Sign Language 3D Virtual Agent
Wanessa Machado do Amaral, José Mario De Martino, Leandro Martin Guertzenstein Angare
wmamaral@gmail.com, martino@fee.unicamp.br, leangare@hotmail.com
Department of Computer Engineering and Industrial Automation,
FEEC, University of Campinas,
13083-852 Campinas, SP, Brazil

Abstract-Accessibility is a growing concern in several research areas. In computer science, programs that provide accessibility are those that enable people with disabilities to use computing resources efficiently. Since virtual environment relies primarily on the visual mode, usually using writing content, it may seem that access for deaf people is not an issue. However, for prelingually deaf individuals, those who were deaf since before learn any 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. The aim of this work is to develop a 3D virtual agent for presenting sign language information. The avatar consists of an articulated model, representing a human character, which is able to articulate signs. The virtual agent control is done through text scripts that describe the signs in particular notation developed in the context of this work. A sign language 3D virtual agent provides important improvement in human–computer interaction for deaf people, and may also be a tool for teaching and research of sign languages.

KEYWORDS: sign language, educational technology, accessibility, computer graphics, virtual reality.

I. INTRODUCTION

Between years 2000 and 2004, estimates of the number of people in the United States with a self-described “hearing difficulty” ranged from 28.6 million to 31.5 million [5]. In Brazil, there are estimated to be 5.7 million people with hearing disabilities [6]. Kennaway [8] 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. The goal of this work is to improve computational accessibility adapting software responses to the deaf needs, providing mechanisms to facilitate the generation, transmission, storage and maintenance of sign language content.

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 with deep knowledge of 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, creating pieces of signing that can be joined together to make signed phrases is nontrivial. Each time any content detail changes, new videos must be captured and edited, increasing the costs. Storing and downloading videos can also be problematic as they are usually large files. For displaying content from 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 a 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.

This work will proceed as follows. Section two discuss the need of a transcription system to provide signed content on virtual environment. Section three presents a notation developed in the context of this work to describe sign languages for computer animation. Section four presents the signing virtual agent. Finally, in section five, we give some conclusions and future works.

II. SIGN LANGUAGE TRANSCRIPTION

To enable efficient production of signed content on virtual environment, it is necessary to make written records of signs.

Several systems have been developed for representing sign languages in written form. Stokoe, the founder of the sign language linguistics, proposed the first notation system for a sign language [13]. 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; and
- Signation: movement.

Other studies [9] showed that facial expressions are also important aspects. They are used to enhance meaning of signs and emphasize the context of articulation. SignWriting [14] was the first notation that incorporated mechanisms to describe facial expressions. Later, the HamNoSys [12] notation was created by a group of hearing and deaf people as a scientific research tool. HamNoSys should be applicable to every sign language in the world. It consists of several iconic symbols covering the parameters of handshape, hand configuration, location, movement, and also non-manual expressions description is under develop, like facial expressions. Refer to [11] for an overview of some sign language notation systems proposed in the literature.

Unfortunately, the transcription systems cited here were not created to generate animations and have limitations [7]. These systems omit information which may be obvious to the human reader, but which are not obvious to a computer program. In most cases, the missing information needs to be written explicitly to complete and detail the transcriptions, like markers in glossing. And many important information must always be filled in by the human reader.

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. In other words, there should be no ambiguity nor omission. 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.

Part of this work is to develop a transcription system to provide signed content in virtual environment. The aim is to achieve an accurate transcription system, to explicitly specify how signs is articulated. The challenge is to ensure the realistic animation of virtual agents.

III. PROPOSED TRANSCRIPTION

The transcription system presented here was first proposed in [1] and [2]. Later, the system was adapted and improved to fit the virtual avatar needs. The transaction system 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 are text files, editable in any simple text editor, and easy to share and store.

Stokoe [13] chimeres was created to represent the aspects of sign: hand location, hand shape and movement. Stokoe shows how these three parts, meaningless alone, fit together to form a linguistic structure, identical to the phonemes of spoken language. So, the aspects of a sign must be articulated together to make sense. Despite the simultaneity of aspects of sign, Stokoe and other notations 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 the notation, and even more by a computer program. The proposed notation combines aspects of signs hierarchically, so that it is possible for the avatar interprets and reproduces these aspects simultaneously.

The root element is sign with name attribute which store the name of the sign described. Liddell [9] studied the structure of sign languages, and divided signs into two segments:
hold, when hands are stopped in space and movement, when hands are moving. Later, Liddell and Johnson [10] divided the movements in local and global.

To facilitate comprehension, the XML description is illustrated by UML diagrams. Figure 1 illustrates part of XML file.

Figure 1: Description of a sign

Based on Liddell studies, the notation presented here first grouped information in three parts: hold, localMovement and globalMovement. A sign can has one or more holds and zero or more movements. There is a sequence in holds and movements of a single sign, and this order must be respected and explicitly stored. 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. In the notation proposed, the order of holds and movements are the same as they appear in XML files. So, the sequence is explicitly stored, with no ambiguity.

The hold element has three children: rightHand, leftHand and face. They are illustrated in Figure 2.

Battison [3] and Friedman [5] identified the palm orientation, which already existed in Stokoe original notation system, but with secondary importance. The palm orientation is described here as an attribute of hand element.

Each hand has four attributes: configuration, orientation, wrist and location. The work of Battison [4] has an important restriction that significantly limit the number of possible combinations of signs articulated with two hands: the symmetry. Symmetry condition states that when both hands move independently, they will have the same hand shape, place, orientation and movement. In the proposed notation the leftHand element has additional symmetry attribute, which is set to yes when the sign has the symmetry condition.

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. Facial expressions have ten attributes, as follows:

- forehead: furrowed.
- eyebrow: up, straight, down, up inside, down inside.
- eyes: open, squeezed, closed, half open, wide open.
- look: up, up and side, sideways, down, down and side.
- cheeks: stewed, sucked, tight, blow.
- nose: frown.
- mouth: closed, closed smile, smile, yawn, kiss, tense, folds around the mouth.
- tongue: visible inside the mouth.
- teeth: above touching the lower lip, lower lip touching.
- predefined: happy, sad,...

The face predefined attribute was created to carry emotional values like happy, sad or angry. This attribute can be used when high accuracy is not necessary, that is, describe facial expression just as happy or unhappy is enough for a good articulation.

The localMovement element has three children elements: fingers, wrist and forearm. In local movements just hands or wrist move, and the hands location in space does not change. Figure 3 illustrates the description of local movements in the proposed notation.
Global movement is trajectory between two holds, in the same sign. The movement direction can be vertical or horizontal, and can be circular (clockwise or counterclockwise), halfCircle, straight or zigzag. It is possible to add contact during global movement, described with child element contact of element globalMovement. The contact has following types: beat, touch, brush, rub (move, and remains on the surface) and pick. The attributes local1 and local2 refer to contact points with the hand, body or face.

GlobalMovement speed attribute can be fast, slow or normal. When not filled, the attribute value is considered normal. The movement may have its speed changed during sign articulation, setting final attribute. In real situations, speed may vary during the signing, for example, at the end of a phrase or to assign intonation to a particular word. In this case, a sign can start running at normal speed and have fast speed at the end of articulation.

IV. IMPLEMENTATION

The 3D model was created to represent, with realistic details, a female character. Figure 4 illustrates the polygonal mesh(a), its control skeleton(b) and the rendered image(c), that is, how the model is seen by the user.

The model is more detailed in the upper torso, especially on hands, that is the most important and used part in articulation. The control skeleton, guided by the transcription system presented in section 2, allows the deformation of model surface creating animation. The model also has textures that reproduces the appearance of skin, hair, iris and clothing.

The model was created with Autodesk Maya and exported as FBX ASCII. The .fbx file is imported on Microsoft XNA development platform, where a C# program reads the XML sign files and creates the animation.

The program interface, illustrated in Figure 5, allows the user to have additional control over the screen view.

Thus, you can zoom in or out or change the camera angle. Additional controls can increases or decrease the articulation speed, so it is possible to create a slow motion animation. This is especially good for learning sign language.

The program interface has an area for data entry. The user can loads a XML file with, and/or type new content for articulation.

The hand configuration settings are described separately in an XML file, which is uploaded by XNA. The XML configuration file stores the fingers joints angles necessary to achieve each possible hand shape. The XNA program reads the files and makes the corresponding rotation.

For example, for B configuration, shown in Figure 6, we have the follow description:
V. CONCLUSION

A sign language virtual agent could increase the computational accessibility by people with hearing loss, thus improving the human-machine interaction for these users.

The software implementation is almost finished and a evaluation phase will be carried out to verify the effectiveness of the software and the transcription system.

It is important to highlight the multidisciplinary nature of this work, since the detailed study of sign aspects may help in linguistic studies of sign languages.

REFERENCES


