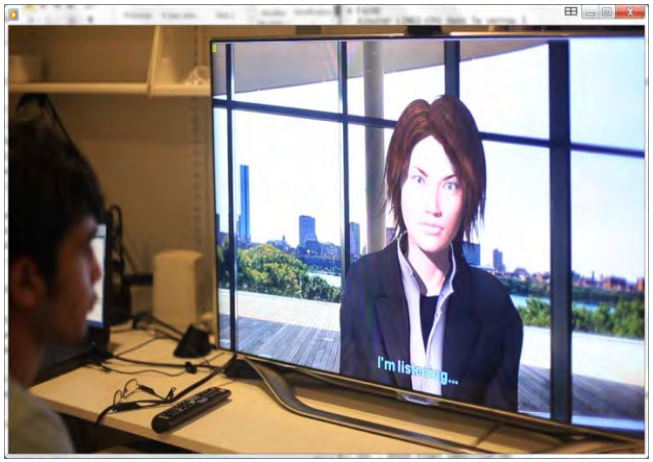
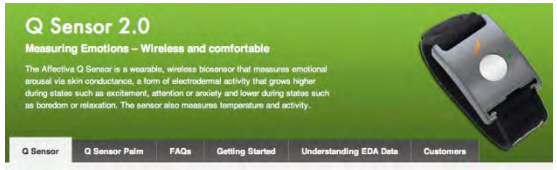
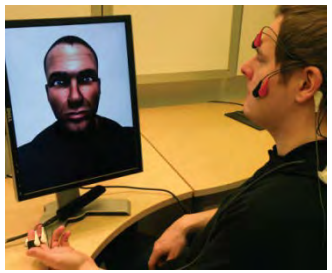
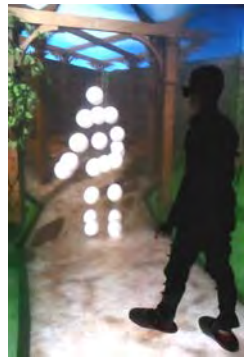
Collaborative Interactions with Virtual Humans

Jean-Claude MARTIN
Univ. Paris Sud / LIMSI-CNRS
MARTIN@LIMSI.FR









INTRODUCTION

Why use virtual humans in collaborative interactions?

- CSCW (cf slides by Michel BeaudoinLafon)
- Real humans are social beings
- Virtual humans are social beings too
  - Groups of virtual humans
  - Professional / private life (e.g. SIMS)
  - Division of labor (e.g. different virtual humans may have different roles)
- Some tasks require the user to be aware to be a part of a group of (real or virtual) humans
- Can virtual humans support cooperative work?

Why use virtual humans in collaborative interactions?

- Virtual humans can
  - Represent another user (Avatar)
    => supporting the user to control his own avatar
  - Be fully autonomous
  - Be semi-autonomous
- Interactions between virtual humans
  - Avatars
  - Autonomous
  - Mixed
- Applications
  - Virtual worlds (Second Life)
  - Training
  - Networked games
  - Mediated communication system
  - Simulations (e.g. crowds)

CSCW taxonomies applied to virtual humans

- Functional taxonomy
  - Communication
    - Multimodal
    - Embodied
    - Social => Emotions
  - Coordination => joint tasks and joint attention
    - Deictic gestures, gaze, proxemics
  - Sharing => via the virtual world
    - Social communication is situated

CSCW taxonomies applied to virtual humans

- Time-space taxonomy
  - Different cases
    - The user interacts with an autonomous virtual human
    - The user interacts with an avatar of another user
    - N users x N virtual humans

	Same place	Different place
Same time	face-to-face conversation	telephone call
Different time	Post-it note	letter

Groups of virtual humans

- Mediated between two humans
  - Bodily communication (CERV)
  - Affective avatars (ANR Project)
  - Email virtual messenger (Moore)
- Teams of pedagogical agents (Baylor)
  - Each agent has a different role: motivator, expert, …
  - Several agents with complementary roles better than a single agent with all the roles
- Teams of agents (Rist and André)
  - Provide complementary opinions about an idea or an object
  - Can have different social status
- Spatial organisation in small groups
  - F formations

Conversational settings for virtual characters

- (Rist et al. 2003)

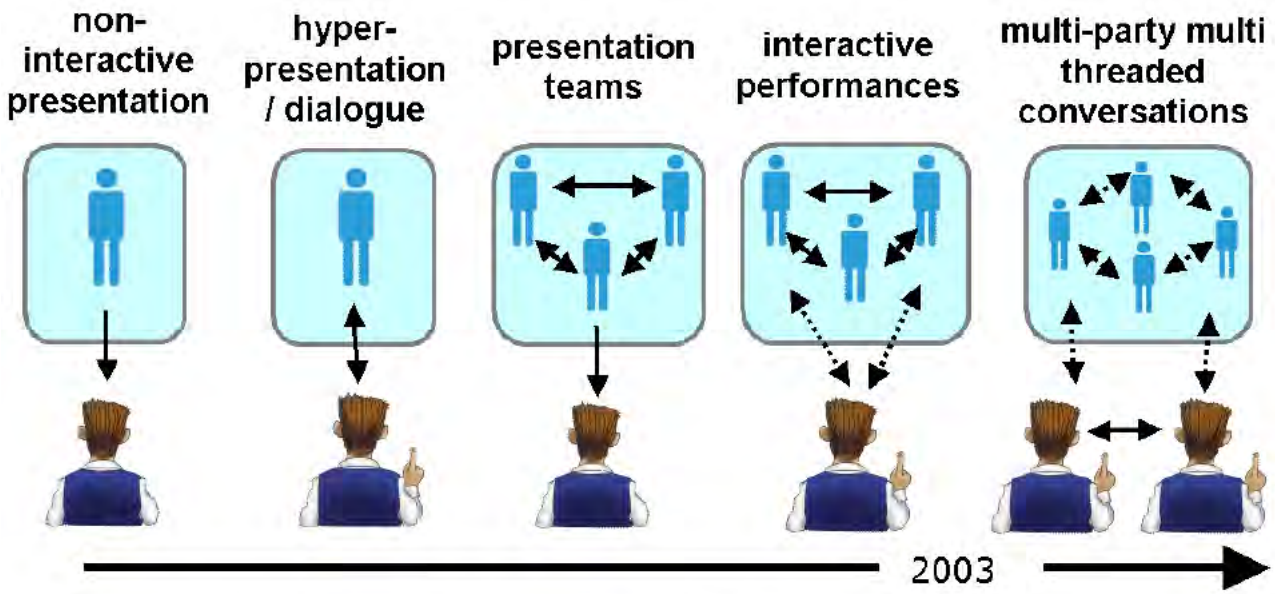


Fig. 1. Character applications with different conversational settings

Interaction – Controling the agent

- Manually controling the expressions of an agent
  - Virtual world
  - Arts

Jacquemin, 2007

Bee et al. 2009

Interaction – Autonomous Agents

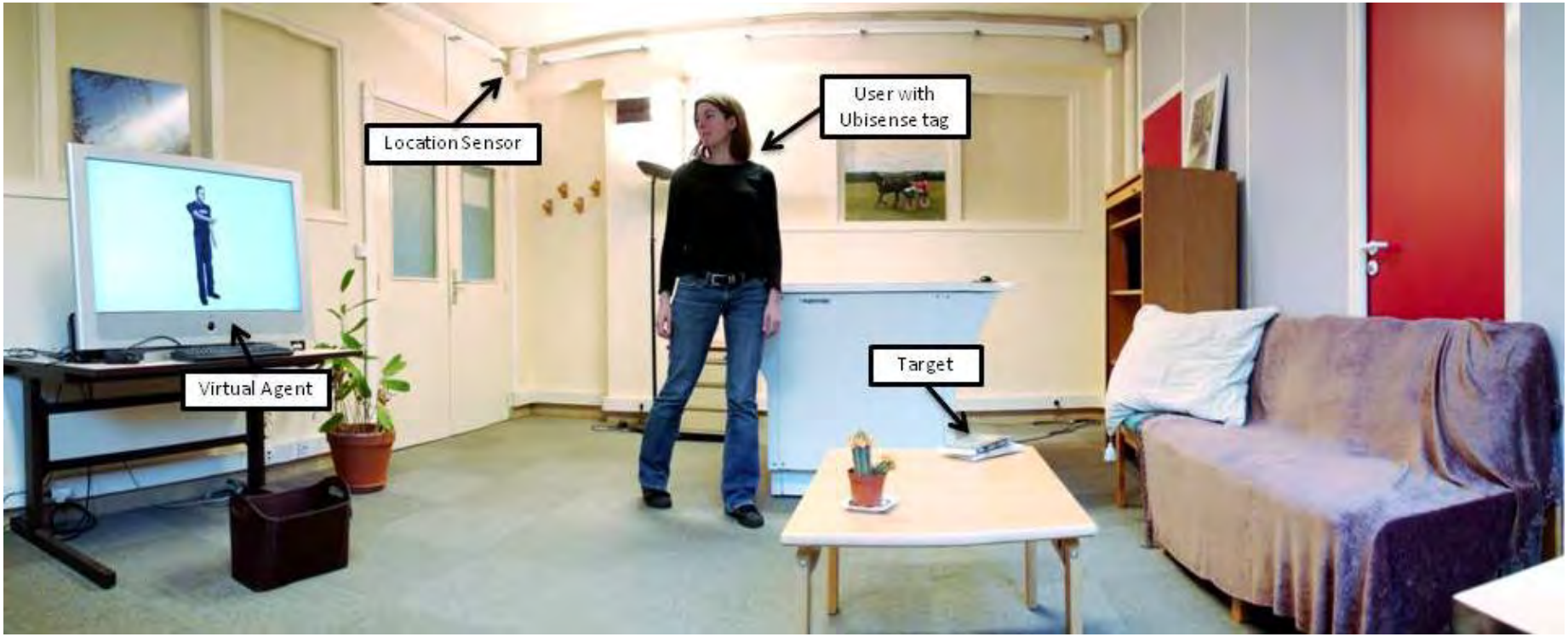
Jörding, Wachsmuth, Becker-Asano, 1997

MAX in the *Heinz Nixdorf MuseumForum* based on the WASABI model of emotions

Swartout et al. 2010

Ada & Grace in the Science museum at Boston

Experimental studies



- Ambient intelligence

Tan, N., G. Pruvost, M. Courgeon, C. Clavel and J.-C. Martin (2011).
A Location-Aware Virtual Character in a Smart Room: Effects on Performance, Presence and Adaptivity.
Intelligent User Interfaces (IUI'2011). Stanford University, Palo Alto, California, USA. Acceptance rate: 40%.

Outline

- VH, Multimodal interaction, applications
- Communicative functions
  - Emotions
  - Deictics and shared attention
  - Feedback (backchannels, formative, summative)
- Modalities
  - Facial animation
  - Bodily animation
- Presence, realism
- Future directions
  - Robotics
- Demo of the MARC platform

VIRTUAL HUMANS & MULTIMODAL USER INTERFACES

Long term vision
Intuitive Multimodal Interaction

(Kaiser et al. 03)

(Lester et al. 00)

(Wahlster 03)

Input

Fusion of speech & Gesture

Ex: references to objects,

PDA

Output

Embodied

Conversational Agent (ECA)

Ex: pedagogical agent:

motivation

via facial expressions

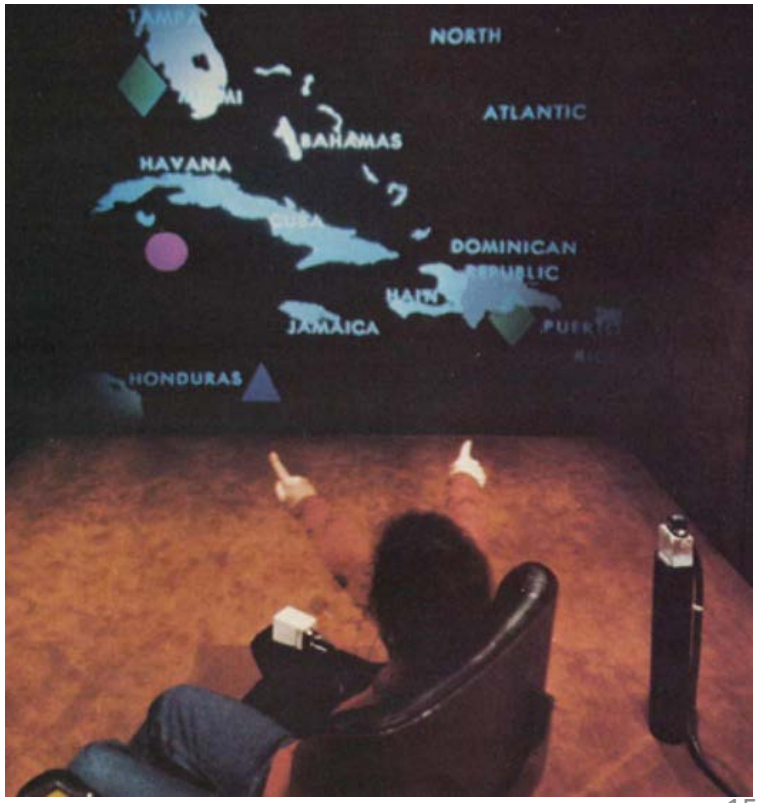
Bidirectional

Seamless real-time coordination of intertwined combinations of input / output

Require representations of multimodal behaviors

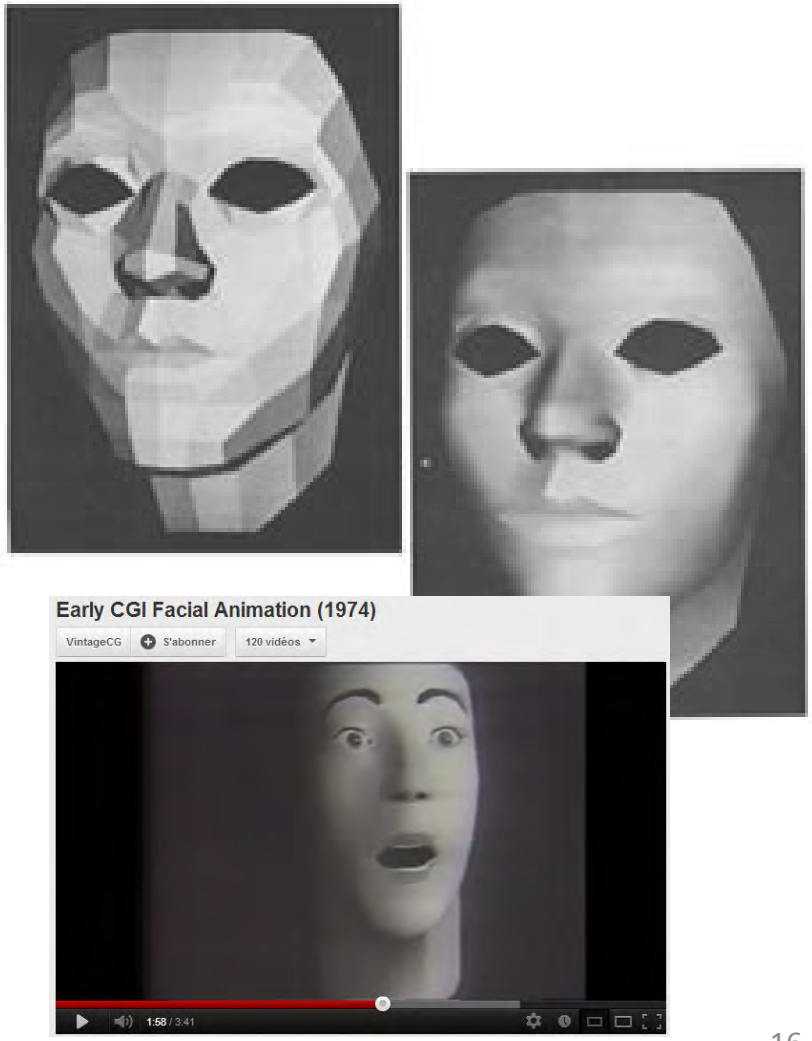
1980 – Bolt

- « Put that there »
- Media / modalities
  - Speech
  - Gesture (joystick + pad), sensor on wrist
  - Objects attributes:
    - color, size, shape
- Application
  - 2D map / 2D shapes



1974 - Parke

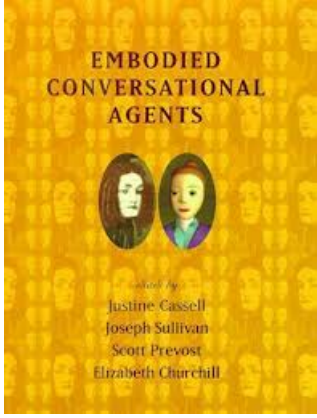
- In 1974, Parke developed a parameterized 3D facial model



Different names …

- … Different meanings with different focus
  - Avatar
    - Representing a user
    - Autonomous avatar
  - Virtual agent
    - Intelligent virtual agent (IVA)
    - Expressive virtual agent
  - Animated agent
    - Animated interface agents
  - Autonomous agent
  - Embodied Conversational Agents (ECA)

2000 - Cassell



Communicative Functions	Communicative Behavior
Initiation and termination:	
Reacting	Short Glance
Inviting Contact	Sustained Glance, Smile
Distance Salutation	Looking, Head Toss/Nod, Raise Eyebrows, Wave, Smile
Close Salutation	Looking, Head Nod, Embrace or Handshake, Smile
Break Away	Glance Around
Farewell	Looking, Head Nod, Wave
Turn-Taking	
Give Turn	Looking, Raise Eyebrows (followed by silence)
Wanting Turn	Raise Hands into gesture space
Take Turn	Glance Away, Start talking
Feedback	
Request Feedback	Looking, Raise Eyebrows
Give Feedback	Looking, Head Nod

Table 1. Some examples of conversational functions and their behavior realization(Cassell, Bickmore, Vilhjálmsson, & Yan, 2000)

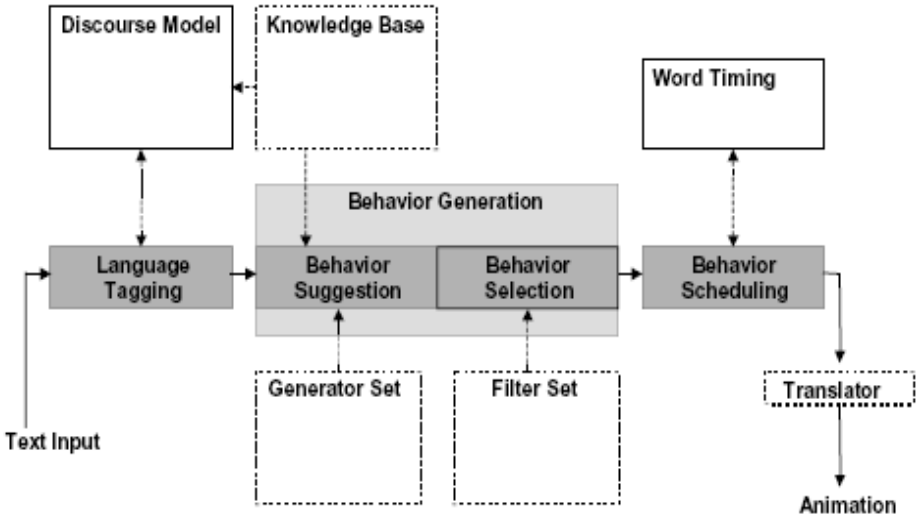


Figure 1. BEAT System Architecture

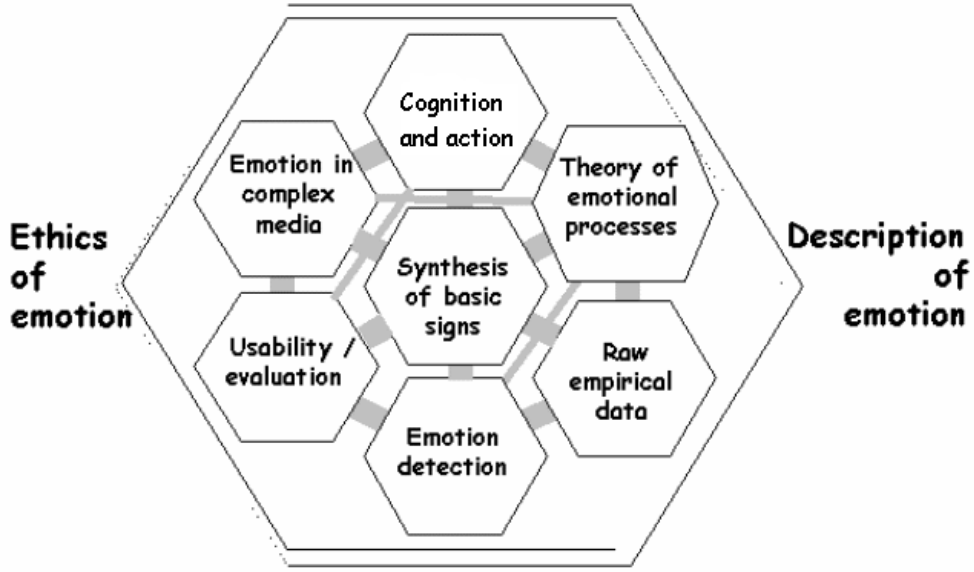
1997 - Affective Computing

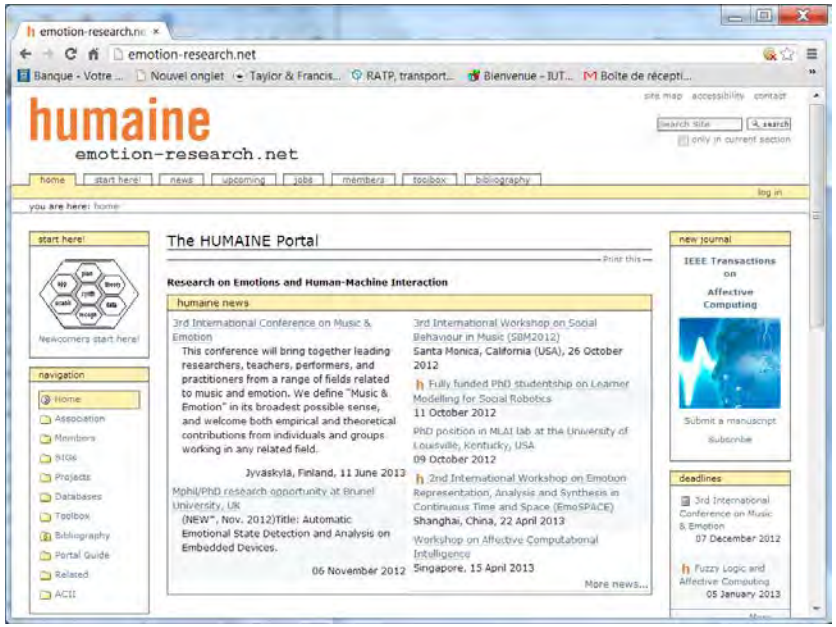
- Picard 1997,
  MIT Media Lab
  - the study and development of systems and devices that can recognize, interpret, process, and simulate human affects

2005 - HUMAINE

Research on Emotions and Human-Machine Interaction

- Network of Excellence
- International association





Applications: pedagogical agents

- Canned animations
- Research interactive prototypes

Table 1. **Sample conversation between AutoTutor and a student**

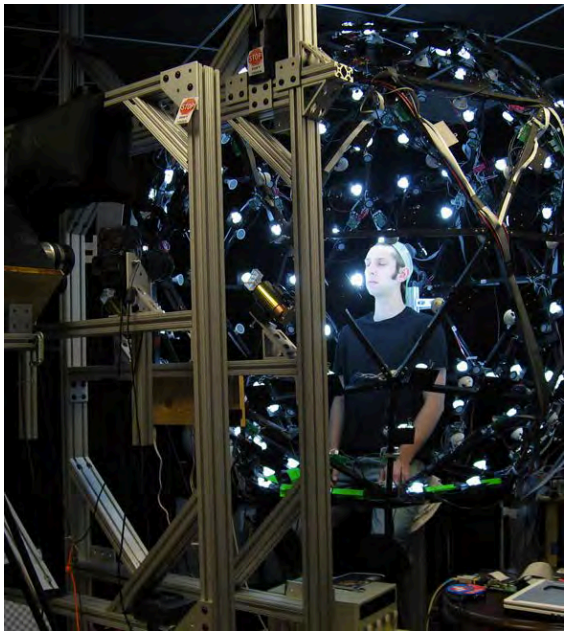
Speaker	Text
AutoTutor	When you turn on the computer, how is the operating system first activated and loaded into RAM? [Main question]
Student	*When the computer is turned on a file is automatically booted up.*
AutoTutor	Anything else? [Pump for information]
Student	*This file contains the operating system.*
AutoTutor	What about storage? [Hint]
Student	*The operating system is then read into RAM and activated.*
AutoTutor	Right. [Positive feedback]. The CPU executes instructions in ROM, which helps the operating system to start. [Assertion]

Applications: therapies and pedagogy

Virtual patients for training doctors

Applications: movie industry

- Based on original research led by Debevec at the University of California at Berkeley and published at the 2000 SIGGRAPH conference, the Light Stage systems efficiently capture how an actor's face appears when lit from every possible lighting direction.
- From this captured imagery, specialized algorithms create realistic virtual renditions of the actor in the illumination of any location or set, faithfully reproducing the color, texture, shine, shading, and translucency of the actor's skin.



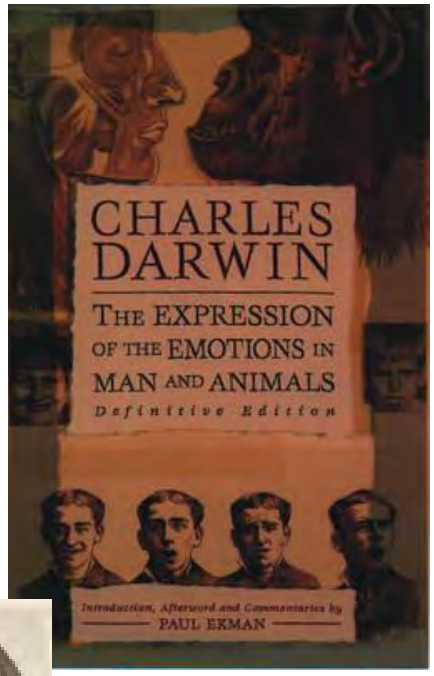
YOUTUBE
Rise of the Planet of the Apes (2011) HD Weta Featurette Making Of Behind the Scenes

FACIAL ANIMATION

Motivation

- Facial expressions are important in everyday interactions
  - Rich and subtle expressions
  - Convey multiple communicative functions
    - emotions, emphasis, turn management, …
- Multiple applications of facial expressions by virtual humans
  - Games
  - Virtual Worlds, Mediated communication
  - Assistance
  - Training
  - Deception detection
  - Therapy

Darwin (1809

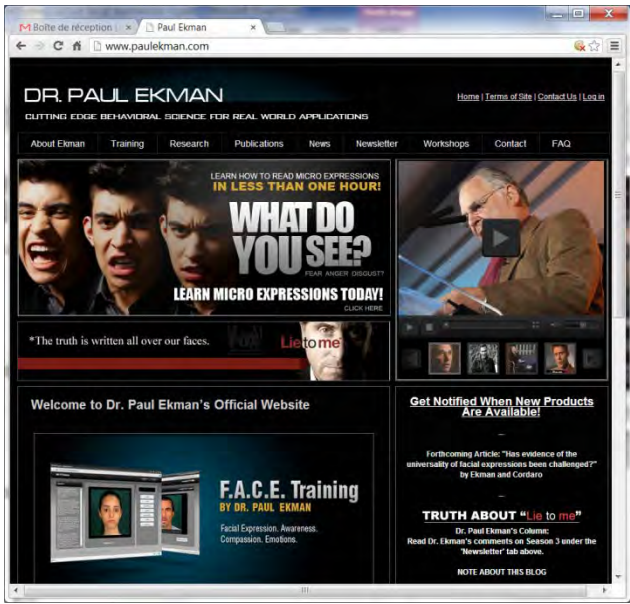


- Darwin traced the animal origins of human characteristics, such as the pursing of the lips in concentration and the tightening of the muscles around the eyes in anger and efforts of memory.
- Believes that emotions are biologically determined and universal to human culture
- Outline
  - General principles of expressions
  - Means of expression in animals
  - Special expressions of animals
  - Special expressions of man
  - Low spirits, anxiety, grief, dejection, despair
  - Joy, high spirits, love, tender feelings, devotion
  - Reflexion, meditation, ill, temper, sulkiness, determination
  - Hatred and anger
  - Disdain, contempt, disgust, guilt, pride, helplessness, patience, affirmation, negation
  - Surprise, astonishment, fear, horror
  - Self, attention, shame, shyness, modesty: blushing

Darwin(1872) The expression of emotions in man and animals, Londres, John Murray, 1872

Ekman (1934 -)

- Measuring facial expressions
- Emotions as universal categories
- His definition of a basic emotion
- His list of basic emotions
  - Anger, Disgust, Fear, Happiness, Sadness, Surprise
- 1 emotion = a familly of facial expressions
- Psychometric tests and training
- Detecting deception



Ekman, Paul (1999), "Basic Emotions", in Dalgleish, T; Power, M, Handbook of Cognition and Emotion, Sussex, UK: John Wiley & Sons

Facial Action Coding System

- The system was developed by Ekman and Friesen, in 1978
- FACS describes facial deformations in terms of “Action Units” (AUs)
- 1 AU can involve several muscles
  - AU 1 is the action of raising the Inner Brow. It is caused by the Frontalis and Pars Medialis muscles
- 1 muscle can involve several Aus
- Some AUs correspond directly to actions of facial muscles others involve movement of the tongue or air filling the cheeks, …
- AUs may be combined to describe any facial expressions

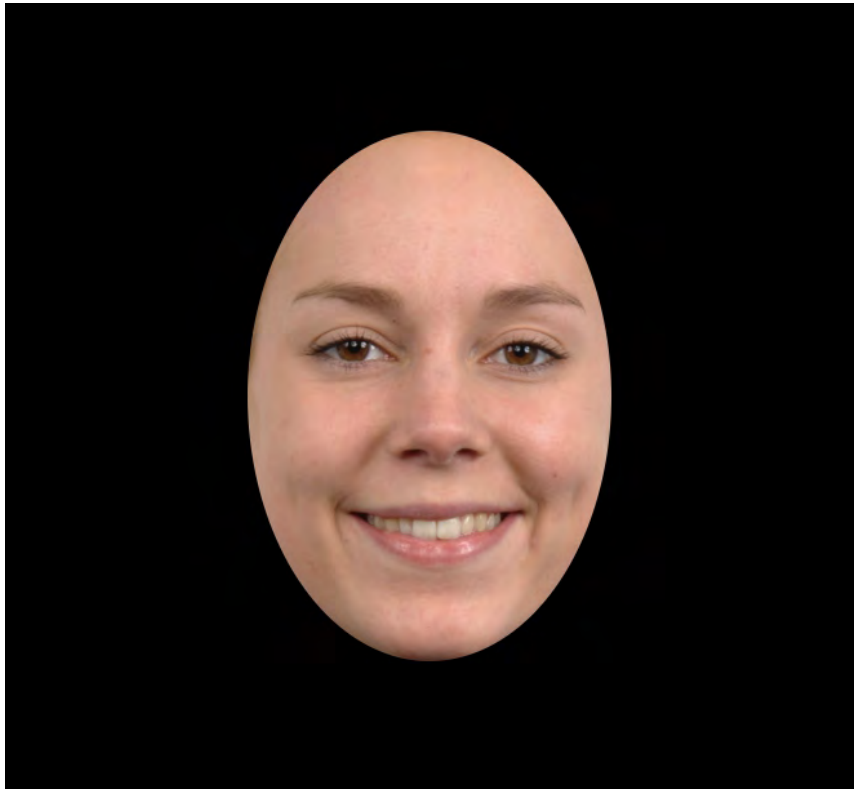
Paul Ekman, Wallace V. Friesen, and Joseph C. Hager. Facial Action Coding System: The Manual on CD ROM. A Human Face, Salt Lake City, 2002

Main AUs

Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
					
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
					
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

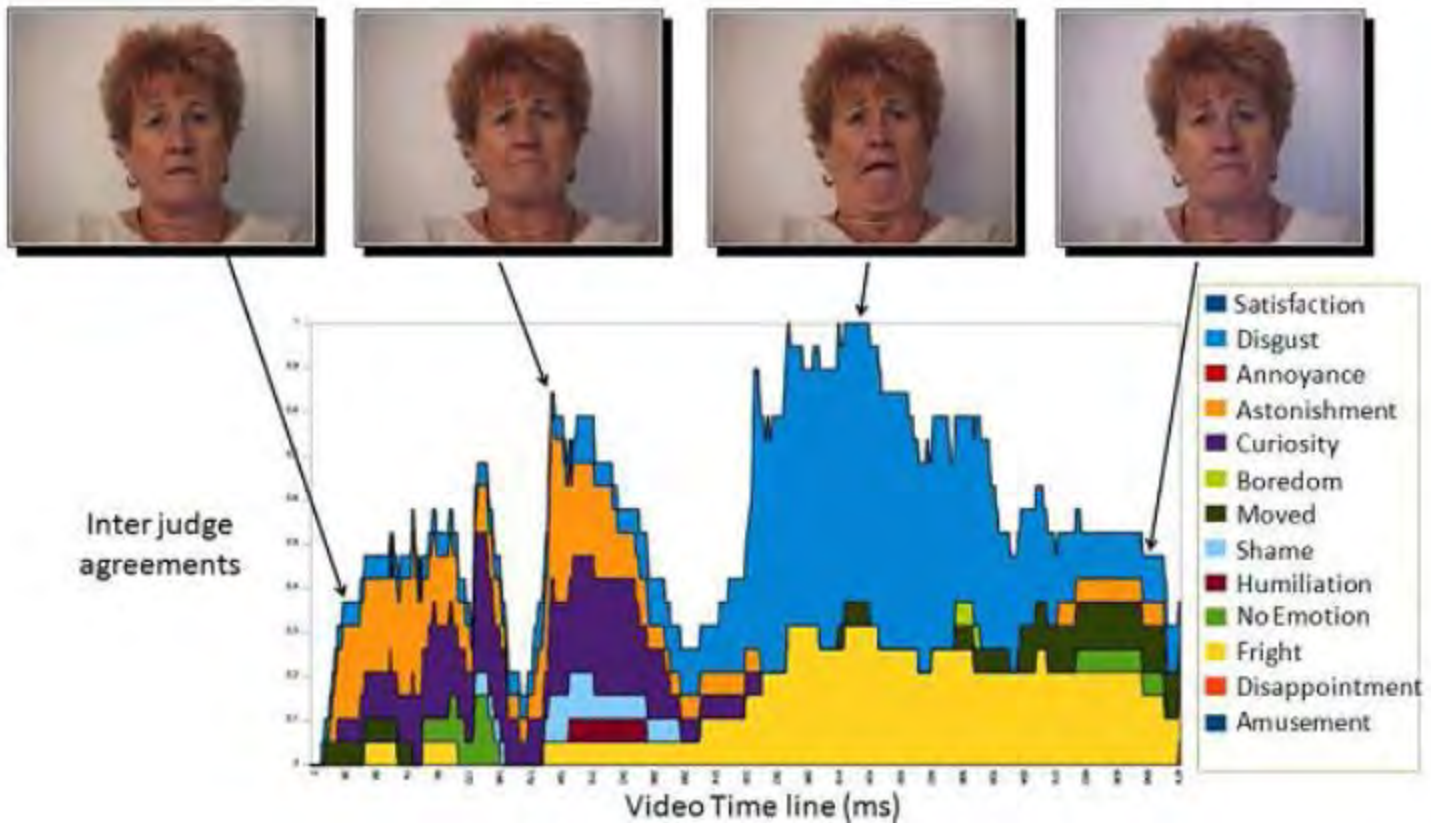
Databases of facial expressions

- Mostly databases of acted static facial expressions
- Ratings
  - Emotion category recognised above 80%
- Experimental studies



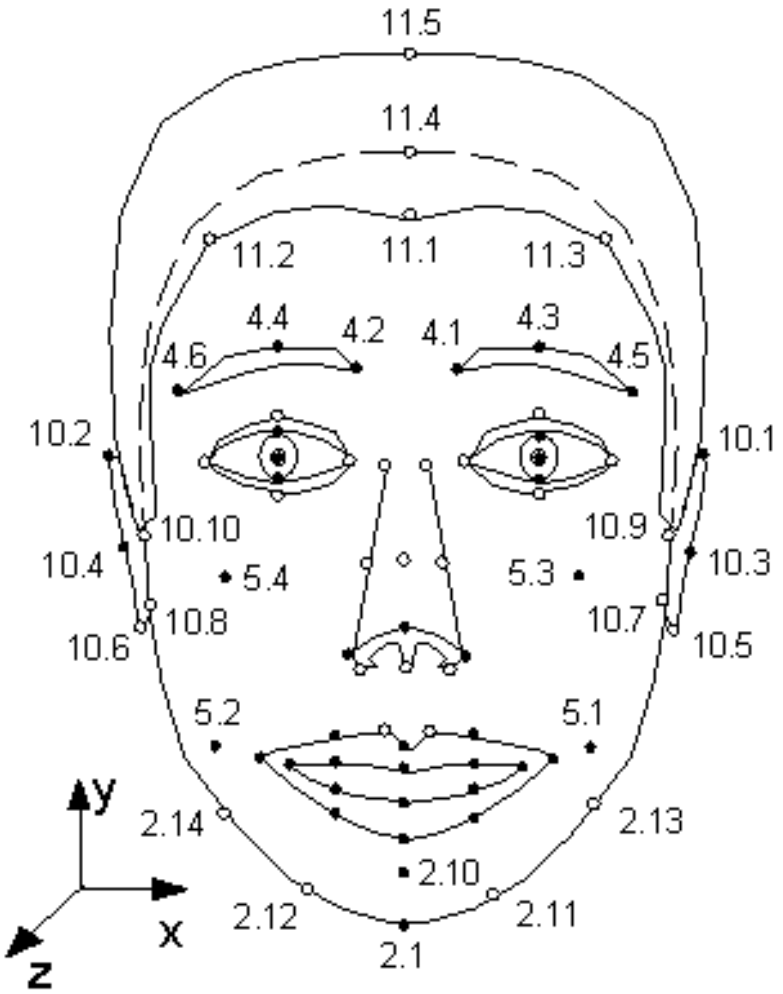
Databases of facial expressions

- Recent databases of dynamic spontaneous expressions (DYNEMO) (Tcherkassof et al.)



MPEG-4 Facial Animation (since 1999)

- 68 facial action parameters (FAPs), defined in terms of face independent FAP units (FAPUs)
- Most define a rotation or translation of one or more feature points
- Same animation can be used on different model, provided the model is properly annotated



Some MPEG-4 feature points

MPEG-4 and FACS

- MPEG-4
  - mainly focuses on facial expression synthesis and animation
  - defines the Facial Animation parameters (FAPs) that are strongly related to the Action Units (AUs)

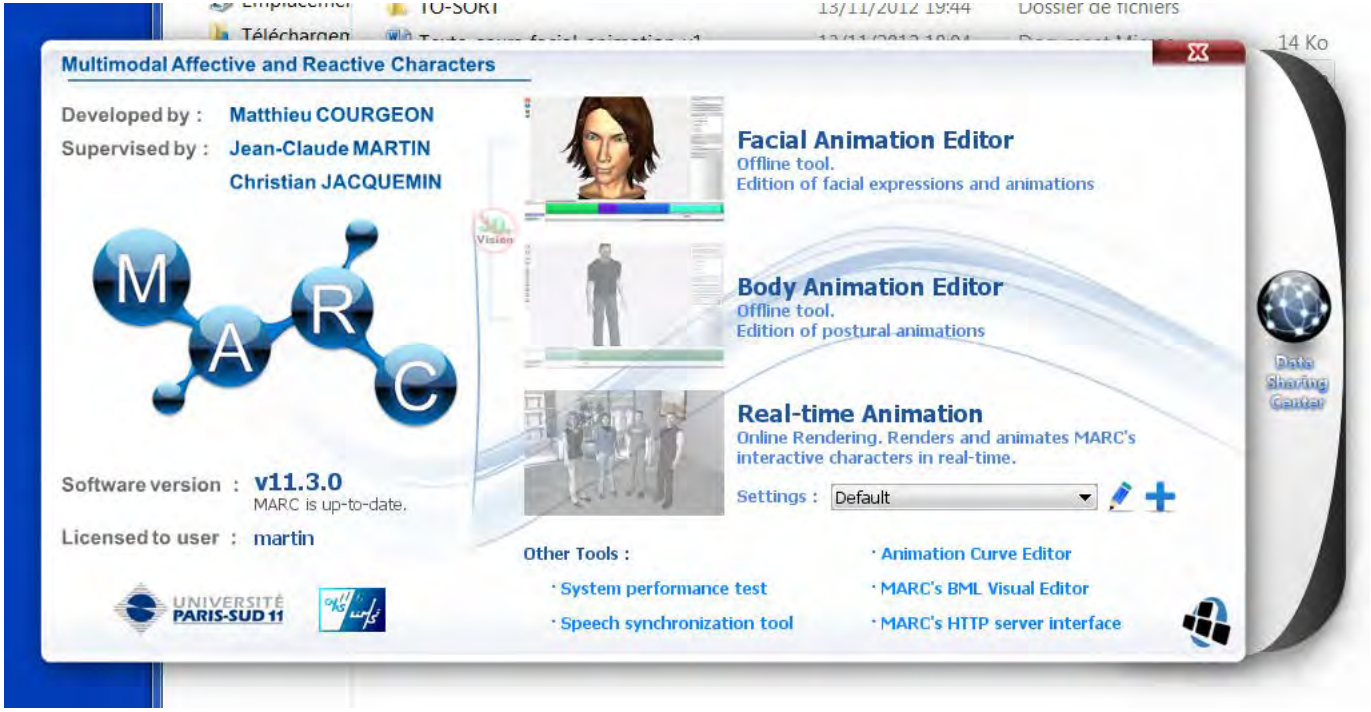
Action Units (FACS AUs)	Facial Action Parameters (MPEG-4 FAPs)
AU1 (Inner Brow Raise)	raise_l_i_eyebrow + raise_r_i_eyebrow
AU2 (Outer Brow Raise)	raise_l_o_eyebrow + raise_r_o_eyebrow
AU9 (Nose Wrinkle)	lower_t_midlip + raise_nose + stretch_l_nose + stretch_r_nose
AU15 (Lip Corner Depressor)	lower_l_cornerlip + lower_r_cornerlip

Table 2: Some of the AUs and their mapping to FAPs [35]

The MARC Platform

- Multimodal Affective and Reactive Character
- Plateform for facial animation
- Libraries of facial expressions
  - Basic emotions (Ekman and Friesen 75)
  - Complex emotions (Baron-Cohen, 2007)
- Eplore the impact of visual realism
  - Realistic rendering (D'eon et Luebke, 2007)
- Real-time interaction with the agent

Facial Animation Editor



EMOTIONS

Why are emotions important for virtual humans?

- Emotions pervades human life everyday
- In order to be « credible », virtual humans need to be able to
  - Recognise user's emotions
  - Express emotions via facial expressions, …
  - Reason about emotions
- Virtual humans' emotions should be properly selected and expressed:
  - Select an appropriate emotion(s) for a given context
  - Select an appropriate intensity
  - Select the appropriate expression(s) in the appropriate modaliti(es)

Multiple definitions (Scherer 00)

- Definition
  - An emotion is an hypothetical construct denoting a **process** of an organism's reaction to significant events
  - An emotion is an episode of interrelated, **synchronized** changes in 5 components in response to an **event** of major significance to the organism
  - **The 5 components of emotions are**
    - subjective feeling
    - motor expression
    - action tendencies
    - physiological changes
    - cognitive processing
- Functions of emotions
  - Expression: Communication, social signalling and interaction strategy
  - Feeling: facilitates self regulation of emotional behavior
  - Adaptation to changes

Emotion theories

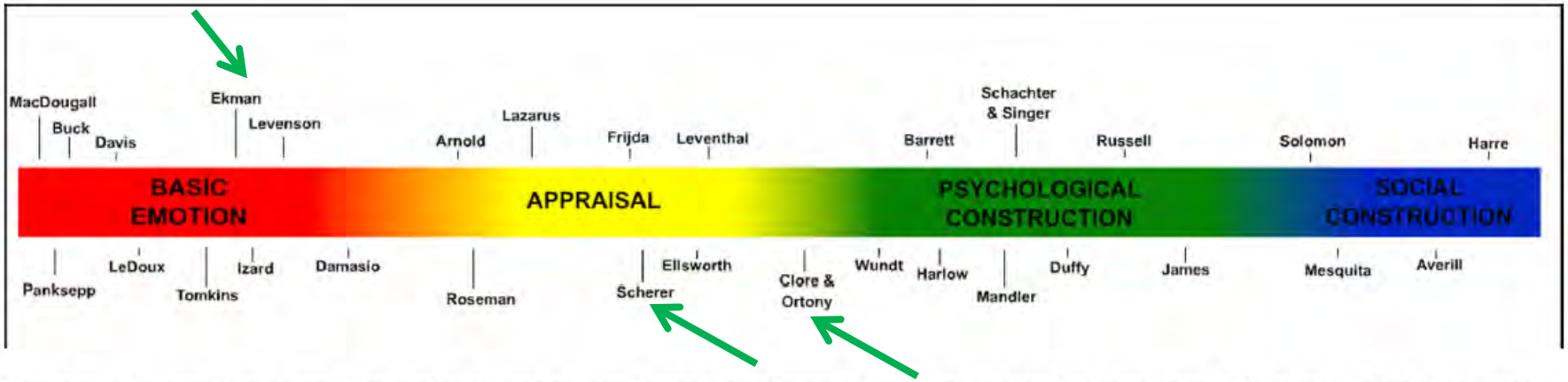


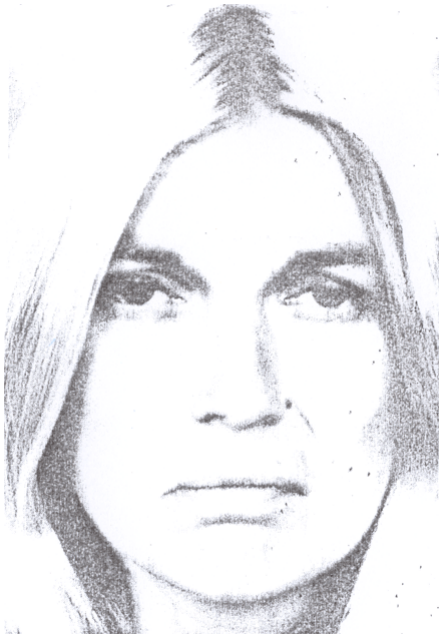
Figure 1. Perspectives on emotion can be loosely arranged along a continuum. We have populated this continuum with representative theorists/researchers drawn from the field of psychology. We distinguish four "zones": (1) basic emotion, in red, e.g., MacDougall (1908/1921), Panksepp (1998), Buck (1999), Davis (1992), LeDoux (2000), Tomkins (1962, 1963), Ekman (1972), Izard (1993), Levenson (1994), and Damasio (1999); (2) appraisal, in yellow, e.g., Arnold (1960a, 1960b), Roseman (1991), Lazarus (1991), Frijda (1986), Scherer (1984), Smith and Ellsworth (1985), Leventhal (1984), and Clore and Ortony (2008); (3) psychological construction, in green, e.g., Wundt (1897/1998), Barrett (2009), Harlow and Stagner (1933), Mandler (1975), Schachter and Singer (1962), Duffy (1941); Russell (2003), and James (1884); (4) social construction, in blue, e.g., Solomon (2003), Mesquita (2010), Averill (1980), and Harré (1986). Given space constraints, as well as the goals of this article, we have limited ourselves to a subset of the many theorists/researchers who might have been included on this continuum (e.g., those who only study one aspect of emotion were not included in this figure).

Gross, J. J. and L. F. Barrett (2011). "Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View." Emotion Review **3(8).**

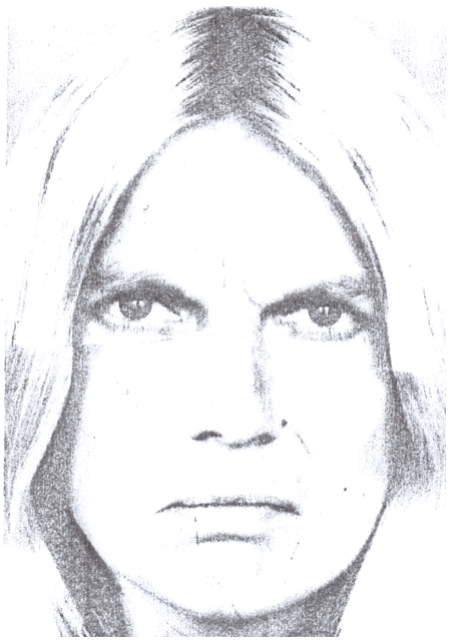
Blends of emotion categories (Ekman 75)

- Theories (Ekman 75, Plutchik 80, Ekman 92, Scherer 84)
- Lost luggage study (Scherer 98)
- Situations involving multiple emotions

Anger



Sadness

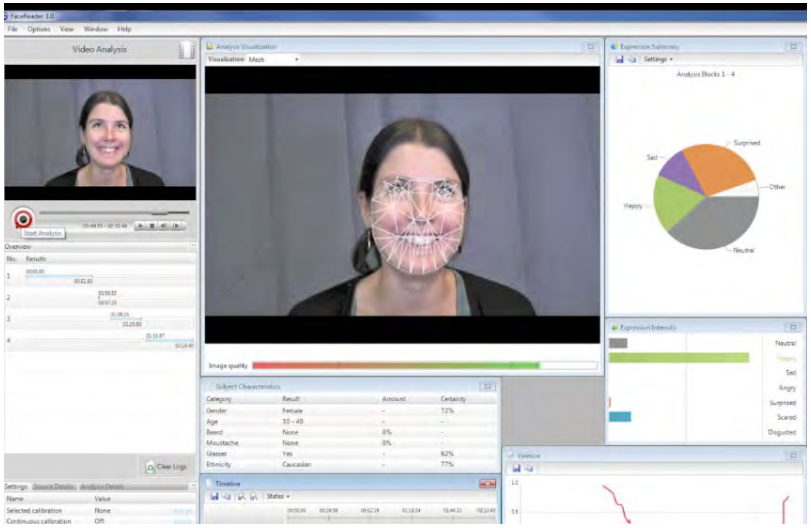


Superposition

Categorical approach

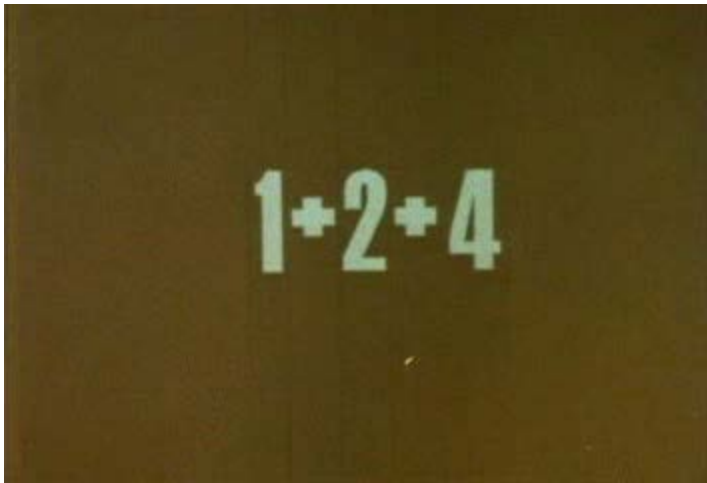
- Controlling an avatar from user's expressions
  - Noldus FaceReader
  - Provides a combination of scores for 6 basic emotions
  - Mapping on the MARC agent from these category blends

  => Not the same expression



Facial expressions
(Ekman 75, Ekman et al. 02)

- **Definition**
  - Rapid signals produced by the contractions of the facial muscles, resulting in temporary changes in facial appearance, shifts in the location and shape of the facial features, and temporary wrinkles
- **FACS**
  - Physically based coding scheme
  - Action Unit(s) ⇔ muscle(s)
- **Synchrony with speech**
  - Phoneme / Word / Phrase
- **Several meanings:**
  emotion, emblems, punctuators

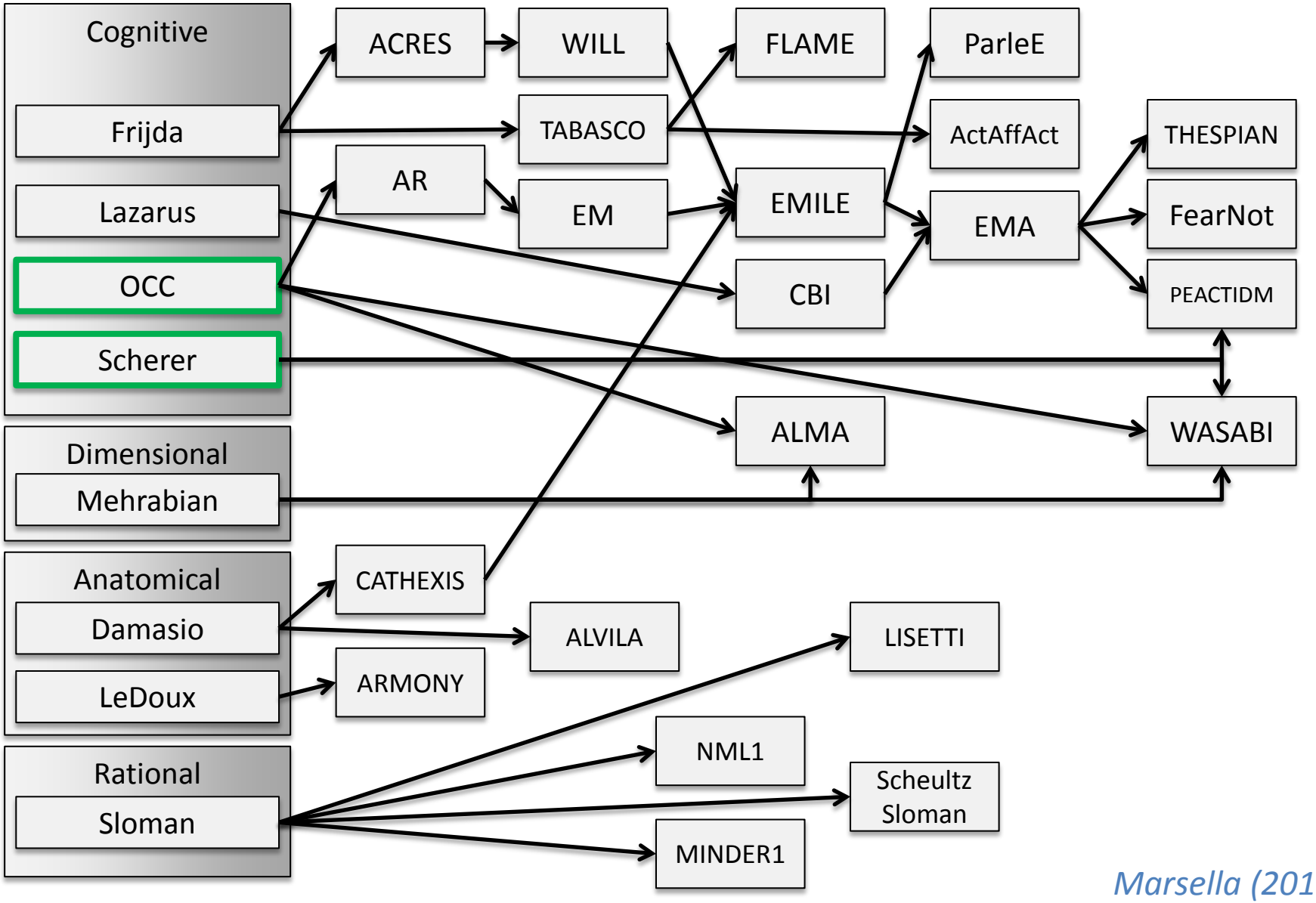


Mapping categories on dimensions

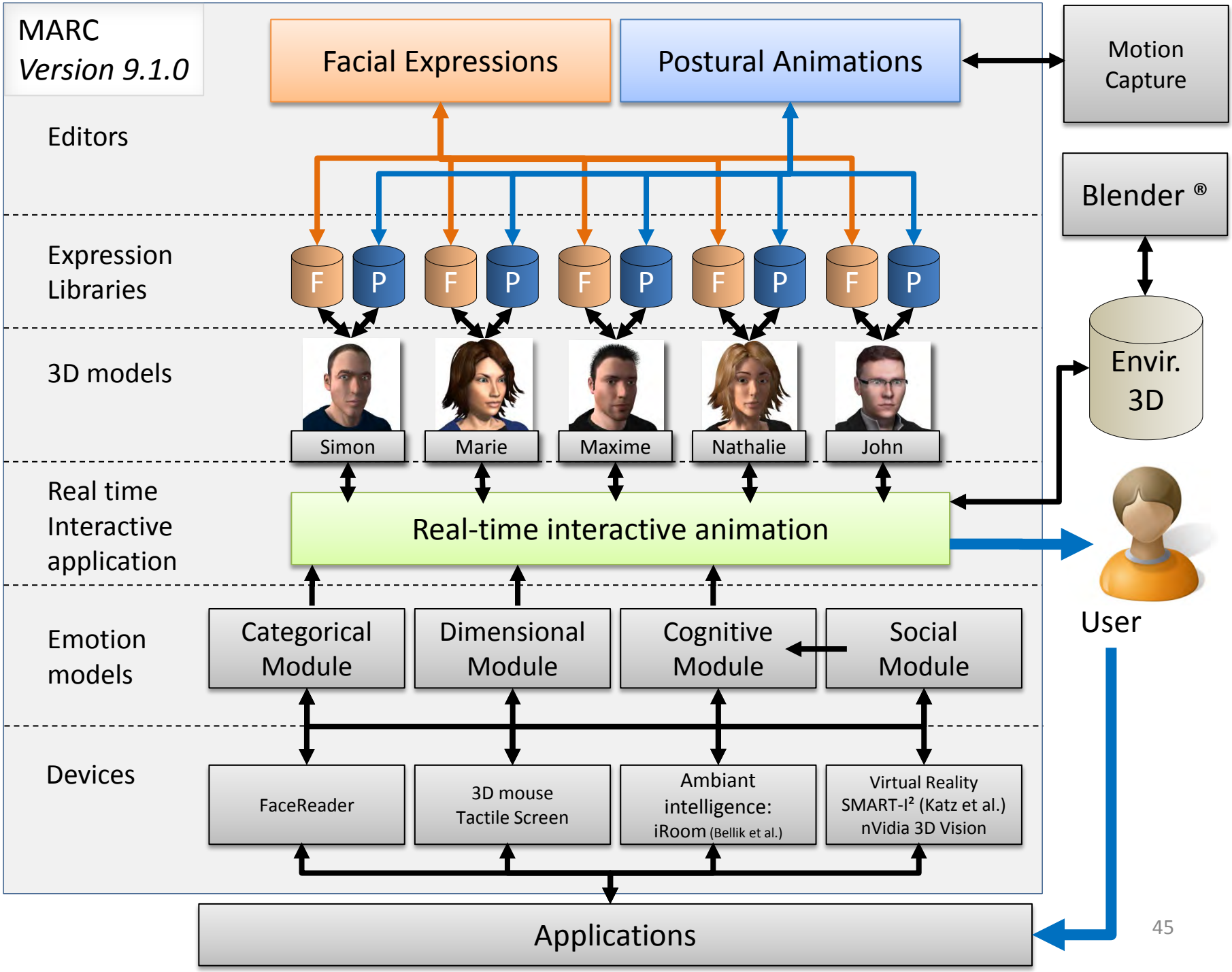
- Lexical studies
- 2 dimensions
  - Valence
  - Arousal

FIGURE 6.5 A two-dimensional representation of emotion terms (vertical dimension: active/passive; horizontal dimension: positive/negative)
(Adapted from Russell, 1980, p. 1167, and Scherer, 1984a, p. 51)

Computational models of emotions

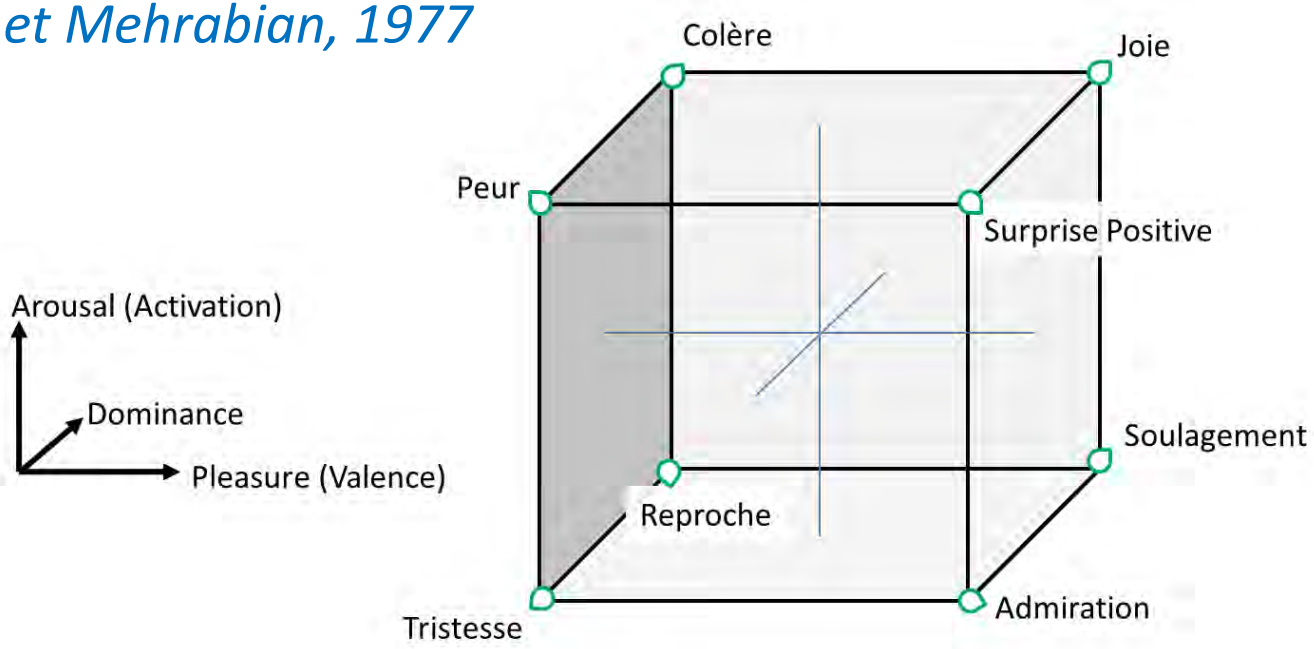


Marsella (2010)


MARC
Version 9.1.0
Facial Expressions
Postural Animations
Motion Capture
Editors
Blender ®
Expression Libraries
F
P
3D models
Envir. 3D
Simon
Marie
Maxime
Nathalie
John
Real time Interactive application
Real-time interactive animation
User
Emotion models
Categorical Module
Dimensional Module
Cognitive Module
Social Module
Devices
FaceReader
3D mouse
Tactile Screen
Ambiant intelligence:
iRoom (Bellik et al.)
Virtual Reality
SMART-I² (Katz et al.)
nVidia 3D Vision
Applications

Implementing a dimensional module in MARC (Courgeon 2011)

- Manual control of the emotional state of the agent in the PAD Space
- *Russell et Mehrabian, 1977*



- Emotional profiles = constrains on the navigation in the continuous space

M. Courgeon, C. Jacquemin, J-C. Martin, (2008) U*ser's Gestural Exploration Of Different Virtual Agents' Expressive Profiles*, in: Proceedings of 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 08), 3, pp 1237-1240, Estoril, Portugal, 12-16 mai 2008

Cognitive theories of emotion

- Most emotions results from a cognitive evaluation of the current situation that the individual is facing (Arnold, 1960, Scherer, 1984)
- Different motor reactions are provoked by this evaluation process (Scherer, 2001)
- The situation might be evaluated according to multiple criteria
  - Laws about action tendencies Frijda, 1988
  - Componential Process Model (CPM) Scherer, 1984,2001,2010

Cognitive Approach to Emotions

Evaluation criteria in the CPM model

Relevance

Is this event new or relevant for me?
Does it affect me directly (or my social group)?

Relation to goals

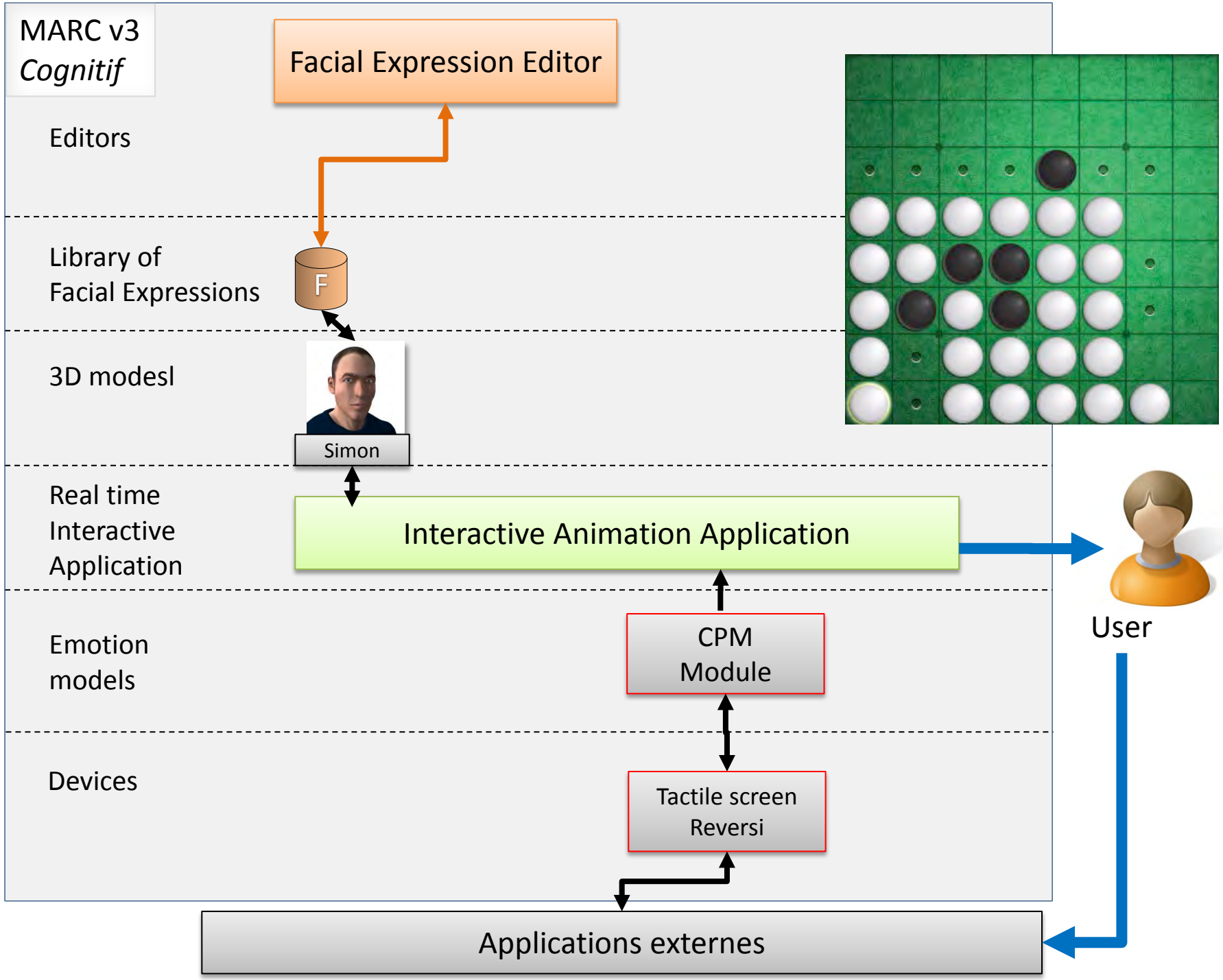
What are the consequences of this evens?
Are they going to affect my well-being or my goals in the short and long term?

Coping potential

Am I able to face the consequences of an event by adapting myself or modifying these consequences?

Norms

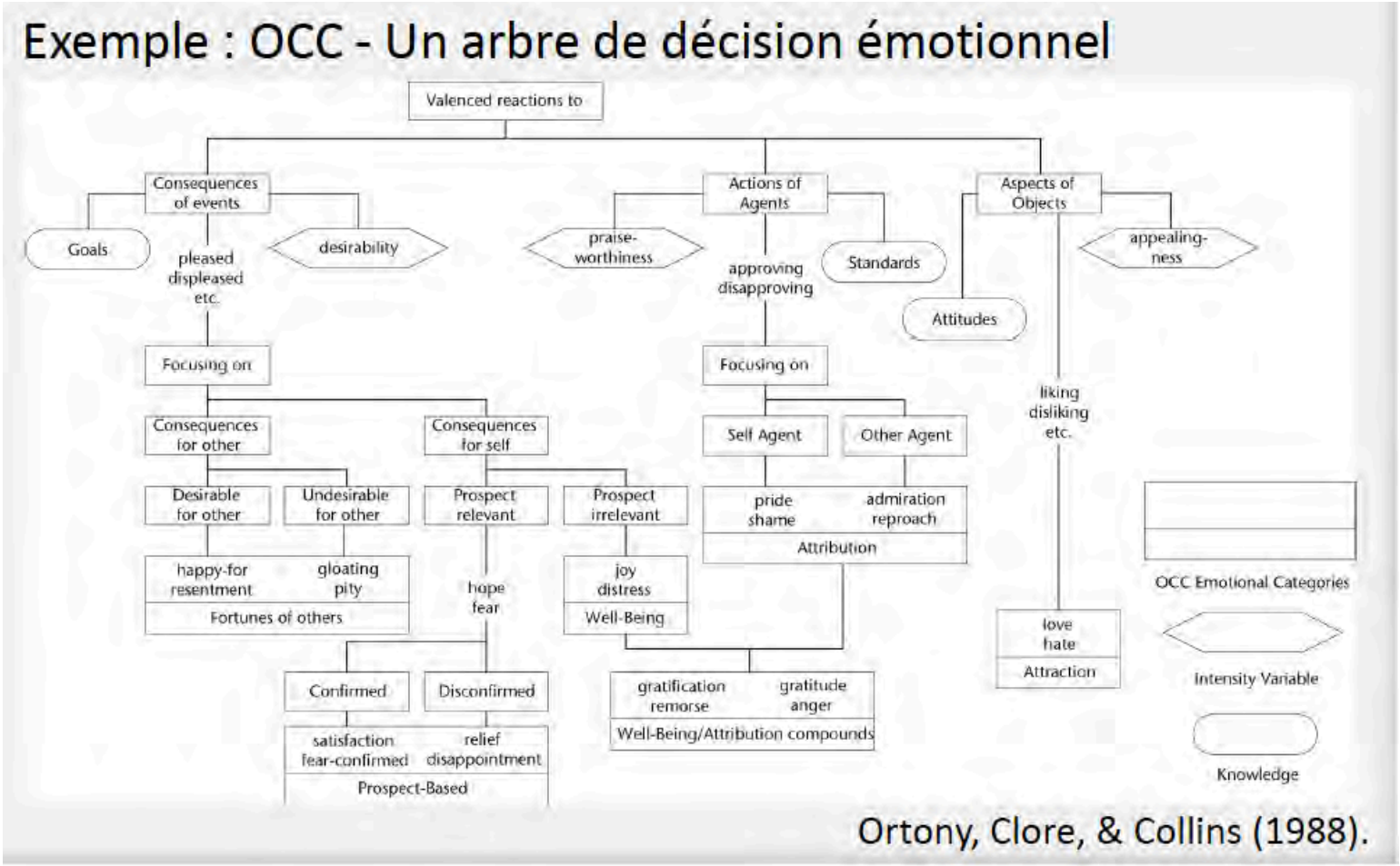
How does this event follow my personal convictions and norms ? And for usual social norms?


MARC v3
Cognitif
Facial Expression Editor
Editors
Library of
Facial Expressions
F
3D modesl
Simon
Real time
Interactive
Application
Interactive Animation Application
Emotion
models
CPM
Module
Devices
Tactile screen
Reversi
User
Applications externes

A model that is different from the CPM model

THE OCC MODEL

OCC


Exemple : OCC - Un arbre de décision émotionnel
Valenced reactions to
Consequences of events
Actions of Agents
Aspects of Objects
Goals
pleased displeased etc.
desirability
praise-worthiness
approving disapproving
Standards
Attitudes
appealing-ness
Focusing on
Focusing on
liking disliking etc.
Consequences for other
Consequences for self
Self Agent
Other Agent
Desirable for other
Undesirable for other
Prospect relevant
Prospect irrelevant
pride shame
admiration reproach
Attribution
happy-for resentment
gloating pity
Fortunes of others
hope fear
joy distress
Well-Being
love hate
Attraction
OCC Emotional Categories
Intensity Variable
Knowledge
Confirmed
Disconfirmed
gratification remorse
gratitude anger
Well-Being/Attribution compounds
satisfaction fear-confirmed
relief disappointment
Prospect-Based
Ortony, Clore, & Collins (1988).

Examples of systems using the OCC model

- Computational models
  - ALMA
  - WASABI
- Virtual agents

DEICTICS AND SHARED ATTENTION

Deictic gestures in human-human communication

- Definition (McNeill 05)
  - Deixis comes from Greek “deiknunai” which means « show »”
  - Pointing is a gesture used to drive the attention of another person so that an object becomes a point of shared attention (Butterworth 03).
  - Concrete pointings (indicate objects or events in the world) vs. abstract pointings (indicate a metaphoric division of space ; used to create a new reference)
- Shape
  - Prototype in USA : hand with extended index (“G”)
  - May involve 2 fingers,the full hand, gaze, lips or any other part of the body (head, nose, elbow) or an object
- Cultural differences

Why are deictic gestures important in virtual humans?

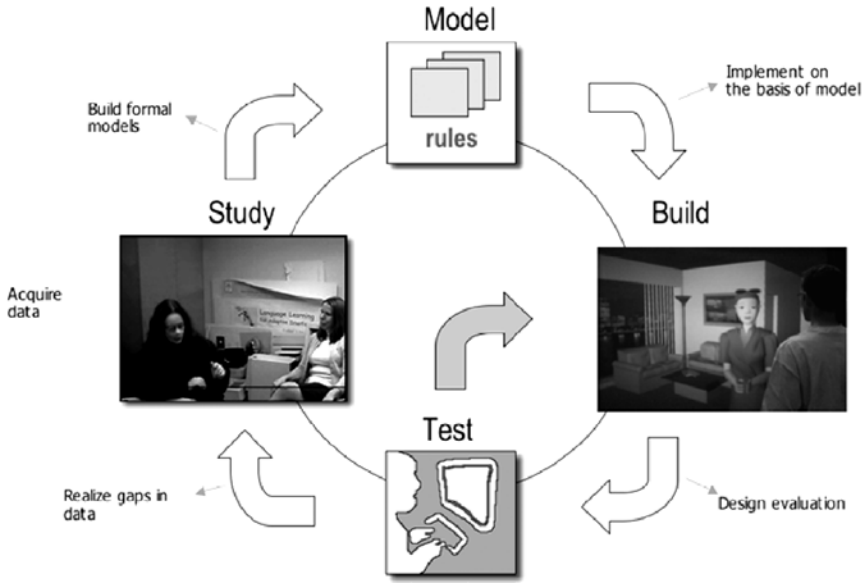
- Showing is another important communicative function for virtual humans
  - pedagogical applications, assistance, game, …
- Deictic should be computed at runtime to ensure a proper consideration of the context
  - User (eg student)
  - Previously pointed objects
  - Grab user's attention
  - Consider ambiguities among multiple objects to show

$\Rightarrow$A (rather) simple illustration of planning multimodal behaviors from a communicative intention

  - When ? Where ? How ?

Informing the design of VH

- Methodology for modeling human conversation and building embodied conversational agents (Cassell 2007)



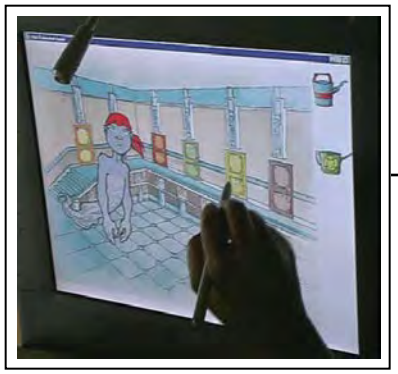
Multimodal Corpus Definition

- Structured and documented set of annotated behavioral data
  - Videos, audio, log files
  - Manual and automatic annotations
  - Documents
    - Protocoles, forms
    - Technical data
  - Measures, representations, models

Digital multimodal Corpora Goals

- Build computational models of situated multimodal behaviors
  - Relations between signs in different modalities
  - Relations between social behaviors and signs
- Enable
  - To model situated knowledge
  - Incremental annotations and measures
  - Integrate manual and automatic annotations
  - Perceptual studies

Annotation process



Video Corpus

Annotation Process

PRAAT

ANVIL
Kipp, 2001

Coding Scheme

Annotations

JAVA JAXP

Metrics

SPSS

Analyses

Questionnaire

Representing gestures

- Phases (Kendon 04, McNeill 92, Kita 98, Duncan, Kipp 04)

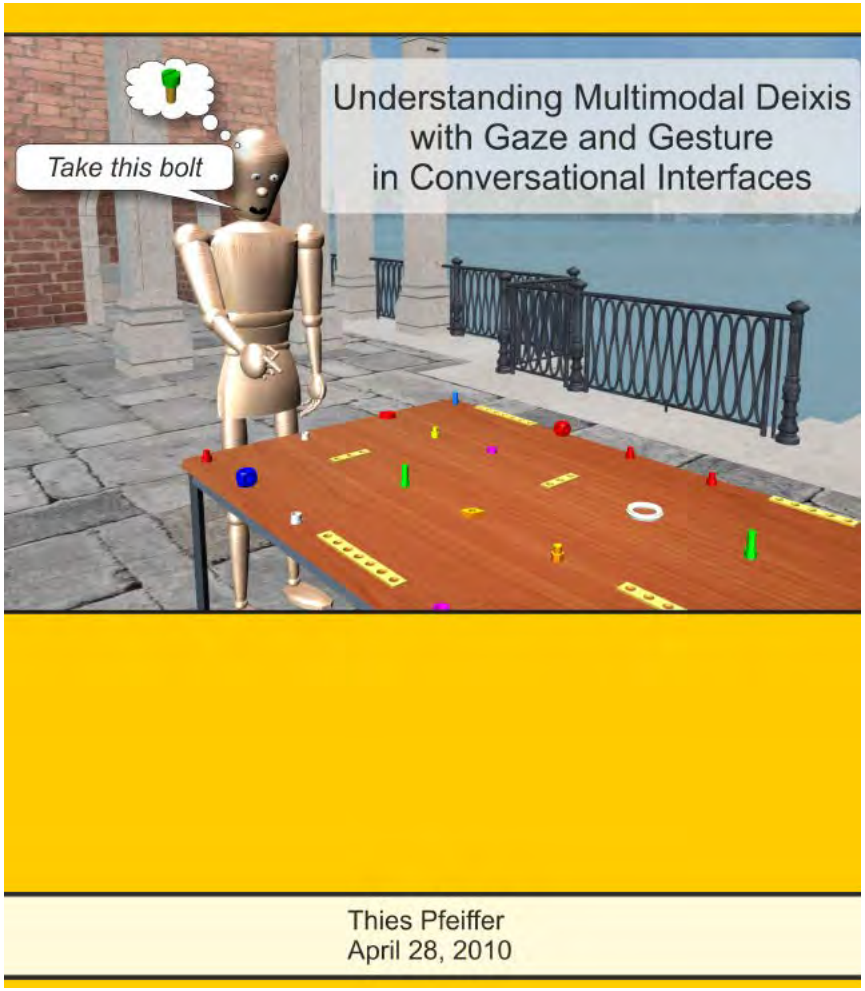
  gesture unit ::= phrase+
  phrase ::= (preparation) expressive-phase (retraction)
  expressive-phase ::= (hold) stroke (recoil) (hold) |
  (hold) beats (recoil) (hold) |
  independent-hold
  retraction ::= retraction | partial-retraction

- Phrases
  - Several classifications
    (Ekman and Friesen '69, Kendon'04)
  - Ambiguities / Overlap
  - Dimensional approach (McNeill 05)

"and he goes up through the pipe this time"

Iconic AND deictic

- Meaning = f(form, time, space, speech) (McNeill 05)
  - Gesture lexicons incl. formational features (Kipp 04, Krehn 06)



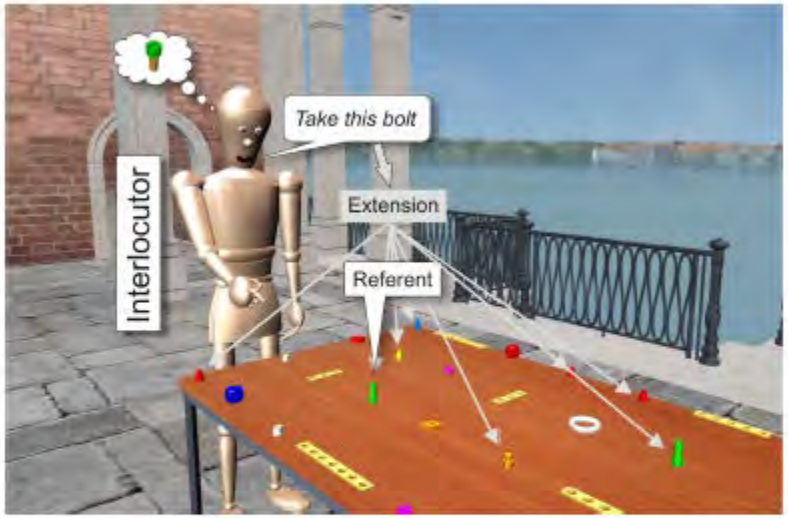


Figure 2.2: *Deictic expressions are used to refer to objects in the world. In the example depicted above, the interlocutor makes a deictic expression as part of a command. The intended referent object is the bolt with the green cap. The potential extension of the deictic expression in the speech alone covers a set of possible referent objects. The manual pointing gesture adds the required information to further restrict the potential extension to the intended object, the referent of the multimodal deictic expression.*

(a) Preparation (b) Stroke (c) Retraction

Figure 2.4: *The main phases of a manual pointing gesture as defined by Kendon (1980).*

(a) Single object intersected (b) Multiple objects intersected (c) No objects intersected

Figure 2.6: *Determining the referent of a pointing gesture is problematic. Besides the ideal case (a), where the referent can, in principle, be identified, at least two other cases exist, (b) and (c), where this is not possible.*

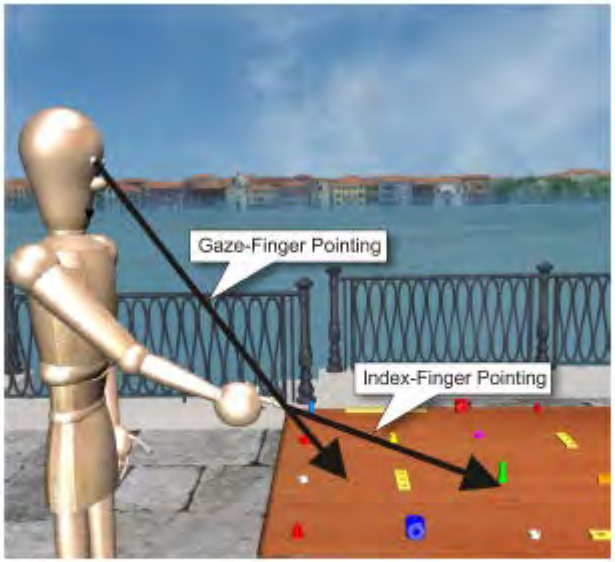


Figure 3.4: *The direction of a manual pointing gesture is not clearly defined. The graphic shows the intuitive interpretation along the direction of the extended finger, coined index-finger pointing. Gaze-finger pointing is an alternative suggestion that takes into account the line of gaze.*

Figure 3.2: *One way to model the extension of pointing is using a pointing cone. The cone is a formalized way to account for the decreasing accuracy of pointing when pointing at distant objects.*

Figure 6.9: *The pointing cone model for pointing is parameterized by the origin (here the tip of the pointing finger), the direction (here GFP/dom) and an angle defining the aperture of the cone.*

A better candidate to model the extension of pointing is the pointing cone (see Figure 6.9), one of the shape-based dereferencing models. Pointing cone models try to take the increasing ambiguity of pointing into account. The equation for the pointing cone (Equation 3.9 on page 55) can be refined for GFP/dom in analogy to the vector extrapolation above:

$$\begin{aligned} G_{cone}: 0 &\geq \vec{y} \cdot \vec{v}_{GFP/dom} - |\vec{y}||\vec{v}_{GFP/dom}|\cos\phi \\ \text{with } \vec{y} &= \vec{x} - \vec{o}_{tip} \end{aligned} \tag{6.2}$$

The equation of the pointing cone model requires an additional parameter ϕ, which is half of the aperture angle of the cone. If ϕ is 0, the pointing cone model is identical to the vector extrapolation model. The aperture angle should be small enough to exclude false positives, i.e. objects that are mistakenly identified as referents. It should also be large enough, that as many referents as possible are correctly identified. In the following, an optimal ϕ for the manual pointing study will be approximated based on the recorded data.

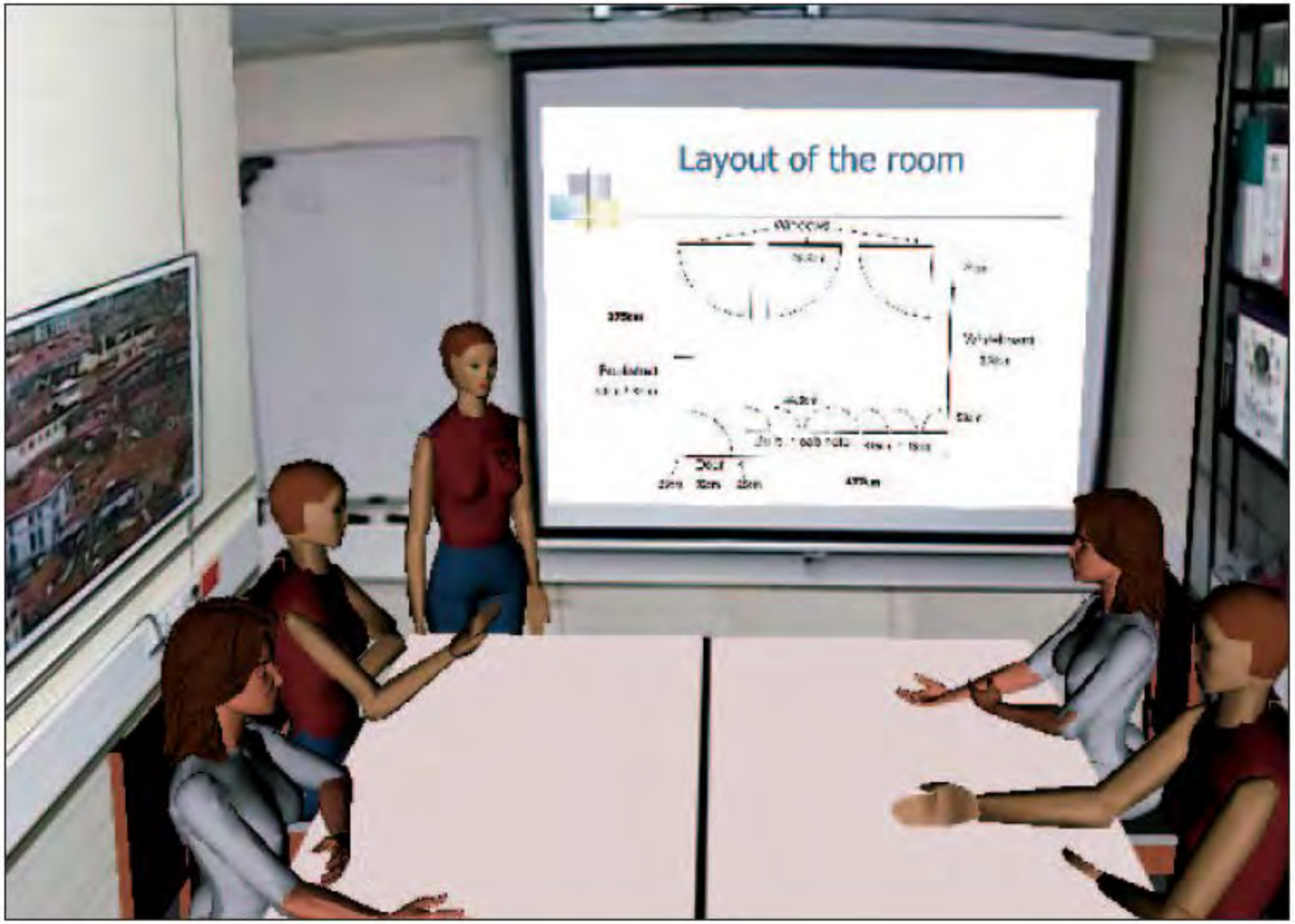


Figure 1. The virtual presenter.

Pointing

A presenter can refer to areas of interest on the sheet by using a gesture with a pointing component. Our pointing model considers several aspects of pointing movement, so that our system can generate the pointing movement given only the intention to point and a pointing target. Like Tsukasa Noma, Liwei Zhao, and Norman Badler's presenter (see the "Presentations by Embodied Agents" sidebar), ours uses its right hand to point to the right and its left hand to point to the left, to keep an open posture. When the preferred hand is occupied, the presenter will gaze at the area of interest instead of directly pointing at it.

Timing. Fitts' law, which predicts the time to move from a certain start point to a target area, is used to model rapid, aimed pointing actions. Fitts' law could thus give a minimum value for the duration of a pointing action's preparation phase. Our virtual presenter uses a 2D derivation[9] of Fitts' law:

$$T = a + b \cdot \log_2\left(\frac{D}{\min(W,H)} + 1\right)$$

where T is the time necessary to perform the pointing action, D is the distance to the object to point to, W is the object's width, and H is its height. a and b both depend on the pointing medium (in our case, the head or arm) and the pointing individual. We empirically determined a and b from a real presenter. We can set their values as static speaker characteristics to create different pointing styles.

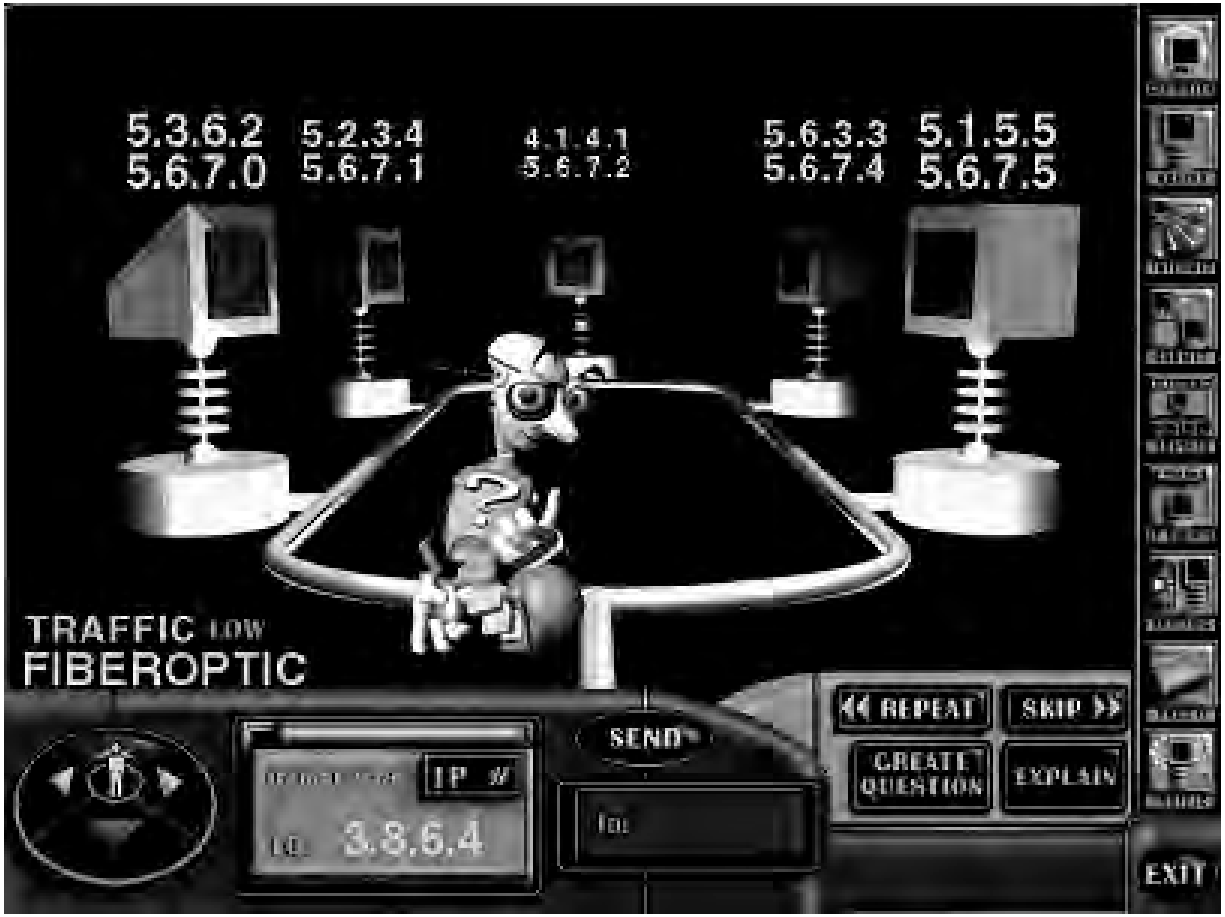
Kranstedt, A., P. Kühnlein, and I. Wachsmuth, "Deixis in Multimodal Human Computer Interaction: An Interdisciplinary Approach," in Gesture-Based Communication in Human-Computer Interaction, A. Camurri and G. Volpe, Editors. 2004, Springer. p. 436-447.

Fig. 1. The pointing game scenario: Comparable settings for empirical studies and VR

Intended empirical studies realized in this VR-setting are aimed at evaluating MAX capacity to interpret and generate situated co-verbal gesture and testing for the naturalness and acceptability of the simulation.

Deictic planning in the Cosmo agent (Lester et al. 98)

- Domain:
  - Internet packet routing



Specifications of the deictic planner

Given:

- For each communicative act created by the explanation planner, the deictic planner is given:
  - *Communicative Act Category*: C
    Examples: `State-Correct`, `Give-Advice`
  - *Topic*: T
    Example: `Address-Resolution`
  - *Gestural Referent*: R_g
    Example: `Computer #15`
  - *Spoken Referent*: R_s
    Example: `SubNetwork #2`
- *World Model*: W (ontology, spatial knowledge, and physical characteristics of objects in environment)
  Example: Ontology of computers and networks, knowledge of the relative locations and proximity of computers to one another, and relative sizes of all environmental and interface objects.
- *Focus histories*: H
  - *Gestural focus history*: H_g
    Example: (`Computer #7`, `Traffic-Information-Label`)
  - *Spoken focus history*: H_s
    Example: (`Computer #12`, `SubNetwork #4`)
- *Current location of agent*: L_A (x-y coordinates)

BODILY INTERACTION

Motion capture databases

- CMU Graphics Lab Motion Capture Database
  - 2600 animations
  - http://mocap.cs.cmu.edu/
- Other web sites
  - http://gfx-motion-capture.blogspot.de/

The .BVH format

- Biovision Hierarchy character animation file format
  - Biovision mocap company
  - Skeleton defined at the beginning
  - Then the animation:
    - 1 frame per line
    - rotation and translation for each bodypart

PP1_tennis1.bvh - Bloc-notes

Fichier Edition Format Affichage ?

```
HIERARCHY
ROOT Hips
{
	OFFSET	0.000000	0.000000	0.000000
	CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation Yrotation
	JOINT LeftHip
	{
		OFFSET	5.160840	-5.226280	-0.522485
		CHANNELS 3 Zrotation Xrotation Yrotation
		JOINT LeftKnee
		{
			OFFSET	0.263405	-12.927100	0.006011
			CHANNELS 3 Zrotation Xrotation Yrotation
			JOINT LeftAnkle
			{
				OFFSET	-0.084503	-14.983900	-1.053590
				CHANNELS 3 Zrotation Xrotation Yrotation
				End Site
				{
					OFFSET	0.225006	-3.726010	-2.164400
				}
			}
		}
	}
	JOINT RightHip
	{
		OFFSET	-5.160840	-5.226280	-0.522485
		CHANNELS 3 Zrotation Xrotation Yrotation
		JOINT RightKnee
		{
			OFFSET	-0.263405	-12.927100	0.006011
			CHANNELS 3 Zrotation Xrotation Yrotation
```

PP1_tennis1.bvh - Bloc-notes

Fichier Edition Format Affichage ?

```
				OFFSET	0.000000	17.257099	0.080036
				CHANNELS 3 Zrotation Xrotation Yrotation
				JOINT Head
				{
					OFFSET	0.000000	1.751380	0.160065
					CHANNELS 3 Zrotation Xrotation Yrotation
					End Site
					{
						OFFSET	0.000000	10.101900	0.080032
					}
				}
			}
		}
}
MOTION
Frames:	16197
Frame Time: 0.008333
-12.055240	36.557915	48.161777	-176.837585	-178.016129	-73.404274
-12.055895	36.557915	48.161953	-176.837585	-178.016129	-73.404274
-12.056196	36.557915	48.162033	-176.837585	-178.016129	-73.404274
-12.056079	36.557915	48.161831	-176.837585	-178.016129	-73.404274
-12.056105	36.557915	48.161839	-176.837585	-178.016129	-73.404274
-12.055349	36.550327	48.164898	-176.837585	-178.016129	-73.404274
-12.055349	36.550327	48.164898	-176.837585	-178.016129	-73.404274
-12.055349	36.550327	48.164898	-176.837585	-178.016129	-73.404274
-12.055349	36.550327	48.164898	-176.837585	-178.016129	-73.404274
-12.055349	36.550327	48.164898	-176.837585	-178.016129	-73.404274
-12.055349	36.550327	48.164898	-176.837585	-178.016129	-73.404274
-12.055349	36.550327	48.164898	-176.837585	-178.016129	-73.404274
-12.055349	36.550327	48.164898	-176.837585	-178.016129	-73.404274
-12.055349	36.550327	48.164898	-176.837585	-178.016129	-73.404274
-12.083359	36.557915	48.078888	-176.837585	-178.016129	-73.404274
```

BEAST: a validated database of static postures of 4 basic emotions

- Bodily Expressive Action Stimulus Test (BEAST)
  - 36 actors / 19 coders

de Gelder, B. & Van den Stock, J. (2011). The Bodily Expressive Action Stimulus Test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. Frontiers in Psychology 2:181. doi:10.3389/fpsyg.2011.0018.

http://beatricedegelder.com/documents/BEAST.pdf

AFFECT ME: a motion capture database

- Acted emotions (VICON)
  - angry, fearful, happy, sad
- Non acted affective states in computer game
  - frustration, concentration, triumphant, defeated
  - Gypsy5 system

A. Kleinsmith, R. De Silva, N. Bianchi-Berthouze, "Cross-Cultural Differences in Recognizing Affect from Body Posture", Interacting with Computers, 18 (6), (2006) 1371-1389

What is *Action Tendency?*

- "A state of readiness to achieve or maintain a given kind of relationship with the environment." (Frijda et al. 1987)

- One component of emotions (Scherer et al. 2000)

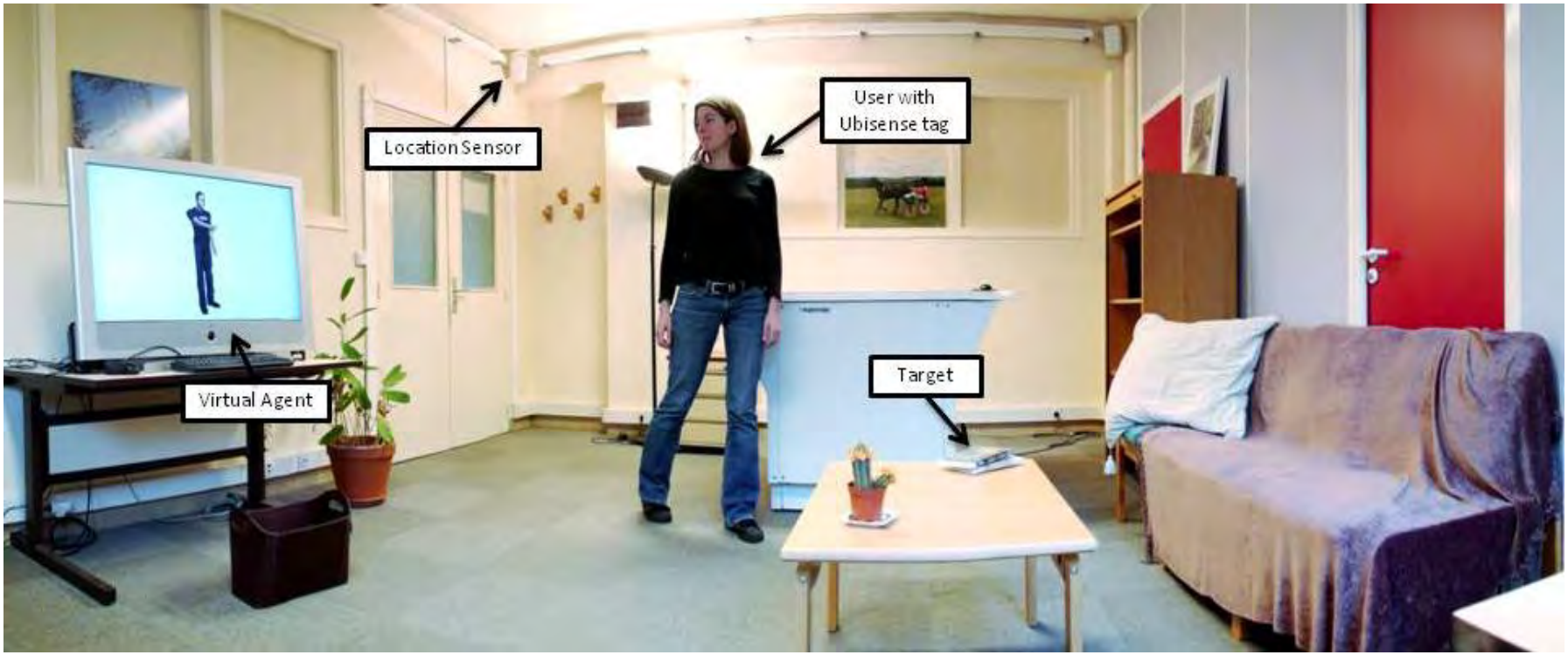
Action tendencies	Postures	
Antagonistic		
Attending		
Disappear from view		

Action tendencies	Description
Antagonistic	I wanted to oppose, to assault; hurt or insult.
Attending	I wanted to observe well, to understand, or I paid attention.
Disappear from view	I wanted to sink into the ground, to disappear from the Earth, not to be noticed by anyone.
Exuberant	I wanted to move, be Exuberant, sing, jump, and undertake things.
In command	I stood above the situation; I felt I was In command; I held the ropes.

Exuberant		
In command		
Neutral		

User in interaction

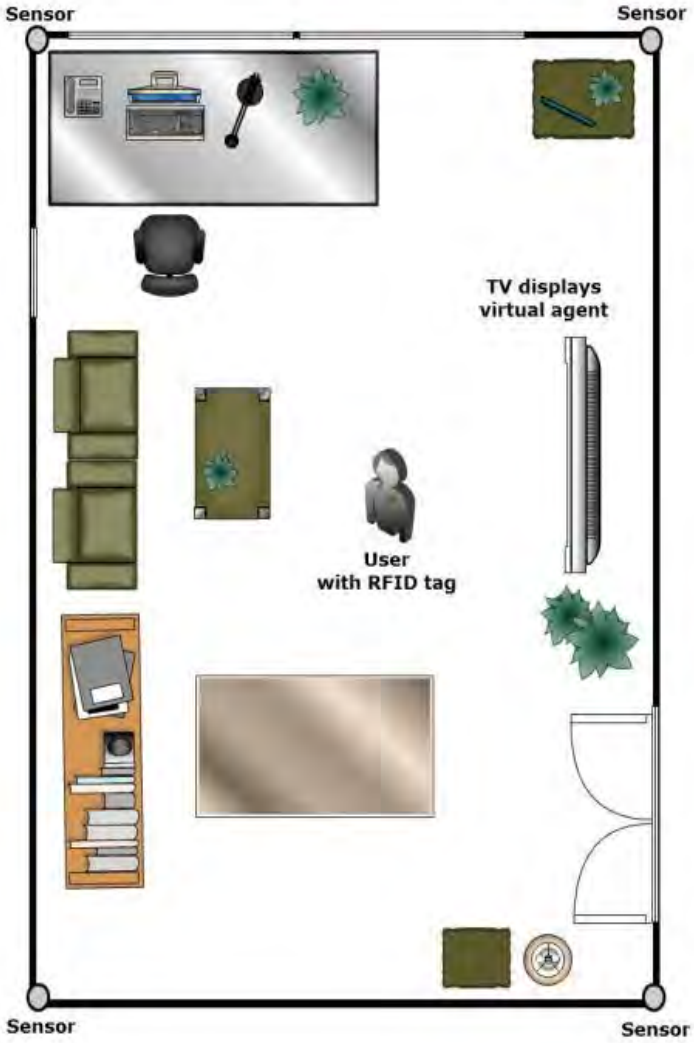
- Searching tasks
  - Continuous data
  - Look at the virtual character
  - Each task is difficult or easy according to the number of distracting objects

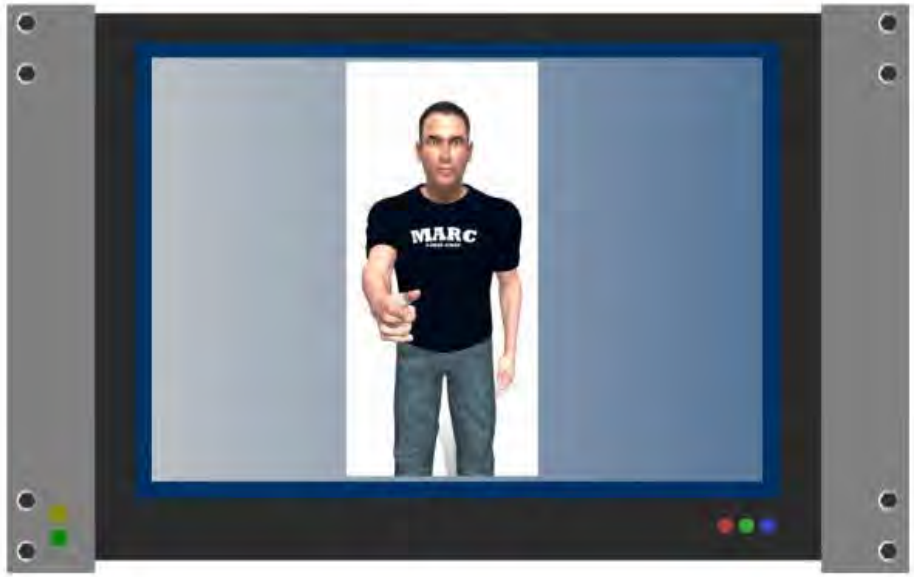


research goals and approach

- Design a location-aware virtual character that helps users find objects in a smart room and is able to adapt its spatial behaviors to users' and objects' locations
- Evaluate the impact of the adaptive character in terms of user perception, task performance and user engagement.

Smart room (IRoom)

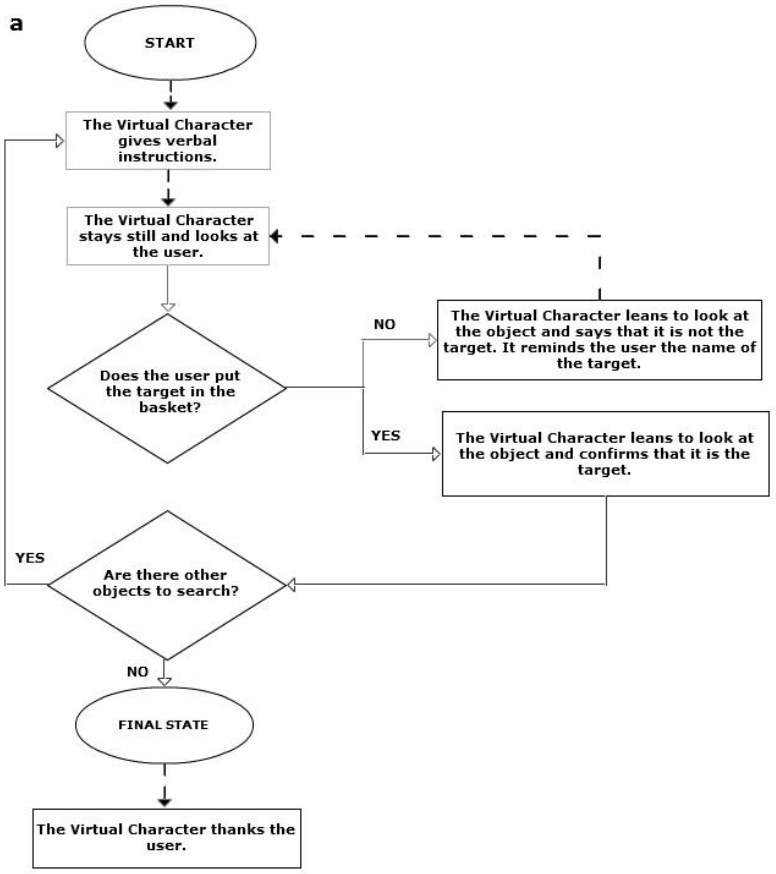




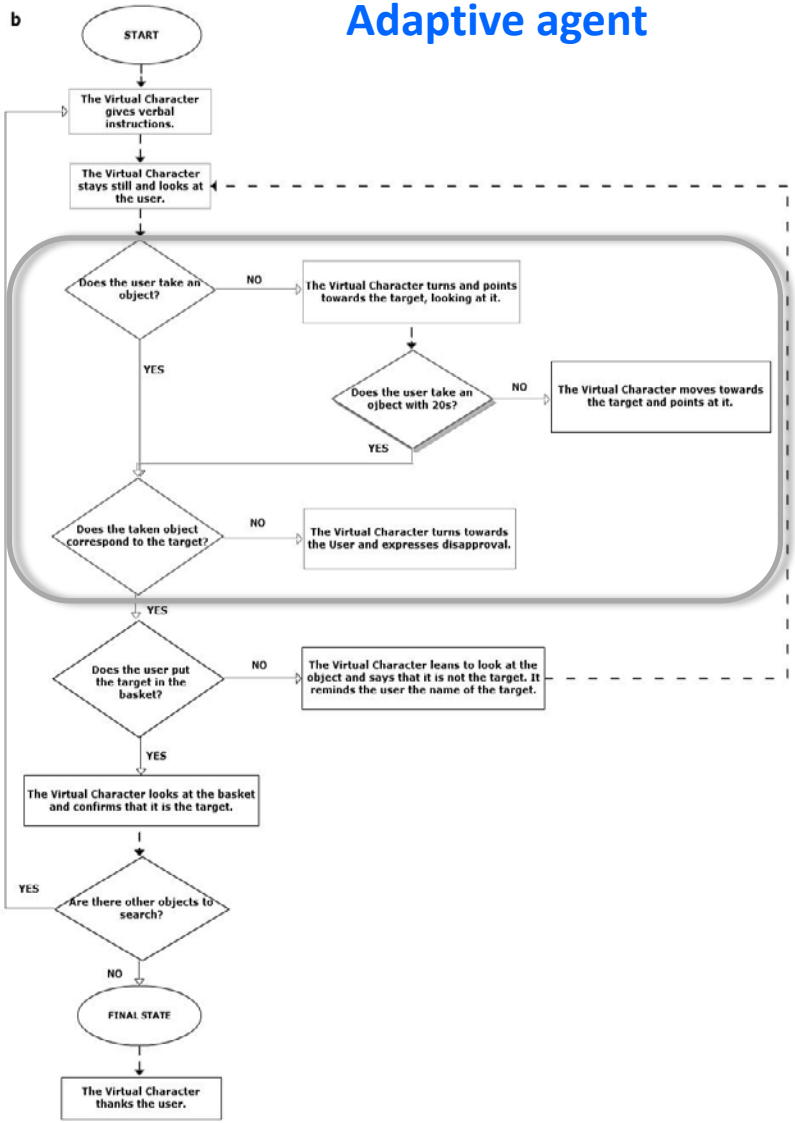
(Bellik & Pruvost)

Interaction design

Non-adaptive agent

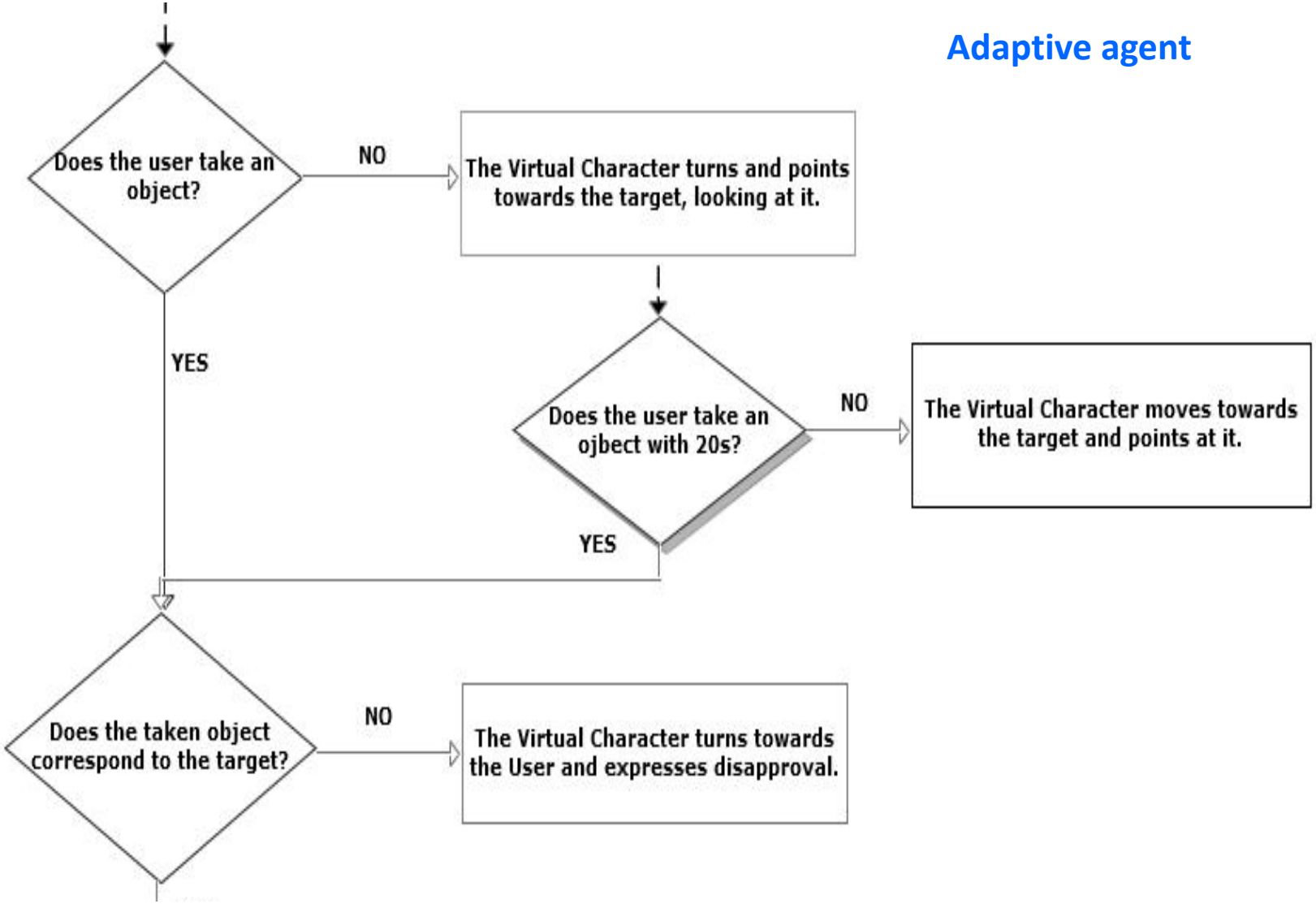


Adaptive agent



Interaction design

Adaptive agent

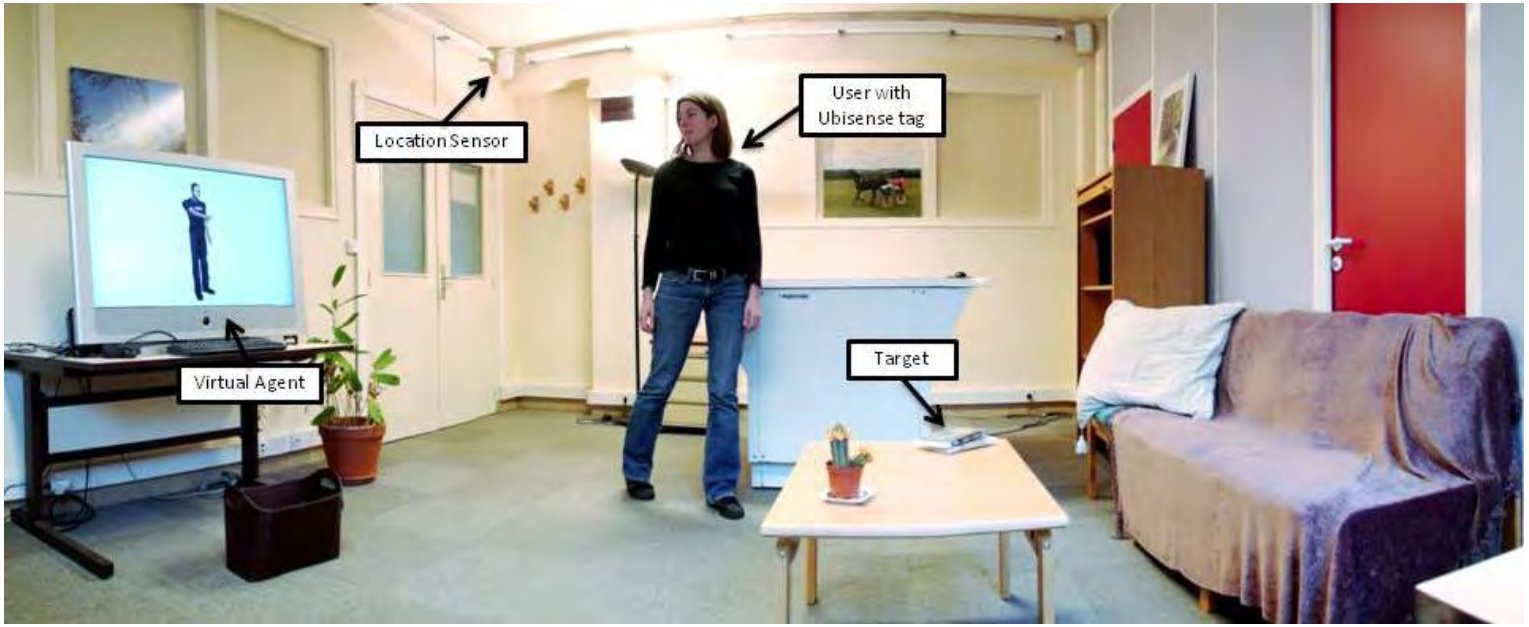


User perception

- Presence
  - Users interacting with the adaptive virtual character felt that the virtual character was watching them and was aware of their presence
- Adaptivity
  - The adaptivity of the virtual character was perceived as significantly higher by users in the adaptive condition than those in the non-adaptive condition.

task performance

measures



results

- No difference in the duration of the whole session across the two conditions
- An effect related to task difficulty
  - Users interacting with the non-adaptive virtual character took more time to complete difficult tasks than easy tasks.

User engagement

measures

- Look around
- Look at an object
- Look at the virtual character for a following instruction
- Look at the virtual character for other reasons

results

- When presented with an adaptive virtual character, the user showed a higher level of visual attention towards the virtual character in terms of action frequency and action length.

MARC in virtual reality

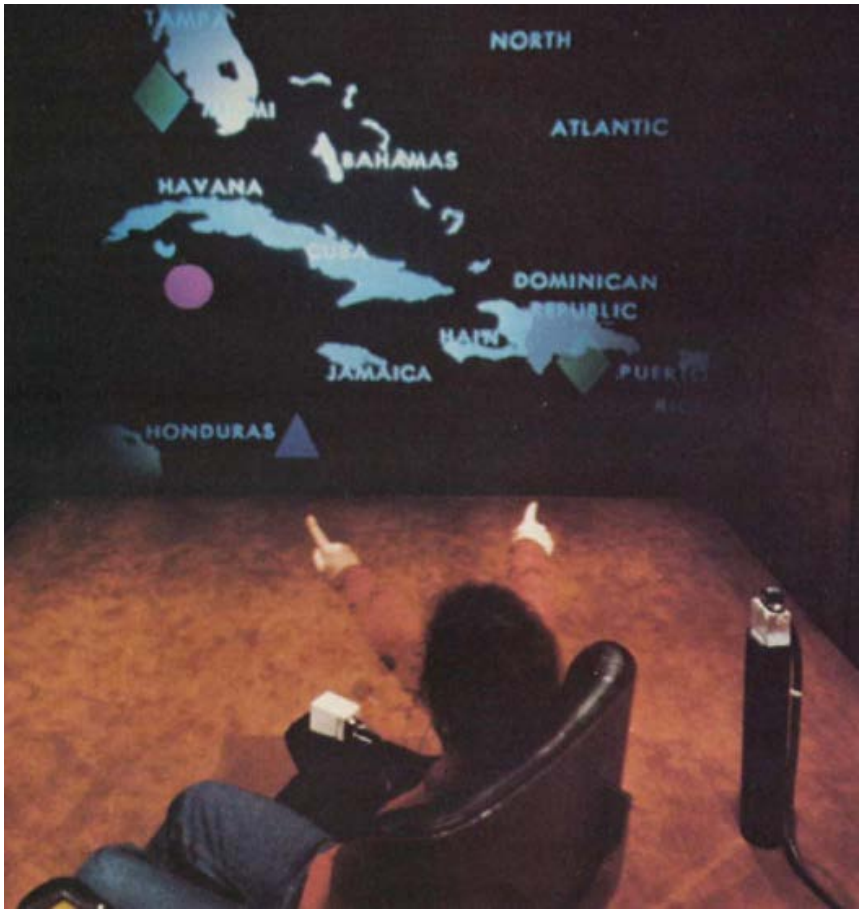
- SMART-I2 (Katz)
- 2 corners
- 3D stereoscopic vision
- Spatial audio rendering

- EVE (Bourdot)
- 3 corners + floor

MULTIMODAL INPUT INTERFACES

« Put that there »
(Bolt 80)

- Media / modalities
  - Speech
  - Gesture (joystick + pad), sensor on wrist
  - Objects attributes:
    - color, size, shape
- Application
  - 2D map / 2D shapes



« Put that there »
(Bolt 80)

- Fusion
  - Speech ("that") notifies the system when it is appropriate to interpret gestures
- Default value
  - for some attributes (size) and none for others (color).
- Limitations
  - Simple 2D shapes, max 5 words per sentence
- Example
  - « move that to the right of the green square »
- Fusion algorithm, management of plurals and ambiguities
  - Not detailled

Why a multimodal input interface?

- Using the modalities that we use everyday in human-human interaction is expected to be
  - Intuitive
  - Efficient (parallel use of several modalities)
  - Robust (redundancy of information)
  - Flexible (equivalence between several modalities)
  - Fun

Main concepts involved in multimodal input interfaces featuring semantic fusion

- Architecture and processing pipeline
- Multiple possible levels of fusion
- Semantic representations
- Uncertainty of input
  - Probabilistic input, mutual disambiguation, modality confidence, congruency/incongruency
- Interactive loops at multiple levels

Architecture and processing pipeline (Abdo El Ali, Dumas et al. 2009)

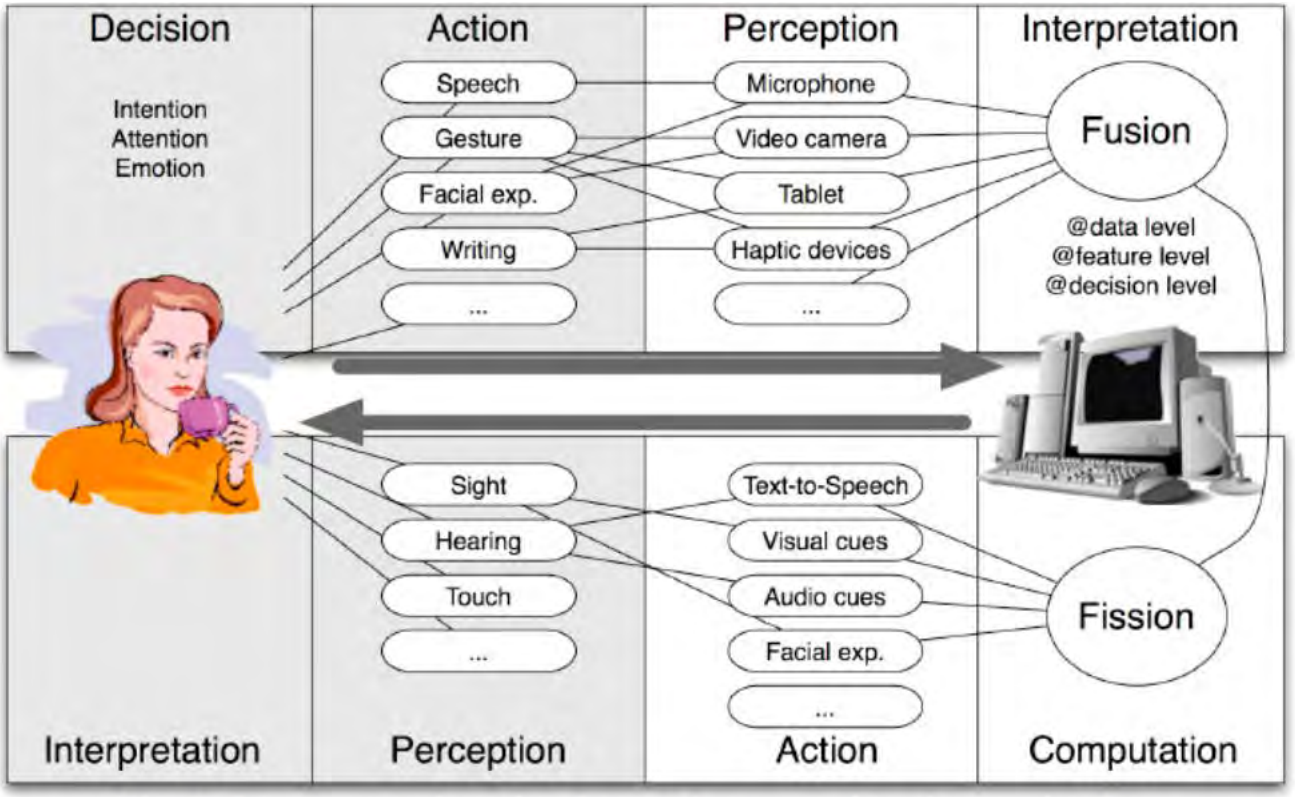
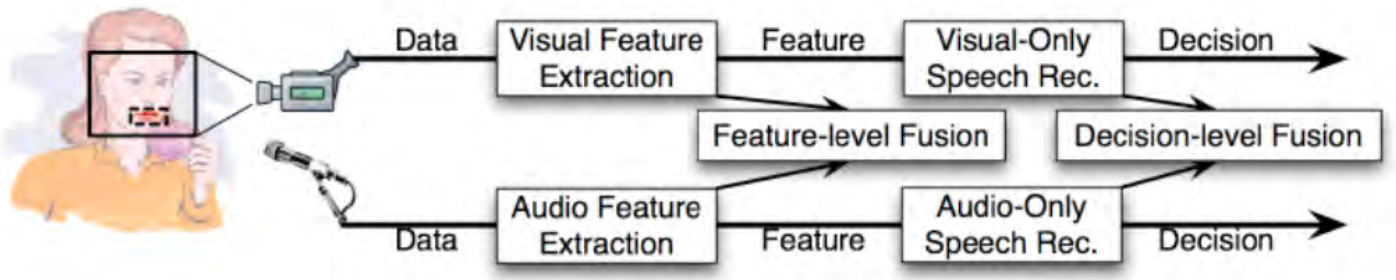


Fig. 1. A representation of multimodal man machine interaction loop.

(Dumas et al., 2009)

- Input FUSION vs. output FISSION

Multiple possible levels of fusion (Abdo El Ali)



Data Level:

e.g., combining 2 webcam video streams, multiple perspectives

Feature level:

e.g., combining speech and lip movements

Decision level:

e.g., combining gestures and speech

Mutual Disambiguation Maven system (Kaiser et al. 03)

Figure 2. VR avatar controlled by tracked user, showing attached regions of interest.

- Media / modalities
  - Input : speech, gesture
    (4 trackers 6DOF hand / wrist / arm / head)
  - Output : 3D head-worn display
- Application
  - Immersive virtual reality environment
- Goals
  - Manage uncertainty in spoken and gestural references

Mutual Disambiguation Maven system (Kaiser et al. 03)

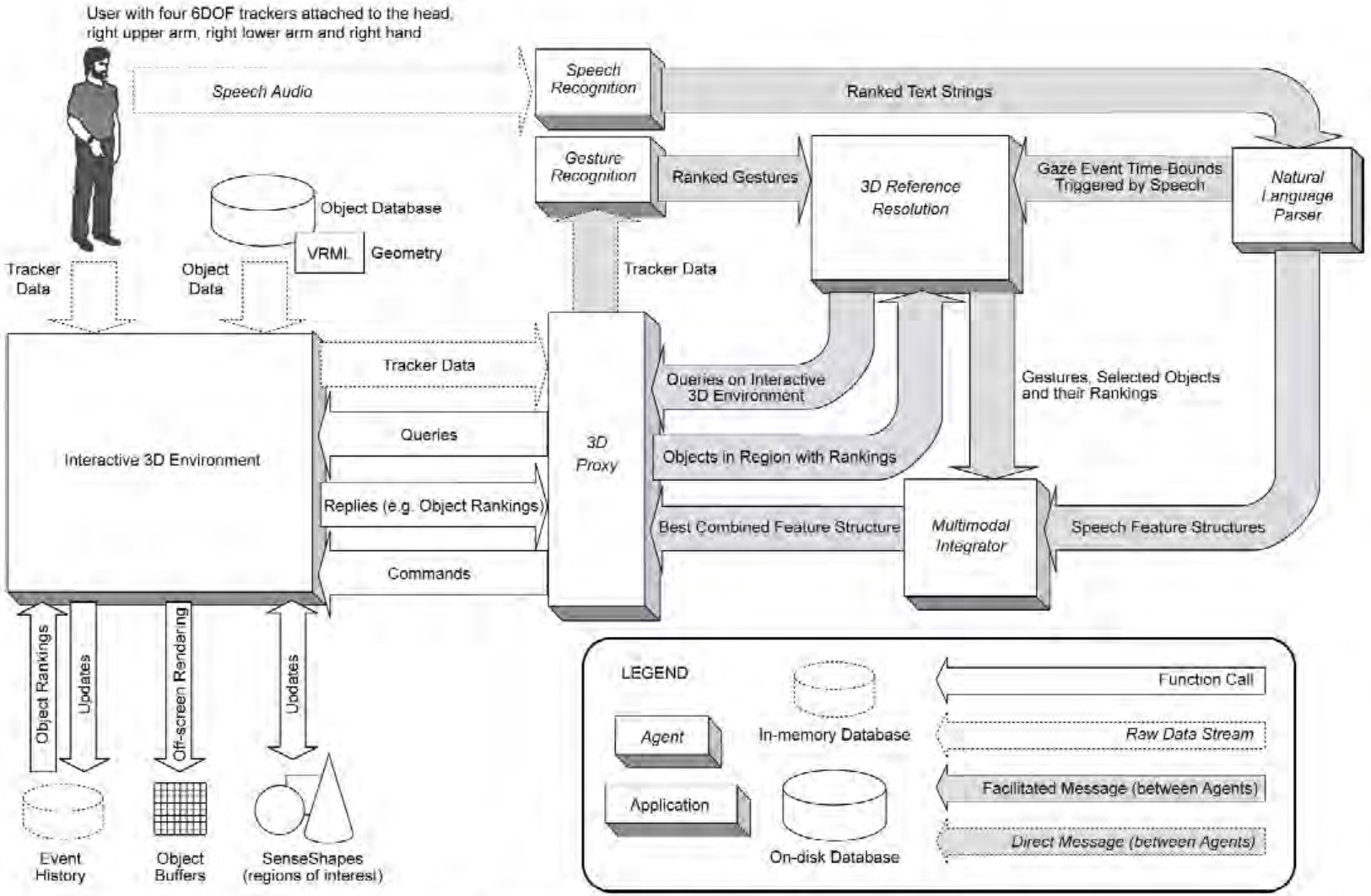


Figure 3. Multimodal interaction architecture.

Mutual Disambiguation Maven system (Kaiser et al. 03)

- Region of interest
  - Volumes controled by the user + intersection with objects
  - Rankings : duration of intersection, stability, visibility, proximity
- History : timestamp events
- Speech
  - Dragon Naturally Speaking, 150 words, context-free grammar
- Gesture
  - Start and end : stability
  - 4 gestures are recognized + look (head) : point, push, twist, rotate

Mutual Disambiguation
Maven system (Kaiser et al. 03)

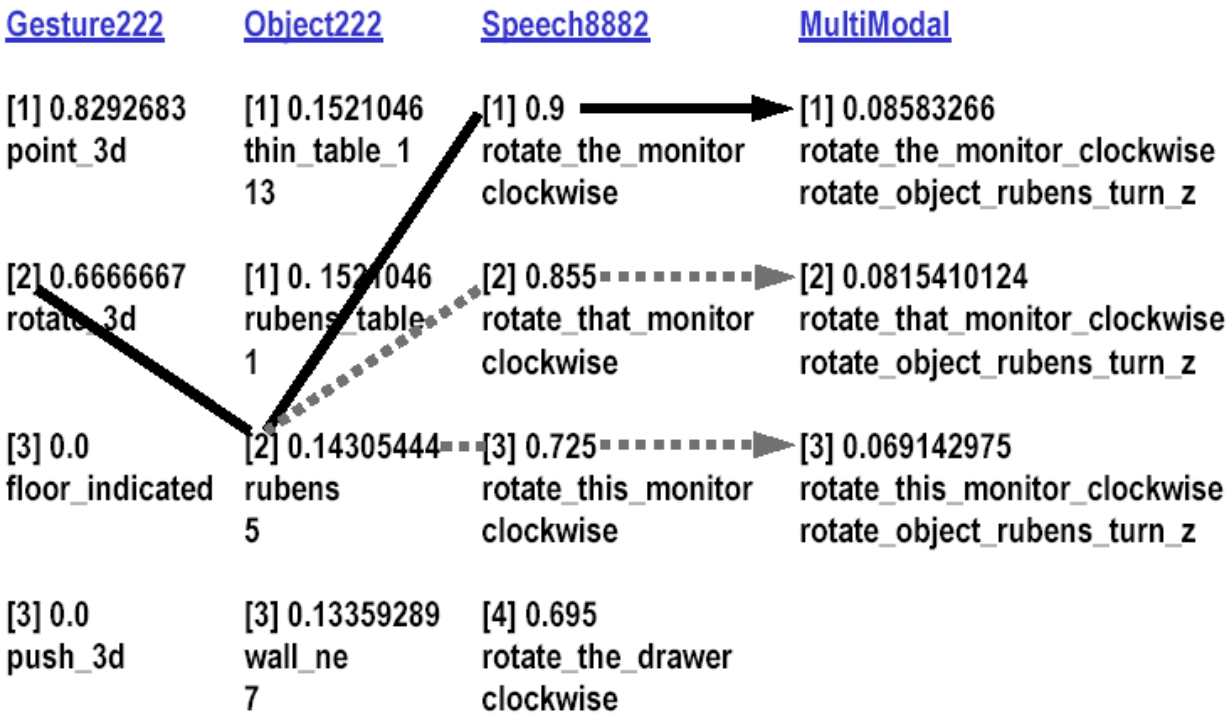


Figure 5. Parallel coordinate plot showing multiple hypotheses for each modality.

Symmetry between input and output modalities
Confidence in each modality
Smartkom (Wahlster 2003)

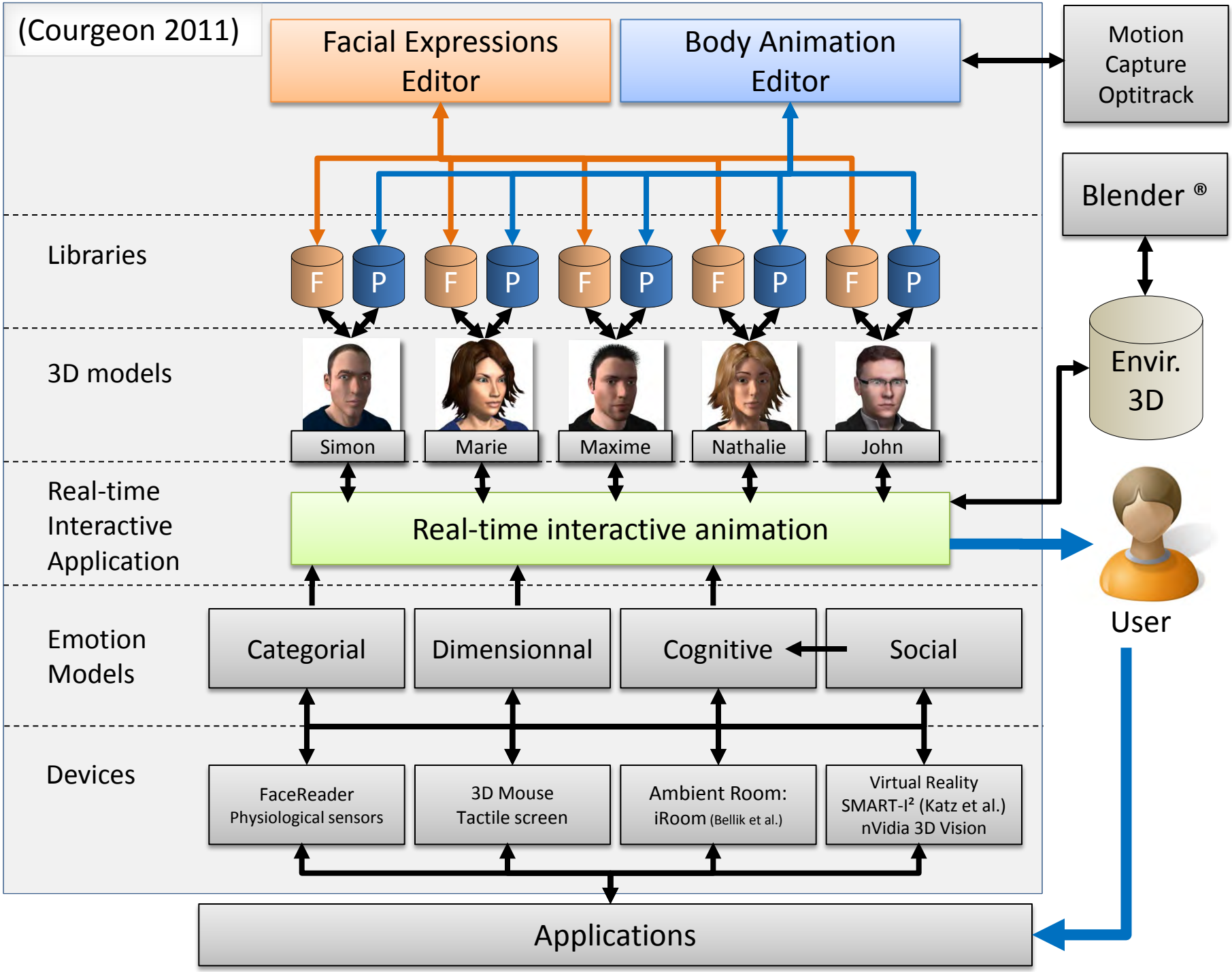
- A collaborative dialogue between the user and the agent elaborates the specification of the delegated task and possible plans of the agent to achieve the user's goal
- Adaptive confidence measures
  - while the user is speaking, the mouth area does not predict emotions reliably, so that the confidence value of the mouth area recognizer must be decreased
- As soon as the modality fusion component finds a referential expression that is not combined with an unambiguous deictic gesture
  it sends a request to the discourse component asking for reference resolution

Fig. 1. Speech and Gesture for Input and Output

THE MARC PLATFORM (COURGEON 2011) – EXAMPLES AND DEMO

MARC
Multimodal Affective Reactive Characters
marc.limsi.fr

- Research goals
  - Affective interactive applications
  - Conduct experimental studies
- Features
  - Several computational models of emotions
  - Designed for real-time interaction
  - Detailled graphical models
  - Facial animation
    - Collaboration Michel-Ange Amorim sur la perception de la dynamique des expressions faciales
  - Body animation


(Courgeon 2011)
Facial Expressions Editor
Body Animation Editor
Motion Capture Optitrack
Blender ®
Libraries
F
P
3D models
Simon
Marie
Maxime
Nathalie
John
Envir. 3D
Real-time Interactive Application
Real-time interactive animation
User
Emotion Models
Categorial
Dimensionnal
Cognitive
Social
Devices
FaceReader
Physiological sensors
3D Mouse
Tactile screen
Ambient Room:
iRoom (Bellik et al.)
Virtual Reality
SMART-I² (Katz et al.)
nVidia 3D Vision
Applications

Component Process Model (Scherer 2010)

- Easily adapted to a computational model
- Suited for interactive application
- Sequential Evaluation Checks

Checks	Definitions (adapted from Scherer, 1997)
Expectedness	Did you expect this situation to occur?
Unpleasantness	Did you find the event itself unpleasant?
Goal hindrance	Did this event hinder you in following your plans or achieving your aims?
External causation	Was another person responsible for the event in the first place?
Coping potential	How did you evaluate your ability to act on or to cope with the event and its consequences?
Immorality	If the event was caused by your own or someone else's behavior, would this behavior itself be judged as improper or immoral by your acquaintances?
Self-consistency	How did this event affect your feelings about yourself, such as your self-esteem or your self-confidence?

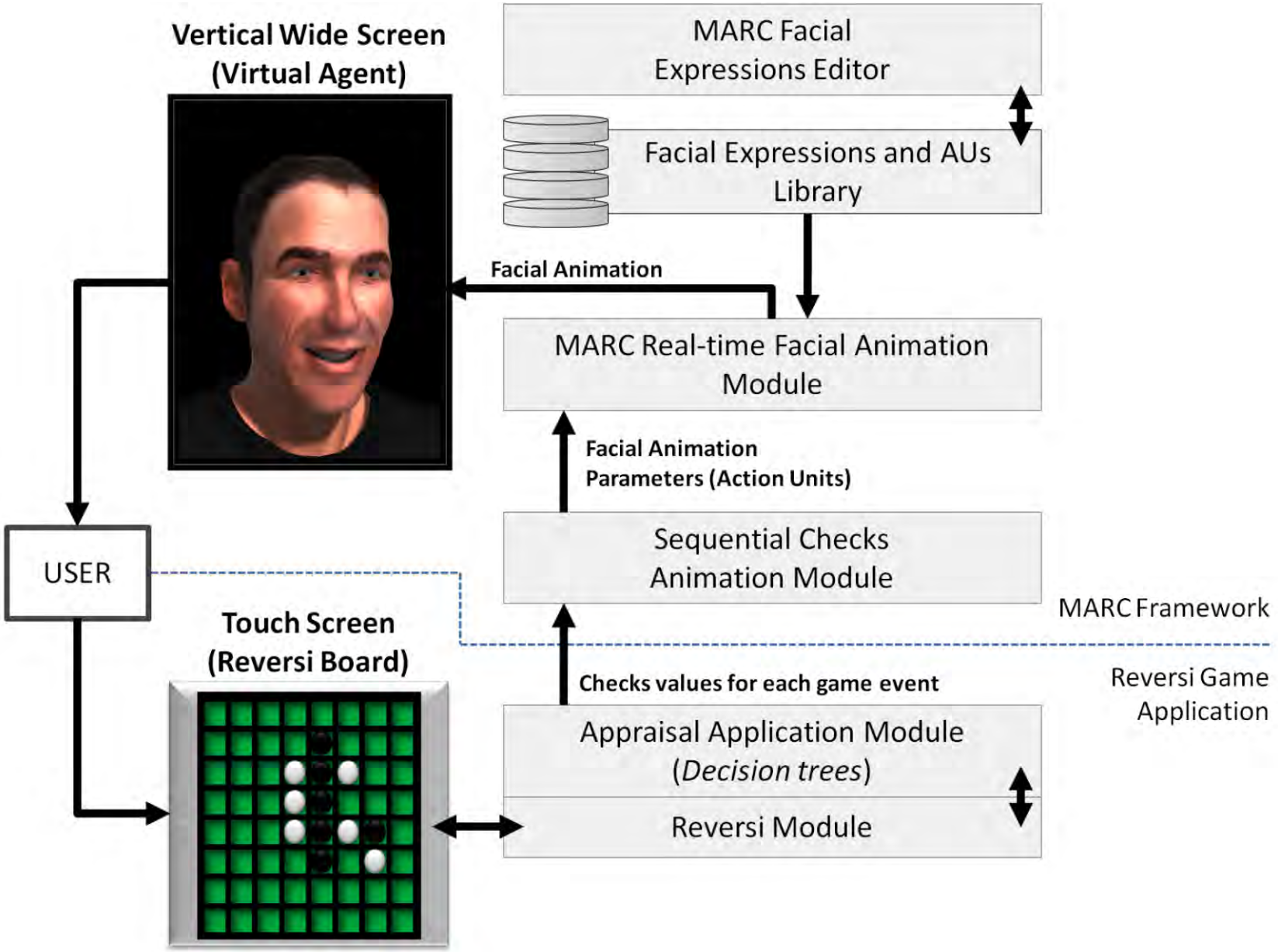


Figure 2. The MARC Framework and its application to the Reversi game.

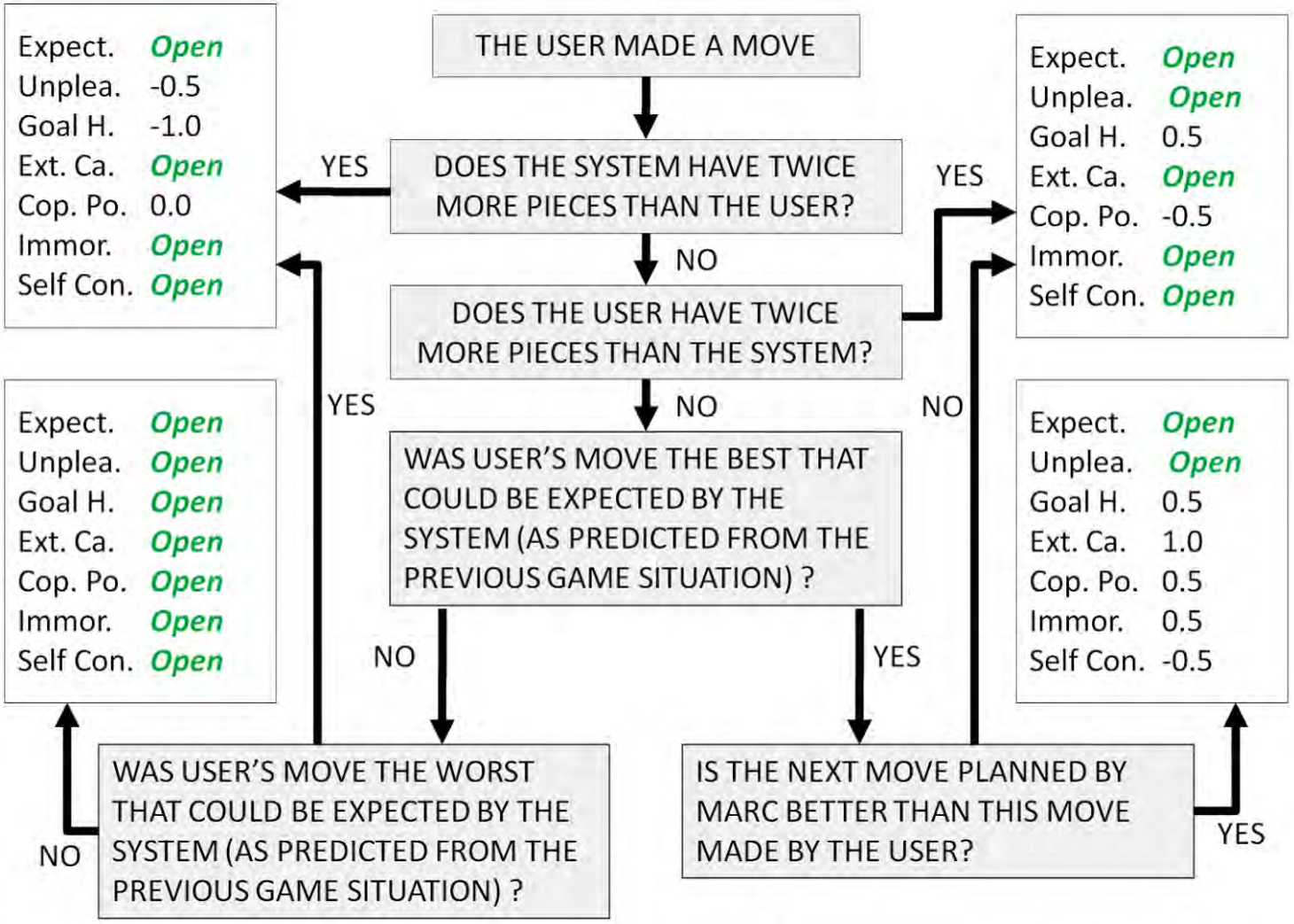
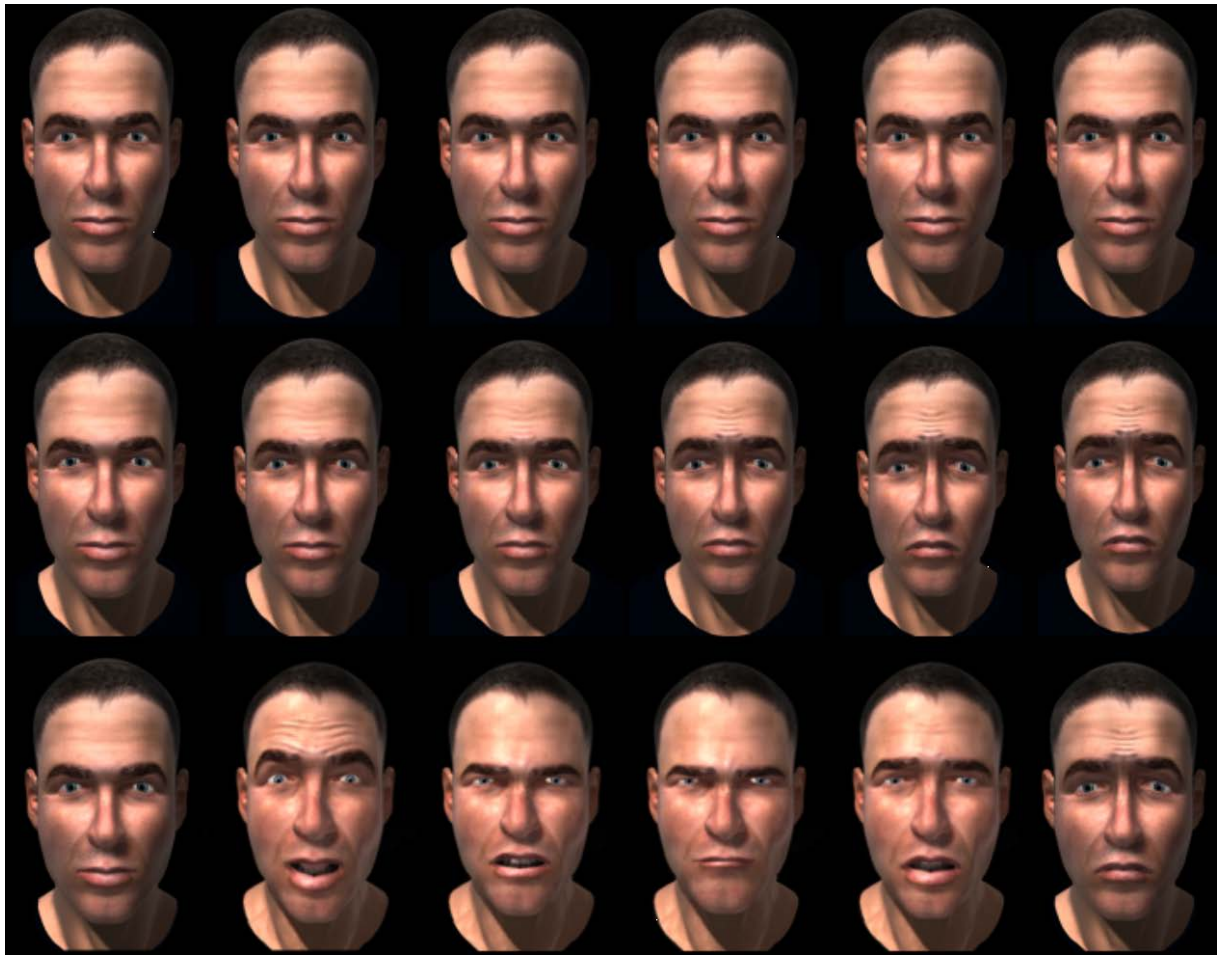


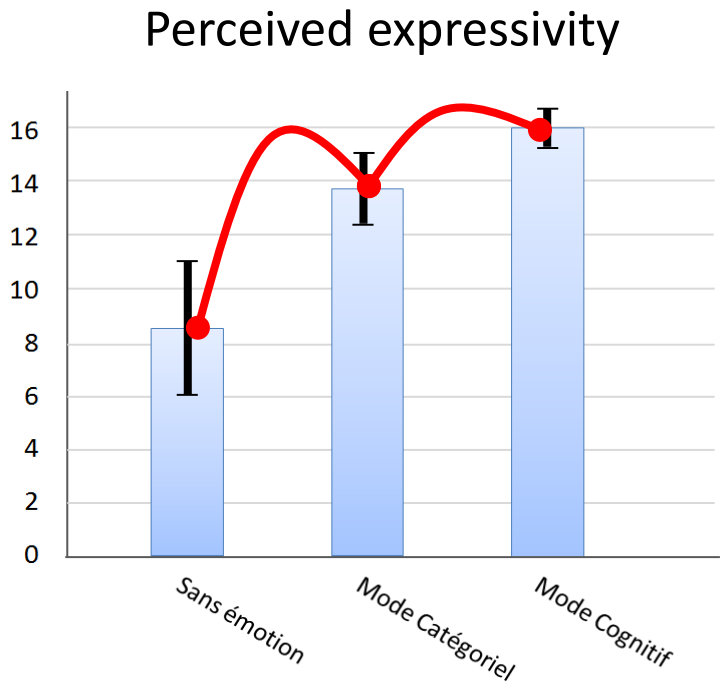
Figure 3. One of the decision trees used to evaluate the current game situation following a move from the user and assigning values to the seven checks. A similar decision tree was used to compute the values of checks following a move from the system. (Expect. = Expectedness, Unplea. = Unpleasantness, Goal H. = Goal hindrance, Ext. Ca. = External causation, Cop. Po. = Coping potential, Immor. = Immorality, Self Con. = Self-consistency). The values that remained open were computed at runtime.

Conditions and facial expressions



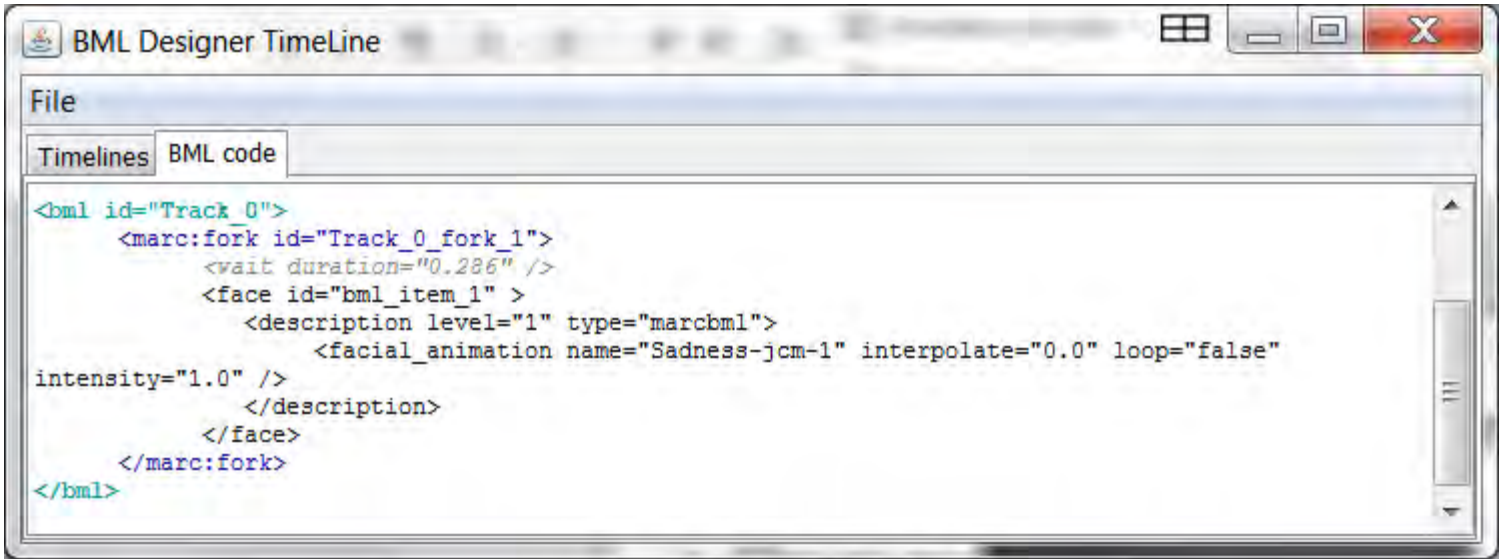
User evaluation

- Goals: study the impact of the facial expressions of the agent on user's
  - Several hypotheses about the impact of the emotion model on Performance / Behaviors / Subjective perception
- Protocol
  - 3 conditions: Categorial / Cognitive / No Emotion
  - 60 users (20 per mode)
- Résultats
  - Impact on gaze behavior
  - Subjective perception of expressivity
    - No Emotion < Categorial < Cognitive
  - The user wins
    - No Emotion < Categorial
    - No Emotion < Cognitive

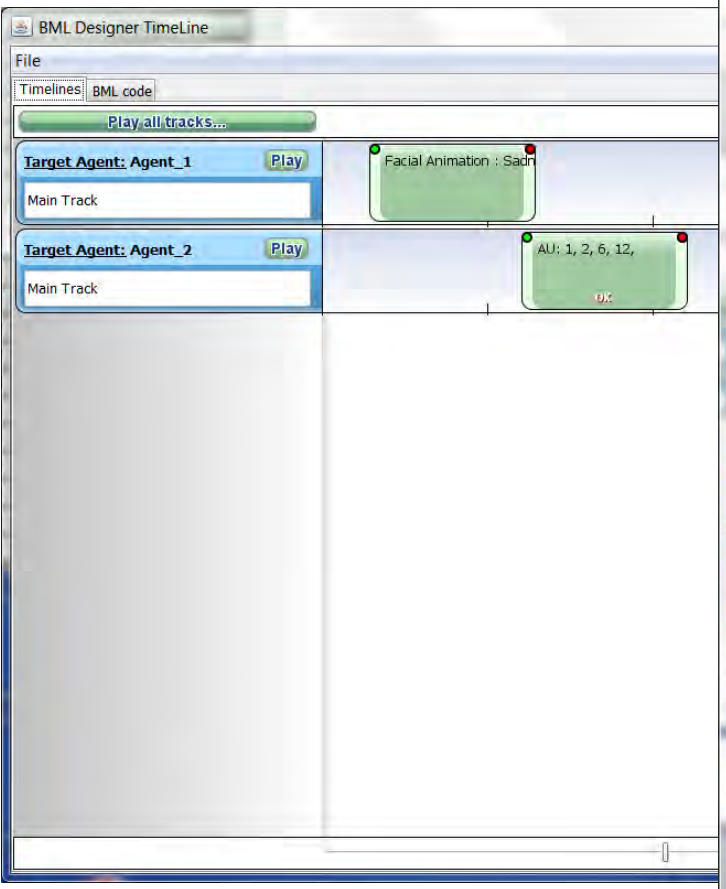


Specify expressions and animation sequence

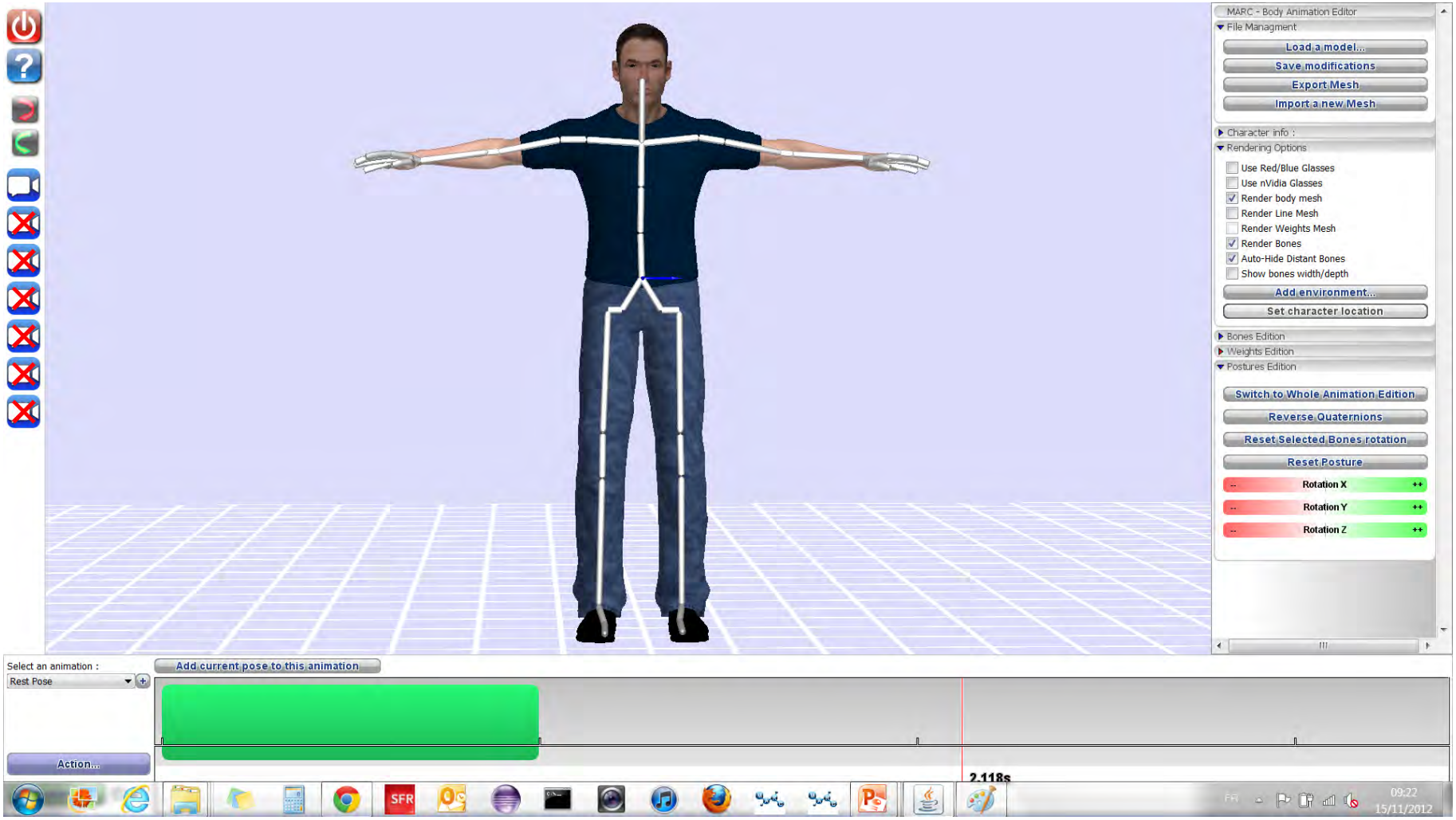
Specify the behavior using a timeline (can be displayed as a BML script)



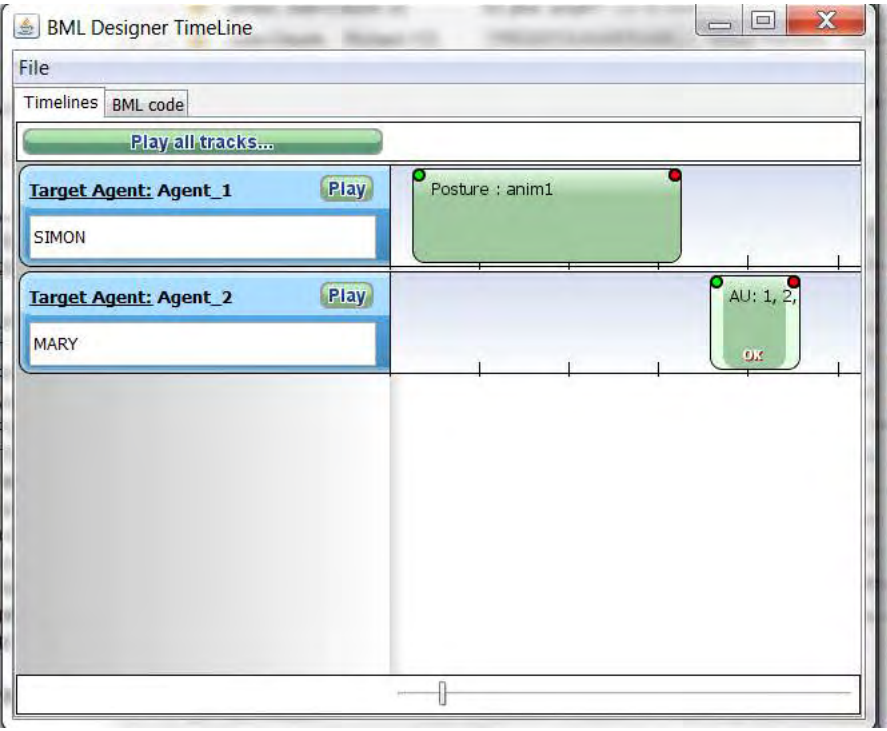
Animate several characters



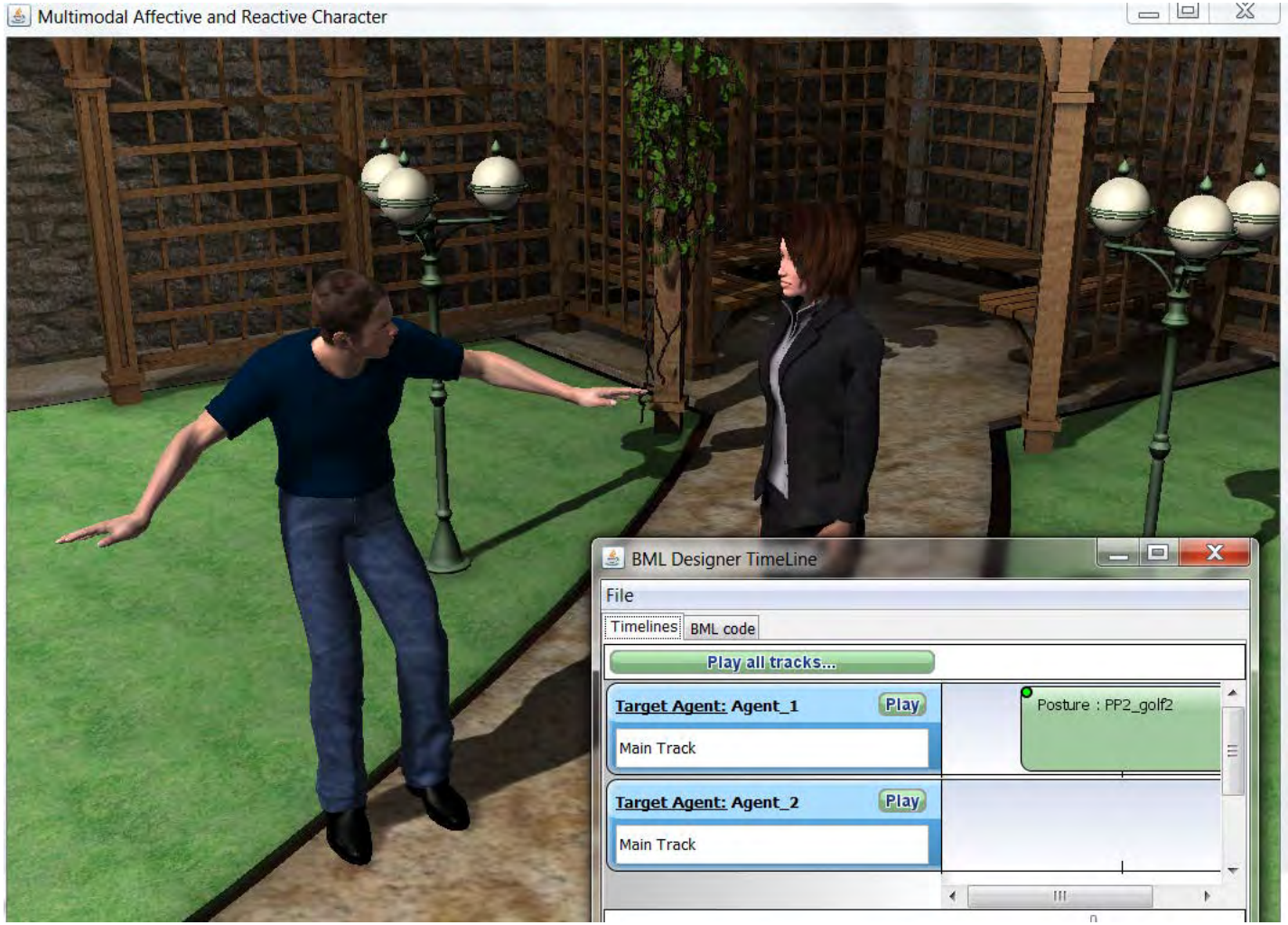
Body animation in MARC
Manual edition of postures



Body animation in MARC
Animating posture at run-time



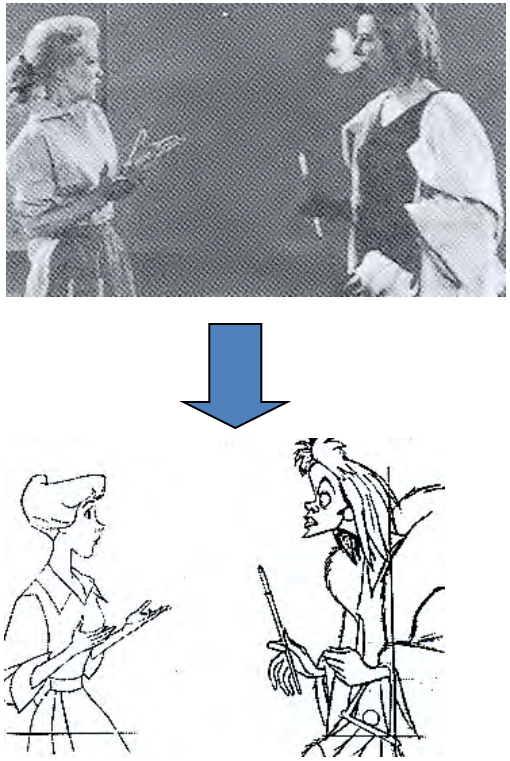
Playing a .bvh file in MARC

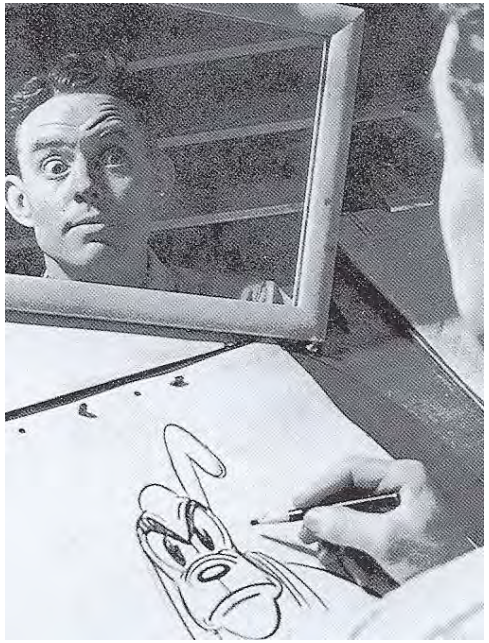


VISUAL AND BEHAVIORAL REALISM

Why inspire from humans?

- Use of live action in drawing





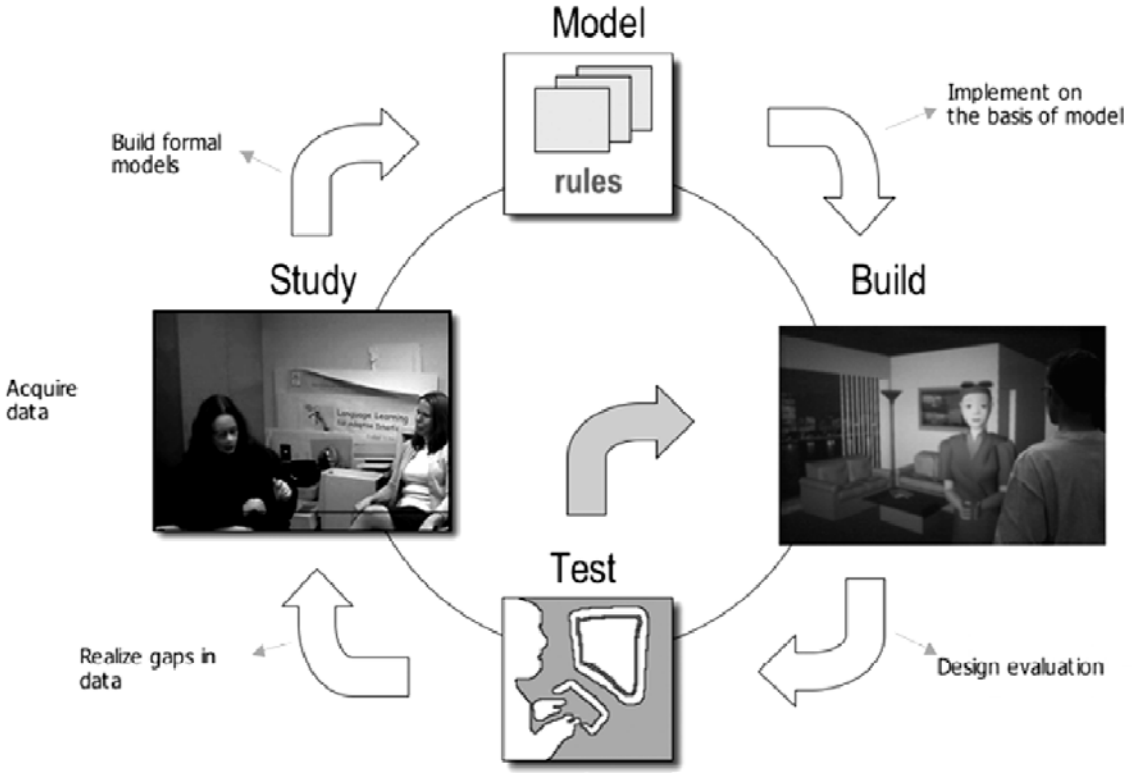
The illusion of life
(Thomas & Johnston 81)

"This was evident in sports films…any person starting to move always began to move with the hips"

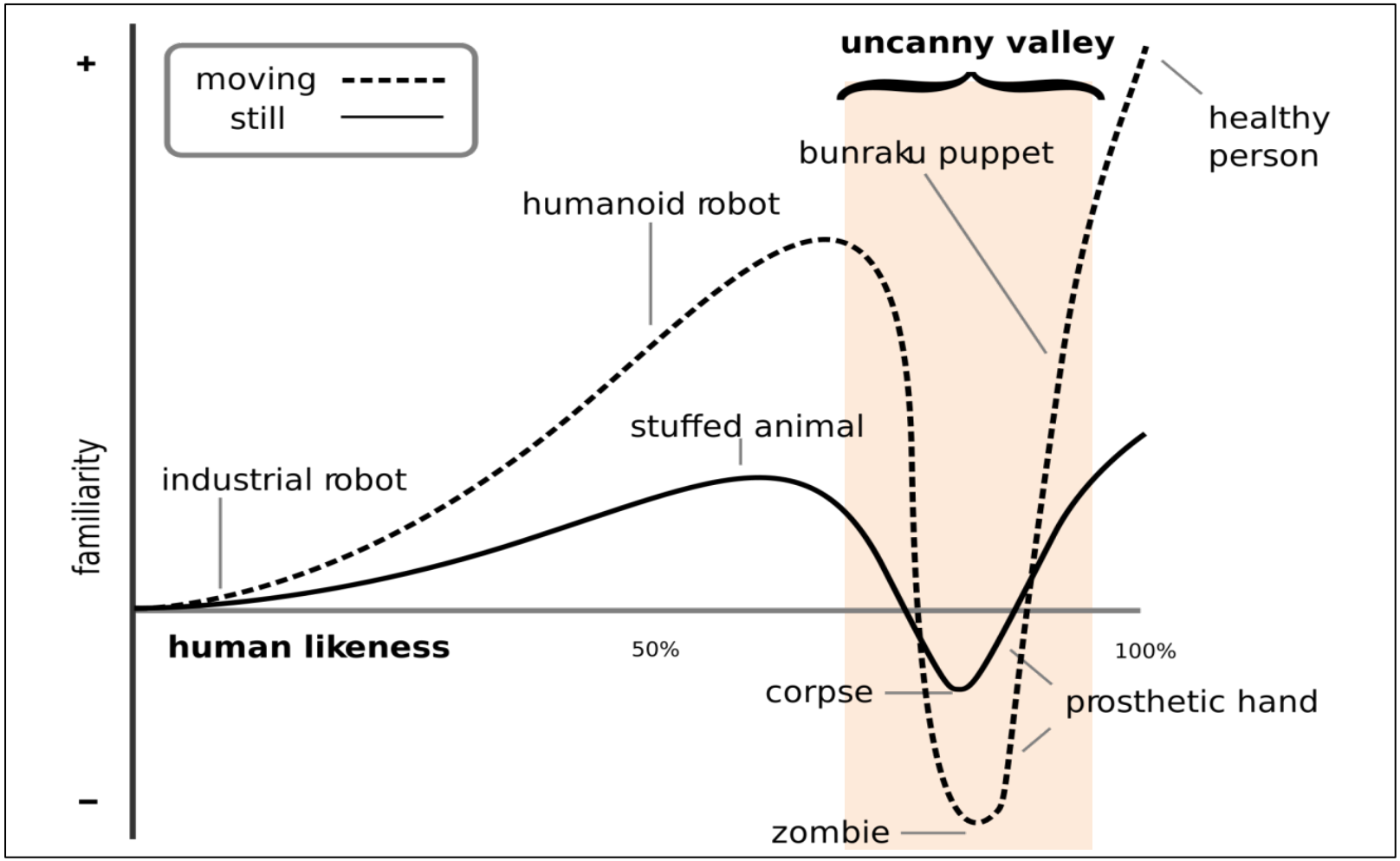
- Life-likeness
- Study human communication

Informing the design of virtual humans

- Methodology for modeling human conversation and building embodied conversational agents (Cassell 2007)



Realism: The uncanny Valley



Mori, 1970

Des observations expérimentales contradictoires (impact de la personnalité ?)

Moore, R. 2012: un modèle mathématique basé sur la frontière entre catégories

Realisme
Doublure numérique temps-réel

- Projet FUI
- Fidélité de la doublure numérique par rapport à un référent
  - Graphique
  - Expressions faciales
  - Vocale
- Thèse Mehdi Boukhris
  - Modélisation
  - Mesures
  - Évaluations
- Post-doc Matthieu Courgeon

FEEDBACK FROM A VIRTUAL AGENT IMPACT ON USERS

Types and timing of feedback

- Formative (immediate ; no score)
- Summative (at the end ; score)

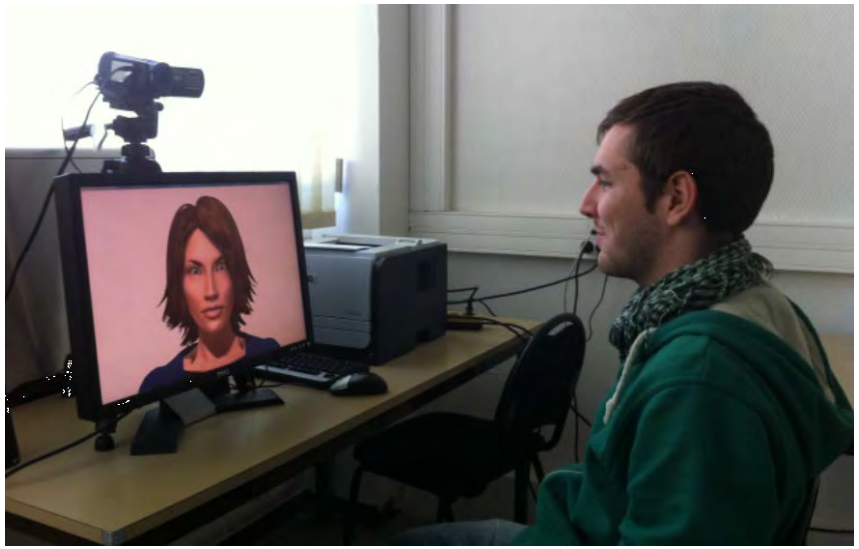
Impact of feedback in pedagogy

- Formative assessment aims to enhance learning through providing appropriate feedback (Shute 2007).
- For feedback to be formative (Shute 2008) and feedback to enhance learning (Hattie and Timperley, 2007)
  the feedback needs to be formulated, delivered and framed in such a way that it invites learners' active engagement with the feedback.

Impact of feedback in eLearning

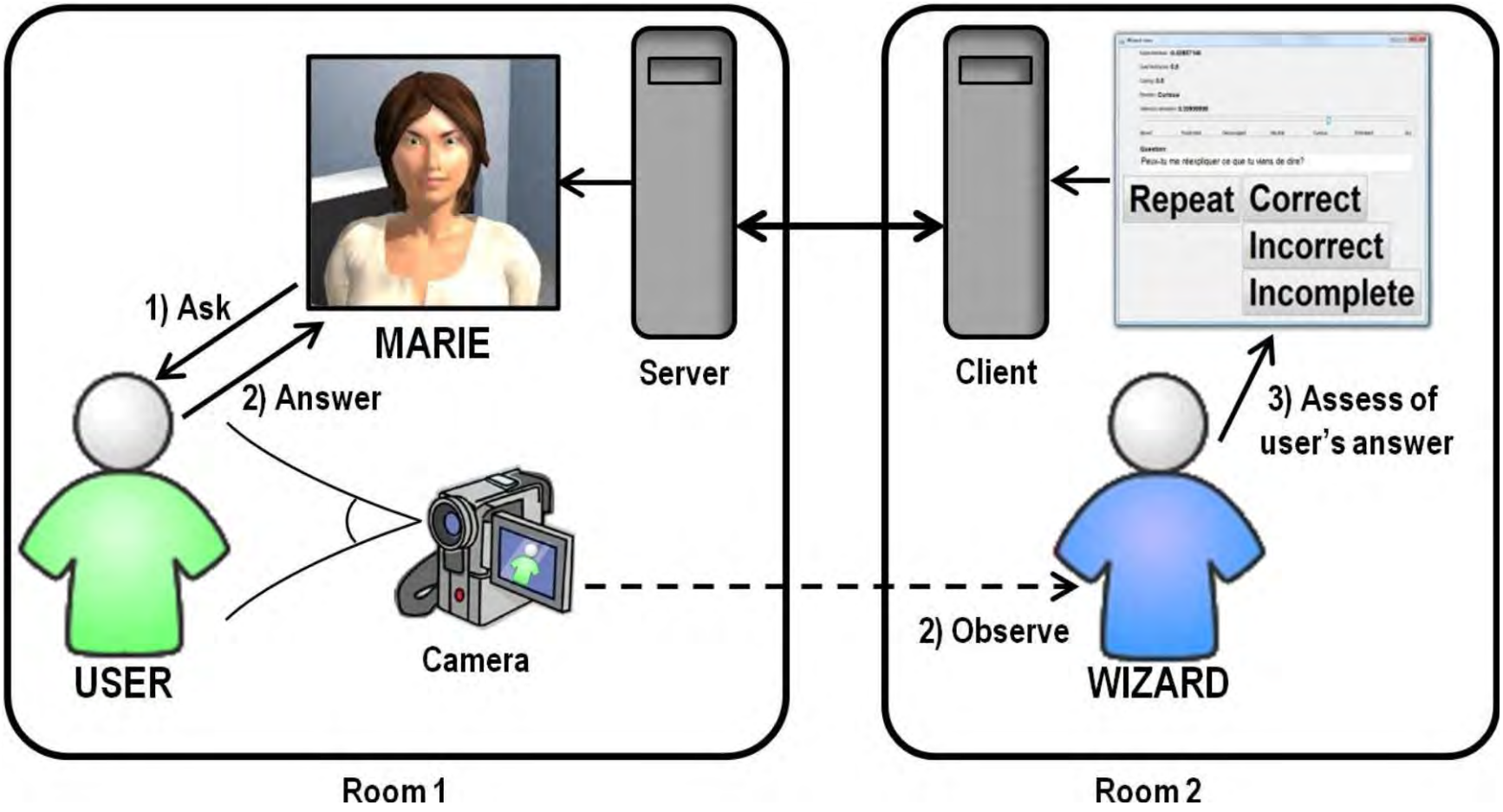
- Feedback can vary according to several dimensions
  - valence and style (e.g., positive vs. negative)
  - Timing (immediate vs. summative)
  - who/what delivers the feedback
- Feyzi-Behnagh and Azevedo (2012)
  - Pre-test and post-test
  - The results revealed that participants receiving immediate feedback significantly outperformed those who did not receive immediate feedback.

Facial Expressions of Feedback from a Virtual Peer (Eyharabide et al. in prep)



- Goal
  - study the impact of the facial expressions of the character when providing positive vs. negative feedback.
- Method
  - a virtual peer
    - asks questions to students,
    - assesses students' answers
    - provides positive vs. negative formative feedback using facial expressions (CPM)
  - Pre and post tests about Java
  - 2 conditions
    - Facial expressions
    - No facial expressions
  - 25 male students
    - Age: min 19, max 24, avg 20

Wizard of Oz set-up



Hypotheses and results

- 3 sets of hypotheses in mind
  - developing the students' knowledge (pre / post test)
    - Overall higher scores at post test
    - Only students in the non-expressive condition were able to improve their knowledge
      - Split attention effect ? Incongruent non expressive speech ?
    - Positive feedback leads students to provide more complete sentences
    - Feedback interpreted on the self-level, rather than the task-level, diverts cognitive resources away from the task and towards the negative affective states such as anxiety that arise due to the self-focus (Gusnard, 2005)
  - students’ behaviors (log files / videos)
  - students’ answers to the subjective questionnaire

Collaboration avec le MIT Media Lab

Job interview (Hocque, Courgeon et al. Submitted)

- Objectifs
  - Automated system for practising job interviews
- 28 interviews analysed
  - Career counselors maintained neutral face during the entire interview process with occasional back channeling behaviors (nodding of head), and then provided feedback after the interview has been completed
  - Types of questions

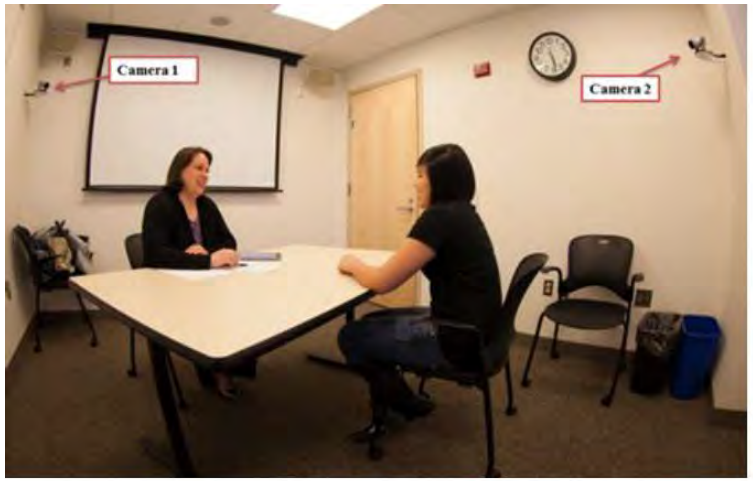


Figure 3. Experimental Setup between the interviewer and the interviewee. Camera 1 is recording the facial expressions and audio of interviewee, whereas Camera 2 is recording the facial expressions and audio of interviewer

Humains Virtuels et Comportements Sociaux

Jean-Claude MARTIN
Univ. Paris Sud / LIMSI-CNRS
Thème Agents Virtuels & Emotions