

# humaine

WP 4:  
From Signals to Signs of Emotion

ICCS, TAU and co.

Plenary meeting, May 25, 2005, Newcastle

# Extraction of Emotional Signs from Signals

---

- ◆ HUMAINE Workpackage 4
- ◆ 9 Participants:
- ◆ *ICCS-NTUA (leader) ,*
- ◆ *TAU, UA,*
- ◆ *UNIGE, QUB,*
- ◆ *TCD, LIMSI,*
- ◆ *FAU-Erlangen, Paris-8.*

# Initial task description

---

- ◆ Four tasks and working groups:
  - ◆ Emotional Speech Analysis and Recognition
  - ◆ Face & Gesture Analysis for Emotion Recognition – Expressive ECA Synthesis
  - ◆ From Physiological Signals to Emotional Signs
  - ◆ Multimodal Emotion Recognition.

# Work already done

---

## ◆ Deliverables Produced

- ➔ D4B: State-of-the-art description and draft potential exemplars (05/2004)
- ➔ D4A: WP4 Workshop on Signals and Signs of Emotion (09/2004)
- ➔ D4C: Potential WP4 exemplar description (01/2005)

# Exemplar Definition

---

## *A Technological Pool of Methods for Multimodal Emotion Recognition*

### ◆ *Focus on:*

- ➔ *extraction of emotional features and signs from the multimodal inputs,*
- ➔ *recognition of the user's emotional state*

### ◆ *Target*

- ➔ *Advance the current state-of-the-art in single and multi-modal emotion recognition*
- ➔ *Take into account knowledge & demands provided by theory, synthesis and interaction.*

# T1: Speech analysis

---

- ◆ *Basic tasks*

  - ➔ Extract F0, intensity, voice quality, word recognition

- ◆ *Segmentation*

  - ➔ Word/phrase/turn boundaries

- ◆ *Feature extraction*

- ◆ *Classification*

- ◆ *Accumulating reference data and defining reference terminology*

## T2: Visual analysis

---

### ❖ *Facial analysis*

- ➔ Extract MPEG-4 FAPs / visemes / gaze
- ➔ Detect movement patterns, temporal correlations

### ❖ *Gesture analysis*

- ➔ Detect arm, wrist, hand movements
- ➔ Extract visual expressivity parameters

### ❖ *Classification*

- ➔ Use a-priori knowledge/rules, adaptation

### ❖ *Reference data*

- ➔ Create/Label emotional benchmarking material.

# T3: Physiological Signal Analysis

---

- ◆ *Affective data processing*

- ➔ Extend and fuse sensor parameters

- ◆ *Segmentation and Feature extraction*

- ➔ Select time intervals
- ➔ Extract most appropriate features

- ◆ *Classification*

- ➔ Statistical / Probabilistic / Hybrid techniques

- ◆ *Reference data*

- ➔ Create and label emotional benchmarking material

## T4. Multimodal emotion recognition

---

- ◆ *Recognition Models- Application dependency*
  - ➔ Discrete / Dimensional / Appraisal theory models
- ◆ *Theoretical models of multimodal integration*
  - ➔ Direct / Separated / Dominant/ Motor integration
- ◆ *Single mode Emotion Recognition*
  - ➔ Advance single-mode recognition state-of-the-art
- ◆ *Modality synchronization*
  - ➔ Visemes/ EMGs&FAPs /SC-RSP&speech
  - ➔ Temporal evolution and modality sequentiality
- ◆ *Multimodal recognition techniques*
  - ➔ Classifiers + context + goals + cognition/attention + modality significance in interaction

## Just to remind you ...

---

- ◆ We're not expected to solve **ALL** the problems in the field
- ◆ We **are** expected to solve some of them
  - And lay the groundwork for future research on the subject

# Workpackage Dependencies

---

- ◆ *Theory (WP3)*

- ➔ *Emotion models, based on particular implementations or applications*

- ◆ *Databases (WP5)*

- ➔ *Provision and labelling of common databases*

- ◆ *Interaction (WP6)*

- ➔ *Expressive ECA generation, extraction of salient features and events*

- ◆ *Cognition (WP7)*

- ➔ *Higher level cues for multimodal emotion recognition*

# The story so far

---

- ◆ *Multimodal signal analysis, extraction of emotional signs and recognition of users' emotional state is a very challenging problem for HCI applications, especially when dealing with naturalistic data*
  - ➔ *Video and audio conditions can be inhibiting*
  - ➔ *Everyday emotions are not extreme, thus not clear, feature-wise*
  - ➔ *Every modality imposes different constraints on time processing and synchronization*
- ◆ *In its 1st year WP4 attempted to gather expertise, interest and efforts of partners and define the WP4 exemplar and future plan of activities*

## The way forward

---

- ◆ Or “what is really missing?”
- ◆ Majority of research still tackles video, audio and biosignals as different, unrelated features
- ◆ The term “audiovisual” processing usually represents an intelligent architecture that fuses features from either modality
  - ➔ But fails to model the interaction *between* modalities

## The way forward

---

- ◆ Audiovisual emotional analysis based on data provided by WP5
- ◆ Speech and physiological data processing
- ◆ Consider theoretical models for emotion recognition w.r.t. representation and multimodality

# Examples and Demos

---



- ◆ Bad lighting and image quality hamper visual feature extraction
  - ➔ But still, results are acceptable
- ◆ Emotional information from the mouth area?
  - ➔ Mouth deformation is the result of the person smiling or talking?

## Some Demonstrations from ICCS

---

- ◆ Spiros & Amaryllis ....

## Some remarks:

---



- ◆ Study of temporal aspects across modalities
  - ➔ Research usually deals with “frozen” instances
  - ➔ FAPs are fine to model expressions in still images, but...
  - ➔ ...no real information on how long an event lasts or which was the previous state

## The “silent movies” we just saw show some openings for multimodal interaction:

---

- ◆ Audio → video:
  - ➔ For example: Information from detection of pauses is vital to detection of mouth shape when there is no phonation
- ◆ Video → audio :
  - ➔ Viseme detection can improve speech analysis (phonemes)

## *And what have the speech guys been up to?*

---

- ◆ Long before HUMAINE ...  
The various members of the Speech Sub Group (SSG) have been very active in studying different aspects of emotion in speech -
  - ➔ Using many different types of emotional databases, obtained in many different scenarios
  - ➔ Using a large variety of speech analysis and classification methods
- ◆ Our challenge: how do we share this large body of experience across the different members?

## From the exemplar proposal:

---

...This title has been chosen to stress the fact that members of the workgroup have acquired large and different degrees of expertise in the various aspects of emotion characterization, analysis and recognition. The aims of the exemplar are to pool the resources of the different members and share their accumulated experience for generating new knowledge and algorithms in the area; the latter will be shared with the rest of HUMAINE WPs for achieving the project goals....

## How have we been approaching this?

---

- ◆ Writing long documents ...
- ◆ Meeting in *Greek islands* ...
- ◆ Discussing plans of action and means of collaboration:
  - ➔ Getting our tools together
  - ➔ And some databases
  - ➔ And some models of emotion
- ◆ Doing all of the above with smart people from WP3 and WP5
- ◆ It's time to start connecting all of these

## Some thoughts on how to continue:

---

- ◆ One very positive direction is using a common database
  - ➔ **Erlangen** has been very kind to contribute their Aibo database (children interacting with a robotic dog, WOZ style)
  - ➔ They also hammered out a collaboration contract (*CEICES*) that other WPs might find useful
  - ➔ *Do we need more databases? **Skeptical databases?** (Probably. The video is beyond current facial analysis capabilities. SAL?)*
- ◆ Some tentative outline for division of labor is called for,
  - ➔ Then: Lock our postdocs in dark dungeons and make them work very hard
  - ➔ *Our postdoc in Israel has been examining some basic tools for basic speech analysis – choosing good tools with good availability will facilitate collaboration among groups*
  - ➔ *Next he'll probably be hard at work on feature extraction*

## Summary:

---

- ◆ WP4 plans for action are outlined quite clearly in the documents we spent a lot of time writing
- ◆ Now we have to carry them out ...