Picture This! Film assembly using toy gestures

Cati Vaucelle
Massachusetts Institute of Technology
MIT Media Lab
20 Ames Street
Cambridge MA 02139
cati@media.mit.edu

Hiroshi Ishii
Massachusetts Institute of Technology
MIT Media Lab
20 Ames Street
Cambridge MA 02139
ishii@media.mit.edu

ABSTRACT
We present *Picture This!* a new input device embedded in children's toys for video composition. It consists of a new form of interaction for children's capturing of storytelling with physical artifacts. It functions as a video and storytelling performance system in that children craft videos with and about character toys as the system analyzes their gestures and play patterns. Children's favorite props alternate between characters and cameramen in a film. As they play with the toys to act out a story, they conduct film assembly. We position our work as ubiquitous computing that supports children's tangible interaction with digital materials. During user testing, we observed children ages 4 to 10 playing with *Picture This!*. We assess to what extent gesture interaction with objects for video editing allows children to explore visual perspectives in storytelling. A new genre of Gesture Object Interfaces as exemplified by *Picture This* relies on the analysis of gestures coupled with objects to represent bits.

Author Keywords
Video, gesture object interfaces, children, play, storytelling.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
We connect to our world using our senses. Every one of our senses is a knowledge shopper that grounds us in our surroundings [1]: with touch, one feels the texture of life, with hearing one perceives even the subtlest murmurs of our existence, with vision one clarifies their instincts. But human senses are not only about perception. We use gesture to apprehend, comprehend and communicate. We speak to ultimately translate and exchange with others. We visualize, record, and playback events using our memory to reflect on our history and to be immersed in experience. We as children and adults are engaged in everyday pretense and symbolic play. We embed and later withdraw from the world, using imagination to project ourselves into situations [35]. Our mental constructs are necessary to reach a deeper understanding of our relationship with our environment [3]. Children are offered stories by adults and are driven into fantasy play. They use toys to externalize and elaborate their mental constructions [8]. With character toys they create interrelationships and plots, a means to expose their social knowledge: knowing about human beings and social relationships [33]. If the toy has an immediately accessible visual perspective, a new world is opened to the child. The toy brings her into exploring visual and narrative perspectives of character props, expanding the discovery of her environment.

We imagine a world in which people play, create and exchange visual narratives with ease and transparency. Motivated by the playful improvisational environment of child storytelling with toys, we have developed a new category of video editing tools progressing towards the child's natural expression of play. In *Picture This!* we combine the activity of play with the video making process. Whereas play emphasizes spontaneity and improvisation, video making necessitates structure and composition. We were inspired by the theater play of Goethe's childhood [35], investigating what technology could add to the narrative and play experience. We use technology to offer visual feedback regarding how the scene looks like from the point of view of an imaginary audience. 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! The video process, supported by gesture induced editing, benefits children in practicing social interrelationships and visual perspective taking.

RELATED WORK
Toys initiate elaboration in play and language. Researchers have found a correlation between open ended play and imagination in writers, poets, and scientists [35]. Eighteenth century German writer Goethe reported treasuring his puppet theater as a child as he envisioned interrelationships and plots between the characters in his later novels. Unstructured or semi structured toys such as blocks, dollhouses and puppets can lead to transformations as varied and creative as a child's capacity to affect them

UbiComp'08, September 21-24, 2008, Seoul, Korea.

[35]. At a period in history when children are invaded with media and technological toys, it is essential to propose devices for authorship with competitive appeal when compared to cherished games that usually do not offer open-ended play. Research has shown that toys serve a fundamental function in the development of children [5, 24, 35]. Literature showed that the ability to move from one's own standpoint to take another person's view is at the center of cognitive growth [26, 67, 40]. As much as people cannot grow without interpersonal connection, without separation they cannot relate [3]. As said Kegan, cognitive growth emerges as a result of people's repeated attempts to solve the unsolvable tension between getting embedded and emerging from embedded-ness [16]. "Dwelling in" and "stepping back" are equally important to get the cognitive dance going [3].

Tangible User Interaction with Video

Picture This! is a tangible video editing system inspired by research on tangible user interface [34] and tangible bits [13] that combine physical objects to digital data. Tangible systems move digital media clips, arranging digital information physically [14], create multimedia stories [20], offer a token-based access to digital information [12, 36, 37], and become a direct physical organization of digital video clips using multiple handheld computers [42]. A set of small displays can be physically manipulated to interact with digital information [21].

Tangible User Interfaces for Children

Picture This! contributes to previous research on tangible tools for children [9, 23]. Researchers envisioned a system that plays visual sequences using tangible objects [31]. Cameras were redesigned to capture both the child and the video of the child to contextualize a recorded visual scene [17]. A tangible environment, informed by the design of a graphical user interface, facilitates the process of capturing and editing videos using physical video tokens [37].

Tangible User Interface for Authorship

Picture This! contributes to the design of authorship tools. Authorship allows children to become active participants instead of observers. TellTale invites children to connect story segments through a caterpillar toy [4]. StoryMat reflects spatial organization of characters in a story [6]. In Jabberstamp, children synthesize their voices in their drawings [29]. In I/O Brush, children use a paintbrush to gather picturesque information from their surroundings and share them with their peers digitally [32]. Our research centers around the design of physical movie-making machines for children thus complementing previous work on supporting children's fantasy and storytelling [6].

Gesture recognition

While tangible systems invite for capturing oral stories and videos, current systems do not benefit from the gesture interaction that children do while playing with toys. In this paper, we propose a gesture language for capturing and editing suitable for children in their toy environment. More specifically, we want to assess whether or not gesture based interaction with character toys for video editing allows children to craft movies that benefit the content of the video being created in exemplifying the toy's perspectives. Computer vision techniques are being implemented in consumer products. Researchers developed a system of gesture recognition to control home appliances [27]. The device is designed to sit on a shelf or table, which has a clear line of sight to the television and the owner. The software recognizes simple, deliberate hand gestures and then sends the appropriate signal to a universal remote control. Modeling computer vision algorithms for finger tracking, researchers control a graphical user interface projected onto a surface, surface that becomes interactive [18]. A common challenge shared between these examples and *Picture This!* is distinguishing between intentional and unintentional gestures.

Tangible Gesture Interfaces

Users move dolls to control parts of a sitcom [19], to manipulate graphic characters in a 3D virtual environment [15] or to control music expression [41]. Users trigger the objects to control visuals on a display and rarely anthropomorphize the objects being manipulated. In Topobo [28], with the rotation of a gear around an axis, the interface records motions. Users regulate their motions only through iterative interactions. In the vein of wind-up toys, with Topobo users create sculptures that can walk around. Construction kits for tangible interfaces are designed to articulate moving skeletons by connecting 3D geometry to physical artifacts [39]. *Picture This!* is inspired by Dolltalk, a gesture based storytelling system [38]. Dolltalk captures, analyzes and interprets a set of gestures in parallel with analyzing changes in voice prosody. Using sensors and audio analysis, the system interprets the narrative structure of a story. The primary goal of Dolltalk is to invite children into discovering narrative perspectives during storytelling play. *Picture This!* brings in the visual perspective complementing the storytelling in Dolltalk and produces movies from the child's own toy environment as she plays.

Gesture Object Interfaces

The movements that one makes with object in hand, not only animate that object, but do something much more. Those movements carve out a context, giving a thing a certain life that is as dynamic as the user is able to imagine and communicate through her gestures. The dynamism in the space of gestures available to an individual is a language. The language of an object that gets associated with gestures is the scope of life that one can impart into the object. Therefore, to interact with a *gestural object*, one needs to understand the scope and flexibility of the gestural space available to give the thing life. In *Picture This!* not only are users making gestures to succeed towards the goal of the application, to make a video, but they are additionally extending an anthropomorphic characteristic directly to the object. The anthropomorphic characteristic is the focus of the user. It is the goal of the application, not

only to provide an input to the screen, but to be an equally valid source of attention on its own. This anthropomorphism or the "identity reinvention" of the controller through manipulation presents itself as a gestural interaction. Gestures scale like a language, have different contexts, different meanings and different results. For instance, in the Nintendo Wii, the controller alternates between being a character on a screen and a tennis racquet. In I/O brush [32], the identity of the camera is reinvented. As the camera becomes a paintbrush it fills children with wander, and they literally want to paint using the entire world palette available to them. We contrast this research with current work in *manipulatory interfaces*, in which the language of manipulation is the scope of functional movement, a physical mechanism to produce a result in the world. Physicality includes the philosophical foundations implied by embodiment. With physical devices we control things by physical body movement, by turning, moving, and manipulating appropriate mechanical devices [25].

We offer a visual representation and logic articulation of our research terms using the characteristics of a semiotic square [10]. We created the gesture object interface semiotic square (see figure 1).

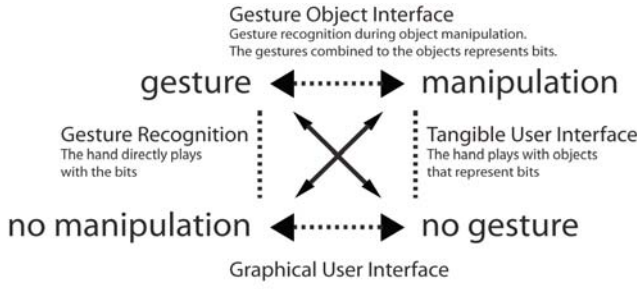

Figure 1. Gesture Object Interface semiotic square

We position the terms in the square to define our research areas, and the specifics of each area with regard to the body. *Gesture + no manipulation:* the field of gesture recognition. It involves the interaction with the hand. The hand can directly play with the bits. *Manipulation + no gesture:* the field of Tangible Bits as exemplified by Tangible User Interfaces [13]. It involves the interaction with the hand. The hand plays with objects that represent bits. The area we position *Picture This!* in is Gesture Object Interfaces, which involves gesture recognition during object manipulation: the gestures combine with the objects to represent bits.

IMPLEMENTATION

Interaction Design

We designed *Picture This!* so that the child's toy becomes a camera person as opposed to having the child hold a camera directly (see figure 2a). As a child plays with the toy that holds the camera, its video feed is projected on a screen in front of her in real time (see figure 2b). This visual flow aims to motivate her in composing a final movie as she plays and explores her visual story.

Figure 2. The toy is the camera person (2a) vs what the toy "sees" from "his" video feed (2b).

Our system consists of the following design features:

An audiovisual device	Two digital video cameras, two vibration sensors and a microphone.
Motion capture (real-time video and sound)	An algorithmic video editing system composes a movie from these inputs.
Motion based editing engine	Assembles the film as its story is being narrated.
Video output display	On a screen and speakers.
Toy prop augmentation	Video cameras and accelerometers. *The children can use Picture This! both as a doll hand-bag or a doll audiovisual recorder. The tool is flexible for a child to take the perspective of props she selected.*
Extensibility	Children spontaneously attach the system to other toys to capture their visual perspectives. *In this paper, we focus on children's interaction with the dolls as they create a movie about a two-character toy conversation.*

Table 1. Design features of *Picture This!*

Because the rationale for our system is to invite children to create and record a movie conversation between toys, we designed two sets of camera bags to be attached to two dolls containing video cameras and embedded electronics. As the two dolls interact with one another, they alternate between their respective visual scenes. The child creates the 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 not break the flow of traditional pretend play with character toys. Thus, our system incorporates the child's gestures with the cameras and toy's accessories as control functions to conduct film assembly. To engage with *Picture This!* children rely on their usual gesture interaction with toys while telling a story and playing with character toys (see figure 3).

Figure 3. Mike, 8 years old, playing with *Picture This!*

Rehearsing mode: The live video feed comes from the camera attached to the doll and is continuously displayed on a screen in front of the child. As the child moves one doll around, the second doll video captures the scene for preview. The video feed outputs from the dolls' cameras alternate between each other. The child rehearses the story video she wishes to create by selecting angles, scene shots and speed.

Recording mode: We implemented a quick interaction language for movie editing with dolls: If a doll wants to be in the video, the doll needs to move. If it wants to be recorded it has to move three times quickly and the doll's conversation partner will start the recording. The first blurry frames will be automatically removed from the recorded piece. Because only specific motions are detected, the child can move the dolls around without interruption of the recording mode.

Stop recording mode: To stop the recording, the doll that is currently being recorded turns horizontally and holds that position for two seconds. The second doll's camera automatically stops the recording and goes back to preview. The system automatically deletes the blurry frames from the horizontal motion. Two dolls alternate back and forth between being cameraman and the actor.

Playback mode: To play back the movie, the two dolls have to be moved in synchrony, in essence, jumping horizontally together. The sequences of video clips are automatically added to one another and the blurry frames from the gesture commands are removed. The final movie is played back on the display for the child to watch her final video composition.

During these modes, a graphic icon indicates the current mode of video action. It works as a feedback to know if the recording/preview or playback is happening. The icon is a 3-inch per 3-inch star: yellow when the child previews the movie, red when the child records the video and green when the child plays back her final movie.

The motions that are detected by the system are anthropomorphized. The dolls need to jump in synchrony at completion and shake for attention, as if the doll wants to say: "film me, film me!" To master the interaction with *Picture This!*, the child needs to alternate between projecting herself onto her toys and being the master-mind of the scene.

System

Hardware

We designed two bag accessories for two toys. Data is transmitted in real time to a microcontroller. Each bag attached to the toy contains a microcontroller, a piezo vibration sensor, a printed circuit board and a tailored video camera with USB connection.

For this first prototype we decided to use digital cameras, because they interface easily with our software. In the future, we will consider a wireless version for *Picture This!* with analog cameras so that children can capture visual perspectives from every object they want in a more flexible context.

We use a Piezo[1] vibration sensor. Piezoelectric materials create electric charge when mechanically stressed. In *Picture This!,* we use a Piezo sensor as an accelerometer. The Piezo is loaded by a mass to offer high sensitivity at low frequencies. We distinguish between horizontal and vertical motions by detecting strong motion into one axis and weaker motions into an orthogonal axis. To amplify the signal from the Piezo sensor, we use a charged amplifier (see circuit diagram in figure 4).

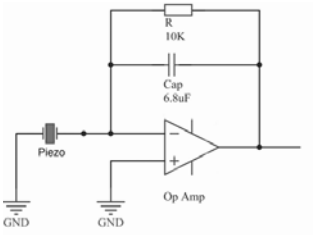

Figure 4. Charged amplifier for Piezo sensor

For a cut-off frequency $1/(2\pi RC)$ at approximately 30hz we use Cf=6.8µF and R=10k. The output voltage will depend on feedback capacitance not the input capacitance[2]. The advantage of using a charged amplifier in our design is that it allows us to use a long cable between the Piezo film sensor (in the bag of the doll) and the electronics. Also the advantage of using a charge amplifier is to minimize charge leakage through the stray capacitance around the sensor. We used accelerometers to detect gestures, but we could use the cameras video feed using optical flow calculation to discriminate between vertical and horizontal motion. This would simplify our sensor package and allow our system to be deployed with common USB web cameras.

We programmed the microcontroller to continuously interface the accelerometers data from two circuits with our software. Our software retrieves the output data of the microcontroller so that each toy carrying an augmented bag can communicate with our program via specific gestures filtered by our software.

Filtering algorithm

The motions we chose to identify support natural character play movements, such as jumping and shaking, with the addition of video control functions to these character play movements. We developed a filtering algorithm for gesture recognition through which angles of motions are detected and interpreted (see table 2).

[1] http://www.meas-spec.com/myMeas/sensors/piezo.asp

[2] MSI Piezo Guide, Piezo Film Sensors, technical manual, Measurement Specialties, Inc. www.msiusa.com

Motions / Input		Output
Doll 1 (D1)	Doll 2 (D2)	
Mode (a) Preview		
Vertical motion	*No motion*	Live video feed from D2
No motion	Vertical motion	Live video feed from D1
Vertical motion	Vertical motion	Continue live video feed
Mode (b) Record This mode blocks (a) until either (c) or (d) is triggered.		
Vertical motion 3 times in < 1s	*No motion*	Live video feed from D2 is recorded, indexed and associated with D2.
No motion	Vertical motion 3 times in < 1s	Live video feed from D1 is recorded, indexed and associated with D1.
Vertical motion 3 times in < 1s	Vertical motion 3 times in < 1s	Nothing happens
Mode (c) Stop Record		
Horizontal motion	No motion	This mode stops the current recording and goes back to (a) with corresponding video life feed from doll.
No motion	Horizontal motion	
Mode (d) Stop Playback		
This mode blocks (a) (b) (c) until the movie played back reaches its end. It then goes back to mode (a).		The software had concatenated the video segments captured in (b). It now plays back the entire movie on display.
Horizontal motion	Horizontal motion	

Table 2. Software architecture in *Picture This!*

USER EVALUATION

Interviews with primary school teachers

Gathering data from educators is essential in our user study, because educators will be the ones to interface between *Picture This!* and the students in a context of visual perspective taking and social knowledge exploration. We want to understand both the context in which our device can be used, e.g. at a school within a specific curriculum, and how the educators could use *Picture This!* as a teaching tool. Thus, prior to testing the system with children, we conducted three interview sessions for a total of five-hours long with two primary school teachers. In addition, we continued post interview via written exchange. The teachers we interviewed both teach 5-6 year old children at a charter school that is K-12. In the past, the teachers have also taught pre K class, third grade and sixth grade.

Teacher Feedback

Without explanation, the teachers moved the dolls as in pretend play, although the content of their play is conversational and turn-based. Their immediate comments pertained to how children can be drawn to conversation with this interface, which will be valuable to discuss social interactions: rather than playing with the toys as children usually do, in *Picture This!* children will be directed towards enacting a conversation between character toys. The teachers declared that with *Picture This!* children will be able to practice a social situation, to see it actually played out, to have time to reflect upon it and to later go back to try it in a different way. The teachers assessed that the video component of the application becomes the purpose and the focus of interacting with it. Creating video from the character toys, instead of just moving the toys around, will focus the attention of their students.

The teachers explained to us that having something physical for the children to play with is important for their learning. As seen in literature, Froebel combined early 19th century technology to play and learning, aiming to help children learn about number, shape and colors [5]. Montessori education [24] built upon Froebel's theory and offers manipulative materials to learn through playful explorations [30]. With *Picture This!* they are getting a combination holding the doll, playing with it and being immersed in watching what they are doing with it. *Picture This!* uses what children are familiar with as well as the way they naturally play. Holding the toys and being part of the action can connect the children between one another and allow them to alternate between "I am acting like this person" and "I am becoming this person". Literature showed that children as young as two years old can clearly demonstrate understanding of an other's point of view with the use of puppets as they naturally projects themselves onto the toys [7] and as one of the teacher explained: "because you can move their arms and you can move their body, it becomes a more real experience".

Teachers could see *Picture This!* being utilized with a wide range of age groups. The teachers expected that children younger than five years old would have difficulty in manipulating the video aspect of the system and that it might still be difficult for children who are 5-6, but that it is an important part of the system to help children practice "hand-eye coordination". The teachers explained that their younger students would love it and be thoroughly engaged by it and that 8 to 12 year olds would be engaged by the fact that they are making a movie with toys. The teachers confirmed that *Picture This!* could help children with the most traditional hand-eye coordination, primarily through what they are looking at in coordination with what they are manipulating, thus determining how they are looking.

An integral part of social skill development for children involves teaching appropriate language based interactions in emotional and difficult social contexts. For example, when a child pushes another child instead of telling her that her feelings are hurt; the teachers remind and assist the child to use language. Teaching positive interactions takes time for the children to practice and form a habit of positive response. The teachers use puppets, stuffed animals and character toys to help the children work through a difficult social situation. If the teachers see that a certain issue has come up a couple of times in the class, they will bring the class together to listen to a puppet interaction. They use toys in pairs pretending they are having a conversation enacting the interaction they were having difficulty with. When the social interaction was distanced from the

students, with the use of the puppets, the students were able to be much more rational and thoughtful about how to handle the situation. The teachers remarked that later on the children use similar dolls and enact situations themselves at school on their own. Enacting a conversation or a social situation with hand puppets allows children to understand someone else's point of view.

A teacher explained that she could introduce *Picture This!* into her kindergarten classroom and using it herself in the fall to model it for them. She would let the children explore it themselves in the winter and use it to do their own problem solving in late winter or spring.

User study with 4-10 year olds

In our qualitative evaluation with users age 4-10, we observe how eight children create a movie with their toys using *Picture This!* The large age range of our users inspires us to discuss the decomposition of our system into usability features based on a child's age. We also interviewed the parents to gather a context for our data. For instance we wanted to know if the children had prior exposure to television, digital media and computers. These observations will help us design a future empirical study by focusing on a specific age group and the system's functionality.

Method

To study children's interaction with *Picture This!* we installed our system at their home, or if the children requested, they came and interacted with it at our research laboratory. The children were asked to bring their own character toys to record a movie with. In the first couple of minutes, children explored the system without explanation. After five minutes, a researcher clarified how to operate the recording and the playback. The children were invited to play as long as they wanted. Eight children worked independently between forty-five minutes for up to two hours. Their interactions were videotaped and transcribed for analysis. Over a period of six months, our qualitative study involved twenty-one hours of transcription and analysis of children's play, the parents and teachers' interviews. One researcher transcribed all the video tapes and analyzed the results.

User study results

Children were extremely methodical and attentive with the use of the video. While in pretend play, they would sometimes stop their story and carefully work on their camera position. Children could spend an hour just playing with the camera view angle and alternating between characters.

They progressed between getting to have the doll in the picture, to a full shot of the doll, integrating specific backgrounds, discovering camera distortions and various camera angles, all facilitated by the size factor of the camera and its context. Children under age six tend to forget about the screen throughout their play, being immersed in their play.

After playing with the toys provided, children took out toys from their bag or from their bedroom. They had selected their favorite toys to be used with the system. Even if in some cases *Picture This!* was too big for their smaller toys, children were determined to make it work. We noticed in our evaluation, that even if some of the children removed the camera from a proposed character toy, they always put it back on top of another one. They did not use the camera detached from a toy as they would have done with a regular video camera. They were keen to explore the visual point of view of the toys they carry around. They found play-like justifications for the wires. A 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 of their mouth sometimes. Like a mask!" Children liked to alternate between dolls for their stories. They regularly changed their outfits as well as accessories. Children moved arms and the bodies of their toys to prepare for some particular movie action. As actors for their videos children used: teddy bears, stuffed animals, Bratz dolls™, Groovy Girls™, action figures such as Spiderman™ and Naruto™, plastic animals such as alligators and polar bears, a cement truck, homemade wands, and a stuffed horse.

The children above eight years old mastered the full system, being able to coordinate the dolls back-and-forth and to control the video, understanding the interaction between preview, record and playback. After twenty minutes of playing, the gestures with the dolls become parts of the children's vocabulary. Throughout our interviews and evaluations, the gesture language that the child learns to interact with *Picture This!* showed to be a motivating element in the video making process.

The main difference between the ten year olds and the eight year olds seemed to be in the motivation to practice more with the system. The ten year olds were willing to practice and revisit their stories over and over. Seeing their first playback motivated them in changing specific dialogs and scenes as well as the integration of visual angles they found funny, e.g. top-down view from an action figure. The flexibility of the tiny cameras allowed the children to capture details they particularly cared about in their movies, e.g. the action of wand in a doll's hand. The children discovered camera angles and visual tricks such as recording the display that appears to infinitely reproduce itself, in addition to working on the conversational aspect of the system and bringing perspectives into play. They were motivated by making a movie about their own toys. Out of the three children, after two hours of playing two of them wanted to continue with their movie, integrating new characters in the scene and testing new camera angles. During playback these older children were more self-critical. They made retrospective observations, like that

their voice should be louder, or that they should have made more flying actions.

At 6 years old, when the child was asked to play with one doll, the doll being exclusively a camera person, there was no confusion for the child. The preview with one doll captivated the attention of the child as she mastered the interaction. The child first focused on discovering visual perspectives and later used the doll to record a video. As she was introduced to the second doll, the child simply took the bag that contained the camera and brought it onto the other toy to repeat the same interaction. She held one toy in each hand, but only one doll was the camera person. Children under six years old could not properly de-couple between the toy as a camera person (see figure 5a) and the same toy not being a camera person anymore (see figure 5b); it was challenging enough to discover the world through the eye of one toy. Continuing with this interaction, the 6 years old child could go in *record mode* and record the video of "the doll who wants to be in the movie". The child played back her story by moving the two toys horizontally together.

Figure 5. 5a: the camera man toy vs 5b: the actor

At four-and-a-half years old, the child was spending most of her time alternating the visual perspective between the toys. Taking a toy in each hand, she told a story about two friends and alternated their visual and narrative points of view. We introduced the record mode, but the child ignored it. She was exclusively working on performing a story with visual scenes coming from the dolls. We questioned the surprising difference between a six year old who could not master the perspective alternation and preferred a dedicated camera toy, and a four and a half year old who managed to alternate between characters in her visual performances. The four and a half year old was only introduced to the preview features while the six year old was introduced to the record feature that seems to have confused her. Also our particular six year old subject had no prior exposure to the camera concept and does not have a TV at home. Six year olds and younger children were still figuring out the hand eye coordination aspect of the system and probably need more time than an hour to access the full functionality of *Picture This!* Under four years old, the child did not move the doll itself, but understood that if an object is presented in front of the camera-man toy, the object appears on the screen. Motivated by this aspect of the system, the child repeatedly inserted objects in front of the camera, trying different objects, smiling and laughing at her results.

Even though an empirical study would quantitatively support our results, we synthesized our observations of children playing with *Picture This!* as follows: *Yes* means above 80% success in a given interaction attempt, *No* means 50% and below. No subject performed between 50% and 80% success. An example of indicator for success in "preview 2 dolls" consists of a child acting a conversation between two characters while synchronizing her gestures and the video preview.

Interaction / Age	Preview 1 doll	Preview 2 dolls	Record 1 doll	Record 2 dolls	Playback final movie
< 4 *1 girl*	No	No	No	No	No
4-5 *1 girl*	**Yes**	**Yes**	No	No	No
6 *1 girl*	**Yes**	No	**Yes**	No	**Yes**
8 *1 girl, 1 boy*	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
10 *2 girls, 1 boy*	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Table 3. Results

The perspective taking in Picture This!

We remarked that the youngest children in our study (under eight 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 fist 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 eight years old) talked to the dolls, giving them directions for the movie, asking them to go away if the doll was too close to the camera, or asking why the doll "wanted to record so badly". 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!". Because the doll was not exactly doing what the child wanted, she took the doll close to her eyes and said, "somebody has a will on their own!". These examples show how 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. They explored conversations about: preferred sports between swimming and football, a doll asking her doll friend to teach her how

to dance, a doll asking a doll to take pictures, two strangers who decide to fight a giant together, a date with a friend, a Ninja fighting a giant, and jokes. The following is an excerpt of a video story by Jeremy, ten years old:

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!”

Real time visual perspectives by playing with video character toys in *Picture This!* allow children to develop visual perspective taking in determining where objects are located relative to another agent, or whether the agent can see a particular object [22]. The high level of concentration exhibited by the children demonstrated how challenging it is for a child to find the right angle and distance between the object and the camera, and between the two objects. An example that shows a child is exploring the hand eye coordination is when she says: “which one is it?” “Oh! it is this one”. She would act using a doll and say: “where is she! where is she!” looking for the doll in the video. A ten year old spent ten minutes trying to have her eyes look directly onto the screen while being able to see herself looking straight at the screen at the same time than using the dolls. It was an interesting challenge, and she could simply have removed the cameras from the dolls to decrease her challenge, but she did not, she wanted to construct a movie in which she would look at the doll straight in the eyes at the same time that wanting to look at her eyes looking at herself (an impossible problem unless the camera doll is at the same location than the display). Children could be confused by which doll was recording, and needed to find visual cues to determine which doll was being recorded, the visual cues consisted in recognizing objects near the doll on the display.

In our tests, we noticed that in the case of a child who does not own a single character toy, even in circumstances when the child is not familiar with characters, when given a toy that through the technology drives the interaction towards a social perspective taking, he takes advantage of that opportunity and plays in terms of characters with perspectives.

DISCUSSION AND FUTURE WORK

With the apparition of the camcorder and its preview display, the relationship between the actor and the cameraman is transformed. The actor has more control over how her actions are represented in the global scene. Through the preview display, the actor is given a real time visual feedback and can adopt different postures accordingly. In *Picture This!* the traditional camera-human relationship is also modified. The perspective effort needed is demonstrated through the spatial and visual coordination, managing the right angle for the right doll at the right moment in time, and this while acting out a story with the toys. The point of focus of the movies being the characters, children are guided towards creating a conversation which provokes a shift in perspective [43]. Children have an object to focus on for their movie that 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 eyes view” to obtain a deeper understanding on their own stories. The children alternate between 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. The ability to mentally manipulate objects, and imagine how an object appear if it is moved, is spatial cognition [11]. We observed that *Picture This!* invites children to practice spatial cognition between imagining the point of view of the toy, trying it out and correcting it.

Children were motivated in seeing “how it looks like out of a toy’s eyes” and expressed the wish to have action figures take video at their house; to have Lego people with an eye socket in which the camera could be inserted; to mount the *Picture This!* system on a racing car to have the point of view of the driver; and to have a waterproof version of *Picture This!* to capture videos under water with the child’s toys. Children proposed to have the *Picture This!* system connected to different kinds of toy accessories. When we asked the children if they would rather have a button to play with their dolls, one child responded that for practicing and alternating camera views the doll should move because it is easier and more fun, but for the recording, maybe a button would be good to be sure the recording starts at the right time, or finally, that the recording could be made continuous by alternating back and forth between the dolls. One child affixed the cameras with a doll scarf onto the two

toys he chose for his movie. Children were working around the system to make it work the way they wanted it to. Because 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, various objects could be given a camera spot and generate movies based on how they are handled. Children projected their actions onto the dolls and anthropomorphized the hardware commands. For instance, if the accelerometer does not respond, it is because the doll wants it that way!

Playing is about spontaneity and improvisation, while editing a movie is about structure and composition. Movie making can have a bit of both. We chose a gesture based interaction for movie making because of its advantage to integrate well with play. The trade-off of our system consists of being a movie making system more than being exclusively about role playing. *Picture This!* invites for the discovery of unique angles and point of view, its gesture based interaction facilitating the movie making flow. We found that *Picture This!* invites children to experience 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.

In the future, we envision the system to allow more editing and playback possibilities, such as the visualization and editing of other's video play stories. We would like to integrate a third camera view. This third perspective could give a visual context to the conversation created. A publication mode of the created movies could allow children to collect, share and revisit their movies. The teachers proposed to create a database of all the successful and unsuccessful interaction videos created by the children using the toys and to retrieve them later on as examples. Finally, we would like to organize a longitudinal study to evaluate the system *in situ*, in a kindergarten environment. In our current study, only two boys out of the eight children played with the system. In a future user study, it will be interesting to compare boys and girls playing with *Picture This!* because girls may be more accustomed to enacting social interrelationships with toys.

CONCLUSION

We presented the design, implementation and rationale for a new input device embedded in children's toys for video composition. By combining movie making during play and the improvisatory element, *Picture This!* naturally extends play to creative outcomes. Integrating a video editing algorithm to automate the editing process in a gesture object interaction allows one to get closer to the object of focus in a captured scene (for instance, a specific character). The video process, supported by gesture induced editing, benefits children in practicing social interrelationships as well as visual perspective taking, thus expanding creativity in video composition. Interviews with primary school teachers, parents and children informed us that *Picture This!* could offer a competitive appeal comparable to children's cherished games that usually do not offer open-ended play. Even though a controlled study would support our qualitative observations, our user study with eight children ages between 4 and 10 years old indicates that *Picture This!* allows children to capture storytelling with physical artifacts at different levels of interaction. For children under 6 years old, only one of the two or more toys should be carrying a camera, while the other toys could converse with each other. For younger children, *Picture This!* functions as a video performance system with video snippets of the child's play. For older children, *Picture This!* allows them to test visual angles and assemble a movie as they play with their toys. Different modes of interaction in *Picture This!* can now be coupled to different age groups. Functionalities can be distinguished with a specific cognitive goal for each mode. The preview mode benefits young children in developing the spatial-visual coordination while playing with their favorite toys and telling stories. The recording and playback modes enable older children to use their social perspective taking visually and through storytelling.

ACKNOWLEDGMENT

Many thanks to the children, parents and teachers who contributed to this work. Thanks also to Adam Boulanger, Edith Ackermann, the reviewers of UbiComp, the members of the Tangible Media Group and the MIT Media Lab for their feedback.

REFERENCES

1. Ackerman, D. (1990). *A Natural History of the Senses*. Random House, Inc.
2. Ackermann, E. (1996). Perspective-Taking and object Construction. In *Constructionism in Practice* (Kafai, Y., and Resnick, M., Eds.). Mahwah, New Jersey: Lawrence Erlbaum Associates. Pp. 25-37.
3. Ackermann, E. (2004). *Constructing knowledge and transforming the world*. In Tokoro, M. & Steels, L. Eds. The Future of Learning. Amsterdam: IOS Press, 15-35.
4. Ananny, M. (2002). Supporting children's collaborative authoring: practicing written literacy while composing oral texts. *Proc. CSCL 2002.*
5. Brosterman, N. (1997). *Inventing Kindergarten*. New York: Harry N. Abrams Inc
6. Cassell, J. and Ryokai, K. (2001). Making Space for Voice: Technologies to Support Children's Fantasy and Storytelling, Personal Technologies, 5 (3): pp. 203-224.
7. Denham, S. A. (1986). Social Cognition, Prosocial Behavior, and Emotion in Preschoolers: Contextual Validation. Child Development, Vol. 57, No. 1.
8. 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, N.J.: Ablex.

9. Frei, P., Su, V., Mikhak, B., and Ishii, H. (2000). Curlybot: designing a new class of computational toys. *Proc. CHI '00.*
10. Greimas, A. J. and Courtès, L. (1979). *Sémiotique - Dictionnaire raisonné de la théorie du langage I.* Carré sémiotique. Paris: Hachette.
11. Henderson, A., Pehoski, C. and Murray, E. (1991). Visual-Spatial Abilities in *Sensory Integration.* F.A. Davis, pp 123-140.
12. Holmquist, L. E. Redström, J. and Ljungstrand, P. (1999). Token Based Access to Digital Information in *Lecture Notes in Computer Science, Springer-Verlag*.
13. Ishii, H. and Ullmer, B. (1997). Tangible bits: towards seamless interfaces between people, bits and atoms. *Proc. CHI'97.*
14. Jacob, R. J., Ishii, H., Pangaro, G., and Patten, J. (2002). A tangible interface for organizing information using a grid. *Proc. CHI '02.*
15. Johnson, M. P., Wilson, A., Blumberg, B., Kline, C., and Bobick, A. (1999). Sympathetic interfaces: using a plush toy to direct synthetic characters. *Proc. CHI'99.*
16. Kegan, R. (1982). *The Evolving Self*. Cambridge, MA: Harvard University Press.
17. Labrune, J. and Mackay, W. (2005). Tangicam: exploring observation tools for children. *In IDC '05*.
18. Letessier, J. Bérard, F. (2004). Visual Tracking of Bare Fingers for Interactive Surfaces. *Proc. UIST 2004.*
19. Lew, M. (2003). Office Voodoo: a real-time editing engine for an algorithmic sitcom. *SIGGRAPH 2003.*
20. Mazalek, A. and Davenport, G. (2003). A tangible platform for documenting experiences and sharing multimedia stories. *Proc. ETP '03.*
21. Merrill, D., Kalanithi, J., and Maes, P. (2007). Siftables: towards sensor network user interfaces. In TEI '07.
22. Michelon, P. and Zacks, J. (2006). Two kinds of visual perspective taking. *Perception and Psychophysics*, 68(2), 327-337.
23. Montemayor, J., Druin, A., Chipman, G., Farber, A., and Guha, M. (2004). *Storyrooms and playsets: Tools for children to create physical interactive storyrooms*, Computers in Entertainment, 2 (1).
24. Montessori, M. (1912). *The Montessori Method*. New York: Frederick Stokes Co.
25. Norman, D. A. (2007). The next UI breakthrough, part 2: physicality. *Interactions 14*, 4 (Jul. 2007), 46-47.
26. Piaget, J. & Inhelder, B. (1967) In The child's conception of space. The coordination of perspectives. Pp 209-246. New York: Norton&Co.
27. Premaratne, P. and Nguyen, Q. (2007). Consumer electronics control system based on hand gesture moment invariants, *IET Computer Vision*. Vol 1: 35-41.
28. Raffle, H., Parkes, A., Ishii, H. (2004). Topobo: A Constructive Assembly System with Kinetic Memory. *Proc. CHI 2004.*
29. Raffle, H., Vaucelle, C., Wang, R. and Ishii, H. (2007). Jabberstamp: Embedding sound and voice in traditional drawings. *Proc. IDC'07.*
30. Resnick, M. (2006) Computer as Paintbrush: Technology, Play, and the Creative Society, in *Singer, D., Golikoff, R., and Hirsh-Pasek, K. (eds.),* Play = Learning. *Oxford University Press. 2006.*
31. Rizzo, A., Marti, P., Decortis, F., Rutgers, J., and Thursfield, P. (2003). *Building narratives experiences for children through real time media manipulation: POGOworld*. In Funology. Amsterdam : Kluwer Academic Publisher.
32. Ryokai, K., Marti, S., and Ishii, H. (2004). I/O brush: drawing with everyday objects as ink. *Proc. CHI'04*.
33. Shantz, C. (1975) The development of social cognition. In E. M. Heterington (Ed.), *Review of child development research* (Vol. 5). University of Chicago Press.
34. Sharlin, E., Watson, B., Kitamura, Y., Kishino, F., and Itoh,Y. (2004). *On tangible user interfaces, humans and spatiality*, Personal and Ubiquitous Computing, 8 (5).
35. Singer, D. and Singer, J. (1990). The House of Make Believe: Children's play and the Developing Imagination. Harvard University Press.
36. Ullmer, B. and Ishii, H, (2000). *Emerging frameworks for tangible user interfaces*, IBM Journal.
37. Vaucelle, C. and Ishii, H. (2007). Interfacing Video Capture, Editing and Publication in a Tangible Environment. *Proc. Interact 2007*, pp 1-14.
38. Vaucelle, C. and Jehan, T. (2002). Dolltalk: a computational toy to enhance children's creativity. *Proc. CHI 2002.*
39. Weller, M. P., Yi-Luen D. E. & Gross, M. D. (2008) Posey: Instrumenting a Poseable Hub and Strut Toy. *Proc. TEI'08*
40. Winnicott, D. (1971) Playing and Reality. London: Tavistock Publishers.
41. Yonezawa, T., Clarkson, B., Yasumura, M. and Mase, K. (2001). Context-aware sensor-doll as a music expression device. *Proc. CHI'01.*
42. Ziegler, F. Mitchell, P. and Currie, G. 2005. How Does Narrative Cue Children's Perspective Taking? *Developmental Psychology*. Vol 41(1): 115-123.
43. Zigelbaum, J., Horn, M. S., Shaer, O., and Jacob, R. J. (2007). The tangible video editor: collaborative video editing with active tokens. *Proc. TEI'07*.