AUTHOR(S):

TITLE:

YEAR:

Publisher citation:

OpenAIR citation:

Dynamic Subtitles: the User Experience

1st Author Name
Affiliation
Address
e-mail address
Optional phone number

2nd Author Name
Affiliation
Address
e-mail address
Optional phone number

3rd Author Name
Affiliation
Address
e-mail address
Optional phone number

4th Author Name
Affiliation
Address
e-mail address
Optional phone number

5th Author Name
Affiliation
Address
e-mail address
Optional phone number

ABSTRACT

Subtitles on television are typically placed at the bottom-centre of the screen. However, placing subtitles in varying positions, according to the underlying video content ('dynamic subtitles'), has the potential to make the overall viewing experience less disjointed and more immersive. This paper describes an investigation into the User Experience of dynamic subtitles. Qualitative data from habitual subtitle users demonstrates that dynamic subtitles can lead to an improved experience, although not for all types of user. Eye-tracking data was analysed to compare the gaze patterns of subtitle users with a baseline of those for people viewing without subtitles. This indicated that dynamic subtitles are less disruptive to the viewing experience than traditional ones.

Author Keywords

TV, Subtitles, User Experience, Accessibility, HCI, Eye-tracking, Attention Approximation

ACM Classification Keywords

H.5.1 Information interfaces and presentation (e.g., HCI): Multimedia Information Systems; K.4.2 Social Issues: Assistive technologies for persons with disabilities

INTRODUCTION

Traditionally, subtitles are positioned so they are centred at the bottom of the television screen. In recent years, however, there has been a noticeable increase in research experimenting with non-traditional placement of subtitles [3, 5, 6]. Broadcasters wish to understand how 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 a more immersive, engaging, emotive and aesthetically pleasing way.

One approach is to change the position of subtitles on the screen, placing each subtitle block so that it takes into account the underlying images. This was done in a limited sense by the DTV4All project [2], who tested 'displaced' subtitles — on the left, centre or right, of the bottom of the screen according to the position of the speaker. Hong *et al.* [3] and Brooks and Armstrong [1] went further, placing subtitles across the entire screen. These subtitles are known as 'dynamic captioning' [3] or 'dynamically positioned subtitles' [1]; this paper uses the briefer term 'dynamic subtitles'.

While traditional subtitles have been studied in some detail, the limited previous research into dynamic subtitling has revealed little about the User Experience (UX) of these subtitles. DTV4All [2] combined eye-tracking, comprehension testing and preference elicitation. They found that displaced subtitles were less popular and resulted in lower comprehension than the more standard use of colour, although participants did fixate more quickly on the text. Brooks and Armstrong [1] focused only on analysing gaze data; they found that the overall dwell time on dynamic subtitles was lower than on traditional subtitles, indicating that more time was spent looking at the drama. Hong *et al.* [3] concentrated primarily on the technicalities of their automatic placement system, but measured subtitle users' comprehension of dynamic subtitles, and two aspects of UX — 'naturalness' and 'enjoyment', with participants scoring each aspect on a scale of 1–10. They found that dynamic subtitles enabled their subtitle users to better enjoy videos and to better recognise the speakers.

We seek to understand the UX of dynamic subtitles in more detail, hypothesising that they will provide an improved user experience by making it easier to follow both the subtitles and the video content. This work seeks to explore that hypothesis, using two experiments. The first extends the initial study of Brooks and Armstrong [1], analysing the original data in conjunction with additional eye-tracking data to discover how

much gaze patterns differed between subtitled and non-subtitled content. While this experiment illuminates how people's viewing patterns differ between traditional and dynamic subtitles, it does not provide any further insight into the UX, e.g,. whether subtitle users like or dislike dynamic subtitles, and why. The second experiment, therefore, investigated the UX of dynamic subtitles, capturing the attitudes of habitual subtitle users towards an example of content with dynamic subtitles, and their thoughts on what makes an effective subtitle placement.

EXPERIMENT ONE

This research uses data from Brooks and Armstrong [1], which is combined with new data and analysed in a novel way. This section summarises their study, describes how the additional data was captured, and explains the analysis process. The results are presented in combination with those from the second experiment.

In Brooks and Armstrong's study [1], 4 clips were taken from 3 episodes of the drama 'Sherlock'. The clips lasted between 1:50 and 2:00 minutes, and 5 versions were created from each: French audio, traditional subtitles; French audio, dynamic subtitles; English audio, traditional subtitles; English audio, dynamic subtitles, and; English audio, no subtitles (baseline case). 24 participants (native English speakers, who did not understand French; participants were not habitual subtitle users) watched the clips, in the same order, on a television in a 'living room' lab. The clips were first presented in one of the 4 subtitle/language combinations, counterbalanced so that 5-6 different participants watched each version. 21 of the participants then viewed one of the clips (chosen at random) in the baseline condition: clips A, B and C were viewed by 6 people, and clip D by 3.

Additional Data Collection

Additional baseline data was collected from 8 people, (convenience sample; 5 male, 3 female, aged 21–55) who watched clip D without subtitles. As [1], participants did not normally use subtitles. The participants were introduced to the study and seated in front of the eye-tracker (configured as [1] and experiment 2, below) to watch the clip as they would normally watch television.

Data Analysis

The hypothesis being tested is that dynamic subtitles allow gaze patterns that are closer to those of viewers watching without subtitles. This analysis therefore compares gaze patterns from viewers watching with subtitles against gaze patterns from the baseline data of people viewing without subtitles (the same baseline data was used for analysis of data from experiment two).

Analysis of eye-tracking data typically compares the gaze statistics for areas of interest (AOIs) in the scene. In this case, however, it is not known, a-priori, where viewers in the baseline condition will fixate. Consequently, while it is possible to define AOIs for the subtitles, it is not for the underlying video content. In order to explore the data, therefore, the scene is evenly divided into chunks, both spatially — as a grid — and temporally — into time slices. Having applied this approximation, it is possible, for each slice of time, to identify which regions of the scene were viewed by participants in each condition. Crucially, the application of regular approximation allows direct quantitative comparison of gaze patterns. In this case it is possible to measure how much the gaze pattern of a subtitled scene differs from that of the same segment without subtitles. By measuring the differences between the baseline and the traditional and dynamic conditions, we can compare the extent to which the subtitle conditions required users to divert their gaze from their usual focus.

In this analysis, the gaze pattern is considered in terms of dwell time. Thus, for each time slice we calculate, for each box in the grid, the *proportion* of total possible attention for that window. If there are n participants, then the total possible attention (A_{total}) is n times the length of the time slice. The attention received by an individual box (A_{box})is the sum of the durations of all fixations for all participants that occurred in that box during the time slice. The proportion of attention for the box is therefore A_{box}/A_{total}, and the gaze pattern for a given slice comprises of an attention value for each box in the grid. The sum of these values across the grid will approach 1, but will be less due to time spent on saccades, or fixations of less than 100ms (which were discarded). It may be lowered further if one or more participants looked away, or the eye-tracker failed to record some data. Note that if fixations overlap more than one time window, only the duration that lies within the window is counted. Similarly, if a fixation lies near to the boundary of two areas, its duration may be split proportionately between them.

For these results, the 1920×1080 pixel scene was divided into an 8×5 grid (resulting in 40 240×216 boxes), and the 115 second clip into 1s slices. The grid size and slice length were determined by the size and duration of the subtitles (subtitles were visible for a mean time of 2.7s, and the mean length of a subtitle block was 550 pixels) — it was necessary to get enough detail to differentiate between areas of the screen and between subtitles, but have the grid/slice combination coarse enough to capture enough data to make meaningful comparisons.

For each temporal slice, a gaze intensity value was calculated for each box in the grid. The intensity of each box represented the proportion of attention received, as described above. To allow for experimental error in gaze position detection, the contribution from those fixations within 20 pixels (approximately 8% of the length of the box sides) of box edges was divided between boxes in ratios proportionate to the edge proximity.

A metric was calculated to reflect the size of the difference of the overall attention pattern for two segments. To do this, a difference grid was calculated, with each box containing the difference between the corresponding

boxes under the two conditions. This grid was smoothed (Gaussian smoothing over the 8×5 grid, with a radius of 1, meant that a shift of attention between neighbouring boxes had a smaller effect on the metric than between distant boxes) and a root mean square value was calculated; these values were linearly scaled to lie between 0 and 5. The difference values calculated in this manner are based on aggregated data, i.e., the difference was comparing the gaze of all participants in one group with all participants in another. This results in a single difference value for each segment of the clip for each condition — a lower value indicates that the gaze pattern was closer to the baseline.

This approach was also used to analyse data collected from the participants of experiment two; the results of both experiments are therefore presented together (following the description of experiment two).

EXPERIMENT TWO
The second experiment was designed to capture qualitative data on the UX of dynamic subtitles from people who habitually used subtitles. In addition, eye-tracking data was collected, to compare gaze patterns across different subtitle placements. The same techniques (and baseline data) were used, although these participants were not shown traditional subtitles. This was to allow a more in-depth exploration of dynamic subtitles.
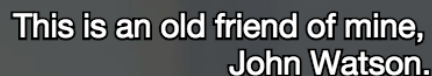
Participants
26 participants were recruited for inclusion in this study. Recruitment was performed by an external agency, and participants were recruited who regularly (i.e., daily) used subtitles at home to watch TV with the sound on. Participants were aged between 22–67 (M = 47.2, SD = 13.6). A mix of gender (7 male, 19 female), socio-cultural/economic backgrounds was also used.

Stimulus
Participants were shown a clip from the TV drama "Sherlock" (Series 1, Episode 1 — clip D in Experiment One). This segment included 3 main characters, plus a fourth who appeared briefly, and contained 34 subtitle blocks. Two characters, Mike, and John Watson, enter a chemistry laboratory, where Sherlock is performing an experiment. Mike introduces Watson to Sherlock, who deduces that the they are both looking for a flatmate.

Dynamic subtitles were authored for the original experiment: each subtitle was assigned a position based on a number of factors: the character speaking the line; the background, and; the position of the previous and subsequent subtitles. All subtitles were displayed as white text (Helvetica Neue, 32 pixels high) with a slim black outline (Figure 1). In order to allow fair comparison, timing remained identical to that authored for the original (traditional) subtitles.

In order to explore the important factors for subtitle placement, alternative positions were authored for 4 of the dynamic subtitles (numbers 3, 19, 24 and 33 from the

Figure 1: The text used to present the subtitles.

sequence of 34 in this clip). Re-authoring of these led to 2 further subtitles being re-positioned (numbers 23 and 25) so that the reader's gaze did not have to jump too far between consecutive subtitles. The original and revised positions of the four subtitles can be seen in Figure 4.

A Framework for Qualitative Data Capture
In order to measure the UX of subtitles in a reliable and meaningful way, a framework is proposed. This was developed following analysis of the literature regarding UX measurement, with particular inspiration being drawn from [4]. This framework provides researchers with a means to comparatively measure the user experience of subtitles[1]. Drawing on existing examples of User Experience Surveys this framework identifies and defines several key aspects of the user experience that can be used to evaluate the UX of variable methods of subtitle display. These aspects are described below.

Aesthetics is 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.

Involvement measures how engaged users are with the 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 a engaging and enjoyable experience. Users with low levels of involvement would feel less involved in the subtitled content.

Attention is awareness of 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.

Familiarity measures how much users feel the current subtitle display matches their expectations. High levels of familiarity indicate a coherence in the relationship between the subtitles and the video content. Low levels of familiarity will indicate a disconnect in what is perceived as routine subtitle practice

Perceived usefulness measures how useful the display of the subtitled content is. Users who perceive high levels of usefulness will see a high levels of value in the subtitle display; users with low levels of perceived usefulness will see low levels of value.

---

[1]The primary purpose of this framework is to provide an overall measure of the user experience when viewing different methods of subtitle display. This framework does not deal with reading rates or comprehension levels.

Perceived usability measures 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.

Endurability is defined as a user's willingness to view subtitled 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 would be less likely to want to use this method again in the future

This framework was used to inform the semi-structured interview, with questions designed to probe participants on each of the aspects above. This ensured comparability of the responses given across the group. The questions used are given alongside the results.

Design and Procedure

The session was run in a usability lab set up as a living room with an adjacent control and viewing room. Sessions were recorded and transcribed. Participants watched the clip on a 47 inch television. A Tobii X-120 eye-tracker was used to record the gaze of participants as they viewed the clip; this was placed on a coffee table 1.8m in front of the television. To facilitate the process of positioning the participants correctly relative to the eye-tracker, participants were seated on an adjustable office chair approximately 0.7m in front of the eye-tracker.

The experiment was started by informing participants that the purpose was to capture their opinions on some subtitles they would see in a short clip. They were seated in front of the eye-tracker and allowed to adjust the television volume to a comfortable level. Participants adjusted the position of their chair to within the range of the eye-tracker. Once comfortable, the eye tracker was calibrated, then recording started and the clip shown. The videos were counterbalanced so that half of the participants saw the video with the re-authored subtitles in their original positions, half with the revised positions.

After viewing the clip, participants were asked for their first reactions. In order to explore what makes a well-positioned subtitle, they were then asked to give their thoughts on the alternative positions for each of the 4 re-authored subtitles. Participants were shown the pairs as still images (using the first frame for which the subtitle was present) on the television screen. They were asked to comment on what they liked and/or disliked about each, and to give a preference.

The final part of the experiment was a semi-structured interview to capture subtitle users perceptions and experiences of, and attitudes towards, dynamic subtitles. The interview designed a series of qualitative probes (detailed in the results, below) to investigate the UX, as explicated in the framework. The semi-structured nature of the interview allowed for comparability of responses across the group whilst also providing participants with the opportunity to offer insight into additional factors or issues deemed important to understanding the UX of dynamic subtitles.

EYE-TRACKING RESULTS

As a test of the difference metric, the data from the revised dynamic subtitles was directly compared to that from the original dynamic subtitles. The difference would be expected to be low for most of the clip, but high during those periods when the subtitles were presented. This is what was found - the median value across the 115 1-second slices was 0.9, with peaks of 2-3 during the changes. Since this worked as expected, the metric was used to compare the overall gaze patterns of people who watched the subtitled clip with those who watched without subtitles (the baseline).

Comparison of dynamic with traditional subtitles was made using the data from experiment one. The aggregated data gave usable gaze data for 5 participants watching with each of the traditional and dynamic subtitles (French audio), and 11 participants watching without subtitles. Figure 2 plots the differences between each subtitle condition and the baseline across the clip, with the filled line indicating which is closer (below the x-axis indicates that the gaze pattern of dynamic subtitles was less different from the baseline). Looking across all slices, the median difference values are 1.9 for the dynamic subtitles (95% confidence intervals ±0.14) and 2.3 for the traditional subtitles (±0.18). This indicates that, on an average slice, the viewers of dynamic subtitles have gaze patterns that more closely resemble those of un-subtitled content than viewers of traditional subtitles.

Comparison of the revised with original subtitle positions was done using data from experiment two and baseline data from experiment one. Figure 3 summarises the results of both experiments. This plot shows the median value, and 95% confidence intervals, for the slices in the clip, divided into those slices where subtitles were present (of which there were 87), and those where they were not (28). In this graph, it can be seen that the difference values for segments without subtitles were all relatively low — this is what would be expected, as the stimulus was essentially the same for all participants in these segments (although there will be some effect from people moving their gaze between the subtitle and the video). In those segments containing subtitles, however, the gaze patterns were all more different than the baseline. In particular, it is notable that traditional subtitles resulted in the largest difference, while dynamic subtitles had smaller differences (the median difference values for segments with subtitles in experiment 1 are 2.78 for traditional subtitles and 1.96 for dynamic subtitles).

The gaze patterns for the two versions of the clip in experiment 2 (which are compared in more detail later) were both closer to the baseline than those from the users in experiment one (Figure 3). There are two factors that might account for this. First, the viewers in the
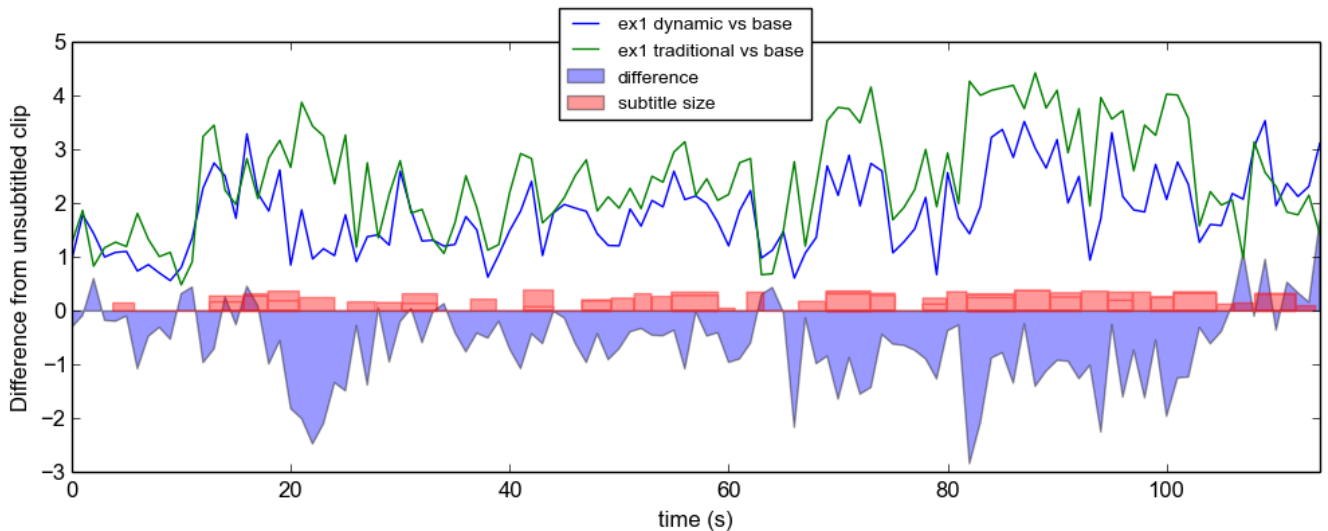
Figure 2: A comparison of how much gaze patterns in the traditional subtitle and dynamic subtitle conditions differed from the baseline. The differences between traditional subtitles and the baseline are shown in green; those between dynamic subtitles and the baseline are in blue. The filled line indicates which was closer: below the x-axis shows that the gaze pattern for dynamic subtitles was closer to the baseline than for traditional subtitles. Red bars indicate when subtitles were visible, with height correlating to the number of characters.
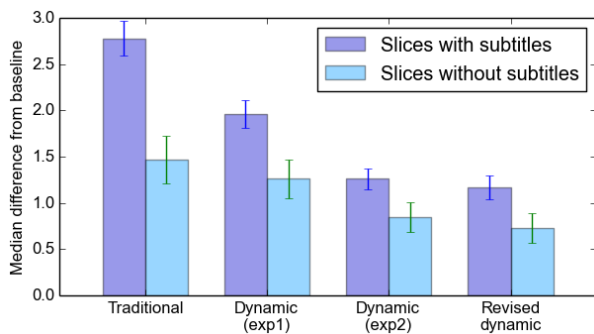

Figure 3: Median difference values for the 1s slices for the different conditions. These are split into values for slices in which subtitles were visible, and those in which there were none.

second experiment were habitual subtitle users; second, participants in the second experiment had the ability (in some cases) to augment their use of subtitles with lip reading and the English audio. These factors may also explain the differences between experiments 1 and 2 for those slices without subtitles — the experienced subtitle users and lip-readers of experiment 2 may revert their gaze to the baseline more quickly than the participants of experiment 1.

QUALITATIVE RESULTS

The qualitative data comprises two parts: the first impressions of participants and their responses to a set of questions aligned to the framework (above). The results

of the subtitle positioning exercise are reported in the next section.

In summary, 5 participants did not like dynamic subtitles, 8 were broadly positive, and 12 were very keen on the idea. Interestingly, the 3 participants who most disliked the dynamic subtitles (P2, P14 and P17) were ones who did not totally rely on subtitles: P2 was slightly deaf in one ear, and used subtitles when the young kids' 'toys are out'; P14 had no diagnosed hearing problem, but liked to use subtitles 'as a double check', and; P17 said 'I don't rely on them'.

First Impressions

Overall, the first impressions of people were mixed. Three participants were immediately negative: they felt that they had to 'follow them round' and found them distracting. For example, P14 stated:

'I hated them, really hated them, I found them really distracting. Every time one flicked up my eye would flick to it, instead of it just being at the bottom where I would just read it when needed. It made me feel tense waiting to see where they would appear.'

Two were mixed, liking aspects of dynamic subtitles, but not seeing sufficient benefit for them to want to change from the familiarity of traditional subtitles, e.g., P9:

'I liked the fact that it was with them as they were speaking. It's not that much of a big problem for me because I'm not totally deaf'; 'If I had to pick between this and familiar ones, I would stick with familiar ground just now'.

Seven others were immediately positive. They identified two main benefits to dynamic subtitles: it was possible to spend more time looking at the video content rather than reading subtitles, and; identifying which person was speaking the dialogue in the subtitle was easier. For example, P5 said:

> 'Loved it. It's there for you, it's next to that person saying it. So you don't need to have the different colours. With this you knew who was talking straight away and you felt more sucked into the television.'

P18 also found identifying the speaker easier, and noted that he was less likely to miss things in the video:

> 'It was really good. . . it gives you a much clearer idea of who is speaking. . . it's more integrated. I can spend more time on the video content. I feel that with this you can see a lot more of the picture as well, not just the words at the bottom, because you are reading and watching at the same time. It makes the experience of watching clearer and easier to understand.'

The remainder of the participants fell somewhere in between, not quite sure what to make of the subtitles immediately after viewing a 2 minute clip for the first time.

Semi-structured Interview

The semi-structured interview took place after the subtitle positioning exercise. The questions that formed the basis for the discussion, and the responses to them, are summarised below. Participants were asked to elaborate where possible.

General Comments

'What do you think are the advantages / disadvantages of having subtitles positioned in different places on the screen?'

The two themes identified in the first impressions — of being able to identify speaker more easily, and of missing less of the video — remained prominent, and were noted by more of the participants. There were also comments about how dynamic subtitles felt more integrated with the programme and 'became part of the story' (P0). For example, P6 and P8 said, respectively:

> 'They seem really well integrated and its easy to switch between the subtitles and the visuals without feeling like it was disjointed.'

> 'It's almost cinema like — you have that feel of being enveloped of it'

More participants commented on the aesthetics, such as P16, who said it was 'aesthetically pleasing', and P20: 'It seems like a very artistic way of doing it.'

Attention

Were you able to follow both the subtitles and the video content comfortably? How does this compare to when subtitles are placed at the bottom of the screen? Does your attention to the video clip differ?

Responses to these questions were largely positive. 16 participants stated that they were able to follow both video and subtitle content, with many noting that the dynamic subtitles were an improvement on traditionally placed ones. For example, P10 stated:

> 'With traditional subs you have to split your attention, but with this because it's so near to peoples faces you can also get a lot of the physical body language of what people are saying'

Others were able to follow the content, but found it more difficult than traditional subtitles (e.g., P19 'would rather have them in a predictable place'; also P20). Two participants (P9, P17) were wholly negative: P17 didn't want to read the subtitles, and found them intrusive.

Aesthetics

Did you find the positioned subtitles appealing to look at? How do they compare to traditional subtitles? Did the positioned subtitles add or detract in any way from the aesthetics of the video?

Although 4 participants (P2, P9, P14, P17) thought dynamic subtitles detracted from the overall aesthetics (e.g., P14: 'Because of their position they detracted from the video'), 15 participants thought they were an improvement. For example, P16 stated:

> 'Compared to traditional subtitles this adds aesthetic value. I'm looking at the whole picture in the few seconds that gives me, but with [traditional subtitles] you have to go down and then back up. This shows you everything that you want to see and is pleasing on the eye. This gives me time to read what is going on and not having to move. I'm just looking straight across.'

P11, also noted how 'I liked them, they were appealing, it reminded me of a comic when you're reading the action and the words'. 4 people (P18, P20, P23, P24) expressed concern about their use in social situations. They thought that they would detract from the aesthetics for co-viewers (who did not rely on subtitles), as they would be harder to ignore.

Usability

Did you have any problems locating the subtitles? Were you able to follow the subtitles comfortably? Did you have any problems identifying the speaker? How did you cope with the subtitles changing positions on the screen? How do reading subtitles placed dynamically on the screen compare to reading the subtitles at the bottom of the screen?

Several people commented that it took a short period of adjustment before they were used to the subtitles ('like a new pair of glasses' - P11). 3 participants (P8, P9, P20) commented on problems locating the subtitles on one or two occasions, while P17 noted that they were 'too

immediate' and difficult to miss. Speaker identification was generally not a problem, although 2 people said that colours could be used to help.

Usefulness
How useful do you find this as a method of displaying subtitles? Do you see any added value in this way of displaying subtitles? Can you think of any instances where having some, or all, of the subtitles displayed like this would be useful or add value? OR equally, any instances where you think they might be unsuitable?

Again, the consensus was that presenting subtitles dynamically was useful, although not necessarily appropriate for all types of programme. Most people thought that it would not be useful for news, which has a relatively static format, although P24 felt that having the words alongside a presenter, if there was space, might be useful. Dynamic subtitles were considered most suitable for drama, or for situations where you have many people talking (e.g., a panel — 'The words can be placed next to the person that owns the speech' — P11.). For example, P8 commented that it was:

> 'Very useful, the added value is that there is less attention processes being spend on just reading... [Normally] I don't know whether the actor has done anything when I've been reading. And reading is compromising the action. This time I'm reading and also catching the movement in the same field.'

P0 said, 'The added value for this is that its more dynamic, it raises my attention to the whole piece, it seems like it's more integrated with the images', while P7 said, 'Would be a big plus to have subtitles this way'. Two participants noted the difference between usefulness and overall appeal — P4 said that dynamic subtitles were 'not useful, but preferable', while P2 said 'Yeah it could be useful... but I don't like it how it is there'.

Involvement
Do positioned subtitles have any impact on how engaged you feel with the subtitled text (and your enjoyment of reading the subtitles)? Do positioned subtitles have any impact on how engaged you feel with the overall video (and your enjoyment watching the video)?

The majority of the participants in the experiment felt that the dynamic subtitles meant that they were more engaged with the content, or enjoyed it more. P14 and P19 felt that they detracted from their enjoyment as they were 'more conscious of them' (P19) or 'I was trying to second-guess where the text would appear'. One of the key benefits of dynamic subtitles that participants identified as increasing their involvement was that they were 'more aware of what was going on' (P13) and able to identify small, but important, aspects of the video that would otherwise have been missed. This was specifically picked out by participants 16 and 18:

> 'I wouldn't have caught a lot of the small social cues if I were watching this with traditional subtitles.'

> 'Normally you are looking down at the bottom of the screen and you miss facial expressions, but with this nearer to the mouth it's easier to see everything.'

Familiarity
Does this method of displaying subtitles feel familiar (or strange)? How does this method of displaying subtitles compare to traditional subtitles?

For P14 ('strange and distracting') and P17 dynamic subtitles felt strange, while for some people they felt natural (P4 — 'feels quite natural', P8 — 'first impression was that this is intuitive', P10 — 'because I read comics it felt familiar', P18 — 'It felt happier; it was more natural'). For some it felt unfamiliar, but something that could be got used to, either quickly (e.g., P7: 'It felt a little bit strange, but only for a nanosecond – as quick as that'), or more slowly (e.g., P20 'It felt new, I feel like I would have to concentrate but I think that would disappear over continued use').

Endurability
Do you think you could you watch subtitled content like this for an extended period of time? Would you want or choose to view content with subtitles like this in the future?

The majority of participants who expressed an opinion (12) stated that they could watch dynamic subtitles for longer periods of time, and that they would choose to watch subtitles like this if they had the option. P7 commented that it was less tiring than traditional subtitles:

> 'Reading subtitles can be tiring, so I've got a limited span, I can watch a couple of films and that's about it. I think that this is a lot gentler on the eye.'

Others were unsure about viewing for longer periods, but would like to try. Only P14 and P17 said that they wouldn't want to watch these subtitles again.

POSITIONING SUBTITLES
The overall preferences for each of the four pairs of alternative subtitle positions (version A, in the original position, and B, in the revised position) are summarised in Figure 5. For two subtitles, the participants were split almost equally, while for the other two, they were more likely to prefer the revised subtitle. More interesting than the preferences, however, are the themes that emerged from the discussions about the various placements. These can be classified as follows.

Speaker identification
One of the key factors in people's preferences was positioning the subtitle so that it could be easily associated with the character who was speaking. This was explicitly mentioned by 8 of the participants. For example P19 preferred the revised version of subtitle 24:
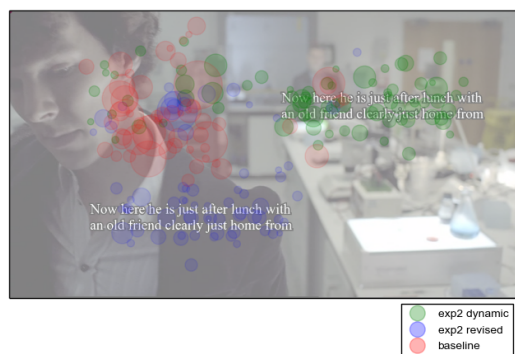
> 'I prefer [B] because you can clearly see that it's attached to Sherlock. It's where he is in the screen — it makes more sense with him being there.'
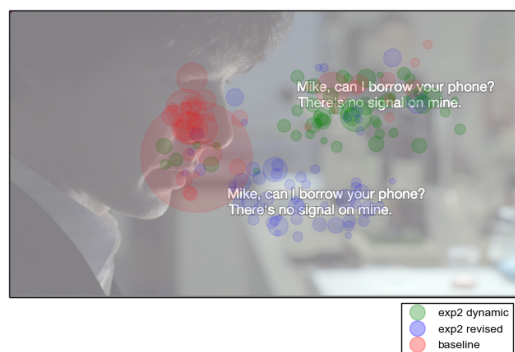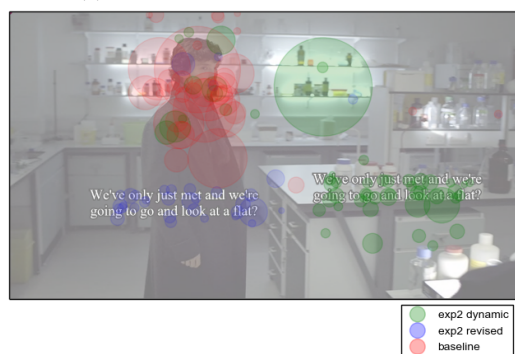
(a) Subtitle 3. Version A is the upper one.
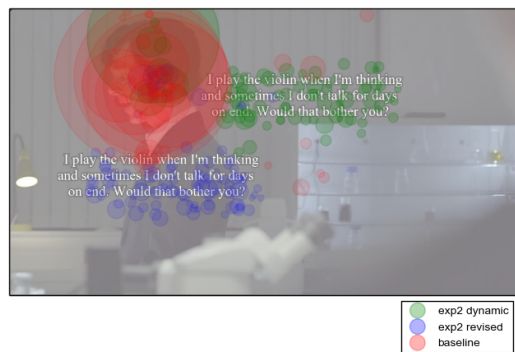
(b) Subtitle 19. Version A is the upper one.

(c) Subtitle 24. Version A is the upper one.

(d) Subtitle 33. Version A is the right one.

Figure 4: Versions A (original) and B (revised) of four subtitles. Overlaid are the fixations made during the lifetime of the subtitle, for people watching with the original subtitle, the revised subtitle, or no subtitle.
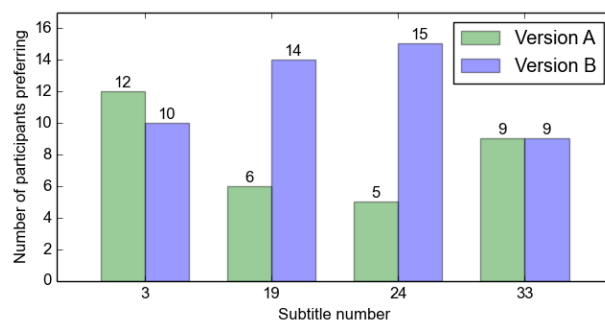

Figure 5: Numbers of participants expressing a preference for the version A (original) or B (revised) of the subtitles.

Participant 10 also preferred this version, as it was clearer to whom the dialogue belonged:

> 'Maybe [B] is better, because it's the speech that is linked with his characters so it makes it clearer that it's him that is speaking'

Five of the participants commented positively on how dynamic subtitles were comic-like or similar to a cartoon, with the text resembling a speech bubble. Some of the appeal of the comic-like presentation is undoubtedly due to the fact that it is an aid to speaker identification; for example, P8 said:

> 'Perfect, it works for me because the face is very important, the face is communication and it's almost like a cartoon bubble that comes out. It's like he's saying it and I'm reading it and it all makes sense.'

However, this style is not necessary for speaker identification (e.g, the subtitle could be placed over the actor's body), and subtitles presented like speech bubbles seemed to have an intrinsic appeal.

Readability
Although most participants said that the subtitles were usable, the qualities of the background were an important consideration when selecting position. When this was mentioned, people either stated that they liked a position because it was particularly clear, or said that they found a position difficult. A plain, dark background was considered good, e.g., subtitle 24B: P0: 'It's the contrast between the text and the colour; P4: 'it's easier to read as its against the dark background', and subtitle 33 P9: 'prefer [B] because of the contrast'. P10 also found subtitle 3A easy to read ('the background is blurred so the words stand out quite well').

In contrast, lighter or more varied backgrounds were more difficult. P20 preferred subtitle 19B, 'possibly because [A] is going across a more varied background'. Similarly, for subtitle 33A, P0 said, 'It's a bit noisy in the background, there's so much other stuff behind the text, and [B] is a lot easier.', while P4 stated, 'I don't mind

the position, but [A is] against a white background so it's a bit more difficult to read.'

Relationship between subtitles and action
Five people felt that the action was, or could be, obscured by the subtitles, particularly if over the actor. Positive comments were made when subtitles were over the background of the scene, e.g,. 'it's in a place where it's just over a blurred bit of background so you're not missing much' (P6 on subtitle 19B), 'that's ok, it's near to the bottom it's not stopping me from seeing the overall scene and the actor isn't obscured' (P14, subtitle 33A). Similarly, some people felt that having the subtitles placed over the actor diminished the experience. For example, three people commented on subtitle 19B:

P9: 'Its like the subtitles are competing with the actor'

P15: I don't like it over his body, it feels like if he starts moving around you don't want to be looking through the writing. You want them to be slightly separate.

P17: 'If it's against them then it's a barrier between you and the actor'

Similarly, P14 commented on subtitle 33: 'I prefer [A], again, this one is across the actor and not giving me a clear view of the action'. This was not an over-riding preference, as these same participants sometimes preferred later subtitles that were placed over the actor (e.g., P15 preferred version B of subtitle 24 'That actually looks quite good down there, which contradicts from my last choice', and P9 preferred 24B and 33B).

Some participants clearly preferred subtitles to be placed over the actor, so that the character and subtitles were co-located. P18 stated about subtitle 19B, 'My gaze is naturally on him so it makes sense for them to be together', and P19 said (of the same subtitle), 'I think that this one is possibly better, in that your attention is focused on the left hand side of the screen'. For the last subtitle, P24 wanted to see the subtitle over Sherlock, even though he was not the character speaking (she did want another cue to indicate who was speaking):

> 'The important thing is to see Sherlock and the action — the director has chosen that shot for a reason. It's the same viewing experience then, it doesn't matter if you look at the subtitles or not, you're still looking at what the director intended.'

The contradictory opinions of the participants highlight the tension between obscuring characters and enabling speaker identification.

Additional Comments
In more general terms, participants P3 and P7 had a preference for subtitles on the right of the screen. Participant 14, who did not like dynamic subtitles, wanted them placed lower on the screen, more like traditional subtitles, where they were less obstructive. Similarly, P17 felt 'for some reason, the higher it is the more it throws itself at you, so I prefer the more subtle one'.

subtitle	original	revised
3	2.1	2.0
19	1.8	2.0
24	2.0	1.4
33	1.5	1.4

Table 1: Difference metric values between the two subtitles and the baseline.

Conversely, P7, P19 and P24 all expressed a preference for subtitles to be placed higher. P10 wasn't keen on the central positioning of 19B, explaining, 'I did photography at college, so I'm thinking about the rule of thirds when I'm going through it'.

There was a slight aversion to subtitles being placed too close to characters, with 7 people commenting on subtitle 19A being too close to Sherlock 'like it's going to hit him in the neck' (P11). P6 and P15 wanted 3B to be placed slightly to the right or lower.

One participant commented on the position of the subtitles in relation to the gaze of the character. P8 preferred subtitle 3 to be closer to the eyes than the mouth: 'For some reason is it to do with the position of the eyes — it's saying to me "look — read". When it's closer to the mouth I want to hear it'.

Participant 24 identified a potential problem with subtitles jumping around the screen leading to 'word tennis'.

Eye-tracking data
A grid representing the gaze pattern for each condition over the lifetime of each subtitle was generated, and the difference between each subtitle and the baseline was calculated. These are presented in Table 1. In subtitles 3 and 33, where users expressed no clear opinion, gaze data suggests that either subtitle led to a similar difference from the baseline. In subtitle 24, the revised subtitle was preferred, and also led to a gaze pattern closer to the baseline. The difference in opinion for subtitle 19 was not matched by any difference in gaze data. The nature of these results means that inferential statistics are not possible, but they are suggestive of a link between a well-positioned subtitle and one which leads to a gaze pattern that is closer to the baseline.

DISCUSSION
The majority of people who watched dynamic subtitles enjoyed the experience, and wanted to try them further. A number were very keen, and would have liked to convert to dynamic subtitles immediately. The main reason for this was that the viewers felt more immersed in the action, and missed less of the video content. Reading the subtitles was a less disjointed experience, and people were more able to follow the action, and pick up non-verbal cues from the actors. The analysis of the eye-tracking data supports this (albeit for people who do not normally use subtitles), finding that people who viewed the clip with dynamic subtitles had gaze patterns that

were more similar to people who viewed without subtitles than those who viewed with traditional subtitles.

The other major benefit was that dynamic subtitles enabled a more explicit link between speaker and text than using colours on traditional subtitles. Most participants were able to connect subtitles to actor even with all text presented in white, although the additional use of colour should be investigated. One of the major use-cases identified by participants was in situations where multiple people were talking, such as panel shows.

A small number of the participants in this experiment did not like this style of subtitle presentation. The data suggested that there was a relationship between a person's reliance on subtitles and their attitudes towards and preferences for dynamic subtitles: those people less reliant tended to be less positive.

The concerns of some participants for the use of dynamic subtitles in social viewing suggest that, ideally, viewers would have the option of whether to have subtitles dynamically placed, or placed in the traditional position. Most people also thought that using dynamic subtitles would not be appropriate for all content; the news was identified by many as a genre for which traditional subtitles were more suitable, due to its relatively static nature.

This experiment has also identified some of the factors that need to be taken into consideration when authoring dynamic subtitles. Identifying the speaker is one of the key benefits, so subtitle position needs to reflect this. Positioning the text as a cartoon speech-bubble would be placed is one option; another is to place the text over the speaker's body. There were divided opinions about this, however, with some people feeling that the subtitle became a barrier in this situation. In either case, the text should not obscure important action, and should not be placed too close to the speaker, particularly to the face. There is perhaps also an argument for placing the subtitles more towards the right of the screen (it is possible that this is because, for subtitles on the right, the viewer starts reading in the centre of the screen, which is likely to be closer to their current gaze). Readability is clearly important, so the effect of the background, particularly if light or varied, needs to be considered. It may be worth exploring the use of font effects to improve readability in such situations.

CONCLUSIONS

In summary, the majority of participants reported that they felt that dynamic subtitles would provide an improvement over traditional subtitles on all aspects of the framework. Some participants (notably those people who were not reliant on the subtitles to follow the dialogue) did not like their first experience of dynamic subtitles, finding them more disruptive than traditionally placed subtitles. It would therefore be desirable for viewers to have the option to revert to traditional subtitles if they, or their viewing companions preferred. For most people, however, it enabled a more immersive experience. They allowed people to relax and enjoy the programme, to follow the dialogue while also picking up more non-verbal cues from the speaker. Eye-tracking data supported this, showing that the gaze patterns of people viewing dynamic subtitles were closer to those of people viewing without subtitles than those of traditional subtitles were. Speaker identification was improved compared to traditional subtitles, although subtitle location may need supplementing with colours in some situations.

> 'With traditional subtitles you feel too focused and can't veg out; with this it makes it a lot easier to relax and watch television.'

As a result of these findings, we suggest that more research effort is directed towards dynamic subtitles. There are areas of the UX of dynamic subtitles that are not fully explored, including their use in social situations. Extending the work of Hong *et al.* [3] into placement algorithms is also necessary, given the time-consuming nature of hand-positioning these subtitles.

REFERENCES

1. Brooks, M., and Armstrong, M. Enhancing subtitles. In *TVX2014* (2014).

2. DTV4All Project. Deliverable D2.5: Final report on mature pilot services. Tech. rep., Brunel University, 2010. http://dea.brunel.ac.uk/dtv4all/.

3. Hong, R., Wang, M., Yuan, X.-T., Xu, M., Jiang, J., Yan, S., and Chua, T.-S. Video accessibility enhancement for hearing-impaired users. *ACM Trans. Multimedia Comput. Commun. Appl. 7S*, 1 (Nov. 2011), 24:1–24:19.

4. O'Brien, H. L., and Toms, E. G. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology 61*, 1 (2010), 50–69.

5. Rashid, R., Aitken, J., and Fels, D. Expressing emotions using animated text captions. In *Computers Helping People with Special Needs*, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds., vol. 4061 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2006, 24–31.

6. Secară, A. 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).