SignSpeak

Scientific understanding and vision-based technological development for continuous sign language recognition and translation

Grant Agreement Number 231424

Small or medium-scale focused research project (STREP)

FP7-ICT-2007-3. Cognitive Systems, Interaction, Robotics

Project start date: 1 April 2009

Project duration: 36 months

Report about the study of the new communication bridge between signers and hearing Community

Major Deliverable D.9.4 – M30

Dissemination Level: Public

Release version: V1.0 – end of September 2011

Authors: M. Carmen Rodríguez (TID), Javier Caminero (TID) and Annemieke Van Kampen (EUD).

Reviewer: Gregorio Martínez Ruíz (CRIC).

SEVENTH FRAMEWORK PROGRAMME

1. Introduction

The SignSpeak project is gathering efforts for making available a new communication bridge between deaf and hearing community. The overall goal of SignSpeak project is a new vision-based technology for recognizing and translating continuous sign language to text, being this the first step to approach this technology at levels already obtained in similar technologies such as automatic speech recognition or statistical machine translation of spoken languages.

In this document an analysis of the communication solution which is completed by the SignSpeak technology is presented. The analysis wants to be realistic, taking into account all the drawbacks and technological problems which could appeared once the integration of the different technologies is carried out. Thus, in Section 2 the communication needs for deaf community which are supposed to be partially fulfilled through SignSpeak are described. Later, in Section 3, a state-of-the-art has been provided for the technologies implied in the communication bridge proposed: text-to-speech, automatic speech recognition, signing avatars and natural signing to text (as SignSpeak).

Furthermore a survey has been carried out in collaboration with the European Union of the Deaf. Five experts from the deaf community have been interviewed about difficulties in the communication between deaf and hearing people, about the possibilities of SignSpeak technology and about some scenario proposals. This work is described in Section 4.

2. Communication needs for deaf community

In the current section we review the main problematic issues about the communication needs for deaf community (Section 2.1) and how they solve these problems in daily life (Section 2.2). Regarding this it is noteworthy the data presented in Section 2.2.3 where a preliminary analysis of the more relevant factor associated to the selection of communication services is carried out. Afterwards in Section 2.3 we propose a compilation of various scenarios which are representative of communication problems between deaf and hearing community.

2.1. General Considerations

Communication for Deaf community revolves around sign language since it is “the only language Deaf people can acquire effortlessly and spontaneously when given the right input” (Wheatley and Pabsch 2010). Unfortunately deaf and hard of hearing signers have serious limitation for communicating with the vast majority of European citizens, who have no sign-language skills and thus, the integration into educational, social and work environments is not complete.

It is worthy to notice that currently there are 31 sign languages in the European Union (EU). 26 of them corresponding to each Member State (excluding Luxembourg which uses a dialect of German Sign Language) and an additional sign language for Belgium, Finland, Spain, Estonia and Sweden. Apart from national sign languages there is also an international communication system often called International Sign (IS). All these languages are mother tongue for the corresponding deaf communities in each country. Although mother tongue is defined as the first language one has acquired, for deaf community is more complex than that. Most deaf children (around 90-95%) are born into hearing families which do not have signing skills at all. The remaining 5-10% (born into deaf families) is luckier and they will acquire a sign language naturally and in similar stages as hearing children do with a spoken language. In summary the national sign languages need to be considered the mother tongues of Deaf people and in order to this, it is necessary to re-define the concept of mother tongue (Wheatley and Pabsch 2010).

Taking into account these peculiarities we realize that although barrier-free communication is a basic human right, deaf people find numerous barriers. Some of these barriers include the presence of an operator (which may be seen intrusive and do not represent parity with hearing people), slow communications connections and lack of awareness of how to communicate with people who are deaf or have speech difficulties. About this issue a recent study with deaf participants was carried out by OfCom, a competition and regulatory authority from UK (Opinion Leader 2011), and their results state the following six general conditions to achieve barrier-free communications:

- Services and equipment should meet the many different needs of those who are deaf and have speech difficulties.
- Use of everyday and mainstream technology should help to ensure that communications methods are up-to-date and that the general public understand how to use them.
- Services should cater for different situations and different needs so that individuals can be offered choice in the ways that they communicate.
- Services should allow deaf people and those with speech difficulties to enjoy the same level of access to their chosen means of communication as that which is available to the wider community.
- Access should be free at the point of use for alternative, specialist services or equipment that aids communication for people who are deaf or have speech difficulties.
- There should be better public understanding of the issues faced by people who are deaf or have speech difficulties and the ways in which their needs can be taken into account.

In addition they identified in this study the most important characteristics or requirements for a communication equipment intended to deaf people, namely: choice (i.e. in the ways they communicate as well as in the equipment and technology), respect (i.e. offering an experience as similar as possible to that enjoyed by hearing people, hold a conversation with no interruptions, hold private and confidential conversations without the involvement of a third party), independence (i.e. being able to call an ambulance, or emergency services), equal access to communication services, awareness and inclusivity (i.e. the development of new technologies into mainstream products and services should accommodate the needs of deaf people).

Having these issues into consideration, SignSpeak project aims to provide deaf people a communication bridge between signers and hearing community. Thus, a new vision-based technology for translating continuous sign language to text is developed in order to provide new e-Services to the deaf people and to improve their communication with hearing people and the other way around. As a consequence of the automation of the services and applications provided by the SignSpeak technology, users' privacy feeling and their confidentiality in the communication process would be improved.

2.2. Communication services

Next, a number of different options used for communication in deaf community and other aspects related with are described in order to know in depth the real communication needs that deaf people might have. Besides, a set of scenarios which includes some difficulties for communication is analysed.

2.2.1. Face-to-face communication

In case of face-to-face communication, deaf people usually communicate through sign language. Sign language is a complex language, having its own grammatical structure and syntax. It is a language of space and movement using the hands, body, face and head. Due to sign language is not well known by the hearing community, sign language interpreters are often needed. They are highly trained, skilled and qualified in sign language and deaf awareness. They should be fully insured and regulated and follow a recognised code of ethical practice.

When you are interacting with an interpreter a set of key points should be considered:

- The hearing person should both address and face the deaf person when they are speaking. They should not face and speak to the interpreter.
- Different from spoken language translating, sign language interpreting operates parallel with speech. There may be a couple of seconds delay but there is no need to stop and wait for the interpreter to translate information.
- Don't use formulas as, "Tell him…", "Ask her…" Speak as it was done directly to the deaf person.
- The interpreter is in the role of facilitating communication. So she or he is not going to express opinion or comment, or to advocate on behalf of the deaf (or hearing) person.
- When communicating through sign language interpreters always speak normally and clearly.

Although sign language is the main communication means for these situations, people may choose different methods of communication for different situations. Thus, a deaf person could prefer to lip-read, that is, to read the lips from the person who is speaking. However, no all deaf people can understand what lipspeakers are saying and it depends on residual hearing (i.e. how much you hear with hearing aids). For those who are able there are also lipspeakers who repeat what is being said to a deaf person but will do so in a way that will enable the deaf person to understand what is being said. Most people do not speak in such a way that their lip movements and facial expression are enabling of lip reading. Lipspeakers can be a helpful option for deaf people who do not sign.

In other situations, as meetings in a noisy environment that can disrupt the effectiveness of hearing aids, then palantypists are used for typing the words being spoken in real time, and these words are directly transmitted to a screen that the deaf person is then able to read from.

2.2.2. Remote communication

Some of the most commonly methods and technologies used when remotely communicating to a deaf person are the following:

Land line

The device used for this purpose is about the size of a small laptop computer with a standard keyboard and small screen to display typed text electronically. The text is transmitted live, via a telephone line, to a compatible device, i.e. one that uses a similar communication protocol (Figure 1).

In certain countries there are Telecommunications Relay Services, so that a deaf person can communicate with a hearing person on an ordinary voice phone using a human relay operator. There are also "carry-over" services, enabling people who can hear but cannot speak ("hearing carry-over"), or people who cannot hear but are able to speak ("voice carry-over") to use the telephone.

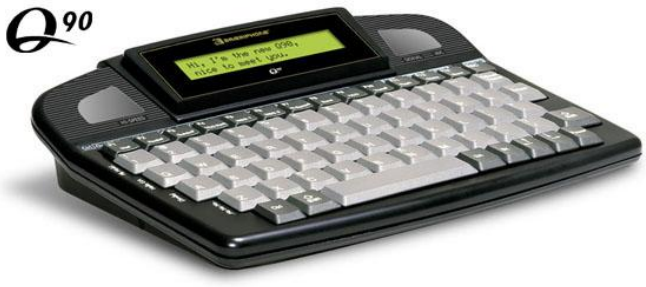


Figure 1. Ameriphone Q90 mobile

In case people with severe hearing loss/deafness still want to use their voice then devices as in Figure 2 are very versatile. This kind of service, called voice carry-over, allows a person who is hard-of-hearing or deaf and does speak to use one's voice while receiving responses from a person who is hearing via the operator's typed text.

Figure 2. Clarity Ameriphone VCO

Another interesting device is CapTel 800i (Figure 3). It has an Ethernet/IP connection which automatically links to the CapTel Captioning Service. Operators at the Captioning Service use voice-recognition technology to transcribe everything the caller says into text, which is instantly transmitted to the CapTel 800i over the Internet connection. The captions appear on the phone's display screen, giving users the ability to hear what they can of the phone conversation and read what they need to in the display screen.

Figure 3. IP-based caption telephone Captel 800i

Finally, in the case users want to communicate each other through sign language, they can use a videophone. They include a big display where the video is showed. It is prepared for full duplex video and audio transmission.

Figure 4. Nortel IP Phone 1535

Mobile phone

Some of the most popular mobile services in the deaf community are:

- Text (SMS) and multimedia (MMS) messages.
- T-Mobile Sidekick was one of the most popular mobile devices in the deaf community. The first T-Mobile Sidekick debuted in 2002. It was a text-based communication medium with a QWERTY keyboard providing real-time e-mail and instant messaging, serving as a telecommunications tool for use both inside and outside the home. T-Mobile's decision to offer a "data only" wireless plan, thereby eliminating charges for unusable voice minutes, also played a role in the device's dominance in the deaf community. However Sidekick hasn't evolved as fast as other smartphones and so, many deaf users are moving to the BlackBerry and other devices.

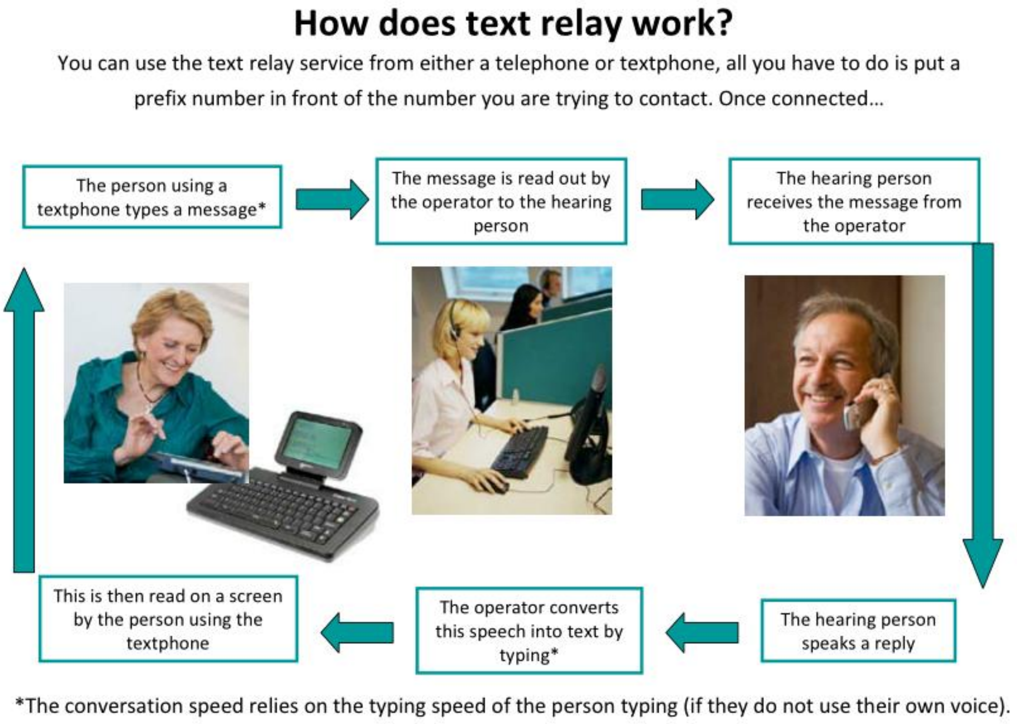
Figure 5. T-Mobile Sidekick and BlackBerry devices

- Smartphone text chat applications (i.e. "Whatsapp", "LiveProfile").
- Smartphone 4G videocalling applications: (i.e. iPhone4 Facetime Videocalling and HTC EVO phone from Sprint).

Internet

- E-mail.
- Messenger, chats, Skype (text and video using webcam).
- Social networks.
- Relay services. There are two types of relay services: traditional and broadband video. Traditional relay services have all communication in text only, through a text-phone or via Internet (see Figure 6). A video relay service uses a videophone or a webcam, and a sign language interpreter. Several companies offer relay services. Most of them offer multiple options (web, video, instant messaging). For example, in Greece, the University of Athens offers a text and video relay service for students who are deaf or hard of hearing to provide remote communication between students and their fellow students and with academic and administration staff of the University[1]. In Sweden a text and video relay service is procured by the National Post and Telecom Agency (financed by national taxes). The text relay service is accessible through the Internet enabling a software application to make it possible to use a computer as a text phone. This service itself is open 24/7.

[1] http://access.uoa.gr/Unit%20Sign.htm



How does text relay work?

You can use the text relay service from either a telephone or textphone, all you have to do is put a prefix number in front of the number you are trying to contact. Once connected...

The person using a textphone types a message*

The message is read out by the operator to the hearing person

The hearing person receives the message from the operator

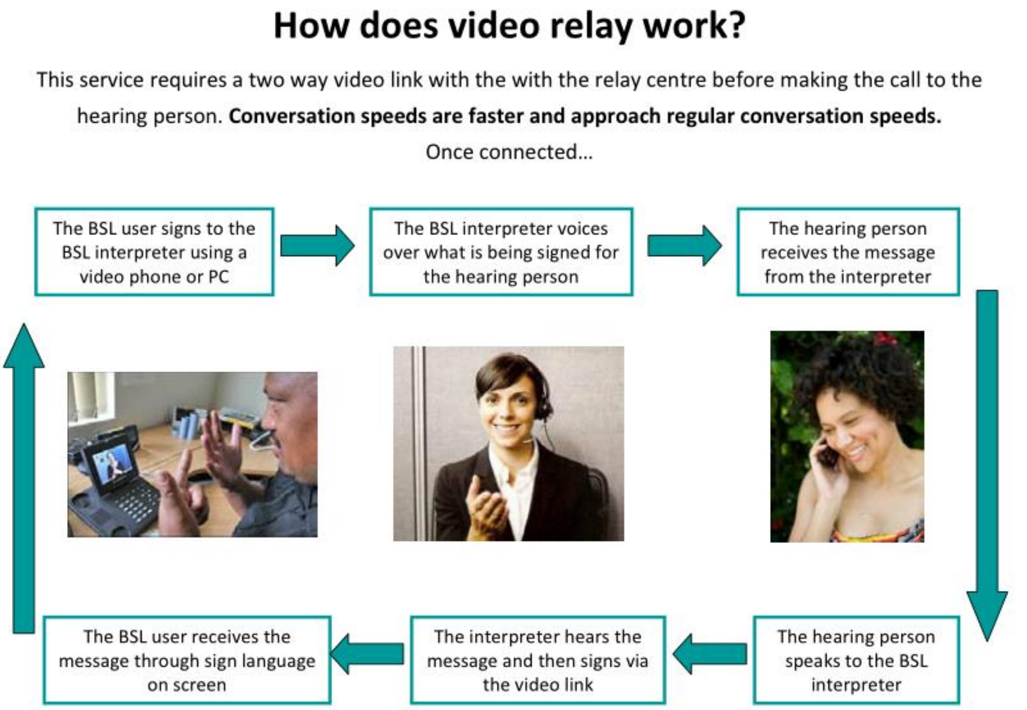
The hearing person speaks a reply

The operator converts this speech into text by typing*

This is then read on a screen by the person using the textphone

*The conversation speed relies on the typing speed of the person typing (if they do not use their own voice).

Figure 6. How a text relay service works (from (Opinion Leader 2011))



How does video relay work?

This service requires a two way video link with the with the relay centre before making the call to the hearing person. **Conversation speeds are faster and approach regular conversation speeds.** Once connected...

The BSL user signs to the BSL interpreter using a video phone or PC

The BSL interpreter voices over what is being signed for the hearing person

The hearing person receives the message from the interpreter

The hearing person speaks to the BSL interpreter

The interpreter hears the message and then signs via the video link

The BSL user receives the message through sign language on screen

Figure 7. How a video relay service works (from (Opinion Leader 2011))

- On the other hand, anyone who prefers to make telephone calls using sign language and the Internet can use IP Video Relay Service (IP-VRS). Using a Video Interpreter and web camera they can communicate with voice telephone users in their preferred language, which is most natural for them. They sign to the Video Interpreter who voices their conversation to the voice user, then signs the voice user's conversation back to them.

2.2.3. Performance and relevant factors

Next we combine the most important findings obtained in (Opinion Leader 2011) and the survey carried out by EUD and TID for this deliverable. The former was focused in the deaf people's acceptance of communication services and the latter covered some aspects related with the real use of technology by some experts from deaf community.

Then, some of the difficulties encountered when communication services are used by deaf people are:

- Communication technologies are not always easy to use. Most of them are designed for hearing people and minorities, as deaf people, find several barriers.
- Text relay could be improved in terms of speed, perceived lack of confidentiality, being impersonal and being off-putting to hearing people. Some participants said that the availability of other modes of communication, such as email, SMS text messaging and Skype, have led some who are deaf or have speech difficulties to reduce or stop using text relay services.
- Text relay lacks of many desirable qualities for providing "real conversation" (accessibility, availability, confidentiality, mobility and allowing a user to communicate as a "real person"). "REACH 112"[2] project is working on the concept of 'total conversation' where you can use video, text and speech at the same time in a call.
- The communication method may vary according to user's personal preferences.
- Sign language users are more positive about video relay services, although they have specific concerns about the possible expense of video relay, difficulty in using the equipment, the need for an appropriately light environment, and lack of confidentiality.
- Although most of smartphones provide video conferencing, the service is too slow and usually it turns out into delays.
- SMS is seen as tedious and prone to provoke miscommunication.
- Video chatting programs are considered not accessible. They usually have very low video quality (which could affect lip reading) and bad synchronization between images (signs) and voice.

Thus, the challenges that arise in order to improve the communication services could be summarize as follows:

- Increase efficiency and speed in communication services.
- Improve ease of use.
- More widespread acceptance of text and email by business/third party organisations (they usually do not think in deaf people when they provide a communication channel).
- Make available different communication methods in different contexts (i.e., family, friends, work colleagues, etc.)
- Not having to rely on a third party to "listen" and communicate for her by the emergency services, businesses and other third party organisations
- Being able to have a real time conversation with someone "That makes you feel like a real person" in the sense of you could "hear" the emotions of the other person.

[2] http://www.reach112.eu

- Speech-enabling devices built into phones, so it is "always there".
- Working on all this, overcoming economic issues (specially now in crisis times). Improving communication between deaf and hearing community would help to create new business opportunities.
- Total conversation: speech, text and sign to let deaf and hard of hearing people and those who have speech difficulties to choose.

Then accessibility, availability, confidentiality and allowing a user to communicate as a "real person" were the most important aspects of communication services. Besides it was also important for communication services to be mobile. Respondents also confirmed the importance of privacy (which is similar to confidentiality), of 24/7 service (which is an aspect of availability), of being able to have real-time conversations (which contributes to being able to communicate as a real person), and of affordable services (with fair charging structures).

Some features in the communication equipment that could be helpful especially for deaf and hard of hearing people were identified, namely: mobility, alerts for incoming calls (including flashing lights, vibrating alerts, and amplified ringers for those with some hearing), compatibility with hearing aids, amplified phones and a voicemail to SMS service (some participants thought that there were considerable benefits of having voicemail converted to text).

In Figure 8 a bar graph is presented which displays the relevance, based on the opinions of the experts who participate in our survey, of different variables regarding what aspects are more important for selecting technology products. The values have been extracted from the second question of the expert survey (Appendix A. Expert survey). Participants could choose the relevance of some proposed factors assessing them from 1 (strongly disagree) to 5 (strongly agree). The bars in the Figure 8 are highlighted with three colours depending on their values (red for low values, orange for values in the range 3-4 and green for those between 4 and 5). Focusing in the analysis of the results, the price of the product is not seen as a fundamental factor. It seems that when the service provided by the technology is really useful, price doesn't matter. Logically this fact, that is applicable to any target population, gains importance for deaf community since technology helps to break down very annoying communication barriers. At the other end, the more relevant factors are: usefulness, ease of use and to have the ultimate technology. Related with the abovementioned great necessity of technology by deaf community, it seems clear that usefulness is a key variable for selecting a device. This may indicate that cutting-edge technology has been associated to a perceived loss of reliability of the technology's performance (a factor which wasn't represented in the survey). And finally, due to the accessibility difficulties which traditionally deaf people have to face in the use of technology products, easiness of use is also lightly highlighted.

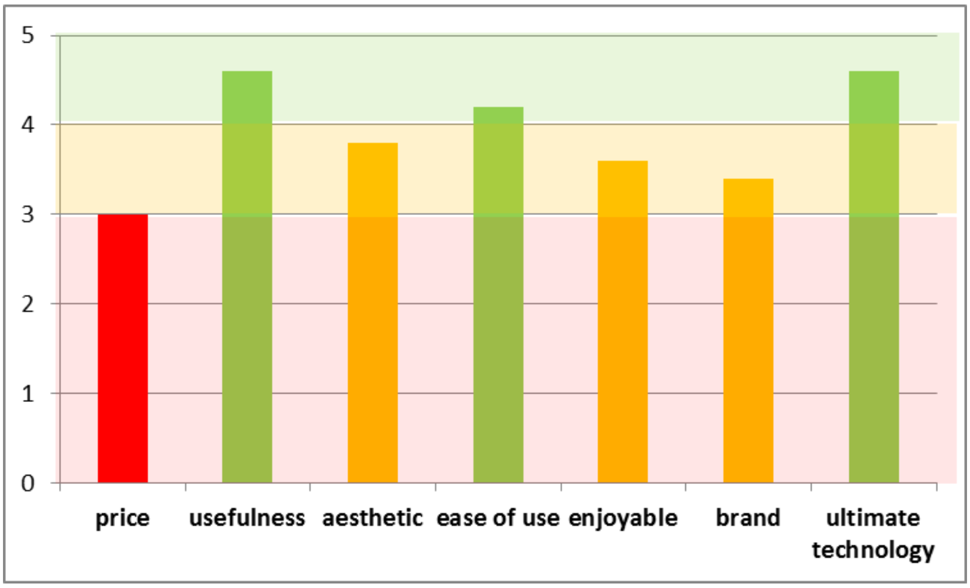


Figure 8. Relevant factors in the selection of technology products

2.3. Scenarios with communication problems

In the expert survey carried out for this deliverable we want to capture the feelings and opinions of deaf community when they face several difficult situations. In order to that, a question about the more unpleasant situations when they have to communicate with non-signers was proposed. Next, some of these scenarios are listed together with the most relevant comments of our experts:

- Telephoning hearing people through relay service in text, which is not my first language.

  "Recently, a project was launched for a relay service using sign language. The problem is that it only opens from 8 to 16. At night if deaf people need to call they do not have access. Then you have to use text, but it implies barriers. Text is not my first language and so expressing oneself by text is hard"

- E-mails via text messages which is not my first language.

  "In relay services you need to think instantly what you write. It could cause misunderstanding. However email gives you time to read what you write so it is easier than relay services. But if you need to communicate just at that moment then relay service is the best option (for example, to cancel a meeting)"

  "To express oneself using sign language gives more freedom and is easier. Deaf people think visually so if they have to write text, they need to translate from sign language to written text."

  "When I send e-mails using text to friends and known people, it is ok. But for formal communications to deaf or hearing person then I have more difficulties. I use to ask my interpreter to write a text while I sign, then I e-mail the resulting translation"

- Video films using sign language are often not subtitled so hearing people cannot understand sign language.

  "For private services (as Facebook) it is up to deaf people if they want to subtitle sign language videos. However public services required these subtitles. It depends on the target population"

  "Deaf people talk about their problems through video films. If hearing people had access to these films then they would become aware about deaf community's problems"

  "Hand op Tafel (web TV program for deaf people) uses subtitles but they are inserted into the videos which are uploaded to the web, not real time. Hearing people also need a motivation to learn sign language"

  "It is the same problem as deaf people who do not have access to voiced films without subtitles"

- Going to public authorities/service (passport service, bank, etc.) where most people cannot sign.

  "It must be accessible. People have rights for a full accessible communication"

  "When I go to the bank, I tell them that I am deaf. Then, they try to talk slower and clearer. If I do not understand them, then I ask them to write down what they are saying. Some deaf people cannot write well because it is their second language. Even for Dutch language it sometimes happens deaf and hearing people have Dutch as their second language"

  "If it is an official talk I use to book an interpreter. Public authorities would take care of the expenses. If not they would exist an unfair situation."

- Relay service almost never opens 24 hours a day.

  "If there is an accident at night, how could we reach 112?"

  "There is an interpreting centre in Germany which serves as a relay service. Unfortunately it has a limited timetable. Hearing people can call 24 hours a day. It breaks UN convention. Also this is a paid service (although I get some subsides from the Government). Hearing people do not need to pay, so I find it unfair"

- Hearing people cannot learn sign language without instruction in their own written or spoken language.

  "*It depends on the teaching methodology you use for teaching sign language. It is necessary innovative approaches beyond 'a sign for a word'. Understanding sign language is not a complete solution, it is necessary to understand experience and culture of deaf people"*

  "I am a teacher of sign language for hearing people. I believe hearing people need to learn visually. So subtitles are not so important. They need to learn using hands. If they are focus on subtitles then they won't think in sign language"

3. SignSpeak: a new communication bridge

In the previous section, we have identified some general communication needs, difficulties and challenges in communication services and a number of important features in communication technologies that deaf people may need when they use a communication service. Having these findings into account, we next focused in the SignSpeak technology. We analyse how this technology would tackle a range of deaf people communication needs. The objective of this analysis is to identify the technological gap, from a user-centred point of view, that SignSpeak covers.

Some of the identified strengths of SignSpeak technology are:

- There is the possibility to transmit emotional content. Users can see the interpreter and they can receive the emotional content of the discussion through the SignSpeak video channel.
- The combination of text and video may improve the accuracy in the communication service.
- Deaf people do not have to rely in a third part in the communication service. Therefore, the confidentiality and the users' privacy feeling are improved.
- The automation of the SignSpeak system may improve the availability of the system: it gives a 24/7 service.
- Having the ability to communicate with sign language, would make users feel more comfortable and confident than when communicating in any other language.
- SignSpeak would be more cost-effective than using interpreters would.

In the other hand there are also some disadvantages associated to this technology:

- There are concerns about the cost of the equipment and cost of the service.
- Technological doubts: quality of the image, translation accuracy sign-to-text (possibility of errors in interpretation), smoothness of the service.
- Concerns about the degree of anonymity offered by the SignSpeak (regarding to the video channel, as it is not appropriated to be seen in particular situations).

Furthermore, when SignSpeak technology is integrated in the communication bridge between hearing and sign language community (Figure 9), it is necessary to take into account the special features of all technologies involved: text-to-speech, text-to-sign-language and speech-to-text.

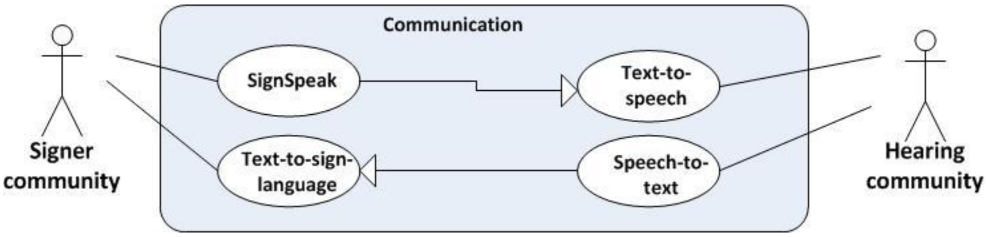


Figure 9. Communication bridge between Signer and Hearing community

As it is showed in Figure 9 SignSpeak technology captures the video information from the signer and converts it into text. In order to do that multimodal processing of the video is carried out and afterwards, the resultant sequence of signs is translated to natural language. Using a text-to-speech (TTS) engine, the intended message is communicated to people who are able to hear. In the reverse way the speech from hearing community is captured and translated to text. Then the text is used by virtual avatars which compose the suitable sequence of signs.

Next we will present a review of the technologies involved in the communication bridge abovementioned. We will give some details about the current performance of these technologies and what are the principal actors into each sector. We will also point at the weak points for each technology and in which work contexts these problems could appear.

3.1. Technology review

3.1.1. Text-to-Speech

A text-to-speech (TTS) synthesiser can be defined as a piece of software which transforms into speech any input sentence in text format (Dutoit 1997). This functionality makes a TTS very useful for communication systems because it avoids pre-recording every sentence or words we plan to use in a service. However, it is important to realize the complexity of this task. For example, there is the problem of the existence of abbreviations (Mr, Ms…) which cannot be transformed to speech directly. Or, also, there are different intonations which could be adopted in function of the word position in the sentence (beginning vs. end), the type of the sentence (declarative vs. interrogative), or the emotional cues (happy vs. sad).

A typical TTS system has two main modules: the first for the analysis of the linguistics and the second for the transformation of the previous module to a sequence of sounds (speech signal). The 'natural language processor' (NLP) is the more complex. Firstly, it has to deal with the abbreviation issue, then a morphological, grammar and prosodic analysis is made in order to get the phonemes which the 'digital signal processor' should transform to output speech. The main three methods to convert the output of the NLP into synthesized speech are: formant, articulatory and concatenative. A formant synthesiser reproduces the changing formants (frequencies and bandwidths) of speech. simulating lung pressure, vocal fold vibration, and so on. And, a concatenative synthesis implies to concatenate pre-stored human utterances. The utterance units could vary from phonemes, diphones or longer units (López-Cózar and Araki 2005).

The freely available TTS systems more popular are the following:

- Festival[3] which offers a general framework for building speech synthesisers and also some examples in various languages (American and British English, and Spanish). Festival allows several methods for building a voice based on concatenative philosophy (Clark, Richmond and King 2007).
- MBROLA[4], initiated by the TCTS Lab of the Faculté Polytechnique de Mons (Belgium), has as primary goal to obtain synthesisers for as many languages as possible. In order to do that MBROLA uses diphone concatenation techniques. They offer more than 30 languages such as German, Hungarian, Italian, Japanese, etc.

The commercially available options are quite numerous. Among them the more relevant are:

- Loquendo TTS[5] has also a wide portfolio of voices (about 30 languages and more than 70 voices). The most remarkable feature is the possibility to add some expressivity, reading the sentences in several different styles and mixing the voice with crying, laughing and so on. They have a version for mobile devices.
- Acapela Group[6] is also a very relevant actor. Their voices can add effects, expressions and can be reproduced with different speeds. Their catalogue of voices is huge. They have a version for mobile devices
- Nuance Vocalizer[7] is a product from Nuance, a prominent American company. They also offer some speech synthesis services focus on specific contexts as automotive. Recently, Nuance has acquired Loquendo.
- AT&T Natural Voices[8] are ones of the most realistic in the market. They have a version for desktops and servers.

Despite of the great performance of the aforementioned systems there are yet critics to the use of TTS for certain applications. Some of these problems are part of the future key technical challenges. Next we cite the most relevant ones:

- Pronunciation of new and rare words: it could happen there are words which aren't available at the corpus or they aren't correctly pronounced (i.e. Blogger, Flickr or words in other languages)(Spiegel 2003).
- Prosody, that is, speaking style related to several factors (i.e. topics, discourse segmentation, etc.). Engines usually don't take into account important topics as paragraphs or changing topics to modulate the resulting speech (Hirschberg 2002).
- Availability in new languages. Although the current catalogue of TTS companies is really wide, including more than 30 languages, there are over 6000 languages in the world. The creation of a new language implies important investments to gather hours of voice, and so some minority languages are not included in the development plan of companies. Furthermore, most of languages have also some special features which can make the development process either

[3] http://www.cstr.ed.ac.uk/projects/festival
[4] http://tcts.fpms.ac.be/synthesis/mbrola.html
[5] http://www.loquendo.com/en/products/text-to-speech
[6] http://www.acapela-group.com
[7] http://www.nuance.com
[8] http://www.naturalvoices.att.com

much easier or considerably harder.

- Conveying emotion and expressiveness. Appropriate tone, pitch accent and suitable speech intensity should be set to create emotional speech. These variables could give different meanings to the speech built from the same text message (Rebordao, et al. 2009).
- Personalization of the voices. Users use to like particular kind of voices (i.e. childish, different accents…) based on their personal preferences, the context application, etc. The direct solution would be to create corpus for each wanted voice, but a strategy which implies the transformation of existing engines would be cheaper and faster (Olinsky and Cummins 2002).

3.1.2. Speech-to-Text

The human voice is generated by the vibration of the vocal cords. The vibration of the cords moves the air and these variations of pressure arrive to the listener's ear. Then the pressure waves are transformed into a signal that is processed by the brain and properly interpreted. The acoustic features of this signal allow the listener to differentiate one sound from another, and that is what an Automatic Speech Recogniser (ASR) tries to accomplish.

A conceptual ASR consists into two modules (Rabiner and Juang 1993): the feature extractor and the classifier. The former obtains, from the voice signal fragmented and applying different signal processing techniques, a set of representative features. Considering these features as a sequence of vectors, then the classifier is, firstly, trained and afterwards, it is used for the recognition of feature patterns. The probabilistic models are built using techniques as Hidden Markov Models or Neural Networks.

The ASR systems can be classified according to different criteria. Based on *allowed speakers* they are: single speaker, speaker independent and speaker adapted speech recognisers. The former can only recognise speech uttered by the same speaker who trained it. This kind of systems has very good performance and can be used for adapting the recogniser to specific pronunciation problems. In the case of speaker independent systems, they can recognise words from different speakers who have trained the system. However the accuracy is worse. Finally, adapted speech recognition implies an initial training to adapt the models to a particular speaker. Based on *allowed user speaking style*, isolated words, connected words and continuous speech system can be considered. The former requires the user makes clearly defined paused between words, in order to define the beginning and the end of the utterance. Connected words systems also require pauses but they are shorter. And, in the latter, there is no necessity to make pauses. So users are allowed to speak normally. Another classification based on *allowed user speaking style* defines read speech and spontaneous speech. Systems which can recognize read speech are able to deal with a formal speaking style while the latter, are prepared to face typical problems of natural speaking as hesitations, grammatical errors and so on. Based on the *vocabulary*, that is the set of words which can be recognised, systems' complexity varies deeply. They are often classified as follows: small vocabulary systems can recognise up to 99 different words; medium vocabulary systems can recognise up to 999 different words; and large vocabulary systems can recognise more than 1000 different words.

Find further details about this issue in (López-Cózar and Araki 2005).

Regarding the different options for selecting an ASR, a list of the most relevant is presented next:

- CMU Sphinx[9] is a set of tools and a series of recognisers (from Sphinx to Sphinx4) which was developed at Carnegie Mellon University. Sphinx is a continuous speech, speaker independent recognition system. In the last version, written completely in Java, a speaker adaptation procedure is provided. They also have a release of the recogniser for mobile devices, called PocketSphinx.

[9] http://cmusphinx.sourceforge.net

- iATROS[10] is a speech recogniser infrastructure that has been adapted to be used in both speech and handwritten text recognition. iATROS provides a modular architecture which implements a Viterbi search on a Hidden Markov Model network. The entire infrastructure is implemented in C.
- RWTH ASR[11] is software which contains a speech recognition decoder and tools for the development of acoustic models. It has been developed by the Human Language Technology and Pattern Recognition Group at the RWTH Aachen University. Speaker adaptation, speaker adaptive training, unsupervised training, a finite state automata library, and an efficient tree search decoder are some of the most notable components.
- Dragon Naturally Speaking[12] from Nuance which has a version for Windows OS and Mac OS (called Dragon Dictate). The software supplies three kinds of applications: dictation, text-to-speech and command input. The dictation app is widely used in legal and medical fields.
- Microsoft Speech API[13] is an interface developed in order to access the recognition services in Windows OS. Recently a cloud version of the recogniser which was named as Microsoft Tellme was released.

However, most of the times, the performance of an ASR system depends drastically on external factors (A. Acero, Acoustical and Environmental Robustness in Automatic Speech Recognition 1992):

- Input level: speakers normally don't speak with the same volume and position respect the microphone. This can affect between different utterances or even in the same utterance.
- Additive background noise: speech recognisers' accuracy varies greatly when the training and the verification are carried out with different noise levels. Anyway when SNR ratio is less than +10dB, the speech is considered so corrupted than even training and testing with the same level of signal is not enough to perform a proper recognition.
- Spectral tilt: it is a channel distortion which could be produced by different sources (i.e. speaker physiology characteristics, speech styles, room acoustics, recording equipment, etc.)
- Physiological differences: due to differences in vocal tract size and shapes. For example, it is known that male voices exhibit lower formants than those of females.
- Socio-linguistic factors: from a linguistic point of view big differences have been established between the diversity speaking styles (Eckert and Rickford 2001). These alterations are usually conveyed to the features used by speech recognizers.
- Interference by the speech of other speakers (the cocktail party effect): most of the recognition systems suppose that there is only one voice in the speech that is target of the analysis. So the presence of other voices causes a dramatic degradation in recognition rates.
- Real-world issues: a set of variables which come from the use of ASR systems in real life and with real people, that is, speaking with disfluencies (i.e. well, uh, er…), using new words, etc.

3.1.3. Text-to-Sign Language

Recently, the virtualization of everyday life and the gaming industry has promoted a great development of the virtual characters field. The improvement of several communication technologies as the automatic speech recognition or the text-to-speech engines makes real to create virtual agents

[10] http://prhlt.iti.es/page/projects/multimodal/idoc/iatros

[11] http://www-i6.informatik.rwth-aachen.de/rwth-asr

[12] http://www.nuance.com/dragon

[13] http://www.microsoft.com/en-us/tellme

which could interact with users. The benefits are obvious: cheaper customer service and 24/7 availability. Furthermore, through this kind of interfaces, users could establish relationships close to those between humans (Reeves and Nass 1996). Thus, virtual agents' behavior can be designed to react to emotional responses of the users and these reactions could also include nonverbal information (emotions, gestures, etc.)

In the research area the proposals are numerous and applied to different contexts. One of the pioneers was the MIT Media Lab. They have mainly worked on two virtual agents: REA and MACK. REA (Figure 10) is an embodied, multi-modal conversational interface agent who acts as a real-estate salesperson. Rea implements some conversational skills (turn taking protocols, requesting and giving feedback, etc.) which let to maintain a natural face-to-face conversation with the user. The goal of Rea is answering user questions about properties in her database and showing users around the virtual houses.

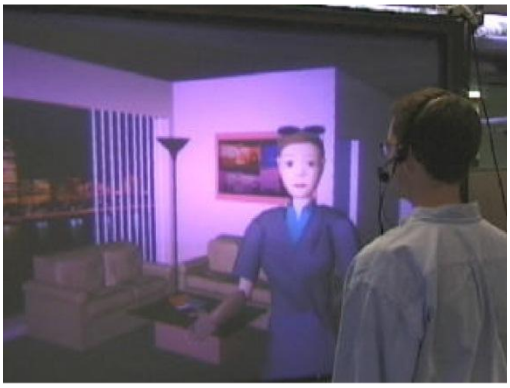


Figure 10. User interacting with REA (Bickmore, Vilhjlmsson and Yan 2000)

MACK (Figure 11) is a conversational agent who can answer questions about and give directions to the MIT Media Lab's various research groups, projects and people. As REA does, MACK uses speech and gestures for communicating with the users. MACK also is able to give indications on a normal paper map that users place on a table between themselves and MACK.

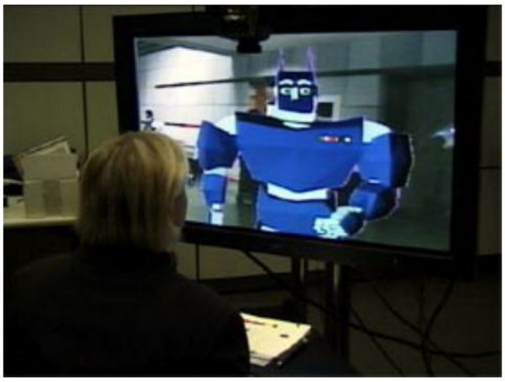


Figure 11. User interacting with MACK (Cassell, Stocky, et al. 2002)

Other examples of conversational agents are: AdApt system which tries to help those people, who are searching apartments in the Stockholm area (Gustafson, et al. 2000); Companions agent who acts as a sociable and emotionally intelligent companion for the user. For example, the user talks about her/his day at the office and the companion is intended to remember appointments or make any other suitable comments (Cavazza, et al. 2010); and FitTracker who was designed to construct, maintain and evaluate long-term relationships between the agent and the users. The selected scenario is an application where user is advised about health habits (Bickmore and Picard 2005). A snapshot of these virtual agents is shown in Figure 12.

COMPANIONS DIALOGUE SYSTEM MONITOR
ENGLISH COMPANION PROTOTYPE
MIT FitTrack

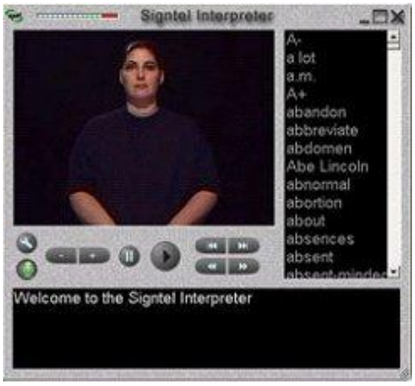
Figure 12. From left to right: AdApt system, Companions agent and MIT FitTracker

Taking into account the great expressivity of some of those virtual agents and the availability of characters with hands, lately, they are being used for expressing messages through sign language. It is important to notice deaf people usually have sign Language as first language. In other hand written a text could mean more difficulties to be read and understood well. Signing avatars could help to widespread the use of sing language in digital media since it has several advantages respect to using videos of human interpreters:

- Videos must be well produced and of high quality, which is expensive for the content provider.
- Each time content changes new videos must be made, increasing the costs further.
- There are continuity issues. Making videos consistent, i.e. using the same signer, in the same clothing and with the same background, so that signed phrases may be joined together, complicates the content maintenance process.
- Storage and download of videos can also be problematic as they are large files. For users on dial up or mobile connections the time and cost involved in download of video sequences may be prohibitive.

Some commercial applications which use avatars or a sort of them are:

- SignTel Application Interpreter[14] which let to translate written and spoken English to American Sign Language (ASL). The system includes more than 30.000 words and 1.400 phrases and idioms which are displayed in pre-recorded videos (see Figure 13). These videos are concatenated for building more complex expressions. They also offer an application for iPhone[15].

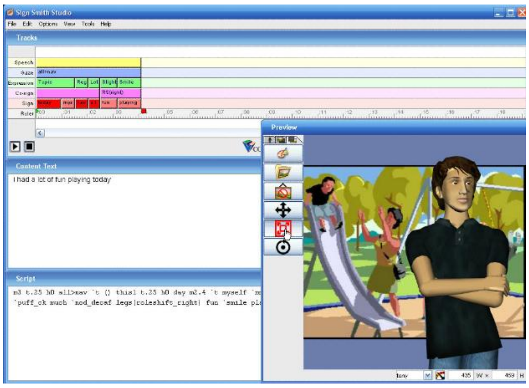


Welcome to the Signtel Interpreter

Figure 13. SignTel application interface

[14] http://www.signtelinc.com
[15] http://www.asl-dictionary.com

- Sign Smith Studio[16], from VCom3D, is an authoring tool that uses SigningAvatar characters for translating text into ASL. This software is intended to create mainly web contents, generating videos which could be embedded in any web page. In order to do that it has a database with more than 10.000 words. Furthermore, they have launched an application ("Sign 4 Me"), based on SigningAvatar engine, which offers English to ASL translation for iPhone devices[17].



I had a lot of fun playing today

```
m3 t.25 h0 all>nav `t () this1 t.25 h0 day m2.4 `t myself `m
`puff_ck much `nod_decaf legs|roleshift_right| fun `smile pl
```

Figure 14. Sign Smith Studio authoring tool interface

- Sys Consulting[18] is a British company which was created from the expertise of School of Computing Sciences, University of East Anglia (Norwich). The offered products related with avatars are quite innovative, between them: "Performing hands" that is a Shockwave plugin to develop the literacy of deaf children. The software includes video stories and educational games designed to help deaf children build grammatical sentences and create stories - in both British Sign Language (BSL) and English; and also, "Say It Sing It", in collaboration with IBM and RNID (Royal National Institute for Deaf People), that automatically converts the spoken word into British Sign Language (BSL) which is then signed by an animated digital character or avatar.

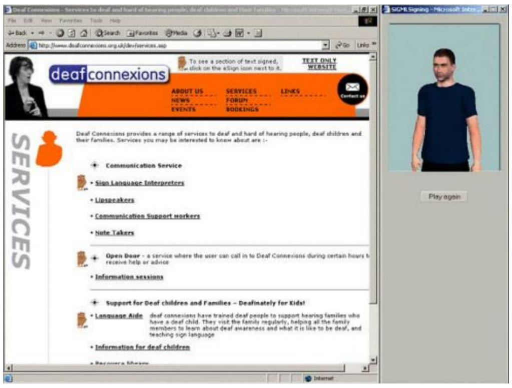
However the variety of industrial applications most of the most relevant work is being carried out in the research community. Next we present some of the institutions which are or have been working in these issues:

- ViSiCAST[19] (Virtual Signing: Capture, Animation, Storage and Transmission) is a European project which started in the year 2000. The goal of the project consists in make it possible Post Office workers could communicate with deaf people. Partner's project, including Institute of German Sign Language and Communication of the Deaf from the University of Hamburg, School of Information Systems from University of East Anglia and Televirtual, designed and implemented a virtual signer called TESSA ("TExt and Sign Support Assistant"). However for the creation of the signing movements was necessary a very complex and expensive motion capture process. To avoid it a new project was intended, eSIGN.

- eSIGN[20] (Essential Sign Language Information on Government Networks) is also a European project started from ViSiCAST results. The basic animation and rendering technology supplied by Televirtual in ViSiCAST was augmented by a piece of software, developed at University of East Anglia, which interpreted a sign definition language. The definition language was SiGML which is based on the HamNoSys notation. Through these improvements the avatar could

[16] http://www.vcom3d.com
[17] http://www.signingapp.com/sign4me_desktop.html
[18] http://www.sys-consulting.co.uk
[19] http://www.visicast.cmp.uea.ac.uk
[20] http://www.sign-lang.uni-hamburg.de/esign

generate synthetically the sequence of signs and be embedded into a web page in the form of an ActiveX control. Thanks to the success of eSIGN project, a new funded project in the European framework was launched in 2009 and it is active yet. This project is called DictaSign[21] and it takes advantage of the experience acquired in past projects in order to develop the necessary technologies that make Web 2.0 interactions in sign language possible. Thus, the computer recognizes the signed phrases, converts them into an internal representation of sign language, and then has an animated avatar sign them back to the users.



Deaf Connexions provides a range of services to deaf and hard of hearing people, deaf children and their families. Services you may be interested to know about are :-

- Communication Service
- Sign Language Interpreters
- Lipspeakers
- Communication Support workers
- Note Takers
- Open Door - a service where the user can call in to Deaf Connexions during certain hours to receive help or advice
- Information sessions
- Support for Deaf children and Families – Deafinately for Kids!
- Language Aide deaf connexions have trained deaf people to support hearing families who have a deaf child. They visit the family regularly, helping all the family members to learn about deaf awareness and what it is like to be deaf, and teaching sign language
- Information for deaf children

Figure 15. Example of a web page providing sign language information using VGuido avatar

- DePaul ASL Project from DePaul University in Chicago has the goal of translating English to American Sign Language. To this end they have implemented a virtual agent with excellent capabilities to reproduce finger movements. They have used it in several applications as for example, supporting translation tasks in security areas at airports in order to inform deaf people about security procedures (Furst, et al. 2002) or teaching ASL to hearing people through an avatar called PAULA (Davidson 2006).

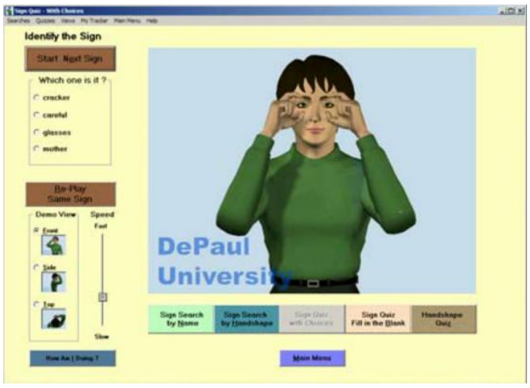


Figure 16. Sign language tutor PAULA interface

- SignSynth[22] is a sign language synthesis application developed by Angus Grieve-Smith from the University of New Mexico. Although the aesthetic of the signing avatar isn't realistic, this proposal is becoming worthy since the application is intended to run in future web browsers. The result is a three-dimensional animation sequence in Virtual Reality Modeling Language (VRML or Web3D), which is automatically rendered by the browser.

[21] http://www.dictasign.eu

[22] http://www.panix.com/~grvsmth/signsynth

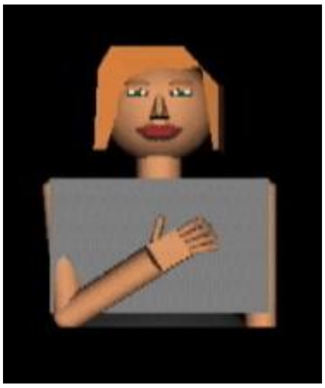


Figure 17. SignSynth avatar

- We can also find TEAM ("Translation from English to ASL by Machine") project from University of Pennsylvania. The system employs rules to build an ASL syntactic structure while an English dependency tree was built during analysis. In order to represent the signs it was employed an avatar which takes a special care in the smooth transition of movements (Zhao, et al. 2000).

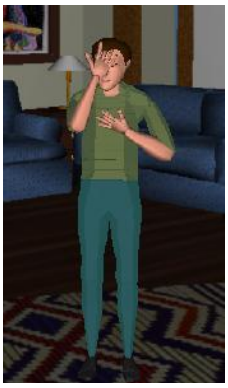


Figure 18. TEAM project avatar

Furthermore there are various initiatives which are focused in specific languages or special features of sign language. For example, a translation system from English to Irish Sign Language (ISL) (Morrissey and Way 2007); from English to South African Sign Language (Zijl and Barker 2003); from Thai to Thai Sign Language (Dangsaart and Naruedomkul 2007); from Spanish to Spanish Sign Language (Baldassarri, Cerezo and Royo-Santas 2009), etc.

In the expert survey, transcribed in Appendix A. Expert survey, drawbacks and usefulness of virtual signers were asked. The main conclusions, related to virtual signers' drawbacks, were:

- Avatars are generally "stiff". They lack facial expressions, mouth movements, nodding, etc. These capabilities are important for a complete and easy understanding of sign language.
- Not fluent signing. Avatars sign slowly and that makes their signing turn out unnatural.
- Taking into account 3D space for signing is necessary.
- Virtual signers are not suitable for expressing complex information. They could be used for simple messages as train timetables or similarly simple information.
- In the case an avatar solution would be adopted it would imply an intensive use of it. Then a lack of options to configure the avatars' style (clothes, characters, etc.) could reduce the usability of them.

About the potential usefulness which was observed by survey experts, the following issues are noteworthy:

- Human interpreters are scarce and expensive unlike virtual signers.
- Avatars are accessible 24 hours a day, 7 days a week.

In summary the generalized opinion of experts is that they prefer a real person instead of an avatar as interpreter. Anyway the confidence about the possible improvement of the technology is really high. Several of our experts named virtual signers from a promotional video showed in WFD Congress 2007 at Barcelona as a good starting point in this sense (see Figure 19 or promotional video[23]).

Figure 19. Virtual agent showed in a promotional video for 15th WFD World Congress

3.1.4. Sign Language to Text

The most challenging technology included in the communication bridge proposed by SignSpeak is the translation of Sign Language to text. That is, capture the movements, expressions and emotions of the signers; identify the signs from the extracted features, and then translate the sequence of them to natural language in order to obtain a message understandable by hearing users.

The means used to capture hand movements can be classified mainly in two groups: instrumented and video-based. Traditionally, for instrumented approaches, it has been used special gloves for capturing the movements when users are signing. Ryan Patterson (Dean 2002) designed and developed a prototype of sign language translator using a glove with 10 sensors. This 17-year-old's award-winning design involves "...sensing the hand movements of the sign language alphabet, then wirelessly transmitting the data to a portable device that displays the text on-screen".

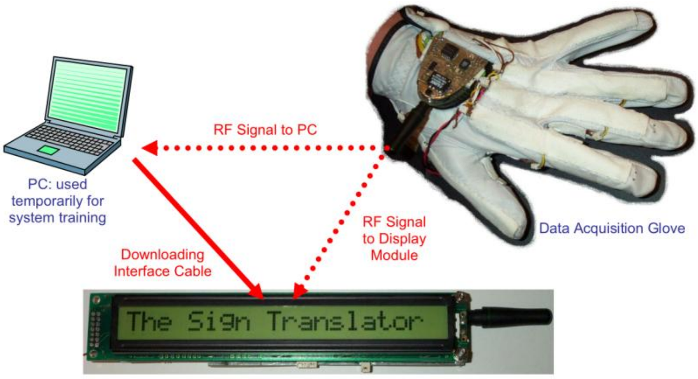

RF Signal to PC
PC: used temporarily for system training
RF Signal to Display Module
Data Acquisition Glove
Downloading Interface Cable
The Sign Translator

Figure 20. Sign language translator designed by Ryan Patterson

[23] http://www.youtube.com/watch?v=bhjvpcGUDTo

As it happens in speech recognition, users should be trained in the system, although the process is intended to be quick. The capabilities of the system are limited since it is only prepared to translate the alphabet. Although, through the training, more complex sequence of movements could be used. There are other examples of capturing systems based on gloves. For example "CyberGlove III"[24] has 18 or 22 sensors and can be connected to the host computer through wireless. They offer also a SDK to analyse and process the movements captured by the glove. "AcceleGlove"[25] is a light-weight glove with seamlessly integrated sensors and an USB connector. It has been used for developing an application ("AcceleSpell") that recognizes the hand shapes of American Sign Language. This device has also been employed with a two-link arm skeleton that detects hand location and movement with respect to the body in other promising experiments as (Hernandez-Rebollar, Kyriakopoulos and Lindeman 2004).

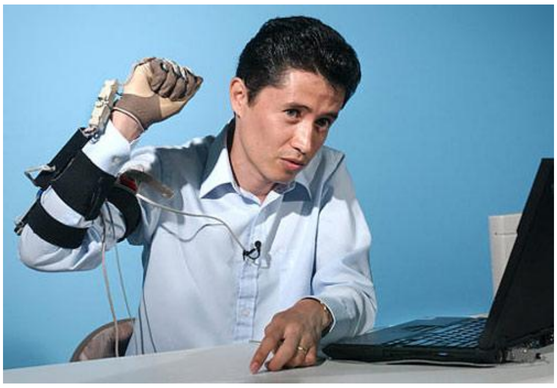


Figure 21. Dr. Hernández-Rebollar showing the set consisted of AcceleGlove and an arm skeleton

Another set of techniques to capture the signs involves using video cameras. The video-based approaches allow the users to avoid any instrumentation or extra device, beyond the camera. The hand positions and trajectory are detected and tracked just processing the images provided by the camera. An early work was presented in (Starner, Weaver and Pentland 1998), they developed a system for recognizing sentence-level continuous American Sign Language (with a 40 word lexicon). The camera could be placed in a desk, obtaining a word accuracy rate of 92%, or mounted in a cap worn by the signer, and thus improving the rate till 98%. Recently, the same researcher group from Georgia Institute of Technology has proposed a multiple sensor approach for disambiguation of noise in gesture recognition (Brashear, Starner, et al. 2003). They use a camera and accelerometers, placed on the users' wrists and torso, with three degrees of freedom. The accelerometers capture information that the camera is not able to, as rotations or some vertical movements. Applying the same concept to the educational field has produced fruitful results (Weaver, et al. 2010). This work describes CopyCat (Brashear, Henderson, et al. 2006) that is an American Sign Language (ASL) game intended to help young deaf children practice ASL skills. In order to capture the gestures the students wear funny coloured gloves with accelerometers mounted close to their wrists. Their most recent research line is going to integrate the Kinect sensor capabilities so no gloves or accelerometers would be necessary[26].

Up to now most of the video-based works described use some elements (accelerometers, gloves, etc.) to facilitate the image processing. In Sign2 project a new conversion system was developed to translate ASL to written and spoken English (Glenn, et al. 2005). It is based on image processing using a set of default "points" set all over the left and right hands (called Points-of-Digital Articulation) which are processed and compared with a database of different letters of the ASL. Other related studies have been carried out during the last years: associated to 5th FWP project WISDOM a vision-based sign

[24] http://www.cyberglovesystems.com

[25] http://www.acceleglove.com

[26] http://www.youtube.com/watch?v=qFH5rSzmgFE (demo video)

language recognition system for mobile use was developed. It was based on a skin color model complemented with pixel level motion information (Akyol and Alvarado 2001); an automatic fingerspelling recognition system which tries to solve finger occlusion problems using a multi-flash camera which avoids unwanted shadows (Feris, et al. 2004); a gesture recognition system able to recognize 46 gestures (ASL alphabet, ASL digits and some typical mouse movements) which only use a wristband as a reference (Lockton and Fitzgibbon 2002); or a system for the simultaneous detection, based on color gradients, of faces and 45 Japanese Sign Language gestures (Terrillon, et al. 2002).

Both approaches, video-based and instrumented, have pros and cons. For instrumented proposals, gloves are usually complemented with other devices, as accelerometers. It means users have to remain close to the radiant source, in the case of a wireless connection, or close and physically tethered to the computer in the case of a wired one. Furthermore, current glove technology is not intended for daily use; the gloves deteriorate quickly with extended use and output becomes increasingly noisy as they break down. In the other hand this kind of solution uses to be more reliable, overall against ambient noise or other adverse background conditions. In video-based approaches, the signer avoids having attached to hers/his body any instrumentation. However the working conditions should be controlled and the amount of data obtained, compared with instrumented systems, is lower.

SignSpeak project wants to go beyond most of the limitations which we have presented. It has been faced from a global planning which implies advances in several research fields and industrial knowledge in order to transfer the technology to daily life of deaf community. As it is showed in Figure 22, the development of the project includes several tasks:

- Creation and improvement of video corpora and their annotations. A complete description of the work is made in public deliverable D7.1. Basically Corpus-NGT (Sign Language of the Netherlands) and PHOENIX (German Sign Language, DGS) are being expanded.
- Scientific understanding of sign languages. The goal is to collect linguistic knowledge on the phonetics and phonology of sign languages, and in particular for Sign Language of Netherlands. It could improve the efficiency of the annotations and shed light on the recognition and translation of signs.
- A new approach to sign language recognition. The main objective is to develop a system which performs isolated and continuous sign language recognition. The core technology is based on Automatic Speech Recognition techniques which have been deeply studied by consortium partners (Stein, et al. 2010).
- Sign language translation to text. That is an automatic machine translation of multimodal input from recognized signs transcribed in gloss notation into a spoken language. The lexical knowledge is being incorporated through morphological analyser Morphisto[27].
- System integration and industrial uses. Parallel work is being carried out for a complete integration of the aforementioned modules. Furthermore, an industrial prospection of the technology obtained from SignSpeak project has been performed. Telefónica I+D's experience in the deployment of this kind of applications and the end user's view from European Union of the Deaf are crucial at this point.

[27] http://code.google.com/p/morphisto

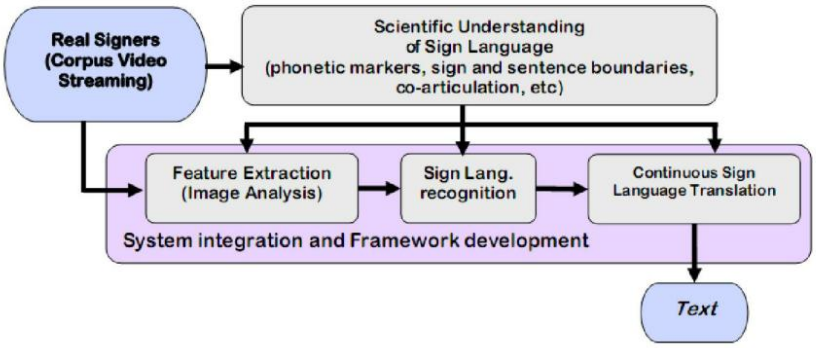

Real Signers (Corpus Video Streaming)
Scientific Understanding of Sign Language (phonetic markers, sign and sentence boundaries, co-articulation, etc)
Feature Extraction (Image Analysis)
Sign Lang. recognition
Continuous Sign Language Translation
System integration and Framework development
Text

Figure 22. Conceptual scheme of the work planned in SignSpeak project

SignSpeak has been developed using different video databases aiming at showing how SignSpeak works by handling databases with different features (i.e. different Signed Languages corresponding to German, Dutch, British, American and Irish), context domain, vocabulary size and recording conditions.

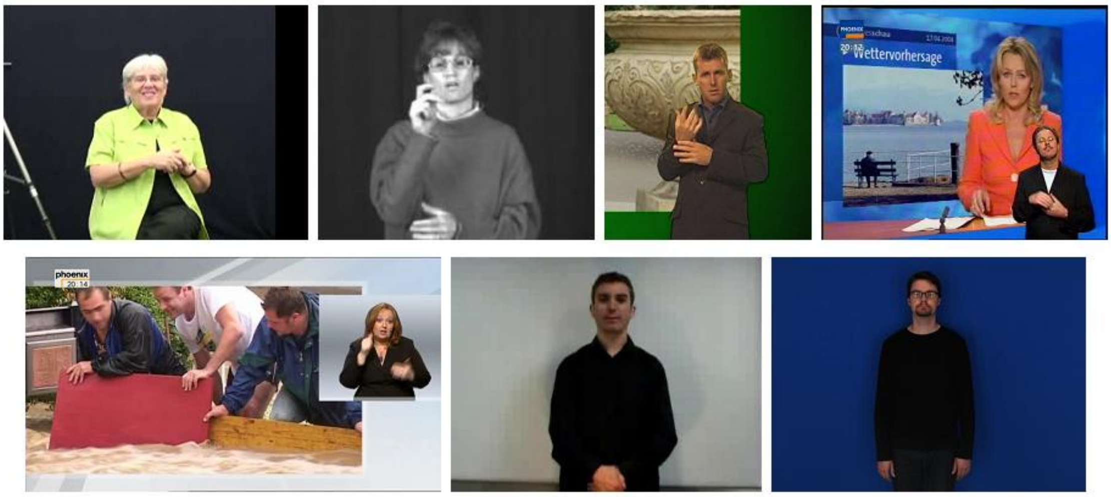


Figure 23. Sample images from different video-based sign language corpora (FLTR): Corpus-NGT, RWTH-BOSTON, OXFORD, RWTH-PHOENIX v1.0 and RWTH-PHOENIX v2.0, ATIS-ISL, SIGNUM

As of the writing of this document SignSpeak has had its best results working with PHOENIX database. This database is being created by one of the partners in the consortium (RWTH Aachen University); in total the database features eight (8) different signers in controlled (TV-studio) conditions performing simultaneous interpretation of German weather forecasts. This controlled context domain implies a smaller vocabulary size and higher repetition of signs (average type-token-ratio) than for the other databases, explaining thus why SignSpeak works better for this database. Additionally, that demonstrates SignSpeak can work with other sign languages and other context domains if adequate and sufficient data (videos and annotations) are provided.

3.2. Relevant factors in SignSpeak

The SignSpeak project is intended to be a first step to achieve a sophisticated technology able to complete the communication bridge between hearing and deaf community. In this preliminary stage the demands about the performance of the technology should be ambitious but bearing in the mind the possible problems which could arise in a realistic scenario.

In the previous sections we have presented the current state of the technologies involved in the communication bridge. So, we have described the basic principles of each technology and the operation conditions where the technologies could diminish their performance.

Following this brief introduction we consider the whole communication framework (TTS, ASR, Text-to-Sign, and SignSpeak) and we discuss about the different factors which are relevant to its proper operation.

3.2.1. User factors

User factors are individual differences that include demographic variables and situational variables that account for differences attributable to circumstances such as experience and training (Agarwal and Prasad 1998). Most of the studies about the performance and acceptance of new technologies don't take into consideration the possible influence of these individual factors, although most of them also admit it turns to certain limitations in the generality of their results. In fact, it is usually the cause of inconsistencies between research works regarding the same issues (V. Venkatesh, M. G. Morris, et al. 2003).

Below a list of user factors which could be relevant to SignSpeak technology is given. We will consider some factors which are important for all technology products and others specific to the context of SignSpeak (some of them were briefly introduced in section 3.1).

Gender

Research has shown that there are differences between men and women regarding the cognitive structures employed during the interaction with technology products (Venkatesh and Morris 2000). For example, in (Meyers-Levy and Maheswaran 1991), it is notice how female and male use different processing strategies for the recognition of advertisements.

In a general way, three major differences have been found for technology acceptance, namely:

- Men have a more pragmatic behavior than women. As it is showed in (Minton and Schneider 1984) men are task-oriented and their motivation lays on achievement needs, what is much related with the concept of usefulness. However, acceptance models are not always so simple and cross-interactions have been found between age and gender (Morris, Venkatesh and Ackerman 2005).

- Women experiment higher computer anxiety and lower computer self-efficacy (Venkatesh and Morris 2000). Thus, ease of use is perceived differently. Anyway, the influence of other multiple factors as cultural level, computer expertise and so on makes the results as tentative in most of the cases.

- Women have a greater capacity of feeling empathy compared to men and thus, they are more easily influenced by others (Venkatesh and Morris 2000). So maybe the opinion of certain people about a technology could affect to women more than men.

Regarding the technologies mentioned in section 3.1, the influence of the gender varies. For *text-to-speech* there aren't any evidences about differences between females and males related to listening synthetic voices. However there are several studies which have showed that the gender of the synthetic voice could trigger some stereotypic responses (Nass, Moon and Green 1997) or even affect intelligibility of TTS engines, generally being male voices more intelligible than female voices (Stevens, et al. 2005). For *automatic speech recognizers*, the influence is closely linked to physiological properties. The shape of the vocal tract determines the unique features corresponding to each speaker. The glottis at the larynx is the source for voiced phonemes and shapes the speech signal in a speaker characteristic way. Thus, the length of male and female vocal tracts is a key factor for the accuracy of ASR systems. Actually, in (Abdulla, Kasabov and Zealand 2001) they use the pitch information to build gender dependent models for recognition. For *text-to-sign language*, that is, signing avatars, gender influences since even experienced computer users interact with their computers according to social rules of politeness and gender stereotypes (Cassell 2000). Unfortunately there are no studies about how the gender of signing avatars could influence in the perception of the technology by deaf community. In (Bailenson and Yee 2005) is showed how users prefer avatars which are similar to them and mimic their behavior. Thus it could be suggested that signing avatars try to mimic the signing style of the users or even adopt the users' gender. For *SignSpeak technology* there are no differences of performance depending on the users' gender, at least from a technological point of view. Taking into account specific aspects of sign language it has been reported differences in vocabulary in older population of Irish Sign Language (Janzen 2005). Above all it was due to social differences and nowadays this situation is hardly applicable.

Intellectual capabilities

The individual competence for the assessment of technologies is a factor that has received some attention. In (Lederer, et al. 2000) it is noticed how the ease of use and usefulness of Web sites could vary significantly if users were "less educated". Regarding this the concept of "profession" usually is associated, considering it an indicator of the users' mental skills or competence (Chau and Hu 2002).

Naturally the perception of the communication bridge by hearing and deaf community will be influenced by the specific capabilities of each individual, since it seems to affect the ease of use and usefulness of a technology. However, to the best of our knowledge there aren't studies related to the referred technologies. Maybe this factor could be reflected into others as for example it would occur when a well-educated user, with a better diction, could get better word accuracy rates in speech recognition tasks.

Experience

Experience is defined differently in several studies. Thus, experience is measured by the time (number of years) a user has been in touch with computers (Venkatesh and Morris 2000) or by means of an ordinal value (i.e. 0, 1, 2…) which tries to reflect the user experience with technology (V. Venkatesh, M. G. Morris, et al. 2003). In other studies the time is measured from the point users are starting to interact with the technology which is under study (i.e. 1 hour after introduction or 14 weeks after introduction) (Davis, Bagozzi et al. 1989). Anyway it seems necessary to deal with a better and a domain specific conceptualization of experience.

Generally speaking, a base of knowledge about technology increase the possibility that new interaction experiences will be related to what is already known and therefore may perceive that ease of use is not a big issue (Cohen and Levinthal 1990). Regarding *TTS technology*, in (Lai, Wood and Considine 2000) a study is carried which compares the comprehension of TTS systems vs. human voice based on several working conditions. They conclude that "subjects who had cause to listen to TTS with some regularity did better for the synthetic voice than for human, as well as doing better than those subjects who had less exposure". For *automatic speech recognition* two main effects have been studied. In (Karat, et al. 2000) is described an experiment where the ASR performance is worst for novice users than expert ones. Furthermore, the latter group of users is more effective carrying out the corrections when the system fails. Besides that in (Koester 2004) they make notice that experience users' expectations are closer to the real performance of an ASR system, and thus it is more difficult they will get disappointed. A similar effect is observed for *virtual agents* acting as interfaces, users who don't have previous experience use to have very high expectations about expected performance of a system (Krämer, Bente and Piesk 2003). In case of *SignSpeak technology* no evidences have been observed about how the experience using the system could favor a better performance of sign language recognizers.

Age

From the point of view of acceptance of technology age is recognized as a key factor. Specially, senior users, who don't have usually great experience with technology and have age-related problems with cognitive abilities, face difficulties understanding and interacting with technological devices (Ziefle and Bay 2008). On the contrary, older users are more inclined to accept technologies when the usefulness is clear and there is a good support of the system (tutorials, help system, etc.) (Arning and Ziefle 2010). Related to the age, three are different interpretations about the usefulness or the acceptance of the technology depending on the context of use (eHealth, leisure, etc.).

Regarding the technologies involved in the SignSpeak communication bridge: for *TTS* the prior studies which contribute to the relation between age and intelligibility are about natural speech. They have demonstrated the existence of negative effects in this sense. For synthesized voice, in (R. and Reichle 2001), they show an improvement of the intelligibility of synthesized sentences for senior as well as young users when a context in discourse is given. However it never reaches younger users' word accuracy (Roring, Hines and Charness 2007). It's important to notice that reductions in synthetic speech perception could be due to sensory impairments associated with advanced age (Baltes and Lindenberger 1997). A similar problem happens if levels of intelligibility are compared between adults and children (Drager, Reichle and Pinkoski 2010); for *ASR*, a typical case is the children's speech recognition. Their vocal characteristics vary significantly from those of adults, and also sentence structures and vocabulary are quite different to adult patterns (Blomberg and Elenius 2003). Since ASR systems are usually modeled with corpora comprising mainly adult speech, the recognition accuracy degrades and specific approaches are needed (Das, Nix and Picheny 1998). Even the error resolution strategies taken by adults and children are usually different: children tend to repeat the same utterance, altering certain phonetic features and adults modify other aspects of their utterances such as lexicon and syntax (Bell and Gustafson 2003); for *text-to-sign language,* recalling the work presented in (Bailenson and Yee 2005), a good approach would be to present a virtual agent with similar age to the user. In fact there are several applications, related to learning with children, which use the concept of 'virtual peer' to interact with students (Cassell, Tartaro, et al. 2007). For *SignSpeak technology* the age could affect in a combine way with gender. As it was explained in Gender section the older generation of deaf people suffers from several differences in lexicon depending on the gender. Also the number of signs that older and young deaf people use, are different due to educational background (older people usually have experienced a lot of lipreading and speaking at school, so they sign less).

Cultural background

Another factor given relatively little attention is cultural background. However the globalization of technological markets makes critical the understanding of cultural influence. A wide accepted definition of culture is "the collective programming of the mind which distinguishes the members of one group or category of people from another" (G. Hofstede 1984). Hofstede also suggested four dimensions which provides a theoretical framework, namely: power distance (degree of inequality among people which the population of a culture considers normal), individualism/collectivism (degree to which people in a culture prefer to act as individuals rather than as members of groups), masculinity/femininity (degree to which values like assertiveness, performance, success, and competition prevail among people of a culture over gentler values like the quality of life, maintaining warm personal relationships, service, care for the weak, etc.) and uncertainty avoidance (degree to which people in a culture feel uncomfortable with uncertainty and ambiguity).

In relation to *TTS* technology, the differences in the interpretation (for different cultures) of human speech features, as rhythm, intonation or emotions, leads to the hypothesis that synthetic speech suffers the same effect (Scherer 2000). Regarding the specific aspect of emotions, the experiment presented in (Burkhardt, et al. 2006)showed evidences about the necessity to use different emotional models for the creation of emotions in synthetic speech, although some of the implemented emotions could be considered as cross cultural. For *ASR* engines the problematic issue is quite clear: the accent. In (Huang, et al. 2001) the accent was identified as one of the principal components of speech variation. Indeed, a degradation of the recognition performance has been observed for the task of recognizing accented speech and non-native speech (Kubala, et al. 1994, Lawson, Harris and Grieco 2003). For *virtual signers* is important to be aware of the fact that some behaviors are interpreted differently across cultures is the first step to avoid inter-cultural misunderstandings and with it, to learn intercultural communication (Hofstede, Hofstede and Hofstede 2005).

And more

Many more user factors could affect the acceptance of a new communication paradigm (i.e. SignSpeak's communication bridge) and the technologies involved (TTS, ASR, virtual signers and Sign Language to text). In the previous sections those most important have been covered. However let us cite briefly some more:

- Level of signers' expressiveness could affect to the identification of signs.
- Users' emotional state could affect to the way they express themselves. It influences the expressiveness conveys to the sign language and, in case of hearing community, how the speech is pronounced.
- Physiology should be taken into account because it could affect the speech production or any limitation of the signer recognizer.

The impact of those factors on the corresponding technologies has been lightly studied and more efforts are necessary to gain knowledge into them.

A thorough review of how TTS performance could be affected by numerous users' factors is given in (Winters and Pisoni 2004). In case of ASR technology a relevant reference is (Benzeghiba, et al. 2007).

3.2.2. Environmental factors

The conditions relative to the context where the interaction is performed are collectively called environmental factors. They include numerous variables as weather conditions (i.e. lighting), noise conditions (i.e. "the cocktail party effect") or location conditions (i.e. mobility, in-car scenario…).

In the case of TTS engines arguably the most harmful effect is that posed by noisy environments. Traditionally, intelligibility of synthetic speech is tested under lab conditions in what is a really ideal performance and clearly does not reflect real-world conditions. In (Bregman 1994, Koul and Allen 1993) the intelligibility of TTS systems are evaluated in presence of different kind of noises. They have found that in these conditions people understand better natural speech than synthetic speech.

For ASR technology the most counterproductive factor is also the noise. Sources of noise are numerous. For example, additive noise from machines, reverberations produced inside a room or different spectral features of microphones. Thus, the performance of a speech recognizer could degrade from 100% accuracy in controlled conditions to 30% for in-car scenarios (driving at 90 km/h) (Lockwood and Boudy 1992). Likewise, the 1% error rate of a system trained under quiet conditions increases to 50% in a cafeteria environment (Das, Bakis, et al. 1993). Furthermore, in presence of noise, users don't speak normally, trying to increase the volume of their voice or articulating in a different way they did for the training. This phenomenon is called the Lombard effect (Junqua, Fincke and Field 1999) and may cause serious differences between the training and recognition processes. In this case, the performance uses to get worse, even if the recognition task is carried out under better conditions than training (Das, Bakis, et al. 1993). A complete review of this issue can be found in (Gong 1995).

Regarding virtual signers, taking into account that deaf users should be looking with attention to the virtual agent, the cognitive load that the environment demands has to be taken into consideration. In order to illustrate this let's imagine an application which is designed for interacting through a tactile interface and at the same time, it uses a virtual signer for communicating the information. Then it is necessary to set the message of virtual signer in such a way it does not coincide with any other visual message or it does not require a high cognitive load interaction simultaneously. For Automatic Sign Language Recognition most of the methods assume controlled environmental conditions, e.g. simple backgrounds or good lighting. For *SignSpeak* there are three main tasks related to the multimodal visual analysis: tracking of hand positions, facial analysis and body pose estimation. All of them need robust tracking algorithms since they should avoid the effect of i.e. signing hands moving in front of the face, or signing hands crossing the other hands. Unfortunately, research work from SignSpeak project, has not yet succeeded in being independent of environmental conditions. Two main constraints have been accepted: firstly, a specific video camera has to be used as video source of the tracking and recognition algorithms and, secondly, the corpus which has been used for training the models is weather forecasting corpus RWTH-Phoenix (Stein, et al. 2010). Due to that the sign language recognition task is closed to weather forecasting context.

3.2.3. Resource-related factors

Whenever a service which uses technologies as ASR or TTS is planned, it is necessary to analyze factors as: what is the availability of computational power/memory space for the devices or what is going to be the quality of communication. Thus, these resources influence the selection of a concrete technique for technologies, the use of a concrete device (desktop environment vs. portable device) or in a worst case scenario, degrading the quality of the user experience.

In regards to text-to-speech technology the challenge consists on developing engines with a low computational cost and limited memory (suitable for embedded devices). Unfortunately, TTS techniques rely usually on a high-load runtime selection and a compilation of speech units from a large speech database (see Section 3.1.1). Some of the approaches to build low computational cost TTS systems reduce the number of speech units to be selected, at the expense of degrading the speech quality (Schnell, et al. 2002, Nukaga, et al. 2006). Recently, speech synthesizers based on Hidden Markov Models (HMM-based TTS) have been proposed (Kim, Kim and Hahn 2006). They represent the spectrum, the excitation, and the duration of the context-dependent speech units through HMMs, thus lessening the computational requirements and allowing them to be adopted by portable devices.

Likewise, ASR suffers the same problems derived from the limitation of resources as memory capacity or computational power. Furthermore, the progress of speech recognition technology has been based on the continuously increasing power of computers. In order to overcome these difficulties two research lines could be identified. In first place some works have tried to develop light ASR engines suitable to be embedded in low resources devices, implementing algorithms computationally efficient (i.e. through fixed-point arithmetic) (Varga, et al. 2002, Novak 2004). Other works try to move the most costly processes to a remote server or infrastructure (i.e. cloud computing) and then, the acoustic features (distributed speech recognizers) or the speech signal (network speech recognition), are sent to this remote platforms to be processed (Ion and Haeb-Umbach 2008, Kiss 2000).

Each approach has pros and cons. Embedded speech recognizers consume a large amount of computational resources and consequently energy, both of which are scarce in some devices. However, they get independency from network connectivity and the possible distortion generated by transmission. These latter problems are suffered by distributed and network solutions.

Regarding virtual signers the requirements for a smooth performance are also very demanding. The rendering and animation engines, mostly due to the high agility agents need for signing, use to demand great graphical resources and most of the times they only work properly on powerful devices. Besides implementations are sometimes implemented using proprietary frameworks such as Microsoft ActiveX or Adobe Flash, what reduces enormously the catalogue of operative systems and devices which are eligible. New approaches are needed which support the use of virtual agents (and signers) in widely use operative systems as Android and maybe, the best way is making use jointly of web technologies (i.e. HTML5) and graphical libraries (i.e. WebGL). "*Three.js*"[28] is an example of it.

SignSpeak technology has very demanding requisites regarding computational power. Its flow network implies several stages with certain complexity. Due to that there is a factor around 20 compared to realtime (for example, the translation of 6 seconds of video will take around 2 minutes) for testing data from the same domain as the data used to train the system.

3.2.4. Dependence on the capture devices

As it was noticed in Section 3.2.2, performance of conventional speech recognizers is not stable when conditions of test environment (i.e. noise) do not match with those of training environment. The same concept is applied for channel variability, that is, when microphone changes. In (Acero and Stern 1990)the recognition accuracy of a speech recognition system (SPHINX) dropped from 85% to 20% when they replaced a professional close-talking microphone used in training by an omnidirectional microphone. Some studies address these important issues and propose algorithms which try to deal with different acoustical conditions (Stern, et al. 1992; Acero 1990).

For *SignSpeak technology* occurs something similar to speech-to-text systems and channel variability concepts could be applied likewise. In this case there is no microphone but a video camera.

3.2.5. User perception and acceptance of the technology

User acceptance of a new technology does not depend exclusively on its technical functionality. User perception of a new technology is built from a set of psychological, social and contextual factors that are related to its use in everyday life applications. In Section 3.2.1 we have given some information about how user factors could affect to the acceptance of technology, however complete models from different perspectives and at various levels have been developed to explain IT acceptance perceptions and behaviours in the last years.

The pioneer and the most widely accepted model is from (Davis 1989). The model consists of two factors (see Figure 24): perceived usefulness and perceived ease of use. Perceived usefulness is the tendency to use or not to use an application since people believe it will help his or her job performance i.e., by reducing the time to accomplish a task or providing timely information. Perceived ease of use is the degree to which a person believes that using a particular system would be too hard, even taking into account the benefits it carries (Davis 1989). However the simplicity of the model has raised doubts about the validity and utility of this theory (Bagozzi, Davis and Warshaw 1992). In reply to it other models have been developed: the Technology-to-Performance Chain Model (TPC) (Goodhue and Thompson 1995)which is based on the idea that a technology has a positive impact in users when it is utilized and it has a good fit with the tasks it supports; the Unified Theory of Acceptance and Use of Technology (UTAUT) Model (V. Venkatesh, M. G. Morris, et al. 2003) which was created with the aim of unifying the majority of existing acceptance models. UTAUT posits that four constructs play a relevant role as determinants of technology acceptance: performance expectancy, effort expectancy, social influence, and facilitating conditions.

[28] https://github.com/mrdoob/three.js

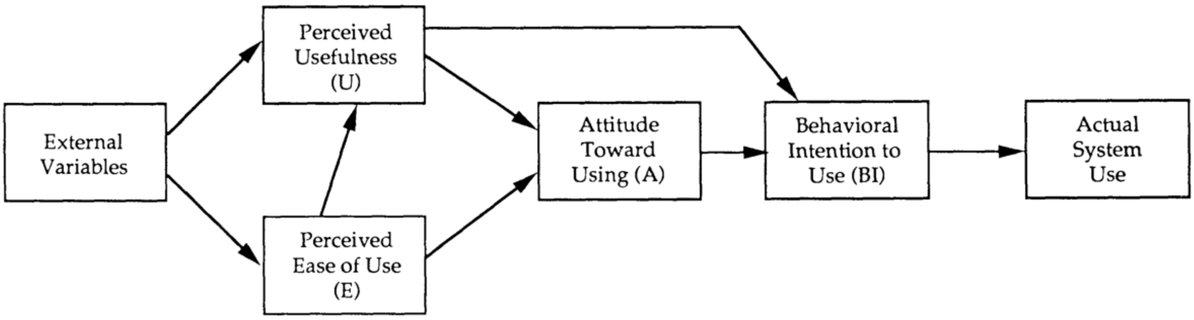

Perceived Usefulness (U)
External Variables
Attitude Toward Using (A)
Behavioral Intention to Use (BI)
Actual System Use
Perceived Ease of Use (E)

Figure 24. Technology Acceptance Model (TAM) (Davis, Bagozzi and Warshaw 1989**)**

4. Exploitation approaches

In this section we are going to propose some exploitation opportunities which could be explored using SignSpeak's technology. Our sources of inspiration have come from:

- TID's experience in the interpretation trial carried out in STEP Project (Gumiel, Serrano and Moya 2010) which has provided firsthand knowledge about how to deal with tele-interpretation scenarios and what are the main communication needs of Spanish deaf community.
- The participation in SignSpeak project which is letting us to establish contact with relevant agents inside the European deaf community and likewise, to access to leading technology for the translation of sign language to text.
- The joint work with European Union of Deaf in order to obtain very qualified opinions about the communication needs of the deaf community and the best approaches for SignSpeak's technology.

Next we will describe TID'S experience in STEP project, a centre for remote interpretation. Later, in Section 4.2, we propose several interesting scenarios for the exploitation of SignSpeak technology. They have been analysed with the collaboration of EUD and 5 experts selected to this task.

4.1. TID experience (STEP Project)

In Spain there are 2.781 Spanish Sign Language interpreters, among them 25.17% are active. According to this, in Spain, the ratio of sign language interpreters is one professional for every 143 people who are deaf or hearing impaired[29].

The deaf community is endowed with an associative structure with dense networks of relationships, organized around institutions and distinctive culture. Culture in the double sense of belief systems, and cultural productions such as narrative, storytelling, humour, sign language poetry, drama and mime, sculpture, painting, photography and films sensitive to the experiences of deaf people. It is a living community, varied and open to all sorts of people whose central element is sign language. In Spain, the Law 27/2007 (October 23rd) is focused on the "linguistic community of the people who use Spanish Sign Language". It recognises the Spanish Sign Language and regulates the support means for the communication of deaf and hearing impaired users (BOE, 2007).

[29] http://sid.usal.es/noticias/discapacidad/35003/1-1/es (in Spanish)

	Age				
	6-19	*20-44*	*45-64*	*65-79*	*80+*
Disability in receiving any sound	4.103 (0,6‰)	17.138 (1,1‰)	25.711 (3,0‰)	33.102 (6,5‰)	22.340 (16,2‰)
Disability in hearing loud sounds	4.948 (0,8‰)	24.088 (1,6‰)	36.870 (4,2‰)	92.788 (18,4‰)	72.042 (52,2‰)
Disability in hearing speech	17.584 (2,7‰)	67.993 (4,4‰)	148.317 (17,0‰)	341.169 (67,5‰)	240.576 (174,3‰)
Totals	**22.102 (3,4‰)**	**90.913 (5,9‰)**	**182.853 (21,0‰)**	**391.001 (77,4‰)**	**274.620 (199,0‰)**

Table 1. The disabilities of people based on their age in 1999[30]

In January 2009 an innovating project, STEP, was created by Telefónica in collaboration with the FAAS (Andalusian Federation of Associations of the Deaf) and promoted by the Andalusia local government. The goal of the project was the development of an interpretation service which supports the communication by phone between deaf and hearing people. The STEP project was motivated by the numerous communication barriers which arise in the access to some services of Public Administration that can be used by phone, such as request information, make emergency calls or request appointments.

As it is showed in Figure 1 the flow of communication starts when a deaf user wants to contact any of the abovementioned services. Then user makes a video call to the STEP interpretation centre and she/he requests to the interpreter to make a call to a public entity. Once the communication is established the interpreter translates to both parts, acting as a communication bridge.

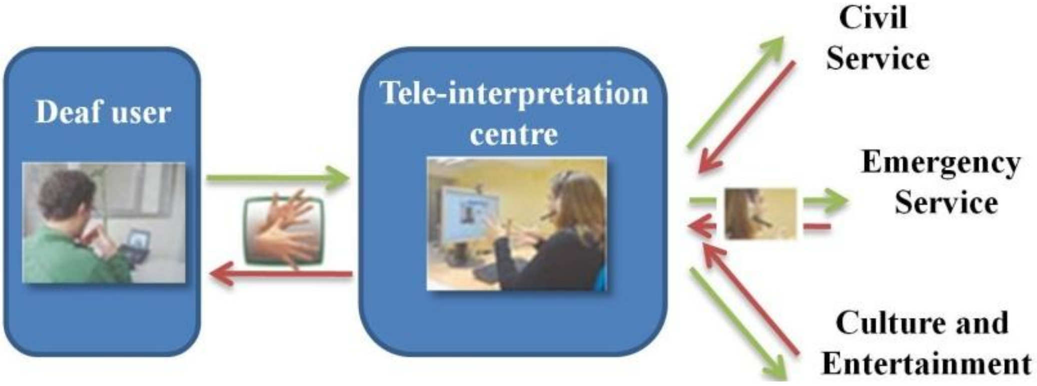


Figure 25. Interaction flow of STEP project service

The trial service has been working for 1 year, with 2 interpreters and around 20 deaf users making an average of 80 calls per month (see Figure 26).

[30] http://www.boe.es/boe/dias/2007/10/24/pdfs/A43251-43259.pdf (in Spanish)

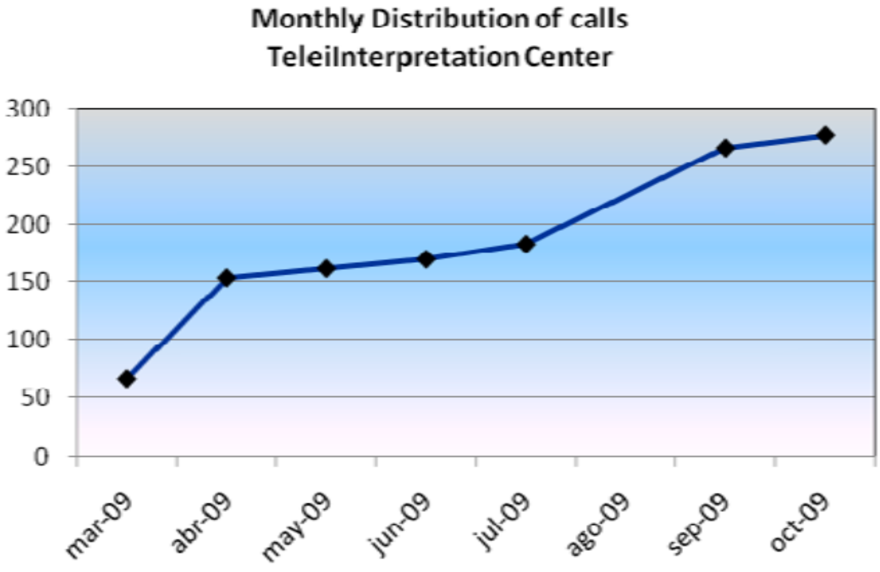


Figure 26. Distribution of call per month

In the beginning the availability of the service was limited to only during mornings. After the numerous requests of the users the timetable was extended also to the afternoon. It is worth to mention that during the trial users were permitted to make calls to other deaf people, without the intervention of an interpreter, just using the video conference capabilities of the service.

Between the most demanding services were requesting butane gas bottles, communicating with their lawyer or requesting the services of a plumber.

Thanks to the experience obtained during the first trial of this project, there are some general points which should be taken into account for future developments:

- There are not enough interpreters to attend the requirements of deaf users. During rush hours, there are almost always users waiting;
- For these busy situations where there isn't an available interpreter, it would be very advisable to have an automatic system for attending the calls;
- An automatic subtitling system wouldn't be a general solution since there are some deaf people which hardly understand reading.

4.2. Scenarios analysis

In the expert survey several scenarios were proposed to expert. These scenarios were created in collaboration with EUD taking into account the communication needs of deaf community and the expected usefulness. All of them started with the same motivation story that was:

> *"John and Mary are a deaf-hearing marriage and they have one child, Susan, who is 7 years old and she is also deaf. This family is bilingual; sign language and spoken language. They have hearing neighbours and family members who cannot sign very well."*

Once this situation was explained the rest of scenarios' details were concreted.

4.2.1. Sign language e-learning

This scenario consists on,

> *"A neighbor girl of Susan is following a course for improving her sign language skills. For doing this course, pupils have to connect to the teacher through Internet (using a webcam). Then, pupils see the teacher in their monitors and the teacher can see all the pupils at their own homes. The teacher gives the lessons using sign-language and, thanks to SignSpeak technology, text subtitles appear at the same time."*

This scenario was proposed in a first moment for SignSpeak project. However for our consulting experts, subtitles don't seem the most suitable mean of teaching sign language, even more when target population is so young. Some comments at this respect were: *"E-learning with subtitles is possible for teaching SL but why not use voice-over. Would a child of 7 years old understand subtitles easily?" ''Subtitles… I do not think they are suitable for children. I would use signing with voice-over. In fact the best option is that children socialize each other, in a direct way, to learn sign language. Children can learn sign language very quickly. They are flexible. They are natural learners."* Visual information (that is, to observe signers without subtitles) is considered the most suitable way to start learning sign language (*"beginners learn better with signing videos without subtitles and then they can watch signing videos with the subtitles to see if they already understand sign language").* Furthermore, in the opinion of some of our experts it should be mandatory to obtain recognition rates very close to 100% since the application is intended to teach. So, *"It must be a good translation, close to 100% success"* although they can imagine this goal is a *"challenge for SignSpeak. So patience needs to be part of this new development process"*

Despite all these comments most of them assess the integration of SignSpeak technology for e-learning as *"positive"* and *"ok, for basic skills"* although it could be used for other goals as "*help to explain why we make a sign in a concrete way. For example, in 'car' we use a driving wheel. Hearing people tend to ask why particular movements become to be a sign".*

4.2.2. VideoSL mail

This scenario consists on,

> *"Mary wants to send an email to several people. Some of them can hear while others can't. She records a video signing and she sends it. SignSpeak technology translates the sign language message into text and then it sends the email with the video and the text message to all the addressees."*

The scenario was planned as a parallelism with Google Voice[31] service. Among the features of Google Voice a transcription service was implemented for getting the text from voice messages. It is a very welcomed service because it lets to users to obtain a quick preview of the voice messages. Following this concept a similar scenario was proposed in order to translate a video with sign language to text. In this case the motivation is twofold. Firstly, it would give the possibility to a preview of the message to those deaf people who are comfortable reading text. Secondly, it would let hearing people (non-signers) to understand the message expressed in sign language.

From the view point of our experts this scenario was well received (*"good!"* or *"suitable"*). Even it was detected a possible learning application: *"Hearing people would learn sign language by reading the text. Text and sign language should be next to each other in the system".* However, again, doubts arise about how good the translation would be in this service ("*Would be the translation correct? Google translations aren't often correct. I would like to know if this translation would be done correctly")* or, even, if different grammars or languages could be supported (*"If signing is recognized then, it would be translated using the right grammar? If I sign with*

[31] http://www.google.com/googlevoice/about.html

different grammars, would the service selected the correct one for translation?").

4.2.3. Answering machine

This scenario consists on,

> *"John is in a congress and makes a video call to home. Nobody is at home, so he leaves a recorded video with his sign language message. The answering machine, through SignSpeak technology, translates the sign language message into text. When Mary arrives home, she realizes there are several messages. As she is busy, she decides listen the messages while preparing the dinner. She listens to her husband's message through a voice synthesizer."*

Video Relay Services are quite extended in deaf community. But what happens if the person you call isn't at home? You could record your video message signing and when the person is back at home she could watch the video. But, it would be interesting to have the chance of selecting the modality/language through the video information is presented? We proposed this situation to our experts.

However the general opinion wasn't so good about this service. Most of the experts considered the service as unnecessary taking into account Mary, the wife, understands sign language. In one of our experts' words who is a Child of Deaf Parents, and raising with sign language at home, *"I would prefer to see him directly signing instead of hearing the voice synthesizer. When I finish cooking, then I would watch the video, while eating! I would enjoy more with this";* Other expert told us: *"Mary should see his (from husband) signing directly. The message comprises a visual communication"; "I think that Mary could see sign language later. We have to take into account that SL could be her mother tongue".* From these opinions it seems we weren't able to express correctly the scenario. Maybe it would be better understood if the person to whom is directed the message has no ability to sign. Then the only option would be the translation of sign language into voice or text. Of course, then, voice or text could be place on the video in order not to lose the visual information (facial expressions, emotions, etc.). As one of the experts commented "*More options mean best services"* in concordance with the concept of 'total conversation'

4.2.4. Play Sign Language

This scenario consists on,

> *"Susan has a game console which includes a camera. She wants to play with her neighbour girl. They love to play an educative adventure game that makes you practice some sign language expressions. Using the video from the camera, SignSpeak technology assesses the quality/correctness of the signs and the game gives Susan feedback about how to improve her sign language abilities. As the neighbour girl gets better, she moves forward the levels of the game. They improve their communication very well through playing the game."*

A more informal and fun way to learn sign language could be through games. We proposed a scenario where SignSpeak recognition technology is used for assessing the quality of different signs which are performed by the players. It would make that, as it was observed before for our experts, hearing and deaf children could learn sign language socializing and enjoying at the same time.

The idea was received by our experts in a very positive way. They remark that *"it is a motivation for hearing people in order to learn sign language", "playing with sign language is the best way to learn it. If it is more formal as in the school, then children would get very bored".* However these games would be only suitable to be initiated in sign language. In case you need to go deeper in the knowledge of sign language *"they need to go to a SL School".*

The idea of games and gestures brings us to commercial systems as Wii or the most recent Kinect. These two gaming systems use the gestures for the interaction with the games. However, as far as our concern, no commercial initiatives have been proposed for these platforms. Although some research work is in progress[32]. Anyway, one of our experts notices this interesting research area and the obvious relation of SignSpeak project with the impulse of this kind of initiatives.

5. Conclusions

This work wants to be a study about the communication needs of deaf community (especially those who use sign language to communicate themselves) and also a preliminary step to the implementation of a complete prototype associated to the communication platform due to the end of the project (Work Package 8).

In Section 2 we have reviewed some of the communication needs suffered by deaf community and what are the current solutions they apply to solve these problems. We have also listed some of the most problematic situations regarding communication for deaf people, namely: telephoning hearing people through relay service in text, sending e-mails via text messages, going to public authorities/service (passport service, bank, etc.) where most people cannot sign*,* or relay service timetables almost never are 24 hours a day.

In Section 3 a comprehensive review of the most relevant factors which affect to SignSpeak performing is presented. Attending to this information it seems clear that the development of SignSpeak communication bridge is a complex task and should take into account a myriad of factors. At this point, we would like to stress that the use of SignSpeak components will give further details about the relevance of these factors.

Eventually the analysis of the experts' interviews has shed light on the adequacy of SignSpeak technology to diverse scenarios. In the case of the e-learning application, experts tell us *"beginners learn better with signing videos without subtitles and then they can watch signing videos with the subtitles to see if they already understand sign language".* The answering machine service arouses a similar feeling, at least for those who can sign very well, since *"I would prefer to see him directly signing instead of hearing the voice synthesizer"* betting for the concept of total conversation. Instead someone who cannot sign well or at all, a service like this would be considered a good idea. By contrast VideoSL mail was considered as good and suitable to SignSpeak technology. As long as the translation was almost 100% correct. Finally the game for practicing sign language was really welcomed since *"playing with sign language is the best way to learn it. If it is more formal as in the school, then children would get very bored".*

6. Bibliografía

Abdulla, WH, NK Kasabov, and D.N. Zealand. "Improving speech recognition performance through gender separation." Citeseer, 2001. 218-222.

Acero, A., and R.M. Stern. "Environmental robustness in automatic speech recognition." *IEEE.* 1990. 849-852.

Acero, Alejandro. *Acoustical and Environmental Robustness in Automatic Speech Recognition.* Kluwer Academic Publishers, 1992.

Agarwal, Ritu, and Jayesh Prasad. "A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology." *Info. Sys. Research* (INFORMS) 9 (1998): 204-215.

Akyol, S., and P. Alvarado. "Finding relevant image content for mobile sign language recognition." 2001. 48-52.

Arning, K., and M. Ziefle. "Ask and you will receive: Training older adults to use a PDA in an active learning environment." *International Journal of Mobile Human-Computer Interaction* 2, no. 1 (2010): 21-47.

[32] http://www.youtube.com/watch?v=qFH5rSzmgFE (demo video)

Bagozzi, R.P., F.D. Davis, and P.R. Warshaw. “Development and test of a theory of technological learning and usage.” *Human Relations* (Sage Publications) 45, no. 7 (1992): 659.

Bailenson, J.N., and N. Yee. “Digital chameleons, Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments.” *Psychological Science* (SAGE Publications) 16, no. 10 (2005): 814.

Baldassarri, S., E. Cerezo, and F. Royo-Santas. "Automatic Translation System to Spanish Sign Language with a Virtual Interpreter." *Human-Computer Interaction--INTERACT 2009* (Springer) 5726 (2009): 196-199.

Baltes, P. B., and U. Lindenberger. “Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging?” *Psychology and Aging* 12 (1997): 12â€“21.

Bell, L., and J. Gustafson. “Child and Adult Speaker Adaptation during Error Resolution in a Publicly Available Spoken Dialogue System.” ISCA, 2003. 613-616.

Benzeghiba, M., et al. “Automatic speech recognition and speech variability: A review.” *Speech Communication* 49, no. 10-11 (2007): 763-786.

Bickmore, C.J., H. Vilhjlmsson, and H. Yan. "More than Just a Pretty Face: Affordances of Embodiment." 2000.

Bickmore, T. W., and R. W. Picard. "Establishing and maintaining long-term human-computer relationships." *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, no. 2 (2005): 293-327.

Blomberg, M., and D. Elenius. “Collection and recognition of childrenâ€™s speech in the PF-Star project.” *Citeseer.* 2003. 81-84.

Brashear, Helene, Thad Starner, Paul Lukowicz, and Holger Junker. "Using Multiple Sensors for Mobile Sign Language Recognition." IEEE Computer Society, 2003. 45--.

Brashear, Helene, Valerie Henderson, Kwang-Hyun Park, Harley Hamilton, Seungyon Lee, and Thad Starner. "American sign language recognition in game development for deaf children." ACM, 2006. 79-86.

Bregman, A.S. *Auditory scene analysis: The perceptual organization of sound.* The MIT Press, 1994.

Burkhardt, F., N. Audibert, L. Malatesta, O. Türk, L. Arslan, and V. Auberge. “Emotional Prosody-Does Culture Make A Difference.” *Citeseer.* 2006.

Cassell, J. “Embodied conversational interface agents.” *Communications of the ACM* 43, no. 4 (2000): 70-78.

Cassell, J., A. Tartaro, Y. Rankin, V. Oza, and C. Tse. “Virtual Peers for Literacy Learning.” *Educational Technology, Special Issue on Pedadogical Agents* 47, no. 1 (2007): 39-43.

Cassell, J., et al. "MACK: Media lab Autonomous Conversational Kiosk." 2002. 12-15.

Cavazza, Marc, et al. "How was your day? An Affective Companion ECA Prototype." Association for Computational Linguistics, 2010. 277-280.

Chau, P.Y.K., and P.J.H. Hu. “Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories.” *Information & management* (Elsevier) 39, no. 4 (2002): 297-311.

Clark, R.A.J., K. Richmond, and S. King. "Multisyn: Open-domain unit selection for the Festival speech synthesis system." *Speech Communication* (Elsevier) 49, no. 4 (2007): 317-330.

Cohen, W.M., and D.A. Levinthal. “Absorptive capacity: a new perspective on learning and innovation.” *Administrative science quarterly* (JSTOR), 1990: 128-152.

Dangsaart, S., and K. Naruedomkul. "Bridging the gap: Thai-Thai sign machine translation." 2007. 191-199.

Das, S., D. Nix, and M. Picheny. “Improvements in children's speech recognition performance.” 1998. 433 -436 vol.1.

Das, S., R. Bakis, A. Nádas, D. Nahamoo, and M. Picheny. “Influence of background noise and microphone on the performance of the IBM TANGORA speech recognition system.” *IEEE.* 1993. 71-74.

Davidson, Mary Jo. "PAULA: A Computer-Based Sign Language Tutor for Hearing Adults." 2006.

Davis, F.D. “Perceived usefulness, perceived ease of use, and user acceptance of information technology.” *MIS quarterly* (Management Information Systems Research Center, University of Minnesota) 13, no. 3 (1989): 319-340.

Davis, F.D., R.P. Bagozzi, and P.R. Warshaw. “User acceptance of computer technology: a comparison of two theoretical models.” *Management science* (JSTOR), 1989: 982-1003.

Dean, Katie. *Wired.com.* 01 28, 2002.

http://www.wired.com/gadgets/miscellaneous/news/2002/01/49716.

Drager, K.D.R., J. Reichle, and C. Pinkoski. “Synthesized Speech Output and Children: A Scoping Review.” *American Journal of Speech-Language Pathology* (ASHA) 19, no. 3 (2010): 259.

Dutoit, T. *An introduction to text-to-speech synthesis.* Vol. 3. Springer, 1997.

Eckert, P., and J.R. Rickford. *Style and sociolinguistic variation.* Cambridge Univ Pr, 2001.

Feris, Rogerio, Matthew Turk, Ramesh Raskar, Karhan Tan, and Gosuke Ohashi. "Exploiting Depth Discontinuities for Vision-based Fingerspelling Recognition." 2004.

Furst, J., et al. "Making Airport Security Accessible to the Deaf." *ACTA Press.* 2002.

Glenn, Chance M., Divya Mandloi, Kanthi Sarella, and Muhammed Lonon. "An Image Processing Technique for the Translation of ASL Finger-Spelling to Digital Audio or Text." *NTID International Instructional Technology and Education of the Deaf Symposium.* 2005.

Gong, Yifan. “Speech recognition in noisy environments: A survey.” *Speech Communication* 16, no. 3 (1995): 261-291.

Goodhue, D.L., and R.L. Thompson. “Task-technology fit and individual performance.” *Mis Quarterly* (JSTOR), 1995: 213-236.

Gumiel, J., M. Serrano, and J.M. Moya. “{Automatic sign language recognition: a social approach}.” 2010.

Gustafson, J., et al. "AdApt-a multimodal conversational dialogue system in an apartment domain." *Citeseer.* 2000.

Hernandez-Rebollar, Jose L., Nicholas Kyriakopoulos, and Robert W. Lindeman. "A new instrumented approach for translating American sign language into sound and text." IEEE Computer Society, 2004. 547-552.

Hirschberg, Julia. "Communication and prosody: Functional aspects of prosody." *Speech Communication* 36, no. 1-2 (2002): 31-43.

Hofstede, G., G.J. Hofstede, and G. Hofstede. *Cultures and organizations: Software of the mind: Intercultural cooperation and its importance for survival.* McGraw-Hill New York, NY, 2005.

Hofstede, G.H. *Culture's consequences: International differences in work-related values.* Vol. 5. Sage Publications, Inc, 1984.

Huang, C., T. Chen, S. Li, E. Chang, and J. Zhou. "Analysis of speaker variability." *Citeseer*. 2001. 1377-1380.

Ion, V., and R. Haeb-Umbach. "A Novel Uncertainty Decoding Rule With Applications to Transmission Error Robust Speech Recognition." *IEEE Transactions on Audio, Speech, and Language Processing* 16, no. 5 (2008): 1047-1060.

Janzen, T. *Topics in signed language interpreting: theory and practice*. Vol. 63. John Benjamins Publishing Co, 2005.

Junqua, J.C., S. Fincke, and K. Field. "The Lombard effect: A reflex to better communicate with others in noise." *IEEE*. 1999. 2083-2086.

Karat, John, Daniel B. Horn, Christine A. Halverson, and Clare Marie Karat. "Overcoming unusability: developing efficient strategies in speech recognition systems." ACM, 2000. 141-142.

Kim, S.J., J.J. Kim, and M. Hahn. "HMM-based Korean speech synthesis system for hand-held devices." *Consumer Electronics, IEEE Transactions on* (IEEE) 52, no. 4 (2006): 1384-1390.

Kiss, Imre. "A comparison of distributed and network speech recognition for mobile communication systems." 2000. 250-253.

Koester, H.H. "Usage, performance, and satisfaction outcomes for experienced users of automatic speech recognition." *Journal of rehabilitation research and development* (REHIBILITATION RESEARCH & DEVELOPMENT SERVICE) 41 (2004): 739-754.

Koul, R.K., and G.D. Allen. "Segmental intelligibility and speech interference thresholds of high-quality synthetic speech in presence of noise." *Journal of speech and hearing research* (ASHA) 36, no. 4 (1993): 790.

Krämer, N. C., G. Bente, and J. Piesk. "The ghost in the machine. The influence of Embodied Conversational Agents on user expectations and user behaviour in a TV/VCR application." 2003. 121-128.

Kubala, F., A. Anastasakos, J. Makhoul, L. Nguyen, R. Schwartz, and E. Zavaliagkos. "Comparative experiments on large vocabulary speech recognition." *IEEE*. 1994. I--561.

Lai, Jennifer, David Wood, and Michael Considine. "The effect of task conditions on the comprehensibility of synthetic speech." ACM, 2000. 321-328.

Lawson, A.D., D.M. Harris, and J.J. Grieco. "Effect of foreign accent on speech recognition in the NATO N-4 corpus." 2003.

Lederer, A.L., D.J. Maupin, M.P. Sena, and Y. Zhuang. "The technology acceptance model and the World Wide Web." *Decision support systems* (Elsevier) 29, no. 3 (2000): 269-282.

Lockton, Raymond, and Andrew W. Fitzgibbon. "Real-Time Gesture Recognition Using Deterministic Boosting." 2002. 817-826.

Lockwood, P., and J. Boudy. "Experiments with a nonlinear spectral subtractor (NSS), Hidden Markov models and the projection, for robust speech recognition in cars." *Speech Communication* 11, no. 2-3 (1992): 215-228.

López-Cózar, R., and M. Araki. "Spoken, Multilingual and Multimodal Dialogue Systems." *John Wiley & Sons, Ltd, section3* 2 (2005): 67-70.

Meyers-Levy, Joan, and Durairaj Maheswaran. "Exploring Differences in Males' and Females' Processing Strategies." *Journal of Consumer Research* (The University of Chicago Press) 18, no. 1 (1991): pp. 63-70.

Minton, H.L., and F.W. Schneider. *Differential psychology*. Waveland Pr Inc, 1984.

Morris, M.G., V. Venkatesh, and P.L. Ackerman. “Gender and age differences in employee decisions about new technology: an extension to the theory of planned behavior.” *Engineering Management, IEEE Transactions on* 52, no. 1 (2005): 69-84.

Morrissey, S., and A. Way. "Joining hands: Developing a sign language machine translation system with and for the deaf community." *Citeseer.* 2007.

Nass, C., Y. Moon, and N. Green. “Are computers gender-neutral? Gender stereotypic responses to computers.” *Journal of Applied Social Psychology* 27, no. 10 (1997): 864-876.

Novak, M. “Towards large vocabulary ASR on embedded platforms.” 2004.

Nukaga, N., R. Kamoshida, K. Nagamatsu, and Y. Kitahara. “Scalable implementation of unit selection based text-to-speech system for embedded solutions.” *IEEE.* 2006. I--I.

Olinsky, C., and F. Cummins. "Iterative English accent adaptation in a speech synthesis system." *IEEE.* 2002. 79-82.

Opinion Leader. "OfCom Relay Services." Marker research, 2011.

R., Kathryn D., and Joe E. Reichle. “Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: Applications for AAC.” *Journal of Speech, Language, and Hearing Research* 44 (2001): 1052â€“1057.

Rabiner, L.R., and B.H. Juang. *Fundamentals of speech recognition.* PTR Prentice Hall, 1993.

Rebordao, A.R.F., M.A.M. Shaikh, K. Hirose, and N. Minematsu. "How to Improve TTS Systems for Emotional Expressivity." 2009.

Reeves, B., and C. Nass. *The media equation: how people treat computers, television, and new media like real people and places.* Cambridge University Press New York, NY, USA, 1996.

Roring, R.W., F.G. Hines, and N. Charness. “Age differences in identifying words in synthetic speech.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* (SAGE Publications) 49, no. 1 (2007): 25.

Scherer, K.R. “A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology.” *Citeseer.* 2000. 379-382.

Schnell, M., M. Kustner, O. Jokisch, and R. Hoffmann. “Text-to-speech for low-resource systems.” *IEEE.* 2002. 259-262.

Spiegel, Murray F. “Proper Name Pronunciations for Speech Technology Applications.” *International Journal of Speech Technology* (Springer Netherlands) 6 (2003): 419-427.

Starner, T., J. Weaver, and A. Pentland. "Real-time American sign language recognition using desk and wearable computer based video." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20, no. 12 (1998): 1371-1375.

Stein, D., J. Forster, U. Zelle, P. Dreuw, and H. Ney. “RWTH-Phoenix: Analysis of the German Sign Language Corpus.” 2010.

Stern, Richard M., Fu-Hua Liu, Yoshiaki Ohshima, Thomas M. Sullivan, and Alejandro Acero. “Multiple approaches to robust speech recognition.” Association for Computational Linguistics, 1992. 274-279.

Stevens, C., N. Lees, J. Vonwiller, and D. Burnham. “On-line experimental methods to evaluate text-to-speech (TTS) synthesis: effects of voice gender and signal quality on intelligibility, naturalness and preference.” *Computer Speech \& Language* (Elsevier) 19, no. 2 (2005): 129-146.

Terrillon, Jean-Christophe, Arnaud Pilpré, Yoshinori Niwa, and Kazuhiko Yamamoto. "Robust FACE detection and Japanese Sign Language hand posture recognition for human-computer interaction."

2002.

Varga, I., et al. “ASR in mobile phones-an industrial approach.” *Speech and Audio Processing, IEEE Transactions on* (IEEE) 10, no. 8 (2002): 562-569.

Venkatesh, Viswanath, and Michael Morris. “Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior.” *MIS Quarterly* (Management Information Systems Research Center, University of Minnesota) 24, no. 1 (2000): pp. 115-139.

Venkatesh, Viswanath, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. “User Acceptance of Information Technology: Toward a Unified View.” *MIS Quarterly* (Management Information Systems Research Center, University of Minnesota) 27, no. 3 (2003): pp. 425-478.

Weaver, Kimberly A., et al. "Improving the language ability of deaf signing children through an interactive American sign language-based video game." International Society of the Learning Sciences, 2010. 306-307.

Wheatley, M., and A. Pabsch. *Sign Language Legislation in the European Union.* Edited by EUD. European Union of the Deaf, 2010.

Winters, S.J., and D.B. Pisoni. “Perception and comprehension of synthetic speech.” *Progress Report Research on Spoken Language Processing* 26 (2004).

Zhao, L., K. Kipper, W. Schuler, C. Vogler, N. Badler, and M. Palmer. "A machine translation system from English to American Sign Language." *Envisioning Machine Translation in the Information Future* (Springer) 1934/2000 (2000): 191-193.

Ziefle, M., and S. Bay. “Transgenerational designs in mobile technology.” *Handbook of Research on User Interface Design and Evaluation for Mobile Technology* 1 (2008): 122-141.

Zijl, L. Van, and D. Barker. "South African Sign Language Machine Translation System." *ACM.* 2003. 49-52.

7. Acknowledgments

This work received funding from the European Community’s Seventh Framework Programme under grant agreement number 231424 (FP7-ICT-2007-3).

Appendix A. Expert survey

A.1 Design of the survey

The discussion guide was made by TID and EUD. After a consensus about the questionnaire was reached the experts were selected by EUD trying to get different profiles. The interviews were performed by EUD and they were recorded.

A.2 Discussion guide for the survey

1. **Background information**

❑ Female

❑ Male

Age:

Nationality:

Home city:

Job title:

Knows sign languages:

Knows written (and spoken) languages:

Mother tongue (first language):

2. **Technology products are important for your life. Which factors to select technology products are important for you? How very important? Please rate from 1 to 5 how strongly you think how important the factors are. 1 is the lowest, 5 is the highest. 3 is neutral.**

	Strongly disagree **(1)**	**Disagree** **(2)**	**Neutral** **(3)**	**Agree** **(4)**	**Strongly agree** **(5)**
its price					
its usefulness					
its aesthetic					
its easiness of use					
that it is enjoyable					
its brand					
that it is the ultimate technology					

3. **What do you use mainly the following devices for? (If you don't use a specific device, leave answers blank)**

Smartphone/Mobile:

❑ Make video calls

❑ SMS

❑ E-mail

❑ Real Time Text (RTT)

❑ Internet

❑ Games

❑ Other:

Webcam:

❑ Recording films for social media (Facebook)

❑ Chatting

❑ Sending films by e-mail

❑ Other:

Personal computer:

❑ E-mail

❑ Internet

❑ Chat box (msn, ooVoo, Skype, Camfrog)

❑ Microsoft Office

❑ Games

❑ Other:

Video phone:

❑ Make the calls through sign language

❑ Other:

Television:

❑ Watching the programs

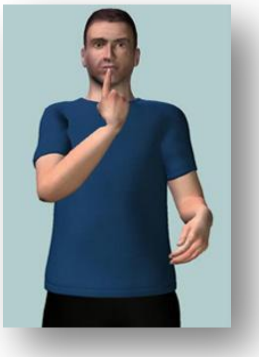
❑ Wii

❑ DVD, video

❑ Other:

4. **Which technology products (for example: called above) do you use mainly? It is possible to fill in more products.**

5. **Virtual signers are agents able to express sentences in Sign Language (as this showed in the figure).**



a. **What drawbacks/problems do you find in them?**

❑ Not natural

❑ Facial expression is not clear

❑ Signs are slow

❑ Prefer a real person instead of an avatar

❑ Fill in if you have more comments:

b. **Do you think they are useful? So yes? Please fill in and add as many comments as you desire.**

❑ Very useful

❑ Very attracted how the avatar works

❑ Easy to use the avatar for various purposes

❑ Fill in if you have more comments:

The **communication between deaf and hearing communities** is a problematic issue: few hearing people know sign language and technology hasn't yet given mature solutions to this. Then, in your experience,

6. **What are the top-3 most unpleasant situations for you to communicate with hearing people?**

❑ Telephoning hearing people through relay service in text, which is not my first language.

❑ E-mails in text that is not my first language.

❑ Video films in sign language are often not subtitled so hearing people cannot understand sign language.

❑ Going to public authorities/service (passport service, bank) where most people cannot sign.

❑ Relay service almost never opens all of 24 hours.

❑ Hearing people cannot learn sign language without instruction in their own written or spoken language.

❑ More comments? Please fill in:

7. **Imagine as many as 3 technology solutions/gadgets which could improve the communication between the deaf and hearing communities (it doesn't mind if you think they are impossible to implement right now). Describe them in a few lines and/or make a simple drawing of your idea. Some examples would be: telephoning by webcam then automatically translates in text or speech, or video films automatically subtitled. More ideas?**

Next, we want to discuss about a specific technology. We would like to encourage you to give so many additional details and explanations as you can. They will be very helpful for us.

SIGNSPEAK TECHNOLOGY

SignSpeak's aim is to translate sign language into text. To this end, a new technology will process a video (pre-recorded or live through a camera) containing a signer and will yield as a result the signed message in text format. Afterwards text-to-speech engines could transform this text message into a synthetic voice.

(View more details in the web of the project: www.signspeak.eu)

Taking into account this brief explanation:

8. What drawbacks would you say this technology could have?

9. And, in your opinion, what strengths could this technology have in order to support the communication for deaf community?

10. We would like to propose to you some examples of scenarios where SignSpeak technology (sign language to text translator) could be used.

John and Mary are a deaf-hearing marriage and they have one child, Susan, who is 7 years old and she is also deaf. This family is bilingual; sign language and spoken language. They have hearing neighbors and family members who cannot sign very well.

a. Scenario "Sign language e-learning"

A neighbor girl of Susan is following a course for improving her sign language skills. For doing this course, pupils have to connect to the teacher through Internet (using a webcam). Then, pupils see the teacher in their monitors and the teacher can see all the pupils at their own homes. The teacher gives the lessons using sign-language and, thanks to SignSpeak technology, text subtitles appear at the same time.

i. What do you think of this scenario?

b. Scenario "VideoSL mail"

Mary wants to send an email to several people. Some of them can hear while others can't. She records a video signing and she sends it. SignSpeak technology translates the sign language message into text and then it sends the email with the video and the text message to all the addressees.

i. What do you think of this service?

c. Scenario "Answering machine"

John is in a congress and makes a video call to home. Nobody is at home, so he leaves a recorded video with his sign language message. The answering machine, through SignSpeak technology, translates the sign language message into text. When Mary arrives home, she realizes there are several messages. As she is busy, she decides listen the messages while preparing the dinner. She listens to her husband's message through a voice synthesizer.

i. What do you think of this service? Only for a hearing-deaf couple?

d. Scenario "Play Sign Language"

Susan has a game console which includes a camera. She wants to play with her neighbor girl. They love to play an educative adventure game that makes you practice some sign language expressions. Using the video from the camera, SignSpeak technology assesses the quality/correctness of the signs and the game gives Susan feedback about how to improve her sign language abilities. As the neighbor girl gets better, she moves forward the levels of the game. They improve their communication very well through playing the game.

i. Do you think a game like this could be attractive for children or even adult people who are learning sign language?

ii. What do you think of this game? How so far will they learn Sign Language through this game?

11. **Once you have realized what the capabilities of SignSpeak technology are, could you give any other example where this technology could be used for improving the communication between deaf and hearing communities? Feel free to express so many ideas as you want and using text or drawings. (This question looks like the question 6) Any comments or ideas about SignSpeak in the general?**

A.3 Videos

It is planned to include some interesting excerpts of the video interviews in the web of the project. Currently we are working on get the necessary permissions from the experts.