Eurographics/ACM SIGGRAPH Symposium on Computer Animation (2004)
R. Boulic, D. K. Pai (Editors)

Graceful Degradation of Hand Gestures

Zsófia Ruttkay and Han Noot

Centre for Mathematics and Computer Science, Amsterdam, The Netherlands

Abstract

Embodied Conversational Agents (ECAs) should interact with the user in a reactive way, using non-verbal modalities parallel to speech. Hand gestures, particularly, should be generated on the fly, and precisely timed to speech. However, this does not always succeed. In this short paper we discuss what can be and should be done if the intended gesture cannot be performed in full detail. First we outline the possible shortcomings and problem categories in hand gesturing. Then we give a conceptual description of remedies for time shortage problems. Finally, we report about implementation of graceful degradation of timing, in the framework of our GESTYLE gesture markup language.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Animation, Embodied Conversational Agents

1. Introduction

1.1. Hand gestures by ECAs

Hand gestures contribute a lot to communication. They may modulate, augment or even substitute speech. Also, an embodied agent not using his hands looks unnatural. For this reasons it is more and more the practice to use anembodied conversational agents (ECAs) [2] which are endowed with full body and limbs capable of gesturing. (Usage of other modalities like facial expressions and body postures are beyond the scope of this paper.) It has become a common practice to control an ECA in high-level terms, indicating the gesture to be performed (e.g. by text markup), and leaving the specification of subtle parameters (e.g. for motion characteristics) to the animation system. Details like amplitude and timing are to be derived by the animation system on the basis of additional information, which may not be even available to the user in advance, as it is the case with e.g. the duration of the utterance generated by some TTS engine.

1.2. What can go wrong?

In practice, both with ECAs and with human beings it may happen that the intended gesture does not get, or even cannot be performed, because of one of the following reasons:

- **Time shortage** The gesture should be performed so fast that it either looks unnatural and/or meaningless, or due to physical constraints of the joints (characteristic of the human body) or due to insufficient computer processing resources for the ECA it is not possible to render the motion as declared.
- **Priorities and interruptions** The hand to perform the gesture is still occupied with another function which has to be completed before the gesture can start. The current gesture's performance may get interrupted before completion, as the hand is assigned to another gesture. (Note that here we do not deal with the issue of scheduling of alternative modalities as a possible remedy to resolve modality conflicts.)
- **Body morphology limitations** The ECA's body model does not contain the body parts with the required motion capabilities. This is common with different level of detail (LOD) H-anim models: it is only LOD2 models which have also moveable finger joints.

Currently, we are working on providing solutions to recover from the above cases in a graceful and automatic way. In this paper we will explore tge problems related to time shortage. To our knowledge, the possibility of graceful degradation has not been looked at in case of gestures. The issue is related to providing different LOD animations [1], also for the human body [4], but there the motivation is either adjustment to the LOD of the body model, or to save computational resources in cases where full detailed motion would not be perceived, because of viewing conditions.

2. Graceful degradation of gestures

When performed completely, a gesture consists of 5 stages: preparation, pre-gold, stroke, post-hold, and retraction. For discussion of the stages and the justifying literature, see [6]. Each of these phases can be performed with different speeds and amplitudes. Hence, there are two ways of influencing the duration of a gesture:

1. Changing the initially prescribed duration of its phases. What freedom one has here, depends a priori on the meaning of the gesture, e.g. all phases of a greeting may be scaled down by the same amount, but for emphasis and especially for conducting, the beat should stay on time which in general excludes uniform scaling.
2. Changing the amplitude of its motion, possibly "per joint". An extreme case is omitting certain motions, e.g. omitting the wave (a wrist motion) in a greeting or omitting all but the finger motion in a counting gesture.

2.1. Strategies

Currently we are developing strategies for gesture degradation, relying on the tactics 1 and 2 above. In designing the strategies, we rely on observation what type of strategies occur among humans for different gestures. We propose a hierarchy of graceful degradation actions, such as:

- **Strategy A:** speed up uniformly all phases (e.g. used with waving).
- **Strategy B:** strongly shorten the durations of hold-phases, if not enough, scale the other phases, but but always so that the end of the stroke lies at a certain time (e.g. used with beat gestures).
- **Strategy C:** scale down all amplitudes. If not sufficient, use faster movements (e.g. used with a shrug)
- **Strategy D:** reduce arm movement, preserve finger movement (e.g. used with counting).

2.2. Fine tuning according to actual computational resources

In order to make a graceful degradation algorithm work on different platforms, we need to know the minimal actual durations of gesture phases. If a (platform-independent) strategy prescribes a certain, increased speed as a remedy, we still must check if the platform can realize this speed. If not, another strategy should be chosen. So, for instance, strategy A can be ideal in principle, but because of resource limitations strategy D has to be chosen instead. In order to take into account the limitations of the current platform too, one should know the 'fastest possible' motions which can be realized. E.g. what is the shortest time on this platform to lower the arm, to open the hand, to do this in parallel etc. One needs separate answers for motions of different complexity (amount of parallelism), and with respect to the (also background) processing load of the operating system.

In order to get a picture about the critical time durations, we will design a series of systematic test scripts for all gestures performed in details to be used in different degradations (arm+hand, hand only, half intensity). Running the scripts will provide a database with duration information, which is to be used as a system-dependent reference in the degradation strategies These test scripts are to be run once after every environment change (new computer, new OS, different ECA etc.). Such an initial testing phase will assure good performance afterwards, when the ECA is to be used. A different strategy could be not to require a first testing stage, but learning about the shortcomings and limitations in course of the gesturing tasks to be performed.

3. Implementation of an experimental platform

As an experimental platform we use our hand gesturing system for H-anim compliant ECAs developed on top of the STEP language [3]. For high-level control, we use our own GESTYLE text markup language [5]. We will use this platform to study the range of acceptable gesture parameters and the effect of strategies on perception of gestures by using an iterative cycle, involving gesture (re-)design and evaluation of the gesture by users.

Up to now we implemented strategies using time-scaling of gesture phases, i.e. (refinements of) strategies A and B. We could "stay in sync" with the speech running the GESTYLE script in different environments and with different ECA bodies. On the other hand, it is apparent that when gestures get performed too fast, they are perceived as unnatural. So we will use also C and D type strategies as an alternative for using unbounded speed increase.

References

1. Carlson D., Hodgins J. Simulation Levels of Detail for Real-time Animation, In: *Proc. of Graphics Interface '97*, pp. 1-8, 1997.
2. Cassell J., Sullivan J., Prevost S., Churchill E. (Eds.) Embodied Conversational Agents, MIT Press, Cambridge, MA. 2000.
3. Eliens A., Huang Z., Visser C. "A platform for Embodied Conversational Agents based on Distributed Logic Programming", In: *Porc. of the AAMAS Workshop on "Embodied conversational agents - Let's specify and evaluate them!"*, Bologna, 2002.
4. Giang T., Mooney R., Peters C., Sullivan C. ALOHA : Adaptive Level of Detail for Human Animation, Europgarphics'2000, Short papers.
5. Noot H., Ruttkay Z. "Style in Gesture", In: A. Camurri, G. Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction*, LNCS 2915, Springer-Verlag, 2004.
6. Ruttkay Zs., Pelachaud C., Poggi I., Noot H. "Excercises of Syle for Virtual Hu-mans", In: L. Canamero, R. Aylett (Eds.), *Advances in Consciousness Research Series*, John Benjamins Publishing Company, to appear.