Non-verbal behaviors expressivity and their representation

PF-star report 3

Maurizio Mancini
LINC-Paragrape
IUT of Montreuil
University Paris VIII
m.mancini@iut-univ.paris8.fr

Bjoern Hartmann
LINC-Paragrape
IUT of Montreuil
University Paris VIII
b.hartmann@iut-univ.paris8.fr

Catherine Pelachaud
LINC-Paragrape
IUT of Montreuil
University Paris VIII
c.pelachaud@iut-univ.paris8.fr

1. Introduction

Embodied Conversational Agents (ECAs) are a powerful user interface paradigm, aiming to transfer the inherent richness of human-human interaction to human-computer interaction. ECAs are virtual embodied representations of humans that communicate multimodally with the user (or other agents) through voice, facial expression, gaze, gesture, and body movement. Effectiveness of an agent is dependent on her ability to suspend the user's disbelief during an interaction. Our work aims at creating one ECA system called Greta, that would be able to interact with the user not only communicating information in a general deterministic way but exhibiting a consistent behavior with her personality and with contextual environment factors. In fact human individuals differ in their reasoning, their set of beliefs, goals, their emotive states, and also their way of expressing such information through the execution of specific behaviors. The behavior of one person depends not only on factors defining her individuality (such as her culture, her social and professional role, her personality and her experience), but also on a set of contextual (such as her interlocutor, the social conversation setting), and dynamic variables (belief, goal, emotion). We call these types of factors 'influences' in the sense that they affect the behaviors to be displayed. Influences may act at different levels: they may act on what to say and when as well on how to say it and to express it. Until now most ECA systems have concentrated solely on defining computational models of behavior selection – *which* behaviors to choose for a given communicative act. To increase believability and life-likeness of an agent, she has to express emotion and exhibit personality in a consistent manner [24]. In this document we will first briefly describe what happens in real life, how people differ in their behavior depending on different personality and situation (section 2). To do that we will introduce our taxonomy of behavioral influences in section 3. A general description of our system's architecture will be presented in 4. After that we will provide an explanation of the two main components of the system, the *Expressivity Specification Module* and the *Animation Engine*.

2. Human Expressivity

Prior to investigation by ECA researchers, the problem of characterizing facial displays and bodily movement has been addressed mainly within the social psychology, performance studies, and biomechanics communities. While we still need to conduct a more detailed survey of biomechanics, some studies in the remaining two areas will be summarized to guide our search for a consensual definition of expressivity.

2.1. Facial Expressivity

Research in face perception is a vast and complex area of study. Some work concentrates on recognizing individual identity [39], facial features, lip movement [26, 34], facial expression of emotion [14]. Other studies try to recognize a face from a background, or to understand the process of recognizing familiar faces, of putting a name to a face [39]. Other important branches of studies look at how emotions are perceived – what the roles of conversational signals and backchannels are [13].

Quite some research has been done on expression and emotion perception. Most of the experiments have the same set up [14]: a set of photographs of posed expression of emotions (or spontaneous expressions as is mostly the case nowadays) is shown to subjects. The subjects are asked to recognize the expressions. The subjects show good results specially in recognizing and distinguishing facial expressions not involving the same facial movements [13]. Ekman and Friesen have found differences in the display of felt and fake facial expressions. Fake emotion displays will be more asymmetric and their timing properties vary from felt expressions[16]. Expression recognition requires an abstraction of the particular face displaying the expression to view only the model-independent facial features.

In order to recognize the subtle changes of facial expressions, several researchers propose to recognize minimal facial signals and combine the signals to recognize the complete facial expressions [11, 7]. That is, rather than trying to recognize the entire facial expression they are working on recognizing singular facial actions. The facial expression is then deduced by combining the several facial actions. This recognition method is based on the Facial Action Coding System, FACS [15], a framework to measure facial signals using minimal action units (AUs).

2.2. Bodily Expressivity

Our approach to bodily expressivity is driven by a perceptual standpoint – how expressivity is perceived by others; not what internal muscle activation patterns underly these signals. Researchers in social psychology have investigated how various influences affect perceived bodily behaviors, mostly through ad hoc measuring instruments constructed by narrowing down an extensive list of choices through coder reliability testing.

Wallbot and Scherer [38] had judges encode their impressions of behavior along the following five categories: slow/fast, small/expansive, weak/energetic, small movement activity/large movement activity, and unpleasant/pleasant. In a later study, Wallbot [37] found that besides static pose configurations, three dynamic dimensions could be reliably identified by oberservers: amount of movement activity, expansiveness/spatial extension, and movement dynamics/energy/power. Gallaher [17] found four significant dimensions of variability in personal encoding style: expressiveness - energetic communication; animation - energy in acts not directly related to communication; expansiveness - use of space, elbow position; and coordination - smoothness, fluidity. Ball and Breese [1] outline Collier's finding on correlations between temporal and spatial tendencies in gesture/posture and personality/emotion – movement frequency and speed were related to emotional arousal, as was the size of overall body outline. A summary of Pollick's research [32] points out limits of dissecting movement features and ascribing discrete values to them: the degree and manner in which this style is dependent on spatial and temporal encoding is not trivial and varies between different movements. The importance of considering human posture as a communication channel has been stressed as well [28, 20] but we do not yet integrate these findings in our own work.

3. Taxonomy of influences

Most of the agents that have been created so far are very generic in their behavior type. Several studies have shown the importance to consider complex information such as cultural factors [23], personality [21], environment setting when designing an agent [19]. These factors affect the interaction a user may have with the agent. User's trust, reliability on what the agent says, persuasion by the agent and so on, are correlated with the agent's tendency to be similar to the user and to show the same referential background. Studies have shown that the way people present information, describe events, talk about feelings are culturally dependent [3]. Of course this is not the only factor of differentiation. Personality also is an important aspect that makes people look and act very differently. Gender, age (child vs teenager), our social role (e.g., mother, doctor), our experience and memory intervene in the manner we interact with others, we talk about things, etc. The point we are trying to make here is that, given a goal in mind (say, talk about illness to a doctor) people differ in their manner of expressing themselves. These differences arise at different levels: the formulation of their thought as well as their expression [4]. Mapping such a reasoning to the agent world, we can say that all the factors that should be considered to create an individual agent, act at different levels of the creation. They have influences on the surface generation and realization level during the dialog generation phase as well as on the selection of the nonverbal behaviors and their expressivity [10].

In the next subsections we present a taxonomy of the different types of influences.

3.1. Intrinsic Influences

At first, we consider that each human being has a set of the conscious and unconscious habits that intervene in the content of her discourse and that define her attitude and her behavior when she talks. These habits, we call the intrinsic factors, derive,

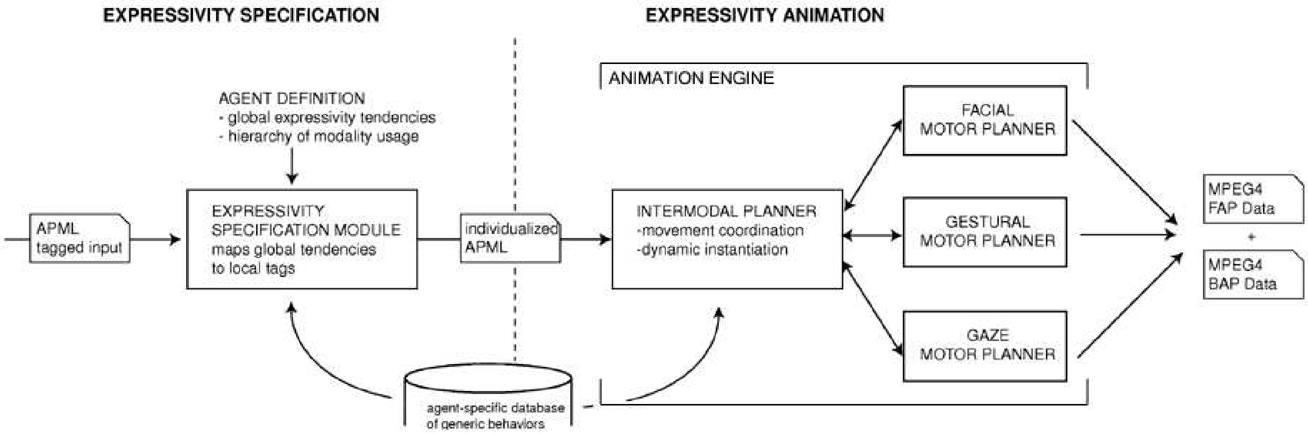

Figure 1. System overview.

amongst others, from her personality, her age, her sex, her nationality, her culture, her education and her experiences [10]. Some nonverbal behaviors are very culturally dependent. This is the case of emblems, gestures that may be directly translated into words. This might be the case with some iconic or metaphoric gestures that may take their origins in the culture (the action of eating will not be represented identically if one is used to eat with a fork or with sticks). On the other hand, not all gestures are culturally dependent. Justine Cassell found that the type of gestures one makes while conversing might not differ over various cultures as much as one would have thought at first. The main discrepancy is more in the quantity of gestures made rather than on the type of gestures itself [5].

3.2. Contextual Influences

Several factors, depending on the context, may increase or decrease the effects of the intrinsic factors. The contextual factors may refer to the environment setting such as the light conditions, the sound intensity, the spatial layout or the function of the conversation site. These factor may affect some speech and behavior characteristics. For instance, in a crowed room, some gestures are simply impossible. Likewise, in a noisy room, a person has to increase the volume of her voice due to pragmatic considerations. She will also be inclined to amplify her gestures. In the opposite, in a religious building or in a museum, a person ought to remain quiet and move silently due to social considerations. Another factor of influences is how the agent is placed in the environment; a person sitting in an armchair will not gesticulate as a person standing in a hallway. These contextual influences, unlike the intrinsic ones, may vary during a dialog session.

3.3. Dynamic Influences

The mental state of the agent affects greatly the way the agent will behave. It modifies the prosody of speech, the amplitude of a facial expression, the movement tempo. For instance, a person does not talk and does not behave in the same way whether she is angry or not. Her relationships with her interlocutor modulates also her behavior: she does not behave in the same way with a friend, an unknown person, an employee, a child or a doctor [8]. The agent's mental state evolves all along the conversation. Her emotion varies through time, her goals and beliefs get modified as the conversation evolved. The agent's mental state if extremely dynamic.

4. System overview

The complete Greta ECA system is shown in Figure 1. Our multimodal agent (see Fig. 2), interprets a text marked up with a language called APML [9] to generate synchronized speech, face, gaze and gesture animations. APML allows to specify communicative functions of the text to be uttered by the agent. Text-to-speech conversion is accomplished through Festival [2]. The timing output from speech synthesis is used to set coincidence constraints for separate engines modeling face/gaze and gesture animation. Both engines produce animation data in MPEG4-compliant FAP/BAP format [35], which in turn drive a facial and skeletal body model in OpenGL. Inside the system an *individualized* version of APML is used to communicate between the two main half of the system. The first part takes the APML tagged text and it adds some extra tags and attributes that holds the information given by the various influences we described in the last section. So the enriched version of APML can be considered as an individualized APML and it will be explained in section 5.4.

Figure 2. Greta.

4.1. Expressivity attributes

Based on the aggregate evidence presented in the studies above, we propose to capture expressivity with a set of six attributes which we describe below in qualitative terms. As part of an indivdualized agent's defintion, personal default values for the expressivity attributes are defined. These values can be supplanted by including expressivity information in communicative function tags within our markup language.

- *Overall activation*: amount of activity (quantity of movement) across several modalities during a conversational turn (e.g., simultaneous use of facial expression and gesture to visualize communicative acts – passive/static or animated/engaged).
- *Spatial extent*: amplitude of movements (e.g., amount of space taken up by body; amplitude of eyebrow raise)
- *Temporal*: duration of movements (e.g., quick versus sustained actions)
- *Fluidity*: smoothness and continuity of overall movement (e.g., smooth,graceful versus sudden,jerky)
- *Power/Energy*: dynamic properties of the movement (e.g., weak/relaxed versus strong/tense)
- *Repetitivity*: tendency to rhythmic repeats of specific movements along specific modalities.

Overall activation is singly float-valued and ranges from 0 to 1. Each of the other attributes is also float-valued and defined over the interval $[-1, 1]$, where the zero point corresponds to the actions our generic agent without expressivity control would perform. Overall activation, Fluidity and Power act on the entire agent animation calculated for a conversational turn, while the other parameters generate only local changes specific to one communicative act. Though not represented in our literature review, the concept of repetitivity was added from its appearance in an annotated gesture corpus [25]. Wachsmuth [36] has previously identified the importance of rhythm in nonverbal behavior but has only used it to guide gesture recognition, not synthesis.

5. Expressivity Specification Module

As we explained before, our system takes as input a text file which has been previously tagged using an XML based language called APML. As shown in Figure 1, the Expressivity Specification Module is responsible of the ”individualization” of the APML input text, that is the introduction of the information about intrinsic, contextual and dynamic influences (see section 3) into APML tags to allow the Animation Engine to properly generate face and gestural animation in a way consistent with that influences. The module works as follows. It uses a database that contains a large list of (meaning, signal) pairs. The system takes as input:

1. a value representing the intrinsic and contextual influences
2. a text to be said by the agent annotated with Affective Presentation Markup Language (APML) tags
3. meaning expressivity: local expressivity value for each meaning tag if necessary
4. agent's definition:
   (a) the value of the agent's global expressivity attributes (see section 4.1)
   (b) the agent's preference of modalities hierarchy

The purpose of APML is to specify the agent's behavior at the meaning level. The type of the tags represents the communicative functions [31]. In the next section we explain what we mean by these terms. At first the module selects a set of possible nonverbal behaviors. That is using the database of all possible nonverbal behaviors, it looks at which signals may be chosen based on the agent's definition parameters.

5.1. Meaning Expressivity

The *meaning expressivity* specifies the strength of the related communicative acts. It may be used to express that a given tag should receive a particular attention. An agent may emphasize more her speech than other communicative function; or she can show more her emotion. This expressivity value allows us to provide such a modulation.

5.2. Agent's definition

An agent is characterized by two sets of parameters; one specifies her overall expressivity while the second one represents her preference to display signals on more modality or on another.

5.3. Modalities hierarchy

An agent may express her communicative acts by privileging one modality over the others. This choice may depend on the agent's tendency to use one modality over another to communicate. She may use mainly her face or her gesture. To model such a characteristic, we define a hierarchy of the modalities by associating to each of them a numeric value that reflects the preferential level an agent has to use this modality. During the instantiation of the tags into a set of signals, the system considers the modalities that best fit the hierarchy of behaviors.

5.4. Individualized APML

Given a text to be said that is marked with APML tags and that have been augmented with information on the expressivity, the module have to instantiate each meaning tags into a set of signals with their appropriate expressivity, and code it into a new enriched individualized APML text. As the first step the expression library is accessed, where each description of signals contains a range of expressivity value (i.e. a minimal and a maximal value) and a reference value. These values indicate the domain where their use is the most appropriated. The *expressionValue* associated to the tag is compared to these attributes. For the facial expressions, it is compared to the intensity attribute. From this information is possible to choose different set of signals for a given meaning depending on its expressivity value. E.g., for the emotion ‘joy’ we can define the signals that represent ‘discreet joy’ and the signals for ‘intense joy’; both sets of signals may be defined independently of each other.

If *expressionValue* is not included between the minimal and the maximal values of a signal, it is associated to the nearest bound. Conversely, if it belongs to several ranges, the signal is chosen according to the distance between the desired value and its reference expressivity.

Then the module has to decide the modality (face, gesture, gaze) in which it will search the signal. In the order implied by the modalities hierarchy it verifies that the modality contains at least a signal which can express the tag meaning, and that the use of this modality do not provoke conflicts [30], i.e. that the modality is not yet used to express another meaning.

6. Animation Engine

We will now describe the second module of our system, which is the right half of the diagram shown in Figure 1. Given the individualized APML document calculated by the Expressivity Specification Module, how can we modify non-verbal (facial and gestural) behavior production to communicate the appropriate expressive content? We will suggest some computational models for treating expressivity in the following sections. In the list of expressivity attributes we have elaborated and presented in section 4.1, only Overall activation acts uniformly on the generation of behavior on a global utterance level. It will therefore be introduced first. Then, for the two modalities considered – face and gesture – we will mention the different available animation parameters and go through the list of dimensions of expressivity and describe possible mappings to animation parameters.

6.1. Intermodal filtering: Overall activation

A filtering is applied by the intermodal planner (see Fig. 1). Each tag carries a summary weight attribute that expresses how important stressing the tag’s content through nonverbal signals is. Communicative functions tags for which this activation attribute does not surpass a given agent’s overall activation threshold are not matched against the behavior database and thus no nonverbal behavior is generated at all. A similar principle of activity filtering was presented and implemented by Cassell et al. in [6].

6.2. Face

A facial expression is not only identified by a configuration of facial muscles but it is also important how this configuration will be temporally activated.
One expression (in the sense of ”muscular configuration”) can assume different expressivity depending on the manner it appears (onset), the time it remains on the face (called the *apex* time) and finally the speed it disappears (offset). We have chosen this specification to represent the temporal characteristics of facial expressions [12].
The APML language that is taken as input to the Greta agent system defines the meaning of a given communicative act [29]; the Greta Animation Engine looks up the meaning in a library of expressions to instantiate the corresponding facial expression. Aiming at adding life characteristics to the agent, we have realized some modifications to the markup language in order to allow it to communicate using a wider set of facial expressions by modifying their temporal characteristics dynamically. Let’s now see how expressivity attributes specified by the individualized APML influence face animation generation.

6.2.1. Mapping expressivity to facial animation

- *Spatial extent*: this parameter determines the quantity of physical contraction of the facial muscles involved in the expression showing process.
- *Temporal*: the temporal parameter is connected to the temporal characteristics of the facial expression calculated by the engine. In general, a value of 0 for temporal parameter means that there should not be any kind of temporal modification to the ”natural” way in which the expressions are performed by the agent; positive values will ”stretch” the duration of the facial expression; negative values will ”short” them.
- *Fluidity*: our idea is that this parameter specifies the overall muscle contraction of the face. As a consequence of that, as the movement gets more abrupt, there would be an increase of the muscles’ speed of contraction. Positive values of fluidity will produce smooth facial movement while negative values will produce more abrupt movement, making expressions appear and disappear very quickly on the agent’s face.

- *Power*: this parameter will affect the physical muscles' contraction as the spatial extent parameter does, but it will influence the intensity of all the agent's turn of conversation, instead of a single expression. The power parameter also affects lip shape. We have defined a qualifier at the animation level to control lip muscle tension. The parameter, called *Tension Degree*, can be set to *Strong*, *Normal* or *Light*. It allows one to choose between different intensities of muscular strain. Such a tension may appear for the expressions of emotions like fear and anger. Such a tension can also appear, for example, when a bilabial consonant (as /b/) is uttered and lips compress against each other, or when labial width increases and lips stretch, getting thinner.
- *Repetitivity*: this parameter aims to express how often a facial expression is repeated. An higher value represents a high repetitivity, for example increasing the number of head nods during the agent's turn of listening or the number of times it raises its eyebrows while performing one expression.

6.3. Gestures

Based on an analysis of the communicative functions contained in the marked up input text, the gesture system chooses a matching prototype gesture to be executed. Subsequently, we dynamically instantiate the gesture according to a set of expressivity attributes and synchrony constraints with speech. To do so, we need a suitable representation of gestures, which is outlined first. Next, different ideas for adaptation and instantiation strategies are described. It is our primary aim to include and preserve the semantic value of each gesture during the computation of expressivity. The synthesis methods are currently being tested in implementation and are subject to change. Notably though, it is already apparent that effective strategies will have to adjust behavior on multiple levels - from abstract planning (whether to search for a gesture for a given text at all), to gesture phase-level modifications (whether or not to repeat a stroke), down to adjusting velocity profiles of key pose transitions. As we have done for facial animation (see section 6.2) we now explain how the expressivity attributes specified by the individualized APML influence gesture animation generation.

6.3.1. Mapping expressivity to gestural animation

- *Spatial extent*: spatial extent has been modelled by expanding (assigning higher values to spatial extent) or condensing (assigning lower values) the entire space in front of the agent that is used for gesturing. This gesture space is represented as a set of sectors following McNeill's diagram [27].
- *Temporal*: starting from the synchronicity constraint on the end of the gesture stroke to coincide with the stressed affiliate in speech (cf. [27]), we can calculate preceding and proceeding gesture phases timings from a invariant laws of human arm movement described in [18] and a high-level notion of how quickly the gesture phases should be performed.
- *Fluidity*: this concept seeks to capture the continuity between movements, as such, it seems appropriate to modify the continuity of the arms' trajectory paths as well as the acceleration and deceleration of the limbs. For gesture animation generation we use a kinematic TCB splines [22] defined for the entire animation [33]. To achieve different fluidity we can then adjust tension, continuity and bias parameters for the TCB splines.
- *Power*: to visualize the amount of energy and tension invested into a movement, we once again look at the dynamic properties of gestures and also at inter-gestural rest phases. Powerful movements are expected to have higher acceleration and deceleration magnitudes. Again, this behavior could be modeled with the tension parameter of the kinematic TCB-spline.
- *Repetitivity*: this attribute controls the tendency to rhythmically repeat a certain gesture along with the structure of the speech. A given hand-arm configuration is held after the initial gesture execution and overlaid with beat gestures to follow prosody of speech.

7. Future work

In this paper we have presented a system that takes into account influences coming from different parameters to compute the behavior of an agent. The aim of our research is to define tools that allow the creation of an individual agent. To overcome the difficulties of modelling which influences complex entities such as culture, personality, gender have on the determination of the agent's behavior, we have developed a computational model of influences. Our system models how a given influence will affect the agent's behavior. The system will determine not only the appropriate signals but also their expressivity. We have defined what means expressivity for each modality. We are currently in the experimentation phase and the values associated to the signals and to the agent are empirical. We aim at defining different strategies to combine signals (using redundancy, complementarity, substitution, and so on). We also have to define more accurately the expressivity parameters for gestures (amplitude, movement repetition, movement tempo...). An evaluation phase to evaluate if the change of expressivity is perceived still has to be done.

8. Acknowledgment

We are very thankful to Myriam Lamolle, Vincent Maya and Christopher Peters for their help.

References

[1] G. Ball and J. Breese. Emotion and personality in a conversational agent. In S. Prevost J. Cassell, J. Sullivan and E. Churchill, editors, *Embodied Conversational Characters*. MITpress, Cambridge, MA, 2000.

[2] A.W. Black, P. Taylor, R. Caley, and R. Clark. Festival. http://www.cstr.ed.ac.uk/projects/festival/.

[3] R Brislin. *Understanding culture's influence on behavior*. Hartcout Brace College Publishers, New-York, 1993.

[4] C.M. Brown and G. Nichols-English. Dealing with patient diversity in pharmacy practice. *Drug Topics*, pages 61–69, September 1999.

[5] J. Cassell. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In S. Prevost J. Cassell, J. Sullivan and E. Churchill, editors, *Embodied Conversational Characters*. MITpress, Cambridge, MA, 2000.

[6] Justine Cassell, Hannes Högni Vilhjálmsson, and Timothy Bickmore. Beat: the behavior expression animation toolkit. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 477–486. ACM Press, 2001.

[7] J.F. Cohn, J.J Lien, T. Kanade, W. Hua, and A. Zlochower. Behond prototypic expressions: Discriminating subtle changes in the face. In *Proceedings of the 7th IEEE Robot and Human Communication Workshop*, Takamatsu, Japan, September 30 - October 2 1998.

[8] B. DeCarolis, V. Carofiglio, M. Bilvi, and C. Pelachaud. APML, a mark-up language for believable behavior generation. In *Embodied conversational agents - let's specify and evaluate them!, Proceedings of the AAMAS'02 workshop*, Bologna, Italy, July 2002.

[9] Berardina DeCarolis, Catherine Pelachaud, Isabella Poggi, and Mark Steedman. APML, a mark-up language for believable behavior generation. In Helmut Prendinger and Mitsuru Ishizuka, editors, *Life-Like Characters*, Cognitive Technologies, pages –. Springer, 2004.

[10] F. deRosis, C. Pelachaud, I. Poggi, V. Carofiglio, and B. De Carolis. From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies. Special Issue on Applications of Affective Computing in HCI*, in press.

[11] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, submitted.

[12] P. Ekman. About brows: Emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog, editors, *Human ethology: Claims and limits of a new discipline: contributions to the Colloquium*, pages 169–248. Cambridge University Press, Cambridge, England; New-York, 1979.

[13] P. Ekman. *Emotion in the human face*. Cambridge University Press, 1982.

[14] P. Ekman and W. Friesen. *Unmasking the Face: A guide to recognizing emotions from facial clues*. Prentice-Hall, Inc., 1975.

[15] P. Ekman and W. Friesen. *Facial Action Coding System*. Consulting Psychologists Press, Inc., Palo Alto, CA, 1978.

[16] P. Ekman and W. Friesen. Felt, false, miserable smiles. *Journal of Nonverbal Behavior*, 6(4):238–251, 1982.

[17] Peggy E. Gallaher. Individual differences in nonverbal behavior: Dimensions of style. *Journal of Personality and Social Psychology*, 63(1):133–145, 1992.

[18] Sylvie Gibet, Jean-Francois Kamp, and Franck Poirier. Gesture analysis: Invariant laws in movement. In Antonio Camurri and Gualtiero Volpe, editors, *Gesture-Based Communication in Human-Computer Interaction - GW 2003*, number 2915 in LNAI, pages 1–9. Springer, 2004.

[19] H. McBreen H. Morton and M. Jack. Experimental evaluation of the use of ECAs in eCommerce applications. In Z. Ruttkay and C. Pelachaud, editors, *From Brows till Trust: Evaluating Embodied Conversational Agents*. Kluwer, To appear.

[20] Jinni A. Harrigan. Listener's body movements and speaking turns. *Communication Research*, 12(2):233–250, April 1985.

[21] K. Isbister and C. Nass. Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*, Forthcoming.

[22] Doris H. U. Kochanek and Richard H. Bartels. Interpolating splines with local tension, continuity, and bias control. In Hank Christiansen, editor, *Computer Graphics (SIGGRAPH '84 Proceedings)*, volume 18, pages 33–41, 1984.

[23] E.-J. Lee and C. Nass. Does the ethnicity of a computer agent matter? An experimental comparison of human-computer interaction and computer-mediated communication. In *WECC'98, The First Workshop on Embodied Conversational Characters*, October 1998.

[24] A. Bryan Loyall and Joseph Bates. Personality-rich believable agents that use language. In W. Lewis Johnson and Barbara Hayes-Roth, editors, *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, pages 106–113, Marina del Rey, CA, USA, 1997. ACM Press.

[25] Craig Martell, Paul Howard, Chris Osborn, Lisa Britt, and Kari Myers. Form2 kinematic gesture corpus. video recording and annotation, 2003.

[26] D.W. Massaro and M.M. Cohen. Perception of synthesized audible and visible speech. *Psychological Science*, 1(1):55–63, January 1990.

[27] David McNeill. *Hand and Mind - What gestures reveal about thought*. The University of Chicago Press, Chicago, IL, 1992.

[28] A. Mehrabian. Significance of posture and position in the communication of attitude and status relationships. *Psychological Bulletin*, 71(5), 1969.

[29] C. Pelachaud, V. Carofiglio, B. De Carolis, and F. de Rosis. Embodied contextual agent in information delivering application. In *First International Joint Conference on Autonomous Agents & Multi-Agent Systems (AAMAS)*, Bologna, Italy, July 2002.

[30] C. Pelachaud and I. Poggi. Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation*, 13:301–312, 2002.

[31] I. Poggi, C. Pelachaud, and F. de Rosis. Eye communication in a conversational 3D synthetic agent. *AI Communications*, 13(3):169–181, 2000.

[32] Frank E. Pollick. The features people use to recognize human movement style. In Antonio Camurri and Gualtiero Volpe, editors, *Gesture-Based Communication in Human-Computer Interaction - GW 2003*, number 2915 in LNAI, pages 10–19. Springer, 2004.

[33] Scott N. Steketee and Norman I. Badler. Parametric keyframe interpolation incorporating kinetic adjustment and phrasing control. In *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, pages 255–262. ACM Press, 1985.

[34] Q. Summerfield. Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London*, 335(1273):71–78, 1992.

[35] G. Taubin. SNHC verification model 7.0 - needs to be updated. Technical report, MPEG-4, 1998.

[36] Ipke Wachsmuth. Communicative rhythm in gesture and speech. In A. Braffort et al., editor, *Gesture-based Communication in Human- Computer Interaction - Proceedings International Gesture Workshop GW'99*, volume LNAI 1739, pages 277–289. Springer, Berlin, 1999.

[37] Harald G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28:879–896, 1998.

[38] Harald G. Wallbott and Klaus R. Scherer. Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, 51(4):690–699, 1986.

[39] A.W. Young and V. Bruce. Perceptual categories and the computation of "grandmother". In V. Bruce, editor, *Face Recognition*, volume 1 of *3*, pages 5–49. Lawrence Erlbaum Associates, a special issue of the european journal of cognitive psychology edition, 1991.