Semantic Audiovisual Entertainment Reusable Objects - Vision, Objectives & Results

Georg Thallinger
JOANNEUM RESEARCH
2nd User Group Meeting
London, June 11, 2008

Motivation

- Digital audiovisual media pervade our life
- Automated processing of content is THE problem
- Need to advance the state of the art in digital media to create audiovisual content
  - using ‘intelligent content’ and appropriate tools
  - with greater quality at lower cost
  - for cross-platform delivery
- Provide audiences with more engaging entertainment and information at home or on the move
- Build on and extend R&D in media technologies, web semantics and context based search & retrieval

Intelligent Content!
Content that listens? (P. Kafno*)

- Brings together multimedia, web & knowledge engineering
- Requires new kind of creativity
- Categorizes and meta-tags content
- Personalized, reactive; encourages re-mix and new directions
- Creates revenues

Overall Vision

- Define and develop ‘Intelligent Content’ for media production
  - consisting of multimedia objects
  - with context-aware behaviour
  - for self-adaptive use and delivery across 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.

Areas of Work & Major Achievements

- Speech
  - Linguistic analysis and tagging
  - Expressive speech generation
- Animation
  - Emotional character animation and lip-synching
- Semantic description (emotion, context)
- Context based search & retrieval
- Experimental productions

Major Achievements Semantics & Context

- Semantic description
  - Ontology for virtual characters/animation production
    - Based on previous work (aim@shape virtual human ontology)
  - Description of emotion
    - based on activation/evaluation model
    - used in all applications
    - animation, speech tagging & synthesis, search & retrieval
  - Ontology-based annotation tool
  - Context-based search and retrieval
  - Focus on user context
  - Integration of traditional approaches with semantic and content-based search

* Paul Kafno, Out of the Box - Expanding opportunities for content creators, Presentation IST Event, Helsinki 2006.
Major Achievements
Speech & Audio

- Corpus of emotional speech recorded
- Duologues (separate recording of each voice)
- Induced emotion
- Audio analysis
- Stress and prosody
- General audio qualities
- Expressive speech synthesis
- Multilingual (Spanish, English)
- Limited domains, extended to general purpose
- Audio/voice transformation
  - Age, sex, artificial, tempo
  - Scalable audio codec

Audio/voice transformation
Age, sex, artificial

Major Achievements
Facial Animation

- Maskle approach to transfer facial animation

Facial Animation

Major Achievements
Facial Animation

- High quality animation
  - Depending on medical state

**Major Achievements**

- Program generator
- Template based
- High-quality, real-time renderer
- Visual shader specification
- Post production
- Audio transformation

**Applications**

- Program generator
-模板基
- 高质量，实时渲染器
- 视觉着色器规范
- 后期制作
- 音频变换

**Lessons Learnt so Far**

- Reusability of 3D-content limited
- Game domain needs different/additional models
  - Lower resolution for real-time rendering
  - Complete 3D-environment
- Different capabilities of devices
- Workflow and tools: real-time generation vs. post-production path
  - Real-time: creates content, uses and generates metadata
  - Post-production: modifies existing content, uses and generates metadata
- No one size fits it all tool
- “Semantic technologies” have to be gently introduced to the media industry
  - Not considered a core technology, yet
  - Industry interested if added value shown

**Semantic Audiovisual Entertainment Reusable Objects**

Thank you for your attention!
Why would you want to do that?
- More kids spending more time on-line
- Audience expects interactivity too
- Direct route to market - circumvents broadcasters

... Also... It’s fun!

So... What is a "Virtual World"?
- An on-line environment (web site) where you can create a persistent identity (account!) and customise your appearance and the items you collect.

- And play games!

Why a Virtual World? (Why not a game?)
- Virtual Worlds sit in at an interesting point between full blown MMoRPG’s (like World of Warcraft) and social network sites (like BeBo or Facebook)

Why interesting?
- WoW style games cost around £30 million to develop and are very high risk
- Major social networks are also very expensive and the market is hugely competitive
- Virtual Worlds are much less expensive and can produce significant ROI
Examples:
- Club Penguin bought by Disney for $700 million
- Neopets bought by Viacom for $164 million
- Habbo turnover ~ $70 million
- Maple Story even larger in Far East
- Sequoia investing $6 million in Stardoll

Also...
- The Virtual World format allows producers to be incredibly experimental with content - the site can include games, videos, audio, downloads, books and magazines - if you think of it, chances are you can add it to a Virtual World

Business models:
- Advertising: okay in limited parts of the site
- Micro-payments (buy your own 'virtual currency')
- Subscription

Caveat Emptor!
- Many companies have seen the potential of Virtual Worlds for Children, somewhere between 150 - 250 are currently in development.
It's going to get crowded... Maybe...

So how do you build a Virtual World?
Several different technologies are needed:
- Database
- Application
- Communication
- Video (optional, but good to have)

Database
Tracks user accounts, users inventory, scores, location information, all aspects of customisation
Application

The front-end user experience, graphics files, game files etc - most likely to be Flash.

Communication

For player communication, chat, screen location of players for multi-player games. Needs to be at the socket level to minimise latency.

Video

Should ideally be a separate server because of the demands of streaming, but depends on how much video you intend to include.

Legal issues

- In the US in particular VW’s for children of 13 and under require compliance with the FTA’s COPPA regulations
- The EU working party for article 39 seems likely to be influenced by COPPA

Implications:

- Strict requirements for moderation and data protection
- Compliance with Home Office guidelines for recruitment of Moderators
- Strict T & C’s and Privacy Policies

So finally... My Tiny Planets

- Create your own world
- Chat with friends
- Play single and multi-player games
- Fly your own spaceship
- Customise your character
- Watch videos
Customise your space suit

Meet at Planet Central

Chat with friends

Play games

Fly your space ship

Fly your space ship
Visit thousands of worlds

Choose one to be your own

Then make it just how you want it

Then make it just how you want it

Then make it just how you want it

Then make it just how you want it
Then make it just how you want it

Games room

V-card

Buy buildings for your planet
Customise your President

...Leave messages for friends...

...And collect plants
Bones & Bones Dailies

Thomson Systems Germany
User Group Meeting
London, 11 June 2008

BONES Dailies

- BONES Dailies is a fast and efficient workflow product for producing sound synchronised and color corrected dailies (rushes). It has workflow benefits for the entire production and post production chain.

Benefits of BONES Dailies (1)

- Faster than real time Image Ingest, non-stop per Lab Roll or Video Cassette
- Same system for SD, HD and Full 2K workflows (Full 4K a future possibility)
- Full control of all Spirit DataCines and Shadow Telecines
- Faster than real time file based Audio Ingest incl. automatic audio slate detection as a background task
- BONES real time non-destructive non-linear color correction incl. ASC with integrated reference store
- Semi-Automatic sound synchronisation, with manual override
- Playout graded or ungraded versions, with or without "Print" look 3D LUT

Benefits of BONES Dailies (2)

- By dividing the process into four, remarkable efficiencies and flexibility can be realized compared to traditional dailies workflows
- Template based Production and Playout configurations
- Existing BONES systems can be upgraded with BONES Dailies license
- Use for Dailies at night, and Data Capture or Restoration in the day
  - BONES Dailies increases the efficiency and therefore saves time and money
  - BONES Dailies provides new service opportunities for Dailies post productions
  - Working with BONES Dailies gives better predictability of the booking time of Telecines/Scanners

GUI Example – Audio Ingest

GUI Example – Image Ingest
BONES open post production software

- Open software framework for film and video post production
- Most modular software concept with a variety of application modules
- Features can be easily added
- Fully OFX compliant open Plug-In API for 3rd party apps
- Flexible customer workflows
- Very cost effective hardware: HP XW 9400 Linux Workstation

BONES modules

- Transfer
  - SD/HD video capture
  - SD/HD video playback
  - 2K / 4K data capture
  - Supports da Vinci/Pandora control
  - Local or SAN storage

- Mover
  - EDL or keycode list based data and video capture
  - Controls scanner and tape drives
  - Supports data watchdog drive CTCRs
  - RS422 drive options

- Scale
  - Resizing, zooming, panning & scanning
  - Format conversion
  - Scalable performance
  - Software: Skylab, Know

- Repair
  - Semi-automatic dirt & scratch detection and concealment
  - Manual retouching
  - Motion estimation based processing

- Framework
  - Project Timeline
  - Browser
  - Plug-in Package
  - Primary Color Correction
  - LUT/trie node for 3DLUTs
  - Feature point Tracking

- OFX Plugins
  - Furnace
  - Primatte Chroma Key
  - Keylight

BONES in the DI Post Production Workflow

- Film Acquisition & Scanning
  - Data Ingest
  - Color grading
  - Dust busting and Restoration
  - Conforming
  - Scaling
  - Format conversion
  - Mastering

- Digital Acquisition
  - File management

Bones and Bones Dailies

- Bones is a very modular and open platform
- With it’s timeline/flowgraph (node based) approach it is relatively easy to integrate new functionalities or plug-ins
- Bones Dailies is a workflow product which has the Bones architecture inherited
- Audio slate detection and visual image slate detection could be used to auto-sync audio with image sequences
Live Demo of Tool

- overview of Bones and Bones Dailies
CHARACTER ANIMATION for Intelligent Authoring

Marco Romeo, Alun Evans (GTI-FBM-UPF)

Contents
- NINOS Platform
- Program Editor
- MASKLE
- Ongoing Research at GTI

NINOS Platform
- Audiovisual authoring
- AV Automatic generation
- Facial animation
- Real-time rendering
- FBX support

Audiovisual Authoring

Automatic Generation
- XML Program Templates
- Animation Blending

Facial Animation
- Semi-Automatic Rigging with MASKLE
Facial Animation
- Morph Targets
- Emotional Palette

Real-Time Rendering
- Open-GL Render Engine
- State of the Art Shading
- Post-production Render Passes
- Real-Time Soft Depth Map Shadows

FBX Support
- Easy File Interchange
- Multi-platform asset creation
- Single Common File Format for all Assets

3rd Party Integration
- LipSynch
- External Database
Ongoing Research at GTI

- Maskle facial rigging
- Animation tweaking
- iPhone integration
Audio Transformation Tools  
Music Technology Group – UPF  
Jordi Janer  
Second User Group Event  
11th June, 2008 London

Audio Tools

> Voice Transformation  
  > Manipulates speech or singing voice with spectral techniques

> Advanced Audio Equalizer  
  > equalize using energy histogram-based on different criteria

> Tempo Transformation  
  > Combines rhythm analysis and time-scaling  
  > Not included in today’s hand-on demonstration

Voice Transformation

Main features:
> Real-time VST plug-in effect
> It uses spectral techniques for modifying the character of the voice
> Transformations include individual control of tuning, excitation and timbre
> Targeted to recording or post-production studios that require voice processing

Voice Transformation

Tuning Transformations:
> Pitch transposition:  
  > Control the amount of transposition applied to the input pitch of the voice expressed in semitones, so a value of +12 means transpose up 1 octave.
> Pitch Quantization:  
  > Quantize or not the input pitch to the closest semitone in the desired tonality, based in the scale selected by a combo box.
> Vibrato transformations:  
  > three sliders control the vibrato depth, frequency and tremolo frequency.

Voice Transformation

Excitation Transformations:
> Robotizer:
  > Add a robot effect to the voice with metallic sound and constant pitch.
> Roughness (*):
  > Controls the amount of rough in the voice.
> Breathiness (*):
  > Controls the amount of breath in the voice.
> Alienator (*):
  > Change the voice to an outer space sounding voice.
> Whisper (*):
  > Transform the voice into a whisper voice.
> Remove Unvoiced:
  > Unvoiced consonants are not synthesized after transforming the voice.
> (*) not yet integrated in the current VST prototype

Voice Transformation

Timbre Transformations:
> Timbre mapping:  
  > a set of controls draw the timbre mapping curve to be applied in the transformation to change the timbre characteristics of the voice.
  > A diagonal line means no transformation and the deviation from the diagonal means stretching or expanding the spectrum, resulting in a change in the timbre shape (formant characteristics).
VoiceTransformation

Addicional Features:

- Presets:
  - Human transformations
    - Female: gives the voice a distinct female character
    - Older: ages the voice to make it sound like an old person
    - Child: makes the voice sound like a child
  - Fiction transformations
    - Ogre: makes the voice resemble a horrible giant beast’s voice
    - Robot: makes the voice sound electronic and metallic
    - Alien: gives the voice a weird out of space personality
    - Comic: makes the voice sound like a cartoon character
- The user can store new presets.
- Rendering mode:
  - uncompromised quality for offline processing

VoiceTransformation

User Interface:

Advanced Equalizer

Main features:

- Real-time VST plug-in effect
- Equalize the audio using energy histograms based on different criteria (e.g., panning, inter-channel phase difference, etc.).
- Spectral Filtering using time-frequency masks.
- Some of its applications include: remixing, source separation, or as an audio effect

Advanced Equalizer

Filter parameters:

- Traditional equalizers are based on the energy distribution in frequency

Advanced Equalizer

Filter parameters:

- We can also equalize the audio using energy histograms based on different criteria:
  - panning, phase relation, amplitude, amplitude variation and frequency
- Example: panning

Video:
Advanced Equalizer

Applications:

> Remaking a piece by:
  > changing the volume of each instrument independently.
  > removing instruments from the mix and then re-adding them processed by some effects.
> Remove/extract an instrument from an audio mix.
  > including removing the voice to be used in karaoke.
> Isolate one or several instruments to perform a musical analysis of a piece or used on other projects.

Advanced Equalizer

User Interface:

Video:

TempoTransformation

Main features:

> Real-time VST plug-in effect
> Combines automatic rhythm analysis and time-scaling in order to transform the rhythm of an audio mix
> the user can vary the tempo or add/subtract swing by using the GUI controls
> Targeted to recording or post-production studios that require to modify existing recordings

TempoTransformation

Tempo control:

> Speed controls the amount of time stretch/expand applied to the output sound playing the sound faster or slower but maintaining the original timbre.
> Swing controls the Swing factor during the playback. It can be either used to add swing or remove swing to an audio mix.
> The average BPM of the audio is shown as “detected” BPM and the current transformed tempo is displayed as “current”.

TempoTransformation

Conducting:

> Some controls allow conducting the audio manually.
  > Using a MIDI controller: the plug-in captures note-on’s and estimates the new tempo from the distances between these note-on’s, so the music will follow the tempo given by the midi controller.
  > Mouse control: the tempo can also be controlled with the mouse, by clicking on the beat button continuously at every beat.
  > Other: preliminary experiments using a Wii controller.
Aim of This Presentation

- Demonstrate some of the emotional avatar R & D conducted by Blitz Games Studios under the EC FP6 SALERO Grant
- Demonstrate Experimental Production work using SALERO developed technology in conjunction with the UK Government TSB Serious Games Grant 17178

TruSim

- TruSim: The training games (“Serious Games”) division of Blitz Games Studios Ltd
- 235 staff in Leamington Spa, UK: One of the world’s largest video games developers
- Games for all formats: PC, consoles and handhelds
- 17 years consistently delivering on time & to budget
- Blitz’ games sold to date: $200++ million
- All games developed using the BlitzTech Engine
- R&D Grant Funding: EC FP6 SALERO & UK TSB Grant

Experimental Production

Triage Trainer

R&D part-funded by an EC FP6 grant and a UK Government research programme - to determine how video games should be applied for effective learning transfer & retention

ALSG – Triage Decision Tree

ALSG - Triage Protocols

- Adult Capillary Refill
- Adult Pulse Rate
- Recognised protocols applicable internationally to both military and civilian situations

TRIAGE SIEVE - STORYBOARD

- IED Explosion
- Busy City Centre
- Example Shown uses Advanced Life Support Group (ALSG) Triage Protocols
TRIAGE SIEVE - FIDELITY
Research - Preference for High Fidelity:
- Human Characters
- Appearance (Face, Body, Injuries etc)
- Movement (Eyes, Hands, Breathing etc)
- Behaviour (Emotions, Reactions etc)
- Environments
- Audio

TRIAGE SIEVE - FIDELITY
Research - Reduce Cost of Production Using:
- High Resolution Medical Scanners
- Intelligent Procedural Animation Generation
  - Procedural Generation of Animation Assets
  - Constant Procedural Dynamic Blending
  - Code-driven Animation

FIDELITY
STEP 1 - THE FACE

Medical Scanner Set Up

Point Cloud Scan & Colour Map

Colour Map Generated From Scan Data
Scan To Model – Colour Map Applied

Scan 80,000 Polygons
Game 2,200 Polygons

FACE & EYES*
* Intelligent Procedural Animation Generation

Realistic Eye Movement

Eye Movement In Model

EMOTIONS*
* Intelligent Procedural Animation Generation

Expressions & Emotions (Face)
VOICE*
* Intelligent Procedural Animation Generation

FIDELITY
STEP 1 - THE FACE
STEP 2 - THE BODY

Muscularity & Movement

Breathing

Hands & Movement
Expressions & Emotions (Whole body)

PHYSIOLOGICAL DISPLAY*
* Intelligent Procedural Animation Generation

Procedurally Generated - Shock & Bleeding

Procedurally Generated – Heads (Ethnicity, age, weight)

FIDELITY
STEP 1 - THE FACE
STEP 2 - THE BODY
STEP 3 – THE ENVIRONMENT

Building The Environment
Modularity & Re-Use of Assets*

- Characters & Clothing
- Injuries
- Environments etc

*This Means Significantly Reduced Costs of Production!

Modularity – Characters & Clothing

Modularity – Injuries

PUTTING IT ALL TOGETHER!

ALSG – Triage Decision Tree

ALSG - Triage Protocols

- Adult Capillary Refill
- Adult Pulse Rate
- Recognised protocols applicable internationally to both military and civilian situations

VIDEO or LIVE DEMO
Acknowledgements – Triage Trainer

External Subject Matter Experts:
- Dr. Kevin Mackway-Jones – Prof. of A&E Medicine, Manchester Royal Infirmary
- Dr. Simon Carley – Consultant in Emergency Medicine, Manchester RI
- Lt. Col. David Vassallo – Consultant General Surgeon, Royal Hospital Haslar
- Prof. Bob Stone – Head of Human Interface Technologies Team, Birmingham Uni
- Les Moseley – Director Coventry University Centre for Disaster Management
- Dr. Sara de Freitas – Director of Research, Serious Games Institute, Coventry Uni.
- James Knight – Research Fellow, Birmingham University
- Dr. Daniela Romano – Lecturer Dept. Computer Science, University of Sheffield

Grant Support from:
- EC FP6 – SALERO Integrated Project
- UK TSB Technology Programme