




## Semantic Audiovisual Entertainment Reusable Objects Vision, Objectives & Results

Silvia Russegger  
JOANNEUM RESEARCH  
1<sup>st</sup> User Group Meeting


Barcelona, February 15, 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

Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 2




## 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



\* Paul Kafno, Out of the Box - Expanding opportunities for content creators, Presentation IST Event, Helsinki 2006


Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 3



## 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.


Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 4



## Areas of Work & Major Achievements

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

Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 5



## Major Achievements Semantics & Context


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

Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 6

## 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)
  - Currently limited domains
- Audio/Voice transformation
  - Age, sex, artificial
- Scalable audio codec





Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 7

## Audio/Voice transformation Age, sex, artificial





Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 8

## Major Achievements Facial Animation


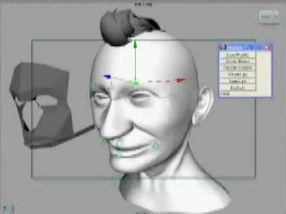


- Maskle approach to transfer facial animation




Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 9

## Facial Animation





Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 10

## Major Achievements Facial Animation

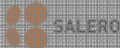
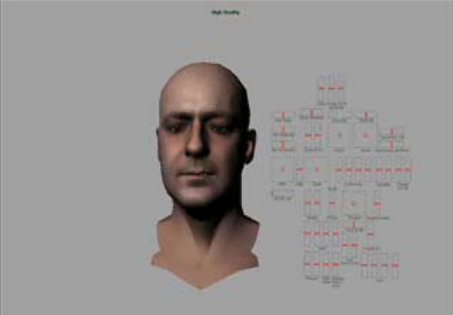


- High quality animation
  - Depending on medical state




Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 11

## Facial Animation





Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 12

## Major Achievements Applications


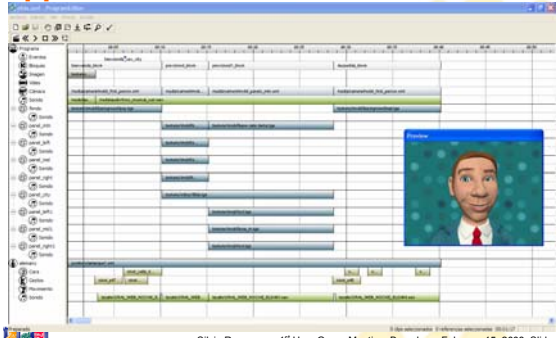


- Program generator
  - Template based
- Visual shader specification
- Post production



IST 027122 Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 13

## Applications


IST 027122 Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 14

[www.salero.eu](http://www.salero.eu)  
[en.wikipedia.org/wiki/SALERO](http://en.wikipedia.org/wiki/SALERO)  
[www.youtube.com/intelligentcontent](http://www.youtube.com/intelligentcontent)

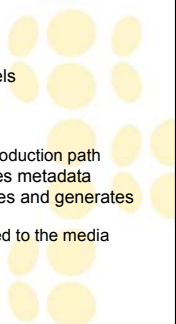



IST 027122 Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 15

## Lessons Learnt



- 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
- "Semantic technologies" have to be gently introduced to the media industry
  - Not considered a core technology, yet
  - Industry interested if added value shown



IST 027122 Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 16

## Semantic Audiovisual Entertainment Reusable Objects




Thank you  
for your  
attention!


With friendly permission, from Pepper's Ghost Productions Ltd.

IST 027122 Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 17



## SALERO's Experimental Productions


Jordi Payo (AM)  
SALERO User Group Meeting BCN 20080215



### Full House – interactive TV political talk show Cross-media format


- Interactive political talk show for TV and cross media.
- Can find information about election candidates and ask questions, etc...
- FULL HOUSE character by answering the questionnaire.
- Help select the best matching candidate.

IST 027122 Jordi Payo (AM) SALERO User Group Meeting BCN 20080215, Slide 2



### Full House – interactive TV political talk show Cross-media format

IST 027122 Jordi Payo (AM) SALERO User Group Meeting BCN 20080215, Slide 3



### Bing and Bong's Tiny Planets: Spark and Socket

- highly production
- methodologies for character creation
- asset re-use
- lip-synch extracted directly from audio files
- character movements, camera configuration and environment designed for maximizing efficiency
- two people in two weeks


IST 027122 Jordi Payo (AM) SALERO User Group Meeting BCN 20080215, Slide 4



### Bing and Bong's Tiny Planets: Spark and Socket



IST 027122 Jordi Payo (AM) SALERO User Group Meeting BCN 20080215, Slide 5





### Triage Simulator

- high fidelity serious game
- teach and reinforce the process of the triage sieve
- Ultra-realistic 3D human models
- physiological system, casualties changes over time, depending on injuries
- Detailed feedback results about the progress

IST 027122 Jordi Payo (AM) SALERO User Group Meeting BCN 20080215, Slide 6


## Triage Simulator



IST 027122

Jordi Payo (AM) SALERO User Group Meeting BCN 20080215, Slide 7

## Hack the Van





- daily TV show with music clips and weather information.
- aimed at young people
- program will be automatically generated
- two versions, role of the characters changes
- Template-based animation
- different audio generation sources
- Reuse of movements
- ...

IST 027122

Jordi Payo (AM) SALERO User Group Meeting BCN 20080215, Slide 8



## Hack the Van 1



IST 027122

Jordi Payo (AM) SALERO User Group Meeting BCN 20080215, Slide 9

## Hack the Van 2



IST 027122

Jordi Payo (AM) SALERO User Group Meeting BCN 20080215, Slide 10



## Thank you for your attention

Jordi Payo (AM)  
SALERO User Group Meeting BCN 20080215






# Ontologies and Media Productions

Silvia Russegger  
JOANNEUM RESEARCH  
1<sup>st</sup> User Group Meeting

Barcelona, February 15, 2008



## Content

- What is an ontology?
- What are the basics of an ontology?
- Why and where should we use an ontology?
- Which services will be provided?

Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 2




## Approach to the Definition of „ONTOLOGY“

*An ontology defines the basic terms and relations comprising the vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary*  
Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., Swartout, W.R. Enabling Technology for Knowledge Sharing. AI Magazine. Winter 1991. 36-58

*An ontology is a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base*  
B. Swartout; R. Patil; k. Knight; T. Russ. Toward Distributed Use of Large-Scale Ontologies Ontological Engineering. AAAI-97 Spring Symposium Series. 1997. 138-148

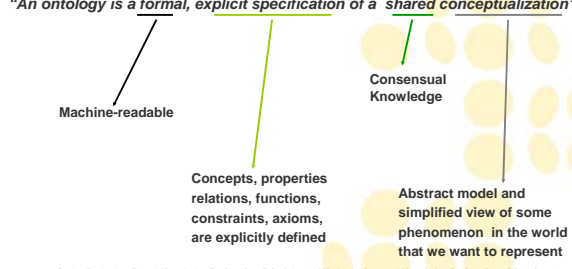
*An ontology provides the means for describing explicitly the conceptualization behind the knowledge represented in a knowledge base*  
A. Bearnaras; I. Laresgotti; J. Correr. Building and Reusing Ontologies for Electrical Network Applications ECAI96. 12th European conference on Artificial Intelligence. Ed. John Wiley & Sons, Ltd. 298-302

Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 3



## Approach to the Definition of „ONTOLOGY“

*“An ontology is a formal, explicit specification of a shared conceptualization”*




Machine-readable

Consensual Knowledge

Concepts, properties relations, functions, constraints, axioms, are explicitly defined

Abstract model and simplified view of some phenomenon in the world that we want to represent


Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 4



## Features of an ontology

- Modelled knowledge about a specific domain
- Defines
  - A common vocabulary
  - The meaning of terms
  - How terms are interrelated
- Consists of
  - Conceptualization and implementation

Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 5





## Ontology Basics

- Ontologies in Computer Sciences
  - Logical constructs with the ability to store knowledge about a given domain
  - Actually a description of participating objects, their characteristics and their relationships
- Basic anatomy of an ontology description
  - Concepts
    - Object and object groups from a specific domain
  - Properties
    - Attributes of concepts
  - Relations

Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 6

### Motivation – Problem

- Possible keywords annotation
  - Man
  - English
  - Comedian
  - Phone call
  - TV show
  - John Cleese
  - etc.
- A search for “British humour” would fail

©BBC IST 027122


### Motivation – Solution




- Using additional ontology input for semantic content retrieval
- Examples:
  - Wales, Scotland, England, Great Britain, United Kingdom, Northern Ireland
  - Athenian drama, Comedy, Tragedy, Satyrs, Humour
  - Relationships: Great Britain is a type of United Kingdom; Comedy is used by Humour; Athenian drama is a type of Tragedy

©BBC IST 027122

### Integration of ontologies in ...




- Asset Management Tools and Production
  - Annotation
    - annotation tool
  - Search and retrieval of media assets
    - search tool
  - Production support
    - intelligent filtering

Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 9

©BBC IST 027122



### Retrieval & Reasoning



- Ontologies provide additional reasoning capabilities
  - Semantic instead of syntactic information retrieval!
  - Retrieval is not limited to a designated occurrence
    - EXAMPLE
      - Search for an unspecific multimedia segment
      - Different segment types can be processed within one single query by selecting the “supertype”

©BBC IST 027122


### Example

- Ontologies are useful NOT for retrieval ONLY
- Problem: ambiguity
  - Search on GOOGLE-Images for “jaguar”:
    - Animal?
    - Motor company?
    - Aircraft type?
- Solution:
  - Disambiguation by providing ontology information together with content description

©BBC IST 027122

### Services provided by an Ontology



- Find super / subconcepts
  - e.g.: Interactive → Environment → Has sub-scenes?
- Find all instances of a specific concept
  - e.g.: BodyAction → Interacting Object: With which objects is it possible to accomplish a fixed body action?
- Find all properties of a concept
  - e.g.: Annotation Use Case → For annotating a character the annotation tool guides the user like a wizard through all values that have to be filled out

Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 12

©BBC IST 027122

## Services provided by an Ontology (contd)



- Find similar instances( similarity metric has to be defined)
  - e.g.: Chair, desk of concept furniture using the same texture
- Check the completeness of a specific concept (annotated asset)
  - e.g.: Is a specific property or character rigged for animation?
- Find concepts that contain a STRING and / or are dealing with other concepts containing this STRING
  - e.g.: What restrictions are there on the usage of a specific asset?
- Complex queries spanning multiple concepts
- Broaden / narrow a query with child / sibling / parent concepts



Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 13

## Ontologies – Summary



- Meta data descriptions provide awesome capabilities for management of AV content
- The “semantic” extension
  - Simplified media declaration and specification process
    - Even with the condition of a more detailed specification in the short run
    - Re-use of existent or formerly defined ontological structures
  - Better and easier possibilities for retrieval



## Semantic Audiovisual Entertainment Reusable Objects



Silvia Russegger, 1<sup>st</sup> User Group Meeting, Barcelona, February 15, 2008, Slide 15

# CHARACTER ANIMATION

Marco Romeo (GTI-FBM-UPF)

# Contents

- NINOS Platform
- Program Editor
- Ongoing Research at GTI

## NINOS Platform

- Audiovisual authoring
- Automatic generation
- Multi-format output
- Easier HQ facial animation flow
- High quality real-time rendering
- Interactive Lighting
- FBX support



Program Editor

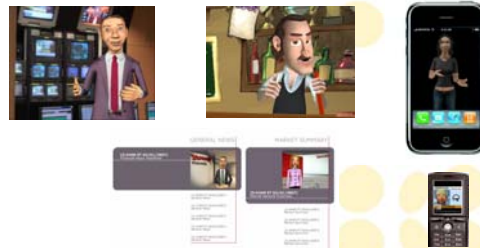
# Audiovisual Authoring



# Automatic Generation

- Program Templates
- XML Parsing
- Database, RSS inputs
- Animation Blending
- Multi-format output (tv, internet, mobile – including WMV, AVI, RM, etc)

# RSS feeds, Multi-format output



## Easier And High quality Facial Animation Flow

- Morph Targets
- Emotional Palette



## High Quality Real-Time Rendering

- Open-GL Render Engine
- State of the Art Shading
- Post-production Render Passes



## Interactive Lighting

- Real-Time Lighting
- Real-Time Soft Depth Map Shadows



## FBX Support

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

NINOS Platform

Program Editor



- Demonstration
- 3rd party integration

Ongoing research at GTI

## Demonstration

- [demo](#)

## 3rd Party Integration

- LipSynch
- Text to Speech
- External Database

Program Editor

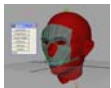
### Ongoing Research at GTI



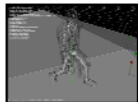
- Maskle facial rigging
- Animation tweaking
- Interactive NINOS
- iPhone integration
- Group Dynamics

## Ongoing Research at GTI

Maskle facial rigging



Animation Tweaking



Group Dynamics



Interactive NINOS



iPhone Integration



Departament de Tecnologies de la Informació i les  
Comunicacions, Institut Universitari de l'Audiovisual i  
Barcelona Media Centre d'Innovació  
Universitat Pompeu Fabra

Passeig de Circumval·lació, 8 - 08003 Barcelona  
Tel : (34) 93 542 2500 y (34) 93 542 2201  
Fax : (34) 93 542 2202

marco.romeo [at] upf.edu

gti-info [at] llista.upf.edu

<http://gti.upf.edu>




# Audio Transformation Tools

Music Technology Group - UPF

First User Group Event  
15 February, 2008 Barcelona



Music Technology Group  
UPF - Barcelona, 2008

# Audio Tools

- > **VoiceTransformation**
  - > Manipulates speech or singing voice with spectral techniques
- > **Tempo Transformation**
  - > Combines rhythm analysis and Time scaling
- > **Advanced Audio Equalizer**
  - > equalize using energy histogram-based on different criteria


Music Technology Group  
UPF - Barcelona, 2008



# VoiceTransformation

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




Music Technology Group  
UPF - Barcelona, 2008



# VoiceTransformation

Tuning Transformations:




- > Pitch transposition:
  - > Control the amount of transposition applied to the input pitch of the voice expressed in cents, so a value of +12 means transpose up 1 octave (+1 octave = +12 semitones = +1200 cents).
- > 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.

Music Technology Group  
UPF - Barcelona, 2008



# VoiceTransformation

Excitation Transformations:



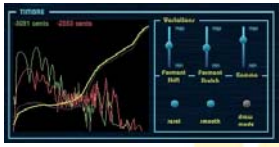
- > **Robotizer:**
  - > Add a robot effect to the voice with metallic sound and constant pitch.
- > **Roughness (\*) :**
  - > Controls the amount of rough in the voice.
- > **Breathiness (\*) :**
  - > Control the amount of breath in the voice.
- > **Alienator (\*) :**
  - > Change the voice to an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

Music Technology Group  
UPF - Barcelona, 2008


# VoiceTransformation

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).

Music Technology Group  
UPF - Barcelona, 2008

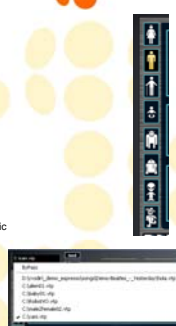


## VoiceTransformation

SALERO

Additional Features:

- > Presets:
  - > Human transformations
    - > Male: gives the voice a distinct male character
    - > Female: gives the voice a distinct female character
    - > Older: ages the voice to make it sound like an old person
    - > Child: rejuvenates the voice to make it sound like a child
  - > Fiction transformations
    - > Ogre: makes the voice resemble a horrible giant beast's voice
    - > Robot: robotizes the voice and makes it 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




Music Technology Group  
UPF - Barcelona, 2008

## VoiceTransformation

SALERO

User Interface:



Music Technology Group  
UPF - Barcelona, 2008

## VoiceTransformation

SALERO

Video:



Music Technology Group  
UPF - Barcelona, 2008

## TempoTransformation

SALERO

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

Music Technology Group  
UPF - Barcelona, 2008

## TempoTransformation

SALERO

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".

Music Technology Group  
UPF - Barcelona, 2008

## TempoTransformation

SALERO

Conducting:



- > Some controls allow conducting the audio manually.
  - > Using a MIDI controller: the plug-in captures note-ons and estimates the new tempo from the distances between these note-ons, 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: experiments with Wii control.

Music Technology Group  
UPF - Barcelona, 2008

## TempoTransformation



User Interface:



Music Technology Group  
UPF - Barcelona, 2008



## TempoTransformation



Video:



Music Technology Group  
UPF - Barcelona, 2008



## 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

Music Technology Group  
UPF - Barcelona, 2008



## Advanced Equalizer



Filter parameters:

- > Traditional equalizers are based on the energy distribution in frequency



Music Technology Group  
UPF - Barcelona, 2008

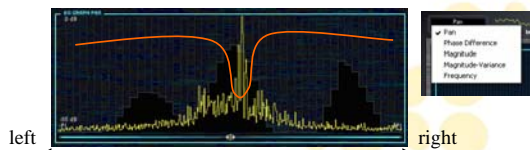


## Advanced Equalizer



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



Music Technology Group  
UPF - Barcelona, 2008



## Advanced Equalizer



Applications:

- > Remixing 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.

Music Technology Group  
UPF - Barcelona, 2008



## Advanced Equalizer



User Interface:



Music Technology Group  
UPF - Barcelona, 2008



## Advanced Equalizer



Video:



Music Technology Group  
UPF - Barcelona, 2008



Thanks!

<http://www.mtg.upf.edu>

Music Technology Group  
UPF - Barcelona, 2008



## Multilingual Speech Synthesis

Carlos Monzo and Xavier Gonzalvo (URL)  
First User Group Event, Barcelona, 15.02.2008

## URL TTS – System core

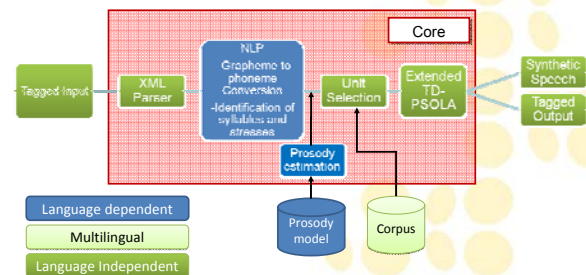
- Language independent core
- TTS system:
  - Currently based on Unit Selection TD-PSOLA [Iriundo et al., 2003]
    - High quality synthesis for restricted domain applications [Alias et al., 2005]
  - Towards statistical approach (HMM-based) [Gonzalvo et al., 2007]
- Input/Output interfaces based on XML

## URL Multilingual speech synthesis system

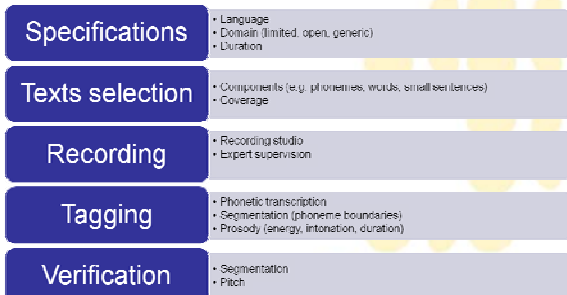
Corpus production for TTS:

- Language independent methodology:
  - Texts selection (e.g. phonetic coverage)
  - Corpus tools (e.g. tagging and integrity testing)
- Recording sessions:
  - Studio facilities

## URL Multilingual speech synthesis system



## Corpus creation diagram



## Corpus creation

- Multilingual =  $n$  languages  $\rightarrow$   $\sim n$  corpus
- Field of application:
  - Restricted domain (e.g. weather forecast)
  - Open domain (e.g. audiovisual productions)
- Tools (domain independent and adapted for multilingual purposes):
  - Texts selection
  - Tagging process
  - Tagging verification

## Corpus tools (i)



- Texts selection:
  - Phonetic balancing
  - Language exception management
  - Corpus duration constraining
- XML Tagging (D6.4.1):
  - Transcription
  - Segmentation (i.e. phoneme boundaries)
  - Intonation (i.e. pitch)



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 7

## Corpus tools (ii)



### Tagging:

- Posterior to audio post-processing
- Phonetic transcription of texts
- Segmentation:
  - Phonemes boundaries
  - Automatic segmentation stage
- Automatic pitch marking [Alias et al., 2006b]



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 8

## Corpus tools (ii)



### Verification:

- Corpus structure
- Segmentation:
  - Number of segmentation marks
  - Minimum units durations
- Intonation:
  - Discontinuities in the pitch contour
  - Maximum and minimum bounds



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 9

## Corpus tools (iv)



### Tools development:

- Core code and user interfaces:
  - C/C++
  - Visual C++
  - Matlab

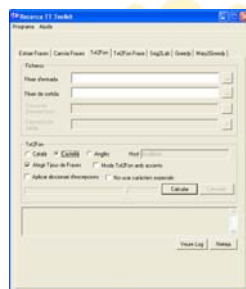


Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 10

## Corpus tools in images (i)



- **RST** (Research Speech Toolkit)
- Texts selection
- Phonetic transcription

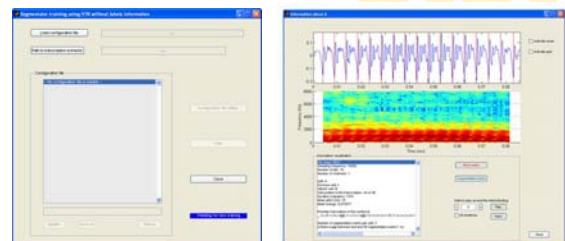


Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 11

## Corpus tools in images (ii)



- **ST** (Segmentation training) and units analysis (**UA**)
  - No previous manual tagging is needed
  - Segmentation based on trained models

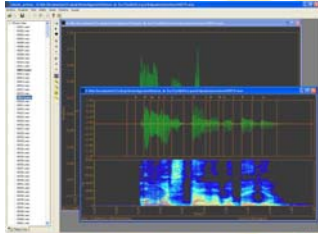


Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 12

## Corpus tools in images (iii)



- **SPI v2.0 (Speech Processing Interface)**

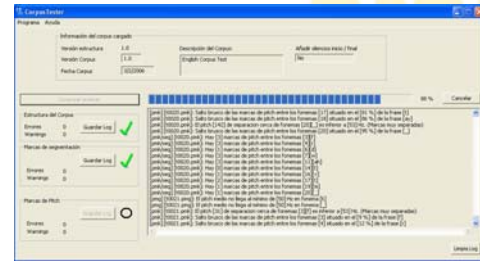


Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 13

## Corpus tools in images (ii)

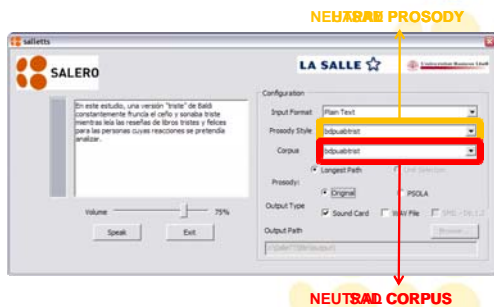


- **Corpus Tester: corpus verification**



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 14

## TTS tools



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 15

## TTS Synch with Animation Experimental Production 1



- **Product Server-Client Oriented (sockets, SOAP...)**

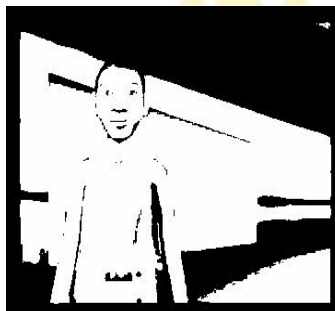


Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 16

## TTS Synch with Animation Experimental Production 1



- XML Tagged output for Lip-Synching and Event Synchronization.
- Fully Automatic



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 17

## URL TTS – Multilingual related TTS improvements (ii)



Spanish HMM-based TTS [Gonzalvo et al., 2007]:

- Statistical approach
- Stable synthesis for general purpose applications
- Reusability: low cost trainable system
- Pursued advantages :
  - Voice transformation
  - Polyglot systems
- Current research focus: improving speech naturalness



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 18

## URL TTS examples



- Restricted domain application (Unit selection):
  - English weatherman   
*"A very good morning to you! How are you today? Would you like to know what the weather will be like today in Helsinki?"*
  - Spanish weatherman   
*"Very good! How are you? Would you like to know what the weather will be like today in Barcelona?"*
- Open domain application (Statistical):
  - Spanish woman   
*"This is a test of emotionated synthesis"*
  - German woman   
*"I'm not really very happy"*



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 19

## URL TTS – Recent related TTS improvements (i)



- Clustering and classification of selectable units [Formiga and Alías, 2007]
- Prosody estimation learnt from corpus [Iriondo et al., 2007]
- Linguistic and Mixed Excitation Improvements on a HMM-based speech synthesis for Castilian Spanish [Gonzalvo et al. 2007]
- Research on naturalness speech improvement by means of voice quality parameters [Monzo et al., 2007]



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 20

## Conclusions



- Language independent:
  - Core (Unit selection and XML tagging)
  - Corpus development methodology (except transcription)
- Unit selection:
  - High quality → restricted domain
  - Multilingual
- Statistical based on HMM:
  - Stable synthesis → open domain (generic)
  - Polyglot



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 21

## Conclusions



- Tools:
  - XML compliant + multilingual methodology
  - Corpus creation and verification
    - RST (Research Speech Toolkit)
    - SPI (Speech Processing Interface)
    - ST (Segmentation Training) and UA (Unit Analysis)
  - TTS:
    - Visual interface
    - Client-server interface



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 22

## References (i)



- Alías, F., Iriondo, I., Formiga, Ll., Gonzalvo, X., Monzo, C., Sevillano, X. 2005. High quality Spanish restricted-domain TTS oriented to a weather forecast application. The 9th European Conference on Speech Communication and Technology (InterSpeech), ISSN 1018-4074, pp. 2573-2576, Lisboa, Portugal.
- Alías, F., Llorà, X., Formiga, Ll., Sastry, K., Goldberg, D.E. 2006a. Efficient interactive weight tuning for TTS synthesis: reducing user fatigue by improving user consistency. In Proc. of the IEEE ICASSP 2006, Toulouse, France, vol. 1, pp. 865-868.
- Alías, F., Monzo, C., Socoró, J.C. 2006b. A Pitch Marks Filtering Algorithm based on Restricted Dynamic Programming. InterSpeech2006 – International Conference on Spoken Language Processing (ICSLP), Pittsburgh, USA, pp. 1698-1701.
- Formiga, Ll., Alías F. 2006. Heuristics for implementing the A\* algorithm for unit selection TTS synthesis systems. IV Jornadas en Tecnología del Habla (4JTH06), Zaragoza, Spain, pp. 219-224, ISBN 84-96214-82-6. (in Spanish).
- Formiga, Ll., Alías, F. 2007. Extracting user preferences by GTM for aiGA weight tuning in unit selection text-to-speech synthesis. The 9th International Work-Conference on Artificial Neural Networks (IWANN'2007), Lecture Notes in Computer Science, ISSN: 0302-9743, Springer Verlag, San Sebastián, Spain.



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 23

## References (ii)



- Gonzalvo, X., Iriondo, I., Socoró, J.C., Alías, F., Monzo, C. 2007. HMM-based Spanish speech synthesis using CBR as F0 estimator, ISCA Tutorial and Research Workshop on Non Linear Speech Processing (NOLISP07), Paris, France, pp. 11-14.
- Iriondo, I., Alías, F., Sanchis, J., Melenchón, J. 2003. A Hybrid Method Oriented to Concatenative Text-to-Speech Synthesis. The 8th European Conference on Speech Communication and Technology (EuroSpeech), ISSN 1018-40-74, Ginebra, Suiza, vol. 4, pp. 2953-2958.
- Iriondo, I., Socoró, J.C., Alías, F. 2007. Prosody modelling of Spanish for expressive speech synthesis. In Proc. of the IEEE ICASSP 2007, Honolulu, Hawaii, USA, vol. 4, pp. 821-824.
- Monzo, C., Alías, F., Morán, J.A., Gonzalvo, X. 2006. Transcripción fonética de acrónimos en castellano utilizando el algoritmo C4.5. Procesamiento del Lenguaje Natural, Zaragoza, Spain, nº 37 (ISSN:1135-5948), pp. 275-282. (in Spanish)



Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 24

## References (iii)



- Monzo, C., Alias, F., Iriondo, I., Gonzalvo, X., Planet S. 2007. Discriminating expressive speech styles by voice quality parameterization. International Congress of Phonetic Sciences (ICPhS07), pp. 2081-2084.
- Tokuda, K., Zen, H., Black, A. 2002. An HMM-based speech synthesis system applied to English. IEEE SSW, Santa Monica, USA.



IST 027122

Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 25



# THANK YOU FOR YOUR ATTENTION

## Multilingual Speech Synthesis

Carlos Monzo and Xavier Gonzalvo (URL)  
First User Group Event, Barcelona, 15.02.2008



IST 027122

Carlos Monzo and Xavier Gonzalvo, First User Group Event, <15.02.2008>, Slide 26