

Extending the SiGML Notation – a Progress Report

John Glauert
School of Computing Sciences
UEA Norwich
Norwich NR4 7TJ, UK
j.glauert@uea.ac.uk

Ralph Elliott
School of Computing Sciences
UEA Norwich
Norwich NR4 7TJ, UK
re@uea.ac.uk

ABSTRACT

We outline work in progress on extensions to the SiGML notation, which is used to describe Sign Language performance at the phonetic level in a manner suitable for animation by a virtual human signer, or signing avatar. These extensions are intended to provide the author or generator of SiGML sign language content with greater flexibility and precision of control, leading to greater authenticity of computer generated performance, while making the production of such content as convenient as possible. A significant influence on our work is the segmental approach to sign language phonetics of Johnson and Liddell. We illustrate aspects of our approach by means of an example.

Categories and Subject Descriptors

I.3.8 [Computer Graphics Applications]: Miscellaneous;
I.6.2 [Simulation Languages]: Miscellaneous

General Terms

Human Factors

Keywords

Sign language, synthetic animation notation

1. INTRODUCTION

SiGML [2, 3] (Signing Gesture Markup Language) is an XML dialect developed for the purpose of specifying the performance of sign language sequences in a form suitable for their presentation by a computer-generated virtual human (VH) signer, or signing avatar. SiGML is closely based on the model of sign language performance embodied in the HamNoSys notation [8], which was developed primarily for the transcription of (real) human signing. SiGML may therefore be characterised, like HamNoSys, as describing sign language performance at the phonetic level. In its scope and rôle, although not in other respects, SiGML is

broadly comparable to Filhol's Zebedee notation [4] for sign language synthesis.

The purpose of the SiGML notation is to support purely synthetic virtual human sign language performance. It originated as the interface notation between the front- and back-ends (or, alternatively, upper and lower levels) of the prototype system for natural-language-to-virtual-human-signing developed in the ViSiCAST project [3]. Since then it has been used as the input notation for self-contained signed content performance systems. In that connection, it is important to note SiGML's close relationship to HamNoSys: because HamNoSys can be represented as SiGML, existing HamNoSys transcriptions and people who know how to write HamNoSys transcriptions can both act as sources of content for our VH signing systems.

Performance by a virtual human signer of SiGML sequences is provided by the JASigning software package [1]. The main component of this package is its animation generation module, Animgen [6], whose function is to generate frame-by-frame animation data for a specific virtual human character from a pair of inputs: first, a description of the salient physical characteristics of that avatar and, second, the SiGML script describing the SL sequence to be performed. The avatar-specific animation data generated by Animgen — skeleton posture, and morph target (mesh deformation) parameters — is rendered on-screen by JASigning using conventional 3-D animation techniques.

Here we outline work undertaken within the framework of the Dicta-Sign project on the development of an extended and modified version of SiGML in order to provide greater flexibility and precision of definition, leading to improved authenticity and expressiveness of the generated animation.

2. EXTENDING SIGML

SiGML follows HamNoSys by decomposing the description of a sign into a structure that represents — in its manual part — features such as hand shape, hand orientation, hand location, and movement, a category which includes not just changes of location, but also changes of shape and orientation. HamNoSys and SiGML also contain operators allowing the description of signs to be abbreviated in various ways, for example by specifying symmetry between the hands in a two-handed sign, and by specifying various forms of repeated movement within a sign. SiGML follows HamNoSys 4 in providing a nonmanual component, specifying the behaviour of facial articulators, such as mouth, eyes, brows, gaze, and non-manual bodily movements such as nods and shakes of the head, and body tilts.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SLTAT 2011, October 23, 2011, Dundee, UK
Copyright 2011 the Authors.

We are currently working on a variety of extensions to the SiGML notation, allowing increased precision and flexibility of definition, thereby facilitating the generation of more authentic and expressive animation of the virtual human signer.

Perhaps the most basic example of increased precision is the provision of explicit control of the timing and synchronisation both of a sign as a whole, and of individual movements within it. The duration of a sign may be expressed either in absolute terms, or as a relative speed-up with respect to its natural or default time.

Another example of increased precision is in the specification of directions in various contexts, such as the direction in which a finger bone points, the direction faced by the palm of a hand, or the direction of a hand movement. HamNoSys provides a discrete range for these directions, such as “upward”, or “upward and leftward”, as well as a “betweenness” operator, which can be applied to a pair of basic directions, such as those just mentioned, to give the direction half-way between them.

The extended version of SiGML allows any direction between two of the basic directions, by attaching a weighting factor — with values in the range 0 to 1 — to the betweenness operator, thus effectively replacing a discrete range of values by a continuous one. Also allowed is the explicit specification of directions by means of vectors in the 3-D space around the signer. The extended notation also includes the capability to define discrete repertoires of directions, locations and sites in signing space, and sizes of movement, using these more primitive, mathematically orientated features.

A significant influence on the developing notation is the approach to Sign Language phonetics of Johnson and Liddell, most recently and fully described in the second of a series of papers presenting their approach in detail [5]. They argue, by means of detailed frame-by-frame analyses of video recordings, that sign language at the phonetic level has an inherently segmental structure. That is, sign language performance can be decomposed into a contiguous sequence of temporal segments, the two most fundamental categories of segment being *postural*, in which one or more articulatory features are maintained for a definite (albeit possibly very brief) period of time, and what they term *trans-forming* segments, that is, time segments in which one posture is replaced by another.

In earlier unpublished versions of their analysis, Johnson and Liddell distinguished four distinct kinds of segment, based on the settings of a pair of binary-valued (on/off) characteristics, *transient* and *dynamic*:

Posture transient, not dynamic (i.e. static)

Detention not transient, not dynamic

Transition, or Trans-form transient, dynamic

Shift not transient, dynamic

These four are often identified by their initial letters, P, D, T, and S. The distinction between the two dynamic kinds of segment can be expressed, at the risk of over-simplification, by saying that a Trans-form is typically a ballistic movement with no discernible features of its own, while a Shift is a slower more deliberate movement controlled by antagonistic muscles and which may have discernible features of its own, such as the fingers changing direction one after the other.

The distinction between the two static kinds of segment is essentially determined by their duration: both represent the maintenance at a particular location of one or more articulatory features for some period of time, but in a Posture that period will be very brief, that is, effectively “instantaneous” (a single frame of video, or perhaps even less at a high frame-rate), while in a Detention the relevant feature or features are maintained unchanged for a sustained period in a more obviously deliberate fashion.

In their presentation of the segmental model cited above, Johnson and Liddell refine this basic analysis, introducing three further segmental categories in addition to the four just described. Each of these three further segmental categories is essentially an extreme variant of one of the earlier four: an **Extended Detention** ([D:]), a **Rapid Trans-form** (T!) or a **Slow Shift** (S:).

For our purposes the exact number of segmental variants is, although not insignificant, an issue of secondary importance. For us the most important aspects of Johnson and Liddell’s model are the notion of a sign as a sequence of (contiguous) segments, each with its own articulatory features, the binary characteristic, *dynamic*, which gives the most basic classification of those segments as postural or trans-forming, and the binary characteristic, *transient*, giving a further subdivision of segments based on the physical manner of their production. This leads to the PDTTS classification given above. (And it accounts for the fact that an earlier form of our extended SiGML notation was designated PTDS-SiGML.)

In the following sections we illustrate some, but not all, of the features of the extended SiGML notation by means of a specific example.

3. EXAMPLE: THE NGT SIGN “INTERNET” IN HAMNOSYS AND SIGML

As our example we use the NGT (Netherlands Sign Language) sign “INTERNET”¹ Figure 1 shows the HamNoSys transcription for the manual part of this sign, broken for convenience into two parts, describing firstly the initial posture and secondly the movement from that posture.

In digesting the description that follows, the reader may find it helpful to refer to Figure 2, which illustrates the virtual human signer performance of this sign, showing the frames for the initial and final postures, and a single representative intermediate frame at approximately the mid-point of the transition between these two postures.

The entire HamNoSys sign definition in Figure 1 starts with a symbol indicating that the sign is two-handed, with left-right symmetry between the hands except where the sequel contains and explicit specification to the contrary. Thus the second symbol specifies that each hand forms a “pinch” (closed C) handshape. Each section in square brackets with an intermediate (skewed) plus-symbol specifies features of the two hands, dominant followed by nondominant — assumed in this case to be right and left, respectively. The first such section specifies the orientation of each hand, with the pinched fingers of the dominant (right) hand facing outwards and to the left, and the non-dominant hand’s pinched fingers facing the opposite direction. The second square bracket specifies a touching contact between the tips of the

¹The HamNoSys definition of this sign was developed for the eSIGN project by Inge Zwitserlood.



Figure 1: HamNoSys for (the manual part of) the NGT sign “INTERNET”

two hands’ middle fingers. The final two symbols on the first line specify that the hands, considered as a unit configured as just described, are located in front of the signer at the level of the shoulder line.

The second line specifies the hands’ movements from this initial posture: the hands make opposing arced movements — outwards with an arc to the right for the dominant (right) hand, inwards with a leftward arc for the nondominant — accompanied concurrently by a handshape change — with both hands’ fingers and thumbs spread — together with an orientation change leaving the hands still facing each other, with the dominant (right) hand facing towards the signer’s body, and with the index fingers pointing upwards and, respectively, leftwards and rightwards.

The left-hand side of Figure 3 shows the so-called H-SiGML representation of this sign. The manual part is simply a (rather lengthy) representation of the manual HamNoSys just described, each HamNoSys symbol being represented by an empty XML element naming that symbol. This is preceded by a non-manual section, specifying that the manual performance is accompanied by a backward tilt of the head, furrowed eyebrows, and the given SAMPA speech mouthing.

The right-hand side of Figure 3 shows the so-called G-SiGML — that is, “Gestural SiGML” — representation of the sign. This is the form used at present as input to Animgen. It represents more explicitly the structure that is implicit in the linear representation of the manual part in the two previous HamNoSys forms. This form shares the same repertoire of hand configurations, orientations, directions, movements etc. as HamNoSys but it uses textual encodings for them, e.g. “u1” for the direction “upward and leftward” that is used in several HamNoSys movement and orientation symbols.

4. NGT “INTERNET” EXAMPLES IN SEGMENTAL SiGML

Here we outline ways in which the NGT “INTERNET” example might be represented in the new form of SiGML under development, which we refer to, somewhat under the influence of Johnson and Liddell as described earlier, as Segmental SiGML or S-SiGML.

The first representation is shown on the left-hand side of Figure 4. This is based quite explicitly on the original HamNoSys, but exhibits its segmental structure explicitly by breaking the HamNoSys string into fragments and embedding them in a suitable XML structure. The notion of a sign is replaced by the more general notion of a “unit”, preceded in this case by an inter-sign transition. For the benefit of the human reader the sign unit includes the complete manual HamNoSys string as in Figure 1, but it should

be emphasised that this HamNoSys string is irrelevant to the animation generation software, that is, to the semantics of the notation, which is concerned purely with the components of the explicit segmental structure that follows. In this case, there is a sequence of three segments, the initial Posture, a Transition, and the final Posture. By convention in the first Posture of a unit all required features are explicitly mentioned, those that not mentioned are assumed to have default/neutral settings; in every subsequent Posture only changes with respect to the previous Posture need be recorded. The nonmanuals have been left at the level of the sign unit, but they could perhaps be attached to the Transition segment.

In our second S-SiGML representation, shown on the right-hand side of Figure 4, the HamNoSys fragments of the previous version are further decomposed and individual HamNoSys symbols representing sites on the hand, orientations, etc. are represented by tags using the latin alphabet, as in G-SiGML. A minor change is the use of a more compact notation for the extended finger direction than the rather cumbersome notation required in the G-SiGML version shown in Figure 3. Each of the Posture elements now has a richer internal structure describing the hands individually, and their configuration relative to each other.

The “hands” component of the initial Posture, in addition to its location subcomponent, has two new sub-components describing the hands individually. The contact between the hands’ finger tips is represented by specifying a location for the appropriate part of the dominant hand. Because left-right symmetry is specified for the sign unit as a whole there is no need for the nondominant handshape to be specified explicitly. Some of the information in the original HamNoSys movement is now transferred from the S-SiGML Transition to the following Posture. Specifically, the directions and sizes of the movements are now represented as location components of that Posture, described relative to the hand locations of the initial Posture.

It is plausible to claim that it is possible to perform automatic conversions, in either direction, between the original manual HamNoSys string for this sign and either of the S-SiGML versions just described. Indeed our Dicta-Sign project associate Christian Vogler has developed software tool that converts the G-SiGML description of a sign — and hence, via our HamNoSys-to-SiGML conversion software, also the sign’s HamNoSys form — to a segmented SiGML form very similar to that shown in the second part of Figure 4. Vogler’s tool extracts and makes explicit the sequence of postural and transforming segments — with all relevant articulatory features attached to each — for the purposes of analysis rather than synthesis, that is, to support HamNoSys-based sign recognition techniques [7].

Our final S-SiGML representation is shown in Figure 5.

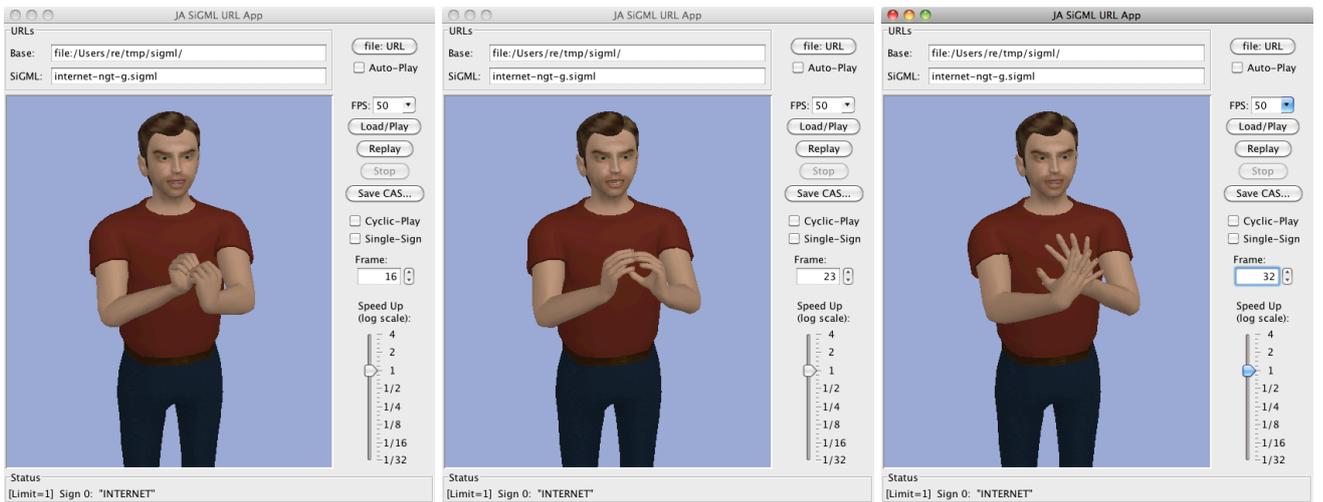


Figure 2: VH Signer performance of NGT sign ‘INTERNET’: initial and final postures, with one intermediate transition frame

This version cannot straightforwardly be derived automatically from the original HamNoSys form. It replaces the Transition and final Posture of the previous version with descriptions that attempt to represent more directly what the observer actually sees, that is, how the observer might describe the signer’s movement — specifically, that the signer’s hands appear to be rotated around a vertical axis passing through the initial point of contact between the hands, this rotation being accompanied by a spreading of the fingers of both hands. The salient characteristics of this rotation are described in the Transition segment, allowing the explicit location information given in the previous version to be removed from the final posture in this one.

5. IMPLEMENTATION ISSUES

We outline briefly some of the issues involved in implementing these extensions by adaptations to Kennaway’s Animgen library [6]. From the standpoint of implementation, the change from discrete to continuous ranges for orientation, direction and location is relatively straightforward, since internally an implementation naturally works with continuous ranges anyway. Indeed for continuous ranges, the definition of descriptive notation that is intuitively appealing to the signed content author is perhaps a bigger challenge than the implementation.

Implementation of an explicitly segmental structure again is not in itself especially problematic. The implementation works internally with such segments, since the HamNoSys model on which the original SiGML was based has an implicitly segmental structure: in HamNoSys, if a sequence of two HamNoSys movements is specified, and if the first has no explicit target, then an implicit postural segment is assumed between (except in those rather rare cases in which a “fused” movement sequence is specified). The basic distinction between transient and non-transient segments is most fundamentally a matter of duration. As we have said, the notation in any case allows timing characteristics to be specified explicitly if desired.

Over and above their timing, the manner in which dynamic segments (Transitions, Shifts) are performed — for

example whether a movement is ballistic or antagonistic — also needs to be considered in order to obtain realistic animation. The challenge here is to characterise the required “envelope” or trajectory mappings — each specifying (numerically) the extent to which a given transition is completed in relation to the proportion of the overall time period allocated for the performance of that transition. Assuming the required repertoire of these mappings is given, either predefined on the basis of empirical corpus analysis, or possibly through a more explicit definitional mechanism, their implementation is straightforwardly accommodated within the existing framework.

6. CONCLUSION

We have outlined the differences between the original SiGML notation for synthetic VH sign language performance, and the extended S-SiGML notation. The most important of these differences are

- Explicit control of timing.
- The substitution of continuous ranges of values for features such as hand orientation, movement direction, and location in signing; space, in place of the discrete ranges inherited from HamNoSys.
- The adoption of an explicitly segmental structure for signs;
- Allowing dynamic features to be ascribed to articulators other than the hands, or to articulator groups other than the hands as a pair (*hand-constellation* in HamNoSys terminology).

S-SiGML effectively generalises the original SiGML: translation from SiGML (or HamNoSys) to S-SiGML is possible. However, in general, translation in the other direction is not possible.

We should acknowledge that we have given rather scant attention here to the nonmanual aspects of SL performance, especially facial nonmanuals. As we have said, SiGML follows HamNoSys 4 in having a range of facial primitives

which be composed in sequence, and concurrently, where that makes physical sense. These features can be attached to an individual S-SiGML segment, where previously they could only be attached to a sign as a whole. Thus the newer notation gives a greater degree of temporal precision for non-manuals than the older one. Johnson and Liddell suggest that, at the phonetic level anyway, nonmanual features fit into the segmental framework. But we consider it an open question whether all nonmanual features of SL performance do so, and we therefore allow explicit timing characteristics to be specified for these features.

Our primary objective in defining this augmented notation is to provide a framework in which the behaviour of a VH signer can be controlled more precisely. This is not unduly difficult, although there are certainly issues about the ways in which this fine-grain control is most conveniently and intuitively expressed. As far as it goes, this fine-grain control is valuable. But it leads towards SL descriptions that are lower-level than is desirable. A continuing and more challenging objective is to identify, through observation and analysis of actual human sign language performance, the appropriate higher-level abstractions that can be defined in terms of these lower-level features without sacrificing realism.

7. ACKNOWLEDGMENTS

We gratefully acknowledge funding under the European Union's 7th Framework Programme, through the Dicta-Sign project (grant 231135). We are grateful also to our colleague Richard Kennaway for his work on the implementation of some of the features described here.

8. REFERENCES

- [1] R. Elliott, J. Bueno, R. Kennaway, and J. Glauert. Towards the integration of synthetic sign animation with avatars into corpus annotation tools. In T. Hanke, editor, *4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, Valletta, Malta, May 2010.
- [2] R. Elliott, J. Glauert, V. Jennings, and J. Kennaway. An overview of the sigml notation and sigmlsigning software system. In O. Streiter and C. Vettori, editors, *Fourth International Conference on Language Resources and Evaluation, LREC 2004*, pages 98–104, Lisbon, Portugal, 2004.
- [3] R. Elliott, J. Glauert, R. Kennaway, I. Marshall, and E. Safar. Linguistic modelling and language-processing technologies for avatar-based sign language presentation. *Universal Access in the Information Society*, 6(4):375–391, 2007.
- [4] M. Filhol. A descriptive model of signs for sign language modelling (abstract). *Sign Language and Linguistics Journal*, 12(1), 2009.
- [5] R. Johnson and S. Liddell. A segmental framework for representing signs phonetically. *Sign Language Studies*, 11(3):pp408–463, 2011.
- [6] J. Kennaway, J. Glauert, and I. Zwitterlood. Providing signed content in the internet by synthesized animation. *ACM Transactions on Computer Human Interaction*, 14, 3(15):1–29, 2007.
- [7] V. Pitsikalis, S. Theodorakis, C. Vogler, and P. Maragos. Advances in phonetics-based sub-unit modeling for transcription alignment and sign language recognition. In *CVPR Workshop on Gesture Recognition, Colorado Springs, USA*. IEEE, June 2011.
- [8] S. Prillwitz, R. Leven, H. Zienert, T. Hanke, and J. Henning. *Hamburg Notation System for Sign Languages—An Introductory Guide*. International Studies on Sign Language and the Communication of the Deaf. IDGS, University of Hamburg, 1989.

```

<?xml version="1.0" encoding="utf-8"?>
<sigml>
<hns_sign gloss="INTERNET">
  <hamnosys_nonmanual>
    <hnm_mouthpicture picture="nEt"/>
    <hnm_head tag="NB"/>
    <hnm_eyebrows tag="FU"/>
  </hamnosys_nonmanual>
  <hamnosys_manual>
    <hamsymmlr/>
    <hampinchall/>
    <hamparbegin/>
    <hamextfingeru/>
    <hampalmd/>
    <hambetween/>
    <hampalmdl/>
    <hampplus/>
    ...
  <hamparend/>
  <hamparbegin/>
    ...
  <hamparend/>
  <hamtouch/>
  <hamshouldertop/>
  <hamclose/>
  <hamparbegin/>
    ...
  <hampplus/>
  <hamparbegin/>
  <hammovei/>
  <hamsmallmod/>
  <hamarcl/>
  <hamreplace/>
  <hamfinger2345/>
  <hamthumboutmod/>
  <hamextfingeru/>
  <hambetween/>
  <hamextfingerur/>
  <hampalml/>
  <hamparend/>
  <hamparend/>
</hamnosys_manual>
</hns_sign>
</sigml>

```

```

<sigml>
  <hamgestural_sign gloss="INTERNET">
    <sign_nonmanual>
      <head_tier>
        <head_movement movement="NB"/>
      </head_tier>
      <facialexpr_tier>
        <eye_brows movement="FU"/>
      </facialexpr_tier>
      <mouthing_tier>
        <mouth_picture picture="nEt"/>
      </mouthing_tier>
    </sign_nonmanual>
    <sign_manual both_hands="true" lr_symm="true">
      <handconfig handshape="pinchall"/>
      <split_handconfig>
        <handconfig extfidir="u" palmor="d" second_palmor="dl"/>
        <handconfig extfidir="u" palmor="u" second_palmor="ur"/>
      </split_handconfig>
      <handconstellation contact="touch">
        <location_hand digits="3" location="tip"/>
        <location_hand digits="3" location="tip"/>
        <location_bodyarm contact="close" location="shouldertop"/>
      </handconstellation>
      <split_motion>
        <par_motion>
          <directedmotion curve="r" direction="o" size="small"/>
          <tgt_motion>
            <changeposture/>
            <handconfig handshape="finger2345" thumbpos="out"/>
            <handconfig extfidir="u" palmor="l" second_extfidir="ul"/>
          </tgt_motion>
        </par_motion>
        <par_motion>
          <directedmotion curve="l" direction="i" size="small"/>
          <tgt_motion>
            <changeposture/>
            <handconfig handshape="finger2345" thumbpos="out"/>
            <handconfig extfidir="u" palmor="l" second_extfidir="ur"/>
          </tgt_motion>
        </par_motion>
      </split_motion>
    </sign_manual>
  </hamgestural_sign>
</sigml>

```

Figure 3: HNS and “Gestural” SiGML forms of NGT sign “INTERNET”

