Advanced Speech Communication System for Deaf People

R. San-Segundo, V. López, R. Martín, S. Lufti, J. Ferreiros, R. Córdoba, J.M. Pardo

Speech Technology Group at Universidad Politécnica de Madrid
lapiz@die.upm.es

Abstract

This paper describes the development of an Advanced Speech Communication System for Deaf People and its field evaluation in a real application domain: the renewal of Driver's License. The system is composed of two modules. The first one is a Spanish into Spanish Sign Language (LSE: Lengua de Signos Española) translation module made up of a speech recognizer, a natural language translator (for converting a word sequence into a sequence of signs), and a 3D avatar animation module (for playing back the signs). The second module is a Spoken Spanish generator from sign-writing composed of a visual interface (for specifying a sequence of signs), a language translator (for generating the sequence of words in Spanish), and finally, a text to speech converter. For language translation, the system integrates three technologies: an example-based strategy, a rule-based translation method and a statistical translator. This paper also includes a detailed description of the evaluation carried out in the Local Traffic Office in the city of Toledo (Spain) involving real government employees and deaf people. This evaluation includes objective measurements from the system and subjective information from questionnaires.

Index Terms: Deaf people, Spanish Sign Language (LSE), Spoken Language Translation, Driver's License renewal, e-Inclusion.

1. Introduction

In 2007, the Spanish Government accepted the Spanish Sign Language (LSE: Lengua de Signos Española) as one of the official languages in Spain, defining a plan to invest in resources in this language. The development of a speech communication system for deaf people described in this paper is part of this Spanish government's plan.

At present, 92% of the Spanish Deaf have significant difficulties in understanding and expressing themselves in written Spanish, and around 47% of the Deaf, older than 10, do not have basic level studies (information from INE, Spanish Statistics Institute and MEC, Ministry of Education). The main problems are related to verb conjugations, gender/number concordances and abstract concepts. Because of this, there are important communication barriers between a deaf person and, for example, a government employee who is providing a service personally. These barriers can cause deaf people to have fewer opportunities or rights. This happens, for example, when people want to renew their Driver's License (DL). A few government employees do not know LSE, so a deaf person needs an interpreter for accessing to this service.

Thanks to organisms like the Fundación CNSE, LSE is becoming not only the natural language for the Deaf to communicate, but also a powerful instrument when communicating to hearing people, or accessing information.

2. State of the Art

Several groups have generated corpora for sign language research. Some examples are: the RWTH-BOSTON-400 Database that contains 843 sentences with about 400 different signs from 5 speakers in American Sign Language with English annotations [1], a corpus composed of more than 300 hours from 100 speakers in Australian Sign Language [2], a corpus developed at Institute for Language and Speech Processing (ILSP) in Greek that contains parts of free signing narration, as well as a considerable amount of grouped signed phrases and sentence level utterances [3], and the British Sign Language Corpus Project that tries to create a machine-readable digital corpus of spontaneous and elicited British Sign Language (BSL) collected from deaf native signers and early learners across the United Kingdom [4].

In recent years, several groups have developed prototypes for translating Spoken language into Sign Languages using different strategies: example-based [5], rule-based [6], full sentence [7] or statistical approaches ([8];[9]; SiSi system).

About speech generation from sign language, in the Computer Science department of the RWTH, Aachen University, P. Dreuw supervised by H. Ney is making a significant effort into recognizing continuous sign language from video processing ([10][11]). The results obtained are very promising.

This paper describes the development and field evaluation of an Advanced Speech Communication System for Deaf People in a real domain: the Driver's License renewal.

3. Database

In order to develop systems involving language technologies, database collection is an important aspect to keep in mind. The database developed in this project has been obtained with the collaboration of the Local Traffic Office in the city of Toledo. It is composed of the most frequent explanations (from government employees) and the most frequent questions (from the user) that were taken down over a period of three weeks.

This local traffic office is organised as several windows (assistance positions) where more than 4000 sentences were annotated and analysed. This analysis showed that including the information from all windows, the semantic and linguistic domain was very wide and the vocabulary very large. In order to define the specific domain for developing the system, the service of renewing the Driver's License was selected.

Finally, 707 sentences were selected: 547 pronounced by government employees and 160 by users. This corpus was increased to 2,124 by incorporating different variants for Spanish sentences (maintaining the LSE translation). The sentences were translated into LSE, both in glosses (capitalised words with a semantic relationship to sign language) and in video, and compiled in an excel file.

The main features of the sentences pronounced by government employees and users are summarised in Table 1.

Government employee	Spanish	LSE
Sentence pairs	1,641	
Different sentences	1,413	199
Running words	17,113	12,741
Vocabulary	527	237
User	**Spanish**	**LSE**
Sentence pairs	483	
Different sentences	389	93
Running words	3,130	2,283
Vocabulary	294	133

Table 1: *Main statistics of the corpus*

For sign representation, a database with more than 400 signs was generated including sign descriptions in glosses, SEA (Sistema de Escritura Alfabética)[12], HamNoSys[13], and SIGML[14].

4. Spanish into LSE translation

The Spanish into LSE translation system converts natural speech sentences into LSE sentences signed by an avatar. This system is made up of three modules (Figure 1). The first one is a speech recognition module that converts natural speech into a sequence of words (text). The second one is a natural language translation module that converts a word sequence into a sign sequence. And the last module represents the signs with VGuido (avatar developed in the eSIGN project [15]).

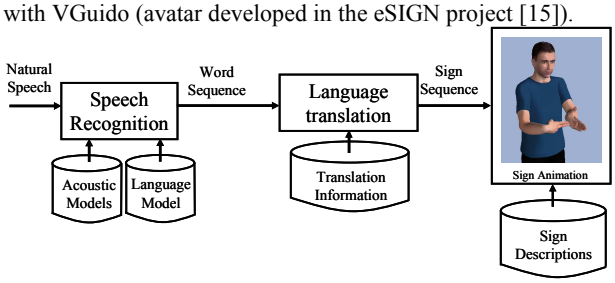

Figure 1: *Spanish into LSE translation module*

The speech recogniser is a HMMs-based continuous speech recognition system, speaker independent, with confidence measures and the possibility to adapt the acoustic models.

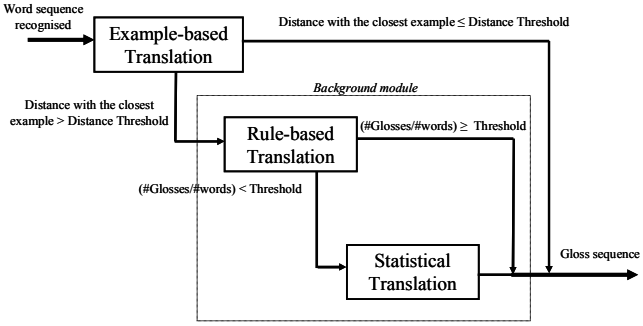

Figure 2: *Diagram of the language translation module combining three different translation strategies*

The translation module has a hierarchical structure divided into two main steps (Figure 2). In the first step, an example-based strategy is used to translate the word sequence: the translation process is carried out based on the similarity between the sentence to be translated and the items of a parallel corpus with translated examples. If the distance with the closest example is lower than a threshold, the translation output is the same as the example. But if the distance is higher, a background module translates the word sequence. For the background module, a combination of rule-based (where a set of translation rules, defined by an expert, guides the translation process) and statistical translators (phrase-based translator and Finite State Transducers) has been used. The first idea was to consider only the rule-based system as the background module but the statistical approaches were also incorporated as a good alternative during system development. The main idea is that the time and effort required to develop a statistical translator (it was possible to obtain a tuned version in one or two days) is considerable lower than a rule-based one (it took several weeks to develop all rules). During rule development, a statistical translator was incorporated in order to have a background module with a reasonable performance. The relationship between these two modules has been implemented based on the ratio between the number of glosses (generated after the translations process) and the number of words in the input sequence. If the #glosses/#words ratio is higher than a threshold, the output is the gloss sequence proposed by the rule-based module. Otherwise, if this condition is false, the statistical approach is carried out. All these thresholds were tuned using a development set. Table 2 summarizes the results (in laboratory tests): SR-WER (Speech Recognition Word Error Rate), SER (Sign Error Rate), PER (Position Independent SER) and BLEU (BiLingual Evaluation Understudy).

SR-WER	SER	PER	BLEU
6.76	10.11	8.45	0.8019

Table 2: *Result summary for laboratory tests*

The Figure 3 shows the visual interface of the module for translating spoken Spanish into LSE.

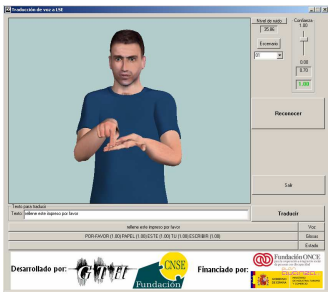

Figure 3: *Visual interface of the Spanish into LSE translation module*

5. Spanish generation from LSE

The spoken Spanish generation system converts a sign sequence (LSE sequence) into spoken Spanish. It is composed of three modules [16] (Figure 4).

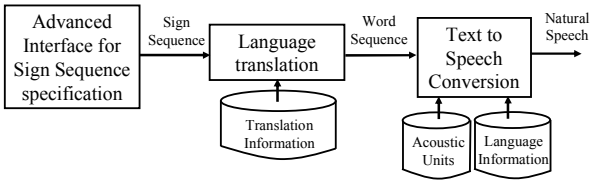

Figure 4: *Diagram of Spanish generation system*

The first module is a visual interface for specifying the sign sequence. This interface includes several tools for sign specification: avatar for sign representation (to verify that the sign corresponds to the gloss), prediction mechanisms, calendar and clock for date or time definitions, spelling, frequent questions, etc. With this visual interface the Deaf can build a sign sentence that will be translated into Spanish and spoken to a hearing person. The sign sequence is specified in glosses but signs can be searched by using specific sign characteristics in HamNoSys notation [13] (Figure 5).

The second module converts a sign sequence into a word sequence with three different strategies combined: an example-based, a rule-based and a statistical translation strategy. The procedure is the same as in the Spanish into LSE translation system. The last module converts the word sequence into spoken Spanish by using a commercial Text to Speech converter. In this project, the Loquendo system has been used (http://www.loquendo.com/en/).

Table 3 summarizes the translation results (in laboratory tests) for rule-based and statistical approaches: WER (Word Error Rate after translation), PER (Position Independent WER) and BLEU (BiLingual Evaluation Understudy).

WER	PER	BLEU
2.36	2.25	0.9113

Table 3: *Result summary for rule-based and statistical approaches in LSE-speech system*

Figure 5: *Visual interface for sign sequence specification*

6. Evaluation and Discussion

This advanced communication system has been evaluated in the Toledo Traffic Office involving a government employee and five deaf people: the speech-LSE system translates the government employee's explanations while the spoken Spanish generator helps deaf users to ask questions.

The Driver's Licence (DL) renewing process at the Toledo Traffic Office consists of three steps: form obtainment, payment, and handing over of the documents. The system was used for all the steps. The evaluation was carried during one day. At the beginning, a one-hour talk, about the project and the evaluation, was given to the government employee and users involved in the evaluation. All the users evaluated the system the same day. In the evaluation, six different scenarios were defined in order to specify real situations: in one scenario, the user simulated having all the necessary documents, three other scenarios in which the user simulated not having one of the documents: Identity Document, photo or the medical certificate, one scenario where the user had to fill some information in the application form, and finally, a scenario in which the user wanted to pay with credit card (only cash is accepted according to existing paying policy).

The five deaf users (three males and two females) interacted with the government employee at the Toledo Traffic Office using the developed system. The user ages ranged between 22 and 55 years old with an average age of 39.7 years. Two of the users said that they worked with a computer every day and three of them used a computer a few times per week. Only two of them had a medium-high understanding level of written Spanish and a good habit of using glosses for sign sequence specification.

Users and government employee tested the system in almost all the scenarios described previously and 25 dialogues were taken down. The evaluation results include objective measurements (Table 4 and Table 5) from the system and subjective information from both user and government employee questionnaires.

MEASUREMENT	VALUE
Speech Recognition Word Error Rate	4.6%
Sign Error Rate (after translation)	8.5%
Average Recognition Time	3.2 sec
Average Translation Time	0.0012 sec
Average Signing Time	4.6 sec
% of cases using example-based translation	95.0%
% of cases using rule-based translation	4.1%
% of cases using statistical translation	0.9%
# of government employee turns per dialogue	8.4

Table 4: *Objective measurements for evaluating the Spanish into LSE translation system*

The WER for the speech recognizer (4.6%) and the SER after translation (8.5%) are quite good. They are better than in laboratory tests. This was possible because for the field evaluation, the acoustic models of the speech recognizer were adapted to the government employee using 50 spoken utterances.

The time needed for translating speech into LSE is around 4,6 seconds (the speech recogniser works in real-time and the translation process is very fast), allowing a fluent dialogue. On the other hand, the example-based translation has been used in more than 94% of the cases showing the reliability of the linguistic study carried out.

MEASUREMENT	VALUE
Translation error rate	2.56%
Average translation time	0,001 sec
Average time for text to speech conversion	1.7 sec
% of cases using example-based translation	92.7%
% of cases using rule-based translation	7.3%
% of cases using statistical translation	0.0%
Time for gloss sequence specification.	16.5 sec
# of clicks for gloss specification	8.5 clicks
# of glosses per turn	2.6
# of user turn per dialogue	4.0

Table 5: *Objective measurements for evaluating the Spanish generator from sign-writing*

As the Table 5 shows, the good translation error rate (2.53%) and the short translation time make possible to use this system in real conditions. Also, the example-based strategy has been selected in most of the cases. This behaviour shows the reliability of the corpus collection.

The subjective measurements were collected from questionnaires filled in by both the government employee and deaf users. They evaluated different aspects of the system giving them a score of between 0 and 5.

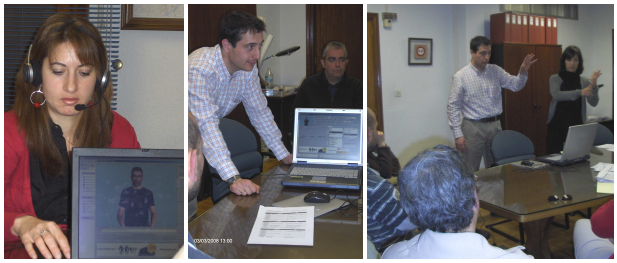

Figure 6: *Evaluation at the Toledo Traffic Office*

About the speech-LSE system, the evaluation from the government employee is quite positive giving a 3.5 score for all aspects considered (including speech recognition). The main problem reported was that it was very uncomfortable to have the screen of the Tablet PC turned to the user. For the future, two screens will be considered. The user assessment was very low (an overall score of 2.6). The worst score was to the naturalness of the sign (1.6). In order to reduce this problem, it is necessary to continue working on the standardization process of the LSE in Spain and integrating new strategies in the developed communication system.

About the LSE-speech system, the government employee has assessed both speech intelligibility and naturalness well, giving a 4.0 score. Users gave a reasonable score to the visual interface (3.2) but they also reported some problems related to the interface complexity. One more time, a significant problem was that gloss notation is not yet standardized enough for LSE (one sign can be represented by two glosses). In order to solve this problem, some users suggested incorporating a sign selection mechanism based on images or gifs.

7. Conclusions

This paper has described the development of an advanced speech communication system for helping deaf people when they want to renew their Driver's License. This paper also includes a field evaluation and a discussion on the main problems that must be solved in order to improve the system.

First, the paper has presented a Spanish into Spanish Sign Language (LSE: Lengua de Signos Española) translation system made up of a speech recognizer, a natural language translator, and a 3D avatar animation module. Secondly, the paper has described a spoken Spanish generator from sign-writing of Spanish Sign Language. This system consists of a visual interface where a deaf person can specify a sequence of signs, a language translator, and a text to speech converter.

For natural language translation, three technological proposals have been evaluated and combined in a hierarchical structure: an example-based strategy, a rule-based translation method and a statistical translator.

The speech-LSE translation system performed very well in speech recognition (4.6% word error rate) and language translation (8.5% sign error rate), but the users did not assess the system with a very good score in the questionnaires. It is necessary to improve the naturalness of the avatar and to make a greater effort for increasing the level of standardization of the LSE. The discrepancies in sign representation or sign sentence grammar are perceived as wrong behaviours.

Finally, the LSE-speech system performed very well in language translation (2.53% Translation Error Rate). The users gave a reasonable positive score but some problems related to the time for practising (with the visual interface) and with the level of gloss standardization were reported. The user needed less than 20 seconds to specify a gloss sequence using the interface. This time is low considering that the user only had a few minutes to practice. This time is higher compared to the time needed by an automatic sign recognition system (3-5 seconds) but the performance is considerably better.

8. Acknowledgements

The authors want to thank the eSIGN (Essential Sign Language Information on Government Networks) consortium for permitting the use of the eSIGN Editor and the 3D avatar in this research work. This work has been supported by Plan Avanza Exp Nº: PAV-070000-2007-567, ROBONAUTA (MEC ref: DPI2007-66846-c02-02) and SD-TEAM (MEC ref: TIN2008-06856-C05-03) projects.

9. References

[1] Dreuw P., Neidle C., Athitsos V., Sclaroff S., and Ney H. 2008a. "Benchmark Databases for Video-Based Automatic Sign Language Recognition". LREC, Marrakech, Morocco.

[2] Johnston T., 2008. "Corpus linguistics and signed languages: no lemmata, no corpus". 3rd Workshop on the Representation and Processing of Sign Languages, June 1. 2008.

[3] Efthimiou E., and Fotinea, E., 2008 "GSLC: Creation and Annotation of a Greek Sign Language Corpus for HCI" LREC 2008.

[4] Schembri. A., 2008 "British Sign Language Corpus Project: Open Access Archives and the Observer's Paradox". Deafness Cognition and Language Research Centre, University College London. LREC 2008.

[5] Morrissey S., and Way A., 2005. "An example-based approach to translating sign language". In Workshop Example-Based Machine Translation (MT X–05), pages109–116, Phuket, Thailand, September.

[6] San-Segundo R., Barra R., Córdoba R., D'Haro L.F., Fernández F., Ferreiros J., Lucas J.M., Macías-Guarasa J., Montero J.M., Pardo J.M, 2008. "Speech to Sign Language translation system for Spanish". Speech Communication, Vol 50. 1009-1020.

[7] Cox, S.J., Lincoln M., Tryggvason J., Nakisa M., Wells M., Mand Tutt, and Abbott, S., 2002 "TESSA, a system to aid communication with deaf people". In ASSETS 2002, pages 205-212, Edinburgh, Scotland, 2002

[8] Bungeroth J., Ney, H.,: Statistical Sign Language Translation. In Workshop on Representation and Processing of Sign Languages, LREC 2004, 105-108.

[9] Morrissey S., Way A., Stein D., Bungeroth J., and Ney H., 2007 "Towards a Hybrid Data-Driven MT System for Sign Languages. Machine Translation Summit (MT Summit)", pages 329-335, Copenhagen, Denmark, September 2007.

[10] Dreuw, P., D. Stein, T. Deselaers, D. Rybach, M. Zahedi, J. Bungeroth, and H. Ney. 2008b "Spoken Language Processing Techniques for Sign Language Recognition and Translation. Journal Technology and Dissability. Volume 20 Pages 121-133.

[11] Dreuw, P., Stein D., and Ney H. 2009. "Enhancing a Sign Language Translation System with Vision-Based Features". LNAI, number 5085, pages 108-113, Lisbon, Portugal, 2009.

[12] A. Herrero. "SEA: Sistema de Escritura Alfabética". Universidad de Alicante. 2004.

[13] Prillwitz, S., R. Leven, H. Zienert, T. Hanke, J. Henning, et-al. 1989. "Hamburg Notation System for Sign Languages – An introductory Guide". International Studies on Sign Language and the Communication of the Deaf, Vol. 5. Uni. of Hamburg.

[14] Zwiterslood, I., Verlinden, M., Ros, J., van der Schoot, S., 2004. "Synthetic Signing for the Deaf: eSIGN". Workshop on Assistive Technologies for Vision and Hearing Impairment, CVHI 2004, 29 June-2 July 2004, Granada, Spain. Spoken R.

[15] Hanke, T., Popescu, H., "Intelligent Sign Editor". eSIGN project deliverable. D2.3. 2003

[16] San-Segundo, R., Pardo, JM., Ferreiros, J. Sama, V. Barra-Chicote R, Lucas, JM, Sánchez, D. García, A. Spanish Generation from Sign Language Interacting with Computers, Vol. 22, No 2, pp. 123-139, 2009.