EXCERCISES OF STYLE FOR VIRTUAL HUMANS

Zsófia Ruttkay
Center for Mathematics and Computer Science
POBox 94079, 1090 GB Amsterdam, The Netherlands
e-mail: Zsofia.Ruttkay@cwi.nl

Catherine Pelachaud
University 'La Sapienza'
via Buonarotti 12, 00185 Roma, Italy
e-mail:cath@dis.uniroma1.it

Abstract

The style of humans is a source of information about the person, her personality and enjoyment in every-day life. We wish to endow virtual characters with style of nonverbal communication. We outline the diversity in gesturing style among humans, and identify the decisive factors. We propose a computational model for defining the styled virtual characters and generate gestures accordingly. The ideas are to be demonstrated on styled facial expressions.

1 Introduction

1.1 Motivations

The title is inspired by Raymond Queneau's famous work, 'Exercises in Style' [21]. In this ingenious literary work, the French author takes a few-line banal story happening in a crowded bus, and in 99 exercises, tells it in different styles. He does it so well, that the reader can see the character acting for his eyes: how he gestures, whether he has a happy face or one of a bitter grumbler.

Another example of the power of style is 'Creature comforts', an Oscar-winning animation film [1], in which animals talk and gesture in the well-recognizable style of some human groups (of certain nationality, social status).

Thus the style is a source of information on the speaker, as well as of variety when communicating with real people. Moreover, a pioneering empirical experiment has shown that such factors as the ethnicity and the personality (introvert/extravert) of a synthetic character – even if manifested in a simple, static feature – do have consequences on their effect on the user [12].

One would like to benefit from style also when confronted with embodied conversational agents (ECAs) [3]. Even if one should not expect a virtual human to act like a blood and flesh real person, the current situation of rather puppet-like, styleless virtual characters should be improved, even if step by step.

1.2 Related work

The first steps have been taken in the direction of expressive ECAs, by endowing them with the capability of showing emotions [8]. Subtle issues like the impact of social role [5, 20] and personality [2] has been addressed. Also non-verbal signals have been used to accompany speech to make an ECA more expressive and believable. However, these works concentrate on modeling the psychological, social and communicative aspects of emotional and cognitive state. Usually the presentational issues are not dealt with as a research topic, but as a practical task for an animator, often only to make a specific application or demonstrator.

The importance of non-repetitiveness has been convincingly pointed out by Perlin [17], who (and several of his followers) used random noise to generate different instances of face and body motion, also in 'idle' state of the character. Badler and his colleagues have developed EMOTE [4], a computational framework inspired by Laban dance annotation, to modify expressiveness of hand, body and face gestures of ECAs. In the computer animation world [24] it has been recognized how important it is to 'add style' to (captured or synthesized) motion. Recently, there have been initiatives to develop XML-base markup languages to encode some of the 'human' aspects of multi-modal communication [25].

1. 3 Our objectives

We are interested in making ECAs which are expressive and individual in their presentation and display. Particularly, we wish to endow ECAs with a **style in their non-verbal communicative behaviors**. A straightforward example for the possibility of using different styles is the communicative act of greeting. Factors like nationality, gender, age, physical state, mood, the unique motion habits and characteristics of

the person all contribute to decide if the greeting will be verbal and/or non-verbal, what facial expression and/or which hand be used and in what way to express the greeting.

We would like to **identify aspects and parameters of style** in human to human interaction, and rely on these findings to build styled ECAs. The presentation aspects can be considered as basically static parameters to define and tune the non-verbal presentational characteristics of an ECA: what type of gestures occur in his repertoire, how he chooses a behavior from several alternatives for a specific communicative function, in what way are the gestures presented.

We are developing a generic **computational model of gesturing style**, which makes it possible to define quickly and easily variants of an ECA animation with different, some group-specific and/or personal style. Our work is at the borderline between the 'mind' and 'body' aspects of ECAs. We are interested in finding out what and how a model of 'style' can be included.

In this article we report on ongoing work towards the above objectives. In section 2, we show how style is manifested in non-verbal communication and what are the decisive factors in using a specific style. In the following section we outline our computational model of styled ECAs. Finally, we discuss the status of our work, and outline further research tasks.

2 Gesturing styles and parameters

From the multitude of manifestations of style, we restrict our focus on style of non-verbal communication, namely facial, hand and body gestures, which accompany (sometimes replace) speech. A taxonomy of gestures and overview of research on human gesturing can be found in [9, 11]. Among the different gesture classes some iconic and emblematic gesture are typical of a **nation or ethnic group**. E.g. the two-fingers V sign for 'victory' should be made with palm outwards in England, while it can be used with both hand positions in the USA [3]. The very same gesture may convey different meaning. E.g. nodding is affirmative in most countries, but not in Bulgary and Greece. There are gestures and gesturing habits typical for **professions**, **cultural** and **social** status. People of different **sex** and **age** are likely to gesture differently.

But there is also much **individuality** in gesturing, considering the expressiveness of a gesture, the meaning and frequency of occurrence of a gesture. Humans keep moving (stretching chin or nose, rubbing hands, changing posture) even when they are 'idle': the type and frequency of 'idle motions' is often characteristic of a person and her personality [10]. While the way of gesturing is typical of a person, it is not entirely predictive and repetitive. On the one hand, a person often has several gestures for a certain communicative meaning. On the other hand, a single gesture is used with slight variations, avoiding the exact repetition of a gesture, which is often not the case with synthetic characters, making their motion looks 'unreal'.

When designing the gesturing style of an ECA, the following aspects have to be taken care of:

1. We need to define the gestures which constitute the **gesture repertoire** of the ECA, and to specify the meaning of the gestures by a **gesture dictionary**.
2. A mechanism should be given to **plan the presentation of gestures** expressing certain prescribed communicative functions: which modalities to use, and which gestures.
3. The chosen **gestures** should be tuned to achieve desired expressive characteristic.

2.1 Gesture dictionary and repertoire

To begin with, we settle our terminology. A **gesture** is some motion involving one or more of the modalities face, head, hands and body. We do not bother about the nature of the gestures, we assume that they are motions either used to express some meaning, or to fulfill biological necessities (breathing, blinking). **Basic gestures** refer to a single facial feature (eyes, eyebrows, mouth) or a single other modality (right/left arm, hand, leg and foot, …). Gestures may be defined by composing 'basic' gestures. For example, a greet gesture may correspond to a 'head nod' and a 'smile'. Examples of basic gestures are:

eye_gesture: look_straight, look_left, look_right, ...
eyebrow_gesture: eyebrow_up, eyebrow_frown, ...
mouth_gesture: corner_mouth_up, mouth_open
head_movement_gesture: head_nod, head_shake, ...
head_direction_gesture: turn_left, turn_right, ...
basic_gesture: eye_gesture; eyebrow_gesture; ...

The compound gestures are formed from basic gestures by two operators:

& parallel composition
\+ sequential concatenation

gesture: basic_gesure, basic_gesture + gesture, basic_gesture & gesture.

The & operator is stronger than +. By parallel composition, the constituting gestures should start at the same time, by concatenation, sequentially, as soon as the previous one has finished. In the definition of a compound gesture, all basic gestures occurring in parallel composition should refer to different modalities or features. That is a gesture may not be composed of two basic 'eyebrow_gesture' with different values such as eyebrow_up and eyebrow_frown. While a gesture may be defined by a basic 'head_direction_gesture' (say 'turn_left') and a basic 'head_movement_gesture' (say 'nod'). The collection

of all basic gestures which an ECA can use is the **gesture repertoire**.

When instantiated, the basic gesture corresponds to an animation of the involved facial features or body parts. A gesture may be instantiated in two ways which vary in complexity. In the simplest case, instantiation of a gesture corresponds to the instantiation of the basic gestures that composed it. In a more sophisticated framework, a gesture is defined as a set of basic gestures linked by constraints [22]. Some characteristics of the gestures, like duration, intensity, on/offset time (for facial gestures), preparation/hold/withdrawal time (for hand gestures) can be specified. This framework enables the modification of gesture definition to produce variants of it with different **expressiveness**. Such a definition of gesture increases the flexibility of the creation process for animation as well as it allows the non-repetition of the final animation for each gesture instantiation [22].

Gestures may have a **communicative function** [9, 11, 18]. We use the taxonomy introduced by Poggi et al. [19] to gather communicative functions in four main groups based on the type of information they contain. Communicative functions provide information about location and properties of objects, concepts or events, about the speaker's affective (pleased, angry) and meta-cognitive state (thinking, listening), believes (agree / disagree, certainty) and intentions (performative, turn-allocation). Communicative functions are identified by a meaning and a signal. It is an essential characteristic of an ECA what gestures it uses to express a communicative function. There can be several alternative gestures expressing the same meaning (see Figure 1).

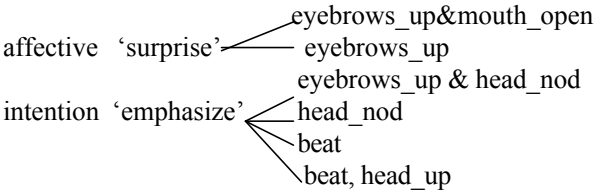

Figure 1. For all communicative functions, there exist a one to many mapping between meaning and signal.

For a communicative act, the mapping between meaning and signal can be the one generally used by an ethnical or professional group, or can be idiosyncratic to the person. In each case, one has to know the person, her culture, her profession and so on, to decode the communicative function of the gesture (e.g. special way of greeting). Besides the mapping of meaning to gestures, there is difference in choosing a particular gesture from the possibilities allowed by the one-to many mapping as the selection is also characteristic of the person. E.g. right-handed people will use right-hand beats rather than left-hand beats, an 'introvert' person uses gestures less often, an Italian person uses beats more often than an Englishman to express emphasis.

We model the characteristics of usage of gestures marking the same communicative act by assigning probabilities to the individual gestures. Taking this characteristic into account too, the gestures used for a communicative act are given by a **gesture dictionary** entry of the following form:

communicative_act: (Gesture$_1$,P$_1$),... (Gesture$_n$, P$_n$)

where *Gesture$_1$,..., Gesture$_n$* are gestures, covering the alternatives of expressing the communicative function, and *P$_1$,... P$_n$* are probabilities of using the specific gesture in this role.

Different gesture dictionaries can be defined and given to set different 'gesturing code': one for gesturing habits of a culture (Italian gesturing) or profession (teacher-like gesturing), one for special gesturing of a certain person. However, one can notice that the use of a gesture or the display of a particular facial expression corresponding to a communicative act may also depend on the conversational context. Ekman [7] defines 'display rules' to embed cultural differences in the display or not display of a facial expression. Poggi and her colleagues [19] argue that the not only the expression of a gesture is cultural dependent but that it also depends on the social context defined by the personalities and cognitive capacities of the conversants as well as on the personal and role relationship existing between them. We adopt such a view here. A gesture is finally chosen not only from the 'cultural' and 'personal' dictionaries but by also considering the social context.

2.2 Style parameters

We wish to specify a language which can express all factors contributing to the non-verbal style. We identify two kinds of parameters:

1. High-level agent defining parameters
2. Low-level gesture modifying parameters

The **high-level agent defining parameters** are used to determine the gesture dictionary to be used (see Table 1). These parameters specify individual gesturing characteristics, personality, the social, cultural and biological aspects of the agent. They are static characteristics and they may be specified only once when the agent is selected. The state characteristics define time-varying physical, mental and emotional aspects. The first two aspects may be set for a certain time interval, while emotional aspects may be triggered by an event, an action or people [14] and whose display may depend on the context [19].

The final gestures, that is the characteristics of the generated motions, can be defined by **low-level gesture modifying parameters**. On one hand, the gestur-

ing parameters (per modality such as hand gesticulation) define characteristics of the gestures of that modality; that is they are defined globally and represent a general characteristic of the agent. On the other hand, the gesture-modifying parameters may refer to a single gesture, or to all gestures of a modality within a time interval; that is they have a local influence and correspond to a dynamic aspect of the agent gesturing. The parameters are to be given in an XML-compliant Markup Language, basically to be used to annotate text to be spoken by the ECA. The markup language is described in the next section.

In general, communicative gestures accompany speech, and are synchronized with speech. The actual frequency and timing of some repetitive gestures (e.g. blink) can be generated on the basis of biological characteristics, while others (e.g. for idle motion) may be defined on the basis of personality characteristics (e.g. if the character is nervous, he will make more frequent idle motions) or by hand.

3 Generating styled gestures

The generation of styled gestures takes a text to be spoken by the character, annotated by two types of tags. One set of tags refers to the communicative functions that accompany the text to be spoken; the other set defines the style of the agent and it influences the selection and fine-tuning of the gesture to express the communicative gestures. Both set of tags are represented by an XML-compliant Markup Language.

As noted before, the tags related to the communicative functions may be generated automatically or placed manually. We are using the APML markup language [5], developed in the framework of the European project, MagiCster[1]. The APML tags are generated automatically by the multi-modal dialog generator module of MagiCster. We are currently developing a Style Mark-up Language (SML) for the tags related to style. For the moment these tags are specified manually. Note that formally the only requirement for the APML and SML is that the gesture dictionary assigns some gesture choice to each communication functions known to the APML. The styled gestures, corresponding to an APML tag, are generated by the by the following steps:

1. Dictionary update.
2. Selection of gesture to be used.
3. Generation of expressive gesture instance.

[1] IST project IST-1999-29078, partners: University of Edinburgh, Division of Informatics; DFKI, Intelligent User Interfaces Department; Swedish Institute of Computer Science; University of Bari, Dipartimento di Informatica; University of Rome, Dipartimento di Informatica e Sistemistica; AvartarME

In the **dictionary update** stage, it is checked which SML parameters have been given that have effect on the dictionary, and the current dictionary is updated. The effect of the static character-defining parameters is decided once, at the beginning of the processing of the marked up text. The probabilities in the resulting dictionary may be further modified at certain points of the text, by dynamical parameters, which have influence on style only for a subset of the text.

The generation of the current dictionary may be based on simple transparent principles. The simplest scenario is, when a hierarchy of parameters is established, and the parameters modify the current lexicon in a sequential order of hierarchy. E.g. a personal lexicon extends/overwrites cultural lexicon. The effect of state parameters can be computed locally, independent of other parameters. E.g. if the character is sad, his gesturing will be less frequent and intense. Cultural-dependent gesture will be set up. Expressiveness of gesture will depend on factors such as personality, age, emotion. Gesture type will be selected considering social, cultural and profession aspects. We are still working on the mapping of high-level tags to the choice of gesture type and its expressiveness. We will also have to face the problem of conflict that may arise due to the setting of different tags. For example, what would be the type of gestures and their expressiveness for an old person whose personality is introvert and that has the tendency to dream awake but that he currently excited due to a good news that just happened? By more complex reasoning it possible to take into account factors like situation-dependent or personality-dependent hierarchy of gesturing codes. E.g. a polite person may confirm to gesturing code expected in a situation (formal English dinner), but in the case of an individual person the personal gesturing code may be dominant in all circumstances. Also, it will depend on the culture if sadness will result in decrease or just, increase of gesturing (according to different mourning habits of cultures).

In the **gesture selection stage**, for each prescribed communicative function, a gesture will be selected to convey it. The selection will be based on the applicable entry of the current dictionary, but also taking into account possible modality conflicts [15]. In case of conflict, the selection from the alternatives in the current dictionary entry should be restricted to gestures with use free modalities (i.e. modalities not in used right now) [23]. If there is not such gestures, the conflict could be handled by resource-allocation [15]. An example of this type of conflict is the agent shaking the head to mean ‘no’ (gesture chosen from the cultural setting), but who also wants to emphasized her say. As head movement is already in used, choosing ‘head nod’ signal to denote ‘emphasis’ gives rise to a another entry from the agent’s gesture dictionary (say an eyebrow raising), the conflict does not subsist any

Tag	attributes	Possible values	M	P	G	E
Physical characteristic	Body	Body model file				
	Gesture repertoire	Directory of gestures			X	
	Dictionary	Dictionary file	X	X		
	Motion_manner	Smooth, hectic, angular…				X
	Handedness	Right, left	X	X		
	Hand_gesticulation	Intense, normal , minimal		X		X
	Face_gesticulation	Intense, normal, minimal		X		X
	Eye_gaze	Fixing, saccade, avoiding, neutral		X		
Personality_traits	Personality	Extravert, introvert, assertive, insecur		X		X
Social_aspects	Culture	Italian, British, American, …	X	X	X	
	Profession	Teacher, salesman,	X	X	X	
Biological_aspects	Age	Old, middle, young, child			X	X
	Gender	Female, male		X	X	X
State	Physical	Tired, sleepy, vivid, relaxed…		X		X
	Mental	Thinking, dreaming …		X		X
	Emotional	Angry, happy, …		X		X

Table 1. Character-defining parameters. M=mapping of communicative functions to gestures, P=probabilities of gestures assigned to a single communicative function, G=gesture repertoire, E=expressiveness of gestures

tag	attributes	Possible values
Gesturing	Intensity	Minimal, normal, intense
	Velocity	Fast, slow, normal
	Manner	Smooth, hectic, angular…
Timing	Gesture	Preparation phase, stroke, hold, end-phrasing
	Face	Onset, apex, offset
	Gaze	Start, end
	Start	Absolute time

Table 2. Gesture defining parameters. They decide the expressiveness of the gesture to be generated.

longer. But if this is not possible, one of the two conflict. If it is possible to express 'emphasis' through movements will have to prevail over the other ones. In the **expressive gesture generation stage**, all the static and dynamic style parameters having an effect on expressiveness of gesturing will be taken into account. E.g. an introvert person will make less articulated gestures, while a typical asymmetric eyebrow-usage will have an effect on all facial signals involving eyebrows. The effect of high-level character-defining parameters on expressiveness is given in terms of low-level gesture defining ones.

The on-the-fly generation of individual expressive gestures is based on our earlier work on the **gesture repertoire principle**: a gesture is defined in terms of constraints on control points of the parameter functions [23]. When style parameters are to be applied to the 'standard' definition of a gesture, they are expressed in terms of modifying certain constraints, e.g. the intensity of the smile will be limited as well as the speed of application and decay. The constraint framework allows the generation of different instances of a gesture, including random variants, as different solutions for the same constraint satisfaction problem.

It may happen that some modalities must be used for different gestures at the same time (like in case of speech and smile). For these modalities, a blend of the contributing gestures is to be produced, either as weighted sum of the contributing gesture parameter functions, or in a more sophisticated way, taking into account the constraints that should hold for the gestures.

4 Current status of research and further work

Up to recently, we have been working on generating facial expressions. We developed a detailed 3D realistic head model [16] and an interactive editor to specify facial expressions [13], also in terms of constraints. Using these tools and a taxonomy and descriptive data of facial expressions, we are currently implementing

the system for styled facial expressions. By the time of the workshop, we will have a demonstrator showing how different facial gesturing styles can be applied to a single talking head. We will also experiment with the effect of non-determinism within a style. We wish to test if the style of facial expressions of a synthetic face is perceived as intended.

There are several issues which need to be clarified before starting to implement a more full-fledged system. Fist of all, culturally different dictionaries need to be defined from psychological and sociological studies to find out what the distinctive gesture types and manners are for certain groups. It is also a question if the gesturing of cultures could be defined in terms of lower-level parameters, like every-day and social values, living conditions. In our outlined framework, it was only the style which had effect on the choice of the gestures. However, the characteristics of the environment (noise, visibility) and the listener should have an effect on the selection of modalities and expressiveness of gestures.

Ultimately, one would like to have an ECA which manifests style also in the verbal modality. There is ongoing work to generate styled NL content [2, 6, 26], and to reflect emotional and mental state in the generation of synthetic speech [25]. On a longer term, it is a challenging task to develop ECAs which have consistent style in all their modalities.

Acknowledgements

We thank Han Noot and Isabella Poggi for their inspiring remarks on the topic.

References

1. Aardman Studios, Creature Comforts, URL: http://atomfilms.shockwave.com/af/content/atom_221
2. Ball, G., Breese, J. Emotion and personality in a conversational agent, in 3, pp. 189-219.
3. Cassell J., Sullivan J., Prevo st S., Churchill E., (eds.) Embodied Conversational Agents, MIT Press, Cambridge, MA, 2000.
4. Chi D., Costa M., Zhao L., Badler N., The EMOTE Model for Effort and Shape, Proc. of Siggraph 2000, pp. 173-182.
5. De Carolis N., Pelachaud C., Poggi I., de Rosis F., Behavior planning for a reflexive agent, Proc. of IJCAI'01, 2001.
6. De Rosis, F., Grasso, F. Affective text generation, In: Paiva, A. (ed): Affective Interactions, LNAI vol. 1814. Springer-Verlag, 2000.
7. Ekman, P., Facial Expressions. In Dalgleish T. and Power T. (eds.), The Handbook of Cognition and Emotion. John Wiley & Sons, Ltd, 301-320, 1999.
8. Gratch J., Marsella S., Tears and Fears: Modeling emotions and emotional behaviors in synthetic agents, Proc. of AA 2001, pp. 278-285. 2001.
9. Kendon A., Human gesture, In Ingold T. and Gibson K. (eds): Tools, Language and Intelligence, Cambridge University Press, 1993.
10. Marsella S., Personal communication.
11. McNeill D., *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press, 1992.
12. Nass C., Isbister K., Lee E-J., Truth is beauty: Researching embodied conversational agents, in [4], pp. 374-402.
13. Noot H., Ruttkay Zs., CharToon 2.0 Manual, CWI Report INS-R0004, Amsterdam, 2000.
14. Ortony A., Clore G.L. and Collins A., *The cognitive structure of emotions*. Cambridge University Press, 1988.
15. Pelachaud, C. Poggi I., Subtleties of Facial Expressions in Embodied Agents, Journal of Visualization and Computer Animation, to appear.
16. Pelachaud C., Magno-Caldognetto E., Zmarich C., Cosi P., An Approach to an Italian Head, Proc. of Eurospeech'01, 2001.
17. Perlin K., Real time responsive animation with personality, IEEE Transactions on Visualization and Computer Graphics, Vol. 1. No. 1. 1995.
18. Poggi I., Mind Markers. In Mueller C. and Posner R.(eds): The Semantics and Pragmatics of Everyday Gestures. Berlin Verlag Arno Spitz, 2001.
19. Poggi I., Pelachaud C. and de Rosis F., Eye communication in a conversational 3D synthetic agent. *AI Communications, Special Issue on Behavior Planning for Life-Like Characters and Avatars*, 13, 3, 169-181, 2000.
20. Prendinger H., Ishizuka M., Social role awareness in animated agents, Proc. of AA 2001, pp. 270-277. 2001.
21. Queneau R., Wright B. (Translator): Exercises in Style, New Directions 1981.
22. Ruttkay Zs., Noot H., FESINC: Facial Expression Sculpturing with INterval Constraints, Proc. of AA'01 Ws on Multimodal Communication and Context in Embodied Agents, 2001.
23. Thòrisson K.R., Layered modular action control for communicative humanoids, Computer Animation'97, IEEE Computer Society Press, p. 134-143, 1997.
24. Thomas F., Johnston O., Disney animation: The illusion of life, Abbeville Press, New York,1981.
25. Virtual Human Markup Language (VHML) http://www.vhml.org
26. Walker, M., Cahn, J., Whittaker, S. Improvising linguistic style: Social and affective bases for agent personality, Proc. of Autonomous Agents 1997.