Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 67 (2015) 263 – 272

Procedia
Computer Science

6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Infoexclusion (DSAI 2015)

Accessible options for Deaf people in e-Learning platforms: technology solutions for Sign Language translation

Paulo Martins[a,b,*], Henrique Rodrigues[b], Tânia Rocha[a,b], Manuela Francisco[c], Leonel Morgado[a,d]

[a]INESC TEC, Rua Doutor Roberto Frias 378, 4200-465 Porto, Portugal
[b]Universidade de Trás-os-Montes e Alto Douro, Quinta de Prados, Vila Real, 5000-801, Portugal
[c]Instituto Politécnico de Leiria, Rua General Norton de Matos, Leiria, 2411-901, Portugal
[d]Universidade Aberta, Palácio Ceia, Rua da Escola Politécnica n.º 141-147, Lisboa, 1269-001, Portugal

Abstract

This paper presents a study on potential technology solutions for enhancing the communication process for deaf people on e-learning platforms through translation of Sign Language (SL). Considering SL in its global scope as a spatial-visual language not limited to gestures or hand/forearm movement, but also to other non-dexterity markers such as facial expressions, it is necessary to ascertain whether the existing technology solutions can be effective options for the SL integration on e-learning platforms. Thus, we aim to present a list of potential technology options for the recognition, translation and presentation of SL (and potential problems) through the analysis of assistive technologies, methods and techniques, and ultimately to contribute for the development of the state of the art and ensure digital inclusion of the deaf people in e-learning platforms. The analysis show that some interesting technology solutions are under research and development to be available for digital platforms in general, but yet some critical challenges must solved and an effective integration of these technologies in e-learning platforms in particular is still missing.

Peer-review under responsibility of organizing committee of the 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2015)

Keywords: Inclusion, e-Learning, Sign Language, Deaf people, Assistive technologies.

* Corresponding author. Tel.: +351 259350332; fax: +351 259350356.
E-mail address: pmartins@utad.pt

Peer-review under responsibility of organizing committee of the 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2015)
doi:10.1016/j.procs.2015.09.270

1. Introduction

E-learning enables apprenticeship while using the Web as a dissemination channel. While it lowers many access barriers, especially geographical and time-related, as well as greater basis on written materials over spoken ones, by itself it does not solve the issue of enhancing inclusion of students regardless of their sensory, cognitive or functional abilities [1].

Specifically, people with hearing impairments, with partial or complete loss of hearing [2], may be prevented from taking part in live (synchronous) cooperative e-learning activities, due to the common reliance on voice communication in such activities. While speech-to-text systems provide an eventual pathway for the deaf people to follow the communication, it does not enable them to participate in the same mode. There are also issues with the use of text content, as will be put forward further ahead.

Not all deaf people have the same communication skills: some individuals communicate using oral language and/or writing and lip-reading, while others use SL. Some are bilingual deaf individuals that use both forms of communication [1] [3].

These communication skills are inevitably associated with hearing loss experience that involves factors such as family profile, which determines the way the deaf child is raised and the contact with SL since birth, the social context, the education inclusion policies that are specific to each country as well as the type of deafness and the psychological affectations that may result from that disability [4].

As suggested by several studies, deaf people that communicate through SL have more difficulties in the access to e-learning platforms because: SL is a visual-spatial language (the gesture recognition and hands/forearms movement are not enough for a correct translation, it is required the facial expressions interpretation); there is added the fact that it is not a "universal" language, but rather several independent ones, with distinct symbols, grammar, and syntax [1] [5] [6] from country to country (e.g., in USA is the American Sign Language - ASL, in Brazil it's the Brazilian Sign Language - Libras, in Portugal it's the Portuguese Sign Language - LGP); and also they present many difficulties in reading and writing texts.

These factors add complexity in the development of a truly inclusive technology that enables the correct SL translation to facilitate bidirectional communication between deaf and hearing, or even in the access to educational contents by deaf students.

This paper is based on a literature review of assistive technologies, methods, and techniques used as SL support, i.e., in the recognition of gestures and facial expressions, in order to allow an effective communication and interaction in e-learning platforms by the deaf people.

Aiming to analyze existing technology on the market and under research, we present a brief description of the latest significant features that are the most referenced in the literature.

It was our concern to divide the present work into six sections. First we introduce the paper in this Section 1. Then, in Section 2, we provide an overview about inclusion of the deaf students in e-learning platforms. In Section 3, we present the most commonly used assistive technologies solutions for SL recognition as input to a digital platform, as well some of the most applied techniques. After, as output from a digital platform, avatars as assistive technologies for sign language presentation are introduced in Section 4. In Section 5, we describe solutions for SL translation than can be applied from text/speech to SL or from SL to text/speech. Finally, in Section 6, we present some final reflections about this study.

2. Inclusion of deaf students in accessible e-learning platforms

When deaf people were first included in the educational process they were thought to have cognitive development difficulties and hence that their condition was associated with a cognitive problem. But the truth was that this happened due to the few stimuli received, the difficulty of communication between deaf and hearing people, and hence the difficulty in knowledge transmission [7].

It is known that cognition development is associated with language learning. This connection is so important that deaf disability is directly classified by the ability of speech and language, in four levels [8]: mild hearing loss (no significant effect on development, since generally the use of an hearing aid is not necessary); moderate hearing loss (may affect speech and language development, but not enough to prevent a person from talking); severe hearing loss (interferes with speech and language development, but with the use of an hearing aid may receive information, using hearing for speech and language development); and, profound hearing loss (without intervention, speech and language are unlikely to occur).

The deaf with severe or profound hearing loss use sign language as a means of expression and communication. Sign language is not a "universal" language but a concept. There are several sign languages, native languages of the deaf community of each country. However, in 1973, a global committee attempted to unify the various sign languages by proposing a standardized system called "Gestuno", meaning "the unity of sign languages". The proposed international sign language consisted of a selection of easy-to-learn gestures from various sign languages. The deaf community does not consider "Gestuno" a real sign language, since it was invented and customized and therefore it is rarely used. In fact, some linguists stressed that “Gestuno” cannot be considered a real language since it doesn't have a own grammar, i.e., introduces a set of vocabulary which is common to different sign languages and, as such, it is basically used in some international events [9].

Sign languages are visual-spatial languages that use gestures (hands) to produce linguistic information but also other non-dexterity markers, such as facial expressions (movement of eyes, mouth, head, eyebrows, etc.), which enrich language structure, adding grammatical elements [10]. This language configuration is very complex since it is formed by a huge amount of gestures, thousands per country, some of them very similar, being extremely difficult to distinguish each other. In addition, the same gesture may have a different meaning in each country or even within each country, since its meaning depends on the cultural context and the way in which it is taught (e.g., school vs. home/street).

Also the translation process (which involves the interpretation of the meaning in a language and the production of a rendering in another, as accurately as possible) can be very complex, because for one deaf person, one phrase is a sequence of gestures, movements and facial expressions, and the technological interpretation process to detect the start or end of a gesture, movement or facial expression, and the beginning of the next one, can be very difficult.

Furthermore, the majority of deaf people have reading and writing difficulties since they use spoken languages as a second language [11]. These aspects arise as a barrier to the digital content access that is presented mostly in text form, having also an interaction impact with the interface platform, with resources and with other users, since the vocabulary used in the Learning Management System (LMS) may not be understood by this kind of users. In this context - e-learning - digital inclusion and web accessibility issues gain significance.

Whereas universality is the fundamental premise of the Internet, access must be guaranteed to all citizens, regardless of their disability or user profile. Based on this assumption we found several approaches to accessible e-learning which may be dependent on the pedagogical model, technology, and the accessibility and inclusion policies followed by each institution [12]. One approach advocates a user-centred model where the platform's interface and the contents are presented according to the definitions applied in the user profile. Other approach is based on the pluralist paradigm: student-centered but focusing on diversity. That is, diversity of the content and of the educational resources, of pedagogical strategies, and technologies, creating a heterogeneous community [13]. Whatever the approach to achieve accessible e-learning, it is essential a strong reflection about technology and, as such, the adoption of accessibility standards to ensure the accessibility of the platforms’ interface as well as of the other web tools and web content used in virtual learning environments [14].

Currently, inclusive education is a reality in many countries, a fact highlighted in the Salamanca Statement [15]. There are some reference schools around the world, created for the bilingual teaching of deaf students, in order to concentrate human and material resources that provide an educational response with quality and equality for these students. However, these schools are concentrated in urban centers and deaf students are integrated in mainstream classes, coming upon barriers in the learning process, because not always the school and the teacher are prepared to attend these students [16]. Thus, it is essential to provide affordable solutions that combine the knowledge transmission and do not impose physical or technological barriers, where individuals with hearing loss can benefit from education likewise other students, emphasizing also the importance of the accessibility issue in web environments.

In the web accessibility field a wide range of problems can be found associated with learning platforms accessibility in web environments. Some of them are also found in the hearing community, namely: there are many graphic and action elements in the platforms interface which cause confusion; some disorganized or unstructured menus should be simplified and follow a logical and consistent navigation, showing just the most important functions [14].

It is also noted that the problem of translation into sign language is not restricted to the spoken or written language, which appears in the e-learning platform interfaces, in the educational resources or even in the communication between pairs that arises in the forums or in other technology mediated tools. Some content provided by teachers or students contains audible information such as music or beats that can be essential for the content message understanding. In the case of musicality, translation into sign language is even more complicated since it involves complex neurological processes that are triggered with the vibrations, the beats and rhythms. These are very hard to be transmitted in distance mode, a whole body interpreter is needed, or the use of colors or other strategies that can convey the purpose of sound used in the content [17].

All these facts highlight the importance and motivation for the study and development of solutions for deaf students in e-learning platforms. As noticed before, sign language translations can be a complex process, so we define a process with three steps that can be involved in the correct translation of sign language: first, recognition (we need a technology that can capture and interpret all gestures, movements and facial expressions); second, representation (normally made by an avatar); third, translation (from text/speech to sign language or from sign language to text/speech). These steps are detailed in the following sections.

3. Assistive technologies solutions for sign language recognition

For the sign language recognition, it must be taken into consideration the gesture recognition and the facial expressions. For gesture recognition there are two most prevalent approaches in the literature: one uses sensors (sensory gloves or miography armbands) and the other is based on the vision approach [18]. Both approaches have the same goal, to provide human-computer interaction through a communication channel. However, physical devices used in these approaches differs, in concrete, the gloves uses sensors and a sophisticated electromechanical integration with the user's body and the vision output devices uses video (moving image and audio, although in this particular case, audio does not apply, because our target audience is deaf people).

The gloves approach emerge when several technology companies worked in order to achieve the demands of the entertainment professionals (in particular in the field of animation) to create technology capable of stimulating various types of physical sensations, thus providing a more efficient human-computer interaction [19].

The most relevant companies are Fifth Dimension Technologies (5DT) and the CyberGlove Systems that, through interactive gloves, created virtual reality systems that can recognize the hand movements of the user, i.e., devices that, through sensors, detect and measure the fingers reflections and spacing between them.

Specifically, 5DT has several types of gloves that are differentiated by the number of sensors, the presence of a wireless kit and whether is or is not optimized for use in magnetic resonance environments. Plus, it includes a kit "plug-and-play", designed to transmit data with the two gloves simultaneously, meaning that a single bidirectional wireless transmitter and a battery can be used for a pair of gloves, saving space and energy and also can be used until four wireless kits (8 gloves) at the same time [20] [21].

CyberGlove Systems is also a global leader in interactive gloves technology and provides 3D motion capture solutions. These solutions allow capturing, in detail, finger, hand and arm movements which allows those who use and interact with digital objects virtually.

On the other hand, the second approach, based on vision, appeared to improve human-computer interaction, with the use of cameras to gestures capture. Microsoft and Leap Motion are the two leading companies in this field with several technologies based in this approach.

Microsoft Digits was created by Microsoft in Cambridge (UK) with the help of researchers from the University of Newcastle and the University of Crete [22]. Digits aims to replace the current use of gloves (this use can be uncomfortable) with technology with high precision and speed than the previous ones. Enables a new way to interact with the digital world, using intuitive hand gestures that incorporate an infrared camera, an infrared generator, diffuse lighting using infrared rays and inertial measurement unit, to know the accurate positions of each finger.

Microsoft Kinect was created initially for games, to compete with Wii from Nintendo and PlayStation 3 from Sony [23]. Kinect is a breakthrough in recent 3D depth cameras, where people are able to interact with games and use their own body in a natural way, being the key to understanding the technology for a human body language (the computer will have to "understand" what the user is doing before giving an answer). There are two Kinect versions, v1 and v2. Kinect v2 in terms of facial recognition, motion control, and resolution is much more accurate than Kinect v1. Kinect v2 uses Time-of-Flight (ToF) technology to determine the features and movements of particular objects. With the use of this technology, Kinect v2 can capture images efficiently, both in a completely dark room as in a well-lit room. Although the first version uses similar technology, Kinect v2 has suffered considerable improvements, since it offers a resolution of 1080p (HD).

Another important existing technology is the Leap Motion sensor, a depth sensor made especially to track the hands' features. David Holz, technical director of the Leap Motion company, and Michael Buckwald, co-founder, created a system that allows users to control a digital environment in the same way that objects are controlled in the real world [24]. The Leap Motion controller, associated with the current API, offers positions in Cartesian space of predefined objects, such as fingertips, pen tip, etc.

Recognition and interpretation of facial expressions are also fundamental in the sign language recognition. Existing technologies for this purpose are based on digital image processing and artificial intelligence where are applied techniques and mathematical models able to interpret the captured information [25]. The Microsoft Kinect is currently one of the most used technologies in capturing moving images.

However, several technological solutions have emerged, with or without the Kinect, but which are based on capturing images through one or more cameras.

Currently, researchers are focusing on adapting the models to three-dimensional scans of the face [26]. For this reason, the use of scanners capable of obtaining high quality 3D images is required. Currently, some companies offer solutions with very positive results, combining some technologies. Emotion Analysis, developed by Kairos company, offers a facial recognition of emotions and expressions through a simple webcam [27]. The solution that provides Affectiva company is also to be taken into account, offering the Affdex application that analyzes the different facial movements that can be undertaken and produces the interpretation of emotions from them [28].

3.1. Techniques for sign language recognition

There are several techniques that had been used for sign language recognition based on gesture recognition and facial expressions. The gesture recognition through computer processing, using the two approaches described above, based on sensory gloves or on vision approach, aims to obtain algorithms able to correctly interpret the information received.

However, several technologies have been developed and currently the sensory gloves were losing position in relation to vision approach, because with sensory gloves only the hand gesture recognition is performed, ignoring other non-manual variations that are included in sign language, like facial expressions, which do not allow the correct translation of sign language.

In literature, two techniques used in sensory gloves are referred. The first is based on the tracking of each glove (hand) movement [29]. It is a complex technique because it requires a very precisely motorization of each hand through image sequences taken within a determined period of time. The other technique is simpler and does not require such a precise tracking of movements because this technique is based on the appearance, i.e., the basis for recognition of sign language gestures is the hand location and its posture [29].

According to Viola & Johns [30], the most common techniques and processes, in most vision-based systems with the objective of gathering information, are: image acquisition - it is the first step of a vision-based system and is the process of acquiring images from video recording devices; image pre-processing - is made the objects identification (face and hands, for example) for extraction features; segmentation - regions of image points belonging to identified objects are isolated; feature extraction - drawing out mathematical characteristics (such as size, location, movement) of objects that compose an image sequence; and, rating - high-level processing that classifies objects by comparing segmented objects characteristics with classes of objects previously established.

In gestures recognition is important to analyze all image features in order to correctly identify the hands or other important elements to make the objects identification. Among used techniques for the body parts detection used in the representation of gestures, it is emphasized techniques based on skin color, in which the human skin is easily segmented when represented by colored images [11] [31]. These techniques based on skin color have advantages and disadvantages, if on the one hand are easier to process, on the other, are subject to erroneous analysis due to lighting variations that the image may be subject.

There are some comparisons between different techniques of Human skin detection in digital images. The values that correspond to the skin tone are explicitly defined in areas of known representation colors: RGB, YCbCr and HSV [32].

Among the techniques and methods used for *objects detection* the following are highlighted: AdaBoost algorithm based [33] [34] [35] uses the Viola & Jones method [30] [36]; the method Normalized Look-Up Table (LUT), according Vezhnevets [37]; and, the Bayes classifier, also indicated by Vezhnevets.

Another important feature to consider is the *motion detection*. A gesture is not necessarily something static, quite the contrary. It is necessary to detect hand movements to recognize a gesture. Cooper et al. [29] consider two essential techniques to detect movements: the tracking of the movement of the hand, where is needed an efficient movements processing through the sequence of images capture; and, the appearance of movement, where it is necessary to trace the movements based in the localization of salient poses, i.e., identifying the gestures of the sign language.

The last feature analyzed is the *gestures spatial and temporal variations*. The same gesture may differ from person to person and even the same person does it differently each time. This is an indicator that the same gesture may have differences in space and time and there is necessary to determine how to process a gesture, i.e., its sequence. This sequence can be determined by analyzing the sequences of images (frames) that constitute a gesture [11].

The techniques and methods commonly used to solve the problem of gestures spatial and temporal variations are: First-order Markov chain model and Hidden Markov model (HMM) [29]; Dynamic Time Warping (DTW) [11] [38]; Statistical Dynamic Time Warping (SDTW) [39]; Finite-state machine [40]; and, Neural networks [40].

The techniques used for facial expressions interpretation focus on two large groups. One uses automated learning techniques and the other algorithms based on mathematical models suitable to the problem. In the first group, among others techniques, stand out Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs). Specifically, ANNs use a computer model to acquire and increase their knowledge through experience. On the other hand, SVMs analyze the data and recognize patterns used for classification and regression analysis[†]. In the second group, stands out the Hidden Markov models (HMMs), based on the probability of occur a given expression (the most likely to occur will be taken into account) [25] [41]. The Deformable Model Fitting (DMF) algorithm is also used and was created as an adjustment to the obtained images from the Microsoft Kinect, to combat an existing problem: the noise of the depth images [42].

4. Avatars as assistive technologies for sign language presentation

Several studies have been conducted in order to provide an assistive technology for deaf people that could provide accessibility and usability in digital platforms for this group of people. In this context, there are different technologies in the area of information acquisition and information output. The state of the art reveals some uniformity in the acquisition of information, whether by capturing gestures or expressions or by the union between the two. In what regards the output information generated, it is found that, depending on the technology field, it is accomplished with customized avatars and/or embedded subtitles.

Avatars are widely used to communicate messages, whether broadcast by voice or gestures. San-Segundo et al. [43] developed a virtual representation of a person (agent), with a considerable level of detail that represents the gestures used in sign language. Also, several research centers have addressed projects to the creation of an agent who correctly can answer using the sign language representation (gestures recognition, hand and forearm movements, and facial expressions). Despite this massive research process in the area, there is a huge problem reported by researchers/developers, the construction of an animation [43]. For these researchers, the effort made in a gesture animation must be drastically reduced. The solution combine the avatar positions created by the developer and by the system, being the final animation generated through an interpolation of the combination of avatar positions, where previously is considered each point of their paths and their intervals of time [43].

Regarding the translation of information to be presented as sign language, as mentioned in the work of Araújo et al. [44], there are various proposals referenced. Some approaches translate spoken words, in a native languages, into sign language (i.e., use voice recognition), other approaches translate written texts into sign language (i.e., use of non-voice recognition) [45] [46] [47] [48] [49] [50]. In short, some gaps and limitations were shown by San-Segundo [43], where the proposals that use voice recognition have a specific domain and need a certain time to complete a translation (8 seconds per sentence), which makes impractical solution for domains in real time, such as television. On the other hand, proposals that do not use voice recognition have difficulty in developing their language constructs, which according to [44] is a non-trivial task and requires a lot of manual work. Adding to this difficulty arises other problem: sign languages are natural and evolve throughout the life, which implies an update on the grammatical construction. Sign language translation is not an easy task and, as pointed by Cooper et al. [51],

[†] In statistics, regression analysis is a statistical process for estimating the relationships among variables.

grammar and double meaning gestures make translation a extremely complicated process.

5. Sign language translation

There are technological contributions for sign language translation, such as academic projects or applications, which should be referred and identified their advantages and disadvantages. It is important to highlight digital platforms that perform recognition signals only by hands movement of those who uses not only hands but also facial expressions and body movements.

VirtualSign is a project involving researchers from the Higher Institute of Engineering of Porto and Universidade Aberta, Portugal, which aims to develop a sign language bidirectional translator that allows Portuguese Sign Language translation to Portuguese Language, specifically using text, and the inverse. Despite being an innovative project, it only recognizes six facial expressions: happiness, fear, anger, surprise, grief and neutral; this fact could lead to an incorrect translation [52].

MocapLab is a company that presents a project of sign recognition using motion capture. In detail, Mocap is a movement recording process and implementation of that movement in a digital model. This process records all movements performed by the hands, face and body, then the data is analyzed in detail and translated into text and/or sound. The major advantage is that the data collected is extremely accurate, however, require quite an effort in pre-processing, the hardware is expensive and motion capture has to be carried out on a controlled environment [53].

ProDeaf and HandTalk are two applications that have the ability to translate text and voice to LIBRAS, the Brazilian Sign Language, and also do the inverse process [54] [55]. These applications have the advantage that they may be used by people of all ages due to its easy usability. However, errors occurs when have to translate long texts. Added to this disadvantage both use the mobile phone's camera to capture the movements performed by the hands, neglecting facial expressions and body movements.

Google Gesture is an application that makes sign language translation to voice, through the use of an electronic bracelet that analyzes user’s muscle activity, hand and forearm position by electromyography, recognizing gestures made, translating them to voice. These devices, as mentioned previously, has the disadvantage of not perform a complete gesture analysis because does not has information on facial expressions and body movement [56].

6. Final consideration

Concerning accessibility of e-learning platforms, the deaf find various barriers, starting with interfaces excessive number of elements (links, images, titles, subtitles, etc.). Furthermore, words and expressions complexity and ambiguity presented to users, with or without hearing disabilities, in these platforms do not help to easy access and use. For the deaf, difficulties in access increase because of their reading and writing problems, as their first language is sign language and not a spoken language. Thus, it is imperative to remove all communication barriers that prevent deaf and hard of hearing users to interact properly through technology. The message to be transmitted has to reach the receiver with the same intent which it was sent.

Sign language translations can be a complex process. So, in order to better understand the process, it can involve three steps: first, recognition (the technology needed to capture and interpret all gestures, movements and facial expressions); second, representation (normally made by an avatar); third, translation (from text/speech to sign language or from sign language to text/speech).

In order to carry out effectively a bidirectional message transmission with digital platforms, several assistive technologies for gesture recognition have emerged, which once properly identified are translated into a desired language, either text or sound. Sensory gloves began to be the target of interest from researchers for gesture recognition. As advantages, these gloves offer high rates of capture that allows to identifying faster movements and gestures made by the user. Now add the possibility of gestures capture performed in any spatial orientation, i.e., without users need to hold their hand into a space reserved for a successful. Also, the absence of occlusion is another great advantage; there is no problem if a portion of the hand hides the other. As regards the disadvantages, the price is significant, it must be dressed, which can become uncomfortable, and cannot be adjusted correctly to the different hand sizes, which may result in erroneous data captured since the sensors are not right placed in the hand, so they will transmit the wrong information to be analyzed.

Notice in the sensory gloves approach, only the hand gesture recognition is performed, ignoring other non-manual variations that are included in sign language, thus do not allow the correct translation of sign language. In fact, it would be easier if gesture recognition for sign language were only limited by the movement of the hands and arms, however, the facial expressions, the movements speed, the slope of the head and the various postures that body can take, are also part of sign language and must be identified together with the hand gesture recognition, for a total and effective translation. Thus, the gloves approach to be used as a unique technology for sign language recognition is of little interest and an ineffective technology.

The vision approach, where the entire body can be captured by a camera, can be a solution for the correct translation of sign language. In this context, various devices can operate for this purpose: Kinect v1, Kinect v2 and Leap Motion. However, these devices despite overcome the limit barrier of capture (capturing gestures and expressions easily), are faced with some problems regarding gestures recognition, as: the amount of existing gestures is huge, in order of thousands, and some are very difficult to differentiate from others; because there are various sign languages, a gesture can be different in each language due to culture, the individual social life and the way gestures were taught (learned in school or acquired in the home/street); and, the sequence of gestures to express a sentence can be difficult to compute because it is difficult to detect where the gesture starts and ends, and begins the next.

Facial expressions are the most direct way to recognize emotions. However, it is hard work recognizing each human characteristic expression, being estimated that there are close to 20,000. Several technology companies have developed solutions that can recognize facial expressions; these solutions use different methods and techniques, but have one element in common: image capture devices. Kinect is the most commonly reported in the literature, however, a simple camera associated with a good mathematical model is capable of producing satisfactory results, but sometimes the image noise is a problem. Currently, the use of scanners capable of obtaining 3D images, using structured lasers systems, produces better results. With this technology high-quality images are achieved for analysis, however the high cost and the time taken to perform a scan are negative points to consider.

Concerning the techniques studied for the gesture recognition, we highlighted three: objects detection, motion detection, and gestures spatial and temporal variations. First, in objects detection, two techniques are more relevant, the one based on skin color and the other based on shapes of the body parts tracked; second, in motion detection, uses hands movements tracking or their movements appearance; and, at last, in gestures spatial and temporal variations, the most applied are Markov models.

The most used techniques to interpret facial expressions are: automated learning (ANNs and SVMs), mathematical models (HMMs) and the DMF algorithm. The literature reports that the use of these techniques is satisfactory. However, SVMs stand out by the strong theoretical foundation, based on the statistical theory, which distinguishes it from the neural networks that have no theoretical model. Another positive feature of the SVMs are the ability to work with high-dimensional patterns, which is ideal when you have a large number of patterns to be analyzed. On the other hand, there are also some disadvantages in relation to the ANNs, such as: the classification speed can be lower than the neural network; it requires much computational complexity; and the acquired knowledge is not easily interpreted. Regarding to the mathematical models, Markov ones are the most used. However, current studies already reporting the use of DMF algorithm associated with the Kinect. This combination has had positives results.

Aiming to overcome the presented difficulties, a technological solution is required to provide, simultaneously, accuracy on gesture recognition, high processing speed, support for real-time applications, and ease of scalability to withstand thousands of existing gestures. This will be the ultimate goal to achieve.

The use of avatars is present in several solutions on the market. Its use is reflected in the voice or text translation for an animated representation of the sign language. Some gaps were detected in the animation performed by avatar movements. It was also noticed that facial expressions were neglected, i.e., the solutions give more emphasis to the hands, arms and forearms movements. When facial expressions were represented, they only translated basic emotions such as: sadness, happiness, irritation, neutral, etc.

The text and voice translation to sign language of an avatar is well explored by technology companies and higher education institutions. The text/voice translation is basically achieved by a process in which the text/voice is received as input, then properly transformed by the computer system and finally translated to movements (sign language signals) by the avatar. During this process we can lose essential characteristics of a language such as syntax and semantics. This might lead to wrong avatar movements and consequently an incorrect translation. Another problem arises given that sign language is the first language for the deaf people and because of that they have difficulty understanding certain words expressed by a non-deaf person. Although this problem is more common in an input reading task, it can also be inherent for the output avatar task, because it will have to find a solution in order to conduct a proper translation to the input received, which will involve having a large number of synonyms.

It is highlighted two types of sign language translation solutions: bidirectional (between deaf and hearing) and unidirectional (from deaf to hearing or from hearing to deaf). A translation can be done by several ways: voice, avatar, just text or a combination of the previous.

With avatars solutions it is possible to present a translation of the hands, arms and forearms movement along with facial and body expressions. For example, we presented Virtual Sign and MocapLab projects. Still, these projects have limitations: Virtual Sign only translates six facial expressions and MocapLab is expensive and the capture has to be performed in a controlled environment. Other solutions such as ProDeaf and Google Gesture HandTalk are also subject of interest because of their ease of use. However, these solutions favor the hand movements, not taking into account the facial expressions and other body movements.

We conclude this analysis believing that some interesting technology solutions are under research and development to be available for digital platforms in general, but have yet some critical challenges to solve and an effective integration of these technologies in e-learning platforms in particular is still missing. Also, there is still no immediate solutions to solve synchronous real-time communication between deaf and not deaf.

In relation to acquisition (gloves and image capture) and presentation (avatars) it seems that the second one may be rapidly developed to become feasible and can be well exploited for asynchronous communication in e-learning platforms.

Although it is difficult to translate texts and sound for sign language in real time, it would make sense to create tools that allow content editing and synchronization in publication, making content accessible to this audience.

Not only the content but also the interface itself should be adapted and simplified for deaf so that it is possible to integrate (or remove) display options for sign language translation.

References

1. Pivetta E, Saito D, Ulbricht V. Deaf and Accessibility: Analysis of a Virtual Learning Environment. *Rev. Bras. Ed. Esp. Marília* 2014;**20**(1):147-162.
2. WHO - World Health Organization. ICD-10: International Statistical Classification of Diseases and Related Health Problems - 10th Revision. Chapter XXI - Factors influencing health status and contact with health services (Z00-Z99). Available: http://apps.who.int/classifications/icd10/browse/2015/en#/Z82.2.
3. Ladd P. Understanding Deaf Culture: In Search of Deafhood. Clevedon: Multilingual Matters; 2003.
4. Neves J. Audiovisual Translation: Subtitling for the Deaf and Hard-of-Hearing. Unpublished doctoral thesis, Roehampton University, United Kingdom; 2005.
5. El-Soud M, Hassan A, Kandil M, Shohieb S. A Proposed Web Based Framework E-Learning and Dictionary System for Deaf Arab Students. *International Journal of Electrical & Computer Sciences IJECS-IJENS* 2010;**10**(1):56-68.
6. Ottaviano S, Merlo G, Chifari A, Chiazzese G, Seta L, Allegra M, Samperi V. The Deaf and Online Comprehension Texts, How Can Technology Help? *Computers Helping People with Special Needs. Lecture Notes in Computer Science* 2010;**6180**:144-15.
7. Padden C, Humphries T. Inside Deaf Culture. Cambridge, MA: Harvard University Press; 2006.
8. WHO - World Health Organization. ICD-10: International Statistical Classification of Diseases and Related Health Problems - 10th Revision. Available: http://trigramas.bireme.br/cgi-bin/mx/cgi=@1?collection=CID10p&maxrel=10&minsim=0.30&text=Surdez.
9. Schein J, Stewart D. Language in Motion: Exploring the Nature of Sign. Washington, DC: Gallaudet University Press; 1995.
10. Emmorey K. Processing a dynamic visual-spatial language: psycholinguistic studies of American Sign Language. *Journal of Psycholinguistic Research* 199: **22**(2):153-187.
11. Han J, Awad G, Sutherland A. Subunit Boundary Detection for Sign Language Recognition Using Spatio-temporal Modelling. The 5th International Conference on Computer Vision Systems; 2007. Available: http://biecoll.ub.uni-bielefeld.de/volltexte/2007/59/pdf/ICVS2007-29.pdf.
12. Brown J, Mirri S. E-learning Accessibility. W3C; 2013. Available: http://www.w3.org/WAI/RD/wiki/E-learning_Accessibility.
13. Francisco M, Neves J, Sousa N, Cadima R, Esperança C, Rodrigues V, Jorge N, Rodrigues C, Mineiro J, Costa S, Maximiano C. Accessible e-learning - practices and research in the Polytechnic Institute of Leiria (Portugal). Accessible E-Learning Online Symposium; 2013. Available: http://www.w3.org/WAI/RD/2013/e-learning/paper2/.
14. Mirri S, Salomoni P, Roccetti M, Gay GR. Beyond Standards: Unleashing Accessibility on a Learning Content Management System. *Springer Transactions on Edutainment* 2011;**V**(LNCS 6530):5-49.
15. UNESCO - United Nations Educational, Scientific and Cultural Organization. Salamanca Statement on Principles, Policy and Practice in Special Needs Education and a Framework for Action. 1994.
16. Gibson H, Small A, Mason D. Deaf Bilingual Bicultural Education. Encyclopedia of Language and Education: Bilingual Education, London, UK: Kluwer Academic Publishers; 1997.
17. Neves J. Music to my eyes… Conveying music in Subtitling for the Deaf and the hard of hearing. In: Lukasz B, Kredens K, editors. *Perspectives in Audiovisual Translation*, Lódz Studies in Language; 2010. p. 123-143.
18. Wang H, Leu MC, Oz C. American Sign Language Recognition Using Multi-dimensional Hidden Markov Models. *Journal of Information Science and Engineering*. 2006;**5**(22):1109-1123.

19. Nielsen J. Usability 101: Introduction to usability. 2012.
20. CyberGlove Systems. Available: http://www.cyberglovesystems.com/index.php.
21. Fifth Dimension Technologies. Available: http://www.5dt.com/products/pdataglove5u.html.
22. Microsoft Digits. Available: http://research.microsoft.com/en-us/projects/digits/.
23. Microsoft Kinect. Available: http://www.microsoft.com/en-us/kinectforwindows/.
24. Leap Motion. Available: https://www.leapmotion.com/.
25. Silva VAS. Extração de Atributos para Reconhecimento de Expressões Faciais. 2007.
26. Zhang Z. Microsoft kinect sensor and its effect. *IEEE MultiMedia* 2012;**19**(2):4-10.
27. Emotion Analysis. Available: https://www.kairos.com.
28. Affdex. Available: http://www.affectiva.com.
29. Cooper H, Ong EJ, Pugeault N, Bowden R. Sign language recognition using sub-units. *The Journal of Machine Learning Research* 2012;**1**(13):2205-2231.
30. Viola P, Jones M. Robust real-time object detection. *International Journal of Computer Vision* 2001;**4**:34-47.
31. Carneiro A. Cortez P. Costa R. Reconhecimento de Gestos da LIBRAS com Classificadores Neurais a partir dos Momentos Invariantes de Hu. Anais da Interaction South America 09. São Paulo: 2009.
32. Nascimento A, Mendonça M, Denipote J, Paiva M. Comparação de Clusters para Detecção da Pele. Biblioteca Digital Brasileira de Computação. 2007.
33. Ong E, Bowden R. A boosted classifier tree for hand shape detection. Proceedings of the Sixth IEEE international conference on Automatic face and gesture recognition. p. 889- 894; 2004.
34. Francke H, Ruiz-del-Solar J, Verschae R. Real-time Hand Gesture Detection and Recognition using Boosted Classifiers and Active Learning. Second Pacific Rim Symposium, PSIVT 2007. Santiago, Chile Proceedings: p. 533-547; 2007.
35. Kadir T, Bowden R, Ong E, Zisserman A. Minimal Training, Large Lexicon, Unconstrained Sign Language Recognition. British Machine Vision Conference. 2004.
36. Viola P, Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features. Conference on Computer Vision and Pattern Recognition. 2001.
37. Vezhnevets V, Sazonov V, Andreeva A. A survey on pixel-based skin color detection techniques. *Proc. Graphicon* 2001;**3**:85-92.
38. Keogh E, Pazzani MJ. Derivative Dynamic Time Warping. *International Conference on Data Mining (SDM'2001)* 2001;**1**,:5-7.
39. Lichtenauer JF, Hendriks EA, Reinders MJT. Sign Language Recognition by Combining Statistical DTW and Independent Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2008;**30**(11):2040-2046.
40. Mitra S, Acharya T. Gesture recognition: A survey. IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and reviews. National Center for Biotechnology Information: v. 37, n. 3; 2007.
41. Osuna E, Freund R, Girosi F. Support vector machines: Training and applications. Technical Report. MIT, Cambridge: 1997.
42. Cai Q, Gallup D, Zhang C, Zhang Z. 3D deformable face tracking with a commodity depth camera. Proceedings of the 11th European conference on computer vision (ECCV'2010). Springer Berlin Heidelberg: Part III, p. 229-242, 2010.
43. San-Segundo R, Montero JM, Macías-Guarasa J, Córdoba R, Ferreiros J, Pardo JM. Proposing a speech to gesture translation architecture for Spanish deaf people. *Journal of Visual Languages & Computing* 2008;**5**(19):523-538.
44. Araújo TMU, Ferreira FLS, Silva DANS, Oliveira LD, Falcão EL, Domingues LA, Martins VF, Portela IAC, Nóbrega YS, Lima HRJ, Filho GLS, Tavares TA, Duarte AN. An approach to generate and embed sign language video tracks into multimedia contents. *Information Sciences* 2014;**281**:762-780.
45. Veale T, Conway A, Collins B. The challenges of cross-modal translation: English to sign language translation in the Zardoz system. *Machine Translation* 1998;**13**:81-106.
46. Zhao L, Kipper K, Schuler W, Vogler C, Badler N, Palmer M. Machine translation system from english to american sign language. *Lecture Notes in Computer Science* 2000;**1934**:54-67.
47. Fotinea SE, Efthimiou E, Caridakis G, Karpouzi K. A knowledge-based sign synthesis architecture. *Universal Access in the Information Society* 2008; **6**:415-418.
48. Huenerfauth M. Generating american sign language animation: overcoming misconceptions and technical challenges. *Universal Access in the Information Society* 2008;**6**:419-434.
49. Anuja K, Suryapriya S, Idicula SM. Design and development of a frame based MT system for English-to-ISL. World Congress on Nature & Biologically Inspired Computing (NaBIC'2009). Coimbatore, India: p. 1382-1387; 2009.
50. San-Segundo R, Montero JM, Córdoba R, Sama V, Fernández F, D'Haro LF, D. Sánchez VLL, Garcia A. Design, development and field evaluation of a Spanish into sign language translation system. *Pattern Analysis and Applications* 2011;**15**(2):203-224.
51. Cooper H, Holt B, Bowden R. Sign language recognition. In: Visual Analysis of Human. Springer: London. p. 539-562; 2011.
52. Escudeiro P, Escudeiro N, Reis R, Barbosa M, Bidarra J, Baltazar AB, Gouveia B. Virtual sign translator. International Conference on Computer, Networks and Communication Engineering (ICCNCE'2013). China, Beijing: p. 290-292; 2013.
53. MocapLab. Available: http://www.mocaplab.com.
54. ProDeaf. Available: http://prodeaf.net.
55. HandTalk. Available: http://handtalk.me.
56. Google Gestures. Available: http://add.berghs.se/case/google-gesture.