Modeling socio-emotional agent

Catherine Pelachaud & GRETA Team

CNRS – LTCI

Telecom ParisTech

catherine.pelachaud@telecom-paristech.fr

Embodied Conversational Agents: an introduction

Embodied Conversational Agents – ECAs are :

- Embodied: body with communicative capabilities
- Autonomous: they can plan what to say, know when to start a conversation, when to answer and when to take the conversation turn
- Interactive: react, adapt, be synchronized
- Expressive: display multimodal synchronized behaviors

Embodied Conversational Agents: Research

Modelling ECAs:

- ECA's cognitive and expressive capabilities simulate human capabilities
- Models are based on theories from human studies
- Several domains: linguistic, phonetics, cognitive science, emotion, psychology, and sociology

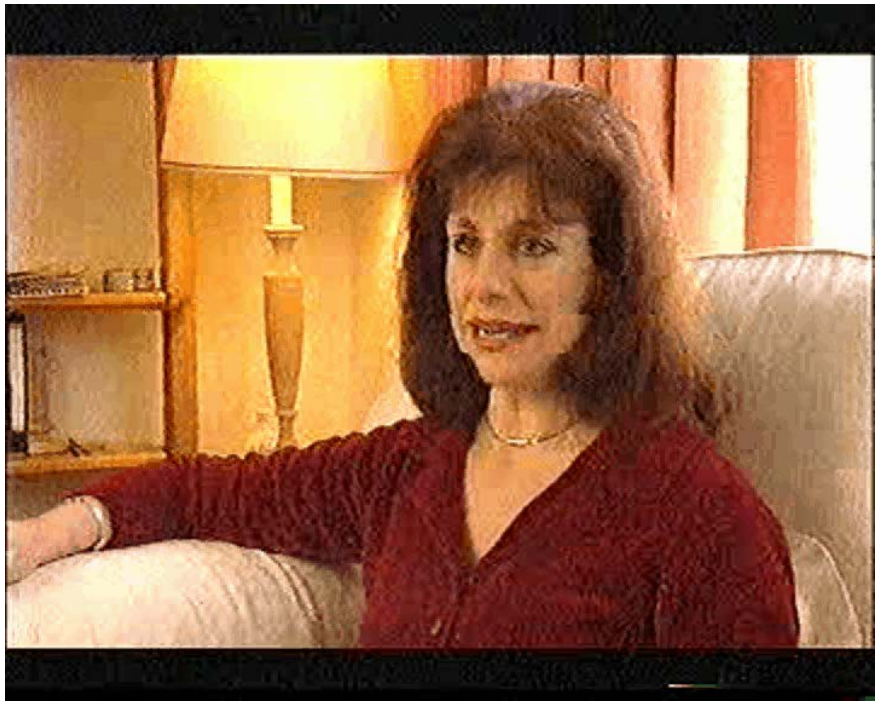
Embodied Conversational Agents: Application

- **Serious games:**
  - Training: job interview, intercultural communication
  - Social ability: bullying, shyness
- **Health & wellbeing**
  - Depression
  - Post-traumatic stress disorder (PTSD)
  - Autism
- **Education: virtual tutors**
- **Rehabilitation: speech therapy**
- **Companions**

Communication

- **Definition of communication (Allwood, 06)**
  - Not simply a transfer function
  - Sharing anything between arbitrary entities where all entities are active, interacting with each other and within a social and interrelational context
  - Not just words, but also
  - Intonation and paralinguistic
  - Facial expression, gaze, gesture, body movement, posture…

Communication



Belfast Naturalistic Database

Research questions & application

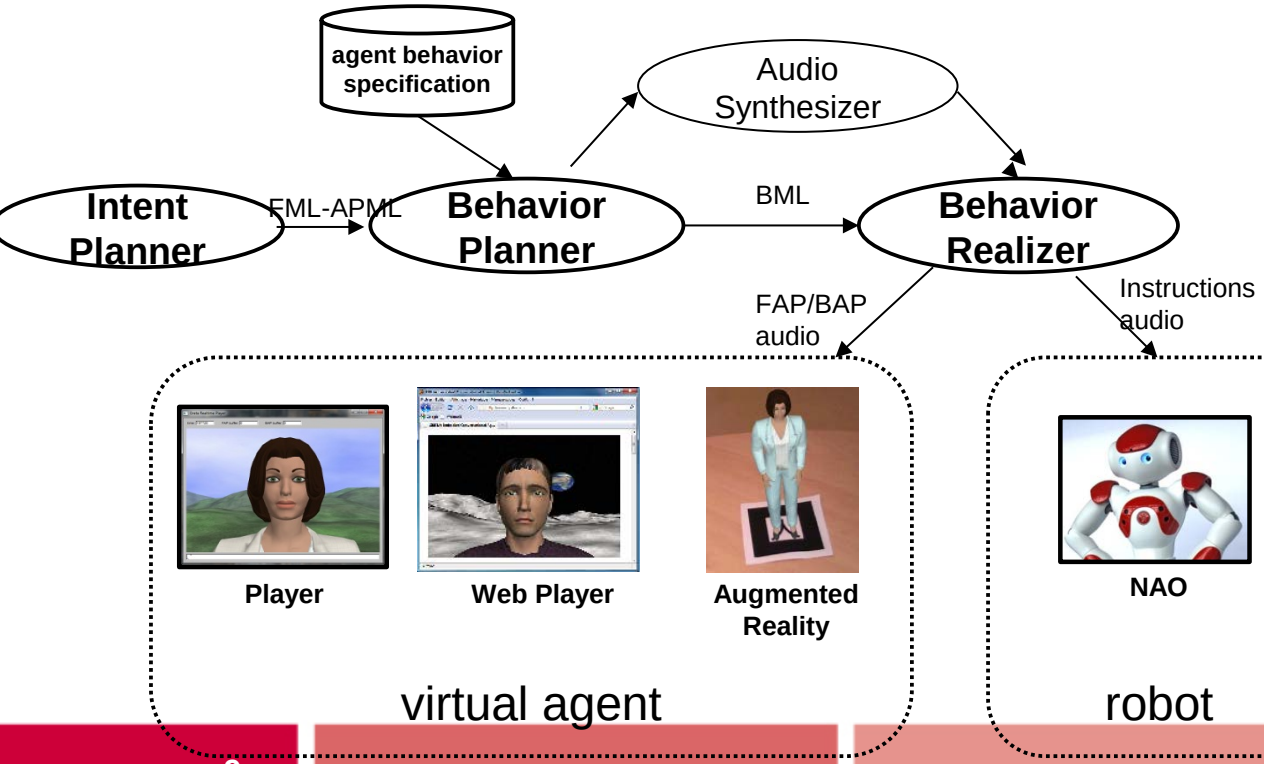
- **Research questions**
  - How to capture variety and subtleness of human behaviors?
    - Multimodal repertoire (communicative, emotional and social signals)
  - How to build a sensitive interactive ECA?
    - Human-ECA interaction
  - How to build an ECA that copies and adapts to user's behavior?
    - Enhance sense of engagement
- **Application**
  - Virtual recruiter with social attitudes

GRETA PLATFORM

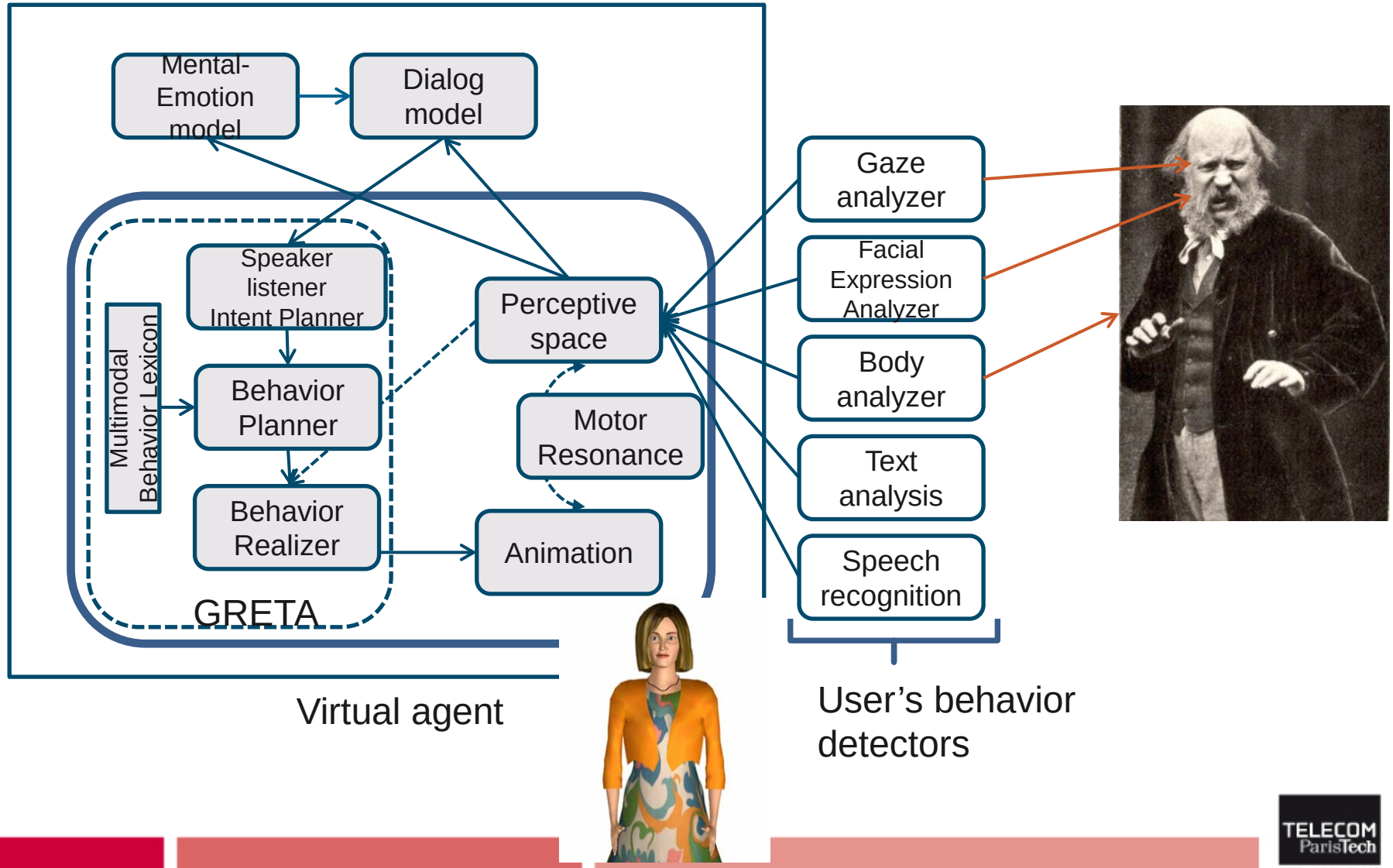
GRETA TEAM

Greta Platform

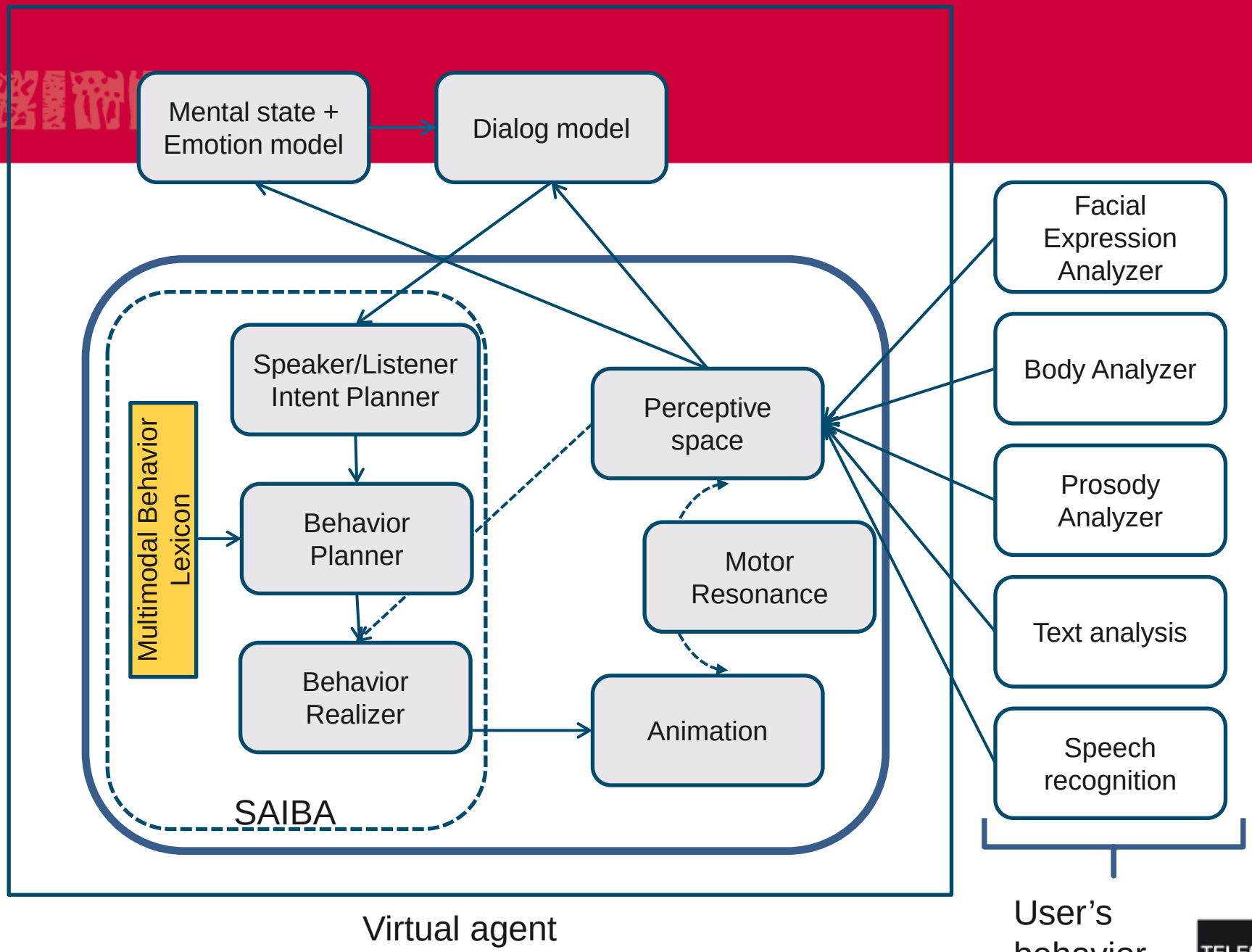
- Embodied conversational agent architecture:
  - Different embodiments and agent representations
  - Multimodal synchronized behaviors linked to intention&emotion



VIB: Human-Agent Interaction Platform



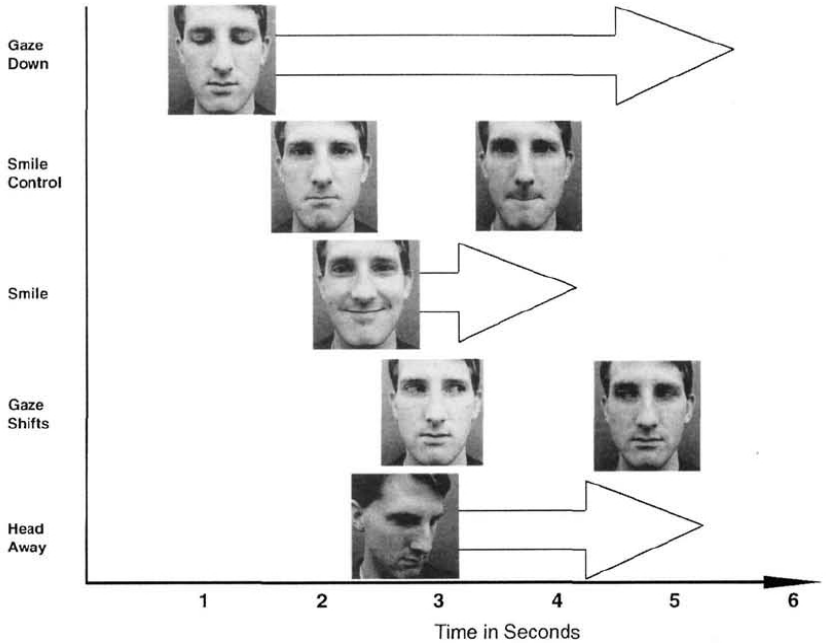
COMMUNICATIVE, EMOTIONAL AND SOCIAL BEHAVIORS: BUILDING A REPERTOIRE


Mental state + Emotion model
Dialog model
Speaker/Listener Intent Planner
Multimodal Behavior Lexicon
Behavior Planner
Behavior Realizer
SAIBA
Perceptive space
Motor Resonance
Animation
Facial Expression Analyzer
Body Analyzer
Prosody Analyzer
Text analysis
Speech recognition
Virtual agent
User's behavior detectors

Multimodal Sequential Expressions – Niewiadomski & Hyniewska

- Go beyond static mono-modal expressions
- Many emotions are expressed by sequences (or combination) of multimodal signals rather than monomodal signals (eg static facial expressions)
- Data obtained from theory and literature:

  (Keltner (1995); Shiota et al (2003); Rozin & Cohen (2003))



D. Keltner, B. N. Buswell
Embarrassment: Its Distinct Form and Appeasement Functions

Corpus of Emotional Displays

- **Annotation of audio-visual recordings from reality shows, hidden camera recordings, Belfast Naturalistic database, EmoTV corpus, DVD Baron-Cohen.**
- **Multimodal coding:**
  - Facial expression (FACS)
  - Gaze
  - Gesture
  - Body

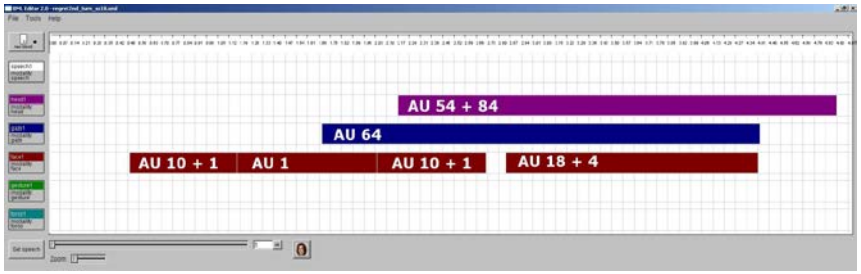
Happiness

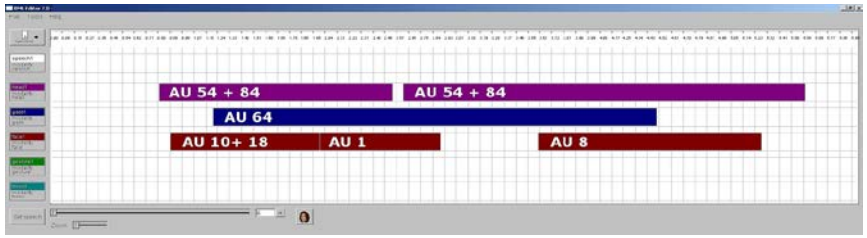
Emotional expression representation

- **Definition of a representation scheme for emotional expression:**
  - signals description across modalities
  - their partial temporal order
  - their constraints
- **Behavior set: set of signals (frown, nod, torso backward, …)**
- **Constraint set: describes the temporal and spatial constraints among the signals in the behavior set**
  - Temporal constraints: start-signal s*i* < end-signal s*j*
  - Spatial constraints: signals s*i* and s*j* cannot co-occur

Expressions of Emotions

Regret

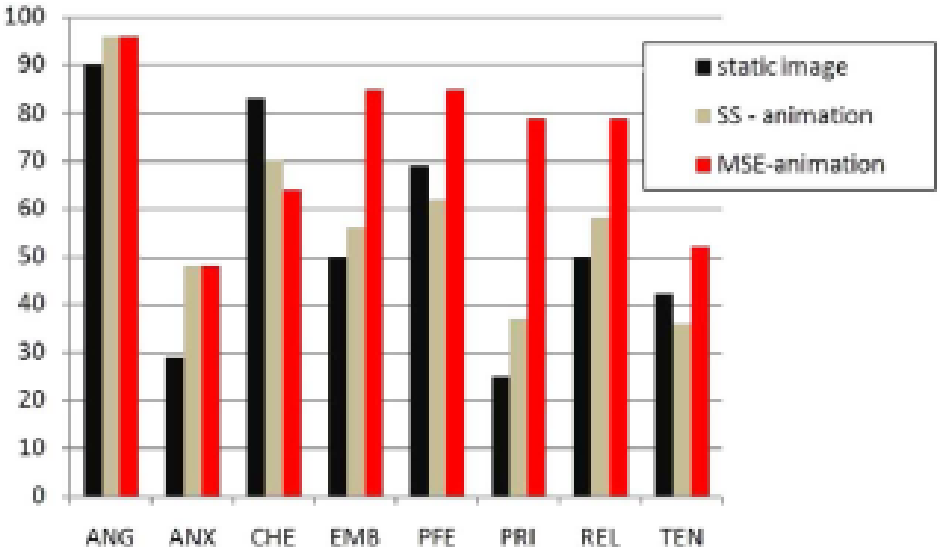




Niewiadomski

Validation study

- **Different conditions**
  - Static image
  - Single signal
  - MSE
- **Study**
  - 48 participants
  - Recognition task
  - 8 emotions:
    - anger, anxiety, cheerful, embarrassed, panic fear, pride, relief, tension
- **Higher recognition rate**
  - MSE
  - Except for Cheerfulness



Social Signals

- **Non adapted socio-emotional displays influence the user's evaluation of the agent negatively**
- **Appropriate socio-emotional displays increase the agent’s believability**

→ **importance of socio-emotional behaviors**

Social Signals

- **Non adapted socio-emotional displays influence the user's evaluation of the agent negatively**
- **Appropriate socio-emotional displays increase the agent's believability**

→ **importance of socio-emotional behaviors**

- **2 studies:**
  - Smile
  - Social Attitudes

Social Signals

- **Smile**

  well-recognized, rapidly perceived (Ekman)

- **Different meanings of smile (up to 50!)** (Hess)

  pleasure, amusement, politeness, embarrassment,…

- **Smile recognition: morphology and dynamism**
- **Effects of (positive) smile**
  - *Perception of a smiling human*: expressive, outgoing, relaxed, sociable, generous, trustworthy, warmer, attractive, intelligent, polite
  - *Perception of a smiling ECA*: convincing, credible, attractive, trustworthy
  - *Effects of smiling ECA on users*: motivation, enthusiasm, decision

(Rehm et al, 2005; Theonas et al. 2008; Krumhuber, 2007; Mehu, 2007; Frank et al. 1993)

Corpus-based approach: crowdsourcing

Magalie Ochs & Brian Ravenet

- **Corpus-based approach:** time consuming, expensive to collect and annotate
- **Crowdsourcing:**

  Outsourcing tasks to an undefined distributed group of people *(Mason et al., 2011)*

  ✓ Online tools for annotated human behaviors (Park et al., 2012)

Toolbox

↓

To create crowdsourcing tool

↓

To build rated corpus of virtual agent's non-verbal behaviors

Users directly configure the agent's non-verbal behavior

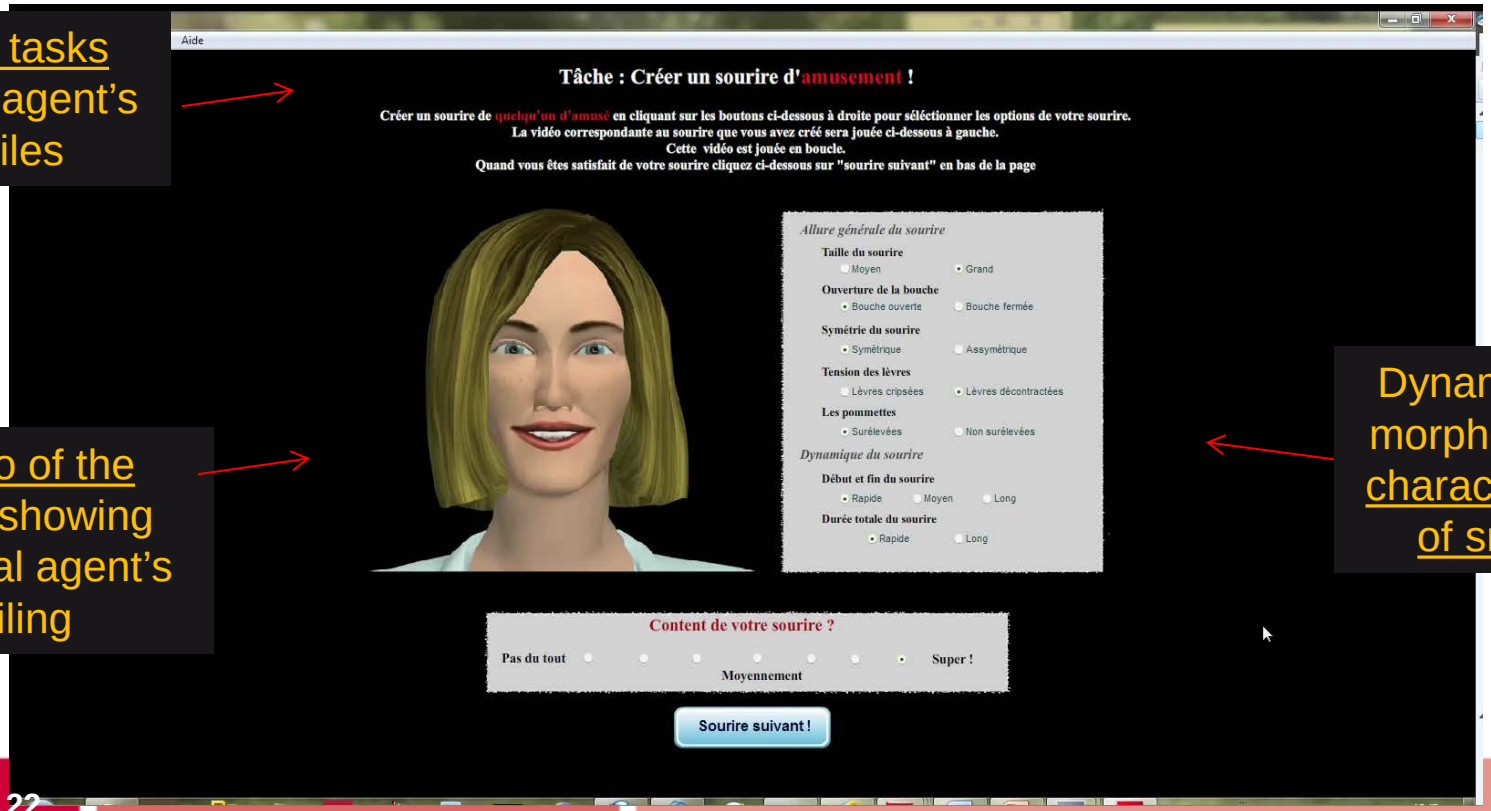
Smile (Ochs)
Social attitude (Ravenet)

Online crowdsourcing tools: *E-Smiles Creator*

- **Online crowdsourcing tool to <u>design virtual smiles</u>**
  - ✓ 192 videos of virtual agents smiles
  - ✓ Different morphological and dynamic characteristics

<u>Users tasks</u>
Design agent's smiles

A <u>video of the corpus</u> showing the virtual agent's smiling



Dynamic and morphological <u>characteristics of smiles</u>

Analysis of the collected data

Corpus of rated smiles (embarrassed, polite, amused)

↓

Oversampling with participants' satisfaction

↓

Machine Learning

↓

Decision Tree

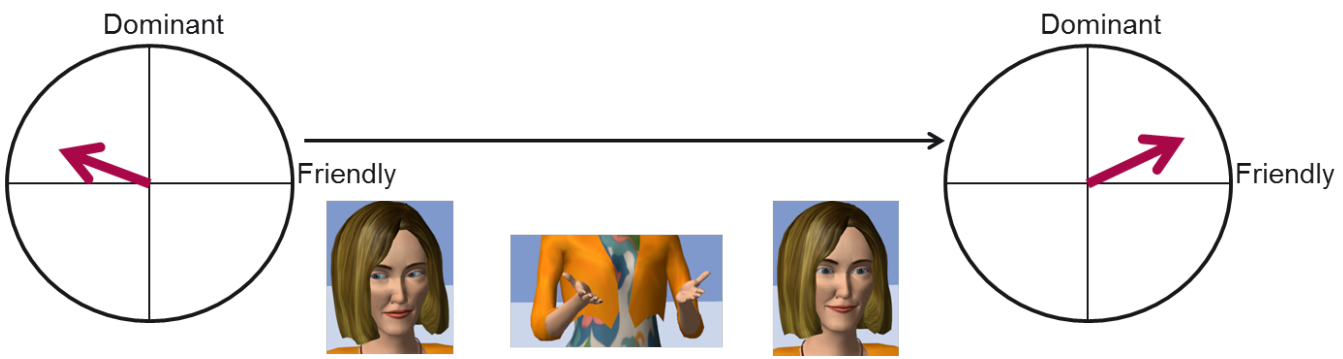
Social attitudes

Matthieu Chollet

- Interpersonal attitudes = « *spontaneous or strategically employed affective styles that colour interpersonal exchanges* » (Scherer, 05)
- Attitudes are expressed through multimodal behavior
  - Numerous gestures, head up… → dominance
  - Smiles, head tilt … → friendliness
- But the meaning of a signal can be changed by other nearby signals: *« No nonverbal cue is an 'island'. It is continually surrounded by a host of nonverbal behaviors which together may delimit and clarify meaning »* (Burgoon, 84)
  - Smile followed by a gaze and head aversion…
  - Frowning and crossing arms after a head tilt…

Objective and Methodology

- Develop attitude expression model for an Embodied Conversational Agent
  - Allow ECAs to vary their attitude through sequences of non-verbal behavior
  - Build sequential behavior planning model

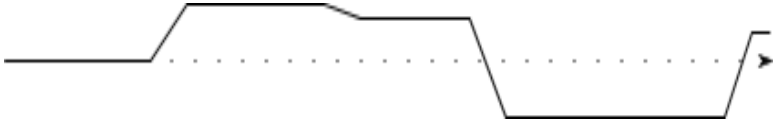


- Methodology
  1. Data collection on attitude expression
  2. Extraction of sequences of behavior conveying attitudes
  3. Implementation of a sequence generation model
  4. Validation

Mining a multimodal corpus for frequent non-verbal signals sequences

Non-verbal signals annotations

Attitude annotations



Non-verbal signals sequences mining methodology

NVB signals sequences expressing attitudes

Computational model of attitude expression

Sentence with communicative intentions (FML)

You are not fit for working with us.

DENY *EMPHASIS*

Attitude

Friendliness increase

Behavior planner
attitude expression -Bayesian network

Non-verbal signals sequence (BML)

Friendly

dominant

REALIZING HUMAN-AGENT INTERACTION:

SEMAINE PLATFORM

SEMAINE Project

SEMAINE
THE SENSITIVE AGENT PROJECT

- **“Sustained Emotionally coloured Machine-human Interaction using Nonverbal Expression”**
- **Sensitive Artificial Listener - SAL:**
  - Subjects dialog with characters
  - 4 Characters with 4 distinct emotional traits
  - Goal of the application is to induce emotions in subjects
- **Our goal: model interactive ECAs with different emotional traits that can:**
  - produce appropriate backchannels
  - sustain an emotionally colored communication with a user

TELECOM ParisTech

Conversation setup

SEMAINE
THE SENSITIVE AGENT PROJECT

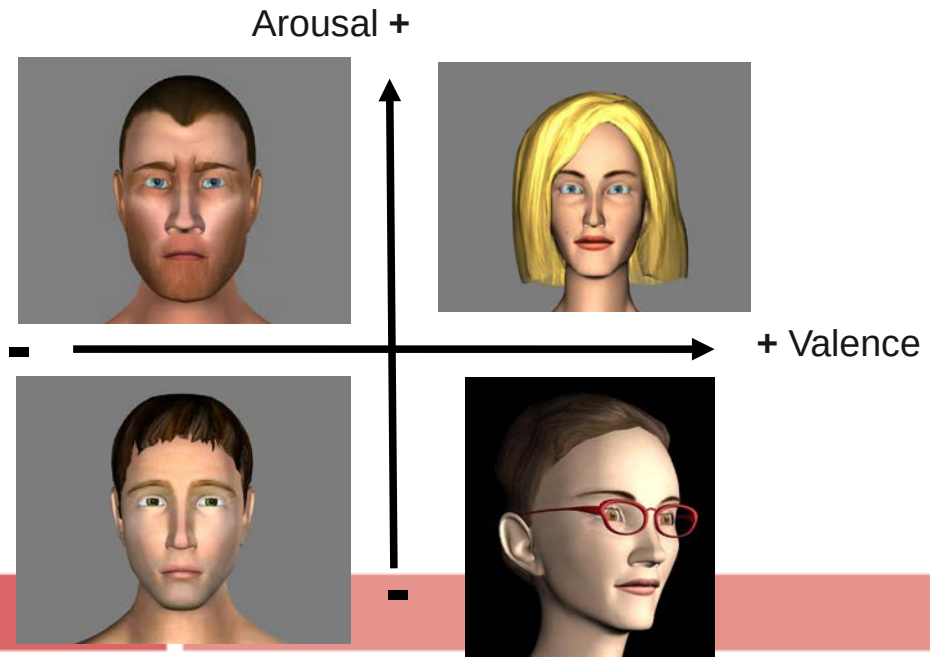
- One human, one ECA

- Analysis and synthesis of user's nonverbal behavior
- Interpretation of their epistemic & affective state

TELECOM ParisTech

The SAL concept

- **Four characters with an emotional agenda**
  - Obadiah tries to make users sad
  - Spike tries to make them angry
  - Poppy tries to make them happy
  - Prudence tries to make them reasonable



Building SAL characters

SEMAINE
THE SENSITIVE AGENT PROJECT

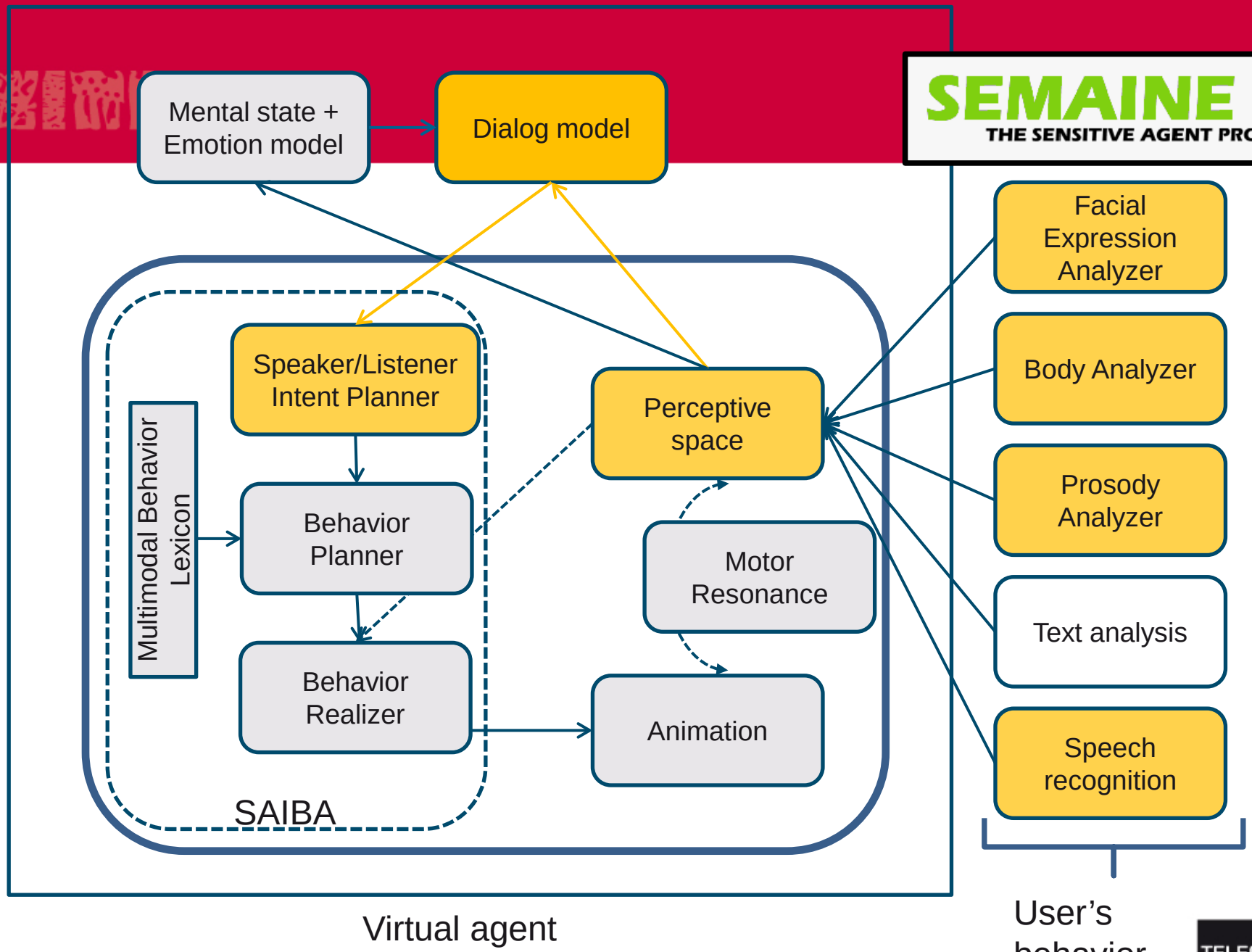
- **Consider trait models of personality**
- **Individual differences in predisposition to experience certain emotions**
- **Use of Eysenck's model:**
  - Neuroticism - negative emotionality
  - Extraversion - propensity towards positive emotion
- **Poppy is extraverted & optimistic**
- **Spike is angry & argumentative**
- **Prudence is pragmatic & practical**
- **Obadiah is gloomy & depressed**

TELECOM ParisTech

Building SAL characters

SEMAINE
THE SENSITIVE AGENT PROJECT

- **System utterances are designed to elicit emotional responses**
- **Baseline – agent's global behaviour tendency**
  - modality preference…
    - e.g. face, head, gaze, gesture, torso
  - …and behaviour expressivity
    - parameters influencing quality of movement - frequency, speed, spatial volume, energy, fluidity, & repetitivity
  - expressivity parameters defined for each modality



SEMAINE
THE SENSITIVE AGENT PROJECT

TELECOM ParisTech

OBADIAH

SEMAINE
THE SENSITIVE AGENT PROJECT

SPIKE

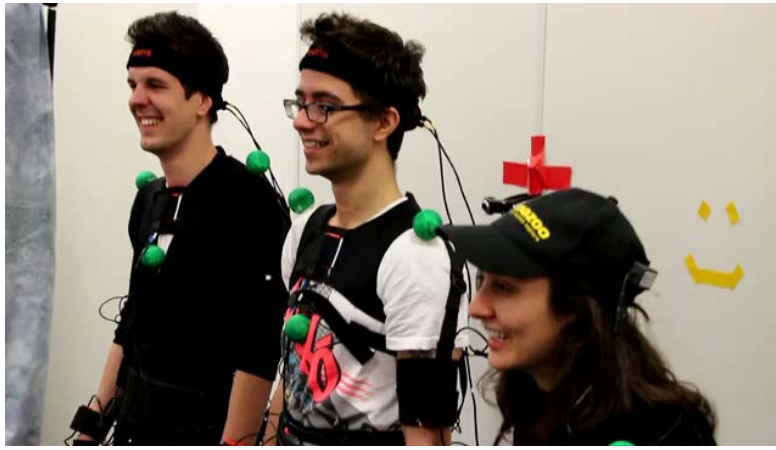
SEMAINE
THE SENSITIVE AGENT PROJECT

IMPACT OF AGENT'BEHAVIOR ON USER'S EXPERIENCE: LAUGHING AGENT

Laughter

- Laughter is an essential social signal in human-human communication.
- It can convey various functions:
  - Feedbacks to humorous stimuli or praised statements
  - Mask embarrassment
  - Social indicator of in-group belonging [Adelsward 1989]
  - Speech regulator during conversation [Provine 1996]
  - Elicit interlocutor's laughter [Provine 1996]
- Endow virtual agent with social attitudes, in particular laughing capabilities
- Focus here on hilarious laughter

Laughter



- **Multimodal signals**
  - Breathing, shoulder, body, head
  - Leaning backward and forward
  - Exhalation and inhalation
  - Rhythmic body pattern
  - Shaking movement
- **Different tasks in a social context**
  - Watching funny clips
  - Play social games
- **Dataset:**
  - Motion capture + audio
  - Segmentation into laughter episodes
  - Pitch and Energy extracted with PRAAT

- **Multimodal animation model of laughter:**
  - Data-driven approach: capture link between audio and multimodal behaviors
    - Input: laughter audio features streams [Urbain et al. 2013]
      - pitch, energy, 12 laughter pseudo-phonemes including duration and intensity
    - Output: data streams of facial/body animation points (FAPs/BAPs)

Laughter: Animation Generators

- **Modeling one dimensional shaking-like movement**
  - Capture rhythmic patterns during laughter [Ruch and Ekman 2001]

    Loop HMM

- **Taking into account the dependency with laughter audio**
  - Inspired by speech-to-animation generators [Hofer2007, Busso2005]

    Transition Parameterized Loop HMM

- **Joint modelling of head and torso motions**
  - Capture strong correlation between head and torso motions occurring in particular during exhalation phases [Ruch and Ekman 2001]

    Coupled Transition Parameterized Loop HMM

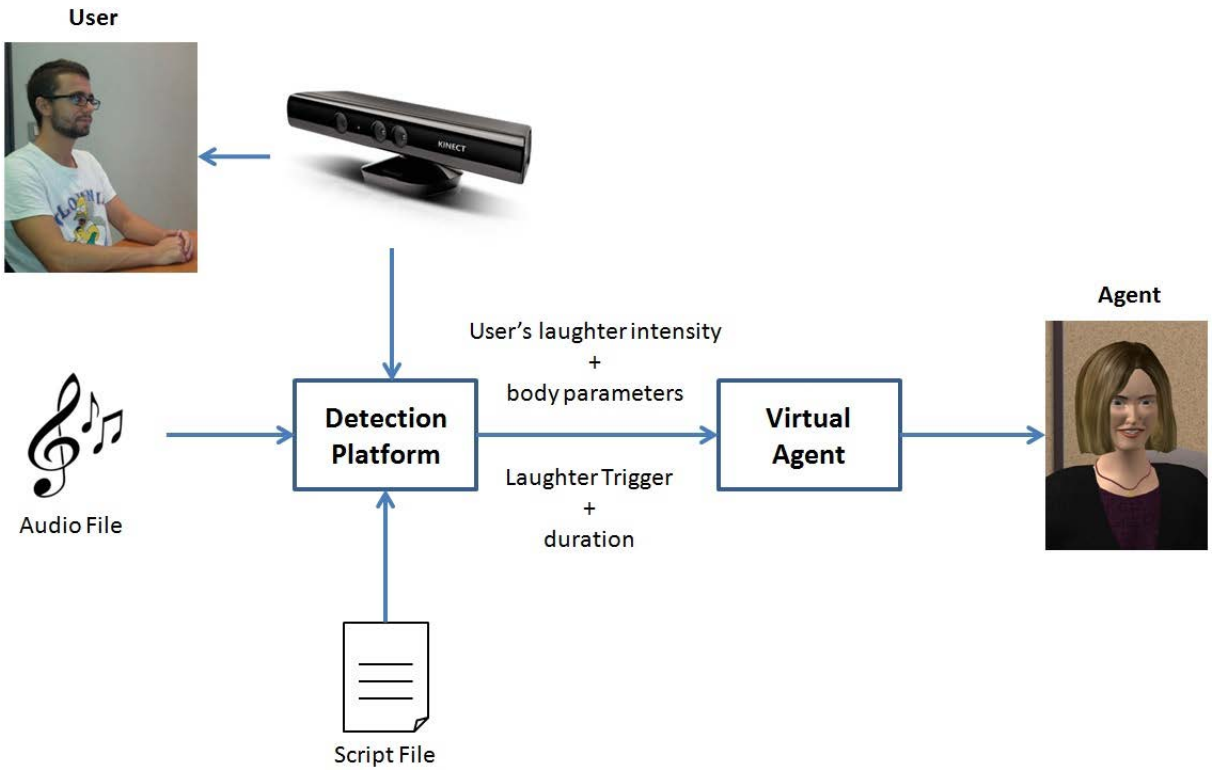
Video Example

LoL – Laugh out Loud

Pecune & Mancini

- Study impact of laughter on human experience during a human-machine interaction
- Effect of dynamic coupling between interactants → enhances sense of engagement (Prepin et al., 2012)
- Behavior expressivity conveys affective content (Castellano et al, 2012)
- Copying paradigm: copying the expressivity dynamically as it evolves in human's performance.
- AIM: Study how a virtual agent able to copy and to adapt its laughing and expressive behaviors on the fly participates in enhancing user's experience in the interaction.

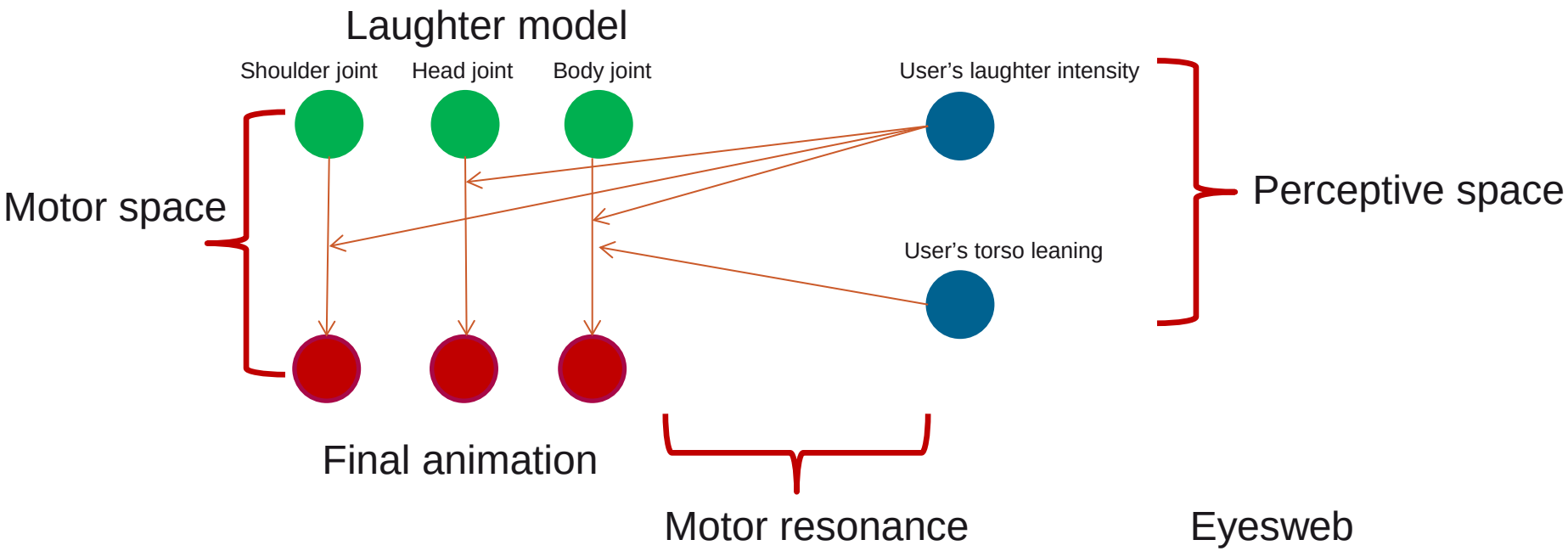
Architecture



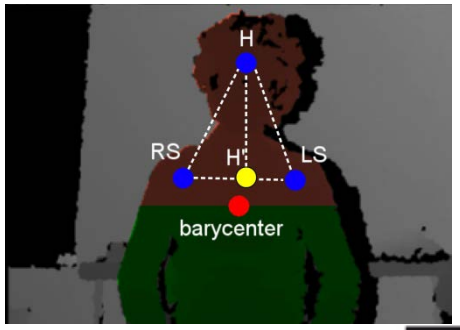
Music creation:

- Follows principles of P. Schickele, « P.D.Q. Bach »: Humor-Evoking Devices (D. Huron)

Copying paradigm

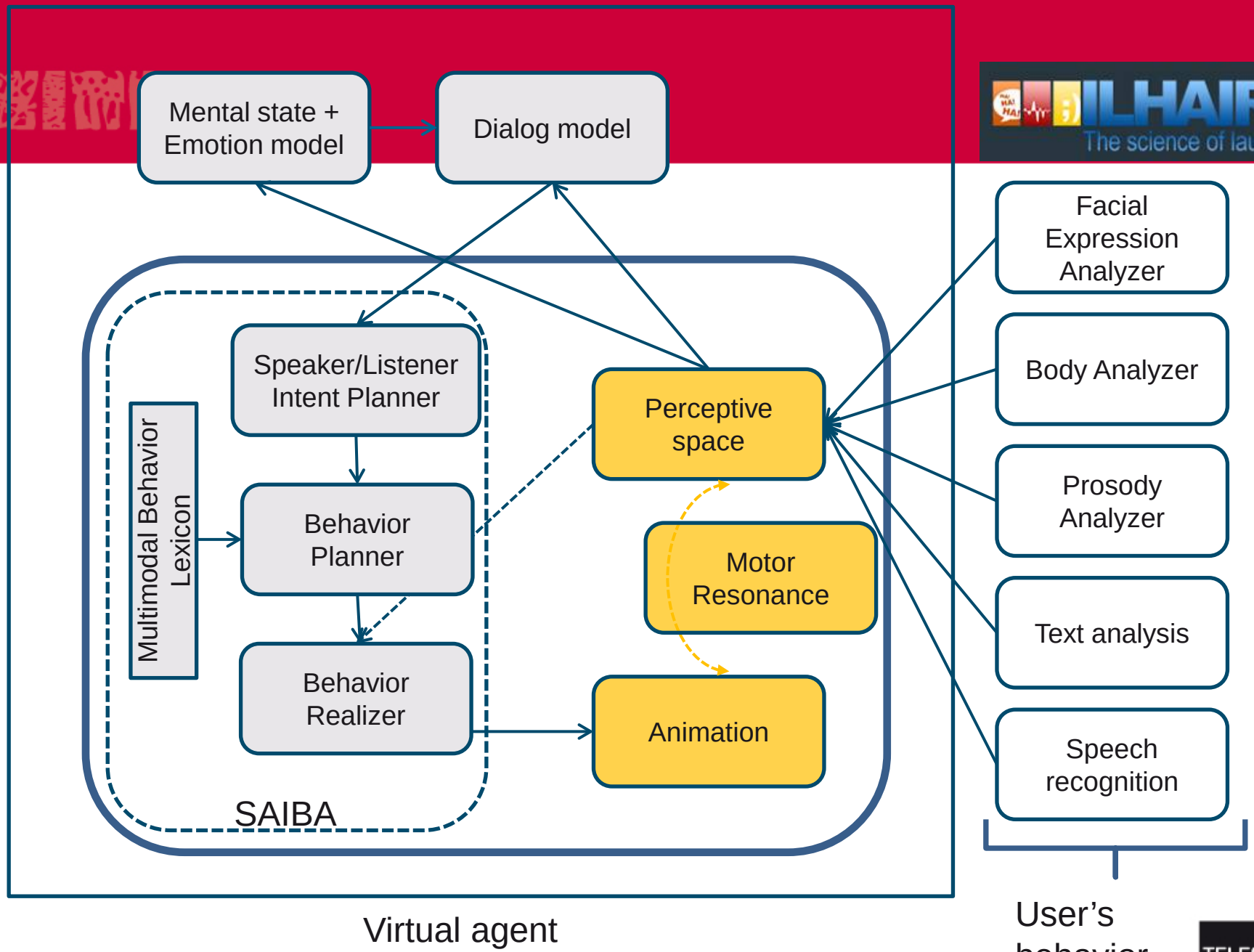


Eyesweb



Motor resonance: 2 influences

- Input computed with Eyesweb (Camurri et al.)
  - User's laughter intensity
  - User's torso leaning
- Output
  - Global influence on agent's laughter motion
  - Agent's leaning motion


Mental state + Emotion model
Dialog model
Speaker/Listener Intent Planner
Multimodal Behavior Lexicon
Behavior Planner
Behavior Realizer
SAIBA
Perceptive space
Motor Resonance
Animation
Facial Expression Analyzer
Body Analyzer
Prosody Analyzer
Text analysis
Speech recognition
Virtual agent
User's behavior detectors

LoL – Laughing out Loud

Video Example

LoL – Evaluation study

Context

- 32 participants listen to funny music first without then with a virtual agent. The agent can laugh performing a prefixed behavior, or copying user's laughter intensity.
- Evaluate user's perception of music funniness
- Evaluate user's mood during the experience
- Evaluate social and spatial presence, and believability of the virtual agent

Discussion

- Participants perceived the music as **funnier** when the agent **was present and copied** user's laughter intensity, than when they listened to the music alone;
- Participants' mood was **more positive** when the agent **was present and copied** user's laughter intensity, than when they listened to the music alone;
- Participant's mood was **more positive** when the agent laughed **mimicking** their laughter intensity, then when the agent laughed performing a prefixed behaviour.

APPLICATION: JOB INTERVIEW SIMULATOR

Tardis: Context

Young people Not in Education, Employment or Training (NEET) EU → ~20%.

- **18-25 y. old**
- **High risk of social exclusion**
- **Lack of social skills for job interviews**

Social Coaching

- **Enhancing skills in job interviews**
- **Time consuming and expensive!**
- **Fear of interacting with a human coach**

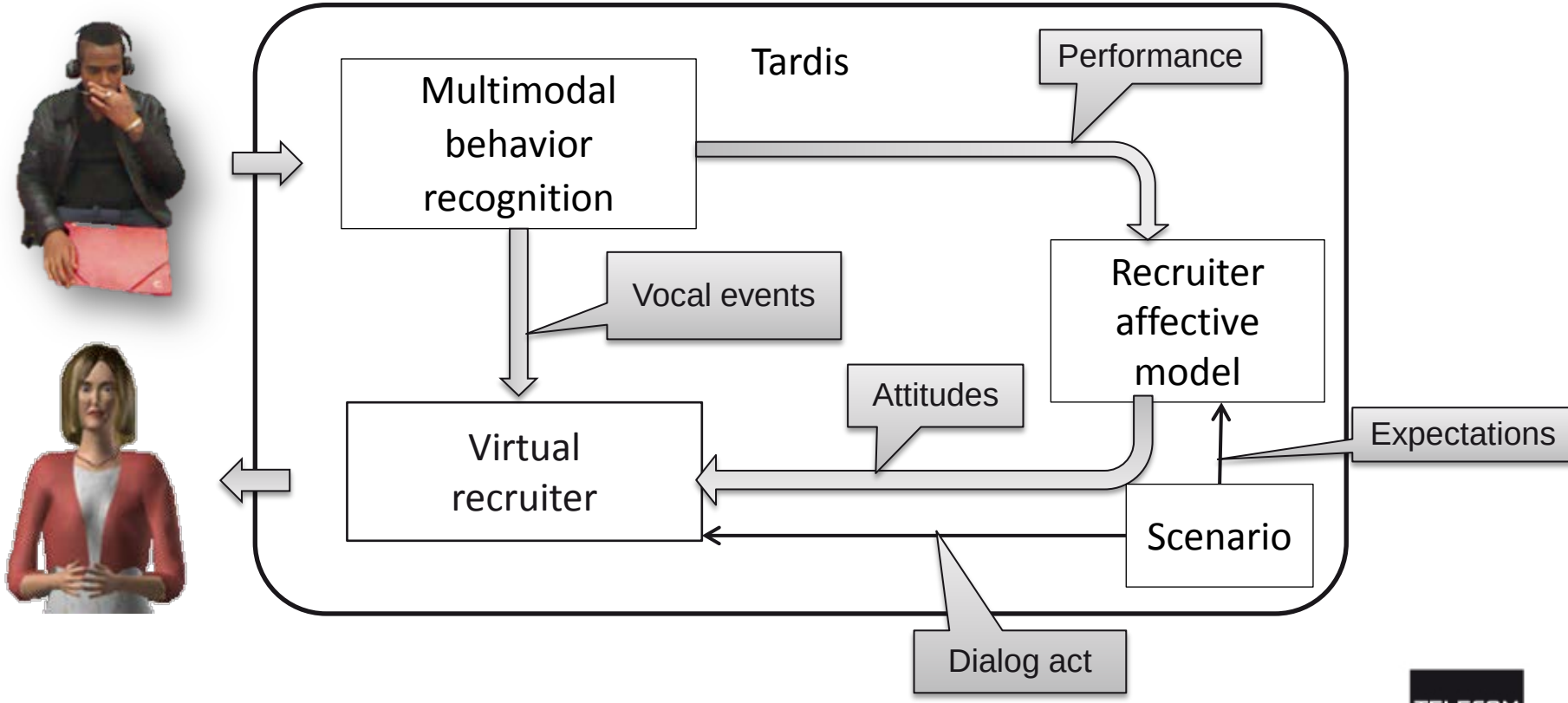
TARDIS platform:

A serious game for practising job interviews with a virtual agent

+ A social coaching tool for inclusion associations

Job interview simulator

- Tardis approach : 2 major components
  1. Job interview simulation
  2. Post-hoc debriefing



Sabouret et al.

Job interview simulator: evaluation

- Tardis job interview simulation platform in use in national associations in FR, UK, DE

- Does the Tardis approach help youngsters improve their job interview skills ?

(by UCL IOE partner of the Tardis project)

- Significant improvement in verbal and non-verbal behavior
  - Depth of answer
  - Eye contact, tone of voice, facial expressions

Conclusion

- Importance of socio-emotional signals
- Richness of the repertoire
- Interactive behaviors
- Dynamic coupling between interactants

Any questions?

Greta Team

CNRS LTCI – Télécom ParisTech - France