Dynamic Subtitles: the User Experience

Andy Brown
BBC R&D
MediaCity, Salford. UK
andy.brown01@bbc.co.uk

Rhia Jones
BBC R&D
MediaCity, Salford. UK
rhia.jones@bbc.co.uk

Mike Crabb
School of Computing
University of Dundee, UK
michaelcrabb@acm.org

ABSTRACT

Subtitles (closed captions) 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 the testing of such subtitles with hearing-impaired users, and a new analysis of previously collected eye-tracking data. The qualitative data demonstrates that dynamic subtitles can lead to an improved User Experience, although not for all types of subtitle user. The eye-tracking data was analysed to compare the gaze patterns of subtitle users with a baseline of those for people viewing without subtitles. It was found that gaze patterns of people watching dynamic subtitles were closer to the baseline than those of people watching with traditional subtitles. Finally, some of the factors that need to be considered when authoring dynamic subtitles are discussed.

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; H.5.2 Information interfaces and presentation (e.g., HCI): User Interfaces

INTRODUCTION

Traditionally, subtitles are positioned so they are centred at the bottom of the television screen. Guidelines for subtitles (e.g., [1]) 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'. These guidelines also recommend 'horizontal displacement of subtitles in the direction of the appropriate speaker', although this seems not to be widely implemented. In recent years, however, there has been an increase in research experimenting with non-traditional placement of subtitles [6, 17, 7, 15, 16, 8]. There are multiple drivers for this, including creativity [6], but the most common is a desire to help viewers associate subtitles with the correct speaker (e.g., [17, 8]). Jenesema [10] noted that the addition of subtitles 'results in a major change in eye-movement patterns', and eye-tracking studies have estimated the amount of time viewers spend fixating on subtitles as between 10–31.8% [4] and 84% [9]. An argument can be made for authoring subtitles in a way that minimises this disruption, so more time can be spent watching the action. From a User Experience (UX) standpoint there is a desire to deliver subtitle content in a more immersive, engaging, emotive [13], aesthetically pleasing and 'contemporary' [6] 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 [7, 8, 3]. These are known as 'dynamic captioning' [7] or 'dynamically positioned subtitles' [3]; in this paper we use the briefer term 'dynamic subtitles'. Hong *et al.* [7] presented a system that automatically recognised the speaker and used visual analysis of the scene to identify a placement for a subtitle; Hu *et al.* [8] extended this with more sophisticated algorithms. Both performed user studies to capture people's views on the placement, although these were not rich, collecting ratings on a scale of 1–10: Hong *et al.* asked participants to rate 'naturalness' and 'enjoyment', while Hu *et al.* asked their participants to rate 'eyestrain level' and 'overall satisfaction'. Both reported that their systems returned better scores than traditional subtitles, although it should be noted that participants in [8] were not habitual subtitle users, a factor which has been found to influence peripheral vision skills [2] and how attention is allocated [5]. Brooks and Armstrong's initial work [3] found that people spent less time reading dynamic subtitles, and more time looking at the drama, but did not explore the UX.

We wish to understand the user experience of dynamic subtitles in more detail, and hypothesise that they could provide an improved experience, making it easier to follow both the subtitles and the video content. This work seeks to explore that hypothesis, by extending the initial study of Brooks and Armstrong [3] in three ways:

- Additional eye-tracking data is collected, and the combined data analysed to discover how much gaze pat-

TVX 2015, June 3–5, 2015, Brussels, Belgium.

ACM 978-1-4503-3526-3/15/06 ...$15.00.
http://dx.doi.org/10.1145/2745197.2745204

terns differed between subtitled and non-subtitled content.

- Habitual subtitle users were asked to view an example of content with dynamic subtitles, and qualitative data was captured about their attitudes towards it.

- The question of what factors determine whether a subtitle is well or poorly placed was investigated.

PREVIOUS EXPERIMENT

This research uses data from Brooks and Armstrong [3], which is combined with new data and analysed in a novel way. This section summarises their study.

4 clips were taken from 3 episodes of the BBC 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).

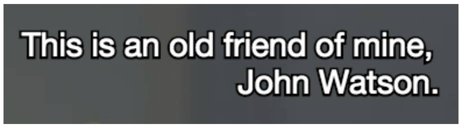
24 participants (native English speakers, who did not understand French; participants were not habitual subtitle users) watched the clips, in the same order, on a television in a 'living room' lab. The clips were first presented in one of the 4 subtitle/language combinations, counterbalanced so that 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 people. The gaze of each participant was recorded using a Tobii X-120 eye-tracker. An initial analysis of the data, in which an area of interest was defined for each subtitle (420 across 4 clips, under 2 conditions), indicated that people spent less time reading subtitles, and more time looking at the drama when using dynamic subtitles than traditional subtitles.

METHOD

The second experiment was designed to collect additional baseline data to combine with that collected in experiment 1, and to capture qualitative data on the User Experience of dynamic subtitles from people who habitually used subtitles as an access service.

Participants

26 participants were recruited for inclusion in this study. Recruitment was performed by an external agency, and participants were recruited who: regularly use the internet to consume news and current affairs information; use subtitles at home to watch TV with the sound on, and; use subtitles on a daily basis. Participants were aged between 22–67 ($\overline{x} = 47.2, \sigma = 13.6$). A mix of gender (7 male, 19 female) and socio-cultural/economic backgrounds was used. In addition, 8 people were recruited (convenience sample; 5 male, 3 female, aged between 21 and 55) to watch the clip without subtitles. As in experiment 1, these people did not normally use subtitles.



This is an old friend of mine,
John Watson.

Figure 1: The text used to present the subtitles.

Stimulus

Participants were shown a 1 minute 50 second clip from the TV drama "Sherlock" (Series 1, Episode 1). 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; Sherlock deduces that Watson has just left the army and is, like himself, 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 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 2.

A Framework for Qualitative Data Capture

User experience is a highly subjective field, focusing on the potential benefits that a user may derive from a product [11]. To be of use to the scientific community, however, the measurement of UX needs to be meaningful and reliable [12]. A standard way of ensuring reliability is to develop a framework that identifies the important components of the UX so that each can be measured. A review of the literature failed to identify such a framework for subtitles, so a new framework is proposed here[1].

The structure of the framework was inspired by [14], while the components were developed from an analysis of the existing literature on the UX of subtitles. These components are described below.

Attention is awareness of what is going on in relation to the subtitled video content. Users with high levels

[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.

of attention would be focused heavily on the video content, while users with low levels would not.

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.

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.

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

Design and Procedure

The session was run in the BBC R&D usability lab in Dock House, Salford which is set up as a living room, and has 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, designed to explore how people felt about viewing content with dynamic subtitles. The questions were aligned to the framework, above, and are detailed in the results, below.

Supplementing baseline data

To supplement baseline data from [3], participants were introduced to the study and seated in front of the eye-tracker (in the same configuration as above). The eye tracker was calibrated, and participants were asked to watch the clip as they would normally watch television.

EYE-TRACKING DATA ANALYSIS

The hypothesis being tested is that dynamic subtitles allow gaze patterns that are closer to those of viewers watching without subtitles, but it is not known, a-priori, where those viewers will fixate. Consequently, while it is possible to define areas of interest 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. Making this calculation twice, for traditional and dynamic subtitles, shows which condition resulted in the smaller change of gaze pattern. A smaller change indicates that the gaze patterns were closer to those for the baseline, suggesting that people's experience of the video content is less disrupted by reading the subtitles.

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 any participants looked away, or the eye-tracker failed to record some data. A fixation that overlaps time slices will contribute its duration to each slice proportionately.

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 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.

QUALITATIVE RESULTS

The qualitative data comprises three parts: the first impressions of participants; their overall views after having performed the positioning exercise, and; their responses to a set of questions aligned to the framework (above).

In summary, 5 participants did not like dynamic subtitles (P2, P9, P14, P17, P19), 8 were broadly positive (P0, P3, P11, P15, P20, P21, P22, P23), and 12 were very keen on the idea. Interestingly, the 3 participants who most disliked the dynamic subtitles 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. 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:

> '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.' (P5)

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

> 'Yeah, 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. . .

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.

General Comments

After capturing the initial thoughts, participants were asked to comment on 4 pairs of alternative dynamic subtitle positions, then asked: *'What do you think are the advantages / disadvantages of having subtitles positioned in different places on the screen?'*

The two themes of being able to identify speaker more easily, and of missing less of the video 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), and:

> 'They seem really well integrated and its easy to switch between the subtitles and the visuals without feeling like it was disjointed.' (P6)

> 'It's almost cinema like — you have that feel of being enveloped of it' (P8)

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.'

Semi-structured Interview

The questions that formed the basis for the discussion, and the responses to them, are summarised below.

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) thought that they would detract from the aesthetics of other viewers, 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... 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?

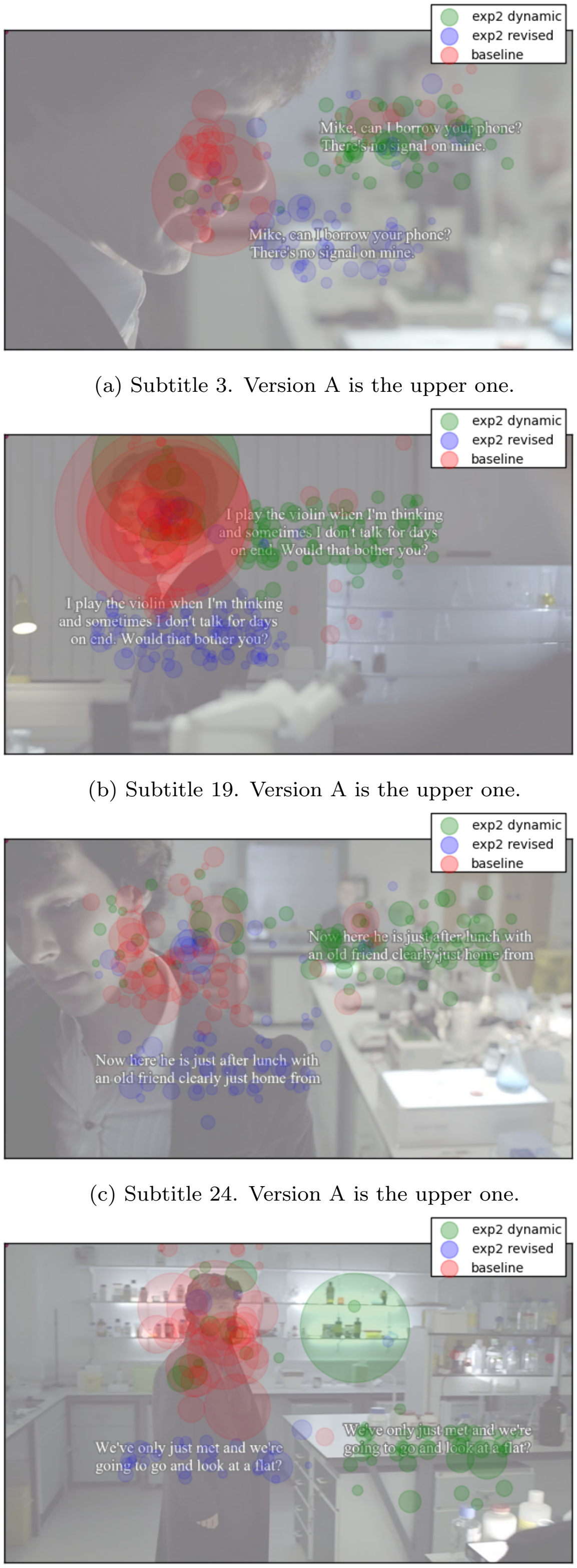
The majority of participants who expressed an opinion (12) stated that they could watch dynamic subtitles for longer periods of time, and that they would choose to watch subtitles like this if they had the option. P7 commented that it was less tiring than traditional subtitles:

> '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 3. 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.



(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 2: 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.

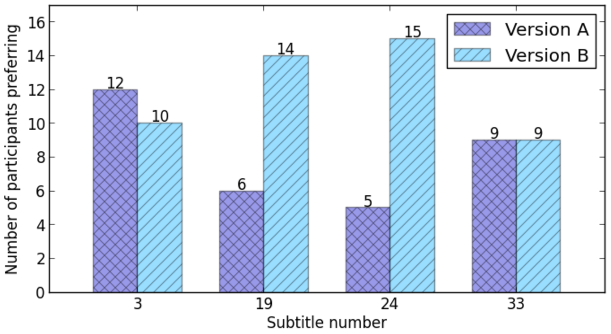


Figure 3: Numbers of participants expressing a preference for the version A (original) or B (revised) of the subtitles.

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 and P10 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.' (P19)
>
> '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' (P10)

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. Although clearly related to speaker identification, the cartoon style is not necessary for it (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., P4 said of subtitle 24B: 'it's easier to read as its against the dark background'. 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. 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'.

Obscuring the action

Five people felt that the action was, or could be, obscured, 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). Similarly, some people felt that having the subtitles placed over the actor diminished the experience, blocking their view. For example, comments on subtitle 19B included:

P9: 'I don't like how its over him. . . Its like the subtitles are competing with the actor in the scene.'

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.

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).

On the other hand, 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, because that placed the subtitle close to the important action:

> '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.'

General positioning

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, where they were less obstructive. P17 felt 'for some reason, the higher it is the more it throws itself at you, so I prefer the more subtle one'. 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.

Eye-tracking data

The eye-tracking data was inconclusive when comparing the revised subtitles with the original ones. 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.

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.

These are presented in Table 1; the only difference of any size was for subtitle 24, for which the revised subtitle was closer to the baseline.

EYE-TRACKING RESULTS

Before full analysis of the data, the difference metric was tested. This was done by comparing the revised and original subtitles; as expected, difference values were low (median difference of 0.9) except when the subtitle positions differed (peaks of 2-3). Having tested the metric, the additional baseline data was combined with the baseline data from Brooks and Armstrong's original work [3], giving usable gaze data for 5 participants watching with each of the traditional and dynamic subtitles (French audio), and 11 participants watching without subtitles. The difference metric was calculated to compare each subtitle condition with the baseline.

Figure 4 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.

Figure 5 summarises the results, showing the difference values for the four conditions: experiment 1 traditional and dynamic subtitles, and experiment 2 original and revised subtitles. 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).

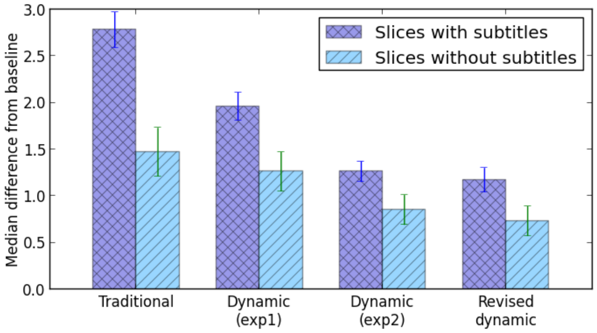


Figure 5: Median difference values for the 1s slices for the different conditions. These are split into values for slices in which subtitles were visible, and those in which there were none.

The results from the second experiment show smaller differences, with no significant difference between the revised and original subtitles. Interestingly, the gaze patterns of viewers watching dynamic subtitles were less different from the baseline in the second experiment than the first. There are two factors that might account for this. First, the viewers in the 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 may revert their gaze to the baseline more quickly than the participants of experiment 1.

CONCLUSIONS

The majority of people who watched dynamic subtitles enjoyed the experience, and wanted to try them further. A number of participants were very keen, and would have liked to convert to dynamic subtitles immediately.

> “This is going to spoil subtitles for me now” (P16)

The main reason was that it meant that the viewers were 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 new analysis of the eye-tracking data from the previous experiment 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.

> ‘I wouldn't have caught a lot of the small social cues if I were watching this with traditional subtitles’ (P16)

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

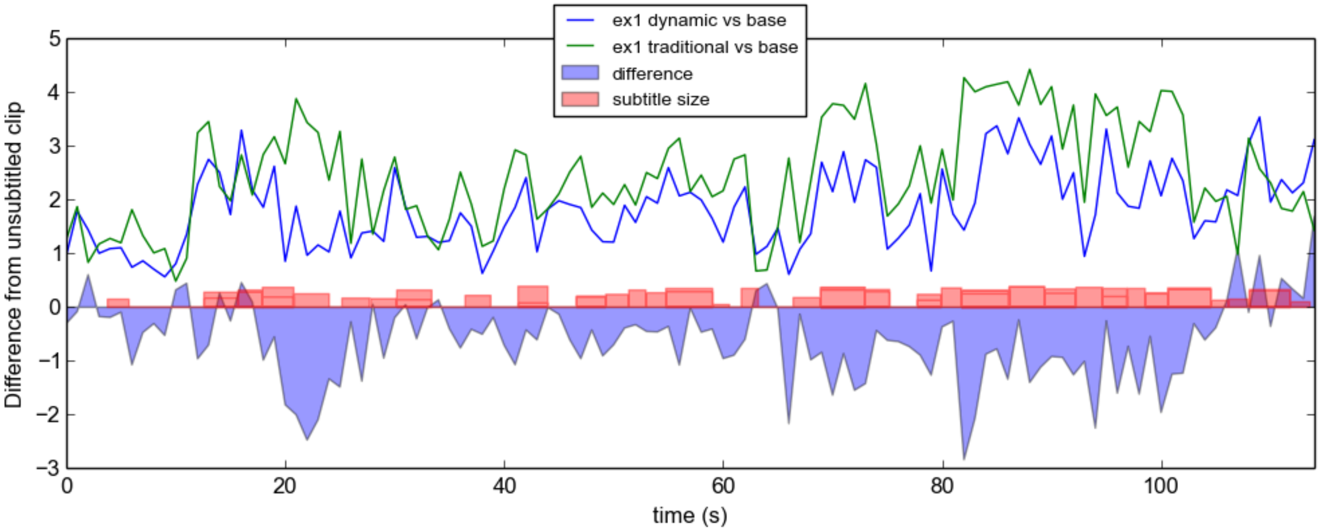


Figure 4: A comparison of how much gaze patterns in the traditional subtitle and dynamic subtitle conditions differed from the baseline. The differences between traditional subtitles and the baseline are shown in green; those between dynamic subtitles and the baseline are in blue. The filled line indicates 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.

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.

> 'I think this would have a huge benefit for a lot of people to make more sense of conversations' (P10)

A small number of the participants in this experiment did not like this style of subtitle presentation — 2 were ambivalent and would prefer to use the subtitles they were used to, while 3 really disliked dynamic positioning. Interestingly, these participants were ones who did not totally rely on subtitles. In contrast, those who were most enthusiastic about the subtitles tended to be those who relied more on the subtitles as an access service.

Two of those people who liked dynamic subtitles themselves expressed concern that co-watchers (who did not need subtitles) would find them more disruptive. This suggests that the ideal solution would be to give viewers the option of whether to have subtitles dynamically placed, or placed in the traditional position at the bottom of the screen. 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. It should be noted, however, that this tended to be an opinion found among those people who were against the idea in general. 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 could be hypothesised 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.

While the participants in this study were positive about the use of dynamic subtitles for Sherlock, and expressed a wish to use them on other content, the conclusions should not be extrapolated too far. The scene contained a maximum of 3 characters on screen at once, and shot-changes were not as frequent as they might be, e.g., in action movies. These factors may well influence the UX of dynamic subtitles, and should be explored further.

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. 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 cant veg out on television, with this it makes it a lot easier to relax and watch television.' (P10)

ACKNOWLEDGEMENTS

The authors would like to thank those who participated in this experiment. In addition, Mike Crabb is support by RCUK Digital Economy Research Hub EP/G066019/1 SIDE: Social Inclusion through the Digital Economy.

ADDITIONAL AUTHORS

James Sandford (BBC R&D, email: james.sandford@bbc.co.uk), Matthew Brooks (BBC R&D, email: matthew.brooks@bbc.co.uk), Mike Armstrong (BBC R&D, email: mike.armstrong@bbc.co.uk) and Caroline Jay (School of Computer Science, University of Manchester, UK, email: caroline.jay@manchester.ac.uk),

REFERENCES

1. Baker, R. G., Lambourne, A. D., Rowston, G., Authority, I. B., Association, I. T. C., et al. *Handbook for Television Subtitlers*. Independent Broadcasting Authority, 1982, 13.
2. Bosworth, R. G., and Dobkins, K. R. The effects of spatial attention on motion processing in deaf signers, hearing signers, and hearing nonsigners. *Brain and Cognition 49*, 1 (2002), 152 – 169.
3. Brooks, M., and Armstrong, M. Enhancing subtitles. In *TVX2014* (2014).
4. Chapdelaine, C., Gouaillier, V., Beaulieu, M., and Gagnon, L. Improving video captioning for deaf and hearing-impaired people based on eye movement and attention overload. *Proc. SPIE 6492* (2007), 64921K–64921K–11.
5. D'Ydewalle, G.and Gielen, I. Attention allocation with overlapping sound, image, and text. In *Eye Movements and Visual Cognition*, K. Rayner, Ed., Springer Series in Neuropsychology. Springer New York, 1992, 415–427.
6. Foerster, A. Towards a creative approach in subtitling: a case study. In *New Insights into Audiovisual Translation and Media Accessibility*, J. Cintas, A. Matamala, and J. Neves, Eds. Rodopi, New York, NY, 2010, 81–98.
7. 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.
8. Hu, Y., Kautz, J., Yu, Y., and Wang, W. Speaker-following video subtitles. *ACM Trans. Multimedia Comput. Commun. Appl. 11*, 2 (Jan. 2015), 32:1–32:17.
9. Jensema, C. J., Danturthi, R. S., and Burch, R. Time spent viewing captions on television programs. *American annals of the deaf 145*, 5 (2000), 464–468.
10. Jensema, C. J., El Sharkawy, S., Danturthi, R. S., Burch, R., and Hsu, D. Eye movement patterns of captioned television viewers. *American Annals of the Deaf 145*, 3 (2000), 275–285.
11. Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P., and Kort, J. Understanding, scoping and defining user experience: A survey approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, ACM (New York, NY, USA, 2009), 719–728.
12. Law, E. L.-C., and van Schaik, P. Modelling user experience–an agenda for research and practice. *Interacting with computers 22*, 5 (2010), 313–322.
13. Lee, D., Fels, D., and Udo, J. Emotive captioning. *Comput. Entertain. 5*, 2 (Apr. 2007).
14. 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.
15. 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.
16. 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).
17. Vy, Q., and Fels, D. Using placement and name for speaker identification in captioning. In *Computers Helping People with Special Needs*, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds., vol. 6179 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2010, 247–254.