Artificial Intelligence for Engineering Design, Analysis and Manufacturing (2009), *23*, 305–316. Printed in the USA.

doi:10.1017/S0890060409000262

Play-it-by-eye! Collect movies and improvise perspectives with tangible video objects

CATI VAUCELLE AND HIROSHI ISHII
MIT Media Laboratory, Cambridge, Massachusetts, USA
(RECEIVED April 24, 2008; ACCEPTED January 11, 2009)

Abstract

We present an alternative video-making framework for children with tools that integrate video capture with movie production. We propose different forms of interaction with physical artifacts to capture storytelling. Play interactions as input to video editing systems assuage the interface complexities of film construction in commercial software. We aim to motivate young users in telling their stories, extracting meaning from their experiences by capturing supporting video to accompany their stories, and driving reflection on the outcomes of their movies. We report on our design process over the course of four research projects that span from a graphical user interface to a physical instantiation of video. We interface the digital and physical realms using tangible metaphors for digital data, providing a spontaneous and collaborative approach to video composition. We evaluate our systems during observations with 4- to 14-year-old users and analyze their different approaches to capturing, collecting, editing, and performing visual and sound clips.

Keywords: Children; Interaction Design; Storytelling; Tangible User Interfaces; Video

1. INTRODUCTION

From an early age, we play, learn, and exchange ideas about our identity using stories; we test our hypotheses by playing with toys, telling stories, and acting in the world. Today, communications technology expands our resources for exploring and sharing our reflections on the environment we live in. With mobile technology, we enter a creative and collaborative world where images, sound, and language mix. Shared movie-making devices can engage people in multidimensional approaches to expressing and exchanging points of view. We imagine a world in which, through play, children create and exchange visual narratives about their lives and their environment.

We connect to our world using our senses. Each sense is a "knowledge shopper" that grounds us in our surroundings (Ackermann, 1990): with touch, we feel the texture of life; with hearing, we perceive even the subtlest murmurs of our existence; with vision, we clarify our instincts. We use gesture to apprehend, comprehend, and communicate. We speak to exchange with others. We visualize, record, and play back events using memory to reflect on our history and to immerse ourselves in experience. We engage in everyday pretense and symbolic play. We embed and later withdraw from the world, using imagination to project ourselves into situations (Singer & Singer, 1990).

Children are offered stories by adults and are driven into fantasy play. They use toys to externalize and elaborate their mental constructions (Fein, 1979). Oral stories in children's play make use of linguistic structure (quoted speech, direct speech, narrator voice) and context (providing spatial and temporal expressions in stories); children's ability to use language to communicate when and where their story takes place is considered a milestone in literacy development (Snow, 1983).

Constructionist educators have demonstrated that young users benefit from systems that support self-expression, because children "learn by making" (Harel & Papert, 1991; Resnick, 2002; Ackermann, 1996, 2004). Still, most commercial technological toys do not provide space for children to tell their own stories; rather, the toys tell stories to the children. Video composition tools could support children's authorship; however, the interaction paradigm in traditional video editing systems is a time line; the objective is to make an immutable "final" cut of a movie. The author can only see the whole once she renders the time line. Now, with digital video, we can change the video-making process so that children can easily use video as an expressive composition tool.

Reprint requests to: Cati Vaucelle, MIT Media Laboratory, Tangible Media Group, 20 Ames Street, Cambridge, MA 02139, USA. E-mail: cati@media.mit.edu

Toys initiate elaboration in play and language. Researchers have found a correlation between open-ended play and imagination in writers, poets, and scientists (Singer & Singer, 1990). The 18th century German writer Goethe reported treasuring the puppet theater he had as a child as he envisioned relationships and plots between characters in his later novels. Toys and storytelling serve a fundamental function in childhood development (Montessori, 1912; Singer & Singer, 1990; Brosterman, 1997), and the ability to move from one's own standpoint to take another person's view in a story is at the center of cognitive growth (Piaget & Inhelder, 1967; Winnicott, 1971; Ackermann, 1996). "Dwelling in" and "stepping back" are equally important to get the cognitive dance going (Ackermann, 2004); when a child makes a movie, she *dwells in* as she creates and tells her story and *steps back* as she watches the movie she just made. However, creating movies requires media composition and narrative skills, which existing user interfaces scaffold poorly for novice users and children (Landry, 2008).

Through design we seek to understand how tangible interfaces for composing movies can empower young users in expressing and sharing ideas, actively "constructing" personal movies driven by storytelling. Indeed, "Children build, make or manipulate objects or artifacts and in doing so are confronted with the results of their actions, learning as they go" (Harel & Papert, 1991). Movie editing systems support personal creation and offer opportunities to convey and reflect on "real-world" experiences. Cell phones, video cameras, and computer game consoles could serve as vehicles for manipulating personal media to co-construct video games, movies, and songs. We base our design exploration on a language of interaction that children are familiar with, adopting play interactions to control a video editing system. The goal is to assuage the interface complexities of commercially available editing software. We aim to motivate young users in telling their stories, extracting meaning from their experiences by capturing video elements to accompany their stories, and driving introspective reflection.

2. RELATED WORK

2.1. Tangible user interaction with video

Our work relates to research on tangible user interfaces (Ullmer & Ishii, 2000) that combine physical objects with digital data. Tangible systems have been built to sequence media clips, arranging digital information physically (Jacob et al., 2002), create multimedia stories (Mazalek & Davenport, 2003), access digital information using tokens (Holmquist et al., 1999; Ullmer & Ishii, 2000), and that use multiple handheld computers to organize digital video clips (Sokoler & Edeholt, 2002; Zigelbaum et al., 2007). Tangible mixing tables enable a performance-oriented approach to media construction (Lew, 2004); for example, the DiamondTouch table (Dietz & Leigh, 2001) invites users to collaborate using shared digital media. These systems variously support capturing and editing movie segments, but none allows users to edit, perform, publish, or share personal movies in a self-contained system.

2.2. Tangible user interfaces for children and storytelling

Our work also contributes to tangible storytelling tools for children (Frei et al., 2000; Montemayor et al., 2004). Rizzo et al. (2003) envisioned a system that plays visual sequences using tangible objects. Labrune and Mackay (2005) redesigned cameras to capture both the child and the video the child is making, to contextualize a recorded visual scene. In I/O Brush, children use a paintbrush to gather pictorial information from their surroundings (Ryokai et al., 2004). For example, TellTale invites children to connect story segments through a caterpillar toy (Ananny, 2002). In StoryMat, a childhood map invites children to collaborate as they act out stories using props (Cassell & Ryokai, 2001). More recently, in Jabberstamp, children synthesize their voices in their drawings (Raffle et al., 2007).

2.3. Gesture–object interfaces

The function of gesture is also critical to our work. The movements that one makes with object in hand not only animate that object but also carve out a context, giving a thing a life that is as dynamic as the user can imagine and communicate through gesture. Therefore, to interact with a *gestural object*, one must understand the scope and flexibility of its gestural space. Gestures scale like a language, have different contexts, meanings, and results. For instance, the Nintendo Wii controller alternates between being a character on a screen and a tennis racquet.

Other work has proposed gestural interfaces for children. In Office Voodoo, children move dolls to control parts of a sitcom (Lew, 2003); and in work on sympathetic characters children manipulate a plush toy to control characters in a three-dimensional virtual environment (Johnson et al., 1999). In the vein of wind-up toys, Topobo (Raffle et al., 2004) records motions: users create sculptures that can walk around.

Our previous work explored interactive objects that use gestures to trigger actions. The Dolltalk storytelling system invites children to discover narrative perspectives during storytelling play. It captures, analyzes, and interprets gestures while analyzing changes in voice prosody (Vaucelle & Davenport, 2002). Using sensors and audio analysis, the system interprets the narrative structure of a story.

3. DESIGN ITERATIONS

In this section we present four design iterations of a video editing system for children. The systems are named *Textable Movie, Moving Pictures, Terraria*, and *Picture This!* We implemented each as a separate system and tested it with children, using lessons learned to guide the next design. We begin by explaining how these four systems relate to our larger research interests, and then devote the remainder of this

https://doi.org/10.1017/S0890060409000262 Published online by Cambridge University Press

section to the iterative design and evaluation of the four systems. We conclude with some general conclusions drawn from our experience building innovative video editing systems.

We use a semiotic square (Greimas & Courtés, 1979; Fig. 1) as a framework to organize our four video systems. The terms in the square identify our research areas and the specifics of each research area with regard to the human body.

The left side of the square (*Gesture + No manipulation*), which represents the field of gesture recognition, involves interaction with the hand. The hand plays directly with bits. However, this paper does not describe a project in this category.

The right side of the square (*Manipulation + No gesture*), which represents the field of Tangible User Interfaces as exemplified by Tangible Bits (Ullmer & Ishii, 2000) also involves interaction with the hand. The hand plays with objects that represent bits. Here we place *Moving Pictures*, a tangible representation of media stories for capture, editing, and performing, and *Terraria*, a joystick that directs the composition of video stories.

The top of the square (*Gesture + Manipulation*) represents gesture–object interfaces, which involves gesture recognition during object manipulation: the gestures combine with the objects to represent bits. We position *Picture This!*, a system that allows children to use their toys to make movies while playing with them, in the gesture–object interfaces area.

Finally, the bottom of the square (*No manipulation + No gesture*) represents graphical user interfaces. We position *Textable Movie*, a text-based video presentation system, here.

3.1. Textable Movie: Making a movie by telling a story

3.1.1. Motivation

Textable Movie, our first design prototype, was intentionally not tangible. Rather, it was intended to inform the design of later tangible platforms for making movies. We wanted to provide an alternative to commercially available video editing software, allowing improvisation and unexpected discovery of media content and to make visual storytelling more playful, engaging, and powerful for young people (Resnick, 2006). Our previous research led to the idea that the projectionist, viewer, and maker could use text input to sequence the projection (Vaucelle & Davenport, 2003). However, early testing uncovered a basic limitation: how would the projectionist/viewer know what words to use? Our response in Textable Movie is that players submit and name their own images.

3.1.2. Scenario of interaction

As the user types a story, media segments appear on the screen, generating a movie. Media segments are selected according to how the user has previously labeled audio and video files in their personal collection. Labeling gives each media file a personal meaning for recall. We incorporated commands to add instant computer graphic effects to the movie being played. Textable Movie enables a user to become a "video-jockey" by mixing, applying effects, and rearranging video samples in real time, and it acts as a projection device for a storyteller. It is not a regular editing tool, but a tool for improvisational multimedia storytelling.

3.1.3. Observations

We organized a 1-week workshop in a cultural center in Dublin, Ireland, with children aged 10–14 who wanted to create movies. Adult mentors, professionals in animation, and documentary filmmaking, demonstrated traditional methods of filmmaking and movie styles. The mentors introduced a decomposition of traditional movies into video segments and showed how one can make a movie by assembling clips, comparing the movies that result when clips are mixed in a different order. Participants created a paper-based storyboard, filmed and digitized their raw movie, and finally used Apple

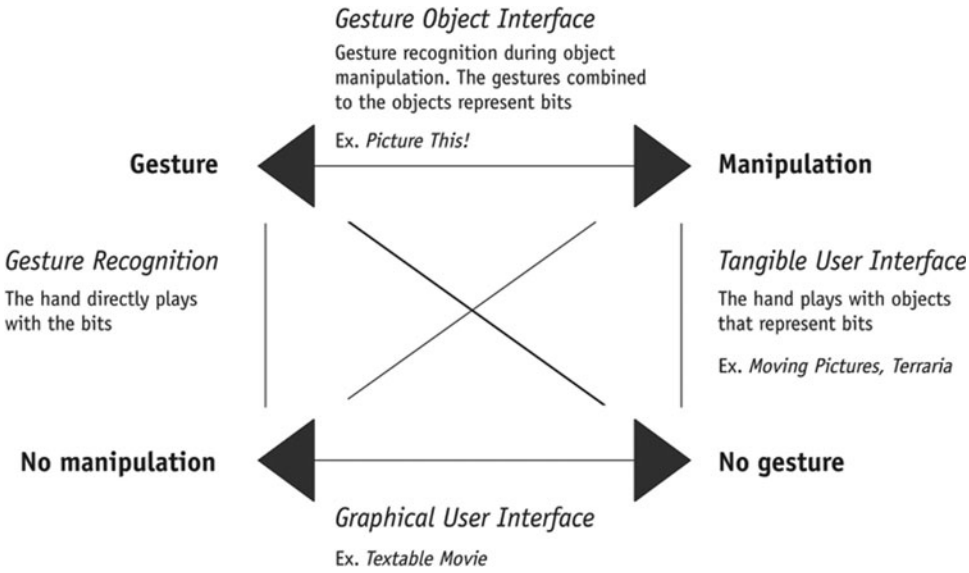

Fig. 1. The iteration design framework.

https://doi.org/10.1017/S0890060409000262 Published online by Cambridge University Press

iMovie™ software to create a palette of movie segments and associated keywords. Children then used the clips in Textable Movie in a visual storytelling performance. We also asked the children to compose a movie from the same video clips using iMovie. Through observations and interviews we analyzed how the children conceived making a movie from planning a video shot, to conceptualizing the editing process and projecting their movie. We also compared how children used Textable Movie and iMovie to compose a movie.

3.1.4. Results

To collect videos, we asked participants to be reporters of their city. The children captured media clips to represent their environment. They were motivated to capture movies, and they respected the content of their storyboard. However, during the editing phase, we lost their attention. When asked to edit their movie to create a final movie in iMovie, more than half of the children stated they preferred continuing the capture process. However, when children were asked to segment and label video segments for the Textable Movie software, they attentively created mappings between text and images. They also composed creative interactions by associating videos with humorous keywords. During the projection phase, they collaboratively created an interactive movie by shouting keywords to type in. The computer keyboard appeared to limit collaborative video making, because only one user at a time could enter the commands offered by the group. The children explored their collective video database, revisiting their keyword matching and recreating video clips as needed. The digitizing and editing phase is necessary for the children to clean their raw data, clarify their video expression and select pieces for use with Textable Movie. If they dropped out of one of these phases, their original vision, as presented in their storyboard, was not followed, and the children did not produce a movie that they were satisfied with.

3.1.5. Lessons learned

Textable Movie reduces the technical difficulties of creating a movie by coupling the performative act of telling a story with editing a final movie. The children's motivation in composing videos with Textable Movie and their telling us that Textable Movie is "more fun because it is more like a game!" reveals a need for an alternative framework in video editing that connects to children's spontaneous play. Textable Movie is not intended to replace iMovie; however, its simplicity of use and immediate response engaged the children in composing a final movie.

Creating a story, acting it out, and making a movie out of it, are three strong motivators for young users to immerse themselves into their environment and later step out of it, observing how it would look from the viewpoint of an audience. We noticed that when the children create a final piece, either an interactive video or a finished movie, they witness their perspectives on their environment, reflect on it with their peers and by doing so are self-critical toward their understanding of the world. Often they ask to revisit their video, shooting clips and remixing them for a final movie. By creating a movie-editing paradigm in which text leads and image follows, Textable Movie provides a natural, fun, and immediate interface to video making. This approach creates a symbiotic relationship between the author's imagination and the stories that she wishes to tell while supporting activities that foster narrative co-construction.

Our next prototype, Moving Pictures, investigated a tangible interface to gather, capture, and edit digital data around the city for later retrieval. In this case, tangible objects become metaphors for captured elements. This physical materialization of a video clip aims to compensate for the lack of an understanding as to how a movie is commonly edited.

3.2. Moving Pictures: A tangible platform for making videos

3.2.1. Motivation

Based on our experience with Textable Movie, and with children as design partners, we implemented a tangible movie making system. This self-contained platform offers children the opportunity to collect video clips from their environment and later compose video using an editing station that provides tangible access to their entire media collection. We aimed to motivate the children to explore the entire process of making a movie. We originally designed Moving Pictures for users aged 10–14. However, because the interaction relies exclusively on manipulating tokens, children as young as 4 years old can play with the system and interact with video clips. To accommodate various age groups and individual characteristics of users, we integrated different layers of complexity, from digitizing the media, performing a movie, to storyboarding a more complex narrative, similar to the video-making process during the Textable Movie workshop.

3.2.2. Scenario of interaction

Moving Pictures is a table top with three radio frequency identification (RFID) readers, a laptop computer, a set of speakers, a display, two cameras built into PDAs with RFID capabilities, and a collection of RFID tokens (Fig. 2). Recorded media is associated with a digital ID and a physical token. The PDA wirelessly sends the mapping between token ID and media to the computer as well as the media files. The computer receives the information and plays the appropriate video or sound segment.

By offering a tangible representation of media elements, Moving Pictures transforms single-user screen-based media sequencing into multiuser physical interaction, adding a collaborative dimension as a direct response to the limiting use of a keyboard in Textable Movie. Conventionally, movie editing consists of assembling short video segments with a soundtrack that unifies the visual composition. In Moving Pictures, users apply sound effects to movie sequences.

https://doi.org/10.1017/S0890060409000262 Published online by Cambridge University Press

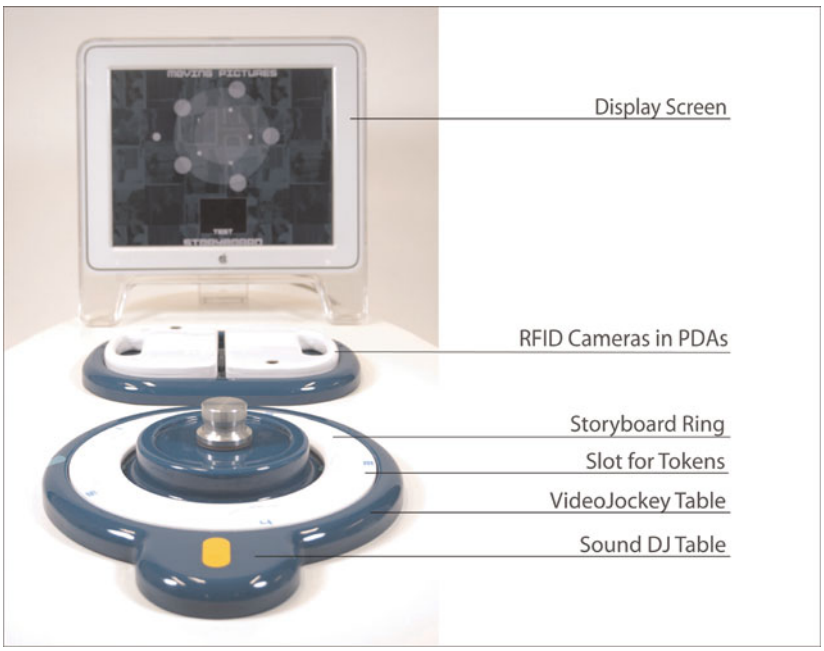

Fig. 2. The final design of *Moving Pictures.* [A color version of this figure can be viewed online at journals.cambridge.org/aie]

Sounds can overlap, or be individually scratched. The soundtrack is recorded as it is performed. The length of the captured movie that can be embedded in the form of a tangible metaphor is limited because a token symbolizes a single shot. Our previous observations with children led to three seamlessly integrated functions.

1. To capture video, users insert a token in the camera, which records a shot.
2. To perform video and sound, once removed from the camera, the tokens are composed on the interactive table. Users place the camera on the table and the material collected is transferred to the computer. Users improvise video compositions using the tokens, and the clips play on the display.
3. To edit videos, five tokens are inserted at a time on the storyboard ring. Rotating the ring on the table plays the corresponding video clips sequentially. When the children are satisfied with the video composition, they export their movie on green tokens. These green tokens can be assembled altogether to construct a longer movie.

3.2.3. Observations

Over a period of 8 months we applied a participatory design approach to implement a functional prototype of Moving Pictures with children age 10–12 as partners (Vaucelle, Africano, et al., 2005). We built a variety of prototypes, including the camera and the final editing, mixing, and performing table (Fig. 3).

To evaluate our system, we organized a 1-week workshop in Ireland with children aged 10–14. Children were motivated and attentive in using the tokens to capture their movie and perform an interactive movie. They were focused during the editing phase to create a final movie. They were careful with the length of the captured clips, which enabled them to practice limited rules in standard video editing without being too conscious of them. Half the children understood the general interaction with Moving Pictures and actively used the tokens for data retrieval.

During the sessions with Textable Movie and Moving Pictures, the children influenced and learned from each other. The collaborative dimension in the physical manipulation of tokens connected children into assessing and testing each other's knowledge and understanding of movie making.

Fig. 3. Camera and table first prototypes versus final models. [A color version of this figure can be viewed online at journals.cambridge.org/aie]

https://doi.org/10.1017/S0890060409000262 Published online by Cambridge University Press

Some children chose to spend more time arranging video clips and adding corresponding sounds, eventually becoming experts at this task. Others specialized in acting or in camera techniques. Throughout the workshop the children created a series of movies. Movie stylistic choices varied from journalistic interviews that were limited to 5 shots; explorations in the city using more than 10 shots; 5 individual shots of the children acting in front of their favorite city place; a more sophisticated 5-shot story with a beginning, a middle, and an end; and a theater play using 10 shots (Vaucelle & Ishii, 2007). The 5-shot story made the most of our tangible environment. The story required a storyboard and it required revising the captured shots. It also engaged children in testing different outcomes using the same shots and in overlapping sounds to create continuity within the soundtrack. Finally, the story became three stories with different endings.

3.2.4. Results

Observing the creative process of the children working on digital media with Moving Pictures, we found that they exhibited the four aspects of *Understanding of the Arts* proposed by Ross and reintroduced by Somers (2000)*: conventionalization*, an awareness and ability to use the conventions of the art form; *appropriation*, embracing, for personal use, the available expressive forms; *transformation*, the search for knowledge and meaning through the expression of "feeling impulses;" and *publication*, placing the result in the public domain. Using Moving Pictures, children made a movie using a series of traditional shots symbolized by physical tokens. They respected their storyboard and they contributed to a video database by expressing their visual narratives for another group of children. Video-jockeying is a spontaneous way to perform final pieces and to integrate selected sounds. It became the physical translation of the projectionist in Textable Movie. Children were engaged in producing all the video stories they created from initial capture to editing their final pieces. Having the digital data represented by physical objects helped the children understand the construction of their movies. Moving Pictures succeeded in engaging children in the entire movie-making process. However, it lacked scaffolding from the children's oral storytelling.

3.2.5. Lessons learned

With Textable Movie, we built a text-based video presentation system that led to its tangible counterpart, Moving Pictures. Our goal was to engage children in editing a final movie in addition to performing an interactive video story. Without a physical metaphor to abstract the editing process, we lost our participants during editing. In evaluating the tangible Moving Pictures system, we noticed that children could only record a limited number of shots at one time, but we witnessed their engagement in manipulating physical objects to interact with and edit digital content. We saw that a video-making system could become closer to the object of attention, for instance a character, a scene, a landscape, with a newly defined interaction technique. In the next iteration, we expanded the idea of connecting video editing to children's spontaneous play, focused on manipulating a single controller.

3.3. Terraria: Real-time video making about toys

3.3.1. Motivation

Computer game controllers, for example, joysticks, can serve to manipulate personal media. We hypothesized that because of children's familiarity with their everyday toys and games, children could be drawn into video making with a joystick. Our next design iteration, *Terraria*, employed a joystick for video capturing, editing, and performing. The joystick controls camera angles, recording, video and sound effects, playback, and projection of the final movie onto a screen.

3.3.2. Scenario of interaction

Terraria's interaction model is based on children's play: playing with toys, adding voices, turning toys into characters, and enabling children to capture their play on video (Vaucelle, Gorman, et al., 2005). Terraria consists of four landscapes with robot props, four video cameras, four joysticks, and five wireless networked computers (Fig. 4). We installed the system for 3 months as part of an exhibition (the system can also be used at home). Young visitors were invited to make movies and to decorate the exhibition space with their interactive creations. The exhibition space forced us to make the tangible video system robust enough to support varied timeframes of use, experimentation, and improvisation of well-structured, sequenced, and live-captured video.

3.3.3. Observations

We first organized pilot studies with 8- to 12-year-old children; later, we exhibited Terraria for 3 months during which we observed children from 4 to 14 years old playing with the system. We saw young users capture and edit their visual stories, prepare the automated toy robot actors, insert audio and visual effects, and soundtracks by selecting songs from a database. The young users found this integrated interface engaging for performing movies in real time.

3.3.4. Results

Children were drawn to give a visual life to their robotic toys. They spent an average of 1 h playing with the robots, recording movies about them and projecting the movie. The exhibition curator reported that the system was a success and by far the most visited and played with exhibit at the museum. The simplicity of use and immediacy of response seemed to engage visitors in creating movies. Both during our user studies and during the exhibition, users recorded videos, and selected soundtracks to fit with their videos and to unify their composition.

3.3.5. Lessons learned

Children captured videos of their toys, selected visual angles, integrated objects, discovered strategies for animation. However, they did not act out social interactions between toys as

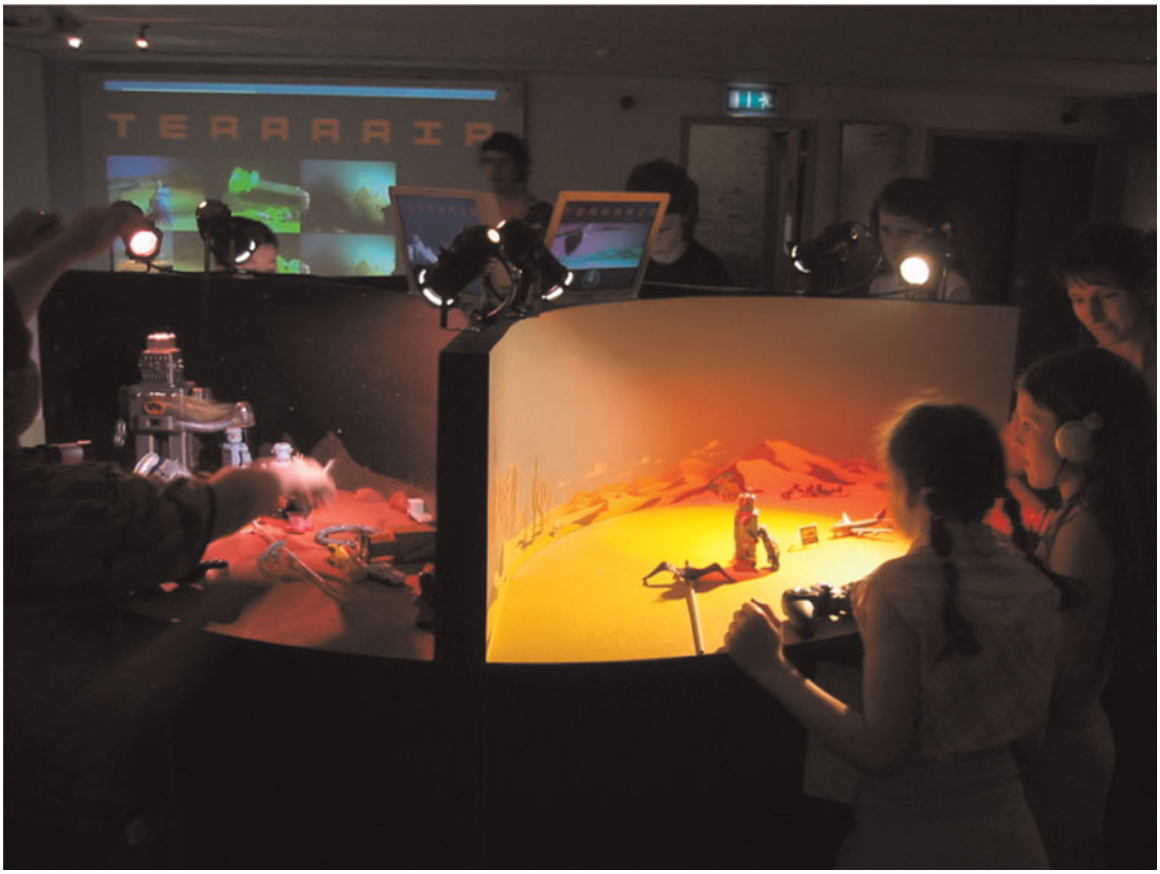

Fig. 4. The *Terraria* landscapes and the projection screen during the exhibition. [A color version of this figure can be viewed online at journals.cambridge.org/aie]

in oral storytelling. This major failure in our system indicated that children focused their attention on the joystick, enjoying grabbing visual components with it, alternating their attention between the joystick and the visual scene. We believe that the children's focus on the manipulatory interface, in this case the joystick, distances them from embedding themselves into the toys. Thus, in Terraria we miss the important component of dwelling in and stepping back from a story, alternating the perspectives of the actor, narrator, and audience, and expressing with words the meaning of a visual scene. Terraria was our introduction to the demands of play in tangible video editing but it failed to engage children in combining their visuals with storytelling. Although we succeeded in motivating children into spending hours making movies, we failed to open the rich space of storytelling within movie making.

3.4. Picture this! Making a movie as the extension of natural play

3.4.1. Motivation

Our next design iteration, *Picture This!*, offers a comprehensive application beyond the scope of assembling visual scenes. We wanted to motivate children to use their toys to tell a story while assembling a movie. We decided to explore video capture from the toy perspective, to create unexplored visual perspectives and to merge storytelling and play to construct movies.

Textable Movie revealed a need for a more interactive form of video editing, Moving Pictures enabled children to create a movie from data capture to making a final piece, and Terraria allowed children to stay in their world of play with toys and robots while making a movie. Picture This! is a video editing tool leveraging the child's natural expression of play while telling stories with their toys.

With character toys, children create interrelationships and plots, a means to expose their social knowledge: knowledge about human beings and social relationships (Shantz, 1975). A toy with an immediately accessible visual perspective opens a new world to the child. The toy brings her into exploring visual and narrative perspectives of character props, expanding the discovery of her environment. The child storyteller enters the world of the movie maker. Cameras become part of a toy system showing how things look from a toy's point of view. They can be integrated in Lego people, car drivers, and even coffee mugs!

3.4.2. Scenario of interaction

Children make gestures while playing with toys, but current systems do not benefit from this. Children project themselves

https://doi.org/10.1017/S0890060409000262 Published online by Cambridge University Press

onto their toys, embedding persons they know in their stories and character toys, adopting a "God's eye view" to obtain a deeper understanding of their own stories. Picture This! offers a gesture language for capturing and editing suitable for children in their toy environment. The children alternate between being actors and movie makers, orchestrating the scene with their favorite props. The playback mode invites children to revisit their movie; as they "step away from their performance" children reflect on the outcome of their spontaneous play and character's conversations. Picture This! invites children to practice spatial cognition by imagining the toy's viewpoint, trying it out and correcting it. Rather than the child holding a camera directly, the toy becomes a camera person (see Fig. 5a). As a child plays with the toy that holds the camera, projecting its video feed on a screen in real time (see Fig. 5b).

This visual flow aims to motivate her in composing a movie as she plays and explores her visual story. As two dolls interact, they alternate between their respective visual scenes. The child creates a conversation using direct speech for the toy characters. The child also uses a narrator voice to introduce the story and contextualize the scene. We chose the interaction to function like a performance to avoid breaking the flow of traditional pretend play with character toys. Our system incorporates the child's gestures with the cameras and toy's accessories as control functions to assemble the movie (Fig. 6).

Picture This! consists of two toys, each with an attached accessory bag that contains a microcontroller, a piezo vibration sensor, a printed circuit board, and a video camera with a USB connection. The microcontroller in each toy detects gestures and communicates them to our program, which continuously retrieves the microcontroller's output. We developed a filtering algorithm for gesture recognition that detects and interprets angles of motions (Vaucelle & Ishii, 2008). The software identifies natural character play movements, such as jumping and shaking, adding video control functions to these character play movements. The motions the system detects are anthropomorphized; for instance, the dolls jump together at completion and shake for attention, as if the doll wants to say: "film me, film me!" To play the movie she just created, the child must move the two dolls in synchrony, jumping horizontally together. The software automatically sequences video clips and removes blurry frames from the gesture commands and plays the movie for the child on the display. To master interaction with Picture This! the child must alternate between projecting herself onto her toys and directing the scene.

3.4.3. Observations

We observed eight children aged 4–10 using Picture This! to create movies with their toys. They interacted with our system at their home, or if they requested, at our research laboratory. The children brought their own character toys, to be fitted out with our system and to record a movie during play. After playing with the toys we provided, children selected their favorite toys from their bag or from their bedroom to be used with Picture This! First, children explored the system without explanation. After 5 min, a researcher explained how to operate the recording and playback controls. The children were invited to play as long as they wanted; they worked independently for between 45 min and 2 h, and their interactions were videotaped and transcribed for analysis.

Children were extremely methodical and attentive with the video. While in pretend play, they sometimes stopped their story and carefully worked on their camera view angle, alternating between characters. They progressed from capturing the doll in the picture, to framing a full shot of the doll, integrating specific backgrounds, discovering camera distortions and various camera angles, all facilitated by the size and context of the camera. Children under age 6 seem to forget about the screen, being exclusively immersed in their play. When some of the children removed the camera from a proposed character toy, they always attached it to another one. They did not use the camera detached from the toy. They were keen to explore the toys' perspectives. They found playlike justifications for the wires. One child said, regarding a rubber band from the camera that covers half the face of his toy: "well it's kind of normal, 'cause they wear something in front

(a) (b)

Fig. 5. (a) The toy is the camera person versus (b) what the toy "sees" from "his" video feed. [A color version of this figure can be viewed online at journals.cambridge.org/aie]

https://doi.org/10.1017/S0890060409000262 Published online by Cambridge University Press

Fig. 6. Mike (8 years old) playing with *Picture This!* [A color version of this figure can be viewed online at journals.cambridge.org/aie]

of their mouth sometimes. Like a mask!" Children older than 8 years mastered the full system, coordinating dolls to control the video, understanding the interaction between preview, recording and playback. After 20 min of playing, the gestures with the dolls became parts of the children's vocabulary.

3.4.4. Results

Although we have not performed a controlled study to validate our qualitative observations, Picture This! seems to allow children to capture storytelling with physical artifacts at different levels of interaction. Functionalities and mode of interaction could be distinguished with a specific cognitive goal for each age group. For children under 6 years old, Picture This! functions as a video performance system with video snippets of the child's play, with only one of the two toys carrying a camera. The preview seems to help them develop spatial–visual coordination while playing with their favorite toys and telling stories. Picture This! allows older children to test visual angles and assemble a movie as they play with their toys and tell stories alternating between direct speech and narrator voice, providing spatial and temporal context. The recording and playback modes seem to enable older children to use their social perspective taking visually and through storytelling.

The youngest children (under 8 years old) transferred their personal characteristics into the toys. For instance, a doll dances because the child takes dancing lessons. Or a doll takes her first picture, because this is the first time the child takes a picture herself. Another child shakes the doll while saying: "Shake! Shake! *I* want to be in the camera!" and she shakes her own body. Older children (over 8 years old) talked to the dolls, giving directions for the movie. A child brought a doll to her face, as if the doll had a mind of its own, to say, "You don't carry your wand like that. You don't put the wand at people like that!" Children navigate from transferring their own lives onto their toys and attributing human characteristics to the toys. All the children in our evaluation developed spontaneous conversations between the character toys, testing their social knowledge and perspective taking. The following is an excerpt of a video story by Jeremy, 10 years old (see Fig. 7 for missing dialogue):

D1: "Hi! My name is Fred what's yours?"
. . .
D1: "Over there in the great yellow mountains, but there is a giant blocking the way. We need to take down the giant so that we can find the treasure."
D2: "sounds good to me, when do you wanna go?"
D1: "how about right now?"
D2: "ok let's go" Narrator voice: and they walked off to the mountains to destroy the giant and get the Peruvian treasure.
D1: "tutututututut" (walking the dolls though the yellow mountains.) Then in front of the giant, the child says with the doll in the video frame:
D2: "hey you evil sid cops, surrender! Face the rest of us! We are superior and strong! We shall take you down!" Then the child uses one of the two dolls to take a video of the giant and says (taking the voice of the giant)
Giant: "I shall take you down first, face the rest of me!"

Playing with video character toys in Picture This! allows children to develop visual perspective taking skills. This entails, for example, determining where objects are located

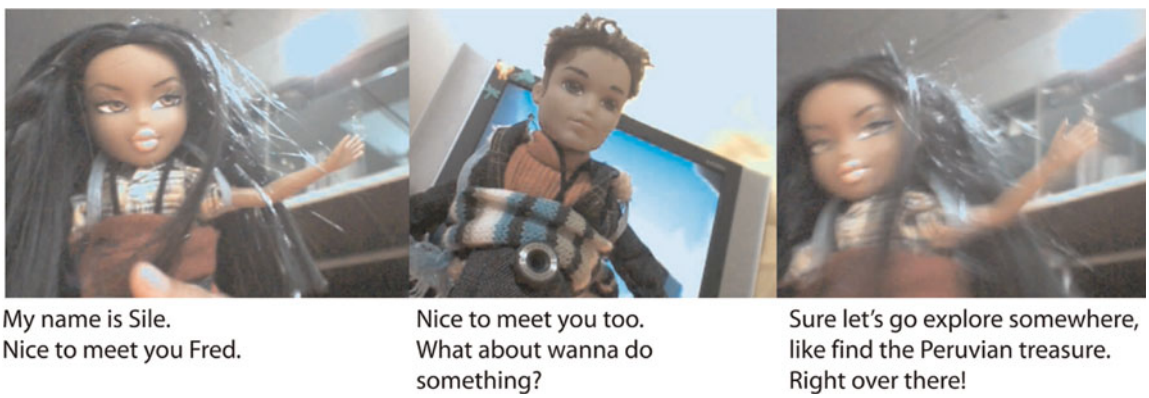

Fig. 7. An excerpt of a movie made with *Picture This!* [A color version of this figure can be viewed online at journals.cambridge.org/aie]

https://doi.org/10.1017/S0890060409000262 Published online by Cambridge University Press

relative to another agent, or whether the agent can see a particular object (Michelon & Zacks, 2006). The high level of concentration the children exhibited demonstrates how challenging it is to find the right angle and distance between the object and the camera, and between the two objects.

3.4.5. Lessons learned

Picture This! modifies the traditional camera–human relationship. The perspective effort needed is demonstrated through spatial and visual coordination, managing the right angle for the right doll at the right moment in time, while acting out a story with the toys. The movies' focus on characters guides children toward creating a conversation, which provokes a shift in perspective (Ziegler et al., 2005). Children have an object to focus on, which allows them to iterate back and forth, stepping back from the scene and immersing themselves into it. Children gradually project themselves onto their toys, embedding persons they know in their stories and character toys, and adopting a "God's eye view" to obtain a deeper understanding of their own stories. The children alternate between being actors and movie makers, orchestrating the scene with their favorite props. The playback mode in Picture This! invites children to revisit their movie; they "step away from their performance" and reflect on the outcome of their spontaneous play and character's conversations.

Visual spatial processing guides our movement. Picture This! invites children to practice spatial cognition—the ability to mentally manipulate objects, and imagine how an object would appear if moved (Henderson et al., 1999)—by imagining the toy's viewpoint, trying it out and correcting it. Children were motivated to see "how it looks like out of a toy's eyes" and they wanted action figures to take video at their home; to make Lego people with an eye socket to hold the camera; to mount the Picture This! system on a racing car to capture the driver's view; and they wanted a waterproof version of Picture This! to capture videos under water with bath toys. All children were keen to keep the Picture This! camera on their favorite toys instead of removing the camera system separately from a character prop.

Playing is about spontaneity and improvisation, while editing a movie is about structure and composition. Movie making can have a bit of both. For this last design iteration, we chose a gesture-based interaction for movie making because of its advantage to integrate well with play. Picture This! trades off movie making with role playing. Its gesture-based interaction invites the discovery of unique angles and point of view, facilitating the movie making flow. Picture This! invites children to experiment with movie editing while playing with their toys. It works as a new mode of video expression and creation through which children are drawn to explore unique visual and storytelling perspectives.

4. CONCLUSION

We have described four design iterations of a tangible video editing system for children. We began with preliminary workshops, where we found that most children preferred recording a video to editing with commercial software. To motivate children to make a movie from beginning to end and to remain focused during video editing we designed strategies for interacting with media content. Our strategies strike a compromise between the powerful capabilities of commercial editing software and the goal of engaging children in making video.

Our first design iteration, Textable Movie, avoids the technical difficulties of commercial editing video software by coupling the task of editing a movie with the performative act of telling a story. Our evaluation of Textable Movie suggested a framework for video editing and storytelling, motivated by playful improvisational storytelling. So in Moving Pictures we aimed to interface video capture, editing and publication, using a tangible element to view, revisit, share, and collaborate on video sequences. Moving Pictures helped children improvise and perform movies collaboratively. Using tokens to retrieve video clips focuses children on editing so they follow their original vision from capturing to editing a final movie. Children are familiar with playing with toys, play-acting character discussions, and enacting toy interrelationships and stories, so in Terraria we used toys to focus children's attention on video composition. Creating video with toys and a game controller, we place children in a familiar realm. The results were pronounced: children spent hours creating, editing, performing movies with the robot toy performers, and projecting their pieces. However, one important component was still missing: children did not tell stories; instead, they merely assembled visual scenes. By combining improvisation with movie making during play, Picture This! extends play to creative outcomes. Automating editing with gesture object interaction allows a child to focus on an object in a captured scene, for instance, a specific character. The video-making process, supported by gesture-induced editing, helps children practice social relationships and take visual perspectives, expanding creative storytelling in video composition.

These design iterations facilitate movie making and engage children in editing a final piece. Except for iMovie and Textable Movie, which do not include an integrated system to capture movies, all our systems offered the following modes of interaction: capturing, editing, performing/playback, composing a final movie, conducting storytelling from the first-person perspective, and storytelling using a narrator voice.

We have synthesized our observations of children's interactions (aged 8–14) with our four movie-making systems and the iMovie video editing system (see Table 1).

Compared with commercial video editing systems such as iMovie, our systems enable children to embed their stories in movies, or to drive their movies by telling stories (in both Textable Movie and Picture This!). Given a choice between an interactive visual storytelling system and the powerful iMovie video editing software, fewer than 50% of the children chose to play back their movies with iMovie. For both Textable Movie and iMovie, fewer than 50% exported a final movie. All the other systems engaged more than 80% of the

Table 1. *Observed interactions with our four interfaces and iMovie with children ages 8–14*

	Capturing	Editing	Playback	Final Movie	Storytelling Perspective	
					1st Person	Narrator Voice
iMovie	No	**Yes**	No	No	**Yes**	No
Textable Movie	No	**Yes**	**Yes**	No	**Yes**	**Yes**
Moving Pictures	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	No
Terraria	**Yes**	**Yes**	**Yes**	**Yes**	No	No
Picture This!	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

children to create and play back a movie. We expected children to conduct storytelling with Terraria using their toys, but fewer than 50% did so, and they used the system only to assemble videos and add visual and audio effects. Terraria did not support storytelling; however, this technology required the least instruction. Children did not narrate when recording with Moving Pictures and iMovie, which would have provided an oral context (spatial and temporal) to their visual stories. While interacting with Textable Movie and Picture This!, participants told stories both from first-person and narrator perspectives. Even though Picture This! was designed to drive movie making by conversational storytelling between toys, children older than 8 years spontaneously integrated a narrator. Implemented sequentially, we learned from each iteration as we moved from a computer–screen–keyboard to a gesture–object-based interface for video expression.

We presented our video-making framework, motivated by the playful improvisational environment of storytelling and integrating tangible technology into video editing systems in the form of toys. We stated the need for a new category of video-editing tools leveraging the child's natural expression of play. Tangible editing systems can engage children in an entirely new video making activity, gaining visual perspectives, driving play, and expanding discovery of their environment. In our tangible movie-making systems, children create story content for editing and performance, and they learn to make a movie "as they go on with their storytelling." With our four iterations of movie-making devices, we redressed a limitation in commercially available video editing software by motivating children to create a final piece. By successfully completing a movie, children can then reflect on the finished piece.

ACKNOWLEDGMENTS

Many thanks to the children, parents, and teachers who contributed to this work in Dublin, Ireland, Umeå, Sweden, and Cambridge, Massachusetts, USA. Thanks to the Media Lab Europe in Ireland where Textable Movie and Moving Pictures was initiated; Trinity College Dublin, CRITE, where we developed Terraria; the Ark in Dublin, Ireland, for organizing the workshops with us; Glorianna Davenport for her advice on Textable Movie and Moving Pictures; Diana Africano and Oskar Fjellström, who contributed to Moving Pictures; and Michael John Gorman, Andrey Clancy, and Brendan Tangney for their support on Terraria. We are also grateful to Adam Boulanger and Edith Ackermann for proofreading the paper and for their valuable insights, and Ellen Yi-Luen Do and Mark D. Gross for editing the final manuscript and for guiding us throughout the journal process. We thank our reviewers, the members of the Tangible Media Group, and the MIT Media Lab for their feedback. We acknowledge the Things That Think Consortium and other Media Lab sponsors for their support.

REFERENCES

Ackerman, D. (1990). *A Natural History of the Senses*. New York: Random House.

Ackermann, E. (1996). Perspective taking and object construction. In *Constructionism in Practice* (Kafai, Y.B., & Resnick, M., Eds.), pp. 25–36. Mahwah, NJ: Erlbaum.

Ackermann, E. (2004). Constructing knowledge and transforming the world. In *The Future of Learning* (Tokoro, M., & Steels, L., Eds.), pp. 15–35. Amsterdam: IOS Press.

Ananny, M. (2002). Supporting children's collaborative authoring: practicing written literacy while composing oral texts. *Proc. CSCL 2002*.

Brosterman, N. (1997). *Inventing Kindergarten*. New York: Harry N. Abrams.

Cassell, J., & Ryokai, K. (2001). Making space for voice: technologies to support children's fantasy and storytelling. *Personal Technologies 5(3)*, 203–224.

Dietz, P., & Leigh, D. (2001). DiamondTouch: a multi-user touch technology. *Proc. UIST '01*, pp. 219–226. New York: ACM Press.

Fein, G.G. (1979). Play and the acquisition of symbols. In L. Katz (Ed.), *Current Topics in Early Childhood Education*, Vol. 2, pp. 211–212. Norwood, NJ: Ablex.

Frei, P., Su, V., Mikhak, B., & Ishii, H. (2000). Curlybot: designing a new class of computational toys. *Proc. CHI '00*, pp. 129–136. New York: ACM Press.

Greimas, A.J., & Courtès, L. (1979). *Sémiotique—Dictionnaire Raisonné de la Théorie du Langage I*. Paris: Hachette.

Harel, I., & Papert, S. (1991). *Constructionism*. Norwood, NJ: Ablex.

Henderson, A., Pehoski, C., & Murray, E. (1991). Visual–spatial abilities. In *Sensory Integration* (Bundy, A.C., Lane, S.J., & Murray, E.A., Eds.), pp. 123–140. Philadelphia, PA: F.A. Davis.

Holmquist, L.E., Redström, J., & Ljungstrand, P. (1999). Token based access to digital information. *Proc. Handheld and Ubiquitous Computing, Lecture Notes in Computer Science*, pp. 234–245. Springer: Berlin.

Jacob, R.J., Ishii, H., Pangaro, G., & Patten, J. (2002). A tangible interface for organizing information using a grid. *Proc. CHI '02*, pp. 339–346. New York: ACM Press.

Johnson, M.P., Wilson, A., Blumberg, B., Kline, C., & Bobick, A. (1999). Sympathetic interfaces: using a plush toy to direct synthetic characters. *Proc. CHI '99*, pp. 152–158. New York: ACM Press.

Labrune, J.B., & Mackay, W. (2005). Tangicam: exploring observation tools for children. *Proc. IDC '05*, pp. 95–102. New York: ACM Press.

Landry, B.M. (2008). Storytelling with digital photographs: supporting the practice, understanding the benefit. *Proc. CHI '08*, pp. 2657–2660. New York: ACM Press.

Lew, M. (2003). Office voodoo: a real-time editing engine for an algorithmic sitcom. *Proc. SIGGRAPH '03*, New York: ACM Press.

Lew, M. (2004). Live cinema: an instrument for cinema editing as a live performance. *Proc. SIGGRAPH '04*, New York: ACM Press.

Mazalek, A., & Davenport, G. (2003). A tangible platform for documenting experiences and sharing multimedia stories. *Proc. ETP '03*, pp. 105–109. New York: ACM Press.
Michelon, P., & Zacks, J. (2006). Two kinds of visual perspective taking. *Perception and Psychophysics 68(2)*.
Montemayor, J., Druin, A., Chipman, G., Farber, A., & Guha, M. (2004). Storyrooms and playsets: tools for children to create physical interactive storyrooms. *Computers in Entertainment 2(1)*, 12.
Montessori, M. (1912). *The Montessori Method*. New York: Frederick Stokes Co.
Piaget, J., & Inhelder, B. (1967). The coordination of perspectives. In *The Child's Conception of Space*, pp. 209–246. New York: Norton & Co.
Raffle, H., Parkes, A., & Ishii, H. (2004). Topobo: a constructive assembly system with kinetic memory. *Proc. CHI '04*, pp. 647–654. New York: ACM Press.
Raffle, H., Vaucelle, C., Wang, R., & Ishii, H. (2007). Jabberstamp: embedding sound and voice in traditional drawings. *Proc. IDC '07*, pp. 137–144. New York: ACM Press.
Resnick, M. (2002). Rethinking learning in the digital age. In *The Global Information Technology Report: Readiness for the Networked World*. New York: Oxford University Press.
Resnick, M. (2006). Computer as paintbrush: technology, play, and the creative society. In *Play = Learning: How Play Motivates and Enhances Children's Cognitive and Social–Emotional Growth* (Singer, D., Golikoff, R., & Hirsh, P.K., Eds.), pp. 192–206. New York: Oxford University Press.
Rizzo, A., Marti, P., Decortis, F., Rutgers, J., & Thursfield, P. (2003). Building narratives experiences for children through real time media manipulation: POGO world. In *Funology: From Usability to Enjoyment* (Blythe, M.A., Overbeeke, K., Monk, A.F., & Wright, P.C., Eds.), pp. 189–199. Amsterdam: Kluwer Academic.
Ryokai, K., Marti, S., & Ishii, H. (2004). I/O brush: drawing with everyday objects as ink. *Proc. CHI '04*, pp. 303–310. New York: ACM Press.
Shantz, C. (1975). The development of social cognition. *Review of Child Development Research* (Heterington, E.M., Ed.), Vol. 5. Chicago: University of Chicago Press.
Singer, D., & Singer, J. (1990). *The House of Make Believe: Children's Play and the Developing Imagination*. Cambridge, MA: Harvard University Press.
Snow, C.E. (1983). Literacy and language: relationships during the preschool years. *Harvard Educational Review 53*, 165–189.
Sokoler, T., & Edeholt, H. (2002). Physically embodied video snippets supporting collaborative exploration of video material during design sessions. *Proc. NordiCHI '02*, pp. 139–148. New York: ACM Press.
Somers, J. (2000). Measuring the shadow or knowing the bird. Evaluation and assessment of drama in education. In *Evaluating Creativity* (Sefton-Green, J., & Sinker, R., Eds.), pp. 107–128. London: Routledge.
Ullmer, B., & Ishii, H. (2000). Emerging frameworks for tangible user interfaces. *IBM Journal 39(3)*, 915–931.
Vaucelle, C., Africano, D., Davenport, G., Wiberg, M., & Fjellstrom, O. (2005). Moving Pictures: looking out/looking in. *Proc. SIGGRAPH '05*, pp. 27–34. New York: ACM Press.
Vaucelle, C., & Davenport, G. (2003). Textable Movie: improvising with a personal movie database. *Proc. SIGGRAPH '03*. New York: ACM Press.
Vaucelle, C., Gorman, M.J., Clancy, A., & Tangney, B. (2005). Re-thinking real time video making for the museum exhibition space. *Proc. SIGGRAPH '05*. New York: ACM Press.
Vaucelle, C., & Ishii, H. (2007). Interfacing video capture, editing and publication in a tangible environment. *Proc. Interactions, Lecture Notes in Computer Science*, pp. 1–14. Springer: Berlin.
Vaucelle, C., & Ishii, H. (2008). *Picture This!* Film assembly using toy gestures. *Proc. Ubicomp '08*, pp. 350–360. New York: ACM Press.
Vaucelle, C., & Jehan, T. (2002). Dolltalk: a computational toy to enhance children's creativity. *Proc. CHI '02*, pp. 776–777. New York: ACM Press.
Winnicott, D. (1971). *Playing and reality*. London: Tavistock Publishers.
Ziegler, F. Mitchell, P., & Currie, G. (2005). How does narrative cue children's perspective taking? *Developmental Psychology 41(1)*, 115–123.
Zigelbaum, J., Horn, M.S., Shaer, O., & Jacob, R.J. (2007). The tangible video editor: collaborative video editing with active tokens. *Proc. TEI '07*, pp. 43–46. New York: ACM Press.

Cati Vaucelle is currently a Researcher and PhD candidate at the MIT Media Laboratory under the guidance of Dr. Hiroshi Ishii. She is a 2008 Rockefeller Foundation New Media Fellow Nominee and 2005 John F. Kennedy Scholar for her studies at Harvard University. She graduated from MIT in 2002 and from Harvard University in 2006. In France, she received dual degrees in computer science (computational linguistics) and applied math (economics), with a minor in fine arts (photography) as well as a master's in computer science (researching graphic design). She also received master's degrees from the MIT Media Lab and from Harvard Graduate School of Design in product design.

Hiroshi Ishii is a Professor of media arts and sciences in and Associate Director of the MIT Media Lab. He received a BE degree in electrical engineering and ME and PhD degrees in computer engineering. He codirects the Things That Think Consortium and founded and directs the Tangible Media Group, pursuing a new vision of human–computer interaction: tangible bits. His team seeks to change the "painted bits" of graphical user interfaces to tangible bits by giving physical form to digital information. Dr. Ishii was elected to the CHI Academy by ACM SIGCHI in 2006.