Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Int. J. Human-Computer Studies 62 (2005) 211–229
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

# Variations in gesturing and speech by GESTYLE

Han Noot[a,*], Zsófia Ruttkay[b]

[a] *Center for Mathematics and Computer Science, INS, Kruislaan 413, 1090 GB Amsterdam, The Netherlands*
[b] *University of Twente, EWi-HMI, P.O. Box 217, 7500 AE Enschede, The Netherlands*

**Abstract**

Humans tend to attribute human qualities to computers. It is expected that people, when using their natural communicational skills, can perform cognitive tasks with computers in a more enjoyable and effective way. For these reasons, human-like embodied conversational agents (ECAs) as components of user interfaces have received a lot of attention. It has been shown that the style of the agent's look and behaviour strongly influences the user's attitude. In this paper we discuss our GESTYLE language making it possible to endow ECAs with style. Style is defined in terms of when and how the ECA uses certain gestures, and how it modulates its speech (e.g. to indicate emphasis or sadness). There are also GESTYLE tags to annotate text, which has to be uttered by an ECA to prescribe the usage of hand, head and facial gestures accompanying the speech in order to augment the communication. The annotation ranges from direct, low level (e.g. perform a specific gesture) to indirect, high level (e.g. take turn in a conversation) instructions, which will be interpreted with respect to the style defined. Using style dictionaries and defining different aspects like age and culture of an ECA, it is possible to tune the behaviour of an ECA to suit a given user or target group the best.


*Keywords:* Embodied conversational agent; Multimodal communication; Style; Mark up language

*Corresponding author. Tel.: +31 20 5924330; fax: +31 20 5924199.
*E-mail addresses:* han@cwi.nl (H. Noot), z.m.ruttkay@ewi.utwente.nl (Z. Ruttkay).

1071-5819/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2004.11.007

## 1. Introduction

### 1.1. About ECAs

Empirical studies suggest that users respond to complex interactive devices as they respond to humans (Reeves and Nass, 1996). This has given rise to the CASA (Computers Are Social Actors) paradigm (Nass et al., 1994). A striking example of such behaviour is reported in Nass (2004). In the experiments, computer software was used for explanatory purposes. After a session with the system, the users were asked to evaluate the software by answering a series of questions. It mattered whether both tasks did run on the same or on different computers. Users rated the service better if they had to do the evaluation on the same computer, which had been used to help them. This parallels the tendency that in every-day life, because of politeness, we express less criticism about a service directly to the person who assisted, than to somebody else. So users expressed politeness towards the computer.

On the other hand, for people the natural way of communicating is speech, accompanied by subtle gestures, facial expressions and postures. These two observations gave rise to human-like characters, so called embodied conversational agents (ECAs) in man–machine communication. It is expected that people, when using their natural communicational skills, can perform cognitive tasks with computers in a more enjoyable and effective way.

An ECA is some creature which resides on the computer screen, which resembles a living creature in look and behaviour, and assists the user in the task at hand (Cassell et al., 2000). Most often human-like characters are used, but agents with embodiments as animals (Isbister et al., 2000) or even animated objects (Microsoft's paperclip) do occur. When utilizing ECAs, many design questions and evaluation issues need to be taken care of (Massaro et al., 2002; Ruttkay et al., 2004, to appear). We mention only a few: How should the ECA look like: 2D or 3D, realistic or cartoon like, what gender and culture does it have, should it posses a complete body or only have a (talking) face? How should it be dressed? What should be its communicative abilities? Does it indicate turn giving/taking, does it show idling behaviour (blinking, drumming its fingers), does it display emotions? What nonverbal signals are used to indicate these states? What are the motion characteristics of the gestures? Is the ECAs nonverbal behaviour fully repetitive, or are some variances possible? Can it adapt to the (static or changing) characteristics of a specific user?

The believability of ECAs highly depends on their nonverbal communicational skills: the richness of the used modalities and gestures, and the correctness and consistency of choosing and performing a gesture. Different persons, depending on their cultural, social and professional background and their personality, use different gestures or exploit different modalities in the same situations while communicating (McNeill, 1991; Kendon, 1993). Also, there is evidence that the user's response to the ECA depends on subtle characteristics like ethnicity and personality of the ECA (Walker et al., 1994; Nass et al., 2002). In general, it seems that the ECA should resemble the user in order to be appreciated most. For instance, the virtual real estate

agent (REA) (Cassell et al., 1999), when it engaged in small talk, induced more trust in users who were extroverts but less in those who were introverts. McBreen et al. (2001) compared the use of formal and casual agents in various retail applications (travel agents, cinema ticket sail and a banking application). They found that for the cinema ticket application an informal agent was preferred and for the banking one a formal agent.

The typical way of one's communication (manifested in facial and hand gestures used, motion and speech characteristics, word choice, etc.) is referred to in everyday language as one's style. The quoted experiments above indicate that:

- style matters,
- what style is to be preferred (for an ECA) depends on characteristics of both the user and the task to be performed.

If we wish to avoid designing a new ECA for every different situation, the desire for ECAs with adjustable gesturing and speech style arises naturally. We have developed GESTYLE for that reason.

In Chapter 2, we first shortly discuss the key concepts of style and gestures. Then in Chapter 3, we introduce GESTYLE and in Chapter 4 highlight the usage of its key constructs by means of examples. Finally, we relate our work to other similar efforts, discuss some experiments and outline further work. An extensive overview (including syntax) of the nonverbal aspects of GESTYLE is given in (Noot and Ruttkay, 2004), while the speech part is covered more extensively in Van Moppes (2002). The concept of style as used here is discussed in Ruttkay et al. (2004, to appear).

## 2. Style in gestures

### 2.1. About style

In its most general form, *style* is manifested in clothing (formal/informal), choice of language (polite/casual), gesturing motion characteristics (expansive/subdued), gesturing frequency and gesturing repertoire (e.g. what to do to signal a greeting), characteristics of speech (rate, intonation, volume), etc. The strict content of the information exchanged is conveyed primarily by what is said. Style 'colours', modulates or augments the verbatim content by the nonverbal and meta-speech signals. However, style is not only to make the conversation more varied and joyful. More importantly, the information expressed by the style tells about the personality, culture or current emotional state of the speaker, his relationship to the interlocutor (boss/employee), and the conversational situation (formal/informal).

The separate manifestations of style are not independent; they are determined by decisive aspects of the person. For instance, culture (in the anthropological sense) influences both 'the looks' and ways of behaving. Typically, an Italian man is well-dressed, has dark hair, talks fast and much, uses hand gestures extensively, several of which are specifically 'Italian', like the hand shape with joined finger tips to start

conversation. So when modelling believable style, consistency should be taken care of, and the aspects which determine or influence the style should be identified. Such further aspects are gender (e.g. postures which are acceptable for one sex but not for the opposite one in a given culture), age (e.g. more fragile movements, different voice characteristics when age progresses), profession (e.g. different motion characteristics for a wood-cutter or a brain-surgeon, also when not on their jobs), etc. In Ruttkay et al. (2004, to appear) we gave an in-depth discussion of decisive aspects of style, related to sociological and behavioural studies.

The specification of the style of ECAs, similar to those of humans, would require that the multitude of aspects and phenomena of human–human communication have been described in a normative way and with a formal precision matching the design parameters of ECAs. There are excellent scientific results for restricted topics (for instance, Kendon, 1993) and practical guidelines like ones for employees of international organisations sent to areas with different cultures. But, unfortunately, there are not enough sources from the fields of social psychology, sociology, cultural anthropology and psycho-linguistics to rely upon for a complete description of, for instance, how a tutor should look like, talk and gesture, given an application domain and a target group of users. Actually, the introduction of ECAs has motivated research in human–human communication too, by posing new, crisply formulated questions, some of which could be answered only by using ECAs as controllable mediums to perform the effects to be tested (see e.g. Krahmer et al., 2003). Thus, our aim is not to present a collection of some pre-cooked styles. What we do rather, is to provide a framework to experiment with ECAs exhibiting style. This framework can be used to easily define styles for ECAs and make them behave accordingly. It can encompass findings form social psychology and other research when it becomes available and meanwhile (or complementary) use knowledge derived from artistic impressions of style as well. Once a style has been defined for an ECA, evaluation experiments can be conducted to investigate the style's believability and its effect on users.

The framework we propose is the GESTYLE language, designed to define style in terms of multimodal behaviour and make an ECA gesture and talk accordingly, with variations in usage of modalities and specific gestures and in subtleties of motion and speech.

To avoid confusion we explicitly note that our concept of style addresses a different domain than the style concept as exemplified in style sheets used in many web applications (e.g. as in XSL or CSS). The latter style deals with *document* style, our style concept refers to the style of *ECA behaviour*. On the other hand, we borrow heavily from the latter field by using the XML/XSL apparatus.

### 2.2. About gestures

In our discussion, we use the concept *gesture* in a very broad sense, for every nonverbal signal with a communicative content. So a facial expression (showing happiness or puzzlement), a hand gesture (like beat to emphasize something, or a hand indicating the size of an object) or a 'thinking' eye-gaze are all gestures using a

single nonverbal modality. Multi-modal gestures use two or more of the nonverbal modalities in a coordinated way, to express a single meaning. For instance, emphasis can be addressed by nodding and a hand beat, as well as by eyebrow-raise, looking at the interlocutor and a beat.

As of the role of gestures in human–human communication, the following functions have been identified (Cassell, 2002):

- Some gestures support prosody and speech punctuation, and hence may *increase the intelligibility of speech*. Common examples are: indication of contrasting or enumerated chunks of information, or a new topic, by hand gestures like 'on the one hand, on the other hand', beats with hand shapes showing numbers to underline enumeration, or signalling importance and emphasis by eyebrows and gaze.
- Gestures can be used to *augment or disambiguate speech*, by indicating location, size, shape and other characteristics of items referred to. "Give me *that* book", accompanied by a pointing arm/hand gesture or in "A small piece of cake for me!", accompanied by indicating size with thumb and index.
- Emblematic gestures *represent concepts or acts* without words and are often much culture-dependent. Think of the different ways of showing victory in the USA and UK.
- Gestures can be indicators of the speaker's *emotional and cognitive state*. E. g. sadness can be expressed by a sad facial expression, head bent down, and the slowness of movements in general, also of hand gestures with other functions.
- Gestures play an important role in *regulating dialogues*. E.g. the intention of turn-taking can be signalled by looking at the interlocutor only, or by also making special hand gestures and/or changing posture.

Note that some gestures assume speech (speech punctuation), but most of them can be also used without speech. GESTYLE is designed to engineer the manifestation of style in the nonverbal and meta-speech characteristics (like intonation, tempo).

We are aware of the fact that besides the gestures, style manifests itself in other characteristics. Visual aspects like clothing, male/female face and body, smooth- or wrinkled face, etc. are not addressed in GESTYLE. Characteristics of the content and structure of one's speech—usage of lexical structures and choosing from synonyms, and even in organizing monologues (a 'talkative' person may be redundant and detailed in his speech) and dialogues (interruption may be a sign of temporary excitement, or a permanent characteristic due to one's extrovert and dominant personality)—are also reminiscent of style, though beyond the scope of GESTYLE.

## 3. GESTYLE in a nutshell

Let us assume, we have a piece of written text, which we want the ECA to utter, such as: "I am offended by your behaviour. I never expected this kind of disloyalty

from you''[1]. We might want the ECA to utter this text with an angry voice, point at its interlocutor on the word ''your'' and put emphasis (both by prosody and by some gestures) on the word ''never''.

In general, we want to indicate parts of the text which the ECA must emphasize, segments where a greeting, refusal or characterisation of some object should be accompanied by some gestures, locations (in the text) where the ECA's emotional state changes, where the ECA wants to take/give turn, etc. GESTYLE can be used to do this with the help of mark up tags embracing pieces of text. For the more computer science minded: GESTYLE is an XML (Extensible Markup Language) compliant text mark up language. This implies that it conforms to a standard for the World Wide Web and hence it can be used with (sufficiently powerful) web browsers.

Besides from being a text mark up language, GESTYLE contains also facilities for defining how the above-mentioned communicative acts, indicated by text mark up, should be performed. For instance, what gesturing is to be done when there is a greeting, how to speak when an angry voice is called for, etc. Finally, the prescriptions on how to express communicative acts are to be given in different dictionaries (style dictionaries), according to a style of some aspect. Alternative dictionaries for one aspect are to be used exclusively (e.g. the ECA should act either as a male or as a female), while different aspects may be given to indicate different 'sources' to determine the final style (the ECA should act as a Dutch male teacher). This feature of defining and using multiple dictionaries is the essential feature of GESTYLE's support for defining style. GESTYLE also makes possible variations in the nonverbal presentation, by allowing alternatives, choices for modality usage and subtle fine-tuning of motion characteristics.

Summing up, GESTYLE can be used to *define* style and to instruct the ECA to *express* some meaning by speech characteristics and also in a nonverbal way in accordance with that style. It makes it easy to generate different presentations by the same ECA, or author the presentation style for a new one. What the effect might be is shown in Fig. 1 below.

The language constructs of GESTYLE are organized in a structured way. At the atomic level there are so-called *basic gestures* (e.g. right-hand beat, nod). Basic gestures can be combined (parallel or sequentially) into *composite gestures* (e.g. two-hand beat, right-hand beat and nod) by *gesture* expressions. All gestures, which can be used by GESTYLE must be elements of one single *gesture dictionary*. Gestures can be fine-tuned by specifying duration, intensity, smoothness and other *motion manner parameters*. At the next level, the *meanings* denote the communicative acts (e.g. show happiness, take turn in a conversation) that can be expressed using gestures. The *meaning mapping definitions* contain the mappings of meanings to alternatives of (usually composite) gestures or gesture expressions (The gestures referred to must be entries in the gesture dictionary). From the alternatives one will

[1]This, as well as the other examples in the paper are not meant to represent utterances from state of the art ECA applications, but rather pieces of texts from real-life dialogues. This is in line with the goal of our work, to bring the ECA's communication skills closer to those of humans.

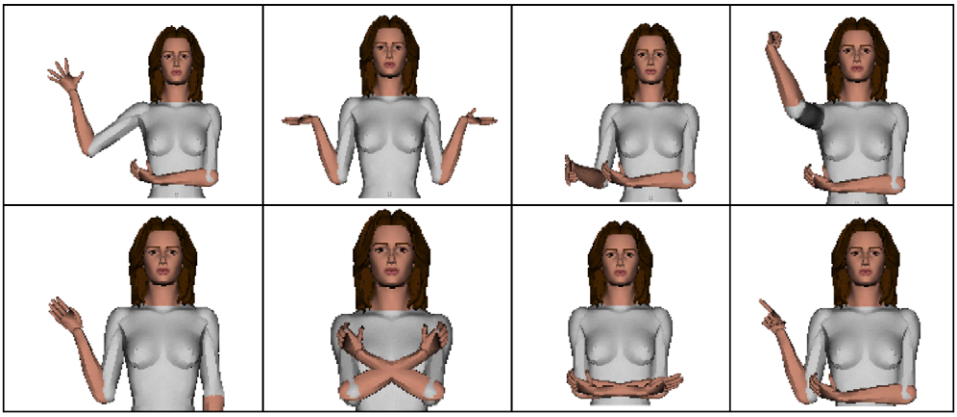

Fig. 1. The communicative acts 'Greeting', 'SorryFor', 'Enumerate', and 'Emphasize' performed by the same humanoid in extrovert and introvert style. The stills shown are taken from a longer demonstrator for GESTYLE which renders in VRML and is programmed in the STEP system Eliens et al. (2002).

be chosen according to their given probabilities, thus giving rise to dynamic variation. The meaning mapping definitions are collected as entries in *style dictionaries*. A style dictionary contains a collection of meanings pertinent to a certain style (e.g. a style dictionary for 'teacher', 'Dutchman', etc.). Style dictionaries can refer to *speech style definitions*, which give details on verbal style. Finally there is the (static) *style declaration*, which specifies the style of the ECA. A style is declared by specifying which style dictionaries to use and how to combine them.

For temporary style changes there are *dynamical modifiers* and *affect states*. They do not, by themselves, produce gestures, but change the way meanings and gestures—occurring within their scope—are performed. For instance, motions may get faster and increase in amplitude, due to some modifiers inserted, or due to bringing the ECA into e.g. an angry affect state. Modifiers are wired into GESTYLE. Affect states have (just like meanings) affect state mapping definitions, which are to be given in the style dictionary. Modifiers operate at a low level, comparable to that of gestures. Affect states are high level; their level is comparable to that of meanings. To sum up, there are four XML documents involved in GESTYLE:

1. The marked up text to be spoken (annotated with style declarations, meanings and possibly gestures). Note: there may also be annotation without text, to cover the silent but gesturing ECA.
2. The style definition file, containing the style dictionaries.
3. The gesture dictionary document, containing the gestures, which can be used in 1 and 2.
4. The speech style definition files, containing definitions of different speech styles.

The main power of GESTYLE is that, in order to instruct the ECA to present a content with a different style, *only the style declaration* in document 1 has to be changed. (Of course, the dictionaries for the different style must have been defined already.)

The overview of the modules relevant to GESTYLE are shown in Fig. 2. The input, that is a piece of text annotated with GESTYLE mark up tags, can be prepared 'by hand'. But also, such a marked-up text can be the output of some

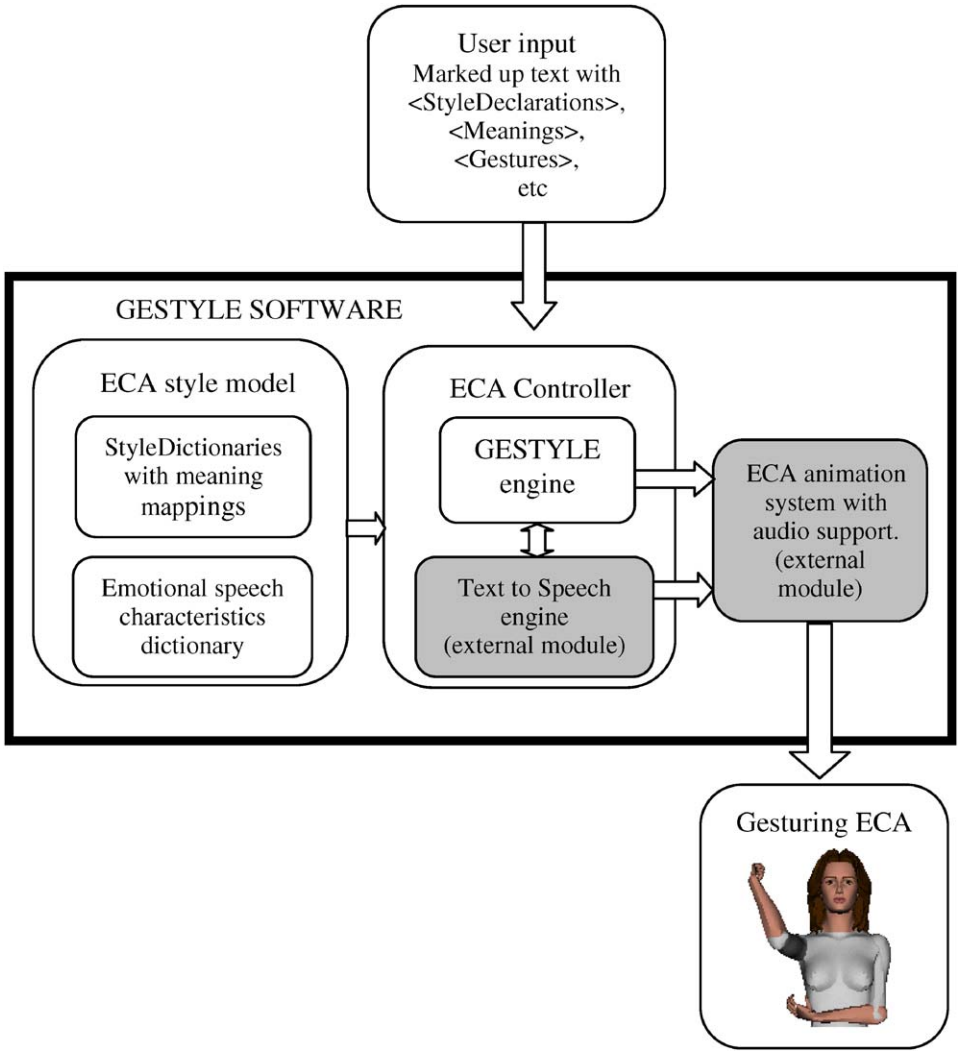


Fig. 2. Overview of the internal and external modules of GESTYLE.

software module (representing the 'brain' of the ECA). Such a module may be a simple monitor (think of the expressions used by MS PaperClip to indicate states of processing), or some complex reasoning module, relying on one or other psychological models of motivations, plans and emotional states. In all cases the decision of what to be said and what meaning to be communicated, is outside the scope of GESTYLE.

There are two modules outside GESTYLE, which are necessary to generate the final gesturing and talking ECA. The ECA animation module should be prepared to generate the basic gestures finally prescribed, as a result of GESTYLE interpreting the mark up tags. In principle, a variety of animation modules can be interfaced to GESTYLE, assuming that they can perform basic gestures (as opposed to systems operating on a key-frame based animation principle). To generate the speech, a TTS engine is needed which is capable to manipulate the characteristics of phonemes according to the speech characteristics used in GESTYLE. While the characteristics are common to most of the TTS engines, not all of them allow access to phonemes. In Section 5.2, we discuss the external modules we have been using with GESTYLE.

## 4. GESTYLE by examples

GESTYLE in principle can be used to control ECAs with a single modality (speech only, hand-gestures only) as well as ones with speech and all the nonverbal modalities (visual speech, facial expression, gaze, head movement, hand gestures, body posture). For our following examples we assume that we have a multimodal ECA, which can speak, gesture by hand and show facial expressions.

### *4.1. Using style to present meaning*

In Example 1, we see a piece of annotated input text ("Hello. I am a gesturing avatar…"), beginning with a StyleDeclaration (line 2 till 6). This StyleDeclaration states that the ECA should behave with the style of an extravert person because we set the value of the personality style aspect to 'Extravert'. (One could have defined other aspects of the ECA, see Section 4.3. Note that the aspect here is a syntactical element, in Section 5.2 we discuss the issues of semantics.) We will come back to StyleDeclarations in Section 4.2, but here already we emphasize the heart of the style concept: by only changing the value of dict to 'Introvert', the same script would be acted out by the ECA in the style of an introvert, with drastic changes in its nonverbal behaviour and speech qualities. After the StyleDeclaration there follows a TextBody. It contains the text, which has to be spoken by the ECA, accompanied by nonverbal behaviours. The most important mark up within a TextBody are the Meaning tags. These denote communicative acts like 'show happiness' (line 8), 'emphasize strongly' (line 9) or mildly (line 14), etc.

The effect of Meanings is determined by the style of the ECA: whether the ECA uses many or few gestures and which ones, performs those with large or small

```
1   ⟨StyledText⟩
2     ⟨StyleDeclaration⟩
3    ⟨OrderedElements⟩
4           ⟨Style aspect="PERSONALITY" dict="Extravert" /⟩
5    ⟨/OrderedElements⟩
6     ⟨/StyleDeclaration⟩
7     ⟨TextBody⟩
8       ⟨Meaning name = "Happy"⟩
9         ⟨Meaningname=" EmphasizeStrong  "⟩ Hello ⟨/Meaning⟩
10        I am a gesturing avatar.
11      ⟨/Meaning⟩
12      ⟨Meaning name = "Sad"⟩
13          Sorry for
14           ⟨Meaning name = "EmphasizeMild"⟩ not ⟨/Meaning⟩
15        beingproperly dressed ...
16      ⟨/Meaning⟩
17        Do you allow
18        ⟨Gesture name="PointAtSelf" intensity="INTENSE"⟩ me ⟨/Gesture⟩
19     in?
20    ⟨/TextBody⟩
21  ⟨/StyledText⟩
```

**Example 1.** Text annotated with GESTYLE mark up tags.

movements, which voice modifications it uses, etc. What the effect might be is shown in Fig. 1, where two alternatives expressing the same Meaning are presented.

This example also shows that Meanings can be nested: form line 8 till line 11 the ECA behaves happy, e.g. by having a happy face and voice. Within the 'Happy' Meaning there is another one to emphasize the text 'Hello'. When the inner Meaning does not just temporarily replace the outer one (e.g. 'Sad' within 'Happy') it should be blended with the enveloping one (e.g. 'EmphasizeMild' within 'Happy').

The example illustrates one more feature: it is also possible to directly annotate with Gestures (line 18). In our case, the ECA is instructed to point at itself. The resulting behaviour is *not subject to style changes*. The same pointing behaviour could be achieved with the style-specific ⟨Meaning name = 'ReferToSelf'…/⟩ tag. Due to the extrovert style to be used, the corresponding pointing gesture will be intense. So the preferred place for using Gestures is in StyleDictionaries; in text input they are only to be used to prescribe specific gestures, for which the current style does not provide an equivalent Meaning.

### 4.2. *Defining styles*

StyleDictionaries contain the definitions of how Meanings are mapped onto Gestures, GestureExpressions and ExpressiveSpeech. In Example 2 a piece of the StyleDictionary of an extravert person is shown (line 1 till 18). At line 2 SpeechMode is set to extravert speech (e.g. high-pitched and fast speech). From lines 3 to 16 it is

specified how to enact Meaning 'Emphasize'. Within the Meaning there are two GestureSpec elements, starting on line 4 resp. 11. This indicates the choice between two ways of making the ECA show emphasize. GESTYLE will choose the first one with probability 0.7 (line 9) and the second one with probability 0.3 (line 14).

Finally let us have a closer look at the first alternative i.e. the GestureSpec at lines 4–10. At line 5, it is stated that there should be mild vocal emphasize. Then at lines 6–8 it is stated that a 'NodAndBeat' gesture should be performed in parallel (the ⟨PAR/⟩ element) with a 'LookAtPerson' gesture.

At line 19, the definition of a StyleDictionary for an Introvert person starts. Here, the Meaning 'Emphasize' is defined too (line 21 till 26), but in this introvert case only as an 'EyebrowRaise' without the gestures of the extravert case.

In the example, the speech characteristics for the style are set too. First we specify a speech style definition (in its function comparable to a StyleDictionary, but now for speech) by the SpeechMode element on line 2. This specification holds for the whole StyleDictionary. It refers to the definition of the actual speech types like 'emph_strong' and 'emph_mild' used in the ExpressiveSpeech elements on lines 5 and 12. For more on speech style definitions, see Section 4.5.

```
1   ⟨StyleDictionary name = "Extravert"⟩
2     ⟨SpeechMode name="ExtravertSpeechMode"/⟩
3     ⟨Meaning name = "Emphasize" CombinationMode = "DOMINANT"⟩
4       ⟨GestureSpec⟩
5         ⟨ExpressiveSpeechemotion="emph_mild"/⟩
6         ⟨U seGest name="NodAndBeat"/⟩
7         ⟨PAR/⟩
8         ⟨UseGestName=”LookAtPerson”/⟩
9         ⟨ProbabilityP="0.7"/⟩
10      ⟨/GestureSpec⟩
11      ⟨GestureSpec⟩
12          ⟨ExpressiveSpeechemotion="emph_strong"/⟩
13          ⟨UseGest name="Beat"/⟩
14          ⟨P robability P="0.3"/⟩
15      ⟨/G estureSpec⟩
16    ⟨/Meaning⟩
18        *
19    ⟨/ StyleDictionary ⟩
20    ⟨ StyleDictionaryname = "Introvert"⟩
21      ⟨Meaning name = "Emphasize" CombinationMode = "COMBINE"⟩
22        ⟨GestureSpec⟩
23          ⟨UseGestname="EyebrowRaise"/⟩
24          ⟨ProbabilityP="1"/⟩
25        ⟨/GestureSpec⟩
26      ⟨/Meaning⟩
27          *
28    ⟨/StyleDictionary⟩
```

**Example 2.** StyleDictionary definition.

```
1   ⟨StyleDeclaration⟩
2       ⟨weightedelements⟩
3           ⟨Style aspect="SOCIAL STATUS" dict="SimplePerson"  weight = "2"/⟩
4           ⟨Style aspect="CULTURE" dict="Brazilian"  weight = "1"/⟩
5       ⟨/weightedelements⟩
6       ⟨orderedelements⟩
7           ⟨style aspect="PROFESSION" dict ="Farmer" /⟩
8           ⟨style aspect="GENDER" dict ="Male"/⟩
9       ⟨/ordered elements⟩
10  ⟨/StyleDeclaration⟩
```

**Example 3.** A StyleDeclaration.

### 4.3. Adding up styles

In Example 3 below, a complex StyleDeclaration is shown. There are style dictionaries dealing with style aspects of 'SOCIAL STATUS', 'CULTURE', 'PROFESION' and 'GENDER', where some are in the 'weighted elements' group, others in the 'ordered elements'. There are two principles to guide how StyleDictionaries add up:

- The style of the ECA (and hence the Meanings it can perform) is primarily governed by one style aspect (e.g. profession), but additional Meanings are available because of an other style aspect (e.g. culture).
- A Meaning can be expressed (i.e. mapped to Gestures ) in various ways because of the different style aspects of the ECA.

More precisely, when a Meaning has to be processed, the ordered elements of the StyleDeclaration are searched first. As soon as a definition for it is encountered, that one is used. If no definition occurs there, the weighted elements are inspected. If the Meaning is encountered more than once there, its various definitions are combined according to the values of the weight attributes.

To sum up, StyleDeclarations change by including references to different StyleDictionaries or by modifying the order or weights of the used StyleDictionaries. In Example 3 the profession dominates the gender, but interchanging lines 7 and 8 would result in the opposite.

### 4.4. Dynamical changes to style

The style of the ECA may change in course of time, due to change in its emotional or physical state, or changes in the situation. E.g. if excited, the ECA tends to use more expressive gestures, and even the dominance of styles may change. For instance, it may "forget" that it is in a public space where the "public social" style is expected, and will start to use its own personal, informal style.

To take care of style changes, the *dynamical modifiers* occur interwoven with the text. We allow three types of low-level modifiers, for indicating changes in:

- 'respecting' different styles,
- usage of modalities, and
- motion manner characteristics of the gestures.

At a higher level we have the *affect state*, which may, according to its definition, indicate changes in all the three items above plus emotional speech characteristics.

Generally speaking, the modifiers change some parameters of the static style declaration and of the meaning definitions in style dictionaries.

*4.4.1. Basic dynamical modifiers*

An ECA's gesture repertoire may change according to the situation. E.g. if the listener turns out to be a superior of the speaker, the speaker will probably adjust its style to more polite. But if it gets very angry, it may fall back to its own, less polite style.

In order to handle such situations, GESTYLE allows to swap two elements (StyleDictionaries) of the static StyleDeclaration, or to change their weights.

In Example 4a, there is an ECA speaking whose style is declared as a diplomat with an extravert personality (lines 3 and 4). We suppose it to be involved in a professional discussion, getting annoyed and switching to a personal, informal style. This is accomplished by the DominanceModifier in line 8, which puts the personal aspect before the professional one.

In Example 4b, we have an ECA speaker who reflects upon the turn the discourse has taken and decides to behave more quietly in order to de-escalate. To that end, it decides to use less hand gestures, accomplished by the ModalityUsage directive on line 2. The modalities to be changed are out of the set of values discussed in Section 2.

```
1   ⟨StyleDeclaration⟩
2       ⟨orderedelements⟩
3           ⟨Style aspect="PROFESSION" dict="Diplomat"/⟩
4           ⟨Style aspect="PERSONAL" dict="Extrovert"/⟩
5       ⟨/ordered elements⟩
6   ⟨/StyleDeclaration⟩
7   Let's stop this formal behavior and get to the essence of the matter!
8   ⟨DominanceModifier aspect="PERSONAL" putbefore="PROFESSION"/⟩
9   As I see it, your arguments are just rationalizations, all you want is just have it your way.
```

**Example 4a.** Usage of dominance modifier.

```
1  OK, This agitatedconversation leadsusnowhere, let's relax  and reconsider the situation
2  ⟨ModalityUsage hands="-40%"/⟩
3  ⟨MannerDefinition modality="HANDS" intensity="-30%" modality ="FACE" intensity="-10%"/⟩
4  So, correct me if I am wrong, but to me your point seems to be….
```

**Example 4b.** ModalityUsage and MannerDefinition.

```
1   ⟨StyleDictionaryname=…..⟩
2     ⟨AffectState name="Angry"⟩
3       ⟨ExpressiveSpeech emotion="Angry"/⟩
4       ⟨MannerDefinition modality="HANDS" intensity="intense" motion_manner="JERKY" /⟩
5       ⟨ModalityUsage modality="HANDS" value="+15%"/⟩
6       ⟨Dominance Modifier aspect="PERSONAL" putbefore="PROFESSION"/⟩
7     ⟨/AffectState⟩
8     ****
9   ⟨StyleDictionary/⟩
```

**Example 5a.** An AffectState entry in a StyleDictionary.

```
1   ⟨AffectState name="Angry"⟩
2       I told youthis for the
3       ⟨Meaningname="Emphasize"ñthird  á/Meaning⟩
4        timenow. I am wonderingifyou have beenlistening
5       ⟨Meaning name="Emphasize"ñ at all. á/Meaning⟩.
6   ⟨/AffectState⟩
7   "More text, nolongerangry ….."
```

**Example 5b.** Using the AffectState markup directive.

#### *4.4.2. Affect state*

AffectStates are to be defined as entries in a StyleDictionary where they are given a name (see Example 5a, lines 2–7). Using this name, they can be used for text markup (see Example 5b, lines 1–6).

The result of the use of the AffectState in Example 5b is that the text on lines 2–5 is uttered using angry speech, the personal style will dominate the professional one and thus a more intense and bigger number of gestures will be used when text is emphasized (5b, lines 3 and 5). Finally, after line 6 all changes brought about by the AffectState disappear again.

### *4.5. Expressive speech*

The last aspect of style we discuss is expressive speech. Meanings may also influence speech properties, like raising pitch level for emphasis. There are speech style definition data sets (see Fig. 2.), which contain SpeechExpression-Definitions as shown in Example 6. The 'properties' element in line 2 prescibes to change speech quality on the global level, by adjusting speed, pitch rate and range (lines 3–5). The 'phoneme_level' element in line 7 prescribes changes of 'micro' properties on the phoneme level, indicating that stressed vowels should last longer (line 8) and final vowels should get an increased pitch (line 9). StyleDictionaries refer to SpeechExpressionDefinitions via their names, as seen in Example 2, line 2. For more details on expressive speech, see Van Moppes (2002).

```
1   ⟨SpeechExpressionDefinition name="happy_speech"⟩
2     ⟨properties⟩
3       ⟨speed_rateñ20á/speed_rate⟩
4       ⟨pitch_rateñ30á/pitch_rate⟩
5       ⟨pitch_rangeñ50á/pitch_range⟩
6     ⟨/properties⟩
7     ⟨phoneme_level⟩
8       ⟨stress_vowel_durationñ20á/stress_vowel_duration⟩
9       ⟨final_vowel_pitch_incñ15á/final_vowel_pitch_inc⟩
10    ⟨/phoneme_level⟩
11  ⟨/SpeechExpressionDefinition⟩
```

**Example 6.** Definition of expressive speech style.

## 5. Discussion

### 5.1. Related work

The synthesis of hand gestures (Cassell, 1998; Chi et al., 2000; Hartmann et al., 2002; Kopp and Wachsmuth, 2000) and their role in multimodal presentation for different application domains (Lester et al., 1999) has gained much attention recently. Particularly, there have been XML-based mark up languages developed to script multimodal behaviour of ECAs, such as MPML (Tsutsui et al., 2000), VHML (Virtual Human Markup Language (VHML)), APML (De Carolis et al., 2002), RRL (Piwek et al., 2002), CML and AML (Arafa et al., 2002), and MURML (Krandsted et al., 2002). Each of these representation languages act either at the discourse and communicative functions level (APML, RRL, CML, MURML) using tags like 'belief-relation' or 'emphasis' or at the signal level (AML, VHML) with tags like 'smile', 'turn head left'. In each case the semantics of the control tags are given implicitly, expressed in terms of the parameters (MPEG-4 FAP or BAP, muscle contraction, joint angles and the like) used for generating the animation of the expressive facial- or hand gestures.

What also differentiates GESTYLE form the previous languages is that it covers both speech and nonverbal modalities. Moreover, GESTYLE does support not only a few predefined tags for emotional speech (as is often the case), but arbitrary emotional speech elements can be defined as part of the style (The TTS system employed should be able to cope with those definitions.)

As far as we know, style has not been addressed in nonverbal communication for ECAs, only the style of the used language was considered (Walker et al., 1997). But there have been ECAs developed sensitive to social role (Prendinger and Ishizuka, 2001), with personality (NECA eShowroom; Perlin, 1995) and emotions (Ball and Breese, 2000; De Rosis et al., 2003).

### 5.2. Implementation and experiments

The first version of GESTYLE with all the features discussed is implemented in XSL and Java. We used MindMaker's FlexVoice TTS system to generate expressive

speech. We made interfaces to two ECA animation systems: to STEP (Eliens et al., 2002) and CharToon (Ruttkay and Noot). The first system allows to animate humanoid agents on the Web. We made a demonstrator of a presentation agent with style manifested in hand gesturing and speech characteristics. The second system we interfaced to was made to design and animate 2D cartoon-like agents. We prepared several animations with GESTYLE by using CharToon as the final animation module, where the talking head is telling the same text, but in different style manifested in speech, facial expressions, gaze and in one case also hand gestures. Some examples with both ECA modules are to be seen at GESTYLE demos.

As of evaluation, one could evaluate two different aspects of GESTYLE: its appropriateness as a software tool, and the appropriateness of the defined styles. As of the first issue, as the system is an in-house research tool (lacking e.g. a user manual), it has not been tested yet by a group of 'ECA authors', but it raised quite some interest among researchers as well as animators.

As of the second issue, we do not intend, ourselves, to carry on 'style studies', we look forward to get input from, and make our tools available for, researchers with relevant background in cultural anthropology and sociology. The same applies for the definition of expressive speech.

But whichever way a style is defined, it is also a question if users perceive the subtle differences at all. We wish to perform tests on the effect of the style of hand gesturing. Earlier dedicated experiments, addressing single modalities (gaze, eyebrow, speech) and sometimes with different user groups, suggested that it is not only 'worth', but necessary to have a tool to fine-tune such subtle aspects of communications (Krahmer and Swerts, 2004; Krahmer et al., 2002a, b; Krahmer et al., 2003).

### 5.4. *Summary and further works*

We have presented GESTYLE, a mark up language for the specification of multimodal ECA behaviour, which allows the user to express style both for speech and for nonverbal modalities. At the core of GESTYLE are the style dictionaries containing mappings of high level Meanings to low level Gestures. As several dictionaries can be used, it is possible to derive a current style by 'adding up' the effect of different decisive factors like gender, culture and profession. GESTYLE also allows to fine-tune gesturing by specifying motion characteristics on three levels: in the style dictionaries, by using modifiers to indicate temporary changes or by inserting gestures explicitly. These features make GESTYLE a powerful tool to quickly define variations in presentation for ECAs, in a modular, systematic and re-usable way.

One would like to cover further aspects of style in GESTYLE. A declaration of the 'look' (gender, age, dress, accessories) could be included in the style definition. We refer to a first experiment, made partly with tools developed by us, where the 'look' already suggested personality characteristics (Smet, 2003). The effect of

(in)consistency in expressing style in look and by facial expressions will be further investigated.

The linguistic style is of major importance. However, the generation of the text to be spoken is outside of the scope of GESTYLE. What could be added is the automatic insertion of nonspeech elements (e.g. 'wow' to express surprise) or text to express emotion (e.g. cursing).

## References

Arafa, Y., Kamyab, K., Kshirsagar, S., Guye-Vuilleme, A., Thalmann, N., 2002. Two approaches to scripting character animation. Proceedings of the AAMAS Workshop "Embodied Conversational Agents—Let's Specify and Evaluate Them! Bologna, Italy.

Ball, G., Breese, J., 2000. Emotion and personality in a conversational agent. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (Eds.), Embodied Conversational Agents. MIT Press, Cambridge, MA, pp. 189–219.

Cassell, J., 1998. A framework for gesture generation and interpretation. In: Cipolla, R., Pentland, A. (Eds.), Computer Vision for Human–machine Interaction. Cambridge University Press, Cambridge.

Cassell, J., 2002. Nudge nudge wink wink: elements of face-to-face conversation for embodied conversational agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (Eds.), Embodied Conversational Agents. MIT Press, Cambridge, MA, pp. 1–27.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., Yan, H., 1999. Embodiment in Conversational Interfaces: Rea", ACM CHI 99 Conference Proceedings, Pittsburgh, PA. pp. 520–527.

Cassell, J., Sullivan, J., Prevost, S., Churchill, E., 2000. Embodied Conversational Agents. MIT Press, Cambridge, MA.

Chi, D., Costa, M., Zhao, L., Badler, N., 2000. The EMOTE model for effort and shape. Proceedings of Siggraph, pp. 173–182.

De Carolis, B., Carofiglio, V., Bilvi, M., Pelachaud, C., 2002. APML, a mark-up language for believable behavior generation. Proceedingsof the AAMAS Workshop "Embodied Conversational Agents—Let's specify and evaluate them, Bologna, Italy.

De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., De Carolis, B., 2003. From Greta's mind to her face: modeling the dynamics of affective states in a conversational embodied agent. International Journal of Human–Computer Studies 59, 81–118.

De Smet, A., 2003. Experimental research to investigate the effect of the 'look' of computer characters on perceived personality. Master Thesis, University of Tilburg, 2003.

Eliens, A., Huang, Z., Visser, C., 2002. A platform for embodied conversational agents based on distributed logic programming. Proceedings of the AAMAS Workshop "Embodied Conversational Agents—Let's Specify and Evaluate Them! Bologna, Italy.

GESTYLE demos: http://www.cwi.nl/~zsofi/gestyle.

Hartmann, B., Mancini, M., Pelachaud, C., 2002. Formational Parameters and Adaptive Prototype Instantiation for MPEG-4 Compliant Gesture Synthesis. Proceedings of Computer Animation. IEEE Computer Society Press, Silver Spring, MD, pp. 111–119.

Isbister, K., Nakanishi, H., Ishida, T. Nass, C., 2000. Helper agent: designing an assistant for human–human interaction in a virtual meeting space. Proceedings of the CHI, pp. 57–64.

Kendon, A., 1993. Human Gesture. In: Ingold, T., Gibson, K. (Eds.), Tools, Language and Intelligence. Cambridge University Press, Cambridge.

Kopp, S., Wachsmuth, I., 2000. Planning and motion control in lifelike gesture: a refined approach. Post-proceedings of Computer Animation. IEEE Computer Society Press, Silver Spring, MD, pp. 92–97.

Krahmer, E., Swerts, M., 2004. More about brows. In: Ruttkay, Z., Pelachaud, C. (Eds.), From Brows to Trust—Evaluating ECAs. Kluwer, Dordrecht.

Krahmer, E. Ruttkay, Zs. Swerts, M., Wesselink, V., 2002a. Pitch, eyebrows and the perception of focus. Proceedings of Speech Prosody, Aix en Province, France, pp. 443–446.

Krahmer, E. Ruttkay, Z., Swerts, M., Wesselink, W., 2002b. Audiovisual cues to prominence. Proceedings of the International Conference Spoken Language Processing, Denver, CO, pp. 1933–1936.

Krahmer, E., Van Buuren, S., Ruttkay, Z., Wesselink, W., 2003. Audio-visual personality cues for embodied agents: an experimental evaluation''. Proceedingsof the AAMAS03 Ws ''Embodied Conversational Characters as Individuals, Melbourne, Australia.

Krandsted, A., Kopp, S., Wachsmuth, I., 2002. MURML: A multimodal utterance representation markup language for conversational agents. Proceedingsof the AAMAS Workshop ''Embodied conversational agents—Let's Specify and Evaluate Them!'', Bologna, Italy.

Lester, J., Voerman, J., Towns, S., Callaway, C., 1999. Deictic believability: coordinated gesture, locomotion and speech in lifelike pedagogical agents. Applied AI 13 (4/5), 383–414.

Massaro, D.W., Cohen, M.M., Beskow, J., Cole, R.A., 2002. Developing and evaluating conversational agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (Eds.), Embodied Conversational Agents. MIT Press, Cambridge, MA, pp. 374–401.

McBreen, H. M., Anderson, J., Jack, M., 2001. Evaluating 3D embodied conversational agents in contrasting VRML retail applications. Proceedings of AAMAS01 Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents, Montreal, Canada.

McNeill, D., 1991. Hand and Mind: What Gestures Reveal about Thought. The University of Chicago Press, Chicago.

Nass, C., 2004. Etiquette equality: exhibitions and expectations of computer politeness. Communications of the ACM 47 (4), 35–37.

Nass, C.I., Steuer, J., Tauber, E., 1994. Computers are social actors. Proceedings of CHI'94, Boston, MA.

Nass, C., Isbister, K., Lee, E.-J., 2002. Truth is beauty, researching embodied conversational agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (Eds.), Embodied Conversational Agents. MIT Press, Cambridge, MA, pp. 374–402.

NECA eShowroom http://www.eshowroom.org.

Noot, H., Ruttkay, Z., 2004. Style in gesture. In: Camurri, A., Volpe, G. (Eds.), Gesture-Based Communication in Human–Computer Interaction, Lecture Notes in Computer Science, No. 2915. Springer, Berlin.

Perlin, K., 1995. Real time responsive animation with personality. IEEE Transactions on Visualization and Computer Graphics 1 (1).

Piwek, P., Krenn, B. Schröder, M. Grice, M., Baumann, S. Pirker, H., 2002. RRL: a rich representation language for the description of agent behaviour in NECA. Proceedings of the AAMAS Workshop on ''Embodied Conversational Agents—Let's Specify and Evaluate Them!'', Bologna, Italy.

Prendinger, H., Ishizuka, M., 2001. Social role awareness in animated agents. Proceedings of Autonomous Agents Conference, Montreal, Canada, pp. 270–277.

Reeves, B., Nass, C., 1996. The Media Equation—How People Treat Computers, Television and New Media Like Real People and Places. Cambridge University Press, Cambridge.

Ruttkay, Z., Noot, N., Animated CharToon faces. Proceedings of the First International Symposium on Non Photorealistic Animation and Rendering, Annecy, pp. 91–100.

Ruttkay, Z., Dormann, C., Noot, H., 2004. ECAs on a common ground. In: Ruttkay, Z., Pelachaud, C. (Eds.), From Brows to Trust—Evaluating ECAs. Kluwer, Dordrecht.

Ruttkay, Z., Pelachaud, C., Poggi, I., Noot, H., to appear. Exercises of style for virtual humans. In: Canamero, L, Aylett, R. (Eds.), Advances in Consciousness Research Series, John Benjamins Publishing Company.

Tsutsui, T. Saeyor, S., Ishizuka, M., 2000. MPML: a multimodal presentation markup language with character agent control functions. Proceedings of the (CD-ROM) WebNet 2000 World Conference on the WWW and Internet, San Antonio, Texas.

Virtual Human Markup Language (VHML) http://www.vhml.org.
Van Moppes, V., 2002. Improving the quality of synthesized speech through mark-up of input text with emotions. Master Thesis, VU, Amsterdam.
Walker, M., Cahn, J., Whittaker, S., 1997. Improvising linguistic style: social and affective bases for agent personality. Proceedings of Autonomous Agents Conference.
Walker, J., Sproull, L., Subramani, R., 1994. Using a Human Face in an Interface. Proceedings of CHI'94, pp. 85–91.