Synthesizing Signing Avatars from Annotated Language Corpus

Carolina Neves, Hugo Nicolau, Luísa Coheur
name.surname@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

June 2020

Abstract

Signing avatars have gain an increase interest in the last years. The anonymity and flexibility provided in the creation of these animations were strong factors for this growth. Nevertheless, the animation of these virtual agents can be a complex task for non-expert users in the area of animation.

In this dissertation, we present a tool for linguists in which we intend to leverage their domain expertise alongside their annotated corpus in the animation of signing avatars. This corpus contains detailed description of Portuguese Sign Language signs through the use of a symbolic representation, more specifically HamNoSys. However, HamNoSys is not machine-readable, requiring the conversion to its XML-Component, SiGML, which will be use in the animation process. Although our tool was developed in the scope of Portuguese Sign Language, it can be used with any other sign language.

We conducted a user study with seven people who use Portuguese Sign Language, to assess the performance of our tool. Results show that synthetic animation still returns very robotic and unnatural signing avatars, which leads to difficulties in understanding the content. In addition, the right velocity and use of facial expressions influence the overall comprehension of these virtual agents. Finally, we describe our achievements, the limitations of our tool and suggestions for future work.

Keywords: Signing avatars, HamNoSys, SiGML, Sign Languages, Synthetic animation

1. Introduction

In contrast to what is widely believed, there is not one universal sign language but many. According to the World Federation of the Deaf[1] there are over 300 sign languages and 70 million deaf people using them around the world. In Portugal, in accordance with the Portuguese Association of the Deaf, about thirty thousand people use PSL as their main source of communication[2]. SL are highly structured languages with linguistic rules distinct from their spoken counterparts [11].

Due to the physical components used in sign languages, online dictionaries for sign languages, in which the entries consist of video recordings of people signing, are very common. Signbank [6] and Porto Editora's online dictionary[3] are two online dictionaries for their own sign languages, Australian and Portuguese Sign Language respectively. On the other hand, SpreadTheSign[4] is an online dictionary with a database which contains content of several sign languages from around the world, including over thirteen thousand signs recorded for PSL.

However, these approaches cannot fulfil the same role as written text. Videos of people signing lack flexibility since they are not easy to edit or reuse, requiring a new recording of the content in most cases. The combination of videos will most likely produce incoherent and unpleasant results [5]. In addition, the creation of a large corpora for sign languages through video is a time-consuming task [31]. Finally, videos do not allow for the author's anonymity to be preserved [20].

Many communication barriers exist for sign language users, and signing avatars (computer animations of humans) have the potential to break down these barriers for Deaf people who prefer sign language or have lower literacy in written language. For instance, these avatars could automatically replace the displayed text with sign language animations, provide anonymity to deaf individuals as well as a more flexible approach in the generation and edition process [27, 20, 1, 19, 16]. In addition, the scalability that is possible to achieve through this generation process allows for a fast

[1] `http://wfdeaf.org/our-work/`

[2] `http://www.apsurdos.org.pt/index.php?option=com_content&view=article&id=43&Itemid=57`

[3] `https://www.infopedia.pt/dicionarios/lingua-gestual`

[4] `https://www.spreadthesign.com/pt.pt/search/`

creation of a lot of content [3]. For these reasons, linguists and computer animators have been jointly interested in developing signing avatars capable of realistic communication in sign languages. Large corpora of intelligible animation data would be valuable research material. It would enable linguists to illustrate and share new concepts, invent new signs as well as develop dictionaries.

The creation of signing avatars requires a mixture of knowledge and expertise hard to find in the same place, such as animators and linguists. Current technologies to animate signing avatars either rely on combining motion capture clips [29, 24, 31, 5], in the use traditional animation tools [3] or in the use of a symbolic representation which encodes the avatar's movements [22, 28, 14].

The automatic generation of avatars' animations through the use of a symbolic representation allows for a cost-effective and simple approach in the process of animating a signing avatar. Nevertheless, the animations originated through this method can be of difficult perception and unrealistic. Producing comprehensible sign language content with avatars remains an unsolved problem due to the great detail required in its implementation. This raises the question: **Can avatar animations generated through an automatic interpretation of a symbolic representation be understood by sign language speakers?**

Our main goal is to leverage the linguists' knowledge in an existing notation to animate signing avatars. Our tool can be used with any sign language and will take advantage of an already existing corpus, with linguists' annotations in HamNoSys, the symbolic representation chosen, and convert it into its machine-readable XML-component, SiGML, therefore avoiding a broken pipeline in the animation process. The information within this XML will be used for the synthesis of the signing avatar.

The main contributions of the paper are: first, develop a tool of easy use, available to any sign language, that animates a signing avatar through annotated content from a linguistic corpus, without the need for any previous knowledge in the field of animation; second, provide linguists with means to help in the process of annotating content; third, provide improvements and feedback on the comprehensibility of the animations created with our tool, obtained from the results of a user study conducted with participants with knowledge in PSL; fourth, analyze the existing technologies for the animation of sign language avatars as well as a review on its advantages and disadvantages to retrieve the most relevant requirements for the development of the system

2. Background

In order to provide the reader with some basic knowledge of sign languages, in this section we present some detailed notions of sign languages components and structure as well as the its annotation.

2.1. Sign Languages

Sign language is not an universal language. Sign languages are natural languages and just like spoken languages they differ from country to country, hence the existence of several of them.

There is always the misconception that sign languages are somehow dependent on spoken languages. That they are spoken languages expressed in signs, which is incorrect. PSL is distinct from spoken and written Portuguese. It has its own morphology, phonology and syntax [12].

For this reason, for Deaf people, knowing the spoken version of their language is considered a secondary language. The majority of the native signers present limited reading skills and deaf adults most commonly present reading ability corresponding to early to mid-primary school level skills [9]. Therefore, the acquisition of information not only online but most importantly in education is an important barrier to break.

2.2. Sign Languages Components

Sign Languages are visual-sign languages. The message is conveyed by gestures and received by the visual channel. Each sentence in SL is constructed from a series of signs, which are arranged according to its proper syntax.

A gesture is considered a sign if it has movements, posture, position and hand shape required to construct a sign. It is a combination of both manual and non-manual components. Non-manual components are all those regarding body and face without considering hands, such as head, eyebrows, eyes, cheeks, mouth, torso and shoulder movements. Every sign is characterized by its configuration, orientation, location, movement and facial expression. These parameters occur simultaneously.

The manual representation of the alphabet is called fingerspelling. Fingerspelling was invented to facilitate the transfer of words from a spoken language to a sign language. It is used with signs not yet documented.

2.3. Annotation of Sign Language Data

Sign Languages are visual languages, therefore, a system that provides a written description of SL would be useful not only to deaf educators and individuals but also to linguists. Annotations provide SL in a written form and are widely used in the linguistic field. Nevertheless, although there are some famous annotations, none was yet accepted as standard [32, 7, 17].

In studies of SL, one of the first steps is the creation of an orthographic or phonetic transcription based on some recorded event, for example, video recordings. However, the process of annotating video recordings can be complicated due to the complexity of SL since both manual and non-manual components are used. For this reason, researchers use annotation tools. These tools allow the addition of text comments, annotations, such as the previously mentioned, or glosses to video recordings. Furthermore, these software also provide multiple tiers of annotation. Each tier can correspond to a different articulatory element, enabling linguists to annotate simultaneous movements. The most popular annotation tools are iLex, SignStream and ELAN and all of them are distributed with free licenses which allow their non-commercial use.

3. Related Work

The past two decades have seen a great evolution in terms of signing avatars. The use of this technology can vary from teaching purposes [8], to linguistic studies [24], to communication between hearing impaired and hearing [2].

The related work reviewed in this section is two-fold: first, we discuss research on synthesizing signing avatars through the use of annotations of sign languages; second, we examine previous attempts on the use of motion capture (MoCap) to animate signing avatars.

3.1. Procedural Synthesis from Sign Language Annotation

While multiple notation systems have been proposed [7, 12, 17], there is no universal standard. Annotations are often in gloss, a form of transliteration where written words in spoken languages are used to represent signs. The Hamburg Notation System for Sign Languages (HamNoSys) [17] is one of the most popular, designed to capture detailed human movements and body positioning for computer modelling. The notation system uses symbols to describe parameters of signs, such as hand shapes, hand orientations, movements, and non-manual components. Although HamNoSys symbols are readable by humans, notation systems typically require intermediary XML-based markup language representations that are computationally compatible and readable [7, 34]. For instance, Signing Gesture Markup Language (SiGML) [34] is compatible with HamNoSys. 1 illustrates how the word “BOLO” is annotated in HamNoSys (left column) and coded in SiGML (right column).

Multiple sign language avatar projects have been

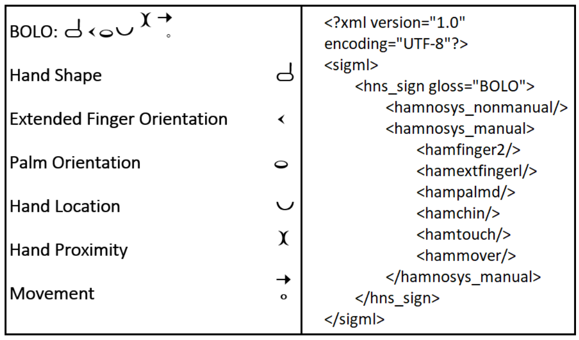

Figure 1: Representation of the Portuguese SL sign for “BOLO” (cake in English) in HamNoSys (left) and its codification in SiGML (right).

using HamNoSys and SiGML. For instance, eSIGN’s (Essential Sign Language Information on Government Networks) [34] avatar uses both notation systems to synthesize signing animations. The avatar will receive the SiGML and will then display the requested signs. More recently, SiS-Builder [9] is a sign language tool that provides a signer friendly graphical user interface. Users can create SiGML scripts either by entering HamNoSys strings of already stored signs or by creating HamNoSys lemmas online, up to a limit of four lemmas. Users may also add non-manual components to the sign. Afterwards, the respective SiGML script is generated and stored. However, the previously mentioned tools are not open source.

3.2. Data-driven Synthesis from Motion Capture with markers and gloves

Data-driven animation methods, or motion capture (MoCap), have shown a considerable uprise in computer animation. It is the most commonly used technique for sign language animation. Motion capture uses live data collected from various data sources which will drive the avatar’s skeleton. A possible data source used for this is through markers on the signer’s body and the use of gloves.

These technologies for animating an avatar can be used in different areas. One of these areas is focused on research. For linguistic purposes, in order to understand details such as timing, co-articulation, spatial references used, non-manuals or inflexions phenomena it is important to have a corpus capture with great detail [24]. This approach is also used in the field of human-humanoid interaction for the project SignCom, to engage in real-time dialogue, in which the animation of the avatar was captured through motion capture [15]. An hybrid approach with annotations is also an alternative. By using MoCap to construct a sign language corpus and the use to handle bone-and-joint system retrieved from the XML [8]. These approaches were implemented using specific and expensive material, various body markers and cameras.

3.3. Data-driven Synthesis from markerless Motion Capture

Motion capture can also be done through markerless devices. Famous devices in this area are affordable depth cameras with skeleton, Kinect, and hand motion tracking, Leap Motion.

Kinect provides a skeleton tracking feature which can be used to create sign languages content [27, 13, 31]. The kinect device is use for the creation of online dictionaries, such as DICTA-SIGN which populates content on the Web through users' videos [10] or more detailed dictionaries that in addition to kinect have manual tasks involving 3D animators and sign language specialists [31]. Besides, it can also be used for teaching oriented applications [5].

Leap Motion is a device smaller than Kinect, which connects to a computer using USB and can sense hand movements in the air above it which are then translated into actions for the computer to perform. For this reason, it was also used for sign languages, more specifically its manual component [25, 2].

The two sensors provide features that complement each other. The Kinects, Kinect v1 and Kinect v2, are two of the most accurate low-cost whole human body motion tracking sensors available whereas Leap Motion is one of the most accurate low-cost hand-tracking sensors [33]. Therefore, studies came with approaches to combine the two devices together [26, 18].

3.4. Applications and Design of Signing Avatars

The impact that these avatars can have in the Deaf community is not solely based on the functionality but also in other features. The focal point of the creation of signing avatars is to use it in sign language. However, deaf people as potential users of these technologies have little or no knowledge about avatars [23]. Facial expressions are a major feature in signing avatars. This feature not only provides expressiveness to the avatar but in sign language it also provides meaning [30]. In addition, the pauses and speed of signing also have major impact in the comprehensibility of a sign [4].

4. HamNoSys

HamNoSys is an alphabetic system describing signs mostly on a phonetic level. The following description of the system is based on [17]. This notation system is a combination of iconic and easily recognizable symbols which cover the parameters of hand shape, hand configuration, location and movement. HamNoSys can be internationally used since it does not rely on conventions that differ from country to country [17, 21].

4.1. Hand shapes

The description of hand shape is composed of twelve basic hand forms and three thumb combinations. These can be combined with diacritic signs for thumb positions and bending. Additional description concerning the fingers involved or the form of individual fingers can be specified. In Figure 1 we can see a simple hand shape for the Portuguese word “bolo” (“cake” in English).

4.2. Hand orientations

The description of the hand orientation is composed of two components: extended finger and palm orientation. The former provides information regarding the direction of an extended finger related to the signer's body (signer's view, birds' view, and view from the right). The later is relative with the former. For a given extended finger, it indicates an orientation of the palm around the shaft of the hand. A practical example is provided in Figure 1.

4.3. Hand locations

The locations of the hand can also be split into two components: The first provides information of the hand location in respect to other body parts, as the second determines the distance of the hand to this location. If the later is missing, a “natural” distance is assumed. In case both components are omitted, a neutral space is assumed. Such space is located in front of the upper part of the body. As shown in Figure 1, both the hand location and hand proximity are provided.

4.4. Hand Movements

Movements can be distinguished between straight, curved and zigzag lines, circles and similar forms. These can either be performed sequentially or cotemporally. Also, repetitions of movements can be specified. In the case of two-handed signs, it is possible to differentiate the actions for each hand.

5. From Annotations to a Signing Avatar

Research shows that the expertise required to work in the field of animation of signing avatars restricts the number of people suitable for this task in addition to its costly implementation. The lack of a proper software available for linguists in this field is still a limitation.

5.1. Approach

Our tool will be based on a synthetic animation approach, where the main focus is to leverage linguistics' expertise and the already existing annotated corpus in the synthesis of sign language avatars by developing a tool which will simply receive their corpus annotated in HamNoSys as input and provide the signing avatar as output.

This work was partially supported by Fundação para a Ciência e a Tecnologia through grants UIDB/50021/2020 and PTDC/LLT-LIN/29887/2017. The annotations in HamNoSys as well as its glosses will be provided by the team of Universidade Católica. The videos which constitute the corpus are performed by Portuguese deaf people, in the range of ten to sixty years old. These videos are diversified, having formal, non-formal, spontaneous or a previous established subject speeches. The annotation tool used for this purpose is ELAN [5].

The architecture of our tool is mainly divided into two parts. The first part consists on the processing of the data obtained from ELAN. The second part, mainly focus on the animation of the avatar. Our tool can receive two distinct inputs. Either it receives an HTML or an SiGML. For the former, upon acquiring the HTML annotations files extracted from the ELAN, our tool will process these and produce their respective SiGML. Afterwards, the content can either be visualized or saved, as illustrated in Figure 2.

5.2. HamNoSys to SiGML

The first part of this project consists on the processing of the data obtained from ELAN, which corresponds to step "Corpus in ELAN (.html)" to step "SiGML file (.sigml)" in Figure 2. The annotated content must be translated into XML, due its advantages in processing. Therefore, annotations in HamNoSys from ELAN will be translated to the XML framework SiGML. This represents an intermediate step in the pipeline of synthetic animation. Even though several tools were previously developed, to the best of our knowledge, none of them is freely available.

5.3. SiGML to Signing

Once the data is saved into SiGML, the following step is to animate the avatar. Due to time restrictions, not all HamNoSys components are fully implemented. The avatar signing is limited to the use of one hand. Nonetheless, our tool is not restricted to any sign language, it will animate any content as long as it is properly annotated in HamNoSys.

As previously described, HamNoSys has twelve basic hand shapes. Each of these twelve hand shapes has its own animation of one frame. From these basic hand configurations, others can be created by combining HamNoSys symbols. The implementation of thumb position is made during run time. In addition, some bending of the fingers are also implemented.

The hand's location was implemented using Inverse Kinematics (IK). By choosing a certain position in space, IK will work on finding a valid way of orienting the avatar's joints so that the end point lands on such position. Besides being more intuitive, this solution will also allow the avatar to produce more natural movements. Each location is identified by its respective sphere, which is use to guide IK.

The hand orientation in HamNoSys is illustrated through two components, extended finger and palm orientation. Through a more practical view, the former defines the direction in which the knuckles of the fingers point towards in relation to the main axis of the avatar. Meanwhile, the the palm orientation describes the rotation of the wrist in relation to the axis of the avatar's hand. The implementation of this component was done using IK, through vectors for both the extended finger component and palm orientation component.

Finally, the hand's movements were also done with IK, by moving its respective sphere with mathematical formulas.

5.4. Integration with PE2LGP

Our tool is incorporated with a previous project also developed for PSL, PE2LGP [13]. PE2LGP provides a platform for the introduction of new signs without specially advanced knowledge in animation, through the use of Kinect or keyboard and mouse , respectively "Sistema 3" and "Sistema 2" in 3. These animations can be later visualised or used in a basic translator from written Portuguese to PSL, respectively menu "Gestor" and "Sistema 1" in 3.

The menu "Sistema 4" in 3 represents our tool. The animation of the avatar will be achieved with Unity by receiving a SiGML.

5.4.1 Visualize Animations

The interface of this feature is fairly simple, available in Figure 4. The button most to the left corresponds to the visualize option of the system. This will forward the user to the file system and will only accept an HTML or SiGML file. The processed data will then be animated by the avatar.

As described in Figure 2, besides animating the avatar, an error log is returned. This log will have a description of easy understanding of eventual notation errors found during the execution. For instance the use of nonexistent symbols or misplaced of symbols, the order of the annotations must follow hand shape - hand orientation - hand location - hand movements, also if the annotation does not have a hand shape defined it will not be accepted as valid either and finally it controls the number of symbols used for extended finger and palm orientation (minimum of one and maximum of two symbols for each).

[5] https://archive.mpi.nl/tla/elan

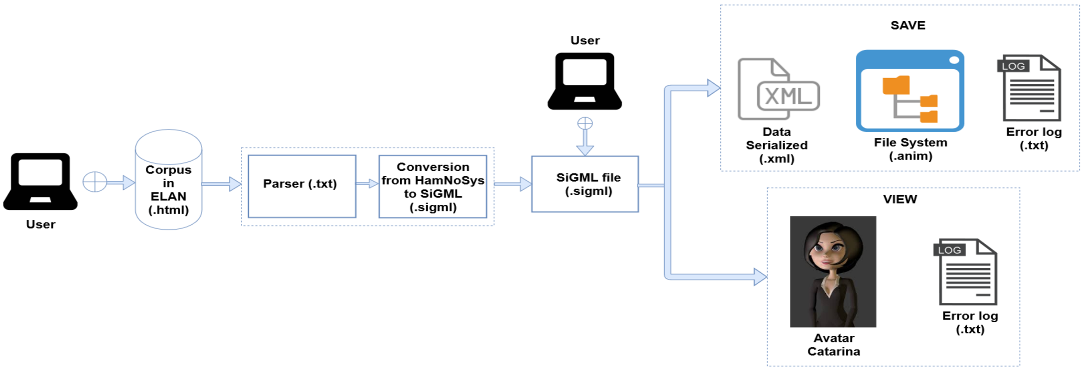

Figure 2: Tool's architecture.

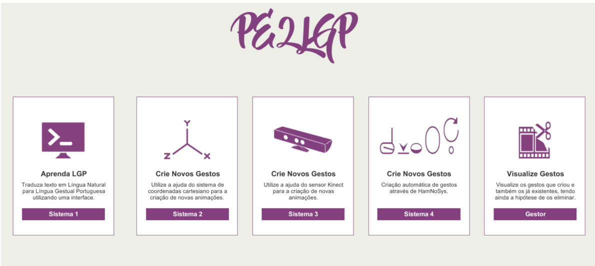

Figure 3: PE2LGP main menu.

Figure 4: Tool's menu.

5.4.2 Save Animations

The other feature of the system consists on saving signs as an animation clip, step "SAVE" in Figure 2. In Figure 4 this option is illustrated by its button (with a disk), next to the "VIEW" option. The pipeline will be the same as with visualize, the only difference residing on the fact that the user will not see the avatar signing, the process will be performed by an avatar in the background, and the animations will be saved, step "Save animations per gloss" in Figure 2. Each animation clip will be identified by its gloss. In case the animation of the gloss already exists, a counter is implemented to create variations to the file name. Therefore, the system allows for **various versions of the same sign**.

The animations will be saved in the file system, represented as step "File System (.anim)" in 2, together with the animations from the previous version of the project [13] and can be visualise in its respective component "Gestor" from 3.

Besides saving the animation clips by gloss, within this component another feature is available, step "Data serialized (.xml)" displayed in Figure 2. In order to help linguists with their annotations, the **serialization of the data** was implemented. Alongside the gloss, its respective HamNoSys codes are saved in an XML file. Just as with the animations, in this XML there too can be more than one annotation per gloss.

L2F, which is a department from INESC-ID focused on the study of natural language with projects in fields such as speech recognition and machine translation, is developing a library which by receiving written Portuguese returns its respective glosses in the correct order, with additional markers that identify the finger spelled words, facial expressions, among other characteristics of PSL. By incorporating this project with our tool, it will be possible for an automatic translation from written Portuguese to PSL.

The architecture of this future feature, not yet implemented, is illustrated in Figure 5. The population of the file system is done with annotations from the Corpus, step "Corpus in ELAN (.html)" in 5, which are converted into a SiGML file. Optionally, the user can do it through a SiGML file directly, as in step "SiGML file (.sigml)" from 5. These files are used to synthesized the avatar signing, and if the user chooses to save the content, "SAVE" in 2 and in 5, this will be stored in the file system. Upon receiving a gloss, a lookup in the file system is performed, as shown in step "Lookup" in 5. This lookup returns the gloss respective animation clip, which is then played by the avatar.

Finally, when saving signs, the system also produces an error log, which has the description of errors occurred during run time, detailed in step "Error log (.txt)" from Figure 2, alongside with alerts in case the gloss in the process of saving already exists, if so, a variation of the gloss is created giving name to the file.

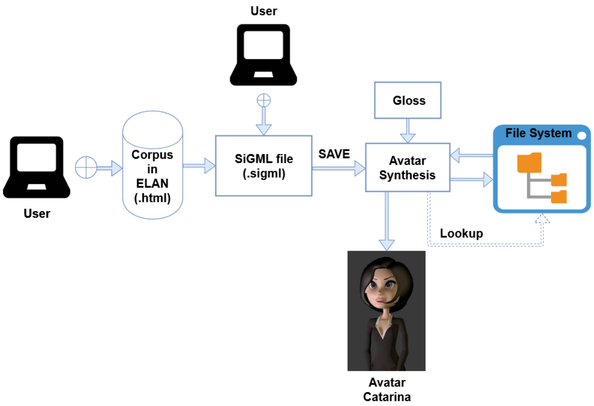

Figure 5: Architecture with lookup feature.

6. User Study

Our user study focused on evaluating the performance of our tool and how its content is signed by an avatar. To assess the signing the content return by our tool we will performed a joint evaluation to compare two avatars: our on progress avatar Catarina and the state of the art online avatar Anna[6].

6.1. Research Questions

This study aims to answer to one main research question:

1. How does our avatar perform in comparison with a state of the art online avatar?

2. Are signing avatars animated through synthetic animation effective as an agent for communication in Portuguese Sign Language?

6.2. Study Preparation

The content used for the evaluation was mostly retrieved from the corpus with some signs annotated based on content from SpreadTheSign[7], in which the content is also signed by native deaf people.

Overall twenty sentences were created for this evaluation. Each sentence was translated from Portuguese to PSL by an interpreter and linguistic. The annotations from PSL to HamNoSys were also conceived together with experts. Afterwards, the twenty sentences and their respective HamNoSys were annotated in ELAN. For each sentence, its respective html was exported.

By running these html files with our tool, the output generated is their SiGML files. For twenty sentences, twenty SiGML files were returned. The JASigning avatar Anna[8] with its default settings and, the on progress, avatar Catarina were animated with these SiGML files.

Both avatars were recorded animating each SiGML. These recordings were also verified by the interpreter and linguistic in PSL of the project, and small corrections were made in the initial HamNoSys annotations. These corrections were applied, new html files were exported from ELAN, and new SiGML were returned and then use to animate the avatars.

For the evaluation of the animations produced by the avatars we created seven different versions of tests for seven users. For each version both the content animated as well as the order of the sentences at which each avatar would animate were chosen randomly[9], considering that two sentences were always assigned to avatar Catarina since avatar Anna could not play these. Besides, the first avatar which appears signing changes for each test.

6.3. Procedure

The questionnaires were created in Google forms. The questionnaire was emailed to all the participants. To obtain participants, the team from Universidade Católica shared a link of a Google form in which those who had interest in evaluating our tool were asked to submit their email. All the user tests were unique, as in, each participant had its own and distinct test. The same questionnaire was never performed for more than one person.

6.4. Study Design and Data Analysis

This questionnaire is composed of twenty four sections. The first section regards general information about the participants. Afterwards, each sentence has its own section, therefore, twenty of the sections, ten sentences per avatar, correspond to the content being animated by the avatars. Besides, after the ten sections correspondent to each avatar, an additional section requires the general evaluation of the avatar in terms of speed, general quality, comprehensibility, naturalness, grammatical correctness, hands' configurations, hands' orientations, hands' locations and hands' movements.

7. Participants

We were able to recruit seven participants (four females). Their ages ranged from 26 to 55 (M=43.86, SD=11.88) years old. five of the participants are deaf with three native. From the remaining four, one uses PSL since birth, two for more than twenty five and ten years, and finally the third's mother language was PSL until the age of 10. Three of the five participants do not relate with any PSL regional variant, one relates with the south's variant, another with both south's and center's variant and the remaining two with the center's variant. Only

[6] http://vhg.cmp.uea.ac.uk/tech/jas/vhg2019/cwa/TwoAvServer.html

[7] https://www.spreadthesign.com/pt.pt/search/

[8] http://vhg.cmp.uea.ac.uk/tech/jas/vhg2019/cwa/TwoAvServer.html

[9] //www.random.org/lists/

one participant has ever used signing avatars but rarely.

8. Results & discussion

8.1. Signing Characteristics

Sign languages are complex languages due to the number of components required to produce a proper signing content and the detail behind each one. We present the results related to the impact of each of the components implemented in this project.

8.1.1 Velocity

The first component evaluated during this study was the perceived appropriateness of velocity of the signing of each of the avatars. Avatar Catarina presents a higher dispersion of the results which demonstrates a high lack of consensus on its velocity quality, from rate 1 to 3. On the other hand, for the avatar Anna, the values suggest that overall testers have a higher level of agreement with each other, with a rate of 1 and 2. All in all, the statistical analysis between both avatars' signing velocity revealed no significant differences between their rates ($Z = -0.447^b, p = 0.655$; Avatar Anna Mdn=1, IQR=1; Avatar Catarina Mdn=1, IQR=2), suggesting that neither velocities truly met the expectations.

8.1.2 Overall quality, Understandability and Naturalness

For the overall quality, the understandability and naturalness, the data is found between rates 1 and 2 for both avatars, with a maximum of rate 3, out of the inter-quartile range for avatar Anna. The overall quality ($Z = -0.557^b, p = 0.564$; Avatar Anna Mdn=1, IQR=1; Avatar Catarina Mdn=2, IQR=1), understandability ($Z = -0.447^b, p = 0.655$; Avatar Anna Mdn=1, IQR=1; Avatar Catarina Mdn=1, IQR=1) and naturalness ($Z = -1^b, p = 0.317$; Avatar Anna Mdn=1, IQR=1; Avatar Catarina Mdn=1, IQR=1) indicate no substantial difference between both avatars

8.1.3 Grammatical Correctness

For the grammatical correctness, although avatar Anna has a higher maximum value, the central tendency is lower than the value for avatar Catarina ($Z = 0^b, p = 1$; Avatar Anna Mdn=1.5, IQR=1,25; Avatar Catarina Mdn=2, IQR=1), indicating a stronger consensus among users for its rate. Considering that the grammatical input (sentence construction and content in HamNoSys) for both avatars was the same, this difference between results must be related to the performance of the avatar in the remaining components of PSL.

8.1.4 Hand's Configuration

The first HamNoSys component evaluated was the hand's configurations. Despite the results acquired with avatar Catarina being more spread out, the median values are equal in both cases, suggesting a central tendency of value 2 in the Likert Scale. As before, the null hypothesis is retained ($Z = -1.414^b, p = 0.157$; Avatar Anna Mdn=2, IQR=1; Avatar Catarina Mdn=2, IQR=2), as there was no considerable statistical variation between the results of both avatars.

8.1.5 Hand's Orientation and Hand's Movements

The results obtained for both the hands' orientations ($Z = 0^b, p = 1$; Avatar Anna Mdn=1, IQR=1; Avatar Catarina Mdn=1, IQR=1) and hands' movements ($Z = -1^b, p = 0.317$; Avatar Anna Mdn=1, IQR=1; Avatar Catarina Mdn=1, IQR=1), show no significant difference between the performance of both avatars. The data is spread out between rates 1 and 2, suggesting a lack of agreement between these two rates amongst testers; however, it portrays the poor performance of these components in both avatars.

8.1.6 Hand's Location

Finally, the evaluation of the hands' locations also showed no significant differences between the avatars ($Z = -1^b, p = 0.317$; Avatar Anna Mdn=1, IQR=1; Avatar Catarina Mdn=1, IQR=1). Again, the central tendency of the results for both is 1, with avatar Catarina achieving a maximum of rate 3.

8.2. Comprehension

In this section, we analyze the number of visualizations of every sentence for each avatar, as well as the percentage of content correctly understood by the testers for each avatar.

8.2.1 Number of visualizations

During our user study, for each sentence, the testers were asked to report the number of visualizations needed to understand the content.

This indicated a higher consistency between testers in the number of visualisations for avatar Anna. On the other hand, for avatar Anna is was also required a high number of visualizations for each sentence.

8.2.2 Content

To evaluate the comprehension of the content, the percentage is calculated based on the number of glosses understood in each sentence, with 100%

corresponding to every single gloss being understood by the participant.

The number of glosses seems to not influence the content's proper interpretation. Half of the above sentences have up to four glosses, which is the highest number of glosses per sentence used in this study. In addition to this, we can gather that the majority of these sentences have the highest values with avatar Catarina with a lower number of visualizations per participant.

Both avatars show some inconsistency in the results (Avatar Anna M=26.85, SD=20.62; Avatar Catarina M=29.05, SD=25.99), from percentages of 0, no content was comprehended, to percentages above 60. This dispersion of the data can be related with the fact that some sentences did not have the same number of visualizations for both avatars, which might lead to misleading results. Overall, the our results showed that there was no significant statistical change between the signing of the two avatars ($Z = -0.240^b, p = 0.811$).

8.3. Participants' Feedback

As part of our study, the participants were also asked for additional comments about the avatars' signing.

With regard to the **performance of the avatar Anna**, comments concerning its velocity, naturalness and facial expressions stand out. The velocity was defined as too fast by some users and the avatar as very robotic. The lack of facial expressions was also noticed.

Taking into to account the **performance of avatar Catarina**, once more the main focus was its velocity, naturalness and facial expressions. Certain content was described as very slow and difficulties arose about whether the content referred to isolated signs or one sentence. The latter might be related to the former, since the low velocity might lead people to question whether the signing has ended. Regarding the naturalness, the avatar viewed as robotic, the lack of flexibility and mobility was noticed and testers described the need to reflect and make an effort while viewing the videos in order to understand what was being signed, due to unnatural performance of the agent. The absence of facial expressions was also noted.

9. Conclusions

We developed a tool which is able to read linguistics' annotations in PSL and animate an avatar directly, without any need of expertise in the field of animation. We performed a user study with seven participants to test our tool. Results show that the avatars are still very robotic and unnatural while signing and that velocity and facial expressions prove to have great weight in sign languages. Although our tool allows for a fast and convenient option to animation signing avatars, the process of animation still requires some work and should be seen as a first step in this field.

1. **How does our avatar perform in comparison with a state of the art online avatar?**

   Both avatars were described as robotic and unnatural. The state of the art online avatar's (Anna) velocity was considered too fast while our avatar's (Catarina) velocity as too slow. Overall, the results from our user study showed that the avatars performed poorly.

2. **Are signing avatars animated through synthetic animation effective as an agent for communication in Portuguese Sign Language?**

   Synthetic animation of signing avatars shows potential in this field, major improvements are still required. As of yet, these virtual agents do not sign naturally, lacking facial expressions and performing at velocities not in correspondence with the expectations.

As of now, our tool is able to synthesize content directly from an annotation tool widely used in the linguist field, ELAN. Nevertheless, several features can be further implemented to improve the overall signing of the avatar, such as the use of facial expressions and the signing speed.

10. Future Work

Although our tool has limitations which might lead to a result not useful in the automatic generation of animations, it has the potential to be used as a tool to accelerate the creation process of manual signs. We foresee the use of our tool to quickly animate a great amount of content that can later be manually improved to return better animations. These corrections could even be used together with **algorithms of artificial intelligence** for the tool to learn how to automatically generate better animations.

Further possible avenues for future research lie in a further and more complex exploration of synthesized animation through HamNoSys, the possibility to control the signing speed, the implementation of collisions and of natural and unscripted movements of the avatar.

References

[1] The dicta-sign Wiki: Enabling web communication for the deaf. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7383 LNCS(PART 2):205–212, 2012.

[2] Desenvolvimento de um tradutor de Língua Gestual Portuguesa. page 107, 2014.

[3] N. Adamo-Villani, V. Popescu, and J. Lestina. A non-expert-user interface for posing signing avatars. *Disability and Rehabilitation: Assistive Technology*, 8(3):238–248, 2013.

[4] S. Al-khazraji, L. Berke, S. Kafle, P. Yeung, and M. Huenerfauth. Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations. pages 259–270, 2018.

[5] J. Bento, A. P. Claudio, and P. Urbano. Avatars on Portuguese sign language. pages 1–7, 2014.

[6] S. Cassidy, O. Crasborn, H. Nieminen, W. Stoop, M. Hulsbosch, S. Even, E. Komen, and T. Johnston. Signbank : Software to Support Web Based Dictionaries of Sign Language. pages 2359–2364, 2017.

[7] A. C. d. R. Costa and G. P. Dimuro. SignWriting and SWML: Paving the way to sign language processing. *Traitement Automatique des Langues de Signes, Workshop on Minority Languages*, pages 11–14, 2003.

[8] D. F. De Conti, C. Z. Bolognini, L. C. Coradine, P. H. d. S. Brito, C. M. Ferreira, Â. B. Benetti, P. D. P. Costa, J. M. De Martino, L. M. G. Angare, W. M. do Amaral, I. R. Silva, E. T. Poeta, and K. M. O. Kumada. Signing avatars: making education more inclusive. *Universal Access in the Information Society*, 16(3):793–808, 2016.

[9] E. Efthimiou, S.-E. Fotinea, T. Goulas, A. Vacalopoulou, K. Vasilaki, and A.-L. Dimou. Sign Language Technologies and the Critical Role of SL Resources in View of Future Internet Accessibility Services. *Technologies*, 7(1):18, 2019.

[10] E. Efthimiou, S. E. Fotinea, T. Hanke, J. Glauert, R. Bowden, A. Braffort, C. Collet, P. Maragos, and F. Lefebvre-Albaret. The dicta-sign Wiki: Enabling web communication for the deaf. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7383 LNCS(PART 2):205–212, 2012.

[11] R. Elliott, J. R. W. Glauert, and J. R. Kennaway. A framework for non-manual gestures in a synthetic signing system. *2nd Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Cambridge*, pages 127–136, 2004.

[12] R. Elliott, J. R. W. Glauert, J. R. Kennaway, and I. Marshall. The development of language processing support for the ViSiCAST project. pages 101–108, 2004.

[13] R. Emanuel, D. Mestrado, and E. Inform. PE2LGP: Do Texto à Língua Gestual (e vice-versa). 2016.

[14] S. Gibet, N. Courty, K. Duarte, T. L. Naour, S. Gibet, N. Courty, K. Duarte, T. Le, N. The, S. System, S. Gibet, N. Courty, K. Duarte, and T. L. E. Naour. The SignCom System for Data-Driven Animation of Interactive Virtual Signers : Methodology and Evaluation. 2012.

[15] S. Gibet, N. Courty, K. Duarte, T. L. Naour, S. Gibet, N. Courty, K. Duarte, T. Le, N. The, S. System, S. Gibet, N. Courty, K. Duarte, and T. L. E. Naour. The SignCom System for Data-Driven Animation of Interactive Virtual Signers : Methodology and Evaluation. 2012.

[16] S. Gibet, F. Lefebvre-Albaret, L. Hamon, R. Brun, and A. Turki. Interactive editing in French Sign Language dedicated to virtual signers: requirements and challenges. *Universal Access in the Information Society*, 15(4):525–539, 2016.

[17] T. Hanke. HamNoSys-Representing sign language data in language resources and language processing contexts. *Proceedings of the Workshop on Representation and Processing of Sign Language, Workshop to the forth International Conference on Language Resources and Evaluation (LREC'04)*, pages 1–6, 2004.

[18] A. Heloir and F. Valenciennes. Towards an Intuitive Sign Language Animation Authoring Environment For the Deaf. 49(0).

[19] M. Huenerfauth, P. Lu, and A. Rosenberg. Evaluating importance of facial expression in american sign language and pidgin signed english animations. page 99, 2011.

[20] H. Kacorri, P. Lu, and M. Huenerfauth. Evaluating Facial Expressions in American Sign Language Animations for Accessible Online Information. 2013.

[21] K. Kaur and P. Kumar. HamNoSys to SiGML Conversion System for Sign Language Automation. *Procedia - Procedia Computer Science*, 89:794–803, 2016.

[22] R. Kennaway. Synthetic Animation of Deaf Signing Gestures. pages 146–157, 2007.

[23] M. Kipp, Q. Nguyen, A. Heloir, and S. Matthes. Assessing the deaf user perspective on sign language avatars. page 107, 2011.

[24] P. Lu and M. Huenerfauth. Collecting a Motion-Capture Corpus of American Sign Language for Data-Driven Generation Research. *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, (June):89–97, 2010.

[25] R. B. Mapari. Analysis of Multiple Sign Language Recognition Using Leap Motion Sensor. 2017.

[26] G. Marin, F. Dominio, and P. Zanuttigh. HAND GESTURE RECOGNITION WITH LEAP MOTION AND KINECT DEVICES. *International Conference on Image Processing(ICIP)*, pages 1565–1569, 2014.

[27] A. Muhammad, M. F. Addenan, M. M. Latiff, B. Haris, S. S. Surip, and A. S. A. Mohamed. Interactive Sign Language Interpreter using Skeleton Tracking. 8:137–140, 1843.

[28] A. Othman and M. Jemni. An XML-gloss annotation system for sign language processing. *2017 6th International Conference on Information and Communication Technology and Accessbility, ICTA 2017*, 2017-Decem:1–7, 2018.

[29] A. Sanna, F. Lamberti, G. Paravati, and F. D. Rocha. A kinect-based interface to animate virtual characters. *Journal on Multimodal User Interfaces*, 7(4):269–279, 2013.

[30] R. G. Smith and B. Nolan. Emotional facial expressions in synthesised sign language avatars: a manual evaluation. *Universal Access in the Information Society*, 15(4):567–576, 2016.

[31] M. Soares, B. Meireles, S. Gonçalves, M. Aschoff, T. Maritan, and V. Becker. A Process for Semi-Automated Construction of Sign Language Dictionaries. *Anais do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres*, pages 132–136, 2017.

[32] W. C. Stokoe and M. Marschark. Sign language structure: An outline of the visual communication systems of the american deaf. *Journal of Deaf Studies and Deaf Education*, 10(1):3–37, 2005.

[33] C. S.-L. Tibor Guzsvinecz, Veronika Szucs. Suitability of the Kinect Sensor and Leap Motion Controller—A Literature Review. 2019.

[34] I. Zwitserlood, M. Verlinden, J. Ros, S. V. D. Schoot, and T. Netherlands. SYNTHETIC SIGNING FOR THE DEAF: eSIGN.