Is Video on the Web for Sign Languages?

Guillaume Jean-Louis Olivrin
golivrin@meraka.org.za

Meraka Institute, CSIR
Pretoria, South Africa

1 Introduction

This position paper treats the feasibility of making Sign Language part of the Web fabric. The position I adopt is that Sign Languages need video as a requirement for communication and informational purposes. There is no single accepted writing system nor graphical notation for Sign Languages and gestures and facial expressions are still too impersonal and computationally hard to fake. We will review the challenges and the requirements of Deaf users and see that these problems are addressable and will allow Deaf people to become first class citizens on the Internet. If Video on the Web "is for Sign Languages", then this paper suggests an action plan to make it a reality.

1.1 Background and context

The current presence of Sign Languages on the Web is poor. The reason for that although we can note the emergence of Web-based video technologies and its popularity, the current solutions do not apply to Sign Languages because Sign Languages are languages, more than just mere video content.

With popular sharing communities like YouTube it is odd that we haven't seen a proper sign language portal yet. Although it is now easy to share videos for informational purposes, it is not appropriate nor geared towards private exchanges of information.

In fact the strangest phenomena is that it's not the Deaf who communicates in Sign Language that creates the most Sign Language video content on the Web, but researchers and teachers. Linguists even have to build expensive Sign Languages video databases and corpora to have some sort of data to carry out their research. Teaching resources are rare and it is not easy to find Sign Language study material on the Web where it should be thriving.

In the community of users, the Deaf are familiar with the features of 3G telephony and use SMS and MMS on a daily basis. In the South African context we even had to opportunity to test IP TV on *cell* phones "before everybody else". Noticing the both the need for tele-health and Sign Language communications with remote and rural villages, William Tucker has proposed the *SoftBridge* [6] technology to allow instant messages in the form of text and video across South Africa.

South African Sign Language (SASL) is present on three of the four public TV broadcasting channels on daily and weekly basis and there is much talk and proposal to make SASL the twelfth official language. In this context, it is important to bring technological solutions and making it possible to implement a language like SASL.

1.2 Thesis

The thesis of this paper states that the Internet is becoming the primary platform for all Sign Language information and communication needs. This is typically attested by the increased popularity of webcams, video podcasts, video blogs and video archiving services such as Google video. However, the Web is evolving from a text based milieu and "rich media" have subsequently been grafted to it. The fundamental problem then is that Sign Language is not only a rich medium but it is a language. Sign languages are languages which are not simply nor easily represented in writing or graphical forms. The primary reason is that they didn't benefit from a mutual acceptance of symbols like some spoken languages have.

Now that the context has been explained and the thesis at hand has been made explicit, I will review the current situation of Sign Language on the web.

2 Some misconceptions about Sign Languages, their users and the Web

The WCAG 2.0 guidelines provide accessible Sign Language for the Web architect The W3C's WCAG 2.0 WAI recommendations [8] regarding the inclusion of Sign Languages on the Web is answered primarily through the use of Synchronised Multimedia Content and SMIL 1 and 2. Although SMIL is a technical solution "de choix" (c.f. Section 3.4) it still hasn't proven itself yet: implementation and usability testing are needed to attest that it fully services Sign Languages.

Deaf are web users Stating that the Deaf are not web users is untrue but stating that they are is not true either. To resume this dilemma let's simply state that many Deaf people have embraced the Web chance as citeX had foreseen in the early days of the Web. A place of no discrimination, accessible to them and full of meeting opportunities.

Deaf are not technophobic There is a psychological association between technology and medical and pathological treatments. In my experience few Deaf people I met had problems or inihibitions using cell phones, computers and emails. But in a South African setting, few people knew about webcams and Internet video messaging. Saying that the Deaf are not technology savvy is further discredited by the fact that Deaf people are keen 3G cell phone users.

Sign Language is well captured by video There is an old preconception that Sign Language isn't in its true form when it's captured in video. The fact that daily interpreted TV news in SASL are being accepted nationaly means that there is no fundation but a old resistance to progress. There is a similar tendence nowdays to ignore and proscribe the usage of graphical notations, even when they present an advantage to promote and learn Sign Language.

The content of a video file or stream can be known before it is "opened". This is what metadata and MPEG-7 profiles have been created for, to make sure that the human and machine readable Web can understand binary contents and build semantics.

Hypermedia does not exist. It is true that before the advent of Flash and SMIL one couldn't provide cross-referencing of video or synchronized material.

The Web architect doesn't only work with text The Internet is made of text yet audio, video and other binary media are popular and well referenced. It doesn't mean Sign Language must be textual but gesture animation scripts would be more adapted than video.

Sign Language isn't a finite subset of rules and gestures which is why pre-recorded video or fixed set of gestures and facial expressions won't do. Sign Languages can be highly iconic and Sign Language videos are to date more widely welcomed and accepted than signing avatars and synthetic gestures.

3 The first elements of a (Web) solution for Sign Languages

I will now address various aspects of Sign Language video on the Web : the technologies that have been adopted, those that aren't tamed yet and those that do not exist but would potentially answer some urgent needs and requirements.

3.1 Strategical integration of Sign Languages in the World Wide Web

One problem with Sign Languages is that they do not quite pass the fitness factor of today's administrative, official, legal needs which consist in writing thing down. The reason for this state of affairs is the cost of processing, storing, exchanging video records as opposed to text documents. Besides the advent of video recorders, Sign Language suffers from not being adapted to paper nor to the computer platform. The fact that Sign Language lives and stands by itself, without the need for a writing system or archiving is a source of pride. This is costly because when time comes to prove the literature and cultural body associated to the Sign Language, there is very little to show to support request for becoming official languages. That Sign Languages are not well documented and suppported by mordern technologies is a great loss in creativity and mutual understanding between the hearing and Deaf in today's societies.

Use-cases In order to open our eyes to the possibilities offered by Video on the Web to the emancipation of Sign Languages, let's look at feasible scenario and use cases.

A. Providing video interpretations for existing web contents Just as there are some legal obligations to make governmental websites accessible in some countries (some states in the USA for example), there is a national responsibility to produce public documents in all official languages [1] Providing interpretation on existing content is probably one of the most straightforward scenarios where textual contents must be made concordant with an interpreted visual content, possibly through timed text and synchronous media.

AA. Website design for Sign Language content This is the situation where Sign Language is the primary format being offered on display (presentation) and operates as the default content (data). Sign language video corpora, Deaf community websites and blogs are often the places where Sign Language is featured as the primary content. YouTube has created a universe where video is the primary currency. Although the technology is applicable to Sign Languages there is no Sign Language specific processing. Higher order sign, gesture and facial expression functions for processing the video content as a language, such as search, topicalization, emphasis,... are not being offered. (more in Section 3.3)

Very few online websites that features a Sign Language are built around Sign Language itself. Online courses [5], online dictionaries (e.g. ASL dictionariess) and linguistic corpora are not organised according the Sign Language natural criteria but alphabetically. Natural criteria could be handshapes, classes of movements, families of similar or related gestures (such as groups, organisation, classes, companies). Typically this means that these bilingual dictionaries are only really usable one-way : from English to Sign Language but not from Sign Language to English. SignBank is the only attempt to create a Sign Language dictionary with signs definitions made explicit in Sign Language itself.

AAA. Sign Language driven web interfaces At this level of integration, all navigational and structural elements of a Web interface are available and controllable via Sign Language. This means that a search function is searching for Parts-of-Signs rather than words (c.f. Section 3.3), that links have video *alt* and *hypermedia* structure and navigation (c.f. Section 3.4) to *browse around.*

3.2 User experience

In the previous section I reviewed realistic use cases and applications of Video to Sign Languages on the Web. Let's complement these use cases with the needs and requirements that define "user experiences".

Ensuring quality of service Quality of service is a term used in Voice over IP technologies to determine the end-user experience qualitatively and quantitatively.

Quality of service for video communications and a visual language needs to evaluate both the message's medium (noise, jitter, delays as in information theory) and the intelligibility (clarity and faithfulness)

[1] South African Sign Language is currently candidate to becoming South Africa's twelfth official language

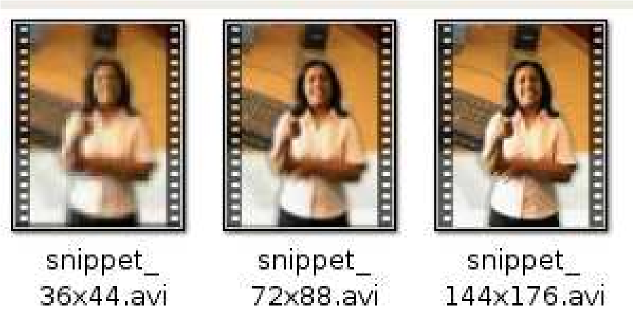


Fig. 1. These stills of Sign Language video clips illustrate that an uncompressed 144x176 video becomes intelligible. Additionally, the 3:4 aspect ratio is adapted to frame the upper-body and signing space.

of the message. Quality criteria of *moving images* are for example the minimum frame rate for Sign Languages. Less than 15fps will impact on the *fluidity* in the act of signing. Compression artifacts must be optimized so as to preserve the perceptual and visual information formed by moving hands and facial expressions. Perceptual qualitative and loss-driven information optimizations have a direct impact on message *readability*, clarity and the comfort of the user in understanding the message. Acceptable delivery times are also essential so that the message gets communicated "in time" and with lowest possible time distortions. An example of the consequence of delivery times causing user aggravation is the perception that two interlocutors would seem to be avoiding eye contact because of the slight delay in communication. This is case applies to real-time duplex communications and servers to illustrate the link between the visual medium and language intelligibility.

Accessibility Accessibility has been a core concept of the W3C and often pretends to address users "with special needs" Deaf people that communicate in Sign Language do not have much further luxury than accessing information and interfaces in their mother tongue.

An important accessibility feature applied to videos and Sign Language is the ability to change the rate of the playback. This accessibility requirement is similar to that of blind people who commonly boost or slow down their speech synthesizers. This makes it possible to adapt the video for various users with diverse levels of literacy in Sign Language. The new WMP [2]10 and Mplayer have this functionality and make it possible to slow down Sign Language. Strangely enough frame-based animation media like Macromedia Flash movies do not make it possible to change the playback rate of an animation during runtime.

The provision for multiple, parallel and alternative tracks of captions (tiers) in the form of customizable closed-captions is also recommended. This should provide both accessibility for other language users (various levels of transcriptions and translations) as well as navigational information (such as marking the end of utterances), c.f. Section 3.3.

Transcoding on demand Presenting the data in a relevant and adapted form is almost as important as making provision for the data itself. In effect the Web was originally conceived with little separation between presentation, logic and data.

Since video is a cumbersome medium, either in files or streams, delivering the right file size, the right screen size, the right compression ratio in the right delays are critical to the end user experience. This means that the audience should be in control of the data format it receives. Different target audiences can be catered for (c.f. Section 3.3) in the way Real Producer handles it, by producing different video

[2] Windows Media Player

bitrate files at creation time. It can also be dealt with in a "just in time" manner, as the user makes a request to the server and for a specific purpose.

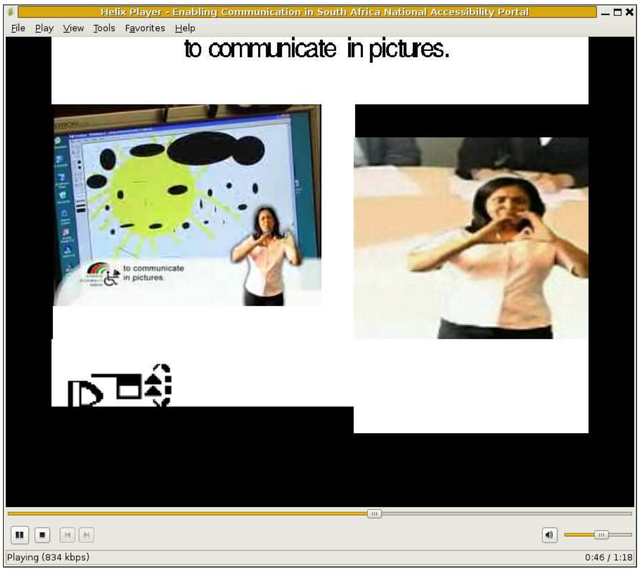
3.3 Sign Languages production for the Web

We have introduced use-cases of Sign Language on the web and described user's needs and requirements for an "home language" experience. This section discusses the means of producing Sign Language content for the web.

3G and IP video Internet connectivity through cell phones has introduced the ability to embed 3GP videos (H263) in MMS [3]. Our Sign Language team has performed a series of tests with camera equiped cell phones to determine if Sign Language on cell phone is acceptable. The resulting 176x144 compressed video clips are not clear nor quite intelligibile enough to communicate in a Sign Language. This is due to the important compression artifacts. Similar understandability tests realised on uncompressed 176x144 video clips showed that it is possible to have enough details to make the language acceptable (c.f. Figure 1). Further usability tests and experiments will follow. If MMS video quality was more indulgent towards Sign Languages, the popularity of SMS amongst the Deaf and the rapid adoption of 3G telephony by South Africans show that MMS based Sign Language would be popular, given that the video compression is optimized for Sign Languages.

MMS is a successful application of a subset of SMIL, a recommendation from W3C's SYMM working group. Again, the parallels between MMS and SMIL-based Web applications are great. Familiarity with SMS and MMS technologies prepares the Deaf to capture and create Sign Language video contents applicable to the Web.

Captioning Closed-captioning media such as audio and video is encouraged yet the tools available to do so often leave outside the Web. There are a few extra captioning functionalities that Sign Languages could use. First, there must be more than one tier (caption stream) displayed at the same time but with different synchronizations. Secondly, these streams must be able to refer to positions in images.


Helix Player - Enabling Communication in South Africa National Accessibility Portal
File
Play
View
Tools
Favorites
Help

to communicate in pictures.

to communicate
in pictures.

Playing (834 kbps)

0:46 / 1:18

Fig. 2. This SMIL movie displays a video (OGG) and an interpretation video overlay. The soundtrack, subtitles and image captions are in separate streams (WAV, Timed-Text, PNG).

[3] Multimedia Messaging Service in mobile devices.

Figure 2 shows an integrated example where SMIL provides a method for hypermedia referencing. An interpretation overlay is added over the video has prescribed by the WCAG 2 guidelines. And all the synchronised media come from various separable sources such as diverse channels for text, audio, and image captions. In this instance control is given to the user. For example it makes it optional to physically receive the audio stream. Sign language users are mostly Deaf and there is no need to create bandwidth overhead with audio data. Simply hitting the mute button doesn't have the same advantage as choosing not to receive audio. In this scenario, the client in charge of the playback does not even need to notify the server that the audio stream is not being listen to : it simply doesn't not load the resource.

In the same way spirit in which text dominates all Web creations, there are few video producing solutions that can make video a structured media. Video productions do not need to be considered as unmovable blocks but should rather be handled as separate logical units —hence the name video editing. Video segments (visiemes) can be segmented and given a structural analogy like text benefit from, letters making syllables and words arranged in sentences grouped as paragraphs, etc ...

Advanced linguistic notations for Sign Languages exist that can be used for captioning and annotating Sign Language specific gestures, poses and facial expressions. These symbol sets, e.g. HamNoSys (SigML) and Sutton Sign Writing (swml), are often XML-encoded and have no unicode representation. The best *Web* solution is to make use of SVG graphics to construct these graphical representations as captions from combinations of XML bits and pieces. The XML-based non-graphical representation of these captions make it possible to search the movie stream for signs, poses, gestures, movements, handshapes and facial expressions. This gives video "search" a *human* touch and function and can be applied in search engines to look for human postures like the ones in sport, martial arts, dance and body expression disciplines. Applied and dedicated to Sign Languages, this means that a sign can be retreive from its parts —Part-of-Sign. Such a functionality exist on the Sign Writing [4] website called *SignPuddle*. Suddently it is possible to look for signs from signs rather than from their *gross* English gloss. It is mostly functional differences that distingues between the following terminology and definitions : captions (e.g. subtitles), transcripts (e.g. longdesc), alternative description (e.g. alt) and gesture scripts (VRML paths, SignBuilder(tm), eSign (signML), STEP). These differences are on the basis of levels of language edition (the textitwording) and the levels of language description (how it is the message expressed).

We said previously that multiple tiers are necessary for Sign Language captioning and for Internet streaming. This is because Sign Languages use a combination of channels in parallel: left hand, right hand, facial expression and various other upper-body parts such as arms, neck and shoulders.

Building Sign Language corpora Available Sign Language content is so scarce that researchers and teachers have to build original collections of Sign Language videos. Apart from TV broadcasts, there are few places that use the web as a platform for publishing Sign Language content. Yet many research and educational institutions try to publish their own corpus of annotated Sign Language on the Web. Where spoken languages often use existing data to create corpora, Sign Language content is often staged information.

This is where the Internet can change the deal: usage of Sign Languages on the web should become convenient and wide spread.

Input method for Sign Languages To make the user experience complete, here is a summary of entry mechanism choices available to Sign Language users for posting Sign Language video on the web.

Webcams can be used either to send real time video which is common for IP messaging, record messages and simply send them over by email. Webcams unfortunately often have in-built limitations that affects the video stream quality (frame rate and compression ratios). Lighting is often a problem and only some webcams are clever enough to center the portrait of moving subjects. There is also the case of overall body movements versus the details of changing handshapes. This may call for multiple video stream, each interested and targetting particular body parts. The technology exists and can be adapted to be more appropriate to Sign Language usage.

Mobile phones were mentioned earlier as candidates for recording Sign Language and submitting the clip on the Internet because these phones can send data to Web URL.

What is missing in all these input methods is the consideration that we are not simply dealing with a video stream but with a language. This means that in the same way one would compose a text, a speech, an article, a letter, a contract . . . in English using spell checkers, capitalisation, various fonts . . . and a logical edition process, inputing data in Sign Language need similar notions of composition rules.

The logical organisation of Web content and its presentation also applies to Sign Languages. Marking-up the beginnings and ends of signed utterances would structure the Sign Language video stream so as to permit cross-referencing, navigation between part of Sign Language and search for signs.

3.4 Web Architecture

SMIL can be used to structure and present video based Sign Language media. SMIL is similar to CSS applied to other media than text. It makes it possible to provide a hypermedia experience, some meta-data for the streams, allows user and system choices to be made during playback and defines alternative layouts, layering and integrations of the various media.

From hypertext to hypermedia Hypermedia is the extrapolation of the hypertext concept that is at the heart of the Web. This means that during the course of a video, one should have the ability to “jump to” other parts of the video itself or to specific parts of other video streams. This functionality is already partially available through different forms and formats:

- Flash SWF allows referencing internal frames as named, anchored, keyframes.
- SignLinks [3] are hyperlinks created by SignEd and that make it possible to reference various times in Quicktime movies
- The Advene platform also provides video navigation as a web service. Advene makes use of custom video annotations and tags.

Refering to and cross-referencing to video time signatures is a useful function. It is also practical to be able to reference particular domains or areas in the video image itself. It can be done in an absolute or relative pixel unit reference frame or more abstractly according to the signing space reference frame. This is useful for the further Sign Language processings that were mentioned earlier in Section 3.3.

Meta-data Meta-data make it possible to announce the content and the format of a video stream or file before it is opened. There are meta-data such as the IMDI metadata descriptor that have been widely used to shelve Sign Language videos.

Meta-data by themselves have the value of an `alt` attribute in standard `HTML`’s `A` anchor elements. By itself, it doesn’t describe the data itself but rather makes explicit by a textual description a reference to data in a specific context such as a anchor link in HTML or a handle in a file system.

Describing the content itself is the job of ressource descriptors such as RDF and MPEG-7 audio-visualprofiles. Creating a MPEG-7 Sign Language profile would make it possible to formally define, model and abtract Sign Language functions and editing processes.

Integration How do you present Sign Language on the Web? How do videos best integrate in webpages? The practicalities of integrating Sign Language videos on the Web are still left to imagination. SMIL gives the power of producing layouts of text and images like CSS does for webpages. Using CSS to arrange video objects on a webpage doesn't seem to make the most of video media on the Web. To be able to manipulate the various media streams in a HTML page —layering, stacking, overlay and transparency effects, a presentation code such as SMIL and a resource descriptor such as RDF would be ideal. For inspirational purposes we can forsee the same kind of inter-media-plays such as the ones in Festoon where live video stream can be embedded in other moving media, and the facial expression in a video stream can be affected and modified by deformation functions.

Here is a list of possible integration strategies :

- Embedded objects such as SWF, RV, MOV;
- The "pop" tricks e.g. pop-in, pop-up, pop-out, which targets separate HTML documents or browsing windows ;
- Browser integration such as the integration of WMP in Internet Explorer as a browser side-pane, or as specific browser plugins as in Mozilla Firefox;
- Thin clients which are browsers dedicated to a specific reading, parse and presentation of HTML or scripts such as Real Media Player with RealScript and HTML.

4 Conclusion

This paper gave an account of why Sign Languages are waiting for Video to become part of the web fabric. The technological components to make Sign Language on the Web a lively solution already exist. Lines of action include the creation of look-up mechanisms for human gestures, facial expressions and poses in video streams, the promotion of structured videos with hypermedia references to spaces and times via SMIL, and finally the digitization of Sign Language video contents with meta-data, Sign Language notations and a RDF profile that defines Sign Language specific functions and playback options.

5 Acknowledgements

Team, NGO and words of thanks

This research wouldn't be complete without the full support and participation of the Meraka Sign Language team, Thibologa. I would like to thank the Deaf community i.e. DeafTV [2], DeafSA [1] and Dr Lynette van Zijl [7] from Stellenbosch University.

The author

Guillaume Olivrin received a Masters in computational linguistics and computer vision from the University of Edinburgh, Department for Artificial Intelligence. In 2005 he joined Meraka Institute's Intelligent Enabling Environments to look at Sign Language accessibility issues, sign-to-text and text-to-sign applications. A Sign Language team was formed including a start-up company, Thibologa [5]. Guillaume follows South African Sign Language certification at Witwatersrand University and has enrolled for a PhD with the University of Stellenbosch in 2007.

Contact details: Telephone: +27 012 841 3253; email: golivrin@meraka.org.za

Home page: http://www.meraka.org.za/∼golivrin

References

1. DeafSA. Deaf Federation of South Africa. http://www.deafsa.co.za.
2. DeafTV. TV broadcast studio. http://www.deaftv.co.za.
3. Daniel G. Lee, Jan Richards, Jim Hardman, Sima Soudain, and Deborah Fels. Creating Sign Language web pages. pages 1–9, 2004.
4. Steve Slevinski. SignPuddle^{TM} Reference Manual. The SignWriting Press, 2007. Center for Sutton Movement Writing Inc., CA, USA, http://www.signbank.org/SignPuddle1.5/.
5. TSLI. Thibologa Sign Language Institution, 2007. http://www.thibologa.co.za.
6. William D. Tucker and Edwin H. Blake. User interfaces for communication bridges across the digital divide. *AI & Soc*, pages 232–242, Aug. 2006.
7. Lynette van Zijl and Guillaume Olivrin. South African Sign Language Assistive Translation. In *Submitted to AT2008*. http://www.cs.sun.ac.za/∼lynette/.
8. W3C Web Accessibility Initiative (WAI). *Web Content Accesssibility Guidelines (WCAG 2.0)*, July 2007. http://www.w3.org/WAI/WCAG20/quickref/#media-equiv-sign.