New features in synthesis of sign language addressing non-manual component

Zdeněk Krňoul

University of West Bohemia, Faculty of Applied Sciences, Department of Cybernetics
Univerzitní 8, 306 14 Pilsen, Czech Republic
zdkrnoul@kky.zcu.cz

Abstract

A sign language synthesis system converts previously noted signs into the computer animation. The animation is created using a specially designed 3D model of the human figure and algorithms transferring the sign to movements of the model. In principle the sign language contains both the non-manual component (shape and movement of hands) and the non-manual component (facial movements, etc.). Notation of the non-manual component was not yet sufficiently explored in terms of an automatic conversion to the animation. In the article we describe both notation methodology of the non-manual component and technical aspects for conversion of symbols to movements of the animation model. In addition an appropriate animation method for the 3D shape of face is assumed. The result is an extended notation supplementing notation of the manual component with the non-manual component. The extended notation preserves the feasibility of an automatic conversion and keeps the original level of generality. In connection with the methodology we present the notations of the basic types of non-manual components of the Czech sign language.

1. Introduction

The sign language synthesis system including 3D human figure, complete animation of arms as well as movements of body, head, and facial gestures is a promising usage of computer technology to reduce communication difficulties for deaf people. The sign language synthesis system is a part of complex systems translating the text to sign language as virtual interpreters, signing tutors, sign language dictionaries and others. For linguistic research, a symbol based synthesis system provides the immediate feedback, verification of entered notations, etc. The research on non-manual signals (NMS), the non-manual component of Czech Sign Language (CSL), uses the SignWriting notation system (SW). However for the manual component, the Czech sign speech synthesis system uses the Hamburg Sign Language Notation system (HNS). The notation method in order to transform NMS to 3D animation has not been defined yet.

NMS are parts of the sign language as the speech of a spoken language is not just expressed words and grammar. There are signs distinguishable only by the NMS and *the specific signs* without the manual component. NMS has at least six different roles (Bridges and Metzger, 1996). Symbolic notation can be used primarily for lexical, grammatical markers, conversation regulators, non-manual modifiers. For the mouth pictures, we can directly use letters of the alphabet instead of symbols (Elliott et al., 2004). Movements of other parts of the face, head and chest should be noted individually. There are already sign language synthesis systems reanimating a data record of speaker of the sign language or systems controlled by a symbolic entry (Elliott et al., 2004; Krňoul et al., 2008). Initial interest was directed to the accurate and realistic animation of shapes and movements of the hands. An extension of the synthesis system involves new algorithms for conversion of NMS to 3D animation. The methodology provides universal notation of the non-manual components of sign languages and guarantees automatic processing of it by a computer system.

Section 2 introduces the concept of notation of NMS by HNS and includes the notation of the basic types of non-manual components of the Czech sign language. Technical aspects of conversion to 3D animation are discussed in Section 3 and Section 4 is the conclusions.

2. Notation of Non-manual Signals

Well-known sign language notation systems are Stokoe, SignWriting (SW), HamNoSys (HNS) (Stokoe et al., 1976; Rosenberg, 1995; Schmaling and Hanke, 2001). In terms of non-manual signals (NMS) SW seems to be the most complex notation system. Notation of NMS has to include not only constructions for facial expressions but also movements of upper parts of the body, head and eyes. Minimal observable actions in the face are also in the detail treated by action units (AU) of Facial Action Coding System (FASC) (Ekman et al., 2002). HNS has very a detailed notation of the manual component but the non-manual component is only adumbrate. In contrast, the structure of signs in HNS is suitable for computer processing. We have a synthesis system creating 3D animation of the manual component from HNS (Krňoul et al., 2008) and consider the collection of HNS symbols of the version 4.0 to be sufficient enough for this notation purpose.

The position of non-manual component in structure of HNS is depicted in Figure 1. HNS does not have symbols for complex gestures but the gestures can be notated by a couple of symbols. We consider NMS to be expressed by one or more *non-manual actions*. One non-manual action describes *the rotation and movement of joints*, or *the movement in the face*. A general notation form has in following order: *a base symbol* and *control symbols*. Furthermore the base and control symbols can optionally be supplemented by additional auxiliary symbols (modifiers).

2.1. Transformation of Joints

Non-manual action for transformation of the joints can be used for movements of stomach, chest, shoulders, head and eyes (eye gaze). We consider these base symbols: ⊜ ⊽ ○ ∞. The base symbol ⊽ (shoulder) may be accompanied by the

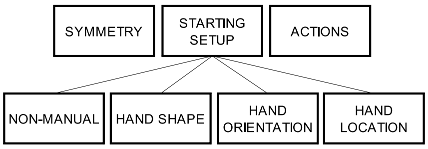

Figure 1: The position of the non-manual component defined by HamNoSys.

symbol ▪ (side modifier) restricting the non-manual action to the left or right side of the body.

The control symbols define both rotation and movement of body part represented by the base symbol. For the rotation, we propose to use symbols of *the finger base direction* originally defined for the manual component. There are 18 symbols defined: ‸ ˥ ›... All combinations of these symbols and the base symbol cannot be interpreted because the joint limits allow only possible poses of human body. We are trying to preserve the original meaning of these symbols. For the base symbols (stomach and chest), the rotation establishes turning in the direction from the body. Hands and head are turning with the chest. In this case, the control symbol determines the neutral pose, is turning on right etc.

The meaning of rotation is not such obvious for the symbol (shoulders). A rotation of shoulders is evaluated for the dominant hand. We consider four main directions: (up/down, forward/backward). The shoulder of the non-dominant hand is determined the left/right symmetry. For the head, the control symbols determine the direction of the nose and eyes are rotated with head (the direction of view). For example, one base and one control symbol is used for head turning to the right: . We can optionally add one more control symbol to notate rotation of joints more precisely. The joints are turned in the same meaning as *the palm orientation*. For example, ; it describes nose direction forward but chin is rotated on right. This optional specifies the rotation especially for chest and head where we expect still a tilt of the joint.

The base symbol may be in combination with symbols for movements as well. The movement will be carried out in the base pose, or in the noted direction. For example, NMS for a head moving from side to side is or only . Rotation of jaw and lower teeth is not considered as movement of the joint but rather as part of a movement in the face. In addition, we propose eye contact with the hands as well. Basic notation is the following short combination of the base symbol and one modifier: .

2.2. Movements in Face

Movements in the face are changes in the shape or the position of a forehead, eyebrows, the area around the eyes, eyelids, nose, cheeks, chin (skin around the chin), and mouth. The base symbols identify parts of the face (locations) that will be changed. The base symbols can be optionally noted in combination with the side modifiers. We consider the following modifiers, ▪ ˆ ˇ, which reduce the base symbol to more detailed parts of the face. The meaning of these modifiers is the same as for locations of the manual component. We can specify the non-manual action for left, or right half of mouth, cheeks, eyebrows and eyelids, and the upper or lower lip, the eyelid, or teeth. If these modifies are not used then non-manual action will be performed for both the left and right half of the face, or both the upper and lower lip, or the eye lid.

For movements in the face, noted control symbol determines elementary movement whereby the shape of the face is deformed. For this purpose, we propose to use symbols for straight movements: ↑ ↗ → .... These symbols determine 18 elementary movements to control the shift of non-manual actions in 3D space. Furthermore the control symbol may be supplemented with the following modifiers: ∘ • * ¯. The size of the non-manual action can be distinguished in three levels: normal, small and large. A modality of the movement may be normal, fast, or slow. For example, "eyebrows go down and near" , or "little inflation of the right cheek" .

2.3. Additional Movements of Mouth

For the mouth, we have three base symbols: . These symbols identify the part of the mouth which will be moved. Furthermore, modifiers allow us to specify more detailed positions. Basically the control symbol may have the same use as the symbols for *the straight movements* for the other parts of the face. We assume meaning of these symbols as direction of a contraction of facial muscles around the mouth as well as a complex articulatory movement.

For the shape of lips, we use the base symbol . The non-manual action describes both mouthing (mouth pictures) and mouth gestures. The shapes of lips for mouthing have already been investigated. The studies confirm the use a combination of three or four key shapes: lip opening, lip protrusion, lip raising, and stretching the lips to the side. A combination of these key shapes allows us to note any form of mouthing. For this purpose, we propose simply to use four symbols for straight movements: lip opening ↓, lip raising ↑, lip protrusion and lip stretch →. The remaining symbols from this group determine other elementary actions applicable for the upper and lower lip or the left and right side of the mouth. There are the directions: up ↑, down ↓, diagonally up ↗, diagonally down ↘, etc. If we do not use the side modifiers then control symbol will have meaning for the right half of the mouth (dominant hand) and the left side will be performed accordance to the left/right symmetry.

For notation of tongue body, we use the base symbol . Again, we can determine its movement in three basic directions: up/down, forward/backward, and left/right. The last base symbol is (teeth). This symbol is recommended to use only for mouth patterns incorporating "uncovered teeth". The rotation of the chin or lower teeth is implicitly included in the non-manual actions describing the lips or tongue positions and does not need to be noted.

The symbols for the straight movements are not sufficient for precise notation of all mouth patterns. Shapes of the mouth often involve contact lips, teeth and tongue with each other. We assume to use two base symbols and one connection symbol (contact). The notation of such non-manual action is intuitive, for example, the upper lip touches the lower teeth. A mutual contact of the same symbol has short notation rather than. For slightly opened teeth or lips, we use the shortcut notation of the base and connection symbol and the connection symbol for squeezed lips.

2.4. Usage of Non-manual Actions

The non-manual actions are placed before the manual component, Figure 1. One non-manual action does not need to be explicitly separated from other symbols. However the notation of two or more non-manual actions has to be always enclosed in parentheses. A composition of several non-manual actions allows us to notate more complex non-manual signals (NMS). We assume the same expression as in the manual component. We propose to use two types of the composition. The first type is used for consecutive non-manual actions in time. For this purpose, we have symbols ( ) (parentheses). The non-manual actions are performed consecutively in the order of their notation. The second type is a composition of non-manual actions expressed simultaneously in time. We consider to use the symbols [ ] (brackets). In this case, all non-manual actions inside produce one fused NMS. Order of non-manual actions is not important.
The combination of these types of composition allows us to note the general NMS. NMS are expressed in parallel to the manual component. Non-manual action begins at the same moment as the movements of the manual component. For example, contact of the lower lip and upper teeth /f/ followed lip protrusion /o/ and simultaneously the head moved slantingly downward and the hand moved in front of the body is noted as: . Notations of the basic types of the non-manual components of the Czech sign language originally expressed in SW are summarized in Table 1.

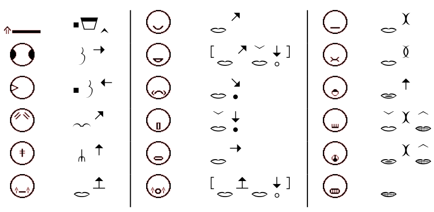

Table 1: Notations of the non-manual component of the Czech sign language, the left column is SignWriting, right column proposed HamNoSys equivalents.

3. Technical Aspects of Non-manual Actions

Technical aspects take into an account problems of the conversion of NMS to the computer animation. Non-manual actions for body joints are expressed by a skeleton structure of the animation model. The principle is same as for the manual component. The animation technique for the face is different. A shape of the face can be created by morph targets, pseudo-muscle actions, control points on the face, or a muscle model (Parke and Waters, 2008). We consider here the shape of the face and tongue as morph targets and the lower teeth as rotation of the rigid body.

3.1. Rules and Rule Actions

The principle of the conversion technique was introduced for the manual component (Krňoul et al., 2008). The schema of the conversion is in Figure 2. This technique automatically carries out the syntactic analysis and creates the parse tree only for the structurally correct entry. Terminal nodes of tree load attributes of particular HNS symbols (descriptors of the symbol). The conversion technique processes the parse tree and reduces its size. Parsing rules join leaf nodes to the parent nodes. Rule actions of the parent nodes integrate attributes from all symbols of the relevant subtree. Next rule actions convert attributes to key frames. The key frames are transformed to the animation frames in accordance with the types and timing of the notated movements.

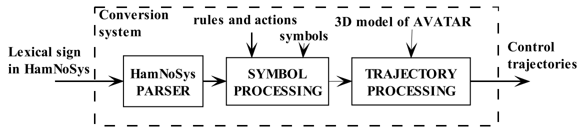

Figure 2: A schema of the conversion system.

The extension of the conversion technique about the new rules and rule actions allows us to accept the input HNS string with non-manual actions. New rules provide the split of parse tree into the manual and non-manual sub-tree, Figure 3. Furthermore rule actions distinguish whether a symbol in the non-manual sub-tree is treated as the base symbol, control symbol, or modifier. The processing of the non-manual sub-tree again takes place in two stages and following order: processing of attributes and processing of animation frames, see Figure 3.

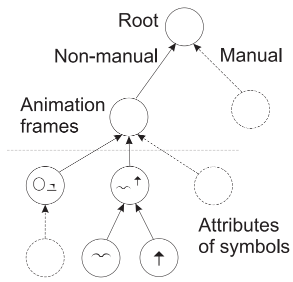

Figure 3: Structure of the parse tree.

The processing of whole non-manual sub-tree precedes the conversion of the manual sub-tree because the animation frames of non-manual actions should be used for the location of the hands. The animation frame consists of two new vectors: *a joint vector* and *a morph vector*. The first one has size $3xN$, where N is the number of joints of the skeleton allocated for the non-manual actions. The second one has a size $1xM$ and stores the morph weights. M is the number of all considered morph targets.

3.2. Rotations of Joints

The transformation of the non-manual action for rotations of joints is completely solved by the rule actions. The rotation of chest, shoulders, the head and eyes must be evaluated individually respecting the geometry of these body parts. For example, the rule action for the chest or head determines rotations individually for each vertebra. The same control symbols in the manual sub-tree are used for the direction of the fingers (a rotation of wrist) and have attribute "Orientation" (Krňoul et al., 2008). Rather than to introduce a new attributes specifically for these non-manual actions, we take an advantage of this attribute for non-manual action. The rule action transfers values of the attribute to proper rotations in the joint vector. To achieve a realistic eye animation, we have to consider both eye gaze and deformations around the eyes. Therefore, the relevant rule action creates nonzero weight of relevant morph target and puts it to the morph vector.

3.3. Morph Targets

The combination of modifiers, base and control symbols defines a list of morph targets. For this purpose, it is advisable to use specialized software (e.g. Poser). The task of rule actions is to determine the type of the morph target and its size. We propose to use one morph target for one non-manual action. Any decomposition of this morph target to the sum of two or more smaller morph targets may be considered for an efficient storage of the animation model and rapid rendering.
The processing of symbols has to identify what morph target is noted. In contrast to rotation of joints, the number of all possible combination of morph targets is very large and a definition of different rule actions loses generality. We propose to extend the description of the symbol by one new attribute "MorphName". The value of this attribute is expected in the definition of control symbol, base symbol, modifiers and connection symbols. Only one rule action processes this attribute to the final name of a morph target, for example: "Right_Cheek_RightMove". In addition, the rule action converts the final name of a morph target to the index, determines the size of processed non-manual action and adds all to the morph vector.

3.4. Processing of Animation Frames

All movements in the face are static gestures that are represented by only one key frame. Two and more key frames in the parallel are processed as the sum of vectors. If non-manual action describes the movement of joint then the rule action will create more key frames (such as head movement from side to side). Finally we assume an interpolation technique to get the animation frames in between the key frames. An illustration of two NMS is in Figure 4.

4. Conclusions

The article addressed the issues of notation non-manual signals (NMS) of the sign language and automatic conversion of NMS to 3D animation. For the notation purpose, we consider the Hamburg Sign Language Notation system. First, non-manual actions are determined by combination of elementary rotations of joints of the upper body and movements in the face. We assume the same symbols and meaning used for the notation of the manual component. For rotations of joints, location symbols are combined with the direction symbols. Movements in the face are described by

Figure 4: The illustration of NMS consisting of following non-manual actions: ([○⌞∞↓ }→][∞̂ ↑ ◡̌ ↓]) (in Poser 8).

the symbols for the location in combination with the symbols for straight movements. The entry of non-manual actions has a general scope and we are not restricted to predefined NMS.
Furthermore, the conversion of NMS to computer animation is discussed. First, it summarizes the conversion algorithm originally designed for the manual component. An extension of the algorithm is described to allow processing of both manual and non-manual components. The conversion of the notation of an eye contact is not yet proposed. We expect, this will be solved in the future in relation with the more general issue of synchronization of the sign speech components.

5. Acknowledgments

This research was supported by the Grant Agency of the Czech Republic, project No. GAČR 102/09/P609.

6. References

Byron Bridges and Melanie Metzger. 1996. *Deaf Tend Your: Non-Manual Signals in ASL*. Silver Spring, MD: Calliope Press.

Paul Ekman, Wallace V. Friesen, and Joseph C. Hager, 2002. *The facial action coding system*. Salt Lake City: Research Nexus eBook.London: Weidenfeld&Nicolson.

R. Elliott, J. R. W. Glauert, and J. R. Kennaway. 2004. A framework for non-manual gestures in a synthetic signing system. In *CWUAAT*, pages 127–136.

Zdeněk Krňoul, Jakub Kanis, Miloš Železný, and Luděk Müller. 2008. Czech text-to-sign speech synthesizer. *Machine Learning for Multimodal Interaction, Series Lecture Notes in Computer Science*, 4892:180–191.

Frederic I. Parke and Keith Waters. 2008. *Computer facial animation*. A K Peters, Ltd. Wellesley, MA 02482, 2 edition.

Amy Rosenberg. 1995. Writing signed languages in support of adopting an asl writing system. Master's thesis, University of Virginia.

Constanze Schmaling and Thomas Hanke, 2001. *HamNoSys, 4.0*. University of Hamburg, in: t. hanke (ed.), interface definitions. visicast deliverable d5-1. edition.

William C. Stokoe, Carl Croneberg, and Dorothy Casterline. 1976. *A Dictionary of American Sign Language on Linguistic Principles*. Silver Spring, 2 edition.