

International Telecommunication Union

ITU-T Technical Paper

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(3 June 2016)

FSTP-UMAA Use cases for assisting persons with disabilities using mobile applications

ITU-T

Summary

The mobile phone industry has evolved rapidly in the last couple of years. Everyone is now capable of buying a smartphone that could make his life different. Smartphones could be very helpful for persons with disabilities (PWDs). They can act as enabler for persons with disabilities to assist them in their daily lives by providing to some extent capabilities for the senses that are missing. . This technical paper presents some applications on smartphones that can facilitate some of daily life tasks of PWDs.

Keywords

Accessibility; Mobile; Application; Disability; Money; Detector; Colour; Tracking; Indoor; Access; Audio; Video; Subtitling; Captioning; Audio; Description.

Change Log

This document contains Version 1 of the ITU-T Technical Paper "Use cases for assisting persons with disabilities using mobile applications" approved at the ITU-T Study Group 16 meeting held in Geneva, 23 May – 3 June 2016.

Editor: Mohannad El-Megharbel
NTRA
Egypt

Tel: +20-100-803-6503
Fax: +20-2-3534-4133
Email: melmegharbel@tra.gov.eg

CONTENTS

	Page
1	Scope..... 1
2	References..... 1
3	Definitions..... 1
3.1	Terms defined elsewhere 1
3.2	Terms defined in this document 1
4	Abbreviations and acronyms..... 1
5	Conventions 2
6	Use cases of mobile accessibility..... 2
6.1	Indoor tracking..... 2
6.2	Money detector 2
6.3	Colour detection..... 4
6.4	Accessing audio visual content..... 5
6.5	Standardization of sign language communication through avatars and mobile phones 6
7	Requirements 6
7.1	Requirements for colour detection..... 6
Appendix I	Information on RGB and HSL colour models 7
I.1	RGB 7
I.2	HSL..... 7
Appendix II	Standardization of sign language communication through avatars and mobile phones 8
II.1	Introduction..... 8
II.2	Presentation of MMS Sign..... 8
II.3.	What should be standardized? 9
Bibliography 12

List of Tables

	Page
Table II.1: A set of languages for describing gestures	10

List of Figures

	Page
Figure 6-1: Identifying money with a mobile phone.....	3
Figure 6-2: Block diagram of the money detection process.....	3
Figure 6-3: Block diagram of colour detection process	4
Figure 6-4: Using a mobile phone for identifying colour	4
Figure 6.5: Sequential screens for ArtAcc��s application.....	5
Figure II-1: General architecture.....	9

Technical Paper ITU-T FSTP-UMAA

Use cases for assisting persons with disabilities using mobile applications

1 Scope

This Technical Paper illustrates the use cases for assisting persons with disabilities using mobile phones. The idea starts from using the mobile phone as a digital assistant for the person with disability (PwD). It should allow PwDs to dispense many hardware assistance tools and act as an enabler to fill in the capabilities of all the senses that are lacked.

These applications are designed to perform specific function(s) that are essential for persons with disabilities to accomplish the activities of their daily lives spanning from reading text, walking on the road alone, identifying colour and money.

2 References

- [ITU-T E.135] ITU-T E.135 (1995), *Human factors aspects of public telecommunication terminals for people with disabilities*.
- [ITU-T F.791] ITU-T F.791 (2015), *Accessibility terms and definitions*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

- 3.1.1 accessibility** [ITU-T F.791]: The degree to which a product, device, service, or environment (virtual or real) is available to as many people as possible.
- 3.1.2 accessibility feature** [ITU-T F.791]: An additional content component that is intended to assist people hindered in their ability to perceive an aspect of the main content.
- 3.1.3 disability** [ITU-T F.791]: Any restriction or inability to perform a function or activity in the manner or the range considered average or accepted functionality, resulting from an impairment or reduction of ability, which can either be permanent or temporary.

3.2 Terms defined in this document

This Recommendation defines the following terms:

- 3.2.1 app**: a self-contained computer program or piece of software designed for a particular platform (usually, a mobile phone) to provide a specific set of functionalities.
- 3.2.2 mobile accessibility**: a set of accessibility features provided in mobile phones (including smartphones).

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AD	Audio description
HSL	Hue, saturation and lightness
NFC	Near-field communication
RGB	Red, green and blue
TTS	Text to speech

5 Conventions

None.

6 Use cases of mobile accessibility

6.1 Indoor tracking

6.1.1 Overview

Currently, visually impaired persons are using many different devices to substitute their disabilities and still these devices do not provide for them full independence. This application enables the blind to be independent and provides them with all the necessary functions in only one device, their mobile phones. It assists the blind to identify the objects around him whether at home or in any public place such as supermarkets, hotels, or schools.

The app in this use case relies on reading tags that are placed everywhere using near-field communication (NFC). These tags can hold much useful information as the product name, its price, expiry date, colour, etc.

The most powerful feature of these tags is that it supports multiple languages, so using the application blind persons can extract the information in their language.

6.1.2 Process description

Integrating different technologies is useful to develop new applications that can add value to people's lives. The indoor tracking app uses the integration between the mobile devices, NFC technology and text to speech (TTS) conversion to facilitate the daily activities of blind persons.

The indoor tracking process is divided into two steps; the first step is attaching the NFC tags to the objects or products, by writing all the needed object's information NFC tags, and placing the tag in the object vicinity.

The blind person initiates the second step by pointing the mobile device near to the object, the App act as an NFC reader that starts reading all the information stored on the tag, and then use the TTS engine to inform the blind with all the recalled information.

6.1.3 Interaction scenario

Mohammad used to go to the supermarket to make shopping with his friend Ahmad. Mohammed usually feels embarrassed because his friend Ahmad spends a lot of time helping him. Mohammad decided to use the new application for indoor tracking.

He went to an accessible supermarket and used his mobile to recognize nearby products, their properties, and prices. The application also helped him to recognize the floor map of the supermarket and guided him in his shopping. This time he felt proud of himself and independent of others.

6.2 Money detector

6.2.1 Overview

One of the main obstacles that face blind persons is making cash payments to others. This problem is due to the difficulty of identifying the value of the paper currency they hold. Another problem arises from the quality of the paper currency (torn, old, etc.). So we developed software that can be used to help the blind when making payments. This software can work with different kinds of currencies and in different environments (daylight or dark night).



Figure 6-1: Identifying money with a mobile phone

6.2.2 Process description

For a blind person it is important to make the currency detectable even under changes in image scale, noise, orientation and illumination. This imposes using algorithms that extract interesting points on the currency to provide a "feature description" of it.

The money detector software is a mobile app that substitutes the eyes of the blind person; it uses an algorithm that extracts the local features vectors from a set of reference currencies and stores them in a database. When a user points the mobile camera to a paper currency, the software captures a live screenshot of the currency and compares each feature of the new image with data in the database, and finds candidate matching features based on Euclidean distance of their feature vectors.

From the full set of matches, subsets of key points that agree on the image and its location, scale, and orientation in the new image are identified to filter out good matches. Finally, the probability that a particular set of features indicates the presence of an image is computed. Image matches that pass all these tests can be identified as correct with high confidence. The value of the currency is fetched from a lookup table. This lookup table maps each currency image to its corresponding value. Using a TTS engine, the blind person is informed by the currency value. This whole process is done on the spot in a very short time that the blind person would feel like it is live direct view of the real world.

Future work will consider additional features such as support for Braille attached devices.

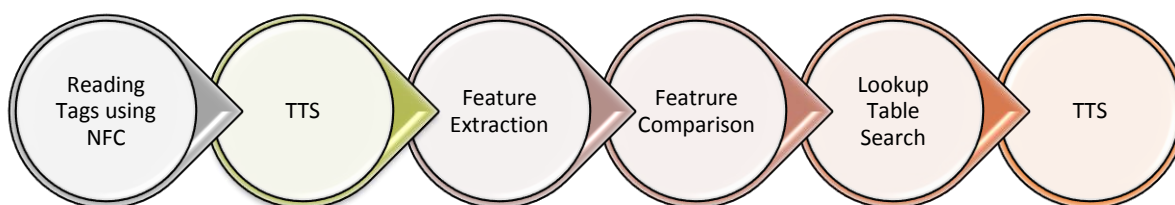


Figure 6-2: Block diagram of the money detection process

6.2.3 Interaction scenario

Ahmad decided to take a taxi, when he arrived at his destination the driver told him the fare is EGP 25. He picked up from his wallet a paper currency and directed the mobile to it, then the mobile told

him that this is an EGP 50 paper currency. Ahmad gave it to the driver who handed back to him the remaining change, an EGP20 and an EGP5 notes. Ahmad used the mobile phone to verify that he received the right amount.

6.3 Colour detection

6.3.1 Overview

Although colours may not have significance to a number of situations blind persons face, the ability to recognize different colours can be helpful. This app aims to help blind persons in detecting colours, using a very simple interface and the TTS, the phone acts as an assistant for detecting the colour of the object used by the blind persons in their daily life.

6.3.2 Process description

The process of colour detection (see Figure 6-3) starts with image acquisition using the mobile phone camera. Then, the image segmentation process is implemented using HSL and the image histogram is created. This histogram is feed to a fuzzy logic processor to detect the colour. Then, the TTS engine converts the colour information into speech for the blind person.

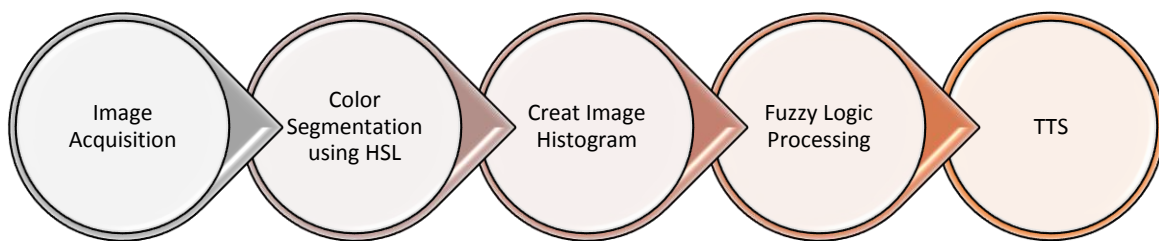


Figure 6-3: Block diagram of colour detection process

6.3.3 Interaction scenario

Mohammad wants to find the best pullover to wear over his black pants, so he starts the program and points the camera to his set of pullovers to detect the colour of each of them, see Figure 6-4. The program tells him the colour of each article, and he finds the green one which he thinks will be the best choice.

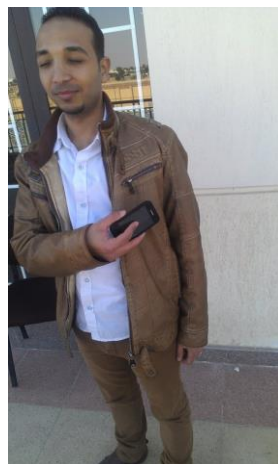


Figure 6-4: Using a mobile phone for identifying colour

6.4 Accessing audio visual content

6.4.1 Overview

New smartphone apps are being developed to increase accessibility to media content. Many applications include three accessibility services which can be used either independently or in any combination. These accessibility services are: audio description [ITU-T F.791], audio subtitling/captioning [ITU-T F.791] and subtitling/captioning [ITU-T F.791].

6.4.2 Process description

The app downloads, stores, and plays or displays accessibility services. It synchronizes with the original media source at any time during play, through different technologies such as fingerprinting or sound stamp.

The app picks up the movie soundtrack using the smartphone microphone and then is synchronised with the accessibility services at the required moment. Navigation screens also change but they look similar such as illustrated in Figure 6.5.

6.4.3 App functionality

Add auto play functionality: Once a user has selected and started a film/ TV programme of their choice on the VOD service and loaded the app on the phone/tablet, the app would recognise the media asset, pull the related accessibility track from the library within the application, download it to the phone/tablet or stream the accessibility track after identifying the exact point in the play out.

Two accessibility services should be allowed to play at the same time, since there are persons who require audio description (AD) with audio subtitling/captions.

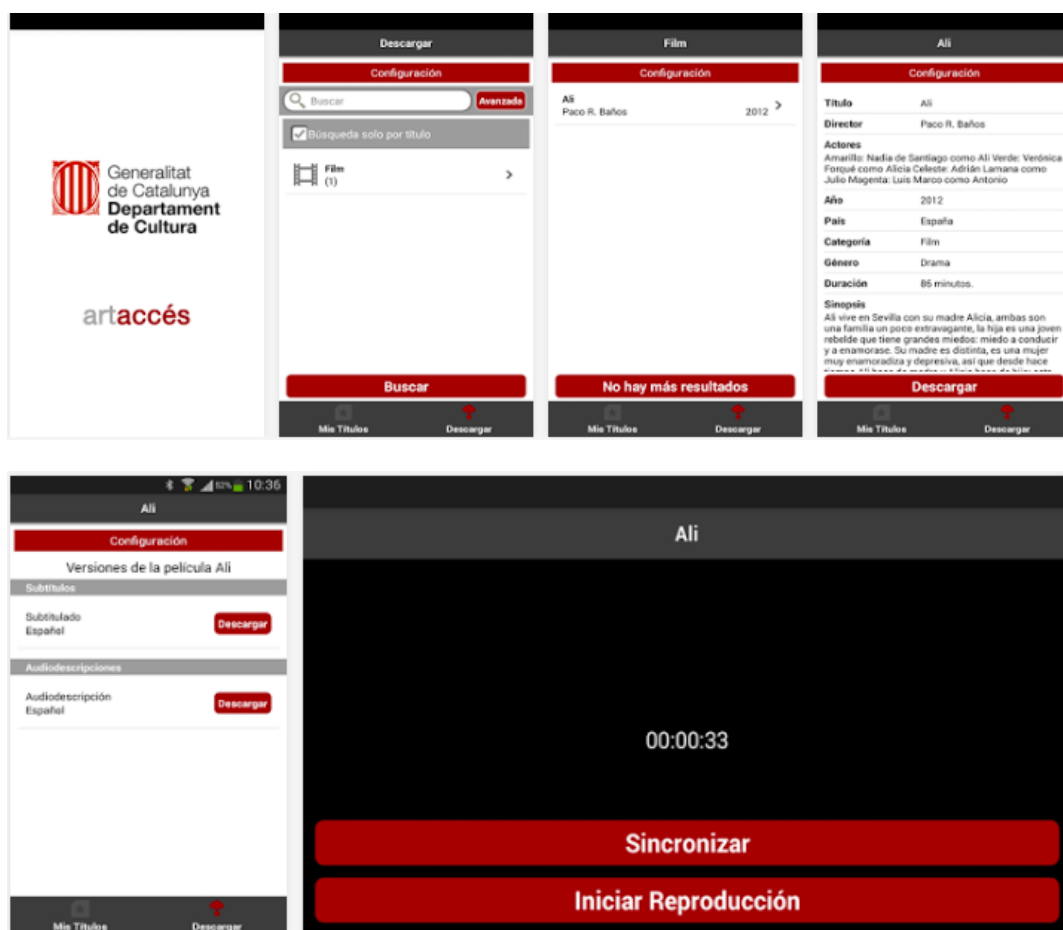


Figure 6.5: Sequential screens for ArtAccés application

- Application navigation
 - Add a search field within the app, which would simplify the process of looking for a title (film or TV).
 - Airplane mode should automatically be turned on when the app had been opened on the device to avoid receiving calls and disrupting the film/ TV programme.
 - One-to-one sessions should be provided to train persons to use the app or else, the option for an in-depth user manual should be made available to users.
- Quality of the sound mix in audio description
 - Sound mix of audio description track and original sound asset should be taken into consideration at two extreme situations: silence and very loud.
 - The AD track should not be put in stereo mode as the blind and partially sighted viewers use only one earplug for listening to it (they also have to follow the original soundtrack of the film played in the movie theatre).
- Synchronisation
 - Synchronisation should take into consideration that some cinemas do not have wireless connection because of piracy laws. Alternative synchronisation should be taken into consideration.
- Interface
 - Interface should be simple to navigate: buttons arrangements, should not be close to each other, colour contrast in the texts should be considered.
- Screen
 - The screen should be black to limit luminosity and avoid interference with other persons. A black screen also guarantees longer battery life.
- Text for subtitles/captions
 - Subtitles/captions should be in a colour contrasting with the black background.
 - Subtitles/captions should be as large as possible to fit with the size of the screen.

6.5 Standardization of sign language communication through avatars and mobile phones

Appendix II contains some considerations for the synthetic generation of sign language communication on mobile phones that were developed at the ITU-T Focus Group on Bridging the Gap: from Innovation to Standards (FG Innovation)¹.

7 Requirements

7.1 Requirements for colour detection

The Hue, Saturation and lightness (HSL) colour space is suggested to be used in the colour detection process for mobile apps, rather than the Red, Green and Blue (RGB), because of HSL's two prominent properties that facilitate the colour detection in different light conditions and without having a requirement of focusing the camera on the detection area:

- Increasing lightness controls the symmetric transition from black to a hue to white
- Lightness controls the decrease of saturation transitions to a shade of grey, thus keeping the overall intensity relatively constant

¹ Home page: <https://www.itu.int/en/ITU-T/focusgroups/innovation>.

Appendix I

Information on RGB and HSL colour models

I.1 RGB

Red, Green and Blue (RGB) colour model is mainly used in computer graphics applications. The RGB colour model is generally built based on Cartesian coordinates. In RGB each colour appears in its primary spectral component of red, green and blue.

The typical range of intensity values for each colour, 0-255, is based on taking a binary number with 32 bits and breaking it up into four bytes of 8 bits each. The fourth byte is used to specify the "alpha", or the opacity, of the colour. Opacity comes into play when layers with different colours are stacked. If the colour in the top layer is less than fully opaque ($\alpha < 255$), the colour from underlying layers "shows through", producing a combined colour.

I.2 HSL

Hue, Saturation and lightness (HSL) colour model are equivalent to human thinking. HSL is the most common cylindrical-coordinate representations of points in an RGB colour model. Developed in the 1970s for computer graphics applications, HSL is used today in colour pickers, in image editing software, and less commonly in image analysis and computer vision.

Based on the standard representation of the HSL colour model, hue represents the original colour itself. Saturation denoted the purity of the colour that is related to the standard deviation around the dominant wavelength. It indicates the range of grey value of the colour, whereas lightness represents the lightness of colour. It described how light or dark the colour is and ranges from 0 (dark) to 1 (bright). It is related to the amount of white in colour.

Appendix II

Standardization of sign language communication through avatars and mobile phones

II.1 Introduction

In recent years, modelling and animating virtual characters become more and more important. The first work of standardization for synthesizing 3D models was initiated in 1997 by the VRML group. The specificities of the animation models of humanoids have been made by the Humanoid Animation Working Group (H|Anim)². It defined the geometry of the human body constituted as a hierarchical structure.

The MPEG4 standard developed by MPEG (ISO/IEC JTC1 SC29/WG11) [b-MPEG4] provides a common framework for multimedia applications by defining complex scenes of audio-visual content and 2D and 3D text or graphic objects, to ensure the compression and transmission of such scenes. It was presented at the end of 1998 and designated as the official ISO / IEC 14496. MPEG4 provides the specification Face and Body Animation (FBA) which sets up a precise representation of the geometry and animation of a virtual character.

II.2 Presentation of MMS Sign

II.2.1 Objectives and problem statement

The main objective is to contribute to the improvement of e-accessibility of the deaf community via the mobile phone. In this context, a new system to transform the text to MMS containing the video sequence of the text in sign language in a format which can be sent to deaf persons through MMS. This transformation is realized in two steps:

- The first consists on the generation of 3D animation of a 3D avatar, by the use of WebSign.
- Then, the MMSSign software transforms the 3D animation to an MMS containing a 3GP video sequence.

The main problems to be overcome in order to realize this ambitious project are:

- Contrary to popular belief, unfortunately, Sign Language (SL) is not universal. Wherever community of deaf persons exists, a specific SL has been developed. Likewise spoken languages, SLs vary from region to region. Many SLs exist in the world and are at the core of local deaf cultures. Some SLs have obtained some forms of legal recognition, while others have no status at all.
- Linguistic treatment: the automatic translation of written text to sign language cannot be considered as a word-to-word translation. As all natural language processing applications, it requires a lot of information: lexical, grammatical, translation equivalences, etc. The most difficult aspect remains the semantic analysis.
- Technologic problems: the animation of the avatar needs sophisticated programming skills.
- Cost problem: the construction of dictionaries of signs is an expensive operation and needs a lot of time.
- The mobile phone does not contain a 3D accelerator of the 3D scene and it does not contain a big memory to allow the animation of 3D scene.

² Website visited 2016-11-22, <http://h-anim.org/Specifications/H-Anim1.1/>.

II.2.2 General approach of MMS SIGN

MMSSign is a converter of textual message to a video sequence in sign language. This system offers to deaf the possibility to use cell phones. In fact, in order to communicate with a deaf person via cell phone, anyone can send him a textual message via SMS or via a dedicated web interface (Figure II-1 (step 1,7)). The text of the message is then processed by our system and precisely by our MMSSign server. If MMSSign server does not respond to the request in a reasonable period of time the server of the operator should return a negative acknowledgement of delivery (Figure II-1 (step 6)).

In the normal work of the procedure, MMSSign server should return an MMS containing the video transformation of the text in sign language (Figure i (step 3)) and the operator of the telecommunication becomes the responsible of the deliverance of the MMS.

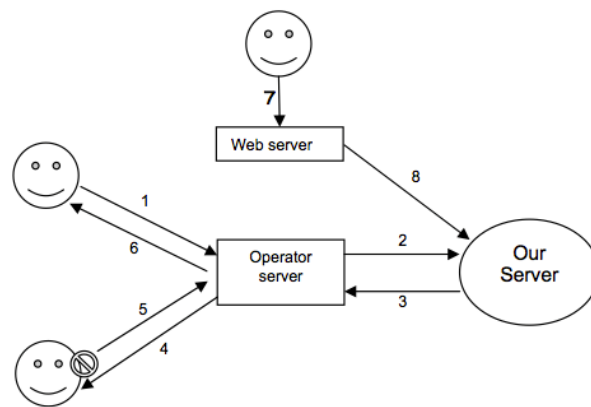


Figure II-1: General architecture

II.3. What should be standardized?

It is possible to standardize:

- Sign language transmission and rendering engines in order to guarantee compatibility between sign languages communication systems.
- The description language for encoding avatar gestures for sign languages.
- Some 3D animation requirements to assure a minimum degree of accessibility to the rendered animation.

II.3.1 Encoding sign languages

The need to create languages to describe the gestures and facial expressions made by virtual characters has encouraged many researchers to propose languages to meet these needs. Indeed, the needs are often based on a set of six criteria: recess body animation, facial animation and sign language support in the first place and emotion, dialogue and respect of [b-MPEG4] in second place.

Many languages are proposed in order to describe avatar animation. However, they are still insufficient to ensure sign language animation needs. In sign languages, gestures are not a simple animation of body and face. The sign should respect many linguistic and physical rules. In fact, a sign language text is a real show, in which, the speaker describes objects, their places in the scene, their shapes, and the relation between them.

For these reasons, it is important to develop and standardize a new language that describes sign languages gestures. The language will define a standard to encode and transmit sign languages through Internet and mobile networks.

Table II.1: A set of languages for describing gestures

Languages	Emotion	Facial animation	Body animation	Dialogue	Mpeg4	Projects	Sign language
HumanML	Yes	Yes	Yes	Yes	-	-	No
CML	Yes	Yes	Yes	-	Yes	SAFIRA	No
AML	Yes	Yes	Yes	-	Yes	SoNG	No
APML	Yes	Yes	No	-	-	Greta	No
DAMSL	-	-	-	Yes	-	-	No
GESTYLE	Yes	Yes	Yes	-	-	-	No
MPML	Yes	Yes	No	-	-	Mobile	No
MURML	-	Yes	Yes	No	-	Max	No
VoiceXML	No	No	No	Yes	-	-	No
RRL	Yes	Yes	Yes	Yes	-	NECA	No
AIML	-	-	-	Yes	-	ALICE	No
VHML	Yes	Yes	Yes	Yes	Yes	InterFac, MetaFac, MentorSys	No
XSTEP	No	No	Yes	-	-	-	No
SigML	No	Yes	Yes	No	-	eSign	Yes
SWML	No	Yes	Yes	No	-	vSign	Yes
SML	Yes	Yes	Yes	No	-	WebSign	Yes

II.3.2 Avatar requirements

In order to obtain the same animation in different players and different avatars, the standardization of avatar's structure and parameters should be done. Below some characteristic that a signing avatar should check.

- The avatar should be a 3D avatar. It should be MPEG4 [b-MPEG4] compliant.
- The avatar should be at least compliant to the Level of articulation two of the H-Anim standard [b-Anim].
- The avatar should support facial animation.
- The possibility to change clothes and body textures and colour.

II.3.3 3D animation requirements

The quality of the avatar and the animation should not be the same as a normal animation of a 3D virtual avatar [b-IMWD]. As an example, the minimum frame rate of the animation should be fixed by the standard according to researcher studies, in order to ensure that the sign will be visible and comprehensible by deaf persons when they received it through Internet and mobile networks. Also, the contrast between hand and clothes colours should be also fixed.

- Colour and luminance contrast are important on the understandability of signs within a video or an animation. Colour blindness affects an individual's ability to distinguish between red and green and the shades in between. Luminance contrast is the difference in the amount of light reflected from the sign compared with the light reflected from the background or surrounding surface. There must be a luminance contrast of not less than 30 per cent between the raised tactile components of the sign and the surface of the sign, and the surface of the sign and the background it is mounted on.

- [b-QMCSL] proposes a resolution for the video of CIF format (352*288) and 3:4 aspect ratios to frame the upper body and signing space [b-QMCSL].
- The minimal frame rate, which has to be higher than 15 frames per second.
- The ability to pause and stop animation at any time whenever it is longer than 5 s (WCAG 2.0 Guidelines).
- The possibility to control the speed of the animation.
- The possibility of zooming and increasing the size of rendered images, so that deaf persons can see the facial expressions and gestures of hands better without lowering the quality of the rendered images.
- The inclusion of subtitles/captions, which are used by deaf and hard-of-hearing to assist their comprehension and the fluency of the sign language interpreter.
- The possibility to move the clip to another part of the screen when desired by user.
- The rapid display of the animation on the web site. Extended waiting time for an animation to load may lead to the confusion of the deaf person, since there is no proper feedback on what is going on.
- The use of sound. Although the sign language video is aimed at deaf persons who do not hear sound, the video can also be used for hard-of-hearing persons who wear hearing aids and still know sign language.
- The possibility to use a transparent background.

Bibliography

- [b-Anim] ISO/IEC 19774:2006, *Information technology – Computer graphics and image processing – Humanoid Animation (H-Anim)*.
<http://www.iso.org/iso/catalogue_detail.htm?csnumber=33912>
- [b-MPEG4] ISO/IEC 14496-2:2004, *Information technology – Coding of audio-visual objects*.
<http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39259>
- [b-IMWD] Matjaž D., Primož K. and Andreas H., *Improving multimodal web accessibility for deaf people: sign language interpreter module*, *Multimedia Tools and Applications*, August 2011, Volume 54, Issue 1, pp 181-199.
- [b-QMCSL] Hellström G., *Quality measurement on video communication for sign language*. 16th International Symposium on Human Factors in Telecommunications, Oslo, Norway, 1997.
- [b-SLMMS] Mohamed Jemni, Oussama Elghoul, Nour Ben Yahia, Mehrez Boulares, *Sign Language MMS to make cell phones accessible to deaf and hard-of-hearing community*, CVHI 2007 Euro-Assist Conference and Workshop on Assistive technology for people with Vision and Hearing impairments, August 28-31, Granada, Spain, 2007
-