Experience With and Requirements For a Gesture Description Language For Synthetic Animation

Richard Kennaway

School of Information Systems, University of East Anglia, Norwich, NR4 7TJ, U.K.
jrk@sys.uea.ac.uk

Abstract. We have created software for automatic synthesis of signing animations from the HamNoSys transcription notation. In this process we have encountered certain shortcomings of the notation. We describe these, and consider how to develop a notation more suited to computer animation.

1 Introduction

Some applications for naturalistic 3D animations require them to be generated in real time. These include 3D chat forums, computer-generated deaf signing, and 3D video games. The large number of possible movements makes it impractical to merely play back pre-composed animations.

In the ViSiCAST project we have used descriptions of signing gestures using HamNoSys, an avatar-independent sign language transcription notation. We have developed software, called Animgen, to synthesize animation data from such descriptions, together with a description of the geometry of the particular avatar. While the basic approach has been successful, it has also shown up some limitations of HamNoSys, and suggested principles for the development of a movement description language purpose-made for animation.

Although this work exists in the context of a project relating to deaf sign languages, this paper is primarily concerned with non-linguistic issues. A description of the role of HamNoSys in the linguistic aspects of the project is given in [7].

2 An outline of the Visicast system

We have reported on our animation of HamNoSys in [11], and will summarise that here.

The general outline of the Visicast system is given in Figure 1. If the original matter to be signed is English text, it is translated to a representation known as a DRS (Discourse Representation Structure), and the tools of Hierarchical Phrase Structure Grammars are used to generate from it a sequence of signing gestures.

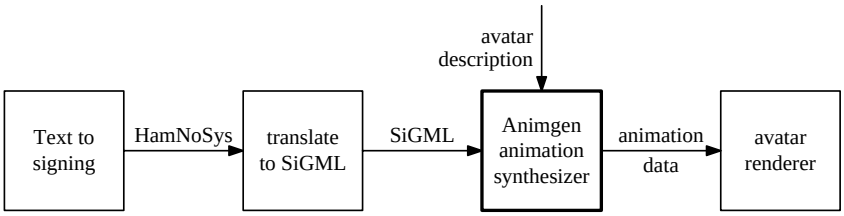


Fig. 1. Block diagram of the generation of synthetic signing animation.

These gestures are described in HamNoSys. This chain of transformations is described in [14], and is represented by the leftmost box of Figure 1.

HamNoSys was developed at the University of Hamburg as a tool for researchers in sign language to make written records of signs [13]. Although not originally designed with computer animation in mind, the ViSiCAST project made use of it as being the only such general-purpose notation available with a substantial body of experience in applying it to multiple sign languages, other notation systems such as Stokoe being specific to particular sign languages. It takes only a minute or two for someone trained in HamNoSys to write a transcription of a sign. This compares very favourably with the time it takes to record signs with motion capture.

As the syntax of HamNoSys is somewhat unwieldy, we designed a version encoded in XML, called SiGML (Signing Gesture Markup Language), and a translator was written from HamNoSys to SiGML. The SiGML representation of a gesture contains exactly the same information as the HamNoSys, but is more amenable to computer processing. As HamNoSys is more widely known, we shall give examples in this paper in terms of HamNoSys.

The next component in the chain is the one we are primarily concerned with in this paper: Animgen, which generates animation from SiGML. After reading the SiGML into its own internal data structures, it endeavours to fill in all the details which the SiGML transcription may have omitted, such as default locations, the duration in seconds of each movement (specified in SiGML merely as fast, slow, or ordinary speed), and so on. For each successive point in time, at intervals of typically 1/25 of a second, Animgen calculates the rotation of each joint of the avatar at that instant. On current desktop machines, Animgen requires about 1 millisecond to calculate each frame of data, i.e. about 2.5% of the available 40 milliseconds for animation at 25 frames per second.

The final component of the chain, the avatar renderer, displays the avatar on the screen in the specified postures at the specified times. For convenience when prototyping, Animgen can also generate output in the form of a VRML file containing the animation data and an avatar in the standard H-anim format.[1]

[1] VRML, the Virtual Reality Modelling Language, is an ISO standard defined at `http://www.web3d.org/fs_specifications.htm`. The H-anim standard for a humanoid structure is defined at `http://www.h-anim.org/`.

3 A brief description of HamNoSys

HamNoSys is designed according to the following basic principles:

1. *Language-independence.*
   HamNoSys is not specific to any particular sign language; it should be able to record any signing gesture from any sign language. This follows from the original motivation for HamNoSys: to provide a written medium for researchers on sign language to record signs. In linguistic terminology, HamNoSys is phonetic, rather than phonemic.
2. *Record posture and movement, not meaning.*
   The meaning of a gesture is not recorded, only the posture and movement. For example, BSL makes frequent use of its fingerspelling signs to represent other words: the sign for "M" also forms part of the signs for "mother" and "microwave". A gesture can thus mean different things in different contexts, but if it is performed in the same way, the HamNoSys transcription will be the same.
3. *Omit irrelevant information.*
   Only those parts of the posture and movement that are significant in creating the sign are recorded. Most signs are made with the hands and the face. What the elbows and shoulders do is not significant; they should do whatever is natural in order to place the hands in the required places. The elbows and shoulders are therefore not notated for most signs. The placement or movement of the elbow is only recorded when it is a significant component of the gesture, for example in the BSL sign for "Scotland" (Figure 2). Version 4 of HamNoSys allows one to specify that a handshape or direction need only be achieved approximately, its exact form being inessential for the sign.
   To some extent, this is in tension with the previous principle, since which aspects of the gesture are important and which are not, which should be performed exactly and which need only be approximate, are determined by what is a constituent of the gesture in that sign language and what is not. We will, however, later argue that the notion of recording the "meaningful" parts of the gesture arises even for gestures in a non-linguistic context.

Like most signing notations, such as Stokoe [15] and the Stokoe-derived notation used in the BSL dictionary [1], HamNoSys describes signs in terms of hand shape, hand position/orientation, and hand movement (leaving aside exceptional signs involving significant use of other body parts).

There are 12 standard hand shapes in HamNoSys (flat, fist, pointing index finger, etc.), and a set of modifications that can be applied to them, changing the bending of individual fingers or the thumb. Position is specified as a set of named locations on the body or at certain distances from it. There is a repertoire of several hundred such locations. Orientation is specified by "extended finger direction" (hereafter abbreviated to e.f.d.) and palm orientation (p.o.). E.f.d. is the direction the fingers would be pointing if they were straight; alternatively, it can be thought of as the direction of the metacarpal of the middle finger (the

Fig. 2. BSL sign for "Scotland"

immovable finger bone within the palm). It has 26 possible values, being the directions from the centre of a cube to its face centres, edge midpoints, and vertexes; additionally, it can be specified as the direction midway between any two of those 26. Palm orientation is one of eight values, corresponding to the directions from the centre of a square to its edge midpoints and vertexes. They are labelled left, up, right, down, and the four intermediate combinations. These designations have the natural meaning when the e.f.d. is forwards; when the e.f.d. points in other directions a more or less conventional assignment is made of palm orientation names to actual palm orientations.

Movement descriptions can be quite complex. A movement of the hand through space can be straight (in any of the 26 directions), curved (the plane of the curve being oriented in 8 different ways about the axis of movement, similarly to palm orientation), circular, or directed to a specific location. Oscillation of the wrists about three different axes can be described, and a movement called "fingerplay", in which the fingers are waggled as if drumming them on a surface or crumbling something between the fingers and thumb. Movements can be combined sequentially or in parallel, and the hands can perform mirrored movements, parallel movements, or independent movements. The "manner" of a movement can be specified as fast, slow, with a sudden stop, or several other styles. The notation[2]  denotes a large, fast rightward movement with a curve convex upward.

Version 4.0 of HamNoSys has been extended to give substantial coverage of facial expressions. SAMPA codes[2] are used to specify speech-like movements of the mouth (frequently used in signing). There is a set of other facial actions such as movements of the eyebrows, or direction of eyegaze. Movements of the shoulders, and tilting of the body or head are also expressible, together with the synchronisation of these with the manual gestures.

[2] Definitions of the HamNoSys symbols used in this paper are given in the appendix.

4 Problems with HamNoSys

4.1 Missing information

Many HamNoSys transcriptions omit information which may be obvious to the human reader, but which are not obvious to a program (because nothing is obvious to a program). In most cases, the missing information can be written explicitly in a more detailed transcription, but some pieces of information are not expressible in HamNoSys at all, and must always be filled in by the reader, human or artificial.

In gestures in which the two hands are placed in some relationship close to one other (what HamNoSys calls a "hand constellation"), there is no way to express the direction from one hand to the other. This must somehow be guessed in every case, but it is difficult to come up with a set of rules which will apply to all cases. In practice, it is usually quite easy to refute any proposed rule by searching through a few dozen randomly chosen entries from the Hamburg corpus of over 3000 HamNoSys transcriptions of DGS signs. In contrast, note that for any particular sign, it is easy for the human reader familiar with HamNoSys to correctly perform it. The problem lies in codifying for computer processing the means by which these judgements are made.

4.2 Extended finger direction

The "extended finger direction" of a hand is always physically present, but at least for Dutch sign language, has been found not to be a phonological constituent ([3]). This is not in itself a problem for HamNoSys, which transcribes on the phonetic level, but it does appear difficult even for those trained in HamNoSys to correctly record e.f.d. In the Hamburg corpus of signs, there are several examples for which the direction that has been notated is not the e.f.d., but the direction in which the hand is pointing, which is often different. An example is the DGS sign for "me" (identical to the BSL sign): the right index finger points to the chest or abdomen of the signer. This has sometimes been transcribed as  . Taken literally (which is the only way a program can take it), this implies the strained posture of Figure 3(a). A correct transcription and performance of the sign is given in Figure 3(b). Contrast the sign for "you", in which the e.f.d. and the direction of pointing coincide (Figure 3(c)).

It is interesting to note that in one introductory textbook on BSL [12], the photographs of signs for "me" and "you/he/she" both show a clear 45 degree bend in the index finger base joint, yet the accompanying line drawings of the handshapes show that joint as being unbent.

4.3 Ambiguity of gestural phonetics

There is a significant disanalogy between speech and gesture in the area of phonetics. For a spoken utterance, the only scope for making differing transcriptions lies in decisions about how to classify the continuously variable elements into the

(a) (b) (c)

Fig. 3. (a) Bad transcription of “me”. (b) Good transcription of ‘me”. (c) Good transcription of “you”.

available phonetic categories, how precise to make these categories, and what aspects to omit as irrelevant to one’s purpose (narrow vs. broad transcription). The different phonetic elements are independent of each other: there is no way to construct, say, a plosive, out of any combination of other elements of the phonetic repertoire. A plosive sound must be notated by one of the symbols for plosives.

This is not at all the case for gesture. There are many geometric elements that can be used to describe a gesture, and every one of them can be constructed out of a combination of a small number of the others. For example, a few geometric elements that one might use are the following:

- direction of the forearm
- e.f.d.
- directions of each finger and thumb bone
- wrist rotation
- bend and splay of the finger and thumb base joints
- bending of the second and third joints of each finger and thumb

Each of these can be defined in terms of others: for example, e.f.d. is determined by wrist rotation and forearm direction; in fact, each of those three can be defined in terms of the other two. Direction of the first bone of a finger is determined by e.f.d. and base joint rotation. And so on. One can arbitrarily pick a basis for gesture space, that is, a selection of geometric elements which are independent of each other, such that every gesture has a unique transcription. However, there does not appear to be any natural choice of such a basis. HamNoSys is one such basis, but as discussed above, some things such as e.f.d. do not appear to be natural choices, at least as regards sign language.

If the aim were simply to give a description of any particular gesture performed on one occasion by a particular person, then any description that accurately reproduced its geometry would do. However, our aim is avatar-independent description, and for that to be possible, the notation must record those aspects of a gesture that would remain the same even if the geometry of the avatar

changed. It is necessary to draw a distinction between, for example, pointing to a specific location on the avatar's body, and pointing in a specific direction. The choice of one or other of those geometric elements when transcribing a gesture is a classification of the intention of the gesturer. These intentions are what must be captured by an avatar-independent notation.

The same issue also arises when considering how to perform the "same" sign in different locations. The BSL sign for "shelf" is illustrated in citation form in Figure 4. "Shelves" can be performed by repeating the sign for "shelf" several times at successive vertical levels. When this is done, it becomes clear that the important direction is that of the straight fingers. When signing "shelf" high up, the hands bend at the finger base joints and the e.f.d. points upwards. The granularity with which HamNoSys can specify e.f.d. is in steps of one quarter of a right angle, and for the finger base joint it is half a right angle. It is thus clear that one cannot accurately produce the desired direction of the fingers at each level by combining values for e.f.d and handshape, even though as continuous variables there would be no problem.

Fig. 4. BSL sign for "shelf"

These considerations apply equally to non-linguistic gesturing. For example, applause is no longer applause if the hands do not contact each other, therefore such contacts must be notated, not a geometry for each arm separately that merely happens to bring the hands in contact for a given avatar.

4.4 Scope

HamNoSys is, by design, limited to the upper body motions required for signing, and in this it has been very successful. Applications for synthetic animation range much more widely, and we wish to have a notation system which could describe the movements required of realistic characters in virtual environments: walking, sitting, standing, manipulating physical objects, conversational gesturing, etc. This is a subject of current research.

5 Iconic and symbolic meaning

It is not the business of a gesture notation to represent the symbolic meaning of a sign, and HamNoSys excludes this. It records only what is physically done. The pointing-index-finger handshape is the same whether it is used to mean "that person there" or "the number 1". There is, however, another sort of meaning to be considered, and which HamNoSys verges upon with its principle of representing only the significant aspects of the gesture. For example, most signs involve only the hands and the face. These are the body parts performing meaningful actions, the rest of the body being only the physical vehicle necessary to support the hands and the head.

This concept can be taken further. The most important aspects of signs are not their concrete geometry, but what we might call their "physical intentions" or "iconic meaning". Pointing with the index finger is the intention of the sign for "that person there". The intention is expressed geometrically in the handshape by the following features: the index finger is extended towards the object pointed at; the other fingers and thumb are curled together into a tight or loose fist. The same intention, when directed towards oneself in the sign for "me" is, as we have discussed earlier, performed with a slightly different geometry, with a bent index finger and an e.f.d. that differs from the direction of pointing. It is this intention which should be notated, rather than the specific geometry that encodes it in a particular context.

It is interesting to compare the latter handshape to one found in one of the BSL signs (there are several) for "caravan"[3] (Figure 5).

Fig. 5. BSL sign for "caravan"

The intention of this sign is a mime of the action of towing something: the right hand is the towing hook of the car (upside down), and the left hand is the

[3] In British English, a caravan is a small house on wheels towed behind a car. Citation forms for this sign vary in different references. The picture is based on that given in [1] (sign 537); [12] shows a version with a much more strongly hooked right index finger.

towing bracket of the caravan. The shape of the right hand in this particular realisation of the sign is almost identical to that of Figure 3(c), but the intention is entirely different: it is not pointing at anything, but hooking into something to pull it. The significant direction associated with the shape is from the inner surface of the index finger towards the wrist. The total range of handshapes that would be correct for the right hand in "caravan" is not identical to the range that would be correct in "me", but there is an overlap.

A gesture notation that recorded these iconic intentions would notate these handshapes differently, although geometrically they might be identical.

6 Principles for a gesture notation

From the above discussion, and our experience thus far with synthesising animations, we arrive at some general principles for a movement notation suitable for this application.

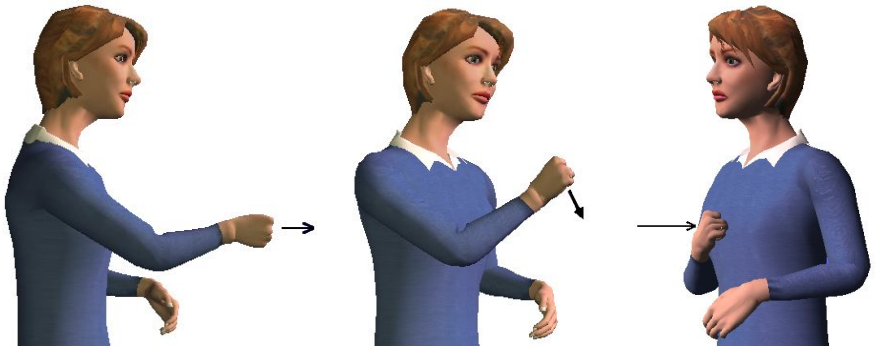
6.1 Intention is primary; geometry is secondary

A transcription of a gesture should start from the intentions that the posture and movement are to express, by which we mean those geometric elements of the posture and movement that must be accurately achieved by the avatar and preserved under change of avatar, as distinct from those elements which merely take whatever values they physiologically must in order to achieve the significant elements.

As a concrete example, with any handshape as used in a given context, one can associate one or more significant axes. For a pointing hand, the significant axis is through the pointing finger or fingers; the orientation of the hand about that axis is usually not significant. For the hooked shape in Figure 5, the significant axis is the direction in which the hand pulls. For a fist, any of three different axes might be significant (see Figure 6). To specify the orientation of the hand, the directions of the significant axes should be given, rather than the direction of some fixed part of the hand. Once the direction of the significant axis or axes have been specified, the rest of the geometry should as far as possible be determined automatically from the geometry and physical constraints of the avatar.

6.2 Express geometry so far as it cannot be deduced from intention

The human ability to fill in what has been left out in an "obvious" way greatly exceeds that of any piece of software. A practical gesture notation for animation must allow geometry to be explicitly specified in more detail than would be necessary for a human reader of the notation. If the software cannot calculate everything that has been omitted in some piece of notation, or produces an answer that the user finds unsatisfactory, then it must be possible to specify more detail to resolve the difficulty. There is a trade-off here between the complexity



(a) (b) (c)

Fig. 6. (a) Axis of punching fist. (b) Axis of hammering action. (c) Axis of tapping chest (BSL sign for "mine").

of implementing the notation and the effort required by a user of the notation to make a transcription.

6.3 Describe the avatar

Avatar description is outside the scope of HamNoSys; in fact, no movement notation of which we are aware makes any reference to the individuality of the avatar. However, software which must produce animation data for a particular avatar must at some point be provided with geometric information about the avatar. The information required includes:

1. The positions of all the joints of the avatar when it is placed in some standard pose.
2. The articulation of each joint: whether it operates as a hinge, a ball and socket, etc., and its limits of movement.
3. The positions of all the feature points nameable in the notation – that is, where on the surface of the avatar are such locations as "the centre of the forehead", "the point of the chin", etc. Each feature point must also be associated with the bone which moves the part of the avatar containing that point.
4. Information about aspects of the avatar's "body language": how fast it generally signs, how sudden are the stops which it makes when moving a hand to a target location, etc.
5. The avatar is assumed to be provided with a set of facial deformations, varying amounts of which can be applied to each frame of the animation. A mapping must be given of each facial element of the notation to the facial deformations (or some combination of them) provided for the avatar.

7 Related work

There are many projects relating to speech synthesis for talking heads, too many to reference here. Their concerns and methods are largely independent of the work described here, which is principally body animation.

VHML (Virtual Human Markup Language) [6] is a programme to devise a notation combining movement synthesis, speech synthesis, emotion, and dialogue, although as yet its more ambitous features have not been implemented. It mainly describes movement at a higher level than we are concerned with.

More directly related are other systems for synthesising animation from movement notations, mostly for deaf signing. These include the research projects GESSYCA ([4]) and SignSynth ([5]), and the commercial systems SigningAvatar (`http://www.vcom3d.com`) and SignTel (`http://www.signtelinc.com`). These are all aimed at particular sign languages (French sign language for GESSYCA, and ASL or various forms of signed English for the others).

Zhisheng Huang and others have implemented a low-level animation system based on the H-anim standard structure for a humanoid [9]. As with HamNoSys, the user can specify postures and movements in broad terms, the program choosing the precise numbers. Animation is currently specified at a very low level, in terms of the rotations of individual joints.

For the dance notation Labanotation ([10]) there is the LINTER software ([8]). This generates movement by linear interpolations, using schematic avatars built from ellipsoids. Its purpose is primarily as a teaching aid to demonstrate to dance students what a given piece of Labanotation means, rather than to produce a naturalistic animation.

8 Acknowledgements

We acknowledge funding from the EU under the Framework V IST Programme (Grant IST-1999-10500). Thomas Hanke, Constanze Schmaling, and Ian Marshall provided many useful comments.

Appendix: Definitions of HamNoSys symbols used in the text

→ Move right. • Large movement. ∩ Bowed upwards. ∗ Fast.

☝ Pointing hand with straight index finger, the other fingers and thumb formed into a fist.

☝̲ As previous, but with the index finger bent at 90° at the base.

☝\☝̲ Midway between the two previous shapes.

⏶ E.f.d. is outwards. ⏷ E.f.d. is inwards.

⏊ E.f.d. is upwards, inwards, and leftwards.

0 Palm faces right. 0 Palm faces left.

References

1. British Deaf Association. *Dictionary of British Sign Language*. Faber and Faber, 1992.
2. SAMPA computer readable phonetic alphabet. `http://www.phon.ucl.ac.uk/home/sampa/home.htm`.
3. Onno Crasborn. *Phonetic implementation of phonological categories in Sign Language of the Netherlands*. PhD thesis, University of Leiden, 2001.
4. S. Gibet and T. Lebourque. High-level specification and animation of communicative gestures. *J. Visual Languages and Computing*, 12:657–687, 2001. On-line at `http://www.idealibrary.com`.
5. Angus B. Grieve-Smith. Signsynth: A sign language synthesis application using Web3D and Perl. In Ipke Wachsmuth and Timo Sowa, editors, *Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop GW2001 (LNAI vol.2298*, pages 134–145. Springer, 2001.
6. VHML Working Group. `http://www.vhml.org/`.
7. Thomas Hanke. Hamnosys in a sign language generation context. In Rolf Schulmeister and Heimo Reinitzer, editors, *Progress in sign language research (International Studies on Sign Language and Communication of the Deaf, vol.40)*, pages 249–264, 2002.
8. Don Henderson. LINTER software for animating Labanotation: see `http://www-staff.mcs.uts.edu.au/~don/pubs/led.html`.
9. Zhisheng Huang, Anton Eliëns, and Cees Visser. Implementation of a scripting language for VRML/X3D-based embodied agents. In *Proc. Web3D 2003 Symposium*, pages 91–100, 2003.
10. Ann Hutchinson Guest. *Labanotation: The System of Analyzing and Recording Movement*. Routledge, 1987.
11. Richard Kennaway. Synthetic animation of deaf signing gestures. In *4th International Workshop on Gesture and Sign Language Based Human-Computer Interaction*, volume 2298 of *Lecture Notes in Artificial Intelligence*, pages 146–157, 1999.
12. Richard Magill and Anne Hodgson. *Start To Sign!* RNID, 2000.
13. S. Prillwitz, R. Leven, H. Zienert, T. Hanke, J. Henning, et al. *HamNoSys Version 2.0: Hamburg Notation System for Sign Languages — An Introductory Guide*. International Studies on Sign Language and the Communication of the Deaf, Volume 5. University of Hamburg, 1989. Version 4.0 is documented on the Web at `http://www.sign-lang.uni-hamburg.de/Projekte/HamNoSys/HNS4.0/HNS4.0de/Inhalt.html`.
14. Eva Safar and Ian Marshall. The architecture of an English-text-to-sign-languages translation system. In G. Angelova *et al*, editor, *Recent Advances in Natural Language Processing*, pages 223–228, 2001.
15. William C. Stokoe, Dorothy Casterline, and Carl Croneberg. *A Dictionary of American Sign Language on Linguistic Principles, rev. ed.* Linstok Press, Silver Spring, Maryland, 1976.