Implementing Expressive Gesture Synthesis for Embodied Conversational Agents

Björn Hartmann[1], Maurizio Mancini[2], and Catherine Pelachaud[2]

[1] Stanford University Computer Science Department, Stanford CA 94305, USA
`bjoern @cs.stanford.edu`
[2] LINC-LIA, University of Paris-8, 93100 Montreuil, FRANCE
`(m.mancini|c.pelachaud) @iut.univ-paris8.fr`

Abstract. We aim at creating an expressive Embodied Conversational Agent (ECA) and address the problem of synthesizing expressive agent gestures. In our previous work, we have described the gesture selection process. In this paper, we present a computational model of gesture quality. Once a certain gesture has been chosen for execution, how can we modify it to carry a desired expressive content while retaining its original semantics? We characterize bodily expressivity with a small set of dimensions derived from a review of psychology literature. We provide a detailed description of the implementation of these dimensions in our animation system, including our gesture modeling language. We also demonstrate animations with different expressivity settings in our existing ECA system. Finally, we describe two user studies we undertook to evaluate the appropriateness of our implementation for each dimension of expressivity as well as the potential of combining these dimensions to create expressive gestures that reflect communicative intent.

1 INTRODUCTION

Embodied Conversational Agents (ECAs) are virtual embodied representations of humans that communicate multimodally with the user 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. To increase believability and life-likeness of an agent, she has to express emotion and exhibit personality in a consistent manner [1]. Human individuals differ not only in their reasoning, their set of beliefs, goals, and their emotional states, but also in their way of expressing such information through the execution of specific behaviors. During conversation, expressivity may manifest itself through gesture selection – *which* types of gestures are displayed – as well as through manner of execution – *how* they are displayed. In this paper we present an augmentation to our GRETA agent architecture that allows for parametric control of the qualitative aspects of gesture execution. Since high-level agent functions such as emotion, personality, culture, role and gender may modify actions in complex and competing ways, and since the nature of these influences is not well understood, we restrict our attention to generating phenomenologically accurate behaviors without claiming to correctly represent internal processes (cf. Nass et al. [2]). The paper is structured as follows: related work is reviewed in section 2, and our method for parameterizing gesture expressivity is reviewed in section 3. After outlining the GRETA architecture in section 4, we devote the majority of the paper to a description of the implementation of the expressivity parameters in section 5. We conclude by describing the results of two evaluation studies of our system and pointers to future work in sections 6 and 7.

2 RELATED WORK

Research in gesture synthesis can be divided into systems that address the problem of gesture selection and systems that address the problem of gesture animation. Gesture selection for agents has mostly been concerned with semantic aspects of human gesturing, often following McNeill's method of classification [3]. Cassell et al. select suitable non-verbal behaviors to accompany user-supplied text based on a linguistic analysis [4]. Tepper et al. cross the boundary towards gesture animation by automatically generating iconic gestures from a parametric model [5]. Noot and Ruttkay address the need for inter-subject variability in GESTYLE [6], which chooses between atomic behaviors based on 'style dictionaries.'

Gesture animation is concerned with realistic movement generation of an agent's arms and hands from an abstract gesture representation language [7, 8]. Often, inverse kinematics techniques are used to calculate wrist trajectories [9]. Other systems allow for modification of existing body animations [10]. Of these, EMOTE by Chi et al. [11] is most closely related to our work as it also introduces an intermediate level of parametrization to obtain expressive gestures. EMOTE implements Laban principles from the dance community, while our system relies on psychology literature to obtain a set of expressivity parameters. EMOTE acts as a generic filter on pre-existing behaviors, while we tie behavior modification into the synthesis stage of gesturing. A more comprehensive comparison between the two systems can be found in [12].

3 EXPRESSIVITY PARAMETERS

We conducted a literature review of social psychology to arrive at a dimensional characterization of expressivity in human bodily movement. A summary of this review has been published in [12]. We regard an intermediate level of behavior parametrization as a useful enabling tool to facilitate the mapping of holistic, qualitative communicative functions such as mood, personality, and emotion to low-level animation parameters like joint angles. Our approach is driven by a perceptual standpoint – how expressivity is perceived by others. That is, we focus only on the surface realizations of movement and do not attempt to model underlying muscle activation patterns.

Based on an aggregation of the most pertinent studies [13–15] and our analysis of a gesture corpus [16], we propose to capture gesture expressivity with a set of six attributes which we describe below in qualitative terms. As part of an individualized agent's definition, personal default values for the expressivity attributes are defined.

- *Overall Activation*: quantity of movement during a conversational turn (e.g., passive/static or animated/engaged).
- *Spatial Extent*: amplitude of movements (amount of space taken up by body)
- *Temporal Extent*: duration of movements (e.g., quick vs sustained actions)
- *Fluidity*: smoothness and continuity of overall movement (e.g., smooth vs jerky)
- *Power*: dynamic properties of the movement (e.g., weak vs strong)
- *Repetition*: tendency to rhythmic repeats of specific movements.

Each of the attributes is 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* is float-valued and ranges from 0 to 1, where 0 corresponds to a complete absence of nonverbal behavior.

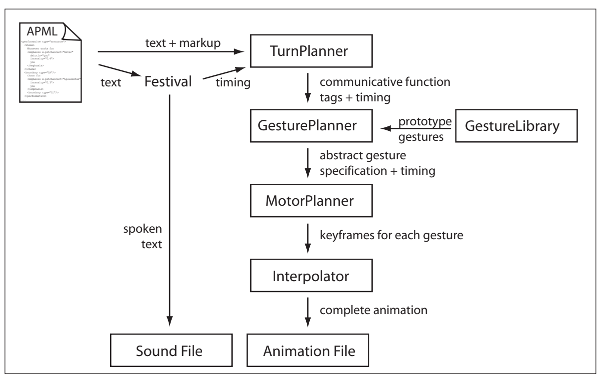

Fig. 1. Agent architecture outline.

4 EXPRESSIVE AGENT ARCHITECTURE

GRETA, our multimodal agent, interprets utterance text marked up in APML with communicative functions [17] to generate synchronized speech, face, gaze and gesture animations. The engines produce animation data in MPEG4-compliant FAP/BAP format, which in turn drive a facial and skeletal body model in OpenGL. We briefly review GRETA's *GestureEngine* [7] (see Fig. 1) here to clarify where expressivity modification are performed. *GestureEngine* first performs text-to-speech conversion through Festival [18] which provides necessary phoneme timing for synchronizing gesture to speech. Communicative function tags which are candidates for gesture matching are extracted in the *TurnPlanner*.

The *GesturePlanner* matches communicative function tags to a library of known prototype gestures and also schedules rest phases when arms are retracted to the body. The *MotorPlanner* then concretizes abstract gestures by calculating key frame joint angles and timing. Finally, a bank of different *Interpolators* generate in-between frames to complete the animation.

To enable the thus-far generic, deterministic architecture for expressivity control, we augmented different stages of the architecture, which we will describe in the next section. Our implementation for gesture instantiation and modification is then presented.

5 IMPLEMENTATION: MAPPING EXPRESSIVITY INTO GESTURE ANIMATION PARAMETERS

Given a particular type of action and a set of values in the expressivity space, how can we modify non-verbal behavior production to communicate the appropriate expressive content? We need a suitable representation for gestures. We strive to preserve the semantic value of each gesture during the expressivity modifications. We hypothesize that effective strategies have to adjust behavior on multiple levels – from abstract planning (whether to search for a gesture for a given text at all), via gesture phase-level modifications (whether or not to repeat a stroke), down to adjusting velocity profiles of key pose transitions.

In the following, let the variables *oac*, *spc*, *tmp*, flt, *pwr* and *rep* stand for the *Overall Activation*, *Spatial Extent*, *Temporal Extent*, *Fluidity*, *Power* and *Repetition* parameter values we are trying to express.

5.1 Example

We introduce a sample dialog in transcript and APML-annotated form that will help clarify the expressivity computations we perform later on. The dialog was transcribed (and slightly edited) from an interview with author/journalist Helen Gurley Brown on the Open Mind television show[3]. We selected the following short utterance – words that coincided with gesture strokes are underlined:

"Whatever works for <u>you</u>, that's for <u>you</u>. But please don't tell me what works for <u>me</u>. Would you just please mind your <u>own business</u> and I'll mind <u>my business</u> and let's get on with the rest of our lives."

In the video, Hurley Brown performs a deictic reference to the interviewer (*<u>you</u>*), overlaid with a beat on the second *<u>you</u>*. A deictic gesture to herself with both hands accompanies the word *<u>me</u>*. After that, a metaphoric rejection is expressed by moving the right arm from shoulder-level downwards and out (*<u>your business</u>*). Finally, a round object in front of her torso is circumscribed to denote *<u>[her] business</u>*. We encoded this segment in APML, but for the sake of brevity only reproduce the beginning here in Figure 2. Text to be spoken by the agent is highlighted in blue.

[3] publicly available through the Internet Archive: http://www.archive.org/

```
01: <performative type="announce">
02:   <rheme>
03:     Whatever works for
04:     <emphasis x-pitchaccent="Hstar" deictic="you" intensity="0.4">you</emphasis>
05:     <boundary type="LH"/>
06:     thats for
07:     <emphasis x-pitchaccent="LplusHstar" intensity="0.4">you</emphasis>
08:     <boundary type="LL"/>
09:   </rheme>
10: </performative>
```

Fig. 2. APML Dialog.

5.2 Gesture Specification Language

In the past, we devised an abstract keyframe based scheme for gesture synthesis [7]. The gesture specification language is a sequence of key poses of the action, each of which describes wrist location, palm orientation and hand shape. Sets of key poses are grouped into the gesture phases defined by McNeill [3]. Our specification language was augmented by attributes defining which features of a gesture carry its semantic meaning and are thus invariable, and which features can be modulated to add expressivity. Description of the temporal aspect of each gesture was made implicit. Where previously kinematics were fixed through the frame times of the key frames, timing is now calculated using motion functions. Let us consider the gesture matched to the deictic pointing towards the user (line 4 of our APML script). This gesture consists of a simple arm movement that halts on the upper torso and a hand configuration that points at the conversation partner. The hand is not immediately retracted, but remains in a post-stroke hold. Figure 3 shows our encoding of this gesture. To conserve space, frames have been arranged horizontally and are to be read from left to right.

```
STARTFRAME                          STARTFRAME                          STARTFRAME
  FRAMETYPE stroke_start              FRAMETYPE stroke_end                FRAMETYPE post_stroke_hold
  ARM XC:fixed YUpperC ZMiddle        ARM XC:fixed YLowerC ZMiddle        ARM XC:fixed YLowerC ZMiddle
ENDFRAME                              HAND form_open thumb_default        HAND form_open thumb_default
                                      WRIST FBInwards PalmTowards         WRIST FBInwards PalmTowards
                                      ADDNOISE                          ENDFRAME
                                    ENDFRAME
```

Fig. 3. Sample gesture definition script.

The postfix `:fixed`, highlighted in red, indicates that a particular element of the gesture must not be modified by expressivity calculations. In the deictic reference, the agent's hand points towards the user who is facing the agent through the screen. Thus the agent should point straight outwards and not besides herself. We thus constrain the lateral X coordinate of the arm goal position to be in the center sector of McNeill's gesture space [3]. Note the absence of explicit timing information. The Gesture Engine calculates default durations. While we lose fine grain control compared to earlier explicit timing information, we gain parametric control over gesture phases as we will describe in section 5.3.

5.3 Expressivity Parameters

We now go through each of the identified dimensions of expressivity and explain how they are implemented. The stages of Figure 1 will be referenced to explain where gesture modification takes place.

Overall Activation A filtering is applied at the level of the *GesturePlanner*, which assigns gesture prototypes to input text mark up tags. Each input tag carries an intensity attribute that captures how important stressing the tag's content through nonverbal signals is – in line 4 of our APML example, the deictic gesture has an intensity of 0.4. 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. Thus, in our example, the deictic gesture will only be matched if the agent has an overall activation threshold ≥ 0.4. A similar principle of activity filtering was presented and implemented by Cassell et al. in [4].

Spatial Extent The space in front of the agent that is used for gesturing is represented as a set of sectors following McNeill's diagram [3]. We expand or condense the size of each sector through scaling. Wrist positions in our gesture language are defined in terms of these sectors (see Fig. 4). Represented by their center coordinates, the location of the sectors can be scaled asymmetrically using a simple matrix for homogenous coordinates. For meaningful scaling, we establish sector center coordinates $\boldsymbol{p}_i$ relative to the agent's solar plexus. Then the modified sector centers are given by:

$$\boldsymbol{p}_i' = \begin{bmatrix} I & \boldsymbol{spc} \\ 0 & 1 \end{bmatrix} \cdot \boldsymbol{p}_i$$

with:

$$\boldsymbol{spc} = \begin{pmatrix} 1.0 + spc \cdot spc_{agent_{horiz}} \\ 1.0 + spc \cdot spc_{agent_{vert}} \\ 1.0 + spc \cdot spc_{agent_{front}} \end{pmatrix}$$

$spc_{agent_{horiz}}$, $spc_{agent_{vert}}$, and $spc_{agent_{front}}$ are individual scaling factors in the horizontal, vertical and frontal directions that can define individualized patterns of space use. To find the location of articulation for a gesture, we first compute a point in the dynamically resized gesture quadrant that matches the gesture definition. We then calculate joint angles needed to reach that target with the IKAN inverse kinematics package [19]. Note that this technique is conceptually similar to EMOTE's kinematic reach space. While inverse kinematics are computationally expensive, they provide the only way of addressing arm movement in terms of goal positions. In a complex articulated joint chain such as a human arm, controlling forward kinematics (i.e., joint angles) directly yields non-linear and unpredictable results. In our example deictic gesture, increasing spatial extent will move the Y and Z goal coordinates away from the agent, while the X coordinate remains unchanged because of the `:fixed` constraint in the gesture definition.

Adjusting the elbow *swivel angle* (Tolani [19]) also directly changes the space taken up by the agent – extended elbows enlarge the body's silhouette. Since we are using inverse kinematics to position the wrist, we can control each arm's IK swivel angle θ for every key position:

$$\theta' = \begin{cases} \min(\theta \cdot (1.0 + 0.5 \cdot spc), \pi/2) & spc \geq 0 \\ \max(\theta \cdot (1.0 + 0.5 \cdot spc), 0) & spc < 0 \end{cases}$$

These modifications are performed at the *MotorPlanner* stage.

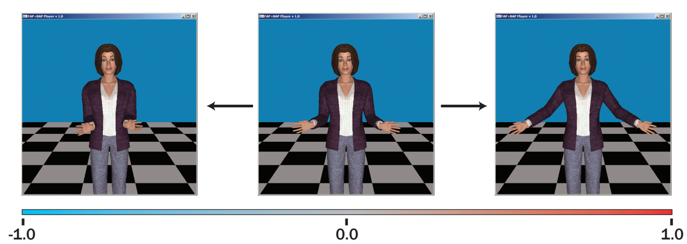

Fig. 4. Spatial Extent - arms extend or contract towards the torso.

Temporal Extent Starting from the synchronicity constraint on the end of the gesture stroke to coincide with the stressed affiliate in speech (McNeill [3]), we can calculate preceding and proceeding frame times from invariant laws of human arm movement described in [20]. During the planning phase, the actual distance traveled by the wrist joint in space is approximated by linear segments through key points. The duration to complete each segment can be derived from a simplification of Fitt's law as

$$T = a + b \cdot \log_2(\|\boldsymbol{x_n} - \boldsymbol{x_{n+1}}\| + 1)$$

The value of the velocity coefficient b has been established as 10^{-1} for average speed movements by Kopp [21]. Using this value as a starting point, the speed of a gesture segment can be adjusted as follows:

$$b = (1 + 0.2 \cdot tmp) \cdot 10^{-1}$$

Since we still have information about which part of the movement corresponds to which gesture phase, we can selectively amplify the stress of the gesture by increasing only the speed of the stroke to accentuate the gesture.

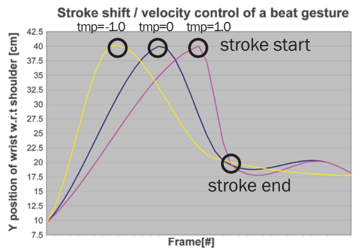

Fig. 5. Temporal Extent - stroke phases are faster or slower.

Fluidity This concept seeks to capture the smoothness of single gestures as well as the continuity between movements (the inter-gestural rest phases). We achieve low-level kinematic control through varying the continuity parameter of Kochanek-Bartels splines [22] used in the *Interpolator* component. Once again, this idea is close to EMOTE timing and fluidity control. In our implementation, we set the continuity parameter *cont* of the spline of the position interpolation spline for the wrist end-effector of each arm to equal the fluidity setting: $cont = flt$.

Fluidity also acts on the *GesturePlanner* level: larger fluidity increases the minimum timing threshold for retracting arms to a neutral position on the sides of the torso in between two gestures. During below-threshold pauses, arms are not retracted. Instead, two neighboring gestures are directly connected by interpolating between the retraction position of a previous gesture and the preparation position of the following gesture. In our example utterance, a low fluidity value would cause the agents arms to be retracted between the gestures accompanying the references to "*you*" and "*me*" (shown in transcript only, not in APML). A high fluidity setting would smoothly interpolate in the pause between gestures.

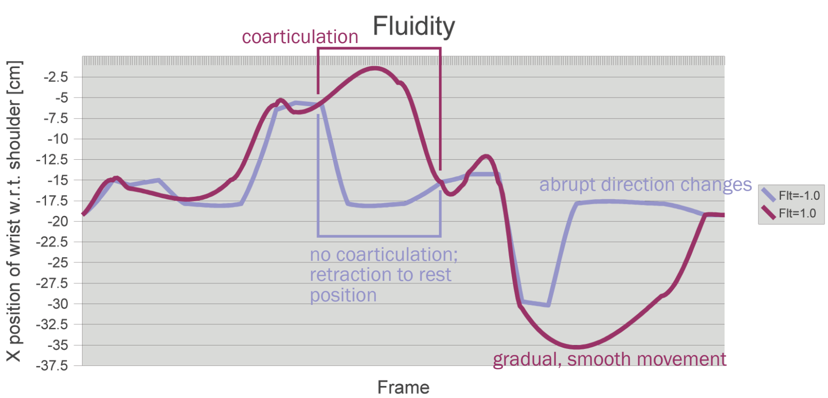

Fig. 6. Fluidity - rest phases and continuity are affected.

Power To visualize the amount of energy and tension invested into a movement, we again look at the dynamic properties of gestures. Powerful movements are expected to have higher acceleration and deceleration magnitudes. However, tense movements should exhibit less overshoot. This behavior is modelled with the tension and bias parameters of the kinematic TCB-spline in the *Interpolator*: $tension = pwr$ and $bias = pwr$. We also hypothesize that tense, powerful performances will be characterized by different hand shapes. If the configuration of the hand is not indicated as fixed in the gesture specification, high power settings will contract the hand towards a fist shape in the *GesturePlanner* stage.

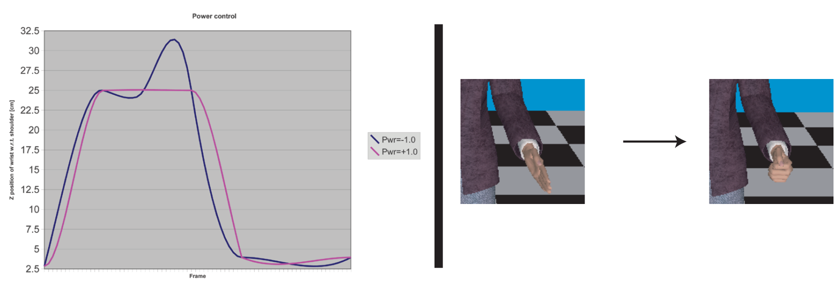

Fig. 7. Power - Overshoot and hand shape are affected.

Repetition We have previously introduced the technique of stroke expansion [7] to capture coarticulation/superposition of beats onto other gestures. Stroke expansion repeats the meaning-carrying movement of a gesture so that successive stroke ends fall onto the stressed parts of speech following the original gesture affiliate. It is possible to control the extent of repetition by selectively increasing the 'horizon' or lookahead distance that the stroke repetition algorithm analyzes. In our example, the original speaker superimposed a beat onto the post-stroke hold of the deictic gesture for *you* during the second occurrence of the term. By increasing or decreasing the repetition parameter, we can encourage or discourage such superposition, respectively.

5.4 Aggregating Parameters

We now show how setting expressivity parameters can generate a qualitatively different animation. Our system represents only a building block towards realizing affective action – exactly how motion quality is changed by the emotional state of an actor is still an open question in experimental psychology. Wallbott [14] had progressed the furthest to establish a mapping from emotional state to behavior quality but much work remains to be done. For now, until a reliable mapping is established, we use qualitative labels that are neutral with respect to emotion and personality, such as “abrupt.” In this case, “neutral” action is modified in the following ways: *Overall Activation* and *Spatial Extent* were disregarded (and thus left to the value 0) since abruptness is less apparent in the quantity of gestures or the amount of space taken up by those gestures. These two parameters are not important to convey abruptness. *Temporal Extent* was increased to 1 to speed up the meaning carrying strokes of all gestures. *Fluidity* was decreased to -1 to create jerky, discontinuous velocity profiles of arm movements and to discourage coarticulation from one gesture to the next – the agent’s arms are frequently retracted to a neutral position to create a disjoint performance. *Power* was set to a high value (1) to force a fist hand shape for beats and rapid acceleration and deceleration between gesture phases. Finally, *Repetition* was minimized (-1) since the rhythmic quality of a repeating movement counteracts the notion of abruptness. If we don’t want to generate a strongly abrupt movement, we can generate *slightly abrupt* behavior by interpolating the pertinent parameters between “neutral” and “very abrupt” settings while leaving other parameters unchanged.

6 EVALUATION

We conducted two evaluation tests. For the first test, we evaluated the following hypothesis: *The chosen implementation for mapping single dimensions of expressivity onto animation parameters can be recognized and correctly attributed by users.* 52 subjects were asked to identify a single dimension and direction of change in forced-choice comparisons between pairs of animation videos. 41.3% of participants were able to perceive changes in expressivity parameters and attribute those changes to the correct parameters in our dimensional model of expressivity. Recognition was best for the dimensions *Spatial Extent* (72.6% of modifications correctly attributed to this parameter) and *Temporal Extent* (73.8Modifications of *Fluidity* (33.9%) and *Power* (32.3%) were judged incorrectly more often, but the correct classification still had the highest number of responses. The parameter *Repetition* (28.0%) was frequently interpreted as *Power*. *Overall Activation*, or quantity of movement, was not well recognized. Overall, we take the results as indication that the mapping from dimensions of expressivity to gesture animation parameters is appropriate for the *Spatial Extent* and *Temporal Extent* dimensions while it needs refinement for the other parameters.

The second test with 54 subjects was conducted as a preference ranking task of four animations with different parameter combinations per trial to test the following hypothesis: *Combining parameters in such a way that they reflect a given communicative intent will result in more believable overall impression of the agent.* In each trial, one clip corresponded to the neutral, generic animation, two clips were variants of the chosen expressive intent (strongly and slightly expressive) and one clip had an inconsistent assignment of expressivity parameters. The subjects were asked to order the video clips from the most appropriate to the least appropriate with respect to the expressive intent. Participants in this second test preferred the coherent performance for the *abrupt* action described above over neutral and inconsistent actions as we had hoped. Similar results were obtained for the *vigorous* action. However, results were more ambiguous for our other test case - *sluggish* action. Two explanations are possible: the problematic implementation of some of the parameters may have led to unrealistic or incoherent animation; alternatively, gesture modification alone may not be sufficient - it may have to be integrated with gesture selection to achieve truly believable expressive action. A person gesturing sluggishly might not use the same gesture types as a vigorously gesturing one.

7 Conclusion and Future Work

We have presented a computational model to add movement quality to a communicative gesture. Six dimensions have been considered. We have performed two tests to evaluate the implementation of each of the six parameters individually and the ability of communicating a given intent when setting appropriately multiple parameters. We plan to refine our computational model, especially for the parameters that had low recognition rate. We are currently investigating how to use a video corpus of people talking emotionally that has been annotated with the perceived emotion and with several information regarding gesture: its type, its description and its expressive quality. We aim at using an analysis-synthesis loop to refine the mapping between emotion labels and expressivity dimensions.

8 Acknowledgments

Part of this research is supported by the EU FP6 Network of Excellence "HUMAINE", IST contract 507422. We are very thankful to Stéphanie Buisine for her help in the gesture evaluation tests.

References

1. Loyall, A.B., Bates, J.: Personality-rich believable agents that use language. In Johnson, W.L., Hayes-Roth, B., eds.: Proceedings of the First International Conference on Autonomous Agents (Agents'97), Marina del Rey, CA, USA, ACM Press (1997) 106–113
2. Nass, C., Isbister, K., Lee, E.: Truth is beauty: Researching embodied conversational agents. In J. Cassell, J. Sullivan, S.P., Churchill, E., eds.: Embodied Conversational Agents. MITpress, Cambridge, MA (2000)
3. McNeill, D.: Hand and Mind - What gestures reveal about thought. The University of Chicago Press, Chicago, IL (1992)
4. Cassell, J., Vilhjálmsson, H.H., Bickmore, T.: Beat: the behavior expression animation toolkit. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM Press (2001) 477–486

5. Tepper, P., Kopp, S., Cassell, J.: Content in context: Generating language and iconic gesture without a gestionary. In: Proceedings of the Workshop on Balanced Perception and Action in ECAs at Automous Agents and Multiagent Systems (AAMAS). (2004)
6. Noot, H., Ruttkay, Z.: Gesture in style. In Camurri, A., Volpe, G., eds.: Gesture-Based Communication in Human-Computer Interaction - GW 2003. Number 2915 in LNAI. Springer (2004) 324
7. Hartmann, B., Mancini, M., Pelachaud, C.: Formational parameters and adaptive prototype instantiation for MPEG-4 compliant gesture synthesis. In: Proceedings of the Computer Animation 2002, IEEE Computer Society (2002) 111
8. Kopp, S., Wachsmuth, I.: A knowledge-based approach for lifelike gesture animation. In Horn, W., ed.: ECAI 2000 Proceedings of the 14th European Conference on Artificial Intelligence, IOS Press (2000)
9. Lebourque, T., Gibet, S.: High level specification and control of communication gestures: The GESSYCA system. In: Proceedings of the Computer Animation 1999, IEEE Computer Society (1999) 24
10. Neff, M., Fiume, E.: Modeling tension and relaxation for computer animation. In: Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation, ACM Press (2002) 81–88
11. Chi, D., Costa, M., Zhao, L., Badler, N.: The EMOTE model for effort and shape. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co. (2000) 173–182
12. Hartmann, B., Mancini, M., Pelachaud, C.: Towards affective agent action: Modelling expressive ECA gestures. In: Proceedings of the IUI Workshop on Affective Interaction, San Diego, CA (2005)
13. Wallbott, H.G., Scherer, K.R.: Cues and channels in emotion recognition. Journal of Personality and Social Psychology **51** (1986) 690–699
14. Wallbott, H.G.: Bodily expression of emotion. European Journal of Social Psychology **28** (1998) 879–896
15. Gallaher, P.E.: Individual differences in nonverbal behavior: Dimensions of style. Journal of Personality and Social Psychology **63** (1992) 133–145
16. Martell, C., Howard, P., Osborn, C., Britt, L., Myers, K.: FORM2 kinematic gesture corpus. Video recording and annotation (2003)
17. DeCarolis, B., Pelachaud, C., Poggi, I., Steedman, M.: APML, a mark-up language for believable behavior generation. In Prendinger, H., Ishizuka, M., eds.: Life-Like Characters. Cognitive Technologies. Springer (2004) –
18. Black, A., Taylor, P., Caley, R., Clark, R.: Festival. (http://www.cstr.ed.ac.uk/projects/festival/)
19. Tolani, D.: Inverse Kinematics Methods for Human Modeling and Simulation. PhD thesis, University of Pennsylvania (1998)
20. Gibet, S., Kamp, J.F., Poirier, F.: Gesture analysis: Invariant laws in movement. In Camurri, A., Volpe, G., eds.: Gesture-Based Communication in Human-Computer Interaction - GW 2003. Number 2915 in LNAI. Springer (2004) 1–9
21. Kopp, S.: Synthese und Koordination von Sprache und Gestik fuer Virtuelle Multimodale Agenten. PhD thesis, Faculty of Technology, University of Bielefeld, Germany (2002) Infix DISKI-265, Berlin: Akademische Verlagssgesellschaft Aka GmbH.
22. Kochanek, D.H.U., Bartels, R.H.: Interpolating splines with local tension, continuity, and bias control. In Christiansen, H., ed.: Computer Graphics (SIGGRAPH '84 Proceedings). Volume 18. (1984) 33–41