AISB'99 Convention

6th-9th April 1999
Edinburgh College of Art &
Division of Informatics, University of Edinburgh

Proceedings of the AISB'99 Symposium on Creativity in Entertainment and Visual Art

Published by
The Society for the Study of
Artificial Intelligence and
Simulation of Behaviour

`http://www.cogs.susx.ac.uk/aisb/`

ISBN 1 902956 01 X

Contents

The AISB'99 Convention

The AISB'99 Convention has provided an opportunity to focus on one of the Cinderella's of the AI world – the study of human creativity. The form of the event – a set of short symposia and workshops running concurrently – has made it possible to bring together a significant number of researchers interested in themes which, though normally considered different areas of AI and Cognitive Science, have in common the theme of creativity.

This, coupled with the contribution of invited speakers of the highest international calibre – Prof. Margaret Boden (Creativity and Evaluation), Dr. Ian Cross (Musical Creativity), Prof. Harold Cohen (Creativity in Visual Art) and Prof. Mark Turner (The Literary Mind), not to mention the many speakers invited to the individual symposia and workshops – has made the Convention an exciting and worthwhile event, as can be clearly seen from the quality of the work presented in this volume.

The Convention included symposia on Musical Creativity, Creativity in Entertainment and Visual Art, Creative Language, Creative Evolutionary Systems, Scientific Creativity, Imitation in Animals and Artifacts, and finally Metaphor, AI and Cognition. Alongside the explicitly creative focus, we are pleased to include in the Convention the 6th Workshop on Automated Reasoning, a Workshop on Reference Standards for NLP, and a Workshop on Teaching Cognitive Science to Undergraduates. The proceedings of all these events are published by AISB, whose Web site can be found at `http://www.cogs.susx.ac.uk/aisb/`.

The organising committee would like to thank Edinburgh Projects, the research wing of Edinburgh College of Art, and the Division of Informatics, University of Edinburgh for their generous support of the event. Our sincerest thanks also go out to the symposium chairs and committees, without whose hard work and careful cooperation there could have been no Convention. Finally, and by no means least, we would like to thank the authors of the contributed papers – including those which were regretfully not eventually accepted.

Andrew Patrizio, Geraint A. Wiggins & Helen Pain

The AISB'99 Symposium on Creativity in Entertainment and Visual Art

The entertainment industry is increasingly recognized as one of the most prominent application areas in which the power of computing plays a key role. So far, research in AI and Alife has mainly applied agent technologies to the domain of entertainment, integrating techniques from a range of disciplines (*e.g.*, animation, film, literature, theatre). The focus of this strand of research is to provide users of virtual environments (mainly games and interactive story-telling applications) with entertaining and engaging virtual counterparts which dynamically adapt to a user's input, or offer multi-variegated expression.

However, not only do we have to get a far better understanding of the subjective and creative aspects of the artificial protagonists within applications, we also have to provide machines with competence in many other entertainment industry tasks presently performed by humans. In spite of pressing demand, research in AI so far has not been active enough in discovering new ways of applying and supporting creativity tasks in entertainment environments.

In addition to inciting new efforts in academia, this symposium intends to raise mutual awareness and establish an ongoing dialogue between research institutions and commercial developers, so as to identify current trends and pave the way for novel developments in this exciting domain.

As well as the papers presented in this volume, the workshop featured two video presentations:

M. Reinhart & V. Widrich	**tx-transform** This is not the ordinary stuff out of which crime stories are made, it is a way to see crime stories in an entirely new way.
Lillian Schwartz	**Beyond Picasso** Lillian Schwartz reordered and combined angular contours, broken planes, and distorted proportions in her own pictorial structures to go beyond Picasso.

The Organising Committee

Organising Committee:

Frank Nack, GMD-IPSI
Elisabeth André, DKFI
Paul Brna, University of Leeds, UK
Kevin Brooks, MIT, USA
Ana Paiva, INESC, Portugal
Paolo Petta, ÖFAI, Autria

EVOLVABLE BELIEVABLE AGENTS FOR BROADCASTED INTERACTIVE GAME SHOWS

Riccardo Antonini
Consorzio Roma Ricerche
Via O. Raimondo, 8
00173 Roma (Italy)
tel. 0039-06-7234606
e-mail: Riccardo.Antonini@UniRoma2.it
http://193.204.117.76/Amuseaw

Abstract

The possibility of interactive digital shows creation at a cost comparable with the creation of 2D web pages is discussed in the present paper. The reduction of cost is due to several reasons. 1) Reduction of bandwith. The mathematical model of the network is discussesd in some detail. 2) A further reduction of cost is decribed. This is made posible by using an asymmetrical channel that uses data broadcasting for 'one-to-many' communication and the Iternet for one-to-one. 3) The creation of believable characters by means of artificial life techniques.The internal model of the characters is described along with the theoretical foundations of the techniques used for the charachter evolvability. A simple interactive interface allows show-developers, as well as participants in shows to cast their own characters without the need of programming. The characters will also continue to evolve indefinitely. Finally, the game show is only a testbed for the technology that could be used for other types of shows as well.

1 Introduction

In that paper we will discuss how we have studied the possibility of "broadcasted interactive games shows" creation. In particular, we will focus on the role of "evolvable believable agents." The discussion of some details on how we have implemented the air and satellite broadcasting of digital game shows will also be given.

The need of evolvable believable agents in games shows comes from the increasing demand of such type of entertainment along with the possibility to exploit to full extent both the interactivity of computers linked to a network and the intrinsic lower bandwidth of the implementation of shows that makes use of digital characters.

The concept of the "believable agent" has been introduced in (Bates, 1994) and our approach to collaborative evolutionary systems has been derived from Dawkins Biomorphs (Dawkins, 1986).

Real game shows has been, historically among the first entertainment shows to be broadcasted by television.

What does "digital play" mean in the framework of the Amusement project? The answer, that comes at the end of a long historical evolution of the entertainment culture, is simple. Bring the best of the past (mostly the possibility of interaction among spectators and between spectators and actors) along with the best of modern technology (namely the possibility of interaction at a distance).

Normally, in a game show, there are three types of fundamental roles: players, a conductor and the audience. The audience can be divided, in turn, into two groups: in site audience and at home audience. This was an "escamotage", implemented since the beginning of the TV era, aimed at solving the dicotomy of having an interactive audience that could be able, for example, to make an applause, (such an audience must have been in the same physical site of the players) and a vast audience (that may have been everywhere else).

Now it is worth noticing that the crowd sitting in even the world biggest stadium is order of magnitude smaller than an average TV audience. The trade off was hence, to have both one interactive (and small in size) and one larger (and completely passive) audiences. Today it is possible to have an audience both large and interactive at the same time. In the digital age we have another advantage that has been somehow overlooked.

Today it is possible not only to convert an analog image of a show into a digital stream of bits, as today's digital TV normally does (the relevant information is, strictly technically speaking, digital, of course, but, logically, the information is organized, simply, as an "analog" of the real play, since they do nothing more than replicate the image of the real play), but also to create, as we do in the Amusement project, a digital show directly from scratch using digital actors.

In the Amusement project a digital play is created from scratch using digital actors as discussed in the following passages.

The digital actors and the environment are replicated only once, at the beginning, in all the "clients" and all that needs to be re-transmitted is their "dynamics."

This type of approach lowers tremendously the bandwidth, reducing it by orders of magnitude. In section 4.3 a mathematical model of the network requirements is described. The associated cost of such a transmission will be so low that an entire new class of productions is now possible. For example, even elementary schools can easily raise the few hundred dollars necessary to transmit worldwide a digital play.

Much in the same way that now almost anybody can publish written text and 2D images whereas before he had to rely on the press and a publisher. This sort of democratization of the press is now possible for the entertainment industry as well. We are at the edge of a new "Gutenberg Galaxy" to paraphrase McLuhan's words. (McLuhan, 1962.) We have in the Amusement project four thousand Italian schools already equipped with the receiving hardware and software and ready to begin the productions. Those schools will be the first but of course, not the only audience of the said productions. The game show has been proposed as a valuable tool for what is now called "edutainment" that is "educational-entertainment").

In order to mantain this possibility of economic game show production, anyhow the digital characters, that are intrinsically cheap to be broadcasted need also to be programmed in a non expensive way.

On the other hand, people expect, from shows in general, the vast richness of characters that cartoons, if not real actors, have let us to be used to. One possible solution is, hence, to program into the characters some sort of "artificial life."

In the Amusement project we have successfully applied "collaborative evolvable procedures" to let the users of a game cast the gestures that they want their avatars to show in the game itself. This type of technique can be applied not only to closed group interactive games, as is the case, for example, of networked computer games, but also to true broadcasted game shows.

2 Implementation of the Collaborative System

How does this happen? First of all, the game can be played between avatars (behind whom there are real players), or between avatars and "bots" (behind whom there are artificial life computer programs).

Anyhow, most of the time they also need a conductor, or moderator of the game. Even in the case in which the game is very well structured and, hence, the conductor can be efficiently implemented as a deterministic "classical" AI algorithm, Artificial Life technique can be used to let its behavior be a "believable character." (Bates, 1994) For example, in the Amusement project the gestures of any character are not defined beforehand, but they evolve genetically. In particular, hence, the gestures of the conductors can be triggered by an internal model relevant to the "algorithm of the game," but they will not always be the same since they are evolving throughout the game. For example, a gesture welcoming new participants, will never be the same in all the game instances. Moreover, the participants themselves will not need to (explicitly) program the gestures needed in the game. This allows them to be concentrated on the game itself. At the same time their expressiveness can be very rich, because they can easily, starting from a first set of default gesture genes, evolve their own gesture sets. In fact, using as selective pressure their preferences for some gesture instead of some others, their gesture set will evolve allowing effectiveness in communicating and or disguising their real moods during the game.

Moreover, the association of a given situation (say, the proximity to an other avatar or an object) to a given gesture can be regarded as a behavior (Kosko, 1992). This association in the Amusement project models is controlled by a genome and hence, is susceptible to evolution itself.

Of special interest to those of us in the world of games and entertainment, is the fact that showing the user's real mood can be counter-productive. In order to provide an avatar with the ability to hide emotions whenever it is necessary, we have allowed the user to define their real mood along with the mood that they want to simulate. The resulting avatar's expression and actions will depend on how gifted it is in dissimulating. The association of a real mood and the expressed mood is controlled in the Amusement implementation, by means of a deterministic, albeit "fuzzy logic" internal model.

The core of the authoring tool generating the gesture by means of evolutionary technique is described in sections 4.1 and 4.2 whereas in section 4.3 we will discuss the details of the broadcasting of interactive game shows.

3 Technical description of the system

In the following a detailed description of the system will be given. The system has been implemented making use of existing interactive virtual worlds' platforms as well as original proprietary technologies. The general cabled networked architecture has been tested using existing commercial or experimental virtual world platforms. The evolutionary engine is a proprietary software source coded in C. The data broadcasting part makes use of proprietary hardware and technologies. Broadcast data transmissions are delivered by Italian state owned channel RAI3 (freq. 11.534 GHz pol. V) on Hot Bird 1, 13 degree east.

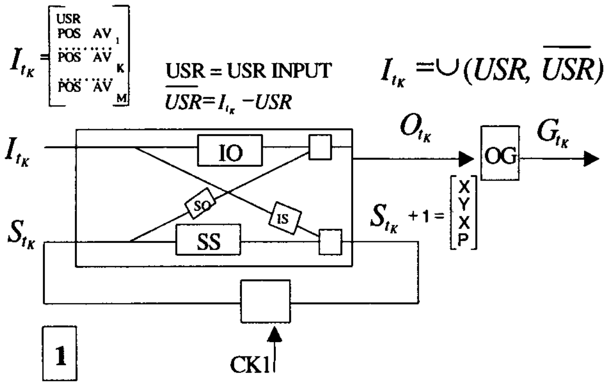
It is needed an analog satellite receiver with radiofrequency output (UHF) tunable on standard terrestrial channels (e.g. channel 21 or 45 or 52); analog satellite receivers with fixed (or trimmable) output. Channel 36 cannot be used. It is better to use at least a 90 cm diameter dish, with 0.8 dB (or better) noise figure LNB.

3.1 Our approach to "intelligent behavior"

Our approach to "intelligent behavior" is based on the chapter "Intelligent Behavior As Adaptive Model-Free Estimation" found in (Kosko, 1992.)

In Kosko's view (the notation is slightly different from Kosko's), a behavior (b), defined as a correspondence between a given situation (or stimulus) ($s \in S$) and a response ($r \in R$), can be modeled as a function b that maps S onto R, being S the set of all the possible Stimuli, and R the set of all the possible Responses. That is:

$$b : S \rightarrow R$$



Regardless of the complexity of the dynamics linking the input (situation) to the output (response), at the most abstract level, it is possible to think of a behavior as a mapping between pairs of vectors (s, r). The correspondence is usually obtained by means of a neural network or a fuzzy system. We have used Fuzzy Associative Memories (FAMs). Hence in our work b is a FAM matrix. The genome generating the matrix b evolves by cumulative selection of small changes analogously to what is described in (Dawkins, 1986) for the morphology of artificial creatures. R is usually called gestures. R also evolves by means of a similar procedure. The user, or the game developer, first uses an interactive procedure of the type described in (Dawkins, 1986) in order to cast the set of gestures needed and hence, by means of a similar procedure the mapping b.

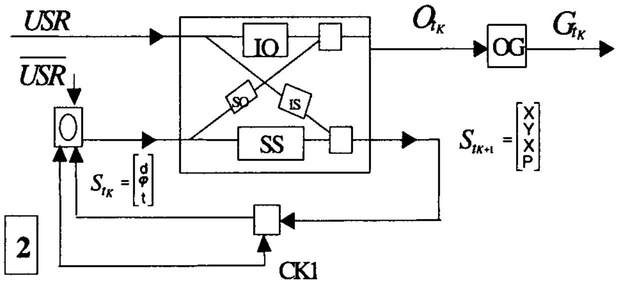
One of the advantages of this approach is that it is model-free. This is particularly useful given the complexity of the task of modelling intelligent behavior. Modeling intelligent behavior, moods and emotions is extremely interesting, but it is out of the scope of this paper.

The term "model-free" should not be taken too literally. What is really model-free is the b function. Some sort of modeling (dimensions of input/output spaces, for example) is obviously necessary.

3.2 How Evolution Arises from Avatars Interaction

Avatars who are in our Active Worlds' world "Amuse" and "Amuse 2" move driven by users instructions (via mouse or other pointing devices) or by internal logic or both, depending on the functioning mode selected. We have defined three different modes called mode 1, 2 or 3 respectively. Except for debugging and experimental purposes their mode is not shown to the user.

Regardless of the functioning mode, the pointing device acts only as global moving device. The user can direct only the global displacement of its avatar and not the movement of its single parts (for example, the expression of "moods" via a predefined, but evolvable gesture or the walking style pattern).

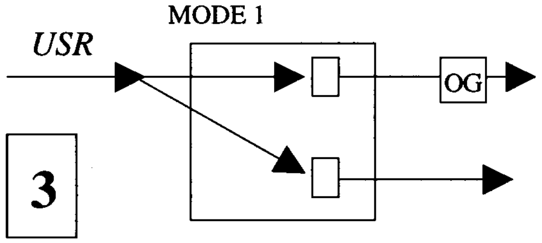


The behavior (i.e. response to a given situation) is completely determined by both the internal state of the avatar and the situation. I.e., if one avatar moves towards another it is up to the user's avatar, to decide whether or not to greet the other avatar and how.

3.2.1 Mode 1 (Puppet Mode)

In this Mode, avatars are driven by user's pointing device commands in order to interact with other avatars. Evolution is still possible for the avatars functioning in this mode but, the outcomes of the evolution will not show until their mode is switched to an higher mode. Anyhow, they will influence both of the avatar's behavior, in the short term, (i.e. they make other avatars react in a certain way) and the evolution of behavior in the long term (i.e. they way they react to a given situation).

PUPPET

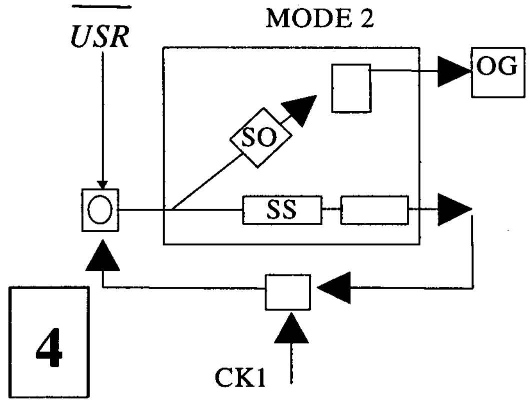


Provided the others are functioning in modes 2 and 3, the influence of mode 1 on mode 2 and of mode 1 on mode 3 will be soon visible. This is mainly a test and a training mode.

3.2.2 Mode 2 (Ro-Bot Mode)

In Mode 2 the avatar accepts no direct pointing device input. It acts according to its internal state and the interaction with other avatars.

(RO)BOT



3.2.3 Mode 3 (Dog Mode)

In this Mode the user merely suggests, by means of a pointing device, what the avatar should do (i.e. pointing towards a given direction or goal). The avatar, in turn, interprets the command and decides how to do the task (in other terms he can "choose how to get there, how to walk and what to do in the meantime.") In principle its freedom could arrive even up to the point of ignoring the user's command. We have not implemented such a mode insofar. We have called it high 3 mode. We have implemented only low 3 mode up to now.

A more formal discussion of the internal model follows.

3.2.4 Formal discussion of the internal model

The generic internal model is illustrated in figure 1. In order to explain the workings of the internal model, we will describe it from the point of view of a generic user.

The user sees the avatar's physical appearance, the displacements and rotations of its whole body *inside* the environment. Such displacements and rotations are called *movements.* Moreover, the user can see the rotations of the various body segments of the avatar. A given set of such rotations is called a *gesture.* During a session in a Virtual World what the user perceives is a combination between movements and gestures.

The user, in turn, can interact with both the environment and the avatar by means of a generic pointing device.

The model has also an internal state that, for the sake of simplicity, we have limited in this first part of the Amusement project to the following variables: x, y, z, p (three cartesian coordinates of the avatar's position plus the angle of pointing). The model accepts two sub-sets of inputs: the user inputs (USR), and the other avatar position and pointing ($\overline{USR}$). The transition from one state to the next is regulated by the clock CK1. The CK1 frequency is usually a sub-multiple of the framerate. The framerate is the frequency at which all the scene rendering is recalculated. Typically the framerate is 10 fr/s. The CK1 frequency is usually 1 Hertz, i.e. 1 every 10 frames. The outputs are generated from the inputs by means of four fundamental functions called IO, SO, IS, SS. The gestures are selected from an ordered, finite cardinality, set by the output by means of the OG function. The output from IO and SO, and from IS and SS, is compounded by means of the operator which does not necessarily have to be a sum.

The structure of IO, SO, IS, SS can be diversified and we will discuss only the implementations that are relevant to our application.

Before proceeding we will introduce some little modifications to the model adopted in figure 1. The relevant schematic diagram is illustrated in figure 2.

In this new version the two sub-sets of the input, USR and $\overline{USR}$, have been made explicit. Moreover, the following transforms have been applied, taking into account only the "horizontal" coordinates x and z. The vertical y coordinate is ignored.

The variable t is the total time of interaction and is calculated by summing the CK1 intervals where d and φ are defined as follows:

$$\varphi = p_2 - p_1$$

$$d = \sqrt{(x_2 - x_1)^2 - (z_2 - z_1)^2}$$

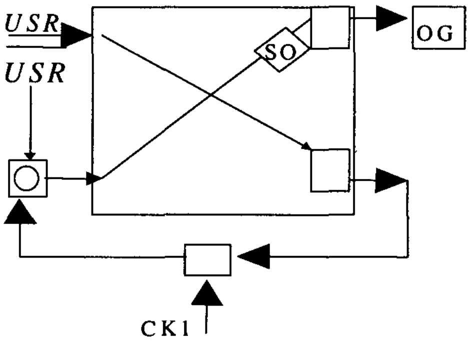
3.2.5 Modes in detail

The simplest case of all is the one we have called *mode 1* (puppet mode, figure 3). In this case SG and SS are null. The link between the output and the input is, hence, direct. The next state, that is the next avatar position and pointing, is specified directly from the user input as well as the the predefined gesture to be performed.

In *mode 2* the avatar is, in reality, a fully independent agent usually called a BOT (taken from the well known word "robot"). The BOT is fully independent and the user has no *direct* way to interact with it. The output and, hence, the relevant gestures, are function of the state by means of SO only. The next state (in practice the next avatar position and pointing), is calculated as a function SS of the position of all the other avatars and/or the present avatar position.

In *low mode 3* the user specifies directly the next avatar state (in practice the new avatar's coordinates), but the gesture is calculated using the state of the avatar. In *high mode 3* the user can still specify directly the next avatar state. But, since now the function SS is present, the new state is a combination of both the user input and the avatar's "will".

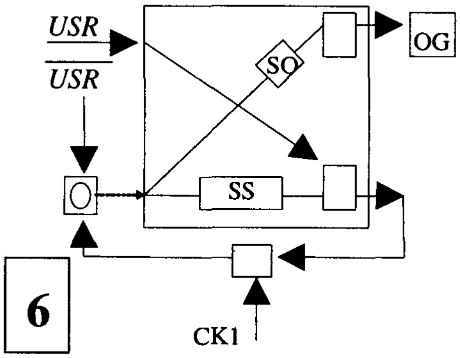
LOW MODE 3



5

The OG block has been made explicit in all modes for the following important reasons.

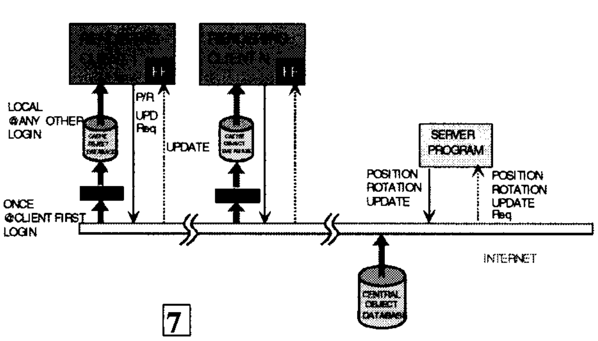
HIGH MODE 3



The OG is, among other things, the memory in which the gestures are stored, regardless on how they have been triggered and/or generated. The SO block is responsible of one the most important parts of the avatar model: the link between a situation, identified with what we have defined here as the "state of the avatar", and a gesture. We will call this association between the situation and the gesture a "behavior" according to (Kosko, 1992). The SO block implements the behavior function b.

3.3 Broadcasting Architechture

The Amusement center has been implemented with a distributed client server architecture. The architecture is truly distributed in the sense that all the different functionalities of the server can be (and, in fact, are) distributed on different Computers, that can be addressed by different IPs, and can be (and, in fact, are) physically distributed at different geographic locations.



7

In the Amusement project the Amusement center server (in the following called for short, simply the Server) is at the most abstract level possible a system with these basic functionalities:

1. Get the information from all users (functionality mode: Many to One).
2. Relay to all users the information generated by all the other users (functionality mode: Many to Many).

In the Amusement project the Amusement center client (in the following called for short simply the Client) is at the most abstract level possible a system with these basic functionalities:

1. Transmit the locally generated information to the Server (functionality mode: One to One).
2. Receive from the Server the information generated from all the other users (functionality mode: One to One).
3. "Render" the information in graphical acoustical and textual modes.

3.3.1 Network Logical description and principles of operation

The logical structure of the network is illustrated in the following block diagram figure 7.

The system is based on a distributed *synchronized* data base. The data base contains the following type of information on both the *avatars* and the *environment*:

1) Geometry
    1.a *Avatar* geometry
    1.b *Environment* geometry
2) Textures
    2.a *Avatar* textures
    2.b *Environment* textures
3) *Avatar* Gestures
4) *Environment* Sounds

The location of the central data base can be anywhere on the Internet (at an address specified by the world administrator). A cache image of the central data base is maintained at any client location. When the client is started, the data base "engine" (of the client), loads data into the rendering "engine" (of the client) from the cached image sending messages to the central data base for update. If no update is needed (i.e. the two data bases are synchronized) the rendering client uses only the cached data. The cached data are relevant only to the "static data", the dynamic data are transmitted through the Internet in the following way.

The rendering client, upon a user request of: translation, rotation or any other attribute change (of an object) sends an update request to the server program.

Note that the server program doesn't need to be running on the same machine as the central data base nor does it even have to be close to it. Indeed they usually are far apart.

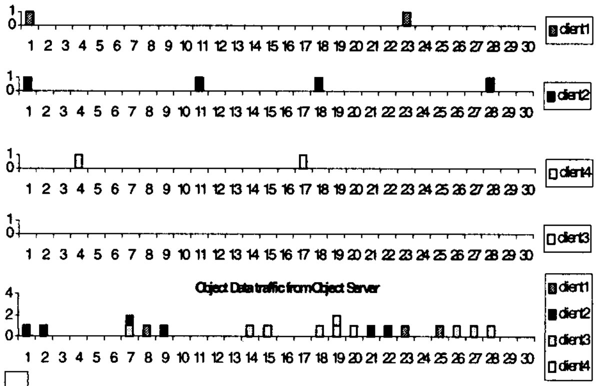
The server program address is identified by a universe data base, maintained by a universe administrator. When a legal, authorized server program is started, on any location on the Internet, the server program asks the universe specified in its ini file to be identified by means of its (that is the server program's) IP address and update the universe data base of IP addresses from which legal authorized worlds are operating.

3.3.1.1 Network requirements

3.3.1.1.1 Standard Cabled Internet Hypothesis

3.3.1.1.1.1 Analytical description from example of (simulated) *Update* data traffic

In order to illustrate the kind of parameters it has to be taken into account in order to estimate the dimentional requirements of a networking able to support the gaming in the Amusement project, an example of a simulated situation of the traffic is reported in the following graphs.



8

First, we will simulate the traffic of Update Requests flowing from clients (towards the server) as illustrated in the first four graphs. The relevant resulting traffic entering the server is illustrated in the last one. The traffic is organized in quanta of constant size.

In the figures the quantum is normalized at the value of one arbitrary unit[1]. At any given time that traffic of Update Requests entering the server is the sum of all the relevant traffic coming out from the clients. (as can be seen in the example figure the traffic from the clients can be zero for many contiguous intervals of time).
The relevant mathematical relationships are reported in the following equation where:

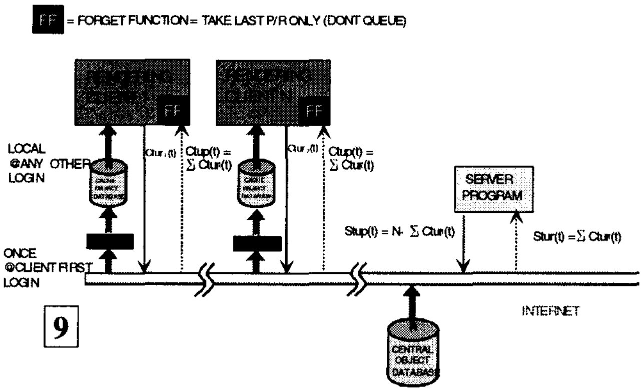
Ctur(t) [Client Traffic of Update Requests] is the traffic of Update Requests flowing from clients at any given time t.

Stur(t) [Server Traffic of Update Requests] is the traffic of Update Requests entering the server at any given time t.

N is the number of clients

$\mathrm{Stur}(t) = \sum_i \mathrm{Ctur}_i(t) \quad i = 1 \text{ to } N$ [eq.1]

As can be seen, the traffic coming out from the rendering client to the server program is very low, being relevant only to translation and/or rotation of avatars.
The relevant traffic, seen from the server, is rather high, albeit still growing less[2] than linearly with the number of clients, as shown in [eq 1] and illustrated in the figures.



On the other side of the net, the server program receives the sum of all clients update request, queues them and re-send, as updates, a copy of all of the updates to all the clients (that he knows that are running on its world). The situation here is dramatic since now the growth is exponential with respect to the number of clients N.
The relevant mathematical relationships are reported in the following equation where:

Stup(t) [Server Traffic of Updates] is the total traffic of Updates copies exiting from server at any given time t.

N is the number of clients

$\mathrm{Stup}(t) = N \cdot \sum_i \mathrm{Ctur}_i(t) \quad i = 1 \text{ to } N$ [eq.2a]

The traffic flow dimensioning formulas can hence label the logical block diagram as follows:
Finally it can be noticed that the traffic entering the client equals the sum of all the update requests (coming from all the clients including itself). Formally:

Ctup(t) [Client Traffic of Updates] is the total traffic of Updates entering any client at any given time t.

N is the number of clients

$$\mathrm{Ctup}(t) = \sum_i \mathrm{Ctur}_i(t) \qquad i = 1 \text{ to } N \qquad \text{[eq.3]}$$

Using [eq. 1] [eq.3] can be written as

$$\mathrm{Ctup}(t) = \mathrm{Stur}(t) \qquad i = 1 \text{ to } N \qquad \text{[eq.4]}$$

[eq. 4] means that the traffic of updates entering any single client coincides with the total traffic entering (not coming out) the server (at any given time t). The traffic coming out from the server is much higher as seen in [eq. 2a].

3.3.1.1.1.2 Analytical description from example of (simulated) *Object* data traffic

The very first time a client logs into a world it requests to the central object database all the information needed for rendering both the environment and the avatars, hence copies them in a cache directory on the same client computer, as described in section two. The traffic happens to be "occasional" and hence occurs in "bursts". The bursts are of the same size of the central object database. Since this traffic occurs only the very first time a client logs into a world, albeit large in peak size (intensity), its time average at the central object database server is really low. In fact the probability that a logon is the logon of a new user is really low (after the world start up period). A simulated example is reported below[3].

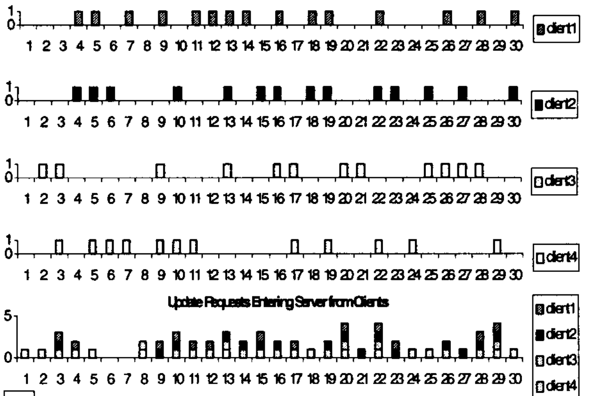
---

[1] In reality the dimension of the quantum is three reals for XYZ coordinates of position and four real for quaternions [quaternions] of rotation. The quaternions can be reduced from four to two since two out of three rotations are (usually) zero being the only usual whole body rotation around the vertical axis only. The triplet of position coordinates can be reduced to the couple of horizontal position coordinates since movement in the vertical direction is rare. The record is usually organized as follows [Length of PositionInfo, PositionInfo, Length of RotationInfo, RotationInfo]

[2] Less than linearly for the following reasons

1. The client receiving the update from all the other clients filters them in a way such he retains meaningful only the most recent one. We may call this function Forget Function or FF. The aim of the FF is to use only the most recent update. Of course if many intermediate updates are lost the movement of the object will be "jerky". Anyhow this strategy prevents the system to be overloaded and eventually stalled.
2. Not all the N clients are sending update requests at the same time.

[3] Traffic to clients is normalized to 1 arbitrary unit. Typical world Object database size 10Mb. Amusement Estimated Object database size 30Mb.



10

A formal discussion follows

Let Stor(t) [Server Traffic of Central Objects Data Base Requests] be the traffic of Central Objects Data Base Requests coming out of the Central Objects Data Base server at any given time t.

$Cto_i(t)$ [Client Traffic of Objects Data (from Central Data Base)] be the traffic of Objects Data Data entering the client upon request at any given time t (equals the size of Central Objects Database).

N be the total number of clients

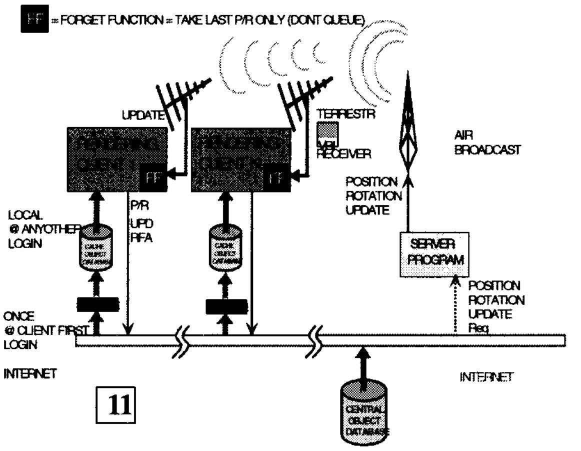
$P_N(t)$ be the probability that N clients will be logging in for their very first time at any given time t (zero for large t intervals).

Hence we may write

$Stor(t) = P_N(t) \cdot N \cdot Cto_i(t)$ i = 1 to N [eq.5]

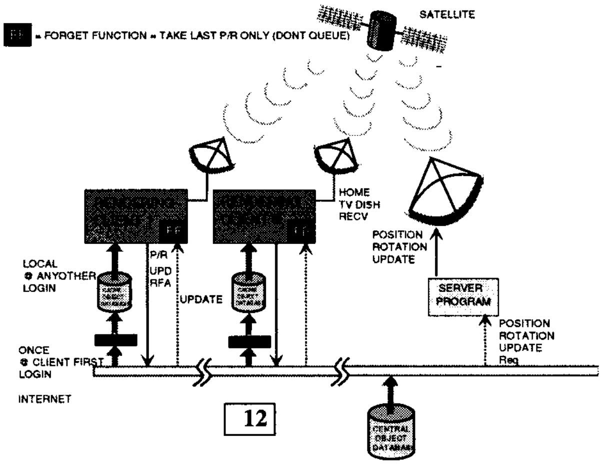
3.3.1.1.2 Mixed Cabled Internet-Air Broadcast *Hypothesis*

As shown in the following figure 11 the update link on the cabled Internet can be substituted by an "air" Broadcast link.



11

The "air" Broadcast link can be either via terrestrial VBI as in the previous figure 11 or via satellite as illustrated in the following figure 12.



12

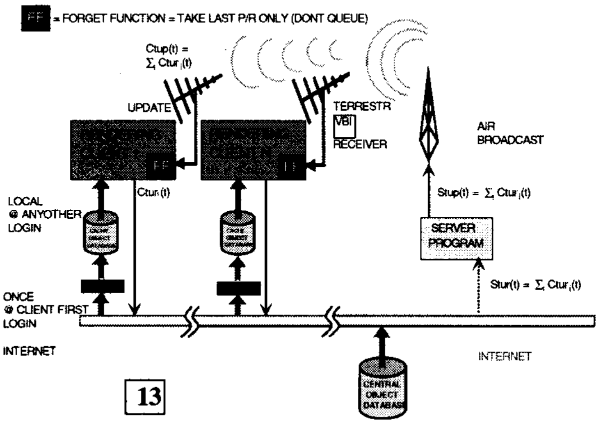
From the logic of the system the two solutions can be regarded as identical and the formal discussion of traffic follows:

Stup(t) [Server Traffic of Updates] is the total traffic of Updates copies exiting from server at any given time t.

N is the number of clients

$Stup(t) = \Sigma_i\, Ctur_i(t)$ i = 1 to N [eq.2b]

In [eq.2b] it can be seen the dramatic increase in efficiency from the case of [eq.2a]. This is due to the drop in the growing rate (of the update traffic exiting the server) falling from exponential back to linear.



13

As far as the individual client authorization procedures are concerned this version of the system is equally (if not more) safe than in the standard situation, because the client receivers will be abilitated, via a hardware based encryption system that can be remotely activated via password release, for any single service/world on an individual basis in real time by the universe and or world manager. All the described procedure will be completely transparent for both the system manager and the user.

The logical block diagram[4] can be now labeled as follows:

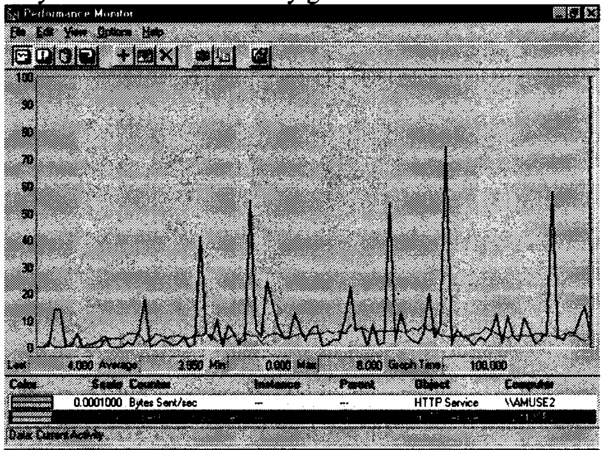
---

[4] It is shown only the case of terrestrial VBI since from this point of view both situations are identical

"Evolvable believable agents" has been successfully used in the Amusement Project in order to implement the Avatars and the non participating characters used in our broadcasted interactive game shows. The advantages that have been illustrated in this paper. As a conclusion we will just emphasize that the new "Gutenberg galaxy" (McLuhan, 1962) will be accessible only if an high degree of automation in the production of the shows can be attained. This can be regarded as the ideal continuation of the process begun with the automation of the printing process at the time of Gutenberg. The web era has further automated the printing process. In order to make a show a further quantum leap has been necessary and this has been possible only with the help of Artificial Intelligence and Artificial Life Technique.

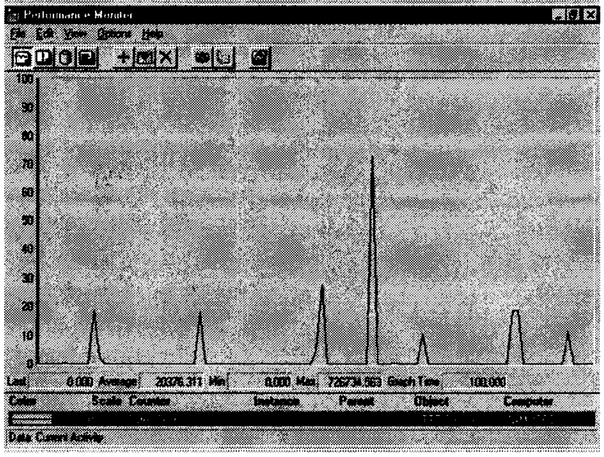
4 Appendix A – Real World Field Data

Object Data traffic from Object Server – 1 client – Worst case = First Logon
Object Data traffic from Object Server – 5 clients – Worst case = First Logon
Please note that the number of active (users) clients may be less than five at any given time.



Last 4.000 Average 3.950 Min 0.000 Max 8.000 Graph Time 100.000

Color	Scale	Counter	Instance	Parent	Object	Computer
▤	0.0001000	Bytes Sent/sec	--	--	HTTP Service	\\AMUSE2



Last 0.000 Average 20376.311 Min 0.000 Max 726734.563 Graph Time 100.000

REFERENCES

1. Bates J., *The Role All Of Emotion In Believable Agents*, Communications Of The Acm, 1994

2. Dawkins R., *Blind Watchmaker*, Longman, 1986.

3. Kosko, B., *Neural Networks And Fuzzy System : A Dynamical Systems Approach To Machine Intelligence*, Prentice Hall, 1992..

4. Mcluhan M., *The Gutenberg Galaxy*, University Of Toronto Press, 1962.

Creativity by Design: A Character Based Approach to Creating Creative Play

Joanna Bryson
Department of Psychology, University of Edinburgh, UK
joannab@dai.ed.ac.uk

Abstract

Creative play requires a fertile but well-defined design space in which to create. This paper describes one possible design process for creating virtual reality play spaces. The methodology is centered on designing AI characters for a *constructive narrative*. Requirements for the characters' agent architecture, and a three-layer design process for producing fertile and æsthetic narratives are described. This paper also discusses lessons learned from participating in a corporate virtual reality research and development effort.

1 Introduction

Creativity is generally agreed to be a Darwinian process involving novel recombination of existing design elements (Simonton, 1997; Boden, 1987). Part of what makes creativity challenging in artificial intelligence is that it often works on multiple levels simultaneously, casting a single element in two disparate roles. This is a problem not only for those planning AI representations, but also for those employed in designing props for encouraging creative play. The problem of designing creative play is to create a rich and interesting design space without limiting the creative potential of participants' experience in that space. This problem echoes the problem of flexibility vs. reliability in both natural and artificial learning systems (McGonigle and Chalmers, 1996).

One solution to this quandary is to allow the players to become constructors of their own experience. This idea has been explored to a much greater extent spatially than socially. Examples here include Internet MUDs, games such as SimCity, and constructive toys. Socially constructive toys also exist, though in fewer variants. Role-playing games, Purple Moon's 'girl games', where the objective is to find a place for the main character in an established society, and the game Creatures, which allows a player to evolve a society, are some examples. However, none of these allow the player to freely create characters and narratives of complex personal interactions. "Like good improvisational theater, cyberspace presents the opportunity for the audience to create its own characters and worlds, to write its own plots and stories, and to essentially become the directors, producers, and actors within their own imaginary worlds." (Pearce, 1997, pp. 345) For this potential to be realized, there must be tools for the creative design of personalities.

This paper describes a design approach for constructive narratives, and a design process for creating a creative play environment using this approach. The design process is separated into three components: a high level, highly artistic design element for creating story and characters; a middle, behavior based design level for creating personality in the agents; and a low level for design of basic behaviors. The middle level, the architectural approach for designing the characters, is particularly critical since it both facilitates and constrains the other two levels. The next section focuses on this level.

2 A Character Architecture for Constructive Narrative

Much research into agents for entertainment concentrates on the problem of combining the concept of a script with the notion of autonomous, reactive characters (Hayes-Roth and van Gent, 1997; Lester and Stone, 1997; André et al., 1998). The constructive narrative approach eliminates this problem by changing the top level creative design from a script to a cast of characters. Removing the script has the advantage of simplifying the addition, substitution, alteration, or removal of characters by the player. It has the penalty of removing a substantial element of structure from the palette of the initial creative designer: time ordered events. This problem has already been addressed by the creators of role-playing and adventure games. Structure is produced through the use of geographic space as well as character personalities. Plot, if desired, can be advanced by knowledgeable characters, found objects, and revealed locations. Personality traits such as loyalty and agrophobia can be used to maintain order despite the presence of a large number of autonomous characters.

Most virtual reality agent architectures are fundamentally behavior based, and at least partially reactive (see Sengers, 1998, for a recent review and critique). This is because the reactive, behavior based AI revolution of the late 1980s was primarily the triumph of a design approach. Behavior based AI is simpler to design than a monolithic intelligence system simply because it requires the decomposition of intelligent behavior into easy to program modules. Specifying that the intelligence should also be reactive removes the complex problems of learning and constructive planning from the agent. Unfortunately, it simultaneously limits the potential complexity of the agent's behavior. Nevertheless, by empowering human designers, the behavior based approach has been more successful than any fully human-specified or fully machine-learned approach.

The limitations of completely reactive systems have been widely recognized, and are addressed in numerous architectures (see for example Hexmoor, 1997). Some authors have proposed that the community has moved beyond both constructive and reactive planning to a new dominant paradigm, situated planning Levison (1996); Kortenkamp et al. (1998). Situated planning architectures generally include reactive behaviors, pre-stored plans, elements of learning, and possibly constrained forms of online planning. Two of the most popular architectures of this paradigm are PRS (Georgeff and Lansky, 1987) and 3T (Bonasso et al., 1997). Both of these have at their center a scripting language for allowing the specification of sequential and hierarchical behavior structures. These structures provide additional information (in the form of internal state) for action selection in situations that might be perceptually identical. This allows the situated planner greater behavioral flexibility than the fully reactive planner, which is dependent on current sensing to select its next action. The script structures also allow for the combination of simple behavior elements into larger modular forms, again simplifying the design task.

The work described below uses a less well established architecture, Edmund (Bryson and McGonigle, 1998). The principle advantage of Edmund over the two architectures mentioned above is that Edmund better maintains and develops the behavior based concept, particularly with regard to specialized perception. The behavior based ideal is for perception and action to be inexorably linked through the behaviors (Brooks, 1991; Matarić, 1997). Another principle is that perception should be specialized to the task (Horswill, 1993; Hallam and Malcolm, 1994). Edmund extends these principles with the observation from the natural sciences that sophisticated perception always requires memory (Pöppel, 1994; Barlow, 1994; von der Malsburg, 1995, e.g.). Behaviors in Edmund are objects, in the software engineering sense of the term. The state of the object serves as specialized perceptual memory. The primitive action and sensing elements referred to by Edmund's scripting language are methods on these behavior objects. Perceptual memory is necessary for many types of perceptual discrimination, as some categories can only be perceived by combining instances of sensory information received over a period of time. Having behavior linked with memory also allows for longer-term learning and other persistent internal states, such as emotions.

Edmund's scripting language allows for four levels of control structure. The most basic level is composed of action and sensing primitives, which interface directly to the behaviors. Next there are two fundamental sorts of conglomerates. The first is an *action pattern*, a simple sequence which runs uninterrupted unless a sense-predicate element indicates radical failure. The second is a *competence*. A competence consists of a prioritized set of elements, whose behavior tends to converge towards the highest priority element, the goal. When a competences is active, it selects the highest priority element which is currently capable of being executed. If that element is the goal the competence finishes successfully, if no elements fire the competence fails. Competences are similar to teleo-reactive plans (Nilsson, 1994), and other reactive planning mechanisms. Their elements may consist of either behavior primitives, action patterns or other competences.

The highest possible level of an Edmund script is a special form of competence called a *drive*. The elements of a drive provide the activation or motivation for a coherent section of the script. A drive's elements, while prioritized like a competence, may also be scheduled. Thus if a high priority element has fired recently, it may be inhibited in order to allow lower priority elements to execute. Drives also maintain overall behavior coherence by being persistent. As described above, competences and action patterns both terminate routinely. Further if a competence selects an element which is itself a competence, the parent competence is replaced by its child in the action scheduling. This feature is called a *slip stack hierarchy*, and allows for indefinite behavior chaining or looping. A drive remembers its initial set of elements, so that if one element terminates it is replaced by the original element. In the case of chaining competences, this allows an Edmund script to 'pop stack' reactively. After a competence terminates, each previous competence of that competence's controlling drive element chain is revisited in order. If the situation has not changed the parent of the terminating competence will be reached, if the situation has changed a more appropriate behavior may be selected. In the meantime, while a particular competence is excuting there is no bottleneck of a long stack to be checked before each individual action.

Edmund's scripting structure differs from 3T's by allowing indefinite chaining of subelements. It differs from PRS by having no external manipulation of the control structure.

For the purpose of the project described below, Edmund has been combined with another character architecture, Ymir (Thórisson, 1999), into a new hybrid architecture called Spark of Life or SoL. Ymir is designed to support multimodal dialogs between human players and artificial characters. It includes a complex scheduling and prioritization system for handling verbal and postural perceptual information, and for producing verbal, gesture and facial expression output, both in real time.

SoL's three most important extentions of Edmund's capabilities gained from Ymir are:

- an extensive encapsulated knowledge of psychosocial data for creating believable and relevant interactions between humans and animated agents,
- a system for selecting between multiple possible expressions of a particular behavior by choosing the one most appropriate given the current physical configuration of the agent, and
- a cerebellum-like system for moving the agent from the current configuration to the next chosen one.

The development of SoL will be described in detail in (Thórisson and Bryson, In preperation). The remainder of this paper discusses the process of designing a constructive narrative. The next section returns to the description of the design process, providing a task decomposition. This is in turn followed by a description of personal experience working with this methodology.

3 Designing Agents for Creative Play

As mentioned in the introduction, creative play consists principally of the novel recombination of established elements. In fact, the evolutionary utility of play is considered to lie in enabling an individual to acquire and rehearse complex behaviors, as well as to learn appropriate situations in which to express them (Bekoff and Byers, 1998; Byrne and Russon, forthcoming). In the relatively pragmatic and demanding field of entertainment, it would be a mistake to attempt to design agents for creative play that were expected to be as self-sufficient as children in developing such skills. Even were the designers' skills and knowledge up to such a task, children themselves take years to acquire such behaviors to any degree of entertaining proficiency.

Similarly, AI developers should not necessarily be expected to be sufficiently skilled artists that they can create the plots and characters necessary for a fully engaging interactive play experience. AI seems to attract (possibly even to require) developers with a hubristic belief in their own ability to replicate the thinking skills of others. However, good artists devote years of attention, and sometimes formal education, to perceiving and constructing the things that make a situation interesting, æsthetic and fun. The following design process places the AI developer as an intermediary between the artistic and the engineering aspects of the project. The AI developer is in the best situation to understand both requirements and restrictions of the overall project, and therefor has considerable responsibility for communication as well as developing solutions.

As a developer, the AI expert is responsible for taking a set of motivations, goals, knowledge, personality quirks and skills, and creating an agent that will behave reasonably rationally. The character should be able to prioritize its goals and display its intentions. It should exhibit both persistence and resolution while at the same time being aware and opportunistic. In short, it should have a recognizable personality. Developing the initial set of character attributes, however, is not necessarily solely the task of the expert in agent development. It *is* necessarily the task of one or more creative artists. The artist's responsibility is to provide well formed and interesting characters, skills and situations, to design potential plots and plot twists. In this, as in most industrial design, it will be best if the artists work in a team with the agent developers, who can help the artists understand the limits of the agent's behavioral and expressive capabilities.

The agent developers are themselves constrained by the platform on which the artificial agent is to be expressed. In virtual reality, these constraints are provided by the graphics environment in which the agent will be designed; in robotics, they are provided by the robots. It is the responsibility of the AI developer to provide requirements for, and understand the constraints of, the underlying platform — just as the narrative developer must understand the capabilities of the agents. Again, the character personality developer may or may not be the correct person to develop the agent's behavioral platform. This does not apply only to 'technical details' because the platform in this context may also provide the basic behaviors, or behavior primitives, for the agents. In this case, the platform developers are also responsible for artistic input to the project, as they need to create believable and attractive behavior environments.

The design process should obviously happened iteratively. Many forms of technical constraint might only be recognized after development has begun. Further, as the system develops, it can provide considerable creative inspiration to the designers. Even more importantly, early users, particularly those coming from outside the project, will discover both shortcomings and unforseen creative potential in the system. All of these sources of information should lead to periods of redesign and renegotiation between the various levels of the project.

4 Case Study: Creating Characters for an Adventure Narrative

The design process described above was developed as part of a blue-sky research effort to create an interactive virtual reality entertainment package that allows a child to engage in creative and constructive play inside the framework of an established action/adventure environment. The funders of this research have given permission for this account, but have asked that they not be identified in print. The project has not been brought to full product; consequently, this paper can only report partial results. However, progress has been sufficient that this case can be used to illustrate the design principles above, and to give some indication of the efforts and difficulties involved. For the purpose of this paper, this project will be called "the castle character project". This phrase will be used to refer to the AI portion of a large-scale, multifaceted research effort.

4.1 High Level Design

In the case of the castle character project, much of the creative environment was predetermined, as it was a virtual version of an active product. Consequently, the general appearance of the characters and their world, and an outline of the characters' personalities, had already been developed. The domain was a magic castle, inhabited by an evil knight and various magical entities. Much of the overall research effort was dedicated to ensuring that simply exploring the space would be intrinsically rewarding. However, this paper will focus on the subproblem of providing fun and interest through interaction with intelligent characters.

The first step to creating and interesting narrative for a set of characters is to understand the constraints of the task and the system. One set of constraints is determined by the technical specifications of the character's environment. In the castle character project, such constraints included the fact that open spaces within the castle were not significantly larger than the characters themselves. Also for technical reasons, the characters had limited flexibility, which constrained their movements and their gestures. Speech recognition was also technically difficult and unreliable. Consequently only one character, who was visibly fixed in place, was chosen for verbal interactions. These interactions were expected to be in the form of questions and answers, so that the character need only recognize a fixed set of queries or requests.

The next set of constraints are those dependent on the expected users of the system. Because the users of the castle character project were expected to be naïve to VR and only exposed to the system for a few minutes, it was considered essential that the characters provide interest whether or not the user deliberately attempted to interact with them. Thus, the characters should interact with each other. They should also react to the visitor in their domain in a way that encouraged exploration, but they should not be too forceful or too intrusive on the user's experience. To maintain interest, the characters should act and interact in such a way that they will generate continuous change. There should be no steady state that the system of characters can reach if the user is being passive.

Because of the constraints mentioned in the previous paragraphs, most of this change had to take the form of simple arrivals and departures, as well as a few gross gestures. This effect was achieved by designing characters with various incompatible goals. For example, a witch could frequently fly around the castle in a quest for intruders. However, when she found the intruder, she could do little other than land near and slightly approach the stranger, and cackle. However, her presence might attract other characters, some of whom might in turn repulse her. By having characters that are attracted by some situations, yet repulsed by either crowds or other characters, the number of simultaneous interactions, and therefor the amount of confusion, can be limited. Also, the amount of free space for character motion can be maintained.

4.2 Encoding Personality

Starting from the descriptions of the characters set by the marketing department of the product, and keeping in mind the constraints determined in evaluating the task, each character was described in terms of three to five goals or drives. Further, the behavior associated with achievement of these goals was visually described. This work was done by a team of in-house artists and external creative consultants, with the AI team participating both creatively and as technically informed resources.

Once the personality of the characters has been sketched, the next step is to code it. Under the Edmund approach described earlier, this involves developing first behavior libraries, and second control scripts. The behavior libraries provide the sensory and action primitives for the system, while the control scripts can be thought of as reactive plans. This process consists of:

- Prioritizing goals or gross behaviors and determining their necessary preconditions. For example, the witch described above has a goal of patrolling the castle from the air. This has a fairly high priority, but the motivation should be reduced by the performance of the act, so that in general she circles the castle only three times. She also has a priority of landing in a room in which she has seen an intruder, once she no longer desires to fly. She also avoids bats.

- Determining behavior primitives and behavior state necessary. For example, the witch has to remember if she saw an intruder on her patrol. Another example: a bat might approach an intruder closer and closer over successive swoops, but back off if the intruder waves their arms. This would require a piece of state within the bats swooping behavior showing how bold it is feeling in order to determine its trajectory. Some characters might be made into friends by playing with them. These would have to remember how friendly they feel towards a particular person. Seeing the user, avoiding the walls of the castle, flying and landing are behavior primitives required by all of these agents.

- Developing and testing the behavior libraries and the scripts.

4.3 Developing Behavior Primitives

In developing behavior libraries, the task of the personality designer connects to the task of environment's architects. For the castle character project, some of the potential difficulties of this relationship were overlooked, and caused some of the greatest difficulties of the AI effort.

There are several possible approaches for building the basic behaviors. One straightforward approach would be for the character developers to program the behaviors from scratch using models prepared by the graphic artists. There is a general problem for this approach: as mentioned earlier, AI programmers are not necessarily artists or students of natural motion. Animals have evolved complex motion behaviors, constrained by physical forces and structures not normally modeled on an artifact, particularly one designed to run in real time, so difficult to take into account. Animals are also constrained by habits of behavior, whether general to a species or specific to an individual. Even if æsthetic motion primitives are achieved by an AI programmer, the process of programming them is likely to have been very time-consuming.

Besides this general problem, the castle project also ran into an avoidable problem. The AI programmers took for granted an agent-oriented view within the graphical environment. That is, we anticipated being able to direct any element forwards or backwards, right or left, up or down, and at a particular speed relative to the rest of the world. However, the graphic artists and modelers were used to working only for the perspective of the camera. Consequently, coordinate frames and even size/distance metrics were not always consistent between the various models. This led to obvious problems in developing allocentric motion routines.

Another potential source of behavior primitives explored on the castle character project were the efforts of a team of animators already working on the project. Animators are trained artists who create many lifelike behaviors in the course of an animation. The idea was to segment animations into sets of behaviors suitable as multiple exemplars of various behavior primitives. As mentioned earlier, the Ymir basis of SoL would be able to select an appropriate instance from such a set and move the character smoothly to that instance. Thus a continuous variety of behavior could be derived from combining and connecting fixed sets of 'canned' behavior. Unfortunately, the animations proved as slow and difficult to develop as the hand-programmed routines. More importantly, the format the animations were produced in was determined to be incompatible with the primary real-time virtual reality environment.

One successful strategy was eventually found: a purpose built animation tool for "quick and dirty" animation segments stored in an appropriate format for the main VR engine. Motion capture is another possible source of natural looking behavior primitives, but it has not yet been explored for this purpose on the castle character project.

5 Discussion and Conclusions

Creativity is analogous to learning. In both processes, something is built by the agent's actions over time. In the case of learning, these actions are often considered implicit, built-in techniques, and the thing changed is usually the agent's own knowledge. In creativity, the actions are more often expected to be explicitly represented learned skills, and the state changed is expected to be external to the agent, so that it can be appreciated by others. However, the lines between these two behaviors are not so clearly drawn. Clark (1996) suggests that the construction of an agent's learned knowledge, particularly of a person's, consists of both memories and artifacts; while human fantasies are creative constructions that are usually completely internal. Skills for learning can be learned, and talent for arts can be inborn.

A constructive narrative is therefor creative on several levels. In creating a creative experience, the goal is to provide both interesting media for expressing the content to be recombined, and tools that facilitate the recombination. If the media provided also includes active creators — for example, agents that autonomously create situations and social dynamics — then the user has the opportunity for highly complex production. This kind of creative experience is currently only afforded to people such as composers and writers of drama, corporate managers and public policy makers. Designing such a system can in itself be a highly creative act, but it is particularly challenging to do so in such a way as to allow the users ample opportunity to express their own creativity. Of course, in the commercial climate, there are often intentional constraints to orient users towards certain forms of creative elements that exploit particular products, but these are not necessarily a burden. People like to create within well-understood spaces and forms: creativity is not chaos.

A creative environment with constantly changing stories and adventures can be developed by using artificial intelligence and design techniques that exploit and express the creativity of the designers. The intelligent agents in these environments are literally agents of creativity rather than being significant creators themselves: they embody the rules and knowledge both invented and learned by their own creators. This paper has presented a design approach for creative environments called constructive narratives, and a design process for creating a creative play space under this approach. The design process focuses on the roles of the various team members in communicating and constructing an interesting reality based around AI characters. The characters are implemented using behavior based techniques, for simplicity of design, combined with situated planning devices, to allow for complexity of characterization and behavior. We have described our experiences in using this process. This work is still in progress — we hope to eventually develop more fully interactive characters, and more open narrative architectures that allow the users to design characters as well.

Acknowledgements

The team who created the project and the environment in the case study above were: Dent-de-Lion du Midi, Christian Greuel, Svend Tang-Petersen, Victor Bonilla, Dan Mapes, Claude Aebersold, Inge E. Henriksen, Preben Kæseler, Kristinn R. Thórisson, and Julian E. Gómez. Their work as artists, visionaries and engineers is gratefully acknowledged. The AI effort was headed by Kristinn R. Thórisson, and included in addition to those mentioned above, Mikkel Arentoft, Michael Thompsen, Henrik Bach Ravn and Randahl Fink Isaksen. Lieselotte van Leeuwen assisted with child psychology; Celia Pearce assisted with concept development; Jacob Buck, Súsanna Thorvaldsdóttir, Michael Nielsen, and Christian George helped with technical implementation. The castle project was developed using Silicon Graphics equipment and expertise. Thanks also to Will Lowe and the anonymous reviewers and lawyers who contributed to the final form of this paper.

References

Elisabeth André, Thomas Rist, and Jochen Müller. Integrating reactive and scripted behaviors in a life-like presentation agent. In Katia P Sycara and Michael Wooldridge, editors, *Proceedings of the Second International Conference on Autonomous Agents*, pages 261–268. ACM Press, 1998.

Horrace Barlow. What is computational goal of the neocortex? In C. Koch and J. L. Davis, editors, *Large-Scale Neuronal Theories Of The Brain*, pages 1–22. mitpress, 1994.

Marc Bekoff and John Alexander Byers, editors. *Animal Play: Evolutionary, Comparative, and Ecological Perspectives*. Cambridge University Press, 1998.

Margert Boden. *Artificial Intelligence and Natural Man*. Basic Books, New York, 2nd edition, 1987.

R. P. Bonasso, R. J. Firby, E. Gat, D. Kortenkamp, D. P. Miller, and M. G. Slack. Experiences with an architecture for intelligent, reactive agents. *Journal of Experimental & Theoretical Artificial Intelligence*, 9(2/3): 237–256, 1997.

R. A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

Joanna Bryson and Brendan McGonigle. Agent architecture as object oriented design. In Munindar P. Singh, Anand S. Rao, and Michael J. Wooldridge, editors, *The Fourth International Workshop on Agent Theories, Architectures, and Languages (ATAL97)*. Springer-Verlag, 1998.

Richard W. Byrne and Anne E. Russon. Learning by imitation: a hierarchical approach. *Brain and Behavioral Sciences*, forthcoming.

A. Clark. *Being There: Putting Brain, Body and World Together Again*. MIT Press, Cambridge, MA, 1996.

M. P. Georgeff and A. L. Lansky. Reactive reasoning and planning. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, pages 677–682, Seattle, WA, 1987.

John C. T. Hallam and Chris A. Malcolm. Behaviour, perception, and action — the view from situated robotics. *Philosphical Transactions of the Royal Society of London*, 349:29–42, 1994.

Barbara Hayes-Roth and Robert van Gent. Story-making with improvisational puppets. In W. Lewis Johnson, editor, *Proceedings of the First International Conference on Autonomous Agents*, pages 1–7. ACM press, February 1997.

Henry Hexmoor. Special issue: Software architectures for hardware agents. *Journal of Experimental & Theoretical Artificial Intelligence*, 9(2/3), 1997.

Ian D. Horswill. *Specialization of Perceptual Processes*. PhD thesis, MIT, Department of EECS, Cambridge, MA, May 1993.

David Kortenkamp, R. Peter Bonasso, and Robin Murphy, editors. *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*. MIT Press, Cambridge, MA, 1998.

James C. Lester and Brian A. Stone. Increasing believability in animated pedagogical agents. In W. Lewis Johnson, editor, *Proceedings of the First International Conference on Autonomous Agents*, pages 16–21. ACM press, February 1997.

Libby Levison. *Connecting Planning and Action via Object-Specific Reasoning*. PhD thesis, University of Pennsylvania, March 1996. School of Engineering and Applied Science.

Maja J. Matarić. Behavior-based control: examples from navigation, learning, and group behavior. *Journal of Experimental & Theoretical Artificial Intelligence*, 9 (2/3):323–336, 1997.

Brendan McGonigle and Margaret Chalmers. The ontology of order. In Les Smith, editor, *Piaget: A Critical Assessment.* Routledge, 1996.

Nils Nilsson. Teleo-reactive programs for agent control. *Journal of Artificial Intelligence Research*, 1:139–158, 1994.

Celia Pearce. *The Interactive Book : A Guide to the Interactive Revolution.* Macmillan Technical Publishing, 1997.

E. Pöppel. Temporal mechanisms in perception. *International Review of Neurobiology*, 37:185–202, 1994.

Phoebe Sengers. Do the thing right: An architecture for action expression. In Katia P Sycara and Michael Wooldridge, editors, *Proceedings of the Second International Conference on Autonomous Agents*, pages 24–31. ACM Press, 1998.

Dean Keith Simonton. Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review*, 104:60–89, 1997.

Kristinn R. Thórisson. A mind model for multimodal communicative creatures & humanoids. *International Journal of Applied Artificial Intelligence*, 1999.

Kristinn R. Thórisson and Joanna Bryson. Flexible behavior-based planning for multimodal characters capable of task-oriented dialogue and action. In preperation.

Christoph von der Malsburg. Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5:520–526, 1995.

The Art and Science of Synthetic Character Design

Christopher Kline and Bruce Blumberg
Synthetic Characters Group, MIT Media Laboratory
20 Ames Street, Cambridge, MA 02139
{ckline,bruce}@media.mit.edu

Abstract

Drawing from ideas in both traditional animation and modern philosophy, we present a methodology for designing synthetic characters. The goal of our approach is to construct *intentional* characters that are both compelling, in the sense that people can empathize with them, and understandable, in that their actions can be seen as attempts to satisfy their desires given their beliefs. We also present a simple, value-based framework that has the flexibility to implement the subsystems necessary for the construction of intentional characters.

1 Introduction

One of the most promising intersections of entertainment and AI research is in the creation of believable *synthetic characters*—simple but complete three-dimensional situated agents who can do and express the right things in a particular situation or scenario. Examples of these types of agents include non-player characters in computer games, digital 'extras' in Hollywood movies, and computer-based artificial pets. Often these characters do not need to perform complex reasoning about the world or build intricate plans to achieve difficult goals. Instead they may effectively play out their roles by reacting to internal and external influences in ways that are both predictable and consistent with the scenario for which they were designed.

In the process of learning to build these types of characters we have often found ourselves struggling with two fundamental problems. First, what kinds of properties or qualities have we, as observers, come to expect from a believable character? Second, given these expectations, what is the 'right way' to go about implementing them?

This paper presents an overview of the lessons we have learned from our experience in building several complex synthetic characters. We begin by discussing a theory of how people go about understanding characters and then identify some subsystems that we have found to be important in building characters that are compelling and easy to understand. Next, we overview several approaches to these subsystems and show how, by separating out the semantic differences of these approaches, we can arrive at the basic activity of each. We then describe a simple value-based framework we have developed for character construction, showing how each subsystem can be implemented with the four components of our framework. Finally, we conclude with some results from our experiments with this framework and suggest directions for future exploration.

2 Expectations of a Synthetic Character

To learn how to build believable characters we look back upon the rich history of traditional character animation. When looking at a character brought to life by a great animator we know exactly what that character is thinking and feeling at every instant and, while we may not know exactly what it is about to do, we can always call upon our perception of its desires and beliefs to hazard a guess. Even when our guess is wrong, the resulting behavior nearly always "makes sense".

Classics like *The Illusion of Life* (Thomas 1981) explain the art of creating believable characters, which is fundamentally the art of revealing a character's inner thoughts—its beliefs and desires—through motion, sound, form, color and staging. But why do these techniques work? The American philosopher Daniel Dennett believes that they work because, in order to understand and predict the behavior of the animate objects around them, people apply what he calls the *intentional stance* (1987). The intentional stance, he argues, involves treating these objects as "'rational agents" whose actions are those they deem most likely to further their 'desires' given their 'beliefs'" (1998).

Desires are the key to understanding and identifying with a character. When we see the wolf look "longingly" at Little Red Riding Hood, perhaps licking his lips, we conclude that the wolf is hungry and wants to eat our heroine. How do we arrive at this conclusion? By applying the intentional stance, of course! Why else would he be acting hungry unless he *was* hungry?

Beliefs are what turn desires into actions, reflecting influences such as perceptual input ("If I see a stream, then I believe I will find water there"), emotional input ("Because I am afraid of that person, I will run away from him"), and learning ("The last time I was in this field I saw a snake, therefore I will avoid the field today"). We understand the actions of characters by inferring how their beliefs influence the ways they attempt to satisfy their desires.

How can we apply both the insights of skilled animators and knowledge of the intentional stance to build a synthetic character that people find compelling *and* understandable? From the standpoint of engineering, we can break these expectations down into a short list of functional subsystems:

- Motivational drives
- Emotions
- Perception
- Action selection

2.1 Motivational Drives

For a character to appear properly motivated it must continue to work towards satisfying its desires while gracefully handling unexpected situations. For example, a creature that is starving may temporarily ignore its hunger in order to flee from an approaching predator. Once the danger has passed, however, the creature should resume searching for food. By biasing action selection towards behaviors that will satisfy the internal needs of the creature, motivational drives provide a way to achieve goal-oriented behavior.

Several researchers have addressed the problem of motivations in the context of building creatures. One example is the work of Blumberg (1996), who used temporally cyclic 'internal variables' in the design of a virtual dog to bias action selection and facilitate external direction of synthetic actors. In another domain, Breazeal (1998) has developed a motivational system for regulating interactions between a robot 'infant' and its human caretaker, with the goal maintaining an environment suitable for learning.

Most approaches agree on the general behavior of drives. Most importantly, they are cyclical and homeostatic—positive or negative deviations over time from the base state of 'satisfaction' represent under- and over-attention, respectively, to a corresponding desire. These desires can be attended to by the successful execution of attentive behaviors like eating, or by changes in external stimuli, such as temperature fluctuations or interactions with other creatures. When unattended to, drives slowly increase over time; the effect of attentive actions is to shift the value of the drive back towards its homeostatic base state.

2.2 Emotions

Emotions bias action selection in much the same way as drives. For example, a creature that is angry may be more prone to violent behavior than one who is happy. However, emotions also bias the quality of the character's motion. If the creature is sad it should walk sadly; if it is fearful it should reach for objects in a manner which conveys its fear. In this way emotion helps observers to form an empathic bond with the creature and makes its behavior appear properly motivated (Thomas 1981).

There are many approaches in the literature to the modeling of emotions and other affective phenomena. In so-called 'appraisal' theories of emotion the individual is said to make a cognitive appraisal of their current state relative to a desired state. For example, Reilly (1996) proposes that fear might be modeled as proportional to "the likelihood of failing to achieve the current goal" multiplied by "the importance of not failing". Others such as LeDoux (1996) argue that emotions can act at a level far below the cognitive, since animals can feel emotions without consciously understanding why. Combining these approaches, Velasquez (1998) presents a framework that models how emotional systems interact with the perceptual, motivational, behavioral, and motor systems.

The general consensus of these models is that, instead of increasing slowly over time as do drives, emotions typically exhibit a large impulse response followed by a gradual decay back down to a base state. By altering the decay term and the gains on stimuli one can adjust the magnitude and slope of the impulse response, shaping the characteristic response of the emotion. Adjusting these parameters across the space of emotions is equivalent to shaping the 'temperament' of the creature. Similarly, by altering the bias term on each emotion predisposes the creature to a particular emotional state, setting its 'mood'. These decay, bias, and stimulus terms represent the influences of a variety of systems[1], which in turn are affected by the current emotional state.

It is perfectly appropriate to model the influences of multiple emotions upon internal processes such as action selection, but it is difficult for human observers to visually perceive more than one emotion at a time. This is why animators tend to emphasize the most important emotion of a character, avoiding "mixed emotions". Because we are designing characters for humans to interact with, it is important for the underlying emotional model to support some notion of a 'dominant' emotion. This dominant emotion can then be used to parameterize motion and expression, giving the observer insight into the internal desires and beliefs of the character.

One example of such a parameterization is the animation system of Rose, Cohen, and Bodenheimer (1997), in which motor commands are specified in terms of verbs ("walk", "reach-for") and adverbs ("sadly", "impatiently"). Through the use of multi-dimensional interpolation, this system can be used to continuously modify a character's motion in order to represent the changing state of one or more emotions (for example, making a character move as if it is mostly happy, but slightly impatient and somewhat tired).

2.3 Perception

Fundamentally, a situated, embodied agent needs a way to "make sense" of the world in which it is situated. By this we mean two things. First, the creature needs a method of sensing the world around it; second, it must have a mechanism for evaluating the salience of incoming sensory information. The combination of a sen-

---

[1] E.g., factors include the neurobiological (e.g., hormones), motivational (intense hunger), cognitive (an impending conference deadline; the perception of a predator), and sensorimotor (posture)

sory stimulus and its corresponding evaluation mechanism is known as a *perceptual elicitor* or what ethologists (Lorenz 1973, McFarland 1993) refer to as a *releasing mechanism*.

Sensory information can be provided to a synthetic creature many forms, most of which fall into the three basic categories: real-world physical sensing, synthetic vision, and direct sensing. Physical devices like the temperature sensors in the motors of the Cog robot (Brooks 1996) and the infrared sensors on the mobile robots of Mataric (1994) are typical of real-world sensors. Synthetic vision techniques attempt to extract salient features from a physical scene rendered from the viewpoint of the creature; examples include the ALIVE system of Maes (1996) and the artificial fish of Tu and Terzopolous (1994). In direct sensing, creatures gain information by directly interrogating the world or an object within the world include; this is the approach taken by the boids of Reynolds (1987) and many video games.

One of the important contributions of Blumberg (1996), building on ideas from Lorenz (1973), Baerends (1976), and McFarland (1993), is the notion that external perceptual influences must be reduced to a form that is compatible with internal influences such as motivations and emotions. Using a consistent internal "common currency" is essential for addressing the issue of behavioral relevance—a piece of rotting food should be as compelling to a starving creature as a delicious-looking slice of cake is to a creature that has already eaten too much. Given this representational paradigm, opportunistic behavior is simply a side effect of the relative difference in weighting between external and internal influences.

2.4 Action Selection

Regardless of the particular implementation, the fundamental issues for any action selection scheme to address are those of adequacy, relevance, and coherence (Brooks 1990). Adequacy ensures that the behavior selection mechanism allows the creature to achieve its goals. Relevance, as noted above, involves giving equal consideration to both the creature's internal motivations and its eternal sensory stimuli, in order to achieve the correct balance between goal-driven and opportunistic behavior. Coherency of action means that behaviors exhibit the right amount of persistence and do not interfere with each other or alternate rapidly without making progress towards the intended goal (i.e., behavioral aliasing).

In an effort to achieve these goals in noisy and dynamic environments, the last two decades of agent research have seen a shift away from cognitivist 'Planning' approaches towards models in which behavior is characterized by the dynamics of the agent-environment interaction. In these environments, *nouvelle AI* researchers argue, collections of simple, competing behaviors that are tightly coupled with sensors and actuators can be more effective than complex planning mechanisms, while exhibiting many of the same capabilities. Examples of these approaches include the Pengi system of Agre and Chapman (1987), the subsumption architecture of Brooks (1986), the spreading activation networks of Maes (1991), and the "Society of Mind" theories of Minsky (1988).

In an attempt to leverage the advantages of both approaches, some hybrid systems like that of Firby (1987) have used a planner to make high-level behavioral decisions while using a reactive system for low-level control during behavior execution.

Inspired by ethological theories of behavior, some systems use a hierarchical organization to break complicated tasks down into specialized cross-exclusion groups (Minsky 1988) in which mutually-exclusive behaviors compete for dominance, using mutual and lateral inhibition to control arbitration (Ludlow 1976). These include most notably the Hamsterdam system of Blumberg (1994) and the work of Tyrrell (1993).

3 A Value-based Framework

In the previous section we talked about some of the important building blocks of a character that acts and emotes in a way that people find understandable and compelling. But how should one go about implementing these subsystems? In our experience we have found it useful to try a variety of approaches; this continual improvisation is made easier when the underlying framework makes it easy to implement and integrate different models.

The traditional approach to building creatures has been to focus on each of these subsystems individually. However, if we step back for a moment and consider them as a whole, two important regularities become apparent. First, there is a high degree of interdependence among subsystems—perception, emotions, and drives influence action selection, and the results of action selection in turn affect the external state of the world and the internal state of the creature. Second, the function of each can be interpreted as a quantitative mechanism. For example, the changing value of emotions and drives indicate the state of internal needs, perceptual elicitors determine the relevance of percepts, and action selection mechanisms choose the most appropriate behavior from among multiple competing ones.

What this suggests is that there is a great deal of common functionality among these subsystems. In many cases the functions performed by these subsystems can be seen as *simply different semantics applied to the same small set of underlying processes.* Consequently, instead of struggling to integrate multiple disparate models for each subsystem, it makes more sense to build them all on top of a framework that provides these shared constructs.

3.1 The Four Components

We have constructed this type of framework from four basic underlying components. The coherency of our framework comes from the fact that our primary internal representation is the floating-point value. In addition to being an intuitive way to think about emotions, drives, and sensory input, value-based frameworks have a number of other advantages. They are relatively easy to implement and fast at run-time, have useful parallels with reinforcement learning and neural networks, and are easily extendable because external semantics are kept separate from internal representation.

Granted, not everything is best represented numerically. However, for the purposes of getting along in the world, the processes which could potentially produce non-numeric representations (sensing and cognition, e.g.) can be seem as means to one end—action. And before any creature takes action it must first decide what action to take, which is a qualitative evaluation. Therefore, for the purposes of action selection, all semantic representations in our system are first converted to a value.

3.1.1 Sensors

In our system, the sensor primitive is an abstract component that operates on arbitrary input and outputs a set of objects appropriate to the sensor's functional criteria. Sensors typically use the external world or the internal state of the character as input. In addition, they may use the output of a different sensor as input; in this manner a directed, acyclic data-flow sensing network may be formed. For example, a `VisibleObjectSensor` could find all the visible objects in the world (through direct sensing, computational vision, or any arbitrary method), passing its output to a `DogSensor` to filter out everything but dogs.

3.1.2 Transducers

The transducer primitive operates on a set of input objects to produce a single floating-point output; transducers are the gateway through which sensor data enters the computational substrate. The values produced by transducers are often objective and the result of basic computations, such as the distance to the first sensed object. However, there is nothing to restrict a transducer from returning a subjective result from a complex computation—reasoning with predicate calculus about a set of input obstacles and returning the best heading in which to move, for example. Chains of sensors and transducers form the perceptual elicitors that allow the creature to react to internal and external situations.

3.1.3 Accumulators

The third primitive in our framework, the accumulator, is the primary unit of computation. Its inputs and gains are typically the output of transducers or other accumulators, and by constructing feedback loops it is possible to create highly connected networks which exhibit useful temporal behavior. The value V_t of an accumulator at time t for N inputs and gains is:

$$V_t = \sum_{i=0}^{N-1} \text{input}_{t,i} \cdot \text{gain}_{t,i} \qquad (1)$$

where N is arbitrary.

3.1.4 Groups

The fourth primitive, the group, is used to organize accumulators into semantic groups and impose arbitrary behavior upon them. For example, a group might force the value of its accumulators to be zero except for the accumulator with the highest value. This abstraction keeps the syntax and configuration of the accumulators independent of their group semantics.

3.2 From Components to Subsystems

As an illustration we will now show one way in which each subsystem can be constructed from the components of our framework.

3.2.1 Drives

Motivational drives can be expressed using an accumulator with a feedback loop whose gain is at least one. Attentive and aggravatory stimulus inputs are given negative and positive gains, respectively, and one additional input-gain pair represents the magnitude of the growth term. The setup in Figure 1 creates a drive in the style of Breazeal (1998).

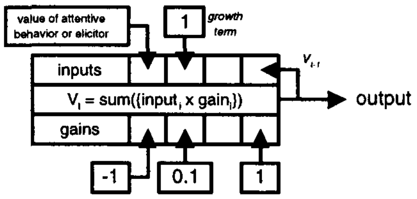


Figure 1: An accumulator-based motivational drive

Assuming that each stimulus$_i$ is a positive-valued stimulus working to satiate the drive, this configuration increases in value over time from a homeostatic base state of zero, according to (2).

$$V_t = V_{t-1} + \text{growth}_t - \sum_i \text{stimulus}_{t,i} \qquad (2)$$

3.2.2 Emotions

Emotions can be implemented with a configuration similar to that used for drives where, instead of acting as a growth term, the input-gain pair biases the homeostatic base state. By limiting the gain on the feedback loop to the range $(0, 1)$ we can effect a gradual decay over time in the value of the emotion. This configuration, show in Figure 2, varies in time according to (3).

Often it is useful to organize emotions into cross-exclusion groups for the purposes of identifying the dominant emotion. By adjusting the inhibition between the competing emotions we can tailor the personality of the creature—making a fearful creature less prone to happiness, for example.

$$V_t = (V_{t-1} \cdot \text{decay}_t) + \text{bias}_t + \sum_i \text{stimulus}_{t,i} \qquad (3)$$

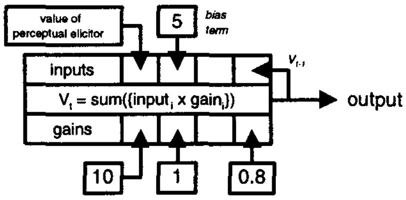


Figure 2: An accumulator configured as an emotion

3.2.3 Action Selection

A behavior is simply an accumulator that is semantically associated with a particular behavioral routine that it executes while 'active'; typically this involves sending a message (e.g., "walk") to an underlying motor system. Their inputs are the outputs of emotions, drives, and perceptual elicitors; whether a behavior is considered 'active' or not is determined by the semantics of its associated group. For example, autonomic behaviors like breathing and blinking might be contained in a group whose policy is to activate any behavior with a value above a certain threshold.

To achieve ethologically-inspired action selection policies, mutually exclusive behaviors can be organized into groups with cross-exclusion and mutual inhibition semantics and forced to 'compete' on the basis of their output values. Hierarchical action selection in the style of Blumberg (1996) and Tyrrell (1993) is easily implementable by associating each behavior with a reference to another group.

This method of implementing action selection has the advantage of making behavior design independent of action selection policy, allowing the designer to use the same behavior in many different contexts. For example, under normal circumstances a character might execute a swallowing behavior at regular intervals; this same behavior, however, might be a sub-behavior with an explicit order in the context of eating a meal. In our framework the same behavior can be used in both situations without requiring the designer to implement or have *a priori* knowledge of policy-specific details (e.g., connections to parent behaviors, execution order, etc.). This flexibility facilitates creating libraries of generic behavioral routines from which a variety of characters can be constructed.

4 Future Work

There are a many areas in which our system could be extended or improved. Most pressing is the need for better character design tools. The framework we have presented was intended as a kind of assembly-level language for building the various components of a complete character. While this flexibility has proven useful and valuable, it is currently tedious to construct complex characters in this fashion. We are currently looking into the development of a high-level behavior language or graphical interface from which we could compile the low-level internal representations discussed in Section 3.

Another area that we intend to pursue is the incorporation of learning. Though we have not yet implemented this in our existing characters, given the similarities between our work and the Hamsterdam system of Blumberg (1996) we are confident that our framework will accommodate a similar model of adaptation.

Figure 3: Two of the characters in *Swamped!*

5 Conclusion

Drawing from ideas in both traditional animation and modern philosophy, we have presented a methodology for designing synthetic characters. The goal of our approach is to construct *intentional* characters that are both compelling, in the sense that people can empathize with them, and understandable, in that their actions can be seen as attempts to satisfy their desires given their beliefs. We also presented a simple, value-based framework that has the flexibility to implement the subsystems necessary for the construction of intentional characters.

The concepts presented here were used to successfully build the many autonomous and semi-autonomous characters in *Swamped!*, an interactive cartoon experience premiered at SIGGRAPH 98. In this exhibit the participants use a *sympathetic interface* (Johnson 1998) to influence the behavior of a chicken character, with the intent of protecting the chicken's eggs from being eaten by a hungry raccoon. The raccoon character is arguably one of the most complex fully autonomous synthetic character built to date, comprised of 84 distinct behaviors influenced by 5 separate motivational drives and 6 major emotions. In addition, the continuously changing emotional state of the raccoon is conveyed through dynamically interpolated character motion and facial expressions.

Acknowledgements

The authors would like to thank Michal Hlavac, Michael Patrick Johnson, Kenneth Russell, Bill Tomlinson, and Song-Yee Yoon of the Synthetic Characters Group, and Andrew Wilson of the Vision and Modeling Group. Their experience, criticism, and hard work were instrumental in the design and implementation of our architecture.

References

P. Agre, D. Chapman. Pengi: An Implementation of a Theory of Activity. In *Proceedings of AAAI-87.* 1987.

G. Baerends. On Drive, Conflict and Instinct, and the Functional Organization of Behavior. In *Perspectives in Brain Research 45.* 1976.

B. Blumberg. Action-selection in Hamsterdam: Lessons from Ethology. In *From Animals to Animats: Proceedings of the Third International Conference on the Simulation of Adaptative Behavior.* 1994.

B. Blumberg. Old Tricks, New Dogs: Ethology and Interactive Creatures. Ph.D. thesis, MIT Media Lab. 1996.

C. Breazeal (Ferrell). A Motivational System for Regulating Human-Robot Interaction. In *Proceedings of AAAI-98.* 1998.

R. Brooks. Prospects for Human Level Intelligence for Humanoid Robots. In *Proceedings of the First International Symposium on Humanoid Robots (HURO-96).* 1996.

R. Brooks. A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation* RA-2: 14–23. 1986.

R. Brooks. Challenges for Complete Creature Architectures. In *From Animals to Animats: Proceedings of the First International Conference on the Simulation of Adaptive Behavior.* 1990.

D. Dennett. *The Intentional Stance.* Cambridge, MA: MIT Press. 1987.

D. Dennett. *Brainchildren: Essays on Designing Minds.* Cambridge, MA: MIT Press. 1998.

R. Firby. An Investigation into Reactive Planning in Complex Domains. In *Proceedings of the Sixth National Conference on Artificial Intelligence.* 1987.

M. Johnson, A. Wilson, B. Blumberg, C. Kline, A. Bobick. Sympathetic Interfaces: Using Plush Toys to Control Autonomous Animated Characters. To appear in *Proceedings of CHI 99.* 1998.

J. LeDoux. *The Emotional Brain.* New York: Simon and Schuster. 1996.

K. Lorenz. *Foundations of Ethology.* New York: Springer-Verlag. 1973.

A. Ludlow. The Behavior of a Model Animal. *Journal of Behavior* 58. 1976.

Maes, P. 1989. "How to Do the Right Thing." *Connection Science Journal* 1(3).

P. Maes, T. Darrell, B. Blumberg, A. Pentland. The ALIVE System: Wireless, Full-Body Interaction with Autonomous Agents. In *ACM Special Issue on Multimedia and Multisensory Virtual Worlds.* 1996.

M. Mataric. Interaction and Intelligent Behavior, Technical Report, AI-TR-1495, MIT Artificial Intelligence Laboratory. 1994.

D. McFarland. *Animal Behavior.* Harlow, UK: Longman Scientific and Technical. 1973.

M. Minsky. The Society of Mind. New York: Simon & Schuster. 1988.

C. Reynolds. Flocks, Herds, and Schools: A Distributed Behavior Model. In *Proceedings of SIGGRAPH 87.* 1987.

C. Rose, M. Cohen, B. Bodenheimer. Verbs and Adverbs: Multidimensional Motion Interpolation. *IEEE Computer Graphics & Applications* 18(5). 1998.

W.S. Reilly. Believable Social and Emotional Agents. Ph.D. thesis, School of Computer Science, Carnegie Mellon Univ. 1996.

F. Thomas, O. Johnson. *The Illusion of Life: Disney Animation.* New York: Hyperion. 1981.

X. Tu, D. Terzopolous. Artificial Fishes: Physics, Locomotion, Perception, Behavior. In *Proceedings of SIGGRAPH 94.* 1994.

T. Tyrrell. Computational Mechanisms for Action Selection. Ph.D. Thesis. Centre for Cognitive Science, University of Edinburgh. 1993.

J. Velasquez.W When Robots Weep: Emotional Memories and Decision-Making. In *Proceedings of AAAI-98.* 1998.

Visual Composition as Optimisation

Patrick Olivier; Nicolas Halper; Jon Pickering; Pamela Luna
Department of Computer Science, University of York, Heslington, York, YO10 5DD
patrick.olivier@cs.york.ac.uk

Abstract

Camera planning is the problem of positioning a camera within a world, such that the resulting image has some predefined set of visual properties. We are developing a graphical presentation planning system which incorporates CAMPLAN, a camera planning subsystem for polygonal graphics. CAMPLAN uses a genetic algorithm to optimise the camera with respect to a set of image objectives (properties of the image). The motivations for CAMPLAN are outlined and an informal evaluation of the system is presented in which we show how successfully more restrictive objectives can impose stylistic regularity over similar graphical scenes. We conclude with a discussion of our current work and outline future directions for graphical presentation planning.

1 Introduction

In computer graphics a virtual camera provides the means of specifying views of its world in a manner simulating that of a real camera. The virtual camera is considered to be an object with its own co-ordinates, orientation and field of view. These parameters determine the view associated with the camera and through geometric projection, and the various operations of the rendering pipeline, the idealised operation of a real camera can be simulated.

For the non-expert, placing the camera so as to produce a desired view of a scene is a difficult task. Furthermore there are a number of domains for which it would be advantageous to be able to automatically place a camera. For example, providing multimedia support for diagnosis and maintenance of engineering artefacts (e.g. communications equipment or automobile engines) requires the generation of depictions of components for every potential diagnostic or maintenance scenario. Similarly, in the entertainment industry, there is a need for the automatic generation of graphics in applications such as action summarises for multiplayer games. Even in their simplest form, both of these tasks, if performed manually, place significant demands on a graphic designer. Furthermore, if we require such presentations to be sensitive user preferences, experience and expertise, manual solutions are infeasible.

This paper describes CAMPLAN, the camera planning component of a graphical presentation planning system that we are currently developing. Section 2 comprises a sketch of existing camera planning systems. In sections 3 and 4 we describe and evaluate CAMPLAN. Finally, in section 5 we point to current and future directions both for the development of CAMPLAN and graphical presentation planning in general.

2 Graphical presentation planning

The camera planning problem resides within the latter two stages of the four stage graphical presentation planning pipeline characterised by Doree Seligmann (Seligmann, 1993):

(a) *Generation of communicative goal:* decide what it is that the image should accomplish. For example, it may be to explain something to the user, to convey the value of some property or to get the user to perform some task.
(b) *Selection of presentation strategy*: determine the visual effect that will be used to satisfy the communicative goal. That is, the types of visual cues and effect that will be used in a picture, for example, emphasise an object in a scene (without specifying how the emphasis will be achieved).
(c) *Selection of presentation method*: given the presentation strategy specify the different ways by which it may be realised. For example, emphasis may be achieved by using an extra light source to increase the brightness of an object, or by requiring the scene element concerned to have a prominent size and/or location in the image.
(d) *Image generation*: based upon the selection of the presentation methods, the graphical model of the artefact must be modified to achieve the specified visual properties. This process will include optimising the camera position to achieve visual properties, but will also involve selecting material and lighting model parameters, and potentially even configuring the spatial arrangement of the scene elements.

There are a number of existing techniques for camera planning, although it is characteristic of such methods that they are highly dependent on the specific application contexts that their originators were concerned with. In Seligmann's IBIS system (Seligmann, 1993) default camera positions were used, specified relative to the objects of interest. Blinn described how vector algebra can be used to position a camera given the desired position of two objects in the viewplane (Blinn, 1993), and this method was used in the compiler for the Declarative Camera Control Language (DCCL) which allowed the specification cinematic idioms in camera planning for animations (Christianson et al, 1996). The limitation of this approach is that it uses point abstractions of the objects, and cannot therefore can account for the range of visual effects that arise from the fact that real scene elements have finite extents (e.g. occlusion between scene elements).

More recently, Bares and Lester have developed CONSTRAINTCAM, a real-time camera visualisation interface for dynamic 3D worlds (Bares et al, 1998) (Bares & Lester, 1999). CONSTRAINTCAM allows the specification and real-time solution of three classes of constraint: viewing angle, viewing distance and occlusion avoidance. Although the expressiveness of this set is limited (e.g. it is not possible to locate objects at particular positions in the image) the strength of the system is the utilisation of solution techniques that allow real-time satisfaction of constraints sets.

The camera planning component of Steven Drucker's CINEMA system (Drucker, 1994) is the principal motivation for CAMPLAN. Drucker formulated the task of finding an image with a particular set of scene element position properties as a constrained optimisation problem whose solution was sought numerically using Newton's method. The range of properties that Drucker allowed was limited to the positions of scene elements (infact, only point idealisations of scene elements) in the image, and relative orientations between scene elements and the camera. Due to the local nature of such numerical methods we have empirically demonstrated that this approach is less effective as a solution method for more complex scenes and large sets of properties.

Lastly, (Jardillier & Languénou, 1998) have reported an approach which uses interval methods to find camera paths which yield sequences of images fulfilling temporally indexed image properties. Although the classes of properties is once again very limited, and the method computationally expensive, the interval-based approach has the advantage of guaranteeing the maintenance of visual properties for the duration of their temporal indexing. This contrasts favourably with techniques which rely on only a sample of the positions along a camera's path.

CAMPLAN is an attempt to address the principal shortcomings of all these approaches: the restriction placed on the range of image properties that may be specified and the unrealistic point-based characterisations of scene elements.

3 The CAMPLAN system

We are developing a presentational graphics system, along the lines of Seligmann's four stage architecture, which will incorporate CAMPLAN as the camera planning subsystem. CAMPLAN extends ideas from (Drucker, 1994) and imposes a division of camera planning into three sub-problems:

- *specification of shot objectives*: shots are specified not only in terms of explicit spatial relationships between the camera and scene elements, but also in terms of the objectives (visual and spatial properties) of the desired image;
- *evaluation of objectives*: for any position of the camera, each objective must be well defined and efficient to evaluate with respect to the underlying graphical modelling paradigm (i.e. geometric abstraction of the graphical model of the scene should be resisted);
- *acquisition of a camera state*: there must be a mechanism by which the camera state (location, orientation and field of view) can be established such that the fulfilment of the specified objectives is maximised.

In the following subsections we describe the requirements and operation of CAMPLAN through briefs account of the resolution of each of these problems. Readers interested in a full characterisation of the implementation of CAMPLAN are referred to (Halper, 1999).

3.1 The specification of shot properties

Shot properties can be explicit spatial relationships between objects in the scene and the camera or properties evaluated over the image itself. Examples of the former include the requirement that an particular object is facing the camera or that an object is in front of the camera's near-clipping plane (although not necessarily "in shot"). However, the use of such explicit spatial relationships between objects in the scene and the camera conflicts with the declarative approach to camera planning that CAMPLAN aims to implement. The use of such properties is in practice restricted to requiring scene elements to reside between the camera's near and far clipping planes.[1]

Image properties (objectives) are properties of the projections of scene elements into the image plane, or relations between properties of the projected scene ele-

---

[1] This is required as most of the image properties discussed remain well defined even when objects are behind the camera. For example, using the standard projection transforms, without clipping, all objects will be projected onto the viewplane whether they are in front of or behind the camera.

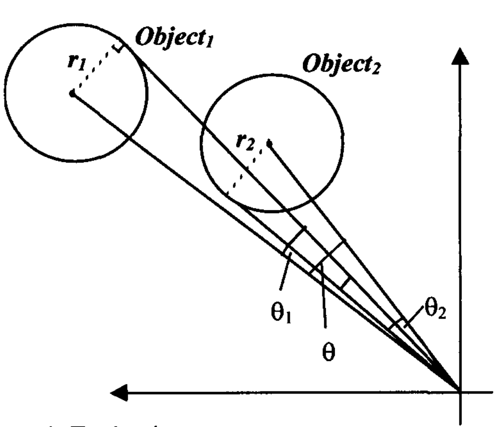


Figure 1. Evaluation of the `NotOccludedBy` property where $object_2$ lies between object1 and the camera.

ments. Table 1 lists a examples of pairs of objectives implemented by the current version of CAMPLAN. The integer following the objective name indicates the number of arguments involved in the specification of the property. The objectives `HorizSize/3` and `VertSize/3` distinguish between the required vertical and horizontal extents of an object, and their 3 arguments identify (1) the scene element for which the property holds; (2) the magnitude of the extent of the projection, in screen co-ordinates; and (3) a tolerance.

The objectives `BetweenX/3` and `BetweenY/3` require the screen coordinates of a scene element to be bounded by specified maxima and minima. More complex quantitative and qualitative relationships between the projected extents of two scene elements may be specified using the `RelObjectsPosition` objectives, including all of the thirteen qualitative relations that can hold between two one dimensional intervals (Allen, 1983) and quantitative variations on these parameterised by the magnitudes of the extents of the objects.[2]

3.2 Evaluation of shot properties

The means of evaluating objectives is dependant on the underlying graphical model of the scene. For example, if all the scene elements are spheres then we can evaluate most of the properties in table 1 very efficiently via closed form mathematical expressions. Consider the `NotOccludedBy` objective which specifies that one scene element is not occluded by another.

Suppose $object_1$ and $object_2$ have the camera co-ordinates (x_1, y_1, z_1) and (x_2, y_2, z_2). In considering the `NotOccludedBy` property, for $object_2$ to occlude $object_1$ it must be between the camera and $object_1$. Thus, assuming $object_1$ is in front of the camera $(z_1>0)$, if either of the constraints: $z_2<0$ or $|\underline{d}_1|<|\underline{d}_2|$, where $|\underline{d}_A| = \sqrt{(x_A^2 + y_A^2 + z_A^2)}$, then $object_2$ cannot occlude $object_1$. If neither of the constraints are satisfied, then consider figure 1. The constraint: $\theta_1 + \theta_2 < \theta$ is satisfied where: $\theta = \cos^{-1}((\underline{d}_1.\underline{d}_2)/(|\underline{d}_1||\underline{d}_2|))$ and $\theta_1 = \sin^{-1}(r_1/|\underline{d}_1|)$, $\theta_2 = \sin^{-1}(r_2/|\underline{d}_2|)$.

Table 1. Image objectives

`HorizSize/3`	`VertSize/3`
`PositionX/3`	`PositionY/3`
`PositionXY/4`	`NotOccludedBy/2`
`BetweenX/3`	`BetweenY/3`
`BetweenObjectsX/3`	`BetweenObjectsY/3`
`RelObjectsPositionX/5`	`RelObjectsPositionY/5`
`InViewport/3`	`EntirelyInViewport/1`
`SizeOnViewplane/3`	`SizeOnViewport/3`
`RelSizeOnViewport/4`	`RelSizeOnViewplane/4`
`OccludedOnViewport/3`	`OccludedOnViewplane/3`

In practice, CAMPLAN implements more computationally expensive evaluation methods over polygonal representations of scene elements. Furthermore, the models themselves allow the specification of the part-whole structure of each scene element which enables the specification of scene properties over named parts of scene elements. For polygonal models occlusion constraints are evaluated in a two stage manner, first over a bounding sphere approximation of the polygonal object (as above), followed by an adaptive scanline visible surface algorithm for which the resolution is dynamically assigned depending on the tolerance of the speci-

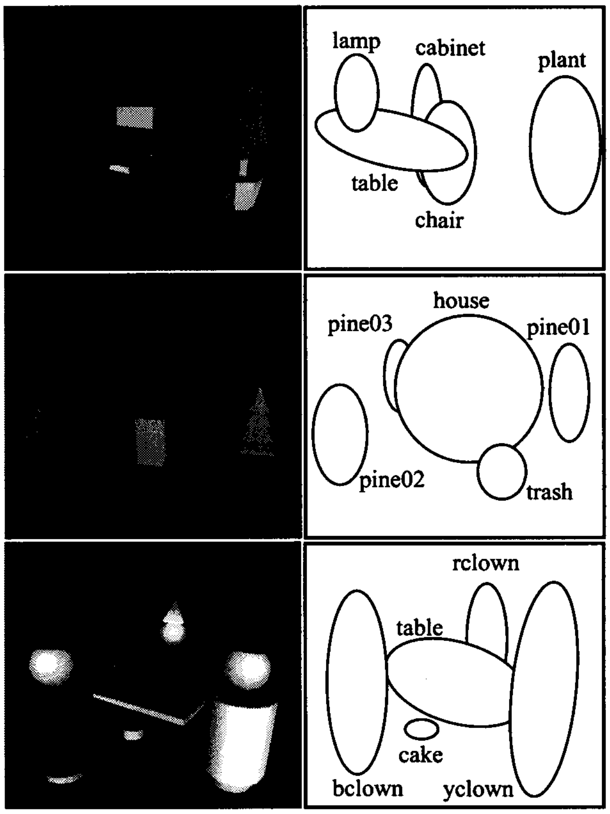


Figure 2. Test scenes and names of scene elements.

---

[2] Although we do not develop this point further, we believe that qualitative relations between image elements constitute a more natural way to specify properties in an image.

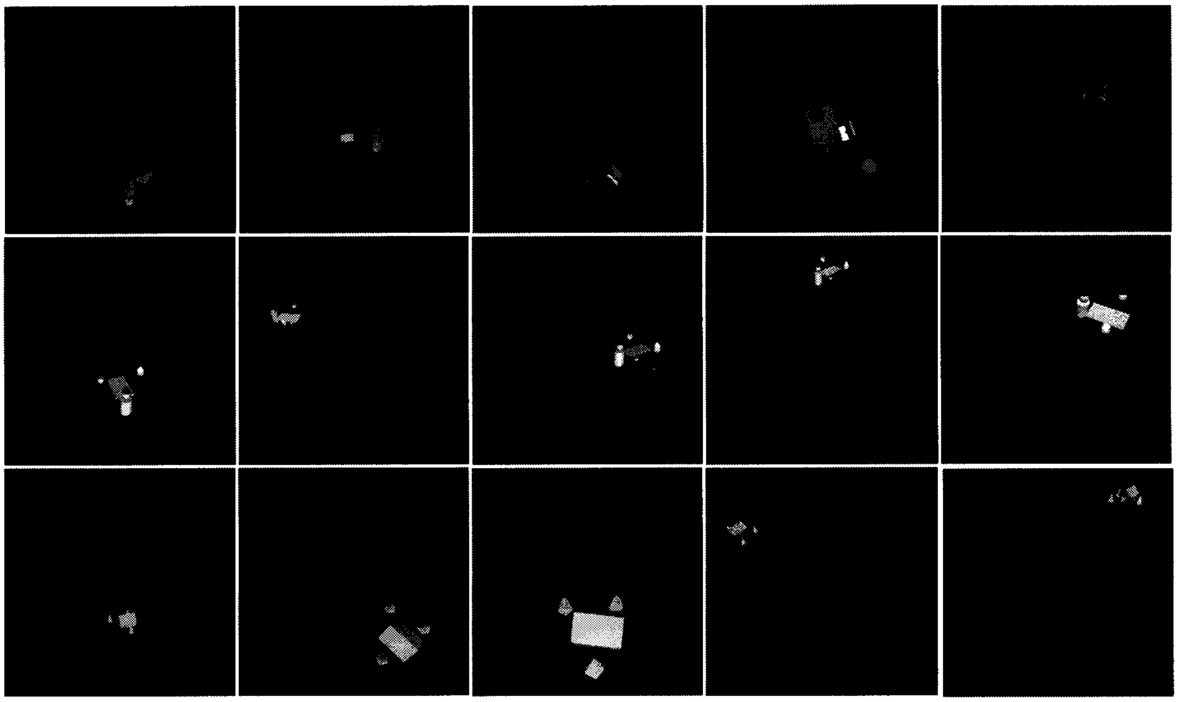


Figure 3. Example output under the constraint that all objects must appear in the viewport.

fied occlusion property. For a more complete account see (Halper, 1999).

3.3 Acquisition of a camera state

Assuming the evaluation methods outlined in the previous section, the process of acquiring a camera location and orientation, that yields an image maximising the fulfilment of the objectives, can readily be cast as an optimisation problem. CAMPLAN uses a genetic algorithm (a non-deterministic optimisation method) to find the optimal camera position. All seven elements of the camera state vector (position/3, angles/3 and field of view) are encoded in the chromosome and the configuration of the genetic algorithm is as follows: the population is randomly initialised within a space specified in the input; selection is by binary tournament; crossover is two point; mutation modifies a gene by a small random perturbation from its current value; and only 10% of the population is replaced in each generation.

The fitness function is a linear combination of normalised values corresponding to the degree of fulfilment of the objectives described in sections 3.1 and 3.2. The results described in the next section have been obtained without any experimentation with the form of the fitness function or any other parameters of the genetic algorithm, other than using a population size and number of generations large enough to give rise to an optimum camera state for which the image maximally fulfils the specified objectives.

4 Examples

Human photographers and cinematographers will casually satisfy communicative goals such as "locate the door" by applying suitable composition styles or cinematic idioms. However, our approach to camera planning depends on the specified image properties defining sets of images that satisfy the attendant communicative goals. Since the images produced by CAMPLAN are intended to function in the same way as human-composed photographs the appearance of styles when similar communicative goals are applied to similar scenes may be anticipated. Hence we attempted to evaluation of the ability of the system to generate recognisable styles. This was done by observing CAMPLAN's satisfaction of progressively more restrictive sets of image properties for the appearance of common styles.

The same sets of properties are applied to each scene in turn and the results of five random runs are given. Isomorphism between the elements of the different scenes are characterised as sets: *A={lamp, trash, cake}*, *B={table, house, table}*, *C={chair, rclown, pine01}*, *D={plant, yclown, pine02}*, *E={cabinet, bclown, pine03}*. The same objectives are applied across the scenes for isomorphic elements. For example, for every image in which *pine01* is involved in the specification of a property, there are images in which *yclown* and *plant* are identically involved.

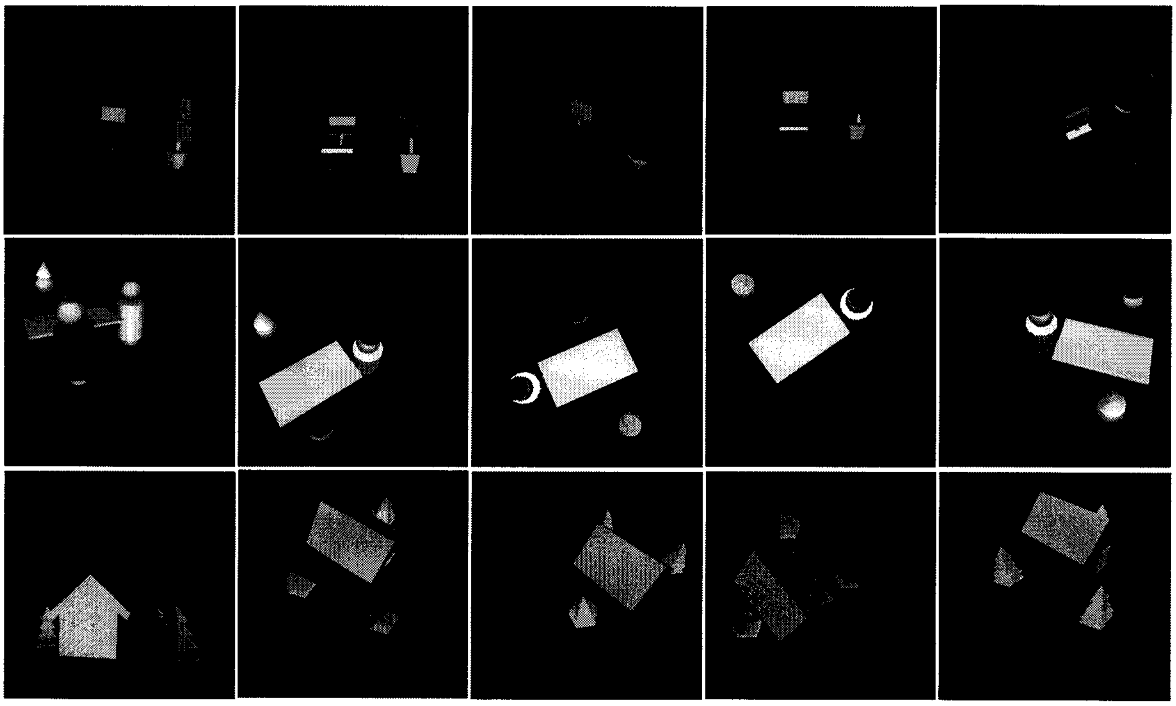


Figure 4. Example output after the addition of the constraint that *width(object B) = screen/2 ± 20%*

4.1 Objective: `EntirelyInViewport`

Figure 3 shows the set of solutions for each scene using the set of objectives given below. These simply specify that all the scene elements must lie in the viewport:

```
EntirelyInViewport objectA
EntirelyInViewport objectB
EntirelyInViewport objectC
EntirelyInViewport objectD
EntirelyInViewport objectE
```

Altough the projections of all five objects are required to be entirely within the viewport, the location of the camera may take any value in the range (±1000, 0..500, ±1000). Since the region of space within which camera location will yield an image where the scene elements are of a reasonable size is small compared to the size of the total space, it is predictable that the resulting images comprise views in which the scene elements occur in the distance.

4.2 Objective: `HorizSize`

The undesirably small size of the scene elements in the figure 3 can be addressed by constraining the screen size of a particular element (e.g. the largest) such that to have particular dimensions in screen space. In CamPlan the height and width of the viewport is two units, and so the additional constraint in the new set given below requires object B (the tables and the house) to be between 40% and 60% of the screen width.

```
EntirelyInViewport objectA
EntirelyInViewport objectB
EntirelyInViewport objectC
EntirelyInViewport objectD
EntirelyInViewport objectE
HorizSize          objectB 1.0 0.2
```

Figure 4 shows example solutions for each of the scenes after the addition of this objective. Although the camera positions are closer to the scene elements than before, there remains a significant variability both in the direction from which the scene is viewed, and in the positions of the scene elements in the viewport.[3]

4.3 Objective: `ObjectCloserThan`

As with all the previous image objectives, when we place a restriction on the direction from which an object is viewed, it must be as independent as possible of implicit knowledge about the scene.

```
EntirelyInViewport objectA
EntirelyInViewport objectB
EntirelyInViewport objectC
EntirelyInViewport objectD
EntirelyInViewport objectE
HorizSize          objectB 1.2 0.2
ObjectCloserThan   objectA objectC
ObjectCloserThan   objectA objectD
```

---

[3] In the furniture scene (top row), the distance of the cabinet behind the table, and the requirement that all objects are in the viewport, results in the fact that views fulfilling this objective having to be shots from in front of the table.

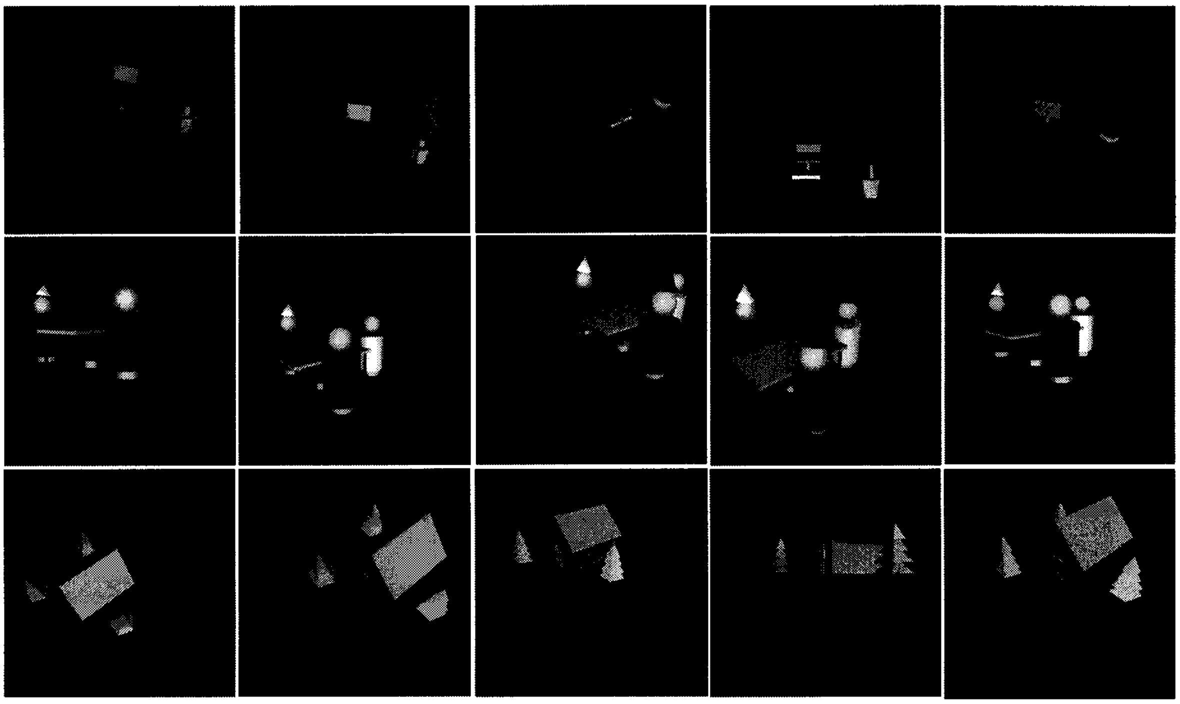


Figure 5. Example output for addition of the constraint that A is closer to the camera than both C and D.

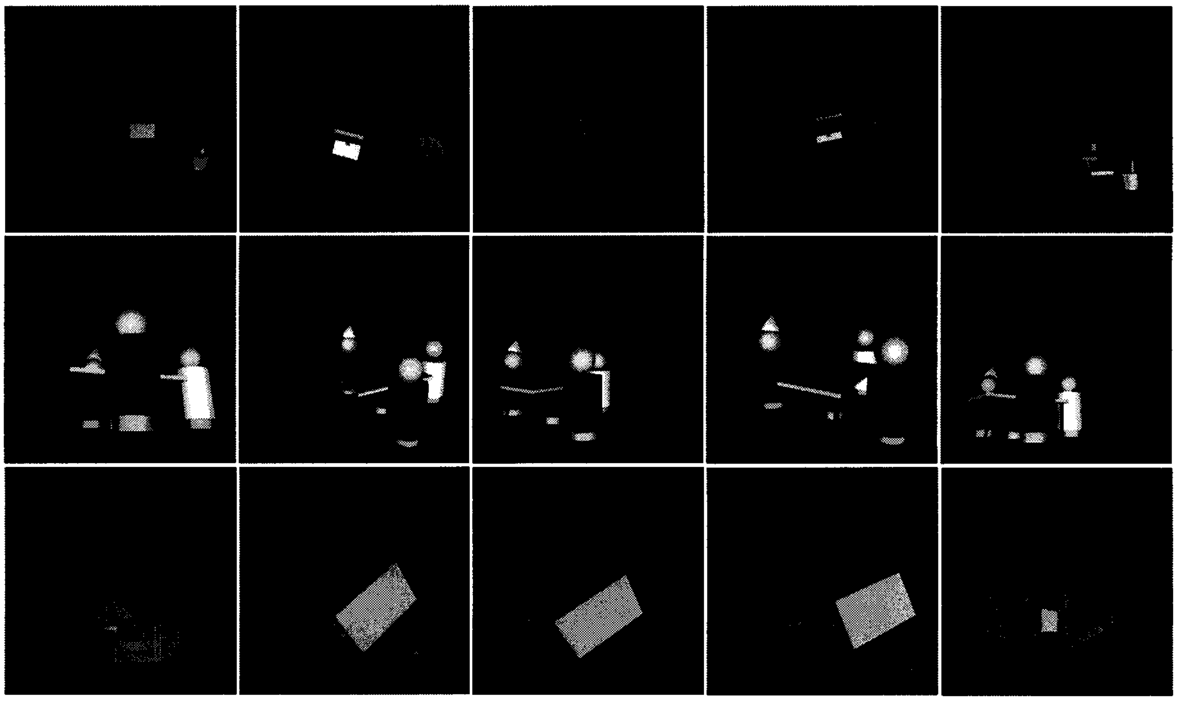


Figure 6. Example output for addition of constraint that the position B must be in lower section of screen.

Thus, rather than referencing orientation information of any particular object, we restrict the orientation of the view by requiring one of the objects to be closer to the camera than two others. The resulting images are shown in figure 5. A degree of compositional regularity has resulted, although there is still some variation in the position of the elements in the viewport.

The addition of this objective relies on a number of assumptions including the fact that the closer object is not significantly larger than its separation from the background objects, and that the three objects (in this case objects A, C and D) are not collinear.

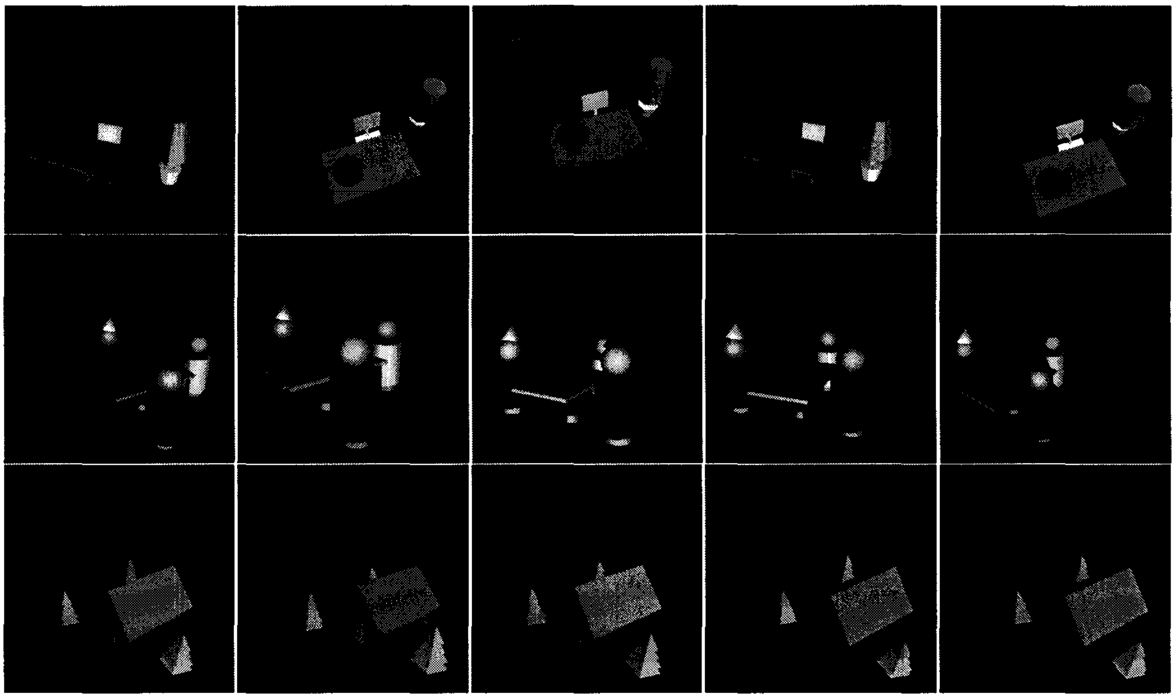


Figure 7. Example output after the addition final occlusion objective.

4.4 Objective: `PositionY`

Restricting the position of the scene elements in the viewport may be achieved by placing a restriction on the image coordinates of one of the scene elements.

```
EntirelyInViewport objectA
EntirelyInViewport objectB
EntirelyInViewport objectC
EntirelyInViewport objectD
EntirelyInViewport objectE
HorizSize          objectB 1.0 0.2
ObjectCloserThan   objectA objectC
ObjectCloserThan   objectA objectD
PositionY          objectB 0.6 0.3
```

In this case the centre of the bounding sphere for object B, (the object with the size restriction) is required to lie in the bottom 15-45% of the image. The resulting images are shown in figure 6 and we can observe a further increase in the compositional consistency. However, it is apparent, from the party scene (middle row) and the house scene (bottom row), that objects may fully, or nearly fully, occlude each other.

4.5 Objective: `OccludedInViewport`

The final objective is the removal of the possibility that any of the scene elements can be either fully, or nearly fully, occluded by another scene element. Thus we require all of the objects to be at least 20% unoccluded. Examples of the images resulting from the addition of these final objectives are shown in figure 7.

```
EntirelyInViewport objectA
EntirelyInViewport objectB
EntirelyInViewport objectC
EntirelyInViewport objectD
EntirelyInViewport objectE
HorizSize          objectB 1.0 0.2
ObjectCloserThan   objectA objectC
ObjectCloserThan   objectA objectD
PositionY          objectB 0.6 0.3
OccludedInViewport objectA 0.0 80.0
OccludedInViewport objectB 0.0 80.0
OccludedInViewport objectC 0.0 80.0
OccludedInViewport objectD 0.0 80.0
OccludedInViewport objectE 0.0 80.0
```

5 A research program

In the following subsections we outline a number of additional aspects of camera planning and presentation planning that we are either currently undertaking or intend to address in the near future.

5.1 Dynamic camera planning

In the preceding sections we have described the application CAMPLAN to planning static shots of scenes. In fact, we have completed preliminary work in extending our framework to the planning of camera paths for static and dynamic scenes. By requiring the path of a camera to be quadratic, between known start- and end-points (established using the static version of CAMPLAN), a camera path can be found by optimising a control point to maximise temporally indexed visual properties specified by a user. Samples of camera positions along the path are evaluated, and the cumulative

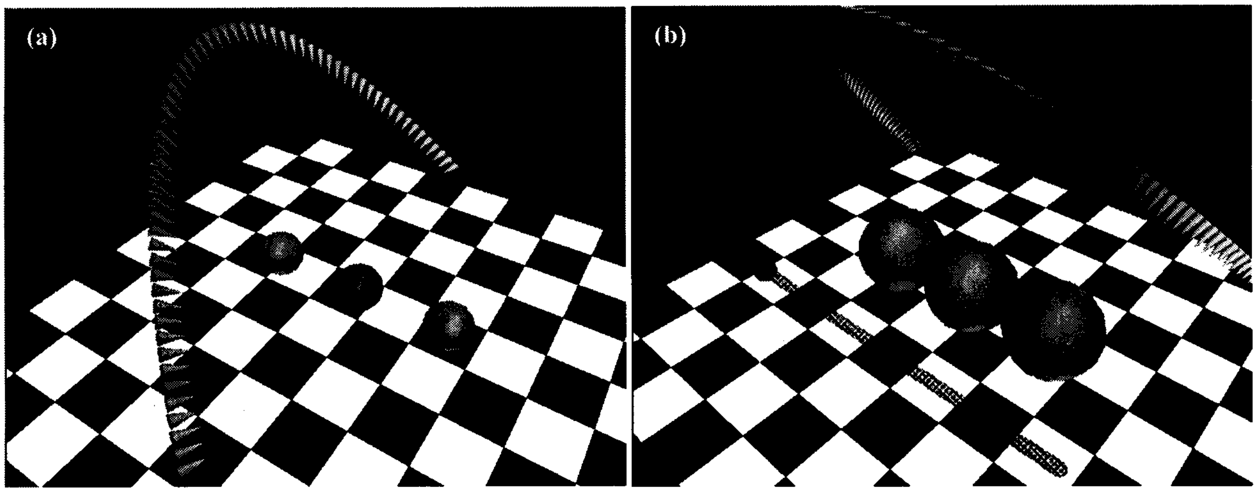


Figure 8. Examples of camera path planning for static and dynamic scenes.

fitness of the path is estimated as a linear combination of the fitness at the sampled positions. In figure 8, the cones indicate the position and orientation of the virtual camera. In figure 8(b) the small sphere moves to the left and a dynamic shot is required that maintains the full visibility and central image position of the moving sphere. In figure 8(a) the focus of the scene shifts from one outside sphere to the other, whilst at the same maintaining the visibility of the central sphere. Visual properties may persist over the whole path, or be temporally indexed. The current system is limited to spheroid approximations of the scene elements and an immediate goal is the efficient extension of this to polygonal worlds.

5.2 Evaluation and design of objectives

The evaluation of section 4 is very limited in its scope. Each scene has the same number of objects, three of which are identical or similar in their spatial characteristics, and one which is significantly larger than the others. Although within these restrictions it is apparent that some degree of stylistic regularity can be imposed, an immediate requirement for research into camera planning is the development of a both a systematic evaluation framework, and a means of eliciting image objectives. As with existing work (Christianson et al, 1996), one potentially fruitful source of image objectives are the accounts of cinematic practice (Arijon, 1976).

5.3 Lighting and material properties

Camera position is just one of the many determiners of the appearance of a graphical scene. Of the remaining aspects, the position and properties of light sources and the material properties of scene elements are very significant. We intend to investigate the application of analogous optimisation techniques to these additional variables. This will in turn require the development of evaluators for properties relating to the illumination of objects, for example, the desired distribution of light on a curved surface, or the contrast between disjoint surfaces that overlap in the image.

5.4 Visual perception and aesthetics

Another source of inspiration for the design of image objectives is the visual cognition literature, and we envisage that sets of cognitively motivated constraints will be useful in tuning results initially derived from objectives motivated by insights from graphic design. For example, cognitive theories of recognisability (Biederman, 1987), depth perception (Rolland, 1995), and figure-ground separation (Koffka, 1935) offer many insights into how the potential for visual ambiguity can be minimised. We also intend to investigate the feasibility of an algorithmic characterisation of concepts from graphic design (Lauer, 1975) and visual aesthetics (Bethers, 1964), for example, unity, emphasis, balance, contrast, pattern, movement and rhythm.

Acknowledgements

We would very much like to acknowledge the efforts of Nick Olivier and Chris McGuire for developing original (spheroid) version of the CAMPLAN software as part of their Masters dissertations at the University of Wales, Aberystwyth. Financial assistance for the continuation of this work is being provided by the University of York, through the Innovative Research and Priming Fund.

References

James F. Allen. Maintaining knowledge about temporal intervals. *Comm. ACM*, 26:832-843, 1983.

D. Arijon, Grammar of the Film Language. New York: Communication Arts Books, Hastings House Publishers, 1976.

William Bares, Joel Gregoire, and James Lester, Real-time Constraint-Based Cinematography for Complex Interactive 3D Worlds, Proceedings of the Tenth National Conference on Innovative Applications of Artificial Intelligence, pp. 1101-1106, Madison, Wisconsin, July 1998.

William Bares and James Lester, Intelligent Multi-Shot Visualization Interfaces for Dynamic 3D Worlds, Proceedings of the International Conference on Intelligent User Interfaces (IUI-99), pp. 119-126, Los Angeles, 1999.

Bethers, Ray. Composition in Pictures. Pitman Publishing Corporation, New York, U.S.A. 1964

Irving Biederman, Recognition-by-components: a theory of human visual understanding, *Psychological Review*, 94, pp. 115-147, 1987.

Jim Blinn, Where are I? What an I looking at? *IEEE Computer Graphics and Applications*, pp. 76-81, 1988.

David B. Christianson, Sean E. Anderson, Li-wei He, David Salesin, Daniel S. Weld, Michael F. Cohen: Declarative Camera Control for Automatic Cinematography. Proceedings of the National Conference on Artificial Intelligence (AAAI), Vol. 1, pp. 148-155, 1996.

Steven M. Drucker, Intelligent Camera Control for Graphical Environments, PhD Thesis, Media Lab., MIT, 1994.

Nicolas Halper, Camera Planning for Polyhedral Worlds, Unpublished MEng Dissertation, Department of Computer Science, University of York, 1999.

Frank Jardillier and Eric Languénou, Screen-Space Constraints for Camera Movements: the Virtual Cameraman. Eurographics '98. *Computer Graphics Forum*, Volume 17, No. 3, 1998.

K. Koffka, Principles of Gestalt Psychology, Harcourt Brace Jovanovich, 1935.

David A. Lauer, *Design Basics*, Holt, Reinhart and Winston, New York, 1979.

Rolland, J.P., D. Ariely, and W. Gibson, Towards quantifying depth and size perception in virtual environments. *Presence*, Vol. 4, No. 1, pp. 24-49, 1995.

Doree D. Seligmann, Interactive Intent-Based Illustrations: Visual Language for 3D Worlds, PhD Thesis, Department of Computer Science, Columbia University, 1993.

"Shaping our creative understanding" and the biological origins of our creative understanding of shape - a universal grammar of aesthetic composition (a position paper)

Prof. Peter D. Stebbing
Fachhochschule für Gestaltung Schwäbisch Gmünd,
Rektor Klaus Straße 100, D-73525 Schwäbisch Gmünd, Germany.
e-mail: stebbing @ hfg-gmuend.de

Abstract

Many artists and designers work intuitively without understanding their creative ability. A substantial part of human creativity is concerned with intuitively producing a shape, gestalt or form which has a unity or harmony regardless of media or meaning. It is proposed that this universal ability has its origins in our evolutionary past when our early hominid ancestors acquired the ability to recognise the limited fundamental relationships which characterise the gamut of organic organisation. These organic relationships and visual cues are thought to be contrast, rhythm, balance and proportion and signified, especially when occurring together, an organism and therefore nourishment or predatory threat. This perceptual ability greatly contributed to the biological success in extending the range of mankind until he had acquired the world as his niche. These perceptual preferences which evolved millions of years ago preadapted mankind with the epigentic rules for an universal aesthetic grammar for creative composition. In most consumers of aesthetic production this perceptual ability remains largely innate and intuitive. However, those concerned with promoting creativity in education and elsewhere need to penetrate these perceptual preferences of our species in order to comprehend both our creative understanding of shape and how we can enhance or shape our creative understanding.

1 Introduction

In discussing the subject of this Symposium-Workshop with the chairman, Frank Nack, he summed up the essence for me with the phrase *"Shaping our creative understanding"* in the context of Entertainment and Visual Art.

Clearly Entertainment and Visual Art is a very broad spectrum. Entertainment ranges across computer and other games, opera, theatre, jokes, poetry, literature, dance etc. whilst visual art is less diverse and includes print, painting, photography, film and video, etc. How then can we understand, enhance and promote creativity in this diverse gamut of human expression and media? More challenging is to consider how we can imagine working and fulfilling our creative expression in media which do not yet exist, but which future technologies will provide? Are there constants in the human condition which have shaped our creative understanding and if so then can we re-shape our creative understanding or merely enhance it?

Currently, in particle physics, scientists are searching for a "deep idea" that unifies the apparent diversity of nature. Some physicists believe that the beguiling patterns discovered in particle physics are clues to a theory of everything. Now "the job of a good physical law or theory is to summarise a range of behaviours in a simple way by pointing out patterns that are common to all of them." Adams (1999). Are there also deep constants to mankind's creative understanding? Is he free or bounded and if so by what? Is there a deep unity underlying our compositional creativity.

We do know that there is at least one common unifying factor in entertainment and the visual arts; it has a common source, namely- our species and its sense of creative perception. Although the question for the workshop is clearly concerned with "the shaping of our creative understanding" I believe that by swapping the words around we arrive at a very important idea and it is concerned with our profound perception of "shape", form, gestalt or composition in terms of the gamut of human expression reflected in this workshop-symposia.

Over the last 15 years I been conducting research into identifying a cohesive grammar for visual composition (Stebbing, 1988) and designing teaching programmes and a curriculum in which visual composition is taught as a cohesive grammar (Stebbing, 1990 & 1993). So far in art and design education there has never been a general concensus of the components of such a grammar in a way which compares with speech. Currently, in both texts and courses I have found a *consistent inconsistency* in both the description and presentation of visual composition. Although this does not seem to unduly concern most educationalists, I believe that this diversity of courses and texts reflects a subjective and incomplete perception rather than an objective understanding of visual composition. This means that many, if not most, art and design students never actually learn a complete grammar for visual composition. Supposing the teaching of linguistic grammar was similarly handled. The resulting diversity of courses and texts would be the product of not all the parts of speech being taught but various combinations of say between any 4 to 7 of the 8 parts of speech. In other words this lack of concensus in art and design education reflects our uncertainty about the nature of visual composition.

This comparison of visual composition with grammar is very important because as we know, grammars are creative generative systems which employ limited means to produce unlimited possibilities (Bohm & Peat, 1989; Chomsky, 1972; Pinker, 1994). Conversely, one of the aims of science is concerned with the compressibility of knowledge, i.e. identifying and understanding the basic rules which are responsible for creating the complexity which we experience around us (Lumsden & Wilson, 1981). Education has a similar aim in that students should be provided with these basic principles which will enable them to be fully creative. Therefore, it is important to achieve a creative understanding of composition in order that we can teach it.

2.1 "limited means provides unlimited possibilities"

It is also important to recognise that the principle of "limited means providing unlimited possibilities" is not a linguistics phenomenon but permeates nature. Let us consider a few examples. Currently, physicists are concerned with trying to understand and reduce the four basic physical forces down to one general unitary theory or GUT (Adam, 1999). These four forces are ultimately responsible for the complexity we observe in the Universe controlling as they do the relationships between the smallest particles of matter. In biology we find the principle of "limited means" in DNA which although it has just four bases has defined millions of life forms. In the retina of the eye just three kinds of cone cells which biologists refer to as red, green and blue types, send impulses to the brain which enable us to see an enormous range of colours and tones (Wilson, 1998). Similarly in the tongue, our perception of taste is the result of impulses from cells responding to only five basic stimuli, namely: sweet, salty, bitter and sour, and the recently reported umami (Ugawa, et al,1998)

2.2 ... but limiting the limited means provides only limited possibilities

Now nature has an important lesson for us. Imagine how DNA, sight or taste would function if one of their "limited means" was missing; one of DNA's bases, or the blue detecting cells in the retina or the cells which react to bitterness. The effects of a missing component of the "limited means" would be lethal in DNA, while the absence of blue cone cells would result in a form of colour blindness and the inability to taste bitterness might result in poisoning. Similarly, it is essential to *both* establish and teach *all* the components of the "limited means" for the grammar of visual composition. A selection is nothing more than a handicap as nature demonstrates.

3. A summary of the biological origin of our sense of composition

I would like to summarise the main argument resulting from my research and also propose a hypothesis which is supported by substantial evidence. Unfortunately, space does not allow this evidence to be reported here. Furthermore, I will briefly indicate how this research can be applied to enhance our creativity in the fields of Entertainment and Visual Art.

One of the major aims of many art forms is to produce a work of cohesive unity. This is not just a western idea but one common to many cultures. Investigations into the visual expression of different cultures and from different times reveals the common use of visual relationships to achieve this visual unity. These relationships are contrast, rhythm (or pattern), balance (and symmetry) and proportion and to which we seem to be particularly sensitive.

When I began my investigations into visual composition it quickly became clear to me that these four components of contrast, rhythm, balance and proportion (subsequently to be referred to as CRBP) were also fundamental components of organic organisation. (I had studied zoology prior to changing to a course in biological illustration.) Although this group of relationships, CRBP, is not recognised by biologists a major part of my research has been to investigate the parallel occurrence of CRBP in both organic and artistic organisation. What is astonishing and not generally known is the extent of the parallel occurrence of these organising principles in both the organic world at every level of complexity and in art and design.

3.1 The hypothesis

The question which now emerges from this and other evidence is, why should the principles of organic organisation also appear to play such an essential role in artistic organisation and expression?

The hypothesis which has emerged from the research is that the ability which evolved in our hominid ancestors to recognise the diversity of organic forms preadapted mankind with an innate grammar for visual (and aesthetic) composition. This innate grammar employs the organic organising principles of contrast, rhythm, balance and proportion. We can reasonably summarise an evolutionary sequence as to how this could occur:

1 CRBP are basic principles of organic organisation,

2 therefore our sense of perception evolved to strongly respond to the visual cues of CRBP since our ancestors' survival depended on the recognition of organic forms for finding nourishment and avoiding predators

3 It is proposed that subsequently CRBP became epigenetic rules. Epigenetic rules are defined by Wilson (1998) as "the regularities of sensory perception and mental development that animate and channel the acquisition of culture" and which are prescribed by our genes.

4 This explains why CRBP are to be found occurring universally in mankind's cultural expression. Furthermore, it appears that these are *constant components of a universal metagrammar of aesthetic (visual) composition.*

If true, this hypothesis provides an explanation for the substantial parallel occurrence of CRBP as organising principles in both artistic and organic form. The question which must now be answered is how did this evolve?

3.2 Preadaption

Preadaptation or exaptation is the evolutionary and "biological hijacking" of a behaviour pattern or part of the anatomy to perform a new function because of changes occurring in the organism's niche. There are many examples of this phenomenon including the evolution of the pentadactyl limb in vertebrates from the fins of fishes. The evolution of the bones in the mammalian ear which contribute to hearing and which evolved from bones in the reptilian jaw. Finally a behavioural example, a part of the sexual display of ducks evolved from their preening behaviour.

3.3 How did this preadaptation come about? A palaeo-ethological approach

Our early hominid ancestors, the Australopithecines, who lived about 5 million years ago (Larick, & Ciochon, 1996) in the Great African Rift Valley were forced to leave their home niche (similar to that occupied today by the chimpanzees) due to climatic changes and had to adapt to a harsher savannah environment. These environmenntal changes would have selected individuals proficient in their ability to recognising the perceptual cues indicating organic form. As James Lovelock (1979) writes:

"Our recognition of living things, both animal and vegetable, is instant and automatic, and our fellow-creatures in the animal world appear to have the same facility. This powerful and effective but unconscious process of recognition no doubt originally evolved as a survival factor. Anything living may be edible, lethal, friendly, aggressive, or a potential mate, all questions of prime significance for our welfare and continued existence."

I propose that our early ancestors acquired the ability to recognise the spectrum of organic forms by responding to the principles of basic organic organisation common across the larger taxonomic groups. The recognition mechanisms which evolved responded to differences (for example figure - ground relationships), the appearance of visual rhythms or patterns (structures resulting from organic processes), balance expressed notably as symmetry, and consistent proportions indicative of growth. Many animals respond to very specific visual cues to identify their food organisms. Insectivores such as toads for example, respond to the visual cues of a dark or black spot which moves. The perceptual leap therefore made by the early hominids was an essential acquisition to freeing mankind from his niche in the African Rift Valley. It enabled him to travel. And that we know he did because today the entire world has become his niche. Mankind is an omnivore par excellence with the potential to be practically either herbivorous or carnivorous. And several millions of years of gathering, scavenging and hunting for food has selected and evolved a perceptual system with these visual preferences. Furthermore, as Boyden (1989), writes:

"A feature of hunting and food-gathering which deserves mention is the fact that these activities appear to be generally enjoyed by the participants. This is to be expected, since in nature all animals enjoy behaviours which contribute to their survival or to their reproductive success. Presumably selection pressures operate against genotypes which do not enjoy, and hence tend to avoid, such activities. This important but somewhat neglected evolutionary principle has wide implications for the study of animal and human behaviour and for the understanding of human health and well-being."

In other words the brain stears and rewards survival-promoting- activities with pleasant sensations. Today, the evidence for our perceptual preferences acquired during our early evolution are to be seen in mankind's aesthetics, arts and entertainments.

3.4 The new demands of culture

The development of living in large social groups bought with it the evolution of culture which placed new demands on our species. Artistic and cultural expression developed in a variety of media. However, despite the diversity of the appearance of mankind's artistic expression across the different cultures there is also a striking sameness at a deep level defined by our past biology. That past was concerned with the search for structures, organisms. Like the kitten which plays with a ball of wool as if it were a mouse so we too play (Dissanayake,1974) with pattern. Intellectually we may not think of food whilst playing with patterns but because of millions of years of responding to pattern for supper and survival our perceptual system is now unable not to be interested and even excited by pattern. The same applies to contrast, balance or symmetry and proportion all factors which scientists have demonstrated that we respond strongly to. So nature has provided us with a visual grammar which we all intuitively or innately possess.

4 A brief overview of the major evidence: the parallel occurrence of CRBP in both organic and artistic expression

The strongest evidence for the hypothesis comes from the parallel occurrence of CRBP in both organisms and art and design. Let us briefly consider this evidence.

4.1 Contrast (Hubel, 1988)

Clearly, without a contrast there can be no perception. Our sense of perception has evolved in such a way that contrast is enhanced in order to help us detect organisms that might nourish us. All organic forms are discrete entities. That is they are separate from their "background", this is the contrast of "figure-ground". In addition organisms possess other contrasts. Plants generally grow up from the horizontal earth, while animals move around on the stationary ground clearly appearing separate from it. It is because of contrast that many plants and animals have attempted to protect themselves from being eaten and have evolved camouflage (Cott, 1966). Indeed, it is reasonable to propose that our system of gestalt perception evolved as a biological arms race in order to recognise hidden organisms. Contrast is as fundamental to the organisation of organic form as it is to artistic form. Itten, (1978) in his programme at the Bauhaus got his students to explore the effect of visual contrasts by juxtaposing different surfaces and materials adjacent to one another. More recently in a study of style description for product design Chen (1997) employed contrasting polarities such as:

"Homogeneous - Heterogeneous
Geometric - Biomorphic
Pure - Impure
Simple - Complex
Balanced - Unstable"
etc

4.2 Rhythm and pattern (Smith, 1986)

I will use the term rhythm as the collective term here for any kind of repetition. However, for there to be a repetition there must be contrast and so rhythm is fundamentally a repetition of contrasts. The four generative operations of repetition, rotation, reflection, and glide reflection (yet another example of "unlimited possibilites" from "limited means") produce the 7 band patterns (1 dimensional) and 17 plane patterns (2 dimensional) and we can recognise them all. We have an innate ability to respond to pattern (Smith, 1986) and also to recognise patterns we have not previously seen. This parallels our linguistic ability to understand sentences that we have not previously heard. This ability is a perceptual wonder without which artistic activity would be unimaginable. The only patterns which we are unable to recognise in nature are the "apparently" chaotic ones which are also the result of nature's rules (Stewart, 1995).

Recent research has found that the variation in brain size in different primates is directly related to vision, notably pattern and colour. The parvocellular system which detects these characteristics is disproportionately large in some primates. Consequently, it is now proposed that it is this ability to see and not the social cleverness of primates which distinguishes them from other animals (Motluk, 1998).

All organisms display rhythm and pattern to varying degrees in both their form, physiology, growth, movement and behaviour (Mestel, 1998). Furthermore, rhythms are repeated at all levels of organic scale and complexity, and in many cases they are fractal. Organic form is the product of a repetition of processes and forms. Currently, there is great interest in developmental biology because it has been found that one kind of gene, the *Hox* genes (Wolpert, 1998), appear to control the modular segmentation in animals ranging from fruit flies to humans. Rhythm and pattern are the most ubiquitous indicators of life because without repetition life would simply be impossible. Wherever organic processes occur patterned structures are necessary for them to happen, for example the patterns of chloroplasts in the cells of leaves. Furthermore, the leaves are arrayed in patterns to catch the sun's light. Consequently, the tree's solar panels, its leaves, often show approximately parallel arrangements in order that their chloroplasts are optimally orientated towards the sun. In visual composition one frequently used harmonising device is the gestalt principle of the same orientation or parallelism. Cezanne used parallelism in his picture 'The Large Bathers' (1906). Ultimately, pattern is about prediction since life without regularity would make organic adaptation and therefore the evolution of life impossible. Form follows process and generally in the organic world a process which is not repeated leaves behind it no form.

4.3 Balance and symmetry (Concar, 1995)

Balance is the mediation or neutralisation of forces. We see it, too, at every level of complexity where organic systems attempt to minimise the forces acting on them by setting one force against another. These forces occur across the physical, chemical, biological, psychological and social milieus. Indeed, balance is one of the most economic ways of neutralising unwanted forces on the organic system; and so balance is universal. The cell attempts to balance the destructive effects of the forces of either too much or too little water acting upon it. We may employ the strategy of balance when carrying our luggage by dividing it between two hands rather than carrying it in one hand and thereby also minimising our energy expenditure. Similarly, most animals have evolved to be bilaterally-symmetrical in the direction of their movement (Dawkins, 1996). Many plants which grow against the force of gravity also exhibit a rotational symmetry in order to balance and minimise the effects of gravity. In art and design symmetrical and asymmetrical balance is used to achieve an aesthetic effect. However, the fact that symmetry pleases us stems not only from other organisms but also from ourselves in which it indicates good health and is sexually attractive. At another level in human culture the idea of balance is associated with legal fairness so that when a judge passes sentence on a criminal the forces of retribution and the nature of the crime balance on the fulcrum of justice.

We can really begin to understand how closely CRBP are interrelated with one another when we consider that organic systems achieve balance between contrasting forces through feedback mechanisms which result in oscillations or rhythms. These rhythms are to be seen at every level of organic organisation ranging from its physiology to the numbers in a community. We ourselves experience the swings of our moods which result from the cycles of hormone secretions etc.

Homeostasis, the constant state of the internal organism, is achieved by an organism's physiology balancing the danger of contrasting extremes to arrive at an optimum. Gregory Bateson (1980) best summed up the physiology of need when he wrote: "Desired substances, things, patterns or sequences of experiences that are in some sense 'good' for the organism - items of diet, conditions of life, temperature, entertainment, sex, and so forth - are never such that more of something is always better than less of something. Rather, for all objects and experiences, there is a quantity that has optimum value. Above that quantity, the variable becomes toxic. To fall below that variable is to be deprived."

4.4 Proportion (Huntley, 1970)

"A ratio involves a relation between two magnitudes, p : q. A proportion involves a relation between two ratios, and potentially between four magnitudes, p : q :: r : s ." (March, 1998). A universal characteristic of life is growth and this process which occurs in the dimension of time results in organisms displaying consistent proportions. The proportion of the components or parts to the whole are normally a balance between contrasting forces resulting in optimal form. Quite simply the parts should not be too large or too small to fulfill their functions. Consequently, repeated parts such as legs or branches of an organism display consistent proportions according to the species, its growth patterns, its needs and niche. The literature on phyllotaxis, the Fibonacci series, and the use of the golden section and other proportions in art and design is well known and does not need requoting here. In the last decade morphogenesis and the development of the pentadactyl limb and the discovery of the *Hox* genes have been key topics of international scientific research (Wolpert, 1998). What may not be realised is that they are also the key to understanding our own aesthetic response to proportion through the imprint of the mother's face and our own hands from our earliest development. This explains our subsequent intuitive response as adults to our proportions, the same proportion which is commonly to be found in nature- the golden section.

4.5 The key to both natural and aesthetic harmony and unity- relationships

The harmony or unity of both the artwork and the organism result from relationships. If we consider the coordination of the organism for a moment; it is a symphony of rhythms which synchronise the body's functions both internally and with the diurnal changes of the environment. Biologists used to search for the temporal organiser which coordinated all of the body's activities, however, today it is now known that all the cells and nearly all organs have their own clocks (the spleen is an exception) (Mestel, 1998). Rhythms in the body are fractal, coordinating our bodies from the cellular level to the conscious actions of walking and running. Similarly, contrast, pattern, balance and proportion are all forms of relationship fundamentally crucial to the functioning of all organisms at all levels or organisation. Loss of relationship and control results in deterioration of the body's functions threatening its survival.

5 the supporting evidence of universality

If our aesthetic preferences are the constants which I am proposing they are and have evolved from our biology then they will have a universal occurrence.

5.1 Universal geographic and historical aesthetic comprehension

If our ability to recognise the diversity of organic forms provided the human species with its visual language then we should be able to appreciate the visual artforms of other cultures. Indeed, while we may not understand the meaning or value attributed by other cultures to the artefacts they create, we can, nonetheless, enjoy at a basic level the organisation of CRBP in the pottery, architecture, sculpture, etc. The art museums the world over can vouch for the success of the blockbuster exhibitions on Chinese, Egyptian and African art. Furthermore, the compatibility of human artistic expression has long been confirmed by the art trade between the west and the east over the "Silk Route". More recently, during the last century and at the beginning of this one, the Japanese (Bergson, 1980) and African arts have had a profound effect on the development of European art providing the impetus which resulted in it exchanging realism for abstractionism.

5.2 Universal occurrence of CRBP in some other media

CRBP occur as the basic organising components not just in visual composition but in other art forms as well, for example, Smith-Autard (1992) writing on the evaluation of dance:

"A few may be pleased by the overall shape of the dance, and see the beginning, middle and end in *proportionate* relationship, and each section as a well *balanced* entity yet carefully blended into a unified whole. Others may feel a sense of pleasure on recognition of the *repetitions* and *contrasts* and follow the design of the dance within these frames of reference."

and on music Westrup (1967) writes:

"Forms are the ways in which at different periods music is cast into intelligible shapes. The fundamental principles remain constant: *balance, proportion, repetition, contrast,* variation, and so on. It is the detailed application of these principles that changes. Styles are the ways in which individuality, or it may be the general feeling of the period, finds expression."

In my most recent research on the Internet I found a Theatre Arts (TA30) course, taught at the Saddleback College in California (1998) included literature analysis. The topic was summarised as follows:

"II. Analysis of Literature
    A. How to analyze:
        1. Unity/Harmony
        2. Variety/*Contrast*

3. *Balance/Proportion*
4. *Rhythm*
B. Semantics and symbolization
(Language)
C. Literary organization
D. Author's attitude"

In the introduction to a paper concerned with architectural design written by Minai (1984) he described his objectives as follows, to:
"Harmonic order, as the law of *opposites*, is reached by increasing uniformity and regularity through reduction of entropy of certain functions (*eg rhythm*), and at the same time increase in randomness and diversity through increase of entropy of *opposite* functions (eg climax) while numerous constraints are conditioning. Finally, 'originality' and 'probability' techniques are introduced as tools for optimising certain random functions, such as axioms (*rhythm*, climax, *balance, proportion*, harmony and functional expression) and thus produce 'harmonic order".

6 A creative universal meta-grammar for aesthetic structure and shape

Evolution unwittingly preadapted us with a universal metagrammar for aesthetic composition and expression. Millions of years of searching for food organisms and avoiding predators has given us a perceptual system which responds to structure and shape resulting from the perceptual constants of CRBP. Consequently, we compose with CRBP in dance, music, literature, architecture, art and design. It is our ability to recognise relationships, as our early ancestors did, regardless of the organic media of feathers, scales, leaves or footprints or some other unfamiliar material, which indicated an organic form. Mankind's survival depended on this ability as successive migrations of our predecessors populated the world. This is why Marshall McCluhan (1969) got it wrong. The media is not the message. Relationships are the message, and these are defined by contrast, rhythm, balance and proportion. These are our perceptual constants defined by the senses which trends and new media can never replace. This is the meta-grammar of composition and shape which we have to understand. Whether one composes with dance steps, computer games, bricks and glass, marble, pixels, paint and canvas, type, sack-cloth and ashes, rusty nails, dead sheep or excrement is of little significance. It is our ability to recognise and understand arrangements which enables us to both appreciate art and detect life forms (even from outer space). We will always search for shapes regardless of the media. Process = Pattern = Structure & Shape = Life form.

7 Some steps from here to...?

In shaping our creative understanding we must first recognise that evolution has already shaped our creativity for us by providing us with the ability to perceive composition and structure in any media. Anything less than shape, structure, gestalt or form, will have a deficit of the relationships of contrast, rhythm, balance and proportion to which we have evolved to respond and therefore a deficit of significance. In shaping our creativity we must therefore understand as completely as we can what we naturally, intuitively, unconsciously and innately enjoy.

The following notes or steps are provided as strategies -
7.1 Comparatively exploring contrast, rhythm, balance and proportion in a selection of works from media which employ different dimensions and are perceived with different senses. A selection for study might therefore include a painting, a film, an opera, a computer game, a dance and a building. By firstly studying the significance of contrast within each of these examples and then comparing the use of contrast between them we will be able to understand and become conscious of the effect of contrast inedependent of the media in which it has been used. This "within" and "between" analysis of contrast should then be repeated for rhythm or pattern, balance and symmetry and finally proportion.
7.2 The second series of exercises, should, using the same examples, look to see and analyse how CRBP are interrelating between each other within each work. So for example we might recognise that a rhythm or pattern of elements on one side of a picture balances with a contrast on the other side of the composition.
7.3 Finally, using the elements of the media in which one works explore creating relationships of contrasts, rhythms, balances and proportions for their own sake. A painter may explore the relationships of point, line and plane. A composer will explore the CRBP relationships between notes, instruments, motifs etc. A writer could explore the relationships of CRBP between characters, location and time in history. A computer games designer may explore the CRBP relationships created between the players, their moves, gains and losses and aims.

Increasing the complexity by mixing the media clearly offers new arenas for creativity but maintaining a clarity of structure will become progressively more difficult. In this case nature is the model for creativity where units at one level of complexity come together to form units at a higher level of complexity. This, in addition prescribes hierarchical structuring or prioritising in creative work otherwise structures will become formless without prioritised aims. Organisms after all have very clear aims- energy acquisition and reproduction.

I have often found that students usually think that it is easier to be creative with unlimited free choice. However, the converse is true as nature consistently demonstrates at every level of complexity that she is able to produce unlimited possibilities from very limited means.

References

Adam, S., *A theory of everything*, Inside Science, no 118, New Scientist, vol 161, no 2174, 20.2.1999

Bateson, G., *Mind and Nature, a necessary unity*, Fontana, Collins, Glasgow, 1980

Berger, K., *Japonismus in der westlichen Malerei 1860-1920*, Prestel-Verlag, München, 1980

Bohm, D., Peat, F.D., *Science, Order and Creativity*, Routledge, London, 1989,

Boyden, *Western civilisation in biological perspective*, Clarendon Press, Oxford, 1989,

Chen, K., *Form language a style description*, Design Studies, vol 18, no 3 pps. 249-274, 1997

Chomsky, N., *Language & Mind*, Harcourt Brace Jovanovich, Inc., New York, 1972

Concar, D., *Sex and the symmetrical body*, New Scientist, vol 146, no 1974, pps 40-44, 1995

Cott, H., *Adaptive colouration in animals*, Methuen Books, London, 1966

Dawkins, R., *Climbing Mount Improbable*, Penguin Books, London, 1996

Dissanayake, E., *A hypothesis of the evolution of art from play*, Leonardo, vol 7, no 3 pps 211-217, 1974

Hubel, D.H., *Eye, brain, and vision*, Scientific American Library, New York, 1988

Huntley, H.E., *The divine proportion*, Dover Publications, New York, 1970

Itten, J., *Design and Form*, Thames and Hudson, London, 1978

Larick, R., & Ciochon, R.L., *The African emergence and early Asian dispersals of the genus Homo*, American Scientist, vol 84, no 6, pps 538-551, 1996

Lovelock, J.E., *Gaia: a new look at life on Earth*, Oxford University Press, Oxford,1979

Lumsden, C., Wilson, E.O., *Genes, mind and culture*, Harvard university Press, Cambridge, Massachusetts, 1981

March, L., *Architectonics of Humanism*, Academy editions, John Wiley & Sons Ltd., Chichester, 1998

McCluhan, M., *Understanding Media*, Sphere Books, London, 1969

Mestel, R., *Slaves to the rhythm*, New Scientist, vol 160 no 2162 pps 32-37, 28.11.1998

Minai, A.T., *Communication systems approach in architectural design*, in *Design Theory and Practice*, pps 37-50, ed by Langdon, R. & Purcell, P.A., Design Council, London, 1984

Motluk, A., *Think your're clever*, New Scientist, vol 160, no 2159, p 20, 7.11.1998, report of recently published work by Barton, R., Proc.of the Roy. Soc. B, vol 265, p 1933, 1998

Pinker, S., 1994, *The language instinct*, Penguin Books, London

Pinker, S., 1997, *How the mind works*, Allen Lane, The Penguin Press, London

Saddleback College, California, *TA 30*, (printed out from the Saddleback College website 30.11.1998) (http://www.sccd.cc.c.us/sc/courses9798/topic80614010.html)

Smith, J.M., *The problems of biology*, Oxford University Press, Oxford, 1986

Smith-Autard, J., *Dance Composition, a practical guide for teachers*, A & C Black, London, 1992

Stebbing, P.D., *Identifying the basis of visual appeal for designers*, ICSID/ Robert Maxwell Competition for new writing on design thinking, unpublished, a shortlisted winning entry, 1988

Stebbing, P.D., *Curriculum für die Akademie des Kunsthandwerks*, Schwäbisch Gmünd, 1990 & 1993 unpublished.

Stebbing, P.D., *There is a universal grammar for visual composition*, Generative Art, ed. C.Soddu, Proc. of the 1998 Milan First int. conf. Generative Art, Milan Polytechnic, Milan, 1998

Stewart, I., *Nature's Numbers*, Weidenfeld & Nicolson, London, 1995

Ugawa, S., Minami, Y., Guo, W., Saishin, Y., Takatsuji, K., Yamamoto, T., Tohyama, M., Shimada, S., *Receptor that leaves a sour taste in the mouth*, Nature, vol 395, no 6702, p555-556, 8.10.1998, "There are five basic tastes; sour, salty, bitter, sweet and umami (the taste of glutamate and similar naturally occurring substances)" 1. Lindemann, B. Physiol. Rev. vol 76, pps. 719-766 (1996)

Westrup, J., *An introduction to Musical History*, Hutchinson Univ. Library, London, 1967

Wilson, E. O., *Consilience*, Little, Brown & Co., London, 1998

Wolpert, L., *Principles of development*, Current Biology Ltd & Oxford University Press, London, 1998

A4SM[1] - An integrated digital movie production environment

Christian Wilk
FAW – research institute for applied knowledge processing
Helmholtzstr. 16, 89081 Ulm, Germany
wilk@faw.uni-ulm.de

Abstract

This paper proposes the development of an integrated digital movie production environment, that effectively is to capture the whole production history of a movie in as much detail as possible. The resulting collection of metadata serves as a rich basis for semantic access to movies and for further investigation and exploitation of conceptual and metaphorical structures. Among other possible application scenarios are the analysis and clustering of this database to gain a deeper understanding of creative and conceptual processes.

1 Introduction

One of the big challenges in multimedia databases is the efficient and content-based access to various levels of detail in visual media, e.g. to camera motion paths, single shots, scene locations or characters. For this to happen, metadata has to be stored and associated with the underlying media components. Automatically indexing and analysing the content of video data by computer software is still in its infancy – due to the lack of the computer's understanding of higher-level conceptual and metaphorical structures – like they are described in [5]. Computers are best in analysing the syntactical structures of visual media – the flow of objects and shapes, colour and brightness etc. In order to allow content-based navigation in visual media, semantically rich descriptions have to be added manually. This nowadays leads to a workload that is twice as much as the whole production time. Analysing and decomposing a movie, based on the finished material, without access to the script and notes, that guided the movie's realisation, takes nearly the same time than to design it. I therefore propose a digital movie production environment, which accompanies and captures the movie production process during its whole lifecycle. It focuses on techniques to intelligently support the technical and creative aspects of digital movie production.

2 Movies

The production of movies is unique in the way it is designed and realized. If you compare the amount of consciousness, that is required to communicate within a certain type of media, time-based visual media is surely the most demanding one. Normally, making conversation requires less consciouss and explicit thinking as writing and in turn less than making a 90 minute feature film. As Linda Seger points it out:

> *"In the cinema the understanding of the material and the understanding of the organization of the material are particularly complex, because the material of the cinema itself demands particular organization, demands particularly extensive and specifically cinematic treatment. The cinema is much more complicated than other forms of art, because the method or organization of its material and the material itself are especially 'interdependent' " [8]*

or in other words:

> *"Film is film. It will always require planning, forethought, and execution. By nature film conveys information, visual and conceptual." [John Daigle, Mon, 07 Dec 1998 at the "film-philosophy" mailing list]*

This visual and conceptual information that film conveys is realized in a linear form, but nevertheless consists of multi-dimensional hierarchical clustered networks of relationships between different kind of media and conceptual entities. These conceptual structures represent the thematic hierarchy of the story and the causal chain of events, as illustrated in figure 1.

---

[1] A4SM - pronounced 'aphorism' - Authoring System for the Syntactic, Semantic and Semiotic Modelling

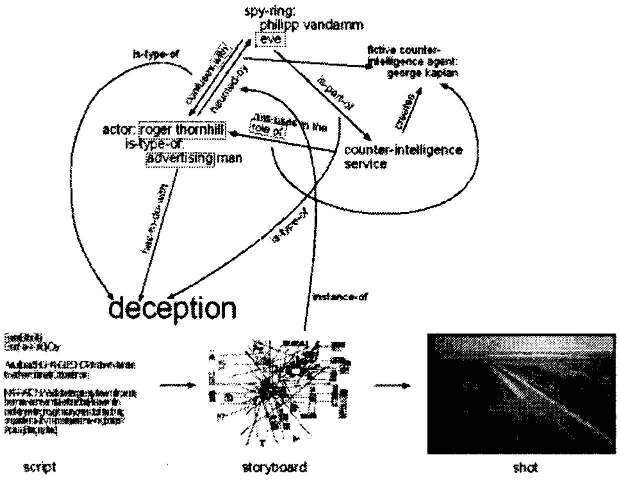


Figure1: Snapshot of the conceptual structure of "North by Northwest", directed by Alfred Hitchcock

They are explicitly designed and constructed by the script-writer, the storyboard artist, the director and the other film specialist involved during movie production. At the moment this preproduction phase is not supported by any means of digital environment. All this information about story decisions, story structure etc. exists only as an unstructured collection of papers, notes, sketches and drawings, which is lost or inaccessible after the production is finished. Ironically, it is this kind of cognitive, content and context based information which today's researchers and archivists try to analyse and reconstruct out of the final product – mostly with limited success.

3 Benefits

Nowadays, when television, movies and reality are blended together, the concepts and ideas that are communicated through media are already a model of reality. There is by no means a chance to distinguish between an evolution of culture without the influence of media on people's minds and thinking.
One intention of movie making is to evoke emotions using various storytelling methods. Therefore its design domain corresponds with everyday human experience. The result of a movie design process is a collection of structured sequences of visual and auditive events associated with conceptual indexes. Together they represent the metaphorical structuring of knowledge in the domain of human life and culture. They can be seen as a model of human perception or partial consciousness. With partial consciousness I mean the fact, that a movie is constructed out of fragments - either representing potential reality or extending reality by imaginable but not possible plots in the context of our actual reality. If we collect and organize the conceptual structures of large amounts of movies we will get a visual thesaurus of human culture and experience. It can be used as training data for AI applications or with further processing enrich the CYC [6] knowledge base of common sense. Among other possible application scenarios it can also serve as a rich basis for further investigations into the nature and genealogy of creative design processes. In [5] Lakoff complains about the lack of a lexicon of the metaphorical structures of our conceptual system:

> *"For example , if you look in a dictionary under "love", you find entries that mention affection, fondness, devotion, infatuation, and even sexual desire, but there is no mention of the way in which we comprehend love by means of metaphors live love is a journey, love is madness, love is war, etc. If we take expressions like "Look how far we've come" or "Where are we now?" there would be no way to tell from a standard dictionary or any other standard account of meaning that these expressions are normal ways of talking about the experience of love in our culture. ... It would be very strange in a dictionary to see "madness" or "journeying" as senses of "love". They are not senses of "love", any more than "food" is one of the senses of "idea". Definitions for a concept are seen as characterizing the things that are inherent in the concept itself. We, on the other hand, are concerned with how human beings get a handle on the concept – how they understand it and function in terms of it.", p. 115-116*

Scriptwriting and storyboarding are ideally suited for gathering such cases due to their creative nature and their explicit construction of concept structures and the design of new metaphorical expressions.

4 Solution

In comparison to the related Informedia project [4], which goal is to automatically segment and analyse video data, A4SM (see figure 2) intends to intelligently support the movie design process by recording the history of the movie preproduction process with simple, but yet efficient end-user indexing and media structure visualization tools. Therefore A4SM complements the Informedia system and supports with its collection of meta-data automatic and semi-automatic analysis methods.

Associated with the underlying database are a collection of diverse tools for the gathering, management and visualization of meta-data. Each tool is independent from the other ones, but accesses the data pool stored in the central database. In order to be successfully accepted by the end users, the A4SM environment has to fulfil some basic requirements:

i. Seamless integration into the production process

ii. Interoperability with standard applications
iii. Availability on various hardware platforms and operating systems
iv. Simple, but intuitive user interface

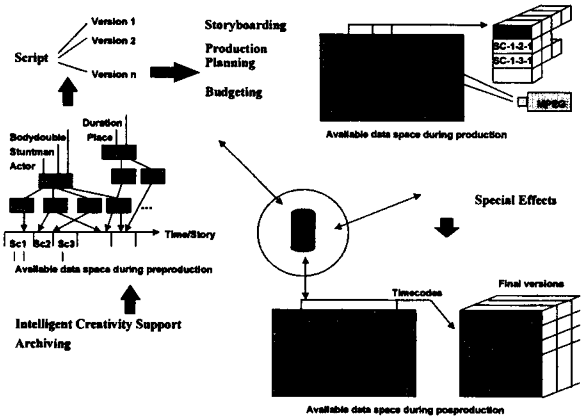


Figure 2: Overview of the A4SM environment

Therefore the framework is realized as a distributed client-server system based on internet technologies and conforming to current and evolving standards like Java (iii.), XML (ii.) and MPEG-7 (ii.) [7].

As an example I will shortly describe the authoring system, which will be used during preproduction for script and storyboard design. Its functionality is based on word-processor and CASE tools technology together with tagging functionality and a concept editor. As the editor will be used mainly by the continuity staff, scriptwriters and directors, it is essential to not constrain the index vocabulary to a predefined lexicon, but to allow them to use their own classification terms. The use of a lexicographic database like WordNet [10] in combination with a knowledge representation system like FramerD [3] is therefore meant to be used for meta-communication, in order to allow easier software based matching of story and discourse structures. Scenarios can be developed either with a text or a graphical editor. The authoring suite is supplemented with a 3D set design environment for the previsualization of motion paths and camera paths (see figure 3).

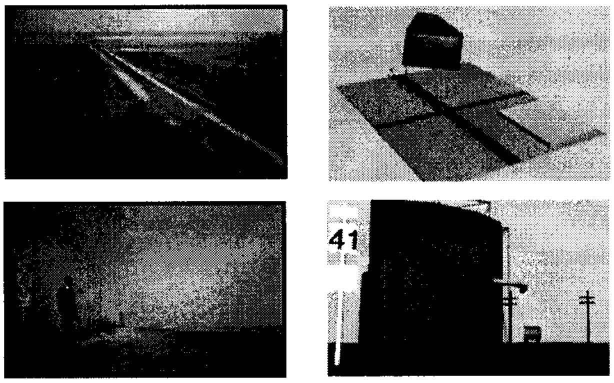


Figure 3: On the left side you can see original still pictures from the movie's scene and at the right its representation in VRML. In the right picture at the bottom the red object represents one shot of the scene. The yellow arrow indicates the direction of the camera

One interface metaphor, that is used both in the authoring suite and in the camera meta-data visualization component, is the visualization of visual perception. This metaphor allows the cognitive compression of time, by transforming a time-based process into a spatial one. If we consider the human visual perception and also the image capturing process of a camera as a projection of the viewing frustum into the surrounding space over time (see figure 4), we can make this model visible by using the computer as a tool to record space.

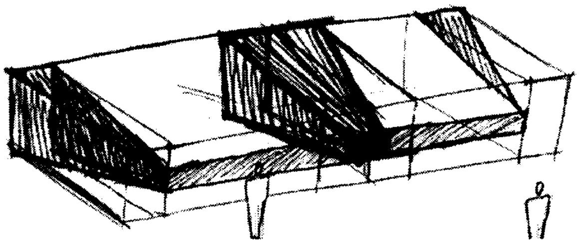


Figure 4: A representation of visual perception by recording the perceived space over time. The cross section of the viewing frustum is extruded along the viewing path.

The model is computed from a set of time varying function curves, e.g. the eye/ camera path, path of the focus, orientation, focal length, depth of field, picture format etc., and generated as a 3d model (see also figure 3). If we use a model of the surrounding space, in which the perception took place, and embed the recorded viewing process into this model we can represent a time-varying process in a static representation. Now, one imagine the film/ video scene by combining both models, without actually having to watch the real video.

5 Conclusion

As A4SM is an environment for effectively capturing meta-data, it can be used with different perspectives in mind. One can see it as a video transcription workbench, an integrated movie authoring tool, a film production management system, a media research toolkit or a basis for a visual thesaurus, which can be used in further analyzing human perception and consciousness.

References

J. Corridoni. *Navigation and Visualization of Movie's Content*, Università di Firenze, 1996

G. Davenport. Smarter Tools for Storytelling: Are they just around the corner? *IEEE Multimedia* 10-14 , Spring 1996

K. Haase. FramerD: Representing Knowledge in the Large. *IBM Systems Journal*, Fall 1996

The Informedia Digital Video Library Project, http://www.informedia.cs.cmu.edu/

G. Lakoff and M. Johnson. *Metaphors we live by*. University of Chicago Press, 1981

D. B. Lenat. CYC: A Large-Scale Investment in Knowledge Infrastructure, *Communications of the ACM*, 38(11): 33-38, 1995

MPEG-7, http://www.darmstadt.gmd.de/mobile/MPEG7/

L. Seger and E. J. Whetmore. *from script to screen. the collaborative art of filmmaking*. Henry Holt 1994

T. G. A. Smith. *If you could see what I mean ... Descriptions of video in an anthropologist's video notebook*. Master's thesis, MIT, 1992

WordNet, http://www.cogsci.princeton.edu/~wn/