

humaine

D6f

**Exemplar: How to Build an
Affective Interactive ECA**

Workpackage 6 Deliverable



Date: 30th November 2006

Revised 6 February 2007 with input from TCD

IST project contract no.	507422
Project title	HUMAINE Human-Machine Interaction Network on Emotions
Contractual date of delivery	<i>November 30, 2006</i>
Actual date of delivery	<i>November 30, 2006</i>
Deliverable number	D6f
Deliverable title	Exemplar: How to Build an Affective Interactive ECA
Type	Report
Number of pages	28
WP contributing to the deliverable	WP 6
Task leader	University of Paris 8
Author(s)	Catherine Pelachaud and WP6 Partners
EC Project Officer	Philippe Gelin

Address of lead author: Catherine Pelachaud

IUT de Montreuil
University of Paris 8
140 rue de la Nouvelle France
93100 Montreuil
France

Table of Contents

1	THE PLACE OF THIS REPORT WITHIN HUMAINE.....	4
2	BRIEF OVERVIEW OF WORKPACKAGE 6 AND ITS EXEMPLAR	5
2.1	The field covered by Workpackage 6	5
2.2	The research objectives.....	6
2.3	Overview of main elements of the exemplar	6
3	ACHIEVEMENTS TOWARD MAIN ELEMENTS OF EXEMPLAR.....	8
3.1	Element 1: Cognitive influences on action	8
3.2	Element 2: Creating Affective Awareness	9
3.3	Element 3: Backchanneling	11
3.4	Element 4: Coordination of signs in multiple modalities	13
3.5	Element 5: Expressivity	16
4	PLANNED PROGRAM OF RESEARCH	20
4.1	Element 1: Cognitive influences on action	20
4.2	Element 2: Creating Affective Awareness	20
4.3	Element 3: Backchanneling	22
4.4	Element 4: Coordination of signs in multiple modalities	23
4.5	Element 5: Expressivity	24
5	REFERENCES	27

1 The place of this report within HUMAINE

The HUMAINE Technical Annex identifies a common pattern that is followed by most of the project's workpackages

The measure of success will be the ability to generate a piece of work in each of the areas which exemplifies how a key problem in the area can be solved in a principled way; and which also demonstrates how work focused on that area can integrate with work focused on the other areas. We call these pieces of work *exemplars*. The exact form of an exemplar is not prespecified: it may be a working system, but it might also be a well-developed design, or a representational system, or a method for user-centred design. (p 4)

To that end, each thematic group will work out a proposal for common action, embodied in one or more exemplars to be built during the second half of the funding period (p.16)

The process will begin with production by each thematic group of a review of key concepts achievements and problems in its thematic area; and drawn from the review, an assessment of the key development goals in the area. This review and assessment will be circulated to the whole network for discussion and comment, aimed both at building understanding of basic issues across areas, and at identifying the choices of goal that would be most likely let the different groups achieve complementary developments. That consultation phase will provide the basis for deliverables in month 11, which describe in some detail a few alternatives that might realistically be chosen as exemplars in each area, and their linkages to issues in other thematic areas. A decision and planning period will follow, involving consultation within and between thematic areas, leading to presentations at the second plenary conference, which will describe a single exemplar that has been chosen for development in each area, and the way work on the exemplar will be divided across institutions. The remainder of the project will be absorbed in developing the chosen exemplar. (p. 21)

The consultation phase has now ended. Near-final plans were presented to the whole network at the Plenary in May 2005, and adjustments have been made accordingly.

This deliverable reports the plans that have now been set out for the remaining 25 months of the project. They are necessarily provisional, because they will be subject to two reviews (in 2006 and 2007) before they are completed.

Work has begun on several aspects of the planned programme. It is also reported. Ethical issues affect the whole of HUMAINE, but rather than repeating essentially similar points in multiple deliverables, they will be handled coherently in a single document, D0o (Science and Society).

2 Brief overview of Workpackage 6 and its exemplar

Within WP6, we are particularly interested in the role of emotion in interaction. This encompasses abilities in three domains, perception, interaction and generation, as an ECA must also be able to perceive and interpret emotion as well as show emotion in order to interact in a desirable emotional manner. Thus, the core WP6 exemplar is to propose the definition of an Affective Interactive Embodied Conversational Agent (ECA) system with capabilities beyond those of present day ECAs. We refer the reader to deliverables D6b, D6c, D6d and D6e for the evolution of the exemplar and in-depth descriptions of the aspects highlighted here.

2.1 The field covered by Workpackage 6

An ECA needs to perceive the user's emotional state and the context; she has to react to the user's action and emotional states as well as to the events happening in the context; she also needs to show emotions, to adapt her behaviour to the user behaviour. She needs to catch the user's attention, to maintain it while she is conversing with the user... WP6 tackles these issues over three main domains:

- Perception domain: In order to interact in proper manner, an ECA must first be able to pay attention to, perceive and interpret her conversational partners (e.g. the user) and the context she is placed in; only then can she hope to adapt her (verbal and nonverbal) behaviour depending on the user, her role, and the social and cultural context she is placed in. At a fundamental level, the ECA should have some notion of conversational partner, emotion, gesture and environment.
- Interaction domain: Interaction is the central theme of WP6. Thus, we are particularly interested in the role of emotion (including emotion proper, mood, interpersonal stance and attitude) in interaction, i.e., how to create an interactive ECA able to not only show and perceive emotion, but also adapt her behaviour to convey some emotion or to induce some emotional state in the human user. Interaction is therefore the bridge between perception and generation.
- Generation domain: Even though the presentation side of ECAs is the most developed in ECA research, further improvement of this area is required, especially as regards a) multimodal integration and the display of emotional behaviours to improve the naturalness of ECAs, and b) the believability of ECAs which is influenced by the consistency of an ECA's behaviour be it in terms of personality, cultural context and situation. Even though decisions on the form of generation (e.g. a particular facial expression or hand gesture) are taken on the basis of the aforementioned domains, it is the generation part that provides the capability for a proof-of-concept implementation and testing of those concepts and essentially initiates interaction with the user and/or other ECAs.

2.2 The research objectives

Following the consultation period, the exemplar proposed for WP6 is Definition of *Affective Interactive Embodied Conversational Agents*. Such ECAs should have several capabilities within the three considered domains:

- Perception Domain:
 - Cognitive influences on action
- Interaction Domain:
 - Create affective awareness
 - Backchanneling
- Generation Domain:
 - Coordination of signs in multiple modalities
 - Produce dynamic expressive visual and auditory behaviours

2.3 Overview of main elements of the exemplar

1. *Cognitive influences on action*

This capability is concerned with certain aspects related to cognition that may influence the actions of an agent, such as gaze behaviours, gestures and linguistic expression. It consists of two key sub-topics: attention shifting related to emotion and the adaptation of politeness behaviours to a user's emotional state.

2. *Creating Affective Awareness*

In this capability, we seek to create connections between a user and an ECA with the aim to maintain user's engagement during an interaction. Sub-topics in this element are creating affective bonds between the user and ECA, imitation and adaptation to generate or alter the behaviour of the ECA.

3. *Backchanneling*

Backchannel utterances and signals can convey simple information of critical importance to the success of communication. Here we seek to investigate backchanneling undertaken by the listener in order to provide the speaker with immediate feedback about the state of the communication, in order to improve the interaction capabilities of an ECA.

4. *Coordination of signs in multiple modalities*

ECAs must show expressivity in a consistent and natural looking way across modalities in order to be believable. Sub-topics in this element consider coordination of multimodal behaviour from corpora, relationships between signs of emotion in

different modalities and gesture repositories for generating appropriate multimodal ECA behaviour.

5. Expressivity

In this capability we are considering expressive aspects of speech and behaviour by looking at sub-topics such as an intermediate level of behaviour parameterisation utilising dimensions of expressivity, expression through acoustic parameters of the voice based on speech synthesis techniques and emotional body gesture accounting for the situation context.

3 Achievements toward main elements of exemplar

3.1 Element 1: Cognitive influences on action

This capability is concerned with certain aspects related to cognition that may influence the actions of an agent, such as gaze behaviours, gestures and linguistic expression. In particular, it investigates attention shifts accounting for emotional stimuli and also the adaptation of an agent's politeness behaviours to an emotional user. These issues are related to WP7 themes and collaboration with WP7 has taken place. Here, we emphasise social aspects.

a. *Emotion related attention shifts*

Emotion related attention shifts are concerned with the guidance and allocation of attention by an agent to emotional or potentially important stimuli within a scene for the purposes of conversation control. These shifts may result in the production of visible phenomenon that are of significance to others in the environment.

Recently, we have considered several aspects of agent interaction related to early stages of appraisal concerned with social perception and attention. We constructed a real-time model of conversation initiation for agents in a virtual environment [Peters 06a]. We have conducted a user evaluation to gain a better insight into human perception of attention and interest behaviours made by an artificial agent [Peters 06b]. The perception of an agents gaze, body orientation and locomotion direction was evaluated from the user's perspective for a number of different situations, featuring both imagery and dynamic behaviour. This evaluation yielded interesting results regarding human perception of attention and emotion related to contrast between body-segment orientations. We investigated some concepts related to the notions of attention and interest with the aim of enabling an ECA to better remember [Peters 06c], initiate and maintain an interaction [Peters et al 06]. An ECA monitors the amount of the attention that an interaction partner is paying and derives an idea of the amount of interest they have in the interaction. Interaction strategy can then altered or terminated if the interaction partner is not showing enough interest in the speaker.

b. *Adapt politeness behaviours to the user's emotional state*

Since emotions have an important impact on perceived face threat, here we seek to create an ECA that can adapt to the user by trying to mitigate such threats through use of gesture, mimics and speech and will also account for the causes of such emotions, rather than just their strengths.

To facilitate this line of research, a multimodal corpus of user-agent interactions was collected within the Gamble testbed. Gamble allows for creating highly emotional situations and at the same time gives us a high degree of control over situational factors (e.g., [Rehm and André, 2005a]). For this study we employed the agent with the ability to fake his facial expression. This was done following Ekman's insights into facial cues to deception. Thus,

different degrees of false emotional displays are possible depending on if the user shall be able to detect the “real” emotion of the agent or not. 24 students participated in this study and were divided into 12 teams. Each team played two rounds of 12 minutes, participants changed position after the first round. Interactions were videotaped and the game progress was logged for the analysis.

3.2 Element 2: Creating Affective Awareness

In this capability, we seek to create connections between a user and an ECA with the aim to maintain user’s engagement during an interaction. Such ECAs should not only perceive the user’s emotion, but also adapt to it, react to it and imitate it.

a. *Creating Affective Bonds*

The creation of a bond between the user and ECA is investigated through monitoring the users level of engagement with an ECA and also by providing an explicit means for the ECA to try influence the users level of engagement in the interaction.

The creation of a bond between the user and ECA is investigated through monitoring the user’s level of engagement with an ECA and also by providing an explicit means for the ECA to try influencing the user’s level of engagement in the interaction.

We can roughly distinguish between the need for emotional displays to improve the user’s awareness of the agent and the user’s overall engagement in the interaction (see below 3.2.d), and the need for suitable measurements of affective bonds. The Gamble corpus study allowed us to tackle these two problems. Different users were confronted with two different conditions in which the agent’s display of emotional facial expressions was varied [Rehm and André, 2005a]. In one condition, users were confronted with an agent displaying its “real” emotions, in the other condition, the agent tried to hide its emotions following Ekman’s observations on facial cues to deceit.

One measurement for affective bonds is the user’s attention towards the agent which can be assessed by analysing user’s gaze behaviours during the interaction. The Gamble corpus was analysed accordingly [Rehm and André, 2005b]. While the users’ behaviours in the user-as-speaker condition were consistent with findings for human-human conversation, we noticed differences for the user-as addressee condition. People spent more time looking at an agent that is addressing them than at a human speaker. Maintaining gaze for an extended period of time is usually considered as rude and impolite. The fact that humans do not conform to social norms of politeness when addressing an agent seems to indicate that they do not regard the agent as an equal conversational partner, but rather as a (somewhat astonishing) artefact that is able to communicate. This attitude towards the agent was also confirmed by the way the users addressed the agent verbally.

b. *Imitation*

Imitation allows the creation of an affective “awareness” and coordination, and to alter affective bonds by allowing agents to be capable of mimicking certain aspects of another, so that synchrony of behaviour may in some circumstances also lead to a synchronisation of emotions.

We have been working on the creation of a virtual agent able to perceive and interpret users’ expressivity and to respond properly to them. Such agents represent a powerful human-computer interface, as they can embody characteristics that a human may identify with and may therefore interact with the user in a more empathic manner. In our system, for each gesture performed by the human, the agent responds with a gesture that exhibits the same quality [Castellano et al. 2006]. We do not aim at copying the user’s gesture shape but only the quality: that is, some of the physical characteristics of movement. To create such a system, we have extended our expressivity model to act either on a sequence of consecutive gestures (the gestures show the same expressivity) or on single gestures (each gesture has their own expressivity).

c. *Adaptation*

In order for an agent to be adaptive, information about the emotional state of the user(s) must be obtained. Here, the user’s expressive gestures and language usage are analysed to infer his/her emotional state – the inferred state is then be used to alter the ECA’s reactions.

At first we have conducted steps towards the development of a computational model for emotion recognition by means of the analysis of the dynamics of movement and gestures. A new library for extraction of relevant motor features to give in input to machine learning algorithms for emotion recognition was developed in the EyesWeb platform. The new set of features starts from movement cues such as quantity of motion, contraction/expansion, fluidity, etc. and extracts information about their dynamics.

A preliminary test of the library was done in a pilot experiment aiming at recognising emotions in music performances by means of the movement analysis of performers [Castellano et al. 06a].

In parallel we have elaborated a scenario whereby an agent senses, interprets and copies a range of facial and gesture expressions from a person in the real-world. Input is obtained via a video camera and processed initially using computer vision techniques. It is then processed further in a framework for agent perception, planning and behaviour generation in order to perceive, interpret and copy a number of gestures and facial expressions corresponding to those made by the human. By perceive, we mean that the copied behaviour may not be an exact duplicate of the behaviour made by the human and sensed by the agent, but may rather be based on some level of interpretation of the behaviour. Thus, the copied behaviour may be altered and need not share all of the characteristics of the original made by the human.

d. *User's engagement*

Based on the analysis of user's gaze behaviour which is also an appropriate measure of user's engagement in an interaction, the Gamble system was improved. The new system also features a tangible user interface (CamCup) and a speech recognition system for input [Rehm and André, 2005c]. Integrating the speech recognition component was an indispensable prerequisite to test the recognition of emotional speech (see 4.2.a.).

A system in which audio-visual feedback is provided to users who are free to move in the space has been developed [Castellano et al. 06b]. The feedback depends on the expressivity of the body movements they perform. The system is based on the integration of two different software platforms: EyesWeb, for analysis of human full-body movement, and pDM, a software developed at KTH, and that allows to render a music performance with different emotional characterisations by manipulating acoustic parameters such as sound level, tempo and articulation. The system analyses the users' motor behaviour, extracts expressive indicators (quantity of motion and contraction/expansion of the body) and controls the sound expressivity manipulation and the visual feedback generation. Visual feedback was designed to respond to a specific expressive motor behaviour of the user, by using appropriately colours changing depending on the emotions associable to the movement characteristics analysed. Acoustic feedback was designed with pDM accordingly to the motor features extracted in real-time with EyesWeb. In the system, the dynamic variations of the motor cues control the dynamics of acoustic cues such as tempo, sound level, articulation.

3.3 Element 3: Backchanneling

ECAs need to become interaction partners. A key feature of spoken dialogue is backchanneling by the listener, providing the speaker with immediate feedback about the state of the communication. Backchannel utterances and signals can convey simple information of critical importance to the success of the communication attempt, e.g., whether the addressee is still listening, understands, or agrees with the spoken content.

Work has progressed along various interconnected lines intended to gain a better understanding of the mechanisms involved in backchanneling, and the set of behaviours and determinants involved in the process. Also, first steps were made towards defining formal models, implementation and evaluation. Steps taken include: (1) data collection and analysis, (2) automatic detection and interpretation of relevant cues of the speaker, (3) designing a framework for defining rules for the agent, working on a feedback lexicon, (4) setting up an evaluation framework and (5) evaluating some aspects of the feedback lexicon.

Work has followed several lines:

- **Conceptually**, we have aimed to clarify the notions of backchannel vs. feedback [Schröder et al 06b], distinguishing the semantic level from the occurrence level. Feedback, as the more general term, encompasses, on the semantic level, both grounding and new information, which can occur either in a separate turn or during the speaker's turn. Backchannels, in this

conception, are a subset of feedback, which typically provides only grounding information and typically only occurs during the speaker's turn.

- We have elaborated a common taxonomy for backchannel signals. The taxonomy is based on the types of information backchannel signals can provide. Each of the classes in the taxonomy can be either positive (e.g. accept) or negative (do not accept). Any combination of the classes can happen: If one is interested, one can check on the quality of the perceived signals; I understand what you say but I disagree, etc.
- A small **corpus** of eight two-person conversations, lasting on the average 15 minutes was collected. This corpus was **annotated** on several levels, mainly looking at gaze and head movements of speakers. The main result of this work has been to gather together examples of behaviours and their determinants that have been described in the literature and to get a feel for differences between individuals.
- Work on the **detection** of low-level features and their classification focuses on tracking head movements and interpreting them in context as many of the back-channels in conversation are elicited and prohibited by head movements of speakers.
- In an **experimental study** [Schröder et al 06b], we investigated the specific subset of *acoustic* emotional feedback occurring during the speaker turn. Assessing the perceptual acceptability of inserting affect bursts as emotional feedback into a speaker utterance, we found a pattern that lends itself to an explanation in terms of social acceptability, or display rules.
- Another **experimental study** was set up where subjects were asked to derive meanings for each *visual* signal: subject could view the signal once only and could not replay the signals. The aim of the test is to evaluate if subjects are able to attach a meaning to a given signal.
- An **evaluation framework** has been set up in which off-line generated back-channel behaviours of an ECA are combined with existing recordings of people so that it appears that an actual conversation between ECA and the real person took place. This framework is used for the evaluation of behaviours that are generated by hand and by rule. A first version of a system to define different kinds of back-channel behaviours was implemented.
- On the level of **software architecture**, a collaboration with the VirtualHuman project was sought within DFKI in order to advance towards the creation of a proof-of-concept ECA system capable of emitting backchanneling behaviour. We could adapt the Conversational Dialogue Engine (CDE) code [Pfleger and Löckelt 06] for future use in emotion-oriented interaction research. The original CDE code developed within VirtualHuman had been tuned towards application in the Soccer domain; we extracted a reasonably domain-independent subsystem, and combined it with the Greta agent as a player technology.

3.4 Element 4: Coordination of signs in multiple modalities

In order to be believable, ECAs must show expressivity in a consistent and natural looking way across modalities. The ECA has to be able to display coordinated signs of emotion during realistic emotional behaviour. Such a capability requires one to study and represent emotions and coordination of modalities during non-basic realistic human behaviour, to define languages for representing such behaviours to be displayed by the ECA, to have access to mono-modal representations such as gesture repositories.

a. *From multimodal emotional corpora to models of coordination between modalities*

We aim at building a model of the coordination of multimodal behaviours during non basic multimodal behaviours observed in non acted data, using multimodal corpora.

Several studies about emotion in the social sciences are based on behaviours performed by actors during in-lab recording when being instructed to act a single emotion. Real life emotions are often complex and involve blends of emotions.

Our work involves several steps:

- **Manual annotation:** In order to study multimodal behaviours during real-life emotions, we have collected a corpus of emotionally rich TV interviews [Abrilian et al. 05a]. Several levels of annotation were manually coded using Anvil: some information regard the whole video (called the "global level"); while some other information are related to emotional segments (the "local level"); at the lowest level, there is detailed time-based annotation of multimodal behaviours. The annotation of multimodal behaviour includes gesture expressivity since it was observed to be involved in the perception of acted emotional behaviours. We adapted the set of expressivity parameters defined in the Greta agent to the annotation of perceived expressivity at several temporal levels (global clip, gesture unit, gesture phrases, gesture phases): activation, repetition, spatial extent, speed, strength, fluidity. We also annotated classical dimensions of gesture studies: gesture phrase and gesture phase. The goal is to enable future studies on the relation between modalities; for example studying the cross-modal relations reported in the literature (e.g. gesture stroke co-occur with or precede the lexical affiliate) are also observed in such rich emotional behaviours. Multimodal emotional profiles are computed from these annotations.
- **Automatic annotation:** In order to validate these manual annotations of movement and global activation, we have explored how manual annotation and image processing can cooperate towards the representation of spontaneous emotional behaviour in such low resolution videos from TV. We observed that the manual annotation of

emotional activation was correlated with the automatic estimation of movement expressivity computed by image processing.

- In order to validate these manual annotations and augment the manual annotations by expert coders we proceeded to two **experimental studies**.
 - The first experiment [Abrilian et al. 06] aimed at studying if the expressivity parameters can be reliably perceived and reported by people. We used 30 clips of EmoTV and a range of subjects. We observed good correlations for spatial extent, activation, repetition, strength. We observed low correlation for fluidity and speed. A Principal Component Analysis organised the clips according to two axes. The horizontal axe was interpreted as opposing clips that obtained high scores and clips that obtained low scores in average of the 6 parameters. The vertical axe was interpreted as opposing fluidity and repetition.
 - The second experiment evaluated the perception of multimodal cues at a global level in a TV clip. A woman is emotionally reacting to a law decision that she considers as unfair (her father and brother are being kept in jail). Forty different coders annotated the clip. We used three groups of coders with respect to gender, age and expertise. We observed some differences between female and male coders: male coders reported more torso movement and brows cues than female coders. Female coders reported more eyes cues than male coders. Future perceptive tests with perception of male emotional reaction (anger)
- The **copy-synthesis approach** aims at replaying the annotated behaviours coded in EmoTV by the Greta expressive agent. Such an approach can be useful to validate the annotation and the representation, iteratively refine the annotation schemes and the expressive agents, and for the specification of blends of emotion in expressive agents. It involves several steps: annotation, extraction, representation, and generation. Several representations are computed from the manual annotations and are used to generate the animation (see next point).
- **Generation of videos:** In the copy-synthesis approach, we have defined two corpus-based approaches to design different Greta animations based on the video annotations [Martin et al., 2006]. The “multiple levels replay” approach involves the level of annotation of emotions, and the low-level annotations of multimodal behaviours (such as the gesture expressivity for assigning values to the expressivity parameters of the ECA, and the manual annotation of facial expressions) [Martin et al., 2005]. The “facial blending replay” approach is identical to the “multiple levels replay” approach except for facial expressions: it uses a computational model for generating facial expressions of blend of emotions [Martin et al., 2006b].

- We used such an approach to evaluate if people detect properly the signs of emotions in different modalities (speech, facial expressions, gestures) when they appear to be superposed or masked. We compared the **perception of emotional behaviours** annotated in a corpus of TV interviews and replayed by an expressive agent at different levels of abstraction [Buisine et al., 2006]. The results provide insights on the use of such protocols for studying the effect of various models and modalities on the perception of complex emotions.

b. *Gesticon*

We need to represent the relationships between the signs of emotion in different modalities, i.e., by designing a representation language for multimodal behaviors.

We have been working on a first specification of such a representation language. The aim of this work is to define a language to describe behavior in a format independent of players and graphics models. The idea of elaborating such a language came up when considering how time consuming and how difficult the creation of gesture shape, facial expression, body posture etc. can be. Having a common description language would allow for the mutualisation of work. Being able to share behavior definition would greatly help the agent community. The representation language developed so far is to be used for defining Gesticons. A Gesticon acts as a behavior description repository. Each entry of this repository is a unit of nonverbal behaviors. Signals are described hierarchically across modalities. Each unit of behavior, also called Form Element, can be of any of the following types: gesture (coordinated movement with arms and hands), hand configuration (hand shape, fingers, orientation of the thumb), facial expression (muscular contraction), gaze (eye and head direction (include neck and shoulder)), head (movement of the head independent of eyes), upper body (movement of the spine and shoulder) and posture (movement of the body elements downward from the hip). The division into this list of Form Elements arises from several considerations: physiology (muscular contraction and joint articulation), existence of studies on communicative non-verbal behaviors, computational factors (same hand shape used in different arm movements).

Gesticon entries provide a) information for planning multimodal behaviour, and b) information that constrains the actual realisation of multimodal behaviour. At the planning stage of multimodal behaviour, information is provided in form of player and context independent form/movement descriptions including rudimentary duration information such as min, max, default, in order to know how far a gesture or gesture phase can be stretched or shrank without changing its meaning/function. This information is important at realiser level, when the behaviours are fine-tuned according to an absolute timeline, which is in the current ECA systems usually defined by the timeline resulting from speech synthesis.

c. *Context dependent emotional body gesture*

We aim to provide ECA's with the capability of modulating its body gestures (in particular upper-body: head, shoulders, arms, hands) according to the emotional state induced by the context –situation- in a virtual environment.

An example of a body gesture that depends on the context is reactive movement. Reactions are unconscious behaviours that are not often implemented in virtual humans. The ability of reacting of virtual characters improves their realism. Based on an observation experiment we have made a semantic model to synthesise the reactive animation of virtual humans [Garcia-Rojas et al. 06]. This semantic model includes not only the animation synthesis and the geometry, but also a description of individualisation of virtual humans through personality and emotion models, age, gender, etc [Moccozet et al. 06]. The mentioned Individual descriptors are used as inputs of a modularised behavioural reactive controller, which sends inputs to the animation engine as well. In our test application we present tree kind of reactive movements of different individualised characters. The synthesis of the animation was made using an inverse kinematics technique, it includes only the upper body motion (both arms and spine). We provided a simple behaviour controller that makes an appraisal of the movement that the character may perform according to the characteristics of the stimuli and to the character's individuality. The advantage of the model is that allows scalability and a modularised implementation of the concepts.

3.5 Element 5: Expressivity

Emotion is not simply expressed through a static facial expression or setting of vocal cords. Acoustic and visual expressions of emotion is dynamic; they evolve through time. The manner a movement is done provides relevant information on the affective state of the emitter. In this capability we aimed to consider the quality aspect of behaviour.

a. Behaviour expressivity

We aim to investigate the generation of phenomenologically accurate behaviours as we focus on developing an intermediate level of behaviour parameterisation, a set of dimensions of expressivity.

Our work was developed in several areas:

- At first we have extended our gesture expressivity model to **facial expressions**. The six dimensions (spatial extent, temporal extent, power, fluidity, repetitivity and overall activation) have been implemented for facial expressions. The parameters act on the intensity of muscular contraction as well as on the temporal parameters of facial expressions.
- **Complex facial expressions:** Then we have elaborated a computational model of facial expressions arising from blends of emotions. It is based on a face partition approach. Any facial expression is divided into n areas. Each *area* represents a unique facial part like brows or lips. The model computes the complex facial

expressions of emotions and distinguishes between different types of blending (e.g., superposition and masking). The complex facial expressions are created by composing the face areas of the two source expressions. Different types of blending are implemented with different sets of fuzzy rules for the computation of the complex facial expression. The fuzzy rules are based on Ekman's research on blends of emotions (Ekman & Friesen, 1975).

- **Copy of expressive motion:** This work focused on the generation of copying behaviour in the Greta agent starting from real expressive motion [Mancini et al. 06]. We analysed human full-body movement for animating Greta. The system we developed takes in input video data related to a dancer moving in the space. Analysis of video data and automatic extraction of motion cues was done with the EyesWeb platform. We considered the quantity of movement and the contraction/expansion. For the animation of Greta, we mapped these motion cues on the corresponding expressivity parameters of Greta.

b. Speech Expressivity

We want to better understand how emotions and related states are expressed through the voice, and how the expression through the corresponding acoustic parameters can be achieved using speech synthesis techniques.

We have worked on the following topics:

- Considerable progress towards **expressive unit selection** has become possible by collaboration with the new basic research project PAVOQUE (PArametrisation of prosody and VOIce QUality for concatenative speech synthesis in view of Emotion expression), which started in June 2006 at DFKI and is funded by the Deutsche Forschungsgemeinschaft. Early work has concentrated on the preparation of a suitable research platform and baseline system for expressivity-related parametrisation of high-quality concatenative speech synthesis [Schröder et al 06a]. Using this technology, we aim to improve on the sound quality which was limited in our previous work [Schröder 06] due to the use of diphone synthesis; the aim is to achieve better quality by combining selection of suitable units based on expressive target costs with a moderate amount of post-processing using signal modification techniques.
- **Extend the research on multicultural diversity emotion display rules:** the multilingual listening experiment described in [Burkhardt et al. 2006] has been conducted in Hungarian, an extended article will appear. The results supported the observations described in [Burkhardt et al. 2006] even though the experiment had to be done with a voice from a different gender.
- Add **new interfaces to the synthesiser** to enable integration in ECA frameworks: interfaces to the MARY speech synthesiser and to the

Festival framework have been implemented that enable researchers to integrate emotional expression with the respective synthesisers.

- Plan the next generation synthesiser which will support graded, blending and time-varying emotional expression: the plan for future extension has been addressed with fruitful discussion during the Humaine meetings, e.g. where the graded and time-variant emotions will be solved by parameter interpolation, the blending of emotions is not so straightforward but of a high scientific interest. A starting point will be experiments with combinations of different parameter-categories denoting different emotion-related states, e.g. combining a sad intonation contour with a voice quality appropriate for anger.
- Use of different acoustic parameters: We have developed a new analysis-synthesis method based on the ARX-LF model [Vincent et al., 2005] which allows independent modification of different acoustic correlates, such as fundamental frequency, intensity, vocal tract information and also voice quality features (Open Quotient, asymmetry and return phase coefficients). A copy synthesis experiment was proposed so as to attest the importance of each of these acoustic parameters in conveying attitudes [Aubergé et al. 06a], [Aubergé et al. 06b]. This experiment clearly showed that the stimuli resynthesised with ARX-LF model were better recognised than those resynthesised with TD-PSOLA. It also confirms the importance of voice quality parameters in expressive speech synthesis, but outlined that the manipulation of the three components of the ARX-LF model, namely the LF glottal waveform, the vocal tract and the residual information should be done in a coherent way.
- We also worked on the **restitution of focus** using corpus-based speech synthesis techniques. The main purpose was to modify the selected speech units so as to make the focus more prominent. These first attempts globally show that such a local approach could lead to interesting results, but also reveal that careful design of prosodic targets is mandatory, not only on the focal part itself but also in the vicinity of this part.
- Progress has been made towards a fuller **understanding of the voice parameters** which are exploited in the expression of various emotions and attitudinal states. The primary approach adopted has involved the use of perceptual experiments using synthetic stimuli generated with source parameter manipulation, producing a diversity of voice qualities [Yanushevskaya et al, 2006]. Of particular interest is the role of different dimensions of the voice in signaling affect: voice quality, pitch contour and loudness, and also the way in which these dimensions interact (paper to appear). The issue of voice expressivity generation is being explored, in a way that is not necessarily linked to a specific synthesis modality, through interaction with WISPR, a parallel EU project at TCD which is concerned with synthesis development [Ní Chasaide et al, 2006].

- Research has also focused on **cross-cultural diversity in the vocal expression of attitude and emotion**. Towards this end perception experiments have been carried out on different language groups using the research paradigm mentioned above. The participants were drawn from the following languages: Irish-English, Japanese, Spanish and Russian and results to date are reported in [Gobl et al, 2006; Yanushevskaya et al, 2006] showing some striking cross-languages differences which would need to be taken account of in affect-sensitive systems.

4 Planned Program of Research

4.1 Element 1: Cognitive influences on action

a. *Emotion related attention shifts*

We will be conducting a further user evaluation study investigating conversation initiation with agents. This will be based not only on gaze and locomotion direction, but also on facial expression of the agent. The differences between user reporting of agent attention, emotion and intention when facial expression is and is not present will be of significant interest to our endeavours in creating perception-based behaviour models. Continuing intended work outlined in [Bevacqua et al 06] and [Mancini et al 06], we will be investigating the notion of a 'switchable' visual sensing module for ECA's, in order for the agents to be capable of processing input information from the real world as well as the virtual environment. Among other uses, this would allow virtual models to be tested and validated with input data from the real world and would also allow for the construction of algorithms designated for the virtual environment that are tightly coupled with real world algorithms.

With WP7 partners we will be continuing investigation into attention and emotion aspects related to facial expression and novelty relation to the early stages of appraisal theory. We are actively enhancing the capabilities of our current models of attention and interaction to be able to support such functionality in real-time [Peters 06d].

b. *Adapt politeness behaviours to the user's emotional state*

The corpus of the Gamble study has to be analysed further regarding the reactions of the user towards emotional displays of the agent. This will give us further insights into which emotional displays of the agent were appropriate. At the moment, emotional reactions are triggered by cues derived from the state of the game at a given moment. The integration of a recognition system for emotional speech will allow for a more specific reaction of the agent. This system will first be tested in a single user scenario (see 4.2.d.).

4.2 Element 2: Creating Affective Awareness

a. *Creating Affective Bonds*

During the Gamble study it was observed that at least in this scenario the agent becomes more believable when it started misbehaving, e.g. by insulting the user or by overtly expressing the negative emotion of rage about the current state of the game. An analysis of the user's reactions to the agent's emotional display will reveal if this is always the case in this scenario. Moreover it was observed that users started talking about the agent instead of talking with the agent, trying to explain the agent's behaviour and reactions to one another. Although this can be interpreted as very impolite behaviour towards the agent, it is a rich source of information to gain insights in the user's ideas about the agent and its performance and has to be analysed in detail.

b. *Imitation*

The animation of the agent is done with Animation Score, a tool we have developed this year. This tool allows for the control of the agent no more from high level specification (eg the communicative functions of the APML tags) but at the level of behaviours. One can specify the behaviours the agent should display through time. Behaviours for different modalities (facial expression, gaze, head movement, gesture) can be specified in Animation Score. The tool is interactive and allows one to visualise in real-time the animation. Moreover one can specify values for all the expressivity parameters on each behaviour. We will use this tool when replaying behaviours on an agent from the analysis of behaviours of a user from a video corpus.

c. *Adaptation*

We will do further tests towards the development of a computational model for emotion recognition by means of the analysis of the dynamics of movement and gestures. Further tests will also be conducted on new sets of affective videos. Part of the work will be done in collaboration in collaboration with WP4.

In the near future we are going to create a system to allow easier implementation of new features for our ECA Greta. The system will also facilitate the integration with other systems and will work in real-time. The first application of this new architecture will be an interface between the Greta and the EyesWeb systems. EyesWeb will be used to continuously analyse video data coming from a video source, like a camera or AVI files and extract expressivity features (only the parameters of amplitude, speed and fluidity of movements will be considered for the moment). Data will be received by Greta which will not perform the same gesture but which will replicate the expressivity (amplitude, speed, fluidity) of the original movement in real-time. The modularity of the implemented system will easily allow us to switch between different modules with the same functions; that is, for example one could use expressivity data coming from pre-defined data files instead of EyesWeb for testing purposes.

We will work next on how to map user's movement expressivity to user's emotional state. Once the emotion has been recognised, the agent, when interacting with a user, can plan affective response accordingly to that. In fact, if the agent can understand the user's emotional state analysing the quality of his gesture, it can decide which facial expression to show. In this model the agent could interpret the user's state to improve the efficiency of communication.

d. *User's engagement*

We will continue investigating how to act on user's engagement through several studies:

- We will conduct experiments with subjects to test our audio-visual feedback system [Castellano et al. 06b]. We aim to evaluate the effectiveness of the audio-visual feedback in reproducing movement expressivity of subjects and the impact on user's engagement.

- A single user test scenario will be developed which will actively react to user's gaze behaviour to increase his/her awareness of the agent. This test scenario will be used to test user's reactions to emotional displays of the agent, e.g. an agent might smile if looked at and frown if the user gazes away. We will also employ a system to estimate the emotional state of the user from its speech input. The agent might e.g., give feedback to the user by mirroring his emotional state.

4.3 Element 3: Backchanneling

We will go toward developed a proof-of-concept backchannel and feedback ECA system. Following the framework for backchannel model presented in Deliverable 6d, two models should be considered: reactive and cognitive models. The former corresponds to simple mimicry emitted very often spontaneously and sincerely; on the other hand the latter reflect a conscious decision to provide backchannel to provoke a particular effect on the speaker or to reach a specific goal. Two main activities will happen:

- **Generation of rules for backchannel behaviours:** We will work on testing, evaluation and updating the back-channel framework. Work on the detection and interpretation of back-channel elicitation cues (head movements and prosodic features such as pitch detection) will first be tested on recorded data to see how well it detects possible back-channel insertion points and back-channel types. Different versions of the rules generating the back-channel behaviours will be defined and tested. Hand-crafted versions will be compared with automatically generated behaviours. Extensive user tests are needed to find out how different back-channel strategies can lead to different impressions of the agent; the way the agent is engaged in the conversation (distracted, engaged, empathic, and so on) and to match different personality types. Furthermore, integration of the various components into a real-time system is on the agenda for the next phase of the project as well. This work will lead to new, improved versions of the architecture, taking up different development cycles.
- **Reactive Backchannel:** The next steps planned concern the use of the adapted CDE architecture to implement a proof-of-concept backchannel and feedback ECA system. A subset of the architecture proposed in D6d for enabling an ECA system to show backchannel and feedback behaviour will be implemented. This will primarily concern the implementation of reactive components which do not require cognitive processing, such as an "audio feature extractor", a "reactive backchannel trigger", a "backchannel selector", and a "backchannel scheduler". To the extent that speech recognition can be used or simulated, we intend to demonstrate the intended function of the "meaning analyser", the "cognitive backchannel trigger", and the "expressed backchannel meaning planner" by the use of mockup modules. In the near future we will study results from the first evaluation study on meaning of visual behaviours signals for backchannel. Later on we aim to pursue another evaluation study related to the level of mimicry and temporal synchrony to get for reactive behaviours.

4.4 Element 4: Coordination of signs in multiple modalities

- a. *From multimodal emotional corpora to models of coordination between modalities*

As the exploratory copy-synthesis approach that we have defined revealed to be successful for the study of levels of representation for blended emotions and expressive behaviours, we aim at going further in this direction. We will work on making the copy-synthesis approach more automatic in order to ease the process of going from the annotations to the replay by the Greta expressive agent. We will also consider the representation languages under definition within Humaine for emotions (EARL) and behaviours (Gesticon).

Another strand of research deals with the automatic analysis of acted multimodal data from Geneva's GEMEP corpus. This research aims for a better understanding and finally a model of the finer grained temporal synchronization between expressive gestures and speech. For this purpose models for the automatic segmentation of speech at the phoneme-level are developed, in order to provide exact information on the location of possible synchronization points. On the other hand, computer vision algorithms are adapted for performing the automatic tracking of hands and faces on a per-frame level, and thus providing the base-level information necessary for all subsequent quantitative analysis of speech-gesture synchronization.

- b. *Gesticon*

We aim to refine the Gesticon specification. At the previous Humaine meeting a number of changes and additions to the current representation format were agreed on, which now need to be fleshed out in a more detailed way.

The most important revision is concerned with the introduction of the concept of different levels of specification (LoS) in the Gesture Repository. We will add the definition of different LoS in the gesture repository. Two types of LoS can be distinguished. On the one hand, LoS may concern the level of precision of the data – stretching from mere animation names for motion-captured data, to fine-grained symbolic descriptions of behaviors. For this purpose for each type of channel (e.g. head, gaze, gesture...) and each level of detail a representational ontology will be defined.

On the other hand, LoS reflects the pertinence of different components of a gesture. E.g. studies have shown that not all dimensions of a gesture are relevant to convey its function or meaning. Thus we are looking for ways to represent this type of information in the Gesture Repository. by employing a typology of primary, secondary, and not relevant dimensions. Such descriptions would allow systems to automatically refine and adjust animations while pertaining their core meaning.

We aim to transform the behavior representation scheme we have been using for the Greta agent technology to the Gesticon formalism. To ensure the portability of the language, i.e. to make sure behavior descriptions can be interpreted identically by different agent geometries and animation players, we are collaborating with OFAI (in particular with the RASCALLI project) where a different agent technology is used.

Currently Animation Score, our tool to control an agent from the specification of its behaviors through time (and not from communicative functions), is extended to work with behaviors described within Gesticon.

c. *Context dependent emotional body gesture*

In a reactive motion scenario of virtual characters, we have considered that an individualisation of the characters influences the type of reaction. The individual descriptors (personality, emotion) and the properties of stimuli are inputs to animation controllers that drive the body's movement (like inverse kinematics).

This reaction synthesis has been modeled inside a semantic representation, and implemented to the upper body of a 3D character. First, we aim to extend the animation to the lower body, and second, we aim to profit the semantic representation to implement two different reactive appraisal processes that include emotion, and evaluate their performance in the 3D character. For these, we need to establish the main inputs to the inverse kinematics algorithm to produce the reactive movements.

To test the realism of virtual humans, we want to build an interactive application. In this application users will make one virtual human react at time. Different reactive movement should be provided according to each virtual human's individualisation.

4.5 Element 5: Expressivity

a. *Behaviour expressivity*

- **Analysis of emotional gestures** We are currently working on refining computer-vision algorithms for extracting features about the spatial and temporal properties of gestures and postures displayed in a video corpus of acted emotions (the GEMEP corpus provided by the Geneva Emotion Research Group). Work on the identification and frame-wise tracking of hands and faces is under way. In addition, other more simple and thus more robust methods for automatically acquiring quantitative and qualitative descriptions of motion data from these videos are to be investigated. These measurements are then to be used for aiding re-synthesis experiments but also for building and testing quantitative models for the influence of emotions on the size, shape and velocity of body movements.
- **Synthesis of analyzed emotional gestures:** We will further develop a system that takes in input videos from the GEMEP corpus with real actors performing gestures in different emotional conditions and controls the Greta animation. The system will analyse the actors' motor behaviour, extract expressive indicators and control the Greta animation so that the gesture performed by the agent will communicate to users the same emotional state expressed by the actors. Perceptive tests with users will allow us to evaluate both the analysis algorithms providing expressive cues and the synthesis approach for the generation of gesture expressivity in ECA.

- **Extension of behaviour expressivity to other domains:** We have conducted an analysis of 2D cartoons where we annotated gesture expressivity of the characters and we have investigated the modulations of this expressivity over time, named breaks. We have observed that these breaks can act as a rhetorical relation of similarity on the one hand, and as a rhetorical relation of contrast on the other hand [Ech Chafai et al., 2006]. After studying the function of modulations of gesture expressivity based on the analysis of 2D cartoons, we aim to answer to the question: if the rhetorical relation of similarity and of contrast are intuitively used by 2D animators to suggest some relations or effects, do the primitive nature of these breaks determine these properties, and is this phenomenon observable in other domains?

b. Speech Expressivity

The fruitful cooperation with the PAVOQUE project will be continued. An **expressive database** for a limited domain unit selection voice will be designed and recorded. Signal analysis and modification algorithms will be pre-selected, evaluated and applied in database analysis and subsequent synthesis. One aspect of using an expressive target for selecting units is the use of copy synthesis: an expressive utterance, recorded with the synthesis voice but not included into the runtime data set, is analysed, yielding an **acoustic parametrisation** at a suitable degree of abstraction. This parametrisation is used as target costs in the unit selection algorithm, yielding expressively similar units when available. This method will help shed light on the question what acoustic parametrisation is a “suitable degree of abstraction”.

A **workshop on “Paralinguistic speech: between models and data”** will be organised at DFKI in August 2007. In this workshop, we intend to build bridges between the two separate camps of “modellers” and “recognisers”, which traditionally have remained quite separate. Encouraging these camps to talk to each other is a first step towards a better mutual understanding, which may ultimately enable the two sub-communities to benefit from each others work.

For the next year the implementation of the issues mentioned in section 3.5b is planned. Furthermore we started work with the W3C incubator group on an **emotional markup language**. The support for this markup language with our tools will also be an issue of our work.

Forthcoming works will be directed towards the **refinement of the focus generation model**, according to the Valence-Intensity-Domain model [Aubergé and Rilliard 06]. We will focus on the two main points: accurate copy of prosodic cues on the focused part and neutralisation in the vicinity of the focused part, in order to avoid conflicting accents, one of the main drawbacks inherent to the use of corpus-based synthesis. Signal processing techniques or/and unit selection will be used to reach the prosodic targets given by the model.

Listening tests will be carried out in order to evaluate the model on the three axes: valence (is the focus perceived or not?), intensity (if the focus is perceived, how strongly is it perceived?), and domain (is the focus perceived

on the good linguistic domain?), as well as in terms of naturalness and acceptability. The perceptual exploration of the mapping of voice parameters to affect will continue and be further consolidated. To supplement and evaluate the output of perception experiments [e.g., Yanyshetskaya et al 2006], production data will be analysed for a subset of the same affects.

A fundamental goal is to understand the **prosody** of the voice in expressive speech, i.e. not just pitch variation, but also the correlated source parameters. Towards this goal it will be important to contextualise the expressive dimension within what is currently understood about the dynamics of prosody. Analytic work is also planned here and there will be interaction with a TCD project on Irish Prosody.

The **cross language/cross cultural** investigations of the past three years will also be consolidated, and developed to yield initial principles for differentiating the universal from the culture-specific.

5 References

[**Abrilian et al 05**] Abrilian, S., Devillers, L., Buisine, S. and Martin, J.-C. EmoTV1: Annotation of Real-life Emotions for the Specification of Multimodal Affective Interfaces. 11th Int. Conf. Human-Computer Interaction (HCI'2005), Las Vegas, Nevada, USA, 22 - 27 July Electronic proceedings, LEA, 2005.

[**Abrilian et al 06**] Abrilian, S., Martin, J.-C., Buisine, S. and Devillers, L. Perception of Movement Expressivity in Emotional TV Interviews. Humaine Summer School, Genova, Italy, September 22-28, 2006.

[**Aubergé and Rilliard 06**] Aubergé, V. and Rilliard, A., Le focus prosodique n'est pas que déictique, le modèle VID (Valence-Intensité-Domaine), JEP'2006, Dinard, France, June 2006.

[**Aubergé et al 06a**] Audibert, N., Vincent, D., Aubergé, V. and Rosec, O. Expressive speech synthesis: evaluation of a voice quality centered coder on the different acoustic dimensions, Speech Prosody 2006, Dresden, Germany, May 2006.

[**Aubergé et al 06b**] Audibert, N., Vincent, D., Aubergé, V. and Rosec, O. Evaluation of expressive speech synthesis, LREC'2006 Workshop on Corpora for research on emotion and affect, Genoa, Italy, May 2006.

[**Bevacqua et al 06**] Bevacqua, E., Raouzaïou, A., Peters, C., Cardakis, G., Karpouzis, K., Pelachaud, C. and Mancini, M. Multimodal Sensing, Interpretation and Copying of Movements by a Virtual Agent, Perception and Interaction Technologies, Kloster Irsee, Germany, June 19-21, 2006.

[**Buisine et al 2006**] Buisine, S., Abrilian, S., Niewiadomski, R., Martin, J.-C., Devillers, L., Pelachaud, C. Perception of Blended Emotions: from Video Corpus to Expressive Agent. 6th International Conference on Intelligent Virtual Agents (IVA'2006), 21-23 August 2006, Marina del Rey, USA. LNAI 4133, Springer, pp. 93-106. BEST PAPER AWARD.

[**Burkhardt et al 06**] Burkhardt, F., Audibert, N., Malatesta, L., Türk, O., Arslan, L. and Auberge, V. Emotional Prosody - Does Culture Make A Difference?, Proc. Speech Prosody 2006.

[**Castellano et al 06a**] Castellano, G., Camurri, A., and Scherer, K. Expressive gesture and music: analysis of emotional behaviour in music performances (in preparation).

[**Castellano et al 06b**] Castellano, G., Bresin, R. and Camurri, A., Audio-visual feedback of full-body expressive movements (in preparation).

[**Devillers et al 06a**] Devillers, L., Cowie, R., Martin, J.-C., Douglas-Cowie, E., Abrilian, S. and McRorie, M. Real life emotions in French and English TV video clips: an integrated annotation protocol combining continuous and discrete approaches. 5th international conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy, 24-27 May 2006.

[**Devillers et al 06b**] Devillers, L., Cowie, R., Martin, J.-C., Douglas-Cowie, E., Abrilian, S. and McRorie, M. Real life emotions in French and English TV video clips: an integrated annotation protocol combining continuous and discrete approaches. 5th international

conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy, 24-27 May 2006.

[**Garcia-Rojas et al 06**] Garcia-Rojas, A., Vexo, F. and Thalmann, D. Individualised Reaction Movements for Virtual Humans, in Proc. Graphite conference 2006. (to appear).

[**Gobl et al 06**] Gobl, C. Yanushevskaya, I., and Ní Chasaide, A. (2006). Cross-language perception of voice and affect. *Journal of the Acoustical Society of America*, 120, 3290.

[**Mancini et al 06**] Mancini, M., Castellano, G., Bevacqua, E. and Peters, C., Copying behaviour of expressive motion, (submitted, Mirage 2007).

[**Martin et al 05**] Martin, J.-C., Abrilian, S. and Devillers, L. (2005). Annotating Multimodal Behaviours Occurring during Non Basic Emotions. 1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005), Beijing, China, October 22-24 Springer-Verlag Berlin.550-557.

[**Martin et al 06**] Martin, J.-C., Caridakis, G., Devillers, L., Karpouzis, K. and Abrilian, S. Manual Annotation and Automatic Image Processing of Multimodal Emotional Behaviours: Validating the Annotation of TV Interviews. Fifth international conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy, 24-26 May.

[**Martin et al 06b**] Martin, J.-C., Niewiadomski, R., Devillers, L., Buisine, S., Pelachaud, C. Multimodal Complex Emotions: Gesture Expressivity And Blended Facial Expressions. Special issue of the Journal of Humanoid Robotics. Eds: C. Pelachaud, L. Canamero. Vol. 3, No. 3, September 2006, 269-291.

[**Moccozet et al 06**] Moccozet, L., Garcia-Rojas, A., Vexo, F., Thalmann, D. and Magnenat-Thalmann, N., In Search for your own Virtual Individual. In Semantics And Digital Media Technology Conference 2006 (to appear).

[**Ní Chasaide et al 06**] Ní Chasaide, A., Wogan, J., Ó Raghallaigh, B., Ní Bhriain, Á., Zoerner, E., Berthelsen, H. and Gobl, C. (2006). Speech Technology for Minority Languages: the Case of Irish (Gaelic). *Proceedings of the 9th International Conference on Spoken Language Processing, INTERSPEECH 2006*, Pittsburgh.

[**Peters 06a**] Peters, C. A Perceptually-Based Theory of Mind Model for Agent Interaction Initiation, Special Issue of the Journal of Humanoid Robotics, Special Edition "Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids", Eds: C. Pelachaud and L. Canamero, 3(3), September 2006.

[**Peters 06b**] Peters, C. Evaluating Perception of Interaction Initiation in Virtual Environments using Humanoid Agents (2006). In Proceedings of the 17th European Conference on Artificial Intelligence (ECAI), pp. 46–50, Riva Del Garda, Italy August 2006.

[**Peters 06c**] Peters, C. Designing Synthetic Memory Systems for Supporting Autonomous Embodied Agent Behaviour, Proceedings of the 15th International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 14–19, September 2006.

[**Peters 06d**] Peters, C. Bottom-Up Visual Attention on the GPU (in preparation).

[**Peters et al 06**] Peters, C., Pelachaud, C., Bevacqua, E., Ochs, M., Chafai, N. and Mancini M. Social Capabilities for Autonomous Virtual Characters, M. Proceedings of iDiG,

International Digital Games Conference, Games Congress 2006, pp. 37-48, Portalegre, Portugal, September, 2006.

[Pfleger and Löckelt 06] Pfleger, N. and Löckelt, M. A Comprehensive Context Model for Multi-party Interactions with Virtual Characters. Proceedings of the 6th International Conference on Intelligent Virtual Agents (IVA 2006), pp. 157 – 168, 2006.

[Schröder 06] Schröder, M. Expressing degree of activation in synthetic speech. *IEEE Transactions on Audio, Speech and Language Processing*, 14 (4), pp. 1128-1136, 2006.

[Schröder et al 06a] Schröder, M., Hunecke, A. and Krstulovic, S. OpenMary - open source unit selection as the basis for research on expressive synthesis, *Blizzard Challenge Workshop 2006*, Pittsburgh, PA, USA, 2006.

[Schröder et al 06b] Schröder, M., Heylen, D. and Poggi, I. Perception of non-verbal emotional listener feedback. *Proc. Speech Prosody 2006*, Dresden, Germany, 2006.

[Vincent et al 05] Vincent, D., Rosec, O. and Chonavel, T. Estimation of LF Glottal Source Parameters Based on an ARX Model, Interspeech'2005, pp. 333-336, Lisbon, Portugal, September 2005.

[Yanushevskaya et al 06] Yanushevskaya, I., Gobl, C. and Ní Chasaide, A. (2006). Mapping Voice to Affect: Japanese listeners. *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden, Germany.