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

Use case report analysing industrial applications where SignSpeak could fit

Major Deliverable D.9.5 – M30

Dissemination Level: Public

Release version: V1.0 – end of September 2011

Author: M. Carmen Rodríguez (TID) and Javier Caminero (TID).

Reviewers: Gregorio Martínez Ruíz (CRIC).

1. Introduction

In the Work Package 8 of the SignSpeak project, Telefónica I+D works together with the rest of the partners to get a communication platform with the SignSpeak technology properly integrated. The aim of this prototype will be to improve the communication between the hearing and the deaf community.

Aside from sign language applications, it would be interesting to analyze in which industrial applications SignSpeak technology could fit (as is stated in the purpose of the WP9), such as for improving human-machine communication by gesture, or for an automatic object and body part recognition and tracking in video streams.

On the other hand, the implementation of vision-based system within industrial applications is becoming common, given that the novel features of such systems provide new ways for solving industrial problems. Therefore, the technological breakthrough of SignSpeak will clearly impact in many applications fields as we could see in a coming point within this document.

In addition, due to the breakthrough in sign recognition and translation, SignSpeak results could be applied in other industrial fields apart from sign language, like in human-machine communication as explained the following sections.

This document is structured as follows. First we provide a short description of the main features of SignSpeak technology and we perform a comparison of this technology with other commercial products. Then we describe the main strategic lines in TID, and associated to them, we identify several applications areas in which SignSpeak technology could fit. To illustrate better the industrial application areas, and taking into account the opinion of five experts in technology, we present five use cases in which it is described how SignSpeak technology could be integrated in other application domains. Finally, we present the main conclusions that may be extracted from this analysis. In the appendix A we present the questionnaire that the experts filled in, and their responses.

2. SignSpeak technology

In this section we will briefly describe the main characteristics of SignSpeak technology. Other documents have been delivered during the execution of SignSpeak project which explain widely the technical details of the developed technology. In this case our intention is to give a background of the technology while we stress those features which would make the application of SignSpeak technology in other areas more attractive.

2.1 Description of the main features

The overall goal of SignSpeak is to develop a new vision-based technology for translating continuous sign language to text. SignSpeak combines innovative scientific theory and vision-based technology development by gathering within a common framework novel linguistic research and the most advanced techniques in image analysis, automatic speech recognition (ASR) and statistical machine translation (SMT).

In this section we will give a brief review about the more relevant features of SignSpeak technology with focus on its exploitation. Detailed information of the goals is gathered in the public deliverable D1.1 and updates about the progress of the work in the annual public reports. All of them are available on SignSpeak website[1].

The main characteristics of SignSpeak technology are:

1. *Multimodal video recognition systems.* As signed languages transmit information not only through

[1] http://www.signspeak.eu/en/deliverables.html

the hands movement but also by facial expressions or head movements, SignSpeak is prepared to take into account different channels. For e.g., the hands movement, along with head nodding as a non-manual feature for identifying negation.

2. *Non invasive devices*. The signer will speak without wearing gloves or other types of sensors/markers. The recognition process will entirely be based on video processing.
3. *Robustness to ambient conditions*. The video corpus includes recordings under different lighting conditions so recognition process is supposed to work well under these circumstances. Some work has been done to avoid problems with transient occlusions.
4. *Contextual translation*. The system will carry out continuous sign language translation within a context, not merely identifying isolated signs.
5. *Multilingual*. Scalability to other signed languages is planned but is dependent on the availability of video corpus of those signed languages.
6. *Software integration*. The different modules which configure SignSpeak prototype have been implemented to be connected easily under a common framework.
7. *Context-domain of the translations*. RWTH-PHOENIX corpus has been selected so the context domain is going to be the weather forecast.
8. *Non real time*. There is a factor around 20 for translation process. A real time factor of 20 means 6 seconds of video records will take 2 minutes for providing the translation.
9. *Vocabulary size around 4.000 words*.

2.2 Comparison with commercial products

In order to gain knowledge about the application areas where SignSpeak techniques (video processing, automatic speech recognition and statistical machine translation) are being used successfully we are going to review some of the most outstanding commercial products which are based on them. This will be useful to find the suitable area of application for SignSpeak technology and to apply similar business strategies which help to the exploitation of SignSpeak's results.

2.2.1 Communication

The SignSpeak project could be considered as a technology which translates a modality (sign language) into another (text or voice). In order to do that, image processing and Statistical Machine Translation (SMT) techniques are used.

In the area of communication services, SMT has been used traditionally for performing translation tasks. These services have become very popular and nowadays they are able to get promising results. Some of the more relevant products are:

- Bing Translator[2] which is a translator service provided by Microsoft. It is based on SMT technology but also relies on language specific rule-based components to identify sentences from one language to another. This hybrid system is called "Linguistically informed statistical machine translation".
- Systran software[3] is one of the oldest SMT companies. It supports the translation of Yahoo! Babel Fish[4]. It combines the strengths of rule-based and statistical machine translation in

[2] http://www.microsofttranslator.com

[3] http://www.systran.co.uk

[4] http://babelfish.yahoo.com

order to reduce the amount of data required to train the software and the size of the statistical models while maintaining high performance.

- Google Translate[5] which uses statistical translation models learned from parallel text, that is, sets of documents and their translations. The huge amount of indexed data owned by Google (which in 2006 had already reached one trillion words) is used for automatically learn a model and differs from other approaches based on linguistic rules.

2.2.2 Video gaming

Gaming industry has experimented a deep growing lately. Mainly this has been due to the development of computing (i.e. higher processing power) and the use of more realistic images. In 2009 the gaming industry moved 57.600 millions of euros worldwide. These figures are already very close to cinema and music business.

In the search of new gaming experiences and more natural means of interaction, video cameras and gesture recognition have been introduced in game consoles. In 2003 EyeToy camera was launched by Sony, for being used with PlayStation 2. It was prepared to let users interact with games using motion, colour detection and also sound (through a microphone integrated in the camera). At the end of 2008 the EyeToy had sold more than 10 million units worldwide. In 2007 Sony presented the successor of EyeToy, the PlayStation Eye, for being used with PlayStation 3. It included a new microphone with better cancellation of noise, improvements in the performing in low-light conditions and better responsiveness. Finally Kinect, the proposal from Microsoft, was launched in 2010 for XBOX 360 console. In its first 60 days a total of 8 million units were sold, being the "fastest selling consumer electronics device". Up to now more than 10 million units have been sold.

Figure 1. Video cameras for video game consoles

It is important to notice that camera devices for controlling games have as direct competitors motion-sensing game controllers as PlayStation Move or Wii. Although video cameras have advantages because they let users to play without any extra component, motion controllers are more accurate under certain conditions and also cheaper.

2.2.3 Augmented reality

One of the leading trends in gaming and also for entertainment applications is Augmented Reality. It applies to 3-D virtual objects which are integrated into a 3-D real environment in real time, using matte techniques to display real and computer generated images in a seamless and integrated fashion. Some examples of video games which integrate AR scenarios are:

[5] http://translate.google.com

- EyePet[6] is a game for PlayStation 3 where a camera is used to embed a virtual pet in the images from real world and even people can interact with him.

Figure 2. EyePet game for PlayStation3

- Fantastic Pets follow a similar concept to EyePet. In this case the game is proposed by Microsoft for playing with Kinect.

Figure 3. Fantastic Pets game for Kinect

- Kinect promo videos propose applications that play with social networks and fashion. For example, a couple of friends could select together what to wear to the high school. Even trying their wardrobe on virtually.

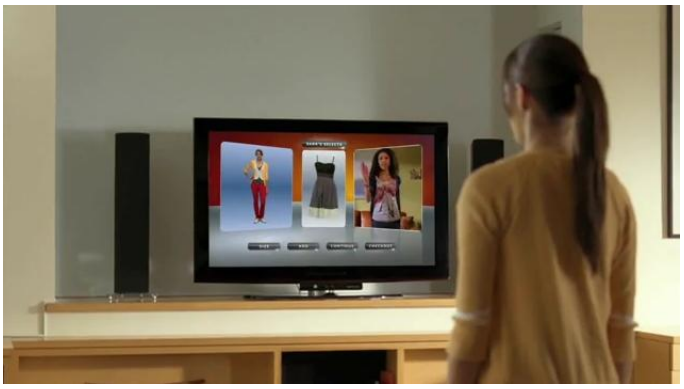

Figure 4. Sample application for Kinect (from promotional video)

- Star Wars Arcade Falcon Gunner uses the iPhone camera to capture what it is going to be the scenario of the proposed adventure. It consists on has taking on the enemy using the famous spacecraft Millennium Falcon.

[6] http://www.eyepet.com

Figure 5. Star Wars Arcade Falcon Gunner for iPhone

In addition to gaming, the number of areas which take benefits from AR interaction is increasing every day. Some of them are:

- Simulation environments: for example to simulate flights or driving activities.
- Emergency services: augmented reality is used to show instructions over real images for evacuations.
- Architecture: in order to visualize planned constructions over the terrain or decoration.
- Support for complex tasks: as for example mechanics or maintenance works, surgery operations, etc. In these cases additional information would be inserted in the vision camp of workers. For example, for surgery, tomography info could be added virtually while surgeon is operating.
- Navigation systems: AR improves the efficacy of navigation systems adding extra information e.g. over the windscreen of a car informing about close roads, traffic jams or simply giving indications over the image of real world.
- Prospection: AR is used for showing an interactive analysis of terrain characteristics.
- Advertising: in order to attract consumers' attention they can show 3D models of products e.g. when a client takes the box which contains the product. Also, to let the clients to try clothes on over their own body.
- Tourism: giving information about monuments, shopping, restaurants, etc. just over the real images which could come from your smartphone.

However, in mobility scenarios, smartphones are ideal platforms for creating augmented experiences. Nowadays smartphones count with very powerful processors, high resolution cameras, connectivity almost everywhere and geolocation. In order to get a glimpse of what applications are being offered we list some of them:

- "Wikitude World Browser"[7]: augmented reality browser based on Wikipedia information and GPS location. You can find more than 350.000 points of interest around the world and display them over the cam view.

[7] http://www.wikitude.com

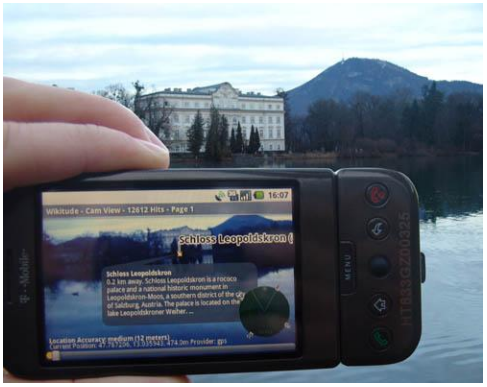

Figure 6. Wikitude snapshot

- "Layar Reality Browser": it is a modern and cool augmented reality browser that displays real time digital information on top of the real world images from mobile camera.

Figure 7. Layar snapshot

- "Google Goggles": it let you launch searches about real world things (books, QR code, famous landmarks, artwork, popular images, etc.), simply taking a picture. It can even translate English, French, Italian, German and Spanish sentences.

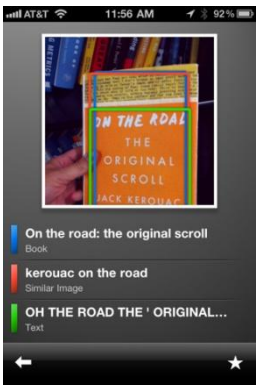

Figure 8. Google Goggles snapshot

- "Wikitude Drive": this application uses the images from the phone's camera to overlay the driving instructions. It permits you to follow the instructions at the same time you are watching the road.

Figure 9. Wikitude Drive snapshot

3. Industrial exploitation

In order to identify the exploitation areas where SignSpeak technology could have some impact, we are going to review the strategic trends which could guide the future development of the upcoming Information Technology (IT) services. For that purpose we will give an overview of the strategic areas upon which Telefónica I+D has business and research units in operation. We believe this approach will provide a realistic snapshot of the current trends and growing areas of the market.

Later, taking into account this perspective and the industrial consulting work carried out by several experts, we will describe what are the opportunities and the application fields where SignSpeak technology could be applied.

3.1 TID trends

3.1.1 Cloud

Cloud computing is defined as the shift from locally installed programs to distant servers where the computation is performed (Hayes, 2008). It is already represented in most of the major companies; in fact 77% of large corporations use the cloud in one of its modalities. However, this is just the beginning since it is expected that in the year 2020 more than 34% of the world total digital information will be stored or transported through cloud services.

Telefónica I+D is putting great efforts into the area of Cloud Computing. It defines differential technologies in order to incorporate them in the products that Telefónica will market. Nowadays, developments are being carried out in different areas such Infrastructure as a Service (IaaS) devices and services, virtualization, cloud content storage services or desktop or PC virtualization. Through these products Telefónica I+D provides a differential strategic value so that customers perceive the value-added of Telefonica's network capacity, of the security of the applications, and of the capacity to migrate applications that are in the customers' private settings to the public cloud.

Some of the main lines within this area are:

- "Virtual Public Cloud" which allows deploying and managing – remotely and automatically – servers and storage infrastructure.
- "Virtual Private Cloud" which allows hosting the user data centres without sharing the infrastructure with another user or company.
- "Virtual Mobile" which is an ecosystem of services related to mobile virtualization to "upload" it to the cloud. Thanks to this technology we can manage several virtual mobiles on the same physical device.

3.1.2 Communication services

It is estimated that there are 1.97 billion Internet users worldwide, who have sent 107 trillion e-mails, have created more than 200 billion websites and over 150 billion blogs and shared in excess of 30 trillion pieces of data of all kinds on Facebook. This trend is seemingly unstoppable and it is changing the way people communicate.

Telefónica I+D is looking for being part of these new communication modes, evolving voice communication services for them to be more personal or converting voice into digital content so it can be organized and searched. The goal is to offer advanced and attractive to customers services trying to break the barrier between the web and traditional communications: voice communications between people.

Some interesting products related to this trend are:

- Giff-Gaff[8] is an innovative virtual mobile network operator built on web 2.0 principles and aimed at the participation and direct involvement of its user community in the success and strategy of the company, rewarding users for that.
- Facebook Call[9] is a mobile application that lets you make calls through Facebook. The friend you want to talk with receives a Facebook chat message advising of the call, and then the call is connected directly from the browser.
- mIO![10] is a Spanish funded project which brings together companies and research organisations. Its aim is to create technologies which allow us to provide widespread services in an intelligent environment and adapted to each individual and his or her life. The mobile phone will be used as a means of interaction, for both services provided by third-party and microservices created and provided by the mobile users themselves.

3.1.3 E-Health

The high cost of citizen healthcare expenses has been for some years now a matter of concern for societies. Moreover, healthcare costs are expected to increase during the next decade. This is partially due to the increase in people's life expectancy worldwide which will be around 70 years in 2020. New lifestyles and availability of goods are leading to the widespread epidemics of conditions such as obesity, all of which will cause an increase in the cost of derived illnesses. In this regard, it is estimated that 2000 million people will suffer from obesity in 2020 and that 80% of them will develop some form of diabetes.

Information technologies and communications have shown their efficiency in providing solutions that reduce medical costs. This is mainly due to the possibility of performing certain activities at the patients' domiciles which in the past were carried out in sanitary environments where the costs are much higher, or bring specialized care services to health centres closer to people. In this sense the integration of technology in daily life of handicapped people or those who don't have technical skills is a key point and so the improvement of technology accessibility is a must. It is related with the concept of Universal Access (Carbonell, 2005) and the different standardization initiatives to set up the design principles for create accessible web interfaces[11].

In this context, Telefónica I+D is working on developing mobile telecare for the telemonitoring of

[8] http://www.giffgaff.com/

[9] http://platform.jajah.com/facebook-call

[10] http://www.tid.es/en/Tecnologia/Pages/TIDInternetTecnologiaFichaProject.aspx?IDElemento=20

[11] http://www.w3.org/WAI Web Accessibility Initative

the elderly, sanitary telecare to provide accessible services of telehospitalization (for both the caring of chronic patients and telerehabilitation) and hospital advanced services for remote control of medical applications and sharing of content and concurrent information. Specifically, "Colabor@" ("Collaborate") is an environment of collaboration, promoted by Spanish institutions and Telefónica I+D, among health professionals. Its focus is to provide tools so that the communication can be as complete as possible.

3.1.4 Financial services

Currently, there are more than 5.000 million of mobile phone users around the world and it is estimated that these numbers will keep on increasing and reaching 6.000 users by 2020. This means that the majority of population has or will have a mobile device and most of them will have access to advanced services which will be connected to Internet. All this and the launch of technologies for close communication such as NFC (Near Field Communication) makes the mobile phone a great candidate as a payment mean.

Telefónica, as other important companies[12], is also involved in this new trend. For this reason Telefónica I+D is working to develop new experiences for payment services. At this point the goal is two-fold, firstly, to make secure processes in order to gain the users' confidence. It is carried out through architectures and technologies like the SIM card which enables the operators to provide these services according to the quality levels required by consumers. Secondly, providing value-added services (i.e. savings on transportation operational costs) make more beneficial the use of innovative electronic payment systems.

3.1.5 M2M

It is a fact that there are many more machines, defined as things with mechanical, electrical, or electronic properties, in the world than people. And furthermore most of them are connected somehow. In fact, in 2014, it is expected that 70% of consumer electronics devices will be connected to the Internet. And in 2020 there will be 50 billion devices connected to the Internet. The concept of M2M (Machine to Machine) tries to take advantage of the communication between all these entities, with no or minimized human intervention (Lawton, 2004).

In this context, Telefónica I+D is developing new platforms to collect and manage large amount of information that causes the use of M2M. Specifically one of the current challenges of Telefónica ID is the so-called "Smart Cities". The associated technological difficulties are numerous, for example, how to manage the increasing number of heterogeneous and geographically dispersed sensors/machines deployed around a city; or how to share the huge amount of heterogeneous information generated among different applications and services; or to create open, easy-to-use, and flexible interfaces in order to get involved the different agents (public administrations, enterprises, and citizens) in the management of all aspects of urban life in a cost-effective way. It is worthy to name three basic concepts needed make real these new paradigms:

- Internet of Things (IoT), defined as a global network infrastructure of physical and virtual "things" based on communication protocols. The devices are seamlessly integrated into the information network.
- Internet of Services (IoS): flexible, open and standardized enablers that facilitate the harmonization of various applications into interoperable services as well as the use of semantics for the understanding, combination and processing of data and information from different service provides, sources and formats.

[12] http://www.google.com/wallet

- Internet of People (IoP): envisaged as people becoming part of ubiquitous intelligent networks having the potential to seamlessly connect, interact and exchange information about themselves and their social context and environment.

An ongoing European project which is expressing all these concepts in a real scenario is SmartSantander[13], led by Telefónica I+D. The city-scale experiment will comprise more than 20,000 sensors and will be based on a real life IoT deployment in an urban setting.

3.1.6 Networks

It is estimated that by 2015 the global Internet traffic will quadruple the current and it will reach 966 exabytes per year. Furthermore, 61% of the Internet traffic will be video and 1 million video minutes will traverse the Internet every second. The number of devices connected to IP networks will be twice as high as the global population in 2015. And In 2015, non-PC devices (smartphones, tablets, etc.) Internet traffic will grow to 15 percent (Cisco, 2011). So it is clear that the new network model is changing rapidly and it is necessary to provide greater capacity (at the lowest possible cost) and greater flexibility to enable new services.

A sample of interesting projects in this area is:

- Genio[14], co-funded by Spanish government and European institutions, is intended to specify home network architecture able to face the problem of heterogeneity of the devices and their interactions as well as to support the management of user needs in relation to the services supported.
- OPTI-TRANS[15] is a FP7 European project which aims to create a mobile Global Navigation Satellite Systems (GNSS) platform allowing travellers to plan their trip efficiently. It will use the information from various public transport authorities and will create a combination of public/private transport in order to display the optimal route.

3.1.7 Security

The immersion of new technologies in daily life of ordinary citizen isn't a trend, it is a fact. But the confidence put on some of these technologies (above all those related with monetary transactions) isn't as solid as it should be. The number of computer networks present in the daily lives of people makes more critical the protection due to the duplicity of sensible data.

In this environment, the role of telecom operators is double: on the one hand as an infrastructure safeguard role and on the other hand as end users trusted entity. In order to play these roles it will be great help software that can verify a user's identity based on the way they interact with a device. Behavioural biometrics is one of the options.

3.1.8 Service platforms

Taking into account the new social paradigms of communication through Internet, open innovation is the new wave for developing services. In this environment it is essential that companies provide elements (APIs, platforms, visibility, etc.) to support this collaboration. There are multiple examples as the Android Market or iPhone App Store. In the case of Android Market, the more recent launched, there were 2.300 apps on March 2009 and now, the number of applications is close to 250.000 with more than 4.500 million of downloads.

[13] http://www.smartsantander.eu

[14] http://projects.celtic-initiative.org/genio

[15] http://www.optitrans.net

In this line, the more promising proposal from Telefónica is called BlueVia[16]. It is a global environment created for a community of developers which allows them to create and commercialise applications based on Telefónica's capabilities (MMS, SMS, call management, etc.) In the research field there are also approaches such as Webinos[17], another initiative corresponding to a FP7 European project, which aims to define and deliver an Open Source platform to enable the development of applications and services based on Web technologies and their consistent and secure use over a broad spectrum of connected devices (mobile, PC, home media and in-car units).

3.1.9 UX and user modelling

The huge availability of different technologies and services makes user interaction more complex and offers lots of possibilities to users. The user preferences have to be taken into account for the creation of new services and so recommendation systems are already commonly used. In fact, 61% of consumers want their communications service provider to offer customized services and plans based on their own consumption habits of mobile broadband.

Telefónica is acquainted with this need and aware that it is increasingly important to know the customers in order to offer differentiated added value services to provide to the users a delightful user experience and an intelligent customization.

In this sense the work lines are two:

- User modelling: to improve business intelligence through a better understanding of customers using new algorithms to obtain user preferences and expected behaviour and thus to make the customisation of their services easier for the provider. This requires managing a huge amount of information, requiring significant technological advances in the area of massive information processing.
- User experience: from the information obtained through users' behaviour and user research, the next step is to apply all this knowledge to design easier and more intuitive interfaces. In this way the introduction of new devices and technologies to general population will be a smoother process.

3.1.10 Video

As it was already mentioned the video will become the main source of traffic on the Internet (61% of the Internet traffic by 2015). Telefónica I+D wants to offer multimedia services with a differential user experience across all connected devices. Specifically, entertainment services around the video, Augmented Reality (AR) interfaces, communication services of high quality video and tele-presence.

A special mention deserves Augmented Reality. It is a term for a live direct or indirect view of a physical, real-world environment whose elements are augmented by computer-generated sensory input such as sound, video, graphics or GPS data. The strategic importance for Telefónica I+D of this sector is represented by the recent partnership established with Layar[18], a pioneeringcompany focused on AR experiences in the field of Internet search.

In order to compare different video proposals and identify their innovation gaps (in function of dimension that really matter to people) Telefónica I+D uses an own conceptual framework called Video Habitat[19].

[16] http://www.bluevia.com

[17] http://www.webinos.org/

[18] http://www.layar.com/

[19] http://www.tid.es/es/Tecnologia/Documents/VideoHabitat.pdf

There are already several Telefonica's video products at the market. For example, "Movistar Videoclub" uses Internet connectivity to provide a service for renting movies, series and other multimedia contents, and "My World" is a platform which provides a set of specially designed interactive services to be enjoyed on TV.

In the research activity there are projects intend to improve the tele-presence as, for example, "3DPresence"[20], a FP7 European project with the goal in mind of implementing a multi-party, high-end 3D videoconferencing concept that will tackle the problem of transmitting the feeling of physical presence in real-time to multiple remote locations. "Vision"[21] is a Spanish initiative with similar objectives and that stresses the importance of studying the human factors involved in the perception of presence.

Regarding Augmented Reality services, Telefónica I+D is working in the integration of "Iris"[22] service, which provides information related to an object just by taking a picture of it, with Layar technology of AR visual search.

3.2 Application areas

In the previous section a compilation of technological trends from the point of view of Telefónica I+D has been given. Somehow these trends could be able to affect the exploitation of technology products over the following years. Together with it in Section 2.1 we have explained briefly the main capabilities of SignSpeak technology and we have reviewed some products with certain similiarities with SignSpeak and that are already in the market (Section 2.2).

Regarding all this information and industrial experts' opinions gathered in Appendix A. Industrial prospection we propose several application areas for SignSpeak technology.

3.2.1 Biometrics

The need of security in order to guarantee the confidentiality of communications and the privacy of data will be a key issue for the future ICT (Nissenbaum, 2010). Accessing to remote personal services (e.g., e-health profiles, cloud storage, etc.) or validating transactions through technologies as electronic wallets are examples of scenarios were a confident user identification would be needed. Biometrics, that is, automatically recognizing a person using distinguishing traits, is an option to face the task of recognizing you as you.

Traditionally most of the biometrics technologies need some special sensor to capture the users' traits (e.g., fingerprint sensors, digital tablets, etc.). There are some of them that rely on images from video cameras, as for example it happens for facial recognition or hand shape. In this case it gets a more comfortable and seamless user experience because they don't need to have access to complex sensors (which are also more expensive). On the other hand the working conditions for capturing properly the users' traits are stricter, above all regarding lighting.

As one of the industrial experts proposed a possible use of SignSpeak technology could be for *"offering access by means of an established sequence of symbols"*. It would permit to users an easy access to personal applications. Moreover in case of mobile devices, which have mostly video cameras integrated, they wouldn't need to add any extra hardware component.

[20] http://www.3dpresence.eu

[21] http://www.cenit-vision.org

[22] http://www.tid.es/en/Tecnologia/Pages/TIDInternetTecnologiaFichaProject.aspx?IDElemento=13

3.2.1 E-Learning

The application of SignSpeak technology for learning context is getting clear when we consider teaching signed languages. This case has been discussed in the public document D9.4 and reader will find further details on it. Beyond this, nowadays, the use of video processing for e-learning is mainly being directed to Augmented Reality approaches. It is thought that learning applications of AR could be beneficial for providing visual and rich examples of complex phenomena (e.g. engineering, environmental applications, etc.) and for getting illustrations, stories, etc. which traditionally live in books to life. This is a wonderful way to engage students of all ages.

From Telefónica I+D point of view this kind of applications would be useful for enriching tutorials of technology products or any other device include in the interaction with a client.

3.2.2 Gaming

In the Section 2.2.2 we have reviewed the most popular game consoles which make use of video for interacting with the game or even for creating the background of the game scenario. It seems clear SignSpeak technology could be re-orientated to this area, as some of industrial experts surveyed for this deliverable have noticed (see Appendix A. Industrial prospection).

A related proposal would be *"training for sportspeople and dancers"* which might be presented in game formatting. Thus, attending to the analysis of video images sportspeople or dancers could know if they are practicing well, or if they have to change their technique.

3.2.3 Healthcare

In the opinion of the questioned industrial experts some of the proposals for applying SignSpeak technology in the Healthcare area would be:

- Virtual surgery, that is, virtual reality simulation of surgical procedures. If SignSpeak had enough accuracy capturing the surgeon's movements (hands and body postures) then these actions could be monitored and alerts would be launched in case some dangerous situation would be detected.
- Augmented reality for surgery. SignSpeak tracking abilities could also be used for letting surgeons to operate superimposing visual information (i.e. dangerous areas for tumours or thermography information) in order to minimize the impact of the surgery.
- Rehabilitation exercises. SignSpeak would recognize the patients' movements and it would give some indications about needing corrections or recommendations of new exercises.

Taking into account the interest of Telefónica I+D in developing initiatives related to healthcare, we find really useful the application of computer vision for improving the relation with the patients. In the Section 4 a use case is described which tries to develop the application of SignSpeak in this area.

3.2.4 Human computer interaction

Nowadays, in the industrial field, the most intense efforts are focused on make usable products. For getting close to this goal is necessary to study the specific features of the targeted users and to design interaction strategies which, taking into account the mentioned users' characteristics, guide people smoothly and easily to the services' goals.

SignSpeak technology could help to improve human-technology interaction regarding two aspects: firstly, using its multimodal video recognition in order to enrich the means of interaction with gestures and natural movements, and secondly, using its tracking abilities and recognition of facial expressions in order to include this emotional information into the flow of the interaction. In case of the gestural interaction some proposals would be:

- Writing for mobile devices. Using movements in the air which could represent letters or even concepts. The advantage would be that we avoided the tedious use of keyboard, typical for this kind of devices.
- Controlling in-home devices as televisions. The user would make gestures for representing each action as: changing TV channels, turning up or down the volume or accessing to any other TV service.

Regarding the affective interaction, that is, the one which *"emotional information is communicated by the user in a natural and comfortable way, recognized by the computer, and used to help improve the interaction"* (Reynolds & Picard, 2001), the application of SignSpeak could be applied in the following situations:

- In conversational systems, in order to analyse the users' emotional state with the aim to apply communicative strategies to show empathy to the user.
- In in-car systems, in order to survey drivers' attention and in case it was detected some dangerous situation (e.g. tiredness), then alert or even act in the vehicle.

Apart from these two means of application of SignSpeak technology it would worth to mention, in a more general way, the usefulness of video processing capabilities to create Augmented Reality interaction scenarios. Augmented interfaces use to be very attractive and to encourage users in the use of applications. Some Augmented Reality applications are proposed for some of the application areas presented in this section and also, with a more commercial point of view, in Section 2.2.3.

3.2.1 Manufacturing

There are several complex tasks which require from workers very specialized trainings and high memorizing skills. With Augmented Reality technology, enhancement of information in a variety of workspaces is possible. For example BMW have developed a concept of mechanics assistant where the worker, using augmented reality glasses, is able to perform maintenance on the company's cars (see Figure 10). The information, visual and through audio, is about the part which needs to be replaced or how to make these replacement operations.

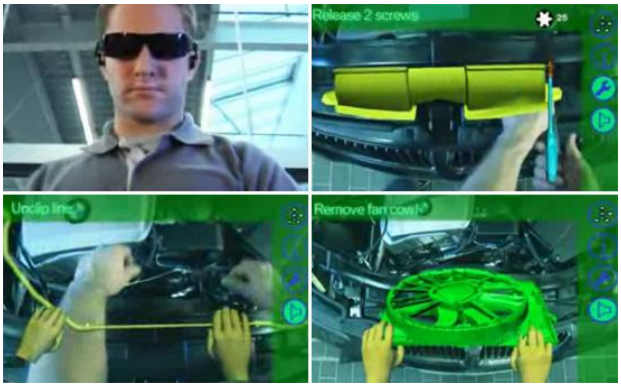

Figure 10. BMW Augmented Reality system for mechanics works

SignSpeak technology could help to create similar applications for manufacturing works and even to take advantage of its capabilities to analyze the hands movements to correct the actions of operators.

3.2.2 Robotics

Independently of the area where robots are intended to work, if we want to give robots real autonomy is necessary to equip them with powerful tools for perception. Computer vision has experimented remarkable advances for some years to now, however more efforts are needed. The ability of SignSpeak for video processing could be used *"for controlling the precision of robots in several industrial environments"* or to permit that "*emotional information is communicated by the user in a natural and comfortable*

way, recognized by the computer, and used to help improve the interaction".

3.2.3 Surveillance and monitoring

Due to security reasons the most developed application of computer vision is video surveillance. However surveillance and monitoring techniques are not only used for find terrorist in a crowd. The possible applications of SignSpeak capacity for monitoring and tracking are quite numerous and very useful. For example, traffic surveillance, where a real-time analysis of traffic on the roads could help to manage the cars in a big city. As one industrial expert proposed *"another possibility could be to watch the driver's behaviour from an in-car system"*. Thus, SignSpeak could alert the driver or even the vehicle about her state (tiredness, nervous, etc.). Also, road safety would be improved if technology would warn driver about other situations, external to the car (i.e., alerts due to close vehicles, pedestrians, etc.)

A use case has been made in case to explain a typical scenario for this application area (see Section 4).

3.3 Integration problems

In the deliverable D9.4 "Report about the study of the new communication bridge between signers and hearing" writers presented an exhaustive review about factors which could affect the performance of SignSpeak technology. They are grouped in the following categories: user factors (referred to the specific users' characteristics as age, experience or cultural background), environmental factors, resource-related factors, dependence on the capture devices and user perception/acceptance of the technology.

We also requested to the expert panel, knowing their limited knowledge about SignSpeak technology but counting with experience in similar technologies, what were the problems they think could arise with the industrialization of SignSpeak.

The main trouble that they point out is their doubts about a proper performance with bad lighting conditions or problems with low contrasts between clothes and background. An expert in artificial vision affirms that *"It would not work properly in outside environments with a high contrast level due to the low quality of the webcam"*. They admit the complexity of the recognition task and that's because they also are suspicious if SignSpeak technology could have a very high computational load and thus not being able to obtain result in real-time.

If we consider deaf community as potential users of SignSpeak technology, the problem, in opinion of the expert panel, is that *"market interest may be limited by the lack of users"*. That is, unfortunately, minority groups are sometimes not a good business for certain companies. Anyway, a complete market research would be needed to gain more knowledge about questions as *"Are there any other solutions to cover the same needs? Which advantages have SignSpeak over the other solutions? Which would be the price of this technology in the market? Are investments companies interested in SignSpeak?"* or other issues as *"privacy"*.

4. Use cases

The oncoming section is intended to propose several scenario samples and describe the possible sequences of interactions between SignSpeak technology and the corresponding external actors.

The aim is to show how SignSpeak technology could be integrated in a real application and find out the responsibilities and behaviour of each component. The intention is not to reveal how the internal parts are structured and technology and actors are seen as "black boxes".

4.1 Use case 1: *"Secure access for smartphones"*

Maria has just bought a smartphone. One of her major concerns with this sort of devices is to achieve a secure access to the applications and the information stored in the device. As she declares *"It is not simply a mobile phone, you can store a big amount of private information here…"* Therefore she decides to use a secured access system which makes use of the SignSpeak gesture and movement recognition technology by means of her smartphone. Thus Maria can be authenticated by the SignSpeak technology using a sequence of gestures or movements (even it could be used the representation of her signature) which would have been trained before. At the end, she is able to perform her personal sequence of movements in order to access private applications, and she gains confidence using the system.

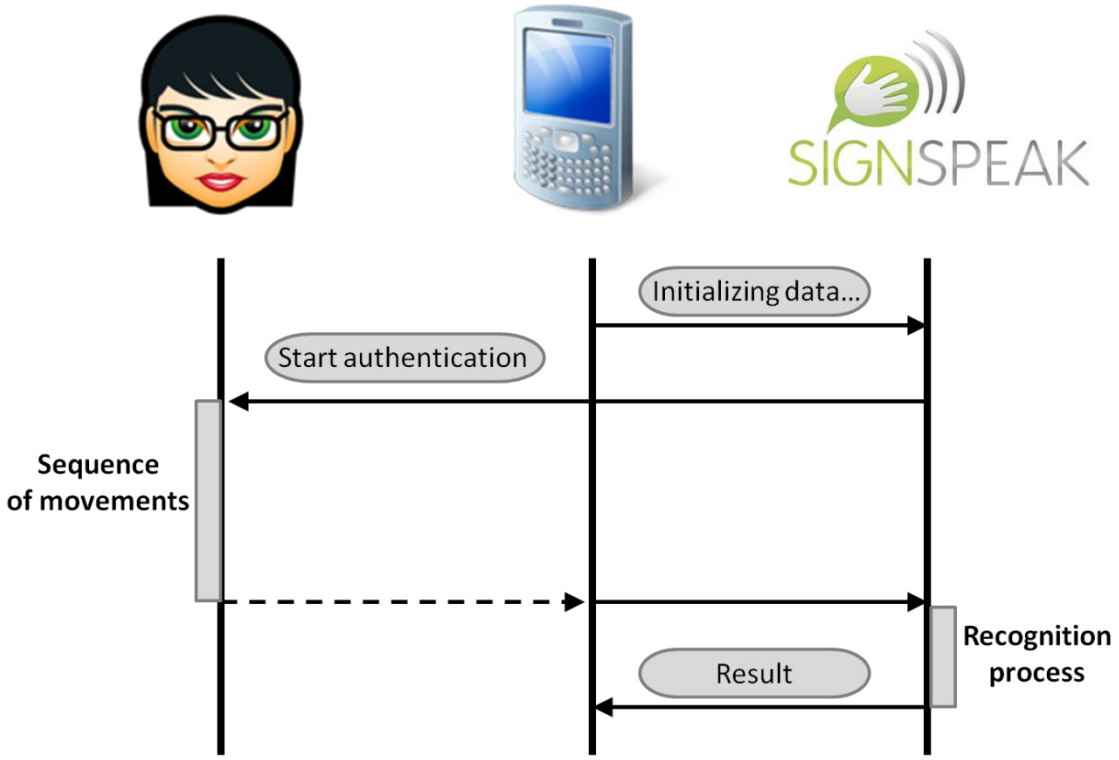

Figure 11. Use case diagram for "Secure access for smartphones"

4.2 Use case 2: *"Physical therapy monitoring system"*

David has been got into surgery due to a broken ankle. After the operation, David is ready to receive a long recovery treatment of physical therapist from his home. Thus, he would not need to go to the hospital periodically during ten long months. Moreover, he would leave a free place in the hospital in order to be occupied by another person.

David uses a PC with a webcam. The PC has a monitoring system installed which is connected to SignSpeak technology. Thus, the doctors can connect to this monitoring system and specify the physical therapist treatment that David should follow. On the other hand, David can start this program of exercises and perform them in front of the webcam. At this point a video connection is established between David and the monitoring system. SignSpeak technology recognizes the movements performed by David and check if the exercises are well done. If he is not doing the exercises properly, SignSpeak technology gives feedback to David providing markers over the video signal which helps to correct the exercises. Finally, with the purpose that the doctors and physical therapists track David's improvements, the system sends the data storage to them.

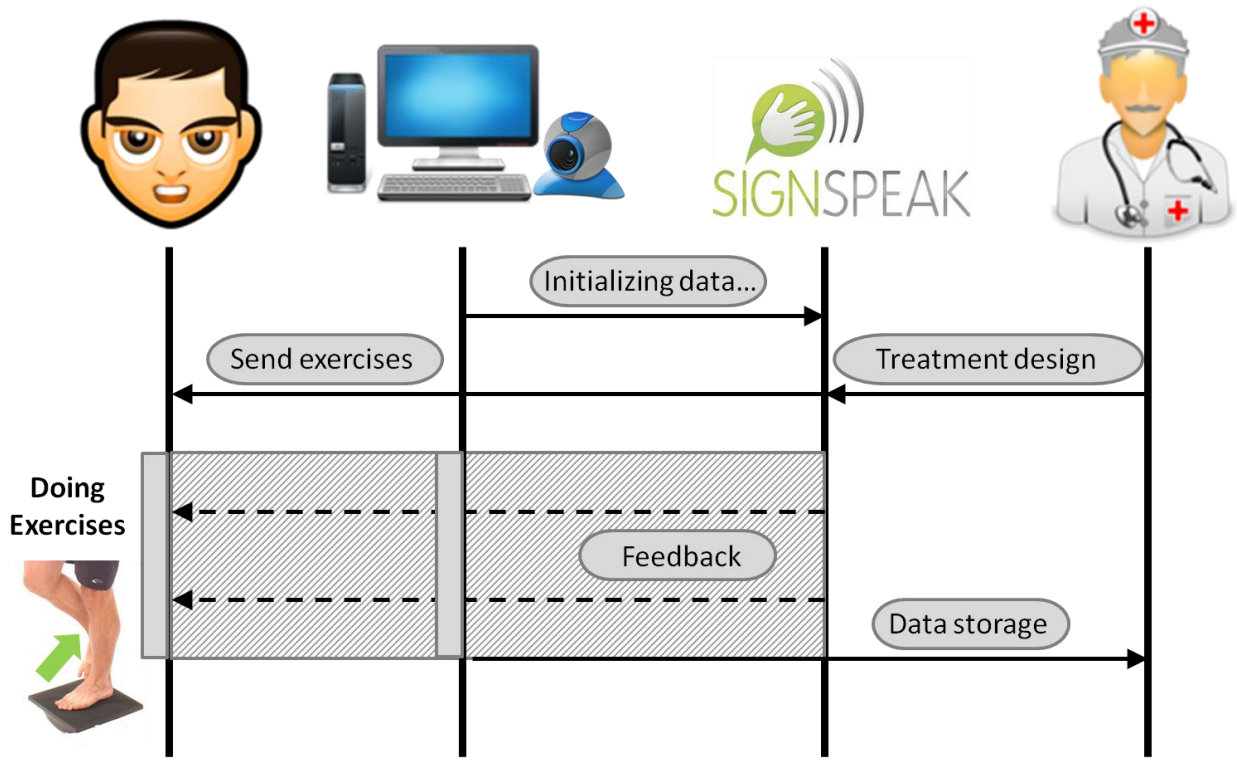

Figure 12. Use case diagram for "Physical therapy monitoring system"

4.3 Use case 3: *"In-home devices control system"*

Ana has just got installed a platform for controlling the electronic devices in her home. This system is connected to the SignSpeak technology, and therefore, is able to control every device connected to the platform by means of gestures and movements. Thus, whenever Ana or another member of her family (for example her mother who find difficult to understand how the TV remote control works) wants to change the channel of the TV, the procedure will be as follows: The TV connects to the SignSpeak technology initializing the system. Ana (or her mother) perform a pre-establish gesture to change the TV channel. While she is doing the gesture, a camera- which is connected to the TV- sends the video signal to the SignSpeak. SignSpeak technology recognizes the movements from the video and associates them a specific control order. Finally SignSpeak sends the order ("change channel") to the TV, and this device executes the action. Now, Ana and her mother are able to control the TV using a set of simple gestures.

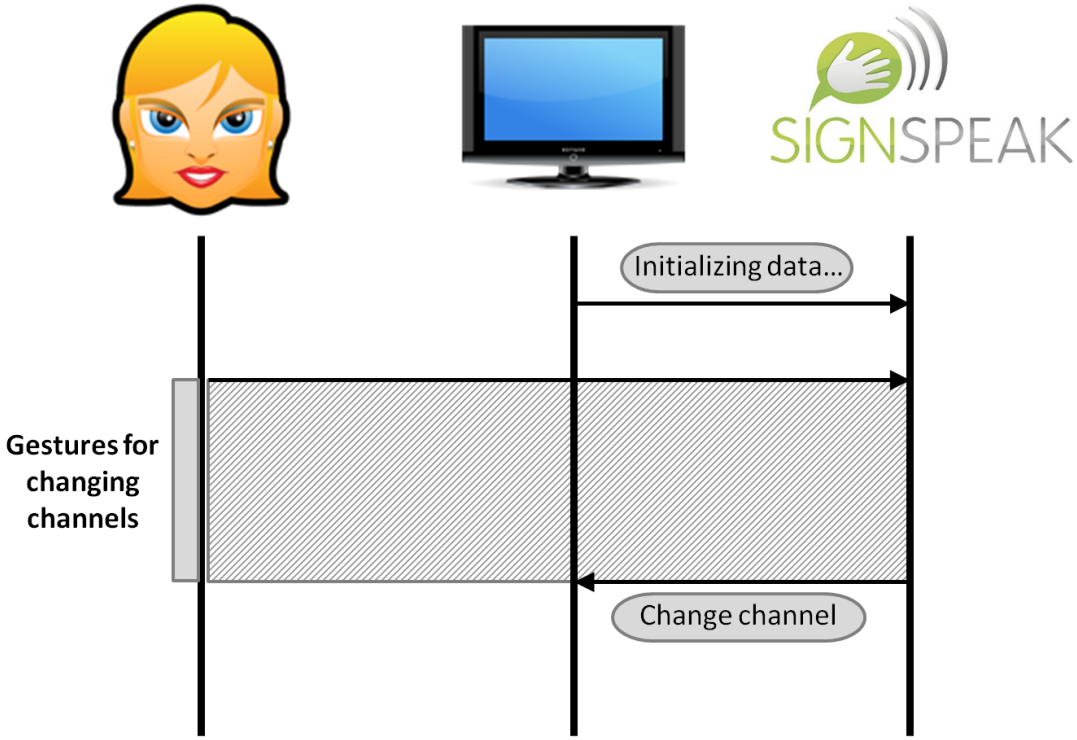

Figure 13. Use case diagram for "In-home devices control system"

4.4 Use case 4: *"Gesture writing system"*

Lately, Daniel is having problems to see clearly at close distances. It means serious difficulties when he is interacting with his smartphone, especially when he writes. SignSpeak technology offers him a new way of interacting with his smartphone, he can make movements which represent letters or even concepts and SignSpeak translates them to text. After that mobile reads him (TTS) what have been recognized in case it would be necessary any correction. Now Daniel is able to use his mobile device every time and without any restriction.

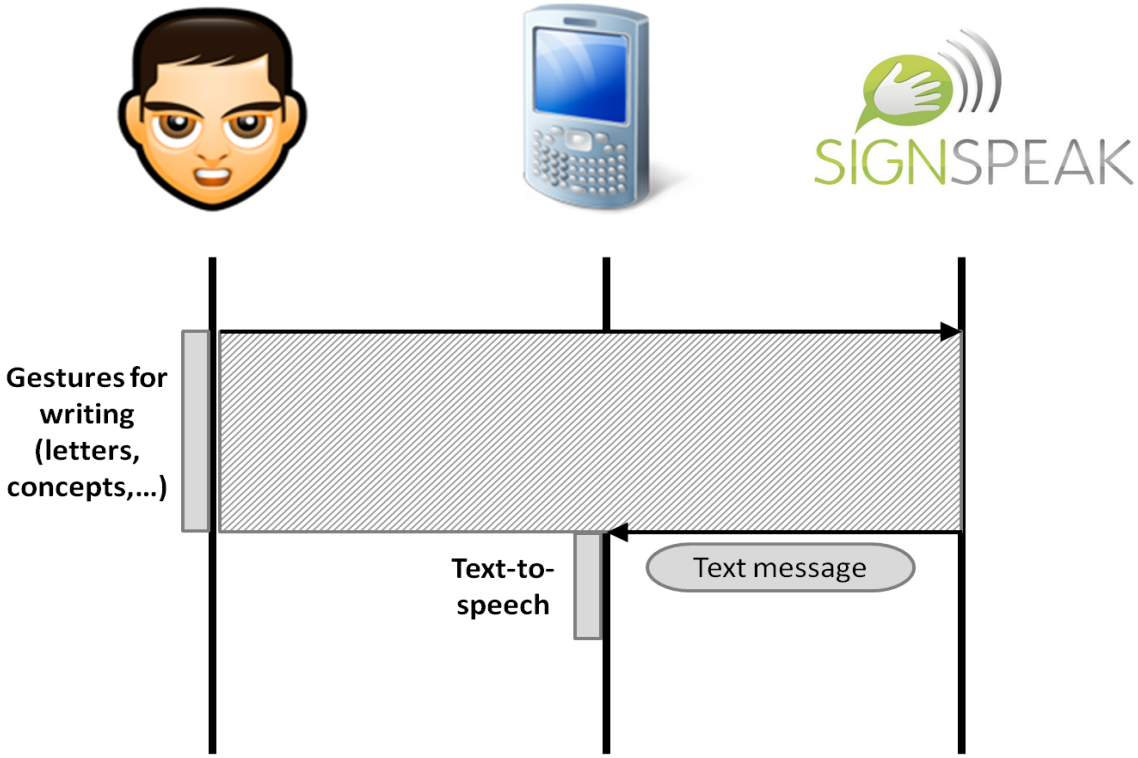

Figure 14. Use case diagram for "Gesture writing system"

4.5 Use case 5: *"Driver attention monitor"*

Rose is a bus driver. She usually does long distance journeys. Therefore she spends many hours driving which may increase her tiredness. Recently, the transport company has installed a system for driver's surveillance in every bus. This system has a camera in the bus that captures a video from the driver. The video stream is sent to the SignSpeak by the bus connection system. Thus, SignSpeak recognition technology is able to determine the tiredness of the driver from the driver's facial expression and her body posture. If SignSpeak detects that Rose is tired or almost falling asleep, it will inform her by means of an alert signal. This way, Rose's driving is calm because she knows that her life and the rest of passengers' life are getting safety.

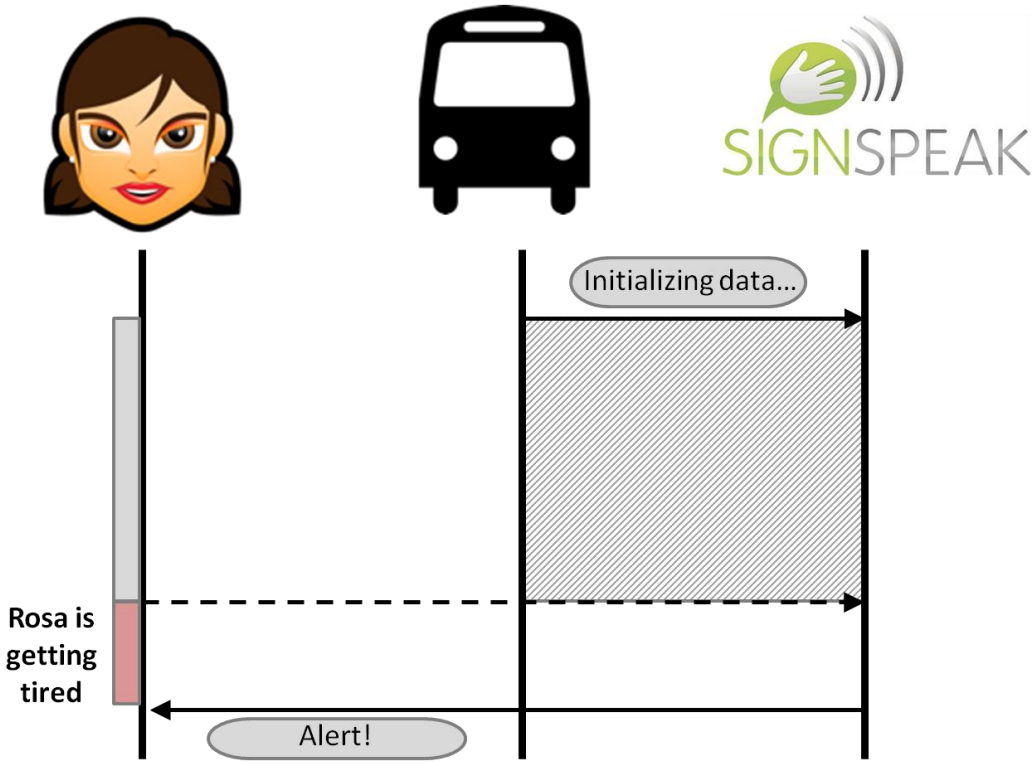

Figure 15. Use case diagram for "Driver attention monitor"

5. Conclusions

In this document we have made a compilation of some commercial products which include a similar technology to the one developed in SignSpeak project. These applications are related to communication (e.g. Automatic Speech Recogniser, etc.), video gaming (e.g. Kinect, EyeToy, etc.) and augmented reality and they make use of video processing and statistical machine translation, as SignSpeak technology.

Next, we have presented the technology trends which, under the perspective of Telefónica I+D, will lead the development of ICT products. In our opinion areas such as Cloud computing, UX, or E-health will guide the commercialization of the future products.

With all this information in the mind, in Section 3.2, we propose a set of areas where SignSpeak could be applied successfully. These areas arise from the needs of the mentioned trends such as secure user's authentication (e.g. in financial services) or remote assistance of patients (e.g. in e-Health services). Thus, application areas as Biometrics and Healthcare appear as very promising for the use of SignSpeak technology. Finally, Section 4 becomes more specific and we suggest five use cases which represent the application areas proposed.

Certainly further work is needed in the study of the industrial application of SignSpeak technology. However this document is a first step in order to look at together, commercial products and SignSpeak technology.

6. Bibliography

Carbonell, N. (2005). Multimodal Interfaces-A Generic Design Approach. *Universal Access in Health Telematics. LNCS , 3041*, 209-223.

Cisco. (2011). *Visual Networking Index: Forecast and Methodology, 2010-2015*. Cisco.

Hayes, B. (2008). Cloud computing. *Communications of the ACM , 51* (7), 9-11.

Lawton, G. (2004). Machine-to-machine technology gears up for growth. *Computer , 37* (9), 12-15.

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford Law \& Politics.

Reynolds, C., & Picard, R. W. (2001). Designing for Affective Interactions., (p. 499).

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. Industrial prospection

A.1 Design and procedures

Taking into account the innovative character of SignSpeak technology the selection of the panel expert was made knowing we want to know the opinion of industrial experts but without wanting to lose the perspective of research world. Thus, we selected three experts from different areas of Telefónica I+D, who are used to work in the process of launching of commercial products and two more experts who were related with research institutions.

In order to collect their opinions we elaborated an online questionnaire and sent it to them. The questionnaire is online in case it wants to be consulted[23]. The questionnaire started with a brief introduction of SignSpeak project as follows:

> The aim of SignSpeak Project is to translate Sign Language (from deaf people community) into text. Therefore a new video processing technology (using both video recorded or live video) is being developed to capture the gestural sequence from a signer and translate her message into text. The main functionality of this technology is to extract the visual characteristics linked to the tracking and the shape of the hands and the face. From this information, classical techniques of speak recognition are used to identify the movements of the user in real time. The main features of this technology are:
>
> - **Multimodal video processing.** There are several information sources which are monitored in a simultaneous way (hands, face, head and trunk). Thus, the system not only tracks the movement of the dominant hand, but also the other hand, the facial expression and the body posture (shoulders, elbows, chest, etc). In addition, the redundancy of the information helps to increase the robustness of the recognition system.
>
> - **Natural video capture.** The sensor for capturing video is a camera similar to a webcam. Another kind of systems normally uses gloves or other types of sensors or markers, which usually decrease the user acceptation level of the technology.
>
> - **Robustness and adaptation to different environments.** The system behaviour would be independent of the background colour, changes of illumination and user characteristics (speed, clothes, etc.)
>
> Taking into account this general vision of the SignSpeak technology, we would like that you complete the following questions from an industrial point of view and considering your domain of specialization.

[23] https://docs.google.com/spreadsheet/viewform?formkey=dHYxZ3hxSUVsb0Zyb2hvTXQtWWxUemc6MQ#gid=0 (in Spanish)

And after that the following questions were asked:

1. In which area are you working at present?
2. Could you specify your application domain?
3. Could you mention a product – which can be purchased nowadays in the market- whose technology would be similar to SignSpeak technology?
4. Taking into account the previous information, and according to your professional specialization, try to imagine up to three manufactured products where the image processing technology of SignSpeak could be included.
5. What kind of problems do you think the integration of SignSpeak technology into a consumer good may have?
6. According to your experience, in which strategic line of Telefónica do you think SignSpeak technology could be more suitable to be included?

A.2 Transcriptions

Therefore the development should be done including the potential real users, thus their requirements and needs would be covered. It could be done as a classical product development but we should take into account the fact that they are a very specific collective, and they may have more specific needs. SignSpeak is a technology that would contribute to increase the accessibility in any interpersonal communication service, or even in services with a multimodal conversational interface.

1. Does not know/does not answer

A.2.1 Expert 1 (Telefónica I+D)

1. Financial Services
2. Java Development
3. New videogame systems (from Xbox-Microsoft or PlayStation-Sony) based on movements
4.
    a. Authentication: offering access by means of an established sequence of symbols. Maybe, apart from the own symbols, it would be possible to extract information from the users behaviour (i.e. how they perform a sequence of symbols). This may help to identify users by means of their behaviour, in the same way of a biometric application. On the other hand, authentication would be dependent on the existence of actual differences between users. Moreover, repetitions patterns in the same user should be found in order to get a properly performance of the authentication service.
    b. Writing in mobile devices. With the aim to avoid the user of the keyboard, this proposal could be a quick way of writing text, but using only one hand.
    c. Interface for controlling videogames. If this system allows mapping symbols to language, it would be possible to re-orientate the system for recognizing actions in order to control a videogame.
5. Market interest may be limited by the lack of users. On the other hand this system offers an alternative to voice recognition technology useful for those users that are not able to speak properly. Alternative methods, such us graphics interfaces and another devices (mouse,

keyboards, etc.) are already being used when a user is not able to use a voice recognition system. In addition, if a high processing load is needed, it would restrict the context or environments in which this technology can be used.

6. Maybe eHealth could be the most suitable area, because they probably have into account the disability domain in their developments.

A.2.2 Expert 2 (Telefónica I+D)

1. Scientific area, subsidies projects
2. Human-Machine Interfaces, Multimodal Systems, Embodied Conversational Agents (ECAs)
3. Interactive systems based on gestures such as Microsoft Kinect or Sony Eye Toy (although these have humblest objectives)
4.
    a. Television (or any other electrical appliance) controlled by user gestures. i.e. turning the TV up or down, or switching the television channel using the user's hand or head gestures.
    b. Physical medicine and rehabilitation systems. The image processing system of SignSpeak would be useful for recognizing the exercises carried out by the patient and also for helping the patient to perform the exercises in a more effective way. It could also give remote feedback to the doctor, who supervises the rehabilitation, about the quality of the exercises performed by the patient.
    c. Augmented reality systems. The recognition of the user's body posture could be used for getting a precise representation of the augmented interface. It could be useful in systems which may have a limited interaction with commonly used tools (i.e., in systems for visualizing 3D objects which use rotations and navigations models such as 3D studio or Maya)
5.
    - Changing lighting environment
    - Situations in which the contrasts between users and their clothes or the background are low.
    - Video Processing demands a high processing load.
    - Privacy issues in video applications (i.e. Face recording)
    - The capturing system would need to be calibrated regarding to the position, focus and frame. It is an issue that should be taken into account when moving the camera or its stand.
6. It should be suitable to be included in the area that I am working in (interactive systems and human-machine interfaces). It also could be integrated in accessibility research lines, TV and video technologies. Another possibility could be the user experience domain, due to the usability components that this type of systems may bring.

A.2.3 Expert 3 (Telefónica I+D)

1. Human-Computer Interactive Systems.
2. Technologies for the development and evaluation of interactive systems.

3. I believe that the products more similar to SignSpeak technology are the systems for controlling videogames using gestures, such as Kinect or GestureTek.
4.
    a. Due to the accuracy in gesture recognition, SignSpeak technology could be used in products for virtual surgery, where a high ability to monitor the surgeon's hand and body movements is needed.
    b. Regarding to the quality in video recognition, SignSpeak technology could be used for controlling the precision of robots in several industrial environments.
    c. Depending on the computational efficiency of SignSpeak technology, it could be used for controlling user devices. An example could be the control of cameras and mobile phones, using the SignSpeak technology as a "biometric" access to the device.
5. The main problems that SignSpeak technology may encounter are:
    - The precision of the technology, particularly in difficult environments, for example those environments where the lighting is complex.
    - The computational load or computer requirements this technology would need in order to work properly in real time.
6. SignSpeak technology could be included in Telefónica, in any development area of the human machine interaction domain, such as Digital Home applications or mobile devices.

A.2.4 Expert 4 (Researcher)

1. Research in multimodal human-machine interaction
2. User-Centred evaluation in multimodal interactive systems.
3. SignSpeak technology could be similar to speech recognition technology. It can also be compared with any technology which could be able to translate information from one modality to another. An example may be the rendering from an SMS to a synthetic voice.
4.
    a. Training for sportspeople and dancers. Something similar to SignSpeak, based on image processing, could be used for analysing both the motor technique in sportspeople and the interpretative technique in dancers. It could also be a useful tool for helping them to improve these techniques.
    b. Traffic surveillance: real-time analysis of drivers' behaviour. It could be performed watching the movement of a vehicle from the highway. Another possibility could be to watch the driver's behaviour from an in-car system. Regarding this topic I could easily think in many applications that would help to improve the Road Safety Awareness (i.e., warnings to the driver, alerts to the nearest vehicles or to the Traffic Division Headquarters, etc.)
    c. In conversational systems, it could be useful as a tool to analyse the users' emotional state (by means of their gestures and their facial expression) with the aim to design communicative strategies to show empathy to the user.
5. This system should work in real time. This system will probably demand a high load in processing data. A deep market research should be done. First of all, Does it cover an relevant need in deaf community? (or even in blind people - as it could serve as a platform to communicate both impairment communities), Are there any other solutions to cover the same

needs? Which advantages have SignSpeak over the other solutions? Which would be the price of this technology in the market? Are investments companies interested in SignSpeak? In addition, we should also think from the beginning in the specific applications that would support this technology. In which devices it could be available, where it could be used, in which contexts, etc.

6. SignSpeak technology could be include in any Telefónica I+D initiative which is about developments in the human-computer interaction such as Digital Home or mobile devices.

A.2.5 Expert 5 (Researcher)

1. Artificial Vision
2. Luminosity characterisation and their influence in matching and tracking algorithms.
3. Regarding to the sign capturing technology, this application would be similar to the Kinect systems and also with the Wii remote control, although the last one is more limited.
4.
    a. A system for capturing movements in games.
    b. Home automation: controlling the in-home devices.
    c. Accessibility in mobile devices.
5. It would not work properly in outside environments with a high contrast level due to the low quality of the webcam.
6. Does not know/does not answer