White Paper
Grant # HD-51031-10
Gesture, Rhetoric, and Digital Storytelling
Director: D. Fox Harrell, Ph.D.; Associate Professor of Digital Media, MIT
National Endowment for the Humanities
August 31, 2011

Narrative Description

The Gestural Narrative Interactive Expression (GeNIE) project was conceived to *develop new narrative models for emerging digital media technologies*. Technologies allowing for gestural input and output have become more prevalent, e.g. the iPhone, Nintendo Wii and 3DS, and laptops with multi-touch-screen input and accelerometers to measure motion. Addressing such technologies, we produced a new interactive narrative model for systems that change emotional tone, theme, perspective, and other content elements in response to embodied user input. We produced *scholarship that examines the philosophical implications and impact of the use of these emerging technologies*. Furthermore, we used this Level I grant to also produce a full prototype. In particular, we developed new interdisciplinary theory and technology exploring the relationship between gestural input and digital storytelling based in media studies, semiotics, psychology, human-computer interaction.

Project Activities

Overview

We created new theory regarding the relationship between gestural input and digital storytelling. Based on this theory, our team produced a platform called the GeNIE system to allow authors to create effective, gesture-driven interactive stories. In this document, the GeNIE system will be referred to variously as “the system,” “the prototype,” or “the platform” depending on the context in which is discussed. The GeNIE system allows for implementation of digital stories that read like animated graphic novels in which gestural input from users causes interactions between characters and narrative events in a storyworld to occur.

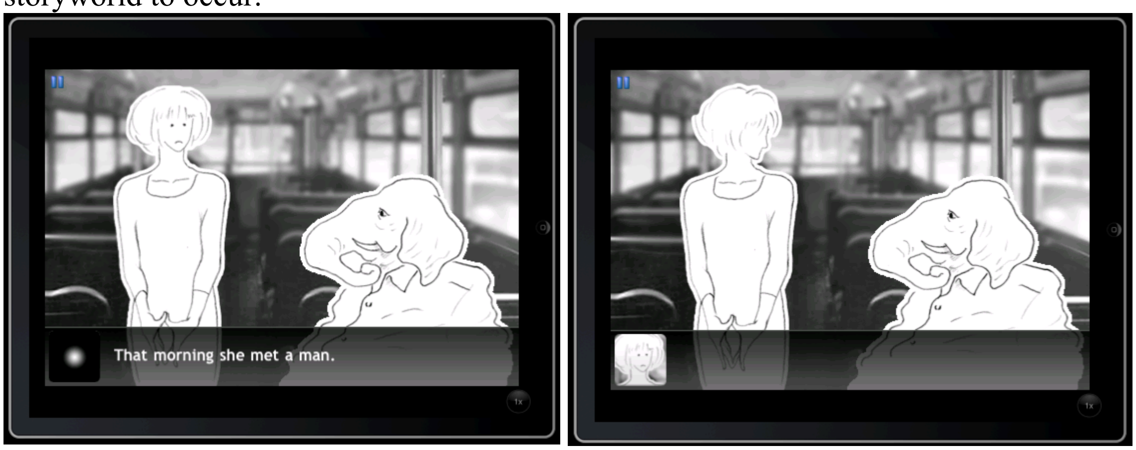


Figure 1: Screenshots from a test narrative implemented on a mobile “smartphone.” Gestural input causes events to occur in a storyworld and drives the narrative forward.

The project consisted of a range of activities to support development of new theory and technology. The PI convened weekly meetings with his collaborators Kenny K. N. Chow and Erik Loyer. These meetings were conducted as teleconferences between Cambridge, Los Angeles, and Hong Kong. The participants set up shared repositories for code and documents. The participants carried out a research process that was a combination of seminar (to discuss theory), scenario-based design (to develop the design), and agile programming (to develop code).

Table 1 summarizes proposed activities and the final status of each outcome. We completed, and in some cases exceeded, the activities proposed. A narrative description of the three main activities undertaken follows the table.

Development Phase	*Proposed Activity*	*Final Status*
June 2010 Weeks 1 - 2	Planning	Completed: Weekly meetings occurred and a design document was initiated.
Weeks 3 - 4	Planning	Completed: Weekly meetings; Analyzed Loyer's *Ruben and Lullaby;* Shared Harrell and Chow's theoretical framework; Design document developed
July 2010 Weeks 5 - 9	Problem Scenario Development	Completed: Weekly meetings; Refined design document; Each participant developed a problem-scenario (a document to describe imagined project needs/early visualization)
August 2010 Weeks 10 - 13	Problem Scenario Development	Completed: Weekly meetings; Significant theory development; Development of wireframes (project sketches and functionality specs); Set up shared document repository online
September 2010 Weeks 14 - 15	Iterative Prototyping	Completed: Weekly meetings; Refined theory and wireframes in shared documents; Set up shared code repository; Began coding using Agile Programming methodology (several month long coding "sprints"); Ordered software development kit
Week 16	Iterative Prototyping	Completed: Weekly meeting; Working Prototype; Refined theoretical framework; Updated design document with theory framed in semiotic terms; Posted overview document online: http://erikloyer.com/clients/sinux/http://erikloyer.com/clients/sinux/
Week 17	NEH Presentation	Completed: Developed Powerpoint presentation; Harrell successfully presented work to NEH in Washington D.C.
October 2010 Week 18	Iterative Prototyping and Theory Development	Completed: Weekly meeting; Added gestures to the prototype; Began implementing narrative events; Development of sample narrative
Week 19	Iterative Prototyping and Theory Development	Completed: Weekly meeting; Added gestures to the prototype; Continued implementing narrative events; Implemented sample interactive narrative; Presented work to MIT Computer Science and Artificial Intelligence Laboratory faculty
Week 20	Iterative Prototyping and Theory Development	Completed: Weekly meeting; Added gestures to the prototype; Refined narrative event model; Refined sample interactive narrative
Week 21	Iterative Prototyping and Theory Development	Completed: Discussed project via email; Completed second iteration of the prototype; Refined design document by incorporating theory from psychology literature on gesture
November 2010 Week 22	Prototype and Theory Development	Completed: Weekly meeting; Debugged second iteration of the prototype; Loyer focused on programming, Chow bolstered gesture theory, Harrell focused on narrative model
Week 23	Prototype and Theory Development	Completed: Weekly meeting; Debugged second iteration of the prototype; Loyer focused on programming, Chow bolstered gesture code, Harrell focused on sample narrative
Week 24	Prototype and Theory Development	Completed: Weekly meeting; Debugged second iteration of the prototype; Loyer focused on programming, Chow bolstered gesture code, Harrell focused on sample narrative
Week 25	Prototype and Theory Development	Completed: Weekly meeting; Finalized second iteration of the prototype; Finalized design document
December 2010 – Spring 2011	Paper Writing and Proposal Development	Completed: Assessed results; Harrell traveled to Hong Kong to work with Chow on the design document; Chow and Harrell's article on the project was accepted to the ACM Cognition and Creativity Conference; Chow created art assets; Presented the work at an NEH Workshop at Harvard University; Implemented an addition to the narrative model that allows the story to progress when users do not perform gesture; Planned extension to system to be completed with Loyer; Preparing for future grant applications

Table 1: Detailed Work Plan and Outcomes

The three major activities, in detail, were:

Theory Seminar:
The PI conceived of the project as building upon his previous research in interactive narrative. Each of the collaborators was able to contribute to different aspects of the project. Harrell shared interactive narrative theory, Chow shared theory of interactive animation (conducted while pursuing his Ph.D. with Harrell as his advisor), and Loyer shared his experience as an artist and developer who has developed successful gesture-driven narratives for the iPhone. Building on this set of experiences, the team developed an interdisciplinary theoretical framework building upon the following:

- Theory and Technology for Interactive Narrative (Harrell, 2007)
- Studies of human gesture (Ekman & Friesen, 1972; McNeil, 1992)
- Studies of human-computer gestural input (Wexelblat, 1994)
- Semiotics (Peirce, 1965)
- Study of narrative (Goguen, 2001; Labov, 1972)
- Study of conceptual metaphor (Lakoff & Johnson, 1980; Lakoff & Turner, 1989)

Building on this theoretical framework we developed our own synthesizing framework discussed in the Accomplishments section below.

Scenario-Based Design
We undertook a design process that began with each of the collaborators envisioning an ideal alpha-level prototype as an outcome of the process. Scenario-based design is a unique method in that rather than waiting to produce system sketches until the concepts are already fleshed out, designers produce functionality sketches as a brainstorming device. This process clarifies goals and illuminates potential challenges. In this case, the team used this process to decide upon the requirements for an engine that can enable developers to produce gesture-driven interactive narratives. Since our aim was to create a platform for others to use (as opposed to a one-off narrative artwork) this was an important part of the process. We also developed the core concepts and graphical style for a test narrative that could effectively convey the system's functionality. The design document was updated iteratively in light of the development of our new theoretical model and our alpha-level prototype.

Agile Programming
We undertook a development process called *agile programming,* which is suited to programming applications in small teams with restricted timeframes. This method allowed us to specify major aspects of the system while maintaining an exploratory attitude toward implementing new features.

The work was divided into three milestones:

- Design: Creating wireframes describing the basic user experience that we wanted to allow the test narrative to exhibit was established.
- Development Sprint 1: Implementing the prototype's core functionality to enable authors to represent narrative models and produce sample stories.
- Development Sprint 2: Implementing the prototype's visual presentation system.

Omissions and Changes in Project
The one major change during the completion of the project was the PI's move to a position as Associate Professor at MIT from a position as Assistant Professor at the Georgia Institute of Technology. There were no omissions or other major changes from the project proposal. We were able to complete more than was proposed in one important regard. Beyond just completing the GeNIE platform as an alpha-level prototype, we were also able to use the system to implement a rough demo on the iPhone as a proof of concept. Key personnel remained constant throughout project development. The work was publicized in several ways to be detailed below. One aspect of the work that we would like to explore more is the relationship of gesture to different cultural models for communication (mentioned in more detail under accomplishments below).

Publicity
Aside from presenting the work at the NEH PI's Meeting, the PI presented on this research at an NEH Digital Humanities Workshop held at Harvard University on January 27th, 2011. The PI also presented the work to the faculties of the Computer Science and Artificial Intelligence Laboratory and the Comparative Media Studies Program at MIT. This work will be presented at the ACM (Association for Computing Machinery) Creativity and Cognition Conference to be held November 3-6, 2011. This conference features published double-blind peer-reviewed proceedings (~23% acceptance rate). This work is also featured in various outreach and promotional materials by the Comparative Media Studies Department at MIT.

Accomplishments
As specified in the proposal, the activities supported by the grant resulted in:

- *New Theory*: We created a taxonomy of relationships between gestures performed by users as input to devices and narrative meanings in digital stories.
- *New Technology*: We built an engine for implementing gesture-driven interactive narratives for mobile devices featuring touchscreens and a sample interactive narrative to instantiate and assess our outcomes.

Quantitatively, we were quite productive. We produced a prototype system, a step beyond the early brainstorming stage typically supported by Level I NEH Digital Humanities Start-Up Grants. Furthermore, this system is a platform that is usable by others, not a self-contained demo. Authors need only create visual assets and a narrative model in an XML format that we devised based on the PI's previous work. It is an input format that does not require significant technical expertise, which was an important design consideration for us. Furthermore, we implemented a demo of a quality to effectively demonstrate our ideas.

Qualitatively, we feel that the system is effective for producing gesture-driven, animated interactive narratives. The demo we constructed effectively demonstrates our model for gestural interaction. The demo implements a simple, but narratively complete scenario authored by the PI with input by the team. To fully exhibit the power of the system, a complete narrative demo should be created. This is planned in future work and would be completed along with development of the platform into a version that is even easier for non-expert programmers to use. We would like to relate the demo narrative to cultural models for expression more robustly. This may be accomplished by integrating this effort with Harrell's National Science Foundation CAREER Award project, of which a portion involves developing culturally-nuanced character representations.

Details on the outcomes follow:

New Theory

We use an expansive definition of "gesture" that encompasses a range of non-verbal communication types including hand gestures, posture, facial expression, and other forms of embodied meaning expression. We have noted that when it comes to digital storytelling, there are two meanings of gesture that are likely to be used. These are:

- Input Gesture: Gesture as a user input mechanism on specific device (such as a user clicking and dragging using a touchscreen)
- Storyworld Gesture: Gesture as narrative act/expression *within* a particular media experience (such as a character in a game pointing at another character)

In C.S. Peirce's classic work in semiotics, he describes multiple types of relationships that representations can have to meanings. (Peirce, 1965) One of these types of relationships is termed "indexical." It describes a function between the representation (representamen) and a meaning (object). The indexical relationship describes a function between these, i.e. how they (often indirectly) relate to one another. For example, "smoke" can be an index for "fire" in a sign where the presence of smoke indicates the presence of fire. This is relevant here, because when gestures performed by a user, such as moving a finger up and down (Input Gesture), causes gestures to be performed by a character, such as nodding her head (Storyworld Gesture), then we can say that the two gestures have an indexical relationship to one another.

So, the task of a gesture-driven interactive narrative system is to *implement a set of indexical relationships suited for effective interactive narrative*.

There are many types of indexical relationships. Based on the references in our theoretical framework, some of those that were most useful to us in developing interactive narrative works were:

- **Pantomimic:** user action is echoed as an avatar action
  Example: Swing device to swing racket
- **Iconic**: user action depicts the form of an avatar action
  Example: "<>" motion with fingers to show hands on hips
- **Metonymic** or **Metaphoric:** user action is associated with the same meaning as an avatar action
  Examples: *Metonymic* -- Shake device to show angry face; *Metaphoric* -- Down swipe to show downcast eyes (SAD is DOWN)
- **Manipulative:** user action tightly manipulates an object
  Example: Drag to flip light switch on/off
- **Semaphoric (non-diegetic)**: user action controls something outside of the storyworld
  Example: Double-click to pause

New Technology

The GeNIE system consists of two components. One component structures narrative events and the other component renders them using animated graphical images and text.

Narrative system component:

The narrative event-structuring component allows authors to represent stories based on a model drawn from the work of sociolinguist William Labov (see Figure 2).

Narrative category	Narrative question	Narrative function	Linguistic form
ABSTRACT	What was this about?	Signals that the story is about to begin and draws attention from the listener.	A short summarising statement, provided before the narrative commences.
ORIENTATION	Who or what are involved in the story, and when and where did it take place?	Helps the listener to identify the time, place, persons, activity and situation of the story.	Characterised by past continuous verbs; and Adjuncts (see A3) of time, manner and place.
COMPLICATING ACTION	Then what happened?	The core narrative category providing the 'what happened' element of the story.	Temporally ordered narrative clauses with a verb in the simple past or present
RESOLUTION	What finally happened?	Recapitulates the final key event of a story.	Expressed as the last of the narrative clauses that began the Complicating Action.
EVALUATION	So what?	Functions to make the point of the story clear.	Includes: intensifiers; modal verbs; negatives; repetition; evaluative commentary; embedded speech; comparisons with unrealised events.
CODA	How does it all end?	Signals that a story has ended and brings listener back to the point at which s/he entered the narrative.	Often a generalised statement which is 'timeless' in feel.

Figure 2: Labov's narrative model derived from studies of oral storytellers conveying personal experiences (Labov, 1972)

Authors using the GeNIE system must produce specifications for stories based on Labov's narrative model. Story specifications use the well-known XML format to make it easily usable by non-expert programmers.

Graphical System Component:

The graphical rendering system uses appropriate animated illustrations to express underlying narrative content. In a sample narrative we used these to affect important aspects of storytelling such as emotional tone. For example, the metaphor gesture of pinching in or pinching out can be used to express an introverted or outgoing feeling as in Figure 3.

Figure 3: Pinching in or pinching out affects the character's emotional state.

In our alpha-level prototype we implemented emotional states by representing both what characters *internally felt* and what they *externally displayed* as shown in Figure 4 (illustration by Chow):

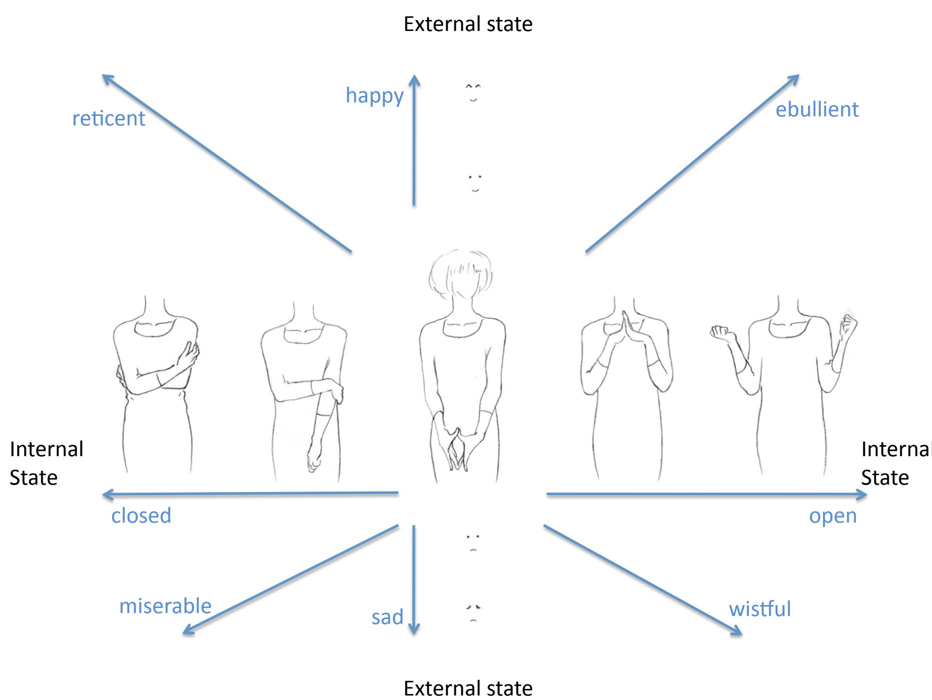


Figure 4: A character's emotional state is displayed via storyworld gestures. The vertical axis shows how the character's facial expression changes to represent externally displayed emotional states. The horizontal axis shows how the character's body language changes to represent internally felt emotional states. Combinations of internal and external states result in emotions such as "wistful" or "ebullient."

Furthermore, the system utilized conventions from cinema in order to shift between gestures of different types that affect storytelling differently (see Figure 5). For example, a close-up shot allows the user to alter their character's facial expressions and reactions in greater detail, while still keeping an eye on the actions and reactions of the other characters. Dialogue balloons appear to indicate character speech.

Figure 5: A design sketch illustrating that touching a hotspot for the first time causes an explanation of the related interaction to appear. Dragging in a "U" shape causes the character to smile (an inverted "U" causes the character to frown).

Similarly, the medium shot allows the user to get a sense of the topography of the space, and allows for manipulation of the player character's overall posture (see Figure 6). Dialogue balloons appear to indicate characters' speech.

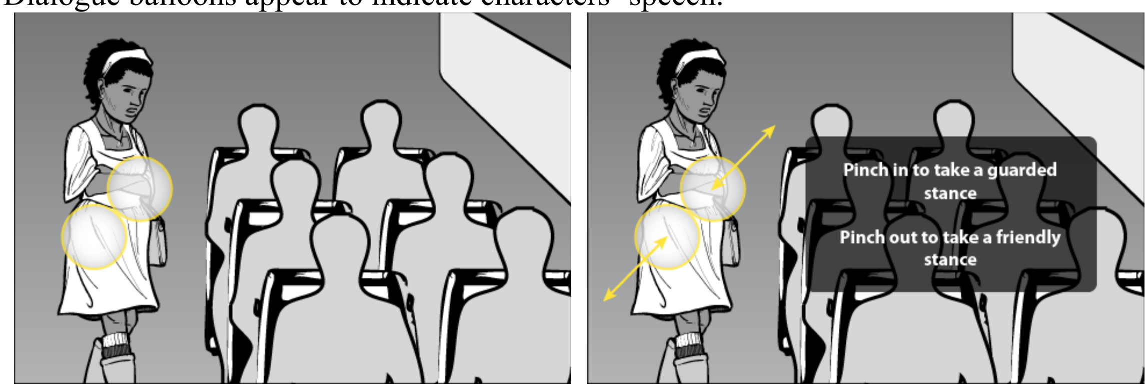


Figure 6: A design sketch illustrating that pinching out causes the character to take a friendly stance, while pinching in causes the character to take a guarded stance.

Culturally Specific Expressive Interfaces

We were also interested in implementing culturally diverse gestural models. The project proposal cited Henry Louis Gate's work on the "speakerly" text in which culturally-specific content associated with oral communication can be conveyed in literary writing. (Gates Jr., 1988) Similarly, we are interested in how culturally-specific gestures can be conveyed via digital media. For example, in some speakerly texts, actions such as eye-rolling, placing one's hands on her hips, or neck snapping have served as markers for a self-possessed "attitude." Gestural walk cycles can convey culturally meaningful differences such as a "cool strut" or "uptight stride." In our sample narrative demo, we implemented a specifically Japanese anime-based gestural model explicitly to exploit the notion of cultural discomfort felt between two characters.

Intuitive Interfaces

At the same time as enabling culturally-specific communication, gestures can also implement relatively universal forms of communication. For example, an aggressive act such as shaking a device can be deployed to serve as a more intuitive interface for many users. As another example, tilting a device from side to side can be used as in intuitive way to switch between two characters.

Audiences

The audience for the GeNIE platform is digital media authors of high school age and older. We are critical of the commercial game world's focus on "hardcore" games that feature play styles typically oriented toward combat and mastery of the games' mechanics. Instead, we are interested in cross-cultural narrative traditions specifically aimed at crossing demographic boundaries such as gender, ethnicity, and cultural background. The PI plans to integrate the system into courses that he teaches at MIT, which reach a broad cross-section of students ranging from computer science to the arts.

Evaluation

Evaluation was carried out informally by the project team members in dialogues with casual test subjects (typically students and colleagues). While early feedback was positive and influenced the description of outcomes above, serious qualitative analysis should be undertaken. In other work supported by the NSF, Harrell has developed evaluation methods for digital media systems (games and interactive narratives) based on well-established techniques from the social sciences, in particular, grounded theory analysis. In further developing this work, a demo could be evaluated using the same method. Anecdotally, users were most impressed by the integration of theory from the humanities with computer science-based interface design strategies. Users found the test demo to be effective in conveying our core aims. Users also were surprised at how intuitive and natural it seemed to become a "puppeteer" for characters and thought it would be a fruitful interaction mechanism for video games aimed at "non-typical" users. Computer scientists found the gestural input taxonomy to be informative for developing interfaces.

A weakness of the system might be that it implements a particular form of interactive narrative. We feel that the insights generalize to both utilitarian systems aimed at practical ends and to more graphically-rich forms of interactive narrative artwork such as in virtual worlds or gaming. However, our demo may bias users to think that the model is only applicable to the particular style of narrative that we constructed. Our theory is quite interdisciplinary, which also constitutes a challenge in conveying it to diverse academic and professional audiences.

A challenge in developing such systems is the necessity to effective demos while maintaining focus on issues central to the digital humanities. We feel that better understanding how digital storytelling systems fit within the broader purview of storytelling in media at large is within the scope of the digital humanities. Our theory helps to better understand digital storytelling systems and to enable development of systems that are more culturally-grounded. Yet, some researchers in computing are purely focused on developing technologies and some researchers in the humanities are purely focused on the critical interpretation of texts. Both of these types of researchers may find that our approach to the digital humanities is quite different than their own. At the same time, we hope that our work can contribute meaningfully to the definition of the digital humanities enterprise.

Continuation of the Project

We would like to continue the project and are considering application for a Digital Humanities Start-Up Level II grant or a Digital Humanities Implementation Grant.

Long Term Impact

As mentioned above, the GeNIE platform will be used pedagogically in Harrell's courses at MIT. The work was extremely well-received by researchers at a Digital Humanities Workshop held at Harvard, which positively impacted the impression of NEH-supported digital humanities work at MIT.

Grant Products

- Prototype: The GeNIE platform (implemented in Objective C for iOS)
- Paper: "Enduring Interaction: An Approach to Analysis and Design of Animated Gestural Interfaces in Creative Computing Systems," *Proceedings of the 8th ACM Conference on Creativity and Cognition*, (with Kenny K. N. Chow), accepted, forthcoming, November, 2011.
- Demo: An interactive narrative called "Rainbus Ecstatic."

References

Ekman, P., & Friesen, W. V. (1972). Hand Movements. *Journal of Communication, 22*, 353-374.

Gates Jr., H. L. (1988). *The Signifying Monkey: A Theory of African-American Literary Criticism*. New York: Oxford.

Goguen, J. (2001). Notes on Narrative, from http://www.cse.ucsd.edu/~goguen/papers/narr.html (accessed 8/31/2011)

Harrell, D. F. (2007). *Theory and Technology for Computational Narrative: An Approach to Generative and Interactive Narrative with Bases in Algebraic Semiotics and Cognitive Linguistics.* Ph.D. Dissertation, University of California, San Diego, La Jolla.

Labov, W. (1972). The Transformation of experience in narrative syntax. *Language in the Inner City* (pp. 354-396). Philadelphia, PA: University of Pennsylvania Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Lakoff, G., & Turner, M. (1989). *More Than Cool Reason -- A Field Guide to Poetic Metaphor*. Chicago, IL: University of Chicago Press.

McNeil, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.

Peirce, C. S. (1965). *Collected Papers of Charles Saunders Peirce*. Cambridge, MA: The Belknap Press of Harvard University Press

Wexelblat, A. (1994). *A feature-based approach to continuous-gesture analysis.* Massachusetts Institute of Technology.

Appendix

Paper attached: "Enduring Interaction: An Approach to Analysis and Design of Animated Gestural Interfaces in Creative Computing Systems," *Proceedings of the 8th ACM Conference on Creativity and Cognition*, (with Kenny K. N. Chow), accepted, forthcoming, November, 2011.

Enduring Interaction: An Approach to Analysis and Design of Animated Gestural Interfaces in Creative Computing Systems

Kenny K. N. Chow
School of Design
The Hong Kong Polytechnic University
sdknchow@polyu.edu.hk

D. Fox Harrell
Comparative Media Studies Program
Computer Science and AI Laboratory
Massachusetts Institute of Technology
fox.harrell@mit.edu

ABSTRACT

This paper provides an interdisciplinary reflection on the nature meaning-making involving users and animated gestural interfaces. In particular, we propose a new model for analysis of creative computing systems incorporating gestural input into dynamically animated interfaces. Our contributions are based on a theoretical framework synthesizing embodied cognition approaches in cognitive science, phenomenology in philosophy, and user interface design in computing. We introduce the term *enduring interaction* to refer to the phenomenon of bodily and conceptually engaging interaction within constantly changing computational environments. Our construct centralizes the issue of how users' motor-sensory experiences inform their construction of meaning in the design of interactive systems. We argue that *creative computing systems*, a class of artifacts including types of hobbyist websites, video games, and computer-based artworks, require a new design perspective quite distinct from user-centric interface design approaches focused on productivity-oriented applications. Using examples including outcomes of the Gestural Narrative and Interactive Expression (GeNIE) project (Harrell, PI; Chow and Erik Loyer collaborators) along with existing prevalent, exceptional, or historically significant artifacts, we articulate a continuum of various kinds of engagement, showing design implications of our perspective, enabling users to use gestural interaction (through multi-touch and gyroscope/accelerometer-based input devices) to result in narratively salient, evocative, and even intimate interaction mechanisms in interactive narrative environments.

Keywords

interface design, embodied cognition, phenomenology, digital media, gestural interfaces

1. INTRODUCTION

"Do you wish to continue reading the paper?" asks a character in an interactive narrative game. You, the player, have four discrete choices to choose from: (1) "yes," (2) "no," (3) "is it a technical paper?" or (4) "is creativity a significant focus of the paper?" Your discrete choice, selected by mouse-click, will result in a discrete response by the system, continuing a conversation. It is a feedback loop involving a turn-based exchange. Now, imagine instead a system in which the same question is asked, but in response you rest your hand on a multi-touch screen. As the character asks questions, you might continuously slide your finger up and down, puppeteering your own character's nodding head to indicate a "yes" response. Or you might use your finger to draw a '^' shape, raising a character's eyebrow to express skepticism, followed by your performance of a pinch-in motion using your fingers to result in your character's mood becoming more withdrawn. The computer-controlled character's facial expression *continuously changes in response to your actions*.

Imagine now, a computer hang-gliding simulator in which you must hold a mobile gaming device with a built in gyroscope ahead of you. Without thinking about what commands to use, you gently tilt the device from the left to the right, signaling shifting of your character's weight rightward, causing the hang-glider to bear into the simulated easterly wind. Seamlessly, continuously, you are *deploying a now habituated movement in response to the computational environment at hand*: the hang-glider has drifted in the direction you intended.

In each of these imagined scenarios, a type of habituated gestural interaction was scaled appropriately to the device in hand. The interface also required response to a continuously changing animated environment. We call this phenomenon, to be defined in Section 3 below, ***enduring interaction***, and argue that accounting for such phenomena is important as recent gestural interface systems grow increasingly robust.

With increased popular deployment of gestural interfaces in systems such as the Nintendo Wii, Microsoft Kinect for Xbox 360, mobile phones with multi-touch screens and built-in gyroscopes and accelerometers, and laptop computers equipped with touchpads, it has become apparent that computational media objects engage users not just in sensory perception, but also in increasing degrees of motor-based bodily interaction. A user whose attention is absorbed in engagement with audio-visual content, for example in a game where one must perform dance-like moves at certain rhythmic intervals, is also driven to take timely action and to react promptly because the artifact invites user participation meanwhile continuing to show a concert-like animated visual presentation. The body of the user is engaged continuously, in terms of attention – like being captivated in a theater performance, and also alternately, in terms of action – like participating in a tennis game.

Furthermore, while there are some accounts focused on usability of such interfaces, there are few accounts of such interfaces for creative computational media that have expressive aims such as hobbyist websites, video games, and computer-based artworks. We turn to an interdisciplinary framework, primarily based on embodied cognition perspectives from cognitive science, phenomenology in philosophy, in addition to user-interface design in computing, to better understand enduring interaction and to suggest new design strategies for systems employing this interaction paradigm, particularly creative computing systems. This paper develops and synthesizes theoretical models from Chow's Ph.D. dissertation work on embodied cognition approaches to animation, Chow and Harrell's work on the same, and recent results of the National Endowment for the Humanities (NEH)-funded Gestural Narrative and Interactive Expression (GeNIE) project (Harrell, Principal Investigator; Chow and Erik Loyer collaborators).

In this paper, we describe processes of expressive meaning-making involving significant motor-sensory interaction. Section 2 provides further motivation for our theoretical framework and grounds our argument in a synthesis of existing theories. In Section 3, we focus on specific ways that bodily motion and temporality are central to these analyses. With emphases on motility and temporality, we explain the idea of *enduring interaction* as a continuum of degrees of interface engagement varying in these two major parameters in Section 4.

2. THEORETICAL FRAMEWORK

When engaging in enduring interaction, the user makes meaning through simultaneous engagement of sensory and motor apparatuses with the system. On the sensory side, the user understands animated visual images at both immediate (perceptual) and metaphorical (conceptual) levels through what we call material-based imagination [1]. On the motor side, as we argue in this paper, the user also makes *conceptual* meaning through performing bodily motor action. A common example is a user running his or her fingers over a touchpad and seeing pages scrolling in a window as a result. That user might be provoked to imagine the scrolling process as reminiscent of panning a camera over the document. The two sides conjoin in a motor-sensory feedback loop, constituting conceptually meaningful interaction between the user and the system. Based upon this loop, users can imaginatively elaborate new meanings from the experience at hand.

Indeed, everyday experiences that provide a basis for imaginative elaboration involve our bodies in motor-sensory connections. For example, a lifestyle application on a tablet computer displaying a digital pond presents fish that do not only swim around, but also respond to user stimuli like pointing or touching – metaphorically suggesting that the fish are alive. This sense that the fish are alive is an imaginative elaboration of the perception of the fish as merely dynamic animated CGI artifacts. Understanding similar motor-sensory loops in familiar human environments where we similarly see artifacts as lively can help in understanding and designing how users engage in creative and expressive systems. The theoretical framework in this section informs our understanding of such systems.

2.1 Meaning between Bodies and Environments: Feedback and Affordances

Various motor-sensory loops exist in interaction between users and systems in different scenarios. When riding a bike, typing on a keyboard, playing computer games with the gamepad, or using other motion-sensing input devices, a user has to engage part or all of the body in "controlling" the machine and "sensing" the feedback. When these actions become conventionalized and second nature, the interaction mechanisms are similar to general cases in which living creatures deploy extensions of the body as control mechanisms. For example, consider a sparrowhawk subtly adjusting its wings to change its trajectory in order to swerve in response to an updraft of wind and nearby building. Our use of interface objects can become so second-nature that they effectively become extensions of our bodies like the birds wingtip feathers.

Quite a few scholars have paid attention to the similarities in interaction mechanisms between engineered systems and physical environments. Norbert Wiener is credited with introducing the term "cybernetics" for the study of control and communication theories in animals and machines [2] (p.19). His awareness of the value of attending to people's ongoing interactions with objects in their environments reminds us that observation of the natural environments and their inhabitants is informative to designing interactive systems because of the forgotten commonalities between two sets of environments. Recall the first time that one attentively uses the mouse to drag a file icon onto a folder icon in a GUI environment. The users keeps moving the mouse while tracking the moving file icon on the display until it reaches the folder. This motor-sensory phenomenon is reminiscent of our physical experience of sliding a paper across a table in an office. We need to "see" something being "moved" from one location to another. In the GUI environment, after being quite attentive during an initial experience, a user no longer notices details of the action after many repetitions. Current users are quite accustomed to operating a wide array of input devices and interpreting sensory feedback from these systems, much like the way our bodies get used to physical environments. Digitally mediated environments absorb the body no less than the physical environment.

In addition, J. J. Gibson's theory of affordances also calls attention to a vital link between bodies and environments. Gibson coined the term "affordance" to mean those environmental features that support what actors can do. Gibson believes that these properties in an environment, whether natural or the artificial (in fact Gibson dismisses this dichotomy in environments as in [3] (p. 130)), are usually directly perceivable, without requiring much learning (p. 143). This idea inspired Donald Norman's asserting that in designing interfaces affordances should be made perceivable and noticeable to users [4] (pp. 22-23). How? As Gibson puts it, "affordances are properties taken with reference to the observer" (p. 143). A property that is "commensurate with the body of the observer" is "more easily picked up." Therefore, Gibson's idea reminds us that accounting for relative scales between bodies and environments should be a more central concern in interaction and interface design. This idea of human-scale embodiment, as we shall show in Section 3.1, applies to not only spatial concerns, but also temporal.

2.2 Meaning in Tools: Dourish's Embodied Interaction

With an awareness of the role of the body in designing interactive systems, Paul Dourish advocates an approach to interaction design grounded in the idea of embodiment. Dourish draws upon notions from phenomenology, particularly Martin Heidegger's "being-in-the-world," to interpret embodiment as people's engagement in the world in order to make meaning of it [5] (p. 126). For Dourish, engagement includes both physical and social interactions. Therefore, to create embodied interactive systems he suggests making use of people's "familiarity" with the mundane everyday world, including practical experiences with physical objects and communication skills in social communities. His book delineates several high-level principles to achieve this goal.

By embodiment, Dourish seems to mean largely the embodiment of tools in the working environment. This bias toward utilitarian situations may originate from Heidegger. As in his well-known hammer analogy describing that a carpenter would only see a hammer available for his tasks (what he calls "ready-to-hand"), Heidegger presupposed a "work-world" in his pronounced dictum [6] (p. 233). It follows that Dourish's proposal also emphasizes the practical and functional meaning of everyday world objects and tools. Although this meaning of "everyday experience" with objects is in line with the major notion of phenomenological embodiment [7] (p. 36), we believe that Dourish's approach can be complimented by approaches that focus to an even greater degree on the body that uses those objects. Our approach focuses less on systems oriented toward productivity (though our ideas apply in those cases too), instead primarily addressing culturally-oriented creative systems. Instead of tools, we also address creative computing systems whose aims are expressive, aesthetic, and/or playful.

2.3 Meaning in Bodies: Merleau-Ponty on Motility

Maurice Merleau-Ponty asserted that the "body is our general medium for having a world" through motor habit [8] (p. 146). Based on Gelb and Goldstein's empirical studies of brain-impaired patients, Merleau-Ponty describes two types of bodily movement, namely "greifen," meaning "to seize," and "zeigen," as "to show" (p. 123). The former refers to concrete movements toward some existing objects, while the latter refers to abstract movements not relevant to any actual situation, but dependent upon one's imagination. For example, an impaired patient may be able to touch her or his forehead because of a mosquito bite, but not able to point to it without preparatory movement. A typical subject can perform a hand-to-forehead salute, but the patient cannot do so without placing himself in an actual situation that calls for it. The patient cannot enter into the situation if forced to only imagine it. This helps to reveal the fact that unimpaired people have turned most abstract movements, like pointing or other gestures, into situated motor habits. Merleau-Ponty believes that it is our body that actually absorbs meaning, in the form of bodily experience (pp. 146-147). In Hubert Dreyfus's words, our body gets involved through practice and acquires bodily skills [9].

Merleau-Ponty posited that what is impaired in the patient is the "power of laying out a past in order to move toward(s) a future." He called this the "intentional arc," which is our disposition, attitude, and *aboutness* toward something (pp. 135-136). The intentional arc exists in both space and time and works beneath the level of conscious conceptualization. Through repeated practice, we enrich our repertoire of actions within the intentional arc [9]. It is the way we acquire bodily skills and build motor habits. Our bodies "absorb" motor knowledge and take care of our everyday motion [9]. Therefore, a typical subject can perform seemingly imaginary abstract movements based on the intentional arc as easily as concrete movements at the immediate cognitive level. This *motility*, as Merleau-Ponty puts it, reveals our consciousness as "not a matter of 'I think that' but of 'I can'" move toward something [8](p. 137). In other words, bodily motion performed in space and time embodies one's intention. This dictum underpins the notion of enduring interaction.

3. ENDURING INTERACTION: MOTIVE INPUT AND ENDURING ENVIRONMENT

The theoretical framework above has been meant to reinforce the importance of attending to issues of embodiment in meaning construction using devices for gestural interfaces. Specific important points that motivate the need for our concept of enduring interaction include that:

- relative scales between our bodies and environments matter not only in physical interaction with the world, but interaction with digital devices,
- meaning is embodied not just in the ways we use tools, but also in how our bodies act upon expressive artifacts like devices for entertainment media,
- our bodies can turn any movements into motor habits that reveal our intentionality toward things.

In this section, we shall further show that meaning is best embodied in movement taking place in an animated computational environment and time frame that closely matches our bodily and perceptual scales.

In creative computational media, this goal can be achieved by mobilizing two sides of the motor-sensory connection: (1) user input contains *motion components that are significant to user intention*; and (2) this user input is in *response to a continuously changing environment*. We call the paradigm of interaction mechanisms consisting of these two mobilized qualities ***enduring interaction*** in contrast to the discrete, conversational style of alternating computer-human interaction, like the classical command-line interface, or the conventional point-and-click mouse-based input mechanisms found in many GUIs.

HCI research has offered an approach called enactive interaction, which is related to ours. Yet, there are also significant differences between the concept of enduring interaction and the concept of enactive interaction. What the terms share is a concept of interface design that centralizes motor and perceptual actions and responses between users and computers. Enactive interface research explores new kinds of interfaces representing a paradigm shift in HCI from symbolic, iconic knowledge representation toward more motor-sensory-oriented forms. [10] The diffence between the terms is that rather than addressing a broad, over-arching agenda, the term enduring interaction has a more narrow focus in an emergent domain – creative computing systems. These types of system often require specific attention to the embodiment of intentionality and temporality.

3.1 Motive Input: Spatiotemporal Embodiment of User Intention

The two types of bodily action mentioned by Merleau-Ponty aid in articulating commonly assumed general divide between mechanisms of user input in HCI, (1) point-and-click (or related types if discrete, command oriented input) versus (2) on-going gestural input. The former corresponds to "grasping," whereas the latter "showing." Our aim is not to detract from conventional point-and-click mechanisms, indeed they can be usefully deployed as shown by various examples provided by Ben Shneiderman's idea of direct manipulation. As Shneiderman puts it, the primary goal of direct manipulation is to bring human-computer activity back to the level of early stages of Jean Piaget's theory of child development, in which children comprehend largely through physical actions. Interfaces based on this principle should be easier for children, not to mention adults, to use [11]. That is to say, direct manipulation requires not much learning and just demands users of the innate capabilities of concrete movements.

On the other hand, the concept of gestural interfaces en vogue in HCI often envisions users performing abstract movements with their empty hands, such as pointing. Early research on gestural input includes Richard Bolt's interface combining finger-pointing and speech-ordering [12], Alan Wexelblat's feature-based approach [13], and others. These research works built upon rigorous reviews of gesture taxonomy and non-verbal human communication.

However, we argue against any hard distinction between the two types of motor input in human-computer interaction. Repeated use of any input mechanisms in digital environments would enrich, to use the terminology of Merleau-Ponty cited above, our "intentional arcs," turning any possible movement into a motor habit. For example, double-clicking a mouse button is a learned (and arbitrary) *gesture* at first, but repeated exercising turns it into a meaningful *habitual action* embodying a user's intention to confirm. Hence, stringently separating gestural input from other more discrete types of motor input might not lead us toward tighter motor-sensory connection and more embodied meaning in interactive systems. Instead, distinctions should be drawn based on whether a given type of motor input contains meaningful (intentional with respect to the computational environment) motion components related to our intentional arc. That is whether the gestural input it a type of situated, embodied, and evocative motion.

Consider the jog dial of a video tape recorder (VTR). When a user spins the dial, the motion, including speed and direction, conveys his or her intention of going forward or backward at variable speed. This kind of motion has already been borrowed as a gesture of hurry in live performance, in which the spinning speed roughly reflects the hurriedness! (Figure 1) Similar motion-based signals can be applied to computer interaction as well, such as circling on the touchpad to browse through a huge image database. Therefore, we call for a specific kind of motor input, called *motive input*, in which motion embodies the intentional arc in space and time with meaning and information. In other words, many types of bodily motion, with repeated use and practice, will be spatiotemporal embodiment of intention.

When one performs a bodily action, body parts such as hands or fingers, move in different directions quickly or slowly, continuously in time, carrying and conveying one's various intentions. We believe interactive systems can be designed for tighter motor-sensory links and potential for more evocative meaning-making by taking these spatiotemporal bodily scales into consideration.

Figure 1: Use of the jog dial of a VTR

3.2 Enduring Environment: Temporal Embodiment of System Behavior

Since Merleau-Ponty's work discussed above, other thinkers have focused on how our bodies interact with worldly situations. Cognitive scientists George Lakoff and and Mark Johnson believe that bodily interaction comprises the basis for much of both everyday and higher level, conceptual thought. They assert that most of our early-acquired cognitive models are based on motor-sensory experiences of the world, called image schemas. For example, we encounter artifacts such as tubs containing water as infants, so we might also project a container metaphor onto abstract concepts, as when we understand a tribe as living "in" the forest [7] (p. 36). The theory of image schemas is a powerful tool for us to understand how we reason based on spatial logic. However, although image schemas are based on motor-sensory actions, which necessarily carry both spatial and temporal dimensions, many examples seem to emphasize spatial relations while downplaying temporality. As is well discussed in the cognitive linguistics literature, because time is an abstract concept, we are used to spatializing time metaphorically. When we say the "flow" of time, we may mean seeing time as a moving object (e.g., a river) from a static observer's viewpoint. In short, we conceive of time in a source-path-goal schema in which we situate our body. It seems that our inclination to reason spatially is a result of the body's spatial existence in the world. However, the body also inhabits time. As Merleau-Ponty puts it, "I belong to space and time, my body combines with them and includes them" [8](p. 140). In this paper we emphasize aspects of the notion of embodiment involving the concept of time.

Merleau-Ponty tends to dismiss the river metaphor of time, saying that time is "not like a river, not a flowing substance" and "it is not the past that pushes the present, nor the present that pushes the future, into being" [8] (p. 411). Merleau-Ponty's doubt regarding the more-or-less ongoing linear ordering of past, present, and future echoes Gilles Deleuze's cinematic philosophy. Deleuze theorizes a perspective of time in his two famous volumes about cinema. On one hand, his work points out that human natural perception introduces "halts" or "fixed points" in our everyday lives [14] (pp. 22-23). We tend to "immobilize" the continuous flow of life, slowing down the intense "flux" of data, in order to think [15] (p. 149). Hence, we see time as connections of those fixed points within an ordered whole. This model is best illustrated by the use of storyboards in filmmaking or in interface design, where they are often called "wireframes." To an interface designer, a wireframe helps understand the connection between input and feedback. In such media, time is just a sequence of alternating stimuli and responses. In other words, humans usually experience time as a discrete order of events based on our natural perception.

On the other hand, Deleuze reminds us "cinematographic perception works continuously," meaning that the camera does not naturally encode these discrete events [13] (pp. 22- 23). His major theses include the notions of the "movement image" and the "time image." The movement image refers to the kind of cinematic images enabled by the moving camera and/or montage, a type of image in which movement is not just about individual changes of an object in space, but is also related to variation or change in the whole [13] (pp. 8-9). Each movement, as Claire Colebrook explains, is "not just a change of place within a whole but a becoming in which the movement is a transformation of the body which moves" [16] (p. 45). For example, a static camera can just capture a fixed camera view of an object changing in the frame, while a moving camera together with shot juxtaposition can show the object's movement actually transforming the whole surrounding holistically. Here, change is individual movement with the whole as background, and transformation refers to the change in state, or "becoming," of the whole environment. The moving camera, together with montage, reveals relative movements simultaneously happening in the whole.

The idea of the "becoming" whole, Deleuze believes, is "to change constantly, or to give rise of something new, in short, to endure" [13] (p. 9). Deleuze uses this common cinematic phenomena to reveal that time is not alternating sections of sensory perception, but constantly changing wholes involving concurrent segments of varying saliency and attention. We add that cinema makes this "becoming" perceptible to humans, that is, scales it appropriately to the human body and human perception.

We believe that cinema, with its hundred years of history, has shaped recent perception of time in creative computational media, making it different from the natural "stimulus-response" mechanism [14] (p. 149). In such media, the ordered perception of time corresponds to the common discrete, conversational style of interaction. For instance, a classical command-line interface introduces "halts," breaking the engagement into ordered, discrete, and alternate segments of input and feedback.

Today's creative computing system can present powerfully evocative meanings by invoking a sense of the "becoming" whole in spatiotemporally embodied interaction. This means that users are engaged in interfaces that highlight perceivable continuity and variation. Hence, we argue, in creative computing systems the idea of embodiment must be reconsidered from the perspective of engagement in a holistically transformative and partially responsive world.

Many expressive interactive creative computing artifacts need to project an illusion of a constantly changing environment, or an *enduring* environment. Such systems simultaneously respond users' motive input while continuing to vary their output.

One exemplar of enduring environments is John Conway's well-known *Game of Life*. The "game" is a two-dimensional cellular automaton in which the cells live, die, and reproduce according to pre-defined rules based on the neighboring cells. The running of the game often looks like an animated work featuring vivid patterns that are constantly changing. Meanwhile, a user is always free to interfere with the growth of cells by putting extra cells in the two-dimensional grid. This motor input is motion-sensitive, because to add cells the user just runs the mouse-cursor, or finger in case of the touchpad or touch screens, across the grid (in some cases clicking is required, other implementations work based on mouse-over).

At the motor-sensory level, the piece is a generative and interactive animation. At the metaphorical level, it projects an image of life, which is constantly changing on the whole because of its rules, and is always differing in various parts because of its openness to intervention. The interaction between the user's motion-based motor action and the animated images of the whole cell habitat exemplify spatiotemporally perceptible embodiment in a "grid" of "becoming." Similar enduring environments make computer interaction more reminiscent of our everyday life experiences. Our notion of enduring interaction becomes a lens for expressing several aspects of why many people find the *Game of Life* to be so compelling to watch and use.

In summary, the necessary conditions of an enduring interaction mechanism include two sides of the motor-sensory feedback loop:

- *Motive Input*: The motion qualities of user input represent user intention, effecting significantly on system output;
- *Enduring Environment*: The system still presents continuous variation on the whole, even without any user input. Meanwhile, motive user input it would show differing output in specific parts in response.

We believe computational artifacts with enduring interaction interfaces hold the potential to make creative computing systems, and even more utilitarian systems, more familiar, intimate, and close to human users in light of our typically situated, enculturated performance of gestures in the world.

4. ILLUSTRATIVE ANALYSES: MECHANISMS OF ENGAGEMENT

Toward creation of human-computer interaction that is more evocative and intimate to human users, we require a new model derived from spatiotemporal embodiment of meaning, including motive input and an enduring environment. We introduce a continuum of mechanisms from the least to the most engaging (see Figure 2):

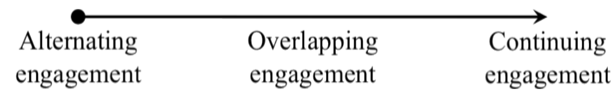


Figure 2: A continuum of Engagement

This continuum provides a model for designing interfaces in non-productivity applications or creative computing systems. To illustrate, we first analyze interaction mechanisms of some existing artifacts, including both mechanical and digital ones. We then discuss how this continuum is manifested in the authors' on-going interactive narrative project GeNIE. The GeNIE project aims to formulate computational models for gesture-based interactive narrative systems and to enable development of multimedia narratives exploring and demonstrating expressive non-verbal digital interaction. The developing computational narrative platform GeNIE is inspired by accounts of how other media represent embodied expressive affordances such as Henry Louis Gates Jr's notion of the "speakerly text" in literary theory [17], William Labov's sociolinguistic narrative model, studies of gestural interaction such as referenced in Subsection 3.1 above, and modes of multi-touch interaction currently implemented on the iOS platform.

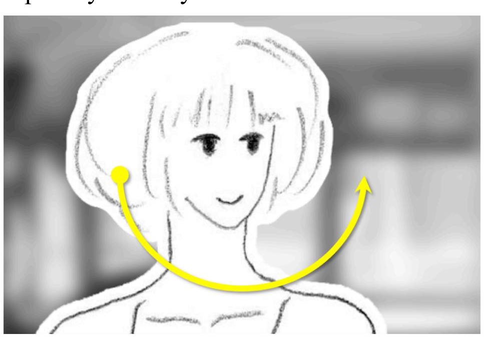
Authors using the GeNIE system must produce formal descriptions of nonlinear narratives in XML format, gestural input functions, and sets of sprite images in order to generate instances of interactive narrative as iPhone or iPad applications. The multi-touch-screen device allows users to perform gestural or motive input to the narrative application while perceiving non-verbal expression (such as smiling, frowning, crossing arms, or extending arms) of the characters in the narrative. The gestural interaction within a GeNIE work constitutes an embodied form of using input intimately tied to the meaningful trajectory of the narrative. This section will refer to the design and implementation of the non-verbal interaction in a sample narrative *Rainbus Ecstatic*, which features a girl's bus journey and poignant dialogue with a stranger augmented with gestures and facial expressions. Below we use this system to exemplify the three types of engagement listed at the start of this section.

4.1 Alternating Engagement

Alternating engagement mainly refers to a conversational style of interaction with artifacts or machines. The mechanism, like verbal dialogue, involves discrete sections of motor input and sensory feedback taking turns alternately. The ordered connection of these sections is the "immobilized" and "fixed" nature of the environment. The use of mechanical typewriters is a typical example. The typewriter responds to each key tap by the typist with a corresponding character strike in sequential order. When a typist performs any one stroke without waiting for the feedback of the preceding tap's completion a jam occurs. In other words, the input and feedback sections cannot overlap. This mechanical constraint leads to strictly alternate sections, resulting in the defining characteristic of alternating engagement. Moreover, alternating engagement involves no motive input. In typing a text, motion qualities like the direction from which the finger hit a key or the speed of reaching a key could have no significant effect on the outlook of the printed text.

The phenomenon of alternating engagement seems to be a result of the mechanics of the apparatus, yet this type of engagement should not be thought of only limited to mechanical artifacts. In fact, examples can be found in digital environments as well, including the command-line environment of MS-DOS or early text-based adventure games, and the point-and-click mechanism in most graphical user interfaces. In these cases, a system always takes a user input, whether it is a tap of the "return" key or a mouse click on a button, and then responds accordingly. The system does not immediately process the next input until it completes the current output. If no input arrives after output, it waits indefinitely. Clearly the alternating pattern is a matter of design conventions, not physics as in the typewriter example.

Alternating engagement can also be seen in gesture-based interaction. In the sample narrative *Rainbus Ecstatic* from the GeNIE project, user input is largely gestural, quite different from conventional mouse click or finger tap input. A user causes the girl protagonist to explicitly express her feeling, represented as "external state" variables, through facial expressions like smile or frown (Figure 3). Drawing a "U" shape results in a smile; an inverted results in a frown. However, the corresponding engagement is alternating because the system must wait for the end of drawing before it recognizes the shape. For shape or icon recognition, the input and output section intrinsically cannot overlap. In the implementation of the GeNIE system, the recognition of input starts after a touch ends, not during a touch moves, determining the type of engagement. To a user of *Rainbus Ecstatic*, it is like giving a command to the protagonist, who then acts accordingly. The interaction does not spatiotemporally embody her or his intention.



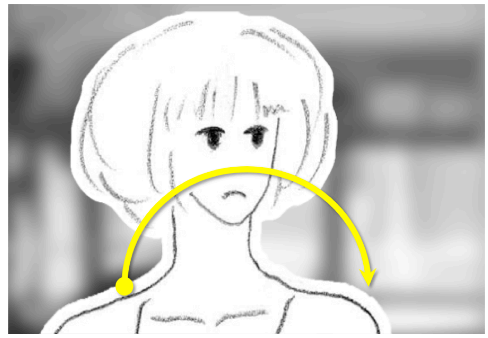


Figure 3: Iconographic gestural input and facial feedback in *Rainbus Ecstatic*: Drawing a "U" shape results in a smile while an inverted one a frown.

4.2 Overlapping Engagement

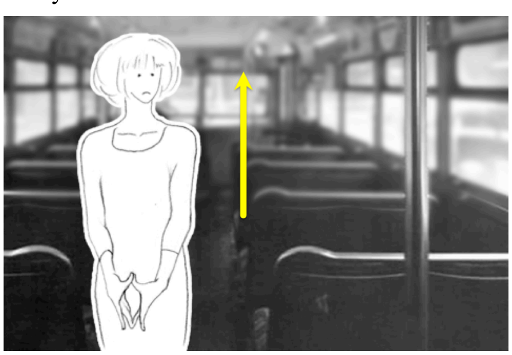
Overlapping engagement differs from alternating in that user input sections seem to overlap with system output sections. An artifact accepts users' motive input, the motion data of which affects output instantaneously. This process results in an illusion of continuous response. A user is simultaneously engaged in performing motor action and perceiving sensory feedback. Some machine interfaces provide good examples. As mentioned, professional VTRs with jog dials allow users to control video playback by rotating the knob. The motion components of the dial affect how the medium is presented. A clockwise spin results in fast-forward, whereas a counter-clockwise spin rewinds the tape. The faster it spins, the faster the tape plays. The case is similar to the mechanics of a zoetrope. The direction of rotating the zoetrope determines the direction of the animation seen through the slits. These machines accept motive input and enable overlapping engagement.

Yet, these machine interfaces present just a partially enduring environment. Although the motor input in the zoetrope or the jog dial is motion-sensitive, a delay of user input could slow down the "becoming" of the output. In case of the zoetrope, the viewer needs to keep spinning the apparatus in order to see the animated effect. The spinning speed determines what the viewer exactly sees through the slits. If the viewer defers the motor action, the animation would slow down and finally halt. Similar to the alternating cases, these interfaces would also wait indefinitely for user input, resulting in an environment not constantly changing and less engaging.

In this regard, many so-called immersive computer interfaces entail similar moderately engaging interaction. These kinds of immersive interfaces, whether 2- or 3D, are ones in which users can navigate by moving the mouse, swiping on the touch screen, or moving fingers on the touchpad. The most classical examples include the interfaces of many first-person shooting games, like *Doom*, in which the player moves the mouse left or right to look around, forward to walk, backward to retreat. Multimedia websites enabled by technologies such as Quicktime VR or Flash often present interactive panoramic views or menus allowing visitors to pan the views or menus left or right with the mouse (e.g., *Out My Window*, an interactive documentary by Katerina Cizek [18]).

Meanwhile, many hand-held devices or tablet computers let users run their fingers on the touch screen to scroll through screens or some large-than-screen canvas. Users of the Mac OS X system are able to scroll a web page or a long document up or down by moving two fingers on the touchpad. Many of the above systems even implement an inertia effect making users feel like physically panning a viewing camera or move a canvas. The direction, speed and even frequency of users' motive input cause immediate visual feedback on the screen. Users are engaged in motor action and sensory perception simultaneously. While the motor-sensory feedback is comparable to that in the jog dial of VTR machines or the zoetrope, these computer interfaces, although evocative, inevitably cease to vary if there is no input.

In *Rainbus Ecstatic*, users can also be engaged in motive input and sensory feedback simultaneously. One can make the girl protagonist express her view by running fingers on the touch screen. Swiping up and down causes the girl to nod, showing agreement to her exchanger. Swiping left and right causes the girl to shake head, meaning dismissal. Furthermore, one can let the girl leak her stance implicitly, being either defiant or submissive, which are represented by "internal state" variables. Pinching out inclines the girl toward an open and receptive posture, while pinching in toward a closed and introverted one (Figure 4a & 4b). Here the gestural interaction in *Rainbus Ecstatic* enables overlapping engagement. The girl is animated in real time to respond to users' motive input. The animation takes place during a touch moves in case of swipe or pinch. When the user swipes fast, the girl nods or shakes vigorously. If the user keeps pinching in and out, the girl changes pose frequently. The user experiences a sense of performing gestures, not only his or her own finger gestures, but also the girl's body gestures, resulting in spatiotemporal embodiment of the user's intention in the virtual story world.



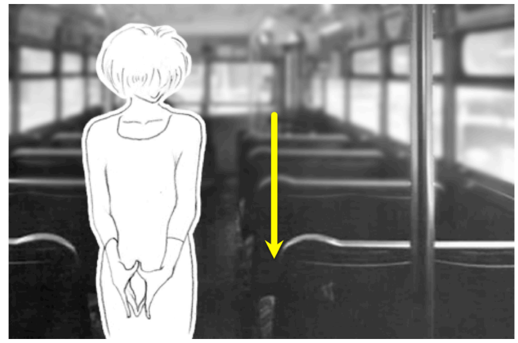


Figure 4a (above): Motive input and gestural feedback in *Rainbus Ecstatic*; swiping up and down makes nodding

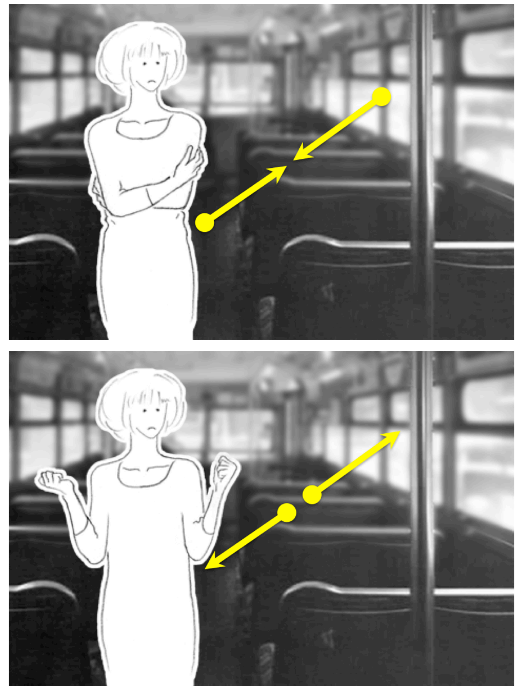


Figure 4b (above): Motive input and gestural feedback in *Rainbus Ecstatic*; pinching in or out makes the girl cross or extend her arms.

4.3 Continuing Engagement

Lastly, continuing engagement describes those systems showing ceaseless transformation, with or without user input. This kind of engagement is continuing in that the whole environment continues to change autonomously at all times. Simultaneously, a user's motive input, at any time might trigger particular variation that would carry on. This "becoming" whole is persistent and divergent. A good analogy is the tea-serving mechanical doll of Edo-period Japan called a karakuri. After winding it up, the automaton paces slowly with a cup of tea to approach its user. When it bows, the gesture cues the user to pick up the teacup. If the user does so, it waits for the return of the cup; otherwise, it turns away, and comes back after a while. Winding it up notwithstanding (which is a type of alternating engagement), the doll is geared to follow its internal rules continuously. The doll might react to its audience's timely motor action or wander around. In short, it behaves differently in different occasions and engages its audience in continuing and differing happenings.

In computational media, examples of this kind of continuing engagement are emerging. One example is the greeting front page *SnowDays* at Popularfront.com [19]. The page displays an outdoor view of snow falling (Figure 5). The downward drifting flakes vary in shape because they are actually other web visitors' individual submissions. A visitor may create a customized snowflake using a simple interactive cutting tool and attaching a message. Once the visitor submits the flake, it is added to a database and now falls in the scene, constituting part of the "becoming" whole. Using a touchpad, a visitor can run a finger across it to "catch" a falling flake and check out the details and the attached message. Yet one has to take timely action, otherwise the target may fall out of the window frame. The mechanism here is a simple example of motive input. This is because direction and speed of finger motion still embody one's intention toward a snowflake. Furthermore, while a flake is held by a visitor, new flakes from others are still continuously added to the database and may enter the scene at any time. At times, a visitor may choose to do nothing but watch as the scene keeps snowing while the background color changes with the time of day. The whole environment is undergoing enduring change no matter regardless of whether the user oversees all flakes or performs close-up actions.

Figure 5: A snapshot of *SnowDays* at Popularfront.com

Another good example is the water-level interface of the Japanese mobile phone N702iS. The interface displays computer-generated images of water that react to user action in real time. When a user tilts the phone, the direction, speed, and frequency of the user's hand motions determine how the water dynamically flows on the screen, yielding an illusion of a water-filled cell phone (Figure 6). Shaking more vigorously leads to other effects like turning off an incoming call. That means the user input is motion-sensitive and embodies the user's intention. Meanwhile, because the water level actually represents the battery level, even when there is no user action, the water level drops very gradually according to the battery consumption. This subtle, but persistent, change inside the virtual container reflects an enduring environment. All in all, the reactive and transformative water image constitutes another example of a "becoming" whole, continuously engaging the user.

Figure 6: The interface of the mobile phone N702iS showing computer-generated imagery of water reactive to user action

In GeNIE, whether to make the story environment continue to change without user input is a matter of aesthetic choice. The narrative *Rainbus Ecstatic* features an uncanny dialogue between the girl and a stranger. After the stranger utters something, the story cannot advance without the girl's response; otherwise it would become just the stranger's monologue. On the other hand, we can implement time-sensitive triggers to create an illusion that the story world still evolves without user input. For example, the time of day may change, the stranger may get off the bus, and another one could get on. The story environment in these cases would change in the absence of dialogue. The goal of such additions would be to implement an enduring environment without contradicting the aesthetic or logic of the intended narrative. Continuum of engagement is not rule of thumb, but rather an important type of phenomenon to consider in the goal of achieving engaging interaction.

In summary, continuing engagement includes both motive input and enduring environments. Overlapping engagement generally involves the former but lacks the latter. If both vital components are missing, an interactive system likely belongs to examples of alternating engagement. A brief look at the examples we raised reveals that the continuum from alternating to continuing aligns with the inclination from tool-based to toy-like artifacts.

5. CONCLUSION AND IMPLICATIONS

In this paper, we have provided a model of enduring interaction that calls attention to important phenomena regarding animated gestural interfaces, with a focus on creative computing applications. Furthermore, in our account a continuum of engagement, we have proposed perspectives useful for designing interfaces for non-productivity-oriented, or creative computing systems. Our approach is mainly grounded in theories of embodied cognition, bolstered by Merleau-Ponty's phenomenological work on motility and Deleuze's cinematic philosophy. The resultant theoretical framework provides a valuable lens for considering animated gestural interfaces since it centralizes two essential components of our everyday lives, namely bodily motion and temporality at a human scale. This lens helps to focus on two issues, motive user input and enduring environment, useful for building embodied, evocative, and engaging interfaces. Through an array of illustrative examples corresponding to a continuum of types of nuanced user engagement with both mechanical and digital devices, we have proposed emerging design orientation in creative computing applications. An orientation that moves away from tools or instruments toward types of interfaces more reminiscent of users' everyday life experiences that involve reactivity, continuity, and divergence.

We believe that the type of engaging interfaces that we propose reflect a phenomenon of "soft anthropomorphism," alternative to the debatable aim of artificial intelligence research to develop agents that fully anthropomorphize the computer by emulating human behavior, rather than *human experience*. By "soft," we mean that idea of enduring interaction represents a step toward making computers holistically "human-familiar," rather than stringently "human-like." This perspective can inform design of creative computing systems that are bodily familiar and cognitively intimate, in other words systems that enable types of engagement close in scale to humans' embodied experiences. Finally, our approach suggests a change of focus from utilitarian issues such as ease-of use regarding types of interaction. Rather, we advocate *an approach centralizing embodied meaning-making* in the study and design of computational media interaction. It is an approach potentially relevant to both everyday work activities and to creative computing systems aimed at enabling expression of, and insight regarding, the nature of the human condition.

REFERENCES

1. Chow, K. K. N. and Harrell, D. F. "Material-Based Imagination: Embodied Cognition in Animated Images," *Proceedings of the 2009 Digital Arts and Culture Conference*, Irvine, 2009.
2. Wiener, N. *Cybernetics : or, control and communication in the animal and the machine*. M.I.T. Press, New York, 1961.
3. Gibson, J. J. *The ecological approach to visual perception*. L. Erlbaum, Hillsdale, N.J., 1986.
4. Norman, D. A. *The Psychology of Everyday Things*. Basic Books Inc., New York, 1988.
5. Dourish, P. *Where the Action Is: The Foundations of Embedded Interaction*. MIT Press, Cambridge, 2001.
6. Moran, D. *Introduction to phenomenology*. Routledge, London, 2000.
7. Lakoff, G. and Johnson, M. *Philosophy in the flesh : the embodied mind and its challenge to Western thought*. Basic Books, New York, 1999.
8. Merleau-Ponty, M. *Phenomenology of perception*. Routledge & Kegan Paul, London, 1962.
9. Dreyfus, H. L., Wrathall, M. A. and ebrary Inc. *A companion to phenomenology and existentialism*. Blackwell Pub., Oxford, 2006.
10. *Enactive network - Enactive Interfaces*. www.enactivenetwork.org/index.php?8/objectives, 2007 (accessed August 7, 2011).
11. Shneiderman, B. *Direct Manipulation: A Step beyond Programming Languages*. MIT Press, Cambridge, 2003 [1983].
12. Bolt, R. A. *"Put-That-There": Voice and Gesture at the Graphics Interface*. ACM Press, New York, 1980.
13. Wexelblat, A. *A feature-based approach to continuous-gesture analysis*. Massachusetts Institute of Technology, 1994.
14. Deleuze, G. *Cinema*. University of Minnesota, Minneapolis, 1986.
15. Colebrook, C. *Understanding Deleuze*. Allen & Unwin, Crows Nest, NSW, 2002.
16. Colebrook, C. *Gilles Deleuze*. Routledge, London ; New York, 2002.
17. Gates Jr., H. L. *The Signifying Monkey: A Theory of African-American Literary Criticism*. Oxford, New York, 1988.
18. Cizek, K. *Out My Window*. National Film Board of Canada/Interactive, 2010.
19. *Snowdays*. snowdays.popularfront.com (accessed August 7, 2011)