

# Hand Gesture Based Interactive Puppetry System to Assist Storytelling for Children

Hui Liang<sup>1</sup>, Jian Chang<sup>2\*</sup>, Ismail K. Kazmi<sup>2</sup>, Jian J. Zhang<sup>2</sup>, Peifeng Jiao<sup>3</sup>

**Abstract** Digital techniques have been used to assist narrative and storytelling, especially in many pedagogical practices. With the rapid development of HCI techniques, saturated with digital media in their daily lives, young children, demands more interactive learning methods and meaningful immersive learning experiences. In this paper, we propose a novel hand gesture-based puppetry storytelling system which provides a more intuitive and natural human computer interaction method for young children to develop narrative ability in virtual story world. Depth motion sensing and hand gestures control technology is utilized in the implementation of user friendly interaction. Young players could intuitively use hand gestures to manipulate virtual puppet to perform story and interact with different items in virtual environment to assist narration. Based on the result of the evaluation, this novel digital storytelling system shows positive pedagogical functions on children's narrating ability as well as the competencies of cognitive and motor coordination. The usability of the system is preliminary examined in our test and the results which showed that young children can benefit from playing with Puppet Narrator.

**Keywords** Serious game · Digital storytelling · Virtual puppet · Hand gesture recognizing

## 1 Introduction

In recognition for their considerable positive effects on pedagogy, educational games or serious games for training purposes have become immensely popular [1-2]. As a modern form of traditional storytelling, digital storytelling systems emerged over the last few years and have demonstrated powerful pedagogical functions, which enable children to express themselves and cooperate with others during narrative performance. Storytelling is essentially one of the original forms of teaching [3], which can be used as a method to teach ethics, values, and cultural norms and differences.

Digital storytelling following the same well-known strategies similar to classical storytelling can help children to acquire several technological skills and work in groups

and strengthen the bonds between each other. As another social benefit, digital storytelling can also help disabled children or students with learning difficulties to remove the barriers of communication with adults and peers and overcome the inability to focus on their feelings or thoughts by providing them with opportunities to play active roles [4]. At present, the major pedagogical benefit gained with digital storytelling is the ability to narrate [5-6].

However, storytelling is not only about narrative. In its basic form, storytelling is usually combined with gestures and expressions. Oral narrative can also be combined with other body movements, e.g. dancing to enhance the storytelling through remembrance and dramatic enactment of stories [7]. From this point of view, storytelling will not only benefit a child's understanding of narrative structures but also fertilize other abilities, such as cognitive competence and physical coordination during performance with the aid of different media.

In this paper, we design and develop a novel digital puppetry storytelling system 'Puppet Narrator' for young children utilizing depth motion sensing technology, which supports hand and finger motions as input but requiring no hand contact or touching. Considering a puppet's operation complexity for young children, we use hand tracking and gesture recognition technologies to simplify operations and provide intuitive interface, in which children can use hand gestures to manipulate virtual puppet to perform story. Contrary to the research [8], in addition to narrative fertilization, we also devote our proposed system to increase children's abilities on the aspects of cognitive development, and motor coordination ability during their storytelling performance with the help of depth motion sensing device. The system usability is preliminarily examined, and the results show that young players can benefit from the novel narrating.

In summary, this work has two main contributions:

- Introduce a novel narrative assistance with gesture control and computer animation by combining motion sensing technology to manipulate a 3D puppet;
- Implement a prototype of the digital storytelling system to help young children to develop their related skills.

The remainder of this paper is organized as follows. Section 2 makes a brief introduction on related research works. Section 3 presents our system design. Section 4 describes system implementation including the architecture, input data processing, motion control, and output. Experiment result and evaluation is discussed in Section 5 and Section 6 concludes the paper.

---

\* Corresponding author: [jchang@bournemouth.ac.uk](mailto:jchang@bournemouth.ac.uk)

1 Communication University of China, Beijing, China

2 National Centre for Computer Animation, Bournemouth University, Poole, UK

3 Basic Medical School Southern Medical University Guangzhou, China

## 2 Related Works

Over the past decade, there has been a considerable growth of advanced virtual reality technologies which provides the possibilities that allowing students to experience virtual learning environment in highly interactive and natural ways. Virtual reality is also considered as one of the most powerful tools for supporting learning process. VR interactive technologies are widely used in digital storytelling systems, such as the tangible interfaces, haptic feedback and wireless handheld orientation sensors etc.

Storytelling Alice [9] introduces computer programming to learners using 3D animated stories. Toontastic [10] can be considered as a collaborative and constructive digital animation creator that is designed to help children capture and share their stories with other children around the world. Wayang Authoring system [11] is a web-based visual story authoring media for children, which enables children to use virtual puppets to create stylised digital stories. Inspired by traditional Chinese shadow puppetry, ShadowStory [12] is designed for children to use a Tablet to create digital shadow puppets and perform story cooperatively on a projection screen. Handheld wireless sensors are used to control the movements of shadow puppets. In Mousawi's [13] research, an iPad was used as a storytelling tool to help teachers and parents evaluate and improve the communication skills of Arabic children. Bonsignore's [14] work supports collaborative mobile storytelling for young children of 8-11 years old, which focuses on mobile reading and authoring to encompass both local and remote mobile authorship practices. In the educational program "From the Ancient to the Modern Tablets", target users can help a digital agent to time-travel back in a 3D Immersive eLearning Environments [15]. Rubart [16] proposed a multi-touch tablet system supporting face-to-face collaborative storytelling which could provide a natural interaction experience following the metaphor of a virtual meeting desk.

Comparing to the traditional oral storytelling, virtual reality technologies present digital content in more compelling and engaging formats, which provide users more interactive and immersive experience to assist narration. Interaction methods primarily used in previous research are limited to the traditional Human Computer Interface (HCI) technologies, such as keyboard and mouse (e.g., [9], [11]), handheld remote controller (e.g., [12]) and touch screen including mobile phone and tablets (e.g., [10-16]). However, at present the young children, also known as the "new media generation" or Digital Natives [17], were born in a richer media environment and they start to interact with new technologies from an early age. Saturated with digital media in their daily lives, they require more interactive learning environments, multimodal feedback, and meaningful learning experiences, which bring new challenges to the current digital storytelling. The new generation of the digital storytelling system is expected to offer a novel and immersive way to captivate learners' interests in a new horizon 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. The development of novel HCI

methods (e.g. depth motion sensing technology) provide us with new possibility for deriving educational benefits from storytelling by creating new ways of enabling interaction through hand gestures or other modalities. Not only oral narration but children's other capacities are expected to be improved during their storytelling performance, such as cognitive competence and physical coordination. Such improvement can be significant especially when digital storytelling system can provide players with more immersive interactions, such as using hand gestures to manipulate avatars' movements in virtual environments.

## 3 General Design

In this section we discuss the system design which involves the novel manifestation pattern as well as the pedagogical considerations we mentioned previously.

### 3.1 Target

Recently, there has been substantial amount of research undertaken on digital storytelling mainly investigating narrative abilities training, such as Toontastic, Kodu, Storytelling Alice, Wayang Authoring, et al [8]. However, as we mentioned before, storytelling is not just narrating. Beside oral narration, the development of other abilities is also vitally important for young children, e.g. space and object cognitive abilities. If powerful user friendly interaction methods, which are more intuitive and natural, are provided by digital storytelling systems, then the additional abilities (cognition and motor coordination) will also be considered.

There is a plethora of research in the field of psychology which is used for development of important skills in children [18-21]. Research suggests that spatial-temporal reasoning and spatial visualization ability is an important indicator of achievement in science, technology, engineering and mathematics [22] and a pre-school child's visual spatial attention ability predicts his future reading skills [23]. Researchers have also postulated a set of so-called "core domains" in cognitive development and suggested that children have innate sensitivity to specific kinds of patterns of information. Those commonly speculated core skills of cognition include: number [24], space [25], visual perception [26], essentialism [27], and language acquisition [28]. As an important aspect of children's psychosocial development, the significance of motor coordination competence has also been recognized in pedagogy a long time ago. Children with poor motor coordination have been found to underachieve educationally and to experience difficulties with peer relationships [29].

Puppet Narrator is developed to target children between 5 and 8 years old, which covers the age group of Key Stage1 (5~7 years old). In UK, the national curriculum is organised into blocks of years called "Key Stages". And also in Arora's study, initial prototype of a narrative learning device "FunPi" has been tested with four children of age group of 6-8 years [30]. Our aim is to endow digital storytelling with a novel interaction method, which is more flexible and immersive. At the same time, our system is

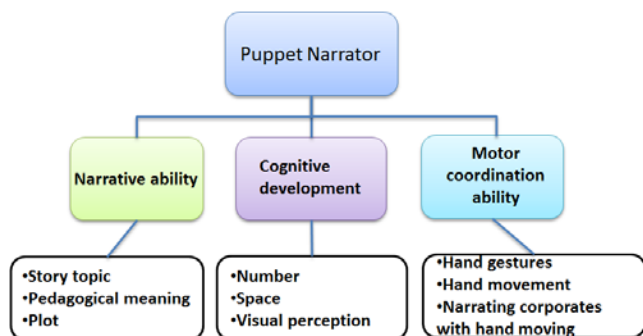
highly educational not only in terms of narrative competence but also cognitive ability and motor coordination. In our system, besides narrative ability training, we also pay attention to the development of the core skills of numerical cognition, spatial awareness and visual perception with the assistance of motion sensing technology in virtual environment. To enhance the children's motor coordination competence, by using depth motion sensor to track and recognize players' hand movement and gestures, our system enables children to use their hand motions to manipulate virtual puppet for interacting during narrating.

Our preliminary conception is as follows: following the provided story plot, children will finish the whole story narration, and at the same time, children can simply use hand motion to control the movement of virtual puppet and interact with playthings in virtual scenario to assist narrating. Through this procedure, their narrative ability will be nourished. Using hand gestures for controlling the avatar, their motor coordination ability will be trained. In interaction with virtual items having different properties and roles in the story, their space and object recognition capability will also be developed.

Based on the above considerations, there are several aspects we should consider in our system design, as summarized below:

- 1) For the purpose of narrative ability training:
  - a) The story topic should be carefully chosen to be familiar to the children and so it can be easily repeated and narrated.
  - b) The story should have pedagogical meaning.
  - c) The story plot should be provided to the children to follow before the story commences.
- 2) For the purpose of cognitive core skills' development:
  - a) Numerical cognitive: players should finish a certain number of goals, denoted by a score within the game.
  - b) Spatial cognitive: players should have a clear recognition of their hand's location in real world as well as the location of avatar and other playable objects in virtual environment. They will develop the ability to map the position of their hands from the real world to the position of the avatar in the virtual world.
  - c) Visual perception: players should have the ability to distinguish different playable objects having different functions and uses.
- 3) For the purpose of motor coordination ability training:
  - a) Players are required to perform hand gestures correctly and in sequence.
  - b) Players are required to move their hands steadily and accurately.

The structure of our training aims is illustrated in Fig.1.



**Fig. 1** Training aims to cover three fundamental abilities: narrative ability, cognitive skills and motor coordination ability, each of which is realised through different related training activities in the digital storytelling practices.

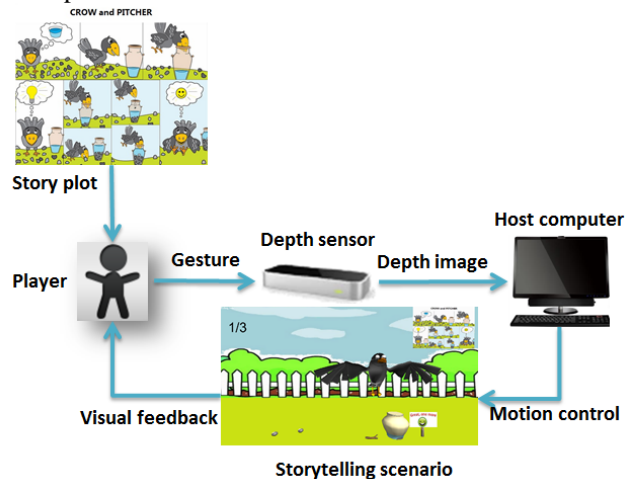
### 3.2 Story Topic

“The Crow and the Pitcher” is one most famous of Aesop's Fables. A thirsty crow found a pitcher with some water at the bottom out of the reach of its beak. The crow picks up pebble stones and drops them into the pitcher to raise the water level until it can drink the water. The fable is made by ancient Greek poet Bianor [31] and then collected by Avianus. The fable emphasizes the virtue of thoughtfulness over brute strength and the value of the crow's persistence. Considering its popularity among young children as well as its positive pedagogical meaning, we choose “The Crow and the Pitcher” as the story topic.

During narration within our virtual environment young children use their hands and a set of hand gestures to manipulate the puppet crow to pick up pebbles and drop them into the pitcher. The crow's actions, such as flying, grasping, and drinking will be presented through pre-recorded animations, controlled by the hand gestures.

### 3.3 Pipeline

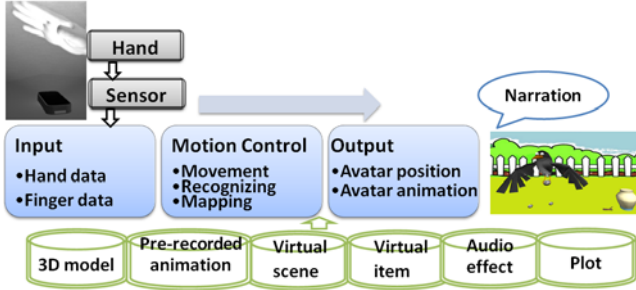
The component level interaction within the system is shown in Fig.2. Firstly, a story plot is provided as storytelling hints to young players. Secondly, players use hand motion to manipulate the avatar through depth motion sensor device, which can automatically track hand motion and recognize hand gestures. Depth image data from the motion sensor is obtained and interpreted into motion control commands by the host computer. Finally, as visual feedback, the 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. Under this novel manifestation pattern, not only players' oral narrative ability but also their cognitive and motor coordination competence is expected to be developed.



**Fig. 2** Component diagram of the system interactive pipeline

## 4 Implementation

Puppet Narrator is mainly composed of three parts: Input, Motion Control, and Output, as illustrated in Figure 3. The input part processes the sensor data captured from motion sensor device through HCI and passes it to the next part. Motion Control interprets the data subsequently and determines avatar's location and posture. The output module updates avatar performance in virtual environment as the feedback.



**Fig. 3** The implementation of the system architecture is mainly composed of three parts: Input, Motion Control, and Output.

We utilized a Leap Motion controller [32] in our system as the HCI sensor device to track hand gestures, which can provide a high fidelity finger tracking through an infrared depth sensor. We utilized the Leap Motion SDK provided by the Leap Motion Co. as the API to access the motion data of hands and fingers from the device. All the 3D models and animations were created in Maya 2014. We integrated and developed the entire system in Unity3D Pro V4.2.

### 4.1 Digital Puppet Crow Design

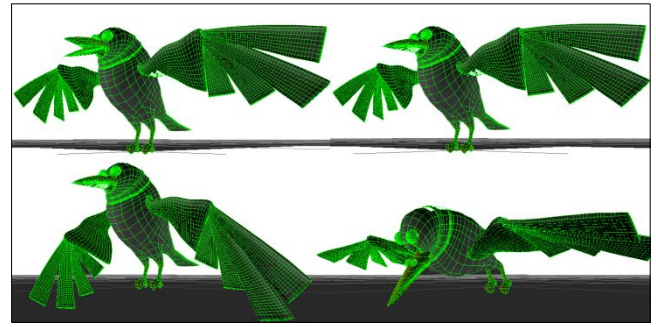
To make our system more appealing for young children, considering puppetry's positive benefits in education, we use a digital puppet as an avatar to assist children's storytelling through animation technology.

1) *Puppet geometry construction*, which describes puppet shape, the rigging system, and its shading materials. The shape of the digital crow is modelled in Maya 2014 simulating the picture of the crow in the plot. Rigging the puppet defines its behaviour with bones connecting joints.

2) *User interaction*, which defines the way players interact with the puppet, describing the performer expressions, and interaction interfaces.

3) *Animation*. The crow puppet uses pre-recorded animations to produce actions, which are triggered by the players' hand gestures.

Fig. 4 shows the sketched crow, the digital 3D crow model and animation screenshots.

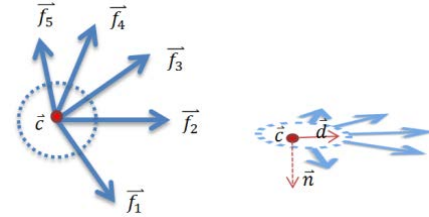


**Fig. 4** The sketch (top-left), the 3D model (top-right) and screenshots of pre-recorded animations (bottom two rows)

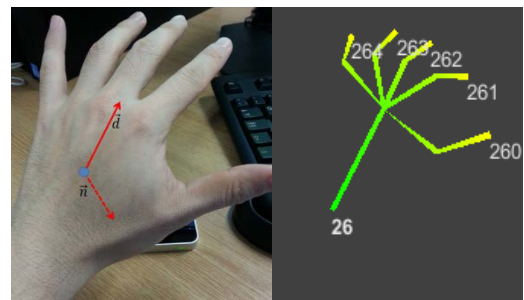
### 4.2 Hand Model

Hand model is shown in the right of Fig. 5 (a), where  $\vec{f}_i$  ( $i \in [1,5]$ ) presents the position of the finger tips ( $i$  is the number of recognized fingers), and  $\vec{c}$  presents the centre of the palm. The plane of the hand is formed with the normal vector  $\vec{n}$ , and the directional vector  $\vec{d}$ . Vector  $\vec{n}$  and  $\vec{d}$  presents the normal and the directional vector of the hand plane. The player's hand and its virtual skeleton mapped are shown in Fig. 5 (b).

For ease of controlling by young children, we define four intuitive motion controls: right, left, move downward and upward, which are mapped to different hand gestures, as illustrated in the upper four rows in Table 1.



(a)



(b)

**Fig. 5** (a) Hand skeleton model; (b) Player's hand and its virtual representation (the numerical values represent the ID of the recognisable fingers)

### 4.3 Input Data Processing

The sensor data provided by Leap Motion controller contains a diversity of information about hands and pointables (such as fingers or finger-like tools defined by Leap Motion Co. [31]), in the virtual scene, which is updated by frame and can be represented as follows:

$$SensorData = \langle FR, H, P, T \rangle \quad (1)$$

where FR is frame rate; H represents the set of hands detected and P represents the set of pointables; while T represents timestamp. Hand data H mainly contains the hand identifier, direction and different value about the palm position and status:

$$H = \langle id, dir, palmInf \rangle \quad (2)$$

And the pointables data P includes the id, direction and position information relative to the hand:

$$P = \langle id, dir, handid, positionInf \rangle \quad (3)$$

#### 4.4 Motion Control

Recent research on neuroscience found that in human brain development there is a strong connection between perception, imagination and movement. According to previous studies [33-35], through the feedback of their own movement, visual observation or the feedback of avatar's (3D virtual character's) motion [36], humans can recognize and coordinate their movements better.

The motion control module is the core of the system, which has two main functions: movement control and recognizing. Movement control function is responsible for mapping the play's relative hand position to the movement of the puppet crow. Recognizing function is in charge of identifying the pattern of hand movement and gestures implicating player's intention. Once a recognizable gesture is detected, recognizing function will trigger a pre-recorded animation or event (e.g. gripping a pebble or flapping wings).

##### 1) Movement Control Mechanism

Since in a storytelling system children mainly focus on narrating, a complex puppet manipulation as if we are stringing a puppet in a real show will be distracting or even hamper narrative [37]. If young children pay much attention on puppet manipulation, they might forget the story line. From this consideration, we design a user-friendly puppet prototype used as a storytelling avatar with a simpler interaction manner and easier motion control mechanism.

For young children to control the movement of the puppet/avatar, the most direct way is converting the translation of their hands position in Leap Motion coordinate system to the position of puppet in virtual environment. Considering the different scale between these two workspaces, a proper transformation matrix should be involved as a scaling into the coordinate translation. Once players' hands is not recognized by the Leap Motion device, puppet will keep its position in the last frame and then resume the movement immediately when hands can be detected again in the following frames. Within each frame, the position of the puppet is decided by the values of two coordinate vectors: the puppet's former coordinate in previous frame and the hand's relative translation generated by player's hand movement in current frame. And the scaling factor considered at the meanwhile, the

puppet crow's coordinate in virtual workspace is computed as follows:

$$P_{crow} = M \cdot S \cdot R_{hand} + P'_{crow} \quad (4)$$

where  $P_{crow}$  is the puppet crow's position in current frame,  $R_{hand}$  is player's relative hand movement in real world in current frame comparing with the previous frame, which can be obtained in formula (2), M is the transformation matrix between the player's workspace and the puppet crow's coordinate system, S is the scaling matrix, and  $P'_{crow}$  presents puppet crow's position in the previous frame.

The algorithm of the movement control mechanism is illustrated below, which takes Leap Motion sensor data as the input and calculates the difference of the data information between current frame and previous frame in order to generate puppet new position in current frame. If motion data is not accessible from Leap Motion controller in current frame (that maybe caused by an unrecognized object/hand gesture or moving out of the detecting range), puppet will stay where it was in the previous frame, unmoved (Line 1 to Line 3). Else, update puppet position according to formula (4) (Line 5 to Line 9).

---

##### Algorithm: Movement control

---

**Input:** *sensorDataCurrentFrame, sensorDataPreviousFrame, puppetPositionPreviousFrame*

**Output:** *puppetPositionCurrentFrame*

**Begin:**

1. **If** *sensorDataCurrentFrame* **is null** **Then**
2. *puppetPositionCurrentFrame* = *puppetPositionPreviousFrame*;
3. **Return** *puppetPositionCurrentFrame*;
- 4.
5. *relativeHandMovement* = *getDiffer(sensorDataCurrentFrame, sensorDataPreviousFrame)*;
6. *transMatrix* = *getTransMatrix(playerWorkspace, puppetCoordinateSystem)*;
7. *scalingMatrix* = *getScalingMatrix()*;
8. *puppetPositionCpreviousFrame* = (*relativeHandMovement, transMatrix, scalingMatrix, puppetPositionPreviousFrame*);
9. **Return** *puppetPositionCurrentFrame*;

**End**

---

##### 2) Types of Hand Gestures

In virtual interactive environment, using hand gestures as an input to control avatar's performance is more natural and intuitive than other HCI methods, such as keyboard or touch screen input. Hand gesture interaction is clearly visible to others and constitutes an expressive action in itself.

When designing the system, the most important consideration is choosing a most natural and intuitive gestural interaction manner to play with the avatar. Considering their simplicity and demonstrated effectiveness, we utilize detectable pointing gestures [23] into HCI.






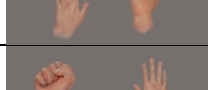

In our development, we have designed different types of hand gestures targeted to different levels of interactive applications.

##### Type-I

Type-I, basic single hand gesture set, is used to control avatar's movement and trigger avatar's simple animation actions with a single hand. This set of hand gestures is

mainly designed for young children by providing a natural and intuitive way to interact with the playthings at a basic interaction level. Table 1 shows the mapping between basic single hand gesture set performed by young children and the puppetry crow's action in virtual environment, which actually includes two different kinds of gestures: navigation gestures (item 1-5) and action gestures (item 6-7).





**Table 1** Type-I, basic single hand gesture set




Hand gesture	Movement	Target Action
	Move right	Fly to the right
	Move left	Fly to the left
	Move down	Fly down
	Move up	Fly up
	Stretch	Hover
	Stretch to grip	Grasp pebble/ stick
	Grip to stretch	Drop pebbles/ stick

**Type-II**

Type-II, advanced single hand gesture set, is used to generate more complex avatar's actions on the base of the Type-I. Such type of gestures can be used for producing more complex interaction. This set of hand gestures is the complement and improvement of the basic single hand gesture set. We defined complex hand gesture interactions to provide more functionalities as shown in Table 2. The targeting users are primarily juniors instead of 5-8 years old young children, who have a better motor control and cognition ability. An intuitive example is to use a single finger tracing a circle in space to manipulate the crow to turn around. Table 2 also shows the mapping between advanced single hand gesture set performed by players and the avatar's action in virtual environment.

**Table 2** Type-II, advanced single hand gesture set

Hand gesture	Movement	Target Action
	Circle	Turn around
	Swipe	Swing wings/head/tail
	Key Taps	Nod head /put down
	Screen Taps	Pick up

	rapid multiple finger taps	Step
	Thumb up	Begin
	Okay sign	End

**Type-III**









Furtherly, both the player's left and right hands are involved in our Type-III hand gesture interaction for more experienced users, called two-handed gesture set. In Type-III, two subsets of hand gestures (Hand-A: navigation gestures and Hand-B action gestures) are designed separately and each set is attached with one hand: navigation gestures are assigned to the left hand for steering tasks and action gestures are allotted to the right hand and trigger avatar's action performance. As illustrated in Fig.6, the Leap Motion controller is placed in front of the player and his hands are held out straight on each side of the sensor device.



**Fig. 6** Two hands' positions in relation to Leap Motion device when using Type-III hand gesture set

Type-III: Hand-A, left hand navigation gestures allow players to continuously control avatar's movement in virtual scenes when moving left hand into different directions or inclining hand upward/downward, as illustrated in Table 3.

**Table 3** Type-III: Hand-A, Navigation gesture set

Hand gesture	Movement	Target Action
	Move forward	Forward
	Stop	Stay/hover
	Swipe right	Turn right
	Swipe left	Turn left
	Move down	Fly down
	Move up	Fly up
	Incline downward	Accelerate
	Incline upward	Decelerate

Type-III: Hand-B, right hand action gestures represent what kinds of action the player wants the avatar to play. Different gestures of the right will trigger pre-recorded animation clips to perform different actions. Action gestures include the action gestures defined in the Type-I basic single hand gesture set (item 6-7) and Type-II advanced single hand gesture set.

Generally speaking everyone has different handedness, for example, some people are right-handed and others are left-handed, or someone has his/her own understanding of the meaning of hand gestures, e.g. using fist to finish the game instead of the using Okay sign. The type and appearance of hand gestures may vary a lot. Of course different kinds of hand gesture sets could be defined for more personal and more complex interaction with the deep motion sensor device. No matter how powerful the interaction is, there is one principle we should follow: intuitive and natural input is the most desired feature.

Leap Motion controller could recognize and track hands, fingers with high precision. Furtherly, the movement patterns of each fingers could be observed individually and certain kinds of movement patterns could be recognized by Leap Motion as gestures which indicated the user's intent or command. Hand gestures are represented by the "Gesture" class and its subclasses, such as CircleGesture, KeyTapGesture, ScreenTapGesture, and SwipeGesture [31]. For example, moving a hand from side to side indicates a swipe gesture. And furtherly there are also some third-party software devoting to extend the gesture recognition functionality of the Leap Motion controller, which enables users to assign actions to different gestures according to their favour [38]. By using Leap Motion API and its third-party software, we could track and recognize the different types of hand gestures as mentioned before. To reduce the degree of difficulty of young children's operation, simple gestures, such as grip and stretch, are recognizable in our system, as shown in the two bottom rows of Table 1.

#### 4.5 Output

The output module updates puppet crow's position by calculation of the motion control module. It also plays pre-recorded animations which have been linked with the recognisable gestures in motion control. Once a recognisable gesture is detected, pre-recorded animation of the puppet crow linked with this gesture will be triggered. The puppet crow can then act as a real puppet and perform a pre-set action responding to the player's hand gestures in the virtual environment. A scenario of a player controlling the avatar by finger gestures is shown in Fig.9. This module provides corresponding feedbacks to players for adjustment of hand movements, which is vital for our cognitive development purposes as well as motor coordination ability training.

## 5 Results & Discussion

### 5.1 Prototype

The plot of "The Crow and the Pitcher" is presented first as the hint of the story, as illustrated in Fig.7. A scenario of Puppet Narrator during playing is shown in Fig. 8.



Fig. 7 Draft of the story plot is provided as the hint before narrating starts.

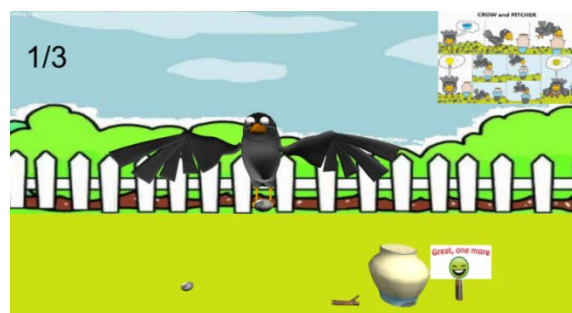







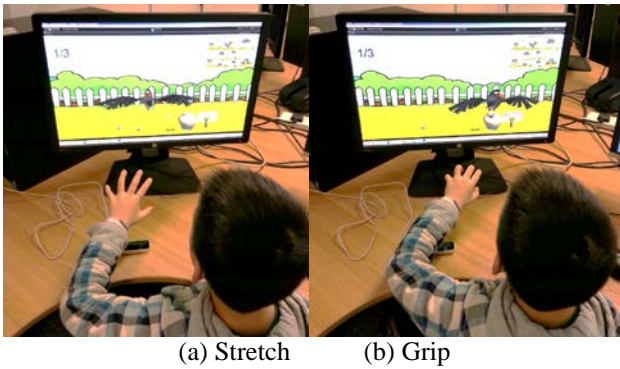
Fig. 8 Screenshot of a scenario of Puppet Narrator during playing

In the scenario, there are five kinds of virtual items, each of which has different function, as shown in Table 4.

Table 4 Virtual items used in the system

Hand gesture	Movement	Target Action
	Pebble	Fill pitcher for water rising
	Stick	Confuse player which needs to be differentiated from pebbles
	Pitcher	Collect pebbles dropped by crow and feed crow water after three successful drops
	Signboard	Provide information to player
	Counter	Calculate how many pebbles are in dropped in pitcher

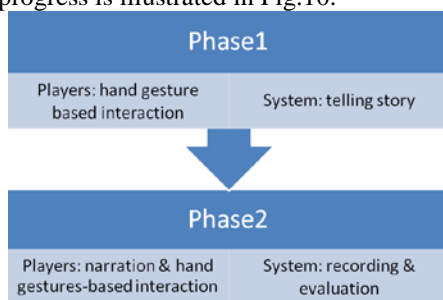
In Puppet Narrator, young children can use their fingers to control the animation of the puppet. In Figure 9(a), we can see that the player stretches his hand and move it to control the movement of the puppet crow by mapping the palm position to the crow's position. In Figure 9(b), we can see that the crow has grasped a pebble successfully and is preparing to drop it into the pitcher.



**Fig. 9** Basic gesture control, an example of using combination of Type-I gestures to steer and manipulate the puppet.

## 5.2 Progressive Storytelling Creation

It is a dilemma that on one hand using hand gestures to manipulate an avatar to assist storytelling can greatly increase young children's enthusiasm for narrating, but on the other hand it also demands higher motor and speech coordination. This becomes a certain challenge to young players from our observation, especially in the initial practice phases. To reduce adaptive difficulty, we have designed a kind of progressive storytelling training model, which includes two practice phases. During the first phase, players only need to focus on the hand gesture interaction and do not need to worry about the story plot narrating. Story plot is narrated automatically by the system following the players' operation, which provides young children the opportunity to master how to use hand gestures to operate the avatars in the virtual scene through the motion sensor device. Once the young players get used to the hand gesture-based interaction, self-narration is added in as a demand for them to fertilise narrative ability as well as motor coordination training during the second practice phase. The concept model of the progressive storytelling training progress is illustrated in Fig.10.



**Fig. 10** Concept model of the progressive storytelling training progress

## 5.3 Supervised Narrative

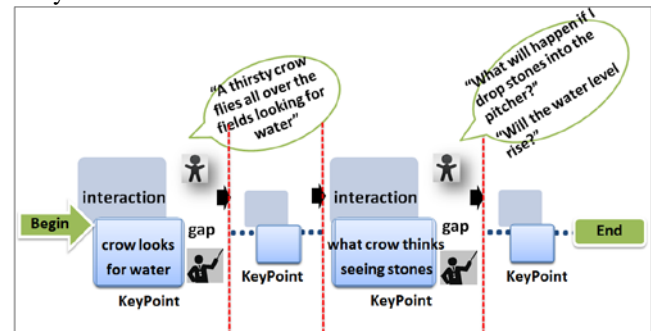
During the training process, necessary supervision and guide of an adult tutor is provided to assist young children's narration. An important aspect is the plot control which is crucial for further development of interactive storytelling. To provide young players the opportunities to narrate when interacting with the virtual puppet, the tutor

will interrupt and ask the player to stop once a certain task or movement is accomplished to allow time for narration.

In our interactive storytelling system, it is not compulsory for young children to follow the direct causality pattern strictly. The entire narration can be considered as a graphic structure which consists of multiple key plot points. The connections between these key points push the plot forward. The key points of the story "The Crow and the Pitcher" are listed below:

- Describe a thirsty crow searching for water.
- Show how the crow tries to reach the water in the pitcher but fails.
- Describe what the crow thinks when seeing the stones.
- Describe how the crow distinguishes between the different uses of stones and sticks.
- Help the crow approach and pick up the stones.
- Help the crow fly to the pitcher and drop a stone into the pitcher to raise the water level.
- Present how happy the crow is when it can drink the water.











The concept model of the supervised storytelling procedure is illustrated in Fig.11. Key plot points accompanied by players' action and interaction compose the main storyline. At the key points, a tutor would gently interrupt the young players' hand gesture interaction with the avatar and remind them to stop at the "gap" to take their time to narrate. Referencing the draft of the plot, the players can narrate from their own understanding of the story.



**Fig. 11** Concept model of the storytelling procedure that supervised by a tutor

The screenshots of the story narrated by a 7 year-old boy is recorded in Fig.12.



Plot	Virtual scene	Narrative record
		A crow is flying around on a hot summer day looking for water...
		He comes across a pitcher. Oops! He can't reach the water.
		He has an idea: collect pebbles and drop in the vase
		The crow drops pebbles in the pitcher. Water level rises.
		He takes lots pebbles and drops them into the Pitcher... He can drink water now!!

**Fig. 12** Example of key points performed by a 7 year-old child

## 5.4 Evaluation

### (1) Participants

A pilot study participated by four 5-8 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.

One adult volunteer (a post graduate student) took part in the experiment as the observer to assess the young players' performance. This observer has research background of HCI and is experienced in VR development. He only provided observations instead of any professional opinion or analysis.

### (2) Method

All participants were trained to familiarise themselves with the gesture controller after detailed explanation of the experiment.

In the experiment, each child has carried out five trials of the "The Crow and the Pitcher" story. Following the provided key story points, all children were engaged with the linguistic game with great interests and built their own stories by interacting with the digital puppets.

During playing, some minor frustrations (or difficulties) were observed, such as picking up the stick by mistake, failing to move the avatar toward the pitcher, 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.

An overview observation is that:

- All of the children could finish the game with few faults after a couple rounds of repeating.
- All of the children could narrate the plot with more complex words of their own language and understanding in the last three rounds.

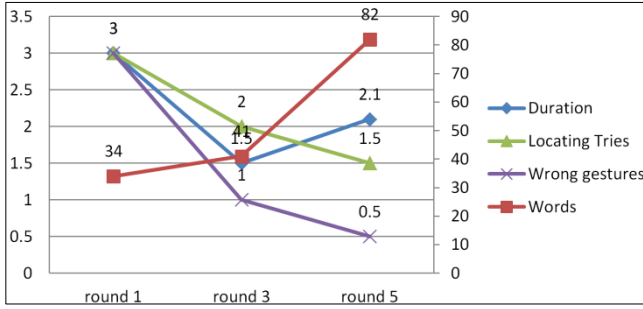
For pedagogical evaluation, two different types of evaluations are conducted: metric-based objective evaluation and observer's subjective evaluation.

Fundamental research has been carried out to evaluate the story narration quality of the elementary school student from long time ago [39-41]. Various variables were examined including story length and syntactic complexity [42-44], the use of specific story grammar components, and the presence of episodes etc. [45], among which the metrics of story length and syntactic complexity are commonly examined in the practice of school-age language development studies [42-44]. In this paper, the variables of story length and syntactic complexity were studied for narrative ability evaluation by using the metric of *Narrative Complexity (M2)* to calculate the number of words produced by the children in objective evaluation phase and the criterion of *Vivid Narration* in the subjective evaluation. The evaluation of other narrative abilities including Story Grammar Components, Episodes and Story Comprehension etc. is relatively complex, which needs the analysis and agreement among independent judges, and children's answers to several questions [45]. Since the main concern in this paper is to provide a novel interactive method to assist narration rather than a case study on the pedagogical evaluation using existing educational tools/systems, only the most commonly used variables were considered.

Event Related Potentials (ERP) is currently used in the study of the cognitive process and motor control in VR. It is the measured brain response to the sensory, cognitive or motor event [46]. For example, ERP method is employed to study cognitive process in a virtual traffic environment through the presentation of traffic signs with different background colours [47]. In the virtual cognitive training [48], subjects' motion and cognitive process was studied by asking participates to think about and furtherly to grasp the virtual objects (i.e., glasses, cup, scissors, mouse, pen, fork) with gesture controller. Based on these previous Event-Related Potential (ERP) research, three events were investigated in our experiment in accordance to our system features and target users: locating hand to grasp pebble, picking up stick by mistake, and performing wrong hand gestures. They were depicted with three metrics respectively: *Number of tries to locate pebble (M3)*, *Number of tries to pick up stick (M4)* and *Wrong hand gestures (M5)*. For more details of Event-Related Potential (ERP) method, please refer to previous research [46-48]. Considering the implementation difficulty in objective evaluating, two subjective criteria for assessing the abilities of the smoothness of hand movement and accuracy of hand/avatar location were evaluated by an observer's subjective judgement.

### (3) Objective Evaluation

The statistics of objective metrics in each round of trial is recorded. The trend plot of average values of metrics is presented in Figure 13, which showed improvement and the details are shown in Table 5.



**Fig. 13** Trend plot of the average values of metrics observed

**Table 5** Experiment result

Metrics	Round 1	Round 3	Round 5
Duration of round ( <i>M1</i> )	3 minutes	1.5 minutes	2.1 minutes
Narrative complexity ( <i>M2</i> )	32 words	41 words	82 words
Number of tries to locate pebble ( <i>M3</i> )	3 tries	2 tries	1.5 tries
Number of tries to pick up stick ( <i>M4</i> )	1 try	0 tries	0 tries
Wrong hand gestures ( <i>M5</i> )	3	1	0.5

Both Fig.13 and Table 5 shows the quality of improvement of young participants' performance over several rounds of trials. Ability to successfully locate the pebble (*M3*) is improved by practicing picking up or dropping pebbles which can be indicated by the decline of locating tries (shown by green line in Figure 11). The average number of tries of picking up the stick by mistake (*M4*) drops from 1 to 0, which means that children can distinguish the stick from pebbles after 2 rounds of training. Wrong hand gestures (*M5*) in the later stages of the experiment is improved over the earlier attempts.

We noticed from the experiment that young children were able to tell the story confidently and use more words and longer sentences in the last two rounds of trials, with some encouragement, which is reflected by metric *M2* (red line in Figure 11). One interesting thing we found is change of duration of narrating the story (*M1*). Initially, participants used 3 minutes on average to complete the whole narration and then they could finish it within 1.5 minutes, which benefits from getting more familiar with the system. Finally, the duration of narration tends to last a little longer. The reason is that once young children get used to the control mechanism and the interaction interface, they tend to pay more attention on the story line and the improvisation, which is vital for storytelling and more suitable for pedagogical purpose as a learning tool for children.

#### (4) Subjective Evaluation

Three subjective criteria are included in the subjective evaluation of the young players' performance: smoothness of hand movement, accuracy of hand/avatar location and vivid narration.

- Smoothness is characteristic of player's coordinated hand gesture movement.

- Location accuracy determines the accuracy of the player's hand that manipulates the avatar to a certain location in the virtual scene.
- Vivid narration requires young children to use rich expressions to create a clear picture and use vocal tones to suit the story. It enables the audience to be fully engaged and emotionally involved in the story.

The evaluator rated the players on each subjective criterion using a scale that ranges from "poor" to "excellent", as opposed to objective evaluations that more often have a numerical score attached to the criteria. The evaluator's records showed the improvement of players' performance during the subjective evaluation experience.

Firstly, young players' hand movements became smoother with practices. The decrease of the jerky hand movement suggested the improvement of the motion control and hand-eye coordination.

Secondly, the ability to locate hand position accurately is enhanced. From the tutor's observation, the times of relocating attempts decreased in the last trails. Children tended to control avatar's movement more easily by gradually adapting the mapping between their hand positions and the gesture controller's coordinate system after several rounds of practice.

Finally, and most importantly, the training led to the ability to present a vivid narration. According to the records, the children tended to use more descriptive words for narration across the five rounds. One typical example is the words used by a 7 year-old boy at the start of the story: in the first trial, the boy used the sentence "A crow is looking for some water." In the third trial, the sentence was rephrased to "A thirsty crow is flying around looking for water." After adding more decorative words, the sentence finally transformed to "A thirsty crow is flying around on a hot summer day looking for water." It is obvious that the narrative ability was enhanced during storytelling. With the improvement of the criteria of the smoothness of hand movement and location accuracy after several rounds of trials, the children were able to pay more attention on the narration and rhetoric, which is vital for storytelling and even more suitable for pedagogical purposes as a learning tool for children.

#### (5) Discussion

During the test, the aspects of children's cognitive competence and motor coordination ability were improved.

Numerical cognition: all the three pebbles need to be dropped into the pitcher before the crow can drink water, which requires a basic numerical cognition to finish the game.

Spatial cognition: pebbles could only be picked up within predefined area and dropped into the pitcher, which means the spatial cognition ability is required in locating the pebble.

Visual perception: there are several kinds of virtual items in the scenario and each of them has different property and usage, which means the players need to distinguish them from others and make correct choices, for example, only pebbles can make water to mount up and the stick will not work.

Motor coordination: Only predefined hand gestures can be recognized by our system, which means players need to perform hand gestures correctly. For controlling the puppet

crow, players need to use hand motions as the input to manipulate the crow in the virtual environment and adjust hand gestures according to the visual feedback received from the crow's responding movement. This requires players to move their hands smoothly and steadily. During the performance, the players' narration often incorporates with hands movement and changing gestures, which also forms motor coordination.

In our experiment, basic single hand gesture set (Type-I) is used as the HCI instead of more complex hand gestures although we have designed two other different types of hand gestures set to provide different levels of interaction. Due to the young participants' limited motor control ability and the cognition capability, we have difficulty to train the young players to use the complex hand gestures. Simple hand gesture is more adaptable for young children by providing a natural and intuitive way for interaction at a basic level. It will be possible to introduce complex gesture set for older age group if a new experiment will be designed or in other pedagogical practices.

One limitation of the user interface comes from the hardware. The Leap Motion device can only detect hands, fingers and tools within the tracking area. Once player's hand moves out of range, the system can easily lose track and the puppet will keep still and stay where it is until the hand moves back within the detective area and is recognised again by the controller. At the same time of providing a better immersive experience to the players, the passive effects, e.g., the cybersickness, should also be taken into account in system design. To alleviate the passive effect, for example, the entire virtual game/story is designed to consist of several key plot points as we discussed in the system design. And at some points (e.g., the key plot points), players were reminded to stop at the "gap" and take their time to narrate. That means there should be not too much time restrictions, especially when the targeting users are young children.

There are limitations with our evaluation considering the limited data pool. According to our observation, there were obvious differences between a five-year-old and an eight-years-old in storytelling performances. And the language skills also varied between genders. These differences are expected to be well observed and discussed in our next research. We plan to exploit our system to conduct indicative experiments with more subjects to identify the pros and cons with quantitative data and analysis. Both independent variables and dependent variables will be measured or observed to provide a clearer understanding. Also, a control group will be involved in the future serving for comparison evaluations. The number of subjects is limited by the difficulty of working with young children, where close supervision and support of parents are necessary. Therefore, a control group is missing in the experiment which may possibly cause ambiguity.

## 6 Conclusion

We have presented a novel digital storytelling system assisted with virtual puppetry, 'Puppet Narrator', providing young children with natural interaction/control and immersive experience when narrating story. The system is

designed to support training of different cognitive skills and motor coordination through storytelling. It has been a novel attempt to include advanced motion sensing technology and computer animation as a medium for this development.

The usability of the system is preliminary examined in our test and the results which showed that young children can benefit from playing with Puppet Narrator from the analysis are promising. However, as only a limited number of subjects are tested, we will need to examine more cases and design a psychological experiment to affirm the conclusion. Further validation and analysis of the effectiveness of this approach is needed, but at the moment our observation and analysis can be supported with success of other parallel development in digital story telling [8].

The story telling in our test is a supervised learning process guided by an adult, which is important for the young ones to accomplish their narrative task and receive proper training. Without the presence of supervision, it is possible that the virtual puppet may distract the story telling that the child would focus on the playing and controlling of the puppetry without practicing their narration. It will require a rewarding strategy in the future development of such system to automatically encourage and reward the players when they accomplish the narrative task properly.

**Acknowledgments** We would like to thank Mr Mark Dodwell and Shujie Deng for proofreading the draft. The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° [623883] - "AniM". The authors also acknowledge partial support from the FP7/2007-2013 ("Dr. Inventor", FP7-ICT-2013.8.1 611383) and the National Natural Science Foundation of China (NSFC, 31200708).

## References

1. Backlund, Per, Maurice H.: Educational games-are they worth the effort? A literature survey of the effectiveness of serious games. In: Proceedings of the 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), pp. 1-8. IEEE Press (2013)
2. Greitzer, F.L., Kuchar, O.A., Huston, K.: Cognitive science implications for enhancing training effectiveness in a serious gaming context. *Journal on Educational Resources in Computing (JERIC)*, 7(3), p.2 (2007)
3. Pedersen, E. M.: Storytelling and the art of teaching. In *English Teaching Forum*, 33(1), pp. 2-5 (1995)
4. More, C.: Digital stories targeting social skills for children with disabilities multidimensional learning. *Intervention in School and Clinic*, 43(3), pp.168-177(2008)
5. Di Blas, N., Boretti, B.: Interactive storytelling in pre-school: a case-study. In: Proceedings of the 8th International Conference on Interaction Design and Children, pp. 44-51. ACM (2009)

6. Di Blas, N., Paolini, P., Sabiescu, A.G.: Collective digital storytelling at school: a whole-class interaction. *International Journal of Arts and Technology*, 5(2-4), pp.271–292 (2012)
7. Eder, D.: *Life Lessons Through Storytelling: Children's Exploration of Ethics*. Indiana University Press (2010)
8. Psomos, P., Kordaki, M.: Pedagogical analysis of educational digital storytelling environments of the last five years. *Procedia-Social and Behavioral Sciences*, 46, pp.1213–1218 (2012)
9. Kelleher, C.: *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, (2006).
10. RUSSELL, A.: 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, (2010).
11. Widjajanto, W.A., Lund, M., Schelhowe, H.: Wayang Authoring: a web-based authoring tool for visual storytelling for children. In: *6th International Conference on Advances in Mobile Computing and Multimedia*, pp. 464-467. ACM (2008)
12. Lu, F., Tian, F., Jiang, Y., Cao, X., Luo, W., Li, G., Zhang, X., Dai, G., Wang, H.: ShadowStory: creative and collaborative digital storytelling inspired by cultural heritage. In: *Proceedings of the the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1919-1928. ACM (2011)
13. Al-Mousawi, Z., Alsumait, A.: A digital storytelling tool for Arab children. In: *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services*, pp. 26–35. ACM (2012)
14. Bonsignore, E., Quinn, A.J., Druin, A., Bederson, B.B.: Sharing stories “in the wild”: A mobile storytelling case study using StoryKit. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(3), p.18 (2013)
15. Mystakidis, S., Lambropoulos, N., Fardoun, H.M., Alghazzawi, D.M.: Playful Blended Digital Storytelling in 3D Immersive eLearning Environments: a Cost Effective Early Literacy Motivation Method. In: *Proceedings of the 2014 Workshop on Interaction Design in Educational Environments*, p. 97. ACM (2014)
16. Rubart, J.: Face-to-Face Collaboration Points in Storytelling using Multitouch Tabletop Systems. In: *Proceedings of the 2015 Workshop on Narrative & Hypertext*, pp. 33–35. ACM (2015)
17. Thompson, P.: The digital natives as learners: Technology use patterns and approaches to learning. *Computers & Education*, 65, 12-33, (2013).
18. Greer, R.D.: *The Teacher as Strategic Scientist: A Solution to Our Educational Crisis?*. Behavior and Social Issues, 1(2), (1991)
19. Greer, R.D.: *Designing teaching strategies: An applied behavior analysis systems approach*. Academic Press (2002)
20. Skinner, B. F.: *The Technology of Teaching* (1968)
21. Vargas, E.A., Vargas, J.S.: Programmed instruction: What it is and how to do it. *Journal of Behavioral Education*, 1(2), pp.235-251 (1991)
22. Wai, J., Lubinski, D., Benbow, C.P.: Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), p.817 (2009)
23. Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., Facoetti, A.: A causal link between visual spatial attention and reading acquisition. *Current Biology*, 22(9), pp.814-819 (2012)
24. Feigenson, L., Dehaene, S., Spelke, E.: Core systems of number. *Trends in cognitive sciences*, 8(7), pp.307-314 (2004)
25. Ness, D., Farenga, S.J.: *Knowledge under construction: The importance of play in developing children's spatial and geometric thinking*. Rowman & Littlefield Publishers (2007)
26. Simion, F., Regolin, L. Bulf, H.: A predisposition for biological motion in the newborn baby. In: *Proceedings of the the National Academy of Sciences*, 105(2), pp.809-813 (2008)
27. Gelman, S.A.: *The essential child: Origins of essentialism in everyday thought*. Oxford University Press (2003)
28. Best, J.B.: *Cognitive Psychology*, 5th ed., pp. 15–17 (1999)
29. Skinner, R.A., Piek, J.P.: Psychosocial implications of poor motor coordination in children and adolescents. *Human movement science*, 20(1), pp.73-94 (2001)
30. Arora, N., Agarwal, N., Reddy, S.R.N.: FunPi: An Interactive Learning Experience Using Story Narration. In *Proceedings of the Sixth International Conference on Computer and Communication Technology 2015*, pp. 398-402. ACM (2015).
31. *The greek anthology*. 1. W. Heinemann (1916)
32. Leap Motion Websites, <http://www.leapmotion.com>
33. Beardsworth, T., Buckner, T.: The ability to recognize oneself from a video recording of one's movements without seeing one's body. *Bulletin of the Psychonomic Society*, 18(1), pp.19-22 (1981)
34. Knoblich, G., Flach, R.: Predicting the effects of actions: Interactions of perception and action. *Psychological Science*, 12(6), pp.467-472 (2001)
35. Knoblich, G., Prinz, W.: Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: human perception and performance*, 27(2), p.456 (2001)
36. Mazalek, A., Nitsche, M., Chandrasekharan, S., Welsh, T., Clifton, P., Quitmeyer, A., Peer, F., Kirschner, F.: Recognizing self in puppet controlled virtual avatars. In: *Proceedings of the 3rd International Conference on Fun and Games*, pp. 66-73. ACM (2010)
37. Von Kleist, H., Neumiller, T.G.: On the marionette theatre. *The Drama Review: TDR*, pp.22-26 (1972)
38. Video game and operating system control with gestures, <http://uwyn.com/gamewave/>
39. Westby, C.: Development of narrative language abilities. *Language learning disabilities in school-age children*, pp.103-127 (1984).
40. Page, J.L., Stewart S.R.: Story grammar skills in school-age children. *Topics in Language Disorders* 5(2), pp.16-30 (1985).
41. Stein, N. L., Glenn, C. G.: *An Analysis of Story Comprehension in Elementary School Children: A Test of a Schema*, (1975).

42. Loban, W.: Language Development: Kindergarten through Grade Twelve. NCTE Committee on Research Report No. 18, (1976).
43. Chappell, G. E.: Oral language performance of upper elementary school students obtained via story reformulation. *Language, Speech, and Hearing Services in Schools*, 11(4), pp. 236-250, (1980).
44. MacLachlan, B. G., Chapman, R. S.: Communication breakdowns in normal and language learning-disabled children's conversation and narration. *Journal of Speech and Hearing Disorders*, 53(1), pp. 2-7, (1988).
45. Merritt, D. D., Liles, B. Z.: Story Grammar Ability in Children with and without Language Disorder Story Generation, Story Retelling, and Story Comprehension. *Journal of Speech, Language, and Hearing Research*, 30(4), pp. 539-552, (1987).
46. Luck, S. J.: An introduction to the event-related potential technique. MIT press, (2014).
47. Liu, B., Wang, Z., Song, G., Wu, G.: Cognitive processing of traffic signs in immersive virtual reality environment: An ERP study. *Neuroscience letters*, 485(1), pp.43-48, (2010).
48. Invitto, S., Faggiano, C., Sammarco, S., De Luca, V., De Paolis, L. T.: Interactive entertainment, virtual motion training and brain ergonomics. In *Intelligent Technologies for Interactive Entertainment (INTETAIN)*, 2015 7th International Conference on, pp. 88-94. IEEE, (2015).