SALERO: SEMANTIC AUDIOVISUAL ENTERTAINMENT REUSABLE OBJECTS

G. Thallinger, G. Kienast, O. Mayor, C. Cullen, R. Hackett, J. Jose,

JOANNEUM RESEARCH Forschungsgesellschaft mbH

ABSTRACT

Broadcasters around the world are in desperate need to automate content production as much as possible. This need is twofold on the one side automatic production for good structured program parts is needed on the other hand production for different target devices is an issue. The EC project SALERO has developed a range of tools enabling automatic template based production of animation clips with virtual presenters over the past years.

In this paper we describe the workflow devised by the project to automate major parts of media production based on 3D content. This is accompanied by a description by the individual tools developed and examples from the experimental productions implemented with this tools.

INTRODUCTION

The SALERO[1] project set out to implement the notion of intelligent content. The concept behind intelligent content is that it is self-descriptive and as such might be automatically processed and adapted to various uses. In order to implement
the and the tools how animated clips are automatically generated based on so called program templates. In the process a template editor is used to define the overall structure of a clip. The template may contain overall structural elements and defines parts of the clip where information is inserted from external sources to influence the rendering of the clip. External information could be the script for the virtual presenter, emotional information for text-to-speech and face animation, sex and age of the presenter used to synthesize the voice accordingly and any further information e.g. props for the presenter or the studio. A rendering engine is used to generate versions for the different target devices as needed.

progress has been made in the area of intelligent media, including the development of advanced algorithms for facial animation, reusable multilingual speech analysis for character animation, expressive speech synthesis, audio transformation, improved tools for game development, establishment of an ontology for modelling virtual characters and the implementation of the second phase experimental productions.
Research in facial animation yielded techniques for emotional animation based on the Maskle approach and using the circumplex model to specify emotions. This is

[1] The research leading to this paper was partially supported by the European Union under contract FP6-027122 (SALERO). More information can be found on the project's website: www.salero.eu.

combined with work on highly realistic modelling and animation of the appearance of persons depending on medical state.
Speech analysis work has produced a method that can be reused across characters, productions and in any input language.
In respect to expressive speech synthesis a limited domain engine has been implemented. The work in respect to audio transformation has yielded different algorithms to alter the quality of a voice (e.g. converting a female voice to a male one, changing the perceived age and gender of the speaker).
An ontology to describe virtual characters for 3D animation and a multimedia annotation and retrieval tool have been developed.

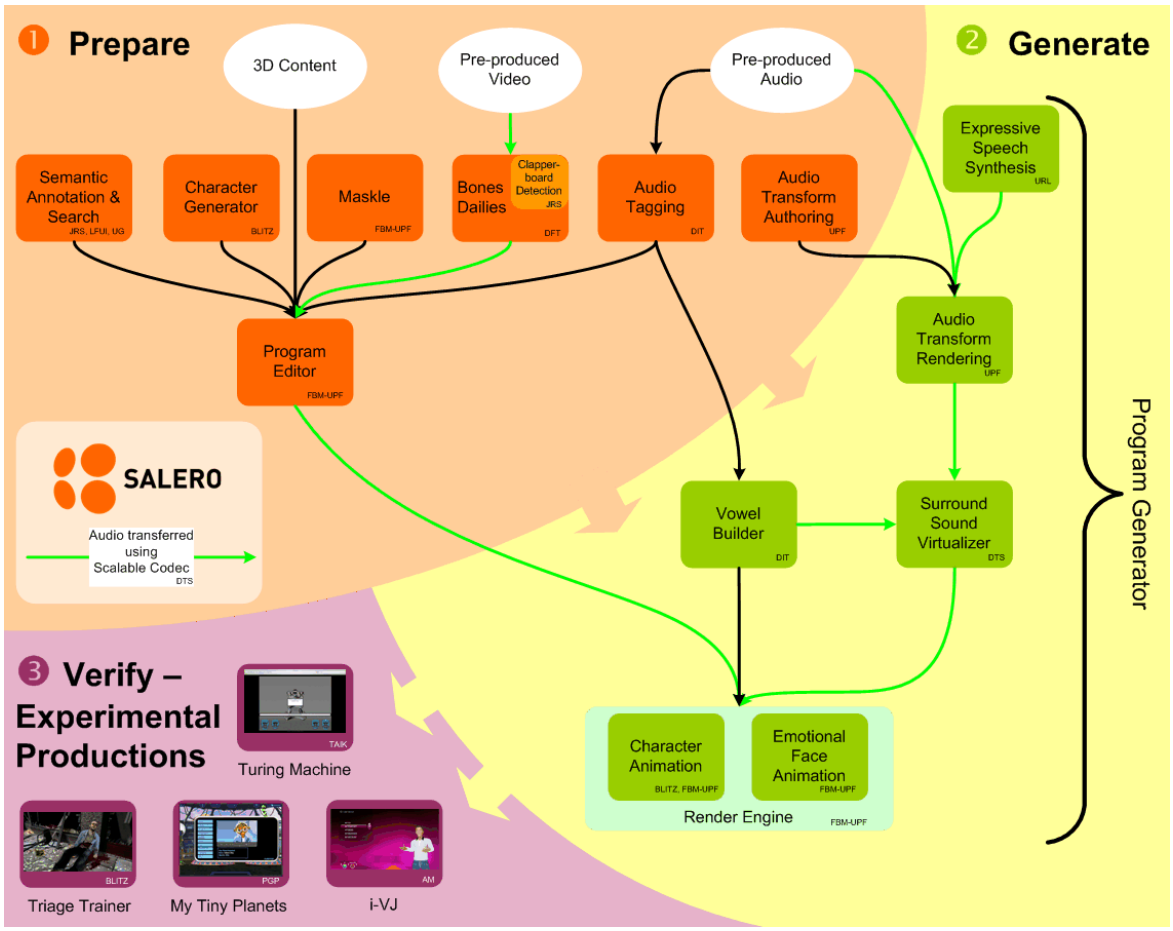

Figure 1 – SALERO Workflow

With *Triage Trainer, Turing Machine and Turing Bot, MyTinyPlanets.com* and *Interactive Video Jockey* four experimental productions were implemented and subsequently evaluated.

INTELLIGENT MEDIA RESEARCH

Media Semantics

The SALERO project researched methods and tools to formally capture and exploit the semantics of media objects in media productions. This is enabled through a semantic workbench which provides lifecycle support for multimedia ontologies. This includes tools to lift existing multimedia metadata to ontologies, to engineer ontologies, tools to annotate media objects using these ontologies and facilities to exploit the annotations in semantic search.

Parts of the semantic workbench are provided as Web Services with the intention to include functionalities for annotation and semantic search into other applications. The workbench is accompanied with a set of ontologies, among those the SALERO Virtual Character Ontology supporting formal description of virtual characters including geometry, animations, and behaviour.

Context-Based Search & Retrieval

The *AspectBrowser* interface was created which allows users to organise their search process by creating aspects, where each aspect represents a single narrow search need within the context of a larger information need. The interface was designed to provide an integrated environment for the searching of images, videos, and the web. Locally indexed image and video collections can be searched using our own custom search backend, which allows search by example using a range of low-level visual features. Web resources (e.g. Flickr, YouTube) can be searched using their web APIs. The AspectBrowser interface also allows the user to take advantage of information found in one collection to search other collections.

PROCEDURAL FACE GENERATION

Blitz Games Studios has built upon previous work in improved animation techniques and procedural generation of head and eye movement to develop an infinitely flexible tool for the rapid creation of realistic people. Work has focused on the face and head and the construction of a deformable mesh to enable the automatic conformation of multiple ethnicities, ages and weights as well as the ability to display emotion.
The system enables additive morph blend offsets on top of existing meshes, which supports blending from one base mesh to up to three ethnic variations of this mesh while retaining features from all of the morphs. The animation bones in the rig are set to fixed points in the base mesh and then procedurally altered with numerical offsets generated by the blends on the base mesh. Animations are additively blended at the root of the animation tree so that any further animations can be blended afterwards while retaining the desired offsets. This way, animations created for the original head are just as valid for any other generated heads.

SPEECH ANALYSIS AND REUSABLE MULTILINGUAL CONTENT

Analysis methods developed in DIT provide means to author the animation for a given speech file in any input language. This speech file can then be used to drive a specific character (or multiple different characters) across productions.
By developing a single authoring workflow, animators can concentrate on character development irrespective of the input speech files that will eventually drive it. Prototypes for online content delivery have been developed in Flash and Flex and are currently being developed for dialog animation.

SPEECH SYNTHESIS & AUDIO PROCESSING

Speech synthesis research by Ramon Llull University was oriented to improve the expressiveness and the quality when the text input does not belong to a limited domain (e.g. weather forecast). In this sense, a new speech database in English was recorded (by Pepper's Ghost Productions), segmented and tagged. The database was initially designed from texts related to the MyTinyPlanets experimental production, but some general texts were added in order to improve

the phonetic coverage. The final goal is to use this voice with the voice transformation tools in different experimental productions with more open domains than MyTinyPlanets.
More precisely, the current version of the text-to-speech tool is corpus-based and it incorporates some improvements related to the unit selection function. Ramon Llull University also worked in other research topics:

- Analysis and synthesis of Voice Quality parameters
- Prosody estimation of expressive speech
- Hidden Markov Models (HMM) based speech synthesis

INTEGRATED APPLICATIONS

Program Editor and Renderer

The *Program Editor* is a tool for automatic generation of audiovisual pieces. Given a set of assets as 3D meshes, virtual characters, animations, virtual sets, etc. it is possible to generate an animated sequence.

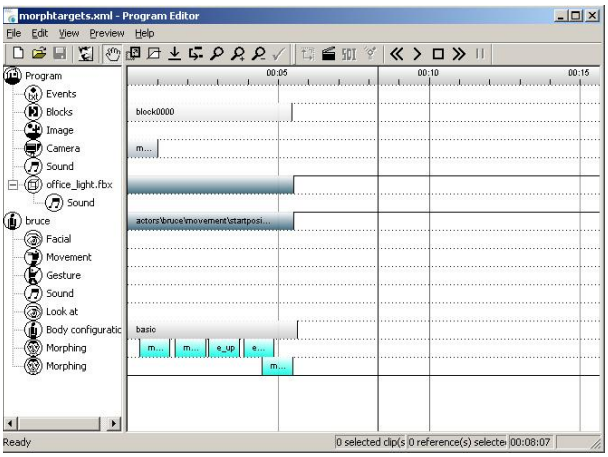

Figure 2 – Timeline editing interface

The sequences can be edited by hand, through the editing timeline or by preparing templates for automatic generation. Given the production requirements, it is possible to create templates that are used to automate the process of creating several videos with dynamic contents.
Once the sequence has been edited or the template for automatic generation is ready, it is possible to render the final output video, thanks to an OpenGL based render engine developed using the GTI-Framework. This render engine uses state-of-the-art algorithms for dynamic lighting and shadowing, complex shaders and post-processing.
Latest enhancements to the Program Editor include improvements to the interface and interaction flow. The results of the research on emotional animation can be accessed thanks to an interface that gives direct access to emotional states.

Bones Dailies

The Bones framework is a next-generation environment for digital intermediate and post production that lets users create workflows tailored to their needs. It serves as the heart of an end-to-end post-production workflow and manages the process from data ingest to finished digital master.
Bones Dailies is a fast and efficient workflow tool for

Figure 3 – Bones Dailies User interface for image manipulation and timeline processing

producing sound synchronised and colour corrected dailies (rushes). The workflow provides a benefit for the entire production and post production chain.
Bones Dailies allocates advanced features for audio and image ingest, metadata capture, colour grading, Audio/Video synchronization and play-out of dailies deliverables. The framework interfaces directly with postproduction products such as film scanners and other high end capture devices. The open environment supports third-party applications as well as render-clusters to deliver optimum rendering times tailored to specific project requirements. Using Bones applications, users can edit and conform transitions, rescale images and convert clips to a variety of output formats (e.g. 4k film, HDTV and SDTV). Where third-party applications are desired, the environment manages the data transfer of images to and from workstations in a facility and collects them for conformance into a completed sequence.
Bones Dailes is used as a platform for the Experimental Productions.

Scalable Codec

Based on the validated approach and framework, the *Scalable Audio Codec* was further developed to accommodate both multi-channel (up to five primary channels) and extended bit-rate range (24-96 kbps/channel). This was delivered as the second version deliverable to the project during the year. Current work continues to develop GUI-based tools and to integrate the compression technology to partners' production tools and work flow. The work at this stage aims at the applications of developed scalable codec in the third stage experimental productions. Meanwhile, parallel work is carried out to develop and enhance a multi-channel to two-channel audio spatial virtualization to support the concept of media intelligence and reusability from the play-back device perspective.

Audio Transformation Tools

The Audio Transformation Tools developed by the Music Technology Group of the Pompeu Fabra University are a set of cross-platform VST plug-ins and offline rendering tools including *Voice Transformation*, *Tempo & Pitch Audio Transformation* and *Advanced Audio Equalizer*.
Voice Transformation allows several singing and speech voice transformations using spectral techniques for modifying the character of the voice. Transformations include: transposition, quantization, vibrato, roughness, breathiness, whisper, timbre mapping and other spectral transformations. This tool is targeted to recording or post-production studios that require voice processing.
Tempo & Pitch Audio Transformation allows to time scale as well as pitch transpose polyphonic audio. This is especially useful to change duration of music to fit video length maintaining the quality of the original audio or transpose pitch in background music of games or film material with high quality.
Advanced Audio Equalizer is aimed for the advanced equalization of commercial music productions. This is a powerful tool able to equalize the audio using energy histogram-based on different criteria. It is not limited to the typical energy distribution versus frequency, but allows working with other useful energy distribution criteria such as panning, inter-channel phase difference, or energy variance. Applications include: Remixing a piece by changing the volume or panning of each instrument independently or removing instruments from the mix

and then re-adding them processed by some effects or isolate one or several instruments to perform a musical analysis of a piece.

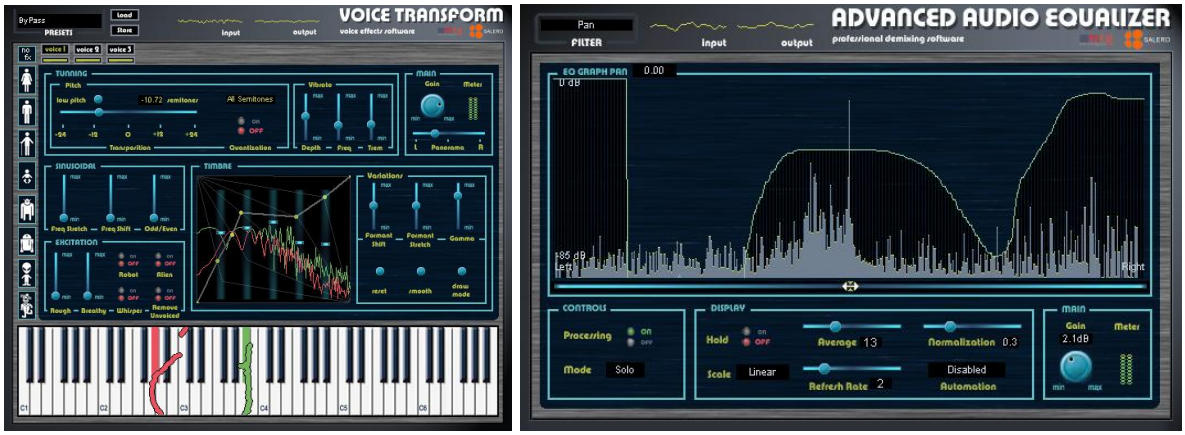

Figure 4 – Audio Transformation Authoring Tools

Game Development Tools

Work on the component based editing interface for shaders has been performed and the interface has been leveraged to create new tools for visually editing both animation blending trees and finite state machines. The state machine system has been targeted at core game logic control, player cause-effect scripting, artificial intelligence decision making and cut-scene animation sequencing when used in conjunction with the animation blending tree editor.
The bi-directional real time network link between game engine and editing interface has been developed to provide rapid and accurate change feedback on the target platform for edits and diagnostic feedback in the editing environment for runtime data state changes.

EXPERIMENTAL PRODUCTIONS

Triage Trainer

Figure 5 – The Triage Trainer application

Triage Trainer supports the development of critical decision making in the life saving skill of Triage. This is the process of prioritising casualties for treatment after a major incident. The training game is set at the scene of an explosion in a high street. The multiple casualties look and behave realistically to make the experience as near real life as possible, which supports training. Trainees' decisions are tracked through the game, when they have assessed all casualties they are given feedback on their performance, compared to the protocol laid down by Advanced Life Support Group, which advised NATO on triage.
The SALERO tools and methodologies used in the production were:

- Improved facial animation processes, animation rigging, procedural generation of eye movement, respiration, blood flow, sweating, pallor and flushing
- Distributed Asset processing tool & re-usable asset processes
- Shader Editor

Trainees using the game in controlled trials performed significantly better in a common assessment exercise than those who had used a tabletop exercise during training. The improved performance may reasonably be ascribed to the immersion achieved with realistic human figures.

Turing Machine

Crucible Studio produced the Turing Machine cross-media theatre and online production in collaboration with Helsinki Skaala Opera. Crucible Studio produced 3D-scenography for the opera performance and a related web-based module, which further explores the life of Alan Turing. The web site that features Alan Turing as an enigmatic artificial chat bot character, Turing Enigma, works by itself or as part of the Turing Machine Opera.

The production investigated repurposing media objects between the two genres, between physical theatre space and an online site. The same objects that appear in on-stage projections also pop out from the Turing Enigma machine.

Tools research found the SALERO emotional framework useful for script-writing and reviewed and improved upon the use of facial animation tools. Creating a linear performance and a non-linear artificial chat bot character from the Alan Turing character was also Crucible Studio's artistic and design research project on storytelling, in which the chat bot as an intelligent media production application repurposed automated dialogue technologies to narrative media.

Figure 6 – Scene from Turing Machine Opera with Merja Nieminen's 3D-scenography and screenshot of Turing Enigma Chat Bot with animated masks

MyTinyPlanets.com

Peppers Ghost Productions produced animation tests for the My Tiny Planets browser based virtual world with characters animated by both Dublin Institute of Technology speech analysis methods and traditional toolsets, and evaluated the integration of these into a variety of on-line delivery mechanisms.

Having established an initial test group of several thousand users, PGP was better able to understand behaviours and requirements of players in virtual worlds, and revise development plans accordingly. It became apparent that animated sequences as a support for game-play, narrative and immersion represent a significant commercial advantage, and will be pursued aggressively during the next phase of development and experimental production.

Figure 7 – Screenshot from MyTinyPlanets.co.uk

The approach of allocating movements and gestures using the DIT speech analysis method proved to fit well as a layer over the top of standard animation procedures such as 'ambient' motion capture files that prevent stilted movements.

Interactive VJ

Activa Multimedia has developed an experimental prototype called an interactive Video Jockey (i-VJ). The production is a digital television broadcast over IPTV. This is in line with two growing trends in digital content, making the TV experience more interactive and allowing for customization.

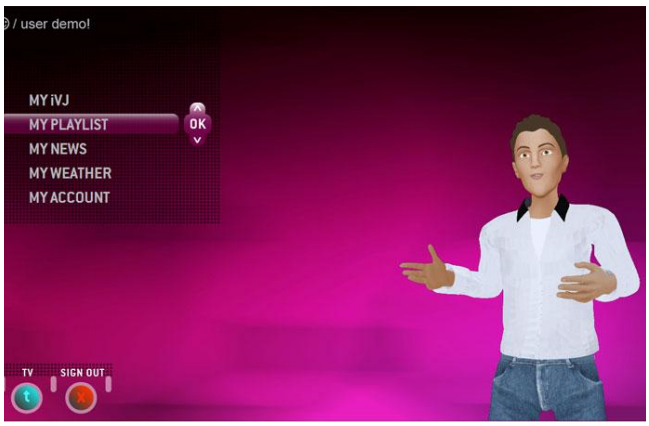

Figure 8 – i-VJ presenting the different services of the application

The main objectives of this experimental production were to explore:

- Customization: users can customize distribution lists and their virtual presenter.
- Interactivity: users can see what they want, when they want, and how they want.
- Automation: control systems that automatically generate content and minimize human intervention.
- Re-use: maximizing the re-use of assets, particularly 3D models and animations.
- Compatibility with different ways of generating audio: a low-cost, highly automated audiovisual production must be compatible with various video sources, ranging from pre-recording to voice transformation and text-to-speech. The efficient use of audio sources crucial is for cost reduction, the re-use of assets and automation of production.
- Automatic and semi-random script modification.

This experimental production is designed as a 24-hour music IPTV channel, a virtual Video Jockey as a 24/7 automatic broadcast and video-on-demand service.

CONCLUSIONS

Based on the promising results to date the focus of work for the final project year is to further integrate the technologies, e.g. a search engine combining textual, semantic and content based search technology. Sets of interoperable modules build the thorough base for the implementation and evaluation of the 3rd phase experimental productions.

REFERENCES

1. A. Evans, M. Romeo, M. Dematei, J. Blat, The Maskle: Automatic Weighting for Facial Animation – An automated approach the problem of facial weighting for animation. International Conference on Computer Graphics Theory and Applications. February 2009, Lisbon, Portugal.

2. T. Bürger, P. Hofmair, G. Kienast, The SALERO Virtual Character Ontology, Workshop on Semantic 3D Media at Third International Conference on Semantic and Digital Media Technologies, December 2008, Koblenz, Germany.

3. O. Mayor, J. Bonada, J. Janer, Voice Transformation from Interactive Installations to Video-Games, AES 35th International Conference: Audio for Games. February 2009, London, United Kingdom.

4. R. Villa, I. Cantador, H. Joho, J. Jose. An Aspectual Interface for Supporting Complex Search Tasks, The 32nd Annual ACM SIGIR Conference, July 2009, Boston, MA (in press)