Towards a Transcription System of Sign Language for 3D Virtual Agents

Wanessa Machado do Amaral, José Mario De Martino
Department of Computer Engineering and Industrial Automation,
FECE, University of Campinas,
13083-970 Campinas, SP, Brazil

Abstract - Accessibility is a growing concern in computer science. Since virtual information is mostly presented visually, it may seem that access for deaf people is not an issue. However, for prelingually deaf individuals, those who were deaf since before acquiring and formally learning a language, written information is often of limited accessibility than if presented in signing. Further, for this community, signing is their language of choice, and reading text in a spoken language is akin to using a foreign language. Sign language uses gestures and facial expressions and is widely used by deaf communities. To enabling efficient production of signed content on virtual environment, it is necessary to make written records of signs. Transcription systems have been developed to describe sign languages in written form, but these systems have limitations. Since they were not originally designed with computer animation in mind, in general, the recognition and reproduction of signs in these systems is an easy task only to those who deeply know the system. The aim of this work is to develop a transcription system to provide signed content in virtual environment. To animate a virtual avatar, a transcription system requires explicit enough information, such as movement speed, signs concatenation, sequence of each hold-and-movement and facial expressions, trying to articulate close to reality. Although many important studies in sign languages have been published, the transcription problem remains a challenge. Thus, a notation to describe, store and play signed content in virtual environments offers a multidisciplinary study and research tool, which may help linguistic studies to understand the sign languages structure and grammar.

KEYWORDS: computer graphics, sign language, XML, accessibility, virtual reality.

I. INTRODUCTION

There are estimated to be 5.7 million people with hearing disabilities in Brazil [3], 20 million in U.S. [5] and 278 million people worldwide [16]. Kennaway [6] shows that the reading performance of deaf children is poor compared to that of their hearing peers. Thus, situations in which information is presented primarily in written form place them at a substantial disadvantage.

The access that deaf people have to virtual content could be greatly improved by the provision of sign language information. Sign language is being displayed on computer environment using video content. However, there are disadvantages to this means of providing information, since it is necessary to use specific equipments and trained people who deeply know the sign language. The maintenance of video content is another problem. There are continuity issues, like use the same signer, in the same clothing and with the same background. Thus, create pieces of signing that can be joined together to make signed phrases is nontrivial. Each time any content detail changes, new videos must be made, increasing the costs. Storing and downloading videos can also be problematic as they are large files. For displayed content on Web sites using Internet connection, the time and cost involved in downloading video sequences may be prohibitive.

A virtual avatar driven by animation software provides an attractive alternative to video. Virtual signing has some advantages. Signed content can be created by one person on a desktop computer. No video capture equipment is required. A virtual avatar can generate real time content, so continuity is not a problem since details of the content can be edited at any time, without having to rerecord whole sequences. Storing the content is further advantage. Disk space demands to store sign description are negligible. Data transmission is improved, since transcription content can be stored in text files, which are smaller than video files and can be downloaded faster. Another advantage is the extra control by user, which is not possible with video. The view angle can be continuously adjusted during playback.

To animate a virtual avatar, the first need is to develop an accurate transcription system, to explicitly specify how sign is articulated. The challenge is to ensure the realistic animation of virtual agents.

II. RELATED WORK

A. Stokoe

Stokoe, the founder of the sign language linguistics, proposed the first notation system for a sign language [14]. The original Stokoe notation is consisted of 55 symbols, divided in three groups, called chemeres, representing aspects of sign:

- Tabula: hand location;
- Designator: hand shape;
- Signation: movement.

Hand shape symbols are based on Latin letters and numerals. Location and movement symbols are iconic. Stokoe shows how these three parts, meaningless alone, fit together to form a linguistic structure, identical to the phonemes of spoken language [11].

T. Sobh, K. Elleithy (eds.), Innovations in Computing Sciences and Software Engineering,
Despite the simultaneity of aspects of sign, Stokoe notation sequentially writes the symbols that represent each aspect. This sequential structure can be read by humans who know the notation and the sign language. However, it is nontrivial to understand by someone not familiar with Stokoe notation, and even more by a computer program.

On the other hand, there is sequentiality in the internal structure of signs. For example, the sign "deaf" in ASL is articulated with two face touches with the index finger, the first one in the region below the ear and the second in the region next to the mouth, with an arc trajectory between them. Change the order of these movements results in a meaningless articulation in ASL.

It is important to note that simultaneity and sequentiality are unresolved issues in existing notations, and they are treated in the transcription system proposed here. Typically, traditional notations sequentially write the sign aspects, which are simultaneous during articulation. In the other side, there is a sequence in holds and movements of a single sign, and this order must be respected and explicitly stored.

There are defined hand shapes in Stokoe notation. To write a sign with a new hand shape, it is necessary to look at the whole hand shape set and matches the new shape with an existing one that resembles most. This can also be problematic, since sign language is not static and new signs may arise requiring new hand shapes.

Stokoe notation does not provide non-manual aspects, such as facial expressions. However, signs showing emotions such as joy or anger need to be accompanied by the appropriate facial expression. Liddell [9] shows that to reproduce a story with characters that has no name, the signers can use facial expressions to refer different characters. Thus, facial expressions play a very important role, and are necessary to effectively convey meaning.

B. SignWriting

SignWriting [13] was created in 1974 and is defined by a combination of iconic symbols to represent hand shapes, body locations, facial expressions, contacts and movement. Hand shapes can have variations of three basic form of the fist: closed, open or flat. The symbols in SignWriting are all shown from the expressive, not the receptive, viewpoint. That is, signs are written with the signer perspective, looking at their own hands. There are also symbols to represent palm, back and hand side. The articulation space is represented in SignWriting using arrows.

C. HamNoSys

A later system, HamNoSys [12], was first made publicly available in 1989 and consists of about 200 symbols covering the parameters of hand shape, location and movement, like Stokoe Notation. HamNoSys is still being improved and extended, and it is possible to note down facial expressions, but their development is not finished yet.

The eSign [7] project uses an XML-based scripting notation, SiGML, based in HamNoSys. The project presents a web browser plug-in to animate signs using an avatar-independent scripting notation. Nevertheless, SiGML has some limitations. There are some omitted information, such as default locations of sign articulation and the duration of each movement, specified on SiGML merely as fast, slow, or ordinary speed [6]. In real situations, speed may vary during the signing, for example, to assign intonation. In this case, a sign can start running at normal speed and have fast speed at the end of articulation. This issue will be treated on this work.

III. Transcription

A. XML and diagrams

Digital information provided in sign language is not common. Sign language can be displayed using video. However, this is not an appropriate alternative for computational environment. Kennaway [6] shows that the traditional transcription systems were not created to virtual animation purposes.

This work addresses a transcription system for sign languages for virtual 3D agents implementation. The main challenge is to develop a model that represents the signs without having to store all existing gestures combinations, which would certainly result in a combinatorial explosion and make the work impossible.

The transcription system proposed in this paper considers the following main limitations of previous approaches: movement speed, sign concatenation, sequentiality and non-manual expressions. Moreover, the information was grouped hierarchically using XML (eXtensible Markup Language) files, which have the advantages below:

- Are text files, editable in any simple text editor.
- Are hierarchical files, validated and consolidated.
- Low cost. Existence of free editors and validators.
- Are files easy to share and store.

To facilitate comprehension, the XML description is illustrated by UML diagrams. Each XML element is represented in diagram as a class. The cardinality between classes means how many times the element may appear in the document. Each attribute is illustrated as a class attribute. The possible values that attributes can assume will be written immediately in front of the attribute name and may refer to a set of values, for example, 0,...,10, and preceded by the equal sign (=), or a value type, string or integer, preceded by a colon sign (:).

B. Notation

As mentioned above, Stokoe identified three aspects: hand shape, location and movement. Later, Battison [1] and Friedman [4] identified the palm orientation, which already existed in Stokoe original notation system, but with secondary importance. Klima and Bellugi [8] identified the hands arrangement, which hand holds the sign and if in active or passive form. Liddell and Johnson [11] grouped the movements in local and global. The work of Battison [2] has two restrictions that significantly limit the number of possible