International Journal of Human-Computer Studies 70.4 (2012): 271-286

DOI: http://dx.doi.org/10.1016/j.ijhcs.2011.11.003

Synthesizing Mood-Affected Signed Messages: Modifications to the Parametric Synthesis

Fernando López-Colino[1], José Colás

Human Computer Technology Laboratory
Department of Technology for Electronics and Communications
Universidad Autónoma de Madrid

Abstract

This paper describes the first approach in synthesizing mood-affected signed contents. The research focuses on the modifications applied to a parametric sign language synthesizer (based on phonetic descriptions of the signs). We propose some modifications that will allow for the synthesis of different perceived frames of mind within synthetic signed messages. Three of these proposals focus on modifications to three different signs' phonologic parameters (the hand shape, the movement and the non-hand parameter). The other two proposals focus on the temporal aspect of the synthesis (sign speed and transition duration) and the representation of muscular tension through inverse kinematics procedures. These resulting variations have been evaluated by Spanish deaf signers, who have concluded that our system can generate the same signed message with three different frames of mind, which are correctly identified by Spanish Sign Language signers.

Keywords: Sign Language Synthesis, Frame of Mind Synthesis, Graphical Interface, Deaf People

Email addresses: fj.lopez@uam.es (Fernando López-Colino), jose.colas@uam.es (José Colás)

[1]Corresponding Author. Tel.: +34914973613

1. Introduction

Since the last decade there has been an increasing interest in computer-animated signing avatars. Using different techniques, systems have focused on representing signed messages. Currently, all sign language (SL) synthesis efforts are focused on generating neutral signed messages. There is no work that studies how emotional signed messages can be synthesized; this line of research must be established.

To start this line of research it is necessary to answer this question: How are emotions transmitted among humans? Coulson states, "The general conclusion is that emotion, insofar as the term refers to the six 'basic' emotions (anger, disgust, fear, happiness, sadness, and surprise) is accurately perceived through the face and voice [...] Recognition of emotion from posture is comparable to recognition from the voice, and some postures are recognized as well as facial expressions" (Coulson, 2004). Mehrabian (1981) states that the expression of emotion during communication relies on both visual aspects (appearance, facial expression and body language) and speech-related aspects (mainly the way the voice is used and, to a lesser extent relevance, the message itself). However, how is emotion specifically transmitted through SL? Obviously, its transmission through voice inflexions is not used among deaf people, which leaves facial expressions and body postures as the only ways to transmit this information, according to these authors. Both facial expression and body posture are a part of the phonologic descriptions of signs, and changing a facial expression to represent an emotion, can modify the meaning of a sign. Therefore, the common approach used to represent emotions in general-purpose avatars cannot be directly applied to a signing avatar. Let us return to an interesting point made by Coulson: "emotions are perceived through voice." Speech is the natural way of communicating among those who can hear, and modifications to speech are used to communicate emotions. Sign language is the natural way for the Deaf people to communicate; is it possible to use a similar approach, including some modifications to SL, to communicate paralinguistic information without modifying the message?

The emotional state of the speaker is reflected by means of variations in prosody. Prosody is usually defined as the rhythm, stress and intonation of speech. However, this is not only applicable to speech; it is also applied to signed communication (Sandler and Lillo-Martin, 2006). In this work, we do not deal with the representation of the basic emotions (Ekman, 1971, 1999) in signed messages. This work does not attempt to emulate how human signers represent affect. We begin the proposed line of research by focusing on stress in SL (Wilbur and Schick, 1987; Wilbur, 1999). We present a study of the modifications that can be applied to SL synthesis to represent different degrees of stress without modifying the message:

The relation between stress and muscular tension was described by Wahlstrom et al. (2003). In this work, we evaluate if particular modifications to the animation of a signing avatar are perceived as being produced by different muscular tensions (which can be related to different stress levels). We have studied if modifications to hand shapes can represent different muscle tensions without modifying the represented phoneme. We have also studied if different elbow positioning, for a defined hand position, are correctly perceived as different muscle tensions in the shoulder complex. The modification observes the position of the hand because it is phonologically relevant.

We have also selected from previous research on visual prosody three modifications that can be included in the synthesis process without interfering with SL phonology (facial expression, movement acceleration and spatial extension, and speed). We have evaluated if the inclusion of these three modifications and the previously stated (hand muscle tension, shoulder complex tension) are perceived as different stress levels, without modifying the content of the message.

This paper is structured as follows. We will review the most relevant work related to visual prosody and SL synthesis in Section 2. Next, in Section 3, we will present our proposal on how mood-affected synthetic signed messages can be created. Section 4 will discuss how the synthesizer modifies the synthesis procedure to create different variations of signed messages. Section 5 will describe the experiments and will present the obtained results to verify individual hypotheses and the evaluation of mood-affected signed messages by native signers. Finally, we will discuss our findings and will suggest future work in Section 7.

2. Background and Related Work

This work focus on two main topics: SL synthesis and the work related to the synthesis of mood-modified gesture. First, we will discuss the different techniques used for SL synthesis, particularly the technique we have used in our system, which is based on phonetic descriptions (the specific details of the synthesizer are presented in section 4). Then we will review the literature related to the studies which deal with the perception of visual prosody.

2.1. Sign Language Synthesis

There are two main approaches to SL synthesis, which are similar to those used for speech synthesis: that based on the concatenation of pre-recorded chunks and that based on abstract definitions of message units. The first approach may use video recordings of human signers (Solina et al., 2001) or animations applied to virtual avatars (VCom3D, 2009; Segouat and Braffort, 2009). These animations are defined manually by expert animators. Although this approach provides more human-like results, the sign language's phonology and its inflections make this approach unfeasible. However, some authors, like Huenerfauth (2009), use this approach to evaluate different characteristics of synthetic signed messages, such as speed, timing and pause duration and their impact on message intelligibility. The most widely used approach to SL synthesis is based on abstract definitions of signs. These definitions are described using different graphical notations, such as HamNoSys (Prillwitz et al., 1989; Hanke, 2004) and SignWriting (Sutton, 1974), or alphanumeric notations, such as SEA (Herrero Blanco, 2004) or the Szczepankowski notation (Szczepankowski and Rona, 1994; Szczepankowski, 1999). These notations describe the basic units that compose each sign. However, like text, they do not describe a signer's mood or prosody.

SLs define different elements within a signed message: the fingerspelling dictionary, established signs, classifier constructions, etc. Although there are many applications that allow for the creation of a fingerspelling animation to spell a word (Segouat, 2009), this is not how deaf people communicate with each other. The number of applications that fully resolves this problem, with respect to SL synthesis, is reduced. Signs are composed of several sequences of phonologic parameters. We have mentioned before that there are many notations that allow for the description of signs at this phonetic level. The literature provides several examples of projects that use these notations for SL synthesis. The ViSiCAST/eSign/Dicta–Sign projects (Kennaway et al., 2007; Elliott et al., 2008; Elliot et al., 2010; Jennings et al., 2010) have defined an XML-based notation, SiGML (Elliot et al., 2004), which is based on the HamNoSys notation and some of its characteristics. The SWML notation (Rocha and Pereira, 2004) follows an approach similar to that of SignWriting, defining an equivalent XML-based notation. Other projects, although based on HamNoSys notation to describe the signs, use a different approach. Instead of using SiGML, Fotinea et al. (2008) describe a module that transforms the HamNoSys definitions into the STEP notation[2] (Huang et al., 2002). This synthesis module was previously used for an educational application developed by Karpouzis et al. (2007).

Independent of notation, these projects use phonetic sign descriptions to generate an animation that, when applied to a virtual avatar, will represent the signed message. The most widely used 3D API is VRML (ISO/IEC 14772-1:1997, 1997; ISO/IEC 14772-2:2004, 2004), which allows for the representation of 3D content on desktop computers and web browsers. Most projects use the H-Anim skeleton structure ISO/IEC 19774:2005 (2005), a standard for human representation on VRML. In this manner, the avatar's animation management is simplified. Although the ViSiCAST project uses a newly developed structure, its skeleton structure is very similar to that of H-Anim.

These synthesizers are usually language-independent, as shown by San-Segundo et al. (2008). This project, which has focused on Spanish Sign Language (LSE), uses eSign synthesizer, which was developed for British and German Sign Language. However, "signing styles" are different between different cultures, and reusing SL synthesizers requires some kind of adaptation.

2.2. Visual prosody

The second main focus of our work is visual prosody, the study of facial expressions and body postures and the perception of emotions from them.

When talking about visual representations of emotions, the first thoughts that come to mind are concerned with facial expressions (Mehrabian, 1981). In the human-computer interaction context, synthetic faces are fundamental for embodied agents. Most extended embodied agents are displayed as talking heads (Beskow, 2004); their basic implementation is focused only on visual speech synthesis by means of lip movements. However, the studies performed by Brave et al. (2005), using real faces as embodied computer agents, demonstrated that including emotion representation capabilities enhanced users' experience. Therefore, including emotion representation in the virtual avatar-based interfaces will improve the human-computer interaction. The

[2]This notation is a general purpose animation notation that allows defining the animation of an H-Anim compliant avatar. This notation focuses on joint rotation definition, so presents a lower level of abstraction compared to SiGML.

literature presents several techniques for defining facial animations (Sloan et al., 2009; Cao et al., 2005; Gralewski et al., 2004); other authors have focused on improving concrete facial aspects such as the eyes (Xiong et al., 2010). We have mentioned that synthetic faces are mainly used in conversational embodied agents. Indeed, there are several works that focus on modeling the relationship between various speech content, different facial expressions and the affective states (Cassell et al., 1994; de Rosis et al., 2003) or even how these interfaces can benefit social groups with special needs (Cole et al., 2003).

It is also worth mentioning robotic faces, such as the one presented by Breazeal (2003). The artificial synthesis of emotive facial expressions is not limited only to virtual avatars; it is also relevant to robotic faces. The work of Cassell (2000), who described the role of gestures in human-human communication, describes how gestures can be indicators of a speaker's emotional and cognitive state. For example, bending the head to one side may express doubt and raised eyebrows usually indicate surprise or disbelief. This aspect of gesture perception has been studied by several authors. Berthouze et al. (2003) studied how humans recognize affective postures. Their results suggested that angry, happy and sad animations are the easiest to recognize; they also reported that modifications to the animation speed is very relevant for affect recognition from body postures. Coulson (2004) presented an experiment using a wooden mannequin capable of representing different body postures; users were asked to identify the emotion represented by each pose. The results showed which sections of the body were the most relevant in representing each of the six basic emotions. Coulson concluded, with an interesting reflection about his experiment, that: "*... static images may not represent an appropriate medium within which dynamically expressed emotions can be communicated.*" Following the same approach for evaluating static body postures, Shaarani and Romano (2007, 2008) extended Coulson's experiments by improving the quality of the test images. They also studied the intensity perceived by users for each emotion and focused especially on happy postures (because is the most easily recognized emotion). The work presented by Neff and Fiume (2006), which focused on the modifications made to the inverse kinematics algorithms to generate different poses to modify the expressiveness of a stance, is also interesting. Finally, the results obtained by McDonnell et al. (2008, 2009) are important to our work. They studied the relation between the perception of emotion in virtual characters and their shape and visual aspects, concluding that they are independent of each other.

With a focus related to Coulson's conclusion (previously stated), Hartmann et al. (2005a) studied different aspects of motion that can be related to the perception of expressiveness in virtual avatars. Their results show that users can identify different frames of mind from an avatar's movements. However, in their study, not every expression was correctly identified by users. By applying this research to embodied conversational agents, Noot and Ruttkay (2005) developed the GESTYLE notation. This notation allows for the definition of different characteristics of an avatar's personality to define animation style and performance during a conversation. Castellano and Mancini (2009) developed an interesting system that enabled an avatar to "mimic" a user's emotional expressions. This system mainly focused on three emotions (anger, joy and sadness). The recognition levels of the gestures performed by the avatar were, for some emotions, greater than the gestures performed by actors.

These studies were performed on conversational embodied agents, developed as complements to speech-based interaction. When applied to SL, the gestures of conversational embodied agents cannot be considered complementary because they are used for the conversation itself.

3. Modifications to the Parametric Synthesis

Signs are phonologically described by seven Phonologic Parameters (PPs) (Corina, 1996; Muñoz Baell, 1999): configuration, or hand shape; orientation, or hand absolute orientation; location, hand positioning within the frontal plane, defined by means of anatomic references; plane, horizontal distance between the hand and the body; contact point, defining the active point of the hand (usually a finger joint) that is used for hand positioning and interaction; movement, or hand displacement; non-hand parameter, comprising facial expressions and body postures.

How can we compare this with speech? When synthesizing speech, different parameters are modified to represent prosody and mood: pitch, tone, speed, etc. These changes do not modify the message, which is still recognizable. Mood-affected SL synthesis follows the same approach: modifying the performance of individual PPs to represent stress.

The plane and location PPs define the position where the hands must be placed during signing. These parameters cannot be altered without modifying their definition because any alteration would imply a different hand position and produce a different sign. The same argument applies to the contact point PP. It defines the location within the hand that is used for positioning and interaction with the rest of the body. It cannot be modified without altering the meaning of the sign.

Different hand orientations can be used to represent different frames of mind. If we were to imagine someone pointing at some direction, the effect produced when his or her palm is faced upwards is different from that when the palm is pointing downwards. However, in SL, these different hand orientations are relevant to the definition of the sign; changing hand orientation denotes different signs. The minimum difference between two orientations is a 45° rotation around any of the $X-, Y-, Z-$axes.

The definition of the other PPs (configuration, movement and non-hand PPs) is not as strict as that for the previous ones. Is it possible that a slight modification of these PPs can be perceived as a mood variation in a synthetic message without altering its content?

3.1. Modifications to the Configuration PP (Modification M_1)

Different hand gestures are socially recognized as representing different frames of mind; a closed fist can be interpreted as an aggressive gesture and an open-hand shape as being pacifist (De Silva and Bianchi-Berthouze, 2004). However, hand-shape is phonologically relevant to the construction a sign; thus, a prosodic modification cannot change hand shape. Therefore, we propose introducing mood-related modifications into this PP by means of hand muscle tension. Variations in muscle tension slightly modify the angles of different finger joints, but the configuration phoneme is still the same. The variation of muscle tension also affects the vasoconstriction of the hand's veins. When the muscles are tense, the flow of blood is restricted to the hand, making the skin turn paler.

In this work we have focused on the effect of varying muscle tension: producing slight changes in the angles of finger joints. We have defined three variations (relaxed, standard and tense) for the seventy-four hand shapes that can be found in the LSE. Our hypothesis H_1, related to this PP, is as follows:

H_1: For a defined configuration phoneme, increasing the value of the joint angle for an extended joint and decreasing this value for a flexed joint will be perceived as an increase of the hand muscular tension. The opposite modification will be perceived as a decrease of the hand muscular tension.

The modifications proposed in this modification M_1 are detailed in Section 4. The experiment E_1 was designed to test H_1.

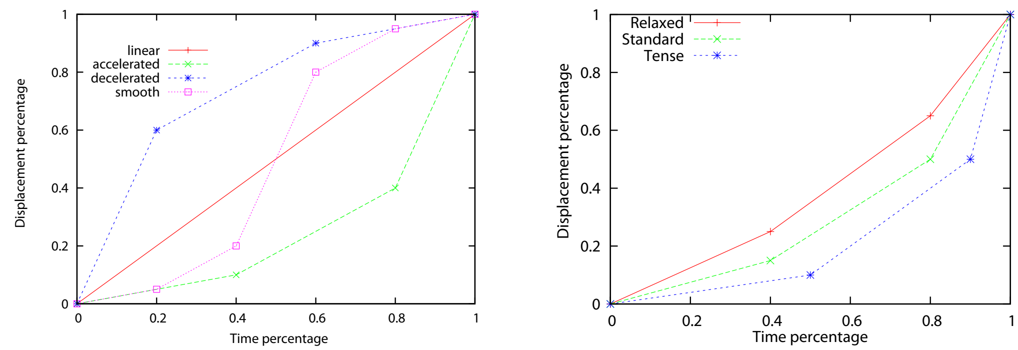
3.2. Modifications to the Movement PP (Modification M_2)

In SL, the tension required to produce a movement partially defines the meaning of a particular sign. As different hand shapes modify the meaning of a sign, altering this tension has the same effect. The movement phoneme is not only defined by its path and speed; describing the muscular tension (acceleration) involved is also required. In Figure 1(a) we represent space-time graphs of a movement, using different accelerations for the same linear movement. All of these movements are used as different movement phonemes of different signs.

A few authors describe how the tension of a movement represents different frames of mind (Neff and Fiume, 2002). Hartmann et al. (2005b) proposed five expressivity parameters: *Spatial Extent*, *Temporal Extent*, *Fluidity*, *Power* and *Repetition*. The duration of a movement is phonetically restricted, so the *Temporal Extent* modification is discarded. *Fluidity* cannot be reduced to only the movement PP, as there are other movements involved in signing[3]. *Repetition* is also phonetically and syntactically relevant (used to represent the plural of some signs). The other two parameters *Power* and *Spatial Extent* are also relevant to the description of a sign, though they allow for some variations. The trajectory of a movement can be slightly modified by making tense movements a bit shorter and relaxed movements longer. *Power* is also relevant, as the phonetic descriptions of signs include references to muscular tension of a movement. However, it is also possible (similar to the configuration PP) to include slight modifications to the acceleration of the movement.

Our second proposal is to include slight modifications to the movements in terms of the acceleration degree and spatial extension. An accelerated movement solely denotes an accelerated movement. However, if the signer's mood is tense, the movement will be more explosive and more accelerated than relaxed signing. The same applies to the spatial extension of a movement: slightly expanding it when the signer is relaxed and contracting it when the signer is tense. In Figure 1(b), we show the space-time graphs for

[3]Both *Temporal Extent* and *Fluidity* are considered in section 3.4 as whole sentence modifiers.



(a) The same linear movement altered by means of the acceleration. The modification defines different movement phonemes which will modify the meaning of a sign

(b) Slight modifications of the same movement tension to perform different frames of mind for the same movement phoneme

Figure 1: Space-time graphs of a linear movement using different accelerations; both space and time are represented using percentages of the total duration

an accelerated movement for different frames of mind. These two parameters were studied by Hartmann et al.. Therefore, we do not present any related experiments. However, we included this modification to the movement PP in the final experiment E_3. The modifications proposed in this modification M_2 are detailed in Section 4.

3.3. Modifications to the Non-Hand PP (Modification M_3)

The last PP we are considering for mood-based modification is the non-hand parameter. This PP is comprised of facial expressions, body postures, head tilts and gazes, mouth movements and eye gazes. Facial expressions and body postures have been studied within the context of visual prosody and some of them are presented in Subsection 2.2. The first approach could have used the results of the research previously discussed when synthesizing prosodic signed messages. In this case, as when dealing with the movement PP, there would be a conflict if we were to merge prosodic body postures and facial expressions into a signed message. Most experiments related to the study of static body postures and their perceived frames of mind involve the position of the hands and the arms. As we mentioned before, the position of the hands are meaningful in SL and are defined by the location, plane and movement PPs, so they cannot be modified without changing the meaning of the sign. The same applies to the positioning of the chest and shoulders; they are elements relevant to the non-hand PP and sign's phonology. Therefore, changes in body posture and hand positioning cannot be used to synthesize a mood-modified signed message.

Facial expression is also an element of the non-hand PP; it is a relevant part of a sign's phonology and SL syntax. Although it is the least studied and the most ignored PP by synthesis systems, it is the most relevant in signed communication (Lu et al., 1998). Caridakis et al. (2007), in particular, applied the five expressive parameters proposed by Hartmann et al. to facial expressions. Both *Repetition* and *Temporal Extent* are discarded because the repetition is linguistically relevant and the non-hand PP must be synchronized with all of the other PPs and global sentence timing. *Fluidity* is a global modification and, as we stated for the movement PP, it should be processed as a global modification. *Power* modifies the acceleration of a movement; this feature has been included within global modification which alters the transition between signs. Finally, modifying the *Spatial Extent* is not trivial, because modifying the spatial extent of the movements of different face features may change the meaning of a sign (e.g., the size of a signed object is assumed to be large when the cheeks are inflated, and even larger if they are extremely inflated). Hence, modifying an avatar's expression to represent a mood during signing cannot be done as an embodied agent's expression can. The modification M_3 limits the mood-affected variations of facial expressions to the eyebrows, eyes and mouth using the research of Caridakis et al. (2007) as base. We propose to define a different neutral facial expression for each mood-variation. The performance of the other elements of the non-hand PP is not modified.

The modifications proposed in M_3 are detailed in Section 4.

3.4. Other SL synthesis-related modifications (Modification M_4)

After some discussion with SL experts, we concluded that timing also affects the prosody of a signed message. *Temporal Extent* is one of the expressive parameters defined by Hartmann et al.. When a signer is tense, the duration of a global message is reduced, which implies an increase in the global speed of his or her sign and a decrease in the transition time between two consecutive signs. On the other hand, when a signer is relaxed, the opposite effect can be observed: lower signing speeds and higher transition times. Therefore, these two aspects were considered when introducing mood-related variations to synthetic signed messages.

We mentioned in Section 3.2, that *Fluidity* cannot be considered for the movement PP alone. The entire signing process includes many movements, although they are not strictly phonemes (Liddell and Johnson, 1989). Therefore, *Fluidity* can be applied as a mood-modifier, using long and fluid transitions to represent a relaxed sign and short, sudden and fast transitions when a signer is tense.

The modifications proposed in M_4 are detailed in Section 4.

3.5. Elbow swivel angle (Modification M_5)

In SL, hand positioning is defined by the location and plane PPs, which after processing, define a spatial coordinate. This coordinate specifies the place where the hand's joint, defined by the contact point PP, must be placed. The orientation PP defines a constraint for the wrist orientation. After processing these four PPs we obtain the coordinates where the wrist must be placed. The synthesizer uses an inverse kinematics algorithm to define the shoulder and elbow angles that make the hand reach the defined coordinates with the hand orientation constraint defined by the orientation PP. When the distance between the objective coordinate and the shoulder is smaller than the arm's total length, the number of available solutions to the inverse kinematics algorithm is virtually infinite. The elbow can rotate around the shoulder-wrist vector without modifying the wrist position; all of the coordinates that the elbow can reach describe a circle. Obviously, not all of the possible solutions to the algorithm are correct: the arm cannot intersect the thorax; the shoulder and the wrist joints must reflect their anatomical restrictions. However, after considering these restrictions, there is still a range of valid solutions for positioning the elbow. The position of the elbow can be defined using the elbow's swivel angle (Φ angle); the value of this angle is $0°$ when the elbow's height is the lowest. We also define the value of this angle as being positive when the elbow is in an exterior semicircle such that the body is in an inner semicircle.

The shoulder complex is composed by the scapula, the clavicle and the humerus and is controlled by five interdependent linkages (Levangie and Norkin, 2001). There are many muscle groups involved in the movement of this joint, including the deltoid, pectoral and dorsal muscles, among others. Different shoulder muscles are involved in the extension/flexion of the elbow, depending on the angle Φ. If Φ is low (near zero), only the triceps is involved, but if Φ is increased, the dorsal and pectoral muscles are used, increasing the strength of the extension/flexion movement[4]. For a stressed movement, muscle tension is increased, causing the muscles of the shoulder complex contract. Hence, Φ will be higher. Our hypothesis H_2 is as follows:

H_2: An increment in the value of the Φ angle, for a defined hand position, will be perceived as an increment in the muscular tension of the gesture. Decrementing the value of the Φ angle will be perceived as a decrement in the muscular tension.

The implementation of this modification M_5 is described in Section 4. Experiment E_2 (see Subsection 5.3) was used to test H_2; the last experiment (E_3) was also used to test H_2 with respect to the synthesis of signed messages.

4. Formalization and Implementation

The previously discussed modifications imply two different updates to our synthesizer (López-Colino and Colás, 2011): describing multiple allophones for the phonemes of the configuration, movement and non-hand PPs and altering the gesture generation process at two different stages, the inverse kinematics stage and when defining the timing of the key frames.

The synthesizer (López-Colino and Colás, 2011) uses a multilevel relational database to store the phonetic descriptions of the signs. This approach provides three main advantages over the existing phonetic notations (e.g., HamNoSys or SEA): a) including 1-to-n relation between phonemes and allophones. The allophones can be described using a set of bone orientations (configuration, orientation and non-hand PPs), anatomic references (location PP) or a set of spatial vectors (plane and movement PPs). b) storing the standard sign duration for every sign. This information is not made explicit in phonetic notations. The database stores the sign duration in its isolated form extracted from a LSE visual dictionary (Fundación CNSE, 2008). c) storing the internal syllable structure of every sign. This information describes different temporal segments, during this segments each PP may remain still or may be transitioning to a different phoneme.

4.1. Realization of the modified PPs

The modifications M_1, M_2 and M_3 proposed that the realization of three different PPs could be modified to introduce mood variations into a synthe-

[4]Consider the push-up exercise; it is harder when the swivel angle is low (elbow is touching the body) and easier when the swivel angle is increased.

sized message. These three modifications imply that for a sign's phonetic definition, different allophones should be retrieved according to the mood variation to be represented. We will now describe how these modifications are implemented in the synthesis process.

Modification M_1 proposed that the configuration PP can be used to represent stress variation. The same configuration phoneme describes different allophones corresponding to different muscle tensions applied to the hand. During the synthesis process, the synthesizer retrieves the phonetic description of the sign, and for each phoneme, the synthesizer retrieves the corresponding hand shape (a set of bone orientations applied to the finger joints). We have modeled and stored, for each phoneme, three different allophones corresponding to three different degrees of muscle tension (normal, tense and relaxed). These three variations were defined introducing the modifications proposed in H_1. The results of the experiment E_1 verified H_1, so the different allophones were correctly assigned to both the corresponding phoneme and the corresponding muscle tension.

Modification M_2 stated that the movement PP is also representative of a signer's mood. Although different researchers have formalized different mood-related parameters that modify movements, we have pointed out that only two of these parameters (*Spatial Extent* and *Power*) can be applied to the movement PP. We have described three different movement allophones for the phonemes used in the example signs. The 'tense' variation's trajectory has been shortened to 90% of the corresponding standard trajectory, whereas the 'relaxed' variation's trajectory has been expanded 110%. A movement's power has also been modified. We fixed the linguistic value of the movement PP's acceleration, so that a modification in a movement's power could not transform a smooth movement into an explosive one. As shown in Figure 1(b), the same movement phoneme, slightly modified, can represent different frames of mind. We have stored different movement definitions for the required movements. The synthesizer uses the corresponding allophone.

The non-hand PP is a bit different (M_3). We have stated that we cannot simply use different facial expressions, as they can introduce relevant phonologic information (e.g., raising the eyebrows defines an interrogative sentence, inflating the cheeks implies "bigger" and showing the end of the tongue implies "smaller"). Head and eye gazes are used for subject agreement. Any modification made in a concrete sign can modify its meaning or the whole sentence syntax. However, we have modified the neutral avatar's expression to different stress levels ('tense', 'standard' and 'relaxed'). Hence, any phonologic or syntactic modification is made mood-independent, but the neutral avatar's facial expression is different. The relaxed and tense facial expressions were selected using an opinion poll:

To select the relaxed and tense facial expressions, we conducted an opinion poll after Experiment E_1. We proposed E_1' users (see next section for details about the experiment) eight different facial expressions, which were quite similar to the neutral face, but with slight differences in the eyebrows, eyelids, nose and mouth (modifying the jaw and the lips). Using a 7-item Likert scale (1 = "very relaxed", 4 = "neutral" and 7 = "very tense"), we selected the three expressions to be used in the synthesis of mood-modified sentences. The selected facial expression for the relaxed variation (score = 2.96) is depicted in Figure 2(a); the standard variation is depicted in Figure 2(b) (score = 3.94); the tense face expression (score = 5.26) is presented in Figure 2(c).

(a) Relaxed (b) Standard (c) Tense

Figure 2: The avatar's neutral facial expressions for the three mood variations.

4.2. Sentence modifiers

Modification M_4 stated that a movement's *Temporal Extent* and *Fluidity* can be used as expressive parameters. Signing communication is based on "Movements and Holds" (Liddell and Johnson, 1989). However, not every movement is a movement phoneme, so these expressive modifications must be applied to the entire signing animation. The current work has only used the variations in *Temporal Extent* to represent expressiveness. *Fluidity* is closely related to the interpolation approach used by the rendering engine. The rendering engine that was used only implements linear and basic spline interpolation; to correctly represent the movement-hold model (Liddell and

Johnson, 1989) using smooth transitions, a Bezier interpolation approach is required. Hence, we have not included the *Fluidity* parameter.

The global signing speed is modified to represent mood variations. The synthesizer stores the standard sign duration and the internal structure (based on the movement-hold model). When the global signing speed is modified, every segment of the internal structure is modified as well. Therefore, the internal timing relation between different signing segments remains. The speed parameter has also been included in other synthesis platforms (Elliot et al., 2010).

The transition between signs is also modified for expressiveness: during the tense mood variation, the sign transition is shorter than the standard mood variation. Additionally, the transitions in the relaxed mood variation are longer than in the standard mood variation. The synthesizer uses this information to set the duration between the last segment of a sign and the first segment of the followign one. The transition time and the average signing speed are presented in Table 1.

Table 1: This table contains the defined duration for transitions between signs and the average signing speed of the different sign variations

Variation	Transition between signs	Average signing speed
Relaxed	$400ms$	83.43%
Standard	$300ms$	100.00%
Tense	$200ms$	134.80%

4.3. Inverse Kinematics

Inverse kinematics is used for hand positioning, due to the linguistic hand orientation restrictions (orientation PP), its position (obtained from location and plane PPs) and shoulder position (derived from the non-hand PP and the syntax); the elbow's swivel angle is the only parameter that allows multiple values. The synthesizer uses a trigonometric approach to provide an initial solution for the inverse kinematics. Hypothesis H_2 proposed that the elbow's swivel angle is related to the perceived muscular tension. We have established three different elbow's swivel angle functions, depending on which tension appearance we want to transmit. These three functions of the Φ angle are depicted in Figure 3. The system, after the modification, will use the Φ value that maximizes the corresponding function, while respecting the limitations on shoulder and wrist rotation limits.

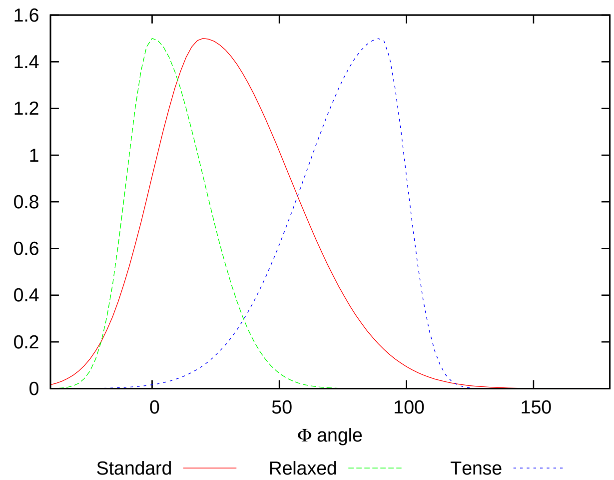


Figure 3: Shoulder complex defined by different degrees of muscular tension depending on the elbow's Φ angle value corresponding to the represented shoulder tension.

5. Individual Modification Experimentation

This section discusses the experiments we have conducted to validate or to refute H_1 and H_2, proposed after the modifications M_1 and M_5. Modifications M_2, M_3 and M_4 apply the results presented by other authors to SL synthesis. These modifications were introduced restricting the application of the expressive parameters to avoid interference with SL phonology. Proposing specific experiments for these three modifications would have been a repetition of the experiments conducted by previous research. Hence, we have included the modifications proposed by these three modifications M_2, M_3 and M_4, directly in the final experiment (section 6).

5.1. Method

The validity of our claims was tested through user evaluations, with respect to SL synthesis (Cox et al., 2003; Huenerfauth et al., 2008; Huenerfauth, 2008) and visual prosody (Hartmann et al., 2005a); user evaluation is the only approach to test a system or a hypothesis of this kind. The main difference between the two topics is the number of users who participate in the experiments. For studies related to visual prosody, it is usual to find between fifty and one hundred users to perform the experiment, but the evaluations related to SL synthesis are generally performed with six or ten users. This is justified because the people who can perform the second kind of experiments belong to a relatively small social group, namely, the Deaf community. The first three experiments can be performed by anyone, so it was easy to find up to fifty users. The last experiment, which is related to the modified SL synthesis (see section 6) and, required knowledge of SL (in our case, LSE), presented more difficult search for the required users (native LSE signers). The collaboration of the FCNSE[5] was fundamental for interacting with the LSE native users who performed the evaluations; some of these users were from the FCNSE linguistic department.

5.1.1. Experimental Design

These experiments were designed as within-subjects experiments to evaluate the perceived muscle tension, applying individual modifications. These two experiments were performed by the same group of fifty university students, researchers and staff (30 men and 20 women, 18 to 66 years old, with an average age of 38.2).

5.1.2. Procedure

All of the experiments were performed using web forms that contained a static image, a video or two videos related to the testing issue and a set of forced answers. All users registered for the study were provided with the URL to access each experiment. Participants were asked to complete each experiment within a week. The structure of all of the experiments was similar, differing only in the layout of the questions and the list of possible answers. The first page of each experiment contained some text that explained the experiment to the users and indicated how they should proceed. On the next page, participants began with the first question. The layout featured an image or video(s) on the left side of the screen. The right side formulated the particular question and, below it, the different possible answers. The system did not allow the users to leave a question unanswered.

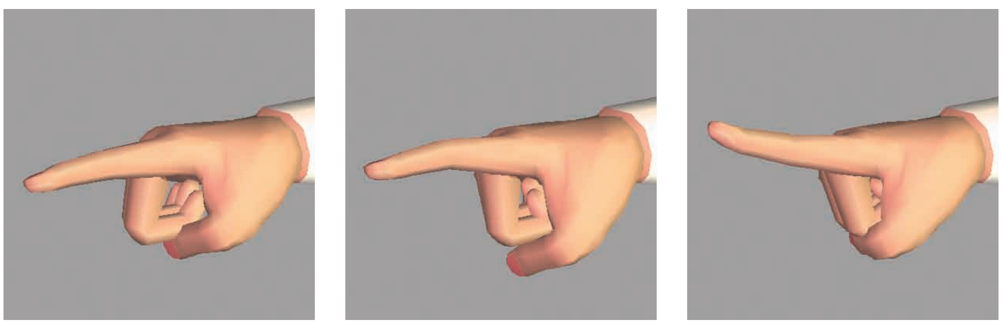
The evaluation method did not restrict the number of times a video could be replayed or the time each user had to answer each question. These elements were recorded by the system; however, they did not provide any relevant information and, thus, were not included in this paper.

[5]The *Fundación de la Confederación Estatal de Sordos de España* (FCNSE) is the national confederation of Spanish Deaf people. It defends Spanish Deaf people's rights and promotes the study of LSE and technologies related to deaf people's accessibility `http://www.fundacioncnse.org/`.

5.2. Evaluation of hand expressiveness (E_1)

The goal of the first experiment was to validate H_1. This experiment evaluated whether an avatar's hand could represent the same configuration phoneme using different muscular tensions by means of variations of the finger joint angle value (factor) and still be recognized as the same phoneme. For this experiment, we selected 10 configuration phonemes from those that are most commonly used in LSE. Every configuration phoneme was presented as a $250px \times 250px$ image, using the three defined levels of muscular tension: relaxed, standard and tense. To verify the other part of the hypothesis, all of the images were verified by an LSE expert, who asserted that all the variations represented the same configuration phoneme.

Each question proposed the three variations (see Figure 4), which were randomly ordered and labeled with a letter. The question asked the user to order them according to the perceived muscular tension. Half of the questions asked the subject to order the variations from the tensest to the most relaxed and the other half of the questions asked the reverse. To perform this task, the user could select from the 3! possible solutions.



(a) Relaxed (b) Standard (c) Tense

Figure 4: Three different representations of the "pointing hand" phoneme.

The results presented in Table 2 show 82% of the answers correctly identified the three muscular tension variations. The tense variation was also correctly identified by 12.2% of the answers, though with the standard and relaxed variations switched. The overall score of the evaluation was 78.4%, calculated subtracting the number of correct answers minus 20% of any incorrect answers. A Mann-Whitney test indicated that the independence against the null hypothesis was significant, $U = 0$, $p < .001$, $r = -.07$. We conclude that hypothesis H_1 is valid.

Table 2: Results of E_1. Left column shows each possible ordering for the tense (T), standard (S) and relaxed (R) variations. Right column shows the percentage of answers for each variation.

	Percentage of Answers
T >S >R	82 %
T >R >S	12.2 %
S >T >R	0 %
S >R >T	2 %
R >T >S	2 %
R >S >T	1.8 %

5.3. Evaluating the elbow swivel angle expressiveness (E_2)

To evaluate the validity of hypothesis H_2, we proposed a set of animations in which the avatar performed four linear movements with both hands: pushing front, pulling from the front, compressing a spring with both hands and pushing outwards (the avatar is between the pushed objects). These four movements were generated using the three different shoulder complex tensions applied to both arms: relaxed (low value of Φ), standard (medium value of Φ) and tense (high value of Φ). The avatar was expressionless and the movements were performed with the same amplitude and speed, so the only difference between them was the Φ angle (factor). The questionnaire contained 24 randomly ordered questions that compared two videos of the same movement, combining the different degrees of tension (4 movements $\times$ 3! tension degrees). Half of the questions asked the users to select which variation was tenser, and the other half asked the users to select the relaxed one. Every question had three possible answers ("the left one", "the right one" and "both are similar"). This formulation of the questions can be considered equivalent to a 3-point Likert scale if we presented the questions as "the left one is tenser, equal or more relaxed than the right one". The obtained results, depicted in Table 3, show that users clearly perceive the different values of the Φ angle as different degrees of tension. Users perceived the higher values of Φ as tense gestures. The overall result was 91.2%, calculated subtracting the number of correct answers minus 50% of incorrect answers. A Mann-Whitney test indicated that the independence against the null hypothesis was significant, $U = 0$, $p < .001$, $r = -.07$. We conclude that hypothesis H_2 is valid.

Table 3: Results of E_3, which compared two shoulder complex tensions. The symbol ✓denotes the correct answers, whereas the wrong answers have been divided between those who answered that both videos showed the same tension (=) or the opposite tension (✗).

	Compared to								
	Tense			**Standard**			**Relaxed**		
Variation	✓	=	✗	✓	=	✗	✓	=	✗
Tense	98 %			85.5 %	9 %	5.5 %	92.5 %	4 %	3.5 %
Standard				100 %			88.5 %	6 %	5.5 %
Relaxed							97 %		

6. Evaluating mood-affected SL synthetic messages (E_3)

The last experiment conducted was the evaluation of the mood-modified SL synthesis itself. In this experiment, we created synthetic sentences that included all of the modifications described above (see Section 3): the modifications to the configuration (M_1), movement (M_2) and non-hand (M_3) PPs; the timing variations in the synthesis (M_4); and the shoulder's complex tension by means of the angle Φ (M_5).

6.1. Experimental Design

This experiment was designed as a within-subjects experiment to evaluate the perceived stress on mood-modified SL synthetic messages. This experiment was performed by a group of 10 LSE native signers (7 men and 3 women, aged 20 to 62 years old, with an average age of 37.1 years). Four of these users work at the LSE linguistic department of the FCNSE. Their collaboration was not limited to performing this experiment, but also included evaluation of the synthetic signing and its correctness. All of the users in this group have a medium comprehension level for reading Spanish, which allowed us to use written instructions to perform the experiment. We mentioned above that evaluating synthetic signed messages requires the collaboration of native deaf signers and obtaining their collaboration is a difficult task. Although the number of deaf users who performed this experiment is low, it is similar to the number of users reported by other authors.

6.2. Experimental data

We proposed five sentences, shown in Table 4, that were translated into LSE by a FCNSE interpreter. These sentences were synthesized using our

SL synthesizer in five different variations. The 'standard' variation was generated without introducing any modification to the synthesizer. The 'relaxed' and 'tense' variations included the modifications related to M_1, M_2, M_3 and M_4: modifying the configuration, movement and non-hand PPs and the global signing speed as we stated in Section 4. The 'very tense' and 'very relaxed' variations also included the modifications related to M_5, which modified the elbow's Φ angle. The five variations for each question were validated by the same FCNSE interpreter so they meant the same. Figure 6 depicts two different signs represented using the 'very relaxed', the 'standard' and the 'very tense' mood variations.

Table 4: Sentences used in E_4.

Sentences
• I have waited for you here for three hours.
• I want you to fix the telescope today.
• I don't want the document, thank you. Would you mind deleting it?
• We will go to the cinema because it is raining.
• In short, the prize will be for your company. Good morning.

6.3. Procedure

The users received an e-mail containing the URL[6] of the form and asked them to perform the evaluation within a week. The first page of the questionnaire contained text that explained the experiment to the users and instructed them how to proceed. The questionnaire contained 25 randomly

[6]The questionnaire is available at `http://www.hctlab.com/evaluation/?idexp=1`, experiment contents can be downloaded from `http://www.hctlab.com/evaluation/expdata/exp1.rar`

ordered questions (5 sentences × 5 variations). The layout of each question featured two videos on the left side of the screen. The right side formulated the particular question and, below it, the different possible answers. Each question compared two videos depicting the same sentence; above, the standard variation was presented, and below, one of the five variations of the same sentence (including the standard) was presented. Users were asked to define the perceived stress level in the second video using a 5-point Likert scale. The forced answers were : "The second video, compared to the first one, seems: a) very relaxed, b) relaxed, c) the same, d) tense, e) very tense". Users were also asked to identify any pair of videos that were considered to represent different meanings. Each user performed the evaluation in less than 10 minutes. A screenshot of the questionnaire is depicted in Figure 5.

6.4. Results

The results of the first question of this experiment are presented in Table 5. The results of the second question reported that each of the 25 pairs of videos showed the same information (see Table 6).

Table 5: This table presents the obtained results for the first question of the SL synthesis experiment

	Selected answer				
Variation	Very Relaxed	Relaxed	The Same	Tense	Very Tense
Very Relaxed	**46,0%**	36,0%	18,0%		
Relaxed	12,0%	**62,0%**	26,0%		
Standard			**100,0%**		
Tense			34,0%	**52,0%**	14,0%
Very Tense				34,0%	**66,0%**

6.4.1. Discussion

The results of this experiment show that it is possible to generate synthetic signed messages representing different stress levels without modifying the meaning of the message. Table 6 shows that the message of the 5 sentences was kept regarding any modification. This result verifies the condition we imposed: "the message must be the same". This evaluation has been performed by LSE native users. Native users are demanding when it comes to synthetic signed contents and are less tolerant to mistakes than interpreters. The sign recognition rates of deaf user have been reported to be lower than

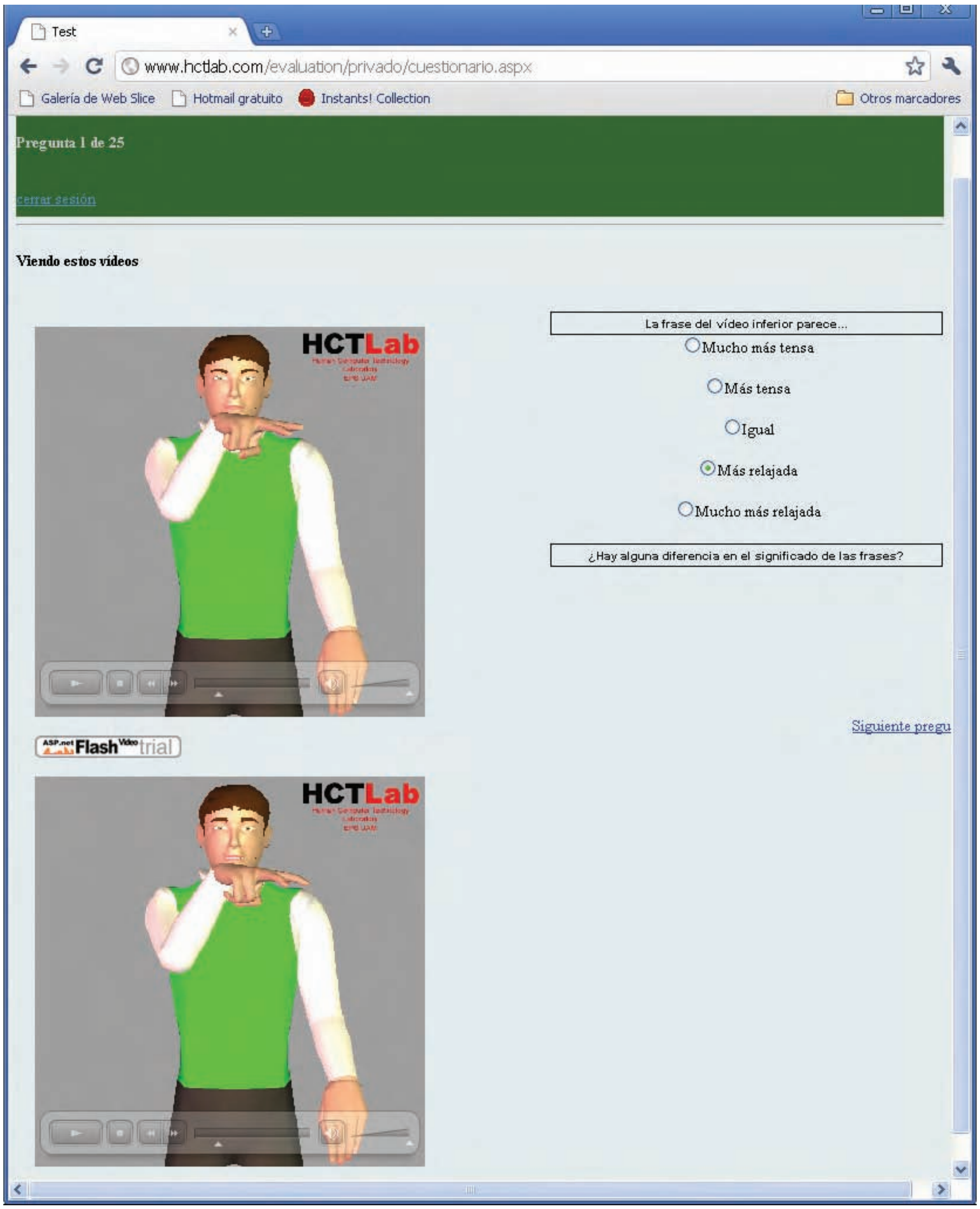


Figure 5: Screenshot of the questionnaire

Table 6: This table presents the obtained results for the second question of the SL synthesis experiment

Sentence	The meaning of the sentences is... the same	different
Sentence 1	10	0
Sentence 2	10	0
Sentence 3	10	0
Sentence 4	10	0
Sentence 5	10	0

the results obtained when the users who evaluated the system were SL interpreters (López-Colino and Colás, 2011). The acceptance of synthetic signed messages is also lower when evaluated by deaf users (Cox et al., 2003). Deaf users provide the baseline of the system.

When the first four modifications (M_1 - M_4) are introduced into the synthetic contents, 74% of the answers correctly identify a relaxed variation and 66% correctly identify a tense variation. None of the users selected the opposite variation. The 'very tense' and 'very relaxed' variations also include the modification M_5. The results show an increase of the perceived intensity. The results of the sentences which include this modification show that the 82% of the answers correctly identify the variation as 'relaxed' or 'very relaxed' and the 100% of the users consider the variation as being 'tense' or 'very tense'. The represented variation and the users' answers have a significant relationship ($\chi^2(16, N = 250) = 403.72, p < .001$).

These results were not expected considering the results of experiment E_2. In E_2, the standard-relaxed comparison was better identified than the standard-tense comparison, contrary to the results of E_3. We believe that the combination of modifications M_4 and M_5 have created this effect. A gesture involving more muscle effort (M_5) performed faster (M_4) is clearly perceived as 'tense'. However, this specific aspect of visual prosody has not been studied.

6.4.2. Experts report

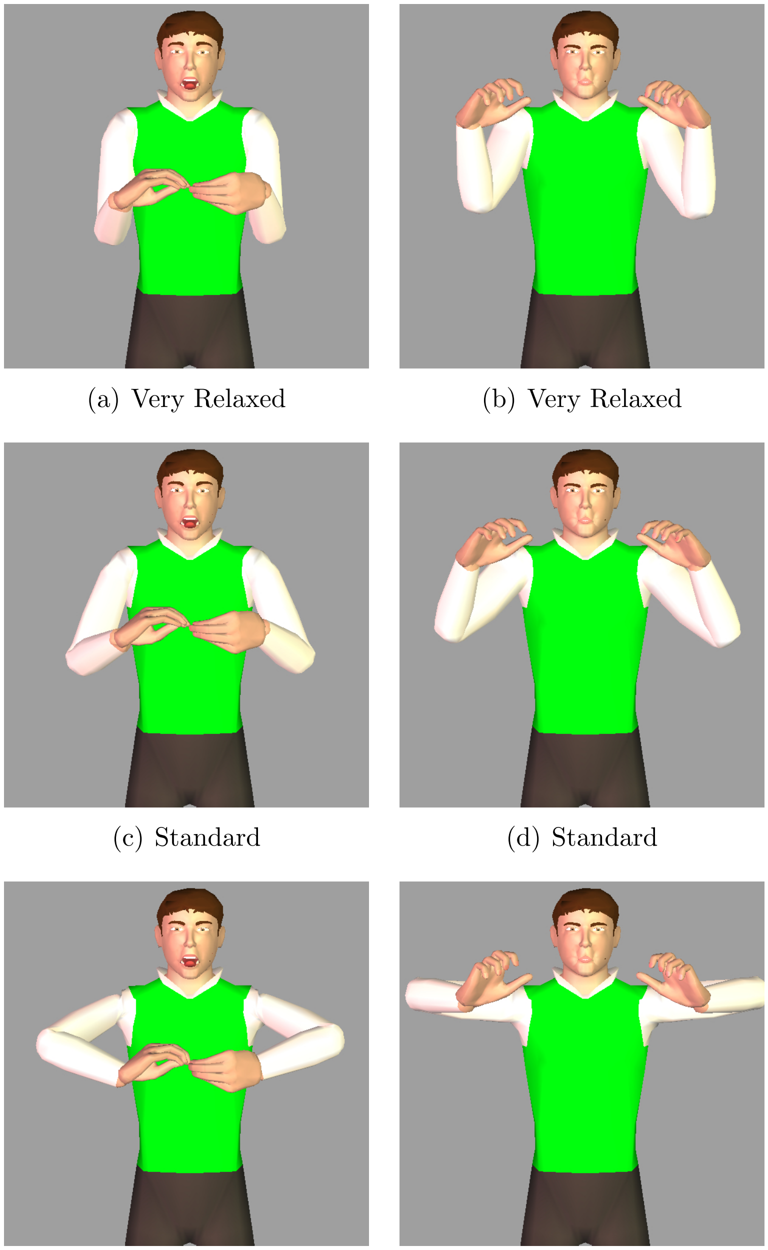
After the experiment, the group of experts from the FCNSE provided a report about the videos. This report confirmed the collected data declaring that the modifications did not modify the message. However, their main concern was the facial expression. They remarked the importance of the facial expression for the signed communication. Although the use of facial expressions was linguistically correct, they requested more intensity in the facial expressions and more animation to the face.

7. Conclusion

This work pertains to the synthesis of mood-modified signed messages. This work presents a study of the modifications and parameters related to visual prosody that can be included in a signing avatar. The objective of these modifications is to represent stress in the synthetic signed message. However, these modifications cannot modify the message. We proposed three modifications related to the alteration of three PPs (configuration, movement and non-hand parameter) to introduce mood-related modifications. We have also consider the signing speed and the shoulder complex tension as parameters to achieve the mood-related modifications. The modifications to the movement and non-hand PPs and the signing speed have been adapted from existing visual prosody research. The modifications to the configuration and the shoulder complex have been introduced in this paper. We present two evaluations to verify that these modifications alone are correctly perceived by users; these experiments can be considered to fall under the visual prosody research area.

The five modifications have been included in our previously developed LSE synthesizer (López-Colino and Colás, 2011). We conducted a ten-person evaluation (within-subject design). It was shown that the related modifications can be included in a signing avatar without modifying the message. The results show the validity of our proposal proving the first approach in synthesizing mood-affected signed messages; users correctly identified different stress levels in the synthetic messages.

Although the proposed modification to the configuration parameter (M_1) is correctly perceived in the individual experiment, the size of the hands in synthetic messages is not enough to perceive muscular tension differences. During the signed communication, hands are only observed during the $4,1\%$ of the time (Lu et al., 1998). These two aspects lead us to consider omitting this modification in future research. On the other hand, the proposed modification to the shoulder's complex tension by means of the angle Φ (M_5) is promising. Including this modification with the other four, made that every user consider that the modified sentence to be more tense or more relaxed.



(a) Very Relaxed (b) Very Relaxed

(c) Standard (d) Standard

(e) Very Tense (f) Very Tense

Figure 6: This figure presents two signs TO_FIX on the left and RAIN on the right) under the three different frames of mind that we have defined: Very Relaxed, Standard and Very Tense

The FCNSE linguistic users showed great interest in our research. The Spanish Deaf community is distrustful of signing avatars. However, the FCNSE experts considered our research an important first approach to improving the perception of signing avatars among the deaf community. They suggested us to focus on facial expressions. Considering existing research on visual prosody and the results provided by Lu et al. (1998) (the face is observed during the $77,1\%$ of the signed communication) improving the facial expression is a priority.

This paper presents possible (but not the only) modifications that, when included in the SL synthesis system, are correctly perceived as different degrees of stress without modifying the meaning of the sentence. However, these modifications are not based on how signers express stress in real signing. This limitation could be a problem when representing sensitive information (e.g., disturbing news) or complex syntax structures. To emulate real signers, it will be necessary the generation of a mood-altered signed corpus, the study of this corpus and the application of the results of the study to the SL synthesis. The results of different naturalness evaluations will help us to understand why the proposed levels were not similarly perceived by the users. It will require more evaluations to find the combination of the proposed modifications and new modifications to maximize the naturalness of the mood-altered synthetic signed messages.

7.1. Future work

This work studied the representation of tension in a synthetic signed message. Although it is an important step towards improving synthetic signed messages, it is only the beginning of a line of research. The next step in mood synthesis should focus on emotional synthesis itself by representing the six basic emotions in a synthetic message. This will require both psychological studies about how emotions modify the signed messages and evaluations of the synthetic signed messages that include these modifications. The existing studies on visual prosody and visual speech will serve as the basis in providing an experimental design.

8. Acknowledgements

The authors would like to acknowledge the FPU-UAM program for its support. We also want to acknowledge the collaboration of those who performed the evaluations, especially those from the FCNSE linguistic department for participating as evaluators and reporting any possible linguistic flaw in the signed messages.

The authors authors are grateful to the editor Dr. J. Riegelsberger for his valuable support and to the four anonymous reviewers for their comments, which have significantly contributed to the improvement of this paper.

9. References

Berthouze, N., Fushimi, T., Hasegawa, M., Kleinsmith, A., Takenaka, H., Berthouze, L., July 2003. Learning to recognize affective body postures. In: Computational Intelligence for Measurement Systems and Applications, 2003. CIMSA '03. 2003 IEEE International Symposium on. pp. 193–198.

Beskow, J., Oct. 2004. Trainable articulatory control models for visual speech synthesis. International Journal of Speech Technology 7 (4), 335–349.

Brave, S., Nass, C., Hutchinson, K., 2005. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. International Journal of Human-Computer Studies 62 (2), 161 – 178, subtle expressivity for characters and robots.
URL http://www.sciencedirect.com/science/article/B6WGR-4F5SB91-1/2/c3731ef2fb0432046a13222809708ed9

Breazeal, C., 2003. Emotion and sociable humanoid robots. International Journal of Human-Computer Studies 59 (1-2), 119 – 155, applications of Affective Computing in Human-Computer Interaction.
URL http://www.sciencedirect.com/science/article/B6WGR-48BKRSM-1/2/1b9513c7b211f468e433ad3b05a29879

Cao, Y., Tien, W. C., Faloutsos, P., Pighin, F., 2005. Expressive speech-driven facial animation. ACM Transactions on Graphics 24 (4), 1283–1302.

Caridakis, G., Raouzaiou, A., Bevacqua, E., Mancini, M., Karpouzis, K., Malatesta, L., Pelachaud, C., 2007. Virtual agent multimodal mimicry of humans. Language Resources and Evaluation 41, 367–388.

Cassell, J., 2000. Nudge nudge wink wink: elements of face-to-face conversation for embodied conversational agents. MIT Press, Cambridge, MA, USA, pp. 1–27.
URL http://portal.acm.org/citation.cfm?id=371552.371554

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M., 1994. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In: Proceedings of SIGGRAPH. ACM, New York, NY, USA, pp. 413–420.

Castellano, G., Mancini, M., 2009. Analysis of emotional gestures for the generation of expressive copying behaviour in an embodied agent. Lecture Notes in Computer Science 5085, 193–198.

Cole, R., Van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., Schwartz, S., Wade-Stein, D., Ward, W., Yan, J., Sep. 2003. Perceptive animated interfaces: first steps toward a new paradigm for human-computer interaction. Proceedings of the IEEE 91 (9), 1391–1405.

Corina, D. P., 1996. Sign linguistics phonetics, phonology and morphosyntax. Lingua 98 (1–3), 73 –102.

Coulson, M., jun 2004. Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. Journal of Nonverbal Behavior 28 (2), 117–139.

Cox, S., Lincoln, M., Nakisa, M., Wells, M., Tutt, M., Abbott, S., Oct. 2003. The development and evaluation of a speech-to-sign translation system to assist transactions. International Journal of Human-Computer Interaction 16 (2), 141–161.

de Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., De Carolis, B., 2003. From greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. Int. J. Hum.-Comput. Stud. 59 (1-2), 81–118.

De Silva, P., Bianchi-Berthouze, N., Jul. 2004. Modeling human affective postures: an information theoretic characterization of posture features. Computer Animation and Virtual Worlds 15 (3–4), 269–276.

Ekman, P., 1971. Universals and cultural differences in facial expressions of emotion. Nebraska Symposium on Motivation 19, 207–282.

Ekman, P., 1999. The Handbook of Cognition and Emotion. John Wiley & Sons, Ltd., Ch. Basic Emotions, pp. 45–60.

Elliot, R., Bueno, J., Kennaway, R., Glauert, J., May 2010. Towards the integration of synthetic sl animation with avatars into corpus annotation tools. In: Proceedings of LREC, 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. Valletta, Malta, pp. 61–64.

Elliot, R., Glauert, J., Jennings, V., Kennaway, R., May 2004. An overview of the sigml notation and sigml signing software system. In: Proceedings of Language Resources and Evaluation Conference. Lisbon, pp. 98–104.

Elliott, R., Glauert, J., Kennaway, R., Marshal, I., Sáfár, E., 2008. Linguistic modelling and language-processing technologies for avatar-based sign language presentation. UAIS 6 (4), 375–391.

Fotinea, S.-E., Efthimiou, E., Caridakis, G., Karpouzis, K., 2008. A knowledge-based sign synthesis architecture. Universal Access in the Information Society 6 (4), 405–418.

Fundación CNSE, 2008. Diccionario Normativo de la Lengua de Signos Española [Spanish Sign Language Normative Dictionary]. Fundación CNSE.

Gralewski, L., Campbell, N., Thomas, B., Dalton, C., Gibson, D., 2004. Statistical synthesis of facial expressions for the portrayal of emotion. In: GRAPHITE '04: Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia. ACM, New York, NY, USA, pp. 190–198.

Hanke, T., May 2004. Hamnosys - representing sign language data in language resources and language processing contexts. In: Heidelberg, S. B. . (Ed.), Proceedings of LREC. Lisbon, pp. 1–6.

Hartmann, B., Mancini, M., Buisine, S., Pelachaud, C., 2005a. Design and evaluation of expressive gesture synthesis for embodied conversational agents. In: AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems. ACM, New York, NY, USA, pp. 1095–1096.

Hartmann, B., Mancini, M., Pelachaud, C., 2005b. Implementing expressive gesture synthesis for embodied conversational agents. In: In Gesture Workshop. Springer, pp. 188–199.

Herrero Blanco, Ángel., May 2004. A practical writing of sign languages. In: Proceedings of LREC. Lisbon, pp. 37–42.

Huang, Z., Eliëns, A., Visser, C., 2002. Step: A scripting language for embodied agents. In: Proceedings of the Workshop on Lifelike Animated Agents. Tokyo, pp. 46–51.

Huenerfauth, M., 2008. Evaluation of a psycholinguistically motivated timing model for animations of american sign language. In: ACM SIGACCESS conference on Computers and Accessibility. ACM, New York, NY, USA, pp. 129–136.

Huenerfauth, M., 2009. A linguistically motivated model for speed and pausing in animations of american sign language. ACM Transactions on Accessible Computing 2 (2), 1–31.

Huenerfauth, M., Zhao, L., Gu, E., Allbeck, J., 2008. Evaluation of american sign language generation by native asl signers. ACM Transactions on Accessible Computing 1 (1), 1–27.

ISO/IEC 14772-1:1997, 1997. Information technology – Computer graphics and image processing – The Virtual Reality Modeling Language (VRML) – Part 1: Functional specification and UTF-8 encoding. International Organization for Standardization, Geneva, Switzerland.

ISO/IEC 14772-2:2004, 2004. Information technology – Computer graphics and image processing – The Virtual Reality Modeling Language (VRML) – Part 2: External authoring interface (EAI). International Organization for Standardization, Geneva, Switzerland.

ISO/IEC 19774:2005, 2005. Information technology – Computer graphics and image processing – Humanoid animation (H-Anim). International Organization for Standardization, Geneva, Switzerland.

Jennings, V., Elliot, R., Kennaway, R., Glauert, J., May 2010. Requirements for a signing avatar. In: Proceedings of LREC, 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. Valletta, Malta, pp. 133–136.

Karpouzis, K., Caridakis, G., Fotinea, S.-E., Efthimiou, E., 2007. Educational resources and implementation of a greek sign language synthesis architecture. Computers & Education 49 (1), 54 – 74. URL http://www.sciencedirect.com/science/article/B6VCJ-4H4V6WF-1/2/7485593c58f51483c6da1d049922c501

Kennaway, R., Glauert, J., Zwitserlood, I., 2007. Providing signed content on the internet by synthesized animation. ACM Transactions on Computer-Human Interaction 14 (15), 1–29.

Levangie, P. K., Norkin, C. C., 2001. Joint Structure and Function: A Comprehensive Analysis, 3rd Edition. F.A. Davis Company.

Liddell, S. K., Johnson, R. E., fall 1989. American sign language: The phonological base. Sign Language Studies 64, 195–278.

López-Colino, F., Colás, J., 2011. Hybrid paradigm for spanish sign language synthesis. Universal Access in the Information Society.

Lu, S., Igi, S., Matsuo, H., Nagashima, Y., 1998. Towards a dialogue system based on recognition and synthesis of japanese sign language. In: Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction. Springer-Verlag, London, UK, pp. 259–271.

McDonnell, R., Jörg, S., McHugh, J., Newell, F., O'Sullivan, C., 2008. Evaluating the emotional content of human motions on real and virtual characters. In: APGV '08: Proceedings of the 5th symposium on Applied perception in graphics and visualization. ACM, New York, NY, USA, pp. 67–74.

McDonnell, R., Jörg, S., McHugh, J., Newell, F. N., O'Sullivan, C., 2009. Investigating the role of body shape on the perception of emotion. ACM Trans. Appl. Percept. 6 (3), 1–11.

Mehrabian, A., Jun. 1981. Silent Messages: Implicit communication of Emotions and Attitudes, 2nd Edition. Wadsworth.

Muñoz Baell, I., 1999. Cómo se articula la lengua de signos española? Confederación Nacional de Sordos de España.

Neff, M., Fiume, E., 2002. Modeling tension and relaxation for computer animation. In: SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation. ACM, New York, NY, USA, pp. 81–88.

Neff, M., Fiume, E., 2006. Methods for exploring expressive stance. Graphical Models 68 (2), 133–157.

Noot, H., Ruttkay, Z., 2005. Variations in gesturing and speech by gestyle. International Journal of Human-Computer Studies 62 (2), 211 – 229, subtle expressivity for characters and robots.
URL http://www.sciencedirect.com/science/article/B6WGR-4FDN8N3-3/2/af87f2ce5dff1a212e4e90e04c5eb6ba

Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Herming, J., 1989. HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductory guide. Signum-Verlag.

Rocha, A., Pereira, G., May 2004. Supporting deaf sign languages in written form on the web. In: Proceedings of LREC. Lisbon, pp. 26–28.

San-Segundo, R., Barra, R., Córdoba, R., D'Haro, L. F., Fernández, F., Ferreiros, J., Lucas, J. M., Macías-Guarasa, J., Montero, J. M., Pardo, J. M., Nov. 2008. Speech to sign language translation system for spanish. Speech Communication 50 (11-12), 1009–1020.

Sandler, W., Lillo-Martin, D., 2006. Sign Language and Linguistic Universals. Cambridge University Press.

Segouat, J., 2009. A study of sign language coarticulation. In: ACM SIGACCESS conference on Computers and Accessibility. No. 93. ACM, New York, NY, USA, pp. 31–38.

Segouat, J., Braffort, A., 2009. Toward the study of sign language coarticulation: Methodology proposal. In: Proceedings of the International Conference on Advances in Computer-Human Interaction. IEEE Computer Society, Los Alamitos, CA, USA, pp. 369–374.

Shaarani, A. S., Romano, D. M., 2007. Perception of emotions from static postures. In: ACII '07: Proceedings of the 2nd international conference on

Affective Computing and Intelligent Interaction. Springer-Verlag, Berlin, Heidelberg, pp. 761–762.

Shaarani, A. S., Romano, D. M., 2008. The intensity of perceived emotions in 3d virtual humans. In: AAMAS '08: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp. 1261–1264.

Sloan, R., Cook, M., Robinson, B., July 2009. Considerations for believable emotional facial expression animation. In: Visualisation, 2009. VIZ '09. Second International Conference in. pp. 61–66.

Solina, F., Krapez, S., Jaklic, A., Komac, V., 2001. Design and Management of Multimedia Information Systems. Idea Group Publishing, Ch. 13: Multimedia Dictionary and Synthesis of Sign Language, pp. 268–281.

Sutton, V., 1974. Signwriting. http://www.signwriting.org/.

Szczepankowski, B., 1999. Wyrwnywanie szans osb nieslyszk acych. WSiP.

Szczepankowski, B., Rona, M., 1994. Szkolny slownik jezyka migowego. Oficyna Wydawnicza Vinea.

VCom3D, 2009. Signsmith studio. Online (www.vcom3d.com).
URL `http://www.vcom3d.com/signsmith.php`

Wahlstrom, J., Lindegard, A., AhlborgJr, G., Ekman, A., Hagberg, M., 2003. Perceived muscular tension, emotional stress, psychological demands and physical load during vdu work. International Archives of Occupational and Environmental Health 76, 584–590, 10.1007/s00420-003-0454-5.
URL `http://dx.doi.org/10.1007/s00420-003-0454-5`

Wilbur, R., 1999. Stress in asl: Empirical evidence and linguistic issues. Language and Speech 42, 229–250.

Wilbur, R., Schick, B., 1987. The effects of linguistic stress on asl signs. Language and Speech 30, 301–323.

Xiong, L., Zheng, N., Liu, J., Du, S., Liu, Y., 2010. Eye synthesis using the eye curve model. Image and Vision Computing 28 (3), 329–342.

10. Vitae

Fernando López-Colino received his Bachelor degree in Computer Science from the Universidad Autónoma de Madrid in 2005 and the Ph.D. degree in Computer Science and Telecommunication Engineering from the same university in 2009. In 2005 he joined the Human Computer Technology Laboratory (HCTLab) focusing his research on Sign Language processing and synthesis applied to mobile devices. He received in 2011 the *Infanta Cristina* award for this Ph.D. dissertation, which contains the results published here.

José Colás is professor within the Computer Architecture and Technology area since 2002. He received his Bachelor degree in Telecommunication Engineering from the Universidad Politécnica de Madrid in 1990 and the Ph.D. degree in Telecommunications from the same university in 1999. In 1993 his group received the *Reina Sofia* award for a research trajectory focused on technologies for In 2001 he founded the Human Computer Technology Laboratory (HCTLab) at the Universidad Autónoma de Madrid. This group received in 2003 the *Infanta Cristina* award for their research related to the new technologies for disability focused on mobile devices. He is the head of the "Multimodal Interaction oriented to Disable people" research at the HCTLab.