Computational Semiotics: Past, Present and Future

Andy Clarke

Kinonet
78 West Kensington Court
London, W14 9AB
andy@kinonet.com

It is now four years since the first COSIGN conference, held at CWI (Centrum voor Wiskunde en Informatica), The Netherlands, from the 10th-12th September 2001. That first event has been followed up by equally successful conferences at the University of Augsburg (Germany) in 2002, the University of Teesside (UK) in 2003 and this one at the University of Split (Croatia) in 2004 - with each one further defining the field of computational semiotics and contributing to its body of knowledge. It therefore seems an appropriate time to take stock, and to both look back over the history of COSIGN and forward to its future.

Since that first conference, we have taken great steps towards establishing computational semiotics as a discrete field, and COSIGN continues to play a leading role in this, providing an important forum for presentation of research in this area, as well as the formal (and informal) exchange of ideas.

So what then is "computational semiotics"? We are still, in many ways, feeling our way towards a definition of this term, but a clear starting point for any definition is the fact that signs and sign systems are central to computing. When we interact with the computer, we can do this only through the manipulation of symbols, which imperfectly represent (a) the real-life object that they stand for; (b) the way in which it is stored and manipulated within the computer; and (c) the way in which it is represented in the code of the application that we are using.

But semiotics is not only relevant at the level of the interface: programming, too, is the manipulation of signs, and often involves finding ways to represent high level concepts and meanings in low level computer languages. Furthermore, there are some problems in computing that explicitly involve semiotics and sign systems. One such problem is the so-called "semantic gap" in visual content representation - the difference between the (relatively poor) level of meaning that can be extracted automatically from an image and stored in the computer and the (relatively high) level of meaning that the user will wish to use in relation to that image (in searches, for example).

Equivalent problems exist in other fields of computing. The creation of interactive narratives, for example, will likewise often require us to come up with systems which are aware of the symbolism of objects and actions (i.e. their connotative meaning) in order to provide the level of sophistication that we have come to expect from stories. There is still, to some degree, a gap between the level at which the computer can store and manipulate the meaning of objects/actions, and the level at which the user wants them to be stored and manipulated, and while there are strategies and techniques for addressing this issue, it nevertheless remains a problem still to be fully resolved.

These examples have barely touched the ways in which the an awareness of semiotics and sign systems can contribute to the study of computers and computer-based media. Even so, they should still give some indication of the scope and importance of computational semiotics, and confirm the relevance of this field.

They should also demonstrate how computational semiotics is not a single, fixed discipline, but a matrix of distinct, yet overlapping, concerns. Two quotes are particularly pertinent here. The first is from our call for submissions, which refers to "the way in which meaning can be created by, encoded in, understood by, or produced through, the computer". We regard computational semiotics as being relevant to all of these fields, and applicable both to the production and to the analysis of computer-based media. The second quote is from the introduction to last years proceedings: "Computational semiotics is not just a set of rules that can be learnt by rote and applied in any circumstance. It is something more elusive than that: a way of thinking rather than a formula".

We believe that the type of problems addressed by COSIGN stretch across the traditional boundaries that exist within computer science, and between computer science and other disciplines. Our intention with the COSIGN conferences is to provide a vibrant forum for the discussion of these issues and to facilitate the crossover of ideas between these separate fields (while still acknowledging the differences that exist between them). This crossover benefits all who attend, and we feel that it validates our decision to make COSIGN a truly interdisciplinary conference. We are grateful that all those who have attended the conferences have also shared this interdisciplinary vision, and it is interesting that a number of our papers - both in this year and in previous years - come from interdisciplinary teams, partnerships, or research groups.

It is clear that the combination of art and science at COSIGN is an important element in providing a stimulating atmosphere in which the exchange of ideas is encouraged. We take care in our selection of artworks, which undergo a peer-review process as strenuous as that faced by the academic papers. An important criteria in our selection is relevance, but our call for submissions, invites the submission of both digital and non-digital art. We believe that restricting ourselves solely to computer-based art would be to place too narrow and simplistic an interpretation on relevance, and thereby lead to a less interesting conference. In previous years, for example, we have exhibited paintings, installations, dance performances, and performance art in addition to the various genres of digital art and net art.

It is important to see, however, that academic research is also a creative act, particularly when (as is often the case in computer science) it involves producing prototype or proof-of-concept systems (or outlining ways in which such systems can or could be built). As a result, the flow of stimulating ideas is not just in one direction. Our intention is that the artists contribute to the conference, but that they also benefit through exposure to ideas, research, technology and practice which is at the cutting-edge of computer science, but which is still relevant to their own interests and practice.

We are particularly satisfied with the exceptionally strong and diverse selection of papers and artworks that we have for COSIGN 2004. Whilst it is difficult to know where to begin this introductory overview of the proceedings, two particular papers - by Timothy Jachna and Alan Peacock - exemplify the diversity of our submissions and the versatility of computational semiotics as a theoretical framework.

In his paper, Timothy Jachna explores the notion of urban semiotics - the idea that the city is a semiotically complex structure and as a result, can be read, even if only to a limited extent, as a "text". Whilst acknowledging both the strengths and limitation of this "reading", Jachna goes on to develop this idea to encompass the notion of cyburban semiotics - the intriguing idea that both this sign system (the city) and the possible reading(s) of it are made deeper and more complex by the fact that the network of the city is now overlaid and intertwined with communication networks (phone, mobile phone, internet, etc.) which in some cases parallel or transcend the structure of the city and in others, undermine or work against them.

But while Jachna uses semiotics at the "macro" level to study the city, Alan Peacock uses it at the other end of the scale, at the "micro" or "nano" level. Here, in his paper "if mouseSign then: Semiosis, Cybernetics and the Aesthetics of the Interactive", Alan Peacock explores the semiotic meaning of one aspect of the computer that is so small and so apparently "natural" that it is all too often overlooked: the relationship between the mouse and the cursor on the screen. In many ways, the user's interaction with the mouse is the fundamental building block to all interaction - as Peacock says in his paper: "Placing the hand on the mouse is a formal declaration of entering into a signing relationship".

The paper by Kristine Jørgensen likewise focuses in detail on one particular aspect of interactive media - in this case, the use of sound. She provides an in-depth analysis of the use of sound in the game *Sacred*, drawing upon notions of perceptual opportunities - as well as upon the work of Frasca (game theory) and Chion (film sound) - to examine the role of sound as a representational system within the videogame. Bernard Perron also covers the use of sound in his exploration of the ways in which videogames - primarily from within the "survival horror" genre - use this and other techniques to foreshadow danger in order to heighten tension.

Both the Peacock and the Jørgensen papers draw upon film semiotics, and this is also an influence on the paper by Grant and Bizzocchi. But while the former two papers use film theory to analyse computer-based media, Grant and Bizzocchi look at a series of projects which have used elements of film theory (particularly those by film semioticians such as Metz) to aid in the creation of systems which edit film clips into meaningful sequences. He then goes on to contrast the theories on which these projects are based with those of the linguistic philosopher J. R. Searle, highlighting both the similarities and the differences between them.

Several of the artworks at COSIGN 2004 deal explicitly with language and communication. The work by Rebecca White, for example, explores the arbitrariness of signs, particularly those which we use in language. As she says in her paper, "Is there any reason why the word bird signifies an animal with wings? A bird could just have easily been named 'tree', 'rain', 'igloo' or 'lamp'".

Using this as her starting point, Rebecca White has produced an artwork in which users invent and define words in a language of her own creation. She has created a set of phomenes and rules which govern the way in which these can be combined to produce words. The user is then able to construct their own "word" according to these rules and add it to the dictionary, including alongside it a "definition" in the form of an monochrome bitmap image. By doing this, the arbitrariness of the connection between the written word, its spoken version, and the definition of the word is made transparent to the user.

The *Universal Whistling Machine* (by Marc Böhlen and J. T. Rinker) likewise provides an interesting take on the communication between man and machine. In this intriguing artwork, one communicates with the machine by whistling a tune to it. *The Universal Whistling Machine* then replies by whistling back its variation on your tune; this, in turn, encourages you to reply with your own variation of its tune, which it again responds to. Attendees at COSIGN 2004 will be able to try the *Universal Whistling Machine* for themselves at the conference.

The communication process is also central to the paper by Marco Schmitt on multi-agent systems. Whilst acknowledging the strengths of the agent-oriented modelling approach that has traditionally been used in this field, Schmitt outlines in his paper the benefits of modelling the communication processes rather than the agents. His paper also describes the COMTE application which uses the principles of communication-oriented modelling to enable the simulation and visualisation of communication processes in multi-agent systems by mapping the flow of message signs.

The paper by Kevin McGee and Johan Hedborg also provides interesting perspectives on agents and our relationship with them, questioning the traditional "master and servant" relationship and proposing a number of other paradigms that may be more appropriate - either for specific applications or in general.

Developments in multi-agent systems and distributed artificial intelligence can lead to benefits in a number of other related fields. The creation of interactive narratives, for example, often also involves the creation of autonomous agents, but places particular demands on these agents as their actions must be co-ordinated so that they form a narrative experience which is coherent, pleasurable, and interesting to the user.

A number of papers offer complementary views on narrativity in virtual environments. The paper by Josphine Anstey, Dave Pape, Orkan Telhan (Department of Media Study, University of

Buffalo) and Stuart C. Shapiro and Trupti Devdas Nayak (Department of Computer Science, University of Buffalo) describes their work in creating the CAVE-based virtual dramas “The Thing Growing” and “The Trial The Trail”. Their paper concentrates on their use of multilayered narrative structures to aid in the creation of psychologically involving stories (“psycho-dramas”).

The paper by Shachindra Nath proposes a more user-centred approach to narrative (in virtual environments and elsewhere) - one that regards narrative as a process, rather than as a structure, and emotion as a core factor in maintaining narrativity, rather than as a by-product of the narrative. An alterative perspective is provided by David Myers in his paper “The Anti-Poetic: Interactivity, Immersion and Other Semiotic Functions of Digital Play”. In it, he expands upon his own previous semiotic analysis of interactive media in the book, *The Nature of Computer Games*, concentrating here upon the principles of early formalist theory and practice - in particular, the concept of “defamiliarization” and its relevance to current interactive media.

Three of the papers at COSIGN 2004 explore the field of artificial life and explore its relationship to computational semiotics. The first, by Theresa Gartland-Jones, provides a outline history of digital art which uses or draws upon theories of artificial life, using this to highlight some of the important issues regarding work in this field. She then goes on to describe the creation and use of an autonomous, evolutionary, drawing system.

The second, by Troy Innocent, deals with his artwork *LifeSigns*. In *LifeSigns*, Innocent has created an “ecology of signs” where “both human and digital agents contribute to the formation of meaning”. Through doing this, he explores the notion of language as an emergent phenomenon - a theme which, in many ways, echoes the way in which man and machine in the *Universal Whistling Machine* artwork find enough common ground to have a “conversation”.

The third paper on artificial life, by Jason Lewis and David Bouchard, also explores the notion of an “ecology of signs”. Here, they describe their artwork *Alien Letter Forms* which takes the form of an environment in which letters (of a text that the user types in) breed, mutate, and die according to the principles of artificial life, their “fitness” being associated with how readable and meaningful they are.

Lewis and Bouchard explain in their paper how they judge the “meaningfulness” of their texts (and consequently their “fitness”), but the broader question of how we handle and assess meaning in computer-based systems is an important one, and central to computational semiotics. The papers by Edward Hartley (University of Lancaster) and by Duc Do and Audrey Tam (RMIT University) both deal with this wider issue, addressing the so-called “semantic gap” in visual content representation. Both papers propose semiotic models to resolve this problem, which remains a critical issue in this field. The paper “ap - fm01” by Martin Howse and Jonathan Kemp explores similar territory, albeit from a radically different perspective.

It is important to remember in our discussions of computational semiotics that we are not usually talking about one single layer of meaning, but in stead about meaning layered upon meaning. The digital artist, for example, works at one level (that of data and code), but the viewer of the artwork only tends to see the result of that activity (what is revealed on screen, or through interaction). Even so, the code and data remain underneath, revealing both the “raw material” of the artwork and the physical and mental processes which acted upon it. As a result, it is sometimes unclear at what level the digital artwork can/should be appreciated.

But all computer systems are multi-levelled in this way - not just digital art. The production of any digital system involves working at several different levels at once - or at least, working at one level while being conscious of the consequences that this has at other planes -, and this presents difficulties. Decisions made at the level of code will have effects at the level of the interface - and ultimately at that of the user’s experience.

This is possibly where computational semiotics has an advantage over other forms of computer-related theory - it is a more multifaceted theory, and therefore allows one to more easily shift from one level of analysis to another while remaining within the same theoretical framework and using the same general terminology. It also provides elements of a common language that makes is easier to communicate across disciplines. The issues that computational semiotics addresses - signs, signification, structures of meaning, etc. - are applicable to all.

This layering of meaning is particularly clear in an artwork such as Christina McPhee’s *Slipstreamkonza* which takes, as it raw material, a stream of climate data measuring the net flow of carbon dioxide on the Konza Prairie of East Kansas (as part of a global project). McPhee takes this raw data and transforms it into image and sound, and her paper explores both the potential strengths of doing this, and the technical and aesthetic problems that she faces.

The various layers of meaning are also apparent in a work such as *(t)Error* by Robert Praxmarer which takes as its raw material semiotically “charged” figures such as George Bush or Osama Bin Laden. Praxmarer regards the strategic subversion of signs as being a key strategy in his art - as he says in his paper, “Fakes, Adbusting and Semiotic Sniping are the ways to express your feelings in a world ruled by global players. The logos and signs of this companies can be found everywhere. Resistance is futile? Not if you undermine their symbolic system, fight with their weapons and use corporate disinformation. That’s what I try to do with my art.” His paper describes the motivations behind *(t)Error* - which are pleased to be able to exhibit at COSIGN 2004 - and places it within the broader context of his other work.

For the first time, we are hosting two separate pre-conference events. The first of these events is a tutorial on computational semiotics provided by Frank Nack (CWI) and Grethe Mitchell (University of East London) - two of the founders of COSIGN. The second is a workshop on recombinant information and its relationship to computational semiotics, which will be provided by Andruid Kerne of Texas A&M University. Andruid Kerne will also be presenting a paper during the main conference on his combinFormation system.

We are also delighted to be able to welcome two distinguished keynote speakers to COSIGN 2004. The first is Peter Bøgh Andersen, Professor at the Department of Information and Media Science, Aalborg University (Denmark), and the author of *A Theory of Computer Semiotics: Semiotic Approaches to Construction and Assessment of Computer Systems*. The second keynote presentation is by Anne Nigten from the V2 Lab in Rotterdam. Her presentation will be on the creative process in interdisciplinary computer arts projects, using the Kurort project by Angelika Oei and René Verouden as an example. A short biography of Peter Bøgh Andersen and Anne Nigten - and a fuller summary of their keynote presentations - is provided elsewhere in the conference proceedings.

Having looked briefly at the history of COSIGN, and outlined the content of the current conference, it is now appropriate to look to the future. Whilst it is difficult to predict future developments in computational semiotics, two developments leave us feeling positive about the outlook of the field. Firstly, during the Summer of 2004, COSIGN became an associate member organisation of the International Association for Semiotic Studies (IASS-AIS). This association was established in 1969 and is the foremost organisation for those studying or working in the field of semiotics and sign systems, having Greimas, Jakobson, Kristeva, Benveniste, Sebeok, and Lotman amongst its founding members. As a result, we feel that membership of the IASS-AIS is a validation of the field of computational semiotics and of the role of COSIGN within this field.

Secondly, we have made the decision to set up a COSIGN Journal. This seems a natural next step in the consolidation of computational semiotics as a distinct field and will allow us to more easily continue the work of COSIGN throughout the course of the year, rather than it being focussed exclusively on the conference in September (though given its success, we will naturally be continuing with the COSIGN conferences). We will be making an announcement regarding the COSIGN Journal at COSIGN 2004, and also on the COSIGN website which remains, as always, the focus for COSIGN-related activity and announcements.

All that remains now is to thank all of the COSIGN presenters and attendees - both past and present - for their contribution to the field of computational semiotics. A conference can only survive and thrive through the strength of its submissions, and one of the most rewarding aspects of organising the COSIGN conferences has been to see how it has, in such a short space of time, produced such a rich, diverse, and interesting body of knowledge.

Psycho-Drama in VR

Josephine Anstey
Dave Pape
Orkan Telhan
Department of Media Study

Stuart C. Shapiro
Trupti Devdas Nayak
Department of Computer Science

University at Buffalo
{jranstey, depape, otelhan, shapiro, td23}@buffalo.edu

ABSTRACT
In this paper, we describe narrative and ludic structures that support interactive drama in immersive virtual reality.

Keywords
Interactive Drama, Virtual Reality, Intelligent Agents

1. INTRODUCTION
The goal of our group is to produce immersive virtual reality (VR) experiences that engage the user as a central protagonist in compelling interactive dramas. Our dramas are designed for CAVEs or CAVE-like systems -- projection-based, 3-D stereo VR displays with one large screen or multiple screens forming a virtual theater. A tracking system and "wand", with joystick and programmable buttons, create the interface between the user and the virtual environment (VE) [12]. Immersive VR puts the user inside the virtual world with the other characters rather than outside, viewing the world on a monitor and manipulating an avatar of herself. In a projection-based system this integrated feeling is heightened as the user automatically uses her own real body to judge the physical proximity, scale and size of any virtual object including computer controlled characters.

We believe that the production of VR drama requires the adaptation and extension of existing dramatic tools, structures and methods; and the appropriation of artificial intelligence techniques for the creation of responsive, believable, intelligent agents that act as characters in the story. This paper describes the narrative and ludic structures that support our VR dramas, (we discuss the agents in as far as they as part of these structures for more details of their architecture and implementation see [7]). In particular we focus on the important role emotion plays in interactive narrative [29], specifically detecting and stimulating the user's psychological/emotional state. The first section briefly touches on the relationship between form and content for interactive fiction. The much longer second section discusses the current state of the dramatic structure we follow. This theoretical structure is evolving in a tight relationship with our practice, specifically the production and exhibition of **The Thing Growing** [2] developed by Anstey and Pape, and the production of our work in progress **The Trial The Trail** [6].

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

2. FORM AND CONTENT
The task of interactive fiction research is not only to develop the necessary computer-based structures and operations to serve as production tools; but also to investigate what kind of stories most benefit from an interactive form, and ultimately what kinds of stories can only be experienced in interactive form. We believe that stories that speak to the construction and struggle for psychic identity are especially good candidates. Although it is argued that full-blown computer-based fiction does not yet exist [11], video games offer a rich site of investigation and insight into computer-based interactive experience. Bob Rehak suggests that a vital driver of video games sits in the unconscious. He retells the history of gaming focusing on how the users' psychological needs shaped the development of the avatar, the representation of self in the game. He suggests that the crucial relationship of computer games is between the user and the avatar, and that ritual destruction of the avatar, and rupture of the identification between the self and the avatar are a major element that create satisfaction in games [22].

Immersive VR is another interactive medium which has been successfully used for work in the psychic arena. Alison McMahan, writing about immersion and presence in virtual environments, notes that VR's responsive and immersive potential make it an effective place of treatment for phobias (fear of flying, arachnophobia, claustrophobia and agoraphobia [20]). In her own work she is creating a VR horror story, using bio-metric devices to test the user's involuntary reactions and bending the story to fit those responses. (In a related argument, Rehak notes that certain film genres, sci-fi, action, horror, are most likely to be re-imagined in video games. Stories in these genres have great metaphorical force to dig deep into emotional terrors and phobic reactions, and speak to the creation and destruction of ego.)

An underlying assumption of our own work is that interactive VR is a unique medium for building stories that access the phantasmagorical, the psychological, the construction of self [3]. Feminist theorist and psychoanalyst Jessica Benjamin suggests that during the process of differentiation from the mother, the child's task is not merely to establish that it is separate but that a step of mutual recognition must occur as the child realizes that the other is also a subject [9]. Anstey and Pape's interactive drama **The Thing Growing** was designed to explore this emotional territory. Unfortunately there is no space here to discuss the differences in terms of identity and agency between video game experiences that play with the relationship between the user and her avatar, and experiences in an immersive medium where that split does not exist.

However, if we do assume that interactive media allows users to explore their own psyche some questions still remain. First: Are people interested in playing with the formation and reformation of their identity? Obviously yes; many critics of new media have analyzed the relationship between cyberspace, and the construction of fluid and multiple personalities [28]. More importantly: How can we structure the user's engagement in psychological dramas or fiction? In the remainder of this paper we describe the process followed by our collaboration in our search for answers to this question.

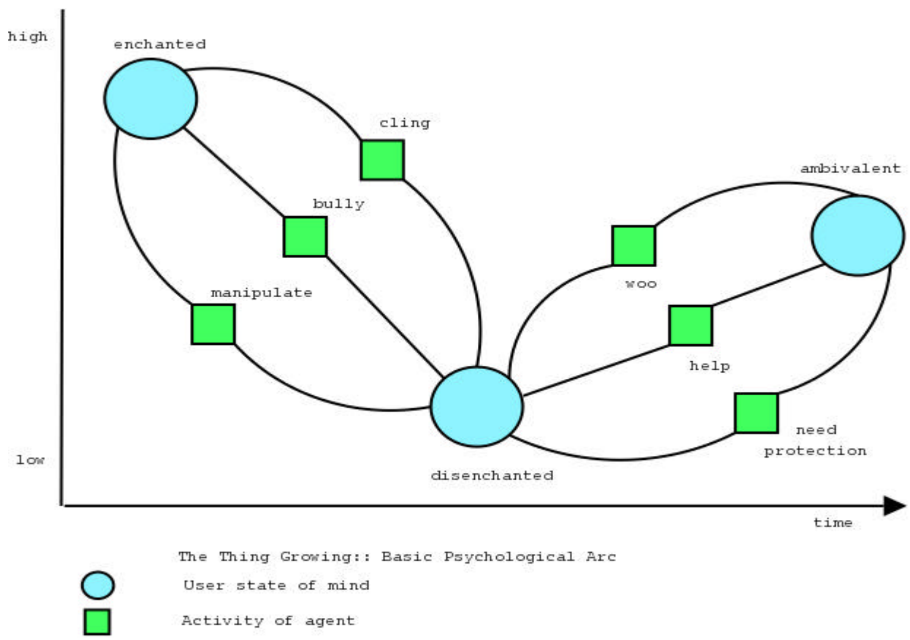


Figure 1: The Thing Growing Psychological Arc

3. STRUCTURING PSYCHO-DRAMA

We build our VR dramas using a two part structure; a psychological substrate where we explicitly determine the emotional states we want to evoke in the user; and an implementation level with three elements - an interactive script, a smart set, and intelligent agents - that turns the psychological plot into a dramatically evolving series of conundrums that the user is faced with. As a play depends on the interrelationship of script, set design, and actors to communicate, so meaning assembles around the user through the co-dependent working of the three responsive elements - the interactive script, smart set, and intelligent agents.

The strategy of using a psychological substrate as the underlying driver for drama has a substantial proponent in Alfred Hitchcock. He explicitly prioritized the psychological, basing his scripts around an emotionally fraught theme. He made famous the concept of the MacGuffin, a plot pretext that supplies a narrative framework of cause, effect and continuing choice to illustrate his characters' psychological development (or unraveling!). [27]. In our case, the implementation level with its three elements make up our “interactive McGuffin.”

Our process of translating a psychological substrate into an interactive McGuffin is influenced by dramatic performances by the Impact Theatre Cooperative in the 80s [10]. In one scenario the cast ganged up on one member insistently mimicking her every word and gesture - the scene generated much tension and dis-ease, and recalled the innocence and savagery with which children act out issues of power, control and identity. The interactive MacGuffin in our dramas tend to contain similar elements that are modeled on childish exaggerations of behavior and are designed to engage at a level beneath that of polite adult intercourse.

3.1 The Psychological Substrate

Our psychological level is an arc representing the ideal emotional route we want the user to travel. Figure 1 shows this arc for our VR drama, **The Thing Growing**. The psychological domain of **The Thing Growing** is differentiating from, yet recognizing the subject-hood of, the other. The plot pretext is a dysfunctional love story which the project simulates between the user and an intelligent agent, the Thing. The Thing is a real-time animated character which speaks to the user. It does not look human, but simulates human-like emotions and gestures. Figure 1 is a simplified version of the arc without details or alternatives. The circular nodes represent the user's state of mind as she reacts to the Thing's activities represented by the squares. The y axis represents the emotional well-being of the user, and the x axis represents time passing. The arc builds a relationship history between the user and the Thing, which corresponds to a love story and moves the user from being enchanted by the Thing, through being distrustful and resentful of it, through some reconciliation. At the end we want the user to feel a certain weight of this shared history, and a certain ambiguity of feeling for the Thing, then she is presented with a choice to kill it - or not.

In **The Thing Growing** we wanted to make the user's feelings as intense as possible so that they felt themselves engaged and present in a relationship of sorts. Therefore the psychological arc, and the MacGuffin that implemented it, were designed to plunge the user into unfamiliar territory, to contain abrupt reversals of feeling, high and low points, expectations and disappointments, and NOT to give her time to analyze her feelings until she had pulled the trigger (or not). We made some assumptions about the user's probable emotional reactions to the Thing. We assumed that if it simulated emotion, the user would react to it emotionally. We assumed that many users would react to power-playing patterns that are common in relationships, and often fall into action-reaction pairs, or sequences of action-reaction pairs. Figure 2 catalogs a few of these pairs, the more co-dependent readers will be able to supply many more.

Agent	**User**
I love you	**I love you**
I cling to you	**I flee**
I abandon you	**I cling to you**
I cajole	**I comply**
I demand	**I deny**

Figure 2: Action Reaction Pairs

For example, when we wanted to move the user from a state of being enchanted by the Thing to being disenchanted, first the Thing announced it was in love with the user, flattered her, and showered her with compliments, then it started being demanding and clinging. We observed that the users' smiles sagged as the Thing changed. We do not pretend that users became as invested in the Thing as it pretended to be in them, but they were evidently moved by its manipulative machinations [5].

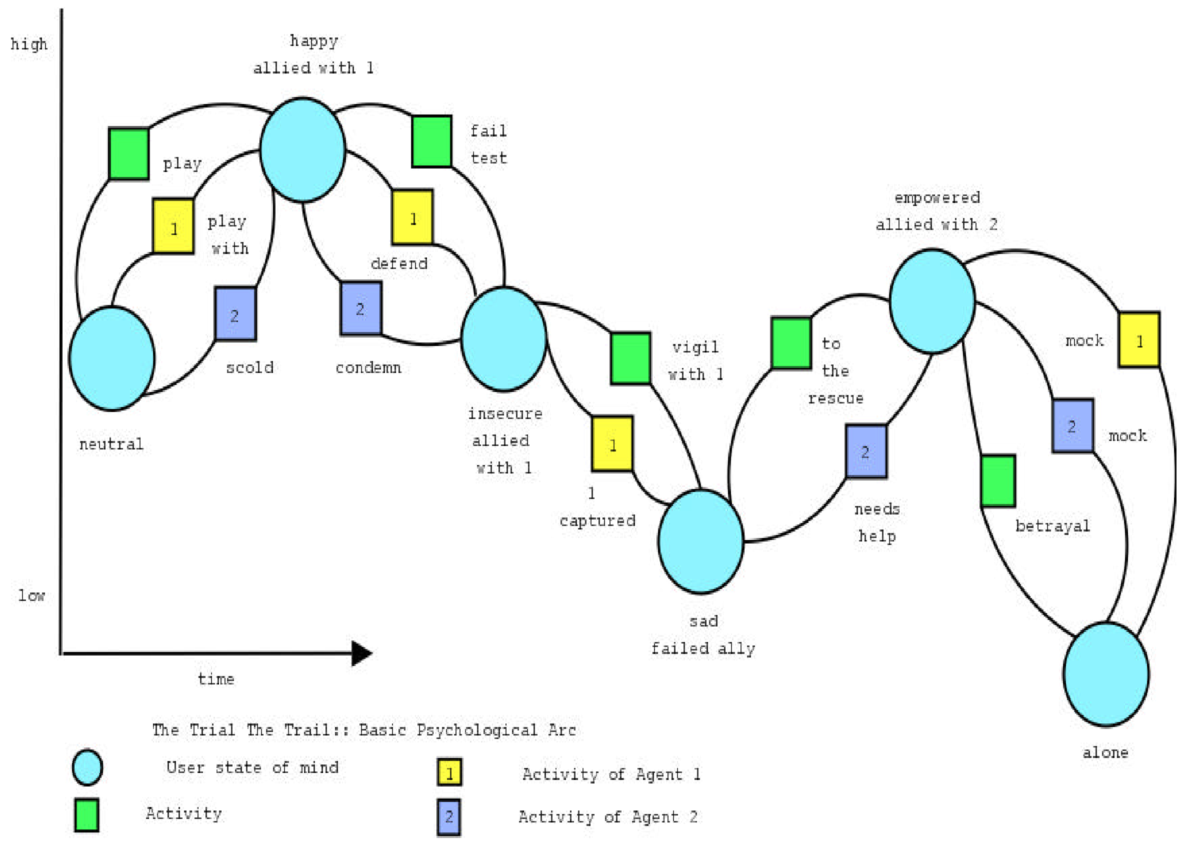


Figure 3: The Trial The Trail Psychological Arc

Our immersive VR display system adds to the psychological charge and effectiveness of our work. In the real world we are affected psychologically by the physical presence of others. If someone is too close it can be intimidating or annoying. These same feelings register in immersive VR. Since we are tracking the user's head and hands, it is also possible for the VR agents to mimic or reflect back gestures that the user makes. Patterns of behavior are conveyed in body language and in mutual body language. That our system can simulate some aspects of this strengthens the psychological level of our work.

Part of the psychological arc of our work in progress **The Trial The Trail** is shown in Figure 3. The psychological domain of this work is the handling of uncertainty and the nature of trust, with respect to other people and to life itself. **The Trial The Trail** has three main characters including the user. The introduction of a third major character allows us to investigate behavior triggered by triangular relationships, much of which involves two characters ganging up against one, changing allegiances, betrayals. The psychological arc for **The Trial The Trail** is more complicated, circular nodes represent the user's state of mind and her alliances with the agents in the drama, empty boxes indicate general activity that is occurring, numbered boxes represent activity of particular agents. The basic structure of moving the user from emotional state to emotional state is the same, however the structure as a whole is made more complex by the separate simulation and stimulation activities of the two agents. The agents play good cop and bad cop, agent 1 sides with the user, while agent 2 gives her a hard time. A reversal is effected when agent 1 is captured and agent 2 needs the user's help to rescue her. A second reversal reveals that both agents are setting the user up and making fun of her. A complication not represented in this diagram, but which will be discussed in the section on the interactive script, is that we want to structure breathing spaces into this story, where the user has time to reflect on her emotional progress.

3.2 The Interactive MacGuffin

The interactive MacGuffin framework supports the “evolution” of the user's state of mind, following the main psychological arc and building alternative responses for the user who deviates. Part of this work is done by a regular MacGuffin; the storyline that provides plot pretexts and narrative rationales and divides the drama into the acts and scenes that move the plot forward as intensely as possible. But we also need a structure that gathers information about the user and folds that back into the evolving narrative so that it becomes responsive. This structure explicitly attempts to move the user from one emotional state to another along the psychological arc. The structure must have a context that will evoke the first state and some form of stimulation to move the user to the second state. We must also be able to test whether she has reached the second state. Our interactive drama consists of a related and unrolling series of these emotional tests which we will call “dramatic snares” or simply “snares”.

Figure 4 shows the parts of the snare. Following the storyline we have chosen, we use the virtual environment (the smart set and the intelligent agents) to build a narrative context. The narrative context provides a set-up that quickly puts the user into a recognizable situation and contains implicit or explicit suggestions/instructions for some activity. The user acts, and the system detects the actions. The narrative context in conjunction with the detection creates implications for a particular action, such as how, whether or not, the user is performing the suggested activity. We use these implications to interpret the user's state of mind, and this information is fed back to influence the narrative that follows. It is important to note that snares can be of varying lengths, they can be assembled into sequences, and they can be nested. In a sense they work very like the acts, scenes, sequences and beats that are typically used to construct drama in plays and films [19] - although as we will demonstrate later they may not always exactly coincide with the dramatic, act and scene, structure of the storyline. The snare structure is similar to Mateas and Stern's discussion of beats as the architectural unit used by their drama manager [18]. Our emphasis is more exclusively on the details of how the user's emotional state may be stimulated and detected. Our work in progress **The Trial The Trail** can be understood as one extended snare, containing within it smaller snares. Each snare is implemented by the interactive script, the smart set, and the intelligent agents. In the following sections we use **The Trial The Trail** to illustrate the snare structure, and then to discuss how these three elements work in the construction of both the snare and the storyline.

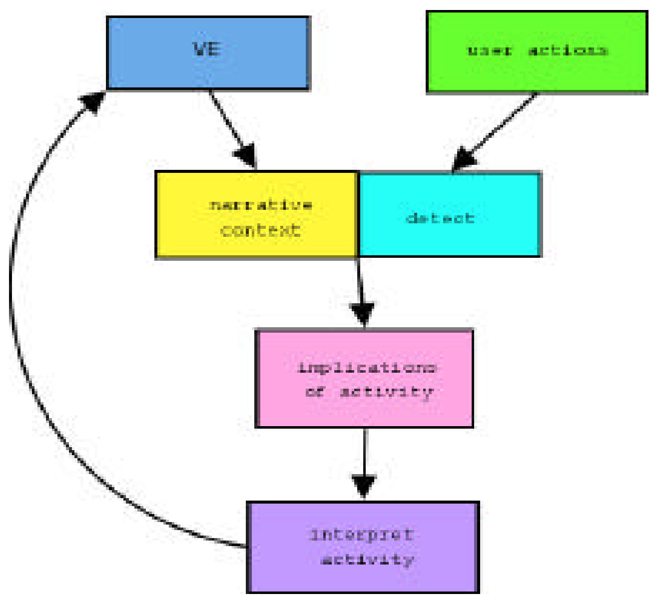


Figure 4: The Snare

3.2.1 The Snare

The Trial The Trail is a quest narrative, the user is given two companions, the actor-agents Filopat and Patofil and told that at the end of her experience she will get her heart's desire. Our first illustration is of a snare that that is the equivalent of a short scene. The narrative context of this snare is a task that the user is set by her companions. They bring the user to a reed-bed where cat-like creatures are playing, and tell her to collect the crowns they are wearing. They show her she must creep up on a creature, crooning softly, then stroke it gently as she takes the crown. This is the activity. Our tracking system allows us to detect whether the user is near the creatures, and the speed and direction of her hand. Implications can be drawn from the tracking information about whether the user is stroking or hitting the creatures. Within the scene this information is used directly to program the cat creatures' responses – stroking leads them to surrender their crowns, hitting makes them dodge away. We can also imply from the user's position in the world whether she is trying to engage in the activity at all. There are several resulting interpretations about the user's state of mind that we can make. First, does the user obey the agents and try to carry out the activity? We interpret this as compliance. Second how does she treat the cat creatures, is she gentle or aggressive with them? Interpreting the user as compliant or disobedient, gentle or aggressive can be used in the set up of future snares.

Our second illustration shows more clearly how a snare can stimulate specific changes in the user's state of mind. This is a three part snare sequence. Part one is the snare described above, where the user is taught how to accomplish the crown gathering activity by her companions, and left to do it. We detect whether she is successful or unsuccessful, gentle or rough. Part one becomes the narrative context for part two where one of the cat creatures suddenly exhibits surprising behavior, clinging to the user and weeping if she tries to get away. We detect how the user treats this clinging creature - does she beat it, does she stay with it? This result is used in the third and final part of the snare where the companions comment on the user's actions. At the start of this sequence the user should be fairly happy. In part one she is moved to a sense of accomplishment and superiority over the dumb creatures who she is essentially tricking into giving up their crowns. This is followed by the reversal in part two where we aim to elicit feelings of guilt, discomfort, annoyance. Then in part three, the user overhears the companions judging her actions. If, for example, she pulled abruptly away from the clinging and wailing creature, Filopat will condemn her cruelty; if she is unable to get away from it, he will laugh at her wimpiness. Patofil will defend her. To sum up, the psychological movement of this snare is from happy to 1. empowerment and superiority, to 2. discomfort and a sense of failure, to 3. feeling judged and defensive. This snare, also fits into the overall emotional scheme of the drama and is the point where the user feels a strengthening alliance to Patofil (agent1 in figure 3).

Figure 5: Storyboard image of the stage in The Trial The Trail with Patofil, Filopat and the user as green figure

3.2.2 The Interactive Script

The interactive script is a blueprint for the dramatic arc (and its alternatives) that implements the psychological arc. It contains the storylines for the entire interactive drama and for the individual snares, linked sequence of snares, and nested snares that comprise the drama. For example, the overall storyline of **The Trial The Trail**, is a warped quest narrative owing inspiration to Tarkovsky's Stalker, Alice Through the Looking Glass, Monty Python and the Holy Grail, Don Quixote. The stages in the quest are linked sequences of snares such as the one described above. The storyline for the first snare in that sequence is the challenge of gathering crowns from cat creatures. The interactive script must be designed to conceal the starkness of the snares that the user is being urged into, give reason for the constraints that the snare is composed of, and motivate the user [16].

It is also the work of the interactive script to assemble the necessary snares into dramatic sequences with schematics for alternative ways of maneuvering the user through them. This means detailing the possible implications of the user's activities given the narrative context; making interpretations about the user's state of mind; and weaving them back into the psychological trajectory. We assume a great deal of iterative user testing so that the final drama will anticipate what most users will do and have a response for them. So it is very much the case that the interactive script evolves during production.

We are still working on the best way of representing the interactive script, but at present it exists as a storyboard divided into acts and scenes which sets out the ideal path through a dramatic arc; and as detailed scene descriptions. Generally speaking the snare structure follows the division of acts and scenes. Specific scenes are nested inside acts, and both scenes and acts would be implemented as snares. However, it is also possible for snares to start in one act and end in the next.

We mentioned above that we want to structure breathing spaces into **The Trial The Trail** to give the user time to consider the implications of her actions in the environment. We now discuss these breathing spaces as they show how the requirements of the psychological level affect the storyline and dramatic structure of the interactive MacGuffin. In this discussion we also demonstrate a snare that passes beyond an act boundary.

Our first, rough, implementation and user-testing of **The Trial The Trai**l, showed that it was too literally a quest adventure; it did not stimulate the psychological questions (handling uncertainty, placing trust) we were interested in [1]. We needed a more ironic and questioning take on quests, goals versus journeys, the real desirability of attaining one's heart's desire, to come through. The tests led us to restructure the storyline of the interactive MacGuffin to contain an explicitly theatrical metaphor which acts as a transition between stages of the quest, and creates a series of entre-actes.

In the VE this translates into two major elements, a curtained stage where the companions explain each part of the quest, and, as the curtain rises and stage flats fly out of sight, an endless vista of the fantasy landscape where the quest itself takes place. The stage, and the banter of the companions when they are on the stage, refer to the over-genial charm of music halls, English pantomimes, game shows (in this game show, the user gets her heart’s desire.) The artifice and theatricality of the endeavor are forefronted, in order to answer the cynic's presumption that she will not get her heart's desire. By contrast, when the curtains are raised the user moves into a fairy-tale landscape in which wishes may come true. This space displaces the stage, yet it hints at interior space. Here Freud and Dali meet Alice in Wonderland. The user is offered the chance to move between immersion in the task at hand, giving full rein to the feelings that brings up, and observation of herself.

In the snare sequence in our example described above the first two parts of the snare take place in the fantasy landscape as one of the challenges in the quest. The challenge is ended by the curtained stage reassembling around the user, signaling the end of the act and the beginning of the entre-acte. The third part of the snare takes place in the entre-acte as the two agents, hidden behind the curtain, discuss the user.

3.2.3 *The Smart Set*

The smart set consists of the visual mise-en-scene, various responsive elements in the virtual environment, and the mechanics that carry the narrative forward. The connotations of the visuals, the effectiveness of the responsive elements, and the deftness of the control mechanisms are important in implementing the snare structure. Our smart set is built using the VR authoring framework Ygdrasil [21] which is based on the OpenGL Performer scene-graph, and provides a framework for extension; application-specific modules (plug-ins) may be added to define behaviors for objects or characters. A text file system is used to translate the interactive script into the actualized virtual environment (the smart set and the intelligent agents): all the models, sounds, their locations and behaviors are described in the text file along with messages to be passed between objects in response to events.

The smart set visually establishes the narrative context for each snare. As in any set design, this does not only mean providing visual elements, but providing elements that metaphorically strengthen the ambiance of a particular moment. In the first illustration of a snare described above the action takes place in a bed of reeds. The reeds are responsive, they flatten as the user, agents, or cat-creatures move through them. They connote games of hide and seek, outdoor fun, an activity the user will want to take part in. Responsive elements of the smart set can also be used in the snare steps that stimulate the user to act, and detect how the user has acted. For example, the cat creatures combine autonomous behavior with

3.2.4 The Actor Agents

The basics of the snare structure came out of our work on the Thing agent, in **The Thing Growing** [4]. In that piece the Thing established a narrative context that included a suggested activity for the user, monitored the user's response with respect to that activity, and fashioned its own response based on the implications of the user's detected actions. Even within that production the snare structure started to overlap the agent proper. Especially when other agents or smart set elements were added it was clear that the agent implemented a larger dramatic structure rather than being identical with it. Splitting the working of the snare from the working of the agent was an important step in our thinking, however below we shall see that there are strong connections between the two. Another step was to abandon the very ad hoc construction of the Thing, in favor of a systematic agent architecture which could produce more flexible and generic actor-agents.

The agents we are creating for **The Trial The Trail** follow the GLAIR [8, 14, 15, 23] agent architecture. Their higher mental functions are built using the SNePS knowledge representation and reasoning system [24, 25, 126], and their embodiment and the virtual world they inhabit are built using Ygdrasil, SNePs and Ygdrasil are connected via sockets. The cat creatures described above might also be called agents, but we do not anticipate them needing the reasoning power of the SNePs mind, so they are implemented entirely in Ygdrasil. Previous GLAIR cognitive agents all had the following capabilities: natural-language input, recognizing a fragment of English, based on a grammar written by the research group; a repertoire of primitive acts and sensory abilities; a knowledge-base of background information, domain knowledge, and acting plans; reasoning to answer natural-language questions, to decide what to do and when; and natural language output.

Our agents have goals and plans based on the interactive script which forms their knowledge base, Their plans often correspond to the stages of the snare. First they help establish the narrative context, second they work with the user on the activity and detect the user's actions. Third, they reason about the user's actions and respond with the next part of the plan or divert into a contingency plan. They have primitive and composite acts; speeches, animations, and movements about the world and with respect to the user. They have senses, detecting the user's presence relative to themselves and other relevant objects, detecting the user's actions and that of other objects in the world. They also have a self-perception that lets them know what they have just said. Currently they do not output natural language or take voice input from the user. Previous GLAIR agents used text input for natural language input and output, however that would not work in our immersive VR set-up. Since current voice-generating software does not do a good job of rendering emotional qualities of the voice, we pre-record phrases for each character creating a library of lines large and flexible enough so there is a response for every eventuality, and which includes redundant phrases so the character is never stuck repeating the same thing. Our research plan does include making use of voice input from the user, however, we are only at the very beginning of that research and the narrative structures described in this paper do not assume voice input.

Our project **The Thing Growing** confirmed for us that human users will be stimulated emotionally by agents simulating emotions. This is vitally important since our endeavor is predicated on moving the user from emotional state to emotional state. Part of an agent's work in helping set up a narrative context is to set an emotional tone which the user may empathize with or react against. This can be done directly by having the agent simulate anger, hysteria, despair. But the user's reaction to the agent's emotion will be colored by her perception of the agent's personality. Thus the user is likely to be worried by the tears of an agent she likes, but may laugh at the tears of an agent she dislikes or distrusts. She may like an agent who has done her a good service, demonstrated affection for her, or holds similar believes.

By the time we get to the point in the drama described in our snare example, Patofil and Filopat have established personalities. These are important elements in the interactive MacGuffin as they take up positions relevant to our psychological terrain and reveal their positions in their attitude to the quest and its challenges. So Patofil is reckless and insouciant, believes the journey is more important that the arrival, and is dubious whether the heart's desire exists. Filopat follows rules, adheres to duty and fervently believes in the quest. Patofil stimulates the user to disobey and to be a little cruel. Filopat provokes defiance to authority, yet also urges humanity and caring. The user is encouraged to side with one, then the other. These alliances implicitly include an adherence to the particularly philosophical position of that agent.

The narrative then moves on to embroil the user in possible implications of taking that position. In the last part of the snare sequence Patofil defends the user against Filopat's criticisms and in so doing questions the importance of the quest. In the next act Filopat punishes Patofil by making her stand all-night vigil with the user. These actions are designed to increase the user's feelings of warmth towards and alliance with Patofil. As the act unfolds Patofil's disobedient and careless attitude leads them both, but especially Patofil, into danger. Patofil's peril leads the user into an alliance with the righteous Filopat. Finally it is revealed that the danger is not real, but fabricated by the two agents in cahoots. This revelation is designed to lead the user to distrust both Patofil,, Filopat and their philosophies.

Figure 7: Patofil standing all night vigil

4. CONCLUSION

To reiterate we structure our interactive dramas with two levels; a psychological substrate and an interactive MacGuffin. The psychological substrate explicitly details the emotional journey we want the user to take; the interior terrain that we want her to visit, experience and perhaps analyze. The interactive MacGuffin details the plot pretexts, storyline and dramatic structure that will implement the psychological substrate. As importantly it contains a basic structure, designed to do the work of moving the user from one emotional state to another which we are calling a snare.

The snare can be the length of an entire drama, an act, a scene, or a small part of a scene. The snare as we have described it has a narrative context and user acts. The narrative context that sets up one snare can be drawn from the previous snare or the larger context of the drama. The user acts are detected and their implications in light of the context are interpreted and used to further the drama. The snare is implemented by an interactive script, a smart set and intelligent agents. The snare structure is designed to invisibly constrain the user, and to be very responsive to the user's actions at a localized level. However, at the highest level of granularity the overall plot has a limited number of alternatives. This is for two reasons; one to avoid exponential growth of story lines; two to tightly focus the user on the psychological terrain we are interested in. Clearly these dramas are not designed to be experienced multiple times, however we do want to accommodate a wide range of users, and to involve them in a dramatic experience that involves their own psyche, and that has a good balance of responding to their input and surprising them. To that end we iteratively test the dramas adding responsiveness at all levels of granularity.

Our hypotheses about narrative and ludic structures for interactive VR drama are constantly tested against our practice. Our work in progress, **The Trial The Trail** is a proving ground for ideas stimulated by the production and exhibition of **The Thing Growing**. During 2003 we started work on building actor-agents that married the SNePs AI system and Ygdrasil authoring tool. At the same time we implemented a rough version of **The Trail The Trail** in VR to test our initial story-line. This was a wizard of Oz version where a human took the role of the agents. This test led to a radical restructuring of the interactive script, and radical redesign of the VE. We are currently implementing act three of our current story-board with GLAIR agents.

5. ACKNOWLEDGMENTS

This research is supported by an IRCAF grant from the University at Buffalo.

6. REFERENCES

[1] Anstey, J. The Trial the Trail: Progress. http://www.ccr.buffalo.edu/anstey/VDRAMA/TRAIL/progress.html. ACM SIG PROCEEDINGS template. http://www.acm.org/sigs/pubs/proceed/template.html.

[2] Anstey, J. Writing a story in virtual reality. In Hocks M. E. and Kendrick K., (eds.), Eloquent Images, MIT Press, Cambridge, Massachusetts, 2003. 283-304. Mackay, W.E. Ethics, lies and videotape... in Proceedings of CHI '95 (Denver CO, May 1995), ACM Press, 138-145.

[3] Anstey, J. and Pape, D. Being there: Interactive fiction in virtual reality. In Proceedings of Consciusness Reframed 3. CAiiA, (Newport Wales, Oct 2000)

[4] Anstey, J. and Pape, D. The Provoking Thing: A VR Relationship. In Emotional and Intelligent II: The Tangled Knot of Social Cognition: Papers from the 2001 AAAI Fall Symposium, Technical Report FS-01-02, AAAI Press. 2001, 3-8

[5] Anstey, J. and Pape, D. Scripting the Interactor: An Approach to VR Drama. In Proceedings of The Fourth Conference on Creativity and Cognition, (Loughborough UK, Oct 2002) ACM Press, 150-158

[6] Anstey, J. and Pape, D. The Trial the Trail: Building a VR drama. In Proceedings of Technologies for Interactive Digital Storytelling and Entertainment Conference, (Darmstadt Germany, April 2003)) Fraunhofer IRB Verlag, 2003, 394-402.

[7] Anstey, J., Pape, D., Shapiro, S.C. and Rao, V. Virtual drama with intelligent agents. In Proceedings of the Ninth International Conference on Virtual Systems and MultiMedia (place and date 2003). International Society on Virtual Systems and MultiMedia, 521-528

[8] Bandera, C., Shapiro S.C., and Hexmoor, H. Foveal machine vision for robots using agent based gaze control. Final Technical Report #613-9160001, Amherst Systems Inc., Buffalo, NY, September 1994.

[9] Benjamin, J. The Bonds of Love: Psychoanalysis, Feminism and the Problem of Domination. Pantheon Books, New York, NY, 1988.

[10] Stan's Cafe. Who were impact theatre co-operative? http://www.stanscafe.co.uk/qimpact.html.

[11] Crawford, C. Interactive storytelling. In Wolf, M.J.P. and Perron, B.. (eds), The Video Game Theory Reader, Routledge, New York, New York, 2003, 259-273.

[12] Cruz-Neira, C., Sandin, D. and DeFanti, T. Surround-screen projection-based virtual reality: The design and implementation of the CAVE,. In Proceedings of SIGGRAPH 93 Computer Graphics Conference, ACM Press / ACM SIGGRAPH, 1993, 135-142.

[13] Glassner, A. and Wong, C. Fiction 2000: Technology, Tradition, and the Essence of Story, In SIGGRAPH 99 Conference Abstracts and Applications, ACM Press / ACM SIGGRAPH, 1999. 161

[14] Hexmoor, H., Lammens, J. and Shapiro, S.C.. Embodiment in GLAIR: a grounded layered architecture with integrated reasoning for autonomous agents. In Proceedings of The Sixth Florida AI Research Symposium, (The Florida AI Research Society, April 1993) 325-329

[15] Hexmoor H. and Shapiro, S.C.. Integrating skill and knowledge in expert agents. In Feltovich, P. J. , Ford, K. M. and Hoffman R. R. (eds), Expertise in Context, AAAI Press/MIT Press, Cambridge, MA, 1997, 383ñ404.

[16] Laurel. B. Computers as Theater. Addison-Wesley, 1993.

[17] Lehmann, F. (ed) Semantic Networks in Artificial Intelligence. Pergamon Press, Oxford, 1992

[18] Mateas, M. and Stern, A. Towards integrating plot and character for interactive drama. In The Human in the Loop Symposium, AAAI Fall Symposium Series, (Menlo Park, California, 2000) AAAI Press.

[19] Mckee. R. Story: Substance, Structure, Style, and the Principles of Screenwriting. Harper Collins, New York, New York, 1997.

[20] McMahan. A. Immersion, engagement and presence. In Wolf, M. J. P. and Perron, B. (eds), The Video Game Theory Reader, Routledge, New York, New York, 2003, 67-86.

[21] Pape, D. and Sandin, D. Alive on the grid. In Proceedings of 6th World Multiconference on Systemics, Cybernetics and Informatics(SCI 2002), 2002.

[22] Rehak, B. Playing at being: Psychoanalysis and the avatar. In Wolf, M. J. P. and Perron, B. (eds), The Video Game Theory Reader, Routledge, New York, New York, 2003, 67-86.

[23] Shapiro, S. C. and Ismail, H. O. Anchoring in a grounded layered architecture with integrated reasoning. Robotics and Autonomous Systems, 43:2-3 (May 2003) 97-108

[24] Shapiro S.C. and Rapaport, W.J.. SNePS considered as a fully intensional propositional semantic network. In Cercone, N. and McCalla, G. (eds), The Knowledge Frontier, Springer- Verlag, New York, 1987, 263-315

[25] Shapiro S.C. and Rapaport, W. J. The SNePS family. Computers & Mathematics with Applications, 23:2-5 (January-March 1992) 243-275].

[26] Shapiro S.C. and The SNePS Implementation Group. SNePS 2.6 User's Manual. Department of Computer Science and Engineering, University at Buffalo, The State University of New York, Buffalo, NY, 2002. Available as www.cse.buffalo.edu/sneps/Manuals/manual26.ps.

[27] Spoto, D. The Dark Side of Genius, The Life of Alfred Hitchcock. Little, Brown and Company, Boston, MA, 1983.

[28] Stone, A.S. The War of Desire and Technology at the Close of the Mechanical Age. MIT Press, Cambridge, Massachusetts, 1996.

[29] Szilas, N. Idtension: a narrative engine for interactive drama. In Proceedings of Technologies for Interactive Digital Storytelling and Entertainment Conference (Darmstadt, April 2003) Fraunhofer IRB Verlag, 187-203.

Context, Convention and Complexity in Film Meaning

Douglas Grant
Faculty of Applied Science
Simon Fraser University
Vancouver, BC, Canada
1.604.731.5457
rdgrant@sfu.ca

Jim Bizzocchi
Faculty of Applied Science
Simon Fraser University
Vancouver, BC, Canada
1.604.268.7437
jimbiz@sfu.ca

ABSTRACT
As Lindley and Srinavasan note regarding film, "Particular cinematic devices can be used to create different connotations or subtextual meanings while dealing with similar diegetic material" [5]. In language, different 'illocutionary forces' create different meaning from the same 'propositional content.' However, linguistic philosopher J.R. Searle thinks this points not to infinite elasticity of meaning, but rather to five kinds of speech acts that encompass meaning production. Each of these analyses, the cinematically-oriented and the language-focused, holds lessons for a computational semiotics that seeks to richly reflect reality (and imagination) *and* be usefully human- and machine-manipulable; context and convention are pivotal to both. This paper examines the productive comparisons and contrasts that the disjunctions and intersections of these approaches afford.

Keywords
genre, illocution, semantic, semiotic, syntax

1. INTRODUCTION
The first part of this paper examines a series of projects during the late 1990s that researched modeling of video semantics and generation of video productions from component clips in keeping with that modeling. The projects sought to embrace the diversity of the cultural dimension of film and the multiplicity of meaning that results from it within a framework that would be usable for non-trivial computational creation of films. In other words, that would be pragmatic and systematic. Our examination looks at how the framework dealt with the complexity that such an undertaking inevitably engenders.

The second and final part of the paper examines a linguistic analysis of meaning production, comparing its approach to complexity and its treatment of context and convention (in relation to meaning) with the framework covered in the first part of the paper.

2. REFLECTING REALITY
Throughout the late 1990s, Craig Lindley and a variety of colleagues, including Uma Srinavasan and Anne-Marie Vercoustre, *inter alia*, carried out a practically-oriented analysis of film semiotics [2] [3] [5] 6] [7] [8]. The object and result of this work for Lindley was the implementation and testing of a research prototype for semantics-driven video synthesis, called FRAMES [4].

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

In these projects, modeling of video semantics was based on multiple 'interpretation paradigms' consisting of five broad levels or types: *perceptual*, *diegetic*, *connotative*, *subtext*, and *cinematic* [3]. An interpretation paradigm is a "set of fundamental assumptions about the world, together with a set of beliefs following from those assumptions (analogous to Kuhn's (1972) concept of a scientific paradigm)" [2]. The paradigms were derived from corresponding levels of codification outlined by film theorist Christian Metz in his book, *Film Language: A Semiotics of the Cinema* [2].

The *perceptual* level involves inherent features of moving image such as line, shape, color, texture, and movement [2]. The *diegetic* level involves the spatiotemporal aspects of everything narratively denoted in a video, including fictional time and space and other elements such as agents (e.g., characters), objects (visual and auditory), actions, and events, all within a particular cultural context [2].

The *connotative* level involves all the 'symbolisms' and metaphorical, analogical, and associative meanings that attach themselves to events and objects (or to relationships between objects) outside of films, i.e., in culture [2]. The *subtext* level involves hidden and suppressed meanings of symbols and signifiers, framed by one of the "great narrative structures that obtain outside of films"; these framings include Nietzschean, Marxian, Jungian, Freudian, and feminist ways of looking at the world, among others [2].

The *cinematic* level accounts for the use of formal film/video techniques to produce particular artistic/formal/expressive results [2]. Lindley, employing Metz' analysis, expounds that this constitutes "the set of…cinematographic systems that, in a specific type of discourse, organise the diverse elements furnished…by the four preceding [levels]" [2]. An important reflection of this for Lindley's position (and for this paper) is that "particular cinematic devices can be used to create different connotations or subtextual meanings while dealing with similar diegetic material" [5].

For film makers and viewers this is an interesting and useful fact, but it also points the way to what seems almost an embarrassment of richness. If part of the power of syntax is that it enables flexibility of meaning for various 'combinatorial' circumstances, the other part of its power is in how it constrains that flexibility within tolerable limits. The problem is that film is less systematic than natural language. Lindley agrees with Metz that a unified syntax for film is impossible because of related differences between natural language systems and film.

> A language system is a system of signs used for intercommunication. Cinema is not a language system since it lacks important characteristics of the linguistic fact. In particular cinematic language is only partly a system, and it uses few true signs…. The preferred form of signs is arbitrary, conventional, and codified, not a characteristic of an image, since an image is not the indication of something else, but the pseudopresence (or close analogy) of the thing it contains. Hence there is a film language, but not a film language system. As Metz states, "Cinematic image is primarily speech—a rich message with a poor code, or a rich text with a poor system. It is all assertion." The meanings imparted by film are primarily those imparted by its subject. Montage demonstrates the existence of a "logic of implication", thanks to which the image becomes language, and which is inseparable from the film's narrativity. The understanding of a film precedes the conventionalization of specific "syntactic" devices. The plot and subject make syntactic conventions understandable. [2]

But Lindley maintains that while these 'realities' prevent a universal syntax they do not preclude a *comprehensive* one [2]. Very specific syntactic and semantic conventions within film genres and styles—the more stylized, the more identifiable—can constitute genre- and style-specific syntaxes, which in turn can constitute a comprehensive one:

> Syntactic conventions that arise within filmic productions interact with the systems of meaning that they are used to codify. These manifestations are recognised in the varieties of type, genre, and style of filmic productions. That is, a partial syntax may emerge within a particular genre executed in a particular style. A comprehensive "syntax of film" must include the full range of such partial syntaxes, and that range and the styles within it are always changing and evolving. [2]

What all this richness means practically is that in order to reduce resulting complexity to manageable levels, a strategy must be adopted of integrating a computer-enabled support environment with human authoring of semiotic data structures [8]; this is what the FRAMES prototype for semantics-driven video synthesis set out to do. Lindley [4] reports that the system handles semantics and syntax for categorical, associational, and abstract film forms well, but that rhetorical and narrative forms require additional models for rhetorical and causal relationships because these "additional 'systems' of meaning…involve specific ways of organising meanings at [the five interpretation paradigm] levels" [5].

FRAMES provides elegantly for a certain level of semantics and syntax such as: identifying and controlling the variable values of semantic model entity (objects and their attributes) and relationship types; incorporating propositional operators of the syntax such as AND, OR, and NOT; and computing termination conditions for a video sequence [6]. It is still challenged, however, by the goal of creating coherent narrative, tending to require restrictive levels of specification in efforts to do so [4]. Lindley and Vercoustre [7] describe research to overcome such problems, including articulation of conceptual and logical models of various types of videos. They state that complex types such as narrative and rhetoric "will require conditional branching within specifications…and extensive rule sets specifying principles of composition for videos in various genres," as well as "detailed development of causal and rhetorical representations [and] techniques for incorporating causal and rhetorical information into the matching process" [4]. This echoes a need for further research back down the whole semantic modeling chain described; for example, in discussing the connotative interpretation paradigm, Lindley and Srinavasan [5] don't specify how connotative annotations will be expressed, but state that "Ongoing work will provide generic ontologies and lexicons at [this and the other] levels." Though the FRAMES-related projects survey rich potential for annotating aspects of cinematic meaning, they don't formulate a rigorous toolset for doing so.

But the devil is not just in the details of specification; higher-order requirements generate something of an impasse between broad objectives of the overall project. Allowing for the reality that cultural and formal context 'multiplies' meaning makes it a significant challenge to systematize the resulting complexity so that it is both readily computable *and* comprehendible by human users of such an embodied system. In short, the FRAMES approach is hard-pressed to reconcile the contextual with the systematic.

3. CONTROLLING COMPLEXITY

John R. Searle differs from Metz and Lindley that a context-independent code is a necessary condition for a language to be systematic, arguing that although natural language is not the former, it can still be the latter. This difference points up a similarity between their positions, as well as what appears, initially, to be a further difference. For Searle [9], language meaning is no more rigidly rule-based than film meaning is for Lindley; but neither is it infinitely elastic. If the reality of human experience is that we construct such things as economic systems, marital arrangements, and political parties within limited categories, then "why should language be more taxonomically recalcitrant than any other aspect of human social life?" This stance would appear to hold the prospect of reconciling contextual richness with systematic rigor, of being neither straitjacketed by arbitrary, rigidly-codified signage, nor subject to unbridled complexity.

Searle's analysis [9], elaborated over decades of thinking in this area, yields five general ways of producing meaning, which he calls categories of illocutionary acts: "We tell people how things are (Assertives), we try to get them to do things (Directives), we commit ourselves to doing things (Commissives), we express our feelings and attitudes (Expressives), and we bring about changes in the world through our utterances (Declarations)." Given this paper's primary concern with film rather than language as such, it's necessary to deal straight off with how Searle's primarily linguistic bent maps and/or doesn't to film. Searle is a philosopher perhaps known best for his work regarding 'speech acts' [9], which are illocutionary acts carried out by using language. However, his linguistic analysis is embedded in a broader philosophy of relations between minds and the world, and it is certainly not the case that all illocutionary acts are speech acts. Searle [9] gives as an example the act of classifying, which can be accomplished in some cases by the physical act of sorting, rather than by a verbal or written speech act. Lindley [2] provides the cinematic analogy for this in the way shots employ angles, scale, and relative placement of objects to suggest interpretation of relations between them. While all illocutionary acts are not speech acts, all 'utterances' are (at least potentially) illocutionary acts, and film production consists of cinematic utterance. The conditions for such utterances being illocutionary acts in the most problematical case, fiction film, are covered later in this paper.

It's also necessary to deal straight off with the remarkable conciseness of Searle's illocutionary taxonomy. In language, verbs are vehicles for illocution; of course given the rich reality of language, there are many more illocutionary verbs for creating meaning than the five generalized categories of results they produce. This would be unproblematic as long as there was simply a many-to-one mapping between verbs and any particular illocutionary act, but it turns out that more than a few verbs can map to more than a single act. Thus, language is not perfectly coincident with Searle's illocutionary categories, and the elegance of his system begins to erode here.

A variety of other things also complicate Searle's analysis. Not only do verbs not fit neatly into illocutionary categories, but according to Searle [9] we often do more than one of asserting, directing, committing, expressing, and declaring in the same utterance. As well, meaning can be articulated in a variety of ways that add complexity to the initially simple picture outlined by Searle; interestingly, some of these are nothing if not cinematic in nature. One example is the perlocutionary act—perlocutionary acts manifest as effects (such as convincing, annoying, amusing, or frightening) that go beyond 'understood meaning' and "may or may not be achieved by specifically linguistic means" [10]. Another is how, not unlike the way different cinematic devices create different meaning while dealing with similar diegetic material, different illocutionary 'forces' create different meaning from the same 'propositional content.' For example, the sentence 'You're standing on my foot!' always has the same propositional content (i.e., that you are, in fact, standing on my foot), but under different circumstances the illocutionary act involved might be to merely assert that, or to direct you to get off my foot, or to express my 'stance' toward you standing on my foot (or, indeed, as pointed out at the beginning of this paragraph, some combination of those).

While none of these phenomena breach the boundaries of the analysis as Searle conducts it (they are in fact part and parcel of the overall construct he posits), they inevitably push those boundaries in a variety of directions in order to encompass the resulting complexity. However, even with these elaborations, Searle's concept of a small number of illocutionary categories is unable to capture the real complexity of the world (and the imaginary) without referring to and grounding itself in something larger. For Searle, this is the capital-B 'Background,' and what it amounts to in many ways is an acknowledgment that context is a prime driver in producing meaning, and must be adequately accounted for.

While roughly equivalent to contextual aspects of Lindley's analysis such as some of the interpretation paradigms, the 'Background' deserves brief explication here under Searle's own rubric in order to profitably compare it with Lindley's view. In order to explain the Background, Searle must first explain "Intentionality":

> My subjective states relate me to the rest of the world, and the general name of that relationship is "intentionality." These subjective states include beliefs and desires, intentions and perceptions, as well as loves and hates, fears and hopes. "Intentionality," to repeat, is the general form for all the various forms by which the mind can be directed at, or be about, or of, objects and states of affairs in the world. [11]

[Note that Intentionality is not *merely* the *intention to do*; this fact is indicated by the inclusion of small-i intention in the list of other subjective (capital-I Intentional[1]) states.]

Searle [11] further explains that Intentional states do not exist and function in isolation, but within a complex "Network" of other intentional states, such as beliefs, and that in addition to this Network of multiple beliefs and other Intentional states, one has "to have a set of capacities and presuppositions that enable me to cope with the world. It is this set of capacities, abilities, tendencies, habits, dispositions, taken-for-granted presuppositions, and 'know-how' generally" that is the Background—in other words, the context in which we operate to produce meaning.

Just as for Lindley the diegetic, connotative, and subtext interpretation paradigms are conditioned by culture, the Background also has a large cultural component. Searle [11] says "Part of the Background is common to all cultures.…Such universal phenomena I call the 'deep Background,' but many other Background presuppositions vary from culture to culture.…I call such features of the Background 'local cultural practices.'" Obviously, if the meaning natural language and cinema produce is grounded in local context, contextual richness is served but systematic rigor will suffer; where it seemed hopeful that Searle's approach might 'contain' the contextual, context once more threatens to overwhelm the systematic by the complexity it engenders.

So, Searle's attempt at being systematic doesn't so much reconcile with context as be assimilated by it, and he ends up pretty much on par in this with Lindley's approach. On close analysis, the two approaches are also illuminatingly similar in how they see the role that convention (and especially genre conventions) plays in producing meaning. In order to see this, it's best to look at how Seale couches fiction in terms of his body of theory.

While the subject of a complete chapter by him in his book *Meaning and Expression*, fiction is not central to Searle's thinking about language, persons, and the world. It is an exception, but an important one that Searle treats as essential to demonstrate the robustness of his theories, and thus the chapter-length study of it. Searle is interested in fiction precisely because while his chief analytical concern is how rules correlate words or sentences to the real world [9], fiction is an elaborate pretence. Sophisticated takes on this pretence argue that fiction authors allow fiction readers to be complicit in it by helping them to willingly suspend disbelief; Searle agrees, but is careful in his use of the suspension of disbelief phraseology. His main interest is that just as there are 'procedures' for successfully describing reality, there are other ones for creating unreality in a compelling fashion. They involve "engaging in a…pseudoperformance which constitutes pretending to recount to us a series of events.…the author of a work of fiction pretends to perform a series of illocutionary acts, normally of the assertive type" [9]. Searle elaborates that in addition to assertions, assertive illocutions may consist of statements, descriptions, characterizations,

[1] Also note that though in most of his works Searle capitalizes 'Intentionality,' as we do here, to distinguish it from small-i intention, he incongruously does not do so in the work that the preceding quote is taken from.

identifications, and explanations, *inter alia*—all devices of the fiction author.

It is clear from Searle's argument that his analysis of fiction, perhaps unsurprisingly, dovetails with narrative form; the chief example of fiction he examines is a narrative form, and his analysis is clearly situated in that domain. This becomes even more clear as he moves on to explain how conventions underpin production of meaning in fiction, with it becoming apparent that the conventions are in large part those that characterize narrative in general and various genres of narrative in particular. Searle [9] describes *non*fiction procedures, the "rules correlating words (or sentences) to the world," as "vertical rules that establish connections between language and reality." From there he moves to suggest that fiction is built on

> a set of horizontal conventions that break the connections established by the vertical rules. They suspend the normal requirements established by these rules. Such horizontal conventions are not meaning rules; they are not part of the speaker's semantic competence. Accordingly, they do not alter or change the meanings of any of the words or other elements of the language. What they do rather is enable the speaker to use words with their literal meanings without undertaking the commitments that are normally required by those meanings….the pretended illocutions which constitute a work of fiction are made possible by the existence of a set of conventions which suspend the normal operation of the rules relating illocutionary acts and the world. In this sense, to use Wittgenstein's jargon, telling stories really is a separate language game; to be played it requires a separate set of conventions, though these conventions are not meaning rules; and the language game is not on all fours with illocutionary language games, but is parasitic on them. [9]

The degree to which the normal rules are suspended, and readers' disbelief along with them, depends on conventions native to particular narrative genres.

> In part, certain fictional genres are defined by the nonfictional commitments involved in the work of fiction. The difference, say, between naturalistic novels, fairy stories, works of science fiction, and surrealistic stories is in part defined by the extent of the author's commitment to represent actual facts, either specific facts about places like London and Dublin and Russia or general facts about what it is possible for people to do and what the world is like. For example, if Billy Pilgrim makes a trip to the invisible planet Tralfamadore in a microsecond, we can accept that because it is consistent with the science fiction element of *Slaughterhouse Five*, but if we find a text where Sherlock Holmes does the same thing,… [9]

In effect, authors both assert and declare ('assertive declarations' are a special category of illocutionary acts that Searle grafts on to his original five [9]) a world or universe, but, importantly, do so within genre and style conventions (note that in this, Searle echoes the claim by Metz, and Lindley, that cinema is 'all assertion'). The correspondence of this created world is most importantly with the genre, rather than the real world. Most importantly—but not exclusively—because created worlds are necessarily partial worlds. Most importantly, still, because while a partial world must rely on Searle's Background to sustain it, the Background is 'fictionally' conditioned by pertinent genre conventions. This allows created worlds to not be comprehensive in and of themselves, but to nonetheless be coherent, with that coherence depending on the conditioning of the Background by genre.

> The author will establish with the reader a set of understandings about how far the horizontal conventions of fiction break the vertical connections of serious speech. To the extent that the author is consistent with the conventions he has invoked or (in the case of revolutionary forms of literature) the conventions he has established, he will remain within the conventions. As far as the *possibility* of the ontology is concerned, anything goes: the author can create any character or event he likes. As far as the *acceptability* of the ontology is concerned, coherence is a crucial consideration. However, there is no universal criterion for coherence: what counts as coherence in a work of science fiction will not count as coherence in a work of naturalism. What counts as coherence will be in part a function of the contract between author and reader about the horizontal conventions. [9]

And coherence is not the only thing that genres and their conventions condition and control. As we have seen in a previous quote from Lindley, "meanings imparted by film are primarily those imparted by its *subject*….The understanding of a film precedes the conventionalization of specific "syntactic" devices. The plot and subject make syntactic conventions understandable." Or in Searle's terms, what film is 'about,' capital-I Intentionally speaking (i.e., its subject), is a prime determiner of what it means. What this amounts to is agreement, on the part of Lindley and Searle, that the important and interesting conventions for producing meaning in fictional narrative are not reductively syntactic. While not as systematic as syntactical convention, generic convention would nonetheless appear to offer a stable foundation for analyzing and describing the production of meaning; however, both Searle and Lindley point out that genres are in fact transitory phenomena. Searle has mentioned in the preceding paragraph that authors feel fairly free to revolutionize literature by establishing new conventions, and Lindley says that regarding the varieties of type, genre, and style of filmic productions, 'that range and the styles within it are always changing and evolving.' In this, they concur with the strongly-held conviction of film theorists David Bordwell and Kristin Thompson [1] that genres are ever evolving.

We will close our analysis of Searle and this paper by highlighting a few points he makes about the pretend illocution of fiction; these provide interesting insights into cinematic production that emerge incidentally out of his thinking about fiction. Searle [9] points out that a common mode of pretending in many circumstances is to "pretend to perform a higher order or complex action by *actually* performing lower order or less complex actions which are constitutive parts of the higher order or complex action." There could hardly be a more succinct expression of how scripting, *mise-en-scene* and framing, and shooting are used (along with editing) to create the impression of a continuous world where only a partial and fragmented one exists. He further elaborates [9] that "in Austin's terminology, the [fiction] author pretends to perform *illocutionary acts* by way of actually performing *phonetic* and *phatic* acts." This reminds us that fiction film rarely employs direct statement, relying instead on partial utterances that are often indirect and largely visual in nature to produce its meanings. And meaning *can* be produced (even when as Searle insists the illocution is pretended) because cinematic utterances can be perceived as having illocutionary 'force' if the audience so receives them, which it does in significant part thanks to cinematic conventions. Lastly, Searle closes his analysis of fiction by addressing the core question: How exactly, in absence of direct statements, can a film (or other work of fiction) 'say' anything? He has no doubt that

> serious (i.e. nonfictional) speech acts can be conveyed by fictional texts, even though the conveyed speech act is not represented in the text. Almost any important work of fiction conveys a "message" or "messages" which are conveyed *by* the text but are not *in* the text. Only in such children's stories as contain the concluding "and the moral of the story is…" or in tiresomely didactic authors such as Tolstoy do we get an explicit representation of the serious speech acts which it is the point (or the main point) of the fictional text to convey. [9]

But while "literary critics have explained on an ad hoc and particularistic basis how the author conveys a serious speech act through the performance of the pretended speech acts which constitute the work of fiction,…there is as yet no general theory of the mechanisms by which such serious illocutionary intentions are conveyed by pretended illocutions" [9]. Here too it seems we must presently be satisfied with particularistic rather than systematic description.

4. REFERENCES

[1] Bordwell, D., and Thompson, K. Film Art: An Introduction. McGraw-Hill, New York, 2001.

[2] Lindley, C.A. A Multiple-Interpretation Framework for Modeling Video Semantics. Paper presented at the ER-97 Workshop on Conceptual Modeling in Multimedia Information Seeking (1997, November 6-7). Retrieved December 17, 2003, from http://osm7.cs.byu.edu/ER97/workshop1/lindley.ps.gz

[3] Lindley, C.A. The FRAMES processing model for the synthesis of dynamic virtual video sequences… in (Proceedings) Ninth International Workshop on Database and Expert Systems Applications (1998), 119-22. Retrieved December 18, 2003, from http://80-ieeexplore.ieee.org.proxy.lib.sfu.ca/iel4/5718/15304/00707390.pdf?isNumber=15304&prod=IEEE+CNF&arnumber=707390&arSt=119&ared=122&arAuthor=Lindley%2C+C.A.%3B

[4] Lindley, C.A. Generic film forms for dynamic virtual video synthesis… in IEEE International Conference on Multimedia Computing and Systems, Vol. 2 (1999), 97-101. Retrieved December 17, 2003, from http://80-ieeexplore.ieee.org.proxy.lib.sfu.ca/iel5/6322/16898/00778148.pdf?isNumber=16898&prod=IEEE+CNF&arnumber=778148&arSt=97&ared=101+vol.2&arAuthor=Lindley%2C+C.A.%3B

[5] Lindley, C.A., and Srinivasan, U. Query semantics for content-based retrieval of video data: an empirical investigation... in (Proceedings) Ninth International Workshop on Database and Expert Systems Applications (1998), 355-360. Retrieved December 17, 2003, from http://80-ieeexplore.ieee.org.proxy.lib.sfu.ca/iel4/5718/15304/00707424.pdf?isNumber=15304&prod=IEEE+CNF&arnumber=707424&arSt=355&ared=360&arAuthor=Lindley%2C+C.A.%3B+Srinivasan%2C+U.%3B

[6] Lindley, C.A., and Vercoustre, A.-M. A Specification Language for Dynamic Virtual Video Sequence Generation… in International Symposium on Audio, Video, Image Processing and Intelligent Applications (1998). Retrieved December 17, 2003, from http://citeseer.nj.nec.com/rd/50688729%2C129657%2C1%2C0.25%2CDownload/http://citeseer.nj.nec.com/cache/papers/cs/2633/http:zSzzSzwww-rocq.inria.frzSz%7EvercoustzSzPAPERSzSzisaviia98.pdf/a-specification-language-for.pdf

[7] Lindley, C.A., and Vercoustre, A.-M. Intelligent Video Synthesis Using Virtual Video Prescriptions (1998). Retrieved December 19, 2003, from http://citeseer.nj.nec.com/cache/papers/cs/2633/http:zSzzSzwww-rocq.inria.frzSz~vercoustzSzDOCSzSzICCIMA-98.pdf/lindley98intelligent.pdf

[8] Lindley C.A., and Vercoustre A.-M. Generic Viewer Interaction Semantics for Dynamic Virtual Video Synthesis (1999). Retrieved December 17, 2003, from http://citeseer.nj.nec.com/cache/papers/cs/8547/http:zSzzSzwww-rocq.inria.frzSz~vercoustzSzPAPERSzSzVisual99.pdf/generic-viewer-interaction-semantics.pdf

[9] Searle, J.R. Expression and Meaning: Studies in the Theory of Speech Acts. Cambridge University Press, Cambridge, 1979.

[10] Searle, J.R., Kiefer, F., and Bierwisch, M. (eds.). Speech Act Theory and Pragmatics. D. Reidel Publishing Company, Dordrecht, Holland, 1980. (Introduction).

[11] Searle, J.R. Mind, Language, and Society: Philosophy in the Real World. Basic Books, New York, 1998.

Unexpected, Unremarkable, and Ambivalent
OR
How the Universal Whistling Machine Activates Language Remainders

Marc Böhlen
University at Buffalo
marcbohlen@acm.org

JT Rinker
University at Buffalo
jtrinker@buffalo.edu

ABSTRACT
In this paper, we describe a situated machine that allows us to explore the idea of the language remainder from an unusual perspective.

Keywords
Language processing, philosophy of language, languages without words, low bandwidth direct communication, interface design and belief, animal-human-machine interaction

1. INTRODUCTION
One often takes for granted the many ways of knowing that are needed in order to make meaningful statements about the world. Logically correct statements are comparatively easy to formulate because the rules of logic are known and formalized; sensible and meaningful statements, as in statements of common sense, are much harder to specify. Computer systems have been struggling since their inception with this problem. What are the requirements for formalizing the automation of making sense? How can one test if "sense" has been made? In this case, people are still the best arbitrators. For this reason, the automation of making sense includes the capacity to interact with people on their terms, to reason about their input on their terms, and to share the output again on their terms. On the input side alone, the problem is exceedingly difficult. The most obvious natural input modality, speech, is difficult to automate. Despite decades of research, no speech recognition system is capable of graceful and robust recognition when dealing with unrestricted vocabularies and untrained speakers [1, 2]. But there is no need for pessimism. One day this problem may be solved. And incremental improvements in the machine learning community may just bring about the complex kinds of data representations we will be hard pressed not to call meaningful.

2. The Language Remainder
While general-purpose meaning may be achieved using current computational approaches, it is questionable whether fine-tuned variations of meaning, layered, contradictory, situation-specific, and temporally limited meaning can be captured. Fuzzy aspects of language such as innuendo defy formal linguistic descriptions and are not even modeled in computational models of language that seek to represent "all aspects of communication" [3]. Assuming that the making of meaning will remain intimately linked to language, it makes sense to ask how language should be represented in machines capable of making sense. Language is more than a database of words and rules by which these words can be combined. Languages are not static, and not fully describable through the grammatical rules that constrain them, however refined such rules may be. Many philosophers of linguistics, semioticians, and writers have pointed this out. Lévi Strauss reflected on the counterintuitive robustness of systems of semantic impoverishment [4]. Eco revealed the follies of those who have tried to uncover, rediscover or create a perfect language, failing because of their inflexibility, utilizing rigid rules [5]. Maturana and Varela spoke of language as a process, not as a static collection of words and rules. They coined the term 'languaging' to better represent the richness and complexities created in the exchange between cognition, language and language use [6]. Lecercle proposed the term 'remainder' as a formal entry into the levels below, above, and adjacent to strait-laced meaning covered by linguists' version of language [7]. For Lecercle, the remainder is the fallout from the intended use of language. It is the essence of poetry and metaphor, but also of miscommunication, word play, and double-entendre. It is the fuzziness and leakage of meaning amongst words.

But how could one possibly attempt to include the language remainder in computational systems? Is it at all possible, given that the rigor of linguistics seems even tighter in the limited corpora of texts, the defined rules and intelligent but blind numerical clustering methods underlying computational linguistics? In order to prevent varied and flavored meaning and language remainders from being filtered out of computation, it might be worthwhile to investigate varied and less structured forms of knowing, unorthodox methods of input, and unexpected flavors of output. Easier said than done. This is not only a difficult problem, but also a poorly defined one. How can one even begin to formulate such issues as tasks, let alone make them computationally tractable? We offer no general solution to this problem. However, we suggest an alternate approach towards the problem. Would it be possible to reduce the complexity of language to a more manageable subset, albeit one that still allows instances of language remainders to exist? Rather than creating a machine that is conceived with hardwired

First published at COSIGN-2004,
14 – 16 September 2004, University of Split (Croatia),

knowledge of a fully structured language, including vocabulary and grammatical system, would it be possible to create a device that is only 'primed' for language? Is the ability to perceive and imitate a limited bandwidth of data that is mutually suggestive by machine and user as communication, a precursor to language? Can meaning arise in such a situation? In this context we offer the following thought experiment:

Imagine walking down a corridor, past an elevator, lost in thought. You hear a whistle. You stop and search in curiosity or disdain for the person seeking your attention. Nothing. You notice an intercom-like device embedded in the wall. Again you hear a whistle. You walk to the device, stare at it. Another whistle. You whistle back. The device whistles again, in a different fashion. You respond, and realize now that you have engaged in a whistle exchange with a machine.

3. Accessing the Language Remainder

We have designed, built, and tested such a device and call it the Universal Whistling Machine (U.W.M.). It senses the presence of moving creatures in its vicinity and attracts them with a signature whistle. Given a response whistle, U.W.M. counters with its own composition, based on a time-frequency analysis of the original. The following paragraphs describe the contexts in which we place this device, and how we succeed and fail to introduce the slippery language remainder into a machine.

Figure 1. U.W.M. in standalone mode.

4. Unremarkable Forms of Expression

How can whistling include the above-mentioned potential of language remainders? Is it too limited a form of expression? Maybe not. Whistling, like other forms of low-bandwidth communication – grunting, coughing, laughing, and humming - are phoneme-less modes of communication. One might consider such modalities to be extreme cases of homophony. The fact that one word can have many meanings has led to interesting debates on the limits of language, such as, - are there more things than words in the universe? If we are structurally confined to the homophony condition, why not make a virtue of the vice? Whistling is extreme homophony.

There are practical reasons that make such low-bandwidth forms of expression interesting. Technically, it is much easier to analyze whistles than it is to analyze spoken language. The frequency spectrum of whistles is far simpler than the frequency spectrum of other vocal emissions. A low-bandwidth, high-contrast signature is typical of whistles and contrasts sharply with voice signals. Because of this, whistles can be readily differentiated from singing [8]. Furthermore, whistling is a low-entropy mode of information encoding, and easier to separate from a noisy background than speech.

But technical details should not guide this discussion. Is there a merit -beyond delivering a tractable signal-processing problem- that makes such low-bandwidth exchanges desirable and interesting?

We believe the answer is yes. Whistling is an underexposed and unremarkable area of expression: It is raw and direct. Whistling is immediate, code and content at once. Whistling is pleasure, admiration, warning, unfiltered desire, cipher, and protest. Emmet Till, a young man of color, was lynched in 1955 after wolf-whistling in the presence of a white woman. Intuitively understood, whistling is transcultural communication below the radar of social etiquette. The idea that language is material seems uncontested in such low-bandwidth expression – here the body speaks on its own terms. Coughing and humming bring body fluids to our lips. If language is material, then it is in the forms of low-bandwidth expression that this becomes most apparent. Language, as material, unfolds. Everything said is said by someone [6], and every utterance is uttered through a mouth. In whistling and humming we revert back to less articulated states of communication. Whistling is simple, unassuming, and highly unremarkable. Yet, as De Landa observes, it is in the unremarkable where human universals are to be found [9].

4.1 Agency via Semantic Impoverishment: Languages Without Words

There are numerous examples of human communication systems based entirely on whistling. This phenomenon was widely reported during the late 1970s in linguistics' circles [10]. Two of the better-known whistling languages are "el Silbo", practiced on the Isla de la Gomera, one of the Canary Islands off the coast of Morocco, and the whistled language of Kuskoy, a remote village by the Black Sea in Turkey that has only recently been connected to the telephone grid. In La Gomera, the skill is still being passed on to youngsters today. These two whistling languages contain a subset of the spoken language with which they coexist.

But why do such whistling systems even exist? Some linguists have speculated that they originate from the need for secrecy and robustness. High-pitched strong whistles travel farther than the spoken word. In La Gomera, the maximum distance that a whistled message can travel has been reported to reach 10km. Skilled whistlers are said to be capable of producing whistles of 130dB (measured 1 meter in front of the whistler [11]).

Whistled languages are generally "reduced" languages, in the sense that not everything that can be expressed in speech can be expressed by whistling. However, they are far closer to languages than to a code or to simple signalization systems. They are speech-like and carry the vocabulary, the grammar, and in many cases the phonology of the local language they have emerged from, especially at the level of prosody.

Whistling occurs across languages and cultures. Whistling is a primitive communication in most human languages. Whistling is a kind of time travel to a less articulated state. Whistling carries the potential for song, pleasure, and secretive message. All people have the capacity to whistle, though many do not whistle well. Since it lacks phonemes, whistling is a pre-language language. Unlike other forms of low-bandwidth encoding, such as hand clapping, whistling is nuanced and offers the advantage of a rudimentary dialogue. Within the laws of digital signal processing, human utterances and machine-based audio signals are more similar than different. Whistling is much closer to the phoneme-less signal primitives compatible with digital machinery than to the domain of spoken language. As such it offers itself an Esperanto of communication not only across language boundaries, but also across species boundaries.

But how should one represent an artificial whistle? And how can one interface with it?

5. Interfaces beyond Navigation and Control

Artifacts are inscribed with involuntary meaning. Artifacts have signatures. They reveal their signature in the materials and techniques used to make them and in the logic and worldview under which they were conceived. In this sense, any artifact is a statement about the world. The computer interface as artifact is no exception and reveals a particular belief system about how we are expected to engage with machines. The usual distance between a human and a computer is arm's length with a keyboard as the mediating device. Robust but rigid, this configuration has come to define the human computer interface. Attempts to make this interface more 'natural' tend to include direct input modalities, i.e., modalities that require no adaptation by the human. Speech and gesture recognition suggest that a person could communicate with a machine as if the machine were human. Despite the fact that the implementation is troublesome, the idea seems appealing. But even when the problem is solved, we may be no closer to language remainder compatible exchanges with machines than before. The problem is not just technical.

In 'Behind the Blip,' Matt Fuller analyzes Harun Farocki's "I Thought I Was Seeing Convicts" [12]. This video consists of video clips, footage from inside the California state prison, Corcoran. The notorious reputation of the Corcoran penitentiary is based on the rumor that prison guards were alleged to have set up fights between members of rival gangs in the exercise yard. Guards then watched the yard via video surveillance and bet on the outcome of the fights. Fuller points out that the "interface" here is not where one might expect it to be. The screen in front of the guards is the site where the visual cues were available, but the structural interface that made the scenario possible is built into the 24-7 surveillance system, the image storage capacity of the video system, the prison architecture, the command structure of prison officers, and the range and caliber of the guns used to enforce order [13].

Interfaces extend beyond the immediate point of sensory exchange between the human and the machine. Interfaces are only meeting points, rhizome-like, between other dependent and informing mechanisms. Interfaces are made up of objects and the use of these objects over time by people with various needs and desires. The effectiveness of an interface is a function of what it allows as well as what it disallows. As with language, there is an interface remainder, something that falls between the cracks, something the user manual does not mention.

6. Situation as Interface

U.W.M. recognizes when people are approaching by analyzing the data from a built in video camera. The machine then waits until a person who has entered its field of perception moves outside of it before emitting a whistle.

We use a blatantly direct method of attention seeking. Moreover, it is a method that is decidedly outside of standard human-machine interface conception. Machines do not initiate exchange and machines do not whistle at people. Here, the device latches onto a particular form of human-human 'interface' for its agency. The awkwardness of this situation can provide the activation energy necessary to ponder alternatives to established interface models. To whom should we complain when a machine harasses us? What then is the nature of this kind of harassment – the whistle itself or the imitation of desire? And the people that decide to turn around and approach the machine are not 'users'. Our machine offers no instructions or visual cues on how to 'use' it. U.W.M. has no interface in the traditional sense. Only those who choose to whistle back at the machine will be able to extract something from it. U.W.M. is not picky. Whether one produces a fine-tuned melody or a rough and graceless whistled sound is of no significance for the machine. U.W.M. works with the whistle it gets.

Figure 2. U.W.M. folded into a wall

In its deployed state U.W.M. is embedded in a wall, and the situation it initiates becomes its interface. This anti-interface creates a window of ambivalence in which the casual pedestrian is offered an unusual experience. Being whistled at by a human might generate feelings of anger, disgust, or pleasure. Being whistled at by a machine forces humans to recognize their preconceptions about the role of the machine vis-à-vis the human. U.W.M. has no goal state other than to generate such a situation, such a discontinuity of expectation, from which an experience akin to the pleasure of the language remainder can result.

In order to make such events likely, we are installing foldable and portable U.W.M.s in quiet, low-traffic spaces of exchange and transition. Restrooms, corridors, and elevator halls (Fig. 2) are examples of transition spaces that suit this experiment. They are multi-purpose spaces, home to a variety of "services" not inscribed into the formal architecture. For instance, people tend to linger around while waiting for an elevator. Some people may wander down corridors lost in thought. The underdetermined nature of activities typical of such locations creates a temporary semantic void that is well suited for the experience U.W.M. offers.

Figure 3. European Robin (*Erithacus rubecula*)

7. Meaning Beyond Human Understanding

As whistled sound producers, machines, humans, and animals share a common denominator. Whistling and song we share with animals. Mammals and birds have the means for making songs and whistles. Ducks, robins, loons, and starlings whistle. White whales and bottlenose dolphins [14] whistle under water. Just as we carry physical remnants of our bodily evolution in us, we have the capacity for whistling. When we whistle, we acknowledge the plane of being underneath phonetically articulated language that we share with other species. Beyond alternatives to computer interfaces, U.W.M. also offers the potential for a new approach to human-animal communication.

U.W.M. is capable of imitating certain bird whistles as easily as it can synthesize human whistles. (See our examples collection for a synthesized robin whistle [15].) Would a songbird imitate a whistle from a whistling machine just as readily as it imitates whistles from birds and certain human-generated sounds? Could this lead to new modes of human-machine-animal exchanges? Can we find alternatives to exchanging information beyond the species boundary? And when we do, which rules shall we apply? [16]. Interaction research to date has centered on human-machine exchange. The worn-out narrative of subduing our surroundings is unconsciously but consistently inscribed in the computer interaction program. How can we open ourselves to the otherness of the animal state, so utterly unlike our own that Agamben calls it 'unopen' [17]. Can the language remainder, made partly accessible by machines like U.W.M., help us open the unopen? St. Francis derived great admiration from his ability to speak to the animals.

8. The Remainder Persists

How do we speak of the unspeakable? Where is a 'right' place to begin an investigation of the informal? Sometimes detours are the direct path. That is what Paul Miller observes on his tour through the art and science of rhythm. Miller starts with sound and works through literature, the visual arts and philosophy, only to land back at mystery of music, informed by the art of the DJ [18]. It was all in the mix from the beginning. Before language there was gesture, then utterance. The past is phonemeless. Pure signal. Before the word forms on your lips, air is pressed through your mouth. Today, the word rules. Tomorrow, the wordless utterance of the machine that is so far beyond human it is animal again, awaits us.

9. Technical Notes

The U.W.M. project is coded in C++ and PD. U.W.M senses people passing by via a low-cost, IEEE1394 enabled CCD camera/sensor at video rate with a public domain camera driver [19]. The data from this video stream is parsed by standard and custom-made machine vision routines. Oblong objects traversing the camera's field of view at the speed of a casual pedestrian trigger the device into whistle mode. The machine waits until a person who has entered its field of perception moves outside of it before a whistle is produced.

Given a whistled response to this invitation, U.W.M. will in turn whistle back at an individual. Sound capture occurs through a USB enabled noise-reducing microphone. Signal sampling occurs at 44.1 kHz. In order to prevent U.W.M. from erratically responding to all kinds of sounds and noises, the audio input is analyzed for its 'whistleness'. We have developed a robust decision mechanism that allows us to react specifically to whistled input. Incoming signals are passed through an FFT based pitch tracker, and analyzed on two passes. Data outside of the three-sigma boundary are discarded as outliers. The filtered data is then reprocessed for the standard deviation, curtosis, median and bandwidth. Select boolean conditionals based on these parameters allow us to exclude most unwanted signals. The device is able to reject background noise, including music (pop, hip-hop, strings and high wind instruments and song from both male and female voices). Cries from young children, however, can elude the current detection mechanism.

9.1 On the Whistleness of Whistles

Other researchers have previously investigated the properties of whistles [20]. Here the aim is different and the return value is different. This machine speaks as spoken to. It whistles back at an input whistle. The schematic below gives an overview of the algorithm we use for U.W.M.'s artificial whistles.

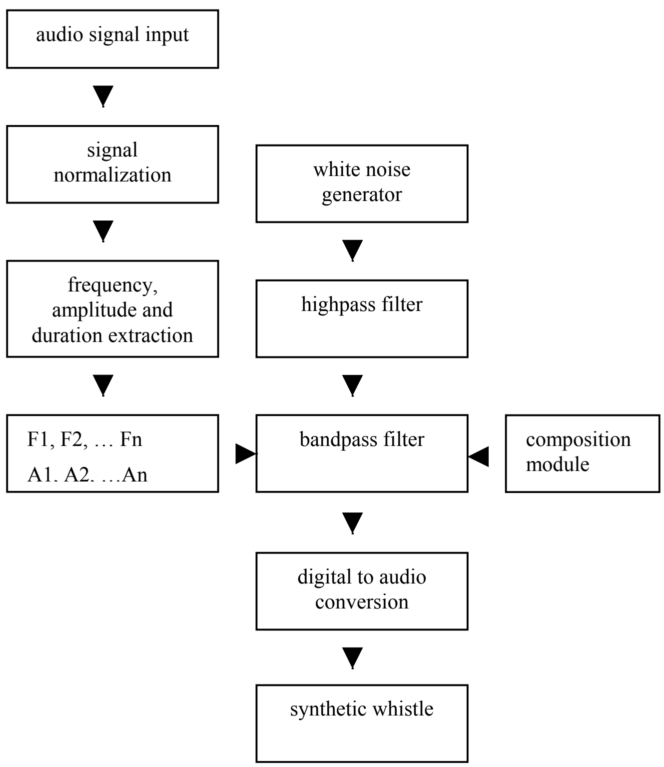


Figure 4. Schematic diagram of the main audio processing components in U.W.M.

The whistle synthesizer is based on the basic spectral characteristics of a human whistle. Most human whistles exhibit a fundamental frequency with very few harmonics, (often only one or two) as well as a band of high frequency noise. Here, we create a whistle through a process of subtractive synthesis. We use white noise as a signal generator for the whistle synthesizer. The noise is passed through a pair of filters in series. The first is a single-pole high pass filter. This filter eliminates most low frequency noise. The second filter is a band-pass filter, which passes a sinusoid at a specified center frequency and attenuates all other frequencies. The center frequency is the pitch for the whistle, and the "Q" (quality factor, similar to bandwidth) of the filter is set proportional to the center frequency. Data from the pitch-tracking device drives the whistle synthesizer. However, the prototype machine can also generate random whistles by supplying its own center frequency and amplitude data to the synthesizer.

9.2 From Synthesis to Transformation

To generate responses to perceived input whistles, raw data is collected from the pitch tracker as frequency and amplitude pairs. High threshold gates on amplitude content as well as other algorithms, including interpolation in the pitch domain and peak detection in the amplitude domain, help smooth out areas in which the pitch tracker fails. This is most notably the case at the ends of whistled pitches and between attacks of whistles where pitched signals are not present. These smoothing operations allow us to extract from the data stream only those elements essential to the whistle itself. In addition to being able to imitate input whistles, we have also created numerous forms of whistle transformations. For instance, adding a fixed pitch interval to the pitch data creates a transpositional transformation. This results in a response whistle that is higher or lower in pitch than the input whistle. Contours of the input whistle can be increased, decreased or inverted to give a semblance of the shape of the input whistle while varying the pitch. Time transformations read the data at a rate different from the capture rate. This creates responses that are slower or faster, and independent of pitch and amplitude and their subsequent transformations. Tempo rubato is created by randomly changing the read-time interval between each index of the pitch and amplitude arrays, thus speeding up some portions of the response whistle while slowing down other portions. The data can be read backwards as well as forwards, essentially reversing the input whistle. All of these transformations can be applied in parallel providing a wide palette of responses that are all based on a user's input whistle. We make use of Miller Puckette's PD environment [21] for these operations.

10. ACKNOWLEDGMENTS

This text is dedicated to the memory of Hope Kurtz. The mockingbird knows.

11. REFERENCES

[1] R. Cole, J. Mariani, H. Uszkoreith, A.Zenen, V. Zue, "Survey of the State of the Art in Human Language Technology", Center for Spoken Language Understanding CSLU, Carnegie Mellon, 1995.

[2] M. Maybury, Language Technology: A Survey of the State of the Art Language Resources: Multimodal Language Resources, MITRE Corporation, 2002.

[3] R. Hausser, "Foundations of Computational Linguistics: Man–Machine Communication in Natural Language", Springer Verlag, 1999.

[4] C. Lévi-Strauss, "The Savage Mind (La Pensée Sauvage)", University of Chicago Press, 1967/1962.

[5] U. Eco, "Serendipities, Language and Lunacy", Harcourt Brace & Company, 1999.

[6] U. Maturana, F. Varela, "The Tree of Knowledge", New Science Library, 1987.

[7] J. Lecercle, "The Violence of Language", Routeledge 1990.

[8] D. Gerhard, "Computationally measurable differences between speech and song", PhD Thesis, Department of Computer Science, Simon Fraser University, 2003.

[9] M. De Landa, "A Thousand Years of Nonlinear History", Zone Books, 1997.

[10] R. Busnel and A. Classe, "Whistled Languages", Berlin, Springer Verlag, 1976.

[11] J. Meyer, “Comparison typologique des langages sifflés: processus cognitifs”, PhD thesis, University of Lyons, (synopsis on the web: http://www.ddl.ish-lyon.cnrs.fr/)

[12] H. Farocki, “I Thought I Was Seeing Convicts”, with text “Controlling Observation” in: Harun Farocki, Nachdruck/Imprint. Texte/Writings, Lukas and Sternberg, 2001.

[13] M. Fuller, “Behind the Blip”, Autonomedia, 2003.

[14] L. Rendell and H. Whitehead, “Culture in whales and dolphins”, Behavioral and Brain Sciences (2001), 24, 309–pp. 382.

[15] M. Böhlen, “More Robots in Cages”, Space&Robotics2000, 4th International Conference on Robotics for Challenging Situations and Environments, p. 206 - 212, Albuquerque, New Mexico, 2000.

[16] www.buffalo.edu/~mrbohlen/uwm.html

[17] G. Agamben, “L’uomo e l’animale” (The Open, Man and Animal), Bollati Boringhieri / Stanford Press 2002/2004.

[18] P. Miller, aka DJ Spooky That Subliminal Kid, Rhythm Science, MIT Press, 2004.

[19] The CMU 1394 Digital Camera Driver is maintained by the Robotics Institute at Carnegie Mellon University. See:

http://www-2.cs.cmu.edu/~iwan/1394/index.html

[20] L. Prechelt, R. Typke, “An interface for melody input “, ACM Transactions on Computer-Human Interaction (TOCHI), Volume 8 Issue 2, 2001.

[21] M. Puckette, "Pure Data: another integrated computer music environment." Proceedings, Second Intercollege Computer Music Concerts, Tachikawa, Japan, pp. 37-41, 1996.

Formal Semantic Models for Images and Image Understanding

Duc Do Audrey Tam

School of Computer Science and Information Technology, RMIT University

GPO Box 2476V

Melbourne, Australia 3001

{ducdo,amt}@cs.rmit.edu.au

ABSTRACT

A number of formal models for images [13,27,28] and models for text and image matching [1] have been proposed, but they have not sufficiently dealt with features with high-level semantics. While formal models are supposed to be precise, their structures should allow for the level of subjectivity involved in interpreting the high-level semantics inherent in images.

In our earlier work, we have shown that by restricting image retrieval to a specific domain, we can use logical reasoning based on common sense knowledge bases and the knowledge extracted from text corpora from the same domain to infer higher level semantics from lower level semantics. The interpretation of these lower level semantics, usually involving objects in the image, is subject to a lower level of subjectivity, hence making it possible to build an image model that is reasonably *objective*.

Based on these observations, we propose that an effective and feasible approach to build high-level semantics into image retrieval is to build semantic models for both the image (the object of meaning) and image understanding (the perception of meaning). The image model will aim to capture image features which are commonly accepted within a certain domain. The image understanding model will include mechanisms for subjective interpretation and will be associated with correspondence functions which measure similarity between instances of these two models. This level of similarity, or the semantic distance, can be called the *semiotic gap*. Using this framework, the image retrieval problem can be deemed equivalent to the problem of defining a correspondence function that delivers the theoretically, or empirically, narrowest semiotic gap.

We propose to construct the formal image model based on the concepts of semiotic structures, and an image understanding model based upon insights into how knowledge inference could assist with image retrieval. In this paper, we present the formal image model and argue why this model is suitable for the retrieval of visual data. An image understanding model, which is under ongoing research, is also briefly discussed with results of some preliminary experiments.

Keywords

Formal image model, semantic model, image understanding, logical reasoning, knowledge base.

First published at COSIGN 2004

14 – 16 September 2004, University of Split (Croatia)

1. INTRODUCTION

Popular image retrieval approaches are either content-based (CBIR) or text-based. Neither approach addresses image semantics very well. CBIR's low-level features, such as colours, textures and shapes, fail to capture the high level semantics inherent in images. Text-based approaches, with retrieval strategies largely borrowed *as-is* from the text retrieval discipline, rely on textual annotations to capture the meanings of images. Proposed methods using either approach have so far not been able to overcome the semantic gap problem [31]. As the semantic gap problem is fundamental to the problem of not being to satisfy a certain information need, the general performance of state-of-the-art image retrieval is very poor compared to what can be currently achieved in text retrieval.

Most work in CBIR has taken the view that low-level features are objective. This however is not entirely true because a photo image may not necessarily represent the *true nature* of the physical environment captured in its visual frame (although philosophically speaking, whether humans ever attain an understanding of the true nature of anything is questionable). For example, the colour of someone's hair in a photo may appear a lot lighter if the photo is well lit. The dominant colour feature for this person's hair expressed in some RGB value may not result in an accurate representation of the true nature. In some situations, a photographer may choose to focus on a particular object and leave out significant parts of the surroundings, resulting in a photo that misleads viewers' interpretation of the image content. Therefore a photo image is, at best, only a subjective representation of the true nature. The level of subjectivity is dependent on the equipment being used, the technical conditions of photography, such as the source of lighting or the film development technique, and the photographer's intention. Even outside the photographic domain, human visual perception already modifies true nature, as the human brain "has no direct access to information about properties of the external environment; it has access only to whatever information is available in the retinal image" [4]. Rossotti 1983 [29] illustrated this point by describing how mackerels appear silvery-green while in fact their scales are colourless, or how rainbow trouts appear to change colours when viewed from different angles. This is not even taking into account certain visual deficiencies such as colour-blindness, true to Kant's philosophy that "all our knowledge begins with the senses, proceeds then to the understanding, and ends with reason" [20].

Text-based approaches rely on language forms for both classification and retrieval. Although the interpretation of these forms can be objective through spelling and grammatical rules, the interpretation of their meanings can be subjective.

Taking into account these observations, we can say that every single visual feature in images is subjective. Therefore, an effective image retrieval model must be able to differentiate between and work with both semantic features and possible representations of true nature. A formal model with mechanisms to store, compare and retrieve features is not sufficient. It must be able to associate features with meanings in some forms of semiotic structures, analyse the entire feature space as a whole and reason with them in order to infer other meanings. This has given rise to the need of a formal semantic model for the image. As the baseline accuracy of CBIR is generally limited and often worse than that of keyword-based methods [23], this semantic model needs to be built on the foundation of textual annotations with provisional support for CBIR features.

2. FORMAL SEMANTIC MODEL FOR THE IMAGE

2.1 Prerequisites

An effective formal semantic model for the image should satisfy the following requirements:

- *It should support all image features, from primitive to abstract.* For the model to be complete, primitive features need to be supported as under certain conditions, they have been proven to be effective for certain search problems.
- *It should map naturally to the human search strategy*. The reason for this is obvious, as the well-researched semantic gap problem [14,31] is caused by two main deficiencies: the annotation deficiency and the query language deficiency. It is unnatural to expect humans to think in terms of low-level features. Squire and Pun 1998 [32] has demonstrated that algorithms which compute the similarity between images often make judgments that fail to match those of the human researcher, unless the human has been well trained in the computer algorithm.
- *It should be able to capture domain-specific high-level semantics*. Cyc is an example of a knowledge base system that helps constrain a search by tying knowledge to different domains, or microtheories [22]. The growth in both quantity and quality of Cyc has partly validated this approach. Technically, microtheories are only namespaces. However, semiotically, microtheories allow disambiguation of concepts given domain constraints. As image collections can be viewed as a knowledge base, they should also inherit this property.
- *It should support the incremental addition and modification of features as the domain grows (either in quantity or in quality)*. Compilers of dictionaries, year in year out, have to deal with the problem of introducing new words or expressions into dictionaries, or taking words out (albeit less often than the former activity) of them. It is the same problem when dealing with a body of knowledge related to a specific domain. Sometimes these domain shifts can be quite overarching, such as the departure of mathematics from the field of philosophy in modern times. In the old days, philosophers such as René Descartes or Alfred North Whitehead were also philosophers, making the study of mathematics part of the study of philosophy.
- *It should support unlimited levels of abstraction granularity*. A typical annotation structure, such as one found in the Lonely Planet image collection [6], that relies on metadata has a fixed number of levels for knowledge categorisation. This is partly implemented as an effort to compromise between information need and annotation efforts. However, the domain in question constantly changes, and as the body of knowledge may unpredictably expand in both breadth and depth, a fixed number of category levels is inadequate.

2.2 Success Criteria

The proposed image model should not be aimed at solving the general information retrieval problem, such as one offered by the Google text search engine [12]. Attempting to do this would fall into the *context dilemma* trap [7]. At the same time, if the theoretical model has any fixed requirements that effectively prevent it from practical adoption, such as requiring an excessive amount of time to annotate each image, it should not be recommended either.

Other success criteria of the model implementation should include:

- *It must be testable with real photo images as opposed to synthetic images*. Low-level features in synthetic images such as colours and shapes are usually well grouped, hence more easily identified, potentially resulting in a positive bias for content-based methods. We assume that a model that works well with real photo images will also work well with synthetic images.
- *Improvement in image retrieval using the model must be measurable*. The reason for this is obvious. These measures may include traditional information retrieval parameters such as recall and precision. Based on our own observations from the usage of the Lonely Planet image collection [24], the general aim is to attain significant and consistent improvement in recall, and *ideally* higher precision, for all image queries.
- *Improvement in image retrieval using the model must be comparable to other methods*. This criterion will depend on the availability of a common image benchmarking framework. The image retrieval community still does not have anything close to the TREC tracks [34] that are widely accepted and used in text retrieval experimentation and benchmarking. The leading image benchmarking framework, the Benchathlon Network [26], is still at its early stage and currently only supports CBIR methods. However, the objectives of the Benchathlon project are consistent with what we need for a suitable benchmarking framework. These include a standard data collection, a set of standard queries, a form of ground truth, a benchmarking engine, a set of performance measures and a standard access protocol.

2.3 Model Building Blocks

In text retrieval, everything can be traced back to terms and documents. A text collection consists of documents and a document consists of terms. At least that is the current basis for most, if not all, text retrieval methods. The content, information quality, and information quantity of a document can be examined by analysing these terms. Despite the issues with word sense and syntactical ambiguities, these terms, belonging to specific subsets of a language or languages, already render themselves semantically interpretable.

There is no such simple equivalence in image collections, even though from a semiological point of view, images can be viewed as part of a language of signs [2]. If each image is viewed as a document, what forms of *visual units* can be deemed equivalent to terms in text? As we are dealing with two completely different forms of media, such equivalence may never be found. Some traditional image annotation approaches that try to associate a whole image with phrases or sentences (the simple captioning approach) are problematic. Frege, according to Haaparanta 1985 [17], observed that neither a word nor a noun phrase nor a verb phrase in isolation does tell us something. Wollheim 1996 [35] suggested that assimilating pictorial art, which can be quite unstructured, with textual annotations in forms of sentences, which are always structured, is not always possible. Because of these observations, many authors have tried breaking down the structure of the image by proposing their ideas of visual units, each aiming to solve problems related to their own domain. In pattern recognition, a field closely related to image retrieval, visual units are regarded as assemblies of low-level physical features such as pixels (with values denoting colours) or lines or shapes. Although these visual units, derived from formal analysis, are part of the basic vocabulary used in most art classes, they cannot be used in themselves to describe artistic or esthetic qualities in a visual piece of work. Bloomer 1990 was quite emphatic in proving this point in stating that “in its simplest terms, saying that all art consists fundamentally of line, shape, and color is like saying that all food is made up of protein, carbohydrates, and fats” [3].

To define visual units that can be used in a semantic model, we work on a basic assumption that domain experts have to be able to think in terms of these visual units. In fact, domain experts should not have to translate their information needs into other forms to be able to perform retrieval. The most straightforward way to achieve this is to define visual units that capture units of knowledge that are relevant to a particular domain. Johnson 1987 [19] proposed a concept called *image schema*, which is a derivation of Kant’s concept of *schema*. In Johnson’s definition, an image schema is “a mental pattern that recurrently provides structured understanding of various experiences, [and] is available for use in metaphor as a source domain to provide an understanding of yet other experiences”. The keywords and phrases to note in this definition are: *pattern*, *structured understanding*, *understanding of various experiences*, and *metaphor*. From these, we can infer that an image schema must:

· be reusable (being a pattern);

· have a structure and provide a structured process for interpretation;

· allow the aggregation of various pieces of knowledge, thus requiring some form of nesting structure;

· facilitate the mental process of comparison and instantiation (with and to other experiences, either relating to or symbolised by the same image schema or not).

With these explicit properties, we propose to use image schemata as visual units for the formal semantic model. Each schema must represent a concept that is significant to a particular domain. The same idea is being used in the Cyc knowledge base, where, for example, a constant called #$VisualImage assumes different meanings in different microtheories (MT’s). In the #$BaseKB MT, it is a type of #$InformationBearingWavePropagation, each instance of which is an event in which visible light is generated in a particular pattern, which (does or might) contain information for an observer; while in the #$UniversalVocabularyMt MT, it is simply classified as a #$TemporalStuffType, which interestingly enough, focuses on the *temporality* facet of this concept. Linguistically, an image schema can also be classified as a kind of source domain [21], which provides a reference point for other concepts.

2.4 Terminology

In discussion of the semantic model, the following terms are used:

· *Image collection*: a collection specific to a particular domain. All images in a collection must be relevant to the domain in question.

· *Image*: an image in a collection.

· *Image schema* (or *schema*): the abstraction of a visual unit in an image. An image contains instances of image schemata. An image schema may contain other schemas. A schema may inherit properties from higher-level schemata. Schemata can thus be grouped and categorised into a tree structure.

· *Image schema instance* (or *schema instance*): an instantiation of a schema within an image. All instances are unique to the image they belong to. For example, we all have a certain mental picture of an “abstract dog” (not any specific dog). The instance of Boo the dog in Picture A is different from the instance of Rover the dog in Picture B. The instance of Boo the dog in Picture A is also different from the instance of Boo the dog (albeit the same dog) in Picture C.

· *Semantic features (*or *features)*: semantic properties of an image, a schema, or a schema instance.

· *Metadata*: properties of an image, not related to or derived from any schemata or instances.

· *Semantic relationships* (or *relationships*): include *schematic relationships* and *instantial relationships*. Schematic relationships denote relationships among image schemata, while instantial relationships denote relationships among schema instances. In the current model, we propose two types of relationships: *spatial relationships* and *interaction relationships*. Instantial relationships do not have to be instances of schematic relationships, as opposed to schema instances, which are always instantiated from some image schemata.

2.5 Semantic Features

Eakins and Graham 1999 [9] proposed that image queries be categorised into three types based on the information need. These same query categories can be extended for image features to make them applicable to annotations. While only three query types were proposed, we propose the following four types of features. Features on all four levels can be present in images, image schemata, and schema instances.

2.5.1 Level 1 Features – Primitive

Level 1 features correspond directly with the primitive features of CBIR, including colours, textures and shapes. These features are automatically indexed and included as part of image annotations.

2.5.2 Level 2 Features – Derived

Level 2 features are those that can be automatically computed from level 1 features. A number of authors have proposed that this is possible. Eidenberger and Breiteneder 2002 [10] conducted an experiment to show that some human-world features such as symmetry, geometry and harmony can be computed from a set of primitive features including edge histogram, colour histogram and dominant colour. However, as the data set being used in this experiment only contains synthetic images, it is hard to guess if similar results can be expected with a data set containing real photos. Zhao and Grosky 2000 [37] even proposed that latent semantic analysis (LSA), a popular text retrieval technique, can be used to transform low-level features to a higher level of meaning. In a collection where certain objects can be consistently expressed with a common set of low-level features, it is easy to see how this approach could be feasible. However, if we are dealing with a collection diverse in topics and visual representations, this technique may not apply very well. As the proposed semantic image model is designed to work with all collections, it should have a built-in mechanism to capture these derived features, instead of just relying on automatic computation.

2.5.3 Level 3 Features – Topical

Level 3 features contain visible objects that humans recognise in the real world. The majority of work on semantics in image retrieval has focused on these objects. In our model, these features correlate directly to the image schemata. It is important to note that, within a certain domain, the set of topical features that are of interest can be finite, thus allowing the definition of a comprehensive list of objects applicable to a certain domain. An example of this definition can be found in the travel domain, as suggested by I'Anson 2000 [18]. This observation implies that we are not trying to define a generic list of objects that apply to all images under all circumstances.

2.5.4 Level 4 Features – Abstract Derived

Level 4 features are derived from features from the lower three levels or from other level-4 features themselves. These features usually represent abstract concepts such as feelings, emotions or interpretations, and thus, are highly subjective. Being subjective implies an inference process; therefore these features are classified as *abstract derived* instead of being simply *abstract*. Some common level-4 features can be derived within a particular domain with a low level of subjectivity. For example, if nudity is considered offensive in a certain domain, a specific rule can be defined that links the *is-nude* property of the *person* schema (a topical feature) to *offensiveness* (an abstract derived feature). In another domain, *offensiveness* may be linked to *violence* (another abstract derived feature). In this case, the *violence* feature is derived from a combination of other features present in the image.

2.6 Semantic Relationships

Queries involving relationships among certain objects potentially present in images are popular in image retrieval. For example, one may want to locate images with groups of people *in front of* a certain type of building, or images with a child *riding* a horse. Therefore, an effective semantic model for the image ideally should have a built-in mechanism to capture these semantic relationships. In the current model, we propose two types of relationships: spatial and interaction. In our future work, we may propose the addition of other types of relationships. It is important to note the role of schematic relationships in particular. Relating image schemata together, schematic relationships represent mental patterns built on individual experience and cultural or educational background. These include, for example, the notions of "dog *chasing* cat" or "big fish *eating* small fish". We assume that some kinds of relationships relevant to a particular domain are stereotypical, and thus, can be captured by schematic relationships.

2.6.1 Spatial Relationships

Spatial relationships among objects are usually the first types of relationships to be considered. This is probably due to the fact that these relationships are in themselves visually related. Relating back to our proposal to use image schemata as visual units and considering their roles in the real world as an aid for physical navigation [19], it only seems natural that spatial relationships among these schemata must also be considered to make navigation more effective. As the definition of these relationships is dependent on the domain in question, spatial relationships need to be either precise or imprecise. Certain domains, such as architectural design, require certain kinds of images, such as drawings, to contain precise spatial relationships, such as the exact metric distance between two objects. Users of other domains are simply not interested in precise spatial relationships. In these domains, the conversion of a certain spatial relationship, such as *is-far-from*, into some metric value will depend on what constitutes common sense in that domain. *is-far-from* in circuit design can probably be thought of in terms of centimetres, while it is more logically expressed in terms of dozens of metres in interior design. A model of spatial relationships such as one proposed by Güesgen 1998 [16] may be appropriate for use in this image model.

2.6.2 Interaction Relationships

Another common form of relationships in image queries represents interaction among certain objects in images. For example, one may be interested in locating images of two men fighting each other, or images of an African performer playing a bongo. As these kinds of relationships may exist among more than two objects, some form of syntactical constraints need to be introduced to avoid the kind of syntactical ambiguity found in sentences such as *Monkeys beat up cows with roses*. In the current model, we propose a form of syntactical constraint based on a structured description proposed by Tam and Leung 2001 [33]. All interaction relationships can be expressed as 4-tuples of Agent-Action-Recipient-Object.

Table 1. Formal rules of the semantic model for the image

Element	Rule
x	y | *instance*(s,i) | *instance*(s) | *instance*(i)
y	i | s
id(x)	UNIQUE-ID
S	∑s
I	∑i
i	∑*metadata*(i) ∑*feature*(i) ∑*instance*(s,i) ∑*rel*(i)
metadata(i)	*metadata_time*(i) | *metadata_location*(i) | *metadata_technical*(i) | *metadata_origin*(i)
metadata_time(i)	(`metadata-time`[:METADATA-SUBCATEGORY] $\mathrm{TERM}_{\mathrm{TEMPORAL}}$*)
metadata_location(i)	(`metadata-location`[:METADATA-SUBCATEGORY] $\mathrm{TERM}_{\mathrm{GEOGRAPHIC}}$)
metadata_technical(i)	(`metadata-technical`[:METADATA-SUBCATEGORY] $\mathrm{TERM}_{\mathrm{TECHNICAL}}$)
metadata_origin(i)	(`metadata-origin`[:METADATA-SUBCATEGORY] $\mathrm{TERM}_{\mathrm{ORIGIN}}$)
feature(x)	*feature1*(x) | *feature2*(x) | *feature3*(x) | *feature4*(x)
feature1(x)	(`feature-1:`PRIMITIVE-PROPERTY $\mathrm{TERM}_{\mathrm{PRIMITIVE}}$ | NUMERIC)
feature2(x)	(`feature-2:`DERIVED-PROPERTY $\mathrm{TERM}_{\mathrm{DERIVED}}$ | NUMERIC)
feature3(x)	(`feature-3` $\mathrm{TERM}_{\mathrm{TOPICAL}}$ | *id*(*instance*(,)))
feature4(x)	(`feature-4` $\mathrm{TERM}_{\mathrm{ABSTRACT}}$)
instance(s,i)	∑*feature*(*instance*(s,i))
rel(i)	*spatial-rel*(i) | *interaction-rel*(i)
spatial-rel(i)	(`rel-spatial`:$\mathrm{TERM}_{\mathrm{SPATIAL}}$ *id*($instance_a$(s,i)) *id*($instance_b$(s,i)))
interaction-rel(i)	(`rel-interaction`:$\mathrm{TERM}_{\mathrm{INTERACTION}}$ *id*($instance_a$(s,i)) *id*($instance_b$(s,i)) [*id*($instance_c$(s,i))])

* $\mathrm{TERM}_{\mathrm{O}}$ denotes a term taken from domain ontology O. For example, $\mathrm{TERM}_{\mathrm{GEOGRAPHIC}}$ is a term from a geographical ontology.

- An *Agent*, which is a schema instance in the semantic role of a person or thing that is the doer of an event, for example, ***girl*** *running*. An Agent is compulsory in an interaction relationship.
- An *Action*, which is a verb in the present continuous tense, indicating what the Agent is doing, for example, *running* or *sleeping*. An Action is compulsory in an interaction relationship.
- A *Recipient*, which is a schema instance, indicates the recipient (indirect object) of an action, for example, *boy giving* ***girl*** *guitar*. A Recipient is optional in an interaction relationship, such as in *child sleeping*.
- An *Object*, which is a schema instance, indicates the direct object of an action, for example, *girl holding* ***doll***. An Object is optional in an interaction relationship, such as in *Mark bullying John*.

This simple structure supports the definition of interactions among all possible schema instances in an image.

2.7 Metadata

Metadata are not exclusive to images. They are present in most forms of information media. Therefore the issues related to metadata should not be solved within the image retrieval discipline. However, as they are part of the proposed model, they are briefly discussed here for completeness purposes.

Units of information are often tagged with metadata to facilitate classification and retrieval. With images, there are certain properties that are not dependent on the content of the image itself but need to be captured as they are important to a particular domain. These may include the name of the photographer, the type of film the photo was captured on, the details of the camera equipment, the film development technique, or the geographic location where the photo was shot. In some cases, certain features can be inferred from certain metadata or vice versa, such as a certain photographer's name (a metadata field) and his topic of photography. The same situation exists between a landmark building (a topical feature) and a geographic location (a metadata field). However, it would be unnecessary to include these particular kinds of relationship explicitly, as they can be inferred from examining all the images within a certain collection. For example if all images with the *photographer* metadata of value Photographer A include a *war* feature, we can establish a relationship between Photographer A and the *war* topic in photography.

2.8 Formal Semantic Model

We propose the following formal semantic model for the image.

Let T be the collection of all textual terms used in image notations, and t_k T be a term in that collection, k {1, ... , u} where u = *size*(T).

Let D be a particular domain relevant to the image collection.

Given I as a domain-dependent image collection, i_l I is an image in that collection, and l {1, ... , w}, and w = *size*(I). As all images must be relevant to domain D, the relationship between images and the domain can be expressed through the following semiotic function:

truth(i | D) ≥ _, where _ is the truth threshold of domain D, which determines the minimum relevance value of any piece of information in the domain. A normalised value of function *truth*(i | D) = 1 signifies that image i is relevant to all kinds of information need within domain D, i.e. it should be returned for all queries. As this case never happens, 1 is only a theoretical value.

Given S as the collection of all schemata relevant to domain D, s S is a schema in that collection. As all schemata must be relevant to domain D, the relationship between schemata and the domain can be expressed through the following semiotic function:

truth(s | D) 8 _. A normalised value of *truth*(s | D) = 1 signifies that the image signifies that schema s is relevant to all queries of type *instance*(s) (there exists at least an instance of schema s) within domain D.

A schema instance can be expressed by the instantiation function *instance*(s). As ∑ s S are relevant to domain D, the collection of all instances is also relevant to domain D. This collection can be expressed as follows:

∑ $instance_p$(s) . ∑ s_q, where:

p P and P {1, ... , m}, and

q Q and Q {1, ... , n}, and

m = *size*(S), and

n = *max*(∑ *size*(*instance*(s_q))).

Metadata can be represented by the following metadata function:

$metadata_j$(i) = $_k$, where:

j J and J {1, ... , v}, and

v = ∑ $metadata_j$(i) x ∑ i_l

Level-1 features can be represented by the following feature function:

feature1(x) = $_k$, where x can be i, s or *instance*(s).

Features on other levels can be similarly defined by the functions:

feature2(x), *feature3*(x), and *feature4*(x).

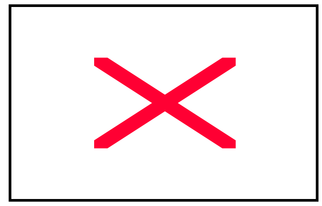
Relationships can be represented by relationship functions of the form *rel*(y_1, y_2, y_3) where y_1, y_2, and y_3 can be either s or *instance*(s).

Thus, an image can be seen as an aggregate of its metadata, features, instances and relationships:

i = ∑*metadata*(i) ∑*feature*(i) ∑*instance*(s,i) ∑*rel*(i)

Table 1 summarises the formal rules of this model.

Figure 1. A formal semantic representation



```
(image
        (metadata-origin:photographer "Richard Mills")
        (metadata-technical:camera "Pentax 6x7")
        (metadata-location:state-province Britanny)
        (metadata-location:country France)
        (feature-1:dominant-colour 0x32FE11)
        (feature-2:orientation horizontal)
        (feature-2:light-direction right)
        (feature-3 #ID1001)
        (feature-3 #ID1002)
        (feature-3 #ID1003)
        (rel-spatial:near #ID1001 #ID1002)
        (rel-spatial:right-of #ID1001 #ID1003)
        (rel-interaction:riding #ID1001 #ID1002)
        (rel-interaction:riding #ID1003 #ID1004)
        (rel-interaction:riding #ID1005 #ID1006)
)
(schema
        (id #SID1001)
        (feature-3 person)
)
(instance
        (id #ID1001)
        (instance-of #SID1001)
        (feature-3 gender-male)
)
(instance
        (id #ID1003)
        (instance-of #SID1001)
        (feature-3 gender-female)
)
```

2.9 Example of Model

There are different ways to relate all the descriptors together in a single model of the image. These include extensible markup languages such as XML, or description logics. We do not believe that a particular representation language is more suitable than others, but as logical reasoning will form part of the image retrieval algorithm based on this model, we use a form of description logics for descriptors to make them more suitable to perform inference functions with in the future.

Figure 1 contains an example which illustrates the model presented in the previous section.

3. IMAGE RETRIEVAL PROCESS AND FORMAL SEMANTIC MODEL FOR IMAGE UNDERSTANDING

This section discusses work in progress and is included here for completeness purposes. It aims to put the model proposed in the previous section into the overall context and explains how it can be used in the retrieval process.

3.1 Information Retrieval as an Interactive Process

Conventional information retrieval paradigms assume that it is the searcher's responsibility to properly communicate the information need to the information retrieval system. This assumption has led to construction of retrieval systems as depicted in Figure 2, in which the smilie symbol represents the searcher and the computer symbol represents the retrieval system.

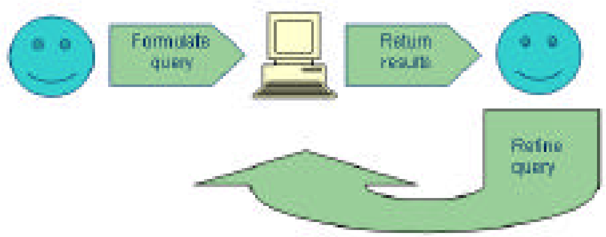


Figure 2. Conventional image retrieval paradigm

The key feature to note in these systems is that the system assumes that the query already contains sufficient information for search to take place. There are issues even with relevance feedback mechanisms where the searcher is asked to refine the choices after being presented with the first set of results. Searchers are usually asked to refine results when their queries result in too many matches. Most users may be, in fact, reluctant to do this because they are already overwhelmed by the amount of information returned by the first search attempt. Another issue is that if the first set of results being returned is deemed of little relevance to the user's information need, it may be perceived as a negative experience for the user and may, thus, dissuade the user from further interaction with the system.

It is quite obvious when analysing information retrieval scenarios between humans that the above model is rather unnatural and awkward. Let's consider a non-electronic environment, a non-electronic library, where patrons call upon the assistance of reference librarians to locate a certain publication on a certain topic. Slavens [30] edited a collection of reference interviews and questions that were manually collected in real libraries. Scrutinising this collection reveals a range of information needs and different methods that librarians used to fulfill an information need. Table 2 outlines a reference interview between a librarian (L) and a patron (P). These roles can be mapped to a typical image retrieval scenario where the user P issues queries and the system L responds with matches deemed relevant to these queries.

From the above reference interview, we have reached the following observations:

- Queries contain certain assumptions that are not necessarily correct. A good retrieval strategy should allow for the detection of assumptions and a method to clarify them if possible.

Table 2. Semantic analysis of a reference interview

Question/Answer in Sequential Order	Analysis of Information Flow
P: Do you have a history encyclopaedia?	P makes an assumption that the topic is historically related. Queries can contain assumptions that are not necessarily correct
L: What was it that you had in mind to look up?	Good strategy to clarify assumptions.
P: I need a chapter on the Gold Rush.	Query contains semantic ambiguity – at least on geographical and temporal levels. "Chapter" may tell the system something about the amount of information required.
L: In the United States?	L forms an appropriate question to help clarify semantic ambiguity related to geographic location.
P: Yes.	Query *clarity* is improved through geographic disambiguation.
L: Do you know when it happened? There were several I think.	L forms an appropriate question to help clarify semantic ambiguity related to time period.
P: The one I want was around 1848.	Query *clarity* is improved again through temporal disambiguation.
L: Is there any other word we could use for Gold Rush?	L tries to extract more information in attempt to improve query *quality*.
P: Sometimes they're called the Forty-Niners. Where would you find this information?	P is able to provide a crucial piece of information in "the Forty-Niners" which may help retrieval precision. The last question, though, may signal a certain level of impatience from P, probably due to being asked too many questions.

- Queries can be semantically ambiguous. These ambiguities can be multi-fold. A good retrieval strategy should allow for the detection of semantic *clashes* and a method to resolve them as early in the process as possible.
- A good retrieval strategy should *encourage* users to build more quality into queries even before attempting search.
- A good retrieval strategy is a collaborative process where the system and the user both contribute to the task of locating the relevant information.

In attempting to engage the searcher in clarifying semantic ambiguities in query terms, the Getty Image Search system [11] reflects this more natural retrieval model to some extent.

Based on this observation, we propose that query analysis should be fundamental to effective information retrieval models in general, and image retrieval models in particular. In Figure 3, we propose a more natural paradigm for image retrieval.

In Figure 3, the "Ambiguous?" conditional can be determined by a function that measures ambiguity and checks if the

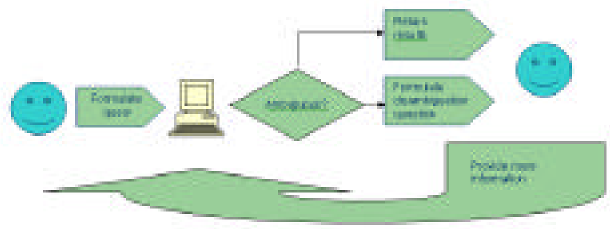


Figure 3. Interactive image retrieval paradigm

ambiguity in the current query is below a certain threshold. Relevant ideas may need to be extracted from the development in the field of query clarity [5] in order to develop this function.

3.2 Image Retrieval as a Reverse Function of the Visual Perception Process

In Figure 4, the features depicted in the right hand frame (perceived shapes and colours) map directly onto level-1

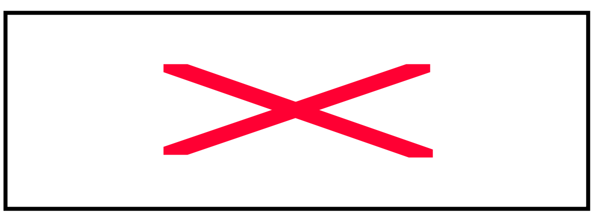


Figure 4. Perception of object shapes

features in the image model. Generalising on this observation, we can argue that the image retrieval process is a reverse function of the visual perception process.

Building an image retrieval system therefore implies building perception and reverse-perception functions for all levels of features in the image model. We call the collection of these functions the *formal semantic model for image understanding*. In order to build this model, we will need to examine how features can be computed from other level features. This is where work such as one proposed by Eidenberger and Breiteneder 2002 [10] (computing level-2 features from level-1 features) and one proposed by Do and Tam 2004 [7] (computing level-4 features from level-3 features) fit in. Through empirical analysis [7,8], we have demonstrated that domain-specific text corpora contain knowledge that can be used in improving image retrieval performance, and that logical reasoning using common sense and domain-specific knowledge bases may be used as a method to infer new features given existing features.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a formal semantic model of the image. For this model to be used effectively in image retrieval, a formal semantic model for image understanding also needs to be developed. Part of this model is the definition of a correspondence function which computes the semiotic gap between instances of the image model and the image understanding model. This particular model will be presented in another paper in the near future.

Related to the image model itself, we have only outlined the concepts and the formal rules for image representations. We have not considered the specification of specific ontologies that are suitable for use with the model. This consideration and, ideally, the recommendation of specific ontologies, if they exist, is part of the ongoing research. These ontologies will need to be structured around a common knowledge framework, such as one based on elementary typics of knowledge proposed by Gudwin and Gomide 1997 [15]. Johnson 1987 [19] classified image schemata into different types such as center-periphery, containment, part-whole, and verticality. This system of classification may be used to make schema instantiations more specific.

Upon completion of both models, experiments will be conducted with different image collections to validate our approach. The authors also expect that the existence of these two formal models may facilitate the creation of a formal image retrieval benchmarking framework for both CBIR and text-based techniques.

5. ACKNOWLEDGMENTS

The authors acknowledge the assistance of Dr. Ron Gallagher and Lonely Planet Pty. Ltd. in providing them with sample images and guidebook data that are used in the experiments referred to in this work and in other related works conducted by the authors.

All photographs used in this paper, unless otherwise specified, are obtained with permission from Lonely Planet Images.

6. REFERENCES

[1] Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D. M., and Jordan, M. I., Matching words and pictures. In *Journal of Machine Learning Research 3*, pp.1107-1135, 2003.

[2] Barthes, R. *Elements of Semiology* (trans. Annette Lavers and Colin Smith). London: Jonathan Cape, 1967.

[3] Bloomer, C. B. *Principles of Visual Perception*. The Herbert Press, 2nd edition, p.211, 1990.

[4] Boothe, R. G. *Perception of the Visual Environment*. Springer-Verlag, New York, p.13, 2002.

[5] Cronen-Townsend, S., and Croft, B. Quantifying query ambiguity. In *Proceedings of the Conference on Human Language Technology (HLT)*, 2002.

[6] Do, D. *Building High-Level Semantics into Lonely Planet's Image Collections*. Industry Presentation, July 2004.

[7] Do, D., and Tam, A. M. The context dilemma and observations on using vocabulary constraints in image retrieval. *Paper submitted to CIKM*, 2004.

[8] Do, D., and Tam, A. M. Querying compound concepts in image collections. In *Proceedings of the International Conference on Intelligent Agents, Web Technology and Internet Commerce* (*IAWTIC*), Gold Coast, Australia, July 2004.

[9] Eakins, J. P., and Graham, M. E. *Content-based Image Retrieval: A report to the JISC Technology Applications Programme*. Institute for Image Data Research, University of Northumbria at Newcastle, January 1999.

[10] Eidenberger, H., and Breiteneder, C. Semantic feature layers in content-based image retrieval: implementation of human-world features. In *Proc. ICARCV*, 2002.

[11] *Getty Images*. http://www.gettyimages.com. Last accessed April 2004.

[12] *Google*. http://www.google.com. Last accessed August 2004.

[13] Grosky, W. I., and Stanchev, P. L. An image data model. In *Advances of Visual Information Systems, Visual 2000, 4th International Conference, Lyon, France*, pp.14-25, 2000.

[14] Gudivada, V. N., and Raghavan, V. V. Content-based Image Retrieval Systems. In *IEEE Computer, volume 28, issue 9*, pp.18-22, September 1995.

[15] Gudwin, R. R., and Gomide, F. A. C. An approach to computational semiotics. In *Proceedings of the 1997 International Conference on Intelligent Systems and Semiotics*, NIST Special Publication 918, pp. 467-470, 1997.

[16] Güesgen, H. W. Spatial reasoning based on Allen's temporal logic. In *International Computer Science Institute Technical Report TR-89-049*, 1998.

[17] Haaparanta, L. Frege's context principle. In *Communication and Cognition 18*, p.81-94, 1985.

[18] I'Anson, R. *Travel Photography: A Guide to Taking Better Pictures*. Lonely Planet Publications, 1st edition, pp.100-204, October 2000.

[19] Johnson, M. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press, 1987.

[20] Kant, I. *et al*. C*ritique of Pure Reason*. Hackett Publishing Company, December 1, 1996.

[21] Lakoff, G. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago, pp. 276-277, 1987.

[22] Lenat, D. B., Guha, R. V., Pittman, K., Pratt, D., and Shepherd, M. Cyc: Toward programs with common sense. In *Communications of ACM, vol. 33, no. 8*, August 1990.

[23] Li, M., Chen, Z., and Zhang, H.-J. Statistical correlation analysis in image retrieval. In *Pattern Recognition 35*, pp.2687-2693, 2002.

[24] *Lonely Planet Image Collection*. http://www.lonelyplanetimages.com. Last accessed June 2004.

[25] Loos, E. E., Anderson, S., Day., D. H. Jr., Jordan, P. C., and Wingate, J. D. (eds.). *Glossary of Linguistic Terms*. Extract from the LinguaLinks Library, SIL International, CD-ROM version 5.0, 2003.

[26] Müller, E., Müller, W., Marchand-Maillet, S., Squire. D, and Pun T. A web-based evaluation system for content-based image retrieval. In *Proc. ACM Multimedia Workshop on Multimedia Information Retrieval (Ottawa, Canada)*, pp.50-54, October 2001.

[27] Mechkour, M. A multifacet formal image model for information retrieval. In *Proceedings of the Final Workshop on Multimedia Information Retrieval (Miro '95)*, Glasgow, Scotland, 18-20 September 1995.

[28] Meghini, C., Sebastiani, F., and Straccia, U. A model of multimedia information retrieval. In *Journal of the ACM (JACM), volume 18, issue 5*, pp.909-970, September 2001.

[29] Rossotti, H. *Colour*. Penguin Books, Great Britain, p.92, 1983.

[30] Slavens, T. P. (ed.). *Reference Interviews and Questions*. Michigan University: Campus Publishers, 3rd Ed., no dates noted for this edition.

[31] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. Content-based image retrieval at the end of the early years. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, December 2000.

[32] Squire, D. M., and Pun, T. Assessing agreement between human and machine clusterings of image databases. In *Pattern Recognition, vol. 31-32*, pp.1905-1919, 1998.

[33] Tam, A. M., and Leung, C. H. C. Structured natural-language description for semantic content retrieval. In *J. American Soc. Information Science*, pp. 930-937, September 2001.

[34] *TREC Tracks*. http://trec.nist.gov/tracks.html. Last accessed August 2004.

[35] Wollheim, R. On the assimilation of pictorial art to language. In *Towards a Theory of the Image* (ed. Jon Thompson), Jan van Eyck Akademie, Maastricht, p.27, 1996.

[36] Zhao, R., and Grosky, W. I. From features to semantics: some preliminary results. In *Proc. International Conference on Multimedia and Expo*, 2000.

Visual Blends: A Computational System Exploring Digital Creative Spaces

Theresa Gartland-Jones
Kingston University London
Faculty of Art, Design & Music
theresa@atgj.org

ABSTRACT

The focus in this paper is the application of evolutionary computing processes in the generation of artworks. I shall ask questions such as, to what extent do artists re-interpret their practice through the context of programming languages, theories of artificial life, or the metaphors they use. Reference will be made to models of creativity, for example, conceptual blending, which may inform our understanding of art systems that evolve over time, exhibit emergent behaviour and how semantic and syntactic issues arise from such processes.

Reference will be made to the development of an autonomous, evolutionary drawing system which combines traditional art studio practice (drawing, painting, montage) with computational computer processes. An important aspect of this research is in defining the high level concepts by which the system carries out artistically significant analogies. The paper will discuss how metaphor, analogy or similarity is relevant to defining relationships between the domains (classification of images) in a computational, emergent system.

Keywords

Art, Evolutionary computing, Semantics, Metaphor.

CONTEXT

In 1966, the Experiments in Art and Technology group was founded in New York by Billy Kluver who stated: "That the goal from the beginning was to provide new materials for artists in the form of technology" [1]. In 1968, an exhibition called 'Cybernetic Serendipity', at the ICA, focused on the ideas and technologies that linked artists and scientists and how the notion of open-ended computer behaviours might lead to unexpected art outcomes generated by machine systems. More recently, books such as "Metacreation: Art and Artificial Life" by Mitchell Whitelaw have focused on ALife (AL) artists. Eduardo Reck Miranda illustrates how he combines the modeling of both physiology and abstract musical ideas: "Alife is a discipline that studies natural living systems by simulating some of their biological aspects."[2] "Information Arts" by Stephen Wilson gives a summary of some of the many types of interdisciplinary creative (computer / art) work from the robotics of Ken Rinaldo, the intersection between the physical body and technology in the performances by Stelarc, to the visual computer generated algorithms of Karl Simms. The research outlines many ways in which a computer can be used to create art from utilizing the bounded components of Photoshop to devising personal metaphors and syntax in an idiosyncratic software design.

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

An idiosyncratic approach towards computer art making was adopted by the artist Harold Cohen about thirty years ago at a time when pre-existing software interfaces did not exist. He asked questions about his perceptual and cognitive art practice and how it might be computationally modelled. He wove traditional art practice, computer science and the semantics of code with philosophy to produce an '*investigative tool*' called Aaron, using Artificial Intelligence processes. "Artificial intelligence (AI) is a cloth woven from three academic disciplines – psychology (cognitive modelling), philosophy (philosophy of the mind), and computer science – with further strands from linguistics, mathematics, and logic". [3]

Aaron draws and paints and is a knowledge–based program where decisions are represented in rule form based on pre-defined boundaries that model, specifically, representational drawing. Aaron selects which subject to paint from the internal symbolic model that has been provided by Cohen (i.e.: a portrait or a still-life). The system's hierarchical descriptions include direct references to making a two dimensional image with semantics that refer to *PLANNING* the picture and planning involves *LINES* and line refers to *CURVES* and *SECTORS* that signify potential pathways for the line to take. Cohen is using familiar visual language components so that:"The encoding and decoding of messages requires access to the same code-book by the image-maker and the image-reader."[4] Cohen states how he is not making art but, exploring for example the boundaries between lines having meaning or not. He uses the term '*standing-for-ness*' instead of: "words like "symbol", "referent", "metaphor", "sign", and so on: words which abound in art theory and art history. An image, I have said, is something which stands for something else."[4] However, by using a familiar coded representation or model of the world, Cohen is, to a certain extent limiting the boundaries of potential meaning to be explored in his investigation and in the audience mediation.

ALife (AL) differs from the knowledge–based approach because it has rules that do not hold a model of the real world, instead, these systems are based on complex interaction of objects in the system. Jon McCormack states: "Many Alife simulations make the distinction between genotype and phenotype. The genotype is commonly referred to as 'code' that creates the phenotype in an interpretation analogous to the view of DNA as a 'code'. [5]. The phenotype is the visual form that the viewer can see on the screen and it is resolved partly by the code structure that defines it and as a direct result of the dynamic interaction over time with other forms in the artwork. This evolutionary approach enables potential for greater emergent behavior in an open-ended system design that allies creativity with life. But, does this mean that an AL computer / artist collaboration is any less bounded in the '*code-book*' of image-making and mediated communication of meaning than a Knowledge-based system such as Aaron?

VISUAL BLENDING OF MEANING

By studying the metaphors used by AI/AL artists I shall discuss the visual ideas that are communicated and in what ways they are bounded in meaning and creative value.

Arthur Koestler, a notable writer on creative processes, quotes Karl Mach who describes an analogy for relaying the idea of being aware of the solution before the problem, similar to Aaron having an image at the outset of the drawing: "The subject who wishes for a tree to be laid across a stream to enable him to cross it, imagines in fact the problem as already solved….he proceeds from the target-situation to the given situation, along a road he will re-trace. [6] The point here is that the search process for solutions to creative problems may be enhanced by using methods that are less bounded by prescriptive rules. Evolutionary computer algorithms may potentially offer artists more flexible search processes and emergent creative behaviour.

The artist and historian John Willats draws on an interesting analogy between language and visual art. Willats refers to modern linguistics and two definitions defined by Fernand de Saussure who stated: "Synchronic linguistics studies the structure or "state" of a language as it exists at a given time within some particular group; while diachronic linguistics studies the changes in language over time." [7] Aaron has information for constructing representational form, adopting a Synchronic approach that represents a particular style. AL is possibly analogous to diachronic principles, because a system evolves over time and therefore adapts to novel developments / interactions.

Cognitive processes are such that: "The solution does not proceed in a single line from target to starting point or vice-versa, but by a branching out of hypotheses – of possible strategies – from one end, or ends, until one or several branches meet. [7] In AL art, the systems typically do not begin with a pictorial model at the outset, so they have a greater emergent potential embedded in their code design. Contemporary languages allow for detailed class interfaces for individual objects. For instance, Java supports the dynamic interaction of objects as containers for ideas or the code structure for a genotype as in AL.

This in theory should facilitate a greater ability for the artist / computer collaboration to be more adaptable to potential combinations of ideas and aims. The artist and writer Paul Brown wrote: "The artistic mind is a 'butterfly' mind that can fly from flower to flower, from source to source, with little respect for logic or scholarship". Therefore: "Theory does not (necessarily) inform creation, although creation, of necessity, informs theory". [8]

Koestler refers to 'stepping stones for thoughts' which are the spaces where the artist compels the audience to exert its imagination. The suggestion here is that creative computational art systems should leave room for interpolation both during the execution of the algorithms and in the perception of the audience. For a system to be about creativity then maybe greater consideration needs to be given to how ideas or metaphor are explored. Lakoff states: "The essence of metaphor is understanding and experiencing one kind of thing in terms of another" [9]

There are many examples where visual metaphors for AL art are grounded in forms derived from molecular structures and biological cells and then united with scientific theories. The theories have developed software models for genetic coding, particle systems, or Embryogeny as in biology which is used in computer science and art for mapping genotypes onto phenotypes. In Biotica, Richard Brown uses biological semantics for his cell structures called 'Bions' which are like particles, with rules regarding their DNA, which are the genotypes expressed in cell-like phenotypes. The inspiration for both the art and science of 'Biotica' included an idea common to AL art which is Autopoiesis systems theory developed in 1970's by Humberto Maturana and Francisco Varela. It explores the notion of how living systems co-exist in a fine balance between the order and chaos modelling of life and cognition. A computer model called Cellular Automata (CA) is based on the Game of Life by Dawkins and manifests itself in the work of Paul Brown's 'Chromos' and Jon McCormack's Eden World projection, both of whom use patterns similar to Islamic tiles as an aesthetic expression of CA. For McCormack it is a lattice work populated by his sonic agents, rock and bio-mass but, the tiles are a means of metaphorically engaging with an alternative aesthetic than the common theme of biological forms in AL: "These minimalist geometric textures suggest abstract landscapes rather than the iconic or literal individual representations that are common in many Artificial Life simulations. "[10]

So, even artists using AL principles perceive limitations in their methods and acknowledge sacrifices in the aesthetic appearance of the visual forms and the potential for emergent meaning from the separation of genotype and phenotype which is not a natural state in nature. But, the design of dynamic interactions in an autonomous assessment process is considered to be a closer parallel to life and possibly the notion of creativity than more determined computer processes. McCormack states that: "What we would like is a system that combines the ability to subjectively evolve towards phenotypes that people find' interesting' without the bottleneck and selection problems inherent in aesthetic evolution".[10] Possibly the processes employed limit visual and metaphorical communication in AL work which are the elements that compel an:

"audience to exert its imagination. " [6]

Conceptual Blending (Fauconnier and Turner, 1990's) offers a model of how 'mental spaces' or concepts are dynamically mixed. Many of the ideas can in principle be linked to Lakoff and Johnson (1980) and their work on metaphor. Together, these writers explore both the cognitive and physical nature of metaphor. They discuss how we associate concepts and how these evolve culturally and physically as in the work of The Neural Theory of Language by Srinivas Narayanan (1997). Lakoff writes about the multiple layering or interaction of concepts which is considered to be necessary for creative thought: "even our deepest and most abiding concepts – time, events, causation, morality, and mind itself – are understood and reasoned about via multiple metaphors. In each case, one conceptual domain (say, time) is reasoned about, as well as talked about, in terms of the conceptual structure of another domain (say, space)". [9]

As an artist, evolutionary computational processes offer the possibility to define idiosyncratic, dynamic layering of concepts in the form of objects within a system. The resulting cause and effect of blended, 'mental spaces' results in emergent visual ideas. Lakoff and Johnson state that new ways of understanding or being creative / novel / innovative are: "not miraculous; they do not come out of nowhere. They are built using the tools of everyday metaphorical thought, as well as other commonplace conceptual mechanisms." [9]

When studying for example Synthetic Cubism, there is a synthesis of commonplace and cultural objects and concepts which when united (blended in a generic space) they inferred new meanings / genre in art. The 'Guitar' (1912 -13) by Picasso blends painting and sculpture in one space, mixing the two genres and found materials associated with everyday objects and then placing them in the abstract context of art. The juxtaposition of the commonplace and exploring boundaries of established disciplines is part of creative practice but, in this context it resulted in a novel approach. It has contributed to the cultural re-appraisal of the boundaries between two and three dimensional form and process in art. Margaret Boden discusses 'conceptual spaces' [11] as generative systems which may contain the rules of chess or at the turn of 20th century, the rules of painting and sculpture. Picasso's tweaking of these spaces may be considered to have resulted in a 'transformational' creative act because he altered the rules for painting. Therefore, the potential for real emergence must depend on the design of systems that enable the cross-fertilization of ideas or concepts as opposed to pre-determined structures that set boundaries for generated forms or image-making.

In an AL piece such as Life Spacies (1997 – 99) by Christa Sommerer and Laurent Mignonneau, a form of '*counterfactual blending*' (Tony Veale: 'Creativity as Pastiche: A Computational Treatment of Metaphoric Blends with Special Regard to Cinematic Borrowing') occurs, meaning a re-visiting of a familiar theme but interpreted in a new context without altering the general structure of our perception of living systems. Tony Veale gives the example of Romeo and Juliet revisited in the form of Westside Story. The premise of Life Spacies is the evolution of an ecosystem as analogous to 'Art as a Living System' which represents the dynamic of the audience/user and the system interaction. So, to what extent is this theme inferred visually and semantically? Living systems or evolutionary processes are bound in Victorian interpretations of survival of the fittest and more recent visualizations of bacterial growth or film footage of plants sprouting from seed to full form in five filmed seconds. The structures of this theme are not challenged but bounded by the evolution of phenotypes that are directly associated with abstracted organic creatures / plant-like forms. A direct semantic link between nature and code are encapsulated in the work of William Latham that began as the paper based *FormSynth* drawings but, later formed the foundation for *Mutator,* a generative computer system. His interest in biological forms resulted in works whose titles such as "*Mutations"* appropriately conjured the mutation and evolution of forms derived from nature. The Mutator interface allows us to 'Marry' forms, to 'Breed' them and the code has containers such as 'LOBSTER' structures who live out a life cycle of birth, 'matures', 'decay' and 'dies'. This is designed on an Interactive Genetic Algorithm basis and therefore retains considerable control over the aesthetic outcome of the work. The difficulty in an autonomous system is controlling the aesthetic assessment process and still facilitating the layering of ideas and diverse visual outcomes.

Casey Reas's at Ars Electronica 2003 discussed the importance of analogy and versatility of varieties of images being animated in an art work: "You can sort of move to different kind of spaces within time by selecting images in a different way. Computational it's not very interesting at all but by carefully selecting certain images and cropping them in certain ways I create different rhythms and patterns in time. The code works on any image but I think it is only significant if it is extremely carefully selected." [12] So, the value of this statement for me is in the idea that code can result in generic application of code semantics to different kinds of image input. The higher level definition for fitness in Reas's autonomous systems enables the flexibility for adjusting to diverse subject content. He suggests a conceptual design that facilitates the separation of the underlying image input from the visual drawing behaviour in the animation MicroImage'. Therefore, allowing flexibility to accommodate subject/content changes.

A COMPUTATIONAL SYSTEM

The computational system under discussion in this section is autonomous, evolutionary drawing software which combines traditional art studio practice (drawing, painting, montage) with evolutionary computer processes. The computer / art collaboration begins with the physical, studio practice where images are made. The images are input into the Map Conversion Interface of the software as shown in Figure 1. The images are converted to Map objects as illustrated by the image on the right in Figure 1. All these images are held and categorized in the system as the foundation drawings for the animation. The Map Layer (Figure 2 on the right) evolves from one group of drawings to the next, selecting images from the groups and applying fitness criteria to their animation.

However, this evolving Map Layer remains invisible to the audience who see only the Animation Layer as shown in Figure 2 (on the left). The image on the right shows the system selecting and evolving between Map objects which include in this example, the Portrait and abstracted GasNets image categories. The drawing aesthetics are separate from the Map Layer but, they are semantically linked through the system design and syntactical code descriptions.

In order to get closer to the '*essence of metaphor'* I have taken a similar approach as suggested in the context of Reas, where the Genetic Algorithm powers the selection of images in the Map Layer and this layer is separated from the expressive drawing behaviours. This enables a generic means of working with diverse image content and drawing expression. Similarly, in traditional studio practice one can hold in the mind for example, two entities, one that relates to the image subject and its associated ideas. The other is what expressive and conceptual decisions you are making in order to communicate that image.

The groups of images are intended to be classified so that they contain a prototype for that group such as portrait but, there are sub-types that may be semantically linked to thumb print images or text conceptually reinforcing the idea of portraiture. The developing rationale for the structure of these ideas has partly been informed by George Lakoff and his work on '*categorization'* as a means of comprehending our experiences and ideas. This in turn is supported by using object orientated programming methods which enable groups of objects and events such as drawing behaviours to interact with each other. I employ genetic algorithms to achieve fitness assessments for the animation cycles. The process is influenced by books such as "the Blind Watchmaker" by Richard Dawkins who proposed "biomorphs" that held code which represented the DNA for forms. However, unlike AL outcomes, I have not absorbed the ideas and externalized them using the visual metaphor of organic forms derived from nature.

In "Evolving Line Drawings", Ellie Baker and Margo Seltzer describe "Drawing Evolver": "a system that mates or mutates drawings selected by the user to create a new generation of drawings. The initial population from which the user makes selections may be generated randomly or input manually. The process of selection and procreation is repeated many times to evolve a drawing." [13]

Their paper references art systems such as those produced by Karl Sims, Latham and Todd. Sims provides a genetic code for defining colour or form. When he runs his computer program, the DNA of the images is altered and takes on a new appearance. The user then decides which are the most successful to include in the next generation. When you run the code and achieve the fitness or the rules/analogies that you stated, then the system is ready to mutate the next generation of bitmaps

From Ellie Baker's model I have taken the principal of including two drawings, or as I specify them, a starting drawing and a target drawing. The images are converted into Map objects in the Map Conversion Layer, which are then mutated in the Map Layer, producing populations to which a fitness measure is applied. The fitness measure may be the direct comparison of the number of black and white pixels in the starting and target drawings. However, when the fitness is reached (i.e. the starting drawing evolves towards the target drawing), I do not then interact with the system in order to select the next generation of drawings or to direct the path or search journey of the animation. The systems mentioned above depend on aesthetic decisions to be made by the user where as this research is developing autonomous ways of making these choices. Therefore, there has to be suitable high level descriptions that result in successful evolutions between image concepts regarding their classification (grouping) and drawing behaviours.

The Animation layer shown in Figure 2 on the left is what the audience actually sees on screen or as a projection. The Map Layer shown in Figure 1 remains invisible, it is there as a topological layer for the lines to map themselves to or propel themselves from. The potential of this design means that I have the ability to build explorative relationships between the drawing Animation Layer and the Map Layer which evolves the images I have input into the system. Should the animation reach 100% fitness for the algorithm, then it would in this scenario take on the form of the face before continuing with the selection process. At about 60 % one can begin to see the face emerging through the drawing as shown in Figure 2. However, similarity assessments are continuously being made based partly on group similarity descriptions. The drawings may break off and select another face to evolve towards or make the similarity decision based on visual drawing expression. A new target image may have similar drawing descriptions or similar visual behaviours which have emerged during the process of animation.

CONCLUSION

The creative interest for me is in the journey that the drawing takes during the animation. But, the challenge is in the descriptions that are defined in order for the animation to successfully reach fitness or break at a particular point. Gombrich presents a drawing in 'Art & Illusion' which shows a '*Rabbit or Duck?'* which is not, fully one or the other, yet our perception allows us to subtly perceive both animals in the same drawing. The visual play of meaning and content in this way is partly what may evolve in my system and I have to semantically tag images, so that they make successful, aesthetic and conceptual transitions.

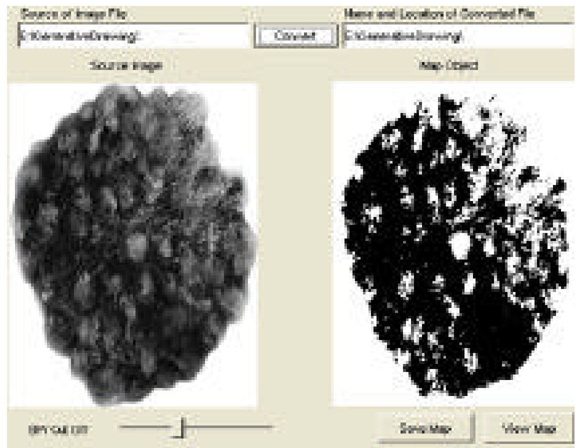


Figure 1. Drawing input and conversion to a Map object in the Map Conversion Interface.

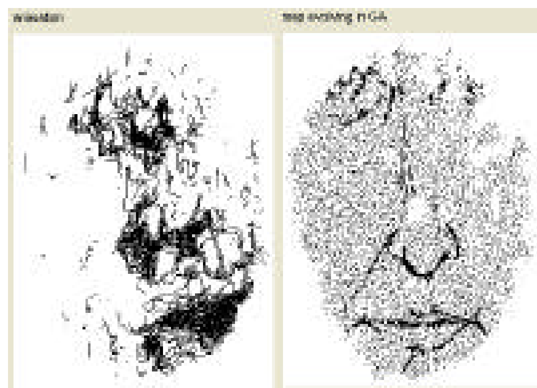


Figure 2. A 60 % evolved Animation Layer screen shot (on the left). The start drawing (as above left in Fig.1) and a portrait image Map object evolving in the Map Layer (on the right).

REFERENCES

[1] Candy, L. and Edmonds, E. Explorations in Art and Technology, Springer, Great Britain, 2002.

[2] E.R Miranda, Composing Music with Computers, Focal Press, 2001.

[3] M. Sharples,D. Hogg, C. Hutchinson, S. Torrance, D. Young, Computers And Thought ; A Practical Introduction to Artificial Intelligence, ; The MIT Press, Massachusetts, 1994.

[4] Cohen,H. "What Is An Image?", http://crca.ucsd.edu/~hcohen/, 1979.

[5] McCormack, J. and Dorin,A. “Ways of Seeing: Visualisation of Artificial Life Environments, Beyond Fitness: Visualising Evolution, Alife 8, Smith, Bird, Bullock (eds), pp. 155-162, 2002.

[6] A. Koestler. The Act of Creation, Hutchinson & Co, London, 1964.

[7] Willats, J. Art And Representation, Princeton University Press, Princeton, New Jersey, 1997.

[8] P.J. Bentley, D.W. Corne, Creative Evolutionary Systems, Academic Press, 2002.

[9] Lakoff, G. and Johnson, M.., Metaphors We Live By, University of Chicago Press, 1980.

[10] J. McCormack, “Evolving Sonic Ecosystems”, Kybernetes: The International Journal of Systems&Cybernetics,2003,vol. 32, no. 1-2, pp.184-202(19).

[11] Boden, M. Dimensions of Creativity, The MIT Press Cambridge, Massachusetts, 1994.

[12] C. Reas, “Code. In conversation with Casey Reas”, Ars Electronica, ww.netart@Kopenhagen.dk, 2003.

[13] E. Baker and M Seltzer, “Evolving Line Drawings”, Graphics Interface ’94 Proceedings, 1994.

Bound Together: Signs and Features in Multimedia Content Representation

Edward Hartley
Computing Department
Lancaster University
United Kingdom
hartley@comp.lancs.ac.uk

ABSTRACT

This paper introduces new extensions to the semiotic model that allow the model to account for image features that characterize an audio, visual or audio-visual object. The treatment in this paper emphasizes visual content description. The framework and the associated construct of *image features* characterizing the *visual object* "binding" to *conceptual terms* used to describe the *visual object* is described and illustrated in terms of UML diagrams. Visual object identification, location and media temporal segmentation approaches are outlined. The setting construct developed by Parkes [17] is then replaced with the introduction of the existent context as the minimal unit for temporal decomposition in content description. This construct forms the basis of the content modeling for the upper level representation scheme adopted in the Automating Video Annotation (AVA) moving picture content annotation prototype tool. The analysis is supported through a worked example developed on a content sequence from an episode of children's television.

Keywords

Computational Semiotics, Multimedia Annotation, Video Annotation, Content Description, Content Representation, Visual Feature Extraction.

1. INTRODUCTION

Multimedia content representation is concerned with constructing models of that are used to facilitate content description or annotation. In general these models have adopted a bottom-up or top-down approach. The bottom-up approach is based on the analysis of still and moving picture features typified by the descriptors in MPEG-7 Visual part [10]. These features typically characterize a visual object in terms of color, texture or both using methods similar to those described in section 3. The top-down approach is typified by the content modeling portions of the MPEG-7 MDS part [12]. In contrast the approach adopted in the Automating Video Annotation (AVA) project at Lancaster is bi-directional in that it combines elements of both approaches. The integration of both approaches has needed new theoretical developments that extend the semiotic model. These new theoretical developments are based on previous revisions and extensions to the semiotic model developed by Hartley *et al.* [8].

First published at COSIGN-2004,
14 – 16 September 2004, University of Split (Croatia)

An outline of the Revised and Extended semiotic models is given in section 2 these introduce the data and description planes into the model. This is followed by a description of the new revisions to the model. These revisions enable the model to account for image features together with denotative and connotative descriptions. To simplify the analysis the emphasis in this paper is on visual media and objects, however the approach can be readily extended to the audio and multimedia domains.

Section 3 outlines an example of image processing techniques for visual object identification and location and an overview of approaches to temporal segmentation. Known limitations of these techniques are then discussed. On the basis of this analysis assumptions are made about the capabilities of these algorithms that allow an extension of the semiotic modeling from section 2 into the time domain.

In section 4 the adoption of the term existent by Chatman [4] in his semiotic analysis of narrative structure in fiction and film is described in relation to the modeling from section 2. This results in the definition of the existent context as the minimal unit of spatio-temporal segmentation under constraints determined by the capabilities of the chosen segmentation algorithms from section 3. It is then shown that the existent context set can replace the setting construct developed by Parkes [17] as the minimal unit of temporal segmentation for moving picture content description.

A qualitative worked example of the application of the model is given in section 5, which is followed by conclusions about the adoption of the approach in section 6.

2. Revising the Semiotic Model

It is well known that the semiotic analysis [5] of textual media distinguishes between the content plane and the expression plane. The content plane contains the meaning carried by the words on the page and invoked in the readers mind. Whereas the expression plane contains the words seen on the page through expressed through the print medium. The analysis then goes on to distinguish between the idea of the words denoted by the text i.e. the base meaning of the word in a given context and the ideas that are conotated by the word from associations in the readers mind. Nack [16] adopted a semiotic approach to multimedia content modeling and description in his work on the Auteur, automatic editing prototype. It was found subsequently by Hartley *et al.* [8] that whilst the semiotic approach to multimedia content modeling was valuable the distinction between content and expression could not fully account for digital multimedia content and its description. This was not due to inadequacies in the semiotic approach to content modeling but is due to limitations in the extent of the model. Revisions and extensions to the model were undertaken which this paper continues to allow features characterizing objects to be included.

2.1 Revised Semiotic Model

The revised semiotic model (see Figure 1) was developed [8] to extend the application of the semiotic approach to multimedia content. The data plane was added to the basic semiotic model to separate the information being processed for rendering or sonifying from its content and the computer based expression mechanism. The description plane was also introduced to distinguish content description from the content itself. The description and data planes are needed because content and its expression are not linked in digital multimedia in the way they are in printed textual media. The traditional semiotic premise is that the text content is the ideas invoked in the viewers mind. This premise is maintained by placing the observer adjacent to the content plane quadrant in the model.

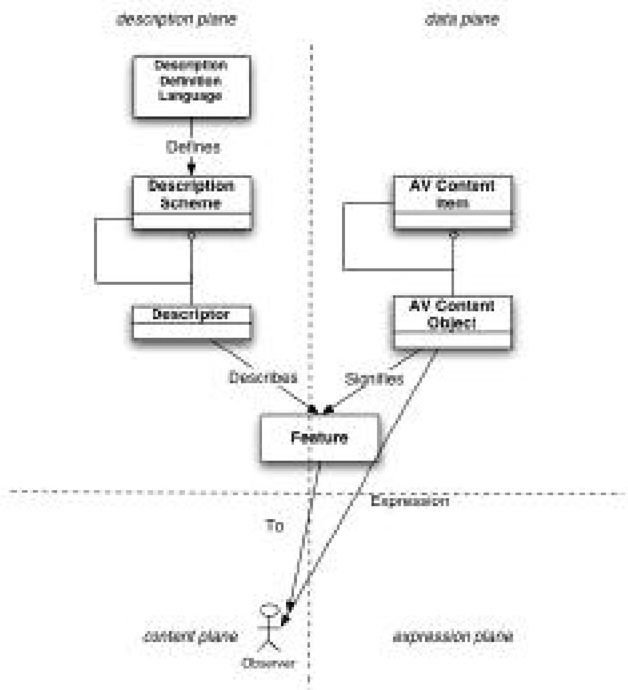


Figure 1.Revised semiotic model.

The model is useful because it provides a conceptual bridge between software modeling and a semiotic analysis of content. The model emphasizes that there is no immediate one to one relationship between the ideas invoked in the observers mind and their textual description. It can be seen that this contrasts strongly with the case of the ideas denoted and expressed by a piece of text. The model was originally developed to clarify terminology being proposed for adoption during the analysis of MPEG-7 [11] requirements. This clarification centered on exposing the terminological deficiencies in the shot and scene break terminology then being considered as the basis for temporal decomposition in MPEG-7. This points to the need to introduce mechanisms to account for the time based nature of much of multimedia content, which is undertaken in section 4.

2.2 Extended Semiotic Model

The extended semiotic model (see Figure 2) introduces the concepts of compressed data and compressed descriptions into the model. This was seen as essential in the context of multimedia content as much of the content of interest is compressed. The assumption that binary descriptions would need to taken into account in the modeling has been vindicated by the development of the Binary Format for MPEG-7 the BiM. This is MPEG-7's XML schema-based compression standard [9]. Hartley *et al.* [8] show the application of the extended model to a short segment of video content.

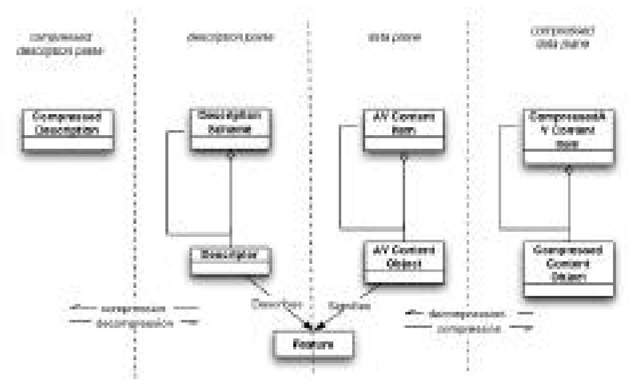


Figure 2. Extended semiotic model.

2.3 Further Revisions to the Semiotic Model

Consideration of Figures 1 and 2 leads to the conclusion that there are still deficiencies in modeling the relationships between the visual feature, the visual content object, the object description and the observer. Further revisions of the semiotic model have now been introduced to accommodate the need to distinguish: the feature sets used to *characterize* the object perceived by the viewer, the visual object itself, the concept invoked in the viewers mind and the schemata describing the visual object. In MPEG-7 [15] terminology the descriptors and description schemes describing the visual object include the features characterizing the object and the descriptive terms. The features *characterize* the visual content object at some level of decomposition of the audio-visual content item as opposed to describing the visual content object in any meaningful sense. The features characterizing an object are typically color or texture metrics. The introduction of image features into the model provides an improved framework for the description of still images. The framework and the associated construct of the image features characterizing the visual object "binding" to the conceptual terms used to describe the visual object is described and illustrated in diagrammatic form in terms of UML (see Figure 3). The model is also consistent with the earlier model. It becomes apparent that to provide an effective meaningful description of visual objects both visual features *characterizing* the visual object and meaningful terms *describing* the visual object are needed. These feature sets and other descriptive components are then combined or "bound" together either by a content describer of by the system. This introduces an important separation of concerns since it allows content modeling to be carried out independently of feature extraction. It also allows meaningful descriptions and feature sets to be instantiated independently without the binding operation being carried out. In implementation the descriptive terms corresponding to the concept denoted by the visual object, the ideas connoted to the describer and the features characterizing the object are all separate class or attribute instances.

The ability to instantiate object description and feature sets without completing the binding operation allows greater flexibility. This is one of the factors that distinguish the approach adopted in the Automating Video Annotation (AVA) project and the associated AVA demonstrator developments at Lancaster from other approaches in that the binding operation is made overt rather being implicit.

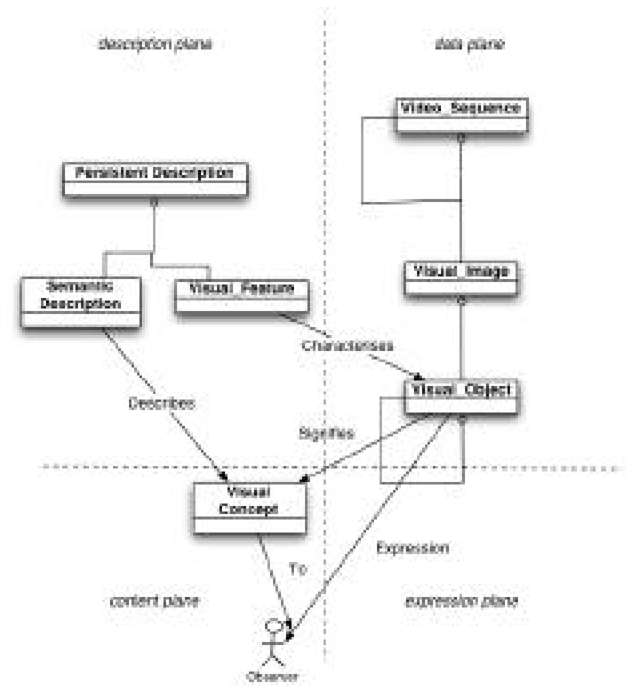


Figure 3. R2 semiotic model.

2.4 Knowledge Representation or Feature Extraction

It will be readily appreciated that there is minimal correspondence between the color and texture metrics referred to above and any meaningful classification in the knowledge engineering sense of the visual objects under consideration. This is supported by noting that only the correspondence between color terms such as red green etc. and physical colors has been investigated in any depth by Berlin [2] and that few descriptive terms exist for texture.

However the use of color and texture is now common and other metrics for visual object recognition and location in content-based retrieval applications have no common usage denotative basis. So typically these applications rely on using a query image and return images similar to the query. This highlights the arbitrary nature of the identifiers for visual objects and their association with descriptive terms when viewed from the feature space viewpoint.

2.5 Modeling Time

The modeling so far has been related to visual objects in images, which whilst allowing for decomposition of time based media into media objects has been primarily static. This is acceptable for still images however in multimedia content representation time dependent media must be modeled as well. To facilitate this modeling the distinction between separate time lines relating to content will be introduced namely; *•capture time, •expression time and •represented time.*

Content capture time is the time at which the content is recorded or captured. Content expression time is the time at which content is expressed to an observer and content represented time is the time portrayed in the content. These separate time lines are needed to allow the model to account for time-based media content. In turn the temporal relationships within each time line can be represented using interval graphs. Separating these time lines allows the entities visible to the observed at content expression time to be modeled in a way that supports Chatman's [4] distinction between process and stasis statements. These are described in section 4 after an outline of data-plane spatio-temporal decomposition techniques is given in section 3.

3. Spatio-temporal Segmentation

The bi-directional approach adopted in the AVA project combines both low-level feature extraction and high-level knowledge representation techniques. The low-level techniques that are used to provide both a temporal and spatial decomposition of moving picture data into temporal segments and visual objects will now be described. Implicit in this account is that some manual segmentation has to be carried out at some point to provide a basis for retrieval. The spatial techniques for object recognition and location are followed by an outline of the temporal techniques.

3.1 Visual Object Identification and Location

There are effectively two tasks that were identified by Swain and Ballard [19] that must be addressed that relate to segmentation of still images namely; visual object identification and location.

3.1.1 Histogram Intersection

Swain and Brown [20] described image histogram intersection as a method for visual object retrieval. More recently the technique has been investigated extensively by Smith [19] and Schiele [18] who both compare the effectiveness of extensions to the basic approach described and describe different combined color and texture approaches. An image histogram is defined (see equation 1) as an n-dimensional vector:

$$H_i(j), j = 1,...,n, \qquad (1)$$

Where n is the number of bins representing the number of grey levels or colors and *Hj* is the number of pixels in the image with the color *j*. Normalized histogram intersection provides a confidence value estimation method and is defined on a pair of histograms designated the image I and the model M each containing n bins (equation 2).

$$H(I_j, M_j) = \frac{\sum_{j=1}^{n} \min(I_j, M_j)}{\sum_{j=1}^{n} M_j} \qquad (2)$$

This method provides an approach that has been used successfully for identifying the the presence of a visual object in a given image.

3.1.2 Histogram Back Projection

Visual object location in a still image can be achieved through histogram back projection. This algorithm first computes a ratio histogram from the model histogram and the image histogram (equation 3).

$$R_i = \min\left(\frac{M_i}{I_i}, 1\right) \qquad (3)$$

This operation results in a look up table in the color space representing how much of the searched object color is present in the image. Then in *back-projection* each pixel (x-y) of color i in the original image is replaced by R_i, and resulting peaks in the distribution of values will represent the expected locations of the object in the image.

3.1.3 Limitations in Histogram Intersection and Back Projection

Schiele [18] identifies that the histogram intersection technique has good immunity to changes in scale and rotation and limited occlusion. His findings indicate that the algorithm can support changes of scale of around 4:1. Ennesser and Medioni [6] have successfully applied a modified back projection algorithm that uses weighted histogram intersection to object location in cluttered scenes. This gives improved results to those obtained using the back projection algorithm. However even in relatively simple content such as the children's TV example considered here there are changes of scale and levels of occlusion are likely to defeat the capability of these algorithms. This is the problem that the binding concept is introduced to overcome.

3.2 Temporal Segmentation

An extensive body of literature exists on shot and other film production effect detection, which is reviewed by Aas et al. [1], Brunelli et al. [3], Koprinska and Carrato [13] and Lienhart [14]. This describes a wide variety of temporal segmentation approaches that map moving picture production effects such as cuts, dissolves and wipes etc. onto metrics derived from the expression plane time distributions of the one or more image features. These features can be extracted from the data or compressed, data planes. The majority of these techniques have difficulty discriminating slow moving objects from the effects that they seek to recognize. From the point of view of automating content description this limits the usefulness of the algorithms. At present in the AVA project a simple color difference threshold based approach has been adopted.

3.3 Open Issues and Assumptions

In the remainder of this analysis the following assumptions will be made about the capabilities of spatial and temporal segmentation techniques.

3.3.1 Spatial Segmentation Assumptions

It will be assumed in rest of this paper that a pair of algorithms can be used to identify and locate a given visual object over a range of scales up to 4:1 at a high level of confidence.

3.3.2 Temporal Segmentation Assumptions

Annotation obtained from production data or one of the techniques outlined in section 3.2 above can be used to provide a *expression* time temporal segmentation. However the level at which this segmentation will correspond to a temporal segmentation of *represented time* dependent on which segments are stasis statements and which process statements in the content. The distinction between stasis and process statements is described in section 4. This is regarded as an issue for future study and it will be assumed that this is a concern for higher level modeling than the instantiation of existents and existent contexts. So the assumption is made that expression time segmentation can be achieved that has some level of correspondence with production of effects.

4. Modeling and Describing Media Objects in Time

In the subsequent discussion some terminology will be adopted from narrative theory applied to moving pictures by Chatman [4] that characterizes all the actors and objects in a narrative as existents. The introduction of this terminology allows the modeling to be extended to include time whilst still maintaining the semiotic basis of the analysis.

4.1 Describing Visual Objects

Chatman's approach is to characterize the actors, objects and scene elements in a narrative as different classes as existents. This terminology is adopted here because it provides a convenient method to identify a base class for all objects visible in an image. He also distinguishes between stasis statements in narrative process statements. Stasis statements are those that do not move the story along but state the existence of something. In contrast process statements are those that denote a change of state. In moving picture sequences both types of statement result in elapsed expression time but not necessarily elapsed narrative or *represented time* in the terminology from section 2. This distinction confirms that there are limitations in the interpretation that the behavior of the temporal segmentation algorithms and supports the assumptions outlined above. In moving picture content where the process statements are not denoted by a voice track a process statement will typically involve motion which again will be considered as an issue for higher level modeling.

4.1.1 Describing Visual Objects in Still Images

Considering only the visual object description portion of figure 3 an existent description is obtained by combining the attributes needed to completely describe a visual object in a still image. The term existent description is adopted to denote such a description.

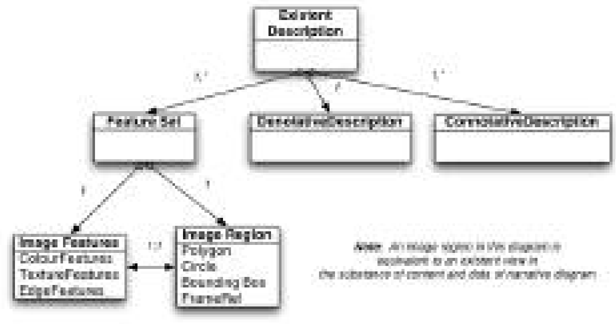


Figure 4. Existent Description.

The attributes that are needed are; (i) a feature set defining the visual features needed to identify and locate the visual object and their boundary, (ii) the objective or denotative description of the object and (iii) any additional connotative terms the describer may choose.

4.1.2 Describing Visual Objects in Time

The existent context is now introduced as the minimal unit of spatio-temporal decomposition for moving picture sequences. The existent context is defined as the set of images from a moving picture sequence that contain a visual object that can be recognized using the same set of image features for a *specific* pair of recognition and location algorithms. Clearly the temporal extent of the existent context is dependent on the capabilities of the recognition and location algorithms in question but the definition is independent of these capabilities. In fact there is scope for competitive evaluation of algorithms for existent context instantiation. So the existent description is extended to include temporal references or frame numbers (see Figure 5). The existent context provides a bridging mechanism between image features characterizing the data plane entities that correspond to a visual object perceived by an observer and the denotative and connotative description of the visual object. It does this in a way that is objectively verifiable whilst being limited by visual object recognition and location algorithm performance. For this reason it is preferred to the setting construct developed by Parkes [17], which is outlined below for comparison.

Visual objects in moving picture content are not just visible for single sequences of frames but may appear disappear or be occluded many times from the observer's viewpoint in any meaningful sequence. The existent context is therefore allowed to reference any number of different feature sets that are needed to support the objects recognition and location throughout the entire sequence. If an object is present in a sequence at a scale or level of occlusion different from that at which the bound existent context was initially instantiated another binding between the description and feature set occurs. This gives a failure driven mechanism for binding the feature sets to the descriptive elements.

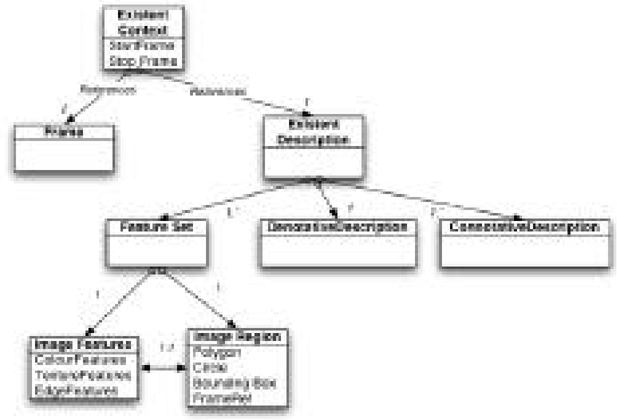


Figure 5. Existent Context.

4.1.3 The Setting

Parkes [17] in his approach to videodisc content description introduced the setting construct. He used the construct to provide a conceptual bridge between entities that were objectively visible in a still image and their objective description. Parkes defined the setting to be a set of images in a moving picture sequence sharing the same objectively visible state. The objectivity was dependent on the describer. The associated setting description was an objective description of that visible state. Underpinning the setting concept was the assumption that a still image of a motion image sequence could only have motion inferred from it rather than the still image being able to denote motion. This is assumption is invalidated by functional nuclear magnetic resonance imaging results [7] showing the same brain areas are excited when still images with implied motion are viewed as are excited when moving image sequences with comparable movement are viewed. The existent context in contrast provides a definition that is verifiable against feature extraction algorithm performance. It is also not dependent upon assumptions about the depiction of motion.

5. Applying the Model

A descriptive account of how the model has been applied to an episode of the UK children's television program Teletubbies[1] will now be given; a comparable numerical analysis is currently in progress. A small number of frames from the 'Favourite Things" episode of this popular program are reproduced. This account serves to illustrate the role of the constructs described in the earlier sections. This content was chosen because it combines simplicity of character, location and plot to be tractable enough to illustrate the concepts under discussion. Yet it is complex enough to be interesting and expose many of the issues present in more complex material. The "Favourite Things" episode in particular was chosen because the quest theme that it realizes is repeated four times involving a different teletubbie in turn.

---

[1] Images reproduced from Teletubbies Favourite Things are copyright BBC/Ragdoll Productions Ltd 1996 and are reproduced with permission.


Figure 6 (1:50).


Figure 7 (3:40).

Many episodes of teletubbies begin with a title sequence where each of the characters is introduced in turn followed by a group shot (see Figure 6). The numbers correspond to approximate time codes for the frames. During the introductory sequence the describer would bind the name of each teletubby to the feature set characterizing each teletubby. Each of the teletubbies is then shown with their corresponding favorite thing. So additional feature sets would have to be instantiated and bound to the favorite things existent descriptions corresponding to the visual objects now visible (see Figure 7). The narrator informs us that LaLa's ball is missing, which is clearly visible in the shot shown in figure 7.

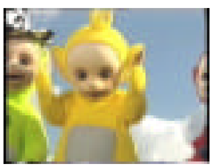


Figure 8 (4:05).

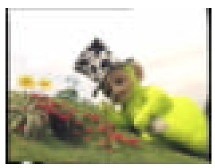


Figure 9 (04:21).

LaLa then expresses distress about the lost ball and there is a significant change of scale (see Figure 8). This results in one teletubby being obscured completely and the other two becoming severely occluded. So the existent context for Tinky Winky would end and additional feature sets instantiated and bound for Dipsy and Po. The search for the missing ball then begins. A number of different locations are visited and a cut occurs between each. The scale/occlusion level in these sequences is consistent with those already instantiated in the earlier sequences (see Figure 9).


Figure 10 (04:31).


Figure 11 (05:00).

In Figures 10 and 11 more locations are shown being visited by the teletubbies. Figure 11 in particular shows a significant level of scale change affecting two of the teletubbies. So a new feature set would need to be bound to the existent descriptions concerned. Several cuts also occur and there is considerable object motion in these sequences so the frames shown are representative only. It has already been noted that motion modeling is not considered in this analysis.

LaLa's missing ball is then shown in a new location requiring that a new existent description and existent context must be instantiated. The modeling of relationships between existents and existent contexts will be considered in further work.


Figure 12 (05:17).

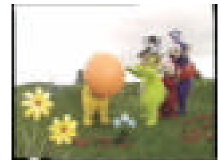


Figure 13 (05:38).

Dipsy finds the ball (see Figure 12). The scales of both the ball and Dipsy are consistent with an existent context to feature set binding for both to objects. The last image shown in this analysis is that of the ball being returned to LaLa by all the teletubbies (see Figure 13). This image is particularly interesting because LaLa is severely occluded and the other teletubbies reenter the picture at different levels of occlusion. In subsequent sequences each of the other teletubbies favorite things is lost and subsequently found. In these stories many sequences already described are repeated. So the previous existent context bindings can be reused in the subsequent realizations of the search story except where the story differs from the one described here.

6. Conclusions

The overt binding of feature sets derived from image processing to denotative descriptions has been introduced. Both feature sets and existent descriptions can be instantiated without being bound to each other. This allows analysis to be carried out without modeling and vice-versa. The existent description and existent context constructs have been introduced. The existent description combines the feature sets with denotative descriptions through the binding mechanism. The existent context has been shown to be a preferable minimal unit of temporal decomposition to the setting. These components are used in the AVA approach to content description at Lancaster. Further extensions of the modeling to accommodate higher-level concepts, object relationships and object motion from a semiotic perspective will be the subject of a subsequent paper.

7. ACKNOWLEDGMENTS

UK EPSRC provided financial support for work contributing to this paper under grants GR L81/604 (MIM) and GR M81/755 (AVA). The author's participation in MPEG-7 standardization work was supported by grants GR S421871/01 and GR S421871/01

Hartley family resources have also provided financial support for this work.

8. REFERENCES

[1] Aas K. and Line L., A survey on: Content-based access to image and video databases. Technical Report Report Number 915, Norsk Regensentral (Norwegian Computing Center), Gaustadalleen 23, Post Office Box 114 Blindern, N-0314 Oslo, Norway, 1997.

[2] Berlin B. and Paul K., Basic Color terms: their universality and evolution. Berkeley University of California Press, 1969.

[3] Brunelli R., Mich O., and Modena C.M., A survey of the automatic indexing of video data. Journal of Computer Vision and Image Representation, 10:78–112, 1999.

[4] Chatman S. Story and Discourse: Narrative Structure in Fiction and in Film. ISBN 0-8014-9186-X. Cornell University, 1978, 1993.

[5] Eco U. A Theory of Semiotics. ISBN 0-253-35955. Indiana University Press, Bloomington, London, 1976.

[6] Ennesser F. and Gerard Medioni G., Finding Waldo or focus of attention using local color information. IEE Transactions on Pattern Analysis and Machine Intelligence, 17(8): 805–809, August 1995.

[7] Goebel G., Khorram-Sefat D., Muckli L., Hacker H., and Singer W., The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. European Journal of Neuroscience, 10:1563–1573, 1998.

[8] Hartley E., Parkes A.P., Hutchison D. A Conceptual Framework to Support Content-Based Multimedia Applications, pages 297–315. ISSN 0302-9743 LNCS 1629. Springer, 1999.

[9] ISO/IEC 15938 Multimedia content description interface Part 1 Systems ISO IEC, 2000.

[10] ISO IEC SC29 WG11. ISO/IEC 15398 Multimedia content description interface Part 5 Visual. ISO IEC, 2000.

[11] ISO MPEG-7 Requirements

[12] ISO IEC SC29 WG11. ISO/IEC 15398 Multimedia content description interface Part 5 Multimedia Description Schemes. ISO IEC, 2000.

[13] Koprinska I. and Carrato S., Detecting and classsifying video shot boundaries in mpeg compressed sequences. In Proceedings of the IX European Signal Processing Conference (EUSIPCO), pages 1729–1732, Rhodes, 1998.

[14] Lienhart R., Comparison of automatic shot boundary detection algorithms. Image Video Processing, VII, 1999

[15] Manjunath B., Salembier P., and Sikora T., editors. Introduciton to MPEG-7 Multimedia Content Description Interface. ISBN 0 471 48678 7. Wiley, 2002.

[16] Nack F. and Parkes A. P., Towards the automated editing of theme orientated video sequences. Applied Artificial Intelligence, 11(4):331–366, 1997.

[17] Parkes A.P., Settings and the settings structure: The description and automated propagation of networks for perusing videodisc image states. Proceedings of the Twelfth Annual International ACMSIGIR on Research and Development in Information Retrieval, 229–238, June 1989.

[18] Schiele B. and Crowley J.L. Recognition without Correspondence using Multidimensional Receptive Field Histograms, 36 (1):31-50, 2000

[19] J. R. Smith J.R. and Chang S.-F., VisualSeek: a fully automated content-based image query system. ACM Multimedia '96, November, 1996

[20] Swain M.T. and Ballard D.H., Color indexing. International Journal of Computer Vision (IJCV), 7(1): 11–32, 1991

ap - fm01

martin howse
ap
studio 22,38-40 upper clapton rd
london e5 8bq uk
m@1010.co.uk

jonathan kemp
ap
studio 22, 38-40 upper clapton rd
london e5 8bq uk
j@1010.co.uk

ABSTRACT
This paper examines a new discursive space which has been exposed by radical software art through the heuristic possibilities opened up by software abstraction.

Keywords
Machinic abstraction, environmental coding, endodata, software art, auto-destructive, distributed systems, reconfigurable computing

1. INTRODUCTION
fm01 is speculative software: that is it both explores the potentiality of possible programming thereby creating transversal connections between data/code, machines and networks, and software that reflexively addresses/investigates and finally reinvents itself as software by its own means (it is a bastard ontology): presenting a release of technology from the ontological idea of its usage, and permit its own ontological structuring as unknowable - placing it in the unknowable Kantian realm of practical reason (like infinity, god, morality, things in themselves) and outside of the operation of pure reason, (a sphere of necessity, of laws of cause and effect, syllogism, the law of identity/singularity, of the excluded middle); and evading the logic of the universal/particular coupling. so for us, the ontogenesis of fm01, isn't to be found in the retelling of a teleological story of image processing culminating in digitization or whatever (just as the immune system doesn't orientate itself to the survival of its host) and as fm01 doesn't take images as carriers of experiences and meanings, we are not interested or know where it sits in the context of expanded film.

instead fm01 attempts to confront between what man regards as being possible [function + value] and what machines present as feasible [in potentia]. technologies are permanently shifting this relation between the possible (potential) and the feasible (functional) and where the construct of the real constitutes such a "negotiation" between the potential and the functional, being constantly reformulated.

fm01 [also] attempts to ask the radical question of how data can be represented for the machinic without the supervenience of meta-data or the demand for the purely functional impacting on abstraction.

fm01 as investigation centres around the domains of:

1. new operational zones created by machinic ontologies [lying on the plane of fascism, outside humanism, because it demands a space outside the operation of pure reason in to the realm of practical reason (things in themselves)].
2. the specific exposition of a radical space for code art [denying the trope that if code defines how data gets handled on a technical level, then meta-code [like belief systems, ideologies and organizing principles] is the philosophically relevant level]

subthemes in relation to these two domains are: a) the decoding and recoding of data under non species dependent symbolic orders b) the form space of code and environment and its autopoeitic engineering. so the relationship we will present between image and generation is not a reenactment of the anthropological semantics of the human eye but is part of an integrative poeisis of processing between orders as extant between geology, plants, machines and humans. the making of images is not a return to imagination of iconic culture but the move to calculation + computation [vilem flusser], the morphospace of numerology. although we're used to being on the winning side of the image an image comes full circle when it is revealed by its own machinic underpinning.

2. ENVIRONMENT
in presenting the fm01 project, there is an underlying assumption that the audience is familiar with some current issues and practices in contemporary software art/cultural practice including digitally expanded cinema. a general synopsis of current paradigmatic thought around issues specifically of digital cinema may be found in Peter Weibel/Jeffrey Shaw's accompanying book for ZKM's 2002/3 Future Cinema exhibition[1], where they

write that, for example, “the biggest challenge for the digitally extended cinema is the conception and design of new narrative techniques that allow the interactive and emergent features of that medium to be fulfillingly embodied”[1], and outline three overlapping areas of recent praxis:

1. modular structures of narrative content which allow indeterminate numbers of permutations or parallelisms (eg Lev Manovich’s “soft cinema”; Jon Jost; Eija- Lisa Atthila; Raymond Tomin’s AVRA software; Jennifer + Kevin McCoy; Marc Lafia’s Max/Msp based “variable montage”).
2. algorithmic generation of narrative/mnemonic sequences/ markers that could be modulated by the user (eg. Martin Reinhart + Virgil Widrich’s “tx transform”; Marnix de Nijs “run, motherfucker, run!”; the work of George Legrady).
3. digitally extended cinema inhabited by audience who become immersed agents and protagonists in its narrative development (eg. Jeffrey Shaw; Michael Naimark; Margarete Jahrmann + Max Mosswitzer’s “rgani-engine-toolz”).

Future Cinema was cast as a show where “the medium is the message” yet was situated in the familiar terrain of the hegemony of the species, the subjugation of machinic cinema embodiments to the plateau of pure reason, and the enslavement of technology to an ontology of relational functionalism. It is paradoxical that Future Cinema’s seeming activism is consistent with a series of static shots from a fixed point of view, between man and machine, as a homogenous relationship [the great Gutenberg program].

3. CODE

there’s a lot of data flying about that isn’t species dependent for its encoding, decoding or recoding, and in many ways we are no longer the sole traders in the realm of the symbolic as a species: for example, within the interaction of human text with machine coding, language is not the exclusive domain of human thought but also that of the internal logic of computers. Fm01 is an attempt within this context to offer a total software environment for the semi-automated production, scripting and editing of endless cinema outside the humanist realm of meta-data. Fm01 is not conceived as an editing engine for the manipulation of generic clips (in an expanding database of all possible scenes rganizatio according to a huge number of elements and relations) but rather offers an enmeshing within script, data streams and environment [WHERE ENVIRONMENT=CODE=ENVIRONMENT IN BEAUTIFUL TRAJECTORIES]. An environment of trajectorial machine nodes - from environment to language these machinic nodes are the environment - consisting of trajectories, macros-functions, relations, levels, scenes, scripts, machine-nodes.

Gdapp (generic data application) commenced coding in late 2003: the idea being that a pure lisp-coded prototype fm01 would meet with the arc described by the nodal, distributed gdapp experiment. An intensely nodal, modular, environmental, self-referencing system which builds on ap02 vm model but extending and multiplying this model/ dissecting it also into a nodal/ geologic strata model. Gdapp describes a globular/dynamic changing node structure of connections, of flow and instructions - a total environment, an operating system which dynamically re-codes itself, is recoded in operation. Layers of abstraction as environment: consisting of a base layer which mobile, almost viral code nodes sit and work upon and an open exchange as opposed to a protective, secure OS model. Gdapp codes promiscuity and open discovery rather than static upload/download or I/O structuring. Gdapp codes through a nodal model: nodes run/write code/data, nodes run other nodes, nodes link other nodes, nodes reflect macro linkages or embed code and linkages within neural model of linkage excitation.

The technologies of vision/representation have been driven by a indexical linking between reality and representation [image and its physical rganiza object] and we are compelled to locate [veracity] within the technologised image [the history of film and relation between celluloid/chemicals/light/scene]. If images have been conceived of only for human species then images that are machine-assisted or automated seeing render redundant imaging as an attempt to reconcile this contradiction, the indexical idea.

Digital images are not immaterial [as some think through the breakage of indexical link]

“hardware looks like programs: it is a configuration. Hence programs, hardware (configurations) and data have the same nature.” [2]

descriptions of cell matrix architecture (Macias 1999) well match the gdapp engineering philosophy. The cell matrix is internally configured by exchanging configuration data with its neighbours

“evolvable computing and some modern systems do not share computational scenario of a standard Turing machine and cannot be simulated on Turing machines.” [2] thus interactivity and infinity come into play with consequences for any theory of abstraction. That it is not a field of equivalence - here time enters the computational equation.

“at each time point evolvable devices have a finite description - however when one observes their computation in time they represent infinite sequences of reactive devices computing non-uniformly”[2]

“evolvable computing is beyond scope of an ordinary Turing machine. It does not violate the Church-Turing thesis because this thesis deals with a slightly different class of computations corresponding to the concept of algorithm. This class of computations is not typical for contemporary computational systems” [2]

the advent of the digital image is thus freed from its material support and mobilizes it through networks and recoded as mathematical information (bits, code) that allows it (the bit-real) to be recomposed infinitely and to flow indefinably as data.

Complex data cannot be comprehended by human reading (too slow). Data is rendered /abridged by images (like the celluloid film strip of Conrad Zuse holding both image/code and data). Digital calculation beyond the individual subject refers neither to the differential symbolic order represented on the screen nor to a world outside this screen (physical reality behind the screen is state and current only); the digital machinery retreats into total abstraction and is catachretic, not metaphoric, a baroque violence rendered to the potentia of the machinic phylum.

So what would be in this philosophy of images. Software routines and their non-traceable/path dependent images are a realm/ environment of data/code [in effect the break between the material support of digital imagery, the break between the indexical link]. And so in fm01 a relational, nodal language of connection will be formulated to descend through levels of scene, shot and frame. Instruction sets (which refer only to trajectories and grammars) are dependent on context/environment but offer the same functionality [non-functionality] across all flattened levels from pixels to major trajectories, from major conceptual conceits and rganizations of nodes to a single script particle.

Fm01 is running invisible code: its self-organising is ceaseless and in contradistinction is environmentally mobile: distributed and polyvalent. Code is an environment of subsystems in search of form or pattern, the orphaned bits without superstructure where nothing corresponds to anything inside nor outside the machineware that applies it. Nothing more than the autonomous movements of data within the clock of the machine - ways of doing things become situated within object orientated fields of influence (marked by a level of analysis, meta/o-nymed) (eg. The micro influenced by the macro, or a “case of” where a process is named) and we see in one sense a blind text/residue stratum ie. Electronic images are effects of a surface right from the start where the surface still appears as material of an object and as intensity of a pixel.

4. ENGINEERING PHILOSOPHY

textural lisp work with an extant written script (a description, a specification for fm01 almost) will describe a trajectory and meet with the arc of gdapp to constitute the first full version of fm01. this textual work (operating on text as stream rather than as file - text as particle physics) is towards creating a compiler/interpreter which actively runs that script. Gdapp is concerned with mobile code nodes which run and enact on data - both approaches lead to fm01 - runs the script creating new text or film - runs that script which acts on and suggests data sets - active across data sets and is inside-embedded within/as data sets.

Components outside an object model - trajectorised components include script, node, language implying an interpreter and particles with velocity and open destination. To begin with the script:

script as making evident of word-background/that-which-is-written within script as it stands - towards what-will-be filmed but without discarding the script. Background material includes data such as how it was written - how it is to be filmed (this is not meta-data but the being-script). Within such a landscape -relation the issue of vectorisation or trajectory is important.

Background/script (film script) stands in same relation to film (filmed) as program to execution - background as revelation of code and script rewriting film rewriting script.

Scene as node which can change/merge/enact upon other nodes/scenes. Functional data/code/scene nodes collaborate/ change each other - for example image analysis nodes/compression nodes

script as node, node as mobile, node as cross-level functional and able to negotiate, coerce and encompass trajectories of nodes, node as interpreter and interpreted ongoing, node as exchange with nodes and collaboration.

Back-foreground relation as question of interpretation (and context =environment) and the ability to alter-construct node trajectories. Nodes define a language which can be shared or argued over - misconceived. Fm01 is thus the description of a language. Language is described by the script. The script is defined in terms of mechanisms with the node as this mechanism which functions in/or constructing/constructs language.

Computer language (created on top of another and bootstrapping) describing an environment which is node, script, interpreter and trajectory (distributed) at the same time. Vectorisation of script and image - as image and in script across combinatorial senses of pages, scenes, shots and also within image and scene (eg. Concentric studio spaces and clues=trajectories) as well as being embedded in components script also defines trajectory (networked and across nodes) of/for software components

instructions or tokens (to use such a word?) are dependent on context - but offering the same functionality across all flattened levels of what could be seen as high and low levels - of pixels and major trajectories (some functions dependent on context = environment). trajectory of node-particle possesses speed and crosses levels as dictated. node is particle, relation and change of space of that particle.

to describe the meeting of both arcs: the co-evolution of cell data-code interpreter elements in lisp (eg. evolution of edge detection). co-evolving mobile modules explore and exchange different functionality and data/code whilst working on both script stream I/O and film stream I/O: simply to replace fixed database software with mobile nodes and changing code/data streams across parallel architectures. script-data-code IS these modules (how module is defined) thus software is really the interpreter or some sort of bootstrap of language and interpreter. nodes are the data, are the software - kein outside, kein addressing, purely trajectorial nodes impacting on an environment which is solely of nodes, of strata. fm01 has purely generic nodes which are just assemblages according to no function - function defined by relations and trajectory - interior (invisible) machinic function.

5. REFERENCES

1. Shaw, J, and Weibel, P. Future Cinema. MIT Press.
2. Sekanina, L. Evolvable computing by means of evolvable components. Natural Computing 00 1-31, 2004

lifeSigns: Eco-System of Signs & Symbols

Troy Innocent
Faculty of Art & Design
Monash University
Australia
Troy.Innocent@artdes.monash.edu.au

ABSTRACT

The nature of the computers as a machine for processing symbolic logic is widely recognised. However, they also have the potential to 'generate meaning' through these same processes, particularly in the domain of constructed, artificial worlds that exist in terms of their own logic and codes. An interdisciplinary model for meaning generation is articulated via the intersection of artificial life and generative systems, computational semiotics, and digital games. This model may be described as a 'generative meaning system'.

A digital media art installation entitled *lifeSigns*, an eco-system of signs and symbols, is presented as a prototypical generative meaning system. This work is a multiplayer digital game set in a world of iconic lifeforms, called 'lifeSigns', that make up an abstract world of form, colour and movement.

Keywords

artificial life, generative systems, digital media language, electronic media art, digital games, iconography, interactive sound design, media creatures, generative meaning system

1. INTRODUCTION

The idea of computer as a processor and manipulator of symbols and logic is widely understood. Aspects of electronic spaces, such as their fluid, mutable nature and immediacy become familiar as they become part of everyday experience. However, these artificial worlds offer other possibilities outside the simulation of the material world. They are systems that possess their own logic and potential forms of expression. Push one or two parameters past the 'normal' limits and you are into the realm of transmutational space where anything can be parameterised, algorithmically generated, or made interactive. There is not one global code or language that defines all electronic spaces - each space is potentially a unique instance of a particular logic or language.

This paper intends to explore an interdisciplinary approach to the construction of artificial worlds that have the potential to generate meaning. In artificial worlds where signs and symbols have agency - could they adapt their form or behaviour so as to change their meaning? Existing theories and systems will be explored to examine the potential of electronic space as a site for the evolution of audiovisual languages. It explores the flux of meaning and constant reconfiguration of language brought about by new technologies of communication. If electronic signs, such as icons, have agency within the digital realm in which they reside, then how does this change our understanding of information, data, language and communication? What happens when the formation of meaning is augmented by computational processes?

These ideas are demonstrated in a recent artwork entitled *lifeSigns: an eco-system of signs and symbols.* In this work, a system for generating, playing and interpreting digital media languages is the basis of an artificial world. Further to this, my interest in these themes has evolved from the development of digital media artworks that explore new communication systems in the digital realm. Through the experience and creation of digital media, I have been interested in what can be described as a 'language of computers' that articulates the unique characteristics and properties of digital media.

2. PAGE SIZE

In order to facilitate the idea of the evolution of language in digital space, the right processes, interactive mechanisms and representations need to be developed. The goal is to realize a space that is capable of change and evolution, user feedback and interpretation, and that can be expressed in terms of codes and structures of communication. Language may then 'emerge' through ongoing interaction and user participation in that space.

2.1 Interdisciplinary model

The conditions required in an environment suited to the evolution of digital media languages need to draw upon the formal languages and models of evolution in artificial life research, the analysis of digital media language from computational semiotics, and the capacity of feedback and gameplay in digital games.

2.1.1 Artificial Life and Generative Systems

Artificial life, a discipline of computer science, encodes life processes into algorithms and data structures that can be implemented in computer software. A genotype defines entities within the system may be mutated or combined with that of other entities in a process similar to biological reproduction, the genetic algorithm. Typically a fitness function will test entities in terms of criteria that may be defined in terms of the system, or entities will adapt their behaviour through learning and adaptation.

Generative systems build structures from simple base elements combined using rules and grammars. Many iterations of these processes result in complex forms that can create novel results not readily apparent when looking at the source material. These systems may be integrated into artificial life models and may also be tested in terms of fitness, combined using genetic algorithms and so on. These systems have been used by artists and designers to evolve worlds and develop new aesthetics.

2.1.2 Computational Semiotics

In order to make use of these systems, the subject of the artificial life simulation needs to be defined in terms of

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

elements and relationships that may be manipulated by the system. Computational semiotics provides methods for relating the 'computer and sign system' that may guide the formal definition of semiotic systems in terms of an alife algorithm. The unique characteristics and opportunities of computer mediated communication are articulated and decoded through the adaptation of semiotic theories to the computational medium. These findings may be used to translate communication processes into formal languages that may be defined in the symbolic code of the computer program.

2.1.3 Digital Games

Digital games combine key properties of digital media, such as interaction, simulation and symbol processing, in a single coherent form. In particular realtime 3D simulation of space offers the ideal medium for the experience of artificial worlds. Many different kinds of worlds can be simulated with a high degree of 'realism' in terms of the players experience and perception. Gameplay offers a structured code of interaction while also allowing exploration and play within the rules of a world. The immediacy of feedback combined with the realism of the simulation result in a high level of engagement that places the player directly within the action of the symbolic game world.

2.2 Generative Meaning Systems

An interdisciplinary model for meaning generation is articulated via the intersection of artificial life and generative systems, computational semiotics, and digital games. The life processes encoded into artificial life software may be connected to systems and codes of signification in digital media to develop models for the evolution of digital media languages. The forms evolved by the system may be represented in realtime 3D simulation – through form, structure, colour, sound, motion, surface and behaviour. The game environment makes the system tangible and provides a medium for feedback. This model may be described as a 'generative meaning system'.

These systems should have the following key characteristics:

1. clearly defined base media elements described in terms of current knowledge on signification and meaning within digital media;
2. rules or grammars for the combination of elements to create larger entities;
3. evolutionary or generative aspects such as the use of iterative processes, genetic algorithms, fitness function;
4. a method or function for the interpretation of meaning of the entities evolved by the system;
5. rules of play that relate to interpretation functions and communication;
6. a representation that encapsulates these characteristics in a coherent, persistent world and enables the interaction of an audience.

Generative meaning systems highlight the potential of digital worlds to manifest new configurations of media and meaning. They demonstrate the intersection of life processes with the evolution of communication forms in terms of what could be described as 'emergent language'. Digital media languages may be evolved and bred rather than designed. Furthermore, these languages may adapt to different semiotic landscapes or evolve through their interpretation and usage. Generative meaning systems are representative of the 'nature of digital media' by demonstrating properties of mutability, interactivity and symbol processing.

2.3 Previous work

This project builds on previous works that have explored dynamic relationships between meaning, artificial life, interaction and electronic space.

2.3.1 Iconica

Iconica [3] is an artificial world made of language, populated by lifeforms made of media. An artificial life model drives the evolution of the world, augmented by human interaction. The endemic language of this world is a system of icons that may be combined using rules to create an endless number of possible meanings. However, the forms within this world are derived from combining elements within a database of pre-existing forms. *lifeSigns* aims to extend this process of meaning generation by using a grammar to create multiple possible iconographic languages.

2.3.2 Semiomorph

Semiomorph [2] is a digital game that explores "semiotic morphism", a "systematic translation between sign systems". [1] The term captures the shape-shifting plasticity of relationships between sound, image, text, and users in virtual worlds; the interactions through which meaning is made, transformed and remade dynamically and synaesthetically in real time. In this work, the gameplay is centred around a typical digital game structure with levels, score, enemies and so on. *lifeSigns* aim to allow 'free play' with the world by removing these constraints - enabling exploration and experimentation within the world.

3. lifeSigns: Eco-System of Signs & Symbols

lifesigns explores the idea of language as emergent phenomena. It intends to establish a digital semiotics that includes dynamic form, evolving structure and synaesthetic representation in its definition. An underlying "abstract machine" defines rules for generating the glyphs in the language, and these rules extend to define their behaviour in a virtual world.

A number of theories on the language of digital media have emerged recently. *lifesigns* investigates five properties of digital media described by the Manovich [6] theory on digital media language.[1] The numerical representation characteristic of digital media is the basis of all processes and interactions within the *lifesigns* system. The living signs, or 'lifeSigns', generated by the system are built from simple atoms, and this modular approach is applied at all levels in the system. The work evolves through multiple iterations autonomously, utilising the capacity of digital media to encode process. Icon's that evolve from these processes display a high degree of variability in their representation, resulting in an emergent language rather than a language that is designed or constructed. lifeSigns are represented as image, sound and motion through the transmutation of the digital code from which they originated. The user may play selected icons and

[1] Lev Manovich proposes five properties in *The Language of New Media*: (I) numerical representation, (ii) modularity, (iii) automation, (iv) variability, and (iv) transcoding.

this causes signals to be transmitted through their geometric structure. These signals create movement and sound in the generated icons that illustrate the internal processes within the system.

3.1 Overview

lifeSigns is a world that generates iconographic languages. It explores the idea of 'emergent language' through computational processes, where both human and digital agents contribute to the formation of meaning.

It explores he interdisciplinary model for meaning generation articulated earlier via the intersection of artificial life and generative systems, computational semiotics, and digital games. The processes encoded into artificial life software may be connected to systems and codes of signification enabling the evolution of digital media languages. The living signs ('lifeSigns') evolved by the system are represented through form, structure, colour, sound, motion, surface and behaviour. The game environment makes the system tangible and provides a medium for feedback.

Navigation and play in the *lifeSigns* world using conventions of digital games and electronic musical performance. Playing each lifeSign reveals its unique form and behavior through synaesthetic animation of image and sound. The languages generated by the work are interpreted by the audience through a cumulative process of assigning meanings during play in the world.

The work explores themes of language and meaning, synaesthesia, process, representation and reality, and the nature of digital media. The system and processes of the *lifeSigns* world will be explored within the context of these themes.

3.2 Installation

The lifeSigns world is an interactive installation that combines two separate views of the space. A large projection shows a map of this world, accompanied by a generative soundscape. Four workstations placed around the edges of the map enable navigation and play of the world. The space appears to be infinite as the player may continuously navigate and search in any direction – there is no up or down. Although it is a three-dimensional representation there is no horizon line or indication of linear perspective, but instead multiple layers of abstract form, colour and movement. When lifeSigns are played they perform animation and music, and their energy is increased thus making them more likely to persist in the world. When multiple users are playing a number of lifeSigns simultaneously, the result is a symphony of artificial image and sound.

The gallery space is extended by the connected web site, which documents all of the icons generated in the world. Visitors to the site can assign meanings to each lifeSign that are transmitted back to the installation.

3.3 Interaction

Interaction with lifeSigns can be summarised into three levels of experience:

1. Ambience: experience the installation space, multi-layered sound, activity in map view
2. Navigation: using one of the workstations to move through the world, observe icon behaviour and form
3. Play: Select and play icons causing them to perform animation and music, give them energy. Multiple players may perform together.

3.4 World

Within the installation space, the *lifeSigns* world is represented in a top-down 'map' that shows all the current lifeSigns, meanings for sequences of connected lifeSigns, and behavioural interaction. This view is shown in Figure 1. In addition to function as a dynamic display of information pertaining to the current state of the artificial world, this map visualizes the idea of a world made of signs and symbols. It can be seen as a kind of iconic cosmos displaying the evolutionary process of the world's audiovisual language.

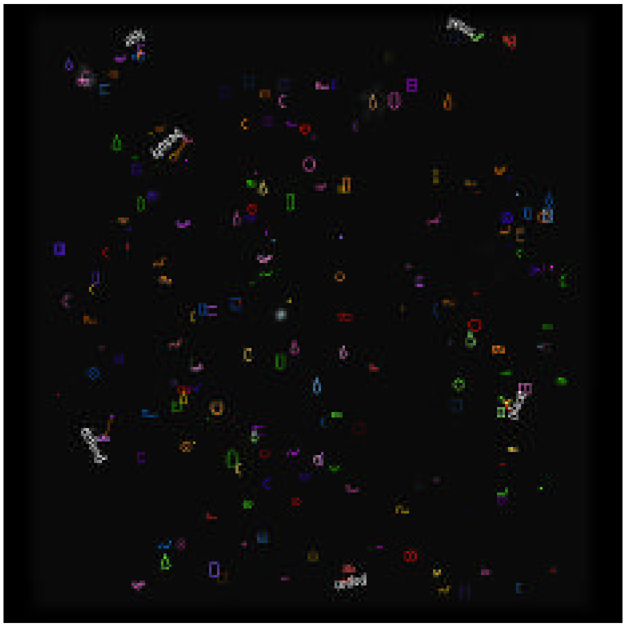


Figure 1. map of lifeSigns world

The world itself is populated by hundreds of lifeSigns that move and interact autonomously. It is a persistent world – data is saved regularly – so that during a period of gallery installation the world population is the result of many generations of lifeSigns. As depicted in Figure 2, the space is simply a large, square volume filled with lifeSigns. As the player freely navigates this space, the experience may be described as immersion in a three-dimensional volume of abstract form and colour. It is a 'persistent world' – the data is automatically saved on a regular basis allowing the same lifeSigns world to run over many generations during its installation in a gallery space.

4. STRUCTURE

Each lifeSign generated by the world is unique. Its form and structure is generated by rules that are stored within a 'dna' string that defines many of its attributes and properties.

Structural patterns are discernable in existing families of signs and symbols. Studies of these patterns have resulted in their use as an indexing tool to search large collections of symbols using features of their form and structure to identify them. The Dictionary of Symbols [5] uses this kind of system to allow the meaning of symbols to be defined simply by deconstructing their form and structure. The system used in this dictionary classifies symbols in terms of soft and hard lines, symmetry and other formal properties

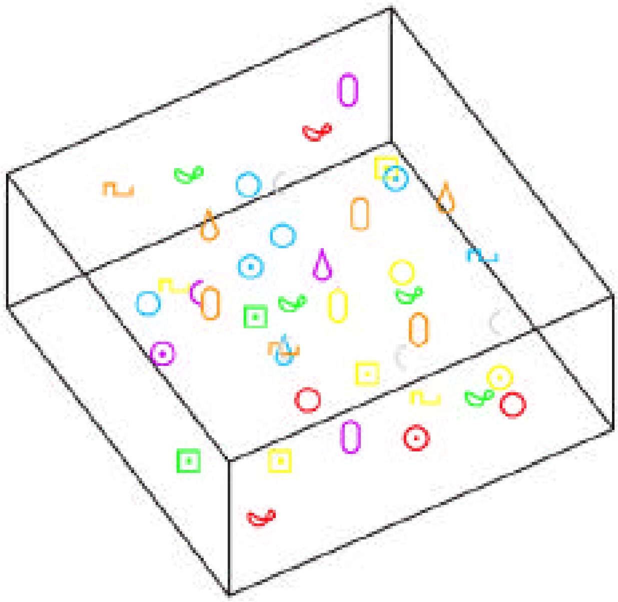


Figure 2. 'volume' of lifeSigns

This system was adapted into the rules used for generating the form of lifeSigns. An atom is the most fundamental component of the system. It exhibits visual and sound effects. Atoms are used to form more complex structures, called lifeSigns. Within this system each lifeSign is built from the atoms in a hierarchical fashion, beginning with a central core shape. An atom is defined as the basic unit within this structure –in Figure 4 a central core shape (the large circle) has six atoms (the smaller circles) connected to it.

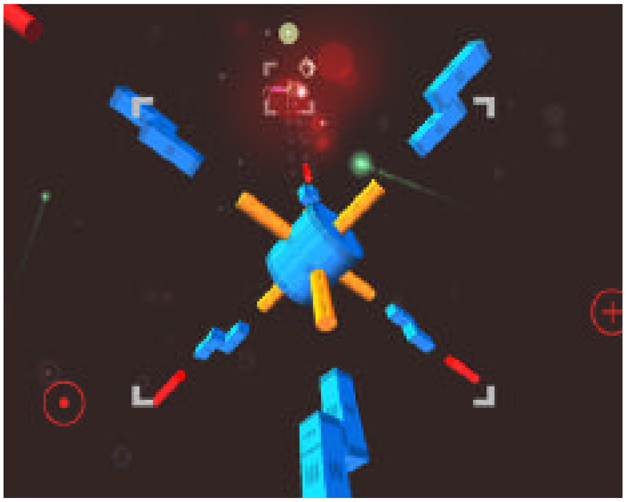


Figure 3. an image of a lifeSign

4.1 Structural Constraints

The structural constraints are defined in terms of a further developed set of criteria for the structural features of signs and symbols. As the lifeSigns exist in three-dimensional space, many of these rules needed to address the new possibilities of that space such as multiple axes of symmetry. Further rules for colour were also added that apply constraints to the palette of each lifeSign.

Fifteen structural constraints are defined in the system that are defined in detail in terms of the software operation, and some constraints interact with one another to create combinations of rules. However, in general terms, there are five main groups of these constraints:

1. *Symmetry:* this occurs in X, Y & Z axes
2. *Construction:* whether or not atoms intersect one another, if they ascend or descend in size through levels of the hierarchy and whether atoms form clusters or long chains of structure.
3. *Geometric primitives:* structural constraints govern the selection of geometric primitives (as described below).
4. *Colour:* different colour palettes may be generated including monochrome, uniform colour, any colour.
5. *Core shapes:* different core shapes may be used, including solids, freeform lines, points, open and closed lines.

These structural constraints comprise the basic, fundamental defining rules of the lifeSigns structure. Other levels of the system build and interact with these principles.

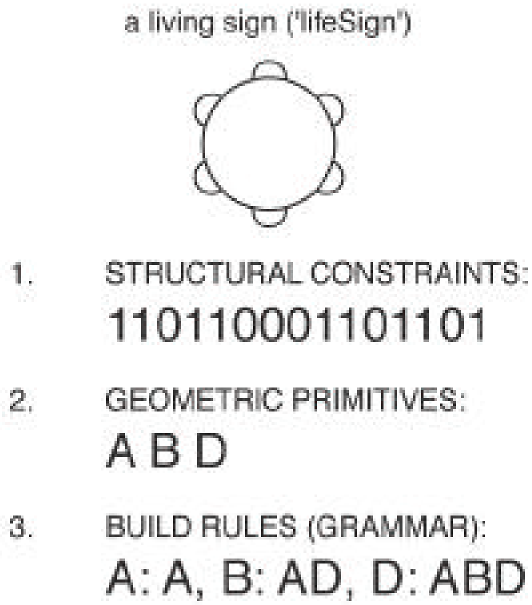


Figure 4. Structural rules for lifeSigns

4.2 Geometric Primitives

Governed by the structural constraints, selections from a set of geometric primitives and cores shapes are made for each lifeSign. This selection defines the basic set of building blocks usable in construction. Forms in the shape library consist of spheres and ellipsoids, cubes and oblongs, cones, cylinders, elbows and corners, torus shapes, squares, and some freeform curves. These are classified into 'soft' and 'hard' forms and different combinations of structural constraints will allow either or both of these to be used within a lifeSign.

It was important to limit the possibilities to those that make sense within the intended aesthetic and communication strategies of the work. In this case, the shape library is made from 'natural' forms of electronic space – the kinds of forms typically generated in 3D modeling software that are useful for constructing computer icons. Alternative shape libraries could also be used. One possibility that was considered during development was the use of gestalt symbols such as a star, a heart, a lightning bolt, and so on in order to create hybrid meanings of a different sort.

4.3 Build Rules

A simple grammar is defined that adds further constraints to the building process. For each shape selected from the library, a rule for what other shapes may be connected to it is defined. Probabilities for the each connecting shape are also generated. In the example depicted in Figure 4 a lifeSign with shapes A, B & D is defined. A build rule is then defined for each of these shapes. In the example, an A shape may only be followed by another A shape, a B shape by either a B or a D, and a D shape may have any of A, B or D connected to it. These rules generate patterns in the levels of structure within a lifeSign. Some patterns of structure generated by the sample rules would include:

AAAAA

BAAAA

BDBDBAAA

DDBDBDD

DAAA

DBDAAA

and so on…

All of this information is stored in the dna string so that it recorded as a characteristic of that lifeSign and passed onto children in further generations of the system.

4.4 Outcome of the system

Codes and relationships defined using the analytical tools of computational semiotics may be translated into genotypes, fitness functions, rules and grammars of artificial life / generative systems. These systems may then in turn generate more complex semiotic entities that would benefit from analysis using computations semiotics. Models of communication in digital games may be connected to the data generated by artificial life / generative systems to make this data perceptible and facilitate user interaction.

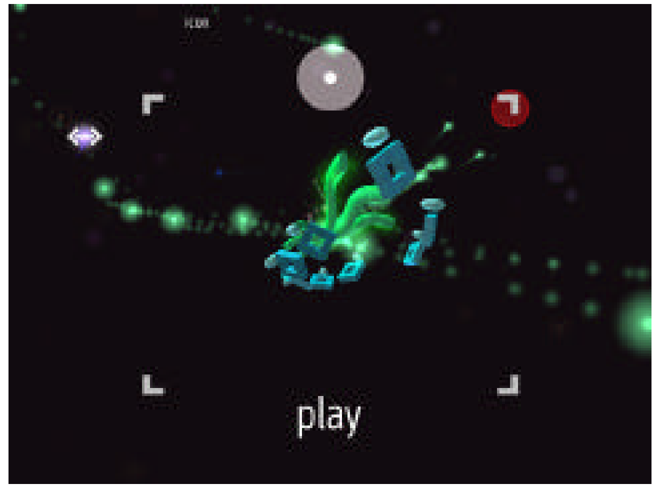


Figure 5. lifeSign in play

5. IMAGE & SOUND

Once each lifeSign is constructed from geometric primitives, additional parameters are defined that relate to its visual appearance, animation and sound. As the lifeSigns exist in a three-dimensional space represented by computer generated images, their visual representation relates to aspects of the 3D engine used to render the world. Translations of geometry, particle effects, texture mapping and 3D shader effects were among the parameters explored in the visualization. The sound or voice of the lifeSigns is generated using musical instruments defined by MIDI controlled synthesizer patches. Parameters of the synthesizer used to create the instruments such as its envelope, filters, pitch and amplitude effects form the parameters of sonic representation.

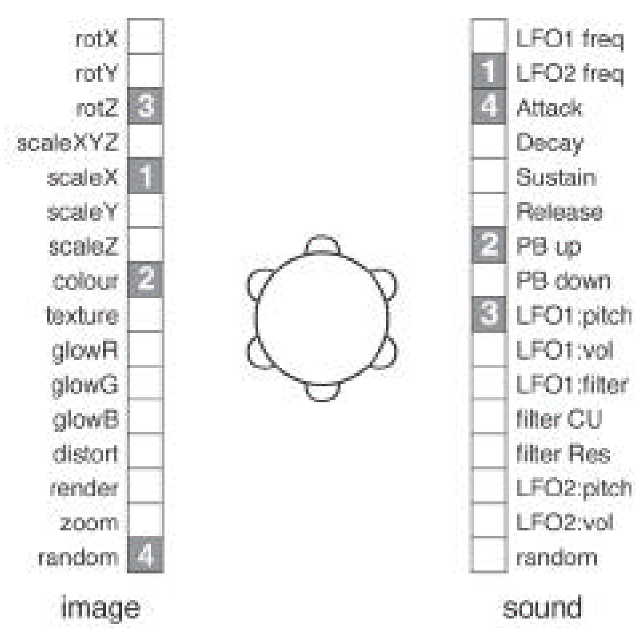


Figure 6. image and sound parameters (with selection)

5.1 Image & Sound Parameters

In this work, the Criterion Renderware 3D graphics engine was used to display the simulation. There are three basic types of image parameters in the system – those that alter the form of the geometry, such as scale, rotation and distortion; others that alter the surface of the forms, including colour, texture and rendering mode; and finally those that project glowing particles from the surface of the forms. An additional 'random' parameter selects a different parameter at random every time the lifeSign is played. During play, different parameters may be combined so that compounded effects may be generated such as scaling a form in the X-axis while changing its colour to blue and rotate in the Y-axis while glowing red.

The sound of each lifeSign is generated by the synthesizer on the Creative Audigy sound card and played back through stereo speakers. The sound parameters may be grouped into four groups – those that change the amplitude of the sound, including LFOs (Low Frequency Oscillators); others that alter the pitch; including pitch bend and more LFOs; parameters that change the frequency of the LFOs themselves; and filter effects. The sound parameters also have the 'random' parameter. Both the image and sound parameters may be performed in realtime and are vary in degree from zero to full effect using a 'continuous control' mode of operation. This allows the players input to be directly mapped to these parameters so that a 3D form may squash and stretch in response to the amount of input received, while simultaneously altering the pitch of a sound in response to the same input.

5.2 Synaesthetic Parameter Selection

Four image parameters and four sound parameters are randomly selected upon the generation of each lifeSign. A sample selection is indicated in Figure 6. These selections are sequentially grouped into image and sound parameter pairs. In Figure 6, for example, the image parameter 'scaleX' is connected to 'LFO2 Frequency' and 'colour' is connected to 'PitchBend Up'. When a player performs the system these parameters are played together as a single image/sound event.

This system allows many different combinations of the parameters to be explored through multiple iterations of the lifeSign generation process. Every lifeSign is unique in its form and structure, and also in how that structure is animated and transmutated into sound through performance of its image and sound parameters. As a result, many different synaesthetic relationships are explored by this system. Changes in the scale of 3D form may be connected to shifts in pitch, the frequency of an oscillator or the amount of attack on an envelope.

All of the selected parameters and groupings are encoded into the dna string of the lifeSign so that they are passed on from one generation to the next. In this way, successful combinations of image and sound persist in the simulation enabling the discovery of new expressive potential within the aspects of digital media used by the system.

5.3 Colour and Timbre

The results of the lifeSign generation process are used to determine other properties of each lifeSign. Their colour and texture is related to a set of defined behaviours in the artificial life model. These are defined in detail in section 6 – the visual effect is simply that each atom in the lifeSign structure may have a different colour. The lifeSigns overall colour scheme is governed by the relevant structural constraints. As a result, the behavioural mix of each lifeSign may be read through its colour and texture. A colour scheme that is predominately red, for example, would indicate a high proportion of the attack behaviour. Mostly green would indicate a more friendly lifeSign. Each behaviour is also represented by an animated texture.

Characteristics of the structure and colour of each lifeSign are used to select from a palette of instruments representing a range of timbres mapped across a simple two-dimensional grid. One axis of the grid varies from 'pure' sounds such as simple electronic tones through to noisy sounds that have complex harmonics and spectral dynamics. The other axis is used to classify sounds from those that are artificial, obviously computer generated or 'electronic', through to natural sounds, including samples of acoustic instruments and human voice. These two aspects of the sound timbre – pure / noise and artificial / natural are mapped to structural complexity and colour respectively. A lifeSign with a simple structure will have a sound on the pure end of the scale, another with a more complex colour scheme will be noisy. Likewise, a lifeSign with a simple colour scheme will have an artificial sound, while one with a more complex colour scheme will sound more natural.

5.4 Sequencing of Image & Sound

When these structures are played, inputs are passed from one atom to another in a hierarchechal fashion. As each atom is 'played' by receiving a signal, it displays a visual effect and plays a note via its instrument. This process is called the propagation of the signal.

There are different styles of propagation that change the way a lifeSign is played, mainly through the initial distribution of inputs to the atoms. These styles are called language, ambient and beatbox. Language mode sends all input to the first atom in the structure, allowing it to propagate down the hierarchy. Only one signal is allowed at a time, resulting in monophonic sound and immediate feedback. It is intended that this mode approximate speech, allowing the lifeSign to 'talk' using its voice and actions. Ambient mode allows multiple signals to propagate up and down the hierarchy, resulting in overlapping patterns of image and sound. Finally, beatbox mode maps inputs directly to separate and specific atoms with no propagation. This results in discrete image / sound events directly in response to play.

5.5 Outcome of the System

A complex system of interconnected parameters that relate to the audiovisual representation of lifeSigns allows the exploration of many permutations of the subset of digital media language expressed through the image and sound parameters defined within the 3D graphics engine and sound synthesizer. Furthermore, these permutations are encoded into the dna string of each lifeSign so that successful combinations of these parameters are inherited by offspring and passed on through generations within the simulation.

6. ARTIFICIAL LIFE

A simple artificial life model governs the overall population and evolution of the lifeSigns world. The lifeSigns have attributes such as energy and meaning that affect their actions in the world. Their primary function is to survive and attract the attention of players who increase their energy by ‘performing’ them.

6.1 Properties

Each lifeSign has a number of properties that relate to the artificial life model. These properties define each lifeSign as an organism or entity in terms of an eco-system of signs and symbols. As a result, they relate to both life processes, such as survival and reproduction, and to semiotic processes, such as signification and the transmission of signals.

Properties defined in the system include:

Energy: each lifeSign starts with a small reserve of energy. This energy is used when it moves about or performs behaviours. Simply being in the world causes this energy to slowly regenerate, so that there is a simple state of equilibrium. However, an excess of energy is required in order for the lifeSign to reproduce. If a lifeSign uses too much energy or loses it by being attacked it will ‘die’ and be removed from the world.

Meaning: all lifeSigns have a list of ‘meaning vectors’ (see also section 6.4) from which their overall meaning is determined by finding the highest value in the list. The amount of this value changes interaction within the system as one lifeSign may have the meaning ‘happy’ with a value of 100, while another may also mean ‘happy’ but only with a value of 10. In terms of the lifeSigns world, therefore there are ‘degrees of meaning’. There may be many different lifeSigns competing for the same meaning.

Compatability: compatability between lifeSigns is calculated using a comparison of their meanings. Friendly and aggressive behaviours may modify this compatability.

Behaviours: Interaction between lifeSigns occurs through behaviours defined in the system. Each atom in the lifeSigns structure is assigned a behaviour, resulting in different combinations that define its overall strategy. One lifeSign may have a combination of ‘leech – befriend – attack’ while another may use ‘mutate – command – join’. The effect of each of these is described in more detail below.

These properties are manipulated by interaction between lifeSigns and through the actions of players within the installation. The primary methods for these interactions are the behaviours, reproduction process, meaning vectors, formation of sequences and play.

6.2 Behaviours

Each lifeSign has a unique combination of behaviours (see Figure 7) that it may use to interact with others in the world. During interaction, different behaviour will be selected depending on the compatability if meaning between two lifeSigns – if their meanings have relevance to one another they will select friendly, helpful behaviour, if they are not compatible then they will select aggressive, destructive behaviour. An inbetweem state may also be achieved where the two lifeSigns have a neutral relationship. Graphic symbols indicate these states on screen during interaction as depicted in Figure 9.

Figure 7. lifeSigns behaviours

These behaviours function as follows:

Attack: decrease the energy of target.

Reproduce: produced offspring with target using a genetic algorithm.

Leech: drain energy of target and transfer it to self

Befriend: express positive compatability to target

Message: state main meaning vector to target eg. ‘I am truth’

Command: tell target to perform particular behaviour

Join: connect to target to expand or create a sequence

Mutate: randomly change one of the audiovisual parameters of target

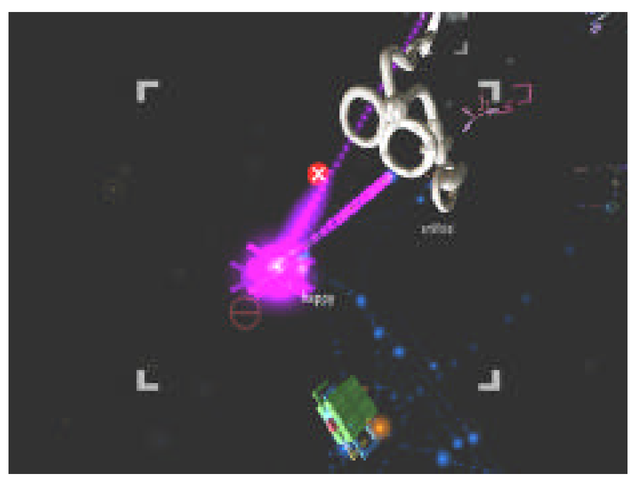


Figure 8. lifeSigns behavioural interaction

The interaction between lifeSigns occurs through the exchange of action signals – this process can be seen occurring in Figure 8. Each behaviour is weighted according to its 'strength' that is determined by how many times it is assigned within the overall 'strategy' of the lifeSign. Action signals are exchanged and may succeed or fail when they reach their target depending on these factors. The dynamics of interaction between the various 'strategies' create different situations between lifeSigns within the world.

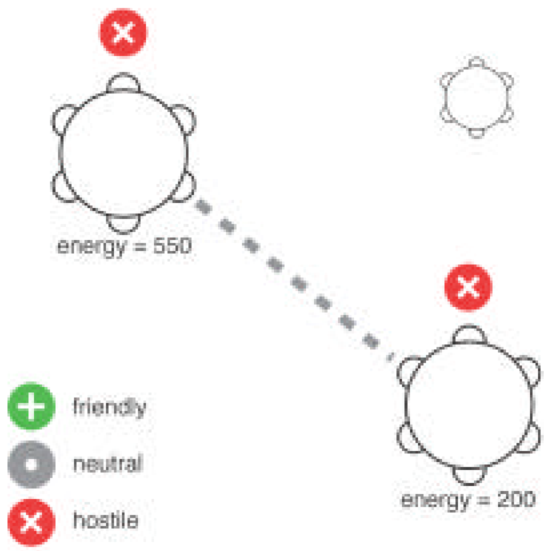


Figure 9. lifeSigns interaction model

6.3 Reproduction

The reproduction behaviour is a special case. If this behaviour is successful, two lifeSigns will produce offspring that combine features of their parents. As features such as structural constraints, build rules, image and sound parameters, behaviour and meaning of each lifeSign are stored in a dna string this can be achieved using a genetic algorithm. The dna string of each parent is used to create a new string combined at a random point of crossover, so that each new lifeSign is a unique blend of its parents.

When the simulation is started, a random batch of lifeSigns is generated. Over time, several generations of this initial population will be created through the reproduction process. Those lifeSigns that are best adapted to their world will survive and reproduce. In this case the parameters for survival include their 'strategy' (a successful mix of behaviours), significance in terms of the meaning system within the world, and being attractive to the players (see also 6.5 Play).

6.4 Meaning Vectors

Each lifeSign has a list of 'meaning vectors' that defines its meaning in the world. This is a list of sixty-four words that describe actions, attributes and structures that are possible in the space. Associations between all of these meanings are expressed as percentage values in terms of the degree of compatibility between meanings. A small network of meanings is depicted in Figure 10 that shows the relative compatibility of a lifeSign with the meaning 'artificial' with other meanings from the list.

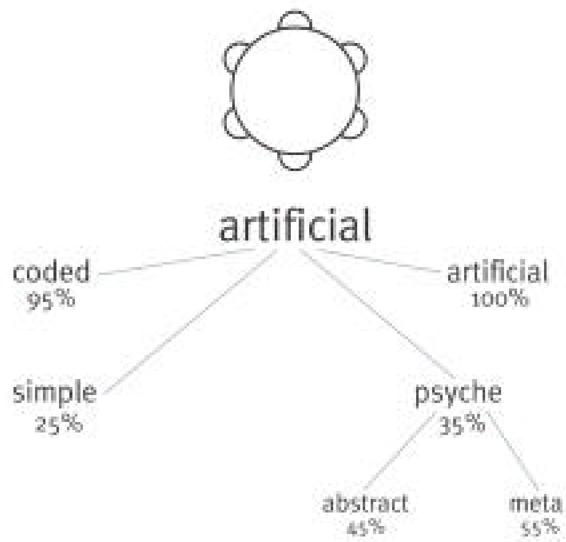


Figure 10. 'meaning vectors' and compatability

There are two main methods by which these lists may be modified. Firstly, the player may directly assign a meaning to a lifeSign in the world by selecting it from an onscreen list. This has an immediate effect, setting the players selection as the meaning with the highest value in the list. However, the process is also cumulative so that previously assigned meanings are also stored in the list, giving each lifeSign multiple 'degrees of meaning' in terms of the system. The second method is via an archive of lifeSigns stored on data-driven web site, separate to the installation space. On this web site, meanings may also be assigned but the process is more democratic in that each user may submit a meaning assignment for a lifeSign and the net result of all user submissions is tallied to arrive at the final list. This list is subsequently synchronized with the installation.

Furthermore, the *lifeSigns* web site supports the gallery installation of the work by archiving and documenting popular and successful signs and symbols that have been generated by the world. Users may search networks of meaning within a database and influence the meaning of icons by voting on their 'meaning vectors'. The web site is an ongoing 'research tool' for exploring the languages that emerge from the *lifeSigns* world.

6.5 Formation of sequences

The join behaviour allows the formation of sequences of compatible lifeSigns. This is indicated by dotted lines in the space showing the links in each sequence. Sequences perform the role of small 'communities' of similar lifeSigns that are able to share resources such as user play input. Being together also makes them more resistant to aggressive behaviour and increases the probability of friendly interactions between neighbouring lifeSigns.

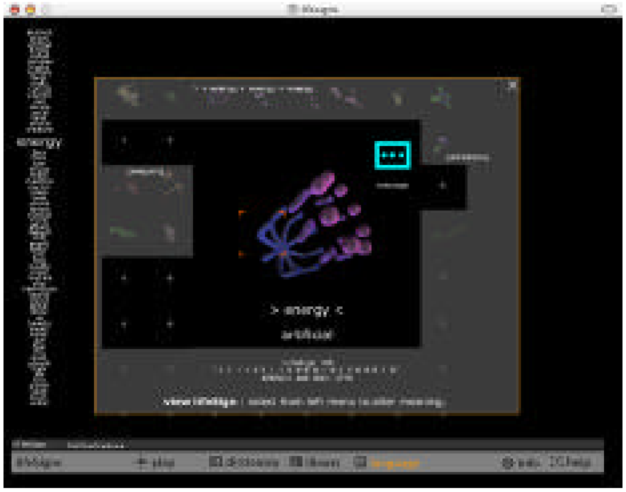


Figure 11. lifeSigns web site

The distribution and population of sequences in the world changes dynamically over time. Some links persist for long periods, other dissipate quickly either through the intervention of players or changes in the relationships between the lifeSigns. The clusters that form are visible on the map view as groupings of similar forms, labeled with the highest meaning in the sequence. In this way, sequences are a useful index to popular and successful meanings in the world – simply by viewing the map it is easy to see the current state of the simulation in terms of both population and distribution of meaning.

6.6 Play

As the players navigate the world, the lifeSigns that they select and play receive energy. More play results in more energy. This increases the likelihood of its persistence in the world and the production of offspring. The process of play itself is rewarding also for the player – this is described in terms of gameplay within section 7.

As a result, the survival of lifeSigns is dependent on this process of 'aesthetic selection'. Those that receive the most user input thrive and reproduce – their form and meaning spread throughout the world. Others that receive little or no user input eventually use up all of their energy and die, making space for the offspring of more successful lifeSigns.

6.7 Outcome of the system

New codes and communication forms may potentially be evolved using genetic algorithms and iterative processes. This provides a methodology for searching the solution space of a given set of semiotic relationships translated into an alife model. Signs and symbols may have agency within a simulation, such as a digital game, to evolve and adapt to suit the conditions of a particular environment. They may modify themselves in reaction to data collected in relation to factors such as their interpreted meaning, popularity or patterns of use.

7. PLAYING THE WORLD

The interface and play of the lifeSigns world are essential in the experience of the world. It is a world that needs to be played and explored in order to perform its function as a generator of iconographic languages.

7.1 Interface/sign system

The interface pictured below is used to play the work. It combines a trackball, used to navigate the world / select icons, and four 'touchpads', used to play the lifeSigns as audiovisual instruments.

Figure 12. lifeSigns interface

To navigate the world, the player may orientate their view using the trackball. The left button causes forward movement through the world. With a lifeSign in view, clicking the right button allows you to 'acquire' that icon. The HUD changes to show you have a lifeSign acquired.

In this mode the controls change. Trackball movement causes rotation about the lifeSign, the left button zooms the viewpoint, and the right button 'deacquires' the lifeSign, returning you to the navigation mode described above.

When an icon is acquired it can be played through the interface, an AKAI MPD16, mapping four of the pressure pads to the four inputs of the icon. When an input is activated the image and sound parameters are performed, following the rules of propagation described earlier.

7.2 Signals

The inputs become signals that travel through the icon and result in further animation and sound. These signals leave the icon and can collide with other nearby icons causing them to play as well. Through this process the user can play the icons like an instrument.

The process of signal transmission is depicted in Figures 13 and 14. Once signals are propagated through a lifeSign, they are transmitted to the next adjacent lifeSign triggering a performance of its image and sound parameters. Multiple signals can be generated, dramatically increasing the image and sound activity in the world as the signals are transmitted across multiple lifeSigns. A hidden feature allows the player to also move and play a lifeSign at the same time, allowing it to move through space spreading signals. Within the installation, up to four players can perform at once allowing multiple performances to occur simultaneously.

Figure 13. lifeSigns at play

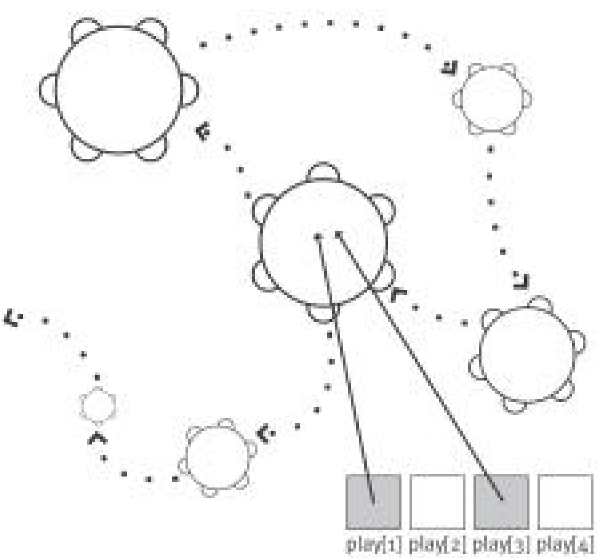


Figure 14. lifeSigns: signal transmission process

8. CONCLUDING REMARKS

8.1 lifeSigns as generative meaning system

lifeSigns allows engagement with the process of 'playing a world'. Interfacing the player as an active agent in the generative meaning system is achieved through the fluid, mutable representation of digital game worlds. The structure and relationships defined in a sign system are given a literal, tangible form in the simulation of the game world. As the simulation itself is essentially a network of signs, the world becomes both interface and representation of its underlying logic and evolutionary processes. So, in general terms the world as sign system is translated across three realms; its behaviour is defined in terms of artificial life / generative systems, its meaning defined in terms of computational semiotics, and its representation manifests in the form of digital games.

As the players may interpret the meaning of lifeSigns in the world, and the cumulative input of all players is collated to determine the final meaning of each lifeSign. This process, combined with the system for generating icons, demonstrates a protoypical 'generative meaning system'. It combines the capacity of the computer to generate a multitude of novel forms using a generative system with the collection of data from human participants to interpret the meaning and significance of the forms generated.

9. ACKNOWLEDGEMENTS

lifeSigns was produced in association with the Film Victoria Digital Media Fund, the Australian Centre for the Moving Image (ACMI), Monash University and Renderware. Troy Innocent gratefully acknowledges the collaboration and contribution of Nicholas Sandow, Shee Zon Chen, Hou Leng Tan, software development/programming; Jeremy Yuille, interactive sound design; Steve Law, sound design; Steve Taylor and Harry Lee, database design/coding.

10. REFERENCES

[1] Goguen, Joseph. Semiotic Morphisms. http://www.cs.ucsd.edu/users/goguen/papers/sm/smm.html Access: February 2 2002

[2] Innocent, Troy. Exploring the nature of electronic space through semiotic morphism. MelbourneDAC, the 5th International Digital Arts and Culture Conference. May 19 - 23, 2003

[3] Innocent, Troy. The language of Iconica in Dorin, A. & McCormack, J. (eds), First Iteration: A Conference on Generative Systems in the Electronic Arts, CEMA, Melbourne. pp. 92-104.

[4] Koza, John R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. USA: MIT Press, 1992

[5] Liungman, Carl G. Dictionary of Symbols, Norton, USA 1991

[6] Manovich, Lev. The Language of New Media, MIT Press, USA 2001

Cyburban Semiotics

Timothy J. Jachna
School of Design, the Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
and
Spatial Information Architecture Lab, Royal Melbourne Institute of Technology, Melbourne, Australia
sdtim@polyu.edu.hk

ABSTRACT
Through interpretation from observations of everyday practice in the use of urban space and digital communications technologies, this paper examines the ways in which the superimposition of digital communications networks on the spaces of cities is influencing the way in which city dwellers socially construct and maintain a symbolic relationship with the city and, through the digitally mediated physical space of the city, with one another.

Keywords
Urban semiotics, digital communications semiotics, interface, urban sociology, spatial perception

1. INTRODUCTION
The notion of 'the urban' emerged from the unique conditions of the city as a living environment, and is associated with semiotic density of physical culture in space. The expansion and evolution of urban regions have historically been facilitated by technological advances – from the iron skeleton frame to broadcast technologies – that have transformed the physical, experiential and semiotic fabric of the city. Though the latest generation of technologies can be depicted as yet another step in this progression, they are also experienced not merely as enabling technologies or connective networks, but as potential sites in their own right, with patterns and structures that often transcend, subvert or modify those of the physical city [5]. The superimposition of digital technologies upon the physical space of the city is restructuring the urban experience and overriding or altering urban semiotic systems and patterns. The digital layer of a city's infrastructure is causing a re-formulation of 'the urban' as a semiotic system. If 'urban' denotes the intensely networked, meaningful living-together and communication of large, heterogeneous groups of people, then the virtual network must be taken as an *a priori* fact of the urban condition in our time, and the most recent layers of spatial and relational experience must be seen as integral facets of the urban experience and the 'space' of the city.

The field of urban semiotics sees the city as a text built on grammars of spatial structures and patterns with semiotic significance. This paper does not pretend to offer a generalized theory of the emerging semiotics of the mediated city, but aims at demonstrating, by examples, the ways in which digital communications media are associated with radical change in the way cities are lived and symbolically understood. This modest demonstration shall concern itself primarily with the two most pervasive and commonplace of digital forums – the cellular phone network and the Internet – and is intended as a survey of actual present lived reality rather than a prognostication or manifesto as to where the trajectory of these trends will, or should, end up.

The analysis starts with observations as to how people 'use' the city, in the broadest definition of the term, and examines the most basic, categorical and generic semiotic structures and practices that come into play in urban life and how they are influenced by digital technologies. The semiotic structures and practices investigated are: addressing systems, wayfinding in the city, categories of urban space, and the space of community. Through examples drawn from the present, the paper demonstrates how digital technologies are transforming the semiotic landscape of cities and defining new criteria and dimensions for the 'legibility' of urban space.

2. THEMES

2.1 Address
An addressing system – be it for urban space or computer memory – is a semiotic tool that serves as an interface that allows for ease of control, navigation and access. Cities are overlaid with multiple addressing schemes, each of which 'formats' the space of the city in a different way to allow users to make use of it. In the 20th century city, street addresses and telephone numbers were the two systems of denotation by which city dwellers most habitually conceptualized and instrumentalized their interactions with urban space and, through it, with each other. Because this paper is not concerned with the semiotics of digital media in isolation but rather with the semiotic practices that arise out of their superimposition on the physical space of the city, the appropriation of the address metaphor for the Internet will not be explored in detail. Of direct relevance for the discussion at hand, however, is the way in which digital communications technologies, like many technologies before them, have symbolically and functionally structured urban space by adding additional layers of addressing upon the city.

While to say that digital technologies in the city – most predominantly the Internet and the cellular phone network – sever the link between an address and a location would be an oversimplification of the issue, the increased mobility, personalization and 'placelessness' of these new layers of addresses has changed the way in which the symbolic relationships between place, person and address in the city are lived and perceived. It is not technologies as such, but rather the patterns of their deployment and employment, that impress themselves upon the semiotics of lived urban space. Cellular phone networks and Internet-based communication are used in

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

ways that support and fuel an increasing focus on the individual, rather than the household, business or neighborhood as the primary social unit of the city. Digital communications allow this primary meaningful social unit to become the primary unit of connectivity [12]. Thus, imposed upon the text of the city as a set of locations (occupied by people) is another layer of addresses corresponding to people (occupying locations), and this latter layer is quickly becoming the space in which most urban dwellers find it useful and meaningful to live their lives and conduct their interactions with one another.

Urbanites increasingly use communications technology to address (speak to) another person directly rather than searching at a specific address (location) at which they hope to find this person. Whereas the typical opening question asked by a phone caller on a fixed line may have been "Is ___ there?", a caller to a mobile phone is more likely ask "Where are you?". Both questions aim at ascertaining the conjunction of a person and a place, but the primary frame of reference is switched. Far from relegating the physical location in the city to irrelevance, the cellular phone is used in ways that allows for real-time coordination of agendas and schedules of people in real-space. Addresses remain the interface to the space of city, but the flow of addresses *cum* people through space (as probes) replaces the fixity of address on a place. The spatial model of digital communications technologies defies definition by an abstract representation of the system, such as a street map or telephone book. Rather, it is contained in the millions of simultaneous locational-information-laden messages that define the state of the system at a given moment.

The conventions used in addressing within digital communications forums introduce other levels to the interplay between address and place. Because cellular phones typically take the area code of the city in which they are registered, users continue to be 'located' in the telespace of their home city, even if they are physically thousands of kilometers from home. Also, the 'at' (@) in Internet addresses introduces a type of *faux* locatedness. Increasingly, at any given time, a person is 'at' several addresses, with the "at" implying different degrees of personalization and locational fixity in each of the various addressing systems (see table 1).

Table 1. Conventions of Addressing

address type	Street address	Telephone number	Cell phone number	E-mail address
addressed entity	plot of land	household, or business	person	person
locational specificity	high (absolute location)	middle (city, district)	low ('base' only)	none
personal specificity	none	middle	high	high
mobility	none	middle (can move withir area code)	high	high

Though addresses in any city are essentially abstract 'tags' denoting a location, they take on a second, connotative, level of signification through conjunction with the qualities of the city on which they are imposed. Streets, street addresses, post codes and telephone number prefixes come to signify not just a location in geometrical space but also carry connotative associations of status, function or history. One need only think of Beverly Hills 90210 in Los Angeles, 5th Avenue in New York or #10 Downing Street in London. Because they are not location-fixed, Internet addresses and cellular phone numbers do not have the potential to support this connotative level of signification. Without prior knowledge of a person, one can make no assumptions about them from their address.

With prior knowledge of a person, however, the digital address becomes connotative not as a stereotype but as a hieroglyph standing concretely and specifically for the person addressed. Addresses become much less like coordinates and much more like names. In a sense, the space of the city is collapsed to a single space, in which one does not have to search in multiple locations for a person, but just call the person's name. The extent of changes in the role-division between spatial semiotics and conversational semiotics as the languages of interaction with and within the city is a topic for another paper.

In a chapter of his book "the Empire of Signs" – a classic text of semiotics – Roland Barthes describes the street addressing system of Tokyo in which streets have no names, and addresses of buildings have no necessary relation to their spatial relations to one another. The rationalized, spatially sequential addressing system of Western cities provides urbanites with a cognitive 'filing system' of sorts, within which spatial locations and relations in urban space can be conceived in the abstract, removed from the actual physical fabric of the city itself. Finding a specific shop or residence in Tokyo, however, relies on experience, memory, engagement with the physical space of the city and exchange of information with other people. The relevance of street addresses as a means of identification in the city is suppressed and the fixed connection between address and relative location, taken for granted in the West, is exposed as imposed upon, rather than inherent in, urban space [1].

As conventional street addresses and fixed phone lines become less relevant and useful for the manner in which people use and understand the city, the phenomenological practice of inhabiting the city described in the case of Tokyo seems to apply increasingly to digitally mediated cities in general. In Barthes' Tokyo this arises from a city of fixed but essentially unaddressed locations. In the modern mediated metropolis, this is the effect of a city of addresses unfixed in space. Each person has become an address unto themselves and the experience of reading the city becomes one of constantly finding and re-finding one another.

2.2 Wayfinding

In "the Image of the City", Kevin Lynch proposed a basic grammar of urban space that he claimed was at the core of what makes urban space 'legible' to a user. His grammar of the city image was based on five basic classes of elements: paths, edges, districts, nodes and landmarks [9]. Gottdiener and Lagopolous criticized Lynch for implicitly reducing the city to the sum of its physical spaces and structures, and for constraining his understanding of the process of using or experiencing the city to the act of moving bodily through the spaces of the city [3]. For all its acknowledged narrowness of scope, Lynch's approach is one of the clearest formulations of what may seem at once the most superficial (in the sense that it describes the clearly visible) aspects of urban semiotics yet also the most difficult to dismiss: the question of how people find their way in urban space and conceive the space of the city and their position within it.

Curiously, Lynch's urban vocabulary of paths, edges, districts, nodes and landmarks can be applied to the description of digital networks as well as cities, with the major difference that, when applied to digital networks, the description remains in the realm of the abstract, because it does not in any way describe the network as experienced by the user. These elements exist in an abstract world that is operated from outside, not a space *within* which one operates. However, it could be argued that, for some of our modes of interaction with the city, the physical space and structure of the city remains as abstracted from our experience of using it as the geometry of the digital network is from our experience of using the Internet or cellular phone.

While Lynch's elements certainly do make up a grammar whereby cities *can* be described, it is debatable whether they correspond to the most meaningful categories by which actual urban citizens *do* perceive and use the space of their city, at least in the modern city. City dwellers are certainly used to shifting between different modes and conventions of interface with the city, of which bodily movement between locations in the physical space of the city itself is only one option. Many of our symbolic navigations through the city take place in the space of abstract projections of the city rather than the streets and squares of the city itself.

As discussed above, we are continuously conditioned by our technologies to conceive and use the city as a set of discrete addresses rather than a set of locations. The Internet presents us with the paragon of the 'pure' address, not attached to any physical space or entity in any experiential sense, signifying itself and the path that leads to it, and the influence of the expectations represented by the use of Internet technology resemble the way contemporary urban citizens tend to cognitively structure their use of the city. The central question one asks in using the modern city is often not "where is it?" but rather "how do I access it?", in other words not where it is located in space but rather which 'path' one needs to call up to connect to it. A path is understood not as a line through urban space that one must traverse but a dimensionless connection between oneself and that which one wants to reach. Not only communications technologies as such, but also transportation infrastructure like expressways and subways, are designed as paths of this second type, with little or no contact with the urban space through which they pass, serving the sole function of joining a network of access points.

The cellular phone has been termed the "compass and beacon" of the users of contemporary cities [7]. By this analogy, the landmark takes precedents above the other categories of Lynch's city image. The landmarks by which one orients and guides oneself through the city are not fixed structures but fellow urbanites (with their cellular phones) who are themselves mobile. Townsend has hinted that navigation in websites may be a model for how people navigate the city "through intangible information cues" [11]. A web site is as much a logical experiential array of meaningfully-linked spaces as any example of architecture or urban design. Though Townsend himself does not elaborate more specifically on the website metaphor, one could follow his lead to equate other cellular phone users in the city as potential 'hotlinks'. The digital mediation of the practice of wayfinding through the space of the modern city is, perhaps ironically, re-introducing an emphasis on the phenomenological and real-time (as opposed to abstract and fixed) dimensions of physical spatial experience. This gives occasion for a reassessment of the relevancy of schemes like Lynch's, which propose that meaning is generated and perceived in the visual and sensory experience of the user of the concrete physical spaces of the city.

2.3 Categories of Space

Although the urban experience is semiotically complex and many-dimensional, semioticians have suggested that the language by which cities signify – the semiotic essence of 'the urban' – can be distilled into terms of binary oppositions such as Ledrut's pairs of ethical, vital, aesthetic and functional values (liberty/constraint, well-being/uneasiness, beautiful/ugly and functional/non-functional respectively) [8] or Greimas' 'axiological micro-universes' (i.e. society vs. individual, euphoria vs. dysphoria) [6]. The suffusion of the physical space of the city with layers of digital technology can be seen to affect all of these oppositions in both subtle and drastic ways. However, I shall focus on two sets of terms that are fundamental to the way in which urbanites perceive the structuring of urban time and urban space, and by which they conduct themselves and lead their lives. These oppositions are those of work/home and public/private.

Each of these archetypes of place is associated with certain activities, rules and codes of behavior. Each place provides the context within which one persona of a person unfolds, based on his or her role within that context: a public persona and a private persona, a work persona and a home persona. A person's enculturation into the urban world is in part a process of learning to read the city by means of such categories and adjust one's expectations and behavior accordingly.

The categories of 'work' and 'home' are referents structuring the time and space of urban lives, in urban time and urban space. The 20th century was characterized by increasingly clear and absolute distinctions between the place of work and the home place, in terms of location, aesthetics, spatial temporal and social organization and roles. There was a space of work and a space of home, and a time at work and a time at home, constrained in space by zoning regulations and in time by the regulated workday, kept at a safe distance from one another in time and space by the necessary commute from one to the other through the space of the city.

Famously, the increasing number of people who work in the so-called knowledge industries, being tied neither to a fixed physical infrastructure nor a specific physical location by their work, make up a growing class of highly mobile professionals. Digital infrastructures, and the related gadgets such as fax machines, laptop computers and of course the ubiquitous cellular phone, play an instrumental role in enabling this lifestyle which has spawned new hybrid typologies of spaces such as the home office and the 'new office' and new spatial practices such as 'hot-desking'. The category 'workplace' ceases to be associated exclusively with a given place, space or position and becomes defined tautologically as the place in which one happens to be working. Much more, it comes to be associated with a set of increasingly mobile and personalized tools and, intangibly, with a state of mind. 'At work' ceases to signify a place and remains as a signifier of a 'state' that could be anywhere and is, increasingly, everywhere. Work and non-work are modes rather than places: shifting qualities of people rather than fixed qualities of spaces.

The sociologist Ray Oldenburg has pointed out the importance of so-called 'third places', neither home nor work, that serve as social condensers in which modes of social interaction take place which are crucial to the maintenance of community and the formation of individual identity [10]. Bars, hairdressers and even prisons fall into this category. In lauding these spaces he also bemoans their disappearance from the urban space of the late 20th century America in which he was writing, due in part to zoning laws that encouraged spatial use segregation.

If the office phone is a metonymy for the workplace, and the home phone for the home, then the cell phone can be seen to constitute a type of 'third place' [7]. Physically, the third space is now potentially anywhere and everywhere: it has subsumed the first and second places as well as all the non-places between. However, the cellular phone network as a 'third place' comes at a price. With a cellular phone, one is potentially always on call. The time and space structure of the day can be interrupted and re-arranged at any time. The 'third place' becomes effectively a potential space of surveillance rather than retreat, reversing the accustomed relationship. The cellular phone also blurs the boundaries between a person's different roles. Certain personal calls would not be made to the office phone and one may hesitate to disturb someone 'at home' (i.e. on the home phone) with a work-related issue. However the cellular phone, being tied to a person rather than a place, presents the ever-present possibility of invasion of one role into the space of another role.

The effects of this become especially evident in the ostensibly 'public' spaces of the city, blurring and complicating the categorization of a space as unambiguously private or public. People in public space on cell phones are communicating, as one is expected to do in public space, but not with those in the same physical space [12]. Indeed, the interactions that they are having may be quite private in nature, of the sort that would not be deemed appropriate 'out in the open' of public space if the conversation were being held with someone standing next to them. In the traditional sense of the terms, public space is understood as the space of communication while private space is the space of withdrawal, but this type of non-co-located private communication in co-located public space is an illustration of the interference patterns caused in categories of space by the superimposition of digital/virtual and physical spaces of communication. It has been commented that public space is thus being fragmented into many private spaces, to the extent where it becomes a 'common living room' [7]. Like the role 'at work', the role 'at home' becomes a mode into which city dwellers may switch at any time, regardless of the space in which they currently find themselves.

2.4 Community

For Kevin Lynch, the district was a primary conceptual category by which one 'reads' the urban fabric [9]. A district refers to a spatially contiguous area within the city that forms a functional and social urban sub-unit, distinct from other districts around it. In common parlance, the word 'neighborhood' probably better captures the sense of identification the actual urban citizen is purported to feel for the district in which he lives and to which he thereby belongs. Even at the time of Lynch's writing, though, the neighborhood as a meaningful semiotic category of urban life was already fading, in that the ostensible 'districts' within many large cities and their growing suburbs no longer corresponded to a social or functional unit of the city. While the district perhaps remained tenable as a descriptive category of surface appearances, it was seen less and less to signify any meaningful social or functional unit.

Even decades before the advent of digital technologies, the typical city dweller would not have seen their social 'community' as identical with their spatially surrounding 'neighborhood'. A sense of community lies in affinity, not proximity, so technologies that enable one to seek, find and maintain meaningful multi-modal contact with far flung people with common interests will be used to the fullest extent of their capability. In the digital dimension, everyone on the network is nominally equally accessible. The group with which one associates becomes more a matter of choice than chance.

Networks of community independent of spatial proximity have always existed [4], but are now being brought to the forefront as the digital technologies that offer the affordances for these communities take up a central role as *sites* (not merely channels) of community interaction. As the physical neighborhood becomes drained of semiotic content, the website as an anchor for a virtual community of choice become (over)invested with it, while online forums such as chat rooms, bulletin boards and newsgroups, with the accompanying websites, form virtual 'third places' with a much higher degree of 'defensibility' against unwelcome incursion than the cellular-phone-as-third-place. While some of these online forums are 'home' to communities who rarely or never meet physically, many others are sites for the support or extension of the communications and interactions in physical space. A website performs the semiotic role in virtual space that the 'neighborhood' may once have performed in physical space. Namely, it presents an integral and self-contained visible presence that may be visited by outsiders but 'belongs' to members of the community that it simultaneously houses and signifies.

The point of access to the communication spaces of the Internet is typically from a computer within the private home, bringing the realm of (digital) community interactions into the (physical) home. This phenomenon is the inverse of the fragmentary privatization of public space by cell phone users mentioned above, with potentially similarly disruptive consequences. The extreme variant of this turning-inside-out of urban space can be seen in the proliferation of pseudo-neighborhoods, termed secessionary 'network spaces' by Graham and Martin [4]. Inward-looking and closed to their surroundings, these housing developments, office complexes, multi-use projects and other building typologies withdraw from interaction with the physical space of the city whilst opening themselves up to the elite neighborhood of similar enclaves around the globe. They are literally connected more to the networked worldwide infrastructure than to their surroundings through the privileged position in spatial patterns of distribution of access to digital networks. Digital communication technologies are certainly not the sole cause of growing polarization in cities, but the distinction between the digital 'haves' and 'have nots' tends to reiterate and exacerbate previously established wealth relations in the city. The neighborhood of the favorably networked expands to a truly global elite neighborhood that exists in dimensionally transcendent virtual space.

3. CONCLUSION

Explorations in urban semiotics must acknowledge that a city is not, in any obvious sense, a message with a sender and a receiver. Ledrut reminds the reader that "the modern city is semanticized by the fact of its social production and use rather than by any communicational intention" [8]. However, it is becoming increasingly pertinent to see the constitutive fabric of the city as being composed of, and by, the exchange of uncountable messages in real-time. A visual bias still characterizes our way of perceiving and conceptualizing the world, supporting a lingering preconception of the city as a physical environment *within* which urban life unfolds rather than the space that unfolds in real-time as an emergent characteristic of more fleeting and less tangible patterns of urban life.

In attempting to formulate a general approach to urban semiotics, Gottdiener and Lagopolous discredited cognitive geography as a way to define the image of the city. They claimed that this method 'asked the wrong questions' and had no clear mechanism for building-up an idea of the communal image of a city from the collection of individual highly personal 'mental maps' [3].

Neither the substance and structure of the physical city nor the messages and forums of the digital urban fabric are sufficient, as stand-alone texts, as an adequate basis for the 'reading' of the contemporary city. As has been demonstrated in the examples above, the day-to-day construction of meaning between urban inhabitants and their city involves a constant interweaving of mediated and unmediated communications and interactions.

One could say that the semiotics of urban life in post-Second World War Western cities was based on one-to-many modes of mass-communication, both in the communications media, epitomized by television and radio, and the physical structures of cities, whose production was driven by programs of mass housing and the paradigm of urban planning as traffic planning. The cities of today are being lived, read and written according to a different media paradigm: the many-to-many structure of the Internet and the cellular phone network. More than any communications media of the past, the use of networks of digital communication are becoming inextricably interwoven with the use and perception of urban space, and vice versa, restructuring and redefining spatial and experiential continuum of urban life and reformulating the field of urban semiotics as cyburban semiotics.

4. ACKNOWLEDGMENTS

The author gratefully acknowledges the support of his colleagues at the School of Design at the Hong Kong Polytechnic University and the Spatial Information Architecture Lab of the Royal Melbourne Institute of Technology.

5. REFERENCES

[1] Barthes, Roland (English translation by Richard Howard). The Empire of Signs. Noonday Press, New York, 1982.

[2] Castells, Manuel. The Informational City. Blackwell, Oxford UK, 1989.

[3] Gottdiener, M. and Alexandros Ph. Lagopolous. Introduction. The City and the Sign: An Introduction to Urban Semiotics. Columbia University Press, New York, 1986, 1-21.

[4] Graham, Stephen and Simon Marvin. Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition. Routledge, London and New York, 2001.

[5] Graham, Stephen and Simon Marvin. Telecommunications and the City: Electronic Spaces, Urban Places. Routledge, London and New York, 1996.

[6] Greimas, Algirdas Julien. For A Topological Semiotics. in Gottdiener, M. and Alexandros Ph. Lagopolous. The City and the Sign: An Introduction to Urban Semiotics. Columbia University Press, New York, 1986, 25-54.

[7] Kopomaa, Timo. Speaking Mobile: The City in Your Pocket. In YTK's Electronic Publications. http://www.hut.fi/Yksikot/YTK/julkaisu/mobile.html

[8] Ledrut, Raymond. The Images of the City. in Gottdiener, M. and Alexandros Ph. Lagopolous. The City and the Sign: An Introduction to Urban Semiotics. Columbia University Press, New York, 1986, 219-240.

[9] Lynch, Kevin. The Image of the City. The M.I.T. Press, Cambridge MA, 1960.

[10] Oldenburg, Ray. The Great Good Place: Cafes, Coffee Shops, Bookstores, Bars, Hair Salons, and Other Hangouts at the Heart of a Community, Marlowe & Co., New York, 1989.

[11] Townsend, Anthony M. Life in the Real-Time City: Mobile Telephones and Urban Metabolism. Journal of Urban Technology, 7, 2:85-104.

[12] Wellman, Barry. Physical Place and CyberPlace: The Rise of Personalized Networking. International Journal of Urban and Regional Research, 25 (2001).

Sounds and Sources in *Sacred*

Kristine Jørgensen
Ph.D. student
Division of Film & Media Studies
Copenhagen University
kristinej@hum.ku.dk

ABSTRACT

This paper looks at sound as a representational system in computer games. Being part of a Ph.D. project on the functionality of sound in computer games, this paper does not present a semiotic answer to how we should understand the relationship between computer game sounds, sources, and the real world. Instead, this paper asks several questions concerning computer game sound and its relationship to its source in the game world, and to sounds in our real world environment. Since computer game sound often has important implications to player actions and reactions, it has a special relationship not only to the player, but also to its source and the game environment. This relationship poses several problematic questions about referentiality and how sound may be part of a representational system. Therefore, game sound as representation is evaluated in relation to the functionality of sound in computer games, not at least according to how sound may influence game play. As an empirical case study, I do an analysis of the soundscape of the recent computer role-playing game *Sacred* (Ascaron 2004). The game sound is studied according to whether it relates to onscreen or offscreen sources. In addition, Clive Fencott's concept of perceptual opportunities in virtual environments will be used as analytical tool in order to describe the functionality of the sound in this game.

Keywords

Computer games, perceptual opportunities, representation, *Sacred*, simulation, sound.

1. INTRODUCTION

All those who have tried playing a computer game with the sound turned off know that sound is an extremely important aspect of the game environment, not at least in relation to successful game play. Game designers consciously utilize sound to influence the game experience and the player's behaviour in different ways, by giving the player information crucial to game play performance. In this sense, sound in computer games is highly communicative. Also, sounds in computer games are pre-produced and synthetically added to the visuals in the game. Thus, no sounds are present "by accident" in the way possible in films, where sound may be directly recorded on location. This makes it easy to see sound in computer games as representational. However, seeing sounds as representational also poses problems in several respects. This raises questions on what exactly sounds may be said to represent: do they represent sounds in the real world, or game-internal sources?

This paper will discuss whether – and how – computer game sound may be seen as a representational system. It will be asked what other kinds of relationships game sound may have to its environment, and in this way the paper tries to illumine the experience of sound both in games and in other contexts. In order to draw on a specific, empirical example, I will study the recent computer action-roleplaying game *Sacred* [1] with reference to what relationship there might be between the soundscape and its functionality in the game environment. Before going on to the analysis of the soundscape of *Sacred*, I would like to discuss how game sounds relate to its in-game sources and to sounds in the real world. The following section will then outline how sound is understood in this paper, and it will also be a fruitful theoretical starting point for a further examination of computer game sound as it will be understood in my Ph.D. project.

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

2. SOUND & SOURCES

Understanding sound as a representational system may seem intuitive. When we hear a sound in daily life, we connect it to its source, regardless of whether we can see the source of not. When we experience sound, we perceive it as causally connected to either a specific object (the duck says quack), or as the result of a process (sound from a motor) or an event (sound from an explosion). When we listen to sound with an analytical ear, we tend to separate sound from the source according to the senses through which they are perceived, namely hearing and sight. And since the source in general is a physical feature, and also more persistent than sound, we are easily led to see the sound as a property of this source; hence it becomes secondary to it. This may give us the impression that sound somehow points to its source: when we hear a sound, we think of its origin. This impression is not at least due to the fact that sound often comes into being as an effect of some kind of manipulation or physical strain towards a source. Hence, we interpret the sound as a representation of the source, since the sound becomes the auditory effect of a process or event arising from the source. In this sense, the sound points towards its source, thereby creating an indexical relationship between sound and source similar to the relationship between smoke and fire. Based on this thought, both in games and in the real world sound may be interpreted as a representation of its source. But we should keep in mind that this is a purely analytical understanding of sound. In everyday experience we do not perceive sounds as distinct or separable from their sources; instead the sound and its origin are experienced in coexistence, where the one cannot exist without the other. Although sound does not have the same physical properties as their sources normally have, sound is experienced as another aspect of the source, just as important as all other aspects, but perceived in a different manner. With this symbiosis in mind, it becomes problematic to say that sounds are representations of their sources.

However, there is one situation in which it may be possible to argue that sounds represent their sources, and this is when we do not have direct access to the source. Here we do not have any choice but experientially separating sound from its source in an analytical way. When we do not see the source of the sound, we have to make hypotheses about what its source may be, based on mental models and previous knowledge about similar sounds in the real world. Especially when the sound is not immediately recognized, we assign the sound to hypothetical sources, making assumptions about what effect the sources may have on us. The sound then becomes the representation of its source. While sounds with onscreen sources have a more direct connection to their sources since sound and visual objects are experienced simultaneously, computer game sounds with offscreen sources may in this sense be experienced as representing their sources. In connection with computer games, it is also important to point out that whether one can see the source of a sound or not creates different reactions in the player. Seeing the source, we get a greater sense of control of the situation, since sound and source merge into one experiential entity, and there becomes no doubt about where the sound originates from. But when we do not see it our reactions will be of a more suspicious and interpretative kind where we try to identify the source, not at least because we are in a game environment filled with monsters and potential dangers. The sound from offscreen sources triggers a sense of alertness in the player, and the sounds may therefore be said to point to both monsters and dangerous situations.

When we look at mediated sound, the feeling that sound is a representational system may become even more convincing, since how the sound is intended to work is the result of a conscious choice from the creators' viewpoint. In addition to imitating real world sounds, all mediated sounds are added for a purpose. Especially game sound creates a strong feeling of being representational. Since the visual environment itself is constructed as graphics on a computer, game sound cannot be directly recorded together with the visuals in the same way as film sound[1]. Instead it must be artificially assigned to the visuals, which obviously is quite different from recording sound simultaneously with images of its real world source. Thus, when a microphone is present sound may be recorded automatically. Where films may use on-location and direct sound that originate directly from the actual source, computer games must take sound from somewhere else, either by recording sound from the real world, or by synthetically creating an appropriate sound. When sounds and visuals are constructed in this way, what is presented becomes more stylized than what direct recording makes possible, and it is therefore also possible to modify and configure what one wants to present to a much greater degree. This makes computer game sound more communicative and purposeful than direct sound, and this adds to the feeling that computer game sound is a representational system.

However, it should not be forgotten that the sound one hears in a computer game, as well as in a film, *is* a real sound, and it is also experienced as such by the player or viewer. Not only is the sound an auditory stimulus for our ears just as any sound in the real world, also, during game play the player will react to the sound as s/he would react to real life sounds, at least as far as the early stages of game play is concerned. After repetitive game play the player will reach a certain phase of mastery or automation of the computer game. Not only until this stage may the player experience the sound as communicative instead of similar to real life perception experienced in present time, since the player now knows the game so well that s/he recognizes all situations and what sounds appear together with them and for what purpose [2]. But at this stage the game will lose its appeal since it has no challenges for the player any more. As far as sound is concerned, this often leads to players turning off the game sound and start playing their own mp3s instead of listening to the game music.

Above we have seen attempts of describing sound as representation of both real world sounds and of their respective sources, and although we partly may explain sound in this way, there is still something missing. Although sound in games may be seen as a representational system in some respects, it is problematic to say that the sounds are representations of real world sounds. What about the sound from a spell of magic? It is also problematic to say that the distinct sound from a cuckoo bird in *Lineage II* [3] is a representation when the sound has all the same qualities as the sound from a real world bird. The idea of representation does not cover everything that is in the experience of playing computer games, nor does it tell us anything about the functionality of sound in games. A representation of something tries to explain or depict the phenomenon in question, but only on surface level. A representation tells us about some characteristics, but it is never exhaustive. What is lost is the actual functionality, or the cybernetics of the phenomenon: how does it work in practice? It seems thus that a representation only presents properties of a phenomenon for us, and gives us general knowledge of it, but a representation does not teach us how to use anything. While representation focuses on interpretation, it is important to understand how sound may be an important contributor to mastering the game.

Perhaps it is better to understand game sound as *simulation* of real world sounds instead of representation? This is in accordance with how for instance Gonzalo Frasca [4] understands computer games in general. He emphasizes that simulations model the behaviour of objects and systems in addition to representing them, hence simulations give important understandings of the functionality of complex systems. This is especially interesting when studying game sound not only because sound is of prime importance for cueing player action, but also because the functionality of sound is more important than what sound may be said to represent. When discussing how simulation focuses on functionality, Frasca exemplifies by a reference to learning how to drive a car. The functionality of a car will be comprehended much better through a simulation than a representation, since the simulation will show in real time and by first-hand experience how the car works, while a representation such as an instructional map or a book only gives the learner a basic idea [5]. Translated to game sound, we may then say that the sound of an explosion in a game is a simulation when it has the same functionality as the sound of

[1] There are interesting exceptions, such as computer games where the visuals consist of photographic or filmic images.

a real world explosion, and when it makes the player act in accordance with it and use it for mastering the motor experience. If a sound is present as illustration, on the other hand, it may be explained as a representation. For instance, if I say “blam!” in order to tell someone what the sound from an explosion sounded like, this may be seen as a representation. But when we hear a the sound of an explosion in a game, it is a simulation because it expects the player to react to it in the same manner as someone would react in a similar real world situation. Thus, situation is important when deciding whether we talk of sound as a representation of a simulation: in isolation it may be easier to see sound as a representational feature than a simulation, since this takes away the importance of functionality. But when studying how the sound may function in a game, it will be more fruitful to see sound as a simulation system.

In short, it seems better to say that sound must be seen in relation to the whole simulation process in a computer game. Even in the case of offscreen sources, what is important is not what the actual source of the sound is, but rather how the sound functions in relation to game play: is it a warning, informational, atmospheric, or something else? This focus on functionality is also in accordance with real world perception. However, when we speak of the functionality of sound, it may be fruitful to draw upon James Gibson’s notion of *affordances* [6]. According to Gibson’s ecological perspective on psychology, any object that we meet in our environment will present for us ways in which we may interact with it. For instance, a door will present us the possibilities for opening and closing. It is hence *openable* and *closeable* in Gibson’s terminology. A chair is *sitable*, while a tree may be *climbable*. But although sound has the ability to influence action and behaviour, it is not straightforward to apply the concept of affordances to the issue of sound. For instance, in one respect the sound accompanying the green light at zebra crossings may be interpreted as *walkable* for the blind. Or in another respect, does the sound tell the blind that it is the street that is *walkable*? It seems here that what is the actual affordance of the sound is to give us information about another feature’s affordances. We may then say that a sound affords information about the functionality or state of things, hence it is *informationable*. Also, hearing sounds from a monster offscreen, the player assumes there is a *run-away-from-able* monster nearby. As long as the sound is offscreen the sound will have an affordance. But when the player hears sound from an onscreen object, sound and source merge into one experiential entity where we cannot speak of the affordances of sound. Instead we must talk of the affordances of the whole complex experience.

This emphasizes the fact that the sound has a special purpose other than merely representing other elements in the world. This may also be illustrated by another kind of auditory perception that has more direct affordances. These are startling sounds and other sounds that make us react intuitively on the basis of bottom-up processes triggered by a reactive part of the nervous system. Such sounds appear suddenly and have the effect of making an individual jump or otherwise being surprised; thus they afford startling. Sounds with certain loud, high-frequency and alarms that have such a piercing sound that one is not able to be in its vicinity without feeling physical pain also may be seen as having their own affordance, namely being *move-away-from-able*. The functionality of such sounds is to create a spontaneous and intuitive reaction, and in this respect it is dubious to talk about sound as representing anything, since there is no kind of interpretation activated in the comprehension of such sounds.

One may ask whether the discussion on affordances only applies to real world sounds, but as stated above, computer game sound is experienced in the same manner as real world sounds, and therefore these thoughts are also valid for games. However, we need to keep in mind what is mentioned above, namely that computer game sound simulates real world sounds by being functional. At the same time, since game environments consist of computer generated constructs instead of registration of physical events by microphone and camera lens, computer games have a certain representational dimension not present in more traditional audiovisual media. Therefore, it is hard to say that computer game sound cannot be seen as representational, at the same time as it poses several problems to claim that it is so. There are at least two answers to the question of sound as representational, dependent on situation. Game sound is in one respect a simulation of real world sounds, and no representation. This has to do with functionality. In another respect, it may be seen as representing the perceived source in the game when the source is offscreen. But since we also deal with functionality here, it is more fruitful and correct to say that sound cues hypotheses about its sources.

3. *SACRED:* AN ANALYSIS

Sacred is an action-roleplaying game, placed closer to *Diablo* on the continuum between the *Diablo* games [7] and the *Baldur’s Gate* games [8]. The game offers quests and monster fighting, and the avatar’s skills and statistics develop during the course of game play, but the personality of the avatar and mental development are not issues in the way we know from *Baldur’s Gate*. The game has an isometric perspective, where everything is seen from above so the graphics engine does not have to deal with perspective differences related to distance. The avatar is thus seen from the third person perspective, and from above. The sound dimension of this game does not seem to differ much from for instance *Diablo*; thus, it seems a good case study of games of this kind of genre. What is interesting, at least from the perspective of those interested in how sound is implemented into the game both on a practical and a functional level, is that most of the sound files are installed together with the game as mp3 files in a separate folder. This makes it possible to go into the different files and listen to them in isolation. They also have titles based on situation that make it easy to identify them. However, not all sounds in the game are available in this way. What lacks from the file folder are sounds related to specific objects – that is, sounds connected to game objects. This is probably due to the fact that these sounds are programmed into the objects themselves and not as separate sound tracks. This is interesting in the perspective of this paper, since it deals with the relationship between sounds and their sources. When sounds are programmed into their sources as if they were originating from them, the relationship between sound and source becomes very similar to the relationship between real world sounds and sources. Although the sounds originally have been pre-produced and assigned to a specific source instead of originating naturally from this source, the fact that they are implemented as a product of their sources makes the relationship feel more natural, since this is what we intuitively expect from sounds. This similarity between real world sounds and sounds in *Sacred* emphasizes the difficulty of dubbing game sound a representational system while real world sound is not. Also, when sound is programmed into objects, it more clearly becomes part of a simulation, where sound and source are merged in order to become as similar to the real world as possible.

However, there are 154 mp3 files available for listening, divided roughly into atmospheric sounds, jingles, sounds related to important non-playing characters (NPCs), and music. Atmospheric sounds are what game designers often call ambient sounds, namely sounds specific for a special area in the game. For instance, the forest will have a different soundscape than the desert, where the forest is dominated by the sound of wind in trees and animals, and the desert will have the sound of wind through an open landscape of sand. In *Sacred*, atmospheric sounds also differ somewhat from whether it is night or day. In *Sacred*, jingles are short[2] musical themes that appear together with a special event. For instance, there are jingles signalling danger, fights, and quests. There are also five short musical themes (about 10 seconds) associated with important NPCs in the game. Last but not least, there are several music tracks for different situations, ranging from one to four minutes. These may be associated with fighting, death, or whether the avatar is in a village or not. The different files are merged, mixed and repeated during game play, so that the player does not hear the same sound in all similar situations. As mentioned, in addition to these sounds ascribed to separate sound files, there are object-oriented sounds. For instance, there is speech from the avatar and NPCs, and there are sounds of weapons, spells of magic, doors, picking up objects from the ground, using objects, etc., in addition to sounds related to the interface, such as opening menus, and when the avatar gains new levels.

On the basis of this sound environment, I will study the soundscape of *Sacred.* The analysis will be concerned about sound as representational system, related to the actual functionality of sound in this game. I will investigate sounds with a source onscreen versus sounds with a source offscreen. I will study sounds that seem natural in the environment of *Sacred*, opposed to sounds that do not seem very natural in this respect. Here I will take advantage of Clive Fencott's concepts of *surprises* and *sureties*, which are two dimensions of what he calls *perceptual opportunities* [9, 10]. Perceptual opportunities are psychological qualities of a virtual environment that work to seek and keep users attention through the human perceptual system [11]. In this way, their function is to somehow influence user engagement and action within the virtual environment. Since computer game sound works along this axis, it will be fruitful to use the concept of perceptual opportunities as analytical tool when studying the soundscape of *Sacred*. Also, Fencott's perceptual opportunities relate to the idea of representation because they are highly communicative, being developed and placed into the virtual environment for a specific purpose related to the users' attention. As far as sureties and surprises are concerned, designers want these to be natural pieces of the virtual environment, thereby simulating features from the real world.

[2] No jingles are more than a minute; most are less than half a minute.

Fencott separates perceptual opportunities into three groups. *Sureties* are elements in the virtual environment that seem natural in this setting, and which are highly predictable. According to Fencott, sound is an important surety in real world, because it "gives important information about the nature and scale of the space we are currently experiencing" [12]. However, sounds may also have other perceptual opportunities. The second group is *surprises*, which do not have the same predictability as sureties, and which arise suddenly but still seem natural to the virtual environment. They seek to draw attention to themselves, thus stimulating users to take action. The last group is *shocks*, which are not part of the natural environment, but arise from errors in the design process. These draw attention to the virtual environment as construct, and are equivalent to software bugs.

3.1 Sound with Offscreen Sources

The film sound theoretician Michel Chion is also concerned about the relationship between sounds that emerge from onscreen and offscreen sources. He calls offscreen sources *acousmatic* sounds [13], borrowing the term from Pierre Schaeffer. An acousmatic sound is heard while its original cause cannot be seen, a definition that of course makes all computer game sounds acousmatic, since all sounds here are pre-produced and artificially added to the visuals. However, in this context we will instead understand it as sound where the source to which a sound is assigned cannot be seen. In *Sacred*, we can separate between three groups of acousmatic sounds, namely sounds from enemy NPCs, ambient sounds, and music. These are all sounds that cue the player to set up hypotheses about the sources of the sounds, thereby being potentially representative. However, music is separated from the other two by the fact that it is nondiegetic; it does not originate from the game universe but is rather an external, commenting background feature that underlines the fictiveness of the game world. Thus, the player does not look for a source within the game. Ambient sounds and enemy sounds, on the other hand, are features of the game world, and the player therefore expects to find their sources somewhere within the game world. But this is not simple in relation to ambient sounds.

3.1.1 Ambient sounds

To find the sources of the ambient sounds is not easy. Ambient sounds in *Sacred* are not connected to specific objects in the same manner as other sounds in the game. Instead they belong to the general atmosphere of the environment, and consist of a mixture of sounds from different potential sources. In the forest the player hears animals, birds, and the wind blowing in the trees, but if the player goes searching for the sources, s/he will never find them since they do not exist as objects programmed into the game. Still it is possible for the player to identify or have hypotheses about their theoretical sources. In this sense, the ambient sounds may be experienced as a representation of their sources, since the player needs to have a mental model of what may create such sounds. In addition, ambient sounds work as a substitute for the sources not present. However, the source is not really important together with the functionality of ambient sounds. The function of ambient sounds is to create an atmosphere and the feeling of a naturalistic environment, hence these sounds clearly simulate a specific feature of real world sounds. As in the real world, ambient sounds are background sounds that do not draw on our attention but instead makes the environment feel alive, and it is this function that is transferred to computer games. In addition, it would be dubious to say that the ambient sounds are a *representation* of an *atmosphere*, since moods are mental states and ideas more than something to be represented. To speak of ambience as representation would therefore have to be in a metaphorical sense. Moreover, neither can we say that ambient sound in *Sacred simulates mood*, but we can say that it is a simulation of sound found in real life locations, which again may represent a certain mood. Here we see again that functionality goes beyond any experience of sound as representation, even though what the sound simulates is not relevant for physical action. We may still say that the sound is *informationable*, since it gives information about what kind of environment the avatar is in. In this case the sound only gives the same information that one receives through the graphics. Related to perceptual opportunities, ambient sound is the prime example of a *surety*, since it is an auditory feature that seems natural to the environment, at the same time as it is highly predictable. There are no dramatic variations in the ambient soundscape, and it works to create immersion and an atmosphere. Thus ambient sound works in the same way as real world ambience; where we normally do not question or try to find the physical source, but expect to have it as omnipresent background sound.

3.1.2 Music

Music stands in special position as acousmatic sound, not only because it is the only nondiegetic acousmatic sound, but also because it experientially tend to blend with the ambient sounds. This is a general tendency in modern computer games, and this trend is transferred also to *Sacred*. Music also has some aesthetic qualities that are hard to describe, not at least when it comes to its relationship to a virtual environment. While the game environment tries to simulate a real environment as far as certain actions are concerned, and not at least as far as the diegetic soundscape is concerned, music does not fit into the simulation in this respect. In my view, music is the kind of sound that we most easily may dub representational, since it has aesthetic qualities, and works to comment game play in a very communicative manner. It also works to create an atmosphere and mood; not by simulating real world sounds, but by presenting music that by the use of certain instruments, tonalities and gamut represents a special mood. In many modern computer games, there is music that signals the entrance of situations or special NPCs; the computer game equivalent of what is known in opera and film as *leitmotifs* [14]. However, *Sacred* does not have leitmotifs; instead it has adaptive music [15] that changes according to the situation. This means that situational music does not start until after the event has started. So when the avatar is in combat, the music may start as forest music, and then continue into ambient sounds before going into combat music. However, as I mentioned above, there are jingles signalling for instance danger. But they cannot be called leitmotifs because they only inform the player that this is a place where there are monsters, not that they are attacking. It only affords information. When studying music as a perceptual opportunity, we find that it is a complex task to describe it according to this concept. Being nondiegetic, music is strictly speaking not part of the environment, although it works most certainly as a psychological quality which is present to keep the player's attention focussed upon the virtual world. Meanwhile, music is directly connected to the environment and what happens within it. Also, music is not a natural part of the game environment, although it most of the time is predictable since music follows certain rules of rhythm and tonality. As already mentioned, it is also a background feature that does not directly call on our attention. In this sense, it is no *surprise*, but it is definitely no *surety* either. It is most fruitfully labelled *shock*, since shocks are not natural to the environment, but influences it and draws attention towards the environment as construct. The difference between Fencott's shocks and music, however, is that music is purposefully added as an aesthetic feature influencing the environment, while Fencott's shocks are errors in the design.

3.1.3 Offscreen NPCs

Acousmatic sounds from hostile NPCs in *Sacred* are of another kind, where sounds are directly linked to their in-game sources. While both ambient sounds and music always remain acousmatic in the game, sounds from NPCs has the ability to get *de-acousmatized* [16], or to become visible. When hearing the voice or footsteps of an offscreen enemy, the player knows that the enemy at the moment is out of range of sight, but that it may be visible any moment. While it still remains offscreen, the player will make assumptions about the sound, for instance that it belongs to an enemy since only hostile NPCs make sounds when they are offscreen, and what kind of enemy it is based on what kind of utterances it makes. These assumptions may then cue the player to take action of some kind. The player may change combat tactics by readying a certain spell of magic or a weapon, or if s/he suspects that the enemy has not yet spotted the avatar s/he may stand still or move away from the sounds. It is important that the direction of the sound gives the player information about the enemies' direction, thus moving towards the enemies may be an option, for instance if s/he wants to kill enemies for experience points or if s/he suspects that these enemies are related to a quest s/he is assigned to. Sounds connected to offscreen NPCs are therefore informationable, since they have the function of informing the player about presence, direction, and type of enemy. In *Sacred*, this works in a similar fashion to how we experience real world sounds from sources we cannot see, and in this way it seems to simulate how real world sounds behave in relation to sources in the real world. However, since the player needs to make hypotheses about the source as long as the source is not seen, and these hypotheses will be part of the player's understanding and interpretation of what might be the source of the sound, we may say that there is a representational relationship between the voice and footsteps and the actual NPC while the NPC is offscreen. But once the sound becomes de-acousmatized, the sound and the source merges into one and the player does not experience them as separate. There seems to be no difference between the experience of real world sounds and game sounds in this respect, because the functionality of the sound is more important than whether it may be seen as representing an object. If sound was regarded a representation in this case, we would have problems deciding on what it actually represents: the source, the direction, or the situation? When we speak of acousmatic NPC sounds, these are best seen as *surprises* in the terminology of perceptual opportunities. Since these sounds are potentially de-acousmatized, they must be defined under the subcategory *attractors*, which are features that draw attention towards spaces in the virtual environment distant from the player. They are often spatialized, hence heard from afar, and may also be partially obscured. Player action is affected by attractors, since they inform the player about presence and direction. In this sense, acousmatic NPC sounds may also be seen as *connectors* that decide the route of the game, affecting where the player moves the avatar [17].

3.2 Sound with Onscreen Sources

3.2.1 Close Encounter: Onscreen NPCs & Events

While voice and footsteps of NPCs may be heard onscreen as well as offsceen, there are also other sounds available when NPCs are in the visual range of the avatar. Even though *Sacred* in theory could have offscreen sounds from events and objects, NPCs will not initiate events or manipulate objects until they are onscreen. Thus, sounds from events such as spellcasting or explosions, do not appear until the enemies are within range of sight of the avatar. Obviously, also sounds from attacks and weapons are not heard until the avatar and NPC are in close range (arrows are of some reason not heard at all). It is then possible to hear very many different sounds, both connected to objects, NPCs and the avatar. When a hit is dodged by the enemy's weapon, there is a sound of metal on metal; when a stroke hits, there is the sound of something hard on something soft; and when there is no hit at all, one hears a "woosh" from steel moving fast through the air. The avatar and NPCs also make grunts when they are hit, and they scream out loudly when they die. We should note that I say that these sounds are onscreen. This is not entirely correct. Although it seems to originate from onscreen agents, it is hard for us to see if a sword hits another sword, the enemy, or does not hit at all. During combat, the player is not always able to for instance monitor the avatar's hitpoint status, and in such a case it is crucial also to hear when the avatar is hit for the purpose of always having a relative idea of whether it is time for healing or running away. Sounds connected to all these events afford informative functions that give additional information to what the player gains through the visuals alone. But the sound does of course also work to create a realistic auditory element in the game. While it would be possible to say that such sounds refer to the physical hit, also here it is more relevant to speak of simulation, since what is important is the fact that the sound influences the player to take some form of action. Since the only senses through which the player may perceive in a game are ears and eyes, it is important to simulate pain and other perceptions through the available senses.

However, in *Sacred* there are also objects and events connected to sounds that do not correspond to sounds from similar real world objects. This fact makes it difficult to see sounds in this game as a representation of sounds in our world. One obvious example is the sound from spells of magic, which we have no real world reference to. However, we do have an idea about magic as a concept from fairy tales, legends and fantasy fiction, and when this concept is simulated in a game it also needs to be accompanied by a sound in order to be convincing. This is because similar features in the real world such as explosions and other violent events are accompanied by sound.

The sounds discussed here connected to onscreen objects may be categorized as *surprises*, since they are sudden sounds that appear naturally from the environment. Although not distant or spatialized, as surprises we may see them as attractors, since they call attention to themselves by informing for instance about relative hitpoint status. However, they are still predictable according to their situation, since we do expect sounds to arise from physical encounters such as a sword hitting a person, and it may be argued that these sounds are not surprises at all, but sureties.

3.2.2 Sound Related to Other Objects

There is also another way in which *Sacred* questions the representative status of sound in games. When the avatar picks up objects from the ground, or equips objects from the inventory menu, the player hears a sound. When the object is a weapon, there is a short metallic sound that reminds of a knife pulled against a grindstone. When the object is armoury, the sound is like dropping a chain mail to the ground. When the avatar picks up or equips magical objects, the sound reminds more of a short strike on a harp. Although the sound here is not equivalent to how we experience sound in real life in these situations, the sounds are added for the purpose of a special function: they are informationable since they notify the player that objects in the inventory are used, so that s/he does not run the risk of using objects involuntarily without noticing. In this sense, they are auditory simulations of physical manipulation. Concerning perceptual opportunities, these sounds are in an awkward position. In a sense they do belong to the environment, since picking up objects and using them are environment oriented; but also, they are sounds of the interface, and thus place themselves somewhere on the threshold between the game space and the player's space. We could label them shocks, on similar reasons as we suggested shock as category for music, but at the same time they are related to the actual diegetic picking up and manipulation of items. We must also admit that implementing these sounds is a way of simulating use and manipulation of items, so it would probably be best to regard them as belonging to the game world. In one sense, we may label them sureties, because often do expect a certain sound when manipulating objects; on the other hand they appear suddenly, and often have the function of attracting attention towards the fact that the avatar uses an item. However, concerning sounds from the interface, there is one interface sound we must label shock, since they are not part of the game world, and this is a subtle click that always appears when the player opens a menu.

The reason why we are convinced by these kinds of simulations, which are not equivalent to the sounds we hear in similar real world situations, is that it is formed a spontaneous relationship between what one sees with what one hears at a specific point in time. This is what Chion calls *synchresis* [18]. Although the sound and its source are not originally connected, they are perceived as such because of their synchronous appearance, and provide thus an explanation of why we accept such connections between sound and a source as described above. Not at least, the idea of synchresis supports the idea that sound and source are experienced as one entity when the source is visible, and not as two separate parts where the one represents the other.

3.2.3 Avatar Sounds

Concerning sound related to the avatar in *Sacred*, these can obviously never be offscreen, since the avatar always is the central focus in the image. In *Sacred*, avatar sounds are limited to a few: non-volatile sounds, and those related to voice or vocal chords. Non-volatile sounds are footsteps, heavy breath, and grunts. Footsteps are present at all times when the avatar moves, and they change according to whether it walks on grass or tiles. Heavy breath appears frequently when the avatar is running around, and grunts indicate that the avatar is hit by an enemy. The function of these sounds is to simulate auditory features within human beings, and thus make the environment more realistic. Some of them, for instance footsteps, are partly informationable, but with no clear relevance for game play. However, heavy breath, for instance, does not seem to be present for other reasons, although one could believe that it suggested that the avatar was tired and needed to rest. Since the functionality of this kind of sound is very limited, it is possible to argue that this is only a simulation of naturalistic human sound reduced to representation. But I do not support this argument since it is only one part of the whole simulation of a human being that is the avatar. The same goes for the avatar's use of voice, which in most cases is atmospheric and does not have any function related to physical game play. Avatar speech in *Sacred* is in general non-informational, instead related to combat situations, where the utterances take the form of threats towards the enemies, which do not react to verbal threats. In this case, the sound has the function of surety, since the sound seems to be a natural piece of the environment, and is predictable, in addition to the fact that it has no direct relevance for game play. But there is an interesting exception to this. If the player moves the avatar in a direction with no relevance for current quests, the avatar will tell the player so by comments such as: "I have no reason to chase enemies in this part of the country". Here we see that the voice has direct relevance to game play, by discouraging the player to take a certain direction. This may be seen as an subcategory of surprise, namely *connector*, since it decides the route of the game.

As mentioned, sounds from onscreen objects and events are more directly connected to their sources than sounds from offscreen sources are, most of all because the player does not have to speculate about the sound's origin. Instead, sounds from onscreen sources are parts of the simulation of objects and events. When we perceive a real world object or event, we perceive it as a whole, where sound is a natural part of what happens. It is not possible in the general course of experience to perceive sound as separated from its source. There is no way we can say that there is an arbitrary relationship between sound and its source, but this is not the same as to say that the sound is a representation of something else. When we say that sound is a representation of something else, this is to reduce the importance of sound to secondary compared to visual perception, and this would be problematic not at least when we regard sound as having important functions related to action and reaction both in computer games and the real world.

4. CONCLUSIONS

Talking about sound as representational system is a delicate issue, and as this paper demonstrates, the discussion comes to no conclusion. Seen in isolation sound may be experienced as a representational system, but when sound is contextualized and given the function to provoke action and behaviour in a player, and thus be a catalyst to mastery of a system, we see that sound is winded up in a larger system of simulation. However, representation and simulation are not mutually exclusive: a simulation may also include representation of some sort, but a representation can only describe simulation, and not demonstrate how the simulation system works.

It is hard to separate a sound from what produces it; they exist in symbiosis, and any separation between the two is an analytical and artificial construct. We hear a sound and know that it always has a source of some kind, but we do not think of it as a representation of its origin. Although there is a cause-effect relation between a sound and its source, it becomes dubious to even see sound as an indexical sign of its source. This would be the same as to claim that clean dishes next to the sink are a representation of dishwashing. To claim that cause-effect relations in general are semiotic systems is a reductive way of seeing such relations, since it degrades the effect to a "signifier" of something else; thus implying that its status is evaluated as less important, or less real. Also, since sound has temporal qualities and can only be studied or registered as a temporal feature, i.e. we cannot pause sound and study it at one exact moment of time as we can with images, sound should be understood as a *process* or an *event*. In the same way as the sound from a computer's fan, ambient sound in a game is a process, since it comes into being as the result of an ongoing process that in itself produces a more or less monotonous sound that is present as long as the process is going on. The sound from a sword that hits a monster in a game can on the other hand be compared to the knock on a door, which is a sudden and short sound resulting from two or more physical objects that meet. This is a different way of seeing cause-effect relations that goes beyond a mere semiotic explanation, and which is helpful when studying symbiotic and multi-perceptual phenomena such as sound comprehension. However, when one hears the sound from a source that cannot be seen, we make hypotheses about the origin of the sound. This way of perceiving sound may have representational qualities, since we must have a mental model in our minds about its source and its implications in this respect. This is especially evident when we have problems identifying the sound according to its source.

This paper has also used *perceptual opportunities* as an analytical tool in relation to sounds in *Sacred*. This has been done because the concept seems to fruitfully combine important qualities of the idea of representation with the functionality aspect of simulation. The concept of perceptual opportunities looks at objects in virtual environments as a medium with communicative features produced for a purpose by designers, yet also as a navigable space where a user may manipulate objects and take actions that matters and influences the system in a way that previously only has been possible in the real world, and not in media constructs. Perceptual opportunities therefore let the virtual environment be both representative and simulative at the same time. However, as we have seen in the analysis of *Sacred* the ideas of *sureties*, *surprises* and *shocks* do not seem to cover all ways in which sound in games function. It may therefore be a fruitful next task to further develop perceptual opportunities into also including the functionality of game sound.

5. REFERENCES

[1] Ascaron Entertainment (2004): *Sacred.* Ascaron Entertainment [software].

[2] Grodal, Torben Kragh (2003): "Stories for Eye, Ear, and Muscles", in Wolf & Perron (eds.): *The Video Game Theory Reader*. London, New York: Routledge, pp. 148-9.

[3] NC Soft (2003): *Lineage II: The Chaotic Chronicle*. NC Soft. Available:

http://www.lineage2.com/pds/game_dw.asp [28.04.2004-20.07.2004] [software].

[4] Frasca, Gonzalo (2001): “Simulation 101: Simulation vs. Representation”, in *Ludology.org* [online]. Available: http://www.ludology.org/articles/sim1/simulation101.html [20.07.2004].

[5] Frasca, Gonzalo (2001): “Simulation 101: Simulation vs. Representation”, in *Ludology.org* [online]. Available: http://www.ludology.org/articles/sim1/simulation101.html [20.07.2004].

[6] Gibson, James (1979): *An Ecological Approach to Visual Perception*.

[7] Blizzard Entertainment (1996): *Diablo*. Blizzard Entertainment; Blizzard North (2000): *Diablo II*. Blizzard Entertainment.

[8] Bioware (1998): *Baldur’s Gate*. Interplay; Bioware (2000): *Baldur’s Gate II*. Interplay.

[9] Clive Fencott (1999): “Content and Creativity in Virtual Environment Design”, paper from *Virtual Systems and Multimedia '99*, University of Abertay Dundee. Available: http://www-scm.tees.ac.uk/users/p.c.fencott/vsmm99/ [02.07.2004].

[10] Clive Fencott, Paul van Schaik, Jonathan Ling and Mohammed Shafiullah (2003): “The effects of movement of attractors and pictorial content of rewards on users' behaviour in virtual environments: an empirical study in the framework of perceptual opportunities”, in *Interacting with Computers*, Volume 15, Issue 1. Available: http://www.sciencedirect.com/ [02.07.2004]

[11] Clive Fencott (1999): “Content and Creativity in Virtual Environment Design”, paper from *Virtual Systems and Multimedia '99*, University of Abertay Dundee. Available: http://www-scm.tees.ac.uk/users/p.c.fencott/vsmm99/ [02.07.2004], p. 5.

[12] Clive Fencott (1999): “Content and Creativity in Virtual Environment Design”, paper from *Virtual Systems and Multimedia '99*, University of Abertay Dundee. Available: http://www-scm.tees.ac.uk/users/p.c.fencott/vsmm99/ [02.07.2004], p. 6.

[13] Michel Chion (1994): *Audio-Vision. Sound on Screen*. New York: Columbia University Press, pp. 71-3.

[14] Michel Chion (1994): *Audio-Vision. Sound on Screen*. New York: Columbia University Press, p. 51.

[15] Whitman, Guy (2003): “Design with Music in Mind: A Guide to Adaptive Audio for Game Designers”, in *Gamasutra* May 29, 2003 [online]. Available: http://www.gamasutra.com/resource_guide/20030528/whitmore_pfv.htm [13.07.2004].

[16] Michel Chion (1994): *Audio-Vision. Sound on Screen*. New York: Columbia University Press, p. 72.

[17] Clive Fencott (1999): “Content and Creativity in Virtual Environment Design”, paper from *Virtual Systems and Multimedia '99*, University of Abertay Dundee. Available: http://www-scm.tees.ac.uk/users/p.c.fencott/vsmm99/ [02.07.2004], pp. 9-10.

[18] Michel Chion (1994): *Audio-Vision. Sound on Screen*. New York: Columbia University Press, pp. 63-4.

Using Composition to Re-Present Personal Collections of Hypersigns

Andruid Kerne, J. Michael Mistrot, Madhur Khandelwal, Vikram Sundaram, Eunyee Koh
Interface Ecology Lab
Center for Study of Digital Libraries, Computer Science Dept.
College Station, TX 77845, USA
+1 979.862.3217

andruid@cs.tamu.edu, mistrot@blueskystudios.com, {madhur, vikrams@tamu.edu}, e0k6690@cs.tamu.edu

A *language* (*langue*) is a social institution and a system of values… that resists the modifications coming from a single individual. In contrast, *speech* (*parole*) is essentially an individual act of selection and actualization, … the combination thanks to which the speaking subject can use the code of the language with a view to expressing his personal thought.

The sign is a compound of a signifier and a signified. The plane of signifiers constitutes the plane of expression and that of the signified the plane of content… The sign-function … is the very unit where the relations of the technical and the significant are woven together."

– Roland Barthes, *Elements of Semiology*

ABSTRACT

This paper develops a semiotic operand, the *hypersign*. The hyperlink is structurally amenable to semiotic analysis. Anchor and destination function as signifier and signified. Especially significant are digital surrogates, such as Amazon catalogue entries, and Google gists, which are systematically produced hypertextual signifier-signified pairings. Hyperlinked surrogates constitute the hypersign form.

We take cues from the art theorist John Berger, the installation artist Hanne Darboven, and the hip-hop DJ Spooky, in theorizing about transformational forms for personal hypersign collections. We provide an architectural overview of combinFormation, a mixed-initiative tool for using composition to re-present personal collections of hypersigns. combinFormation interjects personal speech, or syntagm, into the process of collecting information resources. We illustrate the use of combinFormation with a narrative description of the process of creating a composition of hypersigns, using the source materials that this paper draws from. The result is a form of recombinant information, in which visual bookmarks are transformed through assemblage, to create new meanings. Like the work of Darboven and Spooky, the resulting form interjects a personal sense into the experience of reproductions. In doing so, it recasts Benjamin's aura of the original as a postmodern process of personal collection.

Categories and Subject Descriptors

H.5.4 [**Information Interfaces & Presentation**]: *Hypertext / Hypermedia*: Architectures, Navigation, Theory, User Issues.

First published at COSIGN-2004,
14 - 16 September 2004, University of Split (Croatia),

General Terms

Design, Human Factors, Languages, Theory.

Keywords

semiotics, surrogates, information space, composition, remix, browsing, world wide web, digital libraries, electronic commerce

1. SIGNIFIER-SIGNIFIED-SIGN

Saussure's semiotics are based on the coupling of the signifier and the signified in the sign [28]. The signifier is a label, name, or icon, used to refer. The signified is the content or concept being referred to. The conjunction is tight. It is systemized in the practices of a language, or semiotic code, by a society. We can apply this construct in the analysis of meaning in information age ecologies [16]. Such analysis straddles media, cultures, methodologies, and epistemologies. For an initial example, in declarative programming languages, the signifier is the name of a variable; the signified is its value.

Barthes [1, 2] and Baudrillard [3] extend this formulation to account for n^{th} order nested sign amalgams. Barthes describes such phenomena in terms of *metalanguage*. One example is that a calm and dignified looking person of color saluting the French flag conveys a sense or order and dignity in French globalism [2]. Another example is the use of the term "world wide" web as a means of suggesting that everyone has equal access to digital information. However, 41% of those online are in North America, which has only 6% of the world's population [9]. This means that in actuality people in North America are an order of magnitude more likely to have access to the net than those elsewhere.

2. SURROGATES

Surrogate is a term from library science. A surrogate is "a replacement for an original item, … which gives some description of the item, and how it can be obtained" [7]. Examples of surrogates include entries in catalogues, bibliographic citations, search engine result set elements, and bookmarks. As Greene et al have observed, "in browsing, surrogates provide an important alternative to primary objects as they take far less time to examine," while "in digital libraries and archives, surrogates are crucial for browsing large distributed collections that result from filtering programs or analytical queries of the data space. The need for inventing new types of surrogates underlies much of the research in digital libraries" [13]. Semiotics has a role to play in this research, because it addresses the structure of representations of meaning.

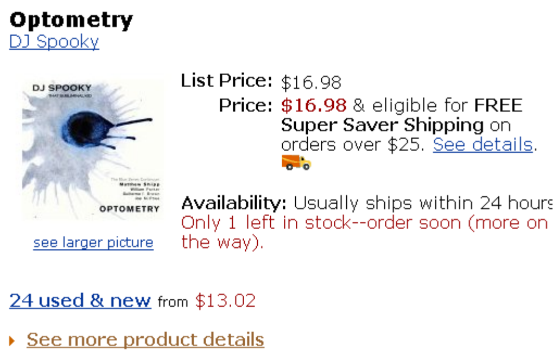


Figure 1: an Amazon surrogate.

Surrogates are often retrieved as members of result sets, in response to search queries, or category-based browse requests. Typical applications, from Google to eBay, present these collections of surrogates in the form of lists. Particularly since people involved in browsing and searching spend so much time interacting with result sets of surrogates, the form of their representation deserves consideration.

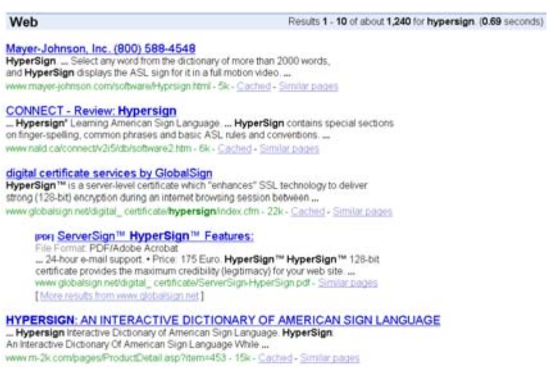


Figure 2: a set of Google surrogates.

3. SEMIOTICS OF THE HYPERLINK

We can use semiotics to analyze the function of surrogates in the context of hypermedia, the web, and digital libraries. Most basically, a hyperlink constitutes a sign conjunction. The link source, known as its anchor text or iconographic image, is a signifier. The href destination is its signified. Aesthetics, style, and semiotics function structurally in the construction of meaning across the hyperlink. For example, according to Tosca, the structure of the hyperlink association can be more straightforward (requiring less thinking), or more lyrical (requiring contemplation to understand). Lyricality, in this sense, can be seen as part of a cognitive process of ideation [11], in which new understandings are arrived at through engagement with the signifier-signified pairing of the hyperlink. The information discovery framework addresses the iterative reformulation of interests and goals, and the emergence of new ideas, during processes in which a person engages with digital information resources [18].

The hyperlink as sign strikes us a relation that is emblematic of the information age. Hypermedia surrogates constitute a special subset of these postmodern phenomena. In their role as descriptive replacements, surrogates are created particularly to function as signifiers; the original object is the signified. The conjunction of surrogate and hyperlink reference object interoperate as a fundamental unit of meaning, that is, as a sign.

Surrogates are created and re-presented according to structural schemas. In this, sets of surrogates from a single source instantiate consistent signifier forms. In addition to Google result sets, commercial electronic catalogs, online catalogues of brick and mortar libraries, and digital libraries are examples. Metadata ontologies play a fundamental role in this bridging between signified and signifier, as do presentation templates. These schemas define the form of a collection, through the plane of signifiers. The uniformity effected by surrogate structures in the presentation of a collection corresponds to the homogenous perspective in terms of which Barthes' identifies typical structuralist activity as the building of a *simulacrum* [1: 95-97], or simulation model, of the objects of investigation. Such phenomena encompass and go beyond branding, and particular experiences of (1) form query, (2) browse choices, and (3) select product. They delimit a metalinguistic form. Thus, the semiotic structure of an Amazon catalogue entry goes beyond selling a particular product, such as a DJ Spooky album (figure 1); it conveys the e-commerce marketplace as a whole. Likewise, the representation of Google result set gists goes beyond the results of a particular query on "hypersign" (figure 2), to stand for the process of effectively searching the web. This role of the hyperlink as sign in the representation of surrogates is so significant in the digital landscape, extending over, across, and through the terrain of browsers, search engines, electronic commerce, and digital libraries, that we call the hypermedia surrogate, *hypersign*. According to the search query of figure 2, this usage appears to be new. It shows that the term hypersign has not previously been used in semiotic or hypertext theory. In this example, the act of *googling*, and the sign of a Google surrogate result set play a standardized role in the validation of innovation.

4. THE SEMIOTIC FORM OF HYPERSIGN SETS

When we make a search query -- to Google, Amazon, or the ACM Digital Library -- we typically get back a set of similar hypersigns, such as a list of representations of articles or DVDs. Such sets correspond to what Barthes calls *system*, which corresponds to *langue*. Such a semiotic system is a homogeneous set of signs derived in association to a common term. The association, rather than an a priori common context, is the set's organizing principle. The search query is exactly this type of associational basis. In this way, the result set represents a language system.

Barthes offers another axis of semiotics, perpendicular to systematic language. *Syntagm* is a combination of co-contextualized signs which correspond to speech (*parole*). This is an individualized set, such as the clothes a person is wearing at one time, or all the furniture in a single house. At first blush, syntagm would seem to correspond to the personal collections of hypersigns that we are interested in: e.g., the shopping cart, the set of references that are associated with the development of an academic paper, or the set of bookmarks to cool stuff on the net that one teenager sends to another.

Yet, we must consider the process through which these personal collections are formed. That is, typically, early in the process of conducting research, a person will form queries to find relevant information resources. The particular results they decide to collect and use can be seen as a residue or crystallization, which emerges from these searches. The search queries formed and hypersign result sets received play an integral role in the process of forming personal collections, and so are, themselves, significant artifacts. Thus, in one type of personal representation of a research collection, surrogates of the most relevant results are foregrounded, over a background formed by the other more or less relevant results. The personal and the systematic are combined. We believe that from a semiotic perspective, the significant hypersign sets people need to collect are both syntagmatic and systematic.

Our research is focused on the semiotic form of the collections that people need and desire to create when they work and play with information. To what extent are the forms of individual and collected hypersign signifiers determined by the automata that produce them? To what extent can humans affect them? What role can expression play in these processes?

We turn to art history and critical theory. John Berger (citing Benjamin [4]) observes that in the age of reproduction, the role of the art museum has changed [5]. Instead of simply presenting paintings as visual artifacts, museums re-present the authority of the original, in relation to reproductions. Berger is unsatisfied with this role. He sketches an alternative paradigm for the museum in the postmodern era:

> Adults and children sometimes have boards … on which they pin pieces of newspaper cuttings, original drawings, postcards. On each board all the images belong to the same language and are all more or less equal within it, because they have been chosen in a highly personal way to match and express the experience of the room's inhabitant. [5: 31]

In Berger's model, the museum is a personal collection of reproductions, which is formed by the individual. The notion of "belonging to the same language," in this context, invokes langue and parole, at the same time. It is certainly personal speech, and thus *parole*/syntagm. Yet the push pin board may be used to collect, for example, a set of impressionist paintings, in which case systematic *langue* is invoked.

In contrast, we observe that the normal forms of hypersigns employed on the web are notable by how fixed they are. This is ironic, since digital media are so amenable to manipulation. Thus, while a user can type a word into a search box, one result set is typically received. There is little generally available in the way of mechanisms that enable one to move forward from the search in a personalized way, that is, in a form that corresponds to speech (parole), and syntagm, assembling hypersigns across queries. Exceptions to this tend to be limited in the expressiveness of their form. The Amazon shopping cart and wish list allow for the construction of personalized lists. Yet the surrogates are textual, and the format is fixed, rather than fluid, like the push-pin board. "My EBay," a more advanced incarnation, still consists of a list of lists (e.g., items I'm watching, items I'm bidding on, items I've bought). The user can customize which lists are displayed, and the order of these lists on the page. The list form does little to aid the user creatively, cognitively or semiotically in seeing how the items can function together as an ensemble or mental set, that is, in functioning syntagmatically. Instead, it challenges the capacities of human working memory [18].

The list and the push-pin board are differentiated by their degree of flexibility. Prior spatial hypertext systems [e.g., 24] give users the ability to manipulate elements spatially. However, they do not facilitate the manipulation of surrogates, in particular. That is, they leave the process of forming personal collections, and of maintaining the links of the surrogate hypersign, to the user.

Installation artist Hanne Darboven creates personal collections systematically [6]. She uses reproduced artifacts as elements. For example, "Friedrich II, Harburg 1986" consists of 398 prints of the same photograph, each annotated with a different handwritten text. The photograph is historic. It depicts a street scene, with a street car, pushcarts, and bicycles. A consistent form of annotation is employed. The resulting elements are re-presented in a grid structure. The form of the re-presentation transforms the

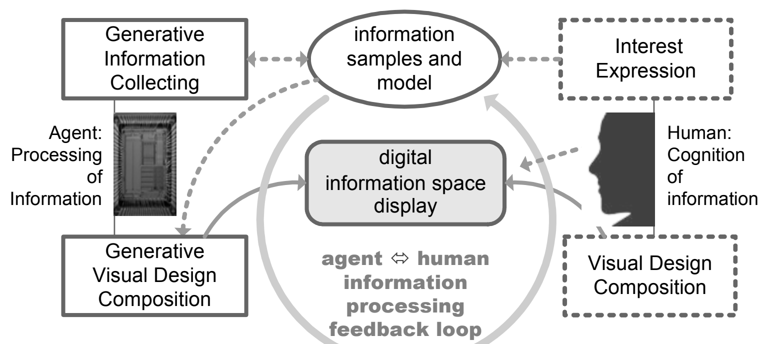


Figure 3: An overview of combinFormation's mixed-initiatives architecture. The agent engages in generative information collecting and visual composition. The user engages in interest expression and composes design. State flows through the model.

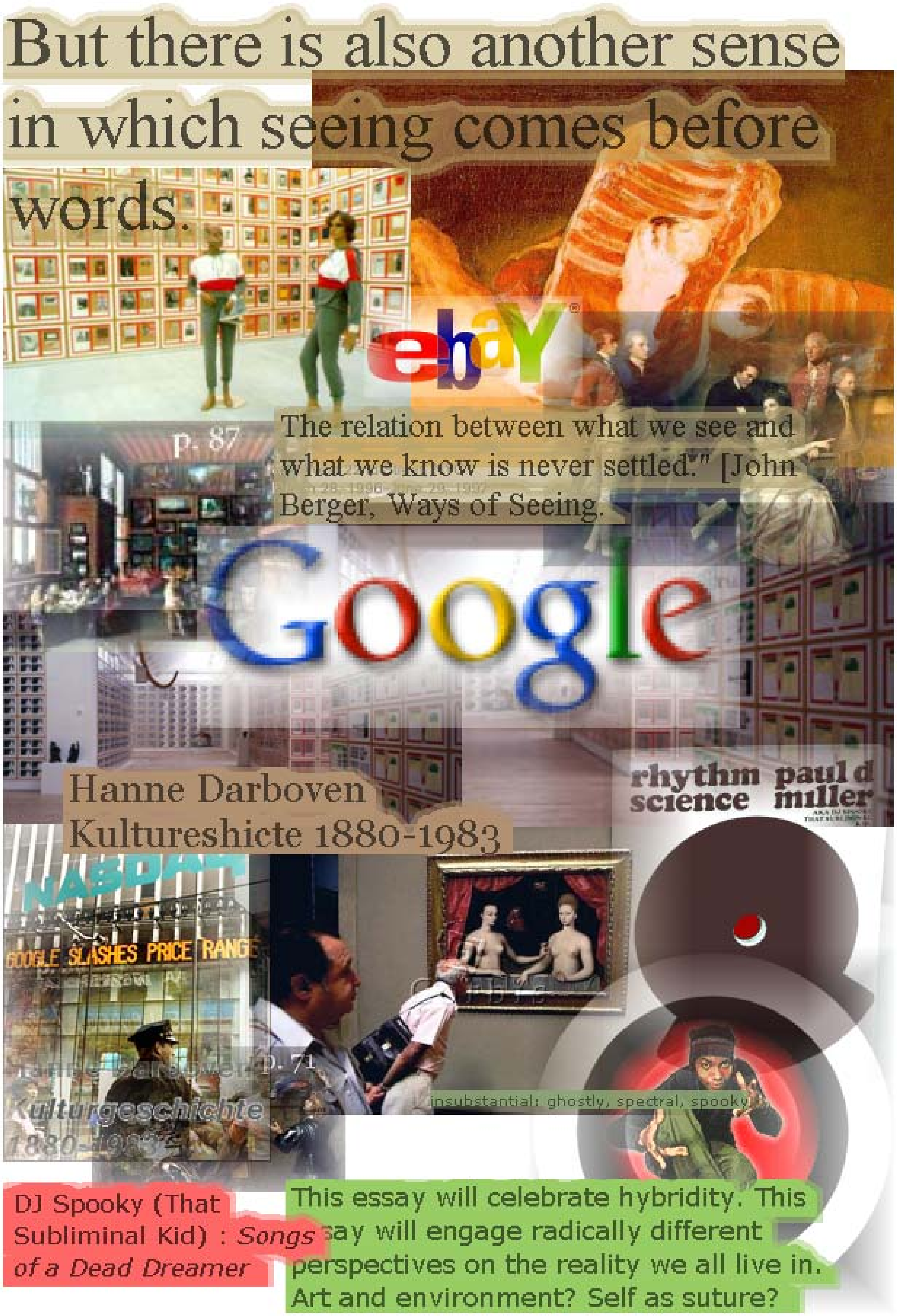


Figure 4: Re-presenting the ideas of this paper as a composition of hypersigns.
Samples include works by and about Hanne Darboven, DJ Spooky, John Berger, Google, and EBay.

reproductions into a new original. While the resulting language is personal for her, the audience receives a beautiful edifice of arbitrary structure, that is simultaneously distinctive and imposing.

Hip-hop musician Paul Miller (aka DJ Spooky) connects the formation of conceptual frameworks with rhythm. He identifies the assemblage of elements of identity, which characterizes the information age through immersion in data, as *multiplex consciousness*. Miller defines DJing in terms of *rhythm science*:

> Rhythm science uses an endless recontextualizing as a core compositional strategy… DJs are griots, and whether their stories are conscious or unconscious, narratives are implicit in the sampling idea.. The mix breaks free of old associations. The languages evolve and learn to speak in new forms, new thoughts… Sampling allows people to replay their own memories… At the end of the day, it's all about reprocessing the world around you [27: 21-27].

The hypersign is a significant type of sample. Surrogates are produced explicitly as samples, to represent the original object. They arrive in our field of browsing as anchors, dressed in hyperlinks. The Interface Ecology Lab researchers are interested in how hypersigns function in people's processes of creating meaning. People encounter hypersigns while browsing and searching. We use them as we work to construct new plans, new ideas, new knowledge. We use them as well, as we play with information, in more freeform processes. We form personal collections when we research, when we write, when we buy, and while we are entertained. We need to discover how to enable users to re-present these collections of syntagmatic and systematic hypersign forms. To this end, the Interface Ecology Lab is engaged in building software to support personal and public processes of hypersign remix.

5. combinFormation: COMPOSING HYPERSIGN COLLECTIONS

Our software gives people tools for re-presenting hypersign collections, whose information elements need to be connected in the mind, across experiences of interaction, cognition, understanding, and use. The focus moves from fixed forms, which are presented for binary one-shot decisions of accept or reject, to malleable forms which can be contemplated and manipulated by the user, forming personal re-presentations. We are developing a transformative digital engine that enables a person to engage in syntagm/parole with digital surrogates, by reappropriating their hypersign forms, and recasting them as personal speech. The basic idea is to gather hypersigns in a compositional space that enables a person to modify and recombine them. This form integrates prevailing language systems of surrogates, with new potentials for personal expression and recontextualization. These compositions can be shared via email and published on the web

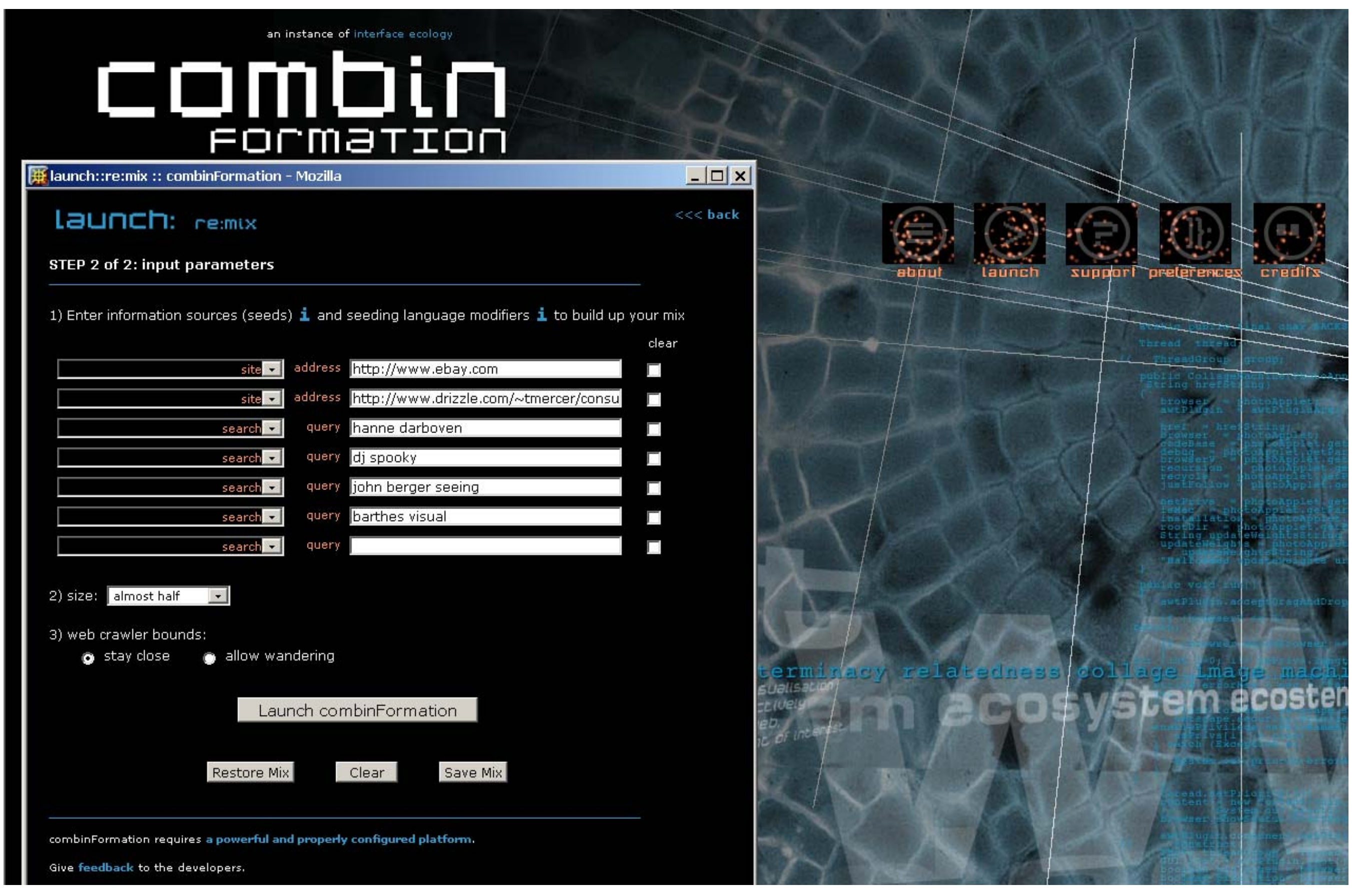


Figure 5: Seeding the information surrogates composition session in combinFormation; launching the re:mix.

[17].

In order to facilitate the person's process of gathering relevant hypersigns, the system integrates a composition agent into the process of creating the information space (figure 3). In this way, combinFormation is a mixed-initiative system that uses composition for browsing, collecting, and arranging information samples from web pages. In the current implementation, these samples are images and sentences of text. The samples act as visual, semiotic, and navigational surrogates for the documents from which they are extracted. When they are brought into the composition space, the images and texts retain the structure of first class samples; that is, they remain connected to their source and hyperlink documents, as hypersigns.

The initiatives mixed in combinFormation are the system's generation of composition, and the user's direct manipulation (figure 3). The system's generative actions -- collecting hypersigns, and composing them visually -- are conducted iteratively, based on a user model. The system presents the ongoing generation of the composition to the user in an interactive information space. In this space, one of the user's initiatives is to directly manipulate the composition through interactive design operations, which enable samples to be displaced, layered, resized, annotated, and removed. Another user initiative is to navigate through the sample as hypersign, connecting to a traditional browser. The user's third form of initiative is to express positive or negative interest in each sample. Expressions of interest affect the model, creating a feedback loop through the information space display (figure 3).

combinFormation uses an underlying grid structure for the layout of elements, as it adds them to the composition. The grid works quite differently than Darboven's [6], because the elements are heterogeneous, rather than homogeneous. In addition to containing different colors and shapes, the elements possess different aspect ratios. They are assigned different sizes, by the composition agent, based on their relative importance. As the space fills up, over time, new elements are expected to overlap prior ones. In the current implementation, the agent places each new element over the contiguous set of grid cells of the assigned size, which corresponds to the lowest weight candidate area. The user can override grid-based layout, by using the grab tool to reposition elements in the space.

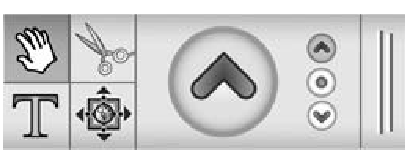


Figure 6: Composition + Interest Expression toolbar with selected operations: grab + positive interest.

The current combinFormation interface includes three floating toolbars. Initially, they are distributed vertically, on the right side of the space. Each toolbar can be dragged to anywhere on the user's screen. The top toolbar (figure 7) controls text-image mix. It defaults to 60% images. The center toolbar (figures 6, 7) enables selection of composition/navigation tool on the left side, and interest expression on the right. These options are orthogonal [19]; that is, they can be combined independently. Interest can be neutral, positive, or negative. The conjunction is displayed in the cursor. At the start of a session, composition/navigation is set to navigate, while interest expression is neutral. The other composition and interest expression actions can be selected by clicking the appropriate icon, or by toggling with arrow keys (composition/navigation via left/right; interest expression via up/down).

Elements are activated for interaction with the expression-interest tool conjunction, by simple mouse over, rather than click. This choice of interaction design style emphasizes fluidity. Activated elements display metadata in the context of the composition (No dialog boxes.). Mouse over activation also provides local tools for initiating search, and an edit palette for element-specific (rather than assemblage-scope) composition actions, such as font change or transparency toggle. Clicking an activated element causes the selected composition and interest expression effects to be processed.

The bottom toolbar (figure 7) is the tape recorder transport. In record mode, the usual initial state, the agent engages in generative visual composition. This can be paused. Prior states can also be traversed with reverse and forward play, and with a history jog-shuttle. These controls address the temporal form of the mixed initiative medium [21].

The composition space is split into 2 regions, delimited by a thin grey and white line (figure 7). The outer area is the *cool space*. This is for the human, only. In the center is the mixed-imitative *hot space*. The composition agent adds elements here, and as the space fills, will also remove elements from here. We have appropriated the language of McLuhan [25]: a hot medium, such as film, saturates its audience with information. A cool one, such as the novel, leaves more of the sense of view forming to the audience member's imagination.

6. HYPERSIGN RE:MIX SESSION

We will now explain the combinFormation experience through a first-person example session narrative. In this session, we will compose research source materials associated with this paper. combinFormation is a client-side web application, implemented as a veneer of DHTML, which gathers runtime parameters, and passes them as it invokes a signed Java applet. To use the software, you access the combinFormation web address [15] in a traditional browser, and then click "launch."[1] The launch page offers several forms for seeding [20] a session. This is the specification of initial documents, which are downloaded, and sampled. Information element surrogates encountered become candidates for the agent's subsequent generative visual composition initiatives. Hyperlinks become candidates for generative information collecting, via web crawling. (See the architecture overview in figure 3.)

In the example case, I have utilized the re:mix seeding interface (figure 5). This particular front-end enables specification of any number of seeds, and allows mixing of seed types, such as web site addresses, and search queries. I specify a couple of web sites (Mercer's paper, "The Consumption of Links" [26], and EBay), and then several searches in areas that have been cited. Then, I click launch.

[1] First time users will likely need to click "support" to install Java, and configure the Java plug-in, prior to using the program.

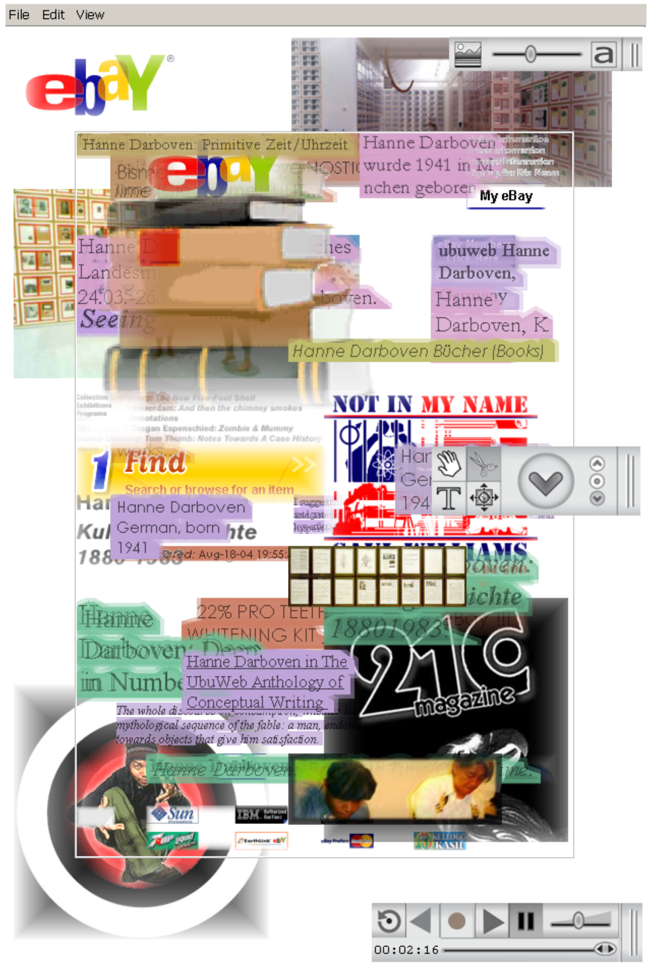


Figure 7: Move desired elements to the cool space, and pause.

The program processes search queries in a tailored manner. It uses Google to perform the searches. It does not actually bring samples from the Google results page – the standard surrogates -- into the composition. Instead, it immediately downloads the first fifteen Google result pages, from their web sites. These get sampled. The samples are collected in a set of candidates, which feeds the composition. Those of the collected samples that are selected for visual composition will function as hypersigns. Similarly, the links from the downloaded pages become candidates to feed the web crawler.

The applet initializes the interface. The space begins to fill. Some excellent images of Hanne Darboven's work enter. I use positive interest expression (up arrow in center toolbar, right side; this is also available with the up-arrow keyboard accelerator), to put some positive interest on one of these elements. I drag it over to the cool space (using the grab tool: center toolbar, left side), in the peripherae. This is a way to make sure that the agent won't remove it, or cover it up. I shift to neutral, so as not to express too much interest, too fast, and drag another Darboven image over. I repeat this pattern -- express some interest and drag to cool space -- the same with a DJ Spooky image. I also drag the EBay logo over, but with neutral interest expression. A bunch of other EBay elements are showing up. They are not interesting to me; I shift

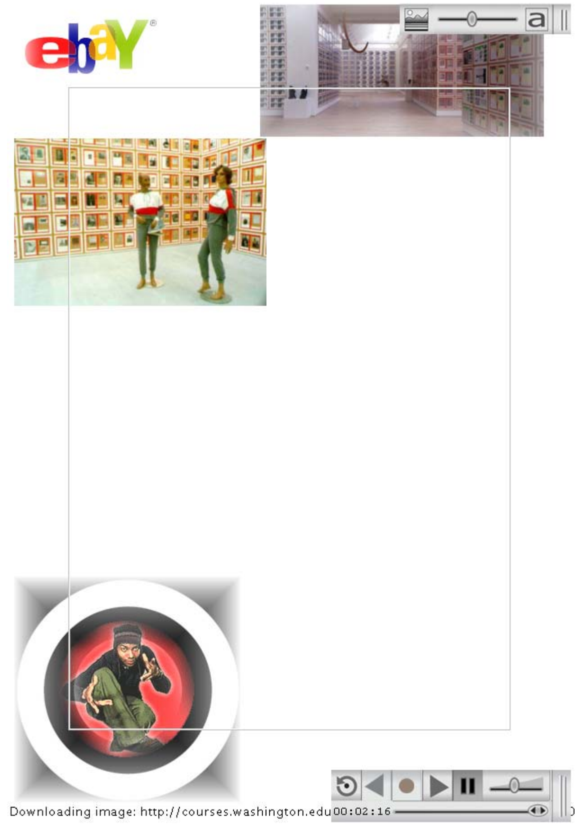


Figure 8: After clearing the hot space of extraneous elements.

the center toolbar left, use the down arrow to express negative interest, and remove several of them.

At this point (figure 7), I pause generation of the space. It seems to me that while I'm getting good stuff, and the session is going well, the center feels crowded with elements that I'm not crazy about, visually. I'm also really wishing for some John Berger material. I cut a bunch of elements, using negative interest. But this feels tiring. So, I press the clear button (lower toolbar, the tape recorder transport, very left). This is the button with the partial circle and arrow pointing counter clockwise. Immediately, all the elements in the center hot space are removed (figure 8). I use File -> Undo to toggle back and forth once, to make sure that I didn't delete any elements that I wanted. I feel satisfied.

I unpause composition generation. An element enters from a site about John Berger's *Ways of Seeing*. I click it a few times with positive interest, to strongly promote collecting related elements. These begin to flow into the space. The system also begins to respond to my expression of interest in DJ Spooky / Paul Miller. I drag the elements I like best out to the cool space. I reach the state shown in figure 9.

In figure 9, an image element with text that reads "Reviews and Features," is selected via mouse over interaction. I have set the composition-interest tool state to cut + positive interest. This is because the subject – reviews of albums by Spooky – is of interest. But the image itself is not visually strong, so I wish to remove it. Immediately after taking the screen shot, I click,

Figure 9: With more relevant elements on DJ Spooky and *Ways of Seeing*. An element is selected, displaying descriptive metadata. The cursor is poised for "positive cut."

simultaneously cutting the element and asking for more related information.

Next, I realize that I want Google depicted in the composition. But even though combinFormation has used Google to conduct the search, I haven't included the look and feel of their site directly in the seeding. On this occasion, Google's presence must be made explicit. I use the traditional browser to find a page in google.com, with the plain Google logo. (This is a bit more work than usual, because the Olympics are on, and most of the Google logos have been altered to celebrate them.) I use drag and drop to bring the Google logo into the center of the composition. I immediately resize this element to be larger, because Google plays such a major role in how people collect information. I simply use mouse over selection, and then drag one of the corner boxes.

I continue to compose with drag and drop. I use the navigate tool to go to a page from the Berger site. This page includes links to larger versions of most of the image examples from the book. I drag several of these into the space, as well. The result is figure 10.

I collect more hypersigns. I recall that there is an image on the New York Times web site this evening, in which a stock ticker

Figure 10: Hypersigns from Google and Ways of Seeing have been composed through human imitative of drag and drop.

shows a message about Google's IPO, and a decline in their initial asking price. The image strikes me as a signifier for the role of economy in shaping our experience of information, and the sometimes hard to comprehend manner in which this works. I drag this image in, as well. I want the new elements to feel more connected, visually, so I use the element editing panel to add compositing to several of them. I then turn generate back on. The space grows extremely dense, resulting in figure 11.

I enter a final stage of work: preparing the composition for re-presentation to an audience. I eliminate many elements, reducing density. I compose elements spatially, with more definite intention. I restore the EBay logo, which I somehow lost track of, with drag and drop. I make some text elements larger, and some elements smaller.

I take a break and return. Now, I make fine changes. I move the Darboven image near the top left up some, so one of the mannequins seems to be standing on the p. 57 label in the proximate Berger image. At the same time, I shift the EBay logo to give it more space, let it breathe. I position this logo clearly in relation to the other mannequin's hand, so it almost seems to emanate from it. The work of Darboven and EBay both involve mechanistic processes of forming and representing collections, which are also personal. Darboven's work can be interpreted as a

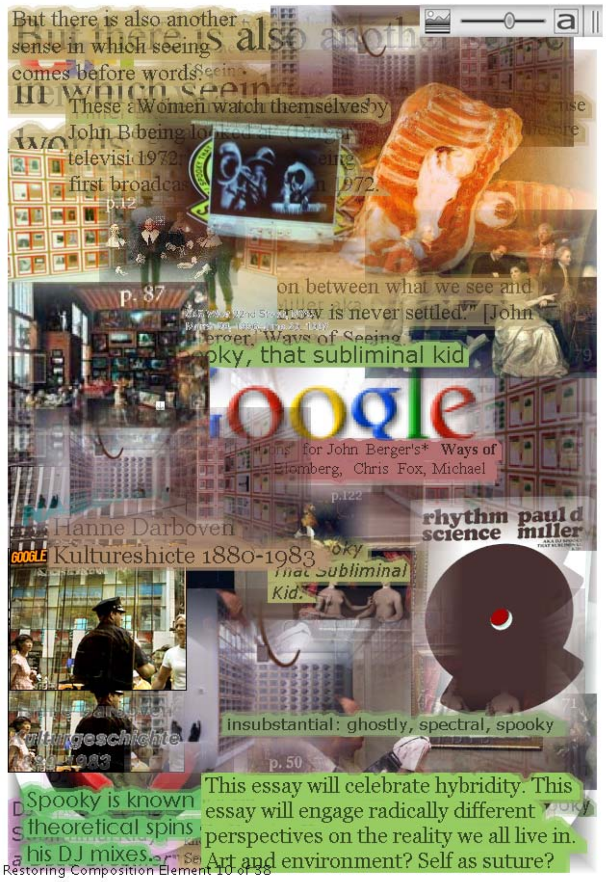


Figure 11 An extremely dense and rich space, after further generation of composition. The agent is retrieving many interesting elements. The composition clearly exhibits the influences both of myself, and of the agent. Two of the toolbars have been dragged out of the way, to permit clearer viewing of and interaction with the composition.

commentary on the role of automata in our lives. EBay, Google, and Amazon are signs of this, as prominent corporate entities.

I alter the opacity of a number of elements. Specifically, I make two of the Darboven images, and one of the images from *Ways of Seeing* (Goya's Still Life with a Sheep's Head) opaque, so they read more strongly. I do this by unchecking "compositing" on the edit palette that pops up on mouse over. (This palette can be seen with a different image, in figure 9.) I also adjust the stacking order of elements, by lifting some with Shift-Click (grab tool). In the composition as a whole, the opaque elements are still strongly composited. This is because for the elements that overlap them, compositing remains selected. These elements are re-presented with transparent borders. The program uses an *alpha gradient* to cross fade elements with "compositing" checked. The way this works is that the extreme border region of a composited element is processed to be more transparent. The center remains opaque. Transparency, also known as alpha, is interpolated across this region. In video editing, this technique is known as *masking*. In commercials and feature films, masking is executed to create perfectly composited scenes, in which the viewer will see elements as naturally part of the same landscape. In combinFormation, since the seams often show; the role of compositing is made visible. This changes its function from spectacle to detournement [8]. The new usage contrasts with Manovich's prior analysis of the prevailing popular media use of compositing in mainstream film and television [23].

The final composition is shown in figure 4. I save this version. The program produces several forms of the composition, when I hit File -> Save. First, it produces an XML file. This file includes textual elements by value, and image elements and source documents by reference. It is a metadocument, that is, a collection of references to other documents [12, 17]. The XML metadocument can be opened again by the program. People can open the saved XML form of the hypersign composition in one of two ways. To access it from a hard disk, use re:open as the launch option. Then, File -> Open. Or, the space can be uploaded to a web server. A special URL syntax can be constructed for hyperlinking to the saved information space [17]. The XML file is passed as an argument to a special seeding web page, which will in turn pass the saved space to combinFormation on launch.

In this mode, publishing the composition of hypersigns is fully enabling for the "reader." That is, the reader receives the composition in the same environment in which it is produced. The hypersigns do much more than enable navigation. Their meanings are connected through the form of the composition. When opened, the composition can be further manipulated. Reading and writing become symmetrical, identical acts, which proceed iteratively. The new reader inherits the interest model of the prior writer, as well as the design. The reader as author can use combinFormation to express new senses of interest, and change the design to suit her own ideas. S/he can also use the program's generative agent to continue the process of information riffing.

Additionally, the program outputs a JPEG image of the composition, and an HTML file, with an image map, that references the JPEG. The HTML form of the composition also functions as a visual and navigational metadocument. It has the advantage that it can be opened in any standard web browser, without requiring Java to be installed and configured. The HTML form is not open to further manipulation by the agent or the user. It instantiates a more typical publishing model of author and reader.

7. DISCUSSION

The saved form of the composition connects the work of Darboven, Miller/Spooky, Berger, Google, and EBay through visual semiotics. It re-presents the ideas of this paper, through demonstration. It can be considered as a supplementary text; or the paper, can be considered as notes on the composition as primary text. The composition is a mix, a form of recombinant information [20]. New meanings are created through the assemblage of found elements. The creative frameworks of generations of visual artists, such Duchamp [22] and, Ernst [30], filmmakers, such as Eisenstein [10], and hip-hop composers, such as Spooky and Public Enemy [28], are invoked.

combinFormation enables authoring by reference with web media. The resulting compositional form is that of the visual metadocument. These are not just images and texts. They are surrogates, which can be navigated, to retrieve original source materials. They are hypersigns, which refer to and re-present the ideas expressed in the source documents. They are recontextualized in the compositional space of detournement.

They are new forms of information, transformations of Google's detritus.

Benjamin decries the loss of the *aura* of the original, as creative works are reproduced in the age of reproduction [4]. Yet, as in the example of Darboven's work, a *postmodern aura* emerges. This is the aura of the personal collection, in which mass produced objects are recontextualized and annotated. Formalism is one means in this process. The appropriated objects become personal when we impose our own forms on them. Composition is a means for transforming personal collections of hypersigns, so they grow syntagmatic, even as a residue of systematic language remains. The mixed-initiatives of combinFormation bring to the public at large tools for engaging in this process of composing hypersigns. The agent's initiative, as well as the logics of underlying dynamic content digital providers such as Google, Amazon and EBay, transmit a re-presented form of prevailing systematic *langue* through this engagement. The user's initiatives interject syntagmatic forms of personal meaning. We believe that by supporting personal expression as such, we will promote people's creative processes of information discovery [18], as well as their sense of empowerment. As developers, the Interface Ecology Lab's underlying goal is to elevate the role of personal speech (*parole*) and creative process in everyday interactions with information.

The composition created to demonstrate the ideas of this essay instantiates the interface as border zone between heterogeneous systems of representation [16]. Semiotics, information science, design, human computer interaction, art history, musicology, and computer science are among the blended methodologies. As the theorist, I am present, developing a first person narrative of my own experience. The composition, itself, connects theory and practice self-referentially, in a strange loop [14], by utilizing materials on the subject at hand. It is part of the environment that it describes. Discourse about theory is instantiated through practice. Through this process and form, the composition+essay functions as an interface ecosystem.

8. REFERENCES

[1] Barthes, R., *Elements of Semiology*, New York :Farrar Straus and Giroux, 1964.

[2] Barthes, R., *Mythologies*, New York : Farrar Straus and Giroux, 1957.

[3] Baudrillard, J., *For a Critique of the Political Economy of the Sign*, New York: Telos Press, 1981.

[4] Benjamin, W., The Work of Art in the Age of Mechanical Reproduction, in *Illuminations*, New York: Schocken Books, 1935.

[5] Berger, J., Blomberg, S., Fox, C., Dibb, M., Hollis, R., *Ways of Seeing*, New York: Penguin Books, 1972.

[6] Burgbacher-Krupka, I., *Hanne Darboven: Constructed Literary Musical*, Stuttgart: Cantz Verlag, 1994.

[7] Burke, M., *Organization of Multimedia Resources*, Hampshire, UK: Gower, 1999.

[8] Debord, G., *Situationist International Anthology,* Bureau of Public Secrets: Berkeley, California, 1981.

[9] Digital Divide Network, Digital Divide Basics Fact Sheet, http://www.digitaldividenetwork.org/content/stories/index.cfm?key=168, 2004.

[10] Eisenstein, S., *Film Form*, Orlando: Harcourt Brace, 1949.

[11] Finke, R., Ward, T., and Smith, S. *Creative Cognition*. Cambridge MA., MIT Press, 1992.

[12] Furuta, R. Shipman, F., Marshall, C. Brenner, D., Hsieh, H.. Hypertext paths and the World-Wide Web: experiences with Walden's Paths. *Proc ACM Hypertext*, 167-176, 1997.

[13] Greene, S., Marchionini1, G., Plaisant, C., Shneiderman, B., Previews and Overviews in Digital Libraries: Designing Surrogates to Support Visual Information Seeking, *JASIS*, 51:4, 2000, 380-393.

[14] Hofstadter, D., *Gödel, Escher, Bach: An Eternal Golden Braid*, New York: Basic Books, 1979.

[15] Interface Ecology Lab, combinFormation, http://ecologylab.cs.tamu.edu/combinFormation, 2004.

[16] Kerne, A. Interface Ecosystem, The Fundamental Unit of Information Age Ecology, *Proc SIGGRAPH 2002 Art and Animation*, 142-145.

[17] Kerne, A., Khandelwal, M., Sundaram, V. Publishing Evolving Metadocuments on the Web, *Proc Hypertext 2003*, 104-105.

[18] Kerne, A., Smith, S.M., The Information Discovery Framework, *Proc DIS 2004*, 357-360.

[19] Kerne, A., Smith, S.M., Mistrot, J.M., Sundaram, V., Khandelwal, M., Wang, J., Mapping Interest and Design for Expressive Mixed-Initiative Information Composition, submitted to *ACM Creativity and Cognition 2004*.

[20] Kerne, A., Sundaram, V., A Recombinant Information Space. *Proc COSIGN 2003*, 48-57.

[21] Khandelwal, M., Kerne, A., Mistrot, J.M.,. Manipulating History in Generative Hypermedia, *Proc Hypertext 2004*, 139-140.

[22] Lippard, L. *Dadas on Art*. Englewd Cliffs NJ, 1971.

[23] Manovich, L. *The Language of New Media*, Cambridge: MIT Press, 2001.

[24] Marshall, C.C., Shipman, F.M., VIKI: Spatial hypertext supporting emergent structure, *Proc ECHT 1994*, 13-23.

[25] McLuhan, M., Lapham, L.H., *Understanding Media: The Extensions of Man*, Cambridge, MIT Press, 1994.

[26] Mercer, T. The Consumption of Links: Hypertext and John Berger's Ways of Seeing, http://www.drizzle.com/~tmercer/consumption/, 2002.

[27] Miller, P.D., *Rhythm Science*, Cambridge, MA: MIT Press, 2004.

[28] Public Enemy. *It takes a nation of millions to hold us back*. New York, Def Jam, 1995.

[29] Saussure, F. de., *Course In General Linguistics*, New York: McGraw-Hill, 1916.

[30] Spies, W. *Max Ernst Collages: The Invention of the Surrealist Universe*. NY, Harry Abrams. 1988.

Alien Letter Forms

Jason Lewis
Concordia University
1455 de Maisonneuve Blvd. West CB206-14
Montreal QC (CA) H3G 1M8
jason.lewis@concordia.ca

David Bouchard
Independent Artist
2715 rue Dézéry
Montreal QC (CA) H1W 2S6
david@deadpixel.ca

ABSTRACT
In this paper we describe Alien Letter Forms, a software environment for letters to live in and evolve into larger texts. Alien Letter Forms builds on technical and conceptual approaches borrowed from research around artificial life systems to create a digital ecology in which text agents live, reproduce, evolve, mutate and die in response to their virtual environment. Individual letters are treated as autonomous agents, where contact between individuals produce offspring composed of new letter combinations up to and including whole words. Evolution is driven further by how well new combinations of letters fit into a set of pre-existing reference texts. The Alien Letter Forms environment is designed to produce a continuously evolving collection of letters that has local stabilities that produce readable text.

Keywords
Digital text, dynamic typography, evolutionary text, emergent meaning, visual language, recombinant text.

1. INTRODUCTION
The current work is the latest installment in a series of experiments focused on exploring the expressive possibilities of computationally active digital texts. By computationally active, we mean texts which make use of computation to operate on the interactive, dynamic and linguistic dimensions of a text. The goal of these experiments is to discover new ways–both creative and technical–of working with one of our oldest communication technologies.

2. MOTIVATION
This project aims to explore a potential biological metaphor for language. Through evolution, simple life forms gave way to more complex ones; similarly, our characters slowly evolve into more complex strings towards formation of words and sentences. An important question we are hoping to answer with this experiment is whether the biological metaphor is relevant and to what extend it can be applied to language. We also hope to provide a means of exploring the semantic space of words as well as the visual space of typography.

We are not interested in, nor make any claims of, developing a scientific tool that simulates the natural evolution of human language. Rather we approach text from an evolutionary perspective as a source of inspiration and as a means of exploring the semantic space of a particular language as well as the visual space of typography.

3. RELATED WORK
The present work draws technical and creative inspiration from concepts of artificial life and emergent behaviour, agent-based behaviour systems and computational text.

Sim's Galapagos [15] is perhaps one of the most interesting attempts so far at applying genetic algorithms to interactive computer art. Projects like the Virtual Fish Tank [9] have been exploring evolution and emergent behaviour between virtual agents in a creative and visually compelling way.

Lewis and Weyers' work on ActiveText [4] specifies a system for coordinating complex dynamic and interactive behaviours between textual elements. The Visual Language Workshop at the MIT Media Lab has been the source of a number of works that use agent-based behaviour systems to support the visualization of rich behaviour-based texts. Ishizaki's dissertation [4] on "typographical performance" anticipates the use of coordinated behaviors to create a rich visual interplay of text in an email system. Wong's thesis [17] on "temporal typography" uses Soo's object-based, behavior-driven architecture [10] to combine dynamics and semantics.

There are a number of creative works which combine computation and text manipulation to create variations of the Surrealist's Exquisite Corpse [11]. Recombinant works such as Lewis' Breeder [5] and Seaman's World Generator [13] recombine texts from a corpus according to pre-marked grammatical attributes. Generative works that compose new texts according to specifications of a rule-base system are exemplified by include Bulhak's Dada Engine [1]. Evolutionary works, such as Schmitzs' Genotyp [12], make explicit use of genetic algorithms to breed new typefaces.

4. ARCHITECTURE
The architecture of Alien Letter Forms is inspired by agent-based simulations. In such simulations, agents are often modeled after real-world organisms. They are autonomous and independent entities which usually have a limited awareness of their environment.

This translates in our system by defining an agent as the most basic unit of written language--a character. Additionally, our agents carry properties of their own such as a velocity and a typeface, which affect their behaviour on screen and their visual appearance. Their position and their proximity to other agents determine what they know of their

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

environment and therefore how they can interact with each other and the space around them. (Figure 1).

Agents take advantage of the NextText library (described in section 8) to model different behaviour that can act on properties like velocity and direction to generate movement patterns (linear motion, erratic motion, etc.). Behaviours can also modify the physical appearance of the agents themselves. These behaviours are programmed as independent units and can be inherited from an agent to the next.

The simulation itself is performed on a time-step basis. At the tick of an artificial clock, each agent is updated in a parallel fashion. The method by which each agent is updated depends on its properties and behaviours but also on its interaction with other agents, which consists primarily of the reproductive mechanism described in section 6.

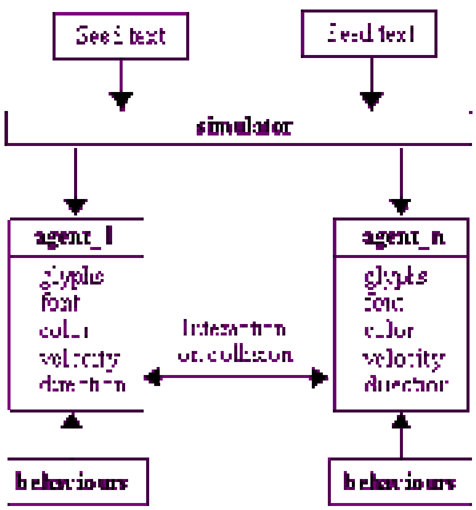


Figure 1. Alien Letter Forms' agent architecture

5. ENVIRONMENT

Agents exist as though they were floating in liquid or in space. They move freely around the environment, but tend to be naturally attracted towards other healthy agents. This movement makes the randomness required for genetic evolution visible to the observer

The environment is seeded with pre-written texts. Each seed text generates a zone of influence which is associated with a region of the screen (Figure 2). Theses zones are dynamic, meaning that seed texts compete for space amongst themselves. Their influence area grows larger as more agents spend time within their boundaries. A future iteration will allow users to submit seed texts to the system in real-time over a network connection.

The seed texts could be texts exhibiting, for instance, a contrast in content, style or even language. The content of the seed texts influences the reproductive process and thus shapes the outcome of the text generation.

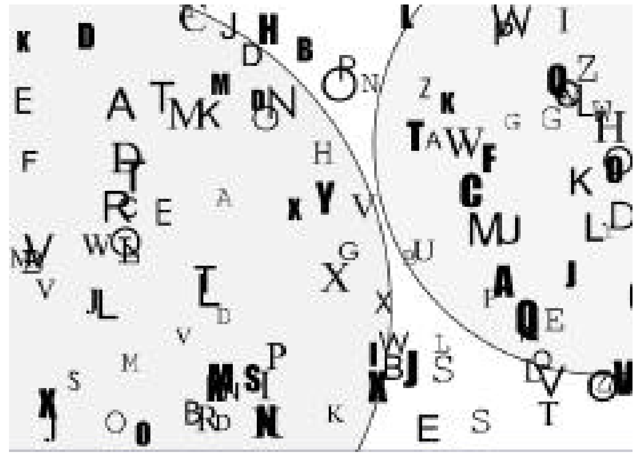


Figure 2. Sketch of the environment filled with random agents in their early stages.

6. EVOLUTIONARY MECHANISMS

In order for evolution to occur, we must set up mechanisms that aim to introduce improvements in the population by each successive generation. This is translated in the system by a reproductive function, which has a small chance of being triggered whenever two agents are within physical proximity of one another. To be coherent with the evolutionary metaphor, our reproductive function implements some common operators of genetic algorithms [8].

6.1 Fitness

Usually, fitness is a function that attempts to rate agents based on how "good" they are at performing a given task. Note that this function should be computationally effective because it will be frequently called. It should also be stochastic in order to allow the system to take unexpected directions.

In Alien Letter Forms, the task is for agents to become readable texts. Some strings will be better suited to this than others; in the English language, for instance, there is an identifiable difference in fitness between a string such as "house" and "zkcoe". In order to represent and measure fitness, agents are rated according to the relative probability of occurrence of their character string within the seed texts. The result of the function is a real value from 0 to 1 which indicates a string's fitness. Additionally, fitness is directly reflected by the agent's properties such as speed.

Fitness is computed using a data structure inspired by Decision Learning Trees [7] which conveniently allows for efficient queries at runtime as well as dynamic updates of the reference texts. Every possible combination of up to four characters has its own path in the tree. The four character limit is imposed by memory restriction since the tree grows exponentially at every branch level. When building the tree, a counter is incremented every time a node is reached. Additionally, the maximum counter is stored at each level of the tree, indicating the number of occurrences of the most frequent letter combination out of all possible strings of a certain length (where the length is the depth in the tree). We use this maximum value as a reference to which all other values are compared in order to generate relative frequencies.

Using such a tree, determining the fitness is as straightforward as following the path given by the characters of the string we are trying to evaluate and then dividing the score contained in the terminal node by the maximum score for its depth. Figure 3 illustrates the tree structure for a partial alphabet and how each level of the tree is associated to a string of corresponding length. The actual tree used in this project includes uppercase and lowercase combinations, spaces and a set of common punctuation characters.

For strings larger than four characters we evaluate their overall fitness by randomly picking a sample substring every time, thus ensuring an element of uncertainty. The use of samples in this case accurately estimates the fitness of the string as a whole over a time span because the sampling area is always changing.

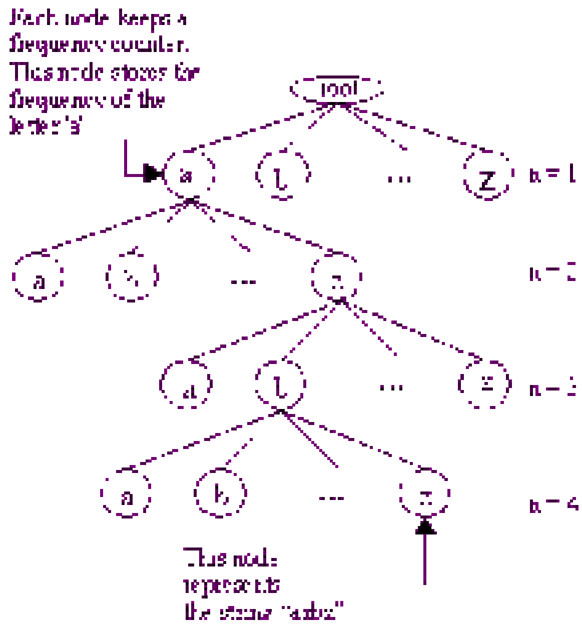


Figure 3. Partial decision tree of maximum depth (n) 4 for an alphabet [a…z].

Fitness acts as a determining factor in the application of the other operators. Whenever a fitness test has to be performed, a random number from 0 to 1 is generated. If it is smaller than the agent's fitness, then the test is a success, otherwise it is a failure. Accordingly, fit agents will have a higher success rate.

6.2 Crossover

The genetic algorithm also implies a crossover operator which simulates reproduction between individual agents. (Figure 4) Traditionally, this operation is performed by taking two agents and randomly recombining them into two children which in turn take over their parent's place in the next generation while preserving some of their properties.

In this case, however, crossover is approached in a different way because we want our organisms to grow larger over time. For this reason we let agents aggregate together during the crossover operation. Therefore, instead of producing two new offspring, agent couples merge together into one larger agent during crossover. The order in which the two agents are aggregated is determined at random.

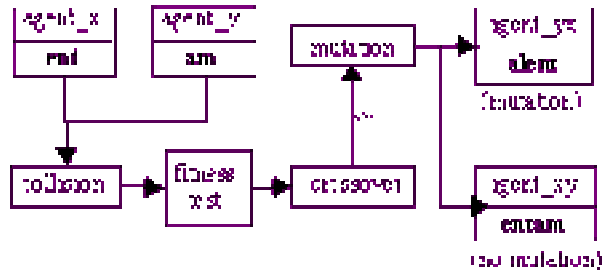


Figure 4. A scenario describing two possible outcomes of a crossover (one with mutation and one without).

6.3 Mutation

Every time two strings aggregate, we introduce a small chance that they might undergo a mutation in the process. All agents are equally likely to be affected by a mutation, regardless of their fitness. A mutation is represented by replacing one of the characters by a new random character. Mutation occurs in order to help prevent the system from stagnation.

6.4 Death

The death operator acts like natural selection. Unlike crossover, which requires two agents to be near each other, death can strike anyone at anytime. Fit agents are less likely to be affected by death than other agents.

To be sure, when struck by death, agents do not disappear entirely but instead break into two smaller strings. This is a consequence of the crossover operator which already reduces the number of agents in the population. Therefore, we chose to avoid destroying large agglomerations of agents at once. Only single-glyph agents can literally die—be permanently removed from the system—as they cannot be further reduced.

6.5 Clone

The last operator, clone, is used to allow fit agents to reproduce themselves by duplication (Figure 5). Like death, clone can happen at any time and is random, although fit agents are more likely to be selected. This operator stimulates the production of good agents in order to improve the overall quality of the population. Also, like during crossover, there is always a slight chance that a mutation might occur in the process.

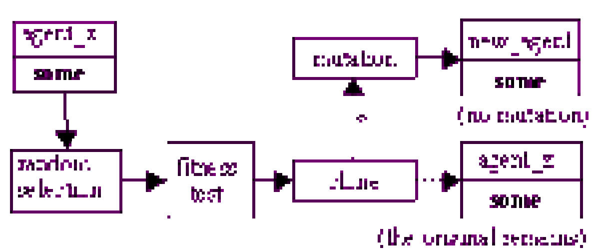


Figure 5. A scenario for a clone operation where no mutation occurs

7. INTERACTION

The Alien Letter Form environment, once it has begun, is largely autonomous and acts without outside intervention. Users can influence the system directly only through the process of agent creation. New agents are introduced to the system from text that users input. In the current system, this is accomplished by typing into a window, from which Alien Letter From pulls characters to turn into agents. The next system will allow users to do this in a more natural fashion by using a microphone and speech recognition engine.

Sentences submitted by users are broken down to letters and born again as fresh agents in the system. In the process, there is an opportunity to parse the input text (before it is taken apart) and extract semantic properties from it, assuming that the user types in or says actual sentences. The occurrence of certain words, for instance, will influence the agent's initial properties such as type of movement behaviour, font or color. Agents can also be given additional properties at birth based on the meaning of the input. An example of this is predator agents.

Predators are integral to most ecosystems yet they have been overlooked in this discussion so far. The threat they pose to other species serves the dual purpose of regulating population growth and stimulating evolutionary change. In order to transpose this phenomenon to language, predator agents are generated whenever a slang word is parsed from the input text. Slang disturbs established linguistic structures. Similarly, our predator agents are given a property that causes them to trigger the death (See 6.4) of other agents upon collision. Note that predators are not affected by death the way regular agents are. Instead, predators have a lifespan dependent on whether or not they can find prey over a given period of time. This mirrors creation of a natural life cycle that will not be dominated by a large population of predators.

8. NEXTTEXT

Alien Letter Forms relies on NextText [9], a Java library which allows manipulation of dynamic and interactive texts. The NextText architecture encourages users to regard text in both its linguistic and visual dimensions simultaneously. Instead of having to shift modes or worse yet, applications, in order to move from one dimension to another and back again, NextText provides an environment in which creators can write as well as paint with words.

NextText employs an object-oriented approach that parses input text into a loosely language-based hierarchy of glyphs, words and phrases. This hierarchy supports the easy applications of behaviors at different levels of the text, allowing the user to specify behavior for an entire passage of text in the same manner as for individual letters. Behaviors can operate on the visual, dynamic, interactive or semantic characteristics of the text. Text components are provided with simple agent-like functionality to allow them to interact with one another as well as with the user.

NextText is the successor to ActiveText, a C++ library with a similar architecture.

9. CONCLUSION AND FUTURE WORK

We have described a software framework created to experiment with genetically evolved texts. We find that the resulting environment generates interesting sub-texts and provides a rich means for re-writing texts. However, the current state of the work represents just the first few steps into this area. Further research and development remains.

One area that will be explored in a future iteration of the project regards the physical appearance of agents. Further work is planned in order to allow evolution of font faces during reproduction. The implementation has been delayed as we research how best to represent glyph outlines such that common characteristics between glyphs can be recognized, extracted and compared. We also need to do additional work on determining how fitness applies to the how fitness applies to the evolution of font faces and how it should be measured. Suggestions include nearness to a historical font class (Modern, Roman, Gothic, etc.) or legibility, although the latter remains rather subjective and therefore difficult to quantify computationally.

In the semantic direction, we are interested in breaking the text generation in two phases, the present one which operates at the character level and a new one which operates at the word level. This approach would require defining a hybrid of grammar-based and genetic models. The presented mechanisms for agent reproduction would be preserved and used to breed readable words. These words could then be assembled together according to a grammar-based system. Once words begin forming, lexical operations could be performed to drive mutations through different lexical spaces such as synonyms, antonyms and heteronyms. The newly formed word agents would employ grammatical operations to link themselves together into phrases and sentences. In order to integrate this new element into the biological metaphor, the grammar itself could be subject to algorithmic evolution.

10. ACKNOWLEDGEMENTS

We would like to thank Taras Kowaliv for his technical and conceptual contribution with regards to the implementation of a multi-agent system and the genetic algorithm. We would also like to thank Alexander Taler for his work on the NextText library. The work described herein was conducted at Obx Laboratories and is further supported by Hexagram, the Centre inter-universitaire des arts médiatiques, and the Fonds québécois de la recherche sur la société et la culture.

11. REFERENCES

[1] Bulhak, A., The Dada Engine. http://dev.null.org/dadaengine/

[2] Ishizaki, S., Multi-agent Model of Dynamic Design: Visualization as an Emergent Behavior of Active Design Agent. In *Proceedings of CHI '96 Human Factors in Computing Systems*, pp. 347 - 354, Vancouver, BC, April 1996.

[3] Leopoldseder, H. and Schopf, C. (eds), Cyberarts 2000. Ars Electronica Center, Linz, Austria, 2001

[4] Lewis, J. and Weyers, A., ActiveText: an architecture for creating dynamic and interactive texts. In the Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology Conference (Asheville, North Carolina, November 1999), ACM Press, 131-140.

[5] Lewis, J., Dynamic Poetry: Introductory remarks to a new medium. M. Phil. thesis, Royal College of Art, 1996.

[6] Lewis, J., NextText Java Library, www.obxlabs.net/research/nexttext/

[7] Mitchell, T., Machine Learning, McGraw-Hill, 1997

[8] Mitchell, M., An Introduction to Genetic Algorithms, MIT Press, 1999

[9] Nearlife, The Virtual Fishtank, http://www.virtualfishtank.com/press/index.html

[10] Pechawis, A., Nation (multimedia performance). http://www.aalab.net/ephemera/resid/1.1Nation/, 2000

[11] Rubin, William S. Dada & Surrealist Art. Harry N. Abrams, Inc., NY, 1968.

[12] Schmitz, Genotyp. http://www.genotyp.com

[13] Seaman, B., Recombinant Poetics: Emergent meaning as examined and explored within a specific generative virtual environment. Doctoral dissertation, Center for the Advanced Inquiry into the Interactive Arts, 1999

[14] Sha, XinWei, Hubbub, http://www.gvu.gatech.edu/people/sha.xinwei/topologicalmedia/hubbub/index.html

[15] Sims, K., Galapagos, http://www.genarts.com/galapagos/

[16] Soo, D., Implementation of a temporal typography system. Masters thesis, Massachusetts Institute of Technology, 1997.

[17] Wong, Yin Yin. Temporal Typography: Characterization of time-varying typographic forms. Masters thesis, Massachusetts Institute of Technology, 1995.

Partner Technologies: an alternative to technology masters & servants

Kevin McGee and Johan Hedborg
Department of Computer and Information Science
Linköping University
581 83 Linköping, Sweden
kevmc@ida.liu.se
johhe386@student.liu.se

ABSTRACT

There are activities where a dynamic, creative *partnership* among equals seems like an appropriate model of empowerment. Not only do good partnerships seem to help people attain or sustain powerful engagement in their current activities, in some cases they seem to enable people to successfully *enter* new activities.

In this paper we describe our initial work to develop *partner technologies*. Using examples from four partner prototypes being developed, we discuss some design patterns based on the work to date, some insights into aspects of partnership, and conclude with a discussion of some future prospects and potential of partner technologies.

1. INTRODUCTION

When it comes to the design of intelligent technologies intended to empower people, much of it is guided by two central metaphors: technologies as servants or as masters. Servant technologies can be empowering because they reduce or remove work that people find difficult, dirty, or dangerous; master technologies can be empowering because they instruct, inform, remind, cajole, nag, or otherwise force people to do things which are important – but which, for whatever reasons, people do not (or cannot) do without this assistance.

For many kinds of activities and contexts these guiding metaphors do indeed seem useful. But there are activities where a dynamic, creative *partnership* among equals seems like the more appropriate model of empowerment – as in the case of musical co-improvisation by jazz groups, where particular collections of individuals mutually inspire and support each other. Not only do good partnerships seem to help people attain or sustain powerful engagement in their current activities, in some cases they seem to enable people to successfully *enter* new activities.

One consequence of shifting the focus to partnerships and partner activities is that it suggests some serious limitations of objectivist accounts of meaning-making. Much of the work in applied AI is based on a traditional semiotic perspective – one in which goals, constraints, evaluation criteria, indeed *meaning* itself are all (objectively) "pre-given" ("in the world" or "in the mind"). For example, "intelligent tutoring systems" are often implemented with an assumption about the objective (and pre-existing) status of information/knowledge that is to be transferred to the student; likewise, search-engines are typically designed with strong assumptions about the ontological status of the "information to be found." These assumptions may (or may not) be appropriate for those systems – but a number of difficult questions seem to arise as a result of trying to apply the techniques used by such systems to other, more improvisational, domains and activities.

As an example, consider the musician who wishes to improvise with other musicians, and the goal (to the extent it is explicit) is for the activity to be as spontaneous, creative, and enjoyable as some of their more successful sessions. What is "similarity" in such cases – and how does one facilitate it? In what sense does the person explicitly represent (if at all) the qualities of "good play" in advance of new sessions – and what happens to such models as the person has more and more experience? Could it be that the process of experiencing more examples, rather than being a process of "refining" what is meant, is more like a process of *elaborating* and *enacting* what is meant – that is, rather than "narrowing in on a definition", it involves some combination of undefined praxis and experience emerges as a co-evolution of activity, participants, and individual experience?

This kind of activity seems different in kind from many traditional computer-science tasks that have (or can be formulated in terms of) clearly defined goals, constraints, and evaluation criteria. In the case of improvisational activities, goals (if that is the appropriate word) may be ill-defined, if at all, and changeable: participants may wish to "improve performance" or "take on as much of the partnership as possible." Similarly, the desire to do "something like" an example activity may not be well-defined, it may not even be *possible* to make it well-defined in any way that it useful for the participants, and it may actually change as a result of participating in the activity.

Some of these challenges are beautifully expressed by Belinda Thom, describing her work to develop musical systems that are intended to co-improvise with human musicians, "Altering the interactive task to one of musical companionship makes it paramount that the agent automatically configure itself so as to reflect its user's particular and momentary style. Simply put, composition – the notion of setting down ideas in advance – makes less sense" [27].

In this regard, a *non-objectivist* [30, 14] approach to the

First published at COSIGN-2004
14–16 September 2004, University of Split (Croatia)
.

study of partnership seems promising – one based on the assumption that meaning is not *pre-given* (either in "the world" or in "the individual"), but rather that it emerges out of *action* [32], in this case, *coordinated action.* And a deeper understanding of partnership can, in turn, inform the development of *partner technologies* – technologies that have the potential to empower people by actively *adapting to* and *evolving with* them. Such technologies would not attempt to lead people to pre-determined goals (by applying predetermined constraints) – nor would they necessarily just execute commands that presume the user has predetermined needs, desires, goals, or constraints to be satisfied.

Although the term *partnership* can be used to describe such things as "partnerships with materials" [24, 25], for the purposes of this paper the focus is upon partnerships that have *pro-active* participants, such as when individuals in a jazz group co-improvise. This kind of improvisation differs in significant ways from from so-called "solo improvisation," but (curiously) has no distinct English term, so we have coined the term *symprovisation* to highlight it.

Thus, there are three main aspects of symprovisation – and their relevance to the development of partner technologies – that we wish to explore here. Specifically, issues that arise when participants wish for an activity to have an *unpredictable* outcome (invention, creativity); to be *sustained* in an appealing way (rather than "attain some goal-state"); or to be *engaging* in ways that are difficult to make precise.

In this paper we describe our initial attempts to develop partner technologies, some design patterns based on the work to date, some insights into aspects of partnership, and conclude with a discussion of some future prospects and potential of partner technologies. Before describing our work, we provide a brief overview of non-objectivist AI and a review some of the work that can be considered broadly relevant to the development of partner technologies.

2. BACKGROUND

There are a number of different non-objectivist concerns, claims, models, and methods. To highlight our interest in *action* – and to avoid the awkwardness of the term *non-objectivist* – in the remained of this paper we will use the term *enactive* [30] to indicate our particular focus.

The work on enactive AI tends to be motivated by insights from either enactive philosophy of mind (constructivism, phenomenology, and the like) – or by concerns about "biologically plausible" models of cognition. One consequence is the exploration of alternative mechanisms, such as neural networks or dynamical systems models (see [30] for examples and review) – or even a reformulation of classical "symbolic" approaches in innovative ways, such as the *deictic* representations of Agre and Chapman [1]. Perhaps the only attempt to combine implementation, enactive models, and some proposal for supporting and empowering people is the controversial work by Winograd and Flores [31] on a model of *decision support systems* influenced by phenomenology, theoretical biology, and speech-act theory.

The issues are many, subtle, and largely unresolved. For present purposes, it is enough to say that much of the debate turns on what can be formalized, the extent to which it can be formalized, and how this should occur. Although it is not possible to give a brief explanation of the key issues, it is important for readers to at least know that most enactive theory is *not* proposing a "subjective" alternative to objective models, where "everything is in the mind" (which is "unconstrained by reality"). Rather, the work is to articulate, and in some cases embody in working systems, an *alternative* to the traditional objective/subjective distinction. Thus, one of the main assertions is that enactive systems are still principled, viable, and robust – but this is not because they have access to objective or universal laws or phenomena.

Turning away from philosophical issues, there is some work in the field of applied AI (artificial intelligence), as well as work on the use of computation to assist in design, that falls broadly within the category of developing partner technologies. Indeed, the organizing taxonomy of cooperative human-machine models used below comes from this latter effort [13]. These categories cut across many domains – there have been partner systems developed for musical improvisation, design support, "just in time" multi-agent decision-support systems, programming partners, and team-sport agents. In the taxonomy that follows, the technology takes a progressively more pro-active role – from "maintaining constraints" to actively attempting to fix situations that it identifies as problematic.

Constraint-Based Paradigms.

These systems are designed such that most of the work is done by the person – while the system participates by maintaining various constraints. In some sense, a spreadsheet is one of the simplest examples of this: various relationships are maintained even as a user makes changes to specific cells. More sophisticated versions of this approach include *programming by example* [5]. An interesting example in this regard is the Logo Turtle [20] which imposes constraints on how it can be moved, but which is also an enabling control metaphor. Constraint-based systems include systems to support human-to-human decision-support [31], collaborative learning [6], and creativity [18].

Critic-Based Paradigms.

These systems have models of good (and bad) praxis and are able to provide different kinds of feedback as the design progresses (for an overview, see [7]). In many ways, such systems help people by "asking questions", raising concerns, and noting contradictions.

Improver-Based Paradigms.

These systems are similar to critic-based systems, but they also include mechanisms for automatically fixing the problems they identify, as in automatic spelling-correction functions in word-processing applications or some of the *programming by example* systems referenced above.

Cooperative Paradigms.

Finally, cooperative systems interact dynamically with humans, transforming their work, and making proposals of their own. One of the earliest proposals for such a model was Negroponte's vision of *The Architecture Machine* [19] (although, in recent years Negroponte has moved away from the metaphor of partnership and speaks instead of computational servants, such as butlers). More recently, there is work on systems that co-create works of art [22], music [27], the user's interface [28], or the "browsing experience" [12]. Additionally, there is work on implementing teams of cooperative player-agents [23] and on the development of joint-control systems; one approach involves extending the subsumption architecture of Rodney Brooks [3], as in the development of semi-autonomous wheel chairs [26] and the simulation of joint-steering system for drivers of automobiles [33].

3. RESEARCH PROBLEM

There are two categories of partner-related activity that are not typically well-served by current research interests nor by conventional AI techniques and mechanisms. First, activities where the goal is to help people discover or invent something unexpected, new, surprising, or interesting – whether it is new artifacts or ideas. Second, activities where the emphasis is on increasing and sustaining ongoing *engagement*, where the *experience* is more important than the attainment of some particular goal. Some of the work on computer-assisted design mentioned earlier is starting to address the first, and some of the work on computational entertainment (games, narrative, and the like) is beginning to take seriously the second (see, for example, [16]).

Researchers working on the development of systems for musical symprovisation and collaborative browsing are among the few attempting to combine the two, though, with its emphasis on implementing explicit models of users and domains, much of that work is broadly objectivist. In this regard, the work of Agre and Chapman requires two brief comments. First, although they use the term "improvisation" (almost apologetically) to describe their model, they do so in order to stress that they are proposing an alternative to *planning* – indeed, they emphasize the *routineness* of many activities and phenomena and are very clear to say that their focus is not on issues of creativity or innovation. And second, although they make explicit reference to enactive theorists and ideas, in their writing there is a frequent implication that they believe their model of improvisation is successful precisely because "the (objective) world" is reliable.

The focus of our research is to contribute to the application of enactive insights to the development of partner technologies symprovisational activities. Our approach is to study examples of good human partnership, implement prototypes that embody possible partner techniques, and, based on the performance of the resulting systems, develop design principles for partner technologies which are then used to guide further study and elaboration of working systems.

For readers not familiar with the discipline of structured symprovisation, it is important to emphasize that it is *not* an activity in which the participants "do whatever they want." In casual usage, *improvisation* may sometimes describe unconstrained invention, but within music, theater, and dance, for example, symprovisation – and improvisation – is very disciplined, structured, and constrained. Indeed, one of the marks of excellent symprovisation is the extent to which real-time invention and creativity manages to resolve constraints. This is widely accepted in the performing arts, but is largely absent in discussions about the development of empowering technology.

To see how different is this perspective from typical AI models of activity, we need only consider the radically different status of *planning*. It is not much of an exaggerations to say that in applied AI, heuristics and other non-algorithmic methods that do not guarantee a definite result are typically considered degenerate; by contrast, in artistic symprovisation, planning is the first and foremost sign of *failure*.

The different attitudes about planning highlight another important distinction: the *enhancement of experience* versus *the attainment of goals*. Crudely put, symprovisation represents an attitude that "we do not choose destinations but headings – with the hope that the travel itself, the destination (if one should materialize), or both are interesting in unexpected and unpredictable ways." Contrast this with a dominant criterion in much of applied computer science: "it is better to arrive reliably and in the most efficient way possible at a clearly-defined destination – or to be informed at once, 'you can't get there from here.' "

For our purposes, these differences have some important consequences for both open-ended ("creative") and goal-oriented partner activities. For open-ended activities, partners need to make valuable contributions to the collective effort of both "what to do next" *and* to the ongoing enactment of "what are we trying to do"? Additionally, good partnerships allow for and support the possibility that even when the activities have more explicit goals, these may not be "shared" by participants in any usual sense of the word.

To explore these differences, the initial architecture we are developing differs slightly from most *cooperative expert systems* or *intelligent tutoring systems*. These architectures typically include components that model *domain knowledge*, various mechanisms for making inferences or drawing conclusions based on the domain knowledge, and facilities for explaining to users how they arrived at their conclusions. In the case of expert systems, the cooperative human-machine interaction is typically predicated on mechanisms for reliably and efficiently attaining pre-determined and well-defined results (or, for identifying new results that meet well-defined constraints). In the case of intelligent tutoring systems, the emphasis tends to shift to mechanisms for eliciting particular classes of pre-determined results by the person being "taught."

Our systems embody various amounts of domain knowledge or explicit constraints, but they are most different in their use of what might be called "difference reduction" mechanisms. In traditional systems, these are used as part of the overall concern to arrive at some pre-determined goal-state(s). In our systems, they are often used as part of *making hypotheses* in the form of *actions* that may satisfy evolving, mutual constraints – as the system has enacted them to that point.

On the other hand, our current approach is similar to work on expert systems in another important way. We are currently trying to identify aspects of *good* partnership – by analogy to the work done in knowledge engineering to identify "domain expertise." Thus, the results we report in this paper are mainly in the form of initial design insights – presented loosely in the form of *patterns* [2] – to potentially important dimensions of partnership. This issue, of making explicit certain aspects of design – indeed the notion of design *at all* – is vexed in the context of non-objective approaches to engineering. We will return to this point in the discussion at the end of this paper.

Before turning to the actual implementations it is important for the reader to have the appropriate expectations. The philosophical issues – and overall ambitions – discussed in the paper so far really *are* our concerns. Nonetheless, applied research on this particular combination of interests is only in its infancy – and the initial implementations are so primitive it would be easy to dismiss them entirely. However, research needs to start somewhere; and as is the case in many symprovisations, the initial work may be most interesting in terms of what it suggests to do next.

4. PARTNER IMPLEMENTATIONS

In this section we briefly describe four Partner implementations developed to further our understanding of partnership: a Typing Partner, developed to explore some aspects of co-adaptation and support for such goals as "getting better"; a Chess Partner, that attempts to address *dynamic reconfiguration* of action in a different way; a Painting Partner, that is intended to support co-creation of art works in a particular style; and a Flying Partner, that explores a technique for the emergence and co-evolution of skilled performance in a flight simulator.

The work on the Typing Partner suggests even goal-oriented activities can benefit from partner techniques that are relevant to sustaining experience. Similar insights come from the work on the Chess Partner, although the skill here is quite different. One of the insights from work on the Painting Partner is how a traditionally-conceived solitary activity can fruitfully be transformed into a partner activity. Finally, the work on the Partner raises a number of issues about "intrinsically motivating" activities, but that also have elements of danger (for both the participants and others).

It may seem counterintuitive that we explore competitive and challenging activities in the context of partnership. However, many kinds of activities – and many kinds of good partnership – involve a combination of support and provocation.

4.1 Typing Partner

We begin with a description of the Typing Partner, which in many ways is the most conventional. It is intended as a partner to help people improve their skill at touch-typing.

The basic model is as follows: the Partner challenges the person by presenting different characters – and combinations of characters – at different speeds; and the person's typed responses *challenge the partner* to present something appropriate to the person's (changing) skill-level.

Visually, the person sees "blocks" of alpha-numeric characters drop from the top of the screen – and the challenge is to type the characters before the blocks hit the bottom. Initially, characters are randomly generated in equal proportions; based on user-inputs, the presentation of characters become more "adapted" to the particular user. Adaptation takes the form of adjusting the a) actual characters presented, and b) the speed of the presentation. The current Typing Partner is basing its activity on two main parameters: typing *accuracy* and *speed* of response. If a user types the wrong key for a letter – or takes "too long" to type the correct key for a letter – the system presents that letter more often (until the system determines that the letter is "too easy").

The Typing Partner is an initial attempt to give computational form to a proposal by Csikszentmihalyi [4] that an important characteristic of *flow experiences* is that they occur "between boredom and anxiety." By this view, one of the contributing factors to flow experiences is that the activities have the appropriate ratio of *difficulty* relative to a person's *skills*: they are not too easy (boredom) nor too difficult (anxiety). This model raises some challenges for those who would like to facilitate flow: the ratio of skill to difficulty is highly individual – and it is a moving target (that is, as individuals become more skilled, what was once difficult becomes easier). Thus, we are experimenting with adaptive models where partners are constantly adapting themselves or the activity to the individual.

In this regard, there are two aspects of Typing Partner that we wish to highlight here. First, the notion that in partnership, *all* the participants are challenged "equally" – that is, appropriate to their skills and interests. Second, the issue of determining appropriate challenge is more subtle than may at first appear, even in activities based on limited parameters of characters, character-combinations, and speed. To give one example, how should a Partner react to the fact that a person keeps hitting the key that is *next to* the key for the falling character? Initially, it seems appropriate for the Partner to use this information to drop more of the same characters, based on the model that "errors are a way of identifying areas of difficulty." On the other hand, at what point do repeated errors become a signal that the task is too difficult?

Although the partner challenge dimension here is slightly more sophisticated than games such as *Tetris*, it suffers from a similar limitation. Namely, that the model of challenge involves "moving up" a simple hierarchy of challenges. As with *Tetris*, there is a certain level at which people will not be able to improve – and above which they will not be able to perform. At that point, neither *Tetris*, the current Typing Partner, nor other games like this, have any mechanism for introducing different *kinds* of difficulties.

Thus, in addition to Csikszentmihalyi's parameters of skill and difficulty, there seems to be an additional parameter related to *stasis* or repetition. Thus, if a person achieves some level of proficiency in an activity where the person stalls at some upper level of skill and difficulty, and if the activity at that level does not change in any other way, boredom will also be likely.

One approach is to add additional parameters. We now turn to an example of a Partner that attempts to address this limitation in a different way.

4.2 Chess Partner

One of the barriers to entry with chess is the complexity of pieces, rules, and situations. Whereas the Typing Partner adapted by changing the speed and frequency of discrete elements (letters), the Chess Partner adapts by creating different kinds of *scenarios*. In particular, the Chess Partner creates different variations on chess – variations with simplified rules as well as piece-movement requirements or restrictions. The central metaphor of the Chess Partner is that of an experienced player who is exploring the kinds of chess scenarios a player can play – and using the results of playing those scenarios to create new scenarios.

Initially, the Chess Partner creates a scenario in which the player must take an opposing piece in three moves, starting from a standard opening position and using the conventional rules of chess. A player is free to move any pieces; if an attempt is made to make an illegal move, the system simply restore the piece to its earlier position (but does not print anything like "you cannot do that"). The Partner keeps track of which pieces are used by the player, and as play continues it makes moves (or reconfigures the board) so that the player is challenged to use different combinations of pieces (both knights, bishops, queen, etc.) When the Partner determines that variations on this scenario are no longer challenging, it adds an additional requirement: the player *must* take an opponent piece if possible during a turn. As the player becomes comfortable with this scenario, the Partner then starts requiring that the player make moves that will lead the most quickly to a captured piece. Variations on this type of scenario are presented in increasingly more complex ways: particular challenges about taking pieces, choosing the more valuable of pieces, positioning pieces to prepare for capturing pieces, and the like. In addition, the Partner may begin to ask the player to, for example, identify all pieces that can be captured in the current move.

It is important to make clear an essential caveat about the Chess Partner: it is quite, well, un-partner-like. The current implementation has a number of specific scenarios hard-coded and is extremely limited in its ability to move between them, following a fairly rigid progression. We are in the process of developing mechanisms that will allow the Partner to invent and present its own scenarios based on the history of play, but given the limitations of the current implementation, we seriously considered omitting a description of it for this paper. In the end, we decided that, even in its current form, work on the Chess Partner raises a number of questions about what it means to adaptively create appropriate *scenarios* – or *patterns of activity* – rather than the more simplistic model of the Typing Partner which has some fairly crisp parameters for adaptively challenging the typist. Some of these issues are discussed later in the paper.

4.3 Painting Partner

When someone wishes to paint "in the style" of another painter, one of the issues that arises is the fact that different people mean different things when they say they want to do something "like that."

The central image of symprovising with the Painting Partner is as follows. Imagine two friends who have decided to create a painting together without speaking to each other. One person makes an initial sketch, then the other makes some additions to the sketch. This process continues in silence, each person elaborating upon the work of the other. In this particular case, the Partner simulates certain aspects of the painter Mondrian's work. One can imagine it in the following way: a child says to a painter, "I would like to make paintings the way you do." It is not clear what the child means by "the way you do" – so rather than asking questions, the painter decides to learn by painting together with the child. In the process, they shape paintings co-adaptively, with the child refining its request (in the form of new paintings that respond to the history of their collaboration) – and with the painter's understanding simultaneously being elaborated.

The current implementation of the Painting Partner has three different "personalities" for working with someone who wants to paint in the style of Mondrian. The first personality is one that co-creates in the style for which Mondrian is most well-known: bright colors and straight lines. The second personality is more representational, helping a person draw specific images (such as trees) in a style that is recognizably Mondrian. Finally, the third personality is a synchronous Mondrian. Whereas interaction with the two other personalities involves turn-taking, synchronous Mondrian paints at the same time as the person, while also attending to the history of the painting to that point; in this sense, painting with this personality is like real-time jazz improvisation.

When symprovising with Abstract Mondrian, a person can draw lines freely, but everything is converted into straight horizontal or vertical lines. It is also possible to draw polygons and ellipses, although these are "normalized" to the grid. And, by selecting different tools from a palette, the person can specify lines to remove, to resize, to change the color or the texture – as well as specifying various fill properties (color, shading, texture) for the rectangles between the lines. (The inspiration for this Partner was Mondrian's painting, *Broadway Boogie Woogie*.) The Partner begins by making random additions and modifications to the drawing made by the person. The Partner notes various aspects of the history of transformations made by each of them, building different hypotheses about what the person wants and likes, occasionally testing the hypotheses by undoing what the person did last, and the like.

Representational Mondrian works much the same, except that it does not transform everything into vertical and horizontal lines. Rather, it "geometrizes" the person's naturalistic drawings to a certain extent – and in the current implementation, the palette is gray-scale. (The inspiration for this Partner was Mondrian's painting, *Grey Tree*.)

Finally, although Synchronous Mondrian is otherwise similar to the other personalities, it does not "wait for a turn." It interacts synchronously with the human painter, painting at the same time.

Much like the Chess Partner, but to an even greater degree, the Art Partner is not trying to "improve" the person's ability nor to "correct" what the person does. Rather, it is trying to *actively contribute* to the joint-painting in a way that is acceptable to the person (and which meets its own constraints). To the extent that these contributions are also hypotheses about "what the person will accept," these are not used to try and change what the person is (or should be) doing, but rather to elaborate and refine an evolving model of "what is mutually acceptable." Of course, the Painting Partner has some internal model of a particular style (in this case, Mondrian's). However, in a very real sense the Partner is not even trying to develop its own explicit model of what the person *means* by "in the style of Mondrian" – and certainly not in any sense that is intended to be later confirmed or correlated with the person's model (if, indeed, such a mental model even exists in any explicitly identifiable form).

4.4 Flying Partner

The Flying Partner is an implementation of a flight simulator in which the airplane has some built-in intelligence.

One of the major concerns in the design of airplanes is the trade-off between maneuverability and control. In order for a plane to turn, it essentially has to be destabilized. Modern fighter jets are designed to be extremely responsive – and in order to attain this responsiveness, they are remarkably unstable. So much so that some of them are designed such that it is not possible for a human to fly them without computer-assistance. The airplane cannot simply maintain control because it is the pilot who knows where to fly – but, the plane cannot simply yield control entirely since the pilot will be unable to maneuver the plane (except within a limited range of control) without crashing it. Add to this the different capabilities and skill-levels of different pilots, various weather conditions, and other factors and a number of interesting partnership challenges arise.

The Flying Partner simulation is designed to address this problem in the following way. The Flying Partner initially has complete control of the plane, keeping it stable and on-course. If there are buffeting winds or turbulence, it is the Flying Partner that maneuvers the plane to regain stability. As time goes on, the Partner progressively releases control, bit by bit. Thus, the human pilot is given control over a few degrees of freedom. The Partner "observes" how the pilot handles different situations and, as the pilot demonstrates competence, the Partner yields more and more control.

Consider two of the major difference between the Flying Partner and the other Partner implementations.

First, it embodies a model of "instant participation." In other words, the pilot can begin flying the plane at once. Compare this to, say, the Chess Partner where the model is one of inventing and presenting simplified versions of the activity. The Flying Partner is designed as a joint-control (cybernetic) system. It has a set of internal correlations that allow it to maintain balance, and the challenge is how to mediate over time between the changing skill of the pilot, the intentions of the pilot (flying a loop), the various forces (weather), conditions (angle and altitude of the plane), and the like.

And this leads naturally to a second important difference: the assumption behind this model is that the Partner will always *remain* a partner in this activity. That is, in the other examples it is possible (and perhaps desirable) to participate in the activity without the Partner. In the case of the Flying Partner, the attentiveness to the pilot "never ends"; each pilot/partner pair will enact a different history – and may continuously enact a different balance of competences in the shared activity.

5. PARTNER PATTERNS

In this section we describe some initial Partner design *patterns*, motivated by and illustrated with examples of the Partner implementations, that are intended to highlight some aspects of partnership that should be present.

Support Instant Participation

One of the things that hinders entry into a new activity or domain is that there are not always good ways to *experience* "what it would be like" to participate in them. This makes it difficult to know whether to commit to the (sometimes) years of preparation and pre-requisites necessary to gain such experience. One approach to this problem is seen in the Chess Partner: create simplified versions of the activity that a novice *can* experience. However, such approaches are not always satisfactory – people want to experience more of the "actual" activity.

An alternative approach can be seen in the design of certain equipment for physical training. As an example, there are machines in which one does "chin-ups" by kneeling on a pneumatic pad that carries most of the person's full weight. Thus, the person has the experience of doing the full-motion range of a chin-up – even though the person does not have the physical strength to do so unassisted. Over time, the person can adjust the amount of assistance provided until the strength is enough to do the activity unassisted.

The Flying Partner is one implementation of this approach, where the technique can be characterized as "progressively and adaptively relinquishing control." Thus, rather than "constructing" progressively more challenging scenarios and situations (as the Chess Partner does), the Flying Partner supports the experience of immediately participating in something like the full activity, while progressively *yielding* more and more of the challenges of that activity to the participant.

This model of partnership can be generalized to such activities as riding a bicycle – or even to the development of partners that can empower swimmers or dancers (in this regard, the work in robotics on swing-dance partners [10] is promising). To flesh out this idea in a bit more detail, consider the case of learning to wind-surf, where one of the dominant aspects is *falling over* – in fact, it dominates to such an extent that people are often motivated to quit. Imagine developing a wind-surfboard that is designed to automatically maintain its balance out on the water, indeed, to maintain its balance in the presence of disturbances by a human rider. Thus, when novices first starts wind-surfing, they actually have the experience of immediate participation. And, as with the Flying Partner, they slowly handle more and more of the challenges of the waves, the wind, and the like. Not only that, but the surfboard can be designed to *initiate* additional challenges for riders who gain enough proficiency with wind-surfing unassisted (and who wish for additional challenge).

Of course, there are a number of technical challenges to be addressed in realizing such a vision. However, the current state of the art – Dean Kamen's remarkable "balancing wheelchair," the *iBot* [17], for example – suggests that it is attainable.

Therefore, for activities where a system can be implemented to perform unassisted, consider elaborating the implementation as a Partner that co-adaptively relinquishes control to the other participant(s).

Elaborate Without Prejudice

In good partnerships, partners know how to work with the ideas, proposals, and contributions of their partners. In other words, good dynamic partnership is often characterized by responses along the lines of "that's interesting – and we could do *this* with that." An important concept emphasized in theater symprovisation is "*yes, and*", as in, "yes, and here's something interesting/useful we can do with the results of what you just did." The idea is that when one participant in a symprovisation does something, the others cannot complain, or stop the activity, or otherwise reject what has been done. They must somehow incorporate it, use it, and build on it. Thus, symprovisation at its best is not a struggle by participants to impose pre-conceived ideas about how the events should unfold or what the final result should be; rather it is a dynamic, real-time creative activity that tries to satisfy certain constraints. Among the most important is the constraint that a symproviser's contribution must "work well" in the context of the current activity (and the history of activity leading up to the current moment); and, crucially, it must contribute in a satisfying way to the context of the immediate future in which the next symprovisational activity will take place.

We see this most clearly in the Painting Partner. Consider how different is its behavior to, say, that of a critic whose role is typically to *pass judgment upon* some performance or result; this can often be characterized as *yes, but* – as in, "yes, but here's what you are doing wrong." The Painting Partner does not reject, deny, or attempt to invalidate the activity of its partner. In this sense, good partners often take the contributions of their partners seriously. And seeing that partners work with their contributions, that those contributions are taken seriously and become useful in the ongoing joint-effort, helps participants develop a powerful sense of the consequences of their actions and their relevance to the ongoing activity. This also differs from what is often meant by "constructive criticism" which implies an outsider giving advice, rather than a participant constructively *contributing* to the elaboration of the joint-project.

Therefore, for activities in which participants *create* things – whether artifacts or ideas – develop partners that elaborate on, rather than critique or reject, the actions of their partners. Said another way, take the actions of partners seriously – and make meaning out of their meaning-making actions.

(Note: it is interesting to consider that even standard servant technologies can use this technique to make them more partner-like. Consider the behavior of some search engines that implement mechanisms one could liken to "assuming the user is asking something meaningful." If the returned results are few, the system assumes this might be a problem. It then looks up more popular results that are similar – by some definition of similar – and when it presents the original results, also presents the option of running the alternative query that will return more results, asking, in effect, "would you like to try this?" Compare this to other systems employed by online merchants that attempt to "suggestive sell" by taking a query that would returns few or no results and simply generating results that contain any of the search-terms. Such systems tend not to feel partner-like, but rather stupid or irritating.)

Allow Partners to Be Challenged

In order to develop symprovisational partners that are capable of rich and expressive interaction is important to implement them in such a way that they can be challenged by the actions of the person. In other words, it is difficult to imagine that people will have fulfilling experiences interacting with systems that are designed with, for example, lookup tables for every possible user-action.

One of the important dimensions of good partnerships is that each participant grows, learns, and changes as a result of the partnership. And this is true even in situations that do not seem at first glance to be partnerships, such as the relationship between a coach and an athlete. We do not mean to labor the point in too much detail, merely to say that the more adaptive the situation between a coach and an athlete, the closer it comes to a partnership in our meaning. In such cases, we would say that all the partners are equal in their relation to the challenges proposed by each other.

Similarly, the work on the Chess Partner is motivated by such concerns. The goal is not to identify and implement some set of appropriate scenarios. Rather, it is to develop a Partner that uses the performance of the player as the creative source for *inventing* appropriate scenarios. That is, to develop a Partner capable of treating the response to the challenges it proposes as challenges *it* must act upon.

Therefore, when activities involve creating something, good partners allow themselves to be challenged by the actions of other participants, rising to these challenge by treating them as sources of inspiration or constraints within which they must contribute.

Respond in Kind

In much of the research on empowerment there is a strong emphasis on making things explicit. Thus, pedagogical research typically looks at ways of explicitly representing, presenting, and helping people to remember and apply explicit rules for different kinds of tasks, whether it is spelling rules, riding a bicycle, or learning a new language. Similarly, much of the work on assistive technologies involves building explicit representations of expert knowledge – and of helping the user by presenting advice and feedback in explicit, often linguistic, form. One obvious alternative is simply to tell people to "just keep trying." Thus, if we take the example of helping someone to learn to ride a bicycle, this usually consists of a combination of "think about this" heuristics and admonitions to "keep trying."

In many activities we find the *actions* of our partners more important and useful than what they *say*. Of course, being explicit is sometimes appropriate, but, on the one hand, there is a risk that the activity drifts away from the activity of concern (to an activity of discussion, analysis, and the like) – and, on the other hand, that the activity of the partner is no help at all.

This is not a proposal for "mindless praxis" – nor even a proposal to always avoid explicit feedback. There are, of course, activities and situations where explicit linguistic understandings and explanations are appropriate, indeed, crucial. However, it is also important to realize that these are not always appropriate – and that, in fact, an overvaluation of such explications can actually hinder praxis for certain activities.

For the most part, the Partner implementations described here represent one alternative: *respond in kind*. The basic idea is that in many cases partners find most useful suggestions and counter-proposals that are similar in kind to the actions they have themselves performed. One aspect of this is to look for and identify, where appropriate, alternatives to strictly verbal or linguistic responses. Another aspect is that these responses may not be entirely symmetrical, as when the Chess Partner plays to create challenging scenarios and the participant plays to answer those challenges. Both these aspects are visible in techniques used by masters in the so-called martial arts [11]. Part of the support consists not in verbal explanations, but rather the creation of new situations in which the person is required to *experience* – and deal with – certain kinds of challenges. Highly skilled teachers are able to adapt the creation of these scenarios to the level and particular challenges of the individual. One could even say that the "cryptic" koans of Buddhist teachers embodies this in its most extreme form, as when it is used to "test the balance" of practitioners [21] with a complex use of language that requires an active demonstration of linguistic competence that is difficult to put into words.

Whatever the form, one of the main advantages of such methods is that practitioners actively work to prevent other participants from "answering back" in another mode, as when a sports-trainer requires an athlete to respond in the form of *athletic activity*. This is perhaps clearest in the case of the Painting Partner. Similarly, the Chess Partner is not designed to explicitly *instruct* the person. Rather, the Partner is mostly silent as it presents the different challenges, allowing players to invent their own models of "good play."

Obviously, none of the Partner implementations described here not very sophisticated. Nonetheless, they do embody these principles in simplified form. The Painting Partner responds to drawing activity with drawing activity of its own – and the Chess Partner responds to chess-playing with chess moves that create contexts for more chess-playing. And none of the Partner implementations allow people to "explain what they mean or want" in any other way than through the target activity.

Therefore, even when the activity is in large part non-linguistic, implement Partners that respond in kind – and require other participants to do the same.

Strive For Mutual Benefit

Good partners act to maximize *mutual* benefit [9]. Obviously, there are deep philosophical questions about the extent to which it is reasonable to describe computational systems in these terms. In what sense, for example, is it meaningful to say that these software programs "satisfy their own needs"? One view is that human beings are essentially computational, and thus, there is no difference in kind between the kind of concern expressed by humans and that expressed by computer programs. An alternative view is that there is such a difference in kind that it strains credulity to speak of computer programs expressing concerns, having interests, or caring about mutual benefit.

We are mindful of this debate and do not wish to participate in it here. Rather we simply wish to note that it seems intuitively plausible that the notion of *mutual benefit* is important to partnership – and that future research on partner technologies should address it in some satisfying way.

The key issue is to embody a particular kind of action: "how can I do something in the current context that takes into account my interests and constraints *as well as* those of my partner?" To give this idea more context, consider the difference between asking colleagues for feedback on a single-author paper – and the kinds of feedback and discussions that often occur between two people who are co-authoring a paper. Of course, the feedback from colleagues is important and valuable, but it is also different in kind than the dynamic of co-authors. One difference, we suggest, is that good feedback between co-authors tends to satisfy mutual interests. And, although we are acutely aware of the limits of this claim, it seems reasonable to say that the Painting Partner embodies this to a certain extent.

Therefore, partners should strive to ensure that their actions are of mutual benefit. In the cases where the activity involves enacting an interesting experience, it should be interesting for all; in the case of inventing or creating something, it should be something satisfying to all.

Coordination rather than Reconciliation

To what extent does good partnership involve *unified frames of reference* – "shared" beliefs, goals, and models? When discussing examples of good partnership, there is a common assumption that "what made the partnership work" has something to do with shared values or habits. It is beyond the scope of this paper to enter this topic in any detail, but only to note that "shared models" are not necessary for good symprovisation. Indeed, the results of research on group-cooperation and conflict-resolution by the Harvard Negotiation Project [8, 9] suggest that one of the contributing factors to conflict is an assumption that participants share the same values and goals – and one way to resolve certain conflicts is to develop solutions that do not attempt to *remove* differences, but rather *address them.*

Coming after a pattern on *mutual benefit* this pattern may seem to be its opposite. However, there is no contradiction. None of the Partner implementations described in this paper attempt to create or propose shared frames of reference. Rather, they act based on certain constraints – and to the extent that they develop hypotheses about aspects of their partners , these are provisional and used as the basis for constructive actions (rather than attempts to "get on the same page").

Therefore, good partners do not entirely take over an activity; and if they must (as the Flying Partner may occasionally need to), they do so in ways that all participants agree are necessary.

Leave No One Behind

Good partners are equal in a significant sense; they make contributions (which may be quantitatively or qualitatively different) that result in a process and product where none of the participants feel as if "they are carrying too much" nor that "the other people have taken over." A Flying Partner that simply takes control and flies the plane on behalf of the pilot is no partner. Nor is a Painting or Chess Partner that does all the work (no matter how successful the results).

Of course, partners need not share the same expertise, skills, or background – but there is some significant sense in which they are working together as equals. Although there may be brief instances when one partner "knows more" than another (or teaches), the result of a good symprovisation is something that all the contributors "own." Who, for example, creates the drawings made with the Painting Partner?

Therefore, the actions of good partners are such that they encourage and support further contributions by other partners.

Maintain History

This final point is less of a design pattern and more of a reminder.

If good symprovisation cannot rely on extensive planning, normative rules, or models of an objective world, what *can* it rely upon? One part of the answer is the joint *history* of participant's activity. This may seem so obvious that it does not require comment, but experience shows that when we emphasize that symprovisation involves "acting in the Now," people frequently assume that this involves "forgetting (or ignoring) the past." And this is true even for readers with a background who may be familiar with work on "reactive systems" [3]. It is important to note that the emphasis on symprovisational techniques does not – indeed, *should* not – preclude memory. Although symprovisation is very much about "acting in the current context" (rather than extensive planning), it makes use of the *history* of the symprovisation in various ways.

We need only look at the infamous "Paperclip" in Microsoft Word to see some of the consequences of having a short memory. For all the complaints about the Paperclip, it is a serious attempt to implement something partner-like. The original model for the Paperclip involved maintaining and using an extensive history of the person's activity; many of the annoying characteristics of the final product can be traced to the fact that, for various reasons, contrary to the researcher's original design, the shipped product maintains an extremely short history of actions.

Similarly, many of the limitations of our current Partner implementations stem from the limited degree to which they maintain and make use of the history of the partnership. Of course, *knowing what (and how) to do* with that history is not trivial. For now, we merely wish to signal our belief that good symprovisational systems will need to make use of their histories in ways that go beyond simply "reacting."

6. CONCLUSION

The work reported here is part of our ongoing research to develop working implementations of partner technologies – technologies that help people enter new domains or activities, or empower them in various ways within their current domains and activities. Although the work to date on partner technologies is still in its earliest stages, and the implementations are so far quite primitive, we are encouraged.

There are a number of obvious ways to continue the work reported here: implementing further the existing prototypes, improving the pattern descriptions, and re-incorporating the insights they represent back into the implementations. This work also involves the exploration of additional classes of domains (music, literature, computer programming, etc.), activities (composing versus performing), and participant epistemologies (blind programmers, dyslexic poets, deaf musicians, etc.) – and, especially, to investigate activities than can or should take place away from the (desktop) computer. Similarly, it will be important to extend the scope of investigations to understand better which kinds of support are applicable in which cases. In other words, in addition to studying different dimensions of partnership (domains, activities, epistemologies) it will also be important to consider the design perspective. For which kinds of activity are the different Partner patterns appropriate – and how would we characterize those activities? The examples in this paper try to indicate this to a certain extent, but much work remains to be done.

This work, in which enactive media studies, design theory, and cognitive science co-inform each other, also raises some larger questions.

It seems that one important aspect of partnership is the ability to elaborate upon the work of one's partners. Contrast this notion of *use* with the notion of *re-use* common in manufacturing and object-oriented programming. By that view, one identifies common components that can be reused in different contexts. But, as famously demonstrated by Kuleshov [15], semiotics and media studies raise a number of issues about what is meant by the transformation of "things" in different "contexts." What are some of the implications for the implementation of partners with the ability to *elaborate upon*, to *repurpose*, to "decontextualize the familiar"? These questions challenge us to follow Kuleshov in an effort to extend media studies, not only as an *analytical* discipline, but as one that is also *generative*.

On a related note, what happens to the notion of *design* as we develop systems that are intended to creatively co-enact? This tension is already present when considering the traditional role of theater-directors with their possible role(s) in improvisational theater groups. Is the role of a director for such groups one of design? If so, in what sense? In literary theory, the "death of the author" is meant to suggest the importance of the *reader* in constructing what is read. It may be that for partner systems we are similarly confronting the potential "death of the developer" – but it is not clear whether (or how) this can give rise to participant creation in a *computational* sense (as opposed to an "interpretive" or "cognitive" sense).

And finally, to what extent is it possible for partner technologies to embody mechanisms for the kinds of phenomena that seem important in symprovisation, such as *shared significance*, *creativity*, and *mutual interest*? It seems very likely that partnerships that are both effective and experientially satisfying are those in which the activity and/or the product of the partnership is important in some non-trivial way to all participants. It is really not clear what this can mean in the case of computational partners. Indeed, for many it is not clear that this is a meaningful concept even in biological-based cognition. Nonetheless, it will be interesting to explore possible cognitive models and mechanisms with the potential to extend or revise traditional formalist representations of creativity and cognition.

It seems appropriate to conclude this paper, which has been concerned with different ways we can empower symprovisation, by noting an intriguing proposal that cognition itself – broadly construed – may be essentially improvisational. In the words of Francisco Varela, "...the nature of the environment for a cognitive self acquires a curious status: it is that which *lends itself* (*es lehnt sich an ...*) to a surplus of significance. Like jazz improvisation, environment provides the "excuse" for the neural "music" from the perspective of the cognitive system involved" [29].

This statement expresses a trend among theoretical biologists to view life and cognition as co-evolutionary processes, and the organism and its environment as a coupled, co-adapting system. Similarly, it may be more useful to frame the study of partnership in terms of mutual adaptation rather than more conventional cognitive models of "adaption as selection for fitness by an independent environment." In future research it will be interesting to explore in more detail some of the ways that co-adaptive models of cognition and symprovisational models of partner technologies can inform each other.

7. REFERENCES

[1] P. E. Agre and D. Chapman. What are plans for? Technical report, MIT, 1988.

[2] C. Alexander. *The Timeless Way of Building*. Oxford University Press, New York, 1979.

[3] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1):14–23, 1986.

[4] M. Csikszentmihalyi. *Beyond Boredom and Anxiety: the experience of play in work and games*. Jossey-Bass, San Francisco, 1977.

[5] A. Cypher, editor. *Watch What I Do: programming by demonstration*. MIT Press, Cambridge, MA, 1993.

[6] P. Dillenbourg, editor. *Collaborative Learning: cognitive and computational approaches*. Elsevier Science LTD, Oxford, 1999.

[7] G. Fischer, A. C. Lemke, T. Mastaglio, and A. I. Morch. Using critics to empower users. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 337–347. ACM Press, 1990.

[8] R. Fisher and S. Brown. *Getting Together: building relationships as we negotiate*. Penguin Books, New York, 1989.

[9] R. Fisher, A. Sharp, and J. Richardson. *Getting it Done: how to lead when you're not in charge*. HarperBusiness, New York, 1998.

[10] S. Gentry and R. Murray-Smith. Haptic dancing: human performance at haptic decoding with a vocabulary. In *IEEE International conference on Systems Man and Cybernetics*. IEEE Press, October 5–8 2003.

[11] E. Herrigel. *Zen in the Art of Archery*. Pantheon Books, New York, 1953.

[12] A. Kerne. Collage machine: Interest-driven browsing through streaming collage. In *Proceedings of Cast01: Living in Mixed Realities*, pages 241–244, September 2001.

[13] S. Kochhar, M. Friedell, J. Marks, S. Sistare, and L. Weitzman. Interaction paradigms for human-computer cooperation in design. In *Conference companion on Human factors in computing systems*, pages 187–188. ACM Press, 1994.

[14] G. Lakoff and M. Johnson. *Philosophy in the Flesh: the embodied mind and its challenge to western thought*. Basic Books, New York, 1999.

[15] R. Levaco, editor. *Kuleshov on Film: writings by Lev Kuleshov*. University of California Press, Berkeley, CA, 1974.

[16] B. Magerko, J. E. Laird, M. Assanie, A. Kerfoot, and D. Stokes. Ai characters and directors for interactive computer games. In D. L. McGuinness and G. Ferguson, editors, *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence*, pages 877–883. AAAI Press / The MIT Press, July 25-29 2004.

[17] B. Metcalfe. More than a wheelchair, the ibot is on the move. *InfoWorld*, November 22 1999.

[18] K. Nakakoji, Y. Yamamoto, and M. Ohira. A framework that supports collective creativity in design using visual images. In *Proceedings of the third conference on Creativity & Cognition*, pages 166–173. ACM Press, 1999.

[19] N. Negroponte. *The Architecture Machine*. MIT Press, Cambridge, MA, 1970.

[20] S. Papert. *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, New York, 1980.

[21] P. Reps, editor. *Zen Flesh, Zen Bones: a collection of Zen and pre-Zen writings*. Charles E. Tuttle Co., Rutland, VT, 1957.

[22] M. Riedl, C. J. Saretto, and R. M. Young. Managing interaction between users and agents in a multi-agent storytelling environment. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 741–748. ACM Press, 2003.

[23] Robocup: The robot world cup soccer games and conferences, http://www.robocup.org.

[24] D. A. Schon. *The Reflective Practitioner*. Basic Books, New York, 1983.

[25] D. A. Sudnow. *Ways of the Hand: the Organization of Improvised Conduct*. Harvard University Press, 1978.

[26] K. A. Tahboub. A semi-autonomous reactive control architecture. *Journal of Intelligent and Robotic Systems*, 32(4), 2001.

[27] B. Thom. Bob: an interactive improvisational music companion. In *Proceedings of the fourth international conference on Autonomous agents*, pages 309–316. ACM Press, 2000.

[28] J. Vanderdonckt, N. J. Nunes, and C. Rich, editors. *Proceedings of the 9th international conference on Intelligent User Interfaces*. ACM Press, January 13-16 2004.

[29] F. Varela. Autopoiesis and the biology of intentionality. In B. McMullin and N. Murphy, editors, *Proceedings of the Conference on Autopoiesis and Perception*, pages 4–13, Dublin City University, 1992.

[30] F. Varela, E. Thompson, and E. Rosch. *The Embodied Mind: cognitive science and human experience*. MIT Press, Cambridge, MA, 1991.

[31] T. Winograd and F. Flores. *Understanding Computers and Cognition*. Ablex Publishing, Norwood, NJ, 1987.

[32] L. Wittgenstein. *Philosophical Investigations*. Macmillan, New York, 1953.

[33] R. Woltjer and K. McGee. A joint subsumption architecture as a framework to address human machine function allocation. In *Proceedings of SIMSAFE 2004: Improving Public Safety through Modelling and Simulation*, June 15–17 2004.

Slipstreamkonza Semiotics: Towards a Telemimetic Sublime in the Data Landscape

Christina McPhee

independent artist/naxsmash group productions
PO BOX 7063
Los Osos, CA 93412
www.christinamcphee.net

ABSTRACT

Slipstreamkonza is an art/science research project that imagines carbon flux and climate change as a semiotic aesthetic of the sublime. Concerning the sense of place and landscape, this work in progress paper thinks through problems of semiotic installation design and 'big data.'

Keywords

sublime, trope, slipstream, glitch, Gaia, remediation

1. INTRODUCTION

"The design of such intimate technology is an aesthetic issue as much as an engineering one. We must recognize this if we are to understand and choose what we become as a result of what we have made."

--Myron Krueger, ***Responsive Environments****, 1977*

As a visual artist, one may turn a gaze to what cannot be 'seen'. Here we move into a zone of the sublime. Sublimity refers to that which is below, beyond or immanent relative to an ontological or cognitive threshold. I assume that there is a way of expressing this indeterminate zone, or invisible condition, in both the realms of the physical and cultural landscape and in the interior, "behind the screen" topology of the electronic sublime.

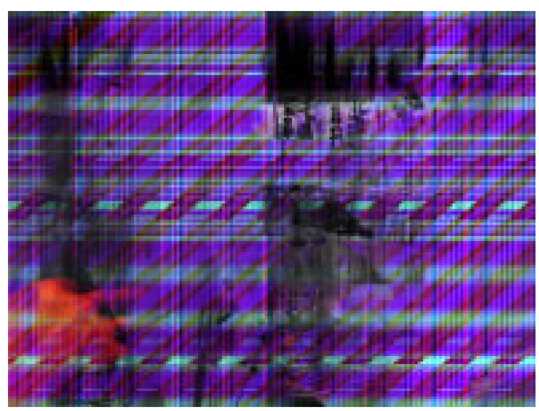
Slipstreamkonza addresses aesthetics of digital data expression of land as a breathing ecosystem. The time based data stream of carbon flux is interpreted as rhythmic, virtual expression of sound and image in net based and spatial installation.

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

2. SITE

On and near Konza Prairie, in eastern Kansas, since 1997, diurnal and annual data are collected as "eddy correlation" or "eddy covariant" flux measurements. From two of the sites, a located on the Rannels Ranch next to the Konza field station, wireless net carries the live data online for collection and analysis. Jay Ham, PhD, agronomist and climatologist, conducts research into carbon flux dynamics relative to models of climate change, at Kansas State University. He is the scientific partner for the present project.

Konza is the Osage term for "south wind." Like breath on a mirror, the metaphor of photosynthesis as *konza* suggests, to this artist at least, the evanescent imprint of an invisible and inaudible (at least on the human scale) dynamic. How to generate a cybernetic process-space that progressively and recursively self reveals, or 'voices' itself? ***Slipstreamkonza*** exists at a distance from, and following behind, and layering into, the semiotic landscape of *konza* itself, that is, the dynamic, time-based measurement and interpretation of the phenomenon of carbon respiration.



3. SLIPSTREAM

In the tallgrass prairie region, solitude opens up many hours of contemplation of invisible realities – the layers of time, memory, human and geophysical dynamics in a single place, such that time is not, to the contemplative mind, a linear vector; rather, a looping suite of simultaneous spatial layerings. This is a kind of 'slipstreaming' -- the artist working at the margins of large phenomena, catching the wind, as it were, from behind a massive, too-big phenomenon much like a sports car can catch the air wave behind a large lorry or truck on the freeway.

The sublimity to which gaze and ears and mind obsessively deflect, or defect, is the sense of landscape as huge, transpersonal, and mostly invisible and inaudible dynamic. Things might be going on just out of sight and earshot. You want to catch the waveforms of that dynamic, to surf the stream. There will, too, arise a sense that if you pay close attention to this dynamic (just to be able to stay on the crest of the wave), you will, at least for a moment, have a sense of the deep structure of a place –its phenomenological essence, if you will, or 'inscape', as Gerard Manley Hopkins once called it. Immersion in ubiquitous computing and the electronic space now turns the imagination towards intimations of a paralleling digital/real landscape, or 'slipstreaming inscape,' which involves some kind of time slippage, or transport-- between an ultimately unrealizable and untouchable Real -- and expressions of that reality in terms of electronic representation, or remediation. The nonlinear and nonfinite phenomenology of digital media, the simultaneous endlessness of reiteration, copy, and reproduction; and the continuous decay, loss, and disappearance of reiteration, copy and reproduction couples with this sense of the invisible dynamic of an unrealizable 'real' in the data landscape. In the slippage, or slipstreaming, between the condition of endless iteration in digital media and the huge volume of dynamic data measurement in scientific exploration of landscape, there exists a semiotic, or transitive zone.

4. SLIPPAGE

Thereby hangs the tale of a 'semiotic' data landscape, if "real" and "sign", never fixed, make a dynamic, Mobius loop between observation, measurement, and representation. You can never really 'be' either 'in' in the electronic space, nor participate without observational distance in the physical landscape. In the slippage between these two conceptual points is a place for the sublime in a 'data landscape. How the data is interpolated into that space becomes an issue of semiotics: the electronic space conceptually is analogous to a space of language, because the data must be interpreted and transformed via arbitrary aesthetic rules through a pipeline of code.

5. KONZA

In 2001, Jay Ham invited me to use the large volume of data associated with his ongoing global climate change studies on the tallgrass prairie. Jay participates as a partner in a global longitudinal study of carbon levels in the atmosphere relative to global climate change. Photosynthesis, during the daylight hours, takes carbon from the atmosphere, and at night, the prairie respires carbon from the surface into the atmosphere. Carbon respiration data is delivered via remote LAN into servers at the research site, and from there, may be transmitted and interpreted at other remote sites, including installation and exhibition locations elsewhere in North America, Europe or wherever sufficient server capacity exists.

The research question that drives this climatologic research also stimulates my search for tools and methods to create a work of art that refers to and embodies an aesthetic of the sublime in the data landscape. This question has to do with a mysterious shortfall, or absence, in the mathematical models we currently use to describe and predict large-scale climate change. Global warming, implicates the increasing atmospheric level of carbon as a primary agent. Nonetheless, the total worldwide carbon budget, which takes into account all known petrochemical usage on an annual basis, shows that terrestrial systems must be absorbing more carbon than we realize. Carbon flux patterns of selected microsystems worldwide, like the tallgrass prairie, may reveal conditions under which more carbon is been absorbed than is being released. From my point of view as a conceptual artist and designer, this discrepancy gives rise to an aesthetic of the sublime, e. g. the representation of something in excess, or outside of a system that cannot be accounted for in that system. With respect to semiotics of representation, the sublime refers to that which is below, beyond or immanent relative to a cognitive threshold.

At Jay's research installations on the prairie, the movement of CO_2 between the prairie and the atmosphere is measured using a method called eddy covariance. This technique requires two instruments: a sonic anemometer and an open-path CO_2 analyzer that operate continuously throughout the year. The sonic anemometer measures the velocity of air in all three Cartesian coordinates by measuring the speed of sound between paired transceivers. Data are collected very rapidly (ten times per second). These data are coupled with results from the gas analyzer (also collecting data ten times per second) that show fluctuations in CO_2, water vapor concentration and fluctuations in air temperature, to calculate the number of CO_2 molecules moving vertically above the surface (towards the surface or away from the surface).

Slippage, or slipstreaming, between the present continuous volume of six million data per day coming from the scientific installations on site, and the remote, or "telemimetic" transport of those data into a sonic and figurative language space is a crucial design problem for ***Slipstreamkonza*** as installation art. It is because you are not 'there' that, paradoxically, you can be telematically present to the data or it to us via a semiotic looping in sonic and visual forms. To date, ***Slipstreamkonza*** has only involved the creation of digital prints and video that include visual and sound abstractions generated from interpolations of code based on saved samples of data, together with photographic documentation of the technological installations on the prairie. Recent exhibitions of the ***Slipstreamkonza*** project have occurred at San Francisco, St. Louis and Berkeley in 2003 and 2004; and online in the magazine SCALE published by the University of California San Diego (2004). Honored and surprised by the intensity of the positive international response to this work in progress, I am driven towards a more thoughtful problematization of the design of the proof of concept for ***Slipstreamkonza***, in anticipation of its further elaboration as installation art. The design questions that arise in the context beg for significant feedback from the computer semiotic community of scholars and artists at COSIGN.

6. SONIC GAIA

As a place of continuous ruin and simultaneous regeneration, the networked space of electronic communications is re-presenting, itself. A semiotic model may offer us the net as a subjective topology, a synaptic process-space. Semiotically, it ‘voices’ itself. A model of the net as a live voice finds some echo in analogy to the Gaia hypothesis on the nature of the physical landscape. As life, Gaia persistently self-represents, or emits information about herself [1]. This is an old idea in new dress. “Day by day pours forth speech,” declares the Psalmist.

The problem is that such a voice doesn’t necessarily make sense, becoming “music of the spheres” or of the land. The stochastic or noise aspect of the sonic expression is important because it emphasizes the inaccessibility of meaning, of what is ‘really’ happening on the prairie. I have begun to design data driven sonic topologies that loop reflexively into audio and video installation, exclusive of overt interactivity. ***Slipstreamkonza*** resists a participatory or interactive art installation because its identity, or ground of being, is in the prairie landscape itself rather than in the installation space. And yet, this prairie landscape, in my view, is unknowable, despite its limited self-expression to the five senses and to the statistical labyrinths of data collection. This anterior reference, to something beyond or behind or below the level of perception, that is motivating a mysterious expression in audio and video conditions, extends an obsession in my lifelong artwork, with ephemerality, absence and memory from photography and painting.

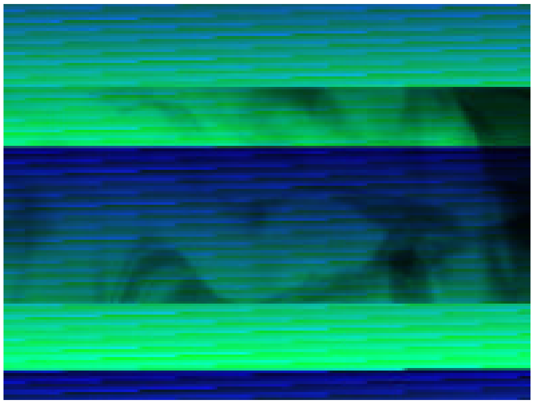
7. SUBLIME

There aren’t any claims here for a pure 'nature'. Once we are in the realm of electronic emulation of data we can no longer claim to be creating a situation to which nature 'directly' responds.

This is true even a work of art that appears to model direct encounter with nature, such as ***The Lightning Field*** (1977), by the American sculptor Walter De Maria in remote southwestern New Mexico.

I see the installation of ***Slipstreamkonza*** in a human scaled, installation space as analogous to a situation of interpretation, or secondary manipulation, of ***The Lightning Fields***. To pursue this analogy: imagine you might put a web cam on at ***The Lightning Fields***, and wait for the electrical storm footage from when the lightening fields actually work (extremely rarely). Then live video could stream into a remote site for installation. This is a three-step interval, lightening storm, web cam recording, streaming packets to installation site. ***Slipstreamkonza*** could be the kind of installation like Lightening Fields wherein very little occurs except in these incredibly rare intervals, maybe the action occurs for a 1/10 sec on Sunday morning at 3 am. As soon as one decides to design for the human condition, however, rather than for the vision machine, one has to address visual and auditory style, timing, delivery -- in a sense, cinema and architecture, and ultimately, the semiotics of data as art.

Thus we arrive at a human design problem. How can one reveal the condition of artificiality as an aesthetic premise in itself, in the installation? How to bring big data into a human scale so that it is visual, sonic and in a scale that is interesting to the primate level of reality? I think too that if you take the data too seriously, as if you can somehow ‘represent’ the reality of the prairie in the installation, that you are at risk of simplifying the content of the prairie into pretty packets of sound and image. Pretty soon it’s just kitsch new media. So what?



So the design program must concern itself with how to critically interpolate, rather than represent; to remediate, rather than to show, a remote physical phenomenon, that of the carbon flux on the tallgrass prairie. I like a kind of weird collaboration with the Gaia hypothesis rather than in an attempt to show or demonstrate the supposed truth of such an hypothesis.

I have chosen to look at the data conceptually as a flawed or entropic formal array, for reasons that honor the artificiality of the installation situation and the incontrovertible aspect of the sublime, i. e. that it cannot be accessed.

8. GLITCH

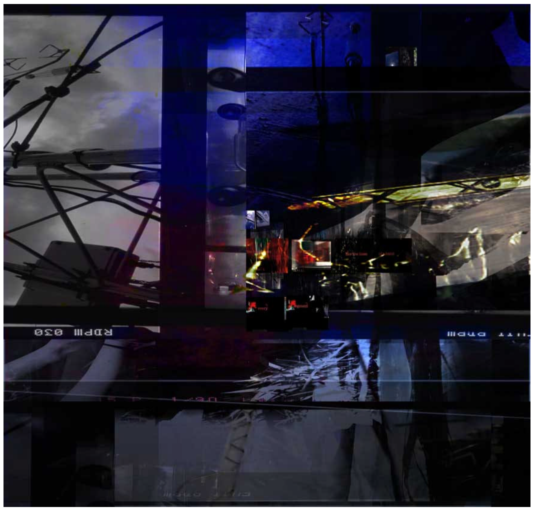
The data has a number of flaws in it, instances of 'flat affect,' such as values like 9999999 or 0000000. These flaws can be appropriated arbitrarily as a part of the aesthetic of the sublime, because the flaws are integral to the data landscape, and because the consciousness of the artist and the tools of the artist are in a condition of indiscriminate immersion in the data. It is indiscriminate because it is impossible to ascertain what preconditions of meaning may be assessed in a purely aesthetic semiotics of the data landscape. Remember that we deliberately discard any attempt at scientific visualization (or sonification as the case may be). Our only recourse is to remember that like the ecosystem of the prairie from which it derives, the data landscape itself may be described as continually subject to entropy, following the second law of thermodynamics. Life itself may be thought of arising, like a phoenix from ashes, as an articulate resistance to entropy. A continuous dialectic between entropy and the architectural self-structuring process of life means that homeostasis is predicated on breakdown, or ruin. Data stream is not always continuous. Scientific instrumentation for measurement and transmission of physical data may fail. Anomalies of landscape data are not always explicable based on known models. Humans struggle with the limitations of their bodies, including, fatigue, inattention, illness and mortality.

A *telemimetic* aesthetic of the sense of place in the data landscape accommodates breakdown of the 'language' of information streams. It is mimetic insofar as it represents itself relative to a precessive content (landscape data) and does so at a distance from itself. Telematic art is asynchronous communication: between the in and out of data feed and interpretation, there is an alteration in time and space. At the point of rupture, in the place between, is to be found the 'sense of place' in virtual topologies. Thus the Platonic view of an anterior, or precedent Form, which comes into consciousness only through a physical expression, is *undermined* by the feedback loop into digital media installation. Even though it is tempting to assign transcendent values to a digital media 'expression' of data, I have come to resist such thinking. I would rather play with the traces of data within artificial structures of semiotic meaning, such as paradox.

I am pessimistic regarding the possibility of creating an aesthetic expression of the data that responds on any level to an anterior reality that the data is supposed to be reflecting. Layers and layers of time and meaning conspire both to create a vivid and sonically exciting array and at the same time, relentlessly resist assignation of cognitive significance. I think it is very important to include the data glitches, the mistakes and miscues of corrupted data as much as the supposedly 'accurate' data, without resort to any kind of precursive truth or reference to the Real, or to an ultimate Platonic form. At the same time, as I am free of any obligation to representation as in scientific visualization, I have been happy to throw out what appears to be corrupted or excessively noisy patterns. I have done this with Java driven images and am working on the sound now. If Gaia has speech, it is an inflected, provisional, medium-specific speech. No claims for representation of reality hold up in the end: even the data is a manufactured event or infinite series of events, and when it is fed back into a feedback loop of audio and video expression, it continues to represent only itself. The installation as autodidactic and autopoetic -- in this perhaps I wonder if it analogizes the carbon data landscape itself.

9. TROPE

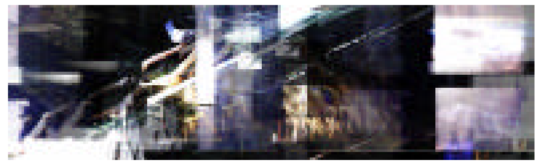
A first attempt at visualizing the data in an arbitrary abstraction was undertaken in spring 2003 using Java scripting to convert arrays of data. In this instance, only RGB values were assigned to median arrays of values. Extremely low and high values were dropped because they would not yield visual content. This choice was arbitrary and driven by artistic taste without regard for scientific visualization. The results were cast into Final Cut Pro as a video of animated stills layered with flashes of the digital print suite. This experience suggested to me a middle path between two kinds of data landscape constructs: one in which the data is assigned abstract numerical values (as in Java) and one in which live photographically and videographically acquired imagery and live sound are acquired on site at the remote data collection installations.



In the video, sound tracks and digital prints of ***Slipstreamkonza*** so far, combination of the two tropes. A trope is, in linguistics, the figurative use of an expression. The two tropes or modes of data, are, one, a kind of 'accumulation and assignation' and two, a kind of 'illumination and acoustic exploration'. These couple or slipstream past and into each other. At the point of slippage is the deep architecture, or design program, for the installation itself. The conceptual precursors of the project itself remain intact, but the aesthetic expression becomes one of arbitrary and ephemeral character, a work or works of art. The looping between these tropes, offers up a sense of place that is neither entirely of the world of generative code nor of the world of documentary photographic and localized sound capture. This sense of place is at the border between two or more incommensurate conditions. Therefore it becomes a third trope, a paradox. A paradox is a proposition that is or appears to be contradictory but expresses some measure of truth. The tension between these sets up clashes as well as harmonics, and, I hope, a baroque range of effects between extreme darkness and light, between articulation and blur, between noise and tonal wave.

10. DESIGN

Three parameters of the aesthetics of this data interpolation design can be addressed in the balance of this paper. One has to do with the scale of the data and its accessible aspects, on immediate ('live'), 30 minute, 24 hour, and annual data flows in compressed timescales. The second has to do with how we might develop sonic and video conditions from the data source. In the second topic, we are using one 24-hour data set from 2002 as a prototype for thinking through possibilities of sound expression as waveforms corresponding to the breathing of the prairie. The third issue has to do with software. When looking at the data access problematic, I have been thinking in terms of using Macromedia Director to handle the data transmission to web and installation spaces. But, I have also been attracted to the possibility of using Pure Data or MAX/MSP/Jitter for the conversion on the fly from data to sonic and visual dynamics.



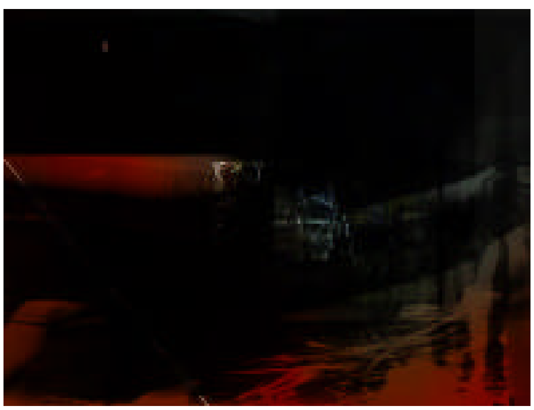
Two designers have recently collaborated with Jay and me to address issues of data access and interpolation, each from a perspective of his own discipline. Will Bauer (Edmonton, Alberta) is an interactive designer who has most recently designed remote data installations for Raphael Lozano-Hammer. Henry Warwick (San Francisco, California) is a code based musician and performance cinematographer who produced the San Francisco Performance Cinema Symposium in 2003.

11. ACCESS

Real-time data flows of 7 or more parameters per site measured at a rate of 10 measurements per second (10 Hz) may be available "live" but we have to be careful not to overwhelm the site's wireless connectivity bandwidth. Additionally, without post-processing of this data, the data make fairly chaotic or turbulent structures. Since it is hard to find patterns because of all the extra "noise" that is filtered out by the post processing, Sipstreamkonza as installation will probably use summaries from the field. Summaries of this real-time data are post-processed by laptops in the field and are stored as files. These thirty-minute data summaries contain the post-processed 10 Hz data plus a number of other, slower changing, parameters - perhaps 60 to 70 parameters in total are calculated/collected every 30 minutes. These thirty-minute summaries do show clear "breathing" patterns over a diurnal cycle. We can download these files as they are produced (i.e. have a new one every 30 minutes) rather than wait for the 24-hour summary files that are currently produced.

Historical data are available, going back to 1997 on the site. This makes for the interesting possibility of also visualizing annual data flows on a compressed timescale (sort-of like time-lapse photography) as part of the interactive environment we create. There are interesting annual and seasonal patterns as well as just daily ones that are perceivable in these data sets.

We are in the process of capturing sound clips and a web-cam image from at least one of the field sites. Due to the bandwidth limitations of the wireless data telemetry equipment, these clips and images may update slowly (e.g. perhaps also every thirty minutes) but they would still provide an interesting reference point with some interesting compositional possibilities arising from use of the sound clips (which Jay says will be mostly "wind" sounds).

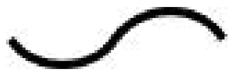


It is possible to make the data files available on a university-based server. Given the size of the data files involved (21 Kbytes for each daily aggregate of the 30 minute data files) the relatively small size of the data files also predicts a net based version via a projector in Macromedia Director to access the web links through FTP or streaming protocols or from a remote computer acting as a reflector (to avoid loading the remote site computers with multiple file requests from many people wanting to view the piece).

12. REMEDIATION

Data can be manipulated to generate information in a number of ways. One way is through scaling or distributing the data over time. Using a twenty-four hour sample, it is possible to take specific data points over (x) time and use it to describe a waveform. Making a waveform is one thing - making one that is sonically useful is quite another. For example, a column of data might have a nice sinusoidal waveform:

but this makes for a very boring sound - pretty much like a flute or making an "oooooooo" sound. Sometimes the sine wave found may not be very strong:



And this would simply be a very quiet "oooooo".

The other waveform that was immediately found was a sawtooth:

This makes a buzzier sound than a sine wave. When combined with a triangle wave, it has a quasi-violin line sound. With a sine wave it makes for more of a "mmmmmmm" kind of sound.

These various waveforms can be set to modulate each other in synthesis. The less rhythmic or wave-like data that forms chaotic or stochastic wave forms

can be used to control the modulation between these other waveforms, or can be modulated upon. This can be done using MaxMSP, where data would be loaded into a field that is managing the output or control of a given MaxMSP module and its effect on another module. Each of these can be given data from the dataset. If one has two out of phase waveforms modulating each other (using FM synthesis or even simply filtering one another) and the sawtooth sweeping another filter range, an entirely different range of sound can be modulated using the more stochastic waveforms, making the sound modulate (pitch / volume / filter) chaotically. The rate of this stochastic change can be further modulated by other data, stochastic or waveform. The waveform can also be slowed down in this way. Take the above wave form which (we'll pretend for the sake of argument) the following values:

5, 9, 8, 2, 1, 9, 8, 2, 9, 4, 6, 3, 5, 9, 8, 1, 6

and if these values were controller values for a MIDI pitch modulation in the key of C, it would be:

G, D^, C^, D, C, D^, C^, D, C, D, D^, F, A, E, D^, C^, C, A

If we then assign time points to each and have it modulate itself over time, it would come up as: (^ indicating an octave up).

GGGGG,D^D^D^D^D^D^D^D^D^,C^C^C^C^C^C^ C^C^,C, D^D^D^D^D^D^D^D^D^, C^C^C^C^C^C^C^C^, DD, C, DD, D^D^D^D^D^D^D^D^D^, FFFF, AAAAAA, EEEEE, D^D^D^D^D^D^D^D^D^, C^C^C^C^C^C^C^C^, C, AAAAAA

Furthermore, the pitch needn't be so strict and note based: it can be quite fluid as a portmanteau can be assigned - and controlled by another data set.

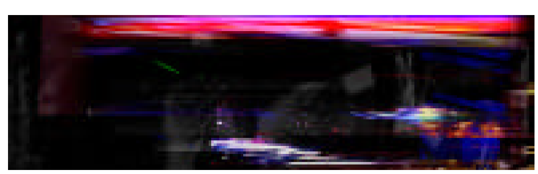
13. IMMERSION

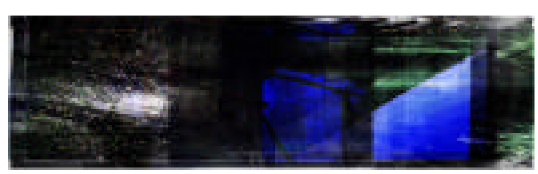
Should all data be used, or should only data that is not corrupt (i.e. isn't pegged at 99999 or stuck at zero)? My feeling is that the anomalous data does not need to be retained, since we are not trying to represent an accurate measurement of carbon flux. Alternatively, the problem becomes, then what level of data filtering is appropriate? Does one keep as much as possible and use as much as possible, or does one pick as choose the data points from the dataset according to an arbitrary sensibility? In the case of the video work I have done so far based on a combination of the java-interpreted code visualizations and photography, that an apparently ‘arbitrary’ or chaotic choice function is preferable. I like to think that this approach allows for the possibility that the carbon flow datascape can influence me formally on unconscious levels of creative work, as when the mind is preoccupied with a work of art while dreaming. A possibility is that human synaptic pathway performs as a layer of dynamic connotation, whether or not I as artist am fully conscious of the same. Or, to put it another way, the landscape represents itself in and as layers of time and human presence. Recursion and flow, between natural data and human/machine is an interpolated topology [2]

As we develop stochastic and waveform values from the data, these can be paired with the live sound clips from the site. The resultant coupled content creates the dissonant slippage or slipstreaming that both allows for and counters the development of a musical motif or ‘song’. Even if each day's data set could be a logical "song" length, given the massive scale of the data, we doubt that the data will vary much from day to day, except for "glitches" in the data. These "glitches” could form components of interest and analysis: when something suddenly goes "quiet" or pins itself at 99999 or (-99999) or gives some vastly anomalous reading, the results could be used for their own meaning/non-meaning slippage.

14. CONCLUSION

Slipstreamkonza desires the extremes of boundary conditions and interlocking means, all in service towards a description of something indescribable, a remote perfection that appears chaotic at our level of resolution but might make formal sense from the perspective of a viewer on the furthest edge of the Milky Way. Slipstreaming, we are immersed in the inexplicable: to all observers, why do the carbon flow data show an incommensurate condition? Where does the atmospheric carbon go but into the slipstreamed topology of pixels and waveforms—the topology of semiotic imagination.





ACKNOWLEDGEMENTS

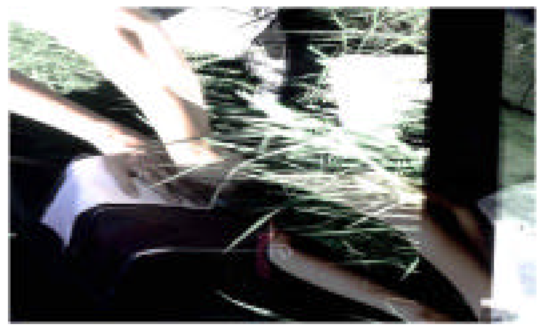
Portions of this paper first appeared in an earlier essay, "Sense of Place and Sonic Topologies: Towards a Telemimetic Sublime in the Data Landscape," published in issue 6/7, SCALE, online at http://scale.ucsd.edu, July-August 2004, Brett Stalbaum, article editor, and Michael Podolak, issue editor, University of California-San Diego; and YLEM Journal, vol. 24, no. 6, May 2004, print edition, San Francisco, California.

My sincere thanks to Jay Ham, PhD, Will Bauer, Henry Warwick, Brett Stalbaum, Molly McPhee and Terry Hargrave for their continued critical interest in the development of this work in progress.

REFERENCES

[1] Geri Wittig has looked at the Gaia hypothesis relative to the discourse on landscape data, holism and science, and includes a bibliography on this topic, at <http://www.c5corp.com/research/complexsystem.shtml>.

[2] Brett Stalbaum asserts that "data's role in the instantiation of the actual may be a matter of virtual informatic interrelations (or external relations between data sets), forming their own consensual domains that heretofore have not yet been observed as such, but which potentially inflect the operation of actual systems via informational transfer between neighboring systems of interrelations." (http://www.noemalab.com/sections/ideas/ideas_articles/stalbaum_landscape_art.html)



The Anti-Poetic: Interactivity, Immersion, and Other Semiotic Functions of Digital Play

David Myers
Loyola University New Orleans
New Orleans Louisiana, USA, 70122
dmyers@loyno.edu

ABSTRACT
The essay examines some of the assumptions of early formalist theory and practice – particularly as regards the widely applied formalist concept of "defamiliarization" (*ostranenie*) – in order to extend the semiotic analysis of interactive media found in Myers (*The Nature of Computer Games*, 2003). That analysis describes new media interactivity as displaying semiotic functions formally similar (but often in functional opposition) to defamiliarization.

The essay argues that, using a cognitive framework, formalist principles and assumptions can be comfortably extended to describe the aesthetic experiences associated with the use of computer-based media (most particularly computer games) – and, further, that the *literariness* of poetic language is formally similar and in opposition to the *interactivity* of digital media.

Keywords
Aesthetics; close reading; computer games; digital media; formalism; interactivity; immersion; poetics; semiotics.

1. INTRODUCTION
This essay discusses the importance of early formalist literary analysis to current analysis and understanding of interactive, computer-based media forms. I am particularly concerned here with an understanding of computer games as aesthetic forms or, as I would like to argue hereafter, as "anti-poetic" forms.

First, a brief summary of early formalism.

2. EARLY FORMALISM

2.1 Important Figures
Historically, the formalist movement is most often associated with two separate bodies of work: that originating within a relatively radical group of Russian critics during the early 20th century – well described in Erlich's *Russian Formalism* [3] – and the publications of the so-called "New Critics," a cadre of (primarily) poetry critics working in United States academia during the 1920s.

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

Wellek [14] assigns Viktor Sjklovsky (*On the Theory of Prose*, 1925), Boris Eikhenbaum (*Melody of the Russian Lyrical Verse*, 1921), Yuri Tynyanov (*Archaists and Innovators*, 1925), and Boris Tomashevsky (*Russian Versification: Metrics*, 1923) leadership in the Russian formalist movement. Sjklovsky's early essays make the explicit claim that "the literary work is nothing but form" and that all art is, in fact, "outside emotion" (as cited in Wellek [14]). While Sjklovsky's views may have been extreme among his fellows, the desire to isolate and analyze literature as a formal derivative of natural language was characteristic of the formalist approach on both continents.

Prominent within the American formalist movement were John Crow Ransom (*The New Criticism*, 1941), Cleanth Brooks (*The Well Wrought Urn*, 1947), and William K. Wimsatt, the author (along with Monroe Beardsley) of the "affective" and "intentional" fallacies (*The Verbal Icon*, 1954). Like their Russian predecessors, American formalists eschewed literary analysis based on either intent of author (the intentional fallacy) or individual and private effect on reader (the affective fallacy). And, despite great differences in cultural backgrounds and political ideologies between the Russians and the Americans, these two early 20th century groups have come to be linked in their common goal of studying scientifically, measuring empirically, and defining objectively the formal properties of "literariness" (*literaturnost*).

2.2 Formalist Methods
America's New Critics introduced the methodology now most closely associated with formalism and still the single most sustaining contribution of formalism to literary analysis: the "close reading" of texts. Close reading consciously avoids all interpretations referring to and depending on elements extrinsic to the text. During a close reading, formalist critics attempt to isolate objective components of texts – e. g., rhythm, meter, and imagery in poetry – that are most characteristic of and fundamental to literary form.

As practiced by the New Critics, this analytical technique is similar to earlier, linguistics-inspired analyses conducted by the Russian formalists. Each is an attempt to introduce scientific methods to the study of literature and, by extension, culture. And, while each was successful in identifying and cataloging meaningful components of human language (see, for instance, Jakobson [6] – about which more later), each also suffers in its inability to move from the analysis of specific components of texts to an explication of more general principles of literature.

In its most isolated and restricted use, engaged solely in the effort to locate literariness, formalist methodology raises uneasy questions concerning the relative importance of (and thus the precise formal relationships between) literary *form* and *content* – or, later, concerning the relative importance of *structure* and *materials*. Close reading begs the question of how much knowledge of context and use of language is required prior to formal analysis. And, indeed, the implicit requirement that formalist critics possess some relatively advanced expertise prior to the application of formalist methods undermines the objectivity of those methods. It is for this reason that New Criticism, in particular, is often regarded – and criticized – as an elitist approach.

Revisionist formalist methods adopted to acknowledge and include the influence of context are of two basic sorts. The first applies entirely different methodology to the measurement of context – critical methods – which subsume the professed scientific objectivity of formalism within social conflict (and, for the Russian formalists, Marxist) paradigms and, subsequently, within increasingly less formal and more *structural* models.

The second retains the objective premise of formalist techniques and applies those techniques to both individual components of texts *and* to the relationships among them. These relationships are then taken as indicative of contextual *systems*. Erlich [3] makes much of the methodological evolution of formal analysis to systems analysis, which served as a precursor to the development of *semiotics*.

> During the 'heroic' period of Russian formalism, the science of signs was virtually non-existent. …But by 1930, … this new discipline was well under way. The theory of language was being fitted into the larger framework of a philosophy of symbolic forms which considered language as the central, but not the only possible system of symbols. (pp. 158-9)

2.3 Formalist Assumptions

> "Poetry is language in its aesthetic function."
> Roman Jakobson, *Modern Russian Poetry*, 1921.

Historically, formalism originated in ideological opposition to existing theories of literature (e. g., symbolism and impressionism), and the methods employed by formalists purposefully ignored pre-existing theoretical contexts. However, formalism involves a linked set of assumptions about the nature of language and literature, which were neither often nor completely acknowledged by early formalist critics.

At the core of both Russian and American formalism is the notion that literature serves a particular aesthetic function apart from that of everyday or conventional or common language. In *Art as Technique*, Sjklovsky describes the purpose of art (including "artistic" or poetic language) as reestablishing the "process of perception." In this function, art "defamiliarizes" those objects to which it refers, creating a sense of strangeness (*ostranenie*). Ostranenie then *re*-engages the process of perception, as that process exists prior to its mediation by language. During this re-engagement, literature functions in a manner somewhat akin to phenomenological "bracketing"; literature defamiliarizes language through a self-referential process with consistent and measurable formal properties.

The above sequence entails a number of assumptions about the nature of language and mind. However, before preceding further, it is first necessary to deal with the potentially misleading term of "perception" in the above – a term which Sjklovsky purposefully disassociated from Aleksander Potebnia's earlier claim that art was "thinking in images." Sjklovsky and other formalists clearly repudiated this particular distinction between practical and poetic language, and, therefore, the "process of perception" referred to by formalists is perhaps better thought of, in a more general sense, as a process of *semiosis*, or, even more generally, as a process of *cognition*.

Given this realization, formalism can be interpreted as an early form of cognitive science with its goal to find formal properties of sign and symbol systems indicative of formal properties of the mind. (Compare, for instance, the relationship between formalism and cognitive science to that between formal and cognitive linguistics.) And the most basic theoretical assumptions of formalism remain consonant with those of cognitive science – with one important omission. Those basic assumptions are these:

- The function of literature is to evoke a subjective but universal human affect (ostranenie), based on common and consistent phenomenological properties of language, i. e., a common and consistent *aesthetic*. This affect is best measured objectively, though indirectly, through measurement of the sign and symbol system (literature) that promotes it.
- While formally and functionally distinct, literature is part of the same sign and symbol system as common language. Literature applies sign and symbol relationships of conventional language in unconventional ways, e. g., in the form of trope or verse. Therefore, the literary function of language is not unique but *derivative* of the common function of natural language. Differences among characteristic types of language (poetic vs. non-poetic) are then differences based on sequence, or syntax, or relative *functions*.
- The primary function of common language is to familiarize (*automisation*); the function of literature (e. g., poetry) is to defamiliarize. Thus, the latter is dependent on and cannot occur without *reference* to the former – again emphasizing the derivative nature of literary form.

These three assumptions – that literature has universal form and affect; that literature is derivative of common functions of natural language; and, related, that literature functions as a self-referential (or language-referential) form – allow the adoption of formalist theory and methods by *semiotics*. However, it remains a bit of a stretch to place early (and current) formalism within the broader context of cognitive science without the further assumption that the subjective experience of literature originates within and is determined by biological properties of the human brain. This assumption – of a *cognitive* aesthetic – was not (and is not) necessarily a part of a formalist agenda. Explanations of the natural-historical origins and causes of human aesthetic experiences remain outside formalist theoretical domains; it is exactly this omission that has allowed the appropriation of formalist methodology by structuralism and other theoretical contexts.

Significantly missing from early formalist theory is a detailed theoretical explanation of just how – and for what reasons – *automisation* takes place. Thus, even when stated in its most positivist guise, formalism remains reactionary and more clearly delineated by its methodology and critique of existing bodies of literary theory than by its own unique theoretical stance. Likewise, formalist theory as it is currently applied to the study of digital media greatly depends upon what set of theoretical assumptions are used to contextualize its findings.

However, given the basic assumptions above, regardless of the origin of universal properties of language, when and if such are revealed through formal analysis, these properties must to some degree reflect universal properties of cognition. That is, the "self" in a self-referential system cannot simply be that system per se, but rather must involve the self-referential process that sustains that system. To isolate form in human language is to isolate some aspect of form in human self (or, using a mechanical analogy, in the system "engine" which generates self).

3. FORMALISM IN DIGITAL MEDIA

Digital media provide rich opportunities for formal analysis due to their reliance on an explicit *code*. And, in fact, most formal analyses of media engage the relationship between media codes and human codes, such as language. There are several flavors of media (code) theory distinguished by the degree to which formal properties of digital code are assigned influence and priority over formal properties of human perception, cognition, and experience.

One of the more recent examples of a purely formalist approach largely unfettered by contextual concerns is Andersen's semiotic analysis of programming languages [2], in which he considers aspects of digital signs "unique to the computer medium" (p. 216). This analysis de-emphasizes all affective and interpretive components of digital media in favor of classifying digital code – i. e., the digital sign and symbol system – solely on the basis of its relationship to other sign and symbol systems. For instance, Andersen assigns a single set of objective properties to an "interactive" sign: an interactive sign accepts input, has mutable features, and can affect features of other signs. When using such a definition, no assumptions need be made or implied concerning the function of interactive signs within human interpretive systems.

Others, however, more strongly emphasize the importance of human interpretive systems and functions in classifying digital signs and symbols – and, correspondingly, give as much attention to the aesthetics as the form of digital media. Manovich, for instance, distinguishes between "transparent" and "non-transparent" digital code with reference to the ability of that code to transform human thought. [8]

> In cultural communication, a code is rarely simply a neutral transport mechanism; usually it affects the messages transmitted with its help. For instance, it may make some messages easy to conceive and render others unthinkable. A code may also provide its own model of the world, its own logical system, or ideology; subsequent cultural messages or whole languages created using this code will be limited by this model, system or ideology. Most modern cultural theories rely on these notions which I will refer to together as "non-transparency of the code" idea. (p. 64)

If digital code – or any alternative sign and symbol system – is indeed transparent, then purely formal analysis such as Andersen's rightfully ignores all supposed distinctions between poetics of texts and poetics of digital media. However, if code is non-transparent – a position media determinism takes to the extreme – then formal analysis is only the first step in establishing the mediating relationship between the study of poetics and the study of cognition. And, in fact, there is an even stronger cognitive-based position: that the transparency of digital media code is an indication of its influence and its origin rather than its "neutrality." That is, the "logical system" provided by digital code is transparent – and relatively intractable – precisely because it parallels analogous systems in human semiosis and cognition.

For the remainder of this essay, I would like to examine some of the common functions of computer games as those functions are related to the early formalist notions of *automisation* and *ostranenie*. Then, based on the formalist assumptions outlined above, I would like to describe the relationship of computer game forms and functions to literary forms and functions.

> It is now widely maintained that the concept of "literariness" has been critically examined and found deficient. Prominent postmodern literary theorists have argued that there are no special characteristics that distinguish literature from other texts. Similarly, cognitive psychology has often subsumed literary understanding within a general theory of discourse processing. However, a review of empirical studies of literary readers reveals traces of literariness that appear irreducible to either of these explanatory frameworks. [9] (p. 121)

3.1 Computer game functions

Aesthetics is most concretely the study of the human senses – or that which livens or awakens or gives pleasure to the senses. When the early formalists attributed the function (or "affect") of ostranenie to poetic language, they did not consider this an affective fallacy for two reasons. First, the formalist position assumes that the effect of poetic language is common and predictable – that is, the function of poetic language has an objective nature and quantifiable form. And, second and related, ostranenie does not affect the individual so much as it affects the raw senses of the species; this assigns a universal – even involuntary and mechanical – quality to the poetic which provokes a single, sense-based aesthetic response: a state of heightened awareness in which, according to Sjklovsky, we "recover the sensation of life."

What might be the corollary of such an affect as regards digital media and, most particularly, computer games? One of the more obvious candidates is media *interactivity* and associated "immersion." Yet, while interactivity is probably the most often cited distinctive formal component of new media, the term regularly eludes precise formal definition. I will not take the time to review this mysterious elusion in depth, but let me offer two representative examples of interactivity definitions – one already mentioned.

As scientist, Andersen [2] prefers to locate the interactive process in objective characteristics of computer-based signs without immediate comment on or concern with the interpretive value of that process; as humanist, Aarseth [1] defines the interactive process more generically as "ergodicity," which is then understood as determined by quantity of reader effort expended rather than by the specific quality of reader affect evoked. Neither of these couch interactivity within a broader *functional* context similar to that of Sjklovsky's, wherein media interactivity might be understood as a derivation, transformation, and/or translation of human semiosis.

I have previously attempted to demonstrate how media interactivity is both derivative and transformative of existing sign and symbol patterns, or human semiosis [12]. That is, the characteristic pattern of new media interactivity entails a specific formal relationship among signs: a temporal sequence of significations during which successive signs are used to construct a context within which subsequent signs are interpreted, valued, and giving meaning. Thus, interactivity is a process of *recursive contextualization*.

Computer games – and, in fact, all types of human play – clearly exhibit formal patterns of recursive contextualization. Computer game experiences are perhaps most distinguished from those associated with literature by their extreme repetitiveness, leading to extended recursive computer game designs (e. g., a continuous progression through endless "levels") and, simultaneously, both extended and discontinuous play. While single computer game play lasts far longer than the time required to watch a single movie or read a single novel, that play is marked by a long series starts and stops, saves and reloads. Paradoxically, then, the computer game seems to retain its novelty and appeal during what superficially appears to be repetitive and monotonous play behavior.

What is the affective function of such repetition?

Prototypically (though there are some exceptions to this), computer games engage the human senses much more directly and immediately than do genres of literature. In many games – e. g., first person shooters – mastery of game mechanics and interface is a necessary prelude to play. Nor is this a temporary impediment, which, once overcome, is no longer important to play. Constant attention to and manipulation of game mechanics is required throughout computer game play, even when these requirements recede from the conscious awareness of players. In fact, game play is more enjoyable precisely when the attention to and manipulation of game mechanics recedes from conscious awareness and the player is fully engaged or immersed in the game. Therefore, one function of repetition and recursion in game play may well be to engage and thereby familiarize the senses, leading to a phenomenological state of "unawareness" or, in early formalist terms, *habituation*.

Another important distinction between game experiences and literary experiences is the degree to which the latter are essentially personal while the former are always at least potentially social. To play is to play *with* some idea, object, or person; and, as computer game technologies have evolved, it has become increasingly common to incorporate multiplayer components into computer game designs. Thus, while computer games may function as private experiences, they also have the potential, unrealized by literature *during the process of reading*, to function as social experiences as well.

This means that the relationship between the computer game player and the digital code is ultimately quite different from the relationship between the reader of literature and the code of language. Literature remains essentially a process of communication in which meanings and values are transmitted (or shared, if you prefer) from person to person through a common code system. Computer game play remains essentially a process of experience in which meanings and values are neither definitively made nor permanently grounded in the digital code of their creation.

Certainly, computer game code is encapsulated to some degree within the rules of the game, but, during play, computer game players both abide by and, on a frequent and regular basis, test, revise, and transcend game rules. Much computer game play is motivated by a sense of mastery of, power over, and movement beyond the rules of the game. This same "movement beyond" the code of language, on the other hand, would quickly render natural language and literature meaningless.

Literature does not formally deconstruct conventional language so much as it calls our attention to it, allowing the defamiliarization process to occur at the level of cognition (or semiosis) rather than at the level of language per se. Unlike literature, however, computer games (and digital media aesthetics in general) are not circumscribed by the embedded rules of a natural language.

Phenomenologically, digital code remains a *simulation* of human sign and symbol systems and, as a simulation, can have neither physical ground nor visceral referent in the language-bound relationship between familiarization and defamiliarization. If the code of language, as the early formalists implied, ultimately refers to and is determined by the relationship between the function of human senses and the function of human sensory processing, then digital code ultimately refers to and is determined by no more or less than *reference* itself.

Thus, while the rules of language bind, restrict, motivate, and focus the literary experience, the rules of digital code have no similar impact on the experience of computer game play. Computer game design may well incorporate (simulate) literary functions, but these cannot serve the same function as literature insofar as these functions are incorporated *into the digital code* (e. g., become a part of the rules of a game). For, once part of the rules of the game, then the rules of language must submit to the same transformations (i. e., referential functions) as all other components of digital code. When simulated by digital code, these rules of language reference something *else*.

Computer games using the rules of language as a design element – notably so-called "interactive fiction" – have found it difficult to combine the distinct aesthetics of literature and play. Several critics now recognized the dissonance of these respective forms, but do not therein acknowledge their fundamental incompatibility. Montfort [10], for instance, describes interactive fiction designs as only "potential" narratives (yet narratives nonetheless); similarly, the literary-inspired analysis of Ryan [13] emphasizes the use of narrative patterns and processes within interactive digital media to generate a variety of "possible worlds." The question remains, however, whether the semiotic process necessary to generate potentials and possibilities is not antithetic to the semiotic process necessary to read and interpret narrative.

Use of defamiliarization techniques within interactive digital media – such as computer games – must, in fact, ultimately fail due to the inability of those techniques to directly reference and viscerally access the *embodied* code of natural language. Any object subject to a defamiliarization process must have been transformed first by a familiarization process; computer games resist such a process. Computer game play familiarizes, but cannot itself, as play, be familiarized. That is, to play with play reduces simply to play. While literature reveals the underlying mechanics of an embodied language, computer games and similarly interactive and playful digital forms reveal only the contents (i. e., the *emptiness*) of a disembodied semiosis.

In summary: four common characteristics of computer games and play distinguish the play of computer games from the reading of literature and are, therefore, associated with a distinctive computer game play aesthetic. These characteristics are 1) the raw mechanics of the digital media interface, 2) a discontinuous and repetitive (i. e., *recursive*) play, 3) the reference to and transformation of game rules during recursive play, and 4) a unique (non-language-based or *disembodied*) relationship between the game player and the game code. Based on these characteristics, computer games are best classified as a supra-literary aesthetic form; and, thus, the computer game aesthetic is, in the sense offered by the early formalists, *anti-poetic*.

Poetic language defamiliarizes conventional values and meanings through reference to the embodied mediation of the senses by natural language. Computer games – and the formal process of interactivity – engage and invigorate our habituated senses through a simulation of human semiosis. This simulated process – marked most definitively by recursive contextualization – displays the formal pattern of a meaning-making *event* without ever terminating (as do the formal patterns of language and literature) in a specific meaning made. From a formalist perspective, all literature tends toward the poetic, which references the habituations of natural language; likewise, all computer games tend toward the simulative, which references only something *else*.

3.2 Computer game forms

Can formal analysis identify those components of digital media and computer games that evoke familiarization? In order to do so, that analysis must focus on relationships among signs and symbols within computer games analogous to those relationships within human semiotic systems, including but not restricted to language.

Currently, a great deal of formal analysis of computer games has focused on the delineation of computer game genres [1] [15], which for the purpose of our discussion here, I will collapse into three broad categories: action/arcade games, role-playing games, and strategy games.

I have argued elsewhere [12] that the fundamental form of these three genres is determined by the semiotic processes associated with their play. Action/arcade games emphasize oppositional relationships among signs; role-playing games emphasize contextual relationships among signs; and strategy games combine these two in a process of *recursive contextualization*. In brief, each genre builds upon the previous, so that strategy games display a culminate form of play in which computer game play is both a mimicry of (something different from) and a model of (something self-similar to) human semiosis.

This peculiar semiotic form associated with computer strategy games – a form which is not what it represents yet which formally represents itself – I call *anticonic*, in opposition to those signs which are what they represent yet do not formally represent themselves, i. e., icons. However, rather than revisit that argument in detail here, I would like again to turn to parallel formalist examples from the early part of last century.

Roman Jakobson was one of the youngest of the early Russian formalists and the member of the original group who perhaps proved most facile in applying formalist principles and techniques within other theoretical disciplines. Also one of the founders of the Moscow Linguist Circle, Jakobson made multiple contributions to linguistics and literary theory. Most pertinent here is Jakobson's classification of literary genres on the basis of their characteristic tropes or, put more generally, characteristic relationships among signs.

Whereas I have previously classified semiotic processes as either oppositional or contextual [12], Jakobson establishes a similar binary division within human semiosis marked by "selection" and "combination." [6] Jakobson then argues, in formalist fashion, that broader literary forms are derivative of these two most basic and fundamental forms. Jakobson associates "selection" with metaphor and, at the level of genre, with romanticism; he associates "combination" with metonymy and, at the level of genre, realism.

It is easy to find parallels between Jakobson's analysis, my own, and that of contemporary computer game critics such as Espen Aarseth, who identifies two formal "master tropes" characterizing not only all computer game play but all "hypertext discourse." In parallel with early formalist claims, Aarseth's tropes display distinct phenomenological affects. The first is *aporia*, a feeling of confusion or helplessness among players – a state associated with the initial awareness and processing of oppositional signs such as those confronted during initial exposure to the physical interface of action/arcade games or, as Aarseth notes, during encounters with difficult puzzles or major obstacles within any game. The second of Aarseth's master tropes is *epiphany*, resulting from the resolution of oppositions (and, thus, the resolution of aporia) through a contextualization process. [1]

Further, Aarseth classifies his tropes as "pre-narrative," existing apart from (or at least prior to) those semiotic processes associated with language and literature. Similarly, Jakobson's analysis implies [7] that, while metaphor is fundamentally an intralinguistic form, metonymy is *metalingual*. From this, we then must assume that any formal analysis regarding such forms must also be metalingual (i. e. non-language-based or determined).

My own analysis concurs with this line: that there exists both a formal and affective distinction between poetic form and computer game form; that this formal and affective distinction establishes the interactive aesthetic form as the more fundamental form (e. g., as either supra- or metalingual in nature); and that the interactive form roughly corresponds to what early formalists referred to as the habituation of the senses. I would, in fact, argue even the stronger position that, as a result of the above, narrative relationships are incongruous and frequently dysfunctional when applied within interactive computer game designs. [11]

Given that the habituation of the senses – part of a familiarization process – occurs prior to the mediation of natural language and, necessarily, prior to the defamiliarization process attributed to literature, it is not unreasonable to transform formalism assumptions into a more fully developed science of cognition by locating the values and meanings of habituation (and, thus, the values and meanings generated during computer game play) within what Grodal [4] calls human "cognitive architecture":

> Media cannot change our innate cognitive and emotional architecture, only invent products that may activate and enhance the innate specifications. (p. 146)

One of the more curious characteristics of the "innate specifications" of play is the absence of an endpoint. That is, play has no built-in terminating function – similar, perhaps, to the circumstance of "unlimited semiosis" (Peirce). And the most basic formal components of story – beginning, middle, and end – seem an interruption within the timeless flow of the game.

Of course, there are designer-imposed, frequently arbitrary, and often disheartening endings to games, but these are seldom greeted with a great sense of player satisfaction. And, more often than not, these endings *ex machina* are incongruous adaptations of otherwise self-similar formal design elements – e. g., the recurring mobs of action/arcade games, the leveling of characters within role-playing games, the multiple contexts, scenarios, tactics, and replays of strategy games, and, indeed, even the repeating cycle of Aarseth's generic aporia-epiphany pair.

Jakobson believed we could learn more about the nature of language from its limitations (e. g., those observed in aphasia patients) [6] than from its achievements; likewise, we may well learn more about the nature of digital media from its failed appropriations of literary form – such as story and narrative –than from its widespread depiction as "hypertext."

3.3 Computer game play

"Play tends to remove the very nature of the mysterious."
Roger Caillois, *Man, Play and Games*, 1961.

Computer game play is a different phenomenological experience from that of reading. That is, computer games and literary forms appeal to different *sensibilities*. And, correspondingly, the aesthetics of play are distinct from the aesthetics of reading. This latter seems obvious, but is not always apparent among those who find close parallels between computer games and literature.

The mutable and transformative properties of play have led some to assume that playful simulations of literary forms (e. g., interactive fictions) function in a manner reminiscent of their originals. And, upon superficial and cursory play, prior to full engagement with interactive media, among players familiar with literary experiences, perhaps this is true; however, upon repeated play, the experience of meaning-making during computer game play colors the values of all meanings made therein.

Computer games function as human sign and symbol systems, and, as such, share a finite set of semiotic patterns and processes with literature, just as literature shares those patterns and processes with natural language and conventional texts. Thus, reading can be considered in this sense derivative of play, just as literature can be considered derivative of natural language. Play is distinguished most fundamentally in that it seems to be at the root of the derivation process and, correspondingly, unaffected by it. Play, in other words, stands alone, rigid in form and dominant in influence.

For these reasons, all signs and symbols within computer games are ultimately interpreted as *icons*. That is, all signs and symbols within computer games are ultimately valued and given meaning only within the system of the game itself. Thus, games tend more toward the simulative (i. e., without regard to referents) than the simulation.

Examples I have used earlier to demonstrate the iconic qualities of computer game signs are the classic games *SpaceWar* and *Hammurabi* [12], though many other examples serve equally well. Though these two games originated as simulations of space combat and city management respectively, during play their signs become disassociated from their real-life referents and more definitively associated with their roles and relationships within the context of the game interface, interaction, and rules.

Whatever might be strange or mysterious concerning a sign or symbol is systematically removed and replaced with the immediate sensation of the sign itself and the accompanying ability to value and understand that sign during play. Games, therefore, do not elicit a sense of awe or wonder as does literature, but rather grant a sense of self-satisfaction and self-identify. This immersive experience of the semiotic self is no doubt part of the addictive-like appeal of computer games.

The willing suspension of disbelief within the literary experience is then replaced in computer game play by the active reinforcement of the phenomenological experience of self. And any interruption in this reinforcement process is more likely to break the player's concentration, focus, and pleasure in completing the specific task at hand than to void any carefully constructed and maintained game "fiction."

For instance, playing game sequences out of order is as common as designing sequences of game play as independent modules, often linked only by the most superficial of narrative overlays (e. g., Pacman-like cut scenes). Literary forms, in contrast, depend greatly on the temporal sequence of their presentations (e. g., their plots), which are intended to have a cumulative effect. Not so with computer game forms. Various parts of computer games may be played more or less frequently, in or out of order, with greater or lesser degrees of enjoyment, regardless of the player's orientation to any supposed (or actually present) literary allusions and forms.

Certainly, an aesthetic of play must recognize player desire for and game design components contributing to some measure of unity and structure, but this function is then accomplished by means appropriate to and consonant with pre-existing human play behaviors. In lieu of story and narrative, for instance, designer-imposed game "winning conditions" frequently serve a unifying function somewhat similar to that of theme or motif in literature. Even in games without explicit winning conditions – e. g., multiplayer role-playing games such as *EverQuest* and *Ultima Online* – players impose their own winning conditions to structure and guide play.

Under certain designer- or player-imposed winning conditions, then, it is conceivable that the computer game's semiotic functions might be forced to resemble those of literature. That is, perhaps either designers or players might impose the same goals, themes, motifs, and the like as those encountered within language and literature. Yet, within the context of digital media and computer games, all such resemblances must remain formal simulations – and therefore distortions – of natural language and its accompanying defamiliarization through poetic form.

Indeed, play may at times "ascend" into literary form, just as literature may at times "descend" into play. Crossword puzzles and puns, for instance, are forms of play which defamiliarize language and thus might be considered lesser or partial forms of literature. There is no corresponding "lesser or partial" form of play, however, which enables the functions of literary forms.

Computer game role-players who bring literary sensibilities into multiplayer games commonly attempt to shape game play in such a way as to construct a story or drama – often to the dismay of other players. The computer game role-player in MMORPGs is in conflict with – and often serves as a source of amusement and/or frustration for – those players who value signs more strictly according to objective game rules (e. g., the min-maxers). And a literary sensibility alone is not sufficient to provide a literary experience.

Most role-play within MMORPGs takes place either among small groups of offline friends or else in periods of short duration wherein large variances in the values and meanings assigned to game signs have only a fleeting impact on individual play. Larger groups (e. g., guilds), which role-play on a regular basis, must adopt a strict regimen of rules and regulations – very similar to those of the broader game of which they are a part – in order to force recursive patterns of play into structures more recognizable as romance or fantasy or similar literary genre. These rules and regulations quickly become more analogous to the form of games than the form of literature, and, insofar as these simulated literary genres remain within an interactive media context, individual play – and a common aesthetic of play – treats these rules and regulations like any other component of play: as objects to be manipulated and transformed.

This is true of all designer- and player-imposed game winning conditions, regardless of their resemblance to literary form. Within computer game play, a game won is only seldom an indication of play over. Computer game winning conditions are most often doled out only in small increments and a single, isolated win is usually an incentive for further play.

Likewise, in competition among human players, single game outcomes are seldom understood as definitive; and, when involving competition with computer AI opponents, the majority of computer game designs provide multiple opponents, multiple levels of difficulty, and/or multiple scenarios of play. Indeed, winning conditions themselves are often made variable as a part of the game design and therein become susceptible to player choices and desires – both inside and outside the formal context of the game rules.

Yet, despite the persistent reference to and eventual transformation of game rules during play, there remain formal constants of play design and desire, or a common and formal play *aesthetic*. Regardless of the specific achievement, or score, or quality of play required to "win," for instance, there is the widespread assumption that game play should be "fair" and that winning conditions should be as equal as possible for all players. And, correspondingly, those players who manipulate game rules and change winning conditions in such a way as to create inequalities of play are normally accused of "cheating" – just as are computer AI opponents who do not obey the same rules as their human counterparts.

But, if game rules must ultimately conform to an aesthetic of play that allows the manipulation and transformation of game rules, to what extent is rules manipulation destructive to the rules (or the code) of the play aesthetic itself? That is, aren't breaking the rules and changing the winning conditions – cheating – logically and necessarily part of the same aesthetic sensibility promoting equality and fairness in play?

Defining a formal aesthetic of play often reveals paradoxical aspects of play such as these, which, although an integral part of human semiosis, the formal analysis of games cannot by itself interpret.

Though some amount of rules manipulation is always expected (and observed) during computer game play, the degree to which this manipulation is characterized as proper or improper, or the precise level of rules manipulation dividing, for instance, the power gamer from the casual gamer, cannot be discovered through formal analysis alone. While formal analysis of game rules and play behaviors suffices to describe the semiotic patterns associated with computer game play, that analysis is silent concerning the values and meanings assigned to those patterns – paradoxical or not – within specific social and cultural contexts. Rewards (or punishments) for play, the relationship of play to work, and the current social status of computer games and their play are all topics indicative of the degree to which play and computer game forms remain part of the social context in which games are created, played, and, in most cases, bought and sold.

Likewise, just as social and cultural contexts can affect values and meanings related to games and play, the principles and techniques of literary form may impinge on game play in the same contextual sense: from the outside in. *During and within play*, however, computer game play displays a rigid integrity of form and a common aesthetic. For this reason, formal analysis of computer games is better focused on the form of computer game *play* than the form of computer game "text" or design.

Early formalist approaches failed to deal with the form and theory of reading in as much detail as they did the form and theory of texts. As a result, theoretical rivals to formalism came to include reader-response theory and hermeneutics, both of which emphasize the power of the reading process and the function (rather than form) of the text *as read*.

While formalism currently retains the ability to reveal, document, and classify objective and recurring characteristics of games most common and significant to game play, the phenomenological process of play – i. e., the *semiosis* of play – remains hidden from direct observation. And, therefore, the relationship between game form and cognitive function remains speculative. However, if human cognitive and/or semiotic functions have formal correlatives in interactive game play – as I have suggested – then it seems at least worth the attempt to apply formalist methods to the study of the subjective experience of play.

This has, sort of, happened before.

As early formalist theory evolved, there were attempts to develop theoretical positions encompassing both the universals of form and the variety of functional structures ultimately determining their effects. Jakobson's "phenomenological structuralism" (see Holenstein [5]) and Ricoeur's "phenomenological hermeneutics" combine the study of reader/player experiences with the formal study of texts. And, for that reason, both appear to have the potential for more valuable insights into the nature of play and games than does a purely formalist approach.

Structuralism, post-structuralism, hermeneutics, discourse analysis, social semiotics, and many other contemporary literary theories have applied early formalist methods to an understanding of values and meanings within social and cultural contexts – and rightfully so. However, as I have indicated here, early formalist theory seems, in retrospect, more appropriate for extension into cognitive science than social science.

In fact, if there are universal properties of human cognition and semiosis and play operating in parallel with and thus reinforcing formal characteristics of games, then there is much to be gained from pursuing a phenomenological hermeneutics in which the "interpretive community" is precisely the biological origin and natural history of the brain.

The most basic argument asserting the primacy of the biological code in determining form is then simply this:

Given two systems of code, one in human cognition and one in digital media, which system could be expected, during some extended period of mutual interaction, to most fully adapt to the other? That is, which system would display the greatest amount of variation and adaptive change over time?

Obviously, it seems to me, the system which is more flexible and capable of change will adapt more quickly and more radically to its environment than will the system which is less flexible and less capable of change. Thus, the code associated with digital media forms is more likely to adapt to the code associated with human interpretive processes – rather than vice versa.

Here, of course, you must assume that there is indeed a common, universal, and biologically determined human interpretive process and an accompanying "code." But this is an assumption implicit in early formalist work and, in my mind, necessary to claim basic and common functions of language such as familiarization and defamiliarization.

Giving at least face validity to early formalist assumptions about the common and universal nature of sensory habituation mediated by natural language, the same simple argument might also be applied to social and cultural "codes." These, too, seem more obviously amenable to sudden change and variation than biological codes governing human cognition. Thus, according to the same logic as above, social and cultural rules and systems are ultimately more likely to display adaptations to human cognitive properties than vice versa.

In any case, formal properties of computer games and computer game play clearly demonstrate the use of sign and symbol systems distinct from the sign and symbol systems of natural language and, most particularly, literature. While reading literature demonstrates that the shared values and meanings of conventional language are only one of many possibilities, playing computer games demonstrates that multiple values and meanings might result from a single semiotic process.

4. REFERENCES

[1] Aarseth, E. J. Cybertext: Perspectives on Ergodic Literature, John Hopkins, Baltimore, 1997.

[2] Andersen, P. B. A Theory of Computer Semiotics, Cambridge University Press, Cambridge, 1997.

[3] Erlich, V. Russian Formalism, 3rd Edition, Yale University Press, New Haven, 1981.

[4] Grodal, T. "Stories for the Eye, Ear, and Muscles: Video Games, Media, and Embodied Experience," in Wolf, M. (ed.). The Video Game Theory Reader, Routledge, New York, 2003.

[5] Holenstein, E. Roman Jakobson's Approach to Language. Indiana University Press, Bloomington, IN, 1976.

[6] Jakobson, R. "Two Aspects of Language and Two Types of Aphasic Disturbances," in Jakobson, R. and Halle, M. (eds.). Fundamentals of Language, Mouton, The Hague, 1956.

[7] Le Guern, M. Sémantique de la métaphore et de la métonymie, Larousse, Paris, 1973.

[8] Manovich, L. The Language of New Media, The MIT Press, Cambridge, MA, 2001.

[9] Miall, D. S. and Kuiken, D. What is literariness? Three components of literary reading, Discourse Processes, 28 (1999), 121-138.

[10] Montfort, N. Twisty Little Passages: An Approach to Interactive Fiction, The MIT Press, Cambridge, MA, 2003.

[11] Myers, D. "Bombs, barbarians, and backstories: Meaning-making within *Sid Meier's Civilization*," in Bittanti, M. (ed.) Ludologica. Videogames d'autore: *Civilization* and its discontents. Virtual history. Real fantasies, Edizioni Unicopli., Milan, Italy, 2004.

[12] Myers, D. The Nature of Computer Games, Peter Lang, New York, 2003.

[13] Ryan, M. Possible Worlds, Artificial Intelligence, and Narrative Theory, Indiana University Press, Bloomington, IN, 1992.

[14] Wellek, R. A History of Modern Criticism 1750-1950. Volume 7: German, Russian, and Eastern European Criticism, 1900-1950, Yale University Press, New Haven, CN, 1991.

[15] Wolf, M. "Genre and the Video Game," in Wolf, M. (ed.). The Medium of the Video Game, University of Texas Press, Austin, TX, 2002.

Narrativity in User Action: Emotion and Temporal Configurations of Narrative

Shachindra Nath

Central Saint Martin's College
of Art and Design
University of the Arts London
s.nath@synesis.tv

ABSTRACT

One of the core problems in Narrative Intelligence is maintaining the narrative nature of event sequences that emerge owing to user participation. This paper challenges the common premises and assumptions about the nature of human action and experience that underlie common approaches to finding a solution to the problem of narrative structuration. An in-depth analysis of the temporality of human action and experience provides important indicators on how the problem can be approached. It is argued that user emotion is not just a by-product of narrative structure, but a critical factor in maintaining narrativity. Finally, it is indicated as to how patterning of emotions can regulate user action and the creation of a subjective experience.

Keywords

Narrative Intelligence, Narrativity, Emergence, User Experience, Action, Emotion, Subjectivity, Immersive, Virtual Environment.

1. INTRODUCTION

It has been proposed that narrative experience in IVEs be viewed in analogy to the concept of *direct experience* [60], [18], [30], [49]. A *direct* experience of a particular object is an experience resulting from direct manipulation of the object, while an *indirect* experience is the process of forming an attitude toward an informational contact [22]. A *direct* experience of a story is living it, and an *indirect* experience of the same story is watching it, hearing it or reading it. The *direct-ness* of the experience also implies the emergence of narrative through action rather than passive experiencing. The story doesn't just happen to the user she *makes it happen* to herself. *Narrativity* and the *narrative process* in IVEs and *emergence*, then, need different approach compared to conventional pre-structured media.

A deeper look at the aspects of user experience is necessary while identifying the elements of narrativity relevant to the medium. Since the experience in the medium is fundamentally distinct, the creation of subjectivity or the meaningfulness of narrative experience would not follow the same process as in the representational media. The structuration of a narrative has its purpose in indicating the creation of this meaning.

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

The medium lends to a user a bi-dimensional position that lies on the continuum of *experience* and *action*. The former is indicative of the 'spectator' mode while the latter assigns an 'actor' position to the user. For the purpose of discussion in this paper, I will adopt the phenomenological distinction offered between the two terms. Experience will be viewed as the more passive, indirect and receiving oriented state while action as the more active, direct and achieving oriented. It needs to be further mentioned that the interaction context of the 'fictional' (world/environment) brings in the aspect of *transportation* [50] to the bi-dimensional position. The user is relocated, as it were, in a different reality (fictional world) with the liminal identity of a character[1]. These positionalities of agency and identity, however, may not be clearly identifiable as such and serve more as conceptual abstractions in the understanding of the narrative process of creation of a subjective experience.

Whether it is centralised or distributed, a system would need to incorporate some manner of affinity toward narrative order. This affinity will be driven by an evaluation of narrative state in terms of its narrative value or *narrativity*. Sustaining narrativity through the structuration process despite user participation is one of the core problems in Narrative Intelligence. This paper will look at the temporality of user experience and view it in the context of *narrativity*. The central thesis of this paper is that the affective aspect of the user experience is not just a 'nice to have' by-product, but a critical factor in the creation and sustenance of *narrativity* in emergent sequences in the story world. In support of this thesis, I will outline some of the core aspects of *narrativity* in conjunction with an analysis of user experience and action.

2. NARRATIVITY

The term *narrativity* was used by structuralists to indicate a group of 'properties characteristic of narrative that distinguishes it from the non-narrative'. Narrativity is the orientation of a narrative that makes it narrative. It is the *narrative-ness* of a narrative [26], [45], [58]. Narrativity

[1] The *in character* performative position, is explained by Schechner in terms of identity as a transition to a liminal state 'in the field between a negative and a double negative… Olivier is not Hamlet, but also he is not not Hamlet: his performance is between a denial of being another (= I am me) and a denial of not being another (= I am Hamlet)' [65]. In performance an individual takes on the identity of the character while simultaneously lending her identity to the character. It is not just identifying with the role but embodying and inhabiting it.

becomes a necessary concept when narrative is viewed as a process [5] rather than structure[2]. It is by scope and definition inclusive of audience/user.

> "…narrativity of a given narrative is not only related to the constitutive elements of the latter and to their arrangement. It must also be related to the context in which the narrative is received and more particularly, to its receiver." [58]

Narrativity 'is involved in the construction of the subject and the conditions of inter-subjective experience' [45]. The narrative process is more than just creating a logical mechanical structure of events; its purpose is to enable a personally meaningful or subjective experience. For a system to identify an emerging temporal sequence or action as narrative, it has to include an awareness of the user's experience of that sequence.

In the context of our discussion, narrativity can be viewed as a function, the purpose of which is to direct participative emergence of events into what can be identified as narrative. Furthermore, the narrativity of a narrative is a matter of degrees. In other words, we may feel a particular narrative to be more 'narrative' than another. Narrativity thus derives from the experiential valuation and conceptual definition of *narrative* itself. Identifying narrativity then depends on what one would consider the criterion of narrative-ness to be. The parameters of the narrativity function therefore depend on the factors we identify as necessary for qualifying a sequence as narrative. In the process-based view we've adopted, the core effects of narrative indicate:

1. the realisation of the core creative intention[3] of the author: the communicating of a subjective view of a (humanised) reality [32], [33], [47], [58].
2. the temporal configuration of experience, action and events in a beginning, middle and end structure [5], [8], [10], [15], [32], [47], [61]
3. the creation of a subjective or personally meaningful experience [8], [15], [45], [66]

From a computing perspective, the biggest challenge derives from the fact that the machine is still largely incapable of apprehending the complexities of human experience and action. The problem of narrative structuration is commonly dealt with by keeping the core basis for the process as a pre-determined 'narrative logic'. This logic is derived from elemental deconstruction of available story structures, and not the deconstruction of its experience. In the following sections, I will discuss the above aspects of narrativity with a view to support my thesis. This paper will present an analysis of human experience and action with a view to identify parameters of user state that could be used by a system to manage narrativity.

2.1 Narrative Reality

The *dramatic mode* assumed by the narrative embodies the functional parameters of the notion of reality that the author intends to make the user experience [49]. Klaus considers the *dramatic mode* to be a function of essential qualities in the world, and the dominant patterns of human experience [33]. It can be argued that even modern reactions to classical dramatic structure like the theatre of the absurd indicate a view on reality. Action in the narrative occurs within the laws that govern this staged reality in a way that best characterises it. In presenting reality through humanised action and situation, it is made relevant to the user as a humanised reality. Narrative effort is an exercise of selecting events and initiating action that are most appropriate (effective) in presenting this reality while leaving out all extraneous matter. The narrative structure resulting from a storyteller's craft strives to achieve a certain unity[4]. Unrehearsed and undirected (naïve) user action [49] on the other hand is assumed to result in many events that are not 'unified'. Since a machine is capable of structuring events through logical processing, computational unity is inherent in the structures that occur in the system. The challenge is achieving a computational unity that enables coherence in human action and experience of the reality construct.

The underlying assumption in this view of the problem of narrativity in interactive systems is that human action is intrinsically non-narrative in its temporal structure. While I do not agree with the view that human experience/action is neither non-deterministic [49] nor non-narrative (see the following section), I do concede that common perspectives on narrativity and narrative structure are inadequate in dealing with the problem. In my view, current computational

[2] Prince proposed his views towards the end of structuralism when it was heavily criticised by post-structuralists for 'having left out many aspects of the experience of narrative' [67]. Following the post-structuralist 'undoing' of the primacy of structural attributes defining 'narrative', the neo-narratology of Chatman and Brooks looks at narrative more as a structuring *process*. The concept of *narrativity* thus gained greater currency and the term became 'associated with the fluidity of structuration rather than structure' [45]

[3] The issue of *intention* in narrative is a hotly debated one. There are perspectives in literary criticism that argue for the work as being representative of the author's intention while others that argue that the text has its own intentionality distinct from that of the author's. For a review on the discussion of intentionality see Richard Kuhn's "Criticism and the Problem of Intention" in *Journal of Philosophy*, vol. 57. p.p. 5-23, Jan 1960. In our discussion *intention* will be viewed in relation to crafting a constructed (virtual) reality as viewed by Klaus [33] and in terms of "what the work sustains as a certain kind of experience, its focal effect" [35]

[4] The idea of unity derives from the Aristotelian concept of unity of action " the structural union of the parts being such that, if any one of them is displaced or removed, the whole will be disjointed and disturbed" VIII.1451a [6]. The concept of unity has since followed many diverse views. It has been viewed as an aid to interpretation, i.e. harmony or coherence of disparate parts and not just simple agreement or consistency. Other views indicate 'equilibrium to express richness and complexity' [53] or 'continuity' [48]. Recent views associated with post-structuralism, revisionism and deconstruction have, however, asserted that complete unity is never achieved by a work of art. R.S. Crane [12] and Walter Davis view unity in terms of purpose and structure: "purpose coincides with structure because it gives birth to it" [14]. Unity derives its occurrence from the human experience of it.

perspectives are ingrained in viewing narrative as structures imposed on the reality of human action by human imagination. Narrative is viewed as an artefact while human action is viewed as a sequence with no inherent narrative structure. Narrative effort of a system is seen as attempting a structural unity and 'efficiency' by avoiding or blocking out seemingly "irrelevant" user actions. Under the approach embodied by this paper, the first step proposed in establishing the conditions for inter-subjective experience is defining the dramatic mode of the reality constructed. The narrative reality defined in terms of the essential qualities the world (beautiful/ugly or orderly/chaotic), and the dominant patterns of human experience, (integrative/disintegrative) not only provides a framework for machine response but also for regulating audience expectation/hypothesis (future) and interpretation of past events[5]. The relationship between the experience of reality and emotions is well established:

> "Emotion is the human reality assuming itself and 'emotionally directing' itself toward the world...Emotion is not an accident, it is a mode of our conscious existence, one of the ways in which consciousness understands (in Heidegger's sense of *verstehen*) its Being-in the-World." [64]

2.2 Temporal configurations in human experience and action

> "…narrative is both metaphysical – narrative has a necessary connection to time – and a cognitive process by which the subject constructs meaningful realties." [45]

Viewing the purpose of narrative action as preparing conditions of a subjective experience of humanised reality necessitates the consideration of temporality. After all, human experience and action occur in time and cannot be viewed dissociated from temporality. The basic premise on which most approaches to solving the problem of narrative structuration in user participation based narrative systems is that user experience is *non-deterministic*. The temporal structure of narrative is seen as a human creation (artefact) that is imposed on sequences in reality. It is therefore assumed that human experience and action is inherently devoid of narrativity. In the discussion that follows, I will argue against such premise and assumptions drawing on David Carr's perceptive review of phenomenological analyses of temporality of experience and action [8]. I will begin with a discussion on the temporal nature and structure of experience to establish the fundamental concepts that will help us understand the phenomenon of action.

Experience, in phenomenology, is used to denote the more passive temporal phenomenon of sensory perception and observation. The consciousness of a present sensation is connected to a consciousness of the past (memory) and an anticipation of the future. An important consideration in understanding the flow of experience is that its relation is not as much to a fragment of the distant past (recollection) or the ideation of a future event (expectation), as to the Husserlian concepts of retention and protention. Retention is 'special sort of memory whose object is the just-past…Present and past function together in the perception of time somewhat as do foreground and background or focus and horizon in spatial perception' [8]. Retention is the "comet tail" that trails behind the present occurrence. Protention or "primary expectation" is to future what retention is to past. They should, however, not be identified as 'short-term' memory or expectation. '[W]hat distinguishes retention from recollection, and protention from "secondary expectation," is not the length of their term but their functioning as horizons from ongoing, present experience.' [8]. The structure of experience is rooted in the consciousness of the present with a gaze into the retentive past and the openness of the protential future. We continuously *reconfigure* the present import of the past based on whether our protential expectations are met or frustrated.

Carr argues that events (as opposed to sensations)6 are the basic unit of experience. Events 'have temporal thickness, beginning and end'; they stand out as identifiable meaningful set of sensations in a sequence. Events are 'experienced as phases and elements of other, larger scale events and processes. These [events] make up the temporal configurations, like melodies and other extended occurrences and happenings… we experience them as *configurations* thanks to our protentional and retentional "gaze" which spans future and past.' [8]. This gaze is what helps us make sense of our current experience in the context of a larger whole. Merleau-Ponty stresses the role of the 'lived body' in the temporal continuity of experience. Experience is lived through from the 'vantage point' of the embodied self [8] [57].

Action is also subject to this protentional and retentional "gaze", although with a different emphasis. It is phenomenologically different from passive experience in that the future expected is brought about by the action one is engaged in. Action embodies an intended result. 'In action the content of my protention is not a state of the world that I *expect*, it is something I *effect*… Since in acting we protend or intend the future goal, rather than just picturing it, there is a sense in which it occupies the center of our concern in action and reflects back upon and determines the present and the past. There is indeed something quasi-retrospective about action, as if we were located *at* the end and from its point of view arranged and organised present' [8]. The flow of consciousness in human action is future focussed, and it is not just attention but intention that is focussed there. The success or failure of an action in meeting an intended outcome makes us reconfigure the present import the past and future intentions. Actions have a beginning, middle and an end and more often than not are part of larger intended configurations. In the flow of life, a unit of these configurations is an action that stands out from the sequence as a meaningful part of a configuration. In the case of action, the future is more vulnerable and fragile

[5] For a detailed discussion on dramatic mode see "Story, Plot and Character Action: Narrative Experience as an Emotional Braid" [49]

[6] Sensations are considered to be the basic unit of experience in abstract analysis. Carr argues that sensations are by themselves 'meaningless' and 'far from being elements of experience'. Sensations are 'theoretical entities or constructs'. [8]

since it depends on the success of the action. To the human agent, however, the future is 'more determined, less open to variation than the passively protended future.' [8]. Furthermore, it can be said that since the action is done by me towards an outcome intended by me, action is more strongly personal in its relevance and effects. Action exercises a more 'retroactive control' on the present since it not only affects how we do things but also how we see things.

Narrative in conventional media is primarily for experience. While experiencing narrative in the present we have a residual memory (retention) of what has happened before in the story while having an 'openness' to a set possibilities in the future. In action, however, the future is not as 'open' to variation. Narrative structuration, therefore, needs to be approached differently to accommodate the way in which narrative configuration occurs in the mind of the user.

It is also necessary to point out that event and action configurations must not be viewed as 'mere sequences', they do not combine in a 'merely additive way', they have an inherent role in larger narrative configurations. Events combine to make up larger-scale events of which they become structural elements. Actions having their own means-end structure become means towards the performance of other larger-scale actions. That is not to say that some occurrences of experiences, events or actions do not belong to larger contexts or have 'no "point" beyond themselves.' Such cases 'seem to stand out by their very intrusiveness and prove to be exceptions' rather than the rule. From a subjective point-of-view (as opposed to the observational) the flow of life and reality constitutes various such large or small configurations and 'it is our tendency to expect such larger contexts that the isolated and intrusive stands out by contrast. Complex events, experiences and actions thus "shape" the sequences of sub-actions and other components that make them up and provide them, at this level too, with the closure constituted by their beginnings, middles and ends'. [8]

The above analysis provides a few design pointers towards addressing the problem:

1. It indicates that the retentional and protential 'horizon' influences what a user is doing and the way she sees things. What the user retains is not short-term memory, but elements of the past and the future that are most relevant to the present.

2. The 'foreground' of the user's consciousness is the present and the most relevant are the retentional past and protended future. The main concern of the user in the flow of action and experience is not the overall/larger structure that stretches farther into the past and the future, but what is happening now. The purpose of structuration, then, is not to achieve an ideal interrelationship or trajectory of events, but to enable a sequence of meaningful presents.

3. The larger or extended configurations are continuously modified in her mind based on the current state. In experience, the structure is discovered by the user through reconfiguration. In action, it is created through reconfiguration. Narrative value (narrativity) in action is perhaps not about achieving a structure similar to narratives in conventional media, but in achieving a pattern of meaningful action units that combine into a meaningful whole.

4. '[T]he temporal span is structured or configured into *events*, in the one case, and *actions*, in the other…The same retentional-protentional grasp which reaches forward and back in time also effects or constitutes a *closure* which articulates time by separating the given temporal configuration from what goes before and after' [8]. Users identify events and actions not based on absolute time, but based on closures or conclusions of a beginning, middle and end structure. It is therefore inappropriate for a system to identify elements or units of a narrative structure based on concepts like 'scenes', 'story beats' or 'moments' [25], [41], [42], [59].

5. There is a narrative structure inherent in human experience and action. The temporal dynamics of this structure, however, is slightly different from the ones obvious in well crafted stories in the way that there is no 'compression' of actual sequences by selecting the most 'interesting' elements. From the design point-of-view, it is necessary to recognise and accommodate this fact. Narrative structuration therefore should not seek to achieve the filtration of action and experience to achieve a unity. Rather it should seek to identify these configurations and allow for subjective differences.

Experience and action are not conceptualised but lived through as parts of a larger whole. In that sense, they have a strong parallel and connection with emotions. Emotions are also lived through and they seem to have similar temporal dynamics and foreground-background configuration [11], [23], [28], [29], [44]. Emotion is a dimension of experience. In fact, it is the first filter through which sensory stimulus passes before it is cognitively processed in the brain [13]. Human emotion is viewed as a thought-action tendency [24].

The above two sections have discussed the first two aspects of narrativity. The reality construct acts as a framework which controls the range of possible action outcomes. The ultimate effect of narrative is the creation of an experience or action that is personally meaningful.

2.3 Subjectivity

> "…subjectivity is engaged in the cogs of narrative and indeed constituted in the relation of narrative meaning, and desire; so that the very work of narrativity is the engagement of the subject in certain positionalities of meaning and desire." [15]

Subjectivity is the experiential effect of narrativity. The term denotes an experiential state. It is a state of being engaged in active creation of meaning and desiring. It is 'our conscious sense of self, our emotions and desires' and it is 'always embodied' [7].

> "The process of assuming subjectivity invests the individual with a temporary sense of control and sovereignty which evokes a 'metaphysics of presence' [Derrida 73] in which s/he becomes the source of meaning."[54].

> "Subjectivity can only be 'had', that is to say, experienced and performed (through the performance one has the experience of subjectivity), in the admission and recognition of one's failure to appear to oneself and within the representational field." [55]

Narrative experience of subjectivity is the meaningfulness of the action/event sequences derived in relation to a notion of self (*identity*), even if it is a constructed notion that is *performed*[7]. Revisionist accounts of narrative following structuralism suggest that narrativity is involved in the construction of the subject and the conditions of subjective experience. Singer proposes that, at every moment, narrative both 'determines a position for the subject to inhabit' and 'submits to the contingencies of determination' [66].

It is not just an experience based in the notion of self, but also an experience in the notion of a self-placed within an inter-subjective context[8]. This is what lends the meaningfulness to a sequence of experiences and qualifies it as narrative. It has been convincingly demonstrated that meaning, understanding and rationality (usually understood as cognitive processes) arise from and are conditioned by the nature and pattern of our bodily experience, including our emotional relationship to the world. [31], [36], [62], [71]

The above discussion on aspects of narrativity has indicated a strong relevance of user emotions in dealing with the problem of managing narrativity in systems. It is amply clear that storytellers guide the audience experience in a multitude of ways, and emotional 'manipulation' is a strong tool at hand to bottle neck their desires, expectations and the construction of meaning [49]. The occurrence of the type of emotions, the context in which they occur, and their temporal dynamics, can indicate narrativity during the experience [50]. In the following section, I will review the role of emotions in narrative structuration as viewed by theorists in film and literary studies. We will also see how emotions can work and influence the flow of action and experience towards narrative configuration.

[7] For a detailed discussion on the performative nature of 'being in a fictional world' see *Narrativity of User Experience: Presence as Transportation in IVE Based Narrative Systems*, [50]

[8] The inter-subjective context is what would lend narrativity to, for example, 'Could you pass the salt?' It would state, depending on the inter-subjective context, not merely a request but also that 'the food is tasteless' or 'I'm about to conduct an experiment'. Under the appropriate conditions, it can act as 'flirtation, rebuke, instruction (examples could be multiplied)' [45] and position the individual in a number of different storylines.

3. EMOTIONS, MOODS AND NARRATIVITY

> "…the *style* of emotional behaviour, the *context* within which it occurs, and the manner in which it unfolds through *time*, are all crucial to its meaning and accountability." [46]

Cognitivist research in the fields of literary and film studies have explained and modelled the role of emotion in narrative structure relevant to their respective media. It should however be noted that most of these studies have approached emotion in what can be called the 'prototypical view' of emotion [68]. Emotion in cognitivism is thought of as being an action tendency (fear indicates tendencies of fight or flight), an orientation towards an object (fear is 'of' something 'believed' to be threatening), and goal directed (fear has a goal of self-preservation)[9]. Bordwell and Branigan's conceptualisations of narrative indicate a fairly simplistic connection between film comprehension and film emotion based in this action, object and goal framework [2], [4]. Cognitivists tend to link emotion with the character's actions, motivation and goals. Emotion is viewed as a result of identification [9], [27], [34], [52], [70] or empathy [16], [72], [73] with the protagonist.
The following issues can be isolated while considering the above approaches to understanding the dynamics of emotion in the narrative process:

1. The nature of emotion considered by the above approaches is a 'witness emotion' and 'witnesses cannot participate in events, nor can they command their movements and views. A film's narration dictates what the viewers see, how they see it and when. Emotion in the film viewer is a response to this predicament' [69]. Emotion in studies of narrative is approached from an indirect experience perspective.
2. The approaches seem to assume a fairly simplistic prototype of emotion biased toward and derived from research focussed primarily on negative emotions [24]. The models of emotion used in understanding their role in narrativity are incomplete.
3. They have focussed on the analysis of character and plot related situations of emotional intensity, and have excluded the effect of cues of low tonality 'that do not advance or retard the narrative progress' [68]. The understanding of emotion dynamics in narrative is restricted since they are viewed in relation to structural elements and not from the perspective of audience experience or action.

It is amply clear that available models of emotion in narrative are inadequate in addressing the problem of narrativity in participative action. Frederickson brings to light the fact that general models of emotions, being based on a large body of research concerning negative emotions, are inadequate in explaining positive emotions [19], [20], [21], [38] [24]. The general models of emotion tend to two presumptions that become questionable when positive emotions are considered - that emotions yield *specific* and *physical* action oriented tendencies. Frederickson points out that certain positive

[9] For a discussion on Emotion in Cognitivism and Arousal Theory see *Art and Emotion* [43]

emotions spark changes primarily in cognitive activity. She proposes instead to view the effect of emotion in terms of *thought-action tendencies.* Negative emotions are known to "*narrow* a person's momentary thought action repertoire. They do so by calling to mind and body the time-tested, ancestrally adaptive actions represented by action tendencies…because positive emotions are not linked to threats requiring quick action [they] *broaden* a person's momentary thought-action repertoire." [24]. Negative emotions are linked to survival instincts, and indicate a higher determinacy of human action. The determinacy is directly related to the intensity of emotion. Positive emotions "go beyond making people "feel good" or improving their subjective experiences of life. They also have the potential to broaden people's habitual modes of thinking and build their physical, intellectual, and social resources…By broadening the thought-action repertoire, positive emotions may loosen the hold of that (no longer relevant) negative emotions gain on an individual's mind and body by dismantling or *undoing* the narrowed psychological and physiological preparation for specific action… positive emotions create physiological support for pursuing the wider array of thoughts and actions called forth" [24].. Emotions also have a profound effect on storage and access of short-term, long-term memory [3], [13], [37], [39], [40]. Intensity of emotion is related to whether a person accesses general or specific Autobiographical Memory [56].

By influencing memory and thought-action tendencies emotions dictate the composition of the retention-protention horizon. Emotion, thus, becomes a powerful tool for both limiting and expanding the possibilities of user action and narrative configurations in emergent systems. Identification of emotional states of the user can also help a machine predict the range of user action-lines possible.

Emotion is considered both a motivating and guiding force in perception and attention. Emotion and mood are known to lend biases to interpretation of situations [29], [11], [23], [28]. Emotions differ from moods in that they are about some personally meaningful circumstance and are typically short lived and occupy the foreground of consciousness. They are more situated in the present consciousness. Moods are typically free floating or objectless, more lasting, and occupy the background of consciousness [51], [63], [24]. Moods are considered to be low tonic levels of arousal within emotional systems that can be induced or changed through a successive stimulation of similar (negative or positive) emotions and feelings. Moods also indicate the 'initial condition' of the dynamic system [44].

By virtue of being embodied, emotion and mood lend subjectivity to an experience and make it undeniably real. They lend meaningfulness to an event or action and influence the process of configuration. Emotion could thus play a critical role in the occurrence and maintenance of narrativity in emergent sequences.

4. CONCLUSIONS

In the above discussion, we have seen a user-centred approach towards solving the narrative structuration problem. Common approaches to defining the problem are based in fundamental misconceptions about the nature of human experience and action. Even though researchers and designers are opening up to a process based view of narrative [1], narrativity is still largely associated with structures of narrative created for experience. Furthermore, user emotion is viewed as a desirable 'by-product' and has not been the subject of much attention. This paper has argued against this paradigm in proposing that narrativity in action can perhaps not be achieved without the manipulation of user emotion.

Indeed, the true value of this proposition would be realised with the presentation of a methodology demonstrating its mechanics in the narrative structuration. While such a presentation is beyond the scope of this paper, it would be worthwhile to conclude the discussion with some indication of its mechanics. Greg Smith's explanation of how emotions work in the narrative process is insightful in demonstrating their role in narrative structuration. He proposes that filmic narratives are composed of emotion cues:

> "Films use emotion cues to prompt us toward mood, a predisposition toward experiencing emotion. Moods are reinforced by coordinated bursts of emotion cues [emotion markers], providing payoffs for the viewer. These payoffs may occur during narratively significant moments (like obstacles) or they may occur in instances that do not advance or retard the plot progress. Cues are the smallest unit [of] a text's emotional appeal… Emotion cues are the building blocks used to create larger narrational structures to appeal to emotions. Mood is sustained by a succession of cues, some of which are organised into larger structures, some of which are not" [68]

The temporal layout of emotion cues is more indicative of 'pattern' [49] rather than 'structure'. Patterns are more congruent with the fluidity of emotion and structuration. The narrative structuration process can be viewed as the unfolding of a pattern of various positive and negative emotional cues and markers. A narrative's temporal placement of emotion cues directed at the user serves the following purposes:

1. It helps establish (in the beginning) and maintain (throughout) the overall mood of the experience.
2. It helps prepare the initial conditions for narratively significant coordinated bursts of emotion.
3. In maintaining a particular mood ('mood congruence'), it lends a predisposition to the construction of the implied reality and inference of the laws governing the narrative world.
4. Emotions and the underlying mood affect the creation of the retention-protention horizon and consequently (re)configurations
5. An orchestration of positive and negative emotion cues thus regulates thought-action, desire and ultimately inter-subjectivity in narrative.

5. REFERENCES

[1] **Aylett, R. and S. Louchart**. "Towards a Narrative Theory of Virtual reality" in *Virtual Reality*, vol.7, p.p. 2-9. 2003.

[2] **Bordwell, David.** "A Case for Cognitivism" in *Iris*, vol. 9, p.p. 11-40. 1989

[3] **Bower, G.H., P.R. Cohen.** " Emotional influences in memory and thinking: Data and theory" in *Affect and*

Cognition: The Seventh Annual Carnegie Symposium on Cognition,Lawrence Erlbaum Associates, 1982

[4] **Branigan, Edward.** Narrative Comprehension and Film. Routledge, NY. 1992

[5] **Brooks, P**. Reading for the Plot: Design and Intention in Narrative. Knopf Press, 1984

[6] **Butcher, S.H.** *Aristotle's Theory of Poetry and Fine Art,* Kalyani, New Delhi. 1985

[7] **Butler, Judith**. *Bodies That Matter*. Routledge, London NY. 1993

[8] **Carr, David.** *Time, Narrative and History.* Indiana University Press, Bloomington and Indianapolis, 1986

[9] **Carroll, Noël.** *The Philosophy of Horror, or Paradoxes of the Heart.* Routledge, NY. 1990.

[10] **Chatman, Seymour**. "Reply to Barbara Herrnstein Smith" in W.J.T Mitchell ed. *On Narrative,* p.p. 260, Chicago University Press, 1981

[11] **Clore, G.L.** "Cognitive Phenomenology: Feelings and the construction of judgement" in L. Martin and A. Tesser ed. *The Construction of Social Judgements,* Lawrence Erlbaum Associates,1992

[12] **Crane, R.S**. Language of Criticism and the Structure of Poetry. Toronto University Press, Toronto, 1953

[13] **Damasio, Antonio R**. Descartes' Error: Emotion Reason and the Human Brain. Putnam, NY. 1994.

[14] **Davis, Walter**. *The Act of Interpretation*. University of Chicago Press, Chicago, 1978

[15] **De Lauretis, Teresa**. *Alice Doesn't: Feminism, Semiotics, Cinema.* Macmillan, London, 1984

[16] **De Wied, Minet, Dolf Zillman, Virginia Ordman. "**The Role of Empathic Distress in the Enjoyment Of Cinematic Tragedy" in *Poetics,* p.p. 91-106, vol. 23, 1994

[17] **Derrida, J**. Speech and Phenomena trans. D. Allison. Northwestern University Press, Evanston IL. 1973

[18] **Ditton, Th.** The Unintentional Blending of Direct Experience and Mediated Experience: The Role of Enhanced versus Limited Television Presentations in Inducing Source Monitoring Errors, unpublished doctoral thesis, Temple University, 1997

[19] **Ekman, P**. "An Argument for Basic Emotions" in *Cognition and Emotion*, vol.6, p.p. 169-200. 1992 b

[20] **Ekman, P.** "Are There Basic Emotions?" in Psychological Review, vol. 99, p.p. 550-553. 1992 a

[21] **Ekman, P**. and E. Rosenberg. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System. Oxford University Press, Oxford. 1998.

[22] **Fazio, R.H. and M.P. Zanna**. "Direct Experience in Attitude-Behavior Consistency" in M.P. Zanna ed. *Advances in Experimental Social Psychology*, vol.4, p.p. 161-202, San Diego Academic Press, 1981

[23] **Forgas, J.P. and S. Moylan.** "After the Movies: Transient Mood and Social Judgements" in *Personality and Social Psychology Bulletin,* vol. 13 no. 4, 1987

[24] **Frederickson, Barbara L. and Christine Branigan.** "Positive Emotions" in Tracy J. Mayne and George M. Bonanno eds. *Emotions: Current Issues and Future Directions.* p.p. 123-151, The Guilford Press, NY London. 2001.

[25] **Gordon, Andrew S. and Nicholas Iuppa. "**Experience Management Using Storyline Adaptation Strategies" in Göbel, Stefan, Norbet Braun, Ulrike Spierling, Johanna Dechau, Holger Diner eds. *Proceedings of the Technologies for Interactive Digital Storytelling and Entertainment (TIDSE) Conference,* ZGDV, Darmstadt, 2003.

[26] **Greimas, A.J.** Sign, Language, Culture, The Hague: Mouton, 1970

[27] **Grodal, Torben**. Moving Pictures: A New Theory of Film Genres, Feelings, and Cognition. Oxford University Press, 1997

[28] **Isen, A.M., K.A. Daubman, and G.P. Nowicki**. "Positive affect facilitates creative problem solving" in *Journal of Personality and Social Psychology,* vol. 52 no. 6, 1987

[29] **Izard, C.E.** "Four Systems for Emotion Activation: Cognitive and Non-Cognitive Processes" in *Psychological Review,* vol. 100 no. 1, 1993

[30] **Jeandrain, Anne-Cécile.** Essay About Telepresence Effects on Persuasion: Three Possible Explanations, presented at Presence 2001-The Fourth Annual International Workshop on Presence, 2001

[31] **Johnson, M**. The Body in Mind: The Bodily Basis of Meaning, Imagination and Reason. University of Chicago Press, Chicago. 1987

[32] **Kermode, Frank**. *The Sense of an Ending,* Oxford University Press, New York. 1968

[33] **Klaus, Carl H**. "Drama" in Robert Scholes, Nancy R Comley, Carl H Klaus, and Michael Silverman ed. *Elements of Literature*. p.p. 773-800, Oxford University Press, 1991

[34] **Kneepkens, E.W.E.M. and Rolf Zwaan.** "Emotions in Literary Text Comprehension" in *Poetics,* p.p. 125-138, vol. 23, 1994

[35] **Kuhn, Richard**. "Criticism and the Problem of Intention" in *Journal of Philosophy*, vol. 57, p.p. 5-23, Jan 1960

[36] **Lakoff, G**. *Women, Fire and Dangerous Things.* University of Chicago Press, Chicago. 1987

[37] **Lang, P.J.** "A bio-informational theory of emotional imagery" in *Psychophysiology,* vol. 16, 1979

[38] **Lazarus, R.S.** *Emotion and Adaptation.* Oxford University Press, NY. 1991.

[39] **LeDoux, J.E.** *The Emotional Brain: The mysterious underpinnings of emotional life.* Simon & Schuster, New York1996

[40] **Leventhal, H.** "A perceptual motor theory of emotion" in P. Ekman and K. Scherer eds. *Approaches to emotion.* p.p. 271-291, Erlbaum, Hillsdale, New Jersey, 1984

[41] **Mateas, Michael and Andrew Stern**. "Towards Integrating Plot and Character for Interactive Drama" in *Working*

notes of the Social Intelligent Agents: The Human in Loop Symposium. AAAI Press, Menlo Park, CA, 2000

[42] **Mateas, Michael and Andrew Stern**. *Architectural, Authorial Idioms and Early Observations of the Interactive Drama Façade*, CMU-CS-01-198, 2002

[43] **Matravers, Derek.** *Art and Emotion*. Oxford University Press, 2001

[44] **Mayne, Tracy J. and James Ramsey.** "The Structure of Emotion: A Nonlinear Dynamic Systems Approach" in Tracy J. Mayne and George M. Bonanno eds. *Emotions: Current Issues and Future Directions.* p.p. 1-37, The Guilford Press, NY London. 2001.

[45] **Mcquillan, Martin**. ed. *The Narrative Reader.* Routledge, London, 2000

[46] **Merleau-Ponty, M.** *Phenomenology of Perception*, Routledge, London. 1992

[47] **Mink, Louis O.** "Narrative Form as Cognitive Instrument" in R.H. Canary and H. Kozicki eds. *The Writing of History*, University of Wisconsin Press, Madison. 1978

[48] **Nassar, Eugene Paul**. *The Rape of Cinderella*. Indiana University Press, Bloomington, 1970

[49] **Nath, S.** "Story, Plot and Character Action: Narrative Experience as an Emotional Braid" in S. Göbel, N. Braun, U. Spierling, J. Dechau and Holger Diener eds. *Proceedings of the Technologies for Interactive Digital Storytelling and Entertainment (TIDSE) Conference 2003*, p.p. 1-18. ZGDV. 2003.

[50] **Nath, Shachindra.** "Narrativity of User Experience: Presence as Transportation in IVE Based Narrative Systems" in S. Göbel, U. Spierling, A. Hoffmann, I. Iurgel, O. Schneider, J. Dechau and A. Feix eds. *Technologies for Interactive Digital Storytelling and Entertainment (TIDSE) Conference Proceedings 2004,* p.p. 55-60. Springer. 2004.

[51] **Oatley, K. and J.M. Jenkins.** *Understanding Emotions.* Blackwell, Cambridge, MS. 1996.

[52] **Oatley, Keith.** "A Taxonomy of the Emotions of Literary Response and a Theory of Identification in Fictional Narrative" in *Poetics*, vol. 23, p.p. 53-74. 1994.

[53] **Ogden, C.K., I.A. Richards and James Woods**. *The Foundation of Aesthetics*. George Allen and Unwin, London, 1922

[54] **Pearson, Roberta. E. and Philip Simpson** eds. *Critical Dictionary of Film and Television Theory.* Routledge, London NY. 2001.

[55] **Phelan, P**. *Unmarked. The Politics of Performance,* Routledge, London. 1993

[56] **Philippot, Pierre and Alexandre Schaefer.** "Emotion and Memory" in Tracy J. Mayne and George M. Bonanno eds. *Emotions: Current Issues and Future Directions.* p.p. 82-122, The Guilford Press, NY London. 2001.

[57] **Priest, Stephen**. *Merleau-Ponty.* Routledge, London, New York, 2003

[58] **Prince, Gerald**. *Narratology: The Form and Functioning of Narrative,* Mouton, Berlin, 1982

[59] **Rawlings, Jeff and Joe Andrieu**. "Player-Protagonist Motivation in First-Person Interactive Drama" in Göbel, Stefan, Norbet Braun, Ulrike Spierling, Johanna Dechau, Holger Diner eds. *Proceedings of the Technologies for Interactive Digital Storytelling and Entertainment (TIDSE) Conference*, ZGDV, Darmstadt, 2003.

[60] **Reeves, B. and C. Nass**. The Media Equation – How People Treat Computers, Television, and New Media Like Real People and Places, Cambridge University Press, Cambridge, USA, 1996

[61] **Ricoeur, Paul.** Time and Narrative Volume 3 trans. Kathleen Blamey and David Pellaur. The University of Chicago Press, London, Chicago, 1990.

[62] **Rose, H**. *Love, Power and Knowledge: Towards a Feminist Transformation of the Sciences*. Polity Press, Cambridge. 1994.

[63] **Rosenberg, E.L.** "Levels of Analysis and the Organization of Affect" in *Review of General Psychology,* vol.2, p.p. 247-270. 1998.

[64] **Sartre, Jean-Paul**. *Sketch for a Theory of the Emotions* trans. P. Mairet. Methuen, London. 1971

[65] **Schechner, Richard**. *Performative Circumstances from the Avant Garde to Ramlila*. Seagull Books, Calcutta, 1983

[66] **Singer, Alan**. "The methods of Form: On Narrativity and Social Consciousness" in *Substance*, p.p. 64-77, vol.41, 1983

[67] **Skinner, John**. *Tell-Tale Theories Essays in Narrative Poetics,* Turun Yliopisto, Turku. 1989

[68] **Smith, Greg M**. "Local Emotion, Global Moods, and Film Structure" in Carl Plantinga and Greg M. Smith eds. *Passionate Views: Film, Cognition, and Emotion*. p.p. 103-126. The Johns Hopkins University Press, Baltimore, London. 1999.

[69] **Tan, Ed S.** "Film Induced Affect as a Witness Emotion" in *Poetics*, vol. 23, p.p. 7-32. 1994.

[70] **Tan, Ed S.** Emotion and the Structure of Narrative Film: Film as Emotion Machine trans. Barbara Fasting. Erlbaum, Mahwah, New Jersey. 1996

[71] **Williams, Simon J.** Emotion and Social Theory: Corporeal Reflections on the (Ir)Rational. Sage, London, CA, New Delhi. 2001

[72] **Zillman, Dolf and Silvia Knobloch.** "Emotional Reactions to Narratives About the Fortunes of Personae in the News Theatre" in *Poetics,* p.p. 189-206, vol. 29, 2001

[73] **Zillman, Dolf.** "Mechanisms of Emotional Involvement with Drama" in *Poetics,* p.p. 33-51, vol. 23, 1994

if mouseSign then: Semiosis, Cybernetics and the Aesthetics of the Interactive

Alan Peacock
University of Hertfordshire
Faculty of Art and Design
UK
e-mail: a.d.peacock@herts.ac.uk

Keywords
Interactivity, hermeneutics, aesthetics, semiotics, sign systems, syntagm, paradigm, cybernetic loop, mouse, interactive art, syntagmurgy

When delivered as a live performance this text is punctuated with the display of interactive media, sometimes with commentary, sometimes without. In this text appropriately illustrative pieces are named. There is no way to word up, describe or account for what is done using the interactive media artefacts. The best I can hope for is that you find those works for yourselves and having experienced them reflect on the text of the paper. Or if you can not find them, then experience some others and reflect on the text. If the words here have any relevance then they should apply to all instances, and not only to my carefully selected materials.

Brunswick Parking Lot [7]

AN OBSERVATION
Thomas Kuhn, talking about the history of 'normal science' makes the point that much time and energy is spent in having to explain things again and again because they are not shared and can not be taken for granted.

Being able to take no body of belief for granted, each writer on physical optics felt forced to build his field anew from its foundations. In doing so his choice of supporting observation and experiment was relatively free, for there was no standard set of methods or of phenomena that every optical writer felt forced to employ and explain. ….. That pattern is not unfamiliar in a number of creative fields today, nor is it incompatible with significant discovery and invention.

Kuhn, [1]

These comments characterise the activities of scientists in the period preceding the establishment of a 'normal science' either before the body of knowledge in a field is established, or at times when the existing body of knowledge is found to be inadequate. 'Normal science' for Kuhn comes at a time when scientists in a given field share a paradigm of ideas about what it is appropriate to measure through experimentation. With that paradigm in place the field develops rapidly, partly because less energy is used up in reworking definitions or advancing and countering competing views of what makes up the field in the first place.

With Kuhn's perspective in mind, at the start of this discussion of semiosis, cybernetics, and the aesthetics of the interactive it is as well to make clear certain terms and ideas.

Firstly, and most importantly, the way in which the word 'interactive' is used. This is because there is a general sense to the word 'interactive', and a specialised sense. In the general sense everything is interactive because we enter into a discourse of signs and cognition with it, a discourse which has been seen as 'hermeneutic' and which is essentially a sign process of syntagmatic-selection (syntagmurgy – see note 1). In the specialised sense, when it is used of particular instances of things which are usually (but not always nor necessarily) mediated through digital technologies, fewer things are *interactive* because the discourse of signs and cognition that goes with those things is unlike the discourse that goes with other things. That discourse includes and involves a dualled-syntagmurgy derived essentially from the cybernetic loop of interactivity and the quandary of physicality that typifies such things.

The number of things that are described by this specialised meaning is, of course, increasing, and the *interactive*, in this sense, has become an important location of cultural activities – economic activities, artistic activities, communication acts, entertainments, functions and distractions of many kinds are now located within the things that are described by this meaning. Indeed, the experience of the *interactive,* in one form or another, is an increasingly common day-to-day activity for many people in the technonomies of the world, and a preferred form of leisure time engagement. So much so that the sign systems (the languages) of the interactive are among the most commonly used languages.

Reminding ourselves of this difference in meaning, and rehearsing that difference, will establish an essential characteristic of the interactive which plays upon all subsequent discussion of it, and will return us to the main point of the discussion which is the way in which the language of Mouse operates and how it contributes to the aesthetics of the interactive, the defining qualities of the experience of the discourse that is the *interactive*, in this specialised sense.

Here, the aesthetics of the interactive is about the experience of experiencing the interactive. It is not about judging the interactive in any terms of beauty derived from arbitrary

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

histories or beliefs about how things should be. The aesthetics of the interactive is about enabling an effective discussion of the experience of the interactive in all its diversity, in all its various forms. It implies that there is a cognate thing, or set of things, that share enough characteristics to be classed together and labelled as the interactive – including such things as web based search engines, medical record databases, word-processors, text adventure games, image editing software, online tax returns, shoot'em'up games, interactive installation artworks, spreadsheets, digital camera photo-album software, soundtoys – and that there is a commonality to the way in which the can be described, discussed and analysed.

Small Fish [8]

DIFFERENTIATION: HERMENEUTICS, SEMIOTICS, SYNTAGMURGY

There is a general sense of the word interactive, a common place meaning, which is a truism – everything is interactive because we interact with it. A painting, for example, is interactive because in looking at it we are in some form of discourse with it. This can be represented as a simple diagram in which two entities exist separately and overlap mutually.
In this diagram the area of overlap stands for the 'moment of discourse', and we would generally identify the two entities as (something like) a text and a reader. Identifying, usually, a set of signs (the text), and a receiver/processor of those signs (the reader).

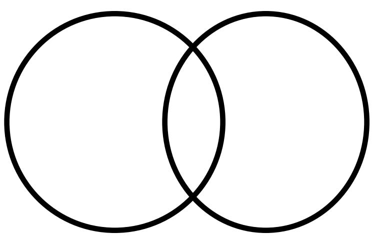


Figure 1

And it is this sense, initially, that all things are interactive because to 'know' some thing is to somehow have experienced a shared discourse with it (or, of course, more accurately with the signs that make it up, or, more accurately still, to have experienced the sensation of the perception and cognition of those signs and to have related that sign system to more or less shared cultural conditions of meaningfulness). In each moment of discourse there is a making of meanings. Those meanings are contingent on the moment, they are fickle and shift from and between each moment of discourse as the meanings play through transient syntagmatic:paradigmatic states. Our reading at any one moment of discourse is dependent on that moment and quite possibly differs markedly from other readings and the readings of others, which are themselves the experience of other moments of discourse.

"*By tradition, the relationship between audience and artwork has been analysed as an active one characterized by "interpretation". According to traditional notions, interpretation changes the artwork (e.g., the reader "acquires" the text, thus changing its meaning into something different). ….. the relationship between the audience and the artwork is "hermeneutic"… this particular form of observational relationship between an audience and the sign object in which the interpretative observation changes the meaning of the sign object. From here stems the notion of a circular process constituted by artwork and observer, the so-called hermeneutic circle.*"

Qvortup [2]

Our general understanding of the continuity of the physical world requires that the text, a painting, say, or a sculpture, is much the same between our moments of discourse, and is, itself unaffected by those moments of discourse – the text has a fixed materiality (which is the paradigm of signs that constitutes all the sign needed to 'read' the text;object), remaining consistently the same for different viewers and for a viewer's own, individual moments of discourse. Unless the text is physically changed in some way, damaged, vandalised, reworked – in which case, as it is different and no longer the (original) text so the discourse is necessarily different – it is understood to have a consistency and fixity across and between viewings. We have no great troubles with this in our quotidian existences. Indeed, the very basis of social worlds can be understood as requiring a clear and certain acceptance of the continuity of the physical world, of the fixity of the materiality of things.
This experience of (or belief in) the material fixity of an object (the text) seems initially to be at odds with the idea of the hermeneutic experience. The fixity of the object contradicts the supposedly 'subjective' nature of the interpretation, and one or the other seems counter-intuitive.
The material fixity of things is bound in, at an existential level, with the signs that make them up – the paradigm of the text. There can be no signs that are not material, and there can be no signs of the text:object beyond its materiality where materiality is bounded by and defined as the presence of the object:sign in the (human) sensorium – I have discussed this previously and elsewhere, and will return to it here, later. Viewed from the perspective of semiotics the 'hermeneutic' process can be described as the making of a selection from the broader paradigm of signs that constitutes the text to form a selection, or syntagm of a particular 'reading', 'interpretaion' or, viewed the other way round, 'utterance'.

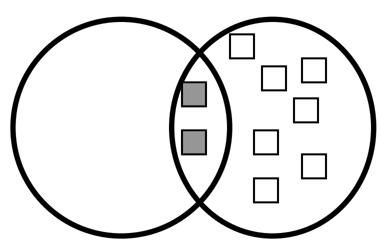


Figure2

In this diagram the 'text' (to the right, my humanist acculturation through European print conventions requires me to place the most important thing to the left, here the 'reader' (individual, human, being, living) is more important that the text (inanimate, reproducible)) is made up of a paradigm of signs – rectangular blocks - and the hermeneutic process includes only a selection (syntagm) of those signs for its interpretation (the greyed blocks in the area of overlap).

This selection, this forming of the syntagmatic sub-set from the paradigm which is all the possible signs (and materiality) of the text will feature in this discussion severally, I will use the term syntagmurgy for it (see note 1). In hermeneutics this syntagmurgic process is one of reception and interpretation, the reading of the (current and continuously changing) syntagm of signs from the text:object.

It is clear that we experience things severally and differently, they differ from moment of discourse to moment of discourse for ourselves, and the interpretation varies. This can be accounted for in one way by changes in ourselves between viewings – psychological changes, life events, present concerns, the transient effects of diet, intoxication, fatigue, the mood of our companions, whatever…… In another way this can be accounted for by different bodies of prior knowledge, varying skills of viewing between different vieings and, between individuals…… Concentration, modified and manipulated by all those thing that affect us, is syntagmurgy, and even though the text is fixed, the discourse is dynamic as the reader continuously shifts and changes within one moment of discourse and between others. In this the reader constitutes a specially fluid text and the signs included in the syntagm of perception (and so, what the text is understood as) vary during and between moments of discourse.

The conditions of the reader create a dynamic process of syntagmurgy. The 'internal' signs of the reader (their knowledge, associations, concerns, interests, etc) all influence and modify the syntagmurgy, and the syntagmurgy modifies the reader by becoming part of the 'internal' signs.

The important point, here, is that the location of syntagmurgy is within the reader. The object:text exists materially, and as a complete paradigm of all the possible signs that can be read as the text:object but not as all the signs that can make up the discourse for the reader, too, brings a body of signs to the discourse. And it is in this sense that we 'interact' with everything, and that interaction is 'hermeneutic'.

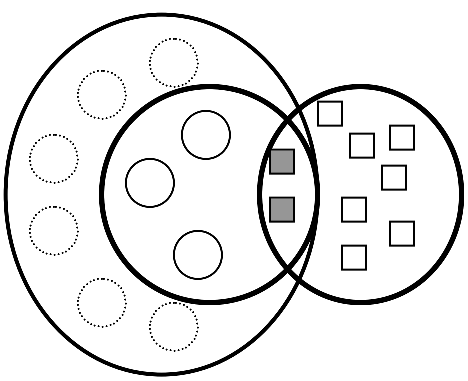


Figure 3

Represented in this diagram by the leftmost ellipse and the circles it contains, several factors influence the syntagmurgy of the reader. The 'hermeneutic' interpretation of the text is dependent on how these factors inform the act of reading the text, which signs from the paradigm that makes up the text are noticed and in which order. These are such things as preoccupations, obsessions, associations, concentration and the mental skills of the prepared or trained mind.

The syntagmurgic process of reading a text necessarily varies from reader to reader as their conditions are particular to them. Although acculturation means that the members of a given community will have a shared general reading of a text (the process of acculturation prepares the mind for all acts of readings, puts in place prepared syntagmurgic processes), individual readings will vary, and cultural readings will vary across time. In this way the material text holds polysemy, its meanings vary even as all its signs are fixed because syntagmurgic processes render signs into readings. All readings then become aberrant readings because there is no one reading which is more definitive or more privileged than any other.

Syntagmurgy is not only a reception of signs. While the reading of text can involve an active searching for signs (an active viewing of an iamge, say, where the eye:mind searches the image, tracks between areas to resolve a discrepancy in the sign systems such as a visual pun where a sign functions indexically for one meaning and iconically for another) in another form, the process of selecting from a paradigm creates utterances. These two syntagmurgies go side-by-side, and turn-by-turn, reception leading to creation leading to reception. In some instances such as conversation, dance improvisation, jazz jamming, performance art, the two forms of syntagmurgy flow across and through a rich interpersonal understanding, fusing the syntagmurgies of reception and creation into one. And in this sense a number of artforms are interactive and often are characterised by an emergent and ephemeral paradigm of signs.

The other meaning of *interactive*, the specialised meaning, and the one which is central to this text and discussion is about the discourse with a particular class of text, ones which are, usually, mediated by digital computing technologies. The thing that marks out this specialised meaning is the particular and peculiar nature of those texts. Where a painting or sculpture, a chair, crab or trumpet are taken to have physical continuity (that is to say they embody the paradigm, and stay the same across time and between viewings and viewers) these texts have no such continuity, or no guarantee of such continuity. Much of what notionally constitutes their paradigm of signs is not material, and that which is material is not necessarily the same for and during any moment of discourse or between viewings. That is to say that much of what will constitute the signs exists as 'coded' information, held, stored, located beyond the human sensorium and although having a physical form (as pits on a CD, or the charges that make up bits in planes of doped silicon or ferro-magnetic particles) they are not material. Even if we could sense the pits on a CD the resulting sensation would not be the same as the sound that would come from the same stream when passed through CODEC hardware and software and through loudspeakers, nor the image that might be made of it. This is the quandary of materiality as it effects the interactive. The encoded data is retrieved, used to generate displays and in that it takes on materiality and meaning. In the interactive that process is, of course, contingent on the reader's actions in the moment of discourse. Literally brought into being by the discourse the *interactive* text is ephemeral and has an emergent syntagmurgy of display which becomes the

paradigm that the hermeneutic syntagmurgy entrs into discourse with . While its paradigm exists at a physical level it only takes on materiality in the moments of its display.
Where in the general sense, the dynamic qualities of the interactive come from the syntagmurgy of a reader's cognition, perception or position in relation to a fixed text, a paradigm of signs embedded in a materiality; in the specialised meaning that same syntagmurgic process of cognition, perception, etc relates to a text that does not have that fixity, it is itself a syntagmurgic instance drawn from an otherwise unknowable paradigm of signs, that only take on materiality in the instance of display. In the interactive the syntagmurgies of reception and creation merge in a way which is unlike that of other sign systems and the discourse can not be reframed by moving through or around a material text because it ain't there.

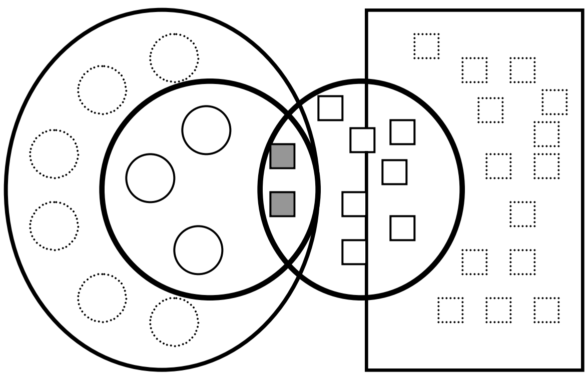


Figure 4

In this diagram the bounding rectangle to the right and the objects it contains represent the data paradigm of the text – the information that can be used to construct a display – the display itself being the material signs of the text, which are represented by the solid line boxes in the right circle. This selection of display sign from the data paradigm is syntagmurgic. It comes about through the overlay of reader's actions with the algorithmic rules of the system. The data paradigm can be thought of as the database that for Manovich is 'traversed'. In interactive artefacts the invocation of elements of the data paradigm as parts of the display syntagm comes about as a consequence of the user's actions, decisions, choices, and in this is Aarseth's idea of the 'ergodic' – the outcome of 'non-trivial' actions in traversing an underlying database.
Taken together, the hermeneutic interpretation that is acts of reading, and the contingent, 'ergodic' process whereby display is formed, constitute a dual-syntagmurgy which characterises the interactive and is unlike other media forms and experiences. It is in this dual-syntagmurgy that the *interactive* (in its specialised sense) is differentiated from the interactive in its more general sense. And with this comes particular experiences and concerns that taken together define the discourse of the *interactive*, and its aesthetics. This dually-syntagmurgic relationship is a kind of two-way hermeneutics – the user experiences a text/system and interprets it, the system experiences a text/user and interprets it. The actions of one provoke the display of the other. The displays of either are the actions of the other. And this describes the loop of communication and control that sits at the heart of all cybernetic systems.

Obsession [7]

INTERACTIVITY, SEMIOSIS, CYBERNETICS

Interactivity here is understood as a 'cybernetic loop' in which actions and displays are irrevocably linked as instances of communication and control, and through which a semiotic process runs (see figure xxx, below). That process is inextricably bound up with spatial and temporal juxtapositions, with real or apparent cause and effect, forming the changes in pattern and of energy that are the basis of semiosis, that create the differences and relationships which are pre-requisites of signification.
In this cybernetic loop user and system mutually modify the display of one and the actions of another in a dynamic dance of syntagmurgies. Cause and effect sweeps clockwise, and while the sweep may be erratic and irregular, it is characterised by the continuity of the semiotic process in which meaning is made, derived, sustained, confirmed, replicated, denied. That process runs along an axis protruding from the page. Viewed sideways on, the cybernetic loop moves, helically, bolt-thread-like along that axis. Actions prompt changes to display, and changes in display prompt actions, and in the moments of action:display signs come into being, meanings are made. Within the loop, display and action blur in the form of rollover images, cursor movements, the tap of a PDA stylus. They spin away in a melee of other feedback loops spawned by acts of display and the display of actions, by sensations and meanings.
In this cybernetic loop display is both communication and control, and action is both control and communication. Although there exists a clear separation and differentiation of states within the loop, within the user (perception, cognition and decision) and within the system (event/conditions, data retrieval/calculation, assembly) there is no clear differentiation between display and action, both are communication both are control, and both are syntagmurgic. The signs of action made by the user, the rolling of a mouse, the tap of keys, the press of buttons, are in fact the display:text viewed by the computer. The computer's actions are in fact the flickering signifiers of light, sound and vibration in and on monitors, loudspeakers, trembling feedback devices, the display:text the user reads.
Here, interactivity is understood to be inherently 'narrative' in that it is time-based and driven by events. Structures of cause and effect, the modification of meanings by juxtaposition in space or sequence, connect the instances of action:display within larger structures of meaning, and those larger structures, themselves, have semiotic function.
Semiosis is located in the moments of action and display, in the 'manners' of the devices in which juxtaposition and cause and effect take place. Signs are patterns of energy, or changes in patterns of energy which are both and at once signs of signs and signs of their meanings which are derived in part from their difference from other signs and from their juxtaposition along with other signs. So, on rollover an area of the screen changes colour. The sign is bound up in changes of energy, in changes of the frequency of light, in changes of the electrical

current which controls the state of a range of cells in the matrix of a TFT display. This area, and not other ones which appear similar, changes. This area, and the ones around it have features of colour that make them appear, quite illusorily, to be embossed, to stand above the plane of the ground on which they apparemtly rest. Shaded darker below and on the right, and shaded lighter above and on the left, the illusion of standing proud from the surface is both perceptual (iconic/indexical) and forms a conventional sign (symbol). On rollover the colours change, darker above and left, lighter below and right, as if the surface recessed (the optical illusion now confirmed by an impossible instant change of state, no transition, no restraints of material form or gavity, no resistance or momentum)……………… and so we could detail the semiotic processes of display, and much work does so. The concern here though will be less with the display (that is the text) presented to the reader than with display (that is the text) presented to the computer, that is with the devices through which the reader acts – the mouse, keyboard, joystick, etc..

Each instance of action:display constitutes a sign and carries a literal denotative meaning ('the mouse has moved forward, and the cursor has moved proportionately across the screen vertically') and a connotative meaning that relates the sign to the larger scale semiotic processes of the work ('the relationship of mouse movement to cursor movement is much the same as it has always been, the system is working, the system is reliable, the information the system displays is believable'), and to underlying cultural myths about how the world and the things it contains relate and work together. Denotation, connotation and myth work together to form affect, they way the signs work on and within the user's experiences, forming an aesthetic of the interactive. These sign processes are embedded and embodied in the 'materialities' of the system, the places where the signs are present in the human sensorium, in the devices of display and the devices of interaction. The possibilities of signs are constrained by the materiality of those devices, by their 'manners' of operation and display. The word 'manners' is used carefully here. It relates to both the 'manner' of operation of the devices as physical constructs, and also to the 'manners' of the devices in the sense of conventionalised and constructed behaviours, in a sense like that of behaving properly, of being well-mannered, or ill-mannered. And by inference this suggests, perhaps, a normative state for those devices, a body of expectations and performances that is to be thought normal. Possibly even 'the way things are'. Myths of the natural embodied in mouse, monitor, action, display, action.

These 'manners' can also be thought of as grammars of use, or codes of use. Associated with those terms are paradigms (of possibilities), and syntagms (of utterances). These ideas refer also to concepts of langue and parole, competence and performance – ideas about a range of potential states articulated into a (necessarily) smaller range of actual states. The paradigm includes all the possible 'signs' that constitute the language of the device. It does not include all possible states of the device as some states may not constitute signs within the (currently normal) language paradigm.

The form, nature, grammars, codes, the 'manners' of devices, flavour interactivity, exist as affect. Medium becomes message in the movement and touch of a mouse and the transient signifiers of the monitor or loudspeakers. Display prompts action, communication contains the possibilities of control. Algorithms of materialisation massage data into syntagmatic instance drawn from the paradigm of all possible displays of that data. Actions evoke instances of display. A syntagmatic utterance of action, of using the mouse in a particular way translates into a selection from the paradigm that is the moment of discourse, and the syntagmatic display

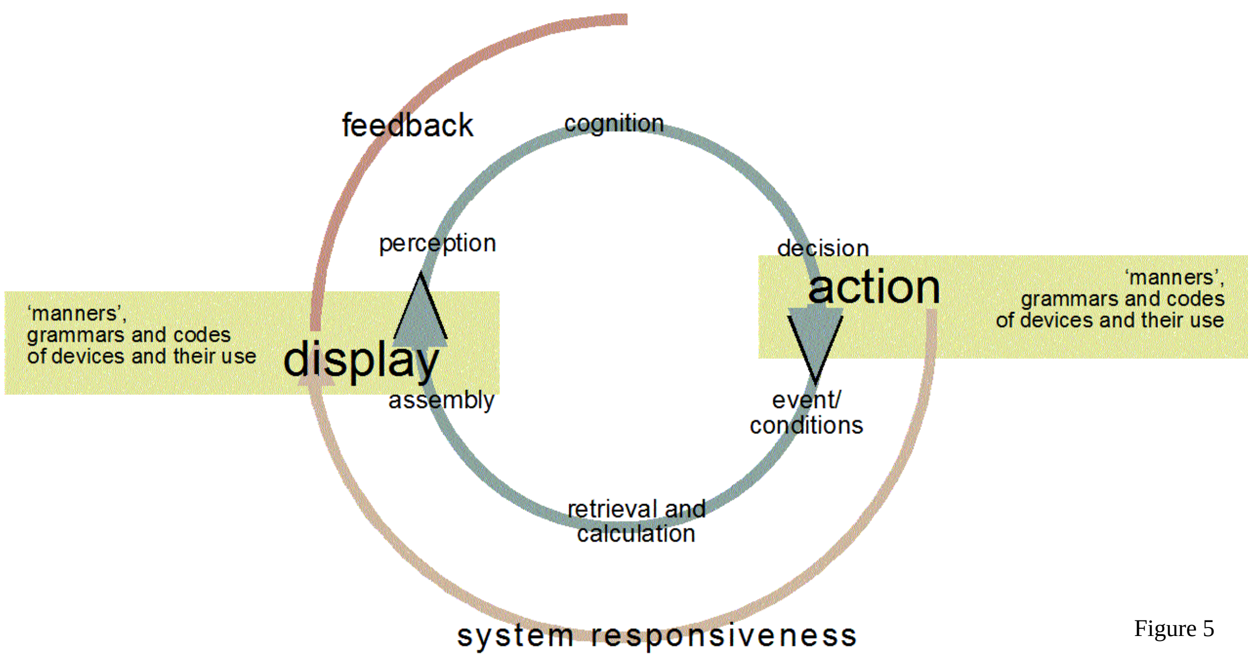


Figure 5

of that selection then forms the paradigm in which using the mouse operates. Along the semiotic axis, where denotation and connotation merge in message and meaning the qualities of interactivity, the 'manners' of the devices, function connotatively. Data brought into being as display takes on meanings both literal and associative.

Audiorom [5]

mouseSigns

Understanding the use of a mouse through the structures of cybernetic loop and syntagmurgy enables a discussion of mouseSigns at the nano-scale of a cursor moving across a screen pixel by pixel, increment by increment, and at the scale of grander narratives of myth, the larger scale semiotic processes that form the aesthetic experience of the artefact as a whole.

The primary signifucations of a mouse relate to movement and to change, by which is meant the operation of a button and the consequences of that act. It is important to differentiate the signs that come from the use of the mouse and the signs that represent that movement on screen. The hand movement of the mouse is noticeably smooth. While the screen display may be visually differentiated and may sign changes in level, texture and even 'limits' to the extent of the current display syntagm the common mouse provides no haptic equivalent of bumping over a line, ascending a buttonised edge, or being constrained in area. The display representation of movement is entirely the outcome of algorithmic operations. The appearance of the representation as a cursor on screen, the way it moves, how it relates to the movement of the hand, whether it jumps five pixels to the right when running over a buttonised image, whether it stops at a certain screen position, are all determined in code. The interplay of algorithms is a syntagmurgic process which leads to the instants of display. The arbitrariness of display then should make more clear the rather limited and conventionalised languages, the manners, that we employ. The possibilities of algorithms are effectively unlimited, the infinity of the paradigm of display is the play of the human imagination.

Similarly, what the system registers is only movement (and that is usually relative rather than absolute) and button actions. It rarely registers the emotion with which the mouse has been moved. And so the signs that are embodied in movement become denatured. Refined into a reading of numbers by the circuitry of mouse and computer the signs we make moving a mouse are rendered as denotation only, their materiality removed, the connotations of the signmaking disregarded a discarded. And yet, of course the sign still works towards ourselves.

The sign-making process of using a mouse is complex – it is, like all sign systems, two directional – it communicates as much to the self (the reader, here) as to the other (the text:system). In the act of making a sign is a feedback loop about making and having made the sign, a sign of the sign itself as a sign. Holding a mouse, and moving it while watching the screen cursor track across is a syntagmurgic act, in which the path of the mouse:cursor is selected from all possible paths (the paradigm) in a series of nano-scale cybernetic loops constantly signifying both the position of the cursor, the relationship between the cursor and other display elements, the relationship between the cursor and its conjectured destination, the modification of action by display

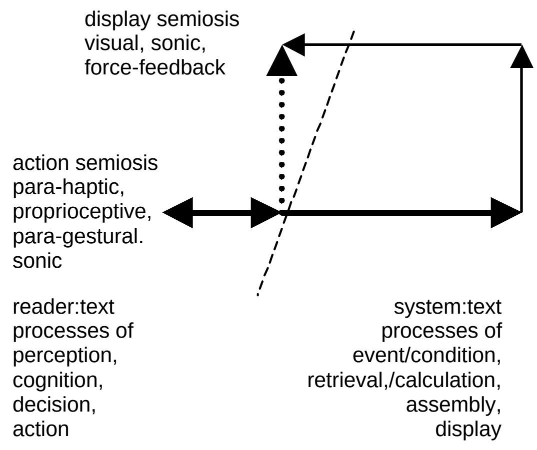


Figure 6

and display by action. Signs about signs and signs about signs that do not yet exist. The reading of these signs is predominantly denotative, but the mode of transit and the texture of the syntagmurgy carry connotative values also. Mouse movement and screen cursor movement are usually integrated together and happen within a time frame that convincingly links action with display, display with action, and connotes values about cause and effect and the reader's 'agency' within the syntagmurgy of the text:system. Such connotations are implicit in the 'manners' of the mouse in the sense of normalcy and myth. A smooth continuity of movement, of unproblematised nano-loops, denote the system functioning normally (a form of phatic communication) and connote certainty, reliability, trustworthiness, not only as an operational process, but also a quality of the experience and of any knowledge gained during it.

The 'manners' of the mouse:cursor are complex, the signs that make up the sign system exist in several imbricated modalities. To use a mouse is to enter into a sign system that is in part

visual

the movement of a screen cursor, rollover effects and other visual feedback cues, the general semiosis of the visual proprioceptive – *muscle tension and the rotation of joints in wrist, fingers, shoulder, back, these sign location on screen as much as the visual does*

haptic, or para-haptic

in that the shape, texture, resistance and yield, smoothness and regularity of the devices movement over surface, the press pattern of buttons, both signify the presence of the mouse and sign values of interactivity

gestural, or para-gestural

because the hand using the mouse takes on particular shape and position within the semiotic process repeatedly and consistently and in this it resembles signing languages

auditory

including the click of buttons, display sounds, the inherent issues of sonic spatiality an the general semiosis of sound.

The signs of mouse read through the various modalities of their signification are about movement and change and these function as signs both within the syntagmurgy of creation and in the syntagmurgy of reception. The signs of mouse function as meta-signs signifying the nature and status of the signs themselves, modelling as moulding the hermeneutic process, embedding mind process in screen display as the wraiths of Sapir and Whorf wander across our discussion.

Using a mouse involves the syntgamurgic processes of reception and creation, and these overlay display semiosis (which includes any way in which data is given material and sensory form) and action semiosis, (the 'self-reading' signs of operation). Usually integrated together and happening within a time frame that convincingly links action with display and so signifies cause and effect, the two co-dependent syntagmurgies of reception and creation spin the cybernetic loop along its semiotic axis. The integration of reception and creation, and the experience of those processes, signs nothing so much as dialogue, responsiveness, agency. In the way that display changes are evoked by action (the signs of cause and effect) and in the way in which this happens at many scales and not only in moments of 'decision', creates a sign field in which the apparent 'coming alive'ness of the display signs the text:system as having attributes of autonomy and intelligence, willingness and resistance, truthfulness and decption.

Because the movement of the mouse and the movement of the cursor map across both the action semiosis and the display semiosis, there is a rich play of how the use of the mouse relates to signifying processes. Using Peirce's categories we can describe and discuss the movement of the cursor on the screen, that is in the display semiosis, as indexical, and in the action semiosis as iconic. Indexical because the cursor is a sign that is existentially linked to the movement of the hand but is not the action itself (much the same as the commonplace smoke:fire analogy, the cursor does not move unless the mouse is moved and the feedback of the movement does not equate to the feedback of an illusory visual surface – the mouse does not bump when the cursor crosses a line). Iconic because the action of moving the mouse itself signs moving the mouse, the sign system of the para-haptic and proprioceptive integrating with the visual. The semiosis of action is a sign of action itself.

The movement of the mouse also signs symbolically. The reader understands through the sign of the cursor, its movement, relative to other signs, and the visual signing of its shape, something about agreed processes of action and cognition. The reader has acquired and uses arbitrary signs of movement (rarely do mouse and cursor track on a 1:1 scale), click, drag, double click, rollover – which seems iconic in the integrated overlay of display and action but is more properly understood as an arbitrary language, acquired and assimilated, made invisible through mastery and familiarity, hidden in myth.

The mouse sign system separates movement and action and in so doing it signs something about the process of thought itself. The cursor is directed to a place and then the button is pressed down, sometimes held down, sometimes released quickly, sometimes double clicked, it does not matter. In this separation of continuous movement and discreet decision cause and effect are structured. The almost idling nano-scale sweeps of the cybernetic loop that are the signs of cursor travel around the screen as the mouse is moved around give way to clicks and decision and the cybernetic loop lurches along its semiotic axis, larger scale changes of display happen, patterns of energy change, temporal and spatial juxtaposition create difference, syntagmurgically form meaningful instances of display, give material form to data retrieval or algorithmic operation.

Placing the hand on the mouse is a formal declaration of entering into a signing relationship. The position of fingers, the rotation of the wrist and forearm, the touch sensations on palm and fingers – these are the first sign of an action semiosis. It is a sign of the signing process that has been initiated. Movement then, scale mapped between mouse movement and display is a sign not only to the system (your action: its reading of a display) but also to the reader. A sign of a process taking place, and that sign denotes the process and connotes its values. The process signed by moving the mouse is a process of searching around, looking around, enquiring. While often metaphored as spatial it is usually undertaken while sedentary. In its sweep of wrist and integrated tracking of the eye, it resembles cognitive acts like purposeful recall, like studying a scene, like the active syntagmurgy of hermeneutics. In Gregory Bateson's image of the blind man and his stick, the mouse is our extended probe by which we get to know about this other world, this representational space. The sign system shifts from looking around to action by, bringing the mouse to halt and changing from free fluid movement of wrist to press action of finger. The action of the finger here is a formal language of operation, arbitrary, almost like alphabetical signs, but where they exist as patterns of light the signs here are muscle actions bounded by time. The duration of a click is a parameter of its signedness and meaning, a means of markedness, difference, affordance. The shift from wrist action to finger action itself signs a difference in mode of operation, denoting decision in place of observation and interpretation. This move – locate – click sequence signs different cognitive states and differing stages on the cybernetic loop and in syntagmurgy.

The move – locate – click process of mouse use becomes a sign for a mental model of decision making and even thought itself. It signs think – decide – act and requires this of the reader. Without this process, meta-signed in the use of the mouse, then, much of the time in the interactive nothing happens. This sign of process is not merely denotative, it does not simply sign the process as process. It is connotative of that process, signing the process with the values of rational mind, decision making, trustworthiness, agency, presence, power. The process of signs and of syntagmurgy means that mouse use, the movement of cursor on screen and the hand behind it, becomes a kind of embodied probe for our thinking process. The movement of the mouse signs mental processes, the cursor becomes the point of our concentration, an outbodiment of the hermeneutic syntagmurgy.

The common mouse process – move-locate-click – deals with one issue at a time. It signs thought as a serial process quite contradicting the structures of associational thought which underlie Ted Nelson's idea of 'hypertext'. This pattern of mouse use is prevalent, dominant, and functions as a myth of normalcy about both the systems and ourselves. It signs up a number of values about the system and its operation, and about the values of the experience and any knowledge gained from or through it. Now among the most common languages used, and that in many diverse cultures and communities, it functions as a powerful myth of normalcy both of system operation and responses to the world. The implicit power relationship of point-and-instruct that has us form unambiguous sentences from menu selections, the look-click-and-its-done operation, are powerful language models for the world. The signs of agency and effectiveness in the dual syntagmurgy of the interactive, where reception and creation overlap, become ways in which we believe the world itself works. All languages become ways of understanding the world.

But it need not be so. Mouse is a distinct and markedly artificial language. Although there is an accustomed normalcy to its use, the nature of the language as being a product of algorithms is open to a poetic exploration and the works of artists and authors can extend, subvert and problematise that normalcy both as a form of explorative enquiry and as a comment on its mythic structures and, ultimately, political values.

Peacock [11]

Within the wide sign systems of the interactive, Mouse is one contributing element to the holistic experience of the reader. Rarely does any one contributing element stand alone, usually they all work together to form the signs which inform the aesthetic of the discourse.

It is important to locate mouseSigns within that holistic. Both in their denotative operational functions and in their connotative and mythic values.

Juvenate [10]

AESTHETIC DOMAINS

Taken holistically interactivity resembles nothing so much as it resembles story-telling because it binds together the syntagmurgy of reception and creation in an act of communication which is inherently 'narrative' in that it is time-based and driven by events. Structures of cause and effect, the modification of meanings by juxtaposition in space or sequence, connect the instances of action/display within larger structures of meaning, and those larger structures, themselves, have semiotic function. And it is in the functions, effects, and affect of that semiosis that the aesthetics of interactivity may be examined and discussed.

The narratives of the *interactive* are of many kinds and it is important to distinguish among them. The main distinction being drawn between those narratives which are predominantly concerned with information and those which are predominatly concerned with the experience of interactivity itself. In the first group are narratives of the interactive where the concern is to give access to an underlying database of information which is understood to be more important than the experience of the interactive itself. This kind of activity is characterised by designs which seek to make the interactive 'transparent', by screen design metaphors that converge with or replicate as on-screen image, a user's 'real world' experiences. Where the word 'functionality' trades alongside 'efficiency' and 'workflow' in a kind of latter day Taylorism.

Of the other kind are interactive artworks and games where the concern is with the experience of the interactive, where the narrative of interaction is the thing in itself. Here there are divergent metaphors, problematised interfaces, toys in the sense of things that are played with for the purpose of playing. This difference is described by Bolter and Gromala as Transparency and Reflectivity and discussed in detail in *Windows and Mirrors.* "*Contemporary culture"* they say, "*is receptive to transparency (the window) and also to an alternative, self-reflective style (the mirror). This latter style, which was truly avant-garde in the early twentieth century, has become the aesthetic of rock concerts in the early twenty-first century. So digital designers and the growing world of digital entertainment need to master both styles. Digital interface design needs to master both styles as well."*

Bolter and Gromala [3]

and they go on to comment "*We might be tempted to think that transparency is for "serious" digital applications, such as productivity software, while reflectivity belongs exclusively to art and entertainment. But it is not that simple*." Making the point that the experience of most users is an oscillation between the two modes, that is to say that the discourse moves from transparency to reflectivity as the user's purpose, task and concentration changes. The oscillation between the transparency of a word processor application contrasts with the reflectivity of the graphic user interface as the user changes from that application to another, as the user's concentration moves from application to system operation, as the syntagmatic processing of interaction proceeds.

Small Fish [8]

We can think of these two kinds of the interactive, the ones concerned with information and the ones concerned with experience, characterised as transparency and reflectivity by Bolter and Gromala, as being points towards the ends of a domain, or two-dimensional field.

information transparency	**the interactive**	experience reflectivity

The discourse that is the experience of the work is located within this domain with a tendency towards one end or the other. "Each design," say Bolter and Gromala, "is a combination of these two strategies – perhaps with more elements of one or the other". So, productivity software, tends towards information and transparency and digital artworks tend towards experience and reflectivity, but both have elements of the other, also. This tendency is the affect of the sign systems operating within the discourse. Transparency is an affect, an aesthetic experience of the interactive, of the dually-syntagmurgic discourse. Reflectivity, likewise. Both derive from a reading of the signs of the essentially dualled-syntagmurgy of the *interactive*, signs which, collectively inform the experience of the interactive and define the aesthetic experience of it, and which are framed by the 'manners' of devices.

The paradigm of an interactive text is unknowable. Because it has a physical rather than material form its extent and structure can never be seen directly and only parts are ever seen at any one time (syntagmatic display). And there is no guarantee that those parts will be seen in the same juxtaposition of sequence or space more than the once of a particular instance. This has a number of consequences which will be discussed here through the notion of domains or two-dimensional fields. A domain represents a condition in the potential experience of the interactive, a sign of the quality of the experience, and is marked at each end by differentiated forms of the experience of that condition. Domains, as thinking structures, relate to but are not the same as binary oppositions in structuralism. What is located within a domain is the discourse of the reader and the text, (although of course it is easier to think of these domains as being features of the text as that mode of ascription is prevalent in the way we think of the world culturally). The domains, or rather the reader's experience of what these domains deal with function within the semiotic processes of the reader:text discourse as signs about the text and the properties of the text.

A primary domain addresses where the reader stands in relation to the text itself. This domain (*the locus of interactivity*) relates to actions of the reader which generate instances of the text (the syntagmurgy of action:display) and is marked at one end by the term *deliberate* and at the other *inadvertent*. The actions of a person entering text through a keyboard characterise the *deliberate* – their actions are purposeful and are directed towards the interactive through a device of interactivity they are aware of. The *inadvertent* is characterised by an artwork installation in which a sensor registers the presence of a person without their knowing it and this triggers an event. A work perhaps like Susan Collins….. banana, or the warrington museum.

deliberate **locus of interactivity** inadvertent

A reader who has inadvertently interacted with a text may work out how the installation operates, say by recognising a PIR detector, and then proceeds to step back and forth into and out of the field of the device causing the installation to respond. In this the reader can be said to have moved through the domain from the inadvertent to the deliberate. Similarly, an inexpert typist………………….

This points out the inherently dynamic nature of the domains and also that within them the location of any discourse is most properly thought of as a generalised tendency. Within any discourse there will usually be a range of locations within the domains (perhaps much the same idea as Bolter and Gromala's 'oscillation' between transparency and reflectivity), and the particular tendency at any one time is an outcome of the dual-syntagmurgy of the discourse.

The first encounter with a text, the initial moment of discourse, is likely to tend towards the inadvertent as the way of engaging with the text is established. Familiarity with similar texts means that the discourse soon relocates towards deliberate. If the discourse includes the unfamiliar object, or unfamiliar processes then such a movement may be slower. If an object at first appears familiar but proves to work in unexpected ways (keystrokes are remapped to an alternative layout, the screen cursor moves erratically, say) then the discourse is abruptly relocated to the inadvertent. Across time, all discourses will drift towards deliberate.

In this domain the mouse:cursor device when operated within its 'normal' manner tends towards the deliberate. It is markedly less demanding of accuracy than the keyboard (the screen cursor generally occupies an area rather than a precise location) but when used as a point-click device it is clearly 'deliberate' and connotes values of control, action, decision. However the mouse:cursor device is relatively easy to subvert through reframed algorithms that, perhaps map horizontal movement of the mouse to verticalmovement of cursor, and/or vice versa, or translate vertical movement to the rotation of a visual display element. These relocate the (discourse of) the device towards the inadvertent, but with the general drift of learning that presses things towards the deliberate. Hidden hotspot areas drifting around the screen, their movement dictated by randomised increments and their direction dependent on the chaos of collisions with sounds triggered by rollover events – such a thing would be inadvertent and could well maintain its tendency long after the general principle is 'understood'.

This domain of deliberate and inadvertent provides an initial classification of instances of the interactive, and, along with other domains which are imbricated within the overall discourse, will direct our attention to the signing pliaces where the discourse happen – to the devices of display and action.

As an interactive text is not materially embodied it follows that all users face a potential uncertainty about what they have seen or missed, how much of the whole they have seen, or missed, how what the have seen realtes to the whole, whether they can find any of it again. An uncertainty quite unlike that of a book or painting which can be 'scanned' or surveyed quickly across or through their material paradigm.

This aspect of the interactive is addressed through two domains. The first is labelled 'missingness' and the second is lablelled 'tmetic anxiety'. This term, tmesis, is derived through Aarseth and Barthes from Aristotle, where its general meaning is to do with an author never knowing how much of a text a reader may have skipped. It clearly relates to the idea of missing-ness, but here it applies not to the author's experience in constructing the text (for there is an important difference between the privileged maker's view of a text and the reader's discourse) but to the experience of the user, reader, player whatever you want. These domains are interconnected. One is about how much the user experiences a sense of missingess in the text, a sense of not having seen important material, of other options or opportunities not taken or followed. The other of these domains is about the user's response to that sense of missing-ness, the state of their anxiety.

incompletedness **missingness** completedness

low **tmetic anxiety** high

Forms of the interactive map across these two domains differently. An information-purposed (transparent) text will necessarily want to locate towards 'completedness' on the 'missngness' domain, and towards 'low' on the 'tmetic anxiety' one because such a pairing is about an affect (the experience of the aesthetic) which includes an acceptance that the information that comes from the system is believable and dependable.

A text adventure game (tending to reflectivity), with its many twisty little passages, would locate towards 'incompleteness'

in the ‘missingness’ domain and could well move around variously within the domain of ‘tmetic anxiety’ at times tending towards the low end , and at others towards the high end as this is about an affect which includes dramatic tensions, which motivates further exploration, and this can be seen to equate with Bolter and Gromala’s ‘oscillation’ between transparency and reflectivity.
Many information based systems include, as a designed feature, a representational overview, plan or ‘site map’ of the constituent parts, and many adventure gamers draw up maps of the worlds they explore. In information based systems such ‘site maps’ are a mechanism for reducing tmetic anxiety and bringing a tendency towards completedness – advantaging ‘transparency’. The affect is reassurance, confirmation, valorisation, ‘if you have missed something, you can find it again, this whole thing can be understood’. The user is made less anxious because less is experienced as missing. The gamer’s map drawn out as she explores has the affect of taking ownership, building mastery, dynamically relocating the discourse within the domains towards both ‘completedness’ and ‘low’ while demonstrating the extent of the task that has been undertaken, shifting from reflectivity to transparency and back again.

Juvenate [10]

The amount of work required during the syntagmurgy of the discourse of using a mouse may have a good deal of effect on the domains of missingness and tmetic anxiety. Tracking the mouse:cursor back and forth across the screen creates a density of decision of process within the nano-loops of action:display, and the repeated cohnitive focussing on different areas may sign the possibility of having missed something.

Those domains of missingness and tmetic anxiety tie in with other domains, mostly clearly with ones which locate the user’s experience of the discourse (although we may describe some of these features as if they are properties of the work they are of course only ever features of the user’s experience) within domains of ‘cursality’ and ‘causality’. Cursality is about the pathedness of the text – about how many choices are to be made, how many paths may possibly be taken, howmany paths have been taken and remain untaken, how the various paths interconnect. As a domain ‘cursality’ maps the user’s experience of complexity of structure, and that experience connotes affects such as reliability, certainty, trustworthiness. But the mapping is not simple – a complexity that is ‘understandable’ is more reassuring than a designed ‘simplicity’ in which the user feels lost. The measure is how the user experiences the connotations of complexity or simplicity within the discourse which is the interactive. Cursality may be a product of the number of choices the user has faced, or the depth of decision they have had to make. Resolving ambiguities, for example, may markedly increase the cursality even though there are few choices. The density of decision making acts as sign for cursality

simplicity **cursality** complexity

An information based system will usually locate towards low cursality. Reducing the number of options speeds operations, creates a relatively uncomplicated map of operations in the user’s mind. And from that comes a ‘belief’ in the trustworthiness of the system – a defining characteristic of what is usually required of ‘transparency’.
Causality, or the consistency of the consequence of actions, is about the way cause and effect operate in the discourse. Actions and displays are bound together through cause and effect and create spatial and temporal juxtapositions of elements from the paradigm of the work. The experience of Causality connotes values such as certainty, understand-ability, trustworthiness, consistency. It is a domain that is marked at one end by redundancy – a clear almost ritualised affect – and entropy. Between the totally predictable and the entirely chaotic, between the transparent and the reflective.

entropy **causality** redundancy

It is doubtful if any discourse can ever locate at the extreme of entropy. That which is truly, deeply, or very entropic is accommodated rapidly by the human mind and its own unpredictability becomes expected and so redundant. Discourse located towards, or tending towards the entropic, is exciting because of the unexpectedness of potential juxtapositions and because of the thrill of the oscillation between states, within the domain. Redundancy brings predictablitity, stability across time and instance and so an affect of reliability, believability. Entropy brings the thrill of the unexpected. Games rely on a tendency towards entropy and employ elaborate mechanisms to maintain it, to counter the dynamic tension of experience leading to redundancy. Randomising sequences, apparent artificial intelligences, simulator models with chaos engines, these strategies of game play ensure a location towards entropy. Mastery of the game moves the discourse towards redundancy and an aspect of the ritualised predictability is the reward of mastery.
Cursality may be signed in the mouse:cursor syntagmurgy through the density of movement and variation in the requirements of movement and, as with causality, in the consistency of cause and effect structures and relationships within the nano-loops and macro-loops of the artefact.

Conclusion

Working together these domains (and others not discussed here) form from the signs of the discourse a ‘view’ of that experience itself. That view is inherently connotative in that it is value laden and mythic in that it signs grander narratives of how the world works. The signs which communicate the view are many and varied, they include visual sign systems, acoustic ones, environmental ones, and ones tied in with the operation of such things as mouse, keyboard, joystick, drawing tablet, etc.
There are well established critical approaches to the discussion of the visual and the sonic. Approaches to the semiotics of graphic design and visual imagery, and to music and sound, which address features such as colour, composition/layout, typographic elements, modes and styles of representational imagery, genre of music, the timbral qualities of non-indexical sounds, the languages of symbolic sound that are brought into play, a reasonably well-established and function much the same across many media. It is important that we develop ways of discussing the relatively

new sign systems that also play within the interactive and which make it so very different to other forms.

Peacock [11]

Notes

1: syntagmurgy – with no apologies for this neologostic act – this is a compound word bringing together *syntagm* (in its sense of a set of signs drawn from a larger paradigm) and the –*urgy* of English words such as dramaturgy, metallurgy, liturgy, which is rooted in the concepts of 'work' – in the sense of bringing forth through labour. So *syntagmurgy* is the act of bringing forth a selection (grammatical in terms of the language paradigm, or not) from a larger set, either unconsciously as part of perception/cognition (as in hermeneutics) or more consciously and deliberately (as in the use of an interactive menu) or inadvertently (triggering a sensor). This term relates to, and resembles, Aarseth's ergodic but grounds the act of selection in bodies of signs and the discourse of readings. The concept of *syntagmurgy* is one of process not of act, the 'bringing forth' is never completed nor does it finish while discourse continues.

References

[1] Kuhn, Thomas S, The Structure of scientific Revolutions, 3rd Edition, University of Chicago Press, 1962
[2] Qvotrup, Lars Digital Poetics, in Digital Media Revisited, ed Listejol, Morrison, Rasmussen,MIT Press, 2003, pp 254
[3 Bolter,] Jay David and Gromala, Diane, Windows and Mirrors: interaction design, digital art and the myth of transparency, Leonardo Book, MIT Press, 2003, pp 68
[4] *in* Hayles, N Katharine, How We Became Post-human, University of Chicago Press, p. 84
[5] *see* Manovich, Lev, The Language of New Media, MIT Press, 2001
[6] *see* Aarseth, Espen J, Cybertexts, Johns Hopkins Press, 1999

Illustrative Interactive Media

[7] Thomas, David and Lucas, Simon, on Erewhon, graphically enhanced CD-Rom, publisher Cooking Vinyl, 1996, COOK CD 105
[8] Furukawa, Kiyoshi, Fujihata, Masaki and Munch, Wolfgang, Small Fish, CD-ROM, ZKM #3, 1999
[9] audiorom, shiftcontrol, audiorom, ARL101, 1999
[10] Glaser, Michelle, et al, Juvenate, CD-ROM, Australian Film Commission, 2000
[11] Peacock, Alan, various works on www.404.org.uk, 2001 -2004

Sign of a Threat: The Effects of Warning Systems in Survival Horror Games

Bernard Perron
Department of Art History and Film Studies
University of Montreal
Canada
(514) 343-7384
bernard.perron@umontreal.ca

ABSTRACT

This paper studies the way survival horror games are designed to frighten and scare the gamer. Comparing video games and movies, the experiential state of the gamer and that of the spectator, as well as the shock of surprise and tension suspense, it focuses on the effects of forewarning on the emotional responses to survival horror games.

Keywords

video games, survival horror, fear, surprise, suspense, game design, gameplay.

> *There is no terror in the bang, only in the anticipation of it.* (Alfred Hitchcock)

1. INTRODUCTION

David Bordwell, one of the most important figures in promoting a cognitive approach to cinema, once wrote that filmmakers were "practical cognitive psychologists" because they take advantage of the ways spectators draw upon everyday thinking while viewing a film (for instance, going beyond the information given by categorizing, drawing on prior knowledge about real-life or films, forgetting some elements in order to remember others, making informal, provisional inferences, and hypothesizing what is likely to happen next) [3]. Such assertions can obviously be made about game designers, too. In comparison with the spectator, it is even more plausible that a gamer may choose to play a video game in alternative ways (freely setting his own goals, testing the limits of the game, playing *with* *the* game instead of playing *the* game) or that a game might be used to other ends (to help overcome phobias, for instance[1]). Regardless of how the game is used, game designers know how to elicit the sort of activities and emotional responses that will create the experience they want the gamer to have. When it comes to survival horror games — the subject of this paper — designers know exactly how the gamer shapes his journey to hell.

First published at COSIGN 2004
14–16 September 2004, University of Split (Croatia)

This paper, therefore, studies the way survival horror games are designed to frighten and scare the gamer. They do so by relying on horror mythology and conventions of horror movies. According to Ed S. Tan, they create both fiction emotions (emotions rooted in the fictional world and the concerns addressed by that world) and artefact emotions (which arise from concerns related to the artefactartifact, as well as stimulus characteristics based on those concerns) [26]. But above all, their design is made to elicit gameplay emotions. That is to say fear, fright or dread that arise from the gamer's actions in the game-world and the consequent reactions of this world. Gameplay emotions come from various actions: exploring, being lost, fighting, being attacked, feeling trapped, dying, using various weapons, being challenged, solving problems, etc. In an overall analysis of the *Silent Hill* series (Konami/Konami, 1999-2003[2]) [21] in which I examined these gameplay emotions, I talked about one of the famous features of the series: the avatar's pocket radio that transmits white noise to warn the gamer that one or many monsters are nearby. I referred to the notion of forewarning, but did not develop this subject. While my observations will stem and borrow from my visits to the town of Silent Hill, I wish to broaden the examination of warning systems by broaching a few other PlaySation games: *Resident Evil* (Capcom/Capcom, 1996), *Resident Evil 2* (Capcom/Capcom, 1998), *Fear Effect* (Kronos Digital Entertainment/Eidos Interactive, 2000) and *Fatal Frame* (Tecmo/Tecmo, 2002).[3] What then are the effects of warnings? I do not mean visual and audio devices informing us of the avatar's health or remaining ammunition (although these warnings are part of a whole), but rather those signals of on-coming monsters off-screen? What are the emotional responses these signs of a threat induce? To answer these questions, I'll be relying in part on empirical psychological research. Psychological approach, results and discussion are indeed very relevant to this study.

2. SHOCK AND TENSION

Generally speaking, survival horror games follow the same formula, and gamers know what gaming experience to expect.

---

[1] The Cyberpsychology Laboratory Of The University Of Quebec In Outaouais (<http://www.uqo.ca/cyberpsy/>) is using *Half-Life* (Valve Software/Sierra Entertainment, 1998), *Unreal Tournament* (Epic MegaGames/ GT Interactive, 1999) and *Max Payne* (Remedy Entertainment Ltd/Gathering, 2001) to treat spiders and heights phobia.

[2] *Silent Hill 4: The Room* is to be released in September 2004.

[3] This list is far from exhaustive. Consequently, my argument remains inductive.

At the plot level, the hero/heroine investigates a hostile environment where he/she will be trapped (a building or a town) in order either to uncover the causes of strange and horrible events (*Alone in the Dark*, *Resident Evil*, *Siren*) or to find and rescue a loved one from an evil force, be it a daughter (*Silent Hill*, *Fear Effect*), a mother (*Clock Tower3*), a wife (*Silent Hill 2*) or a brother (*Resident Evil 2*, *Fatal Frame*). At the action level, in a third-person perspective[4], the gamer has to find clues, gather objects (you cannot do without keys) and solve puzzles. In order to survive with the weapons he has (or will come across), the gamer has to face numerous impure, disgusting, creepy and threatening monsters (zombies, demons, mutated beasts, abnormal creatures, spirits, vampires, etc.). The conflict between the avatar and those monsters is the dominant element of horror.

Since the release of *Silent Hill*, one way of differentiating these games has been to distinguish the more gruesome action-based and quick thrill jump scares of *Resident Evil* from the chilling atmosphere and psychological approach of the Konami series. In fact, this comparison mirrors the acknowledged opposition between horror and terror. As Will H. Rockett puts forward, horror is compared to an almost physical loathing and its cause is always external, perceptible, comprehensible, measurable, and apparently material. Terror, as for it, is rather identified with the more imaginative and subtle anticipatory dread. It relies more on the unease of the unseen. “The most common time of terror... is night, a great absence of light and therefore a great time of uncertainty” [22: p. 100]. Without daylight, certainty and clear vision, there is no safe moment. Terror expands on a longer duration than horror does. By plunging its gamer alone in the dark or in mist and giving him only a flashlight to light his way (and so forcing him to play alongside the imperfectly seen), *Silent Hill* and *Fatale Frame* succeed at creating the fundamentals of terror. Though the young girl Miku, the gamer’s avatar in *Fatal Frame*, suddenly finds herself face-to-face with a spirit or Jill in *Resident Evil* frequently meets up with zombies, these encounters are not the same when the hero can’t clearly see their enemies or their surrounding environment. With the presence of monsters and their unavoidable onslaught, these kinds of games would be more aptly called survival *terror* games. But as long as the contrast between horror and terror relies in great part on the building and, above all, on the sustaining of a feeling of dread, another suitable way for this study to view this contrast is to refer to the another famous distinction.

Crawling with monsters, survival horror games make wonderful use of surprise, attack, appearances and any other disturbing action that happens without warning. According to Robert Baird’s analysis in “The Startle Effect. Implications for Spectator Cognition and Media Theory”, the games have the core elements of the (film) threat scene’s startle effect at their disposal: “(1) a character presence, (2) an implied offscreen threat, and (3) a disturbing intrusion [often accentuated by a sound burst] into the character’s immediate space. This is the essential formula (character, implied threat, intrusion) one finds repeated hundreds and thousands of times since Lewton’s first bus effect” [1: p. 15]. In the aforementioned famous scene of Jacques Tourneur’s *Cat People* that Val Lewton produced in 1942, the spectator is lead to believe that the female character is followed by something from the left, only to be caught off guard by a bus barreling in from screen right. As hostile as the environment might be, it is very unlikely that the gamer (who has embarked upon a lengthy exploration) will not be taken off guard and be surprised. Improving the surprise effect of the long fanged monsters breaks through the cellar and the first bedroom window in *Alone in the Dark* (I-Motion Inc. & Infogrames/Interplay, 1992), the dogs that burst through windows when you cross a corridor at the beginning of *Resident Evil* is considered a classic game startle effect (Figure 1). There is more than one bursting window effect in the *Resident Evil* series. Zombies or creatures can always burst out of window whether it is barricaded or not. In *Fatal Frame*, it is above all the sudden appearances and disappearances of spirits that give you a start. The *Silent Hill* series has few monsters that launch underhand attacks. In the first game, there is also a great scene in the elementary school that gives you a good scare: A cat springs out of a box at the very moment Harry, the gamer’s avatar, is about to open it. The game-world of *Silent Hill* is haunted by sudden noises here and there that have no visible or identifiable source.

Figure 1: A dog bursting through the right window in *Resident Evil* (Capcom/Capcom, 1996).

To trigger sudden events is undoubtedly one of the basic techniques used to scare someone. However, because the effect is considered easy to achieve, it is often labeled as a cheap approach and compared with another more valued one: suspense. As in the well-known example of Alfred Hitchcock, a bomb that suddenly explodes under the table where two people are having an innocent conversation will surprise the spectator for only few seconds at the very moment of the explosion. However, if this spectator is made aware that the bomb is going to explode at any minute, he will participate in the scene and feel suspense for the whole time preceding the explosion. “The conclusion is”, Hitchcock says, “that whenever possible the public must be informed” [27: p. 73]. The shock of surprise is consequently taken over by the tension of suspense.

As Noël Carroll asserts in *The Philosophy of Horror or Paradoxes of the Heart*, suspense is not unique to horror, but

[4] There are also first-person horror games, but they are indeed called such. For instance, *Nosferatu. The Wrath of Malachi* (Idol Fx/iGames Publishing, 2003) is presented as a “first-person shooter survival horror”, and *Clive Barker’s Undying* (EALA/EA Games, 2001) has been categorized among others as a “surviquake horror”.

rather is a key narrative element in most horror stories [7: p. 128]. In Carroll's curiosity theory[5], although the emotions of horror and suspense might be different (the object of horror is an entity – the monster – and that of suspense is a situation), they can coexist and bring about a concerted effect, especially when it comes to one of the most characteristic themes of horror narration: discovery [7: p. 144]. Discovery is also the theme of a large number of survival horror games. In a "drama of corridors" (one of Carroll's expressions that applies quite well to the maze structure of these games and many others[6]), the gamer has to find the virus or the supernatural force responsible for the rise of the monsters. And he can expect to fight a last boss monster at the end. Although suspense can be created in the overarching structure of the plot, it can also be generated during short events or incidents. To borrow, yet again, from Carroll's terminology [8], suspense can arise in regard to the plot's few macro-questions (e.g., will the hero/heroine find the loved one?) or the more numerous micro-questions that connect one fictional event to another. As the tension intensifies when we have to answer these micro-questions (e.g., will the bomb explode under the table while the two people are still talking?), and because it touches the action level of video games, I'm interested in suspense at the episodic level. But still, as Greg M. Smith does regarding film, we have to argue that the primary emotive effect of games is to create a mood, i.e. "a preparatory state in which one is seeking an opportunity to express a particular emotion or emotion set" [23: p. 38]. A fearful mood therefore encourages and prepares you to experience fright, and a good dose of panic bolsters the mood in return. Just as gamers do not like boring games, neither would they appreciate being panic-stricken all the time. It's all about maintaining a good balance.

Suspense becomes significant to the study of the cross-media genre of horror when one looks at its fundamental elements. For Dolf Zillmann, "suspense is conceptualized as the experience of uncertainty regarding the outcome of a potentially hostile confrontation" [30: p. 283]. Three psychologists quoted in Carroll's "Paradox of suspense" give this definition: "We view suspense as involving a Hope emotion and a Fear Emotion coupled with the cognitive state of uncertainty" [6: p. 78]. The notion of uncertainty is, without a doubt, at the core of suspense. When a danger or threat is revealed and you are sure of the situation's outcome, there is no suspense. The more the chances of succeeding are slim, the more the presentation is suspenseful. Suspense is a future-oriented emotion, but also a character-oriented one. Doubt and insecurity are bound to one or a few protagonists .You're made to adopt the protagonist's position to follow the event and to live side by side with him the length of the action. But, studies of suspense have revealed that a character does not only have to be in a distressing situation, he also needs to be liked. Comisky and Bryant's experiment of varying levels of perceived outcome-uncertainty and disposition toward the protagonist confirm that audiences get involved with and become more anxious about a hero with whom they have a strong affinity [9: p. 78]. Bonded with the character that represents him in the game-world, the gamer is visibly driven to have this disposition toward his avatar. Being fond of the protagonist causes more hope for a favored outcome and more fear about the possibility that it might not occur. As a matter of fact, fear emotions are also central to the understanding of suspenseful drama. Again according to Zillmann, suspense in drama is created predominantly through the suggestion of deplorable and dreadful outcomes. "It features people about to be jumped and stabbed, about to walk into an ambush and get shot and about to be bitten by snakes, tarantulas, and mad dogs. The common denominator in all of this is the likely *suffering* of the protagonists. It is impending disaster, manifest in anticipated agony, pain, injury, and death. Suspenseful drama, then, appears to thrive on uneasiness and distress about anticipated negative outcomes. In short, it thrives on fear" [31: p. 136]. This emotional response only evolves during the anticipation of the final result, a rather limited result. Micro-questions raised by expected dangerous and harmful events have, as Carroll remarks [6], only two potential and opposite outcomes. In most cases of survival horror games, the avatar survives the attack, runs away from or kills the monster or he does not.

3. TO BE WARNED

To put the gamer in the wanted emotional state, game designers draw upon the relation between emotion, cognition, and perception. As cognition arouses emotion on the one hand, emotion organizes perception on the other hand. Following Carroll's analogy, emotions can be seen as searchlights. "They direct attention, enabling us to organize the details before us into signification wholes or gestalts. Where the emotional state is one of fear, we scan it for details highlighted as dangerous…" [5: p. 28]. This is much the same as the preparatory state of a mood described by Smith: "A fearful mood puts us on emotional alert, and we patrol our environment searching for frightening objects. Fear makes us notice dark shadows, mysterious noises and sudden movements and thus provides more possibly frightening cues" [24: p. 114]. Undeniably, there isn't a better frightening cue than the sign of a threat by a monster.

In psychology, the concept of threat is associated with the one of "anticipatory fear" and psychological stress [17]. Incidentally, much empirical research has studied the effect of anticipation and the emotional impact of prior information. For instance, relevant to the distinction between shock and tension is an experiment by Nomikos et al. that shows two versions of a film portraying wood-mill accidents. The first without warning and the other one with warning (as shots depicting the victim's finger approaching the whirling blade of a milling machine), demonstrate that: "(a) Long anticipation of a harmful confrontation (suspense) is more disturbing than short anticipation (surprise); and (b) most of the stress reaction occurs during the anticipation or threat period, rather than during the actual confrontation when the subject views the accident itself" [20: p. 207]. In general, studies reiterate these conclusions. Such is the case in the article by de Weid et al. entitled "Forewarning of Graphic Portrayal of Violence and the Experience of Suspenseful Drama" [10], and in Hoffner and Cantor's article, "Forewarning of a Threat and Prior Knowledge of Outcome" [15]. Though I could expose the details of these experiments, I would rather discuss an earlier experiment

[5] Mark Vorobej calls Carroll's solution to the paradox of horror a "curiosity theory" because for Carroll horror appeals to cognitive pleasures associated with the discovery of monsters, the objects of fascination [29].

[6] Furthermore, as I pointed out [21], the play of ratiocination that Carroll associates with horror fiction becomes literal in horror games.

conducted by Cantor, Ziemke and Sparks concerning the "Effect of Forewarning on Emotional Responses to a Horror Film" [4] which is at the root of my remark. Cantor, Ziemke and Sparks show that if, intuitively, prior knowledge about an upcoming frightening event would seem to reduce its emotional impact by decreasing uncertainty about what will happen, it is not what actually happens. In fact, on the contrary, the notion "forewarned is forearmed" does not lead as much to "emotional defenses" or effective coping strategy as to a build up of lasting arousal prior the event [4: p. 22-23]. Using heart rate as the measure of physiological arousal (a method they call into question however) and varying the conditions of the forewarning of forthcoming events in four scenes of the "made-for-television" movie *Vampire* (one version with no forewarning, a second with a vague warning and third one with explicit forewarning), the researchers asked their subjects to rate their anxiety, fright and upset[7]. The following observation resulted from the answers they collected: "[f]orewarning of upcoming events did the opposite of 'forearming' subjects against emotional reactions. Subjects who were given prior knowledge of upcoming frightening events reported more intense fright and upset in response to the movie than did those who had no forewarning. It is interesting to note that reports of fright and upset were intensified by forewarning, but reports of anxiety were not. As will be recalled, fright and upset were expected to reflect responses to specific depicted or anticipated events, whereas anxiety was presumed to denote an uneasiness over uncertain outcomes. Given that forewarning should have decreased rather that increased uncertainty, it does not seem surprising that anxiety ratings were not increased by forewarning" [4: p. 30]. The results also show that forewarning did affect only the two scenes related to disturbing and brutal events. In the final analysis, simple forewarning is not a way of preventing intense emotional upset. It is worse than having no information about an upcoming event. We can understand why designers of horror games take advantage of this technique.

As opposed to the conditions of an experiment, the use of forewarning in an ongoing experience of survival horror game is governed by specificity. Because the gamer controls an avatar, the game narratives tell only what this main protagonist knows (i.e. the narration is restricted in a way that is characteristic of investigation stories). Even when there are different playable characters such as the three mercenaries in *Fear Effect* or even the ten in *Siren* (Sony/Sony, 2004), playing alternatively does not really change what you have to do in each segment. If it did, it might be just finding another playable character as in *Resident Evil 2*. Anticipatory fear is therefore elicited during specific sequences. What's more, because of the antagonism of harmful monsters and the game's confined spaces (usually rooms and corridors), threat is always impending. This is even more the case in the immediate off-screen than in far away places. In fact, the duration of the suspenseful anticipation has to be kept in perspective, though it can sometimes be quite long and it always depends on the (re)action of the gamer (we'll come back to this question later on). Anticipation is not to be counted in minutes, as in Hitchcock's example, or even in tenths of a second. The duration is equal to the short anticipation (4.33 and 6.67 seconds) of the aforementioned experiment carried out by Nomikos et al., rather than with long anticipation (18.75 and 25.75 seconds). Nevertheless, as a video game is defined by the here and now of a situation, the question is still to differentiate the effects of anticipation versus none at all (0 seconds). So as to warn its gamer, survival horror games have various warning systems built on physical cues and/or audio and visual cues either displayed on the screen, presented at an extradiegetic level, or integrated into the game-world. We'll now look at these different types of warning systems in the previously mentioned and chosen games.

Fear Effect has a Fear Meter displayed in the upper left corner of the screen (Figure 2).

Figure 2: A Fear Meter appears in the upper left corner of the screen as Hana is about to face devils in *Fear Effect* (Kronos Digital Entertainment/Eidos Interactive, 2000).

The meter represents the heartbeat of Hana (and of the two other playable characters, Royce and Deke). You see and hear it increase as she becomes more afraid. Since there aren't any health power-ups available in the game, you have to perform well in a stressful situation. The Fear Meter appears when you are about to face human guards or monsters. In the underground hell, the scene in Figure 2, the Fear Meter becomes visible in order to show the threat in a empty space clear just a second before. When Hana goes forward, she is attacked by red devils with hand scythes falling down from the sky. This happens more often then not when her enemies are nearby. The Fear Meter sometimes appears long before she encounters danger.

---

[7] "Consistent with their common dictionary definitions, anxiety was assessed to reflect a non specific sense of uneasiness and uncertainty about what was occurring or was about to occur. Fright was thought to reflect a more direct response to specific threatening events. Upset was used to detect any negative experiential aspect of a subject's response" [4: p. 26].

Figure 3: The Fear Meter remains visible for a while in the village of *Fear Effect* (Kronos Digital Entertainment/Eidos Interactive, 2000)

The Meter remains visible for about two minutes in a search scene in the village, for instance (Figure 3). Although Hana (armed with her pistol as we can see) has faced green zombie-like natives in the left branch, you yet still anticipates more action on the right branch.

Fatal Frame also has a visual device to warn . In third-person perspective (Figure 4) or through the viewfinder of Miku's camera (Figure 5), the screen displays a filament at the lower right corner in the former case and in the upper middle in the other.

Figure 4: At the lower right corner of the screen, the filament turns orange as Miku, in the middle of the frame, is about to enter a room in *Fatal Frame* (Tecmo/Tecmo, 2002).

Coupled with heartbeats, eerie sounds and the controller's vibration, the filament turns orange when spirits are nearby (and blue when you are near a clue). This device is more than essential in *Fatal Frame* as you face incorporeal entities that are otherwise translucent and as you are plunged into darkness with only a

Figure 5: Through the viewfinder, in the upper middle of the screen, the orange filament signals the presence of a spirit, here hanging from the ceiling in *Fatal Frame* (Tecmo/Tecmo, 2002).

flashlight to lit your way. As a typical warning system, the filament turns orange when spirits are in a room with you. But because of their nature, it also glows when the spirits are in another room. While most of the survival horror games segment their spaces making you cross doors through a straight cut or an opening loading screen, *Fatale Frame* frightens you by making you anticipate what you'll face when you open a door. As in Figure 4 where Miku stands in front of the door to the Rubble Room, a bad omen can make you delay your entrance because you're terrified. And when you go into a room with the filament turned on, the situation can be just as suspenseful given that the spirits need to be located. You might switch to the viewfinder of the camera (Miku's weapon against spirits), but you still have to look often for spirits when the filament indicates their location by fading in and out. You have to know that the spirits move and conceal themselves, and that some are attacking while others don not. The Crucified Man in Figure 5 was nowhere to be found in the Naruki Shrine, for instance. You had to look up to the ceiling to find him.

Just after the prologue at the beginning of *Silent Hill*, Harry gets away with a good start when, in a cut-scene identical to the bus scene of *Cat People*, a first window bursts in the background creating a distraction and allowing a flying reptile to burst from the foreground window. From this moment on, Harry is in possession of a great forewarning tool: a pocket radio transmitting white noise that warns the gamer that one or many monsters are nearby (Figure 6). There are no visual displays on the screen of *Silent Hill*. The gaming experience of the whole series is driven by the terrifying static that comes to break the silence. Furthermore, the variations of white noise give information about the monsters and how far away they are.[8] During significant parts of *Silent Hill* and *Silent Hill 2* taking place outdoors in the mist, the duration of the

[8] In the Official Strategy Guide of *Silent Hill 3* [2], there is an explanation of how frequency, pitch and volume affect the radio. A chart gives the range of the radio for each monster: 14 meters for Double Head dogs, 20 meters for an enormous monster called a Closer, 15 meters for a zombie-like nurse, etc.

anticipation is actually extended compared to the parts that take place indoors. In the streets, the static fades in when you advance towards an unseen monster and fades out when you change direction to hurry away. It fades in again along with the monster's own noises when you cannot avoid a confrontation. Frequently this can last more than half a minute. You're always kept on your toes. What's more, when the radio begins to transmit noise and you cannot see outside the light beam of your flashlight, fear seizes you rather intensely.

Figure 6: Harry is off to a good start before getting the red pocket radio at the beginning of *Silent Hill* (Konami/Konami, 1999).

Tanya Krzywinska notes that "[m]any video games deploy sound as a key sign of impending danger, designed to agitate a tingling sense in anticipation of the need to act" [16: p. 213]. In that sense, though there is no specific device in *Resident Evil*, the game warns you in the most classical way by using off-screen sounds(similar to what happens in the underground locations of *Silent Hill* where the radio doesn't work). The moaning of the zombies and the shuffling of their feet indicate that they are nearby in a room or corridor. In fact, most of the time they are waiting just outside the frame, lurking to jump on you. But sometimes, they are farther away. For instance, if you do not move, it takes more than 15 seconds for the first zombie to enter the frame in a scene on the second floor of the Police Station in *Resident Evil 2* (Figure 7), and 5 more seconds for two more zombies to come.

Figure 7: Leon has been waiting long seconds for this first-heard zombie to enter the frame on the right in *Resident Evil 2* (Capcom/Capcom, 1998)

Other examples of this waiting occur on the first floor when zombies moan in an office you'll have to enter, and later on in the Vacant Factory where Leon and Ada hear an approaching zombie for no less than 40 seconds. Be that as it may, the forewarning does not rely only on this technique. Now and then, the search of a room is accompanied by typical, suspenseful extradiegetic music. *Resident Evil* also makes use of a few cutaways. Interestingly, it shows the impending attack of a Hunter twice. The first with a cut to an 18-second, fast traveling shot when this monster initially appears, and the second with a cutaway in the underground courtyard path (Figure 8).

Figure 8: A cutaway to the impending attack of a Hunter in *Resident Evil* (Capcom/Capcom, 1996).

There is also a cut in *Resident Evil 2* to what happens to a reporter (an NPC) in his cell which portends a frightening encounter.

To different degrees, all of the above examples put you in the state of uncertainty. Consequently, and most importantly, compared to the last forewarning of Cantor, Ziemke and Sparks' experiment which precisely described what would happen in the vampire movie scene, the sequences of survival horror games also elicit uneasiness about how uncertain the outcome is. You know that you'll have to face a monster, but you do not know how it will turn out. Not only your fright, but your anxiety as well is therefore intensified. Furthermore, as Torben Grodal stresses about video games, "suspense is interwoven with the interactive and repetitive nature of the game" [14: p. 206]. While aggressions, battles, mutilations and deaths remain final and unchangeable facts in a movie, in a game they are not. Events can be different or, at least, can be triggered in a different order. If you have killed the zombie-like native that was lying on the ground during a first exploration of the right branch of the village of *Fear Effect* (following Figure 3), it will not rise from under the frame the next time you go by. But if you come only once into the right branch, the lying zombie-like native will rise unexpectedly The opposite is also true: "What was surprising in the first game is transformed into a suspense-like coping anticipation in the following games. When the player advances toward the space/time in which the surprising event previously has occurred, say the sudden appearance of a fierce antagonist, it will induce an increased arousal" [14: p. 205-206]. Having to replay a game from the last save point in order to go back and face a boss monster that you have not yet defeated is a great forewarning situation. Replaying a game at the most difficult level, instead of at the normal one, also has the same consequences.

4. I'M SCARED

The connections of the aforementioned key elements of suspense and forewarning with horror seem obvious and definitely help to understand the gaming experience of survival horror games. However, it is necessary to highlight an important distinction between games and dramas or films which was an underlying principle in the preceding two parts of this paper. Though they are not addressing video games directly, Vorderer and Knobloch summarize the matter nonetheless: "According to the [Zillmann] disposition theory, a necessary condition for suspense is that the viewer witnesses the conflicting forces (…) without being able to intervene in the goings-on. If viewers could influence the plot, for example, the fate of the characters, their experiential state would change into actual fear or hope" [28: p. 64].

The spectator of a horror film and the gamer of a horror game are akin in the way that both are always aware that they themselves are not the victim of the monster's assault and that it is someone else doing the *suffering*. But while, ideally, their emotional responses run parallel to those of the characters, their way of feeling fear is different. In a horror movie, Carroll observes [7: p. 17], the emotional responses of the characters cue those of the audience. Both responses are synchronized. The characters exemplify for the spectator the way in which to react to the monsters by the reports of their internal reactions. In that sense, "one of the most frequent and compelling images in the horror film repertoire is that of the wide, staring eyes of some victim, expressing stark terror or disbelief and attesting to an ultimate threat to the human proposition" [quoted in Carroll, 7: p. 243 n. 45]. The spectator is consequently prompted to respond the same way. Often shown in close shots and in shot/reverse shot where both the point of view of the victim and that of the monster are shown, it is the spectator that is forced to witness these bloody confrontations. Furthermore, referring to Zillmann's necessary condition of suspense, the spectator has what Tan and Frijda call witness emotions [25: p. 52]. These emotions are related to Tan's fiction emotions mentioned in the introduction [26]. The spectator sojourns, in the imagination, in a fictional world where he can feel *as if* he were physically present, a world where he runs absolutely no risk. The emotional experience is based on a safe involvement. But since the significance of the fictional character's situation is relevant to his emotional response, the spectator has empathetic emotions. Feeling *with* the protagonist, he experiences empathic distress in seeing, for example, a babysitter terrorized by the idea that a monster is stalking around the house. But whatever happens, the spectator is forced to have an observational attitude, He is controlled by the filmmaker who guides him around as he pleases through the time and space of the fictional world. The spectator cannot participate in the situation. On the brink of finding the action too scary, he only can cover his eyes to defend himself against the horrible sights (though he still hears what's going on).

In a survival horror game, cut-scenes can depict a horrible scene in a filmic way. Since the plot is unfolding through those cut-scenes, it elicits fiction emotions. However, at the action level, a game does not rest on the reports of characters' internal reactions. The third-person perspective always shows the avatar in a long shot, and generally, in a long take, too. What's more, to face the monsters, the avatar is often seen from the back. With the exception of *Fatal Frame,* which shows a close shot of Miku in a short cut-scene before the attack of some ghosts (similar to the *Resident Evil*'s short cut-scenes showing the upcoming attacks of the hunter), there is generally no prior or subsequent reaction shot of a face expressing stark terror and attesting to the threat. Again, *Fatal Frame* allows you to switch to a first-person perspective through the viewfinder of Miku's camera and, in *Clock Tower 3* (Capcom/Capcom, 2003), the camera switches to a first-person view when Alyssa hides from the monsters. But in those case, the effect of the filmic subjective shot structure (which makes you feels *as if* you were in the situation of a character) is replaced by the sense of agency. Janet Murray has defined this characteristic delight of electronic environments in *Hamlet on the Holodeck. The Future of Narrative in Cyberspace*: "Agency is the satisfying power to take meaningful actions and see the results of our decisions and choices" [19: p. 126].You indeed control your avatar in the game-world (and the subjective point of view when it is the case), a control that leads to a mutation in the way you experience the scene.

It is certainly not the avatar that is meant to be scared in a survival horror game, but rather the gamer, i.e. you. If we can still refer to empathy since you experience emotions *with* an avatar, it is clear here that we cannot talk about identification with the character or about *becoming* the character in the game-world.[9] This is because the emotional state of that character is not identical to yours. When a monster bursts through the window, it makes you, not the avatar, jump.[10] Upon the sign of threat, the avatar does not express apprehension. When the visual warning system is displayed on the screen or the audio cues are extradiegetic, these signs are not for the avatar's benefit. Although the various avatars make themselves heard during their fight, scream when assailed and audibly breathe their last breath, they remain impassive on the action level. Whatever situation is faced in Silent Hill, Raccoon City or elsewhere, the avatars keep a "stone face" while responding to your actions. Instead, their reactions are behavioral and external. You are linked and synchronized with them physically. You see their actions and are made to feel their suffering not only as you see them being attacked, but also as you receive feedback from the Dualshock controller as in *Silent Hill* and *Fatal Frame*. Now a typical function found in many games, the controller vibrates every time your avatar is touched or hit. It vibrates throughout a confrontation in the Himuro mansion and goes very fast when touched by a spirit. In *Silent Hill*, to indicate avatars' health status, it also shakes more and more violently as they absorb more damage, echoing the acceleration of their heartbeat. This tactile simulation

---

[9] The notion of identification is not simple to deal with. Much has been written in film studies since its psychoanalytical description. It has been rejected, supplanted, revised and revived. To have a general view of the question, one can check the major literature about this notion.

[10] Manifestly, it is indicative that a gamer would say: "I was scared", not "My avatar was scared" when talking about what happened in a game.

focuses on physical strength for the simple reason that it helps you keep them alive. And that's another departure for video games. In movies, Carroll says, "the fear that the audience emotes with regard to the monster is not fear for its own survival. Our fear is engendered in behalf of the human characters in the pertinent films. We cringe when the Werewolf of London stalks his prey, not because we fear that he'll trap us, but because we fear for some character in the film" [5: p. 38]. Again, you do not fear for your own survival in a horror game either. However, in the game-world, since you merge with your avatar at the action level, and since your main goal is precisely to make him/her survive the threatening monsters, you're indeed made to be afraid that the monsters will trap you, in other words to fear *as if* you were in danger. This time, when the action becomes really scary, you can't simply cover your eyes. Holding your controller, your extradiegetic activity must be to try to overcome the diegetic situation of your avatar.

Fear — as the most commonly referred to emotion in philosophy and psychology, characteristic of an emotion prototype like Greg M. Smith remarks [24: p. 269 n. 4] — helps to distinguish the emotions generated by gameplay from fiction/witness emotions. For psychologist Nico Frijda, whose work has inspired Tan [26] and Grodal [14], emotions can be defined as "modes of relational action readiness, either in the form of tendencies to establish, maintain, or disrupt a relationship with the environment or in the form of mode of relational readiness as such" [12: p. 71]. Emotions are action tendencies. Given that fear is clearly object- and goal-oriented, it provides, as Smith notices once again, a strong action tendency. In the presence of a monster, fear urges you to act in one way or in another to disrupt the relationship. In a horror movie, when the hero/heroine is in danger, you cannot do anything but hope he/she will overcome the threat. Your action tendency is virtual. On the other hand, in survival horror games, you can do something. You can make your avatar act. You actually (even if it is related to a *virtual* game-world) have a repertoire of controls: draw and choose weapons, shoot, attack, guard attack, charge in, turn 180°, run away, use items to replenish life gauge, etc. Those actions give you gameplay emotions, emotions related to the ways you react to the situation. "Video games therefore", asserts Grodal, "simulate emotions in a form that is closer to typical real life experiences than film: emotions are motivators for actions and are labeled according to the player's active coping potentials" [14: p. 201].

5. I HAVE TO COPE

In following with the preceding comments, we have to agree with Grodal who has emphasized [13, 14] that the notion of coping is fundamental to the experience of video games. According to Susan Folkman and Richard S. Lazarus' definition, "coping consists of cognitive and behavioral efforts to manage specific external and/or internal demands that are appraised as taxing or exceeding the resources of the person. These cognitive and behavioral efforts are constantly changing as a function of the appraisals and reappraisals of the person-environment relationship, which is also changing" [11: p. 323]. The appraisals-reappraisals are very significant to determine the emotional effects of forewarning in survival horror games. Anticipatory fear is less important when on-coming monsters are directed by corridors or walks like the outdoors marked paths in *Fear Effect*, but it is greater when the monsters are free to move in the streets of *Silent Hill*. This fear is amplified when you can't clearly see around, specially when you find yourself in the scary places in the vein of the nightmarish world of *Silent Hill* where streets or floors are replaced by rusty grates over bottomless abysses, where walls and floors are splattered with blood and where you hear all kinds of industrial and creepy sounds. Then, we also have to take the way game designers can be playing with you into account. They are intentionally putting you in a state of terror. An example of this happened at one point during my experiential route of fear in *Silent Hill 2*. I was of two minds as to how to get out of the laundry room on the third floor of the Wood Side Apartments in the beginning of the game (Figure 9).The radio was transmitting white noise, but I could also hear footsteps of some sorts and what seemed to be the growling and shrieking of a huge monster. I was too scared to move. When I finally came out, I was very tense and anticipated an encounter with what turned out to be no more than a normal lone Patient Demon. The appraisal of the situation might have been different with a weapon other than just a wooden plank. I would have certainly felt more secure with a gun in my hand. But then, in *Resident Evil*, it is not a handgun, but a bazooka that you need in order to be at ease in front of a Hunter, a monster much faster and powerful than the zombies.

Figure 9: Too scared to get out of the laundry room of the Wood Side Apartments in *Silent Hill 2* (Konami/Konami, 2001).

To explain the relation between coping and emotion, Folkman and Lazarus distinguish two general and interrelated coping processes. The first strategies, called emotion-focused coping, are employed to regulate the situation causing distress. As Folkman and Lazarus talk among others about avoidant and vigilant strategies, another way to understand this is to refer to the degree to which individuals will either monitor (seeking information) or blunt (avoiding information) under threat [18]. Dispositional differences show that monitors (Miller is talking about high monitors/low blunters) scan for threat-relevant information and prefer to attend to information signaling the nature and onset of the shock as well as information about their performance when carrying out a task. Contrarily, blunters (i.e. high blunters/low monitors) tend to avoid informational cues and distract themselves from threat-relevant signals. Using this distinction to study the interaction between forewarning and preferred coping style in relation to emotion reactions to a suspenseful movie, Glen G.

Sparks discusses his findings: “Instead of an increase in negative emotion for all participants due to forewarning, the data indicate that forewarning may operate differently for individuals with different preferred coping styles. Monitors may actually prefer forewarning in order to cope with a scary movie, while blunters may prefer no prior information” [11: p. 337]. Although this experiment and the previously mentioned ones deal with the effects of forewarning outside the time flow of a film viewing and don’t set forth to explain the appeal of such suspenseful movies and horror movies (see studies of sensation seeking for more on this issue), the results still indicate that monitors have more intense emotional reactions when they are forewarned, while blunters do not. Monitors would probably prefer the warning system of a game like *Fatal Frame* which gives audiovisual cues as well as making the controller vibrate when a spirit appears. In fact, I refer to this distinction in order to explain why a few web reviewers have suggested turning off the radio in *Silent Hill* because it detracts from the surprise-factor of *Resident Evil*. The copying preference seems to be one explanation. At least, this demonstrates that, according to the type of gamer you are, the effects of a warning system will be different.

If the preceding remark proves to be questionable because emotion-coping strategies are, above all, used to deal with stressful events the outcomes of which are considered to be unchangeable, the second type of coping process stressed by Folkman and Lazarus is undoubtable. Called problem-focused coping, those coping strategies are directed at altering the situation that causes distress. They are used this time for outcomes that are amenable to change. Thus, in Grodal’s terms, you have active personal coping potentials in video games. And you will undoubtedly make use of them. Among the available types of control, you especially have behavioral control. With the Dualshock controller, you can change the actual terms of the person-environment relationship. A forewarning is an emotional cue, but also a cognitive cue for problem solving. Let’s quickly distinguish two forms of such coping. In survival horror games, a confrontational coping strategy that makes an individual fight back somewhat aggressively when facing a difficulty comes down to killing the monster. When you know that there is a monster nearby, you go to destroy it. This is how fearless gamers are likely to handle threats. In the other way, you can manage the situation in a more rational and planned manner. You appraise more consciously the magnitude of the threat before you face it. You then decide if it’s better to attack or to avoid and escape the monster. A timorous gamer can be expected to react in this way. In any case, the coping process can change through out a game. As Folkman and Lazarus point out: “[d]uring the anticipatory phase of the encounter, cognitive coping strategies can transform a threat appraisal into a challenge through their affect on secondary appraisal [during which you ask yourself what are your options for coping]” [11: p. 321]. One will agree that it is less stressful and much more fun to face a monster (and even more so a boss monster!) when you have the appropriate weapon, plenty of ammunition and first aid kits to recover from damage. It is also reassuring to know that you have mastered all the controls of a game and that you can move freely and (most importantly) quickly in the game-world. With all adequate coping resources, you can interpret the sign of a threat differently.

6. CONCLUSION

Because forewarning intensifies emotional reactions about upcoming frightening events and increases anxiety when there is still uncertainty about the outcome of those events, this paper should have ultimately prepared you to play your next survival horror game. Now it’s up to you to play and cope with your next ludic journey to hell.

ACKNOWLEDGMENTS

This paper has been written with the support of the Social Sciences and Humanities Research Council of Canada (SSHRC) and le Fonds québécois de la recherche sur la société et la culture (FQRSC).

REFERENCES

[1] Baird, Robert. “The Startle Effect. Implications for the Spectator Cognition and Media Theory”, *Film Quarterly*, Vol. 53, No. 3 (Spring 2000), 13-24.

[2] Birlew, Dan. *Silent Hill 3 Official Strategy Guide*, Brady Publishing, Indianapolis, 2004.

[3] Bordwell, David. “Cognition and Comprehension : Viewing and Forgetting in *Mildred Pierce*”, *Journal of Dramatic Theory and Criticism*, Vol. 6, No. 2 (Spring 1992), 183-198.

[4] Cantor, Joanne, Dean Ziemke and Glenn G. Sparks. “The Effect of Forewarning on Emotional Responses to a Horror Film”, *Journal of Broadcasting*, Vol. 28, No. 1 (Winter 1984), 21-31.

[5] Carroll, Noël. “Film, Emotion, Genre”, in G. Smith and C. Plantinga (eds.), *Passionate Views : Film, Cognition and Emotion*, Johns Hopkins University Press, Baltimore, 1999, 21-47.

[6] Carroll, Noël. “The Paradox of Suspense”, in P. Vorderer, H. J. Wulff, and M. Friedrichsen (eds.), *Suspense: Conceptualization, Theoretical Analysis, and Empirical Explorations*, Lawrence Erlbaum Associates, Mahwah, N.J., 1996, 71-90.

[7] Carroll, Noël. *The Philosophy of Horror or Paradoxes of the Heart*, Routledge, New York, 1990.

[8] Carroll, Noël. “An Alternative Account of Movie Narration”, *Mystifying Movies. Fads & Fallacies in Contemporary Film Theory*, New York, Columbia University Press, 1988, 170-181.

[9] Comisky, Paul, and Jennings Bryant. “Factors involved in generating suspense”, *Human Communication Research*, Vol. 9, No. 1 (Fall 1982), 49-58.

[10] de Wied, Minet, Kathleen Hoffman and David R. Roskos-Ewoldsen. “Forewarning of Graphic Portrayal of Violence and the Experience of Suspenseful Drama”, *Cognition and Emotion*, Vol. 11, No. 4 (1997), 481-494.

[11] Folkman, Susan and Richard S. Lazarus. “Coping and Emotion”, in N.L. Stein, B. Leventhal and T.Traboass (eds,), Psychological and Biological Approaches to Emotion, Lawrence Erlbaum Associates, Hillsdale, N.J., 21990, 313-332.

[12] Frijda, Nico H.. *The Emotions*, Cambridge University Press, Cambridge, 1986.

[13] Grodal, Torben. "Stories for Eye, Ear, and Muscles: Video Games, Media, and Embodied Experiences", in M.J.P. Wolf and B. Perron (eds.), *The Video Game Theory Reader*, Routledge, New York, 2003, 129-155.

[14] Grodal, Torben. "Video Games and the Pleasure of Control", in D. Zillmann and P. Vorderer (eds.), *Media Entertainment: The Psychology of Its Appeal*, Lawrence Erlbaum Associates, Mahwah, N.J., 2000, 197-213.

[15] Hoffner, Cynthia, and Cantor, Joanne. "Forewarning of a Threat and Prior Knowledge of Outcome. Effects on children's emotional responses to a film sequence", *Human Communication Research*, Vol. 16, No. 3 (Spring 1990), 323-354.

[16] Krzywinska, Tanya. "Hands-on Horror", in G. King and T. Krzywinska (eds.), *ScreenPlay. Cinema/videogames/interfaces*, Wallflower, London, 2002, 206-223.

[17] Lazarus, Richard S.. "A Laboratory Approach to the Dynamics of Psychological Stress", *The Amercian Psychologist*, No. 19 (1964), 400-411.

[18] Miller, Suzanne M.. "Monitoring and Blunting: Validation of a Questionnaire to Asses Styles of Information Seeking Under Threat", *Journal of personality and Social Psychology*, Vol. 52, No. 2 (1987), 345-353.

[19] Murray, Janet H.. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*, The Free Press, New York, 1997.

[20] Nomikos, Markellos S., Edward Opton, Jr., James R. Averill, and Richard S. Lazarus. "Surprise versus suspense in the production of stress reaction", *Journal of Personality and Social Psychology*, Vol. 8, No. 2 (1968), 204-208.

[21] Perron, Bernard. *Silent Hill. Il motore del terrore*, Edizioni Unicopli, Milan, to be published in Fall 2004.

[22] Rockett, Will H.. *Devouring Whirlwind. Terror and Transcendence in the Cinema of Cruelty*, Greenwood Press, New York, 1988.

[23] Smith, Greg M.. *Film structure and the emotion system*, Cambridge University Press, Cambridge, 2003.

[24] Smith, Greg M.. "Local Emotions, Global Moods, and Film Structure", in G. Smith and C. Plantinga (eds.), *Passionate Views: Film, Cognition and Emotion*, Johns Hopkins University Press, Baltimore, 1999, 103-126.

[25] Tan, Ed S. and Nico Frijda. "Sentiment in Film Viewing", in G. Smith and C. Plantinga (eds.), *Passionate Views: Film, Cognition and Emotion*, Johns Hopkins University Press, Baltimore, 1999, 48-64.

[26] Tan, Ed S. *Emotions and the Structure of Narrative Film: Film as an Emotion Machine*. Lawrence Erlbaum Associates, Mahwah, N.J., 1996.

[27] Truffaut, François. *Hitchcock*, Simon & Schuster, New York, 1985.

[28] Vorderer, Peter, and Silvia Knobloch. "Conflict and suspense in drama", in D. Zillmann and P. Vorderer (eds.), *Media Entertainment: The Psychology of Its Appeal*, Lawrence Erlbaum Associates, Mahwah, N.J., 2000, 59-72.

[29] Vorobej, Mark. "Monsters and the Paradox of Horror", *Dialogue*, No. 24 (1997), 219-249.

[30] Zillmann, Dolf. "The Logic of Suspense and Mystery", in J. Bryant and D. Zillmann (eds.), *Responding to the Screen. Reception and Reaction Processes*, Lawrence Erlbaum Associates, Hillsdall, N.J., 1991, 281-303.

[31] Zillmann, Dolf. "Anatomy of suspense", in P. H. Tannenbaum (ed.), *The Entertainment Functions of Television*, Lawrence Erlbaum Associates, Hillsdale, N.J., 1980, 133-163.

Media Pop Art – Electronic Art as Satire on Everyday Life

Robert Praxmarer
Artist
www.servus.at/cubic
Linz / Austria
cubic@servus.at

ABSTRACT
In this paper I will give an outline of my practice as a young media artist and introduce some of my strategies, thoughts and works. I will try to formulate my subjective view of media culture and media art with a special focus on game art . On the basis of some illustrative works I will introduce the term Media Pop Art, and highlight the importance of socio-political aims in my works.

Keywords
Game Art. Media Art. Pop Art. Activism. Installation. Interactivity. Socio-political Art.

1. INTRODUCTION
My first encounter with a computer was at the age of ten. This early computer (Amstrad CPC 6128) was relatively powerful for the time but not supported by many programmes or games. Thus I soon started programming my own little games and drawing programs. I was overwhelmed by the possibilities it gave me on the one hand but was often frustrated by the strict rules and difficulties writing code that would run as expected. Luckily however I grew up in the 2D era, where I was satisfied programming circles which had to shoot at boxes. The kids growing up now can choose from a myriad of games incorporating complex stories and puzzles coupled with realistic graphical representation of the real world. These kids grow up with these advanced media forms, frequently consuming them in an uncritical way. I see my role as an artist to work with the gaming media, analyze, deconstruct and reconstruct it [6]. In the next section I will write about my project (t)Error, which was my attempt to use the medium computer game to transport a protest message [7]. (t)Error is exhibited at the COSIGN 2004 Art Show. After that I will shortly review the projects Orgasmatron, StencilU and City Obscura and on the basis of these works I will try to formulate my ideas and thoughts about media culture.

2. (T)ERROR – A MIXED REALITY SATIRE ABOUT WORLD POLITICS AND THE GAMING INDUSTRY
Whom would you like to play to get the world back into shape? George W. Bush, Osama Bin Laden, or rather Tony Blair? Starting the computer game (t)Error, the player is spoilt for choice. A camera tracks the player's movements and a projector throws the players outlines in original size on a white wall in front of him, all in realtime (see Figure 1).

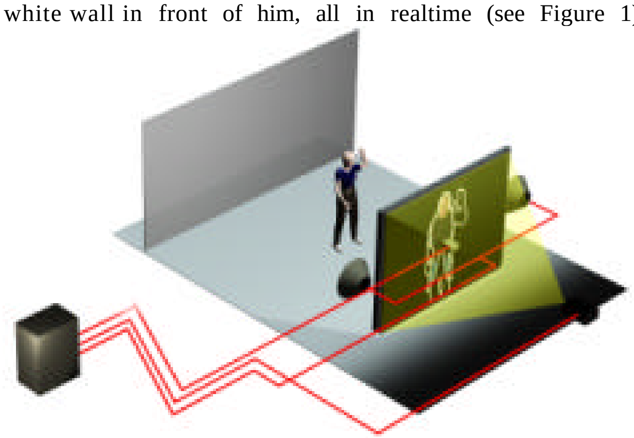


Figure 1. Schematic of the installation

When the player kicks, his shadow will kick. When the player ducks or jumps to avoid bumping into an enemy jet, his shadowy avatar will do exactly the same in real time. The empty silhouette of the player gets filled out - and this is the players choice – with the face of George W. Bush, the flag of a state which supported the Iraq war, or the body of Osama Bin Laden (see Figure 2,3). For gaining a good game score, you will have to fight evilly. Being slipped into this stereotype role, you have to avoid tanks, submarines and fighter planes, you have to collect oil pumps and dollar notes, and you have to kick civilians begging for help. If you succeed in all these matters, you are the lucky winner and rise to the next game level.

Figure 2. Screenshot of (t)Error

First published at COSIGN 2004,
14 – 16 September 2004, University of Split (Croatia)

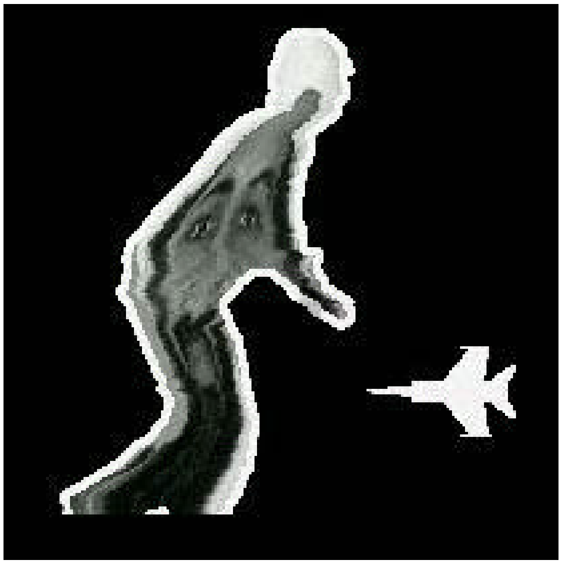


Figure 3. Another screenshot of (t)Error

I always felt that at least for me there is a lack of meaning in a lot of media pieces. I decided to express my feelings about the war in Iraq with a protest computer game and tried to use this genre as a protest medium. In former times a revolution started with a song, in our time console kiddies play games, it's as easy as that. I see a transition in how information and protest works for the younger generation. Fakes, Adbusting und Semiotic Sniping are the ways to express your feelings in a world ruled by global players. The logos and signs of this companies can be found everywhere. Resistance is futile ? Not if you undermine their symbolic system, fight with their weapons and use corporate disinformation. That’s what I try to do with my art. I’ll define this kind of work as Media Pop Art. It takes well known objects, structures, medias, etc. and sets it in a different and sometimes funny context. For me that’s more interesting than to make a serious art work, which in the end doesn’t reach anyone, except some specialists [6].

2.1 (t)Error criticizes the Iraq war

„Art is the cry of distress uttered by those who experience at first hand the fate of mankind.“ [12]

This interactive video installation was my reaction on the Afghanistan campaign and the following Iraq war. It is a political persiflage on world politics, a mixed-reality game, which playfully tempts - through the whole body interaction - to participate, but at the same time the not very playful content of (t)Error wants to provoke reflection and self-reflection.

This play with the aspects of fun and seriousness or game and reality forms the basis of my work and media analysis.

Laughter lures the visitors, and not only an elitist gallery audience, but also the people from the street, who do not know too much about contemporary art. Fun, games and interaction cause motion, it is then much easier to get involved in the game, and so the change from laughter to sadness is more powerful and obvious.

2.2 (t)Error criticizes the gaming industry

To transform this combination of laughing and crying into the world of computer games, and so into the world of our children, was my main aim.

“Fifty percent of all Americans age six and older play computer and video games.” [14]

„In 1998, over one in four American youngsters reported playing games between seven and 30 hours a week. More than one in four homes has a game console.” [4]

This is one of the reasons why I like to show my installation at parties and clubs or in public space (see Figure 4) [10]. This year at the beginning of april, for example, (t)Error was exhibited in Berlin, in the “Club zur Möbelfabrik”. The young audience of this venue participated with enthusiasm. The reflection hit them a couple of minutes later, when the youngsters squatted down besides the installation, tired after the exhausting movements.

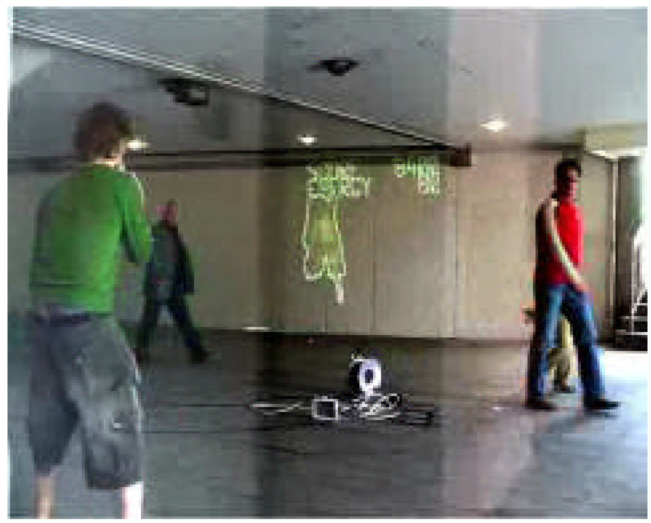


Figure 4. (t)Error as public space intervention

“I just played Osama Bin Laden in the game. I was killing people. Was that fun?” And this is what makes the artistic examination of computer games so precious and powerful: The sudden liberation, which is triggered by the dramaturgy of computer games, the role games par excellence. Nowhere else it is as simple and as easy to slip into the role of authorities, to play with the overthrow of established social systems and experience the dynamism of action and reaction without any danger. At the same time, (t)Error provides its players with a strange closeness and directness: It’s not a joystick that helps our avatars to jump hundred meters high, but everyone is actively playing him/herself.

(t)Error is an attempt to dismantle the gaming industry and the stereotype roles and behavior patterns taught by it. All these are the semiotics our children are confronted with since birth. We need to analyze them to get to know our children’s world.

The simple, pixel graphics of my game was developed following early, famous computer games like for example Space Invaders or Defender.

(t)Error creates a manageable (art) space, a system of signs and symbols, which are well known by the followers of computer games, who always associate these sign systems with fun and leisure time. Society codes and conventions which everyone takes for granted, can now be seen from a different perspective and open up new possibilities for dialogue and debate.

2.3 (t)Error criticizes the media society

Our contemporary life and society consist of brainwashing media flashes, consumption and pseudo-happiness paired with a upcoming depression. My work uses the same media and the same mechanisms however the goal is the deconstruction of this failed system and a reflection of this media junked fake.

Just the unbroken popularity and fascination of computer games speaks for an artistic analysis. There is a whole "culture of external activities that has developed around these games – from trading of customized game patches to demands that the manufacturer provide more levels, to whole web sites devoted to one's avatars [3]. In consequence of her examination, Antoinette La Farge stresses that in the near future, computer games will be as important in the modern cultural landscape as art, music or film.

2.4 (t)Error criticizes the abuse of computer games through military and economy

Another important point of criticism shown in (t)Error lies in the close connection of computer games with military and economy.

„It's now fairly common knowledge that Operation Desert Storm was prepped for by doing simulation strategy exercises down in Florida, and that the US military is currently pumping large amounts of capital into figuring out how to appropriate gaming principles for battle training in massively mulituser SimNet environments." [4]

This list of strange simultaneities and surprising cooperations could go on and on and on.

"According to the Interactive Digital Software Association's 1999 sales information, over 6.1 billion dollars in U.S. entertainment software sales in 1999 went to games." [4]

In a time where professional computer games have more production budgets as Hollywood films, you must question this media, analyze it artistically and bring it up in a topic. A whole generation grows up with games, mostly just having one goal: "You achieve your aims with violence."

Artists and creative must click in at this point, dismantle these games, rearrange them with new ideas and reflect the media itself. The transformation and modification of games are also interesting in the historical context, if you consider that military institutions have taken a great part in the development of this gaming technology. With the second version of a modified 3D Shooter the American Army is at the present time hunting for new recruits; let's hope that this shot was off target.

3. ORGASMATRON

Orgasmatron is an interactive sound installation. It is designed as a glass piano (see Figure 5) [8]. When you move your hand over a glass the vibrator next to a glass starts and generates a certain sound. All the glasses are filled with a certain amount of liquid which so build up a scale. A camera from above detects which glasses are triggered. This installation has been exhibited at the Ars Electronica Pixelspaces Exhibition 2003.

Figure 5. Orgasmatron

I used real sex vibrators and put them in another context. It is my reaction about the floods of spam mail about sex, porn, etc. Towards a sex thrilled society, viagra, penis enlargement, porn. Sex sells it's a simple equation. Orgasmatron is my persiflage of a culture which certainly thinks more about sex then about Mozart. I tried to promote the installation with fancy flyers and in a pop cultural way, to make it look like a clubbing, which worked very well (see Figure 7).

Figure 6. Orgasmatron Flyer

Figure 7. Anonymous with Flyer in his pocket

I printed 700 flyers and they were all gone in 4 days. I don't think this is because people in Austria are so much into media art, it's more about using the references and the same language as youth culture.

4. STENCILU

StencilU is a quite simple but powerful realtime visualization of time and personality [9]. It is an interactive installation, where your silhouette is captured with a camera (see Figure 1) and the frames of the silhouette are made smaller and arrange around the actual silhouette. The small silhouettes arranged are the different time frames. With this technique the motion of yourself floats around as small representations around your big realtime shadow (see Figure 8). Accordingly to the speed of the movements of the user the visual output is inverted, this gives a stroboscope effect, which resembles a club aesthetic and a motion blur is switch on randomly which makes the visualization even more club like.

Figure 8. First version of StencilU

The project was developed at Schmiede 04, a media gathering of vjs, djs, audio producers, coders, hacktivists, sprayers, etc., in one word young individualists who have something to say and do it with their possibilities [13].

The aesthetics in black and white look like pictures in the sprayer/stencil community and therefore I came up with the idea to photograph street art (graffiti tags and stencils) and integrate it in the project (see Figure 9). So your stencils are combined with the street art stencils, respectively icons (communist star, guns, bush, etc.) and result in the visual output.

In this project I used the language of the street and the signs which can be found out there which is for me a visual short story of contemporary youth culture, so it's a very interesting project for me and I would be interested to make it available to everyone with a web cam.

Figure 9. StencilU enhanced with street stencils

5. CITY OBSCURA

City Obscura is an interactive video installation which deconstructs and then reconstructs the urban space and architecture of the city of Judenburg, and was exhibited in July 2004 at the Liquid Music Media Art Festival in Judenburg, Austria [11].

A wide range of different shots of the city's architecture, people and crucial places awaits the visitor's eye. When the spectators start to move in front of the video screen, their movements get tracked and these gestures trigger a process of mingling the digital data streams of the videos itself. You can see the pixels as tiny bits and pieces of colored blobs, like a snow storm, revolving around your own shadowy figure on the video screen and following you wherever you go within the installation. The recipients can be seen as planets which have a gravitational force on the videos (see Figure 10).

Figure 10. Distorted video of the city of Judenburg

When the visitor stops to move around and stands still, the whirring pixels start to compose themselves again to the video shot the visitors have already seen before they started to walk around – but slightly changed

For example, one of the original videos shows us the view of one of the main streets of Judenburg, cars passing by, residents walking off doing their daily shopping. After the deconstruction through the gallery visitors, this video emerges again out of the chaotic data streams – but now the river of Judenburg – the Mur – is running through this main street, water instead of asphalt. A new utopia of Judenburg has developed.

So this points out quite obviously that we only worked with footage we found in and around the city of Judenburg. This was one of the main goals we had to achieve to stress one central thought concerning urban architecture: It is always the residents who create the architecture they are using daily. But there are not too many people who realize the important role they play in generating the infrastructure and atmosphere they are living in and what they can do themselves to make it either better or more functional or worse. This installation attempts to allow people to play with their own surroundings, and this process should lead to a critical consideration of their home town's images and to a reflection of the good and bad parts of the place they are living in. The installation also emphasizes that there is always more than one possibility to design urban space. Architecture grows through the interactivity of its residents - so the real process in everyday life is analogue to the way the installation works. The mixing of the video's digital data streams has its equivalence in the "Brownian motion". – The more movements are generated, the more the particles begin to chaotically move. This principle of chaos and order builds the basis of City Obscura. The Second Law of Thermodynamics can be seen as a consequence of probabilities, which is our metapher for our work. After deconstruction and reconstruction a new, a different structure arises (see Figure 11).

Figure 11. Melting the video material through movement

On the residents of Judenburg, this installation had a very subtle impact as they knew both, the original video of course, but also the new parts of the altered one, and so they saw different parts of their home landscape in a flawless, but newly changed entity, according to the principle of déjà vu. Out of this déjà vu-effect, the artists create a new consciousness of the city in the heads of their spectators and support a critical occupation with something, that was, before, taken for granted and which was, until now, seen as something well-known and established and unchangeable. The main aim of City Obscura is so to show a city – and this can be any city – in its flexibility and mobility. Stand-still means there is no chance to change and no possibility to alter the future. And this installation tries to show us – with its particular interactivity mode – that it is us, the residents, who can feel free, and who have the power, to alter the infrastructure we have to deal with daily, but also to get a different perspective on things we perhaps don't appreciate anymore as we see them day by day.

City Obscura was a more subtle work because the people the piece was intended for were not really into media art. But for us it was clear that the work did its job, when my collaborator Reinhold argued with an 65 year old lady about computer-based art and she thought that she is too old too understand it and there is no meaning and suddenly she saw the city, she has seen so often, differently, and then she was quiet, looked to us and said that she now is able to understand the work, which was one of the best compliments I ever got for a work.

6. MEDIA POP ART - A WAY OF CREATING SOCIO-POLITICAL ART

As the works above show, I try to use the signs and the language of pop culture and everday's life . I believe that with this way, I am able to reach more people and so my works are more likely to have an impact on society.

"Joseph Beuys is perhaps the best post-1960s example of an artist whose work changed society. Rather than adopting a perception of art that is formal and aesthetic only, his concept of social sculpture includes the kind of human action that is directed at structuring and shaping society - Beuys calls it the social organism. When seen in this way, art is not just a material artifact: it is also, and above all, action designed to have social consequences. Beuys' idea of relating plastic creativity to socio-political activities took up the social utopias of the historical avant-garde. Beuys was not primarily interested in including and using media in this context, but many post-1960s artists have both addressed media explicitly, and also used them to pursue concrete socio-political aims. They start by assuming that in a society increasingly influenced by media, an (artistic) change of media content or media structures can contribute significantly to democratizing society. And ultimately, behind this idea there lies the hope that art can change society. " [1].

Guy Debord among others formulated a radical social critique and also a radical critique of media society. It rejected creating aesthetic objects in favor of socially constructing situations aimed not at art but at life." [1]

"The demand not to limit oneself to producing works of art any more, but to raise artistic practice to the level of the technological possibilities offered by modern industrial societies." [5] This demand of Guy Debord is met by modern Computer Based Art. My demand though is that the produced art should use the technology and media with a sensible understanding of how to reach people and provoke the reactions you intended. "Form follows function" is not only the famous Bauhaus motto, in my work it is a credo as well. If I want to communicate something to kids, I will use their medium the computer game.

“As a rule, media installations work as a link between a (media) view of the public sphere and a subjective view of the world.” [2]

7. CONCLUSION

Media Pop Art has some advantages and some drawbacks. One of the benefits for young artists is that you can reach a lot of people with even a simple piece, but for that it must be easy ascertainable. There the drawback arises, that curators and media theorists critic this kind of work whether as immature or not subtle enough.

Another advantage what makes this kind of work so tempting for me is that you recycle your own or other works and make different variations with the same piece, get it a new meaning or looking [9]. Media Art is connected with a lot of work, at least if you belong to the group who is programming their art themselves. Each piece takes you a long time planning, programming, tuning, etc., so it’s quite reasonable to recycle the algorithms you’ve already created and vary the content.

For me there is just no other option to make my art and get some attention for it. Most of the museums are far too conservative to commission computer based art [2] so as a young unknown media artist you have to draw some attention on you and your work, by embedding it in the pop cultural context. Another important point is that I grew up with pop culture and I try to reflect the media and the culture behind it with my works, so therefore my pieces often can be interpreted as media or social critique.

8. REFERENCES

[1] Arns, I. Social Technologies Deconstruction, subversion, and the utopia of democratic communication, 2004. http://www.medienkunstnetz.de/themes/overview_of_media_art/society/1/

[2] Frieling, R. Form Follows Format - Tensions, Museums, Media Technology, and Media Art.

[3] LaFarge, A. An Introduction to the Convergence of Computers, Games, and Art. http://beallcenter.uci.edu/shift/essays/lafarge.html

[4] Nideffer, R. F. Mediating the process of academic exhibitionism. http://beallcenter.uci.edu/shift/essays/robertn.html

[5] Ohrt, R. Situationistische Internationale. *DuMonts Begriffslexikon zur zeitgenössischen Kunst,* 273.

[6] Praxmarer, R. Game Over Mr. President: Artists discover computer games as a protest medium. RebelArt Magazine #1, 78-87.

[7] Praxmarer, R. (t)Error – A mixed reality persiflage on world politics. http://www.servus.at/cubic/(t)error.htm

[8] Praxmarer, R. Orgasmatron – Interactive Glass Piano. http://www.servus.at/cubic/Orgasmatron.htm

[9] Praxmarer, R. StencilU http://www.servus.at/cubic/StencilU.htm

[10] Praxmarer, R., Bidner R. Fun Anyone ? – A public space intervention with (t)Error. http://www.servus.at/cubic/fun_anyone.htm

[11] Praxmarer, R., Bidner R. City Obscura http://www.servus.at/cubic/city_obscura.htm

[12] Reich, W. Arnold Schoenberg: A Critical Biography. Praeger, New York, 1971

[13] Schmiede Media Conference 04 http://www.c-a.at

[14] Statistics about gaming. http://www.theesa.com/pressroom.html

Models for Digital Storytelling and Interactive Narratives

Léonie Schäfer

Fraunhofer FIT
Schloss Birlinghoven
53754 Sankt Augustin, Germany
leonie.schaefer@fit.fraunhofer.de

ABSTRACT

Digital storytelling represents a challenging way of creating a bridge between traditional oral storytelling and the form of communication adopted by modern media and information technology. In this paper I introduce a method for categorization of applications and outline a reference model in digital storytelling. The model will be useful for the analysis of existing, and the development of future applications. In this paper I present three showcases in Digital Storytelling and employ the reference model for analysis.

Keywords

Digital Storytelling, narrative, reference model, team awareness, edutainment, collaboration

1. INTRODUCTION

Digital Storytelling, Interactive Narratives - at a first glance it seems obvious what is to be understood by these terms. However, a closer look reveals a great variety of applications, which differ widely in the way they use, create or tell stories:

In Conversational Storytelling the user interacts with a virtual counterpart in a storytelling context [1, 2]. This approach combines non-verbal communication and discourse management with narrative structures and often employs mixed-reality user interfaces. Story structures find application in areas otherwise unrelated to storytelling. In Information Visualisation, story structures serve as a new form of interface metaphor. Digital storytelling serves as a means to structure and convey huge amounts of data, which otherwise would be difficult to understand and to process at one glance [3].

The advance of digital contents and the development of new types of user interfaces, e.g. in Virtual Reality and Mixed Reality, facilitates new ways of creating and experiencing stories. In these settings, stories emerge through interaction with the application, either on purpose, if the application targets at the creation of stories, or as a side-effect, e.g. in games where the focus is set on playing rather than on following a story. The area most commonly associated with digital storytelling is the field of dynamically generated stories. Storytelling applications in this category aim to entertain the user [4] or have an educational purpose [5]. They target at all age groups and use various types of user interfaces to enable user interaction with the story content.

Digital media material, stored in media repositories like digital photo archives, provides a source for a diversity of stories. The advance of digital photography, video and image processing techniques results in the creation of private media archives to share with friends. The stories inherent in media repositories are not visible at first glance but emerge by interacting with the media content. For example, browsing pictures with friends might result in a story about the people shown on the picture or on the occasion when the picture was taken. The advance of digital media allows a combination of several types of media, e.g. photography with other media types as sound or film.

These are, of course, no definite boundaries for the categories in storytelling. Applications might belong to several categories, depending on their emphasis and components. Each of the categories described provides its own definition of storytelling and its own set of parameters which categorize the application. But what defines a storytelling application in general? Which are the parameters required to categorize an application? Which factors are important for its success? Researchers have taken an overall view on the nature of storytelling[6] or have argued about differences, e.g. between games and storytelling[7, 8]. The discussion on the conflict between story coherence and user interactivity prevails in many publications[9, 10]. However, the question which guidelines to apply in the development of future storytelling applications still remains unresolved.

In this paper I will give an overview of the field of digital storytelling and its major application areas. I will outline and discuss parameters which appear of relevance for a wide range of digital storytelling applications, e.g. the structural degree of the story, the level of control and interactivity, the continuity of the plot, the possibilities for collaboration, the degree of virtuality and the cognitive effort for the user to create a coherent storyline. I will analyse three storytelling examples in order to develop a general model for digital storytelling which can be used as a framework useful for the future development of narrative applications. PhotoStory, StoryNet and DocuDrama, the three example applications presented in the following, relate to each other in their focus on dynamically generated stories, rely on user input but are generated without direct user interaction.

2. RELATED WORK

New media forms and especially Interactive Narratives have become a very popular subject of research. Several people have approached the topic from different angles. Especially the relation between games and narratives has raised much discussion. In the following section I intend to give an overview on different approaches to provide categories, classifications or conceptual frameworks on storytelling and related areas, such as games or collaborative virtual environments (CVE's).

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

Fritz & Fehr have set their focus on games in an educational content. In [11], they present a catalogue of categories to be used for the analysis of edutainment applications. This catalogue combines general questions on the appearance and technical quality of the game with estimation of the quality of the gameplay, analysis of the game effect and dynamics, and an assessment of the pedagogical quality of the game. The catalogue of categories aims to provide a means of making detailed and comparable assessments of computer games. Stauber[12] employed these criteria to analyse a series of games with educational purpose.

Lindley also takes a look at the relationship between narratives and games. He analyses the temporal structure in games and narratives, presents a game/narrative model consisting of several layers and correlates each game level to a narrative level[13]. He describes the user interaction in games as Gameplay Gestalt, a unity of actions which combine perception, cognition and motor interaction. This is opposed to Narrative Gestalt, a pattern of actions, which enables the user to combine story elements into a coherent story.

Manninen examines interaction in collaborative virtual environments which, in his opinion, share many aspects with multiplayer computer games [14]. He describes and classifies different interaction forms in a scheme, e.g. Dramaturgical Interaction as the presentation of the user in a public environment, or Communicative Interaction for finding consensus between participants. The interaction form model aims to provide a conceptual framework for the analysis, evaluation and design of multiplayer games.

Mateas presents a character-based approach in extension of Aristotle's model of drama[15]. Using Aristotle's theory as a basis, he examines Laurel's definition of properties unique to dramatic stories, and discusses Murray's three categories relevant for the analysis of interactive storytelling applications. Laurel defines Enactment, Intensification and Unity of Action as properties of importance [16]. Murray proposes three categories for analysis: immersion, agency and transformation [6]. In his paper, Mateas presents a model based on Aristotle's theory, which has been extended by an interactive component on the level of characters. Façade, a storytelling application developed by Mateas and Stern [17], presents a system which builds upon this model.

Spierling et. al. [1] takes a closer look at the author of interactive storytelling applications. She presents an architecture with four hierarchical levels for authoring. The levels each provide a different degree of agency for the user in the development of the story. At each level the architecture consists of an engine and a corresponding model, e.g. story engine and story model. The engine is responsible for driving the action on that level, the model contains rules which defines this procedure.

In Rules of Play [18], Salen and Zimmerman offer a conceptual framework for game design. They view games from different perspectives, e.g. on Games as Narrative Play.

3. CATEGORIES

Depending on the definition of Digital Storytelling, there is a wide range of applications which belong to this research area. The variety of different storytelling applications is reflected in the variety of definitions of narrative and storytelling. Lindley states that "*In it's broadest sense, a narrative may be regarded as an experience in time that has some kind of feeling of unity and integrity*." [19]. Spierling et. al. states that *"Interactive story telling instead relies on a predefined story, a specific plot concerning facts and occurrences. Only the telling of the story is done interactively."* [1]. Salen and Zimmerman provide the reader with following definition *'A game is a narrative system in which the narrative experience of the player arises out of the functioning of the game as a whole'* [18] (p.419).

I define the term Digital Storytelling as valid for all types of applications which use digital media either to support, to enable the creation or to generate stories. The resulting stories might be told orally or presented with the use of digital means. Figure 1 presents storytelling categories along a vector. The position on the vector defines the degree of oral contribution to the storytelling process. For example, Media Repositories are positioned on the left side of the axis. They provide digital content and user interfaces which enable the creation of stories. The stories themselves are mostly told orally. Dynamically generated stories are positioned on the right side of the axis. They are based on digital content as well, but the story is presented digitally by the storytelling application. Zimmerman regards the terms games, play, narrative, and interactivity as "four concepts, each concept overlapping and intersecting the others in complex and unique ways" [20]. The storytelling categories presented above should similarly be regarded as concepts rather than categories with strong boundaries. Applications seldom belong exclusively to one category, but can be grouped according to their emphasis in storytelling.

Figure 1: A vector of storytelling categories

4. DIMENSIONS

The previous section gave an overview of the wide range of storytelling applications. However, all storytelling applications have factors in common. Digital stories are based on some type of story material. This can be abstract data but also a complete story written in prose. Stories usually follow a conceptual structure, which depends on the consistence of the material they are built upon. The degree of conceptual structure shows consequences in the story's coherence and continuity in the flow of the story. It also affects the cognitive effort required in order to create a mental version of the story. The presentation of digital stories varies according to their degree of spatiality and virtuality. Spatiality defines in how far objects in space, space itself and navigation are of relevance in the evolvement of the story. Virtuality defines the degree of involvement of a virtual environment in the story development. Also of interest is the degree of collaboration between users as part of the story experience, the degree of control the users have over the flow of events, and the level of interactivity the application allows. Finally, immersion and suspension of disbelief are factors which reflect how digital stories appeal to the user.

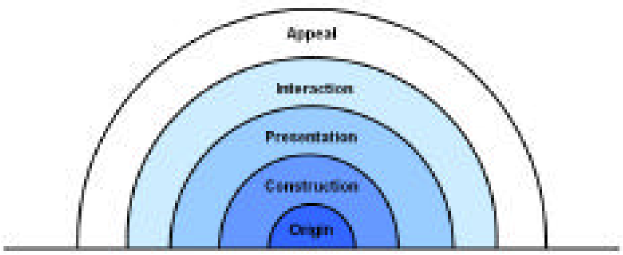


Figure 2: Layers

The dimensions mentioned above can be grouped in five different layers. Figure 2 gives an overview. The innermost layer comprises dimensions which apply to the story material. A layer on story construction forms the next level. On top of that follows a layer which examines the way the story is presented. The outermost layer comprises dimensions which examine the effect the application has on the user. The following section details the different layers and their dimensions.

The dimensions star, see Figure 3, displays all dimensions at one glance in the form of a star. Each axis of the star represents one dimension. The dimension star is a reference model for the analysis of three different applications in digital storytelling.

5. SHOWCASES

5.1 DocuDrama

DocuDrama forms a part of the project Theatre Of Work Enabling Relationships (TOWER). The idea of the TOWER project was to offer to members of a virtual team a platform which would make social encounters possible and provide team awareness [21].

Awareness of the activities of other team members affects the quality of work and provides a positive feeling within a group. Virtual teams which do not work in the same location but at geographically different places usually do not have such a close relationship, since they very seldom meet in person. A variety of tools developed in TOWER offers the user the possibility to become aware of the activities of team members in the work context and to start communication. The TOWER tools provide synchronous awareness about team activities; DocuDrama as part of TOWER offers asynchronous awareness. DocuDrama Conversation employs the TOWER virtual environment as a stage for a replay of activities.

Figure 3: Dimensions Star

Events presented in DocuDrama result from user activities in a BSCW collaborative workspace[22], a team workspace which serves as platform for the collaborative work on project documents. User interactions on documents are stored in a log-file. This data provides the base material for DocuDrama replays.

DocuDrama [23] was developed in the form of three different approaches. The version presented and analysed here is DocuDrama Conversation, which was developed at Fraunhofer FIT. DocuDrama Conversation [24] tells stories about collaborative activities in a virtual workspace. The focus of DocuDrama Conversation is on interaction between users of the workspaces. The two other approaches in DocuDrama focus on the history of documents (DocuDrama Timetunnel) and the progress of a project in general (DocuDrama Project).

The virtual environment of DocuDrama Conversation presents the data-structures of the collaborative workspace in a symbolic form. Geometrical objects in form of coloured boxes present different directories. Sign posts which symbolize the name or topic of the workspace serve as landmark for orientation in the virtual environment. The objects which belong to the respective workspace are placed around the sign posts. Avatars moving around in the virtual environment represent the members of the virtual team using the workspace. The avatars perform symbolic actions depicting the activities of team members.

.

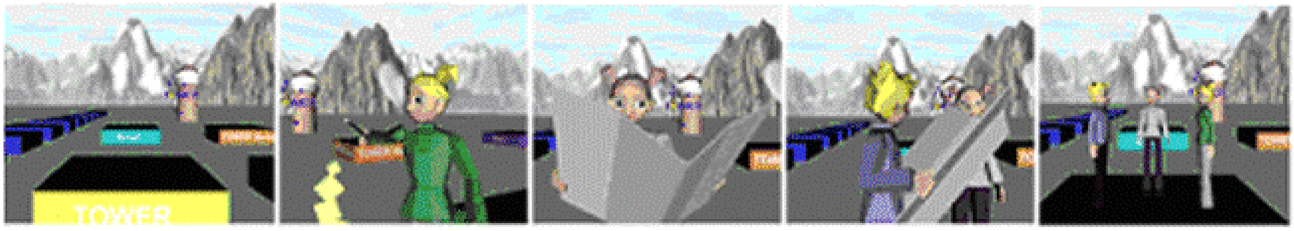


Figure 5: Dialog Scene with three actors

Figure 4 shows a scene in the virtual environment of DocuDrama Conversation The avatar in the front reading a newspaper presents a team member who has opened a document for reading. The avatar is placed on a coloured box which signifies that the document he just opened is a document located in this directory. The moving around between boxes represents a team member switching between directories.

Abstract user data stored in log-files provide the base for stories in DocuDrama Conversation. The challenge is to generate stories from this data, which attract the user's attention and to retrieve from the data information 'between the lines', e.g. the degree of collaboration on a project document.

DocuDrama Conversation allows the selection and sorting of events, e.g. by user, time span or type of event. The events are grouped together in a scene and played out by avatars performing symbolic actions. The avatars turn to each other while performing symbolic actions as they would in a conversation. The play-out of a summary of events on a single document and the presentation by avatars enable the user to quickly grasp the essential information about the work progress on a certain document. The film-like play out in DocuDrama Conversation also enables the user to understand coherences or to become aware of re-occurring patterns in team-members activities.

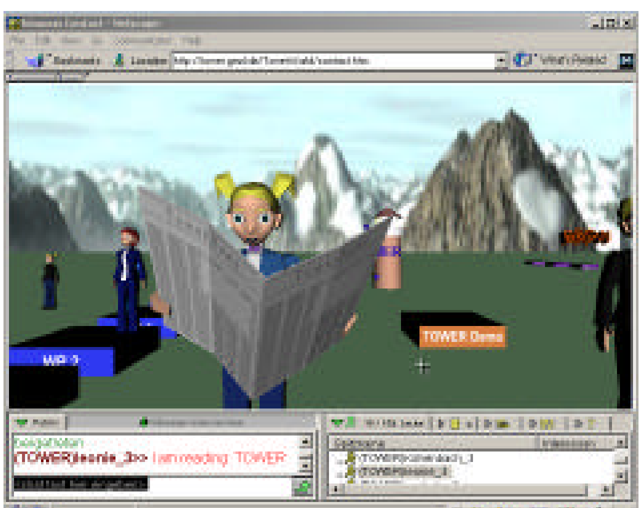


Figure 4: The DocuDrama Conversation virtual environment

Figure 5 shows a dialog scene with three actors in DocuDrama Conversation. Two avatars interact with each other as in a conversation; in turn they perform symbolic actions. In the scene shown above the first avatar shows a Create-Action. Two other avatars subsequently perform a Read-Action, which symbolizes the opening of a document for lecture. Dynamic camera navigation is employed to transfer the user to a central position in respect to the activities. It uses methods borrowed from film making to present the stories in a more exciting and attractive way [25].

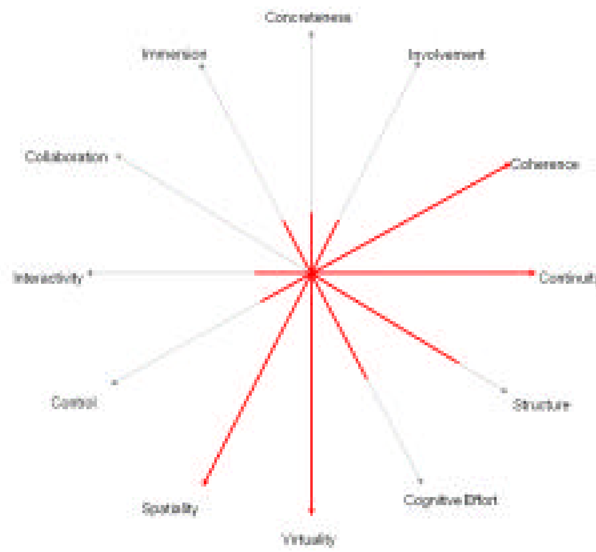


Figure 6: Dimension Star for DocuDrama Conversation

The story material on which the DocuDramas are based is very abstract. In consequence the degree of *Concreteness* is very low. In DocuDrama conversation, the user takes the position of a spectator. He is not involved in the creation of the DocuDrama story content (*Involvement*). The analysis of DocuDrama Conversation shows a high degree of story *Coherence*. The story objects all show a relation to each other. The virtual environment is meaningful to the story through the positions of the avatars; all events which have occurred on a document are grouped together, and the symbolic actions performed are directly interrelated to each other. DocuDrama stories follow a clearly defined *Structure*, although they do not show the three-act-structure typical for dramatic narratives. A fair amount of *Cognitive Effort* is required to follow the story. The fact that there is a conversation taking place is easy to understand, but not necessarily the content of the interaction. DocuDrama stories take place in a virtual environment and use spatial metaphors to convey the story content (*Virtuality, Spatiality*). The user has little *Control* over the development of the story. He only can decide by selection about the composition of the story elements. The application shows a low degree on *Interactivity*. The user only has the option to stop or restart the story one it has been started. DocuDrama Conversation was developed as a single user application without options for *Collaboration*. Owing to the conformity of the presentation the degree of *Immersion* is very low. Figure 6 shows the Dimension star for DocuDrama Conversation.

5.2 StoryNet

StoryNet provides a game-like approach to edutainment. The application targets at participants of seminars on social issues like conflict management or social competence [26].

The plot in StoryNet describes a conflict situation at the workplace. The story is presented to the user in first-person-view. The user takes the part of the main protagonist and is placed in a network of social relationships which typically exist at a workplace. The user has to define his personal goals which he will follow while the story proceeds. The user can revise his decisions at any time. The story does not offer a clear winning or loosing scenario; it is up to the user to judge on the result of his game session. Figure 7 shows scenes from StoryNet.

StoryNet is divided into two conceptual parts. Part one is structured like a tree; each fork requires a decision of the user. The user decides on the progress of the story by emotions, i.e. by control of a set of sliders. In each scene, for each interaction with a story protagonist, the user collects points on conflict and harmony. This way a personal user profile on his relationship with other story characters is developed. The second part of StoryNet shows the result of the user's behaviour. It consists of a collection of scenes which are organized in different levels. At each level the user is provided with a scene which shows the state of the user's relationship to one of the protagonists. The scenes which are presented to the user are selected based on the user's total on conflict and harmony points which he has collected in the first phase of StoryNet. For example, a user behaving moodily and unfriendly in the first part of StoryNet is very likely to receive an unfriendly reaction from the story protagonists in the second phase. If the user shows diplomatic skill in interacting with the story's protagonists in part one, he will probably receive a friendly and welcoming feedback in StoryNet part two.

The first part of StoryNet follows a hierarchical treelike structure. Samsel & Wimberly define this form of plot structure as Bottlenecking [27]. The story starts with a pre-defined sequence of scenes, and ends with a pre-defined scene. In between, on each level the user has several choices on how to proceed. However, since the user interacts by definition of emotions and not by selecting single scenes, the actual story tree remains in the background hidden from the user.

In the second part, scenes are grouped at different levels. This version of plot structure can be labeled Parallel Streaming [27]. Each scene is assigned a number of points on conflict on harmony. The scene which shows the closest match to the

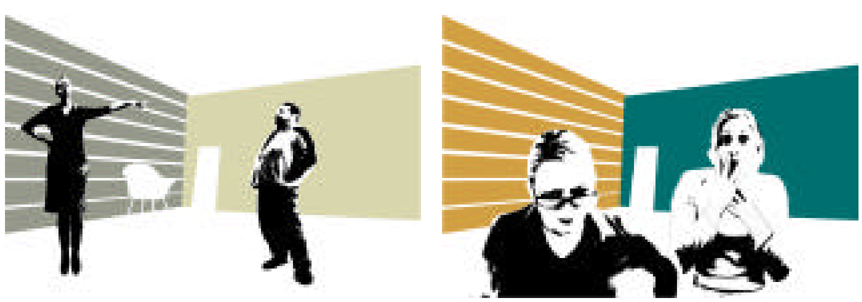


Figure 7. Scenes from StoryNet

Figure 8: Dimension Star for StoryNet

user's number of points is selected for display. In part two, the user has the role of a spectator. The story proceeds like a film, no user interaction is required.

StoryNet is still in the process of development. Pre-user tests which have been conducted on a demo version of StoryNet showed a high interest in the topic. The users enjoyed interacting with StoryNet, and provided ideas for future work, e.g. the development of an overview of the user's current position in the story tree, and a display of the totals of points on conflict and harmony at the end of the first phase.

Figure 8 shows the Dimension Star resulting from the analysis of StoryNet. The story material in StoryNet is very *concrete*; it consists of predefined scenes with images, text and a selection of sliders. The user experiences the story from a first-person-view. He takes the role of the main protagonist in StoryNet. His decisions influence directly the resulting story. The dimension *Involvement* therefore is rated high. StoryNet has been developed as a *coherent* story which can be experienced from different perspectives. For the user it evolves while interacting with the system, but the storyline itself does not evolve. In consequence, the dimension *Continuity* receives a low value. StoryNet has a three act *structure* and this way follows the model of a dramatic narrative. The story is presented similar to a cartoon and therefore easy to follow (*Cognitive Effort*). *Virtuality* and *Spatiality* are both rated low, since StoryNet is a 2D application. Navigation in virtual space or virtual environments in general do not apply. The user can influence the progress of the story by emotion but has no direct *Control* over the next scene presented. StoryNet was developed as single user application for personal training in addition to seminars. There is no option for *Collaboration* with other users. In the first session on StoryNet users got very immersed in the story. The degree of *Immersion* declined in further sessions and is subject of improvement in future versions of StoryNet.

5.3 PhotoStory

PhotoStory [28] still is in its concept phase, nevertheless I think the concept of PhotoStory qualifies for an analysis with respect to storytelling, since the analysis does not focus on implementation details but on features inherent to the PhotoStory concept. PhotoStory aims to provide a collaborative platform for the creation and exchange of stories. Possible applications are the use as a tool for the learning of languages and to support cultural understanding, to provide awareness about team activities, and to offer a means to present a virtual group to the outside world. The intention of creating stories with PhotoStory is twofold. The first intention is to enable communication across borders and to overcome the language barrier. The second intention is more playful. It aims at the random creation of stories based on material available in the media database.

The approach was tested with a group of girls at an open house event at FIT. The girls, aged 8-14, were given of developing a story. First they had to agree on a story subject. The next step was to develop a storyline, taking into account a dramatic arc. The girls then took digital photographs for the story. The pictures were uploaded in a PhotoStory workspace and annotated with keywords. The keywords described the content of the images, author, date, and name of the story. Additionally it was possible to check keywords which described the position of the image within the dramatic arc of the story. The next step in Photostory, though not evaluated yet, will be the retrieval of pictures and their re-arrangement in form of stories. Two options are possible. The first option will allow the retrieval of images for a story by keywords, in the consistence and sequence intended by the author. The second option will allow a random generation of stories, also based on keywords but this time taking into account the keywords' match with the dramatic arc.

Figure 9: An example for a PhotoStory

The dramatic arc, as defined by Frevtag [29], starts with an introduction to the situation, i.e. the location and the people. The arc rises and with it the tension – will the actors succeed or fail? The top of the arc presents the solution to this question. Then the arc descends again with declining tension towards the end of the story, e.g. the happy end.

Figure 9 shows an example for a PhotoStory. It tells the story of a girl which goes for skiing. It is only a short story but follows with its structure a dramatic story arc. The first picture gives an introduction to the location and the actors, the second picture explains the action, the third picture shows the climax – the girl racing on her skis - and the final picture presents the resolution – the girl successfully arriving at the goal.

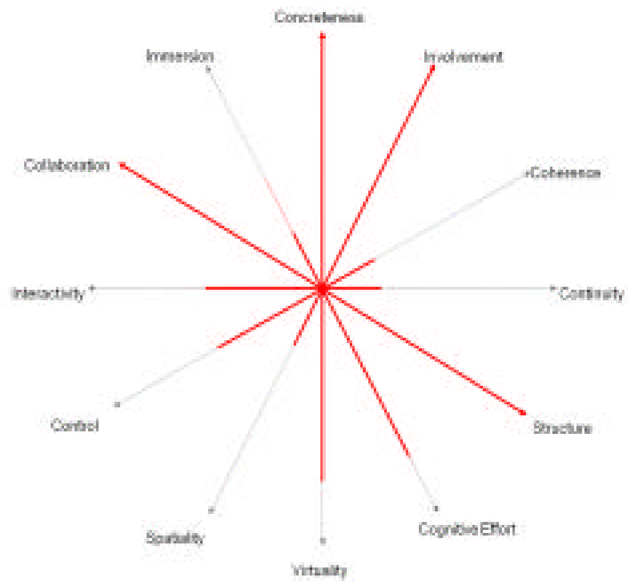


Figure 10: Dimension Star for PhotoStory

PhotoStory uses non-abstract media material which is provided by the users (*Concreteness*). The story results from interaction with the system. The user as story author is responsible for the creation of a story. The user as audience is responsible for the input of keywords which allow the compilation of a story, either re-compiled as intended by the author or consisting of randomly selected media material. The dimension *Involvement* therefore receives a high value. The objects of the story as media objects in a database do not provide any *Coherence* between each other. Furthermore, there is no storyline existent in the background which evolves over time (*Continuity*). However, the degree of *Structure* is very high since the user as author is supposed to create stories which follow the dramatic story arc. The stories generated randomly for the user follow this rule. Regarding the randomly generated stories the *Cognitive Effort* can be very high. *Virtuality* plays a major role, since both the creation and the presentation of the story take place in a virtual collaborative workspace. *Spatiality* does not apply since the media material consists exclusively of images and text. The user as author has a high degree of control on the story content, but as audience he exerts only a low degree of control. This results in a medium rating for the dimension *Control*. PhotoStory offers several possibilities for interaction (*Interactivity*). It is a highly collaborative application (*Collaboration*). *Immersion* is ranked rather low, also comparable to a cartoon rather than a highly immersive book. Figure 10 shows the dimension star resulting from the analysis of PhotoStory.

6. DISCUSSION AND CONCLUSION

In this paper I presented an overview of the wide range of applications in Digital Storytelling. I provided a classification which groups applications in different categories in respect to their degree on oral vs. digital storytelling. I introduced the Dimension Star as a reference model and analysed three storytelling applications with different foci and features. The Dimension Star can be used for analyzing applications in digital storytelling, and will also serve as a design guideline for the development of future systems.

The Dimension Star enables the user to see at a glance the strengths and weaknesses of an application. It allows comparison between different applications in respect to its dimensions. The Dimension Star as a reference model simplifies the classifying and assigning of applications to a category.

It might be argued whether a classification in categories is a useful approach for any type of system. Especially if systems show qualities and features which might belong to several of the presented categories. However, in the first stages of conceptualization, design and development of a storytelling system, such a categorization will provide valuable support. The different dimensions allow are to compare and benchmark developments against successful applications which are representative for the respective category.

Future Work will include a refinement in the definition of the dimensions. It will also include the extension of the Dimension Star for further dimensions to enable comparison and benchmarking in more detail. Furthermore the reference model will be verified by applying it to a wider variety of applications, e.g. games, film or even non-storytelling applications. This will provide the user with a better understanding of the applicability of the Dimension Star as a reference model. The future and long-term goal will be to develop general design guidelines for applications in Digital Storytelling.

7. ACKNOWLEDGMENTS

Many thanks to Bernd Mahr at the TU Berlin and my colleagues at Fraunhofer FIT, especially Uta Pankoke-Babatz and Wolfgang Prinz which provided helpful comments while writing this paper, I also thank Carla Valle, Agnes Stauber, Bozana Bokan, and Zeeshan Javeed for cooperation and support in the development of the showcases.

8. REFERENCES

[1] U. Spierling, D. Grasbon, N. Braun, and I. Iurgel, Setting the scene: playing digital director in interactive storytelling and creation. Computers & Graphics, 2002. 26(1): p. 31-44.

[2] J. Cassell, Towards a model of technology and literacy development: Story listening systems. Journal of Applied Developmental Psychology, 2003. 25(1): p. 75-105. http://www.soc.northwestern.edu/justine/publications/Cassell.JAPD.published.pdf

[3] A. Kuchinsky, K. Graham, D. Moh, A. Adler, K. Babaria, and M.L. Creech, Biological Storytelling: A Software Tool for Biological Information Organization Based upon Narrative Structure. ACM SIGGROUP Bulletin, 2002. 23(2): p. 4-5.

[4] J. Schell, Understanding Entertainment: Story And Gameplay Are One, in The Human-Computer Interaction Handbook, J.A. Jacko and A. Sears, Editors 2003, Lawrence Erlbaum Associates.

[5] W. Swartout and M.v. Lent, Making a Game of System Design Communications of the ACM, 2003. 46(7): p. 32-39

[6] J. Murray, Hamlet on the Holodeck: The Future of Narrative in Cyberspace. Cambridge, MA, USA: MIT Press. 1998.

[7] H. Jenkins, Game Design as Narrative Architecture, in First Person, P. Harrington and N. Frup-Waldrop (eds.), 2002, MIT Press: Cambridge.

[8] J. Juul, Games telling stories? A brief note on games and narratives. Game Studies - the international journal of computer game research, 2001. 1(1). http://gamestudies.org/0101/juul-gts/

[9] E. Adams, Three Problems for Interactive Storytellers, Gamasutra - Features - The Designer's Notebook, 1999, http://www.gamasutra.com/features/designers_notebook/19991229.htm, Last visited: 26.November 2003

[10] K.E. Steiner and J. Tomkins. Narrative Event Adaptation in Virtual Environments. Proceedings of the IUI'04. 2004. Madeira, Funchal, Portugal, ACM.

[11] J. Fritz and W. Fehr, Kriterien zur pädagogischen Beurteilung von Computer- und Videospielen, in Handbuch Medien: Computerspiele: Theorie, Forschung, Praxis, J.Fritz and W.Fehr (Hrsg.), 1997, Bundeszentrale für politische Bildung: Bonn. p. 333-340.

[12] A. Stauber, Computerspiele in der Bildung - Das Game Design im Kontext der Wissensvermittlung, Universität der Künste Berlin, Diplomarbeit,2004

[13] C.A. Lindley. Narrative, Game Play, and Alternative Time Structures for Virtual Environments. Proceedings of the 2nd International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE 2004). 2004. Darmstadt, Germany, Springer Verlag Berlin Heidelberg.

[14] T. Manninen, Interaction Forms and Communicative Actions in Multiplayer Games. Game Studies - the international journal of computer game research, 2003. 3(1). http://gamestudies.org/0301/manninen/

[15] M. Mateas, A Neo-Aristotelian Theory of Interactive Drama, in Artificial Intelligence and Interactive Entertainment 2000, AAAI Press: Menlo Park. p. 56-61.

[16] B. Laurel, Computers as Theatre. Addison-Wesley. 1993.

[17] M. Mateas and A. Stern. Integrating Plot, Character and Natural Language Processing in the Interactive Drama Facade. Proceedings of the TIDSE. 2003. Darmstadt, Germany, Fraunhofer IRB Verlag.

[18] K. Salen and E. Zimmerman, Rules of Play : Game Design Fundamentals. MIT Press. 2003.

[19] C.A. Lindley. The Gameplay Gestalt, Narrative, and Interactive Storytelling. Proceedings of the Computer

Games and Digital Cultures Conference. 2002. Tampere, Finland.

[20] E. Zimmerman, Narrative, Interactivity, Play, and Games, 2004, http://www.electronicbookreview.com/v3/servlet/ebr?-command=view_essay&essay_id=zimmerman, Last visited: 18.August 2004

[21] W. Prinz, U. Pankoke-Babatz, W. Gräther, T. Gross, K.-H. Klein, S. Kolvenbach, and L. Schäfer, Presenting activity information in an inhabited information space, in Inhabited Information Spaces, E. Churchill, D. Snowdon, and F. Frecon, Editors 2004, Springer Verlag UK.

[22] W. Appelt. WWW Based Collaboration with the BSCW System. Proceedings of the SOFSEM '99: theory and practice of informatics. 1999.

[23] L. Schäfer, U. Pankoke-Babatz, W. Prinz, A. Fatah-gen. Schieck, and A. Oldroyd, DocuDrama. Virtual Reality, 2003. 7(1).Springer Verlag UK

[24] L. Schäfer, E.M. Raybourn, and A. Oldroyd. DocuDrama Conversations. Proceedings of the SIGGRAPH - Conference Abstracts and Applications, 2002. San Antonio, Texas, USA, ACM Press, p.171.

[25] B. Bokan and L. Schäfer. Intelligent Camera Direction in Virtual Storytelling. Proceedings of the IVA'03, 4th International Workshop on Intelligent Virtual Agents. 2003. Kloster Irsee, Germany, Springer Verlag.

[26] L. Schäfer, A. Stauber, and B. Bokan. StoryNet: An Educational Game for Social Skills. Proceedings of the 2nd International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE 2004). 2004. Darmstadt, Germany, Springer Verlag Berlin Heidelberg.

[27] J. Samsel and D. Wimberley, Writing for Interactive Media. New York: Allworth Press. 1998.

[28] L. Schäfer, C. Valle, and W. Prinz. Group Storytelling for Team Awareness and Entertainment. To appear in: Proceedings of the NordiCHI. 2004. Tampere, Finland.

[29] G. Freytag, Die Technik des Dramas. Erste Auflage der Neubearbeitung, Berlin: Autorenhaus Verlag. 2003.

Communication-Oriented Modelling – Transforming the Social Visibility of Communication Processes into Cognitive Visibility

Marco Schmitt
TU Hamburg-Harburg
Department of Technology Assessment
Schwarzenbergstr. 95
21073 Hamburg
Tel.: +49 40 42878 3908
marco.schmitt@tuhh.de

ABSTRACT
The visualization of communication processes or of their central structural aspects is of vital importance for the orientation of users and scientific observers of computer-mediated communication. Starting from this hypothesis, we would like to introduce a modeling approach that focuses on the "visible" part of communication: the *message sign*. To demonstrate the C*ommunication-Oriented Modeling* (COM) approach, we would like to proceed in four steps. The first step involves the localization of COM inside the Socionic Research Program, the second step shows the theory of communication behind COM, the third step focuses on the role of the message sign in communication processes, and the final part deals with the concept of social visibility.

Keywords
Communication, Communication-Oriented Modeling, Socionics, Message Sign, Dynamic Networks, Social Visibility.

1. INTRODUCTION
The visualization of communication processes and several of their central structural aspects is of definite value for the orientation of users and scientific observers of computer-mediated communication. Departing from that assumption, we would like to introduce a modeling approach that focuses on the "visible" – in the sense of observable – part of communication: the *message sign*. To demonstrate the C*ommunication-Oriented Modeling* (COM) approach, we would like to proceed in four steps.

The first step involves the localization of COM inside the Socionic Research Program, as a complement and an alternative to the dominating agent-oriented paradigm. Socionics is an interdisciplinary research program between computer science and sociology to build multi-agent systems (MAS) that incorporates social mechanisms to enhance cooperation and coordination of the agents and to achieve a new quality of social simulation. The focus on MAS has lead to a dominance of agent-oriented approaches in Socionics. Inspired by the diminishing role of traceable agents in computer-mediated communication, COM proposes an alternative approach that focuses on the modeling of communication processes, rather than on the modeling of agents.

The second step shows the theory of communication behind COM. This theory consists of a triadic concept of communication with two invisible and transient operations and a visible and more or less persistent message sign. The first of these operations is the *inception*, the production of a message sign, the second is the *reception*, the experience and interpretation of a message sign. The ongoing interplay between actions, interpretations, and observable signs constitutes the communication process as such.

The third step focuses on the role of the message sign in communication processes. The most important feature is the observability of the message sign. It is our contention, that the message sign – as an empirical object – is the hallmark of the constitution of communication processes out of cognitive and physical operations. Only the direct reference to a message sign specifies the operations of inception and reception as being communicative operations. COM uses the dynamic network of cross-referencing between message signs as the basic modeling level.

In the fourth step, we would like to demonstrate how to transform the social relevance of a message sign in a communication process into cognitive visibility for the observer. COM uses a *visibility function* that computes the social relevance of a message sign out of the incoming references it receives from following messages. This social relevance measurement is called *social visibility*. On that basis, it is possible to reconstruct the social visibility of a message sign from a dynamic network of cross-referencing messages and to transform it via different visualization techniques into cognitive visibility. This social visibility of a message could be an important structural information about an ongoing computer-mediated communication process for participants and other interested observers alike.

Finally, we will close this paper with a conclusion that tries to relate the fields of socionics, communication-oriented modeling, and computer semiotics according to their similarities and complementarities. We will illustrate the closeness of our approach to computer semiotics by referring to certain papers of previous COSIGN conferences and show that they elaborate research questions of great interest for communication-oriented modeling.

First published at COSIGN 2004,
14 – 16 September 2004, University of Split (Croatia)

2. COMMUNICATION-ORIENTED MODELING AND THE SOCIONICS RESEARCH PROGRAM

The Socionic Research Program, funded by the Deutsche Forschungsgemeinschaft (DFG), introduces concepts, insights, and mechanisms from sociological theory into the scientific field of Distributed Artificial Intelligence (DAI) and the Research on Multi-Agent Systems (MAS). The interdisciplinarity of this approach is secured by a commitment of computer scientists and social scientists to fuse their research interests into the construction of prototypes of multi-agent systems. This endeavor, to construe artificial sociality in a controllable fashion, has a multiplicity of research agendas. From a sociological perspective, socionics should achieve the clarification of main concepts of social theory and high-quality social simulation. Computer science should profit from new mechanisms of coordination, cooperation and conflict resolution among autonomous agents (for a full programmatic see Malsch [16, 17]).

It seems obvious that socionics is mainly based on the development of agent technologies, agent interaction/ communication protocols, platform development for multi-agent systems and agent-based social simulation. Therefore, socionics followed the path of agent-oriented modeling (AOM). In most cases this is quite consistent with wide branches of sociological theory, where the actor and his observations, evaluations, choices, and actions form the center and kernel of theoretical development (see for instances Coleman, Esser, or Giddens [3, 7, and 9]). The agents build the persistent part of the developed systems and research is focused on agent interaction and the relationships between agents (see Ferber or Weiss for overviews [8, 26]). Despite the considerable achievements of AOM in the realm of distributed and cooperative problem solving, there although exist certain shortcomings of that approach. A central problem according to Malsch & Schlieder [18] is the speech-act based modeling of communication in AOM, because of the inherent limitations concerning mass communication, caused by the sender-receiver pattern of speech-acts, also labeled as the "message sending paradigm" [18]. Large-scale communication processes tend to diminish the role of the sender and the receiver for the communication, so that agent-to-agent-relations are of generally less importance in these many-to-many communication processes, than traditional AOM approaches would assume. Focusing on agent relations, the message sending paradigm also misses the centrality of message-to-message-relations. The referential structure between the messages becomes more and more unclear and unobservable when the load of agents and messages rises to large amounts and has to be channeled through sending and receiving agents. Finally, such large amounts of momentarily participating agents and communicative activities involves an unbearable high modeling complexity. Cutting a long argument short: The message sending paradigm cannot be scaled beyond a certain limit, that falls definitely short of the many-to-many cases of mass communication.

These shortcomings of AOM lead us to shift attention from the interacting agents to the communication events and their referential relations. That is an fairly unconventional approach in socionics as highlighted above; but from a sociological point of view, it seems to be a plausible alternative. By such a shift of attention, we simply follow one of Luhmann's proposals: "If one begins with the possibility of a theory of self-referential systems and with problems of complexity, there is much to suggest simply reversing the relationship of constraint. Sociality is not a special case of action; instead, action is constituted in social systems by means of communication and attribution as a reduction of complexity, as an indispensable self-simplification of the system." [14] Even in sociology such a proposal implies a paradigm shift from action theory to communication theory, so that COM has to rest necessarily on a well-developed theory of communication.

3. COMMUNICATION THEORY

A communication theory with elements from sociological theory (especially the theory of social systems, in the tradition of Niklas Luhmann [14] and the theory of symbolic interactionism, in the tradition of George Herbert Mead[19]) and semiotic theory (in the tradition of Charles Sanders Peirce [20]) constitutes the fundament of COM. We start from some proposals of Luhmann's theory of communication, point out some theoretical problems of major interest to our own approach and develop the basic vocabulary of our theory of communication.

3.1 Points of Departure: Luhmann on Communication

Luhmann suggests to take "communication" as the basic term for understanding social systems and therefore, a theory of communication constitutes the core of his social theory. This approach departs from the common use of "social action" as the basic category for sociological theory, as introduced by Weber [25]. Communication is the operation that constitutes and reproduce social systems and the basic element for the analysis of such systems. COM shares that foundational hypothesis with the theory of social systems. Luhmann gives a threefold interpretation of communication:

1. **Communication as the mode of operation of social systems.**
2. **Communication as the basic and constituent element of social systems.**
3. **Communication as the temporal atom of social systems.**

This third interpretation describes communication as an event, a vanishing moment in the systemic reproduction. A temporal unit that cannot be further divided with reference to the social system. This temporal dimension of communication is of major interest to COM's processual perspective on communication. Two characteristics of this definition of communication as the temporal atom of social systems are essential to the problematic we would like to address here: that communication is event-like and that every communication event is extremely runny. A communication event has to be substituted immediately by the next one or the communication process breaks down.[1] This process of

[1] This was a bit of an exaggeration. Due to the features of certain media of dissemination or other storage devices, some elements of the communication event can be preserved for quite a time. Here rests a grave problem of Luhmann's approach to which COM tries to explicate a possible solution. (see below)

immediate substitution is what communication is all about; it could be called the autopoiesis of communication. [14]

According to Luhmann, communication is “coordinated selection” [14]. Every communication synthesizes three selections: information, utterance and understanding. Therefore, “… communication must be viewed not as a two-part, but as a three-part selection process. “ [14] The information selection actualizes the referential horizon of the communication, by choosing one point of reference and not another one. The utterance is a selection of an expression behavior for the communication. The last selection – that of understanding – is of decisive importance. It is based on the distinction between information and utterance. Understanding – as making a difference between the information and the uttering action – completes a communication event. Communication organizes itself from that last selection backwards. Understanding some event as an uttering of an information coordinates all three selections in the last one and creates the unity of a communication as a singular event. Understanding attributes that utterance of an information as an action event and fixes the communication event at one point of time.

COM focuses on a special problem that arises from the temporal implications of the three-selections approach by Luhmann. There may be an immense time-span between the utterance selection and the understanding selection. Therefore the question is: how is it possible to describe communication as an event, as the temporal atom of social systems, when such an event is elongated to a considerable amount of real-time? Interpreted from a real-time perspective, utterance and understanding are different events and do not constitute a single event.

An additional sphere of highly interesting problems for COM stems from Luhmann’s statement that communication is highly improbable. There are a variety of basic obstructions communication has to overcome. The three improbabilities of communication are: understandability, reachability, and successfulness. Every communicational offer is – without additional assumptions like context – likely to be misunderstood, probably unable to reach its addressee, and unlikely to be accepted and followed. These improbabilities of communication “… operate as thresholds of discouragement.” [14] No communication comes to pass without transforming these improbabilities into probabilities. COM has to address these problems, in order to develop a theory that helps to simulate the stabilization and reproduction of specific types of communication processes.

So far, two central insights from Luhmann’s theory of social systems are vitally important for the development of COM: the paradoxical signification of communication as a temporal atom of social systems and the thesis of the improbability of communication itself. These are the points of departure for the elaboration of our own communication theory.

3.2 Operations: Inception and Reception

The operational level of communication is concerned with the temporal aspects of communication processes. We try to point out possible discrete events that constitute communication. Therefore, COM uses a triadic model of communication (see figure 1.), consisting of two event-like operations – *inception* and *reception* – and a message sign – a specific formation of a medium of communication. The term message sign and its special role in our theory and our modeling approach will be further explained later on. Here, we use message sign as a point of reference for the communicative operations of inception and reception. Only the direct reference to a message sign lifts the two operations from the mental to the social (communicative) level.

Right now, we will focus on this operational level. The two operations have an obvious background in the consciousness of actors that participate in communication processes. Both operations are not directly observable, because every psychic system is a black-box for any other psychic system. Both operations have to be conceptualized in a fashion that makes it possible to infer them from the observable part of the communication process.

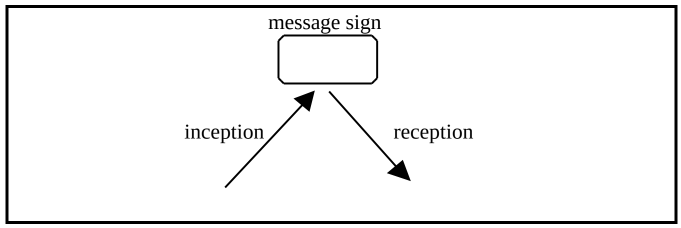


Figure 1. Triadic Concept of Communication

The inception may be conceptualized as the act of uttering, that is the production of a physically manifest sign. The inception is a form of action and is attributed to a person. To avoid misunderstandings: Inception is not the action of sending a message – that would be the case in the perspective of a message sending paradigm – it is the production of a message sign in the form of a “publishing” activity. In short:

inception ← production of a message sign

The reception, on the other hand, stands for the perception and interpretation of a message sign. A reception (re-)constructs some form of information from a message sign. So, there is also some activity involved in the reception, because somebody has to actively interpret and evaluate the signs used in the communication process. In short:

reception ← perception + interpretation of a message sign

Inception and reception as single operations are coupled in two ways. On the one hand, the message sign couples a foregoing inception with a plurality of receptions. On the other hand, actors or persons couple the reception of a foregoing message sign to the inception of a new one. This coupling of operations inside the psychic systems participating in the communication has to be observable as some form of referencing from one message sign to another.[2] To develop a theory of communicative operations right below the level of the Luhmannian synthesis, involves a strategic withdrawal from two central theorems implied by his approach. A possibly indefinite temporal stretching of a communicative event between utterance and understanding will be excluded

[2] As we will see later, COM is mainly interested in the build-up of such referential structures between message signs, rather than in the internal operations of a psychic system. Strictly speaking, these internal information processing of agents belongs to the AOM paradigm.

from our approach, as well as any form of dominance of the understanding selection over the communication process. In re-differentiating the operations of inception and reception as discrete events in the communication process, we address the problem of the definition of the temporal atom of communication, as formulated above. In the next section I will show, how we like to address the second problem – the problem of the improbability of communication.

3.3 Selectors: Significance and Relevance

The theme "improbability of communication" should be addressed on a structural level. COM prefers a bottom-up approach to modeling and simulation, so that an operational structure is introduced. Both operations – inception and reception – are constructed on the basis of a structural homology. They process two structural values, called selectors: *significance* and *relevance*. These selectors determine the probability for the connection of a new message to an old one. Significance and relevance are valuations attributed to a message sign. The inception tries to inscribe these valuations into the message sign, whereas the reception attributes these values in the form of an interpretation of a given message sign under observation.

The simplest mode to capture the notion of significance of a message sign seems to go by using the distinction appropriate/not-appropriate. Significance could be measured by some standard of similarity or correspondence in a thematic or semantic way. This may usually happen by the introduction of binary codes or nominal scales based on simple semantics or ontologies. It seems quite difficult to construe a more formalizable measure like semantic nearness ore sameness. In short:

significance ← some standards of appropriateness of a message sign

Relevance, on the other hand, is captured by using the distinction important/not-important. A gradual measurement of and a formal approach to relevance seems more likely then in the case of significance. A difficulty for the modeling rests in the wide range of possible approaches to measure the relevance of a message sign. As we will show below, there is a striking similarity between relevance and visibility of a message sign, but the concepts arise from different perspectives on communication processes (see below). Whereas relevance is assigned to a message by a single agent and the attributions from different agents may contradict each other, visibility is assigned to a message sign from a process perspective indifferent to the interpretations of the individual agents. In short:

relevance ← some standards of importance of a message sign

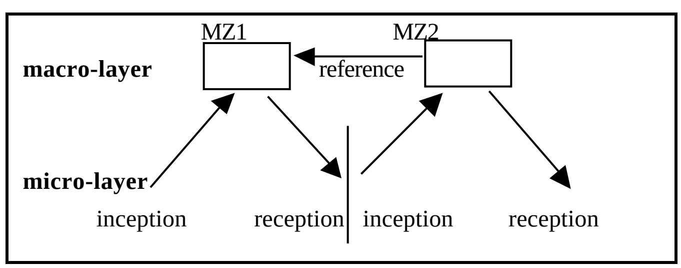
The greatest difficulty concerning these structural values rests in the probable difference between the attributions of significance and relevance from different agents for the same message sign. The integration of these deviating interpretations and perspectives seems to be a great obstacle for our approach.[3] A further problem arises from the difference between the inscription of structural values by an inception and the interpretative valuation by a reception. Therefore, we have to rest on the observable part of communication expressed in message signs.

3.4 Selection Problems and Reproduction Problems in Communication

COM and the communication theory developed here are mainly focused on two problems. Behind those problems stand two different perspectives on communication. The first problem could be summarized in the question: "Where should I connect with my communication proposal?" It is a *problem of selection*. The problem is formulated from the perspective of the single agent. The second problem, on the other hand, concerns the reproduction of an identifiable communication process. The formulation of this problem draws from the perspective of the process itself. It is a *problem of reproduction*. Therefore, we are able to divide two layers inside our theory of communication, where both problems can be addressed separately.

The first layer, or micro-perspective, focuses on the problem of selection. Selection problems are addressed by the receptions and inceptions of agents directed by their attributions of significance and relevance to various message signs. Agents select references to the message signs they value as appropriate and important according to their interests. They may and probably will orient their attributions to the social visibility of the message signs for the whole communication process. The social visibility is a macro-effect produced by the selections of all the agents and is used by them to orient their valuations.

The second layer, or macro-perspective, focuses on the problem of reproduction. Reproduction problems are addressed by following the references between message signs, so that patterns of strong-referenced and weak-referenced message signs emerge from the ongoing communication process. We would like to find out functions that capture the reproduction of such patterns.[4] Where the problem of selection has to be articulated on the operational level of reception and inception, the problem of reproduction has to be captured on the level of the message-to-message-relations indicated by the references (see figure 2.).



[3] As we will show below (in section 5.), we try to surpass this obstacle at the moment by concentrating on a structural value from the process perspective (e.g. social visibility), rather than using the selectors described here. In future work, we will try to integrate the selectors into our simulation tool.

[4] Such functions will be like the visibility functions elaborated in section 5.1 below.

Figure 2. Communication Model and the Problems of Selection and Reproduction

In future research, the central theme of our theoretical work will reside in the possible coupling and de-coupling of the two perspectives, the design of the micro-macro-link in communication. The task is to construct a plausible link between the process-oriented measure of visibility and the agent-oriented structural selectors significance and relevance. If you take up a communication-oriented approach, the agents should not determine the structural effects, but their attributions should have a significant effect on the social structures (referential patterns), especially when they express their attributions through message signs.

4. THE ROLE OF THE MESSAGE SIGN

The message sign is an important feature of our theory of communication, as well as for the visualization of communication processes in COM. In this section, we would like to discuss the semiotic background of the term "message sign" used here, the main attributes of the message sign and the value of that notion for the presentation of structural aspects of communication processes.

4.1 Semiotic Background

In this first section concerning the role of the notion *message sign* in our theory, we will point out some connections to semiotic theories to clarify our understanding and use of the term.[5] From the two main traditions in semiotics – the European-Structuralist tradition based on the work of Ferdinand de Saussure [22] and the American-Pragmatist tradition based on the work of Charles Sanders Peirce [20] (a distinction also common for computer semiotics [1]) – we strongly tend and refer to the American-Pragmatist side.

The term "message sign" is an uncommon notion in semiotics. It is a composite term, that includes the central research object of semiotics – the sign – and a notion of a medium in which the sign and its usage materialize – the message. The two central traditions in semiotics – the Peircean and Saussurean tradition – develop definitions of the sign that share certain similarities but lead into quite different directions. According to Eco [6], Peirce views the sign as something which stands to somebody for something in some way, whereas Saussure defines the sign as the difference between the signifier and the signified. Peirce seems to be interested in the role of signs in the process of semiosis, Saussure, on the other hand, focuses on the inner structure of sign systems. From our sociological viewpoint, we tend to the perspective of Peirce, because our research interests do not reside in language itself, but in the role of language and other media in social (communication) processes. In viewing language as the medium that couples consciousness and communication, the sign system and its inner structure seems to be of less importance.

Coming back to the notion of "sign" in the sense of Peirce implies three characteristic features of the sign that we would like to highlight. According to Peirce, every sign consists of a material quality (it is something), a demonstrative application (it stands for something in a real causal connection), and an idea in a mind (to somebody) [20]. We take-up that definition of the sign, but we would like to add something to denote the communicative usage of the sign. Therefore, we call the observable part of communication "message signs", in order to shed some light on the production and interpretation of signs as a communicative behavior. The notion should include an explicit media reference, a difference of signifier and signified (self- and other-reference) and a reference to some communicative activity. Mostly, a message sign contains not only a basic sign or a single symbol, but a composite sign like a sentence or a whole website on the internet. A plausible alternative to our term could be the notion of "used sign" or "communicative form".[6]

4.2 Persistency and Observability

As described above, we focus on two properties that message signs inhere prior to communication: *persistency* and *observability*. The role of the message sign, as the expression plane of communication, the part of communication that can be observed by a detached third party, is constituted by these two properties. Both aspects point to perception, rather than to communication. In our perspective, that does not lead to a contradiction, because perception is a decisive prerequisite of communication and here rests the central role of the message sign. Communication is forced to attract conscious attention to continue. This attraction of attention is one of the main functions of message signs. To achieve this, some form of persistency and observability has to be realized. Based on these two properties – produced through a formation of a medium – the message sign connects communication and perception, it effects either social and psychic systems.

Persistency describes the physical survival of the message sign and is bound to the material quality of the sign. The persistency is mainly an attribute of the medium which is used to produce the message sign. For the social persistency of message signs we reserve the term "social visibility" (see below), that just couples loosely to the material quality of the message sign.

Observability seems to be more closely related to the visibility of a sign in communication processes. In opposition to persistency, the observability of a message sign is not solely determined by its material quality. It has to be reachable by the senses of certain agents as well. If an operation of reception cannot occur, then we would not speak of a proper message sign. This attribute realizes the potential for a reference to that message sign, but tells us nothing about the attention it actually receives, nor about the attention it is likely to attract.

Therefore, persistency and observability can be viewed as prerequisites for the functioning of message signs in

---

[5] We acknowledge that such a term seems to be quite uncommon in semiotics, but we would like to point out our major interest in actual sign usage in communication and the material realization of that usage as the observable part of communication.

[6] We acknowledge that this direct reference to communicative activity is included in Peirce notion of the "sign", because semiosis is a process of sign usage by a mind or by communication to generate meaning (see for instances the interpretations of Eco [5, 6] and Simon [24]). Our terminological add-on should just highlight our sociological interest in sign usage in communication. To be clear, in congruence with Luhmann [15], the message sign is not a communicative operation (no sign is), but it plays an important role, because it hints towards such operations.

communication and as nothing more. They are also necessary conditions for the observation of communication by third parties. All further aspects and characteristics of communication can only be inferred from this plane and to make such inferences possible could be designated as the role we would like to ascribe to the message sign. One of these aspects is the build-up of a referential structure between the message signs, another aspect is the social visibility of a single message sign that emerges from the referential structure of the communication process.

4.3 The Referential Structure

Communication can be viewed as the build-up of a referential structure between message signs. The referential patterns represent the structure of the underlying communication processes. In COM, we would like to follow and simulate the emergence and stabilization of such referential patterns in communication.

A reference between message signs is established by a reception of a previous message sign that leads to the inception of a new one (see figure 2. above). That these two operations have been come to pass has to be noticeable or inferable from the following message sign, so that we could also say that the reference has to be inscribed into the message sign in some way. We can differentiate a push-model of communication, where the referential structure of the process is produced by operations of reception and inception and a pull-model of communication, where following messages ascribe references to previous ones. These models represent different perspectives on communication and a communication-oriented approach should lean to the latter one, without forgetting or ignoring the former. We would like to try out both approaches to model the emergence of referential patterns in communication in our further research and hope that our communication theory will combine the models in a plausible way.

Apart from the question of the origin of referential patterns in communication processes, we are especially interested in the formation and reproduction of specific patterns over time. This theme of pattern reproduction (as mentioned in section 3.4) is vitally important for the scientific analysis of communication and for the orientation of participating agents alike. In the following section, we will show that such patterns could be described by the distribution of social visibility between message signs and that this distribution is generalizable by a specific visibility function. What we would like to model by the generation of reference patterns in communication are conditions for the reproduction/ stabilization of structures and processes over considerable time-scales, as well as the differentiation of types of communication processes according to their referential structure. The results from the simulation and empirical analysis of pattern formation in communication may even lead to a plausible typology of processes according to the underlying distribution of visibility.

5. SOCIAL VISIBILITY AND COGNITIVE VISIBILITY

As noted above for several times, one of the central features of communication processes – and the central feature of our approach to a communication-oriented modeling – is the construction of a kind of “social visibility” from the referential patterns between message signs. This visibility of message signs in the communication process seems to be a promising structural value for the analysis of communication processes in general. The social visibility of a message sign is a compact expression for the probability that it will be referenced by a new message sign. Social visibility of a message sign signifies its potential for future attraction of attention and results from the attention that the message sign has already attracted during the communication. A case of “preferential attachment” as you may call it.

The first of the two following sections demonstrates the construction of visibility functions to simulate communicative patterns and the distribution of social visibility in different types of communication processes. The second section will develop a description of our simulation tool COMTE and address the problem of transforming the social visibility of message signs into cognitive visibility for the participants by means of a variety of visualization techniques. This transformation is the key to support agents confronted with the problem of selection.

5.1 Visibility Functions

In the following, we will describe visibility as an ordering principle in communication processes, that directs their further development. The actual distribution of visibility between the message signs can be generalized into a “visibility function” for the whole process.

As has been mentioned, we introduce visibility as a principle of order in communication processes, a principle of order that can be generalized from referential patterns between message signs. A single message sign receives a high visibility, when it attracts a lot of attention in communication. Obviously, this involves a circular definition, because a high visibility leads to further attraction of attention. There is a positive feedback-mechanism at work. The visibility functions describe the distributional patterns of visibility in whole communication processes. With this functions we have got formulas to produce certain pattern-specific communication processes.

The basic assumption behind the construction of visibility functions for simulations suppose that the function should depend on an aging factor and on an attention factor for message signs. The aging factor could be given by the medium of communication in use or by the speed of the process itself. The attention factor could be measured by the references a message sign receives from other message signs.

There are two ways to construct a visibility function. On the one hand, you could take empirical data of communication processes and derive a possible visibility function from the referential structure. Or, on the other hand, you could determine what type of communication process or what pattern you would like to simulate and create the visibility function in advance that is able to produce such a process. We would like to proceed in both ways: Analyze empirical data from internet forums, scientific citation indices and some additional forms of computer-mediated communication and simulate various forms of communication processes with an expanding arsenal of usable visibility functions.

Up to the moment, we have got a small range of visibility functions to run simulations with. The most interesting one is a degree-based visibility function with aging. This function includes the desirable factors of aging and attention.[7] We are able to create some interesting referential patterns with this relative simple visibility function. We have simulated three different scientific citation procedures: a historicist mode, a classicist mode, and a modernist mode of citation. In the historicist mode (see figure 3.), aging is a neglectable factor and a "first comes first"-pattern of citation is stabilized. In the classicist mode (see figure 4.), both factors are of considerable influence and a balance between attention and aging is established, so that classic works appear over the whole time-span of the process. Finally, in the modernist mode (see figure 5.), aging is a much more influential factor and only a block of recent message signs is visible in the process.

In the future research, we would like to expand our sample of visibility functions, to include more sociological relevant factors and finally try to supplement and/or substitute the globally working visibility function with a set of different locally working visibility functions or subjective visibility functions for specific agents participating in communication processes. To reach this research goals we have to further integrate the micro-perspective and the macro-perspective on communication on the basis of inferences from the observable part of communication.

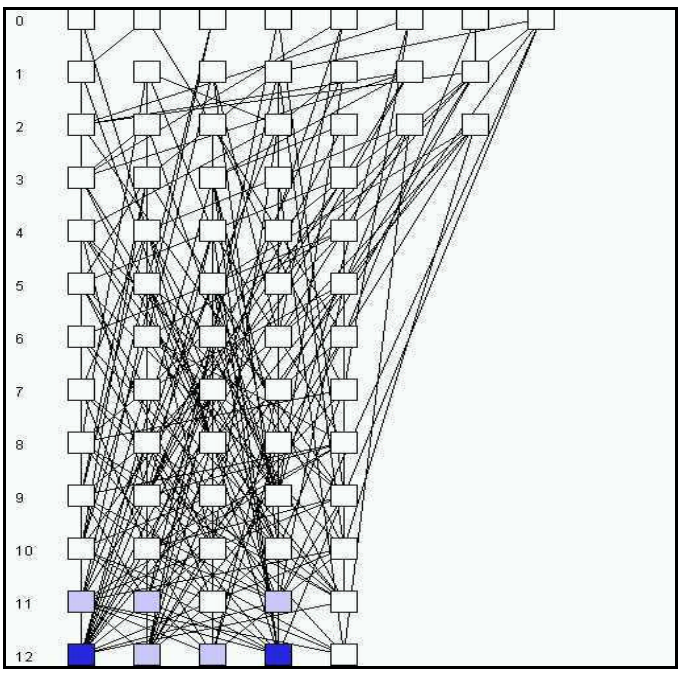


Figure 3. Historicist Mode of Communication[8]

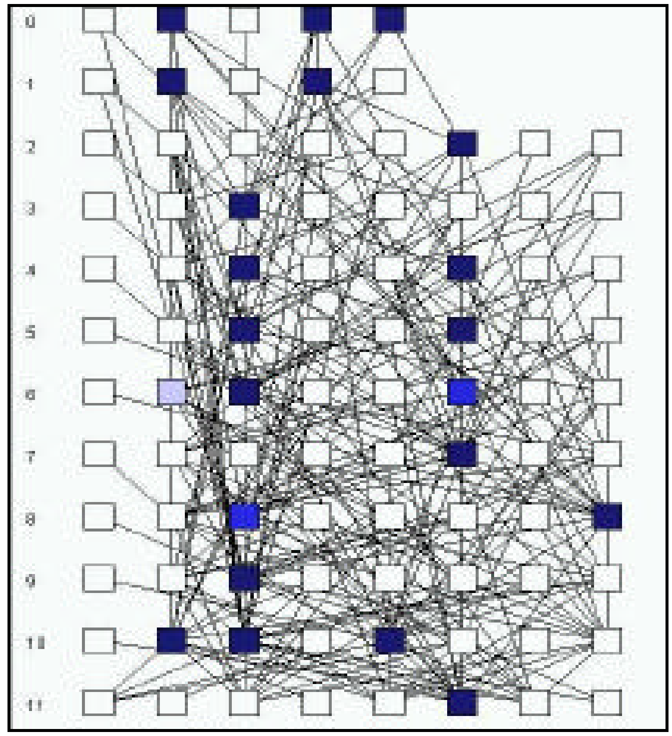


Figure 4. Classicist Mode of Communication

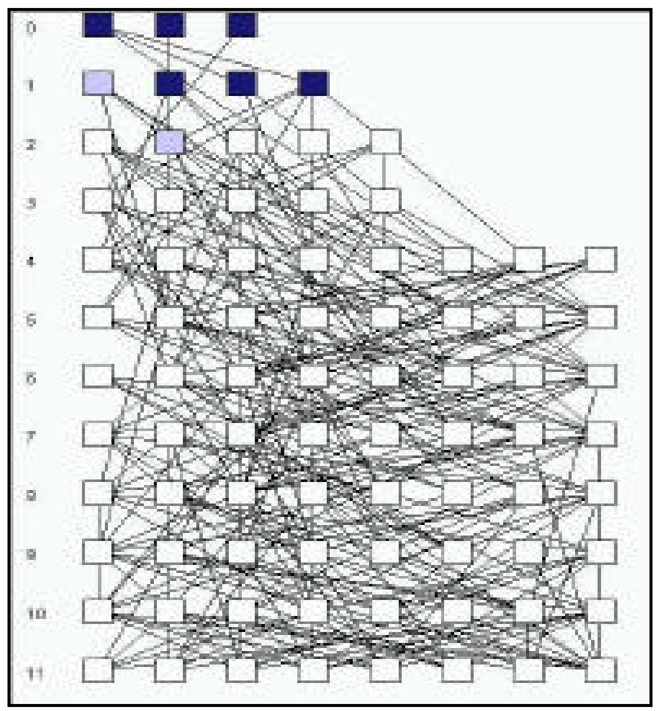


Figure 5. Modernist Mode of Communication

5.2 Simulation and Visualization

We have developed an simulation tool – the COM Test Environment (COMTE) - to simulate, analyze, and visualize communication processes and their basic patterns. COMTE is a prototype and includes, up to now, just the most basic features

[7] We would like to skip the mathematics here and refer interested readers to the project homepage, where a short paper introduces the mathematical basics [4].

[8] A darker shading signifies a higher social visibility.

of the COM theory (see figure 6.). These features include a limited range of usable visibility functions, a distribution function to simulate the production of message signs and the inscription of references by autonomous agents, an interface to change the values of certain factors in the visibility and distribution functions and to change these functions themselves, and a simple visualization of the ongoing communication process and the resulting distribution of visibility between the message signs. A high social visibility is indicated by a darker shading of the message sign. It is possible to filter invisible messages and references to enhance the orientation and to highlight the resulting patterns.

This prototypical simulation tool enables us to analyze and simulate simple patterns in communication. What we would like to do in the future, is to expand this prototype in several directions. One direction is the inclusion of an aggregate level, where a couple of new structural aspects could be visualized; another direction is the differentiation of types of message signs and reference types from an each other; we would also like to include an interface for empirical data and real run-time communication in MAS; and finally, there should be a possibility to show process differentiation in communication. The visualization of these features is vitally important, because the selection problem of the users can only be addressed by some form of transformation of social visibility into cognitive visibility. This transformation is directed via the visualization of the emergence and reproduction of patterns of social visibility by the COMTE tool.

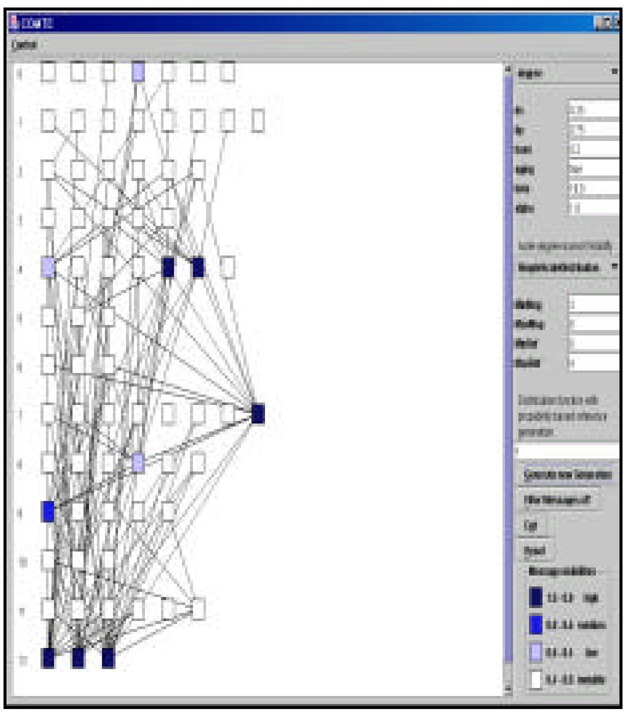


Figure 6. COMTE Interface

6. CONCLUSION

In this concluding section, we would like to point out some aspects of the possible relations between the field of computer semiotics, socionics, and communication-oriented modeling. With reference to some articles published in previous COSIGN-Proceedings [2, 11, 13, 21, and 23] and to Andersens groundwork on computer semiotics [1], we try to outline similarities and complementarities concerning the foundational scientific questions and the possible ways to answer them.

To focus on the central questions: Socionics is mainly concerned with the use of social mechanisms (in the sense of Hedström & Swedberg [10]) or more generally social theories for the development of "better" multi-agent systems; communication-oriented modeling is essentially about simulating the self-organization of communication processes by a modeling approach that focuses on the cross-references between observable message signs; and computer semiotics tries to establish an new branch of semiotics by applying semiotic concepts to computer-based signs.

Beyond the obvious differences of the three fields, there is also a striking similarity: All three approaches deal with computers and signs[9] play an important role in them. Based on that common features, we would like to point out some research agendas of interest for socionics and COM inside the field of computer semiotics (represented by articles published at previous COSIGN conferences).

Of general interest for socionics is the article of Petric et al. [21] with its direct references to the role of signs for the development of multi-agent systems and "socially intelligent agents". As noted above, we think that sign usage is of high significance for the socionic research agenda. The introduction of terms like semiosphere and cultural encyclopedia could lead to a greater concentration on the modeling of environments for agents instead of a purely agent-oriented modeling. This seems in some respects quite similar to our own intentions behind communication-oriented modeling.

Another point of connection between communication-oriented modeling and computer semiotics is a focus on visualization. There were a variety of papers concerning the construction of visual signs that enhance the cognitive visibility of information in computer-mediated communication. The "Crystal Hy-Map™", introduced by David Bihanic [2], gives some interesting hints for the creation of a visual representation of aggregate data on structural aspects of communication, like semantics or process differentiation. Another interesting approach in the area of visualization of meta-data is demonstrated by Kerne and Sundaram [11] and their "CollageMachine" that gives a visual representation of browsing activities by recombining the constituent media elements of a variety of internet documents.

Two further branches of computer semiotics could be of considerable interest for our project. On the one hand, the problem of transforming social visibility into cognitive visibility is not only a question of representing meta-data, but also a question of interface design, a theme of some relevance in computer semiotics (see Scalisi [23] and Andersen [1]). On the other hand, computer semiotics seems to be as interested in the structural aspects of computer-mediated communication as is communication-oriented modeling. So we are quite interested in work like the paper of Lucia Leao [13] that analyses the structure of hypermedia communication from a semiotic point of view.

[9] Signs could be viewed as a vitally important social mechanism and some form of signs or symbols are relevant in a lot of socionic projects (Kron [12] shows some examples).

So that we could finally summarize, that we think that communication-oriented modeling and socionics can profit from the work done in the field of computer semiotics and that both approaches share an interest in sign usage as an organizing principle in communication. Especially COM shares a considerable interest in the question of visualization with computer semiotics, because every support for agents trying to solve their selection problems in computer-mediated communication depends on the transformation from the social to the cognitive level, in our case on the transformation of social visibility into cognitive visibility and the achievement of an enhanced form of coupling between communication and perception.

7. ACKNOWLEDGEMENTS

Firstly, I would like to thank my project partners Prof. Dr. Christoph Schlieder, Dr. Johanna Künzel, and Dr. Klaus Stein at the University of Bamberg and Prof. Dr. Thomas Malsch, Maren Lübcke, Steffen Albrecht, and Rasco Perschke at the Technical University of Hamburg-Harburg. To point out that COM is a project of a community of researchers, I used the personal pronoun "we", rather than "I" throughout the paper.

Secondly, I would like to thank the Deutsche Forschungsgemeinschaft (DFG) for funding and supporting this project and all the other members of the various projects in the Socionic Research Program for the development of an interesting field of research.

8. REFERENCES

[1] Andersen, Peter Bogh. A Theory of Computer Semiotics. Cambridge University Press, Cambridge GB, 1997.

[2] Bihanic, David. A Complete System of Tridimensional Graphical Representation of Information: "Crystal Hy-Map™". COSIGN 2003.

[3] Coleman, James S. Foundations of Social Theory. Belknap Press of Harvard University Press, Cambridge MA, 1994.

[4] Communication-Oriented Modeling. Project-Website. URL: http://www.kinf.wiai.uni-bamberg.de/COM/

[5] Eco, Umberto. Kant and the platypus. Essays on Language and Cognition. Vintage, London, 2000.

[6] Eco, Umberto. Semiotics and the Philosophy of Language. Indiana University Press, Bloomington IN, 1984.

[7] Esser, Hartmut. Soziologie. Allgemeine Grundlagen. Campus, Frankfurt am Main, 1993.

[8] Ferber, Jacques. Multi-agent systems : an introduction to dirtsibuted artificial intelligence. Addison-Wesley, Harlow et al., 1999.

[9] Giddens, Anthony. The Constitution of Society. Polity Press, Cambridge GB, 1984.

[10] Hedström, Peter/Swedberg, Richard (eds.). Social Mechanisms. An Analytical Approach to Social Theory. Cambridge University Press, Cambridge (GB), 1998.

[11] Kerne, Andruid/Sundaram, Vikram. A Recombinant Information Space. COSIGN 2003.

[12] Kron, Thomas (ed.). Luhmann modelliert. Sozionische Ansichten zur Simulation von Kommunikationssystemen. Opladen, Leske+Budrich, 2002.

[13] Leao, Lucia. The Labyrinth as a Model of Complexity: The Semiotics of Hypermedia. COSIGN 2002.

[14] Luhmann, Niklas. Social Systems. Stanford University Press, Stanford CA, 1995.

[15] Luhmann, Niklas. Zeichen als Form. In: Dirk Baecker (ed.). Probleme der Form. Suhrkamp, Frankfurt am Main, 1993.

[16] Malsch, Thomas (ed.). Sozionik. Sozionische Ansichten über künstliche Intelligenz. edition sigma, Berlin, 1998.

[17] Malsch, Thomas. Naming the Unnamable: Socionics or the Sociological Turn of/to Distributed Artificial Intelligence. in: Autonomous Agents and Multi-Agent Systems 3/2001. pp. 155-187.

[18] Malsch, Thomas/Schlieder, Christoph. Communication without Agents? From Agent-Oriented to Communication-Oriented Modeling. In: Lindemann, Gabriela/Moldt, Daniel/Paolucci, Mario (eds.). Regulated Agent-Based Social Systems. First International Workshop, RASTA 2002, Bologna, Italy, July 2002. Revised Selected and Invited Papers, pp. 113-133. Springer, Berlin et al., 2002.

[19] Mead, George Herbert. Mind, Self, and Society from the standpoint of a Social Behaviorist. Edited by Charles Morris. University of Chicago Press, Chicago, 1934.

[20] Peirce, Charles Sanders. Peirce on Signs. Writings on Semiotics by Charles Sanders Peirce. Edited by James Hoopes. University of North Carolina Press, Chapel Hill NC, 1991.

[21] Petric, Mirko/Tomic-Koludrovic, Inga/Mitrovic, Ivica. A Missing Link: The Role of Semiotics in Multiagent Environments. COSIGN 2001.

[22] Saussure, Ferdinand de. Course in General Linguistics. McGraw-Hill, New York, 1966.

[23] Scalisi, Raffaella. A semiotic model for interface design. COSIGN 2001.

[24] Simon, Josef. Philosophie des Zeichens. Walter de Gruyter, Berlin, New York, 1989.

[25] Weber, Max. Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie. 5. Aufl. Studienausgabe. J.C.B. Mohr (Paul Siebeck), Tübingen, 1980.

[26] Weiss, Gerhard (ed.). Multiagent systems: a modern approach to distributed artificial intelligence. MIT Press, Cambridge (Mass.), 1999.

A Semiotic Experiment

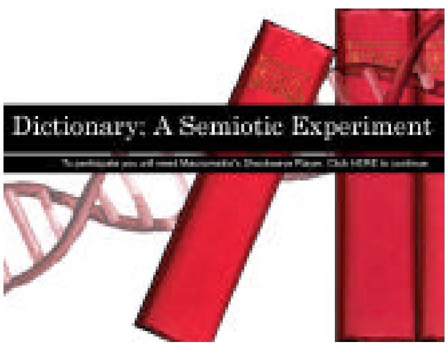
Rebecca W. White
Art Media Studies
Syracuse University
Syracuse, NY13224
rwwhite@syr.edu

ABSTRACT

Words and language are products of societal necessity. Without structured communication devices, we would have misunderstanding and chaos. “Dictionary: A Semiotic Experiment” is an interactive multimedia piece that explores the structure and manufacture of language and the ability of language to portray concepts, thoughts and ideas.

Keywords

Semiotics, Interactive Art, Multimedia, Computer Art, Web Art



Splash Page, *Dictionary: A Semiotic Experiment*

1. INTRODUCTION

Words are essential, yet, arbitrary. We require language to communicate ideas, however the multitudes of languages that exist limit the free-flow of information exchange between all persons of all cultures. Most societies/cultures have their own beliefs and stories that explain this discontinuity of language and very existence of words.

Christians believe that, according to Genesis 2:18-23, God gave Adam domination over the animals and on the sixth day of creation brought each animal before Adam for him to name (all before Eve was ever created). This account is the first in the Bible tackling the creation of words and demonstrates a patriarchal approach to the development of language. Genesis 11:1-9 also explains the transformation from this single language of Adam to many in the story of Babel. The inhabitants of Babylon attempted to build a tower to heaven, so that man would become equal to God. Man’s greed of power led to his downfall, once again, and his ability to understand (one language) was revoked. Many religions and cultures have their own histories that explain or suggest how language came to be.

Whether or not one believes in a creation story is irrelevant, what is important is to understand the arbitrary nature of language parables/myths attempt to explain in the first place. According to Genesis, Adam had free reign to name the animals. Is there any reason why the word bird signifies an animal with wings? A bird could have easily been named “tree,” “rain,” “igloo,” or “lamp.” Our systems of language/communication are sets of signs, symbols and rules collectively accepted and utilized.

2. CREATION

Who creates alphabets, words and grammar? If language is a collective agreement and collectively used, shouldn’t the creation of language also be a collective activity?

If we look at computer languages like JAVA, C, C++, etc. authors can develop extensions to a pre-existing skeleton of terms, commands and rules. There is a mutual understanding of a grammar/ordering of commands that allows the computer or other programmers to understand. Likewise, in English, there is a predetermined alphabet, predetermined words and a predetermined grammar, however the letters and words can be strung together in infinite combinations and compositions. New words are added to the dictionary over time as words infiltrate mainstream culture.

D:ASE [1] utilizes components from both spoken languages and written/computational languages to create a unique procedural language.

2.1 Birth: First Sounds

Alphabets can be made up of letters, or images that signify a specific sound. D:ASE explores the concept of “alphabet” as a pronunciation guide (see **Figure 1**). Each character is a phoneme, a signifier of a specific sound. Like Hieroglyphics the phonemes are images. Each image is a map of pronunciation representing the labial, dental and velar points of articulation used to create a specific sound. For example an English sounding long “o” is represented in **Figure 2** below. The top third represents the open circular shape of lips, the bottom two thirds demonstrate an open airway, where the tongue and palate do not touch.

Figure 1. D:ASE Phonemes/Alphabet

First published at COSIGN 2004
14 – 16 September 2004, University of Split (Croatia)

[1] *Dictionary: A Semiotic Experiment*

Sounds/phonemes are procedurally strung together to form words. Sounds are pre-recorded and pronounced in the feminine voice by the computer. The voice is human, yet, systematically pronounced in a robotic, crisp, explicit and articulate manner to avoid the morphology or simplification a human language would develop over time. D:ASE is not vocalized in the masculine, or technically, the human. The denial of the human/masculine is a rejection of traditional ways, masculine creator and originator that the Judeo-Christian tradition preaches. The feminine and computational voice signifies an "other," a language that is not spoken, yet is heard and visualized.

Figure 2. D:ASE Phonemes

2.2 Growth: Forming a Vocabulary

A vocabulary, or a repertoire of words is not formed overnight. A word is essentially a sign the collective agrees upon to signify a particular thought, idea, object or concept. D:ASE is a database of equality and encourages everyone of every background to create words and include them in the lexicon. Anyone who visits and interacts with D:ASE is entitled to, and has the opportunity to, append the dictionary database (**Figure 3**).

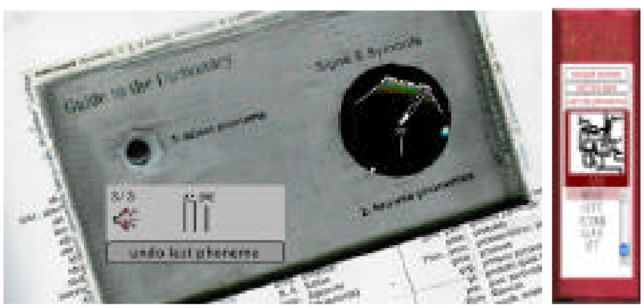


Figure 3. Left: Interface to create words. Right: Dictionary Database Interface with Image Definitions

An individual, however, cannot simply submit a combination of any symbols and call it a word. There are rules, a structure, and a particular alphabet that need to be abided by.

Words are meaningless without definitions. In Western cultures words are defined by other words, which are defined by more words. In many cases words can signify a variety of things; for example, the English word "boat" can signify a gravy boat, a banana boat, a sailboat, a canoe, a kayak, a motorboat, a yacht, or a cruise liner. The clarity of definition, or signified, therefore, is often vague. D:ASE recognizes, reveals and exploits this ambiguity. A user created, black and white pixel image, defines each word. In essence, the word and the definition become a pair of signifiers of a real event, activity or object.

Images help the individual to think beyond the limitations of words. The purpose is to exercise the subconscious visualization of meanings and concepts. It forces the individual to recognize how s/he thinks and conceptualizes ideas and demonstrates that each individual does not envision the same word the same way. The pixel-by-pixel drawing acts as a filter to mediate language. It forces a level of simplicity into each image/definition to maintain a similar level of readability/usability between all definitions that the specificity of a photograph would not permit. Image as definition restricts the easy translation of one language into another. For example, each individual does not visualize an emotion in the same way. Often emotions are understood based on an individual's own life experiences and of those around them. My understanding of grief, for example, may be much more shallow than that of someone who has lost a parent or a child.

A quick and easy experiment to demonstrate how often language is misunderstood or interpreted can be performed by simply asking the next ten people you come in contact with to draw a picture of a shoe. Each of those people have an immediate preconception of what a shoe is, however, among people that conception can greatly vary. A shoe is a general term that is often used to signify something particular. For example, "Go put on a pair of shoes" could mean any number of things: put on pumps, tie-up shoes, sneakers, boots, high-heels, slip-ons, sandals, flip-flops, shoes with buckles, shoes that Velcro, clogs, masculine shoes or feminine ones. Although this is a fairly harmless and obvious example, miscommunication due to the limitations of language happen all the time, and often with negative results. D:ASE highlights this characteristic of language by obliging users to interact with a language s/he is not familiar with and offers no explanation or translation of a word into another language. Users are left to interpret image definitions blindly and creatively, exemplifying the idiosyncratic nature of language, despite cultural attempts to unify, simplify and clarify communication systems.

2.3 Parenthood: Procedural Communication

Computational languages are the best representations of very regular and specific grammars. Rules are integral to the ability of a programmer to write a program that can be read by the computer. If the sequencing is not correct, then the program will fail. Computers analyze systematically and will malfunction when an error has occurred. The computer does not attempt to guess or interpret anything that does not follow procedure. Irregularities and slang make it especially difficult for newbies to understand and utilize language.

Rules and procedure are integral to the clarity of D:ASE. Words are constructed out of three, four or five letters, depending on its type: action, subject, or descriptor.

Action: 3 Letters
Subject: 4 Letters
Descriptor: 5 Letters

Actions are analogous to the English language's verbs, the subject comparable to a noun, and descriptors parallel to an adjective or an adverb. The ordering of words to form sentences is also regular; irregularities would result in further ambiguity and miscommunication. Sentences are given a tense as a whole, rather than specifying the tense of verbs in agreement with nouns. A sentence is indicated as present tense by writing an arrow pointing down. A sentence is placed in the future with an arrow pointing to the right; an arrow pointing left indicates past tense. No arrow defaults as present tense. Descriptor words applying to the subject are placed after the tense indicator and before the subject. The subject is followed by descriptors applicable to the action. The action is always the last word at the end of a sentence (see **Figure 4**).

Figure 4. Procedural Grammar Diagram

A systematic ordering should restrict the confusion complex compositions nurture. Precision is obtained by a standard structural convention. Although the individual defines words in D:ASE, the collective must agree upon and utilize rules of construction to ensure accessibility to the entire population of users.

3. Life and Living

In many ways the concept of language is a belief system like a set of laws or a religion. The growth and existence of words and language is a direct result of its practice and use. Many languages die and have revivals; many words fade into oblivion and several new ones develop. D:ASE is not an exception to this pattern. After the concept of a communication device is created it is dependent upon the society, or the set of users, to cultivate and energize it.

Although D:ASE is the offspring of a woman, that woman no longer has creative control. As the creator she determined rules of government and structural guidelines, however, the innards and the spirit are out of her hands and into the hands of users of all cultures, all backgrounds, and all sexes. The only guaranteed characteristic each user has in common is that s/he has access to the World Wide Web. Once a belief system becomes accessible to a population it can grow, segregate, mutate, or be forgotten, but the creator no longer has an off switch to completely destroy it.

D:ASE is a database, a feminine voice, an experimental language, an alphabet, a systems of signs, an interactive multimedia website, a procedural grammar, an exchange of viewpoints, and a living system.

4. ACKNOWLEDGMENTS

The Art Media Studies Department of Syracuse University hosts a Dictionary: A Semiotic Experiment, on the web at cgr.syr.edu/~dictionary. D:ASE is a Master of Fine Arts project created at Syracuse University by Rebecca W. White in 2003-2004.