Universidade do Minho
Escola de Engenharia

Pedro do Rêgo Ribeiro

Voodoo: A System that Allows Children to Create Animated Films with Action Figures as Interface

Outubro de 2011

Universidade do Minho
Escola de Engenharia

Pedro do Rêgo Ribeiro

Voodoo: A System that Allows Children to Create Animated Films with Action Figures as Interface

Dissertação de Mestrado
Mestrado em Informática

Trabalho efectuado sob a orientação de
Doutor Manuel João Oliveira Ferreira

Outubro de 2011

DECLARAÇÃO

Nome
Pedro do Rêgo Ribeiro

Endereço electrónico: pedroribeiro@gmail.com Telefone: 917298732 / ____

Número do Bilhete de Identidade: 12621932

Título dissertação ☒ / tese ☐
Voodoo: A System that allows children to create animated films whith action figures as interface

Orientador(es):
Manuel João Oliveira Ferreira, Tod Aharon Iurgel

Ano de conclusão: 2011

Designação do Mestrado ou do Ramo de Conhecimento do Doutoramento:
Mestrado em Informática

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE/TRABALHO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, 28 / 10 / 2011

Assinatura: Pedro do Rego Ribeiro

Acknowledgements

This project was developed at the Centro de Computação Gráfica (CCG). I owe my gratitude to all those people who have made this dissertation possible, to be more particular I want to express my deepest gratitude to the following people:

- My dissertation supervisors, Manuel Ferreira, because I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered, and for the help reviewing this dissertation; Ido Iurgel, for helping me to conceive the initial proposal, for the constant support and guidance, insightful comments and constructive criticisms at different stages of my research, and countless valuable pointers to theoretical and practical approaches, and for the help reviewing this dissertation;

- Several people at CCG, for their instrumental help in many steps of the project: Pedro Rego (Avatar2.5D developer), Joel Magalhães (3D modeler and animator), Nuno Cardoso (Avatar2.5D developer), Gustavo Maçães (Augmented reality developer), and everyone else at the company, for the friendly and helpful environment they provided;

- Who have generously provided their help, including Waldir Pimenta and David Ribeiro for help with TeX and formatting this dissertation and Pedro Soares for the help reviewing this dissertation;

Most importantly, none of this would have been possible without the support, and concern of my family.

Finally, this project would also not have been possible without the support infrastructure provided by CCG, which provided me with all resources to carry on my project.

Abstract

Using any kind of dolls as a tangible interface, has the potential to provide a friendly and easy to learn interface that allows children to control virtual characters in a more intuitive way. The research effort in this domain has been motivated by the shortcomings of conventional interfaces, typically mouse and keyboard, which in this context are neither compelling nor do promote immersion.

This dissertation focuses on the design and evaluation of a system which can interpret the behaviors that children give to a doll in order to provide this behavioral information to the virtual characters. With this system, the user (children) gets the role of movie director, directing virtual characters through this natural form of interaction. This dissertation aims to evaluate the hypothesis that dolls behaviors recognition based on the context of a well-known story, may enhance the ability of children in the creation of an animated film (virtual characters animations). Unlike many approaches that use a direct mapping of the doll movements to the virtual character, it is intended to test the mapping based on the crosses between, user behavioral intention, and the context where the doll it is inserted (the role and the location of the character in the story). The results show that the concept of interaction proposed to empower the children with a way to create animated films is actually very intuitive and easy to use. However, due to the technology used, it was not possible to assess to what extent this concept really empowers children to easily and joyfully create animated films.

Keywords: Virtual Characters, Tangible Interfaces, Affective Interfaces, Toy, Children

Resumo

A utilização de qualquer tipo de boneco como sendo uma interface tangível, tem o potencial de oferecer uma interface fácil de usar e mais amigável. Isto permite as crianças o controlo de personagens virtuais de uma forma mais intuitiva. O esforço de pesquisa nesta área, tem sido motivado pelo facto de os utilizadores não possuírem numa interface convencional uma forma imersiva e intuitiva para interagir.

Esta dissertação foca-se na construção e avaliação de um sistema que consegue interpretar os comportamentos que uma criança dá a um boneco no sentido de animar um personagem virtual. Com este sistema o utilizador (criança) assume o papel de um realizador de cinema, dirigindo e dando ordens aos personagens virtuais através de uma forma de interacção especifica. Nesta dissertação pretende-se avaliar a hipótese de que o reconhecimento dos comportamentos de um boneco baseado no contexto de uma historia conhecida, pode potenciar a capacidade das crianças na interacção e criação de filmes 3D (animação de personagens virtuais). Contrariamente a várias aproximações que usam o mapeamento directo dos movimentos do boneco para o personagem virtual, pretende-se neste trabalho testar o mapeamento baseado no cruzamento entre, a detecção da intenção do utilizador, e o contexto onde o boneco está inserido (o papel e localização do personagem tendo em conta a historia). Os Resultados demonstram que o conceito de interacção proposto para capacitar as crianças com uma forma de criar filmes de animação é realmente muito intuitivo e fácil de utilizar, no entanto devido as tecnologias utilizadas não foi realmente possível avaliar até que ponto este conceito realmente potencia a capacidade das crianças para que de uma forma divertida consigam criar filmes de animação.

Palavras-chave: Personagens virtuais, Interfaces tangíveis, Interfaces "Affective", Brinquedo, Criança.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The doll is a companion to the child, "a close friend," during many years and regardless of gender. Playing with dolls provides children with the opportunity of building imaginary worlds that express their thoughts, feelings and fantasies. While playing, children create stories that are often related to their experiences and to other stories. This universal way of children to tell stories with dolls is the motivation to our work: What if it were possible to empower children not only to tell the stories with their dolls, but also to use the dolls to create some lasting story that is stored in a medium that they can watch and listen to again, and show to peers and parents? If playing with the dolls would automatically result in animations of a virtual character, inserted into a story context? This dissertation aims to result in a system that shall enable this.

[Ishii and Ullmer, 1997] presented Tangible User Interfaces (TUI) as a way that allows people to interact with computers through familiar tangible objects, by guiding the interfaces to the human senses and the natural ability that humans have, to interact with the physical world. This kind of interface has been and remains, a source of inspiration for many information systems.

One of the trends that emerged, relies on the use of Tangible interfaces for the manipulation of virtual characters. This trend has been particularly exploited with children as a target user. The metaphor used for the interface is based on the use of dolls (Figure 1.1), a metaphor that disambiguates the input rather than forcing the user to do so, and at the same time, a friendly and easy to learn interface. To this concept I gave the name *Voodoo*, essentially by its connotation, basically it is possible to realize in a single word most of the concept. In this paradigm, users (children) manipulate some kind of doll via physical gestures, then a computer system detects them, alters the virtual character state, and produces a character animation as feedback.

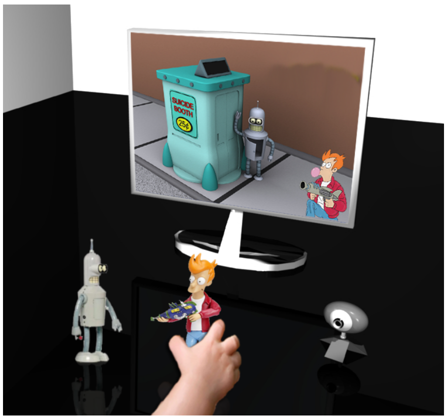

Figure 1.1: Voodoo scheme

1.1 Motivation

Nowadays, there are many tools we use to interact with the virtual world, specifically with virtual characters. We are able, for example, to control virtual characters in a game using a keyboard and mouse, or even via cell phone. However, most of these tools are designed for adults. The creation of an interface geared to children, may involve the use of artifacts with which children are familiar, and augment them with the virtual characters manipulation functionality. The children concentrate on activities that involve manipulating objects around them. Children play using dolls and other accessories to build an imaginary world that express their thoughts and feelings. The creation of a *Voodoo* system to enrich the experience of children when they play with dolls is a great challenge, especially when the aim is to enrich the experience in the pedagogical, recreational and social context.

1.2 Objectives

The main purpose of this project, is to create a *Voodoo* system to support the interaction with virtual characters using the manipulation of dolls, allowing easy and intuitive interaction with virtual characters, in order to maximize the ability of children in creating animated films. The system should allow an extremely simple, intuitive and familiar way to create animated films. The aim in this work is to do a proof of concept, get a first feedback from users regarding the possibility of this concept, making the process of building an animated film, a simpler process.

In this project, the aim is also to analyze and study the strategies that other similar projects have addressed to enhance the experience associated with the manipulation of virtual characters using dolls or similar tangible interfaces. It is equally important to relate these strategies with the model defined in this project. One of the motivations of the project is related to the creation of an intuitive interaction model based on a generic and easily scalable solution. With this assumption, it is imperative to study the usability and interaction models used in similar projects in order to obtain an intuitive interaction system, which enables the reduction of costs of implementation and system maintenance.

This project should be based on the use of virtual characters platforms. The children embody their ideas by creating virtual characters behaviors, through the movements that they offer to the dolls. Therefore, it is necessary to understand the associated experience and if the interaction model created, enhances children's abilities for this activity. It is also an objective of this project the study of the impact, contributions and limitations that the various technologies may offer to the system. Finally, for a better evaluation of the system, with regard to interaction, usability and impact on user experience, it is intended to use a concrete case study. This case study will allow the instantiation of the model developed to a functional prototype.

1.3 Application areas

It is common knowledge that playing with action figures or other types of dolls, is one of children's favorite activities. Adding some kind of interaction with the virtual world, to this activity, can lead to an even more enriching and attractive experience. It is possible to predict what kind of application areas can benefit from a system that enables interaction with virtual characters through dolls. Next is presented a couple of application areas and scenarios for *Voodoo*.

1.3.1 Interactive stories

One of the main areas where *Voodoo* may be employed is interactive storytelling. It is very easy to imagine scenarios in which children, while playing with their dolls, control the unfolding of a story that is being executed and presented by virtual actors.

1.3.2 Directing virtual actors

As explored by [Iurgel et al., 2010] it is possible to apply *Voodoo* so that children can direct an animated film. Children direct the virtual actors through the manipulation of dolls. The dolls could be tracked, and their movements interpreted as instructions to the virtual actor.

The animated films can be inspired by well-known stories, or the story may result from the child's imagination.

Well-known stories The system will work as an assistant providing metrics to create well-known stories like: *The Ant and the Grasshopper*. The virtual actor's behavior depends on the information provided by the child through his interaction and, on the structure of the story. Children could suggest new versions of the story by changing the character's behavior at specific points of the narrative, leading to adaptation on the part of the virtual actor.

Child's imagination In this respect *Voodoo* could offer total creative freedom to the child. In this scenario, the virtual actor follows directly, the manipulation effected by the child on the doll, adjusting his behavior according to their knowledge base and environmental context.

1.3.3 Computer games

With regard to computer games, it is evident the influence that this type of system can have on the gaming experience, especially when referring to very young players. For example, it is easy to imagine a child playing the video game *Sonic* in a scenario where the specific manipulation of the *Sonic* doll, represents a direct or indirect command for the *Sonic* 3D avatar. The doll can surely serve as an input device (interpretation of movements) but not only, the doll may also serve as an output. As suggested by [Johnson et al., 1999] the doll may have *actuators* that will convey the emotional and physical state of the virtual character. For example, *Sonic* could have a mechanism that simulates the frequency of breathing by widening of the doll's thorax. This behavior could very well map out the weariness of the virtual *Sonic*. *Voodoo* may come to be a new paradigm in the field of computer games. A paradigm that can bring many changes, such as, a new and more intuitive way to control the virtual character for young players, or even a way to fade the boundaries between physical and virtual games.

1.3.4 Therapy

It is possible to see the potential of *Voodoo* for therapy when combined with computer games and/or Interactive Story applications. Today, there are several methods of psychotherapy who are already using either computer games or interactive stories to enhance certain treatments, such as those phobias. A system that handles the control of virtual actors by using dolls, can certainly facilitate some treatments, which involve children. A friendly and familiar interface such as a doll can break many barriers associated with stages of treatment. Moreover, a recent study [Mazalek et al., 2009] suggests that if the virtual characters perfectly imitate the movements of the user, this will facilitate both identification and better coordination of the user, with the virtual character. The study also suggests that it is possible, based on common coding theory, that new movements executed by such a 'personalized' virtual character may be transferred back to the user via the perception-action link. Thus improving the user's ability to execute such movements in imagination, and, perhaps, also in the real world. Therefore, virtual characters controlled by dolls can be very important tools for teaching certain movements in fields such as physiotherapy.

1.3.5 Education

In education as in therapy, *Voodoo* can be combined with computer games and/or Interactive Story applications. A good example is a scenario where children interact with an atmosphere of storytelling. The goal is to involve the children emotionally in an improvised virtual drama, in an area of personal and social education. [Figueiredo and Paiva, 2005] have described a system that recreated a bullying scenario to invoke some emotional response in children, in order to educate and prepare them for this social problem.

1.4 Document organization

This document describes the work conducted with the aim to create a system that enhances the ability of children in the creation of an animated film through the use of dolls as interfaces. For this effect, the work was based on the premise that dolls behaviors recognition can be supported on the context of a well-known story. Specifically, was implemented a prototype that embodies this concept in order to obtain a first feedback from users.

The remainder of this document is organized as follows: In Chapter 2, an overview of the main developments in this field of research is provided. Chapter 3 describes in detail the concepts behind the hypothesis that dolls behaviors recognition based on the context of a well-known story, may enhance the ability of children in the creation of an animated film. Chapter 4 describes the first round of participatory design, concretely an informal Wizard of Oz study with two eight years-old participants of both sexes. Chapter 5 presents the work developed in the scope of this research project. Then, Chapter 6 presents some results of the work. Finally, Chapter 7 provides concluding thoughts, summarizing the work outlaid in this document, and setting proposed paths for future enhancements to the approach taken.

Chapter 2

Contextual overview

This chapter surveys the current state-of-the-art. It describes the work performed and, explains the issues and problems encountered when building this kind of system. Additionally, it summarizes the approaches taken so far to overcome these problems. The literature is discussed, with a focus on recent work. However, due to the small amount of work done in the area, some older works are subject of study and discussion.

2.1 Interaction

2.1.1 Distance between interfaces

One of the main features of an interface in the context of *Voodoo*, relates to the distance between input devices and output devices. This means that, depending on the type of task that a user performs with the system, input and output devices may need to be physically distant or, the input device might also serve as the output device. For systems designed to make the user think that the system resides in a single space the distance between interfaces is zero.

To demonstrate this feature, below are described two examples of systems with different distances between their input and output interfaces:

1. the "Fearnot!" [Figueiredo and Paiva, 2005]. Users could manipulate a doll that detects a set of emotions, which the user may want to transmit through a set of gestures. Then, these emotions are mapped to a virtual face, visible on a screen. "Fearnot!" is a system that has input and output interfaces that are perfectly spaced, a doll is used as an input interface and the screen is used as an output interface (Figure 2.1);

2. the "PlayPals!" [Bonanni et al., 2006]. In this system, the users can use the dolls as a communication interface (mobile phone-shaped doll). Basically, in this system, when the child A manipulates his doll, a remote instance doll, that is near child B should move the same way. This system, beyond mapping movements remotely to another doll, also allows for the transmission of sound. The doll speaks and moves remotely. In this system unlike "Fearnot!," the input and output devices are embodied in a single device that is the doll (Figure 2.2).

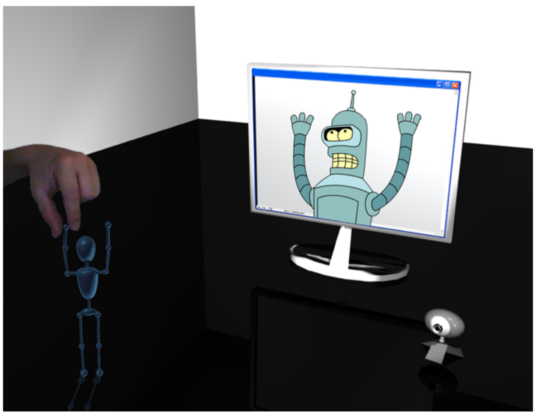

Figure 2.1: System with input and output interfaces perfectly spaced - doll is the input interface and the screen is the output interface

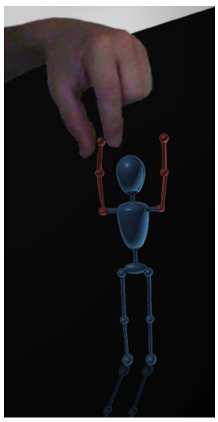

Figure 2.2: System with input and output interfaces integrated in the doll - when the doll's arms are raised they change color

These two are examples of the manipulation of the feature under discussion. However, it is also possible to use a hybrid solution (Figure 2.3), for example: systems that can complement the use of the doll as an input/output device with the use of a screen or projection surface and speakers as output devices.

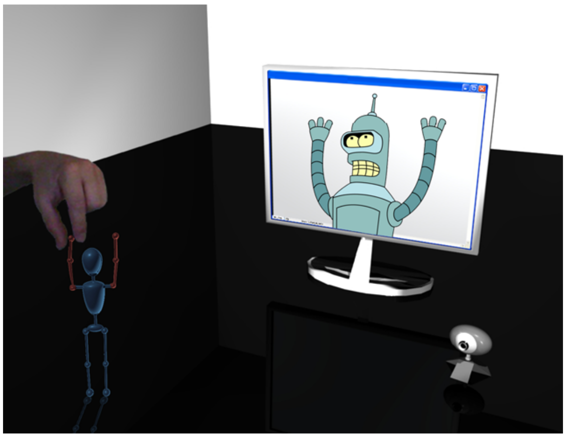

Figure 2.3: Hybrid system - doll is an input and output interface complemented by a screen output

2.1.2 Metaphor

Having a doll as an interface to interact with virtual characters can lead to numerous physical metaphors. There are several features in a doll that can influence and create usage metaphors. Features like the type of doll (action figures, puppetry, plush, etc.), their actions (drop, shake, etc.), shape, temperature, color, size, material, its environment, and others (Figure 2.4). In fact, *Voodoo* is a metaphor, using a physical model of a human being to control a virtual model of one human being. [Fishkin, 2004] grouped the metaphors into two parts, the "metaphor of noun" which appeals to the shape of an object and the "metaphor of verb" which appeals to the movement of an object.

Figure 2.4: Different types of doll, starting from the left we have: Action figures, Puppetry, plush

Starting from the doll metaphor, several other sub-metaphors can be inferred, sub-metaphors that can be of noun, of verb, or both. Below are some metaphors that can be used in *Voodoo*:

1. **the type of doll** - A key feature is the kind of doll, for example, there are such large differences between the assumptions of using a puppet and those of using a plush. A puppet involves the use of handlers to control the body of the doll, while the manipulation of a plush is made directly on the limbs of the doll. Even limiting the focus to the metaphor's underlying puppets, it's easy to find a series of them, like Sock Puppet, Hand or glove puppet, Marionette or "string puppet", Shadow puppet among others. Metaphors associated with the type of puppet, have been exploited in *Voodoo*, for example the "TUI3D" project [Mazalek and Nitsche, 2007] used the metaphor of a string puppet, with the aim of offering a new tool for creating Machinima [Wikipedia, 2011d]. Something that has never been explored in *Voodoo*, was the use of action figures, dolls that personify a character known from fantasy, or real world, such as Action Man, Spiderman or even Barack Obama. Using this type of dolls it is possible to explore a context that is associated with the characters they embody. This can be a very powerful aid when the system has to decide what kind of action the virtual character has to perform. This is so because it reduces the ambiguities that are clearly associated with generic characters;

2. **the actions of the doll** - The most obvious metaphor, concerns the doll's movements that aim to symbolize human movements. It is possible to perform a plethora of actions with the doll, that can be associated with behaviors of virtual characters. From another perspective, there are also numerous virtual actions that can be represented by a physical action on the doll. Actions such as walking, running, jumping, shaking, crying, laughing, etc. These actions may carry two different interpretations depending on the system:

   direct control As explored in [Mazalek and Nitsche, 2007], the system can assume that the motion made on the doll is directly mappable to the virtual character (Figure 2.5);

   intentional control On the other hand, the system can interpret a movement of the doll, as a high-level behavior for the virtual character (Figure 2.6). In the work of [Johnson et al., 1999] this logic was used, the gamers had to control a virtual character through a plush doll. In this system, in order to make the virtual character walk, it is not necessary to grab the plush doll's legs, and put them on the floor one after the other. It is only necessary to move the plush doll, then the system should interpret that action as walking behavior.

In fact, as concluded by [Johnson et al., 1999] intentional control can create navigation problems to people who are more familiar with direct control, like video games, and the solution may lie in something that mixes intentional and direct control.

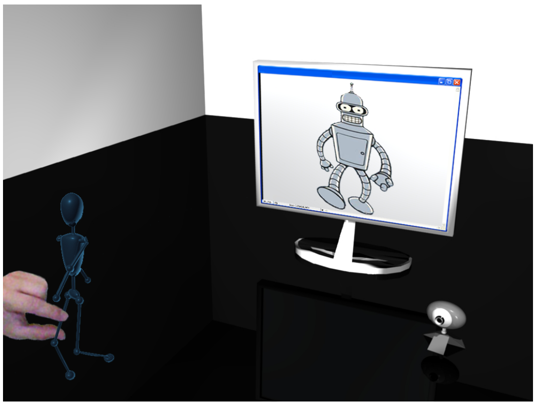

Figure 2.5: Direct mapping - all the body parts positions of the doll are mapped to the body parts of the virtual character

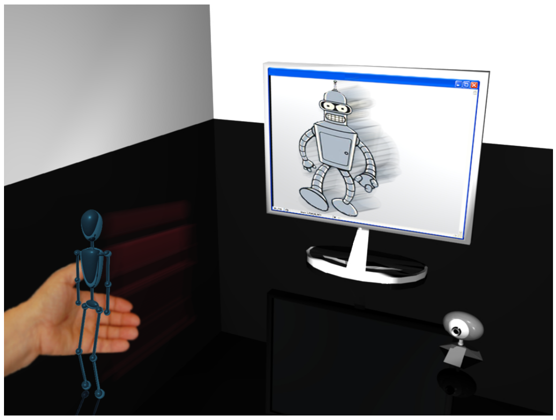

Figure 2.6: Intentional mapping - doll displacement interpreted by virtual character as a walk behavior

Metaphors can become even more powerful if they are the result of the interaction of the doll with its environment. The physical environment where the doll is inserted into may provide a context for defining the behavior of the virtual character. For example, a scenario in which there is a doll of the fictional character Superman and a few crystals of kryptonite. In this context, when the doll is approaching the crystals, the virtual Superman must show weakness. This type of interaction between characters and items from the surrounding environment can also be adopted for interaction between various characters.

2.1.3 Design process

The process of designing a *Voodoo* system aims at finding the best relationship between the actions performed by a user on the doll, and the resulting behaviors of virtual characters presented through output devices. It is possible through this process to find and improve the best metaphors, so that the user can accomplish tasks associated with the purpose of the system [Abras et al., 2004]. It also allows for the acquisition of the best layout in terms of distances between interfaces. Through this process, the designers have the ability to perceive what is the learning capacity of the user. This can be very important in finding the best compromise between what the system should learn from user behavior and, what the user can learn through interaction with the system. For example, the system can be designed so that when the user raises the arms of the doll, the virtual character has a festive behavior. However, in another hypothetical system, this may have to be flexible enough to allow for the parameterization of the virtual character's behavior when the user raises the arms of the doll. Finally, several usability decisions can be made based on this process, decisions about the properties of the doll, such as the type of doll, material with which it is made, dimensions, etc.

This planning process has been used in the design of *Voodoo* systems. For example, the Wizard of Oz method [Kelley, 1983] was used to design "SenToy" [Andersson et al., 2002], a method that is a research experiment in which subjects interact with a computer system that subjects believe to be autonomous but, which is actually being operated or partially operated by an unseen human being. "SenToy" designers resorted to using eight subjects in their tests. In the end, they have succeeded in designing a system that teaches users to determine the correct action on the doll by viewing the reaction of virtual characters. There are many other methods for designing *Voodoo*, simple observation of how children play with dolls, in conjunction with some questions for the acquisition of their intentions during the game, can be enough to design the system.

2.2 Physical construction

The choices made in the aspects of interaction previously mentioned, are the basis for the physical construction of interfaces. There are several variables that must be taken into account in this phase. The vast majority of systems created to date follow the same pattern of interaction, a doll manipulated by a user influences the behavior of a virtual character visible on a screen. Thus, most of the following considerations will be based on this pattern.

2.2.1 Generic vs. specific

One of the first decisions in building the system is attributed to the specificity of the doll i.e.: if it is a generic doll that can be used to impersonate various characters, or if it is a specific doll (action figure) that restricts its activities to a single character. The decision involves choosing, for example, an action figure of Superman to do only and exclusively the role of Superman, or use the generic doll with the ability to represent any role. It is evident the advantage that a generic doll can offer, because with a single doll, it's possible to trigger the activity of several virtual characters. However, this approach also has a disadvantage in the customization of the doll, the virtual character is not like the doll, so the interface is not contextualized, and the metaphor of the object is not so strong. On the other hand, the specific doll (action figure) has the advantage of being able to give better context to the system and user.

2.2.2 Sensing

The sensing is basically the mechanism for capturing the movements of the doll. This mechanism is typically composed by a set of sensors that communicate with the system. Sensing technology is one of the characteristics that most influences the construction and use of the systems.

Outside sensing vs. Inside sensing

To decide which type of sensing to use, it is necessary to choose between a mechanism in which the sensors are out of the doll, or a mechanism in which the sensors are inside the doll. Another possibility would be a hybrid mechanism that combines the use of sensors inside and outside the doll, however, to date, no work in the area has opted for this solution.

Outside sensing Typically when the sensing mechanism is outside the doll, the sensor used is the video camera. The idea is to provide real-time video, to analyze the movements of the doll through computer vision techniques. "Puppet Show" project [Hunt et al., 2006] used an outside sensing solution based on computer vision. The technique they used is called blob tracking, this technique refers to visual modules that are aimed at detecting points and/or regions in the image that are either brighter or darker than the surrounding. These regions could signal the presence of dolls (and relevant environment) or parts of dolls in the image domain with application to doll recognition, and/or doll tracking. Blob tracking is unique in its ability to identify multiple points of interest based solely on the apparent color of the surface. "Puppet Show" leaves a great deal of freedom to the human performers in regards to what exactly they want to use as control elements like colored gloves, t-shirts, any simple colored object can be used to control animations.

Advantages:

- does not require drastic changes in the doll or in its surrounding environment;
- the globalization of video cameras;
- it does not require adaptation by the toy industry;
- is a wireless mechanism.

Disadvantages:

- the area of action is limited to the video camera field of view;
- some techniques are dependent on controlled light conditions;
- doll motion capture is a complex task, due to occlusions made by the hands of users when they manipulate the doll. This problem is avoided, for example, by using the puppet metaphor (there is no direct manipulation);
- activity capture can require large computational resources.

Inside sensing In this case the sensing mechanism is inside the body of the doll. Most *Voodoo* resorts to this mechanism. Unlike the previous approach, capturing the doll's behavior is not dependent of a single sensor (video camera). Inside sensing is based on a network of sensors that are properly placed in the body of the doll, in order to obtain relevant information about the components of the global behavior. These various sensors may have multiple functions, such as obtaining the rotation of the arm, the pressure exerted on a body part of the doll, the temperature of the doll, among many others. The sensor signals are centralized in a microcontroller that does the conversion from analog to digital and, that must be capable of transmitting information to the main computer. Depending on the purpose of the microcontroller and its capabilities, it may perform some data processing, and may have the ability or not to send the information wirelessly. As an example, and just like robots, dolls may even have video cameras in order to recreate their vision. With the "SenToy" doll [Paiva et al., 2002], the creators aimed to infer certain postures of the doll, that enabled the user to demonstrate basic emotions (Intentional control metaphor). For example, putting the doll's hands in front of its eyes means that the desired emotion is fear. To this end, an Inside sensing mechanism was chosen for this project. Basically, a microcontroller has been incorporated into the body of doll, which communicates with the computer via radio. To achieve the previously mentioned example, two magnetic switches were used to detect the movement of putting the hands in front of the eyes, these sensors function via the proximity of the magnetic field. Whenever the user places the hands of the doll in front of its eyes, that gesture is recognized.

Advantages:

- activity capture does not require large computational resources;
- there is a huge amount of local sensors with different functionalities (rotation, power, luminosity, temperature);
- it is possible to integrate a very large network of sensors in a single doll, due to their small sizes.

Disadvantages:

- certain configurations require wired mechanism;
- requires changes in the doll or in its surrounding environment;
- may be vulnerable to certain interference;
- requires adaptation by the toy industry;
- costs.

An alternative scenario that has not been explored, involves the fusion of these two mechanisms, inside and outside sensing. Surely there will be more flexibility when it is necessary to decide which sensors can be embedded in a doll, without creating drastic changes in the typical construction of dolls. Likewise, it is certainly easier to cover a greater range of usage scenarios.

2.2.3 Usability

Typically, tangible interfaces should offer more usability to the users, in controlling a virtual character, a doll interface is more friendly and intuitive than a keyboard. However, it is necessary to ensure that there are no usability problems. To ensure usability, it is necessary to note two important aspects.

Manipulation One of the most important issues is related to the *dimensions* of the doll. If *Inside sensing* is used then, the dimensions will primarily be determined by location and size of the hardware that must be integrated into the doll. However, the handling capacity should never be compromised, even if it means rethinking the hardware options. In the case of *Outside sensing*, the employment of computer vision techniques can also mean that there is a need to parameterize the dimensions of the doll, mainly to avoid the problems of occlusion. Another issue associated with manipulation is the quality of movements. This issue should be mainly influenced by the movement restrictions imposed, by the sensing mechanism.

Appearance In "SenToy" [Paiva et al., 2002], a study was done to see what would be the ideal appearance for the doll. This study focused mainly on features like shape and texture and obtained results that suggest a cuddly doll instead of a hard plastic one is preferable. Thus, the doll's texture should be of a soft fabric, covered with some kind of skin. The study also suggests that users prefer a doll with an unemotional face. Not disregarding the great importance of this study, an ideal scenario, would be one where the users themselves can decide what the appearance of the doll is, unlike the strategy of finding a doll that is perfect for everyone.

2.3 Gesture recognition

Gesture recognition can be seen as a way for computers to begin understanding the body language of a doll. Gesture recognition can be conducted based on a sensing mechanism. As described earlier, this mechanism may operate outside or inside the doll. The gesture recognition makes sense if the system is driven by intentional control. This association of gestures to complex behaviors has been addressed in several studies. The first study [Johnson et al., 1999] used a plush doll as an interface that obtains gestures through interpretation

of sensor data. They used machine learning and gesture recognition techniques to provide the intentional control. During runtime, the system determines the similarity of an input sequence to the set of training sequences (gestures). This study also reports on the use of a context metaphor to minimize the ambiguities and some difficulties related to the need of contemplating the entire range of variations of some gestures, due to the differences in user styles. A special feature in the gesture recognition is the affective computing that is used in the process of identifying emotional expression through computer systems. This was studied and applied in the project "SenToy" [Paiva et al., 2002], where users can perform actions related to emotions through the SenToy doll. That concerns making the virtual character's emotional state change, to exhibit one of six different emotions. Those gestures were chosen as being related to emotions, based on the emotions theory. For example, when the users want to express surprise, they should open SenToy's arms backward, inclining its torso slightly backwards. Rhythm was also explored in this project, the vitality and the intensity of the doll's movement is a way to determine the associated emotion.
Above all, the use of gesture recognition can be very powerful because it is the best model for interaction between human and machine, given that gesture communication is a natural form of communication for human beings. However, it is necessary to take into account certain issues, firstly, users will not behave in a similar way when communicating through the doll as when communicating through their own body. Secondly, users have to be put in a context where they learn how to use the doll, gathering from this, the feedback from their virtual character's reactions, to their manipulation of the doll.

2.4 Behavioral interpretation

The main purpose of interacting with the doll, is in the definition of performance and behavior of the virtual character. To achieve this, certain factors must be taken into account.

2.4.1 Direct control vs. Intentional control

The behaviors of the virtual character depend on the type of control the user has over it. As mentioned previously, control can be intentional or direct. If the system is designed for direct control of behavior, the behaviors system of the virtual character should include the ability to directly map the position of all the body parts of the doll to the body parts of the virtual character. If, for example, the *Sonic* doll's head is tilted then the virtual *Sonic* will also do that. On the other hand, if the system is designed for intentional control, certain complex behaviors must be previously described, so that they can be mapped to a virtual character depending on the intentions of the user. If, for example, a user picks up the *Sonic* doll from one place to another, the system can infer the intention of moving the virtual *Sonic*, activating his running animation.

2.4.2 Virtual character platforms

In most implemented systems, game engines were used for the generation of virtual character performance. In the project "Puppet Show" [Hunt et al., 2006] the choice fell to the Unreal game engine. Technically, the information from doll manipulation was formatted in a single string and sent to an Unreal client through a TCP/ IP connection. This client is then responsible for parsing the information received, and applying the captured movements to virtual characters. This approach enabled the authors to realize, that the control over the characters provided by the game engine is insufficient to create more elaborate performances. The authors' aim was to give users a new interface to create Machinima [Wikipedia, 2011d] and their results were really good but if the system's purpose is the conducting of characters focused on creating animated films, it is expected that there are many limitations in the game engines, because they enhance the action of characters for game situations and not for theatrical performances. Another alternative, still unexplored, is to use virtual actor platforms. A study [Iurgel et al., 2010] explores the use of Virtual Actors (VA) and, explains them to be related, but not identical to, embodied virtual characters. The main capability of a VA lies in its ability to interpret a script (movie or theatre script). A VA platform allows for the direct and intentional control of the VA, because they are capable of controlling an elementary behavior of our body, like raising an eyebrow, and they are also capable of performing a higher level behavior, that may involve a combination of elementary behaviors driven by the personality and knowledge of the actor.

2.5 Considerations

One of the most important issues considered, is the user intention disambiguation. In fact, to get a more accurate result, systems should use all possible information. This includes all sorts of possible context, the character (role) that the doll embodies, the character (role) that interacts with the doll, the environment where the doll is inserted, even emotions that the user is showing, or the interpretation of user speech. This brings us to another issue, the "learning balance", that can oscillate between system and user, so that the system can satisfy the user's intentions the design must surely contemplate the inclusion of the user in the decision process. Systems should not necessarily have a powerful behavior recognition system. If the user knows that one assumption is, for example, indicated by the role of the character, or what his emotional state is, then the system will surely have a much more acceptable performance.

Other issues were addressed, such as the inclusion of actuators on the doll, in order to reduce the boundary between virtual and digital, the problematic of the doll as an element of distraction, among others. There are many interesting areas for future work, such as:

- research and find new usage metaphors, in order to promote the usability of systems, and to reduce the technological complexity in behavior recognition;
- the construction of the doll as an interface of input and output;
- the construction of *Voodoo*, focusing on usability and immutability of any doll on the market;
- research and apply behavior recognition technology, based on computer vision or sensor networks, or even the integration of these two streams;
- explore the integration and research of systems that render virtual character behavior, which have sufficient control points to create complex behaviors.

Chapter 3

Methodology

This research project aims to investigate a new process for creating animated films, based on a new form of interaction, which can maximize the ability of children in this task. Here were discussed and analyzed *Voodoo* systems. As any system that uses tangible interfaces, *Voodoo* is a recent exploitation and, for that reason, it is still faced with many challenges, either by the associated usage assumptions or by the technology needs that it presents.

Some issues associated with the design of such systems were addressed, with the aim of showing possible directions on how to deal with them. Two major open issues that our system shall help clarifying are "disambiguation" and "learning balance."

The first major open issue, disambiguation, refers to the necessity of disambiguating underspecified input: Assuming that the animation system is capable of producing a vast amount of different, fine grained animations that much outnumbers the possible movements that a child can be sensibly expected to make: How can then the doll be, nevertheless employed to create, in a controlled manner, these animations (or a large subset of them)? For example, a system may have the capacity to animate a Little Red Riding Hood (LRRH) virtual character that is suspiciously talking to the Wolf. With plastic action figures, it is impossible to deduce from the movement of the LRRH figure alone that LRRH is "talking," or even that this talking is "suspicious." In general, the movements of a doll tangible interface are ambiguous, and disambiguation is a major issue of such systems (cf. [Ribeiro et al., 2011]). Understanding how to best disambiguate input is the main focus of our work.

The second major issue of doll interfaces refers to the learning balance between system and user. This is the question as to who is expected to adapt more in order to enable a fluid interaction – the user or the system? Doll based systems need not necessarily have a very powerful, adaptive system for recognizing the doll's movements. If for instance the user knows that the system will take into account contextual information about the narrative role of the animated character, or about its current emotional state in the story, then he/she can adapt the doll movements to the expected interpretation, thus easing the recognition and interpretation task (cf. [Ribeiro et al., 2011]).

3.1 Integrating Story Context into Input Interpretation

In order to cope with the problem of ambiguity and underspecification, the system needs to recur to additional context information. This includes (i) the story role that the doll embodies, the (ii) relationship between the story persons, and the (iii) narrative environment of the animation (cf. [Ribeiro et al., 2011]). The concept to be tested is based on the *Ecological approach* that was conceived by [Gibson, 1979], which refers to the study of the interaction between humans and the environment that surrounds it. [Rogers, 2004] notes that several HCI researchers adopted this approach in order to understand how people interact with artifacts. It also states that one of the concepts underlying the *Ecological approach* has been particularly used, namely the *Affordance*. In practice this concept in HCI has been used to categorize the characteristics of objects that let people know how to use them, for example, a spherical doorknob suggests that it must be rotated (Figure 3.1). This concept is used in the tangible interface that the doll represents, especially in regard to action figures that have an associated context. For example, an action figure that embodies the character *Jack Bauer* from TV series *24*, can influence the action figure manipulation. If a child imagines a scene where the character *Jack Bauer*, have to open a door in a stressful situation, probably *Jack Bauer* does not rotate the doorknob (Figure 3.2).

Figure 3.1: Affordance associated to a door knob

Figure 3.2: Jack Bauer Action figure opens a door.© clarktoys.com.

This example of *Jack Bauer* leads us to *Ecological constraints* which is another concept behind *Ecological approach*, it suggests that the actions of a particular person are determined by the environment and not by an internal cognitive process. *Ecological constraints* is the key concept in the approach outlined for this project, and therefore, it is intended to map this concept to the doll. In other words, as the action of a human being is influenced by its surrounding environment, the behaviors of a virtual character should be determined by the environment where the doll is inserted (cf. [Ribeiro et al., 2011]).
How to apply this concept of *Ecological constraints* in this case study? The environment of a particular doll will be influenced by the context associated with a well-known story. The story will influence the information associated with a character or a specific environment in story, or even the information associated with the personal relationship that exists between characters (cf. [Ribeiro et al., 2011]). It is possible to map this logic to a concrete example, as the story of the *Hulk* created by *Marvel Comics*, there are several behavioral information that can be extracted from the character *Bruce Banner*, regarding his interaction with other characters of the story and the surrounding environment. It is known that *Bruce Banner* when angry, becomes physically and psychologically a wild and powerful creature named *Hulk*, and this transformation is more likely when it is with *General Ross*, his enemy (Figure 3.3). *Bruce Banner* takes a more affectionate behavior when it is close to *Betty Ross*, because of their personal relationship (Figure 3.4). Finally, with regard to the environment, *Bruce Banner*

behavioral change can easily happen when it is enclosed in a prison where he will likely turn into the *Hulk* (Figure 3.5).

Figure 3.3: Hulk's behavior influenced by the proximity of general Ross

Figure 3.4: Hulk's behavior influenced by the proximity of Betty Ross

3.2 Solution

As a solution to the approach presented, it is necessary to reinforce the importance of the usage metaphors that can become even more powerful if they are the result of the interaction of the action figure (associated with a character from a well-known story) with its environment (well-known story places). It is proposed the use of a physical map (based on the story) where the action figures can be placed (Figure 3.5). It is anticipated the use of computer vision techniques to determine in real time the environments that surrounds the action figures, the positional relationships between them, as well as the type of movements they are performing. Basing the sensing mechanism in computer vision, allows to avoid drastic changes in the construction of the actions figures, and it offers the advantage of being based in one of the most widely used sensors in the world. It is believed that this information can be supplemented by information coming directly from user behaviors, such as their emotions and speech. Unlike the approaches that use a direct mapping [Mazalek and Nitsche, 2007] [Mazalek et al., 2009] of the doll movements to the virtual character, it is suggested this mapping based on the crosses between, user behavioral intention [Johnson et al., 1999], and the context where an action figure it is inserted.

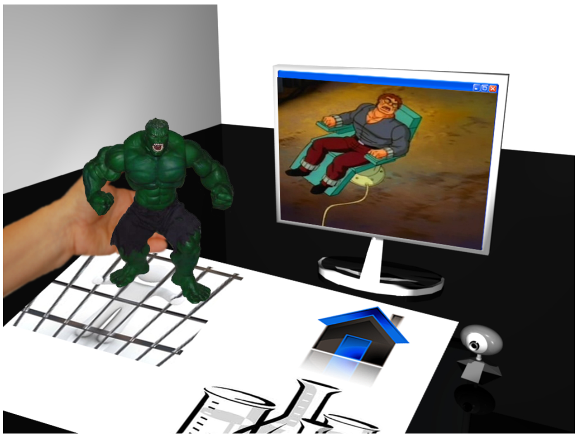

Figure 3.5: Hulk's behavior influenced by the surrounding environment - The positioning of the Hulk action figure on the map influences the behavior of the virtual character on the screen

In order to enable the validation of this theoretical proposal, it is intended to conduct a test case. Specifically, a test case based on the very particular context of the well-known story LRRH. It aims to describe extensively all the preponderant aspects of the system and to isolate the problem. In this dissertation supporting multiple stories must be avoided, because it means a lot of information in the system, such as characteristics of the characters and surroundings, or 3D models for virtual characters and environments.

The next chapter focuses on the process of designing the system. The aim is to find the best mechanism of interaction between the user and the system, based on all the concepts that are intended to be implemented.

Chapter 4

User interface designing

This section describes the planning and execution of the design process, which has already been theoretically discussed in Section 2.1. This section describes two phases: the design of the prototype, and the testing/adaptation of the prototype. The first phase describes the preparation of the test plan, the selection of test participants, the tasks they will have to perform, the formulation of a basic prototype and the installation of the test setup. The second phase is about testing the prototype and adapting it to the needs that resulted from the test. The chapter concludes with the evaluation.

4.1 Design

Defining and maping the user interface usability to a product at an early stage of the project can bring great benefits. The user and their needs are taken into account from the beginning, and the final solution will be the result of these needs. The design process has the purpose of defining solutions to the needs associated with tasks that users have to achieve in a given context. For the design process of the user interface there are methods that have the clear purpose of obtaining information about users, their needs and their tasks, in order to allow the definition of design solutions [Abras et al., 2004]. To define a design solution for this project it was used the Wizard of Oz technique (WOZ). This technique can be used for different purposes, which makes it a versatile instrument not only in Voodoo systems development, but in computer systems development in general. It should be noted that this project does not aim to make an exhaustive interaction design study, rigorous testing would require a high investment. In that sense this study will have the primary purpose of obtaining a first feedback regarding the interaction in order to support the proof of concept that is the main objective of this project.

4.2 Wizard of Oz technique

In short, this technique is an user-based evaluation of an unimplemented technology that is generally unknown to the user, and where a human (Wizard) simulates the feedbacks of the system (Figure 4.1). This technique has often been used to explore design and usability with applications based on natural language, command languages, imaging systems, and ubiquitous computing systems.

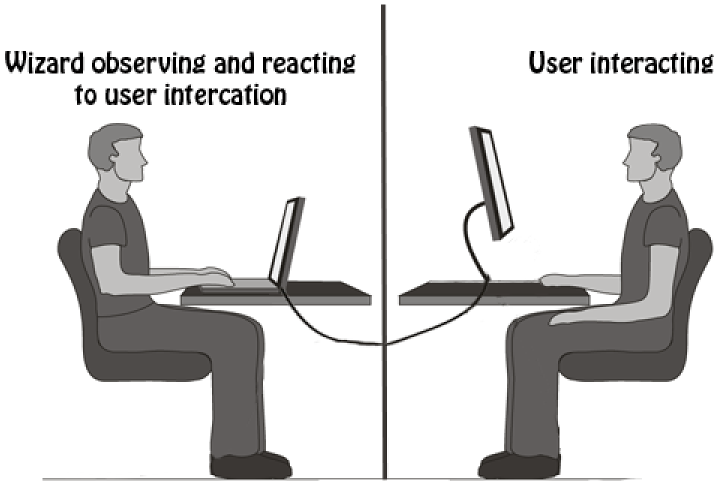

Figure 4.1: WOZ technique scheme

4.2.1 Benefits

The *WOZ* technique can bring benefits such as:

- obtain human responses, about a non-existent human computer interaction;
- test the interaction of a device before building a functional system;

- test design of feedback through output technologies;
- test which input techniques and sensing mechanisms best represent the interaction;
- test heuristics to determine how to create outputs for ambiguous human inputs;
- discover the problems people will have with the techniques and devices;
- investigate aspects of the objects of the system such as exploring the visual affordance (whether the product shows how it can be used).

4.2.2 Disadvantages

- the wizard should be well trained in order to respond in a credible way to the user interaction;
- the training of Wizard is an additional cost;
- if the Wizard is a research member there is a risk that he will improvise beyond the programmed behavior;
- wizard needs to match how a computer might respond, Wizard cannot make errors;
- the wizard may tire over the test sessions, and his reaction time may increase;
- it is very difficult to simulate/evaluate a system where there are many interface elements.

4.3 Designing a Wizard of Oz Study

Designing a *WOZ* study has a major difficulty: the system being investigated is basically unknown. It is necessary to know how children would actually interact with a Voodoo system to be able to build it, but it is impossible to find out that without running a Voodoo system. Therefore, instead of a non-existing system, a prototype is used to only simulate the desired system. As [Nielsen, 1993] suggest, ideally the prototype's performance converges with the desired system's performance after a number of interactions, resulting in a running system.

4.3.1 Test Plan - Development

Study Goals - Definition

In this work it is intended to conduct a case study based on the context of the well-known story LRRH. It is intended to use this case study for the design of the system in order to cover up and study the various aspects of the interaction. As already stated, the aim is to design a system that enables interaction with virtual characters through action figures manipulation. In the Section 3.2 it was defined a high-level approach to the interaction. Essentially this approach focuses on the use of a physical map that contextualizes the action figures. In the first phase of the *WOZ* study is necessary to decide the study use case and make the design based on that decision. Were identified three main case studies which are certainly important:

User Acceptance Studies Create and present to a group of future users a low fidelity prototype simulating an existing technology in order to determine the possible future success of the system. The aim is to evaluate the acceptability of the system through analysis of a survey, and through direct observation of the participant's reaction.

Explorative Dialogue Development The aim is to test the dialogue between computer and human in order to design a new dialogue system based on the results obtained. This approach can be helpful for systems aimed at exploring new forms of interaction or building applications with complex interaction. It is intended to assess aspects of interaction of the solution presented to participants, through a survey and direct observation of the participant's behaviors.

Data Collection Used to collect data from user interaction such as comments, facial expressions, gestures. This collection may be essential to post-analyze the behavior of the user when he performed a certain task. The data record from the experience will be saved using audio visual recording for later analysis.

Type of experiment

One of the decisions to be taken in a *WOZ* study has to do with the control of the experiment. [Bernsen, 1998] presented two possible types of control in a *WOZ* study. The first type is entitled: Controlled experience. In this type, the experiment is performed in a controlled environment, this means that whoever will be tested should be placed in a laboratory environment. This serves to study a system that requires devices that are unique or inaccessible to most people. The second type is predictably entitled: Uncontrolled experience. In this case the experiment is performed in an environment that is familiar to test participants. It is mainly indicated when the hardware requirements are minimal. This type of experiment has the great advantage of not constraining test participants enabling more reliable results. Because no special equipment will be needed to perform the experiment, it was decided that the kind of experiences that best fits, is the Uncontrolled experience. The test participants are children, and given the characteristics of the testing activities ("play with dolls"), it is assumed that it is not appropriate to remove children from their environment, because it can corrupt the experience and reliability of results.

What is intended to measure

The variables that must be studied in the experiment are part of the decisions in the planning phase of the *WOZ* study. In this study the variables to evaluate were divided into three categories:

Assumptions about the participant These variables exist to check the basic knowledge of participants. This knowledge can directly influence the experience. It is intended to know if the participant:

- has knowledge of new technologies;
- knows the well-known story LRRH;
- usually plays with *action figures*;
- wants to use his *action figures* or the personalized action figures of the story;
- is very active.

User experience These variables are used to assess the using experience of the system. It is intended to verify if the participant:

- Understood the story.
- shows confusion
- shows happiness
- shows frustration
- easily associates the *action figures* to the characters of the story
- can clearly see the contexts behind the map and the relationship between the map and story
- think the map is ambiguous

User interaction These variables serve to evaluate aspects of user interface of the system. It is intended to check if the system has:

- ideal screen size, map size and *action figure* size;
- ideal relation between *action figure* size and map size;
- ideal number of contexts in the map;
- ideal distance between interfaces;
- percentage of time looking at screen and percentage of time looking at *action figure*;
- embedded webcam in laptop screen showing good results;
- external and adjustable webcam showing good results.

4.3.2 Test participants - Selection

As mentioned previously, the interaction design is user oriented, this means that test participant are the source of information that feeds the interaction design. It is therefore important to carefully select the participants.

Target Group

The target group identified is composed by children aged between six and ten years-old female and male. This age range coincides with the ages at which children typically play with dolls. It is understood that there is no need that children have to be familiar with new technologies.

Number of Participants

The choice of the number of participants was influenced by the aforementioned reason that there is no purpose to do an extensive study and the fact that an exploratory study does not require a lot of participants. Given these two reasons it was decided to perform an informal *WOZ* study with two eight years-old participants of both sexes.

Test Facilitator

The test Facilitator serves as a bridge to the test participants, is the person who shall transmit the test information to participants. The Facilitator should be familiar with the children or should somehow be able to place the children in a comfortable relaxed environment. To fulfill this goal it was necessary to resort to the aunt of the participating children as facilitator.

4.3.3 Tasks - Definition

The tasks to be performed by children throughout the test have the primary purpose of enabling the evaluation of the system usability. The tasks set for this study should represent tasks that can be carried through the final system. Since there is no intention of conducting an extensive study, it is considered inappropriate to devise a test where the child may actually have the experience of building an animated film based on a story and using an action figure. Such consideration is made because it is complex to create a prototype with a real time controller of virtual characters, and even more complex is to have a trained Wizard that has the ability to perceive the intention of the child and enable the virtual character behaviors just in time. Alternatively, it was decided to devise a similar task that offers a similar experience to the user. In this sense, it was chosen the task of controlling the display of the well-known story of LRRH. In other words, it is the task of navigating through a story, this Implies for example: the manipulation of the action figure in a specific context of the map; enabling the display of parts of the story by putting the action figures needed in the context of the correct map.

4.3.4 Basic Prototype - Creation

It is known that the prototype should be as close as possible to system that it is intended to build. Because this dissertation has the sole purpose of proving the concept described in the Chapter 3, it is intended to build a single-user system, in which children are aged between 6 and 10 years, who can play with up to two action figures at the same time, one in each hand. This does not mean that the collaborative usage aspect in the construction of the system it is ignored, since it is known that in most cases children play together. For the prototype it was used exactly the same type of interaction that is intended for the final system but with a slightly different purpose. The prototype does not allow the real-time construction of an animated film based on the story of LRRH. Instead, the prototype allows the user to exploit the story of LRRH, using the actions figures (featuring the characters in the story) placing these in context through a DIN A4 paper map of the story (Figure 4.2).

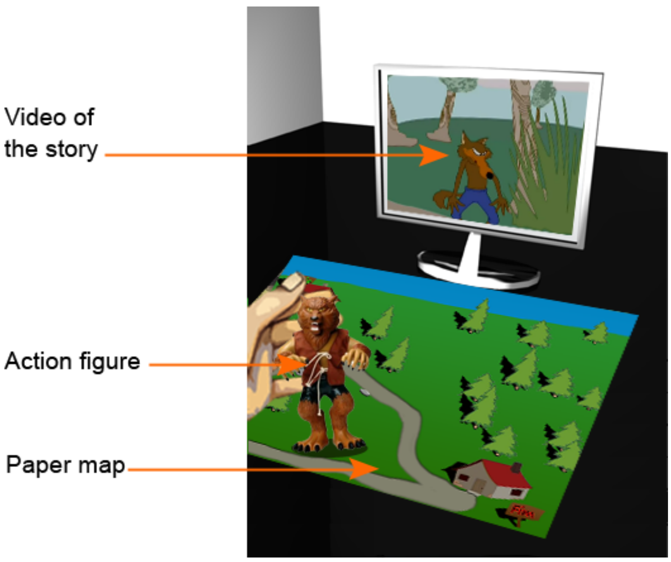

Figure 4.2: Prototype system scheme

In practice, the test participant puts an action figure on the map in the same way that puts a piece in a board game, then the system infers the context of the action figure and plays a movie with the associated part of the story. It should be noted that an extension of the system to allow for collaborative use is easily possible; for this, the size of the paper map and action figures must be increased, and normally larger monitors or a projection would be required to compensate for the increase in viewing distance.
This prototype is divided into two blocks: the configuration block and the interaction block. The configuration block provides typical interfaces (keyboard and mouse) and aims to set up the experience that will result from the interaction with the system. This block was divided into five phases:

- connect a webcam (Figure 4.3);
- choice of the story in which it is intended to navigate (Figure 4.4);
- printing the paper map associated with the story chosen (Figure 4.5);
- adjust the position and orientation of the paper map and webcam (Figure 4.6);
- associate action figures with the story characters (Figure 4.7).

Figure 4.3: Connect Webcam

Figure 4.4: Choice of story

Figure 4.5: Print paper map

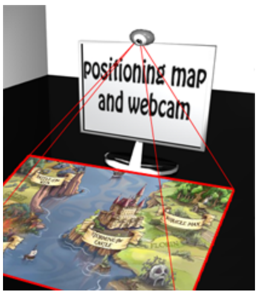

Figure 4.6: Positioning of map and webcam

Figure 4.7: Associate action figure with character

The interaction block is enabled by the result of the configuration block. In this block, the test participant can then abandon the typical interfaces and rely on action figures manipulation in order to explore the story that was chosen in the configuration block (Figure 4.8) .

Figure 4.8: Navigate in the story

To create the prototype (Figure 4.9) it was used a prototyping tool called ForeUI [Foreui]. Through this tool has been possible to quickly and easily build the entire prototype. In the construction of the prototype, all the processes described above have been automated. In practice, all events were programmed for the typical mouse and keyboard interfaces. The Wizard can thus interact with the prototype and dodge the test participants. These think they have control over the system through the action figure, when in reality, it is actually the wizard that analyzes the behavior of the test participant and display parts of the story with the keyboard events.

Figure 4.9: Prototype screenshot - Story exploration phase

To enable the story navigation process it was used the video-sharing website Youtube [Youtube, b]. In practice, given the story that the test participant choose is necessary to have a Youtube video to represent that story. Having the video of the story is necessary to subdivide it in order to associate the display of a sub-video to the event: positioning of an action figure in a particular subarea of the paper map. In the construction of this prototype, it was chosen a Youtube video related to the story of LRRH [Youtube, a]. Then, it was made the analysis, subdivision of the video, and automation of navigation through these sub-videos, using keyboard events.
Besides the story navigation programming, were also programed events to guide the test participant, events with the aim of give more control over the application. Thus were created the following events to inform test participants about the navigation state:

- action figure in a context where there is no story;
- missing action figure so that the story can continue;
- the story continues;
- end of story.

4.4 Evaluation

For the testing phase, the main objective was to allow the evaluation of the variables which have already been discussed in Section 4.3.1. Given these variables, the configuration block (choice of story, map printing, etc.) stays out of the tests. Then, the test focuses entirely on the interaction block, specifically, the task: exploration of story through the interaction based on action figures manipulation and the context of the well-known story LRRH.
The first step in order to enable the WOZ study involves the construction of the prototype system (Figure 4.1). The construction was made in a place known by the test participants in a quick and unobtrusive way. This construction involved the use of a laptop where it was installed the software prototype. This laptop was equipped with a webcam for recording video of the whole interaction made by the test participant. The control of events by the Wizard was enabled by using a wireless keyboard. Finally, were also placed in front of the monitor, the paper map associated with the story of LRRH and the action figures associated with the LRRH and the Wolf characters.

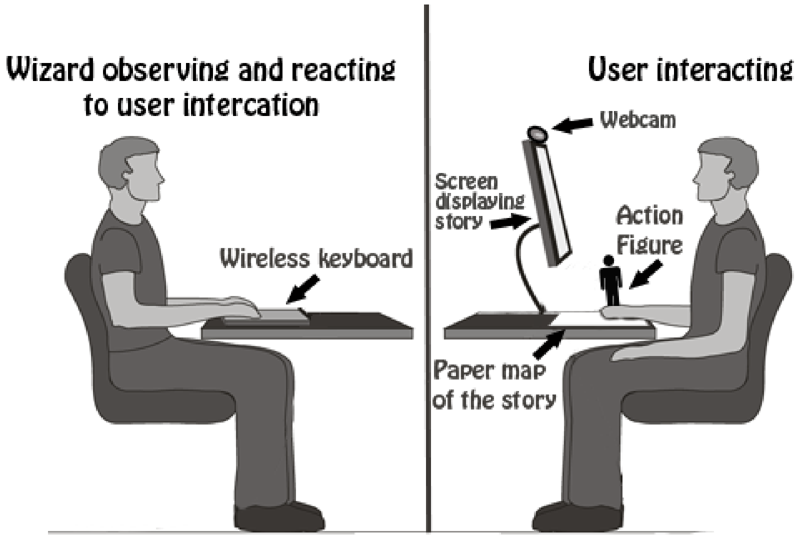

Figure 4.10: WOZ study system scheme

Alongside the construction of the prototype system, the facilitator gave a short background of the purpose of the system to test participants. The facilitator also took this step to get answers regarding the assumptions associated with the children. Assumptions that result from questions such as: Test participant already knows the story of LRRH? Do they often play with action figures? Among other assumptions that were discussed earlier in Section 4.3.1.

Once the prototype and the study participants (Wizard, test participants, facilitator) were ready, it was then possible to perform the tests (Figure 4.11) that ran without any mishap.

Figure 4.11: WOZ study test participant

At the end of the test, the facilitator conducted a survey in order to get user experience and usability answers from the test participants. Moreover, an observation of the children behavior was made, through direct observation and later by the video recording analysis. The (Appendix A) presents the complete survey result.

From this informal *WOZ* study, which aimed to make a qualitative analysis of the prototype system, it was obtained the encouraging result that the kids had no usability problems and generally showed interest and pleasure in using the system. Specifically, it was learned through this study that children without any specific type of learning do not show any difficulty using the system. A curiosity that emerged from this study is related to the fact that the children tested have no experience with information technologies. This could hypothetically influence its easy adaptation to the system. A test participant who already used some kind of information technology could be forced to an extra learning effort, because he already had interaction assumptions, in relation to information systems ("I was not supposed to use a mouse or keyboard instead of an action figure?"). However, it is an objective, check this issue in the final system, involving for that, users with some experience in information technology.

There were even some unexpected behavior as the fact that they manipulate simultaneously more than one action figure in different contexts, namely when the LRRH and the wolf take different paths. This behavior raises some questions: What should be the display when there are parallel actions? What should the system show if the action figures of the wolf and LRRH are moved simultaneously? The system should attempt to infer the most important action? Must the system use the filmmaking technique *Split screen* [Wikipedia, 2011h] popularized in the TV series24 (Figure 4.12)? Should be noted that although these issues are not expected in the single usage, they are too obvious when considering the use of this system in a collaborative way.

Figure 4.12: Split screen technique

Finally, through the construction of the prototype it was learned that a rigid laptop webcam would not be appropriate if the animation were to be watched in the laptop's screen, since the camera needs an inclination of approximately 45 degrees to the paper map, and this would hide the screen.

The next chapter will describe the implementation of the final system. An implementation totally based on guidelines resulting from this informal *WOZ* study, and with the main objective of validating the theoretical proposal described in the previous chapter Chapter 3.

Chapter 5

Implementation

I named the system implemented in this project Voodoo. I decided that the name of the system should be homonymous with the name of the generic concept explained in Chapter 2. Voodoo allows children to take over the role of a movie director, animating virtual characters with action figures. It allows the choice of the story basis and the association between action figures and virtual characters that the child wants to use. Voodoo system is therefore, a generic system that offers flexibility in the management of content associated with the stories, allowing the insertion of new stories so that users can enjoy a wider range of contexts. Once the system is manageable, the manager of the system becomes a new agent of the system. Thus, the logical architecture (Figure 5.1) considers user and manager as main actors.

Specifically, the logic architecture processes consists on:

1. The management of system basis data Section 5.5;
2. The sensing / interpretation of user activity Section 5.2;
3. The performance of virtual characters Section 5.3.

5.1 Physical Architecture

Below (see the Figure 5.2) is presented a diagram that shows the physical architecture of the system. The architecture is based on client-server model and it was implemented in the prototype of this project.

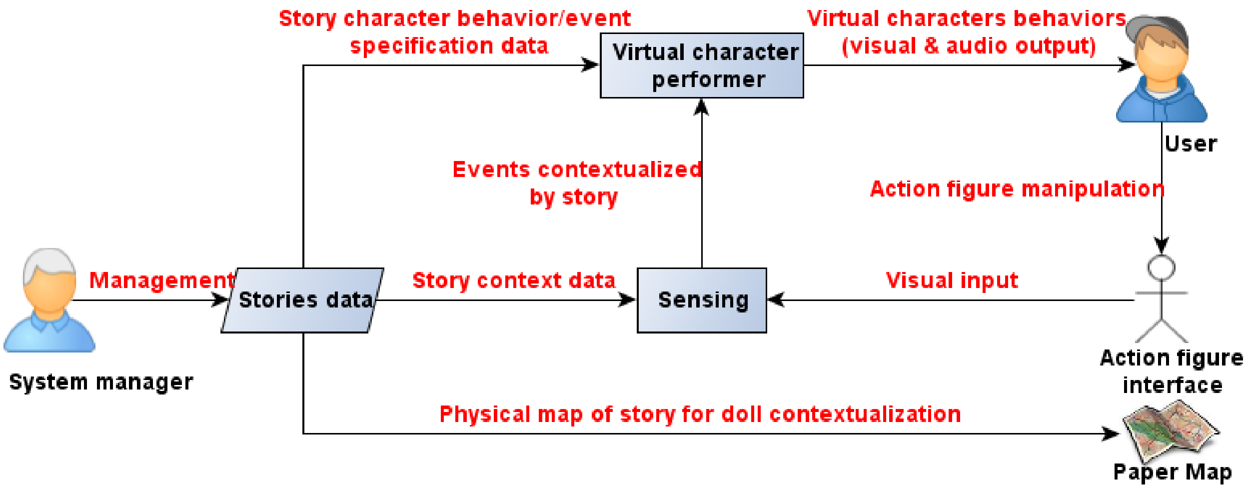

Figure 5.1: Logic architecture

Voodoo web server module is responsible for providing the Voodoo system services to public users and system managers via the internet. These services are represented in the modules Voodoo Client and Voodoo Story Manager which are discussed below. Voodoo client module consists of a computer, a screen, a webcam, the actions figures, and the paper map of the story. This module is responsible for the system configuration, it allows a user to choose the story, print the paper map and choose the actions figures. In addition, this module is responsible for sensing and interpretation of the interaction made by the user through the action figures and paper map, and for generating virtual characters behaviors. The target user of the Story Voodoo manager module is the system manager, this module includes tools for creating and editing stories, introduce 3D character models, and introduce new map images. This module must communicate directly with the Voodoo Web server module to update the content that should be made available to end users (children).

Note that the main development target is to create the Voodoo client module. This should allow the animated films creation using the proposed methodology. However, an effort was made to conceive a system that allows easy manipulation of content (new stories, new maps, etc.), and global access to this new way of creating animated films. The strategy of creating a standalone application was perceived as a successful strategy for the performance and flexibility of the application. However, it is believed that it would not be the best direction to make the system highly and easily accessible. Thus, the strategy was to create a web-based application to ensure global access and an easy system management.

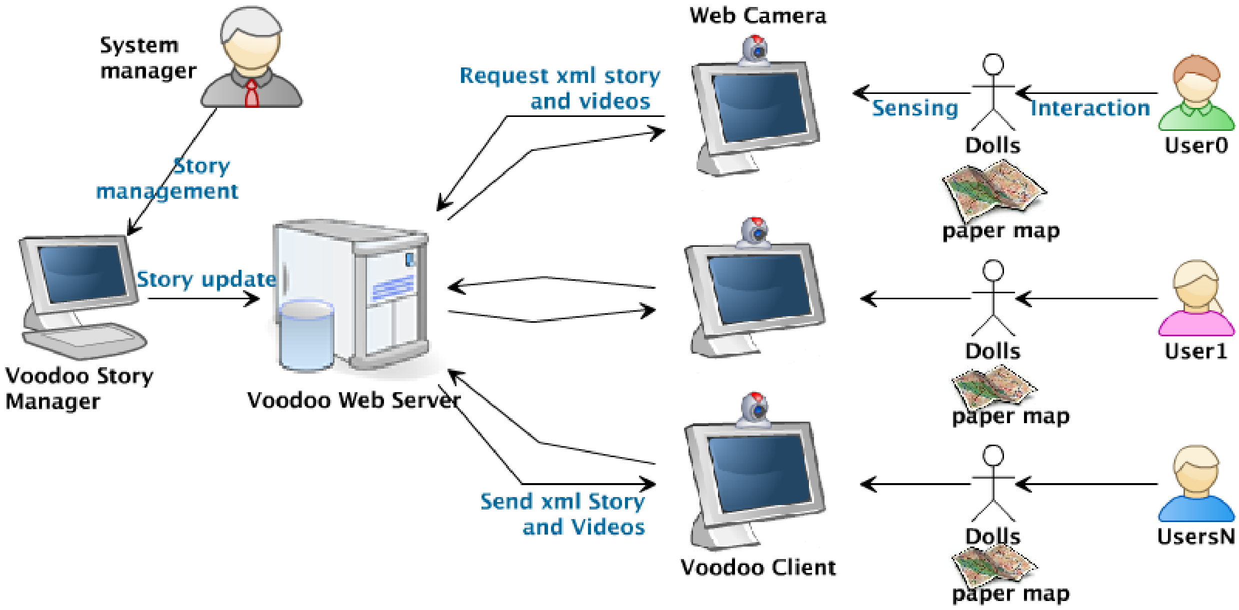

Figure 5.2: Physical architecture

5.2 Sensing

Voodoo relies on a computer vision sensing module (CVSM) to determine in real time the environment that surrounds the action figures, the spatial relationships between them, as well as the type of movements the child is making with them. The CVSM is based on a color blob tracking of the actions figures and on the detection of visual patterns printed on a paper map. These visual patterns allow defining areas of the map; the areas correspond to story contexts, e.g. the house of the grandmother. The localization of action figures into some area of the map is detected through their color blob points. The movements of the action figures above some areas of the map generate in real-time specific animations of virtual characters on the screen. To recognize movements of the action figures, it was implemented a simple movement binarization into two categories: vertical gestures, and horizontal gestures; additionally, the speed of the gestures is calculated. Based on these parameters from computer vision alone, all the further animation details are determined by the aforementioned story contexts: the action figure that was chosen, the area of the map into which the action figure is inserted, and the typical narrative relationship between the figures that are in this same area. Therefore, only two types of gestures plus velocity result in many different animations of the virtual characters on the screen. Technologically this solution involved the construction of a bimodal Adobe Flash [Wikipedia, 2011a] application

(Appendix D). The sensing configuration mode and the event generation mode.

5.2.1 Sensing configuration

The specification of the story chosen by the user is the basis for the sensing module configuration. This specification is contained in a xml file (Appendix F), that specifies all the stories in which users may base their interaction and creation. This file has, for each story, the following information:

1. story Id;
2. story name;
3. path to the image file that symbolizes the story;
4. path to the image file of the story map;
5. path to the visual patterns files;
6. definition of the story contexts (relative position to the ARToolKit visual patterns);
7. story characters;
8. path to the video graphs of the story.

The first step taken by the sensing module is to parse the story specification chosen by the user as a basis for his interaction. The parser creates a *Doll* object for each character specified in the story. The *Doll* object aims to represent the relationship between the action figure and character. Therefore the *Doll* object stores the information of this relationship as the character's name, the action figure position and movement, and events that it triggers. On the other hand, and also as a result of the parsing process, are created for each context of story, the *Context* objects that are responsible for keeping the information about the contexts such as name and position. In order to define the position of the story contexts and update it in the *Context* objects, it was used ARToolKit technology [HITLab] because it allows to obtain real-time 3D position and orientation of various visual patterns that are present in a video frame. Thus, the 3D positions of the story contexts are related to the 3D positions of the ARToolKit visual patterns (Figure 5.3). In fact the main reasons for the choice of Adobe Flash technology was: the fact that the Flash content may be displayed on various computer systems and devices, using the well known Adobe Flash Player; the second reason has to do with the fact that ARToolkit was ported to Adobe Flash - FLARToolkit [Saqoosha, 2009].

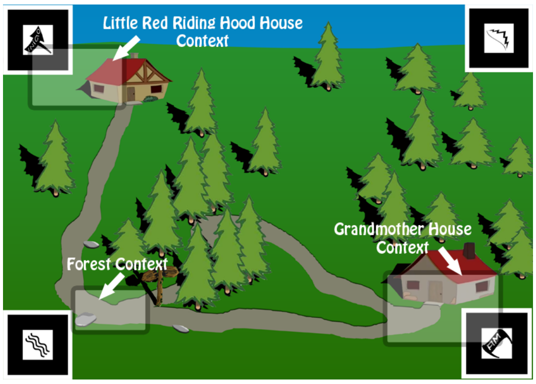

Figure 5.3: Get context position from ARToolKit marker position

Having the story contexts positions, the sensing module divides the screen vertically, according to the number of story characters, the aim is to allow the user to associate the color of an action figure with a story character (Figure 5.4) .

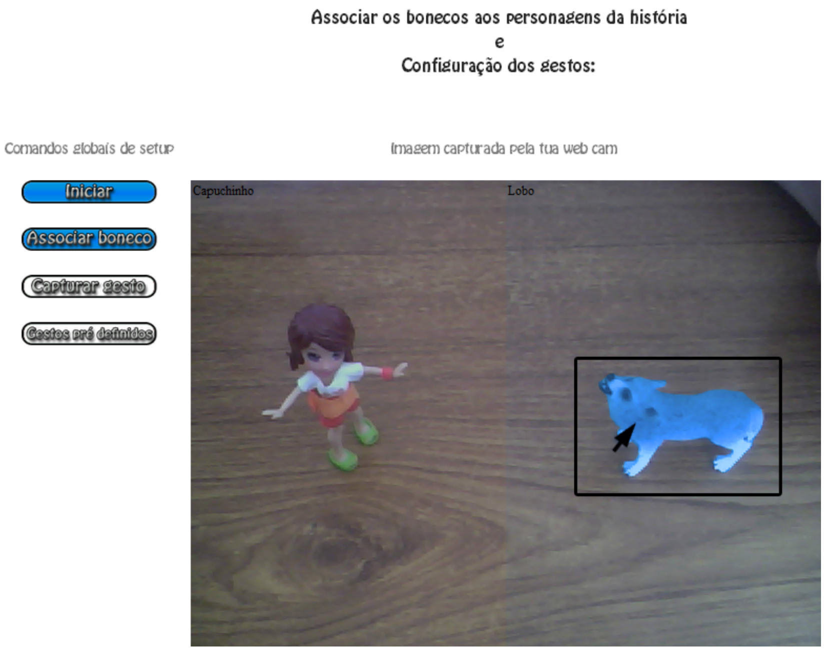

Figure 5.4: Voodoo screenshot - pick and associate the action figure color with the wolf character

From this moment the sensing module only needs to know the relative position between actions figures and the story contexts, and the type of movements of the action figure. The goal is to update the *Doll* objects with this information. When the system encounters an area of a color associated with an action figure, it infers the position (blob position) of the action figure from one point belonging to this area. This point should be centered horizontally but should be as low as possible on the vertical axis. In the picture below(Figure 5.5), the position of the action figure (green) is found between two ARToolKit patterns. Due to the webcam perspective, the typical calculation of the centroid to determine the position of the action figure causes an error, in this way the context that has a red cube would be incorrectly the closest. However, if the centroid is lowered vertically, the error is avoided.

Figure 5.5: Action figure centroid

To obtain the characters position in the plan, it was created a real time process: for each action figure creates a projection of a line that begins at the point corresponding to the 2d position of the action figure (blob position) in order to intersect this line with the plane defined by ARToolkit patterns present on the map (Figure 5.6).

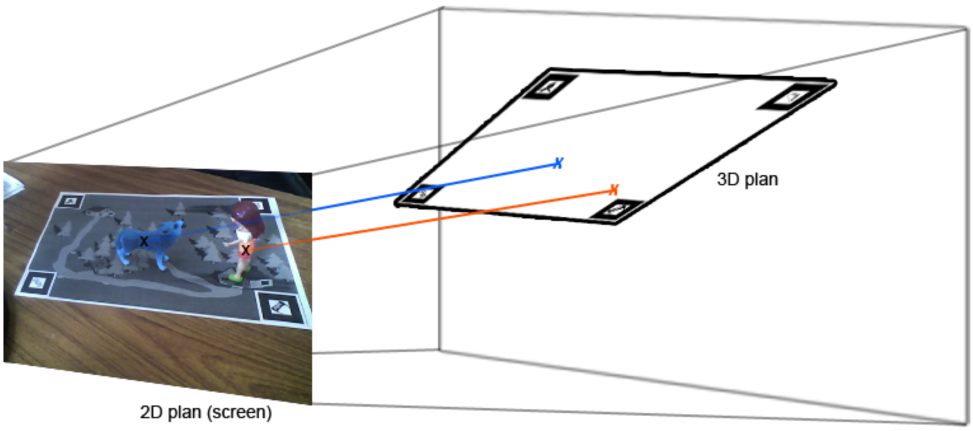

Figure 5.6: Get character position in 3D map

Once calculated the characters positions is then possible to calculate the distances between these positions and the contexts positions in order to determine the nearest context of each character. Regarding the type of movement, this is analyzed in 2D by the sum of 2D points samples, in order to quantify the shift in x and y as well to determine the speed in x and y through the ratio between displacement and time.

It was also explored the gestures recognition with the aim of allowing the user to configure the gestures. This implied the use of a gesture recognition process: $1 recognizer [Wobbrock et al., 2007]. Basically, the user should be able to capture a self-made gesture (action figure movements), and assign it to a kind of virtual character behavior. Although the interaction occurs in a 2D plane (the story map), the movement made with the action figures (gestures source) move in the three-dimensional space which greatly complicates the detection of gestures using this technique. Given that the configuration of gestures is not a priority in the project and considering the complexity of other solutions for the gestures recognition, this functionality was abdicated.

5.2.2 Sensing - events generation

For the events generation (Figure D.2) it was used the software design pattern: Observer, through this pattern it was created a class Observer devoted to the observation of certain events triggered in each *Doll* object. Basically, the events that could be triggered on each *Doll* object are:

1. character present in a new context;
2. character not visible;
3. character with horizontal movement (speed dichotomy - fast or slow);
4. character without horizontal movement;
5. character with Vertical movement (speed dichotomy - fast or slow);
6. character without Vertical movement;
7. character with Diagonal movement (speed dichotomy - fast or slow);
8. character without Diagonal movement.

Also, as a scope of events generation, it was created the *Hearing* class. Its responsibility was to analyze if any user is speaking. The idea was to use this event to trigger some kind of virtual character behavior. In practice, when the user speaks, then the virtual character assumes the speech behavior. Through the action figures movements would always be possible to disambiguate the input (infer what the character should speak, knowing that user

have two action figures in hand). However, this event was not used, the problem lies on the speech behavior construction, particularly in the synchronization between audio and lip movements, at the end of the sentence (the moment when the user stops talking, does not match the time that the virtual character stops moving their lips).
In order to enable the easy monitoring of the system variables, it was created the class Blackboard, using the software design pattern: Singleton. Through this it is possible to store and centralize the current status of all the important system variables, namely, which is the context where the action is unfolding, which is the virtual character more active and what is the state (on or off) of the Virtual character performer component (Section 5.3). Note that both the active context and the active character are updated through the *Doll* objects, and the state of the Virtual character performer component is updated via a connection that exists between the Sensing component and the Virtual character performer component. This connection is made between two Adobe Flash applications, using the *LocalConnection* class that allows the creation of a *LocalConnection* object that can invoke a method in another *LocalConnection* object. Finally, the *Observer* object checks in real time if some event is triggered by one of the *Doll* objects, if the *Observer* object is notified, it is then verified which is the active character and the active context, consulting the *Blackboard* object, in order to enable the events construction. Basically, any *Doll* object can trigger events with the logic: *"character X entered in the context Y and is making the movement Z"*. Thus, the *Observer* object, by querying the *Blackboard* object can create an event with the following nomenclature: *"activeCharacter-activeContext-currentCharacterAction"*, e.g. "Wolf-GrandmotherHouse-FastHorizontal". Once built the event is then sent to the Virtual character performer component using the previously referred *LocalConnection*.

5.3 Virtual character performer

It is expected that this component enables both the specification of the virtual characters behaviors, and the real-time execution and control of the previously specified behaviors triggered by the user interaction. As already stated the system is web-based, thus possible technologies for the realization of the virtual characters' behaviors are more restricted. The use of rendering engines in a web-based paradigm tends to be difficult, either because the solutions provided for this paradigm require the installation of a plugin, either by the graphical results obtained, which do not present a significant visual appearance. Finally, it is considered that tools for the virtual character behavior specification are essential to ensure the behavioral complexity of the virtual characters.

5.3.1 Avatar 2.5D

In the CCG (the institution that provided conditions for the realization of this project) it's currently under development a technology that aims to provide a solution to this component. *Avatar 2.5D* technology arises from the assumption that it is evident the increase of the web technology capabilities in terms of transmission speed. This increase in performance allows the creation of a virtual character technology which aims to increase the behaviors quality and the visual aspect of virtual characters. It is expected that the technology will allow an experience of interaction between users and virtual characters that tends to be richer, more realistic and ultimately more immersive, virtual characters that may present a very realistic appearance, based on physiological and biomechanical principles. Technologically, the solution is based on the pre-rendering to video files of all the possible animations that the virtual character can perform. This approach allows a video rendering with a very good quality, since the rendering time is not restricted to the frame rate of a render engine that typically needs to create 30 frames per second. The character can be very realistic, because it's possible to apply physical simulations of high cost (render time) e.g. clothes and hair. Once the videos are generated, they are saved in the *flv* format, and made available so that a web player can command the performance of the virtual character based on the user interaction input. This new technology faces a series of challenges. These are related to the automatic production of video clips, its organization, and the video transfer issue.

Behavior designer

Behavior designer is a component of avatar2.5D technology, which allows creating and editing the behavior organization of virtual characters. It is an editor for specifying the virtual character behaviors and events that influence the behavioral decision. In this editor, it is possible to define which virtual characters can be used in the system, in which languages they can communicate and establish the association between an event and the resulting behavior performed by the virtual character. As a basis for this component there is a generic graph editor that allows the specification and generation of files in the standard graph format: *Graphml*. With this editor, it is possible to define the events (causes) and their subsequent virtual character behaviors (consequences). The behavior specification is based on a nomenclature that allows the adaptation of a graph to the context: Performance of a virtual character. In practice, the links between nodes in the graph are associated with possible events (e.g. stop speech behavior), and the graph nodes are associated with the virtual character's behaviors that result in videos (e.g. video of virtual character speaking friendly).

Behavior generator

For the generation of the videos (virtual character's behavior) this technology has a generator component which is responsible for interpreting the behavior that a virtual character should play (previously specified in the behavior designer) and produce the videos. The Behavior generator has the basic resources:

1. 3D models of virtual characters associated with a virtual skeleton so that they can be animated;
2. animations based on motion capture [Wikipedia, 2011f] (technique used for body animation) and morph target [Wikipedia, 2011e] (technique used for facial animation);
3. categorization of the animation. This categorization determines the type of animation that each key frame belongs to, e.g. from frame N to frame M, the animation has the Speech category.

The motion capture and morph target animations are associated with the 3D models so that the generator can animate the 3D model according to the previously specified behavior. Finally, this component has the responsibility to organize a set of videos, and create a XML's specification that represents the organization of the videos and events. Technologically, this component is programmed in C++, making use of an external library for the 3D rendering (generation of video) and SAPI [Wikipedia, 2011g] for the speech generation and the Lip sync.

Virtual character performer - Player

To control the behavior's display there is a component developed in Adobe Flash technology that can be easily integrated into a web environment. This component can concatenate video clips from a "universe" of them in order to create a single video. A video that symbolizes the real-time performance of the virtual character. This output component parser the specification of the videos organization made available by Behavior generator component. Therefore, this component infers the video clips to be triggered, based on user interaction.

5.3.2 Use and adaptation of the Avatar 2.5D technology

In this project, it was used the technology Avatar 2.5D due to the advantages already outlined above. Despite all this, the use of this technology required adaptation, because the original technology supports the definition and execution of a performance for a single virtual character in a single context, i.e. the technology allows for example to define the action of a single journalist in the context of a television studio. In practice, when the generator must render a set of videos for the performance of the virtual character, for each frame, it is rendered a character over the context. With the story of LRRH new assumptions are raised, such as LRRH character can perform the speech behavior at his home, in the woods or in his grandmother's house and obviously other characters may be opposite (Wolf or Mother). The possible combinations taking into account the dynamism of the characters and the contexts in which they are embedded make it complicate to render all hypotheses. It is therefore, necessary, pre-render the performances of a virtual character in multiple contexts, without having to render a plethora of videos. Notes that for this project are defined three different contexts (see the Figure 5.7), the LRRH's house, forest, and the Grandmother's house. These three contexts in this framework represent three videos.

Figure 5.7: LRRH story contexts.

With regard to the characters (see the Figure 5.8) a video representation was created for the LRRH, the Wolf and LRRH's Mother.

Figure 5.8: LRRH story characters.

Given this problem, the solution was the definition of a graph for each character and a contexts graph. In this way, the videos associated with the contexts and the characters are rendered separately, resulting in sets of videos for the various story contexts and for each virtual character. The strategy for the Virtual character performer player is based on video layers (see the Figure 5.9), i.e. the first video layer is for the story contexts and the upper video layers are for the characters performances. Thus, it was necessary to ensure the transparency in the characters video layers so that the context video layer does not become hidden. The specification of the graphs files had to include an identifier for the graph (the character's name or in the case of contexts the name "Context"), and also include the depth level in order to define the video layer (see the Appendix G).

Video creation

For the creation of videos (contextualized by the story LRRH) that are necessary to allow interaction between the user and the system, it was followed a process that involved the use of various technologies.

Motion Capture (Vicon IQ) In a first stage, the aim is to capture human movements so that they can be applied in virtual models. This step allows the contextualization of animation, and ensures a fluid and realistic animation. To accomplish this task it was used the Motion capture technology: *Vicon*. This technology requires the existence of some infrastructures such as video cameras, hardware to centralize the video cameras information, a room suitable for the motion capture, a suit (clothing) and a set of optical markers that should

Figure 5.9: video layer solution

be placed in the human body. With regard to software, the process of motion capture is done through the *Vicon IQ* software, where it is possible:

1. calibrate the capture environment;
2. make the capture;
3. post processing the captures in order to make the labeling of 3D points (optical markers capture) over time in order to be associated with a part of the skeleton and eliminate errors introduced such as Gaps (hidden optical markers) and Ghosts (capture of optical markers that do not exist);
4. export the capture information to the Motion file format: C3D, in order to be used in next step.

Retargeting motion capture data to skeleton (MotionBuilder) In a second step it was used the *MotionBuilder* software (see the Figure 5.10). Based on a C3D file produced by *Vicon IQ*, it's possible to do the retargeting of the motion data to a virtual human model, specifically for the joints of the virtual skeleton. It's also possible through this tool to ensure that the transition between animations is made by interpolation in the case of animations that have no linear transition. Finally, export the animation to BVH file format, so that it can be applied to the 3D model, through the rendering software.

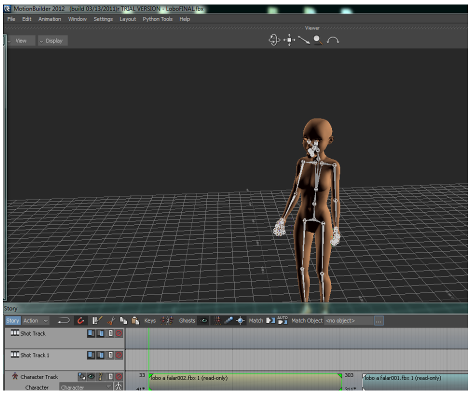

Figure 5.10: Targeting motion capture data to virtual skeleton

3d renderization (Daz3D) For 3D rendering, it was used the software *Daz3D* (see the Figure 5.11). A major advantage is the availability of 3D models that can be easily parameterized (physical structure of the body, age, etc.), thus the contextualization of the models to the LRRH story can be easily done. After the creation of the 3D models for the LRRH story (Wolf, LRRH and LRRH's mother) the MotionBuilder animation was associated to the 3D models. In regard to the facial expressions, the emotional expression of the characters was also parameterized, e.g. angry wolf, friendly wolf. Finally, for the animation renderization, it was configured as output of the render: images with transparency channel (one image corresponds to one video frame) in the file format PNG. The goal is to allow the videos creation with transparency channel.

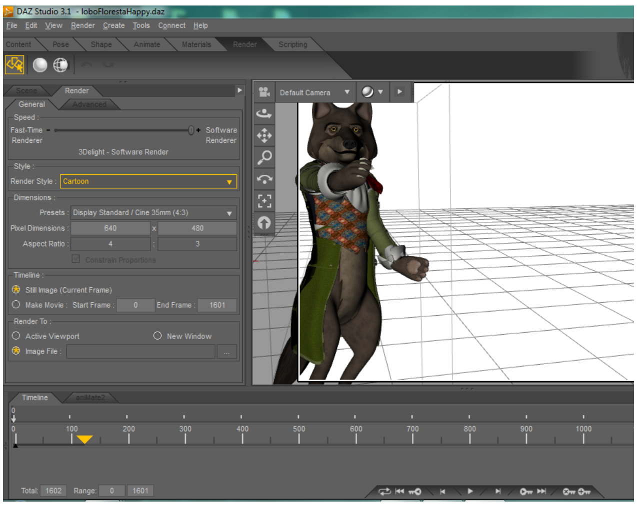

Figure 5.11: 3d renderization process with Daz3D tool

Video renderization (After Effects) The last step is the video rendering of the corresponding characters performances by setting the render output to the video format required by the Virtual character performer player: Flv. This task is made using the post-production video software- *After Effects* (see the Figure 5.12), that have as input the set of images with transparency channel rendered in *Daz3D*.

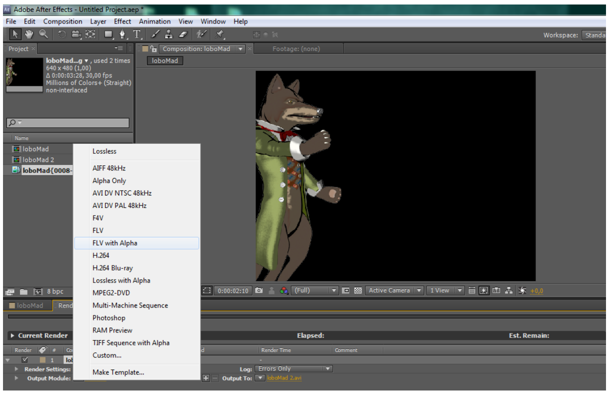

Figure 5.12: Video renderizations

Virtual character performer response to the sensing module events

As previously mentioned Subsection 5.2.2 the Sensing module communicates with the Virtual character performer module through a Adobe Flash *LocalConnection*. When the Virtual character performer module receives the events, the information present in the characters and context graphs files has already been parsed and transformed into temporary data structures. Thus, through these data structures (graph) is possible to search, what video should be triggered depending on the events that are received. Regarding the contextualization of the sensing events to the LRRH story it was done the following:

Table 5.1: Map sensing events to LRRH events

Sensing Events	**LRRH character**	**Wolf character**
Character not visible	character out	character out
Character with slow horizontal movement	talking happy	talking happy
Character with fast horizontal movement	talking very happy	talking happy
Character without horizontal movement	Waiting happy	Waiting happy
Character with slow vertical movement	slightly angry speak	angry speak

Character with fast vertical movement	quite angry speak	angry speak
Character without vertical movement	Waiting angry	Waiting angry
Character with slow diagonal movement	talking very scared	-
Character with fast diagonal movement	talking scared	-
Character without diagonal movement	Waiting scared	-

As can be seen in the previous table, the only characters directly influenced by inputs from the action figures are LRRH and the Wolf. Thus, the LRRH's Mother character is directly influenced by the actions of LRRH, and only appears in the LRRH's house, finally the LRRH's Mother and the LRRH always has the same emotional state.

5.4 Voodoo web application

The final application that integrates components mentioned above is a web based application. It was built based on the result of the system design phase as described in Subsection 4.3.4. The application is divided into two blocks: the System configuration and the Interaction.

5.4.1 System configuration

In system setup, the user must specify how it will interact with the system. For this, the user must define in a first stage which story will serve as basis for his interaction. The application provides a list of stories so that the user can choose (see Figure 5.13). To this end, it's previously done the parsing of the xml file: story.xml (see Appendix F) in order to list the stories that the system can provide. When Voodoo web application receives the story ID that user chose, a JavaScript communication is made with the sensing component, in order to communicate the story ID that will influence the Sensing component configuration. Afterwards and based on the story ID, the application has access to the path of the image corresponding to the map of the story, so that the system can offer the printing service of the map. After printing the map, the user can configure the action figures that he/she will use, and adjust the webcam to the correct angle. This last step of the configuration is done by embedding the sensing component (Adobe Flash application) in the Voodoo web application.

Figure 5.13: Choice of the story in web application

5.4.2 Interaction

After configuring the position of the camera and the action figures association with the characters of the story, the sensing component switches to a "background" mode and becomes invisible. Is then initiated the interaction mode (see Figure 5.14), with this mode the user has the ability to create animations. The interaction mode implies embedding the Virtual character performer player in the Voodoo web application, and communicates the story ID with it. Moreover, the application of sensing is responsible for the events sending for the Virtual character performer player.

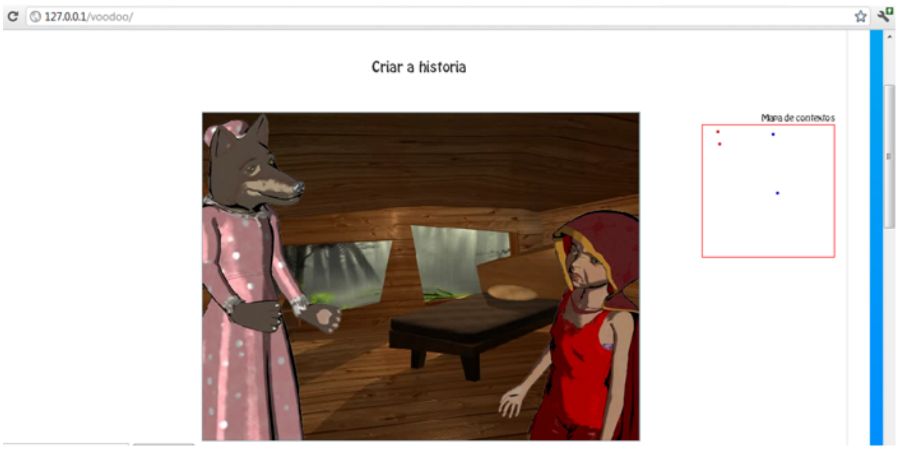

Figure 5.14: Interaction in web application

5.5 Management

For the Story Voodoo manager module was only possible to do a conceptualization (see Appendix C) of its functional specification. Thus, it was envisioned a Back Office environment intended to allow to a Manager to specify the stories, this involves the integration of the following features:

1. generic specification of the story:
    - the name of story;
    - the names of the characters;
    - a representative picture of the story;
2. specification of the story map:
    - map image definition;
    - specification and map integration of the ARToolKit visual patterns;
    - creating and naming contexts;
    - associate the contexts to the ARToolKit visual patterns;
    - creation of videos for each context;
3. creation of the virtual characters' behaviors:
    - behaviors specification for each character through the Designer Behavior (avatar2.5D technology);
    - for each virtual character, associate their behaviors with the gestures made through action figures manipulation;
    - generation of behaviors (video) through Behavior generator (technology avatar2.5D);

Chapter 6

Evaluation

The evaluation of the functional prototype of the system took place in a familiar context to the three test participants (children). There, were carried out several experiences of using the system with the test participants. The functional prototype was used by children individually (one child at a time) and the duration of interaction varied with the predisposition and behavior of children.

At the beginning of the experiment the test participants were instructed about the purpose of the system and the way it should be used. The children were informed about which elements of existing interaction (Action figures, map, story, types of gestures) and how they are related. Likewise, they were informed about the base story and it was found that they all already knew it. Note that children must know the story, because it is considered extremely important that children can use an action figure as having prior knowledge of what it represents (story character), because only then is that predictably that children can enhance the interaction, through the assumption: "I know this character and I know how he typically behaves." After the theoretical introduction to the system, children, one at a time, began freely their interaction with the system, they were observed in real time, and filmed (for further observation) in order to make the analysis of interaction and usage experience.

The functional prototype that was used in the evaluation process, is based on the interaction module of the system, "ignoring" so the configuration part. It was considered very important to check and test the proposed mode of interaction and not how the children could configure the system (choice of story, action figures configuration, map and webcam placement). Thus, based on the case study used throughout this project, the tests were based on the story of LRRH. The action figures used are related with the LRRH character and the Wolf character. It was considered that the minimum necessary to test the system in terms of the number of action figures were: two. It is highly unlikely that the child uses more than two action figures at the same time, and even if they do it, they hardly managed to get their attention on them all. The map used offers three context environments: The LRRH's house, the forest and the grandmother's house. The LRRH character may appear in any of these three environments, the Wolf character can only appear in the forest and grandfather's house and its virtual character representation switches between these two scenarios. The mother of LRRH character may only appear at LRRH's house and only when the LRRH action figure is interacting in this area. The map is a DIN A4 paper and it was printed in grayscale in order to avoid false positives with respect to the action figures detection (confusion between a red roof illustrated on the map and an action figure configured to be detected through the red color).

It should be noted that this project does not aim to make an exhaustive evaluation of the functional prototype, this study will have the primary purpose of obtaining a first feedback regarding the interaction and usage experience.

6.1 Objectives of the evaluation

Given the objectives of the project, it was essentially intended to evaluate the acceptability of the system through direct observation of the participant's reaction (see the Figure 6.1).

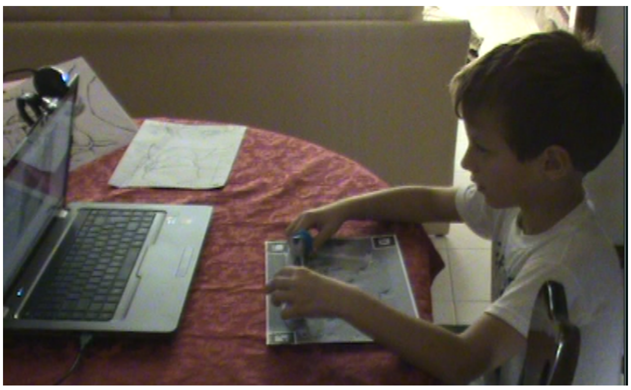

Figure 6.1: Test participant using voodoo

Below are presented in detail the key points of the system evaluation:

1. interaction and usability:
   - usage metaphors;
   - ideal screen size, map size and *action figure* size;
   - ideal relation between *action figure* size and map size;
   - ideal number of contexts in the map;
   - ideal distance between interfaces;
2. Usage experience:
   - understand if the interaction model created, enhances children's abilities for creating animated videos;
   - children shows confusion;
   - children shows happiness;
   - children shows frustration;
   - easily associates the *action figures* to the characters of the story;
   - can clearly see the contexts behind the map and the relationship between the map and story;
   - can easily activate the behavior they intend through gestures;
   - think the map is ambiguous;
3. impact, contributions and limitations that the various technologies may offer to the system.

6.2 Evaluation methods

The evaluation results were obtained using two different methods [Ideo, 2002]: the Observation and the Try it yourself methods. It was decided to select these two evaluation methods because they offer particular potential for the tests conditions. The direct observation of the children behaviors (shadowing technique) allows us to study very directly the various behaviors of children while using the system. This technique has the advantage of not requiring any kind of reflection or introspection on the part of children and has the advantage of not being very intrusive. It was also used video recording equipment to record the behavior of test participants in order to make an further observation with more detail. Finally, with regard to the method Try it yourself, it was considered important that who developed the project could evaluate directly the experience that the actual users might have. Comments or suggestions provided by the test participants were registered and were being analyzed in order to determine where the system can be improved and/or identify possible future developments.

6.3 Results and analysis

The analysis of results had as main objectives, find answers, conclusions and possible opportunities in the interaction / usability of the system and in the user experience. In Appendix H are available the raw results of the evaluation made with the three test participants.

6.3.1 Interaction and usability

Metaphors

Regarding the interaction metaphors proposed in this project, it was observed that children correctly interpreted the metaphor of the system. They perceive easily (by intuition and learning) that the placement of the action figure in a new context will change the scenario where the virtual character will appear, as well as how it looks and behaves. The children even have an immediate tendency to move the action figures in order to give them life or generate new behaviors. However, in this particular field: metaphor of gestures, it was evident that the logic of interaction is not the right one. When a child has two action figures, one in each hand, the wolf and LRRH, the child just can individually activate new behaviors in virtual characters, i.e. it's only possible to move the action figure of the wolf and influence the Wolf virtual character's behavior, it's impossible that the Wolf action figure interaction influences the behavior of the LRRH virtual character. It was verified that the children never showed intent to make a direct control of the virtual character (such as raising an arm of the action figure to raise the arm of the virtual character). Children always interacted through the intentional control Subsection 2.1.2.

Dimensions and distances between interfaces

It is considered that the dimensions of input interfaces (A4 paper map, small action figures) are appropriate dimensions to allow a good interaction, this means that the size of action

figures allows associating without ambiguity a character to a context even when all the action figures are in the same context. It is considered that the distance between interfaces is adjusted. Children initially, visualize contexts in which they intend to place the action figures and then they make gestures while looking at the screen expecting to see the outcome or determine if the behavior that virtual characters have is consonant with their intentions. By observation, it is considered that children provide more time viewing the result of the action figures manipulation on the screen, than viewing the action figures manipulation.
With regard to interaction and usability problems, it was found that the system, in many cases, did not respond immediately to users' gestures. The latency is directly related to light conditions. In practice with low light most webcams automatically increases the exposure time for the composition of each frame, the consequence is a low frame rate. The system need to register, samples of the action figure's position, and each sample can only be achieved by each frame, so if the frame rate is low the system needs more time to determine the child gesture. This caused, frustration, lack of control and influence children's learning. They had difficulty in judging what is the consequence of the action figure manipulation in the virtual character behaviors. It is therefore highly important to achieve synchronization between cause (act done with the action figure) and consequence (the behavior of the virtual character).

6.3.2 Usage experience

It was verified that children liked the animations and even tried repeatedly to activate those they considered funnier. Specifically, in the emotional controlling of the virtual characters there are cases where the children would say in a loud voice what the emotional state they intend to turn on the virtual characters. In another specific case, the children repeatedly tried to activate the animation in which the wolf virtual character attacked the LRRH virtual character. Despite this, children never showed the intention of create a narrative and transforms it, through the system, in a graphical representation on the screen.
In the end, it is perceived that the constraints of the test undoubtedly influenced the user experience, because typically the act of play is a spontaneous behavior, the child plays when he wants, where he wants, and the way he wants. It is impossible to expect this in the test environment.

6.3.3 Technologies - Limitations and contributions

At the technical level from the results of "Try it yourself" the first consideration goes to the system configuration, which was mentioned at the beginning of this section, as not being a target of evaluation. However, it is important to highlight certain aspects that influence the interaction module. Thus, with regard to sensing based on computer vision it is considered that this implies several restrictions.

The technology choice was largely influenced by the type of application that is intended to create, in this project the aim was specifically a web-based application. Another influence for the choice of technology was the use of technology that requires the least effort possible to create the various modules, as the project's main investment is to prove a concept (to prove a system) and not for the development of modules that are highly efficient. Nevertheless, this commitment affected in some extent the proof of concept.

With regard to the real-time tracking of the story contexts through the ARToolKit visual patterns, it is considered that the system had an acceptable behavior. The three-dimensional position of the visual patterns is not constantly calculated either by the light conditions or by occlusions, however, whenever possible the system stores these positions. Since the system requires that the position of the camera and map tend to remain unchanged, the position of ARToolKit visual patterns remains fairly stable. The only rare exceptions occur when the system detects a ghost pattern (false positive). Another issue is related to the map where patterns are inserted, they must be printed on diffuse paper, so that the pattern can be easily detected. If the printed map is made on reflective paper it would not allow the detection of the ARToolKit visual patterns .

With respect to the action figures tracking, this was not always successful. The first reason is related to the blob color tracking. Except in cases where the conditions are extremely well controlled, may be frequent the occurrence of false positives, because the key colors can be detected in various surrounding objects, or even in the user. The second reason relates to the type of material where is made the color capture, a reflective material implies that this may have variations in color, on the other hand, if the material is diffuse it will remain much more stable.

Regarding the detection of gestures, this is also negatively influenced by lighting conditions. The reason lies in the automation existing in typical web cams. Basically depending on light conditions, time of exposure oscillates. With low light the exposure time per frame increases, causing the frame rate drop. Basically, if the frame rate is below 30 frames per second is hardly possible to obtain reliable samples of action figures positions in order to infer what

type of motion is being done.
It is considered that failures associated with false positives of the ARToolKit patterns and action figures colors resulted in an interaction that does not allow full control over the system.
As regards to the Virtual character performer technology is considered that this had a good performance, the main challenges, such as the passage between different animations or multiple layers of action were not an obstacle to the platform, the proof was the pleasure that children had when viewing the animations triggered.

Chapter 7

Conclusion

Tools that facilitate the interactive creation of linear stories, in particular for children, have an enormous entertainment and pedagogical potential. One question regarding such tools is how to deduce complex animations and stories from simple input of intuitive input devices. It is presented an approach where the contextual constraints of a story are integrated to enable a child to create animated films with action figures as input devices.

Under this project, a functional prototype of the Voodoo system was created with the purpose of becoming easier for children to reinvent a well-known story and eventually to create new stories. One of the main project's objectives was, actually, building the system which is considered as an achieved goal. Since the beginning of the project, it was clear the need to prove a concept through a functional prototype, but it was also clear that it would never be possible to be highly efficient, due to the complexity, multidisciplinarity and novelty of the system. The option was obviously the creation of a system, in a pragmatic way, without losing too much time on each module. Although a complete system was achieved, with all its functional modules properly connected, it is clear, as future work, that the system should evolve in the efficiency of all its components.

If the creation of the system was, in some ways, successful, the same cannot be considered regarding the evaluation of the system in terms of its ability to help children to create animated films. Efforts were made to achieve this target through the evaluation of the system but it was impossible to fully evaluate the potential of it due to two main reasons: the fact that the evaluation was made on a prototype that still does not provide the desired stability and efficiency in all its components; the constraints inherent to the evaluation tests. Although these tests were done in a familiar environment, the children never felt comfortable enough

to be abstracted from them. This factor had a negative influence on the evaluation because the children never felt the desire to freely perform their play and possible narratives. It is believed that the solution may be the use of cultural probes [Gaver et al., 1999], which can be appropriate when: it is necessary to gather information from users with minimal influence on their actions; the process/event to be explored takes place intermittently or over a long period of time.
Although it was possible through the assessment to perceive the potential of the proposed concept of interaction, it was found by observation that children can easily understand the logic of interaction and intuitively learn how to use the system. It was also found that the child who has experience of interaction, using typical interfaces (keyboard and mouse), did not present any difficulty associated with interaction assumptions.
Regarding the scalability of the constructed system it is concluded that the system was entirely built to address these needs. It was built to allow easy maintenance and management of contents and to minimize the costs involved. Therefore, for the architecture it was chosen a web-based system to facilitate the access via the Internet to end users and aimed to facilitate the management of the system. In terms of tangible interfaces, these are also on the client side, because he (child) is responsible for printing the map, and for the choice and configuration of the action figures, which he intends to use. It is believed that this model is ideal to enable the system to easily grow in number of stories and certainly to minimize costs because:

- will not require a special manufacturing in order to guarantee the system existence;
- it is not necessary to create specific action figures/maps;
- no need to create and/or combine sensors specialized for the system.

Despite these advantages this model has exposed some difficulties and disadvantages. The first is related to the technological constraints that this model imposes. These restrictions were primarily evident at the sensing level. For example, it was not possible in time to explore and develop a sensing module that was not so limited by environmental conditions (lighting conditions). Despite this technological limitation in sensing, the features that demand high performance can be adapted to web-based paradigm. The proof of it is Avatar 2.5D technology that was used in this project. The challenge was how to have good virtual character rendering in a web based application and this challenge was overcome using a concept based on videos, as explained earlier.

After all, some essays were made on the future of the Voodoo. A first hypothesis aims to maintain the current strategy and opt for the detection and enhancement of the aspects that are limiting the system.
The main problem of the implemented system is located in the sensing module in particular the existence of false positives in the detection of context and action figures. The removal of these limitations is a matter of calibrating the system, capturing the colors in the background image and forcing the user to print a small label with a color that does not exist in the background. This will permit the user to paste the label on the action figure or choose an action figure that has the color required by the system. Otherwise the solution can pass through identify the set of pixels that are different from the background image. This region of interest would typically be associated with the movement of the action figures, so it is possible to avoid the color searching in the whole image.
Another issue is related to the fact that in the current system, the virtual characters' behaviors are triggered based on gestures made in a particular action figure (its related). It is believed that will make perfect sense that when a child handles two action figures, the behavioral control of virtual characters will be based on the composed gesture of two action figures. It is certainly difficult for most children do simultaneously different movements with each hand.

In addition to the natural attempt to overcome the major problems in the current system, it is also natural to essay on other potentially desirable features to add value to the system. It is not clear yet if it is mandatory to enable the child to somehow post-edit the animation. Later, even user emotions could be taken into consideration in order to choose the appropriate animation, as well as some automatic interpretation of what the child is saying. The collaborative use of the system is easily possible; for this, the size of the paper map must be increased, and larger displays would be required. Similarly, it is believed that it will be imperative to provide the user with the configuration of gestures in order to allow the insertion and association of new gestures (relates it with the virtual character behavior).
An interesting feature to investigate is the inclusion of music. [Cohen, 1998] pointed functions of music in multimedia such as induce mood, increase the sense of immersion, and enhance the communication of the narrative meaning. More recently, [Eladhari et al., 2006] present a system for individualized (for each virtual character) adaptive music, where the music is used to reflect the affective process of mood, emotion and sentiment. The inclusion of music in an adaptive manner, can provides to Voodoo the reinforcement of the interaction metaphor and the control over an important component for creating animated films.

From another perspective and abandoning the ideology of the system argued in this paper (creating a system of easy access) other hypotheses can be envisioned:

1. the use of a different type of sensor, namely *Kinect* [Wikipedia, 2011b]. This sensor can configure and track a virtual skeleton based on the capture of a human shape. Knowing that the *Kinect* has this feature and that is feasible at distances greater than 80cm it is believed that the tracking of a virtual skeleton for action figures could be a possibility and may offer a more robust sensing. There is also the advantage that lies on the creation of gestures (less ambiguous and more intuitive), essentially because these are three-dimensional gestures, that can be made over the whole body of the action figure, or just in parts, in fact, this can even allow the fusion of intentional control with the direct control of the system;

2. an alternative that has been previously addressed in the overview section, refers to the change of the action figures so that they can be sensed in a more robust way, implying of course the introduction of sensors in the action figure, besides allowing a more sophisticated tracking, this could permit the introduction of actuators so that the action figure itself could be an output interface;

3. a system directed to the world of books, where the tangible interfaces of the system are part of a pack consisting of a book and action figures. Each page of the book is a map with specific contexts that influence the performance of virtual characters. The action figures will be constructed to optimize sensing. This contextualization to the world of books may represent an opportunity in terms of metaphors, e.g. the notion of timeline represented at the turn of pages for both the moment of creation or to the time of post editing;

4. the passage of the map of the real world into the virtual world, thus the map with their respective contexts, could pass for a multitouch table where the action figures location would be easily calculated. Through this technological configuration would be possible the creation of animated films in a collaborative way;

5. the interaction model developed can also be used in other contexts. A good example would be using it for film navigation: putting an action figure in a particular context of the map would start the film at some point in the timeline. In the film "The return of the king" [Wikipedia, 2011c], if we put the *Frodo* action figure in *Mount Doom* context, the system would play the last part of the film, where the ring is destroyed in the volcano.

Finally, within this dissertation, it was possible to publish this concept [Ribeiro et al., 2011] at the fourth International conference on interactive digital storytelling.

Bibliography

Abras, Chadia; Maloney-krichmar, Diane; and Preece, Jenny. "User-Centered Design". In *In Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 2004. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.94.381. 11, 27

Andersson, Gerd; Höök, Kristina; Mourão, Dário; Paiva, Ana; and Costa, Marco. "Using a wizard of oz study to inform the design of sentoy". In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, DIS '02, pages 349–355, New York, NY, USA, 2002. ACM. ISBN 1-58113-515-7. doi: http://doi.acm.org/10.1145/778712.778762. URL http://doi.acm.org/10.1145/778712.778762. 12

Bernsen, N. O. "Designing interactive speech systems". 1998. 30

Bonanni, Leonardo; Vaucelle, Cati; Lieberman, Jeff; and Zuckerman, Orit. "Playpals: tangible interfaces for remote communication and play". In *CHI '06 extended abstracts on Human factors in computing systems*, CHI '06, pages 574–579, New York, NY, USA, 2006. ACM. ISBN 1-59593-298-4. doi: http://doi.acm.org/10.1145/1125451.1125572. URL http://doi.acm.org/10.1145/1125451.1125572. 8

Cohen, Annabel. "The functions of music in multimedia : A cognitive approach.". pages 13–20, 1998. 69

Eladhari, Mirjam; Nieuwdorp, Rik; and Fridenfalk, Mikael. "The soundtrack of your mind: mind music - adaptive audio for game characters". In *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, ACE '06, New York, NY, USA, 2006. ACM. ISBN 1-59593-380-8. doi: http://doi.acm.org/10.1145/1178823.1178887. URL http://doi.acm.org/10.1145/1178823.1178887. 69

Figueiredo, Rui and Paiva, Ana. "Watch and feel: An affective interface in a virtual storytelling environment". 2005. 6, 8

Fishkin, Kenneth P. "A taxonomy for and analysis of tangible interfaces". *Personal Ubiquitous Comput.*, 8:347–358, September 2004. ISSN 1617-4909. doi: http://dx.doi.org/10.1007/s00779-004-0297-4. URL http://dx.doi.org/10.1007/s00779-004-0297-4. 9

Foreui. "Foreui: Easy-to-use ui prototyping tool". URL http://www.foreui.com/. 36

Gaver, Bill; Dunne, Tony; and Pacenti, Elena. "Design: Cultural probes". *interactions*, 6: 21–29, January 1999. ISSN 1072-5520. doi: http://doi.acm.org/10.1145/291224.291235. URL http://doi.acm.org/10.1145/291224.291235. 68

Gibson, James J. *The Ecological Approach to Visual Perception*, volume 40. Houghton Mifflin, 1979. 22

HITLab. "Augmented reality software library". URL http://www.hitl.washington.edu/artoolkit/documentation/userintro.htm. 44

Hunt, Devin; Moore, Jamie; West, Alex; and Nitsche, Michael. "Puppet show: Intuitive puppet interfaces for expressive character control in medi@terra 2006, gaming realities: A challenge for digital culture", 2006. 13, 18

Ideo. "Method cards for ideo", 2002. URL http://www.ideo.com/work/method-cards. 61

Ishii, Hiroshi and Ullmer, Brygg. "Tangible bits: towards seamless interfaces between people, bits and atoms". In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '97, pages 234–241, New York, NY, USA, 1997. ACM. ISBN 0-89791-802-9. doi: http://doi.acm.org/10.1145/258549.258715. URL http://doi.acm.org/10.1145/258549.258715. 1

Iurgel, Ido A.; Da Silva, Rogério E.; and Dos Santos, Manuel F. "Towards virtual actors for acting out stories". In *Proceedings of the Entertainment for education, and 5th international conference on E-learning and games*, Edutainment'10, pages 570–581, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-14532-9, 978-3-642-14532-2. URL http://portal.acm.org/citation.cfm?id=1881763.1881827. 4, 18

Johnson, Michael Patrick; Wilson, Andrew; Blumberg, Bruce; Kline, Christopher; and Bobick, Aaron. "Sympathetic interfaces: using a plush toy to direct synthetic characters". In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, CHI '99, pages 152–158, New York, NY, USA, 1999. ACM. ISBN 0-201-48559-1. doi: http://doi.acm.org/10.1145/302979.303028. URL http://doi.acm.org/10.1145/302979.303028. 5, 10, 11, 16, 24

Kelley, J. F. "An empirical methodology for writing user-friendly natural language computer applications". In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '83, pages 193–196, New York, NY, USA, 1983. ACM. ISBN 0-89791-121-0. doi: http://doi.acm.org/10.1145/800045.801609. URL http://doi.acm.org/10.1145/800045.801609. 12

Mazalek, Ali and Nitsche, Michael. "Tangible interfaces for real-time 3d virtual environments". In *Proceedings of the international conference on Advances in computer entertainment technology*, ACE '07, pages 155–162, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-640-0. doi: http://doi.acm.org/10.1145/1255047.1255080. URL http://doi.acm.org/10.1145/1255047.1255080. 10, 24

Mazalek, Ali; Chandrasekharan, Sanjay; Nitsche, Michael; Welsh, Tim; Thomas, Geoff; Sanka, Tandav; and Clifton, Paul. "Giving your self to the game: transferring a player's own movements to avatars using tangible interfaces". In *Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games*, Sandbox '09, pages 161–168, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-514-7. doi: http://doi.acm.org/10.1145/1581073.1581098. URL http://doi.acm.org/10.1145/1581073.1581098. 5, 24

Nielsen, Jakob. "Iterative user-interface design". *Computer*, 26:32–41, November 1993. ISSN 0018-9162. doi: 10.1109/2.241424. URL http://portal.acm.org/citation.cfm?id=618985.619982. 29

Paiva, Ana; Andersson, Gerd; Höök, Kristina; Mourão, Dário; Costa, Marco; and Martinho, Carlos. "Sentoy in fantasya: Designing an affective sympathetic interface to a computer game". *Personal Ubiquitous Comput.*, 6:378–389, January 2002. ISSN 1617-4909. doi: http://dx.doi.org/10.1007/s007790200043. URL http://dx.doi.org/10.1007/s007790200043. 14, 16, 17

Ribeiro, Pedro; Iurgel, Ido; and Ferreira, Manuel. "Voodoo: A system that allows children to create animated stories with action figures as interface". In *Proceedings of the Fourth joint* *conference on Interactive digital storytelling*, ICIDS'11, pages 244–247, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-25288-4. 21, 22, 23, 71

Rogers, Yvonne. "New theoretical approaches for hci". *Annual Review of Information Science and Technology*, 38:87–143, 2004. 22

Saqoosha. "Flartoolkit", 2009. URL http://www.libspark.org/wiki/saqoosha/FLARToolKit/en. 44

Wikipedia. "Adobe flash", 2011a. URL http://en.wikipedia.org/wiki/Adobe_Flash. 43

Wikipedia. "Kinect", 2011b. URL http://en.wikipedia.org/wiki/Kinect. 70

Wikipedia. "The return of the king", 2011c. URL http://en.wikipedia.org/wiki/The_Lord_of_the_Rings:_The_Return_of_the_King. 70

Wikipedia. "Machinima", 2011d. URL http://en.wikipedia.org/wiki/Machinima. 10, 18

Wikipedia. "Morph target", 2011e. URL http://en.wikipedia.org/wiki/Morph_target_animation. 50

Wikipedia. "Motion capture", 2011f. URL http://en.wikipedia.org/wiki/Motion_capture. 50

Wikipedia. "Microsoft speech api", 2011g. URL http://en.wikipedia.org/wiki/Microsoft_Speech_API. 50

Wikipedia. "Split screen (filmmaking)", 2011h. URL http://en.wikipedia.org/wiki/Split_screen_(filmmaking). 40

Wobbrock, Jacob O.; Wilson, Andrew D.; and Li, Yang. "Gestures without libraries, toolkits or training: a 1 recognizer for user interface prototypes". In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, UIST '07, pages 159–168, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-679-0. doi: http://doi.acm.org/10.1145/1294211.1294238. URL http://doi.acm.org/10.1145/1294211.1294238. 47

Youtube. "Little red riding hood story: Youtube video", a. URL http://www.youtube.com/watch?v=BGTHyiMc08c. 37

Youtube. "Youtube: video-sharing website", b. URL http://www.youtube.com/. 37

Appendix A

Wizard of Oz study survey

Table A.1: Generic information

Variables	participant 1	participant 2
Participant name	José Pedro	Silvana
Participant sex	Male	Female
Participant age	8	8
Story name	Little Red Riding Hood	Little Red Riding Hood
Experience type	free exploration	free exploration

Table A.2: Assumptions about the participant

Variables	participant 1 answer	participant 2 answer
Has knowledge of new technologies	no	no
Knows the well-known story	yes	yes
Usually play with action figures	yes	yes
Own action figures or story action figures	2nd choice	2nd choice
Is very active	yes	no

Table A.3: User experience

Variables	participant 1 answer	participant 2 answer
Realized the story	yes	yes
Liked the story	moderately	yes
Shows confusion	no	no
Shows happiness	yes	yes
Shows frustration	no	no
action figures are contextualized	yes	yes
showed concentration	moderately	yes
Clearly see contexts behind the map and the	no	no
Direct relationship between map and history	no	no
Think the map is ambiguous	no	no

Table A.4: User interaction

Variables	participant 1 answer	participant 2 answer
Acceptable Screen size 19”	yes	yes
Acceptable Paper map size din A4	yes	yes
Acceptable Action figure size (5 cm)	yes	yes
Size relation between action figure and map	good	good
Acceptable Context number	no answer	no answer
Acceptable distance between interfaces	yes	yes
User looks at action figure percentage	15%	10%
User looks at screen percentage	85%	90%
Adjustable camera	no answer	no answer
Rigid camera	no answer	no answer
Context is sufficiently explicit	yes	yes

Appendix B

Usage workflow

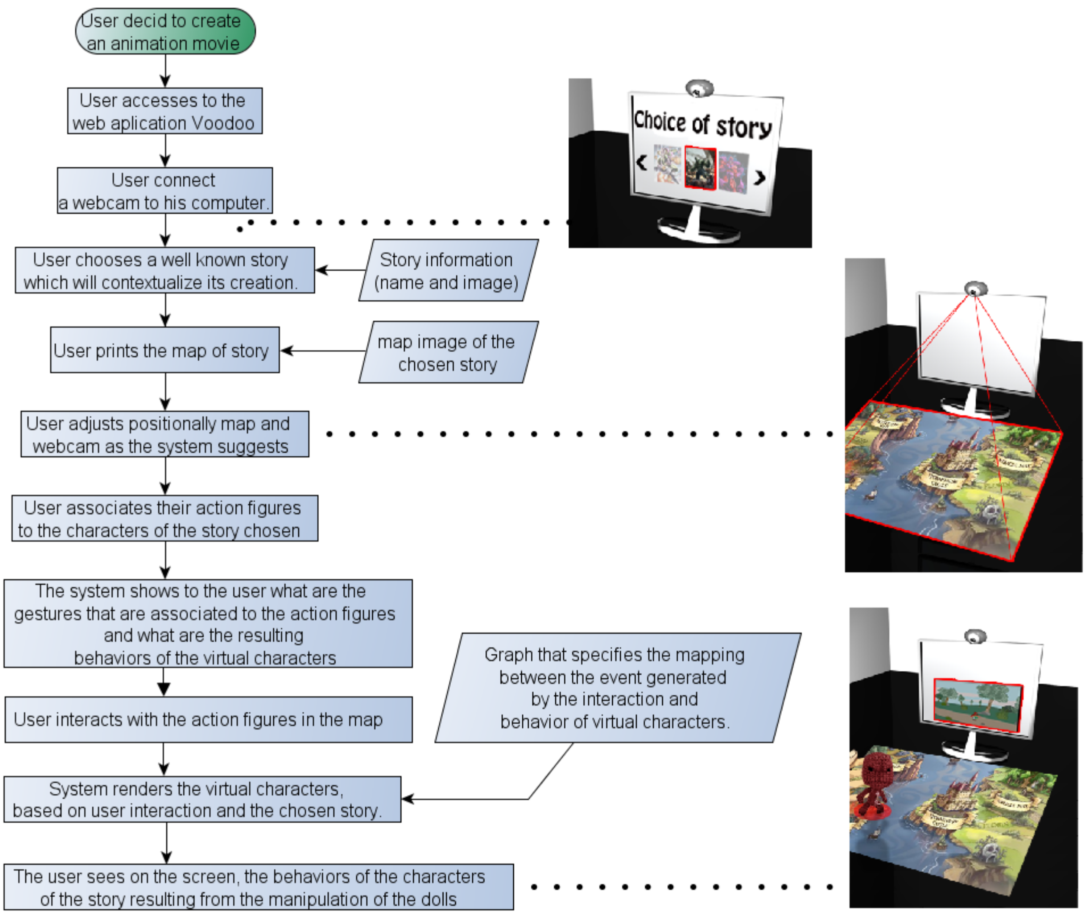

Figure B.1: Usage workflow

Appendix C

Story deploy flow diagram

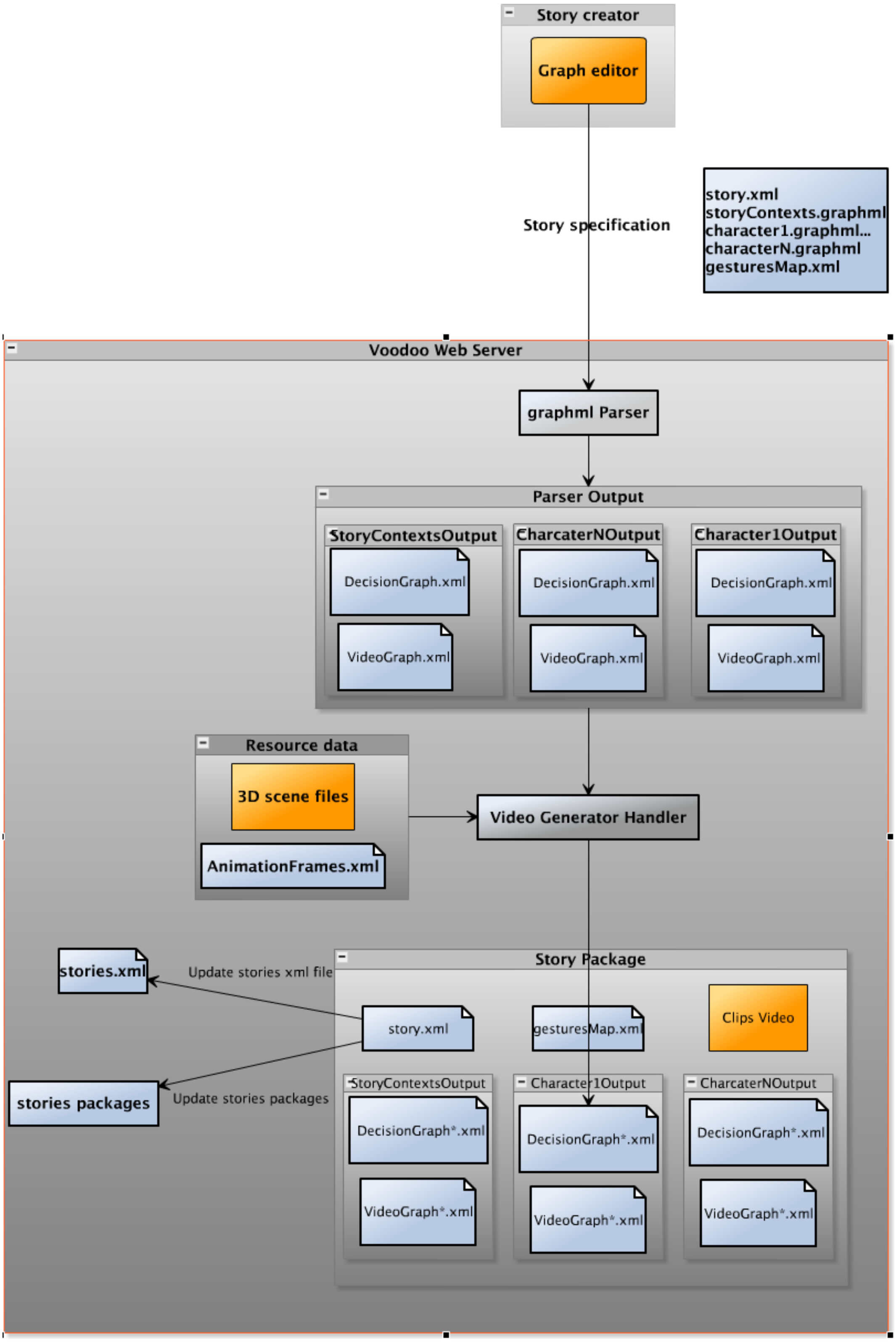

Figure C.1: Story deploy flow diagram

Appendix D

Sensing module

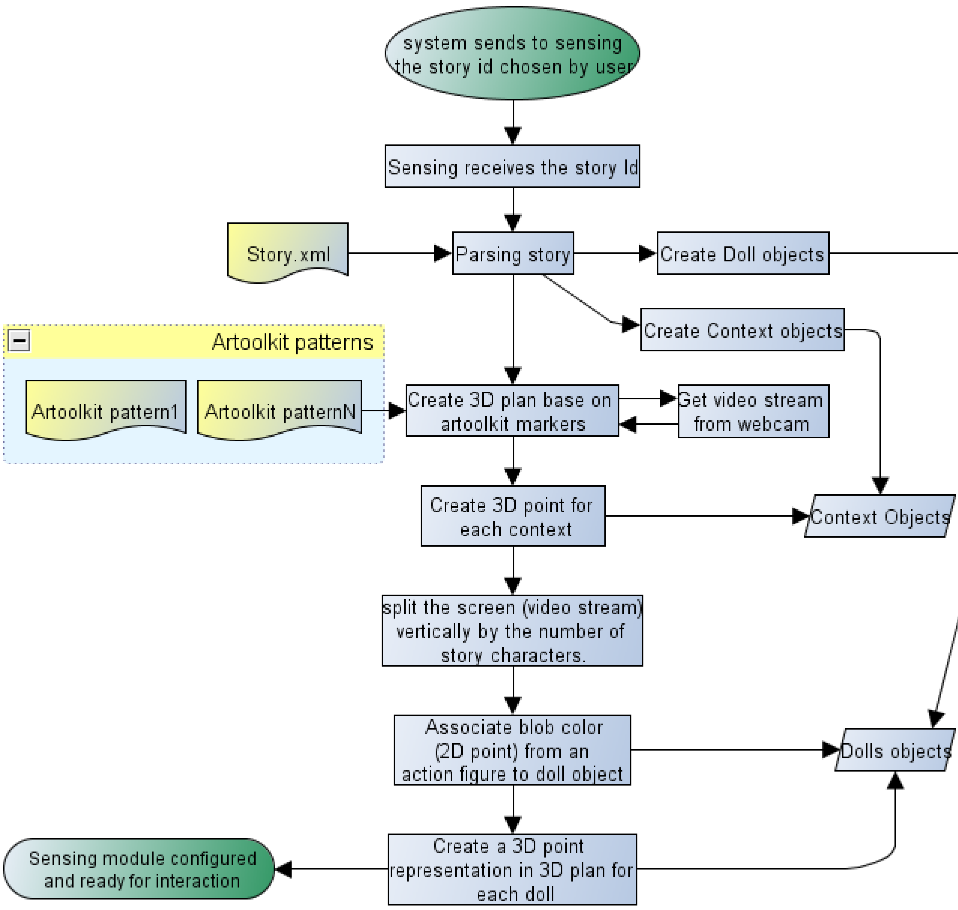

Figure D.1: Sensing module configuration flow diagram

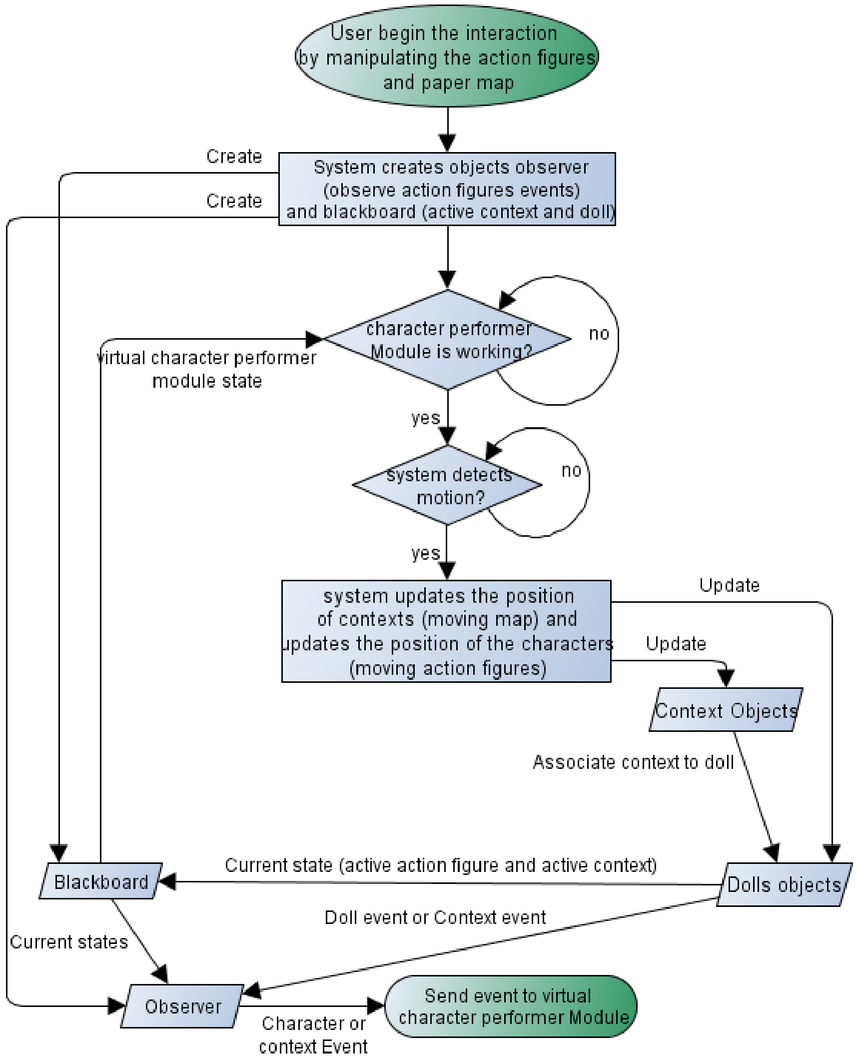

Figure D.2: Sensing module event generation flow diagram

Appendix E

Virtual character performer module

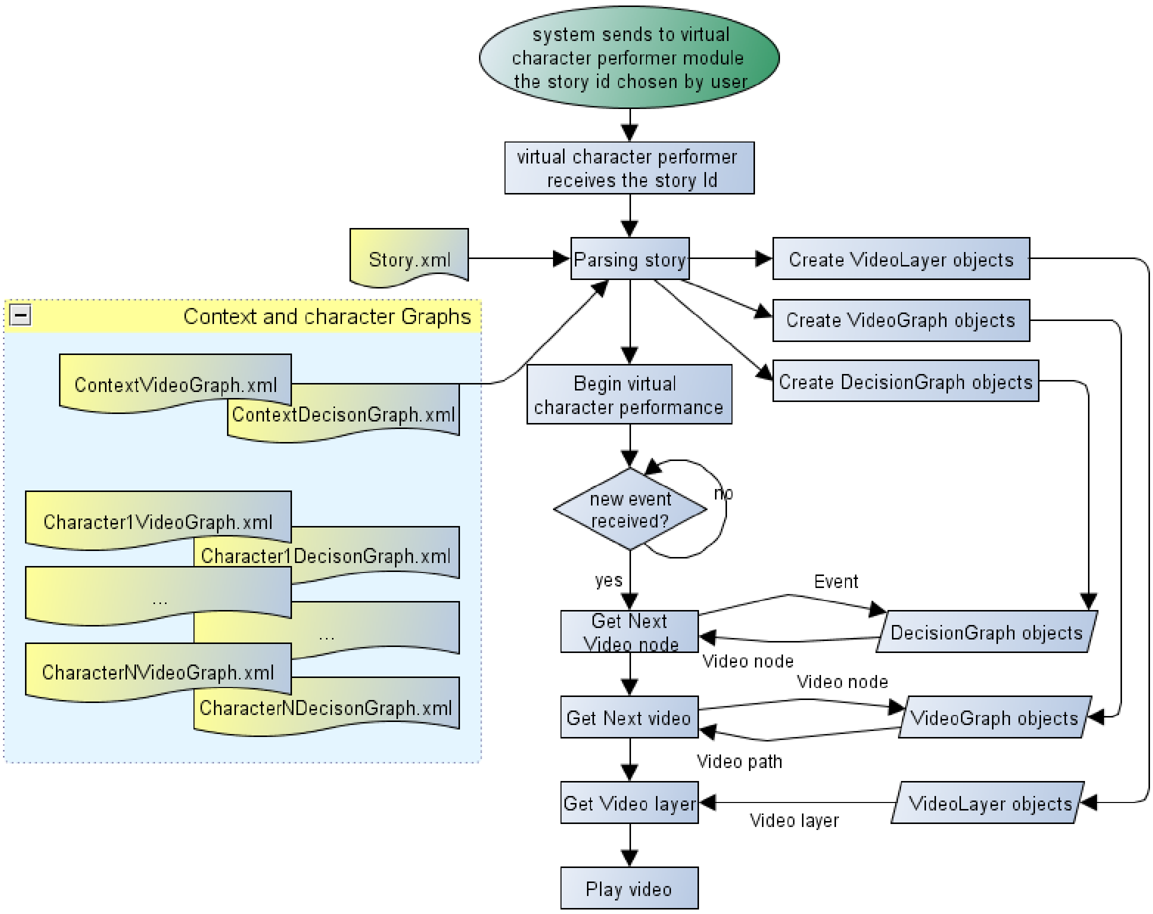

Figure E.1: Virtual character performer flow diagram

Appendix F

Stories specification

Story.xml

```
<?xml version="1.0" encoding="utf-8"?>
<stories>
  <story>
    <id>0</id>
    <nameStory>Capuchinho vermelho</nameStory>
    <img>./img/Stories/Little_Red_Riding_Hood.png</img>
    <printImg>./img/Storys/map_LRRH.png</printImg>
    <markers>
      <marker>
        <cod>0</cod>
        <filePat>./inicio.pat</filePat>
      </marker>
      <marker>
        <cod>1</cod>
        <filePat>./fim.pat</filePat>
      </marker>
      <marker>
        <cod>2</cod>
        <filePat>./mar.pat</filePat>
      </marker>
```

```
</markers>
<contexts>
  <context>
    <nameContext>CasaCapuchinho</nameContext>
    <positioning>
      <marker1>0</marker1>
      <marker2>-1</marker2>
      <influence>-1</influence>
    </positioning>
  </context>
  <context>
    <nameContext>CasaAvo</nameContext>
    <positioning>
      <marker1>1</marker1>
      <marker2>-1</marker2>
      <influence>-1</influence>
    </positioning>
  </context>
  <context>
    <nameContext>Flores</nameContext>
    <positioning>
      <marker1>2</marker1>
      <marker2>-1</marker2>
      <influence>-1</influence>
    </positioning>
  </context>
 <characters>
  <character>
    <nameCharacter>Capuchinho</nameCharacter>
    <imageCharacter></imageCharacter>
  </character>
  <character>
    <nameCharacter>Lobo</nameCharacter>
    <imageCharacter></imageCharacter>
```

```
      &lt;/character&gt;
    &lt;/characters&gt;
    &lt;avatar_info&gt;
        <context id="-1" name="Contexto" videograph="videoContextos.xml"
decisiongraph="decisionContextos.xml"/>
        <actor id="0" name="Capuchinho" sex="M" videograph="videoCapuchinho.xml"
 decisiongraph="decisionCapuchinho.xml" lang="PT" />
        <actor id="1" name="Lobo" sex="M" videograph="videoLobo.xml"
 decisiongraph="decisionLobo.xml" lang="PT" />
        <actor id="2" name="Mae" sex="M" videograph="videoMae.xml"
 decisiongraph="decisionMae.xml" lang="PT" />
     &lt;/avatar_info&gt;
  &lt;/story&gt;
  ...
&lt;/stories&gt;
```

Appendix G

Character performance specification

```
<?xml version="1.0" encoding="utf-8" ?>
<videoGraph id="Capuchinho" zLayer="1">     Character identification and the video layer depth level
    <vertices>
        <vertex id="n1">     Vertex identification
            <video name="out" >
                <hi url="capuchinho/H_Out.flv" />     Video files path
                <lo url="capuchinho/L_Out.flv" />
            </video>
        </vertex>
        <vertex id="n2">
            <video name="madsoftidle" >
                <hi url="capuchinho/H_MadSoftIdle1.flv" />
                <lo url="capuchinho/L_MadSoftIdle1.flv" />
            </video>
        </vertex>
        ...
  </vertices>
  <edges>
        <edge id="e0" start="start" end="n1" />     Edge identification and starting/ending vertices
        <edge id="e1" start="n1" end="n2" />
        <edge id="e2" start="n2" end="n3" />
        ...
  </edges>
</videoGraph>
```

Figure G.1: template of VideoGraph xml file

```xml
<?xml version="1.0" encoding="utf-8" ?>
<decisionGraph id="Capuchinho" zLayer="1">
	<vertices>
		<vertex id="start"/>
		<vertex id="n1"/>
		<vertex id="n2"/>
		...
	</vertices>
	<edges>
		<edge id="e0" start="start" end="n1">
			<event id="start" />
		</edge>
		<edge id="e1" start="n1" end="n2">
			<event id="verticalStop" />
			<event id="verticalSlow" />
			<event id="verticalFast" />
			<event id="horizontalStop" />
			<event id="horizontalSlow" />
			<event id="horizontalFast" />
			<event id="diagonalStop" />
			<event id="diagonalSlow" />
			<event id="diagonalFast" />
		</edge>
		<edge id="e2" start="n2" end="n3">
			<event id="out" />
			<event id="verticalSlow" />
			<event id="verticalFast" />
			<event id="horizontalStop" />
			<event id="horizontalSlow" />
			<event id="horizontalFast" />
			<event id="diagonalStop" />
			<event id="diagonalSlow" />
			<event id="diagonalFast" />
		</edge>
		...
	</edges>
</decisionGraph>
```

Possible events that could change the flow of the performance of the character

Figure G.2: template of DecisionGraph xml file

Appendix H

Evaluation

Table H.1: Generic information

Variables	participant 1	participant 2	participant 3
Participant name	Leticia	Silvana	Duarte
Participant sex	Female	Female	Male
Participant age	7	8	6
Story name	Little Red Riding Hood	Little Red Riding Hood	Little Red Riding Hood

Table H.2: Assumptions about the participant

Variables	participant 1	participant 2	participant 3
Has knowledge of new technologies	no	no	yes
Knows the well-known story	yes	yes	yes
Usually play with action figures	yes	yes	yes
Is very active	no	no	yes

Table H.3: User experience/interaction

Variables	**participant 1**	**participant 2**	**participant 3**
Shows confusion	yes	sometimes	sometimes
Shows happiness	sometimes	sometimes	sometimes
Shows frustration	yes	sometimes	sometimes
action figures are contextualized	yes	yes	yes
showed concentration	sometimes	yes	yes
Clearly see contexts	yes	yes	yes
Direct relationship between map and history	yes	yes	yes
Think the map is ambiguous	no	no	no
tried to make an animated film	no	no	no
can activate the behavior they intend	rarely	sometimes	sometimes
can put the VC in the intended context	sometimes	often	often
Size relation between action figure and map	good	good	good
User looks at action figure percentage	30%	10%	10%
User looks at screen percentage	70%	90%	90%
Context is sufficiently explicit	yes	yes	yes