

# SALERO - Semantic Audiovisual Entertainment Reusable Objects

Werner Haas, Georg Thallinger, Pedro Cano, Charlie Cullen and Tobias Bürger

**Abstract**— The Integrated Project SALERO aims to advance the state of the art in digital media to the point where it becomes possible to create audiovisual content for cross-platform delivery using intelligent content tools, with greater quality at lower cost, to provide audiences with more engaging entertainment and information at home or on the move. SALERO will build on and extend research in media technologies, web semantics and context based image retrieval, to reverse the trend toward ever-increasing cost of creating media.

**Index Terms**— audiovisual intelligent objects, content creation, context aware behaviour

## I. VISION & OBJECTIVES

SALERO's [1] overall vision is to define and develop 'intelligent content' for media production, consisting of multimedia objects with context-aware behaviour 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 of the technical aspects of how the tools that they use actually work.

- Complete realisation of SALERO's vision is a long-term goal. This gives rise to three overarching R&D objectives:
- Address characters, objects, sounds, language sets and behaviours,
- Research into methodologies for creating and finding intelligent content,
- Develop toolsets to create, manage, edit, retrieve and deliver content objects.

Manuscript received August 31, 2006. The R&D work carried out for the IP SALERO is partially funded under FP 6 of the European Commission within the IST Workprogramme 2004 (IST FP6-2004-027122).

W. Haas, G. Thallinger, are with JOANNEUM RESEARCH, Graz, Austria (phone: +43 316 876 1119; e-mail: firstname.lastname@joanneum.at).

P. Cano is with Universitat Pompeu Fabra, Barcelona, Spain (e-mail: pcano@iua.upf.edu).

Ch. Cullen is with Dublin Institute of Technology, Dublin, Ireland (e-mail: charlie.cullen@dit.ie).

T. Bürger is with University of Innsbruck, Innsbruck, Austria (e-mail: tobias.buerger@deri.org).

## II. MAIN INNOVATIONS

### A. Intelligent Content Creation

The first goal is to obtain a better understanding of the relations between media types, genres, workflows and styles as a pre-requisite to the adaptation and transfer of content elements across productions and platforms. To this end, metadata, media semantics and ontologies need to be analysed, researched and developed that define the parameters necessary for the creation and manipulation of semantically aware media objects of various types. Practical methods of context-based information retrieval will be researched that simplify the location and retrieval of characters, sounds, images, movements or behaviours from very large datasets and media storage systems. Improved methods and tools for language processing and speech synthesis, as a means of supporting the generation of multilingual media content, need to be developed.

### Media Semantics and Ontologies

The objective of this research strand is twofold: the main objective is to devise a machine process able description for the semantic features of a multimedia object and the context it should be used in. This will be tackled by building up a set of ontologies taking into account a layered approach – using an appropriate representation technique for every level of meta-information – that relies as much as possible on current description standards for multimedia. The second objective is to design and implement necessary tools and applications to build up, maintain and query ontologies for multimedia objects.

### Media Forms, Programme Styles & Structures

Media objects have different specification needs on different platforms from game consoles and online services to DVD, television and cinema, related to the overall expressive and stylistic objectives of the production. Audience expectations are related to the production genre, be it a western, soap opera, comedy, tragedy, thriller, action-adventurer, a medieval sword & magic MMPORG (Massively Multi-Player Online Role-Playing Game). The genre expectations (whether the engagers are passive watchers of Jerry Seinfeld on television, or active console game players

represented on the screen by Lara Croft) are elegantly expressed by the active questions of Philip Parker [2].

### Context Based Information Retrieval

We expect that the intelligent content elements developed by SALERO will adapt themselves to the context of the production. We will therefore need to research ways of defining, creating (or locating), managing and delivering content objects of different kinds in a range of contexts. We address the context-based retrieval of media objects within a media production environment. The idea is to re-use objects, motion data and other production related data for the creation of new production materials. It is not only about finding and re-using elements at production time, but also using retrieval technologies for the creation of interactive media productions. That means for example that a character may react or adapt to a scene in a way that is based on the users input. A number of factors affect the use of retrieval techniques within media production environments. The most important one is that they should be integrated into the production environment: retrieval should happen as part of the programme development and not as a cumbersome, extra activity. We also need intelligent context sensitive retrieval mechanisms that identify both user context and task context.

### Speech and Language

The aim of this activity is to enable programmes created in one language to be re-purposed and/or synthesised in another language or dialect by researching and developing a 'speech corpus/concordancer'. A speech corpus[3], tagged for various features such as rhythm, pitch contours, intensity contours and emotional dimensions [4] will be used to inform lip-synching, character animation and TTS synthesis stages by establishing emotional rules - initially for English - with which to potentially repurpose 'neutral' or 'nearest match' speech segments in the database for the other language or dialect. Once the rules have been established for English, they will be adapted for Spanish or Catalan.

This requires a framework for defining voice stereotypes for age, genre, emotional dimension etc and a suitable tagging system for corpus transcripts-initially for Catalan, Spanish and English. In the case of English, tagging for speech rhythms and other acoustic features within the recorded speech clips is seen to play a crucial role in developing a more natural corpus. The tagging of the resultant speech corpus will be applied to rule-based analysis, synthesis, lip-synching and character animation.

### Characters, Characteristics & Effects

The research in this activity deals with both visual objects (such as characters) and audio objects (such as effects, speech). It provides grounding work for the linking of visual, audio, and behavioural objects, whose initial intelligence is expected to be increased along the lifetime of the project. It develops along different levels, from the low level provision of basic affordable rendering engines for media; through intermediate level, such as the modelling and animation of

characters; to high level aspects, e.g. programme generators.

### III. TOOLSETS, DEMONSTRATION & TRAINING

Software toolkits, software systems, plug-ins and interfaces will be developed that allow the control of appearances, sounds, semantic behaviour and properties of intelligent content objects for media production and post-production, and can be used in conjunction with existing industry programs. They will be validated and evaluated through a series of experimental productions, based on scenarios defined by artists and creative media professionals.

Results will be promoted by a broad initiative, developing demonstration test beds and training structures for professionals and researchers, as well as by addressing the relevant standardisation bodies.

### IV. RELATED WORK

A number of research groups are dealing with ontology based description of multimedia items often by applying reasoning to low level features extracted, e.g. [5], [7]. Use of ontology languages for media annotation has been investigated in [8], [6] for video and in [9] for audio. Deployment of semantic technologies in media production in tools used every day by the media professional has been rarely investigated.

The SMaRT networking cluster [10] (which SALERO is member of) combines research in the fields of: Semantic Web, Multimedia and Signal Analysis to address emerging research challenges in Semantic Multimedia.

### REFERENCES

- [1] SALERO web page, <http://www.SALERO.eu>
- [2] P. Parker, "The art and science of screenwriting", Intellect, 1998.
- [3] D.F. Campbell, M. Meinardi, and B. Richardson, "Let the corpus speak!", *40th IATEFL Annual Conference and Exhibition*, 9 – 12 April 2006, Harrogate, UK.
- [4] K.R. Scherer, Emotion as a multi-component process: A model and some cross cultural data. *Review of Personality and Social Psychology*, 1984. 5: p. 37-63.
- [5] S. Dasiopoulou, V. Mezaris, I. Kompatsiaris, V.K. Papastathis, and M.G. Strintzis: "Knowledge-assisted semantic video object detection", *IEEE Transactions on Circuits and Systems on Video Technology*, Vol. 15, No. 10, pp. 1210 - 1224, (2005).
- [6] J. Heflin, "OWL Web Ontology Language: Use cases and requirements", *W3C Recommendation*, <http://www.w3.org/TR/webont-req/>, (2004).
- [7] V. Mezaris, Y. Kompatsiaris, N. Boulgouris and M. Strintzis, "Real-time compressed domain spatiotemporal segmentation and ontologies for video indexing and retrieval", *IEEE Transactions on Circuits and Systems on Video Technology*, Vol. 14, No. 5, pp. 606 – 620, (2004).
- [8] J.R. van Ossenberggen; F.-M. Nack; and L. Hardman; "That obscure object of desire: multimedia metadata on the web (Part II)", *IEEE Multimedia*, Vol. 12, No. 1, pp. 54 - 63, (2005).
- [9] P. Cano, M. Koppenberger, S. Le Groux, J. Ricard, N. Wack, and P. Herrera, 2005. "Nearest-neighbour automatic sound classification with a wordnet taxonomy". *Journal of Intelligent Information Systems* Vol.24 .2 pp. 99-111 (2005).
- [10] <http://kspace.qmul.net:8080/kpace/kspacesmartcluster.jsp>