Design and Evaluation of Expressive Gesture Synthesis for Embodied Conversational Agents

B. Hartmann
Stanford University
and University of Paris-8

M. Mancini
University of Paris-8

S. Buisine
LIMSI-CNRS
and LCPI-ENSAM

C. Pelachaud
University of Paris-8

ABSTRACT

To increase the believability and life-likeness of Embodied Conversational Agents (ECAs), we introduce a behavior synthesis technique for the generation of expressive gesturing. A small set of dimensions of expressivity is used to characterize individual variability of movement. We empirically evaluate our implementation in two separate user studies. The results suggest that our approach works well for a subset of expressive behavior. However, animation fidelity is not high enough to realize subtle changes. Interaction effects between different parameters need to be studied further.

Categories and Subject Descriptors

H5.2 [**Information Interfaces and Presentation**]: User Interfaces — *evaluation/methodology*

General Terms

Experimentation

Keywords

Embodied Conversational Agents

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. To increase believability and life-likeness of an agent, she has to express emotion and exhibit personality in a consistent manner [4].

We have previously developed GRETA, a multimodal agent that interprets utterance text marked up with communicative functions [1] to generate synchronized speech, face, gaze and gesture animations. A detailed description of the architecture can be found in [6].

Based on an aggregation of the measurement instruments used by the most pertinent studies [7, 2] and our analysis of a gesture corpus [5], we propose to capture gesture expressivity with a set of six attributes which we describe below in qualitative terms:

Overall Activation: quantity of movement during a conversational turn (e.g., passive/static or animated/engaged).

Spatial Extent: amplitude of movements (e.g., amount of space taken up by body).

Temporal Extent: duration of movements (e.g., quick versus sustained actions).

Fluidity: smoothness and continuity of overall movement (e.g., smooth/graceful versus sudden/jerky).

Power: dynamic properties of the movement (e.g., weak/relaxed versus strong/tense).

Repetition: tendency to rhythmic repeats of specific movements.

Our gesture system then transforms the expressivity specification into low-level animation parameters. The interested reader is referred to for an in-depth description of implementation [3].

2. EVALUATION STUDY

We now turn our attention to the two studies we conducted to evaluate our gesture expressivity model.

106 subjects (80 males, 26 females; aged 17 to 25) participated in our two evaluations studies. All were first and second year French university students.

Each subject completed only one of the two tests. Both tests consisted of observing sets of video clips (two per trial for test 1 - see Figure 1, four for test 2) in a Windows interface and choosing answers to questions about the sets of clips using standard interface widgets.

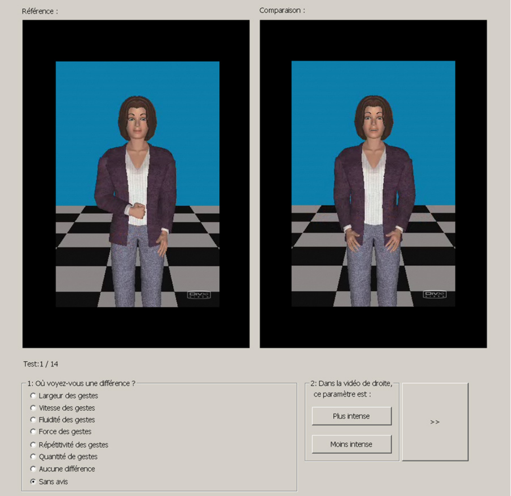
2.1 Test 1

The goal of the first study was to test hypothesis (H1): The chosen implementation for mapping single dimensions of expressivity onto animation parameters is appropriate - a change in a single dimension can be recognized and correctly attributed by users. In this test, subjects (N=52) were asked to identify a single dimension in forced-choice comparisons between pairs of animations.

Figure 2 presents the distribution of users' answers for each parameter. Gray cells indicate when they met our expectations: this diagonal totals 320 answers, which corresponds to 43.1% of accurate identifications of parameters. The chi-square test shows that this distribution cannot be attributed to chance ($\chi^2(35) = 844.16$, $p < 0.001$).

Recognition was best for the dimensions *Spatial Extent* and *Temporal Extent*. Modifications of *Fluidity* and *Power*

AAMAS'05, July 25-29, 2005, Utrecht, Netherlands.
Copyright 2005 ACM 1-59593-094-9/05/0007 ...$5.00.



Référence :

Comparaison :

Test:1 / 14

1: Où voyez-vous une différence ?

○ Largeur des gestes
○ Vitesse des gestes
○ Fluidité des gestes
○ Force des gestes
○ Répétitivité des gestes
○ Quantité de gestes
○ Aucune différence
● Sans avis

2: Dans la vidéo de droite, ce paramètre est :

Figure 1: The interface used for test 1.

		Perceived modification								
		Spatial Extent	Temporal Extent	Fluidity	Power	Repetition	Overall Activation	No modification	Do not know	Total
Modified parameter	Spatial Extent	77	2	5	5	3	3	3	8	106
	Temporal Extent	3	104	7	13	7	1	1	5	141
	Fluidity	2	4	42	10	23	2	34	7	124
	Power	7	8	23	42	9	6	27	8	130
	Repetition	18	12	17	20	35	5	10	8	125
	Overall Activation	7	7	7	17	6	20	41	11	116
Total		114	137	101	107	83	37	116	47	742

Figure 2: Distribution of users' answers as a function of the modified parameter.

were judged incorrectly more often, but the correct classification still had the highest number of responses. The parameter *Repetition* was frequently interpreted as *Power*. *Overall Activation*, or quantity of movement, was not well recognized.

Overall, we take the results of test 1 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.

2.2 Test 2

The second study was designed to test hypothesis (H2): Combining parameters in such a way that they reflect a given communicative intent will result in more believable overall impression of the agent. We considered 3 different types of behaviors: *abrupt* ("brusque"), *sluggish* ("mou"), and *vigorous* ("tonique"). This test (N=54) was conducted as a preference ranking task: the subject had to order four video clips from the most appropriate to the least appropriate with respect to the expressive intent.

For the *abrupt* and *vigorous* qualities, users preferred the coherent performances as we had hoped ($F(3/153) = 31.23$, $p < 0.001$ and $F(3/153) = 104.86$, $p < 0.001$ respectively). The relation between our parametrization and users' perception can also be expressed as a linear correlation, which amounts to +0.655 for the *abrupt* quality and +0.684 for the *vigorous* quality.

Conversely for the *sluggish* quality, the effect of input stimuli was not significant ($F(3/153) = 0.71$, *N.S.*): the overall rating of stimuli was random and the linear correlation was almost null (+0.047). This may be attributable partly to the inadequacy between the specific gestures that accompanied the text and the way a *sluggish* person would behave. This finding points towards the need of integrating gesture selection and gesture modification to best express an intended meaning.

3. CONCLUSION

We have presented an evaluation of an ECA augmentation to enable expressive gesturing with a large number of untrained subjects. The results confirm that our general approach is worthwhile pursuing. However, only a subset of parameters and a subset of expressions were recognized well by users.

This opens up the following paths for further research. First, we can improve the technical implementation of individual parameters to achieve higher quality animation and better visibility of changes to the parameters. Second, we also need to reflect about the interdependence of our expressivity parameters.Third, the integration of gesture selection and gesture modification remains to be implemented. Finally, the study design can be improved upon in the future. A shortcoming of the current test was that only a single utterance was used with varying animations. A wider variety of different situations and utterances is needed to help control for the influence that the particular choice and sequencing of gestures in a single sentence may have on perception of expressivity.

4. REFERENCES

[1] B. DeCarolis, C. Pelachaud, I. Poggi, and M. Steedman. APML, a mark-up language for believable behavior generation. In H. Prendinger and M. Ishizuka, Eds., *Life-Like Characters*, Cognitive Technologies, Springer, 2004.

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

[3] B. Hartmann, M. Mancini, and C. Pelachaud. Implementation of expressive eca gesture synthesis. *Submitted to GestureWorkshop*, 2005.

[4] A. B. Loyall and J. Bates. Personality-rich believable agents that use language. In W. L. Johnson and B. Hayes-Roth, Eds., *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, 106–113, Marina del Rey, CA, USA, 1997. ACM Press.

[5] C. Martell, P. Howard, C. Osborn, L. Britt, and K. Myers. Form2 kinematic gesture corpus. video recording and annotation, 2003.

[6] C. Pelachaud, V. Carofiglio, B. D. 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.

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