Talking Heads for the Web: what for?

Koray Balci

Fabio Pianesi

Massimo Zancanaro

Outline

- XFace – an open source MPEG4-FAP based 3D Talking Head
  - Standardization issues (beyond MPEG4)
- Synthetic Agents – the Evaluation Issues

Xface

An open source MPEG-4 based
3D Talking Head

Xface

- A suite to develop and use 3D realistic synthetic faces
- Customizable face model, and animation rules
  - Easy to use and embed to different applications
- Open Source (Mozilla 1.1 License)
  - http://xface.itc.it
- MPEG-4 Based (FAP standard)

Xface: Modules

- XfaceCore
- XfaceEd
- XfacePlayer
- XfaceClient

XfaceCore

- Developed in C++, OO
- Simple to use in your applications
- Improve/extend according to your research interest

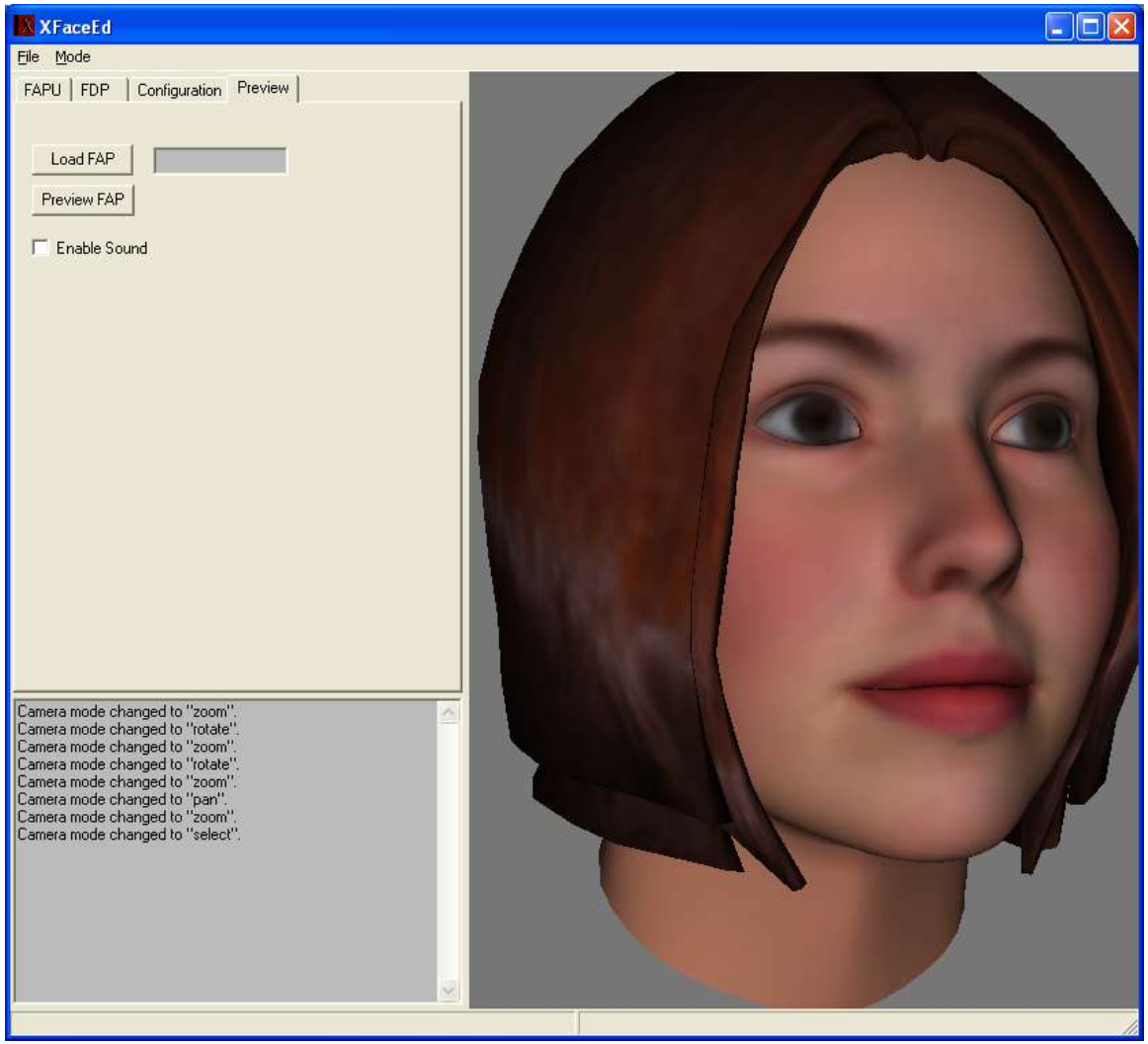
XfaceCore: Sample use

```
// Create the face
m_pFace = new XFaceApp::FaceBase;
m_pFace->init();
// Load a face (and fap&wav similarly..)
Task fdptask("LOAD_FDP");
fdptask.pushParameter(filename);
fdptask.pushParameter(path);
m_pFace->newTask(fdptask);
// Start playback
Task playtask(“RESUME_PLAYBACK");
m_pFace->newTask(playtask);
```

XfaceEd

- Transform any 3D mesh to a talking head
- Export the deformation rules and MPEG-4 parameters in XML
- Use in XfacePlayer

XfaceEd


XFaceEd
File Mode
FAPU FDP Configuration Preview
Load FAP
Preview FAP
Enable Sound
Camera mode changed to "zoom".
Camera mode changed to "rotate".
Camera mode changed to "zoom".
Camera mode changed to "rotate".
Camera mode changed to "zoom".
Camera mode changed to "pan".
Camera mode changed to "zoom".
Camera mode changed to "select".

XfaceEd

XfacePlayer: John

XfacePlayer: Alice

XfacePlayer

- Sample application using XfaceCore
- Satisfactory frame rates
- Remote (TCP/IP) control



XfaceClient

Xface: Dependencies

- Festival for speech synthesis (Uni. of Edinburgh)
- expml2fap for FAP generation (ISTC-CNR, Padova)
- wxWidgets, TinyXML, SDL, OpenGL

XFace Languages

- MPEG4-FAP is a low-level language
- Need for more abstract language

APML: Affective Presentation Markup Language

- Performatives encodes agent's intentions of communication
- Does not force a specific realization
  - FAP will take care of that!

<performative type="inform" affect="sorry-for" certainty="certain**">I'm sorry to tell you that you have been diagnosed as suffering from what we call angina pectoris,**</performative>

De Carolis, B., V. Carofiglio, M. Bilvi & C. Pelachaud (2002). 'APML, a Mark-up Language for Believable Behavior Generation'. In: Proc. of *AAMAS Workshop 'Embodied Conversational Agents: Let's Specify and Compare Them!'*, Bologna, Italy, July 2002.

Problems with APML

- Does not allow different performative on different “modes”
- Lacks of standardization

Can we do that with SMIL?

- Different “modes” associated to different channels
- Performatives as data model

```
<parallel>
    <performative type="inform" channel=”voice”
        affect=”sorry-for”> I'm sorry to tell you that you
        have been diagnosed as suffering from what we
        call angina pectoris,</performative>
    <performative type=”inform” channel=”face”
        affect=”sorry-for”/>
</parallel>
```

Synthetic Agents

The Evaluation Issues

Evaluating expressive agents

- Assess progress and compare alternative platforms wrt
  1. EXPRESSION (recognition): evaluation of the expressiveness of synthetic faces: how well do they express the intended emotion?

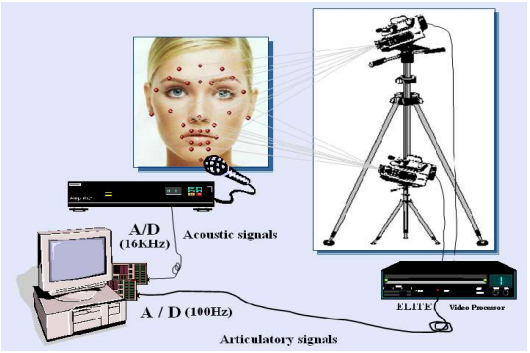
  3. INTERACTION: how effective/natural/useful is the face during an interaction with the human user?
- Build test suites for benchmarking

Procedure

- 30 subjects (15 males and 15 females)
- Within design ; Three blocks (Actor, Face1, Face2)
- Two conditions, randomized within each block:
  - Rule-Based (RB) vs. FAP for synthetic faces
- Three different (randomly created) orders within blocks
  - 14 stimuli per block. 42 Stimuli per subject
  - Balanced order between blocks;

Producing FAP

- ELITE/Qualisys system
- Actor Training
- Recording procedure (example)
  - Announcer
    - <utterance><emotion><intensity>
    - E.g. “aba”, Disgust, Low
  - Actor
    - <CHIUDO> <utterance><PUNTO>
- Example
  - “*il fabbro lavora con forza usando il martello e la tenaglia*”, Happy, High



The Faces: Greta and Lucia

Experiment Objectives and Design

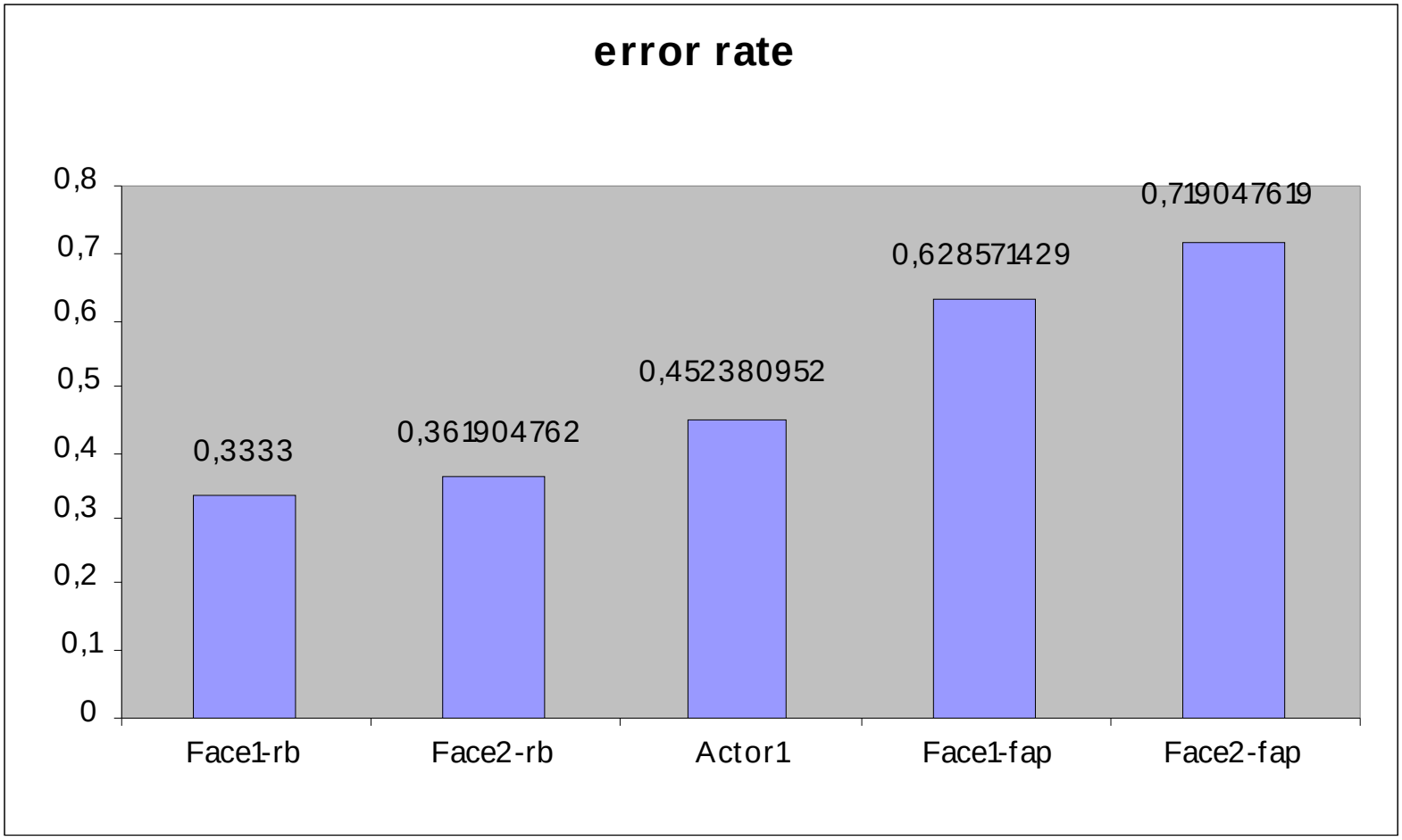
- Comparing recognition rates for **3 FACES**:
  - 1 natural (actor) face and
  - 2 face models (Face1 & Face2),
  - in **2 animation conditions**:
    - Script-based generation of the expressions (RB)
    - FAP CONDITION (face playing actor's faps).
- **Dynamic:** the faces utter a long Italian sentence – audio not available;
- **7 emotional states**: whole set of Ekman's emotions (fear, anger, disgust, sadness, surprise, joy) plus neutral.
- **Expectation**: the FAP condition should be closer to Actor than the SB one

Data Analysis

- Recognition rate (correct/wrong responses)
- multinomial logit model and comparisons of log-odd ratios (z-scores - Wald intervals)
- Errors: information-theoretic approach, measuring :
- number of effective error categories per stimulus and response category
- fraction of non-shared errors on pooled confusion matrices

Results – 1: Recognition rates

	ACTOR	F1-FAP	F1-RB	F2-FAP	F2-RB
anger	90%	27%	53%	7%	23%
happiness	97%	80%	40%	80%	77%
neutral	70%	70%	60%	53%	67%
disgust	13%	20%	53%	17%	17%
surprise	47%	40%	87%	33%	90%
fear	50%	17%	77%	0%	77%
sadness	17%	7%	97%	7%	97%
All	55%	37%	67%	28%	64%


error rate
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0,3333
0,361904762
0,452380952
0,628571429
0,719047619
Face1-rb
Face2-rb
Actor1
Face1-fap
Face2-fap

Recognition Rates – 2: Summary

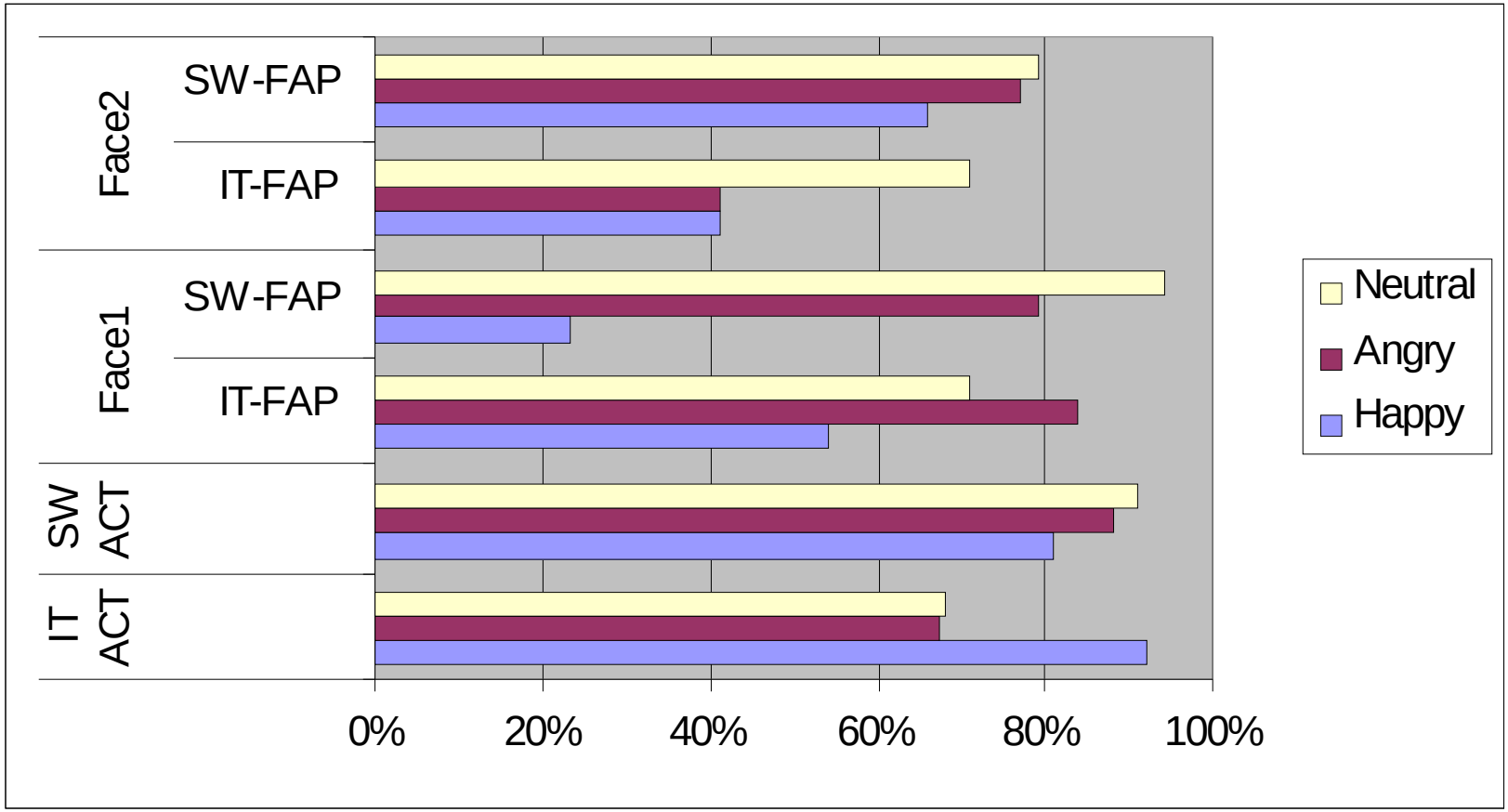
- Actor better than both FAP faces
- The RB mode better than Actor

Logit Analyis

Hit=Face+Condition+Emotion+Face*Condition+Face*Emotion
+Condition*Emotion+Face*Condition*Emotion

- The SB mode is the better, on absolute grounds
- FAP goes closer to ACTOR (if we neglect anger)
  - Both on positive and negative recognitions
  - **FAP faces are more realistic!!!!**
- Recognition rates do not depend much on the particular type of face used (Face1 vs. Face2)

Cross-cultural effect: Italy vs. Sweden



Database of kinetic human facial expressions

- Short videos of 8 professional actors
  - 6 to 12 seconds
  - 4 males and 4 females
- Each actor played the 7 Ekmans' emotions
  - with 3 different intensity levels
- First condition
  - actors played the emotions while uttering a the sentence *"In quella piccola stanza vuota c'era però soltanto una sveglia"*
- Second condition
  - actors played the emotions without uttering
- A total of 126 short videos for each of the 8 actors for a total of 1008 videos.

Related Projects

- PF-Star – EC project FP5
  - Evaluation of language-based technologies and HCI
- Humaine – NoE FP6
  - Affective interfaces and the role of emotions in HCI
- CELECT: Center for the Evaluation of Language and Communication Technologies
  - No-profit research center for evaluation; funded by the Autonomous Province of Trento – 2004-2007

Summary

- Use our Open Source Talking Head:
  - http://xface.itc.it
- Standardization is required at different levels
  - MPEG4-FAP vs. APML vs. SMIL+performatives
- Necessity of Experimental Evaluation
  - When human beings enter into play things are less intuitive!