Hybrid Narrative and Categorical Strategies for Interactive and Dynamic Video Presentation Generation

Craig A. Lindley
CSIRO Mathematical and Information Sciences
Craig.Lindley@cmis.csiro.au

Frank Nack
CWI, Amsterdam
Frank.Nack@cwi.nl

ABSTRACT

There are a number of different approaches for automatically selecting video clips from a video database and sequencing them into meaningful presentations for viewers. The video database represents a multidimensional video hyperspace, and the sequencing algorithms function as (interactive) dynamic linking and path generation techniques within this hyperspace. Sequencing has been based upon either a narrative or a categorical model of video form. Each of these forms has its respective advantages and disadvantages, and varying suitability for different applications. The two primary forms may also be combined into several hybrid forms, both at the same level and at different levels of the syntactic composition of video sequences, to provide more options for authoring interactive dynamic video productions. Narrative, categorical, and hybrid sequence generation strategies can be applied to a variety of media modalities, including the automated generation of behaviour within virtual environments and computer animations

INTRODUCTION

Adaptive video presentation generation involves the creation of user and task specific video presentations from a database of highly recombinable video components. This supports the creation of video programs tuned to viewer needs and preferences, and encourages a potentially high degree of reuse of the video clips in the underlying database. Systems for adaptive video presentation generation have been based upon either a continuity-edited narrative model of video form (1,2), or a categorical model (3,4,5). Each of these methods relies upon a specific model of video syntax, and it is in the terms of that syntax that sequences are explicitly assembled by an algorithm.

Each syntax model is a model of how meanings of a particular kind are created by conjoining subsequences having specific independent meanings. An algorithm that dynamically creates new meanings by clip juxtaposition requires both rules for how that meaning is created from particular clip meanings, and representations of the meanings associated with the clips in the database. Conjoined video clips can have additional meanings (to authors or viewers) encoded in the perceptual structure and interpretation of the audiovisual data, but the sequencing algorithms can only directly process meanings that are explicitly and symbolically represented within the system. Hence perceived meaning is a function of the algorithm(s) used + the audiovisual content of subsequences + the symbolic representations associated with subsequences. This creates a complex authoring task. Very simple (and possibly automatically derived, low level) descriptors might be used to create interesting presentations if the underlying clips are carefully designed. On the other hand, more complex descriptors can provide more explicit control of the meanings of presentations, but then the authorship of the symbolic descriptions becomes a significant creative task. Combined sequencing strategies are of interest for creating a richer syntactic structure for video presentations, and also to provide alternative methods of sequence or subsequence generation when the available video cannot satisfy the sequencing requirements of a single technique.

This paper presents two distinct models of video syntax, namely narrative and categorical forms, and then describes approaches for automated sequence generation for these forms. The discussion is particularly concerned with creating coherent and meaningful video presentations by conjoining previously assembled and fully composited video clips or clip subsequences into linear presentation sequences; real-time compositing and overlaying of independently stored audio and visual data is not considered (6). The paper goes on to describe a number of strategies by which narrative and categorical sequence generation may be

combined. Combining the strategies provides authors of the hypermedia space with a richer language for expressing the combinatorial potential of discrete video clips, and provides viewers of generated presentations with a correspondingly richer potential for interaction during presentation generation processes.

MODELS OF SYNTACTIC FORM FOR FILM AND VIDEO

Different theorists focus on different qualities that characterise narrative (8), and the meaning of the term "narrative" varies from narrow interpretations involving strong spatio-temporal continuity to very broad interpretations in terms of the overall formal, rhetorical, or thematic coherence of a production. Narrative in a broad sense has been the goal of numerous research projects dealing with diverse media, from text (e.g. 9) to interactive 3D systems (e.g. 10) and video. When narrative is understood in the specialised sense of causally interconnected actions and events, it is useful to define the following alternative, non-narrative forms for the organization of cinematic material (derived from 11):

> *Categorical* films use subjects or categories as a basis for their syntactic organisation, typically basing each segment of the film on one category or subcategory. Common examples of stereotyped categorical films include lifestyle and gardening programs, travelogues, and sporting programs.

> *Rhetorical* films present an argument and lay out evidence to support it, with the aim of persuading the audience to hold a particular opinion or belief. Common examples of rhetorical films are television commercials.

Narrative and *rhetorical* film forms are both distinguished from categorical films by their creation of new meanings by the sequential association of initially distinct video sequences. In contrast to this, any basic video component in a categorical film can represent a designated categorical meaning expressed in an annotation irrespectively of what precedes or follows it. A hierarchy of categorical meanings can define the overall form of a categorical film, but any individual subsequence within the film will have the same categorical meaning that it has within the overall sequence.

Each of these forms represents a different (partially codified) syntactic structure for film sequences. In most real films, the forms apply at multiple levels of film structure, a given film sequence may involve multiple forms at the same level, and multiple forms may occur at different levels. These formal models must also be regarded as greatly simplified. However, the simplifications support convenient algorithmic interpretations in the context of interactive video sequencing: computational techniques can be defined for generating cinematic presentations from databases of video material based upon these simple formal models. Properties of coherent cinematic productions that are not explicitly addressed by these models must be addressed by careful composition of the underlying video data.

A CATEGORICAL VIDEO SEQUENCING ALGORITHM

The CSIRO FRAMES project has developed a categorical system for dynamic video sequence synthesis (5). Dynamic presentation synthesis redefines the concept of a video as a presentation instance from a large set of potential instances represented by the underlying database of video clips. Such a presentation instance can be regarded as a *virtual video*, in the sense that a video presentation is perceived as a traditional linear video presentation, but there is no fixed representation of any predefined presentation order of video clips in the system. The generation of dynamic virtual videos in the FRAMES system is based upon annotations of stored video, together with a specification of the videos that are to be created, and queries embedded within specifications expressed using descriptors common to the content models. Video annotations are based upon a multi-level model of video semantics (12,13). Once video components have annotations generated for them, the annotations are stored in a database. The high level structure of a virtual video program is expressed in a virtual video prescription that can incorporate direct references to specific video components, parametric queries based upon exact or approximate matching of annotations to a query expression, and specifications that initiate the generation of a categorical chain of video content. The generation of video sequences by categorical chaining was first demonstrated in the

MIT Automatist system (3,4). The FRAMES prototype extends this concept with the development of a multi-level semantic model for video, a flexible specification language and chaining algorithm, and a weighting mechanism to determine the breadth and depth of semantic categories expressed by a presentation. The FRAMES association specification language supports the specification of annotation types, initial values, soft constraints, weights, and termination conditions for the generation of a categorical video sequence.

Associative chaining in the FRAMES system is a method of generating video sequences based upon patterns of similarity and dissimilarity in annotations. Chaining starts with specific parameters that are progressively substituted as the chain develops. At each step of associative chaining, the video component selected for presentation at the next step is the component having annotations that most match the association specification when parameterised using values from the annotations attached to the video segment presented at the current step. The high-level algorithm for associative chaining is:

1. initialise the current state description according to the association specification. The current state description includes:
   - the specification of annotation types that will be matched in the chaining process,
   - current values for those types (including NULL values when initial values are not explicitly given),
   - conditions and constraints upon the types and values of a condition, and
   - weights indicating the significance of particular statements in a specification
2. Generate a ranked list of video sequences matching the current state description.
3. If no video sequence in the ranked list has a rank > a specified minimum rank, go to 7.
4. Replace the current state description using annotation values from the most highly ranked matching video component: this becomes the new current state description.
5. Output the associated video component identification for the new current state description to the media server.
6. If the termination condition (specified as a play length, number of items, or associative weight threshold) is not yet satisfied, go back to step 2.
7. End.

This algorithm is illustrated by the flow chart shown in Figure 1. Since associative matching is conducted progressively against annotations associated with each successive video component, paths may evolve significantly away from the annotations that match the initial specification. This algorithm has been implemented in the FRAMES demonstrator. Specific filmic structures and forms can be generated in FRAMES by using particular annotations, association criteria and constraints. In this way the sequencing mechanisms remain generic, with emphasis shifting to the authoring of metamodels, annotations, and specifications for the creation of specific types of dynamic virtual video productions. The basic form created explicitly by the chaining engine is categorical. The data model associates typed annotations with video segments. Annotation types can be created that represent category types, and the annotations themselves can be category names. An associative chain is initiated by sending an association specification to the association engine. The specification includes the category types to chain on, as well as initial category values and possible constraints upon values. The rate at which the categories change within each type is determined by the specified weighting attached to the type: the higher the positive weighting, the more slowly the categories will change within that type, while the more negative the weighting, the faster the categories will change. Hence for n category types, the association engine moves through a video annotation search space of n dimensions.
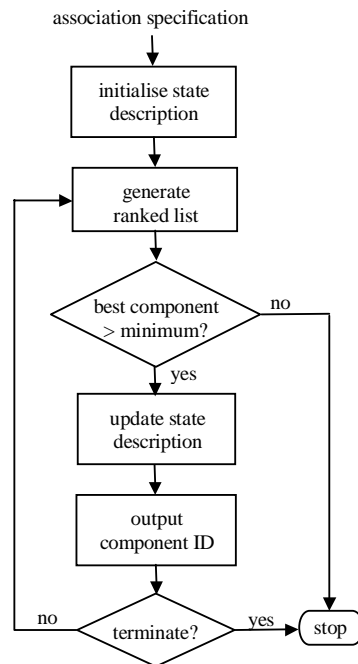
association specification

```
              ┌──────────────┐
              │ initialise state │
              │ description  │
              └──────┬───────┘
                     │
        ┌────────────▼───────┐
    ┌──▶│ generate           │
    │   │ ranked list        │
    │   └────────┬───────────┘
    │            │
    │        ◇───▼────◇        no
    │       ╱ best      ╲──────────┐
    │      ╱ component   ╲         │
    │      ╲ > minimum?  ╱         │
    │       ╲───┬───────╱          │
    │           │ yes              │
    │   ┌───────▼──────┐           │
    │   │ update state │           │
    │   │ description  │           │
    │   └───────┬──────┘           │
    │           │                  │
    │   ┌───────▼──────┐           │
    │   │ output       │           │
    │   │ component ID │           │
    │   └───────┬──────┘           │
    │           │                  │
    │  no   ◇───▼────◇   yes   ┌───▼──┐
    └──────╱ terminate? ╲─────▶│ stop │
           ◇──────────◇       └──────┘
```

Figure 1. Association Engine Flow Chart.

The quality of a video presentation generated by the association engine depends crucially upon the annotation space design (i.e. the category types, specific categories, and the associations created between categories and video clips), and upon the design of the video clips sequenced by the algorithm (5). It is particularly important for this method of video sequencing that the start and end segments of video clips are compatible with the meanings that the system author wishes to convey by their conjunction. In other words, synthesizing a linear video presentation requires a *rhetoric of arrival and departure* (14), i.e. cues that make links between hypermedia components meaningful and coherent from the perspective of the viewer traversing the links. In the case of the linear presentation order of video clips, it may be more appropriate to refer to a *rhetoric of montage*, referring to the system of semiotic codes used to ensure that transitions between clips are meaningful and coherent within the context of the production as a whole. For categorical productions, this may mean avoiding expectations of narrative continuity. For narrative productions, rules must be followed to satisfy expectations of the continuity of action between cuts, and ensuring that discontinuities convey intended meanings.

A NARRATIVE VIDEO SEQUENCING ALGORITHM
As mentioned above, continuity-edited narrative film is concerned with the creation of a pattern of cause-effect relationships among the diegetic events, actions, and situations represented by a film. Basic video components must also be selected and arranged so that the spectator appreciates the intended themes of a production, which are conveyed not just by the presentation of a causally interconnected sequence of events, but also by how those events are audio-visually represented. The computational model for narrative construction is based upon film theoretical analyses of narrative, and film semiotics (15,16,17,18,19,20,21).

Narrative in general is about constructing stories, where a story is a psychological entity that refers to mental or conceptual objects such as *themes*, *goals*, *events* or *actions*. The dynamics within plot construction are twofold. On one hand, the intentions of the narrator must be achieved, and this relies on communication strategies between narrator and receiver organised around surface structures (*expression*)

and deep structures (*content*). *Substance* and *form* can also be distinguished, where substance represents the natural material for content and expression, and form represents the abstract structure of relationships which a particular medium demands (22).

A narrative video sequencing algorithm

The following algorithm for automated narrative video sequencing has been implemented in the AUTEUR system (1). The user provides AUTEUR with the identifier of a start shot and the thematic orientation of the event to be created (such as *humour*).

1 Analyse the content of a user-selected start shot.
2 Develop the story in accordance with the thematic strategy chosen.
3 Establish the appropriate form of presentation for the thematic orientation and story content.
4 Retrieve the appropriate visual material.
5 Edit the material.
6 Present the visual story.

This algorithm can be implemented using a number of interacting planners for story development, a knowledge base, and video database, as shown in Figure 2. The *Video database* is a collection of digitised video material (e.g. stored in MPEG form). The video representation formalism in the knowledge base is designed so that a minimally interpreted representation of a shot (an "objective" description) can be maintained, supporting the reuse of shots to create different meanings in different contexts. The description of each shot is hierarchical, descending from general features to specific details. Each shot description contains a header, descriptions of cinematographic devices, and information about the *shot content.*
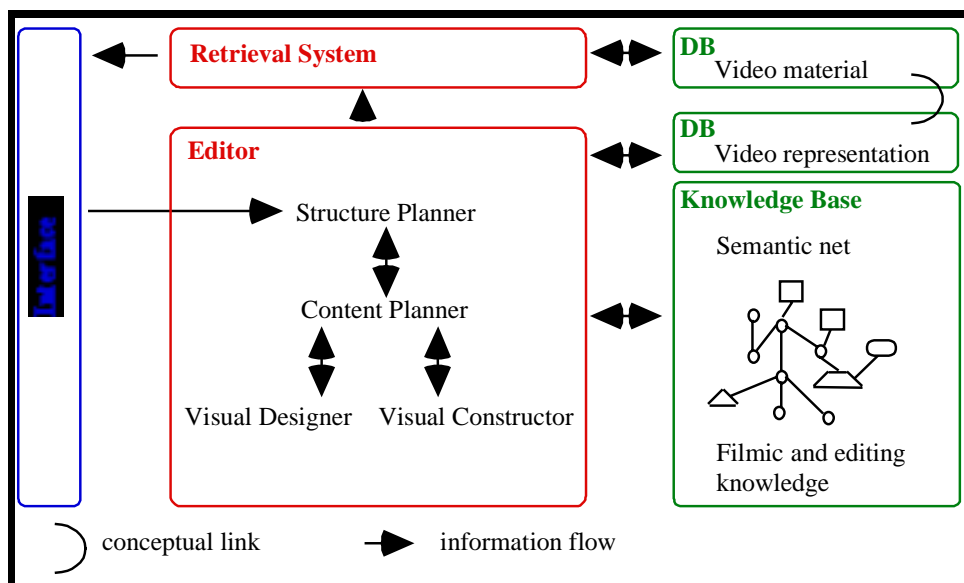

Figure 2. Architecture of AUTEUR

The representational function of a shot is subdivided into two nested structures, the foreground and background, each containing information about mise-en-scène (time, location), the appearance of a character (age, race, etc.), actions (including speed, direction of movement, etc.), the appearance of objects, the events in which objects are involved, and various relationships between different shot elements.

The Knowledge Base

The *knowledge base* contains conceptual structures representing events, actions, and visual features, underpinned by a network of semantic fields supporting:

- *subaction links,* pointing to actions performed concurrently with, and forming a conceptual unity with, a main action. These links are given tags from an arbitrary qualitative modal scale, e.g. necessary, non-essential, etc.
- *opposition links*, specifying antonymous actions
- *synonym links*, indicating co-occurring events
- *ambiguity links*, specifying actions that are subactions of others
- *association links,* pointing to actions associated with another
- *intention links,* linking actions that relate to the goal of an action.

The representation of actions is influenced by that used in Lehnert *et al.* (23), Schank (24), and Schank & Abelson (25). Each action is further described by features such as the objects involved, body parts, spatial relationships between objects and body parts, location, relationships between objects and location, the emotional state of an actor after the action, and possible outcomes of the action.

A scene is represented by an event, denoted by name, the number of actors or objects involved, gender of actors, the intentional state, the main actions involved, and a link to the next higher element within the story structure. The main actions are divided into the three event stages, *motivation*, *realisation* and *resolution*, each containing a sequence of actions for each actor. An event is hence modelled as a structure of composed pre-conditions, main-conditions and post-conditions. The pre-conditions, referred to as the *motivation*, introduce the characteristic objects, locations, activities or moods of characters necessary for the main part of the event. The object, action or the perceived mood of a character suggest certain possible events that can occur, and expectations that may be realised. The appropriate plot structure for the main-condition of the event is described as its *realisation*. A particular realisation may lead to the expectation of certain reactions, which provide additional clarifying information. These post-conditions are part of a phase referred to here as the *resolution* of the event.

Episodes are described by scripts featuring descriptions of the type of characters, their appearance, their mood and a collection of events for the motivation, realisation and resolution stages of an episode. Visual concepts describe how a theme can be visually realised. For example, the concept *pleasure* might be expressed by a smile, or through actions being performed at medium speed, giving the impression that the character is relaxed. The combination of both offers a stronger visual impression than either alone. Finally, the system contains conceptual knowledge of objects, locations, directions and abstract concepts such as time or justice.

The Editor

The Editor is the video selection and sequencing system within the AUTEUR architecture. The architecture of the editor embodies the separation of the two main story layers, i.e. *structure* and *content*. Each layer is provided with its own planning system. The content planner is assisted by two specialised planners, the *Visual Designer* and the *Visual Constructor*. Communication between the different planning systems is based on the memory structures described in more detail in Nack (1).

The Structure Planner

There is initially a purpose to a story, which unites the story's separate elements. This purpose, which can be called an *external reason* (26,27) or *idea*, is the *theme* of the story. To make a story coherent, at least one theme must be represented. However, several overall themes are often represented, and, at the same time, each part of the story may reflect its own theme.

The task of the *Structure Planner* is to establish the relevant themes for presenting the material (based on the content analysis of the material) and then to organise strategies for realising the required theme. The

Structure Planner is involved in the creation process from the outset, providing the analysis of the given start shot, which in turn is used to establish the first *Sequence-Structure*. This analysis, based on the start shot description in the video representation, provides information about the number of actors, groups of actors, objects and groups of objects. Each of these units is related to particular information about their actions, i.e. sequences of actions, actions performed concurrently (i.e. during the same sequence of frames), and single actions. The information acquired is stored in the *Setting, Subject*, and *Action* fields of the first *Sequence-Structure*.

The next task of the *Structure Planner* is to use the results of the start shot analysis to establish the appropriate thematic strategy. The preparation process ends with the completion of the *Sequence-Structure*, by instantiating the fields *Kind*, *Intention*, *Form* and *Appearance*. Once the essential structural elements for the event to be created have been declared, the *Structure Planner* supports the construction process by providing the *Content Planner* with additional information concerning the current event phase. The decision process of the *Structure Planner* is based on feedback from the *Content Planner*. The *Content Planner* provides information about the meaning of the story, the motivated event, mood, action, actors involved, etc.. The *Structure Planner* compares the actual event provided by the *Content Planner* with the suggested content in the original *Sequence-Structure*. The results of this comparison serve as a guide to establishing the *Sequence-Structure* for the following event phase.

The Content Planner

While the Structure Planner deals with the external point, or theme, of a scene, the *Content Planner* is concerned with its *internal* point (26,27), and specifies the content of the scene. It is therefore responsible for the detailed application of a particular thematic strategy. Depending on the strategy and its specification provided by the *Scene Planner*, the Content planner uses conceptual structures gathered from the semantic net of the *Knowledge Base* to attempt to construct a coherent scene.

The Content Planner consists of three independent planners. The *motivation*, *realisation*, and *resolution* planners are specialists for the content generation of their respective event phase. Each planner uses the event-related *Sequence-Structure* and the semantic net of the *Knowledge Base*. Each also collaborates with the *Visual Designer* and the *Visual Constructor* (described below) to establish the visual presentation of their respective part of the scene. If no solution exists, the Structure Planner is requested to change the strategy.

The Content Planner also determines the point of view from which the story takes place, such as from a particular character's perspective. This might indicate the need to change from a third person to first person narrative style. This, in turn, influences the *Visual Designer,* which has to search for shots that succeed as eye-line matches.

The Visual Designer

The *Visual Designer* supervises the retrieval of video material. A content-based query is provided by one of the motivation, realisation or resolution planners (of the Content Planner). The Visual Designer then retrieves the video that is most appropriate in terms of content and style. The mechanisms used are a representation of the permissible relations between shots (based on Vertov's classification - see 28) and various editing codes.

Shots can be juxtaposed in essentially three ways: putting one shot either immediately before, immediately after, or inserting it within, another. The Visual Designer contains rules to determine a favourable shot for a given intention, e.g. an *insert* corresponds with *highlighting*. The most appropriate shot kind for an insert is a *close-up*. The shot type, while important, is only one criterion affecting the retrieval process. A second important requirement is *continuity*, which is determined by examining the content descriptions of the shots to be joined.

The Visual Designer uses editing rules to direct the ordering of shots such as:

- to highlight use a zoom-in,
- to emphasise the dominance of a character use low-angle shots,

Each editing rule has an associated applicability value, which is added to the evaluation value of a shot if the editing rule is applicable. The highest valued shot is ultimately chosen. The output from the Visual Designer is a shot list (called a *Location-Memory-Structure*) for the related event phase, representing the content query in the most appropriate visual terms, which is then transferred via the Content Planner to the *Visual Constructor*. If the Visual Designer is unable to retrieve visual material, as requested by the Content Planner, spatial and general knowledge structures from the knowledge base are used to decompose the query into sub-queries. If no material can be found to realise a particular narrative, the Visual Designer notifies the Content Planner, which alters the storyline until a scene is successfully created.

The Visual Constructor

The *Visual Constructor* receives an annotated shot list from the Visual Designer, and specifies the detailed joining of the shots. The Visual Constructor operates at the *cutting* level. A shot may be truncated if it is too long for the required purpose, and cuts may be motivated if only part of a particular shot is required. The output from the Visual Constructor is an ordered list of shot identifiers, along with frame numbers for each shot, specifying the scene to be displayed.

The Retrieval System and the Interface

The *Retrieval system* adapts the final stream of shot ids and frame numbers (produced by the Visual Constructor) into a file specifying the actual presentation of the scene as a list of MPEG files and associated start and end frames, which can be displayed to the system user or viewer.

Narrative Video Sequencing – an example

A simplified example shows how the different elements of AUTEUR's architecture combine to produce a humorous film. Here we depict shots in the film by using a single image from each shot. Suppose that the query by the user is: go(12, humour), where 12 represents the ID of the start shot shown in Figure 3.

Figure 3. Start shot for the banana skin joke

Preparation phase

The *Structure Planner* first analyses the content representation of the given start shot, based on visual expressions and actions related to an actor. The result is a list of possible moods, each tagged with a certainty value that must be higher than 0.35 for a mood to be considered as an element for the list of assumed moods. For the start shot in Figure 3, the moods pleasure and hurry are identified, each with a value of 0.5.

The *Structure Planner* uses this information to establish the *Sequence-Structure* for the current generation phase. First the most suitable humour strategy is selected. As there is only one person, one action, and there is nothing ambiguous about the action, AUTEUR suggests *misfortune* as the humour type for the joke.

The next step for the *Structure Planner* is to determine the phase of construction. Since the concept of misfortune requires mood deterioration, the *Structure Planner* evaluates the mood of the relevant character. Neither "pleasure" nor "hurry" provide the required certainty value of 0.75 that would indicate that the mood of the character is clearly perceptible from the shot, so the mood is indexed as "to be motivated". Since there is only one character performing one action, there is no need to motivate the action. However, since the conceptual structure of "walk" can be associated with a larger logical sequence, e.g. a meeting, the motivation of an event is also suggested. On completion of the preparation phase, the first *Sequence-Structure* is instantiated.

The *Structure Planner* now informs the *Content Planner* to proceed, by sending a list of supportive humour strategies.

Motivation phase

The first task of the *Content Planner* is to identify the appropriate strategy for the humour type. Based on the set of humour strategies provided by the *Structure Planner*, the *Content Planner* evaluates the available information concerning the visual material. For a single character performing a single action the appropriate humour strategy might be the following:

H-Strategy 4   *If the action portrays an intention (goal), interrupt the action, in a way that is expected by the character, so that the goal is unfulfilled and the character's mood is downgraded or he suffers in some way.*

Given this strategy, the *Content Planner* attempts to create a motivation for a mood and an event. The first aim is to create a visual representation that suggests that the character either feels *pleasure*, or is in a *hurry*. The second aim is to establish an event that corresponds with the chosen mood.

The *Knowledge Base* contains a number of mood concepts, along with related actions. For example, *pleasure* may be associated with *smiling*, *whistling*, and *picking flowers*. AUTEUR selects the action *smiling* as the strongest action providing a visual representation of "pleasure". Traversing the associative links of "smiling" and "walk", AUTEUR infers that a person can walk and smile at the same time. As a result, a query is sent to the *Visual Designer* to retrieve an appropriate visual representation for "Frank walks and smiles".

To find appropriate visual material for the query, the *Visual Designer* uses two knowledge structures: the *Knowledge Base* model of the spatial relationships between two shots, and the conceptual relationships between visual space and narrative functionality. From the representational structure for the cinematic devices in shot 12, which is stored in the *DB of Video representations*, the *Visual Designer* detects that the start shot is of type "long". From the conceptual relationship between visual space and narrative functionality, the *Visual Designer* infers that motivation favours detail, which is related to a decrease of space. This information results in the *Visual Designer* exploring the array of spatial relationships between shots. The *Visual Designer* decides to join a "long" shot and "close-up" shot for the purpose of the motivation of the mood. Now, assume that the *Visual Designer* can retrieve a number of shots from the *DB of Video representations*, which are annotated with the required action, "walk", for the given character (*Frank*), and the facial expression of a smile, as shown in Figures 4, 5, and 6.

The representation of each of the shots is compared with that of the shot to be joined on the basis of spatial continuity, action match, temporal continuity, and stylistic features. Since the intention of the join is a motivation, a zoom-in is stylistically desirable. Based on the evaluation process, the *Visual Designer* chooses the shot represented by Figure 5.

After establishing the mood "pleasure", the *Content Planner* can specify an event by constructing an appropriate goal for the character. Thus, the *Content Planner* traverses the causal links provided by the attribute *Intention* of the conceptual structure for "walk", to detect usable event structures. One such goal

might be to meet another person. Thus, the *Content Planner* attempts to construct the representation for a *meeting*.

Figures 4, 5, and 6. Three possible motivation shots for the banana skin joke

Since the *Content Planner* is currently operating in the "motivation phase", it also targets the motivation parts of the event structure "meeting". In our example, the action of one character is already specified, i.e. "walk", so AUTEUR searches in the motivation field of the conceptual structures for "meeting", for a corresponding action performed by the other character. In this case, a corresponding action is "wait".

This information is used by the *Content Planner* to send a query to the *Visual Designer*, which must retrieve a shot of a character who waits in a similar surrounding to that provided by the start shot (Figure 4.1). The *Visual Designer* may suggest the shot represented by Figure 7.

Figure 7. Event shot for the banana skin joke

Due to the strategy (H-Strategy 4), the *Content Planner* orders the actions according to their appearance in the shot sequence and transfers this information to the *Visual Designer*.

The *Visual Designer* must decide how the shots should be joined. The juxtaposition between "Event shot" and "Start shot" is a simple join of two long shots in the given order. The combination of the motivational intention for the join, in combination with the particular combination of shot types (close-up motivates long shot), implies that the juxtaposition of "start shot" and "motivation shot" should be performed as an insert.

The *Content Planner* now generates a *Location-Memory-Structure*, in which the ids of the shots in their established order are stored. Since all content and stylistic requirements for the visual presentation of the motivation phase have been achieved, the *Content Planner* marks the evaluation value for the motivation as successful.

Finally, the *Content Planner* indicates the status of the ongoing story to the *Structure Planner*. The *Structure Planner* then uses the story status to decide on the next phase in the generation process.

The comparison between the *Sequence-Structure* for the motivation phase and the status of the motivation phase reveals that a new character, the waiting man, has been introduced. Since both characters are still apart, and the new character is passive, the *Structure Planner* infers that the joke is still to be based on the action of the main character. However, the introduction of the new character changes the single-person environment into a person-person environment, indicated by changing the *Intention* field for the motivation *Sequence-Structure* from *event* to *parallel_event*. Also, the fields *Setting*, *Subjects* and *Action* in the motivation *Sequence-Structure* are updated with information about the second character. The humour type (misfortune) remains, and the established strategy (H-Strategy 4) is added.

The next activity of the *Structure Planner* is to decide on the next phase of the generation process. Since the emphasis continues to be on one character, there is no change in action, and there is no change in the strategy type, the *Structure Planner* suggests a realisation phase and then instantiates the relevant *Sequence-Structure*.

Finally, the *Structure Planner* instructs the *Content Planner* to continue the generation process.

Realisation phase

The first task of the *Content Planner* is to retrieve the name of the action or event which forms the basis of the joke. This is indicated by the uninstantiated *Action* field of the realisation *Sequence-Structure*.

To retrieve the required action, the *Content Planner* uses the realisation requirements provided by H-Strategy 4. The aim of H-Strategy 4 is to violate a character's goal under two constraints: the mishap should be simple, and the mishap should be expected by the character. Thus, the *Content Planner* first retrieves the action (walk) of the main character (Frank) from the motivation *Sequence-Structure*. The violation requirement of H-Strategy 4 causes the *Content Planner* to traverse the outgoing opposition links of the conceptual structure "walk". The oppositional links are chosen because the concept of "interrupt" is related to the concept of "perform opposition". The links lead to conceptual structures for oppositional actions for "walk", such as *fall, slip, stumble,* or *collide with*. The *Content Planner* attempts to instantiate one of these oppositional concepts.

Assume the conceptual structure of slip is selected. An important task for the *Content Planner* is to detect if the *result mood* of "slip" can fulfil the required mood deterioration. Since *anger*, *rage* and *astonishment* are related to the emotional token of the opposite classification type of *pleasure*, i.e. *displeasure*, the mood deterioration can be established. Thus, "slip" is a feasible action for foiling the action "walk".

The *Content Planner* can now present the *Visual Designer* with a number of shot queries, considered in descending order of suitability. Suppose that that the query "find a shot where a body part *slips* on an object, where the object is found in the start shot" succeeds. The *Content Planner* must then satisfy the expectation constraint of H-Strategy 4, that the character is aware of the object. Since no retrieved shots portray the line of sight of the character, the *Content Planner* uses the substructure "Bodygesture" from the action representation of the actor, taken from the last shot of the previous *Sequence-Structure* in which the character appears. The comparison of the Relation Object -> Bodypart with the line of sight of the Bodygesture reveals that the character, Frank, is actually not looking to the ground and thus there is no reason for the *Content Planner* to assume that the character is aware that he is about to slip on an object. Thus, H-Strategy 4, which requires an expected mishap, is not applicable.

The *Content Planner* could now continue the search for another suitable opposition action. However, assume that the *Content Planner* investigates weaker mishap strategies, such as H-Strategy 2, which requires an unexpected mishap.

H-Strategy 2   *if the action portrays an intention [goal], interrupt the action in a way that is unexpected by the character, so that the goal cannot be fulfilled and the character's mood is downgraded or he or she suffers in some way.*

This strategy corresponds with the material of the motivation phase and would fulfil the constraint of unexpectedness, since in the shots of the motivation phase the character is not looking towards the ground. As a result, the *Content Planner* changes the strategy ID in the realisation *Sequence-Structure* from H-Strategy 4 to H-Strategy 2, but marks the switch to indicate that the joke can be improved by using a higher order strategy.

The *Content Planner* also attempts to apply constructive H-Strategies related to misfortune, which might improve the joke. An example is H-Strategy 5, which states that "*If the intention of the joke is derisive, reveal the point in advance. (enhanced Schadenfreude)*". As a result, the *Content Planner* sends a query to the *Visual Designer*, requesting a detail shot of the object the character is to slip on, as the spectator will then anticipate the mishap and this is predicted to increase the potential success of the joke. The retrieved shot is that shown on the left of Figure 8.

The *Visual Designer* then analyses the content and style of the potential material, following which the *Visual Constructor* specifies the detailed joining of the material. The final outcome is a two-shot scene as suggested by the stills in Figure 8.

Figure 8. Realisation part for the banana skin joke, generated out of two shots

The *Content Planner* now evaluates the realisation part of the joke. Since all requirements of the strategy could be fulfilled, the evaluation value is in the range "good". Finally, the *Content Planner* updates the *Location-Memory-Structure*, and then indicates the status of the realisation phase to the *Structure Planner*.

The *Structure Planner* once again compares the status of the generation phase (realisation) with the related *Sequence-Structure*. Since the action for the main character is not indicated in the *Sequence-Structure*, the *Structure Planner* updates it with the action provided by the status information (i.e. "slip"). Furthermore, the strategy type must be updated, since the strategy has changed.

The next step for the *Structure Planner* is to decide if a resolution stage is required. Following the concept of *misfortune*, the *Structure Planner* investigates the content representation of the shots generated in the realisation phase with respect to the portrayal of a mood change, either by showing a reaction or a gesture. Such information cannot be found in the constructed material, so the *Structure Planner* suggests a resolution phase, and then instantiates the relevant *Sequence-Structure*.

Finally, the *Structure Planner* instructs the *Content Planner* to continue with the generation process.

Resolution phase
Using the conceptual structure for *slip*, the *Content Planner* constructs a request for video material that portrays an appropriate reaction by the character. The reaction is composed by considering possible resulting states, and, if pertinent, the relevant object. The strategy for choosing between alternatives is based on a preference of reactions to moods, and among competing potential results, the choice is based on the value of the relevant link. Assume that the *Content Planner* sends the following request to the *Visual Designer*: *provide a shot where the character "Frank" looks back at the object (banana_peel).*

In order to ensure a consistent filmic style throughout the scene, the *Visual Designer* attempts to satisfy the request of the *Content Planner* in terms of stylistic aims and the shot history (*Location-Memory-Structure*). Since the realisation phase predominantly makes use of shot types between "medium" and "close-up", the *Visual Designer* attempts to answer the query with a shot within this range of types. Figure 9 shows a frame from a shot that realises this aim with respect to the existing chosen material.

Figure 9. Retrieved shot for the realisation phase of the banana skin joke

Figure 10. The banana skin joke generated by AUTEUR

Once the *Visual Constructor* has established the join, the *Content Planner* evaluates the resolution part of the joke. Since a reaction is shown, rather than a gesture, the realisation is valued as "good". The *Content Planner* then applies the evaluation values gathered for each generation phase to evaluate the humour level of the joke. The overall verdict is "good". However, due to the need to downgrade H-Strategy 4 to the simpler H-Strategy 2, the *originality* is assessed as "average".

Following the *Content Planner's* indication of the successful termination of the joke generation process, the *Structure Planner* seeks options for developing further jokes from the existing story line, or attempts to provide the specification of an appropriate conclusion to the scene. For the above example, the comparison between the *Sequence-Structure*s and the conceptual structure of the event "meeting" reveals that the generated video material neither provides the realisation nor the resolution stage for a meeting. Thus the *Structure Planner* instructs the *Content Planner* to generate one or both of these. If this is unsuccessful, the *Structure Planner* determines that the shot in Figure 7 is superfluous material. This will initiate the sending of a re-editing plan to the *Content Planner*. In the given example, only the motivation *Sequence-Structure* is affected, since no other sequence structure contains information about the character

to be removed. The final version of the banana skin joke, which is roughly 20 seconds long, is suggested by the stills in Figure 10 (read from left to right, top to bottom).

MANAGING COMBINATORICAL COMPLEXITY

Lindley (5) has described principles for the creation of a database and annotation space design that optimise the effectiveness of user interactions with the FRAMES association engine for categorical video productions. This can be done by designing the annotations to fully (the property of *completeness*) and non-redundantly (the property of *minimality*) index the set of video clips, to achieve the kind of maximization of cognitive informational effects for minimum processing effort discussed for hypertext by Tosca (29). This is possible because all possible combinations and permutations of the video components of a categorical production can in principle be meaningful presentations. However, the association engine cannot guarantee narrative continuity between segments, or even simple temporal ordering of the diegetic material represented in consecutive segments. The narrative approach demonstrated by AUTEUR, on the other hand, provides explicit control over narrative continuity. However, for continuity-edited narrative productions, the requirement of continuity imposes severe constraints upon what can be meaningfully conjoined with a particular component. While in principle all possible combinations and permutations of narrative video components could make narrative sense, in practice it is extremely difficult to design a set of components for which the number of coherent continuity-edited narrative presentations is not a small subset of this total combinatorial space. This means that the ratio of coherent presentations to the total number of components in the database will generally be small for narrative productions. For example, in the original AUTEUR experiments, about eighty video clips were used to create three predetermined target narrative structures, with a potential to extend the number of meaningful permutations with only small variations of narrative sequence content.

The range of narratively meaningful combinations could be predetermined and designed (e.g. as a directed network structure of valid continuity-edited sequences). In this case, points of narrative convergence (or strategies for looping, repetition, or reordering; see 30) are required to control the combinatorial explosion of the branching structure. If not, for a presentation in which separate video segments are linked to form a sequence of length n segments, with a branching width of b different narrative options, $\Sigma_{k=1\text{->}n}\ b^{k-1}$ individual video components will need to be created. Limiting b or n reduces the scope of interaction and the number of distinct narrative presentations that can be generated, while as b or n increase, the size of the video database, and hence the video data that must be shot or gathered, edited, etc., increases exponentially in b for individual presentations that may only be of length n components. This is a fundamental problem for the generation of narrative video presentations (31).

Given these respective limitations of narrative and categorical generation, it is natural to seek to merge these strategies in order to create more narrative structure within categorical presentations, or to increase the number of thematically coherent, but not strictly continuity-edited sequences that can be generated within a predominantly narrative system. Merged or hybrid strategies also provide a model in their own right. Strategies for merging these approaches are discussed in the next section.

STRATEGIES FOR MIXED CATEGORICAL AND NARRATIVE SEQUENCING

There are a number of distinct strategies by which categorical and narrative sequence generation might be combined.

1. a narrative sequence generator that cannot proceed due to lack of material and/or exhaustion of rules for satisfying a current goal can use an associative matching step to shift context for lower level goals, after which narrative generation might resume in service of higher level goals.

This strategy might be used within the content planner. The state description that is articulated during the development of a narrative sequence can also be used as video content descriptions and specifications by an association engine. In this case, an associative shift of state is a shift to another state description having a partial match to the state from which the shift is initiated. This will in general represent a state

discontinuity, i.e. a change in the values of at least some state descriptors. The hybrid strategy involves the invocation of an associative state shift in the event of a failure of the narrative generator to articulate a purely narrative sequence. The narrative engine should backtrack to a previously successful step in the generation process, associatively shift the state description, and proceed with narrative generation from the associatively selected clip.

2. a categorical sequence generator that fails to find material matching the current state description at some point in sequence generation can resort to narrative synthesis to create a new sequence from lower level components, where the current state description serves as (part of) the goal for narrative generation. Once a narrative component has been created, categorical sequencing resumes, with a state description modified according to the description of the narrative segment.

In this case, the current state description and the association specification can be used to create a set of state descriptions representing the next most strongly associated state. Those state descriptions can be fed (in decreasing order of associative strength) as goals to the narrative structure planner. The most strongly associated sequence that the narrative generator can assemble from the video database is then used as the next content unit in the categorical video production. This is illustrated on Figure 11. The system seeks to assemble goals as input to the narrative generator, and will terminate if a goal is not satisfied, or else insert the narrative sequence and resume associative processing.
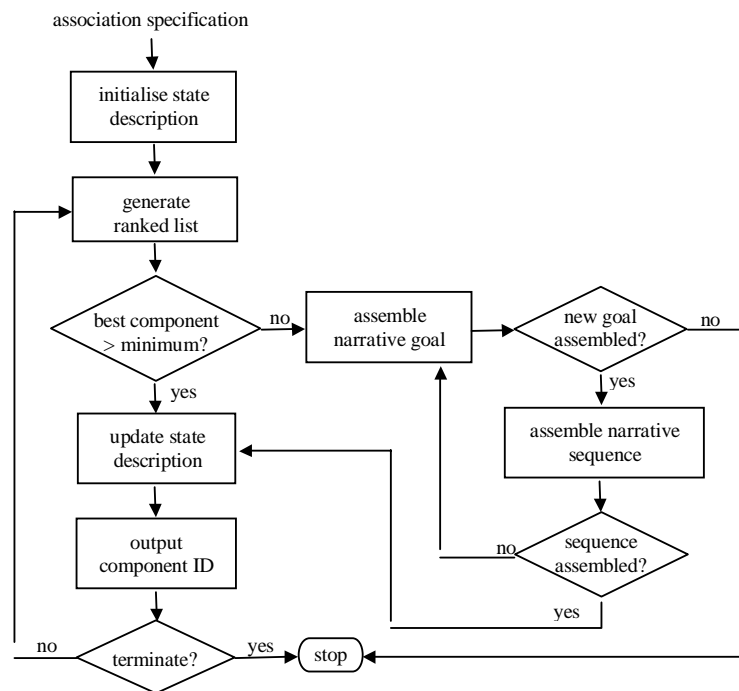

Figure 11. Narrative Synthesis of Associative Components in the event of associative failure.

3. a narrative sequence generator creates a narrative sequence that is then subjected to post-processing by an association engine. This represents a (syntagmatic) reordering of the narrated order away from the diegetic time order of the synthesized narrative material.

This strategy, shown on Figure 12, takes a narrative sequence of clip identifiers produced by the narrative generator, and then applies the association algorithm to the database subset associated with the narrative

sequence. The aim is to define an associative presentation order through the narrative material. This is not completely straightforward, since the clip boundaries defining clips assembled by the narrative sequencer may not be the most appropriate points at which to segment the newly defined presentation for the purposes of syntagmatic reordering. A new virtual database can be defined either by taking some subset of the clip and subsequence boundaries within the virtual narrative presentation, or by defining new subsequence boundaries within the overall virtual presentation. In either case, the new virtual clip set can be based upon analysis of the hierarchical narrative representation created by the narrative engine.
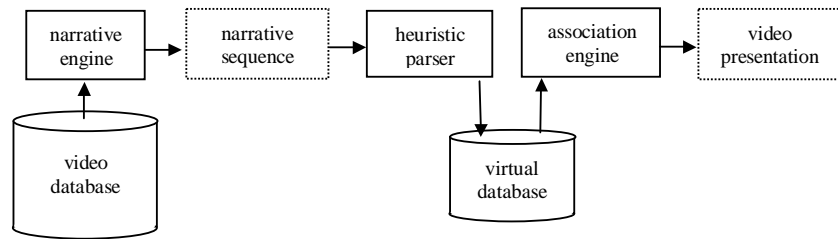
```
┌──────────┐   ┌··········┐   ┌──────────┐   ┌───────────┐   ┌············┐
│narrative │→  │narrative │→  │heuristic │   │association│→  │  video     │
│ engine   │   │sequence  │   │ parser   │   │  engine   │   │presentation│
└──────────┘   └··········┘   └──────────┘   └───────────┘   └············┘
     ↑                              │              ↑
 ┌───────┐                      ┌───────┐
 │ video │                      │virtual│
 │database│                     │database│
 └───────┘                      └───────┘
```

Figure 12. Associative post-processing of narrative sequences.

4. a categorical sequence generator creates a categorical sequence that is then subjected to post-processing by a narrative engine. This represents a (syntagmatic) reordering of the narrated order away from the arbitrary time order of the selected categorical material towards a sequence reflecting the diegetic time order or causal order of the content of those segments.

This is the inversion of strategy 3 above, as shown on Figure 13. In this case, the set of clips identified by associative chaining is used as the database from which the narrative engine seeks to assemble a coherent narrative sequence. This can amount to thematic or topical prefiltering of the material from which a narrative presentation is to be generated, which can simplify narrative generation as well as predetermine the narrative thematically in ways in which the narrative engine and its knowledge base may not be equipped to take into account.
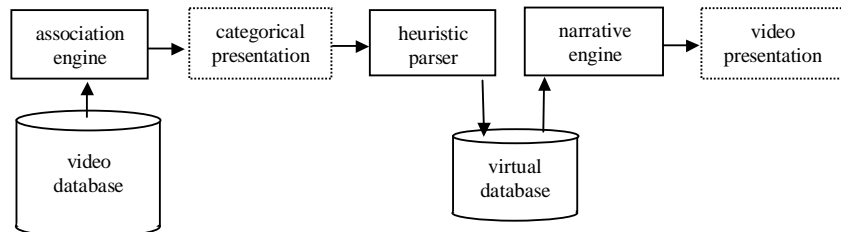
```
┌───────────┐   ┌············┐   ┌──────────┐   ┌──────────┐   ┌············┐
│association │→  │categorical │→  │heuristic │   │narrative │→  │  video     │
│  engine   │   │presentation│   │ parser   │   │ engine   │   │presentation│
└───────────┘   └············┘   └──────────┘   └──────────┘   └············┘
     ↑                                │              ↑
 ┌───────┐                        ┌───────┐
 │ video │                        │virtual│
 │database│                       │database│
 └───────┘                        └───────┘
```

Figure 13. Narrative post-processing of categorical sequences.

5. a narrative sequence generator incorporates specific mechanisms for the explicit insertion of a categorical sequence.

In this case the formalism for encoding narrative and action rules includes a predicate type corresponding to an association specification that can occur within a planning or action generation rule. The most obvious example of when this is useful is when an action intrinsically requires the presentation of a categorical video sequence (e.g. a news reader presents a summary of news highlights).
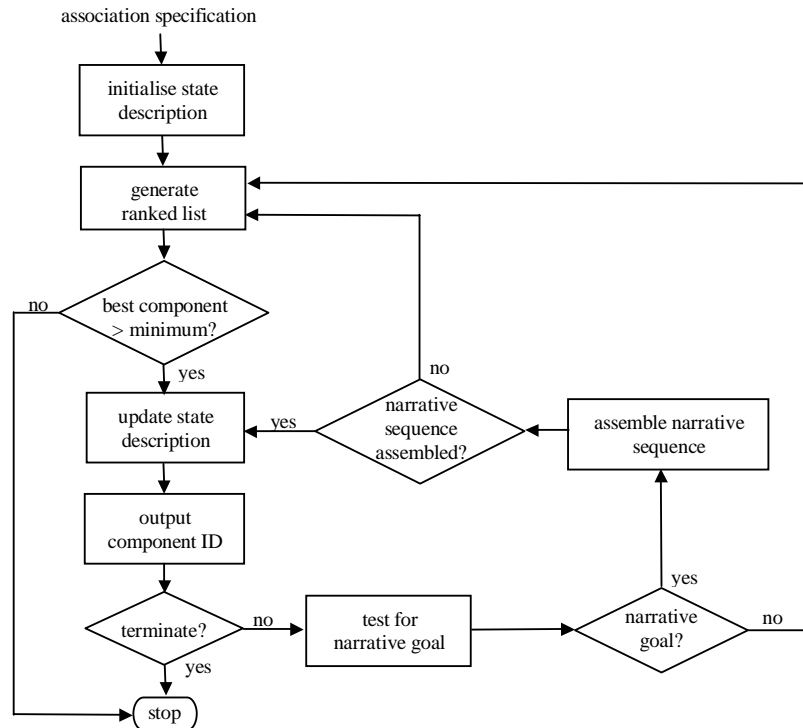
Figure 14. Associative sequence with narrative segment insertion.

6. a categorical sequence generator incorporates specific mechanisms for the insertion of a narrative sequence.

This strategy is invoked during sequence generation, by matching the current state description to ancillary rules for triggering the generation of a narrative sequence; then a particular cycle of the association algorithm uses a synthesized narrative sequence instead of the usual mechanism of selecting a predefined sequence from a video database. A simple mechanism for this is to modify the association algorithm to check the current state description (and specification) against a rule base for detecting when it is appropriate to generate a narrative sequence prior to seeking an associative match. Once the narrative segment has been created, the algorithm iterates as usual. This is shown on Figure 14. It is possible for all of a categorical production's segments to be generated by a narrative engine operating under the higher level control of the association engine, resulting in a categorical production at the top level with synthesized narrative components contained within the overall categorical framework.

7. a narrative sequence generator uses associative mechanisms to structure a narrative presentation as a series of episodes related by theme, topic, etc..

This strategy is very similar to strategy 3 above, but presents the articulation of the categorical structure as an outcome of narrative reasoning. This would occur when the substructure of a narrative goal is intrinsically categorical, so that the decomposition of the goal results in a categorical program framework. That is, decomposition of a narrative goal at a particular level of resolution results in a series of subgoals that are viable state descriptions, but have no narrative interrelationships.

8. a categorical sequence generator uses generated narrative sequences as bridge material to connect categorically distinct segments.

This strategy explicitly interlaces narrative and categorical segments as a high level model of the overall form of the presentation. A mechanism to accomplish this is to use the state description of a current categorically selected segment as the starting state, and the state description of the next segment selected

categorically as the goal state. The narrative generator then assembles a narrative sequence progressing from the current state to the next state, which will be a narrative bridge between categorical segments.

9. the detailed mechanisms for associative matching and action sequencing are integrated.

In this case the progression of selected video material is made by a synthesis of matching on narrative rules according to a high level thematic goal, and matching on patterns of similarity and dissimilarity of associations (represented in annotations) of the individual video segments. This can be envisaged as a kind of fuzzy resolution theorem prover, where predicate matching is partial and weighted with weights modified by an association specification and current state description.

THE AESTHETICS OF HYBRID STRATEGIES

Narrative is a highly conditioned expectation in dominant cinema, with well-defined conventions of establishment, conflict, and resolution (32). However, linear narrative alone leads to a predictable story form, and fails to model many nonlinear pathways in human discourse, behaviour, communication, expression, and cognition. Hybrid continuity narrative/categorical strategies provide a variety of approaches representing different priorities of linear narrative and associative path generation, and a broader range of experiential and expressive possibilities that can be modelled and generated.

Strategies 1 and 2 above propose a primary strategy with the alternate strategy being invoked in the event of failure of the primary strategy to generate a satisfactory solution. This supports a number of interpretations based upon the meaning of the failure of the primary strategy. For a narrative primary strategy, failure can represent the failure of the narrative ideology underlying presentation generation, with the shift into associative linking representing a lapse into "irrational" behaviour under circumstances when the rationalist frame is untenable, the associative episode representing a moment of absurdity in the light of failed reason. Absurdity is itself a rationalist surface interpretation, where the dynamics of ongoing behaviour generation are nevertheless systematic, the system being represented by the principle of association that is invoked. Where the primary structure is categorical, the resort to narrative generation in the event of failure may represent the invocation of instrumentalist reason to cope with deficiencies manifested in the dialectic between a purpose (the association specification under a current contextual state description) and the contingent state of the virtual universe represented by the set of atomic video clips. This is the function of narrative as a discourse of power, finding the world inadequate, and resorting to instrumental construction as a strategy to impose will to create a solution.

Strategies 3 and 7, involving the reordering of a narrative sequence by categorical criteria, or the generation of independent categorically ordered narrative sequences, respectively, may function as strategies for foregrounding thematic elements underlying the narrative. Alternatively, post-processing of a categorical sequence by narrative criteria (strategy 4) may serve to weave the thematic threads of a presentation into a more memorable order (the mnemonic function of narrative), or simply create another level of order to unify the formal integrity of the presentation to a degree beyond that achieved by the initial categorical structure.

The explicit insertion of a categorical subsequence within a primarily narrative sequence (strategy 5) may implement a categorical subsequence as a diegetic element of the narrative, such as the inclusion of a news presentation within a story. Alternatively, the explicit insertion of a narrative sequence by a predominantly categorical sequence generator (strategy 6) represents a more self-conscious form of control over the generated sequence than the 'resort to narrative' approach represented by strategy 2. This might occur when it is known beforehand that the underlying clip database cannot satisfy a particular state description, or where adaptive generation from more primitive elements may more directly serve aesthetic, pedagogical, and/or ideological goals than selection from a predefined database.

Strategy 8, in which a categorical sequencer uses narrative sequences as bridge material to connect categorically distinct segments, can provide a narrative frame for contextualising categorical sequences.

Such a frame can provide high-level formal integrity for a generated presentation, and could include explication of the topical and thematic principles by which the categorical elements have been generated.

Strategy 9, provides a more 'organic' model for the generation of meaning in presentations, potentially modelling the divergence of discourse from a strict linear, rationalist semantic vector, and also from the systematic thematic changes of a pure categorical format. This may represent the meandering of natural conversation (perhaps in the service of higher level but implicit goals of contact, relationship establishment (as discussed by Bickmore and Cassell, 33), or the transmission of information) under circumstances in which the strict linear pursuit of conversational goals is inappropriate (for example, when the agent functioning as the information source must not appear to be too didactic for reasons of relative status). Intrinsic hybrid strategies may also function to model the manifestation of madness (by rationalist definitions), trickster behaviour, or elements of dreams.

IMMERSION AND ENGAGEMENT IN DYNAMIC SEQUENCING SYSTEMS

It was mentioned above that when the annotations within a video clip database fully (the property of *completeness*) and non-redundantly (the property of *minimality*) index the set of video clips (5), cognitive informational effects are maximised for minimum processing effort. An annotation space design that does not have the properties of completeness and minimality can create the more lyrical cognitive effects with increased processing effort (and user interaction) by which Tosca (29) characterizes hypertexts having links with a wide range of weak implicatures. This distinction can also be regarded from the viewpoint of immersion and engagement. Using schema theory, Douglas and Hargadon (34) define *immersion* as the experience of being completely absorbed within the ebb and flow of a familiar narrative schema. *Engagement* is defined as the effect of a work in overturning or conjoining conflicting schemas from a perspective outside the text, a perspective removed from any single schema.

The descriptors associated with video clips in the dynamic video sequencing systems described in this paper are instantiated descriptive schemas (strictly, the database schema is the description schema, and sets of tuples are instantiated descriptions). Association specifications, or narrative rules at various levels of decomposition, are also schemas that are instantiated during the operation of the association or narrative engines, respectively, as they dynamically create pathways through the space of video clips. In these terms, there may be grounds for supposing that immersion and/or engagement as defined by Douglas and Hargadon can be controlled to some degree by the design of schemas as annotations and sequencing rules. Indeed, the discussion presented by Lindley (5) suggests that a nonminimal and/or incomplete annotation space design creates engagement for categorical productions in the same way that Douglas and Hargadon characterize engagement with hypertext narrative. Similarly, a minimal and complete annotation space design maximizes the influence of user interactions on the nature of the unfolding presentation, creating an immersive coherence between the cognitive form of user interaction and the meanings generated in the resulting virtual video. The same principles apply to narrative generation, with the degree of immersion or engagement being determined by the extent to which the schemas expressed in action rules also express normatively anticipated causal connections.

However, the critical role of authorship in determining the relationship between symbolic structures within the system and the media objects that they are used to select and sequence must also be taken into account. Authorship applies to the symbolic structures within the system (content descriptions, interpretations, rules, and the like), to video clips, and to the associations of video clips with symbolic structures upon which dynamic sequence generation is later based. Hence the relationship between schema design and resulting virtual videos is underdetermined by the schema design alone. For example, however conventional the schemas might look, descriptive schemas could be linked to video clips in a highly idiosyncratic way (e.g. visualizing characters as abstract subfields of texture and colour), requiring intensive engagement from a viewer in order to try to understand what at the symbolic level looks like a very ordinary and predictable narrative.

CATEGORICAL/NARRATIVE STRATEGIES FOR INTERACTION IN VIRTUAL WORLDS

While hybrid sequencing strategies provide a rich suite of techniques for interactive video systems, they provide a much richer basis for generating presentations from more primitive recombinable media objects. Complete synthesis of image layers depicting characters and events can be accomplished by computer animation techniques based upon two-dimensional and three-dimensional models (including those derived originally from photographic and filmic material). Hence, ongoing research is addressing the adoption of narrative, categorical, and hybrid sequencing strategies for model-based presentation generation in virtual worlds.

Behaviour generation in virtual worlds can be regarded in terms of four levels of control, as depicted on Figure 15. These levels are identified at a level of authorship at which *a priori* classes of represented entity can manifest different behaviours within a range of possibilities constrained by the available set of behaviour generation representations. The arrows in the figure suggest the inheritance of behaviours from higher levels to lower levels of the model.

Each level is distinguished by a scale of encapsulation, and by a set of primitive components to which it may refer. Each level may be realized by the interpretation of a more or less detailed script or system of scripts, by deliberative planning to achieve high-level goals at that level, or by the execution of reaction rules in response to user actions or external state changes.
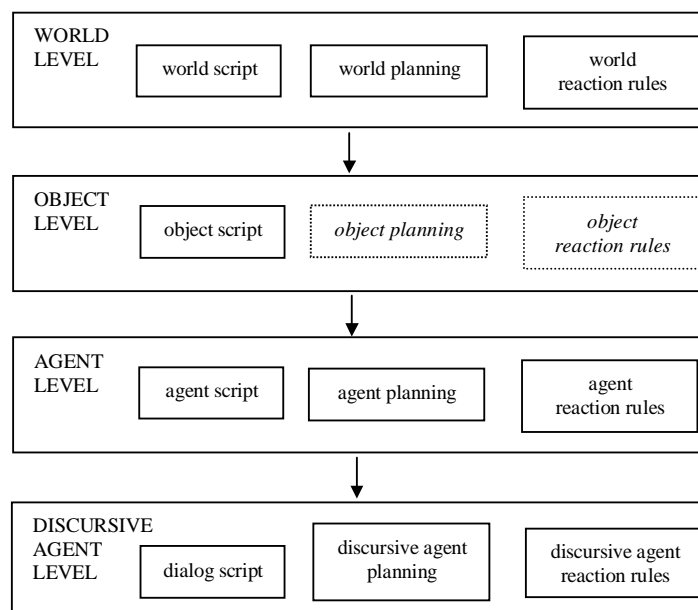

Figure 15. Levels of behavioural control representation for virtual worlds.

The distinction between planning approaches and reaction rule approaches is based upon the extent to which action generation relies upon the use of an inference engine to create behaviours, as opposed to a more direct mapping from virtual sensor descriptions to behaviours. A purely reactive system can only refer to internal state representations, simulated sensory inputs, and behavioural outputs at a given level of the hierarchy. Alternatively, a deliberative approach requires a search over a complex hierarchical or extended linear action plan and world model. Behaviours may also be scripted at higher levels, where the scripting language is closer to a natural language (e.g. 35). These techniques can be combined, for instance, by using a script interpreter to read a high level or natural language expression that is converted

into one or more high level goals that are then passed by the interpreter to a deliberative planner, or to instantiate the parameters of, or activate, a set of reaction rules.

The world level of the hierarchy represented on Figure 15 represents the definition of behaviour from a global perspective. World level control concerns high-level control of the virtual environment, including message transmission between and among objects and agents. Beneath the world level, more variability can be achieved by scripting object and agent behaviours. Objects are passive, with generic physical responses to outside perturbances (e.g. how to bounce when kicked). Agents are objects having autonomous behaviours. Autonomous behaviours are those that originate within the control architecture of the agent itself. Discursive agents are those having some behaviour that serves purely communicative or expressive functions. Discursive behaviours potentially range from reactive responses (e.g. using ELIZA-style language generation rules, or Expressivator-style gesture rules, 36) to the autonomous generation of stories (33) and the ability to use a conversational input as a goal for complex planning (37). More advanced discursive capabilities might include the ability of a virtual agent to interact with another agent or the user through a more complex discourse structure, such as an extended conversation or explanations for problem solving, commentary, post-mortems or tutoring (e.g. 38,39,40). This distinction between discursive and non-discursive agents separates explicitly modelled and controlled communicative behaviour from behaviours represented instrumentally that has only implicit communicative functions.

Different interactive media systems use different subsets of the components identified in Figure 15. This taxonomy can be used as a specification for behaviour generation facilities that could be made available within a comprehensive software architecture for interactive cinema. Scripting/planning/reaction-rule categories form one dimension and world/object/agent/discursive-agent categories form a second dimension. At any of the levels of control system representation, any of eleven sequencing methods may apply (i.e. narrative, categorical, or any of nine hybrid strategies), forming a third dimension. These distinctions along three dimensions form a total of one hundred and thirty two different behaviour generation strategies that could in principle be used independently or in any set of combinations (see Figure 16). Since any particular strategy can be either present or absent, the number of possible combinations of strategies is n = $2^{132} - 1$. Each combination of strategies defines a different potential authoring environment for interactive cinema, and hence a different interactive cinematic form.
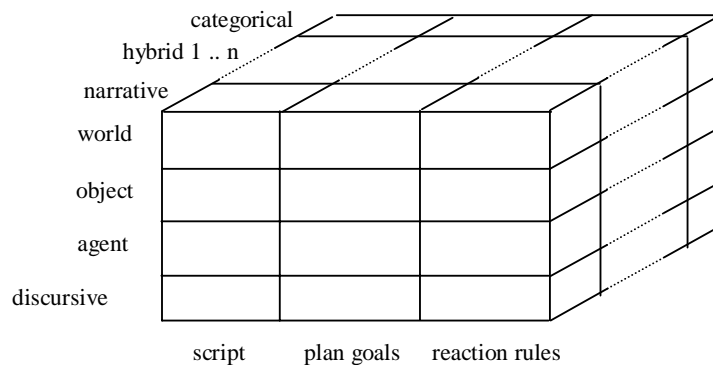

Figure 16. Design subspaces for behavioural control in virtual worlds.

The issue of what interesting interactions, expressions, and/or aesthetic functions might be created or served within the design subspaces depicted on Figure 16 is complicated by the introduction of additional dimensions of variation. For example, Mateas (41) identifies the degree of spatio-temporal *localisation* of control, and the degree to which a system is *generative* of new story structures and elements,. The distinctions within the resulting five dimensions of variation create a very large space of possible scripting

language subspaces for behaviour and story generation, and any particular production could in principle be designed in terms of any subset of this total set of subspaces.

CONCLUSION

This paper has presented a number of strategies for combining categorical video sequence generation with continuity-edited narrative sequence generation. The resulting techniques provide a suite of sequence generation methods that can be integrated into a rich video sequencing environment. These sequencing techniques could also be adapted to generate behaviours in a computer-based animation, or virtual world, system. Providing a rich variety of sequencing methods raises the question of how to author the more complex knowledge base used by the sequencing algorithms. This is essentially a scripting problem, viewing the knowledge base as a text (42) and the result of its processing as expressive AI (43). Addressing the development of suitable knowledge models as creative artefacts requires research at the level of both tools and methodologies to make the task easier for system authors.

REFERENCES

[1] NACK F. and PARKES A.: The application of video semantics and theme representation in automated video editing. *Multimedia Tools and Applications*, [Ed: Zhang, H.], (4,1), 1997, 57 - 83.

[2] NACK F.: *AUTEUR - The Application of Video Semantics and Theme Representation for Automated Film Editing*. Ph.D. Thesis, Lancaster University, UK, 1996.

[3] DAVENPORT G. and MURTAUGH M.: ConText: Towards the Evolving Documentary, Proceedings, ACM Multimedia, San Francisco, California, Nov. 5-11 1995.

[4] MURTAUGH M.: *The Automatist Storytelling System*, Masters Thesis, MIT Media Lab, 1996, http://ic.www.media.mit.edu/groups/ic/icPeople/murtaugh/thesis/index.html.

[5] LINDLEY C. A.: "A Video Annotation Methodology for Interactive Video Sequence Generation", BCS Computer Graphics & Displays Group Conference on Digital Content Creation, Bradford, UK, 12-13 April 2000.

[6] The composition of video presentations from more primitive media objects is explored by Suzuki et al (7), who use counterpoint theory as a basis for image synthesis. Although a scripting language is used to describe the counterpoint composition structure in a static and non-interactive form, similar principles based upon counterpoint theory could be applied to create a dynamic and interactive movie.

[7] SUZUKI R., IWADATE Y., and NAKATSU R.: "Multimedia Montage - Counterpoint Synthesis of Movies", *Multimedia Tools and Applications*, Kluwer Academic Publishers, 2000, 11, 311-331.

[8] STAM R., BURGOYNE R., and FLITTERMAN-LEWIS S.: *New Vocabularies in Film Semiotics: Structuralism, Post-Structuralism and Beyond*, Routledge, 1992.

[9] BRINGSJORD S. and FERRUCCI D. A.: *Artificial Intelligence and Literary Creativity: Inside the Mind of BRUTUS, a Storytelling Machine*, Lawrence Erlbaum Associates, Publishers, 2000.

[10] MATEAS M. and SENGERS P.: "Introduction to NI Symposium", AAAI 1999 Fall Symposium on Narrative Intelligence, http://www.cs.cmu.edu/~michaelm/narrative.html.

[11] BORDWELL D. and THOMPSON K.: *Film Art: An Introduction*, 5th edn., McGraw-Hill.1997.

[12] LINDLEY C. A.: "A Multiple-Interpretation Framework for Modeling Video Semantics", ER-97 Workshop on Conceptual Modeling in Multimedia Information Seeking, LA, 6-7 Nov., 1997.

[13] SRINIVASAN U., LINDLEY C., SIMPSON-YOUNG B.: "A Multi-model framework for Video Information Systems", "Semantic Issues in Multimedia Systems", *8th IFIP 2.6 Working Conference on Database Semantics (DS-8)*, Jan 5-8 1999, Rotorua, New Zealand.

[14] LANDOW G. P.: "The Rhetoric of Hypermedia: Some Rules for Authors", *Hypermedia and Literary Studies*, Delany P. and Landow G. P. Eds., The MIT Press, 81-104.,1991

[15] METZ, C.: "*Film Language: A Semiotic Of The Cinema*". New York: Oxford University Press, 1974

[16] GREGORY, J. R.: *Some Psychological Aspects of Motion Picture Montage*. Ph.D. Thesis, University of Illinois ,1961.

[17] WULFF, H. J.:. Der Plan macht's. In H. Beller (Eds.), *Handbuch der Filmmontage - Praxis und Prinzipien des Filmschnitts* (pp. 178 - 189). München: TR-Verlagsunion, 1993.

[18] PEIRCE, C. S.: *The Collected Papers of Charles Sanders Peirce - 1 Principles of Philosophy and 2 Elements of Logic, Edited by Charles Hartshorne and Paul Weiss*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1960.

[19] JAKOBSON, R., & HALLE, M.: *Fundamentals of Language*. The Hague: Mouton Publishers, 1980.

[20] BORDWELL, D.: *Making Meaning - Inference and Rhetoric in the Interpretation of Cinema*. Cambridge, Massachusetts: Harward University Press, 1989.

[21] ECO, U.: *Einführung in die Semiotik*. München: Wilhelm Fink Verlag, 1985.

[22] CHATMAN, S: *Story and Discourse: Narrative Structure in Fiction and Film*. New York: Ithaca. 1978.

[23] LEHNERT, W. G., DYER, M. G., JOHNSON, P. N., YANG, C. J., & HARLEY S.: BORIS - An Experiment in In-Depth Understanding of Narratives. *Artificial Intelligence*, 20, 15 - 62., 1983

[24] SCHANK, R. C.: *Dynamic memory*. New York: Cambridge University Press, 1982.

[25] SCHANK, R. C., & ABELSON, R.: *Scripts, Plans, Goals And Understanding*. Hillsdale, New Jersey: Lawrence Earlbaum Associates, 1977.

[26] WILENSKY, R.: *Planing and Understanding - A Computational Approach to Human Reasoning*. Reading, Massachusetts: Addison-Wesley Publishing Company, 1983.

[27] WILENSKY, R.: Points: A Theory of the Structure of Stories in Memory. In W. G. Lehnert & M. H. Ringle (Eds.), *Strategies for Natural Language Processing* (pp. 345 - 376). Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1983.

[28] PETRIC, V.: *Constructivism in Film*. Cambridge: Cambridge University Press, 1987.

[29] TOSCA S. P.: "A Pragmatics of Links", *Proceedings of the Eleventh ACM Conference on Hypertext and Hypermedia*, May 30 - June 4 2000, San Antonio, Texas, USA, 77-84.

[30] BERNSTEIN M: "Patterns of Hypertext", *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia*, June 20 - 24, Pittsburgh, PA, USA, 21-29,1998.

[31] Another strategy for increasing the combinatorial space of a production is to dynamically composite multiple layers of audiovisual data. This requires symbolic representations for the meanings of the separate layers, and both intra-media and cross-media rules for how new meanings are created by the synchronic and diachronic juxtaposition of the represented meanings of the atomic media components. The editing process may be simplified, and more combinatorial possibilities may be created, at the cost of a more complex process of description and composition rule design.

[32] DANCYGER and RUSH: *Alternative Scriptwriting: Writing Beyond the Rules*, Focal Press, 1995.

[33] BICKMORE T. and CASSELL J.: "Small Talk and Conversational Storytelling In Embodied Conversational Interface Agents", AAAI 1999 Fall Symposium on Narrative Intelligence, http://www.cs.cmu.edu/~michaelm/narrative.html.

[34] DOUGLAS Y. and HARGADON A.: "The Pleasure Principle: Immersion, Engagement, Flow", *Proceedings of the Eleventh ACM Conference on Hypertext and Hypermedia*, May 30 - June 4 2000, San Antonio, Texas, USA, 153-160.

[35] GOLDBERG A.: "IMPROV: A System for Real-Time Animation of Behavior-Based Interactive Synthetic Actors", in Trappl R. and Petta P. (Eds.), *Creating Personalities for Synthetic Actors*, Springer-Verlag Lecture Notes in Artificial Intelligence (LNAI) 1195, 1997.

[36] SENGERS P.: *Anti-Boxology: Agent Design in Cultural Context*, PhD Thesis, CMU Department of Computer Science and Program in Literary and Cultural Theory, August 1998.

[37] CAVAZZA M., BANDI S., and PALMER I.: ""Situated AI" in Video Games: Integrating NLP, Path Planning and 3D Animation", 1999 AAAI Spring Symposium on Artificial Intelligence and Computer Games, AAAI Technical Report SS-99-02.

[38] ANDRE E. and RIST T.: "Presenting Through Performing: On the Use of multiple Animated Characters in Knowledge-Based Presentation Systems", Proceedings of the Second International Conference on Intelligent User Interfaces (IUI 2000), pp. 1 - 8.

[39] LESTER J. C., TOWNS S. G., CALLAWAY C. B., VOERMAN J. L., and FITZGERALD P. J.: "Deictic and Emotive Communication in Animated Pedagogical Agents", in Cassell J., Sullivan J., Prevost S., and Churchill E. Eds. *Embodied Conversational Agents*, MIT Press, 123-154, 2000.

[40] CASSELL J., SULLIVAN J., PREVOST S., and CHURCHILL E. (Eds.): *Embodied Conversational Agents*, MIT Press, 2000.

[41] MATEAS M.: "An Oz-Centric Review of Interactive Drama and Believable Agents", CMU Technical Report, CMU-CS-97-156, 1997.

[42] LINDLEY C. A.: "A Postmodern Paradigm of Artificial Intelligence", *2nd World Conference on the Fundamentals of Artificial Intelligence*, Paris, 3-7 July, 1995.

[43] MATEAS M.: "Not your Grandmother's Game: AI-Based Art and Entertainment", 1999 AAAI Spring Symposium on Artificial Intelligence and Computer Games, AAAI Technical Report SS-99-02.