SceneMaker: Intelligent Multimodal Visualisation of Natural Language Scripts

Eva Hanser
School of Computing & Intelligent Systems
Faculty of Computing & Engineering
University of Ulster, Magee
Derry/Londonderry BT48 7JL
Northern Ireland
E-mail: hanser-e@email.ulster.ac.uk

100 Day Review Report

Febuary, 2009

Supervisors: Prof. Paul Mc Kevitt, Dr. Tom Lunney, Dr. Joan Condell

Abstract

Performing plays or creating films/animations is a complex and thus expensive process involving various professionals and media. This research project proposes to augment this process by automatically interpreting film/play scripts and generating animated scenes from them. Therefore a web based software prototype, *SceneMaker*, will be implemented. During the generation of the story content, particular attention will be given to emotional aspects and their reflection in the execution of all types of modalities (fluency and manner of action/behaviour, speech, gaze duration and direction, scene composition, timing, lighting, music, camera, set/stage, costumes). Literature on related research areas of Natural Language Processing (NLP) with regard to personality and emotion detection, embodied agents, modeling affective behaviour, visualisation of 3D scenes and digital cinematography is reviewed. Technologies and software relevant for the development of *SceneMaker* are analysed. The project's aims, objectives and development plan are presented. How the scene and actor behaviour changes when emotional states are taken into account (e.g. a happy versus a sad state) will be investigated. Potential unique contributions of this research are the generation of complete scenes from play scripts, the development of a methodology which combines all relevant modalities, influences of expressivity on all modalities and deployment on mobile devices. In conclusion, *SceneMaker* will reduce production time, save costs and enhance communication of ideas providing quick pre-visualisations of scenes.

***Keywords*:** Natural Language Processing, Intelligent Multimodal Interfaces, Film Making/Theatre Production, Affective Agents, Emotional Body Posture Modeling, 3D Visualisation, SceneMaker

1 Introduction

The production of plays or movies is an expensive process involving planning and rehearsal time, actors, technical equipment for lighting, sound and special effects. It is also a creative act which might not always be straightforward, but requires experimentation, visualisation of ideas and their communication between everyone involved (e.g. play writers, directors, actors, camera man, orchestra, managers, costume and set designer). This research proposes a web based software prototype, *SceneMaker*, which will assist in this production process. *SceneMaker* will provide a facility for everyone involved in the creation of dynamic/animated scenes to test and pre-visualise scenes before putting them into action. Users input a natural language text scene script and automatically receive multimodal 3D visualisations taking into account considerations such as aesthetics and emotions. The user can refine the output through an interface which facilitates the control of character personality, emotional states, modalities of output, actions and cinematographic settings (e.g. lighting and camera). Such technology could be applied in the training of film/drama directors without having to continuously utilise expensive actors and actresses. Alternatively it could be used in advertising agencies that regularly need to visualise numerous ideas and concepts. At the Ohio State University a virtual theatre interface for teaching drama students about lighting, positioning on stage and different view points (Virtual Theatre, 2004) was considered very beneficial and had a significant impact on training methods.

SceneMaker will be accessible over the internet and thus will be an easily available tool for script writers, animators, directors, actors or drama students to creatively and inexpensively express their ideas and prove their effectiveness in achieving their desired effect whilst writing or advising directors on set. Successful example scenes can be saved and shared with other scene producers in an online gallery classified by different film/drama genres and scene topics. A Graphical User Interface (GUI) suitable for mobile devices is intended to

facilitate the use of *SceneMaker* on stage or on set. The *SceneMaker* prototype will be developed using appropriate multimodal technology and will extend an existing software prototype, CONFUCIUS (Ma, 2006), which performs automated conversion of natural language to 3D animation.

SceneMaker focuses on the precise representation of emotional expression in all modalities available for scene production and especially on most human-like modeling of body language as it is the most expressive modality in human communication, delivering 60-80 percent of our messages. Actual words only present 7-10 percent of all modalities delivering a message in conversation (Su et al., 2007). Further modalities include voice tone, volume, facial expression, gaze, gestures, body posture, spatial behaviour and aspects of appearance. These facts show the importance of the visualisation of body language in film/play production, but also point out the challenges in deriving information for animation from scripts containing mostly dialogues. Much research is dedicated to detailed modeling of emotion and facial expressions, gaze and hand gestures (Kopp et al., 2008; Sowa, 2008), but body posture has yet to be addressed extensively (Gunes and Piccardi, 2006).

1.1 Research Aims and Objectives

This research aims to solve three research questions: How can emotional information be computationally interpreted from screenplays and structured for visualisation purposes? How can emotional states be synchronised in presenting all relevant modalities? Can compelling, life-like animations be achieved? Therefore this research aims to implement an automated animation system, with a user interface for manual manipulation, catering for affective actor modeling and scene production based on personality, social and narrative roles and emotions. The objective is to give directors or animators a reasonable idea of what the scene they are planning will look like. The software prototype, *SceneMaker*, will be a multimodal content generation system, accessible on mobile devices, which can be applied seamlessly for testing and customizing performances according to the producers' intentions. *SceneMaker* will provide a unique training facility for those involved in scene production. It may also be useful for advertising agencies, which constantly need rapid visualisations of various ideas and concepts.

Section 2 of this report gives an overview of current research on scene production for multimodal and interactive storytelling, virtual theatre and affective agents. In section 3, the project proposal and prototype, *SceneMaker*, are described in detail. *SceneMaker* is compared to related multimodal visualisation applications in section 4. Section 5 concludes the report.

2 Literature Review

Automatic and intelligent production of film/theatre scenes with characters expressing emotional states involves three development stages:

1. The detection of personality traits and the emotional impact of the story from the film/theatre script;
2. The manipulation of the 3D models and their actions according to the emotional findings;
3. The development of an interface for directors, actors or others involved in film making/play production.

This section reviews state-of-the-art advances in these areas and discusses related research projects.

2.1 Detecting Personality and Emotions of Characters in Film/Play scripts

All modalities of human interaction express personality and emotional states namely voice, word choice, gestures, body posture and facial expression. Therefore many projects aiming to create life-like characters integrate affective modeling. Psychological theories for emotion, mood, personality and social status are translated into computable methods. McDonnell et al. (2008) have proven that the perception of Ekman's six basic and universally recognizable facial expressions (happiness, surprise, sadness, disgust, fear and anger) (Ekman and Rosenberg, 1997) also applies to emotional body language and virtual characters independent of their body representation. 3D figures reflect the richness of human bodily expression with varying intensity as perceived in Shaarani and Romano (2008). Scripting languages have been developed to specifically cater for the modeling of affective characteristics. Non-verbal behaviour is automatically modelled from conversational text with the Behaviour Expression Animation Toolkit (BEAT) (Cassell et al., 2001), which can be combined with systems that assign personality profiles, motion characteristics, scene constraints, or the animation style of a particular animator. In SPARK, for instance, BEAT is used to annotate chat messages to automate avatar behaviour in an online virtual environment (Vilhjálmsson and Thórisson, 2008). SCREAM (Prendinger and Ishizuka, 2002) is a web-based scripting tool for multiple characters which computes affective states based on the OCC-Model (Ortony et al., 1988) of appraisal and intensity of emotions, as well as social context. ALMA (Gebhard, 2005), a layered model of affect, implements AffectML, an XML based modeling language which incorporates the concept of short-term emotions, medium-term moods and long-term personality profiles by mapping the five personality traits of the OCC-Model onto the Pleasure, Arousal and Dominance (PAD) mood space. Appraisal rules determine how a character appraises its environment, events and its own or other characters' acts or emotional displays to filter out relevant emotions for display. Based on the analysis of

linguistic and contextual information of dialogue scripts, Breitfuss et al. (2007) automatically add appropriate non-verbal behaviour descriptions. The annotated script is transformed into the Multimodal Presentation Markup Language (MPML) to model facial and body animations and speech synthesis of 3D agents. Su et al. (2007) decode the meaning of story scripts/scene descriptions and classify the affective state of the story characters. A psychological model of personality, the Five-Factor-Model (De Raad, 2000), Ekman's six basic emotions and a model of story character roles are combined through a hierarchical fuzzy rule-based system to control the body language of the characters.

2.2 Embodied Agents and Modeling Personality and Emotionally Influenced Behaviour

Gandalf and REA (Vilhjálmsson and Thórisson, 2008) are both early conversational agents capable of real-time face-to-face conversations with human users and multimodal natural language generation and understanding. They are built on the Ymir architecture which detects descriptions of the human user's behaviour and decides on the agent's goals to be achieved through movement or speech. The Action Scheduler (AS) in Ymir generates the appropriate animation commands. The social interaction intended through dialogue in play scripts is of more importance than the actual actions carried out or topics discussed. For this reason, Paggio and Music (2001) suggest a modality unifying representation to solve such hidden intentions and ambiguities. The virtual human, Max (Kopp et al., 2008), engages museum visitors in face-to-face small talk. Max listens while the users type their input, reasons about actions to take, has intention and goal plans, reacts emotionally and gives verbal and non-verbal feedback. The high-level control of affective characters in Su et al. (2007) is mapped from the output of the Personality & Emotion engine to graphics and animation using Maya as the visualisation environment. Four main body areas are identified for human motion: head, trunk, upper and lower limbs. Possible postural values are supplied to a Dependency Graph to manipulate the shape and geometry of the model, e.g. through the stretch and squash technique. The joint values of the character's skeleton are updated according to the '12 principles of typical animation techniques for believable characters' (Thomas and Johnson, 1981; Disney Animation, 2008) based on physical characteristics of sex, space, timing, velocity, position, height, weight and portion of the body. Sequences can be layered, blended and mixed through non-linear animation. Bickmore (2004) suggests a classification of the intended or unconscious functionalities of body language: prepositional, interactional, attitudinal, affective and relational functions. Greta (Pelachaud, 2005) is modelled as an expressive multimodal Embodied Conversational Agent (ECA). The Affective Presentation Markup Language (APML) defines her facial expressions, hand and arm gestures for different communicational functions and with varying degrees of expressivity (manner). The behaviours are synchronised to the duration of phonemes of speech.

Multimodal annotation coding of video or motion captured data specific to emotion as in Gunes and Piccardi (2006) collects data in publicly available facial expression or body gesture databases, which is useful for instructing subjects on how to perform desired actions.

2.3 Visualisation of 3D Scenes and Virtual Theatre

Live performances such as theatre serve as concise test scenarios for interaction between humans and robots, or virtual agents (Breazeal et al., 2003). The storyline defines the scenario, the script provides constrained dialogues and interaction, the stage or set constrains the environment and 'actors' have to act and react in a compelling and convincing manner, which requires sophisticated perceptual, behavioural and expressive capabilities. Visualisation of scenes from text input is realised in WordsEye (Coyne and Sproat, 2001), which creates static 3D images from specific descriptive texts. In CONFUCIUS (Ma, 2006) multimodal 3D animations of single sentences are produced. 3D models perform actions, dialogues are synthesised and basic cinematic principles determine the camera placement.

The Virtual Theatre Interface project (Virtual Theatre, 2004) offers a web-based user interface to manipulate actors' positions on stage, lighting and view points. PuppetWall (Liikkanen et al., 2008) presents a multi-user, multimodal interface which supports expressive interaction through hand motion tracking, touch-displays and emotional speech recognition to direct puppets, playgrounds and props. Intelligent multimodal virtual theatre productions in real space provide human user interaction through sensitive vests or head-mounts to recognize user movements and speech (Cavazza et al., 2007; RIVME, 2004). The mixed reality project 'Casino Virtuell' (Gebhard and Schröder, 2008) realises an artificial intelligence Poker game with two virtual players whose comments and actions are influenced by their personality and emotional state giving a believable impression of expressive behaviour. Emotions are simulated with the above mentioned ALMA Model and a novel speech synthesis system controls the quality of emotional expression in the characters' synthesised voice. The game plot is generated through the 'authoring toolkit', SceneMaker (Gebhard et al., 2003). SceneMaker treats content in separate scenes (pieces of dialogue) which can be pre-scripted or author written. An author can define the logical scene flow (transitions between the scenes) of the story, in finite state machine graphs. In the Improv system (Perlin and Goldberg, 1996) an author can create virtual actors to interact with human actors in a virtual

theatre. Through an English-style scripting language the author defines the virtual character's behaviour set and personality which act as rules governing how the actor behaves, changes, makes decisions and communicates.

Furthermore, another modality, cinematography, can assist in conveying themes and moods in animations. Through the application of cinematographic rules defining appropriate placement and movement of camera, lighting, colour schemes and the pacing of shots, communicative goals can be expressed. The psychological effects of film techniques are easily understood by viewers and affect their emotional disposition. Kennedy and Mercer (2002) developed an application which automatically applies film techniques to existing animations. Reasoning about the plot, theme, character actions, motivations, emotions and narrative goals described by the human animator, it creates a communicative plan, automatically maps appropriate cinematographic effects from a knowledge base and renders the final animation. De Melo and Paiva (2006) introduce a high-level synchronized language, Expression Mark-up Language (EML), which integrates environmental expressions like cinematography, illumination and music as a new modality into the emotion synthesis of virtual humans. Inspiration for the integration of expressive audio according to relevant visual cues can be found in experimental projects for sound effects (Physically Informed Audio Synthesis, 2008) and music (Rebelo et al., 2005).

2.4 Multimodal Mobile Applications

Technological advances enable multimodal human-computer interaction in the mobile world. SmartKom Mobile (Wahlster, 2006) brings the multimodal system, SmartKom, onto mobile devices such as Personal Digital Assistants (PDAs). Supported modalities include language, gesture, facial expression and emotions carried through speech emphasis. The user interacts with a virtual character, Smartakus, through dialogue. SmartKom supports tasks of 50 different domains for scenarios in the car or for pedestrians with an extendable unified knowledge base. It is possible to place the system architecture and rendering on the mobile device itself or to distribute these from a server via wireless broadband networks. The integration of, and adaptation to, mobile context poses an interesting research question.

The wide range of approaches to modeling emotions, moods and personality aspects in virtual humans and scene environments along with first attempts to bring multi-modal agents onto mobile devices provide a good basis for the *SceneMaker* prototype.

3 Project Proposal

This research will investigate Natural Language Processing methods with regard to extracting theme and mood from film/play scripts. The main aim is to visualise the detected emotions in virtual animations. Ideas will be tested with a software prototype, *SceneMaker*, which will augment short 3D scenes with emotional influences on body language and environmental expression. A scene will be composed of affective actors and multimedia like music, sound, illumination, timing and camera work to support a given mood and theme. A front-end user interface will be created for directors or animators with a focus on high usability and intuitive operation on computers and mobile devices to allow fine-tuning of the automatically created animations.

3.1 Methodologies, Software and Prospective Tools

For the development of *SceneMaker* a Constructionist Design Methodology (CDM) (Thórisson, 2007) or more specifically a Constructionist AI Methodology (CAIM) (Thórisson et al., 2004) will be employed. Existing software tools fulfil sub-tasks and will be modified, combined and extended to construct *SceneMaker*. The Psyclone software platform (Thórisson et al., 2004) may be utilised as it simplifies the design of complex systems and their connections between various input and output modules like e.g. vision, body tracking and graphics. By incorporating the OpenAIR specification (Thórisson, 2007), for information exchange and network routing; Psyclone provides a unified messaging format to communicate between different modules. The architecture of *SceneMaker* will include modules for perception, i.e. text interpretation, decision, i.e. emotion and personality classification, and action, i.e. 3D character animation and further modules for multimodal output like speech synthesis, selection and placement of music, camera and lighting. For the automatic interpretation of the input scripts, *SceneMaker* will build upon the Natural Language Processing Module of CONFUCIUS (Ma, 2006). The syntactic knowledge base (Connexor Part of Speech Tagger (Connexor, 2003), Functional Dependency Grammars (Tesniere, 1959)), semantic knowledge base (WordNet (Fellbaum, 1998), LCS database (LCS, 2000)) and temporal language relations will be extended by an emotional knowledge base. Visual knowledge, such as object models and event models, will be related to emotional cues. Cinematic principles will be classified into expressive categories. EML (De Melo and Paiva, 2006) appears to be a comprehensive XML-based scripting languages to model expressive modalities, but might need to be extended to serve all expressions in *SceneMaker*. Resources of 3D models are available on the internet, for instance Microsoft Agents (MS-Agents, 2008) which have functionalities to model personalities and synchronise speech or H-Anim models (H-Amin, 2001) as used in CONFUCIUS which include geometric or physical, functional and spatial properties.

Animation, transformation and positioning of 3D figures during actions can be achieved through the Jack toolkit (Phillips and Badler, 1988), e.g. moving body and limbs or grasping objects. Maya and 3D Studio Max may be considered for the generation of 3D models. Maya also incorporates an editor for scripting characters as used in Su et al. (2007). The CSLU (Center for Spoken Language Understanding) toolkit (Sutton et al., 1998; CSLU, 2008), a universal speech toolkit, may support *SceneMaker* in text-to-speech synthesis and associated facial animation.

3.2 Project Development

A detailed review of existing affective computing and language processing applications will reveal the most suitable components for *SceneMaker*. A proposed architecture for *SceneMaker* is shown in Figure 1. Different theories on emotion and expression will be evaluated for viability. Test scenarios will be developed based on different genres and animation styles, e.g. plays by Samuel Beckett, which include precise descriptions of stage and props layout and cartoon animations, which employ techniques of exaggeration for expression. The functionality of *SceneMaker*'s components and their accessibility through the editor will be tested in cooperation with professional film directors, comparing the process of directing a scene traditionally with actors or with *SceneMaker*. The effectiveness of the scenes created in *SceneMaker* will be evaluated against hand-animated scenes. A proposed research schedule is given in Table A.1 in Appendix A.

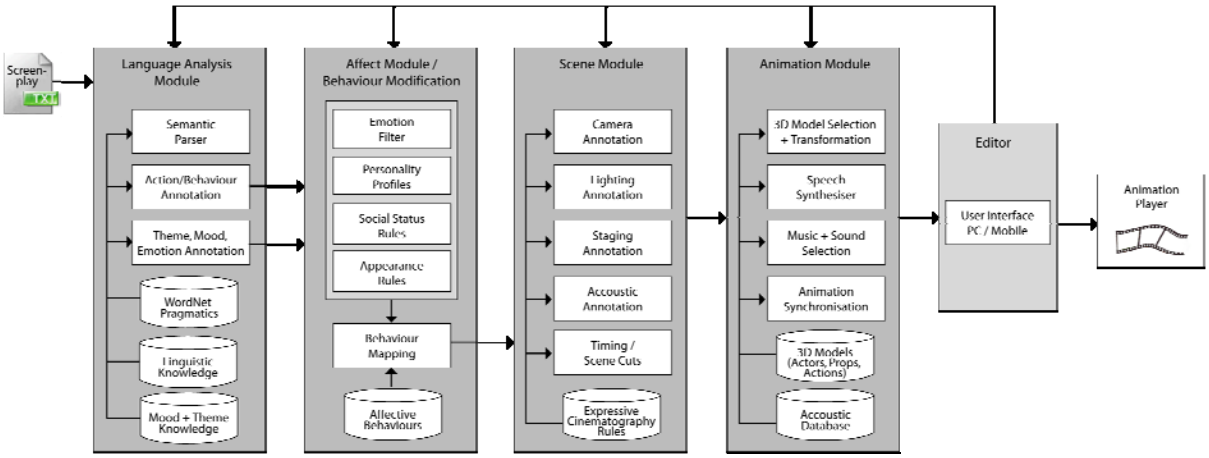

Figure 1: *SceneMaker* architecture

4 Comparison to Other Work

Research implementing various aspects of modeling affective virtual actors, narrative systems and film-making applications is compared to the proposed *SceneMaker* prototype. Table B.1 in Appendix B gives a detailed overview of these related systems. No previous system controls agent behaviour through integrating all of personality, social status, narrative roles and emotions. SCREAM realises most of these aspects, but narrative roles are missing and the input is a specifically scripted language. Only EML combines multimodal character animation with film making practices based on an emotional model, but it lacks consideration of personality types or social roles. *SceneMaker* will bring all relevant techniques together to form a software prototype system for animation production from natural language scene scripts. *SceneMaker* will present a unique user interface for directors or animators to control and adjust the scene production. *SceneMaker's* GUI will facilitate easy access from mobile devices and rapid scene testing on set.

5 Conclusion

This research aims to advance knowledge in the areas of affective computing, digital storytelling and expressive multimodal systems through the development of the software prototype, *SceneMaker*, which automatically visualises affective expressions of screenplays. *SceneMaker's* mobile, web-based user interface will assist directors, drama students, writers and animators in the testing of their ideas. Thus *SceneMaker* will considerably shorten the production time and reduce production costs. Existing systems solving partial aspects of natural language processing, emotion modeling and multimodal storytelling have been discussed. Potential software and tools relevant for the implementation of *SceneMaker* have been investigated. In comparison to similar projects, *SceneMaker* will incorporate an expressive model for multiple modalities, including prosody, facial expressions, gestures, body posture, acoustics, illumination, staging and camera work. *SceneMaker* will also comprise a user interface for manual adjustments. Accuracy of content animation, effectiveness of expression and usability of the interface will be evaluated in empirical tests.

References

Breazeal, C., Brooks, A., Gray, J., Hancher, M., McBean, J., Stiehl, W.D., and Strickon, J. (2003). "Interactive Robot Theatre". In *Communications.* ACM New York, NY. 46 (7), 76-85.

Breitfuss, W., Prendinger, H., and Ishizuka, M. (2007). "Automated generation of non-verbal behavior for virtual embodied characters". In *Proceedings of the 9th international Conference on Multimodal interfaces*. ICMI '07. ACM, New York, NY, 319-322.

Bickmore, T. W. (2004). "Unspoken rules of spoken interaction". In *Communications.* ACM New York, NY. 47 (4), 38-44.

Cassell, J., Vilhjálmsson, H. H., and Bickmore, T. (2001). "BEAT: the Behavior Expression Animation Toolkit". In *Proceedings of the 28th Annual Conference on Computer Graphics and interactive Techniques* SIGGRAPH '01. ACM, New York, NY, 477-486.

Cavazza, M., Lugrin, J., Pizzi, D., and Charles, F. (2007). "Madame bovary on the holodeck: immersive interactive storytelling". In *Proceedings of the 15th international Conference on Multimedia*. MULTIMEDIA '07. ACM, New York, NY, 651-660.

Connexor (2003). http://www.connexor.eu/technology/machinese (accessed December, 2008).

Coyne, B. and Sproat, R. (2001). "WordsEye: an automatic text-to-scene conversion system". *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, Los Angeles. 487-496.

CSLU (2008). Available at: http://www.cslu.ogi.edu/toolkit (accessed December, 2008).

De Melo, C. and Paiva, A. (2006). "Multimodal Expression in Virtual Humans". In *Computer Animation and Virtual Worlds 2006*. John Wiley & Sons Ltd. 17 (3-4), 239-348.

De Raad, B., (2000). "The Big Five Personality Factors". In *The Psycholexical Approach to Personality*. Hogrefe & Huber.

Disney Animation, (2008). http://en.wikipedia.org/wiki/12_basic_principles_of_animation (accessed November, 2008).

Ekman, P. and Rosenberg E. L. (1997). "What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system". *Oxford University Press.*

Fellbaum, C. (1998) "WordNet: An Electronic Lexical Database". MIT Press. Cambridge, MA.

Gebhard, P. (2005). "ALMA - Layered Model of Affect". In *Proceedings of the Fourth International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 05). Utrecht University, Netherlands. ACM, New York, NY. 29-36.

Gebhard, P., Kipp, M., Klesen, M., and Rist, T. (2003). "Authoring scenes for adaptive, interactive performances". In *Proceedings of the Second international Joint Conference on Autonomous Agents and Multiagent Systems*. AAMAS '03. ACM, New York, NY, 725-732.

Gebhard, P. and Schröder, M. (2008) "IDEAS4Games – A.I. Poker im Casino Virtuell". DFKI Newsletter 1/2008.

Gunes, H. and Piccardi, M. (2006).

"A Bimodal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior". In *18th International Conference on Pattern Recognition, 2006*. ICPR. IEEE Computer Society, Washington, DC. 1, 1148-1153.

H-Anim (2001).
Humanoid animation working group. http://www.h-anim.org (accessed December, 2008).

Kennedy, K. and Mercer, R. E. (2002).
"Planning animation cinematography and shot structure to communicate theme and mood". In *Proceedings of the 2nd international Symposium on Smart Graphics*. SMARTGRAPH '02. ACM, New York, NY. 24, 1-8.

Kopp, S., Allwood, J., Grammer, K., Ahlsen, E. and Stocksmeier, T. (2008).
"Modeling Embodied Feedback with Virtual Humans". In *Modeling Communication with Robots and Virtual Humans*. Springer Berlin/Heidelberg. 18-37.

LCS (2000)
Lexical Conceptual Structure Database.
http://www.umiacs.umd.edu/~bonnie/LCS_Database_Documentation.html (accessed December, 2008).

Liikkanen, L. A., Jacucci, G., Huvio, E., Laitinen, T., and Andre, E. (2008).
"Exploring emotions and multimodality in digitally augmented puppeteering". In *Proceedings of the Working Conference on Advanced Visual interfaces*. AVI '08. ACM, New York, NY, 339-342.

Ma, M. (2006).
"Automatic Conversion of Natural Language to 3D Animation". *PhD Thesis, School of Computing and Intelligent Systems, University of Ulster*.

McDonnell, R., Jörg, S., McHugh, J., Newell, F., and O'Sullivan, C. (2008).
"Evaluating the emotional content of human motions on real and virtual characters". In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*. APGV '08. ACM, New York, NY, 67-74.

MS-Agents (2008).
http://msdn.microsoft.com/en-us/library/ms695784(VS.85).aspx (accessed December 2008).

Ortony A., Clore G. L., and Collins A. (1988).
"The Cognitive Structure of Emotions". Cambridge University Press, Cambridge, MA.

Paggio, P. and Music, B. (2001)
"Linguistic Interaction in Staging – a Language Engineering View". In *Virtual Interaction: Interaction in Virtual Inhabited 3D Worlds*. Springer, London, 235-249.

Pelachaud, C. (2005).
"Multimodal expressive embodied conversational agents". In *Proceedings of the 13th Annual ACM international Conference on Multimedia*. MULTIMEDIA '05. ACM, New York, NY, 683-689.

Perlin, K. and Goldberg, A. (1996).
"Improv: a system for scripting interactive actors in virtual worlds". In *Proceedings of the 23rd Annual Conference on Computer Graphics and interactive Techniques.* SIGGRAPH '96. ACM, New York, NY, 205-216.

Phillips, C. B. and Badler, N. I. (1988).
"JACK: a toolkit for manipulating articulated figures". In *Proceedings of the 1st Annual ACM SIGGRAPH Symposium on User interface Software*. UIST '88. ACM, New York, NY, 221-229.

Physically Informed Audio Synthesis (2008).
http://www.sarc.qub.ac.uk/main.php?page=projects&projID=48 (accessed December 2008).

Prendinger, H. and Ishizuka, M. (2002).
"SCREAM: scripting emotion-based agent minds". In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 1*. AAMAS '02. ACM, New York, NY, 350-351.

Rebelo, P., Alcorn, M. and Wilson, P. (2005).
“A Stethoscope for Imaginary Sound: Interactive Sound in a Health Care Environment”. In *International Computer Music Conference Proceedings*. ICMC 05, Barcelona.

RIVME (2004).
http://accad.osu.edu/~sgencogl/mocap/mocap.htm (accessed November, 2008).

Shaarani, A. S. and Romano, D. M. (2008).
“The intensity of perceived emotions in 3D virtual humans”. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*. International Conference on Autonomous Agents. International Foundation for Autonomous Agents and Multiagent Systems. Richland, SC. 3, 1261-1264.

Sowa, T. (2008).
“The Recognition and Comprehension of Hand Gestures - A Review and Research Agenda”. In *Modeling Communication with Robots and Virtual Humans*. Springer Berlin/Heidelberg. 38-56.

Su, W-P., Pham, B., Wardhani, A. (2007).
“Personality and Emotion-Based High-Level Control of Affective Story Characters”. In *IEEE Transactions on Visualization and Computer Graphics*, 13 (2), 281-293.

Sutton, S., Cole, R., De Villiers, J., Schalkwyk, J., Vermeulen, P., Macon, M., Yan, Y., Rundle, B., Shobaki, K., Hosom, P., Kain, A., Wouters, J., Massaro, D., Cohen, M. (1998).
"Universal speech tools: the CSLU toolkit", In *Proceedings of the International Conference on Spoken Language Processing* (ICSLP), Australia. Paper 0649, 3221-3224.

Tesniere, L. (1959).
Elements de syntaxe structurale. Klincksieck, Paris.

Thomas, F. and Johnson O. (1981, reprint 1997).
The Illusion of Life: Disney Animation. Abbeville Press/Hyperion. 47-69.

Thórisson, K. R., Pennock, C., List, T., and DiPirro, J. (2004).
“Artificial intelligence in computer graphics: a constructionist approach”. *SIGGRAPH Computer Graphics.* ACM, New York. 38 (1), 26-30.

Thórisson, K. R. (2007).
“Integrated A.I. systems”. In *Minds & Machines.* Springer Netherlands. 17 (1), 11-25.

Vilhjálmsson, H. and Thórisson, K.R, (2008).
“A Brief History of Function Representation from Gandalf to SAIBA”. In *Proceedings of the 1st Function Markup Language Workshop* at AAMAS, Portugal. 61-64.

Virtual Theatre (2004).
http://accad.osu.edu/research/virtual_environment_htmls/virtual_theatre.htm (accessed November, 2008).

Wahlster, W. (2006).
“Smartkom : Foundations of Multimodal Dialogue Systems”. Springer Verlag.

Appendix A – Research Schedule

Research Activities	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
Perform Literature Review												
100 Day Review and Presentation												
Submission to ISEA2009												
Investigation on User Requirements, Interviews with Actors/Directors												
Submission to MobileHCI 2009 Conference												
In-Depth Review of Systems and Approaches Relevant for Integration into SceneMaker												
Submission to ICMI-MLMI 2009 Conference												
Submission to AICS 2009 Conference												
Confirmation												
Design Automated Scene Production System												
Implementation of Automated Scene Production System												
2nd Year Poster												
Implementation of SceneMaker GUI in Accordance with HCI Guidelines												
Submission to IEEE Pervasive Computing Journal												
Testing and Evaluation												
Submission to ACM Transactions on Multimedia Computing, Communications and Applications												
3rd year presentation												
Thesis write up												

Table A.1: Research Schedule

Appendix B – Comparison to Other Work

System	Year	Category	Input (Perception)					Output (Generation)								Device					Emotions						3D Animation		Author Options		Story Type		Other Comments
			Natural Language Text	Scripting Language	Speech	Gesture/Posture	Mouse/Keyboard/Pen	Text	Speech	Gaze	Facial Expression	Gesture/Posture	Music/Sound	Camera	Lighting	Computer Screen	Mobile Devices	Robot	Mixed Reality	Internet/Network	Personality	Social Roles	Narrative Rolls	Temporal Distinction	Basic Emotions	OCC/PAD	One Agent	Multiple Agents	User Interface	Code/Rules Extendable	Predefined Storyline	Dynamic/User Interaction	
WordsEye	01	Text to Visual		✓								✓		✓	✓	✓				✓							*					*static scenes, not animated
BEAT/SPARK	01		✓						✓	✓	✓	✓				✓				✓						✓			✓	✓		written in Java, XML based, no transition between/interruption of gestures
High Level Control	07		*									✓				✓					✓		✓	✓		**	***			✓			*Scene descriptions **intensity value only *** Maya
Behaviour Generation System	07		✓						✓	✓		✓				✓					✓							✓			✓		written in Java, XML, creates MPML-3D files
SCREAM	02	Scripting Tool		*					✓			✓				✓				✓	✓	✓		✓		**	✓			✓		✓	written in Java, Prolog, MPML *Communicative Acts **intensity value only
Gandalf, REA	96	ECA			✓	✓			✓	✓		✓							✓								✓					✓	
Microsoft Agents	98		✓		✓		✓	✓	✓	*	*	*		*	*	✓				✓	✓	*						✓		✓		✓	*custom animations can be modeled by an animator
SmartKom	04				✓	✓	✓		✓							✓	✓										✓					✓	
The Virtual Human Project/ALMA	05		✓		✓		✓		✓		✓	✓				✓					✓			✓				✓		✓		✓	
Greta/APML	05				✓	*			✓	✓	✓	✓				✓						*			**		✓					✓	*statistical model representing social context **Expressivity through manner of gesture e.g. strength, temporal extent
Jack	88	Virtual Human					✓									✓												✓					
Improv	96			✓	✓				✓		✓	✓	✓			✓				✓	✓							✓				✓	Unix, not intelligent: output as defined by author
Max	08		✓						✓	✓	✓	✓				✓					*						✓					✓	*default probability values for action determine character

System	Year	Category	Input (Perception)					Output (Generation)								Device					Emotions						3D Animation		Author Options		Story Type		Other Comments
			Natural Language Text	Scripting Language	Speech	Gesture/Posture	Mouse/Keyboard/Pen	Text	Speech	Gaze	Facial Expression	Gesture/Posture	Music/Sound	Camera	Lighting	Computer Screen	Mobile Devices	Robot	Mixed Reality	Internet/Network	Personality	Social Roles	Narrative Rolls	Temporal Distinction	Basic Emotions	OCC/PAD	One Agent	Multiple Agents	User Interface	Code/Rules Extendable	Predefined Storyline	Dynamic/User Interaction	
Casino Virtuell/ Cross Talk	08	Virtual Human		✓			✓		✓		✓	✓				✓			✓		✓	✓		✓			✓		✓	✓	✓	
Robot Theatre	03	Interactive Theatre				✓						✓	✓		✓			✓	✓			✓									✓	✓	
Madame Bovary	07				✓	✓			✓		✓	✓							✓			✓				✓					✓	✓	
PuppetWall	08				✓	✓	✓									✓									✓							✓	
Annotated Databases	06	Emotion Data				✓					✓	✓				✓				✓					✓	✓							*input already annotated in APML file
Mood Cinematography	02	Multimodal Storytelling		✓										✓	✓	✓									*		✓		✓				*moods and themes not specified/ determined by animator
Virtual Theatre Interface	04						✓							✓	✓	✓				✓							✓		✓				
EML	06								✓		✓	✓*	✓	✓	✓	✓									✓	✓	✓				✓		*hand and arm gestures only
CONFUCIUS	06		✓						✓			✓	✓	✓		✓											✓				✓		
SceneMaker	08		✓						✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓	✓	✓	▪	✓	✓	✓		✓			

Table B.1: Comparison of Affective and Multimodal Systems and Agents