Heterogeneous Data Sources for Signed Language Analysis and Synthesis: The SignCom Project

Kyle Duarte, Sylvie Gibet

Heterogeneous Data Sources for Signed Language Analysis and Synthesis: The SignCom Project

Kyle Duarte, Sylvie Gibet

Université de Bretagne-Sud, Laboratoire VALORIA, Vannes, France
kyle.duarte@univ-ubs.fr, sylvie.gibet@univ-ubs.fr

Abstract

This paper describes how heterogeneous data sources captured in the SignCom project may be used for the analysis and synthesis of French Sign Language (LSF) utterances. The captured data combine video data and multimodal motion capture (mocap) data, including body and hand movements as well as facial expressions. These data are pre-processed, synchronized, and enriched by text annotations of signed language elicitation sessions. The addition of mocap data to traditional data structures provides additional phonetic data to linguists who desire to better understand the various parts of signs (handshape, movement, orientation, etc.) to very exacting levels, as well as their interactions and relative timings. We show how the phonologies of hand configurations and articulator movements may be studied using signal processing and statistical analysis tools to highlight regularities or temporal schemata between the different modalities. Finally, mocap data allows us to replay signs using a computer animation engine, specifically editing and rearranging movements and configurations in order to create novel utterances.

1. Introduction

As researchers develop expanded studies to analyze the function of signed languages within the larger context of human language, it becomes increasingly apparent that current architectures and technologies provide insufficient resolution for capturing and understanding the movement of the hands, face, and body throughout language production. To respond, the SignCom Project uses both video data and motion capture data as signals for linguistic annotation, for later use in linguistic analysis studies, and for re-synthesis of signed language sequences from a stored set of signs. We believe that this approach is novel in our field, and will provide researchers the data necessary to carry out finer-grained studies on the nature of signed languages.

2. Previous Research

2.1. Signed Language Corpora

Currently, corpora of signed languages exist for Auslan (Australian Sign Language), BSL (British Sign Language), DGS (German Sign Language), NGT (Sign Language of the Netherlands), and SSL (Swedish Sign Language), and researchers have proposed or initiated a number of other signed language corpora as well (Sign Linguistics Corpora Network, 2009).

Though these corpora's linguistic content typically reflects spoken language corpora, their technical devices do not, as data is largely stored in video files, requiring almost all data processing to be done by a human (Johnston, 1998; Crasborn and Zwitserlood, 2008). Among these corpora, differing types and numbers of video cameras have been used (between 2 and 10 standard- and/or high-definition models), offering better or worse data sources for annotators and researchers.

The goals of these corpora have often centered around syntactic analysis, variation studies, and lexicon building, which are important inclusions for most linguistic studies at a corpus level. However, without specialized technological inclusions and narrowed discourse themes, such corpora are not suited for the work of fine-grained signed language phonetic analysis or for signed language generation.

2.2. Motion Capture

Motion capture (mocap) is the process of storing motion as digital information, often by replicating the actions of the bones and joints of the body into a digital skeletal model (Figure 1). Such data can then be manipulated by computer systems to retrieve, analyze, and re-synthesize motion for a variety of research or application scenarios.

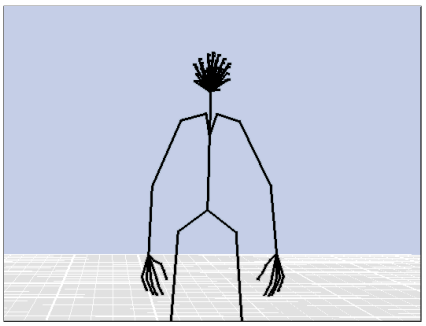

Figure 1: Motion capture technology recreates the skeleton of the subject for quantitative analysis. Breaks in the lines above represent joints, and the lines themselves represent bones. The face is represented with imaginary bones that each control single points on the face, such as parts of eyelids, cheeks, etc.

Currently mocap is used in the fields of movement analysis (sports, music, dance, etc.), biomechanics, character animation in the entertainment industry, and defense sector training simulations (Furniss, 2010). A large motion capture data repository is available at the Graphics Lab at Carnegie Mellon University[1]. Even in such rich

[1]The repository contains 4 GB of data and can be found at http://mocap.cs.cmu.edu.

databases, however, articulators necessary for signed language research, such as the fingers and the face, are not captured.

Finally, to our knowledge, mocap technology has not been actively exploited in the field of signed language linguistics. Such a marriage thus presents exciting future directions for fields of both animation and linguistics.

2.3. Motion Data Retrieval

As motion capture data is somewhat easy to obtain, databases of such data can quickly become unwieldy in the absence of useful indexation methods. Specifically, it becomes necessary to access and retrieve desired data pieces with a minimal computational cost. To date, several methods have been developed to do just this, mostly for the domains of motion synthesis and computer animation.

For our work, we have found it most efficient to combine searchable text data with the raw motion data, dividing motion data into smaller motion chunks (individual signs, as an example) and labeling these chunks with semantic information. Such an approach was first developed for co-speech gestures using manual segmentation (Kendon, 1980), and has been extended to signed language video sequences by Kita et al. (Kita et al., 1998), though signed language annotation on video data had been carried out previous to these studies.

Representing signed language data for the purposes of motion storage and retrieval poses theoretical questions about the depth of annotation, specifically considering the various channels of a multimodal event: hand configuration and orientation, hand placement and motion, and facial expression. Each channel can be segmented temporally, though the channels may not exactly coincide or align, leading to questions about the linguistic and physical interrelatedness of the various channels.

Nonetheless, indexation methods, whether via text annotations of motion data or via representations of what the body parts are doing at a given moment, work to decrease the amount of time necessary to retrieve desired motion chunks. Thus, researchers interested in producing motion with little computational delay focus on improving indexation methods; this indeed is both a previous and present portion of the work of our project (Awad et al., 2009a).

3. SignCom Project Design

The SignCom Project considers the temporal annotation of both video data and corresponding motion capture data for the purposes of language analysis and synthesis. This structure is believed to be unique among the field's project designs to date.

3.1. Video Capture

Video capture of signed language data has been the standard for linguists and archivists for almost one hundred years. Notably, George Veditz recorded a number of sequences in American Sign Language (ASL), including his historical commentary on Deaf education policies, chronicled in his 1913 video on the preservation of "the sign language."

In the 1980s, with the invention of digital video instruments, the field largely shifted to a format of data capture that sacrificed sharpness for the promise of longevity and convenience. In reality, resolution and archival permanence were both diminished with the switch to the VHS and DV video standards, with convenience winning out in the debate. Only recently have high-definition (HD) video cameras been made available in the consumer market, ushering in their adoption in linguistic studies, and finally returning a higher-density picture to captured data.

The notion of frame rate should also be discussed for the purposes of signed language data, as researchers have had to adopt standards that were initially developed to achieve goals contrary to those of the academic community. In order to reduce the bandwidth necessary for moving image transmission across television radio waves, countries use standard frame rates of between 25 and 30 frames per second, depending on historical and technical factors. Though at full speed these images are sufficient to suggest fluid motion, the researcher desires finer-grained data to help distinguish minute changes in hand configuration, as an example. As video technology is developed for the average consumer in mind, it is unlikely that frame rates will increase at marginal expense in the near future; it thus becomes necessary for the signed language research community to consider higher-dimensional data collection methods to better understand the phenomena it wishes to analyze (Piater, 2009).

3.2. Motion Capture

The unique addition of motion capture (mocap) data to signed language elicitation sessions means that the SignCom team has had to develop our own standards for mocap inclusion. We naturally considered pairing mocap recordings with the existing standard for linguistic archives, video data, knowing that parallel recordings would aid in the skeleton reconstruction and data annotation processes.

Our motion capture system uses Vicon MX infrared camera technology to capture the movements of our LSF informants at frame rates that quadruple our existing video data stream. Our setup is as follows:

- 12 motion capture cameras
- 43 facial markers (medium diameter)
- 43 body markers (small diameter)
- 12 hand markers (6 per hand; small diameter)
- 100 Hz capture frequency

The large number of markers used allowed us to capture the movement of known articulators in signed languages, including the arms, hands, torso, and face. By using 12 mocap cameras to detect the position of these sensors, we could minimize the amount of marker occlusions that occurred in the data. For example, when the hands move in front of the face or body, a single straight-on camera would not be able to perceive markers' positions on the opposite side of the hand. By adding additional cameras, angled towards the signer from off-center positions, we were able to reduce marker occlusions to a manageable minimum.

The use of a 100 Hz capture frequency was not random, but verified against current mocap repositories (like that at Carnegie Mellon University [CMU]), and calculated as a compromise between desired data capture and available computing power. Since signed languages involve much more of the body than previous sports or music studies (i.e., facial expressions, body motions, etc.), capturing a signer's motion requires more markers; normally with an increase in marker number, frame rate is necessarily decreased to match the computing power of the motion capture system. However, because technology has advanced since the CMU capture sessions, we were able to retain the 100 Hz frequency and increase the number of markers.

Compared to other mocap studies, 100 Hz is acceptable for this type of work, which falls into the same dexterity and speed category as most sports and dance. Impact events normally require higher frequencies, such as golfing, playing a percussion instrument (appx. 500 Hz), or force feedback gestures (appx. 1,000 Hz). Certainly, it would be possible to record movements at the highest frequency available, but this increases post-processing requirements (filtering, skeleton restructuring, etc.), database size and search, etc. Thus, the 100 Hz frame rate that we are able to capture is by no means a disappointing compromise: compared to the 25 Hz video standard for much of the world, our study stores four times more information per second than existing videotaped studies while maintaining manipulation capabilities by consumer-grade computers.

For the studies that are detailed below, we have generally been satisfied with the amount of mocap data obtained, both in terms of the number of data points and capture frequency. For future sessions, we recommend an increase in the number of facial data points where possible, as signers seem to have an increased command of their facial muscles compared to their actor counterparts (our facial data point model was derived from animated movie mocap standards). Also, using either data gloves or increasing the number of mocap sensors to at least 22 would better replicate the full inventory of the hand's degrees of freedom without applying costly mathematical processing (i.e., *inverse kinematics*) to hand data after recording.

3.3. Data Synchronization

In order to ensure that our many sources of data are synchronized despite differences in frame rates and offsets caused by starting the recording equipment at different instances, we asked our informants to clap before and after each recorded sequence. This action was recorded as both a aural and a visual cue in our video captures, and as a motion cue in our motion captures. After our data elicitation sessions, we align our data sets by matching the timestamps on each data stream for the two claps and adding these offsets to the metadata provided in our annotation files. Of course, this can only be successful if all our data sources are captured at the same time, which they were.

3.4. Segmentation and Annotation

3.4.1. Temporal Segmentation

Both the linguistic and informatics communities take seriously the task of dividing signal data (like videos and mocap of signing sessions) into important sections. Traditionally, linguists have isolated signs from their surrounding transitions to other signs, and computer animators have focused on whole movements, referred to as motion chunks. Of course in both communities, researchers have acknowledged that smaller and larger divisions exist for the movements studied.

To determine the best placement of timestamps, linguistic corpora studies have developed lists of criteria that strive to understand the brain's intentions during the signing event. Normally, sign parts are considered within unified wholes despite temporal asynchronicity among them (for example, the handshape may reach its target before the hand arrives at its destination, as in Figure 2); a similar approach is used for computer animation. On the other hand, we focus on defining signs, or motion chunks, that can later be extracted from the data and used by an animation engine. These chunks are considered separately across the various channels of movement, such that the hand might be timed separately from the arm or the face, even though the brain considers them all part a single meaningful unit.

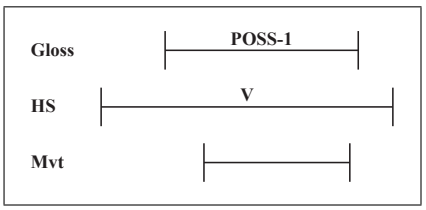

Figure 2: Consider the asynchronous nature of the articulators during a signing event, attested in our data. Linguists have traditionally labeled this sign as a whole unit that functions in a synchronized manner across the sign parts, while we focus more on the individual channels and included in the constructed sign.

For the segmentation of our data, we use the ELAN Linguistic Annotator[2]. The various reasons for choosing ELAN over similar programs are detailed below.

First, many linguists choose to use ELAN for their annotations, especially those who study signed languages. It was our desire to have a file set that was understandable by other researchers should they join our research project in the future. We also felt it important to have a large community of users worldwide to which we could pose questions or with whom we could resolve issues in case any came up during our study. In addition, ELAN requires less manual setup than Anvil, a comparable piece of software largely used in the computer animation community (Kipp, 2010).

Regarding file structure, our goal in searching data was to minimize data retrieval time, and ELAN's file structure presented as a more convincing time-saver than that of Anvil. Specifically, parent-child annotations performed in Anvil are nested in its files' XML data structure, whereas ELAN references parent-child relationships with track IDs and thus keeps its file structure flatter than Anvil's. While

[2]ELAN is a free open-source segmentation and annotation program distributed by the Max Planck Institute for Psycholinguistics (Max Planck Institute for Psycholinguistics, 2010)

chasing the references of an ELAN file can be haphazard at first, the flat file structure allows searches to skip over large portions of data that might be undesired, saving valuable computing effort.
Finally, unlike other linguistic annotators in considerable use, ELAN supports the integration of signal data into the annotation interface, which allows us to annotate based on both traditional video data as well as our mocap data. This serves to give quantitative values to such difficult measures as the distance of the hand from the body when using only one video camera.

3.4.2. Annotation for Linguistic Purposes

Given the theory of symbolic structure that posits than a linguistic symbol is comprised of both a phonological pole and a semantic pole, it is not surprising that annotating language data for linguistic purposes requires a fair amount of both semantic and phonological coding.
The gloss, or more currently the ID-Gloss, is the unique semantic identifier of signs in a corpus (Johnston, 1998). In the NGT corpus, for example, glosses are annotated for both hands, translations are given for the discourse in both Dutch and English, and comments can be left if necessary. The Auslan corpus annotation guidelines specify many more tiers of annotation compared to the NGT corpus, which are well-distributed between semantic and phonological specifications (aspect, referent, movement, orientation, etc.). While we can observe variations in the baseline volume of coding in a corpus, both corpora, and ostensibly others like them, focus on the signs presented during the video discourse and encode information useful for normal linguistic studies.
Corpora designed for linguistic studies may also include discourse-level information, or even codings for prosody or pragmatics. These annotations help researchers study long-duration or cultural influencing factors on the signing event.

3.4.3. Annotation for Computer Animation Purposes

Transitions between signs are largely disregarded in linguistic corpora, perhaps because they are not lexical units and thus cannot be analyzed semantically. Though signed language transitions may not have a semantic pole per the theory of symbolic structure, it is impossible to deny their phonetic values. This is the understanding that the computer animation community has adopted: in order to create convincing animations of meaningful signs, one must also understand what connects signs together in the animation sequence.
In the research of Awad et al., LSF signs were segmented with a transition-inclusive system developed by Kita et al. having previously been tested on signs and co-speech gestures (Awad et al., 2009a; Awad et al., 2009b; Kita et al., 1998). A screenshot of this annotation method is shown in Figure 3. Note that glosses are not necessarily on a one-to-one basis with annotation segments; instead, segmentation has not been forced at sign boundaries but at keyframes that produce desirable motion chunks for later animation.
This approach assumed that any Retraction phase would be highly influenced by the Stroke phase before it, and likewise for Preparation phases. However, this approach ignores the understanding that Preparations and Retractions

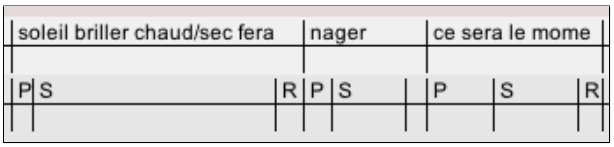

Figure 3: LSF glosses on the upper tier are segmented on the lower tier with (P)reparation, (S)troke, and (R)etraction phases per Kita el al., 1998.

are not clearly defined segments of the signing process, but instead are interactive elements that can overlap during the course of a transition. We believe this notation system works in the context of co-speech gestures because of the more limited form possibilities, especially in the domain of hand configuration. Since signed languages use a larger inventory of hand configurations, incorrectly-formed P/R overlaps in the animation process are more marked and thus less acceptable.

3.4.4. Annotation for SignCom's Aims

The annotation schema for the SignCom Project takes into account the aims of both the linguistic community as well as the computer animation community, since its goals are rooted in both domains. Figure 4 details the hierarchy used in the project's ELAN annotation files.

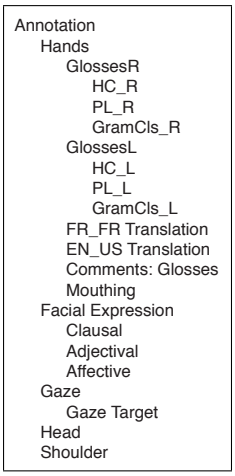

Figure 4: The base tier hierarchy for the SignCom Project.

The structure of our ELAN hierarchy is largely adapted from the Auslan corpus annotation guidelines, and our methods of annotation are inspired by work in progress by Johnson and Liddell on a coding system for signed languages, as in Figure 5 (Johnston and de Beuzeville, 2009; Johnson and Liddell, in progress). However, because of our integration of mocap data into the project, we can also calculate the amount of desynchronization that exists among the various tiers, and the effect such desynchronization has on the ability to animate natural signing.

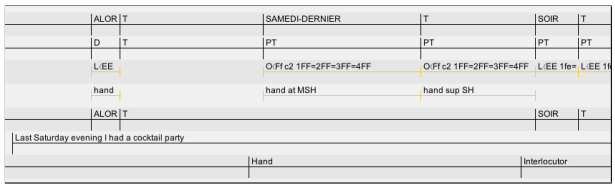

Figure 5: Annotation of an ELAN file in current use in the SignCom Project. The tiers are, from top to bottom, GlossesR, a timing tier, HC_R, PL_R, GlossesL, EN_US Translation, and Gaze Target. The timing tier and the HC_R and PL_R tiers are annotated per the Johnson and Liddell model in progress.

4. Research Application

4.1. Phonological Studies

When considering the simultaneous phonology of signed languages, and the speed with which such phonological bundles are produced, it is immediately evident that current two-dimensional video frame resolutions and their rate of capture (25 to 30 Hz) are too low to accurately capture all the information in a sign stream. Worse, annotating sign production based on such low-resolution data leaves researchers to either leave an annotation entirely blank, or to guess at annotations from either surrounding configurations or personal intuition, thus sacrificing the integrity of corpus annotations.

Motion capture data is our proposed solution to the gap in understanding left by existing video capture technology, providing high-density quantitative data to supplement researchers' already qualitative decisions on annotation and analysis. Motion data can be displayed in ELAN as a time-aligned waveform that gives numerical value to hand position, joint angle, etc. (Figure 6). Motion can be also replayed with a 3D model, allowing the annotator to rotate the figure and see otherwise visually-obscured configurations (Figure 7).

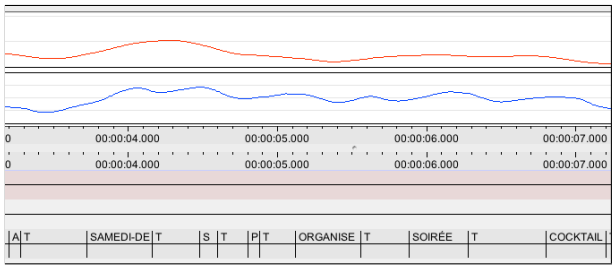

Figure 6: ELAN can be configured to show waveform data that is synchronized with the annotation timeline. The upper waveform represents the hand's X position, and the lower waveform represents its Y position. The annotated tier visible in this figure shows glosses for signs and Ts that represent intersign transitions.

Mocap data can also serve to validate and enforce existing theories on the phonologies of signed languages. Below are studies that are either possible or in progress with mocap-paired language data.

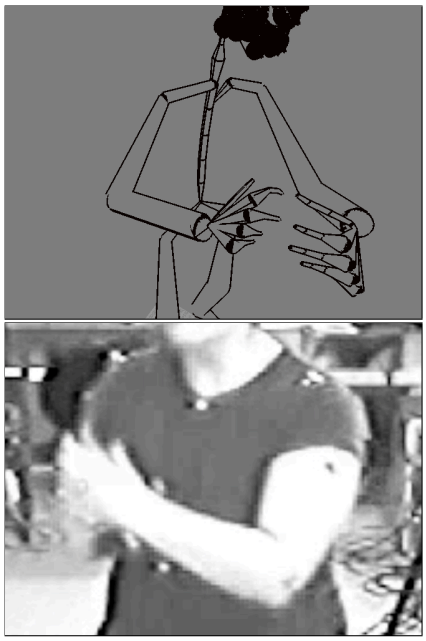

Figure 7: 3D models of the human body (top) can be rotated to see configurations that might otherwise be hidden from view in traditional video camera setups (bottom).

It should be noted that the examples provided below are manually processed; to develop more convincing theories, it will be necessary to process the data statistically, considering exponentially more data points. Also, at the conclusion of the SignCom project, our video, mocap, and annotation data will be made available publicly for other researchers to study.

4.1.1. Hand Configuration Studies

Considering phonological or phonetic notation systems currently in use or development (Johnson and Liddell, in progress; Prillwitz et al., 1989), motion data can serve to validate physical features that a system attempts to encode, or can aid in automatically annotating sign motion with such systems. For the purposes of our project, we have begun work on determining phonological targets for French Sign Language (LSF), specifically in the domains of hand configuration and hand movement.

As is evident in the learning materials published for many signed languages, the concept of having a limited amount of handshapes used in a language is useful for learning signs (Moody et al., 1998). Classically, Stokoe quantified these handshapes in his research on ASL, claiming that there are 19 *dez* ("designator") possibilities that could be modified with any of three diacritics (Stokoe, 2005); in France, too, there has been a history of attempting to quantify handshapes, notably by Cuxac. Learning handshape possibilities is still an integral part of many formal and informal methods of teaching signed languages, just as one might learn the sounds of a new spoken language before learning its words or grammar.

Thus, we will be analyzing how prototypical handshapes that are stored in the brain during language learning can vary from the target form during production. This work

will in the same vein as previous work on phonetic categorization by psycholinguists, but we will be relying on mocap data from production sessions to quantitatively determine phonological category boundaries, instead of relying on respondents' more qualitative analyses of pictures or videos of sign components (Emmorey et al., 2003; Mathur and Best, 2007).

For example, we can compare the citation form of a sign to its production form based on the biomechanical measures that our mocap system provides. Comparing across multiple instances of production of the same target, we can get a statistical sense of how close to target a hand configuration needs to be in order to be considered valid.

Further statistical measures can define when and how quickly a handshape changes during signing events, such as its onset time compared to the rest of the sign, and the duration of the change. Preliminary observations lead us to believe that handshapes are changed more quickly than other sign parts, and arrive at the target configuration before the beginning of its associated sign (consider again Figure 2).

Finally, we suspect that grammatical factors, such as discourse context and grammatical class, will have a role in variations form target configurations. With computer simulations to process our data inputs, such factors should be easily isolatable.

4.1.2. Articulator Movement Studies

Similar studies can be carried out on movement targets for sign production. Figure 8 shows how calculations from mocap data can support existing claims about phonological processes. Shown are two reduplicated signs: one that is reduplicated twice and a second that is reduplicated once. The first instance of the two signs is moved the furthest distance, and is indicated on the graph as the two highest peaks. The reduplications of the signs are shown as the three smaller peaks, a phenomenon that is supported by Liddell and Johnson's phonological principles for reduplicated signs.

Interestingly, despite having been taken from the same signer during separate signing sessions and for different signs, the motions follow the same movement patterns, suggesting mental categorization of movement norms for reduplication. Observed statistically over larger collections of signs, we hope to find additional cases of motion consistency.

Going further, we hope to develop standard motion profiles for signing acts and compare these to existing theories about human motor control on a more general level. Fitts' law, as one example among many, predicts the time required for an articulator to arrive at a target through ballistic motion. Preliminarily, the motion profiles we have extracted during signing do not match these motion theories. Following the forthcoming Johnson and Liddell model, we suspect there will be differences in motion profiles for ballistic, smooth, and protracted movements, and also suspect that intersign movements will vary from intrasign movements. More work will need to occur to develop more accurate signing motion models.

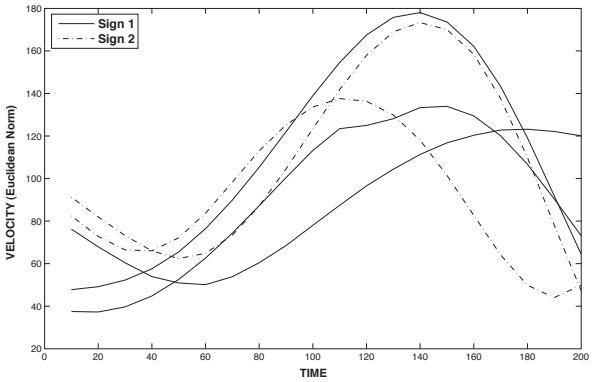

Figure 8: Mocap data can provide insights into movement norms during signing. Here are two reduplicated signs that illustrate phonological principles proposed by Liddell and Johnson whereby movements are reduced in second and subsequent reduplications (Liddell and Johnson, 1989). Sign 1 is reduplicated twice and represented with a solid line; Sign 2 is reduplicated once and represented with a dashed line.

4.1.3. Brain-Articulator Synchronization Studies

Finally, considering once again Figure 2, we hope to make the link between the annotation synchronization choices of the linguistic and computer animation communities, that is to better define the timing relationships between the various articulators, knowing that they appear disjointed but are in fact unified at some cognitive level. This conundrum of multimodal systems has been studied in the French users of Cued Speech, but has not, to our knowledge, been carried out on signed language data (Attina et al., 2006).

4.2. Signed Language Synthesis

Signed language synthesis is the act of generating signed language utterances using a virtual character. This can be done with pre-recorded motion by replaying existing sequences, or by creating new signed language structures or phrases from data containing the building blocks for the desired output.

To achieve this, we first record a generic corpus of signs produced in a specific context, and then annotate the motion data with semantic data, and store both the motion and semantic data in a database for later retrieval (Duarte and Gibet, 2010). For example, stored signs can be retrieved and rearranged to create new syntactic structures that had not been previously recorded, as first demonstrated by Awad et al. (Awad et al., 2009a; Awad et al., 2009b). In those experiments, the authors retrieved whole motion chunks from a signed language corpus on weather signs. These chunks were assembled and transitions were added between them. In our current experiments we are also rearranging signs, but are also considering each of the phonological features of the signing as separate entities that can be isolated on separate channels; this is illustrated in Figure 9.

The appeal of disseminating information through signed language synthesis is the quick, inexpensive manner in which videos can be made in the natural language of the

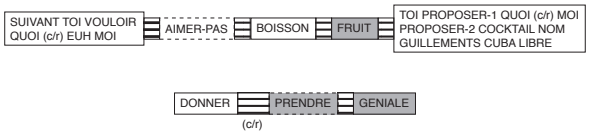

Figure 9: Signs can be rearranged to create novel phrases. Here, signs are retrieved from two different recording takes (white and gray backgrounds) and linked with transitions created by the animation engine (striped background). The sign AIMER ("like") is reverse to create AIMER-PAS ("don't like"), as is DONNER ("give") to create PRENDRE ("take"). Finally, a role shift, shown as *(c/r)*, is included in one transition to ensure discourse accuracy and comprehension.

Deaf community, especially those where written language literacy is low (Holt, 1993). Previous applications of other sign language generation approaches have put virtual signers in train stations to announce train schedules and irregular operations, while others have taken advantage of the anonymizing aspect of animations for politically sensitive topics such as Bible translation (Parkhurst, 2006).

Preliminary trials with our data set have yielded convincing French Sign Language sequences from previously disjointed signs. The majority of our work so far has been to ensure that transitions between concatenated signs contain the same movement dynamics as natural sign sequences. This includes asynchronous timing patterns among the different parts of the sign stream (hand configuration, placement, facial expression, etc.), as well as global orientation issues when signers employ the linguistic notion of role shift.

Perception tests will be added to our work in the near future, in order to judge the success of the rather exploratory field of virtual signers animated from motion capture data. Particularly, we will show a variety of assembling methods and outcomes to judge the credibility of the virtual signer and our animation engine, in terms of intelligibility and comprehension.

5. Conclusion

We have discussed here the generation and analysis goals of the SignCom Project, by gathering various streams of multimedia data to computationally model and analyze signed language data. Specifically, these streams are traditional video data, motion capture data, and annotations of both sources, all of which are synchronized on a single timeline. Aligning various types of data temporally allows us to recall mocap data that would otherwise be devoid of usable meaning, and supports linguistic analysis and signed language synthesis.

6. References

Virginie Attina, Marie-Agnès Cathiard, and Denis Beautemps. 2006. Temporal measures of hand and speech coordination during french cued speech production. In Sylvie Gibet, Nicolas Courty, and Jean-François Kamp, editors, *Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers*, volume 3881 of *Lecture Notes in Computer Science*, pages 13–24, Berlin. Springer.

Charly Awad, Nicolas Courty, Kyle Duarte, Thibaut Le Naour, and Sylvie Gibet. 2009a. A combined semantic and motion capture database for real-time sign language synthesis. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents, IVA'09, Amsterdam*, volume 5773 of *Lecture Notes in Artificial Intelligence*, pages 432–438. Springer.

Charly Awad, Kyle Duarte, and Thibaut Le Naour. 2009b. Gérard: Interacting with users of French Sign Language. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents, IVA'09, Amsterdam*, volume 5773 of *Lecture Notes in Artificial Intelligence*, pages 554–555. Springer.

Onno Crasborn and Inge Zwitserlood. 2008. Annotation of the video data in the Corpus NGT. Technical report, Department of Linguistics and Center for Language Studies, Radboud University, Nijmegen, the Netherlands, November.

Kyle Duarte and Sylvie Gibet. 2010. Corpus design for signing avatars. In *Proceedings of the 4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, Valletta, Malta*, 22-23 May. To appear.

Karen Emmorey, Stephen McCullough, and Diane Brentari. 2003. Categorical perception in American Sign Language. *Language and Cognitive Processes*, 18(1):21–45.

Maureen Furniss. 2010. Motion capture. In *MIT Communications Forum*. http://web.mit.edu/commforum/papers/furniss.html. Accessed 1 February 2010.

Judith A. Holt. 1993. Stanford Achievement Test - 8th edition: Reading comprehension subgroup results. *American Annals of the Deaf*, 138:172–175.

Robert E. Johnson and Scott K. Liddell. in progress. *Sign Language Phonetics: Architecture and Description*. Forthcoming.

Trevor Johnston and Louise de Beuzeville. 2009. Researching the linguistic use of space in Auslan: Guidelines for annotators using the Auslan corpus. Technical report, Department of Linguistics, Macquarie University, Sydney, June.

Trevor Johnston. 1998. The lexical database of AUSLAN (Australian Sign Language). In *Proceedings of the First Intersign Workshop: Lexical Databases*, Hamburg.

Adam Kendon. 1980. Gesticulation and speech: Two aspects of the process of utterance. In Mary R. Key, editor, *The Relation Between Verbal and Non-Verbal Communication*, volume 25 of *Contributions to the Sociology of Language*, pages 207–227. Walter de Gruyter.

Michael Kipp. 2010. Anvil, the video annotation research tool. http://www.anvil-software.de/.

Sotaro Kita, Ingeborg van Gijn, and Harry van der Hulst. 1998. Movement phases in signs and co-speech gestures, and their transcription by human coders. In *Gesture and Sign Language in Human-Computer Interaction:*

International Gesture Workshop, Bielefeld, Germany, September 1997. Proceedings, volume 1371 of *Lecture Notes in Computer Science*, pages 23–36. Springer, Berlin.

Scott K. Liddell and Robert E. Johnson. 1989. American Sign Language: The phonological base. *Sign Language Studies*, 64:195–278.

Gaurav Mathur and Catherine Best. 2007. Three experimental techniques for investigating sign language processing. In *CUNY Conference on Human Sentence Processing*.

Max Planck Institute for Psycholinguistics. 2010. Elan linguistic annotator. http://www.lat-mpi.eu/tools/elan/.

Bill Moody, Agnès Vourc'h, Michel Girod, and Anne-Catherine Dufour. 1998. *La Langue des Signes: Introduction à l'Histoire et à la Grammaire de la Langue des Signes. Entre les Mains des Sourds*, volume 1. International Visual Theatre, Paris, new edition.

Steven Parkhurst. 2006. Spanish Sign Language animation presentation. In *Proceedings of 9th Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil*, December.

Justus Piater. 2009. Video and computer recognition. In *Proceedings of the Sign Linguistics Corpora Network's First Workshop on Data Collection*. Lecture.

Siegmund Prillwitz, Regina Leven, Heiko Zienert, Thomas Hanke, and Jan Henning. 1989. HamNoSys. version 2.0; Hamburg Notation System for Sign Languages. an introductory guide. *International Studies on Sign Language and Communication of the Deaf*, 5:46.

Sign Linguistics Corpora Network. 2009. Sign Language Corpus Wiki. http://www.signlanguagecorpora.org/.

William C. Stokoe. 2005. Sign language structure: an outline of the communication systems of the American deaf. *Journal of Deaf Studies and Deaf Education*, 10(1):3–37. Originally published as Studies in Linguistics, Occasional Papers 8 (1960), by the Department of Anthropology and Linguistics, University of Buffalo, Buffalo, NY.