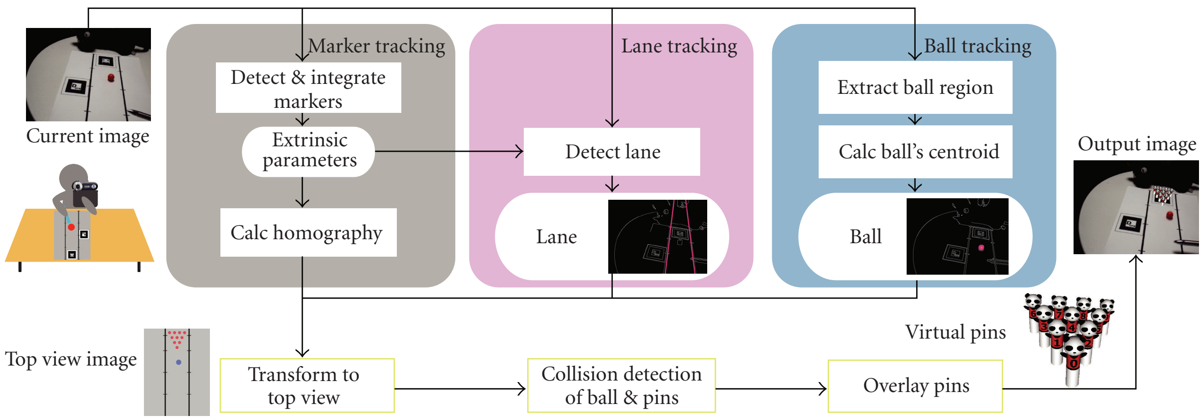


FIGURE 13: Overview of Interactive AR Bowling System.

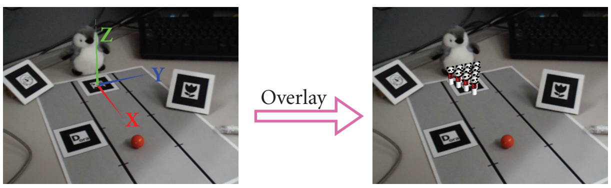


FIGURE 14: 3D coordinate system defined on the real lane model. Virtual pins are overlaid in this coordinate system.

system with overlaying the virtual pins onto the real lane model. Moreover, the user can touch the real ball, so our system achieves a real bowling style. In Matysczok's system, special gloves with physical sensors are also required for user's interaction, however, our system needs only a camera and a PC and use a real ball and lane model.

To realize this kind of bowling system, we have to perform the following tasks. There are two lines and 2D markers on the bowling lane model. These two lines define the lane, which means that the ball should be rolling between the two lines. In case the ball goes out of the lane, it is considered as "gutter." Therefore, the lane and the ball have to be detected and tracked at every frame.

When the ball hits any virtual pins, the pins are knocked down. For generating such virtual pins according to the ball, the geometrical relationship between the real ball and the virtual pins has to be computed interactively. In our method, the ball's position on the input image is transformed onto a top view image, which is the input image seen from top view, to obtain the ball's position to the pins.

Finally, the virtual pins are overlaid onto the input image according to the camera's position and pose, which are corresponding to extrinsic parameters of the camera. The extrinsic parameters are estimated by multiple 2D markers.

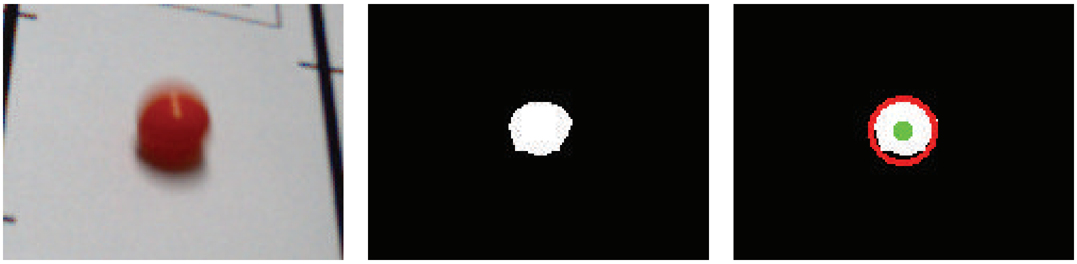
5.1. Overview of Processing

Figure 13 shows a flowchart of our proposed system. First, the images captured by the web camera are applied to three kinds of processing; marker tracking, lane tracking, and ball tracking. During the marker tracking process, AR-Toolkit [4] detects 2D markers placed around the lane model. Then, a 3D coordinate system where the virtual objects should be overlaid is defined on the lane model. Since the relationship of the lane with respect to the marker is fixed by the lane model, the two lines which consist of the lane can be detected by marker detection process. During the ball tracking process, a region of the ball is detected. We assume that the centroid of the region is the ball's position.

After the tracking of the markers, the lane, and the ball, the ball's position is transformed to the top view image to compute the geometrical relationship between the ball and the virtual pins. Then, collisions between the ball and the virtual pins are computed according to their relative position. Finally, the pins are overlaid onto the input image by using the extrinsic parameters computed at the marker tracking process.

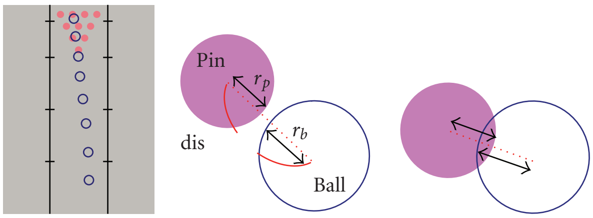
5.2. Marker Tracking

The multiple markers placed around the lane model are detected as same as in our baseball system. The geometrical relationship of the markers are also estimated by the registration method described in Section 3. Then, a 3D coordinate system, where the virtual objects should be overlaid, is defined on the lane model as shown in Figure 14. In our system, to track the trajectory of the ball on the lane, we use a top view image. The top view image is an image which is the input image transformed to the top view point. By this transformation, the ball's 2D motion becomes understandable. Therefore, we compute a homography $\mathbf{H}$ [20] to transform the input image to a top view image. The $\mathbf{H}$ is the planar projection matrix which relates the real-lane model and the lane model area in the input image and can be computed from the corresponding points on the lane model in the real world and the input image. It will be used in Section 5.4.



(a) Original image (b) Ball region (c) Detected ball

Figure 15: Ball detection.



(a) Trajectory (b) Not hit (c) Hit

Figure 16: Collision detection by trajectory of the ball.

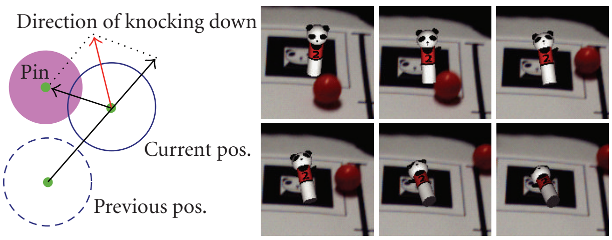


Figure 17: Direction of knocking down.

5.3. Ball Tracking

In this system, we assume that the color of the ball should be quite different from the lane model. In this paper, we use a red ball on a gray lane model as shown in Figure 15(a). For detection of the ball, first, red regions are detected from the input image by dividing it into R, G, B channel images. Figure 15(b) shows the image after dilation and erosion a few times. Finding the minimal circumscribed circle (contour) for the detected region, the center of the circle is considered as the 2D ball's position in the input image as shown in Figure 15(c).

5.4. Transformation to Top View Image

Using homography $\mathbf{H}$ computed at the marker tracking process, the ball's position on the input image is transformed onto the top view image that provides a geometrical relationship between the ball and the pins on the lane model.

As shown in Figure 16(a), the trajectory of the ball can be obtained. This trajectory is used to detect the collision between the ball and the pins, and compute the directions in which the pins are knocked down.

5.5. Collision Detection of Ball and Pins

We assume that radii of the ball and the pins are r_b and r_p, respectively, and define the distance between the ball and each pin as dis. For detecting a collision between the ball and the pins, the distance dis is computed from the top view image at every frame. The collision is detected by comparing distance and radius as in the following equation, and as in Figures 16(b) and 16(c):

$$\{(r_b + r_p) \leq \text{dis} \Longrightarrow \text{Not hit}, \; (r_b + r_p) > \text{dis} \Longrightarrow \text{Hit}. \quad (7)$$

5.6. Overlay Virtual Pins

After the collision detection, the pins are generated with CG and overlaid onto the image. If the collision is detected, the pins are gradually inclined and knocked down. The direction of knocking down is defined by trajectory of the ball. As shown in Figure 17, the direction is computed by a motion vector of the ball, which is decided by ball's positions in previous and current frames, and a vector from the ball to each pin.

The generated pins are superimposed onto the image by the extrinsic parameters computed by 2D markers. The user can see the virtual pins according to the motion of the camera and the rolling ball.

5.7. Demonstrations

In this experiment, four 2D markers on the lane model are placed on the tabletop to estimate the camera motion (extrinsic parameters). Some of them are placed on the same plane as the tabletop; the others are aligned in various directions. The geometrical relationship between every marker is automatically computed by the method in Section 3. One of the markers, which lies between the lines, is also used for defining the 3D coordinate system. The resolution of the captured image is 640 × 480 pixels. The virtual pins are rendered with OpenGL library.

Figure 18 shows the detected lane and ball's trajectory. Both of the lane and the ball can be correctly detected and tracked over all frames by our tracking method, according

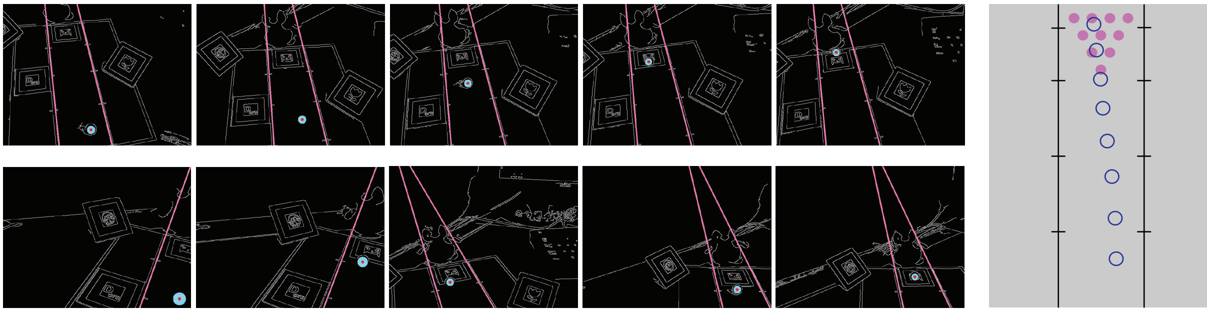


Figure 18: Detected lane and ball and trajectory of ball.

to the camera motion. The ball's position is also successfully transformed onto the top view image by the homography computed by 2D markers.

Figure 19 shows example scenes where the virtual pins are overlaid according to the camera motion. If we use only one marker for overlaying the pins, the registration becomes impossible when the ball is rolling over the marker because the marker cannot be detected. Therefore, we have to use multiple markers. Even though particular markers are not continuously captured over the frames, the virtual pins can be correctly registered on the lane model because of our registration algorithm which can estimate the geometrical relationship of the markers.

Moreover, since the collision of the real ball and the virtual pins are successfully detected, some pins are knocked down by hitting of the ball. The pins existing behind the hit pins are also knocked down as a chain reaction of the front pins by computing the direction of knocking down from the trajectory. This system runs 30 fps, so the user can enjoy the bowling game at video rate.

6. User Study

6.1. Baseball System

AR Baseball Presentation System visually replays the baseball game which was previously played in the other place. In contrast with the usual way to know the game history, such as watching the captured video or reading the recorded scorebook, our AR system can provide the users with much realistic sensation as an entertainment application.

We design our system as AR system which can overlay the CG scene on the real field model in front of the user as well as visualize the recorded baseball game by 3D CG. Using this system, the user can watch the game from the favorite view point by just moving around the field model. Such simple way for watching the CG-represented event using the AR system provides more immersive feeling than using usual CG viewers, in which a mouse or a keyboard is used for changing view points [21].

In this user study, we intend to evaluate how the AR system is effective to enhance the quality of entertainment. There are a lot of factors that may enhance the quality of entertainment such as usability, interactivity, visual effects. Those factors are evaluated by studying "how quickly," "how easily," and "how intuitively" the user can change their view point. The same factors are also evaluated for the usual CG viewer. Then, both of results are compared to evaluate the effectiveness of designing our system as AR system.

In this evaluation experiment, we prepared two kinds of baseball observation systems as shown in Figure 20, one is our AR Baseball Presentation System, another one is created by only CG. In this CG system, the user watches the baseball game with changing the view point by using a keyboard of PC. The rotation and translation about X-Y-Z axis are assigned to each key. In the AR system, they just move around the field model with a handheld monitor. Then, we asked 15 examinees to use these systems, and measured the time each examinee spent on moving the view point to the specified view points as shown in Figures 21(a)–21(d).

Figure 21 shows the average time which the examinees spent to change their view points. We can find that the CG system took much longer time than the AR system. In this experiment, every examinee spent triple to ten longer times to change the view point in CG system than AR system. Because the users only have to move around the field model with a handheld monitor to their favorite positions, our AR system can quickly change the view points.

This result can also be found in Figure 22, which shows answers of questionnaire about changing view points. We asked four questions (a)–(d), and then the examinees rated on a scale of 1 to 5. In the same way as the actual measurement time in Figure 21, most of the examinees felt that our AR system was easier than the CG system to change their view points to their desired positions quickly and intuitively. The questionnaire also asked whether they could change the view point with watching the game. As a result, most of examinees felt that the AR system was easier to change the view point with watching the game. This is because the view point corresponding to the user's own view point, while the view point of the CG system is the virtual camera position. Therefore, designing our system as the AR system is quite helpful for any user to handle this kind of digital contents because the operation is very easy and intuitive.

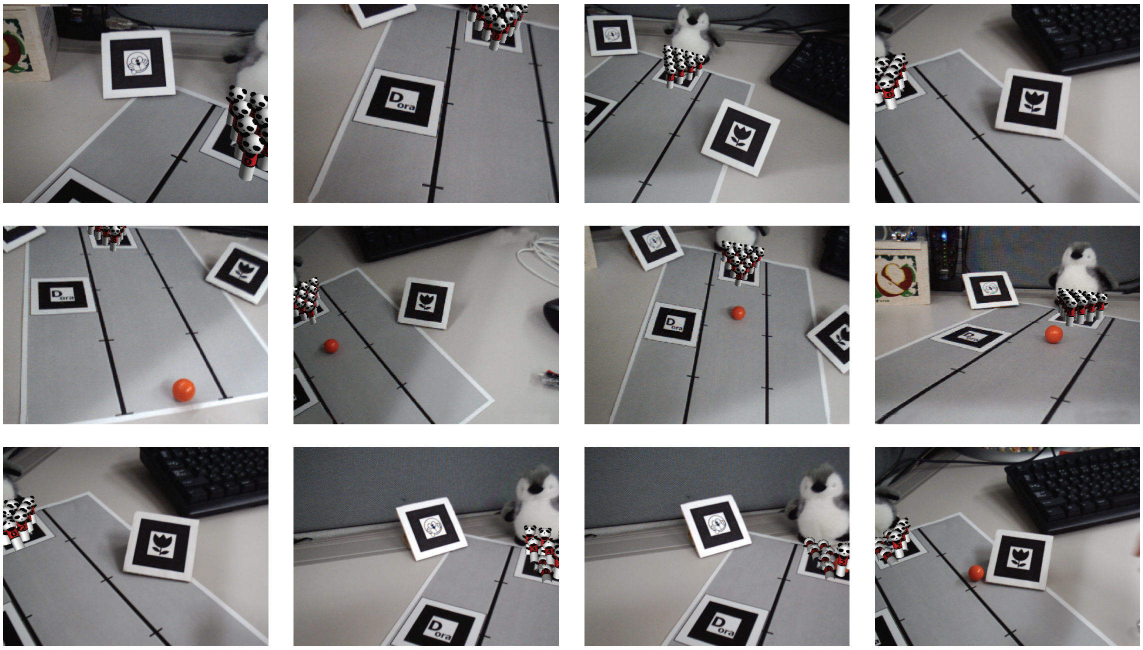
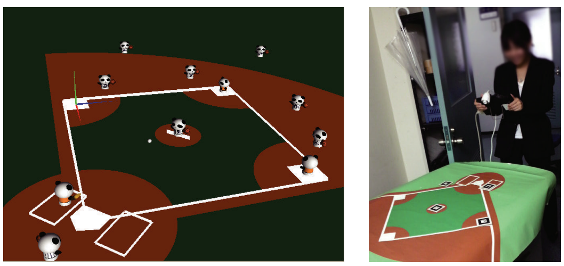


Figure 19: Resulting images on which virtual pins are overlaid.



(a) CG (b) AR

Figure 20: Prepared baseball system.

6.2. Bowling System

AR Bowling System consists of real and virtual objects, such as the real ball, the real lane model, and the virtual pins. In this system, the users physically touch and roll the real ball on the real lane. Such physical communication provides the users with much reality as a tangible interface [5–8].

Therefore, we focus on the effectiveness of such tangibility of our AR system as the evaluation point of the AR Bowling System. The AR system can interact with the virtual world by rolling the real ball with a hand unlike a CG system which entirely consists of virtual objects. This means that we should evaluate how effective the direct touch to the ball for the bowling game as an entertainment application. For evaluating this point, we evaluate "how naturally the users can control the ball," "whether they actually feel that the ball is controlled by themselves," and "whether the game is challenging."

In the same way as the baseball system, we prepared two kinds of bowling system, our AR Bowling System, a CG bowling system, and a real toy bowling game as shown in Figure 23. We also asked the examinees to play the real toy bowling game before using the CG and AR bowling systems. In the CG system, when the user drags the virtual ball on the display by a mouse, the ball starts rolling. The direction of the ball is defined according to user's dragging. The speed of the ball is also defined as the length of dragging of the mouse. In the AR system, the user rolls the real ball on the lane model with a hand and watches the proceeding through the handheld monitor. After playing all the systems, they answered questionnaire by rating on a scale of 1 to 5. The questionnaire items and the results are shown in Figure 24.

Although they could only decide the direction and speed of the ball in the CG system with a mouse, the users could freely control the ball in the AR system; and also they could really touch the ball. As a result, they actually felt rolling the ball by themselves in the AR system. Since the ball is virtual object in the CG bowling system, the users can roll the ball only linearly. Therefore, some users said that they wanted to roll the curve-ball. In order to achieve such a curve-ball in CG system, some random elements have to be included. Since such randomness cannot be handled by the users, however, such system is unacceptable as computer games. On the other hand, the users can roll any kind of ball depending on their skills because the ball of the AR system is real object. Therefore, most of the users felt that the ball's motion of the AR system was more natural like the real bowling game than the CG system. Because there are various ways to roll the ball in the AR system, the game is not too simple to complete. For example, some users sloped the lane; other users used a pen to roll the ball instead of their hands. Therefore, they felt that the AR system was more challenging than the CG system.

By the way, when playing the real bowling game, we asked the examinees to raise and reset the fallen pins by themselves.

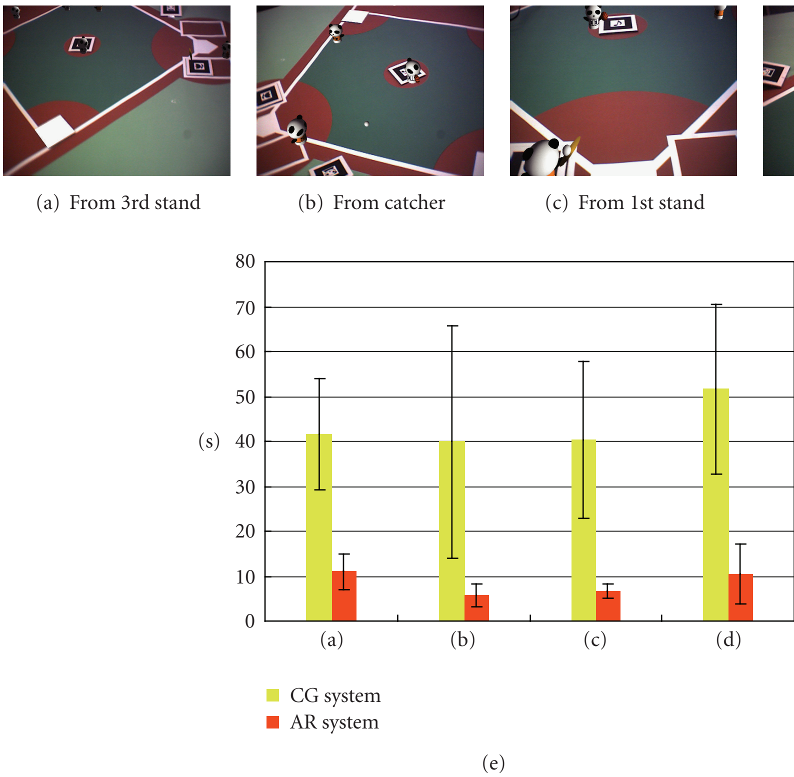
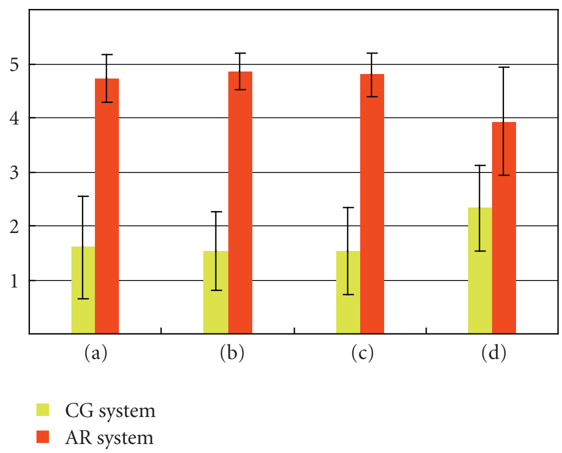


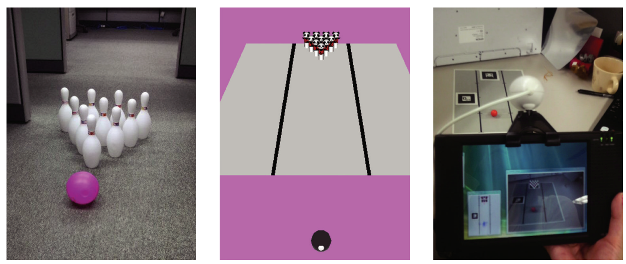
FIGURE 21: Average time the examinees took to change their view points to specified view points.



(a) How fast in changing view-point
(b) How intuitive in changing view-point
(c) How easy in changing view-point
(d) Can see game proceeding during changing view-point

FIGURE 22: Answer of questionnaire about changing view points.

As a result, they felt that it's troublesome to reset the pins every time. As described before, physical communication is very effective for the young users, however, we can afraid that little children cannot arrange the pins very well. On the other hand, virtual bowling game (both of AR system and CG system) do not require such a task because the users only



(a) Real (b) CG (c) AR

FIGURE 23: Prepared bowling system.

have to press 1 button to reset the fallen pins. Therefore, the concept of the AR Bowling is very helpful for any user by allowing the physical communication without troublesome task.

7. Conclusions

In this paper, we have presented two AR applications using vision-based tracking method: AR Baseball Presentation System and Interactive AR Bowling System. Both of the applications can be enjoyed on the tabletop in the real 3D world only with a web camera and a handheld monitor connected to a PC. It is a big advantage for home users that our applications do not require any special device such as positioning sensors or a high-performance PC.

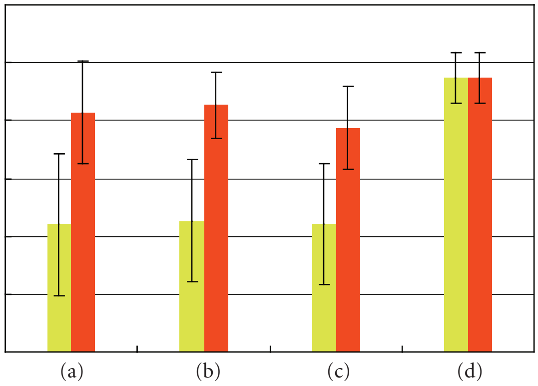


Figure 24: Answer of questionnaire about the reality of playing bowling.

Users can interactively change their view points by moving around the tabletop because of multiple 2D markers. In usual AR applications using multiple 2D markers, users have to measure the distance between the markers. Of course, such extra task is unnecessary in our applications by applying the registration method with the 3D projective space. In contrast with usual CG viewers in which a mouse or a keyboard is used for changing view points, changing view points by moving the users is very intuitive and easy way. Such a facility is very important specially for children.

Using the baseball application, the users can watch a 3D virtual baseball game in front of themselves. It can be a future-oriented 3D game which is represented in movies or animations. The bowling application can interest children because their actions in the real world affect the virtual world.

For the future work, we have to consider that sound is very important element. Since sound can directly interest people, sound is very effective as response of the user's interaction in AR. For example, if the baseball system downloads ambient sound data in the actual baseball stadium with the scorebook data and play the sound according to the game, more realistic sensation will be given to the users. In the bowling system, sound effect of collision between the real ball and the virtual pins is also effective and interesting. So we would like to adopt sound elements in the future.

References

[1] M. Haller, M. Billinghurst, J. L. Leitner, and D. Seifried, "Coeno-enhancing face-to-face collaboration," in *Proceedings of the International Conference on Augmented Tele-Existence (ICAT '05)*, pp. 40–47, Christchurch, New Zealand, December 2005.

[2] A. Henrysson, M. Billinghurst, and M. Ollila, "Virtual object manipulation using a mobile phone," in *Proceedings of the International Conference on Augmented Tele-Existence (ICAT '05)*, pp. 164–171, Christchurch, New Zealand, December 2005.

[3] G. Klein and T. Drummond, "Sensor fusion and occlusion refinement for tablet-based AR," in *Proceedings of the 3rd IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '04)*, pp. 38–47, Arlington, Va, USA, November 2004.

[4] M. Billinghurst, S. Cambell, I. Poupyrev, et al., "Magic book: exploring transitions in collaborative ar interfaces," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '00)*, p. 87, New Orleans, La, USA, July 2000.

[5] H. Ishii, C. Wisneski, J. Orbanes, B. Chun, and J. Paradiso, "Curlybot: designing a new class of computational toys," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*, pp. 394–401, Pittsburgh, Pa, USA, May 1999.

[6] B. Piper and H. Ishii, "PegBlocks: a learning aid for the elementary classroom," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*, pp. 686–687, Minneapolis, Minn, USA, April 2002.

[7] K. Ryokai, S. Marti, and H. Ishii, "Designing the world as your palette," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, pp. 1037–1049, Portland, Ore, USA, April 2005.

[8] J. Zigelbaum, A. Millner, B. Desai, and H. Ishii, "Bodybeats: wholebody, musical interfaces for children," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, pp. 1595–1600, Montreal, Canada, April 2006.

[9] Y. Uematsu and H. Saito, "AR registration by merging multiple planar markers at arbitrary positions and poses via projective space," in *Proceedings of the International Conference on Augmented Tele-Existence (ICAT '05)*, pp. 48–55, Christchurch, New Zealand, December 2005.

[10] A. Henrysson, M. Billinghurst, and M. Ollila, "AR tennis," in *Proceedings of the 33rd International Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH '06)*, Boston, Mass, USA, July-August 2006.

[11] E. Foxlin, Y. Altshuler, L. Naimark, and M. Harrington, "FlightTracker: a novel optical/inertial tracker for cockpit enhanced vision," in *Proceedings of the 3rd IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '04)*, pp. 212–221, Arlington, Va, USA, November 2004.

[12] E. Foxlin and L. Naimark, "Miniaturization, calibration & accuracy evaluation of a hybrid self-tracker," in *Proceedings of the 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)*, pp. 151–160, Tokyo, Japan, October 2003.

[13] E. Foxlin and L. Naimark, "Vis-traker: a wearable vision-inertial selftracker," in *Proceedings of the IEEE Virtual Reality Conference (VR '03)*, pp. 199–206, Los Angeles, Calif, USA, March 2003.

[14] Y. Genc, S. Riedel, F. Souvannavong, C. Akinlar, and N. Navab, "Marker-less tracking for ar: a learning-based approach," in *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '02)*, pp. 295–304, Darmstadt, Germany, September-October 2002.

[15] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana, "Virtual object manipulation on a table-top ar

environment," in *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR '00)*, pp. 111–119, Munich, Germany, October 2000.

[16] D. Kotake, S. Uchiyama, and H. Yamamoto, "A marker calibration method utilizing a priori knowledge on marker arrangement," in *Proceedings of the 3rd IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '04)*, pp. 89–98, Arlington, Va, USA, November 2004.

[17] Y. Uematsu and H. Saito, "Vision-based registration for augmented reality with integration of arbitrary multiple planes," in *Proceedings of the 13th International Conference on Image Analysis and Processing (ICIAP '05)*, vol. 3617 of *Lecture Notes in Computer Science*, pp. 155–162, Cagliari, Italy, September 2005.

[18] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK, 2000.

[19] C. Matysczok, R. Radkowski, and J. Berssenbruegge, "AR-bowling: immersive and realistic game play in real environments using augmented reality," in *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '04)*, pp. 269–274, Singapore, June 2004.

[20] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2003.

[21] GLT ZPR, http://www.nigels.com/glt/gltzpr/.

Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2008, Article ID 357270, 13 pages
doi:10.1155/2008/357270

Research Article

AR Supporting System for Pool Games Using a Camera-Mounted Handheld Display

Hideaki Uchiyama and Hideo Saito

Faculty of Science and Technology, Keio University, Hiyoshi Kohoku-ku, Yokohama-shi, Kanagawa-ken 223-8522, Japan

Correspondence should be addressed to Hideaki Uchiyama, uchiyama@ozawa.ics.keio.ac.jp

Received 2 October 2007; Revised 21 July 2008; Accepted 28 August 2008

Recommended by Adrian Cheok

This paper presents a pool supporting system with a camera-mounted handheld display based on augmented reality technology. By using our system, users can get supporting information when they once capture a pool table. They can also watch visual aids through the display while they are capturing the table. First, our system estimates ball positions on the table with one image taken from an arbitrary viewpoint. Next, our system provides several shooting ways considering the next shooting way. Finally, our system presents visual aids such as shooting direction and ball behavior. Main purpose of our system is to estimate and analyze the distribution of balls and to present visual aids. Our system is implemented without special equipment such as a magnetic sensor or artificial markers. For evaluating our system, the accuracy of ball positions and the effectiveness of our supporting information are presented

1. Introduction

A pool game is one of popular sports in the world and enjoyed by people of all ages. Pool players hit the cue ball by using a long pole and pocket numbered balls into holes around a pool table. As being well known, pocketing balls requires much skills and consideration. It is necessary not only to hit the cue ball precisely, but also to consider the direction and strength of the shot. Especially, it is difficult for beginners to hit the cue ball considering laws of physics such as collisions and reflections.

There are two types of studies about pool games. First type includes studies such as developing a robot that plays pool [1–5], and second type includes studies such as developing a supporting system for beginners [6, 7].

In the studies for developing pool-playing robots, main research topic is that a robot placed above a pool table automatically recognizes ball positions and decides the best shooting way by analyzing positional relationship between pockets and balls for playing pool with human beings. As for recognition of balls, a camera should be hung exactly above the center of the table which provides the sensory input for these robots. Each system estimates ball positions by using color identification with the calibrated camera. The difference among these researches is computation of best shooting way based on analyzing the distribution of balls. Alian et al. [1] used a cost function composed of distances and angles to compute the strength and the direction of the shot and a fuzzy approach to select the best shot. Lin et al. [2] adopted grey decision making which is similar with fuzzy algorithm and can deal with uncertainty and incompleteness of the pool game to find the best shot. Cheng et al. [3] also used the same fuzzy approach as Alian [1], however, Cheng's algorithm included the predictable hitting error model which was developed based on the recorded database of pocketing processes. Chua et al. [4, 5] used another fuzzy approach which was called Sugeno Fuzzy and evaluated the effectiveness of the selected shot based on fuzzy logic.

In the studies on supporting system, main research topic is the way of presenting of visual aids for practice. Jebara et al. [6] proposed an augmented reality-based wearable computer system which used a head-mounted live video display with a camera. However, their visual aids were just drawn lines and the registration of visual aids was not achieved because the camera was not calibrated. They suggested that it was necessary to calibrate the camera to display more effective visual aids. Larsen et al. [7] proposed a framework of an automatic pool trainer. A camera was set above the pool table to estimate ball positions by the same way as the studies for the pool-playing robots and a projector was also set for visual output on the table to users. In addition, users can issue spoken commands to an interface agent acting as a personification of the system.

For a pool-supporting system for games, it is necessary to consider not only the computation of shooting ways suggested in the studies for the pool-playing robots but also the display of visual aids suggested in the studies for pool-supporting systems. As for computation of shooting ways, only current distribution of balls is analyzed in the studies for the robots. However, it is important for pool players to analyze the next distribution of balls after the shot because they can hit the cue ball twice in a row after pocketing a target ball. The players should also consider next strategy such as Pocket that means putting a ball into a pocket or Safety that means disturbing the adversary player to put a ball into a pocket. As for display of visual aids, a head-mounted live video display is not appropriate because it is not natural for players. It is also difficult to install the projector system in a usual environment of a pool hall because the system is huge and takes much cost. It is ideal that special equipment is not utilized.

In this paper, we propose a pool-supporting system that solves these problems as mentioned above. We adopt a camera-mounted handheld display such as a mobile phone because, such device is small and widespread. Our system provides supporting information for games considering the strategies and the next distribution of balls. It is important for beginners to plot strategies of winning the game by the next distribution of balls. The supporting information is presented as visual aids through the display. In our system, ball positions are initially estimated [8]. Then, supporting information for several strategies are computed by simulating ball behavior. Finally, visual aids are drawn on captured images based on a kind of an augmented reality system [9] (this concept is the same as that of the Jebara's system [6]); but they did not draw visual aids and did not register them because the camera was not calibrated in their system. In our system, this registration is achieved since the camera is calibrated by using natural features such as a pool table and balls.

2. Proposed System

2.1. System Overview

The usage and graphical user interface of our prototype system are shown in Figure 1, which is implemented on a Tablet PC that is similar with a mobile phone. Our system is designed for Nine-Ball game which is one of popular pool games in the world among others.

First, a user standing beside the pool table captures the whole part of the table from an arbitrary viewpoint. Next, our system estimates ball positions from the captured image. The user corrects ball positions by using the interface if our system does not precisely estimate ball positions. After

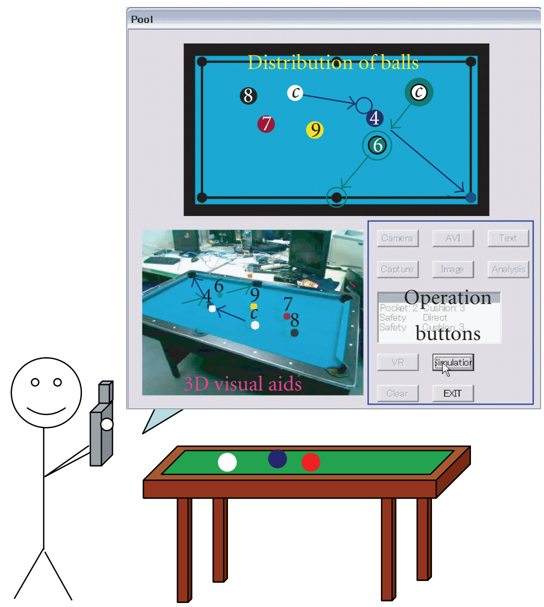


Figure 1: System overview.

getting ball positions, our system computes some shooting ways for each strategy. Since our system provides several shooting ways, users can select one of the ways based on ability, and learn the shooting ways that they cannot notice. When the user chooses one shooting way, the user can watch visual aids such as direction of the shot and ball behavior from an arbitrary viewpoint. To sum up the matter, the user can get supporting information when the user once captures the table. The user can also watch visual aids through the display while the user is capturing the table.

2.2. Algorithm Overview

In our system, it is important to estimate a projection matrix of an image which is captured from an arbitrary viewpoint for estimating ball positions and displaying visual aids on the captured image. A projection matrix represents the relationship between an image coordinate system and a world coordinate system by the following equation:

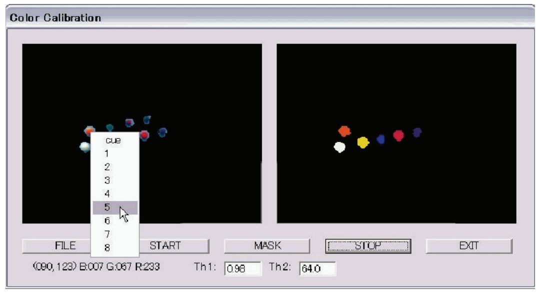
$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \mathbf{P} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{1}$$

where (x, y) is an image coordinate and (X, Y, Z) is a world coordinate, $\mathbf{P}$ is a projection matrix. Our system computes a projection matrix based on a homography that represents the case of $Z = 0$ in (1) by using Simon's method [10, 11].

The algorithm is divided into three sections.

The first section is estimation of ball positions on the pool table [8]. Ball position estimation is based on camera calibration using planar area. First, a green planer area of the table is extracted from a captured image by color segmentation to compute four corner positions of the table in the image. The relationship of four corner positions

(a) Specification of all balls and table area



(b) Specification of each ball area

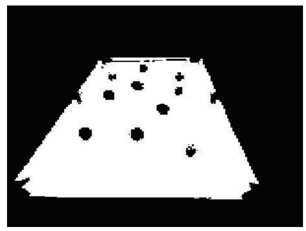
Figure 2: Color acquisition.

between the image and the table provides a homography, which is transferred to a projection matrix by using Simon's method [10, 11]. Then, ball areas are also extracted from the image by color segmentation to compute the center of each ball area, which can be transferred to the ball position on the table using the projection matrix. Finally, a ball number of each area is identified by a voting method.

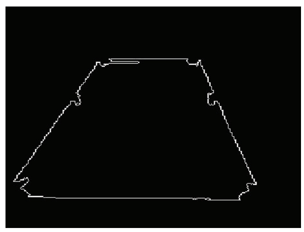
The second section is computation of shooting ways. Our system considers some strategies based on the rule of a pool game and the next shooting way after a shot. First, our system finds ball paths of each strategy in which a player can physically hit the cue ball at a target ball. In each ball path, our system simulates ball behavior by giving several speeds to the cue ball to consider next distribution of balls after hitting the cue ball. Also, each ball path is evaluated by a cost function based on a distance and an angle. As a result, some desirable shooting ways are selected in score order.

The last section is presentation of visual aids [9]. When a user chooses one shooting way, our system displays visual aids such as a direction of the shot and ball behavior on the captured images while the user is capturing the pool table. Our system calibrates the camera frame by frame online and generates background images by removing ball areas from the captured images. Then, the ball behavior is drawn on the background images while the user is capturing the table. Since our system can draw any 3D objects such as a performance of trained player on the image by using this algorithm, our system can draw more plentiful and precise visual aids than those of a previous research [6].

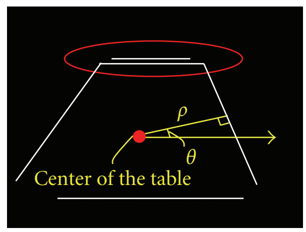
(a) Input image



(b) Table mask



(c) Table contour



(d) Line segments

(e) Table area

Figure 3: Extraction of four corners.

2.3. Preparation for the System

The initial inputs of our system are the sizes of a pool table and balls and the colors of balls. The sizes of a pool table and balls are determined in advance for official rules and are not necessary to be measured. On the other hand, the colors of balls depend on lighting conditions of the room. For this reason, the colors are needed to be measured once.

Figure 2 represents the graphical user interface of the color acquisition system. The mat area seems to be blue in all figures because of color response of our camera. The mat area of the pool table actually seems to be green. For this reason, we use "green" uniformly to indicate the color of the pool table.

The input image of our system needs to include nine solid balls and a pool table. A user segments an area that includes the balls and the table as shown in Figure 2(a). Our system initially decides the color of the table by making RGB histogram of the segmented area and computing a peak of the histogram. Then, our system extracts ball areas by removing the table area, assuming that ball colors are not the same as the color of table. Finally, a user clicks each ball area as shown in Figure 2(b), and our system outputs the colors of the balls and the table as a text file. Since this acquisition of the colors is once measured beforehand if the lighting condition is constant, a user does not necessarily have to measure the colors and another person measures these parameters in a usual environment of pool games.

3. Estimation of Ball Positions

3.1. Estimation of Four Corner Positions

In order to obtain a homography for computing a projection matrix, our system estimates four corner positions of a pool table.

An example of the images which is captured by a user standing besides the table is shown in Figure 3(a). For estimating four corner positions of the green frame area, our system extracts whole part of the green mat areas.

For RGB value of each pixel in the captured image, the inner angle of the two color vectors is calculated by the following equation:

$$\text{Angle} = \arccos\left(\frac{\mathbf{C}_t \cdot \mathbf{C}_p}{|\mathbf{C}_t||\mathbf{C}_p|}\right), \quad (2)$$

where $\mathbf{C}_t$ is the template color as mentioned above, $\mathbf{C}_p$ is each pixel color, and Angle is the angle between the vectors $\mathbf{C}_t$ and $\mathbf{C}_p$. The angle is small if two colors are similar, while the angle is large if two colors are not similar. Each pixel is classified by setting a threshold on 10 degree that is determined via experiments. As a result, a mask of the green mat area shown in Figure 3(b) can be generated.

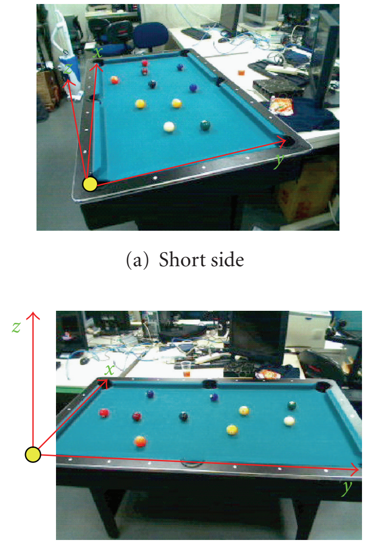
Next, four corners positions of the green frame area are estimated by calculating four intersections of four line segments of the green frame area. Four line segments are computed from contours of the mask as shown in Figure 3(c). Our system applies Mata's method [12] which is implemented on OpenCV [13] for extracting line segments from contours.

In Figure 3(d), more than four line segments are extracted and these segments are clustered into four segments. The space for the clustering is $\rho - \theta$ space denoted in Figure 3(d). For each line, the perpendicular line from the center of the green table area is connected and the connected point is calculated. ρ is the distance between the center and the connected point and θ is the angle between x-axis of the image and the vector connecting the center and the connected point. In the space, line segments have the values of ρ and θ, and they are clustered into four clusters by using k-means clustering method ($k = 4$). In Figure 3(d), two lines in the ellipse become one cluster and the center of the cluster is calculated for one line. In addition, these lines can be extracted clockwise based on the value of θ.

As shown in Figure 3(e), the green mat area is extracted by calculating the intersections of four line segments.

3.2. Camera Calibration

For calibrating a camera, our system computes a homography by applying Simon's method [10, 11] that provides a projection matrix from a homography.



(a) Short side

(b) Long side

Figure 4: Definition of a world coordinate system.

A homography is a matrix which represents 2D-2D relationship by the following equation:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}. \quad (3)$$

A homography is computed from four corresponding points in the same plane because of eight degrees of freedoms.

The relationship between a homography and a projection matrix is represented by the following equation:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \mathbf{H} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \sim \mathbf{P} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix}. \quad (4)$$

A homography is equal to the case that a Z component is 0 in a projection matrix which represents 2D-3D relationship.

However, (5) represents components of a projection matrix, and (6) represents components of a homography. $\mathbf{A}$ consists of intrinsic parameters, $\mathbf{r}_i$ is a component vector of a rotation matrix, and $\mathbf{t}$ is a translation matrix. $[\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3\mathbf{t}]$ represents the rigid transform that consists of extrinsic parameters. As mentioned above, a homography is equal to the case that $\mathbf{r}_3$ is not included in a projection matrix:

$$\mathbf{P} = \mathbf{A}[\mathbf{R}|\mathbf{t}] = \mathbf{A}[\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3\mathbf{t}], \quad (5)$$

$$\mathbf{H} = \mathbf{A}[\mathbf{r}_1\mathbf{r}_2\mathbf{t}]. \quad (6)$$

For computing a projection matrix from a homography, $\mathbf{r}_3$ is calculated by the cross product of $\mathbf{r}_1$ and $\mathbf{r}_2$ which is one of properties of a rotation matrix by the following equation:

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2. \quad (7)$$

For calibrating the camera, our system computes a homography by relating four corner positions in the image

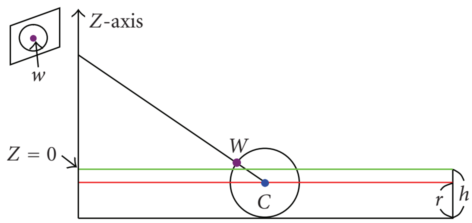
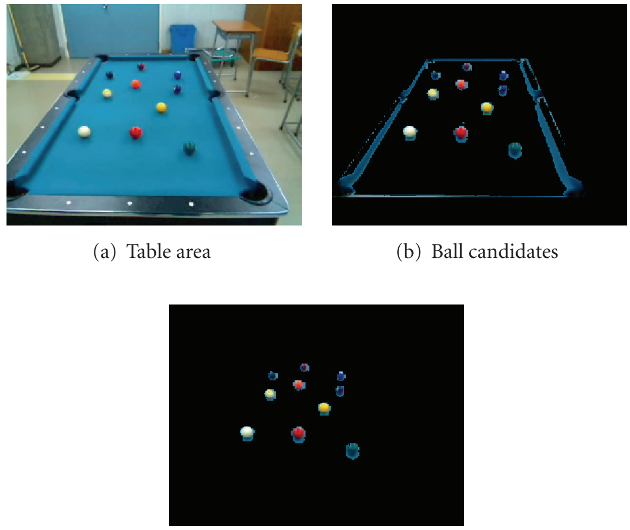


Figure 5: Estimation of a ball position.



(a) Table area (b) Ball candidates

(c) Ball areas

Figure 6: Extraction of ball areas.

with the corresponding corner positions in a world coordinate system. For computing a homography, it is necessary to define a world coordinate system. As shown in Figure 4, there are two probabilities of a side on which a user captures the table. In the case of the short side, the difference between the length of the short side and the long side is small. On the other hand, the difference is large in the case of the long side. By judging the side based on a threshold (in our case, 100 pixels), our system initially identifies on which side a user captures the table. Then, our system defines a world coordinate system as shown in Figure 4(a) or 4(b). To compute a homography, the left below corner in the captured image is related with the origin in the world coordinate system, and other corners in the image are related in a similar way.

Next, $\mathbf{A}$ and $[\mathbf{r}_1\mathbf{r}_2\mathbf{t}]$ are computed from a homography. A homography is represented by the following equation:

$$\mathbf{H} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{bmatrix}, \quad (8)$$

where f is a focal length and (c_x, c_y) is a principal point. Our system assumes that (c_x, c_y) is approximated to a center of a captrued image and obtained depending on the size of a captured image. Also, our system applies one of properties of a rotation matrix by the following equation:

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = 0. \quad (9)$$

Through calculation processing, our system computes a projection matrix from a captured image.

3.3. Estimation of a Ball Position

In (4), (X, Y) is a ball position on the table and Z-axis is vertical to the table. Since the plane of the homography is $Z = 0$ plane, Z of the center of the ball is determined as $Z = -(h - r)$ because the radius r of the ball and the height h of the cushion of the table are known. As shown in Figure 5, we consider the straight line that connects the center of the ball (C) to the center of the circle in the image (w). Then, we define the point W at which the line passes through the ball surface. By getting (x, y) of point w in the image, (X, Y) of point C is computed.

3.4. Extraction of Ball Areas

To compute (x, y) of each ball in the captured image, our system extracts ball areas. Ball areas are extracted in the green mat area because balls exist on the area. For extracting the ball areas, non-green areas are extracted from the green mat area as candidates of the ball areas, and the ball areas are selected from the candidates.

For RGB value of each pixel in the green mat area, the similarity score based on the inner product of the two color vectors and the difference of the two norms are computed by following equaition:

$$E = a \frac{\mathbf{C}_t \cdot \mathbf{C}_p}{|\mathbf{C}_t||\mathbf{C}_p|} + b\left(1 - \frac{|\mathbf{C}_t - \mathbf{C}_p|}{C_{\mathrm{MAX}}}\right), \quad (10)$$

where $\mathbf{C}_t$ is the template color, $\mathbf{C}_p$ is the color of each pixel, and E is the score which is composed of the inner product and the difference of the norms. By setting a on 0.7 and b on 0.3, each pixel is classified by setting a threshold on 0.8, and the candidates of the ball areas shown in Figure 6(b) can be generated.

In the candidates of the ball areas, pocket areas and shadows of the cushion are included because these colors are also different from the color of the green mat. Our system removes these areas depending on the size and the position of each area on the pool table.

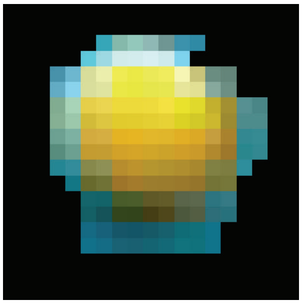
Since the size of the shadow area is larger than the size of a ball, the shadow area can easily be removed from the candidates by using a threshold that is determined from the size of a ball area (from 50 to 400 pixels). The pocket area can also be removed from the candidates by calculating their positions from a projection matrix because the pocket area is out of the area where balls exist.

Figure 6(c) shows the result of ball areas.

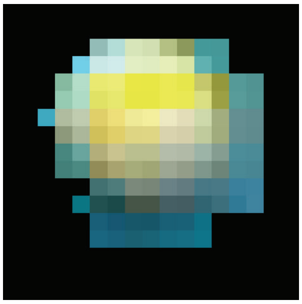
3.5. Identification of a Ball Number

In pool-supporting systems, it is important to identify a ball number as well as a position. Since our proposed system targets NineBall game, our system should identify the ball

(a) Template colors



(b) Ball number 1



(c) Ball number 9

Figure 7: Identification of a ball number.

Table 1: Classification by voting method.

Number	c	1	2	3	4	5	6	7	8
(b)	34	55	0	0	0	13	19	1	1
(c)	67	34	0	0	0	2	11	0	8

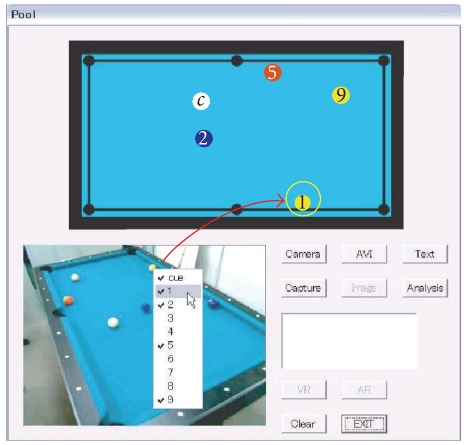


Figure 8: Correction by user interaction.

numbers of white cue ball and balls from number 1 to number 9 as shown in Figure 7(a).

Two examples of the extracted ball areas are shown in Figures 7(b) and 7(c). In addition to the ball color, green mat color and shadow are included in the area. Our system identifies the ball number based on the voting of the closest color pixel to the color template for each ball.

As shown in Figure 7(a), the color of each ball is measured beforehand. In the rule of NineBall game, balls are one white cue ball, eight solid ball from number 1 to number 8 and one stripe number 9 ball. Since both ball number 1 and ball number 9 are yellow, white and the colors from ball number 1 to ball number 8 are used in this voting method.

For RGB value of each pixel in an extracted ball area, the inner angles with each template color are computed, and the counter of the color which gets minimum inner angle gets increased. This means that the counter of closest color with the color of each pixel gets increased. Two examples of the vote are shown in Tables 1(b) and 1(c).

First, ball number 9 which is one stripe ball is identified by the following equation:

$$\frac{b_c + b_1}{b_{\text{sum}}} \geq 0.5 \cap 0.7 \geq \frac{b_c}{b_{\text{sum}}} \geq 0.3, \qquad (11)$$

where b_c is the result of a white ball counter, b_1 is the result of a number 1 yellow ball counter which is same color as the color of ball number 9, and b_{sum} is the sum of all counters. From the former conditional equation, the ball is judged into 4 possibilities such as the cue ball, number 1, number 9, and other balls. The latter conditional equation identifies number 9 ball by the ratio of white. If the ratio of white is over 0.7, the area is white ball. Also, if the ratio of white is less than 0.3, the area is number 1 ball. Otherwise, the area is ball number 9 because the ratio of white of ball number 9 is more than that of ballnumber 1. These thresholds are empirically determined.

In the case that (11) is not satisfied, a ball number that gets lion's share of votes is the number of the area. In Table 1, (b) is ball number 1 and (c) is ball number 9.

3.6. Correction by User Interaction

In the case that some balls stand too close to each other or close to the cushion, the ball areas sometimes cannot be extracted. For instance, the ball area which are close to the cushion that is far side from the camera are not extracted in Figure 8 (the ball with yellow circle). Since the ball area is not extracted, the center of the ball area which is necessary for computing a ball position is not computed. Therefore, the user inputs the center of the ball area and the ball number by using the graphical user interface in our system.

In Figure 8, a user clicks the center of a ball area in the captured image, and the menu of ball numbers is popped up. After the ball number is selected by the user, our system estimates the ball position by using a projection matrix and displays the ball position on the right pool table image in the interface.

4. Computation of Shooting Ways

4.1. Rule-Based Strategies

Our system targets NineBall game which is one of popular pool games in the world. The main rule of NineBall game is that players should pocket balls in an order of the ball's numbers. The common rule of pool games is that a player

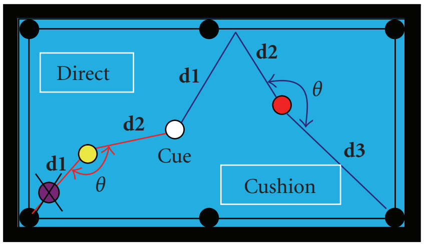
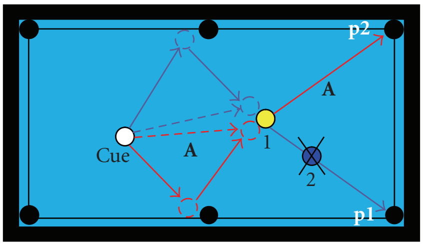
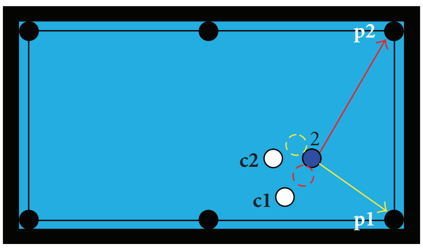


Figure 9: Evaluation of ball arrangement.



(a) Ball path



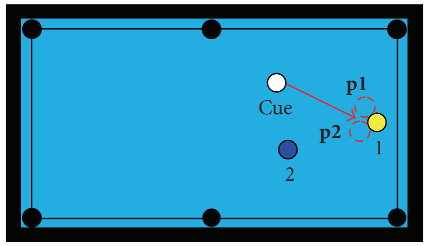
(b) Comparison of two results

Figure 10: Pocket.

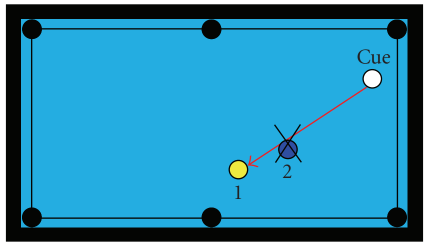
should switch if a player cannot pocket any ball. Based on these two rules, it is necessary to consider shooting ways of Pocket and Safety to win the game.

Pocket means that a player hits the cue ball to pocket balls by considering a position of the next target ball if a target ball can be pocketed, the player can hit the cue ball again. However, a player cannot always hit the cue ball at a target ball because of some special ball arrangement. In the cases, a player targets Safety which means that the player hits the cue ball not to pocket a target ball but to make distribution of the balls complicated to pocket a target ball for another player. This is from the second rule as mentioned above that a current player should switch with the next player if the current player cannot pocket any ball. In NineBall game, it is important for winning the game to judge whether a player can pocket a target ball or not and to consider the shooting way of targeting Safety. For this reason, our system searches some shooting ways of Pocket and Safety considering the next distribution of the balls by simulating ball behavior.

In addition, Free Ball is also one of important events in pool games, which occurs in the case that an adversary player fouls. In Free Ball, a player can put the cue ball anywhere.



(a) How to target safey



(b) Result of p1 in (a)s

Figure 11: Safety.

Thus, our system simulates ball behavior assuming that the cue ball is put on several places.

As for simulation of ball behavior, our system assumes that ball behavior follows particle dynamics and linear motion. Also, our system assumes that a player hits the center of the cue ball.

The algorithm overview is as follows. First, our system searches ball paths of Pocket and Safety for making contact with a target ball directly or using a cushion. Each found path is evaluated by an angle and distances as shown in Figure 9 and as follow:

$$E = a(1 - \cos\theta) + b\prod_{i=1}^{n} \frac{1}{d_i}, \tag{12}$$

where a and b are weighting factors for distances and angles. In a related work [3], these parameters were determined by neural network. Our system empirically determines these parameters as a on 1.0 and b on 0.7. Also, n means the straight pathes for targeting the pocket, $n = 2$ means that the pocket is directly targeted, and $n = 3$ means that the pocket is targeted by one-time cushion bounds. The score E is higher in the case that each distance is shorter and the angle is closer to 180 degree. The score is zero if another obstacle ball is on a ball path as shown in Figure 9. Next, our system simulates ball behavior by giving several speeds to the cue ball. The distribution of balls in each simulation is evaluated by the same way as shown in Figure 9 and (12). As a result, our system provides some desirable shooting ways.

4.2. Pocket

Pocket means that a player hits the cue ball for pocketing a target ball by considering a position of a next target ball. Figure 10(a) represents the ways of making contact with

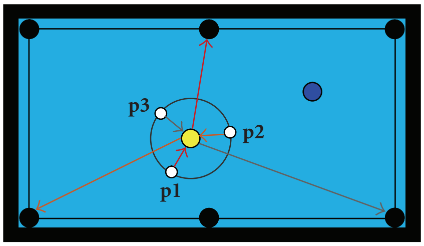
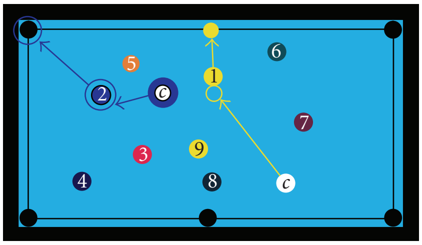
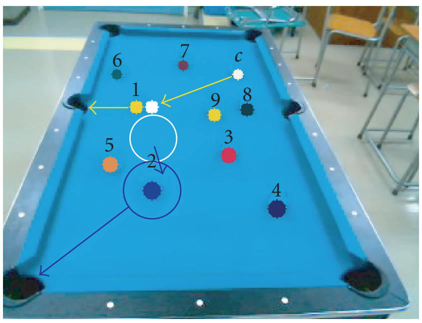


Figure 12: Free Ball.



(a) 2D display



(b) AR display

Figure 13: AR display.

a target ball directly and using cushion. Since our system assumes that ball behavior follows particle dynamics and linear motion and a player hits the center of the cue ball, imaginary points of collision with a cushion and a ball that are dot circles in Figure 10(a) are determined uniquely.

In Figure 10(a), there are two paths of each targeting. Since the paths of pocketing ball number 1 to **P1** are hampered by ball number 2, these two paths are reduced. As for two paths of pocketing ball number 1 to **P2**, the ideal point that the cue ball should be stopped after the shot is searched by simulating ball behavior.

As for simulation of path **A** in Figure 10(a), our system gives the several speed to the cue ball. Figure 10(b) represents two results of the simulation. **c1** and **c2** represents the points where the cue ball stopped in each simulation. As for **c2**, it is difficult to pocket ball number 2 to **P1** because the angle is close to 90 degree. Compared with **c2**, **c1** is better because the ball distribution is straight. In this case, **c1** is a desirable point. Our system performs this analysis on each ball path in a similar way and computes the desirable point that the cue ball should be stopped in each ball path.

(a) Input

(b) Removal of balls

Figure 14: Background.

4.3. Safety

Safety also means that a player hits the cue ball not to pocket a target ball but to make ball arrangement complicated to pocket a target ball for the next player.

In Figure 11(a), it is difficult to pocket ball number 1. For this ball arrangement, a player should target Safety. In Safety, collision point with a target is freely determined such as **p1** and **p2** of Figure 11(a). Figure 11(b) represents the ideal result of Safety. In the case of this ball arrangement, next player cannot target number 1 ball directly. Thus, ball arrangement whose score is zero as mentioned above is a desirable Safety. In Safety, our system searches the case that the average of the evaluation of (12) for each pocket is lowest.

4.4. Free Ball

Free Ball is an event such that a player can put the cue ball anywhere when an adversary player fouls. In Free Ball, the simulation algorithm is the same as Pocket's algorithm assuming that the cue ball is put on the circumference of a target ball such as **p1**, **p2**, and **p3** of Figure 12.

5. Presentation of Visual Aids

Our system provides visual aids of each shooting way by calibrating the camera. Figure 13(a) represents 2D visual aids that are watched above the pool table. Figure 13(b) is a kind of an augmented reality system in which 3D visual aids such as balls and arrows are drawn on the images that a user is capturing, which is called AR display.

For the AR display, a background image is generated from a captured image by replacing ball areas with the color of the green mat as shown in Figure 14.

Visual aids in AR display are drawn based on online camera calibration by the same way as Section 3.2. The coordinate system of supporting information in Section 4 is the same as Section 3.2. For this reason, the world coordinate system of AR display should correspond to the coordinate system of Section 3.2.

In Section 3.2, left below corner is defined as the origin. On the other hand, each corner has a possibility to be the origin because the image is captured from an arbitrary viewpoint. To solve this problem, our system utilizes the estimated ball positions. First, ball positions are estimated assuming that left below corner corresponds to the origin

(a) Viewpoint1 (b) Viewpoint2

FIGURE 15: Online AR display.

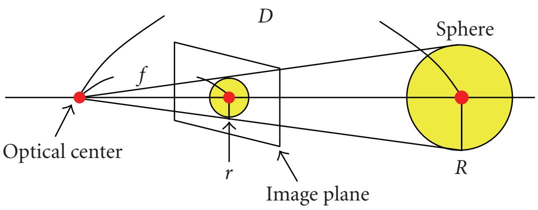


FIGURE 16: Estimation of a ball radius in an image.

which set in Section 3.2. If left below corner is not the origin, ball positions are not equal to the estimated ball positions in Section 3. In a similar way, our system tries estimating ball positions assuming that each corner corresponds to the origin. Our system searched the case that ball positions are equal to the estimated ball positions in Section 3.

Figure 15 represents that a user can see visual aids moving around the table.

As for drawing a ball, our system assumes that a ball is approximated to a circle in the image for drawing a ball. A radius R of a circle is calculated by the ratio of the focal length f from a projection matrix and the distance D between a camera and a ball as shown in Figure 16 and by the following equation:

$$r = \frac{fR}{D}. \qquad (13)$$

6. Experimental Results

6.1. Ball Position Estimation

For evaluating the accuracy of ball positions, the positions measured beforehand are compared with the results of our system. The size of the pool table is 1330 mm × 700 mm which is smaller than the pool table that is used in official games, the radius of a ball is 24 mm and the size of a captured image is 320 × 240 pixels. Development settings are Pentium M(1.0 GHz) CPU and 512 MB RAM capacity.

Figure 17 represents images that captured from different viewpoints. Table 2 represents the comparison between positions measured beforehand and positions estimated by our system. The maximum error is 16 mm, and the average of errors is 10 mm. We consider that this error is mostly caused by the following some reasons. First, our system assumes that a principal point is approximated to the center of a

(a) View1 (b) View2

(c) View3 (d) View4

FIGURE 17: Accuracy of ball positions.

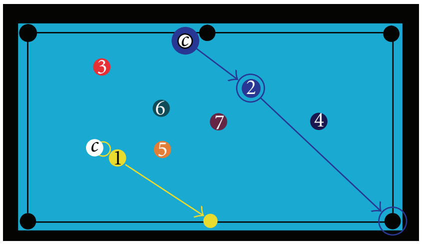
TABLE 2: Estimated results (mm).

Number	c		1	
Coordinate	x	y	x	y
Ground truth	340	404	172	685
View1	340	411	183	691
View2	338	408	173	692
View3	333	398	159	681
View4	352	394	182	678
Number	2		3	
Coordinate	x	y	x	y
Ground truth	389	956	524	815
View1	390	954	531	811
View2	399	948	531	806
View3	385	954	533	817
View4	405	959	531	824

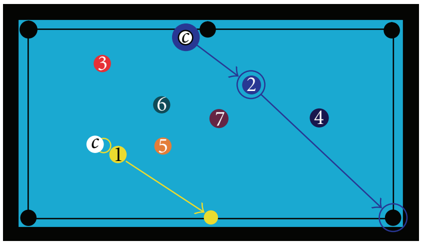
captured image. Second, the center coordinate of a ball area is not center because the ball area includes not only the ball color area but also the border with cushion. The corner positions of the table also have a possibility of including error. The average error is close to 1% of the short-side length of the pool table which is 700 mm. The accuracy is sufficient to provide appropriate supporting information to the users because the error did not influence the computation of supporting information as we indicate in next section.

6.2. Influence of the Error

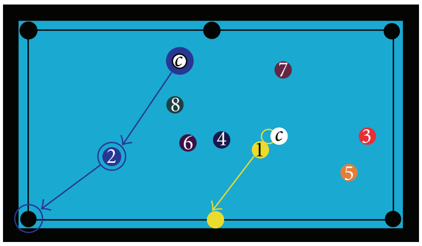
In the previous section, we evaluated that the average error in estimating ball position by the proposed method was about 10 mm. We believe that this amount of error is sufficiently small to provide the user the proper strategy of shooting the cue ball. For testing how this error affects computation of the recommended strategy for the user, we performed



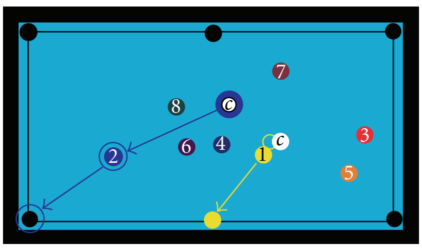
(a) Input1



(b) Match with (a)



(c) Input2



(d) Non-match with (c)

Figure 18: Influence of the error.

computer simulation of computing the strategy based on the ball distribution with 10 mm error for each ball position.

In this simulation, we randomly generate a ball distribution, and then compute the strategy for the generated distribution. We also give 10 mm error with random direction to each ball of the ball distribution, and then also compute the strategy for the distribution with 10 mm error. If the computed strategy for the original distribution is the same as the distribution with 10 mm error as shown in Figure 18(a) and 18(b), we can consider that the error does not affect computation of the strategy. On the other hand, we consider that the error influences to computing the strategy as shown in Figure 18(c) and 18(d).

We tested 50 different ball distributions in Free Ball situation. In those 50 distributions, only two distributions give different strategies for the original distribution and the distribution with 10 mm error. This means that 10 mm error does not give significant influence to compute the strategy from the estimated ball distribution, and it is sufficiently small for the purpose of this paper.

Figure 19: Accuracy of AR Display.

Table 3: Subjective evaluation of free ball.

Arrangement	User (X)	Our System (Y)	Same
A1	0	3	12
A2	3	9	3
A3	0	3	12
A4	2	10	3

Table 4: Subjective evaluation of Figure 20(d).

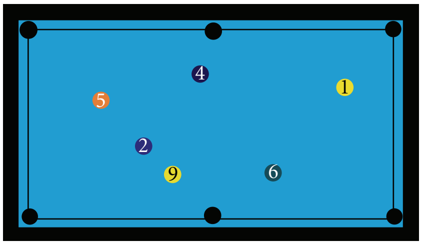
User	Shot1	Shot2	Shot3
2	1	9(2)	3(1)

6.3. AR Display

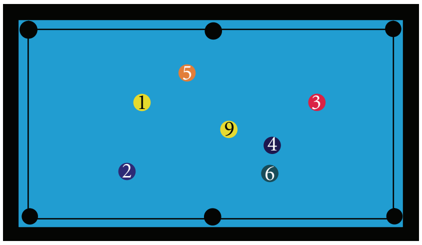
Since the frame rate of AR display is 10 fps, users can see naturally ball behavior drawn on the captured images. For evaluating the accuracy of AR display, the error between projected ball positions and ball positions in the image is computed. In Figure 19, ball positions which are estimated in Figure 17(a) are reprojected on Figure 17(d). The error of the center coordinates between the projected ball and the ball in the image is within 5 pixels on an average. The estimated ball positions include errors as mentioned in Section 6.1. Then, the estimated projection matrix for reprojecting the ball position also includes error. The error of the center coordinates comes from these two error. However, users said that the behavior of the ball which is being pocketed was natural and they did not care about a little error.

6.4. Effectiveness of Supporting Information

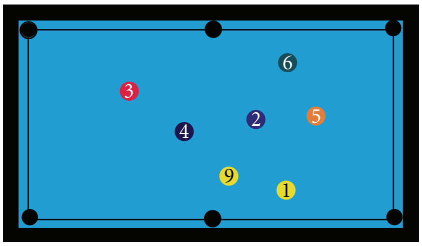
For demonstrating the effectiveness of supporting information, we performed two experiments of subjective evaluation for fifteen beginners who play pool a few times in a year.



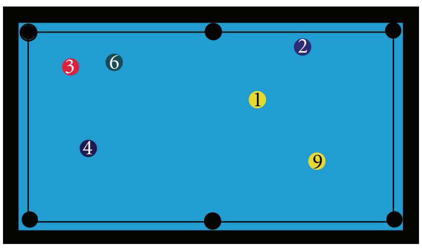
(a) A1



(b) A2

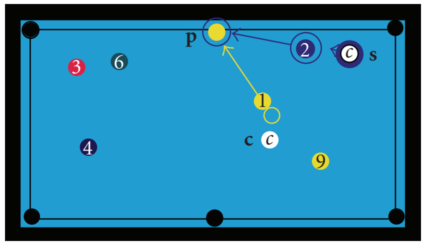


(c) A3

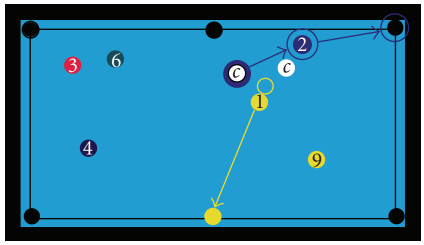


(d) A4

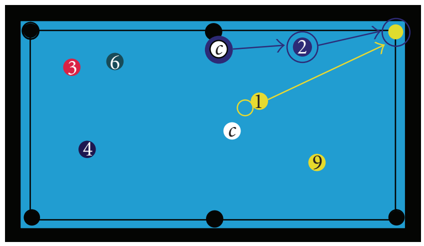
Figure 20: Ball arrangement for free ball.



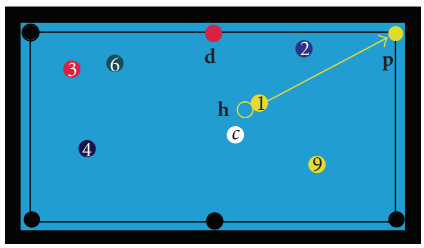
(a) Shot1



(b) Shot2



(c) Shot3



(d) Caution

Figure 21: Supporting information of Figure 20(d).

First, an experiment of Free Ball as mentioned in Section 4.4was performed as follows.

(1) A ball arrangement is given to the examinees.

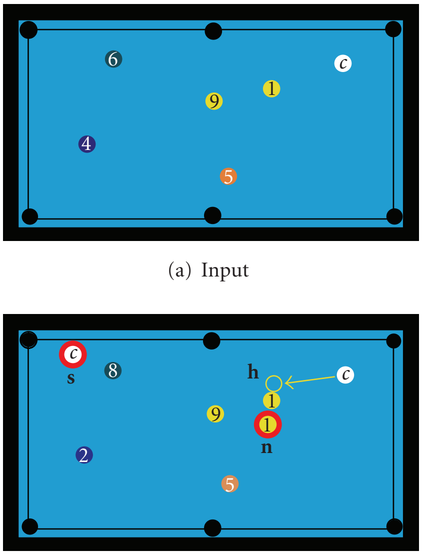
(2) They decide one desirable shooting way (X).

(3) They watch some shooting ways computed by our system (Y).

(4) They select one desirable shooting way from (X) and (Y).

Following sentences are explained by using (X) and (Y).

In Table 3, the term "Arrangement" corresponds to the items of Figure 20, the term "User" represents the number of people who selected their consideration (X), the term "Our System" represents the number of people who selected one of shooting ways computed by our system (Y), and the term "Same" represents the number of people who judged (X) is the same as one of (Y). By counting the number of people, we would like to demonstrate that the shooting ways recommended by our system are actually desirable for the players. This experiment is applied for four kinds of ball arrangement as shown in Figure 20. For each case, the results are shown in Table 3. In the case of (a) and (c), twelve examinees judged that their shooting ways (X) were close to one of the shooting ways of our system (Y). On the other hand, nine examinees in (b) and ten examinees in (d) suggested that one of shooting ways of our system (Y) was more desirable than their shooting way (X). From this table, our system provides shooting ways that are close to the way of



(a) Input

(b) An example of safety

Figure 22: Ball arrangement for Safety.

beginners in the case of simple ball arrangement. Our system can also provide more effective shooting ways than the way of beginners in the case that a shooting way is not simply found.

Figure 21 represents the supporting information of Figure 20(d). In Figure 21(a), **c** is the point that the cue ball should be put, p is the pocket that ball number 1 should be pocketed, and *s* is the point that the cue ball should be stopped for the next shot. In Table 4, each value represents the number of people who selected one of shooting ways as shown in Figure 21. Nine examinees selected Shot2 and seven examinees did not notice this way first. They said that ball number 1 was easily pocketed and ball number 2 was pocketed twice in a row by Shot2.

In addition, our system can provide information of the caution that the cue ball will be pocketed as shown in Figure 21(d). However, **h** is the collision point of the cue ball and ball number 1, **p** is the pocket that ball number 1 should be pocketed, and **d** is the pocket that the cue ball will be put in. Our system can provide this caution because of simulation of ball behavior. In this ball arrangement that is similar with Figure 21(d), there is a possibility that the cue ball is pocketed if ball number 1 is put into **p**. For this reason, Shot3 includes the possibility that the cue ball is pocketed. Every examinee suggested that this information was very helpful because they did not notice whether the cue ball was pocketed or not.

Second experiment is a detection of a shooting ways of Safety as mentioned in Section 4.3. In this experiment, fifteen examinees watched Figure 22(a) and answered whether they could find a shooting way or not. As a result, twelve examinees could not find the way of hitting the cue ball. They commented that they knew Safety but did not know how to do that and learned the way of Safety from our system. Figure 22(b) represents an example of Safety. In fact, **h** is the collision point with ball number 1, **s** is the point that the cue ball should be stopped, and **n** is the point at which ball number 1 should be stopped. It is hard to pocket ball number 1 directly in this distribution of balls and this is a desirable Safety.

7. Conclusions

We have proposed a system for supporting pool game based on analysis of ball positions for NineBall Game. Our system is implemented on a camera-mounted handheld display and does not use special equipment such as a magnetic sensor and artificial markers. A user can capture a pool table from an arbitrary viewpoint and see supporting information and ball behavior which are drawn on the captured images through the display. In our system, ball positions are estimated by computing a projection matrix of the captured image from an arbitrary viewpoint. As for supporting information, our system simulates ball behavior by giving several speeds to the cue ball and provides some desirable shooting ways based on the rule of the pool game. Moreover, supporting information and ball behavior are online drawn on the images while a user is capturing the pool table. In the experimental results, the accuracy of estimated ball positions is enough for analysis of ball arrangement. In addition, fifteen users evaluated supporting information of our system, and the effectiveness of our supporting information was presented.

Acknowledgment

This work is supported in part by a Grant-in-Aid for the Global Center of Excellence for high-Level Global Cooperation for Leading-Edge Platform on Access Spaces from the Ministry of Education, Culture, Sport, Science, and Technology in Japan.

References

[1] M. E. Alian, S. B. Shouraki, M. T. M. Shalmani, P. Karimian, and P. Sabzmeydani, “Roboshark: a gantry pool player robot,” in *Proceedings of the 35th International Symposium on Robotics (ISR ’03)*, Paris, France, March 2004.

[2] Z. M. Lin, J. S. Yang, and C. Y. Yang, “Grey decision-making for a billiard robot,” in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC ’04)*, vol. 6, pp. 5350–5355, The Hague, Netherlands, October 2004.

[3] B.-R. Cheng, J.-T. Li, and J.-S. Yang, “Design of the neural-fuzzy compensator for a billiard robot,” in *Proceeding of the IEEE International Conference on Networking, Sensing and Control (ICNSC ’04)*, vol. 2, pp. 909–913, Taipei, Taiwan, March 2004.

[4] S. C. Chua, E. K. Wong, and V. C. Koo, “Intelligent pool decision system using zero-order sugeno fuzzy system,” *Journal of Intelligent and Robotic Systems*, vol. 44, no. 2, pp. 161–186, 2005.

[5] S. C. Chua, E. K. Wong, and V. C. Koo, “Performance evaluation of fuzzy-based decision system for pool,” *Applied Soft Computing*, vol. 7, no. 1, pp. 411–424, 2007.

[6] T. Jebara, C. Eyster, J. Weaver, T. Starner, and A. Pentland, “Stochasticks: augmenting the billiards experience with probabilistic vision and wearable computers,” in *Proceedings*

of the 1st International Symposium on Wearable Computers (ISWC '97), pp. 138–145, Cambridge, Mass, USA, October 1997.

[7] L. B. Larsen, R. B. Jensen, K. L. Jensen, and S. Larsen, "Development of an automatic pool trainer," in *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '05)*, pp. 83–87, Valencia, Spain, June 2005.

[8] H. Uchiyama and H. Saito, "Position estimation of solid balls from handy camera for pool supporting system," in *Proceedings of the 1st Pacific Rim Symposium on Image and Video Technology (PSIVT '06)*, vol. 4319 of *Lecture Notes in Computer Science*, pp. 393–402, Springer, Hsinchu, Taiwan, December 2006.

[9] H. Uchiyama and H. Saito, "AR display of visual aids for supporting pool games by online markerless tracking," in *Proceedings of the 17th International Conference on Artificial Reality and Telexistence (ICAT '07)*, pp. 172–179, Esbjerg, Denmark, November 2007.

[10] G. Simon and M.-O. Berger, "Reconstructing while registering: a novel approach for markerless augmented reality," in *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR '02)*, pp. 285–293, Darmstadt, Germany, September-October 2002.

[11] Y. Uematsu and H. Saito, "Vision-based registration for augmented reality with integration of arbitrary multiple planes," in *Proceedings of the 13th International Conference on Image Analysis and Processing (ICIAP '05)*, pp. 155–162, Cagliari, Italy, September 2005.

[12] C. Galamhos, J. Matas, and J. Kittler, "Progressive probabilistic Hough transform for line detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, vol. 1, pp. 554–560, Fort Collins, Colo, USA, June 1999.

[13] Intel, http://www.intel.com/.