An Event-Driven, Stochastic, Undirected Narrative (EDSUN) Framework for Interactive Contents

Adam Barclay and Hannes Kaufmann

Institute of Software Technology and Interactive Systems, Vienna University of Technology, Favoritenstr. 9-11/188/2,
A-1040 Vienna, Austria
kaufmann@ims.tuwien.ac.at, adam.br@gmx.at

Abstract. In this paper, we present an extensible framework for interactive multimodal contents, with emphasis on augmented reality applications. The proposed framework, EDSUN, enables concurrent and variable narrative structures as well as content reusability and dynamic yet natural experience generation. EDSUN's main components include a canonical specification of 5-state lexical syntax and grammar, stochastic state transitions, and extensions for hierarchical grammars to represent complex behavioral and multimodal interactions. The benefits of EDSUN in enabling classical contents to support the affordances of AR environments and in complementing recent published works are also discussed.

Keywords: augmented reality, event driven, stochastic transitions, undirected narrative, content, framework, grammar, story telling.

1 Introduction

Interactive environments pose several challenges for multimodal contents. One challenge is the lack of pre-existing rich interactive contents to draw upon, whereas new content creation is expensive and time consuming [5]. Despite the availability of rich classical contents, as found in film and drama, their suitability in matching the corresponding affordances of interactive environments is another challenge. Early storytelling systems and history-centric applications, for instance, often put the user in a passive audience role [6],[7]. This is due to the monolithic non-event driven nature of classical contents. Recent interactive story telling environments [10],[11] rely on hierarchical task networks to support emergent narrative and balance user's needs of both interactivity and realism. A third challenge is multimodal content reusability, where no common format exists to enable cross application content reuse. Furthermore, the interplay between the content author's objectives, choice of narrative, and the user's tendencies to drift off the main path envisioned by the content author, form the basis for the opposing attributes of interactive environment affordances.

In this paper, an extensible framework for interactive multimodal contents is presented, where the mentioned challenges are addressed. In the course of our treatment of the challenges in interactive multimodal contents, the focus is on

S. Göbel, R. Malkewitz, and I. Iurgel (Eds.): TIDSE 2006, LNCS 4326, pp. 13 – 24, 2006.

augmented reality environments, due to the additional constrains present. Table 1 illustrates the relevant affordances in augmented reality environments and EDSUN's approach in supporting them.

Table 1. A taxonomy of relevant AR affordances and the proposed enabling techniques

AR Environment Attributes	Supported Technique	Section
Unconstrained user interactions	Infinitesimal undirected narrative elements.	2
Real-time event driven environment	5-state lexical syntax, canonical form for content presentation.	2.1
Task oriented within an undirected narrative	State coupling through stochastic state transitions.	2.2
New experience generation, complex interaction representation, specialized narrative support	Hierarchical grammars	2.3
Rich content and multiplicity support.	Structured and unstructured variations	2.4

The first component of EDSUN, as detailed in Section 2.1, is an undirected narrative-based canonical form for content representation with a 5-state lexical syntax and grammar. The proposed canonical form enforces the separation between narrative structure and narrative content, and enables content classification and segmentation to support event-driven environments, as well as cross-application reusability. The lexical syntax and grammar model the set of possible user experiences within the AR environment, and script actual user experiences that result from his interactions. The second component of EDSUN, Section 2.2, forms stochastic state transitions that provide the content author with flexibility in incorporating variety of user experiences ranging from exploratory style navigation to task-oriented step-by-step interactions. The third component, hierarchical grammars in Section 2.3, is an extensibility mechanism to support author-specific interaction methodologies such as support for the Branigan cinematic narrative model [1] as implemented in DINAH[4]. Section 2.4 presents structured and unstructured variations, which enable support for variable tone experiences. An example is then presented to illustrate EDSUN's support for variety of experiences from a single source script, and we conclude with a summary of what has been accomplished.

1.2 Related Work

Traditionally, interactive media has adopted a classical narrative approach to content presentation due to the pre-programmed nature of contents within a given environment. The interactive media environment is then viewed as the narrator, and the user is encouraged to participate through dramatic enactment as an avatar of himself or others in the environment as in GEIST[6] or in MacIntyre *et al* [5]. The contents are usually borrowed from film, books, and drama along with their linear and non-linear narrative formats. In this section, several relevant published works, in the area of interactive narrative are reviewed and examined including: Mad Tea-Party[5], GEIST[6], Three Angry Men[7], DINHA[4], CrossTalk[10], Goldfinger[12], and Façade[8],[9].

MacIntyre *et al* [5] applies new media theory and media remediation techniques to propose solutions for AR-based interactive narrative, and in the process, explores the creation of unconstrained story telling systems in Mad Tea-Party. One of the key problems facing interactivity in the AR world as pointed out by MacIntyre *et al* [5] is the need for an unconstrained narrative to bring the story-world to reality. The paper asserts that an unconstrained narrative defines a contradiction: if the user is unconstrained, there is no guarantee that he'll experience the narrative or reach the pre-scripted ending - as a result of not being in the right location at the right time. Consequently, the paper proposes using the affordances of the objects in the virtual and physical world as well as symmetric activities of the interactive objects involved to encourage the user to stay within the limitations of the predefined narrative. The techniques used are based on conventions borrowed from both film and drama – in defining the interaction and in directing the participant. To facilitate subplot branching, the author breaks user interaction up into primitives or basic building blocks that the story can respond to. Procedural nodes along with corresponding procedural rules that match pre-scriptesd subplots are utilized to enhance the variety of user's experiences.

GEIST [6] is a history-based digital story telling system, where the non-linear interactive narrative encourages the user to influence the flow of the story. The digital story telling system in GEIST educates the user about historical aspects of the locations the user encounters. Although the user is allowed to explore the AR environment, interacting with the virtual world is pre-scripted and directed where a story is told based on the participant's proximity to a location of a historical relevance. The act of story telling itself is immotive as the participant assumes a passive audience's role.

Three Angry Men, TAM [7], explores interactive content presentation of multiple points of view. The user can experience one of three jurors' point of view by sitting in a given juror's seat, and listening to the other two jurors "deliver their dialogue". The dialogue stops when the participant decides to switch seats in order to get the point of view of other jurors. Similar to GEIST, dialogue delivery expects the user to assume a passive audience role, which could be envisioned as suitable for a court setting.

Digital storytelling with DINAH[4] addresses the challenge of multiple subplot generation through digital narrative composition. In DINAH, Ventura *et al* [4] uses a metalinear narrative approach, and applies a Branigan cinematic narrative model [1] to auto-compose stories from a database of story clips. The narrative engine in DINAH is constrained through preconditions and post-conditions that obey the Branigan model embodied within each story clip, as well as narrative state vectors that determine the user's progress towards one of the pre-scripted endings. Despite the similarity to Mad Tea-Party [5] in breaking up the story up into primitives or basic building blocks, content segmentation is narrative-specific and subplot-specific. In other words, specific to the story or application with no support for content reusability or new experience generation.

CrossTalk[10] is a virtual character exhibition for public spaces, where interaction between the virtual characters and the user is based on a combination of pre-scripted scenes as well as an hierarchical task network (HTN) plan-based dialog generation. Similarly, Goldfinger [12], which is a multimodal mixed reality role-playing interactive environment, relies on HTN to handle the frequent yet unexpected user interactions by requesting a new action plan to achieve the current task, which in turn, leads the story into a new direction based on user's actions. Both of these examples strive to create a balance between task-oriented environments and free-play through HTN replanning. Despite the flexibility of HTN, it is not possible to implement in each character all possible responses for the permutations of possible interactions [11].

Façade [9] is an interactive first person drama that relies on dialogue generation to maintain the desired interactive narrative. Dialogue generation in Façade is text based, and defines a many-to-few mapping of surface text to discourse acts [8]. Façade's natural language processing (NLP) mechanism defines a set of 24 discourse acts, against which user input is mapped. The response of Façade's characters is the product of an HTN plan that defines a particular narrative path, and the NLP environment that determines the output text.

Although the mentioned works reflect the rich possibilities within interactive multimodal contents, common challenges as has been shown, become apparent, namely: classical contents are (a) application/story specific, (b) monolithic, (c) with limited number of subplots, and (d) support for one type of narrative. As shown in [9], [11],[10], dynamic dialog and scene generation are becoming popular in addressing the affordances of interactive environments, and in enabling a rich user experience. In this paper, a framework is proposed to compliment the mentioned works by defining infinitesimal undirected narrative elements, along with associated lexical syntax, and stochastic state transitions to enable the embodiment of interactive environment affordances and reusability into classical contents.

2 A Framework for Event-Driven AR Content

In formulating a framework for interactive multimodal contents in AR environments, the affordances of AR environments need to be considered, and be embodied within the attributes of AR geared content. As shown in table 1, AR affordances include real-time interactivity, event driven environment updates, and synchronized multi-modality for audio and visual content, as well as geophysical location and orientation.

The challenge in simultaneously satisfying the mentioned affordances for a rich and immersive user experience lies in developing a balance between contradictory attributes. For example, the proposed approach of infinitesimal undirected narrative elements to enable unconstrained user interactions, may not help the user accomplish any given task [5], and may result in plot discontinuity. Similarly, event driven environments are, by definition, a hindrance to narrative continuity. Legacy contents, such as film and classical drama that are often the focus of story telling systems and history-based tourist guide applications, do not lend themselves easily to real-time event driven environments.

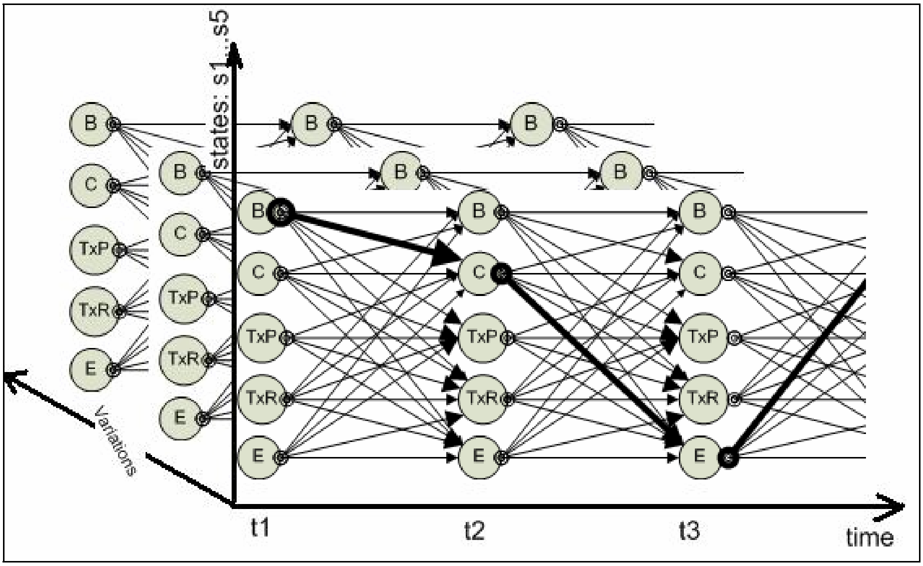


Fig. 1. An illustrative diagram of the universe of infinitesimal undirected narrative elements, mapped to the 5-state canonical form. The user's experience is highlighted by blue arrows, representing the path of stochastic state transitions.

If the present tense could be viewed as an engine that transforms unstructured future events into a structured past tense, an AR environment could be viewed as an engine that transforms a subset of the universe of infinitesimal undirected narrative elements that represent future events into a past represented by a linear narrative. With this in mind, we model the possible set of interactions in the AR environment with a set of infinitesimal undirected narrative elements S, a subset of which, s, forms the user experience and becomes his past, scripted by his own interactions within the AR environment, and is represented by a linear narrative. Therefore, our choice for basing EDSUN on infinitesimal undirected narrative elements lies in its suitability for event driven environments, and flexibility in modeling linear and non-linear narratives, as well as random real-world events.

Despite its flexibility, basing EDSUN on infinitesimal undirected narrative elements poses two challenges. The first is in formulating a methodology for classifying and segmenting contents into basic elements that could be mapped onto the desired set of infinitesimal undirected narrative elements. The second lies in task-oriented applications, where an undirected narrative approach may not guarantee that the user experiences what the content author has in mind. To address the first challenge, a 5-state canonical content representation is introduced in Section 2.1, along with the required lexical syntax and grammar. The second challenge is addressed through stochastic state transitions, in Section 2.2, as an extensible mechanism to control the order and degree of coupling within the corpus of infinitesimal undirected narrative elements.

2.1 Undirected Narrative, Canonical Representation, and Lexical Syntax

The choice of infinitesimal undirected narrative elements lies in both their flexibility to model classical content segments and real world events, which are often unrelated or loosely coupled, and in the fact that a directed narrative, in its linear and non-linear forms, could be built from a set of tightly coupled, ordered, infinitesimal undirected narrative elements. To maintain separation between narrative structure and content, and to support content reusability within and across applications, the segmented multimodal contents need to be mapped onto the set of undirected narrative elements that are in a canonical form. Although variety of published works attempted to break contents up into story clips, behavioral elements, or basic building blocks as in Mad Tea-Party [5], and DINAH [4], the identified building blocks and behavioral elements are story specific, and not in a standardized canonical form to facilitate reusability. In this section, a canonical form defining cross-application infinitesimal undirected narrative model for generic multimodal contents, including the required lexical syntax, states, and grammar are specified.

The proposed lexical syntax specifies 5 states against which multimodal contents need to be segmented and classified, and is shown below in extended Backus-Naur form:

(1) A begin element: an infinitesimal undirected narrative element representing the start of a story line or task subset. The begin element could be used by the content author to determine which stage the user is in, within the AR environment, and is specified as follows:

```
<B>::= {begin-element}|{intro-element}
```

(2) A transaction start element: an infinitesimal undirected narrative element representing the start of an interaction, a story element, or a sub-task, that requires a response. Transaction elements need to be time limited as not to interfere with user's experience, and may be associated with a time-out.

```
<TxP>::= {<tx-initiation-element>}
```

(3) A transaction response element: an infinitesimal undirected narrative element representing the response to an interaction, a story element, or a sub-task, that was initiated with a TxP element. This element could be utilized by the content author to determine success or failure of the user in accomplishing a given task.

```
<TxR>::= { <tx-response-element> }
```

(4) A continuity element: an infinitesimal undirected narrative element that represents any generic interaction that can not be classified as one of the

other 4 states. This element could also be a no-op or a time-wait, and could be used by the content author to maintain continuity of the user's experience across objects and across variations.

```
<C> ::= { <space-filler> }               |
        { <redirection-element> }        |
        { <suggestion-element>  }        |
        { <surprise-element> }           |
        { <interruption-element> }
```

(5) An end element: an infinitesimal undirected narrative element that represents the end of a story line or task subset. The end element could be used by the content author to determine user's progress within the AR environment, as well as success or failure in accomplishing a given task, and is specified as follows: `<E>  ::= { <end-element> }`

With the total set of interactions within an AR environment modeled by a set of infinitesimal undirected narrative elements, as specified above, a given user's interactions within such an environment capture his own unique experience, and is modeled by a linear narrative whose context free grammar, CFG, is specified as follows:

```
<user-interaction-element>::= [<B>]*[<C>]*
                              [[<TxP>]*[<C>]*[<TxR>]*]*
                              [<C>]*[<E>]*
<user-experience> ::=  { <user-interaction-element> }
```

Conversely, the AR environment creates an experience for the user by transforming a subset of infinitesimal undirected narrative elements representing future events into a past represented by a linear narrative, which could be stated as follows:

```
<linear-narrative> ::= [<undirected-narrative-element>]*
```

2.2 Stochastic State Transitions

Transforming a set of unrelated or loosely coupled infinitesimal undirected narrative elements that represent the set of possible user interactions into a past - represented by a linear narrative - requires attribute-based association, or state transitions, between the elements that represent states, of the mentioned set. One extreme would be to organize the infinitesimal elements into an ordered directed linear narrative, where unrestricted user interactions within the interactive environment would be sacrificed. Another extreme would be to assign the state transitions between all the infinitesimal elements equal probabilities, resulting in a random environment, often referred to as a universe of chaos. The proposed framework supports both extremes, reflecting EDSUN's flexibility and extensibility.

Mathematically, the universe of segmented and classified contents S, are divided into subsets $s_{i,}$ where each subset embodies a given objective or an element of a story line, and is assigned a temporal position relative to other subsets, as illustrated by the time line ($t_1..t_n$) in Figure 1. State transition probabilities are then assigned to each

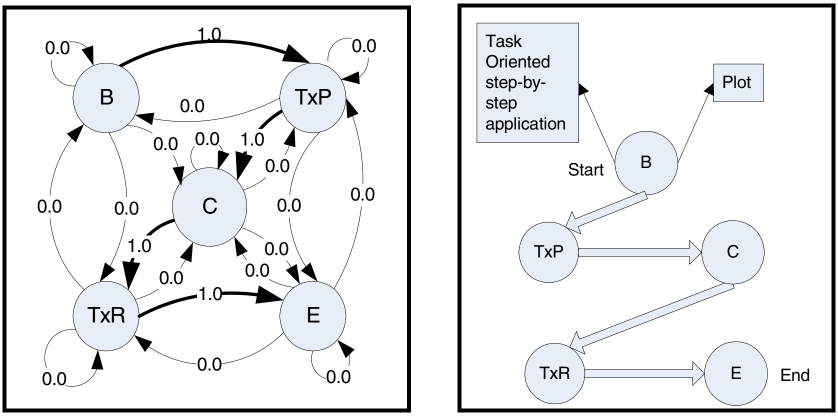


Fig. 2. On the right: an illustration of a classical linear directed narrative. On the left, EDSUN's support for an equivalent linear directed narrative.

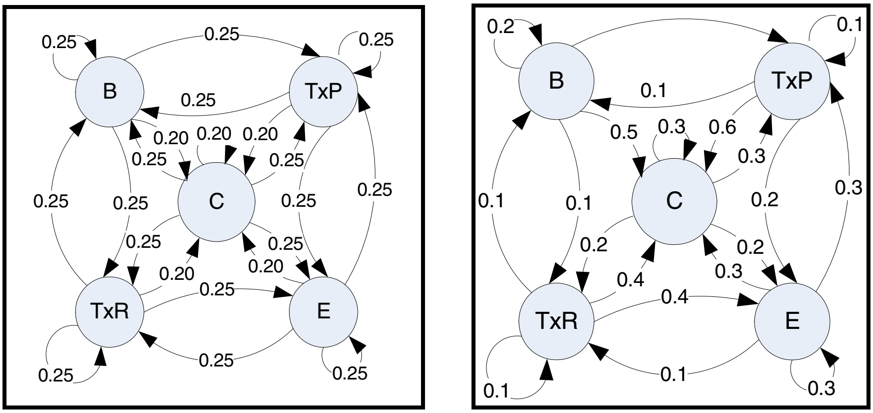


Fig. 3. On the left: EDSUN's support for the universe of chaos, and fora sample model of event-driven AR contents on the right

state transition within a given subset as shown in Figures 2 and 3, as well as to those that connect two subsets based on their temporal positions. The state transition probabilistic values allow for the desired degrees of freedom of a user's interactions, whereas additional external weights can be added, and adjusted externally based on the interactive environment or user's profile to provide for additional fine tuning and enhanced realism of user's experience.

The lexical syntax of section 2.1 specifies 5 states as follows:

$$S = \{ \text{B, C, TxP, TxR, E} \}. \quad (1)$$

To form the user's experience from the set of infinitesimal undirected narrative elements, the likelihood of each state transition is specified as follows:

$A = \{ a_{ij} \}, \quad 1 \leq i,j \leq N;$ where A is the set of state transitions between each of the states in S, and (2)

$$\sum_{j=1}^{N} a_{ij} = 1; \text{ where } a_{ij} \text{ is the state transition probability from } s_i \rightarrow s_j \qquad (3)$$

Therefore, the maximum number of possible unique user experiences in proposed model per content sequence is: $g = (N)^t$ (4)

where N = 5 as the number of states in the proposed lexical syntax, and t is the number of time slice increments that forms the subset of infinitesimal undirected narrative elements for a given user's experience. This provides for a quantifiable measurement of percent coverage of a given AR environment implementation compared with the real world, or the maximum possible set of experiences the user may have.

2.3 Hierarchical Grammars

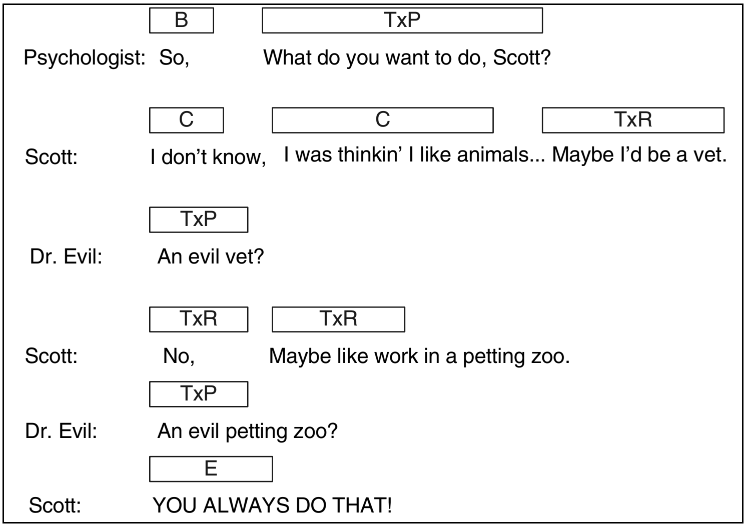
The output of stochastic state transitions of Section 2.1 forms the input set for one or more optional layers of experience specific grammars, such as that of Branigan's cinematic narrative[1] that was implemented in DINAH[4], natural language processing of Façade[8], or plan-based multimodal interaction processing of Goldfinger[12]. Hierarchical grammars enable concurrent support for multiple narratives, complex interactions [2], and languages [3], as well as specialized new experience generation from existing contents [9],[10],[12], all while enforcing the corresponding preconditions, post conditions, and experience continuity requirements. Consequently, EDSUN extensions may maintain and enhance story coherency, or ensure a specific level of dramaturgy. For example, the metalinear narrative approach of DINAH[4], and its auto-generation of story-clip based experiences can be supported here as an extension.

2.4 Structured and Unstructured Variations

To represent environments with rich contents and variety of tones, support for structured and unstructured variations is introduced. Variations are structured when variable contents representing each of the 5 states in equation (1) exist for the lifetime of a given experience, and is represented along the variations axis in Figure 1. Variations are unstructured when variable contents exist for a subset of the 5-states in equation (1) and not for the full duration of a given experience. Unstructured variations are modeled through sub-state transitions within each state where variety of contents exist. Both structured and unstructured variations enable recreating environments with varying tones for a given experience, such as polite, rude, reckless, or insane environments to name a few, as well as representation of historical versus fictional and mythological environments.

2.5 Content Reusability and New Experience Generation – An Example

The example introduced here illustrates content reusability and new experience generation once contents are segmented, and classified according to the 5-state lexical syntax of Section 2.1. Figure 4 contains the original text of a dialog from one of Austin Power's movies titled "International Man of Mystery". Figure 5 contains



Psychologist: So, What do you want to do, Scott?

Scott: I don't know, I was thinkin' I like animals... Maybe I'd be a vet.

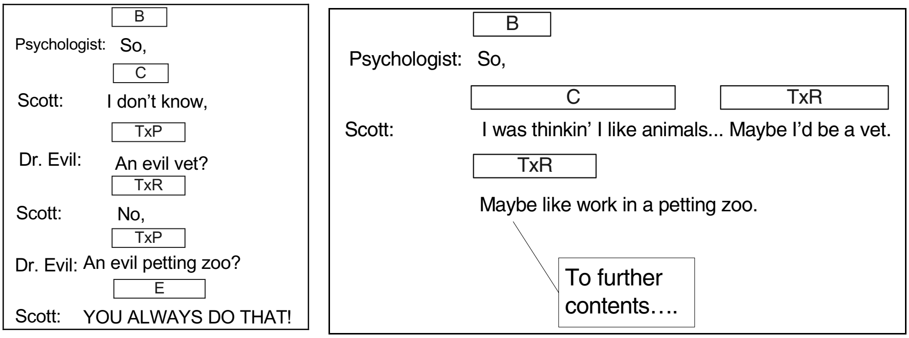
Dr. Evil: An evil vet?

Scott: No, Maybe like work in a petting zoo.

Dr. Evil: An evil petting zoo?

Scott: YOU ALWAYS DO THAT!

Fig. 4. A sample classification of an original dialog from Austin Powers "Inter-national Man of Mystery" – as mapped to the proposed 5-state canonical repre-sentation described in Section 2.1



Psychologist: So,

Scott: I don't know,

Dr. Evil: An evil vet?

Scott: No,

Dr. Evil: An evil petting zoo?

Scott: YOU ALWAYS DO THAT!

Psychologist: So,

Scott: I was thinkin' I like animals... Maybe I'd be a vet.

Maybe like work in a petting zoo.

Fig. 5. An illustration of grammar and state transition compliant variants of original text of Figure 4. (a) On the left, a typical teenager-parent talk emphasizing impatience and lack of context. (b) On the right, a thoughtful detailed dialog with no references to Dr. Evil.

two new experiences generated from the original text, while complying with the grammar and state transitions of Sections 2.1 and 2.2. The left side of Figure 5 presents the experience of a typical teenager-parent's talk, emphasizing impatience and lack of context. The right side of Figure 5 presents a more thoughtful experience with no references to “evil”. This example illustrates the extensibility of EDSUN, and its ability to support different narratives concurrently as well as to generate new experiences from existing contents.

3 Conclusion

In this paper, an extensible framework for interactive multimodal content representation is introduced and examined. The proposed framework, EDSUN, enforces separation between narrative structure and content. EDSUN is based on infinitesimal undirected narrative structure to support unconstrained user interactions, and to enable event-driven interactivity with multimodal contents. EDSUN's canonical format along with the proposed lexical syntax and grammar enable cross-application content reusability and new experience generation as shown in Section 2.5. The associated stochastic state transitions enable concurrent support for different types of narratives, and provide the content author with flexibility in coupling content elements. Extensibility through the proposed hierarchical grammars enable complex behavioral representation and user-defined narrative structures, whereas structured and unstructured variations facilitate the management of rich contents. Based on early conceptual experimentation with EDSUN, the potential it provides as an enabler for rich interactive multimodal contents as well as cross-application reusability is rather promising.

References

1. Branigan, E. “Narrative Comprehension and Film”, Sightlines,ed. E. Buscombe. Vol. 1. Routledge, New York, New York, 1992.
2. Oliver, N. Horvitz, E., and Garg, A. “Layered representations for human activity recognition”, In Fourth IEEE Int. Conf. on Multimodal Interfaces, pages 3-8, 2002.
3. Ivanov, Y., Bobick A., “Recognition of visual activities and interactions by stochastic parsing”, IEEE Trans. PAMI, vol. 22(8), pp. 852-872, 2002.
4. Ventura D., Brogan D., “Digital Storytelling with DINAH: dynamic, interactive, narrative authoring heuristic.”, IFIP First International Workshop on Entertainment Computing, May 2002.
5. MacIntyre, B., Bolter J.D., Moreno E., Hannigan B., "Augmented Reality as a New Media Experience", IEEE and ACM International Symposium on Augmented Reality (ISAR'01), pp. 197-206, 2001.
6. Schneider, H., “GEIST: Mobile outdoor AR-Information System for Historical Education with Digital Story Telling”, CeBIT 2003, and Intergeo 2003.
7. MacIntyre, B., Bolter, J. D., “Single-narrative, multiple point-of-view dramatic experiences in augmented reality”, Virtual Reality, Vol. 7, pp. 10-16, 2003.
8. Mateas, M., Stern A., “Natural Language Understanding in Façade: Surface-text Processing”. In Proceedings of the 2nd Technologies for Interactive Digital Storytelling and Entertainment (TIDSE'04), Darmstadt, Germany.

9. Mateas, M., Stern A., “A Behaviour Language: Joint Action and Behavioural Idioms”. In Predinger, H. and Ishiuka, M. (Eds), Life-like Characters: Tools, Affective Functions and Applications, Springer 2004.
10. Klesen, M., Kipp, M., Gebhard, P., Rist, T., “Staging exhibitions: methods and tools for modelling narrative structure to produce interactive performances with virtual actors.”, Virtual Reality, Vol. 7, pp. 17-29, 2003.
11. Cavazza, M., Charles, F. and Mead S.J., ”Developing Re-usable Interactive Storytelling Technologies”. IFIP World Computer Congress 2004, Toulouse, France.
12. Cavazza, M., Martin O., Charles, F. and Mead S.J., Marichal X. and Nandi A., “Multimodal Acting in Mixed Reality Interactive Storytelling”, IEEE Multimedia, July-September 2004, Vol. 11, Issue 3.