

PE2LGP

PE2LGP: From Text to Sign Language (and vice versa)

Ruben Emanuel Ramires dos Santos

Thesis to obtain the Master of Science Degree in Engineering and
Computer Science

Masters Dissertation in Engineering and Computer Science

Supervisor: Prof^a. Maria Luísa Torres Ribeiro Marques da Silva Coheur
Prof. João António Madeiras Pereira

Examination Committee

Chairperson: Prof. Miguel Nuno Dias Alves Pupo Correia
Supervisor: Prof^a. Maria Luísa Torres Ribeiro Marques da Silva Coheur
Member of the Committee: Prof. Alfredo Manuel dos Santos Ferreira Júnior

November 2016

Chapter 1: Introduction

Portuguese Sign Language (PSL or LGP in Portuguese) is the sign language used by the deaf community in Portugal and it is executed through hand movements, considering body and facial expressions¹. LGP and all sign languages are developed spontaneously and independently within deaf communities around the world (McNeill and Duncan, 2005).

PE2LGP² system (From European Portuguese to Portuguese Sign Language) (Almeida, 2014) started to be developed in 2014 at L2F/INESC-ID by Inês Almeida within her master thesis in Engineering and Computer Engineering.

PE2LGP (made with Blender), translates, in real time, European Portuguese (in textual form) into LGP, being the signs produced by a 3D virtual character. The system contains an interface that lets you interact with the character.

In this thesis we improved and extended PE2LGP. Changes were made in terms of its architecture and two systems were developed in order to allow the introduction of new signs, without the need of having advanced knowledge in the animation area

1.1 Problem and Motivation

Inês Almeida's prototype contains a small quantity of signs. This means that it produces few sentences in LGP. Furthermore, to add information, while involves creating 3D animations, the first version of PE2LGP requires advanced knowledge in the 3D animation field. The goal of this thesis is to eliminate or reduce these limitations.

1.2 Goals

The main focus of this thesis was on creating two systems that reduce the limitations of introducing new signs to the character. On the one hand, we have a manual system, which uses only an interface to directly interact with the model and create signs. On the other hand, we have an automatic system, which explores the capabilities of the Kinect sensor to automate the whole process. Also, the previous prototype architecture was changed. In general, the main goals are:

1. Modification of the Previous Architecture:
 - (a) Exportation of the Prototype to Unity.
 - (b) Enrichment of the Language Component.
2. Creation of the Manual Animations System.
3. Creation of the Automatic Animations System.
4. Unification of the Systems.
5. Evaluation of the Systems.

¹<http://www.disabled-world.com/disability/types/hearing/communication/>

²For more information about this work, please see the full version of this work.

Chapter 2: Related Work

This research began by looking at sign language avatars that were similar to the one used in PE2LGP; then we study some works that make use of the sensors Kinect and Leap Motion.

2.1 Sign Language Avatars

The avatars, in the the digital context, are virtual characters that aim to perform tasks replacing humans. With regard to the sign language, avatars are useful because people suffering from hearing disabilities have significant problems in reading, and therefore Web pages content is not fully accessible (Kipp et al., 2011).

2.1.1 Avatars in Portuguese Sign Language

In Portugal, we highlight the work created by José Bento in 2014, the “Avatars on Portuguese Sign Language” (Bento et al., 2014), which aimed to create a virtual interpreter to express Portuguese Sign Language. Through a Web platform, the virtual interpreter allows us to know several signs used on a daily basis.

2.1.2 Avatars in Brazilian Sign Language

There are 3 main Brazilian entities that had developed avatars within the Sign Language: HandTalk, ProDeaf and Brava.

HandTalk¹ is a startup that consists of a Portuguese translation platform for Brazilian Sign Language (also known as LIBRAS) using a 3D character named Hugo.

Prodeaf², such as the HandTalk, is a startup focused on using avatars as means of communication. However, ProDeaf differs from HandTalk because it uses a crowdsourcing platform.

Brava also differs from others, as the way of creating signs is done via Motion Capture. To complement this approach and helping the capture, it uses a suit and gloves with sensors.

2.1.3 Avatars in other Sign Languages

The work titled “A Prototype Malayalam to Sign Language Automatic Translator” (Joy and Balakrishnan, 2014) presents a system that receives Malaysian text as input and generates the corresponding sign language using an avatar to reproduce the signs.

“Sign 360” consists in an application to learn FSL (French Sign Language). This project also differentiates from others by using the Motion Capture method, which allows to create high quality avatars/signs³.

There are still other works in other languages that follow the same approach. For instance, the project ATLAS (Barberis et al., 2011) that consists in a system to translate Italian to Italian Sign Language, the work of Jemni et al. (2013) that translates Arabic to Arabic Sign Language or the work of Kipp et al. (2011) which is an animation editor that represents signs in Thai Sign Language.

¹<http://www.handtalk.me/>

²<http://www.prodeaf.net/>

³<https://www.youtube.com/watch?v=oRM4nZnokow>

2.2 Creation and Manipulation of Objects using the Kinect

Motion Capture is the process of recording an event of a movement in real time and translate it into mathematical terms in order to obtain a 3D representation. This approach is used in several areas such as: Medicine ([Liu et al., 2014](#); [Das et al., 2011](#); [Gong et al., 2015](#)), Robotics ([Choi et al., 2013](#); [Özgür et al., 2014](#)), Transports ([Palacio and Arévalo, 2014](#)) or Education ([Leite and Orvalho, 2013](#); [Bruciati and Rossignol, 2015](#); [Adell and Castañeda Quintero, 2012](#)), with emphasis on Augmented Reality Games ([Rincon et al., 2016](#)), among others.

With regards to manipulating avatars using Motion Capture, the work of [Bholsithi et al. \(2014\)](#) aims to control avatars from the movement captured by Kinect sensor. To do that, they use Artificial Intelligence libraries such as Open NI with NITE Primesense middleware⁴. In medicine, there is a project ([Gong et al., 2015](#)), that aims to create an avatar that looks like a physical therapist to guide the process of rehabilitation of the patient, since it is not required the physical presence.

2.3 Hand Tracking

Hand Tracking, in general, is a complex and abstract aspect of artificial intelligence and consists in detecting the trajectory of the hand in a sequence of images ([Sharp et al., 2015](#)). For this, it is necessary to know, for each frame, which pixels belongs to the hand. This process is called hand segmentation. In most projects the hand segmentation technique is always the same, which consists in taking advantage of the depth sensor information, assuming that the hand is the closest thing to the sensor and, therefore, everything is ignored.

Hand Tracking can be done by using discriminatory methods ([Keskin et al., 2012](#)), which work directly with the image data (for example to extract features and to use classification techniques), generative methods (or model-based) ([Oikonomidis et al., 2011](#)) which use a 3D model of a hand to recover the pose and, finally, hybrid methods that are a combination of the previous ones ([Erol et al., 2007](#)).

2.4 Resources in Natural Language Processing

NLTK⁵, used in PE2LGP, is a set of libraries and programs for the Python programming language and aims to make the symbolic and statistical natural language processing. NLTK also has graphical demonstrations and samples of data.

Language resources in the LGP are limited. The only existing grammar belongs to [Amaral et al. \(1994\)](#) and is not accessible. A recent study in this context is the one of [da Silva Bettencourt \(2015\)](#), in her work "The order of the words in Portuguese Sign Language: a brief comparative study with Portuguese and other sign languages", which aimed to realize if LGP had a pattern of basic order of the sentence constituents and, if confirmed, what would be. There is also the Portuguese Sign Language dictionary of Ana Bela Baltazar⁶, which allow us to understand how to perform each sign.

⁴<http://structure.io/openni>

⁵<http://www.nltk.org>

⁶<https://www.portoeditora.pt/produtos/ficha/dicionario-de-lingua-gestual-portuguesa/3501376>

Chapter 3: System 1 - Architecture Modification

Initially, PE2LGP was integrated into Blender, including the natural language component (Figure 3.1). Two changes have been made in order to overcome this dependency: export PE2LGP to the game engine Unity, and create an independent natural language processing model (Figure 3.2). Furthermore, Kinect sensor was introduced, used by the automatic system to create signs (chapter 5).

The natural language component can be used as a single stand-alone system, or be integrated in another system, as it is done in this thesis. The choice of introducing a game engine arises from the need to scale the project. Using a game engine lets you export the project to another format.

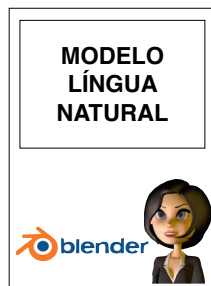


Figure 3.1: PE2LGP Original Architecture

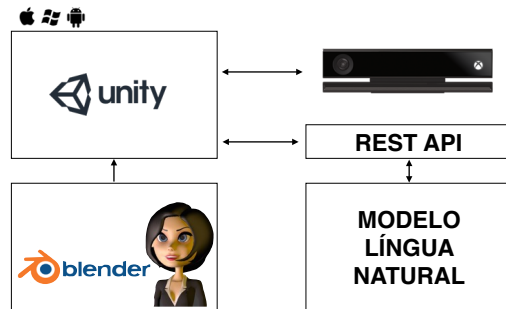


Figure 3.2: PE2LGP Current Architecture.

The interface (Figure 3.3) is used to interact with all the system components. In the interface you can write a word or phrase and get its meaning in LGP, control the speed of the signs, see the sign and its transcription simultaneously, change the size of the transcription text, stop anything that is playing at any time, repeat the current input, move the model around its own axis X and zoom-in or zoom-out.



Figure 3.3: PE2LGP System 1 Interface.

Previously in PE2LGP, the development of the natural language processing component was made through Blender using Python. Therefore, if we need to change that component it would be required to work in Blender. In order to eliminate this dependency, a REST API written in Python was created which handles HTTP requests. Such requests take as argument a string that is analyzed by the natural language processing component, which also had modifications. Each response to an HTTP request results in a JSON file with the string processed.

Chapter 4: System 2 - Manual Creation of Signs

This system allows to create signs, by interacting directly with the model through an interface, thus creating movements, frame by frame, giving rise to animations.

The first step was to create the animation file. It is possible to choose the name of the animation and then insert it in the interface. Then, the chosen name will be validated using two methods.

The next step gives the user the chance to choose both hands configurations in order to create the animation. This way, once again, the user has an interface (Figure 4.1) where it is possible to choose 54 different hands configuration (already implemented in the first version of PE2LGP).



Figure 4.1: Hands Selection

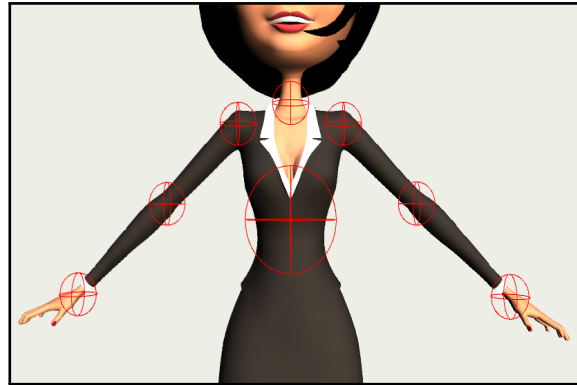


Figure 4.2: Movement Points.

Eight points were defined where the user can move the model (Figure 4.2). The points are: the hands, the elbows, the shoulders, the neck and the belly.

As the feedback given to the user at the selected point was not enough, a 3D object was created, in order to give more real-time feedback to the user (Figure 4.3). With this change, the user knows where the point is rotating and which one is selected. To limit movements a rotation angle has been set for each point, which tries to simulate the human body movement.



Figure 4.3: Rotation Feedback.

To create frames the user must move the model freely and then press a button on the interface for each time that wants to save that frame.

The interfaces makes the process of creation easier and contains the following options: place the model in the initial position eliminating all frames created at the moment, create a frame in the animaiton file, complete the whole process of creation and then return to the main menu, changing at any moment the hands configuration, play animation, stop the animation.

There are three panels in the interface. The first contains general information about the system, the second shows information about the current screen and, finally, the third is to switch the axis.

Chapter 5: System 3 - Automatic Creation of Signs

This system, unlike the previous one, is automatic. This means that it does not require user interaction with the model via an interface, as instead, it uses the Kinect sensor to automatize the process.

After going through the sensor settings and then exploring the Microsoft plugin for Unity, the main focus was to map the points captured by Kinect in any avatar.

The need of introducing voice commands arises due to the fact that the movement of going to the computer to finish the animation is part of the animation itself, which is not what is intended. For this reason, a voice command was implemented in order to stop recording the animation.

The capture was limited to eight points, thus the user has more freedom. This means that the Kinect only needs to capture the upper body to make the system work. In addition, two different cameras were integrated. One of them is the result of the RGB camera sensor, and the other is the same camera but with two features: the captured points are connected through lines and the pixels that are not close to the user are deleted, giving a Chroma Key effect (Figure 5.1).

Initially, the camera was capturing the whole model and was independent, which in most of the cases was keeping the model way from the screen, not making full use of the screen. For this reason, the camera was attached to the model, so the user would see the model using the entire space of the screen. There is a decremental counter with a value of 5 seconds defined empirically with the aim of preventing the recording from starting immediately.

In order to avoid some cases where the capture was wrong at the beginning, (for example when the user is not framed with the camera or when several users are using the system instead of only one), TPose was created which requires the user to open and stretch the arms in order to initiate (Figure 5.1).

Hands capture was not made through Kinect sensor, therefore, the system has an interface, similar to the system 1, that allows to choose the hands configuration and also the animation name.

The interface aims to help the user creating the animation and has the following options in regards to the recording: start, finish, play and stop.



Figure 5.1: Tracking and TPose Example.

Chapter 6: Unification of the Systems

All 3 systems described above were put together into a single application. In addition, it has added a system that allows to view all PE2LGP available animations.

6.1 Main Menu

The main menu has 4 panels (Figure 6.1). It is the first contact with the user and aims to present all the available systems. Beside the panels, each one is represented by an illustrative picture and a brief description. The user can choose any system and return to the main menu at any time.



Figure 6.1: Panels of the Main Menu.

6.2 Animation Viewer

Animation Viewer matches the last panel of the main menu and aims to give the chance to view all the available animations. For each sign it is possible to: play it repeatedly, stop and delete. In addition to these three actions, the user may also use the mouse to interact with the model, namely to move around its own axis X and make zoom-in or zoom-out.

6.3 Usability and Design

The unification of the system was focused in two main points: usability and design. Thus, the concern was focused in making the system looks like one instead of three separate.

A logo to identify the application was created, and it is present in any screen. The phrase corresponds to the acronym of the phrase that represents the application in Portuguese, that is PE2LGP (Português Europeu para Língua Gestual Portuguesa).

The colors were chosen empirically, thus resulting in 4 major: purple, white, black, and beige. The goal was that light colors together with the dark, could not only make contrast, but that does not disturb the model and its animations. In the case of using dark colors it was found that it was sometimes difficult to distinguish the various model elements

In addition to the logo and colors, it was necessary to maintain consistency in other aspects, such as the size and font type, error messages, icons and their respective positions.

Chapter 7: Evaluation

All 3 systems previously described were evaluated, with the main focus on the systems that allow to create signs. In the first evaluation participated a group of 6 people, aged between 18 and 25, both genders, and from different areas. In addition, the systems 2 and 3 had a second evaluation, this time with 6 people involved in LGP, with the aim of evaluating the quality of the signs produced by the systems.

7.1 Evaluation Methods

Initially, PE2LGP and its evolution was briefly presented to the first group. Then, as there were 3 systems, the evaluation was also divided in 3 parts. Each part was followed by sheets containing: a system explanation, the tasks to be performed and, finally, a questionnaire. In the first system the user should test the systems usability, being given some basic tasks, such as checking the meaning of a word and phrase in LGP.

The tasks of the other two systems consisted in the creation of 1 sign, that is, each person would have to create a sign twice, one in each system.

Afterwards, the same signs were carefully created by me in order to be evaluated by the second group with regards to the its quality in the LGP context. The signs, 2 for each word, were sent in questionnaire to people involved in LGP. This group did not know with which system the sign was created. The sign was presented and the user would have to guess it and then classify according with its quality through a scale from 1 to 5.

7.2 Comparison of the Results

There are two important factors that distinguished the systems. On the one hand, the system 3, has a lower creation time than the system 1, since it takes less than a third of the system 2 to create a sign. Additionally, the difficulty in creating a sign may depend on the complexity. However, the system 2 has proved to be the best. In addition, system 3 does not allow to change the hands configuration during the recording, which is a limitation.

As for quality of the signs, the second evaluation of the systems 2 and 3 revealed that the signs produced are inadequate in the LGP context. Still, the system 2 performed better.

7.3 Conclusion

From the results obtained from both focus groups, we concluded that both systems have potential but they should be improved in order to be used by the LGP community. After applying the improvements, it is assumed that these systems can be used by anyone, including its target audience.

Chapter 8: Conclusion

The main purpose of this work was focused in making the system more flexible and streamline the way of creating new signs.

In order to make the system more flexible, a game engine, Unity, was introduced. With regards to the creating of new signs, two systems were created: a manual one, which allows to interact directly with the avatar through an interface, in order to create animations; an automatic one, that tries to automatize the whole process by using the Kinect sensor.

The manual system has the advantage of having multiple levels of freedom in the creation of signs. It is possible to make almost every movement and rotation. In the automatic system, the movements are dependent from the capture of the Kinect sensor. On the other hand, in the automatic system the biggest advantage is the creation time, which is very small compared with the manual system. In addition, system 3 is limited, it doesn't capture the hands in real-time, the user must choose the hands configuration before create the sign.

The evaluation has shown that all systems have quality in terms of usability. With regards to the systems 2 and 3, the signs produced do not have quality to be used in the LGP context.

The greatest difficulties were mainly in overcoming problems encountered during the creation of the two systems, with an emphasis on the creation of the animation itself.

All the objectives defined initially have been achieved. However, future work is required.

Bibliography

- [Adell, J. and Castañeda Quintero, L. J. (2012)]. Tecnologías emergentes, pedagogías emergentes
- [Almeida, I. (2014)]. Exploring Challenges in Avatar-based Translation from European Portuguese to Portuguese Sign Language.
- [Amaral, M. A., Coutinho, A., Martins, M. R. D., and Johnson, R. (1994)]. *Para uma gramática da língua gestual portuguesa*.
- [Barberis, D., Garazzino, N., Prinetto, P., and Tiotto, G. (2011)]. Improving accessibility for deaf people. *Assets (Xiii)*, page 253.
- [Bento, J., Claudio, A. P., and Urbano, P. (2014)]. Avatars on Portuguese sign language. *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)*.
- [Bholsithi, W., Wongwaen, N., and Sinthanayothin, C. (2014)]. 3d avatar developments in real time and accuracy assessments. In *2014 International Computer Science and Engineering Conference*. Institute of Electrical & Electronics Engineers (IEEE).
- [Bruciati, A. and Rossignol, C. (2015)]. Using mocap & chromakey special effects for increasing instructor presence in an online course. *Learning*, 11:45am.
- [Choi, S.-W., Kim, W.-J., and Lee, C. H. (2013)]. Interactive display robot: projector robot with natural user interface. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 109–110. IEEE Press.
- [da Silva Bettencourt, M. F. (2015)]. A ordem das palavras na língua gestual portuguesa: Breve estudo comparativo com o português e outras línguas gestuais. Faculdade de Letras da Universidade do Porto.
- [Das, S., Trutoiu, L., Murai, A., Alcindor, D., Oh, M., De la Torre, F., and Hodgins, J. (2011)]. Quantitative measurement of motor symptoms in parkinson's disease: A study with full-body motion capture data. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6789–6792. IEEE.
- [Erol, A., Bebis, G., Nicolescu, M., Boyle, R. D., and Twombly, X. (2007)]. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, 108(1-2):52–73.
- [Gong, W., Wang, Z., and Zhang, W. (2015)]. Creating personal 3d avatar from a single depth sensor. In *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, pages 1193–1198. IEEE.
- [Jemni, M., Semreen, S., Othman, A., Tmar, Z., and Aouiti, N. (2013)]. Toward the creation of an Arab Gloss for arabic Sign Language annotation. *Fourth International Conference on Information and Communication Technology and Accessibility (ICTA)*, pages 1–5.

- [Joy, J. and Balakrishnan, K. (2014)]. A prototype Malayalam to Sign Language Automatic Translator.
- [Keskin, C., Kiraç, F., Kara, Y. E., and Akarun, L. (2012)]. Hand pose estimation and hand shape classification using multi-layered randomized decision forests. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7577 LNCS(PART 6):852–863.
- [Kipp, M., Heloir, A., and Nguyen, Q. (2011)]. Sign language avatars: Animation and comprehensibility. In *Intelligent Virtual Agents*, pages 113–126. Springer Science Business Media.
- [Leite, L. and Orvalho, V. (2013)]. Inter-acting: Understanding interaction with performance-driven puppets using low-cost optical motion capture device. *Journal International Journal of Advanced Computer Science*, 3(2):65–69.
- [Liu, J., Chuan, H., and Kuan, P. (2014)]. Assessment of range of shoulder motion using kinect. *Gerontechnology*, 13(2):249.
- [McNeill, D. and Duncan, S. D. (2005)]. Grammar, Gesture, and Meaning in American Sign Language (review). *Sign Language Studies*, 5(4):506–523.
- [Oikonomidis, I., Kyriazis, N., and Argyros, A. a. (2011)]. Efficient Model-based 3D Tracking of Hand Articulations using Kinect. *22nd British Machine Vision Conference*, pages 1–11.
- [Özgür, A., Bonardi, S., Vespignani, M., Möckel, R., and Ijspeert, A. J. (2014)]. Natural user interface for roombots. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 12–17. IEEE.
- [Palacio, J. A. P. and Arévalo, S. L. (2014)]. Remote control flying rc helicopter through natural user interface. In *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6. IEEE.
- [Rincon, A. L., Yamasaki, H., and Shimoda, S. (2016)]. Design of a video game for rehabilitation using motion capture, emg analysis and virtual reality. In *2016 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, pages 198–204. IEEE.
- [Sharp, T., Keskin, C., Robertson, D., Taylor, J., Shotton, J., Kim, D., Rhemann, C., Leichter, I., Vinnikov, A., Wei, Y., Freedman, D., Kohli, P., Krupka, E., Fitzgibbon, A., and Izadi, S. (2015)]. Accurate, robust, and flexible real-time hand tracking. CHI.