



Interactive Editing in French Sign Language Dedicated to Virtual Signers: Requirements and Challenges

Sylvie Gibet, François Lefebvre-Albaret, Ludovic Hamon, Rémi Brun, Ahmed Turki

► To cite this version:

Sylvie Gibet, François Lefebvre-Albaret, Ludovic Hamon, Rémi Brun, Ahmed Turki. Interactive Editing in French Sign Language Dedicated to Virtual Signers: Requirements and Challenges. Universal Access in the Information Society, Springer Verlag, 2016, 15 (4), pp.525-539. 10.1007/s10209-015-0411-6 . hal-01367723

HAL Id: hal-01367723

<https://hal.archives-ouvertes.fr/hal-01367723>

Submitted on 9 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interactive Editing in French Sign Language Dedicated to Virtual Signers: Requirements and Challenges

Sylvie Gibet · François Lefebvre-Albaret · Ludovic Hamon · Rémi Brun · Ahmed Turki

Received: date / Accepted: date

Abstract Signing avatars are increasingly used as an interface for communication to the deaf community. In recent years, an emerging approach uses captured data to edit and generate Sign Language (SL) gestures. Thanks to motion editing operations (*e.g.* concatenation, mixing, etc.), this method offers the possibility to compose new utterances, thus facilitating the enrichment of the original corpus, enhancing the natural look of the animation, and promoting the avatar's acceptability. However, designing such an editing system raises many questions. In particular, manipulating existing movements does not guarantee the semantic consistency of the reconstructed actions. A solution is to insert the human operator in a loop for constructing new utterances, and to incorporate within the utterance's structure constraints that are derived from linguistic patterns. This article discusses the main requirements for the whole pipeline design of interactive virtual signers, including: (i) the creation of corpora, (ii) the needed resources for motion recording, (iii) the annotation process as the heart of the SL editing process, (iv) the building, indexing and querying of a

motion database, (v) the virtual avatar animation by editing and composing motion segments, and (vi) the conception of a dedicated user interface according to user knowledge and abilities. Each step is illustrated by the authors' recent work and results from the project Sign3D, *i.e.* an editing system of French Sign Language (LSF) content (<http://sign3d.websourd.org/sltat>).

Keywords Signing Avatar · Interactive Editing · Data-Driven Synthesis

1 Introduction

New communication technologies, such as mobile phones with integrated video, MMS messaging, videoconferencing devices, internet and social networks, have been developed in recent years to improve Deaf accessibility to various media. These new assistive technologies usually contain pre-recorded videos adapted to a visual-gesture communication channel, in which a human signer performs LSF messages. However, the video does not fulfill all the functions of writing languages: it does not allow anonymity and sometimes imposes stringent recording requirements. Furthermore, video lacks flexibility, and operations such as copy / paste fail to deal with transitions in the context of editing new utterances.

One of the current assistive technologies is centered on the generation of signed languages using virtual humans, or avatars. The use of avatars allows more advanced manipulations of SL statements, mostly because the possibilities of content creation with avatars through editing operations are far more advanced and flexible. Importantly, 3D animation presents an opportunity to provide on-the-fly access to otherwise inaccessible content, thus facilitating the comprehension of the signs.

Sylvie Gibet and Ludovic Hamon
IRISA, Université de Bretagne Sud
Campus de Tohannic, BP 573
56017, Vannes, France
E-mail: sylvie.gibet@univ-ubs.fr, ludovic.hamon@univ-ubs.fr

François Lefebvre-Albaret
Websourd
99, route d'Espagne, Bat. A
31100 Toulouse, France
E-mail: francois.lefebvre-albaret@websourd.org

Rémi Brun and Ahmed Turki
Mocaplab
5, cité Riverin, 75010 Paris, France
E-mail: remi.brun@mocaplab.com,
ahmed.turki@mocaplab.com

Avatars also allow for the anonymity of the interlocutor, and can be personalized along with the user’s will.

Most signing avatar projects adopt synthetic animation techniques for their virtual agents. Some of them use keyframe techniques to produce high quality animations [1], or pure synthesis methods to generate virtual avatar animation thanks to inverse kinematic techniques [34,31]. However, these methods have been proven to be tedious for building new signs or utterances. In addition, the resulting movements may fail in convincing audiences with their credibility and expressivity, thus leading to a poor acceptability of the avatar by the deaf community. The authors argue that using captured motion on real signers might help to understand the main features of the sign generation mechanisms, and reinforce the virtual avatar acceptability for the deaf community. Furthermore, such data-driven approaches go beyond the production of realistic animations: they avoid the tedious task of computing all the joint values of the skeleton for an animation sequence, and make possible the use of pre-recorded motion, thus capturing the relevant key-postures and preserving the dynamics of the original motion. Furthermore, composing new motions from editing captured motion can provide new ways of manipulating, transporting, or visualizing sign language utterances that are not contained in the original database.

This paper focuses on a fully data-driven virtual signer. Some important issues have already been studied and implemented in previous works (see for example the whole pipeline production system described and evaluated in [20]). The goal here is to consider the knowledge and skills acquired in previous works, both theoretically and practically, and extract from this knowledge useful information for the development of future applications; more specifically, previous works are generalized and the main requirements and challenges for building an interactive editing system from pre-recorded motion captured data are discussed. Each step is illustrated with current results related to the Sign3D project [4], which aims to build a range of innovative tools for storing and processing high-definition LSF content. Section 2 describes two major approaches to generate new sign language (SL) utterances: the textual-based SL specifying approach dedicated to pure synthesis techniques and the editing operations applied to data-driven animations. The creation of an appropriate corpus and the needed resources to record SL gestures are exposed in section 3. At the heart of the SL editing system is the annotation process which is presented in section 4. Section 5 highlights the key issues for building and accessing an efficient heterogeneous database containing both the captured and annotated motion. Sec-

tion 6 discusses the editing and composing operations of the animation system, while section 7 is centered on the requirements for a dedicated user’s interface. Finally a discussion and some perspectives conclude this article.

2 Related Works

Among recent methods and technologies for editing sign language and animating signing avatars, two major classes of approaches can be distinguished: specifying sign language associated to pure synthesis animation techniques, or editing movement for data-driven animation techniques. In order to produce understandable and accurate animations, both approaches require a fine comprehension of the linguistic mechanisms underlying the production of signed utterances.

2.1 From Specifying Sign Language Utterances to Procedural Synthesis

Different linguistic stages for synthesis mediated by an avatar can be considered, including syntactic, semantic, morphological, and phonetic issues. Until recently, signing avatars have focused on the generation of signing utterances given a phonetic or phonological description of the sign sequence, with methods ranging from notation systems to formalized scripts or dedicated gestural languages. Initial studies on sign languages formed early description / transcription systems, as for example the *HamNoSys* notation [41]. One of the first virtual signers was based on a quantified description of the signing space, associated to a phonological description of hand-arm movements [17]. Starting from *HamNoSys* as initial input description, a similar XML-based notation language called *SigML* has recently been developed [31, 13]. This language has been recently extended by introducing the *PDTs* segmental model of Johnson and Liddell to provide greater flexibility and precision of control [21]. A sign language animation tool also uses inverse kinematics to synthesize an animation from a list of hand location targets [27]. Other researches fall into the category of linguistic modeling for sign language generation. Among these studies, higher-level language processing techniques are used for the formation of signs (*e.g* [38, 40, 25, 16, 14]).

Several major challenges are still to be faced as regards the synthesis of SL gestures mediated by avatars. One of them consists of taking into account the variational aspects of sign languages, and in particular the spatial features that depend on the context of the sentence, and are conveyed simultaneously by parallel chan-

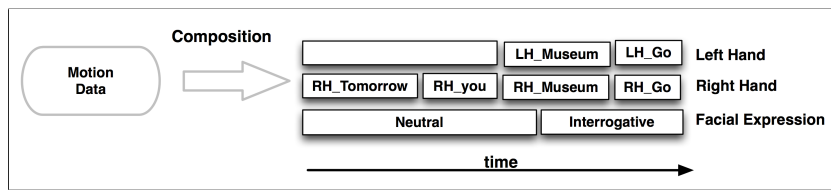


Fig. 1 Composition process from motion data.

nels (*i.e.*, gestures of the two arms and the two hands, facial expressions, and gaze direction). Thus, several studies have emphasized the importance of the synchronization schemes that can be observed between the different channels that compose signs [34, 25, 15]. Furthermore, many avatar systems ignore the importance of facial expressions and gaze direction even though these components are crucial to comprehension in signed languages. Some recent approaches address this important aspect by using pure synthesis techniques [16, 44]. However, one of the main criticisms made by deaf signers regarding virtual agents is the lack of realism and expressiveness of avatars.

Data-driven animation methods can be substituted for the above-discussed pure synthesis methods, making the avatar more expressive and human-like.

2.2 Data-Driven Synthesis

In computer animation, most of the previous work on data-driven methods present editing and composing techniques, with an emphasis on the re-use of motion chunks and the adaptation of captured motion for creating new motion sequences. By using the *motion chunk* term, reference is made to a motion limited to a time interval, compared to the original captured motion with a SL meaning conveyed by the whole body or a sub-part.

The motion processing techniques are mostly based on statistical or optimization methods applied on whole-body motion chunks. If some relevant works rely on qualitative annotations of motion clips (*e.g.* [2]), very few approaches deal with both motion-captured data and their implicit semantic content. Using signs as whole global segments, a first animation system has been developed, which uses a dual database: a semantic one containing the annotations, and a raw motion one containing the captured motion. Initial experiments have demonstrated the feasibility and the interest of the approach for *filling-the-gap* synthesis of novel sentences in French Sign Language [3]. Within the European ViSi-CAST project, motion captured data has also been used to evaluate sequences of signs in British Sign Language [5]. A fully data-driven virtual signer has then been designed and implemented in the SignCom project. The

architecture of the animation system, and its ability to build sentences in LSF has been described and evaluated in [20].

As an added benefit, motion capture data provides analytical material from which to extract specific features or parse generic features of signed languages, such as the dynamics of the movements or the spatial-temporal relationship between channels, etc. Analysis studies have thus been achieved to extract hand motion invariants in LSF [35, 10]. Other works using motion captured data have focused on the characterization of style for whole sequences [24], or segmented sequences on parallel channels [23].

Hybrid approaches combine procedural techniques with the use of data expressed as key-frames or motion chunks [27, 37]. Another study uses a system of procedural techniques to layer data from annotation tiers [45]. In this approach, sparse key-frame data are combined and modified, rather than using motion chunks from a motion capture database.

2.3 Requirements and Constraints of a Sign Language Editing System

One of the key points of an interactive system for editing gestures in LSF is to produce novel utterances from the corpus data by combining motion chunks that have been previously captured on a real signer, and using these data to animate a virtual signer. As gestures in signed languages are by essence multi-channel, *i.e.* meaningful information is conveyed by multiple body parts acting in parallel, it follows that a sign editing system must manipulate motion segments that are spatially decomposed on these channels over time. This composition process is illustrated in Figure 1.

The editing system should be able to retrieve the signs or the elements of signs, also called motion chunks, in the pre-recorded sequences. One of the main challenges is to be able to accurately extract the SL items from the database to compose new sentences through editing operations and animate the virtual signer. A particular requirement to fulfill this objective is to properly annotate SL data and to handle correctly, in terms

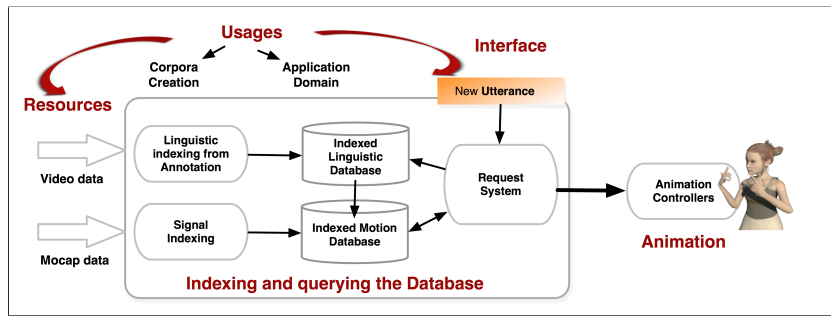


Fig. 2 Editing system.

of linguistics structures, both the raw data and its associated semantics. It must be noted that the sequence of motion elements to be combined does not necessarily respect grammatical LSF rules, the composition process being guided interactively by an expert user. Furthermore, finding the best rules or the best features that enable the extraction of the appropriate sign variant depending on the produced sentence, is part of the motion retrieval process that can be built thanks to user interactions or expertise. The motion chunks must then be concatenated spatially (*i.e.* along the channels) and temporally within an animation system. This last system must synchronize and handle at the same time several modalities involved in signed gestures and produce a continuous flow of animation.

The Sign3D project proposes a full data-base framework to generate sign language sentences from the concatenation of annotated motion captured items [4]. This editing system is based on the following conceptual scheme centered on the indexing and querying of the LSF database to build new utterances and generate animations of the virtual signer (Figure 2). The various modules of this architecture is described below, detailing the challenges and the main constraints and limitations of such an editing system. They include (i) the definition of usages, leading to corpora and resources creation, (ii) the annotation process, (iii) the indexing and querying of the heterogeneous database, (iv) the animation by editing and composing motion segments, and, (v) the human machine interface for the construction of new utterances in LSF.

3 Corpus and Resources Creation

3.1 Corpus

The design of a corpus in the framework of a mocap(motion capture)-based editing system is one of the key points of the approach, because the consistency of the synthesis will depend on its completeness. For the purpose

of corpus design, several questions can be addressed. The first one concerns the corpus definition and the compromise that exists between breadth and depth in its design. The second question concerns the nature of the sign variations that should be included in the corpus for further utterance reconstruction. This variability within signs will condition the editing operations: if the signs involving such inflection processes are contained into the corpus, then the editing operations will be less complex. For example, including many depicting verbs with spatial variation will facilitate the construction of novel utterances with verb declination, without recording new signs. Finally, other questions concern the acted or spontaneous nature of the produced SL utterances.

Depth vs. Breath and Variation vs. Consistency. Is a large number of specific signs necessary to represent the largest slice possible of a language (breath), or should a limited number of signs with many variations (depth) be included in the corpus? This question is closely linked to the specific application. If the objective of the editing system is to have a lexicon that covers a broad area, including several thematic domains, then a corpus with a breadth approach would be suitable. If, instead, the goal is to have a limited vocabulary and use it in different sentences, then the depth approach would be best. In this case, many tokens of the same signs will be preferred in the pre-defined vocabulary, with variations depending on the phrase context. In order to facilitate the editing operations, it will be assumed that a same sign (or gloss) is associated to a unique semantic identifier [30], called ID-Gloss, and that each gloss is characterized by the same gesture realization, possibly with kinematic modulations.

Kinematic variations / Linguistic inflections. Several levels of signs' variability should also be considered when building a corpus. Among them, the following can be included:

- Kinematic variations of body parts: It is useful to incorporate into the corpus a pre-defined set of static

hand shapes, dynamic hand movements, postures, etc. Different facial expressions have to be included also.

- Sign: The same sign can be performed in various places of the signing space [36], with various dynamics, facial expressions, in the context of role shift [7], with or without mouthing, etc. Note that the same body part can be affected by very different sign variations, for instance the mouth shape can be influenced both by the mouthing and by the whole facial expression.
- Succession of Signs: The sign context can be taken into account by capturing the same sign in varying its predecessors and successors (e.g. influence of hand shape and hand placement [42]). The inclusion of such sequencing in the corpus allows for the study of coarticulation.
- Syntax: For composition rules prediction, a lot of examples must be gathered to be representative of the SL variability. For example in the *DictaSign* project, short breaks have been included after presenting the discourse main topic [43]. Other projects have developed some corpora with many instances of depicted verbs [39,9].

Several corpora for virtual signers have already been developed, especially to provide signed information in public areas [42], on the Internet [32], or to broadcast transmission of virtual signing. The German Sign Language *DGS-Corpus* associated to the *iLex* tool which identifies tokens annotated in *HamNoSys*, therefore providing a basis for semantic and phonetic indexing, can also be mentioned. The *Dicta-Sign* project has developed several technologies demonstrated via a sign language-aware Web 2.0, based on sign language resources and language models, with the aim to make, edit, and evaluate avatar-based sign language contributions online, similar to the way people nowadays make text-based contributions on the Web [11]. Within the Sign3D project, twenty sentences centered on information in museums were recorded. These sentences concern messages such as *opening hours*, *entrance fees* or *perturbation messages*. The main challenge, in this context, is to have enough variability for each sign, in order to build new consistent utterances following given structural patterns, by recomposition of recorded motion chunks.

3.2 Resources: Motion and Video Recordings

Signs are by nature very dexterous and quick gestures, that involve several channels (arms, hands, body, gaze and facial expressions) simultaneously. To record such gestures, it is necessary to find a compromise between

motion capture cost, measurement (space and time) accuracy, and spontaneity of the production. For example, the ViSiCAST project uses wired Cybergloves and Polhemus magnetic sensors in order to track hands and arms motions [12]. A headset, equipped with an infrared camera, some emitting diodes and reflectors are used to record facial expressions. The overall system performs captures between 30 and 60hz. However, if the motion capture equipment is too invasive, the signer will not be able to sign in a natural way. Furthermore, capturing all channels with an appropriate frequency rate (> 100 Hz) accurately actually pushes motion capture equipment to their very limits. Given the complexity of the synchronization patterns between the different channels, it is not possible to perform separately the capture on the different channels.

Capturing hand movements is also a real challenge, due to the complexity of the hand's biomechanics, and also because of the occlusion of markers that occurs very frequently. Many different systems have been tested for capturing LSF hand movements, as for example Cybergloves [19]. These systems are very difficult to calibrate, and large deviations have been observed during the recordings, which is unacceptable. Furthermore, such external devices need to be synchronized with the mocap system. Finally, the reconstruction process introduces many errors, due to some approximation in the hand modeling, loss of data, and noise in the capturing process. Other problems may arise when using mocap techniques with few markers [20]: in this latter case, some post-processing (in particular motion retargeting through inverse kinematic techniques) is necessary; this process leads to other errors in the data, with imprecision of the hand configurations, potentially altering the meaning of the signs. It is concluded that the mocap techniques are still the better solution, though it is necessary to use a larger number of markers (around 50, each hand having 22 degrees of freedom). This is the same for capturing facial expressions. Many markers are needed (> 50) to be able to accurately represent the facial deformations. In addition, specific methods have to be developed for further editing operations, both for manual and facial data. The building of the skeleton and the accurate positioning of the joints/surface of the mesh are also crucial to ensure a good final positioning of the fingertips in the absolute frame. Facial expressions must be processed differently, since the movements are very different and the data are not conveyed by similar hierarchical skeleton. It is also clear that the gaze direction is an essential modality that should be modeled in SL. The inability to position markers on the eyes inevitably leads to consider capturing their move-



Fig. 3 Motion capture session: the real signer from the recorded video (left) and the 3D mesh of the corresponding virtual signer (right).

ments by other means (for example by using reflective lenses, or using vision-based algorithms).

The sentences of the Sign3D project have been captured thanks to a Vicon T160 optical motion capture system (16 Megapixel cameras) combined with a head-mounted oculometer (MocapLab MLab 50-W). Markers were placed on the whole signer's upper body, including her face and fingers which allows for a complete performance capture (Figure 3, left). After motion capture, the marker set is rigged onto a 3D virtual signer mesh in order to animate both its skeleton and face (Figure 3, right).

4 Annotation of Sign Language Data

The annotation of sign language gestures from real video and captured motion is at the heart of the sign language editing process. As pointed out by Wolfe *et al.*, this annotation process aims to test and identify linguistic features and also provides a way to generate new gestures [45]. Traditionally, a multi-layer annotation scheme is defined, both spatially and temporally, and the annotators work from videos to achieve this annotation process.

The spatial annotation scheme denotes a hierarchical organization of the tiers, that corresponds to parallel channels. These channels take into account on the one hand, linguistics categories, and on the other hand, physical properties or/and signal characteristics. The physical characteristics may represent body articulators, such as for example hand, head, or upper torso; signal characteristics may represent kinematic features such as velocity extrema. All these characteristics can be useful for animation purposes. It is also important

to be able to represent dependencies and synchronization rules between tiers. For example, the hand shapes occur at times that may depend on hand movements. The aim of temporal segmentation is to define the temporal motion chunks on each track and identify a set of attributes / values for each annotated segment. It is straightforward that the level of details of the annotation scheme greatly influences the way the avatar will be controlled. Indeed, the more finely the signs are decomposed on the spatial and temporal tracks, the more precise the motion retrieval process and consequently the quality of the resulting animation.

4.1 Linguistic annotation

The following describes examples of linguistic annotations. It can be seen that these examples explore different levels of SL description, from coarse linguistic elements (glosses), to more accurate elements such as phonological items or phonetic aspects.

Gloss-based annotation, with or without transitions. The first step in the annotation process consists of dividing videos or motion-capture data into lexical units (or glosses) previously defined in the corpus. In order to create plausible animations of meaningful signs, it is essential to understand what connects signs together in the animation sequence. The authors are convinced that these transitions have to be taken into account in an editing system, and that they should be adapted to the animation context: if the location at the end of a sign is far away (respectively close) from the location of the following sign in the utterance, the transitions should be enlarged (respectively compressed) con-

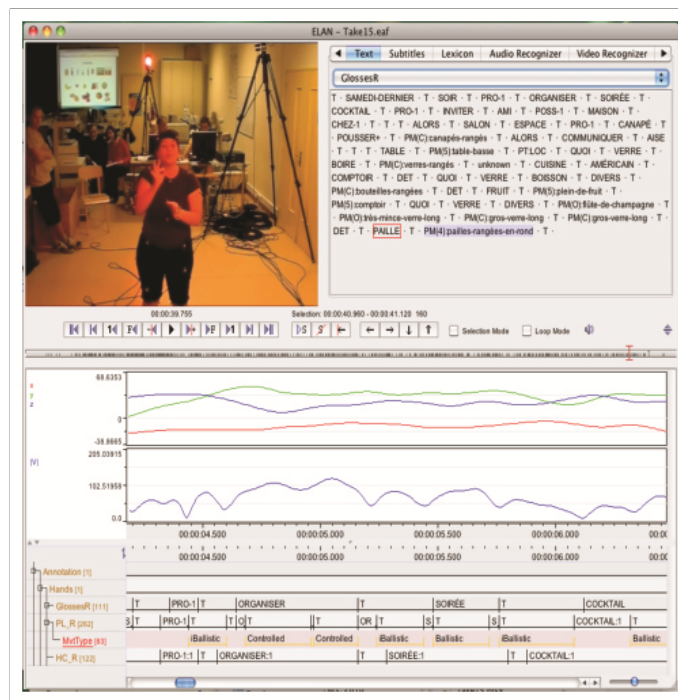


Fig. 4 Phonetic-based annotation [8]; ELAN can be configured to show motion captured data that is synchronized with the annotation timeline [6].

sequently. Under this approach, LSF signs can be segmented with a transition-inclusive system, by separating the Strokes (S) conveying the meaning of the signs from the Preparation (P) and Retraction (R) phases characterizing inter-sign transitions [33,3]. Dealing incorrectly with these transitions may introduce artifacts or non-plausible transitions in the produced animation. It is therefore necessary to introduce co-articulation mechanisms to have acceptable temporal transitions. It should be noted that the gloss annotation is quite limited, since it does not contain information about sign inflections.

Phonological or phonetic based annotation. For a more decomposed description, a phonological annotation scheme can be structured, according to linguistic categories such as handshape, location, hand movement, orientation, facial expression, and gaze direction. The two hand channels should be separated, as some signs can be performed by one or two hands simultaneously, and the motion of one hand can be different from the other one. Furthermore, for each hand, one should distinguish the hand configuration (HC), the hand placement (PL), and the hand orientation (HO). Facial expression can also be represented by one or several tiers, for annotating Mouthing, etc. Such an annotation scheme for animating a virtual signer has been defined in the SignCom project [20,8]. A screenshot of such annotation sequence is shown in Figure 4.

4.2 Physical and signal-based annotation

Physical-based tiers. For the purpose of animation, it is important to identify some sets of joints (physical body parts or articulators) associated with specific tiers. This segmentation level enables the independent control of the articulators, as will be described in the animation section. It is possible in particular to define postures extracted from the mocap database; these postures can be identified during the annotation process, and are given as specific constraints that should be reached by some articulator at some key point of the animation. For example, reaching Hand Configuration targets within a sign may ensure the linguistic coherence of the sign.

Signal-based tiers. In the context of signing avatars driven by mocap data, it is necessary to establish a strong coupling between the motion features extracted from the raw motion and the pre-defined linguistic elements. Mocap data provides raw material from which to extract specific invariant of signed languages, such as the dynamics of the movements (velocity, acceleration, etc.), or the spatial-temporal relationship between production channels, between kinematics and phonetics, etc. These invariants or user-dependent characteristics may be manually identified through an annotation process, or automatically computed through statistical or signal-processing methods, and re-incorporated into the

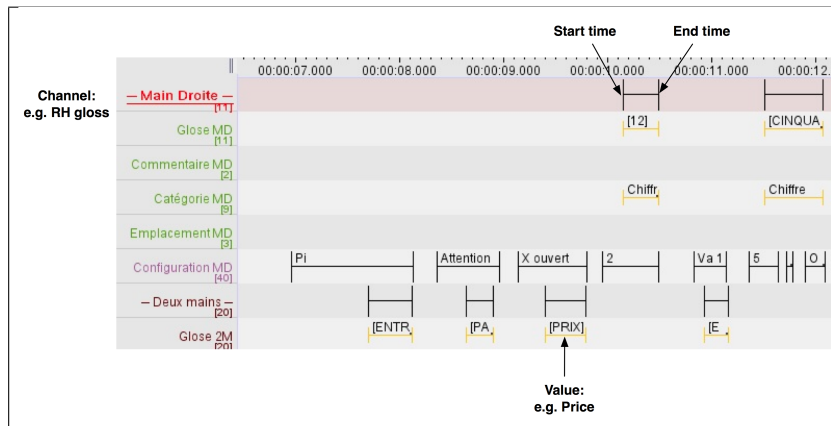


Fig. 5 In the Sign3D project, the LSF glosses are segmented according to a phonological, phonetical and syntactic scheme.

data-driven animation techniques, following an analysis/synthesis approach [18]. For example, velocity extrema computed from motion capture data have been used to accurately position the phonetics elements *PL*, defined by Johnson and Liddell [29,28,8].

Through the authors' involvement in the Sign3D project, the corpus annotation has two main goals:

- Describe the sign features in order to enable a further indexation,
- Formulate, if possible, explicit or implicit rules that will help to choose the best sign variation in context.

During motion capture, a reference video (a frontal view of the signer) is recorded (Figure 3, left). Then, a deaf signer annotates this video with the ELAN software [6], according to a spatial and temporal decomposition. Several channels are created following a phonological, phonetical and syntactic specification scheme (Figure 5). Sentences are then temporally segmented into glosses along channels and labeled by a string conveying its meaning. Other relevant features are annotated (*e.g.* handshape, face expression, body posture, etc.) in order to choose the appropriate sign variant when new utterances are created.

5 Indexing and Querying the Heterogeneous Database

It is necessary to build a heterogeneous database consisting of two parts: (i) a raw database containing raw motion data, (ii) a semantic database containing high level linguistic descriptions, the description level depending directly on the annotation scheme. There must be an exact correspondence between both databases; in particular the timelines must be synchronized. The principle of the dual database and the request mechanism can be represented in Figure 6.

The main technical challenge comes here to define the best indexing method for both linguistic structures (textual annotation) and signal (mocap data), while maintaining coherence and consistency between these two levels of information. Finally the motion retrieval process is closely linked to the editing and composing operations.

5.1 Motion Indexing

The raw motion chunk can be represented by a set of formal parameters including the identifier of the motion sequence, the channel, the start and end times. The raw database must be temporally and spatially indexed. The method used by the Sign3D project is to consider the skeleton articulators/joints and the animation frames, as key points for the indexation process. For example, after serializing the motion data in a unique raw file, the memory addresses, at the begin and the end of each motion, for each joints, can be recorded in a separated indexing file. If the animation timestep between two frames is constant, it is possible to compute, in linear time, the memory address (or the index) of a motion chunk, i.e. a motion simultaneously limited to a specific joints list and/or a frame interval from the raw captured data [20]. The main advantage of such a method lies in the fast retrieval of information that contributes to the rapid loading of the motion chunk. This feature is considered as necessary for an LSF editing software, based on a data driven animation system. A drawback lies in the significant memory space of the indexation file and the raw file, depending, especially, on: (i) the avatar complexity (i.e. the number of joints) and, (ii) the chosen strategy for the key points (i.e. the memory addresses) that must be recorded.

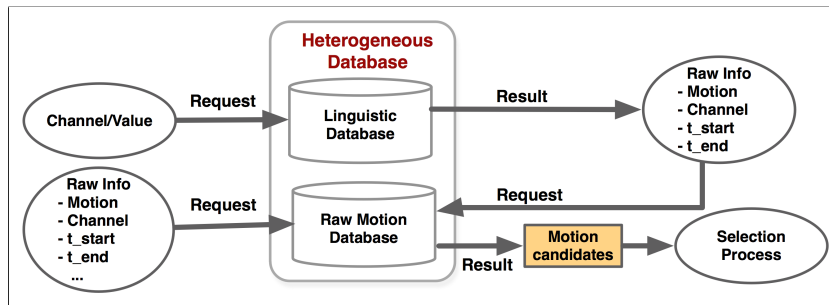


Fig. 6 Heterogeneous database and request principle.

In order to link the raw motion data, and their linguistic descriptions, the semantic part of the database builds a mapping between the annotations and the motion chunks. For that purpose, the semantic part can be made of some advanced hash tables. The hash table values contain the necessary and sufficient parameters to represent the motion chunk (i.e. the corresponding identifier of the motion sequence, the channel, the start and end times) as previously said. The hash table keys defines the indexing strategy and consequently, the underlying query types that will be available for the user. Indeed, if, for example, the keys are the annotation values and/or the channel names, the semantic database can be queried thanks to these two previous attributes. By multiplying this kind of hash table and choosing an appropriate indexing strategy according to several linguistic features, an elementary though efficient query language can be defined, in terms of computer performances and quick information access. The semantic database, built for the Sign3D project, has been indexed by annotation value (*i.e.* glosses), channel name (*i.e.* tiers) and by the couple (annotation value, channel name). This method allows to quickly retrieve any motion chunks annotated according to glosses-based considerations and/or physical/signal-based considerations (cf. section 4).

To complete the semantic mapping, an association must be specified and preloaded, if a channel represents a body part and not a linguistic feature. For example, in the Sign3D project, a text file puts in relation the different channel names to the corresponding joints list of the avatar skeleton. The database will check this association and limit the raw queries results to a sub-skeleton only if the channel belongs to the script.

Nevertheless, a drawback remains in the limited expressivity of such a language. One solution can be the definition of a dedicated query language. The building of a compiler/interpreter is then required, adding a layer and some significant time in the querying process. This last strategy can be perceived as inefficient if one

takes into account the need, for the user, to quickly load and visualize some motion chunk for editing purposes.

With this dual system, a motion chunk can be loaded and extracted by directly examining the raw database with the formal parameters (i.e. name of the motion sequence, channel or set of joints, start and end times). It can also be loaded/extracted by querying the semantic data with a gloss, a channel, or a combination of other linguistic attributes which will produce the formal parameters, corresponding to a segment of movement from the raw database (cf. Figure 6).

5.2 Motion Retrieval

The corpus annotation allows a mapping between the meanings of the signs (and distinctive features) and their realization presented as motion captured data, using the methods described in previous sections. Once the new sentence is composed, the good sign sequence must be selected, which means retrieving in the database the good variant of each sign. Finding and retrieving a desired data from the database requires two successive processes: querying and selecting. Depending on the editing operations, two types of motion retrieval may be considered. The first type, called semantic-based retrieval, consists of applying a first data filtering by querying the semantic database, and then selecting a motion segment among the different candidates. The second type, called signal-based retrieval, consists of finding the best matching subsequence that respects some signal conditions and can be incorporated in the constructed phrase. It is also possible to combine both types of motion retrieval processes.

Semantic-based retrieval. In this case, the retrieving process is achieved by specifying one-condition or multiple-condition queries. Depending of the indexing strategy, it is possible to specify the value of a gloss (the sign *museum* for example), for a whole sign or for a specific channel. The result will be a set of motion

segments representing a specific sign. Finding an acceptable motion that fits the editing operation is somewhat difficult to define and may result in a range of solutions, therefore a selection process should extract the best candidate. Among the possible methods, the following may be used:

- A random choice selection, *e.g.* retrieve a random or a specific motion (the first one for instance);
- An annotation-based selection: it consists of refining the semantic query by a complementary one. For example, the sign may be selected which has been performed to a specified place in the signing space, or with a specific facial expression annotated beforehand. Indeed, these two previous features can change the meaning provided by the rest of the body, according to the parallel aspect of SL. The need to query the database several times and on several channels clearly appears in order to cross the results. For example, if one wishes to retrieve a gesture performed with a given annotated facial expression, it is necessary to do the requests on the two separate channels, and then to select one motion chunk at the time interval corresponding to the responses.

Signal-based retrieval. The aim of the motion retrieval process is in this context to search motion chunks (whole movement or body parts' movement) that best match the request. This process is generally based on a feature-selection process, which consists of automatically computing a set of features, use these features to retrieve the best candidates among possible ones in the mocap database, and then select the best one for further editing. The features can be a set of skeleton postures or joint trajectories, statistical features such as time duration, mean velocity or acceleration, or spatial features characterizing for example targets in the signing space, etc. A similarity measure is also associated to these features, using for instance elastic distances [24]. Two particular cases can be mentioned:

- Transition-based selection: this refers to finding the best transitions that match the passage from one sign (gloss) to another sign (gloss), the whole meaningful signs being directly extracted from the mocap database. With such a process, it should be possible to find the best suitable candidate that can adapt optimally to the previous and successive signs in the constructed utterance [3]. For example, one may select the motion whose first pose (respectively last pose) are closest to the last pose of the previous movement (respectively the first pose of the successive movement);

- Example-based selection: a current approach is to search motion chunks that are very similar to the motion example, the latter being extracted from a database or computed according to an average pattern like in [38,10,35]. The latter methods are generally based on alignment algorithms between sampled signals.

The Sign3D project gives an illustration of a current LSF motion retrieval system. One of the goals is to extract from the database one solution to express the desired information, for example, to sign the following sentence written in natural language: "- Opening hours of the museum: 8:00 a.m. to 5:00 p.m." In SL, the way of signing this sentence is not unique. If one tries to compose the new constructed sentence only from one existing structure (*e.g.* "-The museum is open from 8:00 a.m. to 5:00 p.m."), it is likely that the appropriate motion chunks will not be inside our initial corpus. To bypass this issue, the matching motion chunks in the database from several variants (*e.g.* "-The museum's opening hours are 8:00 a.m. to 5:00 p.m.", or "-The museum opens at 8:00 a.m. and closes at 5:00 p.m.") is helpful. In other words, the goal is to compensate the small variation of the initial corpus by the diversity in the syntactic structures that the system can handle. This approach is being implemented and tested through a study of perceptual evaluation for the generation of sign sequences about time and sign placement. The first results are correct from a syntactical point of view despite the small size of the mocap corpus. Nevertheless, the way of selecting the best sentence among every syntactically correct variants remains an open issue.

6 Animation by Editing and Composing Motion Segments

Creating novel utterances can be achieved by assembling the motion chunks that may be directly extracted from the database (after retrieval and selection processes), through editing operations, *e.g.* splitting, deleting, substituting, inserting, and then composing the resulting movements. The editing operations of a LSF sentence with predefined replaceable parts, may be distinguished from operations applied on components of signs that occur on specific channels (see the annotation process). If the selected motion chunk fits the editing process, then no further processing is needed. On the contrary, it may be necessary to adapt the motion so that it matches the linguistic generation rules while preserving at best the quality of animation, or to synthesize a novel motion that does not exist. Once a list



Fig. 7 Rendering test of the Sign3D project virtual avatar.

of motion chunks is assembled/gathered into a new SL utterances, a succession of skeleton postures is computed and successively displayed by a rendering engine animating the virtual signer.

6.1 Editing a Whole Sentence with Predefined Replaceable Parts

The idea here is to insert into a sentence with some replaceable parts, some original captured motion chunks, while keeping the temporal and spatial properties of the original movements. It is necessary to focus on the adaptation and the optimization of transitions between motion segments. The quality of these transitions can be a deciding factor between similar lists of motion segments from a semantic view point. The easiest way to smoothly achieve the composition is to define interpolation operations, acting both temporally (by concatenating successive motions), and spatially (by blending several body parts' motions). Such an approach has been developed for stating messages regarding incidents of trains (French company SNCF), by substituting sequences of glosses and using rotoscoping or mocap data as animation techniques [42]. The same goal was followed to automatically generate broadcast sentences by changing the prepositions of location or time, or by inter-changing city names [3]. If this approach has already given satisfactory results, it may fail in expressing complex co-articulation effects, when the formation of a novel sentence may spatially affect the structure of the sign components. Moreover, mixing in the same sequence signs from different styles (signers) or prosody may lead to sentences that are not perceptually acceptable.

6.2 Editing and Composing Components of Signs

In order to take into account the different components of signs that convey phonetic or phonological aspects, it

is important to extend the previous system to channel-based editing operations. Among the manipulated components, static components (for example handshape components) will be separated from dynamic ones (for example hand movements). Such preliminary operations have been developed with success in the SignCom project ([9, 20]). For example, in the cooking corpus, different handshapes associated to the sign *GLASSWARE* were mapped to the same hand motion, or motion strokes of some directed verbs were inverted (applied for example to the signs *GIVE* versus *TAKE*, and *LIKE* versus *NOT – LIKE*). Both of these substituted elements were described as segments along a specific channel. However, with such channel-based operations, the temporal synchronization rules that may exist between the channels might not be preserved. A solution would be to incorporate specific controllers that allow to stretch / compress the inserted dynamic movements, using for example Dynamic Time Warping techniques.

6.3 Animation and rendering

The different editing operations, that consist of replaying, adapting, synthesizing, and finally composing, are achieved by specific controllers that are central to the animation system. These controllers perform a temporal interpolation between signs and spatially blend them to produce a succession of skeleton postures which are played in sequence. Using these postures, a 3D rendering engine then animates the resulting virtual signer that ends the workflow. The rendering must be of high fidelity (*i.e.* motion-wise) but not necessarily photorealistic (Figure 7).

In the example of the Sign3D project, one of the most important issue will be to play back accurately each sign and utterance with different 3D rendering models, and to evaluate the comprehensibility of the signs by deaf final users.

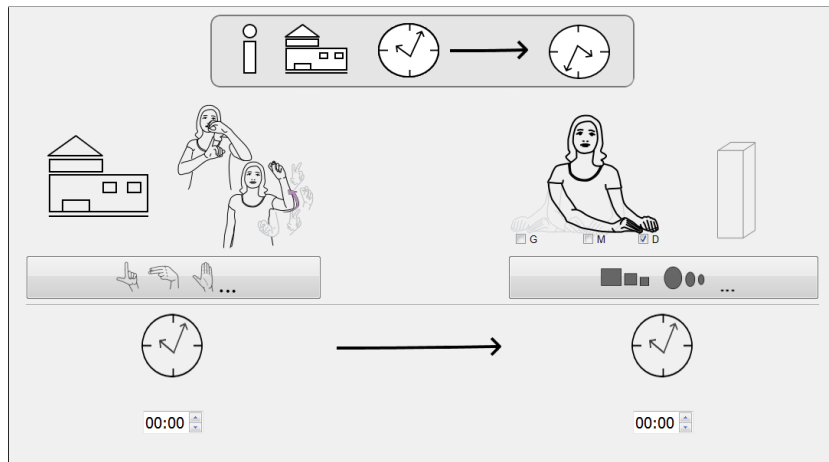


Fig. 8 First version of the composition interface of the Sign3D project.

7 Interface to Construct New Utterances

The SL elements contained in the database have to be combined to generate new utterances that respect sign language organization rules. It is tempting to think of this composition problem only as the substitution of some words in a written sentence, which would lead to a kind of signed French. In order to avoid such a pitfall, most parts of the interfaces will be composed of visual elements (icons, handshapes, sign pictures) only. However, before proposing any interface to make a sentence composition, the following questions must be addressed:

- What set of syntactic structures have to be generated with the interface?
- Does the user know anything about sign language?
- Will the sentences be embedded into longer sequences?

The two first questions are closely related. If the user of the interface is not a signer himself, he/she will have to compose a message using language different than sign language. It is possible to provide a mapping between the interface scripting language and the sign language syntactic structure, however this often leads to composing new sentences by substitution [5]. If the user of the interface has a good knowledge of sign language structure, sentences can directly be built by selecting ordered signs. One must absolutely avoid the pitfall of sentence creation that would only consist of a list of signs directly translated from a scripting language, as signs must be inflected according to the utterance syntax. Three solutions could be considered and freely combined:

- Low level constraints are specified directly through the interface; for instance, select the two signs that have the same placement;
- Pre-defined grammatical structures and the corresponding sets of rules can be selected within the

interface. An example of such an interface has been realized in the *DictaSign* project [15];

- The good sign variants are not specified explicitly, but retrieved from the corpus by using the closest realization. Unfortunately, such an approach would require huge corpora that are out of reach nowadays.

It is also important to insist again on the nature of the generated signs. They are indeed inherently produced by a continuous $3D+t$ system, with many spatial and temporal variational phenomena (co-articulation, classifiers, iconic aspects, placement, etc.). In this context, it is likely that the initial corpus will not contain all signs or sign components necessary to account for this variability. Therefore, it is of prime importance to provide, within the interface, the possibility of building new signs, new facial expressions, or new body postures. Thus, the third issue should then be addressed. If long constructed sequences are considered, then some constraints about the item placement have to be added into the editing process, in order to ensure the co-reference between the same entity used in two sentences: for instance, a museum that will be placed at the right of the signer in one sentence will always be at the same place in the novel constructed sentences.

By taking into account this set of key points, the built interface of the Sign3D project only uses the visual modality (Figure 8). The user can generate information sentences dedicated to public areas while specifying the sign shape and place in the signing space. As soon as the workflow of the Sign3D project is operational from motion capture to synthesis, the system will be submitted to deaf users for evaluation, in order to find the most relevant criteria that enable the composition of novel SL sentences from motion captured data.

8 Discussion and Perspectives

Many questions arise when designing an editing system for generating novel utterances that are played by a virtual signer.

One of the main issues to discuss is the interest of using *pure* synthesis techniques v.s. data-driven ones. Both methods have their pros and cons. Most virtual signers developed with *pure* synthesis show a good level of flexibility, a fine control of the animation thanks to the possibility to parametrize the model, and a precision that is necessary for the linguistic correctness of the produced signs. The main drawback of these methods is the poor realism of the animations, and the lack of expressivity details that are needed for the quality of the movements and the acceptance of the avatar by the Deaf. In addition, the creation of new signs still remains a fastidious and time-consuming task. One alternative response to these problems is to record the movements of deaf signers via *mocap* systems. Human-like movements thus obtained lead to greater acceptability of animations by Deaf. The objectives of using such data-driven approaches are threefold. (i) It is possible to reuse the available data through editing processes, and thus to flexibly cut/paste, adapt, deform, or re-assemble motion chunks in real time; (ii) statistical analysis can be operated on the data, so that to extract invariants in semantic patterns, or to identify kinematic patterns that give some comprehension in SL performances; (iii) beyond the replaying of motion data, it is possible, thanks to learning methods, to generate new signs and sentences and to build good transitions, while preserving the style of the signer.

Though data-driven animation methods significantly improve the quality and credibility of animations, there are nonetheless several challenges to the use of *mocap* data for producing new sentences in sign languages. These challenges are highlighted and discussed below.

First of all, it has been shown that capturing high quality data is necessary to provide signing avatars that are understandable by signing humans. Furthermore, it is necessary to capture simultaneously full-body movements, facial expressions and gaze movements. However, recording *mocap* data is still an expensive and time-consuming process that requires post-processing to check data trajectories, reconstruct the skeleton and adapt the data to the 3D avatar. Consequently, a good recording protocol should be carefully followed.

The second challenge concerns the design of large motion databases in contrast to small ones. There exists a great variability within signs and signing utterances, the same sign being performed differently ac-

ording to the phrase context. Taking into account this variability leads to the necessity to include into the motion database many spatio-temporal variations of the same sign, so that it becomes possible to compose new phrases by reusing existing motion chunks in appropriate contexts. On the contrary, a small corpus implies the development of more complex motion synthesis processes that deal with missing motion segments and deform the movement in accordance to coarticulation effects.

Linked to the previous challenges is the corpus design. A corpus whose purpose is to capture the variability at a high level will allow the extraction of syntactic patterns, and the generation of utterances in accordance with linguistic rules. However, this implies limiting the synthesis to a few sentences whose structure is close to the one already captured, unless a very large corpus with signs performed in many different contexts has been collected. At the opposite, a corpus that captures the variability at a low level should necessitate the development of high level composition processes that respect the rules of SL formation.

As pointed out in the paper, one key issue of the SL editing is the annotation, which determines the quality and accuracy of the motion retrieval process. In particular, one may emphasize the increased accuracy of the reconstructed phrases when dealing with signs / transitions, or phonetical / phonological elements to construct new utterances. The fineness of the annotation is also a guarantee of the high quality synthesis.

Another challenge concerns the indexing and retrieval process. There is always a trade-off between accuracy and efficiency of this process. Two kinds of extraction may be considered: (i) the semantic-based retrieval, directly linked to the annotation. In order to improve the accuracy of the retrieving performance, the decomposition of the signs should be refined, by adding implicitly other spatial channels and temporal segments, and consequently by refining the annotation scheme. In association with this, a more sophisticated request language with multiple conditions requests should be proposed. However, if these additives in the annotation can improve the precision of the sign description, they also limit the efficiency of the retrieval process; (ii) the content-based retrieval process, linked to the raw motion. In order to accelerate the search, the motion should be pre-processed, by reducing its dimensionality (using for example the Principal Component Analysis), or by learning how certain posture sequences are formed. However, manipulation of the raw data must be guided interactively, because it can lead to semantic inconsistencies.

Other important issues have to be considered, as they may affect the future development of SL editing systems. One of these factors concerns *online versus offline* issues: the *offline* specification process of building interactively new SL sentences has to be distinguished from the *online* editing and animating process. The first process involves a user-interface loop, the second depends both on the efficiency and accuracy of the dual database, and on the performance of the animation controllers. The interface must comply with time specifications that are acceptable by the user. This time can be characterized by the duration between the specification of a new sentence and the result of the animation. Finally, the design of ergonomic interfaces should provide new opportunities for data access and communication for the Deaf.

A second factor is about evaluation, which should be integrated at all stages of the editing system (see for example [22, 26] for ASL). The ease and efficiency of the user interface, the efficiency and precision of the indexing and retrieving techniques, and the semantic consistency and quality of the produced animations need to be evaluated.

Acknowledgements The Sign3D project is funded by the French Ministry of Industry (DGCIS: Direction Générale de la Compétitivité de l'Industrie et des Services, Program "Investissements d'avenir").

References

1. Adamo-Villani, N., K. Hayward, J. Lestina, and R. B. Wilbur. 2010. Effective animation of sign language with prosodic elements for annotation of digital educational content. In *Siggraph talks*. ACM.
2. Arikan, O., D. A. Forsyth, and J. F. O'Brien. 2003. Motion synthesis from annotations. *ACM Transactions on Graphics* 22 (3): 402–08.
3. Awad, C., N. Courty, K. Duarte, T. Le Naour, and S. Gibet. 2009. A combined semantic and motion capture database for real-time sign language synthesis. In *Proceedings of the 9th international conference on intelligent virtual agents*. Vol. 5773 of *Lecture notes in artificial intelligence*, 432–38. Berlin, Heidelberg: Springer.
4. Brun, R., S. Gibet, L. Hamon, F. Lefebvre-Albaret, and A. Turki. 2012. The SIGN3D project. [http://http://sign3d.websourd.org/](http://sign3d.websourd.org/).
5. Cox, S., M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt, and S. Abbott. 2002. Tessa, a system to aid communication with deaf people. In *Proceedings of the fifth international acm conference on assistive technologies. Assets '02*, 205–212. New York, NY, USA.
6. Crasborn, O., and H. Sloetjes. 2008. Enhanced elan functionality for sign language corpora. In *Proceedings of lrec 2008, sixth international conference on language resources and evaluation..*
7. Cuxac, C. 2004. Phonétique de la lsf : une formalisation problématique. In *Silexicales - la linguistique de la lsf : Recherches actuelles*, Vol. 4.
8. Duarte, K. 2012. Motion capture and avatars as portals for analyzing the linguistic structure of signed languages. In *Phd thesis, université de bretagne sud*.
9. Duarte, K., and S. Gibet. 2010a. Heterogeneous data sources for signed language analysis and synthesis: The signcom project. In *Proceedings of the seventh conference on international language resources and evaluation (lrec'10)*. Valletta, Malta: European Language Resources Association (ELRA).
10. Duarte, K., and S. Gibet. 2010b. Reading between the signs: How are transitions built in signed languages? In *Theoretical issues in sign language research (tilsr 2010), indianiana, usa*.
11. Efthimiou, E., S. E. Fotinea, T. Hanke, J. Glauert, R. Bowden, A. Braffort, C. Collet, P. Maragos, and F. Lefebvre-Albaret. 2012. Sign language technologies and resources of the dicta-sign project. In *Proceedings of the 5th workshop on the representation and processing of sign languages: Interactions between corpus and lexicon, workshop to the eighth international conference on language resources and evaluation (lrec-2012)*, 37–45. Istanbul, Turkey: European Language Resources Association (ELRA).
12. Elliott, R., J. R. W. Glauert, J. R. Kennaway, and I. Marshall. 2000. The development of language processing support for the visicast project. In *Proceedings of the fourth international acm conference on assistive technologies, assets'00*, 101–108.
13. Elliott, R., J. R. W. Glauert, V. Jennings, and J. Kennaway. 2004. An overview of the sigml notation and sigml signing software system. In *Workshop on the representation and processing of signed languages, 4th int'l conf. on language resources and evaluation*.
14. Elliott, R., J. R. W. Glauert, J. R. Kennaway, I. Marshall, and E. Safar. 2008. Linguistic modelling and language-processing technologies for avatar-based sign language presentation. In *Universal access in the information society*, Vol. 6, 375–391.
15. Filhol, M. 2012. Combining two synchronization methods in a linguistic model to describe sign language. *Human-Computer Interaction and Embodied Communication, GW 2011* 7206: 194–203.
16. Fotinea, S. E., E. Efthimiou, G. Caridakis, and K. Karpouzis. 2008. A knowledge-based sign synthesis architecture. *Univ. Access Inf. Soc.* 6 (4): 405–418.
17. Gibet, S., T. Lebourque, and P. F. Marteau. 2001. High level specification and animation of communicative gestures. *Journal of Visual Languages and Computing* 12: 657–687.
18. Gibet, S., P. F. Marteau, and K. Duarte. 2012. Toward a motor theory of sign language perception. *Human-Computer Interaction and Embodied Communication, GW 2011* 7206: 161–172.
19. Gibet, S., A. Héloir, N. Courty, J. F. Kamp, P. Gorce, N. Rezzoug, F. Multon, and C. Pelachaud. 2006. Virtual agent for deaf signing gestures. In *Amse, journal of the association for the advancement of modelling and simulation techniques in enterprises (special edition handicap)*, 127–136.
20. Gibet, S., N. Courty, K. Duarte, and T. Le Naour. 2011. The signcom system for data-driven animation of interactive virtual signers : Methodology and evaluation. In *Transactions on interactive intelligent systems*, Vol. 1. ACM.
21. Glauert, J. R. W., and R. Elliott. 2011. Extending the sigml notation - a progress report.

22. Hayward, K., N. Adamo-Villani, and J. Lestina. 2010. A computer animation system for creating deaf-accessible math and science curriculum materials. In *Proc. of eurographics 2010, education paper*.
23. Héloir, A., and S. Gibet. 2009. A qualitative and quantitative characterisation of style in sign language gestures. In *Gesture in human-computer interaction and simulation, gw 2007, lecture notes in artificial intelligence, Inai*. Lisboa, Portugal: Springer.
24. Héloir, A., N. Courty, S. Gibet, and F. Multon. 2006. Temporal alignment of communicative gesture sequences. *Computer Animation and Virtual Worlds* 17: 347–357.
25. Huenerfauth, M. 2006. Generating american sign language classifier predicates for english-to-asl machine translation. In *dissertation, university of pennsylvania*.
26. Huenerfauth, M., L. Zhao, E. Gu, and J. Allbeck. 2008. Evaluation of american sign language generation by native asl signers. *ACM Trans. Access. Comput.* 1 (1): 1–27.
27. Huenerfauth, Matt, and Pengfei Lu. 2010. Modeling and synthesizing spatially inflected verbs for american sign language animations. In *Assets*, 99–106.
28. Johnson, R. E., and S. K. Liddell. 2011a. A segmental framework for representing signs phonetically. *Sign Language Studies* 11 (3): 408–463.
29. Johnson, R. E., and S. K. Liddell. 2011b. Toward a phonetic representation of signs: Sequentiality and contrast. *Sign Language Studies* 11 (2): 241–274.
30. Johnston, T. 1998. The lexical database of AUSLAN (Australian Sign Language). In *Proceedings of the first intersign workshop: Lexical databases*. Hamburg.
31. Kennaway, J. R. 2003. Experience with, and requirements for, a gesture description language for synthetic animation. In *Proc. of gesture workshop 2003. LnCS*. Genova, Italy.
32. Kennaway, J. R., J. R. W. Glauert, and I. Zwitterlood. 2007. Providing signed content on the internet by synthesized animation. *ACM Trans. Comput.-Hum. Interact.* 14 (3): 15.
33. Kita, S., I. van Gijn, and H. van der Hulst. 1997. Movement phase in signs and co-speech gestures, and their transcriptions by human coders. In *Proceedings of the gesture workshop on gesture and sign language in human-computer interaction*. Vol. 1371 of *Lecture notes in computer science*, 23–35. London: Springer.
34. Lebourque, T., and S. Gibet. 1999. High level specification and control of communication gestures: the gessyca system. In *Proc. of computer animation*. Genova, Switzerland.
35. Lefebvre-Albaret, F. 2010. Segmentation de la langue des signes française par une approche basée sur la phonologie. In *Phd thesis, université paul sabatier, toulouse*.
36. Lenseigne, B., and P. Dalle. 2006. Using signing space as a representation for sign language processing. In *Gesture in human-computer interaction and simulation, gw, lecture notes in computer science*, eds. S. Gibet, et al., Vol. 3881, 256–260. Springer.
37. Lombardo, V., F. Nunnari, and R. Damiano. 2010. A virtual interpreter for the italian sign language. In *Iva*, 201–207.
38. Losson, O. 2000. Modélisation du geste communicatif et réalisation d'un signeur virtuel de phrases en langue des signes française. In *Phd thesis, université de lille*.
39. Lu, P., and M. Huenerfauth. 2011. Data-driven synthesis of spatially inflected verbs for american sign language animation. *TACCESS* 4 (1): 4.
40. Marshall, I., and E. Safar. 2005. Grammar development for sign language avatar-based synthesis. In *Proceedings uahci*.
41. Prillwitz, S., R. Leven, H. Zienert, T. Hanke, and J. Henning. 1989. *Hamburg notation system for sign languages - an introductory guide*. University of Hamburg Press.
42. Séguat, J. 2010. Modélisation de la coarticulation en langue des signes française pour la diffusion automatique d'information en gare ferroviaire à l'aide d'un signeur virtuel. In *Phd thesis, université paris sud*.
43. Websourd. 2012. Evaluation of the sign wiki. In *Dicta-sign : Sign language recognition, generation and modelling with application in deaf communication, d8.2*.
44. Wolfe, R, P Cook, J C McDonald, and J Schnepp. 2011a. Linguistics as structure in computer animation: Toward a more effective synthesis of brow motion in american sign language. *Nonmanuals in Sign Language Special issue of Sign Language & Linguistics* 14 (1): 179–199.
45. Wolfe, R, J C McDonald, J Schnepp, and J Toro. 2011b. Synthetic and acquired corpora: Meeting at the annotation. In *Workshop on building sign language corpora in north america, washington, dc*.