Methodology for developing an advanced communications system for the Deaf in a new domain

V. López-Ludeña , C. González-Morcillo , J.C. López , E. Ferreiro , J. Ferreiros , R. San-Segundo

A B S T R A C T

A methodology for developing an advanced communications system for the Deaf in a new domain is presented in this paper. This methodology is a user-centred design approach consisting of four main steps: requirement analysis, parallel corpus generation, technology adaptation to the new domain, and finally, system evaluation. During the requirement analysis, both the user and technical requirements are evaluated and defined. For generating the parallel corpus, it is necessary to collect Spanish sentences in the new domain and translate them into LSE (Lengua de Signos Española: Spanish Sign Language). LSE is represented by glosses and using video recordings. This corpus is used for training the two main modules of the advanced communications system to the new domain: the spoken Spanish into the LSE translation module and the Spanish generation from the LSE module. The main aspects to be generated are the vocabularies for both languages (Spanish words and signs), and the knowledge for translating in both directions. Finally, the field evaluation is carried out with deaf people using the advanced communications system to interact with hearing people in several scenarios. In this evaluation, the paper proposes several objective and subjective measurements for evaluating the performance. In this paper, the new considered domain is about dialogues in a hotel reception. Using this methodology, the system was developed in several months, obtaining very good performance: good translation rates (10% Sign Error Rate) with small processing times, allowing face-to-face dialogues.

1. Introduction

There are over 70 million people with hearing impairments in the world. Many of them have either been deaf from birth or have become deaf before learning a spoken language. This fact has serious implications for the education and social inclusion of Deaf people. They are one of the groups of people with the highest level of isolation, suffering substantial exclusion from social networks for the hearing. The main reasons for this exclusion are communications problems: people with hearing impairments cannot access audio content and many Deaf people have limited skills in reading, understanding and writing the dominant languages of the countries in which they live. Deaf teenagers leave school with an average reading age of a 10 year-old [35]. To be deaf means to not being able to hear or comprehend speech and language through the ear. Communication for a person who cannot hear is visual, not auditory. To deny sign language to Deaf people is tantamount to denying them their basic human rights to communication and education, with the resulting potentially severe isolation. For example, figures from the National Deaf Children's Society (NDCS), Cymru, reveal for the first time a shocking attainment gap between deaf and hearing pupils in Wales. In 2008, deaf pupils were 30% less likely than hearing pupils to gain five A*–C grades at General Certificate of Secondary Education (GCSE) level, while at key stage 3 only 42% of deaf pupils achieved the core subject indicators, compared to 71% of their hearing counterparts. Another example is a study carried out in Ireland in 2006, of 330 respondents "38% said they did not feel confidence to read a newspaper and more than half were not fully confident in writing a letter or filling out a form" [3].

In Spain, based on information from INE (Spanish Institute of Statistics) and the MEC (Ministry of Education), around 47% of the Deaf, of more than 10 years old, do not have basic level studies or are illiterate. In real conditions, 92% of the Deaf have significant difficulties in understanding and expressing themselves in written Spanish. The main problems are related to verb conjugations, gender/number concordances and abstract concept explanations. Because of this, only between 1% and 3% of the Deaf have a university level education. This percentage is very low compared to all the population in Spain.

One important cause of frustration for Deaf people is the lack of interpreters. This lack imposes a serious handicap on the involvement of deaf individuals in the wider society. Deaf people cannot access face-to-face services when or where they need them. Developing advanced ICT technologies can contribute to mitigating this deficiency, helping Deaf people to access personal services by allowing natural dialogues between hearing and deaf people.

When developing human–computer interaction systems, it is very important to meet a set of requirements in order to guarantee their usability and user acceptation. In this process, a good methodology is very important for dealing with the main aspects that must be considered. This fact is more relevant when envolving users with any kind of disability. Based on the experience in previous projects, the authors propose a specific methodology for developing an advanced communications system for deaf people focusing on a specific domain. This advanced communications system permits real face to face interactions between hearing and deaf people, allowing a natural dialogue between them. This system is able to translate spoken Spanish into LSE (Lengua de Signos Española) and viceversa: generating speech from LSE.

This paper is organised as follows. Section 2 presents the state of the art. Section 3 describes an overview of the methodology. Sections 4–7 describe the main steps of the methodology: requirement analysis, parallel corpus generation, technology adaptation and field evaluation. Finally, Section 8 includes the main conclusions of this work.

2. State of the art

ViSiCAST and eSIGN [5] have been two of the most relevant projects in speech into sign language translation. The ViSiCAST project focused on producing communications tools allowing sign language communications. This project was structured into three main application-oriented work packages: the first focused on the technical issues in delivery in that specific application area, and two technology work packages, focusing on virtual signing, sign language representation, and sign language synthesis from conventional textual sources. A further evaluation work package was concerned with eliciting feedback from deaf people at various stages within the development of the system.

The eSIGN project aimed to provide sign language on websites. The different tasks of this project are: development of tools needed for creating signed content; improvement in the signed output the avatar; creating the first information sites on the Internet with animated sign language; content creation in all three partner countries; the further development of tools needed for creating signed content; further improvement in the signed output of the avatar and the user involvement and continued evaluation of their tools and the avatar's comprehensibility.

Another example of advanced communications systems for deaf people is the VANESSA (Voice Activated Network Enabled Speech to Sign Assistant) project [36]. This project was part of eSIGN which facilitates the communications between assistants and their deaf clients in UK Council Information Centres (CIC's) or similar environments.

Two recent main research projects that focus on sign language recognition are DICTA-SIGN [12,9] and SIGN-SPEAK [7,8]. DICTA-SIGN aims to develop the technologies necessary to make Web 2.0 interactions in sign language possible. In SIGN-SPEAK, the overall goal is to develop a new vision-based technology for recognizing and translating continuous sign language into text.

The advanced communications system proposed in this paper consists of two main modules: a speech into sign language translation system and a speech generator from sign language.

In recent years, several groups have shown interest in spoken language translation into sign languages, developing several prototypes: example-based [22], rule-based [30,17], grammar-based

Table 1
Spoken language into sign language translation systems.

Ref.	Translation technology	Sign language	Translation performance	Limitations	Our approach in comparison
[4]	Full sentence: the system only recognises a reduced number of pre-translated sentences	British Sign Language (BSL)	Not reported	• It only translates fixed sentences	• Higher flexibility in the sentences to be translated • Combination of different translation technologies
[2]	Phrase-based model	German Sign Language (DGS)	Sign error rate > 50%	• Very small database for the experiments • No field evaluation	• A larger database with Cross Validation test • Combination of different translation technologies • Field evaluation
[22]	Example-based	Irish Sign Language (ISL)	Sign error rate < 40%	• No field evaluation	• Combination of different translation technologies • Field evaluation
SiSi system	Phrase-based model	British Sign Language (BSL)	Not reported	• No field evaluation	• Combination of different translation technologies • Field evaluation
[23]	Example-based and Phrase-based	ISL and DGS	BLEU > 0.5	• No field evaluation	• Field evaluation
[30]	Rule-based translation	Spanish Sign Language (LSE)	BLEU > 0.5	• Very small database • A costly translation technology • No field evaluation	• A larger database with cross validation • Combination of different translation technologies • Field evaluation
[19]	Combination of several translation technologies: memory-based and phrase-based technologies	Spanish Sign Language (LSE)	BLEU > 0.7 Sign error rate < 10%	• Focused on a very specific and limited domain (renewing the Identity Card) • No field evaluation	• A wider semantic domain with several services (hotel reception) • Field evaluation
This paper	Combination of several translation technologies: memory-based and phrase-based technologies	Spanish Sign Language (LSE)	BLEU > 0.7 Sign error rate < 10%	• Focused on a specific domain	

[21], full sentence [4] or statistical [2]; SiSi system http://www-03.ibm.com/press/us/en/pressrelease/22316.wss; [23] approaches. For LSE, it is important to highlight the author's experience in developing speech into LSE translation systems in several domains [30,32,19]. Table 1 describes the main characteristics of the main speech into sign language translation systems, highlighting the contribution of this paper compared to these previous works.

In order to eliminate the communications barriers between deaf and hearing people, it is necessary not only to translate speech into sign language [32] but also to generate spoken language from sign language, giving rise to a fluent dialogue in both directions. A great deal of effort has been made in recognising sign language and translating it into spoken language by using a language translator and a TTS converter. The main efforts have focused on recognising signs from video processing [29]. The systems developed so far are very person or environment dependent [37], or they focus on the recognition of isolated signs [40,38] which can often be characterised just by the direction of their movement. In Lee and Tsai [39], the authors propose a system for recognizing static gestures in Taiwanese sign languages (TSL), using 3D data and neural networks trained to completion. In Karami et al. [13] a system for recognizing static gestures of alphabets in Persian sign language (PSL) using Wavelet transform and neural networks is presented. A system for the automatic translation of static gestures of alphabets and signs in American Sign Language is presented by using Hough transformation and neural networks trained to recognise signs in [24]. In the Computer Science department of the RWTH Aachen University, Dreuw is making a significant effort in recognizing continuous sign language from video processing [6].

Bearing this scenario in mind, the advanced communications system developed in this paper includes the LSESpeak system [20], a new application for helping Deaf people to generate spoken Spanish that includes a spoken Spanish generator from LSE.

3. Methodology overview

The methodology presented in this paper is an adaptation of the Participatory Design methodology: one of the most used User-Centred Design approaches that follows the ISO standard Human-centred design for interactive systems: ISO 9241-210, 2010. Participatory design (previously known as 'Cooperative Design') is a design approach in which all stakeholders (e.g. employees, partners, customers, citizens, and end-users) are involved actively in the design process. The main target is to guarantee that the final designed product meets their needs and it is usable.

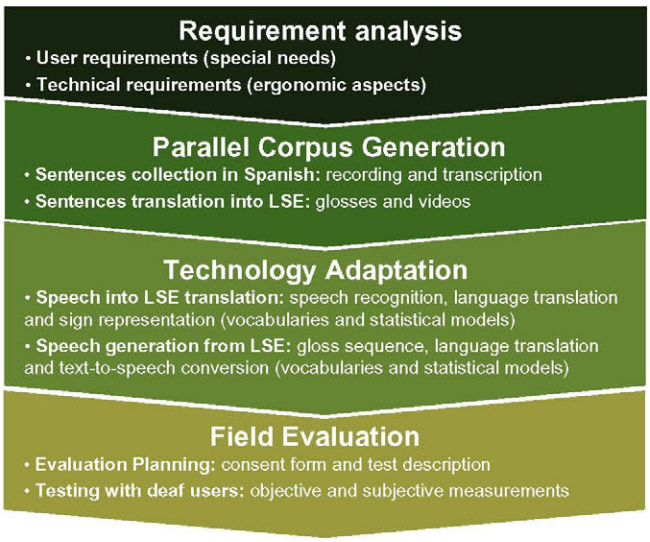

Fig. 1. Diagram of the proposed methodology.

This methodology consists of the following phases or steps (Fig. 1):

- The **requirement analysis** is undertaken with two Participatory Design workshops where end-users (deaf people), researchers and developers work together to define the technical and user requirements. In this step, two workshops were organised for defining user and technical requirements for the specific domain. It is very important at this stage to define and limit the domain of the natural language dialogues.
- The **parallel corpus generation** is carried out in several steps: sentence collection and sentence translation. These sentences must be representative of the specific domain. The translation process must be carried out by several LSE specialists in order to reach an agreement on the best translation.
- During **technology adaptation**, researchers must work together with the users in order to train new models for the specific domain. This training is carried out based on the parallel corpus obtained in the previous step.
- The **field evaluation** consists of an evaluation plan (including several scenarios in the specific domain) and the corresponding tests with deaf people using the advanced communications system. During the evaluation objective and subjective measurements must be obtained and analyzed. At this step, several measurements will be proposed.

These four steps will be described in detail in the following sections.

4. Requirement analysis

This section describes the first step in the methodology: requirement collection and analysis. For this analysis, it is necessary to define clearly the domain in which the advanced communications system will work. In this case, the new domain consists of the spoken dialogues between deaf customers and a receptionist at a hotel reception desk. In these dialogues, any aspect of the hotel may be addressed: check-in, check-out, breakfast, extra activities, etc.

4.1. User requirements

According to the Survey of Disability, Personal Autonomy and Dependency Situations (EDAD, 2008) from INE (Spanish Institute of Statistics), there are 1,064,100 deaf people in Spain. Deafness gives rise to significant communications problems: most deaf people have problems when expressing themselves in oral languages or understanding written texts. Their communications barriers have meant that 47% of deaf population have no studies or are even illiterate (INE – Spanish Institute of Statistics – 1999 y MEC – Science and Education Ministry – 2000/2001). These aspects support the need to generate new technologies in order to develop automatic translation systems for helping in the hearing to deaf people communications.

In order to obtain the user requirements, two Participatory Design workshops were organised including deaf customers, hotel receptionists and researchers from all the project partners.

- The first workshop was organised for data collection and brainstorming on the most frequent needs for deaf customers when they are in a hotel. This workshop was organised in a hotel and the hotel manager gave the team a guided tour around the hotel (Fig. 2). As a result of this workshop, an initial report was drawn up including:

- All the services offered by the hotel: accommodation, parking, restaurant, internet, etc.
- A typical structure and functioning of a hotel: check-in, check-out, scheduling, services, extra activities, accessibility, etc.
- Specific needs for deaf people: visual alarms or visual alarm clock service, etc.
- The second workshop was carried in a meeting room and focused on selecting the most important aspect in this domain (hotel reception). The initial report was analyzed and all the services and characteristics were sorted according to their relevance for deaf customers. After this sorting, the most important (relevance for deaf users) services were selected to be addressed by the automatic system (some services such as background music are irrelevant to them). The result of this meeting was a final report with the selected services and their relevance to deaf people.

This final report is very important for designing the following step: Parallel Corpus Collection. This corpus must include sentences referring to the main services selected in this report.

4.2. Technical requirements

An important challenge of the project is to achieve a minimum level of technical performance, because acceptance depends significantly on this quality. Based on previous experience [30,32], the technical researchers have defined the following technical requirements:

- The speech recognition system must provide a recognition rate of more than 90% in the selected application domain. If that rate is not reached with speaker-independent models, an adaptation process will be performed for each speaker involved in the evaluation in order to guarantee this rate.
- A translation error rate (Sign Error Rate: SER, see Section 6.1.2) of less than 10% is also necessary for the specific domain targeted in the project. These performance constraints are necessary to guarantee a dynamic hearing-deaf dialogue (without many repetition turns).
- Finally, the avatar intelligibility must be more than 90% when representing the signs: recognition rate of deaf people. In order to obtain this intelligibility, as will be shown in Section 6, the sign generation uses techniques based on inverse kinematics and semi-automatic movement capture that allows more realistic movements to be obtained. This approximation requires more time for vocabulary generation, but it is more realistic.

In order to guarantee these technical requirements, a Spanish-LSE parallel corpus with a significant number of sentences in the specific domain will be required. Based on previous experience, more than 500 sentences containing around 1000 Spanish words and 200 signs in LSE are necessary.

5. Parallel corpus generation

This section describes the process for generating the parallel corpus. First, it is necessary to record Spanish sentences from dialogues between customers and hotel receptionists. These dialogues must focus on the main services selected in the previously performed requirement analysis. Secondly, these sentences are translated into LSE (Lengua de Signos Española) in both, glosses and video files. Glosses are Spanish words in capital letters for referring to specific signs.

5.1. Spanish sentence collection in a new domain: hotel reservation

This collection has been obtained with the collaboration of the Hotel "Intur Palacio de San Martín". Over several weeks, the most frequent explanations (from the receptionist) and the most frequent questions (from customers) were compiled. In this period, more than 1000 sentences were noted and analysed.

Not all the sentences refer to the main services selected in the previous step, so the sentences had to be selected manually. This was possible because every sentence was tagged with the information on the service being provided when it was collected. Finally, 500 sentences were collected: 276 pronounced by receptionists and 224 by customers. This corpus was increased to 1677 by incorporating different variants for Spanish sentences (maintaining the meaning and the LSE translation).

5.2. Translating Spanish sentences into LSE (Lengua de Signos Española)

These sentences were translated into LSE, both in text (sequence of glosses) and in video, and compiled in an Excel file (Fig. 3).

The Excel file contains eight different information fields: "INDEX" (sentence index), "DOMAIN" (Hotel reception in this case), "SCENARIO" (scenario: where the sentence was collected), "SERVICE" (service provided when the sentence was collected), if the sentence was pronounced by the receptionist or the customer (AGENT), sentence in Spanish (SPANISH), sentence in LSE (sequence of glosses), and link to the video file with LSE representation.

5.3. Parallel corpus statistics

The main features of the corpus are summarised in Table 2. These features are divided into whether the sentence was spoken by the receptionist or the customer.

Fig. 2. Guided visit to the hotel.

INDEX	DOMAIN	NORMA	SERVICE	AGENT	SPANISH	LSE	VIDEO
1	HOTEL	DE ALOJAMIENTO	SALUDOS	Recepcionista	hola buenos días	HOLA BUENOS DÍAS	videos\1.wmv
2	HOTEL	DE ALOJAMIENTO	SALUDOS	Recepcionista	qué desea	QUERER QUÉ?	videos\2.wmv
3	HOTEL	DE ALOJAMIENTO	SALUDOS	Recepcionista	buenas tardes	BUENAS TARDES	videos\3.wmv
4	HOTEL	DE ALOJAMIENTO	SALUDOS	Recepcionista	hola buenas noches	HOLA BUENAS NOCHES	videos\4.wmv
5	HOTEL	DE ALOJAMIENTO	SALUDOS	Recepcionista	buenos días	BUENOS DÍAS	videos\5.wmv
6	HOTEL	DE ALOJAMIENTO	SALUDOS	Recepcionista	le puedo ayudar en algo	TU NECESITAR ALGO?	videos\6.wmv
7	HOTEL	DE ALOJAMIENTO	CHECK-IN	Huésped	necesito una habitación	YO UNA HABITACIÓN	videos\7.wmv
8	HOTEL	DE ALOJAMIENTO	CHECK-IN	Huésped	necesito una habitación doble	YO UNA HABITACIÓN DOBLE	videos\8.wmv
9	HOTEL	DE ALOJAMIENTO	CHECK-IN	Huésped	necesito una habitación doble con cu	YO UNA HABITACIÓN DOBLE	videos\9.wmv
10	HOTEL	DE ALOJAMIENTO	CHECK-IN	Recepcionista	tiene usted una reserva	TU RESERVA HAY?	videos\10.wmv
11	HOTEL	DE ALOJAMIENTO	CHECK-IN	Huésped	aquí tengo mi reserva	YO PAPEL RESERVA HAY	videos\11.wmv
12	HOTEL	DE ALOJAMIENTO	CHECK-IN	Huésped	la reserva está a nombre de rubén	RESERVA NOMBRE RUBEN	videos\12.wmv
13	HOTEL	DE ALOJAMIENTO	CHECK-IN	Huésped	el número de mi reserva es este	YO NÚMERO RESERVA ESTE	videos\13.wmv
14	HOTEL	DE ALOJAMIENTO	CHECK-IN	Huésped	no tengo reserva	YO RESERVA HAY-NO	videos\14.wmv
15	HOTEL	DE ALOJAMIENTO	CHECK-IN	Huésped	me gustaría reservar ahora	AHORA RESERVAR YO QUER	videos\15.wmv
16	HOTEL	DE ALOJAMIENTO	CHECK-IN	Recepcionista	déjeme ver	A-VER	videos\16.wmv
17	HOTEL	DE ALOJAMIENTO	CHECK-IN	Recepcionista	un segundo	UN-MOMENTO	videos\17.wmv
18	HOTEL	DE ALOJAMIENTO	CHECK-IN	Recepcionista	por favor déjeme su deneí	POR-FAVOR TU DNI DAR-A_N	videos\18.wmv
19	HOTEL	DE ALOJAMIENTO	CHECK-IN	Recepcionista	por favor déjeme la reserva	POR-FAVOR TU PAPEL RESE	videos\19.wmv

Fig. 3. Example of the database

Table 2
Main statistics of the parallel corpus.

	Spanish	LSE
Receptionist		
Sentence pairs	937	
Different sentences	770	243
Running words	6475	3349
Vocabulary	772	389
Customer		
Sentence pairs	741	
Different sentences	594	200
Running words	4091	2394
Vocabulary	594	277

6. Technology adaptation

The Advanced Communications System is made up of two main modules. The first translates spoken Spanish into LSE (Lengua de Signos Española). This module is used to translate the receptionist's utterances. The second module generates spoken Spanish from LSE in order to convert LSE customer questions into spoken Spanish. The corpus presented in the previous section is used for training the models for these two modules. The receptionist's sentences are used for developing the speech into the LSE translation system, while the user questions are used for the Spanish Generation Module from LSE.

6.1. Speech into LSE translation

Fig. 4 shows the module diagram developed for translating spoken language into LSE:

- The first module, the automatic speech recogniser (ASR), converts natural speech into a sequence of words (text). It uses a vocabulary, a language model, and acoustic models for every allophone.
- The natural language translation module converts a word sequence into a sign sequence. For this module, the paper presents and combines two different strategies. The first consists of a memory-based translation strategy: the translation process is carried out based on the similarity between the sentence to be translated and some examples of a parallel corpus (examples and their corresponding translations) stored in the translation memory. The second is based on a statistical translation approach where parallel corpora are used for training language and translation models.
- In the final step, the sign animation is made by using a highly accurate representation of the movements (hands, arms and facial expressions) in a Sign list database and a Non-Linear Animation composition module, both needed to generate clear output. This representation is independent of the virtual character and the final representation phase.

6.1.1. Speech recognition

The speech recogniser used is an HMM (Hidden Markov Model)-based system able to recognise continuous speech: it recognises utterances made up of several continuously spoken words. It has been developed at the Speech Technology Group (GTH-UPM: http://lorien.die.upm.es). In order to adapt the speech recogniser to a new domain it is necessary to adapt the acoustic models, to train a new language model and to generate the vocabulary for this domain:

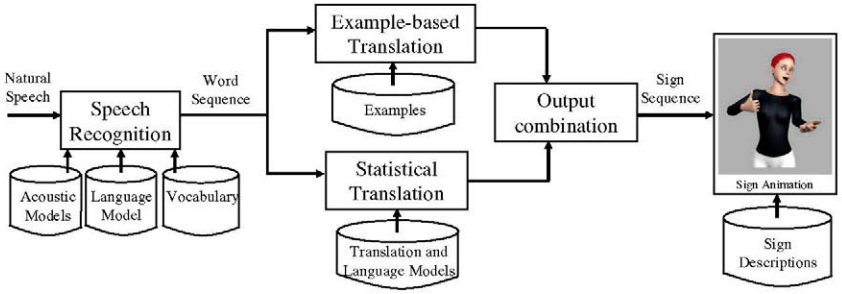

Fig. 4. Diagram of the speech into LSE translation system [19].

- The acoustic models have been adapted to the speaker using the Maximum a Posteriori (MAP) technique [10]. This adaptation process is very important for adapting the acoustic model to a specific speaker who will use the system intensively. This adaptation is highly desirable when a reduced number of speakers use the speech recogniser (as in the case of a hotel reception). As shown in Table 3, the WER (Word Error Rate) is reduced significantly (less than 3%) and the system speed is increased considerably (more than 50% in xRT: times Real Time) when adapting the acoustic models to the speaker. Table 3 includes experiments in laboratory tests.
- When generating the language model or the vocabulary in a new domain from a small corpus, the most important problem is the large number of Out of Vocabulary words (OOVs) and the poor estimation of the language model probabilities. In order to deal with this problem, several variants were included in the corpus, considering these aspects:
- In Spanish, an important strategy is to introduce variants for formal and informal ways of referring to "you" ("usted" or "tu" in Spanish). For example, given the informal form "tu debes darme el pasaporte" ("you must give me the passport"), the system would include "usted debe darme el pasaporte" (with the same translation in English "you must give the passport" and also in LSE).
- Including synonyms for some names, adjectives and verbs.
- Changing the order of expressions like "please" or "thank you": "¿Podrías decirme dónde está el restaurante?, por favor" -> "Por favor, ¿podrías decirme dónde está el restaurante?" ("Could tell me where the restaurant is, please?").
- The language model is based on classes. Instead of considering individual words for estimating the n-g sequence probabilities, the system trains probabilities of word and class sequences. Every class can contain several words. This utility is very interesting when numbers, hours, weekdays or months appear in the domain. With a small corpus, there are not enough sentences to include all possible numbers, hours, weekdays or months. Including these words in classes helps to train the language model better. All the words included in the classes were also added to the vocabulary in order to allow them to be recognised. In this domain, the authors have considered these categories: numbers, hours, weekdays, months, service places (restaurants, shops, public transportation, etc.) and tourist places (historic buildings, museums, attractions, etc.).
- The language model has been generated from scratch using the receptionist's part of the Spanish sentences from the corpus described in Section 5.

6.1.2. Language translation

The language translation module has a hierarchical structure divided into two main steps (Fig. 5). In the first step, a memory-based strategy is used to translate the word sequence in order to look for the best possible match. If the distance with the closest example is less than a certain threshold (Distance Threshold), the translation output is the same as the memory-based translation. But if the distance is greater, a background module based on a statistical strategy translates the word sequence.

The background module incorporates a pre-processing module [18] that permits its performance to be improved. When translating from Spanish into LSE, the number of words in the source and target languages is very different (on average, 6.5 words and 3.5 signs). This pre-processing module removes non-relevant words from the source language allowing a better alignment for training the statistical translation model. The statistical translation module is based on Moses, an open-source, phrase-based translation system released from NAACL Workshops on Statistical Machine Translation (http://www.statmt.org) in 2011.

In order to adapt the translation technology to a new domain, the translation and language models are trained from scratch considering the receptionist's part of the corpus for this domain (Section 5).

6.1.2.1. Memory-based translation strategy. A memory-based translation system uses a set of sentences in the source language and its corresponding translations in the target language, for translating other similar source-language sentences. In order to determine whether one example is equivalent (or at least, similar enough) to the sentence to be translated, the system computes a heuristic distance between them. By defining a threshold on this heuristic distance, the developer controls how similar the example must be to the sentence to be translated, in order to consider that they generate the same target sentence. If the distance is lower than a threshold, the translation output will be the same as the example translation. But if the distance is higher, the system cannot generate any output. Under these circumstances, it is necessary to consider other translation strategies.

The heuristic distance used in the first version of the system was a modification of the well-known Levenshtein distance (LD) [16]. The heuristic distance is the LD divided by the number of words in the sentence to be translated (this distance is represented as a percentage). One problem of this distance is that two synonyms are considered as different words (a substitution in the LD) while the translation output is the same. In recent work [32], the system has been modified to use an improved distance where the substitution cost (instead of being 1 for all cases) ranges from 0 to 1 depending on the translation behaviours of the two words. Additionally, the deletion cost ranges from 0 to 1 depending on the probability of not aligning a word to any sign (this word is associated to the NULL tag). These behaviours are obtained from the lexical model computed in the statistical translation strategy (described in the next section). For each word (in the source language), an N-dimension translation vector ($\bar{w}$) is obtained where the "i" component, $P_w(g_i)$, is the probability of translating the word "w" into the sign "s_i". N is the total number of signs (sign language) in the translation domain. The sum of all vector components must be 1. The substitution cost between words "w" and "u", and the deletion cost of word "w" are given by the following equations.

$$Subs.Cost(w,u) = \frac{1}{2}\sum_{i=1}^{N} abs(P_w(s_i) - P_u(s_i))$$
$$Del.Cost(w) = P_w(NULL) \quad (1)$$

Eq. (1): Substitution and deletion costs based on the translation behaviour.

When both words present the same behaviour (the same vectors), the probability difference tends towards 0. Otherwise, when there is no overlap between translation vectors, the sum of the

Table 3
Speech recognition error and processing time depending on the acoustic model adaptation.

Acoustic model adaptation using MAP	WER (%)	xRT
Without adaptation: speaker independent	7.3	0.73
Using 25 utterances for adapting the models to the speaker	5.2	0.52
Using 50 utterances for adapting the models to the speaker	3.1	0.36

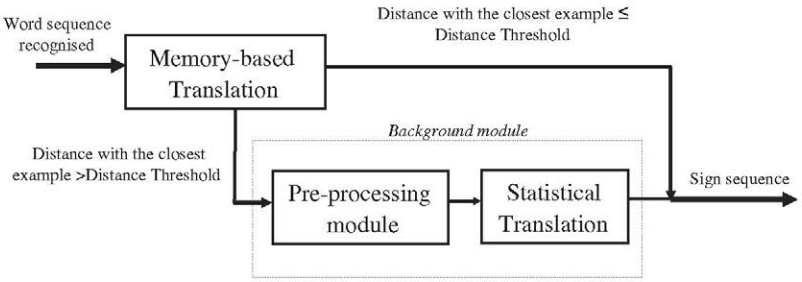

Fig. 5. Diagram of language translation module.

probability subtractions (in absolute values) tends towards 2. Because of this, the 1/2 factor has been included to make the distance range from 0 to 1. These costs are computed automatically so no manual intervention is necessary to adapt the memory-based translation module to a new semantic domain using only a parallel corpus.

The biggest problem with a memory-based translation system is that it needs large amounts of pre-translated text to make a reasonable translator. In order to make the examples more effective, it is necessary to generalise them [1], so that more than one string can match the same example, thus increasing its flexibility. Considering the following translation example for Spanish into LSE:

Spanish: "Habitación número cuarenta y cinco" (Room number forty five).
LSE: "HABITACIÓN CUARENTA_Y_CINCO".

Now, if it is known that "cuarenta y cinco" is a number, a generalised example is saved as

Spanish: "Habitación número $NUMBER".
LSE: "HABITACIÓN $NUMBER".

Where $NUMBER is a word class including all numbers. Notice that other strings also match this pattern. When indexing the example corpora, and before matching a new input against the database, the system tags the input by searching for words and phrases included in the class lists, and replacing each occurrence with the appropriate token. There is a file which simply lists all the members of a class in a group, along with the corresponding translation for each token. For the implemented system, six classes were considered: $NUMBER, $HOUR, $MONTH, $WEEK_DAY, $SERVICE_PLACES (banks, shops, restaurants, etc.) and $TOURIST_PLACES (museums, historic buildings, etc.).

This translation module generates one confidence value for the whole output sentence (sign sequence): a value between 0.0 (lowest confidence) and 1.0 (highest confidence). This confidence is computed as the average confidence of the recognised words (confidence values obtained from the speech recogniser) multiplied by the similarity between this word sequence and the example used for translation. This similarity is complementary to the heuristic distance: 1 minus heuristic distance. The confidence value will be used to decide whether the translation output (sign sequence) is good enough to be presented to a Deaf person. Otherwise, the translation output is rejected and not represented by the avatar. In this case, the receptionist must repeat the spoken sentence again.

6.1.2.2. Statistical translation strategy. The statistical translation module is composed of a pre-processing stage and a phrase-based translation system.

6.1.2.2.1. Pre-processing module. This pre-processing module replaces Spanish words with associated tags (López-Ludeña et al., 2011) using a word-tag list. In this module, all the words in the input sentence are replaced by their tags with the exception of those words that do not appear on the list (OOV words). They are kept as they are considered as proper names. After that, the "non-relevant" tags are removed from the input sentence (Non-relevant words are Spanish words not assigned to any sign). The word-tag list is generated automatically using the lexical model obtained from the word-sign GIZA++ alignments [26]. Given the lexical model, the tag associated to a given word is the sign with the highest probability of being the translation of this word. But this tag is assigned only if this probability is higher than a threshold otherwise it is kept as it is. If the most probable sign is "NULL" and its probability is higher than this threshold, this word will tagged with the "non-relevant" tag. This probability threshold is fixed to 0.4 based on development evaluations. For the words belonging to one of the six classes ($NUMBER, $HOUR, $MONTH, $WEEK_DAY, $SERVICE_PLACES, and $TOURIST_PLACES), the associated tag is the name of the class.

In conclusion, the pre-processing module allows the variability in the source language to be reduced together with the number of tokens that make up the input sentence. These two aspects give rise to a significant reduction in the number of source–target alignments the system has to train in the next step. When having a small corpus, as is the case in many sign languages, this reduction in alignment points permits training models to get better with fewer data.

6.1.2.2.2. Phrase-based translation module. The Phrase-based translation system is based on the software released at the 2011 EMNLP Workshop on Statistical Machine Translation (http://www.statmt.org/wmt11/) (Fig. 6).

The phrase model has been trained starting from a word alignment computed using GIZA++ [26]. GIZA++ is a statistical machine translation toolkit that is used to train IBM Models 1–5 and an HMM word alignment model. In this step, the alignments between words and signs in both directions (Spanish–LSE and LSE–Spanish) are calculated. The "alignment" parameter has been fixed at "target-source" as the best option (based on experiments on the development set): only this target–source alignment was considered (LSE–Spanish). In this configuration, alignment is guided by signs: this means that in every sentence pair alignment, each word is aligned to one or several signs (but not the opposite), and, there are also some words that were not aligned to any sign. When combining the alignment points from all sentence pairs in the training set, all possible alignments are considered: several words are aligned to several signs.

After the word alignment, the system performs a phrase extraction process [15] where all phrase pairs that are consistent with the word alignment (target–source alignment in our case) are collected. In the phrase extraction, the maximum phrase length has

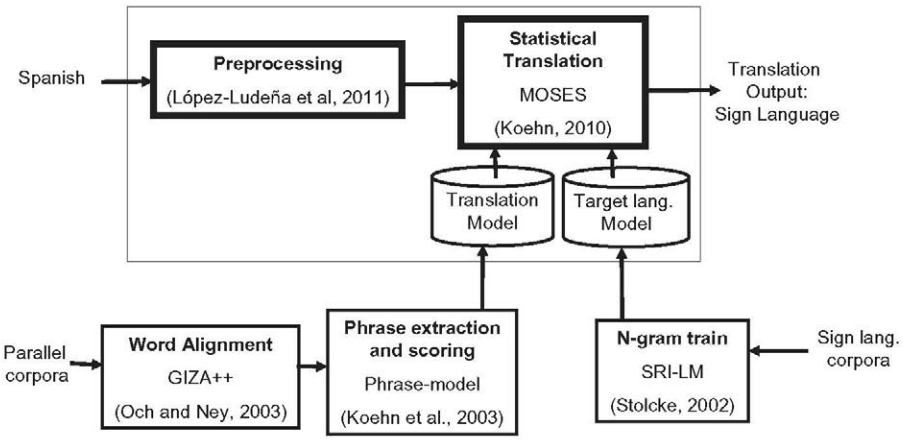

Fig. 6. Phrase-based translation architecture.

Table 4
Result summary for memory-based and statistical approaches.

	SER (%)	±Δ	PER (%)	BLEU	NIST
Memory-based strategy	34.2	1.65	32.4	0.5912	7.124
Memory-based approach (considering a heuristic distance < 30%)	4.2	0.72	3.8	0.9322	10.122
Statistical strategy including the pre-processing module	29.7	1.50	25.9	0.6667	8.132
Combining translation strategies	9.8	1.10	8.3	0.7522	9.222

been fixed at seven consecutive words, based on development experiments on the development set.

Finally, the last step is phrase scoring. In this step, the translation probabilities are computed for all phrase pairs. Both translation probabilities are calculated: forward and backward.

For the translation process, the Moses decoder has been used [14]. This program is a beam search decoder for phrase-based statistical machine translation models. In order to obtain a 3-g language model, the SRI language modeling toolkit has been used [33]. Both translation and language models have considered the six classes used for the memory-based translation module. Words in these classes are translated using a dictionary-based strategy. In this domain, every word in these classes has a unique translation.

6.1.2.3. Translation experiments. In order to evaluate the translation module, some experiments have been carried out using the receptionist's part of the Spanish–LSE parallel corpus described in

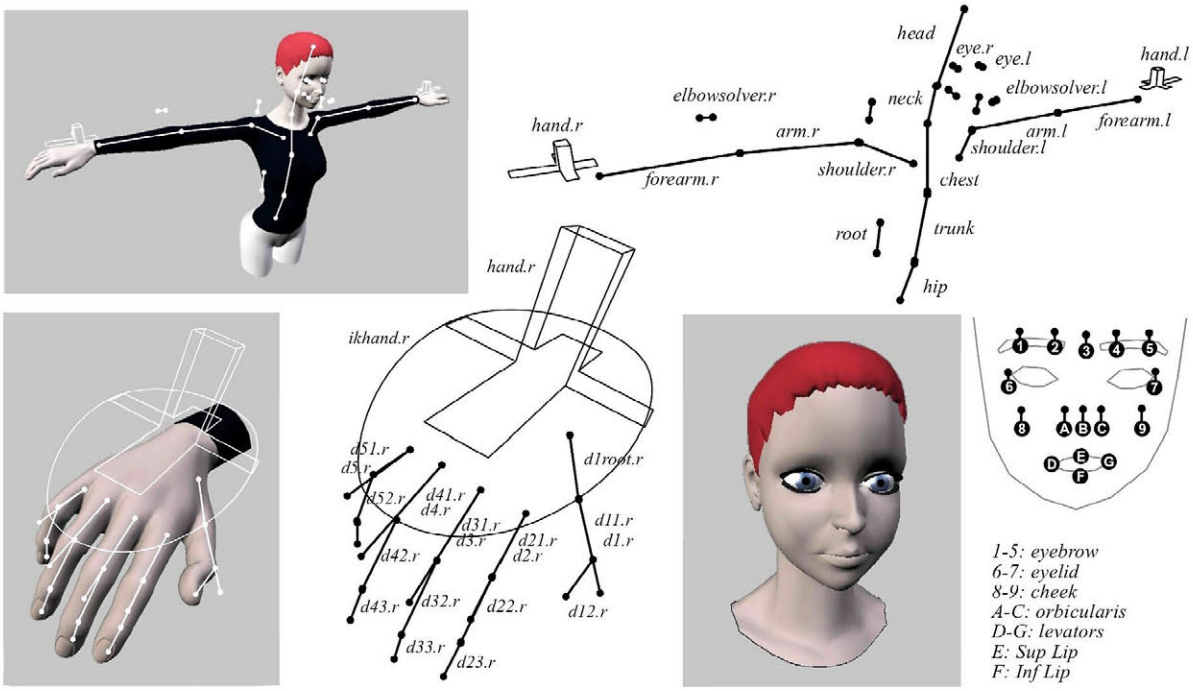

Fig. 7. Main bones and inverse kinematics controls (body, hand and face) of the avatar.

Table 2. The corpus was divided randomly into three sets: training (75% of the sentences), development (12.5% of the sentences) and test (12.5% of the sentences), carrying out a Cross-Validation process. The development set was used to tune the main parameters of the two translation strategies: heuristic distance threshold, GIZA++ alignment and maximum number of words in a phrase. Table 4 summarizes the results for memory-based and statistical approaches considering several performance metrics: SER (Sign Error Rate) is the percentage of wrong signs in the translation output compared to the reference in the same order. PER (Position Independent SER) is the percentage of wrong signs in the translation output compared to the reference without considering the order. Another metric is BLEU (BiLingual Evaluation Understudy; [27]), and finally, NIST (http://www.nist.gov/speech/tests/mt/). It is important to highlight that both SER and PER are error metrics (a lower value means a better result) while BLEU and NIST are accuracy metrics (a higher value means a better result).

For every SER result, the confidence interval (at 95%) is also presented. This interval is calculated using the following formula:

$$\pm\Delta = 1.96\sqrt{\frac{SER(100 - SER)}{n}} \tag{2}$$

Eq. (2): Confidence Interval at 95%.

n is the number of signs used in testing, in this case $n = 3349$. An improvement between two systems is statistically significant when there is no overlap between the confidence intervals of both systems. As is shown in Table 4, all improvements between different approaches are higher than the confidence intervals.

As is shown in Table 4, memory-based and statistical strategies have SER of more than 20%. Table 4 also presents the translation results for the memory-based approach for those sentences that have a heuristic distance (with the closest example) of less than 30% (the rest of the sentences were not translated). In this case, the results increase significantly: SER improvement is greater than the confidence intervals (at 95%). Finally, Table 4 presents the results for the combination of several translation strategies: memory-based (considering a heuristic distance <30%) and the statistical approach with the pre-processing module. As is shown, the hierarchical system obtains better results by translating all the test sentences: SER < 10%. Combining both translation strategies allows a good compromise to be made between performance and flexibility when the system is trained with a small parallel corpus.

For the field evaluation presented in Section 7, the memory-based and phrase-based translation modules were trained using the whole receptionist's part of the corpus.

6.1.3. Sign language representation

The Sign Language Representation module uses a declarative abstraction module used by all of the internal components. This module uses a description based on XML, where each key pose configuration is stored defining its position, rotation, length and hierarchical structure. We have used an approximation of the standard defined by H-Anim (Humanoid Working Group ISO/IEC FCD 19774:200x). In terms of the bones hierarchy, each animation chain is made up of several «joint» objects that define transformations from the root of the hierarchy.

Several general purpose avatars such as Greta [25] or SmartBody [34] have lacked a significant number of essential features for sign language synthesis. Hand configuration is an extremely important feature; the meaning of a sign is strongly related to the finger position and rotation. In our avatar each phalanx can be positioned and rotated using realistic human limitations. This is the most time-consuming phase in the generation of a new sign and, as detailed in the following section; a new approach to increase the adaptability has been created. For each sign it is necessary to model non-manual features (torso movements, facial expressions and gaze). For the upper body control, some high-level IK control has been defined (see Fig. 7).

The skeleton defined in the representation module is made up of 103 bones, out of which 19 are inverse kinematics handlers (they have an influence on a set of bones). The use of inverse kinematics and spherical quaternion interpolation [41] eases the work of the animators in capturing the key poses of signs from deaf experts. The geometry of the avatar is defined using Catmull–Clark adaptive subdivision surfaces. To ease the portability for real time rendering, each vertex has the same weight (each vertex has the same influence on the final deformation of the mesh).

Facial expression is used to indicate the sentence mode (assertion or question) and eyebrows are related to the information structure. In this way, this non-manual animation is used to highlight adjectival or adverbial information. The movements of the mouth are also highly important in focusing the visual attention to make comprehension easier. As pointed out by Pfau [28], non-manuals require more attention from the point of view of the automatic sign language synthesis.

Another advantage of the representation module is the adaptation to different kinds of devices (computers, mobile phones, etc). The rendering phase is often considered as a bottleneck in photo-realistic projects in which one image may need hours of rendering in a modern workstation. The rendering system used in this work can be easily used through distributed rendering approaches [11].

In order to adapt the representation module to a new domain, the main task is to create new signs for the new domain (those necessary to translate the receptionist's sentences). This task needs a lot of time. In order to reduce this time, the system includes a sign editor module to facilitate the construction of the sign vocabulary. In this application, the user chooses the basic configurations of shape and orientations of both the hands (active and passive). The expert chooses the frame and with one interaction picks the closest configuration of the hand. This configuration can be refined later using the aforementioned inversed kinematic facilities. These configurations of the shape and orientation are defined as static poses which contain only the essential parameters that describe the action. This information is stored in XML files. In the current system, 86 hand shapes (23 basic shapes and 63 derived from the basic configurations) were defined. 53 configurations for orientation were also constructed. Fig. 8 shows the first 30 configurations in the sign editor. Thanks to the use of this sign editor, the time required to specify a new sign decreased by 90% with similar quality results. Some examples can be downloaded from http://www.esi.uclm.es/www/cglez/ConSignos/signos/.

It is important to remember that each sign must be made only once and thanks to the design of the representation module, this description of the movement can be reused in different 3D avatars.

6.2. Speech generation from LSE

In order to convert a deaf customer's questions into spoken Spanish, the LSESpeak system was integrated [20]. LSESpeak (Fig. 9) is a new version of an LSE into Spanish translation system [31]. This tool is made up of three main modules. The first module is an advanced interface in which the Deaf customer specifies an LSE sequence. The second module is a language translator for converting LSE into written Spanish. This module has the same structure described in Section 6.1.2. Finally, the third module is an emotional text to speech converter based on Hidden Semi-Markov Models (HSMMs) in which the user can choose the voice gender (female or male), the emotion type (happy, sad, angry, surprise, and fear) and the Emotional Strength (ES) (on a 0–100% scale).

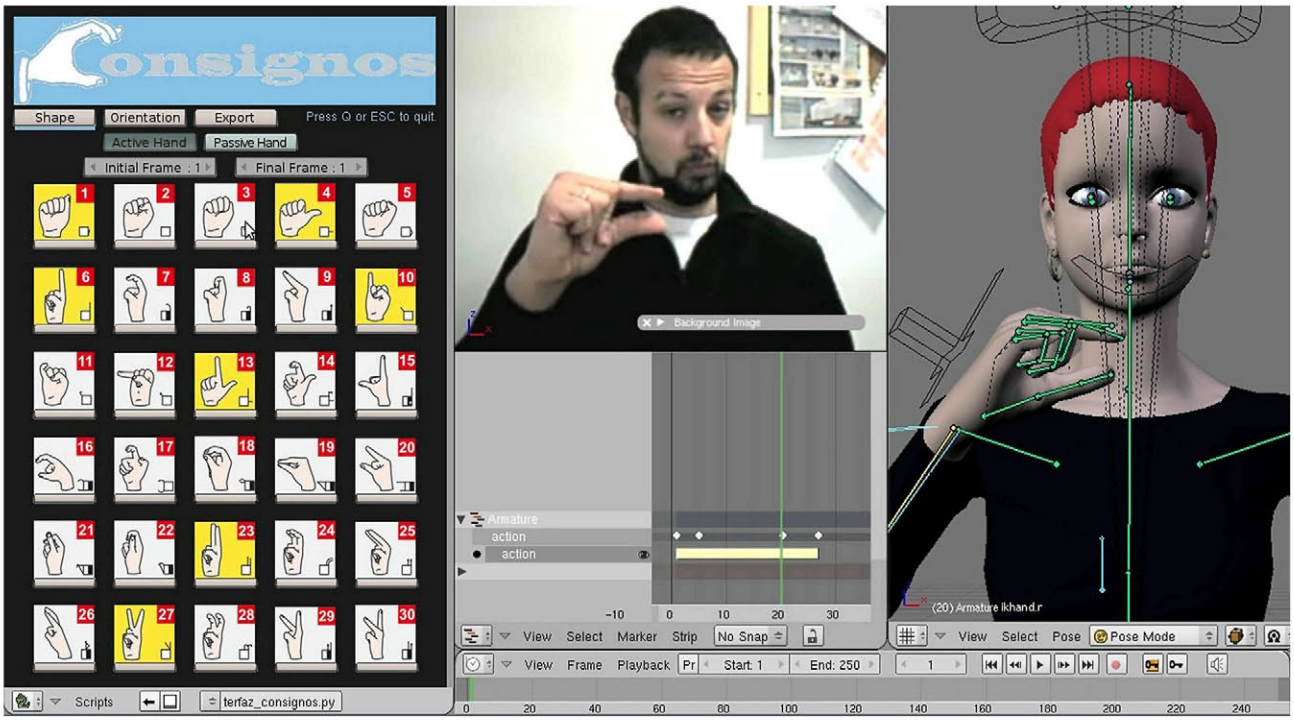

Fig. 8. Sign editor based on the use of predefined static poses for hand shape and orientations.

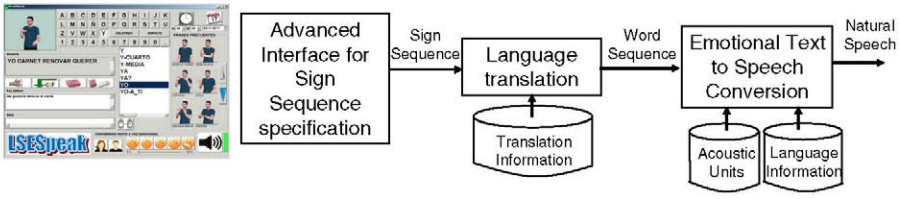

Fig. 9. Module diagram of LSES peak.

In order to adapt the advanced interface to a new domain, all the signs involved in this domain were generated, especially those signs included in the customer part of the parallel corpus (Section 5). A very useful utility incorporated into the interface allows the next signs given a previous sequence to be proposed. When there is a partial sign sequence already specified and the user moves the mouse over the SIGNOS windows, the system displays a popup menu proposing several candidates for the next sign. These candidates have been selected based on a sign language model trained from sign sequences (LSE sentences) of the customer part of the parallel corpus. These probabilities were retrained considering the corpus described in Section 5.

In order to adapt the translation technology to a new domain, the translation and language models are trained from scratch considering the customer's part of the corpus for this domain (Section 5).

For generating the spoken Spanish from text, a text to speech conversion system is general for any domain so it is not necessary to carry out any adaptation process.

7. Field evaluation

The final step of the methodology is the field evaluation. In this evaluation, the communications system must be tested with real users in a real scenario. This step is divided in two main tasks: evaluation design and field evaluation with real users.

7.1. Evaluation design

The evaluation design was carried out in a workshop where end-users (deaf people and receptionists), researchers and developers worked together to define the evaluation plan (Fig. 10) (see Fig. 11).

This plan includes the followings aspects:

- **The main characteristics of the users**: both deaf customers and receptionists. The main selected characteristics were related to their age, their place of residence in Spain, and their ability to use a computer, reading Spanish or dealing with LSE glosses. These characteristics influence the evaluation results: the hypothesis is that users with a higher frequency of using the computer, reading Spanish or LSE glosses will better accept the communications system.
- **Main resources necessary for the evaluation**: In this case, the main resources were two laptops for translating in both directions, a big screen for representing the signs, a microphone and speakers. Additionally, for recording the evaluation process, a video camera was also considered. As sign language is a visual language, video recording is very important for analysing some aspects of the scene. In order to carry out a field evaluation, a real scenario for testing is also necessary. In this case, the hotel reception at the Intur

Fig. 10. Workshop for defining the evaluation plan.

Palacio de San Martín Hotel in Madrid was considered. Finally, depending on the number of deaf customers and the duration of the evaluation, it is necessary to estimate the number of interpreters required during this period.

- **Consent form**: When recording users, it is necessary to design a consent form that the users must fill in and sign before starting the evaluation process. In this form, it is important to highlight that private information may be asked for and a number individual results will be published in some form. The information will be always provided as agglomerative numbers.
- **Scenarios to be simulated**: The receptionist and deaf customers must use the advanced communications system developed in real situations. These situations must be designed in accordance with the main requirements analysed during the first step of the methodology. In this case, five different scenarios were considered: the first two scenarios consisted of the checking-in processes both with and without a previous reservation. The third was the checking-out process. The fourth dealt with questions regarding several hotel services such as restaurant or gym. The last one was related to queries about tourist places close to the hotel. These scenarios were selected based on the most frequent needs described in the requirement analysis report.
- **Objective and subjective measurements**: Finally, it is necessary to specify the main measurements that will be annotated and reported in the evaluation report.

- Objective measurements. Researchers included all measurements related to time processing and accuracy of all the modules included in the communications system. Based on this, a log file was generated (by the system) including this information. To obtain the speech recognition and language translation performances, it was necessary to listen to the audio files recorded during the evaluation.
- Subjective measurements. Traditionally, these measurements are obtained by means of questionnaires filled in by the users. In these questionnaires, the users are asked about several aspects related to the system performance (for example, is the translation correct?) and the user must score them on a numerical scale [32]. A subjective evaluation of sign language involves two main aspects: intelligibility and naturalness. In a research project like this, the first target is intelligibility but, based on previous experience, when asking users to rank general questions, naturalness and intelligibility influence the response. In order to isolate the intelligibility, the questionnaires were redesigned to avoid this aspect: the deaf customers were asked specific questions (instead of general ones) about some dialogues (for example, where is the restaurant?). Three or four questions were considered per dialogue.

7.2. Evaluation results

The evaluation was carried out over one day. At the beginning, the assistance position was installed and a one-hour talk about the project and the evaluation process was given to receptionists and deaf customers involved in the evaluation. The speech recogniser was adapted to the receptionist involved in the evaluation. For this adaptation, 50 spoken sentences (1–2 s) were recorded (see results in Table 3).

The system was evaluated by four deaf customers (two female and two male) who interacted with one receptionist at the reception desk of the Intur Palacio San Martin Hotel. The deaf customer's ages ranged from between 26 and 47 years old with an average age of 36. All the customers said that they use a computer every day or every week, and only half of them had a medium-high understanding level of written Spanish.

Before using the developed system, the deaf customers looked at several signs (10 signs per customer) represented by the avatar and they were asked to identify them considering this specific domain. After that, they were asked to interact with the receptionist using the advanced communications system in the scenarios designed in the preparation of the evaluation plan. After the interactions, the deaf customers were asked several specific questions about the information provided by the receptionist. It is important to comment that for this field evaluation, new dialogues were considered, different from those presented in the laboratory evaluation (Section 6.1.2.3).

For evaluating the advanced communications system, it is necessary to evaluate every module for translating speech into LSE and vice versa. The evaluation of the speech into the LSE translation module includes objective measurements of the system and

Fig. 11. Different photos at the hotel during the evaluation.

Table 5
Objective measurements for evaluating the Spanish into LSE translation system

Agent	Measurement	Value
System	Word error rate (85 utterances)	6.7%
	Sign error rate (after translation)	10.7%
	Average recognition time	3.1 s
	Average translation time	0.002 s
	Average signing time	4.1 s
	% Of cases using memory-based translation	96.5%
	% Of cases using statistical translation	3.5%
	% Of turns translating from speech recognition	95.3%
	% Of turns translating from text	0.0%
	% Of turns translating from text for repetition	4.7%
	# Of receptionist turns per dialogue	7.7
	# Of dialogues	11

Table 6
Subjective measurements in the questionnaires.

	1st (%)	2nd (%)	3rd (%)
Human recognition rate depending on the number of attempts			
Isolated signs: 40 signs in total	87.5	97.5	100.0
Questions about the dialogues: 24 questions in total	62.5	87.5	100.0

Table 7
Objective measurements for evaluating the Spanish generator from LSE.

Agent	Measurement	Value
System	Translation rate (45 translations)	98.0%
	Average translation time	0.001 s
	Average time for text to speech conversion	2.1 s
	% Of cases using memory-based translation	99.0%
	% Of cases using statistical translation	1.0%
	Time for gloss sequence specification.	18.0 s
	# Of clicks for gloss sequence specification.	7.8 clicks
	# Of glosses per turn	2.1
	% Of utility use:	
	– List of glosses	60.0%
	– List of proposed next glosses	40.0%
	# Of turns using the most frequent sign sequences per dialogue	2.0
	# Of deaf customer turns per dialogue	4.1
	# Of dialogues	11

subjective information. A summary of the objective measurements obtained from the system are shown in Table 5.

The WER (Word Error Rate) for the speech recogniser is 6.7% being small enough to guarantee a low SER (Sign Error Rate) in the translation output: 10.7%. On the other hand, the time needed for translating speech into LSE (speech recognition + translation + sign representation) is around 7 s for an agile dialogue. This performance fits the technical requirements defined in Section 4.2.

As regards the questionnaires, Table 6 summarises the recognition accuracy based on the number of attempts for isolated signs and for questions in the dialogues.

For isolated signs, the recognition rate in the first attempt is very high (close to 90%, the technical requirement defined in Section 4.2) but for the dialogues, this percentage was worse, close to 60%. The main problems were related to the recognition of some signs: there were problems in the orientation of several signs and the discrepancy as to which sign to choose for presenting one concept. LSE (Lengua de Signos Española) is a very young language (it has been official since 2007) and there is a very high variability between different regions in Spain. These differences affect not only the signs but also the structure of the sign language sentences.

Finally, some objective measurements of the spoken Spanish generation module are included in Table 7.

As is shown, the good translation rate and the short translation time make it possible to use this system in real conditions. As regards the translation process, the memory-based strategy has been selected in most of the cases. This behaviour shows the reliability of the corpus collection including the most frequent user questions.

The user needed less than 20 s to specify a gloss sequence using the interface. This is not a long time considering that the deaf customer had only few minutes to practice with the visual interface before the evaluation. With more time for practicing, this period would be reduced.

In order to expand this analysis, Table 8 shows Spearman's correlation between some objective measurements from the Deaf customer evaluation and their background and age: computer experience, confidence with written Spanish, and age. This table also includes p-values for reporting the correlation significance. Because of the very low number of data and the unknown data distribution, Spearman's correlation has been used. This correlation produces a number between −1 (opposite behaviours) and 1 (similar behaviours). A 0 correlation means no relation between these two aspects.

As shown, only those results in bold are significant ($p < 0.05$): the questions answered the first time (Table 6) is negatively correlated with age. Although there are interesting tendencies in the other results, it is not possible to extract any significant conclusion due to the small amount of data.

8. Discussion and conclusions

The development methodology presented in this paper consists of four main steps:

- Requirement analysis: user and technical requirements are evaluated and defined.
- Parallel corpus generation: collection of Spanish sentences in the new domain and translation into LSE (Lengua de Signos Española: Spanish Sign Language). The LSE is represented by glosses and using video files.
- Technology adaptation to the new domain: the two main modules of the advanced communications system are adapted to the new domain using the parallel corpus: the spoken Spanish into LSE translation system and the Spanish generation from LSE module.
- System evaluation: the evaluation is carried out with deaf people using the advanced communications system in the specific scenario

Table 8
Analysis of correlations between Deaf customer evaluation and their background.

Evaluation measurement	Computer experience	Confidence with written Spanish	Age
Questions answered the first time (Table 6)	0.43 ($p = 0.120$)	0.22 ($p = 0.114$)	−0.61 ($p = 0.050$)
Time for gloss sequence specification in the LSESpeak system (Table 7)	−0.35 ($p = 0.123$)	−0.13 ($p = 0.214$)	0.52 ($p = 0.077$)
Percentage of times the receptionists had to repeat an utterance (Table 5)	−0.26 ($p = 0.245$)	−0.02 ($p = 0.422$)	0.41 ($p = 0.056$)

The main advantage of this methodology is that the users (hotel receptionist and deaf customers) are involved in the most important steps in this methodology: requirement analysis, parallel corpus generation and field evaluation. Another important advantage is that the technology adaptation is almost automatic from the parallel corpus (vocabularies, language models and translation models). The exception is the generation of the sign language vocabulary (signs). The signs must be modelled using the sign editor. With this methodology, it has been possible to develop the system in several months, obtaining very good performance: good translation rates (around 90%) with small processing times, allowing online dialogues in a face to face conversation.

On the other hand, the main disadvantage is that the methodology is sequential and the technology adaptation depends on the parallel corpus generation. This generation is a bottleneck: if there is a delay in generating the corpus, the entire process is delayed. One way of alleviating this problem is to start sign modelling by hand (the most demanding task), as soon as there are several translated sentences with some signs to model. The vocabularies, language and translation models must wait until the end of corpus generation process but, as the generation process is automatic, these models and vocabularies are available in several hours.

Acknowledgements

This work has been supported by Plan Avanza Exp No: TSI-020100-2010-489 and the European FEDER fund. The authors want to thank discussions and suggestions from the colleagues at INDRA I + D Tecnologías Accesibles, Ambiser Innovaciones S.L., ICTE (Instituto para la Calidad Turística Española), EMT (Empresa Municipal de Transportes) and Fundacion CNSE. The authors also want to thank Mark Hallett for the English revision.

References

[1] R.D. Brown, Automated generalization of translation examples, in: Proceedings of the Eighteenth International Conference on Computational Linguistics (COLING-2000), Saarbrücken, Germany, August 2000, pp. 125–131.
[2] J. Bungeroth, H. Ney, Statistical sign language translation, in: Workshop on Representation and Processing of Sign Languages, LREC 2004, pp. 105–108.
[3] P. Conroy, Signing in and Signing Out: The Education and Employment Experiences of Deaf Adults in Ireland, Research Report, Irish Deaf Society, Dublin, 2006.
[4] S.J. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, Tutt Mand, S. Abbott, TESSA, a system to aid communication with deaf people, in: ASSETS 2002, Edinburgh, Scotland, 2002, pp. 205–212.
[5] R. Elliott, J.R.W. Glauert, J.R. Kennaway, I. Marshall, E. Sáfár, Linguistic modelling and language-processing technologies for avatar-based sign language presentation, in: E. Efthimiou, S.-E. Fotinea, J. Glauert (Eds.), Special Issue on Emerging Technologies for Deaf Accessibility in the Information Society, Journal of Universal Access in the Information Society, vol. 6(4), Springer, February 2008, pp. 375–391.
[6] P. Dreuw, D. Stein, H. Ney, Enhancing a sign language translation system with vision-based features, in: Proceedings of Special Issue Gesture Workshop 2007, LNAI, vol. 5085, Lisbon, Portugal, January 2009, pp. 108–113.
[7] P. Dreuw, H. Ney, G. Martinez, O. Crasborn, J. Piater, J. Miguel Moya, M. Wheatley, The sign-speak project - bridging the gap between signers and speakers, in: 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, 2010, pp. 73–80.
[8] P. Dreuw, J. Forster, Y. Gweth, D. Stein, H. Ney, G. Martinez, J. Verges Llahi, O. Crasborn, E. Ormel, W. Du, T. Hoyoux, J. Piater, J.M. Moya Lazaro, M. Wheatley, SignSpeak – understanding, recognition, and translation of sign languages, in: 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010, pp. 65–73.
[9] E. Efthimiou, S. Fotinea, T. Hanke, J. Glauert, R. Bowden, A. Braffort, C. Collet, P. Maragos, F. Goudenove, DICTA-SIGN: sign language recognition, generation and modelling with application in deaf communication, in: 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010, pp. 80–84.
[10] J.L. Gauvain, C.H. Lee, Maximum a-posteriori estimation for multivariate Gaussian mixture observations of Markov chains, IEEE Trans. SAP 2 (1994) 291–298.
[11] C. Gonzalez-Morcillo, G. Weiss, D. Vallejo, L. Jimenez, J.J. Castro-Schez, A multiagent architecture for 3D rendering optimization, Appl. Artif. Intell. 24 (4) (2010) 313–349.
[12] T. Hanke, L. König, S. Wagner, S. Matthes, DGS corpus & dicta-sign: the hamburg studio setup, in: 4th Workshop on the Representation and Processing of Sign Lan-guages: Corpora and Sign Language Technologies (CSLT 2010), Valletta, Malta, May 2010, pp. 106–110.
[13] A. Karami, B. Zanj, A. Sarkaleh, Persian sign language (PSL) recognition using wavelet transform and neural networks, Expert Syst. Appl. 38 (3) (2011) 2661–2667.
[14] P. Koehn, Statistical machine translation, phD, Cambridge University Press, 2010.
[15] P. Koehn, F.J. Och, D. Marcu. Statistical phrase-based translation, in: Human Language Technology Conference 2003 (HLT-NAACL 2003), Edmonton, Canada, 2003, pp. 127–133.
[16] V. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals, Sov. Phys. Doklady 10 (8) (1966) 707–710.
[17] F. López-Colino, J. Colás, Synthesizing mood-affected signed messages: modifications to the parametric synthesis, Int. J. Human–Comput. Stud. 70 (4) (2012) 271–286.
[18] V. López-Ludeña, R. San-Segundo, J.M. Montero, R. Córdoba, J. Ferreiros, J.M. Pardo, Automatic categorization for improving Spanish into Spanish sign language machine translation, Comput. Speech Lang. 26 (3) (2012) 149–167.
[19] V. López-Ludeña, R. San-Segundo, C. González Morcillo, J.C. López, J.M. Pardo, Increasing adaptability of a speech into sign language translation system, Expert Syst. Appl. 40 (4) (2013) 1312–1322.
[20] V. López-Ludeña, R. Barra-Chicote, S. Lutfi, J.M. Montero, R. San-Segundo, LSESpeak: a spoken language generator for Deaf people, Expert Syst. Appl. 40 (4) (2013) 1283–1295.
[21] I. Marshall, E. Sáfár, Grammar development for sign language avatar-based synthesis, in: Proceedings HCII 2005, 11th International Conference on Human Computer Interaction (CD-ROM), Las Vegas, USA, July 2005.
[22] S. Morrissey, A. Way, An example-based approach to translating sign language, in: Workshop Example-Based Machine Translation (MT X–05), Phuket, Thailand, September 2005, pp. 109–116.
[23] S. Morrissey, A. Way, D. Stein, J. Bungeroth, H. Ney, Towards a hybrid data-driven MT system for sign languages, in: Machine Translation Summit (MT Summit), Copenhagen, Denmark, September 2007, pp. 329–335.
[24] Q. Munib, M. Habeeb, B. Takruri, H. Al-Malik, American sign language (ASL) recognition based on Hough transform and neural networks, Expert Syst. Appl. 32 (1) (2007) 24–37.
[25] R. Niewiadomski, E. Bevacqua, M. Mancini, C. Pelachaud, Greta: an interactive expressive ECA system, in: 8th International Conference on Autonomous Agents and Multiagent Systems, vol. 2, 2009, pp. 1399–1400.
[26] J. Och, H. Ney, A systematic comparison of various alignment models, Comput. Linguist. 29 (1) (2003) 19–51.
[27] K. Papineni, S. Roukos, T. Ward, W.J. Zhu, BLEU: a method for automatic evaluation of machine translation, in: 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, PA, 2002, pp. 311–318.
[28] R. Pfau, J. Quer, "Nonmanuals: their gramatical and prosocid roles" Sign Languages, Cambridge University Press, 2010. pp. 381–402.
[29] M. Porta, Vision-based user interfaces: methods and applications, Int. J. Human–Comput. Stud. 57 (1) (2002) 27–73.
[30] R. San-Segundo, R. Barra, R. Córdoba, L.F. D'Haro, F. Fernández, J. Ferreiros, J.M. Lucas, J. Macías-Guarasa, J.M. Montero, J.M. Pardo, Speech to sign language translation system for Spanish, Speech Commun. 50 (2008) 1009–1020.
[31] R. San-Segundo, J.M. Pardo, F. Ferreiros, V. Sama, R. Barra-Chicote, J.M. Lucas, D. Sánchez, A. García, Spoken Spanish generation from sign language, Interact. Comput. 22 (2) (2010) 123–139. 2010.
[32] R. San-Segundo, J.M. Montero, R. Córdoba, V. Sama, F. Fernández, L.F. D'Haro, V. López-Ludeña, D. Sánchez, A. García, Design, development and field evaluation of a Spanish into sign language translation system, Pattern Anal. Appl. 15 (2) (2011) 203–224.
[33] A. Stolcke, SRILM – An extensible language modelling toolkit. in: Proc. Intl. Conf. on Spoken Language Processing, vol. 2, Denver USA, 2002, pp. 901–904.
[34] M. Thiebaux, S. Marsella, A.N. Marshall, M. Kallman, Smartbody: Behavior realization for embodied conversational agents, in: 7th International Joint Conference on Autonomous Agents and Multiagent Systems, vol. 1, 2008, pp. 151–158.
[35] C. Traxler, The stanford achievement test, 9th Edition: national normin and performance standards for deaf and hard of hearing students, J. Deaf Stud. Deaf Edu. 5 (4) (2000) 337–348.
[36] Judy Tryggvason, VANESSA: A System for Council Information Centre Assistants to Communicate Using Sign Language, School of Computing Science de la Universidad de East Anglia, 2004.
[37] C. Vogler, D. Metaxas, A framework for recognizing the simultaneous aspects of ASL, CVIU 81 (3) (2001) 358–384.
[38] U. von Agris, D. Schneider, J. Zieren, K.-F. Kraiss, Rapid signer adaptation for isolated sign language recognition, in: Proceedings of CVPR Workshop V4HCI, New York, USA, June 2006, p. 159.
[39] Lee Yung-Hui, Tsai Cheng-Yueh, Taiwan sign language (TSL) recognition based on 3D data and neural networks, Expert Syst. Appl. 36 (2) (2009) 1123–1128 (Part 1).
[40] S.B. Wang, A. Quattoni, L.-P. Morency, D. Demirdjian, T. Darrell, Hidden conditional random fields for gesture recognition, in: Proceedings of CVPR, vol. 2, June 2006, pp. 1521–1527.
[41] A. Watt, M. Watt, Advanced Animation and Rendering Techniques, Pearson Education Harlow, UK, 1992.