Interactive Storytelling with Gaze-Responsive Subtitles

Andrew T. Duchowski
Visual Computing
Clemson University
Clemson, SC, USA
duchowski@clemson.edu

Patrícia Araújo Vieira
Brazilian Sign Language, Deaf Studies
Universidade Federal do Ceará
Fortaleza, Brazil
patriciavieira@ufc.br

Ítalo Alves Pinto de Assis
English, Literature and Translation
Universidade Federal do Ceará
Fortaleza, Brazil
italoalves@ufc.br

Krzysztof Krejtz
Institute of Psychology
SWPS University
Warsaw, Poland
kkrejtz@swps.edu.pl

Chris J. Hughes
School of Science, Eng., & Env.
Salford University
Salford, United Kingdom
c.j.hughes@salford.ac.uk

Pilar Orero
Translation and Interpretation
Universitat Autònoma de Barcelona
Barcelona, Spain
pilar.orero@uab.cat

(a) Saccade landing on subtitle box.

(b) Fixation detected on word within subtitle box.

Figure 1: Gaze behavior in relation to subtitles: (c) saccade to subtitle, (d) fixation on specific word within a subtitle.

Abstract

The paper describes an eye-tracking framework for offline analysis of and real-time interaction with gaze-responsive subtitled media. The eventual goal is to introduce and to evaluate gaze-responsive subtitles, which afford pausing of video when reading subtitles. Initial modes of interaction include: *look-to-read* and *look-to-release*. The former pauses video as long as gaze is detected over subtitles, the latter pauses video until gaze falls on subtitles. To avoid disrupted perception of media content, an additional ambient soundtrack matched to the general content of video is proposed. Note that this is potentially revolutionary as it would require an entirely novel approach to film direction. Just as Audio Description is now included in most modern films, ambient sound would also be needed to fill in brief temporal gaps when the user's visual attention is directed toward subtitles. Concomitantly, the eye-tracking framework fosters quantitative analysis of attention to audiovisual content apart from qualitative evaluation on which most of subtitling standardization is based.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**.

Keywords

eye tracking, subtitles, interaction

How to cite this paper:
Andrew T. Duchowski, Patrícia Araújo Vieira, Ítalo Alves Pinto de Assis, Krzysztof Krejtz, Chris J. Hughes, and Pilar Orero. 2025. Interactive Storytelling with Gaze-Responsive Subtitles. In *Proceedings of ACM IMX Workshops, June 3 - 6, 2025*. SBC, Porto Alegre/RS,Brazil, 7 pages. https://doi.org/10.5753/imxw.2025.9779

1 Introduction

The objective of this paper is to introduce and propose evaluation of real-time *gaze-responsive subtitles*, which afford pausing of video when reading subtitles. Making use of eye-tracking methodology, in turn, allows off-line quantitative analysis of captured gaze over subtitled media. Such quantitative analysis of gaze and perception of audiovisual content can yield more meaningful insight into subtitle consumption than traditional qualitative evaluation on which most of subtitling standardization is based. Recording and playback

(a) *1st* ranked: sans-serif font, 80% transparency. (b) *2nd* ranked: sans-serif font in outline.

Figure 2: Top two qualitatively ranked subtitle styles [University Autònoma de Barcelona 2010].

of gaze captured during video presentation allows word-level analysis of visual behavior over scene and subtitles. This allows the possibility of controlled user studies of novel interactive means of subtitle rendering.

Extending traditional 2D cinematic subtitles to interaction requires consideration beyond typical attributes such as their position, color, font, background box, etc. Instead, a framework is proposed that not only embodies such traditional attributes, but also affords dynamic subtitle effects in response to the user's gaze.

1.1 Contributions and Overview

This paper's contribution is a description of the eye-tracking framework that can be used for recording, playback, and interaction of gaze over subtitled media, e.g., in Virtual Reality (VR).

The paper starts with an overview of prior work pertaining to subtitling standards and, in particular, how eye tracking has featured in this context. Next, prior technical developments made in VR are reviewed, providing details of gaze recording and playback. Finally, the novel concept of gaze-responsive subtitles is introduced, highlighting the need for future user-centered evaluation.

2 PRIOR ART CONCERNING CAPTIONS

Prior evaluation of subtitles including style, speed of display, and position, has largely been qualitative. In Europe, progress has been made testing the quality of audiovisual accessible material, part of the Digital Television for All (DTV4All) project [Matamala and Orero 2010]. Primary findings relate to: *subtitle dimensions*, e.g., limited to 32-37 characters per each of two lines of text [Utray et al. 2010]; *font and size*, with sans serif fonts displayed in an 80% transparent box (see Fig. 2); *synchronization and timing*, with subtitles synchronized to appear when related characters begin to speak or when sound information is provided, usually lingering on the screen following the six-second rule [Pereira 2010]; and *position*, with preference favoring appearance at bottom.

These recommendations are based on subjective preference, even though such ratings are known to be influenced by habit or convention [Bartoll and Tejerina 2010]. How subtitles are read is largely uncharted, although it is known that, unlike reading static text, the pace of subtitle reading is in part dictated by the media rather than the reader: there is little time for the reader to regress to an earlier part of a sentence or phrase, and no opportunity to return to previous sentences [Kruger et al. 2015]. An eye tracker allows measurement of when subtitles are read (not just looked at), and also allow pacing of the media content accordingly.

2.1 Tracking Gaze: Current Limitations

Compared to the qualitative research conducted thus far, few inroads have been made into analysis of *process measures*, i.e., eye movements, to corroborate survey responses on which traditional recommendations are based [Uzquiza 2010]. Early eye-tracking studies indicated that the reading of subtitles is an automatic process that is independent of one's familiarity with subtitling, knowledge of the foreign language in the soundtrack, and the availability of the sound track [d'Ydewalle and Van de Poel 1999]. That is, subtitles are read mandatorily, and processed in detail and remembered. The addition of subtitles were shown to alter eye movement patterns with the viewing process becoming primarily a reading process with the amount of time spent on subtitles varying among individuals [Jensema et al. 2000]. As basic reading skills are learned, more attention shifts to captions since more and more information is obtained from this source [Jensema 2003].

Eye movements during reading have been studied extensively [Rayner 1998], but they have mainly been evaluated only qualitatively over video, with results limited to visualization of scanpaths, reminiscent of early work, e.g., that of Yarbus [1967]. Visualizations alone do not offer quantitative comparisons of eye movement elements such as fixations, fixation durations, etc., that are commonplace today [Jacob and Karn 2003; Webb and Renshaw 2008]. Uzquiza [2010] states this is precisely what is called for to better improve subtitles' processing and reception.

However, there is a paucity of gaze analysis over film beyond a handful of notable examples (e.g., Josephson and Holmes [2006]). The reason is likely due to the dynamic nature of video: synchronization of video with eye movement data is problematic (e.g., specification of dynamic Areas Of Interest (AOIs) is for the most part absent in commercial software [Papenmeier and Huff 2010;

Ryan et al. 2010]) and the proliferation of digital video formats complicates playback (e.g., commercial systems are often locked in to use specific codecs, and transcoding of video can be challenging).

Caffrey's [2009] study of gaze over Japanese *anime* exemplifies the methodological problems of commercially-supported eye-tracking analysis over video, highlighting the lack of software available to deal with voluminous data collected.

Martínez and Linder [2010] tested subtitles displayed either in scrolling mode (word-for-word), or in blocks. Quantitative analysis was performed on the number of fixations made per subtitle line and on the ratio of time spent on subtitles versus the remainder of the scene. The general finding was that viewers devoted about twice as many fixations to scrolling (word-for-word) subtitles as they did to block subtitles. Similarly, Perego et al. [2010] evaluated well- vs. ill-segmented two-line subtitles and also found a significantly greater proportion of fixations devoted to text than to the visual scene, regardless of the segmentation quality. A larger number of shorter fixations were issued to the subtitle region while fewer longer fixations were devoted to the visual elements. However, no significant differences were observed in terms of eye movements (numbers of fixations, mean fixation duration, or number of visual transitions between scene/subtitle regions) across conditions of differing subtitle segmentation quality. Lång et al. [2013] used an eye tracker to test properly or poorly synchronized subtitles but failed to show significant differences in number of fixations.

Significant aspects that are missing from prior evaluation of eye movements over video are quantitative analysis as well as the possibility of interaction. Prior work has either focused on a proportion of fixations or failed to produce statistically significant results. Generally, commercial software is ill suited for statistical analysis since this is not the software's primary function. This is not surprising since eye-tracking vendors can hardly be expected to anticipate all possible experimental designs for which their equipment is used.

An analytical approach is needed to quantitatively evaluate gaze distribution, i.e., a framework that will yield quantitative comparison of how viewers distribute their gaze on differing forms of subtitles. Since eye tracking hardware is becoming commonplace (e.g., in Virtual or Augmented Reality (VR/AR) headsets or on the desktop), there is potential for exploiting real-time gaze-based interaction with subtitles, e.g., pausing the audiovisual stream when gaze on subtitles is detected. Since commercial software is not suitable for this task, custom-developed software is required.

3 EYE-TRACKING FRAMEWORK

One objective of the proposed framework is to record and analyze gaze captured over subtitled video. To do so, control must be exercised over the style of presented subtitles, which means control over the video rendering, e.g., using `OpenGL` with proper synchronization established between eye movements usually expressed as time-stamped 2D coordinates (x_i, y_i, t_i) synchronized to the multimedia streams (video, audio, and subtitles). In short, the first step of the project is to develop a custom video player program able to parse `.srt` SubRip subtitles and communicate (e.g., via the network) with an eye tracker. Software to analyze the eye movement streams must be provided along with routines for statistical analysis. The development effort draws on recent work conducted in VR 360°, reviewed below for context.

3.1 360° Video Subtitle Format Parameters

Most recent work on subtitles in VR 360° relied to some extent on prior art featuring creative subtitles [McClarty 2012, 2014], which leads to a more adventurous proposition for immersive subtitles to heighten engagement and immersion, which are at the heart of the new immersive media formats. Though not tested in 360°, creative 2D subtitles have been tested and improve the viewing experience [Fox 2018]. This allows for the subtitles to be rendered relative to either the scene (fixed captions) or relative to the users viewpoint (head-locked). In addition, a number of parameters can be specified either per subtitle or speaker, defaulting to global settings where no parameters are supplied [Hughes et al. 2015].

3.2 The ImAc Testing Framework

As in traditional 2D media, to create subtitles in XR, a subtitle editor and a subtitle player are needed. Although there are commercially available immersive video players offering the ability to play VR 360° video, not many of them support accessible services [Brescia-Zapata 2022; Hughes and Montagud Climent 2020]. The majority of players seem to have inherited from the traditional 2D world, instead of addressing specific features of 360° environments. This scenario served as an inspiration for initiatives such as the European H2020 funded Immersive Accessibility (ImAc) project [1] that explored how accessibility services and assistive technologies can be efficiently integrated into immersive media, focusing on VR 360° video and spatial audio. Under the auspices of this project, an accessible player and subtitle editor were developed.

The accessibility-enabled 360° ImAc player supports audio description, audio subtitles, and sign language, among other features [Montagud Climent et al. 2019]. However, the ImAc subtitle editor is a commercial web-based editor, and its interface is similar to any traditional subtitle editor. The main innovations are related to the Field of View (FoV) in VR 360° video i.e., the extent of observable environment the user is able to see at any given moment.

3.3 The Live Web Testing Framework

One of the challenges of testing immersive subtitles is the difficulty for users to properly evaluate new modalities. This has been addressed by the design and development of a web-based prototyping framework [Hughes et al. 2020b,a].

Fundamentally, 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, usually close to the character speaking. To allow for producing creative subtitles a framework was designed [Hughes et al. 2020b,a]. where subtitle manipulation and evaluation depend on three components of the 360° video display system architecture: a video container, a fake camera container, and a fixed subtitle container.

Along with this XR subtitle simulator, a web-based editor was also developed. This editor allows importing subtitles previously created in `.srt` format or to create subtitles from scratch. Each

[1] `http://www.imac-project.eu`

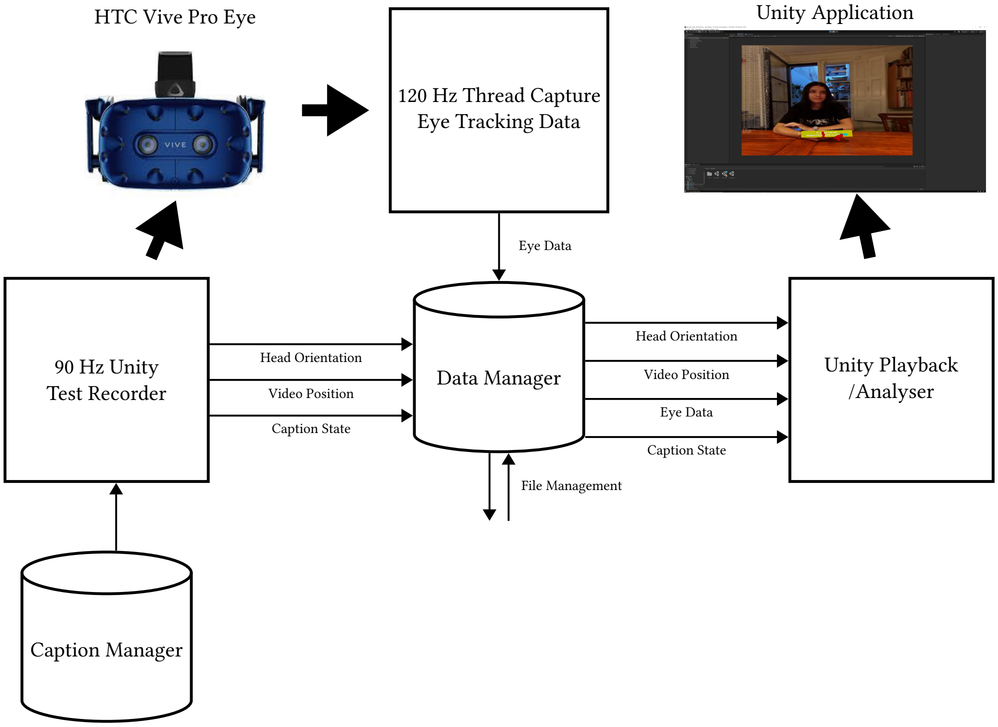

Figure 3: Eye-tracking VR system architecture (from Brescia-Zapata et al. [2022]).

subtitle can be associated with a character, or each subtitle is associated with a position (FoV), that is, the place in the 360° scene where it should appear.

Both the demo player and the editor are open source and can be accessed from a main project area, including a few 360° videos. These tools take inspiration from the player and the editor developed by the ImAc project. The main difference is that ImAc tools are intended to be used by generic audiences (final users), while the demo tools are more focused on research and testing.

3.4 System Architecture

To enable recording of gaze within 360° video, the live web testing framework developed by Hughes et al. [2020b,a] was ported to Unity 3D to display 360° video and to capture data from the eye tracker built in to the VR device. A new system architecture emerged, as depicted by the schematic in Figure 3.

The system architecture was developed to utilize the HTC Vive Pro Eye, which contains a Tobii eye tracker built in to the display. The application uses two Unity assets: one optimized for recording and the other for playback. The linchpin of the architecture is a Data Manager, which stores all test data. It also handles file management and generates output data in a variety of formats as required.

Recording allows for a specified 360° video to be played with the captions fixed in the scene. During the test, each event and data is logged into the data manager as it becomes available and timestamped. To replay a user viewing session, the system records head orientation, video (frame) position, gaze data (raw and analyzed), and the subtitle caption state, i.e., which caption from the accompanying SubRip (`.srt`) file was being displayed and where.

Playback retrieves data from the Data Manager which allows the entire test to be replayed. This affords the opportunity to change the analysis process or include additional Areas Of Interest (AOIs) and the analysis repeated. It also allows for visual analysis by overlaying gaze data onto the video after capture.

3.5 Apparatus

Gaze data is recorded with an HTC Vive Pro Eye VR headset (see Figure 3) driven by Unity 3D. The Vive's eye-tracking accuracy estimation is 0.5°–1.1° [Sipatchin et al. 2020]. The HMD has two AMOLED screens, with a resolution of $2,880 \times 1,600$ pixels in total ($1,440 \times 1,600$ pixels to each eye resulting in a pixel density of 615 pixels per inch, or PPI), with a display refresh rate of 90 Hz and a field of view of 110°.

If data from the eye tracker is collected in lockstep with the display, it will be recorded at 90 Hz. This is the most straightforward approach. However, the eye tracker samples at 120 Hz, and it is possible to record data from it at its 120 Hz sampling rate (how it is done in our current implementation). To do so, a separate thread is used to capture eye movement data that runs concurrently with the display thread.

3.6 Eye Movement Signal Processing

We are particularly interested in how well participants read subtitles Eye movement data shows quantitatively when and where each

(a) saccade flight (b) saccade below text (c) saccade landing on text (d) fixation on word

Figure 4: Recording in VR showing gaze behavior in relation to subtitle: (a) saccade mid-flight, (b) saccade landing below subtitle, (c) saccade to midpoint of subtitle, (d) fixation within a subtitle.

subtitle was processed, e.g., see Figure 1. Our system captures gaze as it traverses subtitles when reading the text displayed within the quadrilaterals that contained them. We are interested in counting fixations as they fall within the subtitles.

Analysis of eye movements relies on fixation detection, which in turn depends on saccade detection. We detect fixations from the raw eye movement signal following Nyström and Holmqvist [2010] and by using the Savitzky and Golay [1964] filter for velocity-based (I-VT [Salvucci and Goldberg 2000]) event detection. The Savitzky-Golay filter fits a polynomial curve of order n via least squares minimization prior to calculation of the curve's s^{th} derivative (e.g., 1^{st} derivative ($s = 1$) for velocity estimation) [Gorry 1990].

The current system architecture allows for detection of fixations falling within arbitrarily defined Areas Of Interest (AOIs), including polygons defined over actors and more importantly over quadrilaterals (quads) used to display subtitles and defined over individual words, see Figure 4.

3.7 Empirical Evaluation

Two experiments were conducted, following a 2×2 mixed design with between-subject subtitle position (fixed or head-locked) and within-subjects subtitle color (monochrome or color) as independent variables. Two videos were created and shown to mitigate memory effects.

In the first study, Brescia-Zapata et al. [2022] found differences in fixation duration on scenes vs. subtitles, as well as differences in ambient/focal scanning of scenes vs. subtitles.

In the second study, Brescia-Zapata et al. [2023] recruited 73 participants from 3 different locations (Spain, Poland, and the UK), split between hearing and Deaf groups. A significant effect of subtitle position was found for hearing viewers (a greater percentage of subtitles was read when they were fixed) but not for Deaf viewers. Effects of subtitle color were also noted (e.g., fewer fixations on color subtitles).

3.8 Summary of Subtitle Testing in VR

The eye-tracking framework is an effective means of testing subtitle usage in viewing VR 360° videos. Beyond examination of subtitle position and color, what was perhaps more interesting was the anecdotal observation of excessive head movement required to locate the subtitles that were relevant at particular points in time of video viewing.

To exploit the 360° video surround, speaking characters were often placed strategically to make use of the entire 360° immersive space. However, this, in turn, necessitated that users make quick head movements in an effort to (1) locate the subtitles of the given speaker (e.g., in fixed subtitle mode), and to then (2) read the subtitles before they disappeared. Consequently, it was discovered that pausing video playback would give viewers sufficient time to accomplish both. However, several design decisions presented themselves when pausing video, leading to two interaction metaphors: *look-to-release* and *look-to-read*. Both of these modes of interaction are described in greater detail below, in the larger context of prior art concerning gaze-responsive subtitles.

4 Gaze-Responsive Subtitles

As with built-in eye trackers in VR as well as with affordable table-mounted eye tracking (e.g., from Gazepoint), novel insights into the subtitle reading process can be obtained and provision of novel means of subtitle interaction can be explored. Kruger et al. [2015] review past eye-tracking research on subtitles in traditional 2D media, e.g., film shown on typical television or monitor screens (see also Bisson et al. [2012] for a review and an example of an eye-tracking study of subtitles). The proposed framework is the first to use eye tracking to actively use subtitles interactively. While a few approaches have appeared concerning gaze-adaptive or other forms of dynamic subtitles [Brown et al. 2015; Kurzhals et al. 2020], no one has yet proposed the use of an eye tracker to detect reading of subtitles as a means of media control.

During our prior work on evaluation of subtitles in VR [Brescia-Zapata et al. 2022, 2023] (see Fig. 4), we noticed that pausing video automatically when gaze was detected over subtitles led to their less hurried reading. Anecdotally, this discovery showed that adapting video pacing to subtitle reading could improve comprehension of information contained within the subtitles so long as viewers had time to process the information. This form of self-pacing was hinted at by Kruger et al. [2015] but it appears that it has never been attempted where gaze acts as the trigger.

The proposed framework is not unlike Vertegaal's [2003] Attentive User Interface, or AUI, but in reverse. An attentive television would, for example, play video when looked at, but would pause when gaze turned away. For gaze-responsive subtitles, at least three technical challenges present themselves: (1) *when* to pause video, (2) *what* to do when the video (and audio) are paused (and when to resume playback), and (3) *how* to detect reading in the first place.

The latter problem is perhaps easier to tackle as it relies on classification of eye movements when reading. One could employ a reading model of eye movement, similar to that of Campbell and Maglio [2001] or based on a transition matrix model as was used to develop models of eye movement synthesis [Duchowski et al. 2019]. The former two problems are address to a certain extent by two proposed interaction metaphors.

First, the *look-to-release* metaphor pauses the video when a new subtitle appears. Video is resumed as soon as gaze lands on the subtitle, i.e., media stops before the newly appearing subtitle is fixated. However, this causes slight disorientation in the sense that the viewer might not initially realize why the video has paused.

Second, video is paused when gaze lands on a subtitle and resumes once gaze exits the subtitle box. This is the *look-to-read* metaphor, which seems more intuitive i.e., media stops when a subtitle is fixated. Note that look-to-read mode of video interruption can be implemented so that media is paused only when a subtitle newly appears and is located by the viewer. This allows the possibility of the user missing the subtitle, e.g., when preferring to attend to the visual content instead. However, this form of interaction is potentially disruptive of the spoken audio track.

Relying on video and audio cessation, both approaches beg the question of how noticeable such an interruption might be. When engaged in reading subtitles, with gaze concentrated on the small portion of the display, pausing video might not be too noticeable, but pausing audio is likely to disrupt aural perception. The proposed solution is to endow videos with an additional soundtrack matched to the general content of video. ***Note that this notion is potentially revolutionary as it would require an entirely novel approach to film direction.*** Just as Audio Description is now included in modern films, *ambient sound* would also need to be provided to fill in brief temporal gaps when the user's attention is devoted to reading subtitles. Because the sound score is a key component of the narrative, *ambient audio* will need to be carefully crafted so as not to distract from the present point in the score.

Development of interactive subtitles stands to fundamentally transform immersive media creation. This is because content creators will need to be mindful of accessibility and inclusion criteria in directing film or developing games. Beyond subtitle standards which the proposed work will help define, creators will need to tailor content to instances when users are reading subtitles, ushering in a novel form of media, i.e., media that is made aware of the user's attention, be it on subtitles or scene. As an example of its potential significant cultural impact, consider the Grammy award category newly created in 2023: Best Score Soundtrack for Video Games and Other Interactive Media, awarded to composer Economou [2023] The key technical achievement centered on creation of composition of ambient sounds (including music) that can loop indefinitely without spoiling gameplay, thus adapting to the player's skill level (an expert only needs to hear the soundtrack for a short period of time compared to that of the novice). This is precisely what is needed for interactive processing of subtitles.

4.1 Empirical Evaluation

Gaze-based quantitative analysis will allow design of preliminary experiments, which can be used to evaluate the following aspects such as effects of (1) shot changes on subtitle perception; (2) various attributes of foreign subtitles, e.g., font, characters per line, 6-second rule, etc.; (3) subtitling color, lines, position. Excluding preference, gaze analytics are applicable to the above tests, provided experimental parameters are recorded, e.g., distance to screen, screen size and resolution, and sampling rate of the eye tracker. These are normally operationalized parameters which are recorded as a matter of course in eye-tracking experiments. As such, the anticipated importance and potential of success of the eye-tracking framework are both high.

In prior work we were able to replicate Yarbus's [1967] classic result showing that the task given to viewers changes their gaze behavior. The novelty of our contribution was our ability to quantify this effect over video [Grindinger et al. 2011, 2010; Vilaró et al. 2011]. Success was limited, however, suffering from two critical deficiencies. First, the machine-learning approach used was too computationally expensive to be practical. Second, the Gaussian-based similarity metric is unconventional and unwieldy.

The Gaussian similarity measure can be replaced by one based on entropy. Computation of entropy is straightforward, yielding a single, normalized value associated with a heatmap, indicating in essence the probability of dispersion (density) of an individual's or a group's distribution of visual attention. Entropy can readily be compared statistically to indicate whether one sequence of entropies (e.g., per video frame per individual or group) differs from another (via traditional analysis of variance, or ANOVA). This type of approach is needed for quantitative analysis of gaze when video is viewed with subtitles present or absent.

Future efforts need to compare and contrast both *look-to-release* and *look-to-read* mechanisms in connection with the ambient soundtrack matched to the general content of the video. Experimental design will need to consider user demographics (e.g., not-impaired users) and alternative methods such as looping a set of frames when paused. Individual user modeling could also be examined.

5 Conclusion

The paper reviewed a first-of-its-kind analysis of eye-tracked gaze on subtitles by individuals immersed in viewing subtitled VR 360° videos. Results were intriguing showing that head-locked subtitles are fixated more than when shown in fixed position. Deaf viewers showed preference for colored subtitles, viewing them with greater frequency than their monochrome version. Moreover, objective gaze data showed a stark contrast between how viewers thought they perceived subtitles and how they actually viewed them. This type of in-depth analysis of subtitle reading would not be possible without detailed examination of gaze. More importantly, perhaps, this work led to anecdotal observations of the benefits of two novel interaction metaphors: *look-to-release* and *look-to-read*. Both modes of interaction rely on temporary stoppage of video and audio. Consequently, new methods of ambient audio (and video) are required. Provision of these carefully crafted additional video tracks in the context of gaze-responsive subtitles is likely to usher in a fresh approach to interactive film direction and immersive storytelling.

References

Eduard Bartoll and Anjana Martínez Tejerina. 2010. The positioning of subtitles for the deaf and hard of hearing. In *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing*, Anna Matamala and Pilar Orero (Eds.). Peter Lang AG, Bern, Switzerland, 69–86.

Marie-Josée Bisson, Walter van Heuven, Kathy Conklin, and Richard Tunney. 2012. Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics* 35 (03 2012). doi:10.1017/S0142716412000434

Marta Brescia-Zapata. 2022. The Present and Future of Accessibility Services in VR360 Players. *inTRAlinea* 24 (1 2022). https://www.intralinea.org/archive/article/2585

Marta Brescia-Zapata, Krzysztof Krejtz, Pilar Orero, Andrew T. Duchowski, and Christopher J. Hughes. 2022. VR 360° subtitles: Designing a test suite with eye-tracking technology. *Journal of Audiovisual Translation* 5, 2 (2022). doi:10.47476/jat.v5i2.2022|

Marta Brescia-Zapata, Krzysztof Krejtz, Pilar Orero, Andrew T. Duchowski, and Christopher J. Hughes. 2023. Subtitles in VR 360° video: Results from and eye-tracking experiment. *Perspectives: Studies in Translation Theory and Practice* (2023), 1–23. doi:10.1080/0907676X.2023.2268122

Andy Brown, Rhia Jones, Michael Crabb, James Sandford, Matthew Brooks, Michael Armstrong, and Caroline Jay. 2015. Dynamic Subtitles: The User Experience. In *TVX 2015* (Brussels, Belgium). doi:10.1145/2745197.2745204

Colm Caffrey. 2009. *Relevant abuse? Investigating the effects of an abusive subtitling procedure on the perception of TV anime using eye tracker and questionnaire.* Ph. D. Dissertation. Dublin City University, Dublin, Ireland. http://doras.dcu.ie/14835/1/Colm_PhDCorrections.pdf (last accessed Sep. 2010).

Christopher S. Campbell and Paul P. Maglio. 2001. A Robust Algorithm for Reading Detection. In *ACM Workshop on Perceptive User Interfaces.* ACM Press, 1–7.

Andrew T. Duchowski, Sophie Jörg, Jaret Screws, Nina A. Gehrer, Michael Schönenberg, and Krzysztof Krejtz. 2019. Guiding Gaze: Expressive Models of Reading and Face Scanning. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, CO) *(ETRA '19).* ACM, New York, NY, Article 25, 9 pages. doi:10.1145/3314111.3319848

Géry d'Ydewalle and Marijke Van de Poel. 1999. Incidental Foreign-Language Acquisition by Children Watching Subtitled Television Programs. *Journal of Psycholinguistic Research* 28, 3 (1999), 227–244.

Stephanie Economou. 2023. Assassin's Creed Valhalla: Dawn of Ragnarok. American National Academy of Television Arts & Sciences 65th Annual Grammy Awards. Best Score Soundtrack for Video Games and Other Interactive Media.

Wendy Fox. 2018. *Can integrated titles improve the viewing experience? Investigating the impact of subtitling on the reception and enjoyment of film using eye tracking and questionnaire data.* Language Science Press, Berlin, Germany. doi:10.5281/zenodo.1180721

Peter A. Gorry. 1990. General Least-Squares Smoothing and Differentiation by the Convolution (Savitzky-Golay) Method. *Analytical Chemistry* 62, 6 (1990), 570–573. doi:10.1021/ac00205a007 arXiv:http://pubs.acs.org/doi/pdf/10.1021/ac00205a007

Thomas J. Grindinger, Andrew T. Duchowski, and Pilar Orero. 2011. Differentiating Aggregate Gaze Distributions. In *Applied Perception in Graphics & Visualization (APGV).* ACM, Toulouse, France. (Poster).

Thomas J. Grindinger, Vidya N. Murali, Stephen Tetreault, Andrew T. Duchowski, Stan T. Birchfield, and Pilar Orero. 2010. Algorithm for Discriminating Aggregate Gaze Points: Comparison with Salient Regions-Of-Interest. In *International Workshop on Gaze Sensing and Interactions.* IWGSI/ACCV, Queenstown, New Zealand.

Chris J. Hughes, Mike Armstrong, Rhianne Jones, and Michael Crabb. 2015. Responsive Design for Personalised Subtitles. In *Proceedings of the 12th International Web for All Conference* (Florence, Italy) *(W4A '15).* Association for Computing Machinery, New York, NY, USA, Article 8, 4 pages. doi:10.1145/2745555.2746650

Chris J. Hughes, Marta Brescia-Zapata, Matthew Johnston, and Pilar Orero. 2020b. Immersive captioning: developing a framework for evaluating user needs. In *IEEE AIVR 2020: 3rd International Conference on Artificial Intelligence & Virtual Reality 2020.* IEEE. http://usir.salford.ac.uk/id/eprint/58518/

Chris J. Hughes, M. Brescia-Zapata, and Pilar Orero. 2020a. Evaluating subtitle readability in media immersive environments. In *DSAI 2020 proceedings.* Association for Computing Machinery (ACM). http://usir.salford.ac.uk/id/eprint/59022/

Chris J. Hughes and Mario Montagud Climent. 2020. Accessibility in 360° video players. *Multimedia Tools and Applications* (October 2020), 1–28. doi:10.1007/s11042-020-10088-0

Robert J. K. Jacob and Keith S. Karn. 2003. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, Jukka Hyönä, Ralph Radach, and Heiner Deubel (Eds.). Elsevier Science, Amsterdam, The Netherlands, 573–605.

Carl Jensema. 2003. *The Relation Between Eye Movement and Reading Captions and Print by School-Age Deaf Children.* Final Report Grant Award Number HH327H000002. Institute for Disability Research and Training, Inc., Wheaton, MD.

Carl J. Jensema, Sameh El Sharkawy, Ramalinga Sarma Danturthi, Robert Burch, and Daviv Hsu. 2000. Eye Movement Patterns of Captioned Television Viewers. *American Annals of the Deaf* 145, 3 (July 2000), 275–285.

Sheree Josephson and Michael E. Holmes. 2006. Clutter or Content? How On-Screen Enhancements Affect How TV Viewers Scan and What They Learn. In *Eye Tracking Research & Applications (ETRA) Symposium.* ACM, San Diego, CA, 155–162.

Jan-Louis Kruger, Agnieszka Szarkowska, and Izabela Krejtz. 2015. Subtitles on the Moving Image: An Overview of Eye Tracking Studies. *Refractory: A journal of Entertainment Media* 25 (02 2015).

Kuno Kurzhals, Fabian Göbel, Katrin Angerbauer, Michael Sedlmair, and Martin Raubal. 2020. A View on the Viewer: Gaze-Adaptive Captions for Videos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20).* Association for Computing Machinery, New York, NY, 1–12. doi:10.1145/3313831.3376266

Juha Lång, Jukka Mäkisalo, Tersia Gowases, and Sami Pietinen. 2013. Using Eye Tracking to Study the Effect of Badly Synchronized Subtitles on the Gaze Paths of Television Viewers. *New Voices in Translation Studies* 10 (2013), 72–86.

Juan Martínez and Gion Linder. 2010. The Reception of a New Display Mode in Live Subtitling. In *Proceedings of the Eighth Languages & The Media Conference* (Berlin, Germany). 35–37.

Anna Matamala and Pilar Orero (Eds.). 2010. *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing.* Peter Lang AG, Bern, Switzerland.

Rebecca McClarty. 2012. Towards a multidisciplinary approach in creative subtitling. *MonTI: Monografías de Traducción e Interpretación* (January 2012), 133–153. doi:10.6035/MonTI.2012.4.6

Rebecca McClarty. 2014. In support of creative subtitling: contemporary context and theoretical framework. *Perspectives* 22, 4 (2014), 592–606. doi:10.1080/0907676X.2013.842258

Mario Montagud Climent, Isaac Fraile, Einar Meyerson, Maria Genis, and Sergi Fernández. 2019. ImAc Player: Enabling a Personalized Consumption of Accessible Immersive Contents. In *Adjunct Publication of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (Hilversum, The Netherlands) *(TVX '17 Adjunct).* Association for Computing Machinery, New York, NY, 3––8. doi:10.6084/m9.figshare.9879254.v1

Marcus Nyström and Kenneth Holmqvist. 2010. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behaviour Research Methods* 42, 1 (2010), 188–204.

Frank Papenmeier and Markus Huff. 2010. DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animations and movies. *Behavior Research Methods* 42, 1 (2010), 179–187.

Elisa Perego, Fabio Del Missier, Marco Porta, and Mauro Mosconi. 2010. The Cognitive Effectiveness of Subtitle Processing. *Media Psychology* 13, 3 (2010), 243–272. URL: http://dx.doi.org/10.1080/15213269.2010.502873 (last accessed Sep. 2010).

Ana Pereira. 2010. Criteria for elaborating subtitles for deaf and hard of hearing adults in Spain: Description of a case study. In *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing*, Anna Matamala and Pilar Orero (Eds.). Peter Lang AG, Bern, Switzerland, 87–102.

Keith Rayner. 1998. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin* 124, 3 (1998), 372–422.

Wayne J. Ryan, Andrew T. Duchowski, Ellen A. Vincent, and Dina Battisto. 2010. Match-Moving for Area-Based Analysis of Eye Movements in Natural Tasks. In *ETRA '10: Proceedings of the 2010 Symposium on Eye Tracking Research & Applications.* ACM, Austin, TX.

Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00).* Association for Computing Machinery, New York, NY, 71––78. doi:10.1145/355017.355028

Abraham Savitzky and Marcel J. E. Golay. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* 36, 8 (1964), 1627–1639. http://pubs.acs.org/doi/abs/10.1021/ac60214a047

Alexandra Sipatchin, Siegfried Wahl, and Katharina Rifai. 2020. Eye-tracking for low vision with virtual reality (VR): testing status quo usability of the HTC Vive Pro Eye. *bioRxiv* (2020). doi:10.1101/2020.07.29.220889

University Autònoma de Barcelona. 2010. *Digital Television For All (DTV4All).* Technical Report. University of Roehampton, UK. URL: http://dea.brunel.ac.uk/dtv4all/ICT-PSP-224994-D25.pdf, (last accessed Sep. 2010).

Francisco Utray, Belén Ruiz, and José Antonio Moreiro. 2010. Maximum font size for subtitles in Standard Definition Digital Television: Test for a font magnifying application. In *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing*, Anna Matamala and Pilar Orero (Eds.). Peter Lang AG, Bern, Switzerland, 59–68.

Verónica Arnáiz Uzquiza. 2010. SUBSORDIG: The need for a deep analysis of data. In *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing*, Anna Matamala and Pilar Orero (Eds.). Peter Lang AG, Bern, Switzerland, 163–174.

Roel Vertegaal. 2003. Attentive User Interfaces. *Commun. ACM* 46 (03 2003). doi:10.1145/636772.636794

Anna Vilaró, Pilar Duchowski, Andrew T. Orero, Thomas J. Grindinger, Stephen Tetreault, and Elena Di Giovanni. 2011. How Sound is The Pear Tree Story? Testing the Effect of Varying Audio Stimuli on Visual Attention Distribution. *Perspectives: Studies in Translatology* (2011). Special Issue on The Pear Stories (to appear).

Natalie Webb and Tony Renshaw. 2008. Eyetracking in HCI. In *Research Methods for Human-Computer Interaction*, Paul Cairns and Anna L. Cox (Eds.). Cambridge University Press, Cambridge, UK, 35–69.

Alfred L. Yarbus. 1967. *Eye Movements and Vision.* Plenum Press, New York, NY.