



Research & Development White Paper

WHP 330

April 30, 2018

Exploring Subtitle Behaviour for 360° Video

A. Brown, J. Turner, J. Patterson, A. Schmitz, M. Armstrong, and M. Glancy

Design & Engineering
BRITISH BROADCASTING CORPORATION

Exploring Subtitle Behaviour for 360° Video

Andy Brown, Jayson Turner, Jake Patterson, Anastasia Schmitz, Mike Armstrong, Maxine Glancy

Abstract

360° video and Virtual Reality are powerful techniques for giving viewers a sense of ‘Being There’, and are becoming increasingly popular. However, giving the viewer the freedom to look around also results in a new challenge for subtitling. We have implemented and tested four different subtitle behaviours with a group of subtitle users, testing against a User Experience Framework for Subtitles. We report user preferences, and have performed statistical analysis on ratings given by participants and thematic analysis on the semi-structured interviews. The discriminatory themes were: effort, missing out, obstruction, distraction and immersion. Making subtitles easy to locate without restricting the viewer’s ability to explore the scene are key. The implementation using a standard head-up display was found to be the best performing, although further research is required to assess its suitability for longer content. We make recommendations for improving our implementation and highlight avenues for future research.

Additional key words: subtitles, accessibility, 360°, attention

White Papers are distributed freely on request.
Authorisation of the Chief Scientist is required for
publication.

©BBC 2016. Except as provided below, no part of this document may be reproduced in any material form (including photocopying or storing it in any medium by electronic means) without the prior written permission of BBC Research & Development except in accordance with the provisions of the (UK) Copyright, Designs and Patents Act 1988.

The BBC grants permission to individuals and organisations to make copies of the entire document (including this copyright notice) for their own internal use. No copies of this document may be published, distributed or made available to third parties whether by paper, electronic or other means without the BBC's prior written permission. Where necessary, third parties should be directed to the relevant page on BBC's website at <http://www.bbc.co.uk/rd/pubs/whp> for a copy of this document.

1 Introduction

Subtitles, or closed-captioning, are a method in which the audio content from a television show or film is shown to the viewer as text to accompany the broadcast. A large proportion of the population use subtitles, with this including users that do so out of both choice and necessity [1]. Subtitles are important in contributing to the understanding of content provided through video — the information they provide aids their users in interpreting and understanding content that they might be otherwise excluded from. The presence, accuracy, and intelligibility of subtitles are key in providing services that are accessible to all.

In recent years, however, new forms of media are emerging, with 360° video and Virtual Reality (VR)¹ starting to enter the mainstream (demonstrated by their prominence in the International Broadcasting Convention² and at the Consumer Electronics Show³ in 2016). One of the key attractions of 360° content is that it offers the viewer freedom to choose where to look, which, although challenging to the filmmaker [2], can enable a very affecting, ‘immersive’ experience. However, it is evident that these new formats should also be made accessible to as wide an audience as possible, and this demands subtitling.

Broadcasters need to present subtitles so they are visually pleasing but also effective: viewers should be able to read and comprehend subtitles at the same time as following the visual scenes that they accompany. From a UX standpoint there is a desire to deliver subtitle content in an immersive, engaging, emotive and aesthetically pleasing way. Traditionally, subtitles are positioned so they are centred at the bottom of the television screen (although there has been experimentation with other positioning techniques [3, 4, 5, 6, 7, 8]). Guidelines for subtitles (e.g., [9]) have long recommended that ‘viewers generally prefer the conventional bottom of the screen position’, while noting that different placement (e.g., top-screen) might be necessary to avoid obscuring important information, and that ‘it is most important to avoid obscuring any part of a speaker’s mouth’.

With 360° video, however, there is no equivalent ‘bottom of the screen’, nor can we know where people will be looking at any given time — the act of *positioning* subtitles becomes one of defining a *behaviour* for them. Existing guidelines can inform how the subtitles look, but in terms of behaviour, they can only act as a familiar reference point that will influence the expectations viewers will have. Specifically, we can identify the following additional challenges that any design will need to consider:

- Nowhere in the scene is guaranteed to be visible to the viewer.
- There will always be something behind the subtitle, and we may not know what this will be.
- A feeling of ‘Immersion’ is seen as an important attribute for 360° experiences, and it will be important that the subtitles disrupt this as little as possible.
- With the possibility of subtitles existing out of a viewer’s field of view, they need to be located in a way that makes sense to the viewer — finding them should not demand significant conscious effort by the viewer.
- Virtual Reality Sickness [10] is a significant problem for many viewers [11], and the behaviour of the subtitles should not exacerbate this.

360° content can be consumed in a number of different ways, each with its own advantages and disadvantages [12, 13]; in this work, we are concentrating on the scenario where the viewer

¹360° video is video where the camera has captured a full (or near-full) sphere of video; when viewed, it is possible to look in any direction at any time, although the viewer cannot move around the scene. VR is characterised by the user entering an environment that is virtually generated in real time and can respond to user’s actions, often including movement.

²<http://www.ibc.org/>

³<http://www.ces.tech/>

is watching the content using a head-mounted display. The core challenge is to design a subtitle behaviour that meets the user experience requirements as set out in the framework developed by Crabb *et al* [14].

2 Related Work

While existing guidelines [9] for Television subtitles (and the research on which they are based) can inform some aspects of how to present subtitles for 360° video, such as the speed and visual appearance, they offer little to inform how these subtitles should behave. We have been unable to find any reference to this particular problem in the academic literature, and have therefore taken a three-pronged approach. First we briefly explore what the literature has to say about attention and gaze patterns; second we look for examples of current practice in 360° video and VR, and; thirdly we look at how text is displayed in augmented reality (AR) systems.

It is widely known that people tend to orient their attention towards faces, both in a natural scene and on screen (Birmingham, Bischof, & Kingstone, 2008). The distribution of gaze between the eyes and mouth depends on task [15], but there remain many fixations on a speaker’s eyes even when observers are trying to lip-read in silence or under high noise conditions [16]. In a conversation setting, Klin *et al.* [17] found *“a clear relationship between speech and fixation, with most gazes directed towards the target who was speaking. However, there was also a significant tendency for looks to precede the change in speaker, such that participants fixated a target about 150 ms before they began to talk. Gaze was also associated with the observer’s judgement of the social hierarchy in the scene, with dominant targets receiving more attention than less influential group members — an effect independent of speaking time”*.

The presence of sound in a scene causes viewers to move their eyes more frequently than when viewing the same scene without sound [18]. Changes in auditory information linked to visual events in a video affect are also found to capture attention [18]. Sound also alerts observers to changes in conversation and attracts attention to the speaker, and leads to more similar gaze patterns between viewers [19]. It also makes people follow the speech turn-taking more closely [20].

In summary, therefore, it can be seen that supporting viewers of 360° scenes who have either limited or no ability is complex. Placing subtitles near the speaker would minimise disruption to their ‘natural’ gaze patterns (i.e., how the viewer would have looked had sound been available), and provide some of the attentional cues that they would otherwise miss. However, in the 360° scenario, this would also need to display subtitles for speakers that are not within the user’s current field of view. Given the complexity of this problem, our initial work in this area will concentrate on solutions that do not dynamically place subtitles according to the position of the speaker. This has the additional advantage of not requiring any additional steps to the subtitle authoring process beyond what is currently done for TV.

Turning now to what techniques are being implemented by practitioners, there is again little to report. The New York Times has published a number of 360° videos, and their example *“The Displaced”* [21] is one of the few examples of subtitles that we have found. In this example, subtitles are ‘burnt-in’ to the video, and appear in three equally spaced positions around the viewer. Best-practice guidelines for game developers [22] are primarily aimed at games played using a screen rather than head-mounted display, so are similar to those for television, advising accuracy, legibility, etc. In terms of positioning, the advice is to position *“in the centre and at the bottom, with a gap between the bottom of the subtitles and the bottom of the screen”*.

There are similarities between reading subtitles over 360° video and Augmented Reality (AR): data (in this case the text of subtitles) is placed over the moving background video. In the field of AR, Orlosky *et al.* [23] explored how to place text notifications (e.g., SMS messages) in an AR headset. Finding that *“users would prefer text that is affixed to visible locations in the environment rather than affixed to a single point on the HMD screen”*, they describe a system that could identify ‘dark, uniform points in the environment’ in which to place text. As part of this work, participants

were asked to view a video and manually select positions in which to overlay text; it was found that “users tend to have a general tendency to place text just below screen center”.

Also in the field of AR, Charrissis and Naef looked at head-up displays for car driver assistance, and found that longer focal lengths for the display (5m) was more comfortable for the user than shorter ones (0.7m and 2.5m) [24].

One final area that is important background for this is VR sickness [10]. Motion sickness is proposed to result from someone’s inability to maintain postural stability [25]. This could potential be induced by reading text fixed in front of oneself while turning the head: the background video appears to move around the static text.



Figure 1: Screenshot from clip 1. The 360° video is projected onto the inside of a 25m diameter sphere, with the viewer at its center. The subtitles are displayed at 4m from the user with a visual style that aims to replicate that of standard subtitling on broadcast television, i.e., black background, fitted to text, sans-serif font.

3 Subtitle Behaviours

Taking into account the existing literature and design considerations discussed earlier, we have derived four modes of subtitling behaviour. These behaviours are have been implemented in Unreal Engine 4⁴ with an Oculus Rift DK2⁵. Equirectangular, non-stereoscopic, videos are texture-mapped onto the inside of a large sphere (25m in diameter) in the virtual environment with a camera positioned and locked to its centre. The camera is able to rotate in three degrees-of-freedom and is controlled by input from the Oculus Rift.

All of the behaviours use blocks of subtitles that are placed and manipulated in the virtual environment. The blocks face inwards towards the camera and are fixed at a distance of 4m⁶. Any movement of the subtitle blocks occurs as rotation around the centre of the sphere.

As shown in Figure 1, we have aimed to replicate the style of standard broadcast subtitles as much as possible [9], namely we have surrounded white sans-serif text with an opaque black fitted

⁴<https://www.unrealengine.com>

⁵<https://www3.oculus.com/en-us/dk2/>

⁶The distances of the video sphere and subtitles from the viewer were derived from informal user-testing, and are parameters that should be investigated more carefully.

box and show a maximum of two lines of text at any given time. The subtitle blocks are formed by taking each section of speech, i.e., each paragraph in the transcript, and then subdividing these paragraphs into roughly equal sections of up to 80 characters in length. These are then divided, where appropriate, into two lines of up to 40 characters. These divisions are adjusted to take account of the language structures such as punctuation, prepositions and articles.

3.1 A: Evenly spaced

This behaviour displays duplicate subtitles at three positions, facing inwards at angles fixed around the centre of the VR environment. As shown in Figure 2, subtitles are located at 0° , 120° , and 240° about the viewer, and at 15° below the horizon⁷. This behaviour is included as an example of current practice [21], and has the added benefit that it could be ‘burnt-in’ to the video using a video-editor.

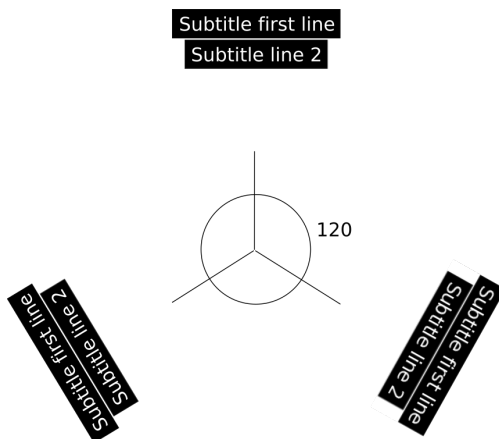


Figure 2: Evenly spaced: Three subtitle blocks are located in the environment at 120° angles around the user (i.e., the centre of the diagram)

3.2 B: Follow head immediately

It is common in VR experiences to show information as part of a ‘head-up display’ (HUD). A HUD typically displays graphics that are fixed in front of the viewer at all times regardless of their posture and pose in a VR environment. Figure 3 demonstrates the ‘*Follow head immediately*’ behaviour, which utilises this same basic HUD mechanic. In this case, the subtitles follow the viewer as they look around the environment — the subtitle appears to the viewer as though the text is static relative to their head. The placement is directly ahead, but 15° vertically below eye-level (i.e., 15° below directly ahead). A potential drawback for this is that it has been reported that overuse of the HUD can cause issues with nausea [10, 11]. Unlike our *Evenly spaced* behaviour, in *Follow head immediately* subtitles are not fixed within the environment and can travel on both horizontal and vertical axis.

3.3 C: Follow with lag

The ‘*Follow with lag*’ behaviour (Figure 4) attempts to address the sickness reportedly associated with the HUD mechanic while aiming to keep subtitles within the viewer’s visual field. When a subtitle appears it is placed directly in front of the viewer and 15° below the horizon. It remains in place within the environment until the viewer’s head rotates beyond a 30° threshold. The subtitle

⁷By *horizon*, we mean the horizontal plane of the real world, by *eye-level*, we mean the plane intersecting the viewer’s eyes and the centre of the headset display.

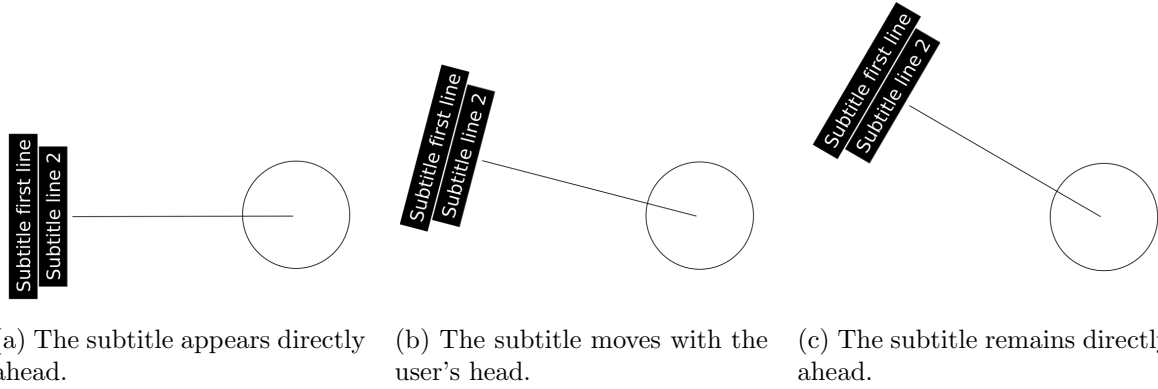


Figure 3: Follow head immediately: As the user turns their head the subtitles stay fixed to the centre of their field of view.

then rotates smoothly (animated over 0.8s with slight easing) to a new position in front of the viewer. For this behaviour subtitles can only travel along the horizontal axis, always remaining at 15° below the horizon — rotating the head to vertical extremes causes the subtitles to be cut from view. The detached nature of this behaviour aims to provide more freedom to move without immediate reaction from subtitles.

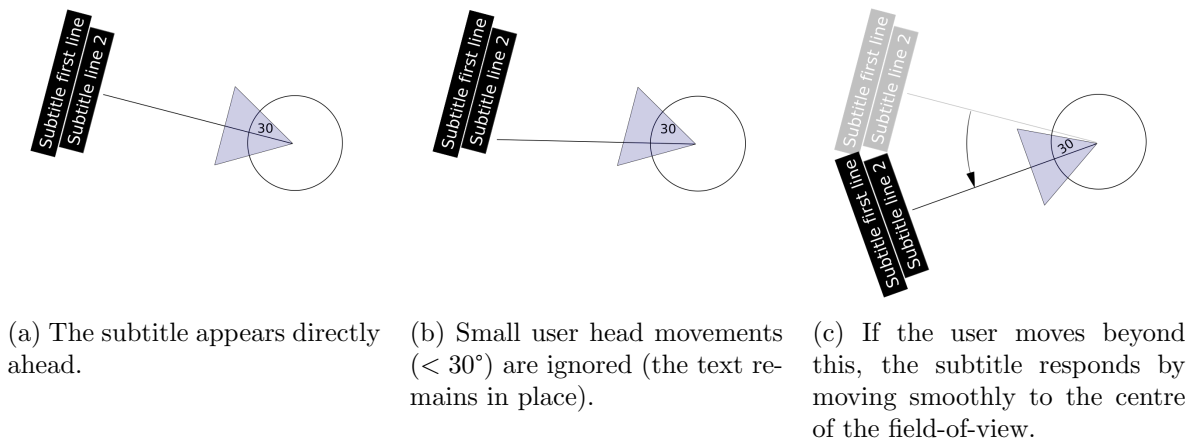
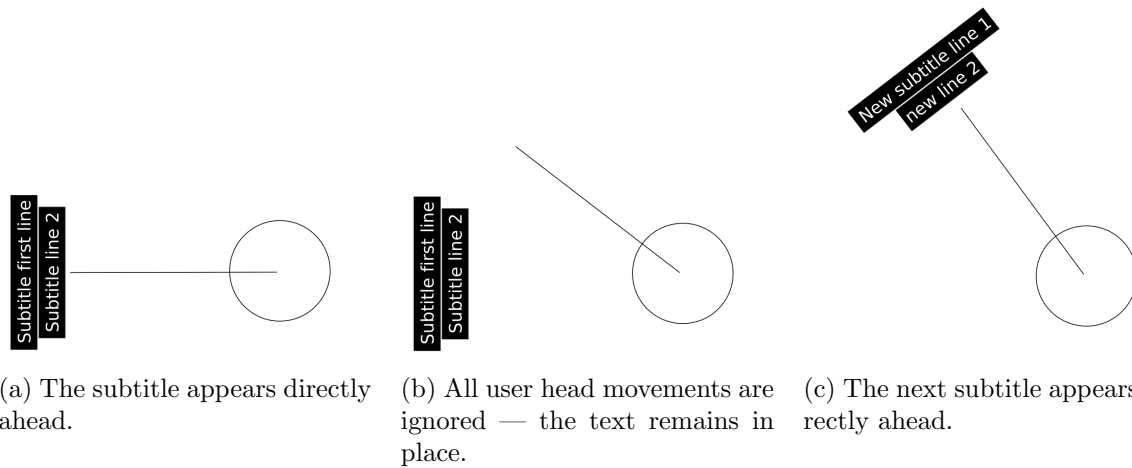


Figure 4: Follow with lag: Subtitles appear in front of the user, but only follow for larger head movements.

3.4 D: Appear in front, then fixed

This behaviour was derived after demonstrating a prototype implementation of *Follow with lag* to a user who was hard of hearing. The user stated that once they had finished reading a subtitle they had a desire to dismiss it, i.e., prevent it from following them around the scene until next subtitle was ready to be displayed. They reasoned that their habitual behaviour lead them to repeatedly re-read the text of subtitles that followed them, so an ability to look away from the subtitle would enable them to explore more freely.

From this feedback we designed the ‘*Appear in front, then fixed*’ behaviour. As shown in Figure 5, each subtitle is placed at the centre of the viewer’s view but 15° below eye-level, in the same way as *Follow head immediately*. As the viewer moves their head, however, the subtitles remain static within the environment (and do not follow the viewer).



(a) The subtitle appears directly ahead.

(b) All user head movements are ignored — the text remains in place.

(c) The next subtitle appears directly ahead.

Figure 5: Appear in front, then fixed: Subtitles appear directly in front of the user, but then remain in place.

3.5 Expectations

From the iterative design process, we had the following expectations about the subtitle behaviours:

A: Evenly spaced offers the simplest solution and does not require interactive mechanisms — the subtitles can be burnt in to the video. The viewer should be able to predict the location of subtitles and always have them in view. On the other hand, there is potential to obscure important content if located in an unfortunate position. This behaviour could work well with all types of content, unless points of interest are located vertically high up and in between subtitles.

B: Follow head immediately uses an established mechanism (HUD) and is very easy to understand. The position, always in front of the viewer, makes the subtitles very easy to locate. This relatively central position might obscure important video elements, but the follow nature means that these can be seen following a small head movement. Anecdotally, having objects fixed to the view can be irritating and prolonged exposure has the potential to cause nausea.

C: Follow with lag toned down reactivity in comparison to *Static-Follow* so it could feel less intrusive when used. Again viewers could, if they wished, explicitly move subtitles and know they will stay in the last place so that they don't obscure parts of the video. It's clear that the responsiveness of this behaviour could require tweaking, the angle of head movement threshold has the potential to be too great or too small. Moreover, the animation duration may need to be carefully considered, especially as viewers will be attempting to read subtitles that animate (move) across the environment. Fast animations may cause motion blur rendering the subtitles unreadable.

D: Appear in front, then fixed was designed to allow subtitles to be read quickly then ignored, and to make it very obvious when a new subtitle appears. The 'fixed within the scene' nature, however, means that if the subtitle covers an item of interest, the viewer will not be able to view that item until the subtitle has gone. Large rotations could cause subtitles to become more difficult to relocate, and if the subtitle appears while the viewer is quickly exploring the scene, they will have to pause exploration to read. As this behaviour's design was driven by feedback from a test viewer it will be interesting to see if it meets their requirements, and if the behaviour is appreciated by other viewers (especially those that are similarly hard of hearing).

4 Method

In order to test these behaviours against the UX framework, we employed a within-subjects design, where each participant saw 4 different video clips, each with one of the subtitle behaviours, and was given the chance to rate and talk about their experiences in a semi-structured interview. This section describes the experimental procedure in detail.

4.1 Stimuli

Six video clips were selected to cover a range of scenarios. In all cases the camera remained stationary.

- 1: Chinatown** A single presenter talks to camera in two shots. The first is in the street, the second inside a shop (1:35 duration).
- 2: Fire Rescue** A single character talks to camera in a fire station, then narrates over some outdoor footage (1:09 duration).
- 3: No Small Talk** Two characters are seated in a cafe having a discussion. The camera is fixed across the table from them, while a book on the table in front of the camera has drawings appear (1:49 duration).
- 4: Argument** Two characters have an argument in an outdoor space, ending in a stage fight. As they argue, the characters move around the camera so that, at times, they are on opposite sides (1:06 duration).
- 5: Planet Earth — Arizona** A narrator describes the scene across several shots outside, then inside a sandstone canyon in Arizona (2:37 duration).
- 6: Planet Earth — Galapagos** A narrator talks about filming on the Galapagos Islands, then introduces the director (on screen), who talks more about the film crew (also visible in the scene) and their roles (1:26 duration).

The six clips and four conditions were counter-balanced using 6 Grecco-Latin squares, according to Table 1. Each participant saw four different clips, each with a different subtitle behaviour. The clip and behaviour combinations were balanced across participants, as was the behaviour ordering. Each clip/behaviour combination was shown 4 times.

4.2 Procedure

Each testing session was run with a single participant, taking place in the User Testing Laboratory in Dock House, Salford. Two facilitators were present, one leading the interview, while the other ran the hardware and software and managed cables. The participant was welcomed and informed about the nature of the study and the procedure for the session. Having signed the consent form, the process was as follows:

1. The participant was seated in a swivelling office chair (reducing the risk of accident and virtual reality sickness [26]) and introduced to the Oculus Rift. A 2 minute clip of an Edinburgh street scene during the Edinburgh International Festival was shown on the headset while the participant adjusted the straps so that it was comfortable. Headphones were given to the participant once the headset was correctly adjusted, and the volume was set to a comfortable level. The option to listen using speakers was also available; this acclimatisation clip contained ambient sound but no speech.
2. The four test clips were shown to the participant:

Participant	1	2	3	4
P01	A1	B2	C3	D4
P02	D2	C1	B4	A3
P03	A4	B5	C6	D1
P04	D5	C4	B1	A6
P05	B6	A1	D4	C5
P06	B3	A4	D1	C2
P07	C4	D3	A2	B1
P08	A5	B6	C1	D2
P09	D6	C5	B2	A1
P10	C1	D6	A5	B4
P11	B1	A2	D5	C6
P12	A2	B3	C4	D5
P13	C2	D1	A6	B5
P14	A3	B4	C5	D6
P15	D4	C3	B6	A5
P16	D3	C2	B5	A4
P17	B5	A6	D3	C4
P18	B4	A5	D2	C3
P19	C5	D4	A3	B2
P20	A6	B1	C2	D3
P21	D1	C6	B3	A2
P22	C6	D5	A4	B3
P23	B2	A3	D6	C1
P24	C3	D2	A1	B6

Table 1: Stimulus matrix for participants, balancing the order of subtitle behaviours (A-D) and their combination with video clips (1-6).

- (a) The behaviour of the subtitles in the clip was explained, and the participant was told what the content of the video was going to be.
 - (b) The participant viewed the clip.
 - (c) The headset and headphones were removed and the participant was asked a number of questions about the experience. Some of these required them to provide a rating on a 7-point Likert scale. (See Appendix A).
 - (d) Once complete, the participant moved on to the next clip
3. The participant was asked to answer some questions about the four behaviours, including which they preferred, and what advantages/disadvantages they could see for each. (See Appendix B).
 4. The participant was given the opportunity to make any additional comments or ask any questions, then given their incentive (£40).

4.3 Likert Ratings

The semi-structured interview was conducted around the themes derived from the UX framework for subtitles [14]. As part of this, the participant was asked for their general opinion about the topic, then asked to mark on a 7-point Likert scale how much they agreed or disagreed with the following statements.

- I enjoyed the clip

- I found it easy to locate the subtitles
- I found it easy to read the subtitles
- I am confident that I followed all the speech
- I found it easy to follow the visual action
- I feel that I understand what was happening in the footage
- I felt immersed in the scene, like I was there

A 7-point scale from *Strongly Agree* to *Strongly Disagree* was used. Participants were asked to complete the scales at points during the interview, each one following a question designed to get a more discursive answer on the theme.

4.4 Participants

24 participants were recruited by an external agency. Participants were requested to cover a range of ages and a mix of genders, to regularly use subtitles while watching television at home, and to be regular users of a smartphone, tablet or computer. There was no requirement for participants to have viewed 360° video or used a headset. In addition, we informed participants that guidelines suggest that headset use is not recommended for anyone who:

- is or may be, pregnant
- has had a seizure
- has an epileptic condition
- suffers from psychiatric disorders
- suffers from a heart condition
- has pre-existing binocular vision abnormalities
- has had a loss of awareness
- suffers from a serious medical condition
- wears a pacemaker or other implanted medical device
- suffers from dizziness and/or sickness
- has a cold/flu at the moment
- has a stress or an anxiety related condition

The final set of participants comprised 11 male and 13 female, with 13 aged between 18 and 40, and 11 between 41 and 70.

5 Results

In this section, we first present the overall feedback on which behaviours were preferred, then look at the results of the Likert scale ratings given by participants during the interviews. We then perform a thematic analysis of the interviews, exploring the themes that emerged from the interviews as well as analysing the comments relating to how the behaviours were implemented. Following this, we revisit the UX Framework for Subtitles, examining how the emergent themes map to its 7 dimensions and considering how the different subtitle behaviours performed on those aspects of the framework that didn't emerge in the thematic analysis. Finally we consider the comments that applied to all behaviours, including the visual style of the subtitles.

NVivo software was used for thematic coding and analysis, while R⁸ was used for statistical analysis.

5.1 Overall Preferences

As part of the wrap-up, participants were asked to compare the behaviours, giving advantages and disadvantages for each, and to select one behaviour that they would prefer to use. Figure 6 summarises the selection, while Table 2 gives precise numbers. Although participants were not explicitly asked which behaviour they liked the least, many of them (13 of 24) made this clear; Figure 6 and Table 2 also include these data.

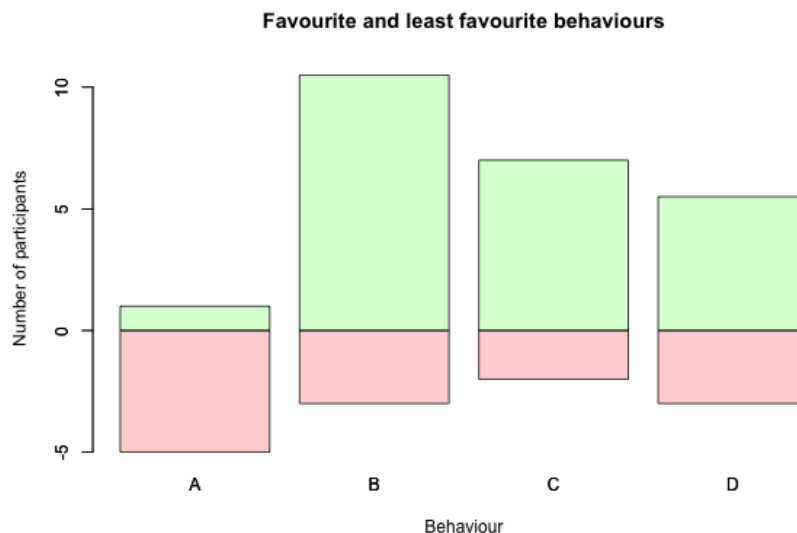


Figure 6: Bar chart showing how many participants selected each behaviour as their favourite or least favourite. Values above the y-axis (in green) show the number who selected the behaviour as favourite; negative values show the number who selected it as their least favourite. Not all participants specified a least favourite.

5.2 Likert Ratings

Table 3 summarises the results of the Likert scales that participants completed after each clip. It reports the median score of the 24 participants, on each scale for each condition (these were balanced across clips). Participants responded to each statement on a seven point scale between *Strongly Agree* (7) and *Strongly Disagree* (1), so a higher score represents a better user experience for the subtitles on that measure. These results are also presented in Figure 7.

⁸<https://www.r-project.org/>

Behaviour	Favourite		Least favourite	
	%	N	%	N
A: <i>Evenly spaced</i>	4	1	38	5
B: <i>Follow head immediately</i>	44	10.5	23	3
C: <i>Follow with lag</i>	29	7	15	2
D: <i>Appear in front, then fixed</i>	23	5.5	23	3

Table 2: Percentages and absolute numbers of people who selected the each behaviour as their favourite or least favourite behaviour. Least favourite was not specifically requested, so was not available for all participants. One participant selected B and D as joint favourite.

	A	B	C	D
<i>I enjoyed the clip</i>	5	6	6	6
<i>I found it easy to locate the subtitles</i>	5	7	6.75	6
<i>I found it easy to read the subtitles</i>	6	7	7	6
<i>I am confident that I followed all the speech</i>	6	7	7	6.5
<i>I found it easy to follow the visual action</i>	6	7	6	7
<i>I feel that I understand what was happening in the footage</i>	6.25	7	6	7
<i>I felt immersed in the scene, like I was there</i>	5	6	6.5	6

Table 3: Median Likert scale scores for each behaviour. 7-point scales were used, from strongly agree (7) to strongly disagree (1).

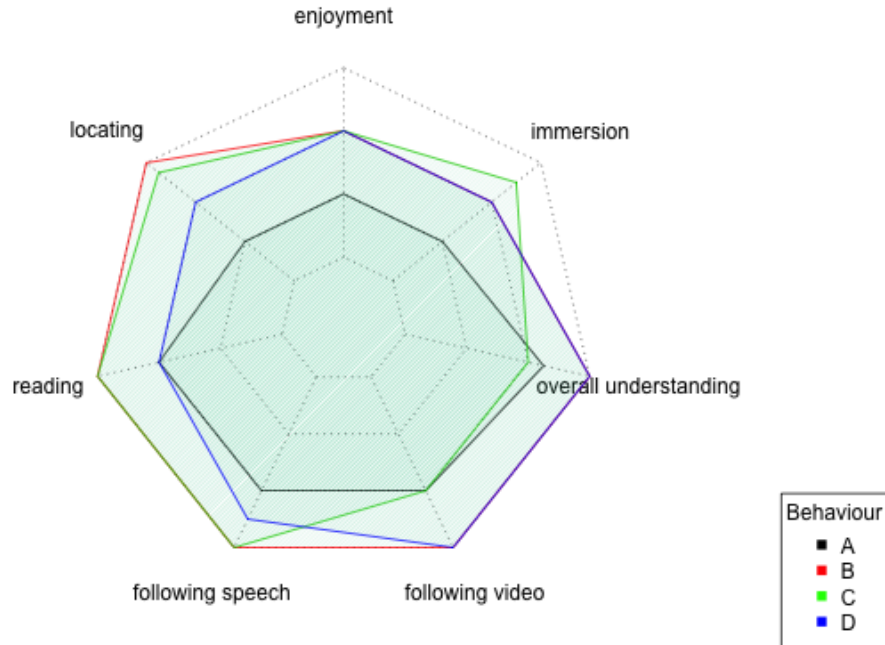


Figure 7: Radar plot showing how each behaviour compared across the Likert scales. Higher values reflect better performance. Note that the scale runs from 4 – 7.

In order to compare the overall UX of the behaviours, the sum of the scores for each clip viewed has been calculated, to give a quasi-continuous measure between 7 and 49 (with a high score indicating better overall UX). Cronbach’s alpha was calculated over the 7 scales as 0.82, indicating high internal consistency between the items. Each condition was viewed by each of the 24 participants; the mean scores are given in Table 4 and presented graphically in Figure 8.

Behaviour	mean	median	sd
A: <i>Evenly spaced</i>	36.7	37.5	8.02
B: <i>Follow head immediately</i>	43.8	44.0	4.29
C: <i>Follow with lag</i>	41.9	43.0	5.41
D: <i>Appear in front, then fixed</i>	39.4	42.5	7.95

Table 4: Means, medians and standard deviations for the sums of Likert scale scores for each behaviour. $N = 24$ in each case. Higher scores indicate better subtitle performance.

Comparing these pairwise with Student’s T-test shows that there are statistically significant differences in the performances: *Follow head immediately* performs better than *Evenly spaced* ($p = 0.0006$) and *Appear in front, then fixed* ($p = 0.025$), and *Follow with lag* performs better than *Evenly spaced* ($p = 0.014$).

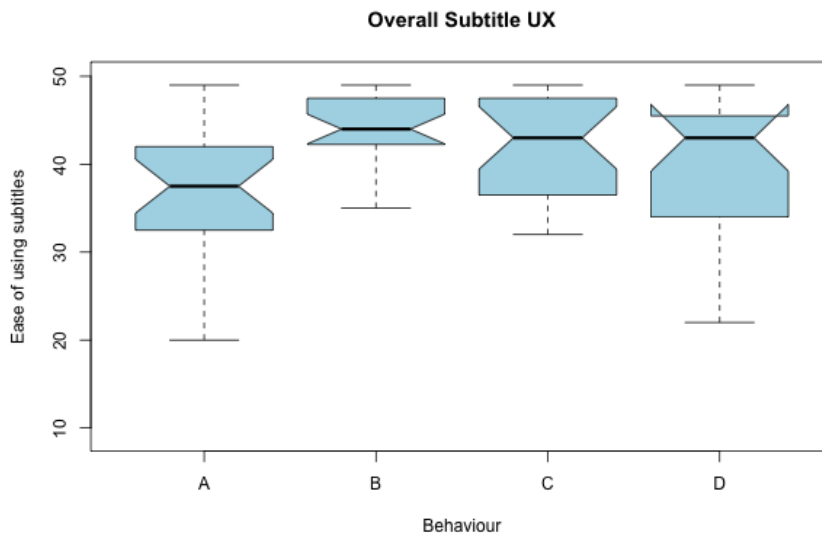


Figure 8: Box plot showing the distribution of sums of Likert scale results, representing the overall user experience of the subtitle behaviours; higher values indicate better performance.

Two-level ANOVA shows a significant effect for behaviour on overall subtitle performance, but no effect for clip, nor any significant interaction between clip and behaviour (Table 4). Looking at the variation of the overall score with clip (Figure 9), it can be seen that there was little effect. Clip 1 (Chinatown) had the highest values, indicating that it may have been the easiest clip to follow, but the differences from the other clips were not statistically significant (using Student’s t-test at $p < 0.05$).

Comparing the results on the individual scales, only that measuring how easy it was to **locate** the subtitles revealed any statistically significant differences. The conditions were compared by calculating the number of times each level on the Likert scale was chosen, and comparing the distributions using Fisher’s exact test. *Follow head immediately* was found to have better scores than *Evenly spaced* ($p = 0.0001$), *Follow with lag* ($p = 0.032$) and *Appear in front, then fixed* ($p = 0.041$). Figure 10 illustrates the results for this scale as a box plot. The ‘overall enjoyment’

	Df	Sum Sq	Mean Sq	F value	p (> F)
Behaviour	3	681	227.04	4.812	0.00379 **
Clip	1	9	6.24	0.132	0.717
Behaviour:Clip	3	90	29.93	0.634	0.595
Residuals	87	4104	47.18		

Table 5: Two way ANOVA exploring the effect of Behaviour and Clip on overall subtitle performance.

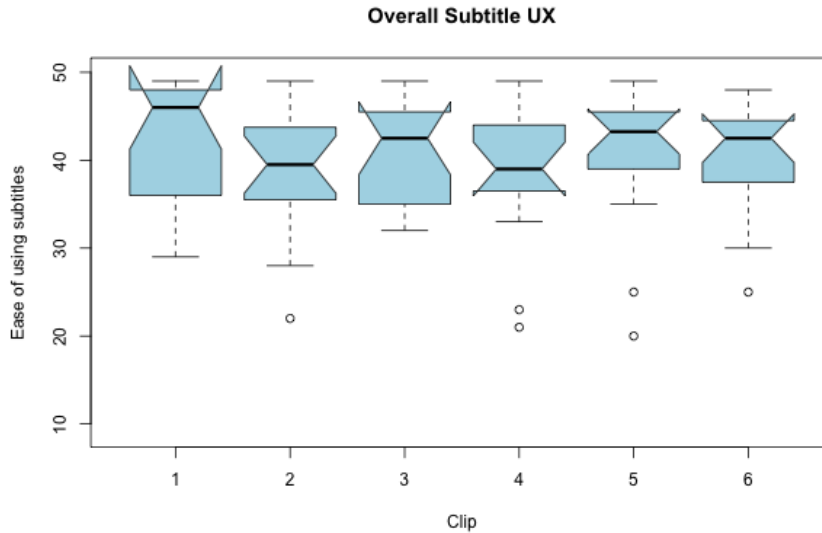


Figure 9: Box plot showing the variation in overall difficulty across clips; higher values indicate greater difficulty.

scale was also suggestive that *Evenly spaced* performed poorly (Figure 11), although this was not statistically significant (Fisher’s Exact test comparing A with B, $p = 0.088$).

5.3 Thematic Analysis

The thematic analysis methodology of Braun and Clarke [27] was used for the analysis of the data collected during the study. While the interviews were structured around questions that were defined by the UX Framework for Subtitles [29, 14], coding of the transcripts was performed inductively, based on participants’ comments on what they thought of the subtitles, what they found easy and what they found difficult. This coding was iteratively refined to identify a set of themes which encapsulate the ways in which the subtitle behaviours affected the users. The following were identified as *differentiating themes*, i.e., themes in which there were differences between the subtitle behaviours (see also Figure 12). These are expressed as difficulties, once these themes had been identified, individual extracts were sorted into positive and negative comments. Prevalance was counted at the level of the number of participants.

Effort: Effort was required to locate or use the subtitles, or to follow the clip.

Missing Out: The viewer felt that they were either missing subtitles or unable to follow the speech, or were restricted in their ability to explore the full 360° scene.

Obstruction: The subtitles were blocking the viewer from seeing the video.

Distraction: The subtitle position or behaviour was distracting, attracting attention in an undesirable way.

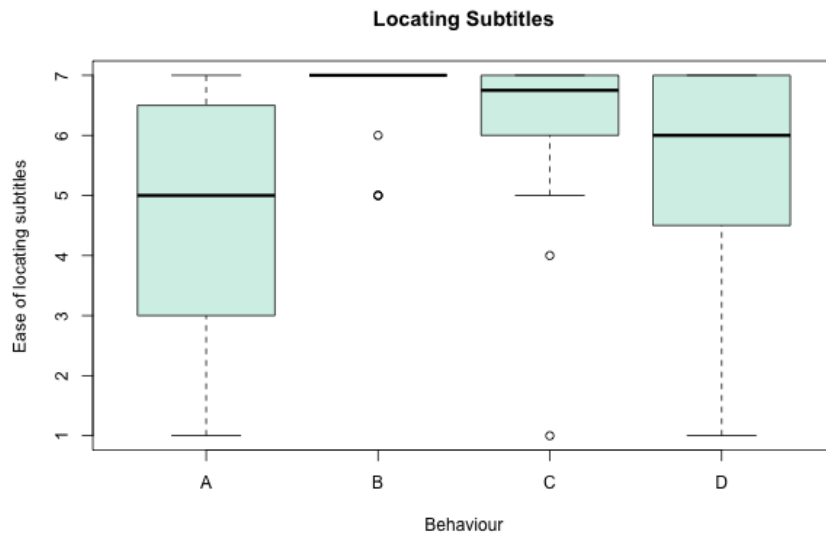


Figure 10: Box plot of the Likert scale results for locating subtitles; higher values indicate that participants found it easier to locate the subtitles. Fisher’s exact test was applied to these results, revealing that behaviour B (Follow head immediately) was significantly better ($p < 0.05$) than each of the other behaviours.

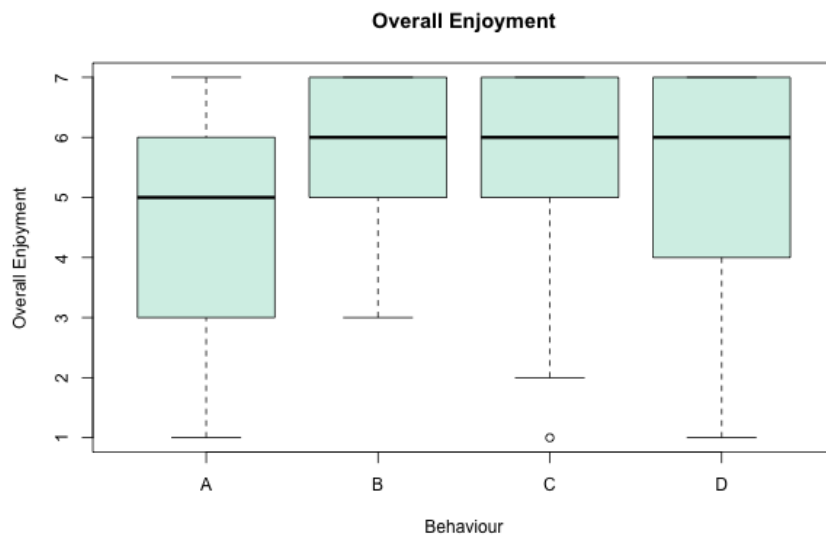


Figure 11: Box plot of the Likert scale results for overall enjoyment of the clip. Higher values indicate **lower** enjoyment. The differences were not statistically significant to the 95% level.

Immersion: The feeling that the subtitle behaviour impacted on how immersed the viewer felt in the scene.

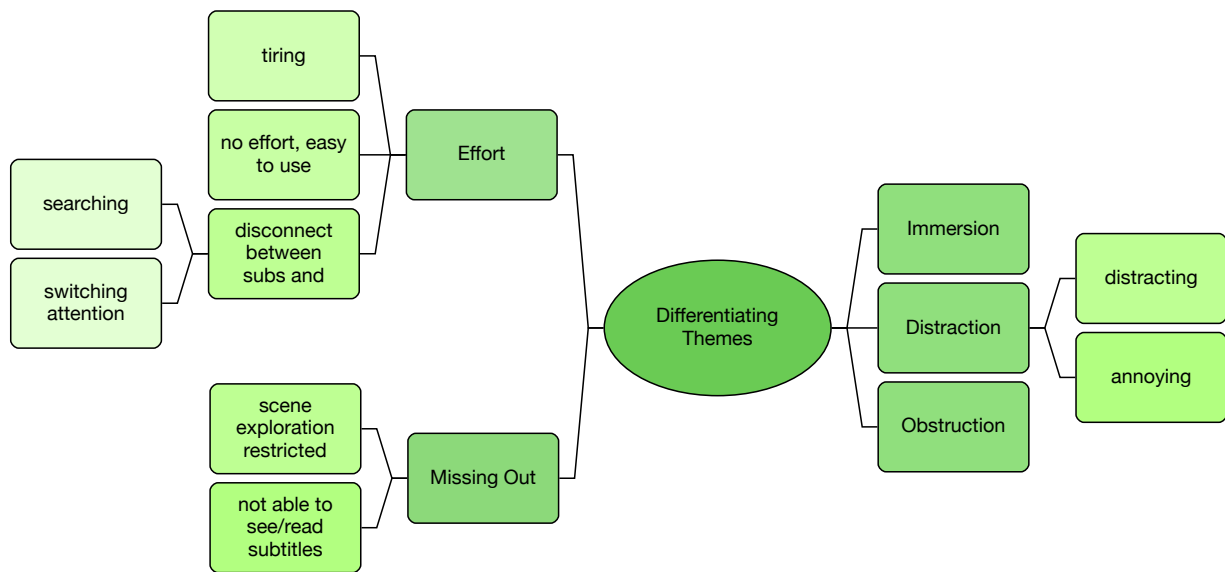


Figure 12: Mindmap representing the differentiating themes identified in the interview data.

In addition, there were some themes in participants’ comments where there was broad consistency across the behaviours. These included:

Speaker Identification: The ability to identify which person in the scene the subtitle is associated with.

Vertical Position: The vertical of the subtitle blocks in the scene/relative to the viewer’s head orientation.

Aesthetics: The visual appearance of the subtitles — size, font, background, etc.

In this section, we summarise the differentiating themes; the common themes are discussed in §5.5.

5.3.1 Effort

One of the key discriminatory themes to be identified was that of *effort*. Participants talked about the effort required, concentration, or using the subtitles being tiring. However, this was not at all consistent across the behaviours, with behaviours A (*Evenly spaced*) and D (*Appear in front, then fixed*) getting more mentions. Participants often ascribed the effort to having to *search* for subtitles in these behaviours:

‘You had to work to find the subtitles, so no good.’ [P10, A5]

‘Whereas this one, I was finding it, then reading it and then watching what they were doing, and then finding the next one. So its like, required effort.’ [P19, D4 (comparing to C5)]

Very few comments were made in this respect about behaviours B and C; the single comment about effort for *Follow head immediately* was P02, who commented that, even with the subtitles following his eye-gaze: ‘if there was 3 or 4 people, a deaf person would find it hard to concentrate

on finding the right person' [P02, B4] (this clip had the two characters circling the camera while having an argument).

Table 6 shows the number of participants who made comments about using the subtitles requiring effort, or about having to search for the subtitles, for each behaviour; Figure 13 represents these graphically.

	A	B	C	D
<i>effort</i>	13 (17)	1 (1)	2 (2)	4 (9)
<i>searching</i>	12 (26)	-	-	5 (8)

Table 6: Number of participants making comments about the effort required to use the subtitles and about having to search for the subtitles. The number of comments is given in parentheses.

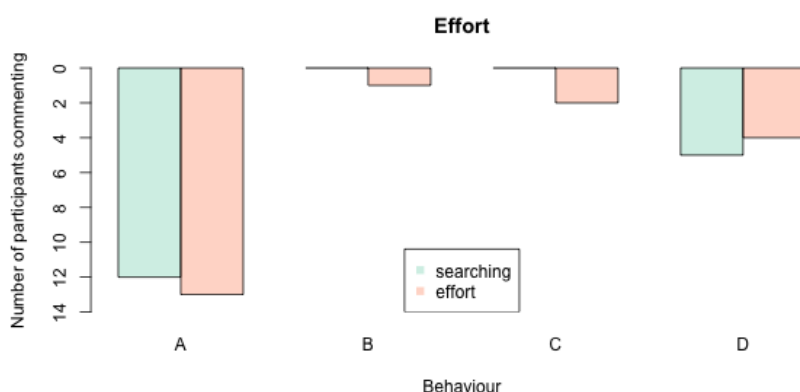


Figure 13: Bar chart showing how many participants commented on having to make an effort or having to search for subtitles.

5.3.2 Missing Out

Closely related to the theme of *searching* is that of *missing out*. Participants were asked how easy they found it to follow the speech, subtitles and video comment, and this often led to them saying that they felt they had missed out, either on the speech or video content.

'I think I kept on top of it. I could have definitely missed bits' [Was that related to subtitles?] 'Yeah exactly. Because they were off-screen' [P15, A5]

'Again because of the position of the subtitles it wasn't easy to see when she was picking a piece of fruit up' [P11, B1]

'If there was something that caught my eye over here and the subtitles were still here, I'd have to look at the subtitles that were there and then miss what was going on.' [P03, C6]

'Difficult to follow really.' [Why was that?] 'It was... where are you looking...? Where do you go with the action? They were both separated. The words were there and the other two people weren't there. I was missing the action. I kind of lost the whole interaction' [P01, D4]

Table 7 gives an indication of the relative problems caused in the different behaviours, giving the number of participants who mentioned missing out in each case.

	A	B	C	D
<i>missing video</i>	12 (22)	4 (4)	6 (8)	9 (12)
<i>missing speech/subtitles</i>	11 (19)	1 (1)	6 (9)	6 (7)
<i>either</i>	15	5	7	11

Table 7: Number of participants making comments about missing out on either subtitles or video. The number of comments is given in parentheses. The ‘*either*’ row gives the number of participants who commented on missing out on one or both of subtitles or video.

Broadly speaking, there were distinct reasons for missing out across the behaviours. For B (*Follow head immediately*) the typical reason for missing out was that the subtitles obstructed their view of the video; for A (*Evenly spaced*) and D (*Appear in front, then fixed*) the main problem was that the subtitle was not positioned so that both it and the region of video of interest were visible at once. For behaviour C (*Follow with lag*) participants experienced a mixture of these problems (although neither to such a great extent) — the lag meant that the position was sometimes too far away from the region of interest and sometimes obstructing it.



Figure 14: Bar chart showing how many participants commented on missing out on either the video or the speech/subtitles.

Another aspect of missing out was the feeling of the subtitles **restricting** where one could look; of not having sufficient **freedom** to look around:

‘you’re very conscious that you don’t want to miss the next subtitle, so you can’t just explore all around’ [P18, A5]

‘I felt like I was moving slower because the subtitles weren’t following as quick. So just because of that, I would feel more free to move quickly if the subtitles came quickly where my eyes were going. But apart from that it was fine.’ [P19, C5]

‘Because I’m having to follow the subtitles, therefore I’m not able to look around and observe what else is going on, because I’m waiting or looking for my subtitle.’ [P24, D2]

On the other hand, there were also many comments about how participants felt free to explore:

‘The fact that they were at the 3 points, I knew where they were, so I could move and know that I’m going to have that information there in front of me. So I felt that over the last one, I could enjoy more and see more and understand’ [P24, A1 (comparing to D2)]

‘100% could look around. He was referring under the rocks, you’ve got them snakes that come out, so I had a little look over there. Again the information followed me and it was all clear for me.’ [P24 B6]

‘I liked the idea of it doing that because it gives you a chance to look about and then the subtitles are still there. That’s good - I like that.’ [P14, C5]

‘You can move away from the talk but you’re not missing anything.’ [P06, D1]

	A	B	C	D
<i>felt free</i>	3 (7)	9 (14)	8 (12)	6 (9)
<i>felt restricted</i>	6 (12)	-	2 (2)	5 (6)

Table 8: Number of participants making comments about the feeling restricted in their viewing or feeling free to explore. The number of comments is given in parentheses.

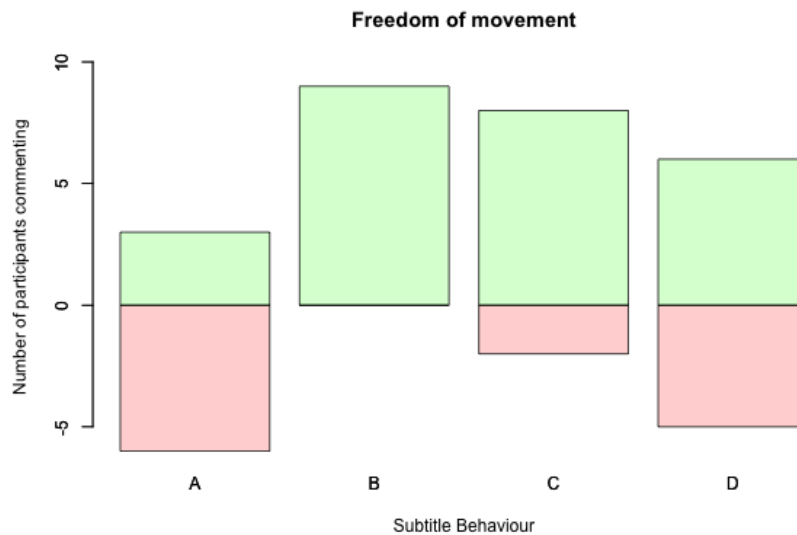


Figure 15: Bar chart showing how many participants commented on being free or restricted in their ability to look around the scene, for each behaviour. Values above the x-axis (in green) show the number who commented on feeling free; negative values (in red) show the number who commented on feeling restricted.

5.3.3 Obstruction

As with television subtitles, participants commented on their dislike of the subtitle text obstructing the video content. This was most significant in *Follow head immediately* (see Table 9), although participants were sometimes able to work around it:

‘They seemed to be too central and they were sort of blocking what you were looking at. The subtitles were taking over the entire screen and it was the focus. Normally when I’ve got subtitles at home they’re at the bottom of the screen and they’re not intrusive, but they seem to be a little bit intrusive in the middle of the screen.’ [P11, B1]

‘But they sometimes obscured some of the other action I might be looking at, but then I could adjust to that, so I could move my head a little bit lower down’ [P15, B6]

Obstruction was commented on less often for behaviours A and D (*Evenly spaced* and *Appear in front, then fixed*), as the subtitle did not move to remain in front of the viewer. However, when the subtitle did obstruct the background, there was no way to see what was behind.

‘They were quite, more centralised than the subtitles that I’m used to, because usually subtitles that I use, when I watch a film, are quite low on the screen, but these were quite centralised. So they were arguably sort of obstructing the view a little bit?’ [P20, A6]

The problem was generally associated with the vertical position of the subtitles, with the majority of participants stating that they would have preferred them slightly lower (see §5.5.2, p25).

	A	B	C	D
<i>subtitles obstructing</i>	4 (4)	8 (20)	3 (8)	5 (8)

Table 9: Number of participants making comments about the subtitles obstructing their view of the video; the number of comments is given in parentheses.



Figure 16: Bar chart showing how many participants commented on the subtitles obstructing their view.

5.3.4 Distraction

There were a number of comments about how the subtitles could distract from the primary content. Four factors emerged as the main causes of distraction:

Appearance: The look of the subtitle blocks was considered quite ‘strong’, particularly the white text on black background (see §5.5.3):

‘The subtitles had the black background and they were really strong. And they were always in front of you’ [P04, B1]

‘Because you know the whole black background sort of thing, it’s very prominent. It’s right there, so you can’t help but look at that rather than look around. So it’s very visible and very clear but kind of detracts from the actual video’ [P12, A2]

Searching: Having to search for the subtitles (see §5.3.1) was a distraction, particularly for behaviours A (*Evenly spaced*) and D (*Appear in front, then fixed*):

‘I certainly did not like the ones that were at 120 degrees away, and you had to keep turning your head to find them - that was very distracting for me.’ [P22]

‘I think I was more distracted by trying to find the subtitles than I was to following what was happening - I couldn’t remember if it was terrapins or iguanas.’ [P17, A6]

Movement: Motion of the subtitles was distracting, primarily in behaviour C (*Follow with lag*):

‘I didn’t like them moving around so much. They were taking my attention moving around so much’ [P09, B2]

‘I found them a bit annoying and distracting when I turned and then they just come by. Where as if they’re there with you all the time it’s easier to read them.’ [P03, C6]

‘It was quite distracting because there was something moving other than what was going on in the clip.’ [P05, C5]

Unpredictability: Participants were distracted when events occurred when they weren’t expecting. This was particularly notable in behaviour C (*Follow with lag*), when the lag before the subtitle moving back into the centre meant that participants weren’t sure exactly when it was going to move:

‘... but if you moved a lot they sort of jumped into the shot. It was kind of a little bit distracting, cause it was “ooh, will they, won’t they?” kind of thing’ [P20, C2]

	A	B	C	D
<i>distracted</i>	9 (10)	4 (4)	8 (15)	5 (7)

Table 10: Number of participants making comments about being distracted by the subtitles. The number of comments is given in parentheses.

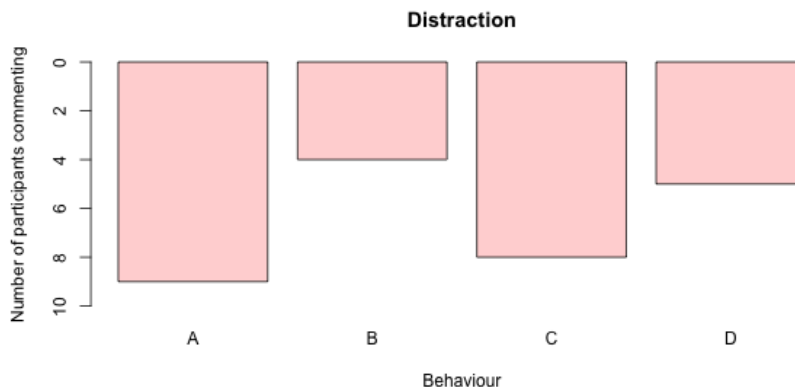


Figure 17: Bar chart showing how many participants commented on being distracted by the subtitles.

5.3.5 Immersion

There was a general consensus that the presence of subtitles had a small negative impact on the sense of immersion, although it was recognised that this was an inherent problem with presenting text over the video.

‘I think it kind of detracts from the immersion feeling. So, because you know when you’re watching movies, the subtitles are there and it kind of reminds you that it’s not reality. And that’s what I kind of attach that to. So the subtitles are there so I know it’s not I’m actually there in that scene’ [P12, A2]

The Likert scale ratings varied slightly across the four behaviours, (see Table 3) and although the differences were not statistically significant, they were matched by the qualitative feedback. Both the appearance of the subtitles and their *behaviour* also had an impact on the level of immersion experienced. The main factor in the latter was related to *effort*: participants commented that any confusion or difficulty using the subtitles ‘took them out of the experience’. For example:

‘You should feel immersed but I wasn’t. It was too awkward to follow them around.’
[P04, C4]

As might be expected from the comments on *effort* (§5.3.1), behaviour A (*Evenly spaced*) had more negative comments about immersion than the other behaviours in this respect. Other factors had an impact, with distracting or obstructing subtitles also reducing the sense of immersion. Overall, however, the visual appearance of the subtitles was the predominant problem cited by participants (see §5.5.3).

5.3.6 Overall Comparison

For an overall sense of the performance of each implementation, the number of participants that commented on having had a problem in each of the themes above for each behaviour can be used as a rough indicator. Figure 18 plots these values for each behaviour for each theme as a radar plot. In this plot, lower values represent better performance, with a low overall area indicating better performance across the spectrum. Note, however, that this overview does not accurately reflect the relationships between themes, nor their relative importance.

5.4 UX Framework Dimensions

The semi-structured interviews were structured around the UX framework for subtitles [29, 14], with the questions covering the 7 dimensions of that framework. From these interviews 5 main discriminatory themes were identified (above). Table 11 gives the definitions of the framework dimensions, and indicates how the discriminatory themes map against these.

Table 11 shows that *Aesthetics*, *Usefulness* and *Familiarity* did not emerge as clearly distinct themes. *Aesthetics* was a topic that participants spoke at length about, although there were no strong differences between subtitle behaviours on this dimension; this theme is discussed in more detail in §5.5.3. In terms of *Usefulness*, there was perhaps a sense that subtitles in 360° video had lower usefulness than on television as it is more of an individual experience that is often consumed using headphones⁹; the different behaviours tested did not appear to have any effects on this dimension, so it is not considered further. Similarly, *Familiarity* emerged neither as a theme in itself, nor as an obvious result of any of the six discriminatory themes; the responses of participants to the questions about *Familiarity*, and how these varied by behaviour, are summarised below. In addition, while the discriminatory themes of *effort* and *distraction* can be mapped as factors that affect the dimension of *Endurability*, we summarise below participants’ responses to this as a specific question.

⁹Two participants (P09, P18) had hearing aids that meant they were unable to use headphones; they used speakers.

Dimension	Definition	Relevant discriminatory themes
Aesthetics	A measure of the visual appeal of the subtitled content. High levels indicate users believe that the content is visually pleasing, while low levels indicate that the content is not visually appealing.	Not apparent in any of the discriminatory themes, although <i>obstruction</i> could be relevant here — some participants complained about subtitles ‘hiding the beautiful scenery’.
Involvement	A measure of how engaged users are with subtitled content. Whereas attention is about focus on the content, involvement is about the depth of engagement with the subtitled content. Users with high levels of involvement would be ‘drawn into’ the subtitled content and would find this to be an engaging and enjoyable experience. Users with low levels of involvement would feel less involved in the subtitled content.	<i>Immersion, distraction, missing out.</i>
Attention	A measure of the awareness to what is going on in relation to the subtitled video content. Users with high levels of attention would be focused heavily on the video content, while users with low levels would not.	<i>Missing out, freedom of movement, distraction.</i>
Familiarity	A measure of how useful the display of the subtitled content is. Users who perceive high levels of usefulness will see a high level of value in the subtitle display. Users with low levels of perceived usefulness will see low levels of value in the subtitle display.	Did not emerge as a strongly discriminatory theme.
Usefulness	A measure of how useful the display of the subtitled content is. Users who perceive high levels of usefulness will see a high level of value in the subtitle display. Users with low levels of perceived usefulness will see low levels of value in the subtitle display.	Not apparent in discriminating themes.
Usability	A measure of the challenge that is faced while engaging with the subtitled video content. Users that report high levels of perceived usability are likely to have found the subtitled content easy to understand, while users with low levels of perceived usability are likely to have found viewing the subtitled content more demanding.	Locating and reading subtitles: <i>effort.</i>
Endurability	A user’s willingness to view subtitled video content using a similar method of subtitle display in the future. Users with high levels of endurability are likely to wish to use this method again, while users with low levels of endurability would be less likely to want to use this method again in the future.	<i>Effort, distraction.</i>

Table 11: The UX Framework for Subtitles and how the discriminatory themes map to its dimensions.

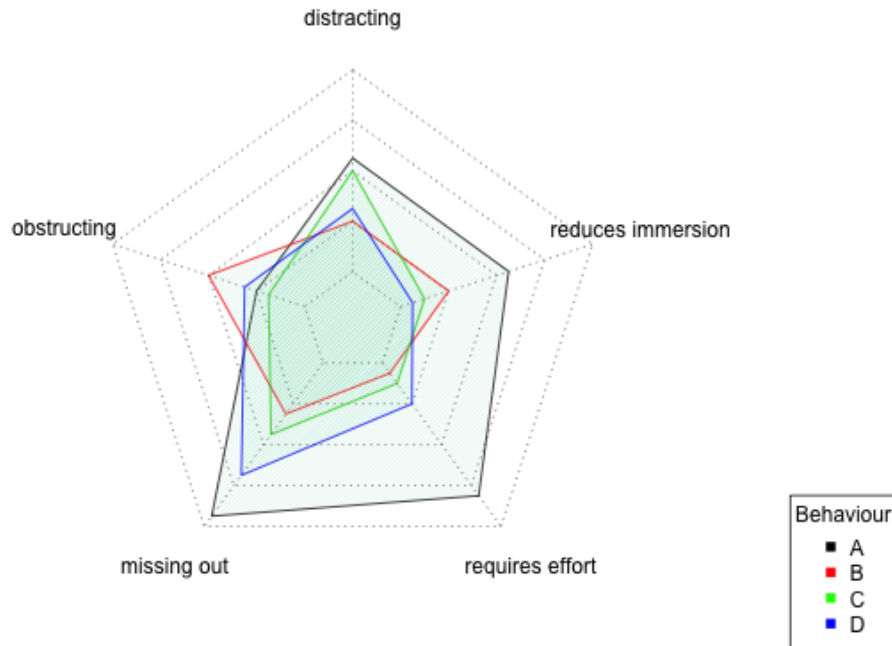


Figure 18: Radar plot showing how each behaviour compared across the themes. The distance from the centre reflects the proportion of participants who commented on the subtitles causing problems relevant to that theme (the grid lines are placed for every 4 participants). **Lower** values reflect better performance.

5.4.1 Familiarity

Although participants were asked how familiar it felt using the different subtitle behaviours, the differences were unclear, as people included both the differences between watching TV and watching 360° video on a headset and the visual style when assessing familiarity. The comments made did not reveal any significant differences between the behaviours, with all 4 being described by some participants as ‘natural’ or ‘like TV’:

‘It’s more likely what I’m used to really’ [P09, A1 (last behaviour viewed)]

‘I think it was the most similar to what I’m used to at home. In terms of when you’re in front of a screen you’re not moving your head, cause you’re just there. With this one, even if I did move my head, you’re still always there, so it just felt the most natural.’ [P19, B2]

‘It felt quite natural this time to normal TV’ [P23, C1]

‘I think the presentation seemed quite similar. If that was on TV you always know where they’re going to be. It’s pretty static. They felt the same in terms of temperament as well, in terms of speed and size’ [P15, D4]

Overall, although the differences were small, behaviours A (*Evenly spaced*) and B (*Follow head immediately*) seemed to be the most natural and familiar to participants.

5.4.2 Endurability

Although this wasn't a theme that emerged naturally from participants' comments, they were asked explicitly whether they would be able to watch subtitles with each different behaviour for an extended period. The answers reflected participants' feelings about the effort required to use the subtitles and how distracting or frustrating they found them. Figure 19 summarises the number of positive and negative responses for each behaviour; not all participants gave a clear answer either way. Overall, behaviour A (*Evenly spaced*) was worst, while the other three were broadly similar. Reasons for not feeling able to endure include:

'I think I'd get frustrated that I was losing bits of what was being said' [P08, A5]

'I wouldn't choose this method because of the effort required to locate the subtitles and follow the scene' [P15, A5]; 'I think I'd be exhausted watching it' [P19, D4]

'I don't think I could because it's too annoying. And confusing. And it made me dizzy I am sure.' [P09, B2]

'I just don't think I would be giving my full attention to whatever I was watching. And unlike other subtitles where you become used to what's going on, I feel like you would never get fully used to it, because it would move round all the time, so it would just remind you that it was there.' [P20, C2]

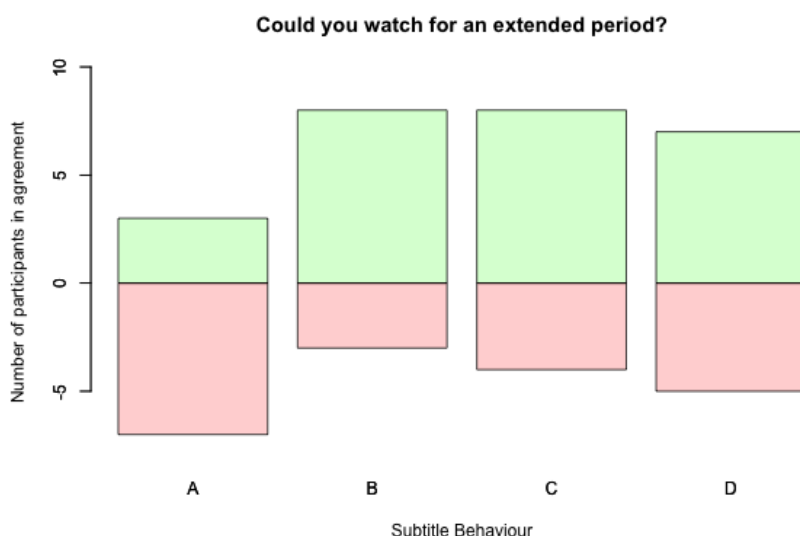


Figure 19: Bar chart showing participants' opinions on whether they could watch content with different subtitle behaviours for longer periods of time. Values above the x-axis (in green) show the number who agreed that they could watch for an extended period; negative values (in red) give the number who disagreed. Not all participants gave an answer.

5.5 Cross-Behaviour Comments

This section presents the themes that emerged that were relevant to all or multiple behaviours.

5.5.1 Speaker Identification

Identifying the speaker was not generally a problem (although none of the clips had more than 2 characters speaking), and the use of multiple colours to differentiate speakers was appreciated.

However, the nature of 360° introduced extra challenges over television. These were down to the resolution making it hard to see who was talking (e.g., ‘*I also find it difficult to see who was talking because it was a bit fuzzy, so I didn’t know who’s mouth was moving*’ [P20, D3]), and the wide field of view (and lack of head-shot close-ups) meant that it could be difficult to associated speaker with colour.

‘I suppose you can determine whether it is a female or man, but if there’s 2 or 3 more people, then you’re not sure who is saying what, because it’s not that clear - so you don’t know who’s mouths are moving (the graphics).’ [P18 B4]

It was also noted that the subtitle behaviours could confuse the viewer. For example, the if the first time the viewer sees a subtitle it is next to the ‘wrong’ person, they could make the wrong association. For example:

‘I think it just happened that the direction I was looking happened to be the woman who wasn’t speaking, so I was like “why is the subtitle there when she’s not the one speaking?”. And then I was like “oh, she’s speaking”, then the subtitle’s in the wrong place.’ [Pilot participant, B3]

Some participants suggested that subtitle position could be used to help identify speakers (see §5.7).

5.5.2 Vertical Position

In *Evenly spaced* and *Follow with lag*, subtitles were positioned at 15° below the horizon, while in *Follow head immediately* and *Appear in front, then fixed* they were placed at 15° below the plane of the viewer’s eyes. There was a broad consensus that this was too high; see Table 12.

code	A	B	C	D
too high	5 (6)	8 (12)	2 (3)	5 (8)
right height	2(3)	2 (2)	-	-
too low	-	1 (3)	-	1 (1)

Table 12: Number of participants making comments about the vertical position of the subtitles. The number of comments is given in parentheses.

Opinions varied about how much lower the subtitles should be, although the majority of comments suggested that they were only slightly too high. For example:

‘Maybe just drop them a tad, a little bit’ [P15, B6]

‘... if it’s placed a little bit lower, then it’s a little more natural to what people are used to at home’ [P19, B2]

Some participants wanted the subtitles to be much lower, however (e.g., ‘*I just seem to prefer them much lower.*’ [P04, A6]), while two participants wanted them a little higher:

‘Locating the subtitles? It was good that — it was central to everywhere you were looking at, but it could be a bit higher’ [P18, B4]

‘Wherever I was looking at they were a little bit too low and I had to sort of move to read them ... I would have liked them probably a little bit more up’ [P11, D5]

Participant P11 thought that there would be individual variation in how high people would want subtitles: ‘*I think there must be an ideal gap with my eyes and the subtitles at the bottom, and I’d like to keep that*’ [P11, C6].

5.5.3 Aesthetics

Font The font used (Roboto) was considered suitable by most participants, although some found it a bit ‘intrusive’, specifically, they found them a bit ‘thick’ (P04, P18, P22), ‘square’ (P04, P13) or ‘unfriendly’ (P04, P13). For example:

‘I’m not sure about this, but I felt that the font was a bit blocky, and rather thick. I think I could have coped with a smaller font. Oh, I don’t know if I mean smaller or thinner? A narrower font somehow. It seemed to me to be quite solid, whereas it could have been... You know, I would have been able to cope with something a little bit less intrusive, in fact.’ [P22]

Text Size The size was also considered roughly right. 12 participants explicitly stated that the size was good, while 3 (P05, P08, P11) wanted larger text. P04 wanted smaller subtitles (*‘No, too big, far too big for me — too big and too obvious looking, although it might just be because of the black background which I really didn’t like.’*).

Background The standard text on black background was accepted by most participants (e.g., *‘the subtitles were very clear, very readable. I’m fine with the black background and white colour.’* [P09]). Although some commented on them being too strong and obvious (particularly P04 and P12), there was a consensus that there was a balance to be made between readability and intrusion.

‘Because you know the whole black background sort of thing, it’s very prominent. It’s right there, so you can’t help but look at that rather than look around. So it’s very visible and very clear but kind of detracts from the actual video’ [P12]

‘if you did just have the text without the background, and the background in one scene was the same colour as the text would be, then you wouldn’t be able to read it, so I feel like, even though it’s not the nicest to look at, the black background is the best, because it’s easiest to read.’ [P20]

There was also the sense, generally related to the background, that the subtitles were ‘unsympathetic’ to the background video, particularly on clips 5 and 6, (e.g., *‘They seemed to look out of place in that scenery’* [P04], *‘aesthetically quite heavy, on quite a lovely moving visual panorama really’* [P09]) and some participants wondered if the subtitles could be better blended in.

Number of Lines Subtitles were presented in one or two lines, depending on their length. While some participants preferred this (P11, P22) and two were happy with it (P18, P20), four participants (P05, P10, P12, P19) stated that they would prefer to only have a single line of text. For example: *‘I think it’s better if it’s in a single line. If it’s a double line it tends not to work very well, it just blocks too much out of the screen.’* [P12].

Speaker colours Participants were all in favour of colouring the subtitles according to the speaker, although one commented that white should be reserved for non-diegetic speech, and three didn’t like the combination of yellow text on a black background.

5.6 Behaviour-Specific Comments

In addition to the themes that emerged across all the behaviours, participants also made comments that were specific to the implementation of a particular behaviour. This section summarises these.

5.6.1 A: Evenly spaced

The particular placement of the three subtitles around the scene had an impact on how effective they were. This was particularly notable in clip 3, where there were two people talking to each other, with the arrangement that they and the camera were seated around a 3-person café table. The precise alignment of the subtitles relative to the speakers was important:

‘Probably move them closer to the action, line it up’ [P15, A3]

‘That was more annoying. Just because it wasn’t well placed in the video. It probably would have worked better if it was. . . It was sort of placed under one person, but with the other it was over there, so it was just really “what’s the point of having it there?”. And it was just on this person, so if I look at this person I can read the subtitles, if I look at that person I can’t.’ [P23, A3]

‘It was weird that they were behind you. Why would they need to be behind you when you’re looking on at a situation?’ [P14, A3]

It was also notable that (even when presented as the first behaviour) some participants didn’t move around and find all three sets of subtitles. For example P05 (A1) found it *‘just a bit frustrating, because I wanted to look around ’cause I knew that there were other things’* — they hadn’t realised that there was more than one set of subtitles

When asked about changes they would like, some participants commented on the number and the position of the subtitles. While 3 subtitles positioned at 120° left significant gaps and meant that participants had to search for them and divide their attention, it was also felt that more subtitles would be ‘too much’, obstructing the view.

5.6.2 B: Follow head immediately

Other than the vertical position mentioned above, there were very few comments about behaviour B. Two participants mentioned the issue of VR sickness:

‘Yes I do. The only thing I think would impact it is I think it potentially has the danger of making me feel a bit of motion sickness. Only because when I’m moving too quickly, the words go a bit too quickly and therefore I’m more likely to feel a bit nauseous than if there was no subtitles’ [P08, B6]

‘There were a little bit dizzying really because they moved around so much with my eyes’ [P09, B2]

P08 commented that they thought that they could watch for longer with behaviour C, with the lag, because ‘I don’t think it would make me feel as funny’.

5.6.3 C: Follow with lag

Behaviour C was the most complex, with a number of parameters that could be varied in its implementation. The main feature of this behaviour was the ‘lag’, where the subtitle didn’t move until the participant had moved a certain distance (30°) away from it. The two key parameters here are the *animation speed* — how quickly the subtitle moves when it catches up; and the *threshold distance* — the angle of movement needed to trigger the follow (where 0° is the same as behaviour B).

Overall, this was more disliked than liked: 6 participants commented negatively on it (P04, P06, P15, P16, P19), while only 3 (P08, P17, P24) explicitly said that they preferred it. As noted above, the main issues were that it gave the behaviour an element of distracting unpredictability, as participants were not quite sure when the subtitle would start moving. P13, P14 and P23

commented on them being a ‘bit slow’, although it is not clear whether this referred to the animation speed or the threshold distance, or the combination.

The other feature of this behaviour was that the follow only went horizontally — when participants looked up or down the subtitle remained at the same level. The general preference was for subtitles to move vertically as well as horizontally — 9 participants (P02, P03, P05, P08, P10, P11, P16, P18, P24) either expressed a wish for this behaviour to include vertical motion, or commented on liking behaviour B because of its vertical movement. Only 3 participants (P13, P19, P21) preferred horizontal-only motion.

5.6.4 Appear in front, then fixed

There were very few comments specific to behaviour D. Despite all participants being briefed before each clip on how the next set of subtitles would behave, it was clear that this behaviour took more explanation and that some participants still did not understand how it worked.

5.7 Participants’ Suggestions

In the questions after each clip, participants were asked how they would change the subtitle behaviour, and during the summing up phase of the interview they were asked if they had any suggestions for changes, or things they would like to try. The suggestions can be grouped into three classes:

Directing Attention: Two participants suggested that subtitles could be used to direct the viewer’s attention to a particular part of the scene or to help them associate the text with the speaker:

‘Maybe if it [point of interest] is up towards the right hand side, maybe put the subtitles in the right hand side, to point you towards what’s happening’, and ‘Maybe the subtitle needs to have an arrow “this way” - you know like the games that you get.’ [P18]

[You could position the subtitles. . .] ‘a bit more, maybe next to the person’s head, like you say, so you can determine who is speaking.’ [P18]

‘I don’t know whether, potentially you could have, if people selected it, a little arrow that pointed you in the direction of where something was happening.’ [P08]

Interestingly, P20 commented on how finding subtitles validated that they were looking ‘in the right place’, although the subtitle positions were not dependent on the clip.

Following Eyes: P04 wanted the subtitles to follow the eye-gaze, not just the head orientation:

‘Can’t subtitles follow your eyes, on whether you need them or not? ’Cause they do that with some games don’t they? [Is that something you’d like?] Yeah well, if it’s possible. It would know where you were looking at and just keep below that. You can do that with new technology I think. You can surely do it with mobile phones they know when you’re looking at it and so it doesn’t switch off till you’ve stopped looking at it.’ [P04].

P18 made a similar suggestion: ‘*From a learning point of view, I think it would be fantastic if subtitles was more modified to where you was looking. It would be a great experience that.*’ [P18, A5].

Matching with scene: Four participants commented on how the subtitle presentation style might vary according to the scene. For example: *‘where it was the Galapagos - the sky was irrelevant - it was the landscape I was looking at, so I could have quite comfortably had my subtitles up at the top, as an aid, and that wouldn’t have interfered at all with my... So yes, it’s depending on what you’re watching isn’t it?’* [P22].

Control: Some participants wanted control over the subtitle presentation, including the ability to change it ‘on-the-fly’. For example: *‘I guess I would like an option to be able to move the height for the subtitles. Yes, probably instead of it deciding for me, if I could decide a myself for that particular TV show, and it would be good if I could do it while I was watching’* [P23].

Customisation: More common was the desire to personalise the subtitle behaviour, particularly when it comes to the vertical position and movement of the subtitles. For example:

‘giving the viewer the control over where they have their subtitles. You know, ‘cause some people — their line of vision is higher — they might want them at the top, they might feel more comfortable’ [P22]

[Would you like the ability to be able to do that yourself - to position them yourself?]
‘Yes, so to be able to try and figure out what you like, cause each person has different preferences, so it would be cool to move it around and see what would be better for you. That would be great, yeah.’ [P12]

Customisation of fonts and colours was also mentioned, as was the ability to turn the subtitles on or off altogether.

6 Discussion

In this section, the performance of the subtitles is compared against the expectations that emerged from the iterative design process, and then against the seven dimensions of the UX framework for subtitles. Finally, we summarise the results of the testing.

6.1 Performance against Expectations

Our user testing was performed with some expectations about how the behaviours might perform (§3.5). We actually found:

Evenly spaced. It was anticipated that viewers should be able to predict the location of subtitles and always have them in view, although with potential for obstruction. In fact, it was found that many people found themselves searching for the subtitles, even in situations where the points of interest were not high and between subtitles. Participants also felt restricted in their ability to explore the scene, and felt they missed out on content. Obstruction was less of an issue, although it was noted that there is no workaround — if a subtitle is covering something in the video there is nothing the viewer can do about it.

Follow head immediately. The expectations were that viewers would understand this behaviour and find subtitles easy to locate. Obstruction was thought to be a potential problem, as was nausea. The testing revealed these expectations to be largely correct, although obstruction was a bigger problem than expected and VR sickness a smaller problem (it is unclear if this effect will reveal itself when viewing longer pieces of content).

Follow with lag. This behaviour was expected to be less intrusive and less likely to induce unpleasant feelings. In fact, participants found that the delay gave the subtitles an unpredictable

nature which was found to be distracting. Obstruction was less of an issue than *Follow head immediately*. The motion of the subtitle returning to in front of the viewer was also found to be unpleasant, and this behaviour was not clearly better than *Follow head immediately* in this respect (although neither attracted many comments). As noted in §3.5, the precise parameters chosen in the implementation of *Follow with lag* could impact the UX quite significantly.

Appear in front, then fixed. It was anticipated that subtitles could be easily ignored if desired, but that obstruction could be a problem. It was also noted that this behaviour might not work well while exploring a scene, and that relocating subtitles could be difficult. What was found was that these expectations were broadly correct: in some scenes these subtitles worked well, but if viewers were looking around then the subtitle location became unpredictable and difficult to locate. This led to viewers feeling that they were restricted in how they viewed the scene and consequently missing out on content.

6.2 Performance against UX Framework

The design of the subtitle behaviours and of the evaluation protocol was done with reference to the UX framework for subtitles [29, 14]; Table 13 summarises how the different subtitle behaviours performed against this.

6.3 Summary

Overall, the ratings (§5.2) show that all four implementations were quite easy to use and performed well across all dimensions of the UX framework. Behaviour A (*Evenly spaced*) performed notably worse than the other behaviours, while behaviour B was the best performing.

The most popular behaviour (§5.1) was B: *Follow head immediately*, followed by C (*Follow with lag*) then D (*Appear in front, then fixed*). The simple HUD implementation of B was easy to understand and made locating subtitles extremely easy. It gave viewers freedom to explore the full 360° scene without missing either video or speech/subtitle content. The two negative aspects of this behaviour were that the position of the subtitle obscured video content and that it could cause motion sickness. The obstruction should be mitigated by positioning the subtitles slightly lower than the 15° below eye-line used here. Motion sickness is still an unknown — it was mentioned by 2 of the 24 participants, and was not a major problem for either, but the effects of viewing longer clips need to be ascertained.

Of the themes that emerged from the testing, the two most important attributes for subtitles seem to be that they are easy to **locate** (see §5.3.1) and that they allow the viewer **freedom to explore** the scene (see §5.3.2). Without these capabilities, viewers miss out on either the visuals or the speech and are required to apply conscious effort to using the subtitles. This results in them getting frustrated and losing immersion, and makes sustained viewing difficult.

As a result, the majority of participants preferred the subtitles to follow their gaze (head orientation) — 73% of people selected either *Follow head immediately* or *Follow with lag* as their preferred presentation style (*Appear in front, then fixed* was also considered to follow the gaze to an extent, since each new subtitle appeared in front of the viewer).

‘Can’t subtitles follow your eyes, on whether you need them or not?’ [P04, A6]

‘And you know what - I want my subtitles to follow me’ [P13, A6]

The consensus was also in favour of the subtitles following in both the vertical and horizontal dimensions (§5.6.3).

The vertical position of the subtitles emerged as a crucial parameter. The majority of these participants was that the subtitles should have been lower than they were here (§5.5.2), although there was considerable variation between individuals.

	A	B	C	D
Aesthetics	Aesthetics did not emerge as a strong discriminating theme. See §5.5.3 for cross-theme comments on aesthetics.			
Involvement	Immersion impacted by the effort required.	Somewhat impacted by obstruction of view, but allayed by complete freedom of view.	Reduced by unpredictability being distracting; mitigated by relative freedom to explore.	Unpredictability reduced immersion, but lack of subtitle movement helped.
Attention	Viewers felt they missed out and struggled to allocate their attention as wished; fixed subtitles restricted exploration.	Some obstruction of the video; complete freedom to explore; some distraction from movement.	Unpredictable movement distracting; some obstruction; freedom to explore	Restricted movement due to subtitles not following; moderate obstruction.
Familiarity	Familiarity did not emerge as a strong discriminating theme. See §5.4.1 for a summary of the comments on this dimension.			
Usefulness	Usefulness did not emerge as a discriminatory theme.			
Usability	Searching required, but to known position.	Always present and visible.	Sometimes delay before reading from lag	Some searching to re-locate; location not always predictable.
Endurability	Searching for subtitles was found to be tiring and require concentration, and would not be comfortable for long periods.	Low effort, but potential to cause motion sickness unknown for longer clips.		Unpredictability and effort of finding subtitles lowers endurance.

Table 13: Summary of performance of behaviours against the UX Framework for Subtitles.

7 Recommendation

Based on the results of this set of user testing, and with the important caveat that testing is required to assess the potential for motion sickness when viewing longer clips, the recommendation is to that behaviour B: **Follow head immediately** gives the best overall user experience. Any implementation of this behaviour should, however, use a lower position for the subtitles, The value for this will need to be determined by user testing, but will probably be around 20° below horizontal. A finer font should be used, again with user testing to ensure readability across a range of devices.

8 Future Work

We recommend that the following work is performed, so that the BBC can provide the best user experience to people consuming its 360° video or VR content.

8.1 Test with Longer Content

The primary question that follows from this study is one of endurance — do these results stand for clips longer than 2 minutes? This is particularly pertinent given that the most popular behaviour was one which was anticipated could cause motion sickness. Although very few participants commented on this, and those only as a minor issue, it is likely that such effects would build up over longer periods of viewing. Testing the UX of this behaviour over longer pieces of content should be a priority. Given the opportunity, it would also be interesting to test the other behaviours over longer periods; *Appear in front, then fixed* in particular was found confusing, but may be a behaviour that participants could learn to understand.

8.2 Vertical Position

The vertical position of the subtitle blocks in all conditions (but particularly in *Follow head immediately*) was generally considered too high. Further testing is required to determine what the best position is. Based on these results, we recommend testing values around 20° below horizontal.

8.3 Visual Presentation

While the primary focus of this work is the subtitle behaviour, there are many other aspects of the design that warrant further investigation and user-testing. The following is a selection of research questions that we have encountered; as with the behaviour, these are attributes that are likely to impact on the user experience as a whole, and not just the ability of the viewer to read the subtitles. Of primary importance are the background of the subtitle blocks and the size of the text.

Depth At what distances from the camera (i.e., the viewer) should the video sphere and the subtitles be placed? The values of 25m and 4m respectively here looked reasonable, but are unlikely to be optimal. This question is likely to become more important and challenging when the video is stereoscopic.

Size The size and resolution of the display can impact the preferred subtitle size [30]. Given the differences between television and a VR headset, therefore, the (apparent) size of the subtitles needs further investigation. Studies should take into account the range of capabilities (e.g., the resolution) of the different display hardware units available and explore if the text is easily legible, or whether users want it smaller, so as not to obscure too much, or bigger, in order to read it more easily? The results suggest that slightly larger text should be tested.

Font The sans-serif font used in this implementation (Roboto) aimed to replicate the style used for television subtitles; this is not necessarily the most appropriate for viewing in an HMD.

Indeed, the results revealed some desire for a ‘finer’, ‘less square’ font. Further testing should be performed to reveal what style of text works best for users, and whether it needs to match the video content.

Background The theme of *obstruction* was present across behaviours, but was particularly notable in the most popular (*Follow head immediately*). Tests should be performed with semi-transparent backgrounds to see if this can reduce the problem of obstruction without compromising readability; it may also help lower the impact of subtitles on *immersion*.

We recognise that there will be interactions between some of these presentation decisions and behaviours, but an iterative approach that captures subjective feedback from users should allow an effective solution to be found.

8.4 Optimise ‘Follow with lag’ parameters

Behaviour C — Follow with lag— was also popular, and it is possible that this could be a better implementation than ‘*Follow head immediately*’ if the optimum parameters (of animation speed and threshold distance) are used. Performing experiments to optimise these, then testing again against behaviour B would be interesting, particularly if longer viewing with the simple HUD does lead to motion sickness problems.

8.5 360° Content Production

The experimental procedure exposed each of 24 participants to 4 of 6 360° video clips. Although their comments were directed towards, and influenced by, the subtitle behaviours, many interesting comments were made about the aspects of the videos that participants felt worked or didn’t work in that medium. An analysis of the data would reveal insights that could benefit producers of 360° video content.

8.6 Dynamically Positioned Subtitles

Finally, this piece of research has concentrated on subtitling behaviours that can be implemented without any additional authoring steps (i.e., a video file and a standard subtitle file). Placing subtitles according to the scene was found to be beneficial to subtitle users [3]; it is likely that positioned subtitles could be even more beneficial in 360° environments, where the low resolution and limited range of camera-shots make speaker identification more challenging. In addition to this, subtitles have the potential to be used for direct attention, which could be particularly helpful when viewers do not have audio cues that can alert them to the presence of a new character, for example. Developing a behaviour that can combine the utility and simplicity of the ‘*Follow head immediately*’ behaviour, with the benefits found for positioned subtitles, and that can also replicate the attention-directing audio cues that subtitle users may miss would be an interesting challenge.

A Appendix 1: Follow-up Questions

The following script was used **as a guide** for the semi-structured interview following each clip:

1. First of all, did you enjoy that?
2. Here is a sheet of scales. As I ask you the questions, I’ll sometimes ask you to give me a rating on the sheet. The first one is to let me know how much you enjoyed the clip overall could you circle the scale and explain why please? [complete *overall enjoyment* scale]
3. Can I ask you what you thought of the subtitle presentation?

4. How easy did you find it to locate the subtitles? [complete *locate* scale]
5. And how easy was it to read them? [complete *read* scale]
6. What did you think of the appearance of the subtitles?
7. (For behaviours B,C,D) How did you feel about the subtitles changing positions on the screen?
8. Is there anything you would have changed, or would like to be able to adjust?
9. Did you have any problems identifying which person was speaking?
10. So overall, how confident are you that you followed all the speech? [complete *follow speech* scale]
11. Could I now ask you about the video how easy did you find it to follow the action, and how free did you feel you were to look around the scene? [complete *follow visuals* scale]
12. Did you feel that the subtitles affected your attention while watching this clip? Was that a good thing or a bad thing?
13. Were there any parts of the clip where you had problems either viewing the video or following the speech?
14. So, taking the visuals and speech together, do you feel like you have a good understanding of what was happening in the clip? [complete *overall understanding* scale]
15. I now want to find out how immersed you felt in the scene, so could you complete the final rating, again explaining your decision. [complete *immersion* scale]
16. Did you feel that the subtitles affected your level of immersion? How? Did this affect your enjoyment?
17. How does viewing content with this method of displaying subtitles compare to traditional subtitles on TV?
18. Do you think you could watch content subtitled in this way for an extended period of time?

B Appendix 2: Wrap-up Questions

After the follow-up questions for the last clip had been completed, the following questions were used **as a basis** for the final wrap-up.

1. How did the different subtitle behaviours compare? What advantages and disadvantages did you find for each?
2. Which type of subtitle behaviour did you prefer: which would you choose for 360 videos in the future?
3. Is there any other way of presenting subtitles that you would like to see?
4. Do you have any view on the visual appearance of the subtitles — the font, colours, etc.?
5. Do you have any other comments?

References

- [1] Ofcom. Television access services. review of the code and guidance, 2006. https://www.ofcom.org.uk/__data/assets/pdf_file/0016/42442/access.pdf, last accessed March 2017.
- [2] Alia Sheikh, Andy Brown, Zillah Watson, and Michael Evans. Directing attention in 360-degree video. In *Proc. IBC '16*.
- [3] Andy Brown, Rhia Jones, Mike Crabb, James Sandford, Matthew Brooks, Mike Armstrong, and Caroline Jay. Dynamic subtitles: The user experience. In *Proc. TVX '15*, 2015.
- [4] A Foerster. Towards a creative approach in subtitling: a case study. In J.D. Cintas, A. Matala, and J. Neves, editors, *New Insights into Audiovisual Translation and Media Accessibility*, pages 81–98. Rodopi, 2010.
- [5] Richang Hong, Meng Wang, Xiao-Tong Yuan, Mengdi Xu, Jianguo Jiang, Shuicheng Yan, and Tat-Seng Chua. Video accessibility enhancement for hearing-impaired users. *ACM Trans. Multimedia Comput. Commun. Appl.*, 7S(1):24:1–24:19, November 2011.
- [6] Yongtao Hu, Jan Kautz, Yizhou Yu, and Wenping Wang. Speaker-following video subtitles. *ACM Trans. Multimedia Comput. Commun. Appl.*, 11(2):32:1–32:17, January 2015.
- [7] Raisa Rashid, Jonathan Aitken, and Deborah I. Fels. Expressing emotions using animated text captions. In Klaus Miesenberger, Joachim Klaus, Wolfgang L. Zagler, and Arthur I. Karshmer, editors, *Computers Helping People with Special Needs*, volume 4061 of *Lecture Notes in Computer Science*, pages 24–31. Springer Berlin Heidelberg, 2006.
- [8] Alina Secară. R U ready 4 new subtitles? Investigating the potential of social translation practices and creative spellings. *Linguistica Antverpiensia, New Series ? Themes in Translation Studies*, 0(10), 2013.
- [9] British Broadcasting Corporation. Subtitle guidelines v1.1.4, 2016.
- [10] Joseph J. LaViola, Jr. A discussion of cybersickness in virtual environments. *SIGCHI Bull.*, 32(1):47–56, January 2000.
- [11] Sarah Sharples, Sue Cobb, Amanda Moody, and John R Wilson. Virtual reality induced symptoms and effects (vrise): Comparison of head mounted display (hmd), desktop and projection display systems. *Displays*, 29(2):58–69, 2008.
- [12] Peter J. Passmore, Maxine Glancy, Adam Philpot, Amelia Roscoe, Andrew Wood, and Bob Fields. Effects of viewing condition on user experience of panoramic video. In *Proc. ICAT-EGVE '16*, 2016.
- [13] Adam Philpot, Maxine Glancy, Peter J Passmore, Andrew Woods, and Bob Fields. User experience of panoramic video in cave-like and head mounted display viewing conditions. In *Accepted for TVX '17*.
- [14] Michael Crabb, Rhianne Jones, and Mike Armstrong. The development of a framework for understanding the ux of subtitles. In *Proc. ASSETS '15*. ACM, 2015.
- [15] Julie N. Buchan, Martin Par, and Kevin G. Munhall. Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience*, 2(1):1–13, 2007. PMID: 18633803.
- [16] Charissa R. Lansing and George W. McConkie. Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics*, 65(4):536–552, 2003.

- [17] Klin A, Jones W, Schultz R, Volkmar F, and Cohen D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, 59(9):809–816, 2002.
- [18] Guanghan Song, Denis Pellerin, and Lionel Granjon. Different types of sounds influence gaze differently in videos. *Journal of Eye Movement Research*, 6(4), 2013.
- [19] Tom Foulsham and Lucy Anne Sanderson. Look who’s talking? sound changes gaze behaviour in a dynamic social scene. *Visual Cognition*, 21(7):922–944, 2013.
- [20] Antoine Coutrot and Nathalie Guyader. How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, 14(8):5, 2014.
- [21] The New York Times. The displaced — 360 vr video — the new york times, 2015.
- [22] Ian Hamilton. How to do subtitles well — basics and good practices, 2015. http://www.gamasutra.com/blogs/IanHamilton/20150715/248571/How_to_do_subtitles_well__basics_and_good_practices.php.
- [23] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. Managing mobile text in head mounted displays: Studies on visual preference and text placement. *SIGMOBILE Mob. Comput. Commun. Rev.*, 18(2):20–31, June 2014.
- [24] V. Charissis and M. Naef. Evaluation of prototype automotive head-up display interface: Testing driver’s focusing ability through a vr simulation. In *2007 IEEE Intelligent Vehicles Symposium*, pages 560–565, June 2007.
- [25] Gary E. Riccio and Thomas A. Stoffregen. An ecological theory of motion sickness and postural instability. *Ecological Psychology*, 3(3):195–240, 1991.
- [26] Omar Merhi, Elise Faugloire, Moira Flanagan, and Thomas A. Stoffregen. Motion sickness, console video games, and head-mounted displays. *Human Factors*, 49(5):920–934, 2007.
- [27] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.
- [28] Barney G. Glaser. The constant comparative method of qualitative analysis*. *Social Problems*, 12(4):436, 1965.
- [29] R. Jones, M. Crabb, and A. Brown. Investigating the user experience of variable subtitle displays. Technical report, BBC Research & Development, 2015.
- [30] T. Vigier, Y. Baveye, J. Rousseau, and P. Le Callet. Visual attention as a dimension of qoe: Subtitles in uhd videos. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6, June 2016.