Visual attention as a dimension of QoE: Subtitles in UHD videos

Toinon Vigier, Yoann Baveye, Josselin Rousseau, Patrick Le Callet

Visual Attention as a dimension of QoE: Subtitles in UHD videos

Toinon Vigier, Yoann Baveye, Josselin Rousseau and Patrick Le Callet
IRCCyN UMR CNRS 6597, Université de Nantes, France
Email: {toinon.vigier, yoann.baveye, josselin.rousseau, patrick.lecallet}@univ-nantes.fr

***Abstract*—With the ever-growing availability of multimedia content produced, broadcast and consumed worldwide, subtitling is becoming an essential service to quickly share understandable content. Simultaneously, the increased resolution of the ultra high definition (UHD) standard comes with wider screens and new viewing conditions. Services as the display of subtitles thus require adaptation to better fit the new induced viewing visual angle. This paper aims at evaluating quality of experience of subtitled movies in UHD to propose guidelines for the appearance of subtitles. From an eye-tracking experiment conducted on 68 observers and 30 video sequences, viewing behavior and visual saliency are analyzed with and without subtitles and for different subtitle styles. Various metrics based on eye-tracking data, such as the Reading Index for Dynamic Texts (RIDT), are computed to objectively measure the ease of reading and subtitle disturbance. The results mainly show that doubling the visual angle of subtitles from HD to UHD guarantees subtitle readability without compromising the enjoyment of the video content.**

I. Introduction

Nowadays, more and more multimedia content is produced, broadcast and consumed worldwide. Therefore, subtitling is an essential service to quickly share understandable content worldwide. For instance, in non-English speaking countries, popular movies and TV series are often available in original language with subtitles long before the dubbed version. It is thus required to take into account style settings of subtitles according to display and device in quality of experience. More specifically, the question of subtitle style and size is raised out with the emergence of ultra high definition television (UHD TV) to replace the current high definition television (HD TV) system. Indeed, UHD TV standard defines new video technologies such as increasing resolution from HD (1920×1080) to 4K (3840×2160) or 8K (7680×4320). Thus, the emergence of UHD potentially provides a better immersion of the user thanks to a wider visual angle with appropriate larger screens [1]. These new viewing conditions must be evaluated in terms of quality of experience of multimedia systems and services such as subtitling in video. In order to accurately measure and quantify viewing behavior and reading of subtitles in videos, visual attention analyses from eye-tracking experiments can be carried out. Several studies were conducted to evaluate speed and ease of reading of subtitles. Most of them were often focused on the impact on subtitle reading of subtitle speed (e.g. [2], [3]), style [4], language soundtrack (e.g. [5], [6]) and personal characteristics of observers such as language, disability, or frequency of subtitle usage (e.g. [7], [4], [6]). Some recommendations have been published by TV channels (e.g. [8], [9]) and some style characteristics have been harmonized:

- Subtitles are displayed on two lines of text;
- Suitable fonts are from the sans-serif typeface, such as Helvetica, Arial or Verdana;
- Characters are approximately between 30 and 40 per line of text.

The current increase of the size of TV screens question these recommendations, in particular subtitle size and placement (e.g. [10], [11]). For example, Brown et al. proposed a dynamic subtitle placement in which subtitles are positioned to minimize the distance from the area of interest to subtitles [10]. Only a few research studies are focused on subtitle size. Lee et al. evaluated the preferred subtitle font size according to viewing distance [12]. They deduced an optimal viewing angle for subtitles around 0.7°. However, results cannot be directly used in Latin script since this study was done in Chinese.

Based on this review, there is a need for studies on subtitle appearance and size in UHD viewing conditions. In this paper, an experiment is described in order to evaluate subtitle readability, reading comfort and intrusion of subtitles thanks to eye-tracking measures. The final objective is to provide guidelines on subtitle insertion in UHD. The paper is organized as follow. Section II presents the experiment design while Section III lists the data and metrics used to analyze the results discussed in Section IV. Finally, conclusion and future work end the paper in Section V.

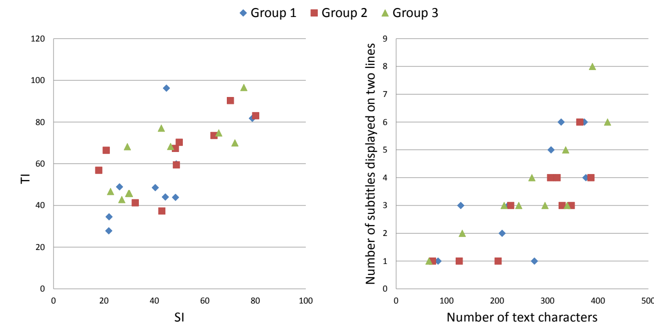
II. Experimental design

A. Stimuli selection

The stimuli used in this experiment are based on 30 Source Video Sequences (SRCs) in English extracted from seven professional Hollywood movies. The 30 segments have been manually extracted in order to select semantically coherent video excerpts containing at least one subtitle composed of two lines of text. The SRCs last 30 seconds which seems long enough for observers to make sense of the subtitled video sequences. For each SRC, corresponding raw Blu-ray subtitles, written in French, have also been extracted. Thus, issues about timing, synchronization, translation or cuts are avoided and not studied in this work. For all SRCs, subtitles are arbitrarily positioned at 8.3% ($= 1/12 \times H_{video}$) from the bottom of the video. The font of all subtitles is Arial and subtitles are written in white with a black outline around the text. The sequences are originally extracted at HD resolution (1,920 $\times$ 1,080) and

978-1-5090-0354-9/16/$31.00 ©2016 IEEE

upscaled at UHD resolution (3,840 × 2,160) using Lanczos-3 algorithm [13], since it is one of the best upsampling algorithm in term of preference of experience [14]. The 30 SRCs are split in three groups. The distributions of the SRCs inside each group are comparable to each other in terms of spatial and temporal information (SI, TI as described in ITU-T P.910 [15]), total number of text characters in the SRC and number of subtitles displayed on two lines in the SRC (Figure 1). Thus, 30 SRCs divided in three SRC groups, denoted SRCG1, SRCG2 and SRCG3, are evaluated in this experiment. In average, the subtitles corresponding to these 30 SRCs are composed of 30 characters per subtitle (6.5 words) and are displayed onscreen for 2.63 seconds at a rate of 148 words per minute. Subtitles are displayed on 62% of the video frames and 48% of them are displayed on two lines.



(a) Video frame-based characteristics (b) Subtitle-based characteristics

Fig. 1. Groups generated for the 30 SRCs based on their (a) video frame characteristics (SI/TI) and (b) subtitle-based characteristics (total number of text characters and number of subtitles displayed on two lines).

B. *Hypothetical reference circuit*

A Hypothetical Reference Circuit (HRC) is a particular set of processing operations. In this experiment, three factors are used to generate six HRCs: the resolution of the SRC (HD/UHD), the size of subtitles, and the maximum number of lines used to display subtitles. The six HRCs are listed in Table I and illustrated in Figure 2. As stated in Section I, no standard settings exist for the display of media subtitles. The characteristics of the subtitles for HD videos used in HRC2 are based on the mean of the characteristics of the subtitles used in this work, extracted from commercial DVDs. Thus, the font size of text characters in HRC2 corresponds to a vertical viewing angle of 1.03° for one line of text in subtitles. For HRC4 and HRC6 composed of UHD videos, the vertical viewing angle for one line of text is unchanged. Consequently, for both HRCs, the vertical ratio for one line of text in term of screen height is half the size of the one in HRC2 (in HD). For HRC5, also composed of UHD videos, the font size of text characters in subtitles is twice the size of text characters in previously mentioned HRCs to preserve the vertical ratio for one line of text in term of screen height compared to HRC2 (in HD). Because UHD screens are wider than HD screens, the display of subtitles on one line only is also tested in HRC6. Finally, sequences without subtitles are inserted in the test in HRC1 (HD) and HRC3 (UHD), in order to compare the impact of subtitles on visual attention.

C. *Experimental setup*

The 4K display used in the experiment is a Panasonic TX-L65WT600E with a 65-inch screen size. In this study, we decided to set the viewing distance for UHD videos to 1.5H (126cm), with H the height of the screened video, as recommended in ITU-R BT.1769 [16]. The viewing distance was set to 3H for HD videos. Both viewing distances are strictly the same because HD videos were shown at the center of the 4K display as illustrated in Figure 3. The soundtrack of the videos sequences was played using external speakers and the volume was set at a comfortable level. The test environment is set as a standard test condition according to ITU-R BT.710 [17]. Because of the wider viewing angle in UHD, observers may move their head and eye-tracking systems may not be accurate enough at the borders of the screen. Therefore, we developed and used a new eye-tracking setup, named Eye Head Tracking (EHT), which is a combination of the mobile SMI eye-tracking glasses and the OptiTrack ARENA head-tracker. Gaze data are recorded at a frequency of 30Hz in binocular mode [18].

D. *Experimental protocol*

Observers are split in three groups with a similar distribution of age, gender and subtitle viewing experience. We restrict the number of occurrences of the same SRC for the same observer to two. During the experiment, each group of observers watches Processed Video Sequences (PVS) composed of the videos from the SRC groups treated with the HRCs indicated in Table II. In this table, “1” stands for the first group of observers, “2” stands for the second group of observers and “3” stands for the third group of observers. SRCG1, SRCG2 and SRCG3 refer to the three groups of SRCs mentioned in Section II-A. Thus, each observer watches 60 PVS during this experiment. For a given observer, the PVS corresponding to his assigned group of observers were selected. The 10 PVS generated using the same HRC were displayed consecutively in a random order. The order of the six groups of 10 PVS corresponding to the six HRCs was displayed randomly. Before each group of 10 PVS generated using a given HRC, the same two 30 seconds long training video sequences treated using the considered HRC were displayed in order to accustom observers to the new subtitles parameters. Thus, recorded data is less sensitive to the adaptation delay and enables the evaluation of the reading easiness. The PVS were separated by two seconds breaks during which a black frame was displayed. For example, a participant assigned to the first group of observers could have watched the following sequence of PVS: the two test videos treated with HRC4, then the 10 SRCs in SRCG1 treated with HRC4 displayed in a random order, then the two test videos treated with HRC6, then the 10 SRCs in SRCG3 treated with HRC6 also displayed in a random order, etc.

Before the experimental session, participants had to sign a consent form and instructions were given. The experiment started after the calibration of the eye-tracking system. Eye-tracking data was recorded in free-viewing conditions. To avoid visual fatigue, the viewers were asked to take a five minutes

[1] The values are computed considering a mean number of 35 text characters per line of text when subtitles are displayed using a maximum of two lines of text and 70 characters otherwise (HRC 6).

TABLE I. THE HYPOTHETICAL REFERENCE CIRCUIT FOR THE SUBJECTIVE EXPERIMENT

HRC	Viewing distance	Presence of subtitles	Font size (px)	Maximum number of lines	Vertical viewing angle for one line (°)	Vertical ratio for one line (% screen height)	Average horizontal viewing angle of total subtitle area[1] (°)	Average horizontal ratio of total subtitle area (% screen width)
1	3H (HD)	No	-	-	-	-	-	-
2	3H (HD)	Yes	58	2	1.03°	10.7%	20.7°	61.7%
3	1.5H (UHD)	No	-	-	-	-	-	-
4	1.5H (UHD)	Yes	58	2	1.03°	5.37%	20.7°	30.8%
5	1.5H (UHD)	Yes	116	2	2.05°	10.7%	40.2°	61.7%
6	1.5H (UHD)	Yes	58	1	1.03°	5.37%	40.2°	61.7%

TABLE II. PROCESSED VIDEO SEQUENCES ASSIGNED TO THE THREE GROUPS OF OBSERVERS

	HRC1	HRC2	HRC3	HRC4	HRC5	HRC6
SRCG1	1	2	3	1	2	3
SRCG2	3	1	2	3	1	2
SRCG3	2	3	1	2	3	1

break after half of the test samples. The total duration of the test was approximately 45 minutes.

E. Participants

68 remunerated viewers participated in this subjective experiment (46 female and 22 male), aged between 19 and 51 (mean$= 23.16 \pm 5.34SD$). 65 participants are French, while the remaining three participants are Moroccan, Greek, and Russian. Participants had different educational backgrounds. Participants also assessed their English language proficiency level and the frequency of subtitled movies viewing (Figure 4). All are non-expert in subjective experiment, image processing or 4K related fields. All participants have either normal or corrected-to-normal visual acuity. Correct visual acuity was assured prior to this experiment. For near vision test, a Parinaud chart was used (French equivalent of Jaeger chart) while for the far vision test, a Monoyer chart was used. Testing both may seem particularly useful for such a subjective experiments with subtitles and UHD videos as the viewing distance is only 1.5H. Ishihara color plates were used to test color vision. All of the 68 viewers passed the pre-experiment vision check.

III. DATA AND METRICS

In this section, data and metrics, calculated from raw gaze data and used for the analysis of results are presented.

A. Fixations and saccades

Fixation points and saccades reflect reading behavior and can be used to evaluate the ease of reading [19]. They are extracted from the raw gaze data following the method explained in [20]. However, because duration of fixations and length of saccades can be shorter when viewers read subtitles, different parameters were chosen to better fit with this experiment. The fixation velocity maximum threshold is set at 20°/s. The minimum time between separate fixations is set at 55 ms and the minimum visual angle between separate fixations is set at 0.33°. Finally, the minimum fixation duration is set at 50 ms. Extracted fixations and saccades are illustrated in Figure 5.

B. Subtitle area-based metrics

For each frame with subtitles, the subtitle area is defined as a box with dimensions equal to the size of the current subtitle with a raised outer border of 2° in visual angle. From raw gaze data, the number of fixations per word and the time spent in the subtitle area is computed for each subtitle and each observer. Then, for each video, the reading duration is defined as the percentage of the time spent in the subtitle area over the video, when the subtitles are displayed. The length of the saccades and the number of backward saccades in the subtitle area are also computed. A backward saccade is defined as an horizontal saccade oriented in the opposite direction of reading. The metrics mentioned in this section are averaged for each video. These values are used for statistical analyses.

C. RIDT

To evaluate the ease of reading, the Reading Index for Dynamic Texts (RIDT) is computed, representing the degree to which each subtitle is read or, in other words, the quantification of reading [21]. For video v, with participant p viewing subtitle s, the RIDT_{vps} is defined as:

$$\text{RIDT}_{vps} = \frac{\text{number of unique fixations for } p \text{ in } s}{\text{number of standard words in } s} \times \frac{\text{average forward saccade length for } p \text{ in } s}{\text{standard word length for } v}$$

In the following, RIDT_v, which is the mean of RIDT_{vps} over all observers and all subtitles for each video, are used for statistical tests. Moreover, RIDT values mentioned in Section IV represent the mean of RIDT_v over all the videos for the HRCs.

D. Saliency metrics

To evaluate the intrusion of subtitles in videos, visual saliency for the different HRCs is compared. Fixation density

(a) HRC1 (b) HRC2

(c) HRC3 (d) HRC4

(e) HRC5 (f) HRC6

Fig. 2. Illustration of the six HRCs using a screenshot from Alice in Wonderland, Copyright the Walt Disney company, 2010.

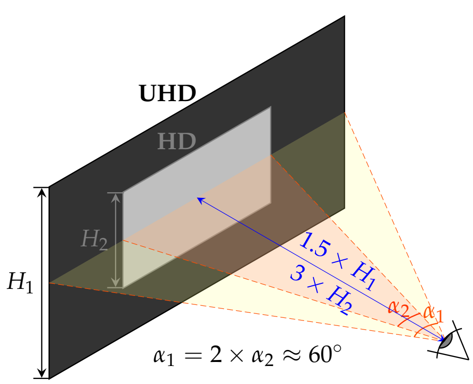
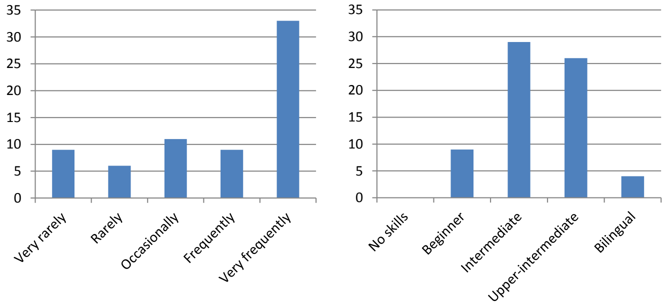


Fig. 3. Viewing distance for UHD and HD videos.

maps are calculated, for each frame, on fixations with a bidimensional gaussian function with $\sigma = 1°$ as recommended in [22]. In order to compare saliency only for the original video content without subtitles, fixation density maps are cropped to discard the bottom of the frame corresponding to a height of 469 pixels in UHD. The fixation density maps of HRCs with and without subtitles (separately in HD and UHD) are then compared by calculating Pearson's correlation coefficient (Cp) and Kullback-Leibler divergence (KLD) as recommended in [22]. The continuous fixation density maps are aggregated per video, thresholded and processed as a binary classifier to keep the top 2%, 5% and 10% salient pixels of the thresholded maps, labeled as fixated. For a given threshold, the binary maps with and without subtitles are compared to compute the precision,



(a) How often are you watching subtitled movies?

(b) What is your English language proficiency level?

Fig. 4. Repartition of the participants' answers to the questions asked after the experiment to measure their English language proficiency level and the frequency of subtitled movies viewing.

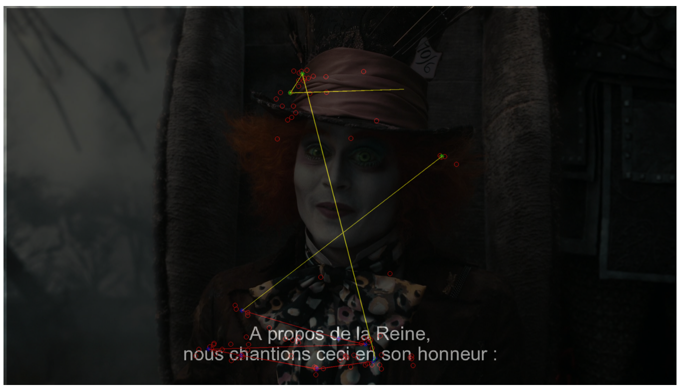


Fig. 5. Fixation points and saccades extracted from the raw gaze data. Red lines illustrate the saccades in the subtitle area.

accuracy and Matthews correlation coefficient (MCC) from the confusion matrices. In this context, the precision corresponds to the percentage of the pixels labeled as fixated without subtitles also classified as fixated in the map with subtitles (for a given HRC). The accuracy and the MCC represent the similarity between binary maps with and without subtitles.

IV. Results and Discussion

In this section, the results are discussed based on the non-parametric Kruskal-Wallis test and Tukey's multiple comparison test, which are used to determine which distribution is significantly different from the others.

A. Impact of subtitle display on reading behavior

The Tukey's test revealed that the reading duration for HRC5 is not significantly different from the one for HRC4 ($p = 0.099$) and HRC6 ($p = 0.677$). Thus the size of the subtitles do not influence the reading duration significantly. However, the Tukey's test also revealed that observers read subtitles displayed on one line (HRC6) significantly faster ($p < .01$) than subtitles displayed on two lines (HRC4). Indeed, the observers spend 29.2% of the time in the subtitle area for HRC4 while this value equals 27.5% for HRC6. It may be due to the fact that the absence of line break for HRC6, thus reducing the amount of backward saccades occurring in the subtitle area.

The RIDT is significantly different for the four HRCs with subtitles ($\chi^2(3) = 282.3; p < .001$). Actually, the Tukey's test revealed that significantly higher RIDT values are obtained for the two HRCs with the highest vertical ratio in terms of screen height (0.68 for HRC2 and 0.66 for HRC5). RIDT values for HRC4 and HRC6 are respectively 0.64 and 0.54. Thus, for subtitles displayed on UHD screens, keeping the same vertical ratio as the one on HD screen while doubling the height of the subtitles seems the best solution to maintain the ease of reading. However, the length of the saccades for subtitles displayed with a vertical viewing angle of 2.05° (HRC5) is almost twice the lengths of the saccades for the subtitles of the other HRCs. It seems consistent with that fact that saccades span the same amount of characters [23]. However, the impact of the amplitude of the saccades over the eye fatigue is uncertain and have to be studied in the future. Contrary to the other HRCs, displaying subtitles on one line of text (HRC6) reduces the RIDT for long subtitles (composed of more than the mean of the number of characters of all subtitles, i.e., 30 characters) compared to short subtitles ($\chi^2(1) = 20.03; p < .001$). Indeed, compared with HRC4, the average length of the saccades in the subtitle area is moderately longer for HRC6 ($231.06px \pm 73.62SD$ for HRC4 and $284.67px \pm 104.33SD$ for HRC6) while the number of fixations in the subtitle area per word is roughly the same (0.63 for HRC4 and 0.60 for HRC6). This result extends the standardized use of two lines of text to display subtitles for UHD.

To summarize, displaying subtitles on one line of text for UHD screens (HRC6) is the worst solution in term of ease of read, in particular for long subtitles. The HRC5, preserving the vertical ratio of subtitles in UHD in term of screen height compared to HD, is the optimal HRC to facilitate subtitle reading. However, the subtitle intelligibility cannot be directly inferred from the measures used in this analysis. Further work is required to address this limitation.

B. Impact of subtitle display on visual saliency

In this section, saliency metrics (see Section III-D) are used to evaluate the intrusion of subtitles in videos. Both HRC5 and HRC6 are significantly closer to HRC3 than HRC4 in terms of Cp ($p < .001$) and KLD ($p < .001$). This tendency is also confirmed by the precision, accuracy and MCC metrics based on the confusion matrix from thresholded binary maps. Respective mean values are indicated in Table III. For all these metrics, the HRC5, which corresponds to the same vertical ratio of subtitles as in HD, seems the least disruptive HRC in term of saliency. However, these metrics mainly evaluate similarities between visual saliency with and without subtitles. Other assessment methodologies are needed to study the subjective experience of the video content (intelligibility, immersiveness, etc.).

C. Other results

Beyond the different HRCS, subjective characteristics also influence the way viewers read subtitles. The time before the first fixation in the subtitle area is significantly different ($p < .001$) from HD (0.35 sec for HRC2) to UHD (close to 0.48 sec for HRCs 4 to 6), whereas there is no significant difference between UHD HRCs. This result shows that the subtitle placement is questionable for wider screens as mentioned in Section I. However, this slight augmentation of

TABLE III. SIMILARITY COEFFICIENTS BETWEEN UHD FIXATION DENSITY MAPS WITH AND WITHOUT SUBTITLES

	Cp	**KLD**	**Precision**			**Accuracy**			**MCC**		
			2%	5%	10%	2%	5%	10%	2%	5%	10%
HRC4	0.568	3.188	0.63	0.72	0.76	0.99	0.97	0.95	0.62	0.70	0.74
HRC5	0.601	2.720	0.65	0.75	0.79	0.99	0.98	0.96	0.65	0.74	0.76
HRC6	0.591	2.773	0.64	0.73	0.78	0.99	0.97	0.96	0.64	0.72	0.76

duration of the visual path between the video region of interest and the subtitle is not so annoying because observers tend to enter in the subtitle area only once and read entirely the subtitles before going back to video content. Indeed, the average number of visits in the subtitle area equals 1.28 for all HRCs. It can also be noted that observers with an upper-intermediate or bilingual English language proficiency level obtain, in average, a significantly smaller RIDT score compared to the other viewers ($\chi^2(1) = 74.48; p < .001$). It can be explained by the fact they do not need to read the subtitles entirely to understand the storyline of the videos but at most likely looking for keywords and/or are visually attracted by the apparition of subtitles [19].

V. CONCLUSION

In this paper, an experiment has been described in order to evaluate subtitle appearance in UHD viewing conditions using eye-tracking measures. The analyses of the reading comfort and intrusion of subtitles allowed us to propose guidelines for the appearance of subtitles in UHD. First, it is not recommended to display long subtitles using one line of text only. Second, vertical visual angle of subtitles can be conserved from HD to UHD, but doubling this angle is preferred to maximize the ease of reading and minimize the disturbance of subtitles. This applies even more that in standard European living rooms, the median distance between TV screens and observers is 2.63 m for a median screen height of 49 cm, namely a median viewing distance of 5.5H [24], far superior from the experimental viewing conditions. To confirm these findings, future work is needed to investigate more varied subtitle sizes and style characteristics in UHD to define the optimal subtitle appearance.

However, currently available metrics from eye-tracking data are not sufficient to appropriately evaluate subtitle disturbance, visual fatigue, understanding of the storyline and quality of experience. To address these limitations, this study could be extended by combining visual attention with subjective self-assessments.

ACKNOWLEDGMENT

This work is part of the UltraHD-4U project financed by the DGCIS through the European CATRENE program CAT111.

REFERENCES

[1] ITU-R BT.2020, "Parameter values for ultra-high definition television systems for production and international programme exchange," *Int'l Telecommunication Union*, 2012.

[2] C. Jensema, "Viewer Reaction to Different Captioned Television Speeds," Institute for Disabilities Research and Training, Inc., Tech. Rep. June, 1997.

[3] K. Rayner, T. J. Smith, G. L. Malcolm, and J. M. Henderson, "Eye Movements and Visual Encoding During Scene Perception," *Psychological Science*, vol. 20, no. 1, pp. 6–10, jan 2009.

[4] A. Szarkowska, I. Krejtz, K. Krejtz, and A. Duchowski, "Harnessing the Potential of Eye-Tracking for Media Accessibility," *Translation Studies and Eye-Tracking Analysis*, pp. 153–183, may 2013.

[5] N. M. Ross and E. Kowler, "Eye movements while viewing narrated, captioned, and silent videos," *Journal of Vision*, vol. 13, no. 4, pp. 1–1, mar 2013.

[6] J.-L. Kruger, E. Hefer, and G. Matthew, "Attention distribution and cognitive load in a subtitled academic lecture: L1 vs . L2," *Journal of Eye Movement Research*, vol. 7, no. 5, pp. 1–15, 2014.

[7] Y. Diao, P. Chandler, and J. Sweller, "The effect of written text on comprehension of spoken english as a foreign language," *The American journal of psychology*, vol. 120, no. 2, pp. 237–261, 2007.

[8] BBC, "Online Subtitling Editorial Guidelines V1.1," 2009.

[9] CHANNEL 4, "Subtitling Guidelines for foreign-language programmes."

[10] A. Brown, R. Jones, M. Crabb, J. Sandford, M. Brooks, M. Armstrong, and C. Jay, "Dynamic subtitles: the user experience," in *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, 2015, pp. 103–112.

[11] S.-B. Park, K.-J. Oh, H.-N. Kim, and G.-S. Jo, "Automatic Subtitles Localization through Speaker Identification in Multimedia System," in *2008 IEEE International Workshop on Semantic Computing and Applications*, jul 2008, pp. 166–172.

[12] D.-S. Lee and M.-L. Huang, "Screen luminance, subtitle, and viewing angle on viewing distance of liquid crystal display high-definition television," in *2012 International Conference on Computer Science and Electronics Engineering*, vol. 3, mar 2012, pp. 106–109.

[13] C. E. Duchon, "Lanczos filtering in one and two dimensions," *Journal of Applied Meteorology*, vol. 18, no. 8, pp. 1016–1022, 1979.

[14] J. Li, Y. Koudota, M. Barkowsky, H. Primon, and P. Le Callet, "Comparing upscaling algorithms from hd to ultra hd by evaluating preference of experience," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014, pp. 208–213.

[15] ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," *Int'l Telecommunication Union*, 1999.

[16] ITU-R BT.1769, "Parameter values for an expanded hierarchy of lsdi image formats for production and international programme exchange," *Int'l Telecommunication Union*, 2006.

[17] ITU-R BT.710-4, "Subjective assessment methods for image quality in high-definition television," *Int'l Telecommunication Union*, 1998.

[18] T. Vigier, J. Rousseau, M. Perreira Da Silva, and P. Le Callet, "A new hd and uhd video eye tracking dataset," in *7th ACM Multimedia Systems Conference (MMSys)*, may 2016.

[19] J. Kruger, A. Szarkowska, and I. Krejtz, "Subtitles on the moving image: an overview of eye tracking studies," *Refractory - A Journal of Entertainment Media*, 2015.

[20] Tobii Technology, "User Manual - Tobii Studio," pp. 1–164, 2014.

[21] J.-l. Kruger and F. Steyn, "Subtitles and Eye Tracking: Reading and Performance," *Reading Research Quarterly*, vol. 49, no. 1, pp. 105–120, 2013.

[22] O. Le Meur and T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behavior Research Methods*, vol. 45, no. 1, pp. 251–266, 2013.

[23] S. C. Sereno and K. Rayner, "Measuring word recognition in reading: eye movements and event-related potentials," *Trends in Cognitive Sciences*, vol. 7, no. 11, pp. 489–493, nov 2003.

[24] K. C. Noland and L. H. Truong, "A survey ok UK television viewing conditions," BBC R&D, Tech. Rep., 2015.