Exploitation of Novel Multiplayer Gesture-based Interaction and Virtual Puppetry for Digital Storytelling to Develop Children's Narrative Skills

Hui Liang*
Communication University of China, China

Jian Chang †
National Centre for Computer Animation
Bournemouth University, UK

Shujie Deng
National Centre for Computer Animation
Bournemouth University, UK

Can Chen
Changzhou University, China

Ruofeng Tong
Zhejiang University, China

Jianjun Zhang
National Centre for Computer Animation
Bournemouth University, UK

Abstract

In recent years, digital storytelling has demonstrated powerful pedagogical functions by improving creativity, collaboration and intimacy among young children. Saturated with digital media technologies in their daily lives, the young generation demands natural interactive learning environments which offer multimodalities of feedback and meaningful immersive learning experiences. Virtual puppetry assisted storytelling system for young children, which utilises depth motion sensing technology and gesture control as the Human-Computer Interaction (HCI) method, has been proved to provide natural interactive learning experience for single player. In this paper, we designed and developed a novel system that allows multiple players to narrate, and most importantly, to interact with other characters and interactive virtual items in the virtual environment. We have conducted one user experiment with four young children for pedagogical evaluation and another user experiment with five postgraduate students for system evaluation. Our user study shows this novel digital storytelling system has great potential to stimulate learning abilities of young children through collaboration tasks.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Prototyping, Interaction styles, User-centered design K.3.1 [Computers and Education]: Computer Uses in Education—Computer-assisted instruction (CAI)

Keywords: virtual reality, interactive storytelling, virtual puppetry, gesture-based control, children learning

*e-mail: lianghui@cuc.edu.cn Currently Hui Liang is also a research fellow at the National Centre for Computer Animation, Bournemouth University, UK for a two year research programme "AniM" [FP7-IIF-623883] as an EU funded incoming fellow.

†Corresponding author: jchang@bournemouth.ac.uk

1 Introduction

Storytelling is essentially one of the original teaching forms and commonly used as an effective method to teach ethics, values, and cultural norms and differences. As the modern form of traditional storytelling, digital storytelling systems have received increasing attention and demonstrated useful pedagogical functions. In the environment of digital storytelling system, young children can express themselves and cooperate with other players. There has been substantial amount of research undertaken on digital storytelling [Psomos and Kordaki 2012], which mainly focused on the ability of narration. Digital storytelling can also help disabled children with learning difficulties to remove the communication barriers by providing them with the opportunities to play active roles [More 2008], as well as support the development of a wide spectrum of cognitive functions and skills [Garzotto et al. 2010].

Virtual Reality (VR) technology is currently used in a broad range of fields, such as games, movies, military simulations, and healthcare. Considered as one of the powerful and promising education tools, VR technology also enables a wide range of interaction options for digital storytelling and provides a variety of tools and techniques for exploiting and strengthening the adoption of digital storytelling. In the virtual storytelling environment, users get engaged, perform learning tasks and experience activities [Hofer and Owings Swan 2006]. *ShadowStory* [Lu et al. 2011], for example, is a digital storytelling system inspired by traditional Chinese shadow puppetry. Children use handheld sensors to manipulate shadow puppetry to perform stories on a projection screen.

At present, the young children, also known as the "new media generation" or Digital Natives [Thompson 2013] were born in a rich media society and they started to interact with new technologies from an early age. Saturated with digital media in their daily lives, they need interactive learning environments, and multiple forms of feedback and meaningful learning experiences [Barnes et al. 2007], which bring new challenges to current digital storytelling. The new generation of the digital storytelling system is expected to offer a novel and immersive way, which is more engaging and interactive, to captivate learners' interests like never before and improve the quality of teaching and learning in the virtual story world.

One of our main concerns is how to provide this "new media generation" an engaging and immersive interaction to accelerate their learning progress by involving novel Virtual Reality technologies. From academia to industry, a wild range of new VR technologies and novel interaction applications are available, which encourage us to exploit potential novel interactivity in the form of digital storytelling for enhanced engagement and

enjoyment. Various technologies and new hardware can facilitate a synthetic, highly interactive VR environment, for instance, motion sensing technology and Microsoft's Kinect device.

It is important to captivate young players' interests in narrating with a vigorous representation. Interactive storytelling has long been synergetic with puppetry. Researchers have explored the possibilities offered by multi-modal and cooperative interactions with virtual puppets in constructing a communicative experience through gestures and voice. Studies in storytelling systems with user created and controlled digital characters have proved that they can improve creativity, collaboration and intimacy among children. Therefore we consider digital puppetry as a proper medium to assist children's narration in terms of emotional stimulation, where virtual puppets act as avatars to perform stories in virtual storytelling environment.

Not only focusing on narration as what previous research did, our research also devotes to increasing children's abilities of cooperation with peers by introducing interaction of multiple players, as well as abilities of cognitive development and motor coordination during their storytelling performance through hand gesture control. Storytelling is also a powerful cognitive tool with potential to support learners in the process of meaning construction.

In this paper, a prototype of an interactive multiplayer gesture-based digital puppetry storytelling system is developed for young children. Two digital puppets are animated to assist storytelling in VR environment, which make the system visually and interactively appealing. The depth motion sensing technology is utilised as the gesture-based HCI. By hand tracking and gesture recognition, young children could manipulate virtual puppets to perform story in an intuitive and natural way. Addition to the development of children's narrative skills, we also devote to improving young children's other abilities including cognitive development, motor coordination ability and collaboration ability through their storytelling performance.

Various VR technologies, such as motion sensing technology of gestural-based control, computer animation for content synthesis and cloud-based multi-player interaction, are integrated in the system. In summary, this work has four main contributions:

- Implemented a prototype of the digital storytelling system by developing and integrating the novel VR technologies;
- Introduced a novel hand gesture based interaction in VR by using deep motion sensor devices to meet the requirements for digital storytelling;
- Designed virtual puppets and their animation for digital storytelling to assist children's narration;
- Introduced the cloud-based network connection for multi-player narration in virtual environment with low latency.

The remainder of this paper is organised as follows: Section 2 briefly reviews the related research works; Section 3 presents our system design as well as the main technologies involved; result is discussed and analysed in Section 4; Section 5 concludes the paper.

2 Related Work

There has been a considerable growth of VR technologies over the past decade. VR has been applied in diverse domains and for a variety of purposes, including pedagogical practices, simulation and training, virtual experiments, and game development. One of its key features is interactivity, i.e. the ability to detect user inputs and modify virtual world instantaneously. Interaction interfaces such as head tracking, motion detection and more immersive or intuitive technologies that are using hand gesture as input has attracted growing interests.

With the advantages of the HCI development, virtual reality technologies are considered to be one of the most powerful tools for supporting learning process [Jonassen 1999; Smeets 2005] by providing a synthetic, highly interactive 3D virtual environment and allowing students to experience virtual learning environments in a natural way. With the help of modern VR technologies, digital storytelling emerged as the modern and challenging evolution of traditional storytelling over the last few years as a powerful teaching and learning tool. Examples such as *FaTe2* [Garzotto and Forfori 2006] and *MyStoryMaker* [McKinley and Lee 2008] have utilised 3D virtual environments in digital storytelling.

Lots of VR interactive technologies are involved in digital storytelling, e.g., the tangible interfaces, mobile-based authoring system, haptic feedback and wireless handheld orientation sensors. *Toontastic* [Russell 2010] can be considered as a collaborative and constructive digital animation creator that was designed to help children capture and share their stories with other children around the world. *Storytelling Alice* [Kelleher 2006] introduced computer programming to learners using the approach of 3D animated stories. *Wayang Authoring* [Widjajanto et al. 2008] enabled diverse storytelling styles under various cultural backgrounds for the children to create digital stories by using virtual puppets.

Picture This! [Vaucelle and Ishii 2008] is a HCI device which embedded a camera and a gesture sensor in children's toys for video composition. It works as a video storytelling performance system allowing children to capture stories with physical artefacts at different levels of interaction among different age groups. Children could craft videos about character toys while their gestures and play patterns was analysed by the system. It could benefit young children from developing the spatial and visual coordination while playing with their favourite toys and telling stories.

ShadowStory [Lu et al. 2011] is a digital storytelling system inspired by traditional Chinese shadow puppetry. It was designed for children to use a Tablet to create digital shadow puppets and perform story corporately on a projection screen. To simplify the control mechanism of the shadow puppetry, a pair of handheld wireless 3D orientation sensors (one in each hand) is used to manipulate the character: tilting the First sensor to move the character left, right, up, and down, and tilting the second sensor leftwards or rightwards to let the character bend the body in the corresponding direction, as if bowing or looking up. This kind of interaction seems not intuitive or natural enough because of the direct manipulation on the puppet using handheld controllers.

Leap Motion controller was released in July 2013, as a non-invasive solution using infrared sensing technology for hand tracking and gesture recognition. It is designed to track high fidelity information of fingers' positions and orientations through a small depth sensor and translate hand movements into computer commands. There already exist several successful applications that have integrated the Leap Motion devices, such as the *AirSpace* [https://airspace.leapmotion.com] which allows users to interact with the Leap Motion as a musical instrument. Leap Motion functions are integrated in our system to offer natural gesture inputs and control the virtual puppets during storytelling.

3 System Design

To create an engaging and interactive storytelling environment for the "new media generation", we needed to resolve two main challenges in our system design: how to provide a novel interaction to young children and how to design and utilise virtual puppets to assist children's narration. To solve these problems, various technologies are integrated in our system, including the motion sensing technology of gesture-based control, computer animation for content synthesis and cloud solution for multiple players.

3.1 Conceptual Model
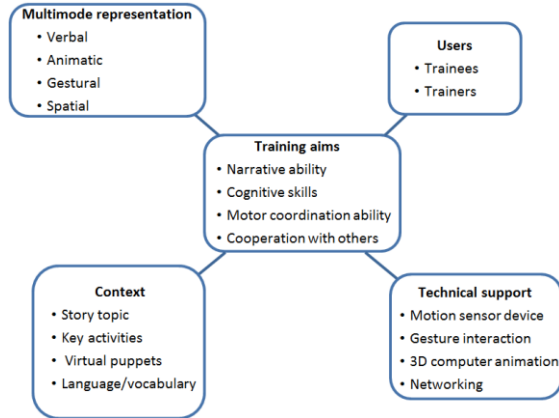
Figure 1: *Components of our conceptual framework*

The components of the conceptual framework of our system are illustrated in Figure 1. The training aim is the kernel component, four other components are involved: users, context, representation and technical support, which will be discussed separately later. In our system, we provide young children with an intuitive interface by utilising hand tracking and gesture recognition technologies. Young children can naturally use hand gestures to manipulate virtual puppet to perform storytelling. This novel interaction provides the children with new horizon of the training of their cognitive skills and motor coordination abilities, as well as the ability of collaborating with others in addition to oral narration.

"Training aims" is the core component of our conceptual framework. Just like the aims of the majority of current digital storytelling systems, narrative ability training is one of our main considerations in system design. There are lots of research works in the psychology field that deal with learning and using technology in educational process, e. g., behaviour analysis [Greer 2002; Vargas and Vargas 1991] and teaching technology in earlier ages [Skinner 1968]. Beside this, some researchers have engaged in the development of some other important skills of the young children, such as cognitive ability development, motor coordination training and the ability of cooperation, which might shed insights on the development of children's other abilities in our system. For example, the abilities of constructing mental representation of spatial-temporal environment and spatial visualisation are important qualities to achieve success in science, technology, engineering and mathematics [Wai et al 2009]. Also preschoolers' visual spatial attention ability predicts their future reading skills [Franceschini 2012]. With poor motor coordination ability, children tend to underachieve educationally and develop difficulties in building peer relationships [Skinner and Piek 2001]. Cooperation is the children's ability to work or act together with others for achieving synergy, as opposed to working

in competition.

With the novel HCI provided by our system, young children could intuitively use hand gestures to manipulate the puppet agents - "crow" and "fox" in our demo, and interact with virtual objects in the virtual scenario. The puppets actions, such as flying, grasping and talking, can only be successfully performed by specific hand movements and predefined hand gestures. Through the procedures of storytelling, their narrative ability will be enhanced. Using hand gestures for controlling the avatar, their motor coordination ability can be trained. In interaction with virtual items that have different properties and roles in the story, the children's space and object recognition capability will also be benefited.

As illustrated in Figure 2. The training aims cover four fundamental abilities: narrative ability, cognitive skills, motor coordination ability and the ability of cooperation with peers. They were realised through different related training activities in the digital storytelling practices.
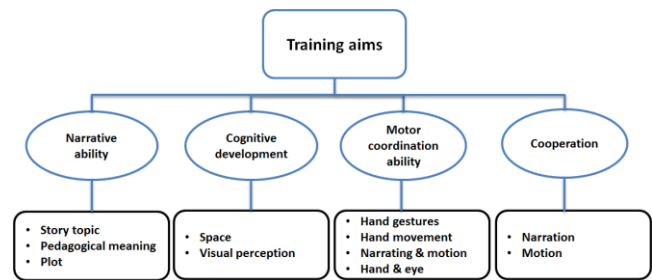

Figure 2: *Training aims*

3.2 Storyline

Aesop's story "the Fox and the Crow" is a great fable that teaches children an important lesson in life – about flattery and deception. Considering its popularity among young children and its positive pedagogical meaning, we chose this story to implement and to provide children an exciting and educational interactive storytelling experience with our proposed framework. The sketch of the plot outline is presented in Figure 3 which is also provided as a useful hint when children are telling the story and interacting with the virtual puppets.
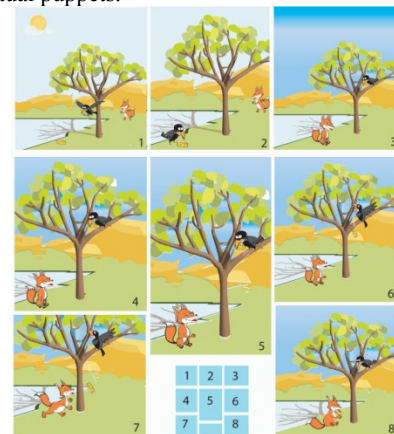

Figure 3: *Sketch of the storyline*

In our interactive storytelling system, players are not required to follow the direct causality pattern. The structure of the entire storytelling can be considered as a graph consisting of multiple story "nodes". The connections between the key "nodes" push the

story forward. We listed the key "nodes" of the story "the Fox and the Crow" as shown below:

- Control the puppet fox and the puppet crow to search for food
- Perform how the fox flatters the crow to defraud the cheese
- Help the crow reveals the fox's attempts to trick her
- Describe how the crow finally gives in and drops the cheese
- Present how regretful the crow is

The concept model of the system storyline is illustrated in Figure 4. The main storyline consists of multiple key plot points accompanied by players' hand movement and interaction. At the key story points, a tutor would gently interrupt and remind the young players to stop and take their time to narrate at the "gap". The players then could organise their words and describe the key points from their own understanding of the story.
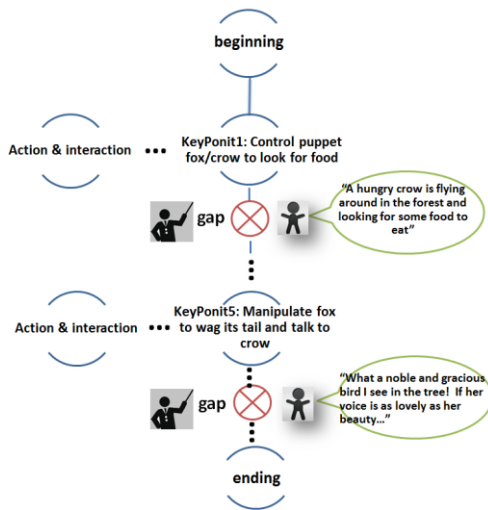
Figure 4: *Concept model of the storyline*

3.3 System Overview

To offer a better interaction and cooperation, our system provides multi-player support which enables two children to play together (with/without supervision) or one on one attention (one tutor and one student) cooperatively. The prototype involves two virtual characters, which provides young children both the opportunity of manipulating the puppet and the opportunity of playing with peers.

We applied Leap Motion as the depth sensor in this system for its specialties in hand tracking accuracy and gesture recognition advantages. It is more cost effective and space saving than other sensors, such as Kinect. However, multiple Leap Motion Controllers cannot be running on one computer. To solve this problem we added networking and multiplayer support using cloud resource to construct a distributed architecture. The concept of the distributed system architecture is shown in Figure 5. In this system, players interact with their own Leap Motion controller at their client computer locally. The control command will be uploaded and exchanged between the local client computers via the remote server, and the whole game scenario will be synthesised and displayed on each client's screen. It enables players to participate simultaneously via internet connection just like the experience of Massively Multiplayer Online Game (MMOG).

The interactive pipeline of a local client is shown in Figure 6. First, a story board is provided as the storytelling hints to the players. Second, the player uses hand motion and hand gestures to

manipulate an avatar for interaction. Depth image data from the motion sensor are obtained and interpreted into motion control commands by the client computer. Finally, as the visual feedback, avatar's responding animation is provided to players and then, players adjust their hand gestures/movement to push the plot forward and narrate the story.
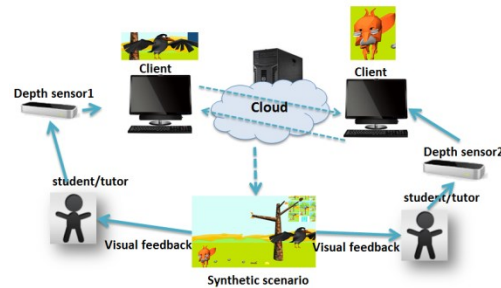

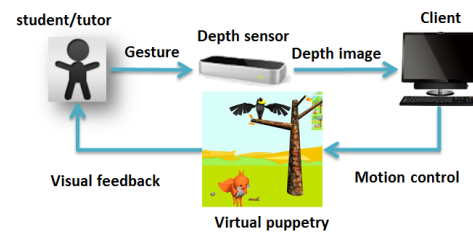**Figure 5:** *Concept model of the distributed architecture*

Figure 6: *Interaction pipeline*

Technique framework is represented in Figure 7. The Photon Cloud [https://www.photonengine.com] is used as the under layer cloud computing infrastructure to provide multiple players supporting, as shown in Figure 5. Photon Unity Networking (PUN) is a Unity 3D client plugin that is introduced to integrate the network functionality on top of Photon Cloud. Unity 3D Game Engine is the core component of the system. As the system implementation platform, it integrates the Phone Cloud framework with digital puppetry model, audio, animation and user's hand gesture interaction and provides technical support for our training aims at the top level.
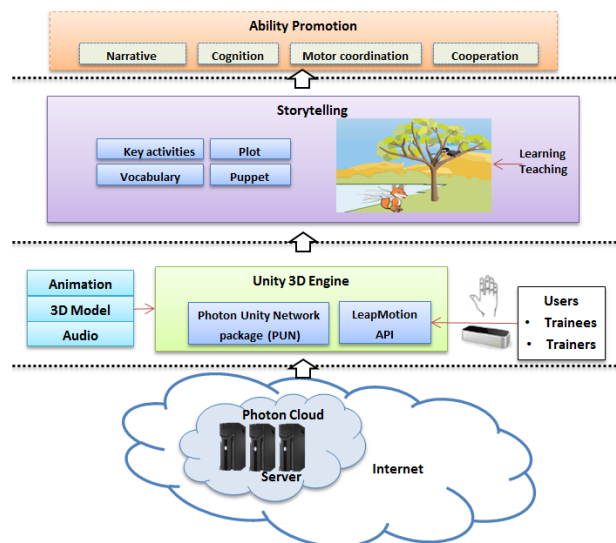

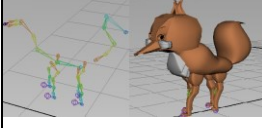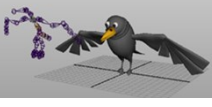**Figure 7**: *An overview of the technique framework*

3.4 Virtual Puppetry Design

Puppetry presents a highly developed form of expressive interaction in our implementation. It depends on motion gestures as the "input" to control a wide range of character's animation. The player and puppet reflects a relationship of actor and avatar.

The benefits of creating and manipulating digital characters for storytelling to improve children's creativity and collaboration have been explored by previous work [Lu 2011]. The digital puppets are developed to make our system more appealing for young children, due to puppetry's positive educational influence as a lively presentation.

As shown in Table 1, stylised cartoon characters of a cute fox and a doltish crow are designed in Maya 2014 according to children's favour, including their sketches, skeleton and 3D models.

Table 1: *Sketch & 3D model of virtual puppetry*

	Fox	Crow
sketch		
3D model		

To offer more engagement and interaction to captivate students' learning interests, we designed characteristic actions for the two puppets, including flying, walking and talking, and these actions can be triggered and performed corresponding to players' hand gestures. To provide real-time response and live visualisation, these actions are integrated as a set of pre-recorded animation clips. Furthermore, these pre-recorded animations associated with puppets' particular actions can only be triggered by specific events of hand movement and hand gestures. The mapping between animation and hand gestures is illustrated in Table 2 in section 3.5.

3.5 Hand Gesture-based Interaction

Many input methods are widely used in HCI, such as keyboard, mouse, touch screen, voice control and hand gesture interaction. We used Leap Motion to capture hand motion. It provides new possibilities for deriving educational benefits of storytelling by creating new modalities of interaction through hand gestures, which can provide more intuitive and natural user experience.

Once the Leap Motion controller has been mounted, a connection between the controller and our system will be automatically established and then the motion tracking data will be passed through. The motion tracking data is presented to our application as a series of snapshots, which is called as the Frames. By programming with the API proved by Leap Motion, the hand positions and orientation as well as the fingertip positions can be accessed.

The motion tracking data accessed from the sensor controller can be represented as a tetrad illustrated as follows:

$$TrackingData = \langle F, HL, PL, GL \rangle \qquad (1)$$

F indicates the frame information provided by the Leap Motion controller. Each frame contains the measured positions and the following information about each entity detected in that snapshot recorded by the controller. HL represents the list of detected hands, PL represents the list of pointables (such as fingers or finger-like tools), and GL indicates the list of recognised gestures. The Leap Motion device recognises certain movement patterns as gestures which indicate user's intent or command.

The tracking device provides model of hands with five fingers of bone structures and each finger is made of four bone joints. We could design specific interactions by using the positions and rotations information of each finger bone provided by the Bone API. The data model of hand motion can be represented as follows:

$$HandMotion = \langle Movement, Gestures \rangle \qquad (2)$$

where $Movement = \langle m_1, m_2, \dots m_i, \dots m_n \rangle$ including different hand movement, such as move up and down; $Gestures = \langle g_1, g_2, \dots g_i, \dots g_m \rangle$ which vary with different puppets; each gesture is composed by a set of joints and their motions.

How to design hand gestures for the purposes of ease of use and intuition is of particular importance for HCI, which involves various considerations, such as the natural language, young players' physical and memorial limitation, recognition accuracy of the input device, and the assumption of the task in storyline. To provide a natural and intuitive way for young children to interact with the virtual toys, we designed our hand gestures vocabulary elaborately. There are in total two sets of hand gestures ("Steering" gestures and "Trigger" gestures) and each set is attached with one hand: "Steering" gestures are assigned for navigation tasks and "Trigger" gestures are allotted for action performance.

The most straightforward way for young children to control puppets' movement is mapping the translation of children's hand position in the gesture controller's coordinate system to the puppet's position in virtual environment, which is called the navigation task. This task is designed to be accomplished by a set of "Steering" gestures performed by one hand (the red hand), as illustrated in Table 2(a). In our definition, such hand is in the state of "Steering Mode". When young players try to control virtual puppets' movement in the virtual environment, they just need to move their stretched palms intuitively. The virtual puppet's position in virtual workspace will move accordingly as shown in Equation (3):

$$C_{avatar} = T \cdot S \cdot R_{hand} + C'_{avatar} \qquad (3)$$

where C_{avatar} represents the virtual puppet's position coordinate in current frame, R_{hand} is the player's relative hand movement in real world, i.e. the displacement of hand positions in current frame and the previous frame, which can be obtained in Equation (1), T is the transformation matrix between the player's workspace and the avatar's coordinate system, S is the scaling matrix, and C'_{avatar} represents avatar's position in the previous frame. For more details of hands movement control mechanism and gesture recognizing, please refer to our previous paper [Liang 2015].

Besides the navigation, virtual puppets' actions are controlled by another set of predefined "Trigger" gestures performed by another hand (the blue hand), as illustrated in Table 2(b). This hand is defined in the state of "Trigger Mode". The assignment of these "Trigger" gestures to computer commands are designed by following the principles of natural language as well as considering the young players' limitations of hand postures. Each "Trigger" gesture representing different puppet's action is associated with different pre-recorded animation respectively, such as grasping to talk.

Table 2(a): *Predefined "Steering" hand gestures*
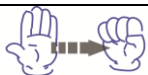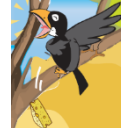
Gestures	Illustration	Description
1. Leftward		Fly/run to the left
2. Rightward		Fly/run to the right
3.Upward		Fly/jump up
4.Downward		Fly/jump down
5. Hover/stay		Hover in the air/stay unmoved

Table 2(b): *Predefined "Trigger" hand gestures*

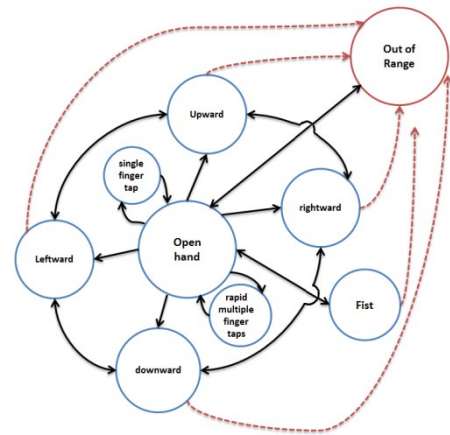| Gestures	Illustration	Description
6. Grasp	Hand open to close	
7. Drop	Hand close to open	
8.Sing/talk	single finger tap	
9.Swipe	rapid multiple finger taps	Waging tail, flattering

Figure 8: *Concept model of hand gestures transition*

The analysis of the possible motion or gesture transitions in our system is illustrated in Figure 8. Additionally, after modified for the specific cases, this kind of concept model could also be adapted in other applications which possibly have different set of hand motion or hand gesture transitions. An obvious benefit is that it could better separate hand motion from hand gestures to reduce the probability of recognition errors as well.

3.6 Cloud Solution to Support Multi-user

Multiplayer supported storytelling will provide approaches and environments for young players to tell stories collaboratively through their actions and allow multiple means of problem solving. Since there were two characters involved in our demo - the "fox" and the "crow", it needed two players to play corporately. It requested the system having the ability to support multiplayers interaction. However unfortunately, the Leap Motion devices do not support multiple sensors connecting to a single computer, which has been a recurring question with no direct answer in Leap Motion developing community.

We solved this problem by integrating a Cloud service provided by Photon Unity Networking to realise one gesture controller to one virtual character. This network has the ability to involve many real entities interacting in real time in the same virtual environment. It presents the solution for synchronising the local manipulation of the virtual puppets to other clients in real time. The storytelling facilitated with multiplayer network connections encourages the players to develop individual and collective strategies. The platform where our digital storytelling system running with features of multiplayer over the cloud is illustrated in Figure 9.
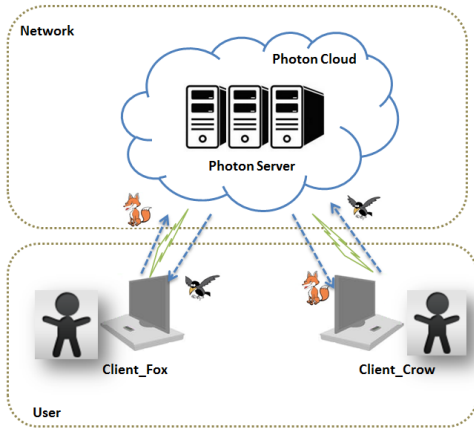
Figure 9: *Multi-user support over the cloud*

For data transmission, firstly, the puppet's information in each frame is serialised, which includes the vectors of puppet's position, the quaternion of puppet's rotation, and the animation state as well. For communication and easiness to be parsed, before transmission, this information is compressed and optimised by binary protocol. After encapsulation, the serialised data is transmitted from the client to the Photon Server using User Datagram Protocol (UDP).

Network Quality of Service (QoS) has a significant impact on our Cloud-based cooperative storytelling system which requires the real time interaction, especially with the virtual puppetry. The whole system response delay could be segmented into three components: the *network transmission delay*, the *server's processing time* and the *client's processing time*, as illustrated in Figure 10. The *network transmission delay* is the summarisation of the time consumed by delivering the local virtual puppet operation data from client to the Phone Server and the time consumed by sending back the synthesised game information from the server to the client. It is usually referred to the network round-trip time (RTT). The *server's processing time* represents the time consumed by the server to handle the incoming data, which is the time interval between the time that the server receives the incoming virtual puppet operation information and the time it responds the client after processing the received information. The *client's processing time* represents the time that the client consumes to decode the feedback data from the server and presents the incoming synthesised game information on the client's screen.
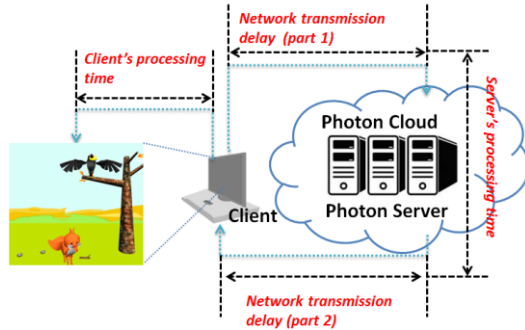

Figure 10: *Latency of the network platform*

To investigate the impact of delay on the cooperative storytelling, a latency test is conducted at the same time with the system experiment (discussed in Section 4.2 in detail). Firstly, we measured the actual readings of the latency as the objective observation in different time periods (in the morning and in the afternoon), the latency mostly varies under 250ms. Further, we tried to find out whether or not players could notice this degradation. The result was that players seldom perceived a delay up to 200ms. However, if the delay exceeded 300ms, they reported the system became annoying and unacceptable. Our studies have shown that the cooperative digital storytelling system which actively requires interaction with other players is sensitive to the network delay when exceeding the tolerable threshold. In fact, according to our observation, in most circumstances, the Photon service could transfer data and messages synchronously in real-time. The latency is small compared to the whole loop of distributed game execution.

There exists some studies focusing on the impact of delay on real-time multiplayer games, but few have addressed the related issue in digital storytelling so far. We are the first that studied the influence of the delay on the cooperative digital narration. Although the result was only obtained from our own system and further analysis was expected, we believe that it is applicable to other distributed storytelling system (such as the mobile based storytelling and the web based storytelling) involving similar types of interaction between players.

4 Results

4.1 Prototype of the Experiment

The experimental prototype of our system is presented in Figure 11. As shown in Figure 11(a), one young boy was controlling the puppetry fox, and on the right, an adult was manipulating the puppetry crow to assist the narration as a tutor. A scenario performed by two children corporately is shown in Figure 11(b). Figure 11(c-e) indicates how excited the young player was when performing storytelling through this novel interaction.

The performance of each story point involves a set of hand gestures, which can only be successfully performed by specific hand movement and touch off puppets' related actions. Then the pre-defined animations associated with puppets' particular actions can be successfully triggered. Two examples of scenario in the virtual storytelling environment and the involved hand gestures are illustrated in Table 3. The screenshots of the story corporately narrated by two young boys is illustrated in Table 4.

(a)   (b)

(c)   (d)   (e)
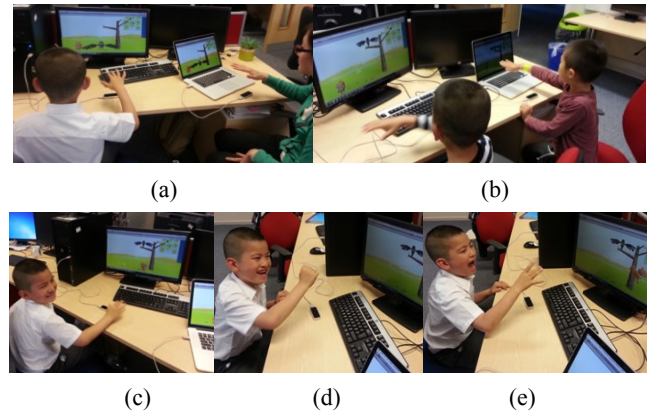
(c) Fox is flattering Crow. (d) Fox is so glad because Crow drops the cheese! (e) Fox is so excited for getting the cheese!!

Figure 11: *Experimental prototype*

Table 3: *Example of Hand gestures used in scenario*
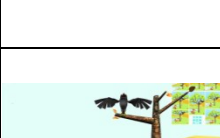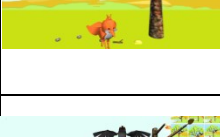
Script	Fox and crow search for food. Crow picks up cheese once he finds it.
Scenario	**Involved gestures**
Key point 1->2	Leftward, Rightward, Upward, Downward, Hover, Talk, Grasp
	Leftward, Rightward, Upward, Downward, Step, Talk
Script	Fox flatters the crow trying to defraud crow's cheese. Crow reveals the fox's attempts.
Scenario	**Involved gestures**
Key point 5->6	Talk
	Leftward, Rightward, Upward, Downward , Stay, Step, Swipe, Talk

Table 4: *Example of key points performed by young players*

Interactive virtual scene	Narration
	(Aside): One day, a fox and a crow were looking for food in the forest… Fox: *I'm so hungry...* Crow: *I need some food… What's that? … Woo!!! Cheese!!!*
	(Aside): The crow found a cheese… Fox: *Wait a minute! That's for me!!!...* Crow: *Got it!!!*
	(Aside): The fox tried flattery… Fox: *Good day, Dear Crow…What a noble and gracious bird...* Crow: ... Fox: *How glossy your feathers… How bright your eye...* Crow: ...
	(Aside): The fox tried again and again… Fox: *Let me hear but one song from you that I may greet you as the Queen of Birds…* (Aside): Tricked by the flattery, the crow opens her beak to sing… Crow: *Caw… Caw*
	(Aside): The cheese was dropped to the ground… Fox: *That's all I wanted! Thank you...* Crow: *My cheese…I'm wrong… I'm so foolish… I would not trust flatterers anymore…*

## 4.2 Pedagogical Evaluation

A pilot study participated by four 5-7 years old young children (mean 6 years and 4 months old) has been conducted for pedagogical evaluation with permission from their parents who were informed about the nature of the study and its purpose. Five adults volunteered in another test for system evaluation which is elaborated in the next section. All participants were trained to familiarise themselves with the gesture controller after detailed explanation of the experiment. All children were interviewed regarding their subjective assessments when they completed the recorded experiments.

In the experience, the children are divided into two groups to narrate the story cooperatively. Each group has carried out five trials of the "Fox and Crow" story. After the learning process, following the provided key story points, all of the children easily got engaged with the linguistic game and built their own story by interacting with the digital puppets with great interests. For the purpose of pedagogical evaluation, we primarily set up some metrics which was recorded as the result of each round of trial. The result is presented in Figure 12, which showed improvement in the following aspects:

Narration ability. As shown in Figure 12(a), the children tended to use more descriptive words for narration across the five rounds. For example, "A crow is looking for some food" in the first round was rephrased to "A hungry crow is flying around in the forest and looking for some food to eat" in the fifth round. Figure 12(b) showed a decrease in narration time followed by an increase. The decrease in completion time showed a learning effect of the interaction manner and the increase indicated that the children were paying more attention on the narrating and rhetoric, which is vital for storytelling and more suitable for pedagogical purposes as a learning tool for children.

Cognition ability. The decrease of times of average locating tries and wrong gestures performed shown in Figure 12(c, d) suggested that both the abilities of spatial cognition and visual perception got practiced and improved.
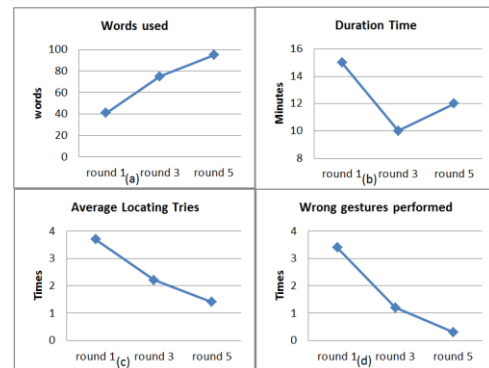
Figure 12: *Average values of metrics in experiment*

Motor coordination. The decrease of times of average locating tries and wrong gestures also suggested that the children could coordinate the position of their hands with the moving puppet more smoothly and steadily with the predefined gestures.

Collaboration ability. The collaboration in our shared virtual environment is twofold: collaborative narration and collaborative manipulation. The collaborative narration involves dialogues between the two players which develops the storyline. In the meantime, to trigger a specific virtual event, it requires both virtual characters to satisfy certain movements, in terms of positions and orientations. For instance, the fox could only pick up the cheese when the crow was on the tree and had already dropped the cheese. Clearly the required movements are gesture-

based control which demands collaborative manipulation between the two players to make sure each character is at the triggering position facing the designed direction.

It was reported by the children that the collaboration increased the task difficulty compared to a single player story, as in the pre-trial practice. The fluency of the storytelling was often interrupted when one of the players made wrong gestures or oriented to a wrong position. However, it was noticed that all the children were more engaged and excited in the collaborative storytelling. Some participants even came up with different tricks to defraud the cheese in each trial. It suggested that the collaboration demanded more attention because the participants needed to react according to their interlocutor's dialogue and action in addition to the narration and manipulation of their own virtual characters.

4.3 System Evaluation

Considering young children's knowledge limitation on the technologies involved in the system, we asked postgraduate students to take part in the system evaluation instead. Five 22-28 years old postgraduate students who were naïve of this experiment participated in the system evaluation test. Each student did six trials with the experimenter, three of which they acted as the "fox" and the other three they acted as the "crow". A questionnaire was given after the trials, regarding acceptability, interactivity as well as their suggestions for improvement. The Likert scales were used to evaluate acceptability and interactivity. An open-ended question was given for suggestions to improve our experiment design.

4.3.1 Acceptability

There are seven criterions given for evaluating the acceptability, ease of use, precision, stability, naturalness, handiness, freedom of movement and effectiveness. The participants are asked to rate each criterion from 1 to 5, indicating "Poor satisfaction" to "Excellently satisfied". Figure 13 shows the result of the mean scores of each criterion about the acceptability of our system. The ease of use, naturalness, freedom of movement, handiness and stability has returned positive feedback with score around four, especially the ease of use and naturalness have achieved the highest satisfaction. This is possibly due to our elaborated hand gestures design as the method of natural interaction. However, the effectiveness and precision received the lowest satisfaction which could be induced by the recognition accuracy limitation of the depth sensor. For example, we have noticed that when the fingers folded over the hand or overlapped each other could lead to false detection and the hand position relative to where the sensor was is also crucial. The perfect position of the tracked hand is when the palm is perpendicular to the top surface of the device. Any other position may affect the accuracy of hand tracking.
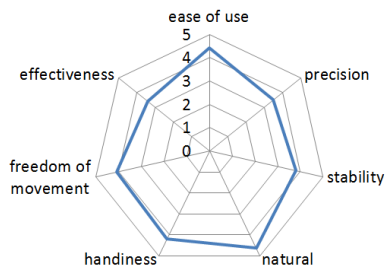

Figure 13: *Mean scores for each criterion of acceptability (1-Poor, 2-Fair, 3-Average, 4-Good, 5-Excellent)*

4.3.2 Interactivity

The evaluation of interactivity was rated in terms of Steering gestures and Trigger gestures with the following criterions: effectiveness, intuitiveness, simplicity, precision and quickness. Similar with the evaluation of the acceptability, participants rated each criterion scaled from 1 (Poor satisfaction) to 5 (Excellently satisfied) whose mean scores are shown in Figure 14.

Overall the Steering gestures returned better user experience than the Trigger gestures. It is probably due to the ease and intuitiveness of the directional navigation design and participants could understand the interaction approach without learning. The Steering has achieved almost perfect scores in effectiveness, intuitiveness and simplicity. Weaker but still satisfying scores were obtained in quickness and precision. Three participants noted that the steering is too sensitive and not very accurate over short distance and four participants mentioned excessive speed. This indicated the sensitivity and moving speed of the virtual characters needed to be tailored depending on users' habit or expectation. The relative lower rates of the Trigger gestures revealed weak effectiveness and precision and slightly better quickness, intuitiveness and simplicity. To improve on the Trigger gestures effectiveness, it requires most simplified gesture design for the ease of recognition by the device. This is a dilemma that complex gestures can enrich the control of virtual characters as well as the story, but it will also cause difficulty of recognition by the device. Movement is probably considered more acceptable than the other interaction modalities because it is the easiest and the most intuitive action for young children without any other complex manipulation.
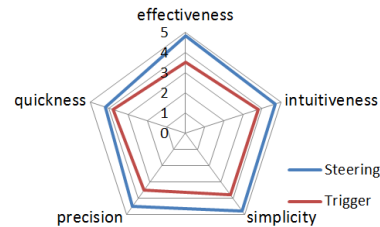

Figure 14: *Average scores for each criterion of interaction modalities (1-Poor, 2-Fair, 3-Average, 4-Good, 5-Excellent)*

4.3.3 Improvement Suggestions

Several participants mentioned the issue of tiredness when raising hands in the mid-air without support. This has been a common issue for all gesture-based applications especially when long periods of operation involved. There is yet no optimal solutions reported but it can be improved by gesture design and even assistance from multimodality. For instance, Song et al. [Song et al 2014] designed a CAD modelling approach combining hand gestures and gaze gestures in zooming tasks.

Although the interaction were reported intuitive, some participants still suggested extending the learning process or improving the hand gestures design, because current experiment settings can cause misoperation and accuracy issues. An ideal intuitive interaction approach is expected no learning process or mild practice. However, it is difficult to achieve because different user has different habits, it will require our system to be adaptive to associate users' intention and their gestures. A richer gesture database is necessary for the Trigger mode and we need to train the system to learn the users' habits for this purpose.

5 Conclusions

In this paper, we have developed a prototype of a novel multiplayer interactive digital storytelling system assisted with virtual puppetry to provide natural interaction/control and immersive experience when narrating stories. The system integrates useful VR techniques and novel hardware to fertilise development of young children's abilities, such as narrative skills, cognitive skills, motor coordination and collaboration through storytelling. It uses the motion sensing technology, computer animation techniques and cloud-based network solution. The development of the prototype demonstrates the possibility of combining novel HCI and digital storytelling to inspire the children's teaching and learning practices.

The use of multiple Leap Motion sensors is creatively implemented, where the multiple sensors' data are collected and exchanged through Photon Cloud. Our method can wildly extend the potential usage of this sensor device and open up many new possibilities with the cloud-based network support.

The pedagogical evaluation of the system is preliminary examined in our experiment and the results from the analysis are promising, which showed that young children can be benefited from the participation. During playing, some minor frustrations (or difficulties) are observed, such as dropping the cheese by mistake, having problem to locate the avatar, or using a second hand to support the main hand in operation etc. Getting around these difficulties also made the children feel rewarded and find the game interesting. Aside from the experiment, we also evaluated the system based on adult participants' feedback.
There are some improving works left to be carried on in the future. More hand gestures signs, such as thumb lift, different swipes: one, two and more fingers in four independent directions (up, down, left, right), and different circles: one, two and more fingers in two independent directions (left, right), could also be investigated to provide more complex interaction. Only a limited number of trails are tested in our experiment, we will need to examine more cases and design a psychological experiment to affirm the conclusion. In the experiment, only postgraduate students participated in the system evaluation. For further improvement, the system evaluation experiment will also be conducted with young children. Further validation and analysis of the effectiveness of this approach is needed.

Acknowledgments

References

Available at https://airspace.leapmotion.com

Available at https://www.photonengine.com

BARNES, K., MARATEO, R. C., AND FERRIS, S. P. 2007. Teaching and learning with the net generation. *Innovate: Journal of Online Education*, 3(4), 1.

FRANCESCHINI, S., GORI, S., RUFFINO, M., PEDROLLI, K., AND FACOETTI, A. 2012. A causal link between visual spatial attention and reading acquisition. *Current Biology*, 22(9), 814–819.

GARZOTTO, F., AND FORFORI, M. 2006. FaTe2: storytelling edutainment experiences in 2D and 3D collaborative spaces. In *Proceedings of the 2006 conference on Interaction design and children*, ACM, 113–116.

GARZOTTO, F., PAOLINI, P., AND SABIESCU, A. 2010. Interactive storytelling for children. *In Proceedings of the 9th International Conference on Interaction Design and Children*, ACM, 356–359.

GREER, R. D. 2002. *Designing teaching strategies: An applied behavior analysis system approach*, 1st Edition, Academic Press.

HOFER, M., AND OWINGS SWAN, K. 2006. Digital storytelling: Moving from promise to practice. *In Society for Information Technology & Teacher Education International Conference*, 1, 679–684.

JONASSEN, D. H. 1999. *Computers as mind tools for schools: Engaging critical thinking*. NJ: Prentice Hall, Inc.

KELLEHER, C. 2006. Motivating Programming: Using storytelling to make computer programming attractive to middle school girls (No. CMU-CS-06-171). CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE.

LIANG, H., CHANG, J., KAZMI, I. K., ZHANG, J. J., AND JIAO, P. F. 2015. Puppet Narrator: utilizing motion sensing technology in storytelling for young children. *In Proceedings of the Seventh International Conference on Virtual Worlds and Games for Serious Applications*. (To be published)

LU, F., TIAN, F., JIANG, Y., CAO, X., LUO, W., LI, G., ZHANG X.L., DAI, G.Z., AND WANG, H. 2011. ShadowStory: creative and collaborative digital storytelling inspired by cultural heritage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1919–1928.

MCKINLEY, B., AND LEE, Y. 2008. Mystorymaker. In *Proceedings of CHI '08 Extended Abstracts* (Florence, Italy, April, 2008), ACM Press, New York, 3219–3224.

MORE, C. 2008. Digital stories targeting social skills for children with disabilities multidimensional learning. *Intervention in School and Clinic*, 43.3, 168–177.

PSOMOS, P., AND KORDAKI, M. 2012. Pedagogical analysis of educational digital storytelling environments of the last five years. *Procedia-Social and Behavioural Sciences*, 46, 1213–1218.

RUSSELL, A. 2010. ToonTastic: a global storytelling network for kids, by kids. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, ACM, 271–274.

SKINNER, R. A., AND PIEK, J. P. 2001. Psychosocial implications of poor motor coordination in children and adolescents. *Human movement science*, 20(1), 73–94.

SKINNER, B. F. 1968. The technology of teaching.

SMEETS, E. 2005. Does ICT contribute to powerful learning environments in primary education? *Computers & Education*, 44(3), 343–355.