Routledge
Taylor & Francis Group

Perspectives
Studies in Translation Theory and Practice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rmps20

Subtitles in VR 360° video. Results from an eye-tracking experiment

Marta Brescia-Zapata, Krzysztof Krejtz, Andrew T. Duchowski, Chris J. Hughes & Pilar Orero

To cite this article: Marta Brescia-Zapata, Krzysztof Krejtz, Andrew T. Duchowski, Chris J. Hughes & Pilar Orero (13 Nov 2023): Subtitles in VR 360° video. Results from an eye-tracking experiment, Perspectives, DOI: 10.1080/0907676X.2023.2268122

To link to this article: https://doi.org/10.1080/0907676X.2023.2268122

Published online: 13 Nov 2023.

PERSPECTIVES
https://doi.org/10.1080/0907676X.2023.2268122

Subtitles in VR 360° video. Results from an eye-tracking experiment

Marta Brescia-Zapata [a], Krzysztof Krejtz [b], Andrew T. Duchowski [c], Chris J. Hughes [d] and Pilar Orero [e]

[a]Department of Translation, Interpreting and East Asian Studies, Universitat Autònoma de Barcelona, Barcelona, Spain; [b]Institute of Psychology, SWPS University of Social Sciences and Humanities, Warsaw, Poland; [c]College of Engineering, Computing and Applied Sciences, Clemson University, SC, USA; [d]School of Computer Science, Salford University, Salford, UK; [e]Department of Translation, Interpreting and East Asian Studies, Universitat Autònoma de Barcelona, Barcelona, Spain

ABSTRACT

Virtual and Augmented Reality, collectively known as eXtended Reality, are key technologies for the next generation of human–computer–human interaction. In this context, 360° videos are becoming ubiquitous and especially suitable for providing immersive experiences thanks to the proliferation of affordable devices. This new medium has an untapped potential for the inclusion of modern subtitles to foster media content accessibility (Gejrot et al., 2021), e.g., for the deaf or hard-of-hearing people, and to also promote cultural inclusivity via language translation (Orero, 2022). Prior research on the presentation of subtitles in 360° videos relied on subjective methods and involved a small number of participants (Brown et al., 2018; Agulló, 2019; Oncins et al., 2020), leading to inconclusive results. The aim of this paper is to compare two conditions of subtitles in 360° videos: position (head-locked vs fixed) and colour (monochrome vs colour). Empirical analysis relies on novel triangulation of data from three complementary methods: psycho-physiological attentional process measures (eye movements), performance measures (media content comprehension), and subjective task-load and preferences (self-report measures). Results show that head-locked coloured subtitles are the preferred option.

ARTICLE HISTORY
Received 16 September 2022
Accepted 1 September 2023

KEYWORDS
Immersive environments; 360° video; media accessibility; subtitles; captions; eye-tracking

1. Introduction

Immersive content is meant to give the user the illusion of being 'physically present' (Slater & Wilbur, 1997), and can provide benefits in a variety of sectors, such as entertainment, communication, learning, arts, and culture (Liberatore & Wagner, 2021; Montagud-Climent et al., 2020). 360° videos – also known as immersive or VR360 videos – offer great potential in providing engaging media experiences. Already in 2017, 49% of the public broadcasters who responded to the European Broadcasting Union (EBU,

CONTACT Marta Brescia-Zapata marta.brescia@uab.cat

2017) report on the use of VR declared offering 360° content. The most popular devices to access this immersive content are head-mounted displays (HMD). According to the newsletter XR today (Greener, 2022), 'in 2020, roughly 57.4 people owned a VR headset in the US, although in 2022, this figure increased by 37.7 million'. The demand for VR headsets is expected to increase due to the adoption of VR technology in enterprise, industry, and education sectors.

These new immersive media environments must be accessible for all to fulfil existing accessibility legislation in most world regions. Standards such as the EN301459 recommend Universal Design when developing any system or product. This requirement is becoming mainstream, with advice from the United Nations Convention on the Rights of Persons with Disabilities (UN, 2006), and now it is also an issue of political will and moral obligation, thanks to the Audiovisual Media Services Directive (AVMSD) and the European Accessibility Act (EAA). These pieces of legislation adopted a user-centric approach, a method that is at the heart of Human Rights towards full democratic participation in society by all, which, in the twenty-first century, depends on access to media. Following the aforementioned legislation and academic research (Romero-Fresco, 2013; Udo & Fels, 2010), any system or process should be designed with accessibility in mind from the outset, leading to a born accessible system that avoids expensive and complex afterthought solutions.

This paper focuses on subtitling, where standardised practices have been adopted in the context of 2D non-immersive media (ISO 20071:23; Matamala & Orero, 2018). Although the trend towards mixed methods in translation accessibility studies is becoming more popular, publications still fail to discuss the mixed-method nature of the study in depth (Hermosa-Ramírez, 2022). Prior reception studies on the evaluation of subtitles in 360° have largely been based on subjective measures, using questionnaires, interviews, and focus groups (Agulló & Matamala, 2020; Agulló & Orero, 2017; Brown et al., 2018; Fidyka & Matamala, 2018; Rothe et al., 2018). Objective, psychophysiological measures (such as eye movements, heart rate, and electrodermal activity) have been largely adopted in the context of tests for 2D non-immersive subtitles (Krejtz et al., 2016; Kruger, 2016; Liao et al., 2021; Szarkowska & Gerber-Morón, 2019), with only a recent contribution by Ibourk and Al-Adwan (2019), who included for the first time eye-tracking technology in a small study on immersive subtitling. The present study adopts a mixed-method design, with psycho-physiological process metrics (eye movements), performance metrics (scene comprehension), and subjective self-reports (task-load and preferences), followed by result triangulation. In this article, a critical review of previous studies related to subtitling in immersive media is presented, followed by the characteristics of the study and the methodology. To finish, the results, a discussion, and conclusions are presented.

2. Background

Studies on subtitles in VR are relatively recent. Understanding the displaying mode for subtitles started at the British Broadcasting Corporation (BBC), which first identified the main challenges when developing subtitles for immersive content (Brown et al., 2017). As a result, four solutions for subtitle rendering were designed (Brown, 2017) and tested with 24 hearing participants (Brown et al., 2018). These four conditions were (a) fixed positioned, evenly spaced, (b) follow head immediately (also called Always Visible), (c)

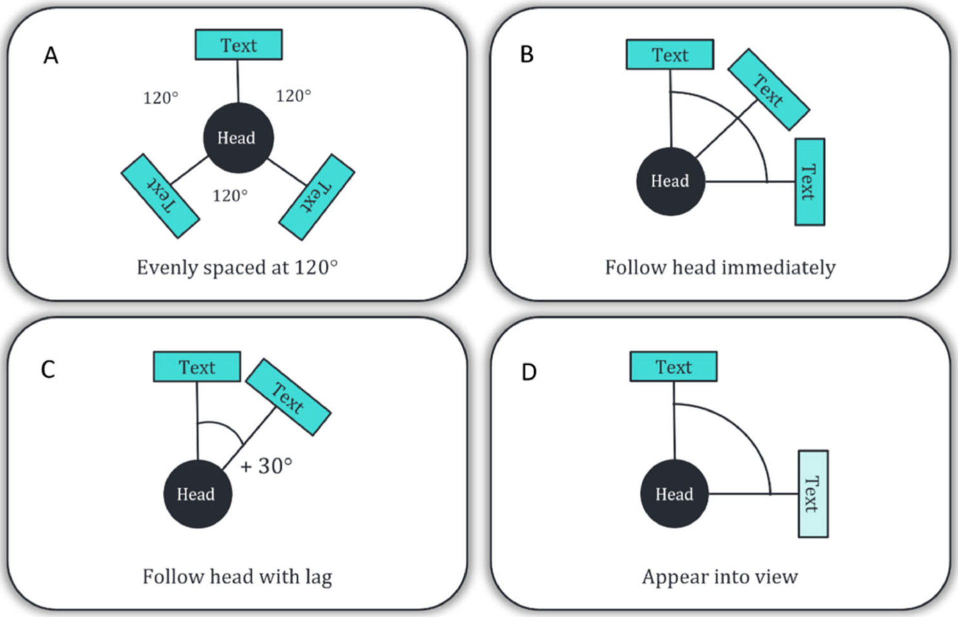

Figure 1. Four static subtitle solutions designed and tested by Brown et al. (2018).
Source: Author own elaboration based on Brown et al. (2018).

follow with lag, and (d) appear in front, then fixed position (see Figure 1). Users reported a preference for subtitles which are always presented in front of the viewer, following head movements: Always Visible subtitles.

Rothe et al. (2018) compared Brown et al. (2018) Always Visible subtitles with a new type called dynamic subtitles: placed near the speaker. They go a step further by measuring simulator sickness and task workload. They tested with 34 hearing participants and received feedback from one deaf participant (although this data was not included in their analysis). This study obtained similar results to those of its predecessor and pointed out that 'additional usage of an eye tracker could lead to more detailed results in the analysis of the viewing direction' (Rothe et al., 2018, p. 214).

The H2020 European-funded project ImAc,[1] based on the results of Brown et al. (2018) and Rothe et al. (2018), added location information within the 3D space to each subtitle (Agulló & Matamala, 2019). Results from Agulló and Matamala (2019) helped to draft user requirements, with a priority for designing a guiding mechanism to help deaf and hard-of-hearing users locate the sound source related to the subtitle. The project designed and tested several guiding mechanisms with six hearing and two deaf participants (Agulló et al., 2019). Results showed two preferred guiding methods: an arrow positioned to the left or right of the subtitle directing the user to the sound source, and a radar circle shown in the user field of view that identifies both the position of the sound source and the relative viewing angle of the user. Guiding mechanisms can also be found in videogames, such as *The Last of Us: Part II* (Myers, 2020), which 'includes a guide arrow to direct the user to the location of the character speaking' (Hughes et al., 2020). It is worth noting that the replication of the Agulló et al. (2019)

study with a larger number of participants ($N = 40$: 27 hearing, 20 with hearing loss, six hearing impaired and seven deaf) showed that always-visible subtitles with arrows were the preferred option (Agulló & Matamala, 2020).

On the one hand, all studies to this point conclude that the preferred visualisation mode is head-locked, centred, bottom subtitle; a trend that replicates conventions established for 2D, non-immersive content as defined in ISO/IEC 20071-23:2018 or UNE-153010:2012, among others. On the other hand, all studies focus on two subtitle features: sound source and subtitle position. All previous studies have three limitations: (a) demography samples are small in number, (b) the methodology is based solely on subjective opinions and answers (in-depth interviews, focus groups, and/or questionnaires), and (c) unstandardised stimuli videos in terms of length of video, language of dialogue in video, language of users, etc.

The most recent work in this field has adopted a new web-based prototyping framework (Hughes et al., 2020). The framework allows for real-time subtitle editing and visualisation in 360° videos. It takes into consideration all the previously reported work and is based on two mechanisms for subtitle rendering: (a) head-locked, where the subtitle is rendered relative to the user's viewpoint, and (b) fixed, where the subtitle is rendered relative to a fixed location in the world, usually close to the speaking character. This framework offers the opportunity of instant evaluation, which was tested by Brescia-Zapata et al. (2022) in a pilot study leading to a full experiment, the results of which are presented in this paper. The experiment received ethics clearance from the authors' home institutions, according to the ethics and privacy regulations of the H2020 EU-funded TRACTION project.[2]

The present study was conceived to further clarify which is the best visualisation mode for subtitles in immersive environments for all kinds of users. The aim is two-fold: (a) to detail the design of controlled experiments testing the influence of subtitles on users' attention allocation, and immersive content comprehension, and (b) to gather feedback regarding preferences of two characteristics of subtitles in immersive content: position and colour. Special attention has been put on the following methodological aspects: (a) diverse sample in terms of demography and subtitle usage habits, (b) a triangulation of psychophysiological, qualitative, and questionnaire methods, and (c) controlled stimuli. It focuses on two hypotheses: the first is related to subtitle positioning and the second is related to subtitle colour.

H1: Head-locked subtitles are easier to follow but more intrusive than fixed subtitles.

H2: Coloured subtitles are helpful to identify a speaking character in a 360° narrative.

3. Method

3.1. Experimental design and independent variables

To test the research hypotheses, the experimental study was conducted with 2×2 mixed design with two independent variables (IV): subtitle position (head-locked vs fixed) and subtitle colour (monochrome vs colour). The first one was treated as a between-subject IV and the later one as a within-subjects IV. The order of the experimental conditions was counterbalanced.

The two conditions for subtitle position were the following.

- (a) Head-locked subtitles which are always visible and are displayed in front of the viewer (see Figure 2), following head movements. Head-locked subtitles are equivalent to static-follow in Brown et al. (2017), to static subtitles in Rothe et al. (2018), and to always-visible in Agulló and Matamala (2020).
- (b) Fixed subtitles which appear near to the speaking characters and remain fixed to the scene (Figure 3). Fixed subtitles are equivalent to 120 degrees in Brown et al. (2017), and to dynamic subtitles in Rothe et al. (2018).

The main reason for testing these two conditions is that results from previous studies (Agulló et al., 2018; Brown et al., 2018; Rothe et al., 2018) only included qualitative data obtained through questionnaires and focus groups.

The two conditions for subtitle colour were the following.

- (a) Monochrome (see Figure 4). All the subtitles for each character were in the same colour white on grey background.
- (b) Colour (see Figure 5). Four colours have been used to identify the different characters: white, yellow, cyan, and green. The colour coding follows the ISO 20071-23:2018 standard.

The objective of examining these two conditions is to evaluate the effectiveness of colour coding, which is suggested by certain guidelines, such as BBC (2022), as a means of distinguishing between speakers and has demonstrated an enhancement in the immersive experience of multilingual 2D content (Szarkowska & Boczkowska, 2022). However, despite its potential benefits, this feature has not been treated as a variable in previous studies on immersive content (Agulló et al., 2018; Brown et al., 2018; Rothe et al., 2018).

Figure 2. Head-locked subtitles attached to the FoV in *Vacations* video stimuli.
Source: Author own elaboration.

Figure 3. Fixed subtitles attached to one position in the sphere in *Vacations* video stimuli.
Source: Author own elaboration.

3.2. Participants

In total, 73 volunteers took part in the experiment at three locations: $N = 24$ in Universitat Autònoma de Barcelona (UAB) (Spain), $N = 24$ in Salford University (UK), and $N = 25$ in SWPS University of Social Sciences and Humanities (Poland). Participants received no incentives for the study. There were no significant differences between the countries in age of the participants $F(2,69) = 2.26$, $p > 0.1$ (see Table 1).

The three locations of the study were chosen in order to ensure the robustness of the obtained results, as each country has a different language translation tradition. Spain has traditionally been considered a dubbing country (Ballester Casado, 1998; Chaume, 2012;

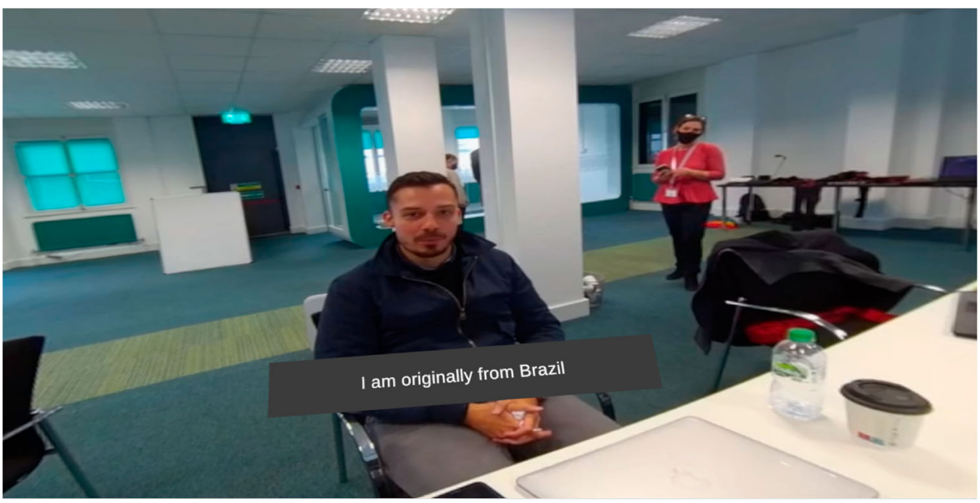

Figure 4. Monochrome subtitles in *TRACTION* video stimuli.
Source: Author own elaboration.

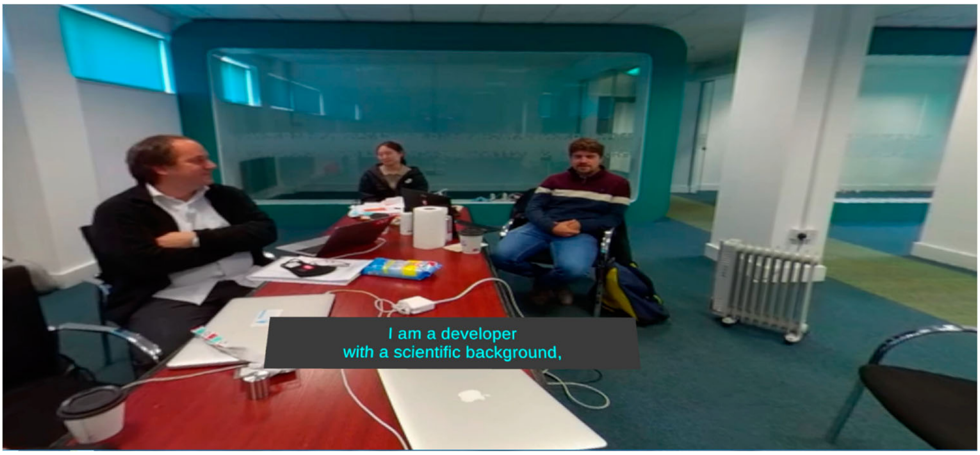

Figure 5. Coloured subtitles in *TRACTION* video stimuli.
Source: Author own elaboration.

Gil Ariza, 2004), although the presence of subtitles has increased in recent years (Matamala et al., 2017). Poland is generally considered a stronghold of voice over (Gottlieb, 1998; Valdeón, 2022), coexisting with other audiovisual translation modes, such as dubbing, subtitling, audio description, and subtitling for the deaf and the hard-of-hearing (Szarkowska, 2009). The UK belongs to a large anglophone audiovisual market and is neither a classical 'subtitling' nor 'dubbing' country (Luyken, 1991), as most audiovisual content is produced in English.

All participants in the sample had above average reading skills. We can also consider them as digital media savvy based on their declared daily usage of digital media and devices although most participants had never experienced VR content before (see Table 1).

Table 1. Participants' demographic characteristics, attitude, and usage of VR technology in each location.

	SPAIN	UK	POLAND
Number (*N*)	24	24	25
Age	($M = 33.96$, $SD = 11.18$)	($M = 29.41$, $SD = 10.34$)	($M = 27.40$, $SD = 6.68$)
Gender	17 female	4 female	19 female
Age difference between genders	$t(21) = 0.52$, $p = 0.61$	$t(22) = 0.70$, $p = 0.49$	$t(23) = 0.94$, $p = 0.36$
Occupation	Students ($N = 11$); Researchers/ lecturers ($N = 11$); Practitioners ($N = 2$)	Students ($N = 18$); Researchers/ lecturers ($N = 3$); Data Analysts/ engineers ($N = 4$)	Students ($N = 16$); Researchers/lecturers ($N = 4$); Journalist ($N = 2$); Analyst ($N = 1$)
Vision	Corrected ($N = 12$), Uncorrected ($N = 12$)	Corrected ($N = 12$), Uncorrected ($N = 15$)	Corrected ($N = 7$), Uncorrected ($N = 18$)
Handedness	Right ($N = 24$)	Right ($N = 21$)	Right ($N = 25$)
VR interest	($M = 4.13$, $SD = 0.76$)	($M = 3.96$, $SD = 0.65$)	($M = 3.00$, $SD = 1.22$)
VR experience	($M = 0.87$, $SD = 1.18$)	($M = 0.81$, $SD = 1.24$)	($M = 1.24$, $SD = 2.01$)
Digital device daily usage	($M = 3$, $SD = 1.38$)	($M = 3.59$, $SD = 1.28$)	($M = 2.91$, $SD = 0.91$)
Attitude to subtitles (answers to 'I always turn subtitles on')	($M = 3.97$, $SD = 1.16$)	($M = 3.88$, $SD = 1.32$)	($M = 4.01$, $SD = 0.97$)

Source: Author own elaboration.

3.3. Experimental procedure

The experimental procedure was the same in the three locations of the present study (see Figure 6). The experimental procedure was executed individually in the laboratory facilities. After signing the consent form, participants started with the demographic and usage and attitudes questionnaire. Next, they were familiarised with the VR headset. The HMD built-in eye tracker was calibrated. The following main part of the experiment consisted of three steps:

(a) watching first video stimuli followed by a comprehension and NASA-TLX questionnaire;
(b) watching second video stimuli followed by comprehension and NASA-TLX questionnaire; and
(c) a questionnaire comparing subtitles in both viewed videos.

After the experiment participants were thanked and debriefed.

A facilitator was present during the experiment, providing information as required, helping the participants to correctly wear the HMD, and assisting with the completion of the questionnaires (see Figure 6, top-right). All the questionnaires were formulated using Qualtrics system and completed using a desktop computer from the laboratory.

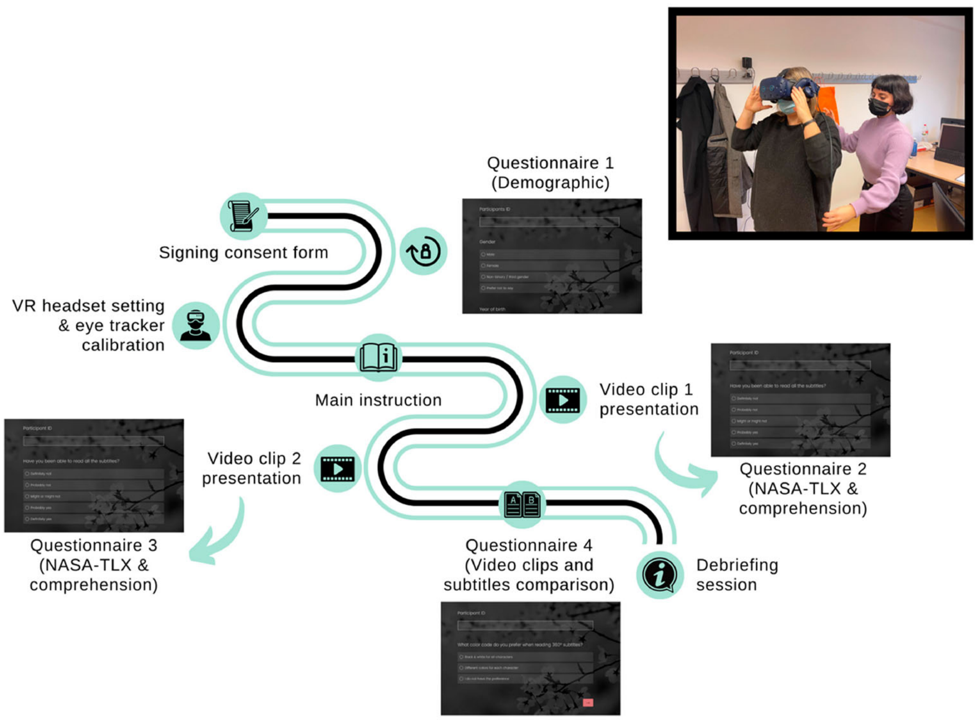

Figure 6. Experimental procedure schematic timeline and settings (top-right).
Source: Author own elaboration.

The order of video stimuli presentation was counterbalanced; thus, participants were presented with the two videos in random order, one with monochrome and the other with coloured subtitles, but the same subtitle position in both cases (either head-locked or fixed). During video stimuli presentation, the participants' eye movements were recorded.

3.4. Experimental materials

3.4.1. Stimuli video

For the present study, two custom 360° stimuli videos were recorded using an Insta360 One X2 camera. The first is called *Vacations* and was of a family, speaking Arabic, discussing their vacation plans. In the second stimuli video, called *TRACTION*, a group of researchers introduce themselves, each speaking in their own language (Spanish, Korean, Catalan, Portuguese, and English). *Vacations* is 1 min and 18 s long and *TRACTION* is 2 min and 32 s long. Both videos followed the same features.

(a) Speaking characters were placed all around the scene, exploiting 360° movement and immersion.
(b) There were no overlaps between speaking characters. If two characters located at different points in the 360° scene spoke at once, it would be almost impossible for the user to read the subtitles.
(c) Language spoken was different to that of participants. This feature was added after the pilot study (Brescia-Zapata et al., 2022), as, during the data collection, several participants claimed that they did not need to read the subtitles because they understood the original audio.

Both videos were subtitled in English, Spanish, and Polish to meet the experimental design (two colour schemes, two positioning schemes). The number of subtitles is the same in all languages: in *Vacations* there are 27 subtitles and in *TRACTION* 39. Participants at UAB were presented with video stimuli subtitles in Spanish, participants at Salford University were presented with video stimuli subtitled in English, and participants at Warsaw University were presented with video stimuli in Polish. Subtitles followed the ISO 20071-23:2018 standard and the font type selected was Liberation Sans. Another feature that differentiates the pilot study from this experiment is the grey background box used in both videos. Although some previous studies proved that it obstructs important parts of the image (Brown et al., 2018), results from the pilot test revealed that 'some of the participants found subtitles with no background hard to read' (Brescia-Zapata et al., 2022).

3.4.2. Questionnaires

Four questionnaires were designed to gather information about demographics, task-load, comprehension, and preference. The first questionnaire had 20 questions to gather demographic data (sex, age, handedness, occupation) and data related to usage and attitudes towards digital media (e.g., 'Which of these devices do you use on a daily basis?'), VR (e.g., 'How often do you watch virtual reality (VR) content for instance, 360° videos?'), and subtitles (e.g., 'When subtitles are available, do you turn them on?', 'How many hours a day do you watch subtitled content?').

After each stimuli video, participants filled out a questionnaire with six multiple choice questions on content comprehension. These questions were customised for each stimuli video. After comprehension questions, participants completed a task-load index (NASA-TLX) (Hart, 2006) consisting of six questions on personal evaluation of task load (e.g., mental load, physical demands, time pressure, etc).

After viewing two stimuli videos and completing the subsequent questionnaires, participants answered three questions on subtitle preferences ('What colour code do you prefer when reading 360° subtitles?', 'Explain in your own words why you prefer the option selected in the previous question', and 'Do you think that the subtitles obstruct important parts of the image?')

3.5. Experimental equipment

The experimental procedure was prepared with a framework developed by Hughes et al. (2020) and ported to Unity 3D. We used the HTC Vive Pro Eye headset, which contains a Tobii eye tracker built in, with 120 Hz sampling rate. 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 × 1,600 pixels in total, a display refresh rate of 90 Hz, and a FoV of 110°.

4. Results

Hypotheses testing was based on a series of Multilevel Linear Models (MLMs) with random effects for subjects and for stimuli videos. Each test we started with the Null model with random effects in it only, then consecutive models added fixed effects for country, subtitle position, and colour. When also needed, the interaction of position and colour were added as fixed factor. Each consecutive model was tested against the previous one with the chi-square test to check if a new fixed term improves the fit of the model. All the statistical analyses were performed in R language for statistical computing (R Core Team, 2020).

H1 was 'head-locked subtitles are easier to follow but more intrusive than fixed subtitles'. To test this hypothesis, we evaluated general comprehension of the content, subjective effort and frustration, and attention allocation over subtitles and scene Areas-of-Interest (AOI). In detail, we expected that head-locked positioning would cause better comprehension of the content, lower frustration, and effort. We also expected that head-locked positioning would cause relatively more fixations on subtitle AOI than fixed positioning.

H2 was 'coloured subtitles are helpful to identify a speaking character in a 360° narrative'. To test this hypothesis, we evaluated comprehension of the content. We expected that coloured subtitles allowed participants to identify speaking characters more easily than monochrome subtitles.

4.1. Influence of subtitle positioning and colour code on content comprehension and self-reported effort

To test the prediction that head-locked positioning would lead to better content comprehension, we ran MLM models on the general comprehension score from the comprehension questionnaires (see Figure 7a and Appendix Table A1 for model detailed tables and

Appendix Table A2 for estimated means for the interaction between subtitle position and colour).

The null model with random effects for subject and video showed $SD = 0.10$ for subject intercepts and $SD = 0.32$ for video intercept, and $SD = 0.99$ for residuals. The intercept of the model was significantly higher than zero ($b_0 = 3.52$, $t(3.12) = 13.70$, $p < 0.001$).

The country fixed effect did not influence the model fit (AIC = 1775.1, $\chi^2(2) = 0.74$, $p = 0.69$). Contrary to expectations, subtitle position did not improve significantly the model fit (AIC = 1778.5, $\chi^2(1) = 2.38$, $p = 0.12$). However, in line with expectations, subtitle colour significantly improved the model fit ($\chi^2(1) = 8.64$, $p = 0.003$). The full model included also the interaction of subtitle colour and position as a fixed effect, but it did not influence the fit ($\chi^2(1) = 1.91$, $p = 0.17$).

The full model fixed effects parameters showed that subtitle colour significantly improved comprehension of the movie ($b = 0.37$, $t(497.99) = 3.05$, $p < .001$). The estimated mean for colour subtitles was ($M = 3.63$, $SE = 0.31$) and for monochrome subtitles ($M = 3.38$, $SE = 0.31$).

Head-locked subtitle position improved comprehension but was only marginally significant ($b = 0.50$, $t(88.23) = 1.93$, $p = 0.06$), while the interaction of subtitle colour and position did not alter significantly the movie comprehension ($b = -0.23$, $t(498.68) = -1.38$, $p = 0.17$).

To test the prediction that head-locked positioning would cause lower effort, a similar analysis on comprehension MLM approach was used (see Figure 7b and also Appendix Table A3 for model detailed tables and Appendix Table A4 for estimated means for the interaction between subtitle position and colour). The null model showed $SD = 2.20$ for subject intercepts and SD = 1.01 for video intercept, and $SD = 1.64$ for residuals. The intercept of the model was statistically significant ($b_0 = 3.68$, $t(2.55) = 4.80$, $p = 0.02$).

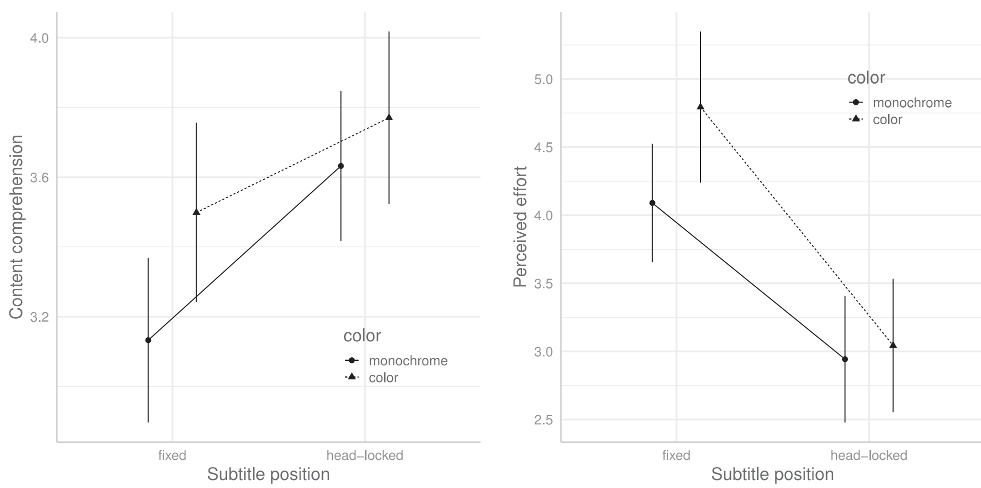

Figure 7. Content comprehension (Figure 7a) and self-evaluation of perceived effort (Figure 7b) depending on subtitle position (fixed or head-locked) and colour. Figure 7a. Content comprehension depending on subtitle position (fixed or head-locked) and colour. Figure 7b. Self-evaluation of perceived effort depending on subtitle position (fixed or head-locked) and colour.

Note: bar height represents the estimated means and whiskers ±1SE.

Comparison of models with fixed factors added showed that country did not improve the model ($AIC = 2247.5$, $\chi^2(2) = 2.84$, $p = 0.24$) but both the subtitle position ($AIC = 2241.9$, $\chi^2(1) = 7.58$, $p = 0.005$) and subtitle colour did ($AIC = 2236.9$, $\chi^2(1) = 7.05$, $p = 0.008$), as well as an interaction term between the colour and position of the subtitles ($AIC = 2234.4$, $\chi^2(1) = 4.43$, $p = 0.04$).

The analysis of the full model fixed parameters showed that colour subtitles in general add significant effort to the task ($\beta = 0.70$, $t(471.01) = 3.37$, $p < 0.001$). Head-locked subtitles lower significantly the perceived effort ($\beta = -1.15$, $t(79.09) = -2.21$, $p = 0.03$). Significant interaction term shows that head-locked subtitles lower the effort mainly when they are presented in colour ($\beta = -0.60$, $t(469.41) = -2.11$, $p = 0.04$).

4.2. Influence of subtitle positioning on visual attention directed toward subtitles

We hypothesised that fixed positioning causes more fixations and longer fixation time over subtitles. To test these predictions, we applied MLM analyses.

4.2.1. Fixation count

The first analysis used proportional fixation count (the number of fixations divided by the number of words in each subtitle) as a dependent variable. The MLM started with the null model with random effects of participants and stimuli video in it. In the next steps, the model was updated with the fixed effects of country, subtitle position, and subtitle colour, and interaction fixed effects of position and colour, position, and country. The models were tested against each other with chi-squared statistics to see if the fixed effect significantly improved the model fit (see Figure 8 and also Appendix Table A5 for model detailed tables and Appendix Table A6 for estimated means for the interaction between subtitle position and country).

The null model showed $SD = 16.20$ for subject intercepts and $SD = 19.30$ for video intercept, and $SD = 13.32$ for residuals. The intercept of the null model was not statistically significant ($\beta_0 = 39.06$, $t(1.04) = 2.83$, $p = 0.21$).

Models' comparison shows the significant effect on the model fit of a country as a fixed effect: $\chi^2(2) = 47.03$, $p < 0.001$, $AIC = 1217.10$, $BIC = 1234.90$. Also, the interaction effect of country and position significantly improved the model fit: $\chi^2(2) = 19.53$, $p < 0.001$, $AIC = 1165.28$, $BIC = 1197.87$. The final models' total psycho-$R^2 = 0.78$.

The full model fixed parameters showed that fixed subtitles significantly lower the proportional fixation count compared to head-locked ($\beta = -12.84$, $t(91.72) = -2.39$, $p = 0.02$). This effect was quantified by country. The pairwise comparisons of means showed that the differences between fixed and head-locked subtitles were significant in Spain ($t(66.6) = -4.22$, $p < 0.001$) and in Poland ($t(65.0) = 2.20$, $p = 0.03$), but not in the UK ($t(65.3) = -1.15$, $p = 0.26$). Note that the difference in fixation count between head-locked and fixed subtitles in the Polish sample was in favour of head-locked, unlike in the Spanish sample.

4.2.2. Total fixation time

The second MLM analysis used proportional total fixation time (total fixation time divided by visibility time of subtitle) as a dependent variable. The procedure of testing

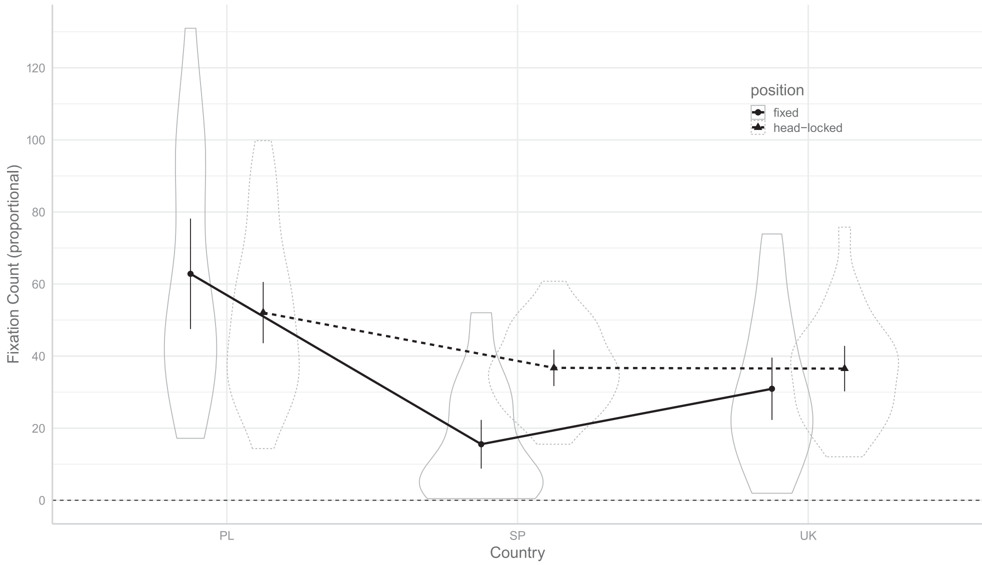

Figure 8. Proportional fixation count on subtitles AOI depending on subtitle position (fixed or head-locked) and country (Poland, Spain, or UK).

Note: bar height represents the estimated means and whiskers ±1SE. Violin shapes represent the distribution of the data points in each experimental condition in different countries.

and modelling random and fixed effect structure was identical to the analysis of fixation count described in subsection 4.2.1 (see Figure 9 and also Appendix Table A7 for model detailed tables and Appendix Table A8 for estimated means for the interaction between subtitle position and country).

The null model (with random effects only) for total fixation time showed $SD = 3537.14$ for subject intercepts, $SD = 6696.95$ for video intercept, and $SD = 3238.18$ for residuals. The intercept of the null model was not statistically significant ($\beta_0 = 10322.92$, $t(1.02) = 2.17$, $p = 0.27$).

The fixed effect of the country significantly improved the model fit ($\chi^2(2) = 16.96$, $p < 0.001$, $AIC = 2808.60$, $BIC = 2826.40$) as well as the fixed effect of subtitle position ($\chi^2(1) = 4.50$, $p = 0.03$, $AIC = 2806.10$, $BIC = 2826.90$). Additionally, the interaction term between country and subtitle position was marginally significant ($\chi^2(2) = 5.83$, $p = 0.054$, $AIC = 2808.10$, $BIC = 2840.70$). Because there was only a marginal model fit improvement, the interaction between country and subtitle positioning is hard to interpret.

However, the parameters of the model with the country and position fixed terms (with pseudo-R^2 (total) = 0.85) clearly show that the fixed subtitles gained more total fixation time than head-locked subtitles ($\beta = 1861.30$, $t(58.07) = 2.11$, $p = 0.04$). Significant effect of country shows also that participants in Spain spent significantly less time reading subtitles than participants in Poland ($\beta = 12718.09$, $t(68.28) = -2.51$, $p = 0.01$). See Figure 9 for the estimated means of proportional total fixation time for both effects.

4.3 Influence of subtitle colour on character recognition

Null multilevel model on character recognition, with random effects only for participants and stimuli video, showed $SD = 0.67$ for subject random effect and $SD = 1.60$ for video,

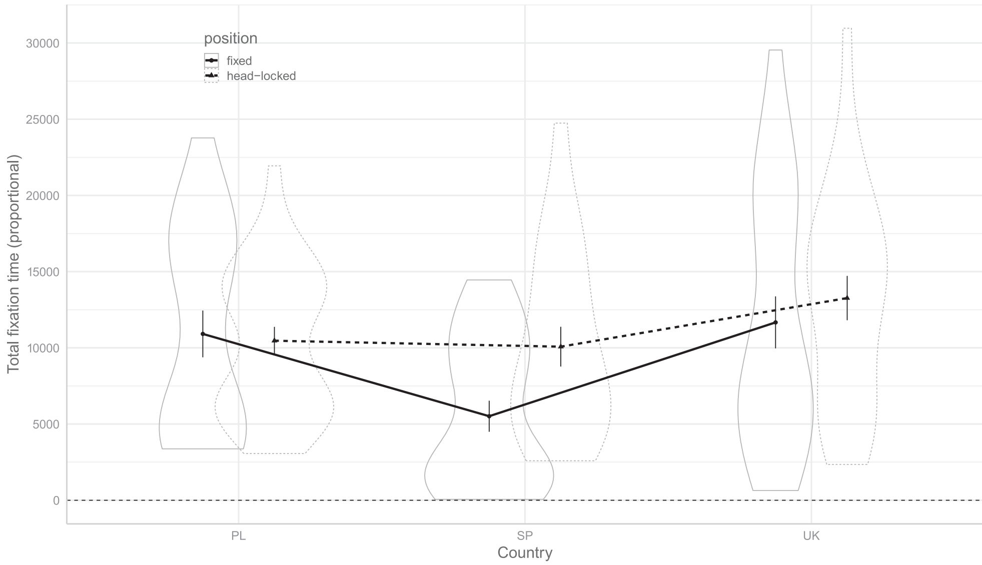

Figure 9. Proportional total fixation time on subtitles AOI depending on subtitle position (fixed or head-locked) and country (Poland, Spain, or UK).

Note: Bar height represents the estimated means and whiskers ±1SE. Violin shapes represent the distribution of the data points in each experimental condition in different countries. The interaction effect is not statistically significant in the full model.

and $SD = 0.69$ for residuals. The intercept of the null model was not statistically significant ($\beta_0 = 1.78$, $t(1.01) = 1.57$, $p = 0.36$).

The models fitting with the fixed effects for country, subtitle position, subtitle colour, as well as interactions between colour and position, and position and country, showed that subtitle colour code significantly improved the model fit ($\chi^2(1) = 12.26$, $p < 0.001$, $AIC =$ 1378.32, $BIC = 1413.14$, pseudo-R^2(total) = 0.86). The model parameters showed that colour subtitles significantly improves the recognition of characters ($\beta = 0.20$, $t(3.51) =$ 3.51, $p < 0.001$). The estimated average recognition of characters with monochrome subtitles was significantly lower ($M = 1.68$, $SE = 1.12$) than with colour subtitles ($M = 1.88$, $SE = 1.12$) (see Figure 10 and Appendix Table A9 for detailed parameters of the model).

Answers to the question 'what colour code do you prefer when reading 360° subtitles?' showed no significant preference between colour and monochrome subtitles ($p > 0.5$). Regarding coloured subtitles, one participant stated that 'colours distracted me from what they [the characters] were actually saying', but, on the contrary, another participant stated that 'it's easier to know which one is talking, especially because you aren't looking to all of them at the same time'.

These general results are in line with those obtained in Poland, but not with those obtained in Spain and the UK (see Table 2).

5 Discussion

The present study examined 360° video, although immersive subtitles can extend to any media asset. It focused on two hypotheses: the first related to subtitle positioning and the

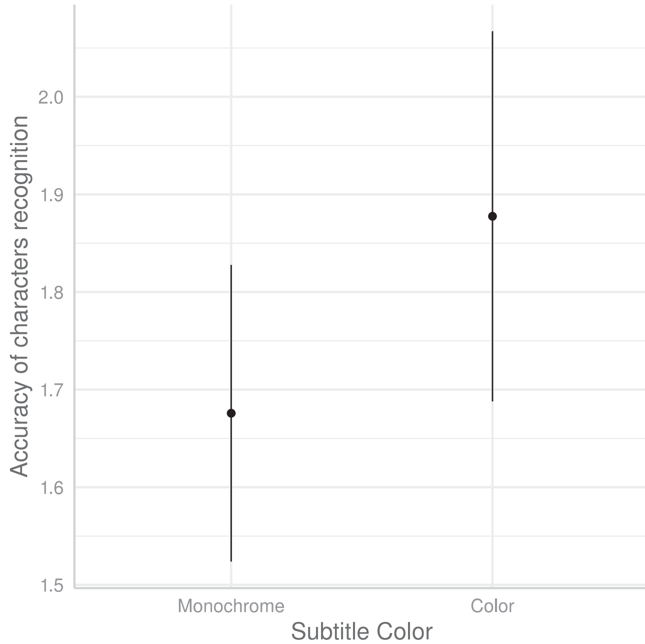

Figure 10. Estimated means of fixed effect of subtitle colour on accuracy of character recognition.

second related to subtitle colour. Regarding subtitle positioning, head-locked subtitles were subjectively evaluated as causing less cognitive effort in comparison to fixed subtitles. However, eye-tracking data showed that head-locked subtitles engaged more attention from the viewer than fixed subtitles. Head-locked subtitles gained significantly higher fixation counts and total fixation time, both of which are proportional to the number of words in the subtitle and visibility time of the subtitle, respectively. These results suggest that head-locked subtitles are more cognitively engaging, despite their subjective evaluation. The latter might be interpreted via the familiarity heuristic (Ashcraft & Radvansky, 2014), which states that people usually prefer things or situations that they are familiar with. However, it is important to note that this conclusion is specific to the Spanish sample. In the Polish sample, the difference in fixation count between head-locked and fixed subtitles favoured head-locked subtitles, which contrasts with the findings in the Spanish sample.

Table 2. Subtitle colour preferences between countries.

	Condition		
	Monochrome	Coloured	No preference
Poland	42.9%	40.8%	16.3%
Spain	39.1%	56.5%	4.3%
UK	58.3%	25%	16.7%
TOTAL	46.8%	40.5%	12.6%

Source: Author own elaboration.
The differences in subtitle colour preference between countries were statistically significant ($p = 0.02$). Spain participants stated a preference for coloured subtitles, while UK participants stated a preference for monochrome subtitles.

Subtitle positioning does not significantly affect video content comprehension. There was only a marginally significant effect of head-locked subtitles improving comprehension. The factor which influenced content comprehension significantly was the subtitles' colour scheme. Use of the different subtitle colours for each character improved character recognition and general comprehension; however, colour subtitles were evaluated by participants as causing more effort. Nevertheless, this is the first experiment testing subtitle colour in 360° videos, and it would be interesting to carry out further research on this topic to confirm our results.

The present article points out interesting evidence of discrepancies between self-report measures and more objective, psychophysiological, and comprehension metrics. Thus, we postulate to incorporate psychophysiological and performance methods and indicators in future studies on immersive media accessibility. It also points out the need for test environments or tools similar to what was proposed by Hughes et al. (2020) and which were used in this empirical user study.

An additional contribution of this study is the comparison of the different audiences from the sample. Replicating the experiment in three countries with different language translation traditions in audiovisual media allowed us to work with a heterogeneous sample and report differences between the participants. We found some significant differences both in perceived effort/frustration and AOI fixations depending on the country. Although it is not possible to define tendencies with the number of participants tested, it might be interesting to consider the preferences of people from different countries and cultures when testing subtitled immersive media.

Even though it was not the aim of the study, some participants mentioned the problem of speaker or audio source localisation when answering the questionnaires. This is in line with earlier work (Agulló et al., 2019; Hughes et al., 2019, Rothe et al., 2018) calling for user-guiding mechanisms such as arrows, radar, or other means. Although arrows have been shown to be the option preferred by users (Agulló, 2020), eye-tracked evaluation of such means holds potential for future work in the field.

6. Conclusions

The paper presented a review of the main studies regarding subtitles in immersive content and the main challenges posed by this new medium. The first aim of the study was to detail the design of controlled experiments involving immersive environments and subtitles. The methodology followed in the three locations where the experiment took place has been described in detail, including the design and IV, the participants, the procedure, the materials, and the equipment. This allows other researchers to replicate the experiment in other locations and/or using different video stimuli.

The second aim was to gather feedback regarding two characteristics of subtitles in immersive content: position and colour. To confirm the hypotheses, feedback on preferences and task-load was gathered, together with eye-movement data. The results have shown that head-locked subtitles, which are always visible and positioned in front of the viewer to align with their head movements, attracted more attention from viewers. However, it should be noted that this finding was observed specifically within the Spanish sample, and the observed trend in the Polish sample pointed in the opposite direction. In all samples, it was observed that participants consistently showed a preference for coloured subtitles as opposed to monochrome subtitles. This preference can be attributed to the enhanced ability to identify the speaking character facilitated by the use of colour. Although colour subtitles were perceived as requiring more effort, particularly when presented in a fixed position, they yielded improved comprehension of the content, especially in the fixed position, and better recall of the characters. Importantly, the impact of colour on comprehension and recall was consistent across the different country samples, indicating that it is not influenced by cultural factors.

The present study had some limitations regarding the type of content. To obtain meaningful results, two videos were recorded in a non-professional environment with non-professional actors. This might have had an impact on the results, and a replication of this study using professional content (e.g., a short movie, a documentary, a news program, or interviews) is encouraged.

Head-locked, coloured subtitles could be considered as good practice guidelines when subtitling in 360° videos. On the one hand, results about the position confirm previous studies, in which results were obtained through Igroup Presence Questionnaires (Agulló & Matamala, 2020). On the other hand, results about subtitle colour are not conclusive and further research in this regard is encouraged.

The study also highlighted the importance of the materials used for testing to ensure the ecological validity of the experiment. Special attention must be paid to the three languages involved in each experiment: the language of the film, the language of the subtitles, and the language of the participants. This problem has already been identified in other studies involving testing of different subtitle implementations, and various solutions have been adopted: the audio was muted (Agulló & Matamala, 2020; Kurzhals et al., 2017) or manipulated (Rothe et al., 2018). When testing with deaf and/or hard-of-hearing participants, one more language needs to be added depending on whether they are oralists or if they use sign language (signers).

To conclude, this is a first-of-its-kind analysis of eye-tracked visual attention on subtitle processing in 360° media content, with a focus on subtitle position and subtitle colour. Detailed examination of eye movements over subtitles combined with self-reports and performance measures has allowed an analysis of visual attention distribution and scene comprehension. We believe that the present work contributes to the novel frontiers on the implementation of accessibility services in immersive media studies.

Notes

1. https://www.imacproject.eu/
2. https://www.traction-project.eu/

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study is part of the article-based PhD thesis of Marta Brescia-Zapata in the Department of Translation and Interpreting at Universitat Autònoma de Barcelona (UAB) within the PhD program in Translation and Intercultural Studies. This research has been partially funded by the H2020 projects TRACTION (under Grant Agreement 870610) and MEDIAVERSE (under Gran Agreement 957252); and it is based upon work from COST Action LEAD ME, CA19142, supported by COST (European Cooperation in Science and Technology). The Commission's support for this publication does not constitute an endorsement of the contents, which reflects the views of the authors only, and the Commission cannot be held responsible for any use which maybe made of the information contained herein. Marta Brescia-Zapata and Pilar Orero are members of TransMedia Catalonia, an SGR research group funded by "Secretaria d'Universitats I Recerca del Departament d'Empresa.Coneixement de la Generalitat de Catalunya" (2021 SGR 00077).

Notes on contributors

Marta Brescia Zapata is a Ph.D. candidate in the Department of Translation, Interpreting and East Asian Studies at the Universitat Autònoma de Barcelona. She holds a BA in Translation and Interpreting from Universidad de Granada and an MA in Audiovisual Translation from UAB. She is a member of the TransMedia Catalonia research group (2017SGR113), where she collaborates in two H2020 projects: TRACTION (Opera co-creation for a social transformation), and GreenScent (Smart Citizen Education for a greeN fuTure). She is currently working on subtitling for the deaf and hard-of-hearing in immersive media, thanks to a Ph.D. scholarship granted by the Catalan government. She is the Spanish translator of Joel Snyder's AD manual 'The visual made verbal', and also collaborates regularly as subtitler and audio describer at the Festival INCLÚS.

Krzysztof Krejtz is a psychologist at SWPS University of Social Sciences and Humanities in Warsaw, Poland, where he is leading the Eye Tracking Research Center. In 2017, he was a guest professor at Ulm University, in Ulm, Germany. He gave several invited talks at, e.g., Max-Planck Institute (Germany), Bergen University (Norway), and Lincoln University Nebraska (USA). He has extensive experience in social and cognitive psychology research methods and statistics. In his research, he focuses on the use of eye-tracking methods and developing second-order eye-data-based metrics that may capture the dynamics of attention and information processing processes (transitions matrices entropy, ambient-focal coefficient K), dynamics of attention process in the context of Human–Computer Interaction, multimedia learning, media user experience, and accessibility. Dr Krejtz is a member of the ACM Symposium on Eye Tracking Research and Application (ACM ETRA) Steering Committee.

Andrew Duchowski is a professor of Visual Computing at Clemson University. He received his baccalaureate (1990) from Simon Fraser University, Burnaby, Canada, and doctorate (1997) from Texas A&M University, College Station, TX, both in Computer Science. His research and teaching interests include visual attention and perception, eye tracking, computer vision, and computer graphics. He joined the School of Computing faculty at Clemson in January, 1998. He has since produced a corpus of publications and a textbook related to eye-tracking research, and has delivered courses and seminars on the subject at international conferences. He maintains Clemson's eye-tracking laboratory, and teaches a regular course on eye-tracking methodology, attracting students from a variety of disciplines across campus.

Pilar Orero Ph.D. (UMIST, UK) works at Universitat Autònoma de Barcelona (Spain) in the TransMedia Catalonia Lab. She has written and edited many books, nearly 100 academic papers, and almost the same number of book chapters, all on Media Accessibility. She is leader and participant on numerous EU-funded research projects focusing on media accessibility. She works in standardisation and participates in the UN ITU IRG-AVA – Intersector Rapporteur Group Audiovisual Media Accessibility, ISO, and ANEC. She has been working on Immersive Accessibility for the past four years, first in a project called ImAc, the results of which are now further developed in TRACTION, MEDIAVERSE, MILE, and has just started to work on green accessibility in GREENSCENT. She leads the EU network LEADME on Media Accessibility.

Chris Hughes is a Lecturer in the School of Computer Science at Salford University, UK. His research is focused heavily on developing computer science solutions to promote inclusivity and diversity throughout the broadcast industry. This aims to ensure that broadcast experiences are inclusive across different languages, addressing the needs of those with hearing and low vision problems, learning difficulties, and the aged. He was a partner in the H2020 Immersive Accessibility (ImAc) Project. Previously he worked for the UX group within BBC R&D, where he was responsible for developing the concept of responsive subtitles and demonstrated several methods for automatically recovering and phonetically realigning subtitles. He has a particular interest in accessible services and is currently focused on developing new methods for providing accessibility services within an immersive context, such as Virtual Reality and 360° video.

ORCID

Marta Brescia-Zapata http://orcid.org/0000-0002-2465-0126
Krzysztof Krejtz http://orcid.org/0000-0002-9558-3039
Andrew T. Duchowski http://orcid.org/0000-0003-1681-7878
Chris J. Hughes http://orcid.org/0000-0002-4468-6660
Pilar Orero http://orcid.org/0000-0003-0269-1936

References

Agulló, B. (2020). Technology for subtitling: a 360-degree turn. Hermeneus. *Revista de Traducción e Interpretación*, *22*

Agulló, B., & Matamala, A. (2019). 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.pdf

Agulló, B., & Matamala, A. (2020). Subtitles in virtual reality: Guidelines for the integration of subtitles in 360° content. *Íkala, Revista de Lenguaje y Cultura*, *25*(3), 643–661. https://doi.org/10.17533/udea.ikala.v25n03a03

Agulló, B., Matamala, A., & Orero, P. (2018). From disabilities to capabilities: Testing subtitles in immersive environments with end users. *HIKMA*, *17*, 195–220. https://doi.org/10.21071/hikma.v17i0.11167

Agulló, B., Montagud, M., & Fraile, I. (2019). Making interaction with virtual reality accessible: Rendering and guiding methods for subtitles. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *33*(4), 416–428. https://doi.org/10.1017/S0890060419000362

Agulló, B., & Orero, P. (2017). 3D movie subtitling: Searching for the best viewing experience. *CoMe – Studi di Comunicazione e Mediazione linguistica e culturale*, *2*, 91–101. https://ddd.uab.cat/record/183656

Ashcraft, M. H., & Radvansky, G. A. (2014). *Cognition* (6th ed.). Pearson Education.

Ballester Casado, A. (1998). *Traducción y nacionalismo. La recepción del cine americano en España a través del doblaje, 1920–1948* [Translation and Nationalism. The Reception of American Cinemain SPain through Dubbing, 1920–1948]. Comares.

BBC. (2022). *Subtitle guidelines*. Version 1.2.1. https://www.bbc.co.uk/accessibility/forproducts/guides/subtitles/. Accessed 25 April 2023

Brescia-Zapata, M., Krejtz, K., Orero, P., Duchowski, A. T., & Hughes, C. J. (2022). VR 360o subtitles: Designing a test suite with eye-tracking technology. *Journal of Audiovisual Translation*, *5* (2), (In press). https://doi.org/10.47476/jat.v5i2.2022.184

Brown, A. (2017, October 26). *User testing for 360° content*. BBC. https://www.bbc.co.uk/rd/blog/2017-10-subtitles-360-video-virtual-reality-vr

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2017, June 14–16). Subtitles in 360-degree video. In *Adjunct Publication of the 2017 ACM International Conference* *on Interactive Experiences for TV and Online Video*. The Netherlands, 3–8. https://doi.org/10.1145/3084289.3089915

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2018). *Exploring subtitle behaviour for 360° video* (White Paper WHP 330). https://www.bbc.co.uk/rd/publications/whitepaper330

Chaume, F. (2012). *Audiovisual translation: Dubbing*. St Jerome Publishing.

European Broadcasting Union (EBU). (2017). Virtual reality: How are public broadcasters using it? https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it

Fidyka, A., & Matamala, A. (2018). Audio description in 360° videos: Results from focus groups in Barcelona and Kraków. *Translation Spaces*, *7*(2), 285–303. https://doi.org/10.1075/ts.18018.fid

Gejrot, E., Kjellstrand, S., & Laurin, S. (2021). New realities unlocking the potential of XR for persons with disabilities. Tech. rep., Funka Nu AB, Stockholm, Sweden.

Gil Ariza, M. C. (2004). A case study: Spain as a dubbing country. *Translation Journal*, *29*. http://translationjournal.net/journal/29movies.html

Gottlieb, H. (1998). Subtitling. In M. Baker (Ed.), *Routledge encyclopaedia of translation studies* (pp. 244–248). Routledge.

Greener, R. (2022, December 19). *Virtual reality statistics to know in 2023*. XR today. https://www.xrtoday.com/virtual-reality/virtual-reality-statistics-to-know-in-2023/#:~:text=In%202020%2C%20roughly%2057.4%20people,%2C%20industry%2C%20and%20education%20adoption

Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hermosa-Ramírez, I. (2022). Physiological instruments meet mixed methods in media accessibility. *Translation Spaces*, *11*(1), 38–59. https://doi.org/10.1075/ts.21020.her

Hughes, C., Brescia-Zapata, M., Johnston, M., & Orero, P. (2020, 14–18 December). *Immersive captioning: Developing a framework for evaluating user needs*. IEEE AIVR: 3rd International Conference on Artificial Intelligence & Virtual Reality, IEEE. https://doi.org/10.1109/AIVR50618.2020.00063

Hughes, C., Montagud, M., and tho Pesch, P. (2019, 5–7 June). Disruptive approaches for subtitling in immersive environments. In *Proceedings of the 2019ACM International Conference on Interactive Experiences for TV and Online Video – TVX '19*. https://doi.org/10.1145/3317697.3325123

Ibourk, E. M., & Al-Adwan, A. (2019). Subtitling virtual reality into arabic: Eye tracking 360-degree video for exploring viewing experience. *Lebende Sprachen*, *64*(2), 286–308. https://doi.org/10.1515/les-2019-0016

International Organization for Standardisation. (2018). *Information technology — User interface component accessibility* (ISO/IEC 20071). https://www.iso.org/standard/70722.html

Krejtz, I., Szarkowska, A., & Łoginska, M. (2016). Reading function and content words in subtitled videos. *The Journal of Deaf Studies and Deaf Education*, *21*(2), 222–232. https://doi.org/10.1093/deafed/env061

Kruger, J. (2016). Psycholinguistics and audiovisual translation. *Target. International Journal of Translation Studies*, *28*(2), 276–287. https://doi.org/10.1075/target.28.2.08kru

Kurzhals, K., Cetinkaya, E., Hu, Y., Wang, W., & Weiskopf, D. (2017). Close to the action: Eye–tracking evaluation of speaker–following subtitles. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, 6559–6568. https://doi.org/10.1145/3025453.3025772

Liao, S., Yu, L., Reichle, E. D., & Kruger, J. L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading*, *25*(5), 417–435. https://doi.org/10.1080/10888438.2020.1823986

Liberatore, M. J., & Wagner, W. P. (2021). Virtual, mixed, and augmented reality: A systematic review for immersive systems research. *Virtual Reality*, *25*(3), 773–799. https://doi.org/10.1007/s10055-020-00492-0

Luyken, G. M. (1991). *Overcoming language barriers in television: Dubbing and subtitling for the European audience*. European Institute for the Media.

Matamala, A., & Orero, P. (2018). Standardising accessibility: Transferring knowledge to society. *Journal of Audiovisual Translation*, *1*(1), 139–154. https://doi.org/10.47476/jat.v1i1.49

Matamala, A., Perego, E., & Bottiroli, S. (2017). Dubbing versus subtitling yet again? An empirical study on user comprehension and preferences in Spain. *Babel. Revue internationale de la traduction/International Journal of Translation*, *63*(3), 423–441. https://doi.org/10.1075/babel.63.3.07mat

Montagud-Climent, M., Orero, P., & Matamala, A. (2020). Culture 4 all: Accessibility-enabled cultural experiences through immersive VR360 content. *Personal and Ubiquitous Computing*, *24*(6), 887–905. https://doi.org/10.1007/s00779-019-01357-3

Myers, M. (June 12, 2020). The last of us part 2 review: We're better than this. Polygon. Vox Media. Archived from the original on June 12, 2020. Retrieved July 6, 2022.

Oncins, E., Bernabé, R., Montagud, M., & Arnáiz Uzquiza, V. (2020). Accessible scenic arts and virtual reality: A pilot study with aged people about user preferences when reading subtitles in immersive environments. *MonTI. Monografías de Traducción e Interpretación*, *12*(12), 214–241. https://doi.org/10.6035/MonTI.2020.12.07

Orero, P. (2022). Translation, accessibility and minorities. In E. Bielsa (Ed.), *The Routledge handbook of translation and media* (pp. 384–399). Taylor Francis.

R Core Team. (2020). *R: A language and environment for statistical computing*. R foundation for statistical computing. http://www.R-project.org

Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking. *The Journal of Specialised Translation*, *20*, 201–223.

Rothe, S., Tran, K., & Hussmann, H. (2018). Positioning of subtitles in cinematic virtual reality. In G. Bruder, S. Yoshimoto, & S. Cobb (Eds.), *International Conference on Artificial Reality and Telexistence and Eurographics Sysmposium on Virtual Environments, ICAT-EGVE 2018*. The Eurographics Association. https://doi.org/10.2312/egve.20181307

Sipatchin, A., Wahl, S., & Rifai, K. (2020). Eye-tracking for low vision with virtual reality (VR): Testing status quo usability of the HTC Vive Pro Eye. *bioRxiv*.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Szarkowska, A. (2009). The audiovisual landscape in Poland at the Dawn of the 21st century. In A. Goldstein, & B. Golubović (Eds.), *Foreign language movies – dubbing vs. subtitling* (pp. 185–201). Verlag Dr. Kovač.

Szarkowska, A., & Boczkowska, J. (2022). Colour coding subtitles in multilingual films – a reception study. *Perspectives: Studies in Translatology*, *30*(3), 520–536. https://doi.org/10.1080/0907676X.2020.1853186

Szarkowska, A., & Gerber-Morón, O. (2019). Two or three lines: A mixed-methods study on subtitle processing and preferences. *Perspectives: Studies in Translatology*, *27*(1), 144–164. https://doi.org/10.1080/0907676X.2018.1520267

Udo, J.-P., & Fels, D. I. (2010). Universal design on stage: Live audio description for theatrical performances. *Perspectives: Studies in Translatology*, *18*(3), 189–203. https://doi.org/10.1080/0907676X.2010.485683

UN. (2006). Convention on the Rights of People with Disabilities. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html

Valdeón, R. A. (2022). Latest trends in audiovisual translation. *Perspectives*, *30*(3), 369–381. https://doi.org/10.1080/0907676X.2022.2069226

Appendices

Table A1. Model parameters for content general comprehension.

	Estimate (*β*)	*SE*	*t*	*df*	*p*
Intercept	3.25	0.33	9.70	11.10	0.00
Country Spain	−0.11	0.30	−0.37	72.02	0.71
Country UK	−0.22	0.30	−0.76	71.94	0.45
Position head-locked	0.50	0.26	1.93	88.23	0.06
Colour colourful	0.37	0.12	3.05	497.99	0.00
Position head-locked X Colour colourful	−0.23	0.16	−1.38	498.68	0.17

Table A2. Estimated means from the model for content general comprehension for interaction effect of subtitle position and colour.

Position	Colour	Est. mean	*SE*
Fixed	monochrome	3.13	0.34
Head-locked	monochrome	3.63	0.33
Fixed	colourful	3.50	0.34
Head-locked	colourful	3.77	0.33

Table A3. Model parameters for self-perceived effort.

	Estimate (*β*)	*SE*	*t*	*df*	*p*
Intercept	4.57	0.90	5.07	4.52	0.01
Country Spain	−0.31	0.62	−0.50	68.96	0.62
Country UK	−1.12	0.61	−1.83	68.45	0.07
Position head-locked	−1.15	0.52	−2.21	79.09	0.03
Colour colourful	0.70	0.21	3.37	471.01	0.00
Position head-locked X Colour colourful	−0.60	0.29	−2.11	469.41	0.04

Table A4. Estimated means from the model for self-perceived effort for interaction effect of subtitle position and colour.

Position	Colour	Est. mean	*SE*
Fixed	monochrome	4.09	1.08
Head-locked	monochrome	2.94	1.07
Fixed	colourful	4.79	1.08
Head-locked	colourful	3.04	1.07

Table A5. Model parameters for proportional fixation number.

	Estimate (*β*)	*SE*	*t*	*df*	*p*
Intercept	64.93	14.23	4.56	1.17	0.11
Country Spain	−47.29	5.15	−9.18	66.55	0.00
Country UK	−31.90	5.11	−6.24	65.29	0.00
Position head-locked	−12.84	5.36	−2.39	91.72	0.02
Colour monochrome	−4.20	3.29	−1.28	69.19	0.21
Position head-locked X Colour monochrome	4.18	4.49	0.93	69.46	0.35
Country Spain: position head-locked	31.95	7.00	4.56	65.80	0.00
Country UK: position head-locked	16.34	6.90	2.37	65.11	0.02

Table A6. Estimated means from the model for proportional fixation number for interaction effect of subtitle position and country.

Position	Country	Est. mean	*SE*
Fixed	Poland	62.8	14.1
Head-locked	Poland	52.1	14.0
Fixed	Spain	15.5	14.1
Head-locked	Spain	36.7	14.1
Fixed	UK	30.9	14.0
Head-locked	UK	36.5	14.0

Table A7. Model parameters for proportional total fixation time.

	Estimate (β)	*SE*	*t*	*df*	*p*
Intercept	10756.03	4916.41	2.19	1.13	0.25
Country Spain	−5401.21	1556.41	−3.47	66.92	0.00
Country UK	756.23	1547.46	0.49	65.83	0.63
Position head-locked	−317.28	1575.22	−0.20	83.65	0.84
Colour monochrome	316.28	806.71	0.39	69.40	0.70
Position head-locked: colour monochrome	−257.78	1100.44	−0.23	69.58	0.82
Country Spain: position head-locked	5010.05	2118.56	2.36	66.45	0.02
Country UK: position head-locked	2041.71	2088.48	0.98	65.86	0.33

Table A8. Estimated means from the model for proportional total fixation time for interaction effect of subtitle position and country.

Position	Country	Est. mean	*SE*
Fixed	Poland	10914	4900
Head-locked	Poland	10468	4856
Fixed	Spain	5513	4880
Head-locked	Spain	10077	4888
Fixed	UK	11670	4878
Head-locked	UK	13266	4878

Table A9. Model parameters for accuracy of character recognition.

	estimate (β)	*SE*	*t*	*df*	*p*
Intercept	1.53	1.13	1.35	1.05	0.40
Country Spain	0.14	0.21	0.68	68.19	0.50
Country UK	0.12	0.21	0.60	68.10	0.55
Position head-locked	0.12	0.17	0.70	68.11	0.48
Colour colourful	0.20	0.06	3.51	500.42	0.00