Real Time Bidirectional Translator of Portuguese Sign Language

Paula Escudeiro[1], Nuno Escudeiro[1], Rosa Reis[1], Pedro Rodrigues[1], Jorge Lopes[1], Marcelo Norberto[1], Ana Bela Baltasar[1], Maciel Barbosa[2] and José Bidarra[3]

[1]*Departamento de Engenharia informática - Instituto Superior de Engenharia do Porto, Porto, Portugal*
[2]*Departamento de Engenharia Electrotécnica - Faculdade de Engenharia do Porto,* Porto, Portugal
[3]*Departamento de Ciências Exactas e Tecnologia - Universidade Aberta, Lisboa, Portugal*

Keywords: Portuguese Sign Language, Education, Deaf, Kinect, Sensor Gloves, Translator, Machine Learning, OpenCv, Support Vector Machines, Dynamic Time Warping, Avatar, Serious Games.

Abstract: The communication with deaf by means of written text is not as efficient as it might seem. In fact, there is a very deep gap between sign language and spoken/written language. The deployment of tools to assist the daily communication between deaf people and the rest may be a significant contribution to the social inclusion of the deaf community. The work described in this paper addresses the development of a bidirectional translator between Portuguese Sign Language and Portuguese text and a serious game to promote the learning of the Portuguese Sign Language. The translator from sign language to text employs two devices, namely the Microsoft Kinect and 5DT Sensor Gloves in order to gather data about the motion and shape of the hands. The hands configurations are classified using Support Vector Machines. The classification of the movement and orientation of the hands is achieved through the use of Dynamic Time Warping algorithm. The translator exhibits a precision higher than 90%. In the other direction, the translation of Portuguese text to Portuguese Sign Language is supported by a 3D avatar which interprets the entered text and performs the corresponding animations. As a complement, we also present a serious game directed to assist in the difficult task of learning the Portuguese Sign Language.

1 INTRODUCTION

Promoting equity, equal opportunities to all and social inclusion of people with disabilities is a concern of modern societies at large and a key topic in the agenda of European Higher Education. The evolution of science and the emergence of new technologies combined with the commitment and dedication of many teachers, researchers and the deaf community is promoting the social inclusion and simplifying the communication between hearing impaired people and the rest.

Despite all the progress, we cannot ignore the fact that the conditions provided by the society for the deaf are still far from being perfect. For example, in the public services, it is not unusual for a deaf citizen to need assistance to communicate with an employee. Another critical area is education. Deaf children have significant difficulties in reading due to difficulties in understanding the meaning of the vocabulary and the sentences. This fact together with the lack of communication via sign language in schools severely compromises the development of linguistic, emotional and social skills in deaf students.

The VirtualSign project intends to reduce the linguistic barriers between the deaf community and those not suffering from hearing disabilities.

The project is oriented to the area of sign language and aims to improve the accessibility in terms of communication for people with disabilities in speech and/or hearing, and also encourage and support the learning of the Portuguese Sign Language.

The work described in this paper has three interconnected modules. These include a translator from gestures in Portuguese Sign Language (PSL) to text, that collects input data from a Microsoft Kinect device and a pair of data gloves, a translator from Portuguese written text to PSL, that uses a 3D avatar to reproduce the animations of the gestures corresponding to the written text, and a module consisting of a serious game designed to improve the learning of PSL. The first two modules are independent from the game. The game is an application that uses the bi-directional translator between PSL and written Portuguese with a specific

Escudeiro P., Escudeiro N., Reis R., Rodrigues P., Lopes J., Norberto M., Baltasar A., Barbosa M. and Bidarra J..
Real Time Bidirectional Translator of Portuguese Sign Language.
DOI: 10.5220/0005448502030213
In *Proceedings of the 11th International Conference on Web Information Systems and Technologies* (WEBIST-2015), pages 203-213
ISBN: 978-989-758-106-9

aim like many other applications that may be developed. Several technologies have been integrated including the Blender modelling software, the Unity 3D and Ogre game engines and the integrated development environment Microsoft Visual Studio together with a set of multi-paradigm programming languages, namely C# and C++.

In the remaining of this paper we briefly describe the Portuguese Sign Language, in Section 2, followed by a revision of related work, in Section 3. Section 4 gives an overview of our proposal, while Section 5 provides the technical details of the translation from sign language to text, Section 6 provides details of the translation from text to sign language and Section 7 describes the serious game proposed to assist learning of the Portuguese Sign Language. Section 8 presents our conclusions.

2 PORTUGUESE SIGN LANGUAGE

The Portuguese Sign Language (PSL) had its origins back in 1823 at the Casa Pia of Lisbon, from the initiative of Pär Aron Borg, a Swedish educator and pioneer in the education of deaf people. Currently, although there are no similarities in terms of vocabulary between the Portuguese and Swedish sign languages, the alphabets of both languages continue to show their common origin. The interest in PSL has shown remarkable growth over time, not only by the deaf community - which now accounts for nearly 100000 persons in Portugal (Associação de Surdos do Porto, 2015) - but also for the whole community involved, such as relatives, educators, teachers, and many more.

The Sign Language, like any other living language, is constantly evolving, increasingly being seen as a language of instruction and learning in different areas, a playful language in times of leisure, and professional language in several areas of work (Morgado and Martins, na).

2.1 Linguistic Aspects

The Portuguese Sign Language involves a set of components that make it a rich and hard to decode communication channel. When performing PSL, we must take account of a series of parameters that define the **manual** and **non-manual** components. By changing one of these parameters, usually the gesture changes or loses its meaning. At the level of manual component, we apply the definition of **dominant** and **non-dominant** (or supporting) hand. Usually for each person, the dominant hand coincides with the hand with greater dexterity. In the execution of gestures, the acting of the dominant hand may possibly differ from the support hand.

The manual component includes:

- **configuration of the hand**. By configuration of the hands we mean the form that each hand assumes while executing the gesture. There is a total of 57 hand configurations, shared between the dominant and supporting hand, which form the basis of the PSL.
- **orientation of the palm of the hand**. Some pairs of configurations differ only in the palm's orientation.
- **location of articulation**. The place of articulation comprises the different areas of the body where the gestures are performed (gestural space). Some gestures are intrinsically connected to a contact point (e.g. eyes, neck, chest, trunk, etc.), others are held in front of the person, without any contact point (as in the case of the alphabet).
- **movement of the hands**. The movement is characterized by the use of one or two hands and by the motion of the hands in the gestural space.

The non-manual component comprises:

- **body movement**. The body movement is responsible for introducing a temporal context. The torso leaning back, upright or forward indicates the communication in the past, present or future, respectively.
- **facial expressions**. The facial expressions add a sense of emotion to the speech, that is, a subjective experience, associated with temperament, personality and motivation. Two distinct emotions cannot occur simultaneously, since the emotions are mutually exclusive.

3 RELATED WORK

Although it is a relatively new area of research, in the last two decades a significant number of works focusing on the development of techniques to automate the translation of sign languages with greater incidence for the American Sign Language (Morrisey and Way, 2005), and the introduction of serious games in the education of people with speech and/or hearing disabilities (Gameiro et al., 2014) have been published.

Several of the methods proposed to perform representation and recognition of sign language gestures, apply some of the main state-of-the-art techniques, involving segmentation, tracking and feature extraction as well as the use of specific hardware as depth sensors and data gloves. Deusdado (Deusdado, 2002) writes about the use of new technologies for dissemination and teaching of sign language, highlighting the use of 3D models (avatars) in the translation of words to sign language. Bragatto, T. A. C. et al. (Bragatto et al., 2006) suggest the use of colored gloves and the application of a low complexity neural network algorithm for recognition of the hands configurations. The model has a recognition rate of 99,2%. Nicolas Pugeault et al. (Pugeault and Bowden, 2011) suggest a system for recognition of the hand configuration in the context of ASL, using the Microsoft Kinect to collect information about appearance and depth, and the OpenNI + NITE framework (Rezende and Tavares, 2013) to detect and track the hand. The collected data is classified by applying a random forests algorithm (Breiman, 2001), yielding an average accuracy rate of 49,5%. Cooper et al. (Cooper et al., 2011) use linguistic concepts in order to identify the constituent features of the gesture, describing the motion, location and shape of the hand. These elements are combined using HMM for gesture recognition. The recognition rates of the gestures are in the order of 71,4%. Vladutu et al. (Vladutu, 2009) propose the analysis of the video stream using singular value decomposition (Klema and Laub, 1980) to extract the representative images by Fuzzy-Logic (Klir and Yuan, 1995). Then the color and shape features are extracted using MPEG-7 descriptors and finally the classification of the hand configuration is made using a Support Vector Machine (SVM) (Steinwart and Christmann, 2008). The authors claim an error rate lower than 10%. McGuire et al. (McGuire et al., 2004) propose an ASL translator using sensors gloves in conjunction with the HMM. With the shown model, for a small vocabulary, they achieve an accuracy of 94%. The project CopyCat (Brashear et al., 2010) is an interactive adventure and educational game with ASL recognition. Colorful gloves equipped with accelerometers are used in order to simplify the segmentation of the hands and allow estimating the motion acceleration, direction and the rotation of the hands. The data is classified using HMM, yielding an accuracy of 85%.

4 THE VIRTUALSIGN

VirtualSign PTDC/CPE-CED/121878/2010 is a research project funded by the Portuguese government through the FCT - Fundação para a Ciência e a Tecnologia. The project has the participation of researchers from the School of Engineering of the Polytechnic of Porto, the Faculty of Engineering of the University of Porto and the Portuguese Open University and an expert in PSL.

VirtualSign aims to contribute to a greater social inclusion for the deaf. Its main motivation comes from a team of university teachers that have realized the difficulties in communicating with deaf students in the context of a class. The creation of a real time bi-directional translator between PSL and text is expected to facilitate the communication with students who have hearing disabilities. In addition to the bi-directional translator, this paper also presents a serious game directed to assist in the learning of the Portuguese Sign Language.

The project bundles three interlinked modules:

1. **Translator of PSL to Text** (Figure 1): module responsible for the capture, interpretation and translation of PSL gestures to text. A pair of sensors gloves (5DT Data Gloves) provides input about the configuration of the hands while the Microsoft Kinect provides information about the orientation and movement of the hands.

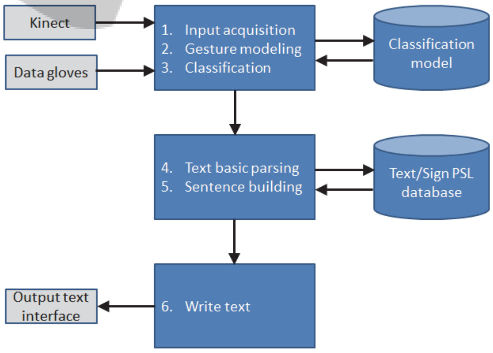


Figure 1: PSL to text translator.

2. **Translator of Text to PSL** (Figure 2): this module is responsible for the translation of text to PSL. The 3D models and animations used in this application to mime PSL were created in Blender. A MySQL database is used to store animation data. The main code is written in C# and all the features are merged together with Unity.

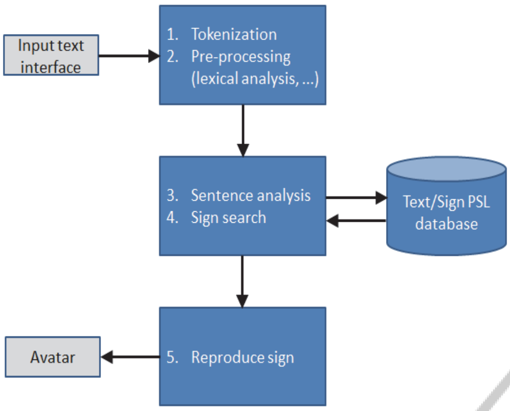


Figure 2: Text to PSL translator.

3. **Serious Game** (Figure 3): Module responsible for the didactic aspects which integrates the two modules above described into a serious game. This adventure game has several challenges that bring the basics of PSL to the scene introducing the player to the PSL alphabet, commonly used words and sentences.

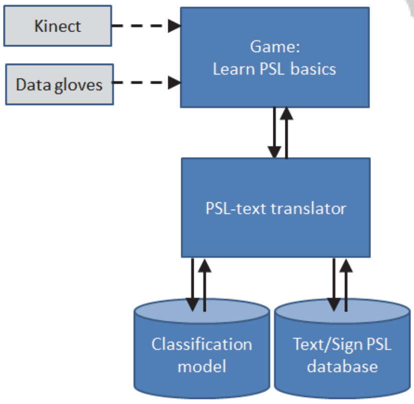


Figure 3: VirtualSign game and translator.

5 PSL TO TEXT

The translation from a visual representation of speech to text is a complex process that involves not only the ability to distinguish between words, but also to identify the beginning and end of each word in a full speech. Similarly to what happens with oral languages - in which the speakers reproduce their understanding of the language by introducing personal characteristics, in particular accent and speech speed- several features bring meaning to the communication in sign language. Gestures in sign language performed by different people usually have significant variations in terms of speed, range, movement etc. These characteristics require the adoption of flexible metrics to identify each word and to delimit the words in a sentence. To simplify we consider that a word corresponds to a gesture in PSL. A gesture comprises a sequence of configurations from the dominant hand, each associated with (possibly) a configuration of the support hand, and a motion and orientation of both hands. Each element of the sequence is defined as an atom of the gesture. The beginning of a gesture is marked by the adoption of a configuration by the dominant hand. The end of the gesture is marked either by the return of the dominant hand to a neutral position or by a configuration change. In the case of a configuration change, two scenarios may arise: the newly identified configuration is an atom of the sequence of the gesture in progress or the acquired atom closes the sequence of the gesture in progress and signals the beginning of a new gesture that will start with the following atom.

5.1 Hand Configuration

In PSL there is a total of 57 hands configurations. However, only 42 of those are completely different, since 15 pairs differ only in the orientation. Such is the case of the letters "M" and "W", as seen in Figure 4, where the configuration is similar and only the palm orientation changes.

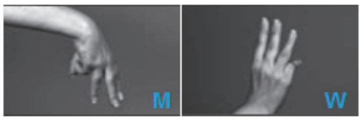


Figure 4: Hand configuration for the letters M and W respectively.

The configuration assumed by the hand is identified through classification – a machine learning setting by which one (eventually more) category from a set of pre-defined categories is assigned to a given object. A classification model is learned from a set of labelled samples. Then, this model is used to classify in real time new samples as they are acquired.

5.1.1 Hand Configuration Inputs

In order to obtain information to identify the configuration assumed by each hand over time, we use 5DT data gloves (5DT Data Glove Ultra, 2011). Each glove has 14 sensors placed at strategic locations of the hand's articulation and provides data at a maximum rate of 100 samples per second. The inputs received from the sensors in the data gloves are stable, but still a low level of noise is present. To improve the robustness of the data acquisition process and reduce the classification burden, we only retain one sample from the sensors output (for further classification) each time the data remains stable for a pre-defined period of time, after having detected a significant change.

5.1.2 Classification

Once having ensured stability of the data, we proceed with the classification of the configuration. During a preparatory stage we have compared the performance of six classification algorithms, namely Random Trees (RT) (Le Gall, 2005), Boost Cascade (BC) (Viola and Jones, 2001), Neural Networks (NN) (Haykin, 2004), K-Nearest Neighbours (KNN) (Cover and Hart, 1967), Naive Bayes (NB) (Rish, 2001) and Support Vector Machine (SVM). For all these algorithms we have used the default configuration of the respective implementation available in the Open Source Computer Vision Library (OpenCV) (Bradski and Kaehler, 2008). To evaluate their performance we have used a dataset composed of 40 samples for each hand configuration (1680 samples in total). To reduce the variance of our estimates we have used 10-fold cross validation. In Table 1 and Table 2 we present the results of the evaluation for each glove (right and left glove).

Table 1: Classification results of the 1680 samples, obtained with the use of the left glove.

%	RT	BC	NN	KNN	NB	SVM
Precision	98,6	82,0	98,1	98,8	97,5	98,6
Accuracy	85,5	95,4	78,1	97,3	97,1	100,0

Table 2: Classification results of the 1680 samples, obtained with the use of the right glove.

%	RT	BC	NN	KNN	NB	SVM
Precision	98,8	86,1	97,2	98,0	98,0	98,1
Accuracy	87,3	96,6	80,4	98,2	96,8	100,0

From these results, we may discard the Boost Cascade algorithm, by far the worst of all. We have also discarded Neural Network due to the high computational cost when in comparison to the rest. This is a serious drawback since we need a classifier to use in real time. The remaining four algorithms, present a high precision and accuracy. Based on these results we have opted to use SVM classifiers. For each configuration we have kept the top three instances and their associated probability, meaning that the application will take into consideration the tree configurations with the highest probability and their probability will be used in the classification to increase the accuracy. These instances were used later to build the classification model for word recognition.

A point to take into consideration is the fact that intermediate (fake) configurations that constitute only noise may occur during the transition between two distinct configurations. As example we can see in Figure 5 the transition from the configuration corresponding to the letter "S" to the configuration corresponding to the letter "B", where we obtain as noise an intermediate configuration associated that matches the hand configuration for number "5" in PSL.

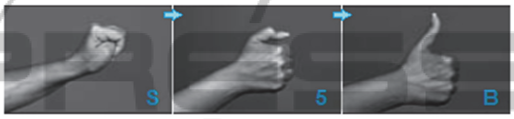


Figure 5: Transition from configuration S to configuration B, through the intermediate configuration (noise) 5.

Intermediate configurations differ from the others by the time component, i.e., intermediate configurations have a shorter steady time, which is a constant feature that may be used to distinguish between a valid configuration and a noisy, intermediate configurations. Thus, we use information about the dwell time of each configuration as an element of discrimination by setting a minimum execution (steady) time below which configurations are considered invalid.

5.2 Hand Motion and Orientation

To obtain information that allows characterizing the movement and orientation of the hands we use the Microsoft Kinect. In order to equalize the sampling frequency between Kinect and the data gloves, we reduced the frequency of sampling in the gloves to 30 samples per second. For each skeletal point, the Kinect provides information about the position in space (x, y, z) over time. Of the 20 points available we only use 6, in particular the points corresponding to the hands, elbows, hip and head. We consider that a gesture is valid only if the dominant hand is positioned above the hip, and a gesture is performed with both hands only if both hands are above the hip. Differences are notorious when there is a significant dissimilarity in the level of proficiency in sign language. The time that it takes to perform the gesture is one of the most prevalent differences. So, in order to remove the influence of the time in the classification of the gesture, we only save information about the motion when a significant movement happens, i.e. when the difference between the position of the dominant hand (or both hands), and the last stored position is greater than a predefined threshold. Therefore, when a significant movement is detected we save an 8-dimensional vector corresponding to the normalized coordinates of each hand (x_n, y_n, z_n) and the angle that characterizes its orientation. If the gesture is performed just with the dominant hand, the coordinates and angle of the support hand assume the value zero. The coordinates are normalized by performing a subtraction of the vector that represents the hand position (x_m, y_m, z_m) by the vector that defines the central hip position (x_a, y_a, z_a).

$$(x_n, y_n, z_n) = (x_m, y_m, z_m) - (x_a, y_a, z_a)$$

The orientation is calculated based on the angular coefficient defined by the straight line that intersects the hand (x_a, y_a) and the elbow (x_c, y_c).

$$Angle = \tan^{-1}\left(\frac{y_a - y_c}{x_a - x_c}\right)$$

In summary, for each configuration assumed by the dominant hand, a set of vectors characterizing the motion (position and orientation) of the hands are recorded.

5.3 Classification of PSL Gestures

A sequence consists of, at least, one configuration of the dominant hand. Each sequence may have a configuration of the support hand and a motion associated to it. A sequence corresponds to one or more words. Figure 6 shows the scheme of the gesture's fulfilment corresponding to the greeting "Bom dia" ("Good Morning" in English), and its representation in the translator's model.

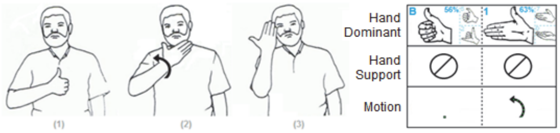


Figure 6: The PSL gesture corresponding to the greeting "Bom dia" and its representation in the translator.

By analyzing the previous illustration we see that the gesture is performed only with the dominant hand (support hand was not used), having occurred the transition between two configurations, being the most likely configurations the ones associated with the letter "B" and number 1 respectively (configurations of the "bom dia" gesture). In the first atom, the motion was nil so we only have one point in space, whereas in the second atom we have an arc-shaped movement.

Once the dominant hand returns to the neutral position (the construction of the sequence is ended) we can proceed with the classification of the resulting sequence. Each gesture in the database (previously built) is successively aligned with a fragment of the sequence in analysis of size equivalent to the gesture size. Each pair of aligned atoms is compared, being given a weight of one third to each of their components (dominant hand, support hand and movement). If the hand configuration matches, it is assigned a value corresponding to P_n multiplied by the maximum value of the component, where P_n correponds to the associated probability obtained previously in the classification of the configuration. If there is no match, a null value is given to the component. The comparison between the series of vectors that describe the movement is performed by applying the sequence alignment algorithm Dynamic Time Warping (Müller, 2017). Finally, the value obtained in the comparison of each pair of atoms is added and normalized by the number of atoms that composes the gesture. The highest ranked gesture shall, in principle, correspond to the gesture actually performed in this fragment.

5.3.1 Evaluation

In order to evaluate the performance of the translator, we performed several sequences of gestures in PSL. Each sequence was formed by a variable number of words from a set of 10 distinct words, in particular, "Ola" (W_0)(hello), "Bom Dia" (W_1)(good morning), "Boa tarde" (W_2)(good afternoon), "Boa Noite" (W_3)(good night), "Sim" (W_4)(yes), "Não" (W_5)(no), "Aprender" (W_6)(learn), "Ar" (W_7)(air), "Maça" (W_8)(aple), "Branco" (W_9)(white). To ensure that the algorithms can distinguish similar gestures, we have used pairs of words that present some similarities between them, such as words that use the same hand configurations differing only in the movement (e.g. "Bom dia" and "Boa tarde"), words that have the same motion but require different configurations of the hands (e.g. "Maça" and "Branco") and words that differ in the number of transitions of the dominant hand (e.g. "Boa noite" and "Aprender"). In the Table 3 we have the resulting confusion matrix.

In terms of results, we have achieved a significant precision in the order of 91.7% with real-time translation, which consists of a very positive result. For upcoming evaluations, it is necessary to expand the vocabulary to ensure that the performance remains at high standards.

Table 3: The confusion matrix. The last column of the matrix corresponds to the number of times there was no valid match for each word.

	W_0	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	X
W_0	27										3
W_1		26	4								0
W_2		1	29								0
W_3		3	2	25							0
W_4					26						4
W_5						27					3
W_6							29				1
W_7								27			3
W_8									30		0
W_9									1	29	0

6 TEXT TO PSL

The translation of Portuguese text to PSL is a quite demanding task due to the specific linguistic aspects of PSL. Such as any other language, the PSL has grammatical aspects that must be taken into consideration. For instance, the facial expression and the body position are relevant for the translation. The different body tilting changes the tense of the speech. When the body is tilted forward the text is being said in the future tense, if it is tilted backwards the text has to be in the past tense.

In order to support the deaf community this module has integrated an orthographic corrector. This feature aims to aid the user to understand any possible mistakes written for the translation, however it won't change what the user wrote, it simply points out that there may be a mistake. The avatar used for the translations was created in Blender (Hess, 2007) as well as its animations. The avatar body is identical to a human one in order to get the best quality in the translation as possible.

Figure 7: Avatar.

6.1 Structure

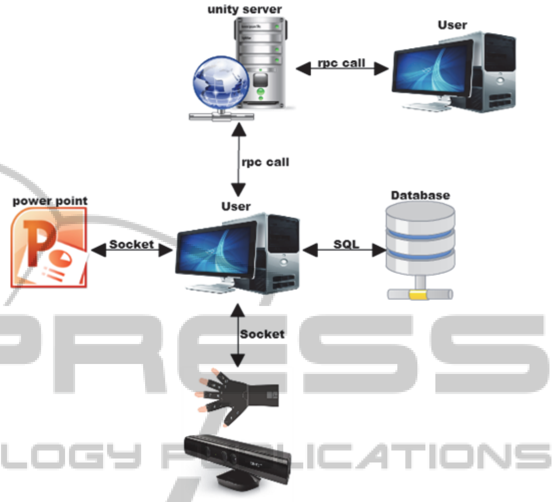


Figure 8: Structure.

The text to PSL translator module is divided in several parts. The connection to the Kinect and data gloves is based on sockets which will retrieve information from the Gesture to text translator in a string and the reason it is represented in the figure 8 is because the gestures performed by the user can be reproduced by the avatar. After the string containing the message is received the text to gesture translator replies to the sender letting it know that the animation was played and the text received was translated. This protocol was also used for the PowerPoint Add-in, however the text sent will not be coming from the translator but from the PowerPoint slide itself. The Add-in will send each word on the slide, highlight it and wait for the reply in order to continue so that the user knows what is being translated at the time. The database contains all the animation names and the corresponding text, based on this it becomes easier to know which gesture to perform based on the text. The database is MySQL (MySQL, 2001). During the translation process the application will search the database for the word that came as input, either written on the program or from other applications. When the text is found in the database there will be an animation assigned to it and that animation will be played. In case the text is not in the database the avatar will proceed to translate that text letter by letter.

A chat was also created within this project so that all the features could be used on it to improve the integration with this community. The chat works using the Unity server and resorts to RPC calls to the server, being able to support up to twenty persons at the same time.

6.1.1 Architecture

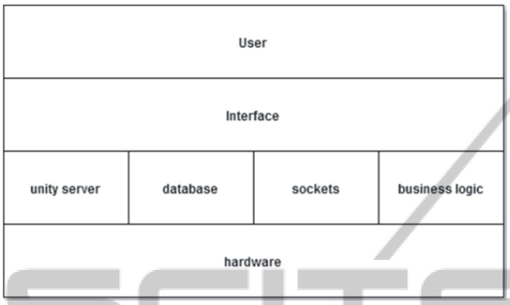


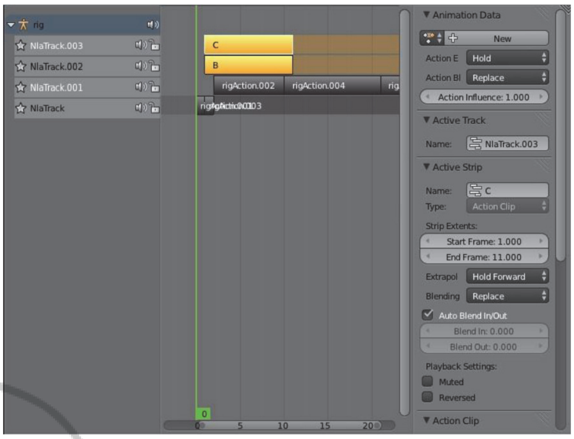
Figure 9: Architecture.

The architecture is organized in two main layers. First, we have the interface layer that allows a user to interact with the functionalities of the module. The second layer is divided into four parts, including the sockets, the web-service, the database and the business layer. The business layer is the layer that implements the functionalities, providing its services to the layer immediately above. The web-service layer is in charge of making connections to the server to allow the connection from multiple devices. The layer sockets is in charge of linking the application with the Kinect so you can get answers from the recognition made by translator and also to connect the translation system with the power point. Finally, there is a database in charge of storing the information necessary for translation, such as the name of the animations. In communication between layers, the topmost layer can communicate with the lower level.

6.2 Gesture Animations

After understanding Blender's features, it was necessary to understand how to perform each animation. First all the animations were made continuously, however at the end of this experience it proved to become very difficult to work with them later. So we decided to make every gesture apart in a new nlaTrack (lines that keep the position of the body in motion frames), thus making the use of the animations easier and making them independent from each other.

As the previous figure reveals to create the animations it is necessary to treat the animation with



▼ Animation Data

New

Action E: Hold

Action Bl: Replace

Action Influence: 1.000

▼ Active Track

Name: NlaTrack.003

▼ Active Strip

Name: C

Type: Action Clip

Strip Extents:

Start Frame: 1.000

End Frame: 11.000

Extrapol: Hold Forward

Blending: Replace

☑ Auto Blend In/Out

Blend In: 0.000

Blend Out: 0.000

Playback Settings:

□ Muted

□ Reversed

▼ Action Clip

Figure 10: NlaTracks.

the nlatracks. In the nlatracks the information referring to the animation was kept, such as the initial frame and all the necessary frames that create the motion for the full animation. All the information was then stored in Animation Data where all nlaTracks are stored. The animation name created was also stored in the same nlaTrack.

As for the creation of frames it was necessary to create a path for frames that Blender could plug in gently and creates the desired effect for the animation.

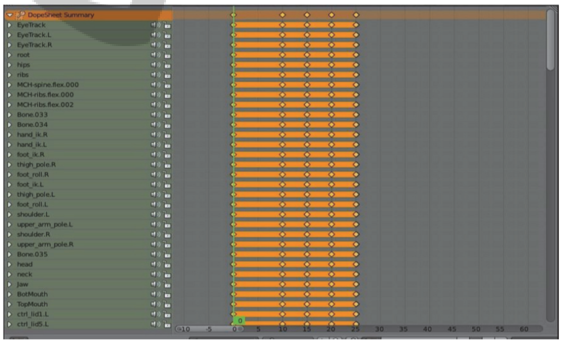


Figure 11: Frames.

With support of the previous figure it is possible to understand how the various frames must be coordinated. The keyframes(frames containing the position of the avatar) were placed, for instance at frames: 10, 15, 20 and 25. That way we can create a fluid and smooth animation without having to deal with the moments between each frame stored in nlatrack. The avatar contains a Rig (armor and avatar bones), this armor is used in order to move the Avatar body as intended.

7 SERIOUS GAME

Learning a new language is always a complex task which requires a considerable amount of effort and time. Resorting to gamification a serious game was developed in order to smooth the learning process of the Portuguese Sign Language. The game has three main scenes, each representing a different part of PSL. The first scene is where the user learns the alphabet. The second scene has several of the most used words and finally the third and last scene contains frequently used sentences. The learning process is based on the observation of the avatar and the performance of the player using the Kinect. There is an inventory system which stores all the gestures acquired by the player, those gestures can be accessed at any time and performed by the avatar, that is how the player will learn to correctly execute the gestures. All those signs are scattered around the scenes and some must be acquired by challenges, sometimes requiring to perform a sign in order to obtain another, or beating a fixed score on a mini-game.

7.1 Non-functional Requirements

Accessibility: In order to play this game the user only needs a computer with a windows OS (XP or higher), however to use all of its features the user also needs the VirtualSign translator together with the Kinect and 5DT gloves.

Deployment: The game was created in Unity (Creighton, 2010) and it's mostly programmed in C# however it only requires Direct3D (Trujillo, 1996). Although not entirely necessary it is recommended to have the graphic card drivers updated.

Usability: As a serious game this application has a very intuitive interface and controls and it's carefully explained how to use them.

Performance: Good performance is always a must especially in a game where the feedback is instantaneous, therefore this game was optimized and tested to ensure that in a modern machine it runs at 60 FPS(frames-per-second) and never drops under 30. The recommended values for games are 40 FPS (Claypool et al., 2006).

Interoperability: The game will be able to connect to the VirtualSign translator and its features will be used in the game checkpoints.

7.2 Game Structure

The performance of the application must always be one of the main concerns. The application structure was created aiming to be the most efficient and intuitive as possible.

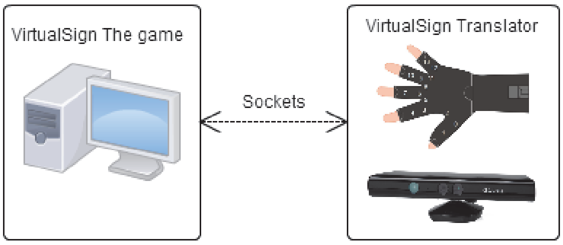


Figure 12: Interaction between The game and the translator.

The communication between the game and the translator is done by sockets connecting both applications (Wegner, 1996). This connection is created upon choosing the option to play with Kinect and stays idle until a player reaches a checkpoint, then it will start waiting for the input. The game waits for the translator to send a message, this message will be the text generated by the analysis and translation of the sign done by the user. The following architecture was implemented:

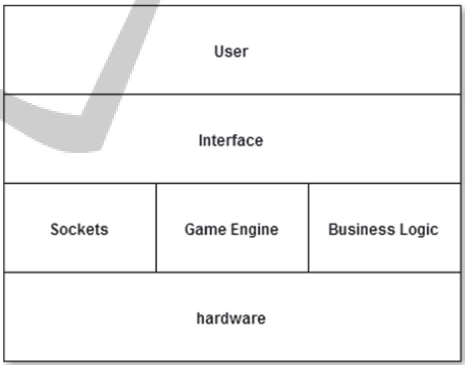


Figure 13: Game architecture.

The project has several layers due to its high rate of complexity. The interface is at a high layer as it is the one responsible for the user interaction with all the game features, therefore the interface communicates with the layers below. In the second lowest layer there are three parts, sockets which is the part responsible for the connection between the translator and the game, the game engine which is the base of the game itself, and finally the business logic which is the main functions of the game.

As shown in the figure above, the project has 7 main packages that contain the scripting for each section of the game. The environment represents the scripts for the random events within the scene, the animation as the name suggests has the base for all the animations, the Interface has the user interfaces

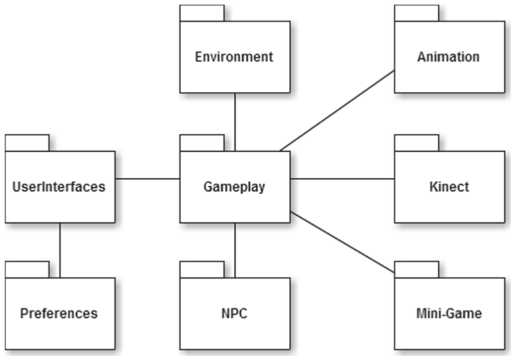


Figure 14: Package Diagram.

such as the main menus, inventory and quest windows. The Kinect is where the connection scripts based on sockets are stored. The NPC(non-player character) contains the scripting for the events that each NPC triggers. The gameplay is where the basic functions are, such as the inventory, movement, interaction and highscore. The preferences is very linear as it is where the audio, graphics and saves are stored and finally the Mini-games which contains all the information and scripts.

8 CONCLUSIONS

With the introduction of the bidirectional translator of Portuguese Sign Language and the serious game, this project aims to contribute to the improvement of accessibility in terms of communication of people with disabilities in speech and/or hearing. The translation module from PSL to text, although evaluated with a small vocabulary, presents very promising results, allowing a translation in real-time with high precision, above 90%. However, there is still a long way to go, mainly due to our objective to include this module in portable devices. We intend to eliminate the need of using the data gloves in the project, replacing them with other solutions that permit the identification of the configurations assumed by the hands and the tracking of movement as well as the detection of facial expressions, increasing the portability and reducing the costs of implementing the translator.

In the text to PSL translator there is still some effort needed in order to cover the Portuguese sign language in full. PSL has about 10000 different words. It is also necessary to improve the spell checker to become more efficient, currently there are 500 words covered by this tool and the corresponding animations.

Finally, the serious game aims to improve the knowledge of its users in order to provide an easier integration within the deaf community. Exploiting the motivation that drives users to play a game and putting it to use in order to make them learn while enjoying the gameplay smoothens the learning process and allows the user to practice PSL while getting rewarded for it.

The applications of the VirtualSign bi-directional translator between PSL and text go far beyond its original motivation linked to teaching and classroom settings. We have a broad range of areas where such a tool will significantly improve the life quality of the deaf and foster the effectiveness of the daily communication with hearing impaired persons.

AKCNOWLEDGEMENTS

This work was supported by the ISEP (Instituto Superior de Engenharia do Porto) and GILT (Games, Interaction and Learning Technologies) under Grant Nos. 60973089. FCT (Fundação para a ciência e Tecnologia) project.

REFERENCES

Larsson, Stig. "A Review of Claes G. Olsson: Omsorg och kontroll-en handikapphistorisk studie 1750-1930 (Care and control-an analysis of the history of disability) Umeå 2010." *Vulnerable Groups & Inclusion 2* (2011).

Associação de Surdos do Porto - *Algumas definições úteis sobre a Surdez*, http://www.asurdosporto .org.pt/artigo.asp?idartigo=77, January 2015.

Morgado, Marta, and Mariana Martins. "*Língua Gestual Portuguesa.*" Comunicar através de gestos que falam é o nosso desafio, neste 25. º número da Diversidades, rumo à descoberta de circunstâncias propiciadoras nas quais todos se tornem protagonistas e interlocutores de um diálogo universal.: 7.

Morrissey, Sara, and Andy Way. "*An example-based approach to translating sign language*." (2005).

Gameiro, João, Tiago Cardoso, and Yves Rybarczyk. "Kinect-Sign: Teaching Sign Language to "Listeners" through a Game." *Innovative and Creative Developments in Multimodal Interaction Systems. Springer Berlin Heidelberg*, 2014. 141-159.

Deusdado, Leonel Domingues. *Ensino da língua gestual assistido por personagens 3D virtuais*. Diss. Universidade do Minho, 2002.

Bragatto, T. A. C., G. I. S. Ruas, and M. V. Lamar. "Real-time video based finger spelling recognition system using low computational complexity Artificial Neural Networks." *Telecommunications Symposium*, 2006 International. IEEE, 2006.

Pugeault, Nicolas, and Richard Bowden. "Spelling it out: Real-time asl fingerspelling recognition." *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011.

Rezende, Lorena A., and Dalton M. Tavares. "OpenNI e Suas Aplicações." *X Encontro Anual de Computação* (2013).

Breiman, Leo. "*Random forests." Machine learning 45.1* (2001): 5-32.

Cooper, Helen, Nicolas Pugeault, and Richard Bowden. "Reading the signs: A video based sign dictionary." Computer Vision Workshops (ICCV Workshops), 2011 *IEEE International Conference on.* IEEE, 2011.

Vladutu, Liviu. "Non-rigid shape recognition for sign language understanding." *WSEAS TRANSACTIONS on SYSTEMS 8.12* (2009): 1263-1272.

Klema, Virginia, and Alan J. Laub. "The singular value decomposition: Its computation and some applications." *Automatic Control*, IEEE Transactions on 25.2 (1980): 164-176.

Klir, George, and Bo Yuan. *Fuzzy sets and fuzzy logic*. Vol. 4. New Jersey: Prentice Hall, 1995.

Steinwart, Ingo, and Andreas Christmann. *Support vector machines*. Springer, 2008.

McGuire, R. Martin, et al. "Towards a one-way American sign language translator." Automatic Face and Gesture Recognition, 2004. *Proceedings. Sixth IEEE International Conference on*. IEEE, 2004.

Brashear, H., et al. "CopyCat: A Corpus for Verifying American Sign Language During Game Play by Deaf Children." *4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*. 2010.

5DT Data Glove Ultra - User's Manual, http://www.5dt.com/downloads/dataglove/ultra/5DT *Data Glove Ultra Manual v1.3.pdf, Version 1.3*, January 2011.

Le Gall, Jean-François. "Random trees and applications." *Probab. Surv 2.245-311* (2005): 15-43.

Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. *Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1*. IEEE, 2001.

Haykin, Simon (2004) Neural Network. "A comprehensive foundation." *Neural Networks 2*.2004.

Cover, Thomas, and Peter Hart. "Nearest neighbor pattern classification." Information Theory*, IEEE Transactions on 13.1* (1967): 21-27.

Rish, Irina. "An empirical study of the naive Bayes classifier." *IJCAI 2001 workshop on empirical methods in artificial intelligence. Vol. 3*. No. 22. 2001.

Bradski, Gary, and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

Müller, Meinard. "Dynamic time warping*." Information retrieval for music and motion* (2007): 69-84.

Hess, Roland. The essential Blender: *guide to 3D creation with the open source suite Blender.* No Starch Press, 2007.

MySQL, A. B. "*MySQL*." (2001).

Creighton, Ryan Henson. Unity 3D Game Development by Example: *A Seat-of-Your-Pants Manual for Building Fun, Groovy Little Games Quickly*. Packt Publishing Ltd, 2010.

Trujillo, Stan. *Cutting-edge Direct3D programming: everything you need to create stunning 3D applications with Direct3D*. Coriolis Group books, 1996.

Claypool, Mark, Kajal Claypool, and Feissal Damaa. "The effects of frame rate and resolution on users playing First Person Shooter games." Electronic Imaging 2006. *International Society for Optics and Photonics*, 2006.

Wegner, Peter. "Interoperability." *ACM Computing Surveys (CSUR) 28.1* (1996): 285-287.