The jovial, the reserved and the robot

Zs. Ruttkay
CWI
Center for Mathematics and Computer Science,Kruislaan 413
1090 GB Amsterdam, The Netherlands
zsofi@cwi.nl

V. van Moppes
Epictoid Bv
Kruislaan 402, 1098 SM
Amsterdam, The Netherlands
vincent@epictoid.nl

H. Noot
CWI
Center for Mathematics and Computer Science, Kruislaan 413
1090 GB Amsterdam, The Netherlands
han@cwi.nl

Abstract

We discuss a language, GESTYLE, to define markup tags representing meaningful behaviors of ECAs with both non-verbal style and speech style explicitly given. In the same framework, these tags can then be used to markup text and hence animate the ECA.

1 Introduction

Up until recently, the embodied conversational agent [4] (ECA) community has been busy with figuring out basic and technical questions, such as how to make a face wrinkled, how to solve the inverse kinematics problems of the human body, or how to make synthetic speech sound better. We took (often from public domain or by courtesy of a colleague) a model to demonstrate the particular aspect. The question of making an ECA appealing and consistent was an 'art' left for the graphics designer and animator, and considered necessary only for commercial applications.

However, in the last few years, some empirical studies have pointed out how important the details are. Humans, unconsciously or not, respond differently to ECAs with different personality, which they derive from (or project into) the synthetic character by using subtle features such as look [28], speech intonation characteristics, and body postures [11]. Moreover, there have been experiments suggesting that the judgment of an ECA depends also on characteristics of the user, such as ethnicity and personality. So when designing ECAs for applications, we cannot discard the importance of these factors. Most recently, some of these got attention in the ECA research [3, 6, 13, 16]. The well-known BEAT system [5], followed by others, have demonstrated the use of multiple nonverbal signals to accompany speech. Badler's body animation system [2] provides mechanisms to parameterize hand gestures to generate different motion styles. Perlin [13] has been using noise to avoid repetitiveness. Signal processing techniques have been proposed to change the manner of a captured motion. However, none of these works addressed the issue of style per se, as we do, and also not in combination with speech.

But then immediately a multitude of questions arise: how to choose the individual parameters (such as look, voice and speech qualities, facial and hand gesturing habits), how to achieve a desired effect, that is, to end up with a 'jovial school teacher' or a 'serious business consultant' type ECA? Who should tell whether the educational ECA should be a jovial school teacher or a serious business consultant?

To begin with, one could rely on some stereotypes for dress and communication etiquette in certain professional and social roles. E.g. a shop assistant should be well dressed, polite, be not too personal or informal and should be economic and to the point in communication. However, just because of the stereotypic image, how refreshing it is sometimes to run into an 'atypical shop assistant' or an informal business director. Well, it depends on the personality of the listener how a deviation from the stereotypic behavior is perceived: trusted and enjoyed more, or just the opposite.

Note that besides the negation of both factors, i.e. trusted less and enjoyed less, two other possibilities exists: trusted less and enjoyed more, or trusted more and enjoyed less. So there may be a multitude of such expectations of the user, which cannot, or should not, all is satisfied.

1.1 The necessity for individual characters

Before concluding that the situation is hopelessly complex, we should realize that the very same complexity exists in every-day human-human communications. And what luck, that we humans are all somewhat different. Not only would life be terribly boring with copies of perfectly performing robots (as depicted in many science fiction books), but these creatures would run into problems all the time which would cause delays or even fatal mistakes in communication, and thus, living.

Think of the simplest example of conversation: how we start, carry on or interrupt a conversation with somebody. We use a multitude of non-verbal signals besides, and sometimes instead of the verbal communication. By selecting a signal e.g. for greeting, we express the relationship to the partner in terms of social power (one greets a high-ranked boss differently than a colleague), gender and culture (e.g. kissing, and the number of kisses, is regulated differently from culture to culture, and also whether the persons are of the same or different sex). But one's way of using

gestures says something about his general personality (jovial people do greet with more intense waving, while a reserved person may never use hand-waving for greeting) as well as his current mood (an enthusiastic wave versus a reserved hand lifting or just a short eyebrow-raise).

Space does not allow us to elaborate further on the necessity of making an ECA individual. Summing up our claims, we must pay attention to the individuality of an ECA, because this factor:

1. provides more fun;
2. influences the efficiency of performing tasks with the help of the ECA.

In order to be able to design and develop individual ECAs, we must have:

1. Categorization of types of individuals and their effects on communication (e.g. introvert persons listen more to an advice given by an introvert person).
2. Prescriptive theories of factors and manifestations of individuality of characters (e.g. introvert/extrovert is a dimension of the personality, which is manifested in certain characteristics of the face, gesturing and speech).
3. Tools to design representations of individual characters, across all their modalities.
4. Control mechanisms that let the character communicate in its individual way, but taking into account dynamically changing factors of the situations too.
5. Thorough testing if our decisions taken along the above steps are reflected by the user communicating with our carefully designed (and expensive) ECA. That is, if our individual ECA indeed has the added values of increased fun and/or effectivety of communication.

1.2 Our related work

We sketched a full picture, in order to place our related and current work in perspective. In our earlier work [18] we addressed the issue of style in nonverbal communication. We claimed that a style layer has to be added, where the more or less individual, but in all cases 'personal' characters can be defined. Our previous work falls under 2 in the above list. We identified the static (such as the social, cultural) and dynamic (such as mood, relationship to listener, availability of resources for listener/speaker) decisive factors of style. We proposed a parameterization of non-verbal style, and a partial, many to many mapping from the decisive factors to some parameter values. E.g. the ethnicity of the speaker determines which gestures can be used to express what meaning. The gender may further refine this mapping (e.g. exclude rude emblems), and influence the motion characteristics (women usually move and gesture more graciously than man). Personal physical and psychological features (dominant hand, extrovert/introvert) may contribute to the fine-tuning of nonverbal communication.

In order to have styled, individual ECAs, we have been developing the GESTYLE language, which serves both the purposes mentioned under 3 and 4. A detailed discussion of the language constructs of GESTYLE is to be found in [12].

In this paper we introduce the main mechanisms of GESTYLE (section 2), and concentrate on an aspect not covered elsewhere. Namely, we show how the tuning of the speech, both the voice characteristics and intonation and rhythmic parameters of speech, as well as their dynamical change to convey some meaning (section 3) are included in GESTYLE. One can think of these facilities as 'speech style', used together with nonverbal stylistic elements. In section 4 we illustrate the usage of the GESTYLE markup language on an example. Finally, in section 5, we tell about ongoing experiments to demonstrate the effect of using GESTYLE, and discuss and further research issues.

2 Nonverbal style with GESTYLE

We have designed and implemented a new, XML compliant language called GESTYLE. It is to serve both of the purposes discussed above: it can be used to *define* style and to instruct the ECA to *express* some meaning nonverbally (too). The novelty of GESTYLE is that it deals with the concept of *style*, as compared to other markup languages which either operate on the signal level [AML in 1, 24, 25] or on the meaning level [CML in 1, 7, 10, 14].

For the ECA, its style defines what gestures it 'knows', and what the habits of using these gestures are, concerning intended meaning, modality usage and subtle characteristics of the gestures. GESTYLE thus allows the usage of high-level meaning tags, which get translated, according to the defined style of the ECA to the low-level gesture tags specifying the appropriate gestures to be performed and possibly also to some parameter values (such as available modalities), see Fig.1.

2.1 The constructs in GESTYLE

GESTYLE is hierarchically organized: At the atomic level there are so-called **basic gestures** (e.g. right-hand beat, nod). Basic gestures can be combined into **composite gestures** (e.g. two-hand beat, right-hand beat and nod) by **gesture expressions.** At the next level, the **meanings** denote the communicative acts (e.g. show happiness, take turn in a conversation) that can be expressed by some gestures. The mappings of meanings to alternatives of (usually composite) gestures are given as entries of **style dictionaries.** A style dictionary contains a collection of meanings pertinent to a certain style (e.g. a style dictionary for 'teacher', 'Dutchman' etc.). A **meaning mapping definition** contains alternative ways of expressing the same meaning by different gestures, each with a certain probability. At runtime these probabilities, taking into account also the fact that some modalities might be in

use in a given situation, determine how a meaning is actually expressed.

Separate from this hierarchy, GESTYLE supports the **manner definition** specifying motion characteristics of gestures (e.g. whether the motion is smooth or angular) and the **modality usage** specifying preference for the use of certain modalities (e.g. use more/less hand gestures). This manner definition is there to express dynamic, situation dependent (e.g. getting tired) changes of style. Finally there is the (static) **style declaration,** which specifies the style of the ECA. A style is declared by specifying a combination of style dictionaries plus, optionally, a manner definition and a modality usage element.

2.2 Meaning mapping in GESTYLE

The intended usage of GESTYLE is the exploitation of the power of declared style: a text, marked up with the same meaning tags, can be presented with different gestures, according to the specified style of the ECA. Meaning tags are available to annotate the text with communicative functions without specifying what gestures should be used to express them. There can be meaning tags to indicate the emotional or cognitive state of the ECA, to emphasize something said, to indicate location, shape or size of an object referred to, to organize the flow of communication by indicating listening or intention of turn taking/giving, etc. The possible categories and tags for meanings are discussed in [12]. From the point of view of the GESTYLE language, all what we assume is that meaning tags are uniquely identified by their name. We are not interested in either the semiotics or the origin of the meaning tags, but only in which nonverbal gestures and speech characteristics can be used to express a specific meaning [15].

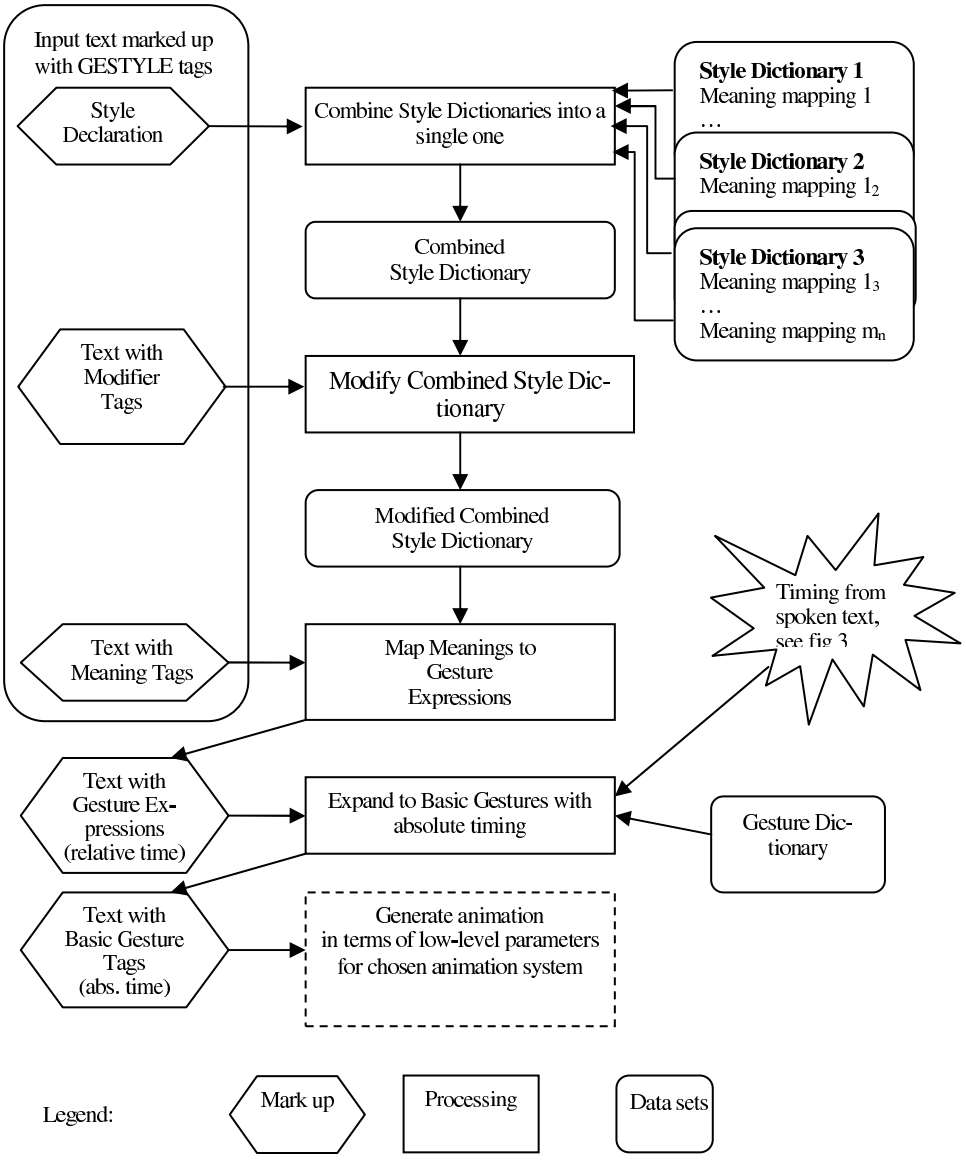

Figure 1 The GESTYLE engine

3 Speech style with GESTYLE

Apart from gestures, meanings can also be reflected by changes in speech. Such changes will result in the synthesis of expressive speech. For example, the meaning 'angry' might not be expressed by 'wildly waving hands about' only, but also by a 'bewildered' or 'angry' voice.

In essence, meanings can map to **expressive speech.** Two types of modifiers of speech characteristics can define the effect of expressive speech:

- **Speech property modifiers** influence overall characteristics of the speech, such as the pitch or speaking rate.
- **Phoneme modifiers** affect phonemes, which are the basic building blocks of spoken words. Thus, modifying characteristics on such a low-level can have more subtle effects on a voice than simple speech parameters. Phoneme modifiers can indicate e.g., that all vowels above a certain utterance duration should be stretched out even longer, or that the pitch for certain phonemes get increased.

For an in-depth discussion of the modifiers see [26]. The definitions for different expressive speech are not hidden; rather, they are to be given in the GESTYLE framework. The specific speech-related tags can be seen as a high-level speech markup language, with infinite extendibility. This differentiates GESTYLE from existing speech markup languages, which often either do not have high-level tags, or only a limited set. Examples of such languages are the Speech Synthesis Markup Language (SSML [23], currently a working draft of the W3C), the Sable consortium's SABLE [19] (an attempt to combine Sun Microsystems' Java Speech Markup Language, JSML, Bell Labs' Spoken Text Markup Language, STML [20] and the aforementioned SSML). There is also the markup format of the Microsoft Speech API, MSAPI [21], which is quite similar to SABLE, though not interoperable.

Yet none of these markup languages have tags on a higher level than being able to specify 'emphasis' on a word or sentence, or being able to give a voice a different 'age' or 'gender' (the effects of which rely entirely on how the text-to-speech engine implements this internally). There is also the Speech Markup Language (SML [22]), which does provide a limited set of emotions as higher level tags. None of the above combines both verbal and non-verbal (i.e., gestures) behavior, however.

Extensibility is the main reason why GESTYLE was developed, instead of using other markup languages. Not only does it allow for high-level tags, but also the user can define their own high-level tags in terms of lower level tags. For example, one could define a happy_speech expressive speech tag with speech properties that give the normal voice a 20% increase in speed rate, and a 30% increase in pitch. This definition can be used over and over in GESTYLE's style dictionaries specifying meaning tags as well as markup tag applied to text, to influence speech directly.

Let's take a closer look at how expressive speech is actually defined in terms of XML. All the definitions are contained in the root tag of an XML document, speech_expression_list. Each definition is placed in a separate element (speech_expression _definition). A speech expression is uniquely identified by its NAME attribute.

Such a speech expression definition consists of two parts: the direct effects the expression has on the speaker's voice (properties element) and the part that operates on the phonemes (phoneme_level element), as described above. An example of such an expression definition is displayed below.

The voice properties part is straightforward: specific parameters can be changed by simply specifying the percentage of increase or decrease in the respective elements. These are parameters that text-to-speech engines are able to accept in one way or another (or, in case a parameter is not supported, it can just be ignored). The phoneme_level part of the definition is a little more complex, consisting of rules prescribing subtle and context-dependent changes in the phoneme sequence. Most of these rules amount to changing the inflection of a voice in the context of a sentence, such as stressing vowels at the end of a sentence.

```
<speech_expression_definition NAME="happy_speech">
        <properties>
                <speed_rate>20</speed_rate>
                <pitch_rate>30</pitch_rate>
                <pitch_range>50</pitch_range>
                <intonation_level>25</intonation_level>
        </properties>
        <phoneme_level>
                <smooth DIRECTION="up" GAP="10">5</smooth>
                <stress_vowel_duration>20</stress_vowel_duration>
                <final_vowel_pitch_inc>15</final_vowel_pitch_inc>
        </phoneme_level>
</ speech_expression_definition>
```

Figure 2 A speech expression definition

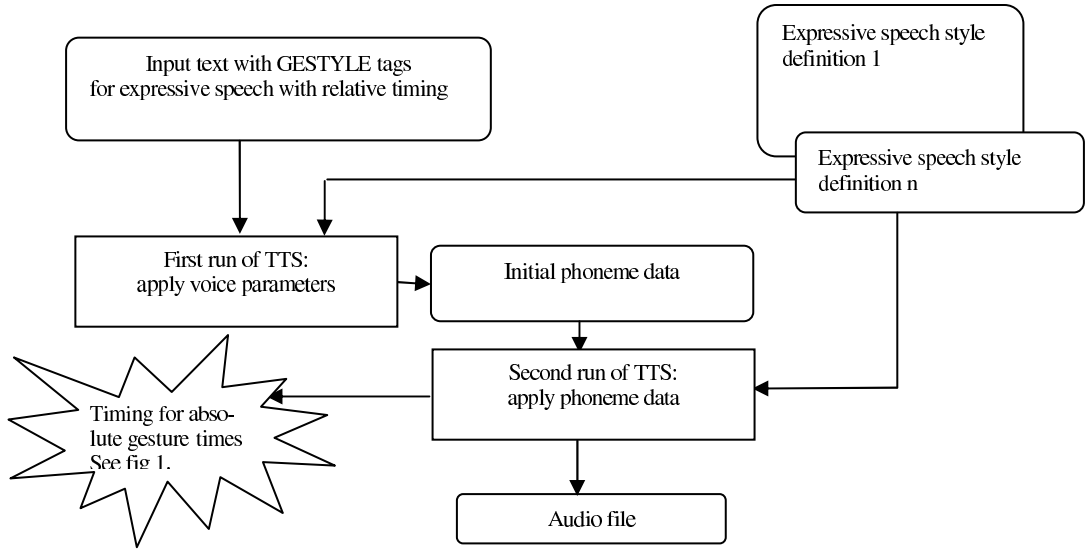

Figure 3 The production of styled speech and timing information

It is not only possible to create an arbitrary number of speech expression definitions in the way described above, it is also possible to make multiple variants of the same speech expressions (reflecting that in case of one person, excited speech means increase in tempo and pitch, while in another case it is only reflected in tempo increase). These alternative definitions should be given in different XML documents, conceptually corresponding to different speech styles. The current global GESTYLE style will dictate what speech style to use, just as it does for gestures.

The global design of the speech style integration in GESTYLE is displayed in Figure 3. It is a two-pass system: once the speech properties have been applied and the voice has been synthesized, there needs to be another pass to apply the phoneme modifiers and re-synthesize the speech again, this time 'for real'. Without this second pass, it would not be possible to apply the subtle effects in changing a voice that you can only achieve by directly operating on phonemes rather than voice properties. This pass uses information (e.g. length) on the phonemes provided by the TTS system.

The actual text-to-speech (TTS) engine used to generate the synthesized speech has not yet been discussed. GESTYLE is largely independent of specific TTS engines (just like the previously mentioned speech markup languages). Of course, there are some requirements that a TTS engine must conform to to make it usable at all. First, there must be a way to really influence speech properties. Just about any engine, commercial or academic, allows changes in speech rate or pitch (although the latter is not always the case with synthesis engines that rely on the concatenation of pre-recorded speech). If certain speech properties are not supported, however, they can be ignored when feeding the text to the TTS engine, with the downside of a lower quality of expressive speech.

To be able to modify the phonemes of a piece of text, the TTS engine, additionally to the binary audio of the generated speech, must provide the phoneme sequence too, with the duration and pitch of each phoneme. The engine must also be able to accept phoneme sequences as input. If the last functionality is not available, expressive speech can still be generated, but the result will be of lower quality, as phoneme modifiers will not be supported. However, the ability to output phonemes and their durations is an absolute requirement. This is rarely a problem with commercial engines, which almost always conform to Microsoft SAPI. Providing phoneme data is part of the SAPI specification. Only the very simplest of TTS do not provide this kind of output.

The expressive speech part of GESTYLE is not only there to influence the synthesized speech. In order to create an animation with absolute timing information on when to start certain gestures, one needs to know when words, or even, phonemes, will be pronounced. The timed phoneme sequence that a TTS engine provides is a necessity to synchronize animation to the speech.

A further, still developing area of text-to-speech, TTS markup and speech expression mappings is having more control over the characteristics of a voice. This covers topics as specifying the intended gender of a voice, or something more exotic such as the breathiness of a voice. It also addresses achieving greater flexibility in specifying intonation characteristics, at the phoneme level.

Another area of research for speech experts is to provide appropriate mappings from speech expression definitions to lower level parameters. Extended and perhaps more intuitive rules for phoneme modifiers also need to be researched and developed.

4 An example: Hamlet marked up

```
1       <StyledText>
2       <StyleDeclaration>
3       <style aspect="biology" dict="Biological"/>
4       <style aspect="personality" dict="Extravert"/>
5       </StyleDeclaration>
6       <TextBody>
7       <Meaning Name="Enthusiastic">
8               <Meaning Name="GetAttention" > What a piece of work is a man! </Meaning>
9               How <Meaning Name="Emphasize">noble </Meaning>in reason!
                                *
                                *
10              world! the paragon of animals!
11      </Meaning>
12      <Meaning Name="Contrast"> And yet, </Meaning>
13      <Meaning Name="Sad">
14              <Meaning Name="PointingAtSelf">to me,</Meaning>
15              what is <Meaning Name="EyeM" gesture_length="300"/> this quintessence of
15              <Meaning Name="Emphasize">dust? </Meaning> Man delights me
                                *
                                *
17      </TextBody>
18      </StyledText>
```

Figure 4a Marked up text

```
1       <StyleDictionary Name = "Extravert">
2               <Speechmode Name="="ExtravertSpeechMode">
3               <Meaning Name = "Emphasize" CombinationMode = "DOMINANT">
4               <GestureSpec>
5                       <ExpressiveSpeech emotion="emph_mild"/>
6                       <UseGest Name="nod_and_beat"/><PAR/><UseGest Name=”LookAtPerson”/>
7                       <Probability P="0.7"/>
8               </GestureSpec>
9               <GestureSpec>
10                      <ExpressiveSpeech emotion="emph_strong"/>
11                      <UseGest Name="beat"/>
12                      <Probability P="0.3"/>
13              </GestureSpec>
14              </Meaning>
                                *
                                *
15      </StyleDictionary>
```

Figure 4b A Style Dictionary

In Figure 4a, we see GESTYLE marked up text. In lines 1-5 there is a StyleDeclaration, specifying the style for the biologically necessary behaviors of the ECA (e.g. blinking) and its personality, in this case extravert. Next comes the marked-up text proper. In this text, only those Meanings can be used which are defined in the style dictionaries (see below) listed in the StyleDeclaration. The meaning tags for GetAttention (line 8), Contrast (line 12) and PointingAtSelf (line 14) are – in the style used – mapped purely to gesture tags. On the other hand, the tags for Enthusiastic (line 8), Sad (line 13) and Emphasize (e.g. line 9) have been defined to influence both the gesturing behavior of the ECA and its speech. In the case of Emphasize, the text so marked will be pronounced with emphasis and will be accompanied by some gesture. Enthusiastic and Sad both influence the speech properties and the motion manner characteristics of gestures performed within their scope. These effects follow from the definition of each Meaning given in the style dictionaries, which are referred to in the StyleDeclaration.

The definition of styles, given in style dictionaries, is also expressed in terms of GESTYLE language constructs. An example is given in Figure 4b. There we see a (part of) the style dictionary for the extravert personality type (line 1). The first element of this dictionary is the Speechmode element, which, through the value of its Name attribute, refers to a speech_expression_definition element as discussed in section 3. Next comes the definition of the Meaning tag for emphasize (line 3 till 14). We see that emphasize can be shown in two ways,

because there are two GestureSpecs in this one Meaning definition, each with its own probability of occurrence (0.7 and 0.3 resp.). Both alternatives define the gestures to be used and also change the speech properties (to mild or strong emphasize, see lines 5 and 10).

5 Further issues

One would like to see the benefits of using GESTYLE in practice. Currently we are experimenting with two ECA models. In the first case, a 2D expressive face with hands is used, (made by our own CharToon [17] animation tool) to demonstrate the variety in using modalities of the face, eye and hands. In the second case, we use a full-body 3D H-anim model [9] (defined in VRML), with a rich repertoire of hand gestures and subtle control over their motion characteristics (see Figure 5). In both cases, expressive synthesized speech (produced with FlexVoice [8]) is used. These experiments show that – although much remains to be done – through GESTYLE it feasible to endow an ECA with different styled communication behaviors.

There are at least two important areas that still have to be addressed: First how do we 'fill' our style dictionaries? In literature, movies and cartoons we often meet stereotypes of a British lord, a Dutch farm lady, a priest, a teacher etc. Discarding of the issue of justification of such stereotypes, our problem is how to capture the characteristics of those stereotypes. We do have a mechanism in GESTYLE to describe behavioral style aspects, but *what to put in the description?* In order to make realistic and useful style dictionaries, one should rely on findings from behavioral psychology and cultural anthropology. The same applies for the definition of expressive speech.

The style should correspond to the impression given by the look of and the language used by the ECA. It would be interesting to have computational means to tune these factors too. However, the design of a character with some individual look has been considered as an art, asking for tedious work from modeling experts. The problems of reflecting style in word usage and dialogue management are in the field of natural language generation and discourse management [27]. These issues are outside the scope of today's GESTYLE, but should be dealt with in the future.

Acknowledgement

The FlexVoice text to speech system we have been using in our work was provided by MindMaker. We are thankful to Anton Eliëns for his advice on software engineering issues.

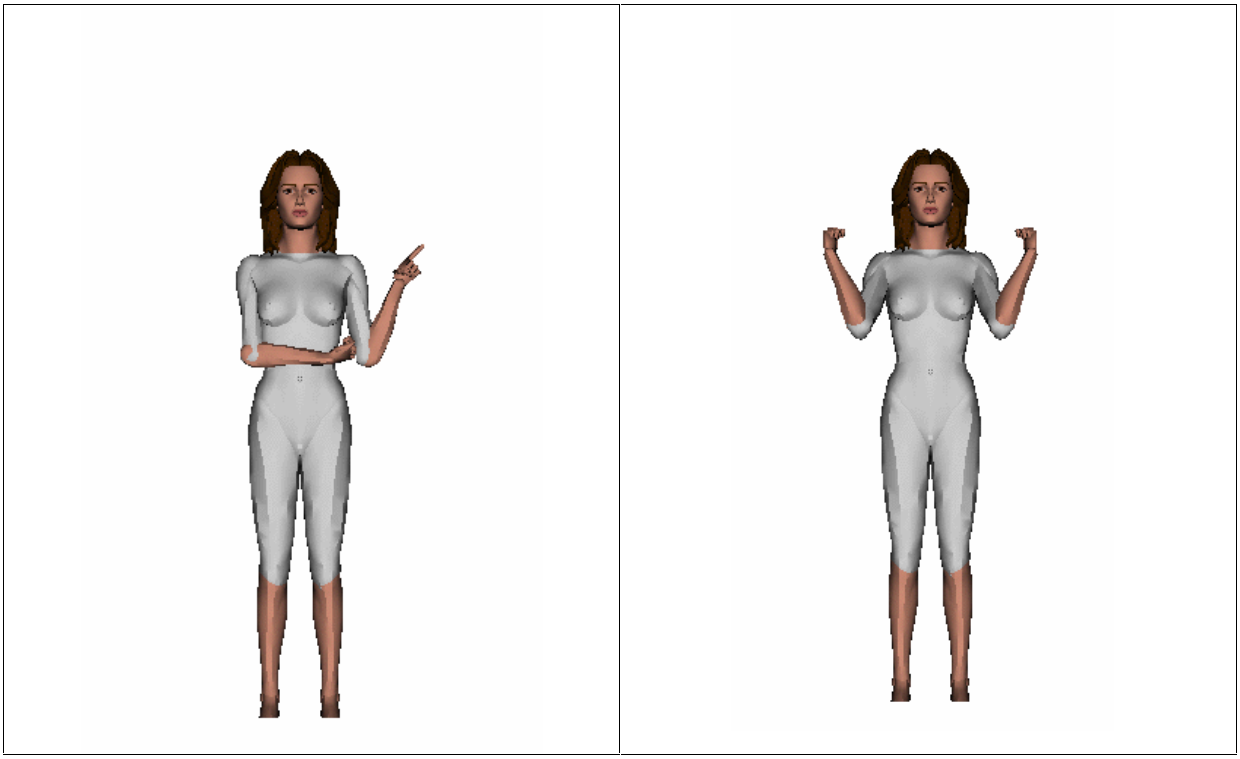

Figure 5 Two hand gestures expressing emphasis in different styles

6 References

1. Arafa, Y., Kamyab, K., Kshirsagar, S., Guye-Vuilleme, A., Thalmann, N. "Two Approaches to Scripting Character Animation", *Proc. of the AAMAS Workshop on "Embodied conversational agents – Let's specify and evaluate them!"*, 2002, Bologna.
2. Badler, N., Bindiganavale, R., Allbeck, J., Schuler, W., Zhao, L., and Palmer., M., *Parameterized Action Representation for Virtual Human Agents.*, In[4], pp. 256-284.
3. Ball, G., Breese, J. "Emotion and personality in a conversational agent", In: Cassell et al. 2000, pp. 189-219.
4. Cassell J., Sullivan J., Prevost S., Churchill E. *Embodied Conversational Agents*, MIT Press, Cambridge, MA. 2000.
5. Cassell, J., Vilhjálmsson, H., Bickmore, T., BEAT: The Behavior Expression Animation Toolkit, Proc. of SIGGRAPH, 2001, pp. 477-486.
6. Chi D., Costa M., Zhao L., Badler N. "The EMOTE Model for Effort and Shape", *Proc. of Siggraph*, 2000. pp. 173-182.
7. De Carolis, Carofiglio, Bilvi, M., Pelachaud, C. "APML, a Mark-up Language for Believable Behavior Generation" *Proc. of the AAMAS Workshop on "Embodied conversational agents – Let's specify and evaluate them!"*, 2002, Bologna.
8. FlexVoice http://www.flexvoice.com/
9. H-anim 2002, Humanoid animation working group: http://www.hanim. org/Specifications/H-Anim1.1/
10. Krandsted, A., Kopp, S., Wachsmuth, I. "MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents", *Proc. of the AAMAS Workshop on "Embodied conversational agents – Let's specify and evaluate them!"*, 2002, Bologna.
11. Nass C., Isbister K., Lee E-J. "Truth is beauty: Researching embodied conversational agents," In: Cassell et al. 2000. pp. 374-402.
12. Noot, H., Ruttkay, Zs. Style in Gesture, *Proc. of Gesture'2003*, Springer-Verlag LNCS Series, to appear in 2003.
13. Perlin K. "Real time responsive animation with personality", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 1. No. 1. 1995.
14. Piwek, P., Krenn, B. Schröder, M. Grice, M., Baumann, S. Pirker, H. "RRL: A Rich Representation Language for the Description of Agent Behaviour in NECA". *Proc. of the AAMAS workshop on "Embodied conversational agents - let's specify and evaluate them!"*, Bologna, Italy. 2002.
15. Poggi I. "Mind Markers", In: Mueller C. and Posner R. (eds): *The Semantics and Pragmatics of Everyday Gestures*, Berlin Verlag Arno Spitz, 2001.
16. Prendinger H., Ishizuka M. "Social role awareness in animated agents", *Proc. of Autonomous Agents Conference*, 270-277, Montreal, Canada. 2001.
17. Ruttkay, Zs., Noot, H. Animated CharToon Faces, *Proceedings of NPAR 2000 - First International Symposium on Non Photorealistic Animation and Rendering,* Annecy, 2000. pp 91-100,
18. Ruttkay Zs., Pelachaud, C., Poggi, I., Noot H. "Excercises of Syle for Virtual Humans", In: L. Canamero, R. Aylett (Eds.), *Advances in Consciousness Research Series*, John Benjamins Publishing Company, to appear in 2003.
19. SABLE http://www.cstr.ed.ac.uk/projects/sable
20. Richard Sproat, R., Taylor P., Tanenblatt M., Isard A., "A Markup Language for Text-To-Speech Synthesis", *Proceedings of Eurospeech 97*, Rhodes, Greece.
21. SAPI http://www.microsoft.com/speech
22. Stallo, J. B. "Simulating Emotional Speech for a Talking Head", *Honours Thesis*, Curtin University of Technology, Bently, Australia, 2000
23. SSML http://www.w3.org/TR/speech-synthesis
24. Tsutsui, T. Saeyor, S., Ishizuka, M. "MPML: A Multimodal Presentation Markup Language with Character Agent Control Functions", *Proc.(CD-ROM) WebNet 2000 World Conf. on the WWW and Internet*, San Antonio, Texas. 2000.
25. Virtual Human Markup Language (VHML) http://www.vhml.org
26. Van Moppes, V. "Improving the quality of synthesized speech through mark-up of input text with emotions", *Master Thesis*, VU, Amsterdam, 2002.
27. Walker, M., Cahn, J., Whittaker, S. "Improvising linguistic style: Social and affective bases for agent personality", *Proc. of Autonomous Agents Conference*. 1997.
28. Walker, J., Sproull, L., Subramani, R. "Using a human face in an interface", *Proc. of CHI'94*, pp. 85-91. 1994.