Synthesizing Signing Avatars from Annotated Language Corpus

Carolina Isabel Martins Neves

Thesis to obtain the Master of Science Degree in

Computer Science and Engineering

Supervisors: Prof. Hugo Miguel Aleixo Albuquerque Nicolau
Prof. Maria Luísa Torres Ribeiro Marques da Silva Coheur

Examination Committee

Chairperson: Prof. Nuno João Neves Mamede
Supervisor: Prof. Hugo Miguel Aleixo Albuquerque Nicolau
Member of the Committee: Prof. Augusto Emanuel Abreu Esteves

June 2020

“Let us fight for every woman and every man to have the opportunity to live healthy, secure lives, full of opportunity and love. We are all time-travellers, journeying together into the future. But let us work together to make that future a place we want to visit.”

Stephen Hawking, “World Economic Forum’s Annual Meeting Speech” at Sustainable Development Summit 2015

Acknowledgments

To Prof. Hugo Nicolau for the support and liberty to explore my own horizons whilst never ceasing to trust the result. For the undeniable presence and patience during the process as well as for always pushing me to do better and advising me for its accomplishment.

To Prof. Luísa Coheur for all the kindness, enthusiasm, openness and advice which describe her nature so flawlessly. Her presence and motivation throughout the project were essential for its completion.

To the rest of my team from Instituto Superior Técnico, Pedro and Matilde, for the company, ideas and laughs alongside the procrastination that is often so much needed. You definitely made it easier. One could not have asked for better group of four people to share this experience with.

To the team from Universidade Católica, for the time and support, which made this project possible, especially to Neide for the availability and kindness which were never at fault.

To my family and my boy for their (eternal) unconditional support, love and patience, without which I would not be standing here today. With whom I have learnt the importance of dedication, persistence and kindness. For always believing in the success and never failing to help for its achievement.

To my friends with whom I have grown and shared the peaks and valleys that come with the journey. Your love and support are crucial pieces to this puzzle which we like to call life.

To my furry babies, my cat and dog, for keeping me sane throughout the whole process. You truly are the most genuine, innocent and unconditional form of love.

To all of you, **thank you** for making this possible. It truly was a challenging and enriching experience. My hope is that at the very least readers can understand the complexity of sign languages and the challenges which deaf people are faced with in their everyday life, for the world craves for empathy.

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 of 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.

Resumo

Ao longo do tempo tem crescido o interesse em avatares gestuantes. O anonimato e flexibilidade devolvidas na criação destas animações é um forte factor para este crescimento. Contudo a animação destes agentes virtuais pode ser uma tarefa complexa para utilizadores não especialistas na área de animação.

Nesta dissertação, apresentamos uma ferramenta para linguistas com a qual pretendemos alavancar o seu conhecimento no domínio em conjunto com o seu corpus anotado para a animação de avatares gestuantes. Este corpus contém descrição detalhada de gestos em língua gestual Portuguesa através de símbolos, mais especificamente HamNoSys. Contudo, HamNoSys não é legível por um computador, o que requer a conversão para a sua componente XML correspondente, SiGML, que vai ser usada no processo de animação. Apesar de a nossa ferramenta ter sido desenvolvida no âmbito de língua gestual Portuguesa, pode ser usada com qualquer outra língua gestual.

Realizámos um estudo com sete pessoas que usam língua gestual Portuguesa para avaliar o desempenho da nossa ferramenta. Os resultados mostraram que animação sintéctica de avatares gestuantes ainda é bastante robótica e pouco natural, o que leva a dificuldades na compreensão do conteúdo gestuado. Para além disso, a correcta velocidade e o uso de expressões faciais influencia a compreensão destes agentes virtuais. Por fim, descrevemos as nossas conquistas, as limitações da nossa ferramenta e sugestões para trabalho futuro.

Palavras Chave

Avatares gestuantes; HamNoSys; SiGML; Línguas gestuais; Animação sintética

Contents

List of Figures

List of Tables

Acronyms

ASL American Sign Language

FK Forward Kinematics

HamNoSys Hamburg Notation System

IK Inverse Kinematics

MoCap Motion Capture

PSL Portuguese Sign Language

SiGML Signing Gesture Markup Language

SL Sign Languages

1

Introduction

Contents

In contrast to what is widely believed, there is not one universal sign language but many. Sign Languages (SL) are visual gesture languages performed through the use of hands, facial, and body expressions. 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 Portuguese Sign Language (PSL) as their main source of communication[2]. SL are highly structured languages with linguistic rules distinct from their spoken counterparts [15].

1.1 Motivation

Considering that sign languages are expressed through hands, facial and body expressions, the most common digital method to present their content is through videos. In 2010 Porto Editora created the first PSL online dictionary[3]. The entries in this dictionary consist of video recordings. Signbank is another online web dictionary for Australian Sign Language which provides a rich lexical database of sign language augmented with video samples of signs [7]. Nevertheless, these dictionaries are only focused on one sign language. 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 [41]. Finally, videos do not allow for the author's anonymity to be preserved [27].

Signing avatars (computer animations of humans), on the other hand, are a feasible option as they provide anonymity to deaf individuals as well as a more flexible approach in the generation and editing process [13, 23, 26, 27, 35] since they require no more resources other than text as input. In addition, the scalability that is possible to achieve through this generation process allows for a fast creation of a lot of content [1]. For these reasons, linguists and computer animators have been jointly interested in developing signing avatars capable of realistic communication in sign languages. A 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.

[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/

1.2 Problem

In spite of presenting great potential in the field of sign languages, 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 can rely on combining motion capture clips [5,31,39,41], in traditional animation tools [1] or it can also be performed using a symbolic representation [21,29,37], which encodes the avatar's movements through symbols.

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?**

1.3 Approach

Keeping the previous mentioned limitations in mind, 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 Hamburg Notation System (HamNoSys), the symbolic representation chosen, and convert it into its machine-readable XML-component, Signing Gesture Markup Language (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.

We will introduce a new solution to the process of animating an avatar through annotated content. We will also evaluate the quality of the signing content performed by an avatar.

With our tool the user can either visualize or save the content synthesized by the avatar. While visualizing the content, a log file with an easy understandable description of notation errors occurred during the annotation process will also be produced. On the other hand, when saving the content synthesized by the avatar, besides storing the animations, a serialization of the data is also performed, which can be used by linguists to assist and facilitate the process of annotation. The animations created with our tool will be saved together with the content produced with a previous version of the tool [18]. Finally, an error log is returned in this option as well, containing alerts, in the case there is repeated content, and errors' description.

1.4 Contributions

The main contributions of this dissertation are:

1. Develop a tool of easy use, available to any sign language, that animates a signing avatar through annotated content from a linguistic corpus, avoiding a broken pipeline on the animation process and without the need for any previous knowledge in the field of animation.
2. Provide linguists with means to help in the process of annotating content.
3. 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.
4. An analysis of 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.
5. Regarding the work carried out under this dissertation, a paper was accepted in the Language Resources and Evaluation Conference (LREC).

   - **HamNoSyS2SiGML: Translating HamNoSys Into SiGML**: Article Publication at Language Resources and Evaluation Conference (LREC) 2020 [Accepted].

1.5 Thesis Outline

In this dissertation, we will contribute to the synthetic animation of signing avatars through the use of annotations. In Chapter 2 we provide some important notions of sign languages and methods of annotations used. In Chapter 3 we present a state of the art analysis from the fields of synthetic animation, to sign language annotations, data-driven animation with markers and markerless, as well as possible uses and designs of signing avatars. Proceeding to Chapter 4 we depict all of our tool's components' implementation. Then, in Chapter 5 we present our research questions, evaluation methods and the collection of the data as well as the interpretation and description of our evaluation results. Finally, in Chapter 6, we deliberate on our current achievements, limitations and suggestions for future work. To support and better illustrate concepts mentioned in the above sections, several appendixes were added to the document.

2

Background

Contents

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. It is unclear how many sign languages currently exist worldwide. Each country generally has its own native sign language, and some have more than one. In Portugal, we have Portuguese Sign Language.

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 [16]. An evident difference between both PSL and spoken and written Portuguese is in its syntax. PSL syntax follows an Object - Subject - Verb (OSV) structure whereas the Portuguese syntax follows Subject - Object - Verb (SVO) [3, 4].

For this reason, for Deaf people, knowing the spoken version of their language is considered as 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 [12].

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 hand configuration is the most obvious parameter in sign languages. It corresponds to the shape of the hand defined by the fingers. The orientation of the hand can be upward, downward, rightward, left, toward oneself and before oneself. The location is the position of the hand according to the body. The movement is the trajectory made by the hand. Finally, contrary to common belief, non-manual components play a huge role in sign languages. These inform the speaker about the speed of the action, size of the object or quantity. They also have enormous importance in interrogations, negations, conditional among others. Considering interrogations, Figure 2.1 presents both “We are going to class” and “Are we going to class?”, in which the difference between the signing of both lies in the end. The interrogative version presents a raise of the eyebrows of the signer. This facial expression is commonly used to indicate a question.

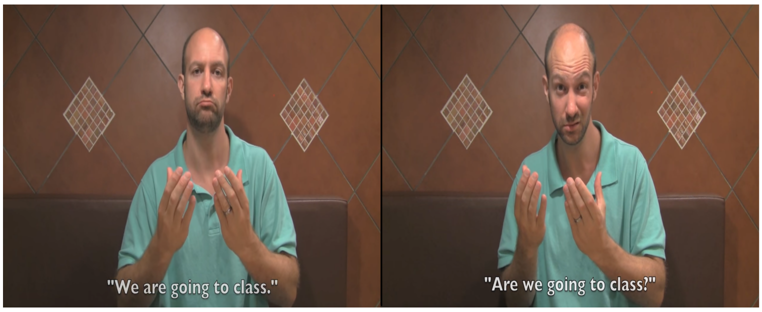


Figure 2.1: Non-interrogative phrase on the left and interrogative phrase on the right[1].

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. One of its most common uses is in names. However, fingerspelling is not considered part of the natural SL system.

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.

Lemmatization is the process of identifying and marking each word in a corpus with its base form. In an English corpus, this would involve, for example, stripping away inflectional morphology on verbs so that all forms of the lemma FORGET-forget, forgets, forgetting, forgot, and forgotten – would simply be marked as representing a form of FORGET. Lemmatization is in the base of glossing. A gloss is represented by capital letters of our own alphabet and do not provide a translation to SL, but instead an annotation where each gloss represents a gesture or sign. It is the result of applying lemmatization to a word. Glossing uses some transcription symbols to inform facial expressions, the number of repetitions of each sign, the need for fingerspelling (fs) among others. An example of an American Sign Language (ASL) gloss is shown in Figure 2.2 for the sentence “I finally found my keys”. The facial expression is defined (“eyebrows up”) alongside with the sign “PAH!” which means “At last! Finally! Success!”. Lemmatization is applied to the verb “found” resulting in its base form “find”. Besides, the first “ME” is

[1]https://www.youtube.com/watch?v=wwuL-hGIudg

indicated with "poss" which stands for the possessive pronoun, in this case, "my". In glossing, each sentence is written in a different line. This representation can be hard to understand since it does not provide any visual representation of the sign, but it is easier to process.

poss — eyebrows up
ME KEY ME FOUND PAH!

Figure 2.2: ASL gloss for the sentence "I finally found my key"[2].

Notations, on the other hand, use special symbols to describe the visual representation of SL. The first attempt in this field was with Stokoe notation [43]. The system described a sign through hand configuration (Dez), location (Tab) and movement (Sig). A serious deficiency of this notation is that it does not have facial expression, mouthing, eye gaze, and body posture.

Later, SignWriting was created. It is a practical writing system for sign languages, composed with a set of graphical and schematic symbols to represent signs [9]. The system tries to emulate the movements themselves. It is kept as a movement writing system which makes it intuitive and usable by common people with no special expertise. As long as the user learns this notation it is an easy way of textually representing sign language. It writes the way the body looks while signing. In contrast to the Stokoe notation, Signwriting allows for the representation of non-manual elements.

Another famous attempt in this field was HamNoSys. The HamNoSys for Sign Language describes signs mostly on a phonetic level [24]. HamNoSys, first version defined in 1984, consists of about 200 symbols covering the parameters of hand configuration, hand orientation, location and movement. These symbols are as iconic as possible and are easily recognizable. For a single sign, it describes the initial posture plus the action changing this posture in a sequence. HamNoSys is now in version 4 and has a new set of systems to encode non-manual behaviour, such as eye gaze, facial expression (eyebrows, eyelids, nose) and mouth pictures, in a degree of detail not previously possible. Some mouth gestures can be seen in Figure 2.3. An advantage of this notation is the possibility to be used internationally since it does not focus on specific national fingerspelling. This system is useful for people who use it frequently, for those who do not it might be hard to memorize.

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 softwares 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

[2] http://tvhsasl.weebly.com/uploads/3/7/5/1/37512505/introduction_to_glossing.pdf

	C01	cheeks puffed
	C02	cheeks and upper and lower lip areas puffed
	C03	cheeks puffed gradually
	C04(C)	one cheek puffed

Figure 2.3: Representation of some available mouth gestures [24].

which allow their non-commercial use.

3

State of the art

Contents

The past two decades have seen a great evolution in terms of signing avatars. The use of this technology can vary from teaching purposes [10], to linguistic studies [31], to communication between hearing impaired and hearing [36].

In section 3.1.1 we are going to present related work at synthesizing signing avatars through the use of annotations of sign languages. Section 3.1.2 and Section 3.1.3 will address approaches using motion capture to animate signing avatars.

3.1 Synthesis of Signing Avatars

In order to achieve an understandable signing avatar, various approaches have been developed and two major classes can be distinguished: Sing Language Annotation and Motion Capture, the latter can be distinguished between approaches with markers and gloves and markerless. All of these will be discussed in detail further ahead.

3.1.1 Procedural Synthesis from Sign Language Annotation

Sign Language Annotation is widely known in the Deaf community. The representation of Sign Language in the written form can be used for a generation when conversational agents are adopted.

XML is a meta-language allowing the definition of platform and application independent languages, dedicated to the storage and processing of information on the Web. To enable computer processing of SignWriting, a markup language was derived from it, SignWriting Markup Language (SWML) [9]. SWML is an XML-based language to allow the computer for independent representation of sign language texts written in SignWriting. In order to edit texts, there already exist some sign language editors such as SignWriter[1].

HamNoSys provides readable symbols to the human eye, but in terms of computer processing, it is not as straightforward. Therefore, using HamNoSys as a basis, other projects were developed. In European project ViSiCAST [17], a new version of the system was encoded in XML: SiGML and a translator from HamNoSys to SiGML was written. An example of both can be seen in Figure 3.1. This XML framework can be used so that the signing expressed in SiGML can drive avatars. Although this language was created in this project, they mainly focused on the creation of signs using motion capture, mentioned in sections 3.1.2. This project was mainly used for applications in broadcast, face-to-face transactions and World-Wide Web (WWW).

The ViSiCAST project was further developed and the eSIGN (Essential Sign Language Information on Government Networks) was created [45]. Since ViSiCAST mainly used motion capture, the eSIGN followed the work on the language previously created, SiGML, and focused more on generating tools

[1] https://www.signwriterstudio.com

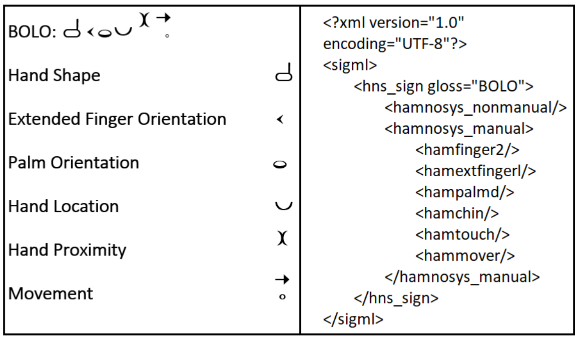


Figure 3.1: Representation of the Portuguese SL sign for "BOLO" (cake in English) in HamNoSys (left) and its codification in SiGML (right).

and contents. In this project, the animation of the avatar is solely done through synthetic animation. The signs start by being transcribed into HamNoSys. Afterwards, these transcriptions are translated into SiGML. Together with the description of the geometry of the avatar, animation data can be sent to the avatar, who will then take request signs. As for the creation of the content to be animated, tools are developed in which simple translations are contained or construct content from scratch. The former allows the translation from Dutch, English and German to their respective sign languages in a restricted set of phrases, phrases mostly in the domain of weather forecasts. The latter provides a platform for sign language users to create content based on the knowledge they have on their own language. The user can retrieve signs stored in a database, organize their order and provide the sign strings in the correct prosody. It can be compared to writing a text on a word processor. The first evaluation showed that the implementation of the non-manual part needed much improvement as well as changes in the appearance of the avatar. The purpose is to enable deaf people's integration into a key component of tomorrow's information society, eGovernment. For the user to be able to access this software s/he needs to install the signing avatar software. This is offered either as a free download or CD-ROM. Once the signing avatar is installed, the user is equipped to view virtual signing in whatever form a website provides it.

A barrier to consider with these notation systems is the facial expression of the avatar. As mentioned above, the actual version of HamNoSys already provides a more systematic treatment of non-manual signing which leads to the same feature in SiGML [15]. This study went in detail in the creation of these non-manuals gestures. The animation engine used to process SiGML to the animation of the avatar was Animgen. Animgen is the component of the software which translates signs described in avatar-independent Gestural SiGML into motion data for any avatar. In order to achieve this, each avatar comes with a set of facial deformations which are called morphs, applied to each frame of the animation. Each frame specifies how much of each morph is to be applied and the overall effect is displayed. To calculate which morphs are to be applied, each facial non-manual of SiGML is encoded into morph trajectories.

Morph trajectories consist of a morph, the maximum amount of that morph to be applied, the time in which this morph is to be increased from zero to that maximum, the time to hold it at that level and the time to decrease it to zero again. These trajectories can be used in parallel or series. Animgen will receive a configuration file which contains the mapping of each facial non-manual to its respective morph trajectories. The other non-manual components, such as trunk, shoulder and head movements, are animated by manipulating the bones of the avatar's skeleton. Each frame contains data describing a static pose of the avatar, together with a time-stamp specifying when the avatar should be placed in that pose. This pose has the configuration of the bones in the avatar's virtual skeleton and the configuration of the avatar's face. It should be understood that here "facial expression" covers only those expressive uses of the face which form part of the linguistic performance, that is, those which play a phonetic role in the given context. This excludes those facial expressions whose role is to convey the signer's attitude or emotions to what is being articulated linguistically.

Besides using gesture annotation for automatically synthesizing deaf signing animations, some approaches also use motion capture for the spine and neck to enhance the realism and provide "ambient motion" [29]. If the animation of the avatar was only based on the calculation of the joint angles for each gesture with a linear interpolation over time the effect would be robotic and unnatural. For this reason, the authors created an avatar with the ability to blend signing animation data with "ambient" motion such as small, random movements, mainly of the torso, head, and eyes, in order to make it appear more natural. For the animation of the avatar, from HamNoSys to SiGML, a signing avatar is generated by determining the joint angles required by each sign and setting the reference values for the joint controllers accordingly at the corresponding times. The blending of motion between signs is automatic, without requiring the avatar to go to a neutral position between signs. The software used to animate the avatar was H-Anim, which specifies a standard hierarchy of joints for a humanoid skeleton and a standard method of representing them in VRML. VRML, known as the Virtual Reality Markup Language, is a standard file format for representing 3-dimensional interactive vector graphics, designed particularly with the World Wide Web in mind. They started by creating a ball-and-stick model and since it is H-Anim compliant, to create a more realistic-looking version, they simply cut and paste from several available H-Anim avatars available online.

Another growing area in which virtual humans are used is in education for deaf students. There still is a lack of either skilled teachers or instructional materials in bilingual education. A Brazilian team targeted this problem and translated elementary school textbooks [11]. The software receives Brazilian text as input and based on a set of rules converts it into Intermediary Language. Intermediary Language is represented by glosses and defines a sequence of signs into animation commands which will be received by the Animation Module. This module has access to parametric sign description that contains the information on how to produce signs. The output is the animation of the avatar. The intelligibility scores of the avatar were equal to or greater than the rates for the video of the real interpreter.

Another approach in this field was developed by Efthimiou et al [12]. They created a Sign Language tool workbench which provides signer friendly graphical user interface. It focused on three major accessibility support features: fingerspelling keyboard, a bilingual dictionary look-up facility and the synthetic signing environment. The fingerspelling keyboard simply provides a set of keys for the GSL[2] alphabet. The look-up facility provides translation from text to sign language in the form of video. Finally, for the synthetic signing environment the user can create SiGML scripts either by entering HamNoSys strings of already stored signs or by creating HamNoSys lemmas online. The user may also add non-manual characters to the sign. Afterwards, the respective SiGML script is created, which can be stored. The animation of the avatar is primarily done through the JASigning software. This software uses Java and OpenGL for rendering and compiled C++ native code for the Animgen component, which converts SiGML to conventional low-level data to the 3D animation. On the whole, the platform was positively scored by the users. The option to compose one's own chosen utterances received high scores and its importance was recognized in the education field. The bilingual dictionary was one of the features that received particularly high scores.

3.1.2 Data-driven Synthesis from Motion Capture with markers and gloves

Data-driven animation methods, or motion capture, 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 linguistics purposes, it is important to understand details such as timing, co-articulation, spatial references used, non-manuals or inflexions phenomena which operate during a SL discourse. With this goal in mind, a project was developed for the American Sign Language [32]. The corpus was obtained with great detail and the signer used expensive equipment such as two cyber gloves, an eye-tracker, a bodysuit and an InterSense IS-900 to calculate the head movement. Besides these devices, they also used three cameras to film the signer from different perspectives: front view, facial close up and side views.

Projects were also developed with the intent of human-humanoid interaction to engage in real-time dialogue, SignCom [22]. The user signs towards a camera by which the system recognizes the signs and gives a culturally and linguistically acceptable response through a virtual agent. The animation of the avatar was captured by twelve cameras, forty-three facial markers, forty-three body markers and twelve hand markers. However, post-processing operations were necessary as well. Each body part was

[2] Greek Sign Language

associated with a channel (lower body, torso, arms, hands, head, face and gaze) and these channels must always be synchronized. The systems have two databases: semantic, for textual annotations, and raw motion database, for motion capture. The semantic database, containing textual information from the annotation process, is queried first. The motion data corresponding to the obtained results is then streamed to the motion composition process. The semantic database is one-to-many since the same annotation can have different motions, corresponding to different contexts. In the raw motion DB, we have the decomposition of the sign in channels. The corpus contains three thematic scenarios: the Cocktail monologue, the Galette and Salad dialogues. In this approach, both body and facial data were recorded at the same time which turned out quite noisy. During their testing, they concluded that the users did not take much attention towards the eye gaze, perhaps because the eyes were too small in the video to be noticeable. The users pointed out the lack of expressiveness and comprehension of the avatar since some times finger contact did not happen when needed. Just like with the SignCom project, the project developed by De Martino et al [11] also used motion capture to construct the sign language corpora. This corpus is then used for handling the bone-and-joint system retrieved from the XML in order to emulate hand, arm, body, and facial movements more naturally.

Since all the equipment for these approaches of motion capture is expensive, some systems were created in order to deal with this problem [8]. Chouhan et al. created an intelligent and affordable system comprising a wired glove interface which translates the gestures performed to textual messages or speech. The advantage is that it implements a low cost wired interactive glove using sensors more easily available to people, such as bend sensors and accelerometer. However, it is still difficult to implement if not by specialists. Furthermore, this glove only focus on the manual aspects of signing and no information is provided on the facial expression of the signer.

Besides the capture of signs, the design of a corpus in the framework of Motion Capture (MoCap) editing system is one of the key points of the approach since the consistency of the synthesis depends on its completeness [23]. This article presents the main requirements for the whole pipeline design of interactive virtual signers. A similar design was used in the SignCom project which was developed by the same author. For the creation of a corpora, its design can either focus more on the depth, limited signs with many variations, or on the breadth, lexicon that covers a broad area. This choice will depend on the objective of the editing system. It is important to have enough variability for each sign in order to build new consistent utterances for each sign by re composition of already recorded motion chunks. As for the resources used it is important to have markers placed on the whole upper body, including face and fingers. The annotation process is the heart of SL editing process. For this process, similar to SignCom, several channels are created following a phonological/phonetical and syntactic specification scheme. Sentences are then temporally segmented into glosses along channels and labeled by a string conveying its meaning. Other relevant features are annotated (e.g. handshape, face expression) to choose the appropriate sign variant when new utterances are created. Since the structure in SignCom was successful they suggest a similar database to be divided into two components: raw database, which contains raw motion data, and a linguistic database, which contains high level linguistic descriptions and returns raw info, which the raw database will receive. The principle of the dual database and the request mechanism can be represented in 3.2. Once the database is populated, the creation of the animation can be done by directly extracting from it or by editing and composing new movements.

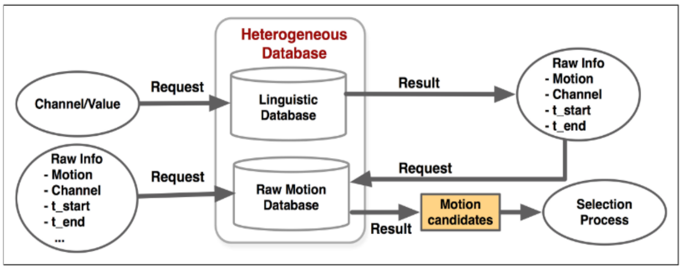


Figure 3.2: Design of the database [23].

All the corpora designed in previous projects, such as HuGEX, SignCom and Sign3D, are described in a more recent paper [20]. HuGEX had two databases: TRAIN database which aimed at providing sentences with predefined replaceable parts and METEO database which aimed at studying the variation in the prosody of the LSF phrases. The Sign3D is an improvement of SignCom. In order to achieve this, they used a higher number of markers. After all the developed projects they were able to achieve a complete concatenated data-driven synthesis pipeline that enables the assembling of motion elements, from signs and parts of sentences to motion chunks retrieved from different channels and body parts (hand movements, hand configurations, body movements, facial expressions, and gaze direction), representing phonetic or phonological components. The purpose was to share motion capture databases with different communities.

As previously mentioned, the ViSiCAST project [17] relied mostly on motion capture using sensors for the hands, body and face for a simplified system.

3.1.3 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 and hand motion tracking. Kinect is one of these cameras. It was initially produced by Microsoft as a gaming accessory for Xbox consoles, however, this device is also famous in the field of research. From this equipment it can be extracted two types of data: the raw data returned by the depth camera and a ”skeleton stream”. The latter can recognize joints in the human body and return real-time motion tracking data. Two versions of this sensor are available, Kinect v1 and Kinect v2. In a recent study [44] authors concluded that the latter is not as widely used as the former, without any apparent reason. Therefore, they assumed that the pros of the first version outweigh the pros of the second: the depth precision is higher, environment color does not affect depth estimation and it has a weak correlation to the temperature. Although Microsoft discontinued the Kinect sensors at the fall of 2017, other similar devices are still on the market, such as the Intel RealSense.

As mentioned above, Kinect became very used in the research field. Studies were conducted using this device to interactively control an armature by tracking body poses [39]. The main contribution of this work is the devising of an interface for animators to associate in real-time natural human movements to both human and non-human virtual characters by using affordable and flexible hardware. The study revealed that the keyframing approach was preferred to animate non-human shaped objects which may be associated with the fact the longer training phase necessary. An advantage pointed by the users was the speed of the envisioned motion capture system and they would prefer to use the Kinect based approach as a supplement rather than only keyframing approach because the overall procedure of creating character animations would be accelerated. Some disadvantages mentioned were the lack of accuracy sometimes in tracking since occluded parts are not tracked and for this reason, manual corrections are often needed.

Considering the feature of skeleton tracking, Kinect is also used to create Sign Language avatars [35]. This study is based on the techniques to develop using this sensor to automatically track the hands' movement and facial gesture to aid in communication between the signer and non-signer. Kinect showed to be able to detect more than one action at the same time as well as facial and voice recognition and to reconstruct and display the joint skeleton in the virtual environment and surroundings.

Portuguese works were also developed [18]. The project developed in this thesis allows an introduction of new signs without advanced knowledge in the animation area. It provides the user with either an automatic creation of gestures or manual creation.

The former is done through a Kinect device and voice detection. The sensor was configured to only detect from the waist up and the user must start in a T-pose so that the device knows which person to detect with mixing other movements of other people that may be present. The application also provided the user with the option to visualize all the animations already available in the system. Kinect does not capture the hands of the signer, this must be done manually.

Given its easy use, this device expanded to the creation of Sign Language dictionaries. In Brazil, a dictionary animated by an avatar was created and is available on more than 1500 websites [42]. This project presents a semi-automatic construction of this dictionary. The process combines automatic tasks, such as motion capture, in this case, a Microsoft Kinect v2 device, with manual tasks involving 3D animators and sign language specialists. The creation of this dictionary started with a data survey of signs in deaf communities and its respective video and motion capture. After this search and using the motion capture obtained 3D animators specialists animate the virtual human with the videos as references if required. After a meticulous inspection of the signs created by a linguist, specialists focused on possible mistakes in the hands' configuration, facial expressions and articulation points between hands and body the signs can be incorporated in the dictionary.

Following the work developed in the eSIGN [45], another EU funded project was developed. This approach focused on crowdsourcing Sign Language dictionaries. The DICTA-SIGN [14] project researched ways to enable communication between Deaf individuals through the development of human-computer interfaces. The user signs to a Kinect device providing input isolated signs or continuous signing. The computer recognizes the signed phrases, converts them into an internal representation of sign language, and then has an animated avatar sign them back to the user. Content on the Web is then contributed and disseminated via the signing avatar. This dictionary also has a feature to look-up signs that provides simple sign-level translation tool for exploring corresponding signs in the four project sign languages. In order to achieve this the user perform sign towards the Kinect and the look-up tool plays back the recognized sign or the closest matches.

Just like technologies mentioned in the previous section, Kinect can also be used for teaching purposes [6]. "Os meus primeiros e-Gestos" is a bilingual dictionary. The creation of content was done by taking advantage of the depth camera feature of the sensor. The look-up for vocabulary is done through the Portuguese Language and returns its equivalent in PSL. The interpreter is an animated avatar and the application contains themes like colours, animals, professions and cities always followed with illustrating images. "Adivinha gesto" is a teaching strand also available in which from a certain amount of signs the user has to choose the right for a specific word. It is also possible to change the avatar's appearance according to the user's preference, a slow-motion visualization of the gesture in high definition and a portal exclusive for administrators.

Leap Motion Controller is another well-known device in the research field due to its hand tracking feature. This device is smaller than Kinect, connected 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. Leap Motion sensor overcomes the major issues in real time environment like background, lightening condition and occlusion. As a result of these features, this sensor is also used for Sign Language communication purposes. Leap Motion was used for Sign Language recognition [33] for American SL, Chinese SL and Irish SL. Some samples for different languages as they appear on the Visualizer tool of Leap motion sensor are shown in Figure 3.3. They created a database using several users belonging to different ages, sex and region. The dataset was trained and tested and all the languages showed an accuracy higher than ninety per cent.

Also in the field of recognition of gestures, in order to test the suitability of the device Leap Motion, some exploration was executed for Australian Sign Language [38]. The use of this system comes from the need for a simpler technology such that a family with deaf children can easily access. This device

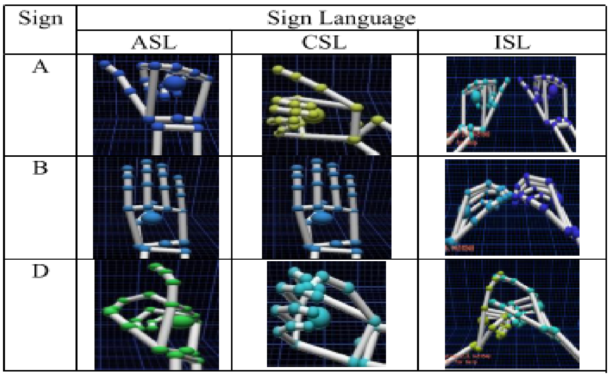


Figure 3.3: Sign samples on Visualizer of Leap Motion [33].

does indeed provide an accurate level of detail since it pre-processes which leads to a faster build and accuracy. It is also able to capture very small movements. Nevertheless, if the hand is not in the direct line of sight of the controller it shows some difficulty in maintaining accuracy. Besides, some positions of the hand cause the controller's detection to deteriorate entirely when the palm is perpendicular to the flat surface. In cases like these, the lack of pre-processing of the data becomes an issue since the hand data is non-modifiable so there is no easy way to correct the detected data. Possible solutions would be to infer the location of the fingers but this would lead to a delay in real time feedback.

Due to its small size, Leap Motion is considered a discrete device. For this reason, it was used to create an application which focuses on capturing and recognizing signs [36]. The idea behind is to provide deaf people with an autonomous way to communicate with everyone without barriers, therefore giving voice to those who do not have it. Once again, post-processing of the data was required. The device presented values above seventy per cent for static gestures. Although the nominal accuracy is of eighty-two per cent, it showed difficulty in maintaining this value while the hand moved.

The two sensors provide features that complement each other. The Kinects 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 [44]. Therefore, studies came with approaches to combine the two devices together [34]. When extracting data from Leap Motion, the device does not return a complete depth map but only a set of relevant hand points and some hand pose features. Besides, the sensor is not always able to recognize all the fingers. Not only fingers touching each other, folded over the hand or hidden from the camera viewpoint are not captured, but in many configurations, some visible fingers could be lost, especially if the hand is not perpendicular to the camera. Although Leap Motion provides a higher level, the data description is more limited. On the other hand, Kinect provides the full depth map. The more complete the description provided by the depth map of the latter is, the more probable it is to capture other properties missing in the former output. For this reason, combining the two devices a very good accuracy can be obtained.

Researches took advantage of how good the two devices work as a team and used them together for Sign Language [25]. This paper describes an effort towards an online collaborative framework allowing Deaf individuals to author intelligible signs using dedicated authoring 3D interface. They used Kinect Faceshift for 3D facial reconstruction and Leap Motion to capture the position and orientation of the user's hands. FaceShift can be used in real-time tracking mode, where the signer's face expression framed by the Kinect is analyzed and used to pilot the avatar's face. To test this feature, users were asked to identify emotions captured with the program and the results were positive. As of the usage of the framework the user can record and edit characters using either pose-to-pose, in which it is possible to adjust the poses of the character by manipulating it as a puppet, or performance capture (puppetry), where the user has direct control over the virtual character movements. The overall architecture of the authoring system is summarized in Figure 3.4. They concluded that with this system, a user without prior computer animation experience is able to perform non-trivial positioning and posing activities as fast as an experienced user would do using a traditional keyboard and mouse interface. When handled by an experienced user the velocity is doubled in comparison with keyboard and mouse input.

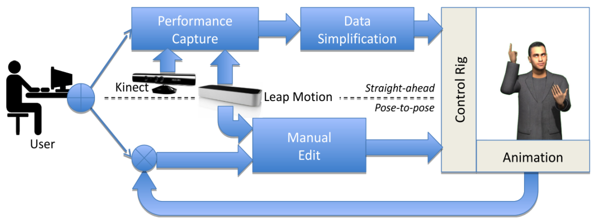


Figure 3.4: The authoring pipeline overview [25].

3.2 Applications and Design of Signing Avatars

The previous sections presented various technologies available to animate signing avatars. However, the impact that these avatars can have in the Deaf community is not solely based on the functionality but also in other features. In this section, we take a closer look at what the future users of this system actually expect from it and the most desired features. Moreover, we intend to understand how the design of the avatar can affect the user's performance.

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. In order to focus more on the avatar performance, first one has to measure the comprehensibility of an avatar. For this reason, to properly access the users' perspective on virtual humans, studies were performed where the main focus was the participants' take in virtual humans [30]. The aim was to access the signing avatars acceptability, shortcomings and potential use cases through focus groups and online studies with the Deaf community as the target audience. Eight people participated in the focus group. In this focus group, the participants could discuss signing avatars in a non-threatening environment during a dual-moderator session. On the other hand, three hundred and seventeen people performed the online study. This study was open to the general public but was only advertised within the Deaf community. The results achieved in both studies were similar. First of all, the mere participation in the user studies increased the acceptance in this technology to a measurable degree. During the focus group phase, by order of importance of the avatar's features voted, we have facial expression, natural movement, mouthing, emotions, body motion, appearance, synchronization of sign and mouthing, charisma and comprehensibility. A curious discovery was that the users showed to prefer a purely hand-made cartoon animation (DeafWorld) to automated avatars (The Forest and Max). Although DeafWorld is a less realistic looking avatar given it is a cartoon, it is hand-made which makes it less stiff and robotic. Nevertheless, the participants emphasized the need of having different avatars for different domains: while a cartoonish character would be suitable for children and entertainment, a more realistic adult avatar is recommended for the use in serious applications. Participants do find a use for these animations in of one-way communication situations. Educational contexts, social network websites and administration pages were the top choices areas for application. It is important to note that an expressed worry during this study was regarding the potential replacement of human interpreters by avatars.

As mentioned in previous studies, the facial expression is a major feature in avatars, mostly signing avatars. This feature not only provides expressiveness to the avatar but in sign language it also provides meaning. Therefore, researches focused on evaluating the contribution of this feature in terms of comprehension and acceptability in sign language avatars [40]. Two avatars were chosen: Anna, a more human-like avatar conceived during the ViSiCAST [17] project and further developed in the eSIGN [45], and Luna, a more caricature avatar, see Figure 3.5. *A Scare in Belfast* was selected and identified as having a high level of Emotional Facial Expressions (EFEs) content since all the seven emotions were present (happy, disgust, anger, fear, contempt, surprise and sad) and was recreated in animated avatar's videos. In total four versions of avatars were used: Anna and Luna with baseline encoding and AnnaE and LunaE with EFE encoding. The participants saw each video two times. The primary focus was to ascertain whether or not the addition of emotional facial configurations increased the understanding ability of a signed utterance, which the results surprisingly showed no. For the baseline avatars, the addition of emotions made very little impact on the score. It is also evident that higher comprehension levels are achieved with Anna and the emotions are more easily understand in the AnnaE avatar. Some participants did mention that Luna's long fingers worked well and Anna's face is better suited to deliver facial expression. Once again it was accentuated the need for a repertoire of avatars to be available for various tasks, an avatar like Anna for more serious tasks and Luna is better suited for children. Participants do not think the avatars looked natural and are missing an appropriate synchronization of manual and non-manual feature and timing at the sign level, particularly for fingerspelling. Overall the study shows that the avatars are an applicable technology that has not yet evolved to a point for mainstream use.

Figure 3.5: Avatars Luna, on the left, and Anna, on the right [40].

By now we are able to understand the importance of non-manual components on sign language. Facial expressions and head movements communicate essential information during ASL sentences. Nevertheless, even when these components are indeed implemented in the avatar its understanding is not always linear and users face some difficulties during the process. In order to address this problem, it was created stimuli and questions to properly measure whether the viewer has correctly understood linguistic facial expressions [27]. A stimuli design is the process of creating animations, short stories and comprehension questions which will affect the scores collected in a study. To achieve this, two alternative methods for stimuli design were presented: English to ASL or ASL originated. The primary difference between the sets is the degree of involvement of a native ASL signer in the creation process and the categories of facial expressions included in each stimulus set. The latter will provide more fluent ASL sentences while the former began with an English sentence followed by its respective translation. Similar to [40] each animation was produced with and without facial expression. The animations were categorized in ”Emotions” and ”Non-emotions”. For English-to-ASL stimuli, in the emotion category, adding facial expressions led to significantly higher comprehension scores. However, there was no benefit from adding facial expressions for the non-emotion categories. Therefore the authors question if perhaps something was lost in translation in this stimuli. For the ASL-originated stimuli, adding facial expressions led to significantly higher comprehension scores for both emotion and non-emotion categories. Overall, designing stimuli in English and then translating them into ASL was not an effective methodology for designing a sign-language facial expression evaluation study. Furthermore, it is possible to conclude that the involvement of native ASL signers in the stimuli design process is important in achieving a high-quality result.

A more recent study from 2018 selected numerous of articles to better understand the aspects necessary for the avatar's design, as well as essential requirements for the virtual environment [19]. As for the virtual environment, they obtained a set of requirements such as a simple interface with easy usability and good communicability. It is also important to explore the use of images and the use of an avatar should be of easy accessibility of the system to the user. Regarding the requirements and characteristics to be taken into consideration during the creation and construction of the avatar they determined as the more relevant the visual of the avatar. The more similarities it has with the user and the more human-like it is, the more probable the avatar is to be accepted. Artistic details such as the hair and skin colour as well as the anatomy of the avatar, body, weight and size, also influence the human aspect of the virtual character. Evidently, the movements performed by the avatar should be as similar to human movements as possible. As for the figure type of the avatar, some researchers defend that having a Self-Avatar is important and can improve the user's confidence in the application leading to the faster and more precise execution of tasks.

In addition to all of the above features mentioned, the pauses and speed of signing have also major impact in the comprehensibility of a sign. A software was developed [2] to partially automate the synthesis of animations in ASL. The goal is given a sequence of ASL words in a message, to be able to automatically identify the pauses, the time-duration of each pause and the variation of speed for each particular word. Minor variations in these three parameters can lead to significant differences in the users' understandability of the animations. They implemented three machine-learning models that worked in cascade, the output of the prior would be the input of the next. The first model was the Pause Insertion which was a classification model to rule if a pause should be inserted after the current word. The second was the Differential Rate which was a regression model to predict the change in signing rate within an ASL sentence. Finally, the Pause Duration which was also a regression model to predict the length of each pause. To evaluate the performance of the system they compared three systems: a baseline system which inserts pauses at sentence boundaries only with uniform pause length and uniform sign speed, 2008 Model which is the current state-of-the-art model for speed and pausing in ASL and ASL-Speed which is the system created. The ASL-Speed outperformed both models. Further tests were performed with Deaf or Hard of Hearing participants with ASL speed and the baseline model. Six out of eight participants preferred the ASL speed animations since it presented an overall good pace, good and correct pauses and regular signing speed.

3.3 Discussion

Signing avatars still need improvements on its various features and it still does not exist an avatar accepted by the Deaf Community. A perceptible avatar needs clear manual gestures with no fingers' occlusions and a facial expression synchronized with these manual gestures [40]. Besides, the correct signing speed is also crucial. It should neither be too fast that becomes imperceptible nor too slow that conveys the idea of a beginner signer [2]. The use of pauses in the appropriate duration and correctly inserted are also essential to an understandable signing.

In the previous sections, different technologies used for the creation and control of signing avatars were described. In order to better compare and evaluate these systems, two tables were conceived: a qualitative evaluation, Table 3.1, and a quantitative evaluation, Table 3.2.

	(Need for) Post-processing	(Need for) Prior calibration	(Need for) Prior knowledge	Manual-gestures
Gesture annotation	X	X	✓	✓
Markers and Gloves	✓	✓	✓	✓
Leap Motion	✓	✓	X	✓
Kinect	✓	✓	X	✓
Kinect and Leap Motion	✓	✓	X	✓

Table 3.1: Technologies for the creation of avatars, qualitative evaluation.

Table 3.1 evaluates the existing technologies to synthesize signing avatars mentioned in our investigation, in a qualitative manner. The lines of the table present the technologies to be discussed and the columns the features we consider most relevant. Gesture Annotation has strong advantages when it comes to working the data. It neither requires post-processing of the data nor the prior calibration of the software with the user. Nevertheless, it does require prior knowledge since the user needs to know the annotation in use, either HamNoSys [24], SignWriting [9] or any other. Markers and Gloves demand high technical skills for recording and post-processing data. This post-processing presents challenges such as the reconstruction of the skeleton and the adaption of the data to the 3D avatar [23]. Besides, prior calibration with the user and its respective markers also has to be executed. Finally, Leap Motion, Kinnect or Leap Motion and Kinnect together demand not only for post-processing of the data but also prior calibration of the devices. However, no prior knowledge is needed since the interfaces available are meant for non-experts [39]. All the technologies provide manual gestures, as expected.

Table 3.2 also evaluates the existing technologies to synthesize signing avatars mentioned in our investigation but in a quantitative manner. The lines of the table present the technologies to be discussed and the columns the features we consider most relevant. We can start by observing that both Gesture Annotation and Markers and Gloves focus on real specific levels of expertise. The former requires linguists while the latter requires animators and people with knowledge in motion tracking technologies in order to know how to properly place the tracking sensors. On the other hand, Leap Motion and Kinect can be used by people with basic computer knowledge [25]. As for the comprehensibility of the avatar, Markers and Gloves provides the best results since it is able to capture a great level of detail [32]. Leap Motion and Kinect return an animation with a reasonable level of comprehensibility and

	Expertise	Understand-ability of the animation	Commod-ity	Cost	Facial Ex-pressions	Flexibil-ity
Gesture annotation	Linguists	*	***	*	*	**
Markers and Gloves	Animators and people with knowledge in motion tracking technologies	***	*	***	***	*
Leap Motion	Anyone with basic computer knowledge	**	**	**	*	*
Kinect	Anyone with basic computer knowledge	**	**	**	**	*
Kinect and Leap Motion	Anyone with basic computer knowledge	**	**	**	**	*

Table 3.2: Technologies for the creation of avatars, quantitative evaluation.
- Doesn't possess *Barely possesses **Possesses reasonably ***Possesses a lot

Gesture Annotation presents the worst results in this field since it has shown to return avatars robotic and stiff [29]. This feature can be related to the facial expression offered by each technology. Since Gesture Annotation returns less detail in the avatar's facial expression the final result will also be less natural and less understandable.

Taking into consideration the commodity provided whilst using these technologies differs. Gesture Annotation presents the best results since the person creating the signed content can do this at his/her own pace with no dependency on the availability of a studio or a camera person [45]. In contrast, Markers and Gloves provide no commodity. It not only requires access to a proper environment but also to proper and very specific devices. Leap Motion and Kinect provide reasonable commodity since the user does not need to perform the signing in any specific environment, s/he can perform it for instance at her/his home. The cost also presents the best results with Gesture Annotation since the software provided by this approach can be easily downloaded with no cost to the user [45]. As one can figure, given the specific material needed for Markers and Gloves, this Motion Capture approach needs great investment from the user. In the majority, the devices needed are not affordable. Leap Motion's prices round the eighty dollars and Kinects' the one hundred dollars, yet, when used together it becomes more expensive and difficult to acquire both devices since the price adds up [44]. Finally, Gesture Annotation obtains better results in flexibility in comparison with motion capture approaches. Editing the gestures created in motion capture either requires more recording or modification of already existing similar signs whereas Gesture Annotation only needs changes in the generator algorithm.

Overall, some conclusions can be withdrawn. Motion Capture with markers and gloves presents an expensive approach for the creation of signing avatars. Although it is the technology which provides the most natural movements of the avatar, it is not affordable and only specialized laboratories will have the devices required for these recordings [17] [32].

In comparison, markerless motion capture technologies are more affordable, however, its cost depends on the chosen approach. A signing only focused on manual gestures can be implemented using a Leap Motion device [34] [38] [36]. A signing focused on facial and body expressions can be achieved using a Kinect device [18] [6] [42]. Either option provides a reasonable price. Nevertheless, while joining both devices' features does provide better results, the price increase makes it a less affordable solution [25].

Two great factors for the growth of these technologies is the expressiveness and human-likeness of the avatar. Motion Capture does provide a more realistic animation at the expense of more costly devices. Besides, when it comes to Motion Capture approaches there is a substantial amount of work involved in setting up and calibrating the equipment, as well as in recording a large number of signs required for a complete lexicon. To build a sentence every sign must be captured. In case new signs need to be added to the corpus, it is also a time-consuming task since there is no easy way to do it. Either the recording is performed again or the modification of existing gestures.

There are also many unresolved challenges with the need to simultaneously record body, hand motion and facial expressions.

Gesture Annotations, on the other hand, provide an easier and more comfortable option. The user only depends on access to a computer. Moreover, repairing mistakes and adapting signed text is rather easy in comparison to making and editing movies, since necessary changes are simply applied through text. Gesture Annotations leverage the expertise of users with knowledge in existing notations in the creation of signing avatars.

The stiffness and more robotic movement of the resulting avatar with this approach can be improved through several features, such as the timing of the signing, and provides an understandable avatar. All in all, Sign Language synthesis is a competing technology that is less costly, more versatile and may prove to the answer to the current lack of access for the Deaf in computer interaction.

4

From Annotations to a Signing Avatar

Contents

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.

In this chapter we will go over our approach in Section 4.1, we will also give an overview of the corpus and the team of our project with Section 4.2 and the overall architecture of our tool in Section 4.3. To better understand the implementation of our work, we will provide a brief explanation of the notation in use, Section 4.4. Following, in Section 4.5, we explain the conversion from the symbolic representation used and its machine-readable XML-component. Section 4.6 describes the major component of our tool, the animation of the avatar through the use of SiGML. We describe the tool's integration with PE2LGP [18] in Section 4.7. Finally the limitations of our prototype as well as suggestions on its improvement in Section 4.8.

4.1 Approach

The focus of our approach 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 tool will be based in synthetic animation, explored in Section 3.1.1. Our choice was based on the advantages provided by this approach in terms of flexibility, commodity and cost, discussed in Section 3.3. It has presented good results in previous works, such as [45], where the possible lack of realism of the returned avatar makes up on the accessible manner provided to the user to work with the signing avatar.

To take advantage of this software the user will only need access to a computer with the required software installed and knowledge in the notation to be used, in this case HamNoSys.

4.2 Corpus

The annotations in HamNoSys as well as its glosses was provided by the team of Universidade Católica. This team's focus is to create a corpus in Portuguese sign language. The team consists of six members, three members with linguistic knowledge in PSL, two deaf PSL experts and one interpreter of Portuguese and PSL. 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 project has twenty-three hours, fourty-five minutes and thirty-nine seconds of transcriptions, with one-hundred-fifty-six signs annotated in HamNoSys.

The annotation tool used for this purpose is ELAN [1], mentioned in Section 2.3. The annotations have

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

information considering the correct translation to Portuguese, the glosses of each sign, their respective grammatical classes and finally its HamNoSys. The exported file extracted from ELAN contains the information about all these different tiers.

4.3 Architecture

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, explained in the following Section 4.5. Afterwards, the content can either be visualized or saved, as illustrated in Figure 4.1. This architecture will be analyzed in further detail in the following sections.

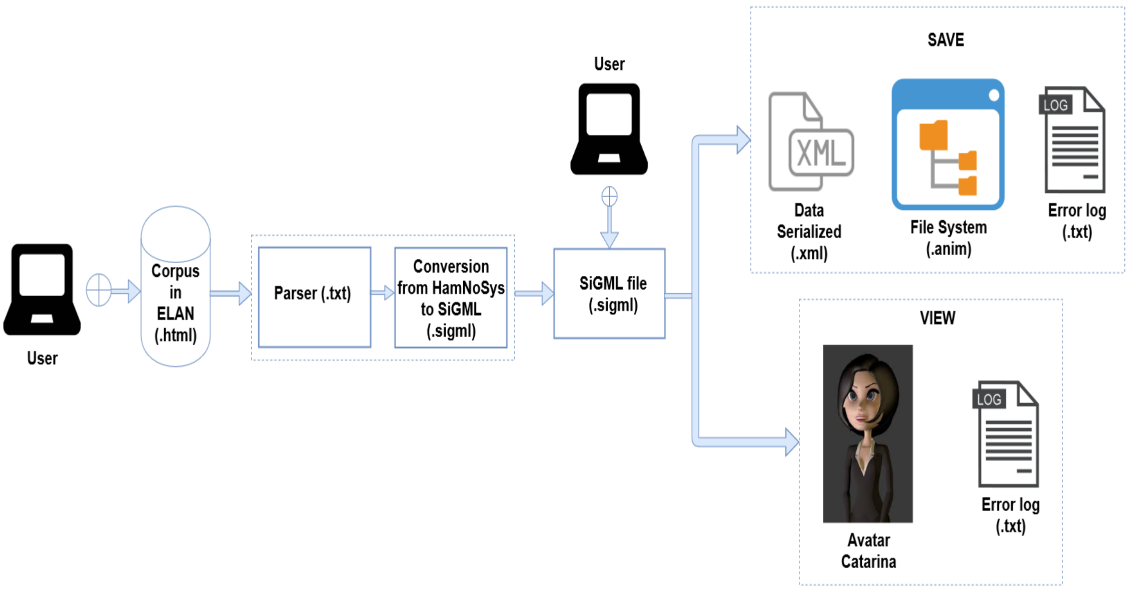


Figure 4.1: Tool's architecture.

4.4 HamNoSys

First of all, to better understand the work developed during this project it is important to have a deeper knowledge of the annotation in use, HamNoSys. As briefly described in 2.3, HamNoSys is an alphabetic system describing signs mostly on a phonetic level. This notation system is a combination of iconic and easily recognizable symbols which cover the parameters of hand shape, hand orientation, location and movement. HamNoSys can be internationally used since it does not rely on conventions that differ from country to country [24, 28]. HamNoSys was the notation chosen by the team of linguists from Universidade Católica to annotate the content.

4.4.1 Hand Shapes

HamNoSys has twelve basic hand shapes. These can be arranged into two groups of six hand shapes: open hand shapes, showed in Figure 4.2, and thumb combinations, showed in Figure 4.3. Nevertheless, these can be further combined with three type of thumb configurations and three type of bending the fingers.

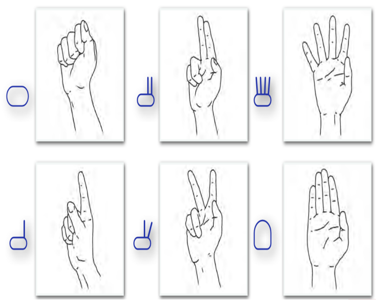


Figure 4.2: Open hand shapes configurations.

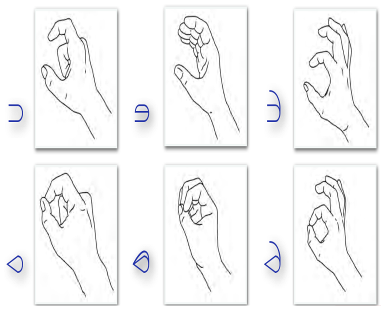


Figure 4.3: Thumb combinations configurations.

By default this bending will be applied to all the fingers. However, for more detailed configurations it is also possible to specify which fingers to bend as well as to specify eventual touching or crossing between fingers. Each finger is defined by six parts: finger tip, finger nail, finger pad, middle joint, base of finger and side of finger. Further detail can also be added with specific thumb locations. This leads to a considerable amount of combinations possible to create with this notation, in Appendix A are defined some. The configurations implemented in this project are highlighted.

4.4.2 Hand Orientations

The description of the hand orientation is composed of two components: extended finger and palm orientation. The former corresponds to the direction in which the knots of the fingers are pointing, in Figure 4.4. The latter is related with the former. For a given extended finger, it indicates the orientation of the palm around the shaft of the hand, in Figure 4.5, in which the dark side of the symbol represents the palm. HamNoSys has a variety of eighteen symbols for extended finger and eight symbols for palm orientation. For both components it is possible to create combinations between two symbols of each category.

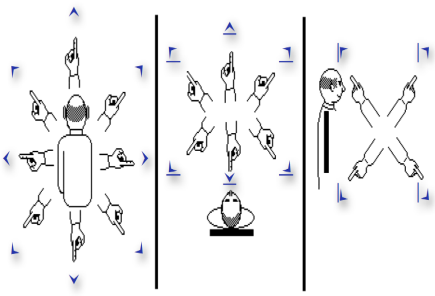


Figure 4.4: Extendend finger directions.

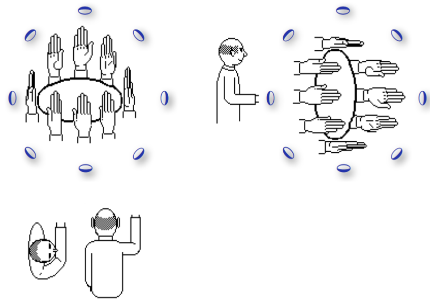


Figure 4.5: Palm orientation values.

4.4.3 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, whilst the second determines the distance of the hand to this location, in Figure 4.6 the last symbol represents touching whereas the black box closeness. If the latter 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. Overall, HamNoSys has forty one basic symbols to define the location, in which seventeen refer to parts of the hand and finger. These can, however, be combined with others for more detailed locations. For instance, the location in front of the chest can be combined to provide more precise information in terms of distance, closer or even touching.

4.4.4 Hand Movements

Movements in HamNoSys can be distinguished between absolute and relative movements. The former describe movements with a specific target location. The final location is known beforehand. Such movements are defined through the change of locations within the same sign and are characterized neither by a size nor a direction. In Figure 4.6 is an illustration of such movement, the initial position is close to the left shoulder (symbols before the arrow) whereas the final position is touching the right cheek (symbols after the arrow). On the other hand, the latter describe movements defined by their direction, size and the end point will be the final position of this movement. An example is illustrated in Figure 4.7 which describes a straight small (empty ball under the arrow) movement down from the top of the head.

Relative movements can be distinguished between straight, curved and zigzag lines, circles and similar forms. These can either be performed sequentially or co-temporally. Within relative movements, it is possible to define movements with no change in the location, such as finger play, which only moves the fingers. Repetitions and speed of movements can also be defined. In the case of two-handed signs, it is possible to differentiate the actions for each hand.

Figure 4.6: Absolute movement.

Figure 4.7: Relative movement.

4.5 HamNoSys to SiGML

As mentioned in Section 4.3, the first part of this project consists on the processing of the data obtained from ELAN, which consists on step "Corpus in ELAN (.html)" until step "SiGML file (.sigml)" in Figure 4.1. 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.

4.5.1 Pipeline

Describing in further detail the step step "Corpus in ELAN (.html)" until step "SiGML file (.sigml)" from the architecture illustrated in Figure 4.1, we have a more detailed pipeline of this system presented in Figure 4.8.

Various versions of this system were developed from reading the input from ELAN, to versions which simply read the input from the command line. The main difference between these are the parser which evaluates the input.

4.5.2 Parser

Firstly, an HTML file of the content to be processed must be exported from ELAN, example of this file in Appendix C. The parser, correspondent to step "Parser" in Figure 4.1, will evaluate this file as input. The HamNoSys symbols received initially cannot be directly read; for this reason, they must be converted into machine-readable content. These symbols are available as an Unicode font with the characters mapped into the Private Use area of Unicode.

The Private Use area of Unicode is a range of code points that are intentionally left undefined so that third parties can define their characters without conflicting with already existing Unicode characters. In this case, we use HamNoSys symbols. Each symbol is associated with a code of exactly four characters. The correct association between both the symbol and its corresponding code is performed by the system in step "Convert HamNoSys symbols to their Unicode codes" (Figure 4.8). We provide an example in Figure 4.9, in which the values in the column *Symbols* are the input received by the program, and its respective code is correspondent to the values in the column *Codes*.

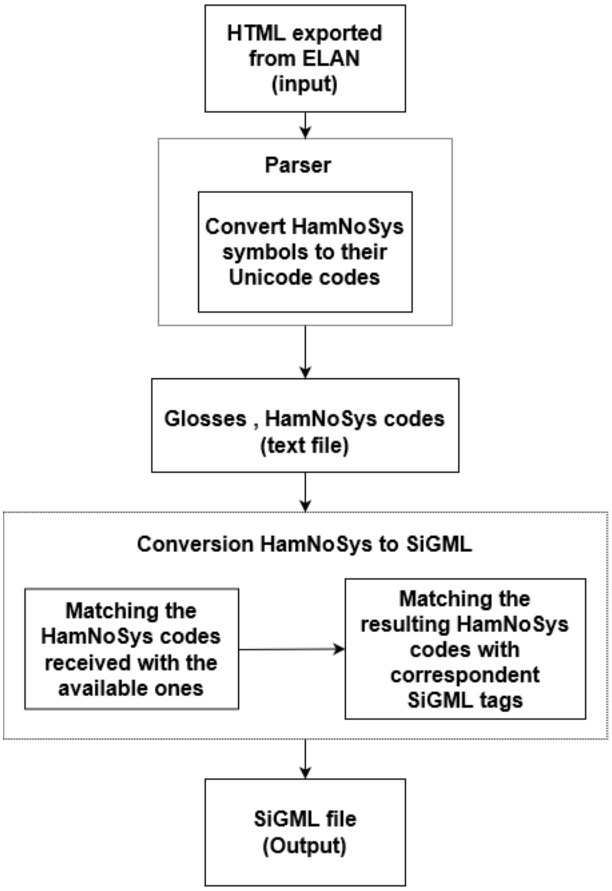


Figure 4.8: Pipeline of the system.

Figure 4.9: Correspondence between SiGML tags, HamNoSys Unicode codes and HamNoSys symbols, in the respective order.

4.5.3 Conversion from HamNoSys to SiGML

The system has access to approximately 210 HamNoSys symbols and their corresponding SiGML tags, provided by the authors of SiS-Builder [12], available in Appendix B with the implemented symbols highlighted. Once all HamNoSys codes received are converted to their respective set of Unicode characters (example in Appendix D) it is possible to match them with those accessed by the program ("Matching the HamNoSys codes received with the available ones", Figure 4.8).

Afterwards, the correct SiGML tag is accessed, which corresponds to the final step "Matching the resulting HamNoSys codes with correspondent SiGML tags" in Figure 4.8, example of a SiGML file in Appendix E. In Figure 4.9 the SiGML tags are present in the column most to the left (*HamNoSys token*).

Once all the associations are performed, a new SiGML file is generated designated as the last step of the architecture in Figure 4.8 and step "SiGML file (.sigml)" in Figure 4.1.

The parser as well as the conversion from HamNoSys to SiGML described in this section were implemented with Python. In the scope of our project, in order to avoid an overall broken pipeline, it is required that these scripts run through Unity software, used in the following steps ("VIEW" and "SAVE") as described in the architecture in Figure 4.1. An initial attempt to merge both Python and C# codes was to use IronPython, which is an open source implementation of Python tightly integrated with .NET framework. Nevertheless, the available version of IronPython is 2.6, whereas our code was implemented in Python 3.7. For this reason, it would be required to convert the code from Python 3.7 to 2.6. Such task is not as straightforward and would require time, time more valued in other tasks. On that account, an alternative solution was created. The project will provide a Python installer and the command line will run these scripts in the background. If required, in order to run the program, the due installation will be performed.

4.5.4 Extending the notation

The conversion can be further extended. Taking into consideration the currently limited facial expressions codes available in HamNoSys, a promising possibility resides in the extension of the notation system for this type of content.

As previously mentioned in Section 4.5.3, the system has access to approximately 210 HamNoSys codes and their corresponding SiGML tags, which are saved in a text file (conversionSpreadSheet.txt). With the aim of improving the content of this document, the user simply has to add entries to this file. Each entry must have the SiGML tag created and their respective HamNoSys Unicode codes.

4.6 SiGML to Signing

Once the data is saved into SiGML, the following step is to animate the avatar. As explained in Section 4.4, HamNoSys has four main components: hand shape, hand orientation, hand location and hand movements. Due to their importance, all these components in signing were implemented. Nevertheless, due to time restrictions, not all components are fully implemented. The avatar signing is limited to the use of one hand. Improvements for the implementation of the second hand are explained in Section 6.3. Nonetheless, our tool is not restricted to any sign language, it will animate any content as long as it is properly annotated in HamNoSys.

In this section we will explain the overall implementation of our tool, with further detail for each the HamNoSys components.

4.6.1 Implementation

As previously described in Section 4.4.1, 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, explained in Section 4.4.1. The implementation of **thumb position** is made during run time. In addition, some bending of the fingers are also implemented. The detailed explanation of this component is found in Section 4.6.2.

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. Section 4.6.3 focus on the description of this component.

As described in Section 4.4.2, 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 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, explained in further detail in Section 4.6.4.

Finally, the hand's movements were also done with IK, by moving its respective sphere with mathematical formulas, described in Section 4.6.5.

4.6.2 Hand Shape

The hand shapes' animations were constructed using Unity manual editor of curves. These animations crafted with this editor are for sign languages, ergo require a lot of detail within its configuration. Nevertheless, due to the lack of freedom and detail offered by this method of creation, the process of creating a good and perceptible configuration of the hand is slow and difficult. First of all, all the avatar's joints must be added manually. Afterwards, as shown in Image 4.10, the edition is done through the manipulation of lines for each joint of the avatar within the desired frames.

The thumb position, as mentioned in Section 4.6.1 is done in run time. Each position describes a specific configuration of the thumb. These configurations are written into the original animation, which is one of the basic hand configurations previously mentioned, during run time.

The implementation of fingers' bending depends on the current hand shape of the avatar. That is,

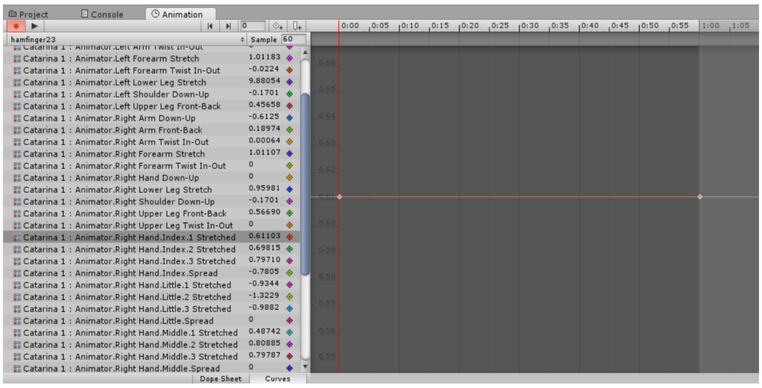


Figure 4.10: Unity editor curves for the animation *hamfinger23*, within the interval from frame 0 to 1, currently editing frame 0 (red line), of the first joint of the index finger (Animator.Right.Hand.Index.1.Stretched).

the same type of bending will play differently if the hand shape has one finger up, the bending is only applied to this finger, or two, in which the bending must be applied to the two fingers up. In addition, the values written for one finger are not the same as for a different finger, the situation repeats itself if the type of bending is different. For this reason, the bending of fingers was performed not by writing into the animation during run time, as with thumb configurations, but by creating new animation clip for each of them. In order to perform it otherwise, a very thorough processing for each situation was required, for instance calculate the precise position of the hand and apply the specific values taking into consideration that positioning. In Appendix A, the hand configurations implemented code independent are highlighted at green, whereas in light orange are the hand configurations implemented through hard code.

4.6.3 Hand Location

Before the implementation of this component, the system was working with Forward Kinematics (FK). FK is achieved through the use of kinematic equations of an avatar to compute the position of the end point from specific values for the avatar's joints. However, in order to apply the locations to the system we had to bring IK into play.

Nevertheless, the original animations created for the hand shape, Section 4.6.2, are not compatible with inverse kinematics. These are generic animations, written with the bones of the avatar, therefore, such animations are characteristics of the avatar in question as shown in Image 4.11.

For this reason, generic animations had to be converted into humanoid animations. The latter are animations written with the muscles, shown in Image 4.12. On contrary to the bones, the muscles are equal between avatars. Consequently such animations will run in any humanoid avatar and enable the use of inverse kinematics. However, with the scripting API available for Unity 5 it is not possible to properly map between the avatar's bones and muscles. Therefore, this conversion can not be performed

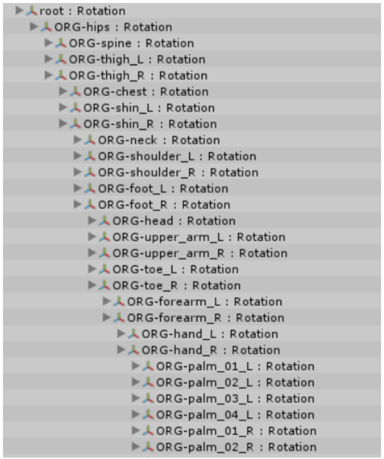


Figure 4.11: Generic animations.

Figure 4.12: Humanoid animations.

automatically. As a result, all the generic hand shapes' animations initially created were discarded and new humanoid animations were created from scratch. It is also worth mentioning that generic animations will not run together with humanoid animations. The state machine can only properly play one type of animation at once.

The implementation of the locations through IK was accomplished with the use of spheres. Each location is marked with a small and invisible sphere, illustrated in Figure 4.14. These spheres are placed in the correct location regarding the avatar skeleton. Initially, these spheres corresponded to the location of the avatar's wrist. Unfortunately, HamNoSys does not specify the exact location of the hand. Therefore, in order to avoid collisions, the location of the hands were calculated considering the position of the fingertip of the index finger.

As explained in Section 4.4.3, a location can be simple, as one HamNoSys symbol identifying it, example in the left side of Figure 4.13 which describes a position in front of the center of the avatar's chest, or a combination of various symbols, for instance three symbols describing the location touching the right side in the line of the chest exemplified in the right side of Figure 4.13. This leads to a great number of possible combinations, each requiring its own specific sphere. There are hundred-nineteen spheres. However, these spheres do not cover all the possible combinations. Due to the lack of the avatar's facial expressions, HamNoSys symbols correspondent to locations such as teeth and tongue are not implemented.

Figure 4.13: Example of the possibility of combinations with HamNoSys symbols.

The Unity's **state machine** represent, through a graph diagram, the actions to be played by a certain character, in this case the signing avatar. These actions will depend on the type of game play but typical

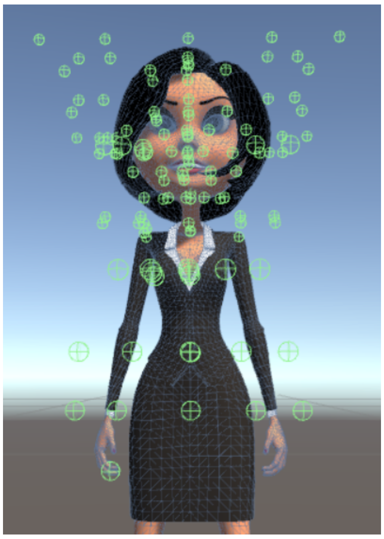


Figure 4.14: Current spheres available to identify locations.

actions include walking, running, jumping. In the context of this project, actions will be defined by the four components of the HamNoSys. These actions are referred to as **states**, and each state is associated with an animation clip. As described in Section 4.6.2, each hand shape has its own animation clip. Therefore, a composite key, with the hand configuration and the gloss, can be viewed as the primary key of a state. Different states can have the same hand configuration; however, the state name must differ. The connection between all states is made with transitions. Taken together, the set of states, the set of transitions and the variable to store the current state form a state machine.

Nevertheless, IK is **independent** from the state machine. Therefore, IK must be associated with a specific state or vice-versa. Both the state machine and IK must be in synchronization otherwise problems, such as a certain location running with the incorrect hand configuration and orientation, will arise. It is necessary that a given location runs **specifically and only** while the animation of its respective gloss is playing.

4.6.4 Hand Orientation

This component was the most challenging to implement. Various versions were developed until the desired goal was achieved. To build a solution that would incorporate both the extended finger and the palm orientation without conflict was challenging.

The first and most naive try followed the logic behind the implementation of the thumb positions, explained in Section 4.6.2. Both the values of the extended finger and palm orientation were saved in its respective text file and were then written to the animation clips. Nevertheless, the values that would work for one position of the hand would not work for a different one. Hand orientations not only change the values of position of the wrist, but also of the forearm. The alternative would be to also pass values for the forearm. On the other hand, changing the values of the forearm would entirely change the sign since it too would interfere with the location of the sign. In other words, on doing so, it would generate a whole different animation, almost independent from the input.

Considering inverse kinematics was already implemented in the project (Section 4.6.3) the next try was to use it to also implement this HamNoSys component. In addition to the location feature available, IK also provides a component to change the rotation of the hand. The implementation of the extended finger was done using this feature. The extended finger is defined with vectors. These vectors define the rotation of the hand in the Z axis. The problem that arose was to incorporate both the extended finger and the palm orientation. Once the former was already correctly working, the result for the latter was not the expected. The palm orientation is dependent on the extended finger in the sense that it can only be applied after the later. The solution was rather complicated since the palm orientation must always have a ninety degree angle with the extended finger, requiring a pre-processing of the original palm orientation's vectors according to the values of the extended finger.

Despite the values being still saved within a text file, they are no longer written to the animation clips. These values are vectors which are applied to the rotation of IK. Hand orientation too requires strict synchronization between the values being evaluated and the state running.

One of the biggest advantages of inverse kinematics is its independence, which makes it so easy to use (once correctly understood). As soon as it has the values, IK does all the work with no need for external control. On the other hand, it may not provide the intended output. IK will always choose the **shortest path**; however, cases exist in which both paths have the same length. For such cases it has to be guaranteed the correct choice. In Figure 4.15, both the paths (A and B) have the same length, reason being that the initial position (0,0,1) and the final position (0,0,-1) have one hundred and eighty degree angle between them. Nonetheless, the choice affects the final output. If the path chosen is A, the final palm orientation will be different than if it is B. For such cases, a pre processing of the data is performed in order to make IK choose the desired path. A new value for the initial point of (0.1, 0, 1) is created (green dot), forcing the IK to choose path C, a minor derivation of path B.

4.6.5 Hand Movement

As described in Section 4.4.4, movements can be distinguished between absolute and relative. The former are achieved through a change in locations, with a specified initial and final location, as exemplified in Figure 4.6. Relative movements according to a specific direction, size and initial location, as exemplified in Figure 4.7.

The implementation of **absolute movements** was rather straightforward. The same state is replicated, which means same hand configuration and orientation, but associated with a different location. The transition between the two states will perform the desired movement.

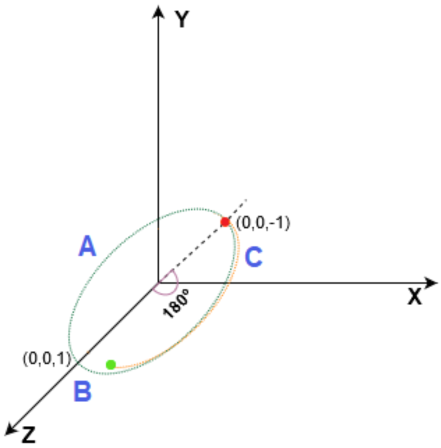


Figure 4.15: Example of when the initial and final point have an 180º angle between them.

In contrast, the implementation of **relative movements** was not as straightforward. These movements are achieved through the use of mathematical formulas. These equations will be applied to the movement of a sphere which is controlled by the inverse kinematics. The movements of the sphere will be the movements performed by the avatar's hand, more specifically by the avatar's index finger's fingertip. However, as previously mentioned, inverse kinematics is independent from the state machine. Considering the mathematical formulas are implemented through the use of IK, these too are independent from the state machine, which once again requires for a strict synchronization of all components.

The Unity's **animator** is the interface responsible to control the Mecanim Animation System, which means it controls the state machine. Since the implementation of relative movements is independent from the state machine, the synchronization of the mathematical formulas and the state machine required that for each movement the animator was paused. This restriction allows for the proper synchronization of all components of a sign. By pausing the animator, it is ensured that during a certain movement, the avatar will only play the correct state, therefore the correct components of the HamNoSys, and not transition to others.

4.7 Integration with PE2LGP

Our tool is incorporated with a previous project also developed for PSL, PE2LGP [18]. 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 Figure 4.16. These animations can be later visualised or used in a basic translator from written Portuguese to PSL, respectively menu "Gestor" and "Sistema 1" in Figure 4.16.

The menu "Sistema 4" in Figure 4.16 represents our tool. Through this menu it is possible to animate

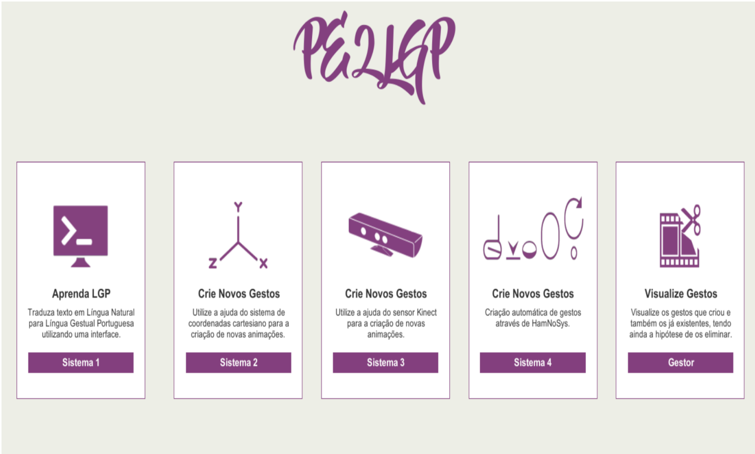


Figure 4.16: PE2LGP main menu.

the avatar through HamNoSys. The animation of the avatar will be achieved with Unity by receiving a SiGML. This file will inform Unity on the movements required for each sign and will be processed as explained in Section 4.6.

4.7.1 Visualize Animations

The animation of the avatar can be used either to visualize the signs or to save the created content. This component consists on the step "VIEW" from Figure 4.1. The interface of this feature is fairly simple, available in Figure 4.17. 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 4.1, 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).

4.7.2 Save Animations

The other feature of the system consists on saving signs as an animation clip, step "SAVE" in Figure 4.1. In Figure 4.17 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 4.1. Each animation clip will be identified by its gloss.

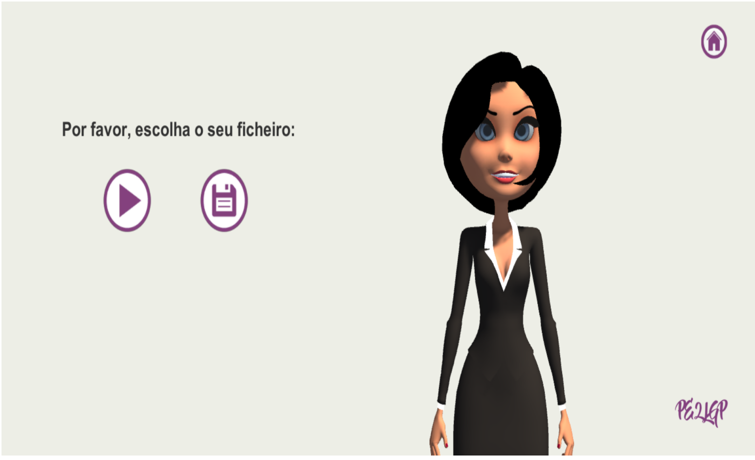


Figure 4.17: Tool's menu.

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**.

Furthermore, following the line of thought from Section 4.6.3, although the animations in use are humanoid animations, in Figure 4.12, the animation clips were saved as generic animations, in Figure 4.11. The reason lies in the fact that with the current version of Unity in use it is neither possible to access the muscle values required to write an humanoid animation, nor to map from bones to muscles. That being the case, considering that to save an animation it is necessary to access the values in which the avatar is positioned, the solution was to use the bones values, used to write a generic animation, which can be accessed. These animations will be saved in the file system, represented as step "File System (.anim)" in Figure 4.1, together with the animations from the previous version of the project [18] and can be visualise in its respective component "Gestor" from Figure 4.16. Due to time restrictions, the tool only save static signs, signs with no more than one keyframe.

Besides saving the animation clips by gloss, within this component another feature is available, step "Data serialized (.xml)" displayed in Figure 4.1. 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, an example of this file is available in Appendix E. Just as with the animations, in this XML there too can be more than one annotation per gloss, which is the case with the gloss "BOM" in the example from Appendix C wit three different annotations even if only one symbol is different from previous annotations. That is, **every** sign saved with our application will have its information written within this file. Statistic and further studies can be developed with the aid of this step, allowing a further analyze of the annotations. For instance, if the same gloss has different HamNoSys annotations, it can be studied to understand if these situations were merely annotation's mistakes or possible derivations of the same sign, for instance, signed in a different region.

L2F, which is a research group 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 4.18. The population of the file system is done with annotations from the Corpus, step “Corpus in ELAN (.html)” in Figure 4.18, 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 Figure 4.18. These files are used to synthesized the avatar signing, and if the user chooses to save the content, “SAVE” in Figure 4.1 and in Figure 4.18, 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 Figure 4.18. This lookup returns the gloss respective animation clip, which is then played by the avatar.

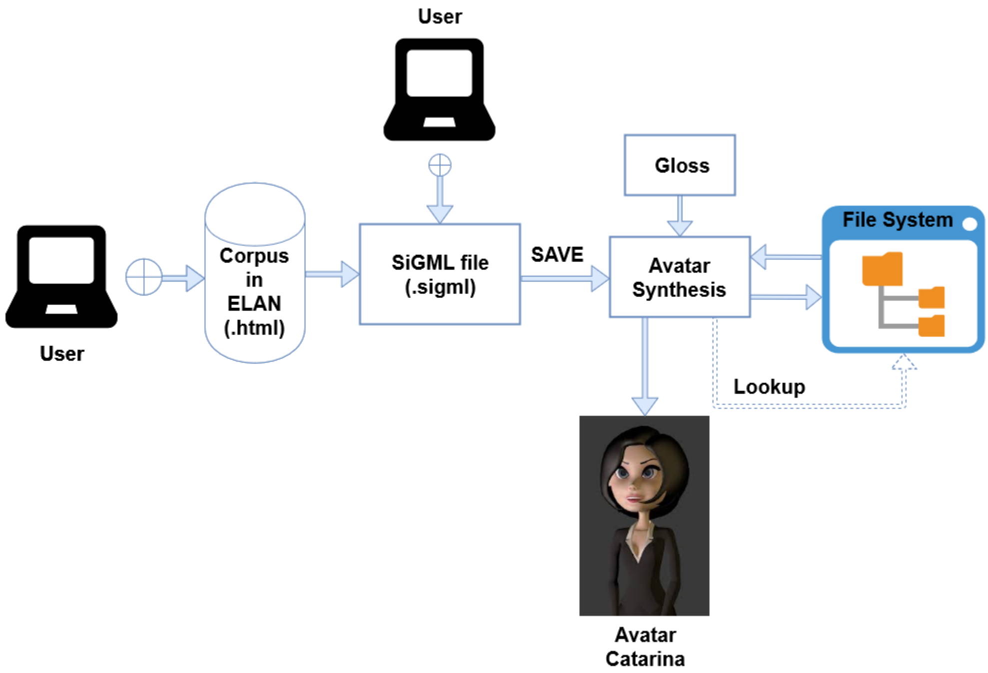


Figure 4.18: Architecture with lookup feature.

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 4.1, 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.

4.8 Prototype limitations

The avatar can only sign with the main hand (right hand). Focusing on just one hand allowed for more detailed implementation. However, a lot of signs require a **second hand** (left hand) in its performance, even if as a support rather than an equal, for instance, a support in which the right hand can touch repeatedly. The left hand should follow the implementation of the right hand, explained in Section 4.6, with its own animations mirrored from the first's hand animations. To incorporate and synchronize both hands, it must be created a new layer of the avatar's animator for the second hand. In addition, the spheres used for the location of the right hand, described in Section 4.6.3, can be reused by the left hand.

Taking into consideration the hand's configurations used in sign languages, explained in Section 4.4.1 and 4.6.2, the detail involved in this component is of major importance to its correct comprehension. By now, however, not all possible hand shapes are implemented, mentioned in Section 4.6.2. In order to automatically produce all the configurations possible to create with HamNoSys, an algorithm should be developed with focus on **moving all hand's components based on a combination of symbols**. An idea for this would be to explore the implementation of IK in the fingers so that they would recognize the different fingers and their different parts, each finger has three components, and touch each other easily and without collisions as well as to bend specific parts of a specific finger.

In regards to the implementation of saving signs with movement (animations with more than one keyframe), the challenge lies with the loss information. The only part lacking of this feature is that when the velocity of these animations is decreased, in order to be properly captured the movements, the avatar does not play the animation of the hand configuration properly, leading to a movement with no hand configuration. As mentioned in Section 4.6.5, the animator is paused for the implementation of movements. Therefore, the problem might lie in the fact that the animator is being paused before the avatar has played the hand configuration required for the sign. The solution should focus on the synchronization of the states machine, the animator and the movements.

Finally, the lack of **facial expressions** is other limitation. Facial expression convey meaning to the signing. Nevertheless, in the development of our tool this field was not cover due to the impossibility of changing the facial meshes of the avatar in use. Nevertheless, this component is being implemented by a team member from the project. Once it is possible to move the avatar's facial meshes, animation clips can be created with the facial expressions necessary. The facial expressions must then be annotated in ELAN and codes for facial expressions were already developed by the interpreter of the project, see in Appendix G. The HamNoSys symbols and their respective SiGML codes must be added to the conversionSpreadSheet.txt and the parser already created can be reused. The processing of this data can be done as in conversion from HamNoSys to SiGML, Section 4.5. The facial expressions should have their own layer in the avatar's animator and must be synchronized with the other layers, for now the layers of the right hand.

5

User Study and Results

Contents

5.1 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 of the content returned by our tool we will perform a joint evaluation to compare two avatars: our on progress avatar Catarina and the state of the art online avatar Anna[1].

In this section we will go into detail on our user study, the research questions we plan on answering, Section 5.1.1. In Section 5.1.2 we explain our the content evaluated was created, the user study procedure in Section 5.1.3, and finally information about our participants in Section 5.1.4.

5.1.1 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?

5.1.2 Dataset

The content used for the evaluation was mostly retrieved from the corpus, described in Section 4.2. In order to create more diversified and complex sentences, some signs were annotated based on content from SpreadTheSign[2], in which the content is also signed by native deaf people. The online dictionary from Porto Editora[3] was not used since, on contrary to the content mentioned above, this was neither signed nor created with native deaf people.

Overall twenty sentences were created for this evaluation, presented in Table 5.1. As shown in the Figure, each sentence was then 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 was their SiGML files. For twenty sentences, twenty SiGML files were returned.

Our evaluation will focused on the comparison of both our on progress avatar Catarina, in Image 5.2, and the state of the art online avatar Anna[4], in Image 5.1, with its default settings. Both avatars were animated with the same content, these SiGML files. Of the twenty SiGML files, the former could only sign animate eighteen.

[1] http://vhg.cmp.uea.ac.uk/tech/jas/vhg2019/cwa/TwoAvServer.html
[2] https://www.spreadthesign.com/pt.pt/search/
[3] https://www.infopedia.pt/dicionarios/lingua-gestual
[4] http://vhg.cmp.uea.ac.uk/tech/jas/vhg2019/cwa/TwoAvServer.html

Nr	Written Portuguese	PSL in glosses	HamNoSys
1	Bom dia.	BOM DIA	[illegible]
2	A menina de óculos vê uma flor.	MENINA OCULOS FLOR VER	[illegible]
3	Tu gostas de animais.	TU ANIMAL GOSTAR	[illegible]
4	O segurança também quer respeito.	GUARDA-SEGURANÇA RESPEITO QUERER	[illegible]
5	A criança é inteligente.	CRIANÇA INTELIGENTE	[illegible]
6	O director pede dinheiro.	DIRECTOR DINHEIRO PEDIR	[illegible]
7	O meu namorado tem olhos azuis.	NAMORADO MEU OLHOS AZUIS	[illegible]
8	O neto come formigas.	NETO FORMIGA COMER	[illegible]
9	O papa é bom ouvinte.	PAPA OUVINTE BOM	[illegible]
10	Ali está pouco sol.	ALI SOL POUCO	[illegible]
11	O estado tem a palavra	ESTADO PALAVRA TER	[illegible]
12	O pai zangou-se com o filho.	PAI FILHO ZANGAR	[illegible]
13	Eu pergunto-te.	EU TU PERGUNTAR	[illegible]
14	O médico ouve uma história.	MÉDICO HISTÓRIA OUVIR	[illegible]
15	A mulher cega é simpática.	MULHER CEGA SIMPÁTICA	[illegible]
16	O jovem quer maçã.	JOVEM MAÇÃ QUERER	[illegible]
17	Ela tem uma casa branca.	MULHER CASA BRANCO DELE	[illegible]
18	Eles comem massa.	ELES MASSA COMER	[illegible]
19	Eu vejo o meu filho.	EU FILHO MEU VER	[illegible]
20	Ele é alemão.	ALEMANHA PAÍS DELE	[illegible]

Table 5.1: Table with content for evaluation in written Portuguese, PSL in glosses and annotated in HamNoSys.

Both avatars were recorded animating each SiGML. These recordings were captured through OBS Studio[5], which recorded the unity and firefox screen, for avatar Catarina and avatar Anna respectively. Afterwards, the edition of the videos, such as cutting and adjustments in the display window so that only the avatars appeared were done with Vegas Pro[6]. 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 used to animate the avatars.

The order of the sentences at which each avatar would animate were chosen randomly[7], considering that two sentences, sentence seven and nine, in Figure 5.1, were always assigned to avatar Catarina since avatar Anna could not play these. In addition, the avatars' order in signing is counterbalanced. The avatar Anna is the first for odd numbers and the avatar Catarina the first for even numbers.

5.1.3 Procedure

The questionnaires were created in Google forms. The study was performed asynchronously and remotely. One version of the questionnaires, example included in Appendix H, was emailed to all the

[5] https://obsproject.com/
[6] https://www.vegascreativesoftware.com/pt/vegas-pro/
[7] //www.random.org/lists/

Figure 5.1: Avatar Anna.

Figure 5.2: Avatar Catarina.

participants.

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 the avatars' order is counterbalanced and the phrases' order random. The questionnaire is composed of twenty four sections, Appendix H. The first section regards demographic 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 signed by the avatars. In these sections, participants are asked to provide the content understood from the video as well as the number of visualizations required with additional comments if desired.

Besides, after the ten sections correspondent to each avatar, an additional section asks for an evaluation of the avatar in terms of speed, general quality, comprehensibility, naturalness, grammatical correctness, hands' configurations, hands' orientations, hands' locations and hands' movements in a 1 - 5 Likert scale, with 1 as terrible and 5 as perfect.

5.1.4 Participants

We were able to recruit seven participants (four females). 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. Their ages ranged from 26 to 55 (M=43.86, SD=11.88) years old. Five of the participants are identified as Deaf and three were native signers. From the remaining four, one uses PSL since birth, two for more than 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 one participant has ever used signing avatars but rarely.

5.2 Results

In this chapter, we present a comprehensive set of results achieved from the user study.

5.2.1 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. This section presents the results related to the impact of each of the components implemented in this project.

Velocity

The first characteristic evaluated during this study was the perceived appropriateness of velocity of the signing of each of the avatars, depicted in Figure 5.3. Avatar Catarina presents a higher dispersion of the results (inter-quartile range lies between the minimum and maximum), 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.

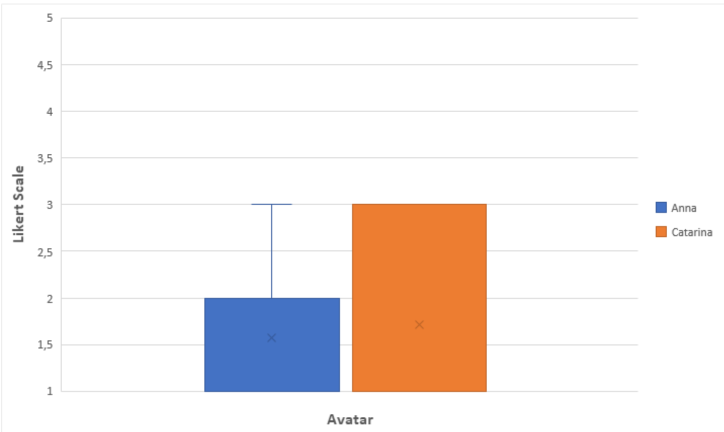


Figure 5.3: Velocity's rate for each avatar. The chart presents the median (line inside the boxes), the 1st and 3rd inter-quartile ranges (boxes), and the maximum and minimum values (whiskers).

Overall Quality, Understandability and Naturalness

In all characteristics of the signing the data, Figure 5.4 and Figure 5.5, 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. For all characteristics, the statistical analysis of both avatars indicated no substantial difference. 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) of the avatars were similarly evaluated.

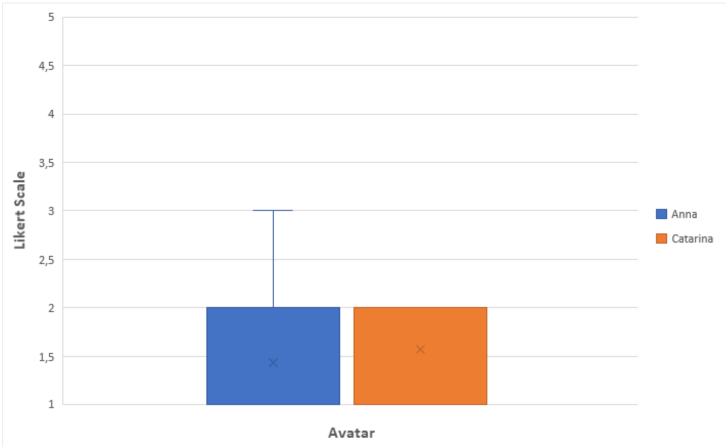


Figure 5.4: Overall quality's rate for each avatar. The chart presents the median (line inside the boxes), the mean (X), the 1st and 3rd inter-quartile ranges (boxes), and the maximum and minimum values (whiskers).

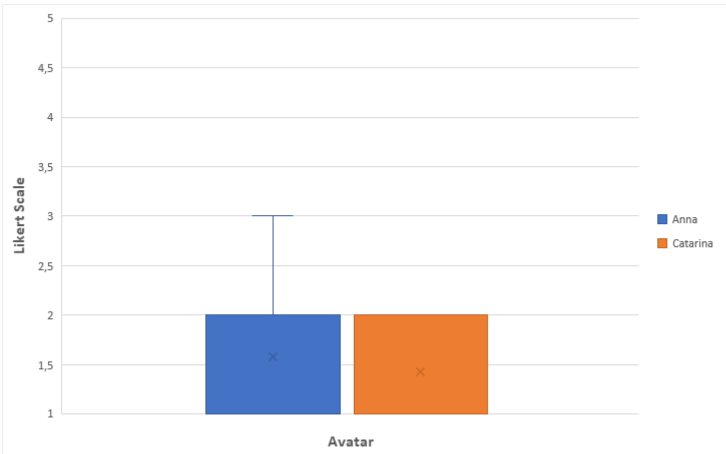


Figure 5.5: Understandability's and Naturalness's rate for each avatar. The chart presents the median (line inside the boxes), the mean (X), the 1st and 3rd inter-quartile ranges (boxes), and the maximum and minimum values (whiskers).

Grammatical Correctness

For this section of the study, the participants had the choice of "I don't know." In this component, one participant chose that option. Therefore, the results showed in Figure 5.6 correspond to 6 participants instead of the original 7.

By interpreting Figure 5.6, 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.

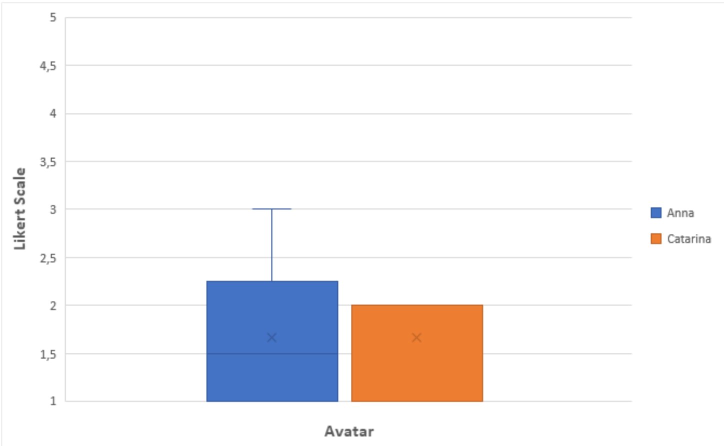


Figure 5.6: Grammatical Correctness' rate for each avatar. The chart presents the median (line inside the boxes), the mean (X), the 1st and 3rd inter-quartile ranges (boxes), and the maximum and minimum values (whiskers).

Hand's Configuration

The first HamNoSys component evaluated was the hands' configurations, Figure 5.7. 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.

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. As displayed in Figure 5.8 and Figure 5.9, 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.

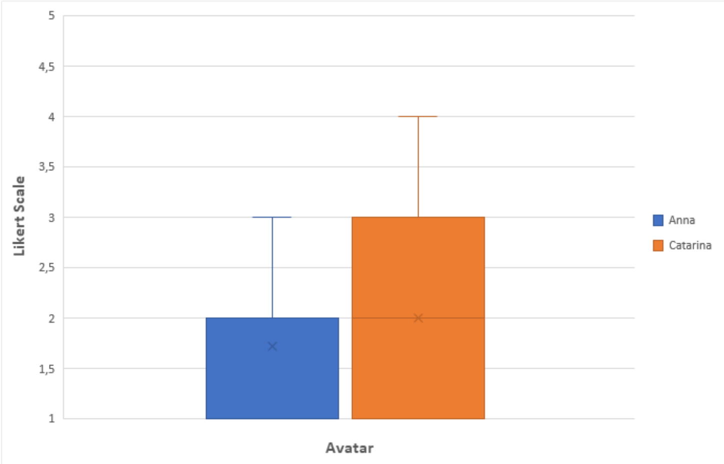


Figure 5.7: Hands' Configuration's rate for each avatar. The chart presents the median (line inside the boxes), the mean (X), the 1st and 3rd inter-quartile ranges (boxes), and the maximum and minimum values (whiskers).

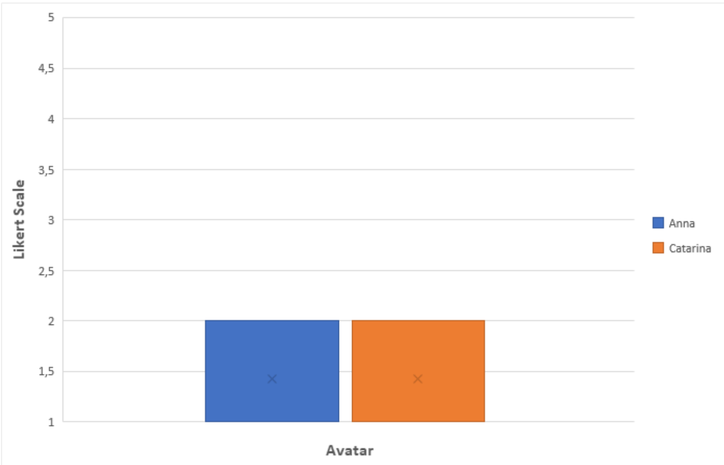


Figure 5.8: Hands' Orientation's rate for each avatar. The chart presents the median (line inside the boxes), the mean (X), the 1st and 3rd inter-quartile ranges (boxes), and the maximum and minimum values (whiskers).

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.

5.2.2 Comprehension

In this section, we analyze the number of visualizations of every sentence for each avatar, Section 5.2.2, as well as the percentage of content correctly understood by the testers for each avatar, Section 5.2.2.

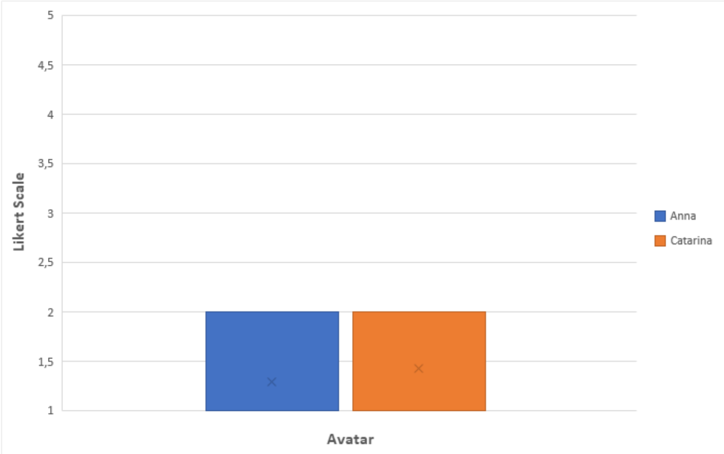


Figure 5.9: Hands' Movements' rate for each avatar. The chart presents the median (line inside the boxes), the mean (X), the 1st and 3rd inter-quartile ranges (boxes), and the maximum and minimum values (whiskers).

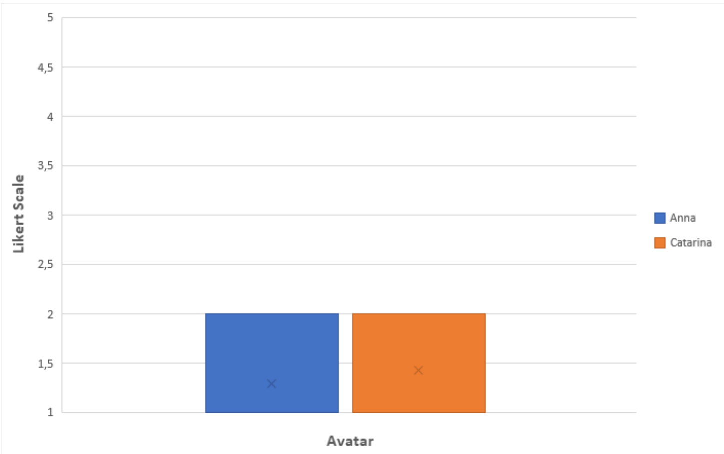


Figure 5.10: Hands' Location's rate for each avatar. The chart presents the median (line inside the boxes), the mean (X), the 1st and 3rd inter-quartile ranges (boxes), and the maximum and minimum values (whiskers).

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 (example of a questionnaire provided in Appendix H).

In the bubble chart below, Figure 5.11, each bubble represents a person. The bubble area is proportional to the number of people who saw the video that number of times. The colour indicates the avatar, and the horizontal and vertical positions relate to a total of visualizations and sentence number, respectively. Sentences 7 and 9 were not signed by avatar Anna,Section 5.1.2.

By analyzing this chart, the colour orange (avatar Catarina) is noticeably more abundant with smaller size bubbles spread out throughout the graph, whereas the colour blue (avatar Anna) is less common but with bubbles in bigger sizes. This indicates a higher consistency between testers in the number of visualisations for avatar Anna. On the other hand, for avatar Catarina, the biggest bubbles are mainly found in the lowest numbers of visualizations, whilst for avatar Anna, these are located in the highest number of visualizations. That is, contrary to avatar Catarina, avatar Anna's videos required a high number of visualizations for each sentence.

The sentences and their order were random for each test. For this reason, the number of viewers might not be the same for both avatars. In the majority of the cases the audience was similar (with weight difference to a maximum of 3 to 1 or vice versa); however, sentences 2, 12 and 13, were only seen by one person with avatar Catarina (orange bubble), as displayed in Figure 5.11.

Figure 5.11: Distribution of the number of visualizations for every sentence in each avatar. The X axis represents the sentence number (from 1 to 20) and the Y axis the number of visualization (from 1 to more than 5).

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. As previously mentioned, sentences seven and nine were not signed by the JA Signing avatar (Avatar Anna).

In PSL, the meaning of a sign can be influenced based on the facial expressions and on the context of the sentence. For the scope of this evaluation, we worked with loose sentences with no concrete context and no facial expressions. For this reason, and in alignment with the interpreter of Portuguese and PSL of the project, the assessment of the content was performed in line with some premises: for the gloss “ELES” the answer “VOCÊS” is also accepted; for the gloss “DELE” the answer “SEU” and “TEU” is also accepted, Appendix I.

Note in Figure 5.12 that the percentage achieved by Avatar Catarina in several sentences shows a positive discrepancy in comparison to Avatar Anna. On the other hand, in five sentences the content comprehended with Avatar Catarina led to wrong answers (0 per cent).

Generally, the sentences with the most positive percentages (above or equal to 60%) were sentence 1, 2, 5, 7 and 12. Going into further detail for each of these sentences, available in Appendix I:

- Sentence 1 - “GOOD MORNING” (“BOM DIA”): Simple content, two glosses, one without and another with movement, “GOOD” and “MORNING” in the respective order. The last gloss seems to have been better understood. Avatar Anna had a higher average number of visualizations per person.

- Sentence 2 -“GIRL GLASSES FLOWER SEE” (“MENINA OCULOS FLOR VER’): Composed by four glosses. Although it was the only gloss with no movement, “GLASSES” was only comprehended by one participant, who saw this content from avatar Catarina.

- Sentence 5 - “KID SMART” (“CRIANÇA INTELIGENTE”): Simple sentence with only two glosses, both with movement. Only two testers identified the second word, both through avatar Catarina. However, Anna had more visualizations.

- Sentence 7 - “BOYFRIEND MINE EYES BLUE” (“NAMORADO MEU OLHOS AZUIS”): Sentence with four glosses, signs have movement and the content was evenly understood by all the participants.

- Sentence 12 - “FATHER CHILD ANGRY” (“PAI FILHO ZANGAR”): All three signs have movement. For the first two glosses, even when not properly understood, participants would wander around words of the same type (between “FATHER”, “MOTHER” and “CHILD”). On contrary to avatar Catarina, this sentence presents good results for avatar Anna, in which the number of visualizations is also higher.

In conclusion, 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, by analyzing Figure 5.12, we can gather that the majority of these sentences have the highest values with avatar Catarina with a lower number of visualizations per participant.

As displayed in Figure 5.13, 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

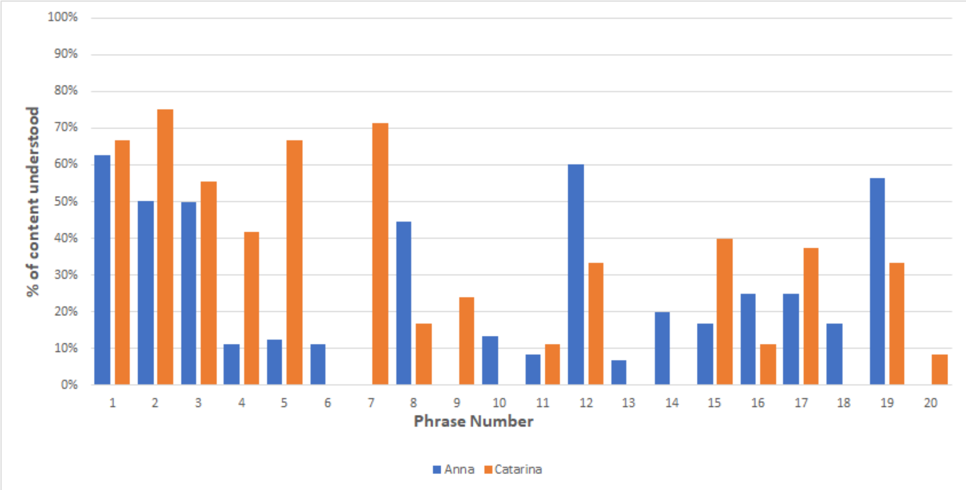


Figure 5.12: Distribution of the understanding percentages for every sentence in each avatar. The X axis represents the sentence number (from 1 to 20) and the Y axis the percentage of correctly understood content.

some sentences did not have the same number of visualizations for both avatars, which might lead to misleading results. For instance, for sentence 13 avatar Catarina had a comprehension value of 0 per cent; however, the content was only viewed by one person. On the other hand, sentence 2 had a high percentage value for avatar Catarina but this also corresponds to the visualization of only one participant. The lack of evenly distributed sentences played a major role in the discrepancy of the results. Overall, the our results showed that there was no significant statistical change between the signing of the two avatars ($Z = -0.240^b, p = 0.811$).

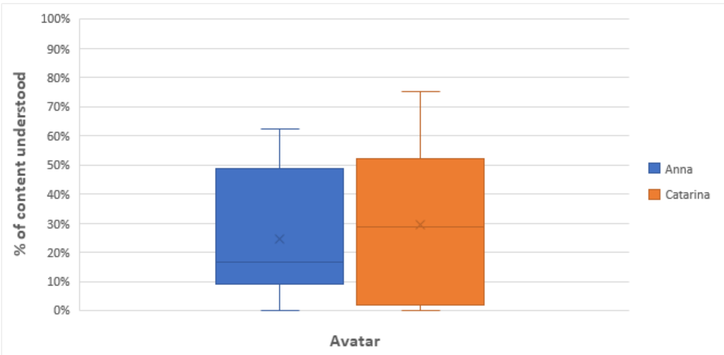


Figure 5.13: Percentage of content understood for each avatar. The chart presents the median (line inside the boxes), the mean (X), the 1st and 3rd inter-quartile ranges (boxes), and the maximum and minimum values (whiskers).

5.2.3 Participants' Feedback

As part of our study, the participants were also asked for additional comments about the avatars' signing. For simplification reasons, these comments will be organized by avatar, their performance and the linguistic component of the sentences used.

Avatar Anna

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 noticed by one tester.

On the other hand, in a **more technical context**, some corrections were made to a few signs:

- Gloss “ANIMAL” (“ANIMAL”): The location was described as wrong.
- Gloss “GET-ANGRY” (“ZANGAR”): This sign should be improved.
- Gloss “GOOD” (“BOM”): The signing of this gloss was viewed as not natural.
- sentence “GOOD MORNING” (“BOM DIA”): The configuration of this sentence was defined as incorrect.

Avatar Catarina

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.

In addition, in a **more practical context**, some observations related to the content were mentioned:

- Gloss “BOYFRIEND” (“NAMORADO”): The location is wrong and the sign is incorrect.
- Gloss “EYES” (“OLHOS”): This sign should be performed with two fingers and not only one.
- sentence “BOYFRIEND MINE BLUE EYES” (“NAMORADO MEU OLHOS AZUIS”): The need for the verb “HAVE”.
- Gloss “BLIND” (“CEGA”): The ending ought to be more natural.
- sentence “WOMAN HOUSE WHITE HIS” (“MULHER CASA BRANCA DELE”): The signs “WHITE” and “HOUSE” are executed repeatedly. It should either only be done once or faster.

5.2.4 Discussion

In this section, we summarized the major results attained and answer our research questions.

User Study Conclusions

The null hypothesis was always retained in the evaluation of the sign languages' components, Section 5.2.1, from which we can conclude that the results were similar for both avatars in all components, mainly fluctuating from rate 1 to 2, suggesting **poor performances**. Overall, the values for **avatar Catarina are slightly more inconsistent**, with higher values for the standard deviation and bigger dispersion of the data.

With regard to the comprehension section of the tests, out of the twenty sentences used for this study, only five of these (one quarter) provided results above sixty per cent, with an overall maximum of seventy per cent. These values can not be deemed good, since they indicate that the majority of the content was not comprehended.

In reference to the opinion component of the study, interesting conclusions can be drawn. Both avatars were described as robotic and the velocity as too fast for avatar Anna and too slow for avatar Catarina. Besides, in sentence 7, which is only signed by avatar Catarina, in spite of receiving several comments concerning its sentence structure, location and configuration of the signs, according to Figure 5.12, it was the sentence with the second higher percentage of content properly understood (71.43%), with the number of visualizations mainly located in the lowest numbers Section 5.2.2.

This may lead to an interesting deduction: **the signing of virtual agents can be separated into two main ingredients, the "attitude" of the avatar and construction of the sentence itself**. The attitude can be seen as the naturalness of movements, whereas the content construction can be defined mainly as the HamNoSys components. In spite of the comprehension for sentence 7 being good, additional comments mentioned problems with more practical issues, leading to the conclusion that the "attitude" of the avatars outweighed these grammatical errors.

Although not supported by our results, other questions can be raised. Some errors were noted by the participants in the additional comments' section, Section 5.2.3, such as the wrong location for the sign "ANIMAL" with avatar Anna and the sign "BOYFRIEND" with avatar Catarina. These might lead to the question of whether **the source of the error** laid on the annotation process in HamNoSys or in the performance of the avatars. In addition, the majority of the signs used in our study had movement. One can argue if an animation of only **static signs**, signs with no movement such as with gloss "BOM", would provide better results. For example, in the implementation of a finger-spelling system. This could be related to the fact that movements are implemented through mathematical formulas, which can lead to stiffness in the movements performed by the avatars. Finally, questions might be raised concerning the **field of view available in the videos** used for the study. The videos are 2-dimensional, whereas sign languages are 3-dimensional languages and information can be lost in this transition. For instance, for sentence 6 “DIRECTOR MONEY ASKS” (“DIRECTOR DINHEIRO PEDIR”), which displayed bad results in the comprehension of the content; in spite of the sign for money being considered simple, the camera view used for the videos was frontal which might lead to participants missing the imperceptible movement performed during some signs, such as with the sign “MONEY”.

All in all, only seven people participate in the user study, which can be consider a **small population** to take strong conclusions from. For instance, conclusions related to the participants' characteristics and the test's results cannot be made. Furthermore, the tests should have been equal for every participant, in order to report more objective conclusions. For instance, to understand whether some errors' were related to the avatar's signing or the annotation of the content (the wrong configurations and locations mentioned in the comments, Section 5.2.3, can be affected by either).

Research Questions

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 both avatars performed poorly and not according to expectations.

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 a more detailed implementation of signing velocities, which are not in correspondence with the expectations. This leads to inefficient agents for communicating PSL.

6

Conclusions

Contents

6.1 Achievements

We presented the implementation of 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. On contrary to other software in the field, our tool allows for a fast creation of a lot of content simply be analyzing linguists' annotations in HamNoSys from ELAN.

We performed a user study with seven participants to test our tool. 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.

6.2 Limitations

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.

Sign languages rely greatly on time components. The same sign signed with different speeds can have completely different meanings [2]. Our user study, Section 5.2.1, showed that on the one hand the avatar Anna signed too fast, while on the other, the avatar Catarina too slow. The lack of proper control over the speed of signing limits the overall potential of our tool. The stiff posture of the avatar was noted as both avatars were described as robotics.

One of the most important features in sign languages is the use of facial expressions, as described in Section 2.2. Our user study, Chapter 5, also showed that participants noticed and commented on the lack of facial expressions in some sentences.

6.3 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. In spite of only the most basic components required for a proper signing were cover in this project, a lot of potential lies within **its growth**. Some ideas were gathered to improve its content.

One of the limitations of our current project is the velocity, as mentioned in Section 6.2. This would be a major improvement would be the possibility to implement a thorough **speed of signing**. The insertion of pauses during a speech alongside its correct duration, as well as the proper speed of each sign and in the transitions between, could vastly improve the comprehension of the final result. This would, however, require a detailed study on these components. Besides, allowing the user to control these velocities could prove useful.

The posture of the avatar during the process of signing should be natural, Section 6.2. The implementation of **natural and unscripted movements** of the avatar would bring great value since it would lose the robotic attitude with which it was described. Taking into consideration that the hand's movements were implemented through mathematical formulas that can lead to more stiff motions, these natural movements should also be extended to include this component.

On contrary to humans, avatars know no limits to their movements. This might lead to possible inhuman results. A crucial feature is the implementation of **collisions**. Resorting to the use of collisions, movements such as the hand penetrating the head or chest will be avoided. Aside from the fact of allowing the programmer to a greater degree of freedom when implementing HamNoSys components such as hand's movements, which, might perform poorly since these were created as generic and a change in the location might lead to a collision between the avatar's hand and other body parts.

The implementation of collisions leads to another important limitation, the location of the hand used in IK. It is implemented considering the fingertip of the index finger. Nevertheless, ideally, this IK should be dynamic with the possibility to specify the finger to which it ought to be applied. For that reason, the development of an **IK algorithm for the hands** would offer a wider support base for the implementation of HamNoSys.

Bibliography

[1] ADAMO-VILLANI, N., POPESCU, V., AND LESTINA, J. A non-expert-user interface for posing signing avatars. *Disability and Rehabilitation: Assistive Technology 8*, 3 (2013), 238–248.

[2] AL-KHAZRAJI, S., BERKE, L., KAFLE, S., YEUNG, P., AND HUENERFAUTH, M. Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations. 259–270.

[3] ALMEIDA, I., COHEUR, L., AND CANDEIAS, S. Coupling natural language processing and animation synthesis in portuguese sign language translation. In *Proceedings of the Fourth Workshop on Vision and Language* (2015), pp. 94–103.

[4] ALMEIDA, I., COHEUR, L., AND CANDEIAS, S. From european portuguese to portuguese sign language. In *Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies* (2015), pp. 140–143.

[5] BENTO, J., CLAUDIO, A. P., AND URBANO, P. Avatars on Portuguese sign language. 1–7.

[6] BENTO, J., CLAUDIO, A. P., AND URBANO, P. Avatars on Portuguese sign language. 1–7.

[7] CASSIDY, S., CRASBORN, O., NIEMINEN, H., STOOP, W., HULSBOSCH, M., EVEN, S., KOMEN, E., AND JOHNSTON, T. Signbank : Software to Support Web Based Dictionaries of Sign Language. 2359–2364.

[8] CHOUHAN, T., PANSE, A., VOONA, A. K., AND SAMEER, S. M. Smart glove with gesture recognition ability for the hearing and speech impaired. *2014 IEEE Global Humanitarian Technology Conference - South Asia Satellite, GHTC-SAS 2014* (2014), 105–110.

[9] COSTA, A. C. D. R., AND DIMURO, G. P. SignWriting and SWML: Paving the way to sign language processing. *Traitement Automatique des Langues de Signes, Workshop on Minority Languages* (2003), 11–14.

[10] DE CONTI, D. F., BOLOGNINI, C. Z., CORADINE, L. C., BRITO, P. H. D. S., FERREIRA, C. M., BENETTI, Â. B., COSTA, P. D. P., DE MARTINO, J. M., ANGARE, L. M. G., DO AMARAL, W. M.,

SILVA, I. R., POETA, E. T., AND KUMADA, K. M. O. Signing avatars: making education more inclusive. *Universal Access in the Information Society 16*, 3 (2016), 793–808.

[11] DE CONTI, D. F., BOLOGNINI, C. Z., CORADINE, L. C., BRITO, P. H. D. S., FERREIRA, C. M., BENETTI, Â. B., COSTA, P. D. P., DE MARTINO, J. M., ANGARE, L. M. G., DO AMARAL, W. M., SILVA, I. R., POETA, E. T., AND KUMADA, K. M. O. Signing avatars: making education more inclusive. *Universal Access in the Information Society 16*, 3 (2016), 793–808.

[12] EFTHIMIOU, E., FOTINEA, S.-E., GOULAS, T., VACALOPOULOU, A., VASILAKI, K., AND DIMOU, A.-L. Sign Language Technologies and the Critical Role of SL Resources in View of Future Internet Accessibility Services. *Technologies 7*, 1 (2019), 18.

[13] EFTHIMIOU, E., FOTINEA, S. E., HANKE, T., GLAUERT, J., BOWDEN, R., BRAFFORT, A., COLLET, C., MARAGOS, P., AND LEFEBVRE-ALBARET, F. 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 (2012), 205–212.

[14] EFTHIMIOU, E., FOTINEA, S. E., HANKE, T., GLAUERT, J., BOWDEN, R., BRAFFORT, A., COLLET, C., MARAGOS, P., AND LEFEBVRE-ALBARET, F. 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 (2012), 205–212.

[15] ELLIOTT, R., GLAUERT, J. R. W., AND KENNAWAY, J. R. A framework for non-manual gestures in a synthetic signing system. *2nd Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Cambridge* (2004), 127–136.

[16] ELLIOTT, R., GLAUERT, J. R. W., KENNAWAY, J. R., AND MARSHALL, I. The development of language processing support for the ViSiCAST project. 101–108.

[17] ELLIOTT, R., GLAUERT, J. R. W., KENNAWAY, J. R., AND MARSHALL, I. The development of language processing support for the ViSiCAST project. 101–108.

[18] EMANUEL, R., MESTRADO, D., AND INFORM, E. PE2LGP: Do Texto à Língua Gestual (e vice-versa).

[19] FERREIRA, M. A. M., AND GARCÍA, L. S. Requirements for Avatar in Virtual Environment of Support Learning in the Literacy of Deaf People in Portuguese Mediated by LIBRAS. 1–6.

[20] GIBET, S. Building French Sign Language Motion Capture Corpora for Signing Avatars.

[21] GIBET, S., COURTY, N., DUARTE, K., NAOUR, T. L., GIBET, S., COURTY, N., DUARTE, K., LE, T., THE, N., SYSTEM, S., GIBET, S., COURTY, N., DUARTE, K., AND NAOUR, T. L. E. The SignCom System for Data-Driven Animation of Interactive Virtual Signers : Methodology and Evaluation.

[22] GIBET, S., COURTY, N., DUARTE, K., NAOUR, T. L., GIBET, S., COURTY, N., DUARTE, K., LE, T., THE, N., SYSTEM, S., GIBET, S., COURTY, N., DUARTE, K., AND NAOUR, T. L. E. The SignCom System for Data-Driven Animation of Interactive Virtual Signers : Methodology and Evaluation.

[23] GIBET, S., LEFEBVRE-ALBARET, F., HAMON, L., BRUN, R., AND TURKI, A. Interactive editing in French Sign Language dedicated to virtual signers: requirements and challenges. *Universal Access in the Information Society 15*, 4 (2016), 525–539.

[24] HANKE, T. 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)* (2004), 1–6.

[25] HELOIR, A., AND VALENCIENNES, F. Towards an Intuitive Sign Language Animation Authoring Environment For the Deaf.

[26] HUENERFAUTH, M., LU, P., AND ROSENBERG, A. Evaluating importance of facial expression in american sign language and pidgin signed english animations. 99.

[27] KACORRI, H., LU, P., AND HUENERFAUTH, M. Evaluating Facial Expressions in American Sign Language Animations for Accessible Online Information.

[28] KAUR, K., AND KUMAR, P. HamNoSys to SiGML Conversion System for Sign Language Automation. *Procedia - Procedia Computer Science 89* (2016), 794–803.

[29] KENNAWAY, R. Synthetic Animation of Deaf Signing Gestures. 146–157.

[30] KIPP, M., NGUYEN, Q., HELOIR, A., AND MATTHES, S. Assessing the deaf user perspective on sign language avatars. 107.

[31] LU, P., AND HUENERFAUTH, M. 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 (2010), 89–97.

[32] LU, P., AND HUENERFAUTH, M. 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 (2010), 89–97.

[33] MAPARI, R. B. Analysis of Multiple Sign Language Recognition Using Leap Motion Sensor.

[34] MARIN, G., DOMINIO, F., AND ZANUTTIGH, P. HAND GESTURE RECOGNITION WITH LEAP MOTION AND KINECT DEVICES. *International Conference on Image Processing(ICIP)* (2014), 1565–1569.

[35] MUHAMMAD, A., ADDENAN, M. F., LATIFF, M. M., HARIS, B., SURIP, S. S., AND MOHAMED, A. S. A. Interactive Sign Language Interpreter using Skeleton Tracking. 137–140.

[36] NEIVA, I. G. S. Desenvolvimento de um tradutor de Língua Gestual Portuguesa. 107.

[37] OTHMAN, A., AND JEMNI, M. An XML-gloss annotation system for sign language processing. *2017 6th International Conference on Information and Communication Technology and Accessbility, ICTA 2017 2017-Decem* (2018), 1–7.

[38] POTTER, L. E., ARAULLO, J., AND CARTER, L. The Leap Motion controller: A view on sign language. *Proceedings Australian Computer-Human Interaction Conference on Augmentation, Application, Innovation, Collaboration - OzCHI '13*, February 2016 (2013), 175–178.

[39] SANNA, A., LAMBERTI, F., PARAVATI, G., AND ROCHA, F. D. A kinect-based interface to animate virtual characters. *Journal on Multimodal User Interfaces 7*, 4 (2013), 269–279.

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

[41] SOARES, M., MEIRELES, B., GONÇALVES, S., ASCHOFF, M., MARITAN, T., AND BECKER, V. A Process for Semi-Automated Construction of Sign Language Dictionaries. *Anais do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres* (2017), 132–136.

[42] SOARES, M., MEIRELES, B., GONÇALVES, S., ASCHOFF, M., MARITAN, T., AND BECKER, V. A Process for Semi-Automated Construction of Sign Language Dictionaries. *Anais do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres* (2017), 132–136.

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

[44] TIBOR GUZSVINECZ, VERONIKA SZUCS, C. S.-L. Suitability of the Kinect Sensor and Leap Motion Controller—A Literature Review.

[45] ZWITSERLOOD, I., VERLINDEN, M., ROS, J., SCHOOT, S. V. D., AND NETHERLANDS, T. SYNTHETIC SIGNING FOR THE DEAF: eSIGN.

A

HamNoSys Hand Shapes

This appendix contains some hand shapes and their respective notation in HamNoSys[1], mentioned in Section 4.4.1.

[1] https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/HamNoSys_Handshapes.pdf

Figure A.1: Some hand configurations described through HamNoSys.

B

HamNoSys Symbols

This appendix contains all the HamNoSys symbols, Unicode codes and their respective SiGML tags, mentioned in Section 4.5.3. The symbols implemented during this project are highlighted

HamNoSys token		HamNoSysUnicode		Old HamNoSys			Comments
hamspace		0020		0020	32		
hamexclaim		0021	!	0021	33	!	
hamcomma		002C	,	0024	36	,	
hamfullstop		002E	.	0023	?35	.	
hamquery		003F	?	0022	34	?	
hamaltbegin		007B	{	00E1	225	{	
hammetaalt		007C	|	0026	38	|	
hamaltend		007D	}	00E2	226	}	
hamfist	D	E000		0032	50		
hamflathand	E	E001		0033	51		
hamfinger2	F	E002		0034	52		
hamfinger23	G	E003		0035	53		
hamfinger23spread	H	E004		0036	54		
hamfinger2345	I	E005		0037	55		
hampinch12	M	E006		003A	58		
hampinchall	N	E007		003B	59		
hampinch12open	O	E008		003C	60		
hamcee12	P	E009		003D	61		
hamceeall	Q	E00A		003E	62		
hamceeopen	R	E00B		003F	63		
hamthumboutmod	J	E00C		0038	56		
hamthumbacrossmod	K	E00D		0039	57		
hamthumbopenmod	L	E00E		0040	64		
hamfingerstraightmod	A	E010		0041	65		
hamfingerbendmod	B	E011		0042	66		

hamfingerhookmod	C	E012		0043	67	
hamdoublebent	6	E013		0045	69	
hamdoublehooked	7	E014		0046	70	
hamextfingeru		E020		0048	72	
hamextfingerur		E021		0049	73	
hamextfingerr		E022		004A	74	
hamextfingerdr		E023		004B	75	
hamextfingerd		E024		004C	76	
hamextfingerdl		E025		004D	77	
hamextfingerl		E026		004E	78	
hamextfingerul		E027		004F	79	
hamextfingerol		E028		0050	80	
hamextfingero		E029		0051	81	
hamextfingeror		E02A		0052	82	
hamextfingeril		E02B		0053	83	
hamextfingeri		E02C		0054	84	
hamextfingerir		E02D		0055	85	
hamextfingerui		E02E		0056	86	
hamextfingerdi		E02F		0057	87	
hamextfingerdo		E030		0058	88	
hamextfingeruo		E031		0059	89	
hampalmu		E038		0060	96	
hampalmur		E039		0061	97	
hampalmr		E03A		0062	98	
hampalmdr		E03B		0063	99	
hampalmd		E03C		0064	100	

hampalmdl	E03D		0065	101		
hampalml	E03E		0066	102		
hampalmul	E03F		0067	103		
hamhead	E040		0071	113		
hamheadtop	E041		0072	114		
hamforehead	E042		0073	115		
hameyebrows	E043		0074	116		
hameyes	E044		0075	117		
hamnose	E045		0076	118		
hamnostrils	E046		005E	94		
hamear	E047		0077	119		
hamearlobe	E048		005D	93		
hamcheek	E049		0078	120		
hamlips	E04A		0079	121		
hamtongue	E04B		006D	109		
hamteeth	E04C		006E	110		
hamchin	E04D		007A	122		
hamunderchin	E04E		007B	123		
hamneck	E04F		007C	124		
hamshouldertop	E050		005F	95		
hamshoulders	E051		007D	125		
hamchest	E052		007E	126		
hamstomach	E053		006F	111		
hambelowstomach	E054		0080	128		(Also 00EE)
hamlrbeside	E058		0081	129		(Also 00EF)
hamlrat	E059		[201A]	130		

hamcoreftag		E05A		[00DB]	158		
hamcorefref		E05B		[0178]	159		
hamneutralspace		E05F		0070	112		
hamupperarm		E060		[0192]	131		
hamelbow		E061		[201E]	132		
hamelbowinside		E062		[2026]	133		
hamlowerarm		E063		[2020]	134		
hamwristback		E064		[2021]	135		
hamwristpulse		E065		[02C6]	136		
hamthumbball		E066		[2030]	137		
hampalm		E067		[0160]	138		
hamhandback		E068		[2039]	139		
hamthumbside		E069		[2018]	145		
hampinkyside		E06A		[2019]	146		
hamthumb	1	E070		[0152]	140		
hamindexfinger	2	E071		[00E8]	141		
hammiddlefinger	3	E072		[00E9]	142		
hamringfinger	4	E073		[00EA]	143		
hampinky	5	E074		[00EB]	144		
hamfingertip	T	E075		[201D]	148		
hamfingernail	U	E076		[2022]	149		
hamfingerpad	V	E077		[2013]	150		
hamfingermidjoint	W	E078		[2014]	151		
hamfingerbase	X	E079		[02DC]	152		
hamfingerside	Y	E07A		[2122]	153		
hamwristtopulse		E07C		[0161]	154		(OBSOLETE)

hamwristtoback	E07D	[203A]	155	(OBSOLETE)
hamwristtothumb	E07E		156	(OBSOLETE)
hamwristtopinky	E07F		157	(OBSOLETE)
hammoveu	E080		161	
hammoveur	E081		162	
hammover	E082		163	
hammovedr	E083		164	
hammoved	E084		165	
hammovedl	E085		166	
hammovel	E086		167	
hammoveul	E087		168	
hammoveol	E088		169	
hammoveo	E089		170	
hammoveor	E08A		171	
hammoveil	E08B		172	
hammovei	E08C		173	
hammoveir	E08D		174	
hammoveui	E08E		175	
hammovedi	E08F		176	
hammovedo	E090		177	
hammoveuo	E091		178	
hamcircleo	E092		192	
hamcirclei	E093		193	
hamcircled	E094		194	
hamcircleu	E095		195	
hamcirclel	E096		196	

hamcircler	E097	197	
hamcircleul	E098	240	
hamcircledr	E099	241	
hamcircleur	E09A	242	
hamcircledl	E09B	243	
hamcircleol	E09C	244	
hamcircleir	E09D	245	
hamcircleor	E09E	246	
hamcircleil	E09F	247	
hamcircleui	E0A0	249	
hamcircledo	E0A1	248	
hamcircleuo	E0A2	250	
hamcircledi	E0A3	251	
hamfingerplay	E0A4	189	
hamnodding	E0A5	227	
hamswinging	E0A6	228	
hamtwisting	E0A7	229	
hamstircw	E0A8	230	
hamstirccw	E0A9	231	
hamreplace	E0AA	104	
hammovecross	E0AD	179	(OBSOLETE)
hammoveX	E0AE	180	(OBSOLETE)
hamnomotion	E0AF	160	
hamclocku	E0B0	39	
hamclockul	E0B1	40	
hamclockl	E0B2	41	

hamclockdl	E0B3	42
hamclockd	E0B4	43
hamclockdr	E0B5	44
hamclockr	E0B6	45
hamclockur	E0B7	46
hamclockfull	E0B8	47
hamarcl	E0B9	183
hamarcu	E0BA	184
hamarcr	E0BB	185
hamarcd	E0BC	186
hamwavy	E0BD	187
hamzigzag	E0BE	188
hamellipseh	E0C0	220
hamellipseur	E0C1	221
hamellipsev	E0C2	222
hamellipseul	E0C3	223
hamincreasing	E0C4	198
hamdecreasing	E0C5	199
hamsmallmod	E0C6	181
hamlargemod	E0C7	182
hamfast	E0C8	204
hamslow	E0C9	205
hamtense	E0CA	206
hamrest	E0CB	207
hamhalt	E0CC	208
hamclose	E0D0	200

hamtouch		E0D1		201
haminterlock		E0D2		202
hamcross		E0D3	×	203
hamarmextended		E0D4		105
hambehind		E0D5		106
hambrushing		E0D6		217
hamrepeatfromstart		E0D8	+	209
hamrepeatfromstartseveral		E0D9		210
hamrepeatcontinue		E0DA		211
hamrepeatcontinueseveral		E0DB		212
hamrepeatreverse		E0DC		216
hamalternatingmotion		E0DD	∼	215
hamseqbegin		E0E0	(	213
hamseqend		E0E1	)	214
hamparbegin		E0E2	[	190
hamparend		E0E3	]	191
hamfusionbegin		E0E4	⟨	236
hamfusionend		E0E5	⟩	237
hambetween	S	E0E6	\	147
hamplus		E0E7		37
hamsymmpar		E0E8	:	48
hamsymmlr		E0E9	..	49
hamnondominant		E0EA		68
hamnonipsi		E0EB		218
hametc		E0EC	...	107
hamorirelative		E0ED	~	108

hammime	E0F0		224
hamversion40	E0F1		20
hamnbs			254
linefeed			10
pagebreak			12
return			13
tab			9
UNUSED			0
UNUSED			1
UNUSED			2
UNUSED			3
UNUSED			4
UNUSED			5
UNUSED			6
UNUSED			7
UNUSED			8
UNUSED			11
UNUSED			14
UNUSED			15
UNUSED			16
UNUSED			17
UNUSED			18
UNUSED			19
UNUSED			21
UNUSED			22
UNUSED			23

UNUSED	24
UNUSED	25
UNUSED	26
UNUSED	27
UNUSED	28
UNUSED	29
UNUSED	30
UNUSED	31
UNUSED	71
UNUSED	90
UNUSED	91
UNUSED	92
UNUSED	219
UNUSED	232
UNUSED	233
UNUSED	234
UNUSED	235
UNUSED	238
UNUSED	239
UNUSED	252
UNUSED	253
UNUSED	255

C

HTML File

This appendix contains an example of an HTML file exported from ELAN, described in Section 4.5.2.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/1999/REC-html401-19991224/loose.dtd">
<html>
<head>
<meta http-equiv="content-type" content="text/html;charset=utf-8"/>
<title>TEST.eaf</title>
<style type="text/css" media="screen">
body {
background-color: #FFFFF2;
font-family: "Arial Unicode MS", Verdana, Helvetica, sans-serif;
font-weight: normal;
font-size: 12px;
color: #000000;
line-height: 15px;
font-style: normal;
}
table {
width: auto; /* change to 100% to horizontally stretch the table */
border-collapse: collapse;
border: 1px solid #666666;
}
table.out {
border: 0px;
width: 100%;
}
tr {
border-collapse: collapse;
border: 1px solid #dddddd;
}
tr.out {
border: 0px;
```

```
}
td {
padding: 2px 8px 2px 2px;
border-collapse: collapse;
border: 1px solid #666666;
}
td.out {
border: 0px;
}
td.label {
width: 68;
font-family: "Arial Unicode MS", Verdana, Helvetica, sans-serif;
font-weight: bold;
font-size: 12px;
color: #444444;
}
td.tclabel {
width: 68;
font-family: "Arial Unicode MS", Verdana, Helvetica, sans-serif;
font-size: 11px;
font-weight: bold;
color: #9E001C;
}
td.tc {
font-family: "Arial Unicode MS", Verdana, Helvetica, sans-serif;
font-size: 11px;
font-weight: bold;
color: #9E001C;
}
td.sdlabel {
width: 68;
```

```
font-family: "Arial Unicode MS", Verdana, Helvetica, sans-serif;

font-size: 11px;

font-weight: bold;

color: #9E001C;

}

td.sd {

font-family: "Arial Unicode MS", Verdana, Helvetica, sans-serif;

font-size: 11px;

font-weight: bold;

color: #9E001C;

}

td.hide {

border: 0px;

}

/* a style class for each visible tier */

tr.ti-0 { /* Glosses */

font-family: "MS Arial Unicode";

font-size: 12px;

}

tr.ti-1 { /* HamNoSys */

font-family: "Hamnosysunicode";

font-size: 12px;

}

</style>

</head>

<body>

<h3>file:///C:/Users/Carol/Desktop/LREC/ELAN/TEST.eaf</h3>

<p>2020 May 22, Fri 15:38</p>


<table class="out">

<tr class="out"><td class="out">
```

```
<table>
<tr class="ti-0"><td class="label">Glosses</td><td colspan="1">MULHER</td></tr>
<tr class="ti-1"><td class="label">HamNoSys</td><td colspan="1">ᑯ‸⦸⦄▪᙮↓</td></tr>
</table>
</td></tr>
<tr class="out"><td class="out"> <br><br></td></tr>
<tr class="out"><td class="out">
<table>
<tr class="ti-0"><td class="label">Glosses</td><td colspan="1">CASA</td></tr>
<tr class="ti-1"><td class="label">HamNoSys</td><td
colspan="1">⬭ᒥ0▪ᗜ↣ᒥ0▪ᗜ᙮↣ᒥ0▪ᗜ↣ᒥ0▪ᗜ᙮</td></tr>
</table>
</td></tr>
<tr class="out"><td class="out"> <br><br></td></tr>
<tr class="out"><td class="out">
<table>
<tr class="ti-0"><td class="label">Glosses</td><td colspan="1">BRANCA</td></tr>
<tr class="ti-1"><td class="label">HamNoSys</td><td
colspan="1">▢‹⦸◡᙮↣⌀][↣⦸◡᙮↣⌀][</td></tr>
</table>
</td></tr>
<tr class="out"><td class="out"> <br><br></td></tr>
<tr class="out"><td class="out">
<table>
<tr class="ti-0"><td class="label">Glosses</td><td colspan="1">DELE</td></tr>
<tr class="ti-1"><td class="label">HamNoSys</td><td colspan="1">⬭‸⊖ᗜ⊥</td></tr>
</table>
</td></tr>
<tr class="out"><td class="out"> <br><br></td></tr>
</table>
</body>
</html>
```

D

Text File

This appendix contains an example of a text file containing the glosses and their respective HamNoSys codes, explained in Section 4.5.2.

MULHER, E002 E00D E020 E03D E049 E059 E0D1 E084 E0C6

CASA, E000 E00D E027 E03E E059 E051 E0AA E027 E03E E059 E051 E0D1 E0AA E027 E03E E059 E051 E0AA E027 E03E E059 E051 E0D1

BRANCA, E001 E00C E026 E03D E04D E0D1 E0AA E03B E04F E0AA E03D E04D E0D1 E0AA E03B E04F

DELE, E000 E00D E020 E03C E051 E089

E

SiGML File

This appendix contains an example of a SiGML use to animate our avatar, mentioned in Section 4.5.3.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<sigml>
	<hns_sign gloss="MULHER">
		<hamnosys_nonmanual/>
		<hamnosys_manual>
			<hamfinger2 hamnosys="E002"/>
			<hamthumbacrossmod hamnosys="E00D"/>
			<hamextfingeru hamnosys="E020"/>
			<hampalmdl hamnosys="E03D"/>
			<hamcheek hamnosys="E049"/>
			<hamlrat hamnosys="E059"/>
			<hamtouch hamnosys="E0D1"/>
			<hammoved hamnosys="E084"/>
			<hamsmallmod hamnosys="E0C6"/>
		</hamnosys_manual>
	</hns_sign>
	<hns_sign gloss="CASA">
		<hamnosys_nonmanual/>
		<hamnosys_manual>
			<hamfist hamnosys="E000"/>
			<hamthumbacrossmod hamnosys="E00D"/>
			<hamextfingerul hamnosys="E027"/>
			<hampalml hamnosys="E03E"/>
			<hamlrat hamnosys="E059"/>
			<hamshoulders hamnosys="E051"/>
			<hamreplace hamnosys="E0AA"/>
			<hamextfingerul hamnosys="E027"/>
			<hampalml hamnosys="E03E"/>
			<hamlrat hamnosys="E059"/>
			<hamshoulders hamnosys="E051"/>
			<hamtouch hamnosys="E0D1"/>
```

```
				<hamreplace hamnosys="E0AA"/>
				<hamextfingerul hamnosys="E027"/>
				<hampalml hamnosys="E03E"/>
				<hamlrat hamnosys="E059"/>
				<hamshoulders hamnosys="E051"/>
				<hamreplace hamnosys="E0AA"/>
				<hamextfingerul hamnosys="E027"/>
				<hampalml hamnosys="E03E"/>
				<hamlrat hamnosys="E059"/>
				<hamshoulders hamnosys="E051"/>
				<hamtouch hamnosys="E0D1"/>
			</hamnosys_manual>
		</hns_sign>
		<hns_sign gloss="BRANCA">
			<hamnosys_nonmanual/>
			<hamnosys_manual>
				<hamflathand hamnosys="E001"/>
				<hamthumboutmod hamnosys="E00C"/>
				<hamextfingerl hamnosys="E026"/>
				<hampalmdl hamnosys="E03D"/>
				<hamchin hamnosys="E04D"/>
				<hamtouch hamnosys="E0D1"/>
				<hamreplace hamnosys="E0AA"/>
				<hampalmdr hamnosys="E03B"/>
				<hamneck hamnosys="E04F"/>
				<hamreplace hamnosys="E0AA"/>
				<hampalmdl hamnosys="E03D"/>
				<hamchin hamnosys="E04D"/>
				<hamtouch hamnosys="E0D1"/>
				<hamreplace hamnosys="E0AA"/>
				<hampalmdr hamnosys="E03B"/>
```

```xml
                <hamneck hamnosys="E04F"/>
            </hamnosys_manual>
        </hns_sign>
        <hns_sign gloss="DELE">
            <hamnosys_nonmanual/>
            <hamnosys_manual>
                <hamfist hamnosys="E000"/>
                <hamthumbacrossmod hamnosys="E00D"/>
                <hamextfingeru hamnosys="E020"/>
                <hampalmd hamnosys="E03C"/>
                <hamshoulders hamnosys="E051"/>
                <hammoveo hamnosys="E089"/>
            </hamnosys_manual>
        </hns_sign>
    </sigml>
```

F

Data Serialized

This appendix contains an example of a text file with the data received to save serialized explained in Section 4.7.2.

```
<?xml version="1.0" encoding="utf-8"?>
<Document>
  <Annotations Gloss="CASA">
    <CodesSet>
      <HamNoSys code="⬭" />
      <HamNoSys code="" />
      <HamNoSys code="‹" />
      <HamNoSys code="0" />
      <HamNoSys code="▪" />
      <HamNoSys code="ᗜ" />
      <HamNoSys code=")(" />
    </CodesSet>
  </Annotations>
  <Annotations Gloss="EU">
    <CodesSet>
      <HamNoSys code="⅃" />
      <HamNoSys code="ˏ" />
      <HamNoSys code="⏶" />
      <HamNoSys code="⌓" />
      <HamNoSys code="ᗜ" />
      <HamNoSys code="↑" />
      <HamNoSys code="↓" />
      <HamNoSys code="" />
    </CodesSet>
  </Annotations>
  <Annotations Gloss="BOM">
    <CodesSet>
      <HamNoSys code="⊲" />
      <HamNoSys code="" />
      <HamNoSys code="0" />
      <HamNoSys code="⊏" />
```

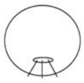
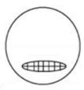
```
    <HamNoSys code="⌔" />
    <HamNoSys code="▪" />
  </CodesSet>
  <CodesSet>
    <HamNoSys code="◅" />
    <HamNoSys code="" />
    <HamNoSys code="ₒ" />
    <HamNoSys code="⌐" />
    <HamNoSys code="ℴ" />
    <HamNoSys code="⌔" />
    <HamNoSys code="▪" />
  </CodesSet>
  <CodesSet>
    <HamNoSys code="⬭" />
    <HamNoSys code="ˏ" />
    <HamNoSys code="‹" />
    <HamNoSys code="ₒ" />
    <HamNoSys code="◡" />
  </CodesSet>
</Annotations>
<Annotations Gloss="OUVINTE">
  <CodesSet>
    <HamNoSys code="ɗ" />
    <HamNoSys code="" />
    <HamNoSys code="‸" />
    <HamNoSys code="⌐" />
    <HamNoSys code="ₒ" />
    <HamNoSys code="ʔ" />
    <HamNoSys code="▪" />
    <HamNoSys code="→" />
    <HamNoSys code="" />
```

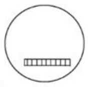
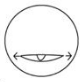
```
    </CodesSet>
  </Annotations>
  <Annotations Gloss="DIA">
    <CodesSet>
      <HamNoSys code="⌂" />
      <HamNoSys code="ˏ" />
      <HamNoSys code="‹" />
      <HamNoSys code="0" />
      <HamNoSys code="◡" />
      <HamNoSys code="↑" />
      <HamNoSys code="ʕ" />
      <HamNoSys code="↣" />
      <HamNoSys code="‸" />
      <HamNoSys code="⬭" />
    </CodesSet>
  </Annotations>
</Document>
```

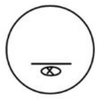
G

Facial Expressions Codes and Symbols

This appendix contains the facial expressions symbols and their respective SiGML codes created by the interpreter of the project, mentioned in Section 5.1.

HamNoSys token	HamNoSysUnicode	HamNoSysUnicode	Comments
faceNeutral	F000		
faceRightCheekFull	F001 1. 		
faceBreath	F002 2. 		
faceLeftCheekFull	F003 3. 		
faceOpenMouthA	F004 4. 		
faceBothCheekFull	F005 5. 		
faceOpenMouthO	F006 6. 		
faceTongueInCheek	F007 7. 		
faceMouthOX	F008 8. 		
faceTongueOut	F009 9. 		
faceMamama	F010 10. 		
faceClenchedTeeth	F011 11. 		

faceTeethBiteLip	F012	12.	
faceTeethFromAbove	F013	13.	
faceRepeatedTongue	F014	14.	
faceSad	F015	15.	
faceHappy	F016	16.	
faceBrrrr	F017	17.	
faceHalfClosedEyes	F018	18.	
faceContractedEyebrows	F019	19.	
faceClosedEyes	F020	20.	
faceEyebrowsUp	F021	21.	
faceLookDown	F022	22.	
faceLookUp	F023	23.	

faceTongueOnChin F024 24. 

faceLipUp F025 25. 

PS:. Os bonecos estão
desenhados em efeito espelho

H

User Tests

This appendix contains the questionnaire used to evaluate our tool, described in Section 5.1.

Avaliação de avatares

No âmbito do Mestrado em Engenharia Informática e Computadores do Instituto Superior Técnico foi desenvolvida uma ferramenta de apoio a linguistas para a animação de gestos anotados em língua gestual portuguesa.

Por forma a avaliar como estes avatares gestuam pedimos a sua colaboração. Este questionário não tomará mais que 15 minutos do seu tempo e as suas respostas permanecerão anónimas, sendo usadas exclusivamente para a avaliação dos avatares. O questionário deve ser feito com a sua conta email uma vez que é disponibilizada a opção de serem carregados vídeos com conteúdo em Língua Gestual Portuguesa.

Esta avaliação consiste na visualização de 20 vídeos por 2 avatares distintos, 10 vídeos cada avatar. Estes avatares não tem expressões faciais, pelo que pedimos que se foque exclusivamente no movimento de braços e mãos.

Agradecemos desde já a sua disponibilidade e colaboração.

* Required

1. Com que frequência usa língua gestual portuguesa? *

   Mark only one oval.

   - Todos os dias
   - 2 a 3 vezes por semana
   - 1 vez por semana
   - 1 vez por mês
   - 1 vez por ano
   - Nunca

2. Já usou um avatar para gestuar língua gestual portuguesa? *

Mark only one oval.

- ( ) Bastantes vezes
- ( ) Algumas vezes
- ( ) Raramente
- ( ) Nunca

3. É surdo? *

Mark only one oval.

- ( ) Sim
- ( ) Não

4. Qual a sua relação com a língua gestual portuguesa? *

Mark only one oval.

- ( ) Nativo
- ( ) Outro

5. Se respondeu Outro na última pergunta, há quanto tempo aprendeu esta língua?

____

6. Selecione a variante regional com que mais se identifica? *

Check all that apply.

- [ ] Norte
- [ ] Centro
- [ ] Sul
- [ ] Açores
- [ ] Madeira
- [ ] Outro

7. Qual a sua idade? *

____

8. Qual é o seu género? *

Mark only one oval.

- ◯ Feminino
- ◯ Masculino
- ◯ Outro
- ◯ Prefiro não dizer

Avaliação da animação de gestos 1-20

Frase 1 de 10 - Avatar Anna

9. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

10. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

11. Comentários adicionais (Língua Portuguesa).

12. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 2-20

Frase 2 de 10 - Avatar Anna

13. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

14. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

15. Comentários adicionais (Língua Portuguesa).

16. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 3-20

Frase 3 de 10 - Avatar Anna

17. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

18. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

19. Comentários adicionais (Língua Portuguesa).

20. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 4-20

Frase 4 de 10 - Avatar Anna

21. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

22. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

23. Comentários adicionais (Língua Portuguesa).

24. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 5-20

Frase 5 de 10 - Avatar Anna

25. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

26. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- ( ) 1
- ( ) 2
- ( ) 3
- ( ) 4
- ( ) 5
- ( ) Mais de 5

27. Comentários adicionais (Língua Portuguesa).

____

____

____

____

____

28. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 6-20

Frase 6 de 10 - Avatar Anna

29. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

30. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

31. Comentários adicionais (Língua Portuguesa).

32. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 7-20

Frase 7 de 10 - Avatar Anna

33. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

34. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

35. Comentários adicionais (Língua Portuguesa).

36. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 8-20

Frase 8 de 10 - Avatar Anna

37. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

38. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

39. Comentários adicionais (Língua Portuguesa).

40. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 9-20

Frase 9 de 10 - Avatar Anna

41. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

42. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

43. Comentários adicionais (Língua Portuguesa).

44. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 10-20

Frase 10 de 10 - Avatar Anna

45. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

46. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

47. Comentários adicionais (Língua Portuguesa).

48. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação geral do avatar Anna

49. Avalie as frases anteriores nas seguintes categorias: *

Mark only one oval per row.

	Não sei	1 (Péssimo)	2	3	4	5 (Perfeito)
Velocidade	◯	◯	◯	◯	◯	◯
Qualidade geral	◯	◯	◯	◯	◯	◯
Compreensibilidade	◯	◯	◯	◯	◯	◯
Naturalidade	◯	◯	◯	◯	◯	◯
Correção gramatical	◯	◯	◯	◯	◯	◯
Configuração das mãos	◯	◯	◯	◯	◯	◯
Orientação das mãos	◯	◯	◯	◯	◯	◯
Localização das mãos	◯	◯	◯	◯	◯	◯
Movimento das mãos	◯	◯	◯	◯	◯	◯

Avaliação da animação de gestos 11-20

Frase 1 de 10 - Avatar Catarina

50. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

51. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

52. Comentários adicionais (Língua Portuguesa).

53. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 12-20

Frase 2 de 10 - Avatar Catarina

54. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

55. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- ( ) 1
- ( ) 2
- ( ) 3
- ( ) 4
- ( ) 5
- ( ) Mais de 5

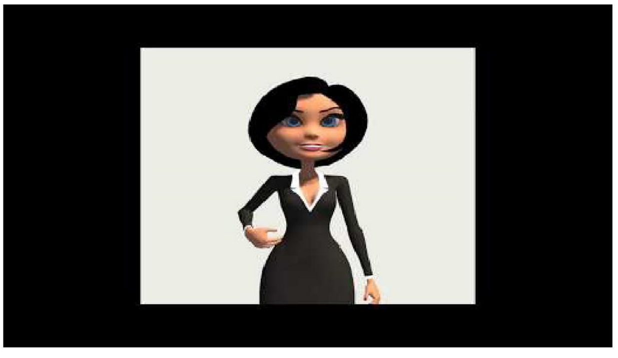
56. Comentários adicionais (Língua Portuguesa).

57. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 13-20

Frase 3 de 10 - Avatar Catarina



58. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

59. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

60. Comentários adicionais (Língua Portuguesa).

61. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 14-20

Frase 4 de 10 - Avatar Catarina

62. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

63. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

64. Comentários adicionais (Língua Portuguesa).

65. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 15-20

Frase 5 de 10 - Avatar Catarina

66. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

67. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

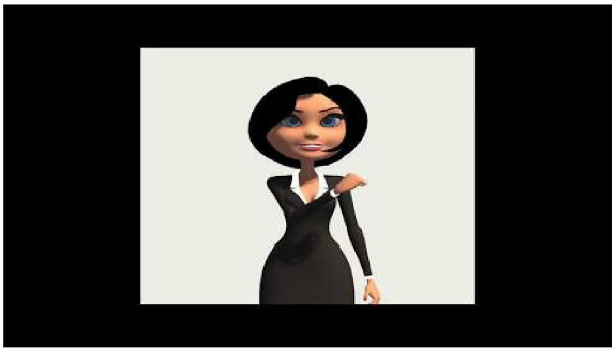
68. Comentários adicionais (Língua Portuguesa).

69. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 16-20

Frase 6 de 10 - Avatar Catarina



70. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

71. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

72. Comentários adicionais (Língua Portuguesa).

73. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 17-20

Frase 7 de 10 - Avatar Catarina

74. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

75. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

76. Comentários adicionais (Língua Portuguesa).

77. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 18-20

Frase 8 de 10 - Avatar Catarina

78. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

79. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

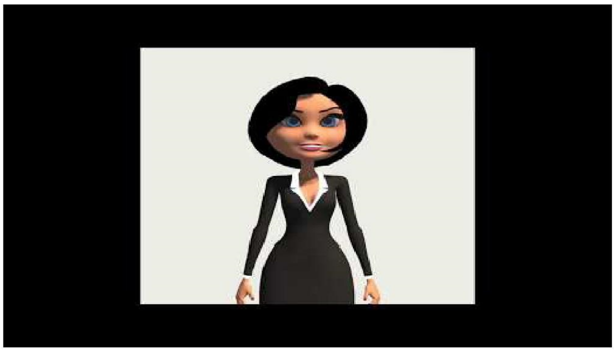
80. Comentários adicionais (Língua Portuguesa).

81. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 19-20

Frase 9 de 10 - Avatar Catarina



82. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

83. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

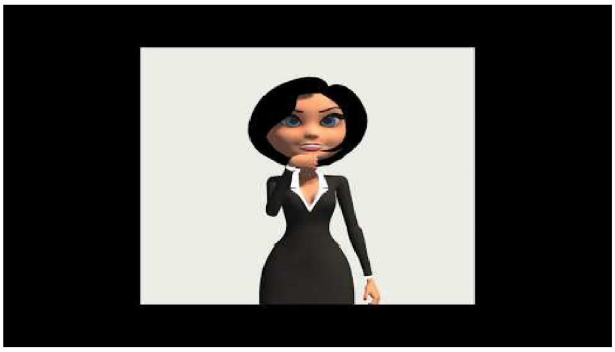
84. Comentários adicionais (Língua Portuguesa).

85. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação da animação de gestos 20-20

Frase 10 de 10 - Avatar Catarina



86. Visualize o vídeo e escreva a frase reproduzida pelo avatar em língua gestual portuguesa. *

87. Quantas vezes precisou de ver o vídeo? *

Mark only one oval.

- 1
- 2
- 3
- 4
- 5
- Mais de 5

88. Comentários adicionais (Língua Portuguesa).

89. Comentários adicionais (Língua Gestual Portuguesa). Por favor faça upload do seu vídeo a gestuar.

Files submitted:

Avaliação geral do avatar Catarina

90. Avalie as frases anteriores nas seguintes categorias: *

Mark only one oval per row.

	Não sei	1 (Péssimo)	2	3	4	5 (Perfeito)
Velocidade	◯	◯	◯	◯	◯	◯
Qualidade geral	◯	◯	◯	◯	◯	◯
Compreensibilidade	◯	◯	◯	◯	◯	◯
Naturalidade	◯	◯	◯	◯	◯	◯
Correção gramatical	◯	◯	◯	◯	◯	◯
Configuração das mãos	◯	◯	◯	◯	◯	◯
Orientação das mãos	◯	◯	◯	◯	◯	◯
Localização das mãos	◯	◯	◯	◯	◯	◯
Movimento das mãos	◯	◯	◯	◯	◯	◯

Fim

I

User test phrases

This appendix contains the table with all the comprehension results from the user tests, from what was understood, its percentage and the number of visualizations required for each avatar, discussed in Section 5.2.2.

Table I.2: Phrases evaluated for each avatar, the content understood, its respective percentage and number of visualizations by all testers.

Phrase Nr	Content signed (PSL)	ANNA What was understood	ANNA % of content understood	ANNA Nr of Visualizations	CAT What was understood	CAT % of content understood	CAT Nr of Visualizations
1	BOM DIA	bom dia	100	2	????	0	1
		Não percebi	0	mais de 5	Bom dia	100	3
		bom dia	100	1	”Bom dia!”	100	1
		Dia	50	3	.	.	.
2	MENINA OCULOS FLOR VER	menina flor ver	75	1	menina óculos flor dois	75	2
		Flor V	50	2	.	.	.
		O menino viu uma flor.	50	mais de 5	.	.	.
		Menina jovem viu uma flor????????	75	4	.	.	.
		???	0	3	.	.	.
		”A menina flores duas.”	50	1	.	.	.
	TU ANIMAL GOSTAR	animal gostas?	100	2	????	0	1
		???	0	3	Tu gostas de animais	100	2
		Gosto de... animal????	66,66	mais de 5	Não percebi.	0	mais de 5
		Só entendi a palavra ”gostar”.	33,33	3	.	.	.
4	(GUARDA-) SEGURANÇA RESPEITO QUERER	Não percebi...	0	2	pessoas respeito querer	66,66	2
		Cinto... respeito...	33,33	mais de 5	Não percebi.	0	mais de 5
		Qualquer coisa cansada ?????????	0	mais de 5	”Segurança respeitar quer”	100	1
		.	.	.	???	0	3
5	CRIANÇA IN-TELIGENTE	menino ??	50	2	O menino é inteligente.	100	3
		não percebi	0	3	Jovem qualquer coisa???	0	3

Continued on next page

Table I.2 – Continued from previous page

Phrase Nr	Content signed (PSL)	ANNA What was understood	ANNA % of content under-stood	ANNA Nr of Visual-iza-tions	CAT What was understood	CAT % of content under-stood	CAT Nr of Visu-aliza-tions
		Não sei	0	mais de 5	”Criança inteligente”	100	1
		????	0	3	.	.	.
6	DIRECTOR DINHEIRO PEDIR	Não percebi.	0	mais de 5	Cortar cabelo pouco por favor	0	2
		???	0	4	????	0	1
		”dinheiro”.	33,33	2	Qualquer coisa.... + obrigada	0	5
		.	.	.	desculpa... faltar???	0	5
7	NAMORADO MEU OLHOS AZUIS	.	.	.	namorado meu olhos azuis	100	2
		.	.	.	O meu namorado tem olhos azuis	100	1
		.	.	.	Os meus olhos são azuis?!	75	3
		.	.	.	Namorado? Meu olhos azuis	100	5
		.	.	.	???	0	2
		.	.	.	”Namorado meu olhos azuis”.	100	1
		.	.	.	Olhos abertos? Abrir os olhos?	25	4
8	NETO FORMIGA COMER	doce comer	33,33	3	Não percebi	0	mais de 5
		”comer”	33,33	3	???	0	1
		Neto telefonar comer...	66,66	mais de 5	”Perguntar formiga comer”	66,66	2
		.	.	.	Não sei	0	mais de 5

Continued on next page

Table I.2 – Continued from previous page

Phrase Nr	Content signed (PSL)	ANNA What was understood	ANNA % of content understood	ANNA Nr of Visualizations	CAT What was understood	CAT % of content understood	CAT Nr of Visualizations
9	PAPA OUVINTE BOM	.	.	.	Papa ouvir telemóvel??	66,66	4
		.	.	.	????	0	1
		.	.	.	Braga telefone	0	1
		.	.	.	Não percebi.	0	mais de 5
		.	.	.	???	0	1
		.	.	.	"Papa ouvinte bom"	100	2
		.	.	.	?	0	4
10	ALI SOL POUCO	??? pouco	33,33	3	???	0	1
		puxar...XXX Não percebo esse gesto	0 0	mais de 5 3	Não percebi nada. .	0 .	mais de 5 .
		"pouco".	33,33	3	.	.	.
11	ESTADO PALAVRA TER	Não percebo...	0	mais de 5	governo pessoa vem aqui	33,33	2
		Não percebi.	0	mais de 5	????	0	1
		.... vocês aqui...???	0	mais de 5	???	0	3
		"ter"	33,33	3	.	.	.
12	PAI FILHO ZANGAR	pai e filho zangam-se	100	1	???	0	1
		menino mãe zangada.	33,33	mais de 5	"Pai zangado"	66,66	mais de 5
		O pai e a mãe estão zangados.	66,66	4	.	.	.
		Como estão pai e mãe?	33,33	2	.	.	.
		O meu filho está furioso ou o pai do filho está furioso ??????????	66,66	mais de 5	.	.	.
13	EU TU PERGUNTAR	??? e	0	5	???	0	2

Continued on next page

Table I.2 – Continued from previous page

Phrase Nr	Content signed (PSL)	ANNA What was understood	ANNA % of content under-stood	ANNA Nr of Visual-iza-tions	CAT What was understood	CAT % of content under-stood	CAT Nr of Visu-aliza-tions
		Buscar (copo)????	0	3	.	.	.
		Eu beber.. buscar copo.. eu???	33,33	mais de 5	.	.	.
		Não percebi.	33,33	mais de 5	.	.	.
		Bebida armário ???????????	0	mais de 5	.	.	.
		Não entendi.	0	mais de 5	.	.	.
14	MÉDICO HISTÓRIA OUVIR	??? história ouvir	66,66	2	Não percebo...	0	mais de 5
		Não percebi...	0	2	?	0	5
		Não percebi.	0	mais de 5	.	.	.
		???	0	1	.	.	.
		”Pode ouvir.”	33,33	2	.	.	.
15	MULHER CEGA SIMPÁTICA	Não percebi...	0	2	mulher cega xxxx	66,66	2
		A mulher viu e ficou desiludida.	33,33	mais de 5	menina cega....XXX	66,66	3
		.	.	.	Menina olhos	0	mais de 5
		.	.	.	???	0	2
		.	.	.	”Mulher cega”	66,66	3
16	JOVEM MAÇÃ QUERER	jovens ??? querer	66,66	3	Não percebi.	0	mais de 5
		Não percebi a frase	0	2	???	0	2
		Não percebo....	0	mais de 5	Jovens....???	33,33	4
		”Jovem, pequeno-almoço.”	33,33	2	.	.	.

Continued on next page

Table I.2 – Continued from previous page

Phrase Nr	Content signed (PSL)	ANNA What was understood	ANNA % of content under-stood	ANNA Nr of Visual-iza-tions	CAT What was understood	CAT % of content under-stood	CAT Nr of Visu-aliza-tions
17	MULHER CASA BRANCO DELE	”mulher” ”Casa” ”tua”	50	4	mulher casa branca sua	75	3
		A casa e o cão na porta?!	25	mais de 5	Menina Casa XXX tua	50	4
		.	.	.	Qual a tua idade?	0	5
		????	0	2	”Senhora empresta”	25	3
18	ELES MASSA COMER	Vocês cor-de-rosa comer...	33,33	mais de 5	Tu ???? comer	0	3
		”Nós fruta comer.”	33,33	4	Qualquer coisa... pessoa???	0	5
		.	.	.	Não percebi.	0	mais de 5
19	EU FILHO MEU VER	Filho meu V	75	3	Eu filho meu ver	100	2
		Vi a minha mãe.	25	mais de 5	Não percebo...	0	mais de 5
		O meu filho viu ?????????	75	4	???	0	3
		”Filhos meus dois.”	50	2	.	.	.
20	ALEMANHA PAÍS DELE	Eu beber buscar copo eu ....????	0	mais de 5	seu	33,33	2
		bebida armário ????????	0	mais de 5	Não percebi.	0	mais de 5
		.	.	.	???	0	2
		.	.	.	???	0	1
		.	.	.	”País teu”	33,33	2