Journal of Visual Languages and Computing 23.3 (2012): 121-136

DOI: http://dx.doi.org/10.1016/j.jvlc.2012.01.003

Spanish Sign Language Synthesis System

Fernando López-Colino[a,*], José Colás[a,**]

[a] *Human Computer Technology Laboratory*
Universidad Autónoma de Madrid
Francisco Tomás y Valiente 11, E-28049 Madrid, Spain

Abstract

This work presents a new approach to the synthesis of Spanish Sign Language (LSE). Its main contributions are the use of a centralized relational database for storing sign descriptions, the proposal of a new input notation and a new avatar design, the skeleton structure of which improves the synthesis process.

The relational database facilitates a highly detailed phonologic description of the signs that, include parameter synchronization and timing. The centralized database approach has been introduced to allow the representation of each sign to be validated by the LSE national institution, FCNSE.

The input notation, designated HLSML, presents multiple levels of abstraction compared with current input notations. Redesigned input notation is used to simplify the description and the manual definition of LSE messages.

Synthetic messages obtained using our approach have been evaluated by deaf users; in this evaluation a maximum recognition rate of 98.5% was obtained for isolated signs and a recognition rate of 95% was achieved for signed sentences.

Keywords: Animated Conversational Agent, Spanish Sign Language, Sign Language Synthesis, Multimodal Interacting Communication

1. Introduction

The aim of this work is to provide signed contents for multimedia applications and to create a Spanish Sign Language (LSE) synthesizer that can be used by application designers and programmers who have no prior any knowledge of Sign Language (SL) synthesis or sign definition.

The existing literature provides several examples of SL translation and synthesis modules. However, the number of web sites and applications that integrate this type of accessible feature is very low. The Spanish government enacted a law [1] that, obligates official web pages to provide signed content to make them accessible to Deaf signing people[1]. Although several years have passed since this law went into effect, most official web pages still lack this type of accessibility. This is likely a result of the fact that the integration of signed content in a web site or an application is an expensive and complex process.

Existing approaches to the creation of SL synthesizers use the XML-based version of a preexisting graphic notation as the description of language parameters. Because these notations define SL at a phonetic level, so the usability of these notations is low because is equivalent to defining the input of a speech synthesizer using phoneme sequences instead of words. The phonetic approach has the further disadvantage that phonetic notations do not describe some relevant aspects of the signs, such as timing and synchronization of different phonologic parameters (see Section 2.3). Graphic notations used in SL synthesizers require an extensive knowledge of the described SL dialect because they describe the message defining all of the parameters of a sign. The approach proposed in this work uses a relational database to store the phonetic information of signs and improve the level of detail in the description of each sign by including the information that other notations missed. The contents of this relational database are defined using tools that have been specifically designed for Deaf people. Because the sign definitions are stored in the relational database, the input notation is gloss-based; this higher level of abstraction speeds the process of defining an input message. However, HLSML also permits the use of phonetic definitions.

The avatar is an important element within an SL synthesizer. Some current approaches use standard avatar definitions, H-Anim, to represent the signed message. This paper describes several modifications of the avatar's skeleton that provide features that are important to the sign synthesis process. These new features include an independent management of all of the sign parameters, easing the look-at and face-to definition, anatomic referencing and collision avoidance. All of these features are mesh-independent.

*Corresponding author. tel: +34 91 497 3613
**tel: +34 91 497 2257
Email addresses: `fj.lopez@uam.es` (Fernando López-Colino), `jose.colas@uam.es` (José Colás)

[1] Different studies show that Deaf people have a low reading comprehension level.

This paper is organized as follows: Section 2 reviews current SL synthesis systems, including those found in complete translation systems. Section 3 presents our approach to the description and synthesis of classifier predicates in LSE. The subsequent sections describe the main modules of the system focusing on the input notation (Section 4), the avatar's skeleton (Section 5) and the database structure (Section 6). After describing each relevant and novel component of the synthesizer, Section 7 presents the entire synthesizer architecture and explains the synthesis process. Section 8 contains the results of user tests and evaluations. Finally, Section 9 summarizes this work.

2. Background and related work

This section reviews current sign language synthesis systems including those found in complete translation systems. The ViSiCAST [2] and eSign projects [3] represent great advances in SL translation from voice or text. The DICTA-SIGN project [4] is an ongoing research project focusing on four European sign languages. All of these projects use HamNoSys notation [5, 6] as a precursor to gesture synthesis and use the same SL synthesis module, which is reviewed below. The goal of the ATLAS project [7] is to obtain machine translation from spoken or written Italian to Italian Sign Language. Because LSE translation is in its early developmental steps, a small number of related works describe, the first efforts on this subject [8, 9, 10, 11].

2.1. Classification of Sign Language synthesizers

Two main techniques have been developed to represent SL messages. The first approach to SL synthesis consists of creating a composition of small segments of video recordings [12]. Because this approach to sign synthesis requires image processing and a great number of pre-recorded sequences, significant storage capacity is necessary. Another drawback of this approach is that the addition of new units to the dictionary requires that the new units be carefully recorded so that they match the previous ones.

The second main approach to SL synthesis uses virtual avatars. The avatar is a 3D-generated human model that is usually animated using a bone structure (skeleton). Many projects [2, 3, 13, 14, 15] use this similar approach to gesture synthesis. The most widely used skeleton structure is H-Anim [16], which is a standard definition for human representation on VRML/X3D [17, 18, 19]. The DICTA-SIGN [4] skeleton structure is very similar to H-Anim, it is based on an earlier design developed in the eSign project [3]. Two different approaches to avatar animation have been used; these differ according to the definition of the animation. The first uses either continuous motion data obtained from an expert signer using motion capture techniques [20] or manual animations created by an expert animator. Although the results obtained with this technique are natural, Kennaway [21] describes several disadvantages associated with it; some of the disadvantages result from difficulties in the adaptation of the recorded data to avatars with different anatomies. An alternative approach to avatar-based SL synthesis uses phonetic definitions of the sign to generate the animation. To appear as natural as the animation obtained using the first approach, this method requires complex algorithms. However, it results in animation quality that is the same over the entire sentence. Furthermore, this method is highly flexible and storage requirements are greatly reduced. The same effect is observed in speech synthesis, where the parametric synthesis is less natural and presents lower quality than the speech synthesis based on the concatenation of prerecorded utterances.

2.2. Sign description approaches

Most SL synthesizers use standard symbolic notation as a basis for of signs. Notations such as HamNoSys [5, 6] and SignWriting [22] are graphic representations of SL and have equivalent computer-friendly versions, SiGML [23] for the HamNoSys notation and SWML [24] for the SignWriting version. In these projects [25, 26, 27, 28, 29], gesture synthesis is achieved by direct conversion from SWML or the SiGML notation into a VRML animation. Grieve-Smith [30] uses the Stokoe notation for defining signs. There is also another representational system called "Szczepankowski's gestographic notation" [31] that is used in Polish Sign Language. This is a textual notation; because it uses regular ASCII characters, it is computer-friendly. However, it does not represent all of the sign phonologic parameters (Section 2.3).

A different approach to SL definition involves the use of a relational database. Crasborn et al. [32, 33] created the database SignPhon as a tool for research into the phonological structure of SLs, in particular of Dutch Sign Language. As the designers state, SignPhon is not designed to allow the synthesis of Sign Languages, but it is an important approach to defining SL using a relational database. Furst et al. [34] presented a preliminary work on SL synthesis consisting of a relational database with two tables, one of which defines the configuration parameter (hand shape); the other stores the relationships between hand configurations required for the synthesis.

2.3. Study of the Spanish Sign Language

This section briefly describes the phonologic parameters that constitute a sign and are required to synthesize it. We will also describe various units that may be can be present in an SL sentence.

2.3.1. Phonetic Model

We have modeled our design on a phonetic model that defines seven phonologic parameters (PPs) [35, 36], each of which functions independently. Every PP is important for the definition of a sign, and changing only one of them can modify the meaning of the sign, as follows:

- The *configuration* PP is defined by the flexion of finger joints, which together describe a hand shape.

- The *orientation* PP is the spatial orientation of the hand; defined by the combination of the absolute directions of the extended finger and the palm.

- The *location* PP is the position of the hand within the frontal plane, defined using anatomic references. This parameter provides X and Y values of the 3D coordinate required for hand positioning.

- The *plane* PP describes the horizontal distance between the hand and the body. This parameter provides the Z value of the 3D coordinate required for hand positioning.

- The *contact* PP specifies the kind of contact between the hands and the body. Our project introduces the definition of this parameter into every sign to define the hand's active joint for spatial positioning. Figure 1 shows two signs that differ only in the *contact* PP parameter.

- The *movement* PP describes hand displacement based on its path and indications for speed and acceleration.

- The *non-manual* PP groups together facial expressions and body postures.

A similar phonetic model was proposed for LSE by Herrero-Blanco [37]. Herreo-Blanco's model merges the position of the hands ("location" and "plane") into a single parameter and defines two kinds of movements, internal and external. The phonetic model considers variations of the "configuration" and the "orientation" as internal movements; a variation of the position of the hands is defined as an external movement. Both types of movements are defined using phonemes of the "movement" parameter. Both models describe the locations of the hands using discrete anatomical references.

The first model has two main advantages. First, the independence of the parameters allows phonologic operations due to morphology (inflection, flexion, repetition, etc.) or syntax (the NMC is used to express questions, negations, etc.). Second, considering the "plane" and "location" parameters separately reduces the number of phonemes and enables morphological modification of the "plane". The latter parameter is related to the temporal aspect, which uses different planes to refer to past, present or future actions. The synthesizer can easily merge the location and plane, thus reducing the number of units to be stored.

2.3.2. Elements in the signed message

An LSE message includes various units. Based on a phonetic criterion, we wish to highlight the following elements:

- The first element is *fingerspelling*, a representation of the alphabet with signs. Each letter is represented using a defined hand shape and orientation. Most letters are static one-handed signs; in some cases they require simple animation.

- *Dictionary signs*, or base forms, represent concepts. Such signs have well-known and static meanings. Describing dictionary signs involves defining all of the parameters, sometimes for both hands in the case of a double-handed signs. The phonetic description of these elements is more complex than that of fingerspelling and, requires more complex management during the synthesis.

- *Classifier Constructions* (CCs) are the last unit. Different studies [38, 39, 40, 41] proposed several approaches to the definition of this unit in different SLs. Herrero Blanco [42] proposed a classification for this unit applied to LSE; we have used the same classification in this work.

3. Approach to Spanish classifier constructions

The previous section presented a classification of signing units (fingerspelling, dictionary signs and CCs) that takes into account both semantics and phonetic complexity. Fingerspelling and dictionary signs have static descriptions, and their synthesis is not modified by semantics. However, with respect to both translation and synthesis processes, CCs require a different approach. Our work focuses on CC synthesis and the description of this semantic unit within the input notation. Because CCs are relevant elements in SL communication, including them in the synthetic sign message will result in an improvement in the naturalness of the message.

Herrero Blanco [42] presented a classification of CCs in LSE. This classification suggests that the following different kinds of CCs exist in LSE:

- *Classifier nouns* are signs that are used as modifiers of the next sign. These signs can be used independently with their own meaning, but they can also be used preceding another sign to define a new concept. Because it can be described as two independent signs, this kind of CC does not interfere with the standard synthesis process.

- *Inflective classifiers* are modifications to one PP of a classified sign. A new phoneme is used instead of the original one. The new phoneme may correspond to the definition of the classifier sign or to the classifier configuration of the classifiers's sign category; in the latter case, the meaning of the classifier sign is added to that of the original classified sign. Figure 2 depicts the sign TO_GIVE and how the configuration is modified in "to give a book" using the book's classifier configuration.

(a) Sign EYE (b) Sign TO_SMELL (c) Sign RED

Figure 1: This figure presents three different signs that can be generated if the contact point is modified.

(a) Sign TO_GIVE (b) Construction "to give a book"

Figure 2: Example of an inflective construction in LSE.

This feature applies only to pairs of classifier and classified signs whose meaning are related to spatial and tangible properties. This CC unit implies modifications to the synthesis of the classified sign. It can be only established during translation, when the classifier sign is defined. Section 4 defines how this type of CC is described in the proposed notation.

- *Iconic classifiers* or *classifier predicates* (CPs) depict the movement, the shape or the location of the objects in space. These productive units depict reality; therefore, their definition must be obtained from the implicit or explicit description found in the translated messages.

3.1. Describing and synthesizing classifier predicates

Huenerfauth [43, 44] proposed the first approach to CP automatic synthesis. This approach used a software system, AnimNL, to generate a 3D scene described by means of natural language (English). However, because there is not equivalent software for Spanish, we propose describing LSE CPs using a sequence of basic units. These units define the Non Manual Component NMC and the configuration, orientation, contact point PPs and the spatial position of each hand. They can also set the avatar's gaze or modify the head orientation to face towards a given point. The latter point and the position of the hands are defined using a spatial coordinate within the articulatory space (Figure 3). Dynamic CCs require hand movement or gazing at different points. A sequence of static units with time gaps between them will define the CCs' animation.

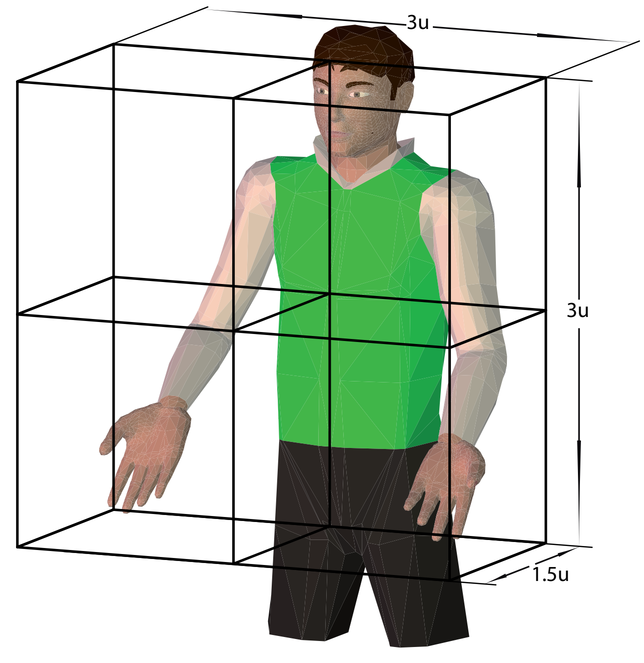


Figure 3: The black wired box shows the coordinate system used to position an element within the classifier articulatory space. The value of u is the length of the avatar's upper arm.

The next section includes an example showing the way this synthesizer describes these structures using HLSML. We refer the reader to [45] for further study and evaluation of the description and synthesis of LSE's classifier constructions.

4. HLSML: New synthesizer input notation

We mentioned in Section 2 that synthesizers that use the phonetic notation-based approach use the SiGML or SWML notations. These notations focus on sign description. Our approach to sign synthesis stores this phonetic information in the relational database (Section 6), and the input notation focuses on message description. This new definition describes a High Level Signing Markup Language (HLSML) that provides multiple levels of representation and fosters collaboration among translation experts without SL phonologic knowledge.

4.1. HLSML definition

The HLSML notation was created with the following five objectives: i) to generate an XML-based notation that can be used by people with minimal SL knowledge; ii) to describe the message independently of the different SL dialects as long as these dialects share a common grammar structure; iii) to define the different units (fingerspelling, dictionary signs, CCs, etc.) in the same specification; iv) to permit the use of prosodic modifiers to change the representation of a sign. v) to define the parallel behavior of non-hand features in the signing process.

The first and second objectives require that the contents in HLSML be completely independent of any SL dialect. SiGML, SWML and the Szc-ze-pan-kow-ski notations define the phonemes of a sign at different levels of detail, but all of them require some knowledge of SL phonology and of specific notation terminology. HLSML aims to reduce these requirements to a minimum by using glosses to represent signed messages. As shown in Section 6, these glosses are the content of the first level of the database, and they are the only compulsory field to recover all the synthesis information. HLSML also allows a sign to be defined by description of its phonemes. This level of abstraction is equivalent to that used in SiGML or SWML, but this feature is only used in HLSML in the few cases for which a sign has not been stored in the database, for example, for dynamic signs used to refer to people or to new concepts, and these signs are only valid during the conversation.

The specification of HLSML provides two approaches for describing a signed message: i) a high-level approach that uses glosses or states the word to be spelled (used for dictionary signs and fingerspelling); ii) a low level-approach that allows phonetic descriptions (used for CCs or signs). The user may choose between these levels depending on the contents of the database. Figure 4 presents a simple sentence in LSE. HLSML defines the sign glosses

```
<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
<sentence value='sentence1' language='es'
                                        tag ='standard'>
 <globalMod speed=0.9 anim='SLERP' timeInter=250/>
 <sign value='ONE'/>
 <sign value='CAR'/>
  <spellSign value='coche'/>
 </sign>
 <sign value = 'RED'>
  <spellSign value='rojo'/>
 </sign>
 <spell value='corsa'/>
</sentence>
</hlsml>
```

Figure 4: Example of HLSML code. This fragment defines a sentence with four elements, three signs and a spell sequence (ONE - CAR - RED - 'corsa'). Required information to synthesize the signs is in the database, thus HLSML only adds modifications in order to modify its representation. The <spellsign> element is recommended inside a <sign> in order to provide an alternative when a *dictionary sign* cannot be recovered from the database.

using Spanish words, but we have used English words in these examples to facilitate their understanding.

Section 3 describes the approach to Spanish CCs, the synthesis of which requires specific approaches to solve two different elements, *inflective classifiers* and *iconic classifiers*. *Introflexive classifiers* require defining a new phoneme; when representing the modified sign, this definition can be obtained from the definition of another sign. HLSML includes the element <signclassifier> to define the alternative parameter of the classified sign and the classifier sign. Figure 5 shows how this item can be used to specify a modification to the sign TO_GIVE (classified sign), which must be signed using the classifier configuration of the sign BOOK (classifier sign).

```
<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
<sentence value='sentence2' language='lse'
                                        tag ='standard'>
 <sign name="TO_GIVE">
  <signclassifier value ="configuration">
   <sign name="clBOOK"/>
  </signclassifier>
 </sign>
</sentence>
</hlsml>
```

Figure 5: Example of an inflective classifier which describes the Figure 2(b)

The classifier predicates are iconic descriptions of specific information. Because their definition is based on semantic information obtained from natural language processing, there is no possibility of storing their representation in the database. Therefore, this kind of CCs must be dynamically generated. Figure 6 is an example of a classifier predicate representing a person walking along a linear path.

In the same way that voice is modified by speaker mood, SL performance can be altered to reflect this parameter. The global speed of a sign, transition between signs, smoothness in transitions or internal movements and sign parameter variations are used for modifying syntax or prosody. Modification of these elements is also used to represent different frames of mind without changing the sign's meaning. SiGML DTD [23] presents the entity 'manner_attribs', which defines six possible variations for movements. These movement variations are part of SL phonology; therefore, they can modify the meaning of a sign, which is not the objective of frame-of-mind emulation.

The SiGML notation has recently been modified to include the two tags: "duration" and "timescale" [46]. Initially, the eSign system obtained the duration of a sign analytically from the HamNoSys definition. The current version of SiGML, like HLSML, allows the user to define a specific duration and speed. The HLSML also allows definition of the duration of the internal segments of the sign. This information is also represented in the database. HLSML includes other tags that modify other aspects of the synthetic signed message, such as timing (e.g., 'timeInter', 'timePrev' and 'timeNext' allow defining the duration of the transition between signs, modifying prosody), animation (e.g., 'anim' defines the interpolation approach that should be used during animation) and frame of mind (e.g., 'variation' defines the allophone for the PP sensitive to variations of the signer's frame of mind). The way in which these parameters alter the standard representation of a sign is not the scope of this paper; they are mentioned here only in order to show the differences between the HLSML definition and other notations. Figure 4 presents some of these attributes, including 'tag', 'variation', 'speed', 'anim', 'timeNext' and 'timePrev'.

Wilbur [47] and Baker and Padden [48] reported that NMC utterances can occur simultaneously with the manual utterances of dictionary signs. For example, in a negative sentence, the head may repeat a movement from side to side, denoting negation, while at the same time the hands are signing the relevant signs whose meaning is negated, such as the verb. Another aspect of the NMC is defining lip movement to simulate speech during the representation of signs.

4.2. HLSML compared to SiGML and SWML

In the previous section, we described HLSML features. This section aims to present a direct comparison between the most extended XML-based signing notations, SiGML and SWML, and HLSML. Fingerspelling can be considered as a special group of signs within the wider group of dictionary signs; therefore, SiGML and SWML will allow the definition of these fingerspelling signs. However, it is easier to state the directive "spell" than to define the whole sequence of single-handed static signs, which are the fingerspelling signs. Because neither SiGML nor SWML

```
1  <!DOCTYPE hlsml SYSTEM "hlsml.dtd">
2  <hlsml>
3  <sentence value='sentence3' language='lse'
4                                         tag ='standard'>
5   <classifierSequence>
6    <classifier time=100 armDivision=3>
7     <hand side="right">
8      <coordinates height=-3 width=0 depth=1/>
9      <configuration name="letter_v"/>
10     <orientation name="vertical_down"/>
11     <contactPoint name="end_point"/>
12    </hand>
13    <eyesLookAt height=-3 width=0 depth=1/>
14   </classifier>
15   <classifierTransition time=500 />
16   <classifier time=100 armDivision=3>
17    <hand side="right">
18     <coordinates height=-1 width=0 depth=5/>
19     <configuration name="letter_v"/>
20     <orientation name="vertical_down"/>
21     <contactPoint name="end_point"/>
22    </hand>
23    <eyesLookAt height=-1 width=0 depth=5/>
24   </classifier>
25  </classifierSequence>
26 </sentence>
27 </hlsml>
```

(a) HLSML code

(b) lines 6–12

(c) line 16–24

Figure 6: Example of a dynamic CCs description using HLSML code. This fragment represents a human-being moving away from the signer. The right hand stays 100 ms at the position $u(0, -1, 2/3)$, then it moves to $u(0, -1/3, 5/3)$ in 500 ms where it stays during 100 ms (u is the length of the upper arm). During these 700 ms the eyes look at the right hand, following its movement.

define CCs, the comparison with HLSML will focus on dictionary signs.

The phonetic description of TO GROW UP has been represented using HLSML in Figure 7 and using SiGML and SWML in Figure 8.

```
<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
<sentence value='sentence4' language='lse'
                                       tag ='standard'>
 <!-- gloss -->
 <sign value="GROW_UP"/>
 <!-- phonetic description -->
 <signDefinition>
  <holdMoveDefinition time="1000"/>
   <configuration>
    <phoneme value="extended" side="right"
              fraction_ini="0" fraction_end="100"/>
   </configuration>
   <orientation>
    <phoneme value="h_i_d" side="right"
              fraction_ini="0" fraction_end="100"/>
   </orientation>
   <location>
    <phoneme value="navel" side="right"
              fraction_ini="0" fraction_end="30"/>
   </location>
   <plane>
    <phoneme value="near" side="right"
              fraction_ini="0" fraction_end="30"/>
   </plane>
   <contact>
    <phoneme value="med_end" side="right"
              fraction_ini="0" fraction_end="30"/>
   </contact>
   <movement>
    <phoneme value="linear_up_med" side="right"
              fraction_ini="40" fraction_end="100"/>
   </movement>
  </holdMoveDefinition>
 </signDefinition>
</sentence>
```

Figure 7: HLSML example

Because HLSML is designed for people with minimal knowledge of sign phonology, it protects internal details from potential inadvertent changes. The examples presented in this section show that HLSML allows different levels of abstraction for message definition.

5. Avatar's skeleton design

Skeleton animation is a common means of avatar animation. The basic skeleton structure consists of a hierarchical definition of "animation bones", the orientations of which define the avatar's posture and modify mesh appearance. The skeleton structure presented in this work simplifies several tasks in SL synthesis by extending its definition with new auxiliary bones (Table 1).

5.1. Auxiliary bones in skeleton design

Most sections of the skeleton are animated using a direct animation approach. The orientations of the finger bones are defined by the configuration. The bones of the body (i.e., waist, thorax, shoulders, neck, etc.) are oriented depending on the NMC. The orientation PP defines the wrist's absolute orientation. The only two bones whose orientation is not prerecorded are the arm bones. The rotations of the shoulder and of the elbow must be calculated so that contact point of the hand achieves a target whose coordinates are defined by means of its location and plane parameters. This calculation must take into account the hand's configuration and orientation.

```
<!DOCTYPE sigml SYSTEM "sigml.dtd">
<sigml>
 <hamgestural_sign gloss="GROW_UP">
  <sign_manual both_hands="false">
  <handconfig handshape="flat"/>
  <handconfig extfidir="ol"/>
  <handconfig palmor="d"/>
  <location_bodyarm
    location="stomach">
  <directedmotion direction="u" size="small"/>
 </hamgestural_sign>
</sigml>
```

(a) SiGML example

```
<swml version="1.0">
 <!-- GROW_UP -->
 <signbox>
  <symb x="38" y="68" x-flop="0" y-flop="0">
   <category>01</category>
   <group>05</group>
   <symbnum>007</symbnum>
   <variation>01</variation>
   <fill>06</fill>
   <rotation>02</rotation>
  </symb>
  <symb x="43" y="40" x-flop="0" y-flop="0">
   <category>02</category>
   <group>03</group>
   <symbnum>001</symbnum>
   <variation>02</variation>
   <fill>01</fill>
   <rotation>01</rotation>
  </symb>
  <symb x="16" y="19" x-flop="0" y-flop="0">
   <category>10</category>
   <group>01</group>
   <symbnum>001</symbnum>
   <variation>01</variation>
   <fill>01</fill>
   <rotation>03</rotation>
  </symb>
 </signbox>
</swml>
```

(b) SWML example

Figure 8: Comparison of SiGML and SWML

A possible solution for resolving this problem was proposed in [49]. However, this method does not observe the independence of the orientation, location and plane PPs. Modifying the location and plane will alter the rotation of the shoulder and the elbow, thus modifying the absolute orientation of the wrist. Hence, the orientation PP would be modified by the location and plane PPs.

In order to provide the necessary independence of the hand orientation, we have inserted a new bone between the hand and forearm that we call the "wrist_position_bone". This bone overrides the inherited orientation from the forearm, maintaining a constant orientation. Therefore, the next bone's orientation will not be modified by shoulder and elbow rotations. The next bone, that we call "wrist_orientation_bone", represents the phoneme of the orientation using the *quaternion* obtained from the database (see Section 6). The position of these bones within the hand definition can be observed in Figure 9. This structure also simplifies the inverse kinematics process. The extended three-bone arm-wrist structure [49, 50, 51, 52] requires automatically handling seven degrees of freedom (DoF) (or relying its definition to a human animator). Our four-bone arm-wrist structure only defines four DoF because hand orientation is defined by the orientation phoneme.

McDonald et al. [49] use the concept 'articulator' to refer to a point in the hand used for targeting. They define the target as the position at which the articulator must be placed. Targets can be positioned in contact with the avatar. The synthesizer must dynamically obtain the position of the required target every time the definition of the sign refers to it, even though this target may have been displaced by an animation. The extended approach depends on mesh vertices to define these references [46]. This implies that mesh deformation must be computed during the synthesis, consuming extra processing time. In the skeleton definition, we introduced several "location bones" whose positions are used to define the required anatomic references. This approach avoids using the mesh definition, thus reducing the synthesis duration while obtaining the same results. The location bones are also used for hand positioning (e.g., the position of the end of the index finger is the position of a "location bone" located at the end of that finger).

Some signs and CCs stipulate that the head or eyes must be facing or looking at a defined imaginary point. This could be achieved by manually defining the orientation of the relevant bones. However, we have defined special "directional bones" for the eyes and the head that are always oriented towards two different invisible objects, one for the head (face_to_object) and another one for the eyes (look_at_object). This minimal modification simplifies the definition of the face-to/gaze-at properties, presented in Section 3.1 and has been found to be very useful.

The last improvement concerning the definition of the skeleton is related not to SL features but to avatar collision detection. Only Kennaway et al. [25] defines a simple collision detection process in the inverse kinematic calculation in order to avoid, for example, the elbow penetrating into the body; however, this detection depends on mesh definition. As mentioned above, the use of auxiliary bones can spare mesh processing. The proposed skeleton includes special bones called "collision bones" that define collision-free volumes (polyhedrons) around the body and head. Using these polyhedrons, the synthesizer will detect if an "animation bone" or a "location bone" is inside of one of these volumes. The system verifies impossible sign definitions and avoids transitions between two correct positions that collide with the head or the body (e.g., the transition from touching one ear to touching the other should avoid going through the head).

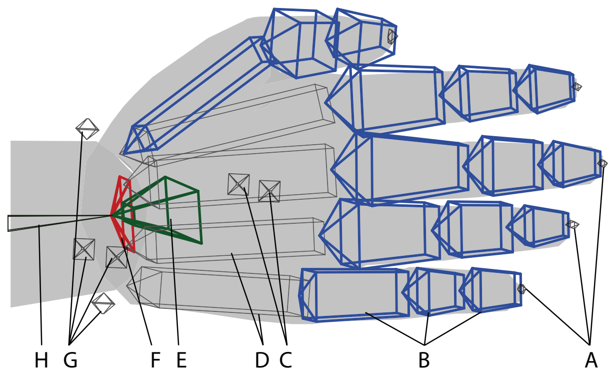


Figure 9: Bone structure of the hand, comprising the following: location bones at the end of the fingers (some of which labeled A); animation bones (labeled B); location bones at the palm and back part of the hand (labeled C); structural bones used to simplify the skeleton design (labeled D); wrist_orientation_bone (labeled E); wrist_position_bone (labeled F); wrist "location bone" (labeled G); forearm bone (labeled H).

Table 1: Table showing all the bones used in this skeleton design. The bones have been classified attending to their positions and functions. It should be noted that some bones are used for different functions; as such, they are included in all relevant fields.

	Position				
Function	Head	Hands	Arms	Body	Total
Animation	23	32	4	5	*64*
Location	36	25	24	13	*98*
Collision	14	0	0	18	*32*
Wrist special	0	2	0	0	*2*
Directional	3	0	0	0	*3*
Structural	3	8	0	5	*16*
Total	*79*	*67*	*28*	*41*	
# skeleton bones					**210**

5.2. *Facial bone animation*

Our design uses bone animation to generate all required facial expressions. We chose this approach instead of the more extended morphing approach [53, 46] for the following reasons: i) using the bone animation approach for both face and body animation simplifies the development of the application and storage of facial expressions (see Section 6); ii) with this approach, the modification and definition of new facial expressions do not require the

(a) Neutral (b) Surprise (c) Anger (d) Relaxation

Figure 10: Example of different face expressions obtained using skeleton animation.

release of a new avatar definition file; iii) this approach requires less storage capacity than the morphing approach, which requires storing every morphing objective. Figure 10 presents several examples of facial expressions obtained with this method.

6. Relational database

As stated in Section 2, recent studies define LSE as composed of seven different phonologic parameters that can be treated independently. We have described how this independence can simplify avatar animation using a specially designed skeleton. This independence has also been considered in the creation of the relational database.

The proposed synthesizer design is based on the phonologic and morphologic theories of SL. Following this approach, the relational database is structured in four logical levels, which are depicted in Figure 11. The first level works as a dictionary entry; this section of the database contains sign entries, some of which may be related to a concrete SL dialect because the database has been designed to represent all the dialectal variations of LSE. This level also defines the default duration of a sign.

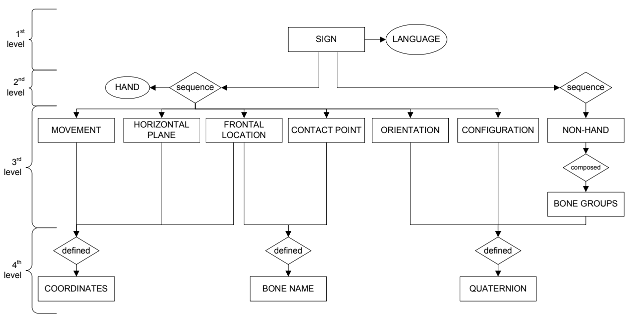


Figure 11: Simplified conceptual diagram of the four levels of database structure. The second level defines two different sequence relations because the NMC PP is the only one that does not require to state a hand.

The second level contains a description of each sign using sequences of the seven PPs; this level comprises seven tables. It should be pointed out that, for each of the six parameters related to the hands, up to two sequences can be found, one for each hand. This level also defines synchronization, if required, between different PPs and between the hands. Because all of these sequences are described using percentages of the sign duration, time scaling is very simple.

The third level contains the single units for each parameter. All hand -related parameters have been defined using one table for this level; however, the NMC parameter is different. This parameter allows the definition of a "complex expression" (e.g., a surprised expression) by means of "simple expressions" (e.g., position of each eyebrow, shape of the mouth or eyelid openness). Thus, at this third level, the structure of the NMC requires two related tables to define the composition of "simple expressions".

The last level of the database contains the information required to synthesize each basic parameter unit defined in the previous level. The definition of the PPs is accomplished through three kinds of elements listed below; each PP uses only one of these elements.

- The *Quaternions* [54] are a common way of defining an orientation in 3D environments. Shoemake [55] presents many advantages of this representation for animating rotations. Configuration, orientation and NMC (mainly facial expressions) generate several relations between a defined joint and a quaternion (e.g., a complete hand shape requires fifteen quaternions, one for each finger joint).

- A *Bone Name* is used in location and contact point PPs. The location is defined in SL using an anatomical point as reference, which is represented by a "location bones" (see Section 5). The location is defined using the name of the bone that depicts the relevant anatomical point. The contact point defines the finger joint that must be used for hand positioning. This joint is defined by means of the name of the "animation bone" related to that joint.

- A *Coordinate* is a simple 3D vector used for plane and movement PPs. When a coordinate is used for the plane, it defines the horizontal distance to the body. The movement PP requires a sequence of vectors that define the path of the movements; when these coordinates are used for a movement definition, they are assimilated to spatial vectors. The acceleration of the movement is modified by defining the duration of the displacement described by each vector. In order to provide independence of the avatar's anatomy, all coordinate values are measured relative to the length of the humerus bone, and the plane uses the position of the chest_location_bone as the origin of coordinates.

The database structure can be better understood using the following voice-based simile. If this database were used for speech synthesis, the first level would store every word or concept, the second would establish the sequence of basic units (phonemes) that forms the word, the third would provide definition of the basic units of the language, and the last would specify how each unit is produced.

In the previous simile, the relation between levels three and four is "1-to-N". It is well known that voice-based phoneme generation varies greatly according to different speakers or mood variations and that SL performance can be altered in the same way. Different representations have been stored in our database for the alterable parameters: 'configuration' (see Figure 12) and 'movement'. Those different representations, as well as sign variations, act as different sign performances that are influenced by the signer's frame of mind. Obviously, these modifications should not decrease the sign recognition rate. Because mood-modified speech synthesizers improve the naturalness of the synthesized speech message, the same result is expected in the signed message.

6.1. Filling the database using specifically designed visual tools

Database filling is a sensitive process because the animation of the avatar and the system's performance depend on it. The complex structure of the relational database necessitated two different stages in the filling process. The first stage defined the contents of the third and fourth levels and the second stage completed the first and second levels.

Because the third level of the database is equivalent to a list of all possible units for each parameter, filling it is not a difficult task. The list provided by Rodríguez González [35] and Muñoz Baell [36] was used, with some additions provided by an LSE expert. This expert also provided a list of the most relevant facial expressions in LSE. Finally, the Spanish visemes [56] were included in order to generate speech-like lip animations. The Phoneme Definition application, shown in Figure 13(a), simplifies the process for quaternion- and coordinate-based parameters. Meanwhile, the configuration and orientation PPs were defined directly; the two-level structure of the database used for the NMC parameter required two steps to define this parameter. This part of the tool is presented in Figure 13(b).

The second stage, which involved the definition of the LSE dictionary, was carried out after inserting every unit defined for each parameter in LSE. Unlike the first stage, this stage required the participation of an LSE expert. The Sign Recording application (depicted in Figure 14) designed for this stage was used by deaf people from the CNSE (National Confederation of Spanish Deaf People) to define several signs. Every application designed to be used by Deaf people (and, especially, pre-lingual deaf people) should minimize text interfaces. The resulting application, the Sign Recording application, uses only images and a drag-and-drop interface to define the signs. The users emphasized the simplicity and flexibility of the sign definition process; it is more precise than a HamNoSys definition, while avoiding its strict grammar.

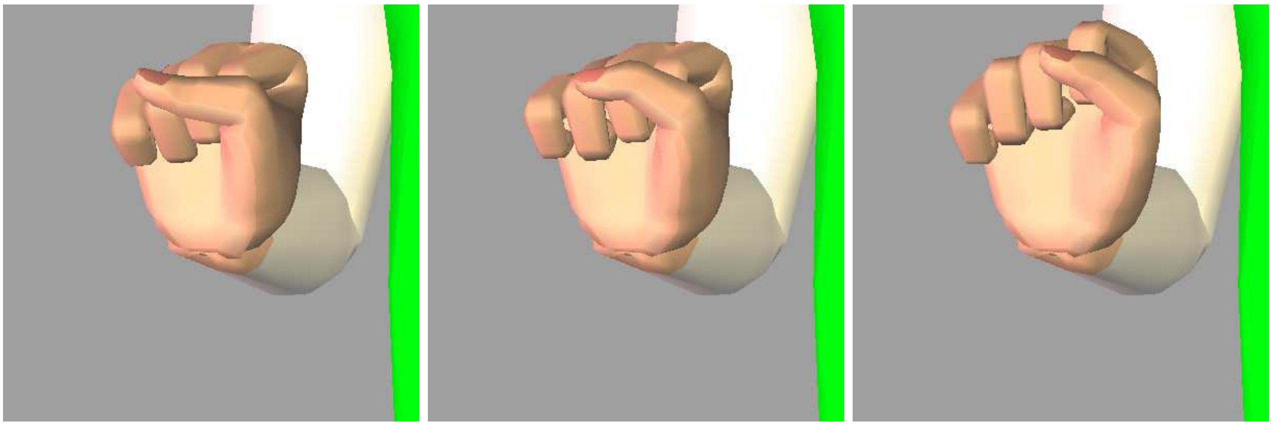
The database currently contains more than four hundred signs comprising fingerspelled letters, numerals, ordinals, pronouns, substantives, adjectives, verbs, adverbs and auxiliary elements (i.e., the neutral position). The contents of the database were used during the user evaluations.

6.2. Non-Manual component

We described in Section 5 how the NMC is completely animated using the skeleton. The NMC defines modifications to both facial expression and body posture. We also demostrated that the system manages different parts of the body as if they are independent (i.e., the position of the eyebrows does not depend on the mouth shape or the gaze direction). The database stores different values for each group. The system can combine these values to generate a great number of facial expressions and body postures. The current contents of the database related to the NMC are presented in Table 2.

Table 2: Database contents related to the NMC.

NMC independent units			
element	# units	element	# untis
head	16	eyes	5
neck	1	nose	3
torax	7	left cheek	4
waist	5	right cheek	4
left eyebrow	10	tongue	5
right eyebrow	10	mouth	30
left eyelid	4	l upper cheek	2
right eyelid	4	r upper cheek	2
# body-related possible postures			$5.6\ 10^2$
# face-related possible expressions			$2.3\ 10^8$



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

Figure 12: Example of different possibilities for the same hand configuration.

(a) Configuration PP recording panel (b) NMC definition

Figure 13: Phoneme Definition application. This application was used for stage one of the database filling process.

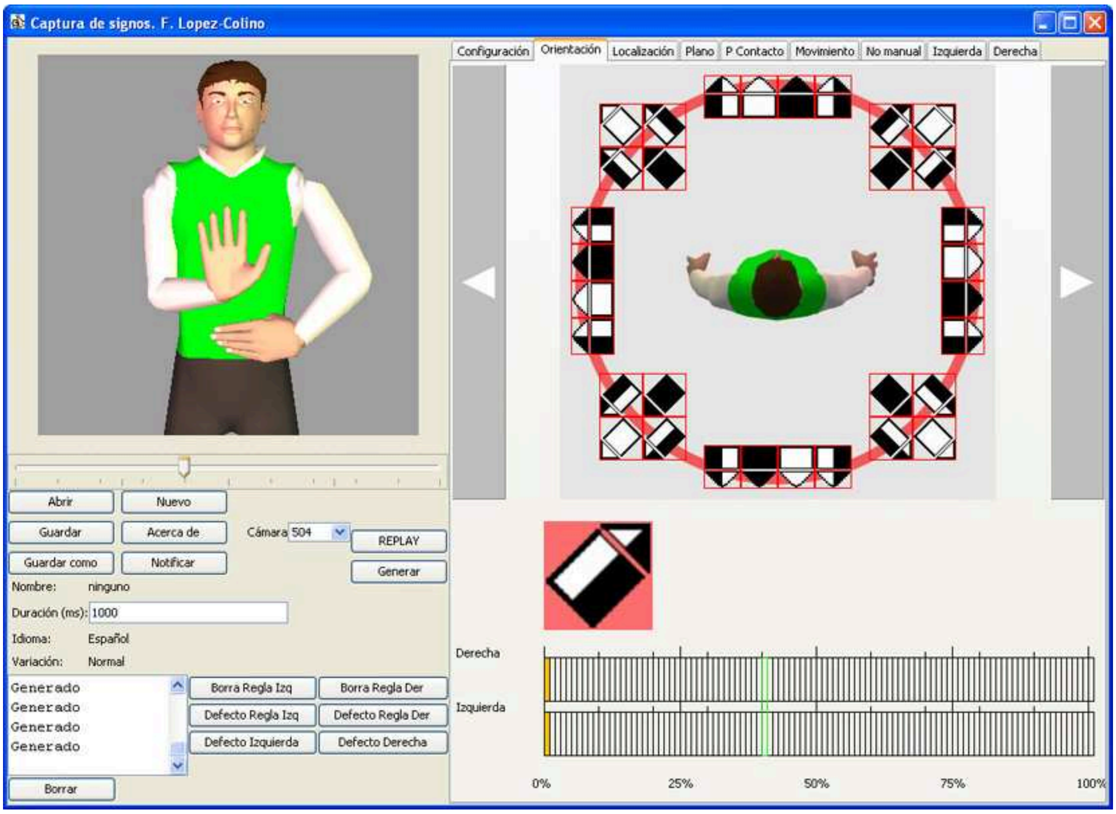


Figure 14: Sign Recording application. This application was used for stage two of the database filling process. It allows the user to specify the different sign parameters independently. The image shows the definition of the Orientation parameter.

7. System architecture and the synthesis process

The most relevant and novel elements of our sign synthesizer have been presented in the previous sections. This section will show how these elements work together to synthesize SL (Figure 15).

When the synthesizer is started and before the generation of any sign message, the Gesture Synthesis module obtains the description of the avatar from the Web Server. This connection is made only once during the execution of the synthesizer program.

The signing process begins when a sign message, defined using HLSML, reaches the synthesizer. This message is parsed in the HLSML parsing module to obtain the sequence of signs, which are defined by means of glosses, fingerspelling sequences and CCs, and the attributes that will modify their representation. The Gesture Synthesis module generates the required queries and recovers each sign description from the relational database. Using these descriptions and the avatar definition, the Gesture Synthesis module defines an animation for each “animation bone”. A bone animation is defined by a sequence of orientations with a time stamp; this sequence is applied to a specific bone. Using the created animations, the Render module generates visual output that can be displayed in real time or stored in a video file for off-line visualization.

Our innovative approach stores the sign definitions in a unique and centralized database. This approach offers

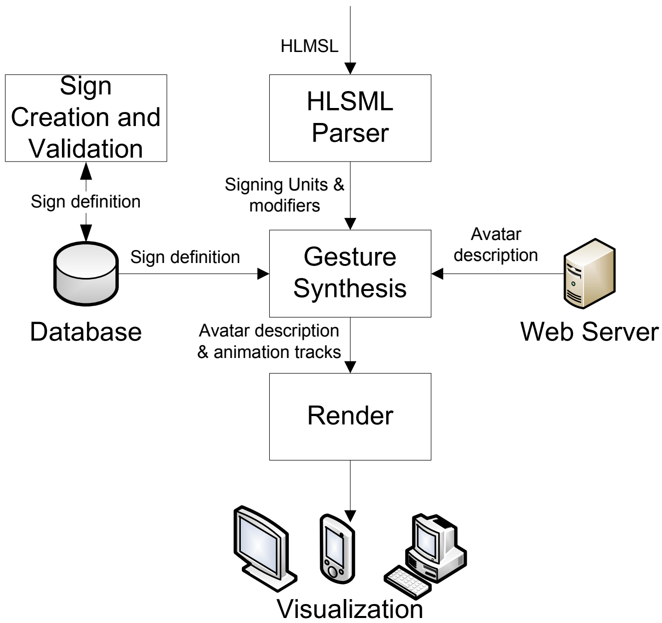


Figure 15: Scheme of the LSE Synthesizer Architecture

some advantages over notation-based approaches. These advantages include the following: i) every modification or addition to the sign dictionary is immediately accessible for every user, while the notation-based approach requires that modifications or additions be distributed for use; ii) a sign definition might not have been verified by an expert, and it might not be accurate; our approach provides an opportunity to ensure the verification of sign definitions. However, our approach has two main disadvantages: i) the synthesis relies on the connection between the Gesture Synthesis module and the relational database; if this connection cannot be established, the synthesis process is not possible; ii) the process of creating new signs is slower than in the notation-based approach, because our proposal requires that sign experts perform the insertion and/or validation of new signs in the database.

8. Results and evaluation

8.1. Implementation of the synthesizer

The synthesis system was created using Java. The 3D graphic library used was the Hybrid Rasteroid 3 API [57], which is a J2SE implementation of the standard Java Mobile 3D Graphics API, JSR-184 [58]. The resulting implementation was run on a PC, Intel Core 2 Duo 2.4GHz with 3.12GB of RAM memory and 512MB of video memory. This system runs under the operating system "Windows XP SP3" and JRE 1.6.0. This system configuration was used to produce all the videos used in the user tests.

8.2. User evaluations

The evaluation group was composed of 10 LSE natives (7 males and 4 females) aged between 24 and 50. Two of these users work as linguistic experts on the FCNSE; thus their comments represent not only the points of view of LSE users but also those of LSE experts. It is important to note that the regional dialect variations that exist in LSE are not likely to have influenced the results of the evaluations because all the users in the evaluation group live in the same city. Such regional variations decrease the recognition rate because the same concept is represented by a different dictionary sign depending on the dialect, as reported in [59]. Although the evaluation group users were LSE natives, they had medium-level knowledge of written Spanish that was sufficient to allow them to understand the written instructions of the evaluations. They are also experienced PC users; so we could therefore send them the URLs of the evaluations by e-mail. Eight of the users were high school graduates; the other three possessed bachelor's degrees.

The evaluations were designed to simulate the standard conditions of the synthesizer, where no additional help is possible. For this reason, all the evaluations were designed as web forms in which, an initial message was used to instruct the user about the concrete objective of the test and how they should manifest the answers. In order to emulate the standard conditions, the users neither previewed the avatar nor performed initial training to adapt them to the avatar's signing. We assume this testing approach is detrimental to the results, but we consider that it represents more clearly the quality and signing comprehension level of the synthetic message.

We carried out three different evaluations. Two of these were related to isolated sign evaluation, and the last dealt with complete sentences. The first evaluation, which represented the first encounter with the avatar, was a free-answer test using isolated signs. The isolated sign evaluation conducted in this test only evaluates morphological aspects of the signing and ignores the message syntactics. In this evaluation, the user was instructed to focus only on the signing performance and not to base his/her answer on the context. Each user was asked to write in the web form the concept that is being represented by the avatar. If the sign was not recognized, the user was allowed to leave the answer empty and proceed to the next sign. Once the user had answered a question and proceeded to the next question, s/he was not allowed to return and correct it. This limitation enhances the "first opinion" approach we wanted to obtain. The evaluation consisted of twenty different signs chosen to represent all kinds of signs, including: single- and double-handed signs and signs with and without non-hand parameter definitions.

The result obtained in the first evaluation showed an isolated sign recognition rate of 90.5%. In considering the the context-free approach, we must note that in LSE some different concepts are represented using the same sign (e.g., WATER and TO DRINK); thus, considering all possible answers, the real recognition rate achieved using our synthesis approach was 96.5%. Figure 16 shows three different signs obtained with our synthesizer.

The second evaluation carried out by the users consisted of a multiple-choice test. This evaluation presented a different set of twenty signs to the users; five possible answers (only one was correct) were presented for each sign. Like the previous one, this evaluation did not allow the users to change their answers to previous questions. The recognition rate on the second evaluation was 98.5%. We must note that, in order to avoid any training effect from it, the users were not notified of the results of the first evaluation. The increment of recognition rate shown in the second evaluation is attributed to the different evaluation approach; the user, aware of the fact that the correct answer is always present, will choose the most plausible option. For this reason, the use of multiple-choice based evaluations is less preferred.

The third evaluation presented ten sentences to the users, which contained between three and ten signs each; the sentences were proposed by an LSE expert and did not contain syntactical or grammatical errors. For the most part, the test sentences consisted of dictionary signs; however, they also included fingerspelling sequences and one iconic classifier that was used to refer to one of the

(a) Sign RUN. (b) Sign YOURS.[a] (c) Sign EQUAL.

[a]This sign is performed only with the right hand, the left hand holds the position from the previous sign. This is commonly observed in humans when signing.

Figure 16: Several signs obtained with our synthesizer.

objects mentioned in a sentence. This evaluation was presented to the users in the form of a multiple-choice test in which, five possible answers, only one of which was correct, were presented for each sentence. The sentences were not topic-related; thus, the users could not use the topic of the previous sentences to advise their choice. This evaluation yielded 95% correct answers. We have included some of the sentences in Table 3.

8.3. Comparison with other works

Previous work [60, 25] has attempted to evaluate SL translation systems as whole by considering sign quality and complexity of interaction with the avatar. San-Segundo et al. [61] proposed a 2D avatar for presenting fifty different gestures from the fingerspelling dictionary; in their evaluation of this system, more than 70% of the letters were correctly recognized in a first attempt and 100% were recognized in a second attempt after the users were presented with their earlier results. The recognition rate of the eSign system is 81% for isolated signs and 61% for complete sentences. Our system yielded a recognition rate of $96.5\%--98.5\%$ for isolated signs and 95% for complete sentences. The increased recognition rate of our system for isolated signs can clearly be measured; however, determination of the recognition rate for complete sentences in the eSign system involved evaluation of both the synthesis and the machine translation system, thus a a direct comparison cannot be presented. Both systems show a decrease in recognition rate when complete sentences are compared to isolated signs. It may be that the larger difference found in the eSign system is due only to the machine translation; alternatively, this difference may be related to the sign concatenation algorithm. Resolution of this uncertainty will require further evaluation and possibly a standard evaluation method.

Table 3: Examples of sentences used in the evaluation.

Sentence:	I have read a book which is very interesting
LSE:	I BOOK position-ix[a]READ PAST, BOOK position-ix INTERESTING VERY
Sentence:	My name is Fernando
LSE:	I NAME spell-FERNANDO
Sentence:	I am very happy because I am walking with you
LSE:	I HAPPY-very CAUSE I WALK NEAR YOU
Sentence:	Your car is as fast as mine
LSE:	CAR YOURS CAR MINE RUN EQUAL

[a] This element is used like a pronoun for later referring.

9. Conclusion and Future Work

This paper presents a novel approach to LSE synthesis. The approach is based on a sign-oriented skeleton design and on storage of the sign definitions in a relational database. The relational database approach and the developed tools presented here allow a highly detailed and simple definition and update of the signs made by Deaf people, who are the main users of the synthesizer. The relational database also makes possible alternative definitions of the signs in order to store mood and prosodic sign variations and different SL dialect representations.

This work presents an architecture that promotes collaboration between sign experts and developers, thus facilitating sign definition and validation. The use of a centralized database in this architecture allows sign definition to be validated by an LSE expert organization (FCNSE), thus ensuring the quality and correctness of the synthetic signs.

We have also proposed the use of HLSML, a multi-level notation that allows the user to define a sign message without any knowledge of how signs are described or generated. This notation makes possible the definition of a sign message in a variety of dialects that share the same grammar rules. Because it describes a gloss-based level, this notation also increases the usability of SL synthesizers. Our synthesizer is the first to integrate multiple phonologically different units of the LSE in the same definition using this notation. The availability of HLSML notation will foster collaboration among translation experts who lack LSE notation knowledge and will facilitate the integration of this LSE synthesizer in multimodal dialog systems.

The recognition rate obtained with this approach improved on the results that have been reported for other SL synthesizers [60, 25, 61] both for isolated signs and for complete sentences. However, as we previously stated, only the isolated sign recognition rates can be compared.

9.1. Future work

The development of the LSE synthesizer described in this work is the basis for further experiments and numerous applications.

In describing the relational database structure, we presented one advantage of this approach for storing the sign definitions, which is the recording of different variations for each PP unit (Figure 12). This method allows the generation of alternative versions of the same sign and makes possible to represent the way in which a sign is performed by signers in different states of mind. Although these modifications have been introduced into the synthesis process described in this work, further study and improvement must be carried out.

The synthesizer has been implemented using a Java 3D Mobile-based graphic library, although a desktop implementation could have used other graphic libraries. We chose this 3D graphic standard because it is included in mobile phones and as a result so the adaptation required to create a mobile version will be easier. We are developing different versions of the synthesizer that are adapted to several mobile devices depending on the hardware and network restrictions. This approach will allow us to provide signed contents in the absence of a desktop computer.

The results of the evaluations show that there are some flaws in our design and that these flaws reduce the sign recognition rate. In order to improve the quality of the synthesis, we are working side-by-side with FCNSE's LSE experts. We are also collaborating with this group of experts to define the elements that must be tested in a synthetic sign message (resolution, mesh detail, colors, speed, etc). Our objective is to define a standard evaluation protocol.

Acknowledgment

We acknowledge the support and advice of the CNSE LSE linguistic unit and the testing users for their collaboration during the evaluations.

[1] Gobierno de España. Ley 51/2003, de 2 de diciembre, de igualdad de oportunidades, no discriminacin y accesibilidad universal de las personas con discapacidad. B.O.E. 3 December, 2003. (Spanish Law).

[2] Andrew Bangham, Stephen Cox, Ralph Elliot, John Glauert, and Ian Marshall. Virtual signing: Capture, animation, storage and transmission - an overview of the visicast project. In *IEE Seminary on Speech and Language Processing for Disabled and Elderly People*, 2000.

[3] Inge Zwiterslood, Margriet Verlinden, Johan Ros, and Sanny van der Schoot. Synthetic signing for the deaf: esign. In *Proceedings of the Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment*, Granada, Spain, June 2004. CVHI.

[4] Eleni Efthimiou, Stavroula-Evita Fontinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Chrisophe Collet, Petros Maragos, and Francois Goudenove. Dicta-sign - sign language recognition, generation and modelling: A research effort with applications in deaf communication. In *Proceedings of the* 4^{th} *Workshop on the Representation and Processing of Sign Lanugages: Corpora and Sign Language Technologies*, pages 80–83, 2010.

[5] Siegmund Prillwitz, Regina Leven, Heiko Zienert, Thomas Hanke, and Jan Herming. *HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductory guide.* Signum-Verlag, 1989. ISBN 978-3-927731-01-1.

[6] Thomas Hanke. Hamnosys - representing sign language data in language resources and language processing contexts. In Springer Berlin / Heidelberg, editor, *Proceedings of LREC*, pages 1–6, Lisbon, May 2004.

[7] Roberto Borgotallo, Carmen Marino, Elio Piccolo, Paolo Prinetto, Gabriele Tiotto, and Mauro Rossini. A multilanguage database for supporting sign language translation and synthesis. In *Proceedings of LREC, 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 23–26, Valletta, Malta, May 2010.

[8] R. San-Segundo, JM Montero, J. Macías-Guarasa, R. Córdoba, J. Ferreiros, and JM Pardo. Generating gestures from speech. *Proc. International Conference on Spoken Language Processing (ICSLP?2004)*, 3:1817–1820, 2004.

[9] Rubén San-Segundo, Roberto Barra, Luis Fernando D'Haro, Juan Manuel Montero, Ricardo Córdoba, and Javier Ferreiros. A spanish speech to sign language translation system for assisting deaf-mute people. In *Proceedings INTERSPEECH-2006*, pages 1399–1402, Pittsburgh. USA, September 2006. Interspeech.

[10] R. San-Segundo, A. Pérez, D. Ortiz, LF D?Haro, MI Torres, and F. Casacuberta. Evaluation of alternatives on speech to sign language translation. In *Proceedings of Interspeech 2007*, Antwerp, Belgium, August 2007.

[11] Rubén San-Segundo, Verónica López, Raquel Martín, David Sánchez, and Adolfo García. Language resources for spanish - spanish sign language (lse) translation. In *Proceedings of LREC, 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 208–211, Valletta, Malta, May 2010.

[12] Franc Solina, Slavko Krapez, Ales Jaklic, and Vito Komac. *Design and Management of Multimedia Information Systems*, chapter 13: Multimedia Dictionary and Synthesis of Sign Language, pages 268–281. Idea Group Publishing, 2001.

[13] European Media Masters of Art. Vsign. http://www.vsign.nl/, 2002.

[14] Informatics and Telematics Institute. Vsigns. http://vsigns.iti.gr:8080/VSigns, 2004.

[15] Seiji Igi, Megumu Ujitani, Mitsue Tamaru, Yasunori Yamamoto, and Shigeharu Sugita. Sign-language synthesis for mobile environments. In *Proceedings of The 11 International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, 2003. URL `http://wscg.zcu.cz/wscg2003/Papers_2003/C19.pdf`.

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

[17] ISO/IEC 14772-1:1997. *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, 1997.

[18] ISO/IEC 14772-2:2004. *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, 2004.

[19] ISO/IEC 19775:2004. *Information technology – Computer graphics and image processing – Extensible 3D (X3D)*. International Organization for Standardization, Geneva, Switzerland, 2004.

[20] Kyle Duarte and Sylvie Gibet. Heterogeneous data sources for signed language analysis and synthesis: The signcom project. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, pages 461–468, Valletta, Malta, may 2010. European Language Resources Association (ELRA). ISBN 2-9517408-6-7.

[21] Richard Kennaway. Synthetic animation of deaf signing gestures. *LNCS*, 2298:146–157, 2002.

[22] Valerie Sutton. Signwriting. http://www.signwriting.org/, 1974.

[23] Richard Kennaway, Ralph Elliot, John Glauert, and K. J. Parsons. Sigml document type definition (dtd). Visicast deliverable, ViSiCAST Project, 2002. URL `http://www.visicast.cmp.uea.ac.uk/sigml/`.

[24] Antnio Carlos da Rocha and Graaliz Pereira. Supporting deaf sign languages in written form on the web. In *Proceedings of LREC*, pages 26–28, Lisbon, May 2004.

[25] Richard Kennaway, John Glauert, and Inge Zwitserlood. Providing signed content on the internet by synthesized animation. *ACM Transactions on Computer-Human Interaction*, 14(15): 1–29, 2007. ISSN 1073-0516. doi: http://doi.acm.org/10.1145/1279700.1279705.

[26] Ralph Elliott, John Glauert, Richard Kennaway, Ian Marshal, and Eva Sáfár. Linguistic modelling and language-processing technologies for avatar-based sign language presentation. *UAIS*, 6(4):375–391, 2008. ISSN 1615-5289. doi: http://dx.doi.org/10.1007/s10209-007-0102-z.

[27] M. Papadogiorgaki, N. Grammalidis, L. Makris, and M.G. Strintzis. Gesture synthesis from sign language notation using mpeg-4 humanoid animation parameters and inverse kinematics. In *Proceedings of the International Conference on Intelligent Environments*, volume 1, pages 151–160, July 2006.

[28] Maria Papadogiorgaki, Nikos Grammalidis, Nikos Sarris, and Michael Strintzis. Synthesis of virtual reality animation form swml using mpeg-4 body animation parameters. In *Proceedings of LREC*, Lisbon, May 2004.

[29] Carlos Machado. Adapting an efficient entry system for signl languages. In *Proceedings of the 4^{th} Workshop on the Representation and Processing of Sign Lanugages: Corpora and Sign Language Technologies*, pages 147–149, 2010.

[30] Angus B. Grieve-Smith. Signsynth: A sign language synthesis application using web3d and perl. In *Revised Papers from the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction*, pages 134–145, London, UK, 2002. Springer-Verlag. ISBN 3-540-43678-2.

[31] Jarosław Francik and Piotr Fabian. Animating sign language in the real time. In *Proceedings of 20th IASTED International Multi-Conference Applied Informatics*, pages 276–281, 2002.

[32] Onno Crasborn, Harry van der Hulst, and Els van der Kooij. Signphon: A database tool for cross-linguistic phonological analysis of sign languages. In *Proceedings of the 6th International Conference on Theoretical Issues in Sign Language Research*, 1998.

[33] O. Crasborn, H. van der Hulst, and E. van der Kooij. Signphon: A phonological database for sign languages. *Sign Language & Linguistics*, 4(1-2):215–228, 2001.

[34] Jacob Furst, Karen Alkoby, Andre Berthiaume, Pattaraporn Chomwong, Mary Jo, Brian Konie, Glenn Lancaster, Steven Lytinen, John Mcdonald, Jorge Toro, Noriko Tomuro, and Rosalee Wolfe. Database design for american sign language. In *Proceedings of 15th International Conference on Computers and their Applications*, pages 427–430, 2000.

[35] M. Ángeles Rodríguez González. *Lenguaje de Signos*. Fundación CNSE, Barcelona, 1992. URL `http://www.cervantesvirtual.com/FichaObra.html?Ref=10207`.

[36] Irma María Muñoz Baell. La querología aplicada a la enseñanza de la lse. *Revista Española de Lingüística de las Lenguas de Signos*, 2:1–18, 2000.

[37] Ángel Herrero Blanco. *Gramática Didáctica de la Lengua de Signos Española (LSE)*. Ediciones SM, 1 edition, 2009. ISBN 978-84-675-9809-4.

[38] Adam Schembri. *Perspectives on Classifier Constructions in Sign Languages*, chapter 1: Rethinking 'classifiers' in Signed Languages, pages 3–34. Psychology Press, 2003.

[39] Scott K. Liddell and Melanie Metzger. Gesture in sign language discourse. *Journal of Pragmatics*, 30(6):657–697, 1998. ISSN 0378-2166. doi: DOI:10.1016/S0378-2166(98)00061-7. URL `http://www.sciencedirect.com/science/article/B6VCW-3VRVF9X-1/2/607db68f9afb4f2eff37a7b6db28dce8`.

[40] Scott K. Liddell. *Perspectives on Classifier Constructions in Sign Languages*, chapter Sources of Meaning in ASL Classifier Predicates, pages 199–220. Psychology Press, 2003.

[41] Dorothea Cogill-Koez. A model of signed language 'classifier predicates' as templated visual representation. *Sign Language and Linguistics*, 3(2):209–236, 2000.

[42] Ángel Herrero Blanco. Una aproximación morfológica a las construcciones clasificatorias en la lengua de signos española. *Estudios de Lingüística de la universidad de Alicante*, 18:151–167, 2004. URL `http://hdl.handle.net/10045/6135`.

[43] Matt Huenerfauth. Spatial representation of classifier predicates for machine translation into american sign language. In *Proceedings of LREC*, pages 24–31, Lisbon, 2004.

[44] Matt Huenerfauth. *Generating american sign language classifier predicates for english-to-asl machine translation*. PhD thesis, Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA, 2006. Adviser Mitch Marcus and Martha Palmer.

[45] Fernando López-Colino and José Colás. The synthesis of lse classifiers: From representation to evaluation. *Journal of Universal Computer Science*, 17:xx–xx, 2011.

[46] Vince Jennings, Ralph Elliot, Richard Kennaway, and John Glauert. Requirements for a signing avatar. In *Proceedings of LREC, 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 133–136, Valletta, Malta, May 2010.

[47] Ronnie Wilbur. *The Signs of Language Revisited: An Anthology To Honor Ursula Bellugi and Edward Klima*, chapter Phonological and Prosodic Layering of Nonmanuals in American Sign Language, pages 213 – 241. Psychology Press, 2000.

[48] Charlotte Baker and Carol Padden. *Understanding Language Through Sign Language Research (Perspectives in neurolinguistics and psycholinguistics)*, chapter Understanding Language Through Sign Language Research (Perspectives in neurolinguistics and psycholinguistics), pages 27–57. Academic Press, 1978.

[49] John McDonald, Karen Alkoby, Roymieco Carter, Juliet Christopher, Mary Jo Davidson, Dan Ethridge, Jacob Furst, Damien Hinkle, Glenn Lancaster, Lori Smallwood, Nedjla Ougouag-tiouririne, Jorge Toro, Shuang Xu, and Rosalee Wolfe. A direct method for positioning the arms of a human model. In *Proceedings of Graphics Interface*, pages 99–106, 2002.

[50] JiHyung Lee, Yoon-Seok Choi, Bon-Ki Koo, and Chi Jung Hwang. An intuitive system for 3d avatar with high-quality. In *Consumer Electronics (ICCE), 2010 Digest of Technical Papers International Conference on*, pages 517 –518, 2010. doi: 10.1109/ICCE.2010.5418776.

[51] Duan Xinyu and Jiang Pin. Three dimension human body format and its virtual avatar animation application. In *Intelligent Information Technology Application, 2008. IITA '08. Second International Symposium on*, volume 2, pages 1016 –1019, 2008. doi: 10.1109/IITA.2008.191.

[52] S. Piotr and K. Bozena. Personalized avatar animation for virtual reality. In *Information Technology, 2008. IT 2008. 1st International Conference on*, pages 1 –4, May 2008. doi: 10.1109/INFTECH.2008.4621655.

[53] Zdenek Krnoul. New features in synthesis of sign language addressing non-manual component. In *Proceedings of LREC, 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 143–146, Valletta, Malta, May 2010.

[54] Sir William Rowan Hamilton. *Lectures on Quaternions*. Royal Irish Academy, 1853.

[55] Ken Shoemake. Animating rotation with quaternion curves. In *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, pages 245–254, New York, NY, USA, 1985. ACM. ISBN 0-89791-166-0. doi: http://doi.acm.org/10.1145/325334.325242.

[56] Oscar Martínez Lazalde, Steve Maddock, and Michael Meredith. A constraint-based approach to visual speech for a mexican-spanish talking head. *International Journal of Computer Games Technology*, 2008(3):1–7, January 2008. ISSN 1687-7047. doi: http://dx.doi.org/10.1155/2008/412056.

[57] Hybrid. Hybrid rasteroid. http://www.hybrid.fi/, 2006.

[58] Java Community Process. Jsr-184. mobile 3d graphics api for j2me. http://www.jcp.org/en/jsr/detail?id=184, 2005.

[59] R. San-Segundo, R. Barra, R. Córdoba, L. F. D'Haro, F. Fernández, J. Ferreiros, J. M. Lucas, J. Macías-Guarasa, J. M. Montero, and J. M. Pardo. Speech to sign language translation system for spanish. *Speech Communication*, 50 (11-12):1009–1020, November 2008. ISSN 0167-6393. doi: http://dx.doi.org/10.1016/j.specom.2008.02.001.

[60] Stephen Cox, Michel Lincoln, Melanie Nakisa, Mark Wells, Marcus Tutt, and Sanja Abbott. The development and evaluation of a speech-to-sign translation system to assist transactions. *International Journal of Human-Computer Interaction*, 16(2):141–161, October 2003. doi: 10.1207/S15327590IJHC1602_02.

[61] R. San-Segundo, J. M. Montero, J. Macías-Guarasa, R. Córdoba, J. Ferreiros, and J. M. Pardo. Proposing a speech to gesture translation architecture for spanish deaf people. *Journal of Visual Languages and Computing*, 19(5):523–538, October 2008. ISSN 1045-926X. doi: http://dx.doi.org/10.1016/j.jvlc.2007.06.002.