
# SceneMaker

## Intelligent Multimodal Visualisation of Natural Language Scripts

**Eva Hanser**
Dipl.-Des. (FH), M.Sc.

**Prof. Paul Mc Kevitt, Dr. Tom Lunney, Dr. Joan Condell**

School of Computing & Intelligent Systems
Faculty of Computing & Engineering
University of Ulster, Magee
hanser-e@email.ulster.ac.uk, {p.mckevitt, tf.lunney, j.condell}@ulster.ac.uk


# PRESENTATIONOUTLINE

**Aims & Objectives**

**Literature Review**

**Project Proposal**

**Relation to Other Work**

**Conclusion and Future Work**

# AIMES & OBJECTIVES

# AIMS


- Automatically generate affective virtual scenes from screenplays/play scripts
- Realistic visualisation of emotional aspects
- Enhance believability of virtual actors and scene presentation
- Multimodal representation with 3D animation, speech, audio and cinematography

# OBJECTIVES

- Emotions and semantic information from context
- Cognitive reasoning rules combined with commonsense and affective knowledge bases
- Automatic genre recognition from text
- Editing 3D content on mobile devices
- Design, implementation and evaluation of *SceneMaker*

# LITERATUREREVIEW

: LITERATUREREVIEW

# SEMANTICTEXTPROCESSING

**INT. M.I.T. HALLWAY -- NIGHT**

**Lambeau and Tom come around a corner. His P.O.V. reveals a figure in silhouette blazing through the proof on the chalkboard. There is a mop and a bucket beside him. As Lambeau draws closer, reveal that the figure is Will, in his janitor's uniform. There is a look of intense concentration in his eyes.**

**LAMBEAU**
**Excuse me!**

**WILL**
**Oh, I'm sorry.**

**LAMBEAU**
**What're you doing?**

**WILL**
**(walking away)**
**I'm sorry.**

Screenplay Extract from ‘Good Will Hunting (1997)’

- Text layout analysis
- Semantic information on location, timing, props, actors, actions and manners, dialogue
- Parsing formal structure of screenplays
  (Choujaa and Dulay 2008)

# COMPUTATIONOF EMOTIONANDPERSONALITY

- Emotion models: Basic emotions (Ekman and Rosenberg 1997)
  Pleasure-Dominance-Arousal (Mehrabian 1997)
  OCC – appraisal rules (Ortony et al. 1988)
- Personality models: OCEAN (McCrae and John 1992)
  Belief-Desire-Intention (Bratman 1987)
- Emotion recognition from text:
  Keyword spotting, lexical affinity, statistical models, fuzzy logic rules, machine learning, common knowledge, cognitive model

# VISUALANDEMOTIONALSCRIPTING

- Scripting Notation for visual appearance of animated characters
- Various XML-based annotation languages:

  EMMA (EMMA 2003)
  BEAT (Cassel et al. 2001)
  MPML & MPML3D (Dohrn and Brügmann 2007)
  AffectML (Gebhard 2005)


```
<GAZE word=1 time=0.0 spec=AWAY_FROM_HEARER>
<GAZE word=3 time=0.517 spec=TOWARDS_HEARER>
<R_GESTURE_START word=3 time=0.517 spec=BEAT>
<EYEBROWS_START word=3 time=0.517>
```

: LITERATUREREVIEW

# MODELLINGAFFECTIVEBEHAVIOUR

- Automatic physical transformation and synchronisation of 3D model
- Manner influences intensity, scale, force, fluency and timing of an action
- Multimodal annotated affective video or motion captured data (Gunes and Piccardi 2006)

AEOPSWORLD
(Okada et al. 1999)

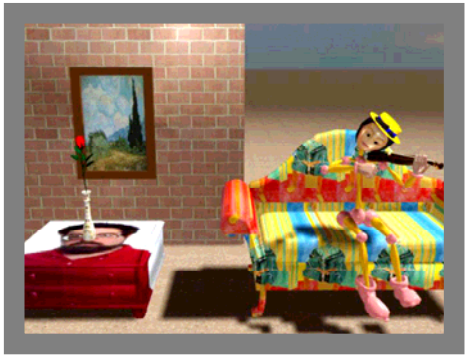
Greta
(Pelachaud 2005)

Personality&Emotion Engine
(Su et al. 2007)

: LITERATUREREVIEW

# VISUALISING3DSCENES

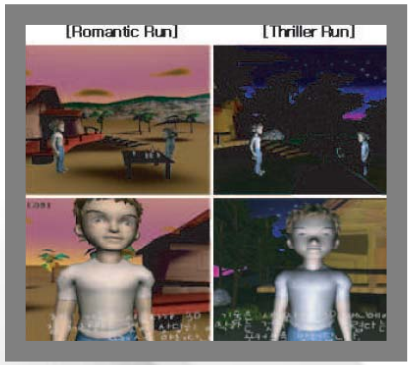
- SONAS – Environment visualisation
  (Kelleher et al. 2001)
- WordsEye – Scene composition
  (Coyne and Sproat 2001)
- ScriptViz – Screenplay visualisation
  (Liu and Leung 2006)
- CONFUCIUS – Action, speech & scene animation
  (Ma 2006)
- CAMEO – Cinematic and genre visualisation
  (Shim and Kang 2008)



WordsEye

ScriptViz

CONFUCIUS




CAMEO

: LITERATUREREVIEW

# AUDIOGENERATION

- Emotional speech synthesis (Schröder 2001)
  - Prosody rules
- Music recommendation systems
  - Categorisation of rhythm, chords, tempo, melody, loudness and tonality
  - Sad or happy music and genre membership (Cano et al. 2005)
  - Associations between emotions and music (Kuo et al. 2005)

# MULTIMODALMOBILEINTERFACES

- Strategies for large 3D graphics data and knowledge bases on small, limited devices:
- Distribution of system architecture between server and mobile device
  - SmartKom Mobile (Wahlster 2006)
  - Multimodal Dialogue (Turunen et al. 2005)
- Rendering on mobile devices
  - M3G (Java API), OpenGL ES, VRML, MPEG-4, H-Anim

SCENEMAKER
EVAHANSER
CONFIRMATION
VIVA

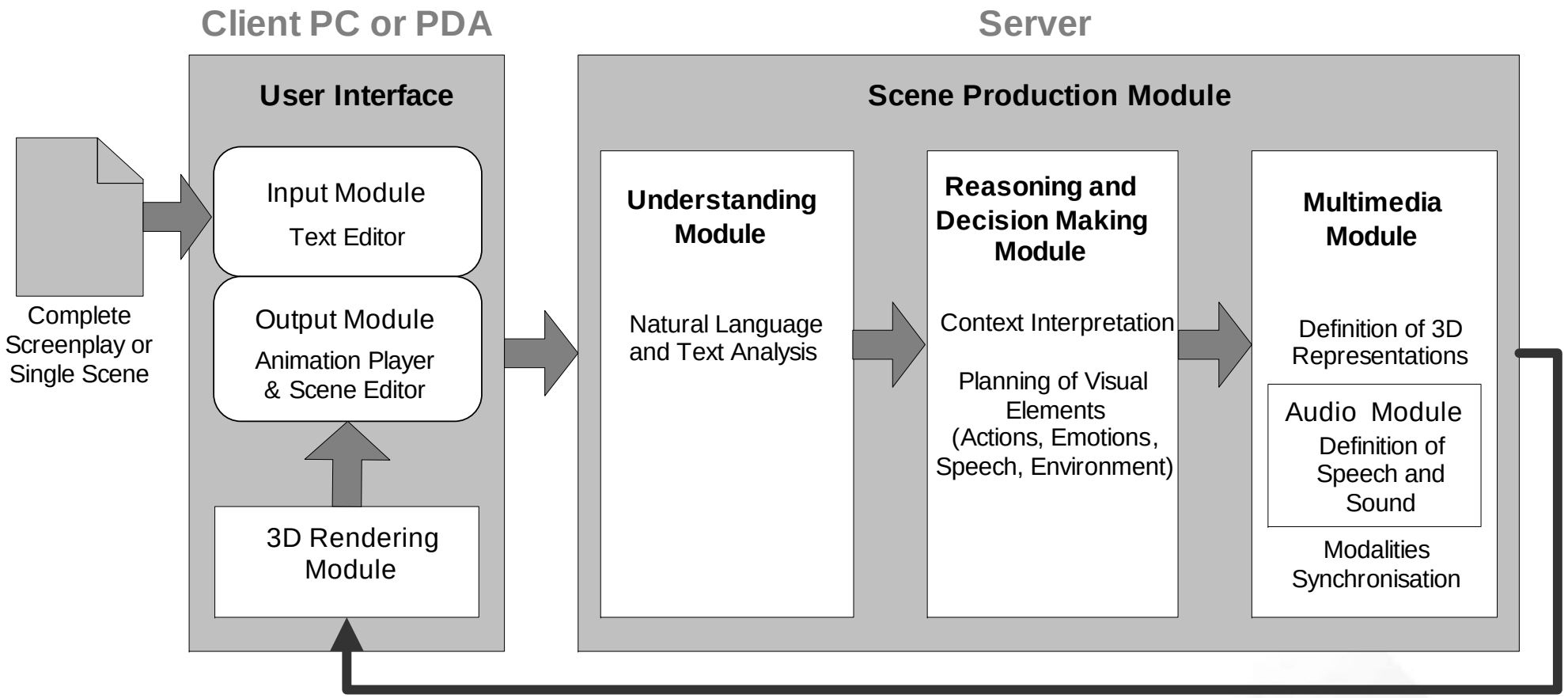
# PROJECTPROPOSAL

# KEYOBJECTIVES

- Context consideration through natural language processing, common knowledge and reasoning methods
- Fine grained emotion distinction with OCC
- Extract genre and moods from screenplays
- Influence on Visualisation
- Enhance naturalism and believability
- Text-to-animation software prototype, *SceneMaker*

: PROJECTPROPOSAL

# ARCHITECTUREOFSCENEMAKER




Architecture of *SceneMaker*

# SOFTWAREANDTOOLS

- Language processing module of CONFUCIUS
  Part-of-speech tagger, Functional Dependency Grammars, WordNet, LCS database, temporal relations, visual semantic ontology

  Extensions :

  Context and emotion reasoning :
  ConceptNet, Open Mind Common Sense (OMCS), Opinmind, WordNet-Affect

  Text pre-processing :
  Layout analysis tool with layout rules
  Genre-recognition tool with keyword co-occurrence, term frequency and dialogue/scene length

: PROJECTPROPOSAL

# SOFTWAREANDTOOLS *CONT*.

- Visualisation module of CONFUCIUS
  H-Anim 3D models, VRML, media allocation, animation scheduling

  Extensions:
  Cinematic settings (EML),
  Affective animation models

- Media module of CONFUCIUS
  Speech Synthesis FreeTTS

  Extension:
  Automatic music selection

- User Interface for mobile and desktop
  VRML player, script writing tool, 3D editing

# EVALUATIONOFSCENEMAKER

## Evaluating 4 aspects of SceneMaker:

| Aspect | Evaluation |
|---|---|
| Correctness of screenplay analysis & visual interpretation | Hand-animating scenes |
| Effectiveness of output scenes | Existing feature film scenes |
| Suitability for genre type | Scenes of unknown scripts categorised by viewers |
| Functionality of interface | Testing with drama students and directors |

# RELATIONTOOTHERWORK

SCENEMAKER EVAHANSER CONFIRMATION VIVA

# RELATIONTOOTHERWORK

| System | Year | Category | Input (Perception) | | | | Output (Generation) | | | | | | | | Emotions | | | | | Story Processing | | | Story Type | | Author Options | | Access | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Text | Speech | Other (Code, GUI) | Other (Vision, Gesture) | Text | Speech | Gaze | Facial Expression | Body Language | Music/Sound | Camera | Lighting | Genre | Basic Emotions | OCC, PAD, BDI | Personality | Social Roles | Narrative Rolls | Layout/Dramatic Analysis | Sentence-based | Context Memory | Linear Storyline | Interactive Story | User Interface | Code/Rules Extendable | Mobile Devices | Internet |
| Improv | 1996 | Virtual Agents | | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | | | | ✓ | | ✓ | | | | | | | ✓ | ✓ | ✓ | | ✓ |
| AESOPWORLD | 1999 | | NL | ✓ | | ✓ | | ✓ | | | ✓ | ✓ | | | | ✓ | | ✓ | | | | | ✓ | ✓ | | | | | |
| Max | 2008 | | D | | | ✓ | | ✓ | ✓ | ✓ | ✓ | | | | | | ✓ | (✓) | | | | ✓ | | | ✓ | | | | |
| Greta | 2005 | | D | | ✓ | | | ✓ | ✓ | ✓ | ✓ | | | | | | ✓ | ✓ | ✓ | | | ✓ | | | ✓ | | | | |
| SCREAM | 2002 | Scripting Tools | D | | ✓ | | | ✓ | | | ✓ | | | | | ✓ | | ✓ | ✓ | | | | ✓ | | ✓ | | ✓ | | ✓ |
| EML | 2006 | | | ✓ | | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | | | | ✓ | | ✓ | | | | | ✓ |
| WordsEye | 2001 | Text to Animation | (NL) | | | | | | | | ✓ | | ✓ | ✓ | | | | | | | | ✓ | | ✓ | | ✓ | | | ✓ |
| Spoken Image | 94/01 | | NL | ✓ | | ✓ | | | | | | | | | | | | | | | | ✓ | | ✓ | | | | | |
| Fuzzy P&E Engine | 2007 | | SP | | | | | | | | ✓ | | | | | ✓ | | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | | | |
| Behaviour Generation System | 2007 | | NL | | | | | ✓ | ✓ | | ✓ | | | | | | | | ✓ | | | ✓ | | ✓ | | | | | |
| ScriptViz | 2006 | | SP | | | | | | | ✓ | ✓ | | ✓ | | | ✓ | | | | | | ✓ | | ✓ | | ✓ | | | |
| CAMEO | 2008 | | SP | | ✓ | | | ✓ | | NS | ✓ | ✓ | ✓ | ✓ | ✓ | NS | | | | | | | ✓ | ✓ | | ✓ | | | | |
| CONFUCIUS | 2006 | | NL | | | | | ✓ | | | ✓ | ✓ | ✓ | | | | | | | | | ✓ | | ✓ | | | | | |
| *SceneMaker* | 2009 | | SP | | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | | ✓ | ✓ |

## RELATIONTOOTHERWORK
# POTENTIALCONTRIBUTIONS

- Context reasoning to influence emotions
  requires common knowledge bases and context memory
- Text layout analysis to access semantic information
- Visualisation from sentence, scene or full script
- Automatic genre specification
- Automatic development of personality, social status, narrative role and emotions

: PROJECTPROPOSAL

# PROJECTPLAN

| Research Activities | 1st year | | | | 2nd year | | | | 3rd year | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2008 | 2009 | | | 2010 | | | | 2011 | | | |
| | Oct-Dec | Jan-Mar | Apr-Jun | Jul-Sep | Oct-Dec | Jan-Mar | Apr-Jun | Jul-Sep | Oct-Dec | Jan-Mar | Apr-Jun | Jul-Sep |
| Literature Review | | | | | | | | | | | | |
| 100 Day Review and Presentation | | | | | | | | | | | | |
| Submission to ISEA2009 | | | | | | | | | | | | |
| In-Depth Review of Systems and Approaches Relevant for Text to Animation Modelling | | | | | | | | | | | | |
| Submission to IMVIP 2009 | | | | | | | | | | | | |
| Submission to KI 2009 | | | | | | | | | | | | |
| Development of detailed System Architecture for SceneMaker | | | | | | | | | | | | |
| Submission to AICS 2009 | | | | | | | | | | | | |
| Confirmation Report and Presentation | | | | | | | | | | | | |
| Submission to 'Artificial Intelligence Review' Journal by Springer | | | | | | | | | | | | |
| Design Automated Scene Production Module of SceneMaker | | | | | | | | | | | | |
| Design User Interface Module of SceneMaker | | | | | | | | | | | | |
| 2nd Year Poster | | | | | | | | | | | | |
| Implementation of Scene Production Module | | | | | | | | | | | | |
| Implementation of User Interface in Accordance with HCI | | | | | | | | | | | | |
| Submission to IEEE Pervasive Computing Journal | | | | | | | | | | | | |
| Test and Evaluate | | | | | | | | | | | | |
| Improve Developed SceneMaker Prototype | | | | | | | | | | | | |
| Submission to ACM Transactions on Multimedia Computing, Communications and Applications | | | | | | | | | | | | |
| 3rd Year Presentation | | | | | | | | | | | | |
| Thesis write up | | | | | | | | | | | | |

: PROJECTPROPOSAL

# THESISOUTLINE

0. Prelims, Abstract
1. Introduction
2. Literature Review
   - Work in area of natural language processing and intelligent multimodal visualisation
3. Theoretical contribution
   - Automatic, intelligent, multimodal and affective text-to-animation generation
4. Description of software prototype
   - Implementation of the text-to-animation system, *SceneMaker*
5. Evaluation of SceneMaker
   - Test results and evaluation of text-to-animation generation with *SceneMaker*
6. Conclusion
   - Summary, Relation to other work, Future Work

Appendices, References

# PUBLICATIONS

Eva Hanser and Paul Mc Kevitt. (2009). "NewsViz: Extraction and Visualisation of Emotions from News Articles". In: ISEA 2009,International Symposium on Electronic Art. Belfast, Northern Ireland, 23 Aug – 1 Sept 2009.

# CONCLUSION AND FUTURE WORK

- Automatic visualisation of affective expression of screenplays/play scripts
- Heightened expressivity, naturalness & artistic quality
- Assist directors, actors, drama students, script writers
- *Focus on semantic interpretation, computational processing of emotions, reflecting affective behaviour and expressive multi-media scene composition*
- Future work: Implementation of *SceneMaker*


Thank you.

# QUESTIONS OR COMMENTS?