



SALERO

Second Annual Online Public Report

SALERO Deliverable 10.5.5



Second Annual Online Public Report

SALERO Deliverable D10.5.5

SALERO identifier: SALERO-JRS-D10.5.5-PublicAnnualReport-v05.doc

Deliverable number: D10.5.5

Author(s) and company: G. Kienast, G. Thallinger (JRS), G. Holmberg (UPF);
M. Yan (DTS); K. Åberg (TAIK); C. Goodman (PGP),
T. Bürger (LFUI); R. Villa (UG); R. Hackett (BLITZ);
F. Alias (URL), J. Blat (FBM-UPF), R. Fach (GVG)

Work package / task: WP10

Document status: Final

Confidentiality: Public

Version	Date	Reason of change
1	2007-10-30	document created, call for input from partners
2	2007-11-14	Interim version, call for further input
3	2007-11-23	Version for internal review by DTS
4	2007-11-26	Comments from internal review
5	2007-11-27	Final version

The work presented in this document was partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th framework programme for R&D.

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

This document contains material, which is the copyright of certain SALERO consortium parties, and may not be reproduced or copied without permission. All SALERO consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the SALERO consortium as a whole, nor a certain party of the SALERO consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.

Table of Contents

1 Project Overview	1
2 Summary of Activities.....	1
3 First Phase Experimental Productions	2
3.1 MyTinyPlanets.com	2
3.2 Full House.....	3
3.3 Triage Trainer	4
3.4 Hack the Van	5
4 Emotional Facial Animation	5
5 Speech Synthesis and Audio Processing	6
6 Intelligent Media Research.....	7
6.1 Media Semantics	7
6.2 Context Based Search & Retrieval.....	7
6.3 Genre Research	8
7 User Involvement, Promotion and Awareness.....	8
7.1 User Group	8
7.2 Open House Day in Helsinki.....	8
7.3 Summer School on Media Semantics	8
7.4 Conference Presentations	8
7.5 Other Events.....	9
8 Conclusions and Future Work	9
9 Further Information	9
9.1 Useful Weblinks.....	9
9.2 Contact	10

1 Project Overview



<http://www.salero.eu>

SALERO aims to make cross media-production for games, movies and broadcast faster, better and more cost efficient by combining computer graphics, language technology, semantic web technologies as well as content based search and retrieval.

SALERO will define and develop 'intelligent content' for media production, consisting of multimedia objects with context-aware behaviours for self-adaptive use and delivery across different platforms. 'Intelligent Content' should enable the creation and re-use of complex, compelling media by artists who need to know little about the technical aspects of the tools that they use.

Based on research into methodologies for describing, creating and finding intelligent content, SALERO will develop toolsets to create, manage, edit, retrieve and deliver content objects, addressing characters, objects, sounds, language sets, and behaviours. The toolsets developed and the concept of intelligent content will be verified by experimental productions.

2 Summary of Activities

In the second year of the project major progress has been made in all areas. Examples are the development of advanced algorithms for facial animation, expressive speech synthesis, establishment of an ontology for media production and the implementation of the first phase experimental productions.

Research in facial animation yielded techniques for emotional animation based on the maskle approach and using the circumflex model to specify emotions. This is combined with work on highly realistic modelling and animation of the appearance of persons depending on medical state.

In respect to speech and audio research, results are twofold. After setting up facilities for recording of duologues (separate recording of each voice), recording of a duologue corpus with induced emotion has started. 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 of the speaker).

A first version of an ontology to describe the universe of animation based media production was defined based on the input by the user partners.

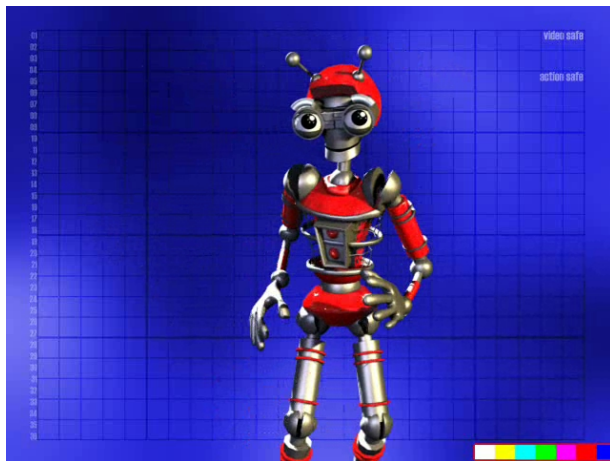
The first phase experimental productions were implemented and subsequently evaluated. Based on the experiences during the implementation of the experimental productions and the results from the evaluation, the specifications for the second phase experimental productions and the updated user requirements were established.

The intermediate results of the research activities and the results from the experimental productions are a perfect starting point for the next experimental productions and the integration steps planned for the next project year.

3 First Phase Experimental Productions

3.1 MyTinyPlanets.com

Using timing and lip-synchronisation a series of test sequences featuring a synthetic character called 'Robbie', a prototype helper avatar for the My Tiny Planets experimental production were produced. This methodology was found to have far greater flexibility and long-term potential than traditional approaches. Spark and Socket were part of a pilot animation for television aimed at pre-school audiences, and were designed to allow for cost-effective but high quality animation sequences within a half-hour format show. The means by which this was achieved was to highly stylise the characters so that facial expression and speech were represented by gross movements of mechanical parts. By contrast Robbie was a far more humanoid character, and capable of more detailed expression (within the domain constraints of animation production). With the ultimate intention being to cost-effectively produce sufficient material to allow the system to respond to help requests from users in a wide variety of different contexts; it is very apparent that enhanced automation methodologies are required in order to provide sufficiently flexible and repurpose-able content to achieve this goal. From this experimental production we concluded that the approach offers positive advantages, especially in the area of tagging and identifying emotional content based on stress and activation. Pepper's Ghost Productions sees this as an important step towards creating animated templates which do not presuppose states of mind on the character which fall into the remit of secondary editorial input. During the next stage of experimental productions we intend to pursue this further using 'streamlined' characters for web and mobile device delivery, wherever possible using the My Tiny Planets site as a delivery and testing channel.



Robbie



Spark and Socket

3.2 Full House

Full House is a demo of an interactive political talk show for TV and cross media. The audience can affect the TV show through Internet and/or mobile connections, find information about election candidates, ask them questions etc. Both the candidates and viewers answer a questionnaire, after which they get an animated 3D Full House character, which corresponds best to their expressed political views. In the demo there are currently three politically archetypal characters: the Red, Green and Blue. These characters appear in all the platforms, including the TV talk show. Through such a character-based representation the system dramatises the answers of the candidates in comparison to the answers of the other candidates – and the audience.



Full House backend

3.3 Triage Trainer

Triage Trainer is a prototype video training game produced by TruSim, a division of Blitz Games Ltd, one of Europe's leading independent development studios. The game tests players' speed and accuracy in carrying out Triage Sieve, where they must identify casualties needing immediate or urgent medical attention at the scene of a major incident. Players must deal with highly realistic casualties, which show distress and display medically accurate symptoms validated by doctors. SALERO has supported the development of technology that enables rapid and cost effective development of high fidelity human models, which react and behave in a realistic manner, significantly improving the production pipeline and adding a new emotional dimension to game play.



Scene from the Triage Training Application

3.4 Hack the Van

This experimental production is a daily TV show based on music clips, with weather information as complementary section. It is aimed at young people. All the sections will be presented by virtual characters. The program will be automatically generated, so it will be produced without any operators involved.

The main character, Sefi talks and moves through randomized actions and pre-recorded voice sentences (to check if it is possible to randomize movements and facial gestures depending on the voice or on semantic tags).



The broadcasted music clips are chosen by viewers by SMS voting, and a robot called Kilo presents the clips, reading the titles from a data base, with a synthetic voice. This section is aimed to check if a synthetic voice is usable in a TV context, depending on the character that uses it and also on the script. Another aim is to test the real-time generation of voice and movements reading of a dynamic database.

This first prototype consists of a short demonstration program with the first scenes of the main presenters and the forecast section. The first prototype has addressed the following technical achievements:

- 100% automatically generated from the automatic animation platform.
- Based on an integration of several SALERO tools.
- Reuse of movements: Automatic adaptation of the same movement in all production characters, independently of the character proportions.
- Automatic random scripts: Every generation of the program changes the script → Enriched automation capabilities.

4 Emotional Facial Animation

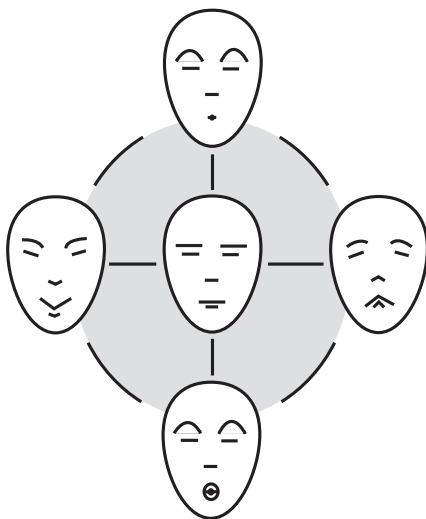
Based on empirical research on emotional facial expression, eye, mouth and muscle movement, as well as exploring the use of medical 3D scanners new approaches and technologies to rapidly produce a highly realistic, facial model, which responds in real-time were developed. The research has supported innovation and improved efficiency in production processes as well as developing new technologies for procedurally generating movement and appearance, which enable the rapid creation of new characters.

The model's movement is driven by animation 'bones', which can easily be adapted to any size or shape of face; while the surface appearance of pallor or flushing, sweating, blood flow, stress and

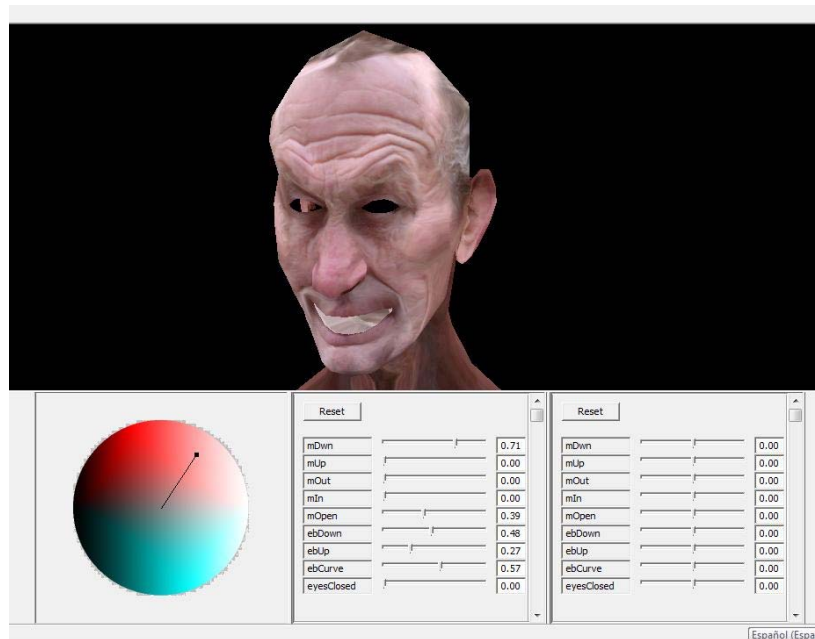
respiration is driven by a coded physiological model. Eye movement can be coded to respond to objects and people in the scene, giving the impression of a realistic response. This work has set new standards in the games industry and is already in use in our commercial productions.

Another strand of research is based on a model whereby almost all facial expressions could be represented by interpolation of a neutral facial pose with other four poses, which take into account the eyes and *mouth* zones. The model is illustrated in the following figure. The neutral pose is depicted in the middle, while the other four poses, which could be labelled as “extreme” – they do not correspond in fact to actual emotions – appear on the corners.

The metaphor for the interface of the prototype implementation has been inspired by two-dimensional colour palettes. The following screenshot illustrates the standalone prototype for the emotional animation, based on these ideas, with an example of an emotional face obtained by using it.



Poses used for interpolation



Standalone prototype

5 Speech Synthesis and Audio Processing

Reusability in compressed audio assets is also studied in the design of a scalable audio codec. This development will ultimately result in an audio compression tool for delivering audio productions from a single stored asset to media of varying data bandwidths. The first version of codec set was delivered in June 2007. By demonstrating a two-level performance-bandwidth scalability from single encoded file, it validated the underlying approach for codec scalability and established the framework for further developments.

The MTG Sound Annotation module contains two particularly innovative parts: the patent pending voice transformation system, and the patent pending sound annotation library able to automatically analyze and classify sounds and music.

The voice transformation system takes any voice signal as input (spoken or sung), performs analysis & transformations such as to change gender or age of speaker/singer. The system is able to perform high-quality transformations in real-time, or up to 8 times faster than real-time (batch-mode) to such a low computational cost that it would fit in a mobile device (PDA or mobile phone). For the invention of the voice transformation system Jordi Bonada was awarded the Spanish “Invention of the Year” award 2007 (primer premi Rosina Ribalta, Fundacio EPSON).

The computational efficiency and scalability/robustness across content databases has been a “showstopper” in commercialization of automatic multimedia content analysis for the past years. During the past year, MTG-UPF has been able to reduce computation cost for the sound annotation library at the same time as increasing accuracy, scalability and robustness. For this, several patents has been

filed. With regard to accuracy aspects alone, MTG-UPF won first and second prize respectively for automatic mood classification and cover detection in the international validation framework MIREX-2007 where the SALERO sound annotation library was used.

The URL team has focused on the following topics:

- Development and evaluation of a Multilingual (Spanish/English) meteorological-domain TTS. The first version of the Speech Synthesis software was delivered in July 2007, after being subjectively evaluated according to the standard MOS test.
- Expressive speech research, working on a Spanish corpus containing five speaking styles (happy, neutral, sad, aggressive and sensual). The corpus contents have been validated by means of automatic emotion identification, after mapping the subjective criteria to the automatic system.
- Prosody modelling for corpus-based and HMM-based synthesis, achieving the first Spanish HMM-based TTS system, which overcomes typical vocoder problems.
- Voice quality parameterizations for style separation and discrimination, applied to the previous mentioned expressive speech corpus.
- Automatic consensus of subjective user criteria (mapped into the cost function of the unit selection module) for enhancing the synthetic quality of corpus-based TTS synthesis.

6 Intelligent Media Research

6.1 Media Semantics

Media semantics may enhance different processes in the games and entertainment industries. A first step towards enhanced production workflows in SALERO was the studying of a typical 3D production workflow in order to determine how formally defined semantics and the use of ontologies may enhance this workflow in different phases of the production. A set of core focal points were identified with respect to that and finally decided to start with the definition of a core ontology for 3D characters in a media production which will provide an agreed view on animated characters between the project partners. An ontology engineering phase was initiated based on input by the user partners and a first core model for the 3D character ontology was built. Ontology engineering in SALERO will be approached with a close feedback loop to the domain experts, i.e. the user and content partners, to ensure the validity of the built model.

6.2 Context Based Search & Retrieval

Context Based Information Retrieval aims to research and develop practical methods that simplify the location and retrieval of characters, sounds, images, movements or behaviours from very large datasets and media storage systems. As a part of this effort, a number of different strands of work has been ongoing in the project, including:

- The development of efficient and effective searching of video databases.
- Research into the modelling of context within the search process, in particular the user's context.
- The development of web interfaces providing a rich environment in which a user can express complex, multi-faceted search needs.

Intelligent Media Objects are important to intelligent, context based search systems in a number of ways, enabling search systems to store and retrieve intelligent media objects. Ontologies may be used to formalise the specification of media objects, while the results of this work in turn can be used to develop ontology based search systems. Such systems can be integrated with more traditional Information Retrieval systems, which approach from the point of view of low-level data (text, image, and

audio processing), to develop the state of the art in identifying concepts and integrating them into search interfaces.

6.3 Genre Research

Media objects have different specification needs on different platforms, related to the overall expressive and stylistic objectives of the production. Audience expectations are related to the production genre, therefore artists and designers in the media industries are highly aware of the significance of style and genre in defining media products to meet the expectations of the end market.

The investigation of media types and genres is central to SALERO's attempts to create tools for industry professionals aiming to reach multiple audiences via multiple media.

- Genre features for character definition (e.g. figure proportions, character status) have been defined as search criterion for the SALERO ontology tool.
- Genre-based work-flows for media production have been described and used for the first experimental productions.
- Genre has been used as a main defining feature in experimental production descriptions as a tool for finding correlations between various types of productions.
- The most popular genres in social media and games have been explored and defined.

7 User Involvement, Promotion and Awareness

7.1 User Group

The user group has been revised and an active participation schedule has been set up for the group members' involvement with the project. The first SALERO user group meeting will take place in Barcelona in relation to the 3GSM mobile world congress, 11-14 February 2008.

7.2 Open House Day in Helsinki

The SALERO Open House event (13 June 2007, in Helsinki, Finland) was intended to give professionals and the general public an understanding of the technological development and productions within SALERO. The agenda included speeches on core SALERO topics and presentations of SALERO's experimental productions.

7.3 Summer School on Media Semantics

The Summer School on Multimedia Semantics took place 15-21 July 2007, at the University of Glasgow. A comprehensive program was designed to bring together leading researchers and practitioners, providing postgraduate students the opportunity to gain a deeper insight into the challenges related to multimedia semantics and the increasingly emerging applications relying on multimedia understanding.

7.4 Conference Presentations

Presentations of Research Papers by SALERO were presented at the following conferences:

- TRECVID 2007 – Evaluation Workshop (November 2007, Gaithersburg, United States)
- Semantic Authoring, Annotation and Knowledge Markup Workshop (October 2007, Whistler, Canada)
- I-MEDIA '07 – 1st International Conference on New Media Technology (September 2007, Graz, Austria)

- ISMIR 2007 – 8th International Conference on Music Information Retrieval
- ICPHS07 – International Congress of Phonetic Sciences (August 2007, Saarbrücken, Germany)
- CCCT 2007 – The 5th International Conference on Computing, Communications and Control Technologies (July 2007, Orlando, USA)
- IWANN 2007 – 9th International Work-Conference on Artificial Neural Networks (June 2007, San Sebastián, Spain)
- CBMI 2007 – International Workshop on Content-Based Multimedia Indexing (June 2007, Bordeaux, France)
- NOLISP 2007 – An ISCA Tutorial and Research Workshop on Non-Linear Speech Processing (May 2007, Paris, France)
- ICASSP'07 – International Conference on Acoustic, Speech, and Signal Processing (April 2007, Hawaii, USA)

7.5 Other Events

Further information about SALERO was distributed at the following events:

- Technical report published by “Centrum voor Wiskunde en Informatica, INS - Information Systems” (April 2007, Amsterdam, The Netherlands)
- UPF presented SALERO related work at the IBC 2007 in Amsterdam, The Netherlands.
- First results from the development work for the post production tool “Bones Dailies” has been shown at NAB 2007 in Las Vegas and IBC 2007 in Amsterdam, The Netherlands
- Aim@Shape workshop at the SAMT 2007 conference, December 2007, Genova, Italy

8 Conclusions and Future Work

With the work of the first two years as a base the project will on the one hand integrate and use the research results in the second phase experimental productions and on the other hand will further explore the open issues in emotional animation, speech synthesis, audio analysis and audio transformation as well as providing new and more advanced tools for the media production workflow.

The investigation of search techniques supporting the media production process encompassing and integrating traditional approaches as well as content based and semantic search will be of major importance for the rest of the project.

Several user group events to present the developments to date and to gather feedback from a bigger community will be organised throughout the next project year.

9 Further Information

9.1 Useful Weblinks

- <http://www.salero.eu>: SALERO project website including a showcase of SALERO's results
- <http://www.youtube.com/intelligentcontent>: SALERO's YouTube channel with results from the first experimental production prototypes
- <http://fullhouse.uiah.fi>: Experimental Production “Full House”

9.2 Contact

JOANNEUM RESEARCH Forschungsgesellschaft mbH
Institute of Information Systems & Information Management

Georg Thallinger

Steyrergasse 17
8010 Graz
AUSTRIA

Email: georg.thallinger@joanneum.at

phone: +43-316-876-1240

fax: +43-316-876-1191