Virtual Avatars Signing in Real Time for Deaf Students

Lucía Vera, Inmaculada Coma, Julio Campos, Bibiana Martínez and Marcos Fernández
Institute of Robotics (IRTIC), University of Valencia, Valencia, Spain

Keywords: Virtual Sign Animation, Speech to Gesture Translation, Virtual Characters, Avatar Animation.

Abstract: This paper describes a speech and text translator from Spanish into Spanish Sign Language, that tries to solve some of the problems that deaf people find when they access and attend specific training courses. In addition to the translator system, a set of real-time avatar animations representing the signs are used. The creation process of such avatars is also described. The system can be used in courses where deaf and hearing people are sharing the same material and classroom, which contributes to improve the integration of this group of people to specific academic areas. The tool has been tested to obtain direct information with a group of deaf people from the Deaf Association of Seville.

1 INTRODUCTION

The constant evolution of computers and portable devices make it possible to develop applications to help impaired people in some aspects of their life. This is the case of deaf people. In the field of computer graphics, some efforts have been made to develop systems that help hearing impaired people in their communication difficulties.

The tools that have been developed help deaf people to communicate with ordinary people through the computer. Some of them are intended to translate text or speech to a sign language by means of virtual avatar animations representing the signs. These systems have two main problems to solve: (1) to translate a speech or written text from a language to a sign language; (2) to create movements for the virtual avatars that can be, at least, understandable by deaf people (Kipp, 2011), and, even better, with a certain level of fluency. These problems are addressed in the present research.

An additional problem is the existence of many different standards, as there is no standard not even in any given country. In this paper a speech and text translator from Spanish into Spanish Sign Language, *Lenguaje de Signos Español* (LSE) is addressed. Moreover, the process of creating a virtual avatar signing in real-time is described. The result of the research is applied to solve some of the problems that deaf people find when they try to access and attend specific training courses.

Thus, when a company offers different training courses, it is difficult for deaf people to enrol in these courses because of the costs associated to hiring a translator. If a special sign language application for deaf students is offered, this can become a good starting point to improve their integration.

This is the main objective of the present research. An automatic system for translating Spanish to LSE has been developed which supports the simultaneous translation of voice and PowerPoint presentations to LSE. There is also an avatar signing in real time (RT), with a special chat between deaf students and his/her teacher who speaks LSE.

The paper is organised as follows: Section 2 shows a review of the literature. Section 3 presents the system architecture. Section 4 and 5 describe the different modules and the interface in detail. Section 6 explains the main problems and our solutions. Sections 7 and 8 show the evaluation results, the main conclusions and future work.

2 RELATED WORK

After a revision of the current literature, we have classified the different efforts made to keep communication between deaf people and hearing people. Some researchers have focused their work on gesture recognition, trying to automatically recognize a sign from a specific sign language (Liang, 1998, Sagawa 2000). Other researchers have

Vera L., Coma I., Campos J., Martínez B. and Fernández M..
Virtual Avatars Signing in Real Time for Deaf Students.
DOI: 10.5220/0004293102610266
In *Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications* (GRAPP-2013), pages 261-266
ISBN: 978-989-8565-46-4

focused their work on developing systems that automatically translate into a sign language, most of which use virtual avatars to represent a speech or a written text in a sign language.

One of the main projects in this research line is VISICAST (Bangham, 2000), which translates English text into British Sign Language. The main effort of this project has been to process natural language by means of a parser that identifies functional words and resolves ambiguities by using an SGML notation (Elliot, 2000). Once a text is translated, signs are displayed using computer animation of a Virtual Avatar in two steps: first, recording a human signer and, then, post-processing the captured sequence. This system can be applied to subtitle television programs or create web pages (Verlinden, 2001).

Other projects of the same research group use virtual signing technology. In eSign, avatars are used to create signed versions of websites (Verlinden, 2005), and in the SignTel avatars are added to a computer based assessment test that can sign questions for deaf candidates. In eSign project (Zwiterslood, 2004), instead of using motion capture to generate the avatar animation, a temporal succession of images is used, each of which represents a static posture of the avatar. The signs are described through a notation for Language Signs (HamNoSys), which define aspects of hand position, speed and gesture amplitude.

Research with Japanese Sign Language (JSL) has been also made, such as Kuroda (Kuroda, 1998) who developed a telecommunication system for sign language by utilizing virtual reality technology, which enables natural sign conversation on an analogue telephone line.

Kato (Kato, 2011) uses a Japanese for a JSL dictionary with 4,900 JSL entries that have an example-based system to translate text. After that, the system automatically generates transitional movements between words and renders animations. This project is intended to offer TV services for deaf people, especially in case of a disaster, when a human sign language interpreter is not available. In this research a number of deaf people have been asked to watch the animations pointing out its lack of fluency fluency as sign language (Akihabara, 2012).

Regarding Spanish Sign Language, San-Segundo (San-Segundo, 2012) describes a system to help deaf people when renewing their Driver's License. Such system combines three approaches for the language translator: an example-based strategy, a rule-based translation method and a statistical translator that uses VGuido virtual animations. Thus, to create a sign animation an XML application which supports the definition of sign sequences is used, which mixes static gestures with movement directions. After that, a movement generation process computes the avatar's movements from the scripted instructions.

After the literature revision, it seems that researchers have not found applications for the same purpose that use virtual avatars signing in real time.

3 SYSTEM ARCHITECTURE

As we mentioned above, this application tries to improve and facilitate the communication between a deaf student and a teacher in an academic area, not only providing a translation of voice and powerpoint presentations in LSE, but also offering a special chat where they can communicate in both directions in an understandable way. This will improve the deaf people's access to specific training courses. For this reason, the architecture implemented is based on a client-server structure. The main application (teacher side) is a system based on a hierarchical architecture, where each module is responsible for a specific part in the translation, communication and visualization process. In the student side (client), the application centres its functionality on the communication process (Figure 1).

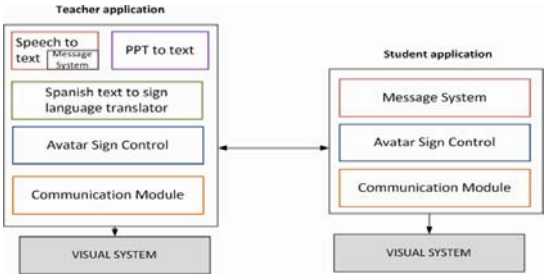

Figure 1: Modules included in each application.

In the following section each module is described.

4 SYSTEM COMPONENTS AND FUNCIONALITY

From the global architecture five main modules are described in the following subsections (Figure 2).

4.1 Speech to Text Module

The Speech to text module contains a Speech

recognizer to analyse the voice input and a Message system to manipulate the text data. The speech recognizer is responsible for generating a collection of words or sentences from a voice input and provides them to the following modules. This module uses Microsoft Speech SDK 5.1, included in every Windows 7 installation by default. This software has the advantage that it does not need to be installed and configured by users and, also, these can be trained as the software is being used.

The main problem found in the recognition process is how to detect when a sentence starts and finishes. The recognizer can provide all the words it collects from the microphone, but it is important to have the complete sentence so as to determine the best method to be applied for translation In Section 6 we analyse the problem and explain the solution.

The Message system is responsible for showing the sentences involved in the translation process in the interface. In the application a "debug mode" is included so that it is possible to control the results of the different steps of the translator. This is useful because the user can detect if there has been any mistake or problem in the recognition step or later, stay informed about all the intermediate steps of the translation process and act accordingly to solve any problem.

4.2 PPT to Text Module

A PowerPoint presentation (PPT) is a very important complementary piece of information for the speech in class. In our case it can help and solve the possible mistakes that the recognizer can make when recognizing oral explanations. Some deaf people cannot read and, therefore, it is necessary to provide a translation of the PPT that gives deaf students all the information taught in class.

The PPT to text module reads the PPT opened by the teacher; then, a collection of sentences in each slide are separated from a collection of words in each sentence. The translation module makes a translation (Section 4.3), one o more possible translations are then available in this module, selecting one as a default (depending on a measure of the probability of success).

We include in our application a Slide Editor with a wizard so as to facilitate the revision and correction of the PPT conversion into LSE, which can be made by the same teacher or with help of a person that has LSE knowledge. Also, by using this wizard it is possible to determine the how the words will be spelled in the translation. The revision changes can be saved to be later used for the PPT translation in the application. A remote control is also provided, which is helpful not only to manage the application, but also the global PowerPoint presentation shown to all students on the main screen

4.3 Text to Sign Language Translator

The text to sign language translator module receives the final Spanish sentence and analyses it so as to generate a correct translation into LSE. This will later be signed by a virtual character in real time. For that translation, two complementary methods are used: a "rule-based method" and an "analogy method".

As we will explain later, the system has a vocabulary including specific and general words and several common sentences, all of which have been classified into different categories to facilitate the translation.

First of all, each word, from the collection of words provided by the previous modules, is classified into one of the pre-established categories. For example: a sentence like "*IN THE COURSE WE WILL WORK WITH COMPUTER*" will be categorized as "*in*<none> *the*<none> *course*<object> *we*<personal pronoun> *will*<future> *work*<verb> *with*<none> *computer*<objet>" and translated as "FUTURE WE COURSE COMPUTER WORK".

Next, we apply the "analogy method", consisting of searching for coincidences with pre-recorded sentences in the vocabulary. If a corresponding sentence is found, it is directly used as the final translation.

In case this method fails, a hierarchical system of rules is applied. As we mentioned before, there is no standard associated to LSE, and, therefore there is no defined grammatical rules that can be applied and codified for translation purposes. In our case, different experts in LSE have extracted a collection of rules that are useful to obtain correct LSE sentences. Using these codified rules, the system tries to find the most adequate one for the sentence being translated, starting from the most particular and moving towards more general ones. When a coherent rule is found, the module applies it and generates the corresponding translated phrase. If after the process there is no specific rule to be applied, we have a general case rule at least to try to give the information to the student. This rule-based methodology allows providing translations of any sentence, giving greater flexibility, only with the limits of the vocabulary available.

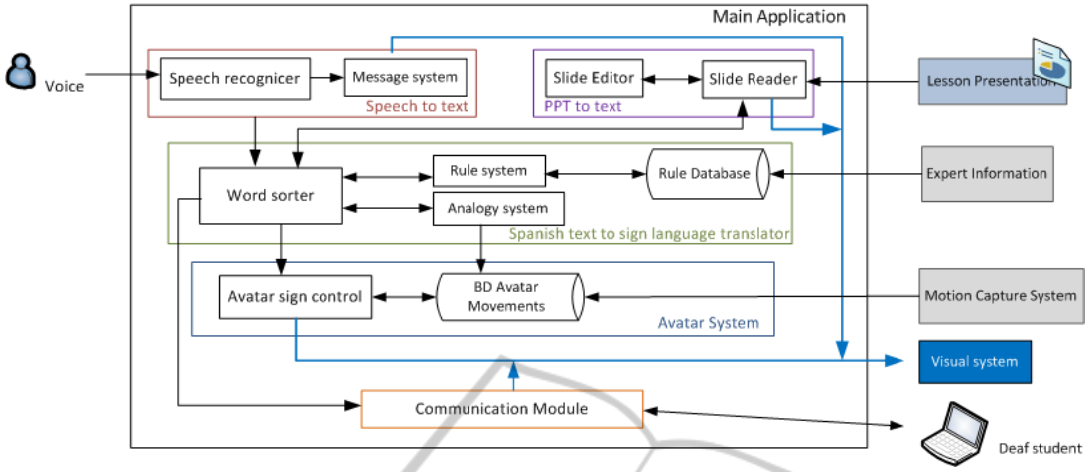

Figure 2: Components included in each module in the Main Application.

4.4 Avatar System Module

The main functionality of the avatar system module is to represent, control and animate realistically the virtual character integrated in the application so as to reproduce the signs required for the translation of the original sentence.

For this purpose, a vocabulary of around 2,000 words has been captured, for general and specific conversations in the academic area of interest. Also, the application can spell words using the alphabet for the translations of names, concepts, etc. To improve and speed the translation, a collection of 150 sentences common in all-purpose and classroom conversations has been included in the vocabulary. This kind of vocabulary is more convenient than those only based on complete sentences, because it has more flexibility when translating any phrase by using words in the vocabulary. It is however true that it is more susceptible to translation mistakes or errors due to the non-standard rules used.

All vocabulary is recorded by an LSE expert using a motion capture system. An 11-infrared cameras setting (NaturalPoint system) and a suit including 34 reflective balls for recording the body motion is used. Also, two data gloves (CyberGlove with 18 sensors) record all fingers. Each movement captured is adapted to the avatar using MotionBuilder and 3d Studio Max, producing a final version of each sign that has no wrong or missed data.

A system based in cal3D and OpenSceneGraph, to integrate, control and animate avatars in real time has been developed. Using this system an avatar can be integrated in the interface, and controlled in RT so as to simulate the sign language speaking (Figure 3).

Once the final sentence translated into LSE is obtained, the Avatar system receives it as a collection of individual words. This must be reproduced fluently in a fluent and connected way to represent a sentence or as a complete sentence (without any intermediate processing).

Figure 3: Avatar simulating the sign “Navigate”.

In the first case, this module has to obtain the individual signs from the motion database and interpolate them so as to connect and reproduce all the words in the sentence sequentially. Reproducing each sign individually will give the impression of a robotic and non-realistic animation, and may hinder the comprehension by deaf people. For this reason, the words comprising the sentence are analysed and the best interpolation points between them are determined, based on the main motion curve of the arms (which motion starts and finishes in a pre-established reference position). Next, all the words in the sentence can be reproduced in realistically and fluently, applying an interpolation between frames of one sign and the following one. To complete the avatar motion, different automatic facial animations have been created so as to provide a more realistic human aspect.

4.5 Communication Module

The communication module is used in both sides and is involved in the transmission of information between the different systems components. The module is important for several reasons: On the one hand, this module is responsible for sending all the control information between applications. On the other hand, this module includes the chat provided that facilitates communication between a teacher and his/her deaf students. The deaf student can send messages in text mode and receives the information in text and LSE format. The teacher sends and receives the information in text mode; this is helpful for people with little or no LSE knowledge.

5 APPLICATION INTERFACE

The application interface is similar in both sides; it includes specific elements for the teacher to start and manage the PowerPoint presentation directly.

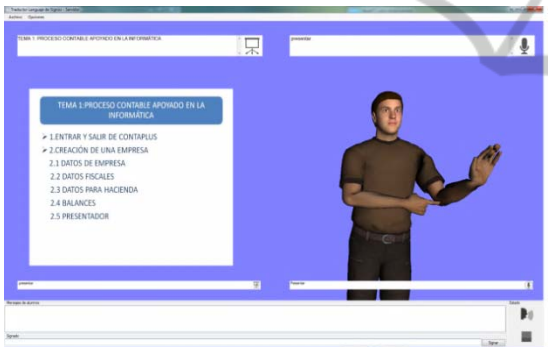

Figure 4: Teacher interface, with the control of PowerPoint and speech translations.

The application general menus allows the user to load PPTs and configure different elements of the interface, such as the avatar costume, background colour, signing speed, text font, main signing hand and interface mode. There are two possible modes that the teacher can activate with the remote control: PowerPoint (left side) or speak mode (right side) (Figure 4). In the PowerPoint mode the user is provided with information about the current phrase to be translated, the slide of the presentation, and the translation of the phrase into LSE. In the speaking mode, the user is given information about the original phrase recognized by the Recognizer (over the avatar) and the translation to the LSE (under the avatar). The lower part of the application has communication facilities (including the chat) and information about the state of the system.

6 PROBLEMS AND SOLUTIONS

During the development of the application important problems have been found that need to be solved so that the application is successful.

In the Speech Recognizer it is difficult to determine when a phrase starts and ends. To solve this problem the recognizer has been configured with different rules, such as a 1-second-silence as a delimiter between sentences .A good methodology serves to separate 90% of the sentences correctly.

Moreover, as the LSE language sign is a non standardized language, it has been necessary to determine a collection of rules to be implemented in the system and apply them to translate the sentences as best as possible. Related with the previous problem, it has been found that the LSE establishes a relation between the words in a sentence that goes from the most general to the most particular. This implies that it is necessary to establish a semantic relation between the different parts of the sentence. For that reason, a property of dimension has been assigned to each word, so the words are ordered in the sentence according to this property.

For the chat system the main problem found has been that some deaf people cannot write. This will be solved including pictograms in the chat.

After the evaluation of the application it has been found that not all the deaf people have the same literacy and need the same speed when signing. The same happens with the visual information provided. For that reason several customized elements in the interface have been created (see Section 5) to facilitate the deaf people understanding and the teacher usage.

7 EVALUATION

To evaluate the system and analyze possible problems a complete test with one teacher and a group of deaf people has been performed. The main objective of the session was to obtain a direct feedback from the deaf people about the possibilities and suitability of this tool for helping them in specific courses. The test consisted of a presentation lesson taught by a teacher (with little knowledge of LSE) to a group of deaf people from the Deaf Association of Seville. The teacher explained a lesson using a PPT, exchanging between voice and PowerPoint. The audience were able to communicate with her using the chat in the application by means of a laptop. After this test, the deaf people comment their impressions about the

application and answered a questionnaire with detailed information about their opinion. The application was improved using the suggestions.

The main conclusions obtained were that they considered useful to be able to select the avatar main hand to sign, because the different signing method of the deaf people and as they have different literacy levels, the speed of the avatar should be configured.

Regarding the sign language must be extended with more vocabulary adapted to the specific areas and updated with neologisms.

As for the application some signs were not clear enough and had little vocabulary, some movements were rough and needed to be smoothened, and the avatar had little facial expression.

At last, the tool presents some advantages and improvements from existing ones. From their point of view, it is a very useful system for on-line courses or as a visual book. In a short time, this application will be useful in secondary school classrooms, to study at home or review Spanish material, because digital written materials are difficult for them (some of them cannot read). Finally, in their opinion it is an incredible system in the area of new technologies, original and it will suppose a big transformation in the training of deaf people, making training and academic courses more accessible for them.

8 CONCLUSIONS AND FUTURE WORK

In this paper an automatic Spanish to LSE translator for academic purposes, from voice and PowerPoint data, was explained, reviewing all the application functionality and the results of the test carried out to obtain direct information from deaf people.

After the test, the system was completed to incorporate some ideas and solve some problems detailed by the deaf group, but we still have some improvements to be done as future work.

It is necessary to improve the avatar facial expression, adding gestures that complete the different signs and make them more understandable. It is important to review the clarity of all the vocabulary, making a deep test with deaf people. The improvement of the chat between teacher and student is been under development, to incorporate a visual interface with signs pictograms and animations in the student side, to facilitate its use by deaf people.

After these improvements, a new test will be necessary to check if the system will be useful enough to incorporate it in some specific courses.

ACKNOWLEDGEMENTS

The authors thank the Junta de Andalucía for their collaboration in the project, to Elena Gándara for her expert collaboration and help in the rules extraction and signs capture and the Deaf Association of Seville for their participation in the test.

REFERENCES

Akihabara. 2012 NHK introduces Automatic Animated Sign Language Generation System. In Akihabara News webpage. http://en.akihabaranews.com/.

Bangham, J. A., et al., 2000. Virtual Signing: Capture, Animation, Storage and Transmission An Overview of ViSiCAST. In *IEE Seminar on Speech and language processing for disabled and elderly people*.

Elliot, R. , Glauert, J. R. W. , Kennaway, J. R. , Marshall, I., 2000. The Development of Language Processing Support for the ViSiCAST Project. In *ASSETS 2000.*

Kato, N., Kaneko, H., Inoue, S., 2011. Machine translation to Sign Language with CG-animation. In *Tehcnical Review*. No. 245.

Kipp, M. Heloir, A. , Nguyen, Q. 2011. Sign Language Avatars: Animation and Comprehensibility.IVA 2011, LNAI 6895, pp. 113-126.

Kuroda,T., Sato, K., Chihara, K., 1998. S-TEL: An avatar based sign language telecommunication system. In *The International Journal of Virtual Reality. Vol 3, No.4.*

Liang,R. , Ouhyoung, M. ,1998. A real-time continuous gesture recognition system for sign language. In *Automatic Face and Gesture Recognition, Proc. Third IEEE International Conference on* pp.558-567.

Sagawa, H. and Takeuchi, M. , 2000. A Method for Recognizing a Sequence of Sign Language Words Represented in a Japanese Sign Language Sentence. In *Proc. of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000* (FG '00). *IEEE Computer Society,*

San Segundo et al, 2012. Design, Development and Field Evaluation of a Spanish into Sign Language Translation System. In *Pattern Analysis and Applications*: Volume 15, Issue 2 (May 2012), Page 203-224,

Verlinden, M. , Tijsseling,C. , Frowein H. ,2001. A Signing Avatar on the WWW. In *International Gesture Workshop 2001*.

Verlinden, M., Zwitserlood, I., Frowein, H. , 2005. Multimedia with Animated Sign Language for Deaf Learners. In *World Conference on Educational Multimedia,Hypermedia & Telecommunications,* pp. 4759-4764.

Zwiterslood, I. , Verlinden, M. , Ros, J. , van der Schoot, S. ,2004. Synthetic Signing for the Deaf: eSIGN. In *Proc. of the Conf. and Workshop on Assistive Technologies for Vision and Hearing Impairment, CVHI 2004,* 29