Play it by eye, frame it by hand!
Gesture Object Interfaces to enable a world of multiple projections.

by

Catherine Nicole Vaucelle

MDesS, Harvard University (2006)
M.S., Massachusetts Institute of Technology (2002)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September, 2010

Author ....
Program in Media Arts and Sciences
September, 2010

Certified by ....
Hiroshi Ishii
Muriel R. Cooper Associate Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by ....
Pattie Maes
Associate Academic Head

Play it by eye, frame it by hand!

Gesture Object Interfaces to enable a world of multiple projections.

by Catherine Nicole Vaucelle

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning, on September, 2010,
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Media Arts and Sciences

Abstract

Tangible Media as an area has not explored how the tangible handle is more than a marker or place-holder for digital data. Tangible Media can do more. It has the power to materialize and redefine our conception of space and content during the creative process. It can vary from an abstract token that represents a movie to an anthropomorphic plush that reflects the behavior of a sibling during play. My work begins by extending tangible concepts of representation and token-based interactions into movie editing and play scenarios. Through several design iterations and research studies, I establish tangible technologies to drive visual and oral perspectives along with finalized creative works, all during a child's play and exploration.

I define the framework, Gesture Object Interfaces, expanding on the fields of Tangible User Interaction and Gesture Recognition. Gesture is a mechanism that can reinforce or create the anthropomorphism of an object. It can give the object life. A Gesture Object is an object in hand while doing anthropomorphized gestures. Gesture Object Interfaces engender new visual and narrative perspectives as part of automatic film assembly during children's play. I generated a suite of automatic film assembly tools accessible to diverse users. The tools that I designed allow for capture, editing and performing to be completely indistinguishable from one another. Gestures integrated with objects become a coherent interface on top of natural play. I built a distributed, modular camera environment and gesture interaction to control that environment. The goal of these new technologies is to motivate children to take new visual and narrative perspectives.

In this dissertation I present four tangible platforms that I created as alternatives to the usual fragmented and sequential capturing, editing and performing of narratives available to users of current storytelling tools. I developed Play it by Eye, Frame it by hand, a new generation of narrative tools that shift the frame of reference from the eye to the hand, from the viewpoint (where the eye is) to the standpoint (where the hand is). In Play it by Eye, Frame it by Hand environments, children discover atypical perspectives through the lens of everyday objects. When using Picture This!, children imagine how an object would appear relative to the viewpoint of the toy. They iterate between *trying* and *correcting* in a world of multiple perspectives. The results are entirely new genres of child-created films, where children finally capture the cherished visual idioms of action and drama. I report my design process over the course of four tangible research projects that I evaluate during qualitative observations with over one hundred 4- to 14-year-old users. Based on these research findings, I propose a class of moviemaking tools that transform the way users interpret the world visually, and through storytelling.

Thesis supervisor: Hiroshi Ishii
Title: Muriel R. Cooper Associate Professor of Media Arts and Sciences

Thesis Committee

Thesis supervisor..

Hiroshi Ishii
Muriel R. Cooper Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Thesis reader..

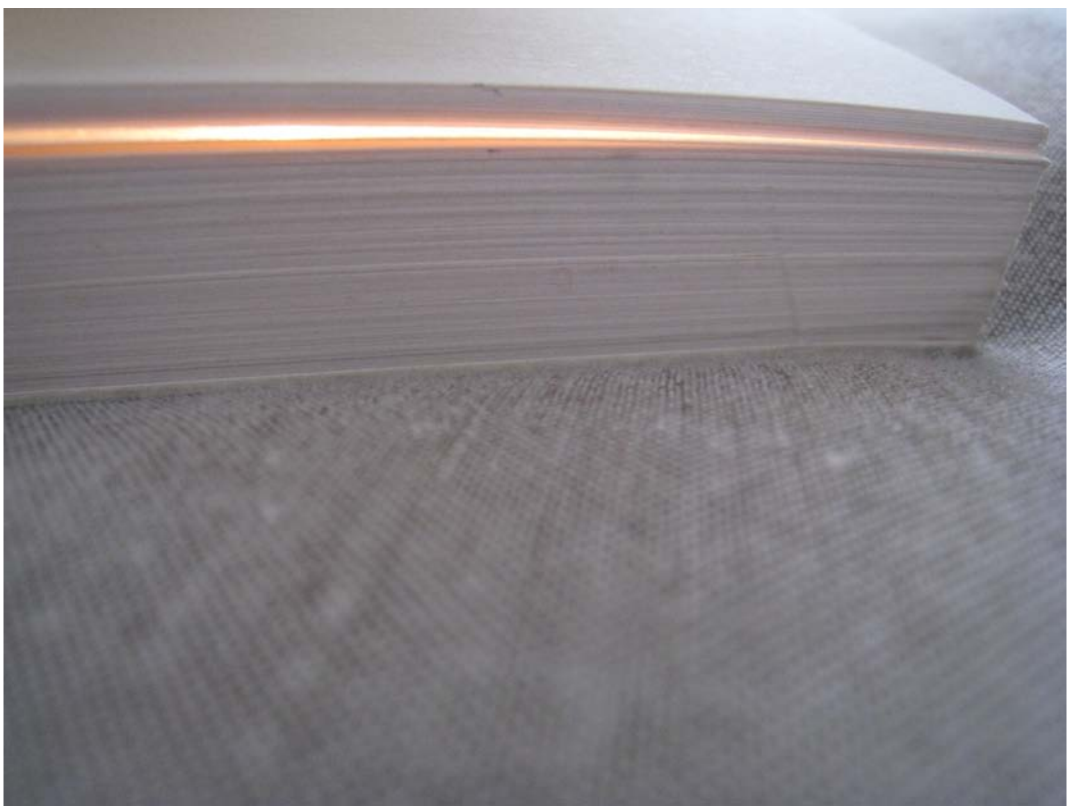
Edith Ackermann, Ph.D.
Professor of Developmental Psychology
Université d'Aix en Provence, France

Thesis reader..

Cynthia Breazeal, Ph.D.
Associate Professor
Program in Media Arts and Sciences

To Adam et Anaïs.

Acknowledgments



La pensée: la magie.
Le livre: le totem.
-Olivier Vaubourg.

I would like to thank many people who accompanied me through this long PhD journey.

First, I would like to thank my advisor, Hiroshi Ishii. He is an incredible mentor. Hiroshi has mastered many perspectives and brings them all in the work: the spark, the intellect, the humanist and the technical. I am so proud to work closely with such a visionary professor and I could never thank him enough for all he has done for me.

I would like to thank my spiritual mother Edith Ackermann. My thesis is replenished with her vision on children, their development and their world at play. Unconditionally present, she helped me push the thesis conceptually, supporting the technical implementations with illuminating interaction scenarios and helping me strengthen the perspective taking mechanisms behind my work.

I would like to thank Cynthia Breazeal, a role model. She merges research to real world questions and pushes the analytical side of my work. Cynthia helped me envision how this work can grow, inspiring my research to an entirely new level.

I am so fortunate to have worked with all of you.

Glorianna Davenport, Natalie Jeremijenko, Ellen Yi-Luen, Joe Paradiso, Michael John Gorman, Richard Boulanger, Rosalind Picard, Pattie Maes and Brendan Tangney for being wonderful mentors. Jacqueline Karaaslanian, Frank Moss, Linda Peterson, Aaron Solle, Felice Gardner, Pat Solakoff, Stacie Slotnick, Amna Carreiro, Jack Driscoll, Tom Lutz, Kevin Davis, Sarah Hunter and Lisa Lieberson for guiding me and supporting me through the student life at MIT! Mirei Rioux for organizing such a nice gathering after my defense! Keywon, Jinha, Adam, Jamie, Sean, Daniel, Xiao Xiao, Hayes, Amanda, James, Mark, Jeff for our tangible discussions!

All the children in my studies and their parents. Diana Africano and Oskar Fjellström for our work on Moving Pictures.

My friends from all over the world: Olivier, Ewa, Sofia and Alexis. Kimiko, Stefan and Kian. Priam. François, Amélie and Paul. Frédo. Fred. Paulo. Susanne and Peter. Nan Wei, Amit, Yasmine, Elisabeth, Hannes, Sumit, Aurélius. Bernd and Mary. Dana, JB and Ella. Craig and Bonnie. Sophie, Fabienne, Chantal, Pat. Bernard. Les Roger, les Vaucelle, les Roberts, les Boulanger! Matthieu, Lauren. Amber. Selene. Joëlle, Arianna, Michael. Jonah, Katherine, and Adrian. Orit and Oren. Marisa and Steve. Etienne, Rémy, Florence, Burak, Alea, Rana, Dale, Barbara, Alyssa. You, my friends who were in constant contact to keep me sane!

My family, near and far.

My wonderful mum, for who I remain her little girl.

Mon amour, Adam. Ma douce Anaïs.
To our days of joy, learning and love!
I dedicate this thesis to you.

Table of Contents

List of Figures

CHAPTER ONE
Introduction

1.1 Collect to connect in the digital age

I am attached to objects: I gather, select, collect, organize and share them with others. Now that I am even more mobile, moving from one continent to another, I cannot carry my entire house on my back, but I must keep up with the collection. Mobile computing, with its digital means for collecting, conveniently allows for this transition. I become increasingly dependent on digital and ubiquitous computing at the same time that my attachment to physical objects must transfer to digital artifact gathering.

I am fascinated with a new kind of collector, whose instinct to collect is transferred into the virtual world and whose digital tools for collecting benefit from sensorial evocative qualities of the collection [Vaucelle, 2008].

I am part of a generation of neo-nomads [Abbas, 2006], carrying the summary of my house on my back, making sacrifices when deciding what to keep and what to leave out. Living away from my homeland, I am separated from my past. Limited by what I can carry, I need to leave behind my keepsakes, the prompts to souvenirs and thoughts that ground me in my emotional memory. I am an orphan of a country that I left for another one, cities that I explored yet abandoned while their memories erode with time. Rimbaud's psychotic ghost *Je est un autre* (I is an other) is the acknowledgment of this exile: to be foreign and live in a foreign country [Kristeva, 1999].

As a child in my home country I collected everything that represented an invitation to explore and understand the world and its habitants: international stamps, rocks, samples of sand from every beach, penpals from all around the world, and leaves from every kind of tree. If I were on a family trip, I would compile a summary of the journey with cards, pictures, flyers and fragments of bottles. I played games, trying to find the most unique rock on the beach, inventing rules that made the most sense for the rarest shellfish to be praised. I cherished these gathered treasures that the family house protected for me, even after I left.

Collectors such as Robert Opie, collector of advertising and packaging, frequently acknowledge that as part of the collecting instinct, one has to be prepared to tolerate the physical space taken by the collected objects [Elsner and Cardinal, 1994] . Now that I become mobile, physical things must go. This results in a selection of memento directed by a mobile life. I cannot keep up with the collection of artifact souvenirs. Is there a new breed of collectors emerging within the generation of nomads?

1.1.1 The digital collection

Digital cameras enable us to take an overwhelming number of shots of the same location. The internet keeps our emails. Online services host our pictures and video fragments. Video games allow us to collect avatars, connect remotely with peers and construct our new identities in a virtual world. It seems that the digital is welcoming a new generation of collectors. The digital object is appropriated, collected, hacked, transformed, possessed and shared. Anyone can capture a memento with a cell phone, or multiple shots of the same scene that end up being accumulated on the computer's hard disc.

Online tools, such as Flickr, precipitate the act of collection, inviting anyone to be a collector. Flickr users collect pictures, diligently select their favorites, organize them into groups, and share them. The audience navigates and searches for specific shots, comments and sometimes annotates directly on the photographs.

Our digital collectors -or neo-collectors- save things for later, re-edit and re-stage collections to show their treasures to chosen audiences and decide whether or not to make them public. Collectors have to decide who to invite and who to leave out. Some collectors decide to never share their collection with anyone.

In the physical collection, the collector saves one object from the whole with an attempt of classifying it and naming it. In the digital world of objects, the objects appear to be accumulated. Additionally, objects can be thematically organized. Our collections take shape by identification and categorization. We look for similarities and nuances between objects. The collection begins at the moment of discrimination [Levy et al., 2006] and the digital collection starts when people select their digital media, edit and beautify them, organize them into groups and present them as part of thematic "sets".

The digital realm can also directly participate in the physical collection. Online trading systems, such as eBay, are designed to invite collectors to share, sell, and exchange their collected material. What can be found, now that eBay has been available for over ten years, is that people collect almost anything. Social networks also call for the collector instinct by reinforcing the need to share contextual information as part of a "user profile": people collect friends, pictures, movies, songs, links and blogs to present themselves. While *Del.ici.ous* offers the abilities to collect links, *StumbleUpon* reinforces the idea of "dénicheur" -term from French art in the 20's- the one that hunts. Here the dénicheur hunts for digital links, links that will become popular as the www grows. For the digital link "accumulator", systems are designed to bridge very different content and allow an automated organization of the gathered material. In the blogosphere, by specializing and discriminating, the blogger collects information and curates media geared towards a specific topic. As much as a curator for a museum is a collector of artifacts, a blogger represents a new generation of curator by collecting and ultimately exchanging information.

Both as children and adults, we invest in the evocative qualities of the display for the collection [Putnam, 2001]. From the stamp book, to carefully designed velvet boxes for rocks, we present each piece as if it were in a museum. The digital realm has not explored this dimension yet: organization and labeling both take place on a very simple online page. The experience of digitally capturing the everyday and making sense of it through the physical act of collection could be combined. The digital could inform the physical. The physical could ground us deeper in our surroundings, and they could both exist independently from one another [Vaucelle, 2008].

The key to bind the physical act of collection and the digital opportunity of representation is metadata. Imagine a scenario where the discovered object grows references beyond the thing held, and the thing seen. The digital world can tie to an infinite number of the object's features, only limited by the technologies used to analyze and link the data. However, even simple features gain new meaning through tagging the collected object and investigating metadata such as location and temperature on gathered objects. This investigation challenges the exclusivity of digital and physical opportunities of interaction, and provides an experience where the physical process of collection is completely married to contextualization via digital means.

The world is being slowly overlaid with and in turn transformed by a virtual world. We bring our memories online: we scan pictures from the past to be "saved" and also shared. We digitize music, movies, books and we are working on ways to integrate more of our natural senses within this space. We make sure every existing physical object has its corresponding phantom online. It is a transition state as if we knew that our physical world was about to end. We feel we will never get it back (or get back to it) again. As Roland Barthes offers, "the Photograph mechanically repeats what could never be repeated existentially" [Barthes, 1981]. The online digital space does something more: it makes sure this existential repetition is being shared with everybody. Nobody can be forgotten nor can we forget: everybody is connected through thematic collections of data organized by locations, genre, or types of friendship connections.

One can easily construct a new self [Turkle, 1995], idealized through the careful selection of representative images, using descriptive tastes and links for personality reinvention, or social network connections to belong to a chosen group. Sharing photographs ensures an identity validation and/or reinvention through the co-observations of carefully selected images. It is as if users start the collection in the physical world for the purpose of the digital collection, grabbing objects, landscapes, colors, and expressions to later digitally capture them. The digital collected object can also exclusively exist in the digital world, e.g. Facebook's popular digital gifts.

1.1.2 The impact of these new collecting habits on the self

The self is never challenged

With a computer, one can be in contact with the entire world without having to move from their parent's house. The computer is the perfect compromise between the teenager's necessity to go outside to become independent and the necessity to keep the protection and security of the parent's house. Teenagers can also escape their fears related to their own body in comparison to the severe criticisms induced by co-located peers! The narcissistic image remains idealized.

Observing how users digitally collect and connect, a trend emerges within modern collectors. In this back and forth between idealized and controlled image, reflected on online tools, the self is never challenged. It is less about others and objects than it is about a constructed self. The modern collector collects herself to a point that glorifies her alter ego and makes sure this ego is never threatened. This narcissist plays with a myriad of images to admire herself and contrasts her appearance through imagery in different scenes, different contexts, surrounded by different persons.

Instant communication messaging systems are our ultimate means of communication. Even though all modern players use instant messaging, the more narcissist among us particularly benefits from such a system because she does not need to develop, be attached, or confronted by a relationship. With instant messaging, she can drop out of any conversation without explanation and avoid ensuing confrontation; she can pretend to be busy or "idle", or she can jump in to receive any required self-support. The super image is maintained. The digital realm works as a portable mirror whose feedback can be used to exclusively enhance the prodigal self. In the digital realm, one's represented self is entirely under one's control.

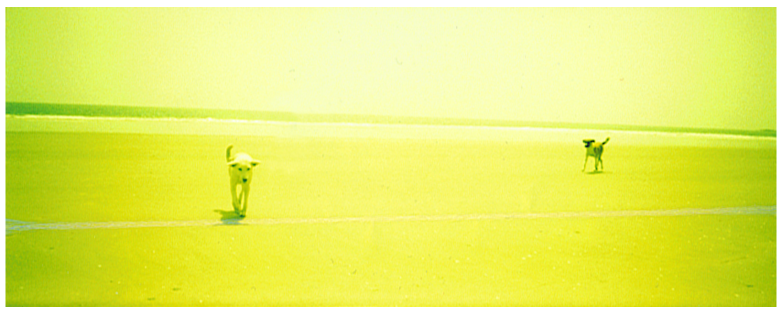
Communication

Communication technologies modified our relationship to space and time and with consequences for the development of our thinking [Gauthier and Moukalou, 2007]. Instead of confronting peers using a verbal exchange, communication is now guided with icons that one needs to only *click* in order to be projected to the other side of the globe.

The modification of the relationship between space and time in communication can explain the modification of our potential to take time to share thoughts. Language is more and more stereotypical, universal and univocal (close to the marketing discourse) which appears in political speech today [Chomsky, 1986b, Chomsky, 1986a, Chomsky, 1995]. The dialog with the computer is a series of keywords, revealing the transformation of the structure and use of language in our occidental society.

Gift giving

While these new collecting habits indelibly invoke the narcissist within, they also reflect our instinct to collect. In the sensibility of the collector, Charles Randall Dean explains that collecting might even pre-date society, and even humanity.



> "*I was on the beach recently and saw a woman walking with her Scottie dog and he had a rock in his mouth. And I said, "That's so nice that your dog is bringing a rock back from the beach for you.". She said, "Oh, this is not for us. This dog collects rocks." And she said he would spend fifteen of twenty minutes on the beach looking around for a rock that resembled the ones he already had, which were roughly hamburger shaped, and put them under the bed.*"
>
> Charles Randall, in [Levy et al., 2006].

Our pets collect rocks on the beach while we collect bits of flowers and thumbnails of friends on our computer. It seems that we all collect by instinct, but that we also use the act of collecting to connect to others.

In the digital world, online users send "virtual" gifts to one another, motivated by unlocking the availability of a new series of digital gifts in return! The digital gifts can be growing flowers, animals or more recently any growing "things". These digital gifts are collected in the spirit of how postcards, sent by family members and friends, used to be collected. Discovering a box of someone's collected post-cards at an antique store hardly seems voyeuristic, as the digital world nets us in one another's highly visible collections of mementos both sent and received from disparate friends. Facebook application developers continuously produce various digital gifts, means to enhance a profile, personalized photos and music albums, aggregations of tastes and personality contests. The collected digital-gifts imply reciprocity, a reciprocal exchange in which the object is tied to the giver [Mauss, 1967]. This relationship is enhanced with tools that directly link to the giver.

In the role-playing realm and in massive multiplayer games such as *World of Warcraft*, the ability to gather, collect and equip avatars with their virtual objects contributes to the game addiction. Players spend hours not only trying to upgrade their avatar's gear to the highest rank, but to equip them with all of the coolest pets that the game has to offer or all of the tabards that can be encountered within the game. The player who possesses the rarest riding mount with which to equip her virtual character is acclaimed. In massive online role playing games, another dimension accompanies the progression within the game, the one of collecting a series of elements, serial gathering facilitated through quest completion and exchanges with other players, e.g. to build a specific outfit, a player needs to exchange rare items with players and depend on others to finally possess the item. Through these virtual exchanges rise connections between players, connections that might affect the player's everyday life, with the feeling of having "friends".

WoW Pod

World of Warcraft (WoW), by Blizzard Entertainment, is the world's most popular massive multi player online role-playing game (MMORPG) with over 11.5 million monthly subscribers. The players navigate their characters (or avatars) through an expansive world of fighting monsters fulfilling quests, joining groups and guilds with other players. The game allows players to customize their avatar: race, gender, and class are selected in addition to skin color, hair color, piercing, tusks, etc. As a player gains experience and wealth in the game she aggressively hunts specific pieces of gear to increase her characters abilities and look more "epic" than less accomplished players. The ability to highly customize a character, to choose what quests the avatar will accept, and the highly social aspect of the game made World of Warcraft not only the most popular video game, but also one of the most addictive.

To engage players and observers in a critical debate, I co-designed the *WoW Pod* [Vaucelle et al., 2010] an immersive architectural solution for the advanced WoW player that provides and anticipates all life needs, see figure 1-1.

Figure 1-1: WoW Pod.

Inside, the gamer finds him/herself comfortably seated in front of the computer screen with easy-to-reach water, pre-packaged food, and a toilet conveniently placed underneath his/her custom-built throne, see figure 1-2.

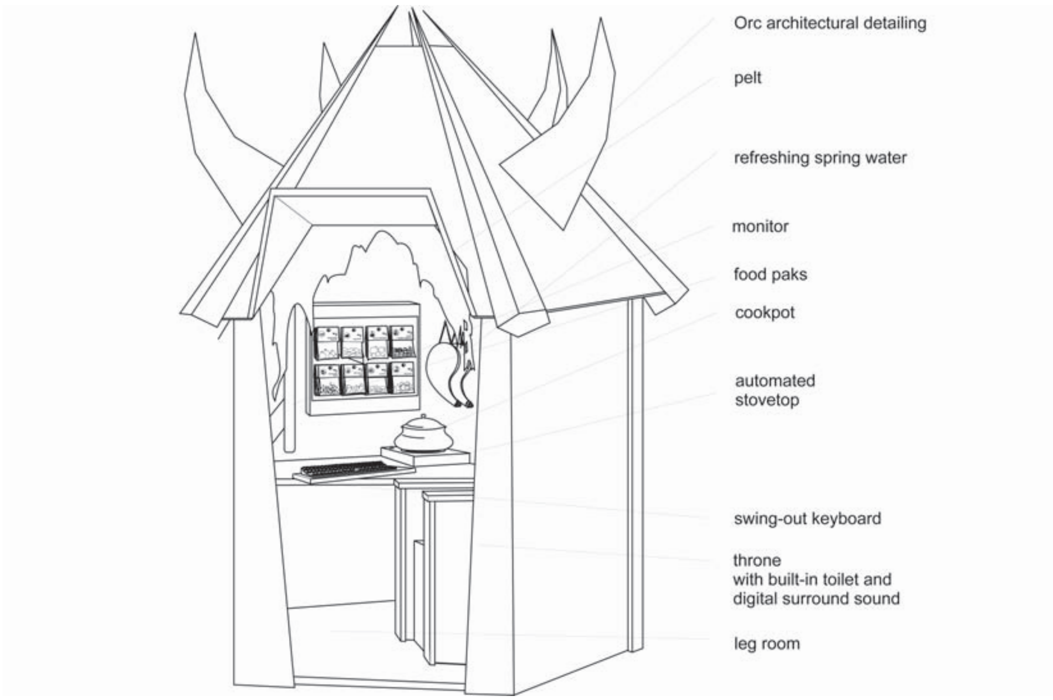


Figure 1-2: WoW Pod's specification.

When hungry, the player selects a food item such as “Crunchy Spider Surprise” or “Beer Basted Ribs” and scans it in. see figure 1-3. WoW Pod then physically adjusts a hot plate to cook the item for the correct amount of time and temperature. The virtual character then jubilantly announces the status of the meal to both the player and the other individuals playing online, e.g. “Better eat the ribs while they're hot!”.

When the food is ready, the system automatically puts the character in AFK (“Away From Keyboard”) mode to provide both player and avatar a moment to eat. The game player can then reach for his/her recently cooked meal, at an arm's length away. When the player resumes playing, he/she might just discover his/her character's behavior is affected by the food consumed in real life - sluggish from overeating or alternately exuberant and energetic.

Figure 1-3: Food-buffs for WoW players.

The exterior of WoW Pod mimics the look of WoW architectural structures, whose swaths of flat, pixellated surfaces digitally recreate the built environment of an imagined past. Upon crossing the threshold and entering WoW Pod, the player finds a tangible simulation of things digital.

It seems that our attachment to physical objects is being projected into this hybrid space. It is as if the new generation instinctively uses the digital collection to catalyze virtual connections. Collecting and connecting go hand-in-hand in social networks and eventually influence the everyday interactions of users.

1.1.3 The form and language of the digital collection

Communication cues

Non-verbal communication cues differ among cultures and may have changed with digital communication. Forty years ago, the distance between individuals drastically affected the dynamics of space interaction [Hall, 1980]. Now that we travel by plane and reach destinations further away from our natal home, we are more sensitive to the dynamics of non-verbal communication and the dynamics of interacting in space. We absorb most of these social interactions in our everyday routines. One of the virtues of the virtual space is to not interact through body language, as if we intrinsically avoid miscommunication and connect with persons we might not be able to connect to without confrontation and misunderstanding in the physical realm. We are now connected instantly to the other side of the planet and the myth of teleportation is virtually achieved.

Connection

The virtual world allows us to be perpetually connected to one another, connected but out of touch! While cell phones inspire us to be contacted at any point in time. The online realm keeps us from disappearing. However, search engines such as Google can sentence us to an electronic death, where we cease to exist online and we cease to be found. This type of extreme digital measure has implications both in the physical world -we cannot easily be contacted and found- and the virtual world -we disappear digitally.

We can now find jobs in other countries, advertise small companies, and develop a sense of omniscience through the ever growing information on the www. With the www, people who have never met before can finally meet. Children with atypical interests can finally find other children with the same hobbies. The digital space might become the land for the discovery of other cultures. People meet others from different countries via social networks. It is the ultimate place for meeting inhabitants of unknown places, cultures and perspectives.

Perspective taking

It is rather banal to say that with the internet we can be more connected. What is interesting is how the styles of our connections change. Our collection extends ourselves by discovering the world and its particularities. With stamps, a child discovers places through the lens of a stamp she later shares with peers. Our digital collection might equally allow us to challenge our points of view, interacting and sharing collections with a much wider audience. The ability to take perspectives is a complex skill that does not necessarily mean putting oneself in someone else's shoes. This following quote illustrates this point; not only can we not know what someone else feels, but we also cannot know that someone cannot know about someone else's life!

> *"One day Soshi was walking on the bank of a river with a friend. "How delightfully the fishes are enjoying themselves in the water", exclaimed Soshi. His friend spoke to him thus, "You are not a fish, how do you know that the fishes are enjoying themselves?" "You are not myself," returned Sohsi, "how do you know that I do not know that the fishes are enjoying themselves?""*
>
> Kakuzo Okakura, in [Okakura, 1964].

Trying to understand someone's life results in a process, an iterative back and forth between discovering, trying to understand, being challenged and reflecting on someone else's life, perhaps never understanding it correctly, but that is not the point. What matters is the process that brings people to open themselves to new perspectives.

The digital space can be perceived as a space where one can lose track of reality, but it can also be used to sharpen perceptive skills by challenging new ideas, visions and ways of thinking. By connecting to someone else, one can relate and create bonds. By creating links, one can project onto someone else's life and later empathize. By communicating with habitants from all over the world, one can perceive things differently. Navigating on the Internet, discussing on role-playing platforms, managing guilds and groups of peers requires a "mindful" state. This mindful state is exposed by being open to novelty, being alert to distinction, being sensitive to different contexts and being aware of multiple perspectives.

The widespread failure to recognize the insights that can be found in all different perspectives may itself constitute a disability. Thus, being confronted by individuals in a dynamic of unpredictable communication , such as through the www, might teach children to adapt themselves to new situations rather than being trapped in their conditional learning. At the heart of many theories of intelligence is a belief that it is possible to identify an optimum fit between individual and environment. Even navigating through the internet requires an education to new technologies. It is not a matter of fitting ourselves to an external norm, rather, it is a process by which we give form, meaning and value to the world [Langer, 1997].

1.1.4 The implications of a digital body

From the digital collection to the "digital" body

> "*We believe that two-way immortality, where one's experiences are digitally preserved and which then take on a life of their own, will be possible within this century.*"
>
> Gordon Bell and Jim Gray, 2001.

The digital offers a variety of tools and algorithms that impact remembrance and facilitate the organization of media. The digital now offers terabytes of data storage capacity carrying almost unlimited digital collections. It acts in the way the family house does for us: we know, may it be illusionary, that it will always be there for us, wherever we go and whenever we want it.

We are able to capture the life of a person, through pictures, video, speech recordings and her interconnections with others. Gordon Bell envisions digital immortality [Bell and Gray, 2001, Gemmell et al., 2006]. The Human Speechome Project [Roy, 2009] analyzes the language development of a single child, by recording, storing, visualizing, and analyzing communication and behavior patterns in several hundred thousand hours of home video and speech recordings. The goal of Lifelogging [Mann, 1994] is to record and archive all information in one's life: text, visual data, audio, media activity. It can even collect biological information from sensors that one wears. In the same vein, SenseCam is a photographic device that, rather than capturing individual images when triggered by the user, automatically captures a series of images [Cherry, 2005]. Isn't that amazing that our digital body motivates us in exercising! Our avatar makes us run [Runner, 2009]. Our virtual model, with her free weight-loss calculator, reflects our body fat. If we want a better looking avatar we need to exercise [VRM, 2008]!

Soon we will be able to have a digital copy of ourselves, a copy that will coexist online starting from birth and that will be able to perpetuate its existence after death, living eternally in a virtual reality. This copy will evolve independently from us, based on our original complexities, and will interact with other digital selves.

Things are getting lost with dematerialization

> "*But I am not afraid to consider the final question as to whether, ultimately—in the great future—we can arrange the atoms the way we want; the very atoms, all the way down! What would happen if we could arrange the atoms one by one the way we want them?*"
>
> Richard Feynman, 1959.

While expanding our experience, some things get lost along the path to dematerialization. While fifty years ago children progressively learned how to build toys under their grand father's supervision, now they think they can be champions in Karate without moving a finger. Whereas, in the twentieth century, kids gathered outside, created groups, and risked their identity by confronting others, now they can stay home while contacting the external world, protected by their parents [Gauthier and Moukalou, 2007].

As much as people cannot grow without interpersonal connection, without separation they cannot relate [Ackermann, 2004]. Kegan contends that cognitive growth emerges as a result of people's repeated attempts to solve the unsolvable tension between getting embedded and emerging from embedded-ness [Kegan, 1982]. "Dwelling in" and "stepping back" are equally important to get the cognitive dance going [Ackermann, 2004]. Typically, we reflect on the *cause-and-effect* of gestures as we observe them being executed in the world. Our gestures evolve via an iterative process of observation, reflection and production. Social gathering in locations and spaces comprise the essential environments by which we introduce and observe gestures. However, these opportunities are reduced as children communicate and play in virtual chat rooms.

The virtual world that we trust, does not connect to all of our senses: at best, the visual and auditory senses are engaged. The body is not being invested. It is as if we privilege a communication in which the body is absent. While our skin stands between us and the world, our hand feels intuitively and precisely what it touches and grounds us deeper in where we stand [Ackerman, 1991]. The virtual disconnects us from our physical existence. Joysticks designed to vibrate, when one bumps into a virtual car, are profligate. It is as if this video game controller says "Remember! You had a body! Let me show you how it feels!"

We are in a transition stage where we are too afraid to invest in our limited physical life and in which the virtual space makes things so much easier: no limitation of space and time, no reality check, no identity threat -we can be who we want to be without actually being the one we want to be. Ray Kurzweil envisions that the nonbiological portion of our intelligence will be trillions of trillions times more powerful than unaided human intelligence! [Kurzweil, 2006] In the future that Kurzweil describes, people will have to resist *intelligence surgery*, the same way we resist plastic surgery today!

> "*NED: You're missing something. Biological is what we are. I think most people would agree that being biological is the quintessential attribute of being human.*
> *Ray: That's certainly true today.*
> *NED: And I plan to keep it that way.*
> *Ray: Well, if you're speaking for yourself, that's fine with me. But if you stay biological and don't reprogram your genes, you won't be around for very long to influence the debate.*"
>
> Ray Kurzweil, 2006.

Our powerful human machine!

Technology brought us a means to organize ourselves, to schedule our time, and to "synchronize" ourselves with one another. However, humans are intrinsically linked to the rhythm of life, not to a planned calendar. People are tied together and yet isolated by hidden threads of rhythm and walls of time [Hall, 1989].

Time is treated as a language, organizer, and message system revealing people's feelings about each other and reflecting differences between cultures. Through repetition comes learning, comes depth of understanding, comes rhythm that affects our entire being. Synchrony in life finds a remarkable analog in the rhythms of music. The pattern of our movements can translate into a beat. Even complex behaviors, hovering around a theme, can imply a meter, a structured pace of time. Without this rhythm, we are not synchronized and we loose our contact with life. When humans interact in the virtual space for a long period of time, they may become desynchronized.

The digital potential of remembering, of allowing us to communicate with the entire planet, and of being always connected, found and synchronized, seems to rely too much on computation rather than human potential. Why doesn't the digital space allow us to make use of our human possibilities rather than replacing our incredible memory with more powerful hard disk space and clever task management software, replacing our subtle intuition and sense of deduction with automatic pattern analysis and behavior prediction?

Marvin Minsky's society of mind reflects confidence about the rule systems that are difficult to capture in Artificial Intelligence and shows how many complex behaviors, choices, and even social interactions can be interpreted by dynamic rule systems [Minsky, 1986]. According to Ray Kurzweil, our memory is far from incredible. It is a simple software dying to be replaced by more powerful computing [Kurzweil, 2006].

Do we really need to be "upgraged"?

It seems that the virtual world makes us dream that we are heroes by enhancing all of our senses, senses that we could exploit in physical reality. In role playing games we are equipped with sonic hearing, super vision, abnormal strength, and ultra sensitive noses! We are able to achieve a myriad of quests and psuedo-accomplishments, relying on our never-ending lives. The virtual space is a pale copy of a dream of ours to be super-human, yet we keep forgetting that we are a powerful machine. Our potential is being projected into virtual worlds, losing track of our capabilities.

With the rise of our virtual selves, are we becoming desynchronized with life?

1.2 Vision

I design tools for supporting inter-personal production of digital media, as well as scaffolding technology for perspective taking. This is a new and important means of understanding relationships with others, their communities, their rules, their habits, and their references to the world. Designing radically new interface experiences for the creation of digital media can lead to unexpected discoveries and shifts in perspective. Inter-personal interaction around the production of media has until recently been the domain of teams creating formal media artifacts such as movies and TV programs as well as a large portion of online media. Can the depth of engagement that is typically part of the formal production approach be designed into systems for everyday engagement?

How can one become the new David Lynch? We need to rethink of technologies that drive our unique assets. It is certainly not by designing technologies that tell the user what is a good from a bad composition. It is about technologies that invite the creator to learn how to break the rules of composition, reflecting on it while producing it, from its simple existence as a media tool to a more profound social means to expose, share, and talk about our identity and culture.

It starts when you are a child. It starts with all the tools, toys, games and interfaces that you are given as a child. We need to think carefully about what our interfaces are inviting us to do. How can we combine the production of new media with the consumption of new media? How do we reflect on what we create? Children spend inordinate amounts of their time online, on social networks, or playing video games. How do we combine the rich space of playing with toys, interacting in our environment with friends, and sharing stories to the mainstream use of video games, TV shows, and social networks for children?

All the tools I design carry with them the potential to transform the way we consume and produce media to access fundamental values such as the ability to take perspectives. I challenge the interface, and the type of content the interface invites us to accomplish, opening a new space for media consumption and production.

I envision a new category of tangible interfaces - gesture objects - to encourage anthropomorphic projection along with geometric and psychological perspective taking. With gesture objects I project myself onto the object, I take perspectives through the object and I change and calibrate perspective taking.

Figure 1-4: Zooming in, zooming out!

1.3 Aims

Back in my days in Paris, I was passionate about photography. I was an expert in experimental photography and filmmaking. I was using photography and movie making to tell stories. My days were full of explorations of secret locations in Paris, unusual scenes, colors and ensembles. I was developing my film, my pictures, in my tiny photo lab, revealing my stories using light, zone of effects, and colors.

Here is a broken toy car that I found in a demolished hotel in Paris, see figure 1-6. I started wondering: what if toys could talk! I started collecting, hacking and recomposing stories with toys that are charged within a period.

This passion for film, photography, toys, and story-building all come together in this thesis. I convey a love for engineering new projects that change a perspective, for people to make their own meaningful stories, to go beyond the making of a movie, for people to express and tell the story about the world around them.

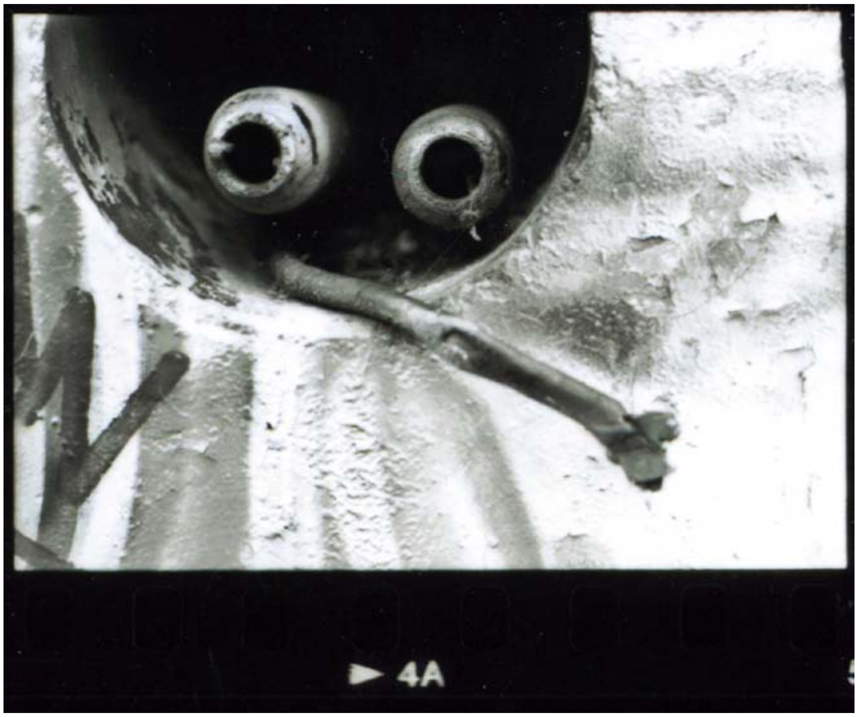


Figure 1-5: Puppet theater

Figure 1-6: What if toys could talk!

Interface design is the key for people to produce both meaningful stories, and powerful new shots that break away from the stereotypical tourist photo.

Figure 1-7: Exploring viewpoints. Camera at the eye level.

When camcorders arrived with a window to just shift the viewpoint slightly, it changed everything, the relationship to the scene, to the actors, and the type of movie created, see figure 1-8.

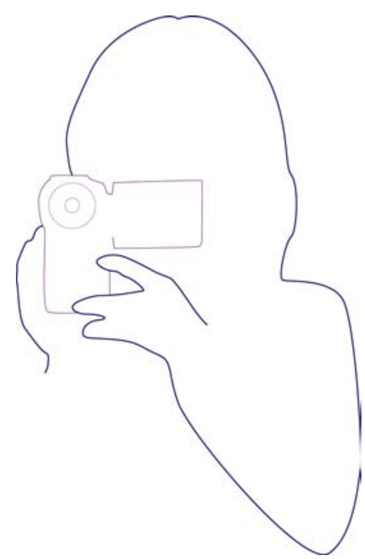


Figure 1-8: Exploring viewpoints. Camera slightly shifted from the eye.

Exploring viewpoint and its relationship to new technology is the exploration of creative affordances and I believe that this area can grow.

So far the viewpoint remains at the eye level, but what if the child's eyes and the "eye" of the camera were not one-and-the same?

What if we moved the viewpoint to the standpoint, where the hand is. Exploring the world visually through our hands might help us break the habit of what we are used to looking at. It is a way to decouple the eye.

Figure 1-9: From the viewpoint to the standpoint -where the hand is.

Picture This! formalizes the shift from the viewpoint to the standpoint in an interaction paradigm where kids are challenged to explore the world visually through their toys.

The goal of Picture This! is to discover atypical visual scenes. User's are guided by their hands as they hold their toys. They visually discover scenes through the toys being held.

Figure 1-10: *Picture This!* formalizes the shift from the viewpoint to the standpoint.

The accompanyig technology relies on gesture analysis to manage the complexity of an open-ended play space where kids have multiple and switching viewpoints available to them.

I am looking for a world where kids connect to the viewpoints of their multiple characters, a world of possible projections. When you have many eyes, you have many different ways to look at a story, at a scene. The child can embody the character who has a certain visual point-of-view! As a design problem, these new ways of seeing necessarily imply new kinds of content creation. Most importantly, if one is comfortably at play in a world of multiple viewpoints, does it draw one into taking on new perspectives, visually, and even socially?

Typically, in an interaction with a video camera, the eye of the beholder is the frame of reference. The eye looks through the camera lens or via the viewfinder and decides what to select in the video while the hand passively holds the camera, or initiates capture and sequencing.

Figure 1-11: Who is the author?

The narrative tools I create shift the frame of reference from the eye of the beholder to her hand, from the viewpoint (where the eye is) to the standpoint (where the hand is). In this shift of reference frame in video, I propose that the object held by the hand becomes the viewpoint and that the hand frames the scene while the eye controls and improvises the play. Play it by eye, frame it by hand! allows a fluid dialog between the eye and the hand by revisiting the role of the hand.

By combining atypical visual scenes to tell their stories, I expect users to practice perspective taking. They will demonstrate perspective taking in their stories as they switch back and forth between expected visual scenes and atypical ones.

By transforming the tangible handle of digital data into an entity that has a “mind of its own”, I translate the object’s view of the world into a story from another perspective. The storymaker can freely acquire and integrate the object’s view into her own story, expanding her own view. The complete *perspective taking mechanism* is as follows: Young users tell their stories, extract meaning from their experience while they are immersed in their stories during play, drive perspective taking, and feed their perspective discovery back into their real time creation.

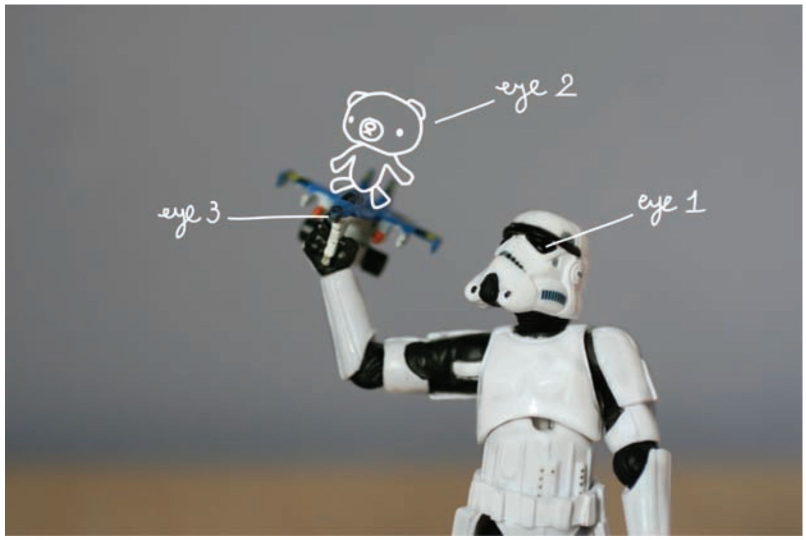


Figure 1-12: It is the recursion of the projection.

My thesis is titled “play it by eye, frame it by hand”: I shift the frame of reference from the eye to the hand.

1.4 Thesis overview

I explore an alternative video-making framework for children with tools that integrate video capture, oral stories, gestures and acting with movie production for perspective taking. I 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.

I 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.

I report my design process over the course of four research projects that span from a graphical user interface to a physical instantiation of video. I interface the digital and physical realms using tangible metaphors for digital data, providing a spontaneous and collaborative approach to video composition.

I evaluate my 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.

I present four systems that I created for movie making: Textable Movie, Moving Pictures, Terraria and Picture This!

Textable Movie is a graphical interface that takes text as input and allows users to improvise a movie in real time. By improvising movie-stories and displaying them for others, I expected that children are challenged in their beliefs about other communities as well as their own. During my observations of teenagers using video in workshops, I noticed that the complexity of traditional video tools presented several drawbacks for communication and collaboration. I hypothesized that Textable Movie would be more powerful if I could construct a physical device that would allow teenagers to easily understand and create videos using traditional cinematic language.

In my next project, *Moving Pictures*, I explore the natural affordances of everyday artifacts and integrated spatial components in the design of a tangible interface. Moving pictures: Looking Out/Looking In is a robust, tangible, multi-user system that invites young users to create, explore, manipulate and share video content with others. Moving Pictures consists of a video station containing a set of two cameras, a number of tokens, a screen and an interactive table. Moving Pictures enables a spontaneous and sociable approach to video creation, selection and sequencing. The station supports multiple input devices and group interaction, encouraging collaborative creation.

Finally *Terraria* is my introduction to the demands of play in tangible video and story making. With Terraria, children make movies with robots using a playstation joystick controller and decorate a museum exhibition space with their movies.

I propose a future generation of narrative tools exemplified by my latest project, *Picture This!*, which combines improvisation with story making during play. I present the design, implementation and studies relevant to Picture This! to establish that 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.

Additionally, with Picture This! the child's toy becomes a camera person as opposed to having the child hold a camera directly.

As a child plays with the toy that holds the camera, its video feed is projected onto a screen in front of her in real time. Picture This! analyzes the child's gestures and conduct film assembly. Automating editing with a 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 opportunities in video composition. Cameras become part of a toy system showing how things look from a toy's point of view.

These four different mechanisms for producing a visual, gestural, tactile and verbal story exist as branch alternatives to the usual fragmented and sequential capturing, editing and performing of narratives available to users of current storytelling tools. I implemented these systems analyzing their level of seamlessness, their ability to motivate children in the production of a final movie while engaging them in cross modal explorations of their gestures, their voice and the visuals around them to compose media. I propose *Play it by Eye, Frame it by Hand!*, a novel interaction technique for making videos accompanied with gestural, visual, tactile and verbal stories. Media interactions can support something more seamless and coherent than what is currently available, constructed silently through hand, eye, gesture and object dialog. A new genre of Gesture Object Interfaces, as exemplified by Picture This!, rely on the analysis of gestures coupled with objects to represent bits.

CHAPTER TWO

Theoretical foundations

2.1 Children and their stories

From an early age, we play, learn, and exchange ideas using stories; we come to know who we are and how things work 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, following us wherever we go. Shared movie-making devices can engage people in multidimensional approaches to expressing and exchanging points of view. I imagine a world in which, through play, children create and exchange narratives about their lives and their environment, and where they can tell their tales with more than words.

At a period in history when children are invaded with media and technological toys, it is essential to propose devices for authorship and open-ended play. Research has shown that toys serve a fundamental function in the development of children [Brosterman, 1997, Montessori, 1917, Montessori, 1912, Singer and Singer, 1990, Singer et al., 2006]. Literature showed that the ability to move from one's own standpoint to take another person's view is at the center of a person's personal and cognitive growth [Piaget and Inhelder, 1967, Denham, 1986, Winnicott, 1971]. And yet, as much as people cannot grow without interpersonal connection, without separation they cannot relate [Ackermann, 2004]. Cognitive growth emerges as a result of people's repeated attempts to solve the unsolvable tension between getting embedded and emerging from embeddedness [Kegan, 1982]. "Dwelling in" and "stepping back" are equally important to get the cognitive dance going, as well as "circling around" a phenomenon [Ackermann, 2004].

Through design, I seek to understand how tangible interfaces for storymaking can empower young users in expressing and sharing ideas, actively "constructing" personal narratives beyond verbal storytelling. "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 and Papert, 1991].

Movie editing systems, combined with gestural technique, may 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. I base my design exploration on a language of interaction that children are familiar with, adopting play interactions to control video making systems. One of the goals is to assuage the interface complexities of commercially available editing software. Optimally, I 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.

2.1.1 Connect to construct

Along with Ackerman[Ackerman, 1991], I suggest that we connect to our world using our senses. Each sense is a "knowledge shopper" that grounds us in our surroundings: 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. 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 play back events to evoke and reflect on our past history and to immerse ourselves 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. Our mental constructs are necessary to reach a deeper understanding of our relationship with our environment [Ackermann, 2004]. Children are offered stories by adults and are driven into fantasy play. They use toys to externalize and elaborate their mental constructions [Fein, 1980]. With character toys they create interrelationships and plots, a means to display their social knowledge: knowing about human beings and social relationships [Shantz, 1975]. If the toy offers its visual point of view, 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.

2.1.2 Talking with objects: from pretend play to role play

When a child talks to her character toys, they do not really listen to her. This is why storytelling with character props is great for an infant. She feels in control, and the interaction does not break the illusion of magical thinking, the illusion of a captive audience. The character toys are the first ones to listen to her stories. She can practice her stories, see figure 2-1.

Figure 2-1: Donald is listening to me!

Later on in their lives, if toddlers play with a voice recorder, they can hear what they said and get an idea of what an audience would hear. The recorder is the externalization of who they are, see figure 2-2. It allows them to define themselves. Listening back to their stories through the recorder, children often exclaim things like: “no it is not me” or “no it is not my voice!” while listening to their recorded voice [Vaucelle, 2002]. This is the beginning of perspective taking through telling and listening to stories.

Figure 2-2: I can listen to myself!

Toys initiate elaboration in play and language. Researchers have found a correlation between open-ended play and imagination in writers, poets, and scientists [Singer and Singer, 1990]. 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 [Singer and Singer, 1990].

According to Piaget, pretend play is an opportunity for the child to secure via fantasy what is not available in reality. For example he describes how his daughter Jacqueline, having been told that she could not play with the water that was to be used for the washing, took an empty cup, went to the forbidden tub of water, and made pretend movements saying, I'm pouring out water' [Harris, 2000].

Children later move into role-play, where they temporarily immerse themselves in the part they create. They talk from the point of view of the creature, taking the mood and tone of voice that is appropriate, give expression to the emotions, sensations and needs for the adopted role.

In their pretense, children incorporate both animate things and animated beings into their play. This role-playing happens with friends, a key form of interaction, and also with hand puppets, dolls, stuffed animals or action figures. In all cases of pretend play, the child may choose to use a prop, and may also choose to use herself as a prop. The puppet show is an embodiment of role-play which also contains a strong layer of gesture to emphasize the "life" of the characters.

If parents install a puppet a theater in their house, the child can hide herself behind the theater and become the puppet. The child offers various interpretations of her stories by gesturing with the puppets while telling a story. In the case of figure 2-3 -a puppet theater designed by Sam Scarborough [Scarborough, 2009]- it is "Mr. the-cat" that is talking so the child feels less responsibility on her stories and can express what she thinks more easily.

Figure 2-3: I can tell my toys to tell my stories!

With the character toy in hand the character gains a life of its own, and to emphasize this life, children move the toy that is talking, externalizing the character, see figure 2-4. The child teleports into the puppets, pitch shifting their voices. The puppet becomes a second self. The child lets the puppet in hand behave. And it is not just figurative. From the airplane that moves in the air, to the little bear that will be her closest friend, gesture makes these toy characters come alive.

Figure 2-4: Children move their toys to make them talk!

Beyond giving life, gestures further reinforce play interactions, see figure 2-5. Children will not only create social interactions, but also close the loop and witness their own understanding of a situation. The child moves the toy that is talking and alternates between the two. The listener is quiet while the speaker is "animated" in the double sense of the term: it is put in motion and gains agency! As the dialog proceeds, the character toys take turn.

Figure 2-5: With two character toys, children move the one that is talking.

2.1.3 Invoking a creature

> "*An invisible character, named and referred to in conversation with other persons or played with directly for a period of time, at least several months, having an air of reality for the child, but no apparent objective basis.*"
>
> Svendsen in [Svendsen, 1934].

When children engage in role-play, they do not simply remain off stage, directors or puppeteers, they enter into the make believe situation they create and adopt the point of view of one of the protagonists within it. At 2-3 years old, children can invoke a creature, an imaginary person that becomes a companion for the child, without the need of a prop. Taylor [Taylor, 1999] reports that 2/3 of a sample of American children, before the age of seven, have either an imaginary companion or an imaginary companion projected onto an external prop. It means there are a lot of imaginary friends out there, approximately 15 million in the United States, and they probably need their own online social network - *imaginary friendster*, of course! Children with imaginary companions proved to be more skilled in assessing how people might feel [Harris, 2000].

When a child plays out a particular character, she needs to have a set of meta-theories: theories about the theories that the character holds. With pretend play, children can notice the gap between representation of reality and reality itself, therefore facilitating their understanding of mental states. Children who engage in role-playing have a predisposition to be able to view a situation from another person's point of view.

What happens as children grow older? When enacting a role, children imagine the world from the point of view of another person. We do the same when we read biography, or when we're projected into a period via a historical novel. We locate ourselves inside the world of the novel rather than the real world and we share the same spatial and temporal framework than the protagonist [Harris, 2000].

Toys and storytelling serve a fundamental function in childhood development [Harris, 2000, Montessori, 1917, Singer and Singer, 1990, Singer et al., 2006, Brosterman, 1997, Birchfield et al., 2006], and the ability to move from one's own standpoint to take another person's view in a story is at the center of human growth [Kegan, 1982, Piaget, 1967, Winnicott, 1971, Ackermann, 1996]. Research shows that successful fiction writers do "live" their characters, and their rich identity is reflected back to their readers [Taylor, 1999].

2.1.4 Perspective taking

Perspective taking is the ability to understand the way people think and feel, and what motivates them to act. Perspective taking is the ability to move from one's own standpoint to take another person's view. When designing new interactions for children, I consider two types of perspective taking that have been distinguished in literature. One is geometric or spatial perceptual perspective-taking. The other one is psychological perspective taking, see figure 2-6.

Geometric or spatial perceptual perspective-taking

Geometric or spatial perceptual perspective-taking involves situations in which the perspectives of two or more protagonists are at odd with ones another. Subjects have to anticipate how a given object will appear from different viewpoints. This type of thinking is critical in developmental research tasks, such as the Tree-Mountain task by Piaget [Piaget and Inhelder, 1967, Huttenlocher and Presson, 1973] or the Cat-Dog Experiment by Flavell [Flavell, 1990]. My works combines the ability to change stand point and viewpoint within movie making and acting by turning the camera and the toys around, sharing and alternating angles of view for the audience, emphasizing and selecting visually what matters and what does not.

Psychological perspective-taking

> "*Some of the most crucial steps in mental growth are based not simply on acquiring new skills, but on acquiring new administrative ways to use what one already knows.*"
>
> Seymour Papert.

Psychological perspective-taking involves situations in which a child knows something -and knows that another does not know. The child's task is to guess what the other may believe, knowing that she doesn't know. Psychological perspective-taking is part of "false belief" experiments [Flavell, 1988, Wimmer and Perner, 1983]. It is also a part of research on children's ability to adjust speech when talking to younger children, demonstrating their ability to modify instructions to match a recipient's perceived abilities [Astington, 1988]. Psychological perspective-taking involves beliefs and knowledge about other people's beliefs and knowledge.

In Perspectives on Perspective-Taking, Flavell [Flavell, 1990] distinguishes between *within-self* or *out-of-self* perspective taking. Within-self perspective-taking is me, as I see things, and as I saw things before, or elsewhere. Out-of-self perspective-taking is me, as I feel and as I think others see things.

Role-play is to psychological perspective what changing stand-point/viewpoint is to geometric perspective taking: both are needed for a child's personal growth and cognitive development [Kegan, 1982, Piaget and Inhelder, 1967, Benson, 1993, Ackermann, 1996, Winnicott, 1971].

Researchers have explained that perspective taking is at the center of human growth [Kegan, 1982]. It allows children to test their hypotheses about the world, working with these ideas while reflecting on them. It allows them to understand other's points of view, thus functioning socially. Perspective taking offers insight into other's motivation and is the beginning of empathy [Piaget and Inhelder, 1967, Benson, 1993, Ackermann, 1996, Winnicott, 1971]. Perspective taking ability can lead to what Seymourt Papert calls: body-syntonic reasoning: "In the programming Logo world, a student could understand (and predict and reason about) the turtle's motion by imagining what they would do if they were the turtle."

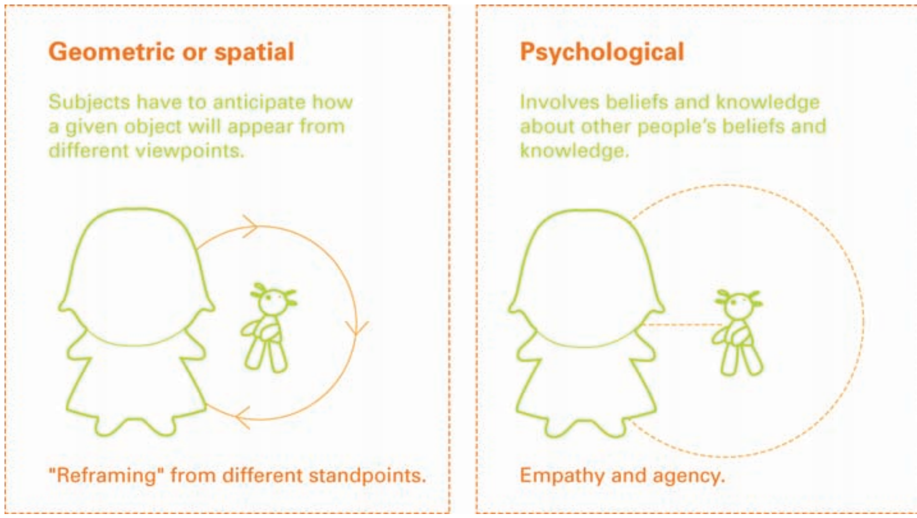


Figure 2-6: Perspective taking in Play it by Eye, Frame it by Hand.

More metaphorically, perspective taking can be the difference between looking at a tree and climbing into a tree. When you look at a tree, you imagine what the tree is like, you have no idea, so you project. When you climb into the tree, you explore the tree world, what it's made of, its secret universe and its rich environment.

Alternating between imagining and experiencing reality is key. Spontaneously, when children hang out in a park they don't want to just look at trees! They want to climb into them, hug them, and touch them to discover their texture and smells. They want to start inventing stories around them.

Alternating between imagining and experiencing is fundamental to our development [Papert, 1993, Dewey, 1938, Schön, 1983]. In this tree-case, perspective taking happens by looking at a tree and eventually climbing into it. We have climbed into trees as children, and can now, as adults, appreciate their beauty fully. We can mentally wander into them. We can walk into parks and feel replenished. Because children have formed a relationship to a tree, children start growing them, taking care of them. It is the beginning of empathy. Perspective taking is fundamental to our development and for children to function socially: to form relationships, create bonds and develop empathy.

"*Children believe that animals understand them and that they can share some of their problems with animals. Animals unconditionally love children without judging them so children can easily share their daily problems with animals.*"

Aysel Köksal Akyol and Vuslat Oǒuz in [Akyol and Oğuz, 2007].

A recent study on perspective taking concludes that the communication skills of pet feeding children improve and their perspective taking skills are affected positively [Akyol and Oğuz, 2007]. They propose that parents support their child's love and interest for their pets. Based on this result, do children playing with artificial pets, such as a Tamagotchi or a robotic pet such as Aibo, perform better in expressing empathy? Can technology help children improve their perspective taking in situations of care and empathy?

2.1.5 Theoretical approaches to storytelling

Decortis researches on the theoretical approaches to storytelling to explore technologies to support them [Decortis, 2008, Decortis, 2005, Bamberg, 1997]. The five main approaches to storytelling are the following:

Cognitive [Stein and Albro, 1997, Stein and Glenn, 1979, Mandler and Johnson, 1977] This approach is centered on the storyteller's ability to organize the content of their stories using a coherent structure.

Interactive [Quasthoff, 1997] This approach is related to the interactive activity between narrator and spectator by decontextualizing the narrator from its narrating activity. This implies a co-participation of the readers in the story.

Constructivist [Bamberg, 1997] This approach indicates that narration belongs to language use. Because the child learns linguistic structures for interpersonal and social ends, as long as the person's life is involved through participation in linguistic practices, a story using the first person voice has the same importance as a story told with a narrator voice (the third person voice).

Cross-Cultural [MacCabe, 1997] This approach is specific to the language itself. The language can be musical, silent, or visual. Researchers investigate the various narrative structures that result from parental influences, gender, and culture.

Socio-cultural [Nicolopoulou, 1997, Vygotsky, 1978, Bruner, 1983a] This approach investigates the tight relationship between play and narration. Narration is an activity in which children build reality and give meaning to experience. Storytelling activity is sociocultural. It is an active interplay between the individual and his/her cultural world.

The usual research questions based on these approaches are:
- How are the episodes linked?
- Does a story have a beginning, middle and end?
- Are the sentences articulated in a logical manner?
- How many events make the story and what connects the story?
- How many words are used to embellish the heart of the action or event?

Decortis [Decortis, 2008, Decortis, 2005] has looked at these theories for use in technological environments and showed that only few researches are centered on the narrative production as a constructive "process". Researchers could look at how a child creates meaning of her experience through her story, what motivates a child to tell a story, or if the child changes her story based on her audience.

Based on this theoretical foundation, my research focuses on perspective taking as a motivational and foundational factor for children to make multimodal stories. I implement technologies that enable children to share and reflect on their new visual discoveries. I seamlessly integrate capturing, editing and performing movies in a single interface to affect the child's storytelling experience. I research the form factor of the video camera as a source for new visual discoveries. For instance, children move the camera from the eye to the hand, and thus explore the world visually through their hands. I propose an interaction design mechanism that provokes children's expression of their own standpoint to experience another person's view. I advocate a new category of gestural and video based story-building tools. I design these tools for the purpose of leveraging the natural expression of play while expressing and alternating between viewpoints and standpoints. I aim to engender new visual and narrative perspectives as part of automatic film assembly during children's play.

2.1.6 Storytelling and literacy

My interest in storytelling started with designing systems for literacy learning [Vaucelle, 2002, Ryokai et al., 2003]. The research was based on the theory of emergent literacy [Teale and Sulzby, 1986]. In the emergent literacy view, aspects of both oral and written language develop concurrently rather than sequentially [Goodman, 1986]. According to this view, literacy learning does not happen only in formal classroom settings, but also in informal settings, in both oral and written modes, and in collaboration and interaction with others.

As explained by Ryokai and myself, [Ryokai and Vaucelle, et al. 2003], the *inside-out* and *outside-in* skills of literacy are being distinguished in research [Whitehurst and Lonigan, 1998]. Inside-out skills are concerned with children's phonological and syntactic awareness, and grapheme-phoneme correspondence, thus facilitating children's ability to decode information within a sentence. Outside-in skills are concerned with children's ability to take the meaning of a sentence from the context in which the sentence is placed (e.g. understanding who 'she' refers in a phrase, 'then she ate the poisoned apple'). Children must bring their knowledge about the world and apply that to the text. These outside-in skills of literacy exemplify children's knowledge about language and how it works in a given context are what drives my storytelling research about the kinds of language activities important for the transition between pre-school to school. Young children's language is initially limited to concrete here-and-now talk.

Early words rely on physically present objects and scaffolding from a familiar conversational partner with whom the child can assume shared knowledge [Ninio and Bruner, 1978, Nelson, 1996]. Thus, the acquisition of outside-in skills, which require gaining independence from physical and temporal context, marks a significant transition in a child's literacy development. Snow[Snow, 1983] introduced the term 'decontextualised language' to refer to language that is not bound to spatial or historical context. Storytelling, then, provides an ideal forum for children to practice decontextualised language since it avoids any laborious writing tasks. Rather than concrete *here-and-now* talk, storytelling encourages the use of *then-and-there* language [Scarlett and Wolf, 1979]. In order to tell a comprehensible story, children must be able to hold the audience's perspective in mind and reconstruct the original context [Cameron and Wang, 1999]. Children learn these skills through interaction with both adults and peers [Ryokai and Vaucelle, et al. 2003].

In this thesis, one of the main purposes is to rethink the trade-off and articulations between decontextualised language and "narration in action" through gesture and performance. I design tools to record and re-assemble images, voice and gesture into narrative sequences.

2.1.7 Narrative perspective taking

Narrative perspective taking is the ability to express the ways the characters think and feel and what motivates them to act, along with the ability to communicate to their audience when and where their story takes place. Both require the ability to symbolize, i.e. take one's own action as an object to think with and evoke objects at a distance. Evocations can be through words, pictures, gestures.

Children embed perspective taking in their stories when they provide contextual information in their stories, for instance when the children specify when and where the story takes place. Narrative perspective taking is also demonstrated by describing the internal states of their story characters, for instance by expressing the emotional state of a character. Children can also introduce their characters in a story with the use of framing clauses, such as, "the lady with the pink hat said". When children use framing clauses to introduce their story characters, they offer the audience cues as to who is talking. These cues are key for the audience to follow the storyline. They are markers of perspective taking. High pitch-sifting or low pitch-shifting a voice while "talking" for a character can also reinforce the framing clauses, becoming a marker of narrative perspective taking in an oral story.

In my work I design systems to engage children in the exploration and expression of narrative perspectives in their stories. Not only does this extra information make the story richer and easier to follow for the audience, but research also indicates that early narrative and perspective taking are a predictor for academic success in pre-school children [Snow, 1983, O'Neill and Pearce, 2001a, O'Neill and Pearce, 2001b, O'Neill et al., 2004].

2.2 Technologies for story-making

Taking this framework into consideration, I explore the potential of new technologies to favor children's narrative competences through perspective taking. I claim that the access to a new frame of reference is a window into a new world. I propose a new class of story-making tools that bring children to discover different points of view through a cross-modal experience of gesture, object, audio and video. These tools provide the structure for perspective taking while motivating children to create finished stories. Our technologies, even when they work within creative environments, such as storytelling or film-making, can do much more to promote social goals. Perspective taking and frame of reference are these kinds of fundamental social goals that have value, and currently, we have no way to structure it in our interactions.

2.2.1 Immersive techniques for perspective taking

Immersive and sensing techniques from designers such as Chris Woebken's animal superpowers, allow children to experiment what it feels like to be a "bird" or "an ant", see figure 2-7. To do so, children wear a custom designed helmet with a camera integrated and their vision is magnified by 50x [Woebken, 2008].

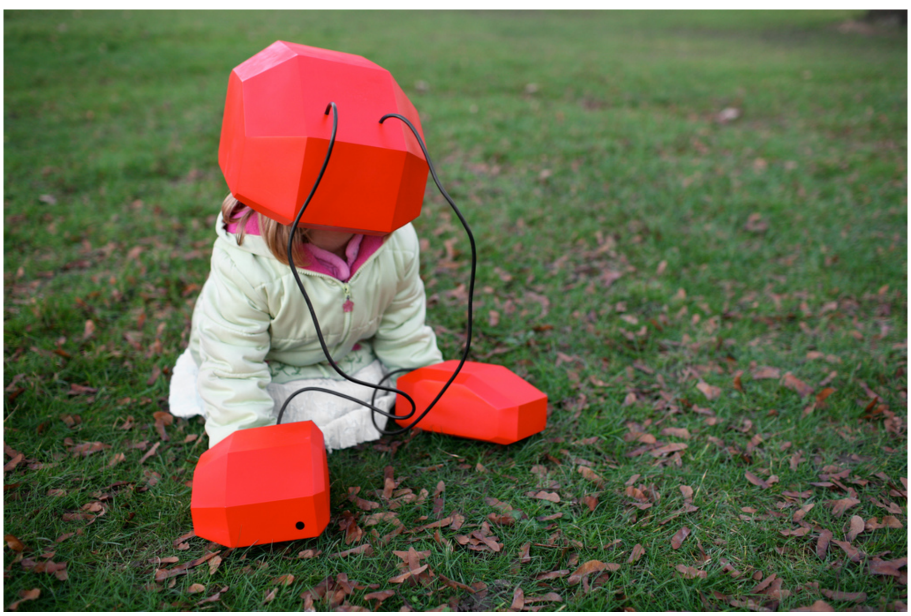


Figure 2-7: Chris Woebken's animal superpowers.

Outfitting a costume to dwell in an elephant's point of view as in Gemma Shusterman's SuperElephant, from her series Endangered Senses, allows people to experience a sense possessed by endangered animals that is not possessed by humans. The elephant-inspired costume investigates the pachyderm's ability to detect infrasonic and seismic vibrations. The wearable has long telescoping sleeves which conceal the arms and hands and connect to the floor [Shusterman, 2006].

The Purple Moon software invites girls to practice their social skills by taking on different roles and perspectives and trying them out in relationships [Laurel, 1997]. In the same vein, Sherry Turkle analyzed our interactions in chat rooms with avatars and explains that social networks allow us to safely experiment with identities and ways of being [Turkle, 1995]. In Turkle's work, the safety comes from the fact that the world is virtual.

These projects invite us to immerse in someone else's point of view; when a child makes a movie, she dwells in as she creates and tells her story and steps back as she watches the movie or re-edits what she just made. Designing an interactive system that seamlessly supports this back and forth dynamic could allow children to iterate while reflecting on their new discoveries, garnered from their immediate creations.

Constructionist educators have demonstrated that young learners benefit from systems that support self-expression, because children "learn by making" [Harel and Papert, 1991, Resnick, 2002, Ackermann, 1996, Ackermann, 2004]. Researchers are currently exploring mixed reality learning environments [Birchfield et al., 2006, SMALLab, 2008] and exploring the use of game design and game-inspired methods to teach critical 21st century skills and literacy [Salen and Zimmerman, 2004, Salen, 2007].

In her project Feral Robotic Dogs, Natalie Jeremijenko hacked robotic toys and transgressed their robotic learning potential towards an environmental "bias". The dogs look at the world through specific filters, e.g. pollution sensors, and bringing back the associated data driving a reaction such as the discovery that one's landscape is highly polluted, see figure 2-8. Thus, through the dog's eye, children are invited to re-discover their environment and make sense of it [Jeremijenko, 2003].

Still, most technological toys are not designed to provide space for children to tell their own stories; rather, the toys tell stories to the children. Storytelling composition tools or "story-builders" could support children's authorship; however, the interaction paradigm in traditional video editing systems often lacks fluidity and has a restricted view of how to build a narrative sequence; in many cases, the objective is to make an immutable "final" cut of a movie. The author

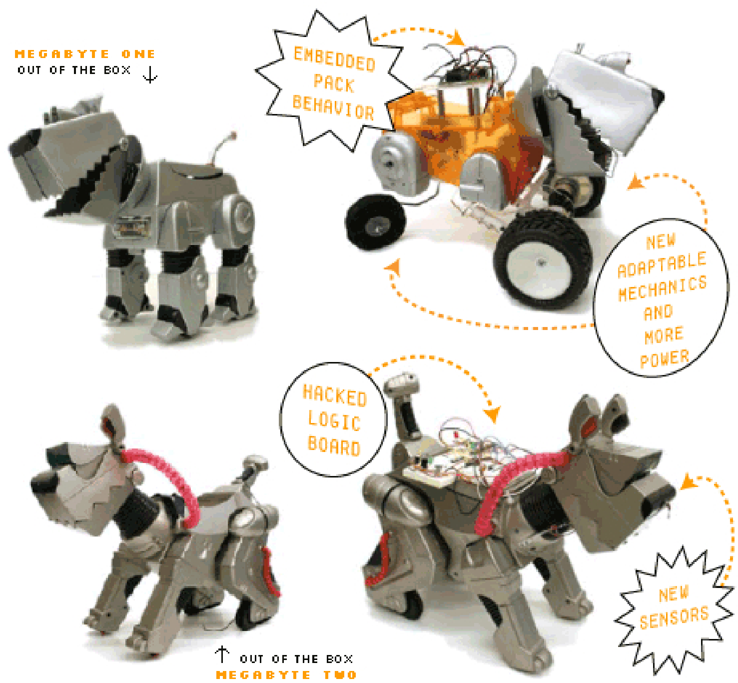


Figure 2-8: Feral Robotic Dogs by Natalie Jeremijenko.

can only see the whole once she renders the timeline. Furthermore, creating movies requires media composition and narrative skills, which existing user interfaces scaffold poorly for novice users and children [Landry, 2008].

A cultural shift occurs with video games; video games seem to have the potential to fulfill an important function in children's play. Analyzing the gaming experience in interactive systems, researchers propose a theoretical framework that also identifies video games as a storytelling medium that provides contexts for social play [Salen and Zimmerman, 2004]. With tangible digital video tools, video making could be coupled to play interactions. This would significantly change the video-making process so that children could more easily use video as an expressive composition tool. It would also help children contextualize speech and decontextualize gesture in entirely new ways.

Oral stories in children's play make use of linguistic structures (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]. Sharing media can engage people playfully and mindfully in multidimensional approaches to shifting and integrating points of view.

Digital cameras allow testing, experimenting and materializing ideas. However, even though people freely experiment with digital media, online sharing media networks such as Flickr are inundated with redundant and highly similar shots. Why bother looking at the same picture of the Eiffel Tower taken by millions of users if one can find a professional quality one! People care about their pictures because they are personally meaningful, despite that the same photos can be boring for their audience or even to themselves later on. This adversely affects the ability of the digital camera, as a tool, to effectively drive narration or tell a story. In contrast to the affordances of current cameras, individuals could be driven to compose media that not only conveys more information to their audience, but also to compose media that could be more personalized, exploring a unique stylistic expression in narration by exemplifying a personal point of view of a scene.

I am developing a new generation of narrative tools through which the video-making process becomes ubiquitous within children's play patterns. My goal is to engage children to use video as an expressive composition tool, video making and assembly being part of their everyday play with objects and toys.

2.2.2 Traditional Media Sequencing

Temporally sequenced media - film, animation (2D and 3 D), radio, TV - are usually created in two or more distinct phases: original recording of media segments and the subsequent editing of these segments into a coherent media artifact or story. In current practice, these segments are then viewed, digitized (if not already digital), labeled and placed in a database or bin for future retrieval. Typical labeling systems include timecode which references recorded time, segment name as in a unique number issued by the camera or a verbal name which is given to the segment by the human after the media has been captured, and keywords. In previous work the media segment is also identified in a digital system by means of an image icon or "micon" [Brondmo and Davenport, 1989]. In the editing phase, one or more of these labels are used to retrieve particular media segments. In digital systems, a time-line is often used to position these segments into a linear temporally sequential arrangement or media realization. Once saved, this realization or artifact can then be viewed on individual desktop screens, on TVs, or projected on a large screen. It can also be sent as a digital artifact to others or transformed into a film artifact.

While this sequential approach has been suitable for professional production of media like cinema, it has several drawbacks for collaborative video creation. For instance, people who are not technically well versed and/or have short attention spans are averse to spending the time required to digitize and label their media. Even when labeled and placed in a bin, the labels do not help the user relate these bits and pieces of media to the "story" they had in mind when they began to record the media. For those people who manage to overcome initial hurdles, reformulating the time-line is difficult not only for novice editors but also often for professionals.

2.2.3 Tangible User Interaction

My work relates to research on tangible user interfaces [Sharlin et al., 2004] and tangible bits [Ullmer and Ishii, 2000] that combines physical objects with digital data. Digital data covers physical objects in a display space [Patten et al., 2001]. Physical objects collapse in the physical world and extends into the digital space of the screen [Lee and Ishii, 2010]. Actuated tabletop displays are now able to render and animate three-dimensional shapes with a malleable surface [Leithinger and Ishii, 2010]

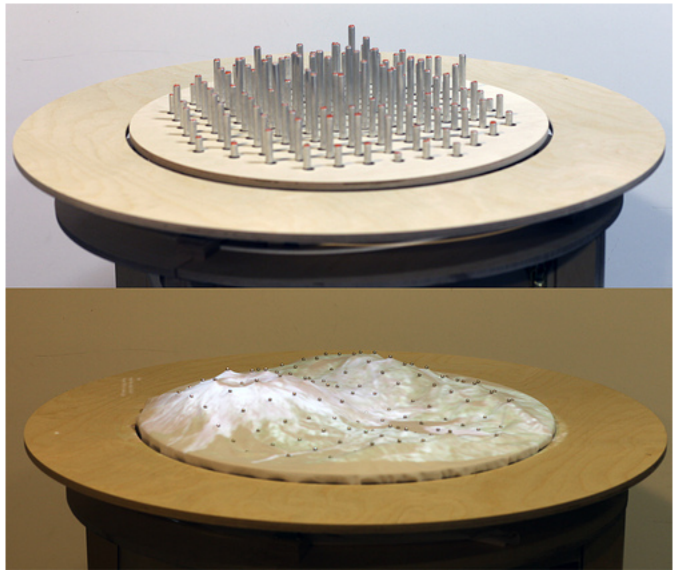


Figure 2-9: Relief by Daniel Leithinger, 2010.

2.2.4 Tangible User Interface for VideoJockeying

Derived from disc jockey (DJ), the term VJ was used for the first time at the end of the '70s. A disc jockey performs pre-sampled sounds in real-time, a video jockey is a live performer of visuals. The mechanism of video jockeying is similar to the mechanism employed by silent film directors in constructing a narrative using visual elements, and later in accompanying these visuals with live music [D-Fuse, 2006].

Robots are also videojockeys when they perform and edit movies in real time. The Filmmaking Robot of Douglas Bagnall edits short films by selecting video footages and taking aesthetic decisions [Bagnall, 2004]. Tangible systems move and sequence digital media clips, arranging digital information physically [Jacob et al., 2002, Ullmer and Ishii, 1999], create multimedia stories [Mazalek and Davenport, 2003], access digital information using tokens [Holmquist et al., 1999, Ullmer and Ishii, 2000], and that use multiple handheld computers to organize digital video clips [Zigelbaum et al., 2007, Sokoler and Edeholt, 2002]. A set of small displays can be physically manipulated to interact with digital information [Merrill et al., 2007]. Tangible mixing tables enable a performance-oriented approach to media construction [Lew, 2004].

2.2.5 Tangible User Interface for co-creation

One of the goals of my work is to rethink what it means to work together, either with other people or toys in the creative process of story-making, or designing narrative sequences together using perspective-taking techniques.

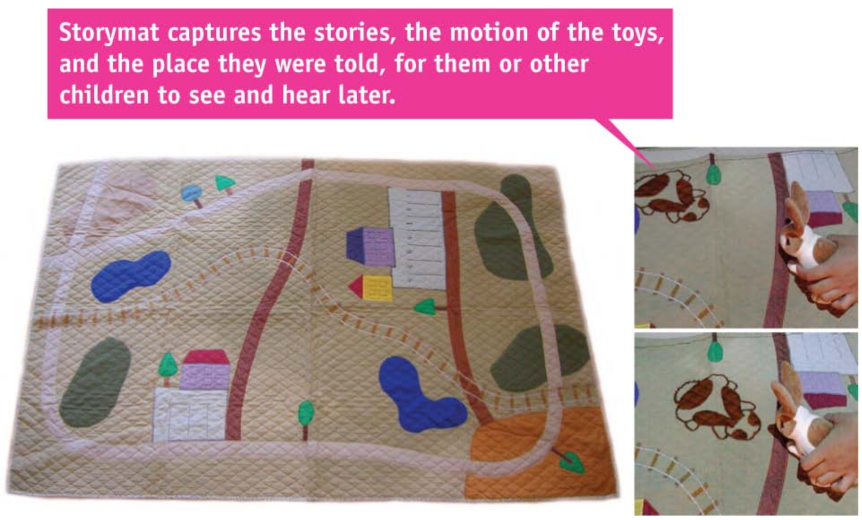


Figure 2-10: Storymat by Kimiko Ryokai, 2001.

A broad range of interactive table-tops have been conceived for collaboration. The DiamondTouch table [Dietz and Leigh, 2001] invites multiple users to collaborate using shared digital media. Tangible User Interfaces are also designed to encourage collaboration between children [Africano et al., 2004, Ryokai et al., 2003]. Tangible mixing tables enable a performance-oriented approach to media construction [Lew, 2004]. In StoryMat, a childhood map is augmented digitally to invite children to tell stories using props. The mat stores children's storytelling play by recording their voices and movements of the toys they play with. These stories revive on the mat as other children play and tell their stories [Cassell and Ryokai, 2001], see figure 2-10. With ClearBoard, users draw together digitally while talking to each other [Ishii and Kobayashi, 1992]. In I/O

Brush, children use a paintbrush to gather picturesque information from their surroundings and share them with their peers digitally [Ryokai et al., 2004].

My research contributes to recent attempts in supporting human-human collaboration with ubiquitous computing [Salvador et al., 2004], especially with research on ubiquitous computing devices for sharing pictures at a distance [Truong et al., 2004]. Researchers have designed a dollhouse, augmented with microphones and cameras, to allow the toy inhabitants of the house to communicate with the inhabitants of a remote dollhouse [Freed, 2010].

2.2.6 Tangible User Interface for Storymaking

My work contributes to the design of tangible storytelling tools for children [Frei et al., 2000, Montemayor et al., 2004]. Much of my research is centered around the design of a tangible movie-making machine for children thus complementing previous work on supporting children's fantasy and storytelling [Cassell et al., 2000].

In Pogo, researchers envisioned a system that plays visual sequences using tangible objects [Rizzo et al., 2004]. Cameras were redesigned to capture both the child and the video of the child to contextualize a recorded visual scene [Labrune and Mackay, 2005]. In I/O Brush, children use a paintbrush to gather pictorial information from their surroundings [Ryokai et al., 2004]. 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 and Ryokai, 2001].

In Jabberstamp, children synthesize their voices in their drawings [Raffle et al., 2007]. Jabberstamp is an interactive craft-based tool that allows children to synthesize their drawings and voices. Children bring their drawings to life with familiar materials like pens and paper. Frst, children make drawings on normal paper, and then press a special rubber stamp onto the page to record sound on their drawings. When they touch the marks of the stamp with a small trumpet, they hear their sounds playback, retelling the stories they have created, see figure 2-11.

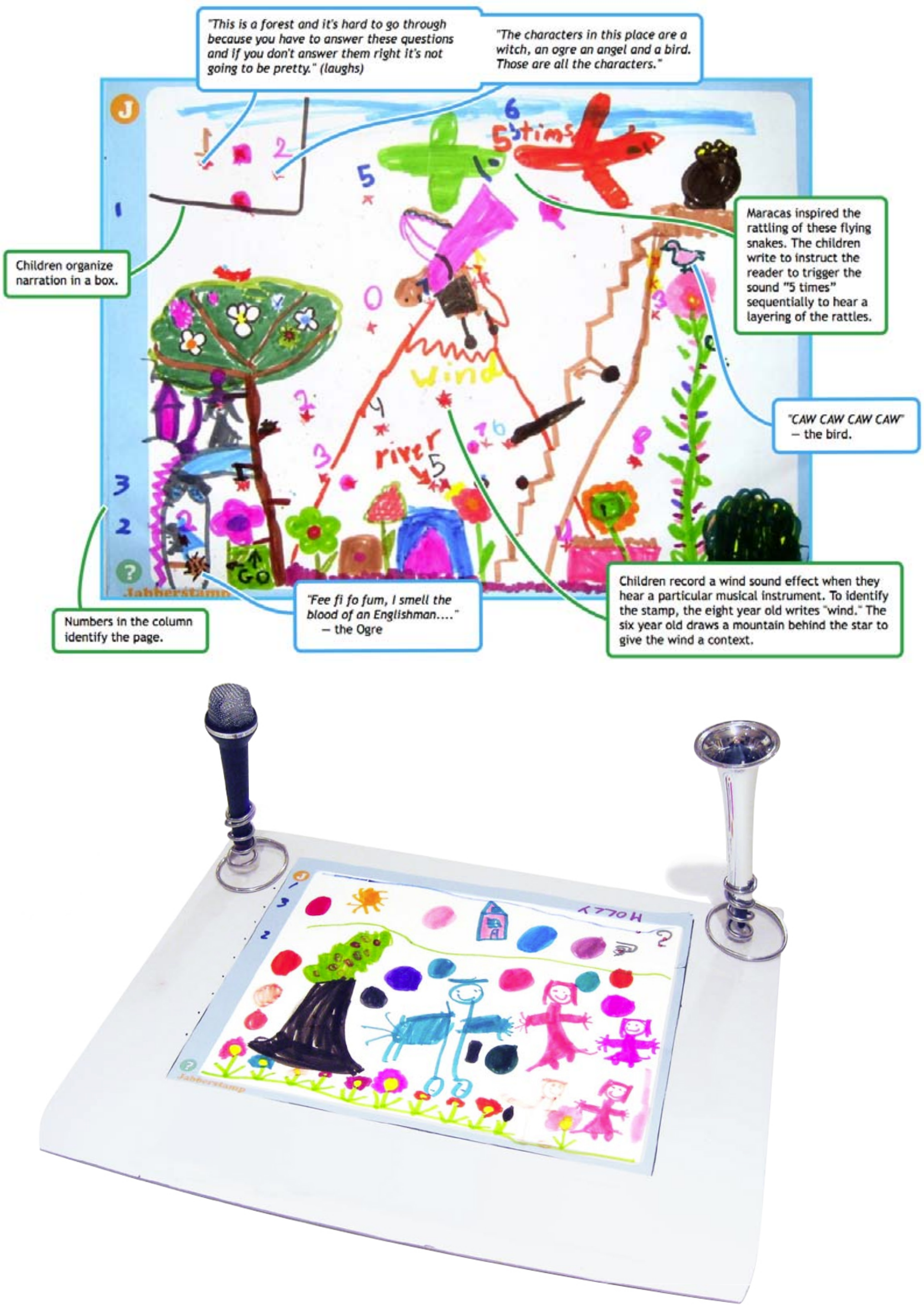


Figure 2-11: Jabberstamp by Hayes Raffle, Cati Vaucelle and Ruibing Wang, 2007.

Yumiko Tanaka's Plable is a traditional looking table with which children build an imaginary world [Tanaka, 2006], see figure 2-12. Designers developed a new concept for movie editing to help children understand the process of editing. It consists of printed movie cards that can be re-arranged in any order. Their bar code is used to identify them on a digital screen [Miyabara and Sugimoto, 2006]. Offering authorship though the interaction with tangible interfaces is rare. It is probably because it requires a flexible interface and a software architecture that takes care of data management. Authorship allows children to become active participants instead of simply observers.

Figure 2-12: Plable by Yumiko Tanaka.

In Flights of Fantasy [Davenport, 2001] everyday visitors in a gallery move blocks around a tangible table-top to edit sequences based on icons that represent story elements. Researchers developed a system that replays visual sequences using tangible objects with a stationary computer for capturing and associating media to objects [Rizzo et al., 2004]. While these systems invite for capturing and editing movie segments, none of them propose to edit, perform, publish, or share final edited movies with peers.

2.2.7 Gesture Object Interfaces

The function of gesture is also critical to my 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.

Gesture interfaces

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 thesis, I propose a gesture language for capturing and editing that is suitable for children in their toy environment. More specifically, I 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 [Premaratne et al., 2006]. 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 [Letessier and Bérard, 2004]. A common challenge shared between these examples and the latest system presented in this thesis, Picture This!, is distinguishing between intentional and unintentional gestures.

Other work has proposed gestural interfaces with objects even though they did not describe their work in these terms. In Office Voodoo, users move dolls to control parts of a sitcom; Office Voodoo is an interactive film installation for two people. It tells the story of Frank and Nancy, two bored Irish officemates, condemned to spend their lives in an office. This infinite film is an algorithmic sitcom inspired from Sartre's play "Huis clos", crossbred with an office life simulator. Two physical voodoo dolls, that represent the protagonists, can be manipulated in order to change the emotions of the characters in the film, see figure 2-13 [Lew, 2003]. In work on sympathetic characters, children manipulate a plush toy to control characters in a three-dimensional virtual environment [Johnson et al., 1999] or to control music expression [Yonezawa et al., 2001].

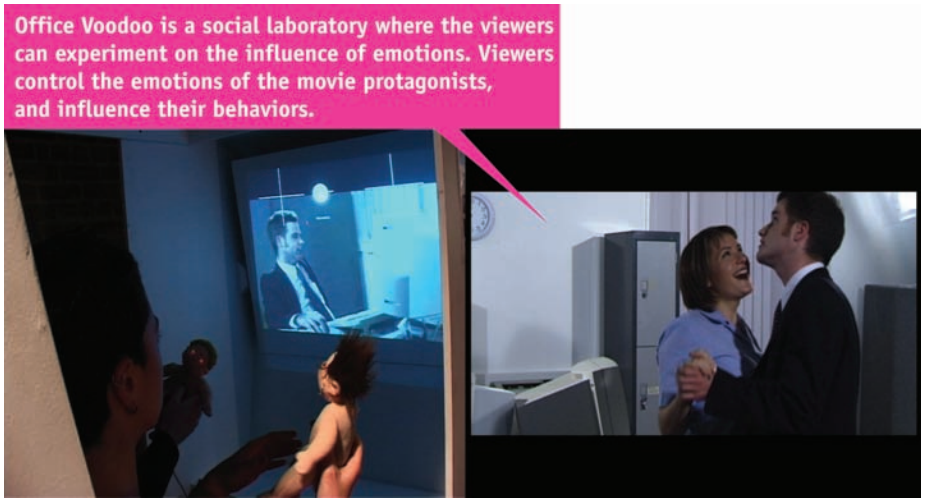


Figure 2-13: Office Voodoo by Michael Lew, 2003

Users trigger the objects to control visuals on a display and rarely anthropomorphize the objects being manipulated.

In Topobo [Raffle et al., 2004], 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 [Weller et al., 2009, Weller et al., 2008].

In OnObject users wear a small device on the hand to program physical objects that respond to gestural triggers [Chung et al., 2010]. OnObject enables novice end users to turn everyday objects into gesture interfaces through the simple act of tagging. Wearing a sensing device, a user adds a behavior to a tagged object by grabbing the object, demonstrating a gesture trigger, and specifying a desired response, see figure 2-14. Following this simple Tag-Gesture-Reponse programming grammar, novice end users are able to transform mundane objects into gestural interfaces in 30 seconds or less.

Figure 2-14: OnObject: gestural programming by Keywon Chung, 2010.

2.3 Systems of Inquiry

Designing a computational object means designing for people. It demands reflecting on the object's critical and aesthetic roles. To this end, researchers propose that designers develop sensitivity to and control of aesthetics, for instance by designing purposeful constraints on communications media [Gaver, 2002]. I rely on ethnographic methods to understand how people use my systems. I spend time with a distinct group of children, interview them and capture videos of their interactions [Taylor, 2009]. Afterwards, I analyze the field materials. This system of inquiry is the evaluation context for the research in this thesis.

Figure 2-15: Participatory design session, Textable Movie, Dublin, 2003.

The methodological approaches I am using to evaluate my work are two design research methods. I rely on these two approaches in two separate phases.

First, I run participatory design sessions with children, during which children are design partners in my projects [Douglas and Aki, 1993, Druin et al., 1998]. My design decisions are coupled with user participation. I combine form, interaction and function throughout my design process informed by research on rich user interfaces [Frens et al., 2004]. With children as co-partners, I design a series of prototypes tested at early stages. I invite the children, as co-designers, to produce movies and stories with the tools designed, and finally, I interview them to gather their insights.

During the second phase, I conduct comparative analysis, revisiting my research prototypes based on how the children understood the systems and made use of them. I use this "post-reflection" to create design guidelines for my next iterations. This phase relies on a comparative analysis of each tool's affordance, using the Cognitive Task Analysis Methodology to address both the environment in which the problem solving takes place as well as the problem solving activities itself [Steinberg and Gitomer, 1996, Steinberg and Gitomer, 1993]: conduction both task analysis: analysis of the task by the researcher, and the problem space: analysis on how learners construct their interpretations of the system they are working on and the process required to manipulate it, that is, their mental model of the system [Jonassen, 1999, Jonassen, 2000, Gardner, 1999]. I compare my three cases in the next chapter to understand how the interface influences the performance of "movie sketching" and enables children's story-making.

For the three cases I have designed, I created three accompanying prototypes. I discuss how, with the successes and failures unearthed with each of these three cases, I gain a new framework to use technology to access perspectives during play. I will be able to distinguish each project per functionalities and outcomes. My final project will be evaluated and compared in regards to the three cases' major findings and lessons learned.

CHAPTER THREE

Design Experiments

The purpose of my work is to put gesture, words and images at the service of narrative competences, beyond movie construction. I aim to enable shifts in perspective (and projection) by varying points of view and by encouraging kids to reframe an event in different ways. The technical tools I implement are designed to facilitate the layering or re-assembly of recorded story segments for further usage.

3.1 Dolltalk: Gesture Object Story-Making

Dolltalk was my first attempt to define the area of Gesture Object Story-Making during which I establish the ability to access perspective as part of gesture analysis built into new play environments [Vaucelle, 2002], see figure 3-1.

3.1.1 Motivation

School curricula are designed with the expectation that students achieve literacy. Typical models support the acquisition of language by encouraging students to learn how to decode information within a sentence. With Dolltalk, I suggest that literacy skills in children up to eight years of age can be obtained along with a broader understanding of language and its representation. Oral storytelling is presented as a way to develop meta-cognitive skills with a focus on character-based narrative, where children must create the perspectives of the characters.

Figure 3-1: Dolltalk.

3.1.2 Scenario of Interaction

Dolltalk is an augmented puppet theater for children to practice pre-literacy skills. It features a virtual agent, an Alien, who asks children to tell stories and plays back the stories to the child using narrative features. Although it was originally designed for a single child, the system can support two children playing together.

The following is a scenario of a child playing with Dolltalk, an excerpt from my user study:

Zia (the Alien): You know that the most popular game on planet Blooper is to race with yellow stars! Kids love to catch and ride on yellow stars. I'm anxious to hear a story about the most popular game on Earth. Can you tell me a story about your favorite game with the two puppets?

Maria, 8, laughs while listening to the story of Zia. Maria quickly grabs the two puppets, swaps out the different features: ears, eyes, nose, etc. to create the characters she wants, and dresses them: one with long hair, one with a scarf.

Maria: (Without moving the puppets) Once upon a time there was two friends and they were happy together. So one of them said to the other,
(Moving the first puppet) Do you want to play?
(Without moving the puppets) And the other one said,
(Moving the second puppet) OK.
(Without moving the puppets) And so they went to the park. One of them went on the swing, the other went on the slide. The one, the person that was on the swing, they got hurt. Then the other one went on the swing and picked her up to, to see if she was alright, and she was. The end.

Then, Maria quickly gives the two puppets back to Zia.

Zia: That's so cool, let me see if I understand, do you mean this?
(normal voice) Once upon a time there was two friends and they were happy together. So one of them said to the other:
(high voice) Do you want to play?
(normal voice) And the other one said,
(low voice) OK.
(normal voice) And so they went to the park. One of them went on the swing, the other went on the slide. The one, the person that was on the swing, they got hurt. Then the other one went on the swing and picked her up to, to see if she was alright, and she was. The end.
Cool! You know kids on planet Blooper love to eat sugar clouds. Kids love to try different flavors of sugar clouds. I'm curious to hear a story about what kids on Earth love to eat. Can you tell me a story about your favorite food with the two puppets?

3.1.3 Design Principles

Dolltalk is a gesture based storytelling system that 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 to discover narrative perspectives during storytelling play. The child tells her story to an animated computer character, using two stuffed animals as props while her story is recorded.

The combination of sound and gestural parameter analysis allows Dolltalk to determine the structure of the narration: from a dialog to a narrator voice. In Dolltalk *easy mode*, when the child is done telling a story, the recorded audio is played back with two different pitches to signify the stuffed animal that was speaking at the time. In a more *advanced mode*, the story is played back using a narrator voice and offers different understandings for the same story using the storyline of the child.

3.1.4 The Dolltalk technology

Dolltalk includes a puppet theater with two modular hand puppets, microcontroller, accelerometer and integrated RFID, see figure 3-2.

The child tells his/her story to an animated computer character, using two stuffed animals as props while their story is recorded. The puppets contain accelerometers that monitor the movement of the toys; statistical analysis of children's play with props allows the system to assume that if a toy is being shaken, then the child is narrating a story segment associated with that toy. The system also computes different types of motions, from a single toy being shaken, to multiple ones and their frequencies over time. In parallel, the system analyzes the speech of the child.

A detailed technical contribution and description is presented in my master thesis [Vaucelle, 2002]. In summary, the Dolltalk software I created is a real time analysis and processing software for audio, RFID and accelerometer data. In real time, I run sound analysis with environmental noise detection and and sensor data analysis. I index the sentences for three voices: two character voices and one narrator voice. I store the audio and gesture data in my software for later retrieval. The playback is then automated: I segment three voices, pitch shift the child's voice, animate a 2D character based on sensor data and sound detection. My algorithm creates character framing clauses played back by the 2D character.

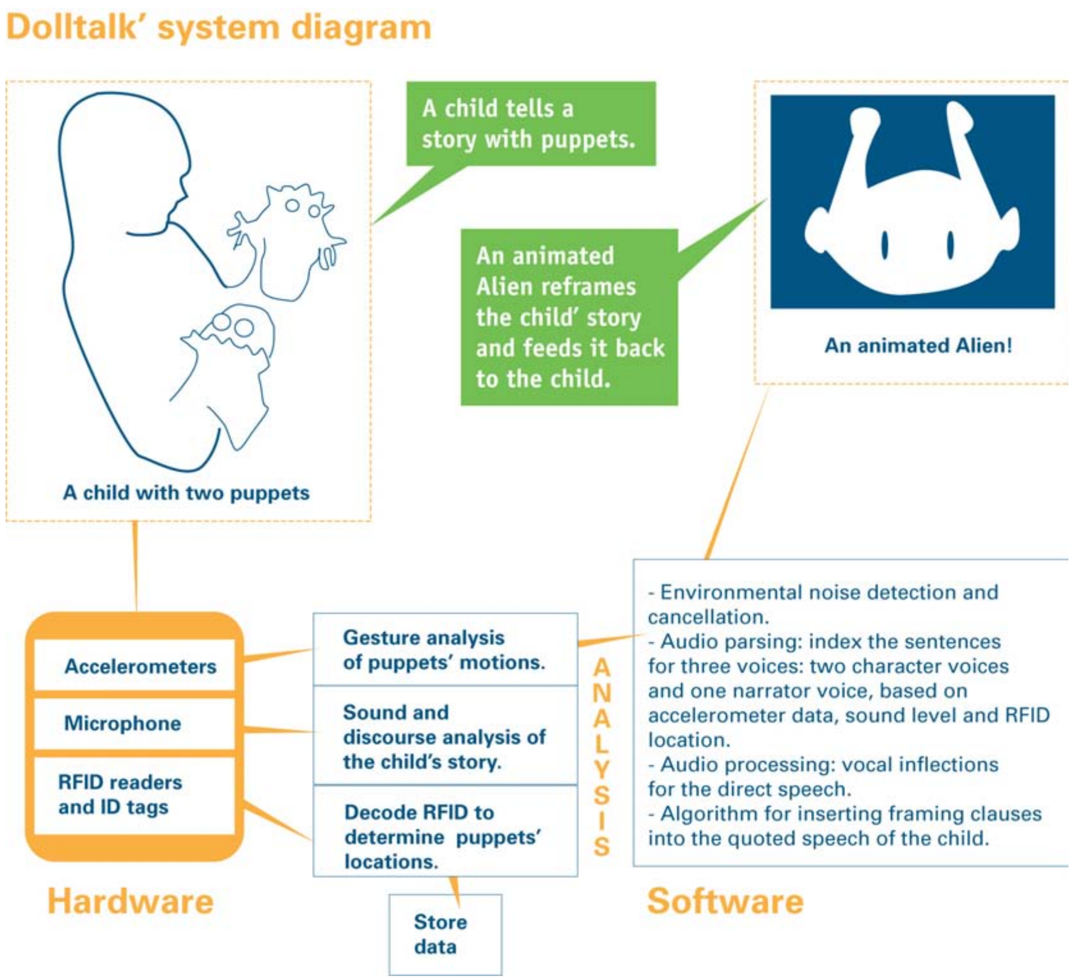


Figure 3-2: Dolltalk specification.

3.1.5 The Dolltalk study

A user study was conducted to understand the short-term effect of Dolltalk on children's elaboration of internal states of story characters.

Setup

To evaluate the system, I ran a controlled study with two passive tests and two active tests: one being the control. The Dolltalk version has modified playback, where the system plays back recorded story segments, pitch-shifting the segments into high and low segments to distance the child from her own recorded voice, as if the Alien is acting out the child's story. The evaluation consists of an in between-subjects study with 16 children 6-8 years old. Each session was videotaped and transcribed for analysis.

The study compares Dolltalk to a regular tape recorder. The tape recorder version includes all aspects used in Dolltalk with no modified playback.

The study includes a pre and post passive test with puppets and theater. Two groups of children participate in three stages of interaction: Passive Pre-test, Active Test, and Passive Post-test. The active test for Group 1 consists of full version of Dolltalk with modified playback of stories. The active test for Group 2 comes as a control. It consists of the *tape recorder* version in which the playback of stories is unmodified. The Passive tests are identical for the two groups with no story playback. This is to evaluate the content of the stories and quantify the frequency of perspective taking markers, before and after the interaction.

Figure 3-3: Children playing with Dolltalk.

I quantify the frequency of quoted speech with framing clauses, e.g. she said: "let's go!, internal states of characters, e.g. "she was sad", and temporal and spatial expressions, e.g. "the one that was on the swing".

3.1.6 Findings

The results show that playing with Dolltalk encourages children to introduce their characters into the story and to express the internal states of their characters much more than with the use of a simple tape recorder. The results also show that playing with the current version of Dolltalk or with Dolltalk in tape-recorder mode encourages children to provide spatial and temporal information in their stories much more than they would without hearing any playback of their stories. My master thesis presents significant results (p =.04) that indicate the current version of Dolltalk encourages children to express the internal states of their characters (Vaucelle, 2002).

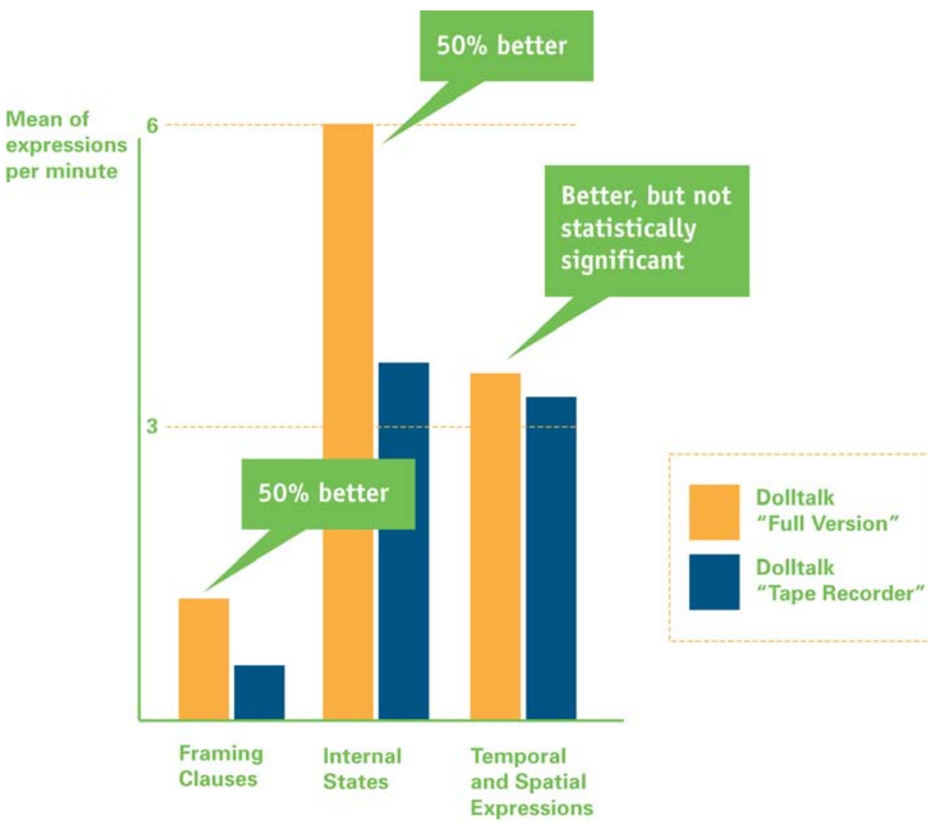


Figure 3-4: Results of the Dolltalk Study.

Both the full Dolltalk system and the tape recorder system encourage children to express the internal states of their characters, provide spatial and temporal information, and frame characters' dialogue. However, the difference between the frequency of internal-state expressions and the frequency of temporal and spatial expressions is small for the tape recorder (3.66 for I.S. and 3.33 for S.T.), while the difference is quite large for the full Dolltalk system (6.03 for I.S. and 3.50 for S.T.). The main feature of Dolltalk seems to be its ability to encourage children to express the internal states of their characters.

In summary, one of the conclusions of this empirical research is that the full version of Dolltalk performs significantly better (above 50%) for the description of internal states of the characters and the framing clauses used to introduce the characters. Dolltalk encourages almost twice the expression of internal states and quoted speech, which are both important markers of narrative perspective taking, see figure 3-4.

With Dolltalk I 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 with toys while analyzing changes in voice prosody. Using sensors and audio analysis, the system interprets the narrative structure of a story.

3.2 Visual Perspectives Auto-Assembly

3.2.1 Motivation

Dolltalk demonstrated that I could build a technology that affects the perspective taking of a child in a storytelling environment. Beyond verbal and gestural perspective I wondered: what about other forms of perspective, and what specifically does the interface have to contribute to different perspective taking behaviors? These findings encouraged me to research interface design to explore technology for perspective taking: what if these toys could not only talk for the child but also look at the world for the child! In other words: what if the child's eyes and the "eye" of the camera were not one-and-the same?

Figure 3-5: What if a toy could offer its visual perspective?

Figure 3-6: Children could make movies about their toys having everyday lives!

With Dolltalk, children are exclusively playing with two puppets, sometimes morphing them into specific protagonists. However, they are not integrating other toys from their environment into their play. They are turn-taking with a virtual agent, the Alien, displayed on a computer screen. The Alien directs the play, alternating between listening and telling stories. Even though Dolltalk facilitates narrative perspective taking, the interaction is highly driven by the Alien; it is not open ended. So what's the point of having the system that brings children to share perspective taking if the children are not fully playing with their toys, giving the best of their natural play!?

Gesture objects were the sign of a spontaneous interface, but the interaction itself seemed too constraining for open-ended play. The children were not producing visual narratives. The interaction paradigm in traditional video making systems often lacks fluidity and has a restricted view of how to build a narrative sequence; in many cases, the objective is to make a "final" cut of a movie.

To understand how the interface design enables children's storytelling using gesture objects and movie making, I designed three experiments with accompanying prototypes. Despite having conducted testing with various age groups, I will focus on my observations and findings with children ages 8-10.

I present three design iterations of a video editing system for children leading to my ultimate system: Picture This! The design iterations are Textable Movie, Moving Pictures and Terraria. I will present motivation, scenario of interaction, technical challenges, and observations. I use the lessons of each system to guide my next design. With the successes and failures unearthed with each of these three experiments, I gained a new framework for the concerted use of technology to drive perspective taking during open-ended storytelling.

I will finally synthesize my findings on two axes: One being the child's outcome, the complete video works with perspective taking, and the second being the seamlessness of the interface, see figure 3-7.

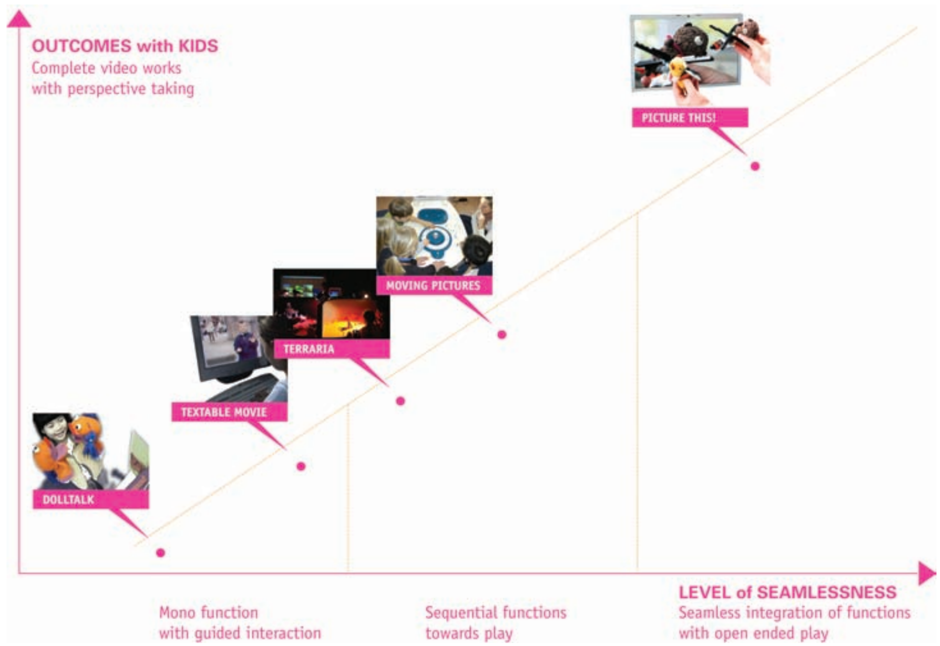


Figure 3-7: The Iterative Design Process.

3.2.2 Textable Movie: Weaving image and "texted" speech into a narrative flow.

Motivation

Textable Movie, my first design prototype, focused on textual annotation, or the typed "voice over" of images as a narrative technique. It was intentionally not constructed as a tangible interface. Rather, it was designed to inform the development of later tangible platforms for movie making. I 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, Singer et al., 2006].

Textable Movie is a fluid interface for weaving image and *texted* voice-over into a narrative flow: a flow based on free associations "on the go" rather than re-assembly or montage after the fact. Textable Movie uses words as triggers of evocative images, and images as triggers for words.

My previous research had led to the idea that the projectionist, viewer, and maker could use text input to sequence the projection [Vaucelle et al., 2003]. However, early testing uncovered a basic limitation: how would the projectionist/ viewer know what words to use? My response in Textable Movie is that players submit and name their own images. My goal with the Textable Movie system is to provide a platform for unexpected discovery of media based on the telling of a personal story.

Scenario of interaction

With Textable Movie, someone can be comfortably installed in a sofa, with her friends, telling stories about her vacations and automatically projecting videos corresponding to what she tells. If she talks about going on an adventure in the woods, the audience will be looking at imagery from her time in the woods. Is it autumn? The image shifts to woods in autumn.

As the user types a story, media segments appear on the screen, generating a movie. Textable Movie evokes the familiar image of characters on TV fantasizing while thought bubbles appear above their heads showing what they are thinking about.

Figure 3-8: Textable Movie: Scenario of Interaction.

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. I 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.

Textable Movie is a graphical user interface that I created for video production [Vaucelle and Davenport, 2004b]. It reduces the technical difficulties of creating a publishable movie by coupling the performative act of telling a story to editing a final movie.

The Textable Movie Technology

Users create a personal mode of interaction with the system by mapping their own keywords to videos and incorporating new video clips and sound samples to their database. Textable Movie relies on text as input, text that is mapped to children's personal databases of videos and outputs the videos made by the children. As the user types a story, media segments appear on the screen, generating a movie. A real-time engine responds to the user's vocal or written keywords by projecting the corresponding movie clips. A set of parameters affects the movie in real time such as zoom, speed, colors, and loop. The same process is used to assemble final movies.

In the framework of computational storytelling, Textable Movie promotes the idea of maker-controlled media and can be contrasted to automatic presentation systems. By improvising movie-stories created from their personal video database and by suddenly being projected into someone else's video database during the same story, users can be surprised as they visualize video elements corresponding to a story that they would not have expected. Users make their own inference about the visual discoveries rather than being passive to an artificial system that would make the inference for them. A movie-editing paradigm in which text leads and image follows provides a natural, fun, and immediate interface to video making.

Observations

Qualitatively observing children videotaping, editing and making a video was necessary to understand how children consider the movie making process, how they use the camera and the type of framing their use. The children, as part of the design team, could directly contribute to my next appication iteration, playing and designing as they go, learning about the challenges of making a movie as well as orchestrating the movie, alternating between being the cameraman, the actor and the audience.

Two groups of ten children between the ages of ten and fourteen participated in our evaluations, however I focused mainly on children age ten year olds to inform my next design iterations [Vaucelle and Davenport, 2004a].

To observe children's mental models of video making, from story-sketching a movie to performing a final movie-story in front of their peers, I observed the children as movie-makers during an informal experimental setting during a one-week workshop, at the Ark, a cultural center in Dublin, Ireland. I relied on two tasks:

- Task 1: Traditional compositional video editing.
- Task 2: Improvisational Textable-Movie editing.

I originally created a methodology for international workshops on creative media making and sharing for the Textable Movie project. The workshop engages teenagers from around the world in digital media making using the Textable Movie tool set. The workshop features a design cycle that begins with concept development and continues onto storyboarding, video production, editing and publication on a public display; as it is realized, participants test and evaluate their video-stories using Textable Movie.

The workshop's global strategy focuses on fostering intercultural visual communication and play. One goal of the international program is to generate a cross-cultural study focused on the creative construction of media by teenagers. Adult mentors, professional animators, and documentary filmmakers, 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, and 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 iMovie$^{\mathrm{TM}}$ software to create a palette of movie segments and associated keywords. Children then used the clips in Textable Movie in a visual storytelling performance in front of their parents and friends. I also asked the children to compose a movie from the same video clips using iMovie$^{\mathrm{TM}}$. Through observations and interviews I analyzed how the children conceived making a movie from planning a video shot, to conceptualizing the editing process and projecting their movie. I also compared how children used Textable Movie and iMovie to compose a movie.

Children prepared their story on written storyboard, then captured their movie sequences with a regular camera, digitized their recorded movie and edited their video story with the iMovie$^{\mathrm{TM}}$ software. They finally used the Textable Movie software to improvise a movie based on their real-time storytelling. Each session was videotaped and transcribed for analysis.

Findings

To collect videos, I 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 followed the content of their storyboard. However, during the traditional editing phase, I lost their attention. When asked to edit their movie to create a final movie in iMovie$^{\mathrm{TM}}$, more than half of the children stated they preferred continuing the capture process.

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.

In sum, the children all captured their movie, but half of them dropped out of digitizing and editing. The digitizing and editing phases are necessary for children to clean their raw data, clarify their original vision (from their storyboard) and select pieces for use with Textable Movie. The result was that the children who dropped out of these phases repetitively captured movie clips. 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.

With Textable Movie, children spontaneously composed by associating videos with keywords. As an example of play: the children designed the database so that if you type in "Tom Cruise", one of their friend appears on-screen.

This is a typical example of the interface being appropriated for play. The children constantly referred to Textable Movie as the "game" and during post interviews they told the researchers that Textable Movie is "more like a game than video making because it's fun!" With Textable Movie, I succeeded in making visual storytelling more playful, engaging, and powerful for young people than in environments with traditional movie making tools.

The children's immediate response towards the system made it comparable to a video game. I created Textable Game, a variation that more directly extends the concepts of Textable Movie to the realm of video games. With Textable Game, teenagers design their action games, exploration games, and mystery games, using their personal video and audio media. They create their own game strategies, rules and scenarios, and become their own video game producers.

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 play. Textable Movie is not intended to replace iMovie$^{\mathrm{TM}}$; 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.

I 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 in 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.

Textable Movie allows children to assemble their movie on the go, it invites them to perform a visual narrative rather than to produce a final piece. Assembly mode, on-the-fly, invites children to revisit their message (children did go back into capturing to get the "right" video clip for their performance).

Their storytelling remained fragmented. Children did not link their sentences to one another. Children used Textable Movie as a "performative tool", to create an effect of surprise: as they typed in a keyword a video appeared on the screen to surprise their audience. Children displayed narrative perspective taking. The content of their stories featured both narrator and first person perspectives. They produced visual compositions with a large digital video camera that were typical of a large camera: frontal views of a scene.

During the course of observations, it became apparent that more fusion between capturing, editing and final production was necessary to allow children to focus their attention on content creation. I started exploring the realm of tangible interfaces.

3.2.3 Moving Pictures: Save it for later!

Moving Pictures investigates a tangible interface to gather, capture, and edit digital data around the city for later retrieval. 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, see figure 3-9.

Figure 3-9: A tangible token, an abstract handle, and yet, I'm holding my movie.

The gains made in Textable Movie were transformed into a more collaborative platform through a tangible interface, see figure 3-10. I pushed the interface towards the fusion of capturing, editing and creating a final movie, functionalities that were not all taken advantage of by users of Textable Movie.

I created a tangible environment, informed by the design of a graphical user interface, to facilitate the process of capturing and editing videos using physical video tokens [Vaucelle et al., 2005a]. Each token encapsulates a story segment in the form of a video clip that the children can replay and recombine at any time.

Figure 3-10: Moving Pictures: a tangible movie sequencing and editing environment.

Motivation

Based on my experience with Textable Movie, and with children as design partners, I 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. I encouraged the children to explore the entire process of making a movie. I originally designed Moving Pictures for users between the ages of ten and fourteen. However, because the interaction relies exclusively on manipulating tokens, children as young as four years old can play with the system and interact with video clips. To accommodate various age groups and individual characteristics of users, I integrated different layers of complexity, from digitizing the media, performing a movie, to storyboarding a complex narrative, similar to the video-making process during the Textable Movie workshop.

The tangible potential, embodied by the direct use of physical video containers for movie creation, presents five opportunities that are critical to this research:

- The Moving Pictures platform privileges improvisation for spontaneous creativity.

- The reinvestigation of the video and sound medium allows a direct and immediate understanding of the effects of combining these elements together.

- Transparent relationship between capturing and editing.

- 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.

- To enhance collaboration at various levels of production: from capturing, experiencing to creating a composed sound and visual story.

Enriching a digital interface with the natural language of physical objects offers an experience that is exciting for users to share with one another. I explore the design of tangible interfaces for supporting inter-personal production of digital media. I synthesize performance and editing to facilitate a flow between improvisation and post-production of a movie.

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.

Scenario of interaction

With Moving Pictures, making movies is a fluid process: one captures videos and edits them using physical objects. It is the ultimate cross between flip-recorders-meets turntables for movies for children.

With Moving Pictures, children can make movies using objects instead of computers: from capturing, editing to VJing. A set of customized cameras allow kids to record their movies. To make a recording they use physical objects, tokens, about the size of a coin, each token representing one movie clip. After recording a series of clips on these tokens, a child moves to an interactive table where by sliding her objects in the table world she can immediately see her movies. Children use a physical rotating storyboard to sequence these tokens and create multi-shot movies. By rotating the storyboard, children play back an edited movie.

Moving Pictures offers three modes of integrated interaction: capture, jockeying and storyboarding, see figure 3-11.

- Video Capture: To capture video, users insert a token into the camera, which records a shot.
- Video Jockeying: 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 collected material is transferred to the computer. Users improvise video compositions using the tokens, and the clips play on the display.
- Storyboarding: 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.

From small clips to longer sequences, children can build up a lengthly and meaningful movie. Sound effects can be inserted at any time. In this version of Moving Pictures, sound effects are not edited, but only applied to movie sequences, at a selected point in time. Several aesthetic decisions were made to evoke the DJ's scratching tables. Sounds can overlap with one another, or be individually scratched. Furthermore, the soundtrack is recorded as it is performed.

The system allows movie composition by enabling users to actively organize the video and sound tokens. Using different tangible affordances, users can move between the three modes of operation: Capturing, VideoJockeying and Storyboarding, see figure 3-11.

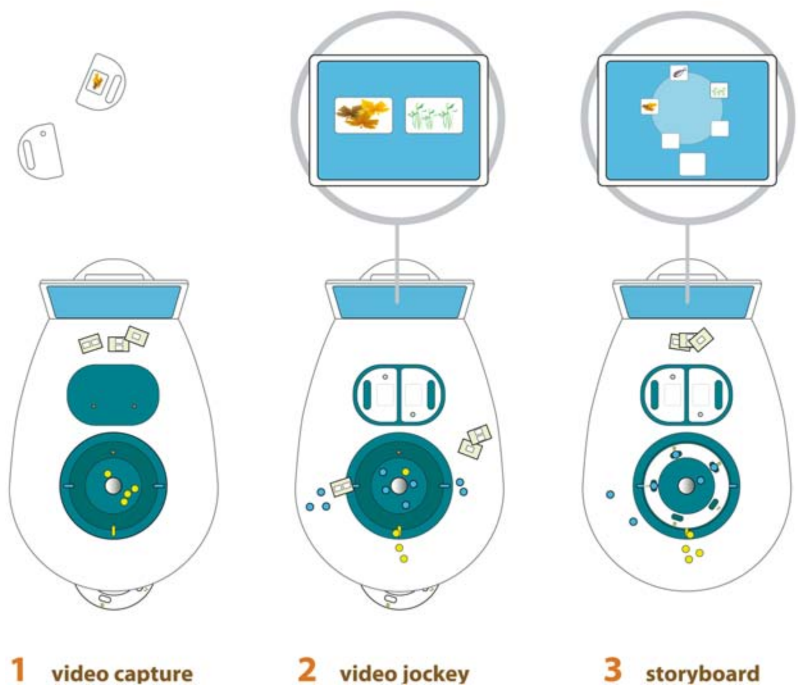


Figure 3-11: Moving Pictures Scenario.

In Shooting mode, see figure 3-12, users insert a token into the camera and then record a shot. They place the camera on the table and the video is transferred wirelessly to the computer. Once removed from the camera, the tokens can be used as a composition element on the table and the resulting video clips can later be combined by the group to achieve a common outcome.

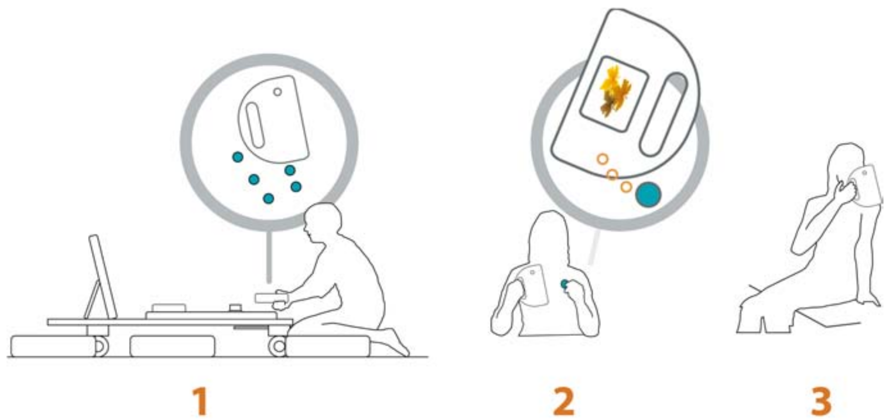


Figure 3-12: Capture with physical tokens.

In VideoJockey mode, users can improvise video compositions using the tokens to play the video clips instantaneously on the screen. As they pass a token over a RFID reader on the table, the computer receives the ID, retrieves the segment associated with it and plays it back on the screen.

In Storyboard mode users can create a structured composition by placing a number of tokens on a Storyboard tool (Storyboard ring) and playing them sequentially, as well as adding sound effects. A graphical user interface guides the users through the steps of composing a final movie.

To switch from VideoJockey to Storyboard mode, the user places the Storyboard ring on the corresponding area of the table. The ring contains slots for the RFID tokens. The Storyboard mode is turned on when the user places the ring on the table which triggers two light sensors that act as an ON/OFF mechanism. In Storyboard mode, users can organize their video segments sequentially by inserting tags into available slots distributed around the ring. By pressing a knob, users can preview and export the final movie.

The Moving Pictures Technology

Moving Pictures is a tabletop 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, see figure 3-13.

Figure 3-13: Moving Pictures: Technology

Recorded media is associated with a digital ID and a physical token. Software written for the PDA wirelessly sends the mapped information between token ID and media to the computer as well as the media files themselves, see figure 3-14. The software written for the computer instantly retrieves the information and plays back the appropriate video or sound segment on a display screen.

More specifically, a token with a digital ID is inserted into a PDA that has a camera built into it, so that the token's ID is permanently associated with the temporal sequence of image and sound as it is recorded by the camera. Once removed from the PDA, the physical token can be used to retrieve the sequence of images from memory, to display this sequence and to place this sequence within a longer media sequence.

Sounds can overlap, or be individually scratched. The soundtrack is recorded as it is performed, see figure 3-15. 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.

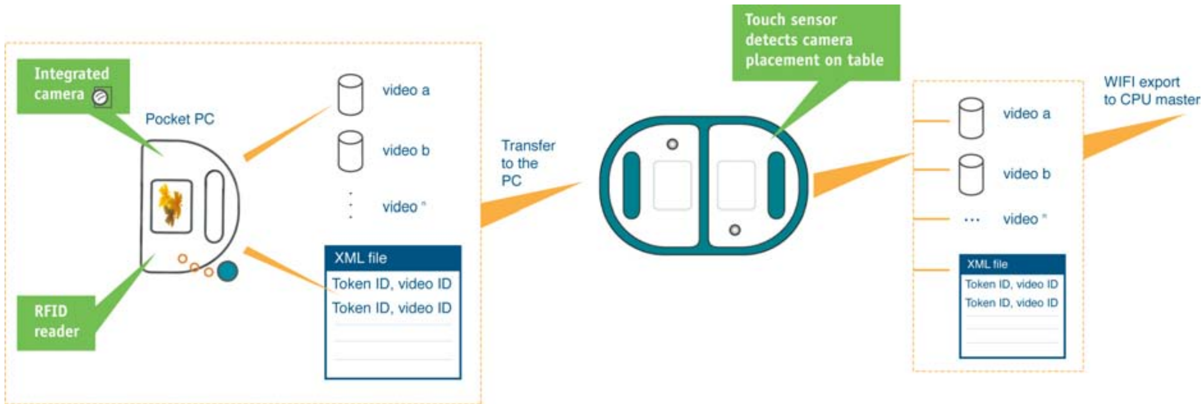


Figure 3-14: Moving Pictures: Token ID to video ID mapping and export.

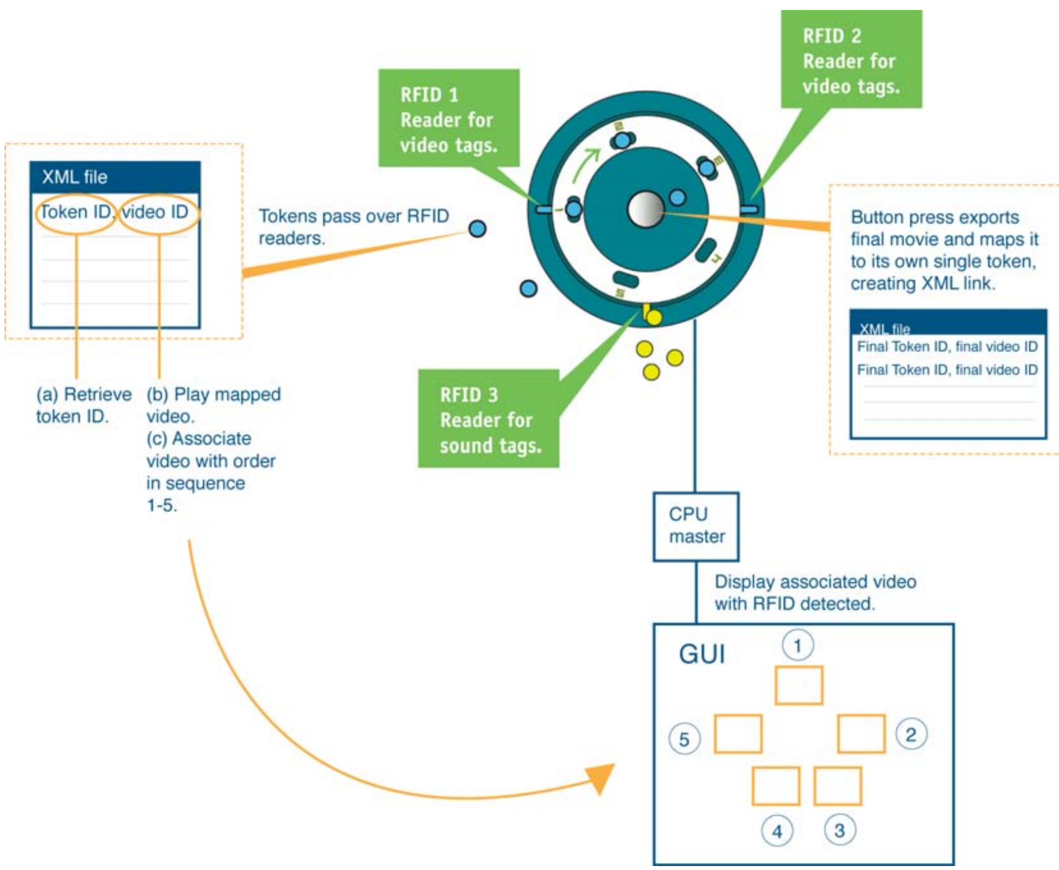


Figure 3-15: Moving Pictures: Token ID retrieval, video playback, video sequencing and final movie exporting.

Observations

The users are children between the ages of ten and twelve. Children have been involved in the project at all stages, as part of the design team and as test users. The final evaluation was made through a cross-cultural workshop involving users from a local school in Umeå, Sweden, and visitors to a children's art and cultural centre in Dublin, Ireland.

Figure 3-16: Children's initial designs are clearly influenced by preconceptions of media editing environments.

The design methodology is based on evaluation results with Textable Movie. As part of the evaluation, I also introduced children to low fidelity design prototypes. The needs and preferences of 10-12 year old children were explored regarding group interaction, attitudes and trends that potentially influence their choice of products. A participatory design approach was applied to implement a functional prototype of Moving Pictures [Douglas and Aki, 1993, Druin et al., 1998]. The first design decisions for a tangible revision of Textable Movie were used as a starting point for the children in participatory design sessions.

Several moderators from Ireland and Sweden organized design sessions with children as co-partners over a period of 8 months, see figure 3-16. Four structured groups of children were involved.

- Group 1 evaluated the concept throughout the entire project.

- Group 2 started participating later on, when a first functional prototype was developed.

- Groups 3 and 4, located in Sweden and Ireland respectively, were involved in a final evaluation. The final evaluation, is based on a cross-cultural workshop with users from a local school in Umeà, Sweden, and participants from a workshop on video making in Dublin, Ireland.

- Children in groups 1, 2 and 3 attended the same school in Sweden. All sessions with group 1 were carried out at the school's after-school club and the children participated voluntarily. Sessions with group 2 were planned within school hours and in agreement with teachers.

During initial sessions we learned about the children's use of video-related hardware and software. This led us to observe the complexity of existing products and notice their effects on children's creativity and group interaction [Vaucelle et al., 2005a]. The Textable Movie system was used as a departure point. Low fidelity prototypes were shown to convey the concept of collaboratively mixing media using tangible tokens on an interactive table.

The moderators progressively introduced the movie making process as well as electronic components that are used to create an interactive movie-making device, see figure 3-17.

Figure 3-17: Testing and experimenting with technical ideas such as RFID tags, Ipaqs and cell phones.

Figure 3-18: Working with children on draft ideas.

Together with the research team, users explored different types of input tools and tested a series of design ideas, see figure 3-19. Children were introduced to a number of cinematic concepts, such as space, time, continuity, point of view and action-reaction sequences. We developed solutions for a spatial, tangible interface that enables a flexible approach to these expressions.

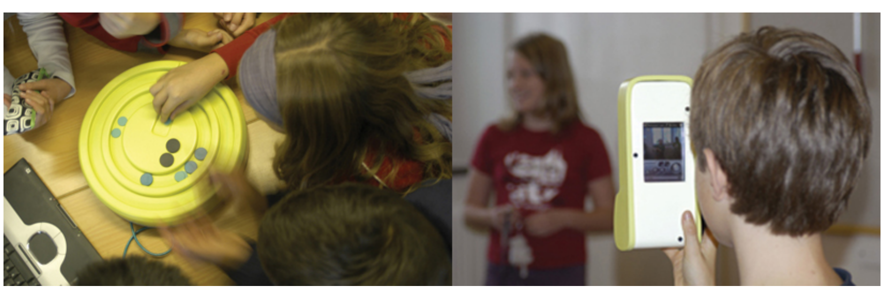


Figure 3-19: First Moving Pictures prototype.

Based on our design sessions with the children as participants, we drew relevant conclusions on a variety of prototypes, including the camera, see figure 3-20, and the final editing, mixing and performing table see figure 3-21.

Figure 3-20: Final camera prototype.

Figure 3-21: Final table prototype.

Besides the structured participant groups, a number of Swedish children from several local schools tested iterations of the prototype during one-day events. Swedish and Irish children from different communities created movie-stories with their own footage.

Findings

With the use of digital cameras, the technical barriers of producing a final still or moving picture are minimized. The possibility to take risks and experiment is encouraged, enabling more expression though the use of visual media. For a majority of the population who do not master the conventions of visual media, some scaffolding, context and constraints may be necessary. Intentionally, in Moving Pictures the user is limited in the length of the captured movie.

The tangible metaphor of a token symbolizing a single shot had to be consistent with common motion pictures language. During my observations, young adults adopted the physical metaphor accordingly. They were careful with the length of the captured clips. It enabled them to practice limited rules in standard video editing without being too conscious of them.

The technical simplification aims at not breaking the creative flow. I also integrated spatial components in the video cameras to take two points of view on the same scene at the same time. This conveys a sense of space such as close-up and large view on a scene. Users establish an exchange using visuals and sound where capturing and editing is made seamless.

The technology itself is not the users' main focus, because we emphasize creative activities generated from organized narratives, visuals and spatial movie content. Each object is designed with an individual digital function. For example, the physical storyboard ring controls a graphical storyboard. This ring contains physical tokens and acts as a rotating device to feed the digital storyboard on a computer screen.

My observations showed that the children were engaged in their use of the Moving Pictures system. The children understood the interaction with the system and were able to improvise movie-stories. Most children participated actively in the use of the tokens, and enjoyed being able to easily retrieve data on the interactive table. The system allowed children to work at different levels of complexity, from simply retrieving data created by others to creating complex final movies with their own footage and added sound effects. This encouraged different kinds of play and varying task distribution within groups of co-creators.

The following outline synthesizes the children's interaction with Moving Pictures:

- General interaction: Half of the children understood it without instruction.
- Confidence in the system: The lack of explanation on what to do with Moving Pictures made the children exchange ideas and explore the system with each other.
- The use of the tokens: The children actively used the tokens for data retrieval.
- The level of complexity of the interaction: Half of the children retrieved data created by others and the other half mixed their own footage with sound effects.
- Our customized cameras versus professional cameras: Children easily remembered how to interact with the cameras. Some children mentioned preferring a smaller camera that they can carry in a pocket.
- Round shape of the table: Children manifested their preference for a round shaped table to interact with as a group as well as to move around. According to the children, a square table would have meant a four user table.
- Table size: Children suggested that the table should be smaller if used in a home. However, the table was too small to accommodate more than eight users at a time.
- Group Size: In group 2 (22 participants, divided in 2 working sub-groups) some children expressed that the group was too large. Not all participants got a chance to interact during the process of previewing and arranging movies and sounds.
- Effectiveness of Group Work: Children recommended the working groups to be smaller, but they also expressed that it was more fun to work in

a large group, even though the work was not very effective. In contrast, some children expressed that working in a group made the work easier and more effective, because participants helped each other to generate ideas for movie making.

- Agreement vs disagreement: Children pointed out that it could be difficult to work with each other if participants disagreed. Many children considered disagreement as being a negative factor in their creative work. They explained how they made efforts to achieve a consensus. They also realized that it was not always possible to keep track of everyone's ideas.

Results from the process of capturing a movie to projecting it.

Over the course of the evaluation, Irish children created a series of movies. Children selected the different themes.

The choices varied from: journalistic interviews that were limited to five shots, explorations in the city using more than ten shots, five individual shots of the children acting in front of their favorite city place, a more sophisticated five-shot criminal story with a beginning, a middle, and an end, and a theater play using ten shots.

The most popular edited movies are the individual shots of the children and the sophisticated criminal story, see figure 3-22.

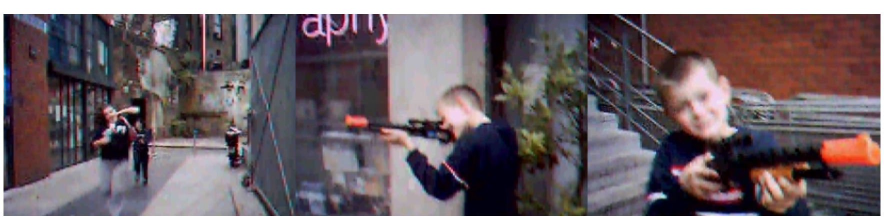


Figure 3-22: An example of children's creation with Moving Pictures.

The students did not want to edit the other country's final movies. Instead, they were excited to watch the variations in the movies and to continue them. This shows potential for cultural exchange through video making [Vaucelle and Ishii, 2007].

I analyzed the two most popular movies to understand how the interfacing of video capture, editing and publication were optimally taken advantage of in Moving Pictures, see table 3.1.

Process	Individual shots	The story
Planning	No storyboard. The children had in mind their favorite place they wanted to be videotaped in front of.	Children spent an hour planning their story, storyboarding and looking for the right spot.
Revisiting the footage	One child revisited the way to jump from one side to the other side of the frame to create continuity within the final movie. His peers were part of his exploration and repeated the same idea.	The shots were constantly revisited, erased and accumulated. Children labeled the token to have the choice of different shots for the same segment of the story.
Visualizing briefly the shots	All of the shots were previsualized and organized.	All of them were visualized and organized.
Editing on the table	No editing of the sequences seemed necessary. Children used their appearance order when they started shooting.	Children enjoyed different outcomes using the same shots. They end up selecting three final movies.
Editing a soundtrack	One specific sound per location. Children did not choose to perform complex sound mixes, but carefully chose their sounds.	Children performed a complex sound mix, overlapping sounds and creating continuity within the soundtrack.
Performance	Children did not try various movies out of the shots, only performed a final movie.	Children kept three favorite movies for videojockeying.

Table 3.1: Analysis of the movies made by the children with Moving Pictures

Lessons learned

The role of the interface: a GUI or a TUI?

Moving Pictures allowed children to collaborate as well as to work individually. Small groups worked better together than large groups, especially when children knew each other from before. In general, the children showed interest in watching films created by other children, but were most engaged when asked to create material themselves. During all tests, the children were very engaged in creative discussions about the content of the scenes they were shooting and the procedures they would use to create a final movie. Even children that seemed to be more withdrawn took part of group discussions and eventually found a role within the team. The system's tangible interface facilitated group work and encouraged participation. The committment we observed from the children during the workshops may relate to our choice of interacting with them as design partners throughout the participatory design process instead of just as passive observation subjects.

We encountered several conceptual limitations related to interacting with the table. The most common issues were related to the Storyboard ring and the relation between the tangible and graphical interfaces. In general, insufficient or delayed system feedback was also a common problem. The number of tokens was too limited to allow an improvisation of a movie in real time. Final movies tended to be the product of a pre-defined storyboard rather than of an improvised association of tokens. With more tokens, and more effects on the video clips, the children would have better taken advantage of the real time approach of the system. Another limitation pertained to the synchronization of sound clips to movies in real time. Not only did children have difficulties in associating a piece of sound to a specific frame in the movie, but they also wanted to record their own sounds, or environmental sounds while capturing video.

Roles assumed by the children while designing and using Moving Pictures

In all sessions, children were free to choose their role in the film-making activity. Driven by their personal interests, they chose to be film script-writers, director, actors, camera-men, or scenographers. Children and moderators discussed and clarified the tasks for each role during the sessions. Most children chose the same role repeatedly. When asked if they thought there should be a leader in the group, children had different opinions. Some appreciated not having a leader and being able to have equal participation. Others said they thought it would be better to have a child taking the leading role.

Several children thought that group members had different skills and this could allow them to learn from each other. During the sessions it was obvious the children influenced and learned from each other. For example, children sometimes helped each other by explaining and showing one another how to perform different tasks. Piaget describes how children influence one another in different ways and how, when it happens, is of great importance for the child [Bukowski et al., 1996]. The children often changed opinions during the sessions, influenced by their friends. This might be a sign of a close collaboration. Prior research shows that when working on film-making, children learned a great deal from each other, which is consistent with what we observed with Moving Pictures [Druin et al., 1998].

As time passed, children became more accustomed to the technologies used in the design sessions and behaved more spontaneously and independently around them. Some children chose to spend a great deal of time arranging video clips and adding corresponding sounds to them, eventually becoming "experts" at this task. Others "specialized" in their acting skills or in camera techniques.

Creativity and learning

Reform movements in education [Rutherford, 1990] have encouraged a shift from didactic instruction to methods resembling real-world problem solving. Visual media are proven to be an excellent medium for such a shift [Smith and Blankinship, 2000]; indeed "visual events provide many opportunities for students to pose questions and reflect on behaviors and process" [Bransford et al., 1990].

In industry, Apple has presented its research on iMovie$^{\mathrm{TM}}$ in the classroom. However by using video editing tools such as iMovie$^{\mathrm{TM}}$ in previous video experiments, I found that young users were particularly unfocused while editing; they would prefer to capture with the camera rather than to edit, analyze and evaluate their footage.

My previous research on Textable Movie, real-time movie making using a personal database of video clips, revealed the usability advantages of tangible interfaces. Textable Movie facilitated the process between capturing and editing for young users, however the interrelationship between capturing and editing needed to be entirely re-designed to convert the information being captured into personal and meaningful content in real time.

Observing the creative process of the children working on digital media with Moving Pictures, I reflected on the four aspects of student *understanding of the arts* proposed by [Ross et al., 1993] and reintroduced by Somers [Somers, 2000]:

- Conventionalisation - an awareness and ability to use the conventions of the art form.
- Appropriation - embracing, for personal use, the available expressive forms.
- Transformation - in which the student searches for knowledge and meaning through the expression of 'feeling impulses'.
- Publication - the placing of the result in the public domain.

Using Moving Pictures, children understood the process of making a movie using a series of traditional shots symbolized by physical tokens. They made a movie adhering to the collaborative storyboard they created. They contributed to a multinational visual database by expressing their visual narratives for use by children living in another country. 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 to produce all of 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 during the entire movie making process.

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. 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.

Technical limitations of Moving Pictures

Moving Pictures is a successful self-contained platform for movie assembly in its ability to engender rules of shot duration and control in structured token environments while providing a limited number of recorded shots at a time. Moving Pictures succeeded in engaging children in the entire movie making process. However Moving Pictures lacked scaffolding from the children's oral storytelling.

Based on my observations with children, I found that Moving Pictures suffers from several limitations related to the problem of how to best digitally support meaningful interactions in the physical space while interfacing video capture, editing and publication in a tangible environment.

First, the scalability of such a system at a networked and international level is flawed. I need to redesign the software technology to centralize the linked data and distribute the nodes of contained data in an organized fashion. To have the technology better assist how an individual moves about the physical space while capturing content, their platform needs to be mediated by a centralized software architecture.

Second, system centralization implies new communication technology to mediate the video platforms and allow them to communicate with one another. The RFID technology in the wireless cameras could be redesigned into a pattern based technology using the video camera of any device.

Lastly, I would like to escape the hardware limitations of commercial video cameras. Users could use any phone, any camera or text based device to exchange material. The system should be designed to generalize despite different input modalities. All of these modifications shift the emphasis of the system from a simple, transparent, video platform, and into an architecture for supporting content generation that reflects the physical environment of the user through multiple information platforms.

Discussion

Moving Pictures succeeds as a self-contained platform for assembling finalized movies, including the editing phase. This is an example of a tangible media success if I compare Moving Pictures to Textable Movie. Children are interacting in the token world using a collaborative tabletop. Moving Pictures interfaces between functions, from shooting to editing to final movie. With iMovie^{TM} and Textable Movie half of the children dropped out of the editing phase. With Moving Pictures, children were engaged in producing all the video stories they created from initial capture to editing their final pieces, however children did not demonstrate the use of narrative perspective taking.

Children preferred two of their creations: the individual shots movie - consisting of sequences of children jumping in different parts of a city - and the sophisticated criminal story in the center of Dublin. I noticed that the most complex features of Moving Pictures were used while editing a sophisticated story.

Reducing the complexity in personal production of digital media and interfacing the process of capturing, editing and performing allows children to experiment effectively with movie sequences, without sacrifying their ability to make complex movies.

Even though the system is not empirically compared with commercialized video editing tools, I conducted prior user testing with iMovie^{TM}. I chose iMovie^{TM} as I found it to be an easy tool to edit movies. Almost all the children were impatiently waiting for their movies to be digitalized in iMovie^{TM}. Some gave up on their original objectives. Other children kept capturing with the professional video camera, not wanting to edit anything.

When editing, children wanted a final movie almost instantly and were confused when they had to erase parts of their movies. For this reason Moving pictures uses raw data captured in small clips. This functionality seems to work better for the children. Having the digital data represented by a physical object helped them materialize their digital data and understand the construction of their movies.

Children were never bored or overwhelmed. They also asked for the ability to incorporate their own sound effects. As it is, Moving Pictures only offers the recording of pre-existing sounds with the video.

With iMovie^{TM}, children never experimented using different endings with their video footages. In our evaluation with Moving Pictures, children captured and revisited their video story elements, edited and experimented with various positions of their shots within the story and this even if the number of shots was only five for the story. With Moving Pictures, children created a soundtrack using sound mixed together.

In early sessions, children asked each other and test leaders for help as soon as they did not understand the instructions. The lack of instructions seemed to be a way to get to know each other better. Research shows that students at various performance levels, working together toward a common goal, can enhance each other's analysis, synthesis, and evaluation of the concepts they are exploring [Gokhale, 1995].

I observed how the children in my studies engage in relationships where they exchange their perspectives and transfer a sense of space through play, collaboration, and storytelling using dynamic media containers and tangible media interfaces. During my observations, we found that even though children in the same age group have similar cognitive development and are effectively "matched" subjects, it is important to design systems that are flexible enough to accommodate individual characteristics.

In terms of the visual content of the movies produced, even though the camera is smaller than a digital video camera, the resultant output movie is typical of a large format camera, in which children did not explore the benefits from a small camera. They did not explore different viewpoints. Moving Pictures succeeded in providing sufficient structure for children to complete entire movies. However it lacked the ability to engage users to incorporate their perspectives or exhibit evidence of perspective taking.

Having the digital data represented by physical objects helped the children understand the construction of their movies. A videomaking 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, I expanded the idea of connecting video editing to children's spontaneous play, focusing on the manipulation of a single controller.

3.2.4 Terraria: Plug-and-Play Movie editing.

Moving Pictures did not offer any identification with the interface. The interface is too generic. I then designed Terraria, a platform that integrates play with character toys such as robots, hypothesizing that children would be driven in storytelling while making a movie because of a character based storytelling setup, see figure 3-23.

Figure 3-23: Terraria: getting closer to robot toys.

Motivation

This last experiment, Terraria, is based on children's play: playing with toys, adding voices, turning toys into characters, and enabling children to capture their play on video. Because of children's familiarity with their everyday toys and games, children could be drawn into video making, building on their cherished objects. Computer game controllers, for example, joysticks, can serve to manipulate personal media. I hypothesized that because of children's familiarity with their everyday toys and games, children could be drawn into video making with a joystick. My next design iteration, Terraria, employs a joystick for video capturing, editing, and performing. The joystick controls camera angles, recording, video and sound effects, playback, and projects of the final movie onto a screen.

Scenario of interaction

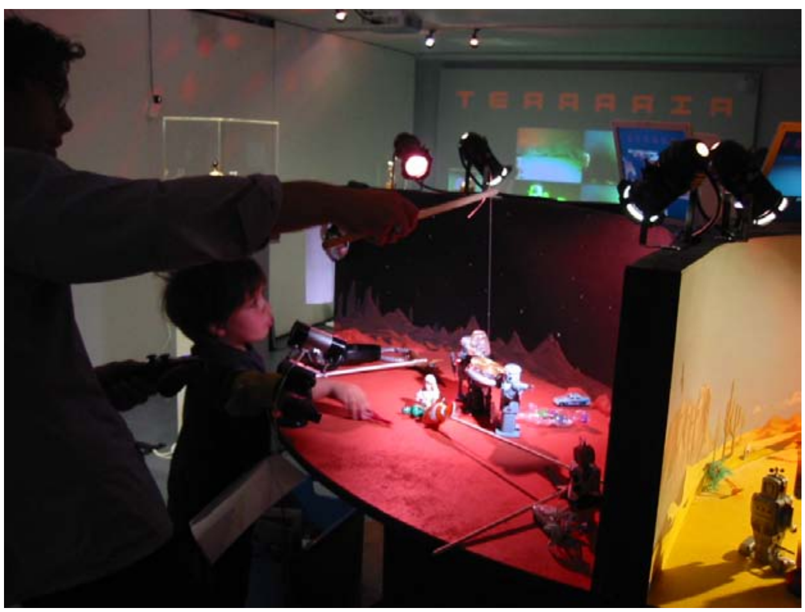


Figure 3-24: Terraria: "Interesting idea Dad, I'll take it under consideration!"

With Terraria children make movies using a joystick and decorate the museum exhibition space with movies they made, see figure 3-24. The interface idea behind Terraria is to see film composition and significant editing and video effects happening via Playstation controllers.

Terraria is setup in a museum space littered with toys. Children come in and pick up robotic toys in stage environments that I prepared with visuals of different settings. Children then use joystick systems to film their toys and movies. These movies get instantly projected within the museum, see figure 3-25.

Figure 3-25: Terraria: interaction

The Terraria technology

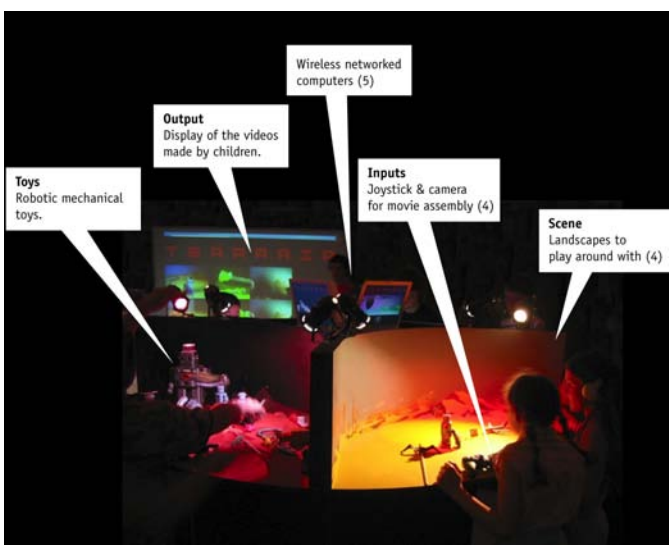


Figure 3-26: Terraria: Technology

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 et al., 2005b]. Terraria consists of four landscapes with robot props, four video cameras, four joysticks, and five wireless networked computers, see figure 3-27.

Terraria employs a joystick for video capturing, editing, and performing of media. I programmed the software so that the joystick controls camera angles, recording, video and sound effects, playback, and projection of the final movie onto a screen, figure 3-28.

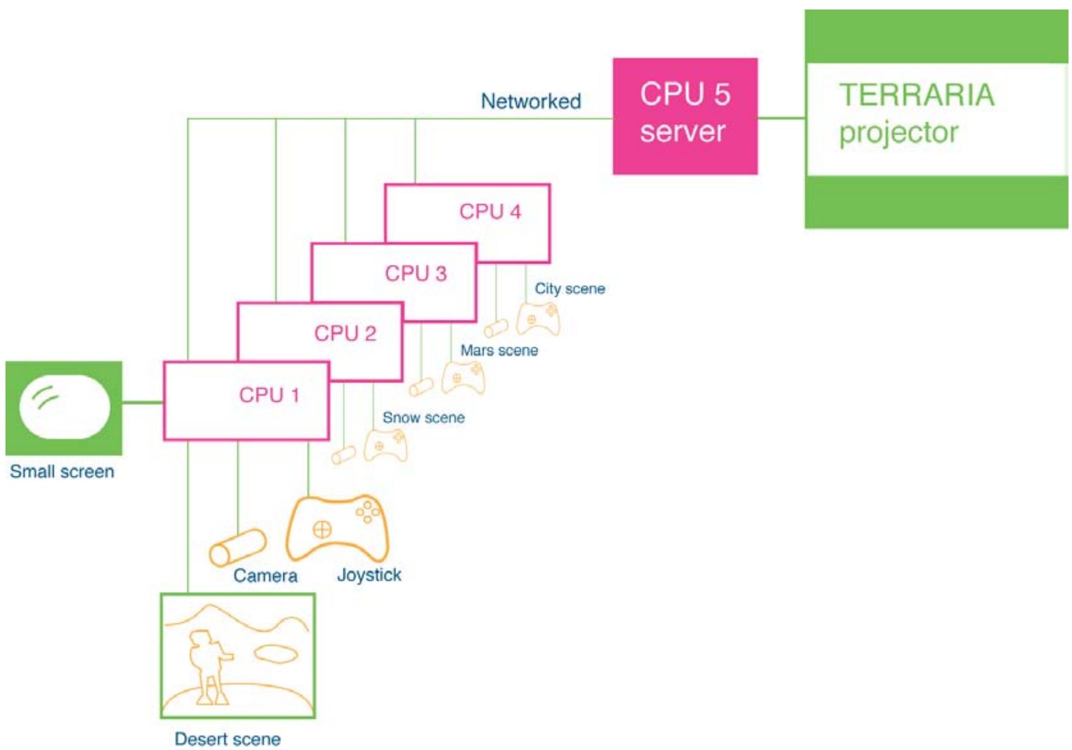


Figure 3-27: Terraria: networked movie creation studios export to the server which maintains the public display of finished movies.

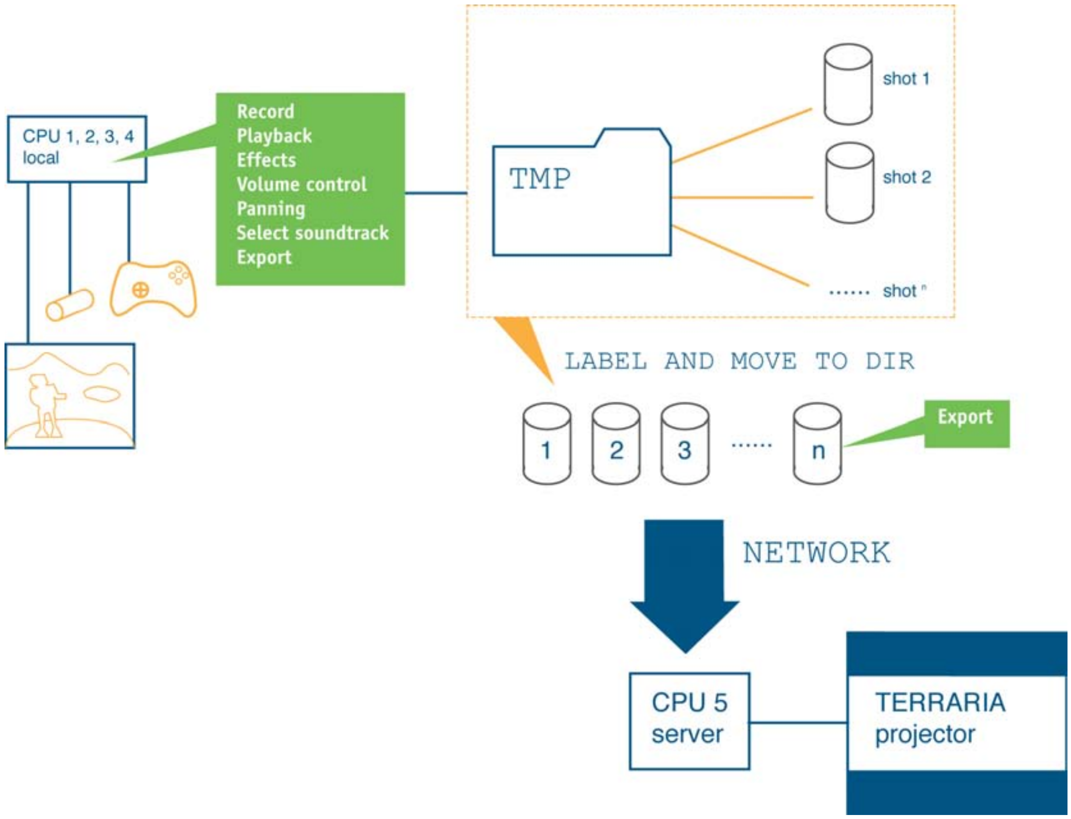


Figure 3-28: Terraria: Recording and editing functionality concatenates shots in temporary, local folder before exporting finished sequences to server.

Observations

Terraria was installed for three months as part of a museum exhibition. Young visitors were invited to make movies and to decorate the exhibition space with their interactive creations. The exhibition space required the tangible video system to be robust to support varied timeframes of use, experimentation, and improvisation of well-structured, sequenced, and live-captured video.

Over a period of three months I observed children from four to fourteen years old playing with Terraria. Five eight to ten year-old children helped refine my design significantly through an iterative process of design and experimentation. The children captured and edited their visual stories, prepared the automated toy robot actors, inserted audio and visual effects, and soundtracks by selecting songs from a database.

Findings

The young users found this integrated interface engaging for performing movies in real time. Children were drawn to give a visual life to their robotic toys and spent an average of one hour each playing with Terraria. Young users captured and edited their visual stories, prepared the automated toy robot actors, inserted audio and visual effects, and soundtracks by selecting songs from a database.

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 my user studies and during the exhibition, users recorded videos, and selected soundtracks to fit with their videos and to unify their composition.

Figure 3-29: An example of children's creation with Terraria.

The interface does not invite the characters to"talk" but rather to be part of the storyline. The camera shots are more traditional: the camera is glued onto the table and the toys just move in front of it. Children consistently bring the robots in front of the frame, and do not create a radically new genre of visuals. What about a robot showing its viewpoint?

Lessons Learned

Children captured videos of their toys, selected visual angles, integrated objects and discovered strategies for animation. However, they did not act out social interactions between toys as in oral storytelling which could have shown markers of perspective taking. The children's focus on the joystick distances them from embedding themselves into the toys. This major failure in Terraria was that children disproportionally focused their attention on the joystick, enjoying grabbing visual components with it, alternating their attention between the joystick and the visual scene. Although I succeeded in motivating children into spending hours making movies, I failed to open the rich space of storytelling within movie making.

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. However, even though I introduced toys as part of the interaction, children did not act out social interactions between toys. Instead they focused on video making.

In Terraria I 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 my introduction to the demands of play in tangible video editing. However, Terraria did not engage children in acting out social interactions with the toys, and children did not exhibit evidence of perspective taking.

CHAPTER FOUR

Gesture Object Auto-Assembly

4.1 The embodiment of the character who has a certain eye!

4.1.1 Design Principles

After Terraria it became clear that if children are intended to bring the rich space of oral storytelling to movie creation, I needed an interface that directly invites the children to do so. Instead of looking at the toy from afar and making a movie, children could make a movie from their toy's views, using their natural gestures with toys to animate the characters in their stories and command the video making assembly.

With character toys, children create interrelationships and plots, a means to expose their social knowledge, knowledge about human beings and social relationships [Shantz, 1975].

I decided to explore video capture from the toy perspective, to create unexplored visual perspectives and to merge storytelling and play to construct movies: fusion of capture and editing of movies to the benefit of perspective taking. Picture This! brings in the child's visual perspective, complementing the storytelling in Dolltalk and producing movies from the child's own toy environment as she plays, see figure 4-2.

Figure 4-1: With Picture This! the child's toy becomes a camera person as opposed to having the child hold a camera directly.

Picture This! puts the eye where it has never been before! A toy with an immediately accessible visual perspective opens a new world to the child. The toy brings her to explore visual and narrative perspectives of character props, expanding the discovery of her environment. The child storyteller enters the world of the movie maker.

With Picture This! the child's toy becomes a camera person as opposed to having the child hold a camera directly, see figure 4-1. Picture This!, goes beyond assembling visual scenes: 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.

4.1.2 Motivation

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.

Picture This!, offers a comprehensive application beyond the scope of assembling visual scenes. I want to motivate children to use their toys to tell a story while assembling a movie. I explore video capture from the toy's perspective, to create unexplored visual perspectives and to merge storytelling and play to construct movies.

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. 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.

Figure 4-2: Picture This! brings in the child's visual perspective producing movies from the child's own toy environment as she plays.

Motivated by the playful improvisational environment of child storytelling with toys, I have developed a new category of video editing tools progressing towards the child's natural expression of play. In Picture This! I combine the activity of play with the video making process. Whereas play emphasizes spontaneity and improvisation, video making necessitates structure and composition. I was inspired by the theater play of Goethe's childhood [Singer and Singer, 1990] investigating what technology could add to the narrative and play experience.

I use technology to offer visual feedback regarding how the scene looks 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 into Lego people, car drivers, and even coffee mugs! The video process, supported by gesture induced editing, invites children to practicing social interrelationships and visual perspective taking.

4.1.3 Scenario of Interaction

The child takes a toy, and puts a camera in the hands of that toy, or somewhere on the toy. Here is a scene, see figure 4-3: the bear has a camera, the helicopter has a camera. As the child moves the toys as part of her play, the gesture analysis system relies on her movements to shift which camera is filming from one camera to the other.

The gestures that I look for are simply to determine which camera should be shooting at one time or another, based on where the action is taking place, who the main character is, etc. I also capture the audio of the child telling her story throughout the play session.

The result is a sequence of shots, taken from multiple cameras, switching amongst each other. The child is directing the switching through gestures, and modifying her storytelling based on who and what is in the visual at a given time.

Figure 4-3: Picture This!: the result is a sequence of shots, taken from multiple cameras, switching amongst each other.

For instance, on the figure 4-3 is the scenario of a child telling a story about two secret agents looking for a treasure in Mount Fiji.

Agent 1: "here I am! Flying above the Fuji mountain, ready to get the treasure! Copy that?"

Agent 2: "Report on what you see inside the Volcano, but be careful 007, we might encounter the evil Sphinx!".

The children can pretend one toy "sees" something the other toy does not know about, reinforcing the toy's independent perspective with what is displayed on the screen.

Children project themselves 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 figure 4-4a, as a child plays with the toy that holds the camera, projecting its video feed on a screen in real time, see figure 4-4b.

Figure 4-4: (a) The toy is the camera person versus (b) what the toy "sees" from "his" video feed.

This visual flow aims to motivate her in composing a movie as she plays and explores her visual story. As two dolls interact, the child alternates between the dolls' 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.

I chose the interaction to function like a performance to avoid breaking the flow of pretend play with character toys. My system incorporates the child's gestures with the cameras and toy's accessories as control functions to assemble the movie.

4.1.4 The technology behind Picture This! Automatic Movie Assembly at the Extension of Natural Play

Figure 4-5: Picture This! analyzes the child's gestures and conducts film assembly.

Because the rationale for my system is to invite children to create and record a movie conversation between toys, I designed two sets of camera bags to be attached to two dolls containing video cameras and embedded electronics.

To engage with Picture This! children rely on their usual gesture interaction with toys while telling a story and playing with character toys. The motions I 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.

I developed a filtering algorithm for gesture recognition through which angles of motions are detected and interpreted. For instance, to playback the video, the two toys need to be shaken horizontally together.

My system consists of the following design features, see table 4.1:

Design features	**Functionalities**
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, I focus on children's interaction with the dolls as they create a movie about a two-character toy conversation.

Table 4.1: Design features and functionalities of Picture This!

The software concatenates the video segments captured, and then plays back the entire movie on display. I have four gesture detection modes: Rehearse, Record, Stop Record and Playback, see figure 4-6.

- 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 captures the scene for preview. The output video feed from the dolls' cameras alternates between the two dolls. The child rehearses the story-video she wishes to create by selecting angles, scenes and speeds.

- Recording mode: I 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 are deleted automatically 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 screen 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 in which mode the dolls are in. 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.

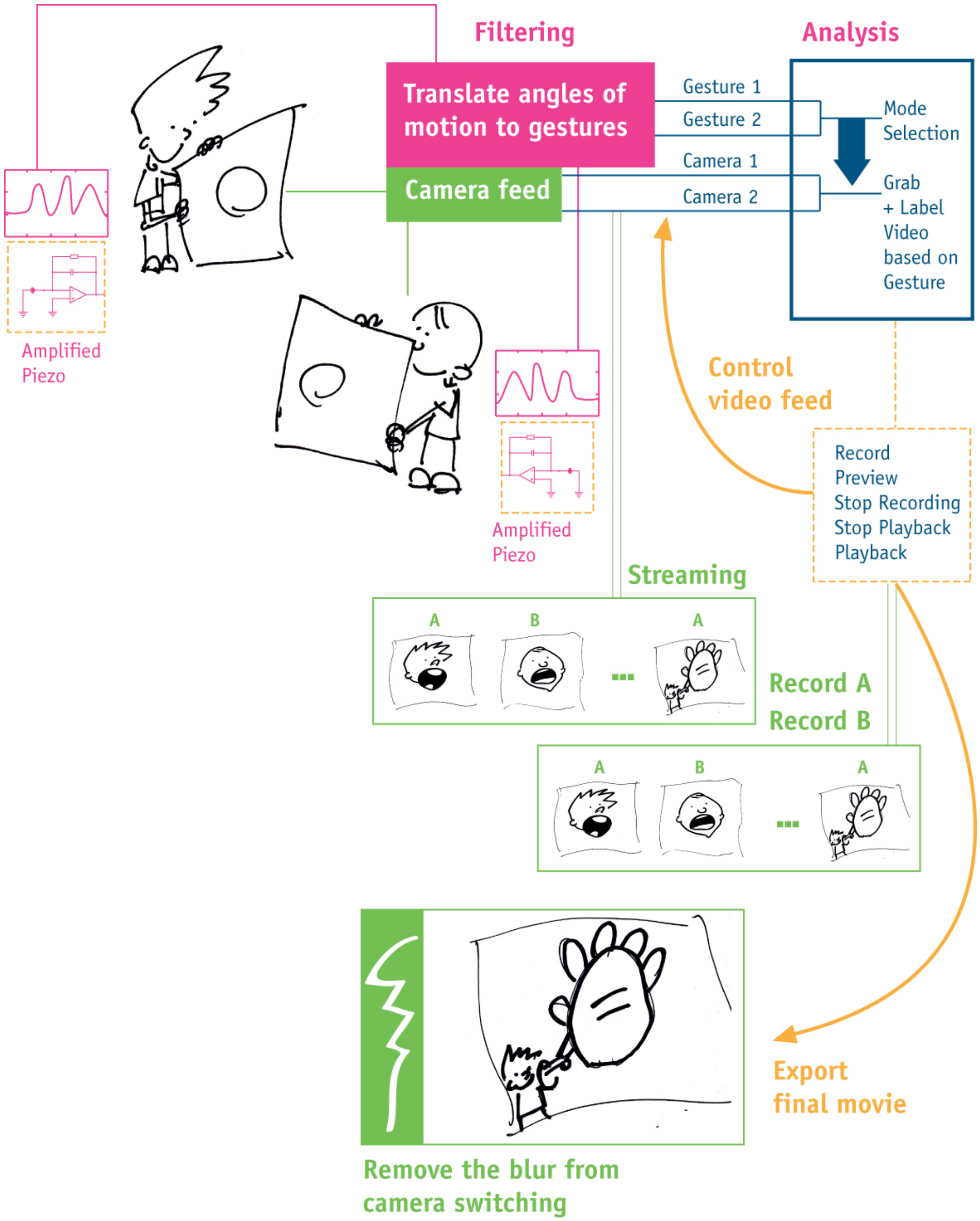


Figure 4-6: Picture This!: recording, preview and playback of multiple -sensor augmented- camera sources are controlled by gesture analysis.

4.1.5 System diagram

As input in Picture This I have the gestures of children playing with the dolls and their voice. A camera, microphone, and Piezo sensor mounted on a printed circuit board communicate to software via a microcontroller. The software I designed conducts automatic assembly based on the gesture analysis of the children playing with toys, see figure 4-7. It outputs different video controls: playback preview, record, and stop record along with the audio which is recorded in parallel. The child can setup her final movie during the play. All of these modes are controlled by the children as they play.

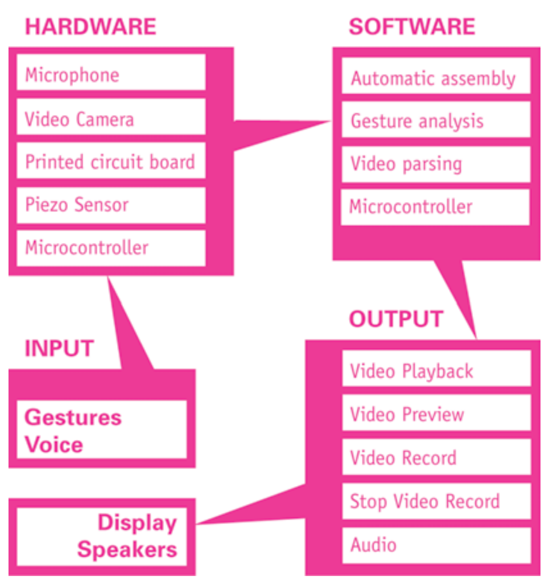


Figure 4-7: Picture This! System Diagram

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 the software, which continuously retrieves the microcontroller's output. I developed a filtering algorithm for gesture recognition that detects and interprets angles of motions [Vaucelle and Ishii, 2008].

The software identifies natural character play movements, such as jumping and shaking, adding video control functions to 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, 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.

4.1.6 Hardware

I implemented a hardware solution to manage the various inputs necessary to characterize gesture-object interactions in the given play environment. The hardware consists of: gesture detection with accelerometers, circuit design to condition the accelerometer signal, and wireless communication with a parent computer to conduct further filtering and program flow control [Horowitz, 1989, Fraden, 2004, Petruzzellis, 1994].

I 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 the initial prototype, I decided to use digital cameras because they interface easily with the software. In the future, I 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.

Picture This! incorporates a piezo sensor as an accelerometer. Piezoelectric materials create electric charge when mechanically stressed. In the piezo sensor, when crystalline structures are stressed do to mechanical strain, a voltage is generated. The change in mechanical stress over time, as a direct correlate of the amount of movement, can be used to define the kinematics of movement such as velocity and acceleration. I am using the MiniSense 100, a piezo vibration sensor loaded by a mass to offer high sensitivity at low frequencies. Impacts containing high frequency components will excite the resonant frequency of the sensor, at 100hz.

I distinguish between vertical and horizontal motions with a single axis accelerometer by detecting small variations in the off-axis motion with the on-axis accelerometer. In doing so, I am able to categorize strong motion in one axis and weak motion in the orthogonal axis.

Signal conditioning, and amplification of the piezo sensor's voltage output, is conducted by a charged amplifier, see circuit diagram 4-8 that I designed for Picture This!.

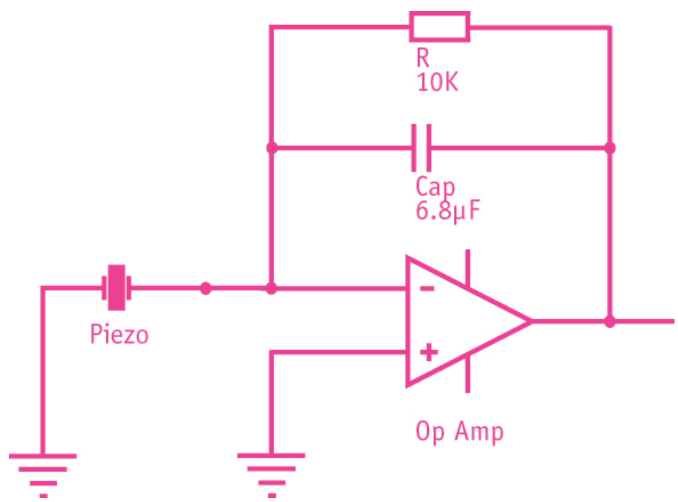


Figure 4-8: Charged amplifier for Piezo sensor

For a cut-off frequency 1/(2RC) at approximately 30hz I use Cf=6.8F and R=10k. The output voltage will depend on feedback capacitance rather than the input capacitance.

The advantage of using a charge amplifier is to minimize charge leakage through the stray capacitance around the sensor. For the future iteration I use surface mounted components in a wireless circuit design which, due to eliminating cable lengths, only requires a voltage amplifier. Additionally, the voltage amplifier has less temperature dependence than the charge amplifier. The output voltage of the voltage amplifier depends on feedback capacitance rather than the input capacitance.

I programmed the microcontroller to continuously interface the accelerometer's data from two circuits with my software. My software retrieves the output data of the microcontroller so that each toy carrying an augmented bag can communicate with my program via specific gestures filtered by my software.

Optical flow method

As an alternative to accelerometers for gesture detection, I could have used the camera's video feed, and analyzed optical flow to detect the vector of a moving toy in a video sequence. This technique would discriminate between vertical and horizontal motion and further simplify the sensor package by eliminating the piezo sensor. Then, the entire Picture This! system could be deployed with common USB web cameras.

However there is a problem with the optical flow method: the child can easily occlude the camera by putting her fingers in front of it, or simply obstruct its line of sight without noticing it during open-ended play. Because the child shakes the doll that will be in the video, the other doll being the cameraman, there are many moments when a child does not have a direct feedback for motion detection occlusion. Due to these concerns, I decided to implement a sensor-based hardware method for prototyping the Picture This! system.

Wireless communication

I connected each toy to a XBee wireless module, which consists of a XBee transmitter and XBee receiver. The XBee receivers are managed by corresponding, battery-powered, funnel I/O boards. The funnel I/O board also receives the output of the piezo sensor, via the charged amplifier circuit. By default the XBee baud rate is 9600, which is synchronized with the software application.

I programmed XBee using of the following steps:

Step1. Download the X-CTU app on windows and run all necessary XBee firmware updates.
Step2. Plug the XBee transmitter into the XBee config tool. Select port, coordinator (computer side), and assign a PANID, ex. 335.
Step3. Do the same with the XBee receiver, on the funnel I/O board.
Step4. Pair XBee transmitter and receiver IDs using the XBee config tool.

Other hardware methods to detect gestures

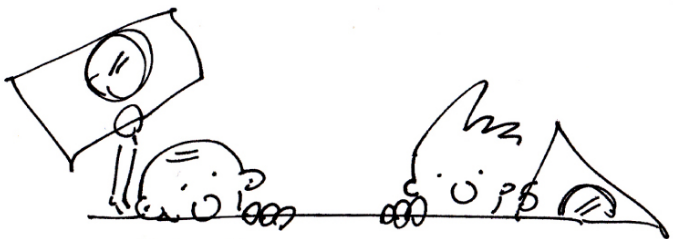
I could have used the following alternative hardware methods to detect gestures:

- Tilt mechanical switch: ON/OFF.
- Capacitive sensing: measuring the distance between an object and its reference point. Change in capacitance related to acceleration.
- Piezo resistive effect: resistance changes with acceleration.
- Hall effect: motion converted to an electrical signal detecting change in magnetic fields.
- Heat transfer: location of heated mass tracked during acceleration by sensing temperature.
- Magnetoresistance: material resistivity changes in presence of magnetic field.

4.1.7 Software

Filtering algorithm

The motions I chose to identify support natural character play movements, such as jumping and shaking, with the addition of video control functions. I developed a filtering algorithm for gesture recognition through which angles of motions are detected and interpreted, see figure 4-9.



Motions / Input: Doll 1 (D1)	Motions / Input: Doll 2 (D2)	Output
Mode (a) Preview		
Vertical motion	No motion	Start live video feed from D2.
No motion	Vertical motion	Start live video feed from D1.
Vertical motion	Vertical motion	Continue current video feed.
Mode (b) Record		
This mode blocks (a) until (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	Continue current recording.
Mode (c) Stop Record		
Horizontal motion	No motion	Stop the current recording and go back to (a) with corresponding live video feed from either D1 or D2.
No motion	Horizontal motion	
Mode (d) Playback		
This mode blocks (a) (b) (c) until the movie played back reaches its end. It then go back to mode (a).		
Horizontal motion	Horizontal motion	Concatenate the video segments captured in (b). Remove the blur from camera switching, and play back the entire movie.

Figure 4-9: Software architecture in Picture This!

4.2 Picture This! goes to school

4.2.1 Interviews with primary school teachers

Gathering data from educators is essential in my user study, because educators will be the ones to interface between Picture This! and the students in a context of exploration of perspective taking and social knowledge. I want to understand both the context in which my 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.

Prior to testing the system with children, I conducted three interview sessions for a total of five-hours long with two primary school teachers. In addition, I continued post interview via written exchange. The teachers I interviewed both teach 5-6 year old children at a K-12 charter school. In the past, the teachers have also taught pre-K, 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 that having something physical for the children to play with is important for their learning. As seen in literature, Froebel combined early 19$^{\text{th}}$ century technology to play and learning, aiming to help children learn about number, shape and colors [Brosterman, 1997]. Montessori education [Montessori, 1912, Montessori, 1917] built upon Froebel's theory and offers manipulative materials to learn through playful explorations [Resnick, 2006].

With Picture This! the children hold the doll, play with it and become immersed in the process. 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 shows 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 [Denham, 1986]. As one of the teachers 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, whereas 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 use it herself, to model it for the students in the Fall of the school calendar. 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.

4.2.2 Observations with Children

In my qualitative evaluation with users, age 4-10, I observe how eight children create a movie with their toys using Picture This! To create a baseline across projects I focus my synthesis on children age 8-10. I also interviewed the parents to gather a context for the data. For instance I wanted to know if the children had prior exposure to television, digital media and computers. These observations will help me design a future empirical study by focusing on a specific age group and the system's functionality.

Figure 4-10: A child playing with Picture This! attached to his Naruto action figure.

Figure 4-11: Tom (10 years old) and Mike (8 years old) playing with Picture This!

I observed eight children aged 4 to 10 using Picture This! to create movies with their toys.

To study children's interaction with Picture This! I installed my system at their home, or if the children requested, they came and interacted with it at the Media Lab research laboratory. The children brought 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. The children used Picture This! with the toys provided for about 20 minutes and played with their own toys for more than 30 min.

Their interactions were videotaped and transcribed for analysis. Over a period of six months, my 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.

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 six seem to forget about the screen, being exclusively immersed in their play with the toys.

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. 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 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.

Children older than eight years mastered the full system, coordinating dolls to control the video, understanding the interaction between preview, recording and playback. After twenty minutes of playing, the gestures with the dolls became parts of the children's vocabulary. Throughout my 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.

At the beginning of each session, the child is learning the system, coordinating between her oral story and who should be focused when. The following is an excerpt of a child's video-story, see figure 4-12.

Children were enthusiastic while playing with the system perhaps to the detriment of attention on the toys themselves! They played along and created rich stories with events, character perspectives, outcomes and endings. After around 20-30 min always the children asked to play with their own toys they brought to the study to be outfitted with the system. As an example, the same child

Figure 4-12: Example of a story with Picture This! when the child plays with the system for the first five minutes

playing with his Naruto action figure would create a much more spontaneous story, closer to natural play in both the motions done with the toys and the stories generated.

When the child knows the system, she brings her own character toys into the world, and begins to explore new visual perspectives. You see her using the cameras to show different types of action. It is also important to note that the story is better.

When the child played with her Naruto action figure, the system made so much more sense! She expressed perspective taking happening not only in the oral stories, but also in the visual stories themselves.

4.2.3 Findings

Picture This! is a new genre of movie making and performing platform, a new platform for media discoveries by exploring visual points of view.

Reconstructing the visual idioms of anime

With Picture This! children record their adventures with their character toys. For instance two action figures might team up for finding the super duper island treasure, see figure 4-13.

Figure 4-13: Robo-team ignite the treasure mission!

With Picture This! children practice alternating viewpoints among the protagonists in their story, see figures 4-14, 4-15 and 4-16.

Figure 4-14: Point of view from the child, "You're not going anywhere."

Figure 4-15: Picture This! Robot: “Must, escape...it can’t end here!”. Visual point of view from the plane.

Figure 4-16: Picture This! Frontal view of the character flying: “This fist was made to protect!”

What is striking with this interface is that child enters the scene visually through their toys holding cameras so that they can literally project themselves onto the toy, see figure 4-15 and 4-16. They create closeups of action shots which resemble their cherished other media, such as anime or video games.

Picture This! draws on participatory design research with over 300 children as part of movie-making-technology studies such as Textable Movie, Moving

Pictures and Terraria. It is the first movie making tool that allows children to transform the way they make a movie by entering the scene, accessing the viewpoints of their characters, and reconstructing the visual idioms of their cherished media such as anime and videogames.

Combining pretend play and role play while changing viewpoints and standpoints

A ten year old girl includes herself in her movie, inserting herself between the two dolls, being a "giant attacked by their magic wands."

The girl is projecting onto the dolls, discovering that she can also be one of the character, and she becomes a character in the story, appearing on the video, turn taking with the dolls and becoming a main character. In doing so, she combines pretend play and role play to change viewpoints and standpoints.

Figure 4-17: Also in between her explorations of visual point of view, she tries to be in the frame and look at the screen straight up and she later becomes a key character in her story.

Figure 4-18: In Picture This! children bring themselves in the movie and the story.

The toy as an interface for video making successfully engaged children towards perspective taking. Children constantly referred to the toys as personified toys, describing the toy's mental states. The interface being a toy with a point of view projected onto a screen. It forces the children to imagine what the toy might be thinking. Because the interface relies on the gestures of the children, the toy is animated as is the speaker: it is put in motion and gains agency.

In a prior study, I observed children playing with toys, attempting to tell a story with the toys, but the stories remained "within the child" [Vaucelle, 2002]. Even in the case of Doll-Talk, with modular puppets to play with, when children were playing with the puppets without technology, they seemed disengaged and bored.

An interface with technology pushes the child to voice her stories, an interface with feedback on her stories pushes her to do better, to make her story richer, and to integrate perspectives. From anthropomorphism to projection, children transfer their viewpoints in the interface. Not only do the toys themselves invite

Figure 4-19: Children's standpoints discoveries with Picture This!.

the children to express perspective taking, but augmented with technology, such as gesture and video, they even "have a mind of their own." For instance in my observations, a child is talking to the doll, blaming her for not being responsive; the child interacts with the toy as if it was a person with intentions.

In the case of Picture This! The toy gains functions as it appears from different viewpoints: from the entire doll to closeup reinforcing particular actions, e.g. its legs or its hand. With Picture This! the child can visually express how the toy character acts, thus reinforcing how the character feels and thinks. The child creates theories *about* the theories that the toy holds. The interface allows a child to express multiple viewpoints visually and within storytelling, because the camera is directly on a toy.

The child practices geometric perspective taking by alternating viewpoints, and psychological perspective taking by reinforcing how the character thinks or feels through the toys' actions.

Capture of storytelling at different levels of interaction

I conducted studies to examine the ability to learn and take advantage of the Picture This! interface environment, by age group. Picture This! allows children to capture storytelling with physical artifacts at different levels of interaction. Functionalities and modes of interaction can now be distinguished by research goals for individual age groups.

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 repeatedly revisit their stories. Seeing their first playback motivated them to change specific dialogs and scenes as and to integrate the visual angles they found amusing, 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 age ten, 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 4-20a and the same toy not being a camera person anymore, see figure 4-20b; 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 4-20: Picture This! a: the camera man toy vs b: 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. I introduced the record mode, but the child ignored it. She was exclusively working on performing a story with visual scenes coming from the dolls. I 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 my 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.

Although I have not performed a controlled study to validate my 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.

Even though an empirical study would quantitatively support my results, I synthesized my observations of children playing with Picture This!, see figure 4-21.

It reads 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 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**

Figure 4-21: Table: interaction functions achieved by age group.

4.2.4 Lessons learned

The perspective taking in Picture This!

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, 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!".

All the children in my 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.

Children navigate from transferring their own lives onto their toys and attributing human characteristics to the toys.

The following is an excerpt of a video story by Jeremy, 10 years old:

D1: "Hi! My name is Fred what's yours?"
D2: "My name is Sile. Nice to meet you Fred."
D1: "Nice to meet you too. What about wanna do something?"
D2: "Sure let's go explore somewhere, like find the Peruvian treasure. Right over there!"
D1: "Over there in the great yellow mountains, but there is a giant blocking theway.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. This entails, for example, determining where objects are located relative to another agent, or whether the agent can see a particular object [Michelon and 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. An example that shows a child is exploring 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 screen). 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 my observations, I 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 social perspective taking, he takes advantage of that opportunity and plays in terms of characters with perspectives.

The interface

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.

Picture This! also 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 movie's 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., 2002]- by imagining the toy's viewpoint, trying it out and correcting it. Research has shown that full body engagement with technologies, a trend called exertion interfaces, engage children's spatial and cognitive abilities [Seitinger, 2009]. In Picture This! children explore their relationship to their body in regards to the video they are making. 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.

Because of this, various objects could be given a camera spot and generate movies based on how they are handled.

Children proposed to have the Picture This! system connected to different kinds of toy accessories. When I 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. 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!

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 me 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 my observations, my 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

	Age group <6	Age group >6
Gesture Object Interaction	Video performance system: video snippets of the child's play, with only one of the two toys carrying a camera at all times.	Full gesture language system: recording and playback modes enable older children to use social perspective taking both visually and through storytelling.
Visual Stories	Exhibited the process of developing spatial–visual coordination while playing with their favorite toys and telling stories.	Tested visual angles and assembled a movie as they played with their toys. Told stories alternating between direct speech and narrator voice, providing spatial and temporal context.
Oral Stories	Told stories using 1st person and narrator voice.	

Figure 4-22: Synthesis

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.

I synthesized my results in two age groups, above 6 and under 6 years old, see figure 4-22.

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, I 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.

In a future version of Picture This!, I envision a system that allows more editing and playback possibilities, such as the visualization and editing of other's video play stories. I would like to integrate a third camera view. This third perspective could give a visual context to the created conversation. 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.

CHAPTER FIVE

Towards a framework: Gesture Objects

5.1 Discussion

Having introduced several studies and their results pertaining to spatial perspective taking in final movie creation, this chapter shifts focus toward the framework, Gesture Objects.

The four iterations of tangible video system described in this thesis capture, perform and edit movies accompanied with children's stories. I started with preliminary workshops, where I found that most children preferred recording to editing their movie. To motivate children to make a movie from beginning to end and to remain focused during video editing I designed strategies for interacting with media content. My strategies strike a compromise between the powerful capabilities of commercial editing software and the goal of engaging children in making finished works.

Initially, I will revisit the salient findings from four design iterations, recast to develop the Gesture Object framework.

5.1.1 Breaking the Sequential Video Making Process

All four iterations, Textable Movie, Terraria, Moving Pictures and Picture This! address significant limitations in commercially available video editing software by motivating children to create final pieces. By successfully completing a movie, children can then reflect on the finished piece. Picture This! achieves the three functionalities of movie making (capturing, editing, performing), but in a different way: it is fully seamless, breaking the sequential video making approach to the benefit of open ended play.

The trajectory to encompass multiple aspects of movie editing while simultaneously engaging children's storytelling begins with Textable Movie. My first design iteration, Textable Movie, avoids the sequential process of commercial video editing software by coupling the task of editing a movie with the performative act of telling a story. In evaluation, Textable Movie suggested a framework for video editing and storytelling, motivated by playful improvisation.

Then, in Moving Pictures I aimed to interface video capture, editing and publication, using tangible elements, e.g. small tokens, to view, revisit, share, and collaborate on video sequences. Moving Pictures helped children improvise and perform movies collaboratively. By using tokens to retrieve video clips, children focus on editing their movies. They follow their original vision from capturing to editing a movie, and produce finished works. The tangible tokens were successful as embodied representations of film segments. Children felt that by manipulating the tokens, they were handling their movies directly. The tokens don't contain any data, they are handles of associated media content. The children made no such distinction. On several occasions, at the conclusion of the workshops, children stole their tokens in an attempt to take their movies with them!

Children are familiar with playing with toys, play-acting character discussions, and enacting toy interrelationships and stories. Creating video with toys and a game controller, I 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. They were able to focus their attention on video composition to an extend not seen in other interfaces. However, one important component was still missing: children did not tell stories; instead, they 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. With the exception of Textable Movie, which does not include an integrated system to capture movies, all of my designs offer the following modes of interaction: capturing, editing, performing/playback, composing a final movie, accompanied to oral stories and sound effects.

I have synthesized and analyzed my observations of children's interactions across my four movie-making systems [Vaucelle and Ishii, 2009].

Given a choice between an interactive visual storytelling system and a commercially available video editing software, fewer than 50% of the children chose to play back their movies with the video editing software. For both Textable Movie and the video editing software, fewer than 50% exported a final movie. All of the other systems engaged more than 80% of the children to create and play back a movie. I 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 tell oral stories when capturing their videos with Moving Pictures. Such oral narratives would have contributed to offering context (spatial and temporal) for their visual stories.

While interacting with Textable Movie and Picture This!, participants told stories both from the first-person and narrator voice. 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, I learned from each iteration: from a computer-screen-keyboard to a gesture-object-based interaction for video and story expression.

My story-building framework integrates the playful improvisational environment of storytelling and tangible technology into video editing systems. There is a 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 my 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."

5.1.2 Functionality versus Transparent Creation Assembly

Moving Pictures is limited in the ability to invite its users to engage in oral storytelling. Several aspects of the interface, including the generic aspect of the tangible tokens for recorded segments, de-emphasize narration. The abstraction of recorded media segments, and their subsequent manipulation in the token world, provides structure for the sequential editing step that has a tendency to alienate would-be child filmakers. It was successful as a platform to engage children to create an entire movie composition. Children remained committed to creating complete works. Typical hurdles to the film creation process, such as editing, did not discourage the kids from completing their film projects. Moreover, Moving Pictures sequentially integrates the critical movie making activities of capturing, editing, and performing.

In comparison, Picture This! is a fully improvisational gesture object interface in which capturing and editing are combined, closer to the play, closer to the character toys, thus encouraging more storytelling and perspective taking than any of my other systems. Picture This! is a compromise between functionality and transparent creation-assembly: do I want many functions or do I want to be openly creative to begin with? One limitation of Picture This! is that by relying on the gesture that sits on top of open-ended play I don't currently have much of an editing functionality. I can preview, start, stop, playback recording, but it is not as functional as Terraria which has video and audio effects, or as Textable Movie that offers a database of media clips.

However, one cannot imagine a professional editing system that scaffolds personal growth or perspective, but one can absolutely imagine a future Picture This! that has more functions. Thus, in my future work, I plan on looking at what sophisticated editing mean in the flow of play.

5.1.3 Level of seamlessness in the interaction towards creative work

Before the four design iterations, culminating in the gesture object framework, I started with Dolltalk. The Dolltalk system establishes the ability to access perspective as part of gesture analysis built into new play environments. However the interaction is built around turn-taking, and as a result, does not invite children to play freely with their everyday toys nor to demonstrate other kinds of perspective taking than narrative perspective taking. Turn-taking is certainly advantageous in certain interaction scenarios, but it is limited in its ability to bind across the capturing, editing, and performing stages of film-making in a cohesive way. I needed something more seamless to offer kids a coherent system to complete movies.

With Textable Movie, I noticed fragmented storytelling, but showing narrative perspective taking. Children did not produce a final movie. Rather, they used the system exclusively as a performance tool.

Terraria was my introduction to the demands of play in tangible video editing but it failed to engage children in combining their visuals with storytelling. In Terraria I 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.

Moving Pictures succeeded in providing sufficient structure for children to complete entire movies. However it lacked the ability to engage users to incorporate their perspectives or exhibit evidence of perspective taking in their token worlds of various media segments.

5.1.4 Prefacing Gesture Object Interfaces

The interface is significantly different in these projects, although they are all tangible. Moving pictures is a very traditional tangible project in the sense that the users handle tokens in a tabletop environment. The tokens represent the movie content, and are too abstract for the children to express perspective taking through the use of tokens . Even though they express ideas, and create sequences, the interface doesn't immediately make sense for the kids in terms of telling a story. It doesn't invite them to tell a story.

The interface is incredibly key, more important than has been previously considered in tangible research. What are our tangible environments inviting us to do! The interface implies what kind of creativity and content user's will readily explore in their systems. Terraria had toys, cameras, a tangible interface that

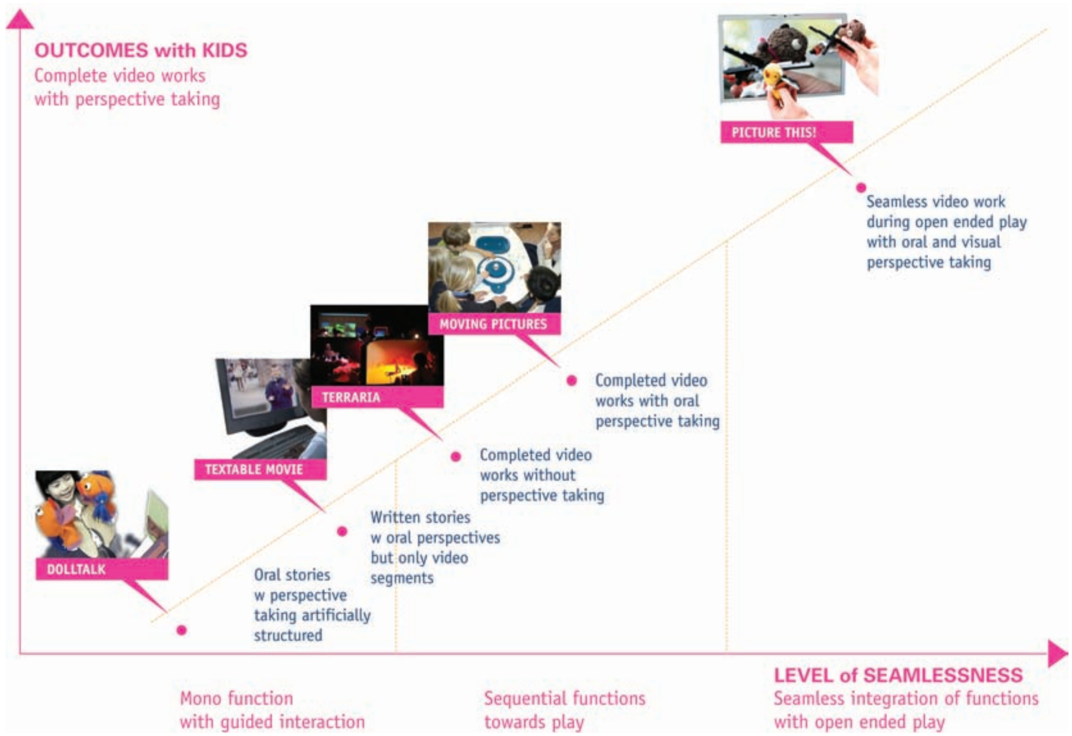


Figure 5-1: Level of seamlessness in the interface

was accessible and encouraging for the kids, but it just was not working in terms of storytelling.

The major interface difference between Terraria and Picture This! is that the camera is directly on the toy, see figure 5-1. The character toy *is* the tangible handle. The interface directly scaffolds storytelling within movie making because it pairs the tangible handle with a *meaningful* object, namely, the character.

The Picture This! interface opens an area where seamless interaction binds between the core functions of movie creation, and further structures both perspective taking and storymaking. Future iterations of Picture This! can be even further modularized and distributed into the open-ended play environment. No longer two character toys with alternating perspectives, but a video and gesture based interface more fluidly integrated into the space of the child.

5.2 Gesture Object Interfaces

The interface is key to perspective taking and the creation of finished creative work in open-ended play environments. As a result of this research, I can generalize to the world of tangible user interfaces and build a new framework as it relates to perspective and information.

5.2.1 Flexibility and Scope of the Gestural Space

The movements that one makes with an 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 is a gestural interaction.

Gestures scale like a language. It has different contexts, different meanings and different results. For instance, with the Nintendo Wii, the controller alternates between being a character on a screen and a tennis racquet. In I/O brush [Ryokai et al., 2004], the identity of the camera is reinvented. As the camera becomes a paintbrush, it fills children with wonder, and they literally want to paint using the entire world-palette available to them.

Gesture Object research differs from 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 [Norman, 2007]. Gesture object interfaces are not burdened by subservience to functionality. They introduce a qualitative component via gesture, in which the language of manipulation is the scope of anthropomorphic extension.

5.2.2 Definitions

A gesture object interface is an object in hand while doing gestures. It is the act of making the gesture that is important for the manipulation of digital data. In a gesture object interface, the interpretation of human gesture is derived from interaction with objects.

As input, we have gesture recognition with object in hand via visual tracking or sensing technologies.

As output, we have gesture control of an object in the digital space that influences the physical world or gesture control of an object in the physical space that influences the digital world.

5.2.3 A Semiotic Square: Positioning the Gesture Objects within HCI

In Human Computer Interaction, I expand on the field of tangible user interfaces to incorporate the gestures performed on tangible objects. This framework organizes tabletop systems, GUIs, and room-scale sensing environments into a coherent space and exposes new opportunities to connect tangibles to social and developmental goals.

I use a semiotic square [Greimas and Courtès, 1986] as a framework to organize my four video systems in terms of tangible interaction, see figure 5-2. The terms in the square identify my research areas and their relationship to the human body [Vaucelle and Ishii, 2008].

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. Gspeak by Oblong industries is an example of an interface in this domain. There is no object in hand, but the gesture recognition controls digital information.

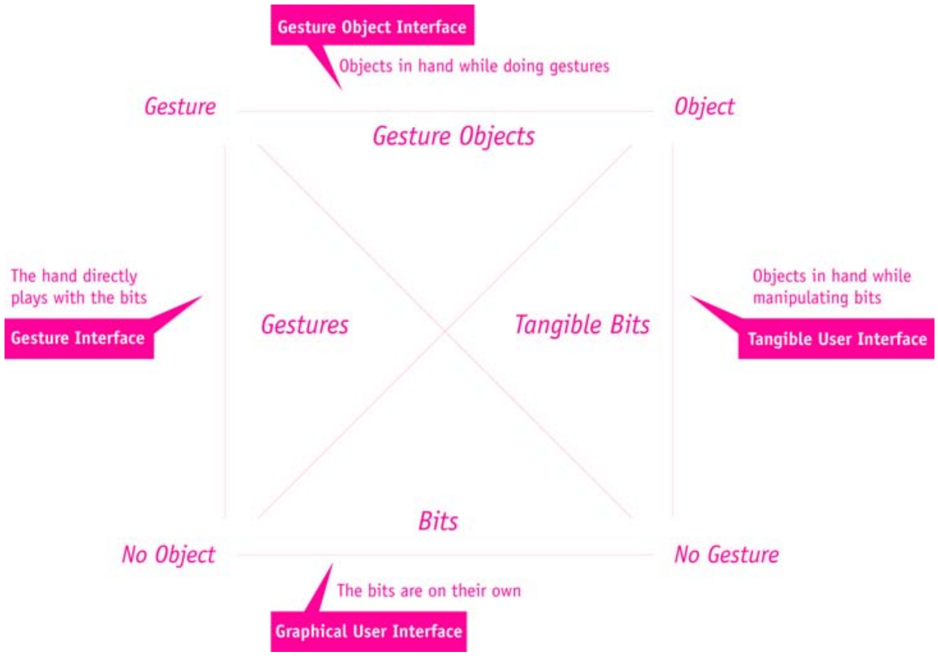


Figure 5-2: A semiotic square for HCI

The right side of the square (Manipulation + No gesture), which represents the field of Tangible User Interfaces as exemplified by Tangible Bits [Ullmer and Ishii, 2000] also focuses on the hand. The hand plays with objects that represent bits. Interfaces such as Moving Pictures, a tangible representation of media stories for capture, editing, and performing, and Terraria, a joystick that directs the composition of video stories, occupy this domain. Moving Pictures, the most canonical example of a tabletop tangible system, uses physical tokens mapped to digital data with no gesture recognition.

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. Picture This! belongs to the Gesture Object Interface category because the system recognizes gestures while the user is holding and anthropomorphizing the object, giving life to the object through her gestures.

Finally, the bottom of the square (No manipulation + No gesture) represents graphical user interfaces. Textable Movie belongs to the Graphical User Interface category because there is neither representative object interaction or

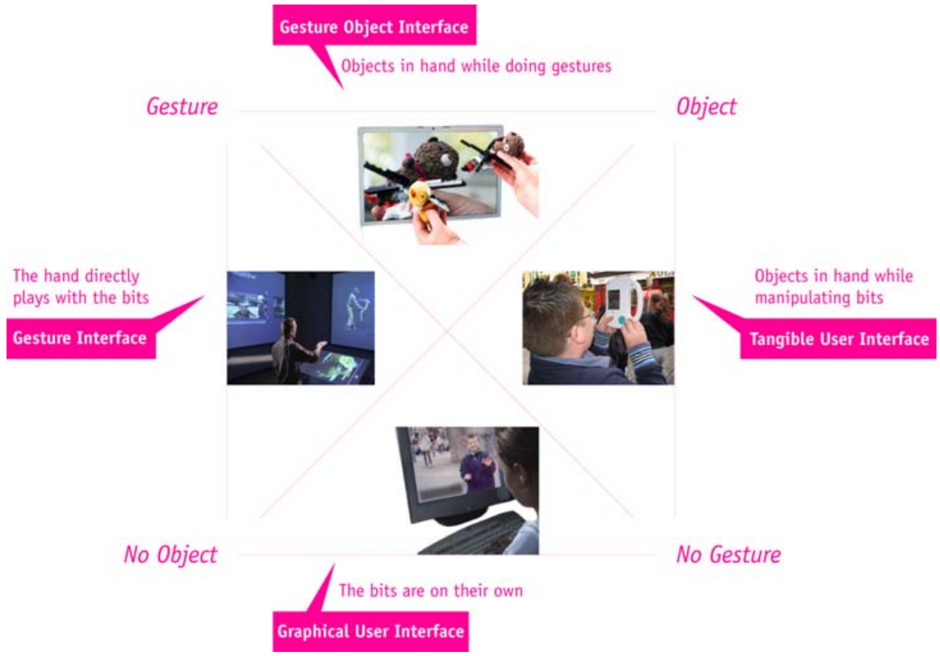


Figure 5-3: The iteration design framework

gesture recognition. Most professional editing systems such as iMovie cohabit this category, see figure 5-3.

5.2.4 Transferring intentions through viewpoints, embedding gestures.

I've demonstrated a new category of gesture tracking systems that privileges interaction with the object. The physical object or the metaphor of a physical object has a meaning, an interaction that we intuitively understand, expect to see, and can produce ourselves.

An object, a glass of water, a pen, a book, any object is a vessel to transfer one's intentions only after technology gives it its own frame of reference. The cameras in Picture This! demarcate a frame of reference. When you transfer intentions to the object, such as "I am looking through this object", the gesture language follows. Since you look through the object, you embed appropriate gestures not only to find visual angles, but also to activate the object, to bring it to life.

These two steps, transferring intentions through viewpoints, and embedding gestures, are what allow an algorithm to ride on top of the cameras in Picture This!. I don't know where users will use the cameras, but their gestures, used as input, will follow as they bring these objects to life by looking through them. The objects themselves are limited without the camera system and supporting technology. Objects do not have a sufficiently rich gesture language without the camera driven frame of reference.

5.2.5 When a gesture gives life to the object

A gesture object is an object in hand while doing gesture. Furthermore, gesture is a mechanism that can reinforce or even create the anthropomorphism of an object, it can give the object life. If I take a pen and I say "hello" by moving it, it is not a gesture object. But, if I move it like a person talking, it becomes personified. In this case, it is the act of moving the object anthropomorphically that makes it alive.

Dolltalk was the introduction to the idea of gesture objects. The gesture with the puppets was the enabling interface. Gesture objects were the sign of a spontaneous interface, but within Dolltalk the interaction itself was too constraining for open-ended play.

Then, Picture This! added flexibility in the pretend play of the children with props. It also added the video dimension, opening a world of multiple projections, asking "which character will be the author?" or "whose eye will tell the story?" Movie making is seamless through the gesture interactions with the props, inviting for spatial geometric perspective taking.

This work brought a new dimension to tangible interfaces: the projection of self in the object being held while the gesture gives life to the object, see figure 5-4.

In this research, I am playing with the boundaries between self and externalization where the interface welcomes perspective taking within movie making, story building and performance.

Figure 5-4: Picture This!: a new dimension to tangible interfaces: the projection of self in the object being held while the gesture gives life to the object.

CHAPTER SIX

Conclusion and Future Directions

6.1 Gesture objects to take, change and calibrate perspectives

Cross-modal interface elements contribute to various perspective taking behaviors. I have presented new technologies to conduct automatic film assembly. My systems interface video capture, video editing and video publication in a self-contained platform.

At each step, I presented the studies to establish principles, which are then used to construct the final project, the centerpiece of my third phase of research, Picture This!

Picture This! is a fluid interface, with seamless integration of gesture, object, audio and video interaction in open-ended play. It allows rapid alternation between novel points of view with which children explore multiple perspectives both in the content of their movies and their supporting narration. Finally, I developed a framework called Gesture Object Interfaces, synthesizing the research as it relates to the field of tangible user interfaces.

I envision a new category of tangible interfaces, gesture objects, which encourage anthropomorphic projection along with geometrical and psychological perspective taking. By projecting the self in the object being held, the gesture personifies the object, giving life to it.

With gesture objects, I project myself onto the object, I take perspectives through the object and I change and calibrate perspective taking throughout the process.

6.2 Picture This: the camera enters the scene ...

6.2.1 Toy play with or without a visual perspective

With Picture This! a child can access her doll's perspectives while creating intricate plots, complex scenes, or showing the secret lives as her dolls just *hanging out*.

She can create a scene where two friends meet and have a discussion complaining about their latest school assignment (or something more along the lines of gossip!)

Half of the joy lies in creating the impromptu plan for the scene, considering the scene, the characters, what they wear, where they are going to be and what they're going to say.

Typically, when children play with toys without visual or audio feedback, they put the toys in a static environment. Perhaps they will incorporate some elements of the environment in their stories, see figure 6-1.

However, Picture This! offers discovery of the environment. The entire environment becomes available through the viewpoint of the character toy. The toy can demonstrate their environment as they own it, as they see it and experience it. Children are driven to access the environment of their toys because it is put in value by the eye of the doll.

Figure 6-1: Doll1: hey sweety wanna stop by the honey tree today? Doll2: yeah sure, but check my cool new bike I got the other day!

Figure 6-2: The toy demonstrates its environment. The child embraces the space not only in her storytelling but through the viewpoint of the toy.

A toy with a camera shows its environment to the child as she creates her stories. The video feedback requires special attention of the children, as they are transported into the viewpoint of the toy. The child directly projects into the toy within the space.

The video feedback reflects to the child what is there. "Is there a boat?! O what is that?" See figure 6-2. Future Picture This! design will expressly focus on the mechanism where the visual feedback of a character toy's viewpoint better transports children into the world of their toys, and increases their meaningful incorporation of the environment into their stories.

Thus, in a future version of Picture This!, children will play outside with their character toys and I will observe how children integrate the scene into their play space both with and without technological feedback.

6.2.2 Reality versus lack of reality

Figure 6-3: The doll is taming the tiger.

Piaget explained that pretend play is an opportunity for a child to secure, via fantasy, what is not available in reality. Toys can participate in the fantasy world, but feedback to the child what that world truly is, from the toy's perspective!

By securing the viewpoint of the character toy, the child is closer to the fantasy. Play becomes a more tangible dream! A little girl can suddenly tame a tiger, see figure 6-3 or a boy can ride the tiger, see figure 6-4, because accessing the toy's viewpoint brings the child closer to the toy's experience!

Figure 6-4: Robo riding the tiger.

6.2.3 Playing with light

Visual feedback also creates new value in various environmental features, such as light.

Playing with light on her character toys, the child can create atmosphere, further designing her scene towards ambiance, see figure 6-6, and to reinforce certain contexts that are lost without the visual component of the camera system.

Similar to the use of the environment to reinforce the internal states of characters, light can become a surrogate for character feelings: angry, moody, sad, joyous, or romantic, see pictures 6-5.

Figure 6-5: By adding red light to the movie making, the child can present a character walking at night.

Figure 6-6: Hey wanna dance?

6.2.4 Scale

Playing with scale in Picture This! a child can pretend the doll is in her environment see figure 6-7, or that the toy is a tiny creature lost in the woods! See figure 6-8.

Figure 6-7: The elfy-doll, at home on her cherished tree!

Figure 6-8: Tiny creature lost in the woods.

6.2.5 Closeup

In Picture This!, the child can zoom in on a character to be exactly on her side, or on her shoulder. The child can capture more subtle expressions than when she takes a God's-eye-view in an environment. Rather than simply thinking, the toy can be *snarky*.

This functionality allows a child to express the psychological perspectives on the toys by dramatizing the close-up. Visual feedback is a perspective learning tool. It gives direct information on how the dolls appear to feel. The child can focus to heighten the impact of a single character's expression in a moment during play.

Furthermore, children can correct the close-up and resultant expression, while seeing the effect of the correction immediately! See pictures 6-9 and 6-10.

Figure 6-9: Doll: "I miss him so much, I wonder when he will be back!"

Figure 6-10: Doll: “hihihi he is really so stupid, trying to get a girl with a flower!”

6.2.6 Mise-en-scene complex perspectives

A visual feedback allows a child to reiterate complex perspectives and mise-en-scene, from picture 6-11 to picture 6-12.

Figure 6-11: A romantic scene.

Figure 6-12: A camera on a toy trully allows a child to achieve complex perspectives: adding a slight angle to the toy watching the scene.

6.2.7 Humor: high level perspectives

Picture This! allows high level suspense to occur as part of subtle themes, for instance, at the service of humor.

By navigating between very close closeup from a shoulder-toy perspective to a wider shot, from picture 6-13 to picture 6-14.

Figure 6-13: Robot: Did I lose her forever?

Figure 6-14: Scene: Unrequited love.

6.3 From anime to action: Viewpoints and transitions enable access of cherished idioms from dominant children's story genres during play.

Children incorporate toys from diverse media genres in their play environments. Whether coming from film, online, video games, cartoons, or the sub-genres within these parent media categories, such as anime, many toys are derived from the predominantly visual environments where children first discover them.

With all of the current tools available to them, including traditional film environments, a cherished Naruto™action figure can only be a facsimile of its animated self. The visual language of anime, in the case of Naruto, is not directly accessible in the play environment.

For all of the design research and technology introduced in this thesis, the majority of children's films are, visually, very traditional. The camera is stationary. Characters enter and depart. The stage is set, and remains static while the scene unfolds in front of the camera.

Picture This! innovates by driving fundamentally new kinds of films. Picture This! allows access to the visual language of cherished genres, see figure 6-15.

Children can incorporate, and reflect on the visual language that is codified by the environments where their characters originate. For the first time, children using Picture This! began to create the complex visual angles, and multiple perspectives, from action shots. Not only were the films different by the incorporation of perspective, but there was ample evidence to support that kids where creating the visual idioms that they associated with the character's parent media.

The gesture layer provides access to the opportunity to explore cherished visual idioms with the associated character toys.

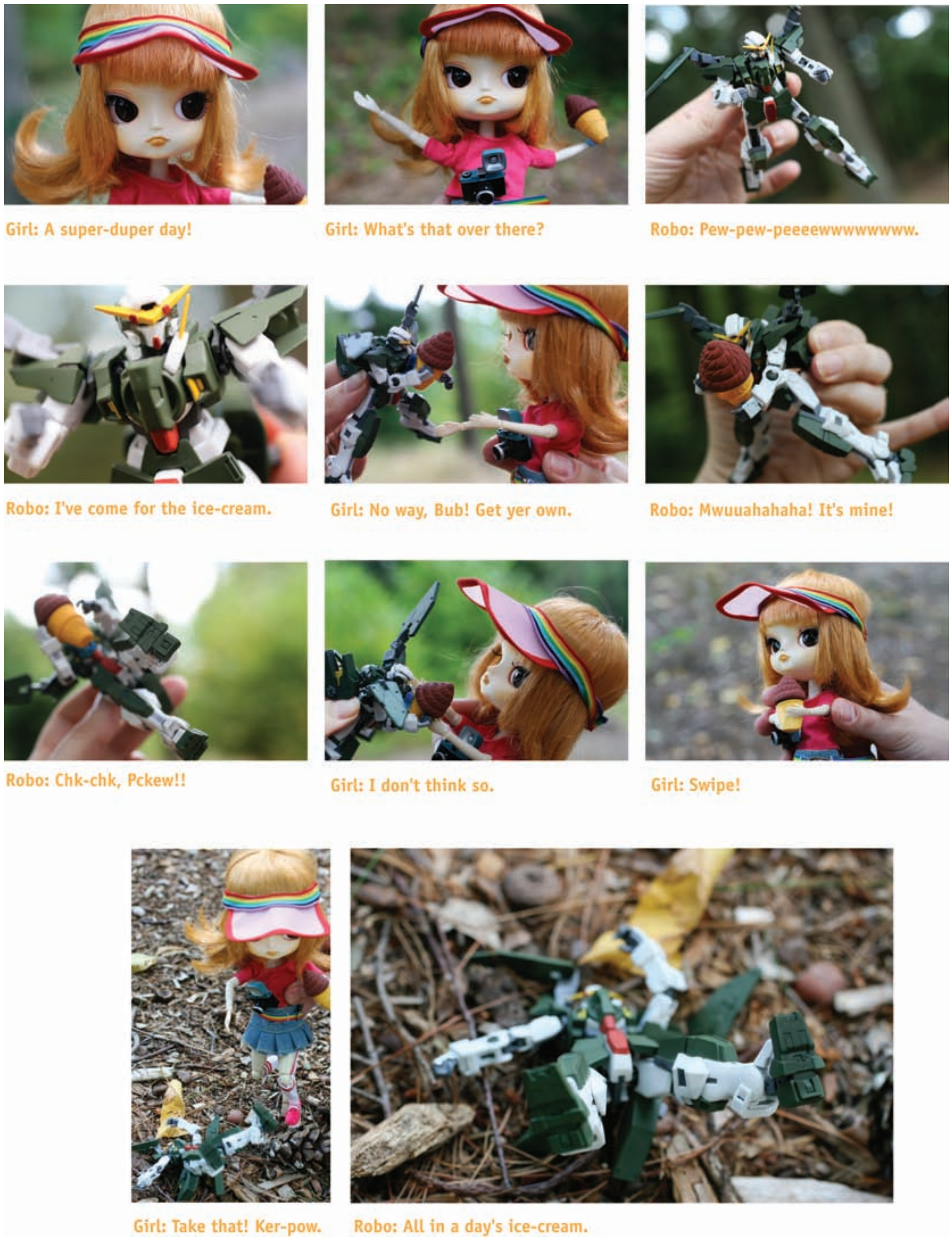


Figure 6-15: Picture This! The results are entirely new genres of child-created films, where children finally capture the cherished visual idioms of action and drama.

6.3.1 Transitions between viewpoints

When a child uses Picture This!, she discovers the character toy's viewpoint through experimentation. She moves the character, with its camera in hand, through an environment, focusing on the video feed, and asking, what is the character observing? The character's viewpoint is directly substituted for her own. The viewpoint is manipulated via the standpoint, where the child's hand and subsequent play-gestures take place. Viewpoint and standpoint relationships are critical to the observed differences I research in the Picture This! system, including the differences in children's story content.

Having established a viewpoint, the gesture analysis component of Picture This! allows the child to transition from one character's viewpoint to another. The act of transitioning, as well as the types of viewpoints that a child transitions from, and to, further enables access to new visual languages as part of Picture This! In Picture This!, privileging the character toy's viewpoint, and transitioning between viewpoints gives the child the tools to access the visual languages of cherished idioms, such as manga, and the style of action that is replete in the dominant media of the child's world.

Comic book artist and theorist Scott McCloud defines a spectrum of transitions that are possible between frames of sequential art works, such as manga [Mccloud, 1994].

Moment-to-moment Transitions mark the passage of very short increments in time.

Action-to-action Transitions demarcate actions taking place sequentially in time, one after another, with a single subject at the center of the viewpoint.

Subject-to-subject While remaining within a scene or an idea, transitions reinforce aspects of the scene or sequential events taking place succinctly, in time, in the scene.

Scene-to-scene Viewpoints transition between vastly different places in space and time.

Aspect-to-aspect The eye wanders between viewpoints that may only be distally related, to reinforce an concept or mood.

Non-sequiter No real, discernible, objective relationship from one viewpoint to the next.

McCloud goes on to chart the distribution of different transitions for various styles of comics, and concludes that diverse creators, whether the acclaimed Stan Lee-Jack Kirby duo from American comics [Lee and Kirby, 2005], Belgian artist, Herge, classic, Will Eisner [Eisner, 2008], or memoir artist, Art Spiegelman [Spiegelman, 1986], adhere to a consistent distribution of transitions. Roughly two-thirds of the transitions are from action-to-action, followed by a varying mixture of subject-to-subject and scene-to-scene transitions to account for the final one-third. The three transition types are similar for both American and Western comics, irrespective of whether the comics are action-adventure, super-hero oriented, documentary, or memoir. Experimental comics break the distribution entirely. But, what is more pertinent to Picture This! is McCloud's observation that Japanese comics, manga, and by association, anime, consist of an entirely different visual language of viewpoint and transition.

Osamu Tezuka [Mccarthy and Tezuka, 2009, Tezuka, 2002], largely credited with creating the manga style that persists today, utilizes action-to-action transitions less than fifty percent, followed by thirty percent of subject-to-subject transitions, and then the rest is given to moment-to-moment, scene-to-scene and aspect-to-aspect transitions. The results are remarkable. McCloud offers that the stories themselves are affected by a sense of timelessness. The viewer is projected into moods, or a sense of place, as part of "labyrinthine" stories. The drastically different visual style persists for many Japanese manga creators.

Since McCloud's "Understanding Comics" was published, during the early nineties, a new generation of children has grown up immersed in the alternative visual languages offered by Japanese media. Many of the most popular cartoons are direct exports of cartoons originally intended for Japanese audiences, and adapted directly from the pages of manga before being exported to western audiences. To say that American media has been influenced would be an understatement. American studios produce new cartoons and stories that incorporate the visual language of manga predecessors.

The character toys that children bring into their play environments originate in the media where archetypal visual languages are predominant. Despite the exposure that today's children have to drastically different visual languages, they have no way to access those languages in their open-ended play environments. The gap between their experience of characters in the originating medium, and their lack of accessible viewpoint in open-ended play, is a new opportunity.

Fundamentally, McCloud's transitions are changes in viewpoint. Picture This! gives children access to their character's viewpoints, and the ability to construct transitions that adopt the idioms of their cherished media genres. Furthermore, the various transitions available to the Picture This! user, are at the service of different kinds of storytelling. In one hypothetical scenario, the moment-to-moment transition defines a certain pacing of a story, which is different both in nature and implementation than the way a child may choose to transition between aspects of a character's scene, all before an action-to-action sequence erupts, involving multiple characters and the objects of the play environment.

In conclusion, dominant children's story genres have visual languages encoded in character viewpoints and transitions that are inaccessible to open-ended play. Picture This! creates value in the moment of transition and viewpoint selection, as a child recreates effects from genres such as anime. The world of the character toy's perspective can be interpreted as the effect of viewpoint and transitions on the various aspects of storytelling, from the conception of time, to the role of the environment, the importance of objects in the environment, or the expression of internal character states.

6.4 Contributions

The gesture object interfaces framework expands on tangible user interaction and gesture recognition, contributing to the field of human computer interaction (HCI). I explore tangibles beyond token-based representations of digital data. By infusing tangible representations of digital data with a gesture layer while the tangible container is "in-hand", the tangible token gains a life of its own. It can be anthropromorphized, as in the play environments that are central to this thesis, maintaining its role as a handle into the digital realm while gaining additional value as an object with a life of its own. Children access this layer of meaning to tell fundamentally different kinds of stories, incorporating the multiple perspectives of their character toys while they simultaneously take advantage of a tangible environment to construct finished movies from multiple camera sources.

As a technical contribution, I create a suite of seamless automatic assembly tools accessible to diverse users where capture, editing and producing a movie are completely indistinguishable from one another. Gestures integrated with objects become a coherent interface on top of natural play. I build distributed, modular camera environments as well as the gesture interaction to control the environment. Finally as an impact, I design new technologies for children to explore geometrical and psychological perspectives while playing with their everyday toys.

Children using Picture This! access the visual idioms of the cherished genre's from which their character toys originate. Transitions and alternating viewpoints can be used to create complex videomaking effects, affecting the sense of time, the expression of emotion, the impression of an environement, and the overall immersion in a scene. Kids can be immersed in the styles of anime, or the dynamic action worlds that helped create their love for certain toys. Their stories can grow as they adopt and adapt the visual languages of their cherished media to create fundamentally new kinds of visual and oral stories.

6.5 Future directions

6.5.1 Towards the atomic revolution of digital kids!

In the vein of Gauthier [Gauthier and Moukalou, 2007, Gauthier, 1999, Gauthier, 1993], I claim that children with difficulties to talk and children who show some reserve towards socialization tend to spend a long period of time on the internet participating in remote communication. With the computer, the relationship between distance and proximity, direct communication, corporeal and indirect, or mediated communication is transformed. It is as if our digital kids privilege a communication in which the body is absent. This transformation might induce difficulties in sharing and exchanging across generations and difficulties in the position that parents take place in the growth of their children.

In the fifty's, children were developing concrete skills, similar to the skills of adults. Now, these concrete skills more directly translate to an imaginary, virtual world. The relationship to the body, the rhythm of lives, is different. Before the industrial revolution, a rhythmic life was imposed due to the constraints of working in the field, outside! Now, we neglect sun's motion in our lives. We eat according to happenstance, find abnormal quantity of food anytime of the day, forgetting that meals can have a social function. The social function of meals is replaced by their nutritional function. The physical constraints related to transportation have been transformed. We walk less, thus transforming our relationship to time and space as well as our relationship to the body: feelings, feeling tired, cold/heat or well-being.

Children have a predisposition to form groups in which learning by imitation is paramount. This helps children to leave the exclusive parental relationship and enter a more complex form of socialization - creating an identity and functional skills.

In psychoanalysis, authors show the importance of a mental construction that needs to connect to the outside world, necessarily developing through gesture and object manipulation [Leroi-Gourhan, 1964, Winnicott, 1971, Mendel, 1992, Montagner, 1997, Gibbs, 2006]. Play is key for social and individual development, a way to measure personal skills in comparison to others at the same time as measuring one's body, a necessary step imposed by the life as an adult.

Urban concentration has reduced the children's possibility to gather outside. The space for play and collective experience is disappearing. The parents themselves lose their everyday corporeal connection, their craftsmanship and their personal space. These transformations impact our ability to measure the consequences of our actions; this can explain a comeback of interest in magical thought in a world where the relationship between cause and effect is less and less certain. Not only has the quantity of available land changed, but also its quality of use has decreased. Before one could close his house with doors and windows, now it is completely impossible. The house walls have not only become porous through modern architecture, but the family remains in communication with the entire world through telecommunication, TV, internet, mobile phone, thus interrupting the paternal order of things.

Distinguishing between the inside and the outside world is more difficult than before (this relationship becomes more and more ambiguous). Distinguishing between private, individual, internal and external realities is becoming difficult.

Creating relationships between children is advantagous because it allows children to realize the humanity's fundamental destiny early on: as a collective of social beings [De Waal, 2005]. It also allows kids to find modes of learning outside of the parental relationship. Jean-Marie Gauthier and Roger Moukakou [Gauthier and Moukalou, 2007] explain that we need to reconsider a child's pleasure in finding groups and learning in a group, reevaluating the quality of learning that can happen within the group. Within groups, it is quite possible that kids can rediscover values of solidarity that are essential to our humanity. In the past, one would constitute a group that opposes itself to another, but today individuals are pushed towards homogeneity. Solidarity as a value is the most compromised, while individuation is assimilated to the general identical. Consumer western society can only live if it destroys values of sharing and solidarity, prefacing individualization.

Concepts of time and space are important for rational thinking. Such concepts are constituted and function via intuition. These intuitive forms of representation are constructed progressively while the child uses his corporeal skills. Corporeal exercise has a direct influence on the essential cognitive functions [Gibbs, 2006, Gauthier, 1999, Gauthier, 1993, Montagner, 1997].

6.5.2 Technical expansions to modify perspective taking, immersion, scale, otherness

> "*Work in digital media is setting the tone for aesthetic expression into the next century.*"
>
> Sefton-Green, in [Sefton-Green, 2000].

The Gesture Object Interfaces I present in my thesis, such as Picture This! or Dolltalk, are not just about movie-making or performing stories. Nor are they just about story-telling or story-building, as we usually think of it, at least in literacy research (focused on words, spoken or written). Instead, these environments integrate narrative and performative aspects to the production of a complete movie. They are enhanced - or augmented - puppet shows, or little doll-theaters, in which children pretend-play, role-play, and take perspectives. They allow players to weave together different modes of expression, such as visual, oral, gestural, and kinesthetic. The players can also take on different perspectives, roles, stances in the world; for instance, they can "speak" their own voice, let their characters "speak" on their behalf, they can be choreographers or enter as themselves into the show. Whatever they, or their characters, see, do, or say, gets projected, thus producing an effect on a screen, that becomes a part of the scene.

This allows for wonderful explorations in the future. I would like to:

- Play with scale by varying the size of the projection, the dimension of the toys in relation to their environment.
- Integrating light, as a contributing factor to a warm/cold atmosphere, a angry/friendly face, creating daytime/nighttime, creating firelight as if the toy was looking at a fire or creating TV light as if the person was watching TV, etc. All these light effects would contribute to a child's ability to express perspectives both from the toy's point of view by re-emphasizing the feeling and moods of a protagonist.
- Play with levels of immersion, i.e. from "miniaturized" stages -or little theaters- to "espaces environnants".
- Let the child move in-and-out of situations, the child as tiny (in the world) or gigantic.
- Offer the possibility for children to watch the final movies made by other children with the same toys, expanding on their own perspectives.

6.5.3 Towards a Gesture Object Interfaces language

I plan to research sophisticated editing in the flow of play. To do so, I want to implement a new stage of Picture This! where the system can be even further modularized and distributed into the open-ended play environment.

I will research a gesture language for children's toy-oriented play. There are so many roles that a toy can play in a child's narrative performance, that I would like to analyze and take into consideration the gestures a child makes when she transitions and embeds the actor, cameraman, director, audience and even when a countryside intervenes into the play. Thanks to this research, I envision embedding multiple eyes in an entire play scene so that a child may orchestrate among viewpoints in a gestural performance with her favorite toys, composing a more complex movie involving her entire playroom!

6.5.4 Perspective in collaborative agents

I will continue in the area of technologies that invite for anthropomorphism and projection, where children can test their boundaries between natural and artificial, control versus lack of control, what is "me" and what is not "me". Research is focused on emotional and sociable robots [Breazeal, 2003], and now, robot's are designed to help someone to be sociable; such a robot could teach, engage or just motivate a child into being sociable and understanding other's feelings and emotions, or, more succinctly, to take multiple perspectives. One compelling research case would define design principles of a robotic companion who could play with a child towards perspective taking.

6.5.5 Inventing a co-participation, co-creation, model of new media creation

In the vein of the anime Denno Coil [Iso, 2007], that envisions a digital world merging physical, microscopic illegals that needs special encodes, kids collecting metabugs to gain more credit, with a digital police that can also be hacked, I would like pursue a new class of anime, TV shows and online games connected to their physical world for children to not only rediscover the joy of living in a group along with the values of solidarity while belonging to a group of peers.

I will rethink the consumption of online media tools, video gaming platforms, TV shows and anime towards a co-creative model. I will involve participants in a child's social sphere, at a school, in afterschool programs to co-create media from their discoveries and activities in the physical environment.

Through a set of activities, children will gather bits of information in the physical space, by talking to their friends, their peers or even children they don't know. This would give them hints on where to go in the online video game to contribute to the next show, the next series of events in a TV show. This would work as a loop of activities and events, drawn from the physical life, encouraging children to communicate and develop strategies with each other.

6.6 The perspective taking gateway

Throughout my work, I explore the anthropomorphic nature of hybrid physical/digital objects and their promise to be catalysts towards unforeseen discoveries. I design objects to offer anima and perspective: from the haptic jacket that an autistic patient wears to express to the doctor what it means to be sensory defensive [Vaucelle et al., 2009a], to the electromagnetic field detector bracelet that transforms invisible information into visible data for a feeling of empowerment [Vaucelle et al., 2009b], or an environment where instead of you taking care of your avatar in massive online games, your avatar turns around and takes care of your biological needs [Vaucelle et al., 2010]. This trend in designing objects with perspectives led me to a fundamental quest in exploring the simplicity of tangible and gestural object interaction combined with the flexibility of digital systems.

Throughout this work, I have maintained a strict definition of perspective - Geometric/spatial perspective and psychological perspective. These dimensions of perspective are particularly important to children and their development. In the world, perspective is much larger. It is part of how we live, and relate to those around us. It is part of how we see others and how we see ourselves. Extending our interest in perspective into the everyday world, new technologies can be imagined that offer perspective in a number of dimensions toward empowering social opportunities, or, less humanistically, powerful new genres of consumer applications.

6.6.1 Healthcare

Physicians chase the ellusive bedside manner, to empathize with their patients, understand their process of disease and recovery, and propose better, more efficient treatments. Many of these functions can be subsumed by perspective. Given the medical information at hand, how can a doctor access a patient's status, in particular during on-going treatments, or where the day-to-day experience of a patient is critically important? These situations abound in psychiatric care, and especially when patients transition from residential hospitals into the community.

During somewhat infrequent check-ups, a doctor might ask whether a patient is adhering to their medication, or the extent to which they fulfill certain behavioral targets. More elaborate methods to access the patient's perspective might include extensive talk-based therapy sessions, a waning model. State-of-the-art, home-based healthcare technologies rarely amount to more than patient monitoring, with no signs of connecting to the patient's underlying experience. I propose technologies to interface and drive perspective between the doctor and the patient.

One such project would consist of an augmented suit to translate perspective between doctor and patient with a resolution extending into the moments of everyday life. Healthcare jackets aren't necessarily anything new [Vaucelle et al., 2009a, Vaucelle and Abbas, 2007], but something that goes beyond passive biomonitoring and enters into the space of patient experience would possibly impact the relationship and understanding between patient and physician. Relational and empathic concerns are *the* significant barrier to compliance, and genuine care.

Language is a prosthesis compared to the sensory experience. It is a weak translation of what a patient goes through, and suffers from. How in pain are you on a scale of one-to-ten? There can only be a superficial answer. The art of therapeutic process relies, in part, on the doctor's ability to divine the patient's experience from whatever data is available, and to then propose appropriate effective treatment.

Technology can translate the sensory experience for the physician to project into the experience of the patient, and to take perspective where that perspective could be advantageous to persciprtion.

The person might not even know what they're going through, but through the relationship to this object, this tool, the doctor gains insite, and can reflect that impression back to the patient. The object as a technology can provide the structure to project and take perspective. The therapeutic relationship can move forward as a result.

6.6.2 The first third-person shooter game

The gaming industry is expected to generate 12.5 billion in revenue in the US alone in the year 2011, and the developer that owns the most popular game properties can easily expect to generate 1 billion dollars in revenue. Despite its size and profitability, the number of different genres of games is extremely limited. The dominant genres are the following:

- Massive multiplayer online games, players assume roles in collaborative online worlds where they compete, cooperate, and strive to accomplish achievements for their online characters. A major MMORPG had 11.5 million subscribers as of December 2008, and has only grown since then.
- First-person-shooters, players take a first-person-view of their character, who then navigates through various maps engaging in combat with a variety of projective weapons and guns. Todays game market is flooded with first-person-shooters. It is fairly common for yearly top-10 lists of the best games released from the previous year to be comprised of 50% first person shooters.
- Simulators, players drive, fly, or otherwise operate vehicles in as close to a reproduction of the authentic experience as possible.
- Real-time-Strategy, players take a god's-eye-view on a theatre of war where they balance troop and the economy to build troops to defeat an opponent who simultaneously tries to accomplish the same.
- Fighers, engage players in one-on-one combat, sequencing strings of moves, or attacks, to best their opponents in tournament-style fighting contests.

The gaming industry, for its size and profitability, rarely innovates to develop new genres of games. Major titles beget other major titles in the same genre. New interface technologies, like Voice over Ip, tend to add functionality as a layer on top of existing genres. Next year's first-person-shooters will have integrated social networking functions.

In future work I will subject the dominant genres of games to an analysis of their perspective taking affordances. In which way to the exising models limit the relationship to the characters being controlled, other players characters, and the computer controlled characters in the game? Reconsidering the perspective taking relationships in these genres will expose opportunities for new games and I could create the first third-person shooter game!

Development in each of these areas benefits from the same methods employed in this thesis. Participatory design would discover the interface opportunities of intended user's in the contexts of use. Comparative analysis would iterate and refine prototypes, developing outcomes both technically and in the social space.

Bibliography

[Abbas, 2006] Abbas, Y. (2006). *Neo-nomads: Designing Environments for Living in the Age of Mental, Physical and Digital Mobilities.* PhD thesis, Harvard University, Graduate School of Design.

[Ackerman, 1991] Ackerman, D. (1991). *A Natural History of the Senses.* Vintage Books, New York.

[Ackermann, 1996] Ackermann, E. (1996). Perspective taking and object construction. In Kafai, Y. and Resnick, M., editors, *Constructionism in Practice*, pages 25–36. Lawrence Erlbaum Associates, Hillsdale.

[Ackermann, 2004] Ackermann, E. (2004). Constructing knowledge and transforming the world. In Tokoro, M., Tokoro, M., and Steels, L., editors, *A Learning Zone of One's Own: Sharing Representations and Flow in Collaborative Learning Enviroments*, number 15-35. IOS Press, Amsterdam, The Netherlands, The Netherlands.

[Africano et al., 2004] Africano, D., Berg, S., Lindbergh, K., Lundholm, P., Nilbrink, F., and Persson, A. (2004). Designing tangible interfaces for children's collaboration. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*, pages 853–868, New York, NY, USA. ACM.

[Akyol and Oğuz, 2007] Akyol, A. K. and Oğuz, V. (2007). An investigation into the perspective taking skills of 6-year-old children in terms of pet feeding and gender variables. *Journal of Humanity and Social Sciences*, 2(1):08–13.

[Ananny, 2002] Ananny, M. (2002). Supporting children's collaborative authoring: practicing written literacy while composing oral texts. In *CSCL '02: Proceedings of the Conference on Computer Support for Collaborative Learning*, pages 595–596. International Society of the Learning Sciences.

[Astington, 1988] Astington, J. (1988). *Developing Theories of Mind.* Cambridge University Press, Cambridge.

[Bagnall, 2004] Bagnall, D. (2004). The filmmaking robot. Available at http://www.halo.gen.nz/robot/, last accessed 2010-07-27.

[Bamberg, 1997] Bamberg, M. (1997). *Narrative Development: six approaches.* L. Erlbaum Associates, Hillsdale.

[Barthes, 1981] Barthes, R. (1981). *Camera Lucida.* Fontana Paperbacks, London.

[Baudrillard, 2006] Baudrillard, J. (2006). *The System of Objects.* Verso, London.

[Bell and Gray, 2001] Bell, G. and Gray, J. (2001). Digital immortality. *Commun. ACM*, 44(3):28–31.

[Benson, 1993] Benson, M. (1993). The structure of four-and five-year-olds' narratives in pretend play and storytelling. *First language.*

[Birchfield et al., 2006] Birchfield, D., Ciufo, T., and Minyard, G. (2006). Smallab: a mediated platform for education. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Educators program*, page 33, New York, NY, USA. ACM.

[Bransford et al., 1990] Bransford, J., Sherwood, R. D., Hasselbring, T., Kinzer, C. K., and Williams, S. (1990). Anchored instruction: Why we need it and how technology can help. In Nix, D. and Spiro, R., editors, *Cognition, Education, and Multimedia: Exploring Ideas in High Technology*, pages 115–141. L. Erlbaum, Hillsdale.

[Breazeal, 2003] Breazeal, C. (2003). Emotion and sociable humanoid robots. *Int. J. Hum.-Comput. Stud.*, 59(1-2):119–155.

[Brondmo and Davenport, 1989] Brondmo, H. and Davenport, G. (1989). Creating and viewing the elastic charles – a hypermedia journal. In McAlesse, R. and Green, C., editors, *Hypertext: State of the ART*, volume Chapter 5, pages 43–51.

[Brosterman, 1997] Brosterman, N. (1997). *Inventing Kindergarten.* H.N. Abrams, New York.

[Bruner, 1983a] Bruner, J. (1983a). *Child's Talk.* W.W. Norton, New York.

[Bruner, 1983b] Bruner, J. (1983b). *Child's Talk*, chapter The Growth of Reference, pages 65–88. W.W. Norton, New York.

[Bruner, 1983c] Bruner, J. (1983c). *Child's Talk*, chapter Play, Games, and Language, pages 43–64. W.W. Norton, New York.

[Bruner, 1986] Bruner, J. (1986). *Actual Minds, Possible Worlds.* Harvard University Press, Cambridge.

[Bukowski et al., 1996] Bukowski, W., Newcomb, A. F., and Hartup, W. W. (1996). *The Company They Keep.* Cambridge University Press, Cambridge.

[Cameron and Wang, 1999] Cameron, C. and Wang, M. (1999). Frog, where are you? children's narrative expression over the telephone. *Discourse Processes*, 28:217–236.

[Cassell and Ryokai, 2001] Cassell, J. and Ryokai, K. (2001). Making space for voice: Technologies to support children's fantasy and storytelling. *Personal Ubiquitous Comput.*, 5(3):169–190.

[Cassell et al., 2000] Cassell, J., Ryokai, K., Druin, A., Klaff, J., Laurel, B., and Pinkard, N. (2000). Story spaces: interfaces for children's voices. In *CHI '00: CHI '00 extended abstracts on Human factors in computing systems*, pages 243–244, New York, NY, USA. ACM.

[Cherry, 2005] Cherry, S. (2005). Total recall life recording software. In *IEEE Spectrum*, volume 42, pages pp24–30.

[Chomsky, 1986a] Chomsky, N. (1986a). *Barriers.* MIT Press, Cambridge.

[Chomsky, 1986b] Chomsky, N. (1986b). *Knowledge of Language.* Praeger, New York.

[Chomsky, 1995] Chomsky, N. (1995). *The Minimalist Program.* The MIT Press, Cambridge.

[Chung et al., 2010] Chung, K., Shilman, M., Merrill, C., and Ishii, H. (2010). Onobject. Available at http://tangible.media.mit.edu/project.php?recid=146, last accessed 2010-07-27.

[D-Fuse, 2006] D-Fuse, D.-F. (2006). *Vj (Includes Dvd).* Laurence King, London.

[Davenport, 2001] Davenport, G. (2001). Flights of fantasy. Available at http://ic.media.mit.edu/projects/FlightsOfFantasy/, last accessed 2010-07-28.

[De Waal, 2005] De Waal, F. (2005). *Our Inner Ape.* Riverhead Hardcover, City.

[Decortis, 2005] Decortis, F. (2005). Survey on narrative and learning environments. *Kaleidoscope network of excellence.*

[Decortis, 2008] Decortis, F. (2008). *L'activité narrative et les nouvelles technologies pour les enfants.* PhD thesis, Université de Paris VIII.

[Denham, 1986] Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers. *Child Development*, 57 n1:194–201.

[Dewey, 1938] Dewey, J. (1938). *Experience and Education.* Macmillan, London and New York.

[Dietz and Leigh, 2001] Dietz, P. and Leigh, D. (2001). Diamondtouch: a multi-user touch technology. In *UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 219–226, New York, NY, USA. ACM.

[Douglas and Aki, 1993] Douglas, S. and Aki, N. (1993). *Participatory Design: Principles and Practices.* L. Erlbaum Associates, Hillsdale.

[Druin et al., 1998] Druin, A., Bederson, B., Boltman, A., Miura, A., Knotts-Callahan, D., and Platt, M. (1998). *The design of children's technology*, chapter Children as our technology design partners, pages 51–72. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[Dunne, 2006] Dunne, A. (2006). *Hertzian Tales.* MIT Press, Cambridge.

[Eisner, 2008] Eisner, W. (2008). *Comics and Sequential Art.* W.W. Norton, New York.

[Elsner and Cardinal, 1994] Elsner, J. and Cardinal, R. (1994). *The Cultures of Collecting.* Harvard University Press, Cambridge.

[Fein, 1980] Fein, G. (1980). Play and the acquisition of symbols. In Katz, L., editor, *Current Topics in Early Childhood Education, Volume 3*, volume 2, pages 211–212. Ablex Publishing, Westport.

[Feynman, 1959] Feynman, R. P. (1959). *There's Plenty of Room at the Bottom.* Annual meeting of the American Physical Society., California Institute of Technology (Caltech). Transcript at: http://www.zyvex.com/nanotech/feynman.html.

[Flavell, 1988] Flavell, J. (1988). The development of children's knowledge about the mind: from cognitive connexions to mental representations. In Astington, J., Harris, P. L., and Olson, D. R., editors, *Developing Theories of Mind*. Cambridge University Press, Cambridge.

[Flavell, 1990] Flavell, J. (1990). Perspectives on perspective-taking. *Paper presented at the 20th Annual Symposium of the Jean Piaget Society.*

[Forty, 1992] Forty, A. (1992). *Objects of Desire*. Thames and Hudson, London.

[Fraden, 2004] Fraden, J. (2004). *Handbook of Modern Sensors*. AIP Press/Springer, New York.

[Freed, 2010] Freed, N. (2010). Toys keeping in touch: technologies for distance play. In *TEI '10: Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, pages 315–316, New York, NY, USA. ACM.

[Frei et al., 2000] Frei, P., Su, V., Mikhak, B., and Ishii, H. (2000). curlybot: designing a new class of computational toys. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–136, New York, NY, USA. ACM.

[Frens et al., 2004] Frens, J., Djajadiningrat, J., and Overbeeke, C. (2004). Rich interaction: issues. ambient intelligence. In Markopoulos, P., Eggen, B., Aarts, E., and Crowley, J., editors, *EUSAI 2004, LNCS*, volume 3295, pages 271–278. Springer, Heidelberg.

[Gardner, 1999] Gardner, H. (1999). Multiple approaches to understanding. In Reigeluth, C. and Carr-Chellman, A. A., editors, *Instructional-Design Theories and Models: A New Paradigm of Instructional Theory*, volume II. Lawrence Erlbaum Associates, Hillsdale.

[Gauthier, 1993] Gauthier, J.-M. (1993). *L'enfant malade de sa peau*. Dunod, Paris.

[Gauthier, 1999] Gauthier, J.-M. (1999). *Le corps de l'enfant psychotique.* Dunod, Paris.

[Gauthier and Moukalou, 2007] Gauthier, J.-M. and Moukalou, R. (2007). *De La guerre des boutons à Harry Potter : Un siècle d'évolution de l'espace-temps des adolescents*. Mardaga Editions.

[Gaver, 2002] Gaver, B. (2002). Provocative awareness. *Computer Supported Cooperative Work.*, 11 (3):475–493.

[Gemmell et al., 2006] Gemmell, J., Bell, G., and Lueder, R. (2006). Mylifebits: a personal database for everything. *Commun. ACM*, 49(1):88–95.

[Gibbs, 2006] Gibbs, R. J. (2006). *Embodiment and Cognitive Science*. Cambridge University Press, Cambridge.

[Gokhale, 1995] Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *journal of Technology Education*, 7(1).

[Goodman, 1986] Goodman, Y. (1986). Children coming to know literacy. In Teale, W. and Sulzby, E., editors, *Emergent Literacy: Writing and reading*, pages 1–14. Ablex Pub. Corp, Norwood.

[Greimas and Courtès, 1986] Greimas, A. J. and Courtès, J. (c1979-c1986). *Sémiotique : dictionnaire raisonné de la théorie du langage*, volume I. Hachette, Paris.

[Hall, 1980] Hall, E. (1980). *The Silent Language*. Greenwood Press, Westport.

[Hall, 1989] Hall, E. (1989). *The Dance of Life*. Anchor Books, Garden City.

[Harel and Papert, 1991] Harel, I. and Papert, S. (1991). *Constructionism*. Ablex Pub. Corp, Norwood.

[Harris, 2000] Harris, P. (2000). *The Work of the Imagination*. Blackwell Publishers, Cambridge.

[Henderson et al., 2002] Henderson, A., Pehoski, C., and Murray, E. A. (2002). Visual-spatial abilities. In Bundy, A., Lane, S., and Murray, E., editors, *Sensory Integration.*, pages 123–140. F.A. Davis, Philadelphia.

[Holmquist et al., 1999] Holmquist, L. E., Redström, J., and Ljungstrand, P. (1999). Token-based acces to digital information. In *HUC '99: Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*, pages 234–245, London, UK. Springer-Verlag.

[Horowitz, 1989] Horowitz, P. (1989). *The Art of Electronics*. Cambridge University Press, Cambridge.

[Huttenlocher and Presson, 1973] Huttenlocher, J. and Presson, C. C. (1973). Mental rotation and the perspective problem. *Cognitive Psychology*, 4:277–299.

[Ishii and Kobayashi, 1992] Ishii, H. and Kobayashi, M. (1992). Clearboard: a seamless medium for shared drawing and conversation with eye contact. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 525–532, New York, NY, USA. ACM.

[Ishii and Ullmer, 1997] Ishii, H. and Ullmer, B. (1997). Tangible bits: towards seamless interfaces between people, bits and atoms. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 234–241, New York, NY, USA. ACM.

[Iso, 2007] Iso, M. (2007). Denno coil. Available at http://www.architectradure.com/2009/10/23/what-a-visionary-anime/, last accessed 2010-07-28.

[Jacob et al., 2002] Jacob, R. J. K., Ishii, H., Pangaro, G., and Patten, J. (2002). A tangible interface for organizing information using a grid. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 339–346, New York, NY, USA. ACM.

[Jeremijenko, 2003] Jeremijenko, N. (2003). Feral robotic dogs. Available at http://www.nyu.edu/projects/xdesign/feralrobots/, last accessed 2010-07-26.

[Johnson et al., 1999] Johnson, M. P., Wilson, A., Blumberg, B., Kline, C., and Bobick, A. (1999). Sympathetic interfaces: using a plush toy to direct synthetic characters. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 152–158, New York, NY, USA. ACM.

[Jonassen, 1999] Jonassen, H. (1999). Designing constructivist learning environments. In Reigeluth, C. and Carr-Chellman, A. A., editors, *Instructional-Design Theories and Models: A New Paradigm of Instructional Theory*, volume II. Lawrence Erlbaum Associates, Hillsdale.

[Jonassen, 2000] Jonassen, H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48:63–85.

[Kegan, 1982] Kegan, R. (1982). *The Evolving Self.* Harvard University Press, Cambridge.

[Kristeva, 1999] Kristeva, J. (1999). Toccata and fugue for the foreigner. In *Strangers to Ourselves*, number 1-40. Columbia University Press, New York.

[Kurzweil, 2006] Kurzweil, R. (2006). *The Singularity Is near*. Penguin (Non-Classics), City.

[Labrune and Mackay, 2005] Labrune, J.-B. and Mackay, W. (2005). Tangicam: exploring observation tools for children. In *IDC '05: Proceedings of the 2005 conference on Interaction design and children*, pages 95–102, New York, NY, USA. ACM.

[Landry, 2008] Landry, B. M. (2008). Storytelling with digital photographs: supporting the practice, understanding the benefit. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pages 2657–2660, New York, NY, USA. ACM.

[Langer, 1997] Langer, E. (1997). *The Power of Mindful Learning*. Addison-Wesley, Boston.

[Laurel, 1997] Laurel, B. (1997). Purple moon. Available at http://en.wikipedia.org/wiki/Purple$_Moon, lastaccessed 2010-07-26$.

[Lee and Ishii, 2010] Lee, J. and Ishii, H. (2010). Beyond: collapsible tools and gestures for computational design. In *CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, pages 3931–3936, New York, NY, USA. ACM.

[Lee and Kirby, 2005] Lee, S. and Kirby, J. (2005). *Fantastic Four Omnibus Volume 1 Hc*. Marvel Comics, New York.

[Leithinger and Ishii, 2010] Leithinger, D. and Ishii, H. (2010). Relief: a scalable actuated shape display. In *TEI '10: Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, pages 221–222, New York, NY, USA. ACM.

[Leroi-Gourhan, 1964] Leroi-Gourhan (1964). *Le geste et la parole. Technique et Langage*, volume Tome 1. Albin Michel. Sciences d'aujourd'hui., Paris.

[Letessier and Bérard, 2004] Letessier, J. and Bérard, F. (2004). Visual tracking of bare fingers for interactive surfaces. In *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 119–122, New York, NY, USA. ACM.

[Levy et al., 2006] Levy, F., C., R. D., Horowitz, G., M., S.-R., and Subkowski, P. (2006). The sensibility of the collector. *Journal of the Philoctetes, Center for the Multidisciplinary study of imagination.*, 1(2).

[Lew, 2003] Lew, M. (2003). Office voodoo: a real-time editing engine for an algorithmic sitcom. In *SIGGRAPH '03: ACM SIGGRAPH 2003 Sketches & Applications*, pages 1–1, New York, NY, USA. ACM.

[Lew, 2004] Lew, M. (2004). Live cinema: designing an instrument for cinema editing as a live performance. In *NIME '04: Proceedings of the 2004 conference on New interfaces for musical expression*, pages 144–149, Singapore, Singapore. National University of Singapore.

[MacCabe, 1997] MacCabe, A. (1997). Developmental and cross-cultural aspects of children's narration. In Bamberg, M., editor, *Narrative Development: six approaches*, pages 137–174. L. Erlbaum Associates, Hillsdale.

[Mandler and Johnson, 1977] Mandler, J. and Johnson, N. (1977). Remembrance of things parsed: story structure and recall. *Cognitive Psychology*, 9(1):111–151.

[Mann, 1994] Mann, S. (1994). Lifelog. Available at http://en.wikipedia.org/wiki/Lifelog, last accessed 2010-07-28.

[Mauss, 1967] Mauss, M. (1967). *The Gift*. Norton, New York.

[Mazalek and Davenport, 2003] Mazalek, A. and Davenport, G. (2003). A tangible platform for documenting experiences and sharing multimedia stories. In *ETP '03: Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence*, pages 105–109, New York, NY, USA. ACM.

[Mccarthy and Tezuka, 2009] Mccarthy, H. and Tezuka, O. (2009). *The Art of Osamu Tezuka*. Abrams ComicArts, City.

[Mccloud, 1994] Mccloud, S. (1994). *Understanding Comics*. HarperPerennial, New York.

[Mendel, 1992] Mendel, G. (1992). *La société n'est pas une famille: de la psychanalyse à la sociopsychanalyse*. La Découverte.

[Merrill et al., 2007] Merrill, D., Kalanithi, J., and Maes, P. (2007). Siftables: towards sensor network user interfaces. In *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 75–78, New York, NY, USA. ACM.

[Michelon and Zacks, 2006] Michelon, P. and Zacks, J. M. (2006). Two kinds of visual perspective taking. *Percept Psychophys*, 68(2):327–337.

[Minsky, 1986] Minsky, M. (1986). *The Society of Mind*. Simon and Schuster, New York.

[Miyabara and Sugimoto, 2006] Miyabara, M. and Sugimoto, T. (2006). Movie cards. Available at http://www.we-make-money-not-art.com/archives/2006/03/japanese-media-1.php, last accessed 2010-07-27.

[Montagner, 1997] Montagner, H. (1997). Les compétences du bébé et du jeune enfant à induire la communication avec leurs différents partenaires. *Communication et Organisation*, n°12 (Induction et communication):33–125.

[Montemayor et al., 2004] Montemayor, J., Druin, A., Chipman, G., Farber, A., and Guha, M. L. (2004). Tools for children to create physical interactive storyrooms. *Comput. Entertain.*, 2(1):12–12.

[Montessori, 1912] Montessori, M. (1912). *The Montessori Method*. Frederick Stokes, New York.

[Montessori, 1917] Montessori, M. (1917). *The Advanced Montessori Method*. Frederick Stokes, New York.

[Nelson, 1996] Nelson, K. (1996). *Language in Cognitive Development: The Emergence of the Mediated Mind*. Cambridge University Press, Cambridge.

[Nicolopoulou, 1997] Nicolopoulou, A. (1997). Children and narratives: toward an interpretive and sociocultural approach. In Bamberg, M., editor, *Narrative Development: six approaches*, pages 179–216. L. Erlbaum Associates, Hillsdale.

[Ninio and Bruner, 1978] Ninio, A. and Bruner, J. (1978). The achievement and antecedents of labeling. *Journal of Child Language*, 5:1–15.

[Norman, 2007] Norman, D. A. (2007). The next ui breakthrough, part 2: physicality. *interactions*, 14(4):46–47.

[Okakura, 1964] Okakura, K. (1964). *The Book of Tea*. Dover Publications, New York.

[O'Neill and Pearce, 2001a] O'Neill, D. K. and Pearce, M. J. (2001a). A new perspective on the relation between preschoolers' narrative ability and later academic competence. In *Biennial meeting of Society for Research in Child Development*, Minneapolis, MN.

[O'Neill and Pearce, 2001b] O'Neill, D. K. and Pearce, M. J. (2001b). A new perspective on the relation between preschoolers' narrative ability and later academic competence. In *Cognitive Development Society Meeting*, Virginia Beach, VA.

[O'Neill et al., 2004] O'Neill, D. L., Pearce, M. J., and Pick, J. L. (2004). Preschool children's narratives and performance on the peabody individualized achievement test – revised: Evidence of a relation between early narrative and later mathematical ability. *First Language, SAGE Publications*, 24(2):149–183.

[Papert, 1993] Papert, S. (1993). *Mindstorms*. Basic Books, New York.

[Patten et al., 2001] Patten, J., Ishii, H., Hines, J., and Pangaro, G. (2001). Sensetable: a wireless object tracking platform for tangible user interfaces. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 253–260, New York, NY, USA. ACM.

[Petruzzellis, 1994] Petruzzellis, T. (1994). *The Alarm, Sensor amp; Security Circuit Cookbook*. Tab Books, Blue Ridge Summit.

[Piaget, 1967] Piaget, J. (1967). The coordination of perspectives. In Piaget, J. and Inhelder, B., editors, *The Child's Conception of Space*, number 209-246. W. W. Norton, New York.

[Piaget and Inhelder, 1967] Piaget, J. and Inhelder, B. (1967). *The Child's Conception of Space*. W. W. Norton, New York.

[Premaratne et al., 2006] Premaratne, P., Safaei, F., and Nguyen, Q. (2006). Moment invariant based control system using hand gestures.

[Putnam, 2001] Putnam, J. (2001). *Art and Artifact*. Thames and Hudson, London.

[Quasthoff, 1997] Quasthoff, U. (1997). An interactive approach to narrative development. In Bamberg, M., editor, *Narrative Development: six approaches.*, pages 51–84. L. Erlbaum Associates, Hillsdale.

[Raffle et al., 2007] Raffle, H., Vaucelle, C., Wang, R., and Ishii, H. (2007). Jabberstamp: embedding sound and voice in traditional drawings. In *IDC '07: Proceedings of the 6th international conference on Interaction design and children*, pages 137–144, New York, NY, USA. ACM.

[Raffle et al., 2004] Raffle, H. S., Parkes, A. J., and Ishii, H. (2004). Topobo: a constructive assembly system with kinetic memory. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 647–654, New York, NY, USA. ACM.

[Reigeluth and Carr-Chellman, 1999] Reigeluth, C. and Carr-Chellman, A. A. (1999). *Instructional-Design Theories and Models: A New Paradigm of Instructional Theory*, volume II. Lawrence Erlbaum Associates, Hillsdale.

[Resnick, 2002] Resnick, M. (2002). Rethinking learning in the digital age. In *The Global Information Technology Report: Readiness for the Networked World*. Oxford University Press, New York.

[Resnick, 2006] Resnick, M. (2006). Computer as paintbrush: technology, play, and the creative society. In Singer, D., Michnick Golinkoff, R., and Hirsh-Pasek, K., editors, *Play=Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth*, pages 192–206. Oxford University Press, Oxford Oxfordshire.

[Rizzo et al., 2004] Rizzo, A., Marti, P., Decortis, F., Rutgers, J., and Thursfield, P. (2004). Building narrative experiences for children through real time media manipulation: Pogo world. In Blythe, M., Overbeeke, K., Monk, A., and Wright, P., editors, *Funology: from usability to enjoyment*, pages 189–199. Kluwer Academic Publishers, Norwell, MA, USA.

[Ross et al., 1993] Ross, M., Radnor, H., Mitchell, S., and Bierton, C. (1993). *Assessing Achievement in the Arts*. Open University Press, Milton Keynes.

[Roy, 2009] Roy, D. (2009). New horizons in the study of child language acquisition. In *Proceddings of Interspeech 2009.*, Brighton, England.

[Runner, 2009] Runner, G. (2009). Game runner. Available at http://www.gamerunner.us/index.htm, last accessed 2010-07-26.

[Rutherford, 1990] Rutherford, F. (1990). *Science for All Americans, Education for a changing future, AAAS Project 2061.* Oxford University Press, Oxford Oxfordshire.

[Ryokai et al., 2004] Ryokai, K., Marti, S., and Ishii, H. (2004). I/o brush: drawing with everyday objects as ink. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 303–310, New York, NY, USA. ACM.

[Ryokai et al., 2003] Ryokai, K., Vaucelle, C., and Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning. *Journal of Computer Assisted Learning. Special Issue: Children and new technology.*, 19:195–208.

[Salen, 2007] Salen, K. (2007). Institute of play. Available at http://www.instituteofplay.com/node/101, last accessed 2010-07-26.

[Salen and Zimmerman, 2004] Salen, K. and Zimmerman, E. (2004). *Rules of Play*. MIT Press, Cambridge.

[Salvador et al., 2004] Salvador, T., Barile, S., and Sherry, J. (2004). Ubiquitous computing design principles: supporting human-human and human-computer transactions. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*, pages 1497–1500, New York, NY, USA. ACM.

[Scarborough, 2009] Scarborough, S. (2009). *Cool Spaces for Kids*. Hamlyn, London.

[Scarlett and Wolf, 1979] Scarlett, W. and Wolf, D. (1979). When it's only make-believe: The construction of a boundary between fantasy and reality in storytelling. *New Directions for Child Development.*, 6:29–40.

[Schön, 1983] Schön, D. (1983). *The Reflective Practitioner: How Professionals Think In Action*. Basic Books, New York.

[Sefton-Green, 2000] Sefton-Green, J. (2000). *Evaluating Creativity. Making and Learning by Young People*. Routledge, New York.

[Seitinger, 2009] Seitinger, S. (2009). Designing for spatial competence. In *IDC '09: Proceedings of the 8th International Conference on Interaction Design and Children*, pages 123–130, New York, NY, USA. ACM.

[Shantz, 1975] Shantz, C. (1975). The development of social cognition. In Heterington, E., editor, *Review of Child Development Research*, volume 5. University of Chicago Press.

[Sharlin et al., 2004] Sharlin, E., Watson, B., Kitamura, Y., Kishino, F., and Itoh, Y. (2004). On tangible user interfaces, humans and spatiality. *Personal Ubiquitous Comput.*, 8(5):338–346.

[Shusterman, 2006] Shusterman, G. (2006). Super elephant. Available at http://atomicbee.com/superelephant.html, last accessed 2010-07-26.

[Singer et al., 2006] Singer, D., Michnick Golinkoff, R., and Hirsh-Pasek, K. (2006). *Play=Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth*. Oxford University Press, Oxford Oxfordshire.

[Singer and Singer, 1990] Singer, D. and Singer, J. (1990). *The House of Make-Believe*. Harvard University Press, Cambridge.

[SMALLab, 2008] SMALLab (2008). Smallab. Available at http://ame2.asu.edu/projects/emlearning/index.php, last accessed 2010-07-26.

[Smith and Blankinship, 2000] Smith, B. K. and Blankinship, E. (2000). Justifying imagery: multimedia support for learning through explanation. *IBM Syst. J.*, 39(3-4):749–767.

[Smith, 2006] Smith, M. (2006). *The Prosthetic Impulse*. MIT Press, Cambridge.

[Snow, 1983] Snow, C. (1983). Literacy and language: Relationships during the preschool years. *Harvard Educational Review.*, 53, 2:165–189.

[Sokoler and Edeholt, 2002] Sokoler, T. and Edeholt, H. (2002). Physically embodied video snippets supporting collaborative exploration of video material during design sessions. In *NordiCHI '02: Proceedings of the second Nordic conference on Human-computer interaction*, pages 139–148, New York, NY, USA. ACM.

[Somers, 2000] Somers, J. (2000). Measuring the shadow or knowing the bird. evaluation and assessment of drama in education. In Sefton-Green, J., editor, *Evaluating Creativity. Making and Learning by Young People.*, pages 107–128. Routledge, New York.

[Spiegelman, 1986] Spiegelman, A. (1986). *Maus*. Pantheon Books, New York.

[Stein and Albro, 1997] Stein, N. and Albro, E. (1997). Building complexity and coherence: children's use of goal-structured knowledge in telling stories. In Bamberg, M., editor, *Narrative Development: six approaches*, pages 5–44. L. Erlbaum Associates, Hillsdale.

[Stein and Glenn, 1979] Stein, N. and Glenn, C. (1979). An analysis of story comprehension in elementary school children. In Freedle, R., editor, *New Directions in Discourse Processing, Advances in discourse processes.*, volume 2, pages 53–120. Ablex Pub. Corp, Norwood.

[Steinberg and Gitomer, 1993] Steinberg, L. S. and Gitomer, D. H. (1993). Cognitive task analysis, interface design, and technical troubleshooting. In *IUI '93: Proceedings of the 1st international conference on Intelligent user interfaces*, pages 185–191, New York, NY, USA. ACM.

[Steinberg and Gitomer, 1996] Steinberg, L. S. and Gitomer, D. H. (1996). Intelligent tutoring and assessment built on an understanding of a technical problem-solving task. *Instructional Science*, 24 (3):221–258.

[Subkowski, 2006] Subkowski, P. (2006). On the psychodynamics of collecting. *International journal of psychoanalysis.*

[Svendsen, 1934] Svendsen, M. (1934). Children's imaginary companions. *Archives of Neurology and Psychiatry*, 2:985–99.

[Tanaka, 2006] Tanaka, Y. (2006). Plable. Available at http://yumikotanaka.net/, last accessed 2010-07-27.

[Taylor, 2009] Taylor, A. S. (2009). Ethnography in ubiquitous computing. In Krumm., J., editor, *Ethnography in Ubiquitous Computing.*, pages pp. 203–236. Chapman and Hall/CRC., Boca Raton, FL.

[Taylor, 1999] Taylor, M. (1999). *Imaginary Companions and the Children Who Create Them.* Oxford University Press, Oxford Oxfordshire.

[Teale and Sulzby, 1986] Teale, W. and Sulzby, E. (1986). *Emergent Literacy: Writing and reading.* Ablex Pub. Corp, Norwood.

[Tezuka, 2002] Tezuka, O. (2002). *Astro Boy.* Dark Horse Comics, Milwaukie.

[Truong et al., 2004] Truong, K. N., Richter, H., Hayes, G. R., and Abowd, G. D. (2004). Devices for sharing thoughts and affection at a distance. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*, pages 1203–1206, New York, NY, USA. ACM.

[Turkle, 1995] Turkle, S. (1995). *Life on the Screen.* Simon and Schuster, New York.

[Ullmer and Ishii, 1999] Ullmer, B. and Ishii, H. (1999). mediablocks: tangible interfaces for online media. In *CHI '99: CHI '99 extended abstracts on Human factors in computing systems*, pages 31–32, New York, NY, USA. ACM.

[Ullmer and Ishii, 2000] Ullmer, B. and Ishii, H. (2000). Emerging frameworks for tangible user interfaces. *IBM Syst. J.*, 39(3-4):915–931.

[Vaucelle, 2002] Vaucelle, C. (2002). Dolltalk : a computational toy to enhance narrative perspective-talking. Master's thesis, Massachusetts Institute of Technology.

[Vaucelle, 2008] Vaucelle, C. (2008). The everyday collector. In *Published in the proceedings of Ubicomp*, volume Tenth International Conference on Ubiquitous Computing. ACM Press.

[Vaucelle, 2009] Vaucelle, C. (2009). Collect to connect in the mobile age. In Abbas, Y. and Dervin, F., editors, *Digital Technologies of the Self*, chapter Identity Gathering. Cambridge Scholars Publishing, City.

[Vaucelle and Abbas, 2007] Vaucelle, C. and Abbas, Y. (2007). Touch: sensitive apparel. In *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*, pages 2723–2728, New York, NY, USA. ACM.

[Vaucelle et al., 2005a] Vaucelle, C., Africano, D., Davenport, G., Wiberg, M., and Fjellstrom, O. (2005a). Moving pictures: looking out/looking in. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Educators program*, page 27, New York, NY, USA. ACM.

[Vaucelle et al., 2009a] Vaucelle, C., Bonanni, L., and Ishii, H. (2009a). Design of haptic interfaces for therapy. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 467–470, New York, NY, USA. ACM.

[Vaucelle and Davenport, 2004a] Vaucelle, C. and Davenport, G. (2004a). An open-ended tool to compose movies for cross-cultural digital storytelling: Textable movie. In *ICHIM 04 - Digital Culture and Heritage / Patrimoine and Culture Numérique*.

[Vaucelle and Davenport, 2004b] Vaucelle, C. and Davenport, G. (2004b). A system to compose movies for cross-cultural storytelling: Textable movie. In Göbel, S., Spierling, U., Hoffmann, A., Iurgel, I., Schneider, O., Dechau, J., and Feix, A., editors, *TIDSE 2004. Lecture Notes in Computer Sciences*, volume 3105, pages 126–131, Heidelberg. Springer.

[Vaucelle et al., 2003] Vaucelle, C., Davenport, G., and Jehan, T. (2003). Textable movie: improvising with a personal movie database. In *SIGGRAPH '03: ACM SIGGRAPH 2003 Sketches & Applications*, pages 1–1, New York, NY, USA. ACM.

[Vaucelle et al., 2005b] Vaucelle, C., Gorman, M. J., Clancy, A., and Tangney, B. (2005b). Re-thinking real time video making for the museum exhibition space. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Posters*, page 24, New York, NY, USA. ACM.

[Vaucelle and Ishii, 2007] Vaucelle, C. and Ishii, H. (2007). Interfacing video capture, editing and publication in a tangible environment. In *INTERACT'07: Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction*, pages 1–14, Berlin, Heidelberg. Springer-Verlag.

[Vaucelle and Ishii, 2008] Vaucelle, C. and Ishii, H. (2008). Picture this!: film assembly using toy gestures. In *UbiComp '08: Proceedings of the 10th inter-*

national conference on Ubiquitous computing, pages 350–359, New York, NY, USA. ACM.

[Vaucelle and Ishii, 2009] Vaucelle, C. and Ishii, H. (2009). Play-it-by-eye! collect movies and improvise perspectives with tangible video objects. *Artif. Intell. Eng. Des. Anal. Manuf.*, 23(3):305–316.

[Vaucelle et al., 2009b] Vaucelle, C., Ishii, H., and Paradiso, J. A. (2009b). Cost-effective wearable sensor to detect emf. In *CHI '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 4309–4314, New York, NY, USA. ACM.

[Vaucelle and Jehan, 2002] Vaucelle, C. and Jehan, T. (2002). Dolltalk: a computational toy to enhance children's creativity. In *CHI '02: CHI '02 extended abstracts on Human factors in computing systems*, pages 776–777, New York, NY, USA. ACM.

[Vaucelle et al., 2010] Vaucelle, C., Shada, S., and Jahn, M. (2010). Wow pod. In *CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, pages 4813–4816, New York, NY, USA. ACM.

[VRM, 2008] VRM (2008). My virtual model community. Available at http://www.mvm.com, last acessed 2010-07-26.

[Vygotsky, 1978] Vygotsky, L. (1978). *Mind in society. The development of higher psychological processes.* Harvard University Press., Cambridge, MA.

[Weller et al., 2008] Weller, M. P., Do, E. Y.-L., and Gross, M. D. (2008). Posey: instrumenting a poseable hub and strut construction toy. In *TEI '08: Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages 39–46, New York, NY, USA. ACM.

[Weller et al., 2009] Weller, M. P., Gross, M. D., and Do, E. Y.-L. (2009). Tangible sketching in 3d with posey. In *CHI '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 3193–3198, New York, NY, USA. ACM.

[Whitehurst and Lonigan, 1998] Whitehurst, G. and Lonigan, C. (1998). Child development and emergent literacy. *Child Development*, 69:848–872.

[Wimmer and Perner, 1983] Wimmer, H. and Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13:41–68.

[Winnicott, 1971] Winnicott, D. (1971). *Playing and Reality.* Tavistock Publications, London.

[Woebken, 2008] Woebken, C. (2008). Animal superpowers. Available at http://www.woebken.net/animalsuperpowers.html, last accessed 2010-07-26.

[Yonezawa et al., 2001] Yonezawa, T., Clarkson, B., Yasumura, M., and Mase, K. (2001). Context-aware sensor-doll as a music expression device. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pages 307–308, New York, NY, USA. ACM.

[Ziegler et al., 2005] Ziegler, F., Mitchell, P., and Currie, G. (2005). How does narrative cue children's perspective taking? *Dev Psychol*, 41(1):115–123.

[Zigelbaum et al., 2007] Zigelbaum, J., Horn, M. S., Shaer, O., and Jacob, R. J. K. (2007). The tangible video editor: collaborative video editing with active tokens. In *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 43–46, New York, NY, USA. ACM.