Analysis of gesture expressivity modulations from cartoons animations

Nicolas Ech Chafai
University of Paris 8 / France
Télécom R&D
n.chafai@iut.univ-paris8.fr

Catherine Pelachaud
University of Paris 8
c.pelachaud@iut.univ-paris8.fr

Danielle Pelé
France Télécom R&D
danielle.pele@francetelecom.com

ABSTRACT
In this paper, we describe a gesture expressivity analysis of a character in a conversational interaction. To determine gesture properties that attract and maintain perceptual attention, we study if there exists some effects in the field of gesture expressivity modulations. We study this role at a low level (gesture phase) and at a higher level (discourse structure) based on a corpus annotation of Tex Avery (MGM) cartoons. First results of the analysis point out synchronization properties between modulations and gesture phase or discourse structure.

Author Keywords
Gesture expressivity, gestures and discourse, corpus annotation, visual attention, 2D and 3D animations.

INTRODUCTION
We are currently developing an embodied conversational agent ECA endowed with an expressive and communicative behavior (Hartmann et al., 2005; Poggi & Pelachaud, 2000). Our aim is to endow an ECA animation system with the capability to attract the attention of a user at specific points of the ECA animation. At the gesture level we have to set which gestures provide semantic information (Kendon, 2004), and which gestures attract the gaze of the other interactant: as Cosnier said in preface of Calbris (2003), there are gestures that carry a meaning, and gestures that manage communication and have pragmatic functions.

We look at how gesture expressivity varies in the animation. We also investigate how gesture expressivity properties can act as a pragmatic tool. Our hypothesis is that gesture expressivity modulations could partly play this role. These modulations could provide, by a sudden change in the perceived behavior of the speaker, some of her intentions to the listener.

Our approach is based on the analysis of traditional animations: animators have developed sharp skills over decades in eliciting empathy and in regulating attentional behavior of spectators through character’s movement and expressions; we aim at taking into account these skills to develop our application.

To get precise data on gesture expressivity modulations, we annotate each expressivity parameter defined in Hartmann et al. (op. cit.), not at a gesture unit or phrase level, but at a gesture phase level. We describe our choices for each of the annotated parameters.

In the remaining of this paper, we first present a state of the art of the works that apply traditional animation features to 3D animation, and that study more specifically gesture abilities to attract listener’s gaze attention. Then we describe our corpus we used to analyze gesture expressivity in cartoons, and we precise our annotation methodology. Finally we describe the results that were observed in the analysis process, and that give a first view on the role played by gesture expressivity modulations during a conversational interaction.

STATE OF THE ART
Some previous works already tried to produce 3D animations based on traditional animation. Several fundamental principles of traditional animation (Thomas & Johnston, 1981) have been applied to 3D animation: Choi et al. (2004) proposed a system able to computationally apply the principle of anticipation on a 3D animation: through the production of a backward movement over the following movements, this principle leads to direct the spectator’s attention towards the place of the action. Lance et al. (2004) studied animators’ abilities to express emotion and empathy in cartoon characters, and built up a system able to generate an expressive gaze for a virtual character. Bregler et al. (2002) captured the animations of 2D objects (deformable or not) by following some feature points; this follow-up allows one to animate in the same way different kinds of 2D or even 3D objects. Not only the movement is identically produced, but it also preserves the same expressivity. But these works do not resolve the question whether imitating 2D animation onto 3D animation is perceptually acceptable by a spectator or not. Lasseter
(1987) pointed out how the principles from 2D animation could be successfully applied in 3D animation; however the perception that the spectator has could change if we limit 3D animation to a 2D imitation and if we do not look at finding to what extent the 2D animation principles could be interpreted. In our work, we try to find some new rules of 2D animation that could be applied in a gestural animation of 3D characters.

In the domain of human gesture study, there exists works dealing more specifically with gesture ability to attract listener’s attention. Eye tracking techniques allow researchers to follow where and when a listener gazes at, and in particular on which gestures he gazes at. This type of disposal was adopted by Gullberg & Holmqvist (1999) to study which are the elements that lead to gaze at a particular gesture; laterality seems to play a preponderant role, as opposed to self centred gestures. With the same kind of disposal, Barrier et al. (2005) have determined that through the use of deictic signals, a speaker is able to redirect listener’s focal attention toward his gestures, or toward a virtual space built by his gestures. In cartoons (Thomas & Johnston, op. cit.) noticed how efficient an animation that could be understood from its silhouette is; this observation complements results from Gullberg & Holmqvist by adding a notion of point of view: a same body gesture can change silhouette type depending from where we are looking at it. Our work aims at determining new criteria that could attract spectator’s gaze attention through some kinds of gesture expressivity properties, and to implement these criteria in an ECA.

CORPUS
We base our corpus on two videos from Tex Avery cartoons (MGM). Each of these videos lasts about ten seconds. Our choice of a low level analysis (described later) leads to a corpus with little data. In regard of our aim to animate conversational agents, we chose sequence showing a conversational interaction between characters; the first one serves as basis for our analysis, the second has been used to verify the results from the first one. One of these videos comes from the cartoon Blitz Wolf (1942): it displays a pig character trying to convince two other pigs to protect their selves against a wolf’s threat. The other video comes from Henpecked Hoboes (1946): in this cartoon, the main two characters are George and Junior who are trying to catch an hen to feed themselves; in the sequence that we are interested in, George explains to Junior the set of actions they will have to perform to reach their goal. Theses two sequences exhibit two different discourse goal: in the first one the pig aims to incitate and advice; in the second one George aims to communicate informations.

ANNOTATION DESCRIPTION
To get precise data on the modulations of gesture expressivity, we annotate the expressivity on a gesture phase level. Kendon defines gesture unit, gesture phrase, and gesture phase, as three different levels in the gesture production (2004, chap. 7). There are different kinds of gesture phase; Kendon organizes them around the phase of stroke recognized as the expressive part of the gesture: preparation, stroke, post-stroke-hold, and recovery. Kita et al. (1997) refine these phases and distinguish: preparation, stroke, hold and independent hold, retraction, and partial retraction.

In our analysis, we are using most of the phases described by Kita et al. For sake of simplicity we consider ‘independent hold’ as having the same function as ‘hold’; no distinction in both terminologies is made. And we add the phase of anticipation: it refers directly to one of the fundamental principles of animation as described in Thomas & Johnston (1981); from our point of view it seems necessary to add this phase in the analysis. Thus, we consider the following set of gesture phases (Kita et al., op. cit., Kendon, op. cit., Kipp, 2003):

- Anticipation: preceding a gesture phase, the arm may produce a backward movement. This happens due to motor constraints, but also to get spectator’s attention focusing on the following movement;
- Preparation: the arm moves to the location where the speaker wants to produce his stroke;
- Stroke: expressive phase of gesture, it is produced synchronously or anticipates the verbal referent;
- Hold: the stroke may be hold for a while;
- Recoil: following the stroke, the arm may recoil to emphasize this stroke;
- Retraction: the arm moves to a rest position;
- Partial retraction: before the arm finishes moving to a rest position, another gesture starts and thus ends up the retraction.

The expressivity parameters we chose for our annotation are those implemented by Hartmann et al. (2005) in their conversational agent Greta. They correspond to: fluidity is

1 Produced right in the middle of WW2, this cartoon is a short propaganda film: the animators are displaying Big Bad Wolf under A. Hitler’s features and are warning how dangerous he is. The main pig figures the judgement value of the American state. Animators are figuring this pig to display to American people what kind of behavior they have to adopt towards WW2: they have to support war effort. Obviously, the title of Blitz Wolf directly refers to the “Blitz Krieg” practiced by Hitler.

2 Refers to George and Lennie characters from J. Steinbeck’s novel “Of Mice and Men” (1937).
the smoothness and continuity of overall movement (e.g. smooth, graceful versus sudden, jerky); power is the dynamic properties of the movement (e.g. weak/relaxed versus strong/tense); spatial extend is the amplitude of movements (e.g. amount of space taken up by body); repetitivity is the tendency to rhythmic repeats of specific movements along specific modalities.

Three values are available for each parameter: positive, neutral, negative. And we define parameters with a set of criteria:

- Fluidity: it corresponds either to the level of continuity between successive phases, or to the movement curvature, or even to presence of an anticipation phase;
- Power: it stands for the shape opening (opened / closed), or the acceleration of the arm, or even for the continuity in tension at the end of the movement;
- Spatial expansion: we define it as the gestural space, or the swivel angle, or even as the point of view from which the gesture is seen i.e. with a high or a low silhouette (Thomas & Johnston, op. cit.);
- Repetitivity: repetition of the gesture stroke.

Both analyzed videos are annotated using Anvil tool (Kipp, op. cit.), which allows us to precise each value of expressivity parameter for each of the gestural phases. Then, we observe and notice the modulations in gesture expressivity: that is, we analyze the variation over time of each expressivity parameter. We are not interested in finding out which particular parameter varies. Rather we concentrate on the variation itself of the parameters.

On the one hand we try to find some kinds of correlations between theses modulations and the production of the corresponding gesture. On the other hand we try to find some correlations between theses modulations and the structure of the verbal utterance, in order to observe if there is any regularity in it.

OBSERVATIONS

We have annotated for each gesture of our corpus the value of each expressivity parameter. When analysing the data we do not consider the annotated value of each parameter as such but we look at the variation over time of these values. This analysis is based on one of the two annotated videos; the second is used to verify the results we obtain. We observe two types of variations that are found over each expressivity parameter. We, now, consider no more the value of each expressivity parameter but these variations which are:

- Irregularities: it corresponds to a brief time period (a single gesture phase) in which the annotated modality has a sudden change of value, and then comes back at the original one just after this phase.

For example, it happens when a character produces a powerful sequence of movements, except for a single phase that is produced with a low power (Figure 1a);

- Discontinuities: it corresponds to a sudden change in the annotated modality. For example, it happens when a character of the animation produces a sequence of movements with a low power, succeeding to a sequence with powerful movements (Figure 1b).

That is, each time a sudden variation in gesture expressivity occurs, it is defined as a discontinuity; but if this variation directly precedes another sudden variation we will speak in terms of irregularities. Figure 1 illustrates graphically these concepts.

![Figure 1a and 1b: Irregularities and Discontinuities](image)

Each occurrence of these two modulation types have been noticed: (4; 8) for irregularities, and (10; 6) for discontinuities. Some invariance appear to inform on their role in a conversational interaction in a cartoon, as described in the following sections. There is differences in results quantity of the two videos; this is partly due to a difference in the quantity of gesture repetitions for each video and the structure of their utterances.

THE FUNCTIONS OF IRREGULARITIES

From the annotation, we observe that irregularities seem to play a role of anticipation by linking similar elements of the enunciative structure as: occurrences of gesture repetitions (2; 7), performatives of a same general class (Poggi & Pelachaud, 2000) (1; 1), gesture phrase (1, 1).

By linking similar structures, irregularities are able to perform the role of an AND connector that allows the spectator to anticipate the behavior the character will...
display. Following the principle of anticipation (Thomas & Johnston, 1981), this property should enhance the visibility of gesture, *ie* to enhance our propensity to gaze at this particular gesture.

**THE FUNCTIONS OF DISCONTINUITIES**

We also observe that discontinuities may perform a relation of contrast. This relation may take diverse forms. It could enhance the emphasis on a specific gesture by contrasting it from the others (6; 1): over a whole sequence produced with low fluidity, only a single gesture phrase (and not phase, that would have lead to an irregularity) has been produced with high fluidity. That leads to an isolation effect of this gesture phrase. It could also contrast the action verbs of the utterance when they are gesturally illustrated (3; 2); each occurrence of these gestures is produced with a specific expressivity. Another form of discontinuity was noticed when the speaker enunciates a new type of general class of performative (1; 2), he changes his expressivity. Thus, discontinuities are a way to oppose different kernels of the enunciative structure.

Theses different functions of discontinuities seem to be closely linked to a relation of contrast between each of the levels they are referring to. This relation is defined as the speaker’s intention that the addressee recognizes, by comparison, similarities and differences of the kernels of the enunciative structure (Mann & Thompson, 1988).

**THE FUNCTIONS OF THE MODULATIONS**

By summarizing the functions performed by the two kinds of modulations in gesture expressivity, it appears that they act at the different levels of the enunciative structure, and that they do not depend on the kind of performative act the speaker enunciates.

Modulations appear as a pragmatical tool. We noticed that irregularities could affect the spectator’s attention through theirs anticipation properties. Discontinuities perform a relation of contrast that suggests an other attentional effect: as Feyereisen (1997) noticed “communication supposes to perform contrasts. A signal is perceived with more clarity if it is distinguishable from noise or other signals” (p. 39).

**CONCLUSION**

We have presented an annotation schema to study gesture expressivity modulations. We annotated a corpus of 2D cartoons and analyzed at the gesture phases level. This analysis leads us to extract some functions of the modulations in gesture expressivity that seem to act as a relation of similarity (irregularities), and as a relation of contrast (discontinuities). We are aware that our study shows some limitations (size of corpus, analysis by one annotator) and its results should be used with precaution.

Nevertheless it appears that the modulations are a kind of pragmatical resource that could have an interest in the animation of ECA, and that could act on the attention that the spectator bears on a gesture. In the near future, an evaluation process will simulate the previous results in an ECA application to test whether they are efficient or not for attracting one’s gaze attention and interest.

**REFERENCES**


