El Mehdi Ibourk and Amer Al-Adwan*

Subtitling Virtual Reality into Arabic: Eye Tracking 360-Degree Video for Exploring Viewing Experience

<https://doi.org/10.1515/les-2019-0016>

Abstract: The recent years have witnessed the emergence of new approaches in filmmaking including virtual reality (VR), which is meant to achieve an immersive viewing experience through advanced electronic devices, such as VR headsets. The VR industry is oriented toward developing content mainly in English and Japanese, leaving vast audiences unable to understand the original content or even enjoy this novel technology due to language barriers. This paper examines the impact of the subtitles on the viewing experience and behaviour of eight Arab participants in understanding the content in Arabic through eye tracking technology. It also provides an insight on the mechanism of watching a VR 360-degree documentary and the factors that lead viewers to favour one subtitling mode over the other in the spherical environment. For this end, a case study was designed to produce 120-degree subtitles and Follow Head Immediately subtitles, followed by the projection of the subtitled documentary through an eye tracking VR headset. The analysis of the eye tracking data is combined with post-viewing interviews in order to better understand the viewing experience of the Arab audience, their cognitive reception and the reasons leading to favour one type of subtitles over the other.

Keywords: Audiovisual translation, Arabic subtitling, virtual reality, eye tracking, viewing perception

1 INTRODUCTION

The core elements of audiovisual translation (AVT) are the same multimodal elements of reality that humans have had to engage with since the start of human

***Kontaktperson:** Amer Al-Adwan, Hamad Bin Khalifa University, LAS Building, Al Luqta St., P.O. Box 5825, Qatar, E-Mail: aadwan@hbku.edu.qa

El Mehdi Ibourk, Ministry of Education and Higher Education, Al-Dafna area, P.O. Box 35111, Qatar, E-Mail: e.ibourk@edu.gov.qa

Open Access. © 2019 Amer Al-Adwan and El Mehdi Ibourk, published by De Gruyter.

This work is licensed under the Creative Commons Attribution 4.0 Public License.

communication. The multi-modal nature of AVT has secured communication and transfer of reality between text makers and the intended recipients by virtue of rapidly changing technologies of projection. The film industry has shifted from wide theatrical screens to smaller and individualized projection interfaces, such as televisions, laptops, tablets and smart phones. At the level of the format, the industry is constantly developing and screening stories in advanced and sophisticated environments, such as virtual reality.

Nowadays, most documentaries are in English, with the exception of few ones that offer interlingual subtitles, which are fixed in a static point in the VR environment. One of the main objectives of this paper is to reflect upon subtitles' design and perception for an optimal viewing experience. Needless to say, drawing on eye tracking technology is key to visualize how subtitles are perceived in a VR environment, by measuring the motion of the eye, its point of gaze and focus relative to an object of interest.

In March 2017, the BBC R&D department designed four subtitling behaviours (types) for a 360-degree video and conducted a user testing experiment to analyse the subtitling preferences of 24 participants. These four subtitling behaviours are: (1) the 120-degree subtitles, which consists of inserting three subtitle blocks located in the VR environment at 120° angles around the user, (2) the Follow Head Immediately subtitles, in which the viewers turn their head while these subtitles stay fixed to the centre of their field of view, (3) Lag-Follow subtitles, which follow the head movement of the VR user only when s/he turns beyond 30 degrees and remain in the centre of the field of view, (4) Appear subtitles, which only follows the head movement of the user when the new subtitle appears. Once on screen, the subtitle will remain fixed in the VR environment until the next subtitle appears in the new position and field of view of the Head Mounted Display (HMD) user. This paper focuses only on the first two subtitling behaviours: The 120-degree subtitles and the Follow Head Immediately (previously referred to as Static Follow subtitles by the BBC team).

The BBC experiment evaluated the effectiveness of the subtitling behaviours based only on the personal preferences of the participants. Although the results of the project show that most participants favoured the Follow Head Immediately subtitles, the animation that shows how this type of subtitles works only illustrates movement at the horizontal level. The BBC online paper (Brown 2017) does not include further information on whether the Follow Head Immediately subtitles do /do not move according to the head movements of the participants in all directions. On the other hand, this paper aims to understand the subtitling preferences of the audience based on a more objective and measureable investigation of how the viewers perceive and interact with subtitled VR content, through eye tracking technology. This will also give us the

opportunity to study the extent of readability of the Arabic subtitles projected on VR content.

2 LITERATURE REVIEW

The study of viewers' perception of a 360-degree documentary supported with Arabic subtitles represents an exploratory research on the integration of audio-visual translation and virtual reality content. The study of subtitling VR content demonstrates the inter-disciplinary nature of audiovisual translation that covers almost all facets of life. Previously known as film translation (Gambier 2003), the field of AVT was distinguished from translation studies given its multi-modality and relevance to all types of audio and visual texts. For a basic understanding of subtitling, Diaz Cintas and Remael (2007: 8) define the process as:

A translation practice that consists of presenting a written text, generally on the lower part of the screen, that endeavours to recount the original dialogue of the speakers, as well as the discursive elements that appear in the image [...] and the information that is contained on the soundtrack.

Such conventional positioning of the subtitles ensures a comfortable processing within the field of view to both read and understand the dialogue and relate it to the audio-visual elements of the film. However, processing the subtitles alone does not guarantee the audiovisual language transfer (Gambier 1996) since the film narrative is not only based on the dialogue.

By reviewing the literature on audiovisual translation, it is observed that dynamism is related to the change of the subtitles' position on the platform for speaker identification, inter alia. It should be distinguished from the definition of subtitling by Gottlieb (1994: 101) as a process that “switches from the spoken to the written mode, and it represents itself ‘in real time’, as a dynamic text type”. Such consideration of subtitles as a dynamic text in the film is extended in the research to include the artistic function of dynamic subtitling in filmmaking.

Describing the subtitling techniques incorporated in selected scenes of the Academy Award winning film *Slumdog Millionaire* (2008), McClarty (2011: 143) states:

Rather than remaining at the bottom of the screen, as is standard subtitling practice, the subtitles have been raised to a more prominent position within the action of the film [...] the point of the subtitles in this film seems to be that the audience gains an overall understanding of the situation and dialogue, rather than a word by word comprehension of each subtitle.

At the level of decision-making, McClarty (*ibid.*) further explains that “these creative subtitles, with all their comedic, narrative and artistic functions, stem from the creative mind of the filmmaker; not the linguistic mind of the translator”. However, the displacement of the subtitles as dynamic texts can be regarded as a subtitling choice, as speaker identification is possible through tagging the name of the character above the dialogue at the lower part of the screen.

2.1 Eye Tracking

Understanding human visual behaviour represents an element of progress in research and development for both the filmmaking industry and audiovisual translation. While film studies examine the interaction of viewers with movies, for example, and the impact of the form and content on developing a desired cinematic experience, a number of experiments were conducted by AVT scholars to study the audience’s processing of audiovisual translated texts based on reception studies and technologies.

The use of technology in research has reshaped data collection and analysis. The tools evolved from post-test questionnaires and naked-eye observations to accurate tools that record data in real time, such as the eye tracking technology. Prior to film and AVT studies, the eye tracking technology was first introduced in educational psychology (Huey 1907), cognitive psychology (Rayner 1998), and in medical and usability experiments to detect the eye movements of individuals. According to Salvucci and Goldberg (2000: 71) “researchers typically analyse eye movements in terms of fixations (pauses over informative regions of interest) and saccades (rapid movements between fixations)”. A number of fixation and saccadic metrics are used to measure the eye gaze movement, frequency, duration, path and scope against an object of interest (Rayner 1998). Objects can either be static or dynamic, and the perception of the latter is examined by monitoring the “smooth-pursuit” eye movements of the participant(s) (Rashbass 1961). The technology is usually processed by infrared eye tracking devices. Eye tracking allows us to measure an individual’s visual attention, yielding a rich source of information on where, when, how long, and in which sequence certain information in space or about space is looked at. (Kiefe et al. 2017: 1)

In his commentary article on eye tracking, Smith (2015) investigates the technology in a number of selected experiments. He (*ibid.*: 2) acknowledges that “eye tracking is a powerful tool for quantifying a viewer’s experience of a film, comparing viewing behaviour across different viewing conditions and groups as well as testing hypotheses about how certain cinematic techniques impact where we look”. However, the technology – according to him – falls short to provide an

accurate study of the viewer's experience of all genres of films. He (ibid.: 10) argues that:

One practical reason why eye tracking studies rarely use foreign language films is the presence of subtitles. As has been comprehensively demonstrated by other authors [...], the sudden appearance of text on the screen, even if it is incomprehensible leads to differences in eye movement behaviour. This invalidates the use of eye tracking as a way to measure how the filmmaker intended to shape viewer attention and perception.

This statement is only valid if eye tracking aims at studying filmmaking in terms of design and impact on a restricted audience. The application of eye tracking technology has not only provided AVT scholars and researchers with an insight on how the viewing experience is developed among non-native audiences, it has also generated quantitative and qualitative data on the impact of subtitles on the viewer's cognitive load (Kruger et al. 2013), reading performance (Kruger and Steyn 2014), and construction of the narrative (Kruger 2012). Fox (2016: 5) used the eye tracking technology to study (sub)titles and to examine "to what extent the placement and design of (sub)titles affect reading time and the visual perception of the image". A similar previous study adopted the eye tracking technology to measure the effects of experimental subtitling procedures on viewer perception of subtitled AV content (Caffrey 2012).

The study of the visual processing of subtitles as dynamic texts (Kruger and Steyn, 2014) through an eye tracking experiment aimed at measuring the viewers' perception of the subtitles by developing a reading index for dynamic texts (RIDT). Kruger and Steyn (2014: 10) note that the index represents "a product of the number of unique fixations per standard word in any given subtitle by each individual viewer and the average forward saccade length of the viewer on this subtitle per length of the standard word in the text as a whole". The RIDT reads as follows (ibid):

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

While p stands for the participant, s and v stand for the subtitle and video, respectively. The application of eye tracking in AVT research, as in this study, indicates that visual behaviour and performance can be accurately measured by taking into account the dynamic features of AVT texts in 2-D platforms such as TV screens or computer monitors. The dynamic variables calculated in the index concern only the subtitles (s) as dynamic texts and the video (v) featuring moving and dynamic images and objects.

However, the measurement of the viewers' processing of subtitles in VR content requires an additional level of dynamism be considered for accurate data

collection and analysis. HMD users are required to move both their heads and bodies to explore the VR content and environment in 360 degrees. The project by Brown et al. (2017) in producing subtitles for a 360-degree video content shows that there is a need to formulate a RIDT for VR content that takes into account the (in)visibility of the subtitles for the viewer and the minimum / maximum time for the viewer's eyes to detect, reach and read the subtitles before they disappear. The issue of eye tracking accuracy at this level becomes multi-dimensional if we take into consideration the stability of the HMD eye tracking cameras. Besides, the possibility of feeling nausea by HMD users can also affect the processing of the subtitles and the visual elements and have a serious impact on the interpretation of the eye tracking results and the viewing experience in general.

Al-Wabil et al. (2010) published the first eye tracking-based experiment article in the Arab world. They studied the interaction of native Arabic-speaking kids with online Arabic books by recording the participants' eye gazes and movements while browsing a computer monitor supported with Tobii 120 eye tracking device. Whereas the first eye tracking study using an interlingual stimuli was conducted one year later by Al-Khalifah and Al-Khalifa (2011). They examined the impact of Arabic language on reading English as a Foreign Language (EFL) through eye tracking technology. Similar to the previous study, the Tobii 120 eye tracking device was used to record the eye movements of 23 female native Arabic-speaking learners of English. The analysis reveals that longer fixations are recorded when the subjects were reading irregular words in English, therefore showing that there is "a strong relationship between the way Arabic is read and the reading of English text for EFL students" (ibid.: 279).

With regard to the application of eye tracking technology in AVT research in the Arab world, it appears that the field remains uncharted, and thus the eye tracking data generated on the processing of the Arabic subtitles in this paper requires a careful analysis for an objective interpretation of the findings. In the absence of a RIDT compatible with VR content, the eye movements of the participants will be recorded against their processing of a 360-degree environment while featuring the 120-degree and Follow Head Immediately subtitles.

3 METHODOLOGY

3.1 The Source Text (360-degree Documentary)

This paper is an exploratory and applied research by nature. In order to study the viewers' processing of two types of subtitles, the 120-degree and the Follow Head Immediately as developed by the BBC R&D (2017), we produced two subtitled

versions of the English documentary “360° Underwater National Park”, produced by National Geographic (NG), in January 2017. Group 1 will watch the clip with the first half supported by 120-subtitles and the second half supported by Follow Head Immediately subtitles. On the other hand, group 2 will watch the same clip, but with a reversed order of the subtitles. One result of juxtaposing these two types of subtitles in one video is that it will offer us the opportunity to analyse the viewing experience of the participants without affecting their answers in the post-documentary interview. By combining the Adobe After Effects software and the Samsung VR player to create – respectively – the above-mentioned subtitles, the viewing experience of the sample will be natural and will not require watching the same documentary twice.

The six-minute documentary features Brian Skerry, a NG photographer, who narrates life in St. Croix’s Buck Island Reef in US Virgin Islands, its fauna and flora and the perils endangering this reserve and the sea turtles living within. The video was first accessed, in October 2017, on YouTube. It is worth mentioning that YouTube offers the possibility of watching 360-degree videos by using cardboards and VR headsets compatible with smart phones or by wearing powerful gaming-oriented headsets such as the Oculus Rift, HTC Vive and Pimax 8K.

360-degree videos can be uploaded by users in their respective YouTube channels. While some of these videos support automatically-generated captions, it was noticed that the subtitles did not follow the viewer’s head movements and only remained at the bottom of the video when shifting to the 360 viewing mode. Therefore, the documentary was downloaded from YouTube to produce the subtitles and to conduct the research.

3.2 The participants

Eight participants were selected for the case study: four females (F) and four males (M); all above 30 years of age. The reason behind setting a minimum age limit is to make sure the participants are accustomed to watching audiovisual materials through the conventional platforms in order to move to the next level of watching audiovisual materials through virtual reality headsets. All the participants are Arab native speakers and understand the functionality of subtitles as a means to access the film content in Arabic. Keeping the level of proficiency in English to the minimum was a requirement for profiling the sample, and therefore better study their perception of the subtitles.

Participants had enough space to allow for a freedom of mobility in all directions. They also sat on rotating chairs while watching the documentary for two reasons: (1) to turn their heads to the desired area of interest in the spherical

environment and (2) to avoid or minimize any possible feeling of nausea. The participants were also selected on the basis of their health conditions. All the participants confirmed that they did not have any vision abnormalities or heart conditions, and all the ladies further confirmed they were not pregnant at the time of conducting the case study.

3.3 Projection of the Documentary for Eye Tracking

The duration of the case study ranged from 15 to 20 minutes for each participant. It included watching the documentary through the Fove Eye Tracking headset, a tethered HMD that projected the documentary from the computer.

The cable of the headset is three meters long, which was comfortably distant from the desk or other pieces of furniture. However, we had to hold the cable up for the entire documentary for each participant to facilitate the movements of their heads and the rotation of their bodies on the chair to the desired direction. Usually, VR companies set up the venue by using roof supports to hold the cables of their hardware during exhibitions. Calibration tests for all participants were run to make sure the infrared eye cameras in the headset were accurately tracking the eye movements and to make sure the position tracking camera, opposite to the participant(s), was working properly.

These tests were necessary to ensure both the documentary was playable in the 360-degree mode and to verify that the accuracy rate of the headset was optimal for all the participants regardless of their eye characteristics and variables. The second step of calibration was to run a test demo where the HMD user was required to gaze at lamps at different locations. The demo test was also an opportunity for the participants to be engaged and feel more comfortable with using a HMD headset prior to watching the documentary.

The participants watched the documentary through a VR player installed on a personal laptop and was projected to the headset. The player is developed by Samsung (version 1.86.3) and is the only player, compared to others, that fits the case study. Although a large number of VR video players are available online for download and use, the tests to verify the mobility of the Follow Head Immediately subtitles showed that not all VR players offer the possibility of running subtitle files that can move in the VR environment.

Despite the fact that the VR player is developed for Samsung smart phones, we managed to run its executive file on the computer as a portable device. The next step was to enable the eye tracking headset as the main display and synchronize it with an eye tracking analysis software called Hairball. The problem with the FOVE eye tracking headset is that it is generally developed for gaming

purposes, to allow players to use their eyes as controllers. The device does not come with a built-in eye tracking analysis software. The eye tracking software needs to be developed by interested individuals through software development platforms such as Unity and Unreal Engine. The mastery of these platforms, let alone software engineering, is currently beyond our scope of competence. We consider the Hairball software a walk-around to design and conduct our case study, as it is an eye tracking analysis software compatible with most eye tracking devices. The synchronization of the FOVE HMD was possible as the software allows pairing with other eye tracking devices.

The next step is to load the Eye Tracker SDK Library file (*.dll) that is included in the FOVE headset's installation folder under the name "Eye Tracking.dll". After synchronization, the eye tracking device is recognized as an eye controller. These steps were important to ensure that the viewing experience of the participants was recorded by the software as an eye tracker data file (*.csv) to be analysed as a media file. These steps, however, are not related – in terms of processes or programs – to the production and incorporation of the subtitles to the documentary.

After the participants watched the documentary, they sat for a semi-structured interview that included the following four questions:

- 1- During the first scene of the documentary, there was a man standing on the boat. What is the colour of his T-shirt?
- 2- Later on at night, did you notice any animals?
- 3- When you were flying next to the airplane, the narrator said that the reserve was inaugurated by the U.S. President J.F. Kennedy. Can you remember in which year?
- 4- After watching the documentary on a VR headset, maybe you have noticed that there were two types of subtitles, what is your impression about them in terms of comfort and obtaining information?

The questions were designed to cover different scenes of the subtitled documentary and to analyze the perception of the subtitles and the viewing experience of the participants. The recorded interviews were conducted in Arabic. The answers to the above-mentioned questions of the interview and the related eye tracking records will be used in the following section as a combined method for data collection and triangulation of the findings.

4 Data Analysis

The study of perception of subtitles in the VR environment is based on the eye tracking recordings of the participants and their responses during the post-documentary interviews. The viewing experience of the participants is analysed by introducing their answers to the questions that were designed to reflect their perception of (1) objects and (2) the features of objects – colour for example – as well as testing (3) if the subtitles are a source of information to understand the content and (4) their impression of watching the 360-degree video supported by two types of subtitles: the 120-degree and the Follow Head Immediately. The eye tracking recordings visualize the viewing experience by capturing the eye movements and areas of focus to understand how the participants watched the 360-degree documentary. Such visualization might help understand the reason(s) for favouring one subtitling type over the other, and the viewer's enjoyment/non-enjoyment of the VR experience with Arabic subtitles.

4.1 The Perception of the Features of Objects

The first question was about the colour of the T-shirt of Brian Skerry, the National Geography photographer and on-stage narrator. The scene where the colour of his white T-shirt is clearly visible lasts from second 00:02 to 00:20. The narrator was positioned above the center of the screen (1280x720 pixels in all footage). P1 answered the question by saying that the T-shirt is 'white, approximately white by 80 %' after being asked about how certain he was. By viewing the eye tracking record for P1, figure (8) shows that his eyes actually crossed the top body part of the narrator but quickly looked down and focused on the white bottom of the boat as it is sailing and causing white bubbles before the appearance of the first 120-degree subtitle “أردت أول الأمر أن أصبح مجرد مستكشف لعالم ما تحت البحار” [Initially I just wanted to be an underwater explorer]. The dominance of the white colour in the area of interest (boat) might show why P1 could not recall the exact colour of the T-shirt. He also did not read all the subtitle before it disappeared, which is the case of all Group 1 participants. Since it was the first 120-degree subtitle in the second half of the documentary, the partial coverage of this specific subtitle could be explained as a matter of surprise.

Figure 1: Gaze Heat Map of P1 in Relation to Question 1

P2, however, stated that the T-shirt is yellow. She added that she was not sure but it was more likely to be of yellow colour. The analysis of the eye tracking footage reveals that she focused on the yellow rectangular-shaped logo of National Geography that first appeared at the center of the spherical environment for four seconds. The eye gazes were subsequently distributed to quickly watch the standing man, the boat, the wide parts of the sea and the island on the background.

Figure 2: Gaze Heat Map of P2 in Relation to Question 1

P3 believed the colour of the T-shirt is ranging between beige and blue. She justified her answer by saying that ‘the dominance of the blue colour in the scene and other colours were more attractive to her eyes than the colour of the T-shirt’. To validate the answer of P3 to the first question, her eye tracking record in that specific scene shows that she had a quick look at the brown island on the top right of the boat on

the background. However, gazing at the island cannot be ruled as the main reason behind her answer. The video also shows that her fixation gazes covered the blue shorts of the photographer during this opening scene. It was also noted that there is another scene in the middle of the documentary where Brian Skerry is wearing a dark blue T-shirt and taking photographs. It might be that she was mixing colours from different scenes while thinking of the exact answer to the question.

P4 of Group 1 – whose assigned clip starts with the 120-degree subtitles – stated that the T-shirt is white. He replied quickly to the question and seemed sure about his answer. To validate his answer, figure (3) shows that his eyes were focusing on the center of the spherical environment where the white colour of both the T-shirt and the boat were within his field of view. It is worth reporting that that P4 only moved slightly to the right and to the left during this first scene of the documentary.

Figure 3: Gaze Heat Map of P4 in Relation to Question 1

The analysis of the same question showed similar details for the second group. Although the first half of their assigned documentary is supported with Follow Head Immediately subtitles, P5, P7 and P8 were not sure if the T-shirt is white, providing answers that range from blue, ‘may be white’ and orange, respectively. The reason behind the responses, although unexpected during my pilot testing given that the scene lasts for 14 seconds before the narrator speaks, is that the three participants were accommodating their vision inside the spherical environment for the first time. Besides, the first subtitle appeared slightly below their eye levels and moved following their head movements, which additionally explains why they focused on the moving subtitle and could not recall the colour of the T-shirt as they moved slightly away from the area of the boat. The confident answer of P6 to the question as white can be seen in his eye tracking footage that shows he did not move away from the central point that covers both the white boat and the T-shirt.

By analysing the participants' eye tracking records and their answers to the question on the feature (colour) of the object (T-shirt), it shows that head-mounted display users do not usually focus on the exact details of objects during the first moments of watching the VR content as they are trying to discover the environment and position themselves accordingly. Recalling a feature that appeared at the onset of the documentary is usually achieved by profiling all memorized images back to the first moments to provide the exact answer. This cognitive effort from the participants to respond to the first question can explain why most of them were undecided. The colour of the T-shirt was visible for 13 seconds before the appearance of the first subtitle. It lasted for 18 seconds in total before the scene changed. This suggests that the recollection of features of objects decreases as new objects, including the subtitles, succeed.

The first question and the related eye tracking records reveal another important element on the perception of objects according to their status of mobility. The characteristics of static objects that do not have a narrative value in the documentary (T-shirt) are easily neglected by the eye when they appear next to bigger objects (the boat and the ocean). The bigger the object is, the more likely the viewer remembers its feature(s).

4.2 Eye Tracking Moving Objects

The study of the perception of moving objects in the VR content is the basis of the second question in the semi-structured interview. All the participants, except P5, mentioned the sea turtles as a response to the second question "During the scene at night, did you notice any animal or animals?" Participants 1, 4 and 6 even elaborated their answers by stating that the turtles broke out of their eggs and that humans need to help them reach the sea (P4 and P7). Others only answered by saying they saw a few turtles, or just one turtle as is the case of P8.

The exception of P5 is worth studying through eye tracking to understand why she could not see the sea turtles as they are dynamic objects of narrative value to the documentary. The records generated by the FOVE headset (Figure 4) show that the concerned participant was reading the subtitle containing the information on the sea turtles while focusing on the narrator.

Figure 4: Gaze Heat Map of P5 in Relation to Question 2

For the sake of an objective analysis of her response to the question, it must be added that P5's eye gazes came across the lady's hands holding the sea turtles and releasing them to the sea. The participant answered the question by saying that she first remembered seeing the corals (a previous scene), adding that she saw the man digging into the sand but could not remember seeing the turtles. She only read about the turtles through the FHI subtitles. The same picture shows that part of her eye gazes heat map covers some turtles that were moving close under the subtitle. A possible explanation of the matter is that the image of the corals and focusing on both the narrator's hands and the moving subtitles distracted her from 'remembering' seeing the dark turtles as physical objects at night. Her answer was therefore based on seeing the subtitle as a source of information and not on the content itself.

4.3 Analysing the Subtitles as a Source of Information

In order to study the relevance of subtitles as a mode of audiovisual translation for 360-degree videos and virtual reality, the research project studies the proposed subtitles as a source of information to understand the content in Arabic. The third question is related to a scene where the narrator gives details on the inauguration of the underwater park in 1961 by the US President John F. Kennedy. By asking the participants when the park was founded by the US president, only P4 gave the correct answer (1961), while both P6 and P8 said that the park was inaugurated in the 1960s. The rest of the participants did not remember the date at all. For this specific question, the interview was designed to make the inter-

viewee elaborate more on the answer if s/he was close to the answer or correct, by explaining whether he heard the information or saw it in the subtitle. The objective of this question is to analyse the viewing experience of the unusual and immersive scene of flying next to the plane and watching the underwater park from the air, while receiving the information uttered by the narrator.

P4 further explained that he heard the narrator saying the year 1961 in English while he was watching both the scene and portions of the subtitle. While both P6 and P8 said that the park was founded in the 1960s, P8 further explained in Arabic that the narrator said ‘jewel’ [in English] to express his fascination for the reserve.

Figure 5: Gaze Heat Map of P8 on the 120-degree Subtitle Related to Question 3

The eye tracking footage of P8 (Figure 5) reveals that the hot zone did not cover the year 1961 as it is located at the end of the second line of the subtitle. There are two reasons: First, the change of the type of the subtitle caused a momentary focus on the center below his eye level, where he expected the Follow Head Immediately subtitle to appear. Secondly, and because he shifted to the 120-degree subtitle, the participant only noticed the edge of the new subtitle to his left and covered a small part of the subtitle before it disappeared. This suggests that P8's answer was probably based on the English narration – by mentioning the English word ‘jewel’ that appeared in a subsequent subtitle, and not based on reading the subtitle as a source of information.

4.4 VR film viewing: An Act of Enjoyment

The immersive nature of the scene and the shift to a second type of subtitle are two elements to consider to understand the viewing experience of the subtitled 360 documentary. Prior to the shift in the subtitles, all the participants started – to different extents – to interact with the spherical environment by moving their heads in all directions and by using the rotating chair to further explore the environment at the horizontal axis. P2, P3, P5 and P7 as well as P4 and P8 were smiling during the first scene of the apparent flying next to the airplane, as an interaction with the aerial scene. With the exception of P8 who saw part of the 120-degree subtitle, the remaining participants of group 2 did not see the subtitle as their heads were down to see the island and the ocean, making the subtitle above their fields of view. All participants in group 1 saw the FHI subtitle as it appeared for the first time in front of them. The clear and moving nature of this subtitle did not go unnoticed. The gaze fixations of these four participants at the new subtitle can be seen in their respective eye tracking video records. The footage reveals that their eye gazes were focusing on two thirds of the subtitle from the right, which explains why they did not recall the year of the inauguration of the reserve from the subtitle. The only exception was P4 who listened to the narrator and thus recalled the date. It shows that the proficiency of the source language plays a key role in shaping the viewing experience of virtual reality. Participants who depended heavily on the Arabic subtitles, especially group 2, started to move quicker during the second half of the documentary compared to group 1 in order to relocate themselves in the environment in search for the new subtitle, which was previously following their head movements and always within their fields of view.

Despite the fact that all the participants were moving in the intervals between the subtitles to explore the environment, the movement of the individuals watching the assigned half with 120-degree was more restricted. They no longer made a full rotation with the chair and often oriented their heads back to the last field of view where the subtitle is expected to reappear when the photographer resumed narration. In general, the eye tracking records of group 1 show greater sense of freedom in motion and a greater scope of view of the second half of the documentary due to the mobility of the FHI subtitles. Their enjoyment of the subtitles considerably changed as they realized that the new subtitle is omnipresent in the VR environment.

Because enjoyment cannot be analysed or measured through eye gazes or heat maps, the answers of the participants to the fourth question of the interview reveal that they enjoyed watching the documentary through a VR headset. Most of the participants stated that they did not focus on the subtitles as they were

attracted by the images and the feeling of being inside the environment. However, they stated that the FHI subtitles moved to their desired fields of view and appeared clear and comfortable to their eyes. P7 revealed that the 120° subtitles would be better and more enjoyable if they were less blurry. The fact that the 120° subtitles were burned to the content resulted in a change (degradation) in their quality, although they appeared clear during production on Adobe After Effects.

4.5 The Impact of the Subtitles on the Viewer's Area of Interest

Virtual reality is designed to ensure maximum immersion, mobility and freedom to watch the content at any desired direction. My observation of the selected sample watching the subtitled documentary shows that the participants, although they gradually started to move their heads and use the rotating chairs to explore the environment, experienced moments of restricted mobility. The eye tracking records of Group 1 (120° > FHI) show that their gazes did not reach distant areas during the first half of the documentary. Very few occasions of a full rotation were recorded. The main sequence of movement was at the horizontal level, either to the right or left before going back to the previous position where the next subtitle is expected to appear. Because the participants did not know in advance that the 120-degree subtitles are evenly located at three positions in the environment, they confined themselves the areas surrounding the previous 120-degree subtitle.

It is worth noting that the gazes of Group 1 were mainly vertical during this half of the documentary. It suggests that the need of the majority of this group to catch the next subtitle restricted their movements to explore other areas of the environment. The locations of the 120-degree subtitles were therefore 'compulsory' areas of interest for the participants to understand the English narration through the Arabic subtitles. The areas of interest of Group 2 (FHI > 120°) during the same half were greater in terms of range and content. Although they did not quickly start to make full rotations during the first two minutes of the documentary, realizing the Follow Head Immediately subtitles were moving according to their directions allowed them a considerable freedom of mobility compared to the first group. The gazes of Group 2 reached further areas with the Follow Head Immediately subtitles. The viewers were free to look in any direction as long as the subtitle always appeared below their eye level.

The shift to the 120-degrees subtitles had a considerable drawback on the viewing experience of Group 2. The video footage of the heat maps of this group is a clear evidence that their gazes were concentrated below eye level, where they expected the FHI subtitle to appear. Despite the fact that they were at different

vertical axis points compared to the central level of the VR composition, the hot zone (in red) was greater below eye level than any of their gazes at the other 120-degree subtitles. It lasted for a minimum of two subtitles for most of them. P7 even expressed orally her confusion over what was happening after she was unexpectedly shifted to a new type of subtitles. Because this was expected to happen, all participants were asked to kindly not remove the headset no matter what happens during watching the documentary.

The perception of the new Follow Head Immediately subtitles by Group 1 (120° > FHI) was effortless and smooth. The participants did not cover the first subtitle in its entirety but they all saw it as it appeared below their eye level, regardless of the area of interest. The partial coverage of the subtitle can be explained, after reviewing the eye tracking footage, that P1, P2 and P4 focused on the new subtitle in front of them and moved sideways to 'catch' the usual 120-degree subtitle before it would disappear. The sequence of their eye gazes was zigzagging between the new subtitle as it remained visible and the empty spaces nearby where the previous subtitle behaviour was expected to repeat itself in the environment.

Group 1 took less time to interact with the new FHI subtitles than did Group 2 with the 120-degree subtitles. I noted that the degree of rotation of Group 1 (120° > FHI) increased during the second half of the documentary. The Follow Head Immediately subtitles enabled Group 1 to explore different areas while the FHI subtitles remains visible. Generally, the eye tracking records did not show a pattern of movements among the participants when they had watched the FHI subtitles, which suggests that the selection of the areas of interest is uncontrolled and left to the choice of our sample. However, the following section (4.6) includes three cases where the participants were redirected to look at different areas to correct their positions to the relevant content after reading two subtitles and watching a particular scene in the documentary.

4.6 Visual Re-orientation Caused by Subtitles / Scenes

The documentary narrates the fauna and flora of Buck Island's natural reserve in the United States of America. One scene was shot at night as it is the perfect time for sea turtles to crack their eggs open and head to the ocean for the first time. The progress of the viewing of my sample from the start of the documentary to this scene resulted in different areas of interest based on their following of the subtitle and/or the visual elements of the documentary. The area where the narrator is seen digging on sand to help the hatchlings emerge is a small luminous section (approximately 40 per cent) of the entire scene in night environment that was in total darkness.

The narrator referred to these animals explicitly in a number of occasions in the original script. Accordingly, the related subtitles appeared in this scene as in the following table:

Table 1: Selected Subtitles with Demonstrative Pronouns to Analyse Possible Impact on the Viewers' Re-orientation in the VR Environment

	Original script	Arabic subtitle
1	There is a lot of stress on sea turtle populations.	تعاني السلاحف البحرية من تحديات صعبة تحقق بها.
2	[When] you are on the beach watching the hatchlings emerge...	عندما تشاهدون تفقيس بيض السلاحف البحرية على الشاطئ...
3	Your hearts go out to these little guys...	فأنكم تشفقون على هذه المخلوقات الصغيرة
4	And these little beautiful nuggets of turtle have to, now, scramble to the ocean.	على هذه الشذرات من السلاحف الجميلة أن تزحف وتشتق طريقها نحو المحيط.

For the purpose of analysing the impact of the subtitles on the orientation of the participants, the eye tracking data in this nocturnal scene will focus on subtitles 3 and 4 that include a demonstrative pronoun. The findings show that P6 started to move to a different area of interest after reading subtitle no. 3 (in Follow Head Immediately mode). The position of the subtitle was over 70 degrees to the left of the actual location of the narrator and the sea turtles. Because the background of the subtitle was dark, her movement to another location as the subtitle was visible can be explained by stating that she reacted to the demonstrative pronoun **هذه** [these] that indicates that the turtles are present in the environment at the time of the narration.

The same analysis applies to P1, P3 and P4, who were watching the scene with 120-degree subtitles. They were reading the subtitles to the left (P1 and P4) and to the right (P3) of the area of the sea turtles. It must be indicated that the narrator and the turtles are clearly visible in this scene under the central subtitle (0°), while the other two subtitles – located at degrees 120 and 240 in the VR environment – have dark backgrounds. The eye tracking footage shows that the hot zone at the position of the demonstrative pronoun at the middle of the subtitle served as a trigger for the concerned participants to leave the subtitle and watch ‘these’ turtles. An interesting, even surprising, result of the position of the demonstrative pronoun is that the three viewers almost covered the entire subtitle, whether in the center or right side of the environment. By reviewing the time of the English narration, I noted that this 120-degree subtitle was slow enough to allow the viewers to relocate their sight and cover part of the right and central 120-degree subtitles and to watch the moving turtles underneath.

The rest of the participants read the subtitles (both types) and watched the sea turtles at different degrees of coverage. After reviewing the records, it seems that our sample was looking for a source of light in the dark environment at the beginning of the scene. The narrator reached the turtles hide using a torch to light his way. The red light attracted the attention, and therefore the visual focus of the viewers. They kept following the acts of the narrator (sweeping and digging the sand with his hand) until the 120-degree subtitle appeared above the turtles (P2), while the Follow Head Immediately subtitle appeared below the eye level of group 2 participants, at slightly different locations, but within the same area of the narrator / turtles in the environment.

Subtitle no. 4 was visible to the viewers as they focused on two out of the three luminous areas that contained visual elements of narrative value. The central area of the environment shows the little turtles running to the ocean. While the lady is seen in a separate light spot to the right carrying some hatchlings and releasing them on the sand, the third area to the far left shows nothing but sand. The demonstrative pronoun in the subtitle did not redirect the viewers to watch other areas as long as the referred object in the subtitle is clearly visible within the field of view. The participants in group 2 adjusted the position of the Follow Head Immediately subtitle by moving their heads up to make the subtitle appear on top of the turtles and the hand of the lady. The participants' focus was covering both the subtitles and the content at the same time, which explains why nearly all the participants, except P5, correctly answered the second question of the interview by stating that they had seen the bale of turtles in this scene.

The evaluation of the viewing experience of the subtitled documentary is not based on the participants' success rate in providing the correct answers. It is rather a data-based description of their perception of both the content in selected scenes and the related subtitles. While the eye tracking technology supported the recording and analysis of our sample's eye gazes, in terms of sequence of movement and focus, the final appraisal of the subtitles is solely expressed by the participants.

5 Conclusion

The excessive integration of software and hardware to conduct the case study has limited the potentials of the eye tracking analysis software. The analysis of the data was limited to the examination of the areas of focus, which could have been extended to measuring the sequence and time spent looking at those areas in the documentary. Other eye tracking analysis software such as the Tobii Pro Studio or

Gazepoint Analysis allow such combined visualization of the eye tracking data, but they are both very expensive and incompatible with the eye tracking headset used in the project.

The objective consideration of the subtitling preferences of viewers requires a thorough examination of how they perceived the subtitles. Data collection obtained through the technology of eye tracking provided measurable insight on VR film viewing and the reasons leading to favouring the Follow Head Immediately over the 120-degree subtitles, which was the case of seven out of eight participants. They considered the former type of subtitles more comfortable to the eye and always present in front of them.

The study of how the film viewing experience is built in virtual environments was also carried out through a post-viewing interview of the participants. Their responses to four questions allowed us to analyse their understanding of the narrative through its visual elements and / or the subtitles. The interviews provided qualitative data to conduct a comparative analysis of the perception of each group to identical scenes but with different subtitles. The analysis also addressed the impact of moving and static subtitles on the viewing experience of the participants.

The results of the analysis show that big objects and moving objects attracted most of the participants' fixations in the documentary. The hot zones in the heat map records explain why the viewers focused more on the sailing white boat (motion) than the white t-shirt of the narrator (static), even though they were closely visible for the same length of time. The same result applies when it comes to the size of objects; more fixations were on the big underwater Elkhorn corals than on the adjacent smaller stones on the soil. One interesting result from the eye tracking footage of our sample is that small objects attract attention, and thus eye fixation, when they are in state of motion even if they are next to bigger static objects. A good example is the small sea turtles scrambling in the vast sand area that covered the environment.

The eye tracking results show that the participants' fixations did not cover most of the 120-subtitles before they disappeared. On the other hand, the Follow Head Immediately subtitles were instantly spotted from the moment they appear in the video. This helped the viewers shift their focus to the subtitled content and cover substantial parts of the moving subtitles before they disappeared. From an eye tracking perspective, the perception of the Follow Head Immediately subtitles resulted in a greater balance of focus between the content and the subtitles.

Finally, although these obtained results were generated from a small pool of eight participants, they can be used as a preliminary framework to examine the patterns of VR film viewing and to arrive at a systematically effective integration of subtitles in terms of format, position and mobility.

References

- Al-Khalifah, Kholod S., and Hend S. Al-Khalifa. 2011. "The Effect of Arabic Language on Reading English for Arab EFL Learners: An Eye Tracking Study." In *International Conference on Asian Language Processing*, Penang: 287–90. DOI: 10.1109/IALP.2011.18
- Al-Wabil, Areej, Ebtisam Alabdulqader, Latifa Al-Abdulkarim, and Nora Al-Twairish. 2010. "Measuring the User Experience of Digital Books with Children: An Eye Tracking Study of Interaction with Digital Libraries." In *International Conference for Internet Technology and Secured Transactions (ICITST)*: 1–7.
- Brown, Andy. "User Testing Subtitles for 360° Content." BBC Research and Development. October 26, 2017. <http://www.bbc.co.uk/rd/blog/2017-10-subtitles-360-video-virtual-reality-vr>
- Caffrey, Colm. 2012. "Using an Eye-Tracking Tool to Measure the Effects of Experimental Subtitling Procedures on Viewer Perception of Subtitled AV Content," In E. Perego, (ed). *Eye Tracking in Audio-Visual Translation*: 223–58. Roma: Aracne.
- Diaz Cintas, Jorge and Aline Remael. 2007. *Audiovisual Translation: Subtitling*. London: St. Jerome
- Fox, Wendy. 2016. "Integrated titles: An improved viewing experience?" In Silvia Hansen-Schirra & Sambor Gruzca (eds.) *Eyetracking and Applied Linguistics*: 5–30. Berlin: Language Science Press.
- Gambier, Yves. 2003. "Screen Translation". Accessed February 27, 2018. https://books.google.jo/books/about/Screen_Translation.html?id=esHsCwAAQBAJ.
- Gambier, Yves. 1996. *Les Transferts Linguistiques dans Les Medias Audiovisuels*. Paris : Presses Universitaires du Septentrion.
- Gottlieb, Henrik. 1994. "Subtitling: Diagonal Translation." *Perspectives: Studies in Translatology*. 2(1): 101–21. DOI:10.1080/0907676X.1994.9961227.
- Huey, Edmund Burke. 1907. "Hygienic Requirements in the Printing of Books and Papers." In *Popular Science Monthly* (70). https://en.wikisource.org/wiki/Popular_Science_Monthly/Volume_70/June_1907/Hygienic_Requirements_in_the_Printing_of_Books_and_Papers <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5678025>
- Kiefer Peter, Ioannis Giannopoulos, Martin Raubal & Andrew Duchowski. (2017). "Eye tracking for spatial research: Cognition, computation, challenges". *Spatial Cognition & Computation*, 17:1–2, 1–19, DOI: 10.1080/13875868.2016.1254634
- Kruger, Jan Louis, Agnieszka Szarkowska, and Izabela Krejtz. 2015. "Subtitles on the Moving Image: An Overview of Eye Tracking Studies." *Refractory: a Journal of Entertainment Media*, 25. <http://refractory.unimelb.edu.au/2015/02/07/kruger-szarkowska-krejtz>.
- Kruger, Jan-Louis, and Faans Steyn. 2014. "Subtitles and Eye Tracking: Reading and Performance." *Reading Research Quarterly* 49 (1): 105–20. DOI: 10.1002/rrq.59.
- Kruger, Jan-Louis, Esté Hefer, and Gordon Matthew. 2013. "Measuring the Impact of Subtitles on Cognitive Load: Eye Tracking and Dynamic Audiovisual Texts." In *Proceedings of Eye Tracking South Africa, Cape Town*: 62–66. DOI: 10.1145/2509315.2509331
- Kruger, Jan-Louis. 2012. "Making Meaning in AVT: Eye Tracking and Viewer Construction of Narrative," *Perspectives*, 20(1): 67–86. Accessed September 26, 2017. DOI: 10.1080/0907676X.2011.632688.
- McClarty, Rebecca. 2011. "Towards a Multiperspectival Approach in Creative Subtitling." *Mon TI* 4. (2012): 133–53. Accessed January 19, 2018. http://rua.ua.es/dspace/bitstream/10045/26944/1/MonTI_04_07.pdf.

- Rashbass, Cyril. 1961. "The Relationship Between Saccadic and Smooth Tracking Eye Movements." *Journal of Physiology*, 159 (2): 326–38. <https://doi.org/10.1113/jphysiol.1961.sp006811>
- Rayner, Keith. 1998. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin*, 124(3): 372–422. DOI: 10.1037//0033-2909.124.3.372.
- Salvucci, Dario D., and Joseph H. Goldberg. 2000. "Identifying Fixations and Saccades in Eye Tracking Protocols." In *Proceedings of the Eye Tracking Research and Applications Symposium*: 71–78. New York: ACM Press. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.68.2459&rep=rep1&type=pdf>
- Smith, Tim J. 2015. "Read, Watch, Listen: A Commentary on Eye Tracking and Moving Images." *Refractory: A Journal of Entertainment Media*, 25(9). <http://refractory.unimelb.edu.au/2015/02/07/smith>