A Comprehension Based Cinematic Generator for Virtual Environments

Bradley A. Cassell
Liquid Narrative Group
North Carolina State University
Raleigh, NC 27695
bacassel@ncsu.edu

R. Michael Young
Liquid Narrative Group
North Carolina State University
Raleigh, NC 27695
young@csc.ncsu.edu

ABSTRACT
Most systems for generating cinematic shot sequences for virtual environments focus on the low-level problems of camera placement. While this approach will create a sequence of camera shots which film events in a virtual environment, it does not account for the high-level effects shot sequences have on viewer inferences. There have been some systems which are based on well known cinematography principles, however these usually utilize schemas or predefined shots which are known to be good choices. In this paper, we present a system which can reason directly about the high-level cognitive and narrative effects of a shot sequence on the viewer's mental state.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information Systems**]: Artificial, Augmented and Virtual Realities; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Plan execution, formation, and generation*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Discourse*

General Terms
Cinematic Generation, Narrative, Discourse, Planning

1. INTRODUCTION
Most work in cinematic generation reasons about low-level frame-by-frame placement of a camera in a virtual scene [1, 4, 5, 21]. However, in film, cinematographers either explicitly or intuitively build shot sequences to manipulate the mental state of the viewer [3]. Initial work on addressing this planning process for automated cinematic generation within 3D virtual environments was done by Jhala and Young [12]. Their work borrows an approach from Moore and Paris [17], Mann and Thompson [14] and others from the natural language generation community where hierarchical templates are used as a means to generate discourse structure.

The goal of our research is generating cinematic sequences for stories within virtual environments from a discourse perspective. Specifically, we want to generate shot sequences that can effectively communicate the mental states of a character when that character is deciding to change his or her course of action within the story. In particular, we focus on cases where characters' observations of their world and internal reflection cause them to change their plans. An important type of shot used in these situations is a shot which seems redundant. For example, shots which show events that have already occurred in previous shots (flashbacks), or shots which contain elements in the story world which have already been presented in previous shots.

In this paper we present arguments for the importance of explicit models of both story and discourse goals in the generation of effective narrative, and give a description of early work that employs these models in an automatic cinematography system that conveys intention dynamics of story characters. Our system is a plan-based discourse generator which has several important parts. First, a plan-based structure for representing the story and discourse is presented. Next, we discuss the computational models of narrative and cognitive psychology model of narrative comprehension that this research is based on. The cognitive model plays an important role in this system. Specifically we utilize a cognitive model of story comprehension during the planning process. Finally, we describe early work on designing a planning process that can reason directly about this model and plan data structure. In the following sections we discuss prior work in this area, the cognitive research our system extends, and the initial work on the system itself.

2. RELATED WORK
2.1 Camera Control in Virtual Environments
To date, there have been two primary types of automatic cinematic generation research: low-level control of the camera to place it and film objects correctly, and high-level reasoning about narrative pragmatics. Most work in automatic cinematic generation has focused on the low-level control of the camera. These approaches mainly use predetermined cinematic constraints as guides to shot sequence creation [4, 1]. Separate from this cinematic work, systems for textual generation have been developed which contain cognitive models that could be useful in cinematic generation based on viewer inferencing [14, 17, 22]. However, little work has been done on designing a system which generates shot sequences based on their pragmatic effect on viewer inferencing.

Workshop on Intelligent Cinematography and Editing, Foundations of Digital Games 2012 Raleigh, North Carolina USA

Christianson et. al.[4] describe a system which uses cinematography idioms – standard sequences of shots that communicate specific actions – to generate a cinematic sequence. The idioms are formalized in their declarative camera control language. The system then uses the idioms to construct shot sequences based on cinematography principles. However, it still relies on these idioms instead of reasoning about the desired effect on the viewer directly.

Christie and Normand [5] designed an interactive system which presents similar shots that are based on cinematic principles to the user. The user then selects the shots for the shot sequence. Virtual environment space is divided into partitions based on the contents of the space. These space partitions are tagged with respect to the cinematographic principles they maintain. The system then searches for best representations of the viewpoints of these spaces and presents them to the user.

El Nasr [6] created a system which not only designs camera shot sequences, but also designs the entire visual experience as a whole. The system has four parts: a character system, a lighting system, a camera system, and a director agent. The three subsystems, character, camera, and lighting, all propose various actions to fulfill a communicative goal and the director agent then unifies these into a consistent plan. This system basses its choices on the action that it is trying to present. The camera system is capable of suggesting several basic shot sequences such as close ups, long shots, pans and other types of shots well known from cinematography.

Bares et. al. [1] have developed a system which allows the user to specify particular camera shots by creating storyboards. The system then takes the storyboard and uses a constraint solver to determine where to place the camera. Another system by Bares [2] can balance a shot based on cinematography principles such as the rule of thirds using a constraint solver. The system is able to position the camera so that various scene elements are balanced in the shot. Various cinematography rules are able to be specified by the user. The system computes the possible shot choices based on an AND/OR tree and then presents them to the user. These examples, however, still leave the high-level shot composition up to the author of the cinematic.

Tomlinson et. al. [21] present a reactive cinematography system. It makes use of an intelligent camera agent that films characters based on how much the characters want to be filmed. Each character communicates how important it believes it is, and the camera agent uses this to help determine when to switch between the characters and what type of shot to use.

The Darshak system [12] uses hierarchical planning to form cinematics using abstract shots which are composed of sequences of more primitive shots. This system approaches cinematic generation from a high-level narrative pragmatics stand point. The system modifies hierarchical planning to use temporal variables. This modification then facilitates planning for temporally restricted events. Darshak's hierarchical plan actions are based on known narrative constructs. This allows for Darshak to build a hierarchical plan which follows a structure similar to narrative theory. By doing this, Darshak is able to generate shot sequences based on cinematic principles which communicate certain cinematic goals. However, it still relies on the pre-written decompositions of these abstract actions to dictate the narrative and cinematographic rules and principles. Darshak does not make significant use of an expressive knowledge representation to characterize a cinematic's effect on the mental state of a viewer.

2.2 Cognitive Model of Narrative

Our system makes use of a cognitive model of narrative comprehension to both model a story's structure and reason about how the information in the story will be stored in a user's memory. Many cognitive models have been developed to describe how people store narrative information and how people recall this information from their memory [7, 10, 13, 30]. The cognitive model we are using is the event-indexing situation model described by Zwann and Radvansky [30].

The *Event-Indexing Situation Model* (EISM) is a cognitive model that describes how people store and recall narrative information. With the EISM people are said to form internal situations as they read or experience a narrative. The model is updated as more information is presented to the reader. The events in the story are segmented around verb phrases in text and character actions in film [27]. The model and the events contained in it are associated and updated along 5 indices:

- *time index* - the time frame in which the event occurs
- *space index* - the space in which the event takes place
- *protagonist index* - whether or not the event involves the protagonist
- *causal index* - the event's causal status with regards to previous events
- *intention index* - the event's relatedness to the intentions of a character

Each event can be connected to other events within the EISM along one or more of these indices. As a person reads or experiences the narrative they update their situation model. This model is separated into an *integrated model* and a *current model.* The integrated model is the combination of all prior events up until the current event. The current model is the situation model for the current event. As the narrative is experienced the integrated model is updated with the information from the current model.

Zwaan and Radvanski [30] discuss how the EISM is stored in human memory based on the memory model of Ericsson and Kintsch [7]. This model of memory has two types: Short Term Working Memory (STWM) and Long Term Working Memory (LTWM). The current situation model from EISM is formed and analyzed in STWM and the integrated model is stored in LTWM. In addition, the STWM contains keys to all unique situation model indices in the developing story which are stored in LTWM. As a new event is experienced, a matching is done between that event's indices and the indices of the current foregrounded events, or events which are

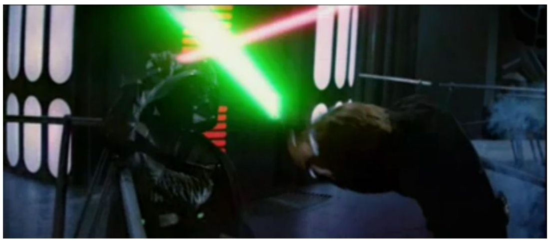

Figure 1: Confrontation Scene between Luke Skywalker and Darth Vader. [15]

Figure 2: Luke tossing away his Lightsaber after stopping his fight. [15]

most salient. If there is no match between the current model and foregrounded events' indices then a lookup is done to LTWM to see which indices in the current model have been encountered before. If this look up is successful, the current model is integrated into the LTWM and the events which have the shared indices become foregrounded, or more salient.

Once this integrated model has been created it can then be used to predict how salient a previous story action is relative to the current situation model. For example, consider a situation where a person experiences story events $e_1, e_2, ..., e_5$ and only e_1 and e_5 are connected with the causal index. Then, when event e_5 is experienced it foregrounds event e_1 because it is connected along the causal index. This would mean that event e_1 becomes more salient right after event e_5 is experienced.

The EISM discusses exactly how events share indices, or how to determine if two events are connected along an index.

- If two events occur in the same time frame they are said to share a *time* index. A time frame is defined by Zwaan [28]: two events share a time index if they are perceived in sequential order and neither event contains an explicit discontinuity in time.
- Two events share a *space* index if they occur in the same spatial region.
- If two events involve the story's protagonist they share a *protagonist* index.
- Two events share a *causation* index if they are related causally either directly or indirectly. A direct causal relation is when an event establishes a condition that is required for the next event. An indirect causal relation between two events e_i to e_n exists if there is a path in the transitive closure of the causal relation from e_i to e_n.
- Two events share an *intention* index if they are part of the same plan of a character to achieve a goal. Goal structures are derived from General Knowledge Structures as identified by Graesser and Clark [10].

In our current work we have developed an operationalization of the EISM as a plan structure. This operationalization forms the foundation of the rest of the system we are working on to generate cinematics from a narrative standpoint.

2.3 Information Redundancy

Walker [22] developed a system which is capable of generating *Informationally Redundant Utterances*, or IRUs. IRUs are defined as utterances which contain no new information but serve specific linguistic purposes. Walker explained that IRUs can:

- *Attitude*: provide evidence to support beliefs about mutual understanding and acceptance.
- *Attention*: manipulate the attention listeners by foregrounding a proposition.
- *Consequence*: augment the evidence of beliefs that support inferences.

Of particular interest to us is the use of IRUs to foreground propositions. In Walker's work, she uses the idea of STWM and LTWM to model a discourse a participant's storage of a discourse. She modifies work on a working memory model by Landauer [13]. In this model, working memory can be thought of as a 3D space, where ideas are located next to each other and the more you dig down, by moving from one idea to the other, the farther away you get from other ideas. Walker augmented this model by adding the functionality to copy an encountered idea to the top, or foreground, of the working memory. Walker uses IRUs for attention purposes to recall ideas into working memory from elsewhere in the model, thus making them more salient. Literature and film employ IRUs extensively The next section will show two prominent examples.

3. CINEMATIC GOALS

Human cinematographers plan shot sequences either explicitly or intuitively to manipulate the mental state of the viewer [3]. One specific case of this is when shots are used to explain the internal mental state of a character. Often, story events are dependent on characters making decisions on a course of action. With written discourse, a writer is able to describe an internal monologue of a character to explain their line of reasoning. In film, however, this more often is accomplished with carefully constructed shot sequences that include informationally redundant shots. When characters deliberate they can be thinking of past events or objects which have meaning. Often, these previous actions or world objects may not be salient in the viewer's mind, thus if they are not explained, a character's decision on how to proceed may not be clear. Consider an example from Star Wars: Return of the Jedi [15], when Luke Skywalker is confronting Darth Vader on the Death Star.

Figure 3: Luke looking at his hand. [15]

Figure 4: Close up of Luke looking at his hand. [15]

Figure 5: Shot of Vader's arm showing its electronics. [15]

Figure 6: Close up of Luke's face. [15]

The shots captured in figure 1 and figure 2 first show Luke fighting Vader and then throwing away his lightsaber after he cuts off Vader's hand. By just looking at these two shots we know that Luke was fighting Vader and for some reason decided to stop and throw away his lightsaber. Without any further explanation we don't know why Luke decided to stop.

Now if we look at the shots left out, shots 3, 4, 5, and 6, we are able to gain a bit more insight. These show Luke looking at his hand after cutting off Vader's hand, then a close up of Luke looking at his hand, then a shot of Vader's arm showing its electronics, then a close up shot of Luke's face. None of these shots introduce any new information. We already knew Vader's hand was cut off, we already knew Luke had a hand. The key here is that the shot of Luke's hand along with the shot of Vader's mechanical arm linked to an earlier shot of Luke's hand being mechanical (from earlier in the movie.) These are redundant shots in that they do not add any information, but help the viewers to recall an earlier fact about Luke's hand. Once Luke cuts off Vader's hand, he sees that it is mechanical and then thinks about his hand being mechanical and realizes he is beginning to become like Vader. We can then conclude that the reason Luke turned off his lightsaber and stopped fighting was because he did not want to continue along that plan.

Note that this example from Star Wars serves to explain the negative effects of Luke's current course of action. Other examples, such as one from the Movie The Twilight Samurai [23], high light the benefits of a new plan. In this example the main character is determined to not get re-married. However one of his childhood friends comes to visit and she shows herself to be very good with the main character's children. A particular scene later uses a redundant shot of her singing with the children (showing that she is good with them) while filming the main characters reaction. It is clear from this shot sequence that he is reconsidering his decision to not re-marry. This is an example of shots that show how a new plan may benefit the main character more than the current plan. These different ways of foregrounding play a role in how to decide which events and objects to foreground to explain the character's mental state.

4. SYSTEM DESCRIPTION

The narrative processes and techniques used to design shot sequences are similar to those used to construct written discourse [3]. The resulting cinematics share much of the same structure as written narrative [3]. One way for cinematic generators to achieve a higher level of communicative capability is for them to reason about the cognitive and narrative effects on the viewers. Researchers [9, 29] have shown that cognitive models can accurately describe how people comprehend and process written narrative constructs. Our system will focus on generating cinematics which can non-verbally communicate the internal mental state of a character. Specifically, we use models of written narrative understanding to depict that a character's mental state is changing (in that they are choosing to change their plan of action). For these cinematics to be successful, they should include camera shots which seem redundant, or camera shots that don't directly add information about the environment. For a system which focuses solely on low-level constraint solving for placement, it is difficult to generate shot sequences which contain shots that don't present new or important information. Our system for shot generation will reason explicitly about belief and intention dynamics in order to effectively communicate these changes in mental state. It will incorporate narrative theories and cognitive models of narrative comprehension with a hierarchical discourse planner. Our system uses the following three components to achieve this:

1. Representations of story and discourse as plan structures
2. A model of the viewer's mental state

3. A planner that uses these structures to reason about the viewer's mental state model

Each of these parts modifies or adds to planning data structures or processes. First, the plan structure for the story needs to encode more information than traditional narrative plan structures if it is going to be used for reasoning about narrative comprehension. We have developed an operationalization of the EISM from cognitive research which was presented earlier for this reason. This operationalization, extends a previous planning data structure used in intentional partial order causal link planning, or IPOCL planning. Next, the viewer's mental state needs to be modeled accurately. This, also relies on the EISM model. The planner will design the discourse based on how it affects the viewer's mental state. A representation of the viewer's working memory is modeled in a data structure based on EISM's. Finally, the planner will be modified to reason directly about the mental state representation of the viewer. The planner will choose shots based on how they modify the viewer's mental state model to foreground specific events and objects so that the viewer will understand the character's mental state when making a decision.

4.1 Story Plans

Plan representations have been shown to be useful for representing narrative structure [12, 19, 25]. In current work we have developed an operationalization of the EISM as a plan structure. This model builds off of Intentional Partial Order Causal Link (IPOCL) plans [19]. IPOCL plans are data structures which can be used to represent stories with causal, temporal, and intentional relationships [19].

A *plan* is a sequence of steps that describes how a world transitions from its initial state, to its goal state [18]. A *state* is a single literal or a conjunction of literals specifying what is true and false in a story world. The *initial state* describes the world at the beginning of a plan. The *goal state* contains which literals must be true at the end. The planning *problem* consists of the the initial and goal states together. A particular IPOCL plan is a solution which transitions from the initial state to the goal state. Characters, items, and places in the story are represented as constants. The actions between the initial and goal states make up the contents of the plan. *Operators* are templates for *actions* which can occur. They are three-tuples $\langle P, E, A \rangle$ where P is a set of *preconditions*, E is a set of *effects*, and A is a set of characters. Preconditions are literals which must be true before the action can be executed, effects are literals which are made true by the execution of the action [8], and characters in A are those which must consent to the execution of that action [19]. Operators are able to contain variables which must be filled with constants during the planning process. The *planning domain* is the set of all available operators . A *step* is an actual instance of an operator which will take place in the story world. In a step, all of the variables in the step's operator have been bound to constants.

Plan steps are partially ordered with respect to time [20]. $s < u$ denotes an *ordering* over steps s and u, where s must be executed before u. A plan must guarantee that all of the steps' preconditions are true before they are executed [16]. A precondition can either be made true by an effect in an earlier step, or be true originally from the initial state. A *causal link*, $s \xrightarrow{p} u$, with step s with effect p and step u with precondition p, explains how a precondition of a step is met (i.e. s makes p true for u). Step u's *causal parents* are all steps s such that there exists a causal link $s \xrightarrow{p} u$.

Frames of commitments are structures included in IPOCL plans to explain a character's actions in terms of individual goals [19]. An *intention* is of the form $intends(a, g_a)$; g_a is a literal that actor a wishes to be true. A *motivating step* causes an actor to adopt a goal and has an intention as one of its effects. A *final step* is a step which achieves some actor goal and has g_a as one of its effects. All steps in the frame of commitment must be causal ancestors of the final step ordered before the final step. A frame of commitment describes the steps an actor takes to achieve some goal, and the step which finally achieves the goal.

The plan produced by a planner is a five-tuple $\langle S, B, O, L, I \rangle$ where S is a set of steps, B a set of variable bindings, O a set of orderings, L a set of causal links, and I a set of frames of commitment. A *complete plan* is guaranteed to achieve the goal from the initial state.

IPOCL plans can be produced by planning algorithms [19]. They model important information about stories which can be modified to operationalize the EISM to predict how well humans remember certain steps.

4.1.1 EISM Operationalization

Our operationalization of the EISM extends the IPOCL plan representation by adding information relavent to the EISM. The IPOCL plan representation already includes many of the features needed to represent EISM structures, and the enhancements we outline are straightforward to introduce.

Events in the EISM framework to steps in an IPOCL plan, assuming that the events center around verbs (in text) or actions (in film). The term *event* and the term *step* are used interchangeably here. Recall that the IPOCL model represents elements of the story, or events, characters, locations and the other entities within the story world.

To extend IPOCL with the EISM knowledge representation, the IPOCL plan is forced to be a fully ordered plan. The modifications to the IPOCL data structure are:

- Time

  Time is represented in the current IPOCL model by a partial ordering of steps. Steps are modeled as executing instantaneously and IPOCL's temporal representation provides a partial ordering over all of a plan's steps' times of occurrence. Our operationalization requires that each operator in an IPOCL planning domain contain a distinguished variable called the *time frame*. In any IPOCL plan, each step's time frame variable must be bound to one of a list of constants that refer to time frames in specified in the domain. These constants can be defined either by the domain creator or automatically.

- Space

  To model the location where an event occurs, out operationalization requires that each operator in an IPOCL plan domain contain a variable labeled *location*. In any IPOCL plan, each step's location variable must be bound to one of a list of constants that refer to locations in the domain. These constants can be defined either by the domain creator or automatically.

- Protagonist

  In the EISM model, the protagonist is single character. To model the protagonist, our operationalization requires that an initial state description contain an entry designating a single character as the protagonist. For any given step in a plan, the protagonist index captures whether or not events involve a story's protagonist. Each step designates a set, A, of consenting characters, thus the IPOCL plan representation already contains elements that can be used to characterize the protagonist index of each step.

- Causation

  The causation index captures whether or not events have a causal relation. The IPOCL plan's causal links can be used to represent both enablement and motivational causal relations. In an IPOCL plan, a causal link's originating step s_i is a necessary but not sufficient condition for the subsequent step s_j through $s_i \xrightarrow{p} s_j$ which represents enablement. A causal link $s_i \xrightarrow{p} s_j$, where s_i is an IPOCL motivating step represents a motivation causal relation.

- Intention

  The intention index indicates that an event plays a role in a character's plan to achieve a goal. The IPOCL plan representation's frames of commitment can already accurately represent the intention index.

These modifications of IPOCL plans to contain EISM information requires that the data structure be a fully ordered plan. However for our use in our discourse planner, we only require that the story planner have the spacial, protagonist, causal, and intentional information completed in the plans, and instead rely on the partial order only. This allows our discourse planner to manipulate the time information directly to form the discourse.

4.1.2 Deliberation Steps

In addition to modifying IPOCL plans, we extend the library of operators that are used to create plans with *deliberation* operations. These operations are used specifically to change a character's plan. They are able to be used once certain preconditions are met just like normal actions, however their effects provide a means to achieve the new goal for the character. In our work we do not currently specify a method for planning with these operations, only that the plan include them at the point of deliberation of a character about a course of action.

4.2 Working Memory

Our system needs to model the viewer's mental state at every point in time during the discourse plan. This will allow the discourse planner to form shot sequences which intentionally manipulate the viewer's mental state model to communicate narrative goals. This viewer mental state model is also constructed using the EISM. A user's current situation model will be structured as an array which holds the currently foregrounded literals added by actions in the discourse plan. These literals will be of the form (bel <literal>), where <literal> is an effect of the story action being filmed, similar to Darshak. The user's integrated situation model will also be an array which holds all literals added to the user's situation model. These literals will be linked to literals in the current model depending on which EISM indices they share. For example, if a literal added to the situation model early in the story is causally connected to a literal in the current situation model, the two literals will be linked through the casual index. Thus, by accessing the literal in the current model, the literal in the integrated model will be able to be retrieved.

4.3 Narrative Comprehension Based Cinematic Planning

The base algorithm to form the sequence of shots will be a decompositional partial order causal link planner that makes use of the hierarchical structures Darshak uses. The end state of the planning problem is to have the viewer's situation model of the story, described in the previous section, in the desired configuration at the end of the shot sequence. Each action in the plan is a shot which modifies the viewer's situation model by adding literals to the current situation model array. These shots either directly add the literals or add them by retrieving them from the integrated model array. When a literal is added to the current situation model array, it is connected to literals in the integrated model depending on its connection in the EISM modified story plan. This allows for the determination of the ease of recall of literals in the integrated model which are connected to the literals in the current model. Each shot action will add or remove support for ideas held within the situation model and move it closer to the goal situation model. The end state of the plan is reached when the situation model contains specific literals which are the communicative goal of the cinematic. The planner will be a forward planning algorithm, where it starts at the beginning state and works toward the goal state. As shot actions get added to the plan, literals in the current situation model would decay out and new literals from the shot action would get added. When literals decay out of the current array they get added to the integrated array. The planner needs to add seemingly redundant shots to foreground them again based on their strength in the situation model. These redundant shots would then affect the viewer's inferences about what story objects and events the character is currently thinking about.

As the planner is adding shots based on steps in the story plan it may need to add redundant shots depending on the steps it encounters. It will choose high level Darshak shots which are currently available based on the previous shot action's effects and decompose them as necessary. However, for each deliberation step in the EISM modified story plan it will need to make sure that the literals in the preconditions of the deliberation shot are foregrounded appropriately. The planner does this by first searching the current situation model array for these literals in the form (bel <literal>). If they are not in the current situation model array, it then uses the connections to the integrated situation model array to find the correct literals to foreground back into the current situation model array. The planner is able to decided which literals to foreground two ways: either negatively or positively. The differences between these two were presented in the examples from Star Wars and The Twilight Samurai. If the planner negatively foregrounds the literals it chooses shot which foreground literals that negate the goals of the characters plan before the deliberation step. If the planner foregrounds literals positively, it chooses shots which foreground literals which support the goals of the character's new plan (i.e. shots which highlight the goal state of the story plan which involve the character in the deliberation step.) Once the shots have been chosen which foreground the appropriate literals to the viewer's current situation model, the planner then chooses a shot which depicts the character deliberating (a close up for example.) It then chooses the next step in the story plan to film until all events in the story plan have been filmed.

5. FUTURE WORK

Several problems still remain to be solved with this approach. One is deciding how to use the event indicies to effect focus decay and ease of foregrounding for ideas. For instance, flashbacks may be able to be generated for certain foregrounding needs. Preventing knowledge is another challenge. Some cinematics, specifically those of the mystery genre, may require withholding certain information. An example would be showing a murder scene, but intentionally hiding the murder's face from all shots. This would require restricting working memory so that it does not contain this information. Work on this system is still in its early stages. The system itself needs to be more fully specified. Once, this is accomplished the system will be created and implemented in a 3D environment utilizing the Zócalo system [26]. The planner will create the discourse plan for the story based off of the story plan, and then send both plans to the 3D environment. Here Zócalo will execute the plans in tandem within the 3D environment to create the cinematic.

To test this system, several experiments are planned. First we would like to determine how much of the effect on the viewer's mental state is dependent on the quality of the characters and objects filmed. To this end we plan an experiment which will compare user experiences while watching a cinematic with characters that have a high fidelity of facial expressiveness, to the exact same cinematics but with characters that have low fidelity of facial expressiveness. Once this is determined, we want to run experiements which test a viewer's comprehension of a character's mental state while that character is making decisions in the story. These experiments will use the QUEST [11] experimentation model to determine a viewer's comprehension of the story.

6. CONCLUSION

Current cinematic generation systems are able to produce cinematics, but they lack the ability to reason about the effects the cinematics have on a viewer's mental state. We have presented support for the use of cognitive and natural language techniques for cinematic generation. A cinematic generator which can reason about these elements of narrative understanding will be able to generate custom cinematics which achieve narrative communicative goals. Specifically, these shot sequences will be capable of communicating the goal of explaining a character's mental state during choice deliberation. In addition, we presented a description of the early stages of a system to do this. Future work will strive to complete this systems and test it's ability to correctly design shot sequences to describe a character's mental state while making decisions.

7. REFERENCES

[1] W. Bares, S. McDermott, C. Boudreaux, and S. Thainimit. Virtual 3D Camera Composition from Frame Constraints. *Proceedings of the Eighth ACM International Conference on Multimedia - MULTIMEDIA '00*, pages 177–186, 2000.

[2] W. H. Bares. A Photographic Composition Assistant for Intelligent Virtual 3D Camera Systems. In *In Proceedings of the Sixth International Symposium on Smart Graphics*, pages 172–183, Vancouver, B.C., Canada, 2006. In Proceedings of the Sixth International Symposium on Smart Graphics.

[3] E. Branigan. *Narrative Comprehension and Film.* Routledge, 1992.

[4] D. Christianson, S. Anderson, L. He, D. Salesin, D. Weld, and M. Cohen. Declarative Camera Control for Automatic Cinematography. In *Proceedings of the National Conference on Artificial Intelligence*, pages 148–155, 1996.

[5] M. Christie and J.-M. Normand. A Semantic Space Partitioning Approach to Virtual Camera Composition. *Computer Graphics Forum*, 24(3):247–256, Sept. 2005.

[6] M. El-Nasr. An interactive narrative architecture based on filmmaking theory. *International Journal on Intelligent Games and Simulation*, 3(1), 2004.

[7] K. A. Ericsson and W. Kintsch. Long-term working memory. *Psychological Review*, 102:211–245, 1995.

[8] R. Fikes and N. Nilsson. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial intelligence*, 2(3-4):189–208, 1971.

[9] A. Graesser, M. Singer, and T. Trabasso. Constructing Inferences During Narrative Text Comprehension. *Psychological Review*, 101(3):371–395, 1994.

[10] A. C. Graesser and L. F. Clark. *Structures and Procedures of Implicit Knowledge: Advances in Discourse Processes.* Ablex Publishing, 1985.

[11] A. C. Graesser and S. P. Franklin. QUEST: A cognitive model of question answering. *Discourse Processes*, 13(3):279–303, 1990.

[12] A. Jhala and R. M. Young. Cinematic Visual Discourse: Representation, Generation, and Evaluation. *IEEE Transactions on Computational Intelligence and AI in Games*, 2(2):69–81, June 2010.

[13] T. Landauer. Memory Without Organization: Properties of a Model With Random Storage and Undirected Retrieval. *Cognitive Psychology*, 7(4):495–531, 1975.

[14] W. C. Mann and S. A. Thompson. Rhetorical

Structure Theory: Toward a Functional Theory of Text Organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281, 1988.

[15] R. Marquand. Star wars: Return of the jedi. *20th Century Fox*, 1983.

[16] D. McAllester and D. Rosenblitt. Systematic Nonlinear Planning. In *Proc. of AAAI*, pages 634–639. Massachusetts Institute of Technology Press, 1991.

[17] J. Moore and C. Paris. Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information. *Computational Linguistics*, 19(4):651–694, 1993.

[18] A. Newell and H. A. Simon. GPS, a program that simulates human thought. *Computers and thought*, pages 279–293, 1961.

[19] M. O. Riedl and R. M. Young. Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research*, pages 164–7, 2010.

[20] E. D. Sacerdoti. The nonlinear nature of plans. In *Proceedings of the 4th International Joint Conference on Artificial Intelligence*, pages 206–214, 1975.

[21] B. Tomlinson, B. Blumberg, and D. Nain. Expressive Autonomous Cinematography for Interactive Virtual Environments. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 317–324. ACM, 2000.

[22] M. A. Walker. Informational Redundancy and Resource Bounds in Dialogue. *Cognitive Science*, 1993.

[23] Y. Yamada. The twilight samurai. *Empire Pictures*, 2002.

[24] R. M. Young. Story and discourse: A bipartite model of narrative generation in virtual worlds. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, 8:177–208, 2007.

[25] R. M. Young and J. D. Moore. DPOCL: A Principled Approach to Discourse Planning. In *Proceedings of the Seventh International Workshop on Natural Language Generation*, pages 13–20. Association for Computational Linguistics, 1994.

[26] R. M. Young, J. Thomas, C. Bevan, and B. A. Cassell. Zócalo: A service-oriented architecture facilitating sharing of computational resources in interactive narrative research. In *In The Working Notes of the Workshop on Sharing Interactive Digital Storytelling Technologies at The Fourth International Conference on Interactive Digital Storytelling*, 2011.

[27] J. M. Zacks, N. K. Speer, and J. R. Reynolds. Segmentation in Reading and Film Comprehension. *Journal of Experimental Psychology: General*, 138(2):307–27, 2009.

[28] R. A. Zwaan. Processing Narrative Time Shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5):1196–1207, 1996.

[29] R. A. Zwaan, J. P. Magliano, and A. C. Graesser. Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2):386, 1995.

[30] R. A. Zwaan and G. A. Radvansky. Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2):162–85, 1998.