Engaging Spect-actors with Multimodal Digital Puppetry

Céline Coutrix[1], Giulio Jacucci[12], Anna Spagnolli[3], Lingyi Ma[1], Matti Helin[1], Gabriela Richard[4], Lorenza Parisi[5], Stefano Roveda[6], Prayag Narula[1]

[1]Helsinki Institute for Information Technology HIIT, Aalto University firstname.surname@hiit.fi
[2]Department of Computer Science, University of Helsinki, firstname.surname@helsinki.fi
[3]HTLab Dept. of General Psychology University of Padova, anna.spagnolli@unipd.it
[4]Educational Communication and Technology, New York University gabriela@nyu.edu
[5]Facoltà di Scienze della Comunicazione, Università di Roma “La Sapienza” lorenza.parisi@gmail.com
[6]Studio Azzuro, stf@studioazzurro.com

ABSTRACT

We present Euclide, a multimodal system for live animation of a virtual puppet that is composed of a data glove, MIDI music board, keyboard, and mouse. The paper reports on a field study in which Euclide was used in a science museum to animate visitors as they passed by five different stations. Quantitative and qualitative analysis of several hours of videos served investigation of how the various features of the multimodal system were used by different puppeteers in the unfolding of the sessions. We found that the puppetry was truly multimodal, utilizing several input modalities simultaneously; the structure of sessions followed performative strategies; and the engagement of spectators was co-constructed. The puppeteer uses nonverbal resources (effects) and we examined how they are instrumental to talk as nonverbal turns, verbal accompaniment, and virtual gesturing. These findings allow describing digital puppetry as an emerging promising field of application for HCI that acts as a source of insights applicable in a range of multimodal performative interactive systems.

Author Keywords

Multimodality, performative interaction, engagement, co-creation, virtual puppetry, field study, museum.

ACM Classification Keywords

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

INTRODUCTION

Puppetry is a 30,000-years-old art [4]. Most puppetry involves storytelling, and its impact is determined by the ability to create a fictional space for the spectators that has aspects in common with magic and with play. If the spectators are invited to act in this fictional space, a truly engaging experience co-created by the puppeteer and the spectators can arise. Thus, spectators are not just passive recipients of the storytelling, but become active characters in it – *spec-actors* in short.

The animation of digital objects or digital puppets does incite imagination of spectators in a particular way. Spectators interacting with such digital “beings” enter also a fictional space. Digital objects can be animated in many different ways and can be transformed in real time to provide novel possibilities for engagement and co-creation.

Computer-mediated puppetry has been used mostly for animation production purposes and not for live public performances. In contrast, in this paper we focus on live digital puppetry with live audience using real-time multimodal animation as a promising area of application for engaging spectators.

We describe a concrete system, Euclide, installed in a science museum in Naples, Italy. Euclide utilizes multimodal inputs: a data glove, a standard keyboard, a MIDI keyboard, and a mouse (see Figure 1, left). The system has a control center from which the puppeteer operates five different stations at the museum (see Figure 1, right). These are audiovisual output installations for the digital puppets, monitored by a camera and microphone. The puppeteer is able to switch between stations swiftly and choose to interact with passers-by. We are interested in characterizing the interactive and performative aspects of this application area by describing its three following elements:

- How the puppeteer makes use of the complex multimodal system,
- The interaction sessions in their structure and lengths on this example of a non-work related interaction,
- How engagement is co-constructed by the puppeteer and “spect-actors”.

Others elements that are of interest are left for future work, such as considering the social context and the meaning making process. Those elements compose are what makes

NordiCHI 2010, October 16–20, 2010, Reykjavik, Iceland.
Copyright 2010 ACM ISBN: 978-1-60558-934-3...$5.00.

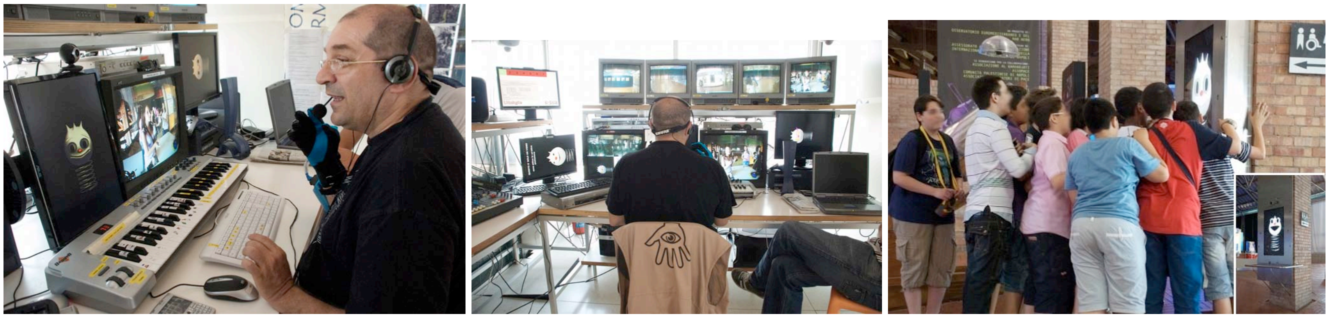


Figure 1. Puppeteer interacting with the multimodal system Euclide in the control room (left and center) and interactive station and young audience interacting with the system (right).

puppetry an interesting application field for HCI. In order to study the first three, we collected composite videos of puppeteer, spect-actors, and puppet over several days, with three professional puppeteers. The findings utilize video analysis employing quantitative and qualitative methods. The variables we considered arose form these research questions and the data, as we explain in the Study section.

RELATED WORK

Real-time Digital Puppetry

There are surprisingly few examples of digital puppetry, given the possibilities offered by current technologies.

Computer puppetry generally is used to refer to mapping the movements of a human performer to an animated character in real time [32]. Dontcheva et al. [13] introduce a novel motion-editing technique that derives implicit relationships between the animator and character. This and other work are generally motivated by the development of tools for animation production. Not only human performers are used to this end. Mazalek and Nitsche [28] addresses production and performative challenges involved in creating machinima through the development of tangible interfaces for controlling 3D virtual actors and environments.

Other examples include development of such production tools for pupils and use in everyday play. Barnes et al. [2] present "Video Puppetry," in which the puppeteer first creates a cast of physical puppets, using paper, markers, and scissors. An overhead camera tracks the motions of the puppets and renders them on a new background while removing the puppeteer's hand.

Liikkanen et al. [26] present PuppetWall, a multi-user, multimodal system intended for digitally augmented puppeteering. This application allows natural interaction to control puppets and manipulate playgrounds comprising background, props, and puppets. PuppetWall utilizes hand movement tracking, a multi-touch display, and emotional speech recognition for interfacing. Here, however, the idea is that the visitors are themselves having fun with the puppets and the puppeteer is not hidden backstage.

Chinese shadow puppetry has been implemented in a system called I-Shadows [14]. The installation allowed children to create stories for an audience. In this theater, the user interacts with the system by controlling a physical puppet of either a hero or a villain, whose movements are interpreted by a vision system that sends the information to the autonomous character's environment. With CoPuppet [6], a system for collaborative puppetry is presented. The CoPuppet project explores the possibilities offered by multimodal and cooperative interaction, in which performers, or even audience members, are called upon to affect different parts of a puppet through gestures and voice.

Some systems are not directly puppetry applications but include interactive aspects whereby digital objects are influenced by spectators. Affective Painting [33] supports self-expression by adapting in real time to the perceived emotional state of a viewer, which is recognized from his or her facial expressions. Cavazza et al. [8] introduce a prototype of multimodal acting in mixed-reality interactive storytelling, in which the position, attitude, and gestures of spectators are monitored and influence the development of the story. Camurri et al. [7] propose multi-sensory integrated expressive environments as a framework for performing arts and culture oriented to mixed-reality applications. They report an example in which an actress's lips and face movements are tracked by the EyesWeb system and her voice is processed in real time to control music.

Frameworks for Performing Media and Spectators

Recently, several researchers have applied different performative or theatrical metaphors to describe the emergence of novel interaction formats and experiences that are related to real-time animated puppetry.

Dalsgaard and Koefoed Hansen [10] observe how the user is simultaneously operator, performer, and spectator. A central facet of aesthetics of interaction is rooted in, as they put it, the user's experience of herself "performing her perception." They argue that this three-in-one situation is always shaping the user's understanding and perception of the interaction, and they address the notion of the performative spectator and the spectating performer.

Reeves et al. [31] present a taxonomy with four broad design strategies for the performer’s manipulations of an interface and their resulting effects on spectators: the “secretive”, wherein manipulations and effects are largely hidden; the “expressive,” in which they tend to be revealed, enabling the spectator to fully appreciate the performer’s interaction; the “magical”, where effects are revealed but the manipulations that caused them are hidden; and, finally, the “suspenseful”, wherein manipulations are apparent but effects are revealed only as the spectator takes his or her turn. Benford et al. [3] extend the above framework for designing spectator interfaces with the concept of performance frames, enabling one to distinguish audience from bystanders. They conclude that ambiguity to blur the frame can be a powerful design tactic, empowering players to willingly suspend their disbelief.

Also central to the discussion is the framework of “Interaction as Performance” [17][18][19]. This framework is based on anthropological studies of performance that have roots in a pragmatic view of experience. In particular, the framework proposes a variety of principles aimed at describing performative interaction. One of these is that of *accomplishment and intervention*. The etymology of the term “performance” shows that it does not have the structuralist implication of manifesting form but, rather, a processual sense of bringing to completion or accomplishing. The concept of *event and processual character* is also key: performances are not generally amorphous or open-ended; they have diachronic structure, a beginning, a sequence of overlapping but isolable phases, and an end. *Expression and experience* is another element of import. According to pragmatist views, an experience is never completed until it is expressed. Also, in an experience there is a structural relationship between doing and undergoing.

These novel frameworks originate from concurrent trends in HCI, including the emergence of installations as a delivery platform for interactive experiences. Installations as also tangible interfaces have the property of providing a stage on which the user becomes at times a performer. Other trends include attention to the fact that performers in general have more and more technology to mediate their interaction with spectators.

The system we present now is a novel multimodal puppetry application that aims at engaging spectators in performative sessions of interaction.

THE SYSTEM: EUCLIDE

Euclide is a virtual puppet (see Figure 2) that has an engaging role in the visit of a science museum in Naples, Italy. The system offers a multimodal interface to the puppeteer in order to animate a virtual puppet and entice the audience.

Figure 1 (left) shows a hidden animator controlling the movements and mimicry of a virtual character through a multimodal interface including a data glove. The animator’s hand movements “activate’’ the virtual character, controlling the mimicking, and digital effects alter the animator’s voice.

The rendering of the character appears on a screen in a second space, the “stage” (see Figure 1). Five stages are scattered about the museum. The animator monitors the audience members via a microphone and a camera and reacts to them (see Figure 1). Therefore, the puppeteer can react and respond to people talking to the character.

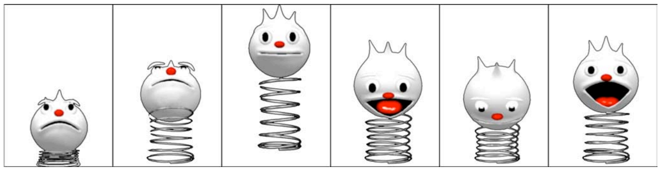


Figure 2. Virtual character of the Euclide system.

The system offers 100 different features to the puppeteer for animating the character, among them jumping (see Figure 2) or dressing like Santa Claus (see Figure 4). To allow use of this great expressive power, with many elements sometimes utilized simultaneously, different modalities are proposed. The interface includes 11 screens, two computer keyboards, two mice, a data glove, a microphone, headphones, and a MIDI keyboard, all in the control room (see Figure 1, left and center). Among these devices, three screens, one computer keyboard, one mouse, the MIDI keyboard, the microphone, and the glove are dedicated to real-time puppetry. The other devices are dedicated to system launch, switching between interactive areas or setting the puppet to inactive in order for the puppeteer to take a break.

Facial Expressions

Using a data glove, the puppeteer can horizontally open/close both eyes by bending the index finger of the glove at i2 (see Figure 3), vertically open/close both eyes by bending the index finger of the glove at i3, and move the eyebrows around the eye by bending the middle finger of the glove at m2.

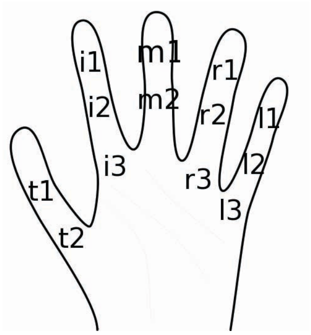


Figure 3. Control points on the fingers of the glove.

To control the mouth of the puppet, the puppeteer can bend the thumb of the glove at t1 to open/close the mouth and bend the thumb of the glove at t2 to raise/lower the corners of the lips. This enables, respectively, making the puppet talk and look happy or sad.

The color of the skin can be controlled by four of the MIDI keyboard sliders, on the basis of the three RGBA channels.

Some costume accessories can be added to the puppet's face, such as hats, glasses, and mustache, by pressing keys in the left portion of the MIDI keyboard (see Figure 1).

Body Movements and Expressions

The puppet has global freedom of movement in 3D space. For this, the puppeteer uses the wheel of the mouse to translate the puppet in the depth dimension. Two sliders on the MIDI keyboard enable translation of the puppet vertically and laterally. The puppeteer can also change the orientation of the puppet to, for example, have its head upside down, by pressing the "M" key on the computer keyboard (see Figure 1).

The body of the puppet can be moved locally by jumping, turning, and bending its spring. Jumping extent is controlled by the little finger of the glove (l3 in Figure 3). The lateral movement of the mouse controls rotation extent around Bit's vertical axis. Bending forward/backward is controlled with the y dimension of mouse movement. Jumping, rotation, and bending maximums are each controlled by one of the keyboard sliders.

As well as the face, some costume items can also be added to the puppet's body, such as a Superman costume, by pressing keys in the left part of the MIDI keyboard.

Background

The background behind the puppet can be modified, from black to live video of the puppeteer or to still images or prerecorded videos. The interface also allows fine-tuning the transparency of the background (via the wheel on the MIDI keyboard), the orientation of the background in the vertical plane ("N" on the computer keyboard), the zoom of the background image (controlled via a potentiometer on the MIDI keyboard), and the 3D position of the background (in the vertical plane and near to distant, each controlled via a potentiometer on the MIDI keyboard).

The design of this complex form of interaction has been driven by the requirements from the museum to use generic and flexible hardware, the puppeteers' requirements for a large number of functionalities, and usability. For instance, the location of the controls we presented here has been studied, to allow the puppeteer to perform particular movements simultaneously. For instance, moving the mouth and the eyes of the puppet have to be enabled simultaneously for puppetry.

Early versions of the system were not as easy to use, and the interface has been improved with puppeteers' assistance. For instance, opening the mouth was tiring before: it was controlled by the index finger and done constantly. To overcome this difficulty pointed out by a puppeteer, the thumb now controls the opening of the mouth.

Also, expressive gestures of the puppeteer were mapped to the expression of the puppet. For instance, clenching the fist makes the puppet look angry, opening the hand makes the puppet look happy, and relaxing the hand makes the puppet seems neutral.

THE STUDY

We studied the system in the science museum in Naples, where it is used for engaging the audience in their visit.

Data Collection

We video-recorded the use of the multimodal puppetry system installed in the museum in winter and spring 2009. We recorded several hours of interaction from four viewpoints synchronized in one video (see Figure 4): the audience from front left and front right, and the real-time image of both puppet and puppeteer from the front. The recording also included sound. In total, three puppeteers interacted with the system. In addition to the recording, there was an interview with the most experienced puppeteer, to examine his use of the system.

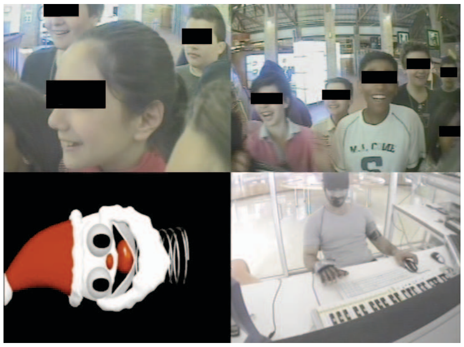


Figure 4: Composite video, in which the puppet is dressed as Santa Claus.

Data Analysis: Procedure and Reliability

Coding Scheme

We employed constant comparison analysis [15] to the video data collected. While initially conceptualized as an inductive process, whereby theory emerges from close reading and analysis of the data, contemporary practice asserts that deductive or abductive processes can also be used in constant comparison analysis [25]. For the purposes of this analysis, we used an abductive process, letting the theoretical constructs emerge from the data as well as from existing theory such as the PAD scale [30] for measuring emotion.

Constant comparison analysis as conceptualized by Glaser and Strauss [15] is suitable for the analysis of multiple data sources, including observation. Constant comparison analysis is a fundamental element of grounded theory [15] and is a means of systematically generating theory by working closely with the data and setting aside previous theoretical assumptions. The coding process is methodical and starts with open coding (creating initial categories), continues with axial coding (parsing out relationships), and finally uses selective coding (coding around core categories) [34]. The theory is developed once core categories are repeated enough in the data for a point of theoretical saturation to be reached. We discussed the coding between three researchers and our final consensual scheme included 35 tracks for annotation. The audio was annotated using Praat software [5] and the video was annotated using Anvil software [23].

Selection of Clips

We selected clips via both a deductive and an inductive process. We applied a deductive process because clip selection was grounded in our research questions and objectives [12]. However, we were aware of a clip's narrative power [12] for understanding the phenomena. In other words, each clip was analyzed to understand whether it fit the scope of our research as well as how illustrative it was. Accordingly, the selection of clips also took an abductive approach, particularly influenced by Goldman's technique of negotiated clip selection [16]. The video material was evenly divided between the three coders.

Agreement Betweens Coders

For reliability of our annotation, we also assessed the inter-coder agreement rate. As the basis, we use Cohen's kappa index, which is a descriptive statistics that summarizes agreement as percentage of the cases on which the coders agree across a number of objects against bare chance. Index at 0.80 to 1.00 is considered very good agreement; 0.60 to 0.80 considered good agreement; 0.40 to 0.60 considered moderate agreement; 0.20 to 0.40 considered fair agreement; and 0.20 or less considered poor agreement [24]. In practice, if our assessed agreement was less than 0.6, we discussed between coders and made revision in the data, then continued testing and revising until the agreement was satisfactory, i.e. greater than 0.6.

FINDINGS

In the presentation of the findings, we focus on how the puppeteers make use of the multimodal system, how the sessions are structured, and how engagement is co-constructed by puppeteer and spect-actors. These findings draw from the statistics distilled from the quantitative and qualitative analysis of interaction sessions.

Multimodal Puppetry

Several features' use is important: the speech, lips, and eye movements, as well as bending and rotation (see Figure 5). This is related to the fact that these are distributed by microphone, different fingers of the data glove, and the mouse. Indeed, the interface for these features has been provided in the data glove and via the mouse in order for these to be easy to use simultaneously. These devices, respectively, are attached constantly to the puppeteer's right hand and in the resting location for the left hand most of the time (i.e., when it is not required for operating a keyboard). These devices are used all the time to make to animate the speech, facial expressions and body of the puppet. Other seldom-used features, such as costume or background change, are handled with less direct devices (e.g., the MIDI keyboard).

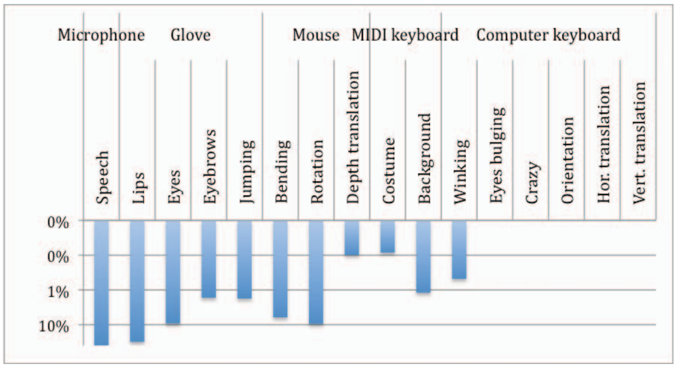


Figure 5: Percentage of the duration of use for each feature (Logarithmic scale).

Figure 5 also shows that third most used features are the animation of the body (bending and rotation). Then comes the use of special effects, such as costume or background change used in specific situations when relevant. Least important features are more complex "physical" actions such as moving around.

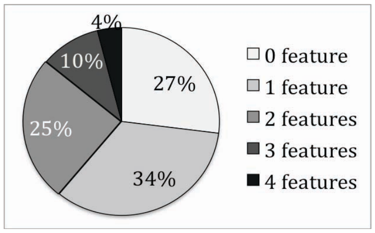


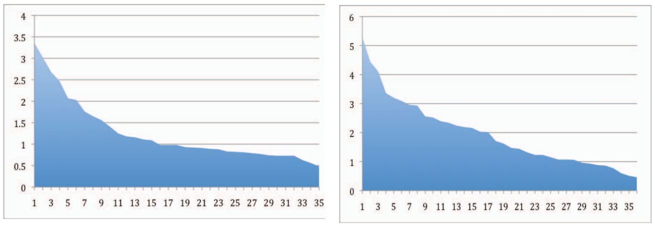
Figure 6: Simultaneous use: no, individual, or multiple features.

In addition, we found that 2 to 4 features are used simultaneously 39% of the time (see Figure 6), whereas no features are used 27% of the time and a single feature 34% of the time. This highlights the actual significant simultaneous use by the puppeteers of the multimodal resources [27].

More precisely, the most used set of features is speech alone, followed by speech and lips together characterizing the talking of the puppet and then lip movement alone to characterize expressions such as smiling. Speech without lip movements is the most frequent combination, meaning that the speech and lips are not fully in synchronization. This points to possible improvements in automating the lips' synchronization.

Figure 7 shows that the distribution of speech and between-speech durations for the puppeteer follows a power law. Indeed, the puppeteers keep segments of speech short, trying to make the audience react rather than monopolizing the conversation (see Figure 7a). In a similar manner, they try to show that the puppet is paying attention to the audience and therefore rarely let it remain silent for long (see Figure 7b).

Comparing speech and silent use of features, we found that the puppet is considerably active also when the puppeteer is silent. Consequently, the use of the features goes beyond a basic animated character as we further explain in section "Resourceful Co-constructing of Engagement".



(a) **(b)**

Figure 7: Distribution of the 35 speech (a) and 36 silence (b) durations for puppeteers, in seconds.

Phase	Audience members' action	Puppet's reaction
Approach	Enter (one or several people, or only a voice)	Stops activity
Testing	Present themselves	Presents itself
Playing	Laugh	Skips happily Asks what is funny
	Say bad words or abuse a bit	Repeats in a mechanical way Cries, complains, and goes away
	Say a keyword	Changes costume Tells a story Sings a song
	Ask questions	Answers normally Answers as if crazy or slow Answers and asks the audience the same
Ending	Greet	Greets

Table 1. Summary of the structure of sessions as reported by the most experienced puppeteer.

Emergence of Performative Structures

We grounded the following analysis on the performative framework presented in the related work.

Engaging the Audience Throughout the Sessions

People in the audience don't talk to each other, 99% of the time, and they don't pay attention to the area outside the interactive space, 98% of the time. They prefer to talk to the puppet (65% of the time). In addition, their pleasure and arousal was never annotated as negative.

In all sessions we analyzed, the number of people in the audience tends to increase until it reaches a maximum near the middle of the session and then decreases towards the end. The increasing phase demonstrates that the system appeals to visitors. The audience also tends to use body movement such as waving in the second part of the session.

Length of Sessions

The length of session seems to follow a power law: four clips last more than nine minutes, five are between two and nine minutes long, nine last 1–2 minutes, and seven are very short. However, we should note that some of the sessions are a continuation: the same group already interacted with the puppeteer. Moreover, in some cases the sessions are interrupted by teachers or parents who want to move on in the visit.

Structure of Sessions

Drawing from previous work on interaction with entertainment-related interactive installations [20], we analyzed the structure of the sessions. Different phases were annotated: Approach, Testing, Playing, and Ending. Approach is the phase in which participants enter the interaction area, observing. Testing is that in which they start trying to interact with the installation, by taking a particular action such as touching the screen in order to find out which actions have an effect on the installation. Playing is the phase in which participants interact with the installation in an aware, active, and involved way. This phase includes the climax or main action of the interaction session. Ending is the phase in which participants have their attention diverted from the installation before they leave. Figure 8 shows that these phases are balanced in the sessions, with the exception of Approach, which could not always be recorded. In this regard, the structure of the sessions as explained by the puppeteer during the interview confirms this distribution: more actions are proposed during the playing phase (Table 1).

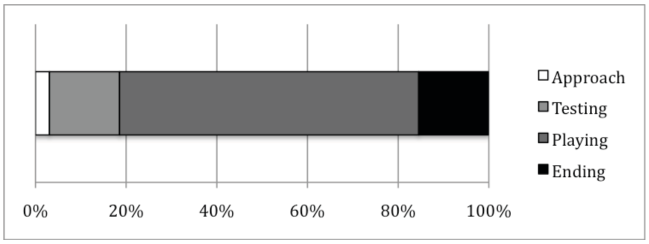


Figure 8: Average relative length of phases in analyzed sessions (100% is the average duration of sessions).

Considering these four phases, we studied the evolution of the perceived emotion of the audience through pleasure and arousal (see Figure 9). For this we considered the facial expressions within the group, as well as their oral expressions and gestures. We found out that pleasure is high during the approach and testing phases while arousal is low. Pleasure is less important ($p<0.01$) during playing phase while arousal is more important ($p<0.01$). During the ending phase, the visitors show more pleasure ($p<0.01$) with less ($p<0.01$), but still positive, arousal. The evolution of pleasure and arousal demonstrates that the audience starts interacting happily but calm, then gets excited and finishes the interaction happily and calm again.

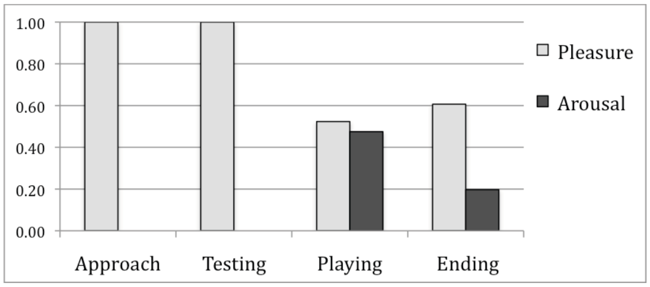


Figure 9: Evolution of pleasure and arousal in the sessions' phases (annotated between -1 and 1).

We also noticed that in all sessions the audience's members carefully take turns for saying goodbye to the puppet, regardless of the way they were interacting during the session, either through a leader or simultaneously. For instance, Figure 10 shows the audience interaction throughout time in a one-minute-long session: They mainly talk all at the same time in a chaotic way. At three occasions a leader emerge in the group. Before they leave, they speak one after the other in order to say goodbye. This ending interaction pattern is recurrent in every session we analyzed.

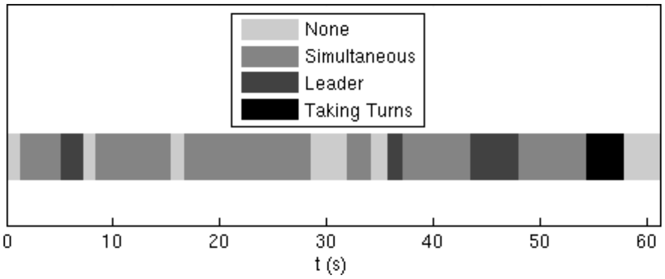


Figure 10: Group verbal interaction with puppet in a one-minute-long session: The audience's members mainly speak simultaneously, interacts through a leader three times, and take turns to talk to the puppet at the end.

Another interesting finding is that the structure of session can also be related to the use of the multimodal interface. Indeed, features like special effects (costume or background change) tend to be used during the Playing phase.

Resourceful Co-constructing of Engagement

The Puppet's Multimodal Resources

The puppeteer acts by using visual effects (eyeglasses, masks, clothes, etc.), and performing virtual movements (jumping, rotating, etc.); the different ways in which these nonverbal resources are used can be distinguished on the basis of their relation to talk, as is done with natural gestures [22]. The following four cases have been observed.

Pure special effects: This is the case in which nonverbal resources represent the predominant modality for interacting with the visitors; they are deployed to address visitors who would hardly be attracted by speech, such as very young babies or a far-away, busy audience.

Nonverbal turns: In this case, nonverbal resources are used to make an individual contribution within a longer interaction that also features verbal exchanges. Examples include wearing some accessories on the special request of a visitor, and subsequently commenting on it, and explaining a detail by displaying a picture of it.

Verbal accompaniment: This is the case in which nonverbal actions, such as acrobatic movements, are accompanied by speech. The communication relies on the nonverbal resource, which dictates its structure and meaning. An example is provided below: the puppet makes a jump commented upon by an accompanying onomatopoeic sound.

Example 1 (Disc 2, video 1, f 1:52)

(Nonverbal events are in double parentheses. Extending over multiple columns indicates that the preceding sound is stretched.)

```
Puppet:   Ho una molla
           I've got a spring
           Mi serve per saltare
           I need it to jump
           ⌈  Dong::   dong:: ⌉
               Dong       dong
           ⌊((Starts jumping))⌋
```

As the puppet starts jumping, the main resource to communicate with the visitor switches from verbal to visual: the onomatopoeic sound "dong dong" refers directly to the jump both in its duration and in its meaning. By producing it, the puppet continues the verbal communication with the audience even during the jump; exclusive room is not given to the visual resources, but the puppeteer signals the change in priority from verbal to visual.

Virtual gesturing: In the final case, priority is given to speech; visual features are used in synch with speech, following its structure and duration. The head movement in example 2 provides an instance of this.

Example 2 (Disc 2, video 1, f 7:57)

(The notation conventions are the same as for the previous example; in addition, events are in brackets when they overlap in time.)

```
Visitor:  Dove sei?
          Where are you?
Puppet:   Sono⌈::((pause))⌉
          I'm
              ⌊((Looking down))⌋
          ⌈qui      dentro,⌉
            inside      here
          ⌊((Looking around))⌋
          non mi vedi?
          can't you see me
```

Circular movement of the head is produced as the puppeteer starts replying to the question "Where are you?" and continues while the sentence is interrupted, filling the gap left by the search for an answer. The movement then stops and restarts with a different function when the answer "inside here" is finally produced: the puppet's head, moving in a circle inside the monitor, shows what "here" refers to and performs a pointing gesture. In natural communication, verbal and nonverbal resources contribute to the creation of one, joint meaning and vary in their mutual dependence [22][29]. The four cases considered here suggest that this holds true for mediated communication as well, adapted to the specific resources of the communication medium – as is the case with the pointing gesture performed with the puppet's head.

Interactional Consequences

The multimodal resources do not just feed communication; they also define who the puppet is. While the repertoire of special effects *per se* is limited and cannot keep the visitor's interest for a long time, talk can sustain the interaction longer. What the puppet does is entertain itself in conversations with the visitors, wherein talk is the main resource and the visual effects are instrumental to it. The visitors in our video recordings inquire about the puppet's personal information (i.e., name, age, family, and sex), about what it can do, and about its general knowledge of the "real" world (TV shows, songs, and popular people). The puppet is so successful in this that most conversations are interrupted forcedly by an external intervention, from a parent or a teacher soliciting the young visitors to leave.

The puppet's resources as a conversation partner distinguish it from other conversational agents, as is apparent from the cases of verbal abuse. In example 3, the puppeteer connects with a puppet located in a new room, with some young visitors trying to interact with it. Immediately after the puppeteer switches on, the conversation in example 3 unfolds.

Example 3 (Disc 1, video 3, f 00:20)

(Same notations as in the previous examples; inaudible sound is between rounded brackets; non-verbal events are in double rounded brackets.)

```
Child:    (Scemo)
           Dumb
Puppet:    Vabbeh adesso basta.
           All right; let's stop this
Child:     ((turns back, surprised))
Puppet:    Ogni volta che scemo, ogni volta
           che scemo.
           All the times dumb, all the times dumb.
           E tu invece come sei?
           And what about you instead, what are you?
Child:     ((goes away))
```

The abuse takes place while the puppet is not animated. The visitor tried to interact with the puppet and to understand what it could do, in an explorative attempt that has already been observed with other performative technologies [20]. Since the visitor did not see any reaction except for a non-motivated friendly look, she judged the character to be "dumb." As the puppeteer started talking (and reacting to the specific, situated abuse), the abuse came rapidly to an end. The appearance of the puppet remained cartoon-like, yet the talk redefined the capacities of the puppet, reconfiguring its social presence.

As is shown by De Angeli & Brahnam [11], visitors probe the cognitive abilities of a virtual character. The conversational agents studied by De Angeli & Brahnam, in this situation, mocked a human speaker and failed to show human conversational competence; in fact, the abuse often focused on the poor quality of the speech. A puppet instead can participate more properly in a verbal exchange, and this is probably responsible for the way in which abuse episodes are concluded. Let's consider the case of irony. In extract 4, below, for instance, the puppeteer recognizes irony, which relies on the ability to understand implicit meaning [1], and is able to defeat the abuse.

Example 4 (Disc 1, video 2, f 06:13)

```
Child:     Ma::: chi e è tuo padre?
           But      who's your father
Puppet:    Ma io ne ho un sacco di papà
           Well, I have a lot of fathers
Child:     E allora tua mamma::: ((turns to a
           friend))
           Then your mom:::
Puppet:    ↑Allora mia mamma? ((increased
           volume))
           Then my mom
Child:     ((pause)) ⌈che lavoro fa?⌉
                     What does she do for a living
Puppet:              ⌊↑Allora mia mamma?⌋
           ((increased volume))
                       Then my mum
Child:     Che lavoro fa?
           What does she do for a living
Puppet:    AAAAAAH
Children:  hahahah
Puppet:    mia mamma è la scheda madre del
           computer
           My mom is the mother board of the computer
```

In this exchange, the visitor replies to the puppet's announcement ("I have a lot of fathers") with a question that implies offense to the puppet's mother ("Then your mom?"). The puppeteer recognizes the implicit meaning in the visitor's words before the sentence is even completed: he repeats the visitor's ironic words twice at a higher volume and in an angry tone, responding to the indirect offense with an indirect threat. The visitor then completes the sentence ("What does she do for a living?"), the puppet reacts with a (funny) scream ("AAAAH!"), all visitors laugh, and the conversation continues normally ("My mom is the motherboard of the computer").

In synthesis, the different resources available to the puppeteer are combined together in multimodal communication with several specificities. The resulting social presence attracts the visitors effectively and avoids the low status that is attributed to other virtual characters.

DISCUSSION AND CONCLUSIONS

We described a system of digital puppetry and reported on a field study with the aim of characterizing a promising application field for HCI. The system we described featured digital puppetry using real-time multimodal animation. The attractiveness for HCI in this area lies in that it provides the possibility to introduce advanced interface techniques. Other examples of digital puppetry, such as those described by Liikkanen et al. [26] utilize expressive (emotional) feature-tracking from voice. This installation utilized a data glove and a variety of other input devices used in a multimodal way. In addition, puppetry provides a case for recent frameworks of interaction that focus on performative situations or installations. In particular, it illustrates a form of interaction where engagement is co-constructed by spectators and puppeteer through the interface.

Multimodal Use of the System by the Puppeteer

We described the use of the system as truly multimodal since several features were used simultaneously. The use of features in ranked by duration served to (1) animate the speech, (2) give the puppet facial expressions, (3) animating the body, (4) use special effects, and (5) perform more complex "physical" actions such as moving around. The puppeteers used mostly the mouse and glove.

We inferred as implications for design the opportunity to automatically animate the lips and synchronize them with the speech of the puppeteer. This might free the puppeteer to concentrate more on expressive or symbolic acts. In addition, the expressive dimension of the interface, e.g. clenching the fist making the puppet look angry, can be further investigated.

Performative Structures for Brief Interactions

In this case, the spectators were mostly pupils and teenagers and sessions lasted more than two minutes, on average. It must be noted, however, that some sessions are interrupted by teachers and others are a continuation of a previous interaction. While improvised, sessions conform to a general structure, which is also reported in interviews with the puppeteers. These structures have been observed to emerge in the use of installations [20, 21]. The groups of spectators generally are attentive to the installation (they did not talk to each other), actively interact with it, and show positive and growing interest as they interact. The puppeteer, therefore, is working with different resources, including a repertoire of gags, to be able to keep spectators engaged for several minutes.

Implications of the found structure of the interaction lie in the design of an extended computer animated puppet dedicated to the four inactive stations in the museum like presented in [9].

Spect-actors and the Co-constructing of Engagement

The puppeteer uses nonverbal resources like visual effects and performing virtual movements. We examined the relation of these nonverbal resources to talk identifying types of use as nonverbal turns, verbal accompaniment, virtual gesturing. However we noticed that in the multimodal resources is talk that can sustain interaction longer. Talk is the main resource and the visual effects are instrumental to it. For example verbal abuses usually addressed to autonomic virtual characters (that generally are attributed a low status) are here resolved through irony by the puppeteer in an effective way. We also observed how the narrative of the sessions emerges from the interaction and contribution of both the puppeteer and the spect-actors.

This area of application is particularly engaging because the spectators are called upon to interact with the puppets in improvised sessions. They express themselves and therefore feel that they have the role of protagonist. Current analysis and frameworks for installations or performing media anticipate some of these themes. Dalsgaard and Koefoed Hansen [10] point to the multiple roles of the user operator, performer, and spectator (see [17], [18]). Jacucci et al. [19] point to a variety of elements characterizing interaction as performance, including the structural relationship between expression and experience. These frameworks ascribe to the user an important role in the construction of the resulting performance. We believe these frameworks can be useful in further analysis of this emergent field.

Beyond the role of the different features of the multimodal system, we showed how puppeteer and spect-actors accomplish engagement in the sessions. In particular, spect-actors have a key part in creating the narrative – the gags are often inspired by what the spectators say. While it is situated and emergent, we described how engagement is the product of particular performative strategies and skills and how it relies on collective accomplishments of the mediated puppet (puppeteer) and the spect-actors.

REFERENCES

1. Attardo, Linguistic Theories of Humor. Berlin: Mouton de Gruyter, 1994.
2. Barnes et al., Video Puppetry: A Performative Interface for Cutout Animation, SIGGRAPH Asia 2008, 124:1-124:9.
3. Benford et al., The Frame of the Game: Blurring the Boundary between Fiction and Reality in Mobile Experiences, CHI 2006
4. Blumenthal, Puppetry and puppets, Thames & Hudson, 2005.
5. Boersma, Weenink, Praat software, http://www.fon.hum.uva.nl/praat/

6. Bottoni, et al., CoPuppet: Collaborative Interaction, Randy Adams, Steve Gibson and Stefan Müller Arisona eds Virtual Puppetry, In Transdisciplinary Digital Art. Sound, Vision and the New Screen, pp. 326-341.

7. Camurri et al. Communicating Expressiveness and Affect in Multimodal Interactive Systems, IEEE MultiMedia, 12(1), 2005, pp. 43-53.

8. Cavazza et al., Multimodal Acting in Mixed Reality Interactive Storytelling, IEEE MultiMedia, 11(3), 2004, pp. 30-39.

9. Coutrix et al., Interactivity of an Affective Puppet, Adj. Proc. of Ubicomp 2009.

10. Dalsgaard, Koefoed Hansen, Performing Perception—Staging Aesthetics of Interaction, ACM TOCHI, 15(3), 2008.

11. De Angeli, Brahnam, I hate you! Disinhibition with virtual partners. Interacting with computers 20, 2008, pp. 302-310.

12. Derry et al., http://visa.inrp.fr/visa/presentation/Seminaires/Journees_inaugurales/Video_Gdlines_JLS5_09Derry.pdf, 2009.

13. Dontcheva, Yngve, Popović, Layered acting for character animation, ACM TOG, 22(3), 2003.

14. Figueiredo et al., Emergent stories facilitated, In Spierling, U., and Szilas, N., editors, ICIDS, Springer LNCS vol. 5334, 2008, pp. 218–229.

15. Glaser, Strauss, The discovery of grounded theory. New York: Aldine, 1967.

16. Goldman-Segall, Points of viewing children's thinking: A digital ethnographer's journey. Mahwah, NJ: Erlbaum, 1998.

17. Jacucci, Interaction as Performance. Cases of configuring physical interfaces in mixed media. Doctoral Thesis, University of Oulu, Acta Universitatis Ouluensis, 2004.

18. Jacucci, G. and Wagner, I., Performative Uses of Space in Mixed Media Environments, In Davenport, E., Turner P., Spaces, Spatiality and Technologies, Springer, London, 2005.

19. Jacucci et al., A Manifesto for the Performative Development of Ubiquitous Media, Proc. 4th Decennial Conference on Critical Computing, pp. 19-28, 2005.

20. Jacucci et al., Bodily Explorations in Space: Social Experience of a Multimodal Art Installation, Proc. INTERACT 2009, Springer, 2009, pp. 62–75.

21. Jacucci, G., Morrison, A., Richard, G.T., Kleimola, J., Peltonen, P., Parisi, L., Laitinen, T., Worlds of information: designing for engagement at a public multi-touch display . In: ACM CHI '10: Proceedings of the 28th international conference on Human factors in computing systems, Pp: 2267-2276, 2010.

22. Kendon, Gesture: Visible Action as Utterance. Cambridge, UK: Cambridge University Press, 2004.

23. Kipp, ANVIL software, http://www.anvil-software.de/

24. Landis, Koch, The measurement of observer agreement for categorical data in Biometrics. Vol. 33, pp. 159-174, 1977.

25. Leech, Onwuegbuzie, Qualitative Data Analysis: A Compendium of Techniques and a Framework for Selection for School Psychology Research and Beyond. *School Psychology Quarterly*, 23(4), pp. 587 -604, 2008.

26. Liikkanen et al., Exploring emotions and multimodality in digitally augmented puppeteering, Proc. AVI 2008, pp. 339-342.

27. Liikkanen, L., Jacucci, G. & Helin, M. (2009) ElectroEmotion: A Tool for Producing Emotional Corpora Collaboratively. In: the Proc. of Affective Computing and Intelligent Interfaces ACII 2009. Amsterdam, The Netherlands IEEE.

28. Mazalek, Nitsche, Tangible interfaces for real-time 3D virtual environments, Proceedings of the international conference on Advances in computer entertainment technology, 2007.

29. McNeill, Gesture and thought. Chicago: University of Chicago Press, 2005.

30. Mehrabian, 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology: Developmental, Learning, Personality, Social, 14, pp. 261-292.

31. Reeves et al., Designing the spectator experience, Proc. of CHI 2005.

32. Shin et al., Computer Puppetry: An Importance-Based Approach, ACM Transactions on Graphics, 20(2), 2001, pp. 67-94.

33. Shugrina et al., Empathic painting: interactive stylization through observed emotional state, Proc. NPAR 2006, ACM Press, 2006.

34. Urquhart, An Encounter with Grounded Theory: Tackling the Practical and Philosophical Issues. In Trauth, E. (Ed.) Qualitative Research in Information Systems: Issues and Trends. Idea Group Publishing, London, 2001.