PE2LGP Animator: A Tool To Animate A Portuguese Sign Language Avatar

Pedro Bertrand Cabral, Matilde Gonçalves, Ruben dos Santos, Hugo Nicolau, Luisa Coheur

Instituto Superior Técnico, Universidade de Lisboa/INESC-ID

{pedro.b.cabral, matilde.do.carmo.lages.goncalves, luisa.coheur, hugo.nicolau}@tecnico.ulisboa.pt, soltex@me.com

Abstract

Software for the production of sign languages is much less common than for spoken languages. Such software usually relies on 3D humanoid avatars to produce signs which, inevitably, necessitates the use of animation. One barrier to the use of popular animation tools is their complexity and steep learning curve, which can be hard to master for inexperienced users. Here, we present PE2LGP, an authoring system that features a 3D avatar that signs Portuguese Sign Language. Our Animator is designed specifically to craft sign language animations using a key frame method, and is meant to be easy to use and learn to users without animation skills. We conducted a preliminary evaluation of the Animator, where we animated seven Portuguese Sign Language sentences and asked four sign language users to evaluate their quality. This evaluation revealed that the system, in spite of its simplicity, is indeed capable of producing comprehensible messages.

1. Introduction

PE2LGP is a project to digitalise *Portuguese Sign Language* (shortened to *LGP – Língua Gestual Portuguesa*), the primary language of the Deaf community in Portugal, through a 3D avatar capable of communicating in it.

Though a living language used by thousands of people, LGP is still largely understudied, with both an absence of linguistic research on it (compared to widely used spoken languages) and a lack of tools and resources for its computational processing. We aim to provide one such resource through this project, in this case through the Animator, a tool that allows users without technical knowledge or animation expertise to create animations of LGP signs for our avatar to use through simple frame-by-frame posing. As part of the larger effort to improve digital support for LGP, the Animator could be used for chatbots, virtual assistants, dictionaries, or, as we have done in PE2LGP, automatic translators.

The motivation for this study was to not only expose this project to the community, but to gain greater intuition of the tool's current performance. We thus present in this paper a description of our Animator and an overview of the role it plays in the greater scope of LGP and sign language research, along with a preliminary study of its capabilities, where, using our tool, we animated seven LGP sentences and asked four users of the language to interpret them and give an appreciation of their quality.

2. State of the Art

Sign languages are visuospatial languages, i.e., the communication is performed using signs produced at determined locations in the three-dimensional space or on the body. Because of the deafs' linguistic isolation, they must sometimes resort to human interpreters, who are not always available, so alternative translation systems are useful. Sign languages have no widely-used written forms (Kaur and Kumar, 2014), so such systems require the representation of a human body to produce messages, such as videos of human signers or 3D avatars.

However, videos of human signers have serious limitations when stringing together signs to compose sentences (Huenerfauth and Hanson, 2009). First, because each sign would have to be recorded in its own video, it would not be possible to have smooth movement from one sign to the next: the signer would have to return to a neutral position and there would likely always be seams. Furthermore, the videos would have to be recorded with the same signer under similar conditions, to avoid abrupt visual changes. Finally, this approach would not allow different signs to be combined: in LGP, for example, a facial expression may be combined with other signs to mark a sentence as interrogative. An avatar-based solution, on the other hand, is more flexible and sidesteps all the problems pointed out in the previous approach. The main concern with an avatar is the work required to create animations that can be combined to produce clear, natural-looking signing.

Several methods exist to virtually recreate of sign language gestures with 3D avatars. They can be clustered into three types: hand-crafted, motion capture and synthesis from a sign notation system (Gerlach et al., 2016).

Motion capture is done by recording signs made by a human using video cameras or other types of sensors and later mapping the human actor's motions onto an avatar. The more detailed the desired result, the costlier and more complex the technology and expertise required. Motion capture frequently requires calibration and must usually be used alongside hand-crafted animation, because the resulting performance often has to be fine-tuned, especially when using cheaper solutions, such as Kinect and Leap Motion (Gerlach et al., 2016). Caroline Guardino and Ching-Hua Chuan attained better results using Leap motion than Kinect and Cyberglove, used in other studies for recognising sign language (Guardino et al., 2014).

The second type of animation consists in creating a system capable of automatically interpreting a phonetic sign language writing system, like HamNoSys (Hamburg Sign Language Notation System) to animate a signing avatar (Zwitserlood et al., ; Elliott et al., 2004). This writing system gives us detailed information about the hands elements

and other human movements that compose a sign (Hanke, 2004), but not secondary movement (unlike motion capture).

Lastly, hand-crafted animation is the oldest of these techniques, widely used, and known to give good results, but also requires intensive work, as someone must manually pose the avatar and adjust the animation until the result is satisfactory. The more realistic and detailed the animations, the more time, effort, expertise and technological sophistication are necessary. Blender[1] and Unity[2] are widely used general-purpose 3D computer graphics software capable of animating avatars. Blender in particular has a vast feature set, but also a steep learning curve (Cano, 2011). In contrast, our Animator is designed specifically for animating humanoid characters, which allows us to restrain its complexity.

3. PE2LGP Description

3.1. Overview

PE2LGP was originally created by Inês Almeida (2014) and further built upon by Ruben Santos (2016) in their master's theses. The project is part of *Corpus & Avatar da Língua Gestual Portuguesa*, a joint effort of researchers at Instituto Superior Técnico and Universidade Católica Portuguesa to create not only an avatar capable of signing LGP, but also the first LGP *corpus*, complete with video, translation, gloss and syntactic annotation. This interdisciplinary approach of linguistics and computer science allows greater cooperation between two otherwise separate projects, with the corpus being used for applications such as machine translation and animation synthesis through HamNoSys.

PE2LGP currently has 5 components, all of which feature our 3D avatar, *Catarina*, as a centrepiece. These are the Translator, the Animator, the Hand Pose Editor, the Kinect Recorder and the Animation Viewer. The Translator component receives a sentence in Portuguese as input, which will then be automatically translated to LGP and signed by Catarina. The Animator, being the focus of this paper, is described in detail in Section 3.2., and it allows the user to create new animations using *forward kinematics* (manipulating each of the avatar's joints individually). The Hand Pose Editor (created in the time between the execution of this study and its final revision) allows users to create and modify hand poses, which are then used in the Animator. The Kinect Recorder's purpose is to create new animations through rudimentary motion capture using a Kinect device. The Animation Viewer is a simple menu which allows the user to view and delete existing animations.

The Animator plays a crucial role in the project as the principal tool for creating signs for our avatar, with the ultimate goal of creating an animation database to be used by the Translator component or any other future component that requires it (such as dictionaries, messaging systems, chat bots or games).

3.2. Animator

Our Animator component makes use of key frames, pre-made hand poses, and forward kinematics. You can see a

screenshot of the animator's interface in Figure 2.

Key frames are the principal poses the avatar will assume throughout the time span of an animation (at a pace of one key frame per second). The user must define a key pose for each key frame one by one and, when the animation is played back, the avatar will not only assume the correct key pose for each key frame, but also automatically interpolate between key poses to generate all the in-between frames (Figure 1). This key frame method allows the user to focus on the essential moments of the sign and only pay attention to the intermediate moments when the situation requires it.

Forward kinematics is used by the system to allow the user to manipulate the avatar into the desired key poses by rotating the avatar's joints so, for example, rotating an elbow will move the forearm but leave the upper arm in place. The joints available for the user to manipulate are the neck, waist, shoulders, elbows, and wrists, which can all be rotated in 3 axes.

By design, the editor does not permit manipulating the joints on the avatar's hands and fingers directly (only its wrists). Instead, the user must choose from a selection of pre-made hand poses, which may be changed each frame and chosen independently for the right and left hand (Figure 3). When this experiment was conducted, these hand poses were limited to a selection of older animations created at an earlier stage of the project but, at the time of writing, a new component called the Hand Pose Editor has been fully implemented, enabling users to create and modify hand poses at will.

The project does not yet support facial expressions, though this feature is a current priority.

4. Evaluation Methodology

Seven simple LGP sentences (Table 1) were created and animated by two engineers with basic LGP training working on the project, using the corpus (Section 3.1.) and two online dictionaries as reference[3,4]. Most signs for these sentences were newly-created using the Animator, but some were already present on the platform's database, such as the letter signs used for finger-spelling[5]. Each new sign took between 10 and 60 minutes to animate, depending on its complexity and the desired quality and detail.

To string together signs to form sentences, Unity's native animation features were used, specifically its *animation controller* system, which employs state machines to allow different animations to be played back and mixed. Without this mechanism, the transitions between signs would have been less smooth, because the avatar would have been forced to return to a neutral position (standing upright, with arms at its sides) between every sign. Using this system, however, is not effortless, as it requires calibration dependent on which two animations the transition involves. Note the distinction between: interpolation between key frames

---

[1] www.blender.org
[2] unity.com

[3] www.spreadthesign.com/pt.pt
[4] www.infopedia.pt/dicionarios/ lingua-gestual

[5] Finger-spelling consists of signing individual letters of the alphabet to spell out words, usually proper names. In Table 1 you can see that the proper name "Júlio" was finger-spelled as J-U-L-I-O.

Figure 1: In the Animator, the user defines only key frames A and B, and the system interpolates between those two key frames to create the in-between frames, as the image illustrates. Note that this is a simplified view: in reality, there would normally be 58 in-between frames, not 4.

Figure 2: Screenshot of the Animator. Notice the red hoop around the avatar's waist, indicating the selected joint and how it will rotate. On the right, you can see the X, Y, and Z buttons, which control which axis is to be rotated. On the left, there are various controls, such as for creating new frames, switching hand poses, and previewing the animation.

Figure 3: Examples of pre-made hand poses, which can be selected in the editor.

(which occurs *within an animation*) and transitions *between two animations*. The former occurs in the Animator, when the sign is created, while the latter is done in Unity's animation controller.

The grammatical correctness of the 7 sentences was then verified by a hearing linguist with intermediate mastery of LGP and a trained LGP interpreter (both native Portuguese speakers). This validation yielded several recommendations, which were then implemented within the boundaries of the Animator before being committed to the next phase. Finally, videos (which are available at this footnote URL[6]) of the avatar signing the sentences were sent to 4 evaluators, who answered an online form about the avatar's quality without being told the sentences' meaning or which signs were being produced. The evaluators described their proficiency as "high", "moderately proficient", "med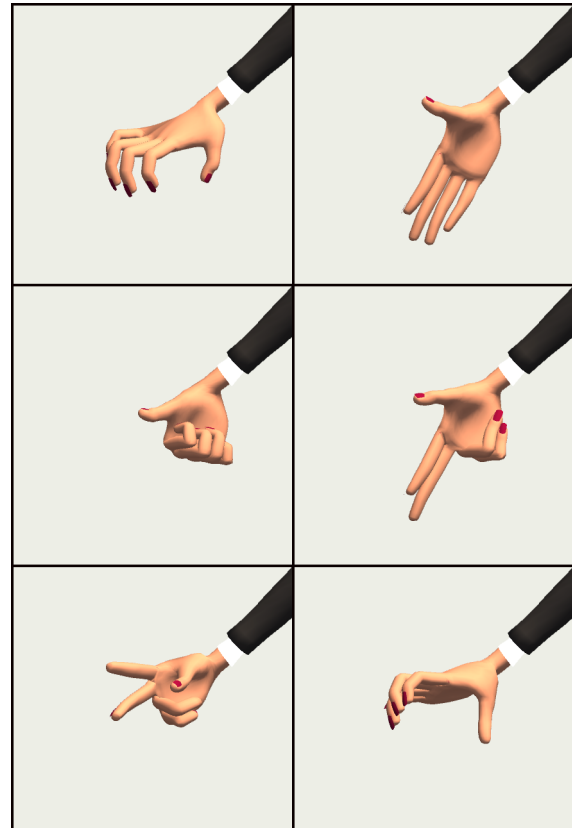ium" and "professional". The form consisted of a set of questions which were repeated for each sentence: the evaluator would view the video of the avatar signing and try to determine the meaning of the sentence. Whether or not the evaluators were able to discern a sentence's meaning was the main factor in measuring how intelligible it was. Next, the evaluator would answer how many times they had to watch the video and rate several aspects of the sentence's quality: speed, overall quality, intelligibility, naturalness, grammar, hand configuration, hand orientation, hand location, and hand movement. These aspects were rated on a scale of 1 to 5 with 1 being "Bad" and 5 being "Good" (speed was the exception, where 1 meant "Too slow" and 5 meant "Too fast".

5. Results

The full quantitative responses to our survey are available in Table 3 and Table 2.

---

[6]tinyurl.com/PlaylistAvatarLGP

Translation	LGP (gloss)
Bom dia (Good day)	BOM DIA GOOD DAY
A rainha é soberana (The queen is sovereign)	MULHER REI IMPORTANTE MUITO CHEFE WOMAN KING IMPORTANT VERY BOSS
O rapaz chegou à cozinha (The boy arrived at the kitchen)	RAPAZ COZINHA CHEGAR BOY KITCHEN ARRIVE
O meu nome é Júlio (My name is Julio)	NOME MEU J-U-L-I-O NAME MINE J-U-L-I-O
O rei pensou e teve uma ideia (The king thought and had an idea)	REI PENSAR TER-IDEIA KING THINK HAVE-IDEA
O meu irmão come peixe (My brother eats fish)	IRMÃO MEU PEIXE COMER BROTHER MINE FISH KING
Ele dança bem (He dances well)	ELE DANÇAR BEM HE DANCE WELL

Table 1: Sentences used in the evaluation.

Naturalness was the lowest-rated aspect, with an average of 1.7. This was to be expected, as the avatar's movements are too machine-like to appear human. The main requirements to improve naturalness would be facial expressions (both grammatical and non-grammatical), configurable key frame interpolation, automatic secondary movement, lower-body movement.

The highest-rated aspect was *Speed*, with 18 out of 28 perfect classifications. Part of the reason for this may be that speed problems are localised to particular signs or transitions. Another factor may be the initial waiting time within the videos, before the avatar begins signing. This waiting time was unintentional and not consistent across the videos (ranging from 4 seconds to less than a second), and we suspect it may create the illusion of a slower animation.

We consider our most successful sentence to be ELE DANÇAR BEM, which was almost perfectly understood by all evaluators with a single viewing of the video and consistently outperformed other sentences across all categories. This sentence is interesting, as it is the only one in this experiment to include non-manual movement (waist and neck motion), and we believe that is why it scored higher in Naturalness than any other sentence.

We consider the least successful sentence to be the second, MULHER REI IMPORTANTE MUITO CHEFE, which the evaluators viewed more times than usual and had trouble understanding (one could not name any signs correctly). One suspected cause for this difficulty is the absence of context.

In the categories of hand configuration, hand orientation, hand location and hand movement, the responses often corresponded to our expectations, where sentences with higher quality signs received better scores. In a few cases, the responses (both quantitative and open-ended) led us to discover more subtle improvements that could be made to the signs (such as bringing the hand closer to the chin in the sign BOM).

In some sentences, we detected that unruly transitions between animations made signs less clear, or at least less natural, by making the avatar's motion too quick or anatomically impossible. A number of problems with individual signs were also detected, most often imprecise hand poses.

6. Conclusion

In this paper, we presented our Animator tool which, as a part the PE2LGP project, aims to be an accessible means to animate signs to be performed by an avatar. We also performed a preliminary evaluation of this system using Portuguese Sign Language sentences which, although too small to yield statistically significant results, provides valuable insight into the current capabilities, limitations and future potential, while demonstrating that the platform is indeed capable of producing comprehensible LGP, which is a positive result, given its simplicity and the complexity of synthesising natural languages.

In the future, we would like to improve the Animator with both quality-of-life features, to make editing animations more comfortable and efficient, and, more importantly, improvements to enhance the tool's capacity, such as custom hand posing, facial expressions, and control over interpolation. For further research, it would be interesting to formally study the Animator's ease of use through user tests, as accessibility is one of its main design goals. Ultimately, this usability should enable the development of a thorough database of animations by using the Animator and its companion components to swiftly bring the first-hand knowledge of LGP users into the platform, to be used as a resource in this and other Portuguese Sign Language projects.

7. Acknowledgements

8. Bibliographical References

Almeida, I. R. (2014). Exploring challenges in avatar-based translation from european portuguese to portuguese sign language. Master's thesis, Instituto Superior Técnico, October.

Category	Evaluator	S1	S2	S3	S4	S5	S6	S7
Number of Views	A	2	>1	3-4	3-4	4	4-5	1
	B	1	7	2	3	2	1	1
	C	1	>1	>1	>1	>1	>1	1
	D	2	5-6	5	1	3	1	1
Speed	A	2	1	1	2	1	2	3
	B	3	1	3	2	2	3	4
	C	4	NA	2	3	2	3	4
	D	3	NA	1	2	2	3	5

Table 2: These were the responses of each evaluator to the questions "How many times did you need to watch the video?" and "How adequate was the sentences's speed?". Unlike the results presented in Table 3, the first question was open-ended, while the second was on a scale of 1 to 5, with 1 labeled "Too slow" and 5 labeled "Too fast".

Category	Evaluator	S1	S2	S3	S4	S5	S6	S7
Overall Quality	A	2	1	1	2	1	2	3
	B	3	1	3	2	2	3	4
	C	4	NA	2	3	2	3	4
	D	3	NA	1	2	2	3	5
Inteligibility	A	2	1	2	1	1	2	3
	B	4	1	3	4	3	4	5
	C	4	2	2	2	2	3	4
	D	4	NA	1	3	2	3	5
Naturalness	A	1	1	1	1	1	1	2
	B	1	1	1	2	2	1	2
	C	2	2	2	2	1	2	2
	D	2	NA	1	2	3	3	3
Grammar	A	3	1	2	3	NA	2	3
	B	5	NA	4	5	4	4	4
	C	4	NA	3	4	NA	4	4
	D	5	NA	5	5	3	5	5
Hand Configuration	A	3	2	1	1	1	1	3
	B	5	2	2	1	3	1	5
	C	4	NA	2	1	2	2	NA
	D	5	NA	1	4	3	4	5
Hand Orientation	A	3	1	1	2	1	2	3
	B	4	2	5	3	2	3	5
	C	4	NA	NA	NA	2	2	NA
	D	5	NA	2	5	4	5	5
Hand Location	A	1	1	2	2	1	1	3
	B	4	4	5	4	2	3	5
	C	4	NA	NA	1	3	2	3
	D	4	NA	2	5	4	4	3
Hand Movement	A	1	1	1	1	1	1	3
	B	3	4	1	2	2	1	5
	C	4	NA	2	2	2	1	3
	D	4	NA	2	4	3	4	5

Table 3: These were the responses of each evaluator for each category and sentence (*S1*, *S2*, etc), on a scale of 1 to 5. In the survey, the response *NA* was written "I don't know", while 1 was labeled "Bad" and 5 was labeled "Good".

Cano, J. L. C. (2011). The cambrian explosion of popular 3d printing. *IJIMAI*, 1(4):30–32.

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

Gerlach, J., Strasly, I., Ebling, S., Rayner, M., Bouillon, P., and Tsourakis, N. (2016). An Open Web Platform for Rule-Based Speech-to-Sign Translation. (August):162–168.

Guardino, C., Chuan, C.-H., and Regina, E. (2014). American sign language recognition using leap motion sensor. 12.

Hanke, T. (2004). HamNoSys-Representing sign language data in language resources and language processing contexts. *Proceedings of the Workshop on Representation and Processing of Sign Language, Workshop to the forth*

International Conference on Language Resources and Evaluation (LREC'04), pages 1–6.

Huenerfauth, M. and Hanson, V. L. (2009). Sign language in the interface: Access for deaf signers. *The Universal Access Handbook*, pages 38–1–38–18.

Kaur, R. and Kumar, P. (2014). Hamnosys generation system for sign language. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2727–2734, Sep.

Santos, R. E. R. (2016). Pe2lgp: Do texto à língua gestual (e vice-versa). Master's thesis, Instituto Superior Técnico, October.

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