We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

11

Maxine: Embodied Conversational Agents for Multimodal Emotional Communication

Sandra Baldassarri and Eva Cerezo
Advanced Computer Graphics Group (GIGA), Computer Science Department, Engineering Research Institute of Aragon (I3A), Universidad de Zaragoza, Spain

1. Introduction

Emotions are fundamental to human experience, influencing cognition, perception and everyday tasks such as learning, communication and even rational decision-making. Human-computer interaction (HCI) systems capable of sensing user affective state and giving adequate affective feedback are, therefore, likely to be perceived as more natural, persuasive and trustworthy (Cassell & Bickmore, 2000). Affective Computing (Picard, 1997) is the research area concerned with computing that relates to, arises from, or deliberately influences emotion. Affective computing expands HCI by including emotional communication together with appropriate means of handling affective information. Last scientific researches indicate that emotions play an essential role in perception, decision making abilities, problem-solving process and learning.

During the last years, a considerable amount of researchers have become interested in the development of new interfaces that follow the human-human model of communication. These interfaces must be able to receive and analyze affective information that comes from different input channels: image, video, audio, tactile, position trackers, biological sensors, etc. In this way, affective multimodal interfaces extend the traditional channels of human perception to better mach the natural communication of human beings (Gunes et al., 2008).

Within this context, embodied conversational agents (ECAs) offer a natural and intuitive interface to perform social interaction. The benefits of interacting through 3D human–like virtual agents able to express their emotions have been widely proved (Cassell & Bickmore, 2000) (Pantic & Bartlett, 2007).

Credible ECAs should be able to properly move and communicate taking into account that human communication is based on speech, facial expressions, body language and gestures. In fact, social psychologists argue that more than 65% of the social meaning of a person-to-person conversation relies on the non-verbal band (Morris et al., 1979) highlighting the importance of emotions and body language in the communication and comprehension of information (Birdwhistell, 1985). Emotions are induced through all body language (Picard, 1997), nevertheless, most of the work in this area is focused in the facial expression of emotions (Cowie et al., 2001) (deRosis et al., 2003) (Raouzaiou et al., 2004) (Zhang et al., 2010) (Gunes et al., 2011). There is a lack of research works that include animation and

synthesis of emotions through the body, probably because of the inherent complexity of determining the characteristics of each emotion. Some researchers include gestures in their work, but only in order to emphasize a verbalized message (Noma & Baldler, 1997) (Haviland, 2004) (Cassell, 2007). There are, however, some applications that include emotional virtual humans but that are usually developed for specific and delimited environments like the tutor Steve (Rickel & Johnson, 2000), an agent used for education and training that has the ability to give instant praise or express criticism depending on the succeed or failure of students' answers. Unfortunately Steve only express gestures through the hands and the head and, although it makes gestures as nodding, refusing or pointing, it doesn't establish any relationship between emotions and gestural behavior. In the work of Prendinger and Ishizuka's (Prendinger & Ishizuka, 2001) emotional virtual agents are used to interact with the users and other agents, but are not very expressive. On the other hand Caridakis et al. (Caridakis et al., 2006) work with more expressive agents but their expressions are limited to gestures captured from video and only hand and head gestures are considered.

There are other very interesting works that focus in non-verbal gestures and also consider the emotional state of the virtual agent (Hartmann et al., 2005) (Mancini & Pelachaud, 2009) (Raouzaiou et al., 2004) to express specific gestures. Su et al. (Su et al., 2007) go further developing a virtual story-teller endowed with personality. The personality influences gestures and the expression of the emotions through the whole body, but, unfortunately, only positive and negative emotions are considered in the final body animation.

With all this in mind we have developed Maxine, a powerful engine to manage embodied animated characters that support multimodal and emotional interaction and that are capable of responding appropriately to the users with affective feedback. The aim is to establish more effective communication with the user, recognizing the user's emotional state, and using a virtual character capable of expressing its emotional state. In this chapter we will focus in the novel features recently added to the Maxine engine: the addition of non-verbal communication capabilities to the virtual characters so that they can express emotions through body postures and gestures. Maxine virtual characters have been used in different domains: a young virtual character that helps special needs children to play with a customizable multimodal videogame; a virtual butler to control a remote domotic hall; and an interactive tutor that helps university students to practice Computer Graphics concepts. In this chapter we will focus in an application in which body language becomes essential: a virtual interpreter that translates from Spanish spoken language to Spanish Sign Language taking emotions into account.

The chapter is organized as follows. In Section 2, Maxine, the platform for managing virtual agents is briefly described. Section 3 details the implementation of the new features added to Maxine virtual characters: the expression of emotions through body postures and gestures. Section 4 presents the affective virtual interpreter to Spanish Sign Language developed focusing in affective issues and, finally, in Section 5, conclusions are presented and current and future work is outlined.

2. Maxine: An animation engine for multimodal emotional communication through ECAs

Maxine is a powerful multimodal animation engine for managing virtual environments and virtual actors. The system is capable of controlling virtual 3D characters for their use as new interfaces in a wide range of applications. One of the most outstanding features of the

system is its affective capabilities. The system supports real-time multimodal interaction with the user through different channels (text, voice, mouse/keyboard, image, etc.) and it is able to analyze and process all this information in real-time for recognizing the user's emotional state and managing the virtual agent's decisions. Virtual characters are endowed with facial animation, lip synchronization, and with an emotional state which can modify character's answers, expressions and behaviours. The new features added to Maxine provide virtual characters with full expressive bodies.

The overall architecture of the system, shown in Figure 1, is composed of 4 main modules described in detail in previous works (Baldassarri et al., 2008) (Cambria et al., 2011): the Perception Module, the Affective Analysis Module, the Deliberative/Generative Module and the Motor Module.

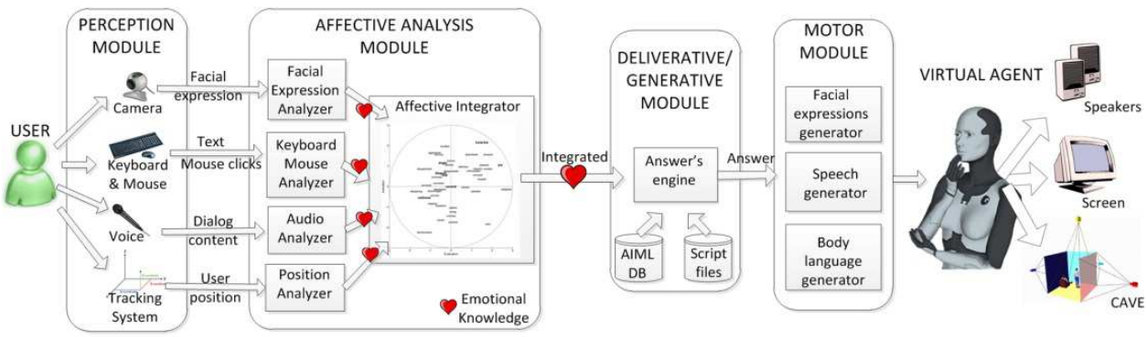


Fig. 1. Maxine's architecture.

The Perception Module consists of the hardware and software necessary to gather the multimodal information obtained during the interaction between the user and the virtual agent: via keyboard, mouse, tactile screens, voice (natural language conversation), communicators, webcam, background microphones, positioning tracker, etc.

Then, from these inputs, the Affective Analysis Module (Cambria et al., 2011) extracts emotional information that it is integrated by a multimodal fusion technique (Hupont et al., 2011). Besides recognizing the user's facial expressions with a webcam (Cerezo et al., 2007) new modules have being added to the system for the detection of affective information. In particular, affective cues are extracted from user's typed-in text, and right now a keyboard pulses and mouse clicks analyzer is being developed to detect states of boredom, confusion, frustration and nervousness of the user. The audio analyzer evaluates the text contents and provides an affective point in the Whissell's space (Whissell, 1989) for each sentence with emotional load. We are also considering the analysis of prosody in speech in a near future.

After gathering and analyzing the multimodal input information and taking into account the dialogue contents, the system has to manage the appropriate virtual agent's reactions and generate the answers to the user's questions. Two kinds of reactions can be distinguished: generative and deliberative. Generative actions are purely reactive: for example, should the user key something in, the virtual presenter will interrupt the presentation; should the user changes his/her position, the 3D actor look/orientation will change; if a lot of background noise is detected, it will request silence, etc. Deliberative character's reaction calls for a more complex analysis which elicits an answer through the user's voice interaction (dialogue contents) and the user's detected emotional state. The Deliberative/Generative Module

basically generates an answer for the user's questions in text mode, and they are based on the recognition of patterns associated to fixed answers. These answers, however, vary according to the virtual character's emotional state (e.g. if the user insults the actor, the later will reply with an angry voice and facial expression), the detected emotional state of the user (captured from video, typed-in text, emoticons), or may undergo random variations so that the user does not get the impression of repetition should the conversation goes on for a long time. The development of this part of the system is based on chatbot technology under GNU GPL licenses: ALICE (ALICE, 2011) and CyN (CyN Project, 2001), and has been modified to include commands or calls to script files which results are returned to the user as part of the answer, when they are executed. This makes it possible, for example, to check the system time, log on to a website to check what the weather is like, etc.

Finally, the Motor Module is in charge of generating the system's outputs and the final animation of the 3D virtual character. The animations which the system works with are derived from two sources: animations from motion capture and animations generated by means of commercial software. Besides general animations that can be regular, cyclic and pose animations, our system allows to work with background and periodic animations, which are secondary animations executed automatically at certain established intervals. One typical example of this kind of animation is used to make an actor blink or register breathing motions. The implementation of secondary animations was done using Perlin's algorithms, based on coherent noise, and using random variation in the execution frequency in order to avoid unnatural movements.

For facial animation, we work with the six basic expressions defined by Ekman: anger, disgust, fear, happiness, sadness and surprise (Ekman, 1999), and animation blending is achieved with the help of the Cal3D library (CAL3D, 2011). The voice synthesis is made using SAPI5 (Long, 2002), a set of libraries and functions that enables to implement a voice synthesizer (in English), and, additionally, the Spanish voices offered by the Loquendo packages. SAPI gives information about the visemes (visual phonemes) that occur when delivering the phrase to be synthesized, what allows to solve the problem of labial synchronization. Visemes are used to model the actor's mouth movements. As the voice generated by text-voice converters usually sounds artificial, is was decided to provide Maxine virtual characters with an emotional voice by modifying the tone, frequency scale, volume and speed (Baldassarri et al., 2009). Script-files-based commands containing the facial expression, the speech and body parameters are generated and executed in real-time to achieve the appropriate body and facial animation (lip-synch and facial expressions) and emotional voice synthesis.

The additions made to the Motor Module in order to achieve more realistic and natural virtual characters improving their body non verbal communication capabilities are described in detail in next section.

3. Body language: Expressing emotions through the virtual agent body

As it has been said before, in order to obtain a more natural interaction with the users, similar to human face-to-face-communication, it is very important that virtual characters can be capable of expressing their emotions in a realistic way. In the early versions of Maxine, emotions were managed basically at a facial level: through the animation of the facial expressions, the modulation of the voice and the lip synchronization, as it was briefly

described in Section 2. Here we present the work done to convert Maxine's virtual agents in credible and humanized characters capable of expressing their emotions through their bodies. For this purpose, the Motor Module of Maxine has been expanded with new features that take into account body language: the generation of emotional body expressions and gestures.

3.1 Emotions through body postures and movements

In Maxine, 3D virtual characters have improved their affective capabilities with the possibility of expressing their emotions through the animation of postures and movements of the whole body. The emotions expressed by the virtual humans' bodies can either be basic or mixed emotions. The basic ones are based on a discrete approach (the six emotions defined by Ekman) while the mixed ones are based in a continuous approach, more suitable for working with the wide range of emotions that can be usually found in social interaction.

3.1.1 Basic emotions

The **basic emotions** considered to be expressed by the virtual characters' bodies are those used in Maxine for facial animation, that is, the six universal emotions proposed by Ekman: anger, disgust, fear, happiness, sadness and surprise, plus the neutral one. Although the problem of facial expression of emotions has been widely studied, there are only few works about the body expression of emotions; probably because of some outward expressions of emotions (body language) have different meanings in different cultures. However, there are some ergonomic studies that establish the more common body postures adopted when one of the six basic emotions is felt. These studies specify the changes in the weight of the body (forward, backward, neutral) and the values of the more important positions and angles of the bones. Our work is based on the study carried out by Mark Coulson (Coulson, 2004) about the expression of emotions through body and movement.

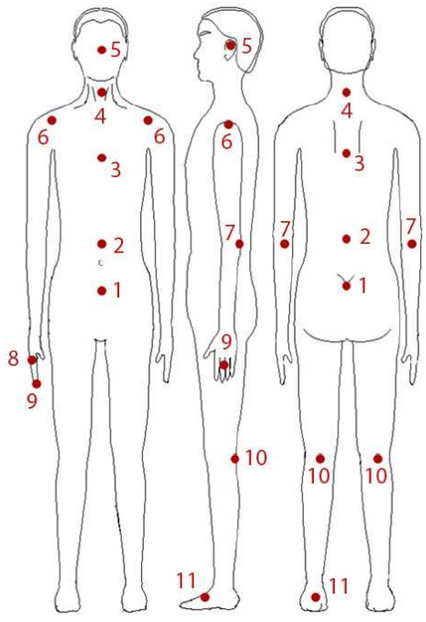


Fig. 2. MPEG4 reference points.

However, for the integration in our system, the values of Coulson had to be adapted to the MPEG4 standard (MPEG4, 2011) (see the positions of the MPEG4 reference points in Figure 2) and, in some cases, they have also been slightly modified, as it is explained later in this section.

The final values of the rotations and movements of each point for showing the different basic emotions can be seen in Table 1. All the values in the table are in degrees refered to the bones' rotations and positive values are considered when the rotation is forward.

	PELVIS (pelvis) ●1	SPINE 4 (abdomen) ●2	SPINE 3 (backbone) ●3	SPINE 2 (neck) ●4	SPINE 1 (head) ●5	RIGHT&LEFT ARM 1(shoulders) ●6		RIGHT&LEFT ARM 1(elbows) ●7		RIGHT&LEFT ARM 2 (hands) ●8		RIGHT&LEFT HAND (fingers) ●9	RIGHT&LEFT LEG 1 (back of the knees) ●10		RIGHT&LEFT LEG 2 (feet) ●11	
						Rotation(1)	Turn(2)	Right	Left	Right	Left		Right	Left	Right	Left
Anger	0º	0º	14º	17º	8º	6º	13º	80º	18º	16º	0º	closed	0º	10º	-	forward
Disgust	6º	20º(3)	0º	0º	-15º 20º(3)	11º	0º	55º	65º	-30º	-50º	opened	0º	11º	-	backward (tiptoe)
Fear	10º	0º	10º	30º	12º	25º	5º	50º	80º	0º	-22º	closed	22º	24º	-	forward
Happiness	8º	-12º	-10º	0º	-5º	-13º	0º	100º		0º		closed	0º		forward	-
Sadness	12 -14º	0º	20º	30º	25º	-20º	0º	30º		30º		half closed	0º	5º	-	forward
Surprise	0º	-17º	-12º	-5º	-13º	10º	0º	90º		0º		opened	25º	0º	backward (tiptoe)	-

* Shoulder's rotation taking collarbone as turn point: positive values indicate up rotation and negative values indicate down rotation.
** Turn of the bone over itself, positive if the turn is forward
***Turn is considered to the sides.

Table 1. Values of the MPEG4 reference point positions in the body expression of the six basic emotions.

Following, some images of different basic emotions generated by our system considering the values described in Table 1, are shown. Figure 3 shows the body of a virtual character expressing the emotions of sadness (left) and fear (right).

It must be said that in some cases, such as in the emotion of anger, the values extracted from Coulson's work give a body posture too masculine (see Figure 4 left). So, for representing animating a female virtual character, some changes have been done and new body postures have been generated (Figure 4 right).

3.1.2 Mixed emotions: From a discrete to a continuous space

The first version of the system was based on the six emotions defined by Ekman (Ekman, 1999) but, in order to add more natural expressions to the virtual agents, the system has been enriched with the possibility of working with mixed emotions, changing from a discrete to a continuous approach.

The Ekman's *categorical* approach, where emotions are a mere list of labels, fails to describe the wide range of emotions that occur in daily communication. There are a few tentative efforts to detect non-basic affective states from deliberately displayed facial expressions, including "fatigue" (Ji et al., 2006), and mental states such as "agreeing", "concentrating", "interested", "thinking", "confused", and "frustrated" (Kaapor et al., 2007) (Yaesin et al.,

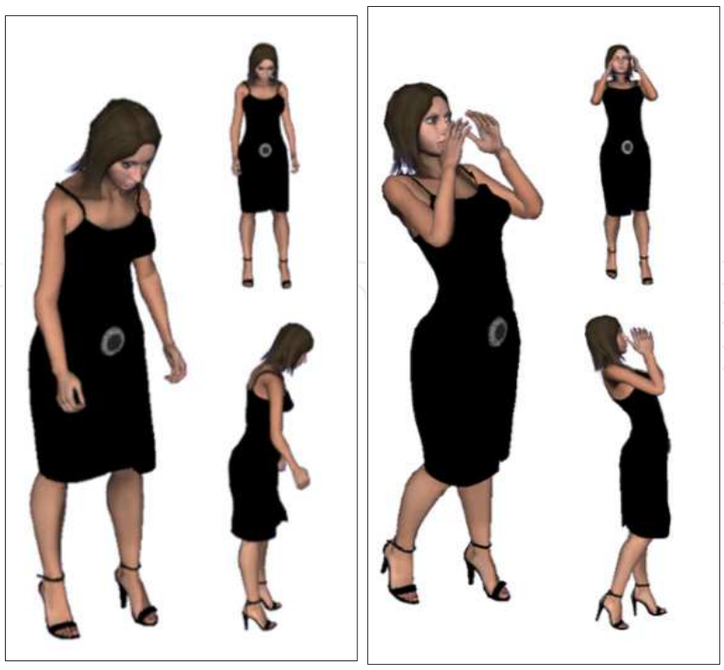


Fig. 3. Body expression of sadness (left) and fear (right).

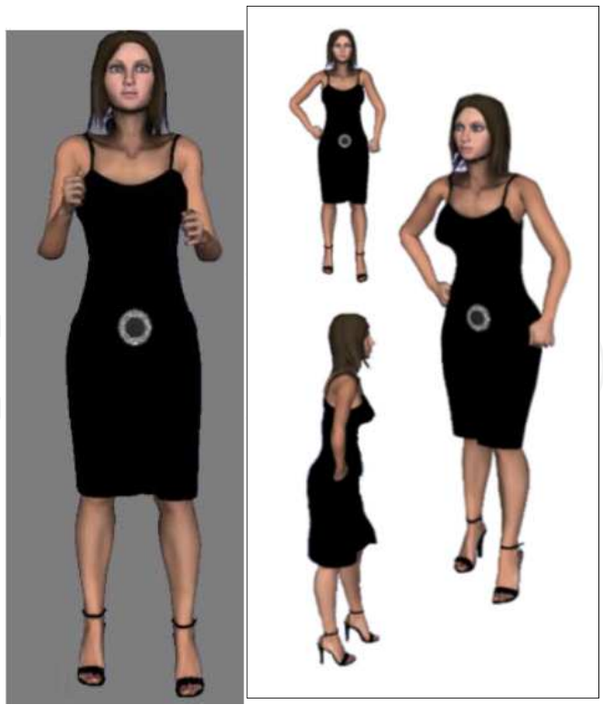


Fig. 4. Body expression of anger considering Coulson's values resulted in a too masculine posture (left). Body posture of anger adapted for a female virtual character (right).

2006). In any case, *categorical* approach presents a discrete list of emotions with no real link between them. To overcome this problem, some researchers, such as Whissell (Whissell, 1989) and Plutchik (Plutchik, 1980) prefer to view affective states not independent but rather related to one another in a systematic manner. In this work we decided to use one of the most influential evaluation-activation 2D models in the field of psychology: that proposed by Cynthia Whissell (Whissell, 1989). She considers emotions as a continuous 2D space whose dimensions are evaluation and activation. The evaluation dimension measures how a human feels, from positive to negative. The activation dimension measures whether humans are more or less likely to take some action under the emotional state, from active to passive. In her study, Whissell assigns a pair of values <evaluation, activation> to each of the approximately 9000 carefully selected affective words that make up her "Dictionary of Affect in Language". Figure 5 shows the position of some of these words in the evaluation-activation space.

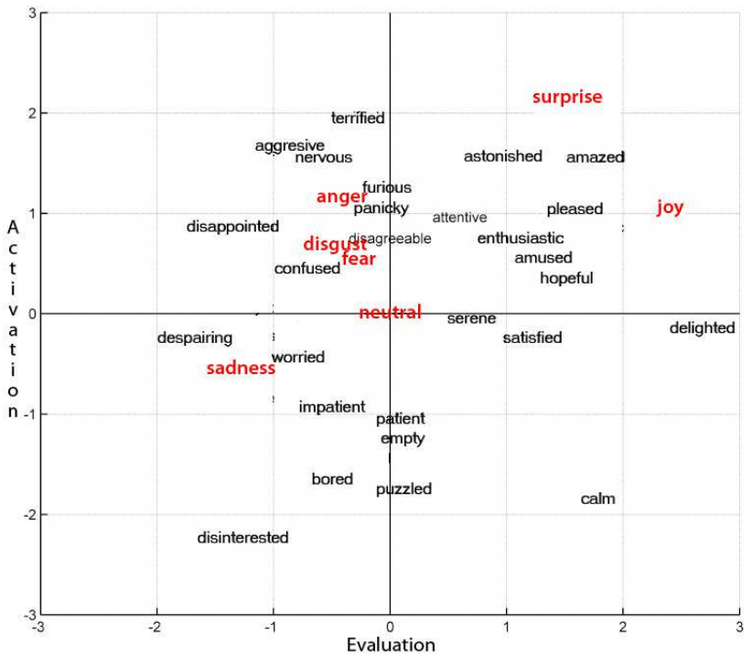


Fig. 5. Simplified Whissell's evaluation-activation space.

It can be seen that the emotion-related words corresponding to each one of Ekman's six emotions have a specific location (x_i, y_i) in the Whissell space (in bold in Figure 5). The final (x, y) coordinates of a mixed emotion can be calculated as the centre of mass of the seven weighted points (corresponding with the Ekman's six basic emotions plus "neutral") in the Whissell's space following (1). In this way the output of the system is enriched with a larger number of intermediate emotional states.

$$x = \frac{\sum_{i=1}^{7} w_i * x_i}{\sum_{i=1}^{7} w_i} \quad \text{and} \quad y = \frac{\sum_{i=1}^{7} w_i * y_i}{\sum_{i=1}^{7} w_i} \tag{1}$$

In order to show the power of our system, we present images of the synthesis of the some specific mixed emotions, defined as combinations of weighted points of the basic ones. Following the equations defined in (1), the "astonishment" emotion shown in Figure 6 (left)

can be represented in the 2D Whissell's space as a weighted combination of surprise, happiness and neutral while the emotion of "displeasure" results of considering a weighted mixed between disgust and happiness (see Figure 6 right).

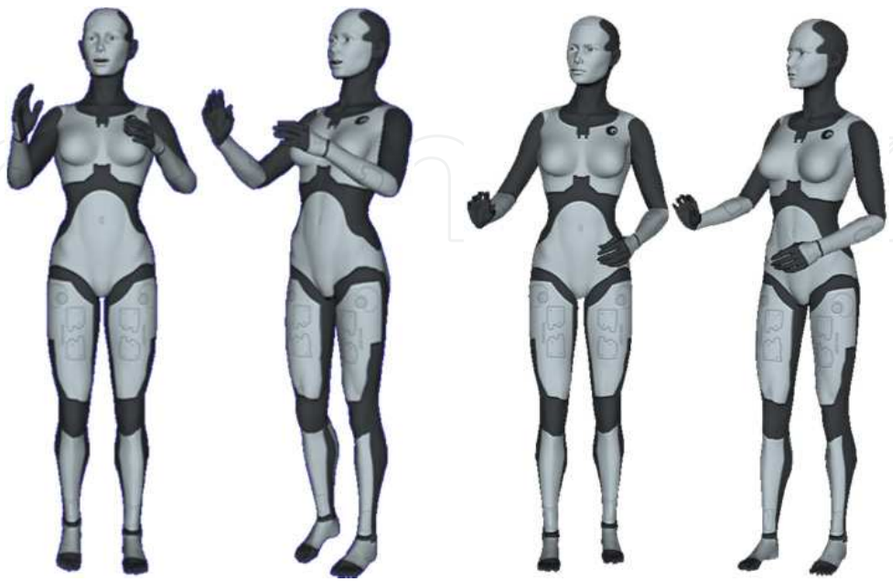


Fig. 6. Virtual character's expression of astonishment (left) and displeasure (right).

3.2 Gestures

Gestures are other way of non verbal communication, and usually comprise different parts of the body like arms, hands, fingers, legs, and feet movement. Gestures allow individuals to communicate a variety of feelings and thoughts, from contempt and hostility to approval and affection, often together with body language in addition to spoken words. However, the meaning of each gesture can vary across cultures, status, occupation, age or gender (Brislin, 1993). In our work we exclude these complex factors and deploy well-known and commonly used gestures in order to be used and understood everywhere. On the other hand, there are many gestures that are performed during social interaction in order to express or reinforce an emotion, such as self-touching gestures. Therefore, we decide to allow the addition of body gestures to the previously generated mixed emotions, based in the work of Su *et al.* (Su et al., 2007).

Summing up, our Gesture module allows the generation of the most commonly used gestures, such as nodding, shaking the head, hiding the face, clenching the fists, saying O.K., or greeting someone with the hand; or allows adding gestures to other body postures and movements, in order to improve the expression of emotions.

Following, some examples are shown. Figure 7 shows a virtual character expressing negation by moving the forefinger, while in Figure 8 the virtual agent is greeting with her hand.

Figure 9 presents a virtual character performing the "astonishment", "confusion" and "displeasure" emotions after having improved them with the gestures of touching the neck, the forehead and folding the arms, respectively.

Fig. 7. Virtual character refusing something by moving the forefinger.

Fig. 8. Virtual character greeting with the hand.

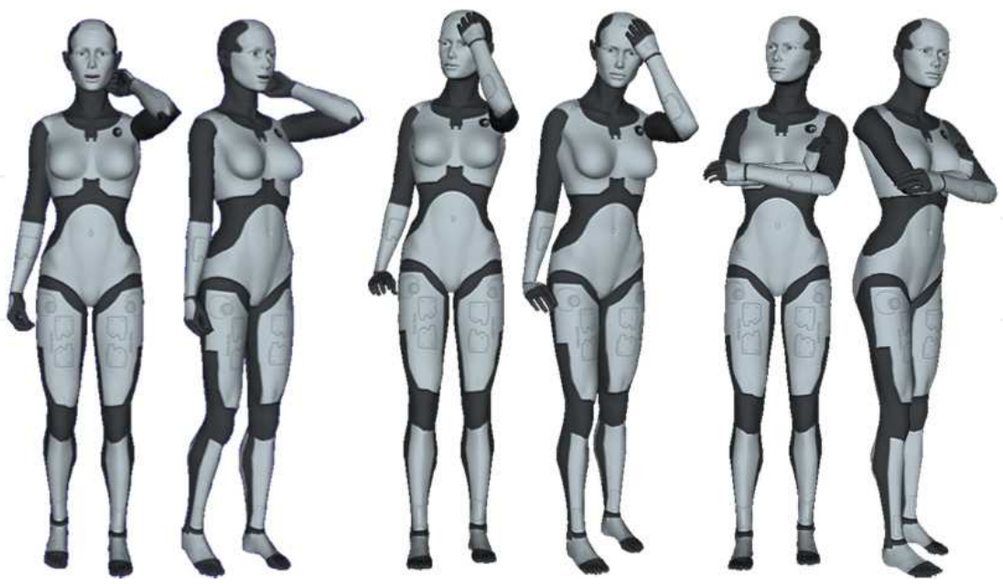


Fig. 9. "Astonishment" emotion was improved with the touch one self's neck gesture (left), "confusion" emotion was improved with the touch one self's forehead gesture (middle), and "displeasure" emotion was improved with the arms-folded gesture.

4. An affective virtual interpreter to Spanish Sign Language

The features previously described have been used in the development of an application that uses emotional virtual characters as virtual interpreters, capable of translating from written or spoken Spanish language to Spanish Sign Language (LSE).

In the last few years, the design of computer application interfaces has evolved in order to guarantee the accessibility of applications to everyone. Regarding the deaf community, a considerable amount of work has been done in the automatic translation into sign languages. These languages, unfortunately, are not universal and each country has its own variety. In fact, most of the work done (Ong & Ranganath, 2005) is based on English grammar. This is the case of the works derived from ViSiCAST (Visicast, 2011) and eSIGN (ESIGN, 2011) projects. Regarding Spanish Sign Language (LSE), San-Segundo et al. (San-Segundo et al., 2008) have developed a translator based on Spanish grammar that uses VGuido, an eSIGN avatar, but their application domain is very restricted (sentences spoken by an official when assisting people who are applying for their Identity Card). None of the previous works take into account emotional states. But, as in face-to-face communication, mood, emotions and facial expressions are an integral part of sign languages (Olivrin, 2008). Words can considerable change their meaning depending on the mood or emotion of the speaker. Moreover, communicating in sign language without facial expressions would be like speaking in a monotonic voice: more boring, less expressive and, in some cases, ambiguous. The system presented in this paper is based on Spanish grammar, takes the emotional state into account and emotions are reflected in the signs performed by an affective virtual character.

4.1 The automatic translation system

An automatic translation system from phrases in Spanish into LSE was developed as an independent module in C++ language. The system considers the syntactical and morphological characteristics of words and also the semantics of their meaning. The translation of a sentence or phrase is carried out by four modules (see Figure 10).

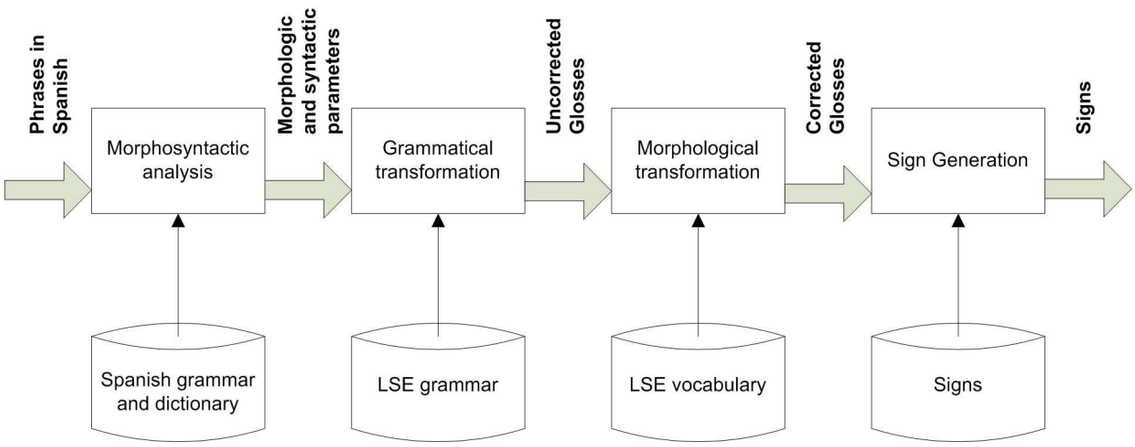


Fig. 10. Automatic translation system developed.

- **Morphosyntactic analysis**: A phrase in Spanish is used as input. A series of parameters containing all the morphological information of the words as well as the relations and

syntactical dependencies among them are drawn from it. This module uses the FreeLing analyzer (FreeLing, 2011), which was migrated to the Windows system.

- **Grammatical transformation**: On the basis of the syntactic information gathered during the previous step, and through the application of grammatical rules, this module generates a series of glosses.
- **Morphological transformation**: Some of the glosses resulting from the previous step could be incorrect. This may occur when the original word in Spanish has no direct correlation to a term in LSE. Sometimes a synonym of that term will correlate directly to a term in LSE, but it can also occur that several signs are required in LSE to render a single word in Spanish. Or sometimes in Spanish several words can be used to express an idea that LSE expresses in a single sign. So, in this step, all the changes in the words are implemented, resulting in grammatically correct glosses.
- **Sign generation**: Once the appropriate glosses have been produced (those which correspond directly to signs), in this step they are translated into a representation format that allows to generate the appropriated animations.

4.2 Adding emotions to the translation process

The emotional state of deaf person influences the way that that person will communicate, just like anyone else. On the one hand, the construction of phrases and the signs that make them up are modified and on the other hand, the final realization of the signs is also modified. In the fully developed system, the inclusion of the user's emotional state provokes changes in two translation phases: grammatical rules are modified, and so is the final transformation of the signs.

- **Grammatical transformation:** As we have seen, each type of block is associated to a series of rules pertaining to its form and function. These procedures are modified to change the way of generating the translations according to emotional swings. Emotion influences meaning and leads to the repetition of certain words, such as the nucleus, or to the appearance of new ones (similar to question tags or pet expressions). However, it can also be the case that certain blocks alter the order of the words within them to emphasize some of them.
- **Sign generation:** Emotion also influences the way in which specific words are signed. Thus, for example, the word "*no*" can be accompanied by different gestures. When the person signing is happy, he or she will move their finger, but if the person is angry, he or she usually resorts to dactylology and signs "N-O". In order to take these cases into account, the dictionary used for final translation of the glosses into the chosen language of representation has been modified, allowing one and the same word to be translated differently depending on the emotional state parameter.

4.3 Managing the virtual interpreter

The automatic translation system previously described has been incorporated in Maxine's system within the Deliberative/Generative Module. The final management of the virtual interpreter is carried out by the Motor Module generating animations that take the face, body, gestures and emotions into account.

The inputs of Maxine can be either written or spoken Spanish, generating phrases, that are passed to the translation system, which returns the signs that must be animated. Thanks to

Maxine's capabilities, the user's emotional state can be captured by a webcam be supplied to the automatic translator to be taken into account. The output consists in an affective virtual interpreter playing the final signs (animations) corresponding to the translation into Spanish Sign Language.

Figure 11 shows images of the virtual interpreter signing interrogative particles in LSE (how, which, how many, why, what and who), in which not only the facial expressions but the position and movement of the hands, and the posture of other parts of the body are involved.

The results of considering the emotional state in the translated signs can be seen in next figures. Figure 12 and 13 show the different ways to sign the word "NO": if happy she/he just moves the head (Figure 12 left) or the head and the forefinger (Figure 12 right), but when she/he is very angry, the face reflect the emotional state and dactylology is used in order to reaffirm the message, signing N-O (see Figure 13).

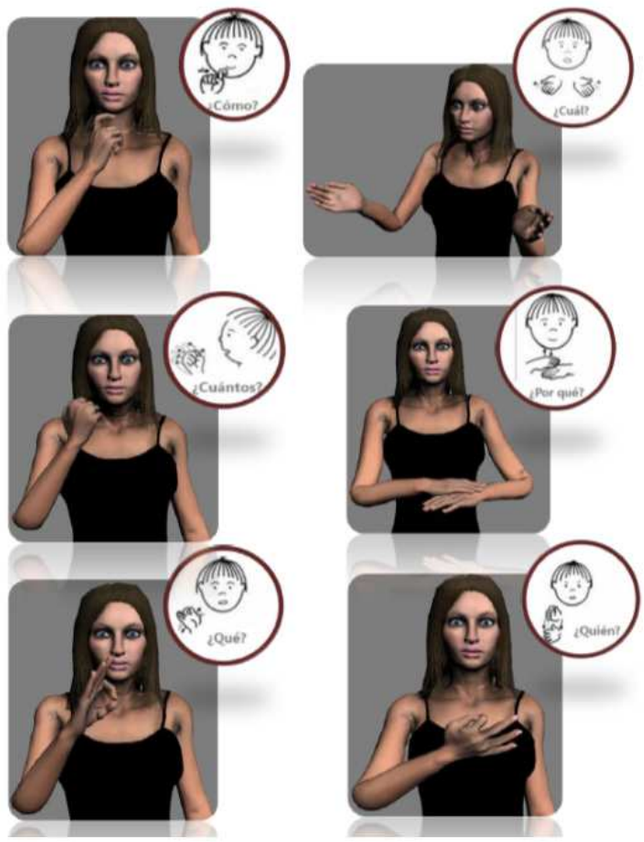


Fig. 11. Affective virtual interpreter signing different interrogative particles: how, which, how many, why, what and who (from left to right and from up to bottom).

Usually, affirmation is also signed in different ways depending on the emotional state of the person: if happy she/he just nods (see Figure 14), but if she/he is very angry, dactylology is used, signing S-I (YES in Spanish), as it is shown in Figure 15.

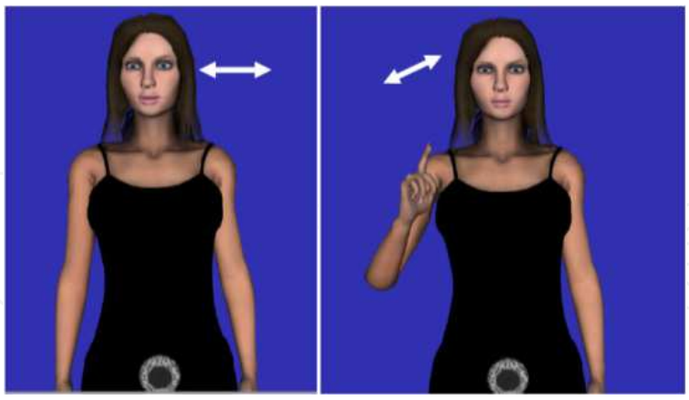


Fig. 12. The affective virtual interpreter says NO, when she is happy, shaking her head (left) or moving the forefinger (right).

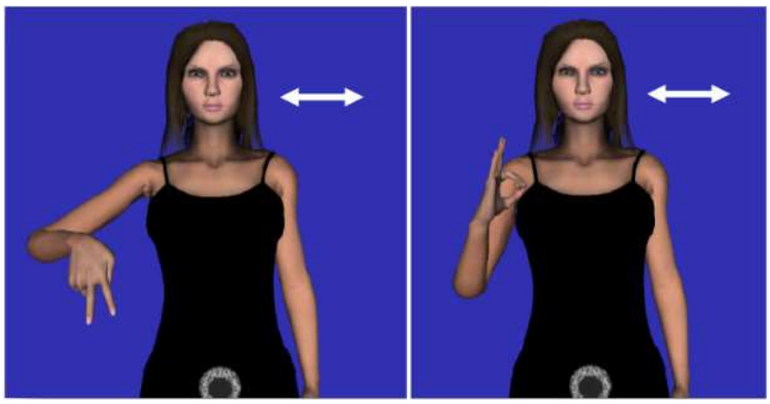


Fig. 13. The affective virtual interpreter says NO, when she is angry, shaking her head, through her facial expression and using dactylology.

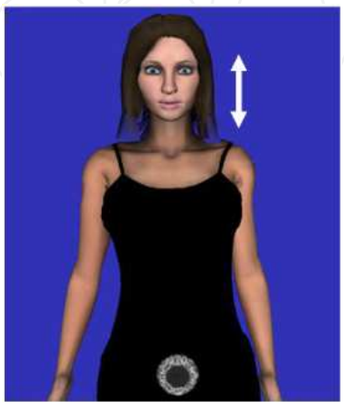


Fig. 14. The affective virtual interpreter says YES, when she is happy, nodding her head.

Fig. 15. The affective virtual interpreter says YES, when she is angry, nodding her head, changing her facial expression and using dactylology S-I (in Spanish).

The results generated with our system have been validated video-recording a real interpreter signing different sentences and comparing them with the same sentences performed by the virtual interpreter in order to verify the visual quality of the sign animations[1], as can be seen in Figure 16.

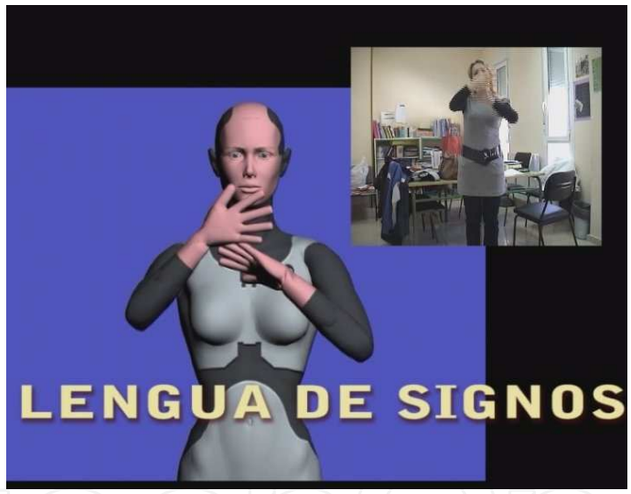


Fig. 16. Virtual and Real interpreters, signing the words "Sign Language" (in Spanish).

5. Conclusion

In this paper we have described the new features added to the Maxine system that allows multimodal interaction with users through affective embodied conversational agents. The virtual characters were already provided with facial expressions, lip-synch and emotional voice. However, in order to improve and obtain more natural human-computer interaction, Maxine virtual characters have now been endowed with new non-verbal communication capabilities: they are able to express their emotions through their whole body, by means of body postures and movements (considering basic and mixed emotions) and/or by corporal gestures.

[1] Videos of the animations can be found at our web page: http://giga.cps.unizar.es/affectivelab/video2.html

The potential of the affective agents provided with these new body language capabilities is presented through an application that uses virtual characters in order to enhance the interaction with deaf people. An emotional virtual interpreter translates from written or spoken Spanish language to Spanish Sign Language (LSE). The final signs performed take into account the interpreter's emotional state.

In the near future we will focus on enriching the interaction between the virtual character and the users including personality models for the virtual character and modifying the expression of emotions accordingly.

6. Acknowledgment

This work has been partly financed by the University of Zaragoza through the "AVIM-Agentes Virtuales Inteligentes y Multimodales" project and the Spanish DGICYT Contract Nº TIN2011-24660.

The authors want to thank the collaboration in this work of Lorena Palacio and Laura Sanz.

7. References

ALICE. (2011). Artificial Intelligence Foundation, Available from http://www.alicebot.org/

Baldassarri, S.; Cerezo, E. & Seron, F. (2008). Maxine: A platform for embodied animated agents. *Computers & Graphics* 32(3): pp. 430-437.

Baldassarri, S.; Cerezo, E. & Anaya, D. (2009). Emotional Speech Synthesis in Spanish for Natural Interaction. *New Trends on Human-Computer Interaction*, Chapter 15, pp. 151-160, Springer-Verlag.

Birdwhistell, R. L. (1985). *Kinesics and context: essays on body motion communication*. 4th Ed. Philadelphia: UPP. University of Pensylvania Press.

Brislin, R. (1993). *Understanding culture's influence on behavior*. Hartcout Brace College Publishers, New-York.

CAL3D. (2011). Character Animation Library. Available from http://cal3d.sourceforge.net

Cambria, E.; Hupont, I.; Hussain, A.; Cerezo, E. & Baldassarri, S. (2011). Sentic Avatar: Multimodal Affective Conversational Agent with Common Sense, *Lecture Notes in Computer Science, LNCS 6456: Toward autonomous, adaptive, and context-aware multimodal interfaces: theoretical and practical issues*. Springer-Verlag, pp. 81-95

Caridakis, G; Raouzaiou, A.; Karpouzis, K. & Kollias, S. (2006). Synthesizing Gesture Expressivity Based on Real Sequences. *LREC Conference. Workshop on multimodal corpora: from multimodal behaviour theories to usable models.*

Cassell, J. & Bickmore, T. (2000). External manifestations of trustworthiness in the interface. *Communications of the ACM*, 43(12), pp. 50-56.

Cassell, Justine (2007). "Body Language: Lessons from the Near-Human" In J. Riskin (ed.) *Genesis Redux : Essays in the History and Philosophy of Artificial Intelligence* . Chicago: University of Chicago Press., pp 346-374.

Cerezo, E.; Hupont, I.; Manresa, C.; Varona, J.; Baldassarri, S.; Perales, F. & Seron F. (2007). Real-time Facial Expression Recognition for Natural Interaction. *Lecture Notes in Computer Science, LNCS 4478: Pattern Recognition and Image Analysis*. Springer-Verlag, pp. 40-47.

Coulson, M. (2004). Attributing emotion to static body postures, *Journal of Nonverbal behavior* 28 (2), pp. 117-139

Cowie, R.; Doublas-Cowie, E.; Tsapatsoulis, N.; Vostis, G.; Kollias, S.; Fellenz, W. & Taylor, J.G. (2001). Emotion recognition in human computer interaction. *IEEE Signal Processing Magazine*, 1, pp. 32-80.

CyN Project. (2011). Available from http://www.daxtron.com/cyn.htm

de Rosis, F.; Pelachaud, C.; Poggi, I.; Carofiglio, V. and 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.

Ekman, P. (1999). *Facial Expression, The Handbook of Cognition and Emotion*. John Wiley & Sons

ESIGN. (2011). Available from http://www.sign-lang.uni-hamburg.de/esign/

FreeLing. (2011). Available from http://garraf.epsevg.upc.es/freeling

Gunes, H.; Piccardi, M. & Pantic, M. (2008). From the lab to the real world: Affect Sensing recognition using multiples cues and modalities. *Affective Computing: Focus on Emotion Expression, Synthesis and Recognition*, pp. 185-218

Gunes, H.; Schuller, B.; Pantic, M. & Cowie, R. (2011). Emotion representation, analysis and synthesis in continuous space: A survey. *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition (FG'11), EmoSPACE 2011*, pp.827-834

Hartmann, B.; Mancini, M. & Pelachaud, C. (2005). Implementing expressive gesture synthesis for embodied conversational agents. *In Gesture Workshop (GW'2005)* pp. 188-199

Haviland J. B. (2004). Pointing, gesture spaces, and mental maps. In *Language and gesture*. Cambridge University Press.

Hupont, I; Ballano, S; Baldassarri, S. & E. Cerezo. (2011). Scalable Multimodal Fusion for Continuous Affect Sensing. *IEEE Symposium Series in Computational Intelligence 2011, (SSCI 2011),* pp. 1-8.

Ji, Q., Lan, P. & Looney, C. (2006). A probabilistic framework for modeling and real-time monitoring human fatigue, *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 36, pp. 862–875.

Kapoor, A.; Burleson, W. & Picard, R. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies*, vol. 65, pp. 724–736.

Long. B. (2002). Speech Synthesis and Speech Recognition using SAPI 5.1. In: *European Borland Conference*, London.

Mancini, M. & Pelachaud, C. (2009). Implementing distinctive behavior for conversational agents, *in Gesture-Based Human-Computer Interaction and Simulation*, Lecture Notes in Computer Science, pp. 163-174.

Morris, D.; Collett, P.; Marsh, P. & O'Shaughnessy M. (1979). *Gestures, their origin and distribution*, Jonahtan Cape, London UK.

MPEG4 (2011). http://cordis.europa.eu/infowin/acts/analysys/products/thematic/mpeg4/coven/coven.htm

Noma, T. & Badler, N. (1997). A virtual human presenter. *Proceedings of IJCAI, Workshop on Animated Interface Agents*. pp. 45-51.

Olivrin, G. (2008). Composing Facial expressions with signing avatars, *Speech and Face-to-Face Communication Workshop*, Grenoble, France.

Ong, S. & Ranganath, S. (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning, *IEEE Trans Pattern Analysis and Machine Intelligence*, Vol 27 (6), pp. 873-891

Pantic, M. & Bartlett, M. (2007). Machine Analysis of Facial Expressions. *Face Recognition*, In-Tech Education and Publishing, Vienna, Austria, pp 377-416.

Picard, R.W. (1997). *Affective Computing*. The MIT Press.

Plutchik R. (1980). *Emotion: a psychoevolutionary synthesis*. New York: Harper & Row.

Prendinger, H. & Ishizuka M. (2001). Simulating Affective Communication with Animated Agents. *Proceedings of the Eighth IFIP TC.13 Conference on Human-Computer Interaction*, pp. 182-189

Raouzaiou, A.; Karpouzis, K. & Kollias, S. (2004). Emotion Synthesis in Virtual Environments. *In Proceedings of 6th International Conference on Enterprise Information Systems.*

Rickel, J. & Johnson W. L. (2000) Task-oriented collaboration with embodied agents in virtual worlds, In J. Cassell, S. Prevost, E. Churchill and J. Sullivan (eds.), *Embodied Conversational Agents*. MIT Press, pp. 95-122.

San-Segundo, R.; Barra, R.; Córdoba, R.; D'Haro, L.F.; Fernández, F.; Ferreiros, J.; Lucas, J.M.; Macías-Guarasa, J.; Monero, J.M. & Pardo, J.M. (2008). Speech to sign language translation system for Spanish, *Speech Communication*, Vol 50, pp. 1009-1020

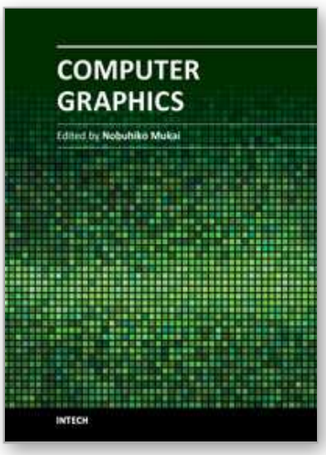
Su, W.; Pham, B. & Wardhani, A. (2007). Personality and emotion-based high-level control of affective story character. *IEEE Transactions on Visualization and Computer Graphics* 13, pp. 284–287

Visicast. (2011). Available from http://www.visicast.sys.uea.ac.uk/

Whissell C. M. (1989). The dictionary of affect in language, *Emotion: Theory, Research and Experience*, vol. 4, The Measurement of Emotions, New York: Academic Press.

Yeasin, M.; Bullot, B. & Sharma, R. (2006). Recognition of facial expressions and measurement of levels of interest from video. *IEEE Transactions on Multimedia*, vol. 8, pp. 500–508.

Zhang, S.; Wu, Z.; Meng, H. & Cai, L.(2010). Facial Expression Synthesis Based on Emotion Dimensions for Affective Talking Avatar. In *Modeling Machine Emotions for Realizing Intelligence*. Springer Berlin Heidelberg, pp. 109-132



Computer Graphics

Edited by Prof. Nobuhiko Mukai

ISBN 978-953-51-0455-1
Hard cover, 256 pages
Publisher InTech
Published online 30, March, 2012
Published in print edition March, 2012

Computer graphics is now used in various fields; for industrial, educational, medical and entertainment purposes. The aim of computer graphics is to visualize real objects and imaginary or other abstract items. In order to visualize various things, many technologies are necessary and they are mainly divided into two types in computer graphics: modeling and rendering technologies. This book covers the most advanced technologies for both types. It also includes some visualization techniques and applications for motion blur, virtual agents and historical textiles. This book provides useful insights for researchers in computer graphics.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Sandra Baldassarri and Eva Cerezo (2012). Maxine: Embodied conversational agents for multimodal affective communication, Computer Graphics, Prof. Nobuhiko Mukai (Ed.), ISBN: 978-953-51-0455-1, InTech, Available from: http://www.intechopen.com/books/computer-graphics/maxine-embodied-conversational-agents-for-multimodal-affective-communication

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen