Evaluating subtitle readability in media immersive environments

Pilar Orero

Department of Translation, Interpreting and Eastern Asia Studies, Universitat Autònoma de Barcelona, Spain, pilar.orero@uab.cat

Chris J. Hughes

School of Computing, Science and Engineering, University of Salford, UK, c.j.hughes@salford.ac.uk

Marta Brescia-Zapata

Department of Translation, Interpreting and Eastern Asia Studies, Universitat Autònoma de Barcelona, Spain, marta.brescia@uab.cat

The advances in VR technology has led to immersive videos rapidly gaining popularity. Accessibility to immersive media should be offered and subtitles are the most popular accessibility service. Research on subtitle readability has led to guidelines and standards (W3C, ISO/IEC/ITU 20071-23:2018). More research into subtitle presentation modes in 360º is needed in order to move towards understanding optimum readability. Evaluating readability for subtitles in immersive media environments requires a flexible and user-friendly framework for both creating the subtitles and presenting the generated subtitle file in a fully functional immersive video player, in order to understand the final view in the environment and assess its quality. This article starts by looking at the readability recommendations in W3C and ISO/IEC/ITU. The second part will describe the new features required in immersive subtitle presentations. The final section will describe the new web-based framework that allows the generation of immersive subtitles where readability may be tested. The framework has adopted a contrast and comparison approach towards instant readability evaluation.

CCS CONCEPTS • Human-centered computing • Interaction design • Interaction design process and methods • Interaction design prototyping

Additional Keywords and Phrases: Evaluation, Subtitle, Caption, Readability, Immersive environment, Framework

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium

DSAI 2020, December 2–4, 2020, Online, Portugal

ACM ISBN 978-1-4503-8937-2/20/12.
https://doi.org/10.1145/3439231.3440602

on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 STANDARDISED SUBTITLE READABILITY MARKERS

The Web Accessibility Initiative (WAI) from the World Wide Web Consortium (W3C) [1] offers strategies, standards and resources to make both interaction and Web content accessible to people with disabilities. It has recommendations for both live and pre-recorded captions/subtitles, but it is more focused on the latter. It has basic information about skills and tools required to create subtitles (transcribing and formatting). Specific guidance on how to transcribe audio to text is also given, but there is a lack of information regarding position and subtitle style, which is reflected in poor support in immersive video players [2].

ISO/IEC/ITU [3] Information Technology - User interface component accessibility Part 23 Visual presentation of audio information (including captions and subtitles) 20071-23:2018 has recommendation beyond language issues on subtitle (7.4) Synchronisation, (7.5) information obstruction, (7.6) Font size, (7.7) Font type, (7.8) Font face, (7.9) Letter cases: lower and capital, (7.10) contrast and use of colour, (7.11) speed, (7.12) number of lines, (7.13) kerning and leading (spaces between letters and lines), (7.14) punctuation, (7.15) spaces between words and phrases, (7.16) transition between presentations, (7.17) sentence segmentation, (7.18) indication of sentence breaks, (7.19) Additional duration for location change. ISO also identifies the Visual Alternative Container (VAC) as a possible influence towards subtitle readability.

2 SOME IMMERSIVE SUBTITLE FEATURES

Some W3C and ISO/IEC/ITU recommendations towards readability are no longer prescriptive since technology these days is increasingly allowing for personalisation at the client side [4]. While in the past sound volume and brightness or contrast were the only two adjustable elements, the change to media objects and the Internet media distribution have opened the possibility to endless subtitle presentation modes [5]. Immersive environments, and in particular 360º media, pose further challenges to readability [6].

2.1 Subtitling guiding modalities

In immersive environments the subtitle sound source may not be necessarily in the user area of view, hence guidance is required [7]. The ImAc project [8] designed and developed several guiding mechanisms, and test results showed two preferred methods:

ImAc Arrow - An arrow positioned left or right, directs the user to the target as in Figure 1 below:

Figure 1: An arrow guides to the sound source

ImAc Radar - A radar is shown in the users view. This identifies both the position of the subtitle, and the relative viewing angle of the user, as in Figure 2 below:

Figure 2: A radar guides to the sound source

A third guiding modality, a big arrow displayed in the centre of the users view has also been implemented as a simple example demonstrating the method for implementing further user-defined guiding modalities, as in Figure 3 below.

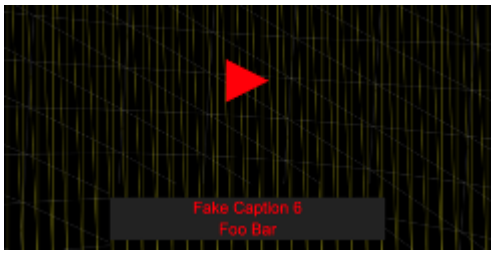

Figure 3: A large arrow guides to the sound source

These three modalities are the default guide modalities demonstrated within the proposed framework and are provided as samples within the JavaScript code in order to allow the framework to be extended to test new ideas towards improving readability.

2.2 Subtitle display modes

Traditional fixed subtitles within the users' view are challenged in the new immersive environments. No standard has yet decided on the best solution. The British Broadcasting Corporation (BBC) was one of the first to perform user testing with immersive subtitles [9]. All their work was based upon projecting traditional subtitles into the immersive environment and they evaluated how successful this could be done in scenarios where the subtitles were:

1. Evenly Spaced - subtitles repeated at 120º intervals
2. Head-locked - subtitles fixed within the users view
3. Head-locked with lag - subtitles follow users view, but only for larger head movements
4. Appear in front and then fixed - subtitles are placed in the position that the user is looking and remain there until they are removed

Results from testing showed the preference for head-locked options. Other tests performed [10] with fixed subtitles and compared to head-locked subtitles found no conclusive results. However, in terms of comfort, fixed subtitles led to a better result. Similar results were found at the ImAc project [11] along the need to guide users to the sound source of the subtitle such as the character speaking (see figs 1 and 2). All the three previous tests were with people with hearing loss. In all cases fixed subtitles in general mean that the user may not always be able to see the subtitle as it may be outside of their view. Recently a W3C Community group [12] focused on developing new standards for immersive subtitles conducted a community survey to gather opinions, but no tests. A small group of users with different hearing levels were asked to evaluate each of the identified approaches for subtitles within immersive environments. Again, the head-locked was clearly identified as the preferred choice. These results for a preference to the existing classic subtitle presentation [4] was also found during the past decade while testing subtitle readability across Europe [13].

One of the challenges to test immersive subtitles and its readability is the difficulty for users to properly evaluate new modalities, and one of the reasons is because of the challenge in terms of cost and time to create new subtitle presentations to enable users to experience them properly. This has been solved by the design and development of a web-based prototyping framework [14, 15].

Within the framework there are three main components: 1) A video container, 2) a fakeCamera container and 3) a fixed subtitle container. The video container hosts a video texture mapped sphere of which the user views from inside. The fakeCamera container is a group designed to replicate the behaviour of a camera which aligns with the users' viewpoint. This allows to always keep the components within the group locked into the users view and it contains a subtitle container, for placing subtitles which are fixed into the users view window. Finally, within the framework there is a fixed subtitle container which is not updated when the user moves. This allows to place a subtitle object into either the fakeCamera group or the fixed-subtitle group depending on whether the subtitle is locked in the scene or the user's view.

A wireframe plane is displayed by default in each view and attached to the fakeCamera to show the user's viewpoint and help provide a coordinated understanding of how the views fit together. The fakeCamera and the fixed-subtitle container both contain a pivot point ensuring that as they are rotated around the origin the subtitle aligns with the video sphere. This allows us to simply position the subtitle anywhere in the video using a spherical coordinate system and by applying a radial distance (r), polar angle (θ) and azimuthal angle (φ) values which are stored in the subtitle file, as illustrated in figure 4.

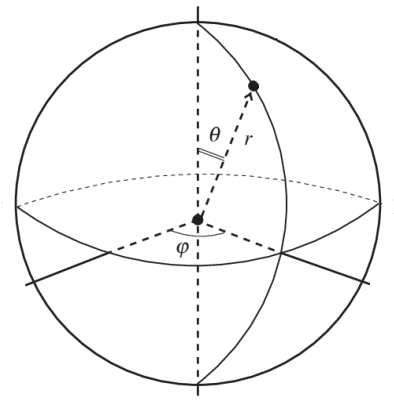

Figure 4: Spherical Coordinate system user to position the subtitle target

This new framework allows for instant immersive subtitle production in the following default modes:

- *Fixed in Scene, Locked Vertical* - The subtitle is positioned at the target, but the azimuthal position (φ) is restricted to 0 so that it remains locked to the horizon.
- *Fixed in scene, repeated evenly spaced* - The subtitle is positioned at the target location then duplicated at 2π/3 rad (120o) intervals around the horizon.
- *Appear in front, then fixed in scene* - The subtitle is rendered in the centre of the user's current view and remains there until the subtitle is updated.
- *Fixed, position in scene* - The subtitle is rendered at the target location.
- *Head-locked* - The subtitle is rendered in the user's viewpoint and is moved in sync with the user to ensure the subtitle remains statically attached to the viewpoint.

- *Head-locked on horizontal axis only* - The subtitle is rendered as headlocked, however the azimuthal position (φ) is restricted to 0, ensuring that the subtitle is always rendered on the horizon.
- *Head-locked on vertical axis only* - The subtitle is rendered as headlocked, however the polar position (θ) is locked to the target.
- *Head-locked with lag, animate into view* - The subtitle is rendered in the *head-locked* position, however as the users' viewpoint changes the subtitle is pulled back towards the *head-locked* position. An animation loop moves the subtitle incrementally causing it to smoothly animate into view.
- *Head-locked with lag, jump into view* - This is the same as above, except the animation time is reduced to 0, forcing the subtitle to jump into the users view.

2.3 Synchronic subtitles

Media consumption behaviour in 360º is no longer linear, with the user enjoying a freedom to decide on the area of attention interest and the time spent there. Hence subtitle duration is no longer determined by the narrative, but triggered by the user who decides where and for how long reads, leading to an extended subtitle reading time. Another issue is the time required to locate the source of sound and the corresponding subtitle. In both cases synchrony is required to avoid subtitle collision, or occlusion. The framework helps to evaluate the readability of the subtitle in these cases, and avoid obstruction.

3 CONTRAST AND COMPARE FRAMEWORK

A framework was developed [14, 15] to allow for fast generation of 360º subtitles with all the ISO requirements, and also for the new immersive subtitles' requirements. The framework was developed with two functionalities in mind towards fast evaluation: contrast and compare, as shown in figure 5.

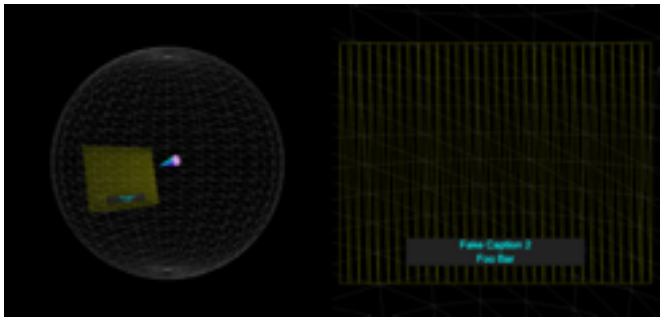

Figure 5: The basic split view of the player allows the user to see the subtitle and view window: left: relative to 360 world and right: from the user's perspective

Thanks to the framework components, 1) A video container, 2) a fakeCamera container and 3) a fixed subtitle container, the subtitler is able to generate a subtitle and simulate a visualization from inside. The fakeCamera container replicates the behaviour of a camera which aligns with the viewer viewpoint. The double screen with the two presentations allows for an immediate understanding of the subtitle effect and both legibility and readability.

3.1 Implementation

Part of the ambition for a framework which is to be generic enough to enable testing in different environments with a variety of devices is portability. Our implementation is based on web technologies allowing it to be used on any device with a web browser. This includes desktop computers, mobile devices and head mounted displays (HMD's).

Three.js [16] is a cross-browser JavaScript library and application programming interface (API) used to create graphical processing unit (GPU) accelerated 3D animations using the JavaScript language on the web without the need for proprietary web browser plugins. In our implementation it provides high level functionality to WebGL [17] allowing us to define our scene as objects and manipulate them within the space.

In addition, we use a WebVR Polyfill [18] - which provides a JavaScript implementation of the WebVR specification [19]. This enables three.js content to work on any platform, regardless whether or not the browser or device has native WebVR support, or where there are inconsistencies in implementation. The polyfill's goal is to provide a library so that developers can create content targeting the WebVR API without worrying about what browsers and devices their users are using. This gives our framework maximum compatibility across a large range of devices. Also, as many devices have limited interfaces, we add an option to automatically play or pause the video when the user enters or leaves VR modes to avoid the need for a play button if controls are available. Our framework allows for the user to switch between several different views or enter VR mode. The default view is a split screen as shown in figure 1. This clearly demonstrates how the subtitles are being rendered and positioned by showing both the user's viewpoint and a representation of the subtitle and view window within the space.

In order to consume 360º video, the scene contains a sphere centred around the origin and it is assumed that the user's viewpoint remains at the origin. When the framework is not connected to a video, the sphere is rendered as a wireframe, however once a video has been loaded the equirectangular video is texture mapper onto the inside of the sphere. As the sphere primitives are generally designed to have a texture mapped to the outside, it is necessary to invert the faces (also known as 'flipping the normals') in order to make this work.

Three.js provides a videoTexture object, which can connect to any HTML5 video object. Therefore, an HTML5 video is embedded in the webpage, with its display set to 'none'. The video playback is then managed through JavaScript manipulating the HTML5 video object.

3.2 Contrast and compare

Fundamentally, from our review there are two primary mechanisms for subtitle rendering. 1) Head-locked where the subtitle is rendered relative to the user's viewpoint and 2) Fixed, where the subtitle is rendered relative to a fixed location in the world, generally at the position of the character speaking.

Three.js allows for the textures to be generated from any HTML5 canvas object. In addition to the hidden video our HTML page contains a hidden canvas element which allows us to render any subtitle using any HTML or CSS styles. This canvas texture is then mapped to a plane and positioned into the scene.

An update is triggered every time a video frame changes, and the player checks to see if the subtitle has changed. If there is a new subtitle then 1) The canvas is updated to the text and style of the new subtitle, 2) The texture is updated and 3) the position of the subtitle is updated. For a fixed subtitle this position is attached to its relative position in the scene and placed within the fixed-subtitle container, however head-locked subtitles are fakeCamera objects which get repositioned each time the users' viewpoint is changed.

For each generated subtitle it is assigned a target location. In the first instance this is the position that is specified in the subtitle file. This concept was first used in the ImAc project where a single location was stored for each subtitle in an extended Timed Text Markup Language (TTML) file [20] and the location is defined in spherical coordinates. Within our player the user can enable the target position to be displayed in order to help with debugging, and understanding, however the subtitles do not necessarily get rendered at this location as the user may have chosen to offset the position, or it may be overridden by the display mode, for example head-locked will always render the subtitle into the users view. On opening our framework uses a random subtitle generator to show what is happening in the current display mode. A text string is generated and given a polar position (θ) between $-\pi$ rad and π rad (-180^o to 180^o) and azimuthal position (φ) between -0.4 rad and 0.4 rad (~-23^o to ~23^o) as subtitles are rarely positioned towards the top or bottom vertical pole.

4 CONCLUSIONS

New media formats, such as immersive environments require accessibility solutions towards a quality of service. In the case of immersive subtitles readability is as important as the quality of the text, where spelling mistakes or grammatical constructions impact as much as the many paralinguistic factors identified by the ISO standard. Evaluating immersive subtitles is needed by those producing creative subtitles, by subtitlers learning the behaviour of subtitles in the new immersive environment, and by researchers looking for the indicators towards quality. The developed framework is a tool with the functionality for contrast and comparison towards evaluating subtitles that will help in user testing and eventually in standardisation.

ACKNOWLEDGMENTS

The content of this article has been enriched by the comments from the W3C Immersive captions CG. This framework will be tested in H2020 957252 MEDIAVERSE, H2020 870610 TRACTION, and H2020 870939 SO-CLOSE. The paper has been partially funded by the the Catalan Agency AGAUR Grant no: 2017SGR113.

REFERENCES

[1] W3C Captions/subtitles. https://www.w3.org/WAI/media/av/captions/

[2] Chris J Hughes, and Mario Montagud. 2020. Accessibility in 360º video players, Multimedia Tools and Applications .

[3] ISO/IEC/ITU Information Technology - User interface component accessibility Part 23 Visual presentation of audio information (including captions and subtitles) 20071-23:2018 https://www.iso.org/standard/70722.html

[4] Lluis Mas, and Pilar Orero. 2018. New Subtitling Possibilities: Testing Subtitle Usability in HbbTV. Translation Spaces 7 (2) 263–284. DOI: https://doi.org/10.1075/ts.18016.man

[5] Chris J. Hughes, Mike Armstrong, Rhianne Jones, and Michael Crabb. 2015. Responsive design for personalised subtitles. The 12th Web for All Conference, 18-20 May 2015, Florence, Italy. https://doi.org/10.1145/2745555.2746650

[6] Belén Agulló, and Anna Matamala. 2019. The challenge of subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group. The Journal of Specialised Translation 32: 217–235. http://www.jostrans.org/issue32/art_agullo.php

[7] Belén Agulló, Mario Montagud, and Isaac Fraile. 2019. Making interaction with virtual reality accessible: rendering and guiding methods for subtitles. AI EDAM (Artificial Intelligence for Engineering Design, Analysis and Manufacturing). doi: 10.1017/S0890060419000362

[8] https://www.imac-project.eu/

[9] http://www.bbc.co.uk/rd/blog/2017-03-subtitles-360-video-virtual-reality

[10] Sylvia Rothe, Kim Tran, and Heinrich Hussmann. 2018. Positioning of Subtitles in Cinematic Virtual Reality. ICATEGVE 2018 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, November 2018.

[11] Chris J. Hughes., Mario Montagud, and Peter tho Pesch. 2019. Disruptive Approaches for Subtitling in Immersive Environments. Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video – TVX '19. https://tvx.acm.org/2019/

[12] https://www.w3.org/community/immersive-captions/

[13] Pablo Romero-Fresco. 2015. The Reception of Subtitles for the Deaf and Hard of Hearing in Europe. Peter Lang, Berlin.

[14] Chris J. Hughes, Marta Brescia-Zapata, Matthew Johnston, and Pilar Orero. 2020. Immersive captioning: developing a framework for evaluating user needs, in: IEEE AIVR 2020: 3rd International Conference on Artificial Intelligence & Virtual Reality 2020, 14th-18th December 2020, Online. (In Press)

[15] Chris J. Hughes. forthcoming. Universal Access: User Needs for Immersive Captioning. UAIS

[16] Three.js (2020) GitHub, https://github.com/mrdoob/three.js

[17] WebGL 2.0 Specification (2020), retrieved from https://www.khronos.org/registry/webgl/specs/latest/2.0

[18] WebVR Polyfill (2020) GitHub, https://github.com/immersive-web/webvr-polyfill

[19] WebVR 1.1 Specification (2017), retrieved from: https://immersive-web.github.io/webvr/spec/1.1

[20] TTML Profiles for Internet Media Subtitles and Captions 1.0.1 (IMSC1), W3C Recommendation 24 April 2018, retrieved from: https://www.w3.org/TR/ttml-imsc1.0.1/